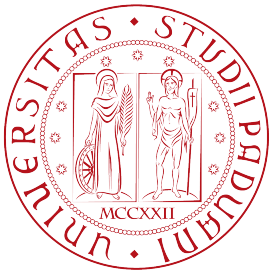


Università degli Studi di Padova

DIPARTIMENTO DI FISICA ED ASTRONOMIA
SCUOLA DI DOTTORATO DI RICERCA IN FISICA
CICLO XXX

Search for resonances in b-jets final states with the CDF II and the LHCb experiments

---

Author:
Emanuele MICHIELIN

Supervisor:
Prof. Donatella LUCCHESI

Director of the Doctoral School:
Prof. Gianluigi DALL'AGATA

Work prepared in collaboration with
Istituto Nazionale di Fisica Nucleare

Universitá degli Studi di Padova

Abstract

Doctor of Philosophy

Search for resonances in b-jets final states with the CDF II and the LHCb experiments

by Emanuele MICHIELIN

This thesis describes the search for resonances decaying into a pair of b-jets performed with the datasets collected by the CDF II experiment at the Tevatron and the LHCb at the LHC. The unique characteristics of these particle detectors, which are highlighted through the thesis with a constant comparison between the two experiments, give the opportunity to search also in the low dijet invariant mass region, unaccessible to the other general purpose detectors at the LHC.

A dataset corresponding to 5.4 fb^{-1} of data collected at CDF II is used to measure the cross section of the inclusive $Z \to b\bar{b}$ process, which turns out to be in agreement with the theory prediction, together with the data-simulation energy scale factor for b-jets. With the same event selection and background modeling then, the limit on inclusive $H \to b\bar{b}$ process is set.

At this point, the study on inclusive $H \to b\bar{b}$ channel is extended with a sensitivity study by using 1.6 fb^{-1} of data collected in 2016 at the LHCb. The future prospects for this search are reported.

Finally, the search for a beyond Standard Model Higgs-like particle produced in association with b-jets and decaying into a pair of b-jets, with the CDF II dataset, is described. No evidence of its production has been seen in the data, hence is set a 95% Confidence Level upper limit on its production cross section as a function of the particle invariant mass in the range from 100 to 300 GeV/c^2.

Contents

List of Figures

List of Tables

Introduction

The discovery of a new scalar particle consistent with the Higgs boson, performed by the ATLAS and CMS experiments [1, 2] at the LHC, has completed the particle zoology of the Standard Model. But there are still many open questions in the particle Fccs which answers implied the introduction of new physics phenomena. The dark matter, the matter-antimatter asymmetry, the naturalness of the hierarchy of mass scales, the neutrino masses are just some of the puzzles that LHC is trying to address [3]. So far, however, there are no clear indications about the theoretical solutions to these problems, nor the experimental strategies to resolve them.

The reconstruction and identification of the spray of particles produced in the fragmentation of a quark or a gluon, the so called jet, and the consequent search for resonances decaying into a pair of jets, gives access to one of the most important channel for the direct search of new physics phenomena. In particular, the possibility to identify and to select b quark initiated jets, the b-jets, allows to increase the signal significance for resonances with a strong coupling to b quarks, such as the Higgs boson.

The main difficulty for these searches arises from the overwhelming background coming from the production of jets from quantum chromodynamics processes. At LHC, the production of jets is so copious that the inclusive searches for resonances decaying to b-jets at the general purpose experiments can look for enhancement in the dijet mass up to 5 TeV/c^2, but at low masses (< 500 GeV/c^2) these searches are limited by the high rate of the jet triggers. Recently, some analyses have started investigating also the region below 500 GeV/c^2 by exploiting events where the two jets coming from the resonance are produced in the boosted regime and they merge in a single fat jet [4].

In this thesis a search for resonances within and beyond the Standard Model, decaying into a $b\bar{b}$ pair in the low invariant mass region is described. The datasets analyzed have been collected at two experimental facilities: the CDF II at the Tevatron and the LHCb at the LHC. The CDF II detector stopped taking data in late 2011, after an incredibly successful campaign lasted more than 20 years. LHCb instead has just began its data taking, considering that is has just collected about 5 fb^{-1} of the 300 fb^{-1} planned. The two have been originally designed for different goals, CDF II is a central general purpose detector while LHCb is a detector specialized for the study of the flavour physics in the forward region. But through all the thesis, the two detectors are compared, showing that the common peculiarities of the two experiments can be exploited. The technologies and the knowledge developed at CDF II, reoptimized and improved can enhanced the physics reach of LHCb, which can profit from a more performing detector and accelerator complex, giving access to a complementary physics to the other LHC experiments ATLAS and CMS.

In particular, the ability to reconstruct secondary vertices at the trigger level, signature of the production of a b hadron inside the jet, was developed for the first time at a hadron collider at CDF II and this gave the opportunity to collect a dataset pure in b-jets keeping low the energy threshold for the jets. At LHC, LHCb has recently developed a jet reconstruction and a heavy flavour tagging algorithms. LHCb is

unique at LHC because it can profit from a leveled luminosity, which leads to less background due to low pile up, and from a very flexible and performing trigger system. Since Run II, new trigger lines for the on-line reconstruction and identification of heavy flavour jets have been implemented with low jet energy threshold. For these reasons, LHCb is the perfect experiment at the LHC for the search of resonance decaying into heavy quarks in the low invariant mass region. In Chapter 2, the CDF II and the LHCb detectors are described. In Chapter 3 the b-jet on-line and off-line reconstruction and identification algorithms for the two detectors are described and compared.

Chapter 4 reports the challenging measurement of the $Z \to b\bar{b}$ cross section at CDF II, which is fundamental to validate the b-jet reconstruction techniques and the whole analysis procedure. The same decay channel has also been studied at LHCb, and a summary of this analysis is also presented. Chapter 5 describes the limit to the Standard Model inclusive $H \to b\bar{b}$ process evaluated both with the CDF and the LHCb datasets. The inclusive mode can be sensitive to any beyond Standard Model phenomena. At the end of the chapter, the future prospects for this analysis at LHCb is reported.

Finally, the search for a Higgs-like particle, decaying into a $b\bar{b}$ pair and produced in association with b-quarks at CDF II is described. Many extensions of the Standard Model foreseen new scalar particles with strong coupling to the b quarks, and this search is of particular interested because of a previous Tevatron combined result which reports a deviation at the level of 2σ from the Standard Model expectations in the $100 - 150$ GeV/c^2 invariant mass range [5].

Chapter 1

The Standard Model of Particle Physics

The Standard Model (SM) of particle physics is a theory which incorporates both quantum mechanics and Einstein's theory of special relativity in the attempt to describe the fundamental particles and their interactions. This is a model based on symmetries. As the Noether's theorem states [6], if an action is invariant under some group of transformations (symmetry), then there exist one or more conserved quantities (constants of motion) which are associated to these transformations. In this sense, Noether's theorem establishes that symmetries imply conservation laws.

The gauge principle invented by Salam and Ward [7] generalizes this idea: by imposing to a given Lagrangian the invariance under a certain symmetry, it is possible to determine the form of the interaction among the particles.

The Standard Model has been built based on this principle, it is a gauge theory based on the local gauge symmetry of the fundamental interactions in particle physics:

$$SU(3)_C \times SU(2)_L \times U(1)_Y . \tag{1.1}$$

The C is a reminder that $SU(3)$ represents the symmetry group of the colored strong interactions of quantum chromodynamics (QCD). The L indicates that the $SU(2)$ group applied to left-handed weak isospin doublets and the Y is a reminder that the $U(1)$ group contains the weak hypercharge singlets. Together, the $SU(2)_L \times U(1)_Y$ groups describe the unified electroweak force.

The search for resonances decaying into b quarks gives the possibility both to explore and test the prediction of the Standard Model and to search for new particles manifestation of a new theory. This chapter first briefly describes the model behind the Z and the Higgs bosons, the Electroweak theory, together with the motivations for a new physics beyond the Standard Model. The physics environment which characterizes the hadron colliders, dominated by the quantum chromodynamics interactions where the resonances are produced in the hard interaction between proton-antiproton or proton-proton, is then described. Finally, a brief review on the measurements performed of the properties of the Z and Higgs boson is reported.

1.1 Electroweak theory

The Electroweak (EW) theory [8, 9, 10] unifies the electromagnetic and the weak interactions as a manifestation of the same force. It is based on the $SU(2)_L \times U(1)_Y$ gauge group.

The electroweak Lagrangian is given by:

$$\mathcal{L} = -\frac{1}{4}W_a^{\mu\nu}W^a_{\mu\nu} - B^{\mu\nu}B_{\mu\nu} + \bar{\psi}i\gamma^\mu D_\mu\psi \tag{1.2}$$

where ψ are the fermion spinors, D_μ is the gauge covariant derivative:

$$D_\mu = \partial_\mu + igW^a_\mu T^a + \frac{1}{2}ig'B_\mu Y, \tag{1.3}$$

with T the weak isospin operator, g and g' are two different electroweak coupling constants and $B_{\mu\nu}$, similar to the electromagnetic field tensor, is given by:

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \tag{1.4}$$

where B_ν is the massless gauge field representing the singlet of $U(1)_Y$. W_μ are the gauge fields of $SU(2)$ and $W^a_{\mu\nu}$, with $a = 1, 2, 3$, the field tensor which is defined as

$$W^a_{\mu\nu} = \partial_\mu W^a_\nu - \partial_\nu W^a_\mu + g\epsilon^{abc}W^b_\mu W^c_\nu \tag{1.5}$$

Experimentally the weak interaction is characterized by a short range, therefore its carriers must have mass. Moreover, the $SU(2)_L \times U(1)_Y$ symmetry does not allow to introduce mass terms for quarks and leptons in a simple way. The problem of the masses in the SM is overcome by means of a electroweak symmetry breaking (EWSB) mechanism [11], which provides a general framework to keep untouched the structure of these gauge interactions at high energies and still generate the observed masses of the W and Z gauge bosons. The masses of all fermions are also a consequence of EWSB since the Higgs doublet couples to the fermions through Yukawa interactions.

1.1.1 The electroweak symmetry breaking

In the Standard Model the electroweak symmetry breaking is achieved adding a new term to the Lagrangian:

$$\mathcal{L}_\Phi = (D_\mu\Phi)^\dagger D^\mu\Phi - V(\Phi) \tag{1.6}$$

where the Higgs field Φ is a self-interacting $SU(2)$ complex doublet (four real degrees of freedom) with weak hypercharge $Y = 1$:

$$\Phi = \frac{1}{\sqrt{2}}\begin{pmatrix}\sqrt{2}\phi^+ \\ \phi^0 + ia^0\end{pmatrix} \tag{1.7}$$

where ϕ^0 and a^0 are the CP-even and CP-odd neutral components, and ϕ^+ is the complex charged component of the Higgs doublet. $V(\Phi)$ is the scalar potential,

$$V(\Phi) = \mu^2\Phi^\dagger\Phi + \lambda(\Phi^\dagger\Phi)^2, \tag{1.8}$$

with μ and λ real constants. $V(\Phi)$ is the most general renormalizable scalar potential and for $\mu^2 < 0$ and $\lambda > 0$ the "mexican hat" potential is obtained: it is characterized by having a continuum set of field configurations minimizing the energy of the system, see Figure 1.1. With this potential the neutral component of the scalar doublet acquires a non-zero vacuum expectation value (VEV) $v \simeq (\sqrt{2}G_F)^{-\frac{1}{2}} \simeq 246$ GeV, derived from the Fermi constant G_F measured in muon decays [12]:

$$\langle\Phi\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}0 \\ v\end{pmatrix}, \tag{1.9}$$

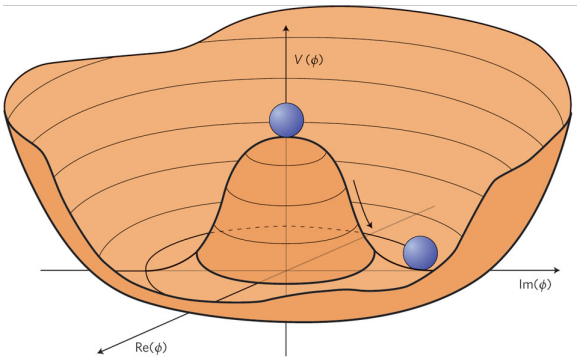


FIGURE 1.1: Shape of the Higgs potential for $\mu^2 < 0$.

with $\phi^0 = H + \langle\phi^0\rangle$ and $\langle\phi^0\rangle \equiv v$, inducing the spontaneous breaking of the SM gauge symmetry $SU(3)_C \times SU(2)_L \times U(1)_Y$ into $SU(3)_C \times U(1)_{em}$. The spontaneous symmetry breaking implies that there is a symmetry of the system (Lagrangian) that is not respected by the ground state. The Higgs field permeates the entire universe and through its self-interactions can cause spontaneous electroweak symmetry breaking in the vacuum.

The Higgs field couples to the W_μ and B_μ gauge fields associated with the $SU(2)_L \times U(1)_Y$ local symmetry through the covariant derivative appearing in the kinetic term of the Equation 1.6, where D_μ is the covariant derivative defined in Equation 1.3. As a result, the neutral and the two charged massless Goldstone degrees of freedom mix with the gauge fields corresponding to the broken generators of $SU(2)_L \times U(1)_Y$ and become the longitudinal components of the Z and W physical gauge bosons, respectively. The Z and W gauge bosons acquire masses,

$$M_W^2 = \frac{g^2 v^2}{4}, \quad M_Z^2 = \frac{(g'^2 + g^2)v^2}{4}. \tag{1.10}$$

The fourth generator remains unbroken since it is the one associated to the conserved $U(1)_{em}$ gauge symmetry, and its corresponding gauge field, the photon, remains massless. Therefore, from the initial four degrees of freedom of the Higgs field, two are absorbed by the $W^\pm$ gauge bosons, one by the Z gauge boson, and there is one remaining degree of freedom, H, that is the physical Higgs boson.

This SM Higgs boson then is a CP-even scalar of spin 0, which mass is given by

$$M_H^2 = 2\lambda v^2. \tag{1.11}$$

where λ is the self coupling parameter in $V(\Phi)$ of Equation 1.8. This quartic coupling λ is a free parameter in the Standard Model, and hence, there is no a priori prediction for the Higgs mass that has to be determined experimentally.

The Higgs boson couplings to the fundamental particles are set by their masses. Therefore, the interaction is very weak for light particles, such as up and down quarks, and electrons, but strong for heavy particles such as the W and Z bosons and the top quark. More precisely, the SM Higgs couplings to fundamental fermions are linearly proportional to the fermion masses, whereas the couplings to bosons are proportional to the square of the boson masses.

As a result, the dominant mechanisms for Higgs boson production and decay involve the coupling of H to W, Z and/or the third generation quarks and leptons. This is the reason why the $b\bar{b}$ decay mode studied in this thesis, for a 125 GeV/c^2 Higgs, is the one with the higher branching fraction. The Higgs boson coupling to

photons is generated via loops, where the one-loop graph with a virtual W^+W^- pair provides the dominant contribution and the one involving a virtual tt pair is subdominant.

1.1.2 Is the Standard Model the end of the story?

With the discovery of the Higgs boson [1, 2], it has been experimentally established that the Standard Model is based on a gauge theory that could a priori be consistently extrapolated to the Planck scale ($M_P \sim 10^{19}$ GeV). The Higgs boson must have couplings to W/Z gauge bosons and fermions precisely as those in the SM to maintain the consistency of the theory at high energies, hence, formally there is no need for new physics at the EW scale. However, as the SM Higgs boson is a scalar particle, and therefore without a symmetry to protect its mass, it has sensitivity to the physics in the ultraviolet. Quite generally, the Higgs mass parameter m may be affected by the presence of heavy particles. The running of the mass parameter from the scale μ to the scale Q reads

$$m(Q^2) = m^2(\mu) + \delta m^2, \tag{1.12}$$

with:

$$\delta m^2 = \sum_i g_i(-1)^{2S_i} \frac{\lambda_i^2 m_i^2}{32\pi^2} log(\frac{Q^2}{m^2}) \tag{1.13}$$

where the sum is over all particles and g_i and S_i correspond to the number of degrees of freedom and the spin of the particle i. This means that light scalars like the Higgs boson cannot naturally survive in the presence of heavy states at the grand-unification, string or Planck scales. This is known as the hierarchy problem [13].

Besides the hierarchy problem, there are a different experimental observations that lead to the conclusion that the Standard Model cannot be a complete theory. For instance, the observation of the flavor oscillation of the neutrinos observed by the neutrino experiments is the evidence for a non-zero neutrino mass, which is not expected in the SM; the baryonic asymmetry, i.e. the imbalance of matter (baryons) and antimatter (antibaryons) in the observed universe, cannot be described within the SM; cosmological measurements show that the Universe is mostly made of components that are not described by this model, called Dark Matter and Dark Energy.

Different models have been developed to solve these problems, and many of these foreseen new particles with a strong coupling to the third quark generation. This is the reason why the search for new resonance decaying into $b\bar{b}$ pair is so important and, in the searched presented in Chapter 6, is let model independent. However, the more popular class of models which tries to address all these problems is based on a new fermion-boson symmetry in nature called supersymmetry [14, 15].

This is a weakly coupled approach to EWSB, and in this case, the Higgs boson remains elementary and the corrections to its mass are screened at the scale at which SUSY is broken and remain insensitive to the details of the physics at higher scales. These theories predict at least three neutral Higgs particles and a pair of charged Higgs particles [16]. One of the neutral Higgs bosons, most often the lightest CP-even Higgs, has properties that resemble those of the SM Higgs boson. Under certain hypotheses on some parameters of the model (m, $\tan\beta$), the couplings of the new neutral Higgs boson to bottom quarks are enhanced.

1.2 Physics at hadron colliders

The searches carry out in this thesis use the data collected at two hadron colliders: the Tevatron and the LHC, where proton-antiproton and proton-proton are collided at high energies. In these collisions, the interactions that occur between the "partons" of the hadrons are described by the Quantum Chromodynamics theory. Also the final state that is used to reconstruct the resonances, the jets, as well as the main background for the analyses, are pure manifestations of Quantum Chromodynamics processes.

The Quantum Chromodynamics (QCD) is a non-Abelian gauge theory based on the gauge group $SU(3)$ and it is the part of the Standard Model that describes the strong interactions binding quarks in hadrons. Quarks are the elementary constituents of the hadronic matter, and they are the only SM elementary particles which experience the strong force. All observed hadrons are either mesons (quark-anti-quark bound states) or baryons (bound states of three quarks).

The QCD Lagrangian can be written as:

$$\mathcal{L} = \sum_q \bar{\psi}_i^q (i\gamma_\mu D_{ij}^\mu - m_q \delta_{ij}) \psi_j^q - \frac{1}{4} F_{\mu\nu}^a F_a^{\mu\nu} \tag{1.14}$$

where the quark (anti-quark) fields are denoted by ψ_i^q ($\bar{\psi}_i^q$) indexed according to the color with i or j and flavor q, m_q is the mass of the quark which is a free parameter of the theory, γ^μ is the Dirac matrix which expresses the vector nature of the strong interaction, $F_a^{\mu\nu}$ is the gluon field strength tensor for a gluon with color index a, and D_{ij}^μ is the so-called covariant derivative,

$$D_{ij}^\mu = \partial_\mu \delta_{ij} + i g_s t_{ij}^a A_a^\mu \tag{1.15}$$

where g_s is the strong coupling constant, A_a^μ the gluon field with color index a and t_{ij}^a the generators of $SU(3)$.

The coupling constant of QCD (α_s) can be defined from the strong coupling constant g_s by $g_s^2 = 4\pi\alpha_s$. Since gluons carry color charge, besides fermionic (quark) loops, gluonic loops are also present. Calculations beyond the tree level, involving gluon and quark internal loops, diverge logarithmically with coefficients of opposite sign. To handle such divergences a renormalization procedure is necessary and a renormalization scale μ, related to the momentum transfer Q in a given process, is introduced. As a consequence the strong coupling constant α_s becomes a function of the scale of the process i.e., a running constant, and it is defined, at one-loop approximation, by the equation:

$$\alpha_s(Q^2) = \frac{12\pi}{33 - 2n_f} \frac{1}{log(Q^2/\lambda^2)}, \tag{1.16}$$

where $\lambda \sim 0.1\,\text{GeV}$ indicates the lower energy limit at which SM should be considered valid (as approximation of more extended theories) and n_f is the number of quark flavors whose mass is greater than the Q^2 of interest [17].

1.2.1 Asymptotic freedom and confinement

The running constant for the strong interactions, as shown in Figure 1.2, is significant at low Q^2 and decreases as Q^2 increases. This property, whereby the strong coupling becomes small for "hard" (i.e., high-momentum transferred processes) processes, is known as asymptotic freedom whereas the increase of the strong coupling constant

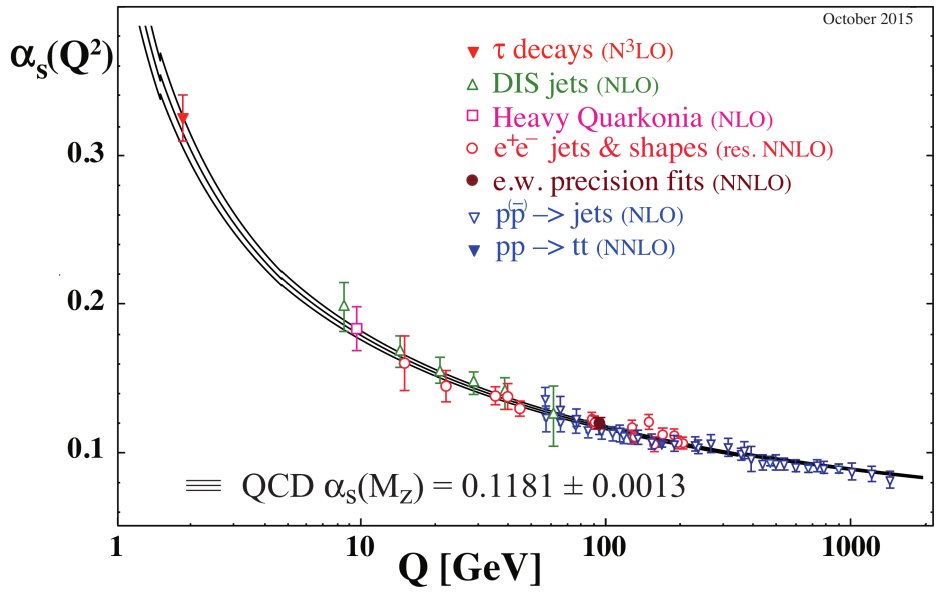


FIGURE 1.2: Summary of measurements of α_s as a function of the energy scale Q, in which the reference scale is chosen to be $Q^2 = M_Z^2$ [12].

with decreasing momentum transfer (i.e., increasing distance), is referred to as "confinement". This behavior is also demonstrated by Eq. 1.16 for $Q \to \infty$ and $Q \to \lambda$.

As a result this confining propriety, with the exception of the top quark which decays before it has time to hadronize, free quarks are not observed since they hadronized on a time scale $\sim 1/\Lambda_{QCD}$. This implies that only colorless (color-singlet) hadronic states can be observed in nature. The low-energy regime is described by lattice gauge theories which provides a way to compute the hadron mass spectrum directly from the QCD Lagrangian [18]. On the other hand, as a consequence of the asymptotic freedom, the perturbation approach becomes applicable for high momentum transfers, where the coupling is weak.

1.2.2 QCD in hadronic collisions

The use of perturbation theory in QCD calculations is made possible by the feature of asymptotic freedom. If a process involves a large momentum transfer, then the running coupling constant α_s may be small enough to justify the use of perturbative techniques. When pQCD can be applied, the factorization theorem states that the cross section of any QCD process can be written as the convolution of basic building blocks such as the quark and gluon distributions in the incoming hadrons, the hard subprocesses describing the large-angle scattering of partons (almost free collinear point-like particles moving in the same direction of the hadron and sharing its momentum), and the fragmentation functions of quarks and gluons into hadrons. This means that any QCD process can be written as the convolution of its high-energy (perturbative QCD) and low-energy (non-perturbative QCD) components [19].

At an hadron collider, where the momentum transfers is very high, the interaction between the hadrons is modeled with the parton model. The model describes the fast moving hadrons as a beam of almost free collinear point-like particles named partons moving in the same direction of the hadron and sharing its momentum so

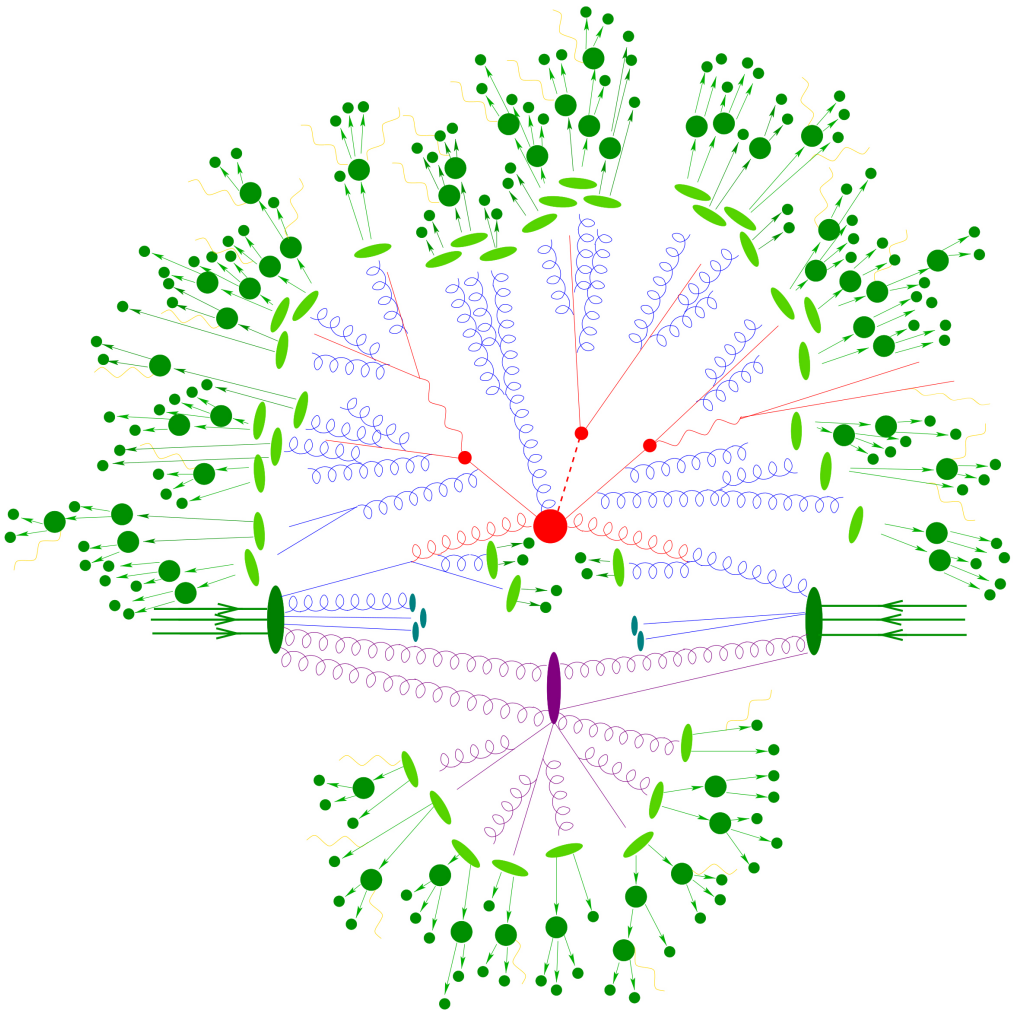


FIGURE 1.3: Representation of a high-energy hadron collision.

that the fraction of the hadron momentum (p) carried by the parton i is:

$$p_i = x_i p, \tag{1.17}$$

where x_i represents the fraction of the hadron momentum carried by the parton.

In this framework, the cross section for the production of X in a proton-antiproton interaction at high energy in a hadron collider can be then described as:

$$\sigma_{p\bar{p}} \to X(s) = \sum_{ij} \int_0^1 dx_1 dx_2 f_i^p(x_1, \mu_F) f_j^{\bar{p}}(x_2, \mu_F) \sigma_{ij\to X}(x_1, x_2, s, \alpha_s(\mu_F)). \tag{1.18}$$

where $\sigma_{ij\to X}$ is the partonic hard scattering and $f_i^p(x_1, \mu_F)$ are the parton distributions functions (PDFs) which are the probability densities to find a parton i with fractional momentum x_i in the proton (or antiproton). The factorization scale μ_F is the value that distinguishes between the two phases, where the perturbative QCD and where the non-perturbative QCD are applied. The same goes in a proton-proton collision.

Figure 1.3 represents how an high energy hadron collision is simulated in the SHERPA [20] Monte Carlo event generator. We can exploit this picture to visualize the different stages of an event in a hadron collider. Matrix element generators simulate the hard part of scattering, shown in red. This stage can be described with perturbative QCD calculations using Feynman diagrams, here the electroweak bosons or the eventual beyond Standard Model particle are produced. The complexity and the number of the diagrams increases rapidly with the number of particles involved.

As the transverse momenta of the shower become smaller and the running coupling constant becomes stronger a progressively larger number of diagrams are involved and the perturbative calculations become unreliable. Moreover, fixed order predictions diverge for soft gluon emission and collinear splitting. This stage is described using a approximate description of the QCD dynamics, the parton shower (PS), and is represented in Figure 1.3 in blue.

The particles in the showers have in average a momentum transfer-square lower then 1 GeV^2, which means that the physics involved is non perturbative and pertains to hadronization, also called fragmentation (represented in green in Figure 1.3). At this stage the partons group into high-mass color-neutral states which subsequently decay into the final-state particles. The tight cone of particles created by the hadronization of a single quark or gluon is called a jet. In particle detectors, jets are observed rather than quarks, whose properties must be inferred. The hadronization is considered a local process, thus the observable hadrons in the final state carry the kinematic and flavor informations of the original partons.

The purple, bottom part of Figure 1.3 shows the secondary interactions between the partons not involved in the primary interaction. They are not negligible and their contribution to the final state of the hard scattering process is known as the underlying event (UE).

Multijet backgrounds

The main background for the searches described in this thesis is the multijet production of heavy flavour, which arises from a large number of production mechanisms [82] for which the rates are not precisely known.

The heavy quark production can be categorized into three types of processes: flavor creation, flavor excitation and gluon splitting. Figure 1.4 shows the Feynman diagrams concerning these processes for the b quarks production.

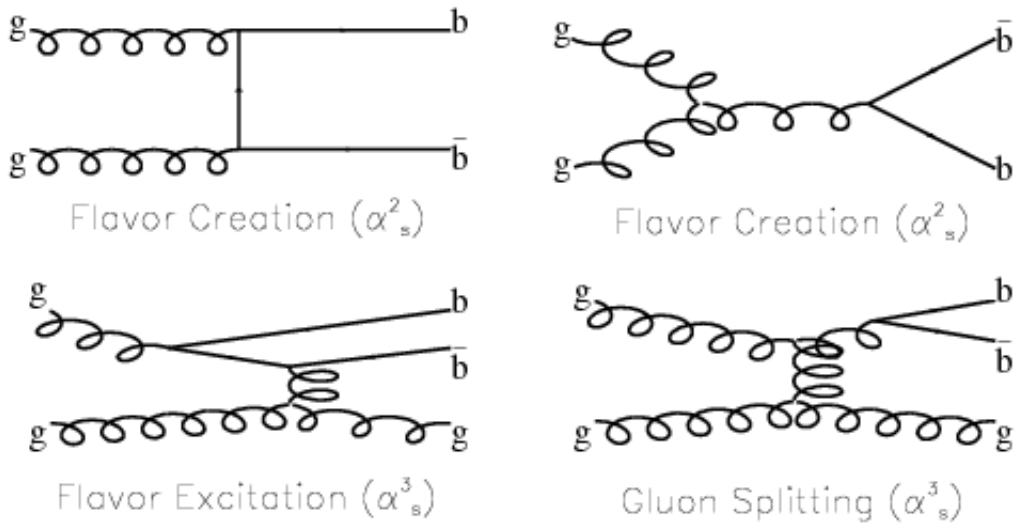


FIGURE 1.4: Main Feynman diagrams of the physics processes that contribute to the bottom production.

Flavor creation, $q\bar{q} \rightarrow b\bar{b} + X$, refers to the lowest-order QCD $b\bar{b}$ production diagrams. This process includes $b\bar{b}$ production through $q\bar{q}$ annihilation and gluon fusion, plus higher-order corrections to these processes. Because this production is dominated by two-body final states, it tends to yield $b\bar{b}$ pairs that are back-to-back in $\Delta\phi$ and balanced in p_{T}.

Flavor excitation, $bq \rightarrow bq + X$, refers to diagrams in which a $b\bar{b}$ pair from the quark sea of the proton or antiproton is excited into the final state because one of the quarks from the $b\bar{b}$ pair undergoes a hard QCD interaction with a parton from the other beam particle. Because only one of the quarks in the $b\bar{b}$ pair undergoes the hard scatter, this production mechanism tends to produce b quarks with asymmetric p_{T}. Often, one of the b quarks will be produced with high rapidity and not be detected in the central region of the detector.

Gluon splitting, $qg \rightarrow qg + X$ followed by $g \rightarrow b\bar{b}$, refers to diagrams where the $b\bar{b}$ pair arises from a gluon splitting in the initial or final state. Neither of the quarks from the $b\bar{b}$ pair participate in the hard QCD scatter. Depending on the experimental range of b quark p_{T} sensitivity, gluon splitting production can yield a $b\bar{b}$ distribution with a peak at small $\Delta\phi$.

1.3 Experimental results on the Z boson

With its direct observation at UA1 and UA2 experiment at CERN in 1983 as a few resonant events in the dielectron and dimuon mass spectrum [21], detailed studies of the Z properties began. Thanks to the measurements performed at e^+e^- colliders LEP and SLC, the Z boson properties and the underlying electroweak theory are known and tested to a precision below the per-mil level [12]. Among these, the overall e^+e^- production cross-section, the Z line-shape and width, the number of neutrino families, and the forward-backward asymmetries of the leptons and quarks.

The fit of the LEP and SLC data gave as results:

$$m_Z = 91187.5 \pm 2.1 \text{ MeV}/c^2 \tag{1.19}$$

$$\Lambda_Z = 2495.2 \pm 2.3 \text{ MeV}/c^2 \tag{1.20}$$

$$\sin\theta_{eff} = 0.23147 \pm 0.00016, \tag{1.21}$$

where $\sin\theta_{eff}$ is obtained from the $\sin\theta_W$ including the effect of radiative corrections.

Although the SLD experiment at the SLC collected much lower statistics, it was able to match the precision of LEP experiments in determining the effective electroweak mixing angle and in particular the rates of Z decay to b- and c-quarks, owing to availability of polarized electron beams, small beam size, and stable beam spot.

SLD measured the parity-violating parameter A_b by analyzing the left-right (back-forward) asymmetry of b quarks in $e^+e^- \rightarrow Z \rightarrow b\bar{b}$ events with different analysis techniques. Similarly was studied the asymmetry parameter A_c from $Z \rightarrow c\bar{c}$ decay. From data collected from 1993 to 1995, in a sample of 150000 Z the asymmetry measurement gave:

$$A_b = 0.910 \pm 0.068(stat) \pm 0.037(syst) \tag{1.22}$$

$$A_c = 0.642 \pm 0.110(stat) \pm 0.063(syst). \tag{1.23}$$

Figure 1.5 shows the electroweak observables related to the Z and W bosons that were precisely measured at LEP and SLC compared with the Standard Model predictions. The agreement with the Standard Model expectations is good for each observable, the only tension is the longstanding puzzle of the $b\bar{b}$ forward-backward asymmetry A^b_{FB}.

Measurements of the Z boson at the hadron colliders cannot compete with the precision measurements performed at the e^+e^- colliders, but the extraction of the Z cross section at different energies gives a strong complementary test of the electroweak theory. Efforts are ongoing also for a precise measurement of the $\sin\theta_W$. The Z decay to muon is very clean and it is also used to perform many studies on the performances of the detector, such as the trigger efficiency and muon momentum scale and resolution. The decay to quarks, and in particular to b quarks, provides a standard candle for searches in final states with a $b\bar{b}$ pair.

The decay to b quarks in addition, gives access to the forward-background asymmetry parameter A^b_{FB}. In case of a high statistic sample, this measurement would be of the primary interest also at a hadron collider because, being the Z-pole in the signal region for the $b\bar{b}$ asymmetry, there is the opportunity to be sensitive to a significant interference effect between New Physics and tree level Z exchange [22].

1.4 Experimental results on the Higgs boson

The Higgs boson is produced in the high energy collisions at the hadron colliders. The main production mechanisms at the Tevatron collider and the LHC are gluon fusion, weak-boson fusion, associated production with a gauge boson and associated production with a pair of top/antitop quarks. Figure 1.6 shows the Feynman diagrams for these Higgs production processes.

- Gluon fusion $gg \rightarrow H + X$ is the main production mechanism at hadron colliders and it is mediated by the exchange of a virtual, heavy top quark. Contributions from lighter quarks propagating in the loop are suppressed proportional to m_q^2. Including the full dependence on the (top, bottom, charm) quark and Higgs boson masses, the cross section has been calculated at the next-to-leading order (NLO) in α_s [23].

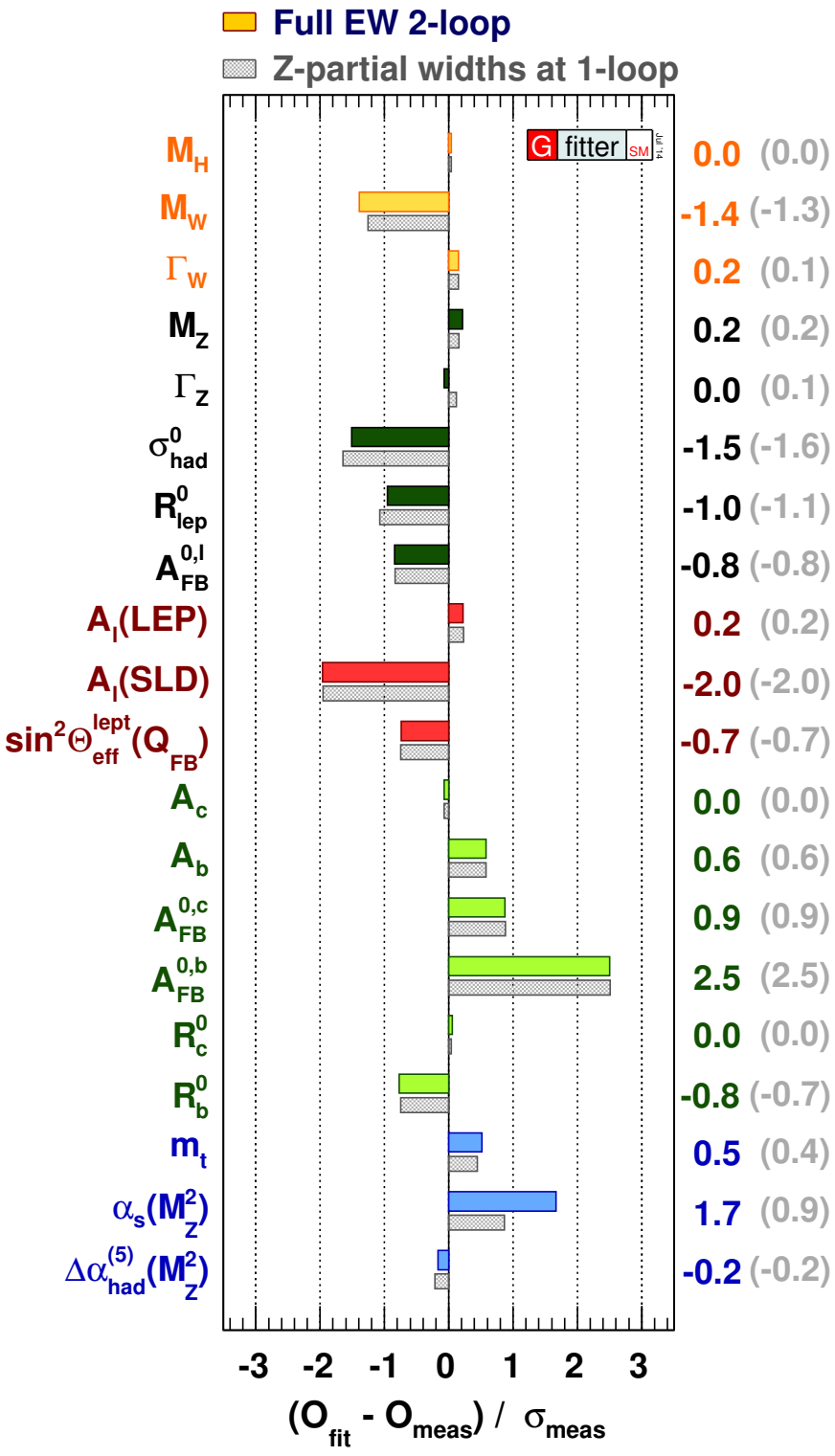


FIGURE 1.5: Comparing fit results with direct measurements: pull values for the SM fit, i.e. deviations between experimental measurements and theoretical calculations in units of the experimental uncertainty.

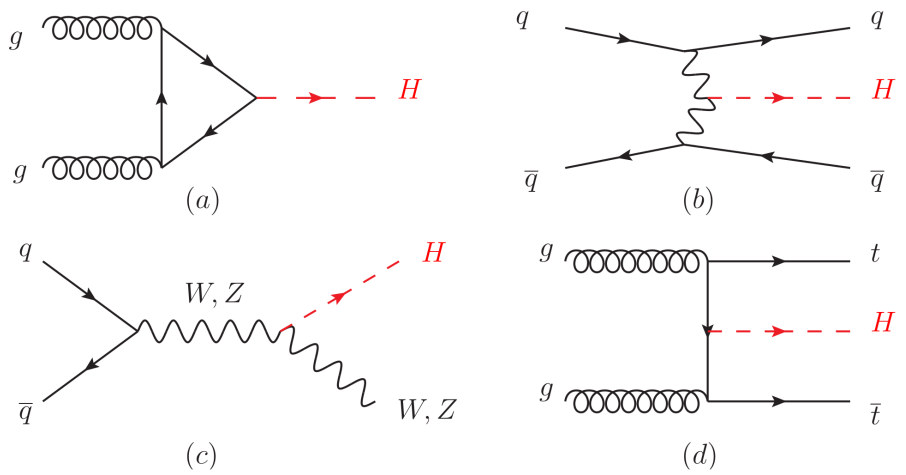


FIGURE 1.6: Feynman diagrams for the main Higgs production mechanisms: gluon fusion (a), vector-boson fusion (b), Higgs-strahlung and associated production with top quarks (c).

TABLE 1.1: Different mechanisms production cross section of the Higgs boson at the Tevatron and LHC.

$\sqrt{s}$ (TeV)	Production cross sections in pb					
	ggF	VBF	WH	ZH	$t\bar{t}H$	Total
1.96	$0.95^{+17\%}_{-17\%}$	$0.065^{+8\%}_{-7\%}$	$0.13^{+8\%}_{-8\%}$	$0.079^{+8\%}_{-8\%}$	$0.004^{+10\%}_{-10\%}$	1.23
7	$15.3^{+10\%}_{-10\%}$	$1.24^{+2\%}_{-2\%}$	$0.58^{+3\%}_{-3\%}$	$0.34^{+4\%}_{-4\%}$	$0.09^{+8\%}_{-14\%}$	17.5
8	$19.5^{+10\%}_{-11\%}$	$1.60^{+2\%}_{-2\%}$	$0.70^{+3\%}_{-3\%}$	$0.42^{+5\%}_{-5\%}$	$0.13^{+8\%}_{-13\%}$	22.3
13	$44.1^{+11\%}_{-11\%}$	$3.78^{+2\%}_{-2\%}$	$1.37^{+2\%}_{-2\%}$	$0.88^{+5\%}_{-5\%}$	$0.51^{+9\%}_{-13\%}$	50.6
14	$49.7^{+11\%}_{-11\%}$	$4.28^{+2\%}_{-2\%}$	$1.51^{+2\%}_{-2\%}$	$0.99^{+5\%}_{-5\%}$	$0.61^{+9\%}_{-13\%}$	57.1

- Vector boson fusion (VBF), $qq' \to qq'H$, is the production mode with the second-largest cross section at the LHC. At the Tevatron collider, VBF also occurred, but for $m_H = 125$ GeV/c^2 had a smaller cross section than Higgs production in association with a W or Z boson. Higgs production via VBF, proceeds by the scattering of two (anti-)quarks, mediated by t- or u-channel exchange of a W or Z boson, with the Higgs boson radiated off the weak-boson propagator. The scattered quarks give rise to two hard jets in the forward and backward regions of the detector.

- The next most relevant Higgs boson production mechanisms after gluon fusion and VBF at the LHC, and the most relevant ones after gluon fusion at the Tevatron collider, are associated production with W and Z gauge bosons $q\bar{q} \to VH + X$.

- Higgs radiation off top quarks, $gg \to t\bar{t}H$, can provide important information on the the top-Higgs Yukawa coupling and gives access to the Higgs decay into bottom quarks.

Table 1.1 gives a summary of the production cross section of the Higgs boson at the Tevatron in $p\bar{p}$ collision, and for different LHC center of mass energies in pp collisions.

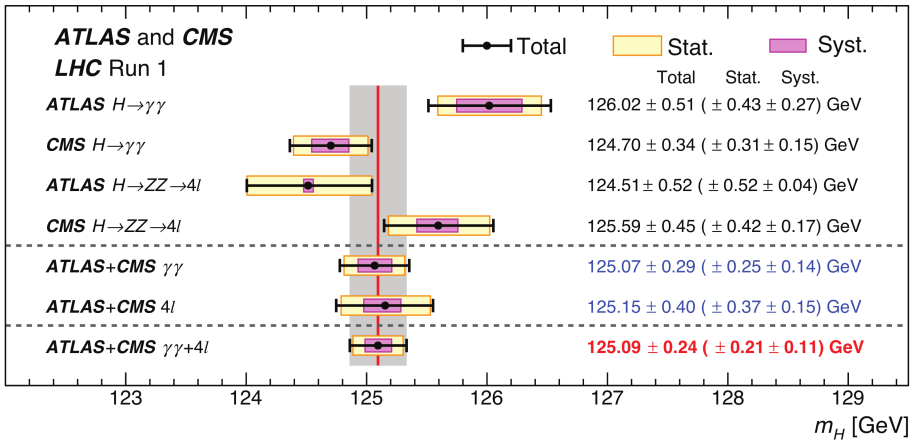


FIGURE 1.7: CMS and ATLAS mass measurements in the $\gamma\gamma$ and ZZ channels, the combined result from each experiment and their combination.

1.4.1 Measurements in the Higgs boson sector

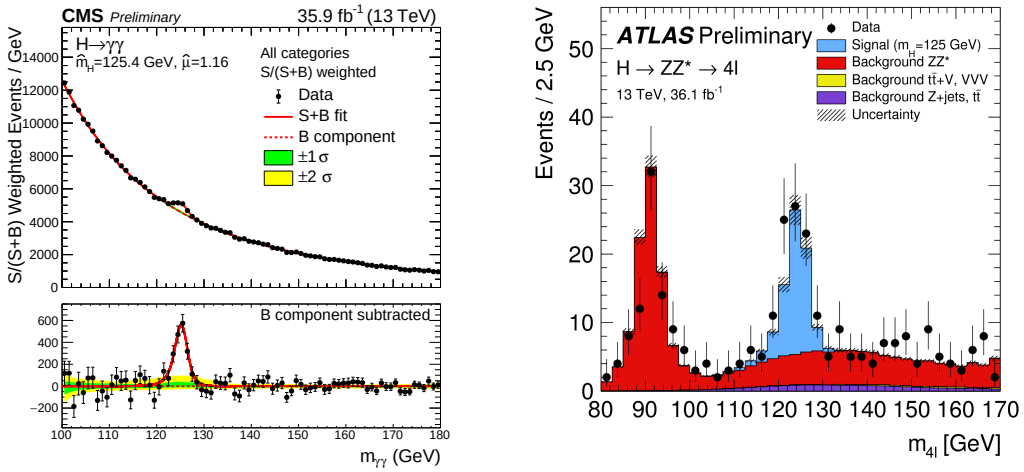
Given the low measured mass of the Higgs, which leads to a natural width of only a few MeV, five decay channels play an important role at the LHC. The sensitivity of a search channel depends on the production cross section of the Higgs boson, its decay branching fraction, reconstructed mass resolution, selection efficiency and the level of background in the final state. The $H \to \gamma\gamma$ and $H \to ZZ \to 4l$ channels, all final state particles can be very precisely measured and the reconstructed m_H resolution is excellent. While the $H \to W^+W^- \to l^+\nu_l l^{'-}\bar{\nu_{l'}}$ channel has relatively large branching fraction, the m_H resolution is poor due to the presence of neutrinos. The $H \to b\bar{b}$ and the $H \to \tau^+\tau^-$ channels have large branching ratio, but they suffer from large backgrounds and a poor mass resolution.

In the $\gamma\gamma$ channel, using the full Run 1 LHC dataset, ATLAS observes [24] an excess over background with a local significance of 5.2σ, while CMS observes [25] its largest excess with a local significance of 5.7σ. In the $H \to ZZ \to 4l$ channel the CMS experiment observes[26] its largest excess with an observed local significance of 6.8σ, while ATLAS observes [27] an excess with a local significance of 8.1σ. A combined measurement of the Higgs mass has been performed by the two experiments, which rely on the two high mass resolution and sensitive channels, $\gamma\gamma$ and ZZ. Figure 1.7 shows the different measurements [28], which combined lead to a mass of:

$$m_H = 125.09 \pm 0.21(stat.) \pm 0.11(syst.)\ \text{GeV}/c^2 \tag{1.24}$$

In 2015 the LHC Run 2 has started, with an increase in the center of mass energy from 8 to 13 TeV. The first measurements performed by CMS and ATLAS in the $\gamma\gamma$ and ZZ channels have been released [29, 30, 31, 32], Figure 1.8 shows the CMS $\gamma\gamma$ and the ATLAS ZZ invariant mass spectrum, with the clear Higgs peak around 125 GeV/c^2 in both channels. Using the new data a new, more precise, measurement of the Higgs mass will be performed.

The best channel for probing the coupling of the Higgs field to the quarks and leptons are $H \to b\bar{b}$ and $H \to \tau^+\tau^-$, respectively. For a Higgs boson with $m_H = 125$ GeV/c^2, the branching fraction to $b\bar{b}$ is about 57% and to $\tau^+\tau^-$ is about 6%. Nevertheless, the presence of very large backgrounds makes the isolation of a Higgs



(a) (b)

FIGURE 1.8: Invariant mass spectrum in the CMS $H \to \gamma\gamma$ (a) and ATLAS $ZZ \to 4l$ (b) analysis channels. The Higgs peak is visible around 125 GeV in both channels.

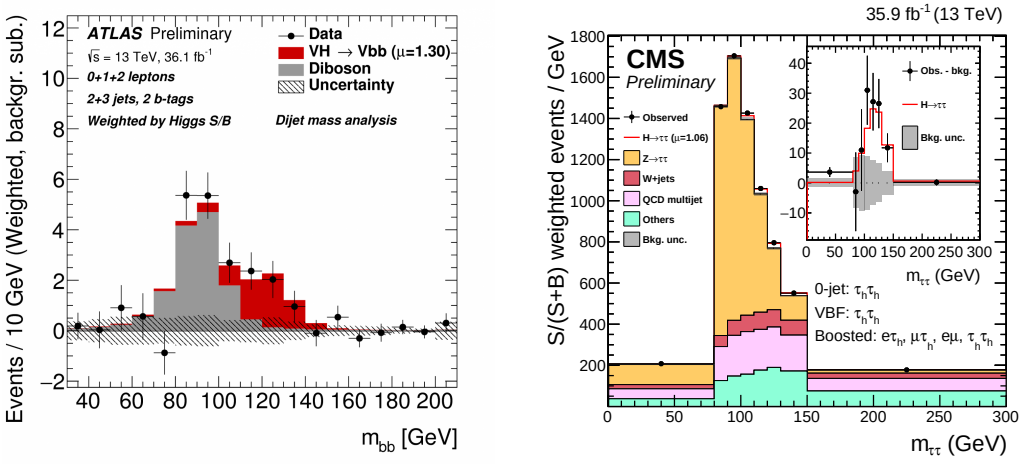
boson signal in these channels quite challenging. With Run 1 data, for H produced associated with a W or Z boson, with H decaying to a $b\bar{b}$ pair, the significance observed by ATLAS is 1.4σ [33], while for CMS it is 2.1σ [34], the combined significance being 2.6σ [35]. In the search for the $\tau^+\tau^-$ channel, ATLAS published a signal significance of 4.5σ [36] and CMS found a signal significance of 3.2σ [37].

Using the available Run 2 data, ATLAS announced a first evidence of the $H \to b\bar{b}$ decay in the VH channel, with a reported significance of 3.5σ [38]. CMS instead, reported the observation of the $H \to \tau^+\tau^-$ with a significance of 4.9σ [39], which rises to 5.9σ when combined with CMS Run1 analysis. Figure 1.9 shows the invariant mass distribution of the $b\bar{b}$ di-jet in the ATLAS $VH(b\bar{b})$ search after the background subtraction, where, besides the $H \to b\bar{b}$ events excess also the $VZ(b\bar{b})$ peak is visible. Figure 1.9 also shows the $\tau^+\tau^-$ invariant mass spectrum in the CMS analysis, which led to the observation of the $H \to \tau^+\tau^-$ decay mode.

Using Run 2 data, CMS has performed a first search for the inclusive $H \to b\bar{b}$ process in the high-p_T regime, where the Higgs is produced so boosted that the two b-jets are reconstructed merged in a single fat jet [4]. This analysis uses the $Z \to b\bar{b}$ decay (first observation in the boosted topology) as a standard candle to validate the background modeling. Figure 1.10 shows the invariant mass distribution of the fat jet, a $H \to b\bar{b}$ signal has been found with an observed significance of 1.5σ.

Also the LHCb experiment has started a program aiming to the measurement of the Higgs boson properties. The first search was performed setting upper limits in the associated production with a Z or a W boson, with the Higgs boson decaying into a b quark pair, and for the first time into a c quark pair [40].

At the Tevatron the CDF and D0 collaborations searched for the Higgs produced via gluon-gluon fusion, WH, ZH, $t\bar{t}H$ and vector boson fusion decaying in the $H \to b\bar{b}$, $H \to W^+W^-$, $H \to ZZ$, and $H \to \tau^+\tau^-$ modes [41]. They observed a total combined significance of 3.0σ at a Higgs mass of 125 GeV/c^2. Figure 1.11 shows the observed and median expected (for the background-only hypothesis) 95% C.L. Bayesian upper production limits expressed as multiples of the SM cross section as a function of Higgs boson mass for the combined CDF and D0 searches in all decay modes.



(a) (b)

FIGURE 1.9: Invariant mass distribution in the ATLAS $VH(b\bar{b})$, after background subtraction, (a) and CMS $H \to \tau^+\tau^-$ (b) analysis channels. An event excess due to the H decay is visible in both plots.

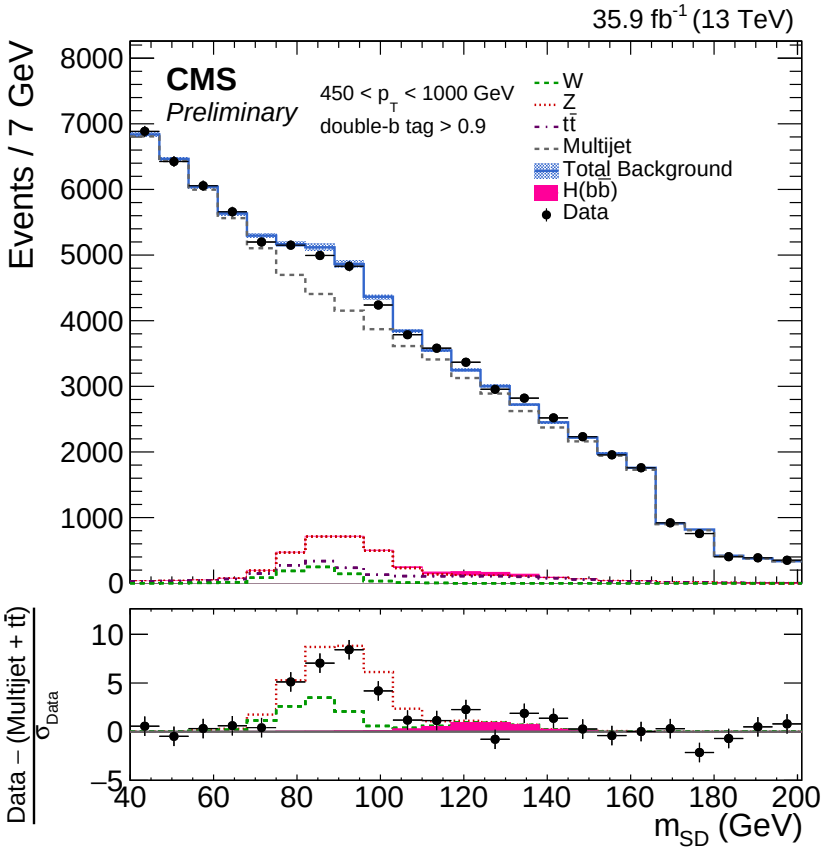


FIGURE 1.10: Invariant mass distribution of the fat jet in the CMS analysis of the inclusive boosted $H \to b\bar{b}$ signal.

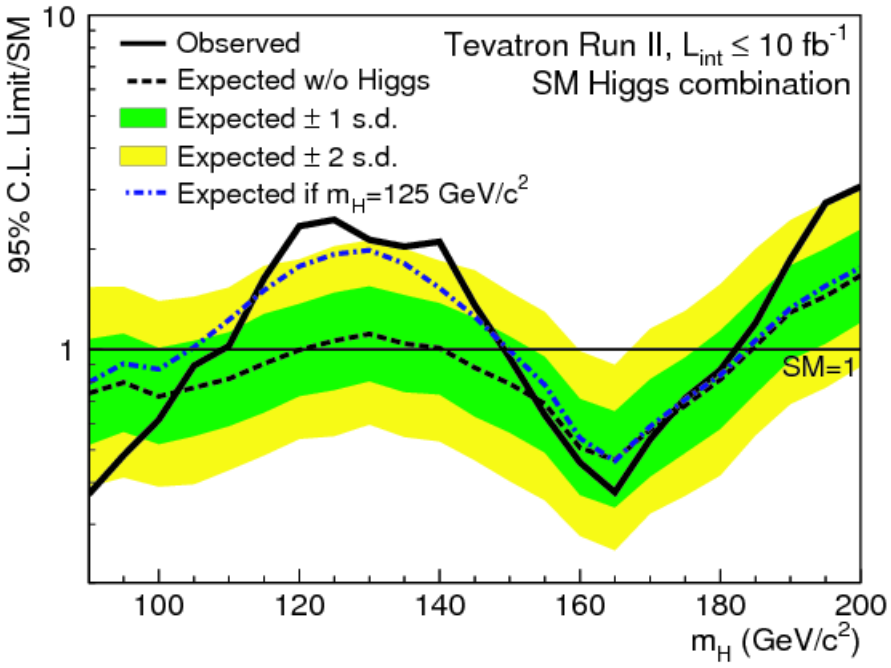


FIGURE 1.11: Observed and median expected (for the background-only hypothesis) 95% C.L. Bayesian upper production limits expressed as multiples of the SM cross section as a function of Higgs boson mass for the combined CDF and D0 searches in all decay modes.

Chapter 2

Experimental facilities

The datasets that have been analyzed in this thesis have been collected in two different experimental facilities: the CDF II at the Tevatron and the LHCb at the LHC.

The Collider Detector at Fermilab (CDF II) is a multi-purpose collider detector which was located at one of the $p\bar{p}$ collision point at the Tevatron and collected data until the end of 2011. It has been the first experiment at a hadron collider developing a trigger processor which gives the opportunity to trigger on events containing a secondary vertex formed by long lived particles [42]. This allowed to greatly improve the capability of selecting events with b hadrons, enhancing the physics reach both in the low energy regime (b hadron flavour physics) and in the high energy regime (physics with b-jets).

The Large Hadron Collider Beauty (LHCb) experiment [83] instead, is a particle detector placed in one of the four pp interaction point at the Large Hadron Collider (LHC) which is currently in data taking. It is a single-arm forward spectrometer, covering the pseudorapidity range $2 < \eta < 5$, and it was designed with the main purpose of performing precision measurements in the flavour physics field.

At CDF II, the good performance on vertexing and the introduction of a trigger processor specialized in identifying secondary vertices led to a campaign of great results in the b Physics. Transferring and optimizing technologies and knowledge, LHCb has been designed in order to be able to perform precision vertexing and track reconstruction at the on-line level thanks to its tracking system and its flexible trigger system.

Besides low energy b Physics, experiment facilities with these characteristics can be exploit to collect datasets enriched of b-jets. Efficient heavy flavour tagging performance then, allow to retain tagged jets without the need to cut hard on the energy of the jets, giving the possibility to cover the low region in the invariant mass spectrum being complementary to the high mass searches performed at ATLAS and CMS.

With this in mind, at CDF II a still not analyzed dataset has been collected with an ad-hoc trigger implemented in the data taking since Spring 2008, to cope with the increased instantaneous luminosity provided by the Tevatron. Also at LHCb a campaign aiming to enhance the Physics reaches of the experiments has started. By performing measurements of quantum chromodynamics processes, electroweak theory, exotic searches and central exclusive production LHCb is now considered a forward general purpose detector. With the introduction of algorithms for the reconstruction and identification of jets, LHCb has showed his capabilities by measuring the $Z \to b\bar{b}$ [87] cross section and it has also started to look for Higgs boson decaying into a b and c quark pair [40]. A sample enriched of heavy flavour tagged jets collected by LHCb will be analyzed to study the prospects for the inclusive $H \to b\bar{b}$ search.

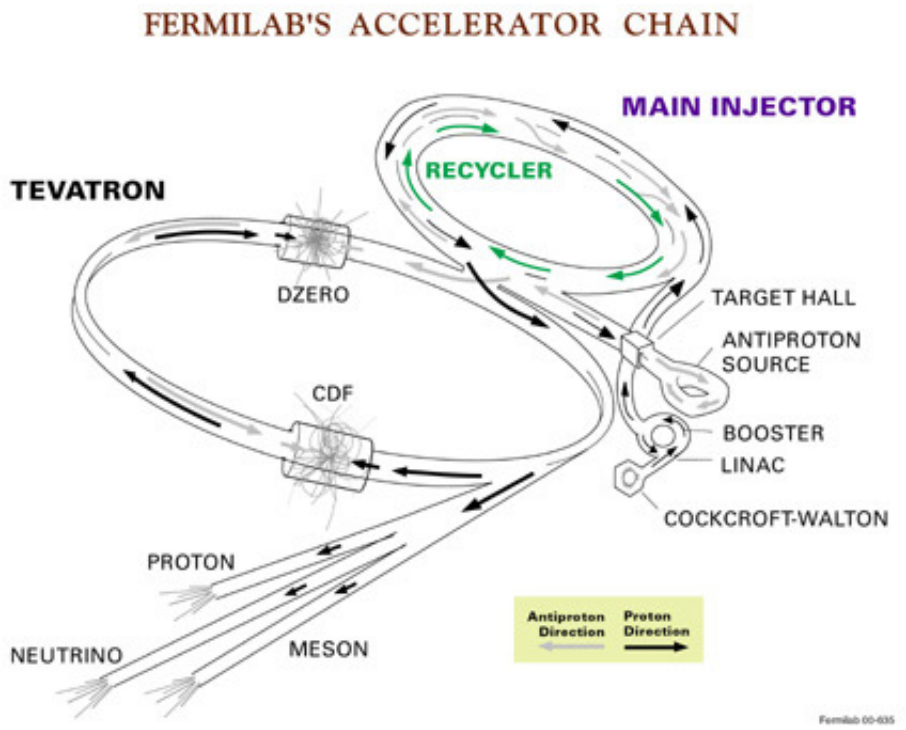


FIGURE 2.1: Schematic view of the Fermilab accelerator chain.

In this chapter, the two experimental facilities are briefly described. First, after a brief description of the Tevatron (Section 2.1.1) accelerator, the CDF II detector (Section 2.1.2) and the trigger system (Section 2.1.3) are described. Then, the LHC accelerator complex (Section 2.2.1), the LHCb detector (Section 2.2.2) and its trigger system (Section 2.2.3) are described.

2.1 The CDF II detector at the Tevatron

2.1.1 The Tevatron acceleration complex

The Tevatron in Batavia, Illinois, USA is the first large-scale superconducting synchrotron in the world. Located at the Fermi National Accelerator Laboratory (FNAL, Fermilab), it was originally named the Energy Doubler since as a proton-synchrotron it was reaching twice the energy of the original facility (the "Main Ring"), it began operation in 1983 in fixed target mode and in 1985 as a proton-antiproton collider.

From 1985 to 1996 various periods of collider or fixed target operations alternate with each other. From January 2001 to September 2011 the Tevatron fully operated as a $p\bar{p}$ collider in the so-called Run II, producing particles interactions at an energy in the center of mass frame of 1.96 TeV. Along the Tevatron ring there were two multi-purpose collider detectors, CDF and D0, that had undergone extensive upgrades during the 6 years long (1996 to 2001) preparations for Run II.

A schematic view of the Fermilab's accelerator complex is shown in Figure 2.1. The major components of the chain preceding the Tevatron Ring were: the Proton Source complex (pre-Accelerator, LINAC, Booster), the Antiproton Source complex, the Main Injector and the Recycler Ring.

The Tevatron was, and still is, the largest circular accelerator of proton-antiproton. In a lenght of approximately 6 kilometers particles were accelerated to collide at

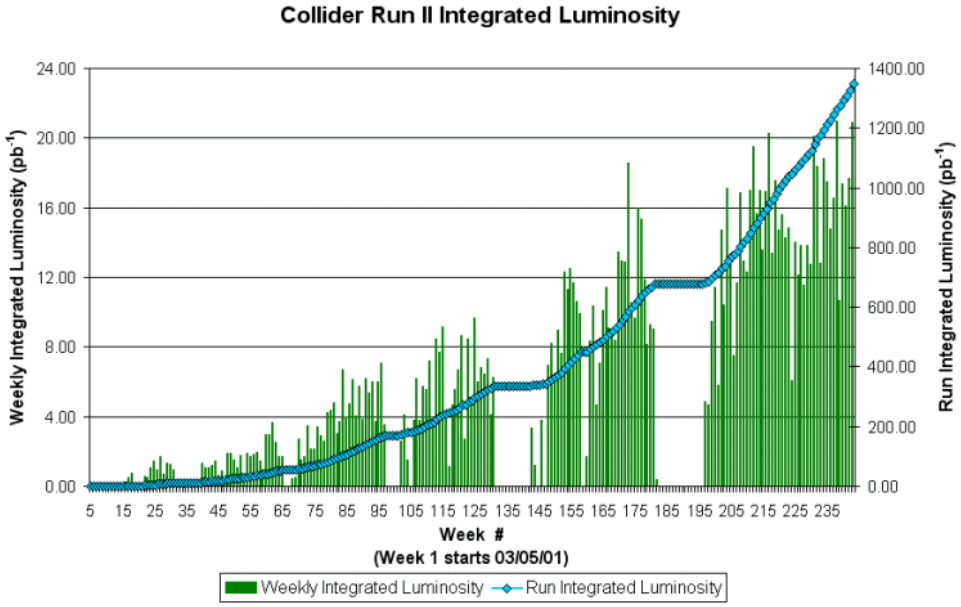


FIGURE 2.2: The weekly and run integrated luminosity achieved during the Tevatron Run II as a function of the time.

1.96 TeV every 396 ns. The beam had a transverse distribution approximate Gaussian with $\sigma_T = 30$ μm.

The instantaneous luminosity achieved by the Tevatron during Run II ranged from about 0.1×10^{32} cm^{-2}s^{-1} up to 4×10^{32} cm^{-2}s^{-1}.

Weekly and run integrated luminosity during Run II are shown in Figure 2.2. The missing periods represent the times when the Tevatron was shut down for upgrading. In total during Run II the Tevatron delivered approximately 12 fb^{-1} to the CDF and D0 experiments.

2.1.2 The Collider Detector at Fermilab

The Collider Detector at Fermilab (CDF) [43, 44], was one of the two general purpose detectors installed at the Tevatron in one of the $p\bar{p}$ collisions point. It was commissioned in1985 (at that time CDF stood for Collider Detector Facility) and significantly upgraded, first in 1989 and then between 1996 and 2001 in order to be adapted to the higher collision rate coming from the increased instantaneous luminosity delivered by the accelerator. The latter upgrade resulted in the detector version referred to as CDF II, the one that is has been used for the analyses presented in this thesis. This version of the detector has been functional until the end of operations in 2011. During Run II, the CDF Detector has recorded about the 85% of the luminosity delivered by the Tevatron, approximately 10 fb^{-1}.

The CDF II Detector, see Figure 2.3, had a cylindrical layout around the beam line, with dimensions of approximately 7.5 m in radius and 15 m in length. The coordinate system was defined such that the z axis lay along the beam line, in the direction of the proton beam at the nominal (z_0) collision point. The positive y direction is defined to point vertically upward. It is often convenient to use a polar

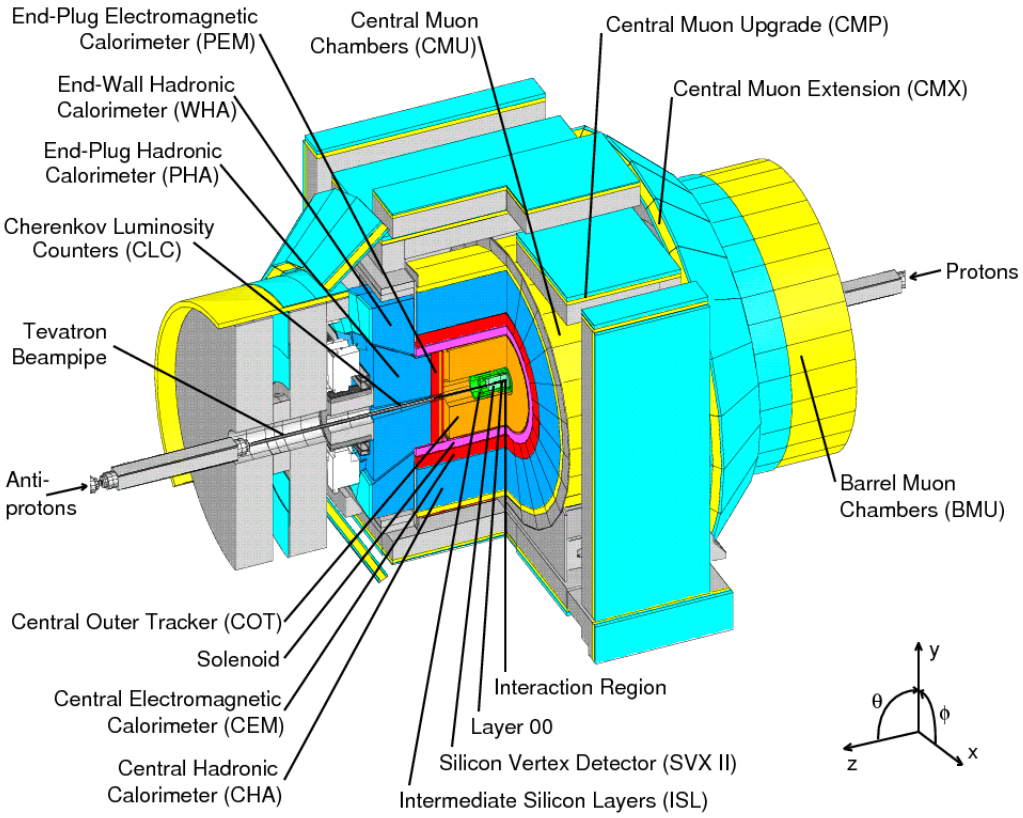


FIGURE 2.3: Isometric view of the entire CDF II detector. Labels indicate the different sub-detectors.

variable invariant under boost along z. This variable is the rapidity:

$$y = \frac{1}{2}\ln\left(\frac{E + p\cos\theta}{E - p\cos\theta}\right), \tag{2.1}$$

with E is the energy, p the momentum and θ the polar angle of the particle. Rapidity is close to 0 when the particle directions tend to be transverse i.e., perpendicular to the beam direction and $y \to \pm\infty$ when the particle is predominantly parallel to the beam axis. For highly energetic particles the rapidity y can be conveniently substitute with the pseudo-rapidity defined as

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right). \tag{2.2}$$

The detector was composed of several sub-detectors, each one optimized for a specific task. Starting from the interaction point and following the path of an outgoing particle within acceptance there were:

- a tracking system enclosed by a superconducting solenoid (1.5 m in radius and 4.8 m in length), which generated 1.4 T magnetic field parallel to the beam axis. The magnetic field was uniform in the tracking region at the level of 1% or better;
- finely segmented calorimeters;
- planar muon drift chambers backed by scintillation counters.

Some of the sub-detectors, as the time-of-flight detector, the small angle spectrometers on beam pipe, are not described in this section because they are not used in the analyses described in this thesis. More details can be found in [44].

The tracking system

The Tracking System, inside the superconducting solenoid which provided an uniform magnetic field of 1.4 T, measured the trajectories and momenta of the charged particles. It was made of an inner silicon tracking system and outer gas drift-chamber as shown in Figure 2.4.

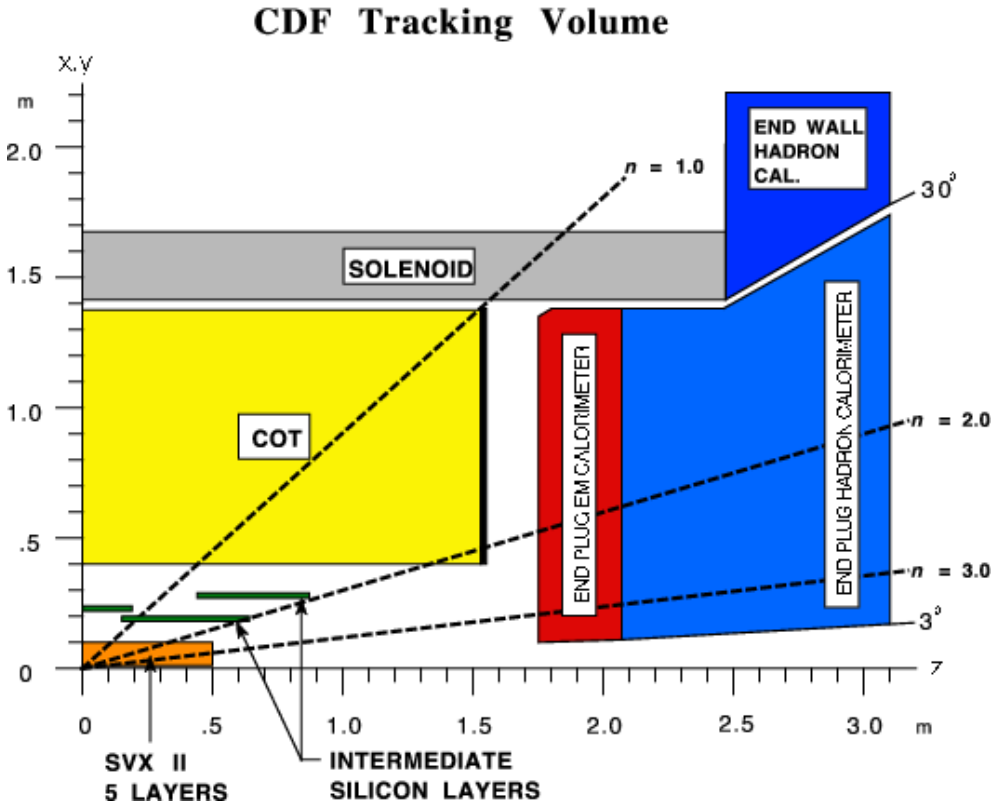


FIGURE 2.4: Longitudinal view of the Tracking System of the CDF II Detector.

The Inner Tracker

The inner silicon detector system was composed of three sub-detectors: the Layer 00 (L00), the Silicon Vertex Detector (SVX), and the Intermediate Silicon Layers (ISL). It was designed with the aim of providing excellent spatial resolution, which is needed to identify displaced secondary vertices which are the signature of events with bottom quarks. All the silicon microstrip sensors had a space resolution of ~ 12 μm in the direction transverse to the beam.

All three detectors had a barrel geometry, for a total of eight layers of silicon sensors (p-n junctions) to guarantee a good track reconstruction even in the event of component failure or degradation.

The innermost detector, placed directly on the beam pipe, was the Layer 00. It consisted of a series of silicon microstrip sensors, single-sided for improved radiation resistance with the disadvantage of providing position information in the $(r-\phi)$

plane only. The L00 detector covered a region up to $|\eta|<4$. L00 purpose was to improve the track impact parameter (IP), the measured distance of minimum approach to the beam axis, resolution (~ 40 μm for tracks with $p_{\mathrm{T}} \geq 2$ GeV/c).

The Silicon Vertex Detector (SVX), which had a coverage up to $|\eta|<2.0$, extends radially from 2.5 cm to 10.7 cm, was located immediately outside Layer 00. The SVX system was composed of 5 layers of double-sided radiation-hardened silicon wafers able to provide information in both the $(r-\phi)$ and $(r-z)$ planes. The double-sided imprint of the wafers allowed for 3D position measurements: one side of the wafer had strips along the beam axis, the other one had either 90° or 1.2° stereo strips.

The outermost Intermediate Silicon Layers (ISL) consisted of three layers at radial distance of 20, 22 and 28 cm from the beampipe. The middle layer was positioned in the central region of the detector ($|\eta|<1$) while the closest and the furthest away provided precision tracking in the $1<|\eta|<2$ region. The three layers consisted of double-sided strips with axial strips on one side and strips with a 1.2° stereo angle with respect to the beam axis on the other side. By combining the information from the SVX and the ISL tracks could be reconstructed in three-dimensions.

The Central Outer Tracker

The Central Outer Tracker (COT) [45] was located around the inner silicon system and it extended from an inner radius of 44 cm to an outer radius of 132 cm and was 155 cm long. It was an open-cell wire drift chamber filled with a gas mixture of Argon, Ethane and CF_4 in proportion 50%, 35% and 15%. The chamber provided full coverage for $|\eta|<1$ and a region of reduced acceptance coverage for $1<|\eta|<2$. The COT was composed of 4 axial and 4 stereo superlayers of azimuthal cells. Each cell had alternated sense and field shaping wires. Within the cell width, the trajectory of a particle was sampled 12 times. The cells were tilted by 35° with respect to the radial direction in order to make the electrons drifting perpendicularly to the radius for optimal momentum resolution, which reached an overall transverse resolution of:

$$\sigma(p_{\mathrm{T}})/p_{\mathrm{T}} = 0.15\% p_{\mathrm{T}}[\mathrm{GeV}/c] \tag{2.3}$$

The calorimeter system

CDF was provided with scintillator sampling calorimeters divided into an inner electromagnetic and an outer hadronic compartment. Both calorimeters were segmented into projective towers (also called wedges), with each tower consisted of alternating layers of passive absorber material (Pb in the electromagnetic and Fe in the hadronic compartment) and plastic scintillator for shower sampling. The light from the scintillator plates was read out through wavelength-shifting bars or plates and light guides by photo-multiplier tubes (PMT).

CDF calorimeters covered the full azimuthal coverage and $|\eta|<3.6$. The calorimeter system included the Central Electromagnetic Calorimeter (CEM) and the Central Hadronic Calorimeter (CHA) in the $|\eta|<0.9$ region, the Endwall Hadronic Calorimeter (WHA) in $0.9<|\eta|<1.3$ and the electromagnetic and hadronic plug calorimeters (PEM,PHA) in $1.1<|\eta|<3.6$, see Figure 2.5.

The central calorimeters, CEM, CHA and WHA, comprised 478 towers, each covering a range of 15° in azimuth and about $\delta\eta = 0.11$ in pseudorapidity.

The CEM [46] was a sampling calorimeter made of 31 layers of 0.5 cm thick polystyrene scintillator tiles (the active media) alternated with absorber layers of 0.32 cm thick lead sheets. The total amount of material corresponded to a depth of

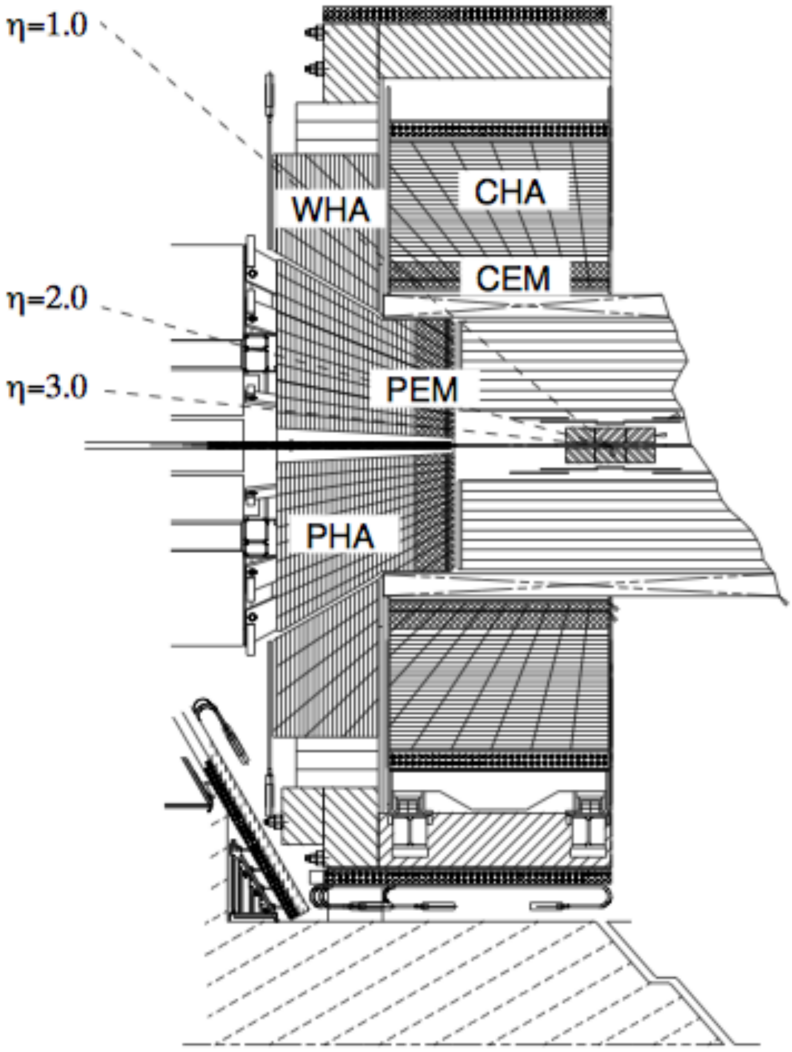


FIGURE 2.5: View of the calorimeter system.

about 18 radiation lengths (X_0). The CEM energy resolution was:

$$\sigma(E_{\mathrm{T}})/E_{\mathrm{T}} = 13.5\%/\sqrt{E_{\mathrm{T}}[\mathrm{GeV}]} \oplus 2\% \tag{2.4}$$

where E_{T} is the transverse energy of a photon or an electron hitting the CEM perpendicularly, and the $\oplus 2\%$ term represents the contribution to the resolution due to the lateral or longitudinal shower leakage, which is at first order constant.

The CHA and WHA were also sampling calorimeters [47]. The CHA surrounded the CEM and consisted of 32 layers of 2.5 cm-thick steel alternating with 1 cm of acrylic scintillator. The WHA extended the CHA and employed the same active and passive medium of the CHA but was made of 15 layers with double thickness of lead (5 cm). The total thickness for both the hadronic parts was approximately 4.7 absorption lengths (λ_0). Resolutions of CHA and WHA for perpendicular particle entrance were, respectively:

$$\begin{aligned} \sigma(E_{\mathrm{T}})/E_{\mathrm{T}} &= 50\%/\sqrt{E_{\mathrm{T}}[\mathrm{GeV}]} \oplus 3\% \\ \sigma(E_{\mathrm{T}})/E_{\mathrm{T}} &= 75\%/\sqrt{E_{\mathrm{T}}[\mathrm{GeV}]} \oplus 4\% \end{aligned} \tag{2.5}$$

The PEM calorimeters [48] shared the same tower segmentation of the CEM in η and ϕ for $2.11<|\eta|<3.64$, but finer in ϕ for $|\eta|<2.11$ ($\Delta\eta = 7.5°$). They were composed by 22 layers that consisted of 4.5 mm thick lead and 4 mm thick scintillator, the total thickness was about $21X_0$. The PEM transverse energy resolution was:

$$\sigma(E_{\rm T})/E_{\rm T} = 16\%/\sqrt{E_{\rm T}[\,{\rm GeV}]} \oplus 1\% \tag{2.6}$$

The PHA, located behind PEM, had the same tower segmentation. The technology was the same as for CHA, with 23 layers of 2 cm thick steel absorber alternating with 6 mm thick scintillator. The total amount of material corresponded to $\sim 4.7\lambda_0$. PHA resolution was:

$$\sigma(E_{\rm T})/E_{\rm T} = 80\%/\sqrt{E_{\rm T}[\,{\rm GeV}]} \oplus 5\% \tag{2.7}$$

The muon detector system

The Muon Detector System [49] consisted of four detectors: the Central Muon Detector (CMU), the Central Muon Upgrade Detector (CMU), the Central Muon Extension Detector (CMX) and the Intermediate Muon Detector Detector (IMU). The detectors were composed of drift chambers and scintillator counters and they covered different region of η. Figure 2.6 shows the spatial coverage of each of the detectors in the (η, ϕ) plane.

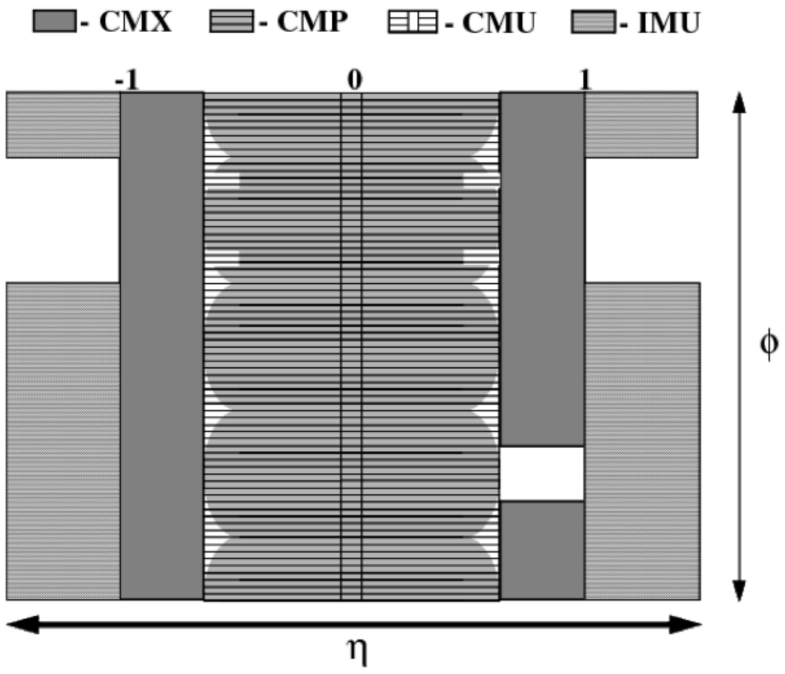


FIGURE 2.6: Coverage of the muon system in the (η, ϕ) plane.

Muon detectors shared common features. They consisted of stack of rectangular drift chamber modules, filled with a mixture of Argon and ethane at 50%, composed of single-wire cells. Stacks were four layers deep with laterally displaced cells from layer to layer to compensate for cell inefficiencies. The difference in drift-electrons arrival-times between neighbor cells provided a typical resolution of 250 μm for the hit position in the transverse plane. Charge division at the wire ends measured the z coordinate with a 1.2 mm resolution. A muon candidate was reconstructed when a

short track segment (stub) in the muon chambers match the extrapolation of a COT track.

The CMU detector was situated right behind the Central Hadron Calorimeter and covered the region with $|\eta|<0.6$. It consisted of a layer of drift chambers with four planes of wires and detected muons with a transverse momentum of at least 1.4 GeV/c. Behind the CMU, the magnet return yoke of the solenoid, made of 60 cm of steel, stopped particle leaking from the calorimeter before the CMP and the CMX detectors. It was very likely that the only charged particles able to penetrate the shielding are muons.

The CMP and CMX detectors detected muons with $p_{\mathrm{T}}>2.2\,\mathrm{GeV}/c$ in the pseudo-rapidity range $|\eta|<0.6$ and $0.6<|\eta|<1.0$, respectively. The CMP detector comprised drift chambers backed by 2.5-thick scintillator counters (CSP), while the CMX detector consisted of large conical sections of drift tubes and scintillation counters (CSX).

2.1.3 The CDF II trigger system

At the Tevatron Run II the bunch crossing frequency was 2.53 MHz, which was much higher than the possible event recording rate (150 Hz).

The CDF trigger system had a three-level architecture providing a rate reduction sufficient to allow more sophisticated event processing one level after another with a minimal deadtime, as shown in Figure 2.7.

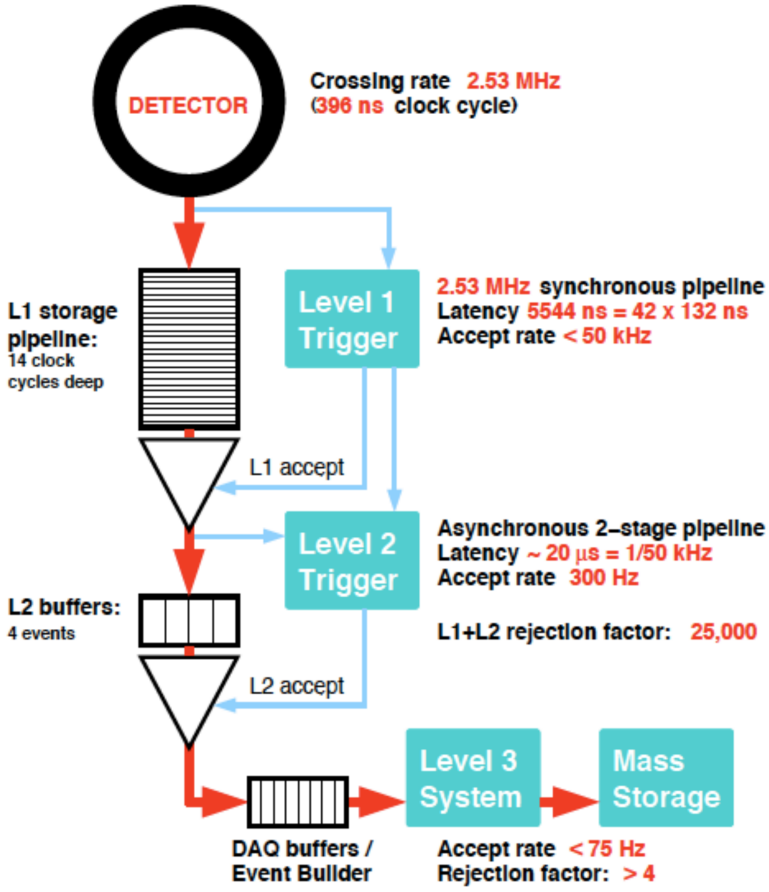


FIGURE 2.7: CDF II readout and trigger scheme.

At Level-1 (L1) raw muons, tracks and calorimeter information were processed to produce a L1 decision. L1 was a synchronous 40 stages pipeline and it was based

on custom-designed hardware. It can provide a trigger decision in 5.5 μm and during normal data taking its output rate was typically below 30 kHz. When an event was accepted at L1, subsets of detector information were sent to the Level-2 (L2) system, where some limited event reconstruction was performed and a decision was taken. The L2 was an asynchronous pipeline and it was based on a combination of custom-designed hardware and commodity processors. Its average latency was 20 μm and its maximum output rate was 1 kHz. Upon L2 accept, the full detector data was readout and sent to Level-3 (L3) processors farm for further processing. Events accepted at L3 were sent to mass storage.

According to the features of the events one would like to select, specific sets of requirements were established by exploiting the physics objects (primitives, such as raw tracks, calorimeter energy deposits, etc.) available for each trigger level. Then links across different levels were established by defining trigger paths, identified by a unique combination of a L1, a L2, and a L3 trigger; datasets were then formed by merging the data samples collected via different trigger paths.

One of the key point that allowed the development of a trigger with an on-line b-tagging algorithm [50] was the Silicon Vertex Tracker (SVT) [42]. It was a L2 trigger processor which, exploiting the potential of a high precision silicon vertex detector, allowed to trigger on tracks with a significantly different from zero impact parameter (IP). SVT combined hits from silicon detectors with tracks reconstructed by XFT. The association was performed by a device based on associative memories. A road was found as a coincidence between hits, four out of five of the silicon layers, and XFT tracks. Overall SVT tracking efficiency was about 80% and the resolution on the IP reach was of 35 μm for 2 GeV/c tracks, comparable with the one achieved for the off-line resolution.

The L1 XFT worked with COT signals at high collision rates, returning track p_{T} and ϕ by means of a fast $r-\phi$ reconstruction. Track identification is performed searching and combining track segments in the 4 axial superlayers of the drift chamber. The upgraded XFT maintained the existing axial system but adding new boards to find track segments also in the outer stereo layers of the chamber. Moreover XFT segments of finer granularity can be sent to L2 where a 3D-track reconstruction can be performed with a good resolution on cotθ ($\sigma_{\mathrm{cot}\theta} = 0.12$) and z_0 ($\sigma_{z_0} = 11$ cm).

The L2CAL system used a fixed cone cluster finding algorithm which prevented fake clusters formation and exploited the full 10-bit trigger tower energy information. A jet was formed starting from a seed tower above a threshold and adding all the towers inside a fixed cone centered at the seed tower and having a radius $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} = 0.7$ units in the azimuth-pseudorapidity space. Jet position is calculated weighting each tower inside the cone according to its transverse energy. The L2 jets energy and position measurements were with nearly equivalent resolution to the off-line one.

2.1.4 The trigger for b-jets

In 2008 a trigger [50] path, called DIJET_BTAG, was implemented in the CDF trigger system. It was optimized for the $H \to b\bar{b}$ search, but it can be efficiently used for every channel with b-jets in the final state. The basic idea was to look for displaced tracks in jets and then make additional selections based on variables distinctive of tracks coming from b hadrons while keeping the cut on the jet energies as low as possible in order to limit the bias on dijet invariant mass distribution. Figure 2.8 illustrates the principle behind the b-tagging algorithms.

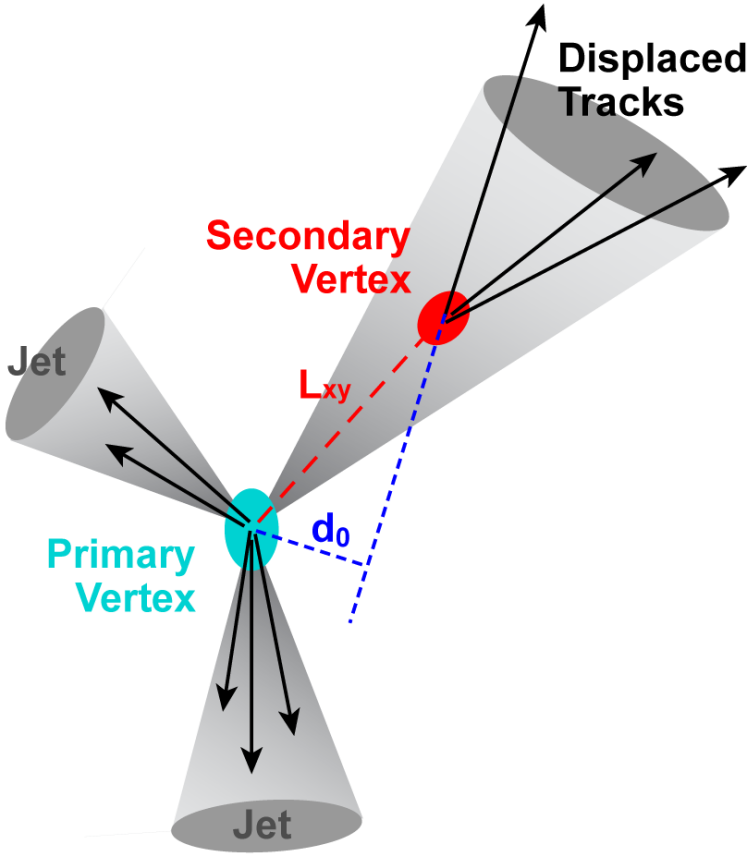


FIGURE 2.8: Sketch showing that tracks coming from b hadron decays have larger IP and form a secondary vertex.

At level 1 the selection was two tracks with $p_{\mathrm{T}} > 2\,\mathrm{GeV}/c$ and 2 CEM towers with $E_{\mathrm{T}} > 5\,\mathrm{GeV}$; the core of the new algorithm was at level 2, where the on-line b-tagging was performed. The requirements at this stage were:

- two central ($|\eta| < 1.0$) calorimeter jets with $E_{\mathrm{T}} > 15$ GeV,
- two XFT-SVT tracks 3-D matched to one of the leading jets ($\Delta R < 0.7$) with:
  - a signed impact parameter $d_0 > 0.9$ μm, the signed is positive if the interaction point of the track with the jet axis is in the same hemisphere of the jet
  - forming a 2-track displaced vertex with $R_b > 0.1$ cm, where $d_0 = R_b \sin(\phi_b - \phi) \sim R_b(\phi_b - \phi)$ with R_b and ϕ_b the b hadron decay length and azimuthal angle.

Figure 2.9 shows the comparison between the efficiency of the cut on signed d_0 for signal and for background. At level 3 the confirmation of the level 2 requirements with off-line quality variables is applied.

The overall efficiency of the trigger (Table 2.1) was 13% for SM $H \to b\bar{b}$ process, 11% for MSSM $Hb \to b\bar{b}b$ process and 5% for $Z \to b\bar{b}$ events. This trigger algorithm ran up to high instantaneous luminosity without being prescaled, thus it provided an overall gain of 40% on signal acceptance.

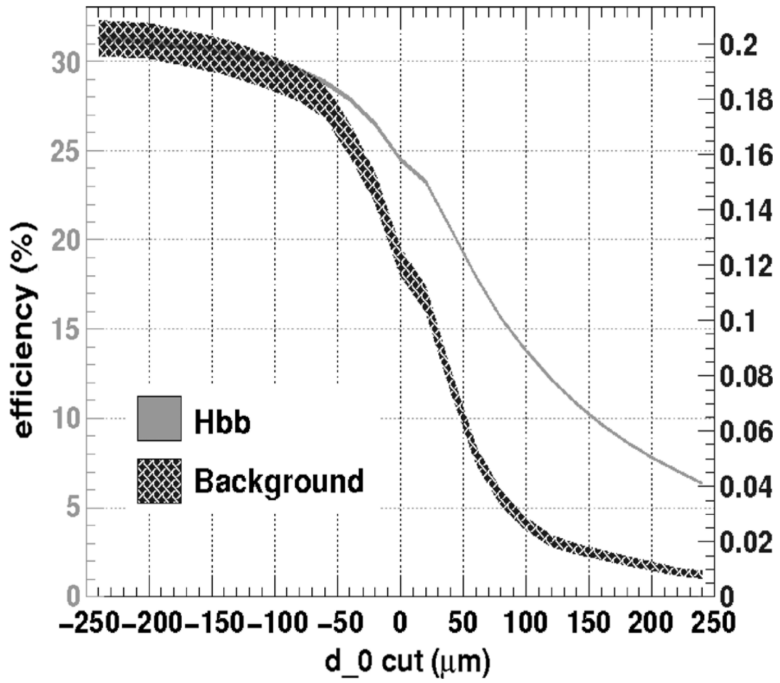


FIGURE 2.9: Efficiency of the cut on signed d_0 for signal (left-hand vertical scale) and background (right-hand vertical scale) events. Both curves are for $R_b > 0.1$ cm.

TABLE 2.1: DIJET_BTAG trigger performances: L1, L2 and L3 trigger efficiencies on different signals.

	SM $H \to b\bar{b}$	MSSM $Hb \to b\bar{b}b$	$Z \to b\bar{b}$
L1 Eff(%)	70	69	38
L2 Eff(%)	14	43	6
L3 Eff(%)	13	11	5

This trigger worked until the end of the Tevatron operations, collecting a total of 5.4 fb^{-1} of data.

Monte Carlo simulation

The Monte Carlo simulation of the physics processes is divided in thee stages:

- the simulation of the hard parton scattering and the subsequent soft interactions;
- the interaction of the produced particles with the simulated detector, correcting for the effect of the on-line data taking conditions;
- the simulation of the trigger and the off-line analysis requirements.

The first step is performed using the PYTHIA 6.216 [65] physics process generator. PYTHIA performs leading order (LO) matrix calculations to simulate the hard parton scattering. In addition, it uses parton showering and the Lund hadronization model to describe the final part of the hadronic collision. The LO predictions

generally describe the principal kinematical features of the processes and, consequently the shaped of related distributions with a good approximation. The normalization is, instead, usually badly predicted because of the lack of contribution from the higher orders diagrams to the total cross sections. The Parton Distribution Functions (PDFs), used in this analysis, is the leading order (LO) CTEQ5L set [68].

The interaction of the final state particles with the passive and active materials of the detector and the response of the various subdetectors is simulated using the GEANT3 [69] software. Data taking conditions are considered into the detector simulation, including time-dependent beam position and the operating conditions of the detector components. Instantaneous luminosity profile is part of the simulation in order to model at best the multiple interactions in the same $p\bar{p}$ collision. Simulated events are generated to reproduce data collected in a given data-taking period.

The output of the detector simulation has the same format as the real data so that every selection and reconstruction algorithm can be applied evenly to the data sample and the simulated events. In order to be able to determine the trigger efficiency on any signal sample a trigger simulation that reproduces the hardware performances has been developed for almost all the online processors except the XFT stereo for which the full simulation is not available. Then, the same reconstruction algorithms and selection criteria are applied. In the simulated samples we are interested only in jets originating from a known generator level quark type. For this reason, in the Monte Carlo simulation we match each jet to a true initiating quark jet (b, c or $usdg$) requiring ΔR between the reconstructed jet and the one at the generation level to be less than 0.7.

2.2 The LHCb at the LHC

2.2.1 The LHC

The Large Hadron Collider (LHC) [88, 89, 90] is the is the world's largest and highest energy particle accelerator. Proposed and realised by the European Organization for Nuclear Research (CERN), it was designed to collide protons, as well as lead ions, at an unprecedented energy and rate, in order to address some of the most fundamental questions of particle physics.

LHC is a synchrotron housed in the tunnel that was once home to the Large Electron Positron(LEP) collider at a depths ranging between 45 m and 170 m. It is a 27 km circumference ring and it is the last part of a chain of accelerators, as shown in Figure 2.10. At the start of the LHC injector chain is a cylinder of hydrogen gas which acts as a source of protons, produced by ionising the gas. The protons are fed into the LINear ACcelerator 2 (LINAC2) where they reach an energy of 50 MeV before they are injected into the Proton Synchrotron Booster (PSB). Once the protons reach an energy of 1.4 GeV they are transfered into the Proton Synchrotron (PS) followed by the Super Proton Synchrotron (SPS) where they are accelerated to energies of 25 GeV and 450 GeV, respectively.

They are finally injected into the LHC in two counter rotating beams. 1232 dipole magnets keep the beams on their circular path, while an additional 392 quadrupole magnets are used to keep the beams focused, in order to maximize the chances of interaction in the four intersection points (IP), where the two beams cross. In total, over 1600 superconducting magnets are installed. Approximately 96 tonnes of liquid helium are needed to keep the superconducting magnets at their operational temperature of 1.9 K. The field in the dipole magnets increases from 0.53 T to 8.3 T while the protons are accelerated from 450 GeV to 6.5 TeV.

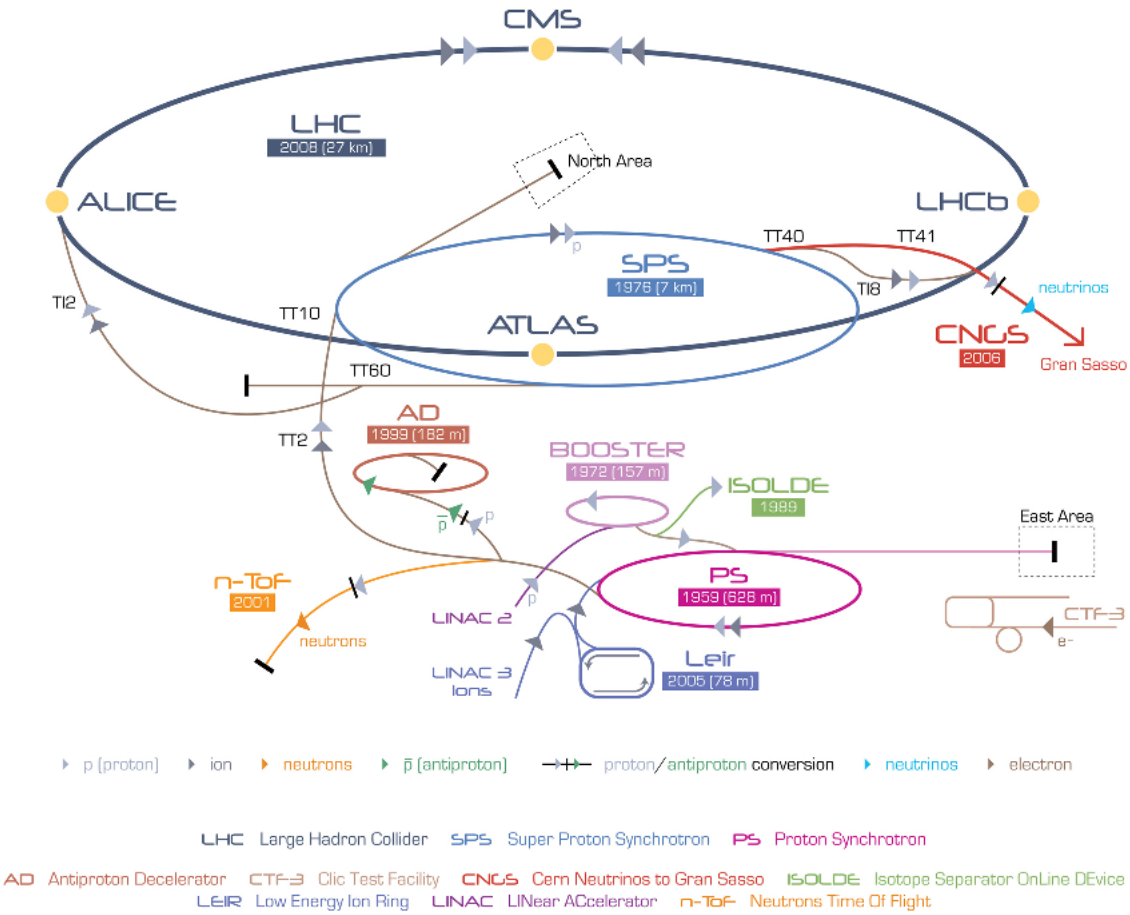


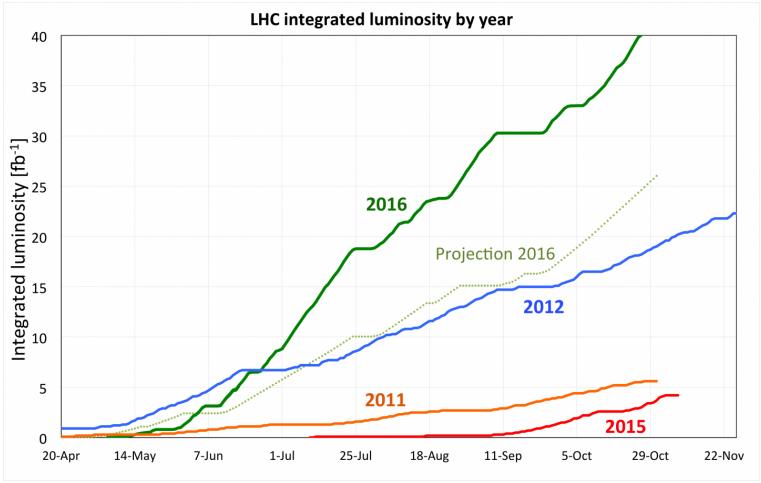
FIGURE 2.10: The LHC accelerator complex at CERN.

The protons complete their long journey when they are brought to collide at one of the four intersection points where the two beam lines cross. It is at these intersection points that the four main LHC experiments are placed to record the remains of the collisions. The Compact Muon Solenoid (CMS) [91] and A Toroidal LHC ApparatuS (ATLAS) [92] are general purpose detectors. A Lead Ion Collision Experiment (ALICE) [93] is designed to examine in particular heavy-ions collisions, specifically looking for and examining the nature of the exotic state of matter known as quark-gluon plasma. The final of the four main LHC experiments is the LHC beauty detector (LHCb) [83].

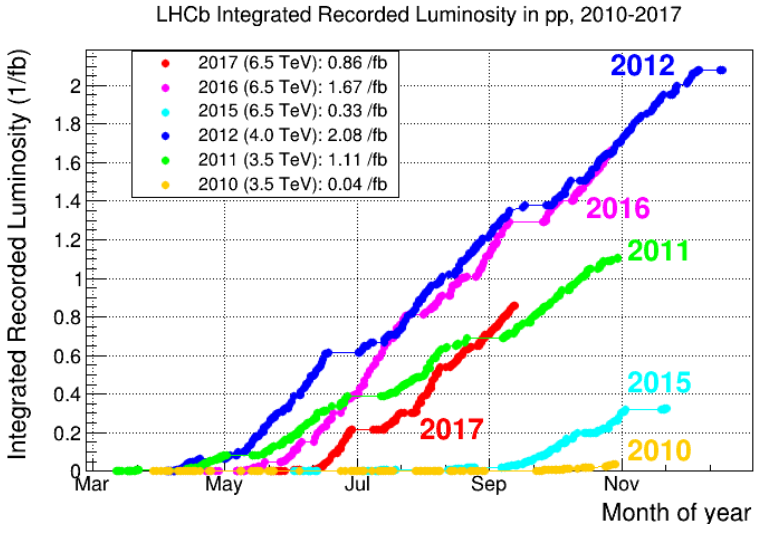
The LHC Run I and Run II

LHC operations are organised in periods of data taking followed by long shutdowns during which maintenance and upgrade can be performed to the accelerators and the detectors. The first period of data taking, known as Run I, happened between 2010 and 2013. After an initial ramp up in energy during 2010, the nominal centre of mass collision energy was $\sqrt{s} = 7\,\text{TeV}$ during 2011 while in 2012 the collision energy was increased to 8 TeV. Between 2013 and 2015 the first long shutdown took place: many interventions were performed to the LHC to enable collisions at 14 TeV and also the other components of the accelerator chain were improved. Run II started in 2015 and will be concluded in 2018 with the beginning of the second long shutdown. In 2015 and 2016 the the nominal centre of mass collision energy has been $\sqrt{s} = 13\,\text{TeV}$.

The LHC is designed to operate at an instantaneous luminosity of $\times 10^{34}$ cm^{-2}s^{-1}, which was first reached in June 2016. Figure 2.11 shows the luminosity delivered by LHC to the general purpose detectors and the luminosity collected by LHCb. It can be noticed that the luminosity collected by LHCb is much smaller than that delivered to the general purpose detectors. The instantaneous luminosity delivered at LHCb is roughly 4×10^{32} cm^{-2}s^{-1}. In fact, the luminosity is intentionally levelled by locally de-focusing the beams. The aim of luminosity levelling is to allow for the more efficient reconstruction of secondary vertices by reducing the number of primary vertices per bunch crossing (1 or 2 in LHCb and approximately 20 in ATLAS and CMS).



(a)



(b)

FIGURE 2.11: The integrated luminosity delivered by LHC (a). The integrated luminosity recorded at LHCb (b).

2.2.2 The LHCb detector

The Large Hadron Collider Beauty (LHCb) experiment is designed to take advantage of the copious production of B mesons at the LHC. Because $b\bar{b}$ pairs are produced predominantly with highly correlated trajectories and are so highly boosted

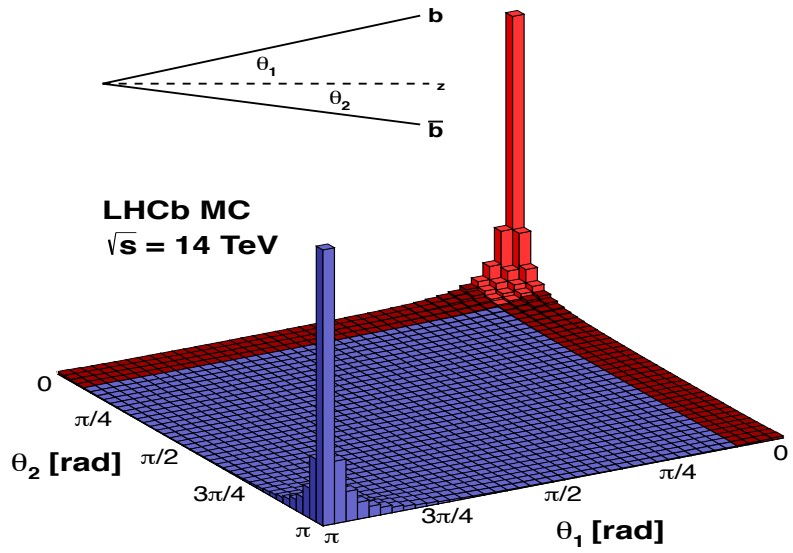


FIGURE 2.12: The production angles, relative to the beam-line, of bb pairs produced in proton-proton collisions at the LHC at $\sqrt{s} = 14$ TeV.

at LHC energies, the polar angles relative to the beam-line tend to be very small, as shown in Figure 2.12.

To take advantage of this, the LHCb detector has its unique forward-arm design, as shown in Figure 2.13. LHCb covers only the region of high pseudo-rapidity, $2 < \eta < 5$, which corresponds to about the 2.5% of the solid angle. Nevertheless, about 27% of b quarks produced in LHC collisions fall within the acceptance of LHCb.

To facilitate a clear frame of reference when discussing the LHCb detector a global coordinate system is defined, and it is also shown in Figure 2.13. The origin is located at the point at which the two LHC proton beams intersect each other and the protons collide. The z-axis is parallel to the line of the proton beams, with positive z pointing into the main LHCb detector, also called the downstream region. The y-axis is in the vertical direction, with positive y pointing upwards, and the x-axis is horizontal, with positive x pointing into the page.

As Figure 2.13 shows, LHCb is comprised of different sub-detectors, each designed for a precise purpose. The design and performance of each of these sub-detectors are here briefly discussed in turn. The LHCb detector as been designed to satisfy some experimental requirements:

- discriminate between primary and secondary vertices as b and c hadrons travel a distance of the order of a few mm before decaying,
- discriminate between particle species, both hadrons and leptons,
- on-line fast and efficientf event selection for discriminating interesting events among the large background.

The first and the third points, which are addressed through the tracking and the trigger systems, are of primary importance for the reconstruction and identification of the b-jets.

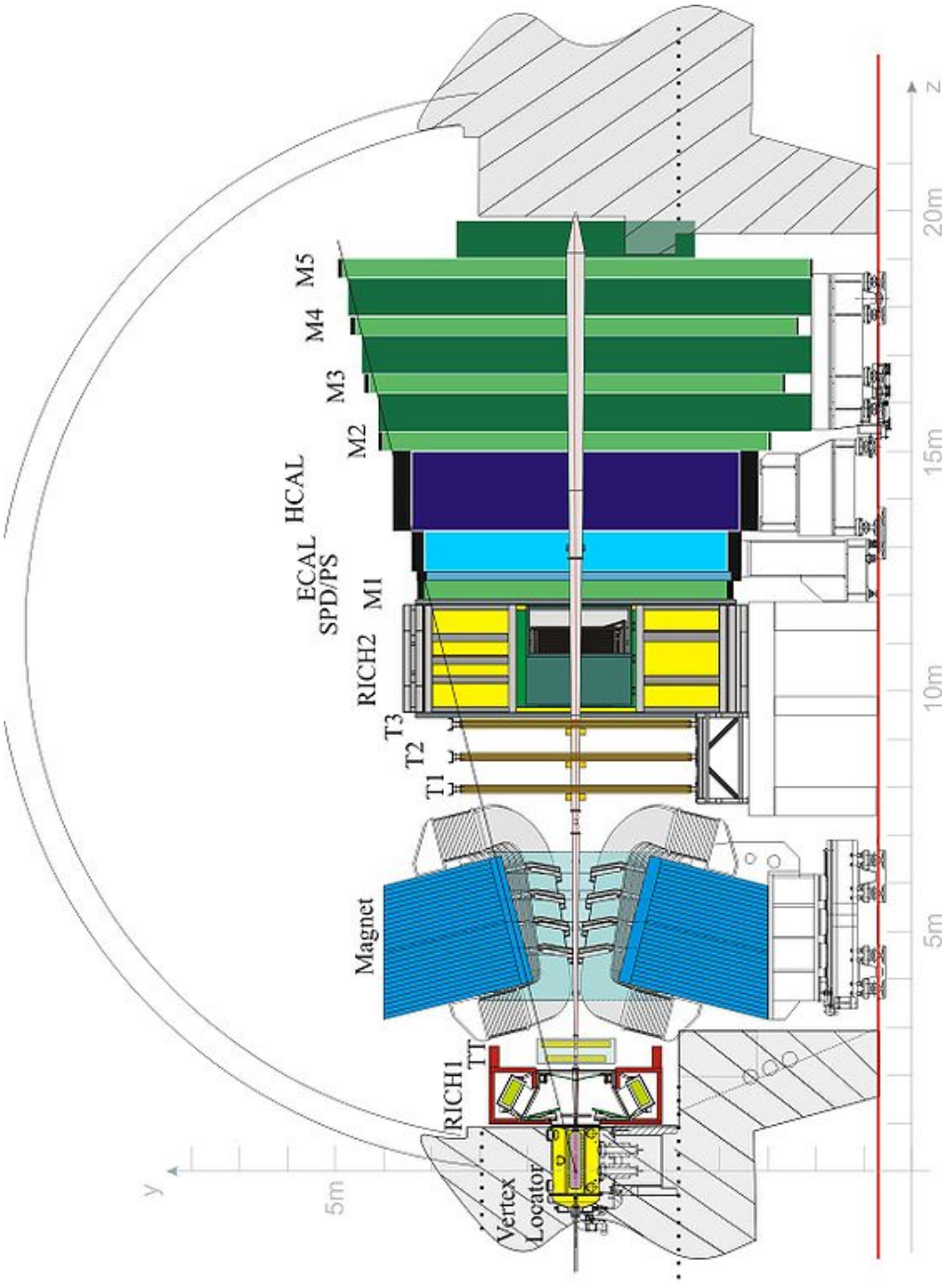


FIGURE 2.13: A side view of the LHCb detector.

The tracking system

The VErtex LOcator

The LHCb VErtex LOcator (VELO) [94] is a silicon microstrip vertex detector designed to provide precise track coordinate measurements close to the interaction region. As its name suggests, it is used to locate the position of any proton-proton collisions within LHCb, the primary vertices, as well as the decay points of any long lived particles produced in the collisions, such as B and D mesons, the secondary

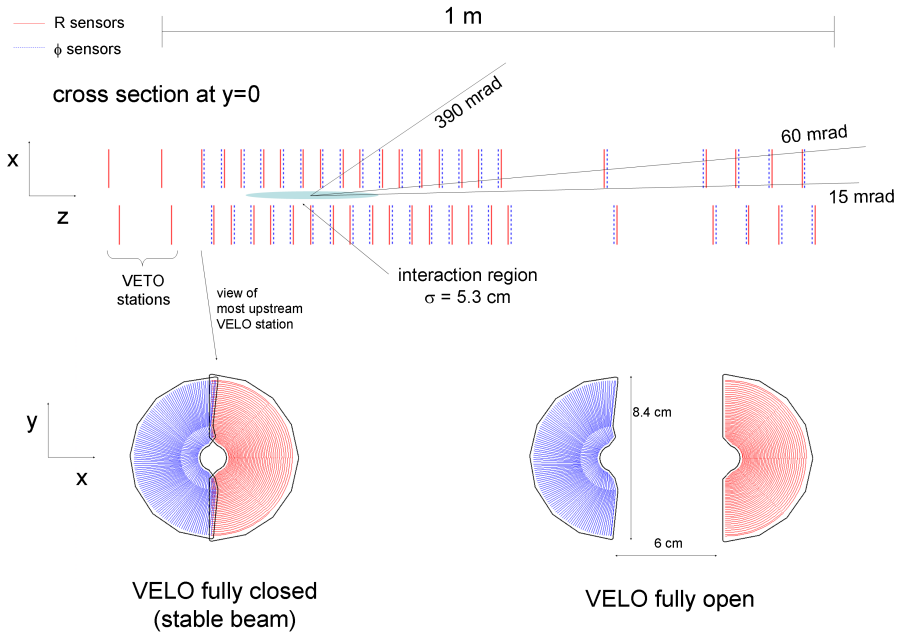


FIGURE 2.14: The layout of the VELO modules and sensors. The R sensors, which measure the radial position of hits, are shown in red, while ϕ sensors, which measure the azimuthal angle of hits, are shown in blue.

vertex. This is the most complex and delicate sub-detector of the whole LHCb. The possibility to perform flavour physics, and also high-p_{T} physics with b-jets, mostly depends on the performance of this detector.

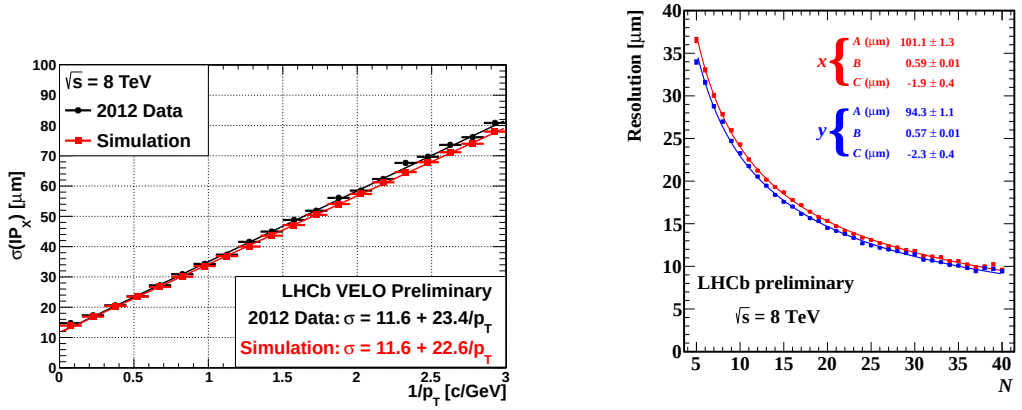
The VELO consists of two sets of 21 modules, located on either side of the beam line (see Figure 2.14). To minimize the extrapolation distance between the first hit of a reconstructed track and the interaction point the active regions of the VELO sensors start at just 8 mm from the beam-line. Each module comprises two semicircular sensors (one R sensor and one ϕ sensor), each approximately 300 μm in thickness and with a diameter of 84 mm. The R sensors are embedded with silicon in concentric semicircles centred on the beam axis, allowing for determination of the r coordinates of track points. The orthogonal coordinates are supplied by the ϕ sensors in which the silicon strips run radially out from the beam axis.

During LHC injection, the width of the beam increases significantly. Therefore it is necessary to horizontally retract each half of the VELO by $\sim$ 3 cm to avoid damage to the sensors. Once the beam is stable, the aperture reduces to $\sim$ 100 μm and the two halves are moved back together so that they overlap slightly in order to ensure coverage of the full azimuthal acceptance and to aid with module alignment. A vacuum is maintained within the VELO to minimise interactions before charged particles reach the silicon microstrips.

VELO performance

The performance of the VELO is of critical importance to the majority of LHCb analyses. The spatial resolution of the impact parameter and of the primary vertex in the x and y direction are shown in Figure 2.15. The impact parameter (IP) resolution improves with increasing particle momentum. The resolution on the x component

of IP achieves a resolution on IP_x of < 36 μm for particles with $p_{\mathrm{T}} > 1$ GeV/c. The excellent IP resolution is reflected in the PV resolution. For a PV using 25 tracks in its fit the resolution on the x coordinate of its position is just 13.1 μm, while the resolution on the y coordinate is just 12.5 μm.



(a) (b)

FIGURE 2.15: Resolution of the x component of IP measurements as a function of $1/p_{\mathrm{T}}$ (a). (b) Resolution of PV x and y coordinates as function of the number of hits.

The dipole magnet

A dipole magnet [95] provides an integrated magnetic field of about 4 Tm. A diagram of the magnet is shown in Figure 2.16. The magnet is designed to fit the detector geometry and the momentum measurements covers the forward acceptance of ± 250 mrad vertically and of ± 300 mrad horizontally. It is located between the TT station and the T1-T3 stations. The generated magnetic field is directed along the y axis, therefore the xz plane is the bending plane.

It is a warm (not super-conducting) magnet consisting of two identical, saddle shaped aluminum conducting coils positioned symmetrically above and below the beam-line. Its polarity can readily be reversed, so as to cancel any asymmetries in the detection efficiency that might fake CP-violation. Throughout data-taking this has been done regularly, and an approximately equal quantity of data has been taken with each polarity. The magnetic field is generated in the positive, or negative, y direction. In order to achieve the required momentum resolution, the magnetic field is measured with a precision of a few parts per 10^{-4}. It is mapped with a fine grid of $8 \times 8 \times 10$ cm covering most of the LHCb acceptance.

The tracking stations

The tracking stations downstream of the VELO are designed to provide measurements of the trajectories of charged particles before and after the magnet, to allow measurement of their momenta. The tracking system is composed by four planar stations, orthogonal to the beam axis: the Trigger Tracker (TT) and the T1, T2, T3 stations. The TT and the inner regions of T1-T3, which formed the Inner Tracker (IT) [96], are subject to very high particle flux, thus they must be very radiation hard, and have sufficiently high granularity as to keep occupancies low enough for reliable pattern recognition. For these reasons they use silicon microstrip sensors with a

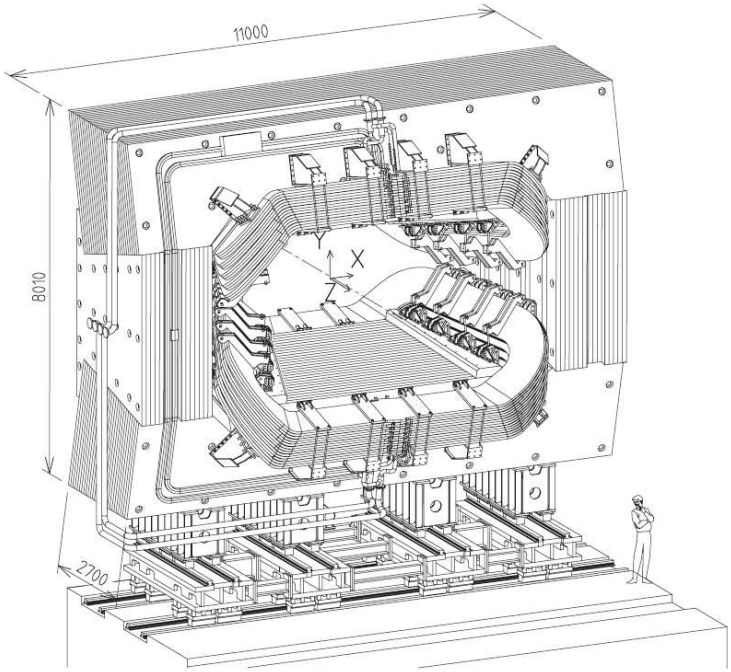


FIGURE 2.16: A diagram of the LHCb magnet.

strip pitch of about 200 μm. Each one of the silicon tracker stations has four detection layers of three different type: the x-type layer has vertical strips, the u- and v-type stereo layers have strips rotated respectively to -5° and $+5^\circ$ with respect to the y axis. The arrangement scheme followed is x-u-v-x starting from the layer closest to the interaction point. The TT station is located downstream of the dipole magnet, 150 cm wide and 130 cm high, it covers the full acceptance of the experiment. Given the proximity to the magnet, the TT station is affected by a residual magnetic field.

The IT located in the centre of T stations is 120 cm wide and 40 cm high. The T stations are placed upstream of the magnet and the acceptance region covered by the IT is approximately 1.3% of the LHCb acceptance. The single hit spatial resolution has been determined to be around 50 μm for both the TT and the IT.

The external region of T stations is called Outer Tracker (OT) [97] and it suffers significantly less irradiation, so cheaper array of straw tube drift-time sensors are used. The OT is designed as an array of straw-tube modules. The boundary with the IT is chosen to limit the occupancy to less than 10% at the nominal LHCb luminosity. The four layers of each OT station follow the same geometry as the IT with the inner two layers rotated by $+5$ and -5 . To ensure a maximum drift time of 50 ns (the time taken for two bunch proton crossing), the tube contain a gas mixture of 70% argon and 30% carbon dioxide, and have an inner diameter of 4.9 mm. Each module is formed by two staggered layers of drift-tubes with an inner diameter of 4.9 mm. The tubes are filled with counting gas, a mixture of Argon (70 %), CO_2 (28.5%) and O_2 (1.5 %).

These are chosen to guarantee a sufficient drift distance resolution of about 200 μm. The OT covers an acceptance of 300 mrad (250 mrad) in the horizontal (vertical) plane.

A schematization of the tracking system with the OT, IT and TT detectors in evidence is presented in Figure 2.17.

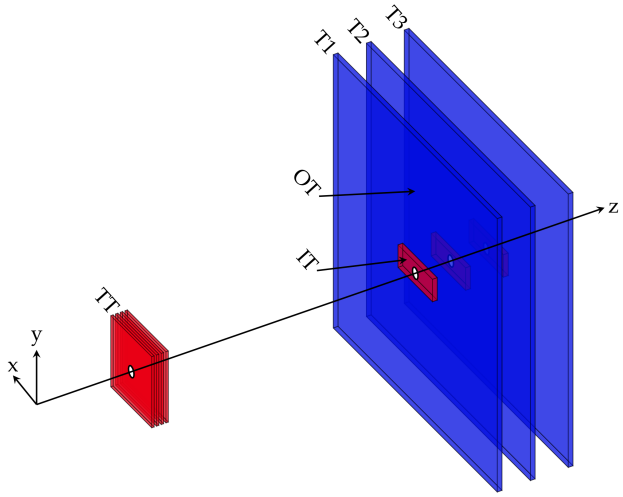


FIGURE 2.17: Layout of the LHCb tracking system.

Tracking performance and track reconstruction

Pattern recognition algorithms are used to reconstruct the track trajectory coordinates in all the tracking sub-detectors. In order to find the best estimate of the tracks parameters a Kalman fit [98] is performed. The tracks are classified in four categories starting from the hits:

- *long* tracks, which have hits in the VELO and in all the T-stations;
- *downstream* tracks, which have hits in the VELO and in the TT station but not in the T1, T2 and T3 stations;
- *upstream* tracks, which have hits in all the T-stations but not in the VELO;
- VELO tracks, which have hits in the VELO only.

The first step of the track reconstruction is the pattern recognition in which independently in the VELO and in the T stations sequences of hits are grouped together and identified as coming from the same track. These VELO tracks and T tracks are then used as input to find long, upstream and downstream tracks. The long tracks are found either by extrapolating the VELO tracks into the T station and looking for matching hits, or by matching directly the VELO tracks with T tracks; TT hits are then added. Downstream tracks are found by extrapolating VELO tracks into the TT while the Upstream tracks combine T tracks with TT information.

The momentum resolution of long tracks has been measured using a data samples of $J/\psi \to \mu^+\mu^-$ decays, collected with a trigger configuration that selects couples of high energy muons (dimuons) [99]. The resolution as a function of the muon track momentum is shown in Figure 2.18. It is of about 0.5% for momentum particles below 20 GeV/c and 0.8% for momentum particles around 100 GeV/c. The invariant mass resolution has been studied using six resonances observed in the dimuon data sample: J/ψ, $\psi(2S)$, $\Upsilon(1S)$, $\Upsilon(2S)$, $\Upsilon(3S)$ mesons and Z boson [99].

The primary vertex is reconstructed from the detected tracks. First tracks are clustered in seeds, using the algorithm described in [100]. Then for each seed the PV position is calculated by minimizing:

$$\chi^2_{PV} = \sum_{i=1}^{n_{tracks}} \frac{d^2_{0i}}{\sigma^2 d_{0i}}, \tag{2.8}$$

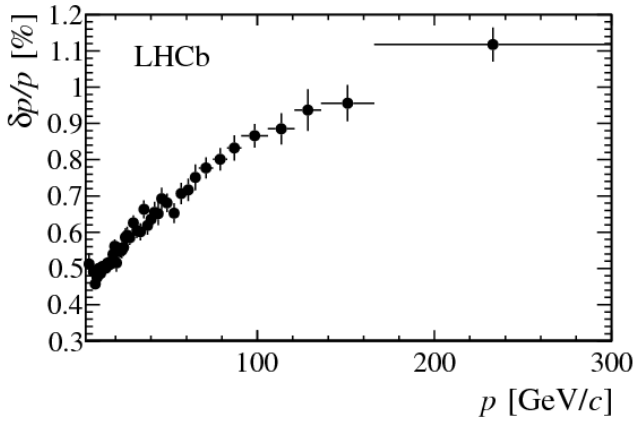


FIGURE 2.18: Relative momentum resolution as a function of the particle momentum, measured in data using muon tracks from J/ψ decays.

where d_0 is the impact parameter of the track and σ_{d_0} its error. If one or more tracks have $d_0/\sigma_{d_0} > 4$ then the one with the highest d_0/σ_{d_0} is removed from the cluster, and a new PV position is calculated by minimizing χ^2_{PV}. If the cluster has less then 6 tracks then it is discarded. The procedure stops if there are no more tracks to discard.

The PV reconstruction efficiency and resolution have been obtained in simulation [100] and it depends from the tracks multiplicity. The average efficiency goes from 90.0% to 97.5%, with a probability to reconstruct a false PV of about 1%. The resolution is of about 8 μm, 10 μm and 50 μm for the x, y and z coordinates respectively.

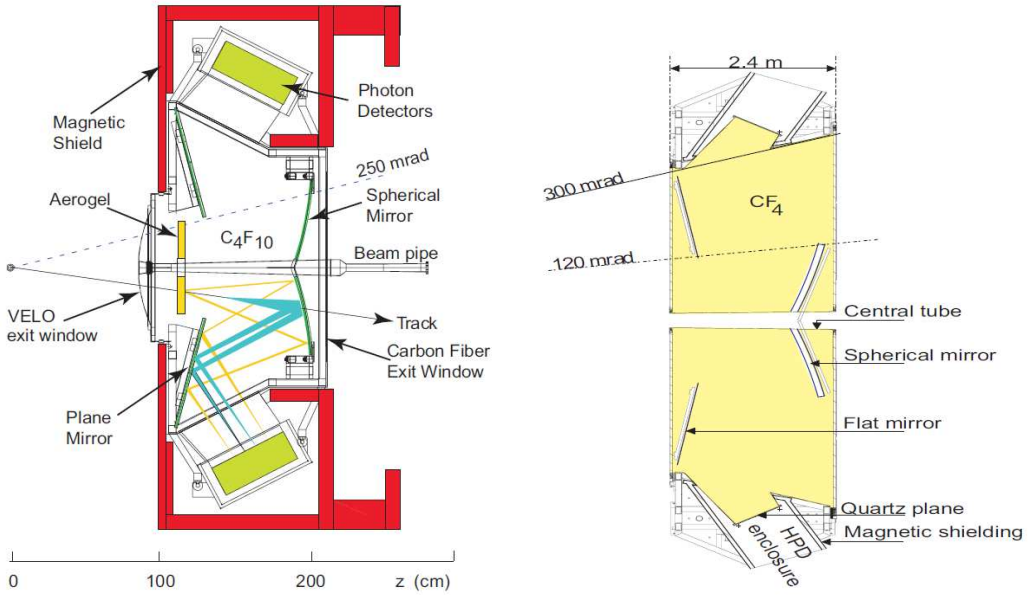
Particle identification

One essential design feature of the LHCb detector is its ability to distinguish different final state charged particles. This is especially important for discriminating between decays that are topologically equivalent but differ only by the species of their final state charged particles. Particle IDentification (PID) is achieved at LHCb by the use of two Ring Imaging CHerenkov (RICH) detectors, a calorimeter system and a muon detector, designed to cover all of the common charged particles (e, μ, π, K, p).

Ring imaging Cherenkov detectors

The two Ring Imaging Cherenkov (RICH) detectors at LHCb provide particle identification for the experiment. There are two RICH detectors in LHCb as shown in Figure 2.19: RICH1 [101] is positioned before the magnet and is designed to perform particle identification (PID) for low momentum particles; RICH2 [101] is situated downstream of the magnet, and is designed to perform PID for high momentum particles. The momentum range covered depends on the radiator material used: RICH1 uses aerogel, with $n = 1.03$, and C_4F_{10} gas, with $n = 1.0014$; while RICH2 uses CF_4 gas, with $n = 1.0005$.

For a given p each different species of charged particle will produce a ring with a different radius. Thus knowing the p of a given track, one can compare the expected rings with the photons observed and so infer the species of the particle that made the track.



(a) (b)

FIGURE 2.19: (a) The RICH1 detector in the x-y plane. (b) The RICH2 detector in the x-z plane (top view).

Such a process has a certain rate of mis-identification, whereby it identifies a track as being of a certain species other than its true identity. The efficiency of the PID algorithm can be checked, on real data, by using decay channels for which the kinematics of the decay products are sufficient to identify them without using the RICH detectors. As an example, the decay $D*+ \rightarrow D^0(K\pi)\pi^+$ is used for $\pi - K$ separation, as applying a tight constraint on $m(D^{*+}) - m(D^0)$ is sufficient to select a very clean signal sample. Figure 2.20 shows the efficiency, as a function of momentum, of correctly identifying a K as a K, and wrongly identifying a π as a K.

The calorimeters system

The LHCb calorimetry system [102] adopts the classical layout of an Electromagnet Calorimeter (ECAL) followed by a Hadronic Calorimeter (HCAL). This system provides several functions:

- it selects hadrons, electrons and photons with significant transverse energies at the first level trigger (L0), making a decision 4 μs after the interaction;
- it provides the identification of electrons, photons and hadrons;
- it measures the energy of neutral particles, such as photons, π^0 and neutral hadrons, which is a fundamental feature for jet reconstruction.

The calorimeters system is composed of a scintillating pad detector (SPD), a preshower detector (PS), an electromagnetic calorimeter (ECAL) and a hadronic calorimeter (HCAL) located in this order downstream of the RICH2 and M1 muon station, before the M2-M5 stations.

The SPD and the PS are two almost identical planes of scintillator pads, separated by a 15 mm thick lead converter, equal to 2.5 radiation lengths. The role of the SPD is

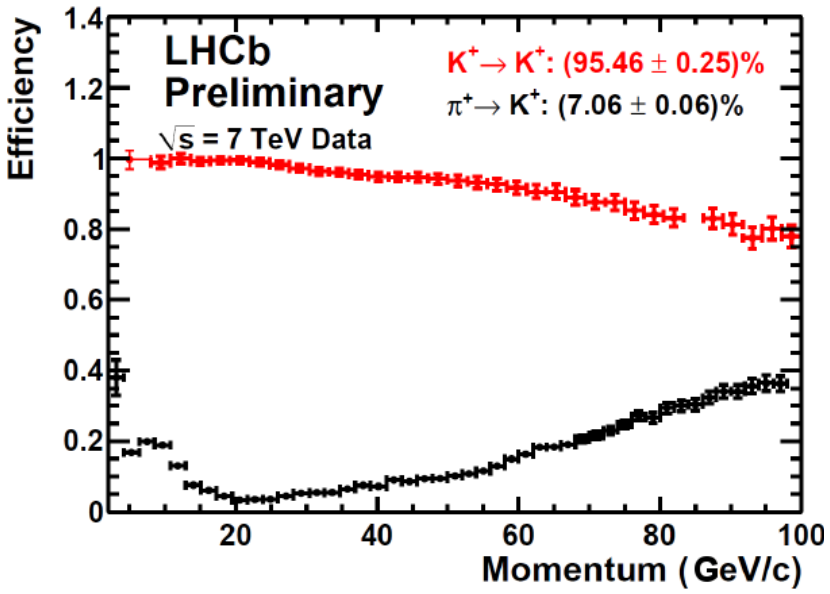


FIGURE 2.20: The efficiency, as a function of particle momentum, with which the RICH detectors correctly identify a K as a K, and wrongly identify a π as a K.

to detect charged particles, and when used together with the ECAL, provides rejection of π^0 and γ backgrounds to e^- signal candidates. In addition, SPD information is used by the Level-0 trigger to reject high multiplicity events. The PS is designed to detect electromagnetic particles which shower in the lead plate (primarily e^- and γ because of their comparatively short interaction lengths) and is used in conjunction with the ECAL to reject $\pi^\pm$ backgrounds to e^- signal candidates. At trigger level a cut on the maximum number of hits in the SPD ($nSPD$) is applied. This requirement is called Global Event Cut and it is used to reject events with multiple interactions.

The ECAL is placed 12.5 m from the interaction point and its acceptance is $\pm$300 mrad horizontally and $\pm$250 mrad vertically, limited in the inner region around the beam pipe at $\pm$25 mrad. A ECAL cell is built alternating layers of 2 mm thick lead tiles and 4 mm scintillator tiles. In depth 66 layers form a 42 cm stack, corresponding to 25 radiation lengths. As for the SPD/PS the scintillation light is transmitted by WLS (wave lenght shifting) fibres and read by PMTs. Photons and electrons deposit their energy in the ECAL, where they are absorbed. The segmentation of the ECAL cells achieved a one-to-one projective correspondence with the SPD/PS pads. The granularity depends on the distance from the beam axis to guarantee an optimal detector occupancy, as shown in figure 2.21. The ECAL energy resolution is given by

$$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus 1\%, \tag{2.9}$$

where the E is the energy measured in GeV and $\oplus$ indicates the sum in quadrature.

The HCAL is made of iron and scintillating tiles, as absorber and active material respectively. In the HCAL neutral and charged hadrons deposit their remaining amount of energy. It is placed 13.3 m from the interaction point covering the same ECAL angular acceptance. The orientation of the scintillating tiles run parallel to the beam axis, and they are interspersed with 1 cm of iron, reaching in total 5.6 nuclear interaction lengths of hadrons in iron. Again the scintillation light is transmitted by WLS fibres and read by PMTs. The HCAL granularity is different from ECAL and it

is shown in Figure 2.21. The HCAL energy resolution is given by

$$\frac{\sigma_E}{E} = \frac{69\%}{\sqrt{E}} \oplus 9\%, \tag{2.10}$$

where the E is the energy measured in GeV. The small HCAL nuclear interaction lenght gives as a result the not so very good resolution (constant term).

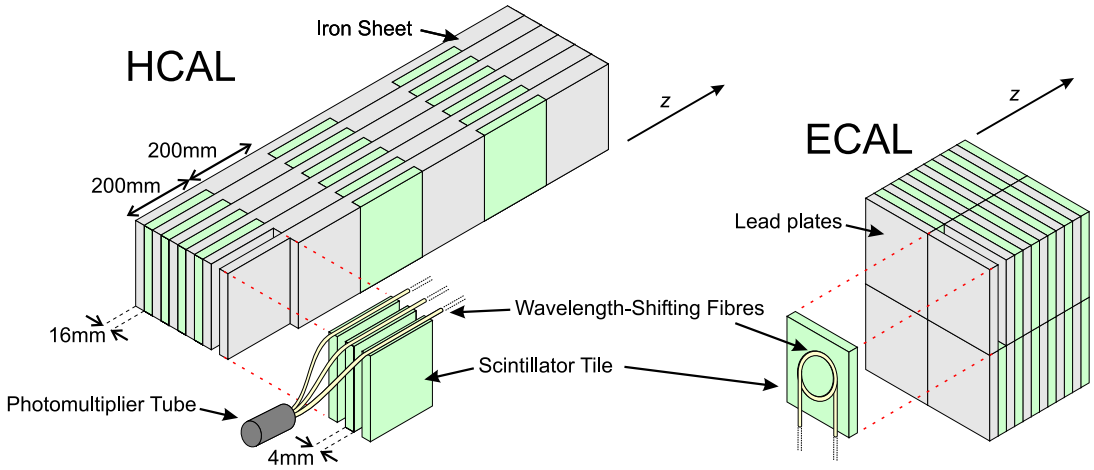


FIGURE 2.21: The internal structure of the HCAL (left), and ECAL (right).

Muon chambers

Five muon stations [103] (M1-M5) are used to provide reconstructed muon tracks. Muon detection is vital for any analyses which contain one or more muon. In addition, the muon stations are used to search for high transverse momentum tracks for the Level-0 trigger. The M1 station is placed before the calorimeters, while the stations M2-M5 are located after the calorimeter, as shown in Figure 2.22. To reach the M5 station, a muon must have a momentum of at least 6 GeV/c.

All of the muon stations use Multi-Wire Proportional Chambers (MWPC), except for the inner region (R1) of M1 (where the particle flux is too high) which employs triple-GEM (Gas Electron Multiplier) detectors, chosen because of their higher radiation tolerance. Both types of chamber use a mixture of $Ar - CO_2 - CF_4$ gas.

Muon system performance

The layers in each muon chamber are taken as a logical OR to determine the presence of a muon. In doing so the GEMs achieve an efficiency of more than 96%, while the MWPCs achieve an efficiency of more than 95%. The correct PID rate is $> 95\%$ and the mis-ID rate is just a few % for all momenta, demonstrating the excellent performance of the muon system.

2.2.3 The trigger system

The LHCb trigger is organized in three levels: the hardware Level-0 (L0) trigger and the software High Level Triggers HLT1 and HLT2. The software trigger has been redesigned during the shutdown between the Run I and the Run II, to allow the most possible wide and precise physics program. This shutdown also gave the

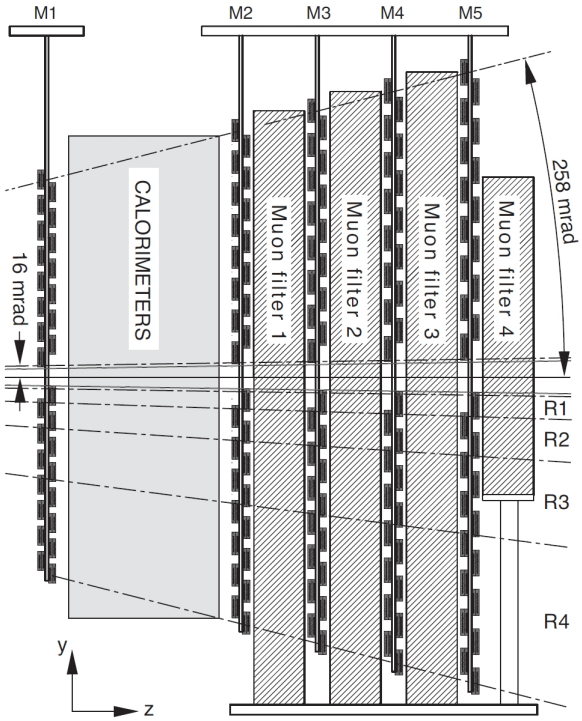


FIGURE 2.22: Side view of the LHCb muon system, showing the position of the five stations.

possibility to implement a new set of triggers dedicated to the on-line selection of jets, and in particular of tagged jets, which was missing in the Run I data taking.

Level 0 trigger

Events with b hadron decays are characterized by particles with high transverse momentum (p_T) or high transverse energy (E_T). The Level 0 trigger uses the calorimeters and the muon system detector to spot these signatures. The calorimeters and the muon system are connected to the Level 0 Decision Unit (L0-DU) that decides if the event is accepted, reducing the rate from 40 MHz to 1 MHz. To reject events with multiple interactions, which need long time to be processed, L0 applies a cut on the number of the hits in the SPD detector (Global Event Cut, GEC). Then the event is accepted if one of these conditions is satisfied:

- *L0-Muon*: a track identified as a muon with a p_T above a given threshold is found;
- *L0-Dimuon*: two track identified as muons with a $p_T^1 \cdot p_T^2$ above a given threshold are found;
- *L0-Hadron*: a HCAL cluster with a E_T above a given threshold is found;
- *L0-Photon*: a ECAL cluster with a E_T above a given threshold is found;
- *L0-Electron*: a ECAL cluster with a E_T above a given threshold anticipated by hits in the PS and at least 1 hit in the SPD is found.

The L0 thresholds have been decided to maximize the trigger efficiencies of benchmark decays. The trigger thresholds used in the 2012 data taking are reported in Table 2.2.

TABLE 2.2: Level 0 trigger thresholds used during the 2012 data taking.

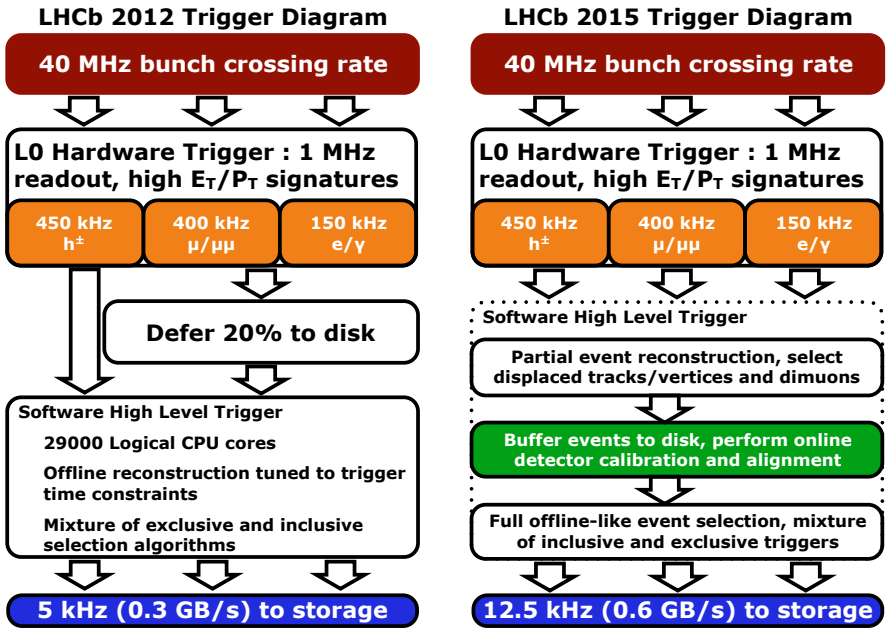
line	p_T [GeV/c]	$p_T^1 \cdot p_T^2$ [GeV/c^2]	E_T [GeV]	$nSPD$
L0-Muon	> 1.8	-	-	< 600
L0-DiMuon	-	$> (1.5)^2$	-	< 900
L0-Hadron	-	-	> 3.7	< 600
L0-Photon	-	-	> 3	< 600
L0-Electron	-	-	> 3	< 600

The High Level Trigger in Run II

Figure 2.23 shows the comparison between the Run I and the Run II trigger scheme. In Run I the output rate of events written to permanent offline storage was 5 kHz. Almost all events accepted by the HLT were recorded in a format containing all raw information from the detector. In order to be performed at the full bunch crossing frequency in real time, the event reconstruction performed in the trigger was simpler than the one used offline. The detector calibration and alignment parameters were obtained offline on triggered data afterward, when all the events were reconstructed offline to achieve the best performance regardless of the timing required. This strategy makes inefficient use of the computing resources and could also cause difficulties in understanding efficiencies due to the differences between online and offline reconstructions.

For Run II, a new trigger scheme has been developed to cope with the almost double visible rate, see Figure 2.23 on the right. The major change to the trigger strategy is that the processing of the second step of the software trigger (HLT2) has been completely deferred, in order to optimize the usage of the event filter farm (EFF) [104] where the software trigger is run. The EFF consists of approximately 1700 nodes with 27000 physical cores; 800 of which newly installed for Run II. The EFF hosts 5 PB of disk space which is exploited as a buffer, distributed in such a way that farm nodes with faster processors get a larger portion of the disk buffer. At an average event size of 55 kB passing HLT1, this buffer allows for approximately 150 hours of HLT1 datataking before HLT2 has to be executed.

The increased computing power in the EFF allows to perform low transverse momentum tracking without impact parameter cuts in HLT1 and the possibility to perform full offline quality reconstruction upfront in HLT2. The first has led to the development, for the first time at a hadronic collider experiment, of a set of HLT1 lines without any lifetime biasing cuts [105]. Physics analyses which can particularly benefit are those which measure the lifetime of B and D mesons which decay into hadronic final states. The latter, combined with an automated alignment and calibration tasks [106] of the subdetectors in real time using data from HLT1, leads to offline quality information inside the trigger software. Offline quality information at the trigger level means that in Run II it is possible to perform physics analyses directly with the information calculated by the HLT event reconstruction using the special stream of data known as Turbo stream [107]. The information calculated by the trigger is used to directly perform physics measurements without the need for further offline reconstruction, saving CPU and storage resources.



(a) (b)

FIGURE 2.23: The LHCb trigger scheme in Run I(a) and in Run II (b)

At HLT2 level the b hadrons are selected by inclusive triggers, based on multivariate techniques, which look for a two-, three-, or four-track vertex with significant transverse momentum, significant displacement from the primary vertex and a topology compatible with the decay of a b hadron. For Run II these algorithms have been reoptimized [108] and the relative gain in efficiency with respect to the Run I topological trigger is expected to be between 10% and 60% depending on the decay mode. For what concern c hadrons, many exclusive selections have been implemented fully exploiting the offline quality information already available in the trigger. As already anticipated, for the Run II a new set of triggers for the collection of jets have been introduced. The one requiring two tagged jet in the final state is used to set the limit on the inclusive $H \to b\bar{b}$ process and will be described in details in Section 3.3.3.

Offline quality information at the trigger level is the first step toward an upgraded trigger system [109] which, thanks to some sub-detectors upgrades, will be able to collect data at ten times its initial design luminosity. The readout architecture will be upgraded by removing existing L0 hardware trigger whose efficiency is limited for hadronic channels at high luminosity. The new readout system will record every LHC bunch crossing and send data to a trigger selection process performed entirely by software running in a computing farm. The installation and commissioning of the upgraded detector and readout system is planned for the LHC Long Shutdown II (2018–2019).

Simulation

The production of Monte Carlo simulated samples at LHCb is centralized and it is done through some softwares based on ROOT [110] and on the Gaudi framework [111]. In the first phase the proton proton interaction is simulated using a specific software, GAUSS [112], which makes use of different generator softwares

like PYTHIA [113], POWHEG [114] or ALPGEN [115] and uses EvtGen [116] to simulate B mesons decay; the interaction of particles with the detector and beam pipe materials is simulated with GEANT4 [117].

The response of the different sub-detectors, including all the instrumental effects like the readout electronic noise and inefficiencies, are then simulated by the BOOLE [112] software through GEANT4 [117].

Chapter 3

Jets at CDF II and LHCb

We are interested in final states with b quarks. But how it was described in Section 1.2.1, because of the confinement quarks show themselves only through the particle produced in the hadronization. They manifest themselves in detectors as collimated streams of charged particles in the tracker, or as concentrated energy depositions in the calorimeter and they are called jets.

In this chapter, after a brief description of the different jet definitions that have been developed (Section 3.1), the reconstruction and the identification algorithms implemented in the on-line and off-line software of the CDF II (Section 3.2) and of the LHCb (Section 3.3) experiments are described.

3.1 The jets

The link between the jets of hadrons measured in experiments and jets of partons for which theoretical results are obtained is not trivial. One source of ambiguity comes from the hadronization, which is a non-perturbative process and cannot be controlled precisely in theoretical calculations. The other reason is that jets at hadron colliders are always produced in a very busy environment and full theoretical control over the radiation prior to, or following, the hard scattering is impossible. This is why a precise and robust jet definition is needed.

An accord as to some general properties of jet definitions, the "Snowmass accord", was set out in 1990 [56] by a group of influential theorists and experimenters, and reads as follows:

> Several important properties that should be met by a jet definition are:
>
> 1. Simple to implement in an experimental analysis;
> 2. Simple to implement in the theoretical calculation;
> 3. Defined at any order of perturbation theory;
> 4. Yields finite cross sections at any order of perturbation theory;
> 5. Yields a cross section that is relatively insensitive to hadronization.

Different algorithms have been developed with the aim of clustering jets following this prescription, and in particular to build jets infrared-safe and collinear-safe. Infrared and collinear (IRC) safety is the property that if one modifies an event by a collinear splitting or the addition of a soft emission, the set of hard jets that are found in the event should remain unchange. Figure 3.1, from [57], schematically shows the importance of a IRC-safe jet definition: IRC-unsafe jet definition brakes singularity cancellation between the real and virtual diagrams and leads to an infinite cross section.

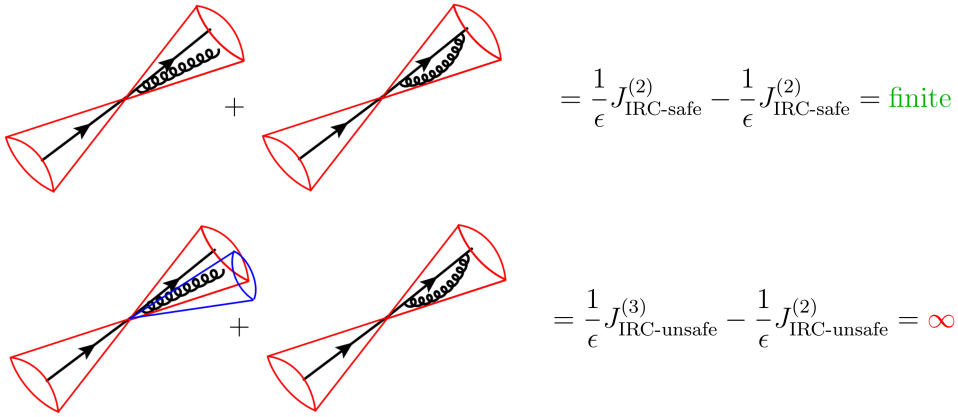


FIGURE 3.1: Importance of the infrared and collinear safety of a jet algorithm. In the top row, an IRC-safe jet definition is used and the corresponding cross section is finite at any perturbative order. In the bottom row, an IRC-unsafe jet definition which leads to an infinite cross section.

Cone algorithms

Cone algorithms have been historically the first way to form the jets. They work by associating particles with trajectories inside a cone of radius R in (η, ϕ)-space around a seed which exceeds a given threshold.

Cone algorithms with initial seed are computationally efficient but they suffer for IRC unsafety. The problems originate from the need to use seeds that are identified with final state particles. Such procedure is manifestly IRC-unsafe, as an emission of a soft or collinear parton changes the set of initial seeds, which in turn, for a non-negligible fraction of events, leads to a different set of the final-state jets. Improvements to overcome the IRC unsafety have been proposed, like the Midpoint algorithm, which introduces seeds in the "midpoints" of the stable cones and iterate the algorithm. A definitive solution that assures infrared and collinear safety is represented by the Seedless-Cone algorithms which avoid the use of seeds and iterations but are heavily time-consuming. Quite recently, a new seedless algorithm (SISCone) [58] that has speeds comparable to the seeded cone algorithms has been developed.

Sequential recombination algorithms

The sequential recombination algorithms dominate almost exclusively in the jet measurements at the LHC. They represent a bottom-up approach by starting to combine the closest particles, according to a distance measure which can be generally written as:

$$d_{ij} = min(k^{2p}_{T,i}, k^{2p}_{T,j}) \frac{\Delta R^2_{ij}}{R^2}, \qquad d_{iB} = k^{2p}_{T,i} \tag{3.1}$$

where d_{ij} is a distance between the particles i and j and d_{iB} is a distance between the particle i and the beam. $k_{T,i}$ is the transverse momentum of the particle i, $\Delta R^2_{ij} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}$ is called radius parameter and it is the geometrical distance between the particles i and j in the rapidity-azimuthal angle plane. The parameter p defines the algorithm from the sequential-recombination family: $p = 1$ for the k_T

algorithm [59], $p = 0$ for the Cambridge/Aachen algorithm [60] and $p = -1$ for the anti-k_T algorithm [61].

Each procedure of finding jets with the sequential-recombination algorithm consists of the following steps:

1. Compute distances between all pairs of final-state particles, d_{ij}, as well as the particle-beam distances, d_{iB}, using the measure from Equation 3.1.
2. Find the smallest d_{ij} and the smallest d_{iB} in the sets of distances obtained above.
   - If $d_{ij} < d_{iB}$, recombine the two particles, remove them from the list of final-state particles, and add the particle ij to that list.
   - If $d_{iB} < d_{ij}$, call the particle i a jet and remove it from the list of particles.
3. Repeat the above procedure until there is no particles left.

The k_T algorithm starts from clustering together the low-p_T objects and it successively accumulates particles around them. The C/A algorithm is insensitive to the transverse momenta of particles and it builds up jets by merging particles closest in the $y\phi$ plane. The anti-k_T algorithm starts from accumulating particles around high-p_T objects, just opposite to the behaviour of the k_T algorithm. In the anti-k_T algorithm, the clustering stops when there is nothing within radius R around the hard center. For that reason, anti-k_T leads to jets that take circular shapes in the $y\phi$ plane. For this reason the anti-k_T algorithm became a default choice at the LHC. In fact jets with regular shapes allow for reliable interpolation between detector regions separated by dead zones.

3.2 b-jets at the CDF II

3.2.1 The off-line jet object

Jets at CDF II are reconstructed from clusters of calorimetric towers using the JetClu [62] iterative algorithm with a fixed cone radius parameter $R = 0.7$. At first, all the towers with E_{T}^{i} greater than 1 GeV are considered seed towers:

$$E_{\mathrm{T}}^{i} = E_{EM}^{i} \sin\theta_{EM}^{i} + E_{HAD} \sin\theta_{HAD}^{i} \tag{3.2}$$

Starting with the seed tower with the highest transverse energy, the towers within $R = 0.7$ from it in the (η, ϕ) space are gathered together to form a pre-cluster. This procedure is repeated, starting with the next unused seed tower, until the list is exhausted. The E_{T} weighted centroid is then formed from the towers in the pre-cluster and a new cone is formed around this centroid. All towers with energy above the lower threshold of 100 MeV within this new cone are added to the cluster. Then, a new centroid is calculated from the set of towers within the cluster and a new cone drawn. This process is iterated until the centroid of the energy deposition within the cone is aligned with geometric axis of the cone (stable solution).

Since each tower can belong to only one jet, in case of jet overlap two clusters are merged if the total energy of the overlapping towers is greater than 75% of the energy of the smaller cluster. If the shared energy is below this cut, the shared towers are assigned to the cluster that is closer in $\eta - \phi$ space. This process is iterated again until the list of clusters remains fixed.

Massless four-vector momenta are assigned to the towers in the clusters for EM and HAD components with a magnitude equal to the energy deposited in the tower and the direction defined by a unit vector pointing from the event vertex to the center of the calorimeter tower at depth that corresponds to the shower maximum. The cluster location and the transverse energy are calculated with the following definitions:

$$E_{\mathrm{T}}^{\mathrm{jet}} = \sum_i E_{\mathrm{T}}^i \tag{3.3}$$

$$\phi^{\mathrm{jet}} = \sum_i \frac{E_{\mathrm{T}}^i \phi_i}{E_{\mathrm{T}}^{\mathrm{jet}}} \tag{3.4}$$

$$\eta^{\mathrm{jet}} = \sum_i \frac{E_{\mathrm{T}}^i \eta_i}{E_{\mathrm{T}}^{\mathrm{jet}}} \tag{3.5}$$

Jet energy correction

The direct measurement provided by the calorimeter, commonly referred as raw energy, requires several corrections to take into account for detector and physics effects, such as the presence of inactive material, the non-compensating nature of the calorimeter and multiple $p\bar{p}$ interaction per beam crossing. Furthermore, additional corrections need to be applied to associate the jet energy to the original parent parton.

Corrections used by the CDF collaboration [63], often referred to as jet energy scale (JES) corrections, include all the possible corrections due to both instrumental effects and to radiation and fragmentation effects and can be parametrized as follows:

$$E_{\mathrm{T}}(E_{\mathrm{T,raw}}^{\mathrm{jet}}, \eta) = (E_{\mathrm{T,raw}}^{\mathrm{jet}} \cdot \mathrm{L1}(E_{\mathrm{T,raw}}^{\mathrm{jet}}) - \mathrm{L4}) \cdot \mathrm{L5}(E_{\mathrm{T,raw}}^{\mathrm{jet}}) \tag{3.6}$$

where L1, L4 and L5 are the correction level factors. The corrections carried by each of the correction factors are described in the following:

- L1 refers to the correction of the calorimeter's non-uniform response along the η coordinate. A lower response arises in the poorly instrumented regions due to the physical separation between the different calorimeter components: at $\eta = 0$, where the two halves of the central calorimeter join, and at $|\eta| = 1.1$, on the border between the central and the plug calorimeters. The η-dependent correction is obtained by using the so called dijet balancing method, based on the assumption that the two leading jets in dijet events are equal in p_{T} in absence of hard QCD radiation. Corrections are determined separately for data and Monte Carlo and for different p_{T} bins.

- L4 represents the amount of energy coming from different $p\bar{p}$ interactions which take place in the same bunch crossing (pile up), that have to be adequately subtracted when falling within the cone of a cluster. This correction is derived from minimum bias data, by measuring the transverse energy contained in a random cone of radius $R = 0.7$ and by parameterizing the value as a function of the number of primary vertexes of the events. The number of $p\bar{p}$ interactions per bunch crossing follows a Poisson distribution and the mean $\langle N \rangle$ increases linearly with the instantaneous luminosity: it is one for $\mathcal{L} = 0.4 \times 10^{32}$ cm^{-2}s^{-1} and reaches the value of 8 for $\mathcal{L} = 3 \times 10^{32}$ cm^{-2}s^{-1}. Consequently this kind of correction has increased its relevance over the time of CDF data taking.

- at L5 the absolute correction aims to transform the measured jet energy into the energy corresponding to the underlying particle jet. The procedure is based on MC simulations and its accuracy depends on how well the response of the calorimeter to the single particle is modeled (calorimeter simulation) and on how well the multiplicity and the p_{T} spectra of the particles inside a jet are reproduced (fragmentation simulation).

After these correction the calorimeter jet is corrected to particle level, i.e. the energy of the jet corresponds to the energy of the physical particles resulting from hadronization process of the parton. Two more corrections, L6 and L7 are needed to reach the parton level. These two corrections are independent of the CDF detector and correct for the energy from the particles that originate from soft interactions of partons not involved in the hard interaction (called "underlying event") and for the energy that leaks outside the jet cone.

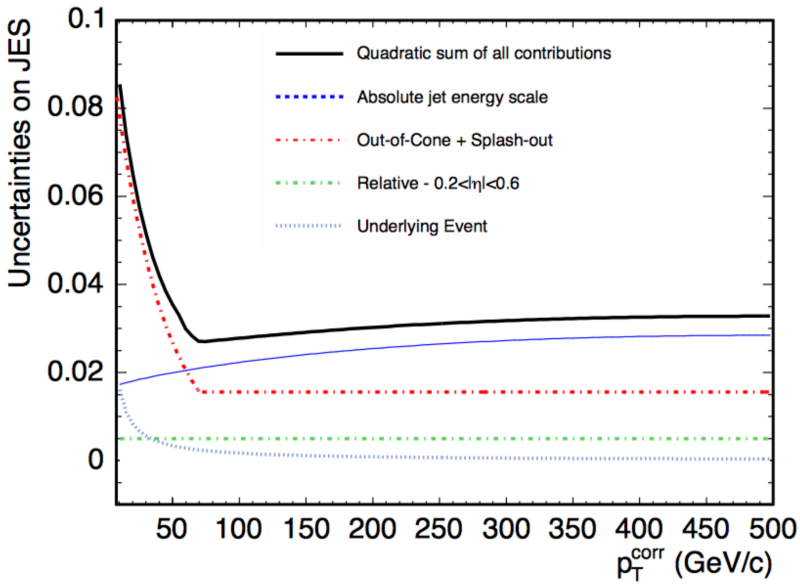


FIGURE 3.2: Systematic Uncertainties on the JES corrections as a function of the corrected jet p_{T} in the pseudorapidity range $0.2 < |\eta| < 0.6$.

The total systematic uncertainty on the final correction and the relative fraction of each correction is shown on Figure 3.2 as a function of the corrected p_{T} of the jet.

Identification of jets coming from b-quarks

The dataset collected with the DIJET_BTAG trigger is a sample enriched in b-jets, but for the purpose of maximizing the heavy flavour content of the final sample, an off-line b-tagging algorithm is used. This algorithm is called SecVtx [64] and exploits, as for the trigger, the long lifetime and high-mass of b hadrons which form a well displaced vertex from the primary $p\bar{p}$ interaction point.

SecVtx searches for secondary vertices in jets with uncorrected $E_{\mathrm{T}} > 15\,\mathrm{GeV}$ and pseudorapidity in the range $|\eta| < 2$. The algorithm first selects charged tracks within the jet cone that have been measured in the silicon detector with sufficiently good position accuracy and that have a large significance of their impact parameter with respect to the interaction point. The algorithm then uses these tracks in a recursive procedure to reconstruct a common point of origin for at least three of them. If

the reconstruction fails, tighter requirements are imposed on the tracks, and a fit accepting two-tracks vertices is attempted. Reconstructed vertices are rejected if their transverse distance from the interaction point corresponds to the location of material of the innermost silicon layer (1.2 cm $< r <$1.5 cm) or if it is greater than 2.5 cm.

Once a secondary vertex is found, the two-dimensional decay length L_{xy} is calculated as the projection onto the jet axis, in the $x - y$ plane, of the vector pointing from the primary vertex to the secondary vertex. Secondary vertexes corresponding to the decay of b hadrons are expected to have large positive L_{xy} while the secondary vertexes from random mis-measured tracks are expected to be less displaced from the primary vertex. To reduce the background from false secondary vertexes, a good secondary vertex is required to have $|L_{xy}/\sigma_{L_{xy}}| > 7.5$, where $\sigma_{L_{xy}}$ includes the error on the primary vertex and is of $\mathcal{O}(100\ \mu\text{m})$.

3.2.2 SecVtx and on-line b-tagging efficiency

The efficiencies of the SecVtx and the on-line b-tagging implemented in the DIJET_BTAG trigger for jets with $R = 0.7$ have been studied with an ad-hoc procedure. Results are compared to Monte Carlo (MC) simulation using a sample of generated events with a luminosity profile which reproduces the running conditions of the data in order to determine the data to MC scale factor.

The b-tagging efficiency ϵ_{tag} is defined as the fraction of tagged jets which is actually identified as being generated from the fragmentation of a b quark:

$$\epsilon_{\text{tag}} = \frac{N_b^{\text{tag}}}{N_b^{\text{tag}} + N_b^{\text{untag}}}. \tag{3.7}$$

In a MC dataset, the true b content of a jet sample can be easily determined at the generator level. Nevertheless, this quantity is difficult to estimate in data, since in this case the population of true b-jets is not a priori known.

The basic idea is to exploit the muon coming from the semi-leptonic decays B$\rightarrow \mu$X. The transverse momentum of the muon relative to the jet direction ($p_{\text{T,rel}}$) is large compared to that one coming from a generic muon. In fact, because of the large masess of the b hadrons, leptons from their deacys tend to have larger $p_{\text{T,rel}}$ than leptons from charm decays or fake secondary vertecies. Using this feature, the b fraction in a sample can be measured by fitting the $p_{\text{T,rel}}$ using two template distributions: one for muon coming from B decays and one for any other generic muon.

Once the b fractions in tagged (f_b^{tag}) and untagged samples (f_b^{untag}) are determined, the tagging efficiency can be easily calculated:

$$\epsilon_{\text{tag}}^{\text{data}} = \frac{f_b^{\text{tag}} \times N^{\text{tag}}}{f_b^{\text{tag}} \times N^{\text{tag}} + f_b^{\text{untag}} \times N^{\text{untag}}} \tag{3.8}$$

The data/MC scale factor, which parametrizes the differences between the simulated jets and the jets reconstructed in real data, is then defined as the ratio between the tagging efficiencies measured in data and in the simulation relative to the same physics process:

$$SF_{\text{tag}} = \frac{\epsilon_{\text{tag}}^{\text{data}}}{\epsilon_{\text{tag}}^{\text{MC}}} \tag{3.9}$$

The data used for this study has been collected with the MUON_CMUP8 and MUON_CMUP8_DPS trigger paths, which require a muon with a $p_{\mathrm{T}} > 8$ GeV/c, corresponding to an integrated luminosity of about 5.5 fb^{-1} of data. Three di-jet MC samples were generated using PYTHIA v. 6.216 [65], for a total of 25M events, with a lepton filter applied to the dijet generation to significantly increase the statistics of QCD heavy quark production available at the analysis level.

Two back-to-back ($\Delta\phi > 2.0$), L5 energy corrected jets are required: one jet referred as *muon jet* must contain a good not-isolated muon. The second jet, referred as "away jet", has to be "taggable", i.e. it must contain at least two good SecVtx tracks.

SecVtx efficiency and scale factor

First, the SecVtx off-line tagger efficiency have been evaluated both in data and in simulation. To get the flavour composition of the data sample, fits are performed using the $p_{\mathrm{T,rel}}$ templates. The template for muons coming from true b-jets is taken from the b MC sample by matching the *muon jet* to a b-quark at the generator level (ΔR(jet, b-quark)< 0.7). The non-b templates are taken from 3 different sources:

1. charm muon jets, extracted from the MC samples by matching the *muon jet* to a c-quark at the generator level (MC charm template);
2. muon jets from data, where the muon is not matched to a stub (anti-match template);
3. muon jets from data, where the number of SecVtx good tracks in both the *muon jet* and the away jet is zero (Pass1=0).

Figure 3.3 shows the $p_{\mathrm{T,rel}}$ distribution obtained with these three different sources compared with the $p_{\mathrm{T,rel}}$ distribution of the b tagged jets. The difference in the fit result using the three non b templates is taken as systematic uncertainty on the b-tagging efficiency due to the non-b component.

The numerical results of the fits are reported in Table 3.1. The b fraction and the number of tags are using in Equation 3.8 to give the SecVtx efficiency on data. The results are summarized in the first part of Table 3.1. The scale factors are then determined from Equation 3.9 and reported at the bottom of Table 3.1 and shown in Figure 3.4 as a function of the jet E_{T}. The uncertainty on the results includes the statistical and the systematic error.

TABLE 3.1: Summary of the SecVtx tagging efficiency on data and on simaluation, with data/MC scale factor. The three different data efficiencies values are evaluated using different non-b templates, as explained in the text.

	SecVtx
Data efficiency (MC charm template)	0.307± 0.003
Data efficiency (anti-match template)	0.296± 0.002
Data efficiency (Pass1=0 template)	0.326± 0.003
Average data efficiency	0.305± 0.001±0.015
MC efficiency	0.355±0.002
Scale factor	0.86±0.01±0.04

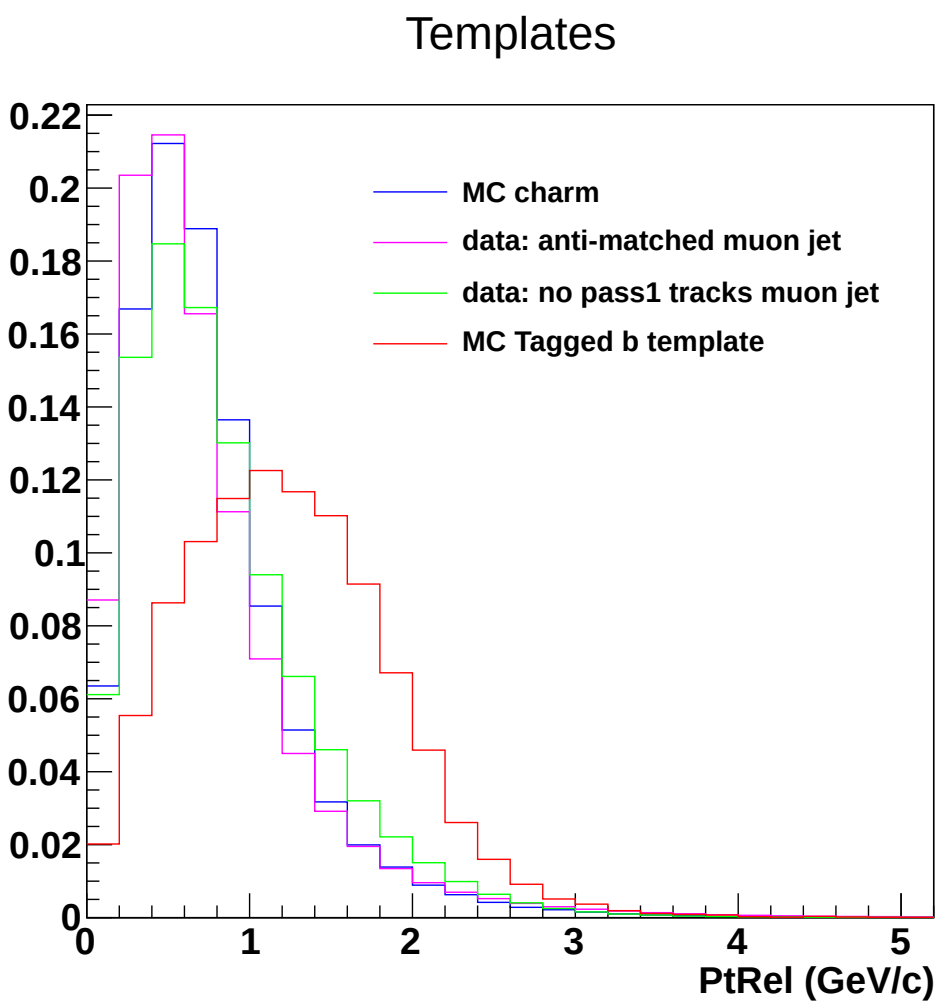


FIGURE 3.3: Comparison of $p_{\mathrm{T,rel}}$ distributions for b tagged MC jets and non-b jets obtained with different techiques as explained in the text.

DIJET_BTAG trigger efficiency and scale factor

The evaluation of the DIJET_BTAG trigger efficiency and of the relative data/MC scale factor is split in two parts: the request of two calorimetric clusters and the on-line b-tag of at least one jets. The data/MC scale factors are then separately calculated under the assumption that the scale factors are independent one from the other. The full scale factor is then represented by the product:

$$SF_{\text{trigger}} = SF_{\text{calo}}(\text{jet1}) \times SF_{\text{calo}}(\text{jet2}) \times SF_{\text{b-tag}}(\text{jet2}) \tag{3.10}$$

where the jet which satisfies the Level-2 on-line b-tagging is labeled as jet2.

The efficiency of the calorimeter jet requirements is calculated using the jets opposite ("away") to the muon-jet in the muon sample which does not have any trigger bias. It follows the characteristic turn-on curve, as shown in Figure 3.5 for data and MC, as a function of the raw jet E_{T}. The bin-by-bin ratio of the two efficiencies, the data/MC scale factor, is fitted to the following function:

$$SF(E_T) = A \times freq(B \times (E_T - C)) \tag{3.11}$$

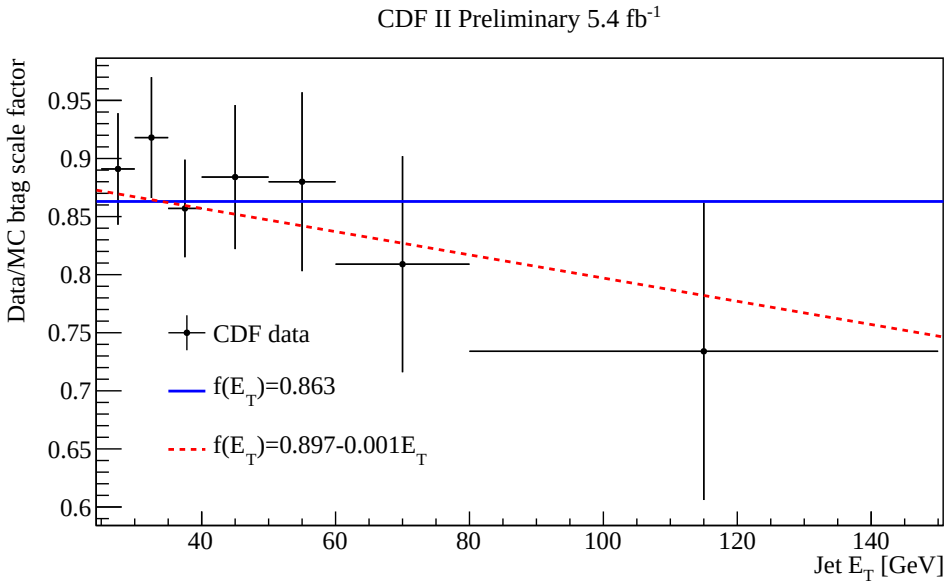


FIGURE 3.4: Data/MC scale factor distribution as a function of the jet transverse energy for SecVtx tags with result of the fit superimposed: in blue the constant line and in red the straight line.

where the frequency function is defined as

$$freq(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp -t^2/2dt \tag{3.12}$$

The results of the fit are

$$A = 0.9974 \pm 0.0005$$

$$B = 0.159 \pm 0.008$$

$$C = 10.8 \pm 0.5$$

Having provided a scale factor parametrization which is a function of the raw E_T, it can be used independently of the jet corrections which are applied at the off-line level.

The efficiency of the on-line b-tagging is evaluated together with the off-line SecVtx tagger. In fact, the on-line tagger affects the response of the SecVtx algorithm, and in this way there is no need to calculate a specific SecVtx scale factor for the individual jets which pass the trigger selection.

The on-line b-tagging has been emulated using off-line variables and the flavour composition of the data sample is obtained fitting the muon $p_{\mathrm{T,rel}}$ distribution. The trigger plus SecVtx efficiency, evaluated using Equation 3.8, is summarized in Table 3.2 using the three different non-b templates. The bottom part of Table 3.2 reports the average of the three measurements with the systematic error (the same sources considered also for the SecVtx alone efficiency determination), the MC efficiency and the data/MC scale factor. Figure 3.6 shows the data/MC scale factor for the combined trigger and SecVtx b-tag as a function of the jet E_{T}.

3.3 b-jets at LHCb

The LHCb environment is characterized by highly boosted jets and, in case of events with multiple jets, resolve and separate one from each other may be difficult. But jet

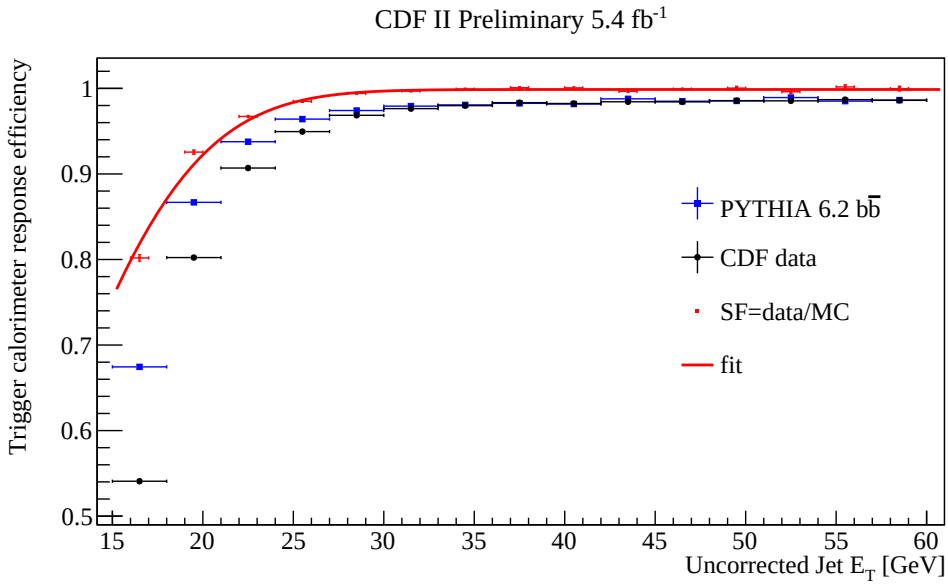


FIGURE 3.5: Calorimeter cluster-offline jet matching efficiencies and data/MC scale factor, as a function of the raw jet E_T.

TABLE 3.2: Summary of the combined on-line and SecVtx tagging efficiency on data and on simaluation, with data/MC scale factor. The three different data efficiencies values are evaluated using different non-b templates, as explained in the text.

Data efficiency (MC charm template)	0.466±0.015
Data efficiency (anti-match template)	0.444±0.010
Data efficiency (Pass1=0 template)	0.486±0.015
Average data efficiency	0.46±0.007±0.002
MC efficiency	0.724±0.01
Scale factor	0.68±0.01±0.03

physics in the forward region is of the primary interest because it probes a different kinematic region to that at the general purpose detectors, like the CDF II. The acceptance for heavy resonances decaying into jets pairs is low, but it gives the possibility to explore lower masses with respect to CMS and ATLAS. Electroweak measurements with jets instead, probe two Bjorken regions previously unexplored at low x and high Q^2, where x is the parton longitudinal momentum fraction and Q^2 the momentum transfer, giving an important input for the description of the Parton Distribution Functions.

3.3.1 Jet reconstruction

At LHCb, the calorimeter cells are coarse and saturate at high energies. On the other hand LHCb has an excellent tracking system, with an efficiency of about 97% on charged tracks with a momentum greater than 5 GeV/c and a resolution of about 0.5%. For these reasons it is necessary to involve both sub-systems, resulting in a hybrid algorithm which can exploit the tracking system to remedy the lack of performance of the calorimeters.

For jets with $E > 15$ GeV, the jets energy composition is given by [118]:

- charged particles which contribute for about 60% of the jet energy,

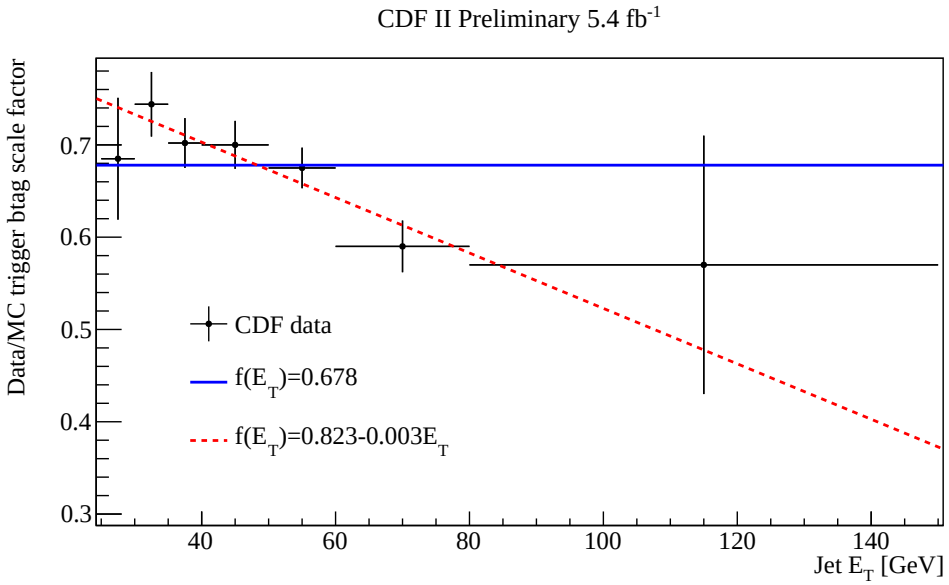


FIGURE 3.6: Combined trigger and SecVtx b-tag data/MC scale factor distribution as a function of the jet transverse energy with result of the fit superimposed: in blue the constant line and in red a straight line.

- photons and π^0 contribute to about 30% of the energy,
- neutral hadrons (neutrons and K_L^0 that reach the calorimeter) contributes to a bit less than 10%,
- neutrinos which only contribute for about 1%.

The particle composition of the Monte Carlo jets is reported in Figure 3.7 as a function of jet p_{T} and η, obtained from a minimum bias simulated sample at center of mass-energy of 7 TeV.

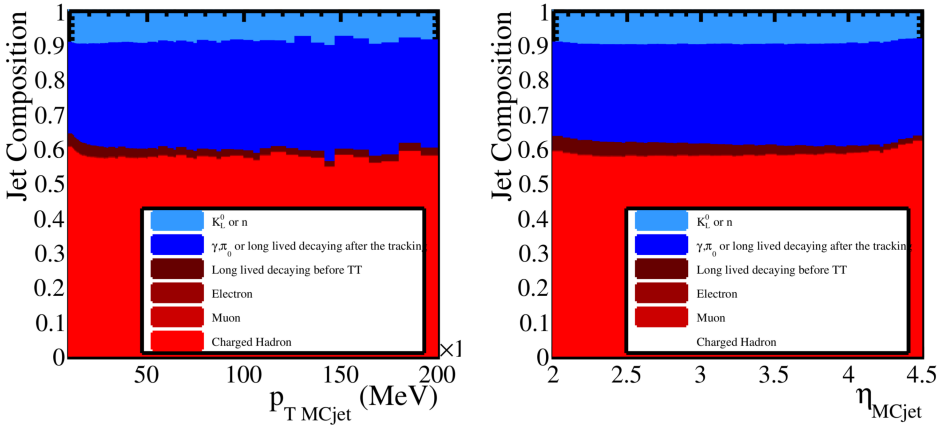


FIGURE 3.7: Composition of Monte Carlo jets as a function of jet p_{T} and η for minimum bias events.

The jet reconstruction algorithm at LHCb consists of mainly two stages:

1. the particle flow [119]: the first stage of the reconstruction, when the information from the tracking system and calorimeter clusters are selected as input for

the jet. Reconstructed objects of various types are taken as input, e.g. tracks, and an output particles are produced. The algorithm ensures that there is no overlap between the basic reconstructed objects that were used to build the particles output, i.e. tracks and calorimetric clusters. Additionally, an attempt is made to recover neutral energy which overlaps with charged particle deposits in the calorimeters.;

2. the jet clustering using the anti-k_T algorithm.

The particle flow

The particle flow algorithm is tuned to optimize the jet energy resolution, using information from tracks where possible, and only uses calorimeter information where no track information is available (for example, for neutron and photon contributions). The particle flow steps are the following:

The procedure start with different selection requirements that are applied to tracks depending on the category they belong: $long$, $downstream$, $upstream$ or VELO tracks (see Section 2.2.2). The cuts on the quality and on the p_{T} of the input tracks that are applied are listed in Table 3.3. The χ^2 is obtained from the Kalman fitter, the P_{ghost} is defined as the probability for a track to not being associated to a real particle and $\sigma(q/p)$ is the momentum resolution.

TABLE 3.3: List of requirements applied to tracks by the particle flow algorithm.

	$long$	$downstream$	$upstream$	VELO
p_T [MeV/c]	-	-	> 100	-
χ^2	< 5	< 1.5	< 1.5	< 10
P_{ghost}	< 0.5	-	-	-
$\sigma(q/p)(q/p)$	> 10	> 10	> 2	-

At this stage, tracks from particles that should not be included in the jet reconstruction, such as muon tracks from $Z \to \mu\mu$, are banned. Also the tracks corresponding to composite particles such as B, D, Λ and K_s^0 hadrons, are banned from further usage the and the composite particle is used as inputs instead.

Next, calorimetric clusters not associated to tracks are stored as inputs for the jets reconstruction. Neutral particles detected by the ECAL are mainly photons and π^0 decayed into two photons. Requirements are applied to select ECAL clusters isolated from the tracks and to ensure their quality before entering the jet clustering step. The following observables are considered:

- the likelihood for the photon hypothesis (PhotonID);
- the cluster transverse energy E_T;
- a χ^2 for each track-cluster combination ($\chi^2_{track-cluster}$): it evaluates how much the cluster is likely to be originated by the particle associated to the track.

The selection of isolated HCAL clusters is performed by applying different $\chi^2_{track-cluster}$ requirements for different cluster energy thresholds: a $\chi^2_{track-cluster}$ greater than 25 is required for HCAL clusters with energies below 10 GeV while a $\chi^2_{track-cluster}$ greater than 15 is required for energies above 10 GeV. No particle identification requirements are applied on the HCAL clusters selection.

Neutral recovery

A jet contains many particles produced at small angles with respect to each other. Simply taking isolated neutral particles will underestimate the number of neutral particles in a jet, as some will fall close to tracks. The main advantage of the particle flow approach is that it enables the recovery of this neutral content.

The neutral recovery [119] is performed on not isolated neutral particles and it is an iterative procedure:

1. The ECAL cluster with the best χ^2 track-match is found for the track of a given particle. If the χ^2 track-match is less than 25, the cluster is associated with the particle.

2. Step 1 is applied to the same particle, but now matching an HCAL cluster.

3. If the particle has an associated ECAL or HCAL cluster, a *SuperCluster* consisting of the particle and its associated clusters, is formed:

   a The deposited calorimeter energy 4-vector for the *SuperCluster* is the summed energy-weighted position 3-vectors ($\vec{x}_i^{\text{cal}}$) and energy (E_i^{cal}) of the clusters: $\vec{E}^{\text{dep}} = \sum_i E_i^{\text{cal}}(\vec{x}_i^{\text{cal}}, 1)$

   b The expected calorimeter energy (E^{exp}) for the particle is calculated as a function of momentum if it is an electron or charged hadron. It is evaluated using an empirical parametrization of E/p, where E is the cluster energy and p the track momentum.

   c The expected calorimeter energy 4-vector for the particle ($\vec{E}^{\text{exp}}$) is the deposited calorimeter energy 4-vector multiplied by $E^{\text{exp}}/E^{\text{dep}}$. If $\frac{E^{\text{exp}}}{E^{\text{dep}}} > 1.8$, then $\vec{E}^{\text{exp}} = (0,0,0,0)$ is set.

   If no association is found the algorithm restars from Step 1.

4. If there are more particles, the *SuperCluster* is stored and the algorithm returns to Step 1 for the next one. If there are no more particles the *SuperClusters* sharing any calorimetric clusters are iteratively merged until each one associated to a particle is uniquely assigned to a *SuperCluster*.

5. After merging, the energy-weighted 4-vector for each *SuperCluster* is calculated as $\vec{E}^{\text{spr}} = \vec{E}^{\text{dep}} - \vec{E}^{\text{exp}}$, if $E^{\text{spr}} < 0$ the *SuperCluster* is discarded.

6. All remaining *SuperClusters* are added to the particle flow output.

The parametrization of E/p used in step 3[b] to get E^{exp} as a function of p is obtained through a procedure called E/p calibration. The calibration for the Run I has been performed on a data sample of pp collisions at 7 TeV, collected using a minimum bias trigger configuration. From this sample isolated tracks matched to a calorimeter cluster have been selected. Requirements are applied to remove the background from minimum ionizing particles, like muons, that release a small amount of their energy in calorimeters:

Track-cluster objects are divided in different categories with the following features (track-cluster objects can belong to one or more of these categories):

- the track is associated to a cluster in ECAL but not in HCAL;

- the track is associated to a cluster in HCAL but not in ECAL;

- the track is associated to a cluster in ECAL, independently of the presence of a HCAL cluster;
- the track is associated to a cluster in HCAL, independently of the presence of a ECAL cluster;
- the track is associated to a cluster in HCAL;
- both a ECAL and a HCAL cluster are associated to the tracks.

A separate E/p calibration is performed for each of these categories. Moreover the calibration is done separately for tracks identified as hadrons or electrons. In the calibration procedure the mean value of E/p is computed in different intervals of p as the mode of the E/p distribution. Then the E/p dependence from p is fitted with empirical functions. For the first five categories listed above, where only one cluster from ECAL or HCAL is considered, the function used is:

$$\frac{E}{p}(p) = a_1 e^{-a_2 p} + a_3.$$

For the last category E is computed as the sum of ECAL and HCAL energies and the function used is:

$$\frac{E}{p}(p) = (a_1 + a_2 p + a_3 p^2) e^{-a_4 p} + a_5.$$

Few examples of this parametrization are reported in figure 3.8, compared with the E/p vs p obtained using a Monte Carlo sample of pp collision at 7 TeV with a minimum bias trigger configuration.

Jet clustering

The output particles selected by the particle flow are ready to be clustered into jets. At LHCb the clustering of the list of input particles is handled using the FastJet [120] interface which implements the anti-k_T [61] algorithm. This algorithm has the advantage of being infrared and collinear safe. Jets are reconstructed using a radius parameter of $R = 0.5$, which has been set in order to maximize some figures of merits, like the jet energy resolution or the jet identification efficiency.

If more than one PV is found in the event, charged particles made out of tracks with Velo segments only those associated to the considered PV are used. All neutral particles of the event made out of calorimeter clusters together with charged particles made of downstream track are used as inputs in all the procedures instead, since they contains no information on the interaction they belong too.

Jet energy correction

The reconstructed jets are corrected for calorimeter clusters response, energy losses, particles going outside of the detector or outside of the jet cone and energy contributions from noise and pile-up particles. At LHCb, this is done by using the Monte Carlo simulation.

The jet energy correction is a single factor that depends on the jet transverse momentum, pseudorapidity, azimuthal angle ϕ, fraction of charged particles in the jet cpf and number of primary vertices nPVs. It links the energy of reconstructed jets, $E(\mathrm{jet}_{reco})$, with the energy of associated Monte Carlo jets, $E(\mathrm{jet}_{MC})$:

$$E(jet_{MC}) = k_{MC} E(jet_{\mathrm{reco}}). \tag{3.13}$$

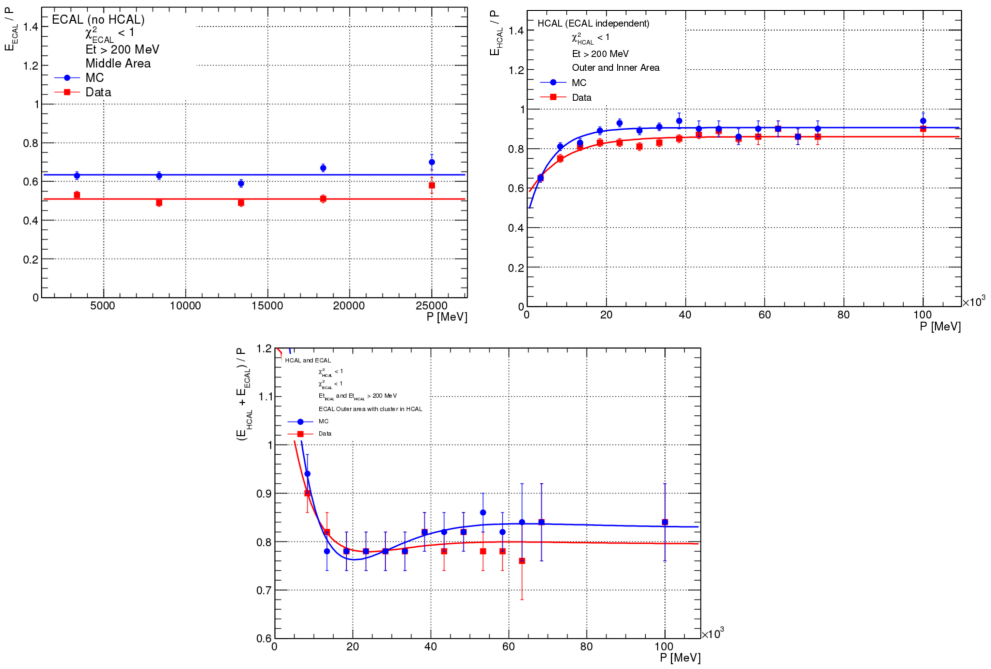


FIGURE 3.8: E/p calibration for hadrons. The plots show the calibration for different categories that are, from left to right and top to bottom: the category where a ECAL cluster is associated to a track but not to a HCAL cluster; the category where a HCAL cluster is associated to a track, independently from ECAL; the category where both a ECAL and a HCAL clusters are associated to a track.

The k_{MC} is parametrized using multivariate regression techniques implemented in the TMVA package [122], using the variables listed above as inputs, by exploiting simulated events of b, c, light quarks and gluons jets at a center of mass energy of 7 TeV and 8 TeV. This correction factor is extracted in bins of the relevant variables (bins of width 1 GeV are used for reconstructed p_{T}, bins of width 0.1 are used for the reconstructed pseudorapidity, and three bins in the number of primary vertices are used, corresponding to one, two, and three and more primary vertices).

Figure 3.9 shows the mean k_{MC} in different intervals of η, ϕ, cpf and nPVs.

The correction factor increases at low and high η. In these regions the neutral content of jets is underestimated, as this region is close to the edge of the detector geometric acceptance, where the calorimeter efficiency to detect neutral particles is reduced. The correction also changes in the rest of the detector, as the calorimeter cell size varies across the detector.

Uncertainties on the jet energy correction

Different sources of uncertainty on the jet energy scale were considered [121]:

- the statistical uncertainty on the evaluation of the jet energy scale, typically less than 1%;
- the main systematic uncertainty is associated with the multivariate method used to find the correction factor. It was determined by effectively finding the correction factor using an alternative approach. A Gaussian was fitted to the peak of the $(p_{\mathrm{T}}^{\mathrm{reco}} - p_{\mathrm{T}}^{MC})/p_{\mathrm{T}}^{MC}$ distribution in bins of the jet p_{T} and the number

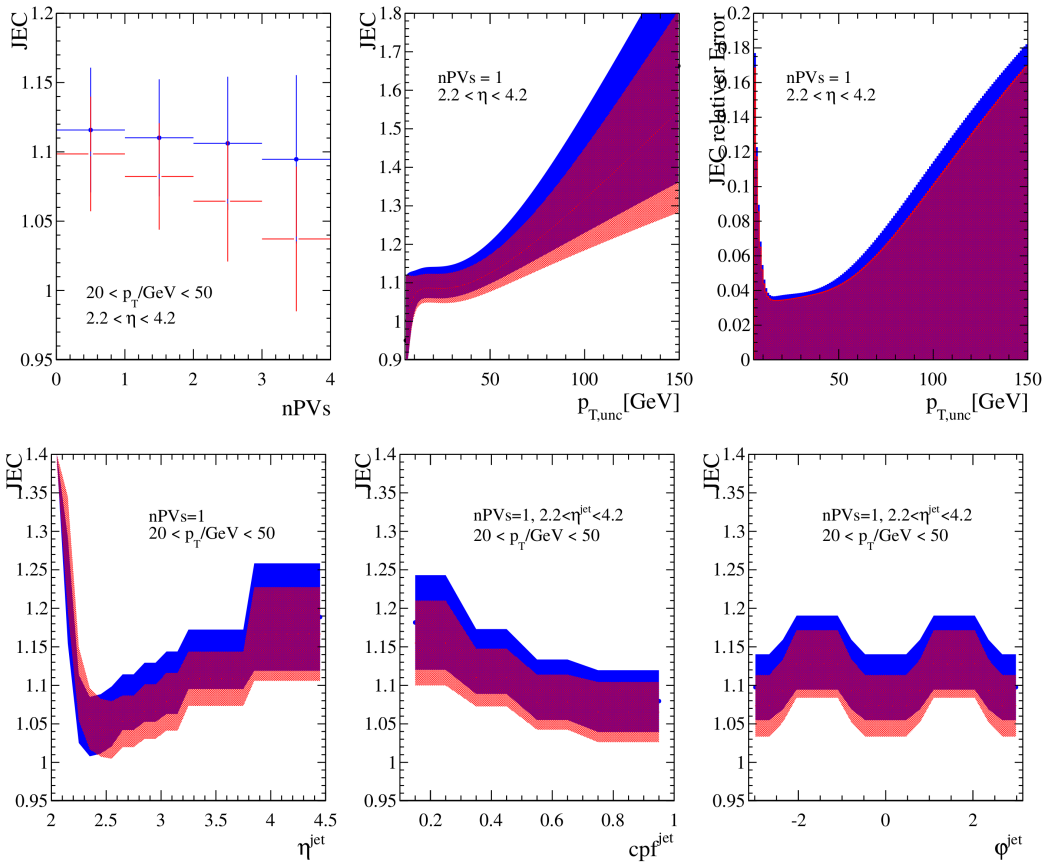


FIGURE 3.9: k_{MC} in different intervals of η, ϕ, cpf and nPV for $R = 0.5$ (blue) and $R = 0.7$(red).

of primary vertices. The difference between the mean of this Gaussian and zero sets the uncertainty on the energy scale from the method. This uncertainty tends to be largest at low and high p_{T};

- in order to determine the statistical uncertainty on the evaluation of the jet energy scale the correction was evaluated on different sub-samples. The mean correction in these sub-samples sets the overall correction, whilst the uncertainty on the mean defines the statistical uncertainty on the correction. This uncertainty is typically less than 1%

- The systematic uncertainty associated with the flavour content of the jet is evaluated by determining the correction factor by selecting exclusively bottom, charm or light jets. The difference in response is typically about 2%, and is reasonably constant.

The total contribution of these different sources is typically $2 - 3\%$.

What is still missing is contribution to the scale uncertainty which arises from differences between simulation and data, the same that has been evaluated for CDF in Section 4.3.2. A mis-modeling of the detector response to jets in simulation leads to a wrong energy correction applied to reconstructed jets in the LHCb data. This has been studied at LHCb, factorizing the uncertainties coming from different sources[123]:

- energy scale uncertainty associated with charged particle reconstruction, which is evaluated with muons from the J/ψ at low momentum and from the Z

at high momentum, by determining how much the measured track momenta must be scaled to recover the known particle mass;

- energy scale uncertainty associated with ghost tracks (tracks not associated to real particles), which is calculated by changing the requirement on the ghost probability;
- energy scale uncertainty associated with neutral particle reconstruction, by studying the discrepancy between data and simulation in the charged and neutral particle fraction variables.

Typical uncertainties are in the range 2-4% for jets in most regions of the LHCb phase space. These values are in agreement with the jet energy scale uncertainty evaluated in-situ using the recoil jet in Z+jet data events.

The measurement of the $Z \rightarrow b\bar{b}$ cross section at LHCb [87], so it has been at CDF, led to the determination of the residual jet energy scale between data and MC for high p_{T} b-jets with an uncertainty of the 2%.

3.3.2 Identification of the jets flavour

The tagging algorithms to identify jets coming from b and c quarks implemented in the LHCb software are here described.

The SV tagger

The flavour jet tagging algorithm implemented at LHCb is called SV tagger [124], and as suggested by the name, it searched for secondary vertex inside the jet. The peculiarity of this algorithm is that, by exploiting the great detector vertexing performance, it can efficiently select and distinguish c-jets from b-jets.

The tracks used as inputs to the SV tagger algorithm are required to have $p_{\mathrm{T}} > 0.5$ GeV/c and $\chi^2_{IP} > 16$, which is defined as the variation of the χ^2 obtained from the PV fit when the track is removed from the fit result. This last requirement is rarely satisfied by tracks reconstructed from particles originating directly from the PV. Hadronic particle identification is not used and, instead, all particles are assigned the pion mass. In contrast to many other jet-tagging algorithms, tracks are not required to have $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.5$ between the track momentum and jet axis, since for low p_{T} jets tracks outside of the jet cone help to discriminate between c and b-jets

Selected tracks are used to build all possible 2-body SVs in the 3-dimensional space. Fits are performed to determine the SV position. The two tracks associated to a SV are combined to form a particle which flight direction is defined as the vector that points the SV from the PV. Its four-momentum is defined as the sum of tracks four-momenta. The 2-body particles have to fulfill the following requirements:

- the distance of closest approach (DOCA) between the two tracks must be less than 0.2 mm;
- the χ^2 associated to the SV fit must be less then 10;
- the invariant mass must be greater than 400 MeV/c^2 and less then the B^0 mass (5279.4 MeV/c^2).

The lower mass requirement is used to remove strange-hadrons decays.

At this stage tracks are allowed to belong to multiple SVs. Next, all two-track SVs with $\Delta R < 0.5$ with respect to the jet axis, where flight direction is taken as the PV to SV vector, are collected as candidates for a so-called linking procedure. This procedure involves merging SVs that share tracks until none of the remaining SVs with $\Delta R < 0.5$ share tracks. The resulting n-body particles are called tag$_{SV}$.

To further remove the light jet contamination, selection criteria are applied to tag$_{SV}$:

- $p_T > 2$ GeV$/c$;
- z-position $< 200\ mm$;
- flight distance$/p < 1.5\ mm/(\mathrm{GeV}/c)$;
- the flight distance χ^2, defined as the χ^2 obtained from the PV fit if the tag$_{SV}$ tracks are added to the fit result, must be above 5σ;
- if tag$_{SV}$ is formed by only two tracks and a mass compatible with the K_S it is rejected;
- the tag$_{SV}$ must have at most one track with $\Delta R > 0.5$ from the jet axis;
- an important quantity for discriminating between hadron flavours is the so-called corrected mass defined as

$$M_{cor} = \sqrt{M^2 + p^2 \sin^2\theta} + p \sin\theta, \tag{3.14}$$

  where M and p are respectively the invariant mass and the momentum of the tag$_{SV}$, θ is the angle between the tag$_{SV}$ momentum and flight direction. M_{cor} must be greater than 600 MeV$/c^2$ to remove any remaining kaon or hyperon. This quantity is reliable even in case where not all final state particles are reconstructed, in fact $p \sin\theta$ is the transverse component of the missing momentum with respect to the particle flight direction.

A few percent of jets contain multiple SVs that pass all requirements; in such cases the SV with the highest p_{T} is chosen.

To further remove light jet contamination and to distinguish b-jets from c-jets a Multivariate Algorithms (MVA) is used. Two boosted decision trees (BDTs) [122] are employed: one for the heavy/light jets separation (BDT$_{bc|udsg}$) and the other for the b/c jets separation (BDT$_{b|c}$). Simulated samples of heavy quark/light quark jets and b/c jets generated with PYTHIA 8 are used as signal/background samples for the BDTs training. The observables in input to the BDTs are those related to tag_{SV} that provide the highest discrimination power between the different flavours, they are:

- the tag$_{SV}$ mass M;
- the tag$_{SV}$ corrected mass M_{cor};
- the transverse flight distance of the 2-body particle closest to the PV within those that form tag$_{SV}$;
- the fraction of the jet p_T carried by tag$_{SV}$, $p_T(SV)/p_T(jet)$;
- the number of tracks that form tag$_{SV}$;
- the number of tracks that form tag$_{SV}$ with $\Delta R < 0.5$ from the jet axis;

- the total charge of tracks in tag$_{SV}$;
- the tag$_{SV}$ flight distance χ^2;
- the sum of χ^2_{IP} for all the tracks in tag$_{SV}$.

The two-dimensional BDT outputs distribution are shown in figure 3.10 for the b, c and light quarks MC samples generated with PYTHIA 8 for 8 TeV collisions. Cuts on these two observables can be applied to enrich the jets samples of a specific hadron flavour.

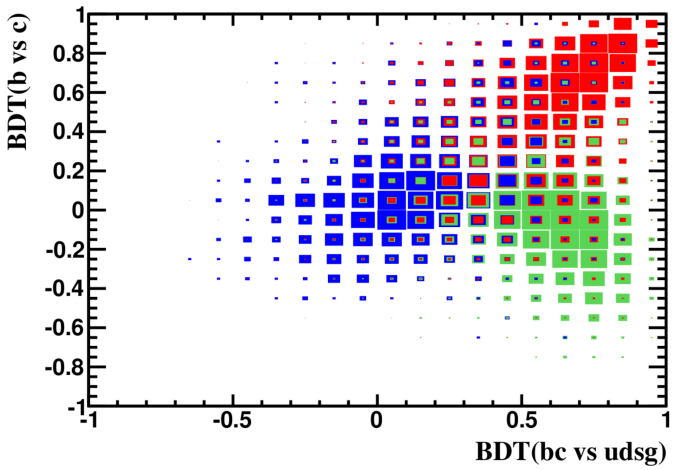


FIGURE 3.10: Two-dimensional BDT$_{bc|udsg}$ and BDT$_{b|c}$ distribution for b, c and light jet, obtained from simulation. In blue the light jet, in green the c jet and in red the b jet distribution.

SV tagger performance

The SV-tagger performance [124] has been first evaluated on simulation, then efficiencies for b- and c-jets have been measured in data and compared with expectations from simulation. For this study b, c and light partons samples generated with PYTHIA 8 for pp collisions at 8 TeV have been used. The tagging efficiencies and the misidentification probability have been determined as a function of the reconstructed jet p_{T} and η. The following conclusions are obtained:

- for jets with p_T greater than 20 GeV the mean SV-tagger efficiency, without further cut on the BDT$_{bc|udsg}$ and BDT$_{b|c}$ distributions, on b-jets is of about 60% and of about 25% for c-jets;
- for jets with p_T less than 20 GeV the b(c) tagging efficiencies are significantly lower;
- for $2.2 < \eta < 4.2$ the tagging efficiencies are almost constant;
- the light parton misidentification probability is of less than 0.1% for low p_{T} jets and increases to about 1% at 100 GeV/c.

Figure 3.11 shows the (b, c)-jet efficiencies versus the mistag probability of light-parton jets obtained by increasing the BDT($bc|udsg$) cut.

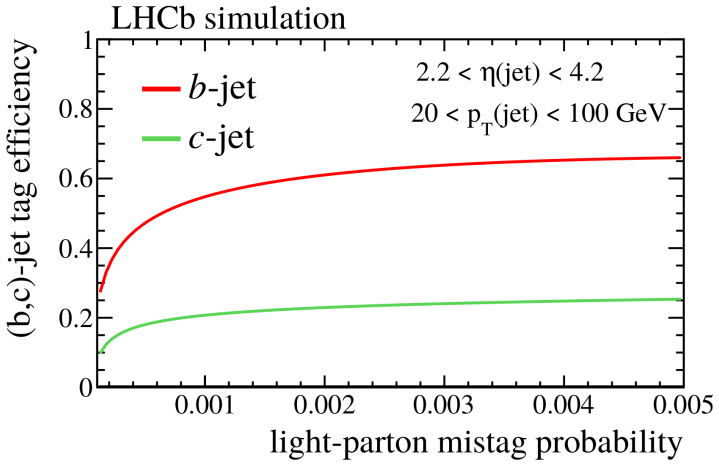


FIGURE 3.11: Efficiencies for SV-tagging a (b, c)-jet versus mistag probability for a light-parton jet from simulation.

To evaluate the efficiency on data, as already seen in Section 3.2.2, both the number of tagged (b, c) jets and the total number of (b, c) jets must be determined. These values have been evaluated by using combined fits to different data samples, enriched in b, c or light jets, including also the systematic uncertainties. The different data samples that have been used are:

- b+jet: an enriched sample of b-jets, collected with the topological trigger in pp collisions at a centre-of-mass-energy of 8 TeV;
- d+jet: an enriched sample of b- and c-jets, due to the $b \to c$ transition in the B hadrons decays, collected with the charm trigger in pp collisions at a centre-of-mass energy of 8 TeV;
- μ+jet: a sample where a high p_T displaced muon is selected to enrich the content of b and c-jets, collected with the high p_T muon trigger in pp collisions at a centre-of-mass-energy of 8 TeV;

The number of tagged b- and c-jets has been found by selecting SV-tagged jets and using 2-dimensional fits to the $\text{BDT}_{b|c}$ and $\text{BDT}_{bc|udsg}$ variables. The $\text{BDT}_{b|c}$ and $\text{BDT}_{bc|udsg}$ distribution are obtained using MC events.

The total number of b-, c-jets is determined by fitting the χ^2_{IP} distribution of the highest-p_{T} track in the jet ($\chi^2_{IP(max-p_{\text{T}})}$). For light -parton jets the highest-p_{T} track will mostly originate from the PV, while for (b,c) jets the highest-p_{T} track will often originate from the decay of the B/D hadron. The templates of $\chi^2_{IP(max-p_{\text{T}})}$ distributions for b, c and light jets are built using MC events. The $b(c)$ tagging efficiency is calculated as

$$\epsilon_{b(c)} = \frac{N^{tag}_{b(c)}}{N^{tot}_{b(c)}}. \tag{3.15}$$

Systematic uncertainties are included in the fits and they are mainly associated to the BDT templates, to the mismodeling $\chi^2_{IP(max-p_{\text{T}})}$ and to the gluon splitting, that can create $b\bar{b}$ and $c\bar{c}$ couples with high probability to be tagged. The measured efficiencies and the data/MC scale factors for the SV tagger in different p_{T} intervals are reported in table 3.4 and in Figure 3.12. The uncertainties bars reported take into account of statistical and systematic uncertainties.

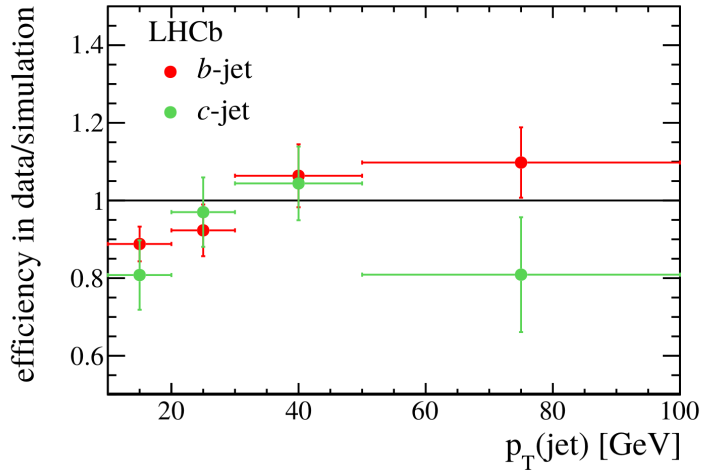


FIGURE 3.12: Ratio between the SV tagging efficiencies measured in data and those obtained from simulation, for b and c jets.

TABLE 3.4: SV tagging algorithm efficiencies measured on data.

		ϵ(data)/ϵ(simulation)			ϵ(data) (%)	
jet p_{T} (GeV)	jet η	b jets	c jets	(b,c) jets	b jets	c jets
10–20	2.2–4.2	0.89 ± 0.04	0.81 ± 0.09	0.91 ± 0.04	38 ± 2	14 ± 1
20–30	2.2–4.2	0.92 ± 0.07	0.97 ± 0.09	0.97 ± 0.04	61 ± 3	23 ± 1
30–50	2.2–4.2	1.06 ± 0.08	1.04 ± 0.09	0.97 ± 0.04	65 ± 3	25 ± 1
50–100	2.2–4.2	1.10 ± 0.09	0.81 ± 0.15	1.05 ± 0.06	70 ± 4	28 ± 4
20–100	2–2.2	1.00 ± 0.07	1.12 ± 0.10	1.05 ± 0.03	56 ± 2	20 ± 1

3.3.3 Jets in Run II

A trigger for jets

During the Run I data taking the selection at the on-line level of events with jets was based on the so called topological triggers, see Section 2.2.3, which are designed to inclusively select events with b hadrons. These algorithms allow to collect samples with a good purity of b-jets, but with a not so good efficiency and with the drawback of strongly suppress the events with c-jets.

For the Run II data taking a new strategy for the on-line selection of events with jets have been implemented. The reconstruction of the jets, always based on the FastJet [120] interface with the anti-k_T [61] algorithm and with $R = 0.5$, is now possible at the HLT2 level. A re-optimized and more flexible version of particle flow has been implemented [119] in the HLT2 software. In view of a sensitivity study on the $H \to b\bar{b}$ process, the most interesting line is the so called "Hlt2JetsDiJetSVSVLine".

As suggested by the its name, the line selects events with two jets both tagged with a secondary vertex; in this way the retention rate is kept enough low not to prescale the line. First, the secondary vertices are searched by reconstructing any two-body displaced vertex using as input the tracks that fired the HLT1 trigger lines which look for high impact parameter tracks. The formed 2-body particle four-momentum is defined as the sum of tracks four-momenta, assuming the K mass. The 2-body particle position is obtained with a vertex fit to the tracks, while the

flight direction is defined as the vector that points its position from the PV. The additional requirements are:

- input tracks $p_{\mathrm{T}} > 500$ MeV/c
- input tracks $P_{ghost} < 0.2$
- minimum distance of tracks χ^2 IP > 16
- χ^2 of the SV < 10
- the SV is 5σ away from the PV

The jets, with a $p_{\mathrm{T}} > 17$ GeV/c, are then reconstructed. If the event has at least two jets with both of them with a secondary vertex the event is selected by the trigger line.

Run II jets validation

As explained above, a new configuration for the reconstruction of jets in the HLT has been introduced in Run II. This mainly differs from the off-line Run I jet configuration for a looser selection of the input tracks and calorimetric clusters in the particle flow algorithm. This leads to a faster algorithm, but with slightly worse performance on the resulting jet resolution. For this reason, we keep using the Run I configuration for the off-line jet reconstruction for the sensitivity study on the $H \to b\bar{b}$ process. Figure 3.13 shows the comparison between the resolution of jets reconstructed using Run I and HLT (performed off-line) configurations, evaluated using Run II data condition Monte Carlo events and defined as the standard deviation of $\sigma = \frac{p_{\mathrm{T}}(jet_{\mathrm{reco}}) - p_{\mathrm{T}}(jet_{\mathrm{MC}})}{p_{\mathrm{T}}(jet_{\mathrm{MC}})}$ as a function of $p_{\mathrm{T}}(jet_{\mathrm{MC}})$. In Figure 3.13, the blue (TURBO) and yellow (PSI) points refer to other two configuration where the neutral energy recovery is not applied at all.

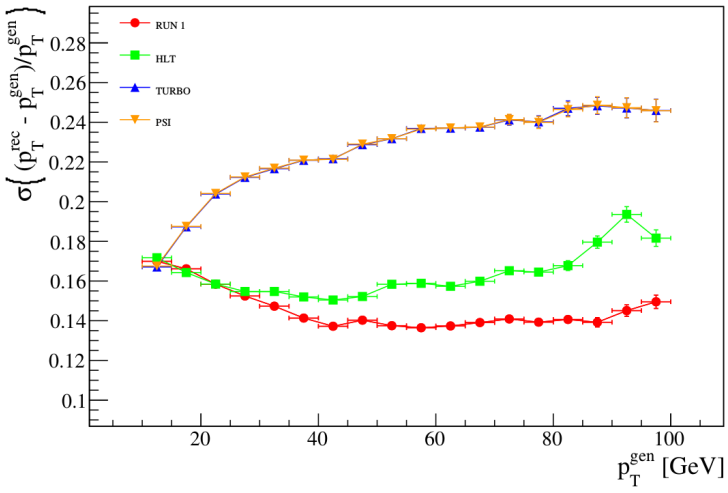


FIGURE 3.13: Resolution of different jet reconstruction configurations. PSI configuration can be considered equal to the TURBO configuration.

Some tests have been performed to validate the Run I jet reconstruction applied to Run II data. First, in order to check if any degradation of the calorimeter system

has occurred, the jet asymmetry variable $\mathcal{A}$ has been check as a function of the data taking time. In events with at least two jets, the jet asymmetry is defined as:

$$\mathcal{A} = \frac{p_{\mathrm{T}}(jet1) - p_{\mathrm{T}}(jet2)}{p_{\mathrm{T}}(jet1) + p_{\mathrm{T}}(jet2)} \tag{3.16}$$

where $p_{\mathrm{T}}(jet1)$ and $p_{\mathrm{T}}(jet2)$ refer to the randomly ordered transverse momenta of the two leading jets. This variable is directly proportional to the jet resolution, in fact:

$$\sigma(\mathcal{A})^2 = |\frac{\partial \mathcal{A}}{\partial p_{\mathrm{T}}(jet1)}|^2 \sigma(p_{\mathrm{T}}(jet1))^2 + |\frac{\partial \mathcal{A}}{\partial p_{\mathrm{T}}(jet2)}|^2 \sigma(p_{\mathrm{T}}(jet2))^2. \tag{3.17}$$

Assuming that $p_{\mathrm{T}}(jet1) = p_{\mathrm{T}}(jet2) = p_{\mathrm{T}}$ and $\sigma p_{\mathrm{T}}(jet1) = \sigma p_{\mathrm{T}}(jet2) = \sigma p_{\mathrm{T}}$, and that both leading jets are in the same η region, the fractional p_{T} resolution can be expressed as a function of the variance $\sigma(\mathcal{A})$ in the following way:

$$\frac{\sigma p_{\mathrm{T}}}{p_{\mathrm{T}}} = \sqrt{2}\sigma(\mathcal{A}) \tag{3.18}$$

The resolution derived corresponds to the true jet p_{T} resolution only in the limit of exactly two jets in the event without additional radiated energy. For this preliminary study, this effect is taken into account only putting a veto on events with a third jet with $p_{\mathrm{T}} > 10$ GeV/c.

The dataset used for this study corresponds to 100 pb^{-1} of data collected in 2016 collected with the Hlt2JetsDiJetSVSVLine trigger. Both jets are required to have $p_{\mathrm{T}} > 17$ GeV/c, $2.2 < \eta < 4.2$, to be SV tagged and back-to-back in azimuth with $\Delta\phi > 2.9$. Figure 3.14 shows the jet resolution so calculated as a function of the data taking month. It can be seen that the resolution has a dependency on the data taking period, indication of a degradation of the calorimeter system, but still negligible.

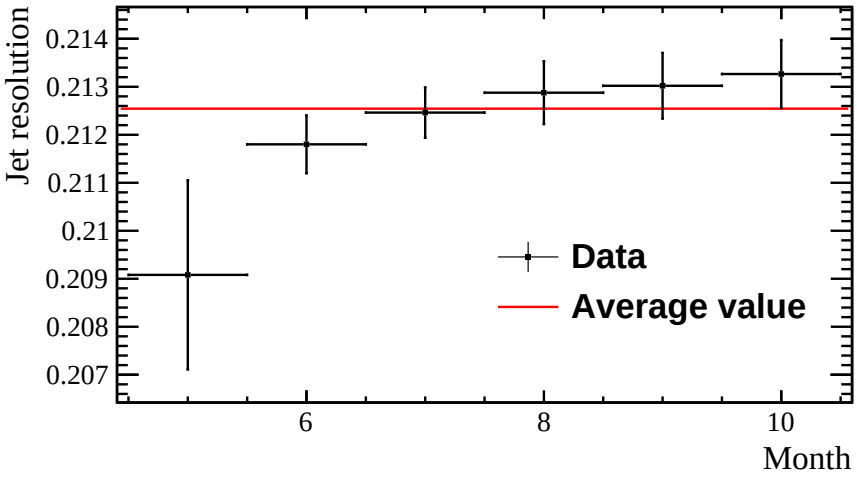


FIGURE 3.14: Jet resolution as a function of the data taking month.

By exploiting the study the p_{T} balance of Z+jet events instead, it is possible to test the jet energy scale response in Run II data. In events where the Z boson and jet are produced back to back in ϕ, the Z boson and the jet should balance in p_{T}. In these events the distribution of the ratio $p_{\mathrm{T}}(jet)/p_{\mathrm{T}}(Z)$ is expected to be well modeled at generator level by simulation, so that any discrepancy between data and simulation can be assigned to the detector response. The dataset has been collected in 2015, at a

center of mass energy of 13 TeV. The Z boson is required to decay into two muons with a $p_{\mathrm{T}} > 20$ GeV/c and $IP < 0.04$ mm. Only jets with a p_{T} greater than 20 GeV/c are considered and the ΔR between the jet and the muon has to be > 0.5. Finally, only events where the jet and the muon are back to back in azimuth, $\Delta\phi > 2.7$ are selected. The $p_{\mathrm{T}}(jet)/p_{\mathrm{T}}(Z)$ distributions in data and simulation are overlaid and plotted in Figure 3.15. Good agreement is seen between data and simulation. This agreement provides confidence that the jet reconstruction is well modeled in simulation also in Run II.

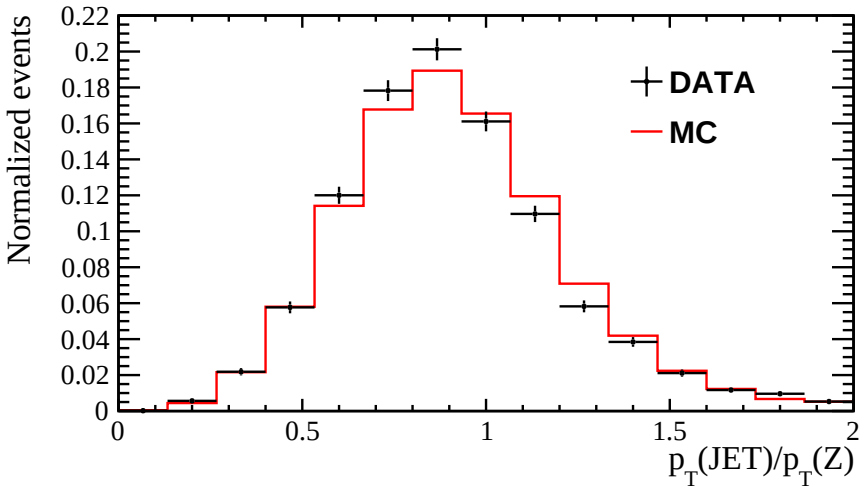


FIGURE 3.15: The p_{T} balance of selected Z+jet events for $p_{\mathrm{T}}(jet) > 20$ GeV/c. The data is shown as the black, with simulation in red.

Chapter 4

Search for the Z resonance into a pair of b-jets

In this chapter it is presented the analysis of the dataset collected at CDF II looking for the Z resonance decaying into a pair of b-jets. As a comparison, at the end of the chapter, the analogous measurement performed at LHCb [87], recently published, is briefly reported (this last analysis is not a directly result of my work).

The Z boson has been studied at hadron colliders predominantly using the leptonic decays into muons or electrons. In fact, the hadronic decays are much more difficult to separate from the overwhelming background arising from generic jet pairs produced by the multijet QCD production. Only the decay to b-quark pairs is observable because the $gg \to gg$ background is significantly reduced. Thus, the identification and the measurement of the inclusive $Z \to b\bar{b}$ process represents a real challenge, but on the other hand it gives the unique opportunity to have a standard candle to validate the analysis procedure, the background modeling and to measure the b-jet energy scale (JES), a factor which evaluates the discrepancy between the effect of detector response and energy corrections in real and simulated hadronic jets.

The measurement of the $Z \to b\bar{b}$ process has been reported before by CDF itself [66], using 584 pb^{-1} of data and by ATLAS [67]. These measurements also require a third recoil jet to define discriminating variables to separate the signal from the QCD background, as well as it has been done for the LHCb measurement. Also CMS released a measurement of the $Z \to b\bar{b}$ channel [4], but in the boosted topology with the two jets merged in a single fat jet. The selection described in this thesis is inclusive in the sense that it does not make any request on the event topology.

Section 4.1 describes the characterization of the samples used in the analysis, so the real data, the event selection and the the Monte Carlo simulated samples. Section4.2 reports the analysis key point: the background modeling; templates as a function of the invariant mass distribution of the dijets are built for different jet flavour compositions. The signal extraction with the measurement of the cross section and of the jet energy scale is described in Section 4.3 while in Section 4.4 the systematic uncertainties are studied. The $Z \to b\bar{b}$ analysis performed at LHCb is then summarized in Section 4.5.

4.1 Dataset and Monte Carlo simulations

4.1.1 Data sample and event selection

Figure 4.1 shows the production cross section of different processes both at the Tevatron and at the LHC as a function of the collision energy at the center of mass frame. This figure helps to give a general picture of the main processes that are involved in a

hard proton-antiproton or proton-proton collision. The physics that we are studying happens with a cross section about 8 order of magnitude smaller than the total cross section and about 4 order of magnitude smaller than the main background for these searches, the QCD production of b-jets pairs. Thus, as a first stage, it is essential to reject as much as possible the contamination from jets originating from light quarks and gluons ($usdg$) and charm quarks (c). Then, we have to keep in mind there are no real handles to discriminate between the b-dijets originating from the signal and the ones originating from QCD processes, except for the invariant mass of the dijets. The event selection is based on these two concepts.

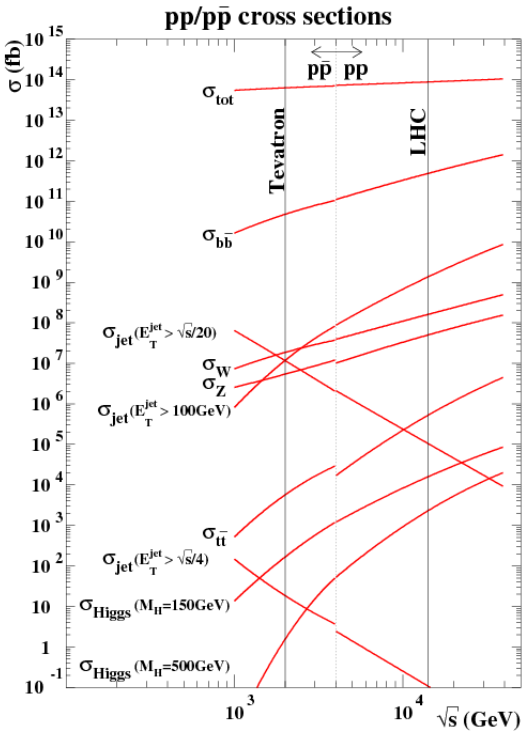


FIGURE 4.1: The production cross section of different processes at the Tevatron and at the LHC as a function of the collision energy at the center of mass frame.

The dataset corresponds to 5.4 fb^{-1} of data collected at CDF II, from spring 2008 up to the end of the Tevatron operation in September 2011, using the DIJET_BTAG trigger path already described in Section 2.1.4.

At the off-line level, the hadronic jets are reconstructed by using the standard CDF II prescription described in Section 3.2.1. The iterative jet cone clustering algorithm JetClu [62], with the cone radius $R = 0.7$ is used. The standard CDF jet energy correction package up to the L5 is then applied to the raw jet energy. The signal is searched in events that are required to have:

- at least two central jets, $|\eta| < 1$, with $E_{\mathrm{T}} > 22$ GeV;
- at least one jet that fired the on-line b-tagging;
- at least two jets with a secondary vertex identified by the tight SecVxt b-tagging algorithm.

This sample is referred through the thesis as the double tag sample. The cut on the E_{T} permits to keep the peak in the invariant mass of the QCD dijets well below the

Z peak but at the same time to avoid the steep rise of the turn-on curve, difficult to model, of the calorimetric response at the trigger level (see Figure 3.5). The request of two tight SecVtx tagged jets allow to remove most of the remaining non-b contamination, especially from the jets originating from light quarks and gluons.

Another selection is applied to the dataset to identify a sample which is used for the modeling of the multijet QCD background:

- at least two central jets, $|\eta| < 1$, with $E_{\mathrm{T}} > 22$ GeV;
- at least one jet that fired the on-line and the tight SecVxt b-tagging algorithm.

This sample is referred as single tag sample, while the jet that fires the DIJET_BTAG trigger and has a tight SecVtx tag is called b-tag trigger jet.

4.1.2 Monte Carlo simulation

The analysis uses Monte Carlo simulated events, generated following the prescription described in Section 2.1.4, to evaluate the efficiency and the acceptance of the $Z \to b\bar{b}$ and $H \to b\bar{b}$ signals, to determine the dijet invariant mass for the signals and to extract the tagging algorithm response to the different jet flavor.

To study the response of the off-line SecVxt tagging algorithm for different jet flavour, simulated QCD events have been produced. The QCD $b\bar{b}$ sample is composed of 14 millions of generated events produced with the configuration:

- MSEL set 1, this produces QCD dijets using Flavor Creation, Flavor Excitation and Gluon Splitting production mechanisms;
- two b-quarks filtered at generation level;
- underlying event modeled using the Rick Field tune A [70];
- $\hat{p}_{\mathrm{T}}$ (the transverse momentum in rest frame of the hard interaction) minimum 18 GeV/c;

With the same procedure, but only with the Flavor Creation production mechanism active, a sample of 10 millions of $c\bar{c}$ has been produced. A $q\bar{q}$ generic dijets sample has been generated with different minimum $\hat{p}_t$ and then weighted in order to obtain enough events after the trigger. The minimum $\hat{p}_t$ are: 18, 40, 60, 90, 120, 150 and 200 GeV/c.

For what concern the signal samples, about 5.9 million of $Z \to b\bar{b}$ signal events are generated with the same version of Phytia; the Z is then forced to decay into a $b\bar{b}$ couple and the two b-jets invariant mass is required to be greater than 30 GeV/c^2. About 1 million $H \to b\bar{b}$ events have been generated, with the Standard Model Higgs generated with a mass of 125 GeV/c^2.

4.1.3 Monte Carlo validation

In this analysis the background estimation is performed using a data driven procedure, since MC simulation is not completely reliable in describing QCD processes. Nevertheless simulated events are necessary to determine efficiencies as a function of kinematical variables, therefore it is necessary to demonstrate that the Monte Carlo is able to reproduce data at least for those variables.

The first validation is performed by using an independent sample, the single muon dataset, which has already been used to determine the b-tagging efficiencies

and the data/MC scale factors, as described in Section 3.2.2. The b purity of this sample is greater than 95%. In the MC sample the *muon jet* is matched to b-quark at the generator level. Figure 4.2 demonstrates a good agreement between data and MC for the jet transverse energy, the jet η, the muon transverse momentum and number of good displaced tracks selected by SecVtx.

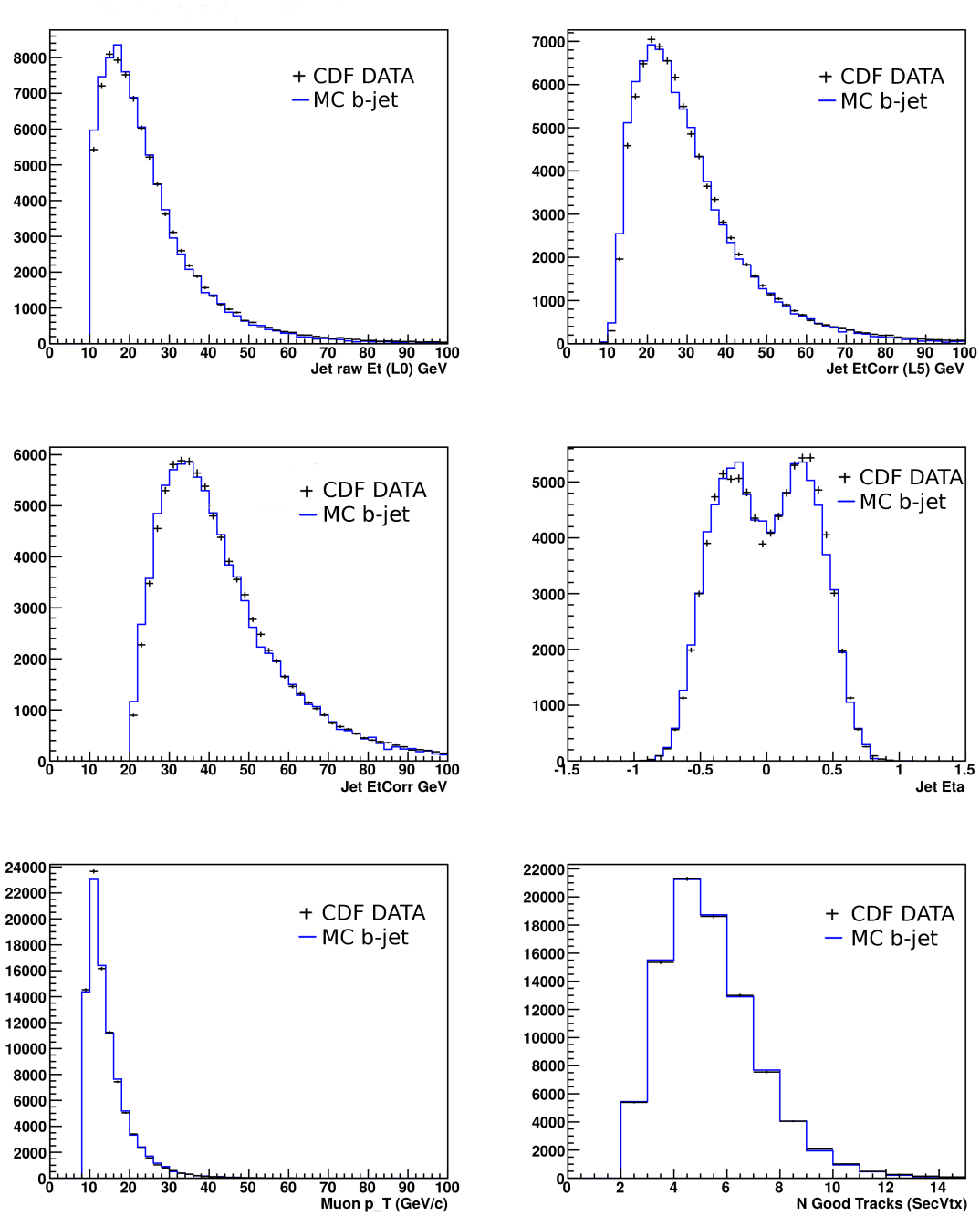


FIGURE 4.2: Variables distribution of data (dots) and MC (blue solid line) for *muon jets* with a tight SecVtx. From top left, clockwise, the jet E_{T} before any correction, the jet E_{T} before off-line selection, the jet η, the number of displaced tracks selected by the SecVtx, the muon p_{T} and the jet E_{T} after off-line selection. The histograms are normalized to the same number of events.

A second validation is performed by using the data and the QCD $b\bar{b}$ Monte Carlo simulation exploited for the analysis: events that survived the trigger and the off-line selection are compared. Figure 4.3 shows the transverse energy and the η distributions of the leading and the sub-leading E_{T} SecVtx tagged jets, data and Monte Carlo superimposed. The agreement is good, meaning that the kinematical properties of the jets are well reproduced, as expected.

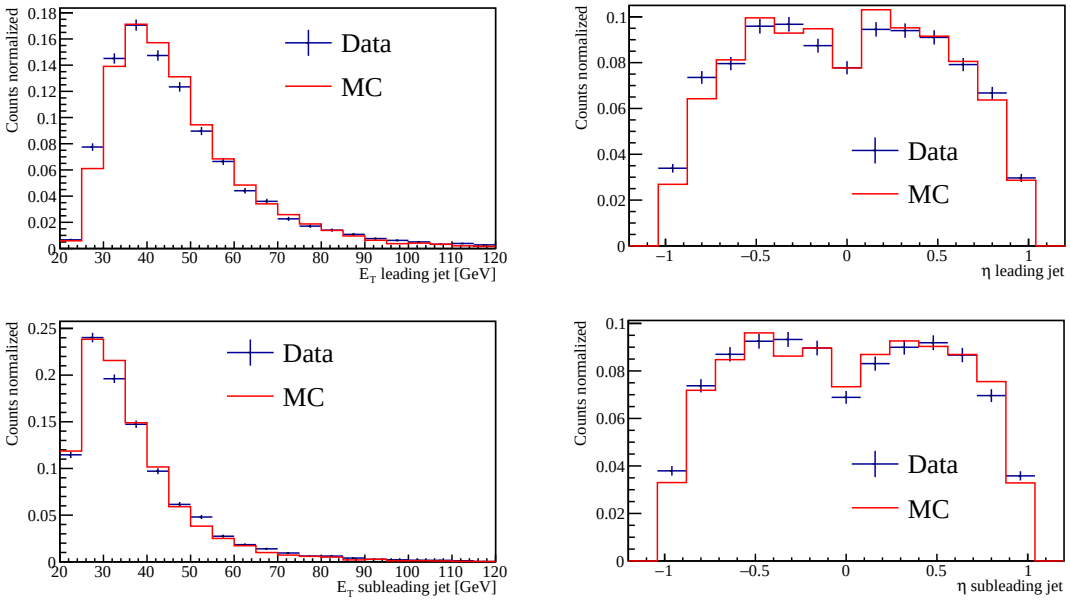


FIGURE 4.3: The E_{T} and the η distribution for data in black dots and for QCD $b\bar{b}$ Monte Carlo simulation in red solid line. Top plots are for the leading E_{T} jet and the bottom one for the sub-leading.

4.2 Background modeling

The sample exploited to measure the $Z \to b\bar{b}$ cross section and to set an upper limit on the $H \to b\bar{b}$, the double tag sample, has at least two tight SecVtx tags. The background is predominantly constituted by QCD heavy quarks jets. These events arise from QCD mechanisms for which a precise prediction of the production rates is difficult to obtain. It is not possible then to simply use the Monte Carlo simulation, scaled to the Standard Model predictions, to determine the background composition of the double tag sample.

Moreover, the expected signal over background ratio (S/B) is very low, about 2%, and very little can be done after selecting a clean dijet system in which both jets are from b-quark. The modeling of the background thus, to be as reliable as possible, has also to include all the biases introduced by the trigger and the off-line selection.

For all these reasons the only way to model and evaluate the background contribution is to use a data-driven method. This procedure has already been validated and applied at CDF to search for beyond Standard Model resonances in a sample of three b-tagged jets [71]. The method proceeds as follow:

- the probabilities, as function of jet E_{T} and η, to tag a $b-$, $c-$, light-quark initiated jet as a b jet are determined by using the Monte Carlo simulated data. These per jet probabilities represent the efficiency to tag (b, c or light quark initiated jet) as a b jet and are referred as tagging matrices;

- starting from the single b-tag jet data sample, the flavour-dependent bias introduced by the SecVtx tagging is simulated on the non-tagged jet by weighting it with the tagging matrices for b, c, light quark jets;
- the invariant mass of the dijet formed by the b-tag trigger and the flavour tagged-simulated jets, is calculated under the different jet flavour hypotheses.

With this method only the shapes of the background are determined, normalizations are found in the final fit and are one of the results of the analysis.

This procedure works under some assumptions. First, when building the background templates, the b-tag trigger jet has to be initiated by a real b-quark. The sample composition of the single tag sample is studied in Section 4.2.2 and an ad hoc cut on the secondary vertex mass will be required to increase the sample purity. Second, in the single tag sample, used to model the background, the contamination due to the $Z \to b\bar{b}$ signal has to be negligible with respect to the QCD multijet production. The signal over background ratio in this sample is only 0.1%, thus it can be neglected.

4.2.1 The tagging matrices

The first step to of the background modeling is the building of the tagging matrices, a 2-dimensional parametrization of the response of the SecVtx b-tagging algorithm jet-flavour dependent.

The probability to tag as b-jet a jet initiated by a b-, c-, light-quark is defined as

$$\frac{N(tag)}{N(jets)} \tag{4.1}$$

where $N(jets)$ is the number of jets originating by a b-(c- and light-) quark and $N(tag)$ is the number of SecVtx tagged jets. These numbers are determined using the QCD $b\bar{b}$, $c\bar{c}$ and light quarks Monte Carlo simulated samples described in Section 4.1.2, after having associated the jet to the quark flavor by using the generator level information.

Events have to pass the trigger selection and at least one jet has to have one SecVtx tight tag in order to reproduce the bias introduced by the various requirements. To keep a high statistics in the $b\bar{b}$ Monte Carlo sample the trigger selection is reduced: no cut on signed impact parameter d_0 is applied.

In figure 4.4 the dependence of the tagging probability for b-jets as function of jet E_{T}, η, number of tracks selected by SecVtx and jets multiplicity in the event is shown for Monte Carlo b-matched jets. A high number of variables used to parametrize the tag probability should help to better describe the tagging response, but with the limited statistics it is necessary to limit to two quantities in order to be not dominated by the statistical fluctuations of each bin. Jet E_{T} and η are chosen because they are uncorrelated and are well modeled by the Monte Carlo simulation.

Summarizing, the tagging matrices are determined as follow from each $b\bar{b}$, $c\bar{c}$ and light quark Monte Carlo samples:

- for each event a b-tag trigger jet is identified in each Monte Carlo sample;
- the distribution of the *other jets* in the event, excluding the b-tag trigger jet, as a function of the jet E_{T} and η is found, these jets are matched to the generator level quark flavour;

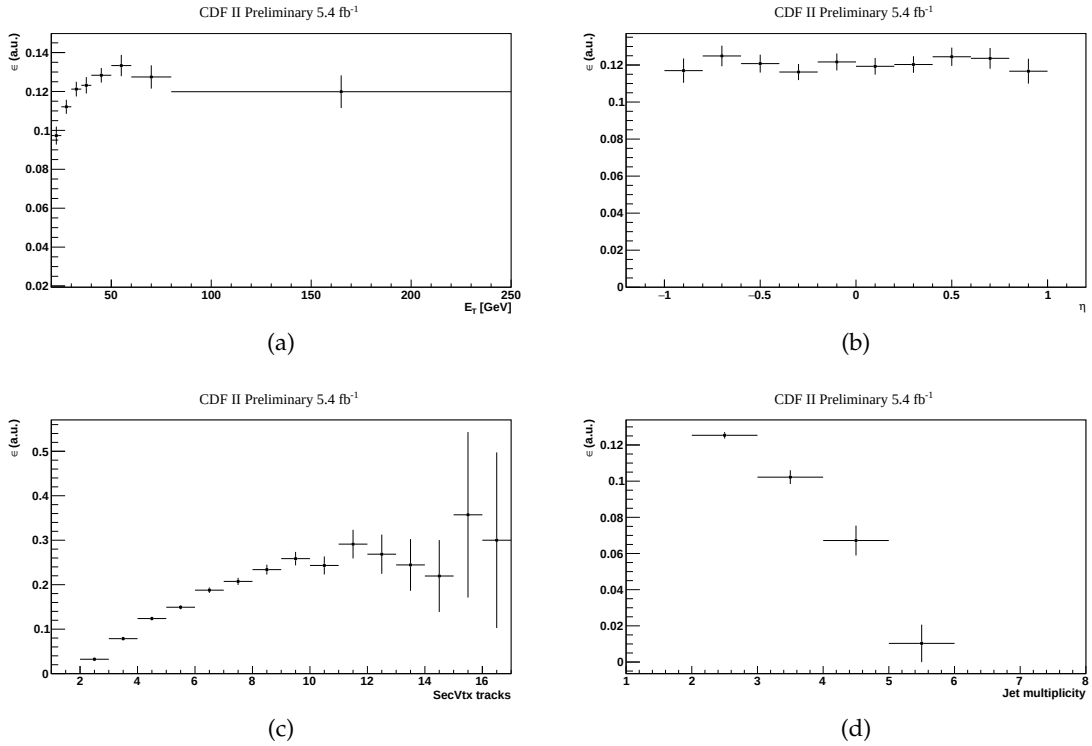


FIGURE 4.4: Jet SecVtx tagging probabilities as function of the jet E_{T} (a), η (b), number of SecVtx tracks (c) and jet multiplicity (d) for Monte Carlo b-matched jets.

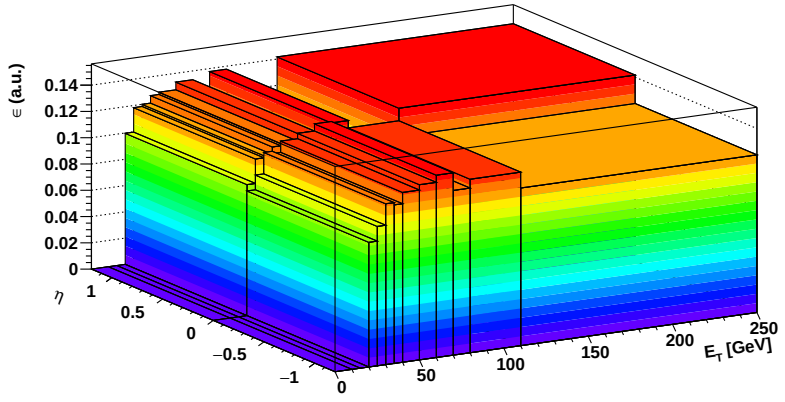
- the ratio between the distribution of the SecVtx tagged *other jets* and the distribution of all the *other jets* constitutes the tagging matrix.

Figure 4.5 shows the tagging matrices for b (a), c (b) and light quark (c) initiated jets. The probability to tag a light quark jet increases linearly with the E_{T}, while it reaches a plateau for b and c quark initiated jets. This is a known effect and it is due to the fact that the mis-tag rate depends on the number of SecVtx candidate tracks, which scales as the jet E_{T}.

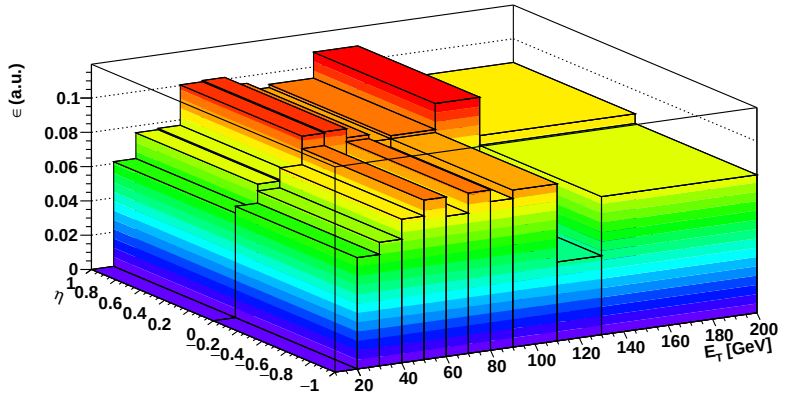
4.2.2 Heavy flavor content of the single tag sample

The flavour composition of the data selected depends on the number of the b-tagged jets, though it is expected to be dominated by b quark initiated jets also in the single tag sample. The TagMass, defined as the invariant mass of all tracks originating from the secondary vertex assumed to be charged pions ($M_{\pi} = 139\,\mathrm{MeV}/c^2$) is used to determine the heavy flavour content of this sample, in fact it carries the information about the heavy hadron that formed the secondary vertex. Light quarks and gluons, which can generate a secondary vertex tag only due to track mis-measurement, have low TagMass distribution. Hadrons originating in b-jets have larger invariant mass with respect to those originating in c-jets, so the TagMass distribution of the latter is distinguishable from the former.

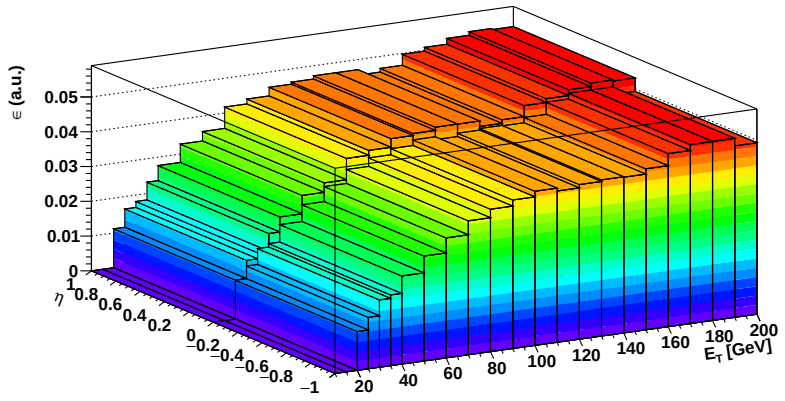
Figure 4.6 shows the TagMass distribution of the b-tag trigger jets, which is fitted with a binned maximum likelihood as sum of three contributions: b quark, c quark and light quarks initiated jets. The templates of the components are obtained from



(a)



(b)



(c)

FIGURE 4.5: The tagging matrices for b (a), c (b) and light quark (c) initiated jet as a function of E_{T} and η of the jet.

the TagMass distribution of the b-tag trigger jets in the $b\bar{b}$, $c\bar{b}$ and light quarks jets Monte Carlo samples described in Sec. 4.1.2.

The result of the fit shows that the composition of the sample made of jets that are

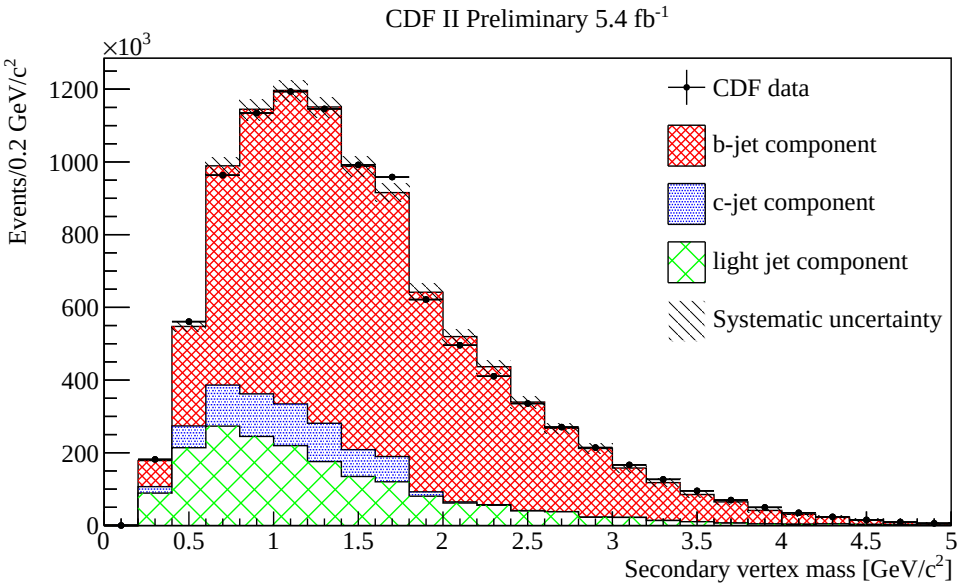


FIGURE 4.6: Fit to the invariant mass distribution of the charged tracks of the secondary vertex of b-tag trigger jets. The flavour composition of the sample is extracted with the fit.

selected by the on-line and the off-line b-tagging algorithms is: $(75 \pm 2)\%$ of b-quark, $(7 \pm 1)\%$ of c-quark and $(18 \pm 2)\%$ of light quarks initiated jets. The uncertainty is the quadratical sum of the statistical and the systematic uncertainties due to the statistics of the Monte Carlo templates.

Figure 4.6 shows that at high values of the TagMass the sample is made of almost pure b-quark jets. In fact, by requiring TagMass > 1.8 GeV/c^2, the b-quark component is 96% and the rest 4% is light quarks. With this requirement, we can select a pure sample of jets originating from b quarks.

4.2.3 The background templates

In order to be able to describe the different components beside the signal, of the double tagged sample, the following invariant mass templates are needed:

- the $b\bar{b}$ produced in QCD processes invariant mass distribution;
- the bc and bq (where c and light quark initiated jets are mis-tagged by the SecVtx algorithm) invariant mass distribution.

The contribution with multiple non-b jets is expected to be negligible and it is not considered. With the tagging matrices (see Section 4.2.1) and the study on the flavour composition of the b-tag trigger jet we have all the ingredients to build these background templates.

The starting point is the data sample with at least one b-tag trigger jet with a value of TagMass variable grater than 1.8 GeV/c^2. With these requirements we select real b-jets, as shown in Section 4.2.2. The tagging matrices are then applied to the other jet of the event (simply by reweighting it) to simulate the effect of the SecVtx b-tagging. In this way the expected invariant mass shape of the different backgrounds is simulated using data itself where the signal is present but its contribution is less than 0.1%. The configurations considered are:

Bb - bB - Bc - cB - Bq - qB

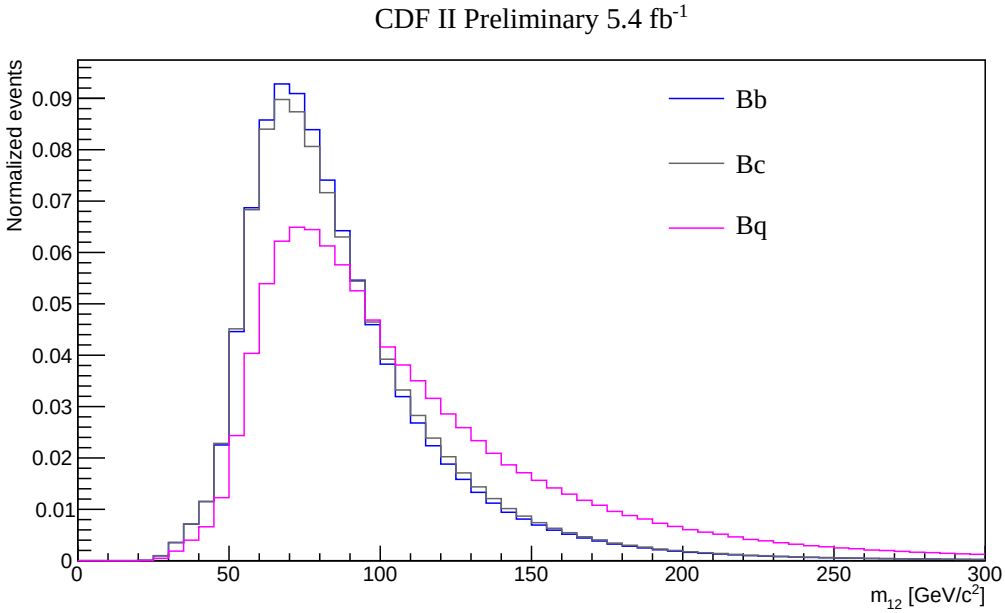


FIGURE 4.7: Invariant mass background templates. Comparison among the templates when the b-tag trigger jet is the leading one and the different flavor matrices are applied to the other jets.

The uppercase B indicates the b-tag trigger jet, lowercase letters give the flavor hypothesis obtained by reweighting the event by using the tagging matrices, where q indicates the light quark jets. The order of the letters follows the E_{T} ordering of the jets, for example bB means that the b-tag trigger jets is the second-leading one.

Figure 4.7 compares the templates when the b-tag trigger jet is the leading one and the different flavor matrices are applied to the other jets. The invariant mass spectrum (m_{12}) obtained with the light quarks jets is the harder, as expected since the mistag probability increases with the energy of the jet, while the Bb and the Bc templates are almost identical.

4.3 $Z \to b\bar{b}$ signal extraction

The search for the $Z \to b\bar{b}$ examines the invariant mass distribution, m_{12}, of the two leading jets of the double tag sample for an enhancement riding atop the continuum background. The search is accomplished by exploiting a maximum likelihood binned fit to data: fixed template shapes for the different background components and for the signal are used, the normalizations of the components are the free parameters of the fit. As described in Section 4.2.3, a cut on the TagMass of the b-tag trigger jet of the background templates is required. The same cut is then applied also in the double tag sample.

After the selection, the double tagged sample is made of 925338 dijet events. The dijet invariant mass signal template is obtained from fully simulated Monte Carlo $Z \to b\bar{b}$ events, 36548 events pass the trigger simulation and the off-line selection. In Figure 4.8 the $Z \to b\bar{b}$ signal template is shown superimposed to the data sample.

Since the invariant mass distribution templates built with the b-tagging and the c-tagging matrices are very similar and the fit is not able to distinguish them, they are merged assuming a 7% contribution (taken from the fit to the TagMass distribution

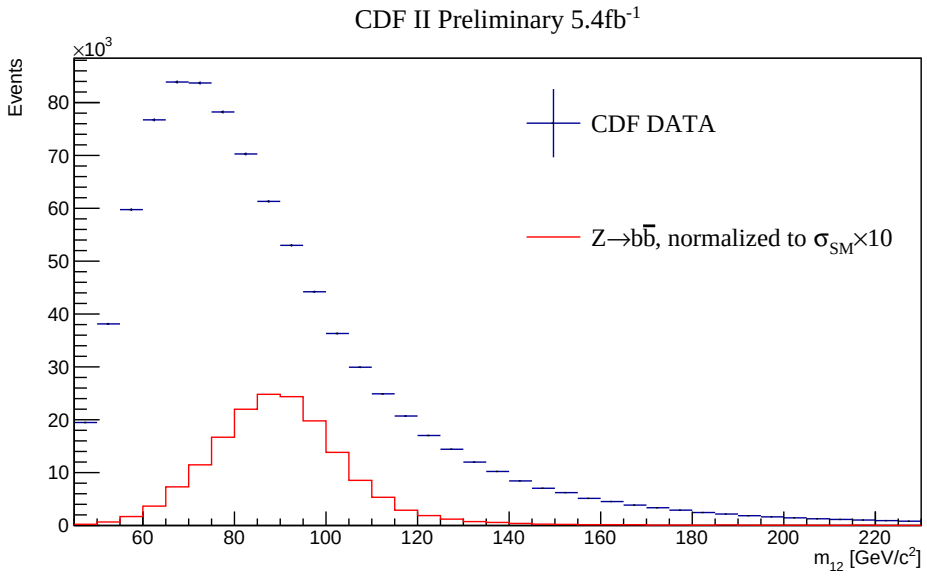


FIGURE 4.8: The double tagged sample m_{12} distribution after the selection. In red the $Z \to b\bar{b}$ template, normalized to the Standard Model expectation magnified 10 times.

of the b-tag trig jets in Section 4.2.2) of Bc and cB events to the Bb and bB templates respectively. A systematics will be assigned for this assumption. Thus, to model the QCD multijet background 4 different templates are used:

(Bb+Bc) - (bB+cB) - Bq - qB

The fit is performed using a binned maximum-likelihood function, defined as:

$$\mathcal{L} = \prod_{i=1}^{N} \frac{n_{\rm s}^i P_{\rm s}(m_{12}^i) + \sum_{\rm b} n_{\rm b}^i P_{\rm b}(m_{12}^i)}{n_{\rm s}^i + \sum_{\rm b} n_{\rm b}^i} \tag{4.2}$$

where $\mathcal{L}$ is the product, over all bins, of the probability that the events in the ith bin with invariant mass m_{12} are described by the 4 background p.d.f $P_{\rm b}(m_{12})$ plus the signal p.d.f. $P_{\rm s}(m_{12})$. The free parameters are the number of signal ($n_{\rm s}$) and background ($n_{\rm b}$) events that are constrained to be greater or equal to zero.

4.3.1 Fit Results

The fit to the double tag data sample is performed. Figure 4.9 shows data with the result of the fit superimposed. The results for signal and background yields are listed in Table 4.1 where the uncertainties are statistical only. The fit returns a sizable signal component and the light quark component compatible with zero, indicating that the sample is constituted mostly by $b\bar{b}$ jets. The goodness of the fit is estimated by calculating the χ^2/NDF, which is found to be 0.7.

Signal significance

The significance of the signal is measured by computing the p-value, i.e. the probability that the background fluctuates to create the observed signal. 50 millions pseudo-experiments have been generated in the background only hypothesis, H_0,

TABLE 4.1: Signal and background yields as returned by the fit to the double tagged sample.

CDF II Preliminary 5.4 fb^{-1}

Component	Fitted yield in events
$Z \to b\bar{b}$	$(16.5 \pm 1.2) \times 10^3$
Bb+Cb	$(68.1 \pm 1.1) \times 10^4$
bB+bC	$(19.4 \pm 1.3) \times 10^4$
Bq	$< 175\ (1\sigma)$
qB	$< 61\ (1\sigma)$

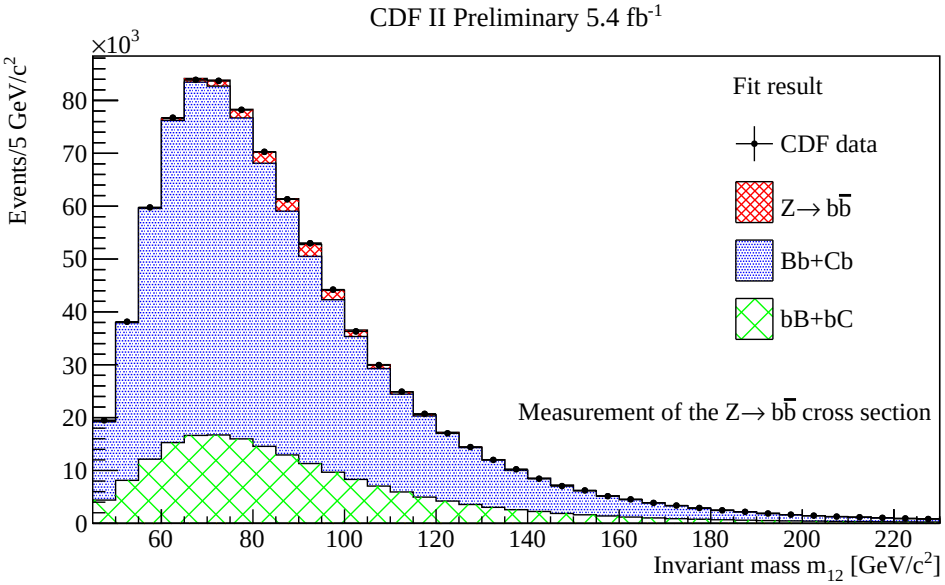


FIGURE 4.9: Double tagged events invariant mass distribution with the result of the fit. In red the fitted $Z \to \bar{b}b$, in blue the Bb+Cb and in green the bB+bC background. Capital letter indicates the b-tag trigger jet.

and additional 50 millions in the background plus signal hypothesis, H_1. Systematics uncertainties, which will be described in Section 4.4, are introduced as nuisance parameters.

The Wilks theorem [72] is exploited to compute the significance. The test statistic T_s employed is the usual likelihood ratio under the H_1 hypothesis ($\mathcal{L}_1$) and under the H_0 hypothesis ($\mathcal{L}_0$). This ratio can be approximated with the difference of the fit to the data χ^2 under the two hypotheses:

$$T_s = -2ln\frac{\mathcal{L}_1}{\mathcal{L}_0} = \chi^2(data|H_1) - \chi^2(data|H_0). \tag{4.3}$$

More signal-like pseudo-experiments would have lower values of the test statistic. The test statistic is evaluated for each of these 50 millions pseudo-experiments, and the expected significance is defined as the probability for a background only pseudo-experiment to have a T_s less than the T_s observed in data.

The significance is expressed in terms of σ, the standard deviation of a Gaussian distribution. The conversion from the p-value to the number of σ is done by finding the value x such that the integral of a Gaussian distribution with mean zero from $x\sigma$

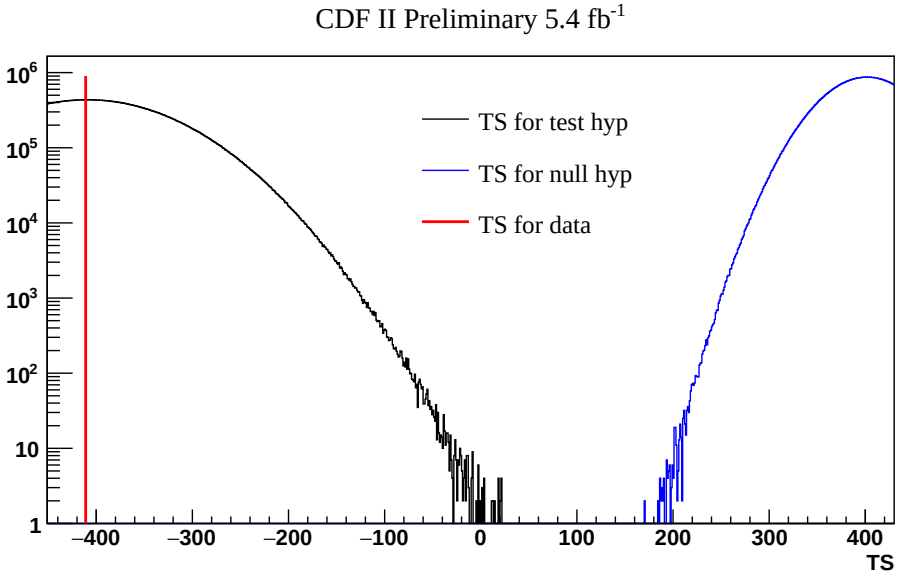


FIGURE 4.10: Test statistic distribution for signal-like (black) and background-only (blue) pseudo-experiments. The red line represents the observed test statistic value for the data.

to ∞ is equal to the p-value. For example, a p-value of 2.7×10^{-3} corresponds to a significance of 3σ.

Figure 4.10 shows the distribution of the T_s for background only and signal+background pseudo-experiments, as well as for the data. None of the 50 millions background only pseudo-experiment has a smaller T_s than the one evaluated in the data. Hence, we can conclude that the investigated Z signal has an observed significance greater than 5σ.

4.3.2 Jet energy scale determination

Since a sizable signal of $Z \to b\bar{b}$ decays is established, a measurement of the residual energy scale for b-jets between data and Monte Carlo is possible. The procedure is based on the fit to the data with the Z invariant mass template constructed varying the jet energy. Jet energy of the $Z \to b\bar{b}$ Monte Carlo jets is multiplied by a factor k, which varies between 0.90 and 1.10 in steps of 0.01. This range largely covers the possible variation of this parameter since jets at CDF are well reconstructed and previous analysis [66] have shown that for b-jets a minor correction is needed.

By varying k, 21 different dijet mass signal templates are built. The fit, described in Sec. 4.3.1, is performed to data sample for each signal template and the χ^2 calculated. The value of k which correspond to the minimum of the χ^2 distribution represents the value of the jet energy scale between data and Monte Carlo b jets. The statistical error is calculated taking the width of k interval corresponding to $\chi^2 = \chi^2(k_{\min} + 1)$. The measured JES is then $k = 0.993 \pm 0.022$.

Fit stability is verified by using pseudo-experiments. Given that the signal invariant mass distribution peak is very close to the background one, we wanted to check if this could bias the resulting b-JES. We build a data-like distribution by summing background p.d.f. and signal p.d.f. with jets energy multiplied by the k factor. Then for a given simulated input b-JES factor pseudo-data templates are generated by drawing n_b background events from the background p.d.f. and n_s signal events

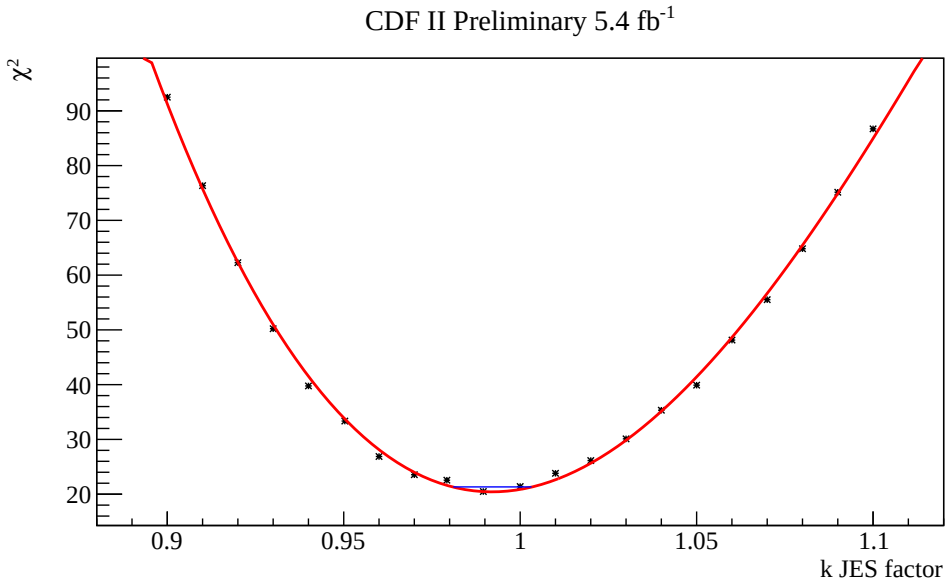


FIGURE 4.11: The parabolic shape of χ^2 of the different fits as function of the JES parameter k. The straight line correspond to the width of k interval corresponding to $\chi^2 = \chi^2(k_{min} + 1)$

from the signal p.d.f. The input b-JES factor varies from 0.95 to 1.05. The pseudo-data is fitted using the standard procedure.

Figure 4.12 shows the mean fitted output scale factor as a function of the input scale factor. No bias from our fitting procedure is observed and the results obtained shows that the b-JES factor is not affected by the fitting procedure.

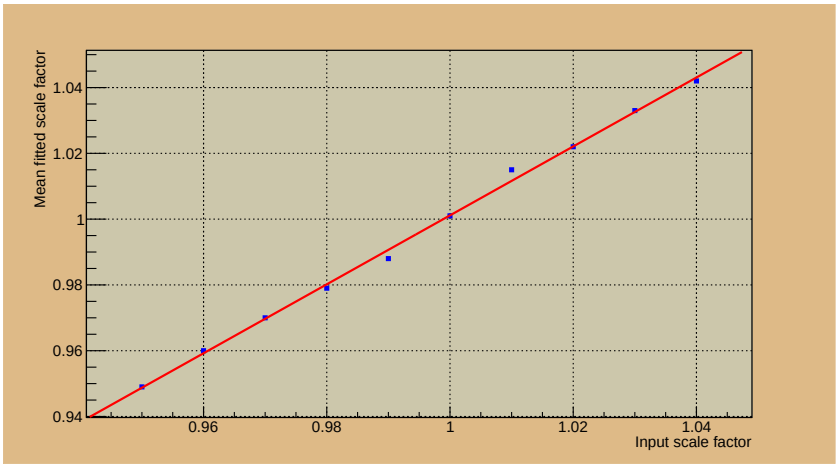


FIGURE 4.12: Mean fitted b-JES factor as function of the input scale factor

4.3.3 Cross section measurement

The fitted Z yield can be used to measure the cross section for Z boson production multiplied by the branching ratio of the decay to b-quark pairs. The comparison

with the theory prediction gives a powerful check on the correctness of the analysis results.

Figure 4.13 shows the double tagged events invariant mass distribution after the background subtraction compared to the Monte Carlo $Z \to b\bar{b}$ signal template. The agreement between the data after the background subtraction and the Monte Carlo signal simulation is very good. From the fitted number of signal events we extract the cross-section using the formula:

$$\sigma_Z \times B(Z \to b\bar{b}) = \frac{N_{\text{sig}}}{\epsilon_{\text{kin}} \cdot \epsilon_{\text{trig}} \cdot \epsilon_{\text{tag}} \cdot SF_{\text{trig}} \cdot SF_{\text{tag}} \cdot \mathcal{L}} \tag{4.4}$$

Table 4.2 summarizes the definition and the values used to calculate the cross section of each quantity.

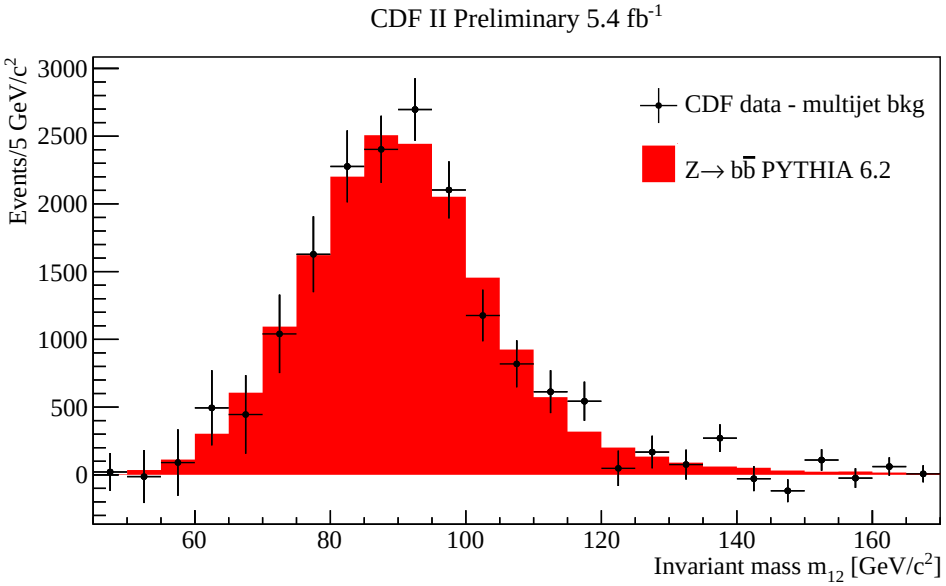


FIGURE 4.13: Double tagged events invariant mass distribution after the background subtraction. The Z peak is clearly visible and it is in good agreement with the MC signal template, in red.

TABLE 4.2: The parameters definition and value used in the cross section measurement.

Quantity	Description	Value
ϵ_{trig}	Trigger efficiency	6.4%
ϵ_{kin}	Kinematic cut efficiency	56%
ϵ_{tag}	Tagging efficiency	11%
SF_{trig}	Scale factor trigger	0.68
SF_{tag}	SecVtx tagging scale factor	0.86
$\mathcal{L}$	Luminosity	5.4 fb^{-1}

The efficiencies are evaluated using the Monte Carlo simulated signal sample, therefore they are multiplied by the scale factors to take into account the trigger and SecVtx b-tagging data/Monte Carlo differences.

The calculated cross section value is then

$$\sigma_Z \times B(Z \to b\bar{b}) = 1.11 \pm 0.08 \text{ nb} \tag{4.5}$$

where the uncertainty is statistical only.

4.4 Systematic uncertainties

Both the $Z \to b\bar{b}$ cross section and Jet Energy Scale measurements are affected by the systematic uncertainties. Some of them are related to data/Monte Carlo differences, while other are related to the signal extraction procedure. They are summarized in Table 4.3 and discussed in more detail in the following.

TABLE 4.3: Summary of the systematic uncertainties.

CDF II Preliminary 5.4 fb^{-1}

Source	Affecting JES	Affecting the cross section
Luminosity		5.9%
Background template statistics	0.004	2.3%
c-quark component in $b\bar{b}$ templates	0.005	2%
Signal Monte Carlo statistics	0.002	3%
b-tag energy dependence	0.004	5%
b-tag scale factor		5%
Trigger and b-tag combined scale factor		4%
JEC		1.4%
FSR		2.6%
PDF		1.1%
Total	0.008	11.4%

Systematics related to the background modeling

Three sources of systematic uncertainty are associated with the background modeling. The first one is due to the finite statistics of the background templates. To estimate the size of this effect we perform pseudo-experiments by smearing the number of events in each bin of the background templates and measure the resulting bias on the fitted b-JES and signal yield. The value is found to be ± 0.004 for the JES and 2.3% for the signal yield. As described in 4.3.1 then, a fixed percentage of 2% of c-quark jet component is added to the Bb and bB templates. This value is evaluated from studies on the TagMass and a systematic is set varying this percentage from 0 to 10%, which leads to an uncertainty of ± 0.005 on JES and of 2% on the signal yield.

Systematics related to the scale factors

Figures 3.4 and 3.6 show that the data/Monte Carlo b-tagging scale factors have a small transverse energy dependence, while they have been considered flat through the analysis. We fit the distribution of the scale factors and we implement the energy dependence by weighting the two jets in the signal Monte Carlo events according to these fits. The absolute difference in JES and signal yield is taken as systematic uncertainty and results ± 0.004 and 5% respectively. The data/Monte Carlo scale factors are also applied to simulated events in the evaluation of the signal selection efficiency. They are affected by systematic uncertainties which propagate to the cross section measurement.

Other systematic uncertainties

The systematic uncertainty related to the finite statistics of the MC signal template is evaluated through pseudo-experiments and it results to be ± 0.002 for JES and 3% for

the signal yield. The systematic uncertainty on the signal efficiency due to the CDF jet energy correction is estimated by shifting the energy of the Monte Carlo jets by $\pm 1\sigma_c$ of the standard jet energy correction and it is found to be 1.4% on cross-section. The effect of decreased or increased Final State Radiation on signal efficiencies has been evaluated generating $Z \to b\bar{b}$ signal samples with different FSR tunings and this gives a systematic uncertainty of 2.6% on the cross-section. For what concern the systematics due to the choice of the Parton Density Functions we generated a new $Z \to b\bar{b}$ Monte Carlo sample using the CTEQ6L set. The difference in acceptance is then taken as systematic uncertainty: 1.1% on the cross-section. The systematic uncertainties due to the choice of the Final State Radiation and of the Parton Density Functions are not evaluated for the JES, because the JES describes the agreement of the jet energy between data and the Monte Carlo simulation generated with the standard CDF parameters.

4.4.1 Final results

At this point, all the steps needed to extract the cross section from the dataset have been performed. The measured cross section, including the systematic uncertainties, is:

$$\sigma_Z \times B(Z \to b\bar{b}) = 1.11 \pm 0.08(stat) \pm 0.13(sys) \text{ nb}, \tag{4.6}$$

which is consistent with the NLO theoretical calculation [73] combined with the measured $Z \to b\bar{b}$ branching ratio, which predicts a cross section value of $\sigma_Z \times B(Z \to b\bar{b}) = 1.13 \pm 0.02$ nb. The good agreement with the theoretical prediction is a strong check that the analysis procedure, and in particular the background modeling, is correct.

The measured jet energy scale for b-jets, including the systematic uncertainties is:

$$\text{JES} = 0.993 \pm 0.022 \pm 0.008. \tag{4.7}$$

The value is compatible with 1, which means that the b-jets simulation energy scale is well calibrated.

4.5 The $Z \to b\bar{b}$ at LHCb

The measurement of the $Z \to b\bar{b}$ cross section, and the consequent determination of the energy scale for b-jets, has been performed also at LHCb. This analysis proves that the LHCb b-jet reconstruction procedure works well and that this experiment can be exploited to search for resonances in $b\bar{b}$ final states. An exhaustive description of the analysis can be found in this paper [87]. The analysis uses 2 fb^{-1} of data collected at LHCb at a centre of mass energy of 8 TeV. It proceeds through the following steps:

1. a b-dijet selection is applied to the dataset collected in 2012, requiring two reconstructed jets identified as originating from a b quark;
2. a recoil jet selection is applied. The recoil jet can be used to define discriminating observables to separate the $Z \to b\bar{b}$ from the QCD background;
3. a multivariate technique, the uGB method see Section 5.2.4 for more details, is employed to build a discriminator uncorrelated with the dijet invariant mass;
4. the MVA output is used to define a control region with low $Z \to b\bar{b}$ contribution where the QCD background mass model can be probed;

5. A simultaneous fit to the b-dijet invariant mass distributions in the signal and control regions is performed to determine the $Z \to b\bar{b}$ yield and the jet energy scale factor, k_{JES}.

The final fit result is shown in Figure 4.14. The measured jet energy scale factor is:

$$k_{\text{JES}} = 1.009 \pm 0.015(stat) \pm 0.011(sys), \tag{4.8}$$

which is compatible with unity, which demonstrates that the LHCb simulation reproduces accurately the b-jet energy, while the measured $Z \to b\bar{b}$ cross section in the LHCb fiducial region is:

$$\sigma(pp \to Z)\mathcal{B}(Z \to b\bar{b}) = 332 \pm 46(stat) \pm 59(sys) \text{ pb}, \tag{4.9}$$

in good agreement with the NLO prediction, $\sigma(pp \to Z)\mathcal{B}(Z \to b\bar{b}) = 272^{+9}_{-12}(scale) \pm 5(PDF)$ pb. The dominating systematic uncertainty is the one related to the heavy-flavour tagging efficiency.

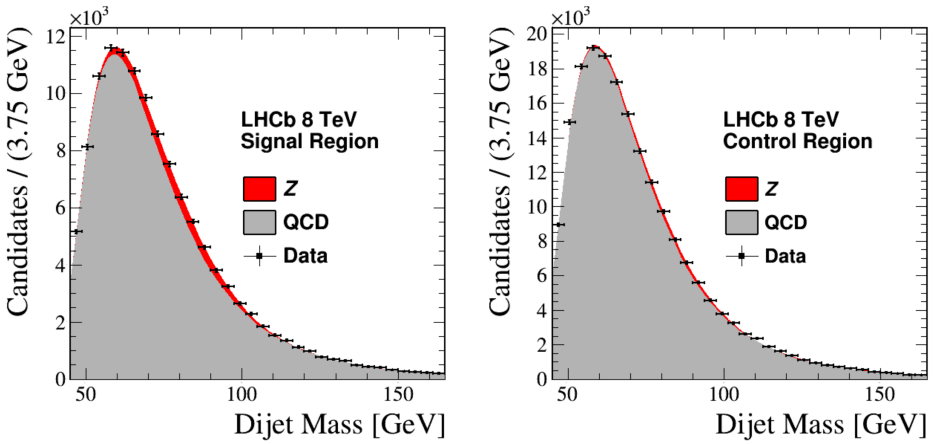


FIGURE 4.14: Simultaneous fit to the dijet invariant mass distribution of $Z \to b\bar{b}$ candidates in the (left) signal and (right) control regions.

Chapter 5

The inclusive Higgs production

The measurement of the inclusive $H \to b\bar{b}$ process, although of the great interest because it can be a probe of production mechanisms beyond the Standard Model, has always been considered a too difficult channel due to the overwhelming QCD background. The major effort in this direction at the general purpose experiments at LHC is focused on the boosted production, where the QCD background can be kept under control [4].

The measurement of the $Z \to b\bar{b}$ cross section at the CDF II, described in Chapter 4, has been possible thanks to the implementation of an ad-hoc trigger for the selection of events with b-jets and to the smaller QCD multijet background production due to the lower center of mass energy with respect to the LHC. The same event selection and background description used for is here exploited to search for the inclusive Standard Model Higgs decaying into a b quark pair. The production cross section for this process is about 3 order of magnitude smaller with respect to the $Z \to b\bar{b}$ decay, far beyond the reachable sensitivity. No signal is then expected, so a limit on the production cross section times the branching ratio for the inclusive $H \to b\bar{b}$ is set for the first for a completely inclusive selection.

If CDF II dataset gives the possibility to set the first limit on the inclusive $H \to b\bar{b}$ process, on the other hand the Higgs production cross section at the Tevatron was far too little to allow an observation, also with much more collected statistics. The possibility to reconstruct, select and identify b-jets at LHCb, as described in Chapter 3, opens the prospect of an observation of the inclusive $H \to b\bar{b}$ channel without further requirements on the events topology. With respect to the other general purpose experiment at LHC in fact, LHCb can profit from low p_{T} triggers, low background due to pile up thanks to the leveled luminosity and from a very good performing flavor jet tagging algorithm. All this, leads to a data sample of pure b-jets which also covers the low invariant mass spectrum and in an unique forward region. With respect to CDF instead, LHCb can profit from the higher centre of mass energy delivered by the LHC, which leads to better a signal over background ratio. The drawback for this search performed at LHCb is the small signal acceptance.

In this chapter a limit on the inclusive $H \to b\bar{b}$ process is performed with both the CDF II and the LHCb datasets. Section 5.1 reports the analysis using the CDF II facility, while Section 5.2 describes a first sensitivity study of this search at LHCb in the near future. This kind of feasibility analysis is important especially thinking about the next detector upgrade. It can also serve to push for a development and re-optimization of the jets tools for the next data taking. Section 5.3 in fact, describes the future prospects for this search.

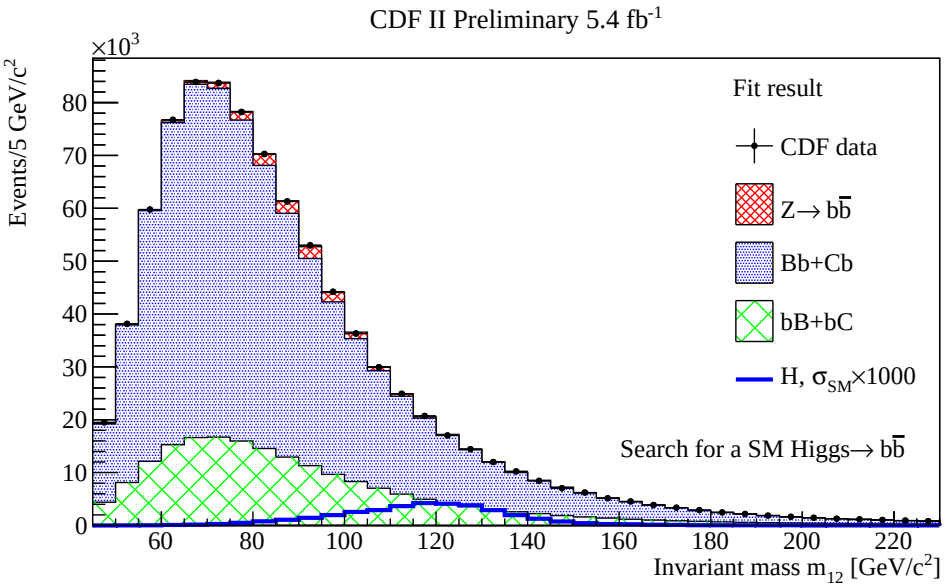


FIGURE 5.1: Invariant mass distribution of the double tagged data sample with the result of the fit that includes the $H \to b\bar{b}$ decay. The normalization of the Higgs signal is set to $\times 10^3$ the expected SM cross section for illustrative purposes.

5.1 Standard Model Higgs into $b\bar{b}$ at CDF II

The same dataset, with the same event selection used for the $Z \to b\bar{b}$ analysis at CDF, can be used to extract information about the inclusive production of the Higgs boson into a b-quark pair. The search is performed fitting the double tagged sample by using the templates described in section 4.2.3 for the QCD multijet and the $Z \to b\bar{b}$ template from the simulation as backgrounds.

At Tevatron the predicted total Standard Model Higgs production cross section is 1.23 ± 0.22 pb [12], while the branching ratio into a pair of b-quarks is $(58.4 \pm 3.3)\%$ [12], more than 3 order of magnitude smaller with respect to the $Z \to b\bar{b}$ process. A Monte Carlo simulated signal sample, see 4.1.2, has been used to evaluated the trigger and off-line selection efficiency, which turns out to be of 1.5%. Given the integrated luminosity of 5.4 fb^{-1}, we expect about 36 signal events in the final dataset. The background under the signal region, defined as the 2 times the signal resolution σ (19 GeV/c^2) around the nominal Higgs mass, is made of about 670 k events. Due to the extremely small S/B, no Higgs events are expected to be found.

Figure 5.1 shows the result of the fit to the double tagged sample, with the Higgs component magnified 1000 times with respect to the Standat Model expectation. The fit strategy is the same explained in Sec. 4.3.1 with the $H \to b\bar{b}$ template, taken from the simulation, added. The normalizations of all the components are the unconstrained parameters of the fit. The fit returns 0 ± 91 Higgs event, while the other background yields remain as in Table 4.1.

5.1.1 Cross section times branching ratio limit

Since no signal has been found, a 95% confidence level (C.L.) upper limit on the inclusive production of the Standard Model Higgs is set using a modified frequentist CL_S method [74]. The limit calculator is based on the MCLIMIT package [75]. The

compatibility of the data with the background only and background plus signal hypothesis is tested by constructing a test statistic, t, which in this case is the difference in χ^2 of the fits in the background only and background plus signal hypotheses.

The CL_S are defined as:

$$CL_S = \frac{p_s}{1 - p_b} \tag{5.1}$$

where p_s is the p-value (the probability to observe a value of the test statistic equal to or more extreme than the observed one in data) in the signal plus background hypothesis and p_b is the p-value in the background only hypothesis. They can be calculated from the probability density functions ($p.d.f.$) for the test statistic in the two hypotheses. The 95% confidence level upper limit on the signal cross section times branching ratio is obtained by finding the value of $\sigma \times \mathcal{B}$ for which $CL_S = 0.05$. The $p.d.f.$s can be generated by using toy Monte Carlo pseudo data, spanning on different $\sigma \times \mathcal{B}$ values for the signal plus background hypothesis, and null signal in the case of the background only hypothesis. For each pseudo data sample a value of the test statistic is found in the two hypotheses. Generating a large amount of pseudo data samples, the distributions of the possible values of the test statistic in the two hypotheses are found as well as the CL_S for the observed data as a function of the signal $\sigma \times \mathcal{B}$.

The upper limits in the case background only hypothesis, which is known as the expected median upper limit is calculated by generating a large set of background only pseudo data samples and calculate the upper limits on the signal $\sigma \times \mathcal{B}$ at 95% confidence level for each of them, as if they were real data.

The systematic uncertainties, listed in Table 4.3 for backgrounds and in Table 5.1 regarding the signal, can affect both the normalization and the shape. They have been calculated as described in Sec. 4.4 and they are introduced in the limit calculator as nuisance parameters.

TABLE 5.1: Summary of the systematic uncertainties on the $H \to b\bar{b}$ cross section limit.

CDF II Preliminary 5.4 fb^{-1}

Systematic uncertainty source on $H \to b\bar{b}$	Variation	Type
Luminosity	5.9%	Rate
b-tag scale factor	5%	Rate
Trigger and b-tag combined scale factor	4%	Rate
Final State Radiation	2.9%	Rate
Jet Energy Correction	2.0%	Rate/Shape
Parton Density Functions	1.4%	Rate
Signal Monte Carlo statistics	-	Shape
Total	9.5%	

The expected and observed CL_S obtained as a function of the ratio between the cross section upper limit and the Standard Model cross section are presented in Figure 5.2. The observed(expected) upper limit at 95% C.L. on the $p\bar{p} \to H \to b\bar{b}$ is found to be 33(46) times the standard model cross section.

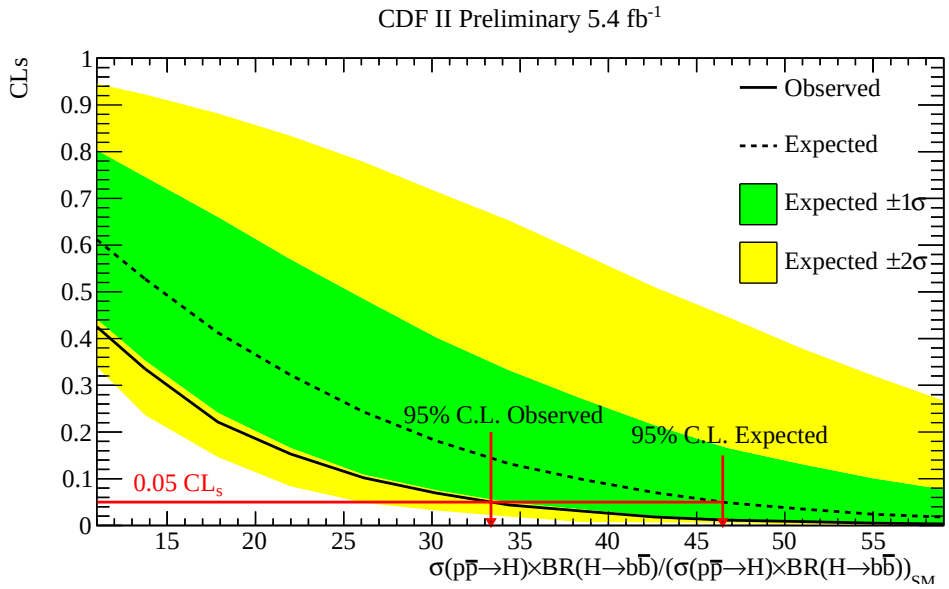


FIGURE 5.2: Observed (black solid line) and expected (black dashed line) CL_s as function of the cross section times the branching ratio normalized to SM $H \to b\bar{b}$. Green and yellow regions are respectively the 1σ and 2σ bands on the expected CL_S. The 95% C.L. limits are indicated by the red vertical arrows under the red horizontal line.

5.2 Standard Model Higgs into $b\bar{b}$ at LHCb

5.2.1 Data and simulated samples

The dataset corresponds to 1.64 fb^{-1} of integrated luminosity collected in 2016 at a center-of-mass energy of $\sqrt{s} = 13$ TeV. The sample has been collected by requiring events where at least one jet passes any L0 and HLT1 selection and at least two jets pass the dedicated HLT2 jets trigger "Hlt2JetsDiJetSVSVLine" described in Section 3.3.3, which requires two HLT jets with $p_{\mathrm{T}} > 17$ GeV and both tagged with a secondary vertex. At the off-line level, jets are reconstructed using the Run I standard LHCb tools described in Section 3.3.1. From now on, these jets are referred as Std jets.

A gluon fusion $H \to b\bar{b}$ Monte Carlo simulated sample is used to study the signal properties an to determine the trigger and off-line selection efficiency. The Higgs is produced with a mass of 125 GeV/c^2 and the Standard Model properties. The simulation has been performed with the LHCb framework GAUSS [112]. About 200k events have been generated with PYTHIA 8 [113] requiring the two b quarks from the H decay inside the LHCb acceptance, $2 < \eta < 5$.

The main background for this search, beside the QCD heavy quarks jet production, is the $Z \to b\bar{b}$ decay. About 200k $Z \to b\bar{b}$ events have been generated with PYTHIA 8 [113], requiring the two b quarks inside the LHCb acceptance, in order to study the selection efficiency for this process. Other backgrounds, like $t\bar{t}$ and $W \to q\bar{q}$ contribute with just few events under the signal region so they are neglected in this study.

A simulated sample of the $b\bar{b}$ QCD background is used to validate the parametrization used to described the QCD background while setting the limit in data. This sample is the weighted sum of sub-samples generated with PYTHIA 8 with different lower threshold in the p_{T} of the higher jet in the event.

5.2.2 Events selection

The events selected by the HLT2 "Hlt2JetsDiJetSVSVLine" trigger are mainly events with at least two heavy quarks initiated jets. Further requirements are applied, in order to select the dijet candidates to increase as much as possible the signal significance before setting the limit on the $H \to b\bar{b}$ cross section.

First, only candidates where the two jets fired the HLT2 trigger are used. To do so, the HLT jet builder algorithm is run off-line and HLT jets so reconstructed are required to have a secondary vertex inside the jet cone. In this way the trigger is simulated off-line. At this point, one Std jet is said to fired the trigger if it is matched to one SV tagged HLT jet. The matching is performed by requiring that $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.5$.

Further requirements are applied to the two matched Std jets:

- both jets have to have a positive SV tagger decision, see Section 3.3.2; this requirement almost perfectly overlap with the HLT2 requirement,
- they have to be associated to the same PV
- $p_{\mathrm{T}}(jet1, jet2) > 20$ GeV/c to reduce the QCD background at low dijet invariant mass but preserving most signal events,
- $\Delta\phi > 1.5$,
- $2.2 < \eta(jet1, jet2) < 4.2$ to ensure that the entire jets cones are inside the fully instrumented LHCb acceptance.

Only events with a dijet invariant mass between 60 and 180 GeV/c^2 are considered. The signal region is then defined as the region with dijet invariant mass between 80 and 150 GeV/c^2.

At this stage, the selected sample has a not negligible component of $c\bar{c}$ pairs, about 20%. We can increase the b-jets purity by setting a cut on the $\mathrm{BDT}_{b|c}$ SV tagger variable described in Section 3.3.2. The signal significance as a function of the $\mathrm{BDT}_{b|c}$ cut has been computed, leading to the further requirement:

- $\mathrm{BDT}_{b|c}(jet1, jet2) > 0.1$

The two selected jets, $jet1$ and $jet2$, form the b-dijet candidate. If more than one b-dijet candidate is found in the same event the one with the highest $p_{\mathrm{T}}(jet1) + p_{\mathrm{T}}(jet2)$ is kept. Finally, only b-dijet candidate within the mass range $80-180$ GeV/c^2 are considered.

The signal efficiency is evaluated with the Monte Carlo simulation sample. The trigger efficiency is about 15%. Adding the off-line requirements, the efficiency drops to 7.2% with respect to the number of generated events. In Table 5.2 the cumulative efficiencies of the selection cuts on the $H \to b\bar{b}$ signal are listed.

5.2.3 Yields prediction

The predictions of the $H \to b\bar{b}$ signal and of the $Z \to b\bar{b}$ background yields after the full selection have been obtained using the formula

$$N^{exp} = \mathcal{L} \cdot \sigma \cdot \mathcal{A} \cdot \epsilon, \tag{5.2}$$

where:

TABLE 5.2: Cumulative efficiencies of the selection cuts on the $H \to b\bar{b}$ signal.

Requirement	Efficiency
L0	83%
HLT1	75%
HLT2	14%
$(jet1, jet2)$ SV tagged	14%
$p_{\mathrm{T}}(jet1, jet2) > 20$	13%
$\Delta\phi(jet1, jet2) > 1.5$	12%
$2.2 < \eta(jet1, jet2) < 4.2$	9%
$\mathrm{BDT}_{b\|c}(jet1, jet2) > 0.1$	6%

- $\mathcal{L}$ is the integrated luminosity
- σ is the theoretical cross section, scaled for the branching ratio. The value for the H signal is $\sigma \times \mathcal{B}(H \to b\bar{b}) = 29.6 \pm 3.6$ pb, taken from [12]. The value for the Z background is $\sigma \times \mathcal{B}(Z \to b\bar{b}) = 8.5 \pm 0.3$ nb, obtained scaling the theoretical $Z \to \mu\mu$ cross section at NNLO computed in [125] for the corresponding measured branching ratios taken from [12].
- $\mathcal{A}$ is the LHCb acceptance factor calculated using PYTHIA, requiring the decay products in the LHCb acceptance ($2 < \eta < 5$). The Higgs acceptance is 0.051.
- ϵ is the selection efficiency, obtained as the number of selected Monte Carlo events divided by the number of generated events.

The efficiencies are corrected for the data/Monte Carlo differences by using the scale factors for trigger and tagging. The SV b-tagging responses has been evaluated as described in Section 3.3.2 and it is listed, as a function of the p_{T}, in Table 3.4.

The trigger scale factor is factorized for the three different trigger level selections: L0, HLT1 and HLT2. The L0 and the HLT1 scale factors have been evaluated using a tag and probe method. The same $(Z \to \mu^+\mu^-) + jet$ event selection, described in Section 3.3.3 for the validation of the jet energy scale in Run II, is required on 1.6 fb^{-1} of 2016 data. The $Z \to \mu^+\mu^-$ candidate is the tag and the recoil jet candidate is the probe, which is also required to be SV tagged. The L0(HLT1) efficiency is thus obtained as the ratio between the number of probe jets selected by the L0(HLT1) trigger divided by the total number of probe jets, as a function of the jet p_{T}. Figure 5.3 shows the L0 efficiency both for the data and the simulated sample. It is clear that the L0 efficiency in the simulated sample is higher with respect to data. A scale factor for the L0 data/Monte Carlo difference is then evaluate. We required that at least one of the two selected jets to pass the L0, the combined efficiency can be obtained as follow:

$$\epsilon = 1 - (1 - \epsilon(jet1))(1 - \epsilon(jet2)) \tag{5.3}$$

Considering the efficiency flat as a function of the p_{T}, then $\epsilon(jet1) = \epsilon(jet2)$, which leads to a scale factor for L0:

$$SF_{\mathrm{L0}} = \frac{2\epsilon_{\mathrm{data}} - \epsilon^2_{\mathrm{data}}}{2\epsilon_{\mathrm{MC}} - \epsilon^2_{\mathrm{MC}}} = 0.93 \tag{5.4}$$

Figure 5.4 shows the HLT1 tagged jet efficiency both for the data and the simulated sample. In this case there is a good agreement between data and simulation, so there is no need for a dedicated scale factor.

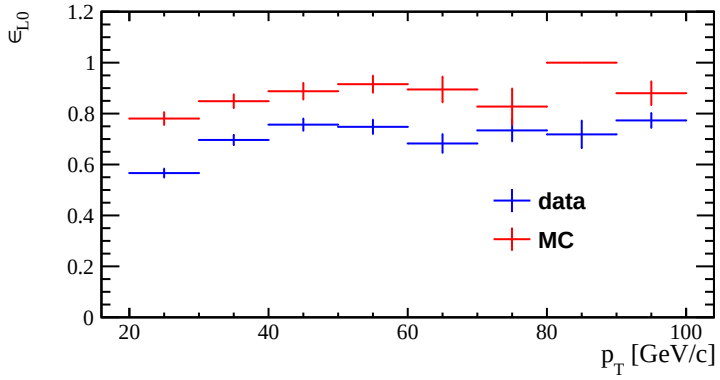


FIGURE 5.3: L0 tagged jet efficiency for data and Monte Carlo samples as a function of the jet p_{T}.

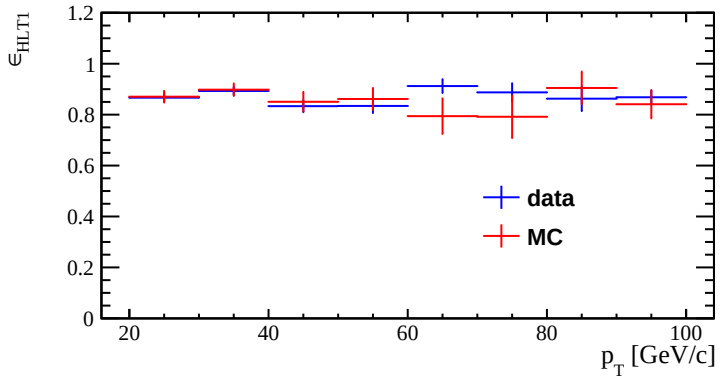


FIGURE 5.4: HLT1 tagged jet efficiency for data and Monte Carlo samples as a function of the jet p_{T}.

For the HLT2, the only source of possible data/simulation disagreement can come from the HLT/Std jet matching procedure described in 5.2.2. To evaluate the efficiency of the jet matching procedure a tag and probe method is again used, by exploiting a sample of tagged dijets. One SV tagged and HLT matched jet, the tag, is selected. Another jet with $\Delta\phi > 2.7$ with respect to the first, the probe, is then required. The matching efficiency is defined as the ratio between the number of HLT matched jets and the number of total probe jets. Figure 5.5 shows the HLT matching jet efficiency for data and Monte Carlo as a function of the jet p_{T}. The efficiency is almost one, as expected, and the data/simulation agreement is good.

The final number of expected $H \to b\bar{b}$ events, considering all the efficiencies and scale factors, is 156. The number of expected $Z \to b\bar{b}$ events is 175k.

5.2.4 Multivariate technique selection

It is of primary importance then to increase as much as possible the signal significance in the final data sample avoiding to introduce any bias in the dijet invariant mass distribution. To do so, the discriminating power of a multivariate technique classifier (MVA) is exploited.

In particular, a Boosted Decision Trees (BDT) classifier is used for this purpose [126, 127]. The structure and the weights of the BDT defined in [126, 127] are determined with an iterative procedure called training, performed using a signal sample and a background sample. A so called uniform Gradient Boost method (uGB) [128], which

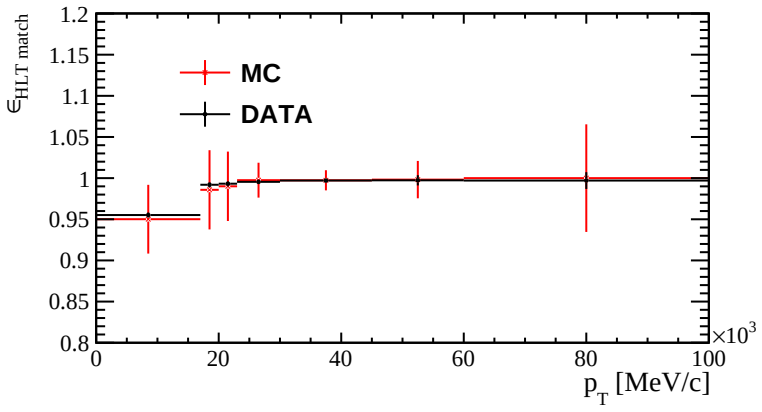


FIGURE 5.5: HLT matching jet efficiency for data and Monte Carlo as a function of the jet p_{T}.

is applied to the BDT, is used to have the classifier uncorrelated with the invariant mass distribution, where the signal is searched.

In order to explain the uGB method several quantities are defined:

- $\vec{x}$: the input observables;
- $\vec{y}$: the observables we want uncorrelated from the BDT output, different from $\vec{x}$;
- γ_i: it is equal to +1 if the i-th event in the training is a signal event or -1 if it is a background event;
- the tree response r_{ij}: it is equal to +1 if the i-th event is classified by the j-th tree as signal or -1 if it is classified as background;
- the tree weight w_j, which is assigned to each tree j in the ensemble of trees (forest) at each stage of the BDT training;
- the event score s_i:
$$s_i = \frac{\sum_j^{N_{tree}} w_j r_{ij}}{N_{tree}}, \tag{5.5}$$
where N_{tree} is the number of trees in the forest;
- the AdaBoost loss function:
$$L_{ada} = \sum_i^N exp(-\gamma_i s_i), \tag{5.6}$$
where N is the number of events in the training;
- the flatness loss function, defined by dividing the $\vec{y}$ space in bins (b):
$$L_{flat} = \sum_b f_b \int |F_b(s) - F(s)|^p ds, \tag{5.7}$$
where $F(s)$ is the cumulative distribution of the score for the events in the training, $F_b(s)$ is the cumulative distribution of the score for the events in the bin b, f_b is the fraction of signal events in the bin and p is a parameter to be decided;

- the loss function:
$$L_{loss} = L_{ada} + \alpha L_{flat}, \tag{5.8}$$
where α is a parameter to be set.

L_{loss} is a function of the event scores s_i. The scores depend from the weights w_j, therefore L_{loss} can be treated as a function of w_j. At each stage of the BDT training the w_j are chosen as the values that minimize L_{loss}. Minimizing L_{ada} provides discrimination, while minimizing L_{flat} provides uniformity with respect to $\vec{y}$. In this analysis the dijet invariant mass m_{jj} is the only observable one wants uncorrelated from the BDT output, therefore $\vec{y}$ is a 1-dimensional space.

The two parameters, p and α, have to be optimized in the algorithm. The latter represents the ratio between the discriminating power and uniformity of the classifier, as shown in the L_{loss} definition. The values $p = 1.5$ and $\alpha = 15$ have been decided after some tests, verifying the classifier discriminating power and uniformity with respect to the dijet mass.

Input variables

For the training of the multivariate method, the discriminating input variables for the uGB training have to be decided. The Higgs signal is so diluted in the data sample that a small fraction (0.4%) of the dataset is extracted and used as a representative of the QCD background. The Monte Carlo $H \to b\bar{b}$ simulated sample is used as the signal sample.

By comparing the background and the signal distributions, 6 input variables have been chosen. Variables too correlated to the invariant dijet mass, such as the jet p_{T}, have been discarded. The input variables are:

- the invariant mass of the leading and the sub-leading jets which peak at higher values for the signal with respect to the background;
- the $\Delta\eta$ and the $\Delta\phi$ between the two jets. Especially the $\Delta\eta$ gives a good discrimination, in fact jets from resonances tend to be more collimated;
- the jet asymmetry $\mathcal{A}$, as defined in Section 3.3.3;
- the helicity angle, defined as the angle between the dijet momentum in the laboratory frame and the momentum of the jet with the highest p_{T} in the couple in the dijet rest frame. It is connected to the spin of the resonance.

Figure 5.6 shows the distributions of these observables for the background (data) and the signal.

The uGB output response

The training is performed using about 25k events for the signal and the background. From these training samples, 20% of the events are randomly selected and used as test samples to check if the uGB suffers of overtraining. The distributions of the classifier for signal and background are shown in figure 5.7, superimposed with the distributions of the test samples. Since the training and test distributions are compatible within the statistical errors, it is possible to exclude a significant overtraining.

The figure of merit that leads our choice on where to set the cut on the uGB response is the signal significance, which is defined as:
$$S(x) = \frac{N^H_{exp}(uGB > x)}{\sqrt{N_{data}(uGB > x)}}, \tag{5.9}$$

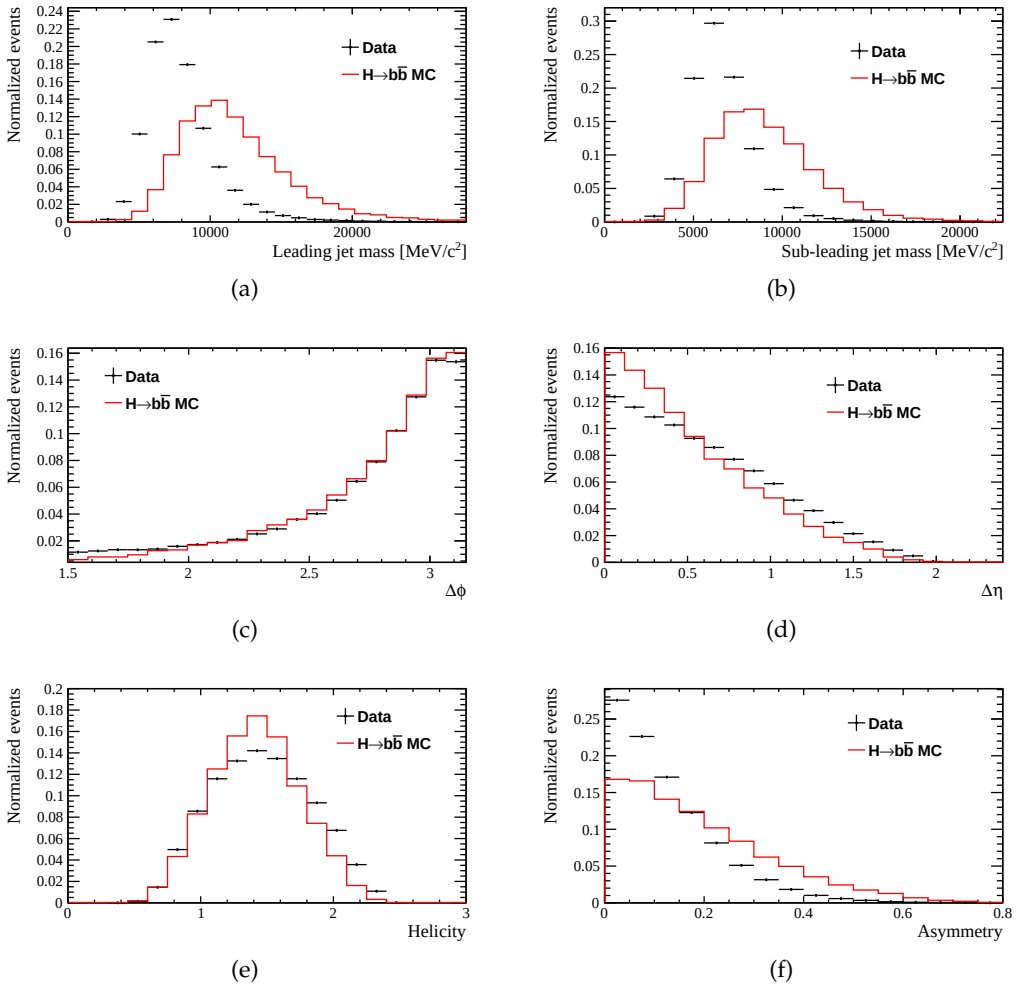


FIGURE 5.6: Distribution of variables used to discriminate the signal from the background, for data and signal Monte Carlo events.

where $N^{H}_{exp}(uGB > x)$ is the number of expected Higgs events that pass the $uGB > x$ cut and $N_{data}(uGB > x)$ is the number of selected data events for $uGB > x$ in the signal invariant mass region. Figure 5.8 shows the signal significance as a function of the uGB response. The x value that maximizes S is chosen to define the signal region ($uGB > x_s$ with $x_s = 0.480$), where the limit is computed. The expected yields of the $H \to b\bar{b}$ and $Z \to b\bar{b}$ in the signal region ($uGB > x_s$ and $80 < M_{12} < 150$ GeV/c^2) are reported in Table 5.3, as well as the number of events in data.

TABLE 5.3: Expected yields in the signal region.

Process	Signal region exp. yield
$H \to b\bar{b}$	126
$Z \to b\bar{b}$	109987
Data	3515906

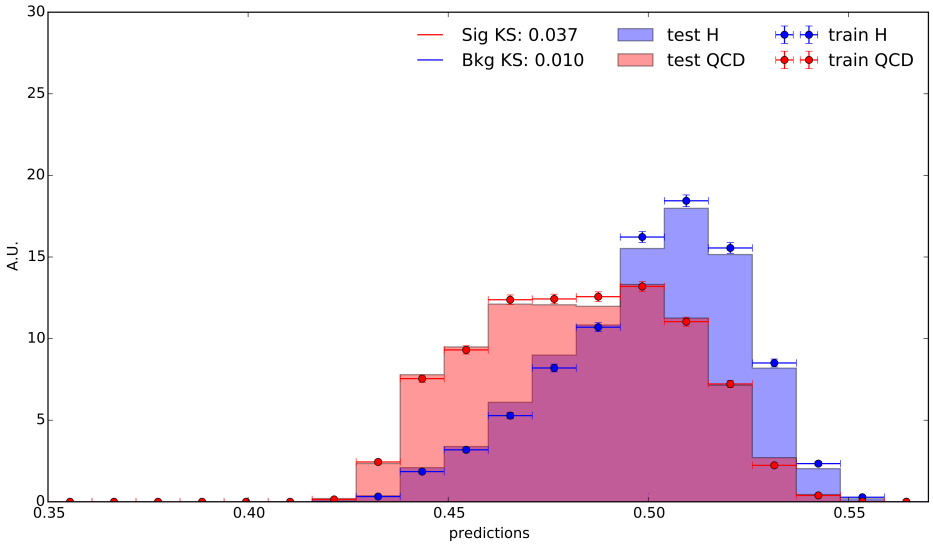


FIGURE 5.7: Distributions of the classifier obtained from the uGB traning, for signal and background. Distributions of the test samples are superimposed.

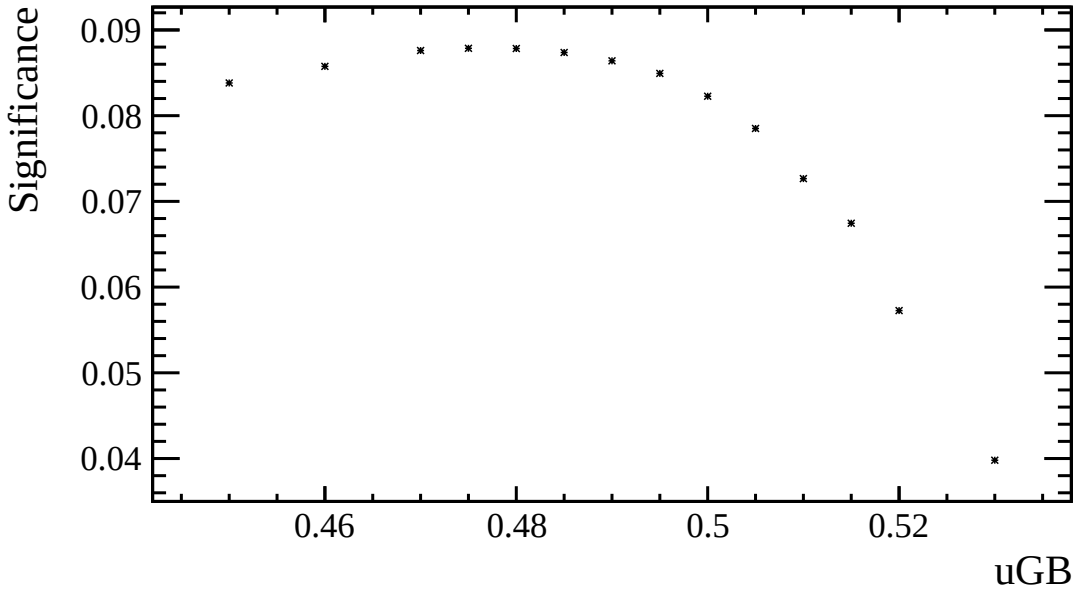


FIGURE 5.8: The signal significance as a function of the uGB response.

5.2.5 Upper limit on $\sigma(pp \to H)\mathcal{B}(H \to b\bar{b})$

The most discriminating variable, and thus the one that is used for the limit computation, is the invariant mass of the dijet system. The use of the uGB method guarantees that the multivariate technique-based events selection does not sculpt the invariant mass background shape.

In this preliminary study, the background description in the signal region is obtained through an analytic model whose parameters are found by fitting the dijet mass distribution of data events in the [60,80] GeV and [150,180] GeV sidebands.

This model is composed by a combinatorial QCD $p.d.f.$ plus a $Z \to b\bar{b}$ $p.d.f.$. Other backgrounds are considered negligible.

Background model

The combinatorial background mass shape is parametrized with a function which is the product of a modified exponential with an acceptance function:

$$P_{QCD} = \frac{b_1[(m_{jj} - b_2)]^{b_3}}{1 + b_1[(m_{jj} - b_2)]^{b_3}} exp[-b_4(m_{jj})^{b_5})] \quad (5.10)$$

where b_1, b_2, b_3, b_4, b_5 are free parameters in the fit and m_{jj} is dijet invariant mass. To validate this model for the QCD background shape also in the signal region, the invariant mass distribution of the Monte Carlo simulated QCD $b\bar{b}$ sample is fitted. The $RooFit$ package of ROOT [129] is used Figure 5.9 shows the result of the fit to the QCD Monte Carlo sample using as model the $p.d.f$ defined by 5.10. The $\chi^2/n.d.f.$ is 2.3.

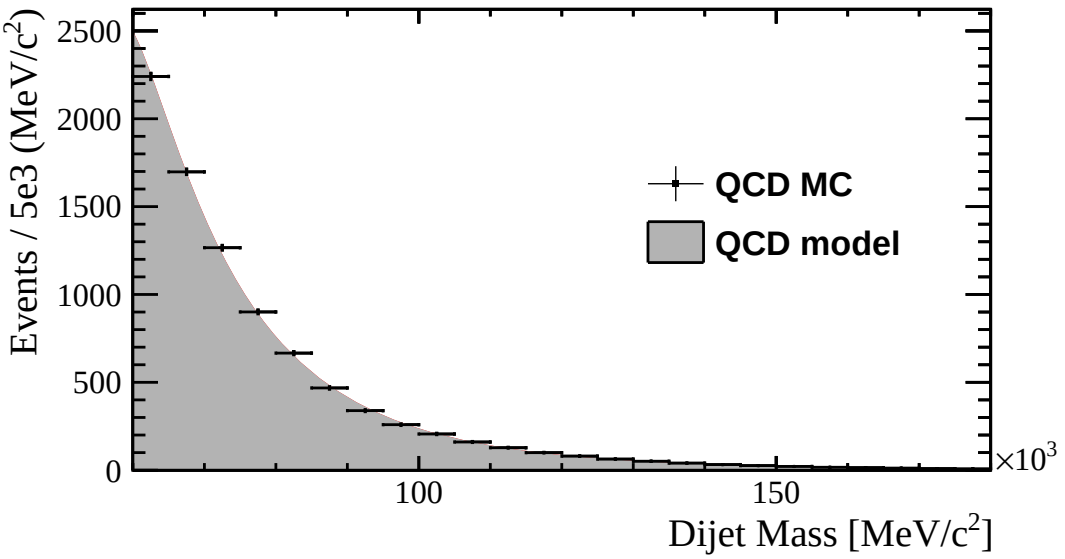


FIGURE 5.9: Fit to the dijet invariant mass of $b\bar{b}$ QCD simulated events using the QCD parametrization model.

The $Z \to b\bar{b}$ dijet invariant mass distribution is modeled using the simulated sample. The distribution is asymmetric with a tail at high values of mass. A triple gaussian model is used to describe it:

$$Z(m_{jj}) = [f_1 G(m_{jj}; \mu_1, \sigma_1) + f_2 G(m_{jj}; \mu_2, \sigma_2) + (1 - f_1 - f_2) G(m_{jj}; \mu_3, \sigma_3)], \quad (5.11)$$

where $G(m_{jj}; \mu_i, \sigma_i)$ are gaussian distributions with mean μ_i and width σ_i, f_i are their relative fractions. The parameters of the model are found with a fit, which is performed using the $RooFit$ package of ROOT [129] with the unbinned maximum likelihood technique. In the fit N_Z, f_i, μ_i and σ_i are free parameters. MC events are weighted to take into account of the differences with data in the b-tagging efficiencies. The dijet invariant mass distribution of Monte Carlo events is shown in Figure 5.10 superimposed with the fit result. The values returned by the fit for the parameters and their statistical uncertainties are presented in Table 5.4. The reconstructed invariant mass of dijets simulated events has a mean of 80 GeV/c^2, below

the known Z-boson mass. The reduced mean is due to parton radiation outside the jet cone, missing energy, and residual biases in the reconstructed jet energy that are not recovered by the jet energy correction. The same applied for the Higgs boson.

TABLE 5.4: Value returned by the fit for the parameters and statistical uncertainties obtained for $Z \to b\bar{b}$ dijet invariant mass.

Parameter	Fitted Value	Stat. Uncertainty
f_1	0.17	0.02
μ_1	64 GeV/c^2	1 GeV/c^2
σ_1	8.3 GeV/c^2	0.5 GeV/c^2
f_2	0.81	0.03
μ_2	83 GeV/c^2	1 GeV/c^2
σ_2	11.8 GeV/c^2	0.4 GeV/c^2
μ_3	112 GeV/c^2	5 GeV/c^2
σ_3	25 GeV/c^2	2 GeV/c^2

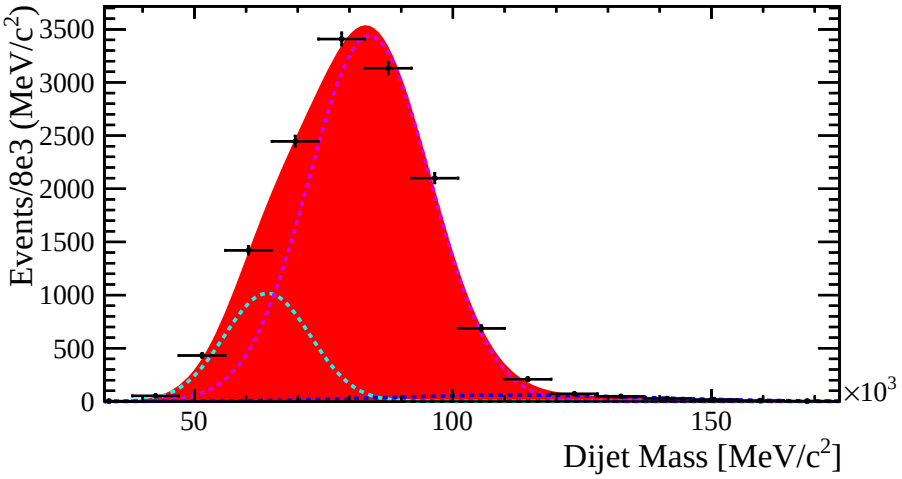


FIGURE 5.10: Fit to the $Z \to b\bar{b}$ simulated events. Three gaussians are used to model the Z shape.

The free parameters of the final background model (QCD+Z) are the QCD background coefficients and the QCD yield. These values are found with the fit to the [60,80] and [150,180] GeV/c^2 data sidebands. The result is shown in Figure 5.11, where the extrapolation in the [80,150] GeV/c^2 region is also plotted. The $\chi^2/n.d.f.$ is 1.8 and the QCD background parameters obtained from the fit are reported in Table 5.5.

Signal model

As for the $Z \to b\bar{b}$ background shape, the signal invariant mass model is obtained from a fit to the $gg \to H \to b\bar{b}$ Monte Carlo sample. Simulated events are weighted for data/Monte Carlo b-tagging scale factors. The signal model is a triple gaussian $p.d.f.$, the same parametrization defined in 5.11. The fit result is shown in Figure 5.12 and the fitted parameters are reported in Table 5.6.

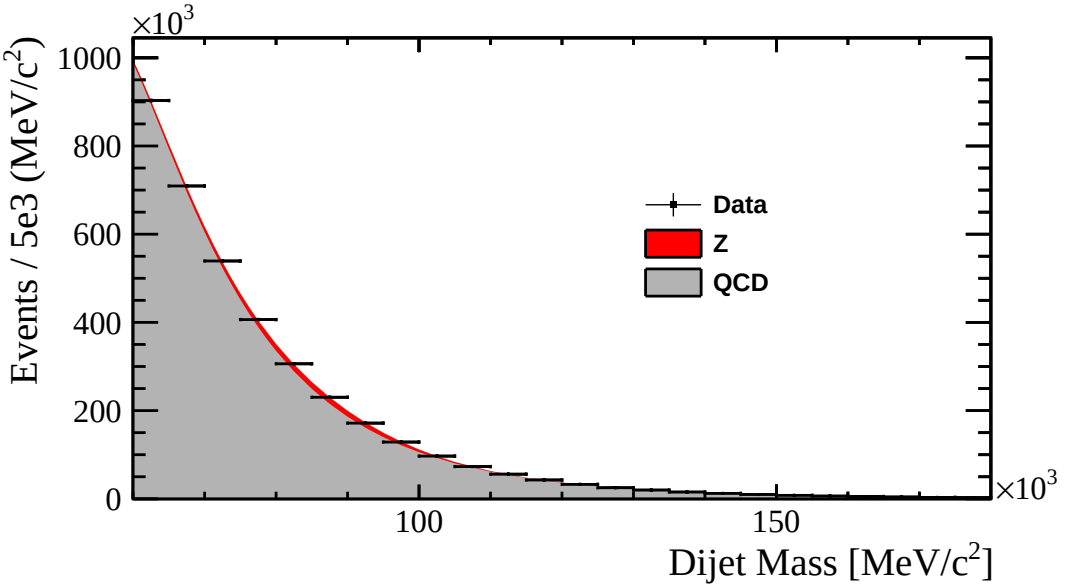


FIGURE 5.11: Data events superimposed with the background distribution obtained from the fit to the [60,80] GeV/c^2 and [150,180] GeV/c^2 dijet invariant mass sidebands.

TABLE 5.5: Value returned by the fit to the dijet invariant mass for the parameters and statistical uncertainties of the background model.

Parameter	Fitted Value	Stat. Uncertainty
b_1	3.1×10^{-2}	3.6×10^{-6}
b_2	2.4×10^{-5} MeV/c^2	1.1×10^{-7} MeV/c^2
b_3	7.5×10^{0}	0.4×10^{0}
b_4	1.0×10^{4} (MeV/c^2)$^{-1}$	4.8×10^{0} (MeV/c^2)$^{-1}$
b_5	5.1×10^{-1}	2.8×10^{-5}

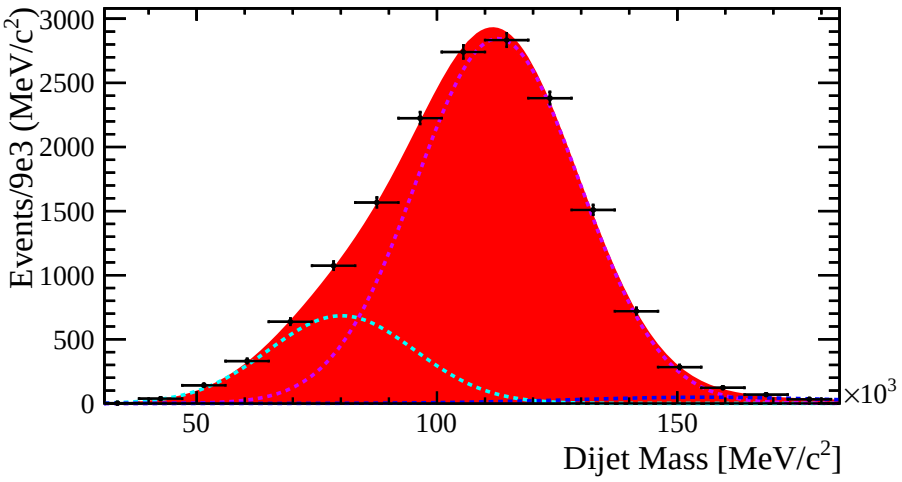


FIGURE 5.12: Fit to the $H \to b\bar{b}$ simulated events. Three gaussians are used to model the H shape.

TABLE 5.6: Value returned by the fit for the parameters and statistical uncertainties obtained for $H \to b\bar{b}$ dijet invariant mass.

Parameter	Fitted Value	Stat. Uncertainty
f_1	0.17	0.02
μ_1	80.6 GeV/c^2	3.6 GeV/c^2
σ_1	15.3 GeV/c^2	1.1 GeV/c^2
f_2	0.81	0.03
μ_2	113 GeV/c^2	1 GeV/c^2
σ_2	17 GeV/c^2	1 GeV/c^2
μ_3	156 GeV/c^2	3 GeV/c^2
σ_3	27 GeV/c^2	22 GeV/c^2

5.2.6 Upper limit for the $H \to b\bar{b}$ process

At this stage everything is set up for the upper limit calculation. A 95% confidence level (C.L.) upper limit on the inclusive production of the Standard Model Higgs is set using a modified frequentist CL_S method. The limit calculator is based on the MCLIMIT packag and it is the same already used in Chapter 4 and 6 to set the limit on the Standard Model Higgs and the Higgs-like particle at CDF II. For more details about the CLs method and the limit calculator see Section 5.1.1.

The measurement of the $Z \to b\bar{b}$ cross section described in Section 4.5 showed that the dominating systematic uncertainty for these kind of searches is the tagging efficiency. It has been introduced in the limit as nuisance parameters. The weights of data/Monte Carlo corrections for b-tagging are varied within their uncertainty listed in Table 3.4 in the H signal and Z background templates determination and in the efficiency estimations (the latter varies the number of expected signal events for a given cross-section). Also the theoretical uncertainty of the Z boson production cross section followed by the $Z \to b\bar{b}$ decay has been taken into account as the yield of the $Z \to b\bar{b}$ background is varied within the theoretical uncertainty, which is of the order of 3% [125]. The systematics related to the signal model are smaller and they have been neglected in this preliminary study.

Finally, the observed and expected CL_S, evaluated with this preliminary analysis procedure, are shown in Figure 5.13 as a function of normalized to the standard model expectation. The observed (expected) upper limit is 15.1 (20.6) times the Standard Model expectation. The observed and expected upper limits are compatible within the 2σ uncertainty bar. Turning these values into an upper limit on the cross section times the branching ratio in the fiducial phase space of the LHCb acceptance gives:

$$\sigma_{fiducial}(pp \to H)\mathcal{B}(H \to b\bar{b}) < 39.0 \text{ pb, at } 95\% \text{ of C.L.} \tag{5.12}$$

5.3 Future prospects

The limit shown above can be considered a first feasibility test for the inclusive $H \to b\bar{b}$ search at LHCb. The luminosity used for this study corresponds to about one half of the total luminosity which is supposed to be collected in the whole Run II. The upper limit scales with the inverse of the signal significance:

$$S = \frac{N_H}{\sqrt{N_{b\bar{b}} + N_H}} \approx \frac{N_H}{\sqrt{N_{b\bar{b}}}}, \tag{5.13}$$

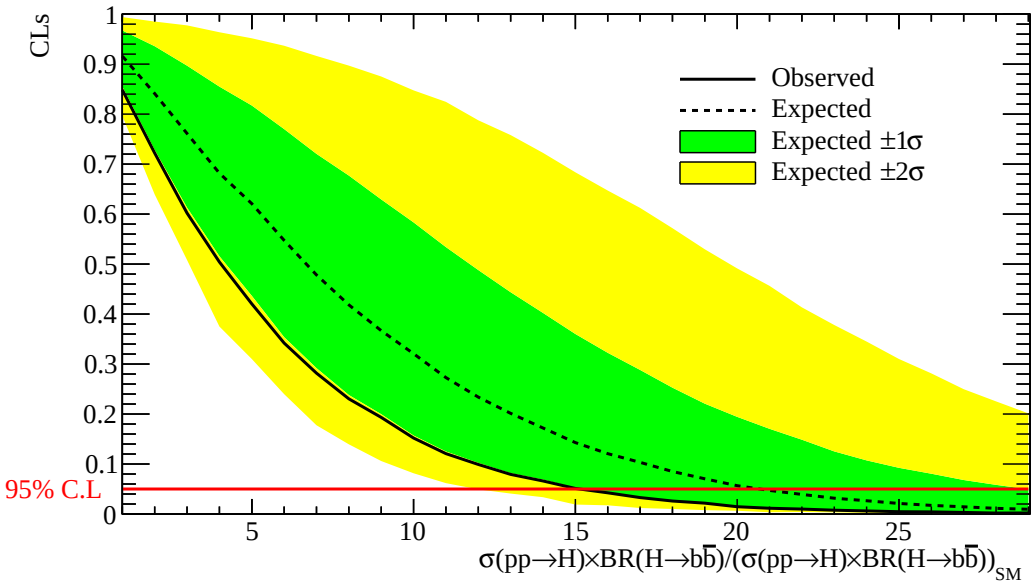


FIGURE 5.13: Observed (black solid line) and expected (black dashed line) CL_S as function of the cross section times the branching ratio normalized to SM $H \to b\bar{b}$. Green and yellow regions are respectively the 1σ and 2σ bands on the expected CL_S.

which means that, without any further improvement, at the end of the Run II the calculated limit should improve of a factor $\sqrt{2}$. The integrated luminosity that LHCb should collect at 13 TeV to reduce the inclusive limit to 1 times the Standard Model cross section, when the $H \to b\bar{b}$ 3σ evidence would be possible, is 365 fb^{-1}, to be compared with the 300 fb^{-1} that are actually foreseen.

Work to better understand the tagging efficiencies is needed, as now these are the sources of the main systematic uncertainty. Furthermore, from 2021, the Run III LHC phase is starting with an increase in the center of mass energy from 13 TeV to 14 TeV. For this new data taking, LHCb will be upgrade, and in particular a new VELO with better performance on the vertexing will be installed [130]. This would lead to even better performance of the heavy flavour jet tagging, improving the overall Higgs significance and opening also the possibility for the first competitive inclusive $H \to c\bar{c}$ limit.

Also from the software and the analysis point of view some improvements can be implemented, which are already being under studying with the Run II data, to increase the signal sensitivity. First, a new jet energy correction is being developed for the Run II jets. This can lead to a better resolution of the signal mass, which is directly proportional to the signal significance.

On the same topic, work is ongoing to develop a multivariate technique-based (BDT) regression which can help to correct the energy measurement of the b-jets. In fact, the b-jets need further corrections than the light jets. The reason is that a third of the b hadrons decay leptonically and semileptonically, and due to the lepton number conservation many undetected neutrinos are produced in these jets, leading to a worse energy reconstruction. This technique was first introduced at CDF [131]. In the regression technique, the correlations of a target observable (the jet p_{T}) with observables related with the kinematics, the structure and the secondary vertex of the jet are used to get a "predicition" function for the target observable. Exploiting simulated events, where the true value of the jet p_{T} is known, this regression can

be trained and then applied to real data for further correction of the b-jet energy and getting a better resolution on the $H \to b\bar{b}$ signal. Figure 5.14 shows the result of the preliminary study of an improved b-jet energy correction with a regression technique at LHCb, where the dijet invariant mass distribution of simulated Higgs events before and after the regression application are displayed.

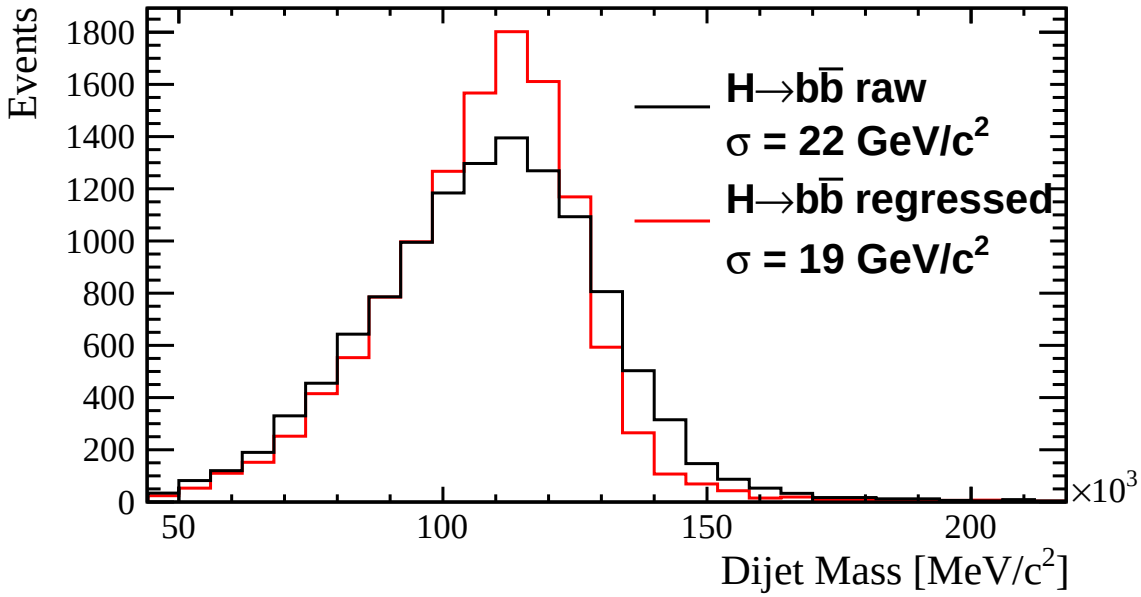


FIGURE 5.14: The dijet invariant mass distribution of simulated $H \to b\bar{b}$ events before and after the application of the regression b-jet energy correction.

Beside the upgraded VELO, the jet tagging performance could profit from new and more powerful algorithms. In fact, studies are ongoing with the aim of introducing deep learning techniques, as it has been done for example for the identification of highly boosted W bosons at CMS [132], which can provide better performance in the jet heavy flavour tagging.

With all these improvements, LHCb can actually be the third experiment to measure the Higgs boson within the end of the LHC era, the first to measure the Higgs in the forward region and probably also the first to measure it in the fully inclusive channel.

Chapter 6

Search for Higgs-like particle in multi-b final state at CDF II

A sample full of b-jets gives an unique opportunity to search for new physics signals, many extensions of the Standard Model in fact predict new particles with a strong coupling to b-quarks. In particular, the discovery of the Higgs boson [1, 2] with measured properties in agreement with the expectations of the Standard Model does not exclude the existence of a new neutral scalar particles ϕ. Higgs-like particles decaying into b-quark jets are foreseen for example in the minimal supersymmetric extension of the Standard Model (MSSM) [76], in two-Higgs-doublet models (2HDM) [77] and dark-matter models involving mediator particles with a large coupling to b quarks [78, 79].

In this chapter, a search for a narrow neutral scalar particle ϕ decaying into b-jets in multi b-jets final states at the CDF II detector is described. To improve the sensitivity, since in the inclusive search the signal would be overwhelmed by the QCD multijet backgrounds, the analysis relies on the case where the ϕ boson is produced in association with one or more b-quarks. Final state with at least three b-quark jets represents a powerful search channel, with the third b-quark jet providing additional suppression of the large multijet background. Because of the various possible theoretical frameworks, the analysis is kept model independent, i.e. no particular theoretical model is tested and the upper limit is set on the production cross section $\sigma(p\bar{p} \to \phi b) \times \mathcal{B}(\phi \to b\bar{b})$.

The characterization and the studies performed on the dataset collected with the DIJET_BTAG trigger in Chapter 4 make this dataset the natural sample where to search for new physics phenomena coupling to b-quarks. The analysis procedure follows the same ideas of the analysis presented in Chapter 4, which has been validated by the measurement of the $Z \to b\bar{b}$ cross section. In particular, thanks to the very low energy threshold at the trigger level, this data sample gives the unique opportunity to search in the low mass region ($100 - 300$ GeV/c^2), covering a complementary phase space with respect to the CMS and ATLAS experiments.

Searches for heavy resonances decaying into $b\bar{b}$ jets, and produced in association with b-quark jets have already been performed by the CDF [71] and D0 [80] experiments at the Tevatron collider and by the CMS [81] experiment at LHC. The two Tevatron experiments have combined their results, reporting a deviation, at the level of 2σ, from the Standard Model expectations in the two b-quark jets invariant mass around $100 - 150$ GeV/c^2 [5]. The result of the limit is shown in Figure 6.1. With the analysis presented in this part chapter of the thesis we can confirm or exclude this excess. The CMS collaboration has excluded a resonance compatible with a Minimal Supersymmetric Standard Model particle and has set an upper limit in the $(\mathrm{M}_A, \tan\beta)$ parameter space.

A future extension of this analysis also at LHCb, once the uncertainties on the trigger and tagging efficiencies will be better under control, can lead to a big improvement to the new Physics sensitivity.

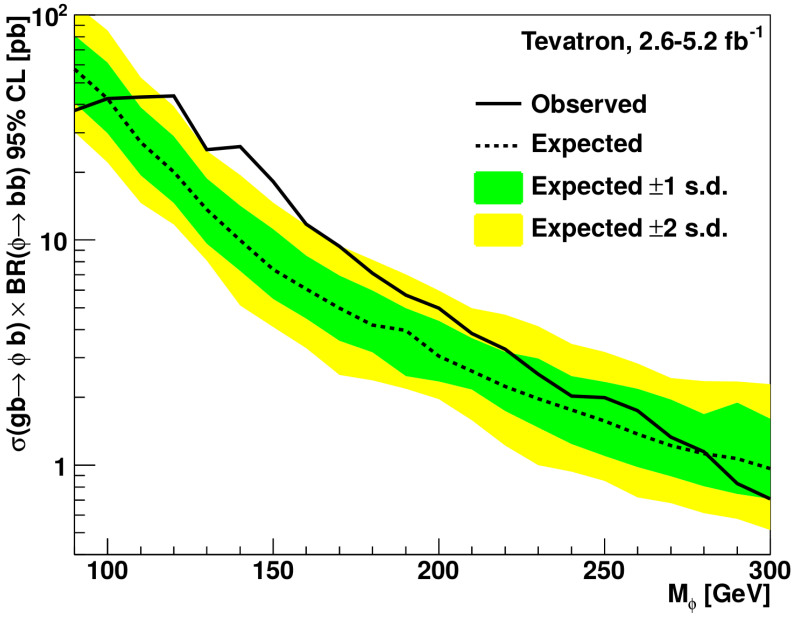


FIGURE 6.1: Model independent 95% C.L. upper limits on the product of cross section and branching ratio for the Tevatron combined analyses.

In Section 6.1 the data selection and the Monte Carlo simulated samples are described, while in Section 6.2 the evaluation of the background is explained. The fitting procedure used to measure signal and background events is reported in Section 6.3 with the systematic uncertainties in Section 6.4. Section 6.5 describes the upper limit on the $\sigma(p\bar{p} \to \phi b) \times \mathcal{B}(\phi \to b\bar{b})$.

6.1 Dataset and Monte Carlo simulation

The signal signature is a bump in the invariant mass distribution of the two leading b-jets in a sample of three b-jets. The event selection follows the one described in Section 4.1.1 used for the measurement of the $Z \to b\bar{b}$ cross section and the upper limit on the $H \to b\bar{b}$ process, but it requires one additional b-tagged jet in the final state.

Thus, the dataset correspond to 5.4 fb^{-1} of data collected with the DIJET_BTAG trigger path. At the off-line level, three central jets ($|\eta| < 1$), reconstructed with the JetClu [62] cone algorithm with the radius parameter $R = 0.7$ and with a L5 corrected transverse energy $E_{\rm T} > 22$ GeV, are required. In order to increase the heavy flavour content of the sample, these three jets have to be selected by the SecVtx [64] b-tagging algorithm (see Section 3.2.1). This sample is referred as triple tag sample.

Also in this chapter a notation to identify jets selected by the on-line and the off-line SecVtx tagging algorithms is useful:

- **b-tag trigger** jet is a jet that fired the DIJET_BTAG trigger and has a tight SecVtx tag;

- **b-tag** jet is a jet that has tight SecVtx tag;

6.1.1 Monte Carlo sample

Monte Carlo simulated samples are used to evaluate the efficiency and the acceptance of the ϕb signal process at different ϕ mass points and to extract the SecVtx tagging algorithm response to the different jet flavor.

TABLE 6.1: The list with the invariant of the Monte Carlo simulated signal samples.

m_ϕ GeV/c^2	Number of generated events
100	850k
120	1.4M
140	1.4M
160	1.4M
180	850k
200	850k
220	850k
240	850k
260	850k
280	850k
300	850k

Signal samples for a variety of Higgs masses are generated using PYTHIA 2.216 [65] MSUB=121 which correspond to the $gg \to b\bar{b}h^0$ process, with a p_{T} cut of 15 GeV/c on a quark which can be either the b or the $\bar{b}$. The mass of the two leading jets in the event, m_{12}, which is used to separate signal from background, is shown in Figure 6.2 for four values of the neutral scalar mass, $m_{12} = 100, 140, 180, 220, 280\,\mathrm{GeV}/c^2$.

Using these signal samples the final selection efficiency is evaluated. It varies from 3.7‰ to 8.7‰ as and it is shown, as a function of the invariant mass of the generated neutral scalar in Figure 6.3.

The detailed description of the QCD $b\bar{b}$, $c\bar{c}$ and light quark Monte Carlo simulated samples, which jets kinematics distributions have been validated on data, see 4.1.2.

6.2 Background modeling

It is possible to obtain more than two heavy quarks in the final state by combining the QCD heavy flavor quarks production described in Section 1.2.2 in a single event. Given the many possible final states with heavy quarks, it is not possible to rely on direct calculation of the multijet production. In addition, the m_{12} spectrum of the background is also affected by biases introduced by the trigger and displaced-vertex tagging requirements that we have to reproduce in our modeling. For these reasons, the background is built using the same data driven technique as the one described in Chapter 4.

Other possible process that can contribute to the background are the $Z + jet$ and the $t\bar{t}$ productions. We can neglect this two process, we expect few events from these sources ($< 1\%$ of the total events) and their contribution are already represented in the double tagged events used to construct the background templates.

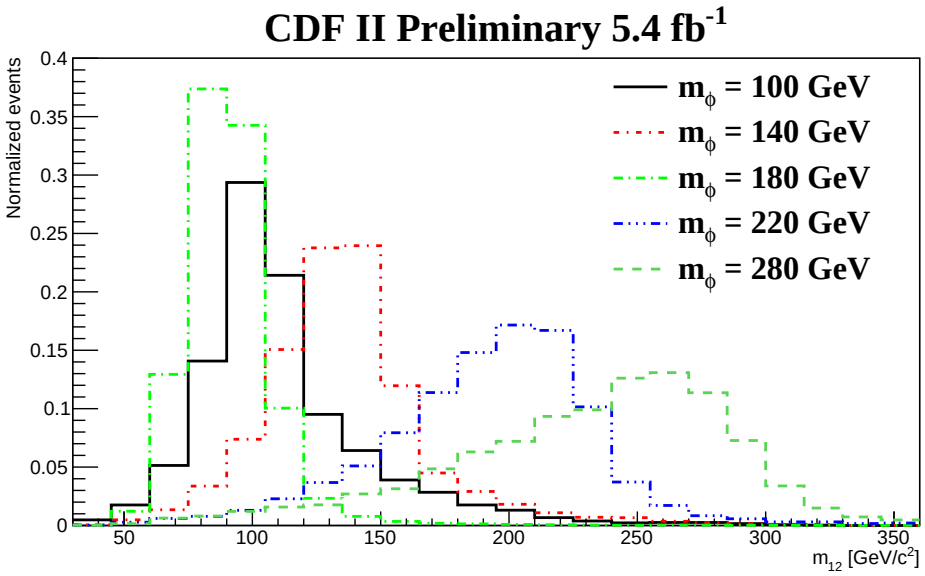


FIGURE 6.2: Invariant mass of the two leading jets in MC signal events for different mass points.

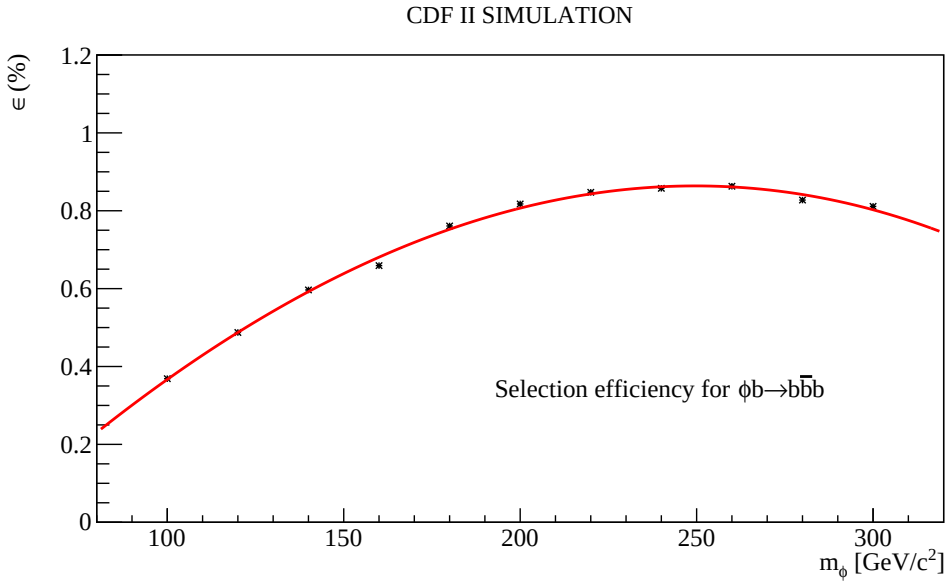


FIGURE 6.3: Signal acceptance as function of the mass of the neutral scalar particle.

6.2.1 Data driven background templates

As in the measurements described in Chapter 4 we search for the signal in the double tag sample by modeling the different multijet backgrounds starting from the single tag sample, in this chapter, where the beyond Standard Model signal is searched in the triple tag sample, the QCD backgrounds templates are built starting from the double tag sample. In this way, kinematical biases due to the trigger are automatically taken into account.

This can be done because, as it can be infer from the double tag sample composition studies explained in the previous chapter and from the search for a Minimal Supersymmetric Standard Model Higgs [71] already performed at CDF II, the triple

tag sample predominantly contains at least two real b-jets. The double tagged sample is then the natural starting point to build the different heavy flavour multijet background templates used to describe the triple tagged data. The effect of requiring a third tag, whose efficiency depends upon the flavor of the jet, can be simulated using a parametrization of the SecVtx response evaluated using Monte Carlo samples, exactly as in Chapter 4. The flavor composition of the triple tagged jets sample is ultimately determined by fitting the data.

The tagging probabilities, used to parametrize the response of the SecVtx algorithm, represent the efficiency to tag a b-, c- and light quark initiated jet as a b-jet as function of its E_{T} and η. The tagging matrices built in Section 4.2.1 are constructed per jet, under the assumption that the probability to tag a jet depends only on its kinematics and not on the event topology. Therefore, the tagging matrices determined for the $Z \to b\bar{b}$ and the $H \to b\bar{b}$ analysis can be used also in this search.

The events in the double tag sample, with an additional third untagged jet, are organized in two categories, bbX and Xbb. These categories depend on the E_{T} rank of the untagged jet which is represented by the upper-case letter X, with the caveat that no distinction is made between the two leading jets. The ranking in descending E_{T} of the three jets is incorporated in the nomenclature adopted here, e.g. Xbb means a sample of events where the third leading jet and either one of the two leading jets is tagged. From these categories six background templates are constructed by weighting the untagged jet with the tagging matrices for the different flavour hypotheses, light quark (Q = $udsg$), charm (C) or beauty (B):

Bbb - bbB - Cbb - bbC - Qbb - bbQ

6.2.2 The x_{tags} variable definition

In a sample with three tagged jets, the fraction of events with at least one light quark initiated jet is higher with respect to the double tag sample. To better discriminate among the multijet production mechanisms thus, a second variable derived from the TagMass is introduced in the fit templates. As already pointed out in Section 4.2.2, the TagMass is sensitive to the flavor of the parton initiating the jet. Light quarks and gluons, which can generate a secondary vertex tag only due to tracks mismeasurements, as well as c quarks initiated jets have Tagmass distribution peaking at lower values with respect to b quark initiated jets. In Figure 4.6 the distributions for jet coming from b, c and light quarks are shown.

The variable x_{tags} allows a better separation between backgrounds with high and low TagMass values. We built x_{tags} following the recipe described in Reference [71]. Because no distinction is made between the two leading jets in the flavor classification scheme, x_{tags} is constructed to be symmetric under their interchange, as it is m_{12}.

We define the x_{tags} variable as:

$$x_{\mathrm{tags}} = \begin{cases} min(m_{3,\mathrm{tag}}, 3) & : \quad m_{1,\mathrm{tag}} + m_{2,\mathrm{tag}} < 2 \\ min(m_{3,\mathrm{tag}}, 3) + 3 & : \quad 2 < m_{1,\mathrm{tag}} + m_{2,\mathrm{tag}} < 4 \\ min(m_{3,\mathrm{tag}}, 3) + 6 & : \quad m_{1,\mathrm{tag}} + m_{2,\mathrm{tag}} > 4 \end{cases} ,$$

where $m_{1,\mathrm{tag}}$ is the TagMass of the leading jet, $m_{2,\mathrm{tag}}$ is the TagMass of the second leading jet and $m_{3,\mathrm{tag}}$ is the TagMass of the third leading jet.

Table 6.2 summarizes how the x_{tags} variable is built. The $m_{1,\mathrm{tag}} + m_{2,\mathrm{tag}}$ provides the sensitivity to bcb and bqb components versus the other components, the $m_{3,\mathrm{tag}}$ separates out bbc and bbq.

	$m_{tag,1} + m_{tag,2}$ [GeV/c^2]		
$m_{tag,3}$ [GeV/c^2]	$0-2$	$2-4$	$4-\infty$
$0-1$	0	3	6
$1-2$	1	4	7
$2-\infty$	2	5	8

TABLE 6.2: Schematic definition of the x_{tags} variable as it results from the combination of the $m_{tag,1} + m_{tag,2}$ and $m_{tag,3}$ values.

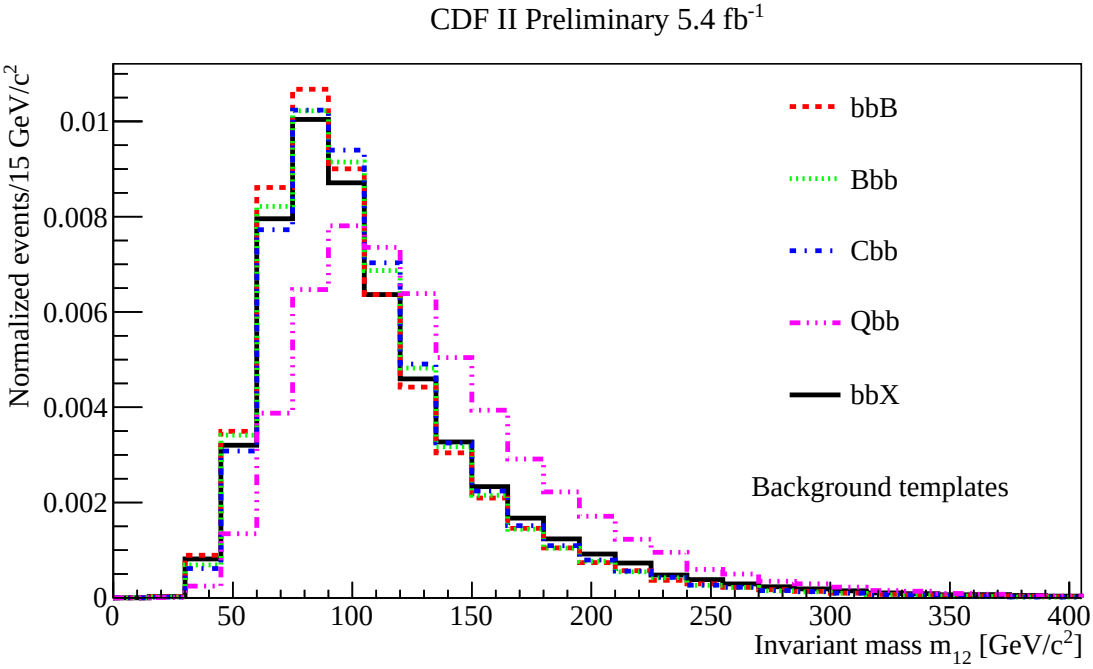


FIGURE 6.4: Invariant mass distribution of the two leading jets m_{12} background templates variables.

In order to compute x_{tags} for the background components we need to simulate not only the bias on the invariant mass m_{12} due to the requirement of the third tag, but also its expected value of TagMas. Therefore the tagging matrices are parametrized also as function of the TagMass in addition to the jet E_{T} and η. Because the untagged jet in double-tagged events does not have any Secondary Vertex to compute x_{tags}, we assign all possible values of the TagMass for that jet, with the caveat of weighting it with the proper weight taken from the tagging matrices. For each weighted TagMass value we then compute x_{tags} and fill the background template histogram. By construction then, each event has multiple entries in the background template, each with the same value of m_{12} but with varying values for x_{tags}.

Figure 6.4 and Figure 6.5 show the m_{12} and x_{tags} distributions of heavy flavor multijet background components respectively. The average of the bbC and bbQ templates (bbX) is used in the fit because they are too similar and the fit cannot discriminate between them. The double tag sample, where the untagged jet is the third in transverse energy, is composed of about 130k events, the Xbb double tagged sample is composed of about 140k events.

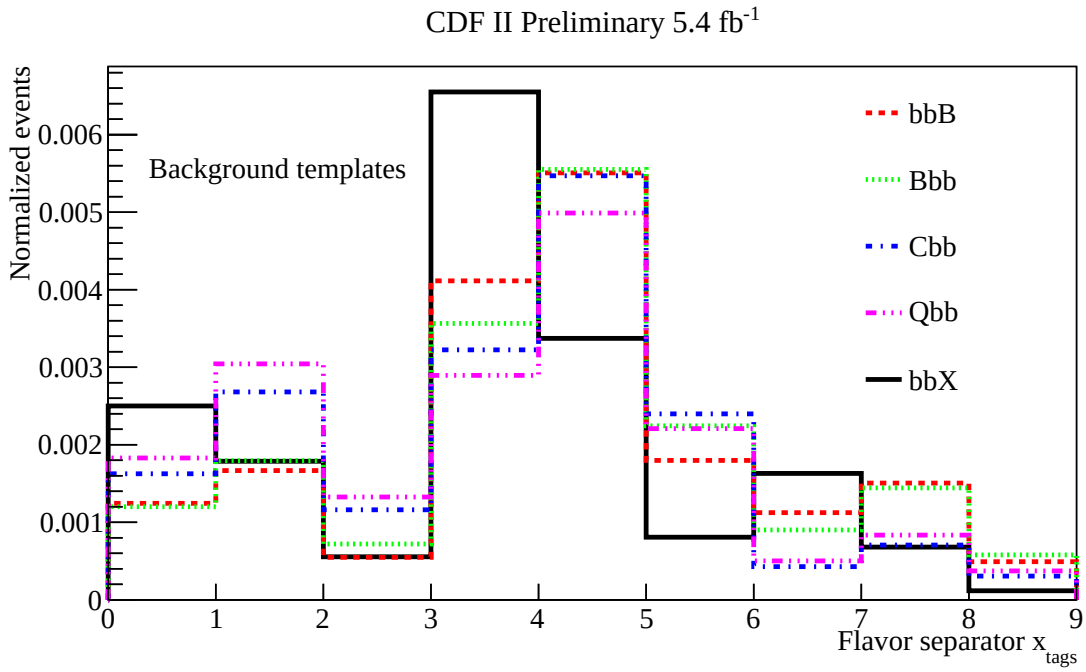


FIGURE 6.5: The flavour discriminator x_{tags} background templates variables.

6.3 Fit description and results

Once we have built the background templates for the two discriminating variables m_{12} and x_{tag}, we have all the ingredients that we need to perform the fit to the triple tag data sample to search for a new scalar particle ϕ. We fit the background and signal templates to the data using a 2-dimensional binned maximum-likelihood fit. The likelihood function is a joint probability of the Poisson likelihood for each bin $\nu_{ij}^{n_{ij}} e^{-\nu_{ij}}/n_{ij}!$, where n_{ij} is the number of observed events in the i-th bin of m_{12} and the j-th bin of x_{tag}, and the expectation in that bin ν_{ij} is given by:

$$\nu_{ij} = \sum_b N_b f_{b,ij} + N_s f_{s,ij}; \tag{6.1}$$

where b represents the five background templates, $f_{b,ij}$ and $f_{s,ij}$ are the bin contents of the various backgrounds and of the neutral scalar signal, and the five N_b and N_s are the free parameters of the fit which represent the normalizations of each component.

6.3.1 Fit to the control sample

Before performing the fit to the triple tag sample, we can test our background modeling exploiting a control data sample. This control sample is the one made of events with two positive and one negative SecVtx tag, where the negative tag can be on any of the three jets. A jet is negatively tagged if its secondary vertex is on the opposite side of the primary vertex from the jet direction. These negative tags are predominantly fake tags from light-flavor jets and are a product of the finite position resolution of the tracking system. This sample then is expected to be almost a pure sample of bbq and qbb events.

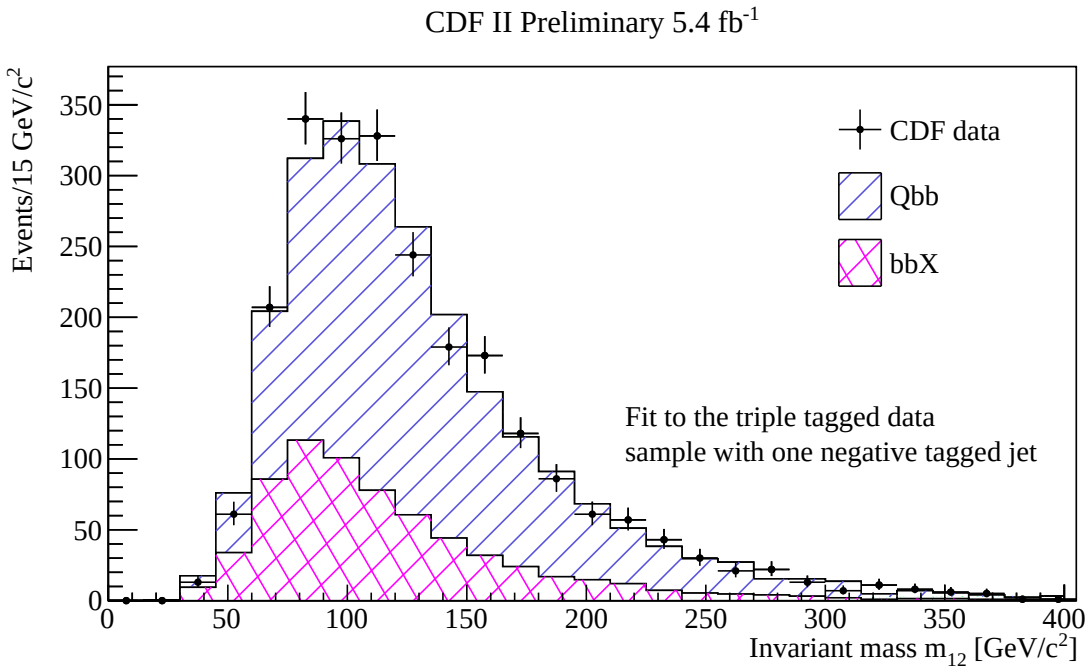


FIGURE 6.6: Invariant mass distribution of the two leading jets in the two positive and one negative tagged jets with the result of the fit projected into the m_{12} variable.

In order to verify the background parametrization, the selected data sample is fitted by using the two dimensional binned maximum-likelihood described previously with m_{12} and x_{tags} as variables. Figure 6.6 shows the result of the fit of the data projected on the m_{12} variable. Data is described by a combination of the bbQ and Qbb templates only, as expected. Table 6.3 reports the result of the fit.

TABLE 6.3: Background yields as returned by the fit to the control data sample.

CDF II Preliminary 5.4 fb^{-1}

Component	Fitted yield in events
Bbb	$< 180\ (1\sigma)$
bbB	$< 205\ (1\sigma)$
Cbb	$< 98\ (1\sigma)$
Qbb	1704 ± 131
bbX	657 ± 191

6.3.2 Signal extraction

Once the background templates are validated, we can focus on the fit to the 5616 events in the triple tag sample, where the signal is searched. First, a fit in the background hypothesis only is performed. Figures 6.7,6.8 show the projections in m_{12} and x_{tag} of the result of the fit. No systematic uncertainties have been included in the fit, which has a goodness-of-fit $\chi^2/d.o.f.$ of 0.8.

The anti-correlation between the Bbb and bbB background is high, -0.973. This leads to big fit uncertainties for these two backgrounds. The limit setting procedure, see Section 6.5, is based on pseudo-experiments rather than relying on the curvature

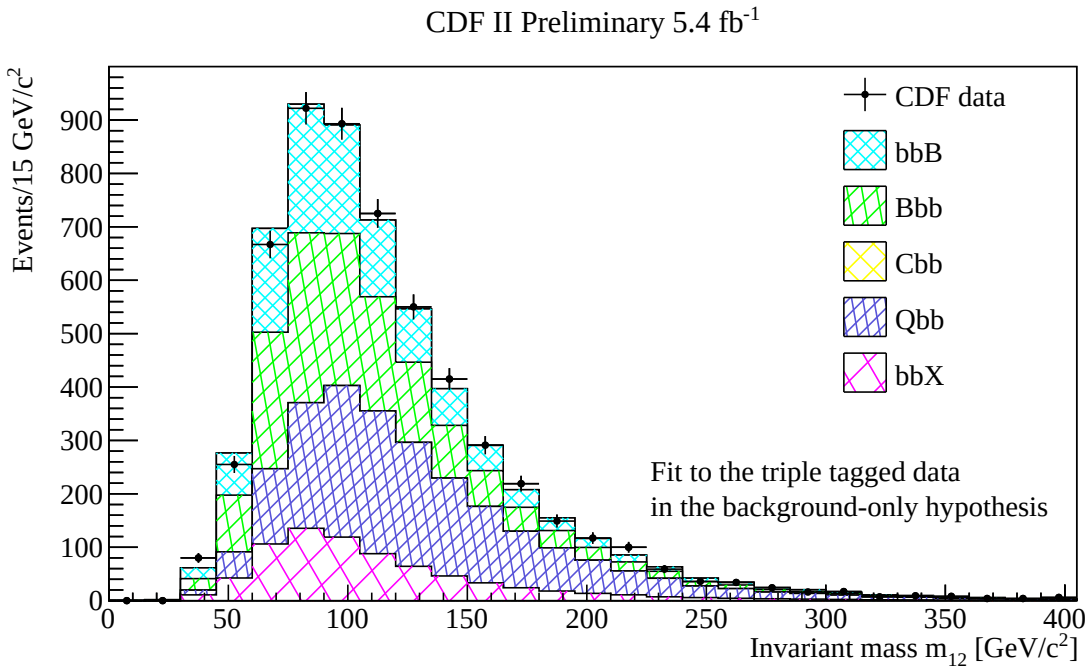


FIGURE 6.7: Triple tagged events fit results projected into the invariant mass m_{12} variable, under the background only hypothesis.

of the likelihood, so as long as the number of Higgs signal events returned by the fit is stable, at the level observed, this does not affect our results.

An a priori estimate of the background components is performed starting from the double tagged sample. The study at Monte Carlo generator-level presented in [71], where the relevant analysis conditions are equal to the ones used here, predicts that in events with two b-jets, the third jet is from b-quark in the 2% of the time and from c-quark in the 4% of the time. The Monte Carlo study demonstrated that these numbers do not depend on the jets energy ordering. Scaling the number of double-tagged events by these values we can predict the number of events with three real b-jets and two b-jets and one c-jet. By normalizing these values by ϵ_f, which is the tag efficiency for a jet under the flavor hypothesis f extracted from the tagging matrices, we can infer the number of bbB, Bbb, Cbb and bbC events.

The Qbb and the bbQ components of the bbX template, are verified by using the fraction of negative tagged jet in the triple tagged sample. Table 6.4 compares the estimated values for the background components with the fitted ones.

The predictions match the results of the fit except for the Cbb, where the expected 550 events seem to be included in the Bbb and bbB fitted components. The c-jets component is absent. This could be due to the cut that the trigger apply on the decay length of the b hadron, which is very inefficient for jets originating from c-quarks.

The search for a neutral scalar Higgs-like resonance, ϕ, is performed in the mass range of $100 - 300$ GeV/c^2. Figures 6.9,6.10 show the result of the fit including a signal with a mass of 160 GeV/c^2. In this case, the goodness-of-fit $\chi^2/d.o.f.$ is 0.75 and 134 ± 70 signal events are found.

In order to correctly evaluate the significance of the fitted signal events, the systematic uncertainties have to be considered. The fitting procedure explained above is applied to a set of pseudo-experiments generated including the systematic uncertainties to test the background only hypothesis and the background plus signal one, see Section 6.5.

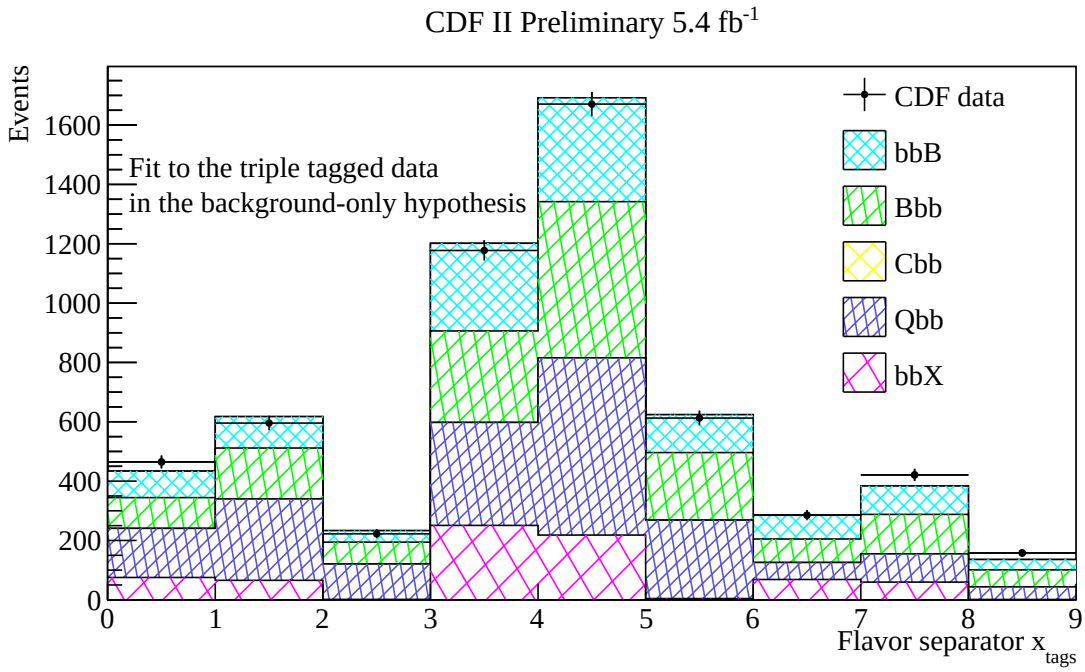


FIGURE 6.8: Triple tagged events fit results projected into the flavor separator x_{tags} variable, under the background only hypothesis.

TABLE 6.4: Events yields as returned by the fit to the triple tagged sample in the background only hypothesis, compared to the Standard Model expectations calculated as explained in the text.

CDF II Preliminary 5.4 fb^{-1}

Background component	Best fit in the background only hypothesis result	Expected events normalizing the double tagged sample
bbB	1227 ± 891	950
Bbb	1672 ± 738	1280
Cbb	$< 90\ (1\sigma)$	550
Qbb	1964 ± 169	1820
bbX	742 ± 293	1080

6.4 Systematics uncertainties

Systematic uncertainties affect both the signal and the background description. They can modify the normalization of the fit results, denoted as rate, and also the m_{12} and x_{tag} distributions of the templates, denoted as shape. Table 6.5 summarizes the systematics uncertainties considered.

Systematics on signal

The systematic related to the on-line and the off-line b-tagging algorithms are taken from the studies performed using the muon $p_{\text{T,rel}}$ method explained in Section 3.2.2. The systematic uncertainty on the signal efficiency due to the CDF jet energy correction is estimated by shifting the energy of the Monte Carlo signal jets by $\pm 1\sigma_c$ of the standard jet energy correction. This is going to affect both the acceptance and the shape of the signal. The acceptance changes from 7% to 4%, depending on the mass

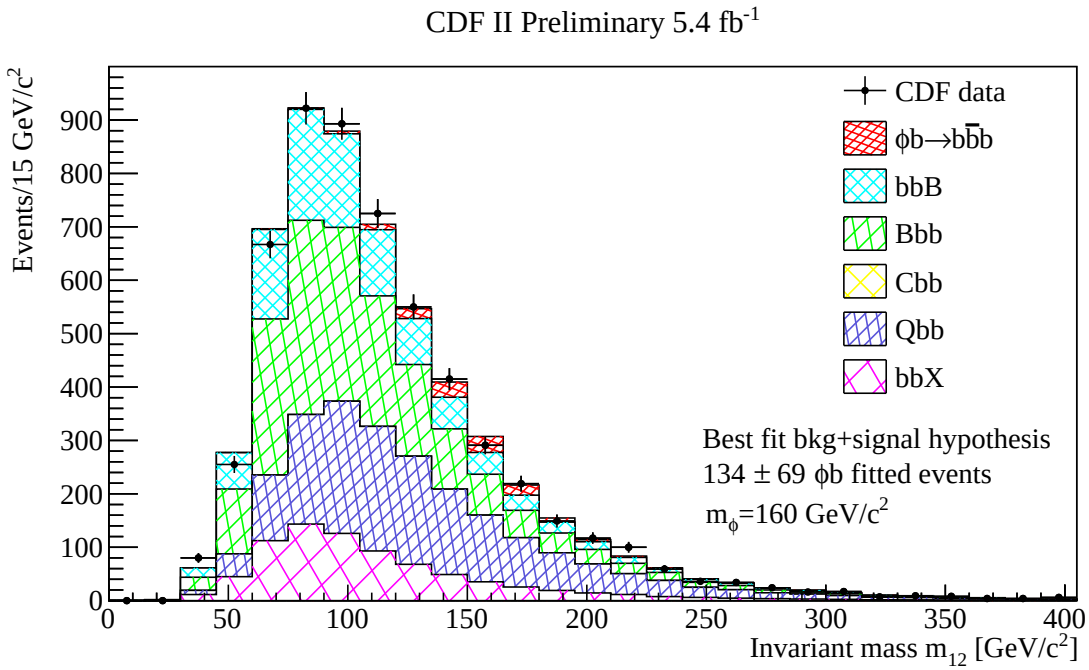


FIGURE 6.9: Result of the fit to the triple tagged data projected into the m_{12} variable. A signal component with a mass of the ϕ of 160 GeV/c^2 is added to the background templates.

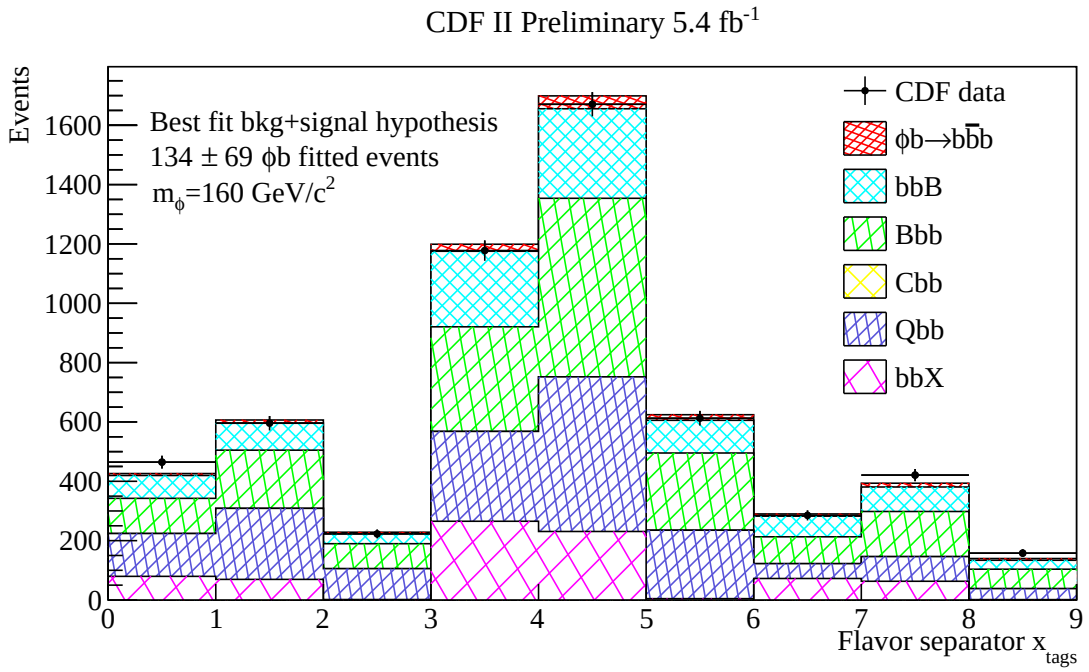


FIGURE 6.10: Result of the fit to the triple tagged data projected into the x_{tags} variable. A signal component with a mass of the ϕ of 160 GeV/c^2 is added to the background templates.

of the ϕ. Figure 6.11 shows the effect of the $\pm 1\sigma_c$ variation on the 200 GeV/c^2 mass point signal template. The uncertainty due to the choice of the Parton Distribution Functions has been evaluated, as in Section 4.4, by generating simaluted samples using the CTEQ6L set and taking the difference in acceptance as systematic.

TABLE 6.5: Summary of systematic uncertainties.

CDF II Preliminary 5.4 fb^{-1}

Source	Variation	Target	Type
Luminosity	5.9%	signal	rate
Offline b-tag	5% per jet	signal	rate
Online and offline b-tag combined	4%	signal	rate
JES	$7-4\%$	signal	rate/shape
x_{tag}	3%	signal	shape
PDFs	2%	signal	rate
Template stat. uncertainty		background	shape
Heavy flavor normalization	5%	background	rate

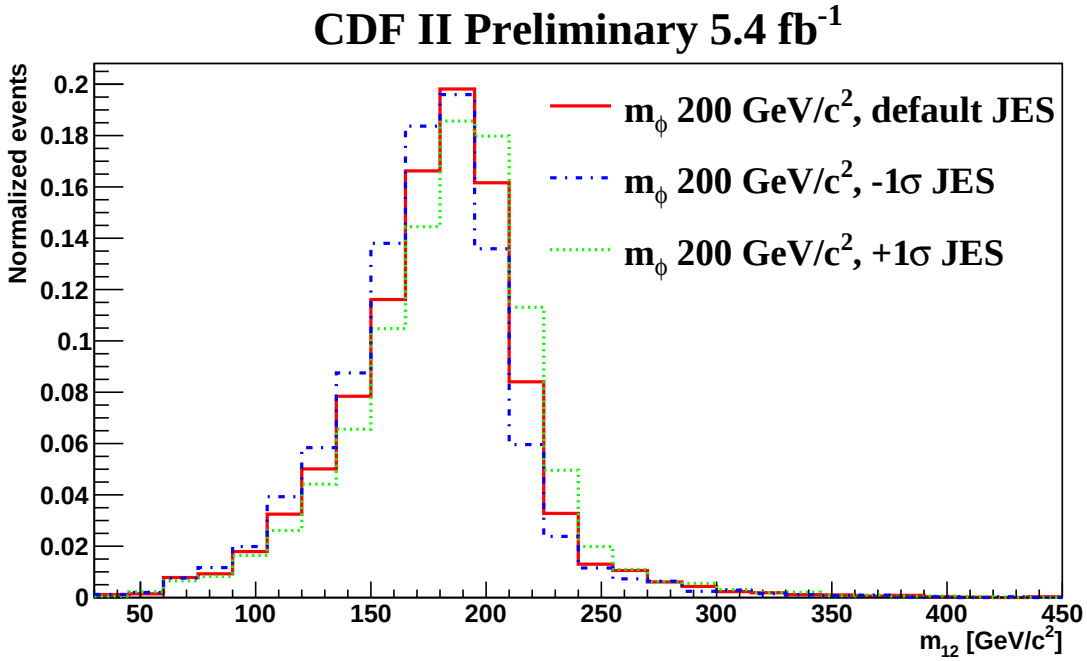


FIGURE 6.11: Effect of the $\pm 1\sigma$ jet energy scale variation on the 200 GeV/c^2 mass point signal template.

Systematics on background

The uncertainty due to the limited statistics of the templates is taken into account in the limit calculator, following a Poisson statistics for each bin. The mass of the SecVtx tags, the TagMass, used to build the x_{tag} variable, has been varied by $\pm 3\%$ following the studies done by the high-p_{T} CDF b-tag group. The starting value of the normalization used for the heavy flavour in the limit calculator depends on the off-line tagging algorithm, a 5% uncertainty is then applied.

6.5 Upper limit on the cross section

In order to compute the significance of the fitted signal events and set the 95% confidence level upper limits on the production cross section times the branching ratio of a narrow scalar as a function of the mass, we use a modified frequentist CL_S

method [74], as we did in Section 5.1.1 to set the limit on the Standard Model Higgs. The limit calculator is based on the MCLIMIT package [75].

Simulated experiments are generated based on the background templates starting with the normalization listed in Table 6.4. The predictions for the numbers of each background type and for the signal are randomly varied for each simulated experiment according to the systematics in Table 6.5. These generated pseudo-experiments are then fitted under the background only hypothesis and in the background plus signal hypothesis.

The CL_s are evaluated for each mass point between 100 and 300 GeV/c^2, and for each mass point the number of signal events corresponding to the 95% confidence level is found. This number of events is then translated into $\sigma \times \mathcal{B}$ using the signal acceptance showed in Figure 6.3, the integrated luminosity and the data/Monte Carlo scale factors for the on-line and off-line b-tagging algorithm. The resulting observed limits and the median expected 95% C.L. limits as a function of the mass of the scalar particle are shown in Table 6.6 and in Figures 6.12,6.13. Figures 6.12,6.13 also show the $\pm 1\sigma$ and $\pm 2\sigma$ bands of the expected limits.

All points of the observed limit are within 2σ band of the expected limit, indicating that there is not statistically significant excess. The result improves the previous combined limit of CDF and D0 and classifies the 2σ excess in the 100-160 GeV/c^2 mass range as statistical fluctuation.

CDF II Preliminary 5.4 fb^{-1}

	95% C.L. upper limit on $\sigma \times \mathcal{B}r$ [pb]	
m_ϕ [GeV/c^2]	Expected	Observed
100	15.2	15.9
120	10.3	12.1
140	6.9	9.3
160	5.3	7.7
180	4.1	5.4
200	3.3	4.4
220	2.8	3.7
240	2.4	2.8
260	2.2	2.1
280	2.0	1.8
300	1.9	1.6

TABLE 6.6: Median expected and observed limits on $\sigma \times \mathcal{B}r$, in pb.

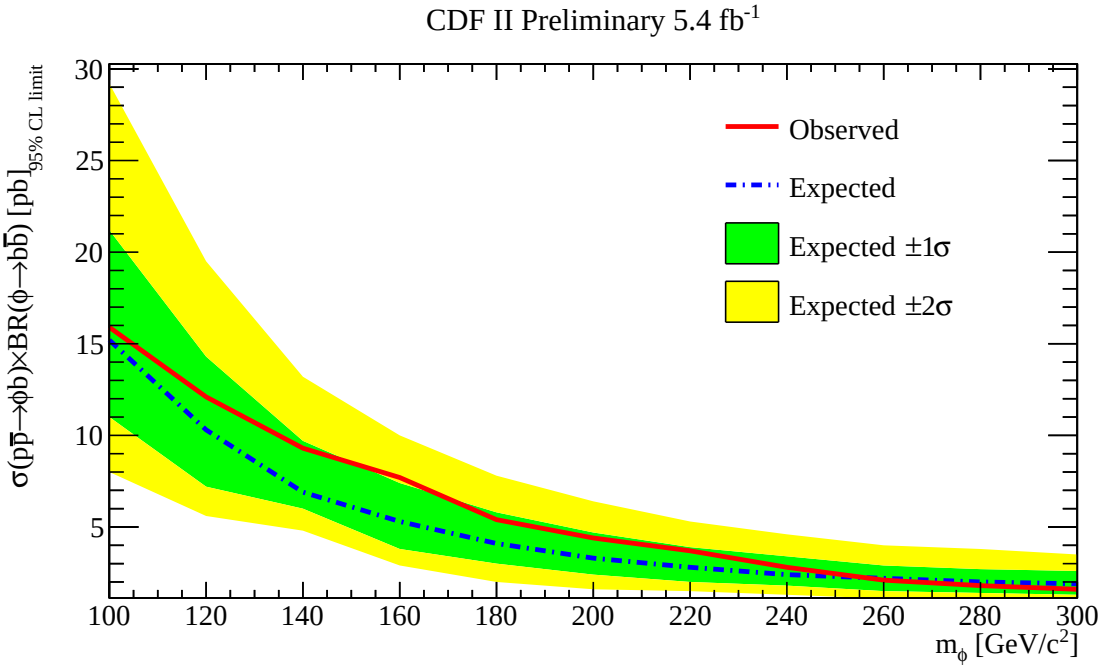


FIGURE 6.12: The observed 95% C.L. upper limits on the cross section times the branching ratio.

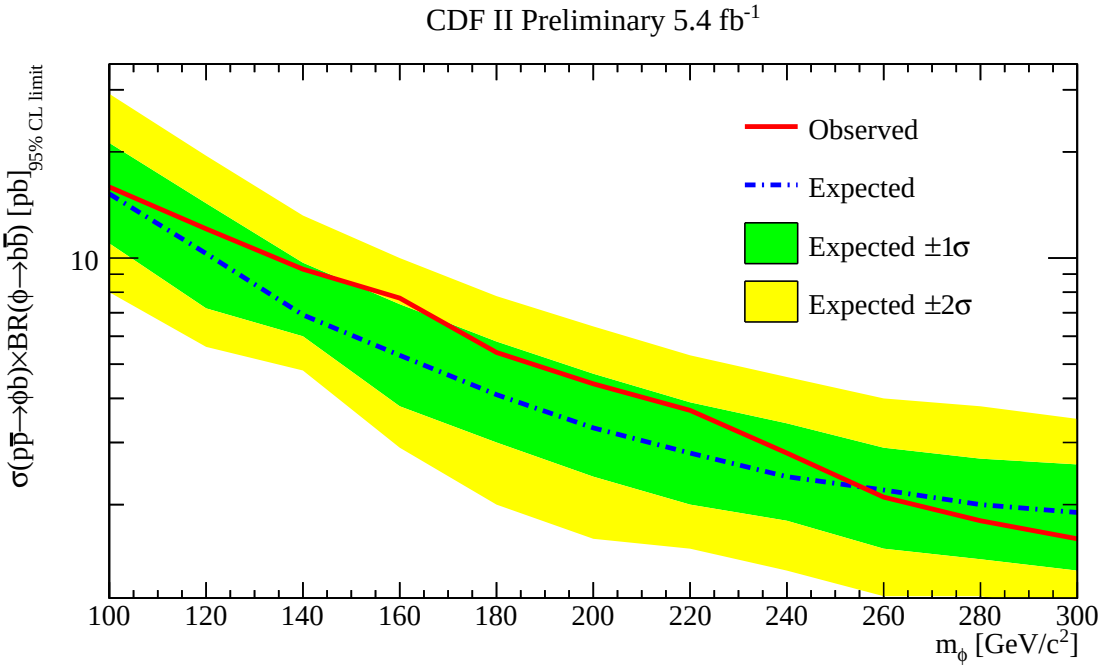


FIGURE 6.13: The observed 95% C.L. upper limits, log scale, on the cross section times the branching ratio.

Conclusion

During these first years of LHC operations remarkable results have been accomplished, one for all the discovery of the Higgs boson. At the general purpose detectors, the searches for Physics beyond the Standard Model with b-jets in the final state are now pushing the limit to new mass values never reached before. But, at low masses, these searches are limited by the possibility to trigger on low p_{T} jets.

This thesis has reported a complementary analysis of data coming from two different particle detectors, the CDF II at Tevatron and the LHCb at LHC, for the search of resonances decaying into b-jets, and in particular of the Higgs boson, in the low mass region. It has been highlighted how the possibility to reconstructed secondary already at the trigger level allow to collect a dataset pure in heavy flavour jets keeping the jet energy thresholds low.

First, we analyzed 5.4 fb^{-1} of data collected at CDF II. The analysis technique has been validated with the measurement of the $Z \to b\bar{b}$ cross section and of the jet energy scale. The measured cross section of $p\bar{p}$ collisions at 1.96 TeV is found to be:

$$\sigma(p\bar{p} \to Z) \times \mathcal{B}(Z \to b\bar{b}) = 1.11 \pm 0.08(stat) \pm 0.13(sys) \text{ nb}, \tag{6.2}$$

in good agreement with the NLO theoretical prediction. The measured jet energy scale factor then, which has to be applied to the jet four-momentum in simulation to match the real distribution, is:

$$\text{JES} = 0.993 \pm 0.022 \pm 0.008. \tag{6.3}$$

The same event selection and background modeling has been used to set the limit on the inclusive $H \to b\bar{b}$ process, which is limited by the poor signal over background ratio. The observed upper limit at 95% C.L. is:

$$\sigma(p\bar{p} \to H) \times \mathcal{B}(H \to b\bar{b}) < 40.6 \text{ pb} \tag{6.4}$$

which corresponds to 33 times the Standard Model prediction.

Adding a third b-jet to the requirement, the same dataset has been exploited to searched for new Higgs-like scalar particles decaying into a $b\bar{b}$ pair and produced in association with another b quark. This search can be included in different theoretical scenario and it is of particular interest because of a previous Tevatron combined result which reports a 2σ excess in the same channel. An upper limit in the $100 - 300$ GeV/c^2 invariant mass range has been set, classifying the 2σ excess as statistical fluctuation.

At LHC, the best experiment for this kind of search in the low mass range is LHCb. Even though it has been designed for flavour physics measurements, it can now be considered a general purpose forward detector. The leveled luminosity provided by LHC and the great performance of the tracking system allow the reconstruction and the identification of low p_{T} heavy flavour jets. This feature has been exploited in this thesis to perform a sensitivity study on the inclusive $H \to b\bar{b}$ process using 1.6 fb^{-1} of data collected in 2016, which leads to an observed 95% C.L. upper limit in the LHCb fiducial phase space of:

$$\sigma(p\bar{p} \to H) \times \mathcal{B}(H \to b\bar{b}) < 39.0 \text{ pb}, \tag{6.5}$$

corresponding to 15 times the Standard Model prediction.

In this thesis it has been shown how the technologies developed at the Tevatron, which data can still be used today to cover a phase space complementary to the ATLAS and CMS experiments, if re-optimized and improved for the new and more challenging LHC environment can give access to decay channels that were though too difficult to explore. In particular, all the knowledge that was developed for the reconstruction and the identification of b-jets can help the future searches for resonances at LHCb. The limit set on the inclusive $H \to b\bar{b}$ process using 2016 data shows that, with the expected upgrade of the detector and the trigger system together with the improvements on the analysis side and tagging algorithms, LHCb can be the third experiment to measure the Higgs boson and the first in a totally inclusive mode.

Bibliography

[1] Serguei Chatrchyan et al. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Phys. Lett.* B716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].

[2] Georges Aad et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Phys. Lett.* B716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].

[3] Pran Nath et al. "The Hunt for New Physics at the Large Hadron Collider". In: *Nucl. Phys. Proc. Suppl.* 200-202 (2010), pp. 185–417. DOI: 10.1016/j.nuclphysbps.2010.03.001. arXiv: 1001.2693 [hep-ph].

[4] *Inclusive search for the standard model Higgs boson produced in pp collisions at* $\sqrt{s} = 13$ TeV *using* $H \to b\bar{b}$ *decays*. Tech. rep. CMS-PAS-HIG-17-010. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2266164.

[5] T. Aaltonen et al. "Search for Neutral Higgs Bosons in Events with Multiple Bottom Quarks at the Tevatron". In: *Phys. Rev.* D86 (2012), p. 091101. DOI: 10.1103/PhysRevD.86.091101. arXiv: 1207.2757 [hep-ex].

[6] E. Noether. "Der Endlichkeitssatz der Invarianten endlicher Gruppen." In: *Mathematische Annalen* 77 (1916), pp. 89–92. URL: http://eudml.org/doc/158716.

[7] Abdus Salam and John Clive Ward. "On a gauge theory of elementary interactions". In: *Nuovo Cim.* 19 (1961), pp. 165–170. DOI: 10.1007/BF02812723.

[8] S. L. Glashow. "Partial Symmetries of Weak Interactions". In: *Nucl. Phys.* 22 (1961), pp. 579–588. DOI: 10.1016/0029-5582(61)90469-2.

[9] Abdus Salam. "Weak and Electromagnetic Interactions". In: *Conf. Proc.* C680519 (1968), pp. 367–377.

[10] Steven Weinberg. "A Model of Leptons". In: *Phys. Rev. Lett.* 19 (1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264.

[11] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Phys. Rev. Lett.* 13 (9 Aug. 1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.321.

[12] J. Beringer et al. "The review of particle physics". In: *Phys. Rev. D* 86 (2012), p. 010001.

[13] Kenneth G. Wilson. "Renormalization Group and Strong Interactions". In: *Phys. Rev. D* 3 (8 Apr. 1971), pp. 1818–1846. DOI: 10.1103/PhysRevD.3.1818. URL: https://link.aps.org/doi/10.1103/PhysRevD.3.1818.

[14] J. Wess and B. Zumino. "A Lagrangian Model Invariant Under Supergauge Transformations". In: *Phys. Lett.* 49B (1974), p. 52. DOI: 10.1016/0370-2693(74)90578-4.

[15] Hans Peter Nilles. "The supersymmetric standard model". In: *Lect. Notes Phys.* 405 (1992), pp. 1–46. DOI: 10.1007/3-540-55761-X_25.

[16] John F. Gunion et al. "The Higgs Hunter's Guide". In: *Front. Phys.* 80 (2000), pp. 1–404.

[17] Ian J R Aitchison and Anthony J G Hey. *Gauge theories in particle physics: a practical introduction; 4th ed.* Boca Raton, FL: CRC Press, 2013. URL: https://cds.cern.ch/record/1507184.

[18] S. Durr et al. "Ab-Initio Determination of Light Hadron Masses". In: *Science* 322 (2008), pp. 1224–1227. DOI: 10.1126/science.1163233. arXiv: 0906.3599 [hep-lat].

[19] R. Keith Ellis, W. James Stirling, and B. R. Webber. "QCD and collider physics". In: *Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol.* 8 (1996), pp. 1–435.

[20] T. Gleisberg et al. "Event generation with SHERPA 1.1". In: *JHEP* 02 (2009), p. 007. DOI: 10.1088/1126-6708/2009/02/007. arXiv: 0811.4622 [hep-ph].

[21] G. Arnison et al. "Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at s**(1/2) = 540-GeV". In: *Phys. Lett.* 122B (1983). [,611(1983)], pp. 103–116. DOI: 10.1016/0370-2693(83)91177-2.

[22] Benjamín Grinstein and Christopher W. Murphy. "Bottom-Quark Forward-Backward Asymmetry in the Standard Model and Beyond". In: *Phys. Rev. Lett.* 111 (2013). [Erratum: Phys. Rev. Lett.112,no.23,239901(2014)], p. 062003. DOI: 10.1103/PhysRevLett.112.239901,10.1103/PhysRevLett.111.062003. arXiv: 1302.6995 [hep-ph].

[23] Marco Bonvini et al. "On the Higgs cross section at N^3LO+N^3LL and its uncertainty". In: *JHEP* 08 (2016), p. 105. DOI: 10.1007/JHEP08(2016)105. arXiv: 1603.08000 [hep-ph].

[24] Georges Aad et al. "Measurement of Higgs boson production in the diphoton decay channel in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector". In: *Phys. Rev.* D90.11 (2014), p. 112015. DOI: 10.1103/PhysRevD.90.112015. arXiv: 1408.7084 [hep-ex].

[25] Vardan Khachatryan et al. "Observation of the diphoton decay of the Higgs boson and measurement of its properties". In: *Eur. Phys. J.* C74.10 (2014), p. 3076. DOI: 10.1140/epjc/s10052-014-3076-z. arXiv: 1407.0558 [hep-ex].

[26] Serguei Chatrchyan et al. "Measurement of the properties of a Higgs boson in the four-lepton final state". In: *Phys. Rev.* D89.9 (2014), p. 092007. DOI: 10.1103/PhysRevD.89.092007. arXiv: 1312.5353 [hep-ex].

[27] Georges Aad et al. "Measurements of Higgs boson production and couplings in the four-lepton channel in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector". In: *Phys. Rev.* D91.1 (2015), p. 012006. DOI: 10.1103/PhysRevD.91.012006. arXiv: 1408.5191 [hep-ex].

[28] Georges Aad et al. "Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments". In: *Phys. Rev. Lett.* 114 (2015), p. 191803. DOI: 10.1103/PhysRevLett.114.191803. arXiv: 1503.07589 [hep-ex].

[29] *Measurements of Higgs boson properties in the diphoton decay channel with 36.1 fb^{1} pp collision data at the center-of-mass energy of 13 TeV with the ATLAS detector*. Tech. rep. ATLAS-CONF-2017-045. Geneva: CERN, July 2017. URL: http://cds.cern.ch/record/2273852.

[30] *Measurements of properties of the Higgs boson in the diphoton decay channel with the full 2016 data set*. Tech. rep. CMS-PAS-HIG-16-040. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2264515.

[31] *Measurement of the Higgs boson coupling properties in the $H \to ZZ^* \to 4\ell$ decay channel at $\sqrt{s}$ = 13 TeV with the ATLAS detector*. Tech. rep. ATLAS-CONF-2017-043. Geneva: CERN, July 2017. URL: http://cds.cern.ch/record/2273849.

[32] Albert M Sirunyan et al. "Measurements of properties of the Higgs boson decaying into the four-lepton final state in pp collisions at sqrt(s) = 13 TeV". In: (2017). arXiv: 1706.09936 [hep-ex].

[33] Georges Aad et al. "Search for the $b\bar{b}$ decay of the Standard Model Higgs boson in associated $(W/Z)H$ production with the ATLAS detector". In: *JHEP* 01 (2015), p. 069. DOI: 10.1007/JHEP01(2015)069. arXiv: 1409.6212 [hep-ex].

[34] Serguei Chatrchyan et al. "Search for the standard model Higgs boson produced in association with a W or a Z boson and decaying to bottom quarks". In: *Phys. Rev.* D89.1 (2014), p. 012003. DOI: 10.1103/PhysRevD.89.012003. arXiv: 1310.3687 [hep-ex].

[35] Georges Aad et al. "Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV". In: *JHEP* 08 (2016), p. 045. DOI: 10.1007/JHEP08(2016)045. arXiv: 1606.02266 [hep-ex].

[36] Anna Bozena Kowalewska. "Higgs τ-lepton Yukawa Coupling Measurement and the τ Embedding Method for Background Estimation". In: *Acta Phys. Polon.* B47 (2016), pp. 1691–1696. DOI: 10.5506/APhysPolB.47.1691.

[37] Serguei Chatrchyan et al. "Evidence for the 125 GeV Higgs boson decaying to a pair of τ leptons". In: *JHEP* 05 (2014), p. 104. DOI: 10.1007/JHEP05(2014)104. arXiv: 1401.5041 [hep-ex].

[38] *Evidence for the $H \to b\bar{b}$ decay with the ATLAS detector*. Tech. rep. ATLAS-CONF-2017-041. Geneva: CERN, July 2017. URL: http://cds.cern.ch/record/2273847.

[39] *Observation of the SM scalar boson decaying to a pair of τ leptons with the CMS experiment at the LHC*. Tech. rep. CMS-PAS-HIG-16-043. Geneva: CERN, 2017. URL: http://cds.cern.ch/record/2264522.

[40] "Search for $H^0 \to b\bar{b}$ or $c\bar{c}$ in association with a W or Z boson in the forward region of pp collisions". In: (Sept. 2016). URL: http://cds.cern.ch/record/2209531.

[41] T. Aaltonen et al. "Higgs Boson Studies at the Tevatron". In: *Phys. Rev.* D88.5 (2013), p. 052014. DOI: 10.1103/PhysRevD.88.052014. arXiv: 1303.6346 [hep-ex].

[42] A. Bardi et al. "The CDF online silicon vertex tracker". In: *Nucl. Instrum. Meth.* A485 (2002), pp. 178–182. DOI: 10.1142/9789812776464_0019.

[43] F. Abe et al. "THE CDF DETECTOR: AN OVERVIEW". In: *Nucl. Instrum. Methods A* 271 (1988), pp. 387–403.

[44] Patrick T. Lukens. "The CDF IIb detector: Technical design report". In: (2003). FERMILAB-TM-2198.

[45] Anthony Allen Affolder et al. "CDF central outer tracker". In: *Nucl. Instrum. Meth.* A526 (2004), pp. 249–299. DOI: 10.1016/j.nima.2004.02.020.

[46] L. Balka et al. "The CDF Central Electromagnetic Calorimeter". In: *Nucl. Instrum. Meth.* A267 (1988), pp. 272–279. DOI: 10.1016/0168-9002(88)90474-3.

[47] S. Bertolucci et al. "The CDF Central and Endwall Hadron Calorimeter". In: *Nucl. Instrum. Meth.* A267 (1988), pp. 301–314. DOI: 10.1016/0168-9002(88)90476-7.

[48] M. G. Albrow et al. "The CDF plug upgrade electromagnetic calorimeter: Test beam results". In: *Nucl. Instrum. Meth.* A480 (2002), pp. 524–546. DOI: 10.1016/S0168-9002(01)01238-4.

[49] C. M. Ginsburg. "CDF Run 2 muon system". In: *Eur. Phys. J.* C33 (2004), S1002–S1004. DOI: 10.1140/epjcd/s2004-03-1795-6.

[50] Silvia Amerio et al. "Triggering on B-jets at CDF II". In: *IEEE Trans. Nucl. Sci.* 56 (2009), pp. 1690–1695. DOI: 10.1109/TNS.2009.2020405.

[51] A. Abulencia et al. "The CDF II eXtremely fast tracker upgrade". In: *Nucl. Instrum. Meth.* A572 (2007), pp. 358–360. DOI: 10.1016/j.nima.2006.10.204.

[52] A. Bhatti et al. "The CDF level 2 calorimetric trigger upgrade". In: *Nucl. Instrum. Meth.* A598 (2009), pp. 331–333. DOI: 10.1016/j.nima.2008.08.035.

[53] Dmitry Bandurin. "QCD measurements at the Tevatron". In: *International Workshop on Future Linear Colliders (LCWS11) Granada, Spain, September 26-30, 2011*. 2011. arXiv: 1112.0051 [hep-ex]. URL: http://inspirehep.net/record/1079578/files/arXiv:1112.0051.pdf.

[54] S. Adomeit. "Hard And Soft QCD Physics In ATLAS". In: 2012. URL: http://cdsweb.cern.ch/record/1478972/files/ATL-PHYS-PROC-2012-200.pdf.

[55] M. Voutilainen. "Hard QCD Results on Jets and Photons at CMS". In: *Proceedings, 14th Workshop on Elastic and Diffractive Scattering (EDS Blois Workshop)on Frontiers of QCD: From Puzzles to Discoveries: Qui Nhon, Vietnam, December 15-21, 2011*. 2011. URL: http://www.slac.stanford.edu/econf/C111215/papers/voutilainen.pdf.

[56] John E. Huth et al. "Toward a standardization of jet definitions". In: *1990 DPF Summer Study on High-energy Physics: Research Directions for the Decade (Snowmass 90) Snowmass, Colorado, June 25-July 13, 1990*. 1990, pp. 0134–136. URL: http://lss.fnal.gov/cgi-bin/find_paper.pl?conf-90-249.

[57] Sebastian Sapeta. "QCD and Jets at Hadron Colliders". In: *Prog. Part. Nucl. Phys.* 89 (2016), pp. 1–55. DOI: 10.1016/j.ppnp.2016.02.002. arXiv: 1511.09336 [hep-ph].

[58] Gavin P. Salam and Gregory Soyez. "A Practical Seedless Infrared-Safe Cone jet algorithm". In: *JHEP* 05 (2007), p. 086. DOI: 10.1088/1126-6708/2007/05/086. arXiv: 0704.0292 [hep-ph].

[59] Stephen D. Ellis and Davison E. Soper. "Successive combination jet algorithm for hadron collisions". In: *Phys. Rev.* D48 (1993), pp. 3160–3166. DOI: 10.1103/PhysRevD.48.3160. arXiv: hep-ph/9305266 [hep-ph].

[60] Yuri L. Dokshitzer et al. "Better jet clustering algorithms". In: *JHEP* 08 (1997), p. 001. DOI: 10.1088/1126-6708/1997/08/001. arXiv: hep-ph/9707323 [hep-ph].

[61] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The Anti-k(t) jet clustering algorithm". In: *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph].

[62] Gerald C. Blazey et al. "Run II jet physics". In: *QCD and weak boson physics in Run II. Proceedings, Batavia, USA, March 4-6, June 3-4, November 4-6, 1999*. 2000, pp. 47–77. arXiv: hep-ex/0005012 [hep-ex]. URL: http://lss.fnal.gov/cgi-bin/find_paper.pl?conf-00-092.

[63] A. Bhatti et al. "Determination of the jet energy scale at the collider detector at Fermilab". In: *Nucl. Instrum. Meth.* A566 (2006), pp. 375–412. DOI: 10.1016/j.nima.2006.05.269. arXiv: hep-ex/0510047 [hep-ex].

[64] Darin E. Acosta et al. "Measurement of the $t\bar{t}$ production cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV using lepton + jets events with secondary vertex $b-$tagging". In: *Phys. Rev. D* 71 (2005), p. 052003.

[65] Torbjorn Sjöstrand, Stephen Mrenna, and Peter Skands. "PYTHIA 6.4 Physics and Manual". In: *J. High Energy Phys.* 05 (2006). We use PYTHIA version 6.216., p. 026. DOI: 10.1088/1126-6708/2006/05/026.

[66] Julien Donini et al. "Energy Calibration of b^- Quark Jets with $Z \to b\bar{b}$ Decays at the Tevatron Collider". In: *Nucl. Instrum. Meth.* A596 (2008), pp. 354–367. DOI: 10.1016/j.nima.2008.08.133. arXiv: 0801.3906 [hep-ex].

[67] Georges Aad et al. "Measurement of the cross section of high transverse momentum $Z \to b\bar{b}$ production in proton–proton collisions at $\sqrt{s} = 8TeV$ with the ATLAS Detector". In: *Phys. Lett.* B738 (2014), pp. 25–43. DOI: 10.1016/j.physletb.2014.09.020. arXiv: 1404.7042 [hep-ex].

[68] H. L. Lai et al. "Global QCD analysis of parton structure of the nucleon: CTEQ5 parton distributions". In: *Eur. Phys. J.* C12 (2000), pp. 375–392. eprint: hep-ph/9903282.

[69] R. Brun et al. "GEANT3". CERN Report No. CERN-DD/EE/84-1. 1987.

[70] Rick Field and R. Craig Group. "PYTHIA Tune A, HERWIG, and JIMMY in Run 2 at CDF".

[71] T. Aaltonen et al. "Search for Higgs Bosons Produced in Association with b-quarks". In: *Phys. Rev.* D85 (2012), p. 032005. DOI: 10.1103/PhysRevD.85.032005. arXiv: 1106.4782 [hep-ex].

[72] S. S. Wilks. "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses". In: *Annals Math. Statist.* 9.1 (1938), pp. 60–62. DOI: 10.1214/aoms/1177732360.

[73] A. D. Martin et al. "Uncertainties of predictions from parton distributions. 2. Theoretical errors". In: *Eur. Phys. J.* C35 (2004), pp. 325–348. DOI: 10.1140/epjc/s2004-01825-2. arXiv: hep-ph/0308087 [hep-ph].

[74] Alexander L. Read. "Presentation of search results: The CL(s) technique". In: *J. Phys.* G28 (2002). [,11(2002)], pp. 2693–2704. DOI: 10.1088/0954-3899/28/10/313.

[75] Thomas Junk. "Confidence level computation for combining searches with small statistics". In: *Nucl. Instrum. Meth.* A434 (1999), pp. 435–443. DOI: 10.1016/S0168-9002(99)00498-2. arXiv: hep-ex/9902006 [hep-ex].

[76] Hans Peter Nilles. "Supersymmetry, Supergravity and Particle Physics". In: *Phys. Rept.* 110 (1984), pp. 1–162. DOI: 10.1016/0370-1573(84)90008-5.

[77] G. C. Branco et al. "Theory and phenomenology of two-Higgs-doublet models". In: *Phys. Rept.* 516 (2012), pp. 1–102. DOI: 10.1016/j.physrep.2012.02.002. arXiv: 1106.0034 [hep-ph].

[78] Eder Izaguirre, Gordan Krnjaic, and Brian Shuve. "The Galactic Center Excess from the Bottom Up". In: *Phys. Rev.* D90.5 (2014), p. 055002. DOI: 10.1103/PhysRevD.90.055002. arXiv: 1404.2018 [hep-ph].

[79] Asher Berlin, Dan Hooper, and Samuel D. McDermott. "Simplified Dark Matter Models for the Galactic Center Gamma-Ray Excess". In: *Phys. Rev.* D89.11 (2014), p. 115022. DOI: 10.1103/PhysRevD.89.115022. arXiv: 1404.0022 [hep-ph].

[80] Victor Mukhamedovich Abazov et al. "Search for neutral Higgs bosons in the multi-b-jet topology in 5.2fb^{-1} of $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV". In: *Phys. Lett.* B698 (2011), pp. 97–104. DOI: 10.1016/j.physletb.2011.02.062. arXiv: 1011.1931 [hep-ex].

[81] Vardan Khachatryan et al. "Search for neutral MSSM Higgs bosons decaying into a pair of bottom quarks". In: *JHEP* 11 (2015), p. 071. DOI: 10.1007/JHEP11(2015)071. arXiv: 1506.08329 [hep-ex].

[82] D. Acosta et al. "Measurements of $b\bar{b}$ azimuthal production correlations in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV". In: *Phys. Rev.* D71 (2005), p. 092001. DOI: 10.1103/PhysRevD.71.092001. arXiv: hep-ex/0412006 [hep-ex].

[83] A. Augusto Alves Jr. et al. "The LHCb Detector at the LHC". In: *JINST* 3 (2008), S08005. DOI: 10.1088/1748-0221/3/08/S08005.

[84] Roel Aaij et al. "Study of J/ Production in Jets". In: *Phys. Rev. Lett.* 118.19 (2017), p. 192001. DOI: 10.1103/PhysRevLett.118.192001. arXiv: 1701.05116 [hep-ex].

[85] Roel Aaij et al. "Measurement of forward $t\bar{t}$, $W + b\bar{b}$ and $W + c\bar{c}$ production in pp collisions at $\sqrt{s} = 8$ TeV". In: *Phys. Lett.* B767 (2017), pp. 110–120. DOI: 10.1016/j.physletb.2017.01.044. arXiv: 1610.08142 [hep-ex].

[86] Roel Aaij et al. "Measurement of the exclusive production cross-section in pp collisions at $\sqrt{s} = 7$ TeV and 8 TeV". In: *JHEP* 09 (2015), p. 084. DOI: 10.1007/JHEP09(2015)084. arXiv: 1505.08139 [hep-ex].

[87] Roel Aaij et al. "Measurement of forward $Z \to b\bar{b}$ production in pp collisions at $\sqrt{s} = 8$ TeV". In: (2017). arXiv: 1709.03458 [hep-ex].

[88] Oliver S. Bruning et al. "LHC Design Report Vol.1: The LHC Main Ring". In: (2004).

[89] O. Buning et al. "LHC Design Report. 2. The LHC infrastructure and general services". In: (2004).

[90] M. Benedikt et al. "LHC Design Report. 3. The LHC injector chain". In: (2004).

[91] S. Chatrchyan et al. "The CMS Experiment at the CERN LHC". In: *JINST* 3 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004.

[92] G. Aad et al. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *JINST* 3 (2008), S08003. DOI: 10.1088/1748-0221/3/08/S08003.

[93] K. Aamodt et al. "The ALICE experiment at the CERN LHC". In: *JINST* 3 (2008), S08002. DOI: 10.1088/1748-0221/3/08/S08002.

[94] R. Aaij et al. "Performance of the LHCb Vertex Locator". In: *JINST* 9 (2014), p. 09007. DOI: 10.1088/1748-0221/9/09/P09007. arXiv: 1405.7808 [physics.ins-det].

[95] "LHCb magnet: Technical design report". In: (2000).

[96] "LHCb: Inner tracker technical design report". In: (2002).

[97] R Arink et al. "Performance of the LHCb Outer Tracker". In: *JINST* 9.01 (2014), P01002. DOI: 10.1088/1748-0221/9/01/P01002. arXiv: 1311.3893 [physics.ins-det].

[98] R. Fruhwirth. "Application of Kalman filtering to track and vertex fitting". In: *Nucl. Instrum. Meth.* A262 (1987), pp. 444–450. DOI: 10.1016/0168-9002(87)90887-4.

[99] Roel Aaij et al. "LHCb Detector Performance". In: *Int. J. Mod. Phys.* A30.07 (2015), p. 1530022. DOI: 10.1142/S0217751X15300227. arXiv: 1412.6352 [hep-ex].

[100] M. Kucharczyk et al. "Primary Vertex Reconstruction at LHCb". In: (2014).

[101] M. Adinolfi et al. "Performance of the LHCb RICH detector at the LHC". In: *Eur. Phys. J.* C73 (2013), p. 2431. DOI: 10.1140/epjc/s10052-013-2431-9. arXiv: 1211.6759 [physics.ins-det].

[102] "LHCb calorimeters: Technical design report". In: (2000).

[103] A. A. Alves Jr. et al. "Performance of the LHCb muon system". In: *JINST* 8 (2013), P02022. DOI: 10.1088/1748-0221/8/02/P02022. arXiv: 1211.1346 [physics.ins-det].

[104] Markus Frank et al. "Deferred High Level Trigger in LHCb: A Boost to CPU Resource Utilization". In: *J. Phys. Conf. Ser.* 513 (2014), p. 012006. DOI: 10.1088/1742-6596/513/1/012006.

[105] Matthew William Kenzie and Vladimir Gligorov. *Lifetime unbiased beauty and charm triggers at LHCb*. Tech. rep. LHCb-PUB-2015-026. CERN-LHCb-PUB-2015-026. Geneva: CERN, Oct. 2016. URL: http://cds.cern.ch/record/2110638.

[106] Giulio Dujany and Barbara Storaci. "Real-time alignment and calibration of the LHCb Detector in Run II". In: *J. Phys.: Conf. Ser.* 664.LHCb-PROC-2015-011. CERN-LHCb-PROC-2015-011 (Apr. 2015), 082010. 8 p. URL: https://cds.cern.ch/record/2017839.

[107] R. Aaij et al. "Tesla : an application for real-time data analysis in High Energy Physics". In: *Comput. Phys. Commun.* 208 (2016), pp. 35–42. DOI: 10.1016/j.cpc.2016.07.022. arXiv: 1604.05596 [physics.ins-det].

[108] M. Williams et al. "The HLT2 Topological Lines". In: (2011).

[109] D. H. Campora Perez et al. "The 40 MHz trigger-less DAQ for the LHCb Upgrade". In: *Nucl. Instrum. Meth.* A824 (2016), pp. 280–283. DOI: 10.1016/j.nima.2015.10.047.

[110] R. Brun and F. Rademakers. "ROOT: An object oriented data analysis framework". In: *Nucl. Instrum. Meth.* A389 (1997), pp. 81–86. DOI: 10.1016/S0168-9002(97)00048-X.

[111] G. Barrand et al. "GAUDI - A software architecture and framework for building HEP data processing applications". In: *Comput. Phys. Commun.* 140 (2001), pp. 45–55. DOI: 10.1016/S0010-4655(01)00254-5.

[112] M. Clemencic et al. "The LHCb simulation application, Gauss: Design, evolution and experience". In: *J. Phys. Conf. Ser.* 331 (2011), p. 032023. DOI: 10.1088/1742-6596/331/3/032023.

[113] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. "A Brief Introduction to PYTHIA 8.1". In: *Comput. Phys. Commun.* 178 (2008), pp. 852–867. DOI: 10.1016/j.cpc.2008.01.036. arXiv: 0710.3820 [hep-ph].

[114] Stefano Frixione, Paolo Nason, and Carlo Oleari. "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method". In: *JHEP* 11 (2007), p. 070. DOI: 10.1088/1126-6708/2007/11/070. arXiv: 0709.2092 [hep-ph].

[115] Michelangelo L. Mangano et al. "ALPGEN, a generator for hard multiparton processes in hadronic collisions". In: *J. High Energy Phys.* 07 (2003), p. 001. DOI: 10.1088/1126-6708/2003/07/001. arXiv: hep-ph/0206293 [hep-ph].

[116] D. J. Lange. "The EvtGen particle decay simulation package". In: *Nucl. Instrum. Meth.* A462 (2001), pp. 152–155. DOI: 10.1016/S0168-9002(01)00089-4.

[117] S. Agostinelli et al. "GEANT4: A Simulation toolkit". In: *Nucl. Instrum. Meth.* A506 (2003), pp. 250–303. DOI: 10.1016/S0168-9002(03)01368-8.

[118] William Barter et al. *Jets reconstruction and performances at LHCb*. Tech. rep. LHCb-INT-2012-015. CERN-LHCb-INT-2012-015. Geneva: CERN, May 2012. URL: https://cds.cern.ch/record/1446552.

[119] Philip James Ilten. *Neutral energy recovery in Turbo*. Tech. rep. LHCb-INT-2016-049. CERN-LHCb-INT-2016-049. Geneva: CERN, Dec. 2016. URL: https://cds.cern.ch/record/2238428.

[120] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "FastJet User Manual". In: *Eur. Phys. J.* C72 (2012), p. 1896. DOI: 10.1140/epjc/s10052-012-1896-2. arXiv: 1111.6097 [hep-ph].

[121] Murilo Santana Rangel et al. *Jet energy studies for RunI simulation*. Tech. rep. LHCb-INT-2013-036. CERN-LHCb-INT-2013-036. Geneva: CERN, June 2013. URL: https://cds.cern.ch/record/1556336.

[122] Andreas Hocker et al. "TMVA - Toolkit for Multivariate Data Analysis". In: *PoS* ACAT (2007), p. 040. arXiv: physics/0703039 [PHYSICS].

[123] William Barter. *Estimating Differences in the Jet Energy Scale between Simulation and Data*. Tech. rep. LHCb-INT-2015-023. CERN-LHCb-INT-2015-023. Geneva: CERN, June 2015. URL: https://cds.cern.ch/record/2026389.

[124] Roel Aaij et al. "Identification of beauty and charm quark jets at LHCb". In: *JINST* 10.06 (2015), P06013. DOI: 10.1088/1748-0221/10/06/P06013. arXiv: 1504.07670 [hep-ex].

[125] Georges Aad et al. "Measurement of $W^{\pm}$ and Z-boson production cross sections in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Phys. Lett.* B759 (2016), pp. 601–621. DOI: 10.1016/j.physletb.2016.06.023. arXiv: 1603.09222 [hep-ex].

[126] Leo Breiman et al. *Classification and regression trees*. Chapman and Hall/CRC, 1984. ISBN: 0412048418, 9780412048418.

[127] Byron P. Roe et al. "Boosted decision trees, an alternative to artificial neural networks". In: *Nucl. Instrum. Meth.* A543.2-3 (2005), pp. 577–584. DOI: 10.1016/j.nima.2004.12.018. arXiv: physics/0408124 [physics].

[128] Alex Rogozhnikov et al. "New approaches for boosting to uniformity". In: *JINST* 10.03 (2015), T03002. DOI: 10.1088/1748-0221/10/03/T03002. arXiv: 1410.4140 [hep-ex].

[129] Wouter Verkerke and David P. Kirkby. "The RooFit toolkit for data modeling". In: *eConf* C0303241 (2003). [,186(2003)], MOLT007. arXiv: physics/0306116 [physics].

[130] Karol Hennessy. "LHCb VELO Upgrade". In: *Nucl. Instrum. Meth.* A845 (2017), pp. 97–100. DOI: 10.1016/j.nima.2016.04.077. arXiv: 1604.05045 [physics.ins-det].

[131] T. Aaltonen et al. "Improved b-jet Energy Correction for $H \to b\bar{b}$ Searches at CDF". In: (2011). arXiv: 1107.3026 [hep-ex].

[132] CMS Collaboration. "Boosted Top Jet Tagging at CMS". In: *CMS-PAS-JME-13-007* (2014).