Alma Mater Studiorum · Università di Bologna

**Scuola di Scienze**
**Corso di Laurea Magistrale in Fisica**

# MULTIVARIATE ANALYSIS METHODS IN THE SEARCH FOR THE HIGGS BOSON PRODUCED IN ASSOCIATION WITH TOP PAIRS AT THE ATLAS EXPERIMENT AT LHC.

**Relatore:**
**Chiar.mo Prof.**
**NICOLA SEMPRINI CESARI**

**Presentata da:**
**GIULIA UCCHIELLI**

**Correlatori:**
**Dott. ANTONIO SIDOTI**
**Dott.ssa SARA VALENTINETTI**

**Sessione II**
**Anno Accademico 2013-2014**

**Sommario**

Il Large Hadron Collider, situato presso i laboratori del CERN di Ginevra, è il più grande acceleratore di particelle al mondo. Una delle linee di ricerca più importanti riguarda lo studio del bosone di Higgs, l'ultima particella scoperta dagli esperimenti ATLAS e CMS.
A causa della bassa sezione d'urto di produzione, lo studio delle proprietà del bosone di Higgs richiede in primo luogo la disponibilità di rilevanti statistiche. Affinchè tali studi siano efficaci è necessario evitare che la segnatura dei decadimenti del bosone di Higgs sia oscurata dal numero e dalla varietà dei processi di fondo prodotti nelle collisioni $pp$ ad LHC.
Particolare rilevanza pertanto assume lo studio di metodi di analisi multivariata che, rispetto alle tradizionali tecniche *cut-based*, possano migliorare la selezione del segnale di un Higgs prodotto in associazione con una coppia di quark top, in uno stato finale dileptonico (canale $t\bar{t}H$).
La statistica finora accumulata non è sufficiente per fornire un numero significativo di eventi nel canale $t\bar{t}H$ tuttavia le metodologie impiegate, oggetto della presente tesi, rappresentano uno strumento essenziale con l'aumentare della statistica prevista nella prossima presa dati.

**Abstract**

The Large Hadron Collider, located at the CERN laboratories in Geneva, is the largest particle accelerator in the world. One of the main research fields at LHC is the study of the Higgs boson, the latest particle discovered at the ATLAS and CMS experiments.
Due to the small production cross section for the Higgs boson, only a substantial statistics can offer the chance to study this particle properties. In order to perform these searches it is desirable to avoid the contamination of the signal signature by the number and variety of the background processes produced in $pp$ collisions at LHC.
Much account assumes the study of multivariate methods which, compared to the standard cut-based analysis, can enhance the signal selection of a Higgs boson produced in association with a top quark pair through a dileptonic final state ($t\bar{t}H$ channel).
The statistics collected up to 2012 is not sufficient to supply a significant number of $t\bar{t}H$ events; however, the methods applied in this thesis will provide a powerful tool for the increasing statistics that will be collected during the next LHC data taking.

# Contents

# List of Figures

# List of Tables

# Introduction

During the last 60 years, much effort has been devoted to the search for a theoretical model able to explain the fundamental interactions among fundamental particles. The beginning of the 1900 has been a prosperous age for the discovery of new particles originated from the interaction of cosmic rays with the Earth atmosphere or, after some time, producted in the newly born particle accelerators.
The Standard Model of particle physics represents the most complete theory and, at the same time, it is greatly supported by experimental evidences. This theory predicts the existence of a particle responsible for giving mass to all the fundamental massive particles of the Standard Model: the Higgs boson. The measurement of the properties of this new particle is essential for contemporary Physics.
The Higgs mechanism provides the mediators of the weak interaction with mass through the *spontaneous symmetry breaking* phenomenon preserving the local gauge invariance of the theory.
The Large Hadron Collider, located at CERN laboratories in Geneva, is the largest particle accelerator in the world and it is designed to reach energies of 14 TeV in the center of mass allowing the search for processes at the energy frontier such as the Higgs production. On the $4^{th}$ July 2012 the CERN laboratories announced the discovery of a 126 GeV mass resonance with properties similar to the SM Higgs boson with a confidence level of more than $5\sigma$.
Among the main Higgs boson production mechanisms at LHC there is the $t\bar{t}H$ channel, in which the Higgs boson is produced in association with a top quark pair. Although the production cross section for the $t\bar{t}H$ channel, which is around $\sigma(t\bar{t}H) \sim 130$ fb, is the smallest one compared to the other production mechanisms, it deserves direct attention for many reasons. First of all, it is the only process which, up to now, has not been observed and its production cross section still remains unmeasured, but can only be constrained through theoretical considerations. In fact, the small cross section and the integrated luminosity collected during the 2012 data taking (20.3 $\text{fb}^{-1}$) predicts an expected number of $t\bar{t}H$ events which is about $10^3$. The

limited statistics collected for this process is at the moment not sufficient to perform any evidence or observation but only to test methods for selecting signal events and estimating background.
Furthermore, the $t\bar{t}H$ channel permits a direct measurement of the quark top-Higgs Yukawa coupling which is expected to be of the order of one since the top quark is the heaviest particle (173.02 GeV) of the Standard Model.
Moreover, the $t\bar{t}H$ production cross section will grow of a factor 4.7 in LHC Run II during which the center of mass energy raises to 13 TeV. This will be the larger relative increase with respect to the other Higgs production mechanisms.
For these reasons the $t\bar{t}H$ channel, the one under analysis in this thesis, is becoming of particular interest in the study of the Higgs boson production.
This channel presents a quite complex signature with a high jet multiplicity and a varying number of leptons in the final state. The different final states under analysis are determined from the different decay modes of the top quark pair and of the Higgs boson. The analysis presented in this work requires in the final state the presence of two leptons with same sign and a minimum of two jets. This signature represents only one of the possible multilepton signatures which provide a very clean channel, especially when the jet and bjet multiplicities become high. The study of this channel cannot be performed without modelling the different background processes which contribute to signal contamination and worsen its selection. Backgrounds are mainly divided into reducible and irreducible and presents different characteristics and event topologies. While the formers present a final state whose signature actually mimics the signal one, the latters can fake signal events only when one or more objects in the final state are misreconstructed. The understanding of how these processes can be reduced becomes fundamental in the search for rare events such as the $t\bar{t}H$ production. The aim of this analysis is to present an alternative to standard cut-based analysis using multivariate methods. Among them, *Boosted Decision Trees* and *Neural Networks* are the methods studied here. The advantage in using multivariate methods stands in the possibility to improve the signal selection (or signal significance $S/\sqrt{S+B}$ ) with respect to cut-based analysis: the selections applied to the signal and background samples, by multivariate methods, are performed by learning from data and by minimizing a "cost" function between the obtained and the desired output on a training sample.
Due to the lack of statistics in the $t\bar{t}H$ channel and since the analysis on real data has still to be finalized, this analysis has been performed on Monte Carlo sample which have been produced in order to simulate the data collected during 2012 with the ATLAS detector with an integrated luminosity of 20.3 $fb^{-1}$.

The first chapter will focus on the theoretical basis of the Standard Model and the Higgs mechanism. In the second chapter a description of the LHC collider, as well as of the ATLAS detector, will be provided. The third chapter will be devoted to the main production mechanisms as well as the decay modes of the Higgs boson; a detailed description of the $t\bar{t}H$ channel will be provided. In the fourth chapter the main methods of multivariate analysis and their characteristics will be discussed. Finally the analysis procedure and the results of the analysis will be presented in the fifth chapter.

# Chapter 1

# The Standard Model and the Higgs Mechanism

## 1.1 Introduction to the Standard Model

The Standard Model (SM) of particle physics represents the theoretical frame which describes, from the 70's on, the fundamental particles (fermions, which are further divided into leptons and quark) and their interactions (mediated by gauge bosons), up to distances of 1fm=$10^{-15}$m, in terms of a specific type of relativistic quantum field theory: a gauge theory.
The SM is based on a local simmetry extending the gauge invariance of QED to a number of charges defined in abstract spaces, thus forming a bigger group of simmetry. The Standard Model symmetry group

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

is defined by the product of the single interaction ones. In equation 1.1, $SU(3)_C$ is the non-Abelian group associated with the strong interaction caused by the colour charge; its mediator are eight massless gauge bosons (*gluons*), which interact with quarks according to quantum chromodynamics (QCD). $SU(2)_L \otimes U(1)_Y$ is the simmetry group which describes the electroweak interaction, better known as Glashow[1]-Weinberg[2]-Salam[3] (GWS) theory. The $SU(2)_L$ group, associated with the weak isospin, corresponds to three gauge fields, while only one is connected with the hypercharge Y, and thus to the $U(1)_Y$ group. Although the gluons, as well as the photons, are massless, the weak bosons have masses of the order of 100 GeV. This experimental evidence, however, is in conflict with the predictions of the Standard Model, where all the gauge bosons should be massless.

The solution to this problem is given by the *spontaneous simmetry breaking*, a mechanism which introduces a neutral scalar field, known as the Higgs field, which gives mass to all the particles of the Standard Model. The power of this theoretical mechanism lies in the solution to the mass problems without spoiling the gauge invariance of the model and in the prediction of a new particle: the Higgs boson whose mass, however, cannot be predicted by the theory.
This chapter will be devoted to the fundamental concepts of the gauge theories and to the electroweak unification. A general description of the lagrangian of the three interactions will also be given. Finally, the spontaneous simmetry breaking will be treated as well as the discovery of the Higgs boson.

## 1.2 The Particles of the Standard Model

On the basis of their collective behaviour, the elementary particles of the Standard Model can be divided into two main categories: fermions and bosons. Fermions consist of quark and leptons, which follow the Fermi-Dirac statistics as they are spin 1/2 particles and they thus obey the Pauli exclusion principle. On the basis of weak interactions, quarks are further divided into three generations:

$$\begin{pmatrix} u \\ d \end{pmatrix} \begin{pmatrix} c \\ s \end{pmatrix} \begin{pmatrix} t \\ b \end{pmatrix}$$

and are identified as the costituents of hadrons. They are the only elementary particles provided with charges (2/3 for u,c,t and -1/3 for d,s,b), which are fraction of the electron charge ($1.6022 \times 10^{-19}$C). This characteristic differs from all the other observable particles whose charge is an integer multiple of the electron one. These fundamental constituents possess different quantum numbers: *colour*, *barionic number* and *flavour*. Colour (*red, blue and green*) is actually responsible for *confinement*: quarks cannot be observed in free states but only in a combination which nulls the colour charge. The barionic number, which is 1/3 for all quarks, is additive and conserved by all the three interactions. Finally, quarks have six different flavours *u, d, c, s, t, b*, that are conserved in all interactions except the weak force.
As well as quarks, leptons are divided into three families by weak interactions:

$$\begin{pmatrix} e \\ \nu_e \end{pmatrix} \begin{pmatrix} \mu \\ \nu_\mu \end{pmatrix} \begin{pmatrix} \tau \\ \nu_\tau \end{pmatrix}$$

whose ranges in mass are substantially different. The SM regards neutrinos as massless particles, in contrast with the experimental evidence of *neutrino oscillations*. According to this phenomenon, neutrinos should have mass but,

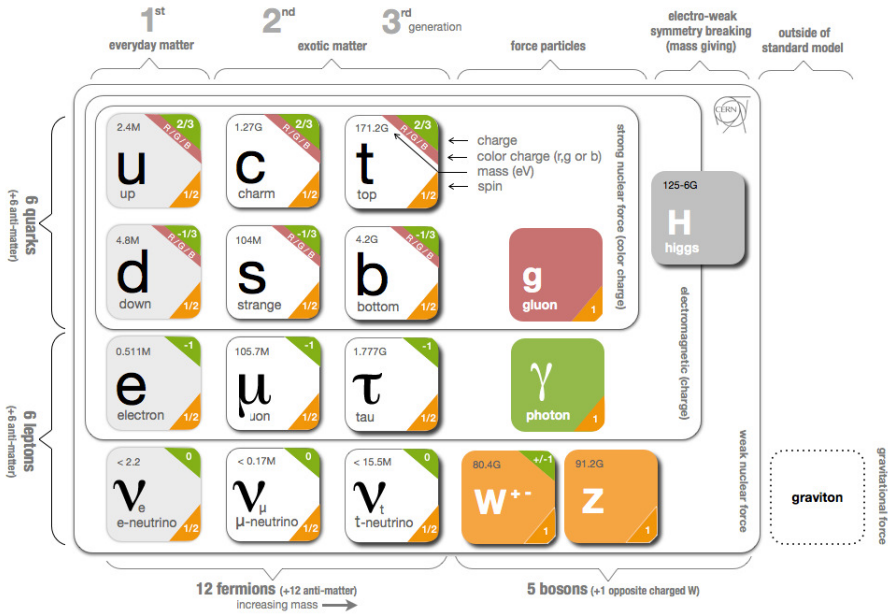



Figure 1.1: Main components of the Standard Model. They are divided into fermions (left) and gauge bosons (right). The Higgs boson, responsible for the spontaneous symmetry breaking and the graviton, which is not a particle of the SM, are also illustrated. Each particle is reported with its mass value or, in the neutrino case, with an upper limit on the mass, provided by the Particle Data Group [4].

up to now, only limits are provided by direct measurements [4]. Neutrinos are not provided with charge: they can thus interact only by means of the weak force. By contrast, $e$, $\mu$ and $\tau$ have negative charge, which allows them to interact electromagnetically and via weak force. The additive *leptonic number* is associated to leptons. This number is 1 for all leptons and it is globally conserved by all the interactions. Actually there are three different leptonic numbers: electronic, muonic and tauonic, but they are not individually preserved in weak interactions, as happens in neutrino oscillations for example.
Figure 1.1 shows all the compontents of the SM: on the left the three quark (top) and lepton (bottom) families are presented. On the right, the five gauge bosons as well as the Higgs boson and the graviton, which however is outside the SM, are shown, as mediators of forces.

The Standard Model is based on quantum field theories [5], according to

which particle interactions are described through the exchange of the field quanta. These quanta, called gauge bosons, follow the Bose-Einstein statistics and have integer spin. Every fundamental interaction has its mediators whose properties are illustrated in table 1.1. Within the Standard Model gravitation is not taken into account since, at the moment, it cannot be described in terms of a quantum field theory.

| Mediator | Interaction | Mass (GeV) | Spin |
|---|---|---|---|
| $\gamma$ | Electromagnetic | 0 | 1 |
| $W^{\pm}$ | Weak | 80.34 | 1 |
| $Z^0$ | Weak | 91.19 | 1 |
| *gluons* | Strong | 0 | 1 |

Table 1.1: For each gauge boson, the main characteristics, as well as the associated force, are listed [4].

Since the electromagnetic and weak interactions present similar properties, during the 60's they were unified in the *electroweak force*, as it will be shown later. However, as follows from table 1.1, only the weak vector bosons are not massless. In order to preserve the gauge invariance of the model, it is necessary to introduce a spontaneous symmetry breaking, which is better known as *Higgs mechanism* and which is going to be treated in detail in section 1.4.

## 1.3 Gauge Theories

The foundations for the present understanding of gauge theories were laid by Maxwell in 1864 with his equations for the description of the electromagnetic field. It is in fact true that the electromagnetic potential, which is introduced in order to generate fields, is not uniquely defined by construction. This freedom to choose many potentials that describe the same electromagnetic fields is called *gauge invariance*. [5]
What is crucial in gauge theories is the notion of symmetry, or the invariance of the lagrangian of the system under the trasformation of a certain group of symmetry. When the theory changes in every point of the space-time of a same quantity under certain transformation, its equation presents a global symmetry with respect to that property. On the other hand, when the system is invariant under trasformations which depend on a parameter

which is locally defined, its lagrangian presents a local symmetry. SM is based on lagrangians which are invariant under local symmetries.

### 1.3.1 Quantum Electrodynamics : QED

Quantum Electrodynamics describes the interactions between charged particles. This theory is the one which is most complete and better supported by experimental evidences. It is thus a model for all the other theories of SM. QED is based on Maxwell equations [6], reinterpreted in quantomechanical and relativistical terms. The free matter lagrangian density can be written as:

$$\mathcal{L} = \bar{\psi}(i \not{\partial} - m)\psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} - \frac{\lambda}{2}(\partial_\mu A^\mu)^2 \tag{1.2}$$

which is invariant under Lorentz transformations.
In equation 1.2 the first term is the Dirac lagrangian for the propagation of the free fermion $\psi$; the second term is the lagrangian for the electromagnetic field, where $F^{\mu\nu} = \partial^\nu A^\mu - \partial^\nu A^\mu$. The last term is a *gauge fixing* term which is necessary to erase any unuseful degrees of freedom. What is missing in 1.2 is the interaction term. This can be derived by applying a local phase trasformation of the group U(1), associated to the electric charge, to the wave function and requiring the theory invariance:

$$\psi \to \psi' = e^{-i\alpha(x)}\psi \tag{1.3}$$

where $\alpha(x)$ is a real parameter function of x. Equation 1.2 is not invariant under the transformation 1.3. Invariance is restored by replacing the derivative operator with:

$$\partial_\mu \to D_\mu = \partial_\mu + iqA_\mu \tag{1.4}$$

which is called *covariant derivative*. After the introduction of this new operator, the lagrangian becomes locally invariant and takes the form:

$$\mathcal{L}_{QED} = \bar{\psi}(i \not{\partial} - m)\psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} + ieA_\mu\bar{\psi}\gamma^\mu\psi \tag{1.5}$$

where $\gamma^\mu$ are the Dirac matrices [5]. It naturally follows that the gauge principle is responsible for the introduction of an interaction term between the charged fermion and the gauge boson of the electromagnetic field, the photon. The fundamental QED interaction vertex between fermions and photon is shown in figure 1.2.
As it will be shown later, for every generator of a group of symmetry for which local invariance is requested, a new gauge boson will be introduced. Since U(1) has only one generator, the only particle that is introduced is the photon $\gamma$.

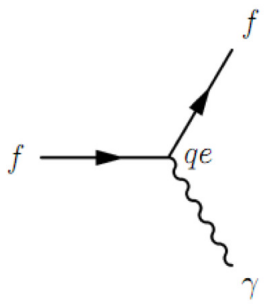



Figure 1.2: Fundamental interaction vertex of quantum electrodynamics between a fermion and the photon. The coupling constant for the process is the electron charge $q_e$.

### 1.3.2 Quantum Chromodynamics : QCD

Quantum chromodynamics describes strong interactions between quarks and gluons in hadrons. The colour charge is responsible for this interaction and, since there are three differents states of colour, quarks are represented by a three component spinor $\psi$. Colour states transform as a SU(3) group triplet whose components are unitary matrices. The eight generators of the fundamental rapresentation of SU(3) are usually defined by the Gell-Mann matrices $\lambda^a$ [7], which play the same role of Pauli's matrices $\sigma^i$ for the group SU(2). An arbitrary element of the SU(3) group is thus described by $3 \times 3$ matrices in the form of $U = e^{-i\alpha_a \frac{\lambda^a}{2}}$.
Starting again from Dirac lagrangian and using the gauge principle, it is possibile to write the lagrangian for the SU(3) group as [7]:

$$\mathcal{L} = \bar{\psi}^a (i \not{D}^{\alpha\beta} - m\delta^{\alpha\beta})\psi^\beta \tag{1.6}$$

where $\not{D}^{\alpha\beta}$ is the covariant derivative in the fundamental rapresentation of the SU(3) group:

$$D_\mu^{\alpha\beta} = \delta^{\alpha\beta}\partial_\mu + ig_s A_{\mu a}(T^a)^{\alpha\beta}. \tag{1.7}$$

In this last equation $T^a$ is a generic generator of the group, $\delta^{\alpha\beta}$ is the Kroneker symbol, $g_s$ is the coupling costant of QCD and $A_\mu$ is the gluonic propagator of strong interaction. In analogy with QED, also quantum chromodynamics introduces a number of gauge bosons equal to the number of generators of the group of symmetry; for QCD there are eight particles called gluons. As presented in the previous paragraph for QED, it is possibile to write the

propagation of the strong field as:

$$-\frac{1}{4}F^a_{\mu\nu}F^{a\mu\nu} \tag{1.8}$$

but the remarkable difference is that the new definition of covariant derivative changes the rapresentation of the field $F^a_{\mu\nu}$:

$$F^a_{\mu\nu} = \partial_\nu A^a_\mu - \partial_\mu A^a_\nu + g_s f^{abc} A^b_\mu A^c_\nu. \tag{1.9}$$

where $f^{abc}$ are the fine structure constants of the group. In non-Abelian theories, such as QCD, the group generators do not commute between themselves and, in fact, the third term in equation 1.9 is different from zero. It follows that, although in QED photons cannot interact with themselves, in QCD the term which appears in the definition of the field requiring local invariance leds to terms of *self interaction.*
Finally, the QCD lagrangian can be written as

$$\mathcal{L}_{QCD} = -\frac{1}{4}F^a_{\mu\nu}F^{\mu\nu a} - \sum_{f=1}^{6}\bar{\psi}_f(\gamma^\mu D_\mu - m_f)\psi_f \tag{1.10}$$

where the index *f=(u, d, c, s, t, b)* stands for the quark flavour. Actually, it can be proved that in this equation is necessary to introduce a term of *ghost lagrangian*, $\mathcal{L}_{ghost}$, which counts for the presence of new "ghost particles" associated to gauge bosons. Ghost particles only appear in loops and not in initial and final states. Consequently, they cannot be considered "real particles", which are experimentally observable. Ghost terms can be considered as corrections of the more general formalism of gauge fixing. As already said, QCD is a non-Abelian theory. The gluon self interaction term, which appears in the QCD lagrangian, have also an effect on the expression of the strong coupling costant $\alpha_s$ [5]:

$$\alpha_s(q^2) = \alpha_s(\mu^2)\left[1 + \frac{\alpha_s(\mu^2)}{12\pi}\log\left(\frac{-q^2}{\mu^2}\right)(2n_f - 11N) + O(\alpha_s^2)\right] \tag{1.11}$$

where $\mu^2$ is a spacelike renormalization point, $n_f$ the number of quark flavours appearing in the loops, N the number of colours. In equation 1.11, there are two competitive terms: the fermion loop tends to enhance the effective coupling costant at short distances or large values of $\mu^2$. On the contrary, the contribution due to the three-gluon interaction is of the opposite sign and tends to decrease the strenght of the interaction at short distances; it corresponds to an *antiscreening* effect. The existence of a regime in which

$\alpha_s(q^2) \ll 1$ implies a realm in which QCD perturbation theory should be valid. This property of non-Abelian theories is known as *asymptotic freedom*: at very short distances, quarks behave nearly as free particles within hadrons. By contrast, the growth of the coupling constant at large distances indicates the existence of a range in which strong interactions become dominant. The strong-coupling regime is of key importance for quark confinement and it is responsible for hadronization. On a phenomenological point of view, it is the only domain in which it is possible, by indirect methods, to observe the existence of quarks.
When evaluating QCD previsions on different cross sections, virtual-loop corrections and singularities, at different perturbative orders, have to be taken into account. For instance, the cross section for a certain process can be written as [4]:

$$\sigma(Q) = \sigma(Q)_{EW}(1 + \delta_{QCD}(Q)) \tag{1.12}$$

where $\sigma(Q)_{EW}$ is the pure electroweak prediction for a certain problem, and $\delta_{QCD}(Q)$ is the correction due to QCD effects. If the center of mass energy Q$\ll M_Z$, the QCD correction can be written as:

$$\delta_{QCD}(Q) = \sum_{n=1}^{\infty} c_n \left(\frac{\alpha_s(Q^2)}{\pi}\right)^n + \mathcal{O}\left(\frac{\Lambda^4}{Q^4}\right) \tag{1.13}$$

where $c_n$ are the coefficients of constant expansion terms and $\mathcal{O}\left(\frac{\Lambda^4}{Q^4}\right)$ is an extra "power correction". It follows from equation 1.13 that the coefficients of $\alpha_s^n$ increase order by order. Calculations in perturbative QCD tend to converge more slowly than would be expected just on the size of $\alpha_s$. The lowest order ($n = 1$) of perturbative calculation, at which a quantity is non-zero, is called *Leading Order (LO)*. Given an observable that is non-zero starting from $n$ particles, *next-to-leading order (NLO)* corrections, which depends on $\alpha_s^2$, have to be computed. Conceptually, *next-to-next-to-leading order (NNLO)* and NLO corrections are similar, except that a further order in $\alpha_s$ has to be added. The increse in the perturbative orders results in an incresing complexity of loop calculations relative to the tree-level calculations, which is the lowest level of perturbation theory. The accuracy of QCD predictions at LO calculation is to within a factor of two. It is also limited by non-perturabative corrections, estimated from the difference between Monte Carlo events at the parton level and after hadronization [4].

### 1.3.3 Theory of Weak Interactions

Weak interactions occur between any of the fermions of the SM. However, the intensity of the weak force is relatively small ($10^{-5}$ times the intensity of

the strong interaction) and thus this force can be hidden by electromagnetic and strong phenomena. It is easier to observe weak interactions in neutrino exchanges, or in *flavour changing charged currents* such as the $\beta$ decay of the neutron which requires a transition from a quark *down* to an *up* one.
The first attempt to a theoritical description of weak interactions was made by Enrico Fermi in 1933. Fermi gave explanation to nuclear $\beta$ decay as a punctual interaction of four fermions. Transitions between particles had been described as vectorial (V) currents in analogy with the electromagnetic ones. Nevertheless, experimental evidences of parity violation led to the modification of Fermi's theory in order to include these observed phenomena and to mantain the Lorentz-invariance of the Hamiltonian. The weak current required an axial (A) term, which conserves its sign under parity transformations. The current acquired thus a vector-axial form of the type [5]:

$$J_\alpha(x) = \sum_l \bar{\psi}_l(x)\gamma_\alpha(1-\gamma_5)\psi_{\nu l}(x) \tag{1.14}$$

and succeed to explain parity violation. Writing separately the V and A terms:

$$J^V_\alpha(x) = \sum_l \bar{\psi}_l(x)\gamma_\alpha\psi_{\nu l}(x) \tag{1.15}$$

$$J^A_\alpha(x) = \sum_l \bar{\psi}_l(x)\gamma_\alpha\gamma_5\psi_{\nu l}(x) \tag{1.16}$$

it is possibile to see that only the vectorial term changes sign under the parity transformation so that in weak interactions, where products of the V·A terms appear, parity is not conserved. In more details, weak interaction operates on certain elicity states, for massless particles or in ultrarelativistic regime, defined by the chirality projector operators:

$$\psi^L(x) = \frac{1-\gamma_5}{2}\psi(x) \tag{1.17}$$

$$\psi^R(x) = \frac{1+\gamma_5}{2}\psi(x) \tag{1.18}$$

which respectively create "left-handed" and "right-handed" helicity states. The first equation enters directly the V-A current and thus only left-handed particles can weakly interact. Since weak interaction violates parity, it is not symmetric for the exchange of helicity states. It is then convenient to divide every fermionic family into a left-handed doublet and a right-handed singlet. By the introduction of the *weak ispospin* of SU(2), which is the group of symmetry associated to weak interactions, left-handed leptons behave as a

doublet of isospin, while right-handed ones are singlets. These considerations have important consequences in making weak interaction a gauge theory. When applying the covariant deritivative of the SU(2) group in the Dirac lagrangian $\mathcal{L} = \bar{\psi}(i \not{D} - m)\psi$, terms of the type

$$\bar{\psi} \not{\partial}\psi = \bar{\psi}_L \not{\partial}\psi_L + \bar{\psi}_R \not{\partial}\psi_R \quad (1.19)$$

$$\bar{\psi} \not{A}\psi = \bar{\psi}_L \not{A}\psi_L + \bar{\psi}_R \not{A}\psi_R \quad (1.20)$$

$$m\bar{\psi}\psi = m\bar{\psi}_L \not{\partial}\psi_R + \bar{\psi}_R \not{\partial}\psi_L \quad (1.21)$$

appear in the equation. As visible, while the first two terms involve only one state of helicity, the third one mixes left-handed and right-handed particles. V-A theory does not allow mixed state of helicity and, as a consequence, leptons must be massless. Nevertheless, as it will be soon explained, leptons will obtain their mass through the mechanism of spontaneous symmetry breaking.

### 1.3.4 The Unified Electroweak Model of Glashow-Salam-Weinberg

During the 60's, Glashow, Salam and Weinberg started to look for a possibile unification of two of the fundamental forces: the weak and the electromagnetic interactions. At first, Glashow in 1961 [1] pointed out that weak interactions, as well as in QED, could have been mediated by gauge vector bosons and that the Hamiltonian for this interaction could have been written as:

$$H(x) = g_W J^{\alpha\dagger}(x) W_\alpha(x) + g_W J^\alpha(x) W^\dagger_\alpha(x) \quad (1.22)$$

where $g_W$ represents an adimensional coupling costant and the field $W_\alpha$ is associated to the vector boson which is the mediator of the force.
Weinberg in 1967 [2] and Salam in 1968 [3], reached indipendently the same conclusion and led to the construction of the unified electroweak theory. They introduced a lagrangian, invariant for gauge transformations, of SU(2) weak isospin T and U(1) weak hypercharge Y: these two quantities are connected to the electric charge by equation

$$Q = T_3 + \frac{1}{2}Y \quad (1.23)$$

know as Gell-Mann–Nishijima formula, where $T_3$ is the third component of the isospin. In the GWS theory, fermions of the Standard Model are divided into isospin multiplets where fermions $\psi_L$ form a doublet and $\psi_R$ a singlet. Table 1.2 shows the electroweak fermionic multiplets. Left-handed weak isospin doublets are provided in parentheses while right-handed singlets form

separated states. Conventionally, the strong interaction $u$, $c$ and $t$ eigenstates as well as the rotated, under the CKM matrix transformation, $d'$, $s'$ and $b'$ eigenstates represent left-handed quarks. The electric charge Q for same doublet states differs by a unit and the difference Q-$I_3$ = $Y_W/2$ is the same inside each doublet (-1/2 for left-handed leptons and +1/6 for quarks).

| | Fermionic multiplets | | | I | $I_3$ | Q | $Y_3$ |
|---|---|---|---|---|---|---|---|
| Leptons | $\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L$ | $\begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L$ | $\begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L$ | 1/2 | $+1/2$<br>$-1/2$ | 0<br>$-1$ | $-1$<br>$-1$ |
| | $e_R$ | $\mu_R$ | $\tau_R$ | 0 | 0 | $-1$ | $-2$ |
| Quarks | $\begin{pmatrix} u \\ d' \end{pmatrix}_L$ | $\begin{pmatrix} c \\ s' \end{pmatrix}_L$ | $\begin{pmatrix} t \\ b' \end{pmatrix}_L$ | 1/2 | $+1/2$<br>$-1/2$ | $+2/3$<br>$-1/3$ | $+1/3$<br>$+1/3$ |
| | $u_R$<br>$d'_R$ | $c_R$<br>$s'_R$ | $t_R$<br>$b'_R$ | 0<br>0 | 0<br>0 | $+2/3$<br>$-1/3$ | $+4/3$<br>$-2/3$ |

Table 1.2: Electroweak fermionic multiplets. The left-handed weak isospin doublets are provided in parentheses while the right-handed singlets form separated states. As a convention, left-handed quarks are the strong interaction $u$, $c$ and $t$ eigenstates as well as the rotated, under the CKM matrix transformation, $d'$, $s'$ and $b'$ eigenstates. The electric charges Q for states of the same doublet differs by a unit; the difference Q-$I_3$ = $Y_W/2$ is the same inside each doublet (-1/2 for left-handed leptons and +1/6 for quarks).

The lagrangian associated to the free propagation of the weak and electromagnetic fields can be thus written as the sum of the one related to the scalar U(1) field $B_\mu$ and the three SU(2) vector fields $W^i_\mu$:

$$\mathcal{L} = -\frac{1}{4} F^i_{W\mu\nu} F^{i\mu\nu}_W - \frac{1}{4} F_{B\mu\nu} F^{\mu\nu}_B \tag{1.24}$$

where the field tensors are defined as

$$F^i_{W\mu\nu} = \partial_\mu W^i_\nu - \partial_\nu W^i_\mu + g\epsilon^{ijk} W_{j\mu} W_{k\nu} \tag{1.25}$$

$$F_{B\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu. \tag{1.26}$$

In the Standard Model $\vec{W}$ is a triplet; the two charged states combine as

follows:

$$W_\mu^\pm = \frac{1}{2}(W_\mu^1 \mp W_\mu^2) \tag{1.27}$$

while the neutral component $W_\mu^3$ mixes with the field $B_\mu$ thus forming the electromagnetic field $A_\mu$ and the field associated to the neutral boson $Z_\mu$. In fact, they can be written as:

$$A_\mu = B_\mu \cos\theta_w + W_\mu^3 \sin\theta_w \tag{1.28}$$

$$Z_\mu = -B_\mu \sin\theta_w + W_\mu^3 \cos\theta_w \tag{1.29}$$

where $\theta_w$ is the Weinberg angle. The value of this angle is $\sin^2\theta_w = 0.23108 \pm 0.00005$ [4], which corresponds to an angle of 30°, and it is obtained by parity violation experiments. This angle, as it is shown in figure 1.3, is also related to the coupling costants of the Standard Model by the relation:

$$\tan\theta_w = \frac{g'}{g} \tag{1.30}$$

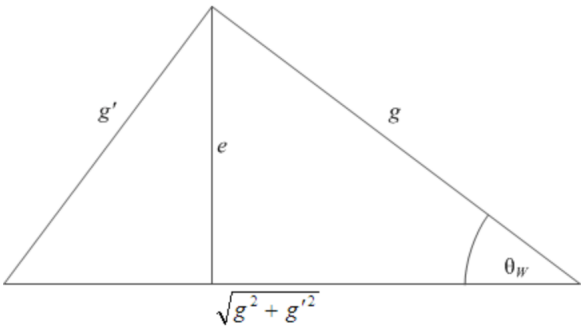



Figure 1.3: Diagram of the relation between the Weinberg angle and the SM coupling costants. The coupling constants $g$ and $g'$ form two cathets of a right-angled triangle, while the electric charge $e$ is the height. One of the two acute angles (on the right) is the Weinberg angle $\theta_w$.

where $g$ and $g'$ are respectively the weak isospin and the hypercharge couplings. In order to make this equation invariant under the gauge transformations of the group SU(2)⊗U(1), it is necessary to introduce also the covariant derivative which, in its most general form, is

$$D_\mu = \partial_\mu + iY\frac{g'}{2}B_\mu + ig\frac{\tau^i}{2}W_\mu^i \tag{1.31}$$

where Y is the hypercharge and $\tau^i$ are the Pauli's matrices. The Dirac term of the lagrangian for free fermions, using the definition 1.31, for the group SU(2)⊗U(1) becomes

$$\mathcal{L}_{fermions} = i[\bar{\psi}_L(\not{\partial} + ig\frac{\tau^i}{2}\gamma^\mu W^i_\mu + i\frac{g'}{2}Y\gamma^\mu B_\mu)\psi_L + \bar{\psi}_R(\not{\partial} + i\frac{g'}{2}Y\gamma^\mu B_\mu)\psi_R]. \tag{1.32}$$

Finally, the electroweak lagrangian which describes massless bosons and fermions can be written as the sum of the free field lagrangian and of the fermionic term (1.32):

$$\mathcal{L}_{EW} = \mathcal{L} + \mathcal{L}_{fermions}. \tag{1.33}$$

However, as will be shown in the next chapter, the terms corresponding to the masses of bosons and fermions are still missing in this lagrangian. They will appear into the equation thanks to the mechanism of spontaneous symmetry breaking, also known as the Higgs mechanism.

## 1.4 The Higgs Mechanism

Up to now, as it has been shown in paragraph 1.3.4, gauge bosons have to be massless as well as the photon in QED. This theorical assumption, which is necessary to preserve the symmetry of the theory, with respect to gauge transformations, does not explain why weak interactions are short-ranged and impossible to observe outside the nucleus. Fermi himself, in its theory of weak interactions, suggested that the weak interaction potential should have been of Yukawa form, implying massive mediators.
Fermi's hypothesis was confirmed by the discovery of the two weak vector bosons by the UA1 [8] and UA2[9] Collaborations at CERN in 1983. As a matter of fact, mass term does not appear in the lagrangian as it is has been constructed in the previous section. In the lagrangian, terms as

$$M_W^2 W^\mu W_\mu; \qquad M_Z^2 Z^\mu Z_\mu; \qquad -M_l \bar{l} l \tag{1.34}$$

are expected.
They, however, cannot surely be added by hand since they will spoil the local gauge SU(2)⊗U(1) invariance of the lagrangian. The only possibility for the preservation of the gauge invariance is a mechanism thanks to which these terms will naturally appear as a result of the spontaneous symmetry breaking. Through this theoretical device, the photon will continue to be massless, while the three vector bosons responsible for the weak interactions will acquire mass. The spontaneous symmetry breaking will thus break the symmetry for the group SU(2) and U(1), preserving the local symmetry for

$U(1)_{em}$. In order to understand better the Higgs mechanism [10], it is necessary to explain what is a broken symmetry.
Let's consider a lagrangian $\mathcal{L}$, for a certain system, and its symmetries. If the system has a non-degenerate energy level, the energy eigenstate will be unique and invariant for $\mathcal{L}$ symmetries. On the contrary, if that level is degenerate, the correspondent eigenstates are not invariant under those transformations. They will, in fact, mix with each other.
In every system, the lowest level of energy is the ground state. If it is non-degenerate it will possess the same symmetries of the lagrangian. On the contrary, if the ground state is degenerate, there will not be only one eigenstate for the rapresentation of the lowest energy level. Any of the degenerate states of the fundamental level will possess the symmetries of the lagrangian anymore. The realization of an asymmetric state is known as spontaneous symmetry breaking. The asymmetry is obtained not by adding further terms in the lagrangian such as 1.34, but by the arbitrary choice of a degenerate state.
The vacuum state, in quantum field theory, is the one with the lowest energy. Applying the spontaneous symmetry breaking to the vacuum state means to consider this state as degenerate. This degeneracy must be created by a potential that, in case of Higgs mechanism, has the form:

$$V(\phi) = \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2 = \mu^2 \mid \phi^2 \mid + \lambda \mid \phi \mid^4 \tag{1.35}$$

where $\phi$ is a complex scalar field that can be written as:

$$\phi(x) = \frac{1}{\sqrt{2}}(\phi_1(x) + i\phi_2(x)) \tag{1.36}$$

with $\mu$ and $\lambda$ arbitrary real parameters and $\lambda > 0$. The behaviour of the Higgs potential is illustrated in figure 1.4 as a function of the scalar field $\phi$ in $\mu^2 > 0$ (left) case and $\mu^2 < 0$ (right) case.

The structure of the Higgs potential when $\mu^2 < 0$ is shown in more details, in a three dimentional rapresentation, in figure 1.5. When $\mu^2 > 0$ the potential has only one absolute minimum in $\phi(x) = 0$. In order to have a degenerate vacuum state it is thus necessary to consider the case in which $\mu^2 < 0$. As shown in figure 1.5, the potential has a relative maximum in $\phi(x) = 0$ and a full circumference of minima in

$$\phi(x) = \phi_0\sqrt{\frac{-\mu^2}{2\lambda}}e^{i\theta} \tag{1.37}$$

with $0 < \theta < 2\pi$ and constant $\phi_0$. The phase of rotation indicates one of the possible directions of the complex plan $\phi$. Choosing a specific direction

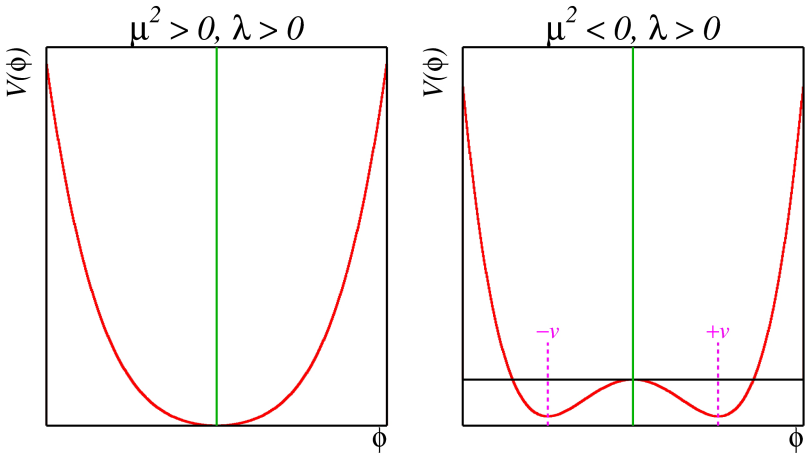



Figure 1.4: Higgs potentials V($\phi$) as a function of the complex scalar field $\phi$ in $\mu^2 > 0$ (left) and $\mu^2 < 0$ (right) cases.

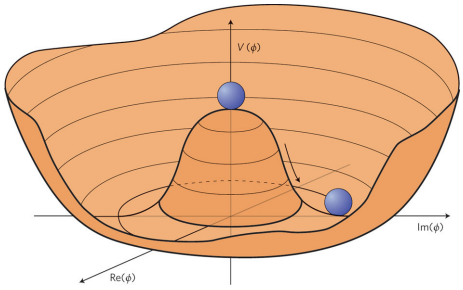



Figure 1.5: Higgs potential as a function of the complex scalar field $\phi$ in $\mu^2 < 0$ case which is responsible for the spontaneous symmetry breaking.

$\phi$ means to operate a spontaneous symmetry breaking for the rapresentation of the vacuum state. Without loss of generality, it is possible to choose $\theta = 0$ and thus equation 1.37 becomes :

$$\phi(x) = \phi_0 \sqrt{\frac{-\mu^2}{2\lambda}} \equiv \phi_0 \frac{v}{\sqrt{2}} \qquad \text{with} \qquad v \equiv \sqrt{\frac{-\mu^2}{\lambda}}. \tag{1.38}$$

It is now possible to apply the Higgs mechanism to the SU(2)⊗U(1) group by defining a complex doublet of SU(2) isospin. The doublet has Y= 1 hypercharge and 1/2 up and -1/2 down components of isospin:

$$\Phi \equiv \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}$$

where $\phi^+$, $\phi^0$ are defined in terms of four scalar fields $\phi_i$ such as:

$$\phi^+ = \frac{1}{\sqrt{2}}(\phi_1(x) + i\phi_2(x)) \tag{1.39}$$

$$\phi^0 = \frac{1}{\sqrt{2}}(\phi_3(x) + i\phi_4(x)). \tag{1.40}$$

Using these fields, the Higgs lagrangian becomes [5]

$$\mathcal{L}_{Higgs} = (D_\mu \Phi)^\dagger (D^\mu \Phi) - V(\phi) + \mathcal{L}_{gauge} \tag{1.41}$$

where the expression for $D_\mu$ is the same as the covariant derivative of the GWS model. The ground state can be written as:

$$\mid \Phi_0 \mid^2 = \frac{-\mu^2}{2\lambda} \equiv \frac{v^2}{2} \tag{1.42}$$

and it is also possible to choose the four scalar fields $\phi_i$ in order to create the spinor $\Phi$

$$\Phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1(x) + i\phi_2(x) \\ \phi_3(x) + i\phi_4(x) \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}.$$

A unitary gauge for the group SU(2)⊗U(1) has been chosen, which leaves only the third scalar field $\phi_3 = v + h(x)$ with $T_3 = -\frac{1}{2}$; all the other fields are null. The field $\Phi$ in its fundamental state $\Phi_0$ brokes the symmetry of the Higgs lagrangian $\mathcal{L}_{Higgs}$ for the SU(2)⊗U(1) group. The choice of this ground state keeps the local invariance for the $U(1)_{em}$ group:

$$\Phi(x) \to \Phi(x)' = e^{iQ_{em}\alpha(x)}\Phi(x) \tag{1.43}$$

when $Q_{em} = 0$, $\Phi(x)' = \Phi(x)$. Finally, the Higgs lagrangian can be written as [5]:

$$\mathcal{L}_{Higgs} = \frac{1}{2}\partial_\mu h(x)\partial^\mu h(x) + \mu^2 h(x)^2 \to H \tag{1.44}$$

$$-\frac{1}{4}W^1_{\mu\nu}W^{1\mu\nu} + \frac{1}{2}\left(\frac{g^2v^2}{4}\right)W^1_\mu W^{1\mu} \to W^+ \tag{1.45}$$

$$-\frac{1}{4}W^2_{\mu\nu}W^{2\mu\nu} + \frac{1}{2}\left(\frac{g^2v^2}{4}\right)W^2_\mu W^{2\mu} \to W^- \tag{1.46}$$

$$-\frac{1}{4}Z_{\mu\nu}Z^{\mu\nu} + \frac{1}{2}\left(\frac{g^2v^2}{4\cos^2\theta_w}\right)Z_\mu Z^\mu \to Z \tag{1.47}$$

$$-\frac{1}{4}A_{\mu\nu}A^{\mu\nu} \to \gamma \tag{1.48}$$

From this equation, the W and Z bosons acquire mass defined as:

$$M_W = \frac{1}{2}vg \qquad\qquad M_Z = \frac{1}{2}v\sqrt{g^2 + g'^2} \tag{1.49}$$

but the photon remains massless since there is no term in equation 1.48 which is proportional to $A_\mu A^\mu$. Furthermore, there is a new scalar boson which completes the frame of the Standard Model: it is the Higgs boson and its mass can be written as:

$$M_H = \sqrt{2\lambda v^2}. \tag{1.50}$$

### 1.4.1 The Free Parameters of the Theory and the Masses of the Bosons

The Higgs mechanism provides a solution to the problem of the boson masses. On the other hand, this theoretical model also presents many free parameters, which are not constrained by theoretical considerations and thus have to be discovered by particle physics experiments.
As it has already been shown in section 1.3.4, in the GWS theory the coupling costants $g$ and $g'$ are connected to the electromagnetic one $e$ by means of the Weinberg angle $\theta_w$:

$$e = g\sin\theta_w = g'\cos\theta_w. \tag{1.51}$$

The boson masses are also connected to the weak and strong coupling costants, as visible in figure 1.49. The Weinberg angle can be consequently written as:

$$\cos\theta_w = \frac{g}{\sqrt{g^2 + g'^2}} \qquad and \qquad \sin\theta_w = \frac{g'}{\sqrt{g^2 + g'^2}} \tag{1.52}$$

and finally a relation between the bosons masses and the Weinberg angle arises

$$\frac{M_W}{M_Z} = \cos\theta_w. \tag{1.53}$$

The Weinberg angle, which has been considered as a free parameter of the theory, is thus connected to two other quantities which are directly measurable: $M_Z$ and $M_W^\pm$. A direct measure of $\theta_w$ represents a proof to the GWS model. Nevertheless, the different masses of the two weak gauge bosons point out that $\theta_w$ must be different from zero and thus responsible for the mixing of the $B_\mu$ and $W_\mu^\pm$ fields.
The masses of the bosons can also be described using the Fermi's costant, since the weak coupling costant $g$ is written as:

$$g^2 = \frac{8G_F M_W^2}{\sqrt{2}} \tag{1.54}$$

in the *Intermediate Vector Boson* theory, which first tried to describe the weak interaction through bosonic propagators. The important result of this equation is the possibility to obtain a prevision for the vacuum expectation value of the Higgs field by writing:

$$v^2 = \frac{1}{\sqrt{2}G_F} \simeq 246\, GeV^2 \tag{1.55}$$

known the value of the Fermi's costant, which is measured by experiments on the muon mean life. Furthermore, as the masses of the bosons can be written as:

$$M_Z = \left(\frac{\alpha\pi}{G_F\sqrt{2}}\right)^{\frac{1}{2}} \frac{1}{\sin\theta_w} \quad and \quad M_Z = \left(\frac{\alpha\pi}{G_F\sqrt{2}}\right)^{\frac{1}{2}} \frac{2}{\sin 2\theta_w} \tag{1.56}$$

a prevision for the range of masses is possible in order to experimentally discover the two gauge bosons. This theoretical prevision was later found to be in agreement with the experimental data provided by the UA1 and UA2 Collaboration at CERN which declared the discovery of W and Z bosons in 1983. Unfortunately, the value of $\lambda$, which appears in the expression for the Higgs mass in 1.50, is not constrained by theorical considerations.

### 1.4.2 The Yukawa Coupling Terms for Fermions

The values of the Yukawa couplings of fermions with the Higgs boson $g^f$ can be measured known $v^2$ and the value of the fermion masses, which can be evaluated with high precision. Same considerations are applied to neutrinos and quark coupling constants [7]. According to the GWS model, illustrated in paragraph 1.3.4, fermions, as well as gauge bosons, should be massless. However, as it has just been demonstrated, the masses of the bosons come up by their direct coupling with the Higgs particle. In a similar way, fermions acquire mass, without gauge symmetry violation, by the coupling with the Higgs boson, as shown in figure 1.6.
The masses are thus introduced in invariant terms, which are the *Yukawa couplings* with the Higgs field. These terms couple the fermionic doublet $\psi_L^f$ and singlet $\psi_L^f$ with the Higgs field $\Phi$ through a coupling costant $g^f$. The lagrangian for the Yukawa term is:

$$\mathcal{L}^f_{Yukawa} = -g^f[\bar{\psi}^f_L\phi\Psi^f_R + \bar{\psi}^f_R\phi\Psi^f_L] \tag{1.57}$$

$$= -\frac{g^f v}{\sqrt{2}}(\bar{\psi}^f_L\psi^f_R + \bar{\psi}^f_R\psi^f_L) - \frac{g^f}{\sqrt{2}}(\bar{\psi}^f_L\psi^f_R + \bar{\psi}^f_R\psi^f_L)h(x) \tag{1.58}$$

$$= -M^f\bar{\psi}^f\psi^f - \frac{g^f}{\sqrt{2}}\bar{\psi}^f\psi^f h(x) \tag{1.59}$$

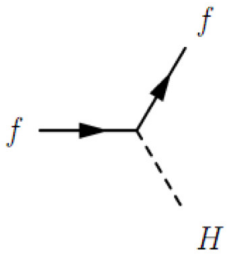



Figure 1.6: Feynman diagram of the direct coupling of the Higgs boson with a fermion of the Standard Model.

and the mass of the fermion is thus given by:

$$M^f = \frac{g^f v}{\sqrt{2}}. \tag{1.60}$$

Moreover, the second term in 1.58 describes the interaction of any fermion-antifermion couple with the Higgs boson. It represents a vertex term of the type:

$$H\bar{f}f \rightarrow \frac{1}{2} ig \frac{M^f}{M_W} \tag{1.61}$$

which is proportional to the masses of fermions.
The same scenario happens for quarks, with the difference that the mass eigenstates are not the same of the interaction eigenstates: only these lasts couple with the Higgs boson. Consequently, the Yukawa lagrangian for quarks has to be written as [5]:

$$\mathcal{L}^q_{Yukawa} = -g^d_{ij}(\bar{u}_i \bar{d'}_i)_L \Phi d_{JR} - g^u_{ij}(\bar{u}_i \bar{d'}_i)_L \Phi^\dagger u_{JR} + h.c. \tag{1.62}$$

where $u_i$ and $d'_i$ are the generic weak eigenstates for the three quark families.

## 1.5 The Discovery of the Higgs Boson: ATLAS and CMS

On 4 July 2012 the two LHC experiments at CERN, ATLAS [11] and CMS [12], announced the discovery of a new particle in the mass region of 126 GeV. It was consistent with the Higgs boson, but further tests have to be taken into account. On 8 October 2013 Peter Higgs and Françoise Englert were awarded with the Nobel Prize "*for the theoretical discovery of a mechanism*

*that contributes to our understanding of the origin of mass of subatomic particles, and which recently was confirmed through the discovery of the predicted fundamental particle, by the ATLAS and CMS experiments at CERN's Large Hadron Collider."* ATLAS and CMS performed their researches with similar integrated luminosity. The recorded luminosities for both experiments, referred to 2011 and 2012 periods, with which the measurement has been performed are presented in table 1.3. They both observed an excess of mass in different Higgs decay channels combining the measurements of the 2011 and 2012 data taking periods. They analized data from $H \to \gamma\gamma$, $H \to \tau\tau$, $H \to WW^*$, $H \to ZZ^*$ and $H \to b\bar{b}$ decays. In figure 1.7 ATLAS and CMS results are shown. On the left the ATLAS experiment distribution of the four-lepton invariant mass, $m_{4l}$, compared to the background expectation in the 80-250 GeV mass region for the combined $\sqrt{s}$=7 TeV and $\sqrt{s}$=8 TeV data [11] is presented for the $H \to ZZ^*$ decay channel . The expectation for a SM Higgs boson at 125 GeV is shown. On the right, the CMS diphoton invariant mass distribution in the $H \to \gamma\gamma$ decay channel [12] is shown. In both cases, the red lines represent the best fit for signal and background. Final values of Higgs mass, as well as its significance, is reported in table 1.3.

Thanks to the excellent momentum resolution of the ATLAS detector,

| | ATLAS | CMS |
|---|---|---|
| $\mathcal{L}_{int}$ in 2011 ($\sqrt{s}$=7 TeV) | 4.8 $fb^{-1}$ | 5.1 $fb^{-1}$ |
| $\mathcal{L}_{int}$ in 2012 ($\sqrt{s}$=8 TeV) | 5.8 $fb^{-1}$ | 5.3 $fb^{-1}$ |
| Most significant Channel | $H \to ZZ^*$ | $H \to \gamma\gamma$ |
| Measured mass (GeV) | 126 $\pm$0.4(stat)$\pm$0.4(syst) | 125.3 $\pm$0.4(stat)$\pm$0.5(syst) |
| Significance | 5.9 $\sigma$ | 5.0 $\sigma$ |

Table 1.3: The integrated luminosities, for the 2011 and 2012 periods, with which the measurement has been performed and the most relevant decay channel that brings to the Higgs boson discovery are listed with the final resulting mass and its significance for both the experiments.

searches for a SM Higgs boson in the $H \to ZZ^*$ decay channel provide good sensitivity over a wide mass range (110-600 GeV). The candidate selection requires a pair of isolated leptons (muon or electron) with opposite charge and same flavour. The expected cross section times branching ratio for the $H \to ZZ^* \to 4l$ process with an Higgs boson expected mass of $m_H$=125 GeV is of 2.2 fb for $\sqrt{s}$=7 TeV and 2.8 fb for $\sqrt{s}$=8 TeV. The main background contributions come from ZZ*, Z+*jets* and $t\bar{t}$ production, where charged lepton candidates arise either from the decay of heavy flavour hadrons or from the misidentification of jets. These background contributions are estimated by using MC simulations for $ZZ^*$ normalized to the theoretical cross section

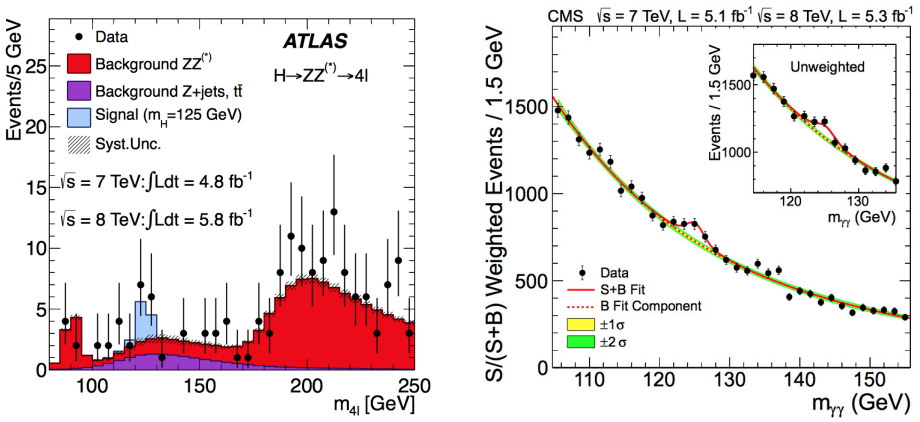



Figure 1.7: Left: ATLAS experiment distribution of the four-lepton invariant mass, $m_{4l}$, compared to the background expectation in the 80-250 GeV mass region for the combined $\sqrt{s}$=7 TeV and $\sqrt{s}$=8 TeV data in the $H \to ZZ^*$ decay channel [11]. The expectation for a SM Higgs boson at 125 GeV is shown. Right: CMS diphoton invariant mass in the $H \to \gamma\gamma$ decay channel [12]. The red lines represent the best fit for signal and background.

for this production mode. For the Z+jets and $t\bar{t}$ backgrounds, methods using control regions are implemented.
The search for a peak corresponding to the Higgs boson, performed by the CMS Collaboration in the $H \to \gamma\gamma$ decay channel, focuses on the narrow mass range of 110-150 GeV. The photon pair candidates must lie in the fiducial region of the detector ( $|\eta|$<2.5 and 1.44<$|\eta|$<1.57) and have to satisfy kinematic selection criteria: $p_T > m_{\gamma\gamma}/3$ or $p_T > m_{\gamma\gamma}/4$ respectively for the leading and the subleading photons. Multivariate techniques are applied for selection and classification of the events. The main background to this type of signal is the QCD production of two photons. There is also a background due to the presence of one or more photons that originate from the misreconstruction of jet fragments. It took almost 50 years for the Higgs boson to be discoverd from its first theorization in 1964, but the observation of this particle is a milestone in particle physics. Further studies, especially during the Run II of LHC, at a center of mass energy of $\sqrt{s}$=14 TeV, will increase the knowledge on this, and maybe other, particles in nature.

## 1.6 Beyond the Standard Model

The Standard Model, as mentioned, provides a theoretical frame of the high energy physics world and describes, with high precision, the experimental results of particle physics. In fact, up to the energy scale of the weak inte-

raction (few hundreds of GeV) data and theory are in agreement. However, there are some experimental, as well as theoretical, hints that new physics could lie at the TeV energy scale.

The search for a quantum field theory aiming at unifying gravity with the other three known forces has been the main subject of several researches for many years. This theory is called the *Grand Unified Theory* (GUT) [13] which assumes the same intensity for the coupling costants up to an energy scale of $10^{16}$ GeV. In fact the Standard Model does not take into account the phenomena which take places at low energy, or big distances, which are subject of the gravity interaction. Furthermore, the graviton, which represents the gauge boson mediator for gravitation has not been yet observed leading to the belief that gravity cannot just be naively added to the Standard Model.

### The Hierarchy Problem

The process of spontaneous symmetry breaking through which particles acquire their mass, as illustrated in paragraph 1.4, has been executed at the *tree-level* contribution. Actually, radiative corrections to the mass of a particle should be computed when one-loop corrections to the propagator are considered [13]: when a fermion couples with a scalar particle, corrections to the fermionic mass should be computed, considering one-loop contributions to the fermion propagator due to the scalar particle, as illustrated in figure 1.8 for a fermionic loop on a scalar particle (right) and a scalar loop on a fermionic particle (left).

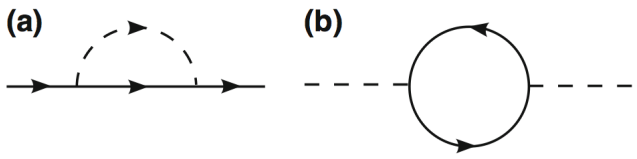



Figure 1.8: (a): Example of a fermionic loop on a scalar particle. (b) Illustration of a scalar loop on a fermionic particle.

The mass of the fermion should be rewritten as:

$$m_f = m_f^{(0)} + \delta m_f \tag{1.63}$$

where $m_f^{(0)}$ is given by the Higgs mechanism and the adding term

$$\delta m_f = -\frac{3\lambda_f^2 m_f}{64\pi^2} \log \frac{\Lambda^2}{m_f^2} + .. \tag{1.64}$$

depends on the introduction of a *cut-off* to the theory $\Lambda$. Addictional terms in equation 1.64 that diverge when $\Lambda \to \infty$ or are independent on $\Lambda$ are ignored. As well as for fermions, also the scalar particle mass should be corrected in order to account for these scalar-fermion loop terms. The correction for the scalar particle is thus:

$$\delta m_s = \frac{\lambda_f^2}{8\pi^2} \left[ \Lambda^2 - 6m_f^2 \log \frac{\Lambda}{m_f} + 2m_f^2 + .. \right] \tag{1.65}$$

and is quadratically divergent with the cut-off $\Lambda$. For this reason, nothing can actually protect the mass of the scalar particle of having very large corrections. Assuming the theory to be valid until the Planck scale, i.e. scale at which quantum corrections to gravity become important, corrections to the scalar particle will be of such order of magnitude. However, the mass of the particle and its corrections must be of the same order of magnitude, otherwise the theory parameters have to be adjusted (*fine tuned*) in order to agree with observations. For the Higgs boson the same problem occurs. Since unitarity requires this particle to have a mass smaller than 1 TeV, quadratic divergences would indicate a large problem, considering corrections would have a much larger order of magnitude than the mass at lowest order. A counter term to cancel these quadratic divergences could solve the problem. However, such large cancellations would be highly un-natural. Conversely, the cutoff should be adjusted down to 1 TeV. The result would be a theory that ceases to being valid at this energy scale. This is known as the *hierarchy problem* and it is one of the reasons for which it is believed that new physics could indeed lie above the TeV scale.

### CP Violation

CP symmetry is the combination of charge conjugation and parity symmetries. This transformation states that nature shoul be invariant when replacing the particle with its antiparticle and with the substitution of left-handed particles with right-handed ones. Although both of these symmetries are respected by the electromagnetic force, CP is violated by weak interactions. CP violation was first observed in 1964 by Cronin and Fitch [14][15] in $K^0$ oscillations and then confirmed by experiments with B mesons. On the strong interaction side, the QCD lagrangian also contains a term that could

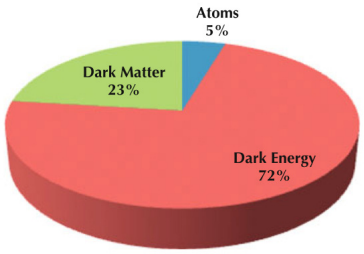



Figure 1.9: WMAP satellite data shows that the content of the present Universe is formed by approximately 5% of baryonic matter, 23% of dark matter and 72% by dark energy. Dark energy should be different from dark matter and responsible for the experimentally observed acceleration of the Universe.

be responsible for CP violation. Since events which involve strong interactions and violate CP have not been yet observed, this term is set to zero. Also this fine tuning is considered un-natural and this is know as the *strong CP problem*.

### Dark Matter

A different indication of new physics at the TeV scale would be the evidence of dark matter. Dark matter is defined to be an undetermined type of matter that does not emit or reflect electromagnetic radiation. In fact, the subatomic composition of this new type of matter is not yet known, neither its interactions. Nowadays, several observations have confirmed dark matter evidence, for example analysis combining galaxy cluster dynamics, supernova data and *Cosmic Microwave Background* (CMB) radiation with Big Bang nucleosyntesis. Moreover, precise measurements have been performed and recently confirmed the existence of dark matter at larger scales than the size of galaxies and clusters. Approximately 5% of the known matter in the Universe is believed to be formed of atoms, while 23% of dark matter and the rest of dark energy, as shown in figure 1.9 from data obtained by the WMAP satellite. In principle, dark matter could be formed of either baryonic, non-baryonic content or by a combination of both. Nevertheless, there are reasons to believe that dark matter could not consist of only baryonic content. Studies have shown that there is more dark matter in the Universe than the maximum number that can be calculated. In fact, the theory of Big Bang nucleosynthesis and acoustic peaks in the CMB can be used to determine the fraction of baryons in the Universe and to predict the percentage of barionic atoms in dark matter. For this reason, it is widely believed that a significant part of the dark matter content is likely to be exotic, composed by non-baryonic atoms. From the SM of particle physics, only the neutrino

would fulfill all the requirements of a non-baryonic dark matter candidate. Since it does not carry electric charge and it interacts very weakly with the particles of the SM, the cross section for neutrino production is very small. However, the mass density of neutrinos in the Universe has been calculated and is not enough to explain the fraction of matter that contributes to the cosmic average density. Besides, studies of galaxies structure affirm that neutrinos with a very small mass cannot be entirely responsible for the measured dark matter composition. As a solution, there could exist different types of neutrinos not yet observed, such as sterile neutrinos, to account for part of the dark matter content. In this case, clearly new physics is expected.
If, however, dark matter is non-baryonic and not formed by neutrinos, it should be composed by new particles, for example axions and WIMPs (*weakly interacting massive particles*). The axion is the hypothetical particle postulated by Peccei and Quinn to solve the strong CP problem that appears in QCD, as explained previously. They are supposed to be scalar particles generated from the spontaneous symmetry breaking of a new symmetry. WIMPs are massive particles, predicted by some theories Beyond the Standard Model (BSM). Examples of BSM theories are *supersymmetry* or theories with *extra dimensions*. Candidates for WIMPs should interact only via the weak or gravitational forces and not through the strong force. The *neutralino* or the *gravitino* could represent WIMPs in different variations of supersymmetric (SUSY) theories. SUSY theories [13] not only predict a stable particle to be the dark matter candidate, but also solve several other issues, for example the hierarchy problem. It is therefore one of the most studied BSM theories giving rise to different branches. Finally, WIMP candidates could also be new and unpredicted particles coupled to the SM only via gravitational interaction.

All the problems, and possible theorical solutions, that have been discussed here could indicate the existence of new constituents of fundamental particles. Moreover, an extended version of the SM would be necessary to incorporate such exotic particles whose masses could be accessible at the Large Hadron Collider energy range.

## 1.7 Main Research Fields at LHC

The Large Hadron Collider (LHC) is the highest energy particle accelerator, in operation since 2009, and equipped with four experiments for different studies. LHC might open a window on the physics beyond the Standard Model.

ATLAS and CMS, two multipurpose experiments, will be able to measure quantities such as the couplings of the Higgs boson with gauge bosons and fermions. The increasing statistics expected in Run II will also give the chance to better explore the physics of quark top which decays before hadronization (due to its heavy mass).
LHCb is a dedicated experiment which focuses on the $b$ quark physics and CP violation measurements in order to understand why our universe is composed of matter and not of antimatter.
LHC is designed to operate also in lead-lead collision mode. Collisions at LHC generate temperatures more than 100,000 times hotter than the centre of the Sun. Under these extreme conditions, which should be similar to those immediately after the Big Bang, protons and neutrons "melt", freeing the quarks from their bonds with the gluons in a *quark-gluon plasma* state. The existence of such a phase state and its properties are key issues in the theory of quantum chromodynamics, for understanding the phenomenon of confinement. The ALICE experiment studies the quark-gluon plasma in its expansion and cooling, in order to evaluate how the particles that constitute the matter of our universe today rose up.

# Chapter 2

# LHC and the ATLAS Experiment

## 2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is the largest particle accelerator in the world. Located at the border between France and Switzerland, it has a circumference of 27 km and it is located in an underground tunnel between 50 and 175 m under the surface, which previously hosted the *Large Electron-Positron Collider* (LEP). LHC is designed to accelerate beams of protons and nuclei (lead) up to a center of mass energy of 14 TeV for protons and of 2.76 TeV/nucleon, yelding a center of mass energy of 1150 TeV, for lead ions. After the shutdown, started at the beginning of 2013, during Run-II LHC will reach a center of mass energy of $\sqrt{s}$=13 TeV and an instantaneous luminosity[1] of $2\times10^{34}$ cm$^{-2}$ s$^{-1}$ [16]. An overview of the machine performance during the 2010-2012 period is presented in table 2.1 [17].

The behaviour of the luminosity delivered to the ATLAS detector during stable *pp* beam collisions is shown in figure 2.1 for 2010 (green), 2011 (red) and 2012 (blue) runnings as a function of time. The maximum mean number of events per beam crossing[2] $\mu$, corresponding to the mean of the poisson distribution on the number of interactions per crossing, is shown in figure 2.2 as a function of time, during *pp* collisions in 2010, 2011 and 2012. In *pp*

[1]The number of interaction in every bunch crossing is proportional the machine luminosity $\mathcal{L}$, according to the relation $R = \sigma\mathcal{L}$, where R is the crossing rate and $\sigma$ the total cross section for a particular process.

[2]It is calculated from the instantaneous luminosity per bunch as $\mu = \mathcal{L}_{bunch}\sigma_{inel}/f_r$, where $\mathcal{L}_{bunch}$ is the per bunch instantaneous luminosity, $\sigma_{inel}$ the inelastic cross section and $f_r$ is the LHC revolution frequency.

| Parameter | 2010 | 2011 | 2012 | Design |
|---|---|---|---|---|
| Beam energy (TeV) | 3.5 | 3.5 | 4 | 7 |
| Bunch spacing (ns) | 150 | 75-50 | 50 | 25 |
| Max number of bunches | 368 | 1380 | 1380 | 2808 |
| Protons per bunch | $1.2\times10^{11}$ | $1.45\times10^{11}$ | $1.7\times10^{11}$ | $1.15\times10^{11}$ |
| Peak luminosity ($cm^{-2}s^{-1}$) | $2.1\times10^{32}$ | $3.7\times10^{33}$ | $7.7\times10^{33}$ | $1\times10^{34}$ |
| Pile up | 4 | 17 | 37 | 19 |

Table 2.1: Overview of performance-related parameters during LHC operations in 2010-2012 [17].

collision mode, two proton beams circulate in opposite directions and collide every 25 ns. They travel into two separate ultrahigh vacuum chambers at a pressure of $10^{-10}$ Torr. In order to keep the beams into circular trajectors, 1232 superconducting dipole magnets generate a magnetic field of 8.4 T at a current of 11.85 kA and a temperature of 1.9 K. Other 392 superconducting quadrupole magnets produce a field of 6.8 T to focalize the beams.
The beams are not continuous but have a minimum time separation of 25 ns, compatible with the Data Acquisition limits of the experiments. Each bunch contains up to $10^{11}$ protons for an overall lenght of 7.55 cm and a transverse dimension of few mm, which is reduced to 16 $\mu m$ at the interaction point. In every collision, elastic and inelastic collisions prevent the interacting protons from continuing to circulate in the beam pipe in phase with the original bunches. A resulting effect of these collisions is that the beam luminosity decreases with time following an exponential behaviour:

$$\mathcal{L} = \mathcal{L}_0 e^{\frac{-t}{\tau}} \tag{2.1}$$

where the time costant $\tau$ is $\sim$15 h. The beam can thus circulate for hours without requiring a refill. The luminosity is measured by dedicated luminosity detectors, among them is LUCID (*LUminosity measurement Using a Cherenkov Integrating Detector*) which is going to be described later.

### 2.1.1 The Acceleration Chain

In order to reach the designed center of mass energy of 14 TeV, protons will be accelerated by a chain of subsequent accelerators as shown in figure 2.3. These accelerators are:

- **Linac2**: is a linear accelerator for protons and ions up to an energy of 50 MeV. Particles are injected in the following accelerator with a rate

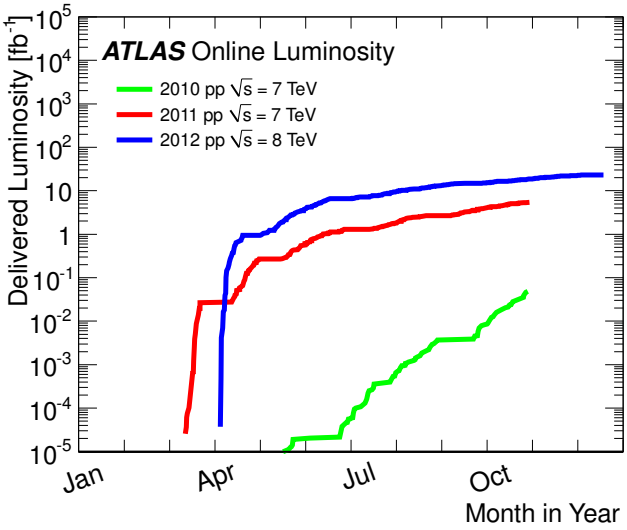



Figure 2.1: Delivered luminosity recorded by the ATLAS detector during stable beams for $pp$ collisions during 2010 (green), 2011 (red) and 2012 (blue) as a function of month in year. The online luminosity is shown.

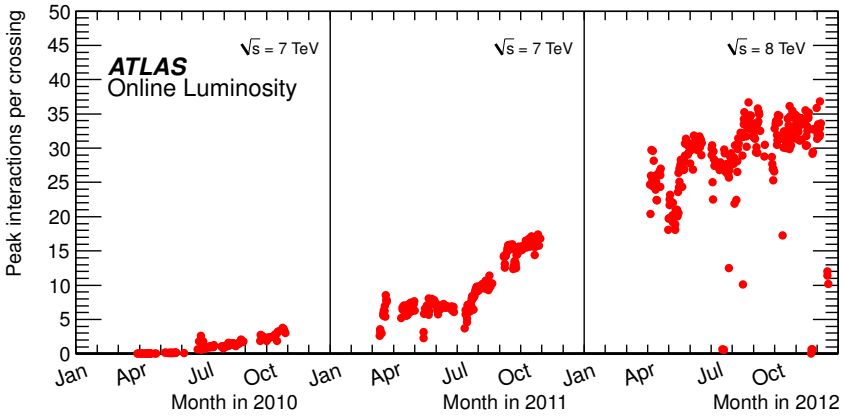



Figure 2.2: Maximum mean number of events per beam crossing versus day during the $pp$ runs of 2010, 2011 and 2012. The average value for all bunch crossing is shown in a lumi-block. The online luminosity measurement is used for this calculation. Only the maximum value during stable beam periods is shown.

of 1 Hz. The pulse duration ranges from 20 to 150 $\mu s$ depending on the number of required protons.

- **Proton Synchrotron Booster (PSB)**: the beam coming from Linac2

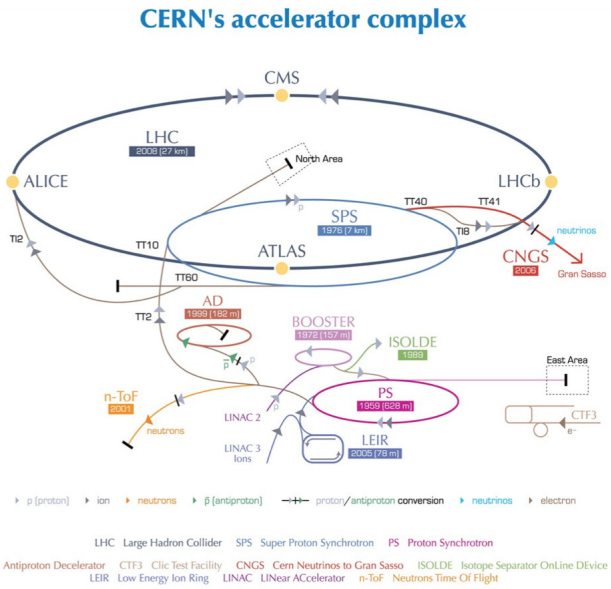



Figure 2.3: Scheme of the CERN accelerator complex.

is accelerated up to an energy of 1.4 GeV. This machine is composed of 4 superimposed rings. Five bunches circulate in each ring and they are then focused and sent through a magnet deflector into a single line for injection into the next accelerating element.

- **Proton Sincrotron (PS)**: in this accelerator protons reach an energy of 28 GeV. It is designed to separate the bunches by 25 ns.
- **Super Proton Synchrotron (SPS)**: it is the final injector for proton and heavy ions, bringing the energy up to 450 GeV.

Lastly, protons are injected in the LHC ring, where they reach a final designed energy of 7 TeV for each circulating beam.

As already said in section 1.6, LHC is equipped with four experiments installed along the tunnel:

- **A Toroidal LHC ApparatuS (ATLAS)**: it is a multi-purpose experiment working at high luminosity ($\mathcal{L} = 10^{34} cm^{-2} s^{-1}$). It is one of the LHC experiments that discovered the Higgs boson and, once reached the designed luminosity, will be able to discover signs of new physics beyond the Standard Model.

- **Compact Muon Solenoid (CMS)**: it is a multi-purpose experiment working at high luminosity. Together with ATLAS, CMS announced the discovery of the Higgs boson. It shares intents with ATLAS but by means of different technologies.

- **LHCb**: it is designed to investigate the flavour physics of the B mesons and how are they involved in CP violation. Since the production and the decay vertices of the B-mesons are more difficult to reconstruct if more than one interaction for bunch crossing is present, LHCb works at lower luminosity ($\mathcal{L} = 10^{32} cm^{-2} s^{-1}$).

- **A Large Ion Collider Experiment (ALICE)**: as illustrated in section 1.6, it is dedicated to the study of the *quark-gluon* plasma state produced in heavy ion collisions. Due to the high track density that follows from the high nucleus-nucleus cross section, ALICE works at smaller luminosity, up to $\mathcal{L} = 10^{27} cm^{-2} s^{-1}$.

### 2.1.2 LHC Detectors Requirements

In hadron-hadron collisions, interactions involve partons so that there is a non-null probability of multiple interactions in a single collision. These events are usually referred to as *underlying events* and are due to the *Initial State Radiation* and to the *Final State Radiation* [34], which are gluon radiations in the initial and final states. Furthermore, hadron collisions produce a high number of particles in the final state, which are grouped in jets. Since many of these events have also a large Lorentz boost, particles inside jets are almost collinear. Furthermore, since the cross section for jet production is dominant over the rare processes, it is important to identify experimental signatures of the rarest physics processes.
For these reasons, fine granularity is one of the most important requirements for the LHC detectors.
New physics processes can be searched focusing on different signatures, such as missing transverse energy or secondary vertices. To cope with these requirements, detectors must be able of high particle identification. In addiction to these properties, detectors must satisfy [16]:

- fast, radiation-hard electronics and sensor elements, due to the experimental conditions at the LHC. In addition, high detector granularity is needed to handle the particle fluxes and to reduce the influence of overlapping events.

- good electromagnetic calorimeter resolution for electron and photon identification. Full-coverage hadronic calorimetry for accurate jet and missing transverse energy ($E_T^{miss}$) measurements is also required;
- accurate charged particle momentum resolution and reconstruction efficiency of the inner detector. For *off-line* tagging of $\tau$ leptons and *b-jets*, vertex detectors close to the interaction region are required to be able to resolve also secondary vertices.
- good muon identification and momentum resolution over wide ranges and ability to determine unambiguously the charge of high $p_T$ muons;
- large acceptance in pseudorapidity ($\eta$) and good azimuthal angle ($\phi$) coverage;
- efficient triggering system on low transverse momentum objects with good background rejection. This is a necessary prerequisite to achieve an acceptable trigger rate for most processes of interest.

ATLAS and CMS detectors are designed to satisfy these requirements in order to perform precise measurements on SM theory and to discover new physics processes. In the next chapter the ATLAS detector will be analyzed in more detail.

## 2.2 The ATLAS Experiment

The ATLAS experiment is a multi-purpose particle detector located 100 m underground in the interaction point 1 along the LHC tunnel. The detector has cylindrical symmetry, is 44 m long and has a diameter of 22 m [16]. It can be divided into a central barrel and two end-cap regions at both sides of the detector, as shown in figure 2.4. Because of the multi-purpose nature of the experiment, ATLAS is composed of many different sub-detectors. Moving outwards from the interaction point, the originated particles encounter the Inner Detector (ID), the electromagnetic and hadronic calorimeters and finally the muon spectrometer. Moreover, ATLAS possess a magnetic system that bends the trajectory of charged particles to measure their momenta. A schematic overview of the sub-systems and their different particle identification modes is given in figure 2.5.

The coordinate system used in ATLAS is the one commonly used in all hadronic colliding experiments: the interaction point is defined as the origin of the coordinates and the z-axis runs along the beam line. The $xy$ plane is perpendicular to the beam line, with the x-axis pointing to the center of the

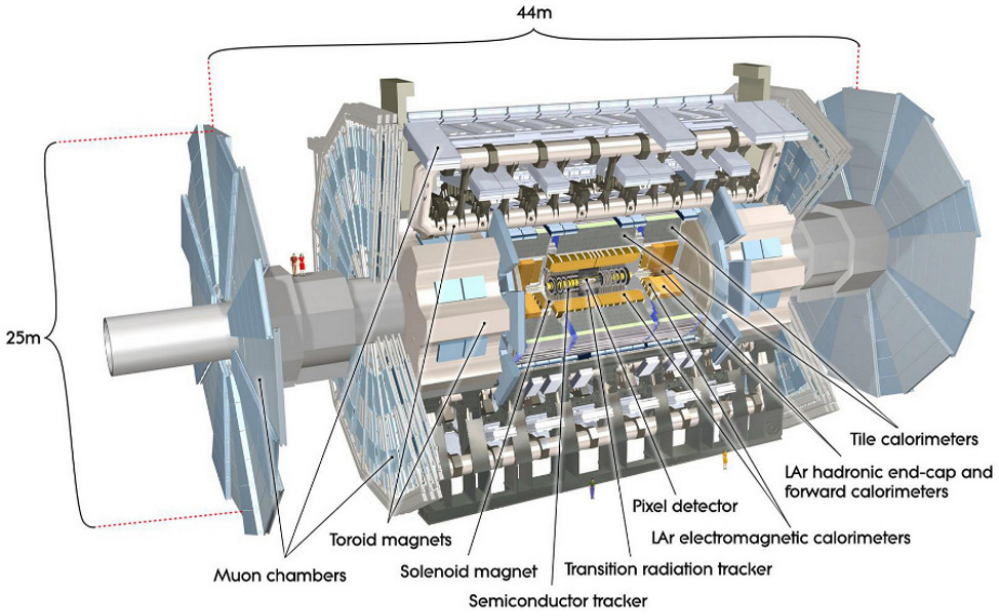



Figure 2.4: Scheme of the ATLAS detector.

ring and the y-axis pointing upwards to the surface. This plane is always referred to as "transverse plane" and it is useful for the reconstruction of kinematic quantities; for example the transverse momentum is cinematically invariant for the Lorentz boost along the z-axis and it is conserved in $xy$ plane. Partons, which interact in $pp$ collisions, carry in fact an unknown fraction of the longitudinal proton momentum. On the contrary, the fraction of the transverse momentum component is negligibly small compared to the longitudinal one, therefore an approximate conservation of momentum in the transverse plane is assumed:

$$\sum p_T \simeq 0 \tag{2.2}$$

where the transverse momentum is defined as :

$$p_T = \sqrt{p_x^2 + p_y^2}. \tag{2.3}$$

The coordinate system is often described through polar coordinates: the azimuthal angle $\phi$ is the one measured around the beam axis, while the polar angle $\theta$ is measured with respect to the beam axis. The radial coordinate $r$ is defined as the distance from the beam line.
It is useful to identify some variables which are invariant under Lorentz tran-

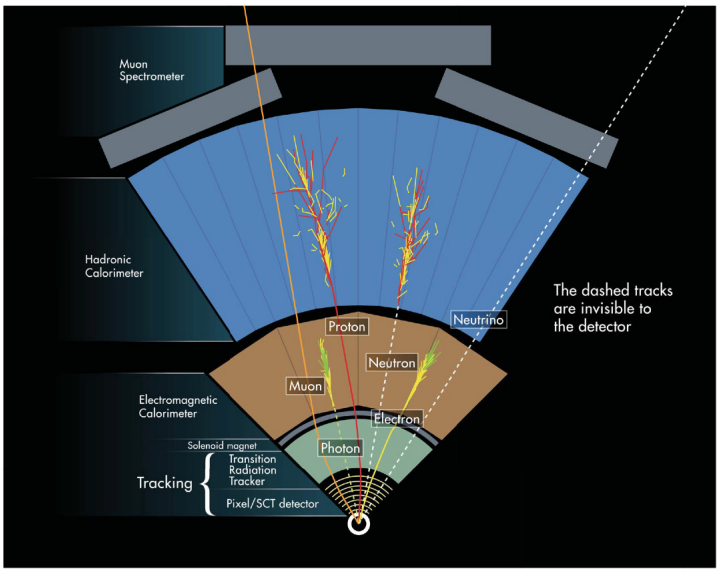



Figure 2.5: Schematic view of the layers of the ATLAS detector. For each kind of particle, the different interactions with the sub-detectors are shown.

sformations: the *rapidity*, for example, is defined as:

$$y = \frac{1}{2} \ln \left( \frac{E + p_z}{E - p_z} \right) \tag{2.4}$$

where E and $p_z$ are the energy and the z-axis momentum component of the particle. This quantity is invariant under Lorentz transformations along the z-axis. Moreover, for particles whose speed is very close to the speed of light, it also useful to define the pseudorapidity:

$$\eta = -\ln \left( \tan \frac{\theta}{2} \right) \tag{2.5}$$

in which it is not necessary to know the energy or the mass of the particle but only its angular position. From now on, particles will be measured by means of $\eta$, $\phi$ and $z$. Further, a distance measurement in the $\eta - \phi$ plane is introduced:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \tag{2.6}$$

where $\Delta\eta$ is the difference in pseudorapidity of two objects, and $\Delta\phi$ the difference between their azimuthal angles. This quantity is often used to define a cone around a reconstructed object in order to measure its isolation.

### 2.2.1 ATLAS Magnetic System

A charged particle $q$, with momentum $p$ that enters a magnetic field B, is deviated with a trajectory of radius $\rho$ such as

$$p\left[\frac{GeV}{c}\right] = 0.3 \times \rho[m] \times q \times B[Tesla]. \tag{2.7}$$

ATLAS has three superconductive magnetic field systems, as shown in figure 2.6, kept at a temperature of 4.8 K [19].

The first one is the **Central Solenoid (CS)**, a superconducting solenoid

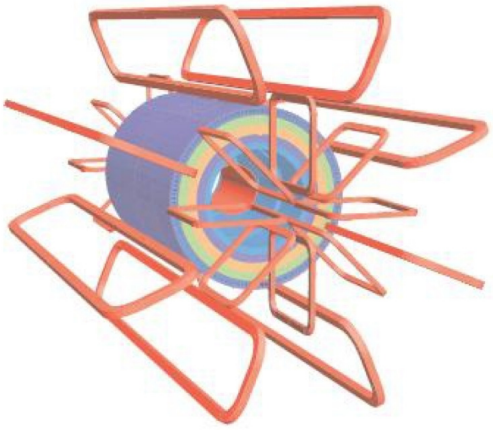


Figure 2.6: Magnetic system of the ATLAS detector.

which surrounds the Inner Detector cavity and provides a 2 T field. It has a radius of 1.2 m, lenght of 5.3 m and it is parallel to the beam axis. It is optimized in order to minimize the amount of material in front of the electromagnetic calorimeter. The **Barrel Toroid (BT)** is composed by 8 rectangular coils in a cylindrical configuration. With a total lenght of 25 m, an outer diameter of 20.1 m and inner diameter of 9.4 m, it provides a magnetic field of 1.5 T. It deviates particles in the region of $|\eta| \leq 1$. Finally, the **End-Cap Toroid (ECT)** is composed by 8 rectangular coils in a single cylindrical vessel. It has an outer diameter of 10.7 m and an inner diameter of 1.65 m. Its total lenght is 5 m. The configuration of the ECT is chosen in order the close the magnetic field lines produced by the BT. The field is thus orthogonal to the beam axis and has a value of 2 T. The ECT magnets bend the particle trajectories in the region $1.4 < |\eta| < 2.7$.

### 2.2.2 Inner Detector

The first sub-detector encountered by particles produced in the collision is the Inner Detector (ID) [20]. It is located near the beam pipe, inside the Central Solenoid, and close to the interaction point. The layout of the detector is presented in figure 2.7. It is designed to reconstruct charged particle

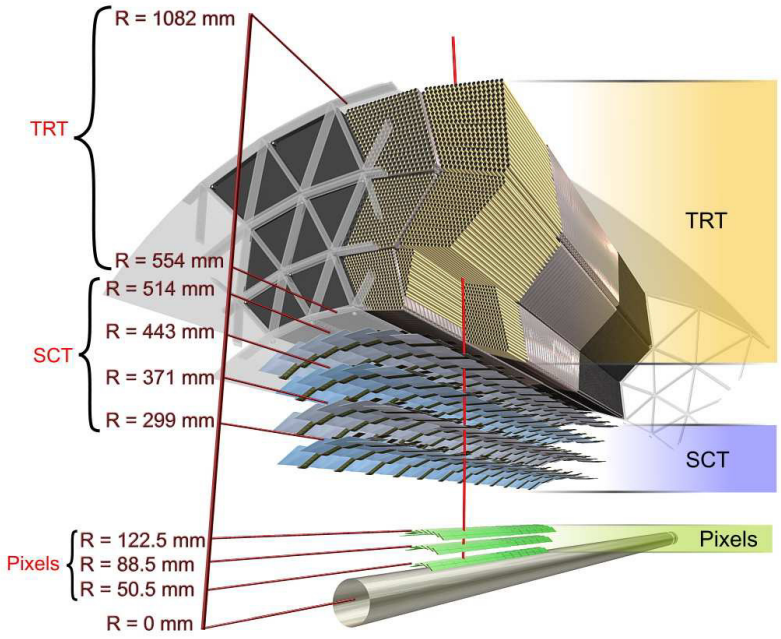



Figure 2.7: A section view of ATLAS Inner Detector: from the interaction point particles first encounter the Pixel detector, then SCT and finally the TRT detector of the ID.

tracks and their production vertex. It is decisive in the identification of long lived particles (e.g. C or B-mesons) and to distinguish electrons from other particles like photons or charged hadrons.
Given the very large track occupancy produced by LHC collisions, the granularity of the detector must be very fine in order to make high precision measurements.
The resolution of the track parameters provided by the detector can be

parametrized [21] as follows:

$$\sigma(d_0) = 11 \oplus \frac{73}{p_T\sqrt{\sin\theta}}\,(\mu m) \quad (2.8)$$

$$\sigma(\Phi_0) = 0.075 \oplus \frac{1.8}{p_T\sqrt{\sin\theta}}\,(mrad) \quad (2.9)$$

$$\sigma(z_0) = 87 \oplus \frac{115}{p_T\sqrt{\sin^3\theta}}\,(\mu rad) \quad (2.10)$$

$$\sigma(\cot\theta) = 0.7\times10^{-3} \oplus \frac{2.0\times10^{-3}}{p_T\sqrt{\sin^3\theta}} \quad (2.11)$$

$$\sigma\left(\frac{1}{p_T}\right) = 0.36 \oplus \frac{13}{p_T\sqrt{\sin\theta}}\,(TeV^{-1}) \quad (2.12)$$

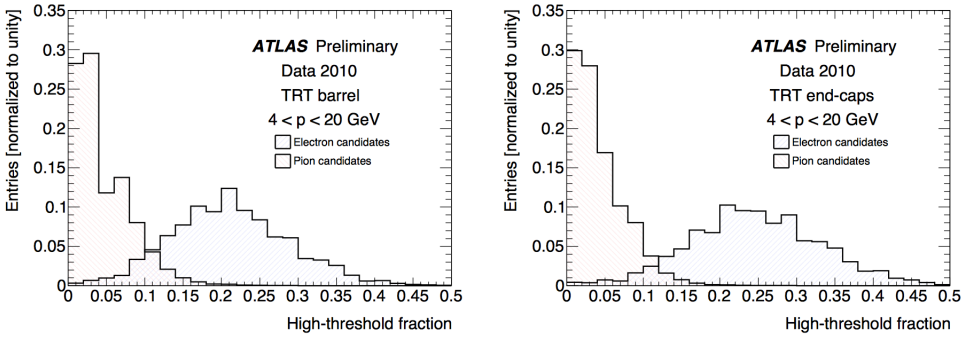
where $d_0$, $\Phi_0$, $z_0$ and $\theta$ are respectively the distance of the closest approach to the beam axis and the parameters that define the direction of the track in the transverse and longitudinal planes at the nearest point from the z-axis. The ID has cylindrical symmetry, with an outer radius of 105 cm, and covers the region up to $|\eta|<2.5$. It can be divided into three parts: a barrel section, which covers $\pm80$ cm of a concentrical region with respect to the interaction point, and two identical end-caps perpendicular to the z-axis.
From a technical point of view, high granularity is realized through different solutions.

The inner part of the ID is composed by **Pixel Detector (PD)** [22], made of a silicon wafer with 46080 semi-conductor $50\times300$ $\mu$m$^2$ pixels, followed by microstrip detectors. This PD has three layers of silicon pixels, placed at 50.5, 85.5 and 122.5 mm from the center of the detector, and five rings on each side with an inner radius of 11 cm and an outer radius of 30 cm, to complete the angular coverage. The readout of this part of the detector is made of almost 80.4 million channels.

The **Semi Conductor Tracker (SCT)** [23] is a detector designed to provide precision measurements of momentum, impact parameter and vertex position in the intermediate radial range. The SCT barrel section is made of four layers of microstrip modules placed at 299, 371, 443 and 514 mm from the center of the detector. Each silicon detector is $6.36\times6.40$ cm$^2$ large, with 768 readout strips of 80 $\mu m$ pitch. Each module is made of four single-sided p-n silicon detectors. On each side of the module, two detectors are wire-bounded together to form 12.8 cm long strips. The end-cap detector modules have a similar structure. While pixels, thanks to their geometry

have a good 2-dimension coverage, microstrips have a better resolution along one privileged coordinate. The spatial resolution of the SCT is 17 $\mu m$ along the $r - \phi$ direction and 580 $\mu m$ in the z direction.

The outer part of the ID is the **Transition Radiation Tracker (TRT)** [29]. It is based on the use of straw detectors. Each straw has a 4 mm diameter and a maximum lenght of 144 cm in the barrel region. The TRT tubes are arranged in 36 layers. The tubes are filled with a Xenon gas mixture that ionizes when charged particles pass through it. In the middle of each tube a gold-plated 30 $\mu m$ thin tungsten wire collects the ionization low energy signal. On each wire the signal is then amplified, shaped and discriminated according to two adjustable thresholds, a low-threshold (LT) of about 300 eV and a high-threshold (HT) of about 6-7 keV. The two thresholds allow for simultaneous measurement of tracking information and identification of characteristic large energy deposits due to the absorption of transition radiation photons. Each layer is interspersed with a polypropylene fiber which stimulates the emission of transition radiation from ultrarelativist particles. This process causes the emission of X-rays, which are absorbed by the Xenon present in the gas mixture, and results in a high energy signal which can thus be distinguished from the ionization signal by the voltage intensity. The spatial resolution of the TRT is of 130 $\mu m$. The TRT is also used to distinguish electrons from other charged particles, such as pions [29]. Candidates of photon conversions into electron-positron pairs are required to have two tracks with a minimum of 20 TRT hits and four silicon (SCT and Pixel) hits. The conversion vertex is required to be well reconstructed and to be at least 60 mm away from the primary vertex in the radial direction.
Pion candidates are selected from reconstructed particle tracks that have a minimum of 20 TRT hits and four silicon hits. Further selection criteria are applied to reject electrons, protons and kaons. Any track that does not have a hit in the innermost Pixel layer or that is reconstructed as a part of a photon conversion candidate is excluded. In addition, any track with a measured dE/dx above 1.6 MeV $g^{-1}$ $cm^2$ [29] in the Pixel detector is excluded in order to reduce the contamination from protons, or kaons, at low momentum. The first step towards establishing electron identification with the TRT is to observe the expected raise in the average number of HT hits with the Lorentz factor $\gamma$. Figure 2.8 shows the average HT fraction, defined as the ratio between the number of hits that exceed the high threshold and the total number of hits on track, for different $\gamma$ ranges. The different shape of the curve leads to a good discrimination between electrons and pions [29], since the distribution for electrons is clearly shifted to higher values. The mis-identification probability reaches a maximum of 12% in the $0 < \eta < 0.625$.




(a)TRT barrel (b)TRT end-caps

Figure 2.8: The high-threshold fraction for electrons (blue) and pions (orange) candidates for TRT barrel (left) and end-caps (right). The HT fraction is defined as the ratio hits on track that exceed the high threshold to the total number of hits on the track. The different shape of the curve leads to a good discrimination between electrons and pions [29], since the distribution for electrons is clearly shifted to higher values.

### 2.2.3 The Electromagnetic Calorimeter

The main goal of the electromagnetic calorimeter (EM) [16] is to measure destructively the energy of electrons and photons and to discriminate these two particles from the rest of produced ones. The EM is designed to provide a good spatial resolution for the reconstruction of photons in a narrow cone. This feature has been of fundamental importance in the discrimination of the $\pi^0 \rightarrow \gamma\gamma$ background from the $H \rightarrow \gamma\gamma$ signal in the Higgs discovery. This detector can measure the direction of the photon on the z-axis, which has been used in the Higgs discovery.
The EM is composed of a sequence of lead aboserber plates, as passive medium, and liquid Argon gaps as active material. Due to the Pb large electromagnetic cross section, the incident particle interacts via *bremsstrahlung* and pair creation, causing an electromagnetic shower. The Argon is thus ionized and a signal related to the energy of the incident particle is read out by an electrode.
An illustration of the electromagnetic calorimeter is provided in figure 2.9. The calorimeter is mainly divided into a central barrel, which has a cylindrical covarage in pseudo-rapidity of $|\eta| < 1.475$, and two end-cap elements which cover the range $1.375 < |\eta| < 3.2$, perpendicular to the beam axis.
The central barrel is divided into three compartments with different characteristics and cell dimensions. The first compartment identifies with high

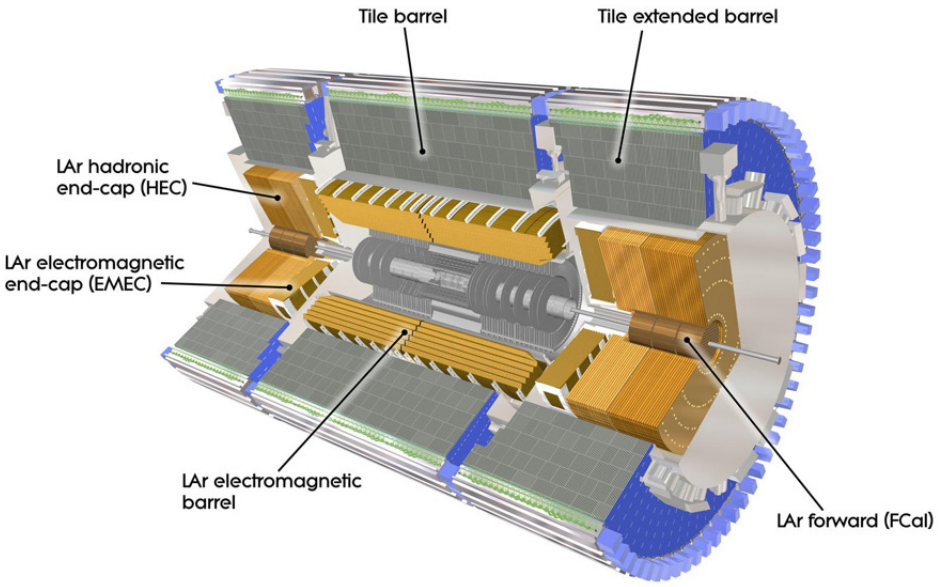



Figure 2.9: Schematical view of the electromagnetic and hadronic calorimeters.

precision the angle and the position of the electromagnetic cluster. It is realized using very narrow cells (4 mm wide in $\phi$). The second compartment (with cell dimensions $\Delta\eta$= 0.025 $\cdot\Delta\phi$=0.025) measures the released energy. The last compartment, made by cells of dimension $\Delta\eta$= 0.025 $\cdot\Delta\phi$=0.05, is used to measure the energy of those clusters that, due to their high energy, are not all contained in the central compartment. This last region is important in the discrimination between electromagnetic and hadronic clusters, since electrons and photons rarely have enough energy to reach the outest part of the electromagnetic calorimeter. The total thickness of this detector is >24 radiation lenghts ($X_0$) in the barrel region and >26 $X_0$ in the end-caps.
The electromagnetic calorimeter must be able to identify photons and electrons with energy which ranges from 5 GeV to 5 TeV. In order to provide good resolution, energy measurements must not deviate from linearity more than 0.5%. The energy resolution of the detector, obtained by experimental measurements, has been fitted with the expression:

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus b \tag{2.13}$$

where $a$ is the stocastic term and $b$ the local constant term reflecting local non-uniformities in the response of the calorimeter. The energy resolution

has been studied using electron beams of energy varying from 10 to 245 GeV [24]. The fraction of total electron energy collected within the whole cluster amounts to more than 90%. The fit results gave the value for the EM resolution:

$$\frac{\sigma_E}{E} = \frac{(9.4)\%}{\sqrt{E(GeV)}} \oplus (0.1\%) \tag{2.14}$$

in agreement with dedicated Monte Carlo simulations.

### 2.2.4 The Hadronic Calorimeter

The main aim of the hadronic calorimeter (HCAL) [16] is to detect jets originated from hadronic showers and to measure their energy. In addition, the HCAL energy measurement is fundamental for transverse missing energy determination, a key ingredient for physics Beyond Standard Model searches. The HCAL is divided into a *Hadronic Tile Calorimeter (HTC)*, two *Hadronic End-Caps Calorimeter (HEC)* and a *Forward Calorimeter (FCAL)*. The scheme of the detector is illustrated in figure 2.9.
The entire hadronic calorimeter covers the region $|\eta| < 4.9$ and every subdetector is designed to measure the jet energy by using different methods. The HTC is made of iron, as passive material, and plastic scintillators as active material, with a structure similar to the electromagnetic calorimeter. It covers the region of $|\eta| < 1.7$. The interaction of hadrons with iron generates an hadronic shower and the light signal produced by the scintillator is proportional to the number of secondary particles produced in the interaction and hence to the energy deposit. The signal is taken and amplified by photomultipliers which convert it into an electrical signal. The HTC has been designed to have an energy resolution of:

$$\frac{\sigma_E}{E} = \frac{50\%}{\sqrt{E(GeV)}} \oplus 3\%. \tag{2.15}$$

After several calibrations with pion beams [25], the measured energy resolution:

$$\frac{\sigma_E}{E_\pi} = \frac{(52.7 \pm 0.9)\%\sqrt{GeV}}{\sqrt{E(GeV)}} \oplus (5.7 \pm 0.2)\%. \tag{2.16}$$

was found, in agreement with Monte Carlo previsions.
In the HEC, which covers the range $1.5 < |\eta| < 3.2$, the active medium consists of liquid argon. The two end-caps are contained in the same cryostat of the electromagnetic calorimeter, even if they are indipendent. The resolution of the HEC has been measured in beam tests with electrons, pions and

muons with energies up to 200 GeV [26]. The analysis of the data taken with electrons gives an energy resolution of:

$$\frac{\sigma_E}{E_e} = \frac{21.4 \pm 0.1\%\sqrt{GeV}}{\sqrt{E(GeV)}} \tag{2.17}$$

while from pion tests:

$$\frac{\sigma_E}{E_\pi} = \frac{70.6 \pm 1.5\%\sqrt{GeV}}{\sqrt{E(GeV)}} \oplus (5.8 \pm 0.2)\% \tag{2.18}$$

in very good agreement with Monte Carlo simulations.
The last sub-part of the detector, the FCAL, situated very close to the beam pipe, covers the region $\mid \eta \mid$<4.9. It is made of liquid argon, iron and tungsten. The FCAL allows the detection of hadronic jets at angles of less than 1 degree. The design resolution for this detector is:

$$\frac{\sigma_E}{E} = \frac{100\%}{\sqrt{E(GeV)}} \oplus 3\% \tag{2.19}$$

and after calibration tests with electrons and pions [27], the following expressions have been found:

$$\frac{\sigma_E}{E}_e = \frac{28.5 \pm 0.1\%\sqrt{GeV}}{\sqrt{E(GeV)}} \oplus (3.5 \pm 0.1)\% \tag{2.20}$$

$$\frac{\sigma_E}{E}_\pi = \frac{94.2 \pm 1.6\%\sqrt{GeV}}{\sqrt{E(GeV)}} \oplus (7.5 \pm 0.4)\%. \tag{2.21}$$

It is very important for the hadronic calorimeter, in order to avoid the loss of statistics and thus to spoil the good resolution of the detector, to be thick enough to contain all the hadronic shower. Consequently, the thickness of the calorimeter has to be tuned to minimize the punch-through into the muon system and to provide good resolution for high energy jets. The total thickness of the detector is of 11 radiation lenghts ($\lambda_0$).

Calorimeters play a crucial role in LHC physics since their intrinsic resolution power, in contrast with the other detectors, increases with energy, as can be seen in equations 2.16, 2.18 and 2.21. Both the calorimeters measure the energy deposits released by the interacting particles such as electrons, photons and jets. The simplest algorithm for jet reconstruction is a *cone algorithm*. The energy of the hadronic jet is calculated by adding the energy released and measured from all the cells contained in the cone of radius $\Delta R$.

The geometry of the calorimeter is optimized in order to obtain the best performance: the energy resolution improves with the increasing R. A too wide cone, however, would lead to a signal degraded by electronic noise and to a greater difficulty in discriminating events. Among these *"cone"* algorithms the *Anti-KT method* uses a different metric definition for the calorimetric cone. The Anti-KT algorithm is also infrared safe [30] and it has been implemented for the reconstruction of the jet used for this analysis.
By measuring the clusters energy, it is possible to measure, by subtraction from the energy of the entire process, the missing transverse energy. The measurement of transverse missing energy is fundamental in identifying particles that interact with very small cross section with matter, as for example neutrinos or particles predicted by BSM theories, such as SUSY or *lightest supersymmetric particles (LSP).*

### 2.2.5 Muon Spectrometer

Excluding undetected neutrinos and BSM particles, only muons can escape from the HCAL ATLAS calorimeter. Because of the large mass with respect to the electrons, the electromagnetic interaction rate is less for muons than for electrons or photons. Muons represent an important signature for both SM and BSM processes. The Muon Spectrometer (MS) [28] is a sub-detector designed for the measurement of muon energy and momentum. The scheme of the MS detector is illustrated in picture 2.10.

Magnets are arranged outside the calorimeter in order to originate a toroidal field whose force lines are concentric and perpendicular to the beam. Muons that traverse the magnetic field change their trajectory allowing the measurement of their momentum. For muons with $p_T$ >30 GeV the measurement of the momentum is more precise than the one obtained with the inner tracker. However, for lower momenta the measurement is less accurate, because of the fluctuations due to the energy loss in the previous layers of the detector, of the order of a few MeV/mm.
The muon spectrometer has an outer diameter of 22 m and it is composed by two sub-systems: the *trigger chambers* and the *precision chambers.*

#### The Trigger Chambers

The *trigger chambers* [31] are fast detectors which make rough measurements of the muon momentum. They are made of **Resistive Plate Chambers (RPC)** and **Thin Gap Chambers (TGC)**.

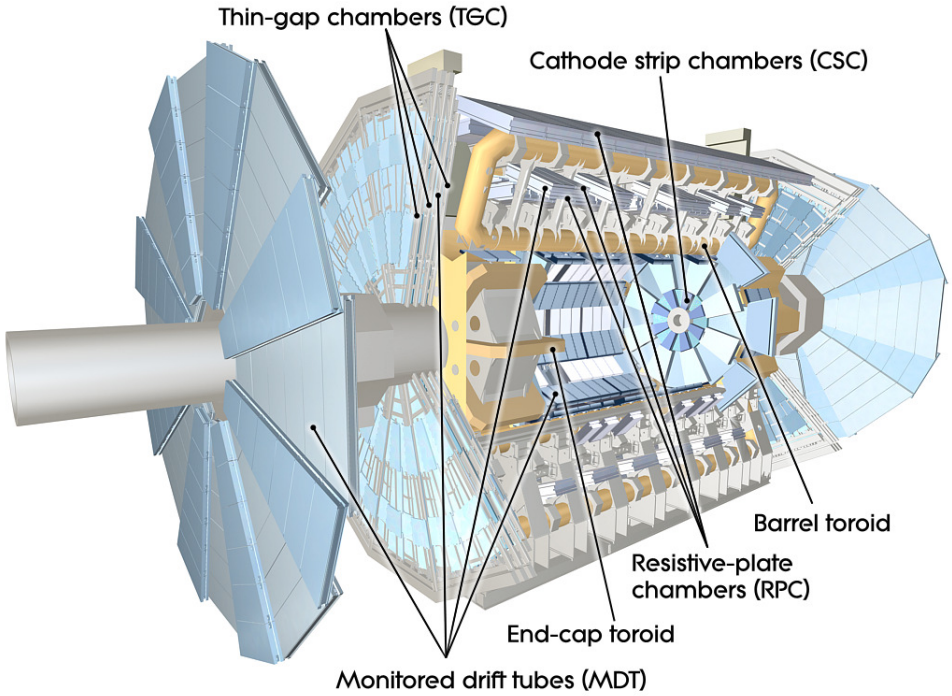



Figure 2.10: Scheme of the muon spectrometer complex.

The RPC are gaseous parallel electrode-plate detectors, with a good spatial resoution of 1 cm in two coordinates and an excellent time resolution of 1 ns. Each of the two rectangular layers which form the RPC is read out by two orthogonal series of pick-up strips: the $\eta$ coordinate is measured by strips which are parallel to the MDT wires, while the $\phi$ coordinate is measured by orthogonal strips. This sub-detector works in the avalanche regime: when a charged particle passes inside the chamber, the primary ionization electrons are multiplied into avalanches by an high electric field, typically 4.9 kV/mm. The signal is read out on both sides of the chamber through capacitive couplings of the strips.
The end-cap region of the trigger chamber is equipped with very thin multi-wire chambers, the TGC. The anode-catode spacing is smaller than the anode-anode spacing and leads to a very short drift time, which is less than 20 ns. To satisfy the requirement of 4 ns time resolution and a good performance in an high particle flux, the TGC works in saturation regime. TGC are filled with a highly quenching gas mixture (55% $CO_2$ and 45% n-pentane $C_5H_{12}$). The spatial resolution of these detectors is 4 mm in the radial direction and 5 mm in the $\phi$ coordinate. The TGC are also used to improve the measurements along the $\phi$ coordinate obtained from the precision chambers.

Both the trigger chambers cover the range $|\eta|<2.4$.

**The Precision Chambers**

These parts of the detector are used to reconstruct the trajectory of the muons. Among them there are the **Monitored Drift Tubes (MDT)** and the **Cathod Strip Chambers (CSC)**.
The MDT are drift chambers of two multi-layer drift tubes, with diameter of 30 mm and alluminium walls, filled with a gaseous mixture of argon and carbon dioxide, kept at a pressure of 3 bar. A high potential difference cable, with respect to the walls, is placed on the axis of the tube. Three or four layers of tubes are superimposed. By measuring the drift time in a single tube it is possible to reconstruct the trajectory of the particle, with a resolution on distance of about 80 $\mu$m. This detector measures only the z coordinate in the barrel region and covers the region $|\eta|<2$.
The CSC are multi-wire chambers with strip cathodes for the measurement of muon momenta in the region $1.0<|\eta|<2.7$. The CSC wires are composed of parallel anodes which are perpendicular to 1 mm large strips of opposite polarity. The anode-catode distance equals the distance between the anode wires, typically of 2.5 mm. The time resolutions is about 7 ns. The obtained spatial resolution is of 60 $\mu m$ in the $\phi$ direction and of the order of one cm in $\eta$.

The arrangement of the Muon Spectrometer makes a particle cross three stations of chambers. It is important to underline that the MS measures momentum and path of all charged particles which cross it, not only muons. For this reason other particles, such as pions, able to overcome the calorimeter, could be also recorded as muons.

### 2.2.6 Trigger and Data Acquisition System

The Large Hadron Collider, with a design luminosity of $10^{34}$ cm$^{-2}$s$^{-1}$ and the high frequency collisions, will have an output of about one billion events per second, a number of data impossible to menage without applying some filters. The aim of the trigger and acquisition system (TDAQ) is to choose and readout only the "interesting" events produced in pp collisions, following the scheme illustrated in figure 2.11. That system is divided into three *on-line* event selection levels: **Level 1 trigger (LVL1)**, **Level 2 trigger (LVL2)** and **Event Filter (EF)**. Every consecutive level requires more and more tightening conditions with respect to the previous level. The initial collision rate of 40 MHz has to be reduced to 300 Hz in order to be stored. The TDAQ has to operate a compromise between the amount of data that have to be

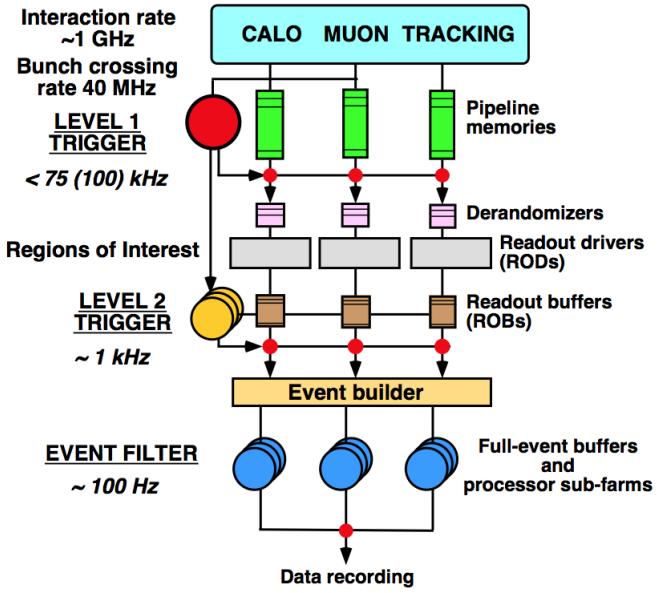



Figure 2.11: ATLAS trigger system.

rejected and the detector efficiency, without spoiling the possible discover of extremely rare events.

### Level 1 Trigger

The first trigger level LVL1 operates a preliminar selection of the events, based on information provided by the two calorimeters and by the muon spectrometer in every bunch crossing. LVL1 decides which events have to be kept for further processing, with a latency of 2 $\mu$s. It is implemented with custom hardware, such as ASICS and FPGAs. The LVL1 output rate is from 75 kHz to 100 kHz.

### Level 2 Trigger

The second level trigger LVL2 reduces the storage rate from 100 kHz to 1 kHz, with a latency ranging from 1 ms to 10 ms, depending on the type of event processed by the trigger. The events selected by the LVL1 are recorded in the Read Out Buffers (ROB); they are then processed by the LVL2 trigger, implemented in a farm of commercial PC. To reduce the amount of data from the ROBs to the LVL2, LVL2 processes only data from Region Of Interest (ROI), which are data fragments from regions of the detectors close to LVL1 reconstructed objects.

#### Event Filter

The Event Filter operates once the reconstruction of the event has been made by the *Event Builder (EB)*, which collects data from the ROBs for each event accepted by the LVL2. The algorithm used by the EF, which is also implemented in a farm of commercial PC, are basically the same used in the off-line event selection, where the calibration and alignement costants are still not the final ones. The latency of the processes executed by the EF is of 2 s. At this step of the data selection, every event can be reconstructed from the primary interaction vertex. The writing rate on memory of the Event Filter is of the order of magnitude of 10-100 MB/s.

#### Data Acquisition System

The Data Acquisition System (DAQ) receives and buffers the data from the detector-specific read out, called *Read Out Drivers (ROD)*, at the LVL1 trigger rate and transmits the data to the LVL2 trigger when requested. After passing the level 2 selection criteria, the DAQ builds the event and moves it to the Event Filter. Finally, the DAQ forwards the final selected events to the mass storage. Besides, the DAQ provides an interface for configuration, control and monitoring of the ATLAS detectors during data taking.

### 2.2.7 Forward Detectors

The Forward Detectors [32] have been developed and projected on a later stage during the construction of the ATLAS detector. Among the Forward Detectors (some of them are reported in figure 2.12) there are:

- **Minimum Bias Trigger Scintillator (MBTS)**: placed at 356 cm from the interaction point, it covers the region $1.9<|\eta|<3.8$. It is used to select events from low luminosity collision runs and to provide the covered $\eta$ region with a detector sensitive to low momentum particles.
- **Beam Conditions Monitor (BCM)**: it is 1.84 m away from the interaction point and covers the region $3.9<|\eta|<4.1$. The main aim of this system is to detect signals of beam instability, in order to prevent the whole experiment from damages. Furthermore, thanks to its good time resolution, it measures the interaction rate, and distinguishes, from the arrival times, true collisions from the background giving an instantaneous relative luminosity measurement.
- **Luminosity Measurements Using Cherenkov Integrating Detector (LUCID)**: placed at 17 m from the interaction point, it makes

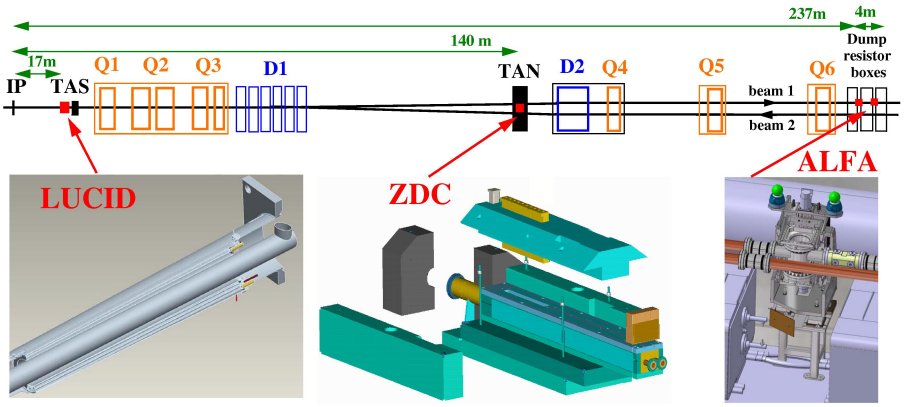



Figure 2.12: Position of some of the ATLAS Forward Detectors with respect to the interaction point. MBTS and BCM are installed near the IP.

use of the Cherenkov effect for relative luminosity measurements in the region 5.6<| $\eta$ |<5.9.

- **Zero Degree Calorimeter (ZDC)**: it is designed for relative luminosity measurements during heavy ion runs. It is placed at 140 m from the interaction point and covers the region | $\eta$ |>8.3. It is used to measure neutral particles (such as photons and neutrons) very close to the beam axis. It is important in high energy physics for the measurements of forward cross sections.

- **Absolute Luminosity for ATLAS (ALFA)**: this detector, whose aim is to intercept the small-angle scattered protons, is inserted in the so-called "Roman Pots" [33], cavities located at a 240 m distance from ATLAS in the LHC tunnel on both sides of the collision point. ALFA covers the region 10.6 <| $\eta$ |<13.5. It has been designed mainly for the measurement of the proton cross section in elastic scattering collisions in the coulomb nuclear interference region. On this measurement depends the absolute luminosity evaluation in ATLAS.

# Chapter 3

# The Higgs Boson Physics and the $t\bar{t}H$ Channel

This chapter will be devoted to the discussion of the Higgs boson production mechanisms at LHC, as well as its main decay channels which make possibile, by the reconstruction of the particles in the final state, the detection of the Higgs boson in the ATLAS experiment.

## 3.1 Higgs Boson Production at LHC

The Higgs boson production cross section for the different mechanisms depends on the available energy in the collisions. At the LHC centre of mass energies reached in 2012 ($\sqrt{s}$=8 TeV), the four principal production mechanisms are the *gluon-gluon fusion*, the vector-boson fusion, the *Higgs-Strahlung* and the $t\bar{t}H$ associated production.

### 3.1.1 Gluon-Gluon Fusion (ggF) : $pp \to gg \to H$

The *gluon-gluon fusion* accounts for 87% of Higgs boson production at $\sqrt{s}$= 8 TeV at LHC. It involves a triangle loop of heavy quarks, such as tops and bottoms, as illustrated in figure 3.1. The theoretical calculation of the cross section for this process is computed at the Next to Next to Leading Order (NNLO) [36]. Among the NLO corrections for the ggF cross section calculation, "*virtual*" and "*real*" loops have to be considered: while the former does not affect the initial or final states, leading to a process identical to the LO one [37], the latter involves corrections due to addictive partons in the final states ($gg \to Hg$, $qg \to Hq$ and $q\bar{q} \to Hg$). Two examples of loops

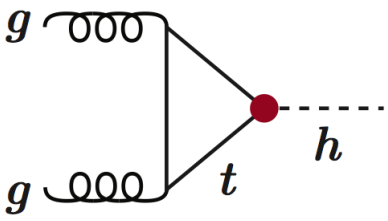



Figure 3.1: Feynman diagram at the *leading order* (LO) for the gluon-gluon fusion production mechanism.

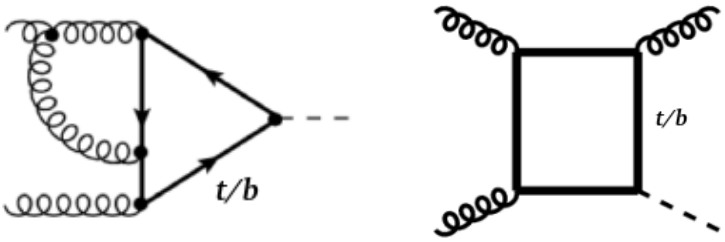



Figure 3.2: **Left**: virtual corrections at NLO for the ggF production of the Higgs boson. In the Feynman diagram the exchange of a virtual gluon is represented. **Right**: real corrections at NLO. The Higgs boson is produced through a quadratical loop of top/bottom quarks. A gluon remains in the final state contributing to a different final state signature ($gg \rightarrow Hg$ process).

are illustrated in figure 3.2: on the left, virtual corrections at NLO for the $ggF$ production of the Higgs boson are due to the exchange of a virtual gluon; on the right, real corrections at NLO are due to a quadratical loop of top/bottom quarks where a guon in the final state remains and contributes to a different final state signature. The total contribution of NLO corrections is not negligible and it can reach the 60%-90% depending on the Higgs mass. The uncertainty which affects the $ggF$ production cross section measurement is of 15-20% and mainly depends on the chosen energy scale of the *parton distribution functions.*

### 3.1.2 Vector-Boson Fusion (VBF)

The vector-boson fusion mechanism:

$$pp \rightarrow qq \rightarrow V^*V^* \rightarrow qqH \tag{3.1}$$

(where V* indicates virtual vector bosons) represents the second major contribution to the Higgs production at LHC (approximately 8% for SM Higgs

boson of $m_H$=125 GeV).

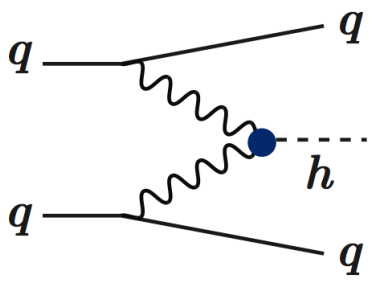



Figure 3.3: Feynman diagram at the leading order for the vector-boson fuction production mechanism.

In VBF processes, shown for example in the Feynman diagram of figure 3.3, two high $p_T(> 20GeV)$ jets arise very close ($\sim$ mm) to the beam direction in the forward and backward regions of the detector with respect to the interaction point from the hadronization of quarks. The jets originate from elastically scattered protons and should be found in the pseudorapidity range $\mid \eta \mid \geq 2$ in the end-cap regions of the ATLAS detector. This channel has a distinguishable signature which makes easier the separation between signal and background. The production cross section has been calculated with full NLO QCD and EW corrections and approximate NNLO QCD corrections [38].

### 3.1.3 Higgs-Strahlung (VH)

Another mechanism of SM Higgs boson production is known as *Higgs-Strahlung*, where the Higgs boson is irradiated through an *off-shell* W/Z boson, whose Feynman diagram is visible in figure 3.4:

$$pp \rightarrow q\bar{q} \rightarrow VH \tag{3.2}$$

where V stands for W,Z.
This kind of process is the main production mechanism at $e^+e^-$ colliders (e.g. LEP, TLEP, NLC) but the third one, in order of relevance, at LHC. QCD corrections range between the 25% and the 40% [39]. Figure 3.5 shows a NLO process for the VH production.

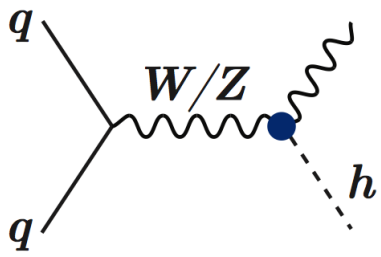



Figure 3.4: Feynman diagram for the Higgs-strahlung process at the first perturbative order.

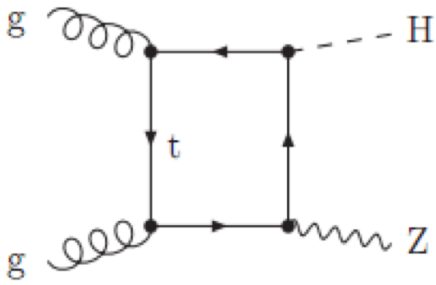



Figure 3.5: Feynman diagram for a NLO VH process. It involves a quadratical top quark loop, which can also contribute to the $gg \rightarrow H$ channel.

### 3.1.4 $t\bar{t}$ Associated Production ($t\bar{t}H$)

The $t\bar{t}H$ production channel, as shown in figure 3.6, is the channel under analysis in this thesis. It is an important production mechanism since it provides direct measurement of the Higgs coupling with the top quark, the two most massive elementary constituents of the SM, close to the EW breaking energy scale. Due to its large mass, the top quark Yukawa coupling is expected to be near one. The cross section for this process is comparable to the VH one. LO predictions for this process are affected by large uncertanties due to the strong coupling costant renormalization. The corrections for the leading orders [40] show an increse of 20 % in the cross section for this process, as a function of the Higgs boson mass. This channel will be discussed in further details latter in this chapter.

Figure 3.7 shows the behaviour of the Higgs boson production cross section as a function of its mass for a center of mass energy of 8 TeV. The coloured bands in figure represent the theoretical uncertanties on the cross section value calculated at NLO. As visible from the figure, for a SM Higgs boson with

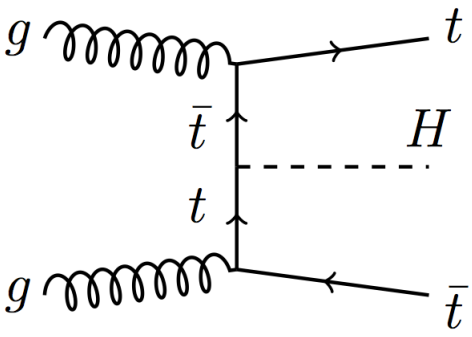



Figure 3.6: Feynman diagram for the $t\bar{t}H$ process for the Higgs boson production.

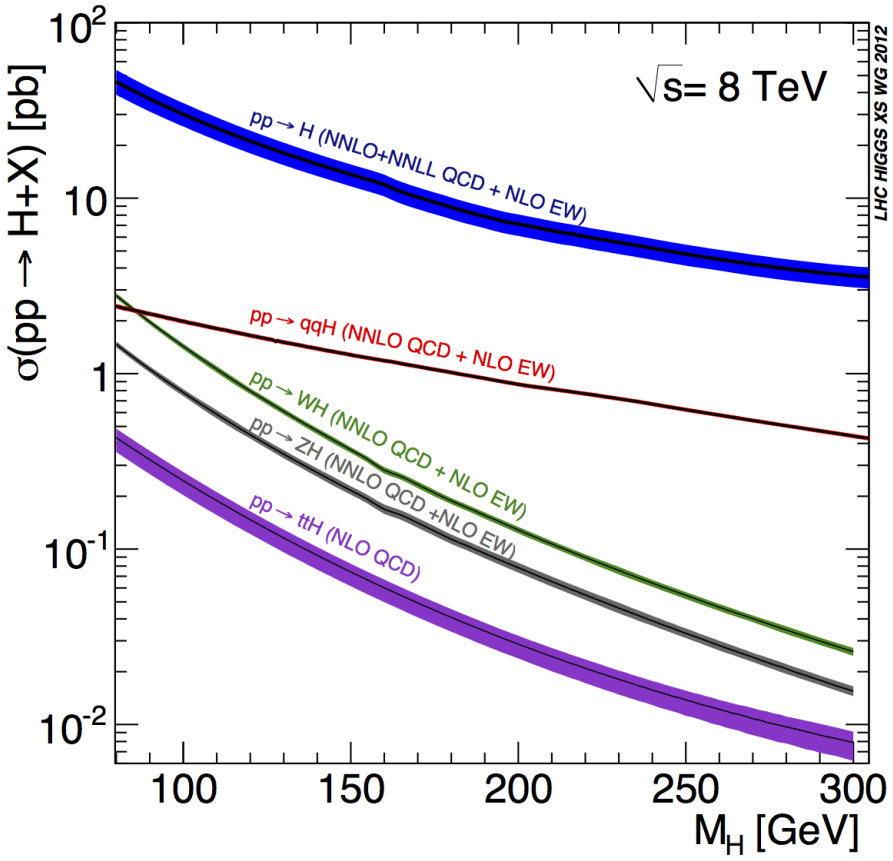



Figure 3.7: Behaviour of the Higgs boson production cross section as a function of its mass. The figure is presented for a center of mass energy of 8 TeV. The coloured bands represent the theoretical uncertainties on the cross section value. They had been calculated through perturbative methods at next to leading orders.

| $m_H$=125 GeV | | | |
|---|---|---|---|
| Process | Cross section (pb) | Scale uncertainty | PDF + $\alpha_s$ uncertainty |
| ggF | 19.27 | +7.2%/-7.8% | +7.5%/-6.9% |
| VBF | 1.58 | +0.2%/-0.2% | +2.6%/-2.8% |
| WH | 0.70 | +1%/-1% | +2.3%/-2.3% |
| ZH | 0.41 | +3.1%/-3.1% | +2.5%/-2.5% |
| $t\bar{t}H$ | 0.13 | +3.8%/-9.3% | +8.1%/-8.1% |

Table 3.1: Cross section values of the different Higgs production mechanisms for a center of mass energy of 8 TeV. For each process, the systematics due to scale factor, PDF and $\alpha_s$ uncertainties are also reported.

$m_H$=125 GeV, the major production mechanism is the $ggF$ (indicated with $pp \to H$ and represented with a blue line); it is about two order of magnitude, in logarithmic scale, larger than the other three production mechanisms. Numerical values for the production cross section of the different processes, as well as the systematics due to the renormalization scale, PDF and $\alpha_s$ uncertanties, are provided in table 3.1.
During the 2012 data taking, according to the values provided in table 3.1, with a total production cross section of 44.2 pb at 8 TeV and a total integrated luminosity of 20.3 fb$^{-1}$, the expected number of produced Higgs boson is about $10^5$. Unfortunately, only a small fraction of these events can be detected, due to the efficiency of the detector in separating background from signal events. For this reason, more refined off-line analysis techniques are needed to better descriminate signal from backgrounds. The discrimination is even more compelling in studying final states coming from small production cross sections.

## 3.2 Higgs Boson Decay Channels

The coupling costants of the Higgs boson with the gauge boson W and Z are strictly dependent on the square of their mass. In the fermionic sector the coupling costants are linearly dependent on the masses. From SM, the relations between couplings and masses are:

$$g_{VVH} = 2\sqrt{\sqrt{2}G_F}M_V^2 \tag{3.3}$$

$$g_{f\bar{f}H} = \sqrt{\sqrt{2}G_F}M_f \tag{3.4}$$

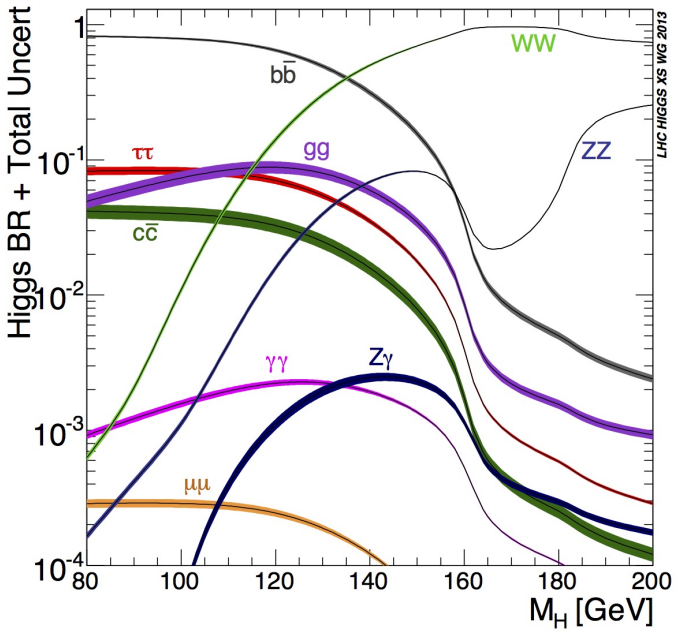



Figure 3.8: Standard Model Higgs boson branching ratios as a function of Higgs mass. The coloured bands correspond to the current level of theoretical uncertainty in the calculation.

the Higgs boson thus prefers to decay in SM particles with larger mass, compatible with the kinematical constraints. The Higgs boson branching ratios as a function of the Higgs mass are shown in figure 3.8.

### 3.2.1 Quark Pair Decay Channel

In the quark pair decay channel, the partial decay width must be corrected with a term accounting for the colour factor and for QCD corrections [41]

$$\Gamma(H \to q\bar{q}) = \left[\frac{G_F}{4\pi\sqrt{2}} M_H m_q^2 \beta^3\right] \left(1 + \frac{4\alpha_s}{3\pi}\Delta_H^{QCD}\right) \tag{3.5}$$

where $\beta$ is the speed of the quarks in the Higgs frame. In equation 3.5, QCD corrections cannot be ignored. In fact, for $M_H$=100 GeV/c$^2$, the BR for the $b\bar{b}$ decay channel decreases by a factor 2, including the $\alpha_s$ QCD corrections. Among the quark decay channels, the $b\bar{b}$ has the highest BR, as visible in figure 3.8. Despite the charm quark is heavier than the tau, the decrease of the $H \to c\bar{c}$ process, due to QCD, is the mechanism through which the BR of $H \to \tau\tau$ is bigger than the one for the $H \to c\bar{c}$ process. Although $H \to b\bar{b}$ is the main decay channel for the Higgs boson, it is not accessible

by all the production mechanisms. For the $ggF$, in fact, it would be totally overwhelmed by background processes, such as $Z \to b\bar{b}$ and $qq \to b\bar{b}$, whose cross section is many order of magniture bigger, as visible in figure 3.9.

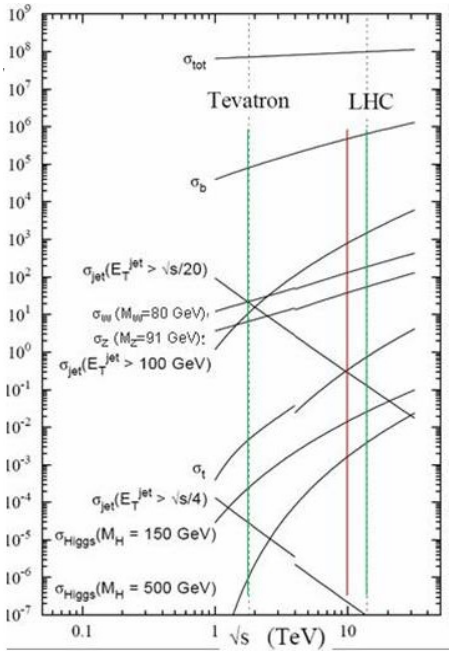



Figure 3.9: Production cross section for different processes as a function of the center of mass energy, for Tevatron and LHC. The SM Higgs production cross section has the lowest value with respect to the other illustrated processes, with only exception for the $\sigma_{jet} > \sqrt{s}/4$ production cross section that presents a decreasing behaviour and reaches its minimum value for a center of mass energy of about 10 TeV.

### 3.2.2 W, Z and $\gamma$ Boson Decay Channels

The partial decay width of the Higgs boson decaying into vector boson W and Z is given by[41]:

$$\Gamma(H \to V\bar{V}) = \delta_V \frac{G_F M_H^3}{16\pi\sqrt{2}} \beta(1 - 4x + 12x^2) \tag{3.6}$$

where $\delta_W$=2, $\delta_Z$=1 and $x = (M_V/M_H)^2$. Since the mass of the Higgs boson is of 125 GeV, one or both the two bosons are produced off-shell. The Higgs boson can also decay into *loop-induced* $\gamma\gamma$ decay channels, as shown in figure 3.10, or Z$\gamma$. Although the BR for the $H \to \gamma\gamma$ decay channel is of few ‰, it has a very clean signature and, together with the $H \to ZZ^*$ decay channel, it has been the main decay channel used for the initial announcement of the Higgs boson discovery.

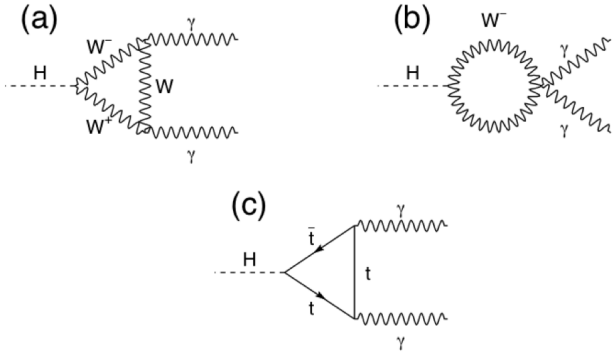



Figure 3.10: Feynman diagram for the decay channel of the Higgs boson into two photons.

### 3.2.3 Lepton Pair Decay Channel

At the lowest lewel, the partial decay width of the Higgs boson into a lepton pair of mass $m_l$ is given by [41]:

$$\Gamma(H \to l\bar{l}) = \frac{G_F}{4\pi\sqrt{2}} M_H m_L^2 \left(1 - \frac{4m_l^2}{M_H^2}\right)^{\frac{3}{2}}. \tag{3.7}$$

As far as equation 3.7 depends on the mass of the fermion, the channel with the highest BR is the one in which the Higgs boson decays into the heaviest pair of leptons: $H \to \tau^+\tau^-$. The dimuonic decay channel $H \to \mu^+\mu^-$ reaches the order of magnitude of about $10^{-4}$, as visible in figure 3.8, and might be accessible for LHC Run-II, while the remaining dielectron channels and neutrino decays are not experimentally relevant.

## 3.3 The $t\bar{t}H$ Channel

Since the top quark is heavier than the Higgs boson, this latter cannot decay into a top quark pair and thus top-Higgs coupling cannot be measured using this decay channel. However, the Higgs boson coupling to top quarks can be experimentally constrained through measurements involving the gluon fusion production mechanism. The ggF process proceeds in fact via a fermionic loop in which the top quark provides the dominant contribution (see figure 3.1), assuming that there is no physics beyond the Standard Model contributing to the loop. Current measurement of Higgs boson production via gluon fusion are consistent with the SM expectation for the top quark Yukawa coupling [42], within experimental uncertainties.

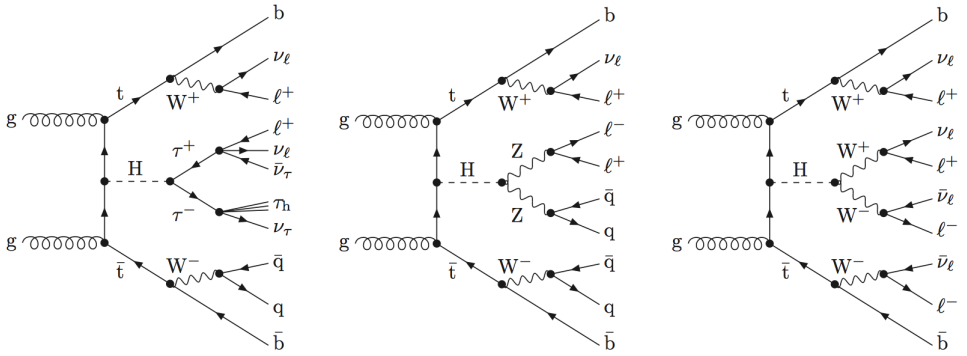



Figure 3.11: Examples of leading order Feynman diagrams for $t\bar{t}H$ production at $pp$ colliders, followed by Higgs boson deacay respectively in $\tau\tau$ (left), WW* (center) and ZZ* (right). The three diagrams are also respectively examples of the two same-sign leptons signature (left), the three leptons signature (center) and the four leptons signature (right).

In order to probe the top quark Yukawa coupling, it is necessary to study a process whose final state decay products can be reconstructed as a Higgs boson associated with a top quark pair in the initial state. In figure 3.11 examples of leading order Feynman diagrams for th $t\bar{t}H$ production at $pp$ colliders, followed by Higgs boson decay in $\tau\tau$ (left), WW* (center) and ZZ* (right), are provided. The three diagrams are also respectively examples of the two same-sign leptons signature (left), the three leptons signature (center) and the four leptons signature (right), according to the final decay of the W/Z boson.
The measurement of the $t\bar{t}H$ production rate provides a direct test of the coupling between the top quark and the Higgs boson. Moreover, many new physics scenarios predict the existence of heavy top quark partners that would decay into a top quark and a Higgs boson: observation of a significant deviation in the $t\bar{t}H$ production rate with respect to the SM prediction would be an indirect indication of unknown phenomena.

Since the top quark decays with nearly 100% probability into a W boson and a b quark, the experimental signatures for Higgs production in association with a top quark pair are determined by the decay of the W boson. The different final states of the W pairs combined with the different Higgs decay channels give a wide variety of complex final states. For example, when both the W bosons decay hadronically, the resulting final state with six jets (two of which are *b-jets*) is referred to as the all-hadronic final state. If one of the W bosons decays leptonically, the final state with a charged lepton, a neutrino and four jets is called lepton + jets. Finally, when both W bosons decay leptonically, the resulting dilepton final state has two charged leptons,

two neutrinos and two b-jets.

As concerns the decay channels for the Higgs boson, the process $H \to b\bar{b}$ contributes almost for the 60% of the total Higgs boson decay width. The first searches for $t\bar{t}H$ production were performed by the CDF [43] and D⊘ [44] experiments at the Tevatron collider, which only put limits to the SM Higgs boson production. However, the inclusive searches for the process $H \to b\bar{b}$ are very challenging, since the production cross section for the inclusive $b\bar{b}$ production, which is of 10 $\mu$b, is about 7 times larger than the signal production cross section, as visible in figure 3.9. The associated production of the Higgs boson with a top quark pair helps to separate the signal from the overwhelming QCD background, by the reconstruction of all the final states for this channel.
The search for the $t\bar{t}H$ production using the multilepton final state looks for the secondary decay products from the Higgs $H \to WW$, $H \to ZZ$ and $H \to \tau\tau$ decay channels and is sensitive to the leptons originated from the decay of the W bosons produced in the top quark decays.
In the next sections, the multilepton signature, used in this analysis, will be discussed in more details.

### 3.3.1 Multilepton Signature in the $t\bar{t}H$ Processes

#### Signal Description

The $t\bar{t}H$ final state can be observed through different topologies regarding the Higgs boson and the top quark decay modes [45]. As shown in figure 3.11, the multilepton final state mainly comes from three Higgs boson decay channels: $\tau\tau$, WW$^*$ and ZZ$^*$. As regards the top quark, it decays essentially in W$^\pm b$, while each W$^\pm$ boson decays either leptonically ($e^\pm$, $\mu^\pm$ or $\tau^\pm$), with missing transverse energy in the final state, or hadronically, leading to many possible different topologies. To avoid a too large reduction in acceptance and efficiency due to the many particles involved in the final state, the multilepton final state is not necessary exclusive. The multilepton final state can be divided into five main channels, according to the number and flavour of leptons. The five channels are orthogonal to each other and are expected to be sensitive to all of the three major Higgs decay channels. These five states contain either exactly two tight isolated leptons ($e^\pm$ or $\mu^\pm$) with the same charge[1], either three leptons with a total charge of one, or four leptons with a null total charge. Events are also sorted regarding the

[1] In this work, however, also the final signature with two opposite sign leptons will be taken into account.

presence of an additional, or two additional, $\tau^{\pm}$ leptons[2]. Assuming a perfect lepton reconstruction, only the $H \to WW$ channel can produce a same-sign di-leptonic event. The three lepton channel is mainly sensitive to $H \to WW$ decay mode: the 70% of the three lepton events are in fact due to this decay, the remaining fraction of 30% consists of $H \to \tau^{+}\tau^{-}$ and $H \to ZZ$. Finally, half of the four lepton events are expected to be from $H \to WW$ decays. Channels with hadronic $\tau$s are expected to select mainly the $H \to \tau^{+}\tau^{-}$ events, but also $H \to WW$ and $H \to ZZ$ with a W or a Z decaying into $\tau$. The contribution of the main Higgs decay modes to the 5 multilepton $t\bar{t}H$ signatures, at generation level, are reported in table 3.2.

| Signature | $H \to WW$ | $H \to \tau\tau$ | $H \to ZZ$ |
|---|---|---|---|
| Same sign | 100% | - | - |
| 1 $\tau$ | 15% | 74% | 11% |
| 3 leptons | 71% | 20% | 9% |
| 2 $\tau$ | 16% | 82% | 2% |
| 4 leptons | 53% | 30% | 17% |

Table 3.2: Contribution of the main Higgs decay modes to the 5 multilepton $t\bar{t}H$ signatures at generation level [45].

At the generation level, in the same sign channel the $t\bar{t}H$ final state contains 6 quarks: the event is thus characterized by a large jet multiplicity. In the three leptons channel, the $t\bar{t}H$ final state contains 4 quarks: the typical jet multiplicity of the event is of 4. In the four leptons channel, the final state contains a small number of quarks: zero in the $H \to W^{+}W^{-}$ case, two or four in the $H \to ZZ$ case. The $1\tau$ channel, as well as the $2\tau$ channel, typically presents four quarks in the final state. For all the channels, two jets are originated by b-quarks.

### Background Description

The main backgrounds generated at LHC energy for this type of signal can be divided into two main cathegories called reducible or irreducible:

- reducible backgrounds are events with a non-prompt or a fake lepton which is selected as a prompt lepton. These processes cannot lead to a final state which is compatible with the signal signature without a mis-reconstructed object. In particular, for the $t\bar{t}H$ final state, events with

[2]One of the two leptons comes from the top quark decay, while the other one comes from the W boson originated from the Higgs decay

a prompt lepton with mis-reconstructed charge and events in which a secondary muon, from a light or heavy flavour quark decay, is mistaken for a primary lepton from a Higgs boson can be wrongly identified as signal events. Top quark decays are also relevant ($t\bar{t}$(+ jets)) as well as Z+jets and $W^+W^-$. The estimation of this kind of backgrounds is made through the use of data-driven techniques.

- irreducible backgrounds are events with the same final state of the signal. The main backgrounds of this cathegory are $t\bar{t}V$(V=W,Z), tZ, $W^{\pm}Z$ and $W^{\pm}W^{\pm}$. They are usually modelled using Monte Carlo simulations; they are eventually checked in data control regions with a level of accuracy compatible with the accumulated statistics.

Example of reducible backgrounds for the $t\bar{t}H \rightarrow b\bar{b}$ process (a) and irreducible ones (b) are reported in figure 3.12. In the former, the jets in the final state can be distinguished from *b-jets* by using *b-tagging* selections, while in the latter the background presents the same signature of the signal. The cross section for the most relevant irreducible background processes, as well as their systematic uncertainties, are listed in table 3.3.

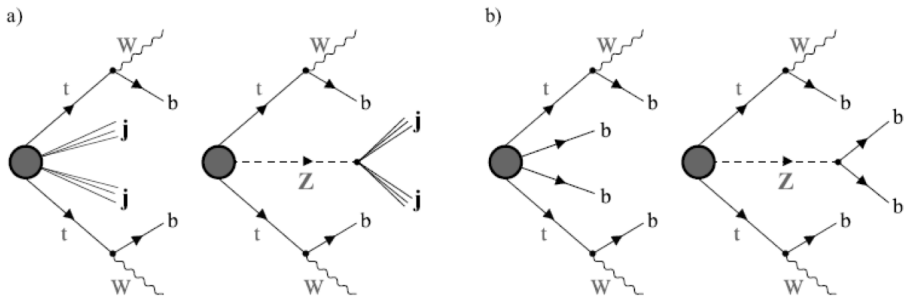



Figure 3.12: a) Example of reducible backgrounds for the $t\bar{t}H \rightarrow b\bar{b}$ process. The jets in the final state can be distinguished from b-jets by using b-tagging selections. b) Example of irreducible backgrounds for the $t\bar{t}H \rightarrow b\bar{b}$ process with 4 b-jets in the final state. The background presents the same signature of the signal.

The jets in the $t\bar{t}H$ hadronic final state lead to larger uncertainties in jet reconstruction compared to leptonic final states. On the contrary, the multileptonic final state can offer better performances in the selection of the events. In fact, in these channels the number of final jets and, by consequence, the overlapping probability substantially decrease. The identification of all the jets in the final state is much performant.

| Process | Cross section (fb) | Scale (%) | PDF(%) |
|---|---|---|---|
| $t\bar{t}W^+$ | 142.6 | 10 | 10 |
| $t\bar{t}W^-$ | 60.5 | 11 | 8 |
| $t\bar{t}Z$ | 205.7 | 12 | 9 |
| $tZ$ | 160 | 4 | 7 |
| $\bar{t}Z$ | 76 | 5 | 7 |

Table 3.3: Monte Carlo cross sections and uncertainties for $t\bar{t}V$ and $tZ$ processes. Uncertainties are symmetrized [45]. Cross sections for the $t\bar{t}W$ and on-shell $t\bar{t}Z$ processes have been evaluated at NLO [53]. The corresponding PDF uncertainties have been evaluated with MC@NLO.

## 3.4 Monte Carlo Samples

The Monte Carlo simulation of physical processes plays a crucial role in testing the response of the detector. The simulation starts with the random generation of the process under consideration. After the collision between protons has took place, partons emit gluonic radiation and, reached an energy of about 1 GeV, they start haronization (*parton showering*), leading to the production of particles with null total colour charge.
After the simulation of the physical process, a Monte Carlo generator mimics the response of the detector. The same algorithms used on the data samples reconstructed from the experiment are also implemented on the simulated samples in order to compare simulation to real data to eventually discover new physics.
The Monte Carlo generators used for the signal description are listed in table 3.4 for all the decay channels. The cross section and the luminosity of the samples are also presented [50]. The samples are generated with inclusive Higgs boson decays, where the branching ratios have been set to the *LHC Higgs Cross Section Working Group* recommendation for $m_H$=125 GeV [35] as well as the inclusive cross section (129.3 fb) at $m_H$=125 GeV. The matrix element calculation is performed at NLO and the generators used for the Higgs signal simulation are PowHel [46] and Pythia 8 [47] for showering.
Monte Carlo generators used for the background description are listed in table 3.5. The cross section and the luminosity of the samples are also presented.
The used background sample corresponds to irreducible background processes, $t\bar{t}Z$, $t\bar{t}W^{\pm}$ and $W^+W^-$, and to reducible $t\bar{t}$, single top production, W+jets and Z+jets processes. For all the background processes the same selection cuts applied to the signal processes are required. All the signal and background processes contain a different number of jets and leptons in the

| Monte Carlo signal samples | | | | |
|---|---|---|---|---|
| Process | Generator | Cross Section (fb) | $\mathcal{L}$(fb$^{-1}$) | Detector Simulation |
| $t\bar{t}H$ →allhad + H | PowHel+Pythia 8 | 59.09 | 2146.5 | Full |
| $t\bar{t}H$ →ljets + H | PowHel+Pythia 8 | 56.63 | 2238.9 | Full |
| $t\bar{t}H$ →ll + H | PowHel+Pythia 8 | 13.58 | 9332.0 | Full |

Table 3.4: Monte Carlo generators used for signal description. The cross section and the generation luminosity for each channel is also listed.

final state; processes with a number of leptons different from two, which is the main requirement in the final state, should thus have a lower impact on signal contamination.
Monte Carlo samples are generated to mimic the data collected with the ATLAS detector during the 2012 data taking period, which correspond to a total integrated luminosity of 20.3 fb$^{-1}$. All the samples have been weighted with the relative $k$-factor, which is the ratio between the calculated next-to-leading order $\sigma_{NLO}$ cross section and the leading order $\sigma_{LO}$ one

$$k_{factor} = \frac{\sigma_{NLO}}{\sigma_{LO}}. \tag{3.8}$$

and scaled to the data luminosity.

## 3.5 Object definition

The physical objects used for this analysis are muons, electrons, jets and missing transverse energy $E_T^{miss}$. In this section the selection criteria required for particles and the other physical quantities will be illustrated.

### 3.5.1 Electron Definition

Isolated electrons produced in many interesting processes can be subjected to large backgrounds from misidentified hadrons, electrons from photon conversions or non-isolated electrons originating from heavy-flavour decays. It is thus important to efficiently reconstruct and identify electrons over the full acceptance of the detector, mantaining a significant background rejection. The electron candidate is reconstructed by the standard algorithm of the ATLAS experiment [54]. A precise measurement of the electron reconstruction and identification over the $E_T$ range from 7 to 50 GeV has been performed with the ATLAS detector using a sample of isolated leptons from $W \to e\nu$, $Z \to ee$ and $J\psi \to ee$ events. This off-line procedure first requires an EM cluster with a total transverse energy above 2.5 GeV. From

| Monte Carlo background samples | | | | |
|---|---|---|---|---|
| Process | Generator | Cross Section (fb) | $\mathcal{L}$(fb$^{-1}$) | Detector Simulation |
| $t\bar{t}$W+Np0 | MadGraph | 122.84 | 2633.2 | Full |
| $t\bar{t}$W+Np1 | MadGraph | 62.98 | 6347.7 | Full |
| $t\bar{t}$W+Np2 | MadGraph | 48.95 | 7450.8 | Full |
| $t\bar{t}$Z+Np0 | MadGraph | 90.70 | 4409.9 | Full |
| $t\bar{t}$Z+Np1 | MadGraph | 60.78 | 6581.1 | Full |
| $t\bar{t}$Z+Np2 | MadGraph | 53.29 | 5580.1 | Full |
| $t\bar{t}$(dilepton) | PowHeg+Pythia 6 | 26554.16 | 1334.1 | Full |
| Inclusive $t\bar{t}$ | MC@NLO | 238.060 | 8.50 | Full |
| $t\bar{t} \to qql\nu$+Np0 | Alpgen+Herwig | 19.193 | 96.79 | Full |
| $t\bar{t} \to qql\nu$+Np1 | Alpgen+Herwig | 20.288 | 92.63 | Full |
| $t\bar{t} \to qql\nu$+Np2 | Alpgen+Herwig | 13.085 | 93.41 | Full |
| $t\bar{t} \to qql\nu$+Np3 | Alpgen+Herwig | 8.692 | 91.46 | Full |
| Single top tchan | AcerMC+Pythia6 | 28433.15 | 271.0 | Full |
| Single top schan | Powheg+Pythia6 | 1817.64 | 660.1 | Full |
| Single top Wt | Powheg+Pythia6 | 2337.67 | 424.8 | Full |
| ZZ→4e | Powheg+Pythia 8 | 0.074 | 16488.72 | Full |
| ZZ→2e2$\mu$ | Powheg+Pythia 8 | 0.171 | 11319.42 | Full |
| ZZ→2e2$\tau$ | Powheg+Pythia 8 | 0.171 | 11048.64 | Full |
| ZZ→4$\mu$ | Powheg+Pythia 8 | 0.074 | 16398.82 | Full |
| ZZ→2$\mu$2$\tau$ | Powheg+Pythia 8 | 0.171 | 10960.77 | Full |
| ZZ→4$\tau$ | Powheg+Pythia 8 | 0.074 | 9846 | Full |
| ZZ→ee$\nu\nu$ | Powheg+Pythia 8 | 0.168 | 1787.06 | Full |
| ZZ→ $\mu\mu\nu\nu$ | Powheg+Pythia 8 | 0.168 | 1785.71 | Full |
| ZZ→ $\tau\tau\nu\nu$ | Powheg+Pythia 8 | 0.168 | 1785.71 | Full |
| WW→ $ll\nu\nu$ | Sherpa | 5.679 | 5.68 | Full |
| Z→ $ee$+Np0 | Alpgen | 711.770 | 7.23 | Full |
| Z→ $ee$+Np1 | Alpgen | 155.17 | 6.96 | Full |
| Z→ $ee$+Np2 | Alpgen | 48.745 | 6.75 | Full |
| Z→ $ee$+Np3 | Alpgen | 14.225 | 6.29 | Full |
| Z→ $ee$+Np4 | Alpgen | 3.76 | 6.49 | Full |
| Z→ $\mu\mu$+Np0 | Alpgen | 712.11 | 7.56 | Full |
| Z→ $\mu\mu$+Np1 | Alpgen | 154.770 | 6.72 | Full |
| Z→ $\mu\mu$+Np2 | Alpgen | 48.912 | 6.73 | Full |
| Z→ $\mu\mu$+Np3 | Alpgen | 14.226 | 3.43 | Full |
| Z→ $\mu\mu$+Np4 | Alpgen | 3.784 | 6.45 | Full |
| Z→ $\tau\tau$+Np0 | Alpgen | 712.1 | 157.58 | Full |
| Z→ $\tau\tau$+Np1 | Alpgen | 154.95 | 270.45 | Full |
| Z→ $\tau\tau$+Np2 | Alpgen | 48.767 | 156.25 | Full |
| Z→ $\tau\tau$+Np3 | Alpgen | 14.184 | 173.44 | Full |
| Z→ $\tau\tau$+Np4 | Alpgen | 3.796 | 106.13 | Full |
| W→ $e\nu$+Np0 | Alpgen | 8044.9 | 0.34 | Full |
| W→ $e\nu$+Np1 | Alpgen | 1579.1 | 1.58 | Full |
| W→ $e\nu$+Np2 | Alpgen | 478.51 | 7.88 | Full |
| W→ $e\nu$+Np3 | Alpgen | 133.73 | 5.01 | Full |
| W→ $e\nu$+Np4 | Alpgen | 35.475 | 7.05 | Full |
| W→ $e\nu$+Np5 | Alpgen | 10.656 | 6.57 | Full |
| W→ $\mu\nu$+Np0 | Alpgen | 8046.3 | 0.42 | Full |
| W→ $\mu\nu$+Np1 | Alpgen | 1579.1 | 1.57 | Full |
| W→ $\mu\nu$+Np2 | Alpgen | 477.41 | 7.81 | Full |
| W→ $\mu\nu$+Np3 | Alpgen | 133.91 | 7.54 | Full |
| W→ $\mu\nu$+Np4 | Alpgen | 35.982 | 7.09 | Full |
| W→ $\mu\nu$+Np5 | Alpgen | 10.4 | 1.92 | Full |
| W→ $\tau\nu$+Np0 | Alpgen | 8028.7 | 0.39 | Full |
| W→ $\tau\nu$+Np1 | Alpgen | 1578.7 | 1.49 | Full |
| W→ $\tau\nu$+Np2 | Alpgen | 477.8 | 7.78 | Full |
| W→ $\tau\nu$+Np3 | Alpgen | 134.01 | 7.54 | Full |
| W→ $\tau\nu$+Np4 | Alpgen | 35.258 | 7.08 | Full |
| W→ $\tau\nu$+Np5 | Alpgen | 10.646 | 6.11 | Full |

Table 3.5: Monte Carlo generators used for background description. The cross section and the generation luminosity for each channel is also listed. MadGraph generator samples use Pythia 6 for parton showering.

Monte Carlo simulations of W and Z leptonic decays, the efficiency of the initial cluster reconstruction is expected to be about 97% at $E_T$=7 GeV and almost 100% for electrons with $E_T$ >20 GeV. Within the tracking volume, tracks with $p_T$ >0.5 GeV are extrapolated from their last measured point to the middle layer of the EM calorimeter. The $\eta$ and $\phi$ coordinates of the impact point are compared to a corresponding seed cluster position in that layer. The matching results successful if the distance between the track impact point and the EM cluster barycentre is $\mid \eta \mid$<0.05. In the case where more than one track is matched to a cluster, the track with the smallest $\Delta R$ distance is chosen. If any track is associated to the cluster, it is classified as an unconverted photon candidate. After the cluster-track matching, the cluster size is optimized in order to take into account the overall energy distribution in the different regions of the calorimeter. The electron cluster is required to be in the fiducial region of the barrel or end cap calorimeters ($\mid \eta \mid$< 2.47). Electrons in the transition region 1.37<$\mid \eta \mid$<1.52 are vetoed. To be identified as an electron, the further following requirements have to be fulfilled:

- the $p_T$ of the electron measured in the electromagnetic calorimeter has to be larger than 20 GeV;
- the track associated to the electron has to be reconducted to a primary vertex, with a significance on the impact parameter (which is the ratio between the transverse impact parameter and the measured error) of $\mid d_0^{sig} \mid < 4\sigma$ and a longitudinal impact parameter $z_0 \sin\theta$ <1 cm.

These cuts are particularly important to suppress the heavy flavour and conversion backgrounds. After the optimization, for the electron is required the *Very Tight Likelihood Identification* operating point, which is an electron identification criteria provided by the E-gamma Performance Group [48]. Additional relative calorimeter and tracking isolation cuts, as well as requirements on track impact parameter, are applied [45]. All the tracks within a cone size of $\Delta R$=0.2 around the electron candidate, with momentum greater than 400 MeV, contribute to the isolation energy. The relative calorimeter isolation variable is required to be $E_T^{cone}/p_T$ <0.05, as well as the tracking isolation variable $p_T^{cone}/p_T$.

### 3.5.2 Muon Definition

Muon reconstruction is performed according to different criteria (leading to different muon "types"), based on to the information collected by the ID, the MS and the calorimetric system. The off-line muon reconstruction for

this analysis uses the MuID (Chain 2[3] [51]) family algorithms, which include muons from both MuID or MuGirl algorithms. *Combined* muons are obtained by fitting the *standalone* requirement (a track from the MS) with an inner detector track, after appropriate corrections due to the energy loss of the particle in the calorimeters. In order to be identified as "tight", muons have to be combined or standalone muons with at least three MDT and CSC hits. For the combined muon, the track reconstruction is performed indipendently in the ID and MS and a combined track is then formed from the successful combination of the two tracks [52]. Furthermore, combined muons from Chain 2 are also required to have at least six silicon hits or at least two hits in the muon trigger detectors. In the absence of segments in the inner station, at least two muon trigger chamber hits are required. The transverse momentum of the combined muons has to be $p_T$ >2.5 GeV and the particle has to lie in the acceptance region of the detector ($\mid \eta \mid < 2.5$). These requirements are provided by the Muon Combined Performance group [52]. To be identified as a muon, in the dilepton analysis, further requirements have to be fulfilled:

- the transverse momentum of the muon has to be $p_T$ >20 GeV;
- the significance on the impact parameter has to be $< 3\sigma$ and the longitudinal impact parameter $z_0 <$ 1cm;
- leptons have to satisfy the isolation criteria.

The isolation requires $E_T^{cone}/p_T$, $p_T^{cone}/p_T$ <0.05. Since the angular distance between the charged lepton and the b-quark decreases as the top quark Lorentz boost increases, a better definition of isolation is required. This new isolation requirement has been developed in the context of boosted top quark searches and it helps also in the resolved regime for top quark measurements. A *mini-isolation* criterium is defined as:

$$I^{\mu}_{mini} = \sum_{tracks} p_T^{track} \tag{3.9}$$

where the sum runs over all tracks (except the matched lepton track) that have $p_T^{track}$ >1 GeV, pass quality cuts and have $\Delta R(\mu, track) < k_T/p_T^{\mu}$. Here, $p_T^{\mu}$ is the muon transverse momentum and $k_T$ an empirical scale parameter, fixed to 10 GeV, optimized for multijet background rejection.

[3]Chain 2 is an algorithm for muon identification which runs on the full muon spectrometer to search for track patterns. Track candidates are built from segments associated to the same pattern and compatible with a curved track.

### 3.5.3 Jet Definition and "b-tagging"

Jets are reconstructed in the hadronic calorimeter using the anti-$k_T$ method [30], which has the property to be an infrared and collinear (IRC) safe algorithm. In fact, soft radiation can provide irregularities in the boundaries of final jets. The knowledge of the typical shape of jets is quoted as facilitating experimental calibration of jets and simplifying theoretical calculations, as well as eliminating some momentum-resolution loss caused by underlying events and pile-up contaminations.
The distance definition $d_{ij}$ between two entities (particles or pseudojets) for the anti-$k_T$ method is:

$$d_{ij} = min(k_{ti}^{-2}, k_{tj}^{-2})\frac{\Delta_{ij}^2}{R^2} \tag{3.10}$$

where R, $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$, $k_{ti}, y_i$ and $\phi_i$ are respectively the usual radius paramenter, the transverse momentum, rapidity and azimuthal angle of the particle $i$. Since the distance depends on the inverse transverse momentum, it is much larger when computed between similarly seperated soft particles than between a hard particle and a soft particle. If a hard particle has no hard neighbours within a distance of 2R, it will accumulate all the soft particles within a circle of radius R, resulting in a perfectly conical jet. If another hard particle is present such that R$< \Delta_{ij} <$ 2R, there will be two hard jets, but it is not possibile for both of them to be perfectly conical. The key feature of this algorithm is that soft particles do not modify the shape of the jets, while hard particles do. In figure 3.13 the behaviour of the anti-$k_T$ algorithm (bottom right plot), compared to other clustering algorithms, is presented: a parton-level event together with $\sim 10^4$ random "ghost" particles have been clustered with the $k_T$ (top left), Cam/Aachen (top right), SISCone (bottom left) and anti-$k_T$ algorithms. For each partonic jet, the region within which the random ghost are clustered into the jet is shown. For the $k_T$ and Cam/Aachen algorithms, the region depends on the specific set of ghosts and the jet jagged borders are a consequence of the randomness of the ghost, since the algorithm is adaptive to soft particles. For the SISCone algorithm, single-particle jets are regular, while composite jets have more varied shapes. Finally, with the anti-$k_T$ algorithm, hard jets are circular with a radius R and only the softer jets have more complex shapes.
The distance parameter used for this analysis is 0.4. Events which contain LooserBad[4] jets are vetoed. The local hadronic calibration is used for the

[4]The LooserBad jet quality requirement is applied in order to eliminate those jets which are not associated to real energy deposits in the calorimeter. They arise from various sources such as hardware problems, LHC beam conditions and cosmic-ray showers.

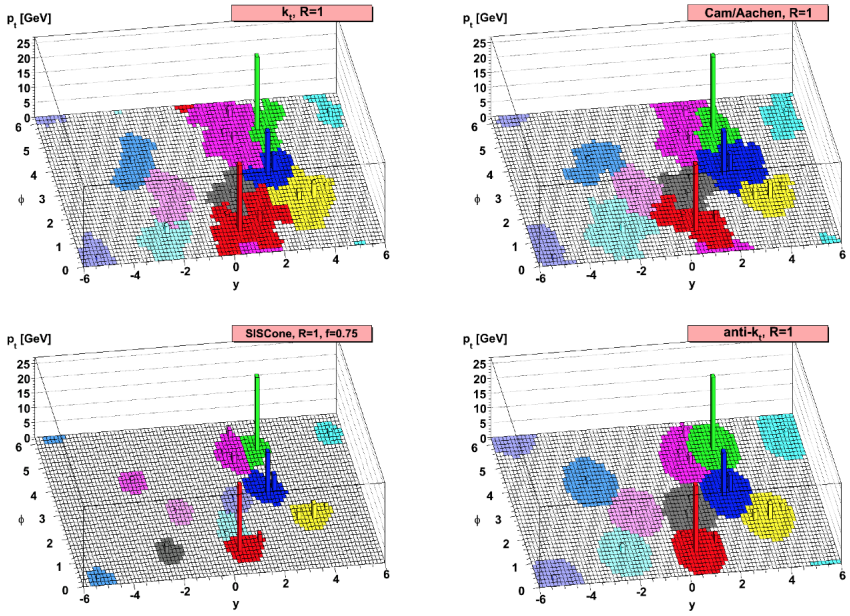



Figure 3.13: A sample parton-level jet event (generated with Herwig [55]), together with many random soft "ghost" jets, is clustered with four different jet algorithms (clockwise: $k_T$, Cam/Aachen, SISCone and anti-$k_T$) illustrating the "active" catchment areas of the resulting hard jets. For kt and Cam/Aachen the detailed shapes are partially determined by the specific set of used ghosts and change when the ghosts are modified. The SISCone algorithm place a boundary roughly in the midway between the green and pink jets, while anti-$k_T$ generates a circular hard jet.

jet energy scale and corrections have been applied to account for energy increase due to pileup. Jets with $\mid \eta \mid < 2.4$ and $p_T < 50$ GeV are required to be associated with the primary vertex. The fraction of the jet $p_T$ belonging to tracks coming from the primary vertex (*jet vertex fraction (JVF)*) must exceed 0.5 (or there must be no tracks associated to the jet). *B-jets* are tagged using a Multi-Variate Analysis (MVA) method called MV1 based on information of the inpact parameter and of the position of the displaced hadron decay vertex inside the jet. The output of the tagger is required to be above 0.7892, which corresponds to a 70% background yield reduction.

### 3.5.4 Missing Transverse Energy

The missing transverse energy (MET) refers to the energy of the particles which are not detected but can be deduced from the conservation laws of energy and momentum. MET is measured by applying the energy conser-

vation in the plane transverse to the beam direction, in order to detect the presence of neutrinos. Neutrinos, in fact, almost do not interact with the matter of the detector, leaving no sign of their transit which could be used for their reconstruction. However, the missing transverse energy is not determined only by the presence of neutrinos. There are many other parameters that affect the energy balance in a experiment: the detector is not fully hermetical, the electronic noise in the calorimeters and in the MS, the presence of muons in some detector regions and the devices required for the signal read-out and, finally, the pile-up. Since all these factors overstimate the effective value of the MET, many algorithms for the reconstruction and minimization of the MET are applied, e.g. by using the topological calorimetric clusters for noise suppression. The missing energy vector is defined as:

$$\vec{E}^{miss} = \sum_{\text{calo cells}} E_i \vec{u}_i \tag{3.11}$$

where $\vec{u}_i$ is the unit vector between the collision point and the position of the energy deposit observed in the $i^{th}$ cell of the calorimeter. The missing transverse energy is the transverse component of $E^{miss}$.

### 3.5.5 Overlap Removal

Overlap Removal (OR) is applied to all fully identified objects, discussed above, that lie in the same overlapping cones in $\Delta R$. In more detail:

- when a muon and an electron, both passing the previous selection criteria, are in the same cone of radius $\Delta R < 0.1$, the muon is retained and the electron is removed. A muon, in fact, passing through the detector, loses only a minimal fraction of its energy (*minimum ionizing particle*), about 3 GeV in all the calorimeters, its track can be thus wrongly reconstructed has belonging to an electron;

- if an electron and a jet are in the same cone of radius $\Delta R$ <0.3, the electron is selected and the jet removed: it is in fact possible for an high energy electron to reach the hadronic calorimeter and to be wrongly identified as a jet;

- two methods to suppress the overlap between muons and jets where tested: a costant distance of $\Delta R$ <0.4 or a $p_T$ dependent distance $\Delta R < 0.04 + \frac{10GeV}{p_T}$ can be chosen. In both cases, when a jet and a muon belong to the same cone of radius $\Delta R$, the jet is retained and the muon is rejected. The variable overlap removal was found to improve the sensitivity in the two lepton channels, compared to a

costant distance criteria: the $p_T$ dependent overlap removal will be thus used in this analysis.

## 3.6 Event Selection

On the reconstructed object, on which the previous reconstruction and identification criteria have been required, the following pre-selection criteria are also applied:

- **LAr Error**: some events cannot be used due to the background noise in the electromagnetic liquid argon calorimeter;
- **Trigger Selection**: events must pass at least one of the single unprescaled lepton triggers: EF_$e$24vhi_medium1 or EF_e60_medium1 for the electron/gamma stream and EF_mu24i_tight or EF_mu36_tight for the muon stream.

In the dilepton same sign channel, events are required to have two light leptons with same charge, verifying the conditions already mentioned above. In addition, at least one lepton has to match the single lepton trigger. The signal is characterized by a large jet multiplicity (at least four) and by the presence of b-tagged jets. The analysis is parametrized in terms of different variables such as $p_T$, $\eta$, MET, etc, and the aim of this work is to find among them the more discriminating variables by using different multivariate analysis methods. In order to enrich the sample in statistic, some of the requirements for the object definition are relaxed. The analysis implemented in this work requires the presence of two leptons, muons or electron, with same sign or opposite sign. On opposite sign events, in order to reduce the contribution arising from Z+jets backgrounds, a Z veto has been applied. The jet multiplicity range is reduced in order to have events with at least two jets and at least one b-jet. The transverse momenta of the two leptons must be greater than 10 GeV. The isolation requirements for both electrons and muons are loosen to $E_T^{cone}/p_T$=$p_T^{cone}/p_T$ <0.1. Any further conditions are required on the MET or on the transverse mass calculated for the analyzed objects.

# Chapter 4

# Multivariate Data Analysis Techniques

In high-energy physics, with the search for ever few signals in ever larger data sets, to extract a maximum of the available information from the data has become essential. Multivariate classification methods (MVA= MultiVariate Analysis) are becoming a fundamental ingredient to most analyses. Also the multivariate classifiers themselves have significantly evolved in recent years. Statisticians have found new ways to tune and combine classifiers to further gain in performances.
Typical areas of application are background suppression (*classification*) and parameter estimation (*regression*) [56], where a physical quantity is extracted from a set of measured observables. The reason to apply multivariate methods is, in most cases, simply the lack of knowledge about the mathematical dependence of the quantity of interest on the relevant measured variables. When no mathematical model is available, or the known models are insufficient, statistical training methods provide a better description of data. Multivariate methods also performe better than other alternatives (such as the Matrix Element [57]) since they employ fewer computational sources. Typical problems that the multivariate methods address to are:

- signal to background discrimination, or selection of those variables which give maximum signal/background discrimination;
- reduction of the dimensionality of the feature space and semplification of the problem;
- finding the regions of interest in the data.

MVA algorithms have been developed because of the vast amount of data and of challenging scientific problems and they are constructed to *learning from data*. The primary goal of learning is to be able to respond correctly to future data. Multivariate methods can be classified according to the characteristics of the training process into two main classes:

- **supervised training**: a set of training events, comprising feature vectors (inputs) and the corresponding targets (or desired outputs) are used;
- **unsupervised training**: no outputs are given and the algorithm has to find them by itself (e.g. a classification of input data into few classes of similar events).

The $t\bar{t}H$ channel has a quite difficult signature in the final state, with a high jet multiplicity and up to four isolated leptons, and low statistics has been acquired with a center of mass energy of $\sqrt{s}$=8 TeV. For this reason, distinguishing signal from background processes is hard and a multivariate analysis will show to achieve better results with respect to standard cut-based analysis. Before looking at how these methods have been applied to the $t\bar{t}H$ analysis at LHC, some basic statistical concepts will be introduced and an overview of the main multivariate techniques used in this thesis will be provided. The methods illustrated in the following sections are impletemented in the *Toolkit for Multivariate Analysis (TMVA)*, the ROOT-integrated framework used for this work [72].

## 4.1 Event Classification

In a typical high energy physics analysis, a class of interesting events, regarded as signal, has to be found among background events. An event is represented in a n-dimentional space by a vector $\mathbf{x}$, whose components are measured quantities (for example particle energy, momenta...) [58]

$$\mathbf{x} = (x_1, x_2, ..., x_n) \tag{4.1}$$

MVA methods perform multidimentional selection in the n-dimentional space. For example, let first consider the distributions of two variables, $x_1$ and $x_2$, that represent two of a possible large number of quantities measured for each event. In figure 4.1 the scatter plots of the two variables for the two data sets, signal (blue circles) and background (red triangles), are represented with different signal/background discrimination approaches. Figure 4.1(a) represents the so-called *cut-based* approach. After choosing some suitable

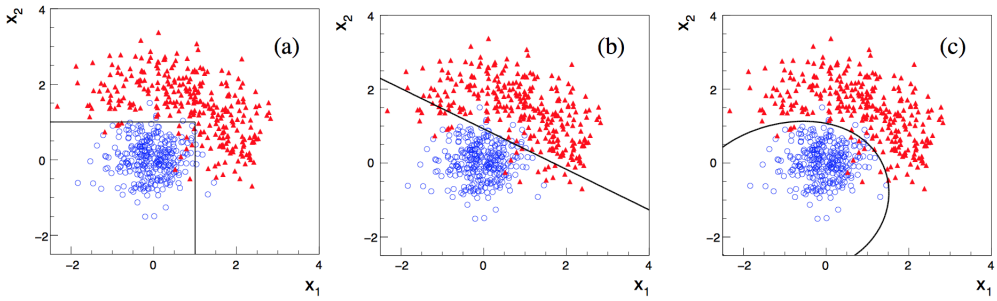



Figure 4.1: Scatter plots of two variables corresponding to two hypothesis: signal (blue circles) and background (red triangles). A possibile event selection could be based on (a) linear cuts, (b) linear boundary or (c) a nonlinear boundary. [59]

cut values $c_1$ and $c_2$, signal is selected by requiring $x_1 < c_1$ and $x_2 < c_2$. Another possible decision boundary is shown in figure 4.1(b), made of diagonal cut. For certain problems, a linear boundary has optimal properties but in some cases neither the cut-based approach nor the linear solution is as good as the nonlinear boundary shown in figure 4.1(c).
The decision boundary represents a surface in the n-dimensional space of the input variables and can be written in the form $y(\mathbf{x}) = y_{cut}$, where $y_{cut}$ is a costant. Events for which $y(\mathbf{x}) \leq y_{cut}$ are accepted as corresponding to the signal hypothesis. On the contrary, events for which $y(\mathbf{x}) \geq y_{cut}$ are rejected. The function $y(\mathbf{x})$ can be used as a *statistical test*. After the determination of the functional form of $y(\mathbf{x})$, the probability density functions (pdfs) for $y(\mathbf{x})$ under both the signal and background hypothesis, respectively $p(y \mid s)$ and $p(y \mid b)$, can be determined. The decision boundary becomes thus a single cut on the scalar variable $y$. Figure 4.2 shows, for example, the distribution of $p(y \mid s)$ and $p(y \mid b)$ as a function of the scalar statistical test $y(x)$. It is clear visible the $y_{cut}$ that has to be applied to reject background processes.
To quantify the goodness of the selection, it is possible to define the *efficiency* as the probability P that a signal event falls in the region of interest. The signal and background efficiencies are thus:

$$\epsilon_s = P(\text{accept event} \mid s) = \int_{-\infty}^{y_{cut}} p(y \mid s)dy, \tag{4.2}$$

$$\epsilon_b = P(\text{accept event} \mid b) = \int_{-\infty}^{y_{cut}} p(y \mid b)dy. \tag{4.3}$$

Considering background as the *null hypothesis*, the background efficiency represents the significance level of the test, also called **type I error**. Being the signal process the alternative, the signal efficiency is called the power of the

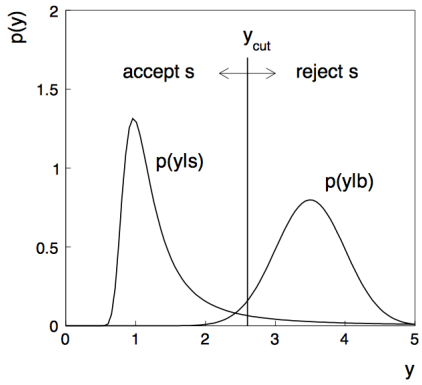



Figure 4.2: Distribution of $p(y \mid s)$ and $p(y \mid b)$ as a function of the scalar statistical test $y(x)$. It is clear visible the $y_{cut}$ that has to be applied to reject background processes.

test: it is the probability to reject the background hypothesis if the signal hypothesis is true (**type II error**).

Physicists seach for a test with maximal discriminating power with respect to a broad class of alternative hypothesis. For two signal and background hypothesis, there is a well defined optimal solution to the problem. The ***Neyman-Pearson*** lemma [61] states that the maximum relative power for the signal hypothesis for a given significance level (background efficiency) is obtained by defining the acceptance region such that, for $\mathbf{x}$ inside that region, the ratio of the joint probability density functions for $\mathbf{x}$ given the signal (s) or background (b) hypothesis ($f(\mathbf{x} \mid s)$ and $f(\mathbf{x} \mid b)$) (*likelihood ratio*) [60]

$$\lambda(\mathbf{x}) = \frac{f(\mathbf{x} \mid s)}{f(\mathbf{x} \mid b)} \tag{4.4}$$

is greater than or equal to a given constant and it is smaller than the fixed constant elsewhere outside the acceptance region. Equation 4.4 represents the statistical test that provides the highest signal efficiency for a given background efficiency, or equivalently, for a given signal purity. In realistic problems, the functions $f(\mathbf{x} \mid s)$ and $f(\mathbf{x} \mid b)$ are not known analytically. Usually theoretical previsions on signal and background distributions rely on Monte Carlo simulations of the $\mathbf{x}$ of the events. Because of the multivariate nature of the data, $\mathbf{x}$ could contain even hundreds of components and it is not trivial to construct a test with a power comparable to that of the likelihood ratio. When likelihood ratio cannot be used explicitly, a variety of other multivariate classifiers could efficiently separate different types of events.
The classifiers employed in this analysis will be presented in the following

sections.

## 4.2 Multivariate Classifiers

### 4.2.1 Linear Cut Optimization

The simplest and most common classifier for discriminating signal from background events is the application of a group of rectangular cuts on selected variables. This classifier is not actually a multivariate analyser but a sequence of univariate ones, because no combination of the variables is achieved and a cut on a variable does not depend on the value of another one [72]. A linear cut returns only a binary response: signal or background. It maximizes the background rejection at given signal efficiency and scans over the full range of the latter quantity. The optimization is performed with the use of multivariate parameter fitters.
At first, the cut optimization starts with building binary search trees for signal and background. For each variable, statistical properties like mean, root-mean-squared (RMS) and ranges are computed to search for the optimal cut. An estimator is required to quantify the goodness of a given cut ensemble. Maximising the estimator means minimizing the background efficiency, $\epsilon_B$ (background rejection is defined as $r_B = 1 - \epsilon_B$) for each signal efficiency $\epsilon_S$. Optimization methods rely on the assumption that one minimum and one maximum requirement on each variable is sufficient to optimally discriminate signal from background. Signal and background efficiencies are derived by counting the training events that pass the cuts and dividing this number by the original sample size. Another way to compute efficiencies is to parametrize the probability density functions of all input variables and to achieve continuous efficiencies for any cut value. However, this method expects the input variables to be uncorrelated. Each generated cut sample corresponds to a point in the $(\epsilon_S, r_B)$ plane. The $\epsilon_S$ dimension is finely binned and a cut sample is retained if its $r_B$ value is larger than the value which is already contained in that bin. A reasonably smooth efficiency curve can be thus obtained if the number of input variables is not too large since the required number of MC samples grows with powers of $2n_{var}$, where $n_{var}$ is the number of input variables. Prior information on a variable distribution can be used to reduce the number of cuts that need to be sampled.
If variables with excellent signal from background separation exist, applying cuts can be quite competitive with more involved classifiers. However, cuts are known to be underperforming in presence of strong nonlinear correlations and/or if several weakly discriminating variables are used. In this latter case,

a true multivariate combination of the information will be rewarding.
In this work, linear cuts will be used for the first event selection procedure and to cathegorize the events as "Opposite Sign (OS)", where the two leptons in the final state have opposite charge, or, on the contrary, as "Same Sign (SS)". Then, the following MVA methods will be applied to the different final states of the $t\bar{t}H$ channel.

### 4.2.2 Neural Newtorks and Multi Layer Perceptron

The development of Artificial Neural Networks (NN) was inspired by the research on the central nervous system and the neurons (axions, dendrites and synapses) which constitute their information processing elements. Currently, the approach stimulated by biological research has been extended to an approach based on statistics, mathematics and optimization theory.
Neural Networks are efficient models for statistical pattern recognition [56]. The main idea is to find a non-linear function $f : \mathbf{x} \to \mathbf{y}$ which, trained on examples, can model relationships between inputs in a $R^n$ space and outputs in a $R^m$ space. The word "network" arises from the fact that function $f$ is a composition of other functions $g_i$ which can also be compositions of other functions $h_i$ and so on. This structure can be represented as a network in which each function is represented by a node and the arrows show the dependences between functions. In most applications, the functions are the nonlinear weighted sums of functions $g_i$:

$$f(\mathbf{x}) = \alpha \left( \sum_i (\omega_i g_i(\mathbf{x})) \right) \tag{4.5}$$

where $g_i$ might be a composition of functions acting on the input vector $\mathbf{x}$ and $\alpha$ is a predefined function called the *activation function*. Largely used activation functions are:

$$f(x) = wx + w_0 \qquad \text{linear function} \tag{4.6}$$

$$f(x) = \tanh(wx + w_0) \qquad \text{hyperbolic tangent function} \tag{4.7}$$

$$f(x) = \frac{1}{1 + e^{wx+w_0}} \qquad \text{logistic function.} \tag{4.8}$$

The most striking feature of neural networks is their ability of training, which means to find an optimal function $f$ using a set of observations. This ability requires the definition of a "cost" function (which can be interpreted as the well known $\chi^2$) so that $C : F \to \Re$, where $F$ is the space of all the functions and $\mathcal{R}$ the space which contains the resulting outputs. For the optimal

solution $f^*$ the relation

$$C(f^*) \leq C(f) \qquad \forall f \notin F \tag{4.9}$$

is fulfilled. The training algorithms search in the solution space in order to find a function $f$ that has the smallest cost. Since training is based on a set of observations, the cost function itself must depend on observations, otherwise it would not model the data. Only a limited number of observations is available and the cost function is itself an approximation. Function C is minimized over a sample of N observations rather than over the true data distribution. In many applications the cost function is based on the $\chi^2$ minimization:

$$C(f) = \frac{\sum_{i=0}^{N}(f(\mathbf{x}_i) - \mathbf{y}_i)^2}{N} \tag{4.10}$$

where $\mathbf{y}_i$ is the desired output.
Let's see in more detail how a neural network is implemented in TMVA. It is clear that the behaviour of an artificial neural network is determined by the layout of the neurons, the weights of the inter-neuron connections and by the response of the neurons to the input, described by a neuron response function $\rho$. The number of input neurons is determined by the number of the input variables. While in principle a neural network with $n$ neurons can have $n^2$ directional connections, the complexity can be reduce by organizing the layers allowing only direct connections from a given layer to the following layer in the so called *multi-layer perceptron (MLP)* architecture. This kind of neural network is the one implemented for this analysis. For a classication problem with $n_{var}$ input variables, the input layer consists of $n_{var}$ neurons that hold the input values $(x_1; ...; x_{n_{var}})$ and one neuron in the output layer that holds the output variable that is the neural net estimator. The neuron response function $\rho$ maps the neuron inputs $(i_1, .., i_n)$ onto one neuron output. It is usually separated into a $\mathcal{R}^n \mapsto \mathcal{R}$ *synapse funcion* $k$ and a $\mathcal{R} \mapsto \mathcal{R}$ *neuron activation funcion* $\alpha$ so that $\rho$=$\alpha \circ k$. The form of the two functions can have for example the following forms: [72]:

$$k \;:\; (y_1^{(l)}, .., y_n^{(l)} \mid w_{0j}^{(l)}, ..., w_{nj}^{(l)}) \rightarrow w_{0j}^{(l)} + \sum_{i=1}^{n} y_i^{(l)} w_{ij}^{(l)} \quad \text{Simple Sum} \tag{4.11}$$

$$\alpha \;:\; x \rightarrow \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad \text{Tanh} \tag{4.12}$$

where the first index $i$ runs over the number of neurons and the second index $j$ on the number of hidden layers. The number of hidden layers and their neuron number have been changed during this analysis, although the Weirstrass

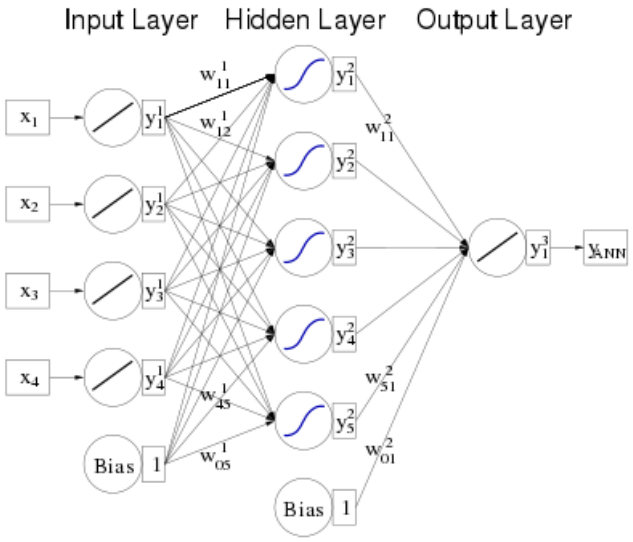



Figure 4.3: Multilayer perceptron structure. The input variables $\mathbf{x}$ are lineary transformed into the $y_i^1$ variables. The input layer neurons are connected to the nodes in the hidden layer, indicated with $y_i^2$, by the connection weights $w_{ij}^1$, while the weights $w_{ij}^2$ connect the hidden layer nodes with the output layer $y_{ANN}$.

theorem, applied to neural networks, assures that for a MLP a single hidden layer is sufficient to a approximate a given continuous correlation function to any precision, provided that a sufficiently large number of neurons is used in the hidden layer. The structure of a multilayer perceptron and of a single layer interconnection, are respectively illustrated in figure 4.3 and 4.4.
The output of a network with a single hidden layer with a Tanh activation function and a simple sum for function $k$ is:

$$y_{ANN} = \sum_{j=1}^{n_h} tanh\left(\sum_{i=1}^{n_{var}} x_i w_{ij}^{(1)}\right) \cdot w_{j1}^{(2)} \tag{4.13}$$

where $n_{var}$ and $n_h$ are the number of input variables in the input layer and in the hidden layer respectively, $w_{ij}^{(1)}$ the weight between input-layer neuron $i$ and hidden-layer neuron $j$ and $w_{j1}^{(2)}$ is the weight between the hidden-layer neuron $j$ and the output neuron.
During the learning process the network is trained with N events $\mathbf{x}$. For each training event the neural network output $y_{ANN}$ is computed and compared to the designed output $\hat{y} \in [0, 1]$ (where 1 is for signal and 0 for background events). An *error function* E measures the agreement between the network

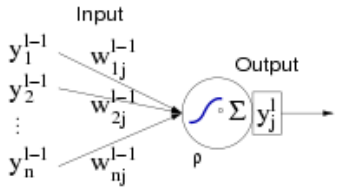



Figure 4.4: Single neuron $j$ in layer $l$ with $n$ input connections. The incoming connections carry a weights of $w_{ij}^{(l-1)}$.

response and the desired one so that:

$$E(\mathbf{x}_1, ..., \mathbf{x}_N \mid \mathbf{w}) = \sum_{a=1}^{N} \frac{1}{2}(y_{ANN,a} - \hat{y}_a)^2 \tag{4.14}$$

where $\mathbf{w}$ is the ensamble of the adjustable weights in the network. The set of weights which minimize the error function can be found using the *gradient descent method*, provided that the neuron response function is differentiable with respect to the input weights. The general behaviour of this method is to start from a random set of weights $\mathbf{w}^{\rho}$ and to update them by moving a small distance in the $\mathbf{w}$-space into the direction -$\bigtriangledown_w$E where E decreases most rapidly:

$$\mathbf{w}^{(\rho+1)} = \mathbf{w}^{(\rho)} - \eta \bigtriangledown_w E \tag{4.15}$$

where the positive number $\eta$ is the *learning rate*. The weights connected with the output layer and with the hidden layer are respectively updated at each event by:

$$\Delta w_{j1}^{(2)} = -\eta \sum_{a=1}^{N} (y_{ANN,a} - \hat{y}_a) y_{j,a}^{(2)} \tag{4.16}$$

$$\Delta w_{ij}^{(1)} = -\eta \sum_{a=1}^{N} (y_{ANN,a} - \hat{y}_a) y_{j,a}^{(2)} (1 - y_{j,a}^{(2)}) w_{j1}^{(2)} x_{i,a}. \tag{4.17}$$

### Supervised Training and Overtraining

When too many parameters in the fitting function are introduced, the NN tends to adapt perfectly to the experimental points, but it would not fit another, statistically indipendent, set of data points. The fitting function does not generalize well but simply trains a particular set of training points. This effect is called *overtraining*. One way to check if overtraining occurs is to divide data into training and validation sets and to perform the training only

on the training sample. The cost function for both training and validation sets should be calculated periodically during the training. When the cost for the validation sample becomes greater than for the training sample, the training should be stopped [65]. Overtraining can be avoided by stopping training early or simplifying the network by removing part of the nodes from the hidden layers ("*pruning*").

### 4.2.3 Boosted Decision Tree (BDT)

A decision tree is a binary tree structured classifier. Repeated left/right (yes/no) decisions are taken on one single variable at time, until a stop criterion is fulfilled. The phase space is thus split into many regions eventually classified as signal or background, depending on the majority of training events that end up in the final *leaf* node. In figure 4.5 a schematic view of a decision tree is presented. A sequence of binary splits is applied to the data, using discriminating variables. Each split uses a cut off variable which should give the best separation between signal and background. At the end of the BDT the leaves are labelled as signal or background, depending on the majority of events in the respective nodes.

In case of *regression trees*, each output node is a specific value of the target variable that the regression function is trying to estimate. The *boosting* of a decision tree extends this concept from one tree to several trees which form a *forest*. The trees are derived from the same training ensemble by reweighting events and finally combined into a single classifier which is given by an average of the individual decision trees. Boosting makes the response of the decision trees, with respect to fluctuations in the training sample, more stable thus enhancing the performance with respect to a single tree.
Whereas a cut-based analysis is able to select only one hypercube as the region of phace space, the decision tree is able to split the phase space into a large number of hypercubes, each of which is identified as either signal-like or background-like. The path along the tree represents an individual cut sequence that selects signal or background, depending on the type of the leaf node.
A weakness of decision trees is their instability to statistical fluctuations in the training sample, from which the tree structure is derived. For instance, if two input variables have similar separation power, a fluctuation in the training sample may cause the tree growing algorithm to split on one variable, while another could have been selected without that fluctuation. The whole tree structure results alterated and the classifier response too.

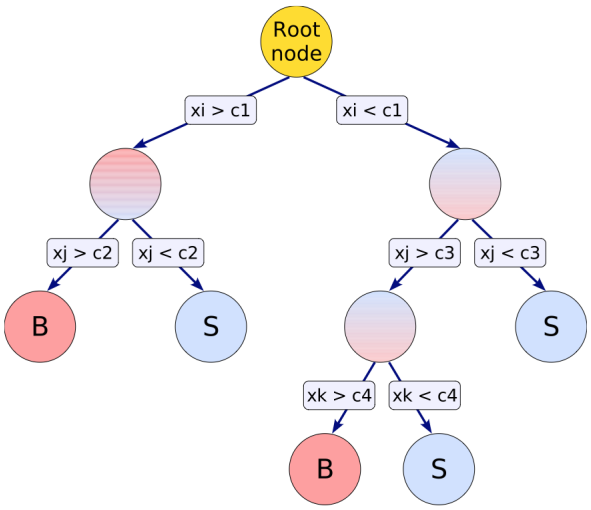



Figure 4.5: Schematic view of a decision tree. Starting from a root node, a sequence of binary splits using the discriminating variables $x_i$ is applied to the data. Each split uses a cut off variable which in that node gives the best signal/background separation. This variable may be used at different nodes, while others might not be used at all. At the end of the decision tree the leaves are labelled with "S" (signal) or "B" (background), depending on the majority of events in the respective nodes.

This instability problem is solved by constructing a decision tree forest and classifying an event on a majority vote of the classifications from each tree. All of the trees are derived from the same training sample, with the events subjected to the boosting procedure that modifies their weights in the sample. In many situations, the boosting performs best if applied to trees that do not have much individual classification power. They are the so-called "weak classifiers", small trees limited in depth. Boosting almost completely eliminate the tendency of overtraining for simple decision trees which are usually grown to a large depth and then "pruned". An example of overtraining is shown in figure 4.6: in 4.6(a) an extremely flexible classifier has menaged to enclose all of the signal events and exclude all the backgrounds. However, if that decision boundary is applied to a statistically independent data sample, the contortions that led to good performances on the training sample will not work so well, as visible in figure 4.6(b). The error rate calculated from the same set of events used to train the classifier understimates the rate on a statistically independent sample.

Among the main boosting algorithms there are the *AdaBoost* [69], *Gradient Boost*, *Bagging* and *Randomized Trees* [70].

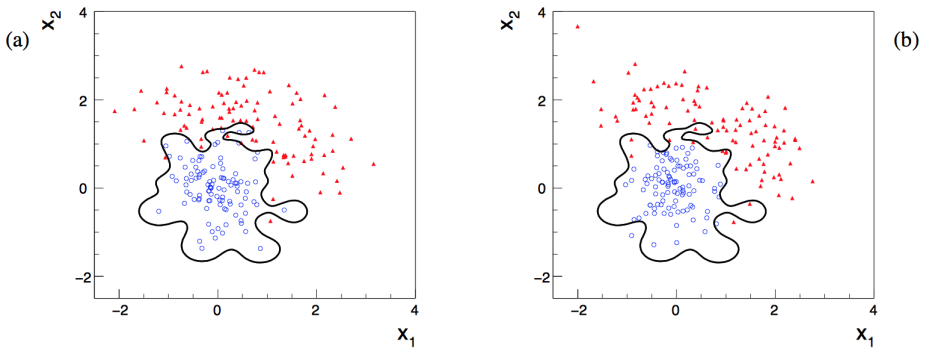



Figure 4.6: Scatter plot of two input variables for events classified as signal (blue circles) and background (red triangles). The decision boundary determined by a particularly flexible classifier is shown as a black line. The decision boundary is much performant on the training sample (left). However, if it is applied to a statistically independent data sample (right), the contortions that led to good performances on the training sample will not work so well. The error rate calculated from the same set of events used to train the classifier understimates the rate on a statistically independent sample.

### Boosting

The idea behind boosting is to make a sequence of classifiers that work on progressively "difficult" events. Instead of using only one high performance classifier, it is preferable to create an ensemble of classifiers, that have a "boosted" performance as a whole. The type of boosting chosen for this analysis is the Adaptive Boost (AdaBoost) where misclassified events during the training of a decision tree are given a higher event weight in the training of the following tree. Starting with the original event weights while training the first decision tree, the subsequent tree is trained using a modified event sample where the weights of previously misclassified events are multiplied by a common *boost weight* factor $\alpha$. The boost weight is derived from the misclassification rate, *err*, of the previous tree:

$$\alpha = \frac{1 - err}{err}. \tag{4.18}$$

The weights of the entire event sample are then renormalized such that the sum of weights remains constant.
The result of an individual classifier is $t(\mathbf{x})$, with $\mathbf{x}$ the data set used for training: $t(\mathbf{x}) = +1$ for signal and -1 for background. The boosted event classification $y_{boost}(\mathbf{x})$ is then given by:

$$y_{boost}(\mathbf{x}) = \frac{1}{N_{trees}} \sum_{i}^{N_{trees}} ln(\alpha_i) \cdot t_i(\mathbf{x}) \tag{4.19}$$

where the sum runs over all the trees in the collection. Small values of $y_{boost}(\mathbf{x})$ thus indicate a background-like events, while large values a signal-like event.

### Training a Decision Tree

The training, building or growing of a decision tree is the process that defines the splitting criteria for each node. At the root node level, an initial splitting criterion for the full training sample is determined. This cut results in two subsets of training events, each going through the same algorithm to determine the next splitting iteration. The procedure is repeated until the whole tree is built. At each node, the variable which provides the best separation between signal and background is chosen and cuts on. The splitting node operation ends once reached the minimum number of events which is specified in the BDT configuration. Since a cut that selects predominantly background is as valuable as one that selects signal, the criteria are symmetric with respect to the event classes. The separation criteria have a maximum where the samples are fully mixed, which correspondes to a purity[1] of $p$=0.5, and falls to zero when the sample consists of one event class only. Tests have revealed no significant performance disparity between the following separation criteria:

- **Gini Index**, defined as $p(1-p)$;
- **Cross Entropy**, defined by $-p\ln(p) - (1-p)\ln(1-p)$;
- **Misclassification error**, defined by $1 - max(p, 1-p)$;
- **Statistical significance,** defined by $S/\sqrt{S+B}$;
- **Average Squared Error**, defined by $1/N \sum^{N}(y-\hat{y})^2$ for regression trees where y is the regression target of each event in the node and $\hat{y}$ is its mean value over all events in the node (which would be the estimate of y that is given by the node).

Since the splitting criterion is a cut on a single variable, the training procedure selects the variable that optimises the increase in the separation index between the parent node and the sum of the indices of the two daughter

[1]The purity of data sample is the fraction of signal events in that sample. It is defined as: $p = \frac{\sum_{i\in s} w_i}{\sum_{i\in s} w_i + \sum_{i\in b} w_i}$ where s and b refer to the signal and background event types and $w_i$ are the correspondent weights. Hence, pure background have zero purity.

nodes, weighted by their relative fraction of events. In principle the process may continue until every leaf contains only signal or background events, which corresponds to perfect discrimination. However, such a decision tree would be overtrained and it has to be pruned.
For simple decision trees, pruning has been found to be better than interrupting the node splitting at an earlier state: appareantly insignificant splits can prove to be good at a later stage down the tree. For BDTs pruning is instead unnecessary since the tree depth should be strongly more limited than any pruning algorithm would do.
Decision trees are insensitive to the inclusion of poorly discriminating input variables. While NNs deal more difficultly with such additional variables, the BDT training algorithm will ignore non-discriminating variables as, for each node splitting, only the best one is used. However, this technique is generally less performant with respect to others like neural networks, despite if its simplicity. However, in examples with more complex correlations or real life examples, the BDTs often outperform the other techniques. This may be due to a lack of available training events that would be needed by the other classifiers, or the optimal configuration of the neural network has not been specified.

In table 4.1 some properties of different MVA methods are compared.

| MVA method | | | | |
|---|---|---|---|---|
| Criteria | | Cuts | NN | BDT |
| Performance | Linear Correlation | $\star$ | $\star\star$ | $\star$ |
| Performance | Nonlinear Correlations | $\circ$ | $\star\star$ | $\star\star$ |
| Speed | Training | $\circ$ | $\star$ | $\star$ |
| Speed | Response | $\star\star$ | $\star\star$ | $\star$ |
| Robustness | Overtraining | $\star\star$ | $\star$ | $\star$ |
| Transparency | | $\star\star$ | $\circ$ | $\star$ |

Table 4.1: Assessment of MVA method properties. The symbols in the table stand for "good" ($\star\star$), "fair"( $\star$) and "bad" ($\circ$) attributes.

Simplicity has to be sought in separate considerations only if it not affects the discrimination power. For problems that require a high optimization degree and a large number of input variables, NN or BDTs are appropriate. In this thesis, for the $t\bar{t}H$ analysis at LHC, the BDT and MLP multivariate methods will be applied and compared, as presented in the next chapter.

# Chapter 5

# Multivariate Analysis on $t\bar{t}H$ Channel

The search for the Higgs boson is closely related to the study of the different background processes that overwhelm signal events.
The study of the $t\bar{t}H$ channel is particularly relevant since it provides a direct measurement of the top-Higgs Yukawa coupling, as well as the coupling of the Higgs boson with the particles produced in its decay. Despite the small production cross section for this channel, the $t\bar{t}H$ channel will gain, more than any other production channel, from the center of mass energy increase in Run II at LHC. Going from $\sqrt{s}$=8 TeV to 13 TeV, the production cross section of the $t\bar{t}H$ channel will in fact raise by a factor of four.
In 2012, with an integrated luminosity of 20.3 fb$^{-1}$ and a $t\bar{t}H$ production cross section of 130 fb at $\sqrt{s}$=8 TeV, the expected number of $t\bar{t}H$ events is of the order of magnitude of $10^3$. Although the reconstruction efficiency of the ATLAS detector provides high performances, different background processes can mimic the $t\bar{t}H$ signature in the final state. In this chapter, the study of the different multivariate methods for signal and background discrimination, as well as their performances, will be applied to the $t\bar{t}H$ channel and the response of the different MVA methods will be compared. Since the Run I analysis are still ongoing, the data in the signal region are "blinded". For that reason and due to the lack of statistics recorded till now, which does not permit precise measurements, in this analysis only MC samples have been used. This study will be tested on the real data that will be collected during LHC Run II, when the center of mass energy will reach $\sqrt{s}$=13 TeV. The decay channel under examination is the one with the multileptonic final state.

## 5.1 Signal and Background Description

The combination of the different decay modes for the $t\bar{t}$ pair and the Higgs boson leads to a complicated final signature with high jet multiplicity and at least two b-jets. The final state under study is characterized by the presence of two leptons, one produced by the $H \to WW^* \to l\nu qq$ decay and the second by the leptonic decay of one of the top quarks ($t \to Wb \to l\nu$+jets) as visible in the Feynman diagram in figure 5.1
The branching ratios for the $t\bar{t}$ pair is 46% in the full hadronic final state, 43.5% in the semileptonic channel and 10.3% in the dileptonic final state. The channel selected in this analysis searches for the semileptonic decay of the $t\bar{t}$ pair in order to have a clear signature from the high $p_T$ lepton.

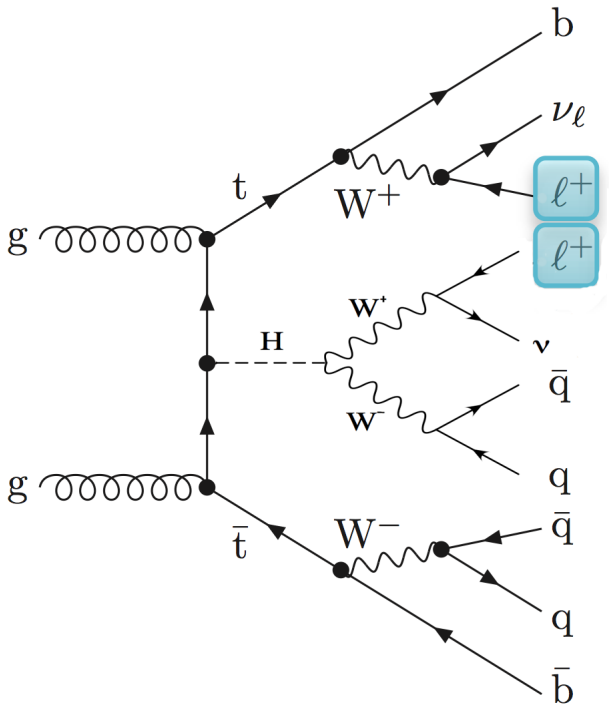



Figure 5.1: Feynman diagram representing the signature searched in the $t\bar{t}H$ channel final state. The final state requires the presence of two leptons (electrons or muons) with same or opposite charge, one arising from the W decay from the Higgs boson while the other is generated by the leptonic decay of one of the top quarks.

The lepton[1] pair in the final state can have same sign (SS), as depicted in

[1]From now on, "leptons" will stand for electrons or muons

figure 5.1, or opposite sign (OS). Since most of the background processes predict OS lepton pairs, the SS channel is expected to have the best signal over background ratio. Nevertheless, to get the maximum sensitivity for the analysis, the OS channel state will be considered as well. The Monte Carlo samples and the generators used in the analysis have already been described in section 3.4 and their features provided in table 3.4.

Background processes differently contaminate the signal region weather they have a SS or OS pair of leptons with same flavour in the final state. The main irreducible backgrounds for both OS and SS channels are the $t\bar{t}W$ and $t\bar{t}Z$ productions, with a 10-20% theoretical uncertanties on the production cross section for these processes. The W or Z boson can, in fact, decay into leptons and lead to the same signature of the signal.
Another important irreducible background arises from the diboson production ($W^+W^-$, $WZ$, $ZZ$ $W\gamma$..). These backgrounds can be however reduced by the requirement of a high jet multiplicity in the final state. Diboson production backgrounds are further reduced with the request of additionally jets. In addition, different reconstruction techniques can mitigate the effect of pile up in enhancing the jet multiplicity. $W^+W^-$ processes produce OS lepton pairs in the final state, while the other diboson processes produce SS pairs.
The main reducible background contribution in the SS channel comes from $t\bar{t}$ and single top production where the charge of one of the two leptons is mis-identified. The main source of electron charge misreconstruction is due to Bremsstrahlung: the interaction of the electron with the detector material produces a trident electrons ($e^{\pm} \rightarrow e^{\pm}\gamma^* \rightarrow e^{\pm}e^+e^-$) whose EM cluster is identified with the wrong electron track, leading to a misidentification of the charge. The fraction of trident electrons depends on the amount of material that electron traverses and, thus, on $\mid \eta \mid$. In addition, jets in any process could fake a lepton with any charge. W+jets events also represent thus one of the reducible backgrounds in the SS final state. For OS events, in addition to the above mentioned processes, also Z+jets events represent a huge background. This is, however, mitigated by a *Z veto* cut on the invariant mass of the same flavour OS lepton pairs.

### 5.1.1 Pre-Selection Criteria and Cutflow

The same pre-selection criteria have been applied to the signal and background MC samples. The final samples, resulted from the pre-selection requirements, have been used for the multivariate analysis as presented in the following sections. Multivariate methods should be in fact applied on a sub-

sample of events that have already passed some preselection. Since, however, MVA methods exploiting correlation among the input variables usually have better discriminating power that cut-based selection, the event pre-selection for MVA methods can thus be relaxed compared to the cut-based analysis. First of all, events must pass at least one of the single unprescaled lepton triggers: electron triggers have a $p_T$ threshold of 24 GeV and 60 GeV for respectively isolated and non isolated reconstructed tracks; muon triggers have a $p_T$ threshold of 24 GeV and 36 GeV for respectively isolated and non isolated reconstructed tracks. On the reconstructed objects the isolation requirement for both electrons and muons are loosen to $E_T^{cone}/p_T$=$p_T^{cone}/p_T$ <0.1 (the isolation requirement is $E_T^{cone}/p_T = p_T^{cone}/p_T$ <0.05 for the standard analysis). Three final channels are then considered in the analysis: events with in the final state two electron ($ee$), or two muons ($\mu\mu$), or an electron and a muon ($e\mu$). These three channels have been at first separately analyzed. Further pre-selection criteria applied to signal and background samples require that:

- the transverse momenta of the leading leptons, as well as the transverse momenta of the subleading ones, must be at least of 10 GeV;
- the events are divided into SS and OS cathegories. As already said, the inclusion of OS events further enhances the statistic of the signal sample, although a cut on OS event would greatly reduce most of background. Multivariate methods will be separately applied both on SS and OS samples;
- each event must contain at least two jets and at least one b-jet in the final state.

For OS events with leptons of the same flavours ($ee$ and $\mu\mu$) in the final state, the dilepton invariant mass $m_{ll}$ is calculated in order to suppress the contribution of the background from Z boson production. The invariant mass of the lepton pair must fall in fact outside the 81<| $m_{ll}$ |<101 GeV/$c^2$ mass range.
The number of MC generated events as well as the following cutflow selection are presented in table 5.1 for the $ee$ SS channel for both signal ($1^{st}$ column) and background samples normalized to a luminosity of 20.3 fb$^{-1}$. After the fulfillment of the different pre-selection cuts the largest contribution is due to $t\bar{t}$ events.
Table 5.2 shows the cutflow in the $\mu\mu$ SS channel. Also in this case the number of MC generated events and the following cutflow both for signal ($1^{st}$ column) and background are presented. At the final stage of the selection the largest contribution is due to $t\bar{t}$ events although it is visible that, with

| **Cutflow SS events $ee$ channel** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Signal | Background | | | | | | |
| | $t\bar{t}H$ | $t\bar{t}Z$ | $t\bar{t}W$ | $t\bar{t}$ | Diboson | Single top | W+jets | Z+jets |
| Generated | 16626±129 | 3333±58 | 4613±68 | $6.6{\cdot}10^6{\pm}2.6{\cdot}10^3$ | 86065±293 | 34383±185 | $6.19{\cdot}10^8{\pm}2.5{\cdot}10^4$ | $4.7{\cdot}10^7{\pm}\ {\cdot}6.9^3$ |
| N jets>2 | 6057±78 | 1115±33 | 1799±42 | $2.63{\cdot}10^6{\pm}1.6{\cdot}10^3$ | 5861±77 | 1092±33 | $9.7{\cdot}10^6{\pm}3.1{\cdot}10^3$ | $2.3{\cdot}10^6\ {\pm}1.5{\cdot}10^3$ |
| N bjets>1 | 5428± 74 | 946±31 | 1496±39 | $2.19{\cdot}10^6\ {\pm}\ 1.5{\cdot}10^3$ | 306±17 | 9158±96 | 567946±754 | 98766±314 |
| 2 leptons $p_T$ >10 GeV | 139±12 | 51±7 | 46±7 | 82997±288 | 38±6 | 1.6±1.3 | 39±6 | 15107±123 |
| SS events | 28±5 | 5±2 | 15±4 | 583±24 | 8±3 | 0.74±0.86 | 23±5 | 110±10 |

Table 5.1: Cutflow for SS events in the $ee$ channel. The number of MC generated and the following cutflow events are presented for signal ($1^{st}$ column) and backgrounds normalized to $\mathcal{L}$= 20.3 fb$^{-1}$. After the fulfillment of the different pre-selection cuts the largest contribution is due to $t\bar{t}$ same sign events. Only statistical errors are shown.

respect to SS sample in the $ee$ channel, the number of $t\bar{t}$ events in the $\mu\mu$ channel is reduced since the charge flip probability for muons is negligible.

| **Cutflow SS events $\mu\mu$ channel** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Signal | Background | | | | | | |
| | $t\bar{t}H$ | $t\bar{t}Z$ | $t\bar{t}W$ | $t\bar{t}$ | Diboson | Single top | W+jets | Z+jets |
| Generated | 16626±129 | 3333±58 | 4613±68 | $6.6{\cdot}10^6{\pm}2.6{\cdot}10^3$ | 86065±293 | 34383±185 | $6.19{\cdot}10^8{\pm}2.5{\cdot}10^4$ | $4.7{\cdot}10^7{\pm}\ {\cdot}6.9^3$ |
| N jets>2 | 6057±78 | 1115±33 | 1799±42 | $2.63{\cdot}10^6{\pm}1.6{\cdot}10^3$ | 5861±77 | 1092±33 | $9.7{\cdot}10^6{\pm}3.1{\cdot}10^3$ | $2.3{\cdot}10^6\ {\pm}1.5{\cdot}10^3$ |
| N bjets>1 | 5428± 74 | 946±31 | 1496±39 | $2.19{\cdot}10^6\ {\pm}\ 1.5{\cdot}10^3$ | 306±17 | 9158±96 | 567946±754 | 98766±314 |
| 2 leptons $p_T$ >10 GeV | 195±14 | 66±8 | 64±8 | 122811±350 | 45±7 | 1.3±1.1 | 16±4 | 22168±149 |
| SS events | 38±6 | 5±2 | 22±5 | 233±15 | 8±3 | 0.71±0.84 | 11±3 | 0 |

Table 5.2: Cutflow for SS events in the $\mu\mu$ channel. The number of MC generated and the following cutflow events are presented both for signal ($1^{st}$ column) and backgrounds normalized to $\mathcal{L}$=20.3 fb$^{-1}$. After the fulfillment of the different pre-selection cuts the largest contribution is due to $t\bar{t}$ events. The errors associated to the event numbers are statistical only.

Finally, table 5.3 shows the cutflow in the $e\mu$ SS channel.

| **Cutflow SS events $e\mu$ channel** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Signal | Background | | | | | | |
| | $t\bar{t}H$ | $t\bar{t}Z$ | $t\bar{t}W$ | $t\bar{t}$ | Diboson | Single top | W+jets | Z+jets |
| Generated | 16626±129 | 3333±58 | 4613±68 | $6.6{\cdot}10^6{\pm}2.6{\cdot}10^3$ | 86065±293 | 34383±185 | $6.19{\cdot}10^8{\pm}2.5{\cdot}10^4$ | $4.7{\cdot}10^7{\pm}\ {\cdot}6.9^3$ |
| N jets>2 | 6057±78 | 1115±33 | 1799±42 | $2.63{\cdot}10^6{\pm}1.6{\cdot}10^3$ | 5861±77 | 1092±33 | $9.7{\cdot}10^6{\pm}3.1{\cdot}10^3$ | $2.3{\cdot}10^6\ {\pm}1.5{\cdot}10^3$ |
| N bjets>1 | 5428± 74 | 946±31 | 1496±39 | $2.19{\cdot}10^6\ {\pm}\ 1.5{\cdot}10^3$ | 306±17 | 9158±96 | 567946±754 | 98766±314 |
| 2 leptons $p_T$ >10 GeV | 322±18 | 51±7 | 111±11 | 204276±452 | 30±4 | 3 ±2 | 113±11 | 148±12 |
| SS events | 66±8 | 10±3 | 37±6 | 858±29 | 14±4 | 1±1 | 83±9 | 7±3 |

Table 5.3: Cutflow for SS events in the $e\mu$ channel. The number of generated and the following cutflow events are presented both for signal ($1^{st}$ column) and backgrounds normalized to $\mathcal{L}$=20.3 fb$^{-1}$. After the fulfillment of the different pre-selection cuts the largest contribution is due to $t\bar{t}$ same sign events. The errors associated to the event numbers are statistical only.

The pre-selection criteria have been applied on the OS sample as well. The resulting number of events at the MC generation level and after the pre-selection cuts are reported in tables 5.4, 5.5 and 5.6 for $ee$, $\mu\mu$ and $e\mu$ OS

channels respectively.
The errors associated to the values obtained from the cutflow are statistical only.

| **Cutflow OS events $ee$ channel** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Signal | Background | | | | | | |
| | $t\bar{t}H$ | $t\bar{t}Z$ | $t\bar{t}W$ | $t\bar{t}$ | Diboson | Single top | W+jets | Z+jets |
| Generated | 16626±129 | 3333±58 | 4613±68 | $6.6{\cdot}10^6{\pm}2.6{\cdot}10^3$ | 86065±293 | 688273±830 | $6.19{\cdot}10^8{\pm}2.5{\cdot}10^4$ | $4.7{\cdot}10^7{\pm}\ 2.2{\cdot}7^3$ |
| N jets>2 | 6057±78 | 1115±33 | 1799±42 | $2.63{\cdot}10^6{\pm}1.6{\cdot}10^3$ | 5861±77 | 221130±470 | $9.7{\cdot}10^6{\pm}3.1{\cdot}10^3$ | $2.3{\cdot}10^6\ {\pm}1.5{\cdot}10^3$ |
| N bjets>1 | 5428± 74 | 946±31 | 1496±39 | $2.19{\cdot}10^6\ \pm\ 1.48{\cdot}10^3$ | 306±17 | 163620±404 | 567946±754 | 98766±314 |
| 2 leptons $p_T$ >10 GeV | 139±12 | 51±7 | 46±7 | 82997±288 | 38±6 | 746±27 | 39±6 | 15107±123 |
| SS events | 111±11 | 47±7 | 31±6 | 82459±288 | 30±5 | 723±27 | 15±4 | 14996±122 |

Table 5.4: Cutflow for OS events in the $ee$ channel. The number of MC generated and the following cutflow events are presented both for signal ($1^{st}$ column) and backgrounds normalized to $\mathcal{L}$=20.3 fb$^{-1}$. After the fulfillment of the different pre-selection cuts, the largest contribution is due to $t\bar{t}$ events. Only statistical errors are shown.

| **Cutflow OS events $\mu\mu$ channel** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Signal | Background | | | | | | |
| | $t\bar{t}H$ | $t\bar{t}Z$ | $t\bar{t}W$ | $t\bar{t}$ | Diboson | Single top | W+jets | Z+jets |
| Generated | 16626±129 | 3333±58 | 4613±68 | $6.6{\cdot}10^6{\pm}2.6{\cdot}10^3$ | 86065±293 | 688273±830 | $6.19{\cdot}10^8{\pm}2.5{\cdot}10^4$ | $4.7{\cdot}10^7{\pm}\ 2.2{\cdot}7^3$ |
| N jets>2 | 6057±78 | 1115±33 | 1799±42 | $2.63{\cdot}10^6{\pm}1.6{\cdot}10^3$ | 5861±77 | 221130±470 | $9.7{\cdot}10^6{\pm}3.1{\cdot}10^3$ | $2.3{\cdot}10^6\ {\pm}1.5{\cdot}10^3$ |
| N bjets>1 | 5428± 74 | 946±31 | 1496±39 | $2.19{\cdot}10^6\ \pm\ 1.48{\cdot}10^3$ | 306±17 | 163620±404 | 567946±754 | 98766±314 |
| 2 leptons $p_T$ >10 GeV | 195±14 | 66±8 | 64±8 | 122811±350 | 45±7 | 1034±32 | 15±4 | 22168±149 |
| SS events | 157±13 | 61±8 | 42±6 | 122578±350 | 37±6 | 1010±32 | 5±2 | 22168±149 |

Table 5.5: Cutflow for OS events in the $\mu\mu$ channel. The number of generated and the following cutflow events are presented both for signal ($1^{st}$ column) and backgrounds normalized to $\mathcal{L}$=20.3 fb$^{-1}$. After the fulfillment of the different pre-selection cuts the largest contribution is due to $t\bar{t}$ events. The errors associated to the event numbers are statistical only.

| **Cutflow OS events $e\mu$ channel** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Signal | Background | | | | | | |
| | $t\bar{t}H$ | $t\bar{t}Z$ | $t\bar{t}W$ | $t\bar{t}$ | Diboson | Single top | W+jets | Z+jets |
| Generated | 16626±129 | 3333±58 | 4613±68 | $6.6{\cdot}10^6{\pm}2.6{\cdot}10^3$ | 86065±293 | 688273±830 | $6.19{\cdot}10^8{\pm}2.5{\cdot}10^4$ | $4.7{\cdot}10^7{\pm}\ 2.2{\cdot}7^3$ |
| N jets>2 | 6057±78 | 1115±33 | 1799±42 | $2.63{\cdot}10^6{\pm}1.6{\cdot}10^3$ | 5861±77 | 221130±470 | $9.7{\cdot}10^6{\pm}3.1{\cdot}10^3$ | $2.3{\cdot}10^6\ {\pm}1.5{\cdot}10^3$ |
| N bjets>1 | 5428± 74 | 946±31 | 1496±39 | $2.19{\cdot}10^6\ \pm\ 1.48{\cdot}10^3$ | 306±17 | 163620±404 | 567946±754 | 98766±314 |
| 2 leptons $p_T$ >10 GeV | 321±18 | 51±7 | 111±11 | 204276±452 | 30±5 | 1797±42 | 114±11 | 148±12 |
| SS events | 255±16 | 40±6 | 74±9 | 203418±451 | 15±4 | 1748±42 | 31±6 | 141±12 |

Table 5.6: Cutflow for OS events in the $e\mu$ channel. The number of MC generated and the following cutflow events are presented both for signal ($1^{st}$ column) and backgrounds normalized to $\mathcal{L}$=20.3 fb$^{-1}$. After the fulfillment of the different pre-selection cuts, the largest contribution is due to $t\bar{t}$ events. The errors associated to the event numbers are statistical only.

In the same sign channel, the applied cuts manage to suppress a large contribution of almost all the different background processes, as visible in tables 5.1, 5.2 and 5.3. The situation is quite different in the opposite sign channel,

since most of the background processes produce OS leptons in the final state. The signal over background ratio, in fact, is larger in the SS channel than in the OS one. The aim of this analysis is to provide a method which can achieve better performances on signal/background discrimination. The standard cut-based analysis for the $t\bar{t}H$ channel, at the moment, only requires SS events in the final state because of the large background contamination in OS events. This study presents a preliminar study also on OS events since, if the MVA techniques should provide a good signal from background discrimination, it does not seem unrealistic to reinclude also different flavour OS lepton pairs in the standard analysis, allowing thus an increase in the sensitivity of the $t\bar{t}H$ search.

## 5.2 Principal Observables and Shapes

The first step in MVA analysis is to identify a set of input variables representative of the event. These variables should be able to discriminate signal events from the various background processes. The selected variables are listed below, in a decreasing order according to their single "approximate"[2] discriminating power. As will be soon explained, dividing the samples according to the lepton flavours in the final state is not necessary. The analysis has been performed using sixteen variables:

- $p_{T1}$ and $p_{T2}$: transverse momentum of the leading and subleading lepton;
- $N_{jets}$ and $N_{bjets}$: number of reconstructed jets and b-jets passing the cutflow and b-tagging criteria;
- $p_{T1}^{bjet}$ and $\eta^{bjet}$: transverse momentum and pseudorapidity of the selected b-jet;
- $H_t$: scalar sum of lepton and jet transverse momenta;
- $\Delta R$: distance between every jet pair ($\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$); $\Delta R_{max}$: maximum $\Delta R$ event by event; $\Delta R_{max}^{p_T}$ (distance between the two most energetic jets in the event) has also been used;
- $m_{jj}$: invariant mass calculated for the two most energetic jets in the event;

[2]The term "approximate" refers to the fact that these variables have been selected by looking at their distributions when superimposing signal and background ones.

- $p_T^{3jet}$: transverse momentum of the third, if present, most energetic jet of the event;
- $\Delta\eta_{max}$: maximum pseudorapidity difference calculated between every pair of jets present in the event.

Finally, three variables related to the shape of the hadronic part of the event have been used: *sphericity*, *centrality* and *aplanarity*. Event shape variables are an ideal way to help in distinguishing between the actual signal of a process and the backgrounds. Sphericity, for instance, is defined as a measure of the summed squared transverse momenta with respect to the event axis. It is deduced by defining the sphericity tensor as:

$$S^{\alpha\beta} = \frac{\sum_i p_i^\alpha p_i^\beta}{\sum_i \mid p_i \mid^2} \tag{5.1}$$

where $\alpha, \beta$=1,2,3 are matrices constructed by using the x,y and z components of the momentum vector of the jet. The normalized eigenvalues of these matrices can be found and they have to satisfy the normalization conditions: $\lambda_1 \geq \lambda_2 \geq \lambda_3$ and $\lambda_1 + \lambda_2 + \lambda_3$=1. Sphericity can be thus defined as a linear combination of two of the three eigenvalues:

$$S = \frac{3}{2}(\lambda_2 + \lambda_3). \tag{5.2}$$

A value of S very close to 1 characterizes an isotropic event (where no jet is produced) in the detector reference frame while S$\simeq$0 classifies the event as *dijet-like*.
Aplanarity can be used to distinguish spherical from planar and linear events. It is defined as:

$$A = \frac{3}{2}\lambda_3 \tag{5.3}$$

1.5 times the smallest eigenvalue of the momentum tensor. The expected values for aplanarity are A$\simeq$0 for an event which lies on a plane and A$\simeq$0.5 for an isotropic one.
The last event shape variable used for the analysis is centrality. It can be defined as:

$$C = \frac{\sum_i p_{Ti}}{\sum_i E_i} \tag{5.4}$$

where $p_{Ti}$ is the jet transverse momentum and $E_i$ is the measured energy of all the jets of the event. This variable thus contains information about the relative locations of particles, as well as their location relative to the transverse direction.

The main observables described above have been used for the parametrization of the multivariate analysis.
Each variable has been first studied separately for the three $ee$, $\mu\mu$ and $e\mu$ channels both for signal and for background processes and then compared. Figures 5.2 and 5.3 show the event shape of eight of the sixteen input variables used for the analysis respectively for the SS and OS channel. The three cathegorizations according to the lepton flavour composition are superimposed while no division between signal and backgrounds has been applied. More variable distributions can be found in figures 20 and 21 reported in appendix 5.5.1. All the distributions present no remarkable differences in the event shape for the $ee$, $\mu\mu$ and $e\mu$ channels. Consequently, the multivariate analysis have been performed without this further classification.

Figure 5.4 and 5.5 show respectively for SS and OS states, the eight most discriminating input variables distributions for $t\bar{t}H$ signal (black dots) and different background processes (coloured lines). In all the plots the events have been weighted according to cross sections of the processes and normalized to the 2012 integrated luminosity of 20.3 fb$^{-1}$. Since no difference have been found in the division on the three $ee$, $\mu\mu$ and $e\mu$ channels, only the distribution for $ee$ and $e\mu$ cases are respectively presented. More distributions can be found in figure 22 and 23 reported in Appendix. According to these distributions, some of the variables present high discriminating power. For example, $H_T$ results quite powerful, especially in the OS case in discriminating diboson production, $t\bar{t}$, Z+jets backgrounds and signal events. Overmore, the number of jet Num$_{jet}$ also provides good discrimination of the signal events from reducible backgrounds while, however, irreducible backgrounds with $t\bar{t}$ pair production also presents a high jet multiplicity which can mimic the signal one.

## 5.3 Multivariate Analysis

### 5.3.1 Input Variables and Correlations

One potential drawback of multivariate methods could be that they underperform in presence of strong correlations between input variables that may result in performance losses. Consequently, before applying the different multivariate methods to the input variables, the correlations between input variables have been evaluated.

Figure 5.6 show the correlation matrix for signal (top plot) and background

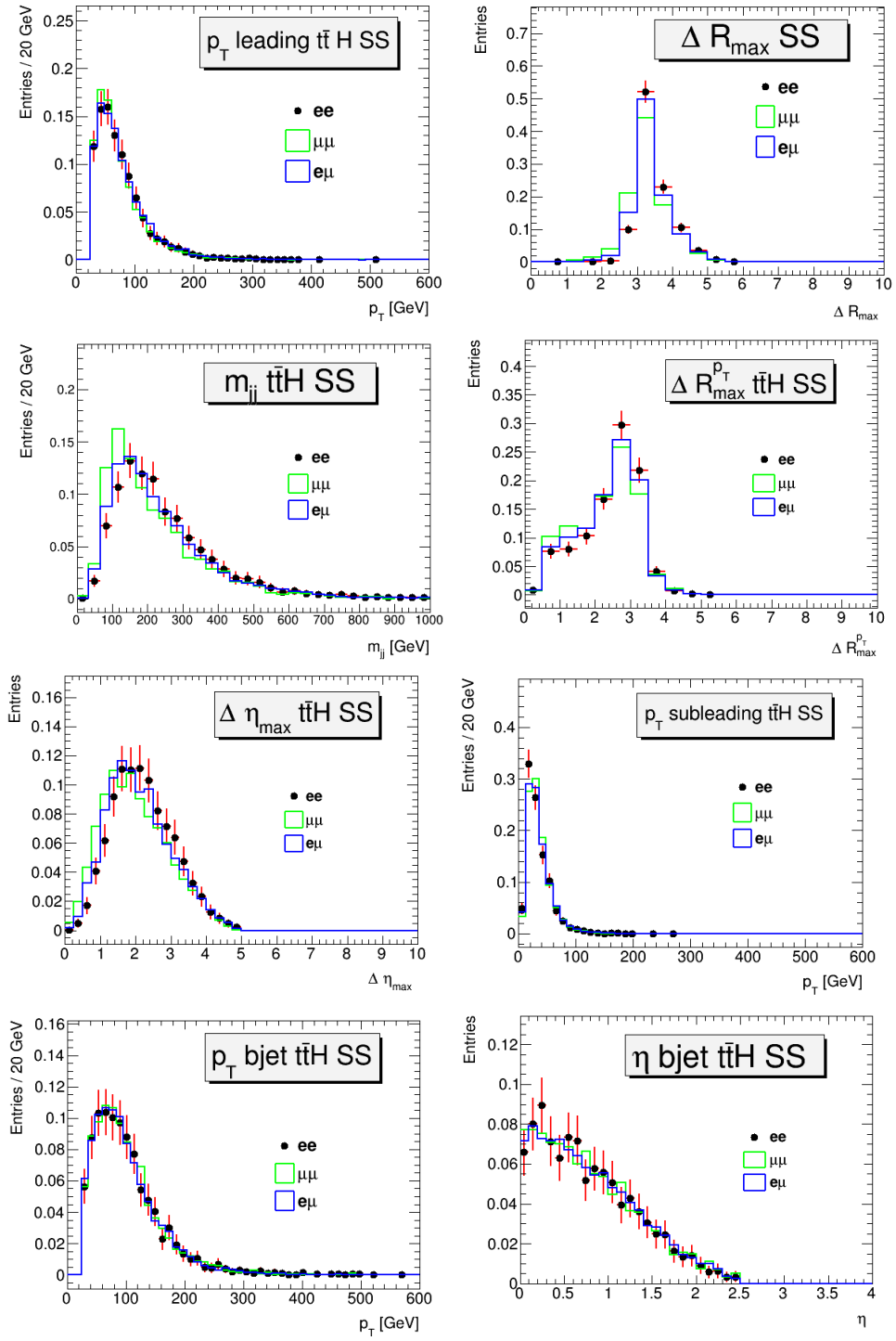



Figure 5.2: Variable distributions for eight of the sixteen variables used in the MVA analysis. Each distribution shows the superimposition of the $ee$ (black dots), $\mu\mu$ (green line) and $e\mu$ (blue line) samples for the $t\bar{t}H$ channel in SS case without distinction in background or signal events. No remarkable differences are present in the shapes of the three different channels. The events have been weighted according to the cross sections of the processes and normalized to the 2012 integrated luminosity of 20.3 fb$^{-1}$.

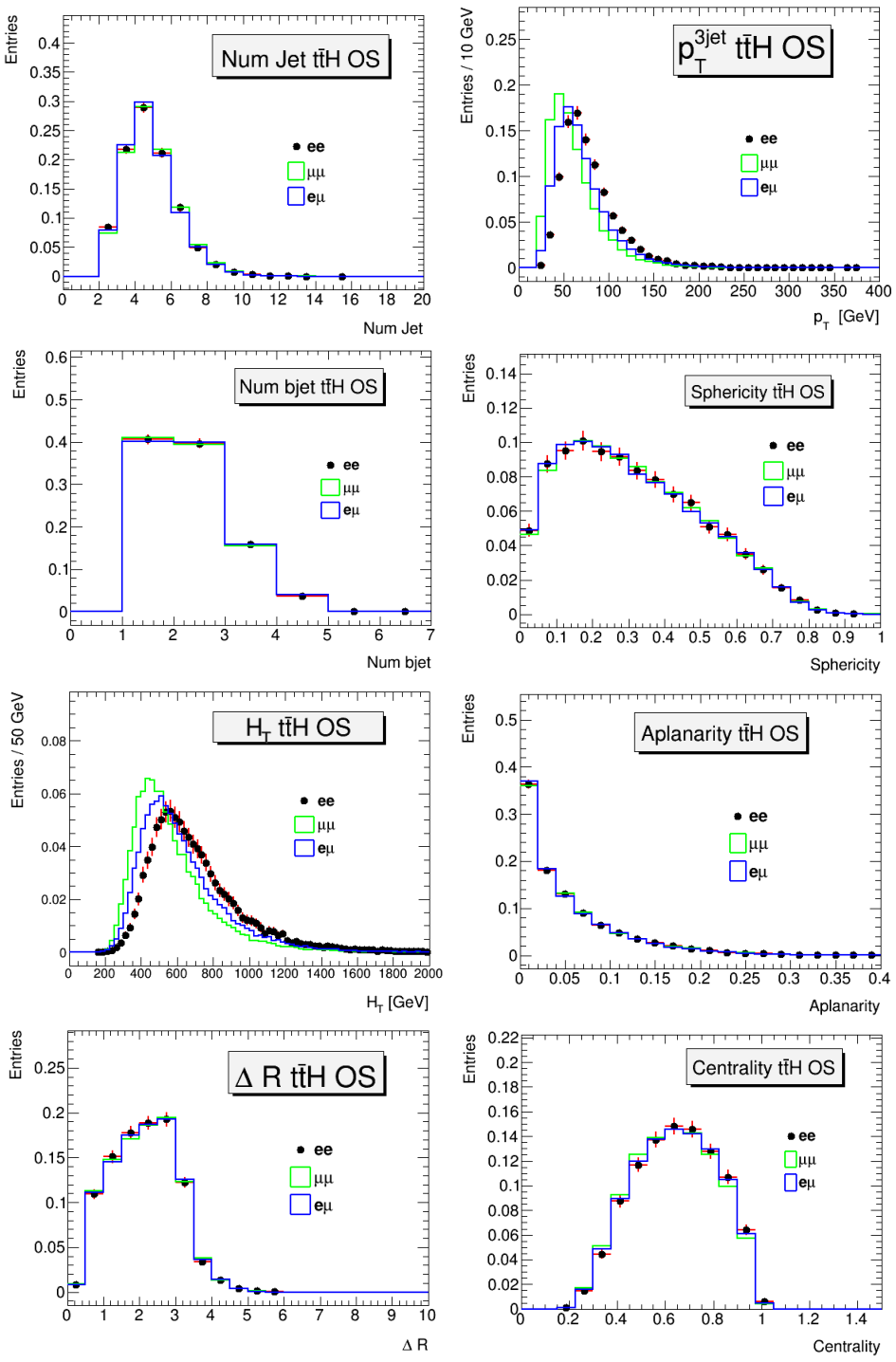



Figure 5.3: Variable distributions for eight of the sixteen variables used in the MVA analysis. Each distribution shows the superimposition of the $ee$ (black dots), $\mu\mu$ (green line) and $e\mu$ (blue line) samples for the $t\bar{t}H$ channel in the OS case without classification in background or signal events. No remarkable differences are present in the shapes of the three different channels. The events have been weighted according to the cross sections of the different processes and normalized to the 2012 integrated luminosity of 20.3 fb$^{-1}$.

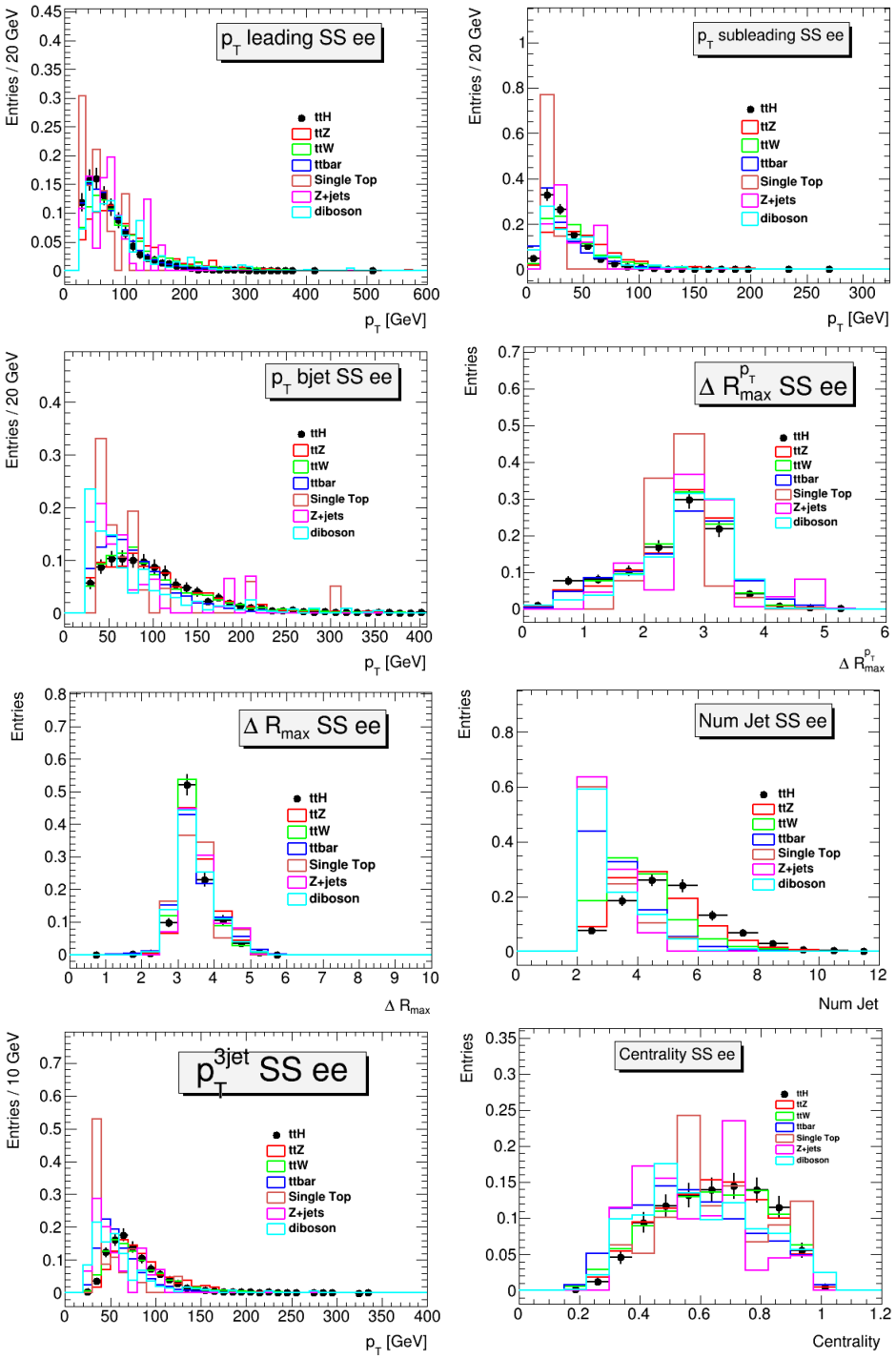



Figure 5.4: Variable distributions for the eight most discriminating input variables for the SS case in the *ee* channel. The distributions are made superimposing the $t\bar{t}H$ signal (black dots) and the different background processes (coloured lines). The events have been weighted according to the cross sections of the processes and normalized to the 2012 integrated luminosity of 20.3 fb$^{-1}$.

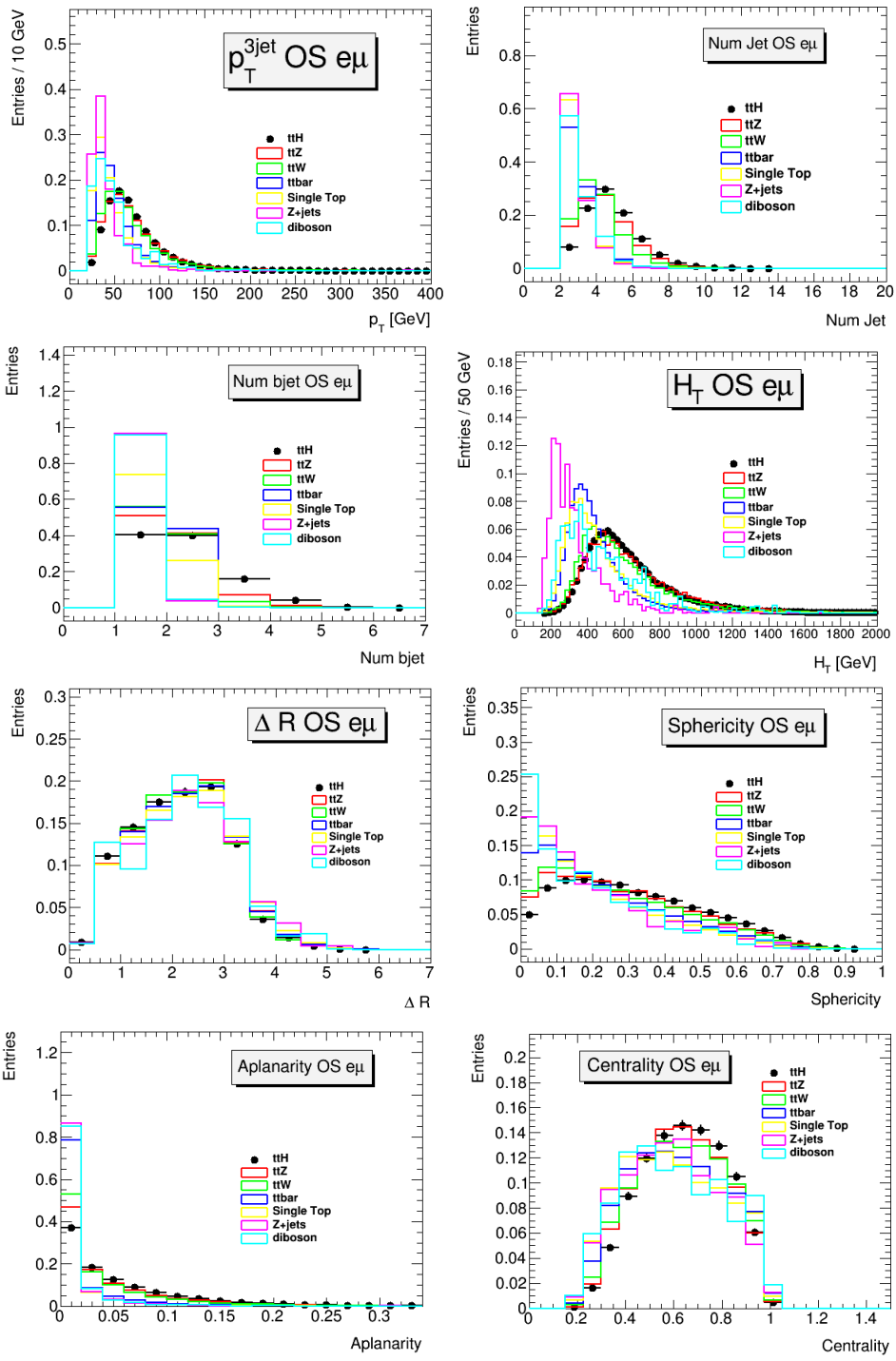



Figure 5.5: Variable distributions for the eight most discriminating input variables for the OS case in the $e\mu$ channel. The distributions are made superimposing the $t\bar{t}H$ signal (black dots) and the different background processes (coloured lines). The events have been weighted according to the cross sections of the processes and normalized to the 2012 integrated luminosity of 20.3 fb$^{-1}$.

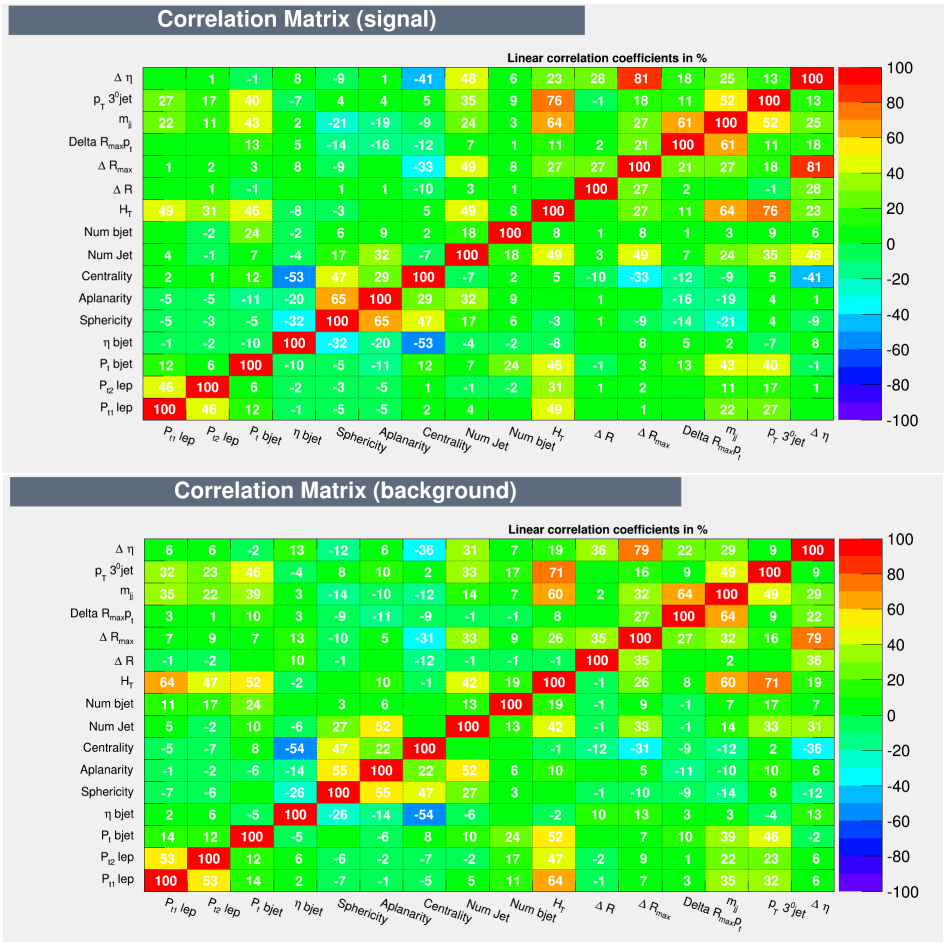



Figure 5.6: Correlation matrix for signal (top plot) and background (bottom plot) events computed for the sixteen MVA input variables. Linear coefficients among the different variables, for signal events, are for most pairs of variables smaller than 40%, with exception for those variables which are derived from one another. For example, $H_T$ is strongly correlated to those variables, such as the transverse momenta of the leptons and of jets, which define it, as expected. Sphericity and aplanarity present also higher correlations, as expected since they are both defined as a linear combination of the momentum matrix eigenvalues. Linear coefficients for background events present the same features seen for the signal sample but with lower values.

(bottom plot). The linear correlation coefficients have been computed for all the combination of the sixteen variables used as inputs for MVA methods.

Linear coefficients among different variables, for signal events, are in average inferior to the 40%, with expected exception for those variables which are derived from one another. $H_T$, for example, results strongly correlated to those variables, such as the tranverse momenta of leptons, jets and bjets, which enter the sum. Furthermore, sphericity and aplanarity present also higher correlations, as expected since they are both defined as a linear combination of the momentum matrix eigenvalues. For background events, linear coefficients present the same features as signal ones but with lower values. Tables for linear correlation coefficients for signal and background are reported in Appendix.
All the input variables have been normalized: maximum and minimum values for the variables to be transformed are determined from the training events and used to constrain linearly the input variables in the range [-1,1]. Such transformation is useful to simplify and make more effective the training phases of MVA methods.

## 5.4 Performance of the MVA Methods

In this section, a comparison of the different MVA methods applied separately on SS and OS events will be provided. The total number of input events for signal and background are respectively 212958 and 3438740. The training has been performed by assigning an average weight of one per event for signal and imposing the sum of weights for background equal to the sum of weights for signal. This results in 1000 events for signal and 1000 for background which, although representing an unphysical situation, do not affect the relative performance of the different MVA methods. Both for signal and for backgrounds, the number of events has been equally divided into two different samples, one for training and the second for testing. During the testing phase, signal and background events have been considered as a unique mixed sample.
The MVA methods have then been tested or using a single background sample containing all the processes or dividing the backgrounds into irreducible (indicated with "Irr" and including $t\bar{t}$, single top, W+jets and Z+jets production) and reducible (indicated with "Red" and consisting of $t\bar{t}V$ and diboson production) ones.

### 5.4.1 Boosted Decision Tree

As already explained in section 4.2.3, a BDT is a forest of decision trees where repeated left/right decisions are taken on one single variable at time

until a stopping criterion is fulfilled. The phace space is thus divided into many regions that are eventually classified as signal or background.
The result of an individual classifier is $t(\mathbf{x})$, with $\mathbf{x}$ the data set used for training, $t(\mathbf{x}) = +1$ for signal and -1 for background. The main parameters which can be tuned for BDT are the number of trees in the forest, the number of input variables, the separation type, the number of points in variable range used in finding optimal cut in node splitting. These parameters have been changed and trained on SS and OS samples, in order to find which combination of them results in the largest improvement in signal over background $S/\sqrt{S+B}$ significance.
BDT has been trained using three different groups of input variables:

- all the sixteen input variables described in previous sections;
- 11, randomly chosen, of the sixteen input variables;
- discarding the five more correlated variables ($m_{jj}$, $\Delta R_{max}$, centrality, aplanarity and $p_T^{3jet}$) according to figure 5.6.

Figure 5.7 shows the BDT response for different parameter settings for both opposite sign (left) and same sign events (right). BDT has been tested using 50, 100 and 150 decision trees while the number of variables randomly reduced from 16 to 11.
Tables 5.7 report the BDT performances obtained using the different settings, the first one for OS while the second one for SS sample. For each configuration of the parameters the corresponding value of the optimal cut for signal from background discrimination is provided, as well as the corresponding signal efficiency $\epsilon_S$, background rejection $r_B$ and signal over background $S/\sqrt{S+B}$ significance. The errors associated to $\epsilon_S$ and $r_B$ are the statistical errors associated to a binomial distribution. The results show an increase in the significance performance with a growing number of trees and of input variables. Figure 5.8 shows signal efficiency (blue continuous line), background efficiency (red line), signal purity (blue dashed line), significance (green line) and signal efficiency times purity (blue dotted line) for the different parameter settings. It is clearly visible that the significance reaches a maximum value and increases as the number of input variables, as well as of decision trees, raises.
In order to verify a possible discriminating power loss in presence of high correlations between variables, a test has been performed with the exclusion of the 5 more correlated variables from the training. Figure 5.9 presents in the top plots the BDT responses while in the bottom plots the signal efficiency (blue line), background efficiency (red line), signal purity (blue dashed line),

significance (green line) and signal efficiency times purity (blue dotted line) on the left for OS events and on the right for SS ones. The response of the BDT classifier is presented for 150 trees in the training with the exclusion of the most correlated variables. The last rows of $1^{st}$ and $2^{nd}$ 5.7 tables report the numerical results obtained for not correlated variables for OS and SS channels respectively. The comparison of these results with the one with full sample of variable shows no loss in the BDT performances due to the presence of correlated variables for both channels.
Finally, another test have been performed only in the SS channel in order to control how BDT performances change when background events in the training and test sample are sorted into reducible and irreducible backgrounds. Figure 5.10 presents in the top plots the BDT responses while in the bottom plots the signal efficiency (blue line), background efficiency (red line), signal purity (blue dashed line), significance (green line) and signal efficiency times purity (blue dotted line) on the left for signal and reducible backgrounds events and on the right for irreducible ones. The BDT performances show better discriminating power when trained with reducible, with respect to irreducible, backgrounds. The $3^{rd}$ and $4^{th}$ tables in 5.7 show signal over background significance provided by the BDT classifier for same sign events divided into reducible (up) and irreducible backgrounds (down). The optimal cut for signal from background discrimination is provided, as well as the corresponding signal efficiency $\epsilon_S$ and background rejection $r_B$. The results show that the signal over background ratio is higher for a background sample which contains reducible backgrounds only.
The performances of the different BDT classifier have been finally compared studying the background rejection versus signal efficiency curves. Figure 5.11 shows the background rejection versus signal efficiency curve obtained using the different parameter settings for the BDT method for OS (top plot) and SS (bottom plot) events. For OS events, all the different BDT configurations show similar shapes in the signal from background discrimination. For same sign events, the BDT performances increase for a data sample which containes only reducible background processes, as already mentioned.

| BDT Performances OS for 1000 Signal and Background Events | | | | |
|---|---|---|---|---|
| Parameters | Optimal Cut | $\epsilon_S$ | $r_B$ | $S/\sqrt{S+B}$ |
| 50 Trees, 16 vars | -0.128 | 0.845 ± 0.011 | 0.732 ± 0.014 | 25.38 |
| 100 Trees, 16 vars | -0.1061 | 0.852 ± 0.03 | 0.729 ± 0.014 | 25.42 |
| 100 Trees, 11 vars | -0.1135 | 0.853 ± 0.011 | 0.725 ± 0.014 | 25.40 |
| 150 Trees, 16 vars | -0.0673 | 0.845 ± 0.011 | 0.745 ± 0.014 | 25.48 |
| 150 Trees, No corr | -0.0812 | 0.846 ± 0.011 | 0.739±0.014 | 25.42 |
| **BDT Performances SS on 1000 Signal and Background Events** | | | | |
| 50 Trees, 16 vars | -0.1241 | 0.850 ± 0.011 | 0.659 ± 0.015 | 24.62 |
| 100 Trees, 16 vars | -0.0806 | 0.851 ± 0.011 | 0.6584 ± 0.015 | 24.63 |
| 100 Trees, 11 vars | -0.0731 | 0.850 ± 0.011 | 0.658 ± 0.015 | 24.61 |
| 150 Trees, 16 vars | -0.0676 | 0.861 ±0.011 | 0.639 ± 0.015 | 24.63 |
| 150 Trees, No corr | -0.0581 | 0.846 ± 0.011 | 0.661 ± 0.015 | 24.58 |
| **BDT Performances SS on 1000 Signal and Reducible Background Events** | | | | |
| 150 Trees, 16 vars | -0.0645 | 0.869 ± 0.011 | 0.718 ±0.014 | 25.61 |
| **BDT Performances SS on 1000 Signal and Irreducible Background Events** | | | | |
| 150 Trees, 16 vars | -0.3302 | 0.984 ± 0.005 | 0.0756 ± 0.008 | 22.53 |

Table 5.7: **$1^{st}$ and $2^{nd}$ tables**: Performances of the BDT classifier as a function of the number of trees and of input variables for opposite sign ($1^{st}$ table) and same sign ($2^{nd}$ table) events when trained with 1000 signal and 1000 background events. For each variation of the parameters the corresponding value of the optimal cut for signal from background discrimination is provided, as well as the corresponding signal efficiency $\epsilon_S$, background rejection $r_B$ and signal over background $S/\sqrt{S+B}$ significance. The reported values are referred to the BDT cut for which the significance reaches a maximum. The results show an increase in the significance performance with a growing number of trees and of input variables without loss of discrimination power excluding the most correlated variables from the training.
**$3^{rd}$ and $4^{th}$ tables:** BDT classifier performances for same sign events divided into reducible ($3^{rd}$ table) and irreducible backgrounds ($4^{th}$ table). The optimal cut for signal from background discrimination is provided, as well as the corresponding signal efficiency $\epsilon_S$, background rejection $r_B$ and signal over background $S/\sqrt{S+B}$ significance. The results show that the classifier works better on a background sample which contains reducible backgrounds only.
For each table, the errors associated to $\epsilon_S$ and $r_B$ are the statistical errors associated to a binomial distribution.

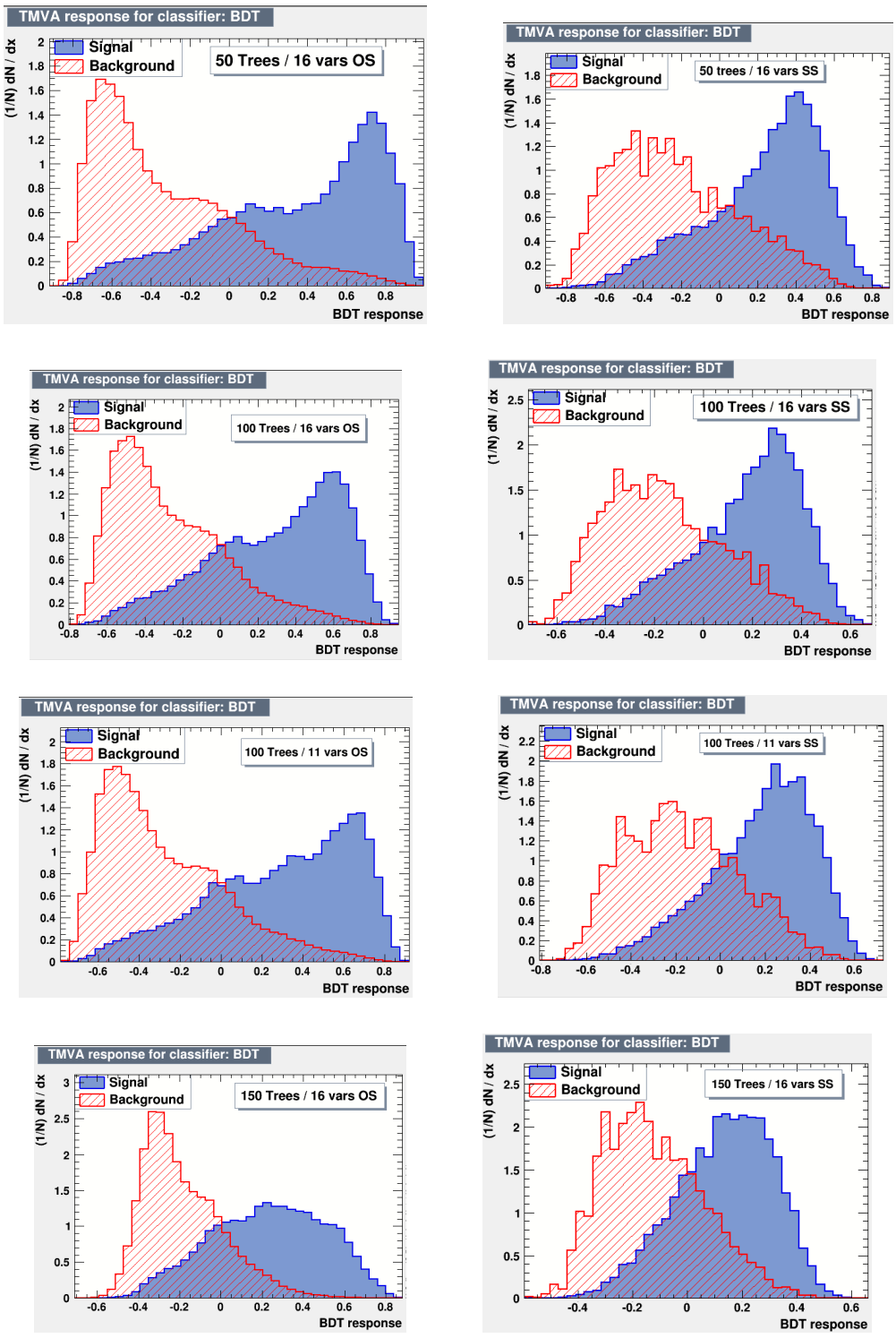



Figure 5.7: BDT response for different parameter settings for both opposite sign (left) and same sign (right) events. BDT has been tested using 50, 100 and 150 decision trees while the number of variables reduced randomly from 16 to 11.

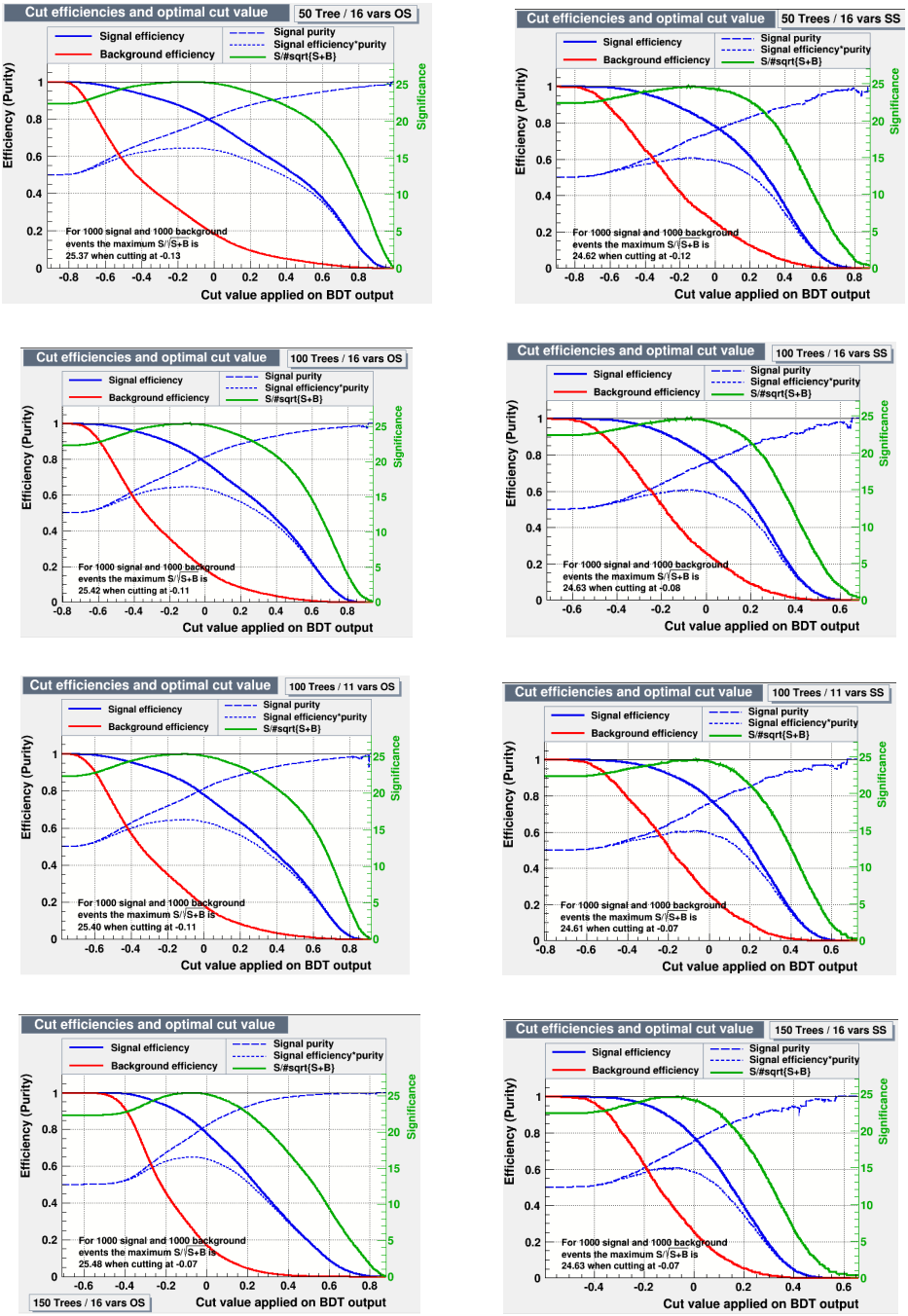



Figure 5.8: Signal efficiency (blue continuous line), background efficiency (red line), signal purity (blue dashed line), significance (green line) and signal efficiency times purity (blue dotted line) for the different BDT parameter settings. It is clearly visible that the significance reaches a maximum value and increases as the number of input variables, as well as of decision trees, arises. Significance reaches a maximum value for 150 trees and 16 input variables.

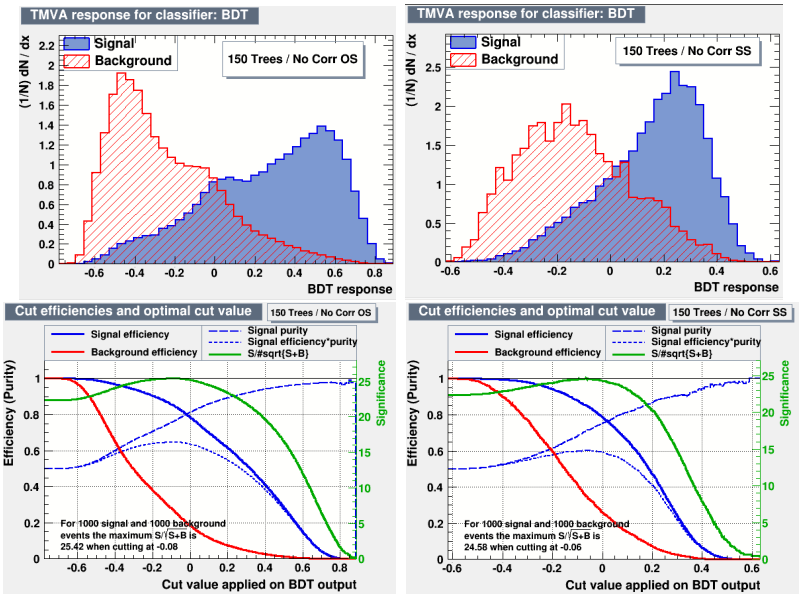



Figure 5.9: BDT responses, signal (blue continuous line) and background (red line) efficiencies, signal purity (blue dashed line), signal efficiency times purity (blue dotted line) and significance (green line) for OS (left plots) and SS (right plots) for a classifier with 150 tree in which the most correlated input variables have been excluded.

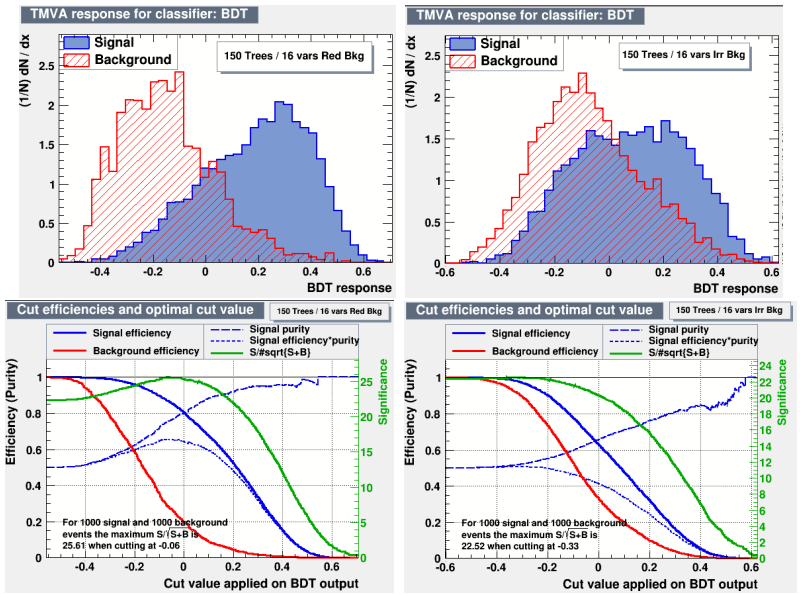



Figure 5.10: BDT responses, signal (blue continuous line) and background (red line) efficiencies, signal purity (blue dashed line), signal efficiency times purity (blue dotted line) and significance (green line) presented for reducible (left plots) and irreducible (right plots) backgrounds for SS events. The BDT performances show better discriminating power when trained with reducible, with respect to irreducible, backgrounds.

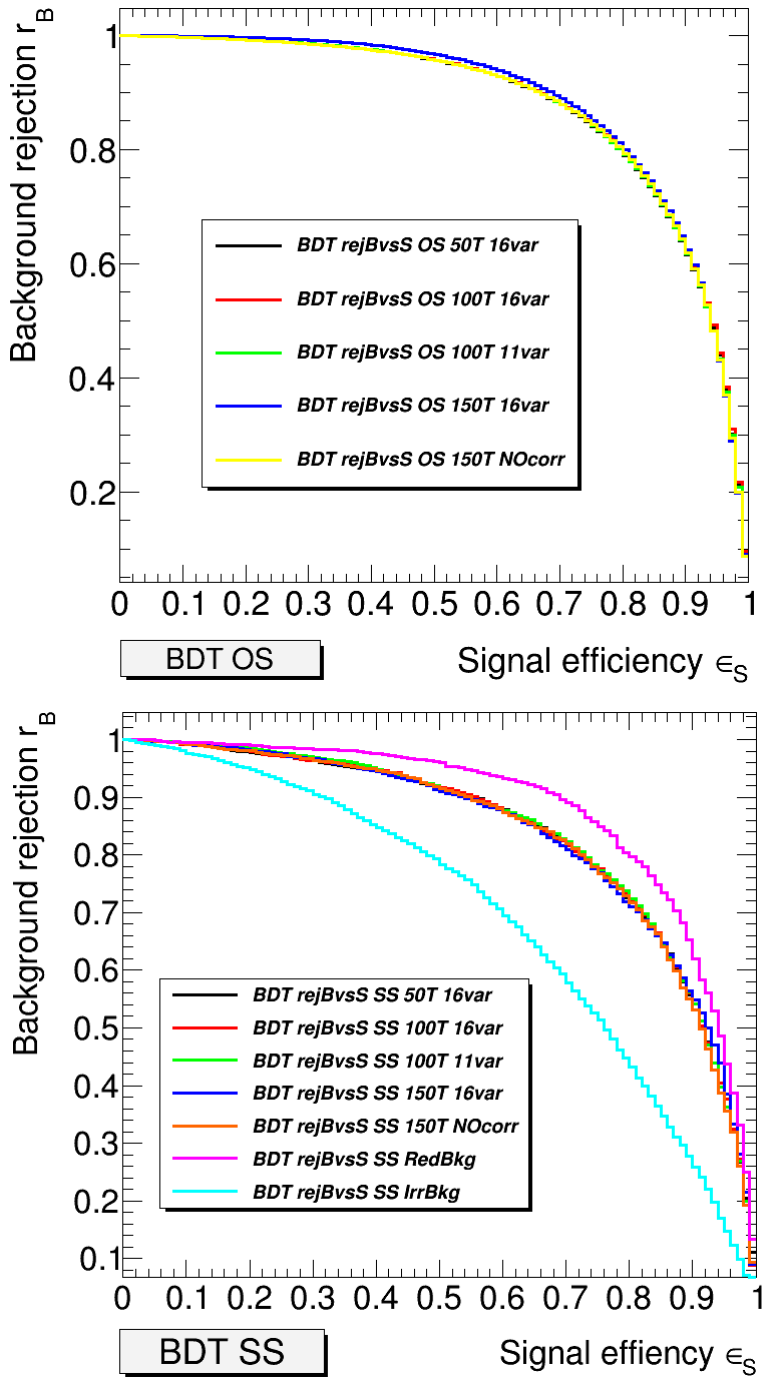



Figure 5.11: Background rejection versus signal efficiency curve for OS (top plot) and SS (bottom plot) events compared for the different BDT parameter settings. For OS events, all the different BDT configurations show similar shapes in the signal from background discrimination. For SS events, the BDT performances increase for a data sample which containes only reducible background processes (pink curve).

### 5.4.2 Multi Layer Perceptron (Neural Network)

As discussed in section 4.2.2, the behaviour of an artificial neural network is determined by the layout of the neurons, the weights of the inter-neuron connections and by the response of the neurons to the input, described by the neuron response function $\rho$. The multi-layer perceptron (MLP or neural network) used in this analysis presents a number of input networks equal to the number of input variables plus a bias, one hidden layer, composed by the 16 input variables plus one bias and 5 extra nodes, and one output node. Figure 5.12 shows the neural network achitecture scheme: the sixteen input variables, and a bias node, are connected to one hidden layer with 21 neurons. The colour shades of the connection lines provide information on the strenght of the connections from the input neurons and the hidden ones. As for the BDT, also MLP has been separately trained on SS and OS events.

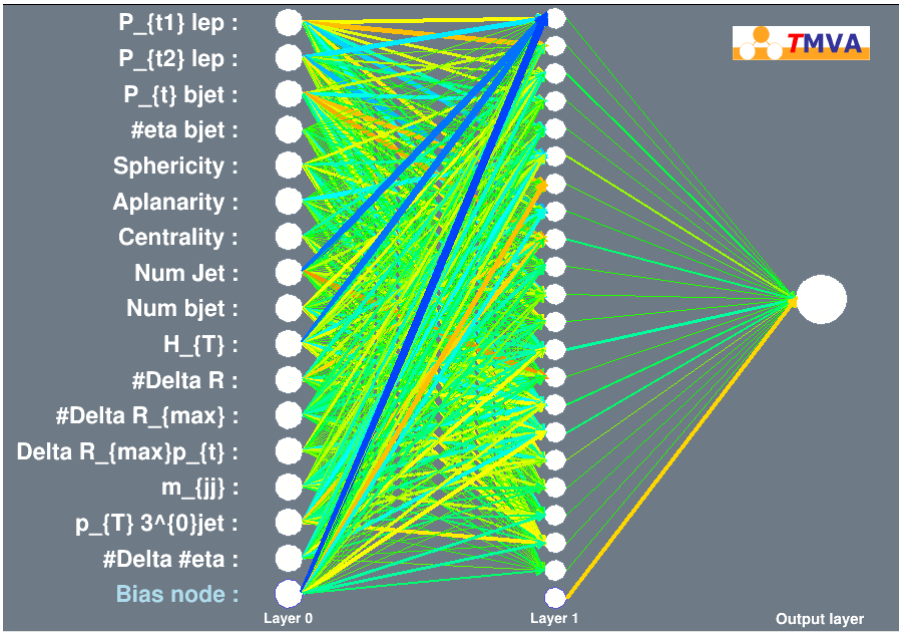



Figure 5.12: Neural Network achitecture scheme: the sixteen input variables, and a bias node, are connected to one hidden layer with 22 neurons. The colour shades of the connection lines provide information on the strenght of the connections from the input neurons and the hidden ones. The MLP also ranks the different variables by sorting them on the base of their discriminating power. For example, $H_T$ and $\mathrm{Num}_{jet}$ provide the highest discrimination power between signal and background. Other discriminating variables are $p_T^{bjet}$, $p_{T1}$ and $p_{T2}$, represented with orange line.

For both the SS and OS channel, the MLP has been trained and tested with samples containing:

- signal and only reducible backgrounds;
- signal and only irreducible backgrounds;
- signal and all background processes.

Tables 5.8 shows the signal over background significance provided by the MLP classifier for same sign ($1^{st}$ table) and for opposite sign ($2^{nd}$ table) events as well as the optimal cut for signal/background discrimination, the signal efficiency and the background rejection for the three samples above mentioned.
In figure 5.13 the MLP responses for OS (left) and SS (right) events for signal and irreducible backgrounds only (top plot), reducible background only (middle plot) and all the background processes (bottom plot) are presented. The results show how the MLP discrimination powers are larger when, as expected, only reducible background are present on the sample with respect to the sample which only containes irreducible backgrounds. However, it is remarkable that when all the background processes are trained and tested together, the MLP discriminating power almost remains constant. The results present the same trend in OS and in SS events although higher significance values are obtained for OS events.
In figure 5.14 the signal and background efficiencies (blue line and red line respectively), signal purity (blue dashed lines), signal efficiency times purity (blue dotted lines) and significance (green line) for opposite sign (left plots) and same sign (right plots) events for the MLP classifier are presented. This latter response is presented for signal and irreducibles backgrounds only (top plots), reducible backgrounds only (middle plots) and all the background processes (bottom plots).
Figure 5.15 shows the background rejection versus signal efficiency curve for SS (top plot) and OS (bottom plot) events for the different training configurations. It is visible that, when the training is performed on events which contains only irreducible backgrounds, a significant performance loss is present.

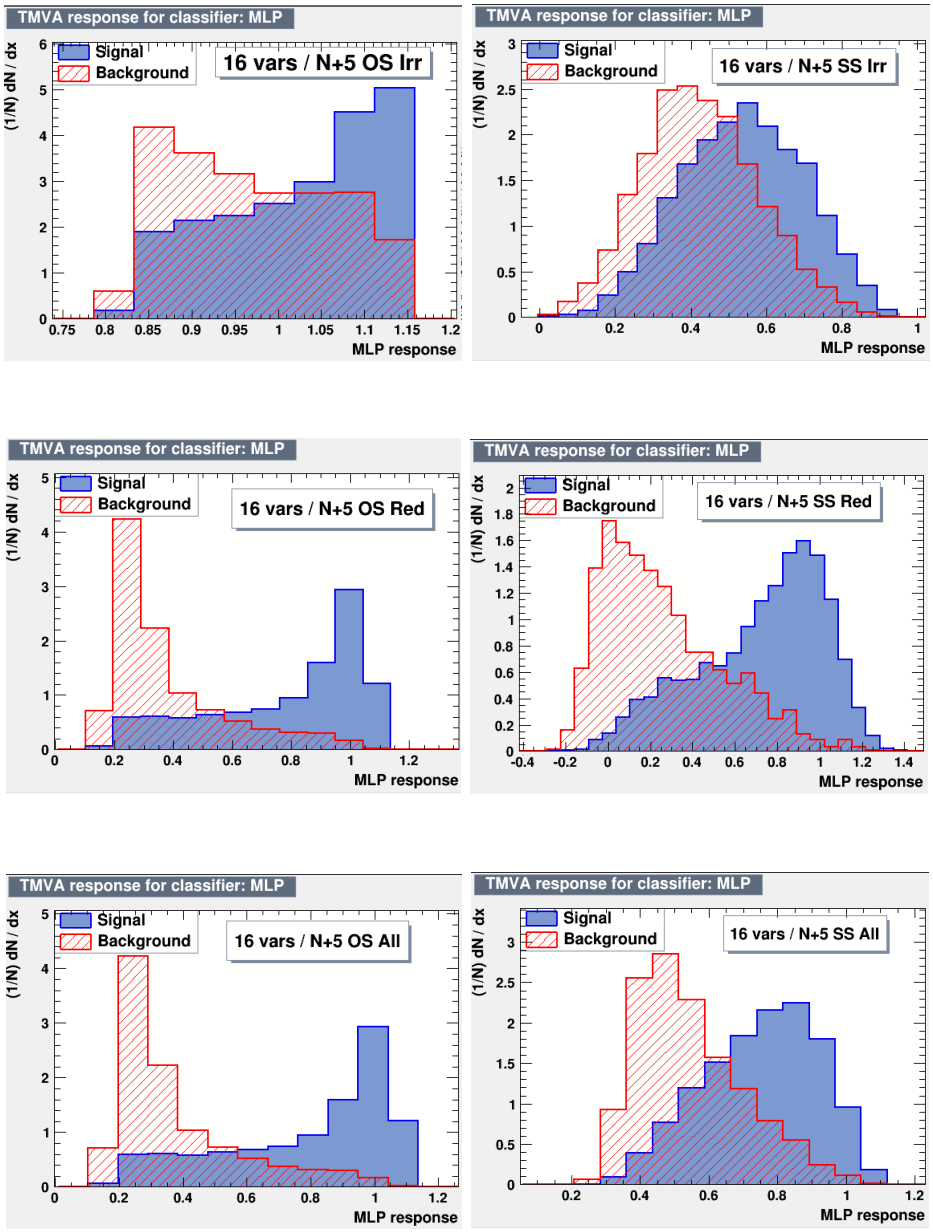



Figure 5.13: MLP responses for opposite sign (left) and same sign (right) events for signal and irreducible backgrounds only (top plot), reducible backgrounds only (middle plot) and all the background processes (bottom plot). The MLP discrimination performances show an increase with training sample with only reducible backgrounds, although no remarkable performance losses can be underlined when the reducible backgrounds are reintegrated in the training sample (bottom plot).

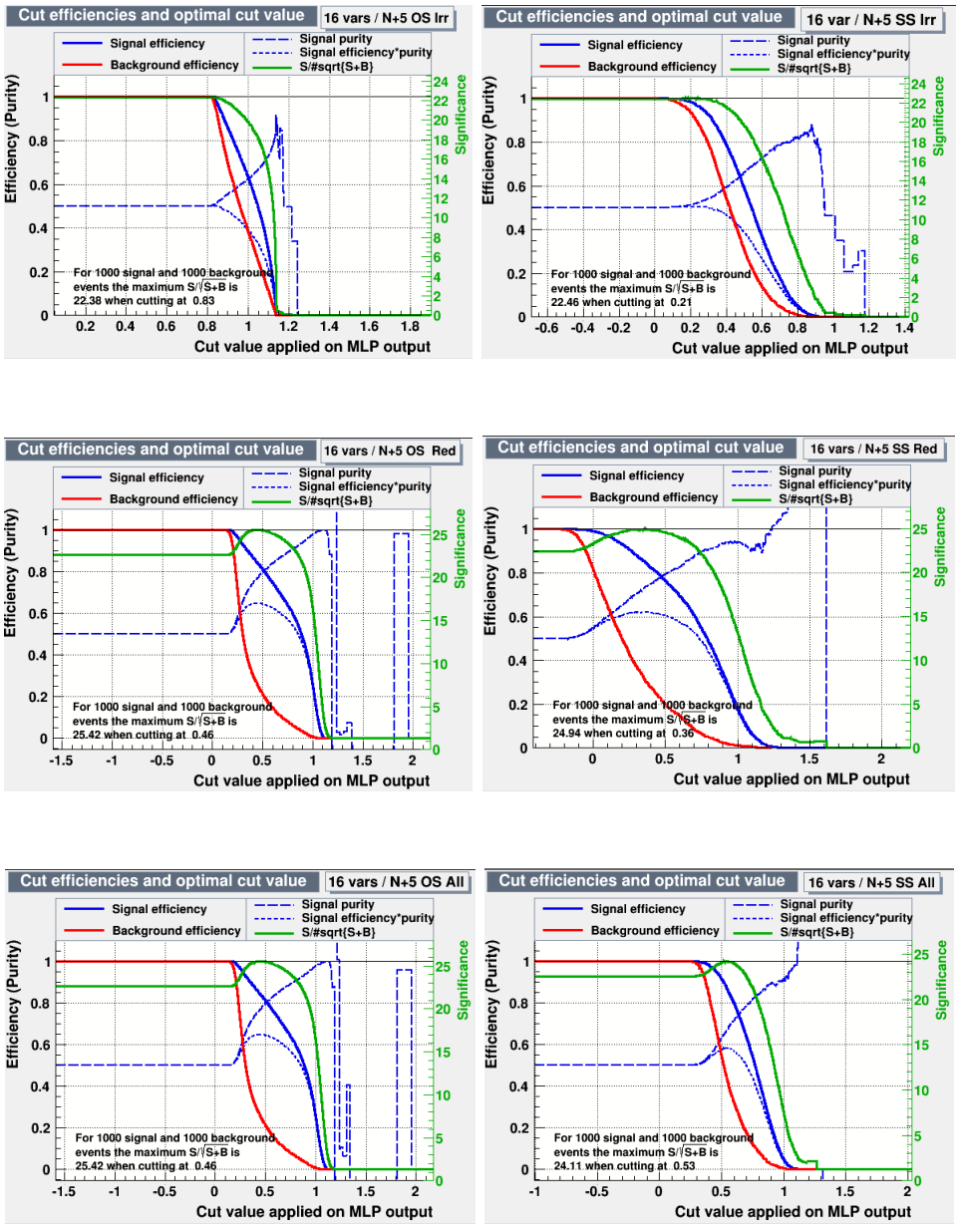



Figure 5.14: MLP signal and background efficiencies (blue line and red line respectively), signal purity (blue dashed lines), signal efficiency times purity (blue dotted lines) and significance (green line) for opposite sign (left plots) and same sign (right plots) events. This latter response is presented for signal and irreducible backgrounds only (top plots), reducible backgrounds only (middle plots) and all the background processes (bottom plots).

| **MLP Performances OS on 1000 Signal and Background Events** | | | | |
|---|---|---|---|---|
| Parameters | Optimal Cut | $\epsilon_S$ | $r_B$ | $S/\sqrt{S+B}$ |
| 16 vars, N+5 Irr | 0.83 | 0.995 ± 0.002 | 0.020 ± 0.004 | 22.38 |
| 16 vars, N+5 Red | 0.4604 | 0.835 ± 0.012 | 0.756 ± 0.014 | 25.42 |
| 16 vars, N+5 All | 0.4604 | 0.835 ± 0.012 | 0.756 ± 0.013 | 25.42 |
| **MLP Performances SS on 1000 Signal and Background Events** | | | | |
| 16 vars, N+5 Irr | 0.2120 | 0.980 ± 0.04 | 0.0778 ± 0.009 | 22.46 |
| 16 vars, N+5 Red | 0.3582 | 0.849 ± 0.011 | 0.6903 ± 0.015 | 24.94 |
| 16 vars, N+5 All | 0.5257 | 0.89 ± 0.01 | 0.5311± 0.016 | 24.11 |

Table 5.8: Signal over background significance provided for the MLP classifier for same sign (**1$^{st}$ table**) and for opposite sign (**2$^{nd}$ table**) events as well as the optimal cut for signal/background discrimination, signal efficiency and background rejection. The network has been trained for three different samples containing signal plus reducible background only, signal plus irreducible background only and signal plus all background processes. The results show how the MLP discrimination powers are larger when, as expected, only reducible backgrounds are present with respect to the sample which containes only irreducible backgrounds. However, it is remarkable that when all the background processes are trained and tested together, the MLP discriminating power almost remains constant. The results present the same trend in OS and in SS events.

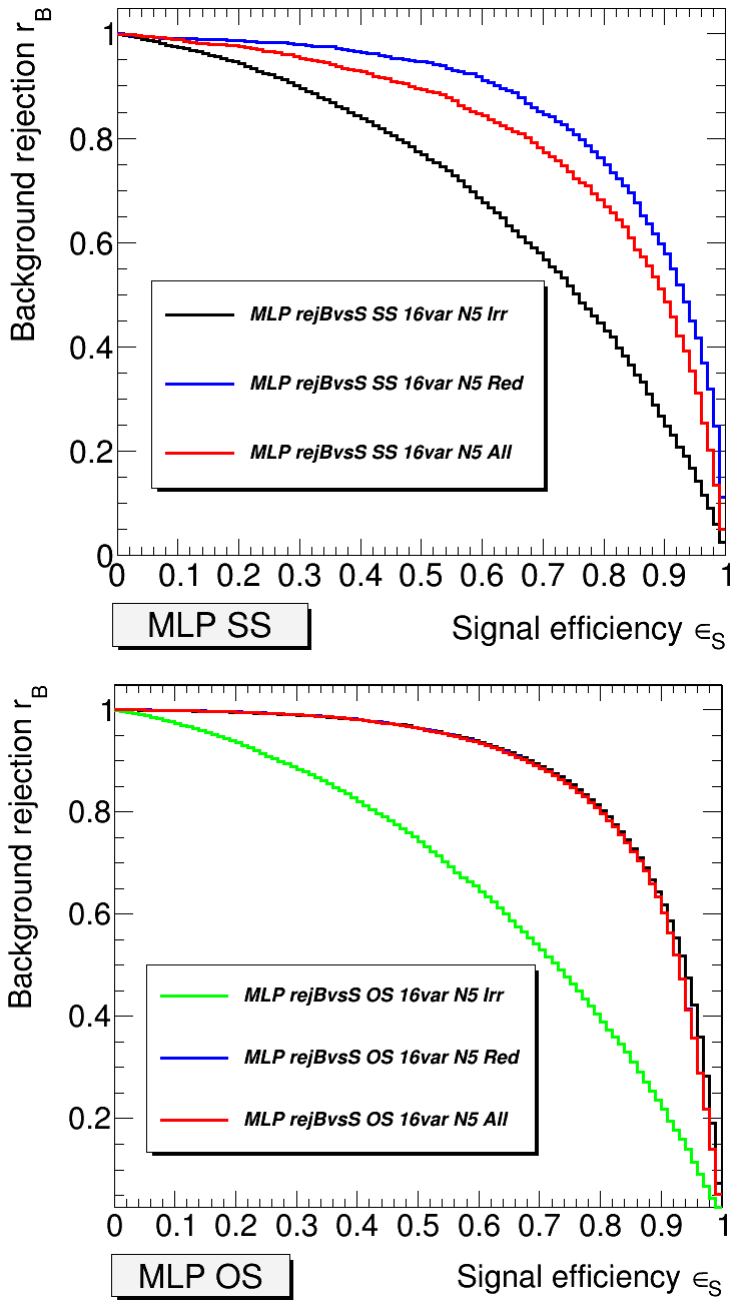



Figure 5.15: Background rejection $r_B$ versus signal efficiency $\epsilon_S$ curve for SS (top plot) and OS (bottom plot) events for the different MLP tests performed using samples with signal plus reducible backgorund only, irreducible background only and all background processes. A significance performance loss is present in both channels when the training in performed with a sample containing signal and irreducible backgrounds only.

## 5.5 BDT and MLP Comparison

In the previous sections the multivariate analysis performed with BDT and MLP on the $t\bar{t}H$ channel and its main reducible and irreducible backgrounds has been performed. For the BDT, among the different tested configurations, the one which maximizes the significance, with respect to the other settings, is the one with 150 decision trees and 16 input variables. Correlations among variables have demonstrated not to be relevant in the performances of the BDT and, in fact, significance is not affected by the exclusion of the most correlated variables. Tests have been perfomed for both OS and SS events separately.
For the MLP method, the selected configuration has 1 input layer composed by the 16 input variables and 22 hidden neurons. This neural network has been tested on different samples in which the $t\bar{t}H$ signal has been trained separately with only reducible, only irreducible backgrounds at first, and finally, with all of them, for both OS and SS events separately.
For both the MVA methods, the results have been reported for an equal number of 1000 signal and background events which, although does not represent a physical situation, provides indications on which multivariate method shows to be more performant in signal from background discrimination.
Table 5.9 reports the comparison between the performances obtained using the best BDT and the best MLP classifier configurations for SS (top table) and OS events (bottom table). For BDT, the best performance is achieved with 150 trees and 16 variables, while for MLP the highest significance value is obtained with 1 input layer and 22 hidden neurons trained with reducible backgrounds only. For both channels the significance obtained from the two MVA methods are similar.

Figure 5.16 shows the comparison between background rejection versus signal efficiency curve for the best BDT and MLP classifier configurations for SS (top plot) and OS (bottom plot) events. While in the OS channel, the curve of the two MVA methods have similar shape, in the SS channel better performances are obtained for the MLP classifier trained with a sample containing signal and reducible backgrounds only.
As already explained in the previous sections, most of the background processes are expected to produce opposite sign leptons in the final state. Since the standard analysis at the moment only takes into account same sign events, a final control of the MLP and BDT classifier responses has been performed in order to verify the discriminating power of $t\bar{t}H$ signal from the reducible and irreducible backgrounds for this channel only.

| 1000 Signal and Background SS Events | | | | |
|---|---|---|---|---|
| Method | Optimal Cut | $\epsilon_S$ | $r_B$ | $S/\sqrt{S+B}$ |
| Best BDT | -0.0676 | 0.861 ±0.011 | 0.639 ± 0.015 | 24.63 |
| Best MLP | 0.3582 | 0.849 ± 0.011 | 0.6903 ± 0.015 | 24.94 |
| 1000 Signal and Background OS Events | | | | |
| Method | Optimal Cut | $\epsilon_S$ | $r_B$ | $S/\sqrt{S+B}$ |
| Best BDT | -0.0673 | 0.845 ± 0.011 | 0.745 ± 0.014 | 25.48 |
| Best MLP | 0.4604 | 0.835 ± 0.012 | 0.756 ± 0.014 | 25.42 |

Table 5.9: Comparison between the best BDT and the best MLP classifier configurations for SS (top table) and OS events (bottom table). For BDT, the best performance is achieved with 150 trees and 16 variables, while for MLP the highest significance value is obtained with 1 input layer and 22 hidden neurons trained with only reducible backgrounds. The results report the optimal cut for signal to background discrimination, the signal efficiency, the background rejection and the signal over background significance. All these values are reported for an equal number of 1000 events for signal and background.

Figure 5.17 presents MLP (top plot) and BDT (bottom plot) responses for SS events. The signal (in red) and the different background contributions (other colours) are reported and divided into $t\bar{t}V$, $t\bar{t}$, single top and diboson productions. Other backgrounds are omitted because their contribution was negligible. Events have been normalized to the data luminosity of 20.3 fb$^{-1}$. The main background contamination arises from $t\bar{t}V$ and $t\bar{t}$ + single top productions, whose response, for both MLP and BDT, presents the same shape. The signal output obtained from the MLP classifier is near 1, while around 0.5 for the BDT classifier.
Figure 5.18 shows the scatter plot of the reducible and irreducible backgrounds for the MLP response to SS events: $t\bar{t}H$ signal is represented with black triangles while backgrounds with differentt coloured markers. Signal events are concentrated near the (1,1) point in the (Red, Irr) plane. With a final cut around 0.8 on the "Red" axis and around 0.5 on the "Irr" axis, the diboson background contamination and the major part of $t\bar{t}V$ and $t\bar{t}$ and single top backgrounds can be rejected. Figure 5.19 shows the same result for the BDT response for SS events for signal only (top left), $t\bar{t}V$ (top right), $t\bar{t}$ + single top (bottom left) and diboson (bottom right) production. As visible from these plots, signal events are concentrated around the point (1,1) in the bidimensional (Red, Irr) plane. On the contrary, the different backgrounds, according to their reducible or irreducible type, are clustered in different regions of the plane. With appropriate cuts on (Red, Irr) plane, the major contribution from background processes can be rejected.

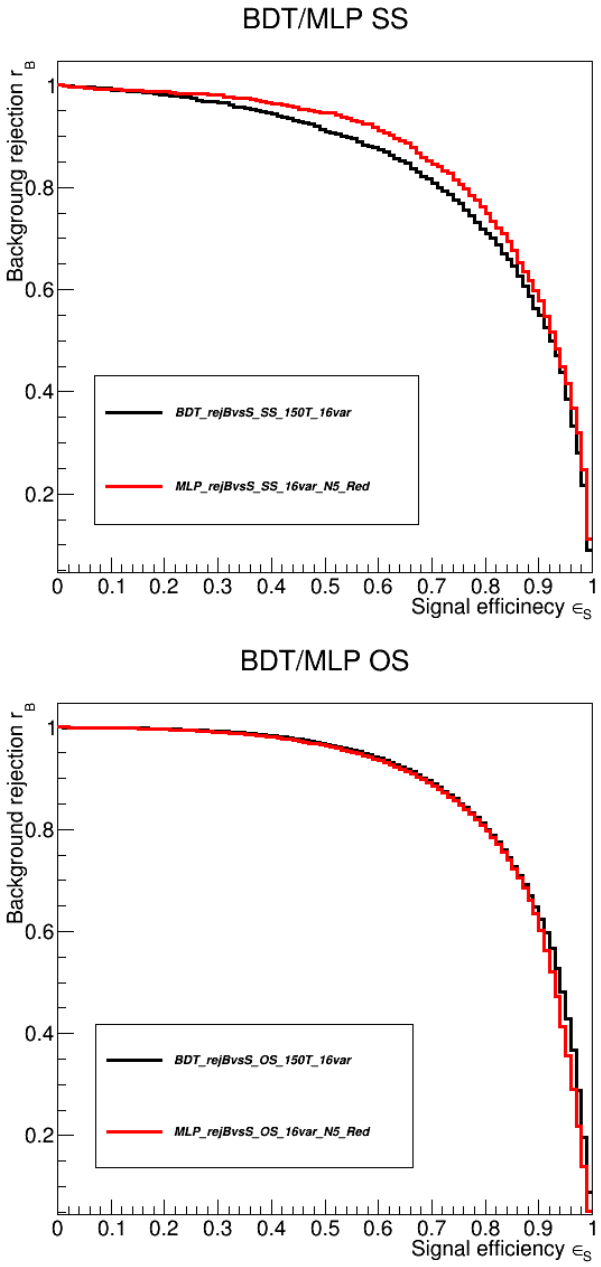



Figure 5.16: Comparison between background rejection versus signal efficiency curve for the best BDT and MLP classifier configurations for SS (top plot) and OS (bottom plot) events. While in the OS channel, the curve of the two MVA methods have similar shape, in the SS channel better performances are obtained for the MLP classifier trained with a sample containing signal and reducible backgrounds only.

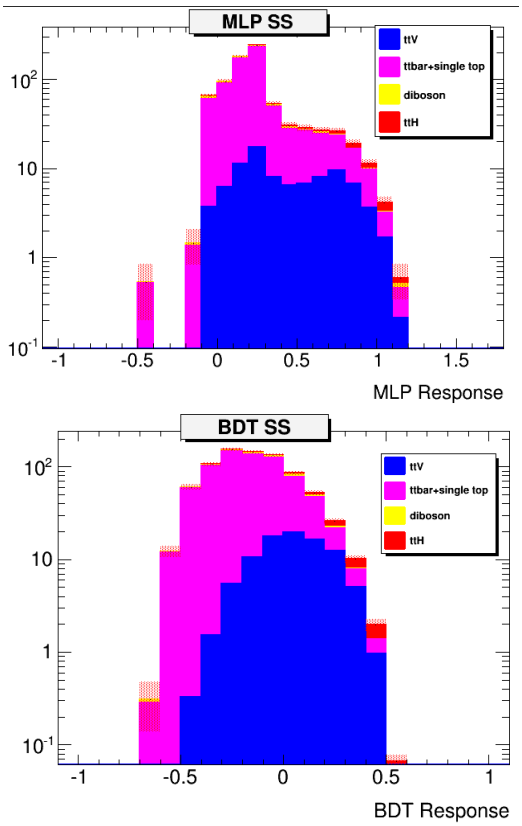



Figure 5.17: MLP (top plot) and BDT (bottom plot) responses for SS events. The signal (in red) and the different contributions of the background processes (other colours) are reported and divided into $t\bar{t}V$, $t\bar{t}$, single top and diboson productions. The other backgrounds are not reported because their contribution was negligible. Events have been normalized to the data luminosity of 20.3 fb$^{-1}$. The main background contamination arises from $t\bar{t}V$ and $t\bar{t}$ + single top productions, whose responce, for both MLP and BDT, presents the same shape. The signal output is located near 1 for the MLP classifier, while around 0.5 for the BDT.

### 5.5.1 Remarks and Future Development

The MVA methods studied here effectively enhance the signal background discrimination in the $t\bar{t}H$ multilepton search with respect to a standard cut-based analysis. Various methods have been studied. The methods are promising for the SS channel, while the OS channel is still overwhelmed by large background. Systematics have to be studied yet. With respect to the cut-based approach in which just a few number of events passes all the required cuts, MVA methods provide a continuous distribution output that is much more powerful if applied to a final fit to extract the signal strenght. This method would also be applied to the top-Higgs Yukawa coupling mea-

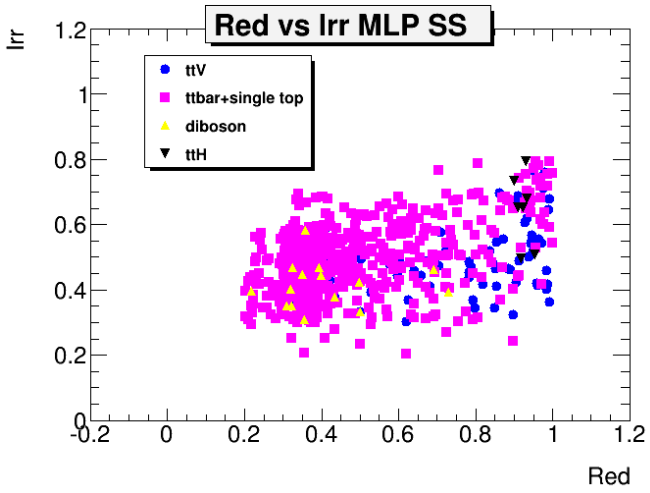



Figure 5.18: Scatter plot of irreducible versus reducible backgrounds for the MLP response in the SS channel for $t\bar{t}H$ signal (black triangles) and $t\bar{t}V$, $t\bar{t}$ + single top and diboson production (coloured markers). Signal events are concentrated in the bottom right region. Cutting at 0.8 in the Red axis and at 0.5 in the Irr axis, the major contribution from diboson background can be rejected.

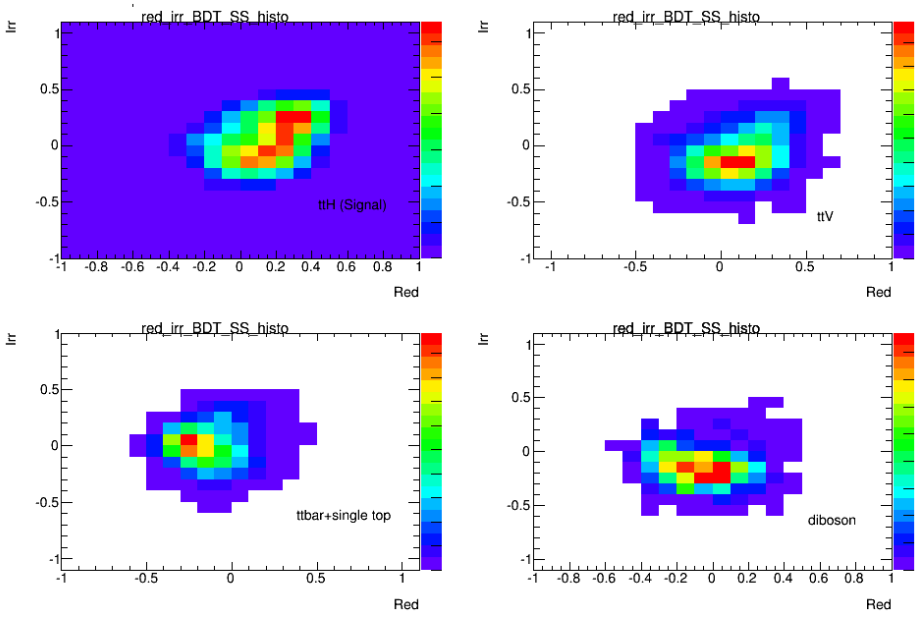



Figure 5.19: Scatter plot for the irreducible versus reducible backgrounds for the BDT response in the SS channel for $t\bar{t}H$ signal (top left) and $t\bar{t}V$ (top right), $t\bar{t}$ + single top (bottom left) and diboson (bottom right) productions. Signal and background events lie in different regions of the (Red, Irr) plane. Appropriate cuts on this plane can thus reject the major contribution from background processes.

surement that would be performed on the data that will be collected during LHC Run II.

# Conclusions

The Large Hadron Collider is designed to reach a center of mass energy of 14 TeV. It is the largest hadron collider and, during its second Run, will reach a center of mass energy of 13 TeV. Thanks to the increase of center of mass energy from Run I to Run II, LHC could be able to explore processes which, due to their low production cross sections, are among the rarest in physics and not available till now due to lack of statistics.
The $t\bar{t}H$ production represents one of these rare processes, with a production cross section of 130 fb. It still needs further investigation since the center of mass energy reached up to 2012 led to an expected number of events around $10^3$. Even with an ideal detector, precision measurements for this process, such as the Yukawa top-Higgs coupling and the experimental measurement of the cross section, certainly need further statistics. However, the theoretical $t\bar{t}H$ production cross section will gain the most, with respect to the other Higgs production mechanisms, from the energy increase which is going to characterize the LHC Run II.
Particle physics experiments constantly have to collect and manage large data sets. With a hundred of input variables, the search for interesting signal events, covered by the different background processes, can be an hard task.
The aim of this work has been to compare different multivariate methods which could improve, with respect to the standard cut-based analysis, the signal selection for the $t\bar{t}H$ channel. The final state which has been taken into account is the dileptonic one, in which two leptons, electrons or muons, with same or opposite charge are produced. One of the lepton in the final state is produced by the decay of the Higgs boson into W bosons ($H \rightarrow WW^*$), while the other one from the semileptonic decay of the $t\bar{t}$ pair.
Since the Run I data analysis is still ongoing with blinded signal region and due to the lack of statistics, this analysis was performed on MC samples only. During this study, different kinematic variables that characterize the event have been studied separately for signal and background samples and then selected for the implementation in the MVA methods according to their signal

from background discriminating power. The two methods which have been studied in this work are the Boosted Decision Trees (BDT) and the Multi Layer Perceptron (MLP). For each method, different tests have been performed using different settings of input variables and classifier parameters in order to find which configuration maximizes the signal over background ratio $S/\sqrt{S+B}$. These tests have been performed saparately for same sign (SS) and opposite sign (OS) class of events. The two MVA methods have been trained with the same sample containing 1000 signal and 1000 background events. The performance of the best configuration for the BDT classifier has been compared to the performance of the best configuration for the MLP, for both same sign and opposite sign channel. The following table summarizes the optimal cut for signal to background discrimination, as well as the signal efficiency, the background rejection and the signal significance for the two MVA best configurations. As visible, MLP performs better on SS sample while BDT on OS sample.

| SS Events | | | | |
|---|---|---|---|---|
| Method | Optimal Cut | $\epsilon_S$ | $r_B$ | $S/\sqrt{S+B}$ |
| Best BDT | -0.0676 | 0.861 $\pm$0.011 | 0.639 $\pm$ 0.015 | 24.63 |
| Best MLP | 0.3582 | 0.849 $\pm$ 0.011 | 0.6903 $\pm$ 0.015 | 24.94 |
| OS Events | | | | |
| Method | Optimal Cut | $\epsilon_S$ | $r_B$ | $S/\sqrt{S+B}$ |
| Best BDT | -0.0673 | 0.845 $\pm$ 0.011 | 0.745 $\pm$ 0.014 | 25.48 |
| Best MLP | 0.4604 | 0.835 $\pm$ 0.012 | 0.756 $\pm$ 0.014 | 25.42 |

This work is a preliminar study which will be extended with the implementation of other multivariate techniques (for example k-Nearest Neighbour or Support Vector Machine). Finally, the best MVA method willl be applied to the data that would be collected during the LHC Run II, when sufficient statistics will be recorded to performe precise experimental measurements.

# Appendix

Table 10 shows the linear correlation coefficients among the input variables used for the MVA analysis, for both signal and background events. They are in average inferior to the 40%, with exception for those variables which are derived from one another. For istance, $H_T$ is strongly correlated to those variables, such as the transverse momenta of the leptons and of jets, which define it, as expected. Sphericity and aplanarity present also higher correlations, as expected since they are both defined as a linear combination of the momentum matrix eigenvalues.

Linear Correlation Coefficients for Signal

| | $p_{T1}$ | $p_{T2}$ | $p_T bjet$ | $\lvert\eta_{bjet}\rvert$ | S | A | C | Num Jet | Num bjet | $H_T$ | $\Delta R$ | $\Delta R_{max}$ | $\Delta R^{pT}_{max}$ | $m_{jj}$ | $p_T^{3jet}$ | $\Delta\eta_{max}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $p_{T1}$ | +1.000 | +0.460 | +0.129 | -0.019 | -0.051 | -0.053 | +0.022 | +0.041 | -0.008 | +0.492 | -0.006 | +0.015 | -0.003 | +0.227 | +0.270 | +0.008 |
| $p_{T2}$ | +0.460 | +1.000 | +0.062 | -0.020 | -0.032 | -0.052 | +0.010 | -0.015 | -0.025 | +0.317 | +0.014 | +0.022 | -0.007 | +0.114 | +0.173 | +0.019 |
| $p_T bjet$ | +0.129 | +0.062 | +1.000 | -0.105 | -0.055 | -0.113 | +0.126 | +0.077 | +0.244 | +0.465 | -0.017 | +0.039 | +0.138 | +0.432 | +0.405 | -0.012 |
| $\lvert\eta_{bjet}\rvert$ | -0.019 | -0.020 | -0.105 | +1.000 | -0.326 | -0.202 | -0.534 | -0.042 | -0.022 | -0.082 | +0.005 | +0.087 | +0.059 | +0.022 | -0.071 | +0.090 |
| S | -0.051 | -0.032 | -0.055 | -0.326 | +1.000 | +0.652 | +0.477 | +0.175 | +0.064 | -0.038 | +0.015 | -0.093 | -0.150 | -0.214 | +0.041 | -0.090 |
| A | -0.053 | -0.052 | -0.113 | -0.202 | +0.652 | +1.000 | +0.296 | +0.324 | +0.093 | -0.003 | +0.017 | -0.003 | -0.167 | -0.197 | +0.042 | +0.011 |
| C | +0.022 | +0.010 | +0.126 | -0.534 | +0.477 | +0.296 | +1.000 | -0.077 | +0.024 | +0.059 | -0.109 | -0.334 | -0.130 | -0.091 | +0.060 | -0.420 |
| Num Jet | +0.041 | -0.015 | +0.077 | -0.042 | +0.175 | +0.324 | -0.077 | +1.000 | +0.181 | +0.499 | +0.033 | +0.493 | +0.071 | +0.241 | +0.359 | +0.485 |
| Num bjet | -0.008 | -0.025 | +0.244 | -0.022 | +0.064 | +0.093 | +0.024 | +0.181 | +1.000 | +0.086 | +0.016 | +0.086 | +0.017 | +0.037 | +0.097 | +0.068 |
| $H_T$ | +0.492 | +0.317 | +0.465 | -0.082 | -0.038 | -0.003 | +0.059 | +0.499 | +0.086 | +1.000 | -0.009 | +0.278 | +0.113 | +0.643 | +0.761 | +0.235 |
| $\Delta R$ | -0.006 | +0.014 | -0.017 | +0.005 | +0.015 | +0.017 | -0.109 | +0.033 | +0.016 | -0.009 | +1.000 | +0.276 | +0.024 | -0.004 | -0.011 | +0.282 |
| $\Delta R_{max}$ | +0.015 | +0.022 | +0.039 | +0.087 | -0.093 | -0.003 | -0.334 | +0.493 | +0.086 | +0.278 | +0.276 | +1.000 | +0.214 | +0.274 | +0.187 | +0.816 |
| $\Delta R^{pT}_{max}$ | -0.003 | -0.007 | +0.138 | +0.059 | -0.150 | -0.167 | -0.130 | +0.071 | +0.017 | +0.113 | +0.024 | +0.214 | +1.000 | +0.618 | +0.115 | +0.185 |
| $m_{jj}$ | +0.227 | +0.114 | +0.432 | +0.022 | -0.214 | -0.197 | -0.091 | +0.241 | +0.037 | +0.643 | -0.004 | +0.274 | +0.618 | +1.000 | +0.525 | +0.252 |
| $p_T^{3jet}$ | +0.270 | +0.173 | +0.405 | -0.071 | +0.041 | +0.042 | +0.060 | +0.359 | +0.097 | +0.761 | -0.011 | +0.187 | +0.115 | +0.525 | +1.000 | +0.134 |
| $\Delta\eta_{max}$ | +0.008 | +0.019 | -0.012 | +0.090 | -0.090 | +0.011 | -0.420 | +0.485 | +0.068 | +0.235 | +0.282 | +0.816 | +0.185 | +0.252 | +0.134 | +1.000 |

Linear Correlation Coefficients for Background

| | $p_{T1}$ | $p_{T2}$ | $p_T bjet$ | $\lvert\eta_{bjet}\rvert$ | S | A | C | Num Jet | Num bjet | $H_T$ | $\Delta R$ | $\Delta R_{max}$ | $\Delta R^{pT}_{max}$ | $m_{jj}$ | $p_T^{3jet}$ | $\Delta\eta_{max}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $p_{T1}$ | +1.000 | +0.536 | +0.144 | +0.025 | -0.078 | -0.016 | -0.055 | +0.056 | +0.113 | +0.641 | -0.017 | +0.079 | +0.036 | +0.353 | +0.327 | +0.064 |
| $p_{T2}$ | +0.536 | +1.000 | +0.126 | +0.060 | -0.064 | -0.023 | -0.079 | -0.023 | +0.174 | +0.479 | -0.020 | +0.091 | +0.020 | +0.227 | +0.239 | +0.060 |
| $p_T bjet$ | +0.144 | +0.126 | +1.000 | -0.059 | -0.007 | -0.061 | +0.088 | +0.102 | +0.249 | +0.521 | +0.005 | +0.074 | +0.103 | +0.399 | +0.462 | -0.021 |
| $\lvert\eta_{bjet}\rvert$ | +0.025 | +0.060 | -0.059 | +1.000 | -0.268 | -0.147 | -0.542 | -0.068 | -0.001 | -0.027 | +0.107 | +0.132 | +0.039 | +0.040 | -0.041 | +0.130 |
| S | -0.078 | -0.064 | -0.007 | -0.268 | +1.000 | +0.558 | +0.477 | +0.279 | +0.040 | +0.008 | -0.016 | -0.102 | -0.098 | -0.144 | +0.080 | -0.124 |
| A | -0.016 | -0.023 | -0.061 | -0.147 | +0.558 | +1.000 | +0.230 | +0.523 | +0.069 | +0.103 | -0.008 | +0.057 | -0.113 | -0.102 | +0.102 | +0.063 |
| C | -0.055 | -0.079 | +0.088 | -0.542 | +0.477 | +0.230 | +1.000 | -0.004 | +0.007 | -0.019 | -0.126 | -0.319 | -0.097 | -0.121 | +0.023 | -0.369 |
| Num Jet | +0.056 | -0.023 | +0.102 | -0.068 | +0.279 | +0.523 | -0.004 | +1.000 | +0.134 | +0.420 | -0.011 | +0.338 | -0.011 | +0.148 | +0.334 | +0.319 |
| Num bjet | +0.113 | +0.174 | +0.249 | -0.001 | +0.040 | +0.069 | +0.007 | +0.134 | +1.000 | +0.197 | -0.014 | +0.096 | -0.012 | +0.071 | +0.171 | +0.070 |
| $H_T$ | +0.641 | +0.479 | +0.521 | -0.027 | +0.008 | +0.103 | -0.019 | +0.420 | +0.197 | +1.000 | -0.017 | +0.263 | +0.088 | +0.605 | +0.713 | +0.193 |
| $\Delta R$ | -0.017 | -0.020 | +0.005 | +0.107 | -0.016 | -0.008 | -0.126 | -0.011 | -0.014 | -0.017 | +1.000 | +0.358 | +0.007 | +0.023 | +0.001 | +0.361 |
| $\Delta R_{max}$ | +0.079 | +0.091 | +0.074 | +0.132 | -0.102 | +0.057 | -0.319 | +0.338 | +0.096 | +0.263 | +0.358 | +1.000 | +0.277 | +0.321 | +0.170 | +0.799 |
| $\Delta R^{pT}_{max}$ | +0.036 | +0.020 | +0.103 | +0.039 | -0.098 | -0.113 | -0.097 | -0.011 | -0.012 | +0.088 | +0.007 | +0.277 | +1.000 | +0.643 | +0.099 | +0.220 |
| $m_{jj}$ | +0.353 | +0.227 | +0.399 | +0.040 | -0.144 | -0.102 | -0.121 | +0.148 | +0.071 | +0.605 | +0.023 | +0.321 | +0.643 | +1.000 | +0.494 | +0.291 |
| $p_T^{3jet}$ | +0.327 | +0.239 | +0.462 | -0.041 | +0.080 | +0.102 | +0.023 | +0.334 | +0.171 | +0.713 | +0.001 | +0.170 | +0.099 | +0.494 | +1.000 | +0.094 |
| $\Delta\eta_{max}$ | +0.064 | +0.060 | -0.021 | +0.130 | -0.124 | +0.063 | -0.369 | +0.319 | +0.070 | +0.193 | +0.361 | +0.799 | +0.220 | +0.291 | +0.094 | +1.000 |

Table 10: Linear correlation coefficients of input variables for signal and background events. They are in average inferior to the 40%, with exception for those variables which are derived from one another. For istance, $H_T$ is strongly correlated to those variables, such as the transverse momenta of the leptons and of jets, which define it, as expected. Sphericity and aplanarity present also higher correlations, as expected since they are both defined as a linear combination of the momentum matrix eigenvalues.

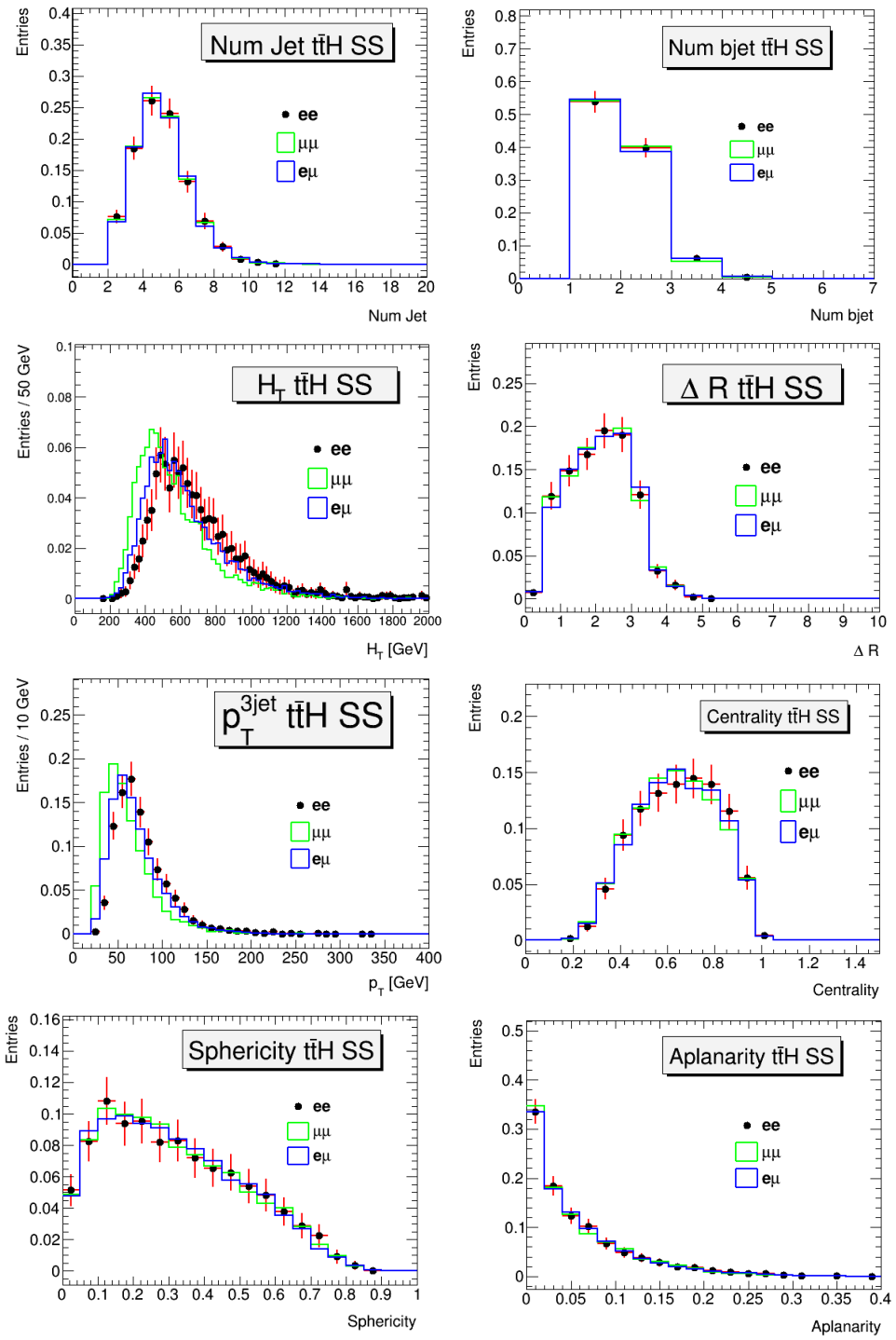



Figure 20: Variable distributions for eight of the sixteen variables used for the MVA analysis. The plots show the superimposition of the $ee$, $\mu\mu$ and $e\mu$ channels in same sign events. It is clearly visible from these distributions that no remarkable differences can be found in the shape of the variables for the three different channels. The events have been weighted according to the cross section of the processes and normalized to the 2012 integrated luminosity of 20.3 fb$^{-1}$.

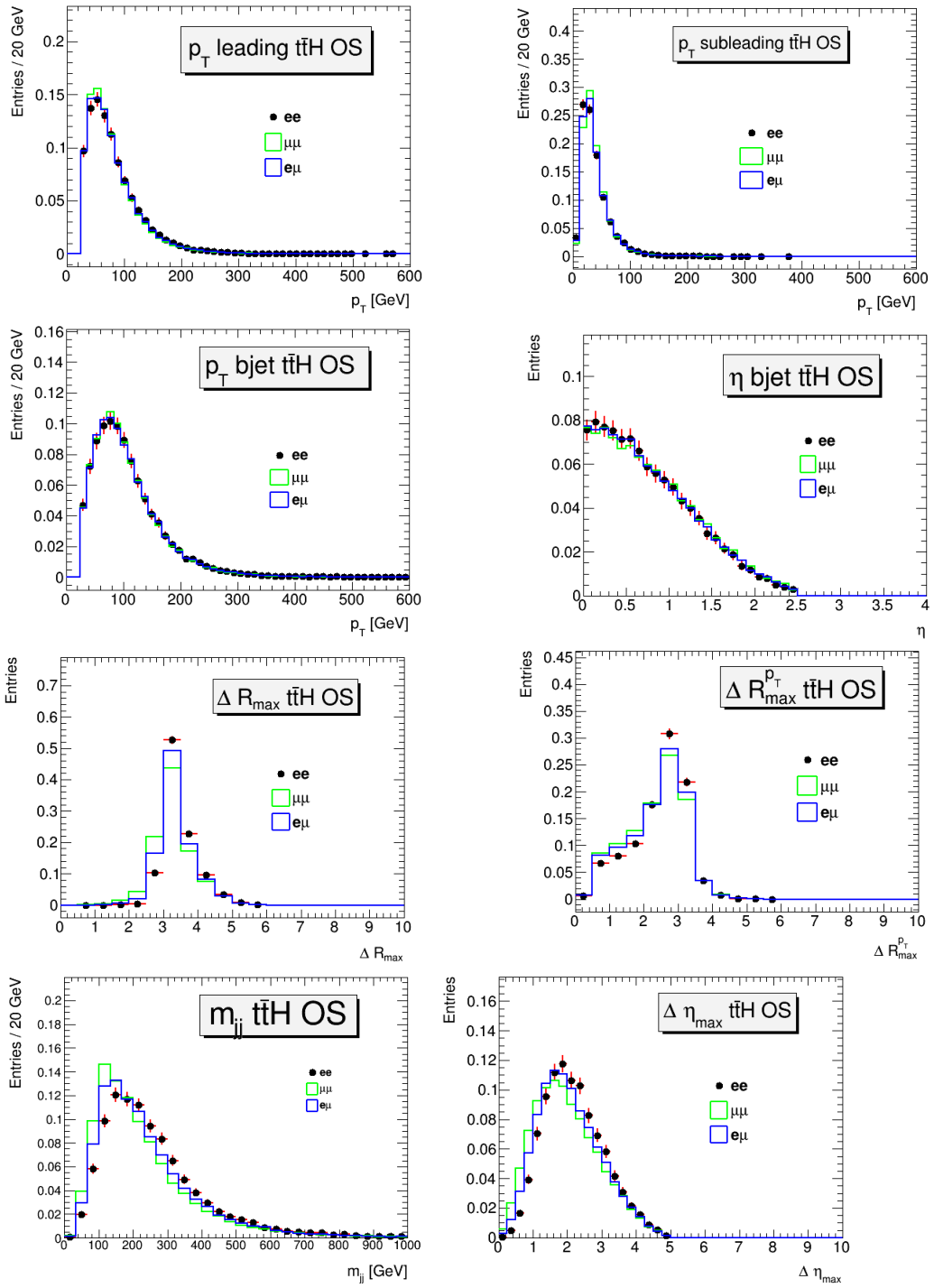



Figure 21: Variable distributions for eight of the sixteen variables used in the MVA analysis. Figures show the superimposition of the $ee$, $\mu\mu$ and $e\mu$ channels of OS events. The three different channels result in similar event shapes of the variables. The events have been weighted according to the cross section of the process and normalized to the 2012 integrated luminosity of 20.3 fb$^{-1}$.

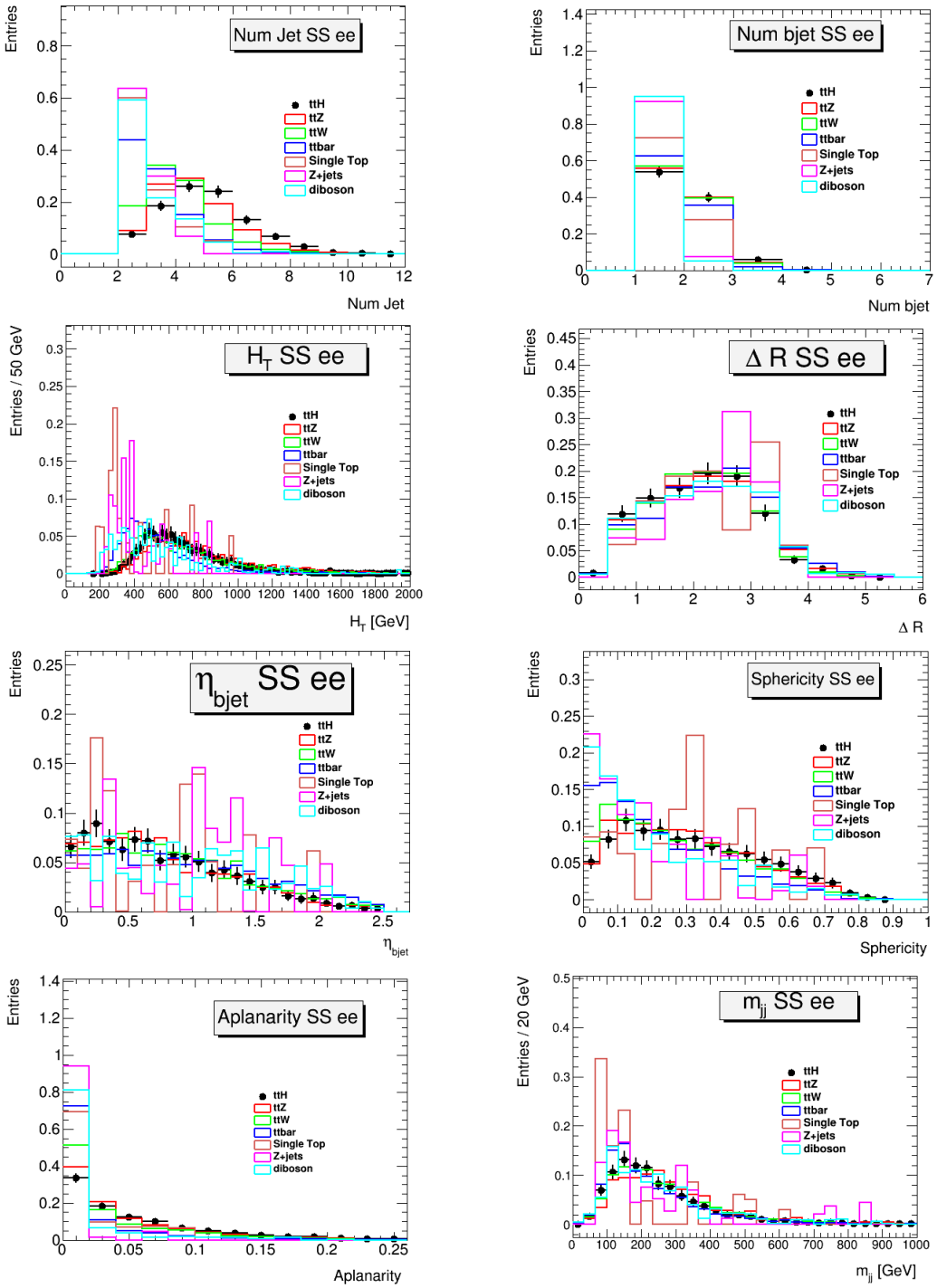



Figure 22: Variable distributions for eight of the sixteen variables used in MVA analysis for SS events in the *ee* channel. The $t\bar{t}H$ signal and the different background processes are superimposed for each variable. The events have been weighted according to the cross section of the process and normalized to the 2012 integrated luminosity of 20.3 fb$^{-1}$.

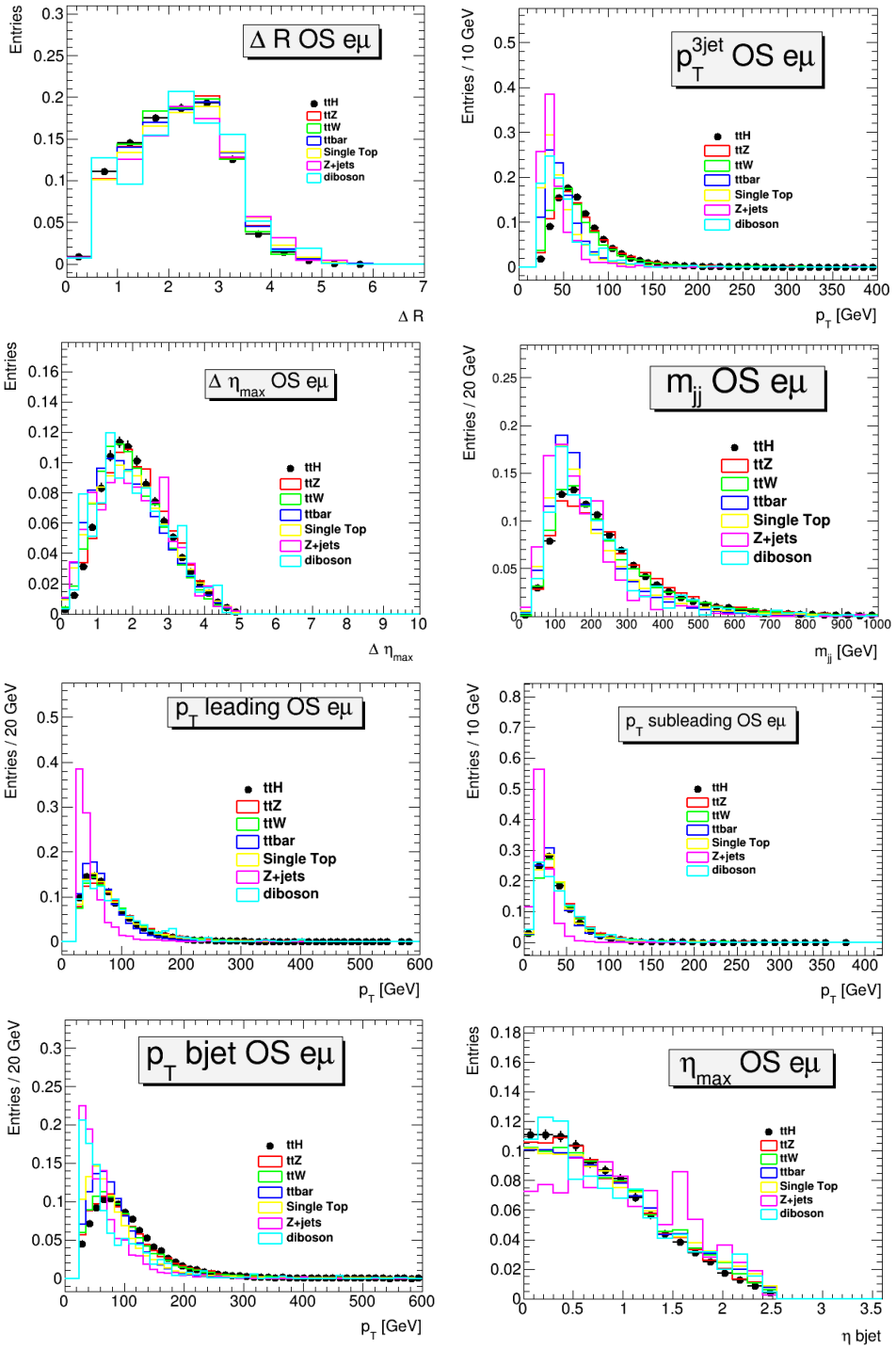



Figure 23: Variable distributions for eight of the sixteen variables used in MVA analysis for OS events in the $e\mu$ channel. The $t\bar{t}H$ signal and the different background processes are superimposed for each variable. The events have been weighted according to the cross sections of the process and normalized to the 2012 integrated luminosity of 20.3 fb$^{-1}$.

# Bibliography


[1] S.L. Glashow *Partial Symmetries of Weak Interactions. Nucl.Phys. 22. 579-588* (1961).

[2] S. Weinberg *A Model of Leptons. Phys. Rev. Lett. 19, 1264.* (196 7)

[3] A. Salam, *Elementary Particle Theory, Ed. N. Svarholm* (1968).

[4] J. Beringer et al. *Particle Data Group, Phys. Rev. D86, 010001* (2012).

[5] C. Quigg, *Gauge Theories of the Strong, Weak and Electromagnetic Interactions. Westview Press.* (1983).

[6] I.Jr Aitchison, A.J.G. Hey, *Gauge Theories in Particle Physics, vol. 1., IOP.* (2003).

[7] I.Jr Aitchison, A.J.G. Hey, *Gauge Theories in Particle Physics, vol. 1., IOP.* (2004).

[8] The UA1 Collaboration, *Volume 126, Issue 5, Pages 398–410*, 7 July 1983.

[9] The UA2 Collaboration, *Phys.Lett. B122 103-116*, (1983).

[10] P.W. Higgs, *Spontaneous Symmetry Breadown without Massless Bosons, Phys. Rev. 145, 1156.* (1966).

[11] The ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC, Phys.Lett. B716,1-29.* (2012).

[12] The CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC, Phys. Lett. B 716, 30.* (2012).

[13] P. de Aquino *Beyond Standard Model Phenomenology at the LHC, Springer.* (2014).

[14] D. Griffith, *Introduction to Elementary Particles, Wiley-VCH, Weinheim*(2008).

[15] J.H. Christenson, J.W. Cronin, V.L. Fitch, R. Turlay, *Evidence for the $2\pi$ decay of the $K_2^0$ meson. Phys. Rev. Lett. 13(4), 138–140* (1964).

[16] The ATLAS Collaboration, *ATLAS: Detector and physics performance technical design report. Volume 1. CERN-LHCC-99-14, ATLAS-TDR-14* (2008).

[17] CERN Courier, *The LHC's first long run: High-quality beam from the injectors and full exploitation of options in the collider underpinned the LHC's performance in 2010–2013.* (19 August 2013).

[18] The ATLAS Collaboration, *Luminosity Determination in pp Collisions at $\sqrt{s}$=7 TeV Using the ATLAS Detector at the LHC*, (11 January 2011).

[19] The ATLAS Collaboration, *ATLAS magnet system: Technical design report - ATLAS Collaboration CERN-LHCC-97-18.*

[20] The ATLAS Collaboration, *ATLAS inner detector: Technical design report. Vol. 1.*

[21] The ATLAS Collaboration, *ATLAS TDR 14, CERN/LHCC/99-14, 5* (25 May 1999).

[22] The ATLAS Collaboration, *ATLAS pixel detector. Technical report, CERN/LHCC, 1998.*

[23] J. N. Jackson, *The ATLAS semiconductor tracker (SCT). Nucl. Instrum. Meth., A541.* (2005).

[24] The ATLAS Collaboration, *Construction, assembly and tests of the ATLAS electromagnetic barrel calorimeter, CERN-PH-EP/2005-034* (2005).

[25] J. Abdallah, on the behalf of the Tile Calorimeter community, *ATLAS Tile Calorimeter Commissioning and Performance* (2009)

[26] The ATLAS Collaboration, *Construction, assembly and testing of the ATLAS hadronic end-cap calorimeter, 2007 JINST 2 P05005* (2007).

[27] R.S. Orr, *The ATLAS Forward Calorimeter, 13th ICATPP Conference* (2011)

[28] The ATLAS Collaboration, *ATLAS muon spectrometer: Technical Design Report, CERN/LHCC.* (1998).

[29] The ATLAS Collaboration, *Performance of Particle Identification with the ATLAS Transition Radiation Tracker* (2011).

[30] M. Cacciari, G.P. Salam, G. Soyez, *The Anti-KT Jet Clustering Algorithm*, (2008).

[31] The ATLAS Collaboration, *ATLAS High Level Trigger, Data Acquisition and Controls, CERN/LHCC/2003-022, Geveva, CERN.* (2003).

[32] The ATLAS Collaboration, *ATLAS Forward Detector for Luminosity Measurements and Monitoring Letter of Intent, CERN-LHCC/04-10, LHCC I-014.*

[33] CERN Courier, *Roman Pots for the LHC*, (28 March 1999).

[34] N.Armesto, H. Ma, M.Martinez, Y. Mehtar-Tani, C.A. Salgado *Phys. Lett. B717 280-286.* (2012)

[35] S. Dittmaier, C. Mariotti, G. Passarino, R. Tanaka, *Handbook of LHC Higgs cross sections: Higgs Properties. Report of the LHC Higgs Cross Section Working Group.* (2013).

[36] S. Catani, D. de Florian, M. Grazzini and P. Nason, *JHEP 0307 (2003) 028.*

[37] S. Bentvelsen, E. Laenen, P. Motylinski *Higgs production through gluon fusion at leading order.* NIKHEF 2005-007.

[38] M. Ciccolini, A. Denner, S. Dittmaier *Electroweak and QCD corrections to Higgs boson production in vector-boson fusion at LHC.* (2007).

[39] O. Brein, A. Djouadi, R. Harlander. *NNLO QCD corrections to the Higgs-strahlung processes at hadron colliders. Phys.Lett. B579, 149-156* (2004).

[40] W. Beenakker, S. Dittmaier, M. Kramer, B. Plumper, M. Spira, P.M. Zerwas *NLO QCD corrections to $t\bar{t}H$ production in hadron collisions. Nucl.Phys. B653, 151-203* (2003).

[41] A. Djouadi, J. Kalinowski and M. Spira*HDECAY:a Program for Higgs Boson Decays in the Standard Model and its Supersymmetric Extension* (1997).

[42] F. Maltoni, D. Rainwater, S. Willenbrock *Measuring the top-quark Yukawa coupling at hadron colliders via $t\bar{t}H$, $H \to WW$. Phys.Rev. D66,034022* (2002).

[43] CDF Collaboration, *Search for the Standard Model Higgs Boson Produced in Association with Top Quarks Using the Full CDF Data Set. Phys. Rev. Lett. 109.* (2012)

[44] D⊘ Collaboration, *Search for the Standard Model Higgs boson in the $t\bar{t}H \to t\bar{t}b\bar{b}$ channel. D⊘ Conference Note CONF-5739.* (2008).

[45] The ATLAS Collaboration, *Search for $t\bar{t}H$ in the multilepton final state: backgrounds and their estimation.* (2014).

[46] M.V. Garzelli, A. Kardos, C.G. Papadopoulos and Z. Trocsanyi, *Standard Model Higgs boson production in association with a top anti-top pair at NLO with parton showering, Europhys. Lett., 96:11001,* (2011).

[47] T. Sjostrand, R. Corke. *Pythia 8 worksheet. [http://home.thep.lu.se/ torbjorn/pythia8/ mergingworksheet8160.pdf],* (March 2012).

[48] The ATLAS Collaboration, *Supporting document on electron efficiency measurements using the 2012 LHC proton-proton collision data. Technical Report ATLAS-COM-PHYS-2013-1295, CERN, Geneva,* (September 2013).

[49] The ATLAS Collaboration, *Electron reconstruction and identification efficiency measurements with the ATLAS detector using the 2011 LHC proton–proton collision data,Eur. Phys. J. C,74:2941,* (2014).

[50] The ATLAS Collaboration, *Search for Standard Model Higgs boson production in association with a top quark pair in multilepton signatures with the ATLAS detector: object and event selection,* (2 June 2014).

[51] The ATLAS Collaboration, *Identification of muon candidates in pp collisions at $\sqrt{s}$=900 GeV with the ATLAS detector,ATLAS-CONF-2010-015,* (2010).

[52] The ATLAS Collaboration, *Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton–proton collision data, CERN-PH-EP-2014-151* (2014).

[53] M.V. Garzelli, A. Kardos, C.G. Papadopoulos, and Z. Trocsanyi, *$t\bar{t}W^{\pm}$ and $t\bar{t}Z$ Hadroproduction at NLO accuracy in QCD with Parton Shower and Hadronization effects. JHEP, 1211:056,* (2012).

[54] The ATLAS Collaboration, *Supporting document on electron efficiency measurements using the 2012 lhc proton-proton colli- sion data. Technical Report ATLAS-COM-PHYS-2013-1295, CERN, Geneva, September,* (2013).

[55] G. Corcella et al., *HERWIG 6.5 Release Note arXiv:hep-ph/0210213,* (2005).

[56] M. Wolter, *Multivariate analysis methods in physics,* (2007).

[57] F. Fiedler, A. Grohsjean, P. Haefner, P. Schieferdecker ,*The Matrix Element Method and its Application to Measurements of the Top Quark Mass, Nucl.Instrum.Meth.A624:203-218,* (2010).

[58] C.M. Bishop, *Pattern Recognition and Machine Learning, Springer, New York* (2006).

[59] G. Cowan, *Topics in statistical data analysis for high-energy physics, CERN Yellow Report CERN-002, pp.197-218*, (2010).

[60] G. Cowan, *Statistical Data Analysis, Oxford Science Pubblications,* (1998).

[61] T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning, 2nd ed. Springer, Berlin*, (2009).

[62] E. Parzen, *Estimation of a Probability Density Function and Its Mode, Ann. Math. Statistics. V. 33. P. 1065-1076.* (1962).

[63] P.J. Werbos, *Beyond Regression: New Tools for Prediction and Analysis in the Behavioural Sciences. Ph.D. Thesis. Boston: Harvard Univ.* (1974).

[64] D. E. Rumelhart, G. E. Hinton, R. J. Williams, *Learning Internal Representations by Error Propagation Computational Models of Cognition and Perception. Cambridge: MIT Press, V. 1, Ch. 8. P. 319-362* (1986).

[65] W.S. Sarle, *Stopped Training and Other Remedies for Overfitting Proc. of the 27th Symp. on the Interface of Comp. Science and Statistics, P. 352-360,* (1995).

[66] C. Goutte, *Note on Free Lunches and Cross-Validation, Neural Computation, V. 9. 1211-1215,* (1997).

[67] V. Vapnik, A. Lerner, *Pattern Recognition Using Generalized Portrait Method, Automation and Remote Control, V. 24,* (1963).

[68] V. Vapnik, A. Chervonenkis, *A Note on One Class of Perceptrons, Automation and Remote Control, V. 25,* (1964).

[69] Y. Freund and R. E. Schapire, *A decision-theoretic generalization of on-line learning and an application to boosting, J. Comput. Syst. Sci. 55, 119–139* (1997).

[70] L. Breiman, *Random Forests, Technical Report, University of California* (2001).

[71] J. Friedman, T. Hastie and R. Tibshirani, *The Elements of Statistical Learning, Springer Series in Statistics,* (2001).

[72] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, H. Voss *TMVA: Toolkit for Multivariate Data Analysis with ROOT, Users Guide. CERN-OPEN-2007-007,* (2013).