Study of the Higgs boson discovery potential, produced by the vector boson fusion with the ATLAS detector and commissioning of calorimetric transverse missing energy

D. Varouchas

To cite this version:

D. Varouchas. Study of the Higgs boson discovery potential, produced by the vector boson fusion with the ATLAS detector and commissioning of calorimetric transverse missing energy. High Energy Physics - Experiment [hep-ex]. Université Paris Sud - Paris XI, 2009. English. NNT : . tel-00452166

HAL Id: tel-00452166
https://theses.hal.science/tel-00452166
Submitted on 1 Feb 2010

LAL 09-116
Septembre 2009

THÈSE

présentée le 18 septembre 2009
par

Dimitris Varouchas

pour obtenir le grade de

Docteur ès sciences
de l'Université Paris XI Orsay

Potentiel de découverte par le détecteur ATLAS du boson de Higgs produit par fusion de bosons vecteurs et mise au point de la mesure de l'énergie transverse manquante calorimétrique

soutenue devant la commission d'examen composée de:

M.	C.	Anastasiou	
M.	G.	Bernardi	Rapporteur
M.	K.	Cranmer	
M.	D.	Rousseau	Directeur de thèse
Mme.	I.	Wingerter-Seez	Rapporteur
M.	G.	Wormser	Président

LAL 09-116
September 2009

DISSERTATION

presented the 18th of September 2009
by

Dimitris Varouchas

to obtain the degree of

Doctor of Science
of the Université Paris XI Orsay

Study of the Higgs boson discovery potential, produced by the vector boson fusion with the ATLAS detector and commissioning of calorimetric transverse missing energy

Committee in charge:

Mr.	C.	Anastasiou	
Mr.	G.	Bernardi	Rapporteur
Mr.	K.	Cranmer	
Mr.	D.	Rousseau	Dissertation advisor
Mrs.	I.	Wingerter-Seez	Rapporteur
Mr.	G.	Wormser	President

Contents

Introduction

The Standard Model of elementary particles is the theory that describes three of the four fundamental interactions (the strong, weak and electromagnetic ones) in a coherent framework. It has been developed during the last century and has proven to be extremely successful providing excellent description of all the phenomena observed in the particle physics domain up to energies explored by LEP (Large Electron-Positron collider) and Tevatron. Nevertheless, the origin of particle masses remains an open question. The electroweak symmetry breaking mechanism as introduced by Peter Higgs [1] and others [2, 3] provides an elegant answer. However, it predicts also a yet undiscovered particle, the Higgs boson. The LEP experiment excluded a Higgs boson with a mass lighter than 114.4 GeV[4], while the Tevatron experiments recently excluded the existence of a Higgs boson with a mass close to 170 GeV[5] in 95% of C.L. Precision tests of the electroweak interaction favor a light Higgs boson with a mass below 185 GeV[6]. Chapter 3 gives an introduction to the Standard Model followed by an introduction to Higgs mechanism.

In the year 2010, the Large Hadron Collider (LHC) at CERN is scheduled to start its operation. After a few run-periods, it is scheduled to reach its nominal energy and luminosity performance and then, $p-p$ collisions with a center of mass energy of up to $\sqrt{s}$=14 TeV and a luminosity of up to 10^{34} cm^{-2}s^{-1} will open up a new era in high energy particle physics, allowing a potential discovery of the Higgs boson over its entire allowed mass range. Therefore, the LHC is expected to provide an answer to the question whether the Higgs mechanism exists in nature and it is responsible for the electroweak symmetry breaking which explains the way particles acquire masses. In particular, the general-purpose detector ATLAS (A Toroidal LHC AparatuS) is designed to discover the Higgs boson and to study a wide range of other physics subjects and new potential phenomena, such as supersymmetric particles [7], extra-dimensions particles [8] or mini black holes [9]. Chapter 1 summarizes the design and the properties of the LHC and ATLAS detector followed by the ATLAS performance aspects focusing on the necessary constituents of VBF $H \to \tau^+\tau^-$ analysis, which is the main topic of the present document. These constituents are the electron/muon/hadronic tau identification and reconstruction, the missing transverse energy and hadronic jets reconstruction and finally the b-jet tagging efficiency (identification of jets originating from a b-quark)

Chapter 4 presents a study of the Standard Model Higgs boson decay into a τ lepton pair which further decays into leptons and/or hadrons produced via Vector Boson Fusion (VBF) mechanism. This process is one of the most promising for the discovery of a light Higgs boson with a mass below ~150 GeV. The VBF production mechanism provides a set of specific characteristics which allow the rejection of large background expected in LHC. This characteristic signature consists of only two highly energetic jets in the forward regions of the detector while the Higgs decay products are preferably emitted into the central detector region. In the same chapter, details of this Higgs decay channel and the most important background processes are discussed, as well as, the specific cut criteria which can be used to discriminate signal from background events. Considering the electroweak character of Vector Boson Fusion process, which implies a limited QCD radiation, no other jet activity is expected a fact which allows the application of a central jet veto (CJV) cut to suppress potential background. A study of CJV cut optimization is also presented.

Finally, an important aspect of VBF $H \rightarrow \tau^+\tau^-$ analysis is the presence of missing transverse energy ($\not{E}_T$) due to the neutrinos originating from the τ decays. A good understanding of $\not{E}_T$ variable will be of crucial importance, since $\not{E}_T$ resolution is directly reflected in the Higgs boson mass resolution. This was the motivation for performing a commissioning analysis of ATLAS calorimeters studying $\not{E}_T$ variables focusing on real noise data obtained during the single-beam/cosmic data-taking period of fall 2008, presented in chapter 2. Noise in ATLAS LAr (Liquid Argon) and Tile calorimeters is addressed, being a very important input for every physics analysis especially related to jet and $\not{E}_T$. An understanding and a parameterisation of the noise is discussed, providing at the same time a useful calorimeter commissioning tool, which allowed the observation and the correction of several unexpected features. Part of the results obtained by this study are already officially approved by the ATLAS collaboration [10].

Chapter 1

ATLAS experiment in LHC

ATLAS (**A** large **T**oroidal **L**HC **A**pparatu**S**) is one of the six particle detector experiments (ALICE, ATLAS, CMS, TOTEM, LHCb, and LHCf) constructed at the Large Hadron Collider (LHC). ATLAS is designed as a general-purpose detector in order to measure the broadest possible range of signals. This is intended to ensure that, whatever form any new physical processes or particles might take, ATLAS will be able to detect them and measure their properties. These different types of new physics that might become detectable in the

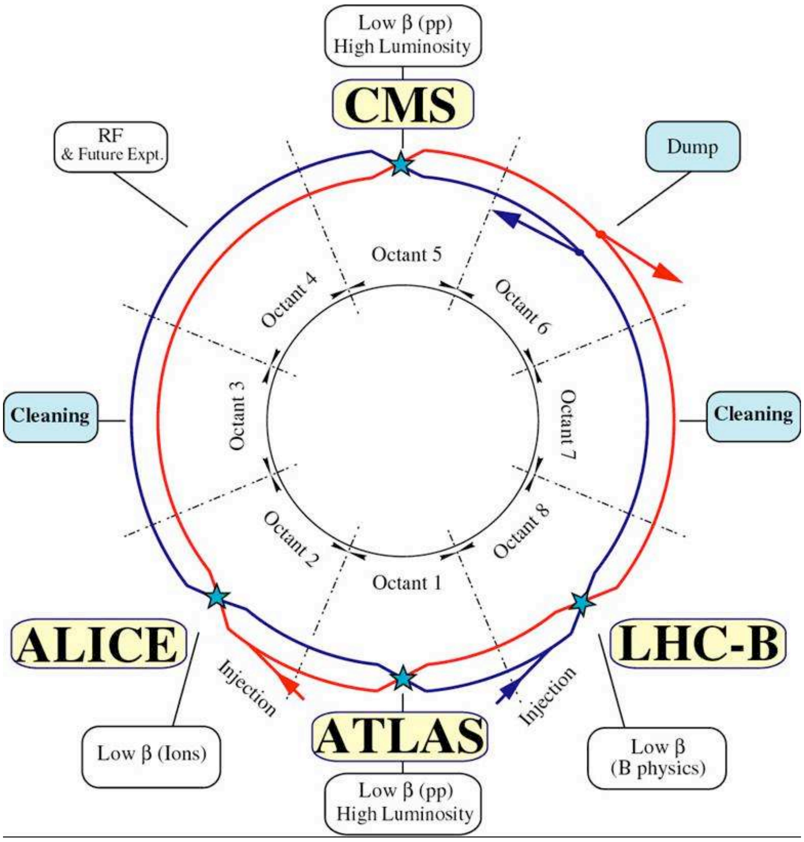


Figure 1.1: *An overview of LHC illustrating the positions of the 4 main experiments including ATLAS.*

energetic collisions of the LHC such as confirmations or improved measurements of Standard Model (through the discovery of Higgs boson), super-symmetric particles or exotic particles, are some of the examples of new physics that might be detected in ATLAS. In the following chapter LHC accelerator is first presented, followed by a detailed description of ATLAS experiment emphasising in Liquid Argon calorimeter. The main aspects of ATLAS physics and detection performance are also highlighted.

1.1 LHC (Large Hadron Collider)

The LHC (figure 1.1) is the world's largest and highest-energy particle accelerator, intended to collide opposing particle beams, of either protons at an energy of 7 TeV per particle, or lead nuclei at an energy of 574 TeV per nucleus. It lies in a tunnel 27 kilometres in circumference, located 175 meters deep, beneath the France-Swiss border near Geneva. After LEP (Large Electron-Positron Collider) which was operating until the year 2000 and TeVatron, which is collecting collision data at the moment, LHC will be the new particles collider devoted to discovery Physics.

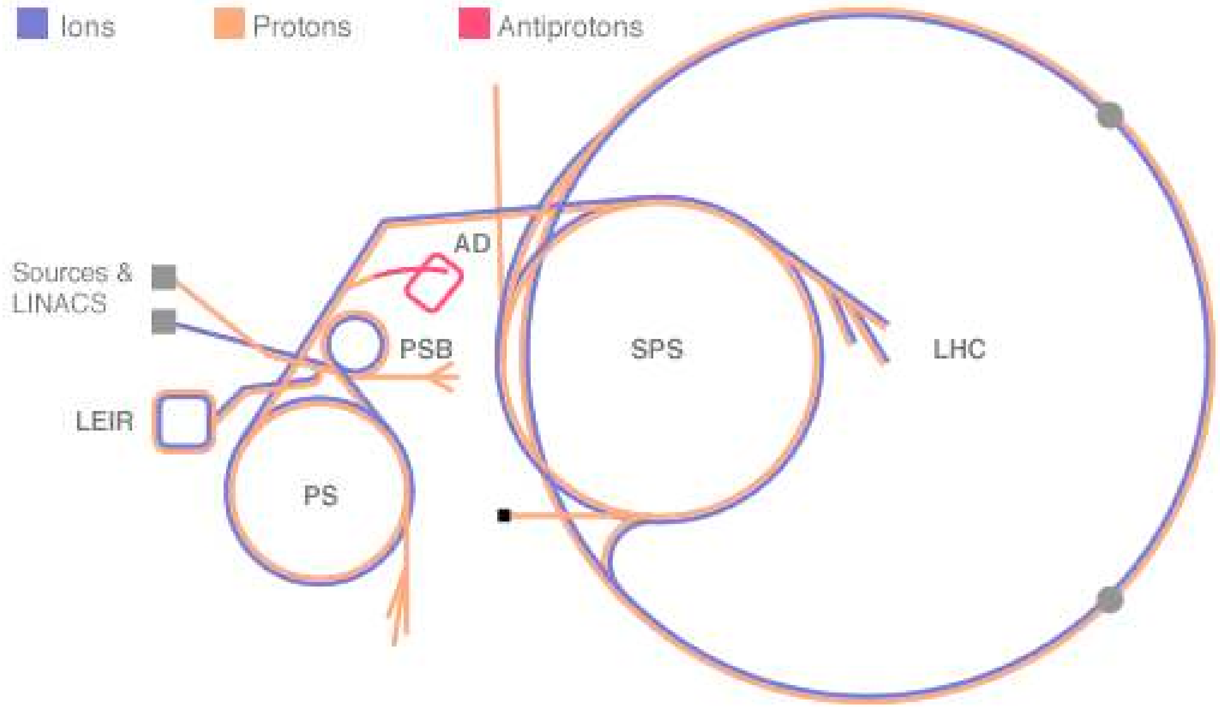


Figure 1.2: *The CERN accelerator complex. Protons follow the following path before enter LHC: (i)Linac, (ii)PSB (Booster), (iii)PS, (iv)SPS and finally (v)LHC.*

The main characteristics of LHC are presented in table 1.1. Protons before injected in the main ring to reach the energy of 7 TeV, are being accelerated through a pre-acceleration system which is summarised as followed (figure 1.2):

i. Linac: an acceleration up to 50 MeV

Circumference	26.7 km
Injection energy	450 GeV
Beam energy	7 TeV
Magnetic field of a dipole	8.4 Tesla
Nominal luminosity	$10^{34} cm^{-2} s^{-1}$
Luminosity life-time	10 h
Beam life-time	22 h
Number of bunches	2808
Number of particles per bunch	1.15×10^{11}
Time interval between two bunches	$25ns$
Space interval between two bunches	7.5 m
Bunch length	7.55 cm
Transverse beam size at the interaction point	16.7 μm

Table 1.1: *Nominal LHC characteristics at high luminosity period.*

ii. PSB(Booster): an acceleration up to 1 GeV

iii. PS: an acceleration up to 26 GeV

iv. SPS: an acceleration up to 450 GeV

v. LHC: an acceleration up to 7 TeV

The number of produced events is given by the equation:

$$N_{event} = \sigma_{event} \cdot L \tag{1.1}$$

where σ_{event} is the cross section of the event and L the luminosity of LHC, a quantity that only depends on the beam parameters and is defined as follows [11]:

$$L = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi \epsilon_n \beta^*} F \tag{1.2}$$

where:

- N_b : the number of particles per bunch ($\sim 10^{11}$ for high luminosity, $\sim 10^{10}$ for low luminosity).
- n_b : the number of bunches per beam (2808).
- f_{rev} : the revolution frequency (40 MHz).
- γ_r : the relativistic gamma factor.
- ϵ_n : the normalised transverse beam emittance (3.75).
- β^* : the beta function at the collision point (0.55 m).

- F : the geometric luminosity reduction factor due to the crossing angle at the interaction point (285 μrad).

The beam life time is expected to be $\sim$ 22 h, however luminosity is being reduced with a factor of 2 after 10h and therefore a new injection to LHC is needed. There are two operation periods foreseen for LHC, for the nominal energy at the center of mass:

- A low luminosity period for the first years ($L = 10^{33} cm^{-2} s^{-1}$). One year of this operation ($\sim$ 100 days) corresponds to an integrated luminosity of $\mathcal{L}$ =10fb^{-1}.
- A high luminosity period which will follow ($L = 10^{34} cm^{-2} s^{-1}$) with a respective integrated luminosity of $\mathcal{L}$ =100fb^{-1}/year.

Apart from ATLAS experiment which is described in details in the following sections, three other major detectors are placed at the different interaction points as shown in figure 1.1. ALICE[12] will study heavy ion collisions focusing in quarks-gluon plasma analysis. LHCb[13] is designed to study $CP-$ violation and perform precision measurements exploiting B-meson decays and finally CMS[14] is a general-purpose detector similar to ATLAS, but using different technologies. CMS is constructed around a unique solenoidal magnetic field and the tracking detector is based entirely on Silicon technology. Its electromagnetic calorimeter is homogenous, constituted by $PbWO_4$ crystals.

1.2 ATLAS detector

Before issuing the technical details of ATLAS[15, 16], the presentation of some definitions and conventions is required. The nominal interaction point is defined as the origin of the coordinate system, while the beam direction defines the z-axis and the $x-y$ plane is transverse to the beam direction. The positive x-axis is defined as pointing from the interaction point to the centre of the LHC ring and the positive y-axis is defined as pointing upwards. The side-A of the detector is defined as that with positive z and side-C is that with negative z. The azimuthal angle ϕ is measured around the beam axis, and the polar angle θ is the angle from the beam axis. The pseudorapidity is defined as

$$\eta = -log(tan\frac{\theta}{2}) \tag{1.3}$$

The transverse momentum p_T , the transverse energy E_T , and the missing transverse energy E_T^{miss} (or $\not{E}_T$) are defined in the $x-y$ plane. The ΔR cone in the pseudorapidity-azimuthal angle space is defined as

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \tag{1.4}$$

Due to particles properties expected to be observed in the high luminosity of LHC, respective requirements are set to ATLAS design. ATLAS will be able to detect hadrons due to the high cross-sections of QCD interactions and all the range of the known particles in order to perform studies in electroweak interactions, and flavour physics. The top quark will be produced at the LHC at a rate of a few tens of Hz, providing the opportunity to test its couplings and spin.

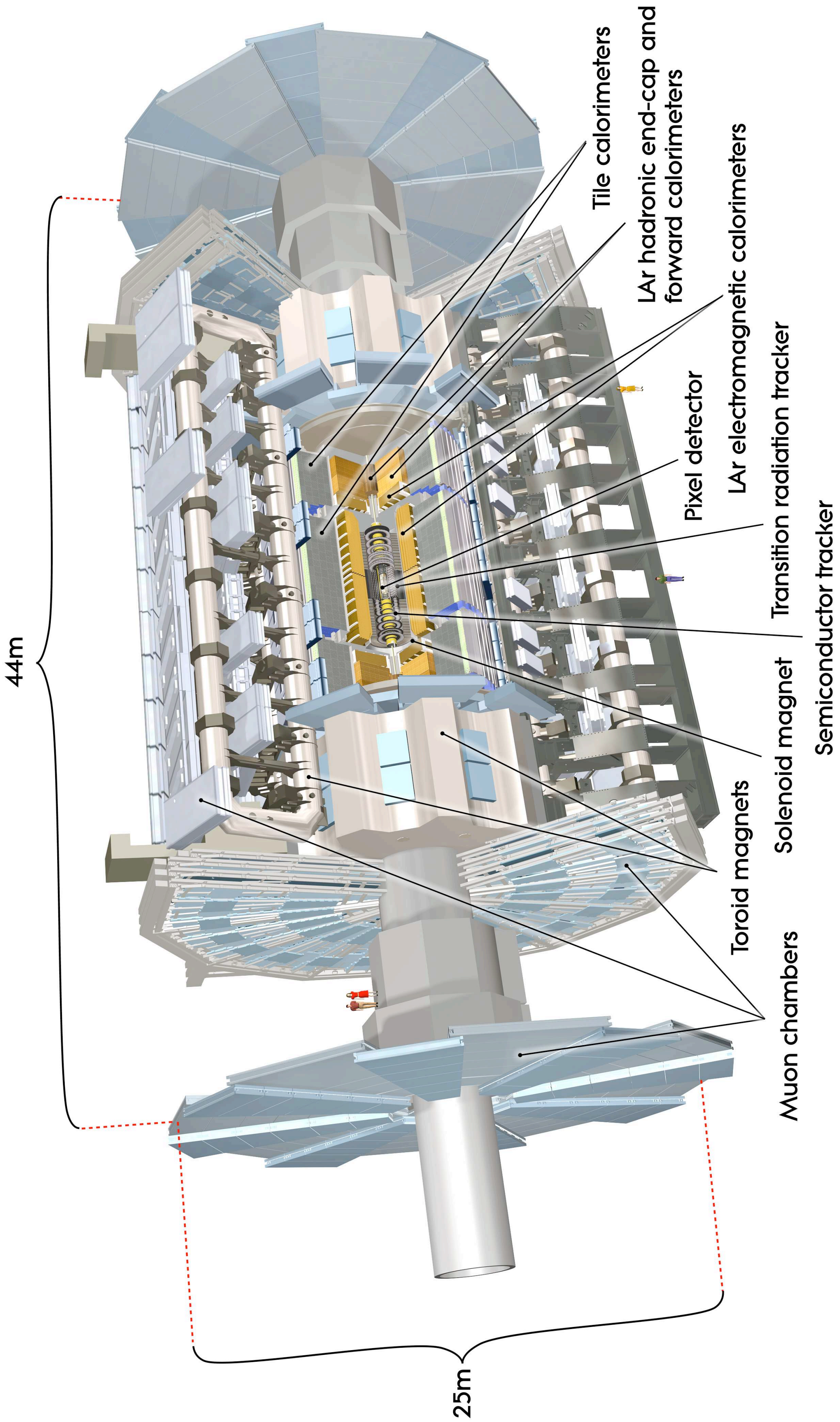


Figure 1.3: *ATLAS detector. The dimensions of the detector are 25 m in height and 44 m in length. The overall weight of the detector is approximately 7000 tonnes.*

The most important benchmark in the ATLAS performance has been set by the potential discovery of the Standard Model Higgs (see chapter 3 and [17]).Higgs boson searches are a challenging task since they demand combination of a high range of production and decay mechanisms of Higgs boson due to its low cross section and the high background. Searches for the Higgs boson beyond the Standard Model will be also feasible in ATLAS detector and, for such particles as the A and $H^{\pm}$ of the minimal supersymmetric extension of the Standard Model[7], an excellent ATLAS performance (see section 1.3) is required.

New heavy gauge bosons W' and Z' could be discovered for masses up to $\sim$ 6 TeV. To study their leptonic decays, high-resolution lepton measurements and charge identification are needed in the p_T -range of a few TeV. Another class of signatures of new physics may be provided by very high-p_T jet measurements. Furthermore, the decays of supersymmetric particles, such as squarks and gluinos, which would involve cascades which, if R-parity is conserved, always contain a lightest stable supersymmetric particle (LSP) will be also possible to be observed. As the LSP would interact very weakly with the detector, the experiment would measure a significant missing transverse energy, $\not{E}_T$, in the final state. The rest of the cascade would result in a number of leptons and jets. In schemes where the LSP decays into a photon and a gravitino, an increased number of hard isolated photons is expected.

Several new models which propose the existence of extra dimensions leading to a characteristic energy scale of quantum gravity in the TeV region will be tested in ATLAS. In terms of experimental signatures, this could lead to the emission of gravitons which escape into extra dimensions and therefore generate $\not{E}_T$, or of Kaluza-Klein excitations which manifest themselves as Z-like resonances (Z') in a mass range of a few TeV. Other experimental signatures could be anomalous high-mass di-jet production and miniature black-hole production with spectacular decays involving democratic production of fundamental final states such as jets, leptons, photons, neutrinos,W's and Z's.

Due to the experimental conditions at the LHC, the detectors require fast, radiation-hard electronics and sensor elements. Furthermore, a detector with granularity is required to deal with the variate particle fluxes and to reduce the influence of overlapping events. In addition, large acceptance in pseudorapidity with full azimuthal angle coverage is required. Summarising all the LHC experimental physics requirements, ATLAS (figure 1.3) detector is designed to fulfil the following expectations

- Tracking Detector
    - Good charged-particle momentum resolution and reconstruction efficiency are essential. Very high vertexing resolution is also required to contribute in tagging of τ-leptons and b-jets detecting secondary vertices.
- Calorimetry system
    - Electromagnetic calorimeter (EM): Very good performance needed for electron and photon identification and measurements.
    - Hadronic calorimeter: Full-coverage will provide accurate measurements for jets and $\not{E}_T$ calculations, elements that play a significant role in almost every physics analysis.

- Muon Spectrometer
  - Good muon identification and momentum resolution over a wide range of momenta and the ability to determine unambiguously the charge of high p_T muons are fundamental requirements of ATLAS detector.
- Trigger system
  - A highly efficient trigger system is required being capable to detect a wide range from low to high transverse-momentum objects rejecting sufficiently the background. Trigger implementation is done in two main steps: *(i)* at hardware level with dedicated sub-detectors and *(ii)* at software level.

The vector boson fusion Higgs search analysis described in this document rely on almost all aspects of the ATLAS detector, making use of electron and muon identification, charged particle reconstruction for hadronic tau decay identification, and finally jets and $\not{E}_T$.

1.2.1 Tracking Detector

The ATLAS Inner Detector (often called Inner Detector, ID) is designed to provide hermetic and robust pattern recognition, excellent momentum resolution and both primary and secondary vertex measurements [18] for charged tracks above typically 0.5 GeV.

The ID layout (figure 1.4) is contained within a cylinder of length 3.5 m and of radius of 1.2 m, within a solenoidal magnetic field of 2 T. The ID consists of three independent but complementary sub-detectors. Starting from the inner radii to larger, it consists of: (i) Pixel detector: a sub-detector designed for precision measurements very close to interaction point. (ii) Silicon microstrip sensors (SCT): a sub-detector complementary to Pixel detector dedicated to precise track and vertexing measurements. (iii) Transition-radiation tracker (TRT): it provides a continuous tracking to enhance the pattern recognition and improve the track momentum resolution. It also contributes to the identification of electrons.

The ID is designed in such a way that charged particles in general hit 3 layers of the Pixel detector (sec. 1.2.1.1), 8 layers of SCT (sec. 1.2.1.2), ~31 layers of TRT (sec. 1.2.1.3).

1.2.1.1 Pixel Detector

The pixel detector consists of 3 barrel layers of a radius of 50.5, 88.5 and 122.5 mm respectively and 3 end-cap layers that reach until 200 mm of radius. Its geometry is such that all charged particles independent to their η hit always 3 of these layers. This detector gives the highest spatial resolution for tracks and vertices and therefore the size of pixels has to be kept as small as possible giving a granularity $R - \phi \times z$ of $50 \times 400 \mu m^2$ in a total number of $\sim$ 80.4 million pixels. The resulting resolution is 10 μm in $R - \phi$ and 115 μm en z. This detector is particularly important for the identification of secondary vertices and thus of the quarks b. The innermost plane ($b-layer$) is exposed to a very high radiation level and might thus be replaced after a few years. To reduce the radiation damages and also to obtain the adequate performances, the detector is maintained at a temperature between -5 and -10 ^0C.

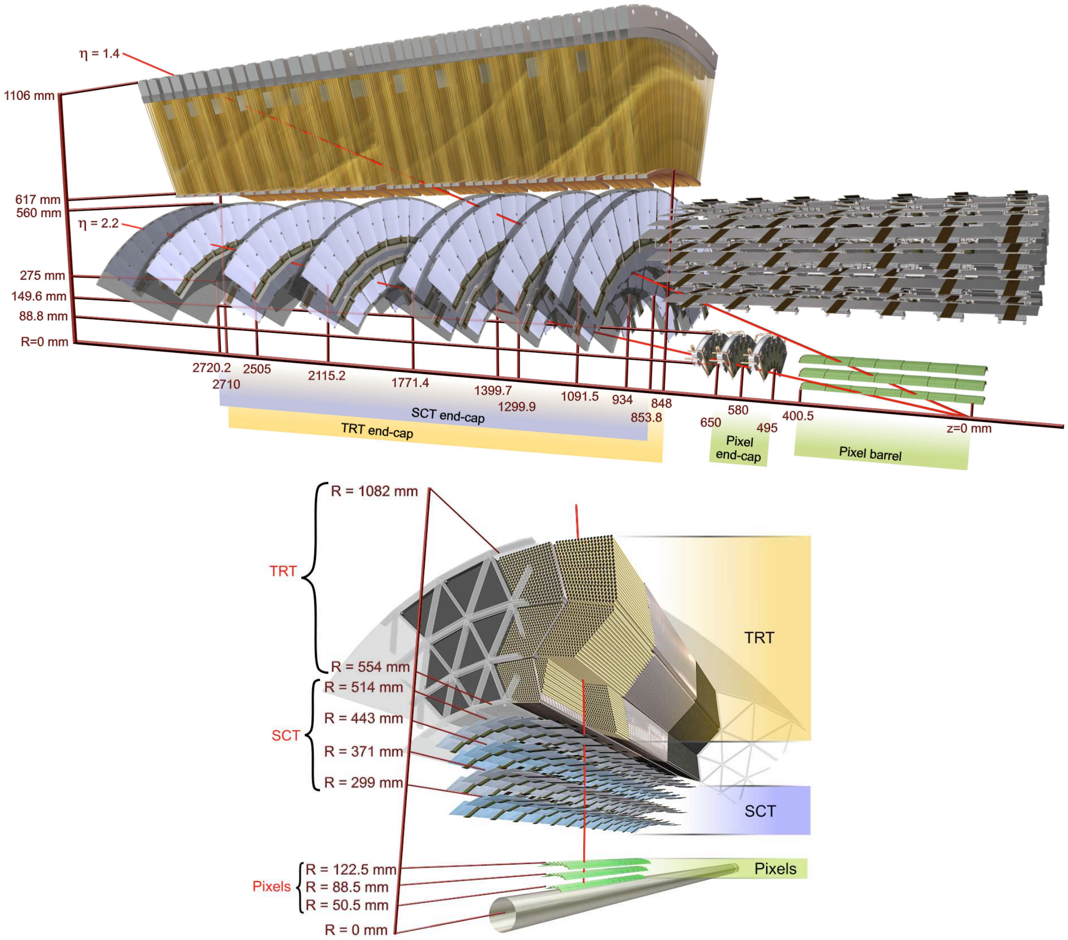


Figure 1.4: *Drawing showing the sensors and structural elements traversed by charged tracks for End-Cap part of inner detector (above) and barrel part (below). Track traverses successively, the three cylindrical silicon-pixel layers, the four cylindrical double layers of barrel silicon microstrip sensors (SCT), and approximately 36 axial straws contained in the barrel transition-radiation tracker (TRT).*

1.2.1.2 SCT

The SCT inner sub-detector is composed by eight strip layers placed in 2-by-2 structures resulting to four space points for each track that crosses them. In the barrel region, this detector uses small-angle (40 mrad) stereo strips to measure both coordinates, with one set of strips in each layer parallel to the beam direction, measuring $R-\phi$. They consist of two 6.4 cm long chained sensors with a strip pitch of 80 μm. In the SCT end-cap region, the detectors have a set of strips running radially and a set of stereo strips at an angle of 40 mrad. The mean pitch of the strips is also approximately 80 μm. The intrinsic accuracies per module in the barrel are 17 μm in $R-\phi$ and 580 mm in z and in the disks are 17 μm in $R-\phi$ and 580 mm in R. The total number of readout channels in the SCT is approximately 6.3 million.

1.2.1.3 TRT

A large number of hits (typically 31 per track) is provided by the 4 mm diameter straw tubes of the TRT, which enables track-following up to $|\eta| = 2.0$. The TRT only provides transverse information , for which it has an intrinsic accuracy of 130 μm per straw. In the barrel region, the straws are parallel to the beam axis and are 144 cm long, with their wires divided into two halves, approximately at $\eta = 0$. In the end-cap region, the 37 cm long straws are arranged radially in wheels. The total number of TRT readout channels is approximately 351.000 .

1.2.1.4 Inner detector Performance

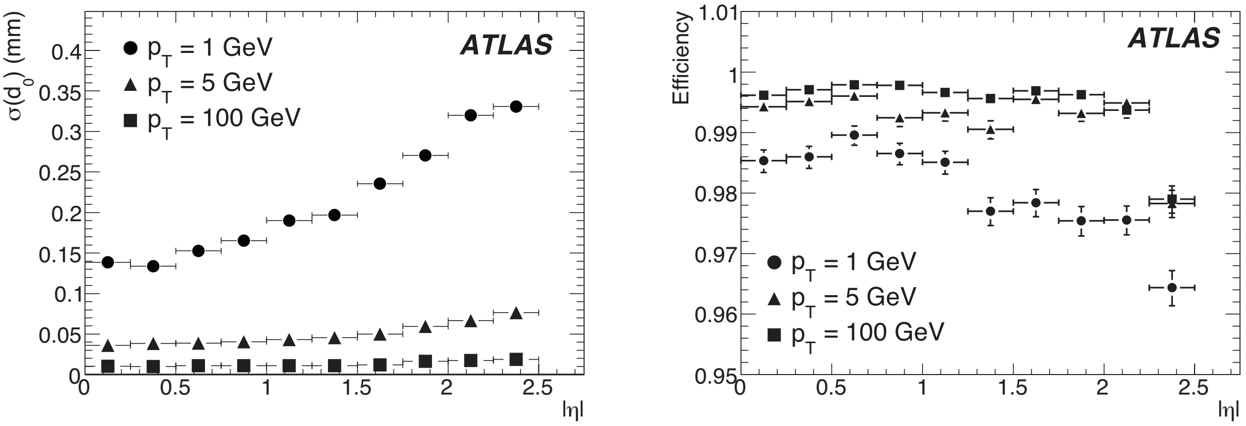


Figure 1.5: *(left) Transverse impact parameter, d_0, resolution as a function of $|\eta|$ for pions with p_T = 1, 5 and 100 GeV. (right)Track reconstruction efficiencies as a function of $|\eta|$ for muons with p_T = 1, 5 and 100 GeV [19].*

Figure 1.5 shows the efficiencies for reconstructing isolated muons and the transverse impact parameter resolutions for isolated pions, assuming the effects of misalignment, miscalibration and pile-up to be negligible. The TRT measurements are included in the track fits for tracks with $|\eta|$ <2.0, beyond this range, there are no further TRT measurements. As expected, a very good efficiency is achieved and it becomes larger and more uniform as a function of $|\eta|$ at higher energies. Similarly, moving to higher values of $|\eta|$ resolution becomes slightly worse.

1.2.2 Calorimetry

A layout of the ATLAS calorimeters is presented in figure 1.6. These calorimeters provide a coverage of $|\eta| < 4.9$, using different technologies suited to the widely varying requirements of the physics processes of interest and of the radiation environment over this large η range. Over the η region matched to the inner detector (section 1.2.1), the fine granularity of the EM calorimeter is ideally suited for precision measurements of electrons and photons. The coarser granularity of the rest of the calorimeter is designed in such a way to satisfy the physics requirements for jets and transverse missing energy ($\not{E}_T$) measurements.

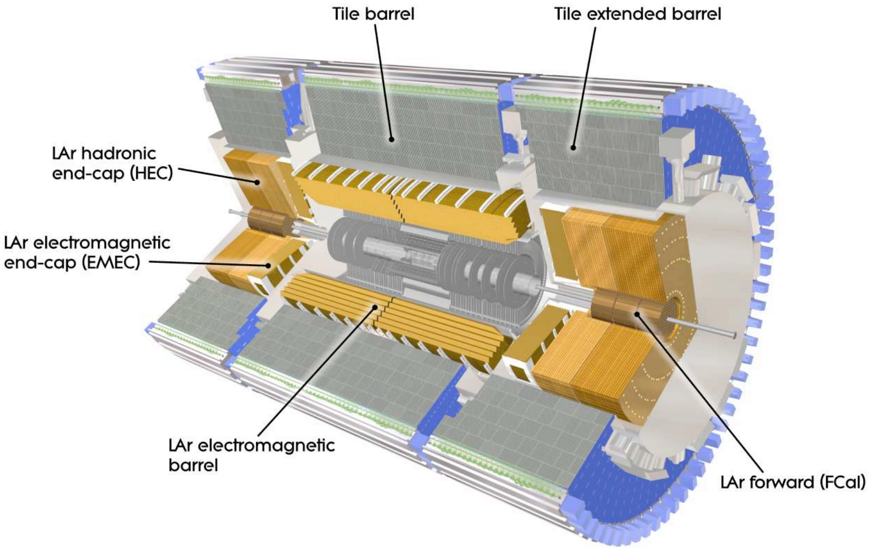


Figure 1.6: *Cut-away view of the ATLAS calorimeter system pointing out all the main calorimeter sub-detectors.*

Calorimeters must provide good measurement of electromagnetic and hadronic showers avoiding as much as possible the punch-through into the muon system. Therefore, calorimeter depth is an important design aspect. The total thickness of the EM calorimeter is > 22 radiation lengths (X_0) in the barrel and > 24 X_0 in the end-caps whereas for the hadronic Tile calorimeter, the $\sim$10 interaction lengths (λ) of active calorimeter are sufficient to provide good resolution for jets. The total thickness, is 11 λ for $\eta = 0$ and has been found both by measurements and simulations to be sufficient to reduce punch-through well below the irreducible level of prompt or decay muons as well as to keep the shower leakage in low level. Together with the large η-coverage, this thickness will also ensure a good $\not{E}_T$ measurement.

ATLAS calorimetry constituents (figure 1.6) are the following:

- Electromagnetic Calorimeter
  - LAr electromagnetic barrel
  - LAr electromagnetic end-cap
  - LAr forward calorimeter (inner part)
- Hadronic Calorimeter
  - Tile barrel
  - Tile Gap
  - Tile extended barrel
  - LAr hadronic end-cap

  - LAr forward calorimeter (outer part)

Following the technology used for the different sub-detectors of ATLAS calorimeters, the description which follows is divided in LAr calorimeters and Tile calorimeters.

1.2.2.1 LAr

The LAr calorimeter is divided into one EM barrel part ($|\eta| < 1.475$), two EM end-cap components ($1.375 < |\eta| < 3.2$), two hadronic end-caps ($1.5 < |\eta| < 3.2$) and two forward end-caps[1] ($3.10 < |\eta| < 4.83$). LAr calorimeter technology demands a very low temperature for operation therefore all the LAr components are housed in special cryostats (see later). EM barrel and the two EM end-caps consist of three successive layers and a presampler detector in the innermost region. The presampler covers the the region $|\eta| < 1.8$ and it provides a measurement of the energy lost in front of the electromagnetic calorimeters allowing the correction for the energy lost by electrons and photons upstream of the calorimeter. The hadronic end-caps and the FCal consist of four and three successive layers respectively. The main parameters of LAr calorimeters are presented in table 1.2.

Liquid Argon: a sampling calorimeter

The principle of a sampling calorimeter is discussed in this section. Such a calorimeter is composed in an alternative way by:

- *active medium*, in which an electrode collects the charges from the ionisation of the medium while particles are crossing it.
- *passive medium* (commonly called *absorber*), which consists of a dense material (lead in ATLAS calorimeter) and in which the particles lose most of their energy. That allows to reduce the size of the calorimeter since it shares with the active medium the task of making the particles loose their energy (what is actually the goal of the calorimetry in order to be able to measure particles' energy with the highest possible accuracy).

The successive layers of active medium and passive medium (figure 1.7) have to be placed in such a way that the particles cross them rather perpendicularly.

The *sampling fraction* is a quantity which defines the energy fraction deposited only in the active medium by the ionising particles, with the energy deposited in the whole calorimeter:

$$f_{sampling} = \frac{L_{active}(\frac{dE}{dx})_{active}}{L_{active}(\frac{dE}{dx})_{active} + L_{passive}(\frac{dE}{dx})_{passive}} \tag{1.5}$$

where L_{active} is the thickness of a layer of the active medium and $L_{passive}$ is the one of a layer of the passive medium. The closer the sampling fraction is to 1, the more accurate is the energy measurement of the calorimeter; for ATLAS electromagnetic Liquid Argon calorimeter $f_{sampling}$ is about 0.2. To obtain the appropriate sampling fraction for electrons, it is necessary to multiply to $f_{sampling}$, a factor called e/μ ratio, which takes into consideration the fact that contrary to a minimum ionizing particle, many low energy photons existing in

---

[1] FCal

	Barrel		Endcap	
LAr EM Calorimeter				
Granularity $\Delta\eta \times \Delta\phi$ versus $\|\eta\|$				
Presampler	0.025×0.1	$\|\eta\| < 1.52$	0.025×0.1	$1.5 < \|\eta\| < 1.8$
1^{st} layer	0.025×0.1	$\|\eta\| < 1.40$	0.050×0.1	$1.375 < \|\eta\| < 1.425$
	0.025×0.025	$1.40 < \|\eta\| < 1.475$	0.25×0.1	$1.425 < \|\eta\| < 2.5$
			0.1×0.1	$2.5 < \|\eta\| < 3.2$
2^{nd} layer	0.025×0.025	$\|\eta\| < 1.40$	0.050×0.025	$1.375 < \|\eta\| < 1.425$
	0.075×0.025	$1.40 < \|\eta\| < 1.475$	0.25×0.25	$1.425 < \|\eta\| < 2.5$
			0.1×0.1	$2.5 < \|\eta\| < 3.2$
3^{rd} layer	0.050×0.025	$\|\eta\| < 1.35$	0.050×0.025	$1.5 < \|\eta\| < 2.5$
Number of readout channels				
Presampler	7808		1536 (both sides)	
1^{st},2^{nd} & 3^{rd} layer	101760		62208 (both sides)	
LAr Hadronic end-cap				
$\|\eta\|$ coverage			$1.5 < \|\eta\| < 3.2$	
Number of layers			4	
$\Delta\eta \times \Delta\phi$			0.1×0.1	$1.5 < \|\eta\| < 2.5$
			0.2×0.2	$2.5 < \|\eta\| < 3.2$
Readout channels			5632 (both sides)	
LAr Forward Calorimeter				
$\|\eta\|$ coverage			$3.1 < \|\eta\| < 4.9$	
Number of layers			3	
Granularity $\Delta\chi \times \Delta y$ (cm)			FCal1: 3.0×2.6	$3.15 < \|\eta\| < 4.83$
			FCal2: 3.3×4.2	$3.20 < \|\eta\| < 4.81$
			FCal3: 3.3×4.2	$3.29 < \|\eta\| < 4.75$
Readout channels			3524 (both sides)	

Table 1.2: *Granularity, η coverage and read-out channels of LAr calorimeter system.*

an electromagnetic shower produced in the calorimeter, are absorbed due to the photoelectric effect. Considering the material Z^5 dependance of the photoelectric effect cross section, these photons are mainly stopped in the absorber, resulting in a reduction of the measured signal. The e/μ ratio is determined by measurements during test-beam and it is found to be 3/4 [20]. In the case of ATLAS, the selected active medium is liquid Argon. This choice is justified by the following arguments:

- signal uniformity, allowing the application of simple calibration methods (using injected current).
- stability and high resistance in radiations, during the full duration of ATLAS operation.
- a high density[1], higher than the one of a gas or a liquid which optimises the ionization process.

---

[1] 1392.8 kg/m^3 (1.013 bar at boiling point)

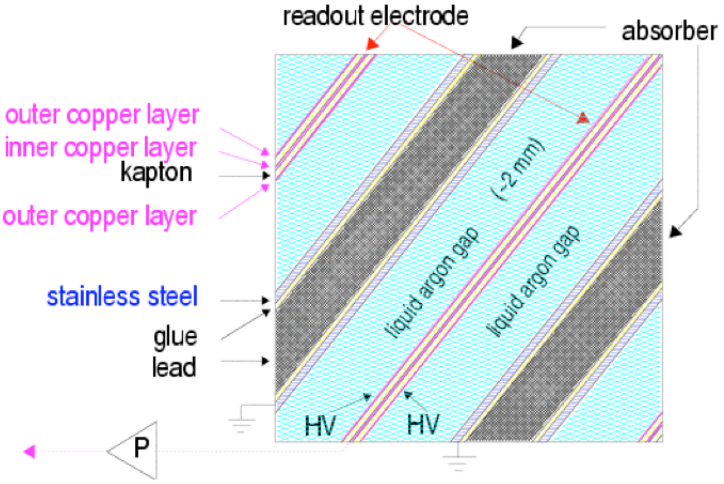


Figure 1.7: *Diagram showing the alternative layers of lead absorbers, gaps of liquid argon and the electrodes in the punts parts of the LAr calorimeter.*

The main difficulty that had to be over-passed was that LAr calorimeter operates at a very low temperature (between 87 and 89 K), resulting to a placement of the whole LAr calorimeter into a cryostat, which introduces dead matter not useful for particles detection, in which particles loose part of their energy. Special corrections are being applied to overcome this issue.

An accordion geometry has been chosen for the absorbers and the electrodes of the barrel and endcap electromagnetic calorimeters (see figures 1.8 and 1.9). Such a geometry provides naturally a full coverage in ϕ without any cracks, and a fast extraction of the signal at the rear or at the front of the electrodes. In the barrel, the accordion waves are axial and run in ϕ, and the folding angles of the waves vary with radius to keep the liquid-argon gap constant, thus a constant sampling fraction. In the end-caps, the waves are parallel to the radial direction and run axially. Since the liquid-argon gap increases with radius in the end-caps, the wave amplitude and the folding angle of the absorbers and electrodes vary with radius. All these features of the accordion geometry lead to a very uniform performance in terms of linearity and resolution as a function of ϕ.

The signal readout in LAr calorimeter consists of cells, which are defined by pads etched on the central foil in each gap. The arrangement of these pads provides a semi-pointing geometry. The size of the readout cells is in $(\Delta\eta \times \Delta\phi)$ is presented in table 1.2 and it can vary depending on the LAr layer they are found. As can be seen from figure 1.8, the first layer is finely segmented along η, as for example in the barrel where there are eight strips in front of a middle cell. One can note however the coarser granularity of the first layer in the edge zones of the barrel and end-caps, as explicitly given in table 1.2. The second layer collects the largest fraction of the energy of the electromagnetic shower, and the third layer collects only the tail of the electromagnetic shower and is therefore less segmented in η.

A high voltage of 2000 V is applied to the electrodes, whereas the absorbers are connected to ground resulting in a voltage of 0V, for the LAr electromagnetic barrel, while a varying high voltage is applied to the LAr electromagnetic end-cap electrodes, depending on the radius. The reason is that the gap size between liquid argon and absorbers in end-caps,

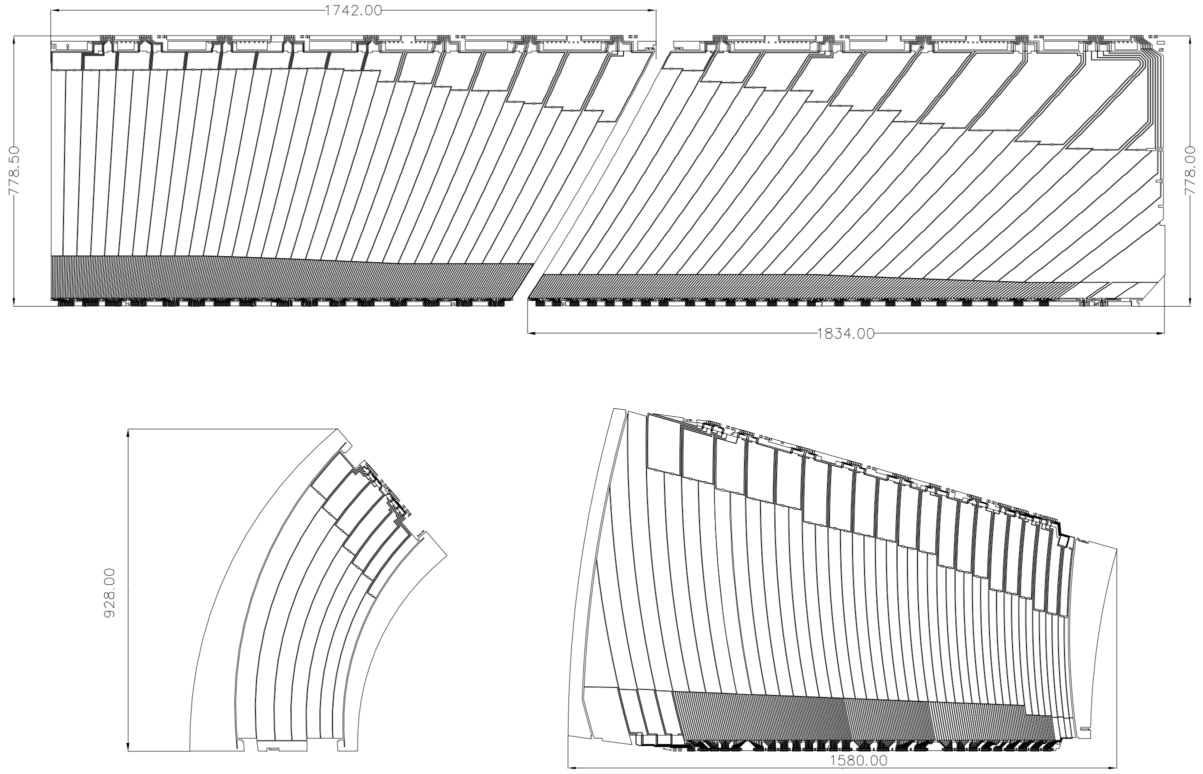


Figure 1.8: *Layout of the signal layer for the four different types of electrodes before folding. The two top electrodes are for the LAr EM barrel and the two bottom electrodes are for the LAr EM end-cap inner (left) and outer (right) wheels. Dimensions are in millimetres. The two or three different layers in depth are clearly visible.*

varies according to the radius resulting in a varying sampling fraction. This effect is thus compensated to first order by the application of a respectively varying high voltage. High voltage is supplied to the electrodes by a single connection at the back of each electrode for a $\Delta\eta$=0.2 sector, and it is distributed to all cells through resistors.

As already mentioned, the absorbers are made of lead plates, to which two stainless-steel sheets (0.2 mm thick) are glued to provide mechanical strength. The lead plates in the barrel have a thickness of 1.53 mm for $|\eta| < 0.8$ and of 1.13 mm for $|\eta| > 0.8$. The change in lead thickness at $|\eta| = 0.8$ limits the decrease of the sampling frequency[1], as $|\eta|$ increases. In the end-cap calorimeters, the plates have a thickness of 1.7 mm for $|\eta| < 2.5$ and of 2.2 mm for $|\eta| > 2.5$. The readout electrodes [21] are located in the gaps between the absorbers and consist of three conductive copper layers separated by insulating polyimide sheets (figure 1.7). The two outer layers are at the high-voltage potential and the inner one is used for reading out the signal via capacitive coupling. The segmentation of the calorimeter in η and in depth is obtained by etched patterns on the different layers, as shown in figure 1.8. Each barrel gap between two absorbers is equipped with two electrodes, one type for $|\eta| < 0.8$ and another one for $|\eta| > 0.8$. Similarly, each end-cap gap between two absorbers is equipped

---

[1] the number of successive layers of liquid argon and absorber crossed by an electromagnetic shower

with one type of electrode for $|\eta| < 2.5$ and with another for $|\eta| > 2.5$.

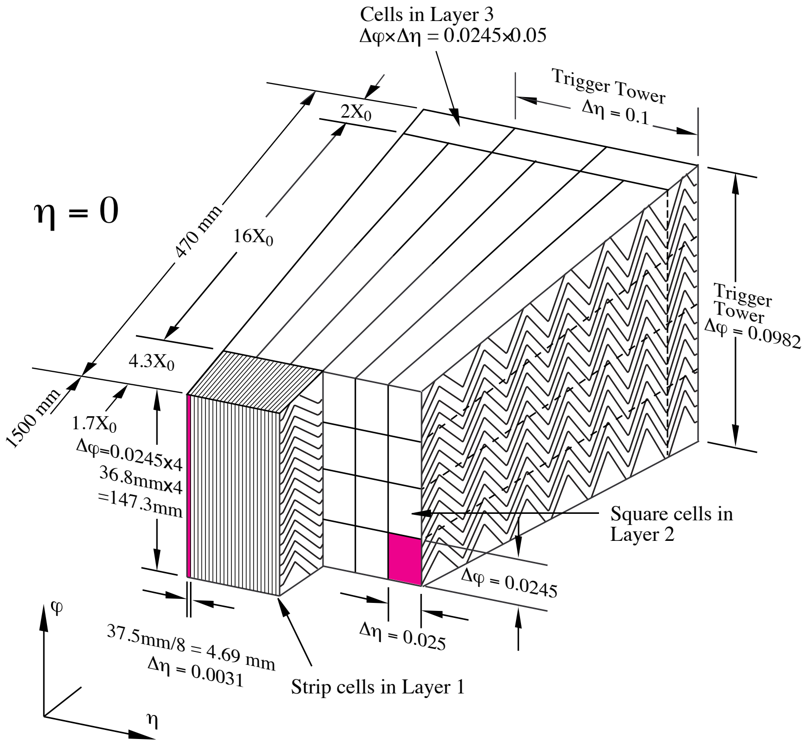


Figure 1.9: *Sketch of a LAr EM barrel module where the different layers are clearly visible The granularity in η and ϕ of the cells of each of the three layers and of the trigger towers is also shown.*

The hadronic end-cap calorimeter (HEC)[43],[23] is a liquid-argon sampling calorimeter with a flat-plate design, which covers the range $1.5 < |\eta| < 3.2$. The HEC shares each of the two liquid-argon end-cap cryostats with the electromagnetic end-cap (EMEC) and forward (FCal) calorimeters (see figure 1.10). The HEC consists of two wheels in each end-cap cryostat: a front wheel (HEC1) and a rear wheel (HEC2), each wheel containing two longitudinal sections. The wheels are cylindrical with an outer radius of 2030 mm.

An important aspect of the HEC is its ability to detect muons and to measure any radiative energy loss.

The forward calorimeters (FCal) are located in the same cryostats as the end-cap calorimeters and provide coverage over $3.1 < |\eta| < 4.9$ (see figure 1.10). FCal is composed by three layers (see figure 1.11) which are located at high η, at a distance of approximately 4.7 m from the interaction point, they are exposed to high particle fluxes. This has resulted in a design with very small liquid-argon gaps, which have been obtained by using an electrode structure of small-diameter rods, centred in tubes which are oriented parallel to the beam direction.

1.2.2.2 Energy reconstruction in LAr calorimeter

As described in previous section, the particles interact with the calorimeter matter which are the absorbers and the active medium. At the active medium (LAr in ATLAS) the signal

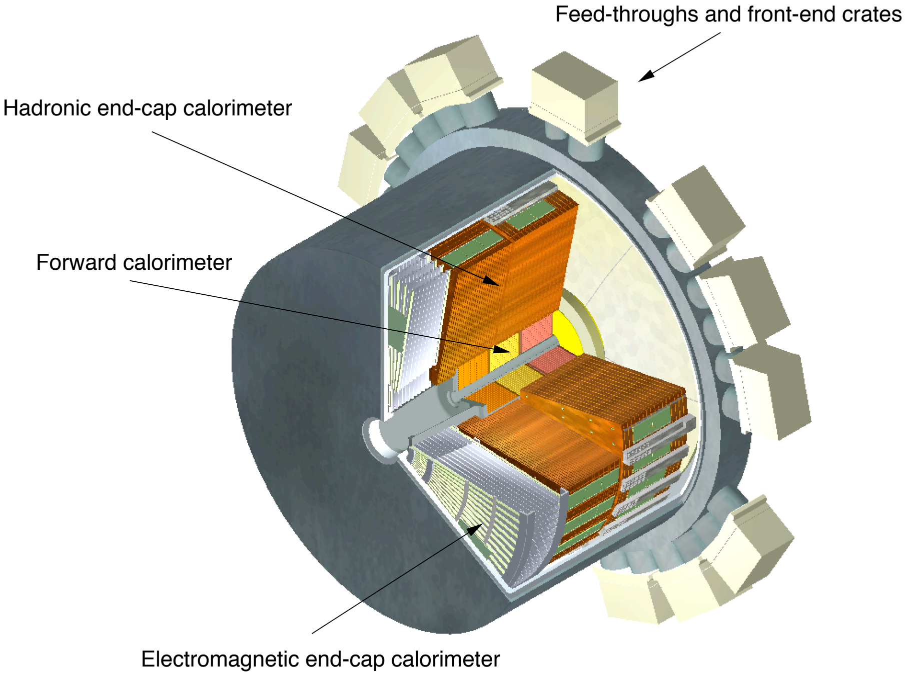


Figure 1.10: *Cut-away view of an end-cap cryostat showing the positions of the three LAr end-cap calorimeters. The outer radius of the cylindrical cryostat vessel is 2.25 m and the length of the cryostat is 3.17 m.*

is generated and it is collected by the read-out electrodes. The overall readout architecture of the LAr calorimeters is sketched out in figure 1.12 and described as following:

Triangular signal

The electric field applied between a read-out electrode and the absorber, collects the electrical charges produced by the ionization of liquid Argon. Thus, the induced current has a triangular form in function of time (see Figure 1.13):

$$I(t) = \frac{Q_0}{t_d}(1 - \frac{t}{t_d}) \tag{1.6}$$

where t_d is the average drift time of the electrons in Argon, about 450 ns, and Q_0 is the total charge deposit in Argon. This signal passes afterwards through an electronic card in the Front End Board (FEB).

Pre-amplification and shaping

Passing through the FEB's, the triangular signal is amplified by pre-amplifiers to obtain a signal superior of the electronics noise level. Afterwards, signal is being transported to special shapers which are bipolar filters type CR - RC^2 which give the bipolar pulse shape of figure 1.13. The goal of such a procedure is to make the whole time integral to be zero.

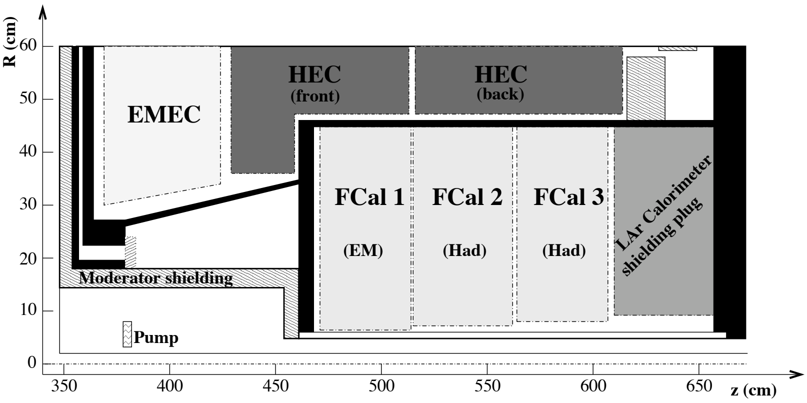


Figure 1.11: *Schematic diagram showing the three FCal modules located in the end-cap cryostat. The diagram has a larger vertical scale for clarity.*

Such a configuration allows the average signal to be null even in case of pile-up.

Sampling

The bipolar shape is sampled every 25 ns (every bunch crossing in nominal LHC configuration) as shown in figure 1.13. These samplings are digitised by an analogical-numerical converter (ADC) and then they leave the FEB'*s* being transmitted to the ROD. The LAr calorimeter records only the 5 first samplings in nominal configuration, but for special studies more samplings can be recorded.

Gain

Another function of the shapers described in previous paragraph is to amplify even more the signal, using 3 different types of linear gain:

- High gain: a factor of 93
- Medium gain: a factor of 9.3
- Low gain: a factor of 1

Signal is digitised for the three gains, in parallel. From these 3 gains, only one is finally kept according to certain thresholds in ADC counts on the Medium gain. This allows to balance precision and large dynamic range.

Calibration Board

The purpose of the calibration card is to establish a correspondence ADC counts↔ current. During calibration run (typically during period without beams), it injects a signal at the detector output in the liquid Argon. The calibration signal is produced in the following way: a digital-analogical converter (DAC), produces a voltage, converted then into continuous current feeding an inductance (RL); the discharge produces a decreasing exponential

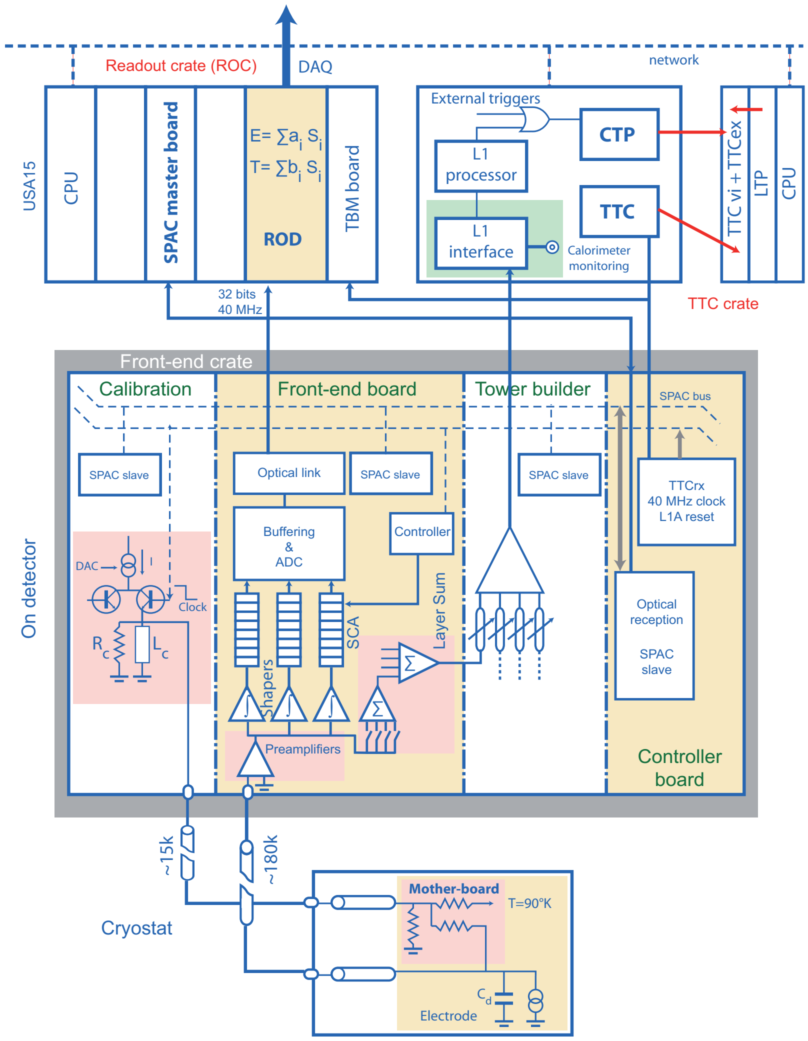


Figure 1.12: *Block diagram depicting the architecture of the overall LAr readout electronics [16].*

distribution mimicking the triangular signal. The signal is then injected to the electrodes through an injection resistance (R_{inj}) and then passes by all the read-out electronics chain until its conversion into ADC counts.

The electronics gain of each channel is determined through a procedure called ramp where for a given number DAC values (∼10), the average ADC for certain number of events is computed. The fit of DAC in function of ADC provides this gain, done independently for the high, medium and low cases. Typically, the non-linearity with respect to this slope is of the level of 0.1 %.

Knowing the correspondences ADC↔DAC(ramp), DAC↔tension and tension↔current (in function of R_{inj}), it is then possible to obtain the global correspondence ADC↔energy. This last factor current↔energy is obtained by test beam and simulation measurements.

It is important to highlight that the signal produced by the calibration card is exponential, similar to the triangular signal of the calorimeters but even after the shaping the two signals

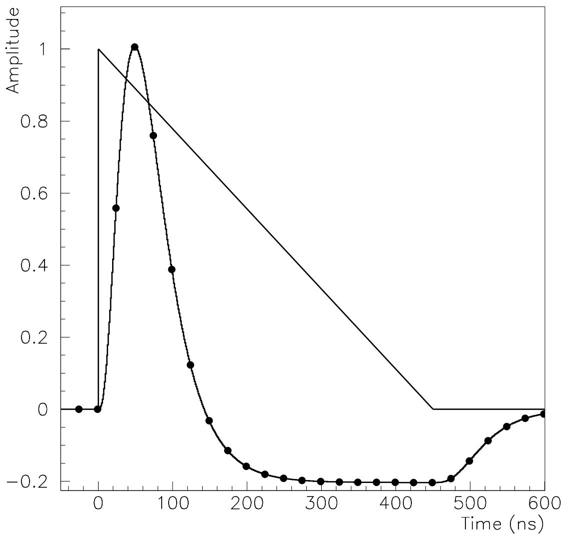


Figure 1.13: *Amplitude versus time for triangular pulse of the current in a LAr barrel electromagnetic cell and of the FEB output signal after bi-polar shaping. Also indicated the sampling points every 25 ns.*

(ionization and calibration) are not exactly identical.

Pedestal and Electronics noise

In absence of input signal in the calorimeter, the signal at the output of the electronics chain is called pedestal. For each channel, the signal of an event varies from one event to another through a Gaussian distribution, following the effect of the noise produced by the whole read-out electronics chain which treats the signal. This noise, which depends on the shapers gain, is called electronics noise. It is measured through the RMS of the signals of a given channel. Such a measurement is performed only to one pedestal for a given gain. Similarly, the pedestal can be measured by taking this time the average value of the signals. Approximatively, it has a value of 1000 ADC counts, allowing the measurement of negative part at the bipolar shape of the (figure 1.13). Pedestal and Electronics noise are defined as the following:

$$Pedestal =< E_{ADC} > \tag{1.7}$$

$$Electronics\ noise = \sqrt{< E_{ADC}^2 > - < E_{ADC} >^2} = \sigma \tag{1.8}$$

where E_{ADC} is the signal energy in ADC counts, only produced by the electronics.

ROD

ROD boards (ReadOut Driver), after having received the information from ADC and information from the calibration board, perform the procedure of *optimal filtering* (see next paragraph). Then ROD boards send the signal to ROB modules (ReadOut Buffer), in the form of a triplet: energy, time, and data quality. ROB'*s* are the input of the data acquisition system (DAQ). If the event is picked by the trigger system, this information is written on disk. For each event, all cells information is written out without any zero suppression.

OFC method

To minimise the influence of electronics noise and the pile-up effect, and to ensure also the condition that a shift on time of the signal will not alter the energy measurement, the ADC samplings are combined using special weights as coefficients, which are calculated by the optimal filtering method[24]. Thus, these coefficients are called optimal filtering coefficients (OFC). OFC method is described in following paragraphs. The observed signal is expressed in function of time:

$$S(t) = Ag(t+\tau) + n(t) \sim Ag(t) + A\tau g'(t) + n(t) \tag{1.9}$$

where A is the amplitude of the signal and g defines its form (Figure 1.13), g' is the time derivative of g, and τ is the time shift which is taken as small enough so that the above approximation described in equation 1.9 is valid, $n(t)$ is the noise contribution. In time samplings terms (t_k), where t_k corresponds to the time of beam crossing (see previous paragraph **Sampling**), equation 1.9 can be written:

$$S_k \sim A_{g_k} + A\tau_{g_k} + n_k \tag{1.10}$$

The optimal filtering method consists of minimising the variance of the two following quantities U,V:

$$U = \sum_k a_k S_k \quad , \quad <U>= A \tag{1.11}$$

$$V = \sum_k b_k S_k \quad , \quad <V>= A\tau \tag{1.12}$$

The coefficients a_k and b_k are called Optimal Filtering Coefficients (**OFC**). The equations 1.11,1.12 lead to the following conditions, taking into consideration as well that the noise has an average value equals to 0 ($<n_k>=0$).

$$\sum_k a_k g_k = 1 \qquad \sum_k a_k g'_k = 0 \tag{1.13}$$

$$\sum_k b_k g_k = 0 \qquad \sum_k b_k g'_k = 1 \tag{1.14}$$

These conditions allow us to calculate the variances of U and V :

$$Var(U) = \sum_{j,k} a_j a_k AC_{jk} \qquad Var(V) = \sum_{j,k} b_j b_k AC_{jk} \tag{1.15}$$

where AC_{jk} correspond to the elements of the auto-correlation matrix noise (filled from the variance of $n_j n_k$). By minimising these variances, using the method of Lagrange multipliers, the following expression for the OFC can be obtained:

$$a = \frac{(g' \cdot Rg')Rg \ - \ (g \cdot Rg')Rg'}{(g \cdot Rg)(g' \cdot Rg') \ - \ (g \cdot Rg')^2} \tag{1.16}$$

$$b = \frac{(g \cdot Rg)Rg' \ - \ (g \cdot Rg')Rg}{(g \cdot Rg)(g' \cdot Rg') \ - \ (g \cdot Rg')^2} \tag{1.17}$$

where the matrix R is the inverse of the auto-correlation matrix of AC. The energy and the time shift are reconstructed from the different samplings using the equations 1.11 and 1.12. In ATLAS LAr it is planned to use 5 samplings centered on the maximum of the bipolar curve (figure 1.13). The energy in MeV is finally measured by the following formula:

$$E = CONV_{ADC \rightarrow MeV_{EM}} \times \sum_i (Sampling_i - Pedestal) \cdot OFCa_i \tag{1.18}$$

The OFC are computed for each cell and stored in a database. Specific computations are done for cosmics data taking as explained in the following chapter (section 2.3).

The calculated energy, in MeV, is the particle energy deposit only in the active medium which is the LAr. To pass the initial particle energy at the interaction point, it is necessary to multiply with the $f_{sampling}$ and the e/μ ratio presented in previous section.

The energy is by convention at the electromagnetic scale, thus appropriate for electrons and photons. Further calibration for e.g. jets is applied at a later stage in reconstruction.

A summary of the most important characteristics of LAr calorimeter components is presented in the following table 1.3. The stochastic and constant terms for the energy resolution are measured in older test beam studies [16].

1.2.2.3 Tile

The Tile calorimeter is placed directly outside the EM calorimeter envelope (figure 1.6). Its barrel covers the region $|\eta| < 1.0$, and its two extended barrels the range $0.8 < |\eta| < 1.7$. It is a sampling calorimeter using steel as absorber and scintillating tiles as active material. Between the Tile barrel and the Tile extended barrel, the Tile Gap module is placed, a small Tile sub-detector which is formed only by scintillating tiles. Radially, the tile calorimeter extends from an inner radius of 2.28 m to an outer radius of 4.25 m. It is segmented in depth in three layers, approximately 1.5, 4.1 and 1.8 interaction lengths (λ) thick for the barrel and 1.5, 2.6, and 3.3 λ for the extended barrel. The total detector thickness at the outer edge of the tile-instrumented region is 9.7 λ at $\eta = 0$. The tile calorimeter is segmented in three-dimensional cell structure, creating a projective geometry for trigger and energy reconstruction (figure 1.14). $\Delta\phi \times \Delta\eta$ granularity equals to 0.1×0.1 in the first two samplings and 0.1×0.2 in the outermost sampling. The energy resolution σ_E/E was measured in test beam test with pions resulting to a stochastic term of 56.4% and to a constant term of 5.5%.

LAr components	Coverage η	Energy Resolution σ_E/E	Radiation / Interaction Length
EM Barrel / End-Cap	$\|\eta\| <3.2$	$10\%/\sqrt{E(GeV)} \oplus 0.7\%$	22-30 X_0
Hadronic End-Cap	$1.5< \|\eta\| <3.2$	$50\%/\sqrt{E(GeV)} \oplus 3\%$	~11 λ in total
Forward Calorimeter	$3.1< \|\eta\| <4.9$	$100\%/\sqrt{E(GeV)} \oplus 10\%$	~11 λ in total

Table 1.3: *Summary of LAr calorimeter main characteristics. The stochastic and constant terms are provided for the energy resolution.*

1.2.2.4 Noise in Calorimeters

Given the number of calorimeter read-out cells (~180k), a significant contribution of noise is caused by the reading of cells signal. This noise is mainly dominated by the noise deriving from the pre-amplifiers [27] and it is measured for each cell in randomly triggered events. For the cosmics analysis shown in this document, the noise was measured from random trigger stream on run 92048 for LAr and on run 90272 for Tile (see section 2.2.1 for random trigger definition). Figures 1.15 and 1.16 show this measured electronic noise at cell level (σ_{noise}), as a function of η for all longitudinal compartments of LAr and Tile calorimeters.

In the EM part of LAr calorimeters, a noise variation of 10 to 100 MeV is measured whereas for the hadronic part of LAr calorimeter the respective range varies from 100 to 600 MeV. For Tile noise, the variation is comparable to the one of EM LAr and it varies from 20 to 70 MeV in mean values.

1.2.3 Muon Spectrometer

A layout of ATLAS muon spectrometer is shown in figure 1.17, which was designed having as a goal to achieve resolution of approximatively 10 % for a 1 TeV muon. Muon momenta measurement is based on the magnetic deflection of muon tracks in the large superconducting air-core toroid magnets, instrumented with separate trigger and high-precision tracking chambers. Over the range $|\eta| < 1.4$, magnetic bending is provided by the large barrel toroid. For $1.6 < |\eta| < 2.7$, muon tracks are bent by two smaller end-cap magnets inserted into both ends of the barrel toroid. Over $1.4 < |\eta| < 1.6$, usually referred to as the transition region, magnetic deflection is provided by a combination of barrel and end-cap fields. This magnet configuration provides a field which is mostly orthogonal to the muon trajectories, while minimising the degradation of resolution due to multiple scattering and in average value equals to 0.5 T.

In the barrel region, tracks are measured in chambers arranged (MDT) in three cylindrical layers around the beam axis; in the transition and end-cap regions, the chambers (CSC) are

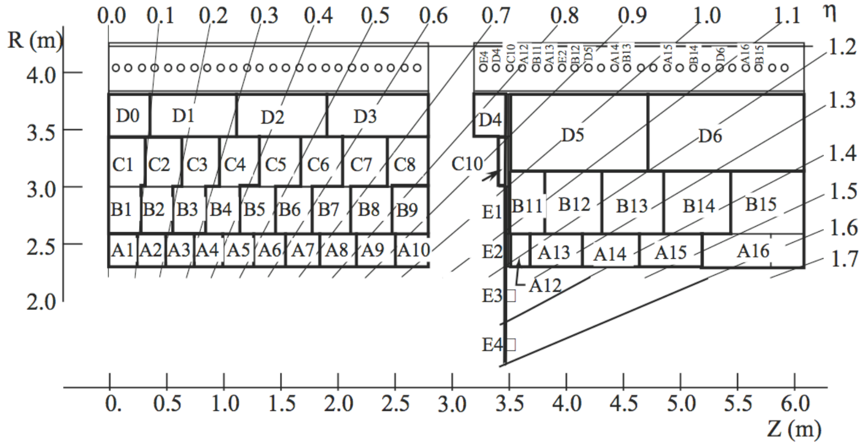


Figure 1.14: *Cell granularity for Tile Calorimeter shown in th r-z plane. Note the 3 longitudinal layers (A, BC, D).*

installed in planes perpendicular to the beam, also in three layers.

1.2.3.1 Monitored Drift Tubes (MDT) and Cathode Strip Chambers (CSC)

The muon precision momentum measurement is performed by the Monitored Drift Tube chambers (MDT) at the barrel region ($|\eta| < 2.0$) and by Cathode-Strip Chambers (CSC) in the end-cap region ($2.0 < |\eta| < 2.7$). MDT are typical drift tubes (figure 1.18) of a diameter of ∼3cm. In the center of the tubes a wire of tungsten-rhenium with a diameter of $50\mu m$ and holding a tension of $3kV$ collects all the electrons resulting from ionization. These tubes are placed in the Muon Spectrometer forming MDT chambers which consist of three to eight layers of drift tubes. The layers distance is 5 m in average values and the resolution achieved is in average 800 mm per tube, or about 35 mm per chamber. The CSC system consists of two disks with eight chambers each (eight small and eight large). The resolution achieved is $40\mu m$ in (R) and 5 mm in ϕ.

1.2.3.2 Resistive Plate chambers (RPC) and Thin-gap chambers (TGC)

RPC and TGC are sub-detectors which provide the first level (L1) of muon trigger for barrel end end-cap regions respectively. As shown in figure 1.19 MDT are embedded inside two layers of RPC and CSC are embedded in two layers of TGC respectvely. These trigger chambers provide a very fast response within a spatial resolution in the order of a few μm.

1.2.4 Trigger System

The proton-proton interaction rate at the design luminosity of $10^{34} cm^{-2} s^{-1}$ is approximately 1 GHz, while the event data recording, based on technology and resource limitations, is limited to about 200 Hz. This requires an overall rejection factor of 5×10^6 against minimum-bias processes while maintaining maximum efficiency for keeping interesting events which

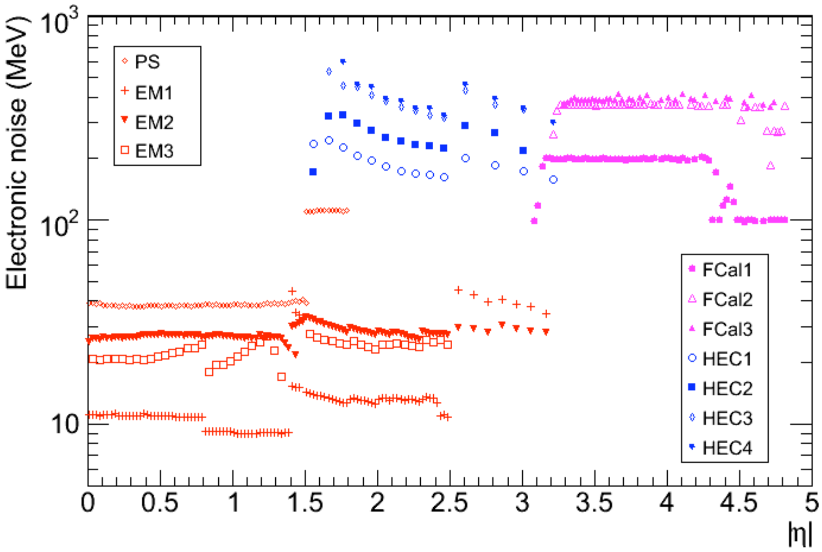


Figure 1.15: *Electronic noise (σ_{noise}) at EM scale in individual cells of the various longitudinal layers of the calorimeters as a function of η for LAr sub-detectors [25]. Noise values are averaged over φ and positive and negative η values.*

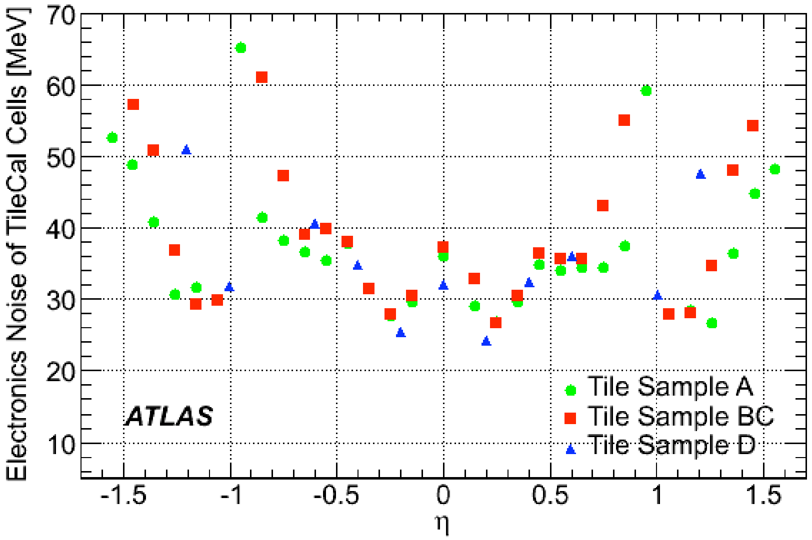


Figure 1.16: *Electronic noise (σ_{noise}) at EM scale in individual cells of the various longitudinal layers of the calorimeters as a function of η for Tile layers [26]. Noise values are averaged over φ.*

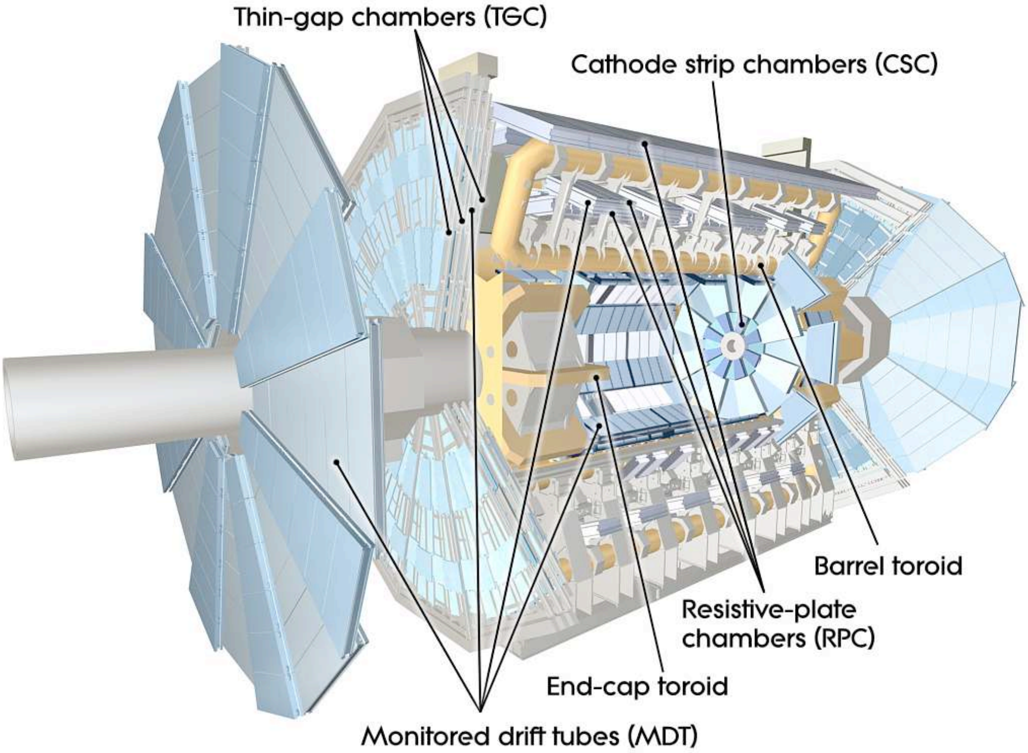


Figure 1.17: *ATLAS muon spectrometer layout.*

lead to new physics. This rejection is achieved by the ATLAS trigger system.

ATLAS trigger consists of three levels of event selection: Level-1 (L1), Level-2 (L2), and event filter. The L2 and event filter together form the High-Level Trigger (HLT). The L1 trigger is implemented using custom-made electronics, while the HLT is almost entirely based on algorithm level.A block diagram of the trigger is shown in figure 1.20.

The L1 trigger searches for signatures from high-p_T muons, electrons/photons, jets, and τ-leptons decaying into hadrons. It also selects events with large missing transverse energy ($\not{E}_T$) and large total transverse energy. The L1 trigger uses reduced-granularity information from a subset of detectors: the Resistive Plate Chambers (RPC) and Thin-Gap Chambers (TGC) for high-p_T muons, and all the calorimeter sub-systems for electromagnetic clusters, jets, τ-leptons, $\not{E}_T$, and large total transverse energy. The maximum L1 accept rate which the detector readout systems can handle is 75 kHz, and the L1 decision must reach the front-end electronics within 2.5 ms after the bunch-crossing with which it is associated. The L2 trigger is seeded by Regions-of-Interest (RoI). These are regions of the detector where the L1 trigger has identified possible trigger objects within the event. The L2 trigger uses RoI information on coordinates, energy, and type of signatures to limit the amount of data which must be transferred from the detector readout. The L2 trigger reduces the event rate to below 3.5 kHz, with an average event processing time of approximately 40 ms. The event filter uses offline analysis procedures on fully-built events to further select events down to a rate which can be recorded for subsequent offline analysis. It reduces the event rate to approximately 200 Hz, with an average event processing time of order four seconds. The HLT

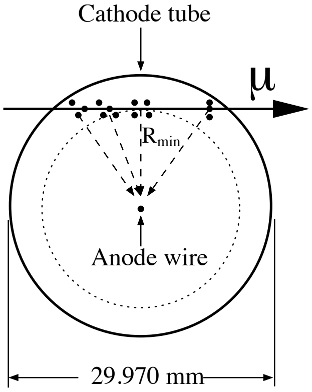


Figure 1.18: *Cross-section of a MDT tube.*

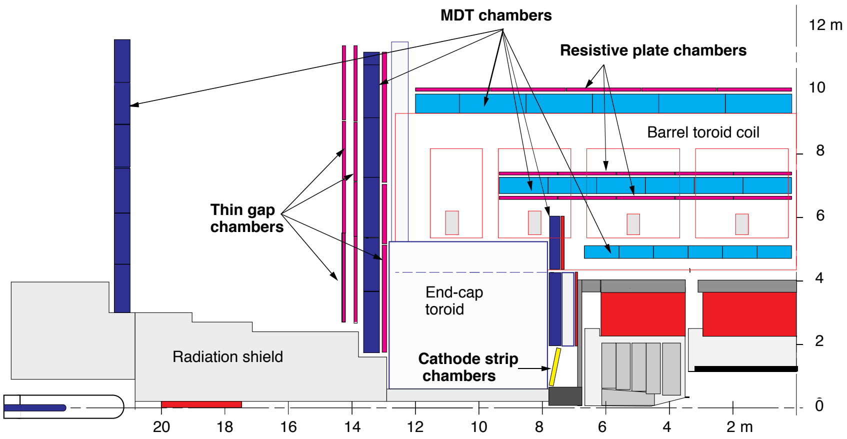


Figure 1.19: *ATLAS muon spectrometer cross-section.*

algorithms use the full granularity and precision of calorimeter and muon chamber data, as well as the data from the inner detector, to refine the trigger selections. Better information on energy deposition improves the threshold cuts, while track reconstruction in the inner detector significantly enhances the particle identification (for example distinguishing between electrons and photons). The event selection at both L1 and L2 primarily uses inclusive criteria, for example high-E_T objects above defined thresholds. One exception is the L2 selection of events containing the decay of a b-hadron, which requires the reconstruction of exclusive decays into particles with low momentum.

Figure 1.21 illustrates the expected event rates for several physics processes at the nominal LHC luminosity of $L = 10^{34} cm^{-2} s^{-1}$, superimposed with the trigger event rates inputs and outputs. It is obvious that the level of the HLT output rate (at the order of kHz) is compatible with the expected event rate of all interesting physics processes at the LHC, such as electroweak, Higgs, SUSY, etc.

1.2.5 ATHENA: ATLAS software framework

ATLAS software framework called ATHENA [28] (figure 1.22), is an enhanced version of the original C++ based software framework GAUDI [29], initially developed by the LHCb

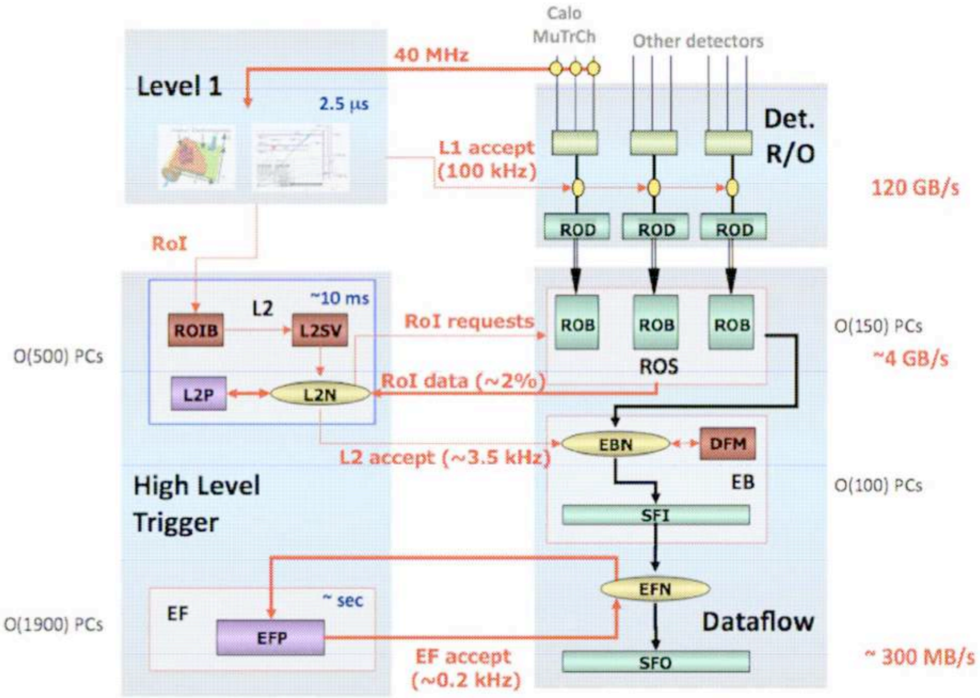


Figure 1.20: *ATLAS trigger block diagram.*

collaboration. ATHENA is a common software design of large scale projects where numerous types of internal and external software components need to be encompassed in a single application. The component library structure allows that modules are loaded as shared libraries at job configuration level.

Three main basic building blocks can be named, which form the pillars of the Athena architecture:

- The “Service” class is designed to provide dedicated functionality throughout the execution of the program. One of the important realization of a “Service” is the transient data store, “StoreGateSvc” (or simply StoreGate). The instance of “Service” classes is handled by a central “ExtSvc manager” that regulates initialization and finalization and the facility is uniformly provided to all ATHENA components.

- The “Algorithm” class represents the primary algorithmic part of an ATHENA application. It is dedicated to actions, which are taken exactly one time at every event and classes derived from the “Algorithm” class need to be registered to the central “ApplicationMgr” that steers initialization, finalization and the execution of the “Algorithm” at every event.

- The “AlgTool” class provides more flexible solution for smaller pieces of algorithms that are typically invoked multiple times within different contexts. “AlgTool” instances are called through an “Algorithm” that either owns “AlgTool” (called private in this case) instances or retrieves them through the central “ToolSvc” where all public tools are registered. This pattern allows “AlgTool” classes to be instantiated multiple times

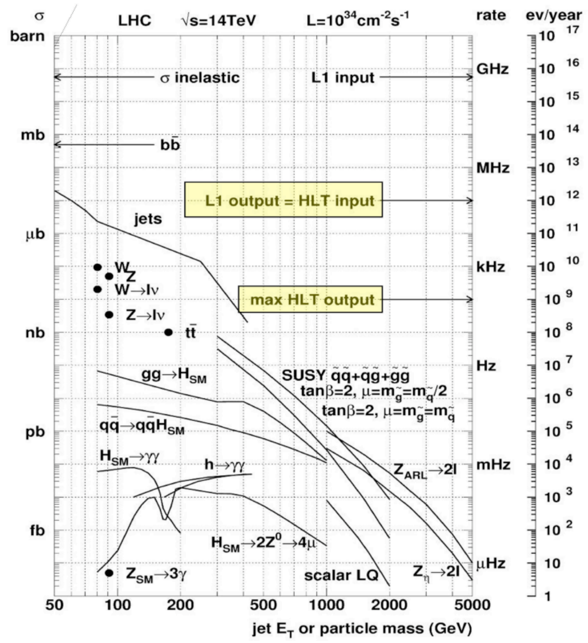


Figure 1.21: *Expected event rates for several physics processes at the LHC design luminosity [19]. Process* $q\bar{q} \rightarrow q\bar{q}H_{SM}$ *correspond to Vector Boson Fusion Higgs production discussed later in this document.*

with different configurations or once with same configuration but used multiple times from different "Algorithms".

"Algorithm" and "AlgTool" are usually written in C++ since it is advantageous in terms of computing efficiency for it produces compiled binary libraries. On the other hand, robust configuration capabilities are provided in Athena by Python scripting language [31]. So-called "Python bindings" enable configuration of C++ "Algorithm" and "AlgTool" from the Python interpreter. Being an interpreted language, Python is equipped with a dynamic scripting environment, which favours rapid development and interactivity. In addition, it is a language with support for high-level dynamic data types and a design concept such as object-orientation.

Generally, the "Algorithm" is responsible for retrieving input data collections from and writing the output data to the transient event store, "StoreGateSvc". On the other hand, modularisation of analysis can be achieved by taking advantage of light-weight "AlgTool" classes, which are self contained collection of small algorithms that can be dynamically chained together through "Algorithm" using run-time configuration.

1.3 ATLAS performance

As mentioned in the introduction, the physics analysis part of this document describes a Higgs boson search produced via the vector boson fusion (VBF) decaying in the di-tau channel. The special topology of VBF combined with the presence of τ's implies an excellent performance of ATLAS detector concerning the identification and reconstruction of different

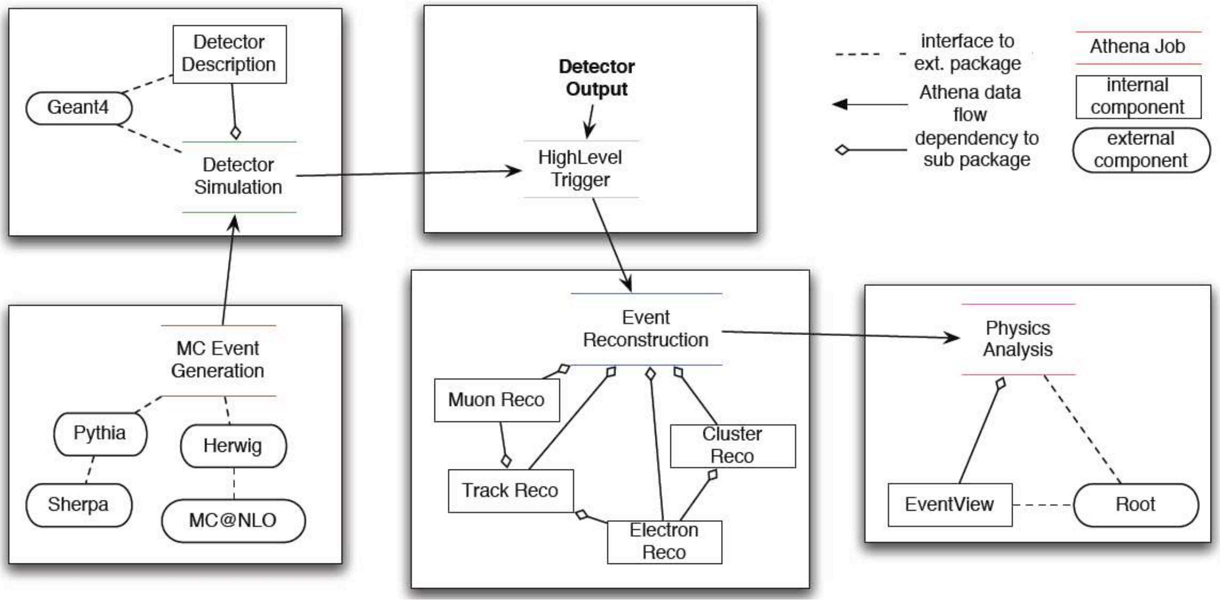


Figure 1.22: *Various computing tasks in Athena components model. [30]*

particle objects and a high efficiency in tagging technics especially related to jets. Higgs boson production via VBF as shown in chapter 4, requires a very good jet reconstruction: *(i)* in the forward regions of the calorimeter (high values of $|\eta|$) to probe the two tagging VBF jets, and *(ii)* in the central part of the calorimeter where a jet veto cut is applied. The importance of a robust jet finding algorithm is therefore illustrated.

Moreover, τ leptons as Higgs boson decay products imply electrons, muons, tau-jets and neutrinos in the final state of the event. High efficiencies in electron and muon identification are necessary for obtaining a high number of signal statistics considering the low cross section of the process. Neutrinos are reflected in missing energy in the transverse plane ($\not\!\!E_T$), so an accurate $\not\!\!E_T$ reconstruction will be a crucial part of VBF $H \rightarrow \tau^+\tau^-$analysis giving the possibility to include neutrinos in Higgs boson mass reconstruction. Finally, the identification of jets either as tau-jets or b-jets will also play an important role tagging hadronic decays of τ lepton increasing signal statistics and furthermore, rejecting background which could fake signal events.

ATLAS was designed taking into consideration all physics requirements including those mentioned above, providing a full and complete reconstruction and identification of all measured quantities in such a high-energy experiment. In the following sub-sections the reconstruction of electrons, muons and taus is discussed. The reconstruction of jets, the measurement of missing transverse energy and the b-tagging are also addressed. All these quantities are essential ingredients not only for VBF $H \rightarrow \tau^+\tau^-$study but for almost all physics processes studied in ATLAS. The reconstruction and identification algorithms described in the following are based on release 12 of ATLAS software.

1.3.1 Electrons reconstruction and identification

ATLAS electromagnetic (EM) calorimeter is designed in such a way that it is able to identify efficiently electrons and photons within a large energy range (5 GeV to 5 TeV), and to measure their energies with a linearity better than 0.5%. In this section, the focus will be given to the electron reconstruction and identification but similar methods are applied for the respective procedures related to photons.

The procedure to measure the energy of an incident electron in the ATLAS EM calorimeter has been described in details in ref. [15]. One of the key ingredients for the description of the detector performance and especially of EM calorimeter which mainly participates in electron studies, is the amount and position of the upstream material and its geometry. The amount of material in front of the calorimeter (figure 1.23 for the as-built detector is significantly larger than was initially estimated, this leads to larger energy losses for electrons and a large fraction of converted photons.

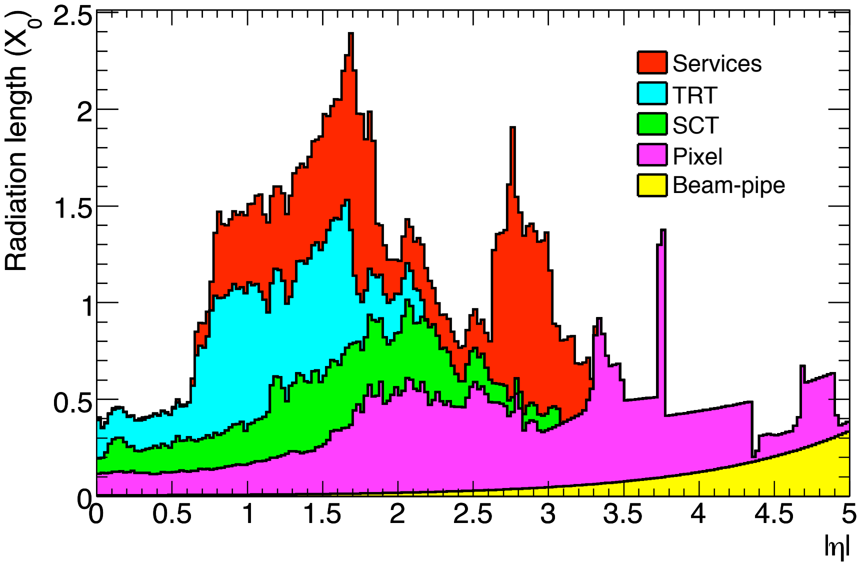


Figure 1.23: *Material distribution (X_0) at the exit of the ID envelope, including the services and thermal enclosures. The distribution is shown as a function of $|\eta|$ and averaged over ϕ [16].*

A *sliding window* algorithm [32] is used to identify and reconstruct electromagnetic clusters. Rectangular clusters are formed with a fixed size, in such a way that their position corresponds to the maximum amount of energy deposited inside them (a minimum energy of 3 *GeV* is required). The optimal cluster size depends on the particle type being reconstructed and the calorimeter region. Electrons need therefore larger clusters than photons due to their larger interaction probability in the upstream material and also due to the fact that they bend in the magnetic field, radiating soft photons along a range in ϕ. Several series of these kinds of clusters are then built by the reconstruction software, corresponding to different sliding window sizes. These clusters are the starting point of the calibration and selection of electron candidates.

For each of the reconstructed clusters, the reconstruction tries to find a matching track within a $\Delta\eta \times \Delta\phi$ range of 0.05×0.10 with momentum p compatible with the cluster energy E ($E/p < 10$). If one is found, the reconstruction checks for presence of an associated conversion. An electron candidate is created if a matched track is found while no conversion is flagged. This early classification allows to apply different corrections to electron candidates and it is the starting point of a more refined identification based on shower shapes and on respective cuts. Three levels of electron quality are defined (loose, medium, tight):

- **Loose cuts**: This set of cuts performs a simple electron identification based only on limited information from the calorimeters. Cuts are applied on the hadronic leakage and on shower-shape variables, derived from only the middle layer of the EM calorimeter. This set of cuts provides excellent identification efficiency, but poor background rejection.

- **Medium cuts**: This set of cuts improves the background rejection quality, by adding cuts on the strips in the first layer of the EM calorimeter and on the tracking variables. Strip-based cuts are adequate for $e - \pi^0$ separation. The tracking variables include the number of hits in the pixels, the number of silicon hits (pixels plus SCT) and the tranverse impact parameter. The medium cuts increase the jet rejection by a factor of 3-4 with respect to the loose cuts, while reducing the identification efficiency by $\sim$10%.

- **Tight cuts**: This set of cuts makes use of all the particle-identification tools currently available for electrons. In addition to the cuts used in the *medium cuts* set, cuts are applied:

    - on the number of vertexing layer hits (to reject electrons from conversions)
    - on the number of hits in the TRT
    - on the ratio of high-threshold hits to the number of hits in the TRT (to reject the dominant background from charged hadrons)
    - on the difference between the cluster and the extrapolated track positions in η and ϕ

    Two different final selections are available within this tight category: they are named tight (isol) and tight (TRT) and are optimised differently for *isolated* and *non-isolated* electrons. In the case of tight (isol) cuts, an additional energy isolation cut is applied to the cluster, using all cell energies within a cone of $\Delta R < 0.2$ around the electron candidate. This set of cuts provides, in general, the highest isolated electron identification and the highest rejection against jets. The tight (TRT) cuts do not include the additional explicit energy isolation cut, but instead apply tighter cuts on the TRT information to further remove the background from charged hadrons.

Figure 1.24 compares the distributions expected from $Z \to e^+e^-$ decays and from the QCD di-jet sample for one of the main basic discriminating variables described above, which is the $\Delta\eta$ between cluster and extrapolated track position.

The electron efficiency as a function of $|\eta|$ and E_T for loose, medium, and tight electron cuts is shown in figures 1.25(a)-(b). Here, only electrons with $|\eta| < 2.5$ and $E_T > 5$ GeV

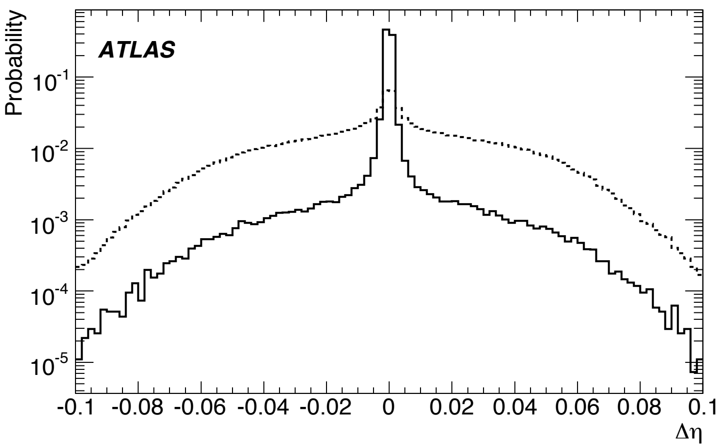


Figure 1.24: *Difference in η ($\Delta\eta$) between cluster and extrapolated track positions for electrons from $Z \rightarrow e^+e^-$ decays (solid line) and for QCD di-jets sample (dotted line) [19].*

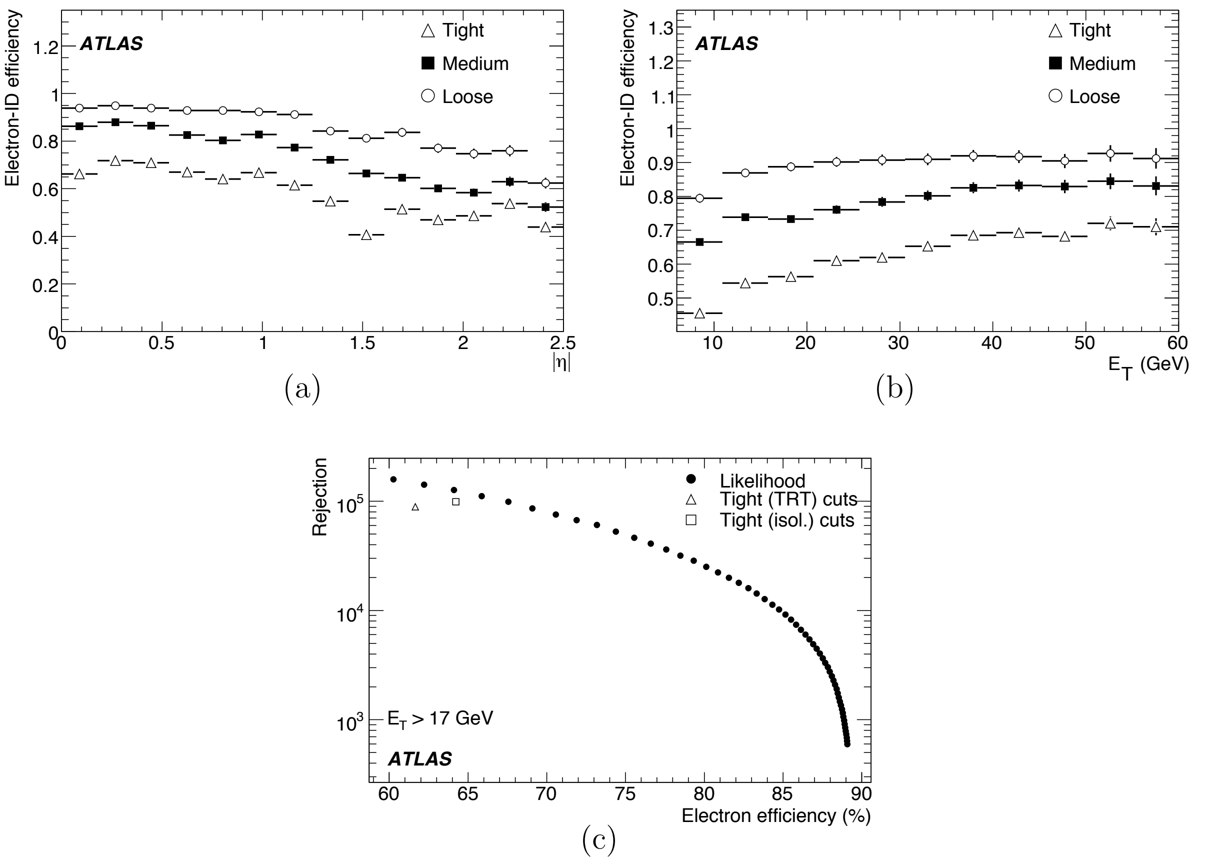
are considered. The drop in efficiency at low E_T is mainly due to the loss of discrimination power of the cuts at lower transverse energies. The loss of the efficiency is in $|\eta|$ ~1.5 is due to the transition region between the barrel and end-cap calorimeters.

In addition to the standard cut-based electron identification described above, several multivariate techniques have been developed and implemented in the ATLAS software. These include a likelihood discriminant, a discriminant called H-matrix, a boosted decision tree, and a neural network[19]. Using a likelihood discriminant method for instance, significant gain is observed in the efficiency and rejection with respect to the cut-based methods. Figure 1.25(c) shows the rejection versus efficiency curve obtained using the likelihood discriminant method, compared to the results obtained for two sets of tight cuts. The multivariate method provides an improvement in rejection of about 20-40 % with respect to the cut-based method for the same efficiency of 61-64 %. Furthermore, it provides a gain in efficiency of 5-10% (tight cuts) for the same rejection. Multivariate methods providing in general a more refined and accurate identification will only be used once the detector performance is understood. For the early data, the robust cut-based methods will be used.

1.3.2 Muons reconstruction and identification

ATLAS employs a variety of strategies for identifying and reconstructing muons. The direct approach is to reconstruct standalone muons by finding tracks in the muon spectrometer and then extrapolating these to the beam line. Combined muons are found by matching standalone muons to nearby inner detector tracks and then combining the measurements from the two systems. Tagged muons are found by extrapolating inner detector tracks to the spectrometer detectors and searching for nearby hits. The current ATLAS baseline reconstruction includes two algorithms for each strategy. The collections (and families) of reconstructed muons are referred by the names of the corresponding combined algorithms: Staco [33] and Muid [34]. The Staco collection is the current default for physics analysis.

As already mentioned, ATLAS muon spectrometer is designed in such a way to provide a resolution (figure 1.26) of 10% for a muon of 1 TeV. In performance terms, given the



(a) (b)

(c)

Figure 1.25: *Electron identification efficiency as a function of* $|\eta|$ *(a) and* E_T *(b) for electrons with* E_T *>5 GeV from* $H \rightarrow$ *eeee decays. Jet rejection versus isolated electron efficiency obtained with a likelihood method compared to the results from the two sets of tight cuts. Only electrons with* E_T *>17 GeV are selected in this case (c). [19].*

magnetic field produced by the toroids ($\sim$ 0.5 T) and the size of muon chambers, a curve distance of $\Delta \sim 500$ μm is expected and therefore a 10% resolution results to a δ=50 μm spatial resolution.

1.3.2.1 Standalone muon candidates

The standalone algorithms first build track segments in each of the three muon stations and then link the segments in order to form tracks. This extrapolation also takes into consideration the effect of muons multiple scattering and muons energy loss in the calorimeter. Standalone algorithms have the advantage of slightly greater $|\eta|$ coverage out to 2.7 compared to 2.5 for the inner detector (see later), but there are holes in the coverage at $|\eta|$ (see figure 1.27) near $\sim$0.0 (cabling area) and $\sim$1.2 (transition between barrel end end-cap parts of muon spectrometer). Very low momentum muons (around a few GeV/c) may be difficult to reconstruct because they do not reach the external stations. Muons produced in the calorimeter, e.g. from π and K decays, are likely to be found in the standalone reconstruction and form a background of *fake* muons for most physics analyses.

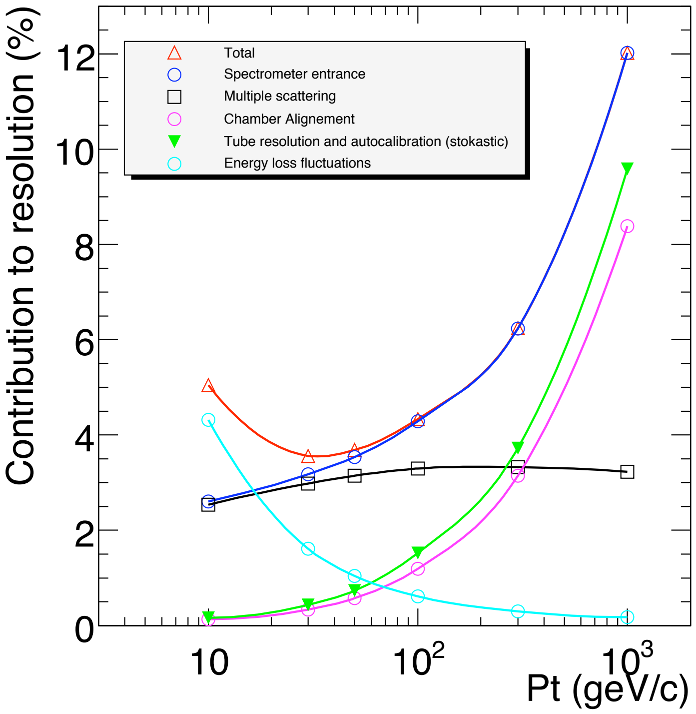


Figure 1.26: *Contributions to the momentum resolution for muons reconstructed in the Muon Spectrometer as a function of transverse momentum for* $|\eta| < 1.5$ *[19].*

1.3.2.2 Inner detector as a starting point

The primary track reconstruction algorithm for the inner detector is described in Ref. [35]. Space points are identified in the pixel and microstrip detectors, these points are linked to form track seeds in the inner four layers, and tracks are found by extending these seeds to add measurements from the outer layers. This strategy is expected to give high detection efficiency over the full detector acceptance, $|\eta| < 2.5$.

1.3.2.3 Combined muon candidates

Both of the muon combination algorithms, Staco and Muid, pair muon-spectrometer tracks with inner detector tracks to identify combined muons. The match chi-square (χ^2_{match}), defined as the difference between outer and inner track vectors weighted by their combined covariance matrix:

$$\chi^2_{match} = (\vec{T}_{MS} - \vec{T}_{ID})^T(\mathbf{C_{ID}} + \mathbf{C_{MS}})^{-1}(\vec{T}_{MS} - \vec{T}_{ID}) \tag{1.19}$$

provides an important measure of the quality of this match and is used to decide which pairs are selected. Here $\vec{T}$ denotes a vector of (five) track parameters, expressed at the point of closest approach to the beam line, and $\mathbf{C}$ is its covariance matrix. The subscript ID refers to the inner detector and MS to the muon spectrometer (after extrapolation accounting for energy loss and multiple scattering in the calorimeter). Figure 1.28(a)-(b) shows the

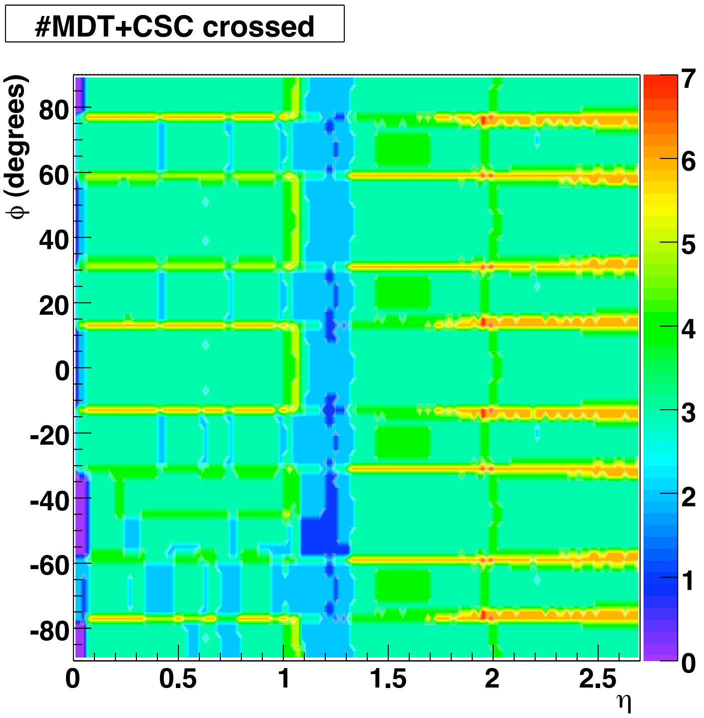


Figure 1.27: *Number of detector stations traversed by muons passing through the muon spectrometers a function of* $|\eta|$ *and* φ *[16].*

combined $t\bar{t}$ direct muon efficiency and fake rates as a function of η for low luminosity for Staco and Muid. When matching inner detector and muon spectrometer tracks, both muon identification algorithms calculate a χ^2 factor (equation 1.19) which serves as a discriminant for separating real and fake muons. The fakes include pion or kaon decays in or near the calorimeter. These fake rates as a function of muon efficiency for different muon p_T is shown in figure 1.28(c)-(d) for Staco and Muid. The higher Staco fake rates for the high values of efficiency come from looser cuts during reconstruction.

1.3.2.4 Tagged muon candidates

The spectrometer tagging algorithms extrapolate all inner detector tracks with sufficient momentum out to the first station of the muon spectrometer and search for nearby segments. These algorithms simply use the inner detector track to evaluate the muon kinematics, so the inner track and spectrometer hits are not combined to form a new track. This is not very important in the low-p_T regime that these algorithms were originally intended to address. They were being further developed to allow extrapolation, to other and multiple stations providing the possibility to include the spectrometer measurements in a track refit.

In general, these tagging algorithms were developed to identify low-p_T muons, which have such a low p_T which is not sufficient to make them reach the muon spectrometer. Furthermore, these algorithms are also adequate for muons that interact a lot through Brehmstralung in calorimeters causing electromagnetic showers (muon energy loss) and lose part of their energy. In this case it is rather difficult starting an identification from muon spectrometer since numerous hits are produced from the EM shower.

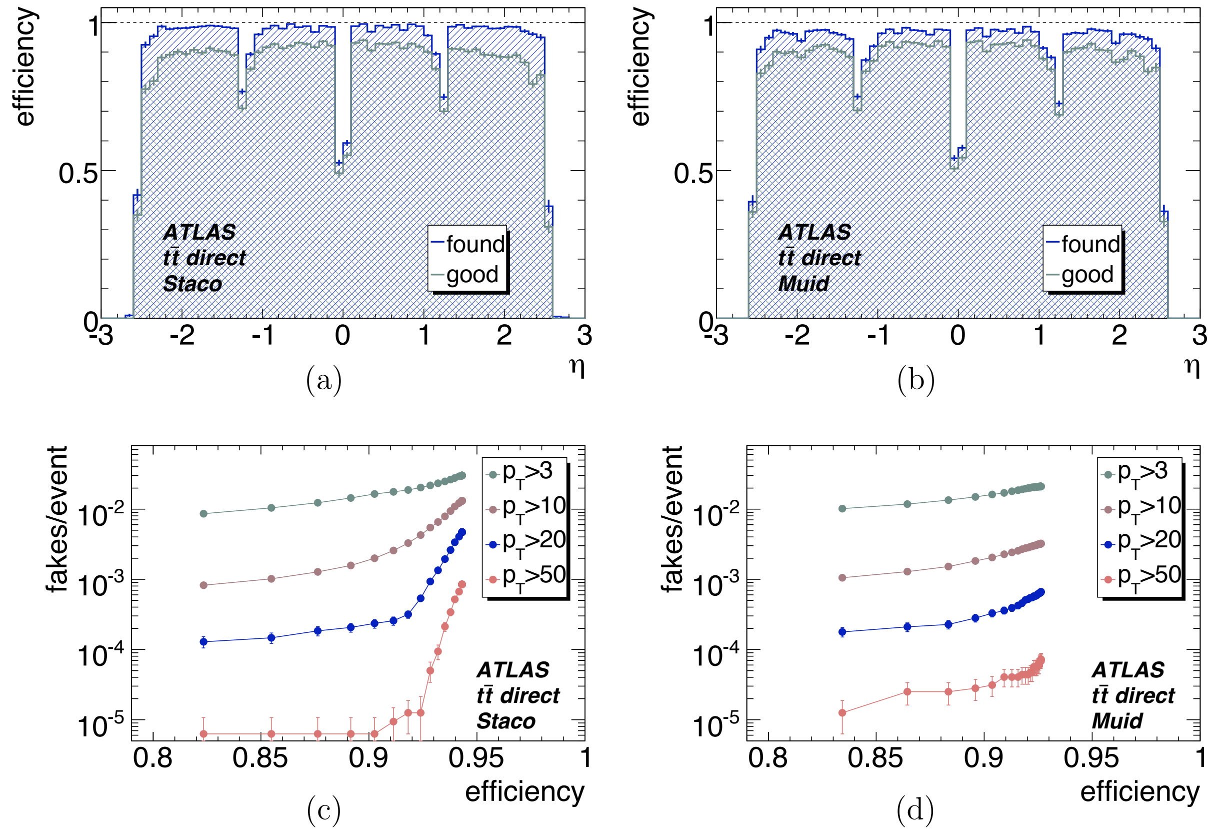


Figure 1.28: *Combined muon efficiency and fake rate for Staco (a) and Muid (b) as functions of true* η *for direct muons in* $t\bar{t}$ *at low luminosity. The upper curve (blue) is the efficiency to find the muon while the lower curve (green) additionally requires a good match between reconstructed and true track parameters. Fake rates as a function of efficiency as the muon* p_T *threshold is varied. Results are shown for Staco (c) and Muid (d) [19].*

1.3.2.5 Merging muon candidates

The muon efficiency may be increased by including muons found by multiple algorithms but special attention must be drawn to remove overlaps, i.e. cases where the same muon is identified by two or more algorithms. This is done when muon collections are created. Standalone muons that are successfully combined are not recorded separately. In those cases where a standalone muon is combined with more than one inner detector track, exactly one of the muons is flagged as *best match*. In the Staco collection, the tagged and combined muons do not overlap by construction, while in Muid collection, overlaps are removed by creating a single muon when both have the same inner detector track.

1.3.3 The lepton τ identification and reconstruction

Tau leptons are expected to play an important role at the LHC. They will provide an excellent probe in searches for new phenomena such as the Standard Model Higgs boson at low masses, the MSSM Higgs boson or Supersymmetry (SUSY). In particular for this document, the decay of Higgs boson in two taus ($H \rightarrow \tau^+\tau^-$) is studied (chapter 4, therefore, understanding taus reconstruction, their selection methods and thus efficiencies is an essential part of the analysis.

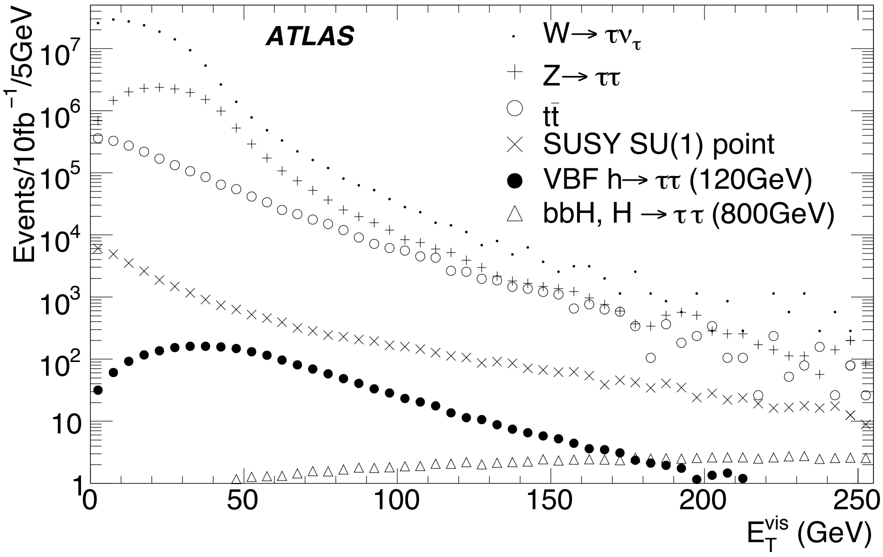


Figure 1.29: *The visible transverse energy of τ leptons from different physics processes: top quark decays, W/Z production, Standard Model vector boson fusion Higgs boson production for $m_H = 120GeV$ with $H \rightarrow \tau\tau$, for τ leptons from low energy Supersymmetry with a light stau (SU1 sample), heavy Z' bosons, and heavy Higgs bosons from bbH production in the MSSM with $tan\beta = 2$ for a mass of 400 GeV [19].*

Tau leptons are massive particles with a measurable lifetime ($\sim 10^{-14}s$ or ct = 87.11 μm) undergoing electroweak interactions only. The production and the decay of τ leptons are well separated in time and space ($\Gamma_\tau/m_\tau \sim 10^{-11}$), providing potential for unbiased measurements of the polarisation, spin correlations, and the parity of the resonances decaying into τ leptons. The interesting transverse momentum range of τ leptons ranges from below 10 GeV up to at least 500 GeV. Experiments at the LHC will thus have to identify them in a wide momentum range. The low energy range should be optimized for analyses related to W and Z boson decays to τ and also to Higgs boson searches and SUSY cascade decays. The higher energy range is mostly of interest in searches for heavy Higgs bosons in MSSM models and for extra heavy W's and Z's gauge bosons. Figure 1.29 shows the transverse energy spectrum of the visible decay products of τ leptons from different processes of interest normalized to the predicted cross-section with which they will be produced at the LHC and to an integrated luminosity of 10 fb^{-1}.

τ leptons decay hadronically in 64.8% of all cases, while in 17.85% and 17.36% of the cases they decay to an electron and a muon respectively. From the detection point of view, hadronic modes are divided by the number of charged πs among the decay products into single-prong (one charged π) and three-prong (three charged πs) decays (figure 1.30). In general, one- and three-prong modes are dominated by final states consisting of $\pi^\pm$ and π^0. Therefore, the reconstruction of τ leptons is in general understood as a reconstruction of the

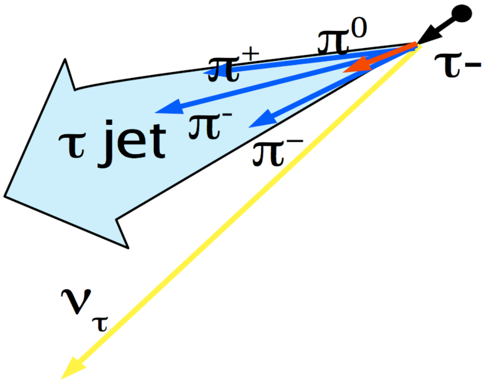


Figure 1.30: *A schema of a hadronic* τ *decay jet* $(\tau^- \rightarrow \pi^0\pi^-\pi^+\pi^-\upsilon_\tau)$.

hadronic decay modes, since it would be difficult to distinguish leptonic modes from primary electrons and muons. The availability of various decay modes makes τ leptons a rich but not totally unique signature. Hadronically decaying τ leptons are distinguished from QCD jets on the basis of low track multiplicities contained in a narrow cone, characteristics of the track system and the shapes of the calorimetric showers. Isolation from the rest of the event is required both in the inner detector and the calorimeter. From this information, a set of identification variables is built, to which either a traditional cut-based selection or multivariate discrimination techniques are applied. Table 1.4 summarizes the principal τ decay modes which are fundamental characteristics in the τ reconstruction as described below.

Decay modes	Branching ratio
leptonic (e and μ)	$\sim 35.2\%$
1-prong ($\tau \rightarrow\ n\pi^0\ \pi^{\pm}\upsilon$)	$\sim 49.5\%$
3-prong ($\tau \rightarrow\ n\pi^0\ \pi^{\pm}\pi^{\mp}\pi^{\pm}\upsilon$)	$\sim 14.6\%$

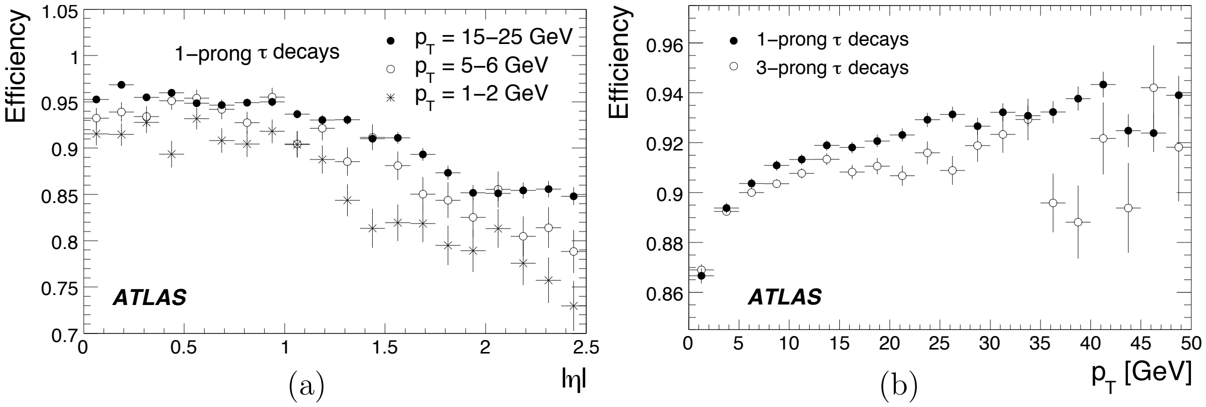
Table 1.4: *Tau lepton* (τ) *principal decay modes, where* n *is a multiplier factor such as* $n = 0, 1, 2, ...$

The inner detector provides information on the charged hadronic track or the collimated multi-track system reconstructed in isolation from the rest of the event. These tracks should neither match track segments in the muon spectrometer nor reveal features characteristic of an electron track (e.g. high threshold hits in the Transition Radiation Tracker). In the case of a multi-track system, they should be well collimated in (η, ϕ) space and the invariant mass of the system should be below the τ lepton mass. The charge of the decaying τ lepton can be directly determined from the charge(s) of its decay product(s).

Calorimetry provides information on the energy deposit from the visible decay products (i.e. all decay products excluding neutrinos). Hadronically decaying τ leptons are well collimated leading to a relatively narrow shower in the electromagnetic (EM) calorimeter with, for single-prong decays with one or few π^0s, a significant pure electromagnetic component. On average, in this case, about 55% of the energy is carried by π^0s present among the decay products.

The calorimeter and tracking information should match, with narrow calorimeter cluster being found close to the track(s) impact point in the calorimeter. Furthermore, the invariant mass of the cluster should be small and the cluster should be isolated from the rest of the event. The algorithms for the reconstruction of hadronically decaying τ leptons are considered as combined reconstruction algorithms, as they use components provided by algorithms specific to different subdetectors like track reconstruction in the inner detector or topological clustering of the energy deposits in the calorimeter. At present, two complementary algorithms have been implemented into the ATLAS offline reconstruction software.

- The calorimetry-based algorithm starts from clusters reconstructed in the hadronic and electromagnetic calorimeters and builds the identification variables based on information from the tracker and the calorimeter.
- The track-based algorithm starts from seeds built from few (low multiplicity) high quality tracks collimated around the leading one. The energy is calculated with an energy-flow algorithm based only on tracks and the energy in the electromagnetic calorimeter. All identification variables are built using information from the tracker and the calorimeter.



(a) (b)

Figure 1.31: *Reconstruction efficiency for tracks from charged πs for one- and three-prong hadronic τ decays from $W \to \tau\nu$ and $Z \to \tau\tau$ signal samples as a function of the pseudorapidity $|\eta|$ for three different ranges of track p_T (a) and of the transverse momentum of the track (b) [19].*

Summarizing, a hadronic decay of a lepton τ has the following characteristics: collimated calorimeter cluster, low charged tracks multiplicity, displaced secondary vertex and single-prong or three-prong decays.

The efficiency for track reconstruction in τ decays is defined as the probability for a given charged π from a τ decay to be reconstructed as a track. With respect to the reference tracking performance of the detector established for single muons in the low p_T range a drop due to hadronic interactions (a charged π interacting with the material of the inner detector) is expected. In the higher p_T range a degradation is caused by the strong collimation of the multi-track system for three-prong decays (figure 1.31(b)).

Good quality tracks reconstructed with p_T as low as 1 GeV are required by the track-based algorithm, while the calorimeter-based algorithm accepts any track with p_T >2 GeV.

To complete the τ reconstruction some quality criteria are then added: good quality tracks are required to satisfy a specific χ^2 threshold, to have a number of pixel and SCT hits >7 and transverse impact parameters d_0 <1 mm. For the leading track in addition the number of low threshold TRT hits has to be larger than 10 in a pseudorapidity η range up to 1.9, while for the second or third track the presence of a b-layer hit and ratio of the of high-to-low threshold hits of smaller than 0.2 are required. Both requirements were added to minimize the number of accepted tracks from conversions. A dedicated veto against electron tracks being used as leading tracks is not applied at the reconstruction level. This will be taken into consideration separately as part of the event selection analsysis, as it will be the case in VBF $H \rightarrow \tau^+\tau^-$analysis (chapter 4).

Figure 1.31 shows the reconstruction efficiency for p_T =1-50 GeV using the standard quality selection. Adding the additional quality criteria as described above, the overall efficiency for reconstructing good quality tracks from τ lepton hadronic decays is reduced to 82-83%. The reconstruction efficiency is slightly higher for tracks from single prong decays compared to three-prong decays, where tracks could be very collimated particularly for boosted τ leptons, such as τ leptons in $H \rightarrow \tau^+\tau^{-1}$.

1.3.4 Missing transverse energy ($\not{E}_T$)

Since neutrinos do not interact in the detector, they deposit no energy. However, it is possible to calculate the sum of all neutrino energies in an event in the plane transverse to the beam axis by using momentum conservation. The colliding protons do not have transverse momentum components and therefore the sum of the transverse momenta of all final state particles has to vanish as well. Thus, the sum of the transverse momenta of all visible final state particles equals the negative sum of the transverse momenta of all neutrinos:

$$E_T^{visible} + E_T^{invisible} = 0 \Longrightarrow E_T^{miss} = -E_T^{visible} \tag{1.20}$$

A very good measurement of the missing transverse energy $\not{E}_T$ is essential for many physics studies in ATLAS. Events with large $\not{E}_T$ are expected to be the key signature for new physics such as supersymmetry and extra dimensions. A good $\not{E}_T$ measurement in terms of linearity and resolution is also important for the reconstruction of the top-quark mass from $t\bar{t}$ events with one top quark decaying semileptonically. Furthermore, it is crucial for the efficient and accurate reconstruction of the Higgs boson mass when the Higgs boson decays to a pair of τ-leptons (see chapter 4).

The transverse missing energy $\not{E}_T$ [19] in ATLAS is mainly reconstructed from energy deposits in the calorimeter and reconstructed muon tracks. Other $\not{E}_T$ contributions have their origins in the underlying events, pile-up and electronics noise. Additional energy corrections due to dead matter are taken into consideration. There are two kinds of algorithms for $\not{E}_T$ reconstruction in ATLAS that use different aspects of energy deposit classification and calibration.The Cell-based algorithm origins in the energy deposits in calorimeter cells that survive a noise suppression procedure. The cells can be calibrated using global calibration weights depending on their energy density. This procedure is robust and it will be used at

---

[1] $m_H/2 >> m_\tau$

initial data taking for commissioning reasons. Further information regarding the Cell-based $\not\!\!E_T$ can be found in chapter 2

For the further LHC collision data analysis, the $\not\!\!E_T$ object-based algorithm will deal with the reconstructed, calibrated and classified objects in the event such as e/γ, muons, τs, jets, etc. The energy remnants outside these objects that are still deposited in the calorimeter are classified as low p_T deposit from charged and neutral pions and being calibrated accordingly, they are taken into the $\not\!\!E_T$ consideration. This is a more refined method adequate for physics analyses that deal with $\not\!\!E_T$ quantities.

The noise in calorimeters presented in the previous section is a very important factor of $\not\!\!E_T$ calculation (if all cells are naively summed it would contribute by 13GeV to the width of $\not\!\!E_T$ distribution) especially for events with low energy deposit which can be comparable with the noise deposit and therefore be suppressed within the two algorithms. This noise suppression is common for the Cell- and Object-based [19] algorithm and is described in the two following subsections.

1.3.4.1 Standard Noise Suppression Method ($\not\!\!E_T$-*base*)

This method is based on using calorimeter cells fulfilling the following criteria:

$$|E_{cell}| > n \cdot \sigma_{noise} \tag{1.21}$$

where σ_{noise} is the cell-by-cell noise presented in section 1.2.2.4. The threshold n is being optimized in studies including $\not\!\!E_T$ resolution, the scale of $\not\!\!E_T$ and energy deposits in calorimeter and a value $n = 2$ is chosen. In the assumption of a noise which can be modeled through a Gaussian Probability Density Function, a selection of $n = 2$ as an energy threshold to the previous equation implies a selection of only 4.6% of cells (cf figure 1.32), in empty events.

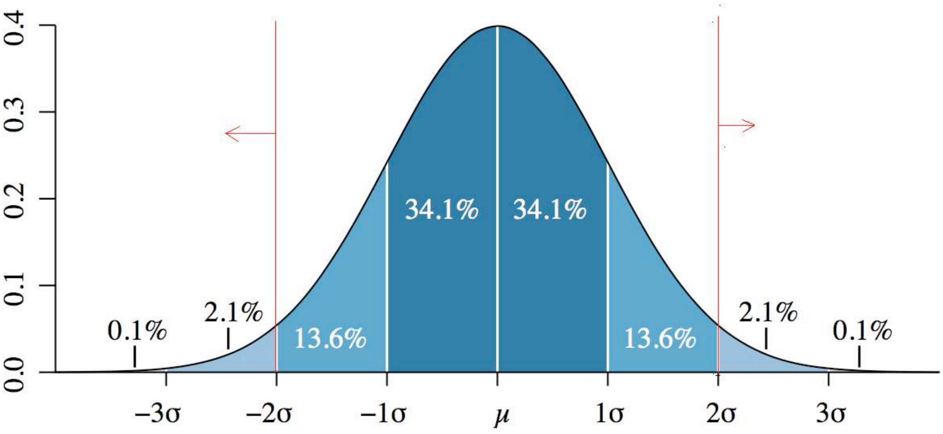


Figure 1.32: *A Gaussian PDF. The red lines indicate the limits imposed by equation 1.21 for the used threshold of* $n = 2$.

1.3.4.2 Noise Suppression using TopoClusters ($\not\!\!E_T$-*topo*)

Typical energy deposit in the calorimeter stems from electromagnetic or hadronic showers that spreads over several adjacent cells. Hence a topological 3 dimensional clustering of calorimeter cells was devised as follows. An iteration procedure is being applied searching cell-seeds for the so-called topo-cluster that fulfill the criteria:

$$|E_{cell}| > n_{seed} \cdot \sigma_{noise} \tag{1.22}$$

After having found the cell-seed a new iteration procedure searches all the neighbor-cells around the seed and adds to the cluster those which fulfill:

$$|E_{cell}| > n_{neigh} \cdot \sigma_{noise} \tag{1.23}$$

Finally, after having completed the layer of neighbor-cells a final layer of cells is added to the cluster adding all the surrounding cells ($|E_{cell}| > 0 \cdot \sigma_{noise}$). After Monte-Carlo and test-beam studies it was found that the optimal configuration for the n_{seed} and n_{neigh} is 4 and 2 respectively (referred to as 4/2/0 configuration). This configuration was used in the following study and it is the default used by ATLAS collaboration. Figure 1.33 illustrates an example of a topo-cluster. When doing noise studies, and in the assumption of a gaussian modeled noise (figure 1.32), the number of selected cells is greatly reduced in the $\not{E}_T$-*base* method since there are only 0.4% of the cells that can be seed cells for the default configuration of 4/2/0.

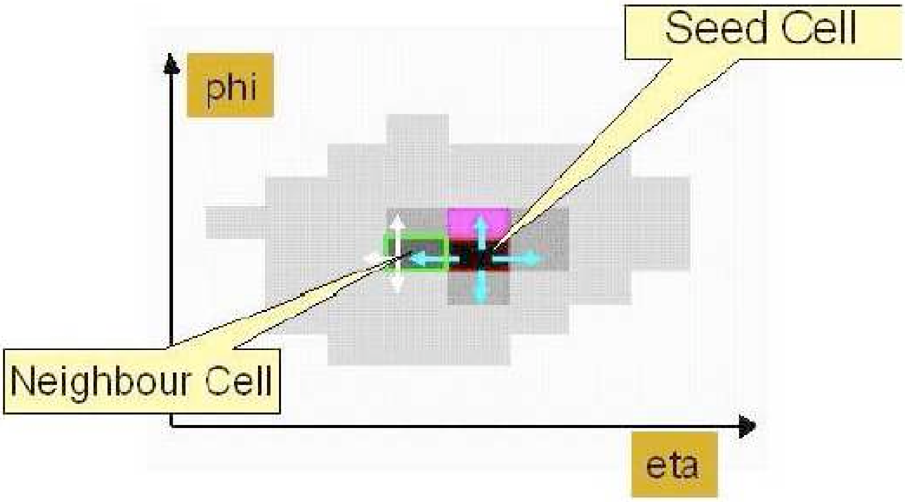


Figure 1.33: *A 3D topocluster is formed around a seed cell above a given seed-threshold (4*σ_{noise}) and neighbouring cells are added to it if their energy is above a given neighbour-threshold (2*σ_{noise}).*

As is explicit in the formulae, the cells are selected based on the absolute value of their energy, in order to avoid an overall bias (there would be a positive bias if only upwards fluctuations were chosen). This implies in particular that negative energy clusters are possible.

From test beam studies,it is found that $\not{E}_T$-*topo* shows a better global $\not{E}_T$ resolution compared to the $\not{E}_T$-*base* . Results on noise $\not{E}_T$ performance obtained using both suppression methods is the main purpose of the following study.

1.3.4.3 $\not{E}_T$ resolution

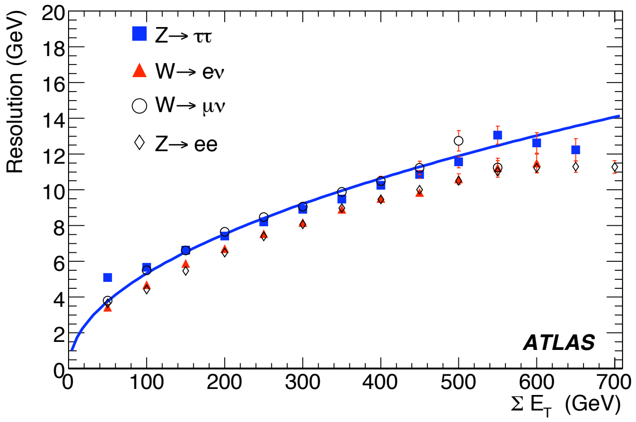


Figure 1.34: *Resolution of the two $\not{E}_T$ components with refined calibration as a function of the total transverse energy, $\sum E_T$ [19].*

Figure 1.34 shows the expected resolution of the missing transverse energy measurement from the detailed detector simulation as a function of the absolute value of the missing transverse energy. The curve correspond to the best fit of σ= $0.53\sqrt{E_T}$ through the points from Z→ $\tau\tau$ events. The reason why $\not{E}_T$ resolution follows this curve is that it is completely dependent on the resolution of ATLAS calorimeter. As shown in previous section, the most dominant component of $\not{E}_T$ contribution is the calorimetric one, and resolution in calorimeter is defined as $\sigma_E = \alpha\%\sqrt{E}$ (section 1.2.2), which finally explains the $y = c\sqrt{x}$ behavior of $\not{E}_T$ resolution.

For an $\not{E}_T$ of ∼50 GeV, which is a typical value in the case of $H \to \tau^+\tau^-$, the resolution is in the order of 10%.

1.3.5 Jet reconstruction

High quality and highly efficient jet reconstruction is an important tool for almost all physics analyses to be performed with the ATLAS experiment, similarly to $\not{E}_T$ reconstruction discussed in previous section. Typically, an absolute systematic uncertainty better than 1% is desirable for precision physics like the measurement of the top quark mass, and the reconstruction of some SUSY final states and it is the achieved performance of ATLAS jets algorithms.

The principal detector for jet reconstruction is the calorimeter system (section 1.2.2), with its basic components illustrated in figure 1.6. All the different features of calorimeter described in previous sections allow a high quality jet reconstruction in the challenging environment of the proton-proton (pp) collisions at the LHC.

1.3.5.1 Jet Algorithms in ATLAS

One important characteristic of the jet defined within the jet algorithm used, is its size represented by a cone in ΔR:

Cone of a jet

R_{cone} is defined by the following formula :

$$R_{cone} = \sqrt{\Delta\eta^2 + \Delta\phi^2} \tag{1.24}$$

where R_{cone} is an indicative quantity for the size of the formed jet. It actually refers to the resolution in (η, ϕ) that a jet algorithm has in order to reconstruct a jet. Two jets that have a distance lower than the R_{cone} will finally reconstructed as one jet. The typical default values used in ATLAS are 0.4 and 0.7. Figure 1.35 illustrates an example of a different R_{cone} value in the jets algorithms. The two arrows corrspond to the (η, ϕ) directions of two jets and the cones correspond to the reconstructed jets. At the left, a jet algorithm with a smaller R_{cone} is used so the two jets are well separate, while is not the case for the image at the right where R_{cone} has a higher value.

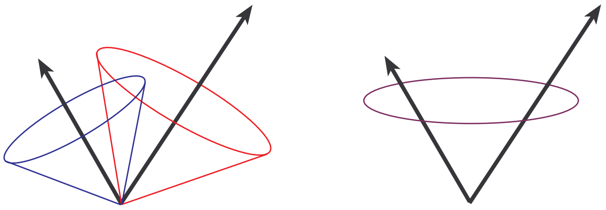


Figure 1.35: *Jet reconstruction algorithms using different R_{cone} values.(left) A jet algorothm with a smaller R_{cone} can discriminate between 2 jets that are found close in ΔR while (right) only one jet is reconstructed in the case of a higher R_{cone}.*

Several jet collections are built during event reconstruction in ATLAS, varying the input to the jet finder and the jet finding algorithm. The inputs considered for the jet finder are calorimeter towers (calotowers) and topological clusters (topoclusters).

Calorimeter towers

In case of the towers, the cells are projected onto a fixed grid in pseudorapidity (η) and azimuth (ϕ). The tower bin size is $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ in the whole acceptance region of the calorimeters, i.e. in $|\eta| < 5$ and $-\pi < \phi < \pi$ with $100 \times 64 = 6.400$ towers in total. Projective calorimeter cells which completely fit inside a tower contribute their total signal. Non-projective cells and projective cells larger than the tower bin size contribute a fraction of their signal to several towers, depending on the overlap fraction of the cell area with the towers (see figure 1.36 for illustration). Following a cone-based jet finding algorithm cells within the towers as selected to from jets, as follows:

1. Finding a maximum E_T^{cell} in $(\eta - \phi)$ space use the cell as a seed for a jet candidate if $E_T^{seed} > E_{threshold}^{seed}$ where $E_{threshold}^{seed}$ is a parameter for the jet definition. The default $E_{threshold}^{seed}$ value for ATLAS cone-based jet algorithm is set to 1 GeV.

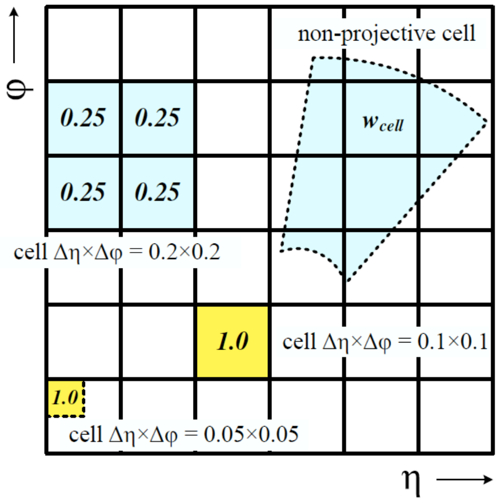


Figure 1.36: *Calorimeter cell signal contributions to towers on a regular* $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ *grid, for projective and non-projective cells. The signal contribution is expressed as a geometrical weight and is calculated as the ratio of the tower bin area over the projective cell area in* η *and* ϕ*.*

2. Make a cone of radius R around the seed center and sum the E_T of cells inside the cone. R is also a parameter of jet definition and as mentioned before, typical values are 0.4 and 0.7 for ATLAS jets.

3. A cone with $E_T > E^{cone}_{threshold}$ is a jet, where $E^{cone}_{threshold}$ is another parameter of the jet definition, indicating the minimum E_T contained in a jet. The default value used for the VBF $H \to \tau^+\tau^-$analysis presented in chapter 4, is 15 GeV.

4. Repeat the steps 1, 2, and 3 until there is no more seed found.

Topological clusters

Topological clusters are built according to criteria that identify significant energy deposits in topologically connected cells, similarly to the $\not{E}_T$*-topo* noise suppresion method (see section 1.3.4.2). Three different levels of signal significance are applied to the seed, the neighboring and surrounding cells. Currently, the settings of $(4, 2, 0)$ in units of σ_{noise} are used. Similarly to what mentioned in previous paragraph, a so-called topo-jet is formed when the cluster$E_T > E^{cone}_{threshold}$. As it will be further discusses in the chapter 4, topo-jets show better performance in the forward regions of the detector and this is the reason why they are used in the VBF $H \to \tau^+\tau^-$study.

Figure 1.37 illustrates the different approach of the two jet algorithms. It refers to a QCD sample of 4 jets in the final state. Cluster jets provide a more refined noise suppression and as a consequance less cells are finally selected to form a jet. The main common feature of all jet finder implementations in ATLAS is the full four-momentum recombination whenever the constituents of a jet change, either through adding a new constituent, or by removing one, or by changing the kinematic contribution of a given constituent to the jet. Also, in the ATLAS reconstruction software framework ATHENA, the same jet finder code can be run on objects like calorimeter signal towers, topological cell clusters in the calorimeters, reconstructed tracks, and generated particles and partons.

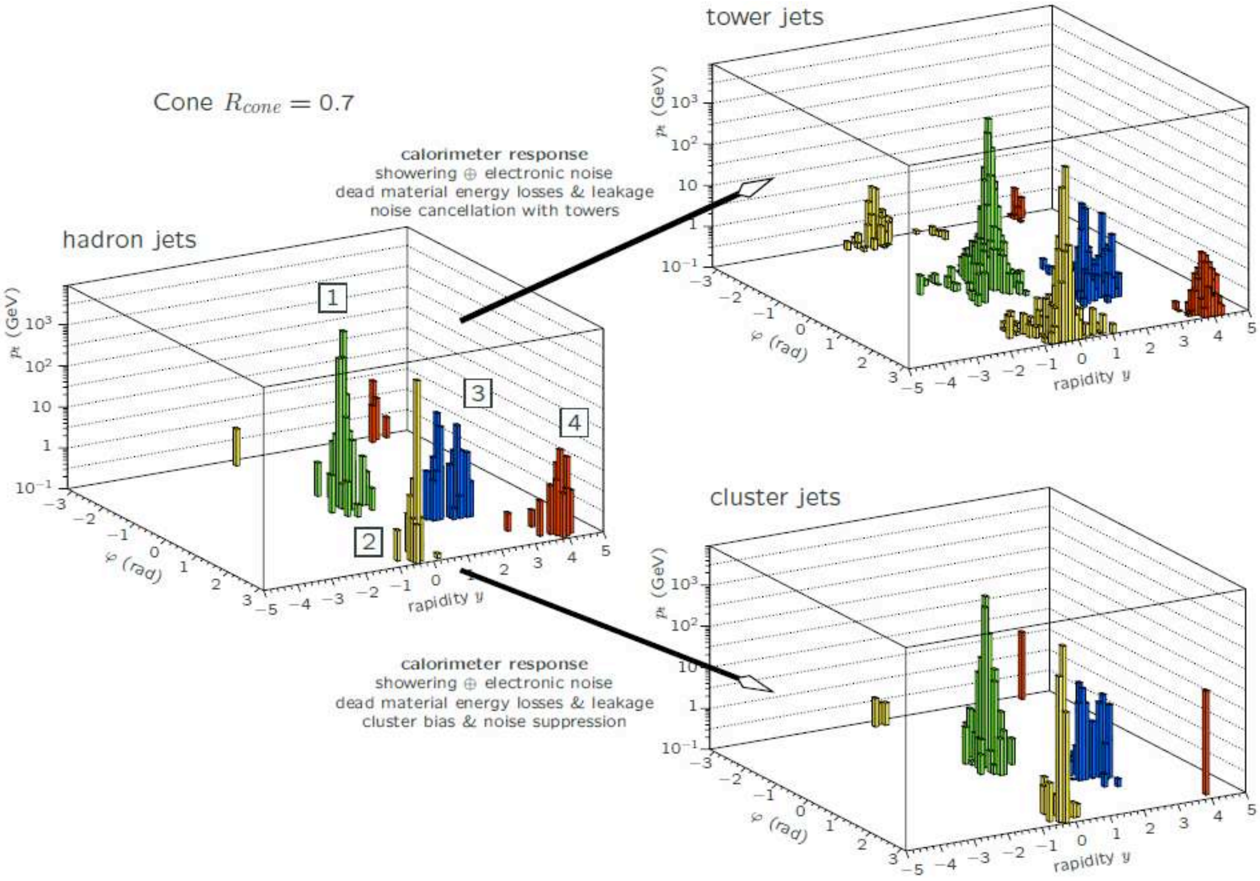


Figure 1.37: *A simulated QCD event with four jets in the final state, as seen at particle level and in the ATLAS calorimeters when using towers or clusters [19].*

Two calibration approaches are developed in ATLAS, the *global* and the *local* schemes [16]. The global scheme uses H1-style weights [16] to correct calorimeter cells after jet finding. The weights are based on the energy density in a cell. This calibration is specific to each type of jet finder. In the local scheme, jets are built from pre-calibrated topoclusters, which already include hadronic calibration, as well as dead material and out-of-cluster corrections.

1.3.6 b-tagging

Another important aspect of ATLAS performance is the ability of identifying jets originating from the hadronization of b quarks, commonly called b-tagging. This aspect is particularly useful to select very pure top samples, to search and/or study Standard Model or supersymmetric (SUSY) Higgs bosons which couple preferably to heavy objects or are produced in association with heavy quarks, to select or veto the large dominant $t\bar{t}$ background for several physics channels and finally to search for new physics: SUSY decay chains, heavy gauge bosons, etc.

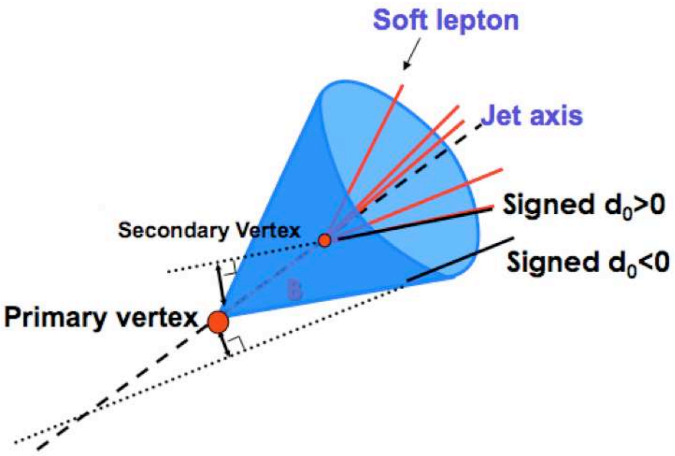


Figure 1.38: *A schema illustrating a b-jet. Primary and secondary vertices are shown as long as the soft lepton, a product decay of the b-quark.*

The identification of b-jets takes advantage of several of their properties which allow to distinguish them from jets which contain only lighter quarks. First the fragmentation is hard and the b-hadron retains about 70% of the original b quark momentum. In addition, the mass of b-hadrons is relatively high ($> 5GeV$). Thus, their decay products may have a large transverse momentum with respect to the jet axis and the opening angle of the decay products is large enough to allow separation. The third and most important property is the relatively long lifetime of hadrons containing a b quark, of the order of 1.5 ps ($c\tau \sim 450\mu$m). A b-hadron in a jet with $p_T = 50GeV$ will therefore have a significant flight path length $< l >= \beta\gamma c\tau$, traveling on average about 3 mm in the transverse plane before decaying. Such displaced vertices can be identified inclusively by measuring the impact parameters of the tracks from the b-hadron decay products. The transverse impact parameter, d_0, is the distance of closest approach of the track to the primary vertex point, in the $r-\phi$ projection. The longitudinal impact parameter, z_0, is the z coordinate of the track at the point of closest approach in $r-\phi$. The tracks from b-hadron decay products tend to have rather large impact parameters which can be distinguished from tracks stemming from the primary vertex. The other more demanding option is to reconstruct explicitly the displaced vertices. These two approaches of using the impact parameters of tracks or reconstructing the secondary vertex will be referred as spatial b-tagging. Finally, the semi-leptonic decays of b-hadrons can be used by tagging the lepton in the jet. In addition, thanks to the hard fragmentation and high mass of b-hadrons, the lepton will have a relatively large transverse momentum and also a large momentum relative to the jet axis. This is the so-called soft lepton tagging (the lepton being soft compared to high-p_T leptons fromW or Z decays).

Figure 1.38 illustrates a b-jet showing the displacement of the secondary vertex which corresponds to the b quark decay vertex. In ATLAS, a sophisticated algorithm takes into

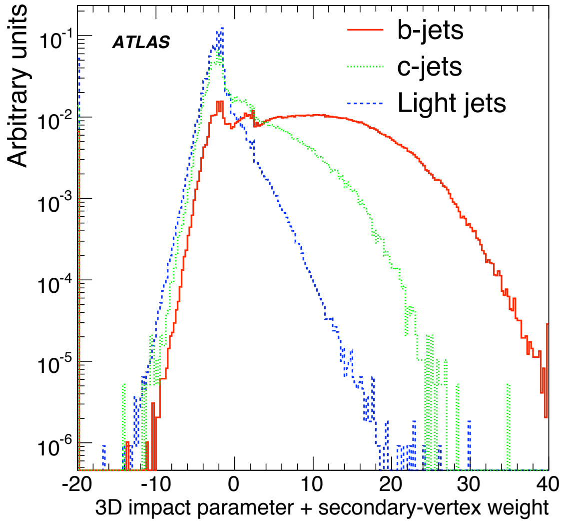


Figure 1.39: *Jet b-tagging weight distribution for b-jets, c-jets and light jets [19].*

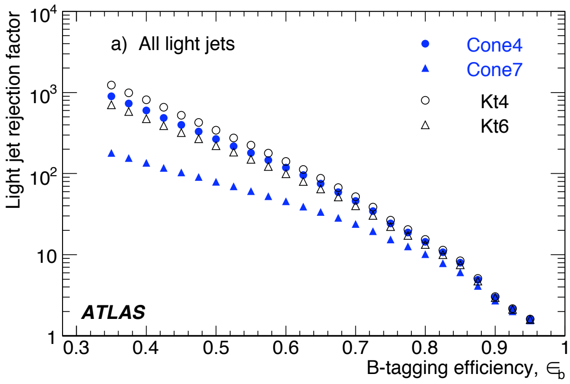


Figure 1.40: *Rejection of light jets versus b-tagging efficiency for the default b-tagging algorithm applied on reconstructed jets with different algorithms: cone algorithm with size* ΔR*=0.4, 0.7 or* k_T *algorithm [19].*

consideration the spatial taggers and the soft lepton taggers described below and based on a likelihood ratio method, it gives b-tagging weights to jets in order to discriminate b-jets from other jets originating from lighter quarks or gluons. In figure 1.39 an example of the discrimination power of this weight is shown for b-quark jets, c-quark jets and a sample of several light jets. Impact parameters and secondary-vertexing are taken into consideration for its calculation.

All b-tagging results are given for jets reconstructed with a cone algorithm of size $\Delta R = 0.4$. In this study, electrons faking jets were removed. Figure 1.40 shows the rejection of light jets versus the b-tagging efficiency obtained with the default tagging algorithm, for several jet algorithms run on $t\bar{t}$ events. As observed, the results are different for different jet definitions. The reason is that in principle, broader jets could be more easily contaminated by neighbouring tracks originating from distinct partons whose showers could not be resolved: light jets for instance could be contaminated by heavy-flavour decay products. However, this effect should be marginal since the maximum track-jet distance for association is kept to $\Delta R = 0.4$ in all cases.

1.4 LHC early planning

At the time of writing this document (summer 2009), the LHC is scheduled to start operating at the late autumn of 2009 for one year. The plan is to start with a beam energy at 3.5 TeV and following the evaluation of this operation it is expected to rise the energy up to 5 TeV per beam. An energy of 3.5 TeV per beam corresponds to an integrated luminosity of $\sim$100 pb^{-1} which can be translated into the observation of more than 25.000 events of $Z \rightarrow ee$ for instance. It is clear that a sufficient number of Standard Model events will be produced with this expected luminosity and thus, several Standard Model oriented physics analysis's will take place. Lead-ions are expected to be accelerated for the first time by the end of 2010, terminating the first complete running period of LHC. A shut down will take place next and finally, an operation at the nominal energy per beam (7 TeV) is scheduled afterwards, with an estimated starting date at the late 2011 or 2012.

Chapter 2

ATLAS Calorimeter Commissioning

2.1 Introduction

Since the year 2008, all the ATLAS sub-detectors are installed in their final position and progressively, they have been all integrated in cosmic data acquisition as well as single beam data acquisition in September 2008. Several trigger menus were used during this period such as Minimum Bias triggers [19], L1Calo triggers [19], Cosmic Muon triggers [36] and Random Triggers. The following commissioning study is focused on random triggers and the noise measurement is done through calorimeter transverse missing energy variables, which allow a global study of combining LAr and Tile calorimeters.

The transverse missing energy reconstruction has been described in details in section 1.3.4, while in this chapter the focus is on the calorimeter part of $\not{E}_T$, which is the most significant contribution to $\not{E}_T$.

2.1.1 Calorimeter $\not{E}_T$ motivations

$\not{E}_T$ is crucial for many physics analyses in ATLAS and a good measurement of the missing transverse energy is of high importance. Several examples in the early data such as $t\bar{t}$, W, Z bosons (figure 2.1), τ leptons require a precise measurement of $\not{E}_T$ resolution due to the presence of neutrinos in order to achieve an accurate mass reconstruction of the final state. Furthermore, for SUSY searches (figure 2.1) or for Higgs searches especially in the Higgs boson decay to a pair of τ-leptons (see chapter 4) a good $\not{E}_T$ measurement is very important. Not only, it plays a significant role in the mass reconstruction of new particles but it also contributes in suppressing SM background (see figure 2.1 at the right).

For commissioning, calorimeter $\not{E}_T$ variables have the advantage of being global variables combining LAr and Tile calorimeters. Furthermore, when an $\not{E}_T$ study is performed in random triggers different tasks are being accomplished:

- several problems can be spotted, such as noisy cells or even calorimeter regions that shouldn't give significant energy contributions

- calorimeter energy readout in noise is being validated through Monte Carlo, and thus a control of calorimeter based tools such as jet/$\not{E}_T$, is achieved.

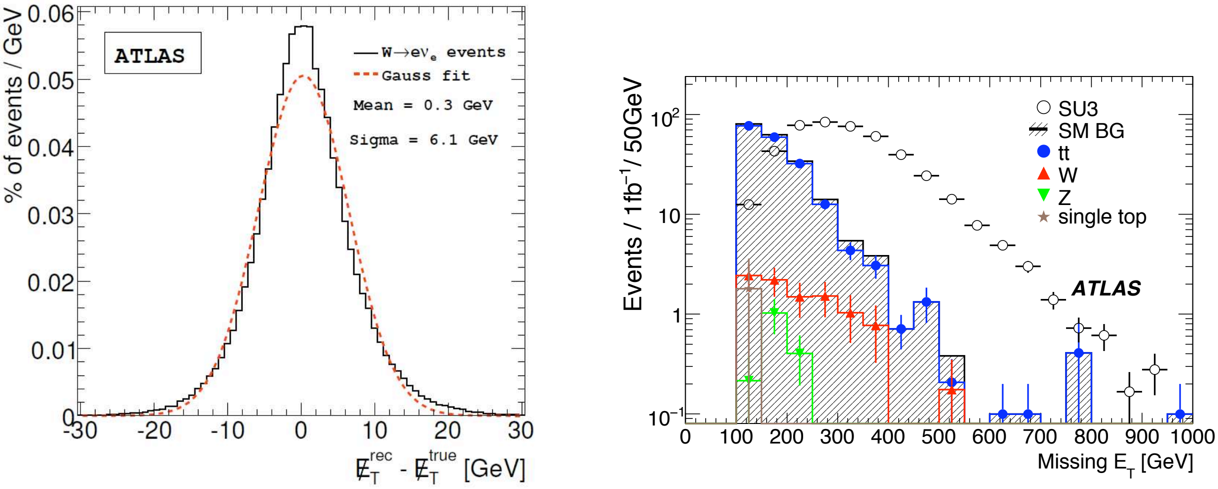


Figure 2.1: **Left plot:** *The* $\not{E}_T$ *resolution obtained from* $W \rightarrow e\nu$ *events.* **Right plot:** $\not{E}_T$ *distribution for the background processes for an example of SUSY benchmark point (SU3) in the one-lepton mode for an integrated luminosity of 1* fb^{-1}*. The hatched histogram shows the sum of all Standard Model processes [19].*

- a good test of calorimeter operation a few months before the LHC start-up.
- coherent noise can be looked for which is not present in MC. A small shift of each cell which is systematic could result to a signifciant overall shift in the $\not{E}_T$ distributions.

Figure 2.2[19] illustrates $\not{E}_T$ resolution as a function of $\sum E_T$ for LHC low luminosity phase[1]. One could expect a noise $\not{E}_T$ resolution which does not exceed 2 GeV, since in an empty event described by random triggers, the $\sum E_T$ per event is not expected to exceed a level of ~10 GeV.

2.2 ATLAS data taking

2.2.1 Random triggers

As described in section 1.2.2, the calorimeter system is composed by around 180k read-out cells giving a significant contribution in electronics noise. This noise should be understood and thus modeled in order to successfully be removed when reading signals in ATLAS calorimeters. The adequate trigger fro performing noise studies is the random trigger, which is designed to serve commissioning tasks. It refers to empty events (often called *zero bias* events) where the only contribution should be the electronics noise. Its commissioning use is mainly to search for signs of energy excesses, or other unexpected effects which will conclude to a problematic read-out channel. During cosmic data taking, a random trigger is fired with a frequency of ~1 Hz without requiring any usual physical trigger criteria. Possible overlap with other trigger streams is not removed, but the possibility of such an overlap is very low ($< 10^{-6}$) and therefore, it is neglected. During collisions, it is foreseen that the

---

[1] early run phase

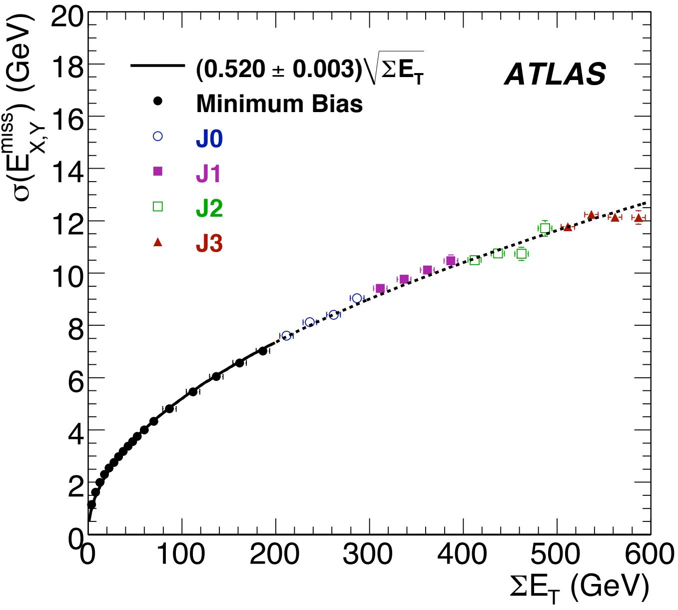


Figure 2.2: *$\not{E}_T$ resolution in QCD dijet events (J0-J3 : different p_T thresholds) together with $\not{E}_T$ resolution from minimum bias events as a function of $\sum E_T$. The integrated luminosity is of the order of $10^{-5}pb^{-1}$.[19]*

random triggers will be also synchronized to empty bunch crossings, therefore overlap with collision triggers will be avoided. The random trigger apart from allowing to monitor noise time fluctuation, it will provide a measurement noise auto-corellation and thus, it will be useful a tool in incorporating the noise in Monte Carlo samples tuned to data.

All the analysis to be presented in following chapter is based on the random triggers stream.

2.2.2 ATLAS cosmic rays data taking

Since the early months of year 2008 all ATLAS sub-detectors were installed in their final position and since the beginning of September 2008 they are all integrated in the data acquisition being ready for the first collisions. Just after the LHC incident [37] which caused the postponement of LHC collisions data an ATLAS "exposure" to cosmic muons was performed for detector commissioning reasons with a duration of two months. As a consequence, thousands cosmic muons were triggered, as long as millions of random triggers. Cosmic events were first recorded in stable conditions with the full-scale detector and it was possible for the first time in ATLAS to exercise the object reconstruction in two different contexts: pure noise (figure 2.3) and localized calorimeter energy deposit caused by muon hard bremsstrahlung (figure 2.4).

In figure 2.3, the homogenous dispersion of signal in ATLAS calorimeters is observed, which reflects the capture of a random trigger. On the other hand, figure 2.4 illustrates an accumulated energy deposit from a cosmic muon localized at the regions of calorimeter

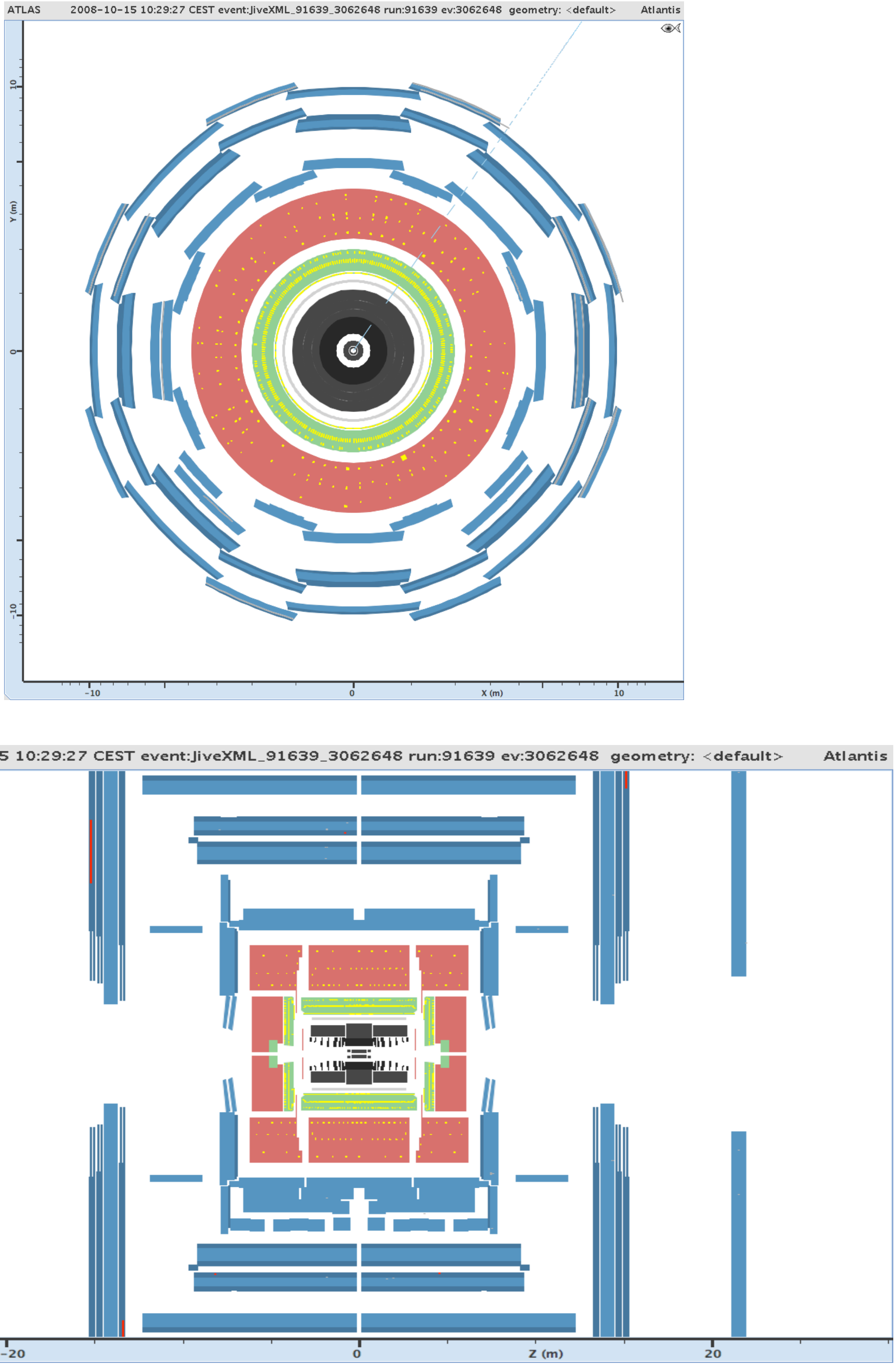


Figure 2.3: *ATLAS event display of a randomly trigger event. The green lines show the muon segments. The yellow boxes indicate energy deposits in the Tile Calorimeter and the direction of the reconstructed jet is shown by the gray marker (lower image). The light blue dotted line is the direction of* $\not{E}_T$ *(upper image). No track was reconstructed in the inner detector for this event. The X-Y projection (upper image) and the R-Z projection (lower image) are shown. Both negative and positive energy is illustrated with the same (yellow) color. The absolute value of the energy is used for the cells with a negative energy.*

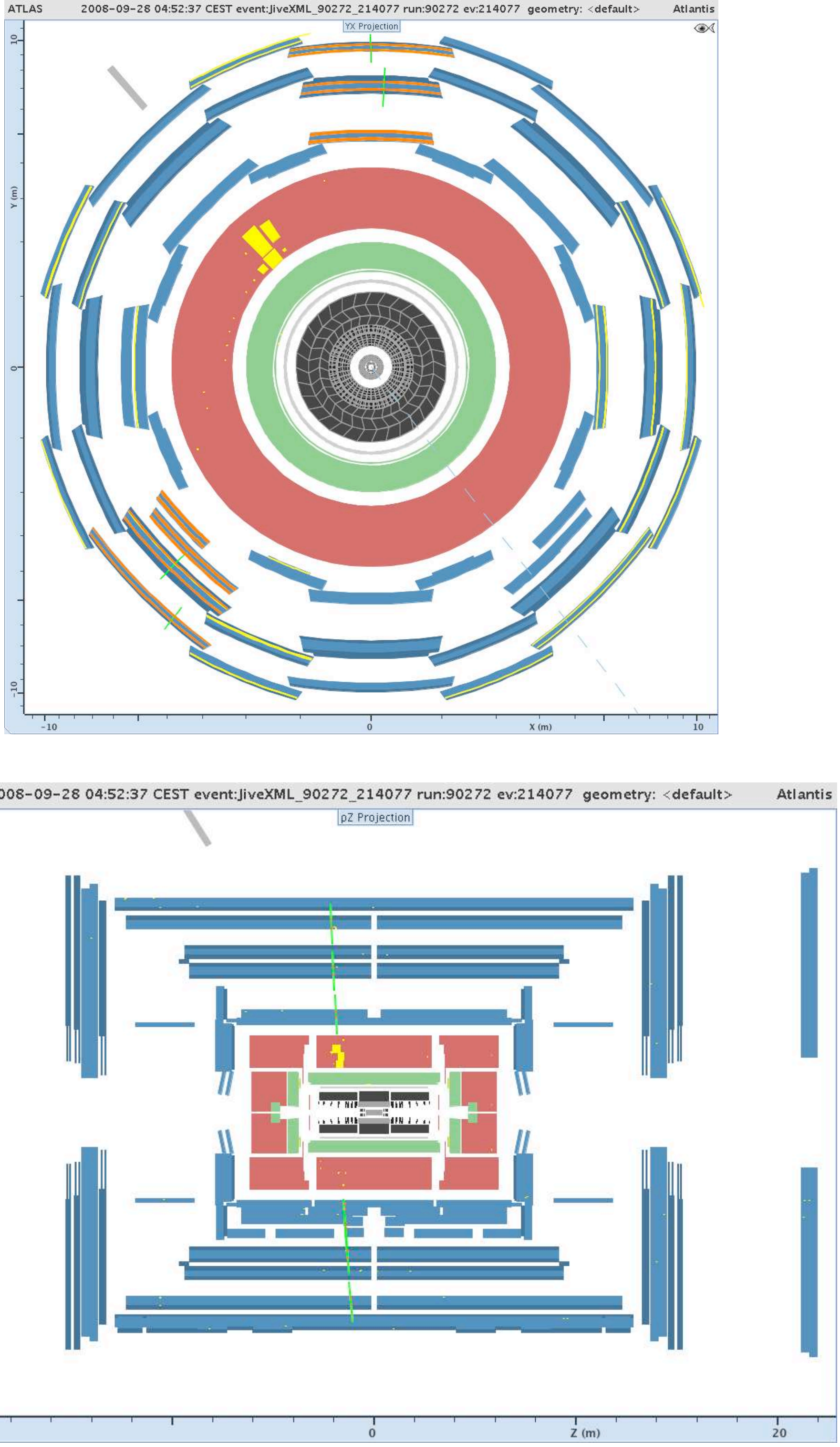


Figure 2.4: *ATLAS event display of an event with a single cosmic muon passing through the detector and depositing more than 1 TeV in Tile Calorimeter cells, through hard bremsstrahlung. The green lines show the muon segments. The yellow boxes indicate energy deposits in the Tile Calorimeter and the direction of the reconstructed jet is shown by the gray marker (lower image). The light blue dotted line is the direction of* $\not{E}_T$ *(upper image). No track was reconstructed in the inner detector for this event. The X-Y projection (upper image) and the R-Z projection (lower image) are shown.*

that correspond to its track. It is to be noted that the energy dispersion of figure 2.3 is also present in the high muon momentum case, but it is negligible and thus, they are not illustrated as in figure 2.3.

The most interesting cosmic runs of these 2 months of data taking were reprocessed during the period of Christmas 2008 using the most adequate ATLAS reconstruction tools. All results presented in the following sections are by default obtained using outputs from ATLAS algorithms using this reprocessed data. Those events were reprocessed using one single database containing all the calibration constants concerning in particular pedestal shift and noise measurements (section 1.2.2.4). Finally, results obtained studying the same runs but modifying the default reconstruction settings will also be shown.

2.2.3 Reference runs

For the results to be presented in following sections, a reference run was selected from the cosmic data taking period of September - October 2008. It is the run 91639 which started at 18:51 of 14^{th} October and ended the 15^{th} October 2008 at 09:36 completing ~14 hours of run, having all the different partitions of ATLAS detector on. This night run resulted in a ~50.000 events statistic of random triggers.

Some other runs containing random triggers were also selected during the same period to perform calorimeter $\not{E}_T$ studies in order to obtain results related to time evolution [38]. From the LAr calorimeter point of view, all examined runs were 10-samples runs and from these 10 sample, 5 are used for the energy reconstruction as it will be the case for LHC collisions data.

2.3 LAr Energy reconstruction during cosmic data taking period

The method to reconstruct the energy in LAr cells has been described in previous section. As shown section 1.2.2.1, a good reconstruction is obtained with the method of Optimal Filtering Coefficients. However, OFC can be applied to events with fixed phase upon small variations. Concerning LHC collision data, the bunch time crossing will be known and therefore the time phase as well. However this is not the case for cosmic data taking where cosmic muons impinge the calorimeter asynchronously. In this scenario, Optimal Filtering can still be applied and it is possible to obtain an accurate reconstruction introducing an iterative procedure, specially developed for this reason, which is defined as follows:

2.3.1 Iterative procedure

The iteration procedure consists of an estimation of the initial phase that indicates the index of the maximum sample. If the maximum sample is the central sample (section 1.2.2.2) the first iteration begins with Optimal Filtering Coefficients calculated for

$$\tau_0 = 0 \; ns \quad , \tag{2.1}$$

while, if the maximum sample differs from the central sample, the coefficients are calculated for

$$\tau_0 = 25(i_c - i_{max}) \; ns \quad , \tag{2.2}$$

where i_{max} and i_c show the index of the maximum and the central sample respectively.

Afterwards the amplitude and phase are calculated in each iteration as

$$A_k = \sum_{i=1}^{n} a_i \mid_{\tau_{k-1}} S_i \quad , \tag{2.3}$$

$$\tau_k = \frac{1}{A_k} \sum_{i=1}^{n} b_i \mid_{\tau_{k-1}} S_i \quad , \tag{2.4}$$

where k is the iteration index starting from 1. It has been observed that the amplitude converges within three iterations for Optimal Filtering weights calculated in steps of 1 ns [40].

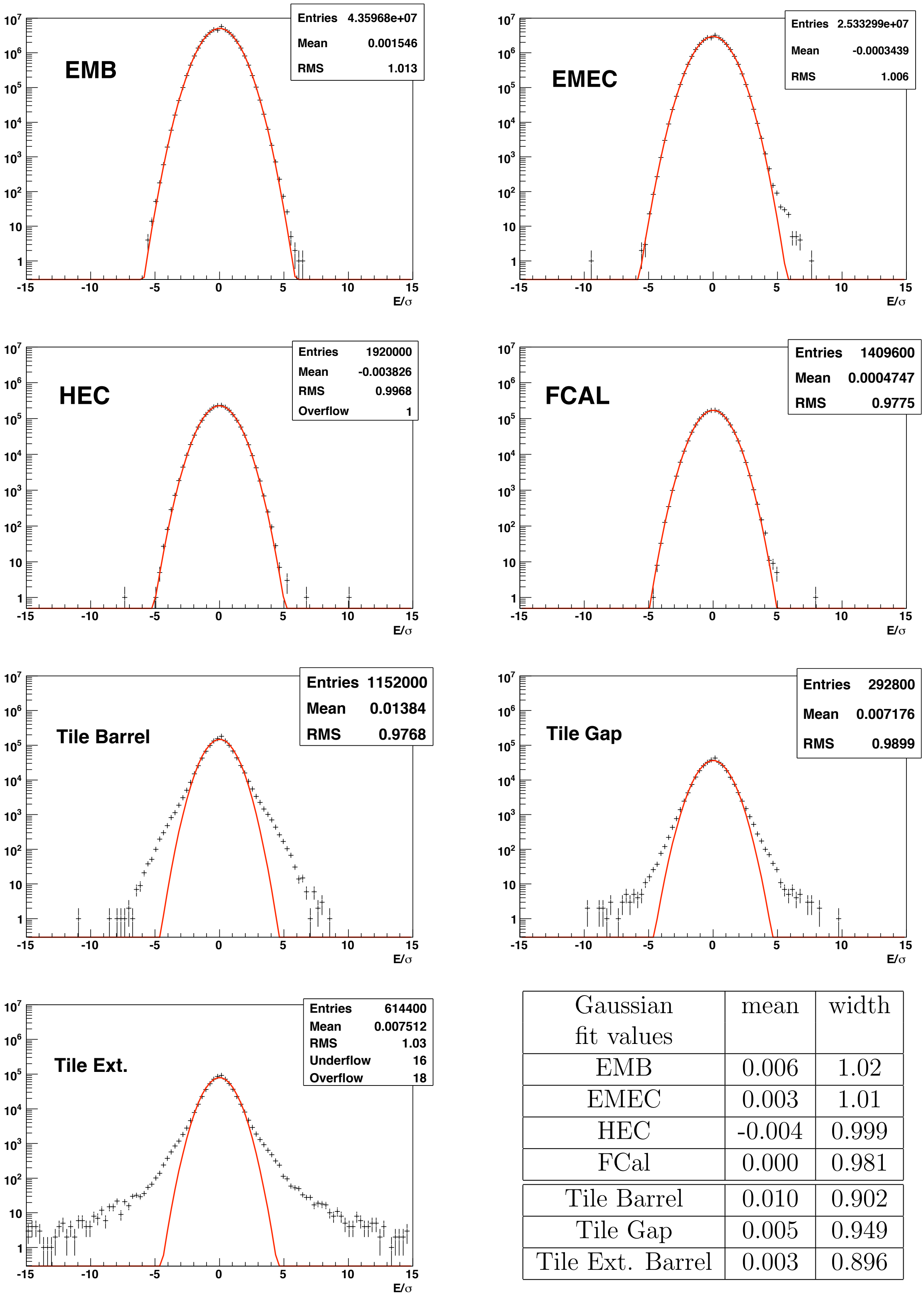
This iterative procedure provides an accurate value of the amplitude for cells with large signal. However, for cells with small or no signal, the iteration procedure can favor fluctuations towards positive values overestimating the amplitude. In order to avoid this bias to positive values the reconstruction algorithm should identify the events without signal and apply a different procedure. In ATLAS LAr reconstruction algorithm a condition is imposed in order to identify real signal from noise signal (empty events), therefore the iteration is activated if :

$$A > n \cdot P_{RMS} \tag{2.5}$$

where A indicates the sample amplitude in ADC counts, and P its electronics noise expressed in ADC counts similarly. A typical value of the n is 3 or 4 and in the following section, the effect of the n_{iter} value is studied. For the cases where $A < n_{iter} \cdot P_{RMS}$, the signal is qualified as noise and therefore no iteration procedure is performed.

2.4 Stability of cell noise and pedestal

One single set of calibration constants, including pedestals and noise, is used to process all cosmic runs of period September-October 2008. A first global check of the pedestal and the noise can be obtained by looking at the distribution of E/σ_{noise} of all the calorimeter cells for a few hundreds of events of the reference run (figure 2.5). Data (scattered points) are fitted with a simple Gaussian distribution (red line). At the bottom right part of figure 2.5, a table is presented with the fit values of mean and width of these Gaussian fits. In all the liquid argon sub-detectors, the mean is ∼0 and the width ∼1, a proof that noise in these sub-detectors follows a Gaussian behavior, as expected. A slight discrepancy is shown in the right tail of EMEC, which caused by one cell with 60 MeV shifted pedestal in the first layer (at $(\eta, \phi) = (-1.816, -2.990)$, FT24, slot 4, Ch.67.)[38]. On the contrary, sizable non Gaussian tails are observed in Tile sub-detectors for a reason which is under investigation. A possible explanation of this effect (or part of it) could be the presence of coherent noise



Gaussian fit values	mean	width
EMB	0.006	1.02
EMEC	0.003	1.01
HEC	-0.004	0.999
FCal	0.000	0.981
Tile Barrel	0.010	0.902
Tile Gap	0.005	0.949
Tile Ext. Barrel	0.003	0.896

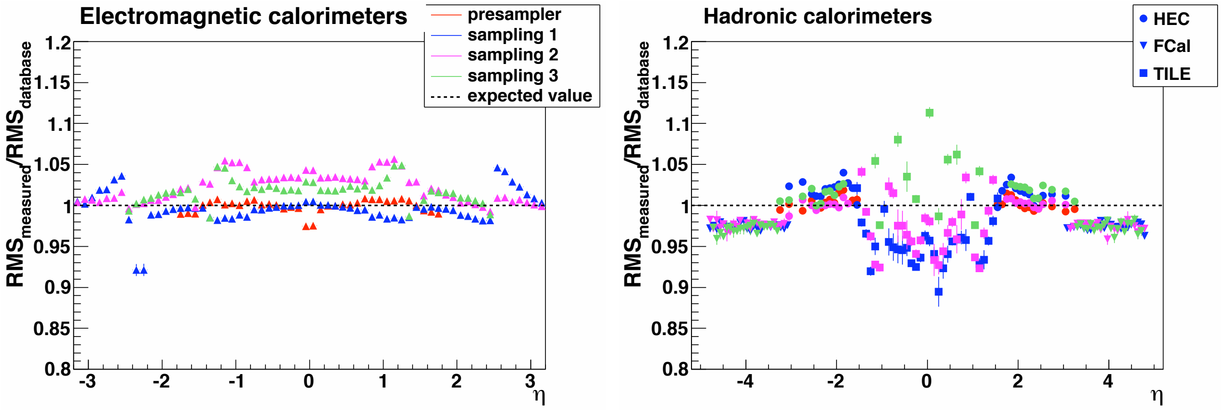
Figure 2.5: *E/σ_{noise} distributions for all the calorimeter cells, per sub-detector. The markers correspond to cell energy measured in a sample of 400 randomly triggered events of reference run (91639). The red lines correspond to Gaussian fits, whose results are indicated in the bottom right table.*

in Tile sub-detectors [39]. No clear connection is shown and therefore until the finalisation of the studies within the Tile calorimeter community, this is an assumption.

Since it is important to have an understanding of the noise/pedestal level in order to perform the $\not{E}_T$ and jet reconstruction, the time stability[38] of these quantities is discussed in following section. It is to be noted, that measurements of all calibration constants, including pedestal shifts and noise will be taken in a day-by-day basis allowing more realistic and accurate measurements of energy reconstruction, during collisions data taking.

2.4.1 Noise level and stability

The noise is calculated for each cell as the RMS of the energy distribution, without demanding any requirements related to the cell energy. As presented in section 1.2.2.4, electronics noise values, called σ_{noise}, were computed for each cell, using the random trigger stream of run 92048 (22 October 2008). As a check, the same σ_{noise} quantities are recomputed with the reference run taken one week before. The ratio of σ_{noise} measured in both runs is shown, as a function of η for every layer of the calorimeter system for the electromagnetic calorimeter (figure 2.6(a)) and for hadronic calorimeter (figure 2.6(b)). A fair agreement below 5% is observed for the liquid argon sub-detectors, while 10% is the respective one for Tile sub-detectors. Differences observed in the LAr detectors are probably linked to a change of the default phase in the energy computation (section 1.2.2.2) between the database noise processing and the current processing. It should be noticed that they are more pronounced (3-5% systematic excess) in the second and third layers of the EM calorimeter. The systematic drop of 5-10% observed in the Tile calorimeter ($|\eta| < 1.5$) is linked to Tile non-Gaussian noise behaviour in this detector.

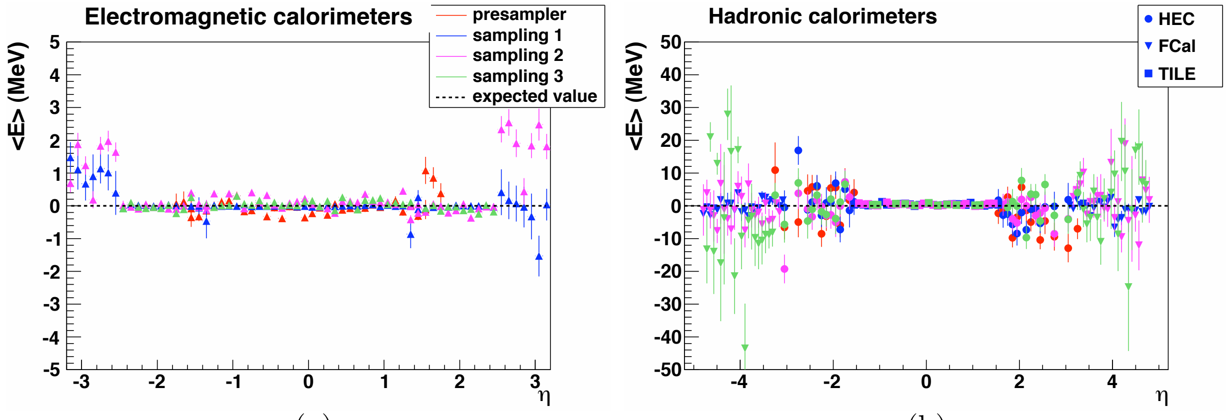


(a) (b)

Figure 2.6: σ_{noise} *, i.e. RMS of cell energy distribution measured in reference run (91639), compared to the official database values in the electromagnetic (a) and hadronic (b) calorimeters. Values are averaged over* φ *and 3000 events are used. For both plots, the different colors correspond to the different samplings as illsutrated in the legend of plot 2.6a [38].*

2.4.2 Pedestal stability

Pedestal values included in the official ATLAS database, used for were computed at the beginning of the period from dedicated calibration runs per sub-detector (13^{th} of September 2008). In the following, the so-called "effective pedestal" is computed for each cell as the mean of the energy distribution ($< E >$) for the reference run. These recomputed quantities are shown as a function of η in every layer of the calorimeter system for the electromagnetic calorimeters (figure 2.7(a)) and the hadronic calorimeters (figure 2.7(b))[38]. Apart from very small zones, all values are within ±0.5 MeV for electromagnetic and within ±10 MeV for the hadronic calorimeter. These variations are compatible with the measured noise variations (figures 1.15 and 1.16).



(a) (b)

Figure 2.7: *Mean cell energy* $< E >$ *measured in reference run (91639) for the electromagnetic (a) and hadronic (b) calorimeters. Values are averaged over* φ *and 3000 events are used [38].*

2.5 $\not{E}_T$ observables

$\not{E}_T$ was defined in section 1.3.4, and it was shown that it consists of a calorimetric term, a muon term and some extra corrections due to the existence of dead matter in ATLAS detector where particles loose part of their energy, which cannot be measured. $\not{E}_T$ can be written as:

$$\not{E}_{x,y}^{Final} = \not{E}_{x,y}^{Calo} + \not{E}_{x,y}^{Muon} + corrections \tag{2.6}$$

For the needs of this study, a focus is done in the most dominant contribution of $\not{E}_T$, which is the calorimetric term and for the following sections,all $\not{E}_T$ notations will refer to the respective $\not{E}_T^{Calo}$ notations. For the geometrical cell angles θ and ϕ, E_X^{miss} , E_Y^{miss} and φ^{miss} are defined as follows:

$$E_X^{miss} = -\sum_{i}^{cells} E_i \cdot sin\theta_i \cdot cos\phi_i \quad , \tag{2.7}$$

$$E_Y^{miss} = -\sum_i^{cells} E \cdot sin\theta_i \cdot sin\phi_i \quad , \tag{2.8}$$

$$\phi^{miss} = atan\frac{E_Y^{miss}}{E_X^{miss}} \quad , \tag{2.9}$$

where the sum is done over the cell energy for each event. The "minus" sign is explained in 1.20 and E_X^{miss} , E_Y^{miss} and φ^{miss} refer to the x-, y-component of $\not{E}_T$ and to the angle defined by these two, respectively. It is to be reminded that (x, y) coordinates define the transverse plane where $\not{E}_T$ is defined perpendicular to the z beam axis. Furthermire the $\sum E_T$ variable is defined:

$$\sum E_T = \sum_i^{cells} E_i \cdot sin\theta_i \quad , \tag{2.10}$$

which is the sum of the transverse cells energy per event. $\sum E_T$ as it will be discussed later, is a very sensitive variable and thus, useful for commissioning studies. The reason is that unexpected cell energy behavior may compensate within the frame of E_X^{miss} or E_Y^{miss} variables, while it will cause a bias at the $\sum E_T$ variable, due to the additive character of the transverse projection of cell energy. Finally the $\not{E}_T$ variable used in physics analysis is defined as follows per event:

$$\not{E}_T = E_T^{miss} = \sqrt{(E_X^{miss})^2 + (E_Y^{miss})^2} \quad , \tag{2.11}$$

Those are the $\not{E}_T$ observables to be discussed in the following sections. It is to be noted that the sum in energy for equations 2.7, 2.8 and 2.10 is performed fore those cells, which are selected using either the $\not{E}_T$*-topo* or $\not{E}_T$*-base* noise suppression methods described in sections 1.3.4.1 and 1.3.4.2.

For the needs of this analysis, a software tool was developed which allowed the reading of the entire number of the calorimeter read-out cells (∼180k channels), obtaining information such as the energy, the (η, ϕ, θ) coordinated, the σ_{noise} , etc, in a cell-by-cell basis.

2.6 Toy Monte Carlo

For a better understanding of data, a toy Monte Carlo is built on a simple assumption: the energy of each cell is randomly taken in a Gaussian distribution centered on 0 with a standard deviation corresponding to the cell database noise (figures 1.15 and 1.16). The coherent noise is neglected. This building algorithm is implemented in ATHENA code, which allows to run *topoclusters* and $\not{E}_T$ algorithms on the randomized cells. Another advantage of implementing it in ATHENA: since several cells are marked as problematic and are masked when reconstructing the energy, they will remain masked when calling the toy Monte Carlo. The ATLAS software tool created for this purpose is called CaloCellRandomizer and it is now implemented in ATHENA official release, starting from ATHENA release 15.0.0.

CaloCellRandomizer acts at cell level and generates cell energy values derived from a Gaussian probability density function (PDF) which is centered at zero and has a width equal to the σ_{noise}:

$$P_{rand}(E) = \frac{1}{\sigma_{noise}\sqrt{2\pi}} e^{-E^2 \big/ 2\sigma^2_{noise}} \tag{2.12}$$

The σ_{noise} is unique for each cell and comes from an output of $CaloNoiseToolDB$. It is measured as the RMS of the cell energy distribution in one randomly triggered run (section 1.2.2.4). Cells that were masked, remain masked with zero energy. The energy randomization occurs at ATHENA reconstruction level and then all default algorithms such as *topocluster* algorithm are run taking as input this new energy. To be noted that CaloCellRandomiser can act independently randomizing the cells energy only in some of calorimeter sub-layers (an application is presented in section 2.8.4 for instance).

Another functionality of this tool is to simulate possible pedestal shift in calorimeter noise. As a result, the user can introduce an offset of the mean value of the Gaussian PDF which generates the energy values. This shift can be introduced for each sub-calorimeter if needed:

$$P^{shift}_{rand}(E) = \frac{1}{\sigma_{noise}\sqrt{2\pi}} e^{-\left(E-\mu_{i,shift}\right)^2 \big/ 2\sigma^2_{noise}} \tag{2.13}$$

where $\mu_{i,shift}$ can correspond to the shift introduced by user for EMB, EMEC, HEC, FCal concerning LAr and for Tile Barrel, Tile Gap, Tile Extended Barrel concerning Tile.

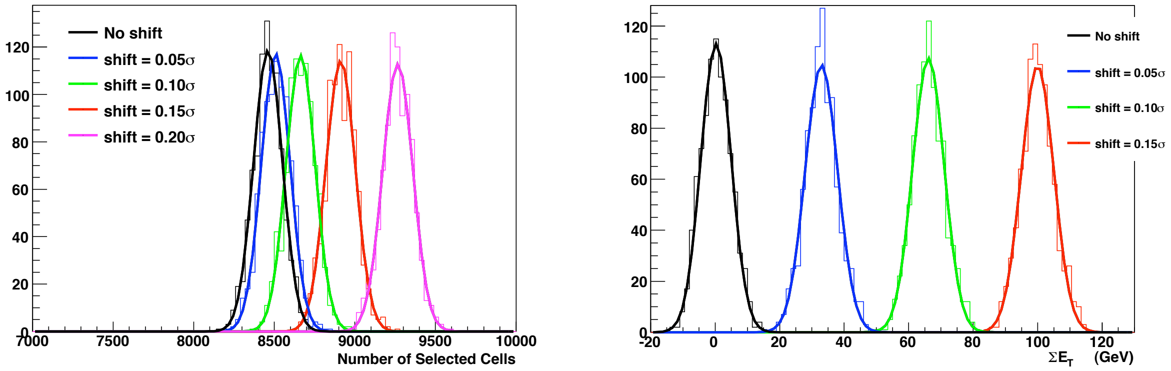


Figure 2.8: *$\not\!\!E_T$ toy Monte Carlo distributions for $\not\!\!E_T$-base : (left) Selected cells ($|E| > 2\sigma_{noise}$) and (right) $\sum E_T$. An overall shift of the mean value is applied (eq. 2.13) for (i) $\mu = 0.05\sigma_{noise}$ (blue), (ii) $\mu = 0.1\sigma_{noise}$ (green), (iii) $\mu = 0.15\sigma_{noise}$ (red), (iv) $\mu = 0.20\sigma_{noise}$ (purple). Distributions are compared with simple toy Monte Carlo case where no shift is applied (black) as described by equation 2.12.*

Figure 2.8 show an application of the toy Monte Carlo when applying an overall shift of the mean value at all the cells of the calorimeter for the $\not\!\!E_T$-*base* for a sample of 900 events. As it is extracted from the figure, when the shift gets bigger a positive bias is observed in distributions of selected cells per event and $\sum E_T$. No impact is observed in $\not\!\!E_T$variable since the shift is uniform in ϕ.

Table 2.1 contains the mean/RMS values of the histogram which describes the number of selected cells per event in addition to the theoretical expectation values. The expected

	Toy MC # cells / event	Expected # cells / event
No Shift	8461 ± 90	8463
Shift = $0.05\sigma_{noise}$	8509 ± 89	8513
Shift = $0.10\sigma_{noise}$	8663 ± 91	8664
Shift = $0.15\sigma_{noise}$	8912 ± 92	8916
Shift = $0.20\sigma_{noise}$	9266 ± 94	9269

Table 2.1: *Number of cells with $|E| > 2\sigma_{noise}$ per event for each subdetector in random events of reference run 91639. The expected numbers derived from the Gaussian noise model are in brackets.*

values are extracted by equation 2.13 when calculating the integral:

$$I = 1 - \int_{-2\sigma}^{2\sigma} P_{rand}^{shift}(E)dE \tag{2.14}$$

which corresponds to the probability $|E| > 2\sigma$ and multiplying it with the total number of cells in the calorimeter (180k). An excellent agreement is observed, a proof that the toy Monte Carlo incorporates successfully a shift of the mean value functionality.

The measured mean values of the $\sum E_T$ distributions are presented in table 2.2. A bias of $\sim +33$ GeV is added for every $5\%\sigma_{noise}$ shift of the mean value, a fact that shows the sensibility of $\sum E_T$ variable to possible shifts of pedestal. More details will be presented in next section.

	Toy MC $\sum E_T$ (GeV)
No Shift	$\mu \sim 0$
Shift = $0.05\sigma_{noise}$	$\mu \sim 33.1$
Shift = $0.10\sigma_{noise}$	$\mu \sim 66.0$
Shift = $0.15\sigma_{noise}$	$\mu \sim 100.0$

Table 2.2: *Measured mean values of $\sum E_T$ histograms which correspond to left plot of figure 2.8.*

Another extension of CaloCellRandomizer allows to parametrize energy using a toy Monte Carlo based on a double Gaussian pdf. Similarly to the simple case described below, the energy of each cell is randomly taken separately by two Gaussians (either by one either by the other one) which are characterized by different σ's. Such a model, would be more adequate to describe possible tails that exceed a Normal Distribution (see section 2.9).

It should be noted that quantification effects due to (i) the energy being derived from ADC counts and (ii) the cell energy saved in integer MeV in RDO are not simulated. The goal is to describe electronic noise in ATLAS calorimeters based on a Gaussian model and thus have a reference which will contribute to a better understanding of ATLAS calorimeters.

2.7 Cell based reconstruction method

In this section, results obtained with the simplest approach of noise suppression: cells enter the $\not\!\!E_T$ computation if their absolute energy is 2 times above the σ_{noise} ($\not\!\!E_T$*-base*), are discussed.

2.7.1 $\not\!\!E_T$ reconstruction

The first result presented, before investigating $\not\!\!E_T$ variables distributions, is the number of cells satisfying the noise suppression cut $|E| > 2\sigma_{noise}$, since the resolution of the $\not\!\!E_T$ distributions depends on this number. A total number of around 8000 EM and 600 Hadronic calorimeter cells are selected per event. The number of selected cells is shown in figure 2.9 as a function of η and in table 2.3 per subdetectors.

	EM Barrel	EM Endcap	Tile	HEC	FCal
# cells / event	5250 (4950)	2950 (2900)	250 (230)	220 (220)	140 (160)

Table 2.3: *Number of cells with* $|E| > 2\sigma_{noise}$ *per event for each subdetector in random events of reference run 91639. The expected numbers in derived from the Gaussian noise model are in brackets.*

They are compared to the expectations from the toy Monte-Carlo[1]. Data and expectations are consistent, except for the EM barrel where the $\sim$ 5% higher number of cells is explained by the underestimation of the database noise (section 1.2.2.4). As expected, data and expectations are in very good agreement when measured noise is used in function of η (figure 2.9). The drop appearing in $|\eta| = 1.5$ is explained due to the existence of the

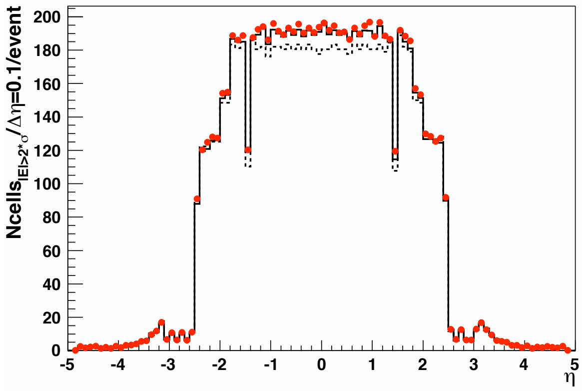


Figure 2.9: *Average number of cells with* $|E| > 2 * \sigma_{noise}$ *normalized to one random event of the reference run, per bin of* $\Delta\eta = 0.1$. *Red dots represent the data. Dashed line is the expected value with the database noise whereas the solid line is the expected value with the measured noise [38].*

---

[1] In this case, it is simply computed from $N_{cells}.\sqrt{2/\pi}.\int_2^{+\infty} e^{-t^2/2}dt = N_{cells}.(1-Erf(2/\sqrt{2})) \simeq N_{cells} *$ 4.55%.

transition region between calorimeter barrel and calorimeter end-cap.

Figure 2.10 shows the E_X^{miss} , E_Y^{miss} , $\sum E_T$ and φ^{miss} distributions for the reference run illustrating the measured data (scattered points) and the toy Monte-Carlo (solid line) in the full statistics sample of 50k events (see section 2.2.3). E_X^{miss} , E_Y^{miss} , $\sum E_T$ data distributions are Gaussian of width 3.4, 3.4 and 5.1 GeV respectively, in fair agreement with the simulation, which are shifted in the terms of mean value with the exception of E_X^{miss} . E_Y^{miss} mean value appears to be -0.6 GeV and $\sum E_T$ mean value equals to 1.7 GeV while they are all expected to be zero. The slight negative shift of $<E_Y^{miss}>$ is reflected to the φ^{miss} distribution where despite an expected flat behavior (solid line of Gaussian model), a *sine* effect is appeared. To better understand these discrepancies, it is interesting to look at the variation of these results within the full time period of data taking (Figure 2.11 left). The mean value of E_Y^{miss} is stable with a constant shift of -0.7 ± 0.3 GeV whereas for $\sum E_T$ it varies from 0 to 10 GeV depending on the run time. Looking at the results per sub-detector (figure 2.11 right) it is concluded that the former is due to EM calorimeter and HEC while the latter comes mainly from the EM barrel calorimeter. In both cases, a possible explanation can be the pedestal variation as a function of time (section 2.4). To examine this hypothesis, the effective pedestal was recomputed in four representative runs

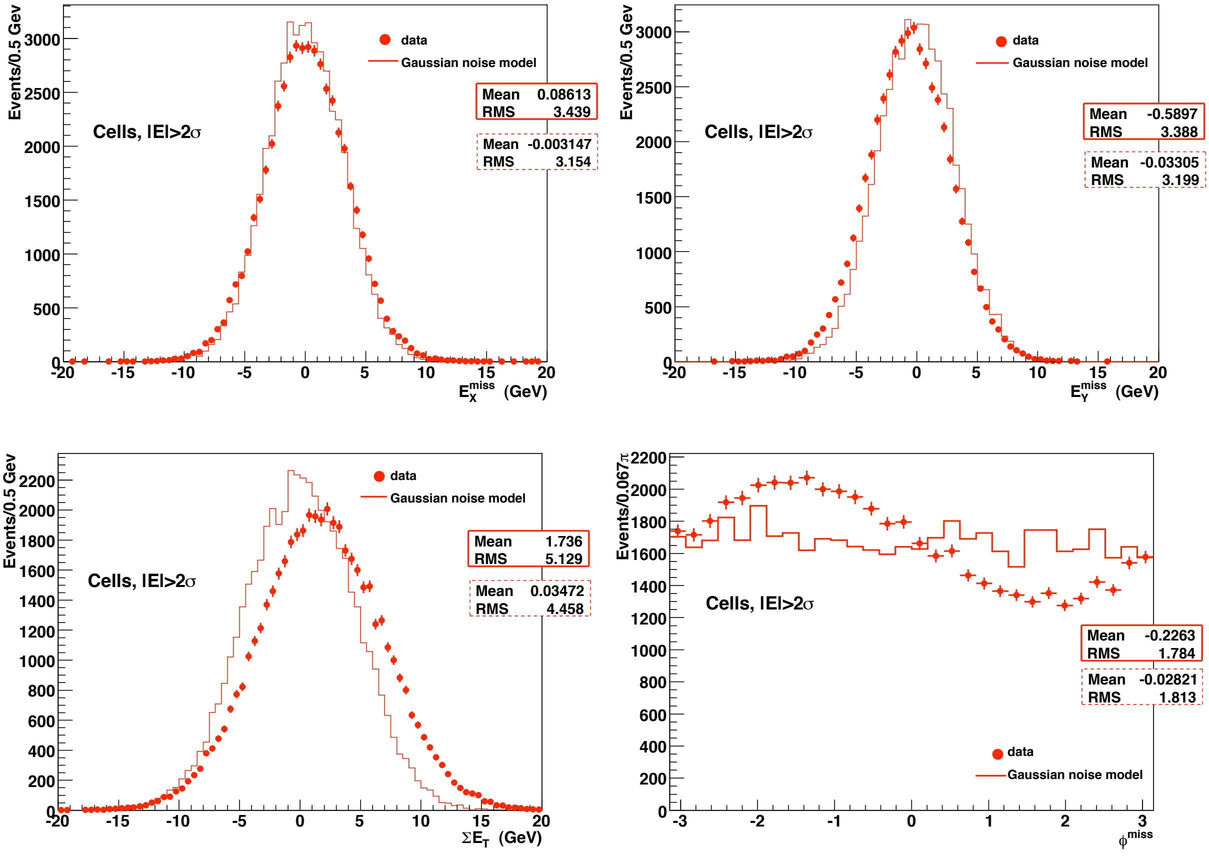


Figure 2.10: E_X^{miss} , E_Y^{miss} , $\sum E_T$, *and* φ^{miss} *distributions computed with* $|E| > 2\sigma_{noise}$ *noise suppression cut for random events of the reference run (91639). Red dots represent the data, whereas the solid line is the toy Monte-Carlo.*

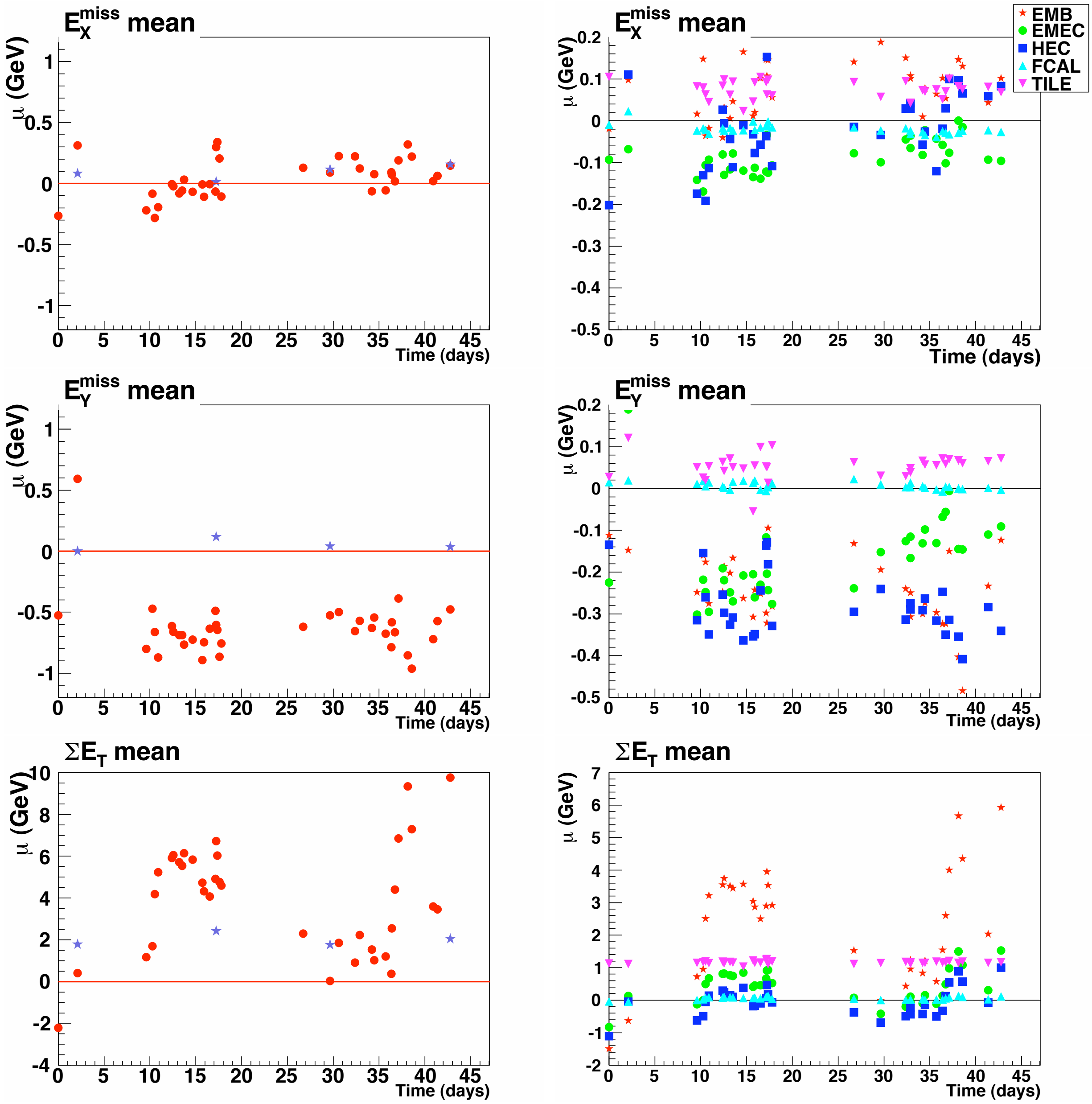


Figure 2.11: *Left:* E_X^{miss} *,* E_Y^{miss} *, and* $\sum E_T$ *Gaussian mean computed for* $|E| > 2\sigma_{noise}$ *noise suppression cut as a function of time, for randomly triggered events. Blue stars show the results obtained when recomputing the effective pedestal for some randomly triggered runs. Right: Same results per sub-detector of the calorimeter system [38].*

taken 2, 17, 30 and 43 days after the 10^{th} of September (as it will be when processing the LHC data). $\not\!\!E_T$ variables recalculated with this new input are represented with blue stars on figure 2.11 (left). E_Y^{miss} mean values are compatible with 0, which consequently flattens the φ^{miss} distribution of Figure 2.10 bottom right. $\sum E_T$ time variation is also suppressed and a stable residual $\sum E_T$ shift of +2 GeV is observed.

To investigate further the residual shift of +2 GeV presented in $\sum E_T$ distribution, the reference run was reprocessed using a 5σ threshold for the iteration on the energy reconstruction instead of the standard 4σ (section 2.3, equation 2.5). The effect at the $\sum E_T$ distribution is presented in figure 2.12 at a 10k events sample. The shift in the $5\sigma_{noise}$ case is -0.4 GeV,

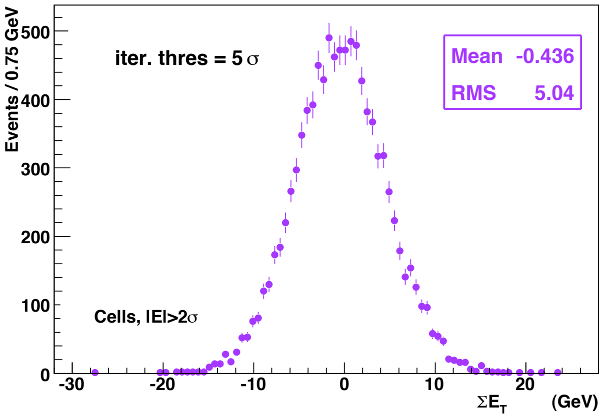


Figure 2.12: $\sum E_T$ *distribution computed with* $|E| > 2\sigma_{noise}$ *noise suppression cut for random events of the reference run (91639) modifing the iteration threshold from 4 to 5* σ_{noise} *. No pedestal correction is applied.*

which means that the $\sum E_T$ mean value is shifted 1.7-(-0.4)=2.1 GeV when modifying the iteration threshold from 4 to 5 σ_{noise} . This 2.1 GeV that remains is compatible with the 2 GeV stable shift observed in all the runs when correcting the pedestal (figure 2.11 left, bottom plot). Therefore, if applying both corrections, the results are valid with the expectations in which a $\sum E_T$ distribution is centered to zero.

To conclude, figure 2.13 shows the $\not\!\!E_T$ distribution of the reference run compared to the toy Monte-Carlo, the variable used in physics analyses. A good agreement is obtained in the core of the distribution, reflecting the very good understanding of the noise in the calorimeter system. Unexpected tails, representing 0.04% of the events, appear for $\not\!\!E_T > 15$ GeV.

Trying to explain the unexpected tails, a parallel study[38] was performed investigating the existence of jets, in the same reference run. The outcome of this analysis showed the existence of a few unexpected topo-jets. Normally, no jets are expected in empty events. This has been traced back to an abnormal coherent noise in the presampler calorimeter around the region (η,ϕ)=(-1.0,2.7). This introduced a fluctuation towards high energies of several presampler cells of this region, seeding topoclusters with a very high number of cells: of the order of 1000 cells in the jet, resulting to a jet reconstruction.

Following the clear proof of a coherent noise in presampler, this region of the calorimeter was entirely removed in a private reconstruction when recomputing $\not\!\!E_T$ distribution. The result is shown in figure 2.14, where the tail observed in figure 2.13 has now disappeared.

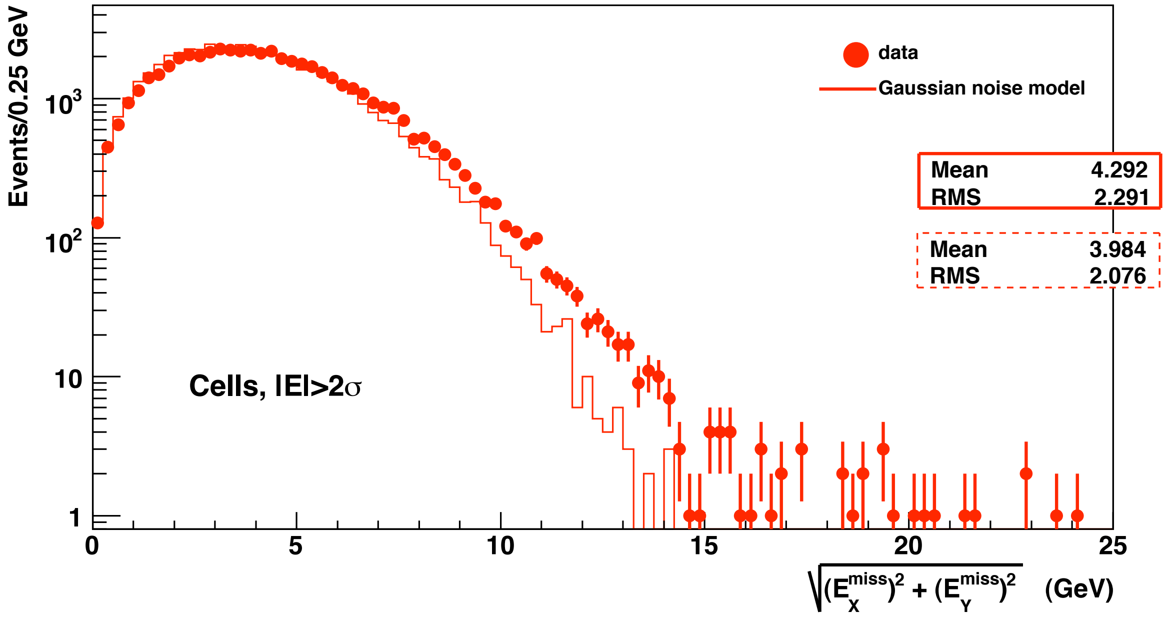


Figure 2.13: *$\not{E}_T$ distribution computed with $|E| > 2\sigma_{noise}$ noise suppression cut for random events of the reference run (91639). Red dots represent the data, whereas the solid line is the toy Monte-Carlo.*

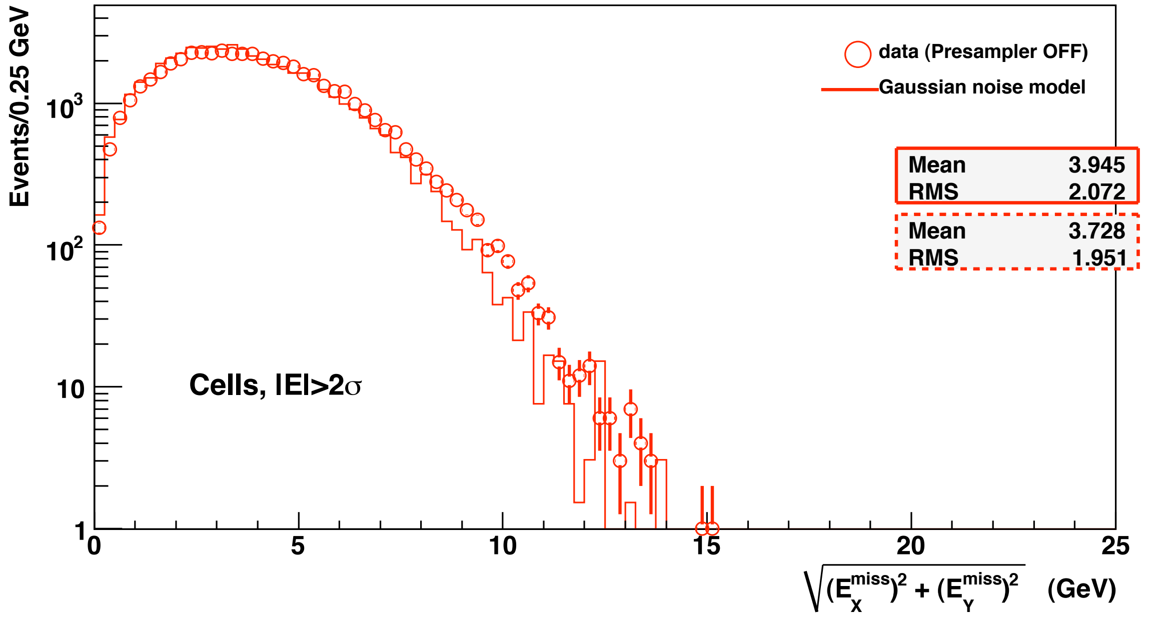


Figure 2.14: *$\not{E}_T$ distribution computed with $|E| > 2\sigma_{noise}$ noise suppression cut for random events of the reference run (91639) without the EM presampler. Red open dots represent the data, whereas the solid line is the toy Monte-Carlo.*

The source of this noise was found to be a faulty HV cable which has been replaced and therefore the problem was fixed.

2.7.2 $\not\!\!E_T$ resolution contributions

An important part of this analysis was also to understand the different calorimeter layers contributions to the final $\not\!\!E_T$ resolution (2.3 GeV when presampler is included). Figure 2.15 shows the E_X^{miss} resolution per calorimeter layer: full and open markers represent the data and the Gaussian noise model expectations respectively. It is worth noticing that the main contribution to the Gaussian width comes from the second layer of the EM barrel calorimeter and the HEC. This is a combined effect of the level of the noise in each cell (figures 1.15 and 1.16), the number of cells and the geometrical $sin\theta$ which is multiplied to the energy in order to project the value to transverse plane. These are the reasons for instance why HEC and FCal show so different contribution in the E_X^{miss} resolution while they represent the same level of noise (figure 1.15). Finally, the E_X^{miss}, E_Y^{miss} and $\sum E_T$ widths evolutions with time [38] are shown in figure 2.16, as well as the respectives per sub-detector. The E_X^{miss}, E_Y^{miss} Gaussian widths are constant with time inside ±3%, and it is shown that the pedestal correction (points represented with blue stars) do not influence the width behavior contrary to the mean values. For $\sum E_T$, a slow increase of the width, of the order of 10% is observed and is explained by the EM barrel width increase (figure 2.16 bottom, right).

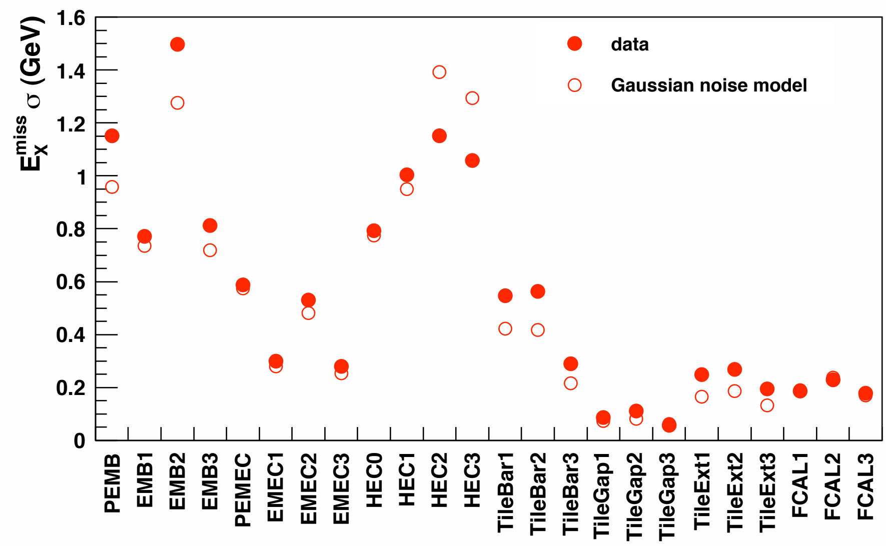


Figure 2.15: E_X^{miss} *Gaussian width* σ *as a function of the different layers of the calorimeter system for data (full dots) and toy Monte-Carlo (open symbols).*

2.8 Cluster based reconstruction method

In this section, another noise suppression method, based on topological clustering and presented in section 1.3.4.2 ($\not\!\!E_T$*-topo*) is used for $\not\!\!E_T$ reconstruction.

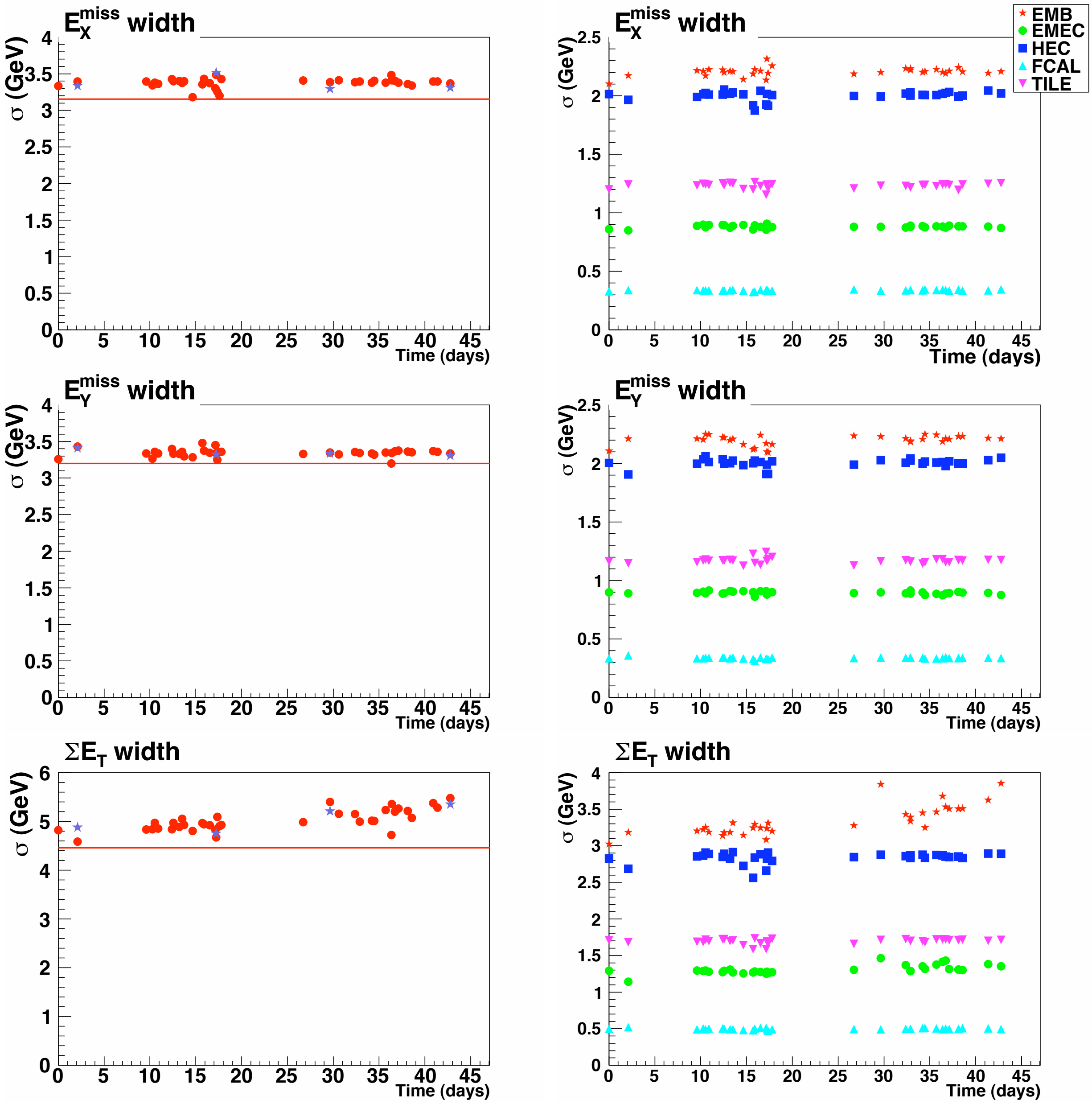


Figure 2.16: *Left:* E_X^{miss} *,* E_Y^{miss} *, and* $\sum E_T$ *Gaussian width* σ *computed for* $|E| > 2\sigma_{noise}$ *noise suppression cut as a function of time, for randomly triggered events. Solid line is the expected value and blue stars show the results obtained when recomputing the effective pedestal for some randomly triggered runs. Right: Same results per sub-detector of the calorimeter system [38].*

2.8.1 Performance of topological clustering

To more efficiently suppress the electronic noise, $3D$ topological clusters or topoclusters (see details in section 1.3.4) are formed around seeds satisfying $|E| > 4\sigma_{noise}$ (0.006% probability instead of 4.55% for $2\sigma_{noise}$ in the cell based method). A good check consists in measuring the number of cells satisfying $|E| > 4\sigma_{noise}$ for each sub-detector and compare it to the numbers derived from the Gaussian noise model. This is shown in figure 2.17 as a function of η and in table 2.4 per subdetector. Results from figure 2.17 and table 2.4 can be compared with those respective from $\not\!\!E_T$*-base* (figure 2.9, table 2.3). One could see that in $\not\!\!E_T$*-topo* method less cells are selected as expected, due to the tighter cut of topocluster seeds ($|E| > 4\sigma_{noise}$). Focusing on 2.17, a discrepancy between data and expectations in the central part of the detector is due to non Gaussian Tile noise behavior. The existence of these tails in Tile (figure 2.5), makes the possibility of finding a topocluster seed higher than the one extracted from a Gaussian model. This is the reason why in table 2.4 when focusing on Tile calorimeter, the Gaussian model predicts 0.3 cells/event with $|E| > 4\sigma_{noise}$ while 13.9 are found in data.

	EM Barrel	EM Endcap	Tile	HEC	FCal
# cells: $\lvert E\rvert > 4\sigma_{noise}$ / event	10.2 (6.9)	6.1 (4.0)	13.9 (0.3)	0.4 (0.3)	0.2 (0.2)

Table 2.4: *Average number of cells with* $|E| > 4\sigma_{noise}$ *per event for each subdetector in random events of reference run 91639. The expected numbers derived from the Gaussian noise model are in brackets.*

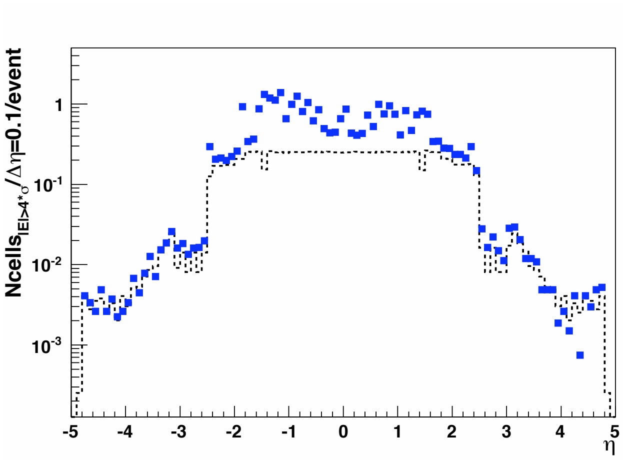


Figure 2.17: *Number of cells with* $|E| > 4\sigma_{noise}$ *normalized to one random event of the reference run, per bin of* $\Delta\eta = 0.1$*. Blue dots represent the data, whereas dashed line is the expected value with the database noise.*

2.8.2 $\not\!\!E_T$ reconstruction

Similarly to section 2.7, $\not\!\!E_T$ variables are calculated using the $\not\!\!E_T$*-topo* noise supression. Figure 2.18 shows $\not\!\!E_T$ variables computed with topocluster cells. Because approximatively 10

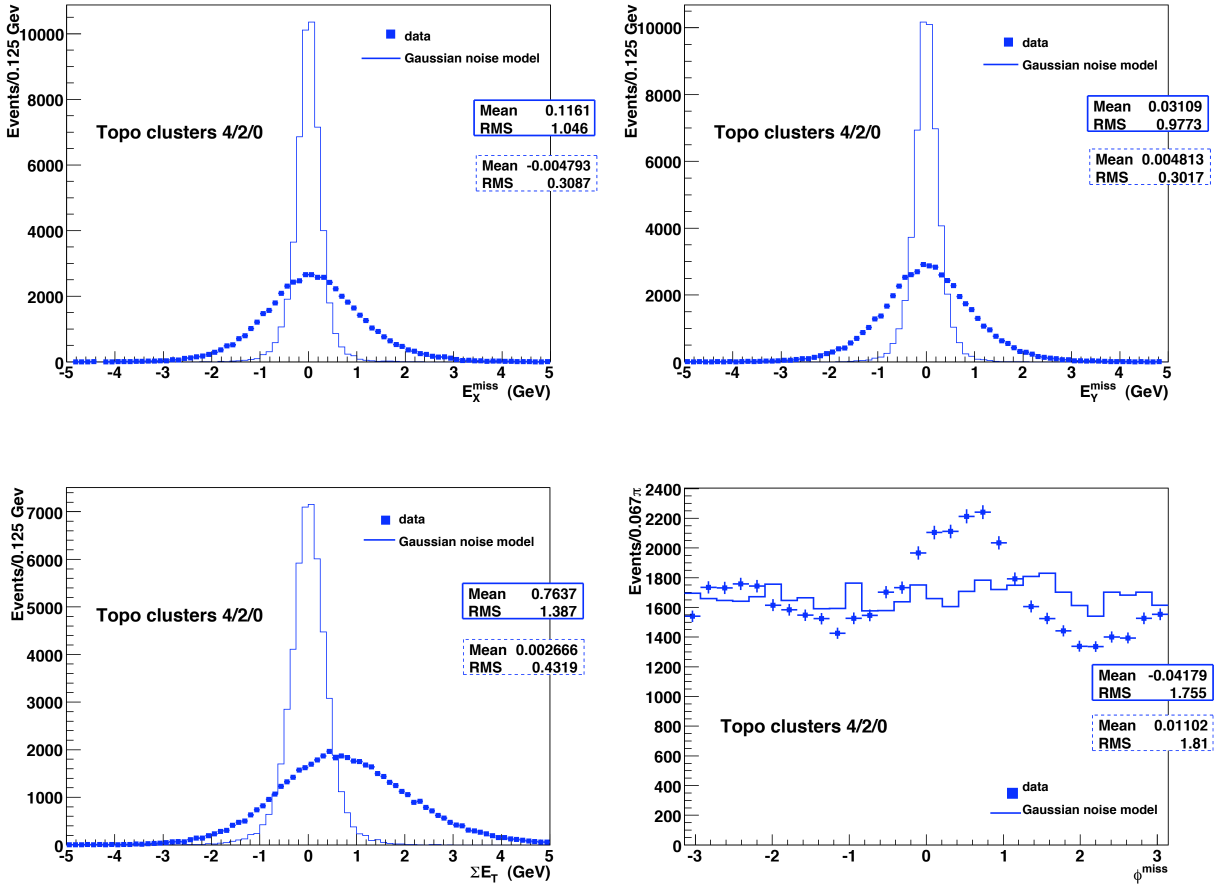


Figure 2.18: *E_X^{miss} , E_Y^{miss} , $\sum E_T$ and φ^{miss} distributions computed with topocluster cells for random events of the reference run (91639). Blue dots represent the data, whereas the solid line is the toy Monte-Carlo.*

times less cells are selected for $\not\!\!E_T$ variables computation compared to cell-based results, the expected Gaussian width is divided by $\sqrt{10} \sim 3$. Nevertheless, this is far from the 0.3, 0.3 and 0.43 GeV expected by the Gaussian noise model for the width of E_X^{miss} , E_Y^{miss} and $\sum E_T$, respectively. Furthermore, even though mean values of E_X^{miss} and E_Y^{miss} are compatible with the expected zero, a slight shift of 0.76 GeV is observed in $\sum E_T$ distribution. The reason for the $\sum E_T$ shift is explained in previous section (pedestal shift & iteration threshold equals to 4 instead of 5) but since $\not\!\!E_T$-*topo* demands a tighter noise suppresion and less cells are selected, the effect of bias is less visible (0.76 GeV shift for $\not\!\!E_T$-*topo* compared to 1.74 GeV shift for $\not\!\!E_T$-*base*). Another disagreement is also present in φ^{miss} where an unexpected peak is observed at around φ^{miss} ~0.8. As it will be shown later this featureis linked to the Tile non Gaussian tails.

Computing these distributions with topoclusters containing only LAr cells (figure 2.19), a fair agreement between the Gaussian width extracted from the toy Monte Carlo and data is observed, showing that the Tile noise non Gaussian behavior explains the discrepancy seen in Figure 2.18.

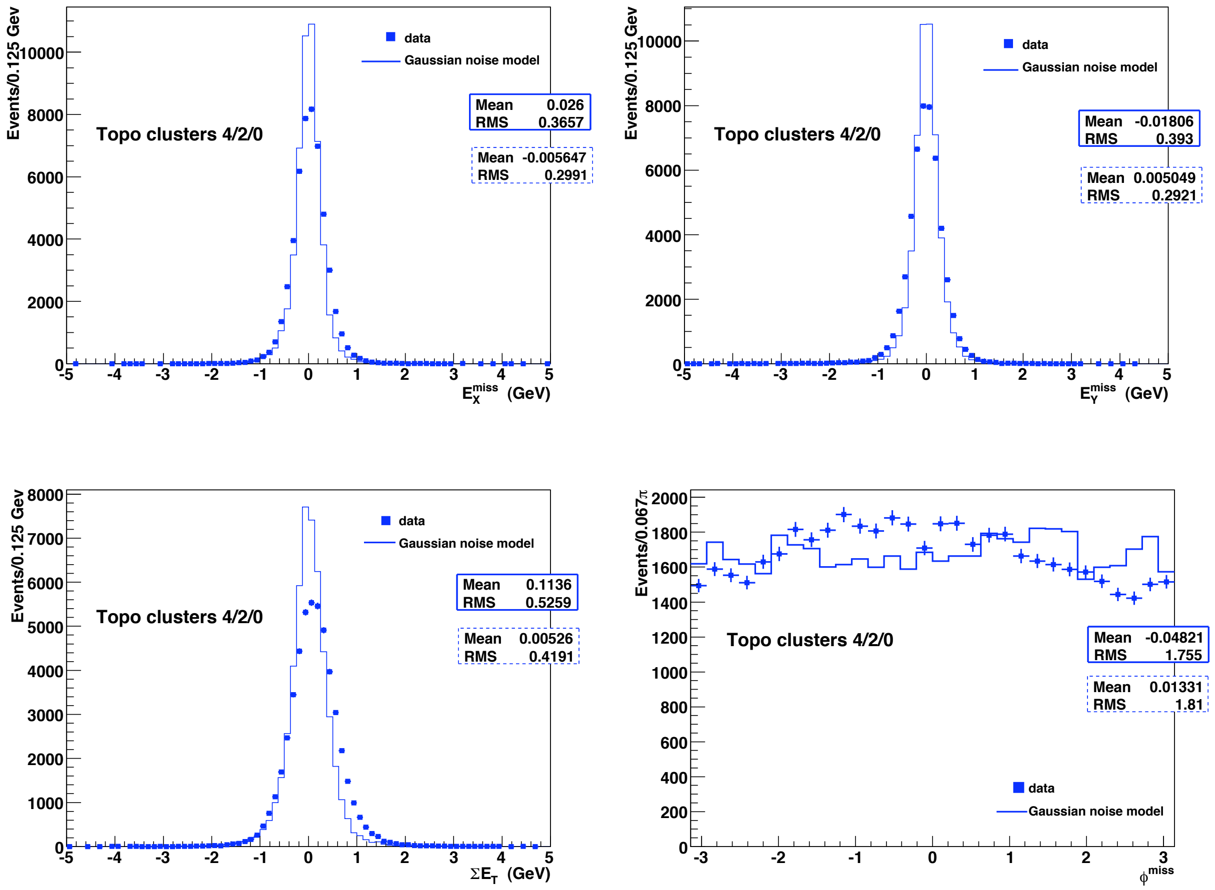


Figure 2.19: E_X^{miss} , E_Y^{miss} , $\sum E_T$ *and* φ^{miss} *distributions computed with topocluster cells taken only from LAr sub-detectors for random events of the reference run (91639). Blue dots represent the data, whereas the solid line is the toy Monte-Carlo [38].*

As only small deviations from 0 of the mean values of the distributions are observed, the time stability of these results[38] is shown as the deviation of the Gaussian mean from its average value over the time (Figure 2.20). Variations are observed in $\sum E_T$ time evolution around its mean value, but they are considerably smaller than the the width of the distribution ($\sim$ 1.4 GeV). A reasonable stability of all $\not\!\!E_T$ variables mean values inside ± 0.2 GeV over 1.5 month of data taking is observed.

Figure 2.21 shows the $\not\!\!E_T$ distribution of the reference run compared to the Gaussian noise model. Because of the Tile noise non Gaussian behavior in the data, no matching with expectations is observed and furthermore similarly to $\not\!\!E_T$-*base* case, tails are observed after 7 GeV. The tail is due to the coherent noise in presampler, a feature presented in previous section, and when removing presampler contribution in $\not\!\!E_T$ measurement the lower plot of figure 2.21 is obtained which excludes the tail feature.

To validate the assumption that the Tile calorimeter non Gaussian behavior is responsible for the disagreement of data and toy Monte Carlo, as for E_X^{miss} , E_Y^{miss} , φ^{miss} and $\sum E_T$, the $\not\!\!E_T$ plot is reproduced incorporating only the LAr cells (figure 2.22) with and without the presampler to avoid the $\not\!\!E_T$ tails feature. In this case a reasonable agreement is observed for the bulk of the distribution. Moreover, similarly to $\not\!\!E_T$-*base* presented in previous section, unexpected behavior of tail is corrected when presampler is removed. In summary, when considering only cells from LAr sub-detectors without the presampler ($\sim$ 173 000 cells) to build topoclusters, a good agreement is obtained between data and simulation (figure 2.22 lower). As mentioned, before this presampler feature, which causes the tail in $\not\!\!E_T$, was understood and solved in hardware level. It is no longer present since the summer of 2009.

2.8.3 $\not\!\!E_T$ resolution contributions

The relative time variations[38] of E_X^{miss} , E_Y^{miss} and $\sum E_T$ Gaussian widths are shown in figure 2.23. The widths show a remarkable stability remaining constant with time inside ±0.1 GeV range. Some variations observed mainly to the $\sum E_T$ width evolution with time can be neglected since it is rather small compared to the time average value of $\sum E_T$ width (1.372 GeV).

Figure 2.24 shows the E_X^{miss} width contribution per calorimeter layer (to be compared with the respective of $\not\!\!E_T$-*base* , figure 2.15).

As expected, due to fewer selected cells contributing to the $\not\!\!E_T$-*topo* calculation, all E_X^{miss} widths for LAr sub-detectors are smaller compared to the respective ones of figure 2.15 for $\not\!\!E_T$-*base* with the exception of Tile sub-detectors. A very visible excess is observed in the two first samplings of Tile barrel, a fact that proves that Tile gives the dominant contribution for $\not\!\!E_T$ resolution when using the topoclusters noise suppression even though LAr sub-detectors contains about 173k cells compared to 5k cells of Tile sub-detectors. The high difference between Tile data and Tile toy Monte Carlo is explained by the non Gaussian tails.

As a summary, one could say that a Gaussian based approach such as topoclusters, shows

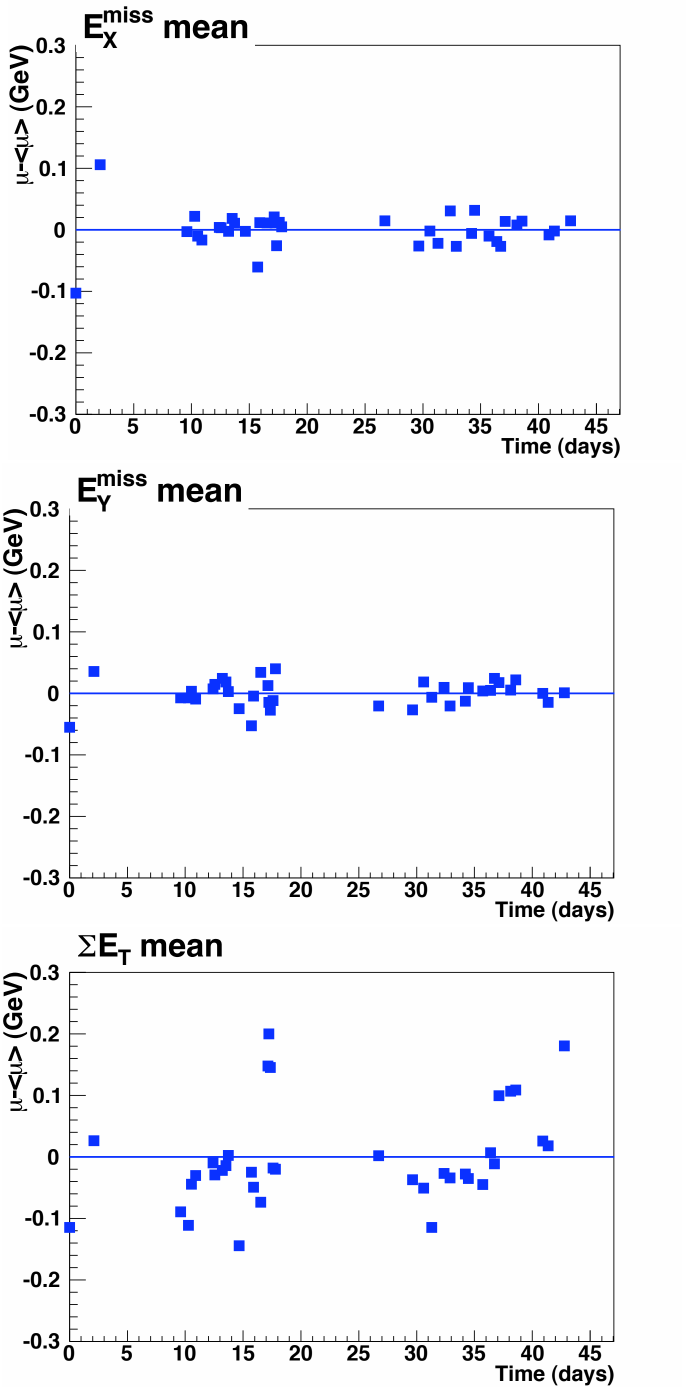


Figure 2.20: *Relative* E_X^{miss} *,* E_Y^{miss} *, and* $\sum E_T$ *Gaussian mean* μ *offset computed with topocluster cells for random events as a function of time. To be noted: <*E_X^{miss} *>=(0.103 ± 0.005) GeV , <*E_Y^{miss} *>=(0.023 ± 0.004) GeV , <*$\sum E_T$ *>=(0.780 ± 0.014) GeV.*

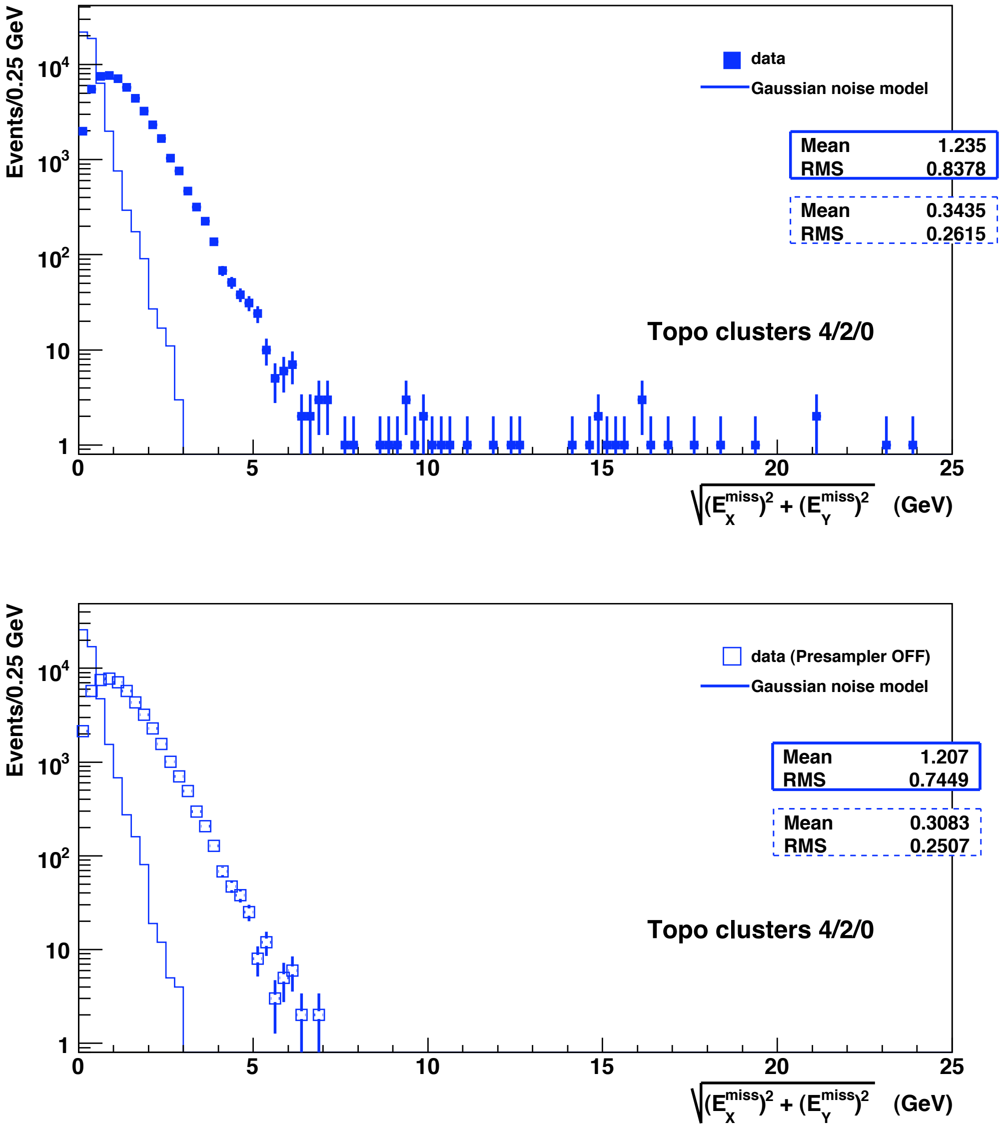


Figure 2.21: *Upper: $\not{E}_T$ distribution computed with topocluster cells for random events of the reference run (91639). Blue dots represent the data, whereas the solid line is the toy Monte-Carlo. Lower: same distributions computed without the EM presampler.*

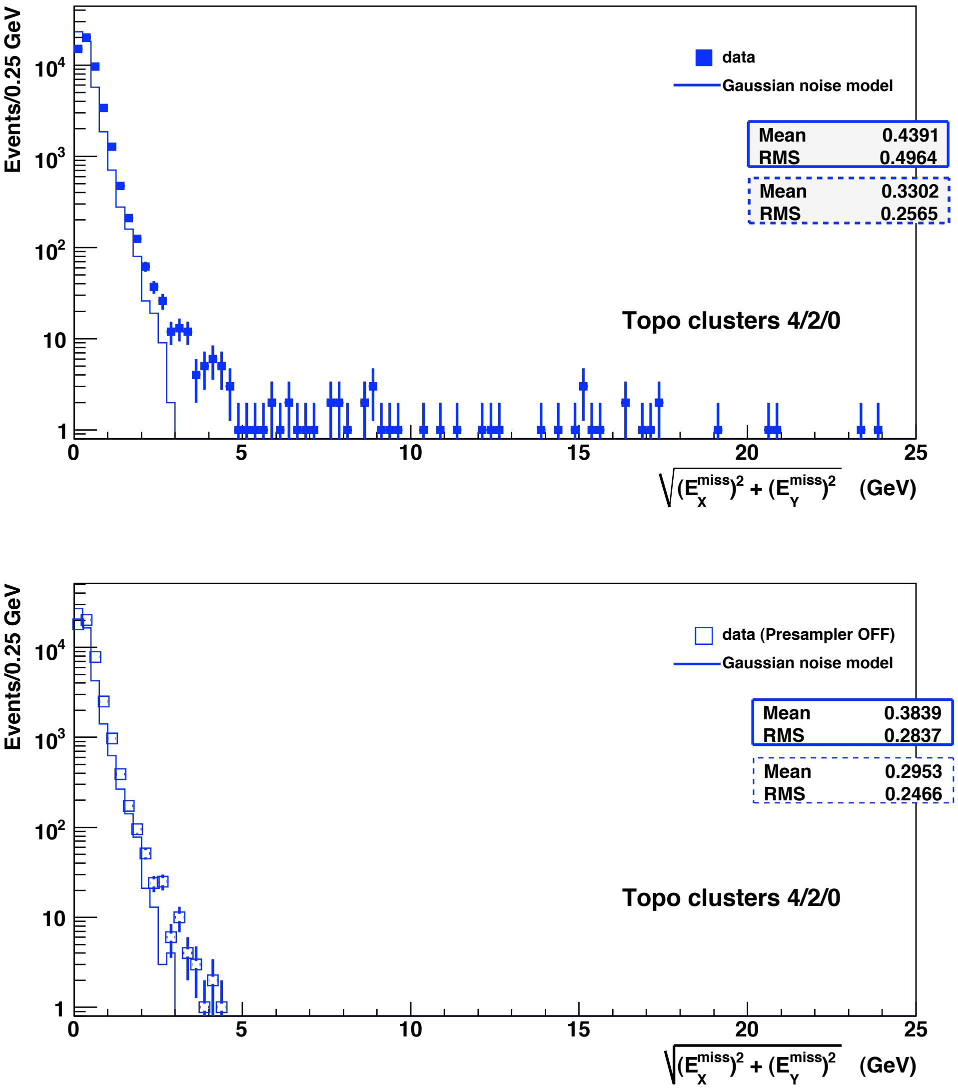


Figure 2.22: *Upper: $\not\!\!E_T$ distribution computed with topocluster containing only LAr cells for random events of the reference run (91639). Blue dots represent the data, whereas the solid line is the toy Monte-Carlo. Lower: same distributions computed without the EM presampler.*

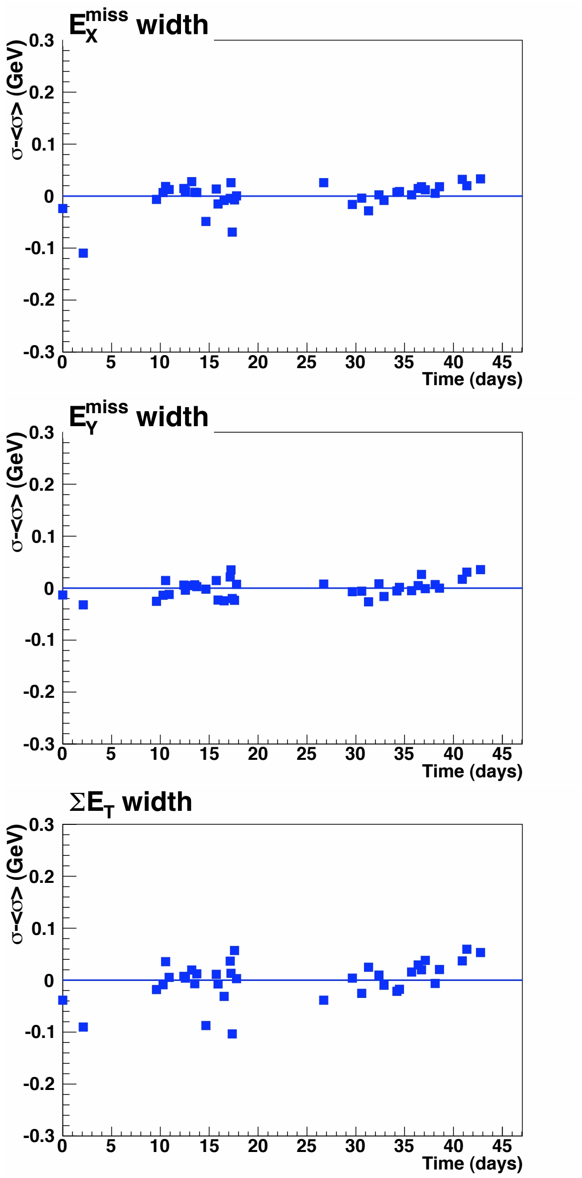


Figure 2.23: *Relative* E_X^{miss} *,* E_Y^{miss} *, and* $\sum E_T$ *Gaussian width* σ *offset computed with topocluster cells for random events as a function of time. To be noted:* $<\sigma(E_X^{miss})>$*=(1.000* $\pm$ *0.005) GeV ,* $<\sigma(E_Y^{miss})>$*=(0.932* $\pm$ *0.003) GeV ,* $<\sigma(\sum E_T)>$*=(1.372* $\pm$ *0.006) GeV.*

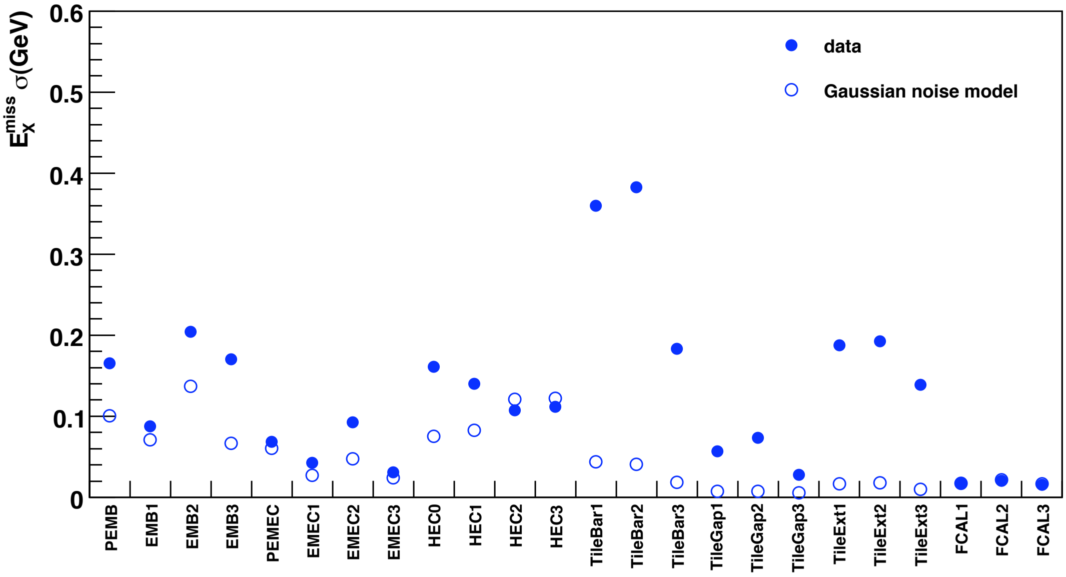


Figure 2.24: E_X^{miss} *Gaussian width* σ *as a function of the different layers of the calorimeter system for data (full dots) and toy Monte-Carlo (open symbols) for* $\not{E}_T$*-topo* .

a poor performance in the calorimeter $\not{E}_T$ due to the non Gaussian tails of Tile. Cell-by-cell double Gaussian description is being considered by the Tile community. An attempt of a similar approach is presented in section 2.9.

2.8.4 Tile non Gaussian behavior affects LAr

In this section possible influence of the non Gaussian Tile behavior to clusters found in LAr is studied. The idea is simple: since a higher number of seeded topoclusters in Tile is expected with respect to a Gaussian prediction, then it is likely that some of those topoclusters are being propagated in neighbor LAr layers affecting finally the $\not{E}_T$*-topo* in LAr, since Tile calorimeter is adjoint to LAr (see figure 1.6).

To test this, a 900 events sample of random triggers of single beam run 88128 (section 2.2.3) was reprocessed, locally replacing only the energy of Tile cells with the output of the CaloCellRandomiser. As a result, the deduced sample was composed by the measured noise data for LAr sub-detectors and by a Gaussian, un-corelated noise for Tile sub-detectors. The plots of E_X^{miss} distributions are presented for LAr EM barrel, LAr EM end-cap and LAr hadronic end-cap in figure 2.25[1].

---

[1] $\not{E}_T$ in one sub-detector system means the somme of the energy of selected cells which are contained in the studied sub-detector system. In the case of topoclusters, it is likely that a cluster is seeded in one layer and it is propagated in one or more neighbor layers

As it is shown, a difference is observed between the data and the results obtained with the randomized Tile. A better resolution is achieved when energy in Tile follows a Gaussian model, a proof that Tile non Gaussian behavior affects the neigbor LAr layers. For a more quantitative representation the widths of the distributions are presented in table 2.5.

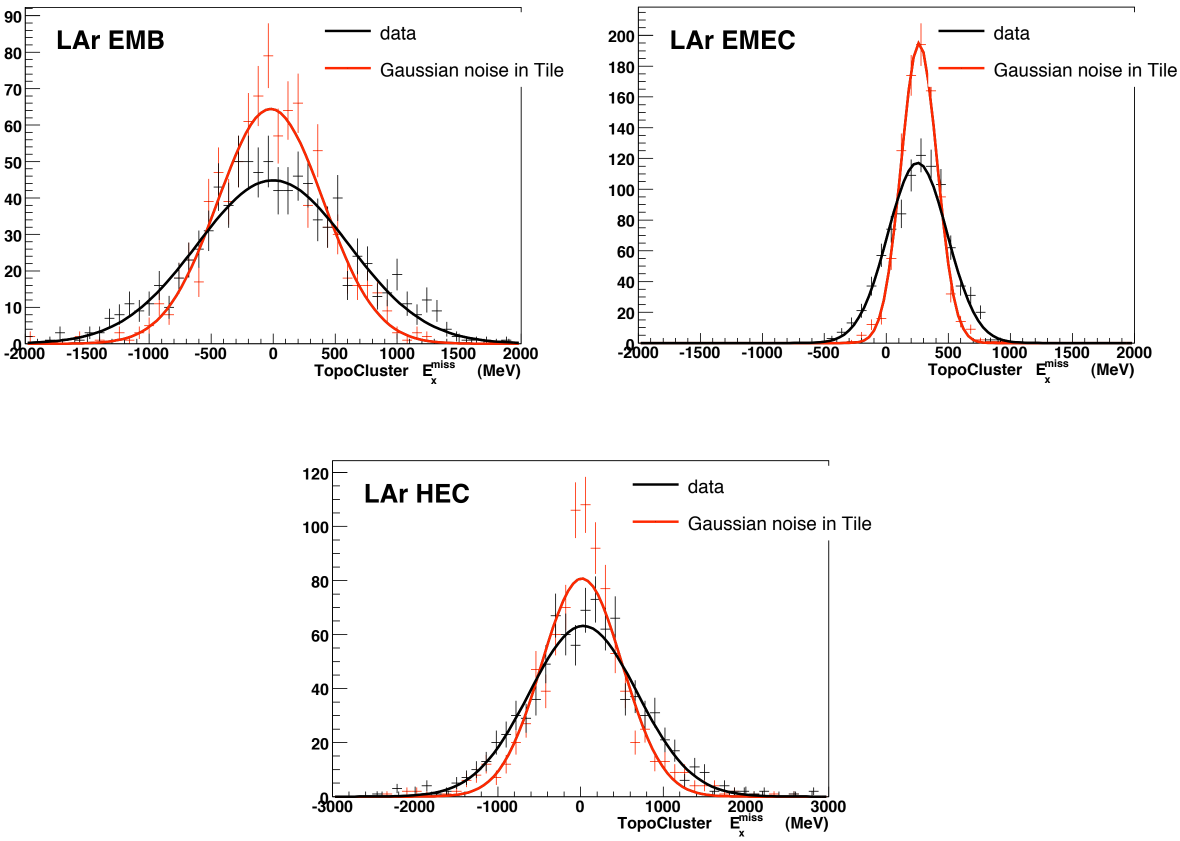


Figure 2.25: E_X^{miss} *distribution of random stream of single beam event 88128 (black line), superimposed with the repsective* E_X^{miss} *distribution where Tile energy is Gaussianly randomized (red line) for LAr EMB, LAr EMEC and LAr HEC.*

A difference which varies from ~25% (HEC) to ~40% (EMEC) is found in the width of LAr E_X^{miss} distributions. Therefore when moving to the whole calorimeter respective plot figure (2.26) the difference in resolution is not only because of changing of the energy in Tile but also due to the affected LAr sub-detectors

LAr forward calorimeter results are not presented since there is a perfect agreement between data and randomised Tile, a clear indication that non Gaussian Tile does not affect this sub-detector, a fact explained by geometrical reasons, since several LAr calorimeters layers are found between Tile and FCal, therefore a seed topocluster in Tile will not reach the FCal (see figure 1.11).

An impact of Tile in LAr when calculating $\not{E}_T$*-topo* is observed and therefore a special care should be given to this aspect when a new paremetrisation for Tile noise will be proposed.

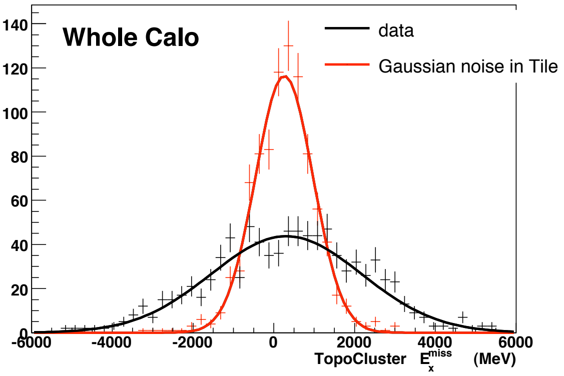


Figure 2.26: *E_X^{miss} distribution of random stream of single beam event 88128 (black line), superimposed with the repsective E_X^{miss} distribution where Tile energy is Gaussianly randomized (red line) for whole calorimeter.*

Not only a model should take into consideration a new noise description in Tile but also its influence on LAr clusters.

	data E_X^{miss} width (GeV)	Gaussian noise in Tile E_X^{miss} width (GeV)
LAr EM barrel	0.62	0.43
LAr EM end-cap	0.24	0.14
LAr hadronic end-cap	0.65	0.49
Whole Calo (LAr & Tile)	1.91	0.72

Table 2.5: *Width values corresponding to E_X^{miss} distributions of figure 2.25.*

2.9 Tile noise parameterization using a double Gaussian PDF

As described in previous section, a toy Monte Carlo was developed in order to validate results observed in random triggers. Following the assumption that noise follows a simple Gaussian distribution, this toy Monte Carlo was developed in a basis of normal distribution *PDF*. It is already shown that the noise could be modeled through this model in LAr sub-detectors but it is not the case for Tile. Figure 2.27 illustrates the square root of the variance (σ) of a simple Gaussian fit for noise in the different Tile sub-detectors as a function of eta. One would expect points which will slightly fluctuate around one in the scenario that noise in Tile would be Gaussian, but divergences from one are observed which in some cases reach almost 20%.

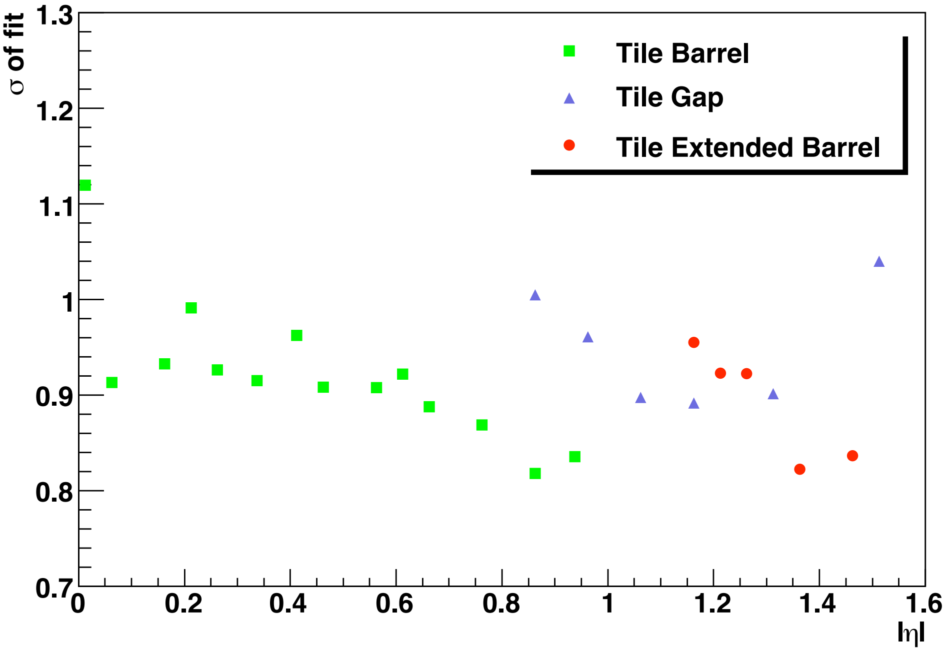


Figure 2.27: *The σ value of a simple Gaussian fit for the 3 different sub-detectors of Tile as a function of $|\eta|$.*

Following the failure of describing noise in Tile with a simple Gaussian PDF, and also following the Tile non-Gaussian tails (figure 2.5) observed, a double Gaussian *PDF* is tested. A first step is to try to fit noise in Tile with this new *PDF*. Results are shown in figure 2.28 where the two *PDF's*: simple Gaussian (red line) and double Gaussian (blue line) are used to fit random triggered data in Tile. Results are presented for each Tile sub-detector obtaining information from all Tile cells in a 400 events sample of reference run. As illustrated in figure 2.28 a double Gaussian *PDF* provides a more adequate description of data compared to the simple Gaussian one but nevertheless, some points starting from $|E/\sigma| > 4$ form tails which are not contained to the fit, resulting to a final result which is not sufficient in order to obtain a good parametrisation.

A more refined fit using a double Gaussian *PDF* was then used. The goal was to construct a model which would take into consideration the fluctuating η behavior of noise in Tile (figure 1.16). There the double Gaussian parametrisation was applied for each $|\eta|$ bin of 0.1. Some characteristic examples of these fits are shown in figure 2.29. To check the quality of this fit the plot of figure 2.30 was produced. The integral:

$$S = 1 - \int_{-4}^{4} \frac{E}{\sigma}\, dE \quad , \tag{2.15}$$

which implies the existence of a topocluster seed is calculated and plotted as a function of the different $|\eta|$ bins for data (blue triangle points) and for double gaussian fit (red circle points). In the calculation of data integral over/under-flows are taken into consideration. In a reasonable description, one would expect the two different quantities to give similar results.

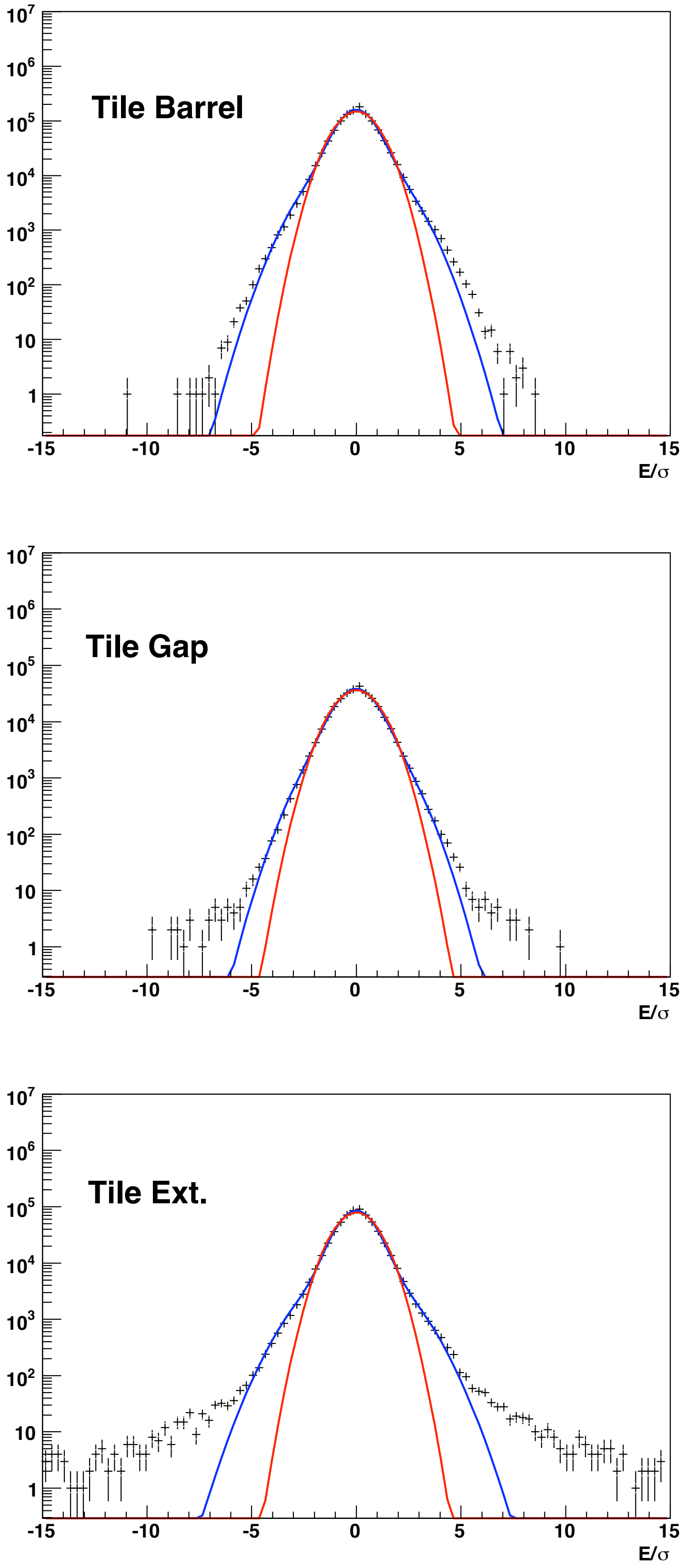


Figure 2.28: *Fitting energy distribution in 3 different Tile sub detectors with a simple Gaussian (red line) and a double Gaussian (blue line) distribution. Scale in y-axis is logarithmic.*

Looking in more details 2.30, it is obvious that the two quantities do not give similar results for all $|\eta|$ bins and as expected, the shape of η dependence follows the one described in figure 1.16.

Nevertheless, a new parametrisation was tested using as an input the fit values of these double Gaussian distributions fitted in 16 different $|\eta|$ bins. Results are shown in figure 2.31 in a sample of 900 events of the reference run. The RMS values of those plots are shown in table 2.6.

The number of clusters is shown in table 2.7. Focusing on the results presented in table 2.6, an improvement of $\sim$ 20% is observed in the RMS for Tile when comparing the toy Monte Carlo of simple and double Gaussian *PDF*, in 16 η bins. A factor of 12 which was the difference between real data in Tile and the simple Gaussian toy Monte Carlo, it is reduced to $\sim$ 2 when comparing real data with the double Gaussian model. Similar are the results, when examining RMS values for the whole Calo: RMS of the double Gaussian model is increased 54% with respect to the RMS simple Gaussian model and the difference between real data and Gaussian parameterizations is reduced to a factor of 1.8 for the double Gaussian with respect to $\sim$3 which is the case for the simple Gaussian model.

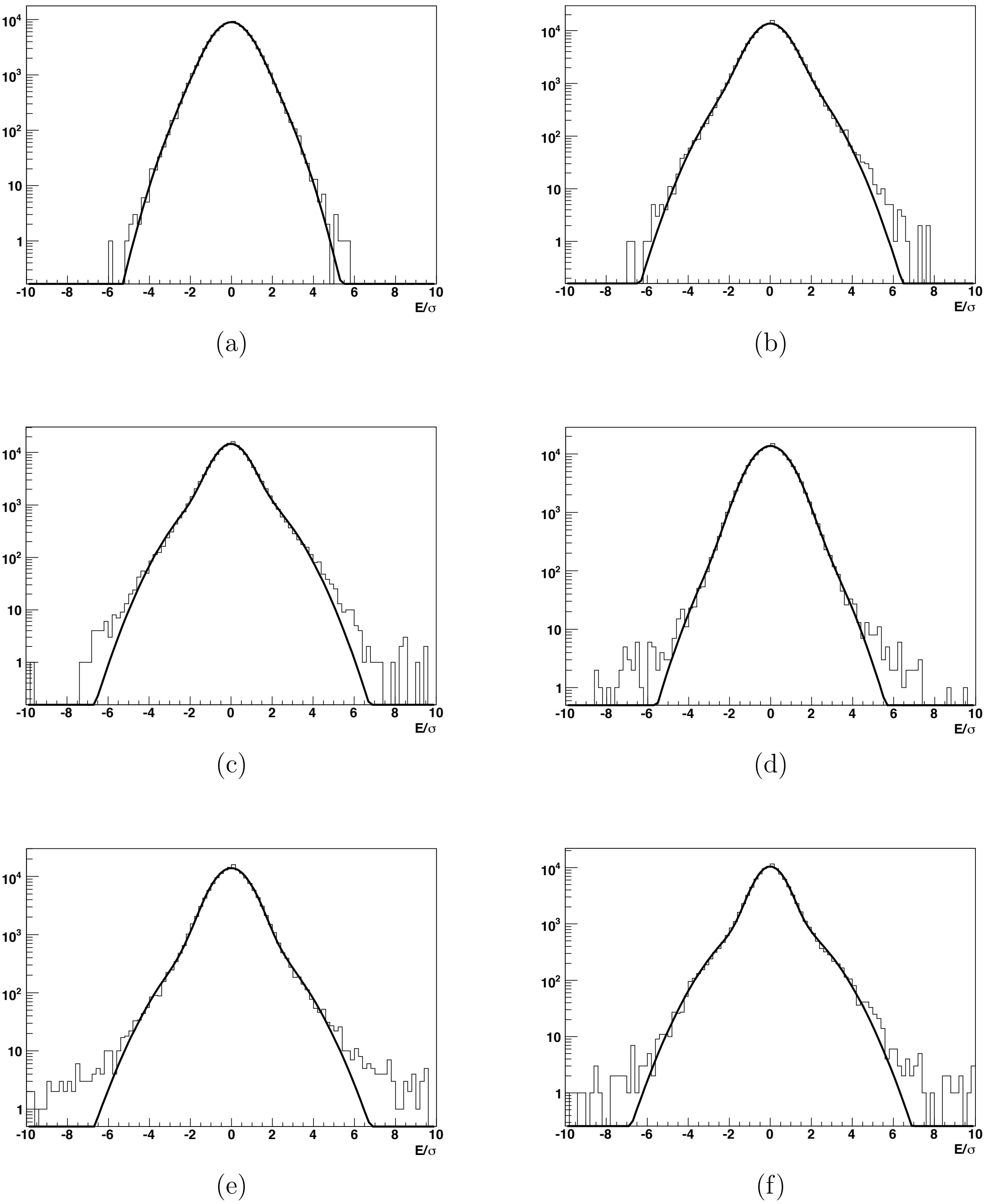
	Tile	Whole Calo
Data histogram RMS	0.88 GeV	1.05 GeV
Simple Gaussian model RMS	0.07 GeV	0.31 GeV
Double Gaussian model RMS	0.37 GeV	0.57 GeV

Table 2.6: *Resolution of E_X^{miss} distribution for data, toy monte carlo of simple Gaussian and toy monte carlo of double Gaussian (RMS of plots in figure 2.31).*

	# clusters / event
Data	30.8
Simple Gaussian model	11.8
Double Gaussian model	20.1

Table 2.7: *Number of celles with $|E| > 4{\cdot}\sigma_{noise}$ (topoclusters seeds) for data, toy monte carlo of simple Gaussian and toy monte carlo of double Gaussian.*

The correction achieved is also reflected in the number of clusters per event observed (table 2.7). Starting from 11.8 clusters per event, a result obtained by the simple Gaussian model, 20.1 events per cluster are observed when applying the double Gaussian model. An improvement of almost a factor of 2 which is not still sufficient to explain the 30.8 clusters per event observed in data, is finally accomplished. As mentioned in previous section, several studies are being performed within the Tile community and the goal is to find and adequate parametrization of noise in order to avoid such problems observed in topoclusters and to fully



(a) (b) (c) (d) (e) (f)

Figure 2.29: *Fitting noise Energy distribution in Tile using a double Gaussian pdf for different bins of* $|\eta|$. *(a):* $0.1 < |\eta| < 0.2$, *(b):* $0.4 < |\eta| < 0.5$, *(c):* $0.8 < |\eta| < 0.9$, *(d):* $1.1 < |\eta| < 1.2$, *(e):* $1.2 < |\eta| < 1.3$, *(f):* $1.4 < |\eta| < 1.5$.

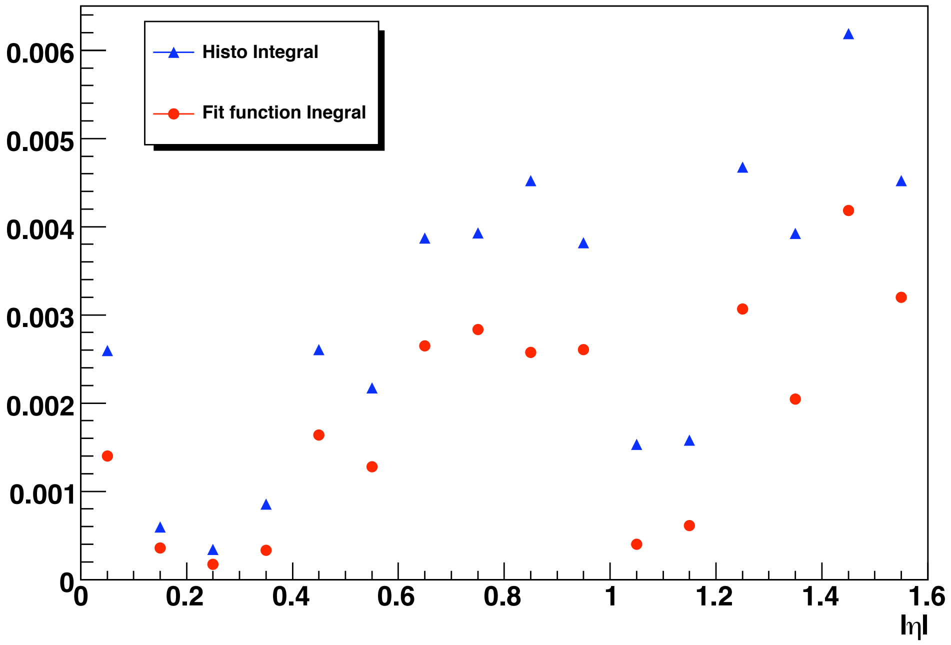


Figure 2.30: *Average number of cells per event with* $|E| > 4 \cdot \sigma_{noise}$ *(topo-cluster seeds) for data (blue triangles) and double Gaussian model (red circles) as a function of eta.*

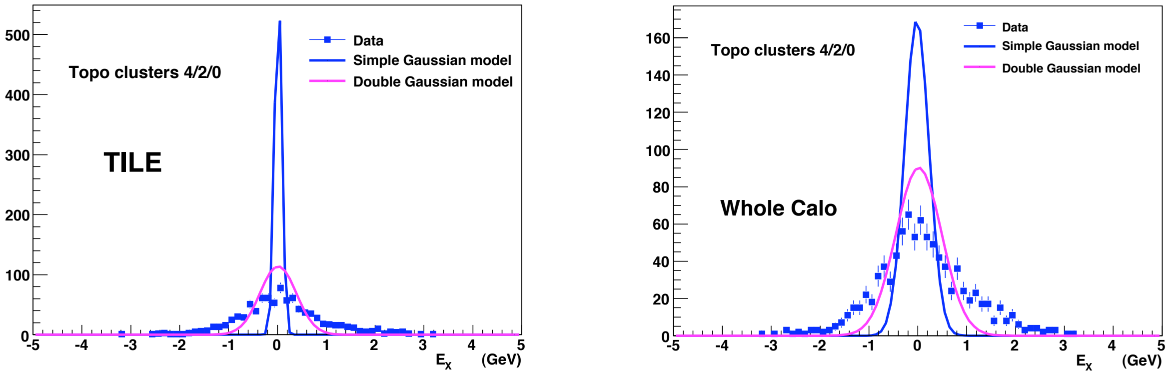


Figure 2.31: E_X^{miss} *distribution for TILE calorimeter (left) and whole calorimeter (right), superposed with fits of simple Gaussian and double Gaussian model.*

understand the non-Gaussian tails. One improvement which is being tested is a cell-by-cell fit given that the effect is not ϕ symmetrical.

Apart from understanding the noise in Tile, which will allow a realistic and accurate description of it, the way topoclusters are taking into account this noise, is an aspect that should be re-addressed. As described in section 1.3.4, topoclusters noise suppression method, select cells with the assumption that their energy is $|E| > n\sigma_{noise}$. This assumption implies that noise follows a symmetrical Gaussian distribution and therefore, a new non-Gaussia behavior of potential noise should be taken into consideration by topoclusters and perform a respective cut in order to select cells. For the double Gaussian model, a possible solution would be that the topoclusters choose an “effective” σ_{noise} which would be calculated from the two σ_{noise} values, which characterize the double Gaussian distribution.

2.10 Effect of finding correct time phase in LAr reconstruction for cosmics

To distinguish cells with signal and cells without, a specific noise threshold was initially chosen of n_{iter}=3 (eq. 2.5, see section 2.3). Results concerning $\sum E_T$ variable, obtained with this threshold are illustrated in figures 2.32 for all different LAr sub-detectors and in figure 2.33 for the whole ATLAS calorimeter, superimposed with the respective distribution when altering the threshold to n_{iter}=4. n_{iter}=3 was the default value before of the data reprocessing at the period of Christmas 2008 (see section 2.2), while it was changed to n_{iter}=4 for the reproccessing. The reason of this modification is shown in figures 2.32 and 2.33, where one could say that a significant shift, which is observed for the n_{iter}=3 case (before reprocessing), it is corrected when changing this threshold to n_{iter}=4 (after reprocessing). The mean values of the $\sum E_T$ distributions are contained in table 2.8. The offset is mainly significant in LAr EMB and EMEC where shifts of 11.4 GeV and 3.3 GeV are respectively observed. An overall shift of 17.8 GeV (Figure 2.33) is corrected. The noise suppression method used was the $\not{E}_T$*-base* since higher number of cells are selected compared to $\not{E}_T$*-topo* and thus the shift effect is more visible. After this study, the later configuration of 4σ was selected to be the default in LAr community, beginning from Christmas 2008 reprocessing.

	EM Barrel	EM Endcap	HEC	FCal	Whole Calo
$<\sum E_T>$ for n_{iter}=3	11.2 GeV	3.2 GeV	1.8 GeV	0.8 GeV	17.9 GeV
$<\sum E_T>$ for n_{iter}=4	0.9 GeV	0.0 GeV	-0.3 GeV	0.0 GeV	1.7 GeV

Table 2.8: *Mean values of the* $\sum E_T$ *distributions for different LAr sub-detectos and whole calorimeter.* $\not{E}_T$*-base is the noise suppression method used and results refer to the reference run.*

The remaining shift of the 1.7 GeV is explained in section 2.7.1 and as shown, is due to a pedestal shift, combined with a +2 GeV bias resulting from the 4σ threshold. Changing to 5σ (figure 2.12), the +2 GeV bias is corrected.

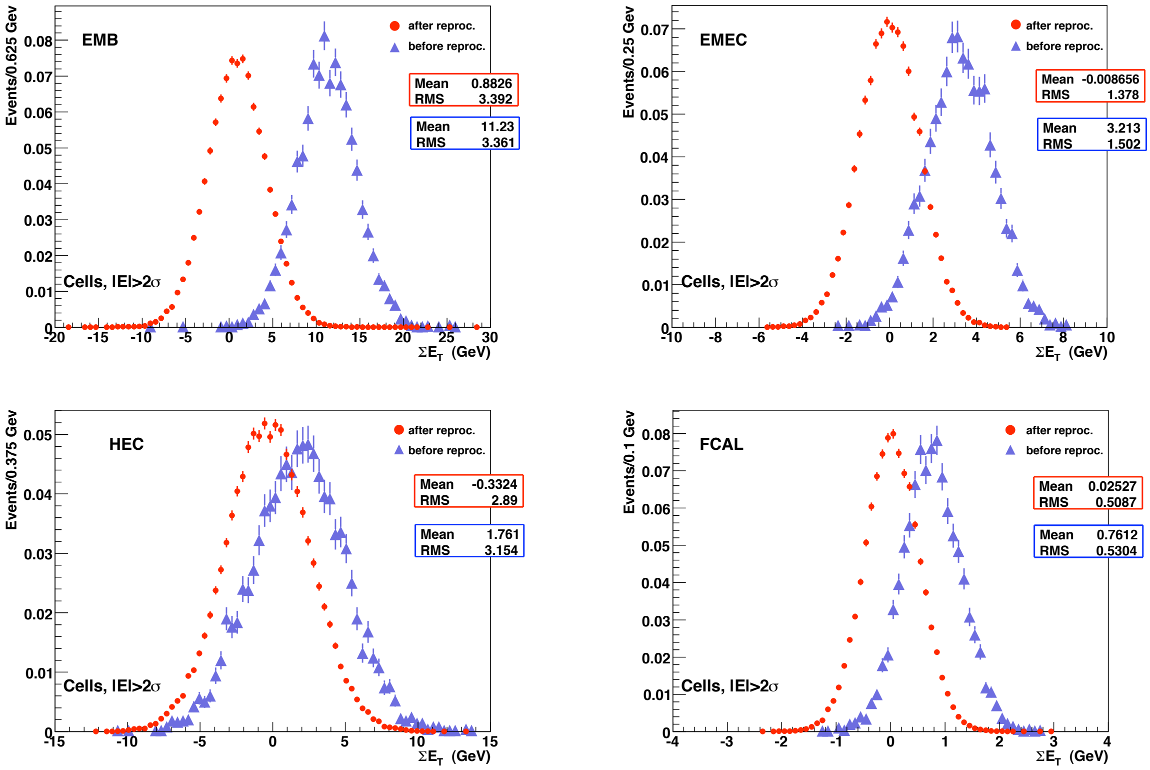


Figure 2.32: $\sum E_T$ *distributions for LAr sub-detectors computed with cell based method in random triggers of run 91639. The red circles correspond to LAr reconstruction after reprocessing (4*σ *threshold), while blue triangles correspond to LAr reconstruction before reprocessing (3*σ *threshold). Number of underflows and overflows is indicated when non zero.*

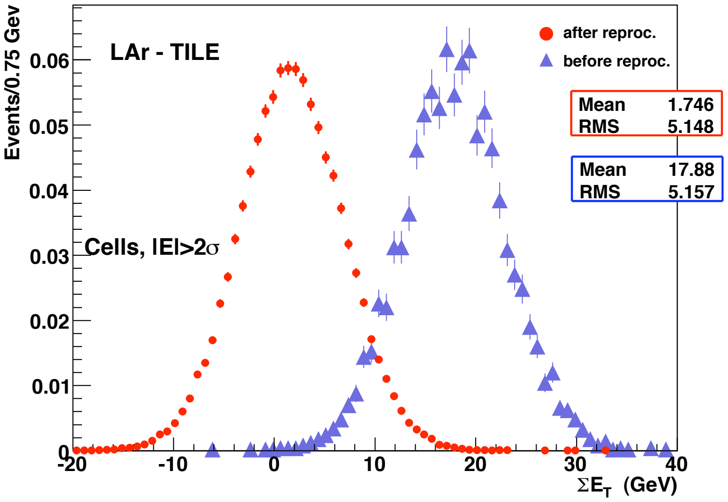


Figure 2.33: $\sum E_T$ *distribution for ATLAS calorimeter, computed with cell based method in random triggers of run 91639. The red circles correspond to LAr reconstruction after reprocessing (4*σ *threshold), while blue triangles correspond to LAr reconstruction before reprocessing (3*σ *threshold).*

During normal data taking, pedestal shift will be be measured in a daily base and no iterative procedure will be applied to find the correct time phase. The correct time will be easily extracted since time phase will be synchronized with the known time of LHC bunch crossing. Therefore, these problematic features will be absent when dealing with the collision data.

2.11 Problematic cells in ATLAS calorimeters

One difficulty in ATLAS calorimeters, is to control the quality of the signal reconstruction of the 180k cells. For the analysis of the runs taken in September-October 2008, several front-end boards were not working resulting in a loss of ∼ 1800 cells (1% of the total number of cells). Moreover, 1875 problematic cells were tagged per sub-detector, as shown in Figure 2.34 [38]. Depending on the problems spotted, different "cell pathologies" are defined in ATLAS calorimetry systems, as following [41, 42, 43, 44]:

LAr

- **deadReadout:** dead channel at the readout level (e.g. at the FEB level)
- **deadPhys:** dead channel at the detector level
- **deadCalib:** channel, which calibration pulse is significantly distorted, but the channel is otherwise OK for real data
- **distorted:** grab-bag of assorted wave distortions (amplitude, width, shape etc...) but not so bad that the channel can be considered as dead
- **lowNoiseHG(MG,LG):** a noisy channel in high gain (medium gain, low gain) which is more than 5 standard deviations from expected noise, but not so large that this channel should be masked.
- **highNoiseHG(MG,LG):** a noisy channel in high gain (medium gain, low gain) which is quite significant and one should consider masking it (more than 10 standard deviations from expected noise).
- **unstable:** features of these channels are unstable with respect to time.
- **short:** channel has a short
- **peculiarCalibrationLine:** a calibration which has a particular calibration-related problem (leaks signal in other channels) only in low gain.
- **sporadicBurstNoise:** sporadic bursts of noise seen during cosmic runs.
- **missingFEB:** channels for which a FEB was missing in the readout

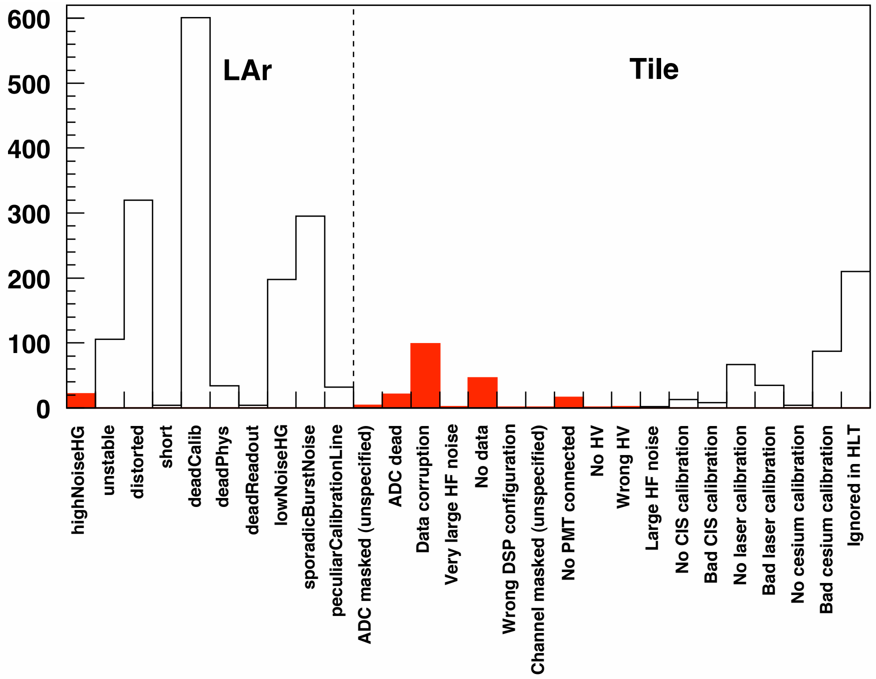


Figure 2.34: *Problematic cell flag categories for Liquid Argon sub-detectors (10 first columns) and Tile (18 last columns). It should be noticed that one cell can enter several categories. Filled bins indicate the categories masked to compute jet and* $\not\!\!E_T$. *Flags with no bin filled are not shown [38].*

Tile

- **ADC dead:** Analog to Digital converter is dead, only zeros or pedestals gets out of the ADC.
- **Data corruption:** Digital data is corrupted (CRC errors, bcid errors, DMU memory parity errors, etc. have been detected), but the fraction of affected events is low and it doesn't justify the DB masking.
- **Very large HF noise:** The average of the events of the RMS of the ADC 7 samples distribution is very large .
- **No data:** No data comes out of the ADC (this description applies to an entire digitizer or above problem, otherwise it is "ADC dead").
- **Wrong DSP configuration:** The online DSP configuration used to collect data was not correct. Data is not usable.
- **Channel masked:** Wildcard for unspecified problems, as for the ADC's.

- **No PMT connected:** The PMT is not connected to the read-out or is dead.

- **No HV:** No High Voltage is applied to the PMT.

- **Wrong HV:** A wrong HV value is applied to the PMT.

- **Large HF noise:** RMS of the first sample distribution is large

- **No CIS calibration:** Set for channels with a long-term problem in CIS calibration, determined based on the history of observed problems. It is suitable to be used in data analysis to reject channels that are not reliably calibrated with CIS.

- **Bad CIS calibration:** Set for channels with 5.2 pF capacitor problems. Monocis cannot be used for these channels.

- **No laser calibration:** The PMT doesn't receive Laser light from Laser box.

- **Bad laser calibration:** PMT receives light, but it is either too low or too light and the PMT cannot be calibrated.

- **No cesium calibration:** PMT doesn't respond to Cs scan (issue of the integrator read-out).

- **Bad cesium calibration:** PMT responds to Cs scan, but the response is either too low or too high and the PMT cannot be calibrated.

- **Ignored in HLT:** useful for very hot or strange channels that might bias HLT decision. This is practically the only useful online problem flag. It should contain all channels bad due to offline problems.

- **General ADC masked:** Wildcard to mask problems not fitting any other description

For the $\not{E}_T$ reconstruction, cell tagged "high noise" in Liquid Argon detectors are masked for all calorimeter reconstruction algorithms. In practice, the masking in LAr cells, means that their energy is set to zero. For Tile calorimeter, cells with hardware problems (High Voltage, Data acquisition, ...) or/and corrupted data are masked. These masked cells represent only 0.1% of the total (206 cells), mainly located in the Tile barrel. This assesses the very high quality of the calorimeter system over its whole coverage. More quantitative information on the impact of those masked cells on $\sum E_T$ variable is discussed in the following subsection.

2.11.1 Influence of problematic channels on $\sum E_T$

The influence of the masked bad cells on the $\sum E_T$ $\not{E}_T$-*base* distribution is presented for randomly triggered events of the reference run (91639) in a sample of 1000 events. Figure 2.35 compares $\sum E_T$ distribution with and without bad cells masking: the $\sum E_T$ distributions (i) are wider, in particular for Tile (3.1 GeV w.r.t. 1.3 GeV) and to a less extend for the EM barrel (3.7 GeV w.r.t. 3.3 GeV), (ii) are shifted for Tile (by almost 10 GeV), (iii) show more tails, in particular EM barrel and Tile and (iv) some underflows appear in EM barrel. Commenting the last point, this behavior is expected, since EM barrel is the sub-detector which contains the biggest number of bad cells[41] due to the fact that it is the sub-detector with the highest number of cells. A bad cell is interpreted as a "noisy" cell resulting to unexpected high energy contributions. This is the reason why some points appear away from the normal Gaussian distribution producing the underflows.

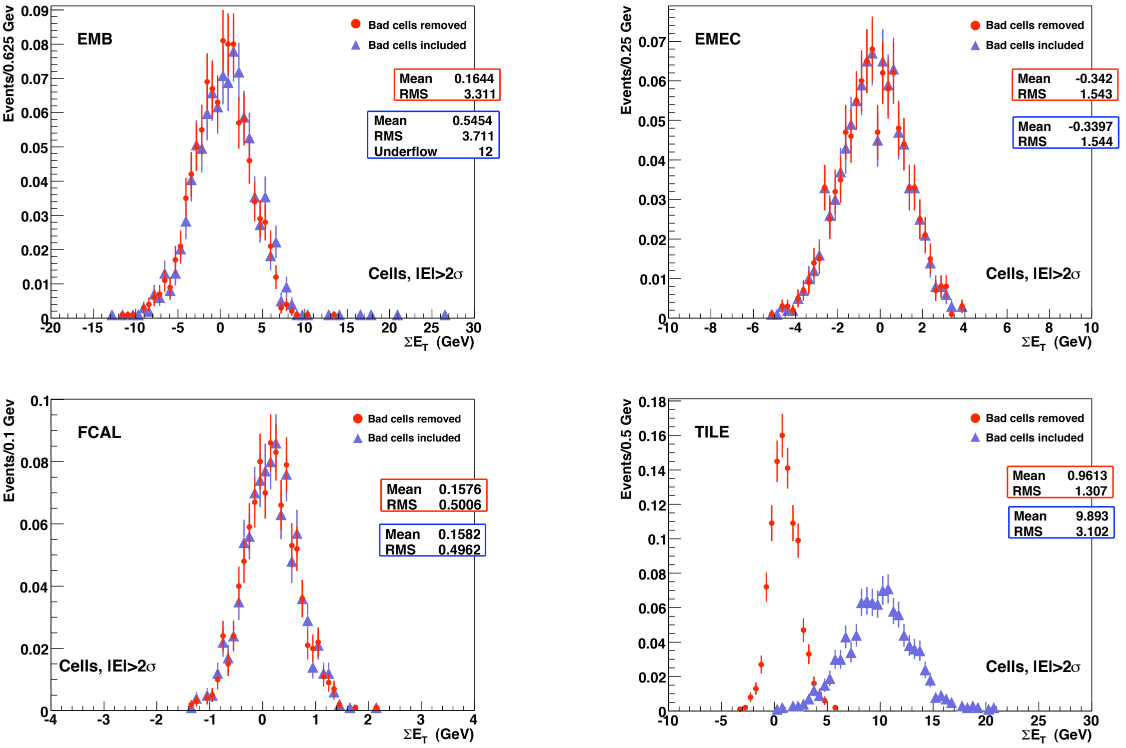


Figure 2.35: $\sum E_T$ *distributions for LAr EMB/EMEC/FCal and Tile sub-detectors and the whole calorimeter system, computed with cell based method in random triggers of run 91639 in a sample of 1000 event for the case where problematic cells are included (blue markers) and the case where problematic cells are excluded (red markers).*

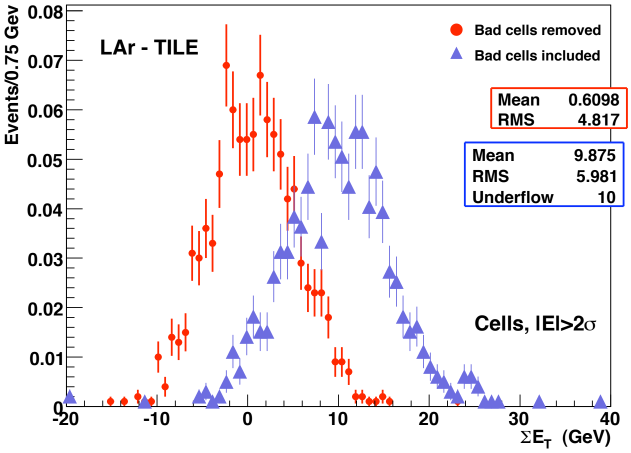


Figure 2.36: $\sum E_T$ *distribution for ATLAS calorimeter, computed with cell based method in random triggers of run 91639 in a sample of 1000 event for the case where problematic cells are included (blue markers) and the case where problematic cells are excluded (red markers).*

2.12 Calorimeter commissioning conclusion

Summarizing the study presented in this chapter, a good understanding of noise in the ATLAS calorimeter cells is achieved, performing a calorimeter commissioning through $\not\!\!E_T$ variables. More than 100k events taken in the period between September and October 2008 with the full partition of ATLAS detector, were examined focusing to the random trigger stream. The importance of detecting "noisy" cells and of understanding fully the energy reconstruction aspects (pedestal shift, iteration threshold) was highlighted.

A Gaussian behavior of noise in all LAr calorimeters is found with the exception of a small region in presampler, which was solved later in hardware level, whereas a region of non-Gaussian tails in Tile was presented. This study was performed during the fall 2008 single beam/cosmics data taking and it was approved by the ATLAS community as public cosmics results[10]. This approval resulted in the plots presented in figures 2.37, 2.38 and 2.39

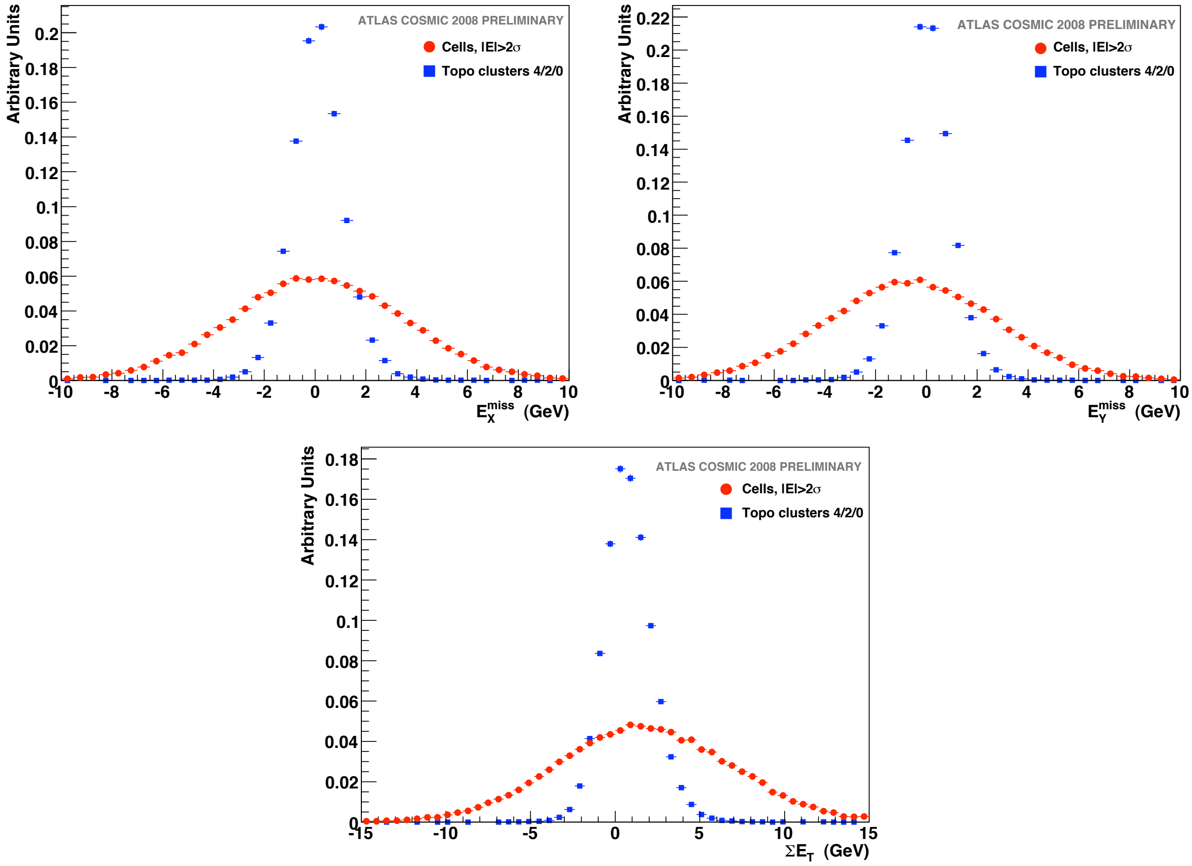


Figure 2.37: *$\not\!\!E_T$-base (red circle markers) and $\not\!\!E_T$-topo (blue square markers) E_X^{miss} , E_Y^{miss} , $\sum E_T$ distributions. ATLAS official approved plots.*

Figure 2.37 shows the E_X^{miss} , E_Y^{miss} and $\sum E_T$ distributions superimposed for $\not\!\!E_T$-*base* and $\not\!\!E_T$-*topo* method. Very good Gaussian shapes are presented and as a general remark one could say that $\not\!\!E_T$-*topo* shows a better resolution, a feature which is due to the fact that a higher number of cells is selected for $\not\!\!E_T$-*base* . The slight shifts with respect to the distributions width, observed in the mean values are addressed in section 2.7.1 and they are corrected when applying the corrected “effective pedestal” shift, and the appropriate iteration threshold (5σ).

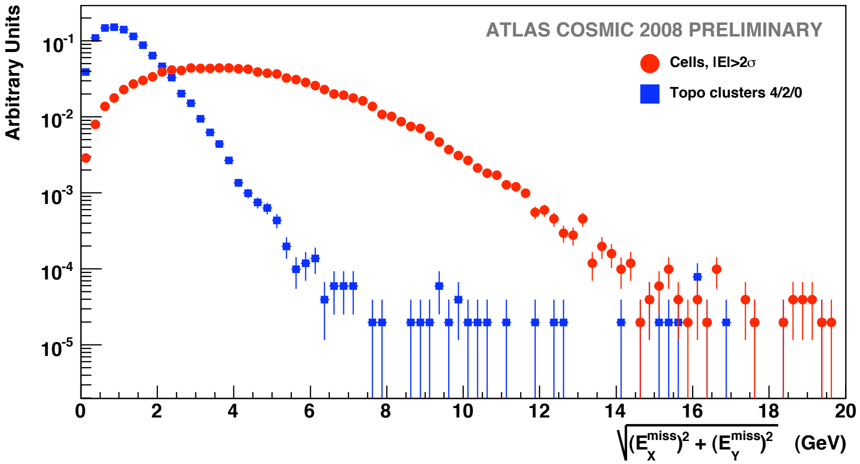


Figure 2.38: *$\not\!\!E_T$ distributions both cell-(red circle markers) and topocluster-based (blue square markers) methods. ATLAS official approved plots.*

Figure 2.38 illustrates the $\not\!\!E_T$ variable used in physics analysis for $\not\!\!E_T$*-base* and $\not\!\!E_T$-*topo* method. As explained in previous paragraph, $\not\!\!E_T$*-topo* has a better resolution due to the tighter noise suppression. Nevertheless, an $\not\!\!E_T$ contribution originating from noise can reach 4-6 GeV for $\not\!\!E_T$*-topo* and 8-13 GeV for $\not\!\!E_T$*-base* . The tails observed in both $\not\!\!E_T$ methods as shown in section 2.7.1 are understood and fixed. They were caused by a coherent noise in LAr presampler originating from a HV cable in hardware level.

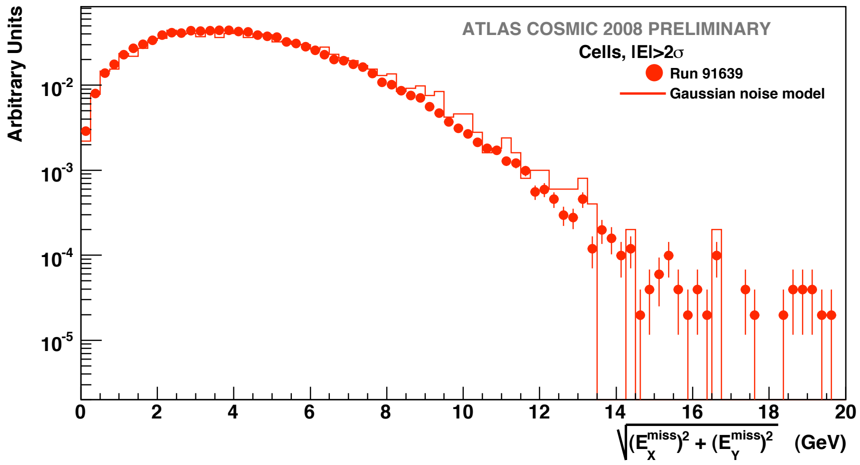


Figure 2.39: *The expected $\sum E_T$ distribution obtained by a randomisation of the cell energy with a Gaussian noise of width σnoise, superimposed on the measured cell-based $\not\!\!E_T$ distribution. ATLAS official approved plots.*

Finally, the $\not\!\!E_T$ distribution for $\not\!\!E_T$*-base* is superimposed with a Gaussian noise model in figure 2.39. A good agreement is shown, a fact which proves a Gaussian behavior of

ATLAS calorimeters for the $\not{E}_T$*-base* case (selected cells if $|E/\sigma| > 2$). The reason why the respective plot for $\not{E}_T$*-topo* is not shown is traced back in section 2.8 where is shown that the non Gaussian noise in Tile generates more topocluster seeds ($|E/\sigma| > 4$) than a Gaussian model would predict and it was decided that a better understanding of this behaviour was needed before making this plot public.The respective plot for $\not{E}_T$*-topo* is shown in figure 2.21.

Chapter 3

Theory

The Standard Model is a successful theory incorporating the present understanding of fundamental particles and their interactions. It is based on a spontaneously broken local $SU(3) \times SU(2) \times U(1)$ gauge theory describing the strong, weak and electromagnetic interactions. A brief descriptions of the Standard Model is given in section 3.1 including the Higgs mechanism which describes the origin of the masses of the fundamental particles. Further details can be found in [17]. Section 3.2 deals with the Higgs boson predicted by the theory and limits on the Higgs boson mass from theory and experiments are discussed in section 3.3. In sections 3.4 and 3.5, the production mechanisms and decay channels of the Higgs boson in the Standard Model are described.

3.1 The Standard Model of particle physics

In the Standard Model (SM), three types of fundamental particles are distinguished according to their spin:

- Fermions with spin $\frac{1}{2}$. The fermions f are the matter constituents and are grouped further into doublets of leptons l and quarks q. Both lepton and quark doublets exist in three generations with increasing mass:

  Leptons

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

  Particles in the second line are of charge -1, while the neutrinos are chargeless. The SM assumes that the neutrino masses are 0.

  Quarks

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

The u (up), c (charm) and t (top) quarks are of charge $\frac{2}{3}$, while the d (down), s (strange) and b (bottom) quarks are of charge $-\frac{1}{3}$. The quarks carry a color charge that is related to the strong interaction. They cannot exist freely, since they have to form bound states so that the resulting state is "colorless". Since the existence of states with more than three quarks is generally not considered to be proved, there are two possible combinations: either a quark is bound with an anti-quark forming a meson or three (anti-)quarks are forming a (anti-)baryon. These two particle types are grouped under the name hadrons. As mentioned above the particles of the first generation are the least heavy ones. Except the neutrinos the particles of the 2^{nd} and 3^{rd} generations are unstable, thus giving the first generation the privilege to form all ordinary matter.

- Bosons with spin 1. They are the vector bosons of the gauge fields mediating the three fundamental forces: the electromagnetic, the weak and the strong interaction. In more details:
  - The electromagnetic interaction affects all particles of non-zero electromagnetic charge. Its associated vector boson is the photon (γ). The associated charge is the electric charge.
  - The weak interaction affects all fundamental fermions. It has three associated vector bosons, W^+, W^- and Z. The associated charge is the weak charge.
  - The strong interaction affects only the quarks and is carried by eight gluons (g). The associated charge is called "color".

  The properties of the vector bosons are summarized in table 3.1.

The intermediate vector bosons

Boson	Mass (GeV)	Electric charge	Weak charge	Strong charge	Associated interaction
γ	0	0	0	0	electromagnetic
Z^0	91.1876 ± 0.0021	0	1	0	weak
$W^\pm$	80.398 ± 0.025	± 1	1	0	
g	0	0	0	1	strong

Table 3.1: The intermediate vector bosons of the Standard Model, see [45].

- The Higgs boson with spin 0 is associated with the spontaneous breaking of the electroweak gauge symmetry. It has not yet been observed experimentally and its potential discovery in the $\tau^+\tau^-$ decay channel is the main topic of the present document.

The Standard Model has been extended from models developed in the 1960's by Glashow, Winberg and Salam [46]. It is based on the gauge symmetry $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$. $SU(3)_C$ is the group of color symmetry, described within the frame of Quantum Chromo Dynamics (QCD), $SU(2)_L$ the one of the weak isospin symmetry and $U(1)_Y$ the one for the hyper-charge symmetry. The symmetry $SU(2)_L \otimes U(1)_Y$ is broken spontaneously by the

Higgs mechanism ($SU(2)_L \otimes U(1)_Y \rightarrow U(1)_Q$). With the organization of the quark fields (Q) and lepton fields (Ψ) into multiplets,

$$Q_L = \begin{pmatrix} U_L \\ D_L \end{pmatrix} \ , \ u_R, d_R \ \ and \ \ \Psi_L = \begin{pmatrix} \nu_L \\ l_L \end{pmatrix} \ , \ l_R, \tag{3.1}$$

where L and R indicate left and right chiralities. The electroweak part of the SM Lagrangien is

$$\mathcal{L}_{EW} = i\{\,\overline{Q}_L\gamma_\mu\mathcal{D}^\mu Q_L+\overline{u}_R\gamma_\mu\mathcal{D}^\mu u_R+\overline{d}_R\gamma_\mu\mathcal{D}^\mu d_R+\overline{\Psi}_L\gamma_\mu\mathcal{D}^\mu\Psi_L+\overline{\nu}_R\gamma_\mu\mathcal{D}^\mu\nu_R+\overline{l}_R\gamma_\mu\mathcal{D}^\mu l_R\,\} \tag{3.2}$$

with the covariant derivative

$$\mathcal{D}^\mu = \partial^\mu + i\frac{g}{2}\tau_j W_j^\mu + 2ig'YB^\mu \quad , \tag{3.3}$$

where τ_j are the Pauli matrices in the $SU(2)_L$ space, Y is the hyper-charge, g is the coupling constant associated with the gauge fields W_j ($j = 1, 2, 3$) that are related to the weak isospin symmetry group $SU(2)_L$ and g' is the coupling constant associated with the gauge field B that is related to the hypercharge symmetry group $U(1)_Y$. The charged vector bosons $W^\pm$ are combinations of the fields $W_{1,2}$:

$$W^{\mu\pm}(x) = \frac{W_1^\mu \mp iW_2^\mu}{\sqrt{2}} \tag{3.4}$$

Also, by introducing the weak mixing angle θ_W[1], one can define the fields associated to the Z boson (Z^μ) and the photon (A^μ) as combinations of W_3^μ and B^μ:

$$Z_\mu = \cos\theta_W W_\mu^3 - \sin\theta_W B_\mu \tag{3.5}$$

$$A_\mu = \sin\theta_W W_\mu^3 + \cos\theta_W B_\mu \tag{3.6}$$

However, this theory has one severe problem: to conserve $SU(2)$ symmetry, the weak gauge bosons $W^\pm$ and Z^0 have to be massless which is in contradiction to observations.

3.2 The Higgs mechanism

The solution to the problem of massive gauge bosons was provided by P.W.Higgs [1] and others [15, 16] in the year 1964. Based on the work of Nambu [2] and Goldstone[3], they developed a mechanism, the Higgs mechanism, in which massive gauge bosons can be accommodated by introducing a complex scalar field of the form :

$$\Phi = \begin{pmatrix} \phi_+ \\ \phi_0 \end{pmatrix} \quad , \tag{3.7}$$

described by the Lagrangien:

$$\mathcal{L}_H = \mathcal{D}^\mu\Phi^\dagger\mathcal{D}_\mu\Phi - V \tag{3.8}$$

---

[1] or Weinberg angle

where $\mathcal{D}^{\mu}$ is the covariant derivative from equation 3.3 and V stands for the Higgs field potential:

$$V = \mu^2\Phi^{\dagger}\Phi + \lambda(\Phi^{\dagger}\Phi)^2 \qquad . \tag{3.9}$$

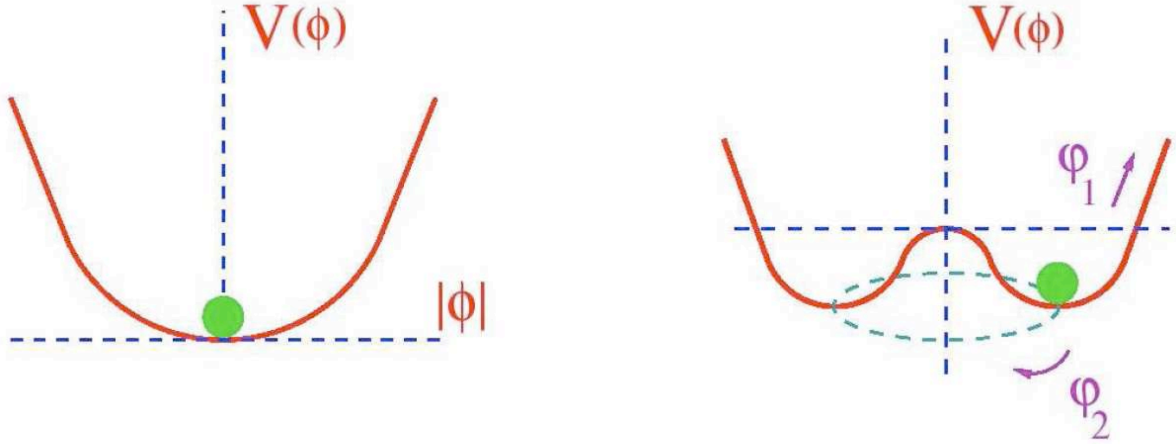


Figure 3.1: *Illustration of the Higgs potential for μ^2 >0 (left) and μ^2 <0 (right). In the latter case, spontaneous symmetry breaking occurs when the system moves to one of the ground states [47].*

To ensure the existence of stable ground states, the potential is bound from below by requiring λ >0. Figure 3.1 illustrates the potential for μ^2 >0 and μ^2 <0. In the latter case, the minimum of the potential is not at $|\Phi_0| = 0$ but at:

$$|\Phi_0|^2 = \frac{-\mu^2}{2\lambda} \equiv \frac{v^2}{2} \tag{3.10}$$

with v called the vacuum expectation value of the scalar Higgs field. The $SU(2) \times U(1)$ symmetry is spontaneously broken when one of the ground states for μ^2 <0 is chosen, by choosing a non-zero value for the Higgs potential in the vacuum state. The Lagrangian remains invariant under local $SU(2) \times U(1)$ gauge transformations. A usual choice is to set $\Phi_+ = 0$ and to make Φ_0 real:

$$\langle 0|\Phi|0\rangle = \left(\begin{array}{c} 0 \\ \frac{v}{\sqrt{2}} \end{array} \right) \qquad , \tag{3.11}$$

with

$$v = \sqrt{\frac{-\mu^2}{2\lambda}} \quad . \tag{3.12}$$

In this case, one can parameterize excitations from this ground state by

$$\Phi = \frac{1}{\sqrt{2}} e^{i\varsigma i(x)} \left(\begin{array}{c} 0 \\ v + H(x) \end{array} \right) \tag{3.13}$$

with a scalar field $H(x)$, the massive Higgs field which describes radial excitations from the ground state changing the potential energy, and massless scalar fields $\zeta_i(x)$, the Goldstone bosons, corresponding to angular excitations without potential energy change. This parameterization can be eliminated by a local $SU(2)$ gauge transformation leading to the following parameterization of the scalar field :

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H \end{pmatrix} \tag{3.14}$$

Introducing the field parameterization 3.14 into the Lagrangian 3.8 and using equations 3.4, 3.5 and 3.6, the following kinematic terms are obtained for the W^+, W^- and Z^0:

$$\frac{g^2v^2}{8} W_\mu^+ W^{\mu+} + \frac{g^2v^2}{8} W_\mu^- W^{\mu-} + \frac{g^2v^2}{8cos^2\theta_W} Z_\mu Z^\mu \tag{3.15}$$

and thus the W,Z bosons masses are obtained:

$$M_W = \frac{vg}{2} \tag{3.16}$$

$$M_Z = \frac{vg}{2\cos\theta_W}. \tag{3.17}$$

To obtain the masses for quarks and leptons, one can write the Langrangien for the Yukawa interaction with the Higgs field:

$$\mathcal{L}_{YU} = Y_{ij}^d \overline{Q}_{L_i} \Phi\, d_{R_j} + Y_{ij}^u \overline{Q}_{L_i} \tilde{\Phi}\, u_{R_j} + Y_{ij}^l \overline{L}_{L_i} \Phi\, l_{R_j} + h.c. \;, \tag{3.18}$$

where i and j run over all generations. Concentrating on the quarks part only, the quark mass matrix can be introduced:

$$M_{ij}^{u,d} = \frac{Y_{ij}^{u,d} \cdot v}{\sqrt{2}} \tag{3.19}$$

and the respective part of the Lagrangien becomes

$$\mathcal{L}_M = M_{ij}^d \, \overline{d}_{L_j}^{int.} \, d_{R_j}^{int.} + M_{ij}^u \, \overline{u}_j^{int.} \, u_{R_j}^{int.} + h.c. \;, \tag{3.20}$$

introducing quadratic,kinematic terms for the quarks. The i and j are running again over all generations.

The Glashow-Salam-Weinberg theory includes the Higgs mechanism. It predicts the W and Z boson and their properties, like mass and decay width. In 1983, the three weak gauge bosons have been discovered by the UA1 and UA2 experiments at CERN [48, 49, 50]. Their properties have been measured precisely at LEP in the nineties with no departure from standard model prediction. The precise measurements allowed an estimation through radiative correction of the top quark mass, before it was discovered at the Tevatron in 1995 with consistent properties, providing a confirmation of the Standard Model validity.

3.3 The Higgs boson mass

The Higgs boson obtains its mass via self-coupling. The value of m_H

$$m_H = \sqrt{2\lambda}v \tag{3.21}$$

is not predicted by the SM and thus, has to be measured. However, several theoretical and experimental constraints limit the possible mass range.

3.3.1 Theoretical limits on the Higgs boson mass

Several consistency requirements of the theory set upper and lower bounds on the Higgs mass in the SM depending on the energy scale Λ up to which the SM is valid and no new interactions or particles appear. Values of Λ up to the Planck mass scale $M_{Planck} = 10^{19}$ GeV are considered, since above this energy threshold, quantum effects of gravity become strong and thus the description of particle interactions in terms of quantum field theory breaks down (due to the non-renormalizability of gravity). Here, only the results are summarized (for further details see reference [51]):

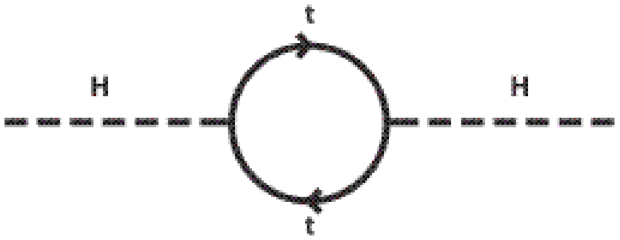


Figure 3.2: *Feynman diagram of Higgs and top quark loop.*

- Unitarity of the electroweak interactions, in particular of the $W^+W^- \rightarrow W^+W^-$ scattering amplitude, limits the Higgs boson mass to m_H <1 TeV.
- The requirement of finite self-coupling of Higgs bosons, including Higgs and top quark loops (figure 3.2), restricts the Higgs mass with an upper bound depending on Λ. m_H <600 GeV for Λ=1 TeV and $m_H \leq$180 GeV for Λ=M_{Planck}.
- To ensure the stability of the Higgs ground state, the Higgs potential (equation 3.9) has to have a lower bound ($\lambda(\Lambda)$ >0). This results in a low limit on the Higgs boson mass of m_H >55 GeV for Λ=1 TeV and $m_H \geq$130 GeV for Λ=M_{Planck}.

Figure 3.3 illustrates the theoretical upper and lower bounds for the Standard Model Higgs boson as a function of Λ [51]. As a result, the measurement of the Higgs boson mass will constrain Λ. A Higgs boson mass of m_H= 500 GeV for instance, implies that the SM breaks down already at a much lower energy scale than M_{Planck} .

For $\Lambda = 1$ TeV, the following theoretical limitation for the m_H range is obtained:

$$55\ GeV \leq m_H \leq 600\ GeV. \tag{3.22}$$

3.3.2 Experimental limits on the Higgs boson mass

A lower bound of the Higgs boson mass comes from direct Higgs boson searches at the Large Electron Positron Collider LEP at CERN. The combined results of LEP experiments (ALEPH, DELPHI, L3 and OPAL) on the direct searches for the Higgs boson at LEP conclude that the Standard Model Higgs boson must be heavier than $m_H \geq$114.4 GeV at 95% confidence level (C.L.) [52].

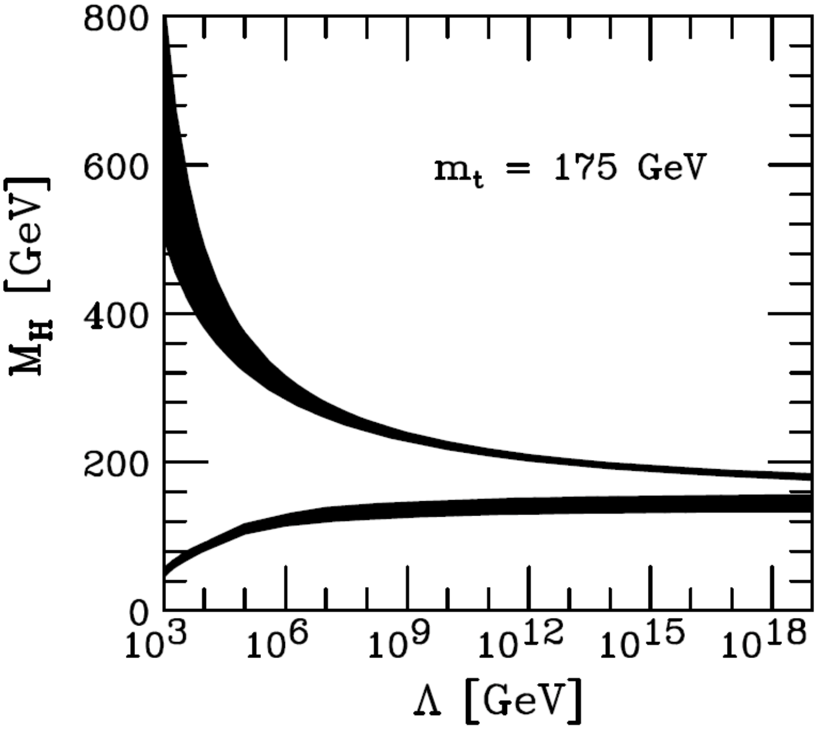


Figure 3.3: *Upper and lower theoretical bounds on the Higgs boson mass as a function of the energy scale* Λ *up to which the SM is valid. A top quark mass of* m_{top} = *175 GeV and a QCD coupling constant* $\alpha_s(M_Z)$=*0.118 are assumed. The bands indicate the theoretical uncertainties [51].*

The range of possible Higgs boson masses can be further constrained by combining the precision measurements of electroweak observables at LEP and at the Tevatron $p\bar{p}$ collider at Fermilab, with the experiments CDF and D0. Higher order corrections to the calculation of electroweak processes like $e^+e^- \rightarrow Z \rightarrow f\bar{f}$ include Higgs loops and thus depend on the Higgs mass. The measurements of LEP and Tevatron are combined in [53] with a precision sufficient to constrain the Higgs boson mass. The most recent result of this calculation is given in figure 3.4 which shows $\Delta\chi^2 = \chi^2 - \chi^2_{min}$ of the global least-squares fit of the Standard Model predictions to the electroweak data as a function of the Higgs boson mass. It can be seen that the fit to electroweak precision measurements favors a small Higgs boson mass. The most probable Higgs mass according to the fit is m_H=85^{+39}_{-28} GeV, taking the lower Higgs mass limit from direct searches at LEP into account the 95% confidence level upper limit being m_H≤154 GeV [53].

The latest results from direct Higgs boson searches at the Tevatron collider exclude the Standard Model Higgs boson with a mass between

$$160\ GeV \leq m_H \leq 170\ GeV \tag{3.23}$$

at a confidence level of 95% [54] (figure 3.5).

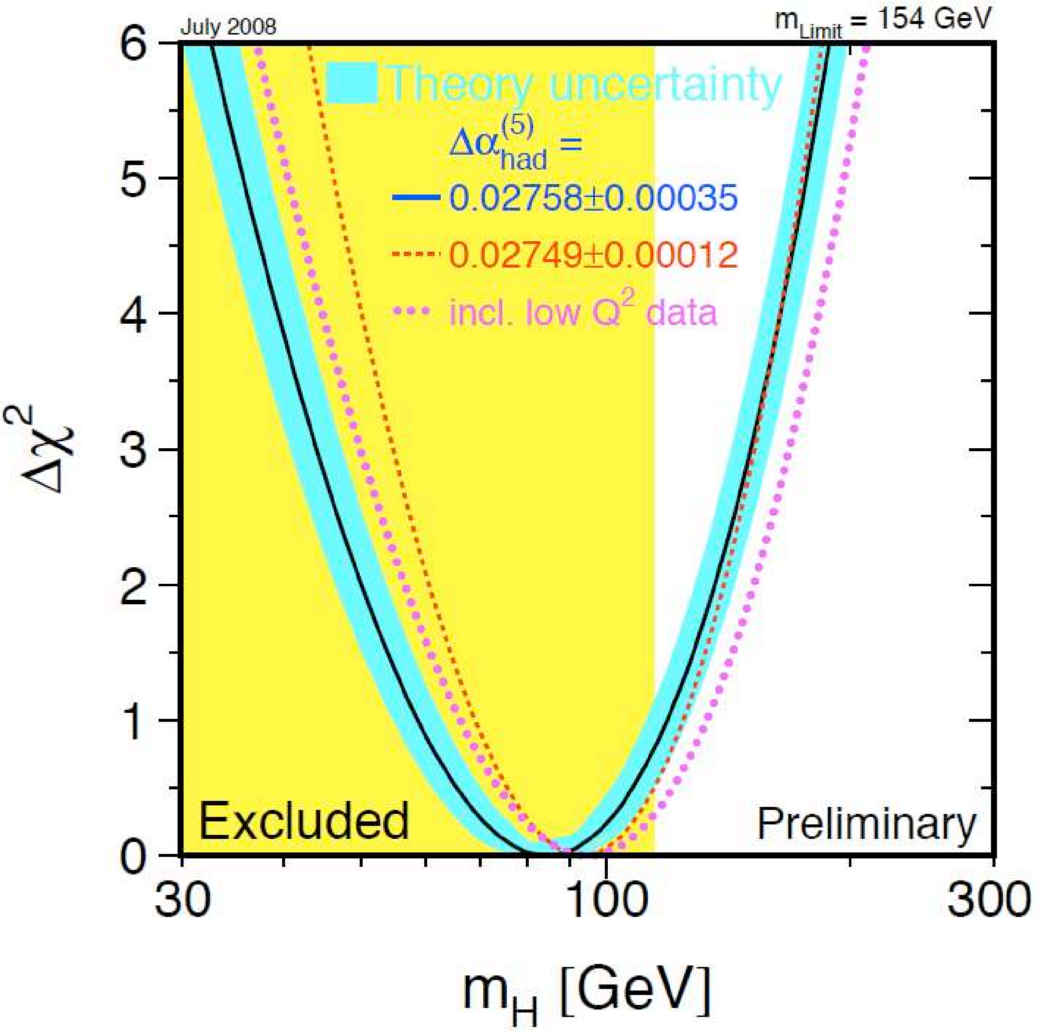


Figure 3.4: $\Delta\chi^2 = \chi^2 - \chi^2_{min}$ *from the global least-squares fit to electroweak precision measurements as a function of the Higgs boson mass* m_H*[53].*

3.4 Higgs Boson production mechanisms

At the LHC, the SM Higgs boson can be produced by the following processes:

a. Gluon fusion.

b. Vector boson fusion. This production mode is addressed in the present document.

c. Higgs production associated with a W/Z boson (Higgsstrahlung of W/Z).

d. Higgs production associated with a pair of top quarks.

The Feynman diagrams of these processes are presented in figure 3.6. The cross-sections of these processes depend on the unknown Higgs boson mass. Figure 3.7 shows the Standard Model Higgs boson production cross-sections for LHC at $\sqrt{s}$=14 TeV in the Higgs mass range 100 GeV<m_H<500 GeV . As illustrated, the gluon fusion dominates over the whole

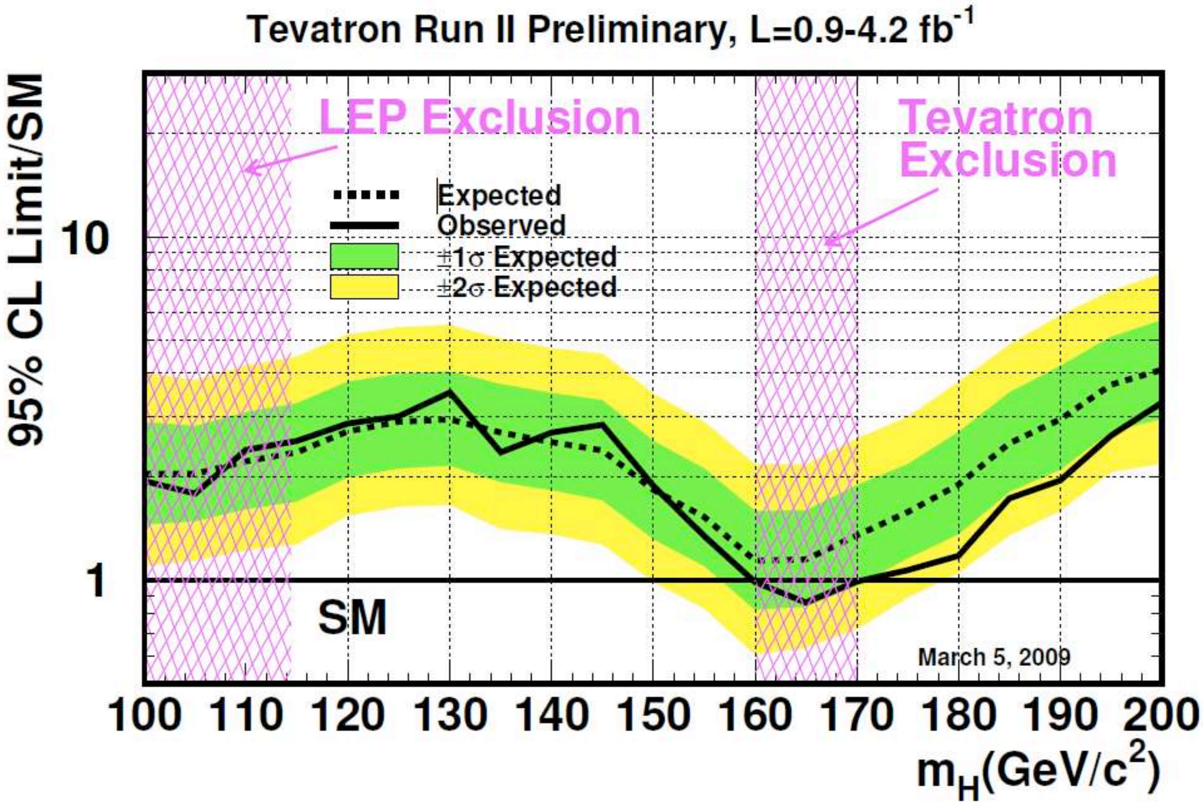
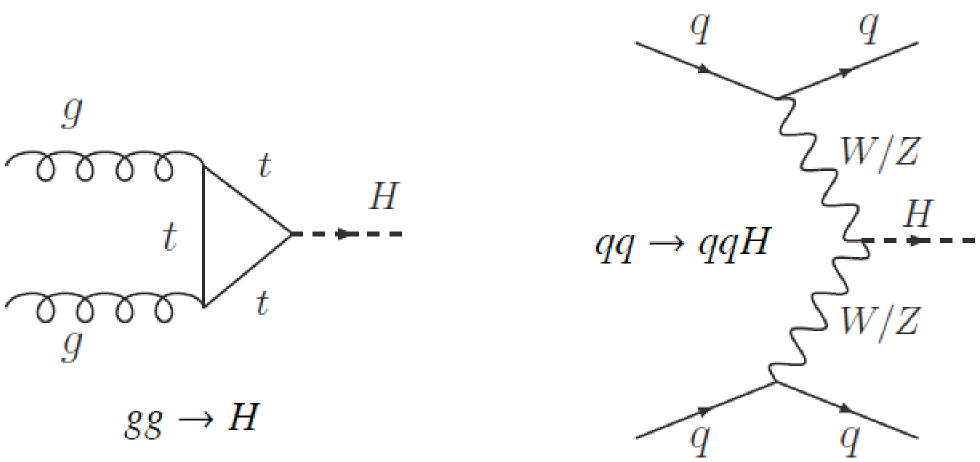


Figure 3.5: *Observed and expected (for the background-only hypothesis (the background-only hypothesis is defined in 4.11.2)) 95% C.L. upper limits on the ratios to the SM cross section, as functions of the Higgs boson mass for the combined CDF and D0 analyses. The limits displayed in this figure are obtained with the Bayesian calculation [54].*

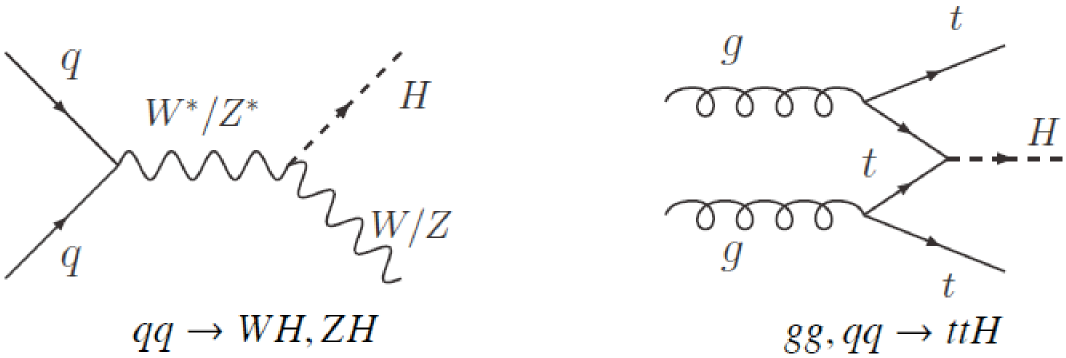
mass range. The vector-boson fusion cross-section is roughly one order of magnitude smaller. However, the two outgoing quarks in this process form two characteristic jets in the very forward regions of the detector at high pseudorapidity values $|\eta|$, which provide a very good signature for background suppression. This main characteristic triggers a VBF Higgs study, an analysis presented in the following chapter.

The other production processes have much lower cross-sections. However, they can still be exploited for specific searches. For instance, requiring two additional b quarks from the top quark decays of the associated Higgs boson production $gg, q\overline{q} \rightarrow ttH$ is essential to suppress the background in searches for Higgs boson decays into a $b\bar{b}$ pair.

Considering that the maximum Higgs boson production cross-section for low Higgs masses is σ_H ~50 pb and the total pp cross-section at the LHC σ_{tot} ~125 mb (figure 1.21), a major challenge of the LHC experiments becomes clear: compared to other pp reactions, the Higgs boson signal is suppressed by ten orders of magnitude. Very detailed and accurate studies should be performed in order to understand and reject the background processes.



(a) Gluon fusion **(b)** Vector-boson fusion



(c) Higgsstrahlung of W/Z **(d)** Associated production with a top quark pair

Figure 3.6: *Tree-level Feynman diagrams of the Higgs boson production processes in pp collisions at the LHC.*

3.5 Higgs Boson decay channels

Also the branching ratios of the Standard Model Higgs boson decays depend on m_H as unknown parameter. Since the Higgs boson couples to other particles proportional to their masses, it decays dominantly into the most massive particles accessible. Feynman diagrams of the SM Higgs boson decays are shown in figure 3.8: the tree-level decays into a fermion pair or a real or weak gauge boson pair and, via loops, the decays into massless photons or gluons.

Figure 3.9 illustrates the branching-ratio calculations of the most important decay channels. A summary of the most important decays is given in the following:

- $H \to b\bar{b}$

  Since the coupling of the Higgs boson to fermions is proportional to the fermion mass,

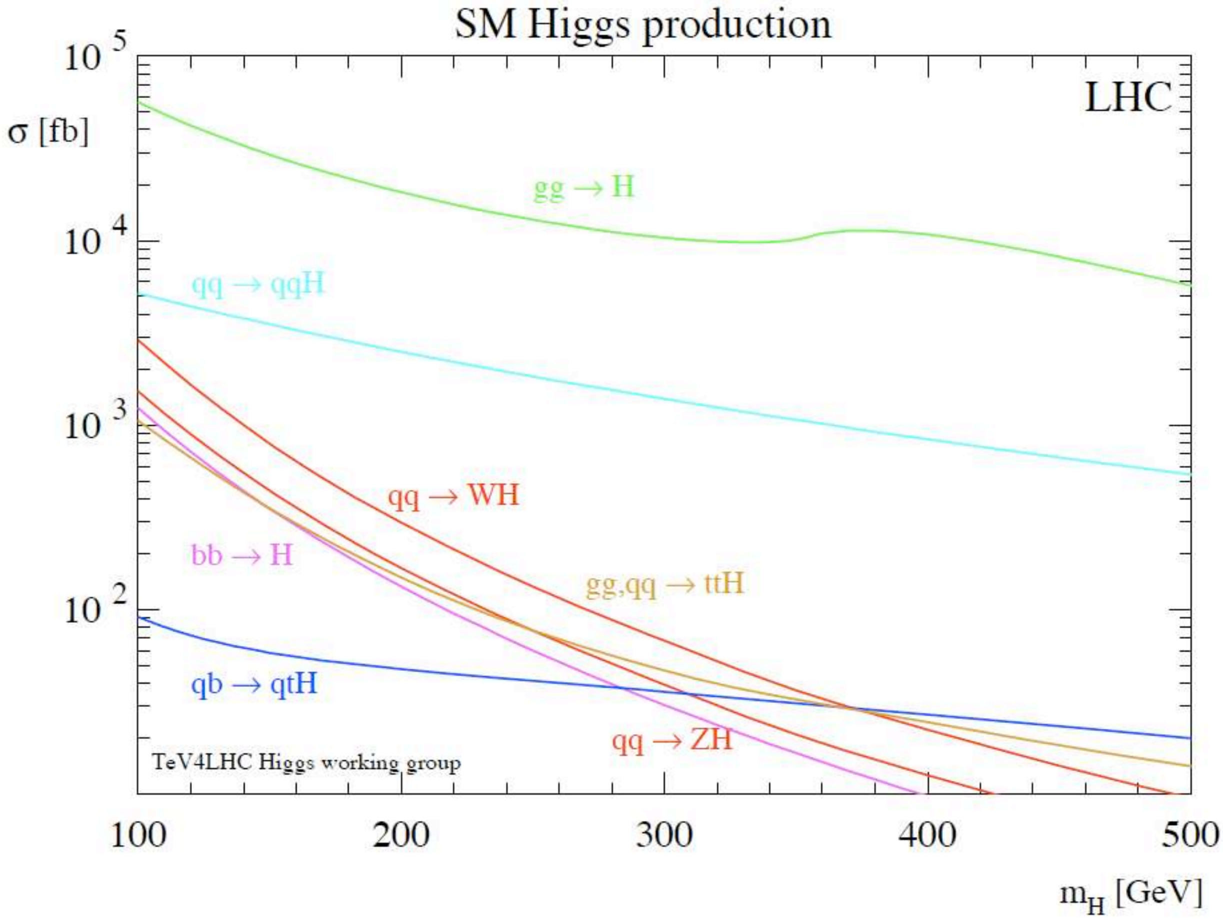


Figure 3.7: *Standard Model Higgs boson production cross-sections as a function of the Higgs boson mass* m_H *at the nominal LHC center-of-mass energy of 14 TeV [55].*

the decay into a $b\bar{b}$ pair has the largest branching ratio for $m_H \leq 140$ GeV. However, the discovery potential for this decay channel suffers from very large QCD background.

- $H \to gg$

  The problem of large QCD background is even more pronounced in the case of the decay into a gluon pair which has the second largest branching ratio in this mass region. The huge QCD jet production cross-sections make it practically impossible to identify $H \to gg$ decays at the LHC.

- $H \to \tau^+\tau^-$

  The Higgs boson decay into a τ-lepton pair also suffers from high background mainly from $Z \to \tau^+\tau^-$ decays. However, if one exploits the signatures of vector-boson fusion Higgs production, this decay channel is one of the most promising ones for $m_H \leq 140$ GeV. The investigation of the discovery potential of the ATLAS experiment

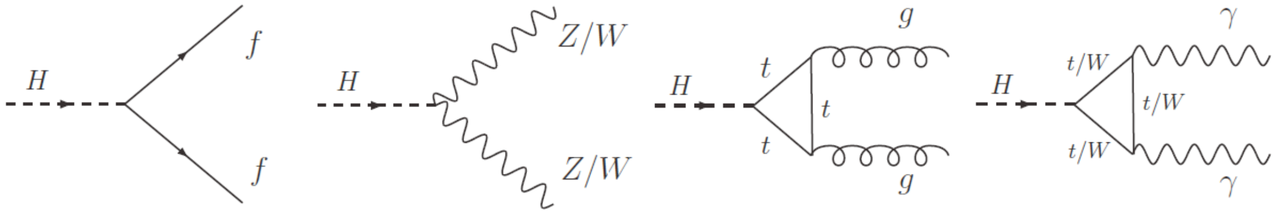


Figure 3.8: *Feynman diagrams of the Standard Model Higgs boson decay channels.*

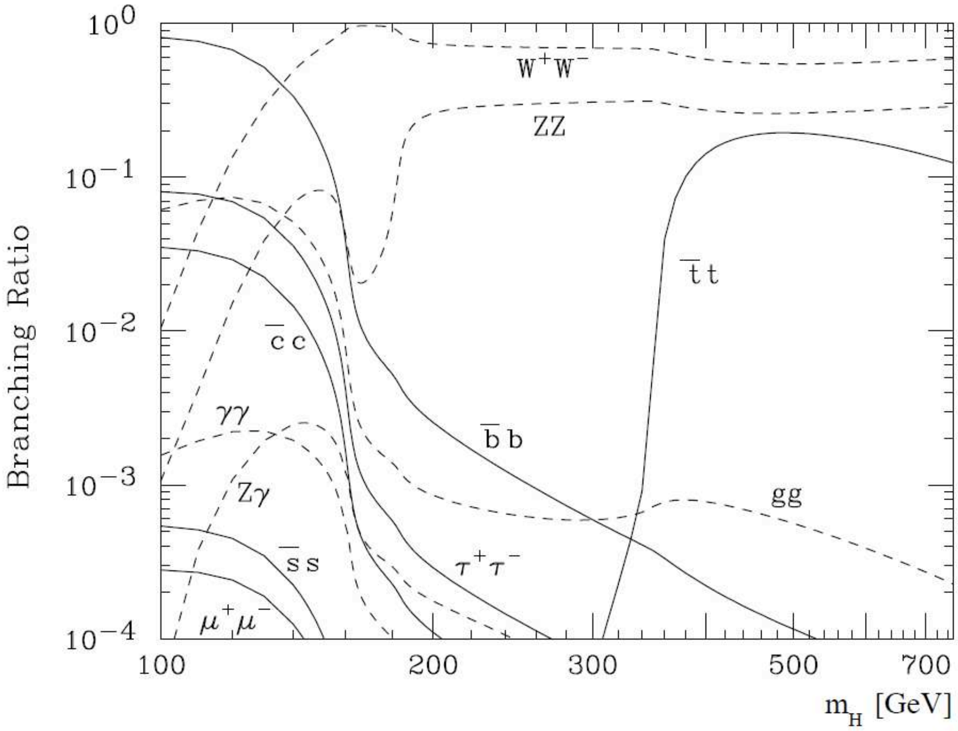


Figure 3.9: *Branching ratios of the Standard Model Higgs boson decays as a function of the Higgs boson mass m_H [56].*

for this decay channel is the main subject of the present document.

- $H \rightarrow \gamma\gamma$

  Another important decay channel for low Higgs masses is the decay $H \rightarrow \gamma\gamma$. Although it has only a very small branching ratio, its discovery potential is high due to the clean signature of two energetic photons and the high Higgs mass resolution in this channel.

- $H \rightarrow W^+W^-$ The branching ratio of $H \rightarrow W^+W^-$ rises towards the threshold for real W-pair production. For m_H~160-180 GeV, the Higgs boson almost exclusively decays into W^+W^-. Unfortunately, the best identifiable leptonic decays of W bosons involve neutrinos, making an accurately Higgs boson mass reconstruction a real challenge.

- $H \rightarrow ZZ$ Above $m_H \geq 190$ GeV, the decay $H \rightarrow ZZ$ is the most promising Higgs discovery channel. The further decay of the Z bosons into electron or muon pairs provides the cleanest signature and an excellent Higgs mass resolution. Therefore, the decay $H \rightarrow ZZ \rightarrow llll$ is known as the "golden signature" for the Higgs boson.

- $H \rightarrow t\bar{t}$ The $H \rightarrow t\bar{t}$ decay becomes kinematically possible above $m_H \geq 350$ GeV. However, due to the high background rate in this decay channel and the branching ratio being about ten times smaller than for $H \rightarrow W^+W^-$, the decay $H \rightarrow t\bar{t}$ is not considered as a Higgs boson discovery channel at LHC.

Chapter 4

Vector Boson Fusion Higgs in the di-tau channels

The search of the Standard Model Higgs boson decaying into a τ-pair in association with two jets in the forward parts of the detector, which is produced by the Vector Boson Fusion (VBF) process, is presented in this chapter. Special attention is drawn to jet studies, an important element of the analysis, which is expected to a play a key role. All the cuts are also presented and several cut-flows of signal and background indicate a significant potential for a discovery at the low mass range, optimized for a Higgs boson mass of 120 GeV. The importance of jet-veto cut study is high-lighted in addition to some cut extensions.

The largest part of the study presented in the following chapter is performed in the context of the ATLAS "Computing System Commissioning" (CSC) effort, a test of the experiment data processing infrastructure which was the opportunity for a major update of all physics analyses [19].

4.1 Introduction

The search for the Higgs boson and the source of electroweak symmetry breaking is a primary task of the Large Hadron Collider (LHC). At the previous chapter, the Standard Model Higgs boson was introduced and its phenomenology at LHC was presented, while in this chapter the Vector Boson Fusion (VBF) $H \to \tau^+\tau^-$decay channel will be presented in details. Vector Boson Fusion (also called Weak Boson Fusion) is one of the four production mechanisms of a Standard Model (SM) Higgs boson in LHC (see previous chapter 3). VBF is a purely electroweak process described by the Feynman diagram presented in figure 4.1. Two quarks originating from the proton beam are scattered through the exchange of two vector bosons ($W^{\pm}W^{\mp}$ or $Z^0\overline{Z^0}$). The fusion of these bosons generates the Higgs boson

$$qq \to qq(W^{\pm}W^{\mp}, Z^0\overline{Z^0}) \to qqH \tag{4.1}$$

which is accompanied by the two scattered quarks.

The VBF topology has some very specific characteristics, which as it will be shown, allow a good signal significance, rejecting background:

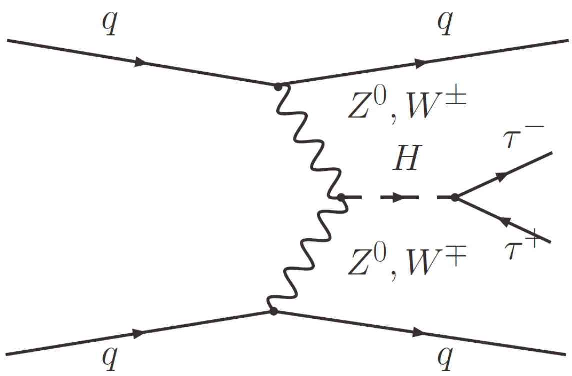


Figure 4.1: *First order Feynman diagram of a SM Higgs boson produced via VBF and decaying to a pair of* $\tau^+\tau^-$ *leptons.*

- The two scattered quarks will hadronize and produce two hard jets at the forward regions of the calorimeter (high $|\eta|$ values). These jets are often called forward jets or VBF tagging jets. The detection of these two jets, requiring a significant η gap in between combined with their high-p_T expected, gives a reasonable starting point selecting VBF events and rejecting potential background.
- The electroweak character of VBF, implies no color exchange between the interacting quarks. As a consequence, light QCD-activity is expected apart from the two forward jets and therefore no other jets are expected, unless jets are produced due to the Higgs boson decay products. Applying several constraints as described later, this feature allow the application of a central jet veto cut, a powerful cut which also contributes to the background rejection.

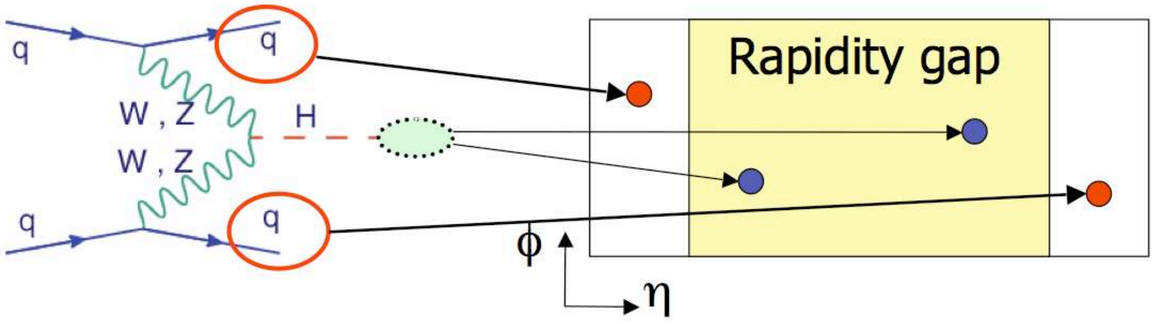


Figure 4.2: *A Higgs boson produced via VBF results to two forward jets in* η*, a significant* η *gap in between and the Higgs boson decay products in this* η *gap. The schema at the right corresponds to an* (η, ϕ) *plane.*

Summarizing these two characteristics, only two hard jets are expected highly separated in η ($\Delta\eta_{jj} >\sim 4$) and all Higgs boson decay products are expected in between these two forward jets in terms of η. For the studied decay mode ($H \rightarrow \tau^+\tau^-$), when removing the τ products, no other jets are expected at final event topology. All these topology aspects are illustrated in figure 4.2, where VBF Higgs boson decay products and VBF tagging jets are matched with the potential detectable objects in an (η, ϕ) plane.

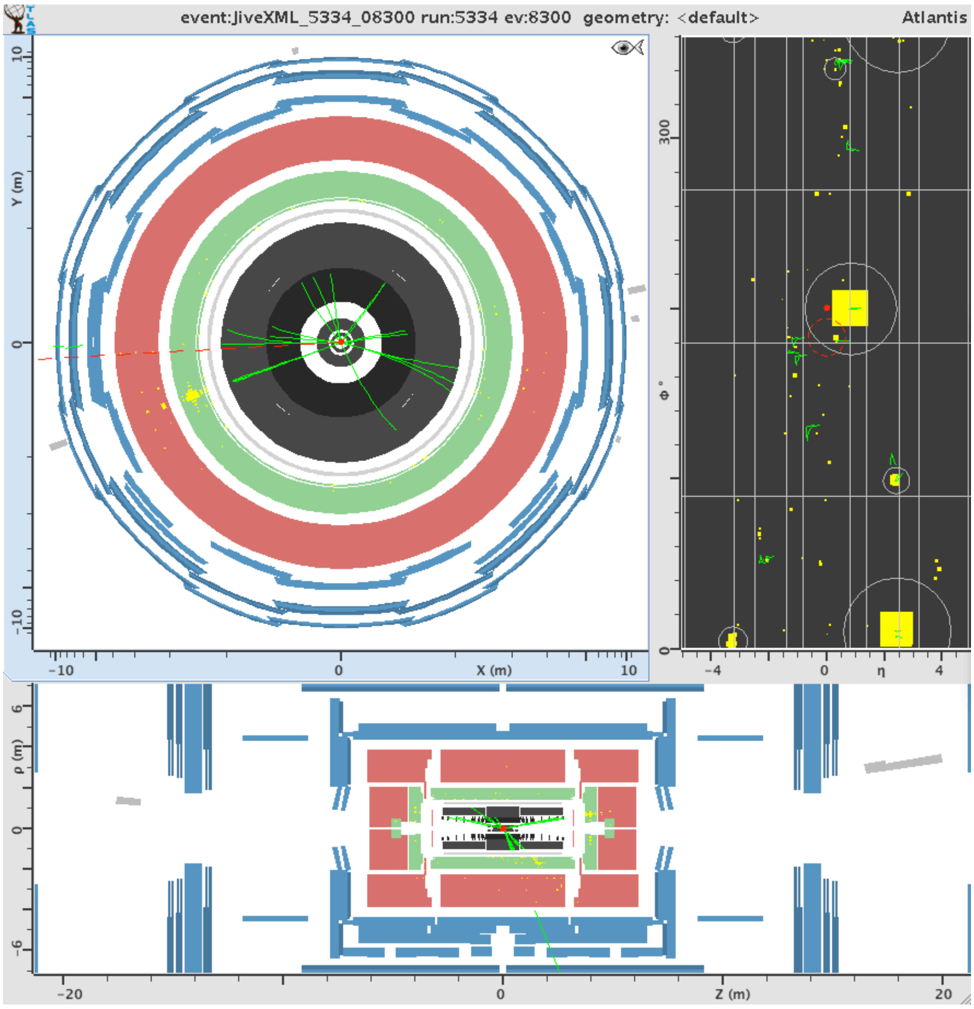


Figure 4.3: *An ATLAS event display for a* $H \to \tau^+\tau^- \to$ *(hadron +* μ*, for a Monte Carlo generated event. Three different views are shown:* (x, y) *plane (upper left),* (η, ϕ) *plane (upper right) and* (r, Z) *plane (bottom). Particle tracks are illustrated with the green lines, while the yellow color represents the energy deposit in the calorimeters.*

Figure 4.3 show an ATLAS event display of a $H \to \tau^+\tau^-$Monte Calro generated event. The two VBF tagging jets are visible in high η values, and the Higgs boson decay products (a hadronic τ and a muon at this example) are equally visible within the η gap defined by the two VBF tagging jets.

The analysis to be presented in following sections, has been done using state-of-the art Monte Carlo generators, full GEANT-based simulation of the ATLAS detector with realistic misalignments and distortions applied to the expected material in the detector, utilization of our current reconstruction algorithms, and, where possible, incorporation of pileup interactions. This analysis requires excellent performance from every ATLAS detector subsystem; the presence of τ decays implies final states with electrons, muons, hadronic tau decays, and missing transverse momentum, while the Vector Boson Fusion production process introduces jets that tend to be quite forward in the detector. Due to the small rate of signal production and large backgrounds, particle identification must be excellent and optimized specifically for this channel. Furthermore, triggering relies on the lowest energy lepton triggers or exceptionally challenging tau trigger signatures.

4.1.1 VBF $H \to \tau^+\tau^-$search at the LHC: motivation

As presented in chapter 3 a SM Higgs boson, if it exists, will be mainly produced via the mechanism of gluon-gluon fusion, since this mechanism is characterized by the highest cross-section. Nevertheless, due to the high QCD background of LHC, such a Higgs boson will be difficult to be detected in some channels even if their branching ratio is relatively high (see chapter 3) e.g. $H \to b\bar{b}$ or $H \to \tau^+\tau^-$. The only decay channel favoring a gluon-gluon fusion Higgs is the $H \to \gamma\gamma$, which even if it has a low branching ratio, it is promising due to its clean photon signature[19]. The second higher cross-section of Higgs production mechanism occurs for the VBF (figure 3.7), which makes possible the study of the decay channels that have a high branching ratio in the low mass region. VBF production mode represents 10 - 20% of the total SM Higgs boson production in the low mass region (115-145 GeV/c^2) and furthermore, searches for the VBF Higgs boson tend to have reasonably high signal-to-background ratios, making them more robust to systematic uncertainties. Within the Standard Model, the ability to observe the Higgs boson in multiple production and decay configurations makes it possible to measure the Higgs boson coupling to fermions and vector bosons [57]. VBF it is the only production mode that makes possible (in the level of signal-to-background ratios) to study the Higgs boson coupling to τ leptons. In addition,the VBF processes provide a tool for measuring the Higgs boson spin and CP properties [58, 59]. In the context of the Minimal Supersymetric Standard Model, (MSSM), the branching ratio of a Higgs boson decaying to photons is generally suppressed, which makes the search for Higgs boson decaying to taus very important. The complementarity of the coupling of the light and heavy CP-even, neutral Higgs bosons of the MSSM to taus makes it possible to cover most or all of the $m_A - tan\beta$ plane by reinterpreting the results for a Standard Model Higgs boson decaying into taus in the context of the MSSM [60, 61].

4.1.2 Signal signature

ATLAS performance in reconstructing τ lepton was discussed in details in section 1.3.3. In a $H \to \tau^+\tau^-$study the decay of τ is an important analysis ingredient, since it defines the final Higgs boson decay channel. As already mentioned, a τ decays either leptonically (e or μ) or hadronically resulting to jets in the detector (mainly originating from pions). According to the τ decay mode, there are 3 possible channels for the study of $H \to \tau^+\tau^-$as described in table 4.2. At the same table the respective branching ratios (table 4.1) are illustrated. Even though the highest decay probability (41%) is obtained when studying the $hadron - hadron$ (hh) channel, this study is outside the scope of the analysis which will follow. Several constraints, mainly due to the difficulties in triggering a hadronic τ because

τ Decay modes	Branching ratio
leptonic (e and μ)	$\sim 35.2\%$
1-prong ($\tau \rightarrow\ n\pi^0\ \pi^\pm v$)	$\sim 49.5\%$
3-prong ($\tau \rightarrow\ n\pi^0\ \pi^\pm\pi^\mp\pi^\pm v$)	$\sim 14.6\%$

Table 4.1: *Tau lepton (τ) principal decay modes, where n is a multiplier factor such as $n = 0, 1, 2, ...$*

of the QCD jets background make the hh analysis very specific. Therefore, the attention will be drawn to the *lepton – lepton* (ll) and *lepton – hadron* (lh) channels.

Decay channels	Channel description	BR
lepton-lepton channel	$qq \rightarrow qqH \rightarrow qq\ \tau^\pm\tau^\mp \rightarrow qq\ e^\pm e^\mp v_\tau v_\tau v_e v_e$	$\sim 3.2\%$
	$qq \rightarrow qqH \rightarrow qq\ \tau^\pm\tau^\mp \rightarrow qq\ \mu^\pm\mu^\mp v_\tau v_\tau v_\mu v_\mu$	$\sim 3.0\%$
	$qq \rightarrow qqH \rightarrow qq\ \tau^\pm\tau^\mp \rightarrow qq\ e^\pm\mu^\mp v_\tau v_\tau v_e v_\mu$	$\sim 6.2\%$
lepton-hadron channel	$qq \rightarrow qqH \rightarrow qq\ \tau^\pm\tau^\mp \rightarrow qq\ e^\pm hadrons\ v_\tau v_\tau v_e$	$\sim 23.2\%$
	$qq \rightarrow qqH \rightarrow qq\ \tau^\pm\tau^\mp \rightarrow qq\ \mu^\pm hadrons\ v_\tau v_\tau v_\mu$	$\sim 22.6\%$
hadron-hadron channel	$qq \rightarrow qqH \rightarrow qq\ \tau^\pm\tau^\mp \rightarrow qq\ hadrons\ hadrons\ v_\tau v_\tau$	$\sim 42.3\%$

Table 4.2: *Decay channels for the study of a SM $H \rightarrow \tau^+\tau^-$followed by the respective branching ratios [45]*

One of the important feature of the final topologies, is the presence of neutrinos (see table 4.2): four neutrinos are present in the ll-channel and three in the lh-channel. Since neutrinos escape the detection, the efficient reconstruction of $\not{E}_T$ becomes a benchmark point of the analysis and as described later, $\not{E}_T$ resolution affects directly the invariant mass of the Higgs boson m_H.

4.2 Background

Every final state which is composed by two or more jets and two τ-lepton decay products is a potential background. Monte Carlo studies show that the most important background contributions are caused by Z + jets (irreducible background) and $t\bar{t}$ production (reducible background). Furthermore, W(W) + jets and QCD multijet events can fake signal events and therefore, are also considered as background. A summary table with all background cross sections is available in section 4.3.

4.2.1 Irreducible background: Z + jets

Irreducible background is a final topology which has exactly the same signature as the signal:

- Two forward jets
- High $\not{E}_T$

- Two isolated leptons (ee or $e\mu$ or $\mu\mu$) for ll-channel
- One isolated lepton (e or μ) and one τ-jet for lh-channel

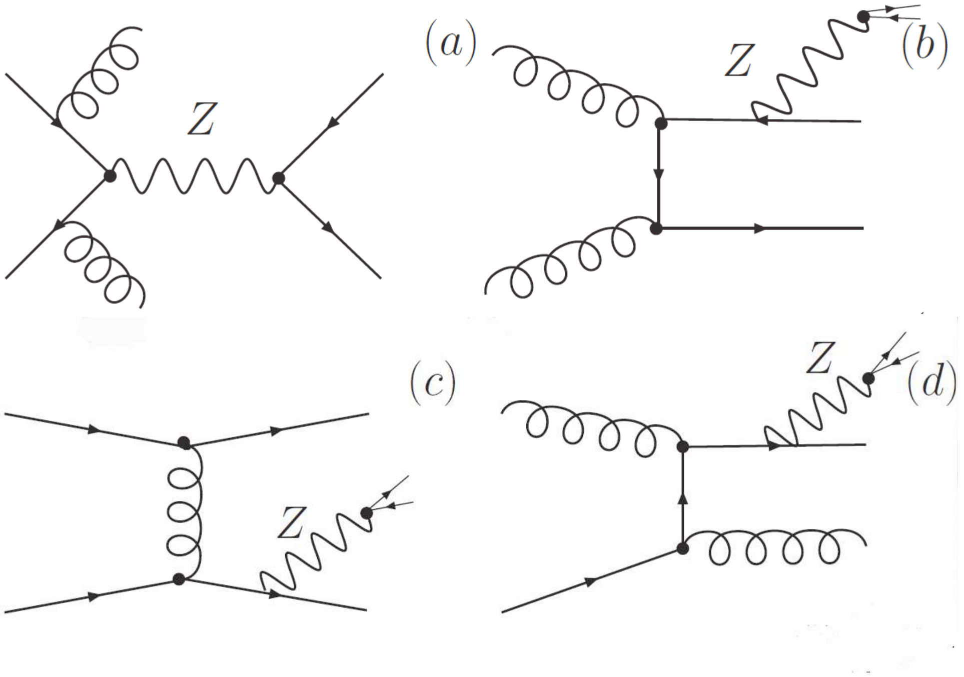


Figure 4.4: *Feynman diagrams at tree level of the irreducible background Z plus jj (Zjj), commonly called QCD Z plus jets processes.*

The main contribution of this kind of background is coming from the QCD corrections of the Drell-Yann process $qq \to (\gamma^*, Z) \to \tau^+\tau^-$ radiating 2 jets in the final state. In figure 4.4 the Feynman diagrams of these processes are illustrated :

a. $q\bar{q} \to gg(\gamma^*, Z) \to gg\tau^+\tau^-$

b. $gg \to q\bar{q}(\gamma^*, Z) \to q\bar{q}\tau^+\tau^-$

c. $qq \to qq(\gamma^*, Z) \to qq\tau^+\tau^-$

d. $qg \to qg(\gamma^*, Z) \to qg\tau^+\tau^-$

Apart from the Zjj processes, the processes Zj have to be included as well. In these cases, there is one hard jet produced and another one softer which originates from QCD radiations:

a. $qg \to qZ \to q\tau^+\tau^-$

b. $q\bar{q} \to gZ \to g\tau^+\tau^-$ (figure 4.5)

Similarly, processes containing more than 2 jets ($Zjjj$, $Zjjjj$, etc) could also be included. They are all based on Feynman diagrams described in figures 4.5 and 4.4 adding CQD

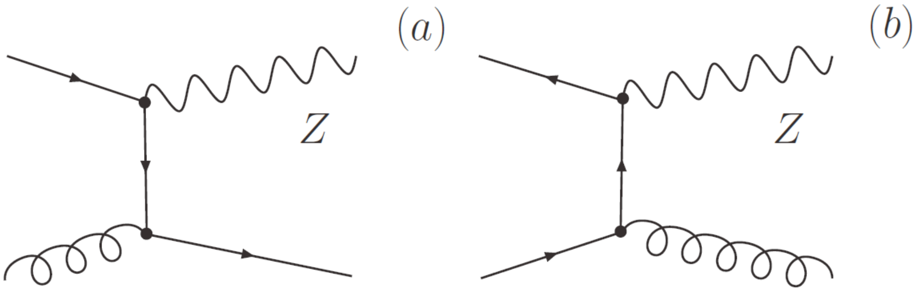


Figure 4.5: *Feynman diagrams at tree level of the irreducible background QCD (Zj). The second jet which fake a VBF event would originate from QCD radiations.*

radiation which conclude to one or several soft jets. All these processes are commonly called QCD Z plus jets.

In addition, an important contribution at the Z background is originating from its electroweak (EW) production, where Z is produced via the scattering of quarks through the exchange of a weak boson in t-channel. At the final state, a Z boson is expected and two quarks which will hadronize producing hard jets (the decay $Z \rightarrow \tau^+\tau^-$is assumed):

$$qq' \rightarrow qq'\tau^+\tau^- \tag{4.2}$$

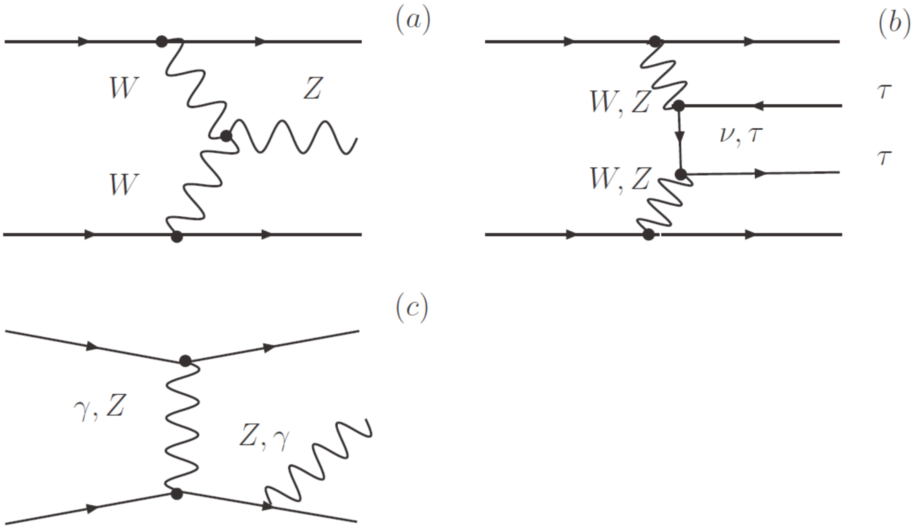


Figure 4.6: *Feynman diagrams at tree level of the irreducible background EW Z plus jj (EW Z plus jets).*

Examples of Feynman diagrams describing the EW Zjj are shown in figure 4.6. This background is more difficult to reject since it is characterized by the same event topology aspects with the one of VBF Higgs boson.

In summary, the Z + jets background suppression depends mainly on the production mode (EW or QCD) and on Z decay channel. The Z boson decays into a lepton pair (ee, $\mu\mu$ or $\tau\tau$, each with a branching ratio of ∼3.4%), into $\nu\nu$ pairs (20.0%) or into hadrons (69.91%). Decays of the Z boson into neutrinos or hadrons do not contribute to the background, since no charged leptons are present in these cases, a principal requirement of ll- or lh-channel. The decays into electrons or muons have no neutrinos in the final state and can thus be suppressed very efficiently by requiring a certain amount of missing transverse energy. Conclusively, only $Z \rightarrow \tau^+\tau^-$ events significantly contribute to the background.

Concerning the production mode, the event topology in the QCD production mode differs from vector-boson fusion Higgs production mainly because of different jet kinematics. As a consequence, adequate jet kinematic cuts have been proposed providing a good suppression of these backgrounds, as described in section 4.6. The most challenging background is the EW Z + jets, however, since its cross-section (1.7 pb) is more than 1000 times smaller than the one of QCD Z production (2.03 nb), QCD Z + jets remains the main background source which affects signal topology.

4.2.2 Reducible background: $t\bar{t}$, W(W) + jets, QCD

The reducible background is characterized by the same signature as signal, but contrary to the irreducible background described in previous section, it occurs after the misidentification of some particles or additional jets. Therefore, such a background may originate from any process which contain 2 jets or more, one isolated lepton (electron or muon), missing transverse energy and a light jet originating from a light quark or a b-quark which will be misidentified as a hadronic tau jet (section 1.3.3).

$t\bar{t}$

The most important reducible background is the top-pair ($t\bar{t}$) production. Feynman diagrams in a tree level of this process are presented in figure 4.7 considering $q\bar{q} \rightarrow t\bar{t}$ (figure 4.7a) and $gg \rightarrow t\bar{t}$ (figures 4.7b, 4.7c, 4.7d) production modes. The relative importance of both amplitudes (gluon-gluon and quark-antiquark scattering) depends on the center of mass energy of the collision and nature of the beams: at the LHC the gluon scattering process dominates (∼90% of the cases)[19]. Each top quark decays into a W boson and a b-quark with a branching ratio of almost 100%. The two b-jets originating from the b-quarks can give a similar signature as the VBF jets and the electrons, muons or τ leptons, decay products of W boson, can be misidentified as Higgs decay products. The W boson may decay into leptonic final states (e, μ, τ plus neutrinos) with a branching ration of 32.4% or to hadrons with a branching ratio of 67.6%[45] resulting to both leptons and jets in the final state faking VBF Higgs boson signal events. An efficient b-jet veto (section 1.3.6) would thus provide a good rejection of $t\bar{t}$ background and similarly a good hadronic tau-id (section 1.3.3) would prevent the misidentification of a jet originating from W or b-quark, as a tau jet. The most challenging aspect of this background is that the leptonic decay of W involves neutrinos, resulting in missing transverse energy. A cut in $\not{E}_T$ will have thus, no effect in rejecting this background contrary to $Z \rightarrow l^+l^-$ events. Nevertheless, due to different kinematics $t\bar{t}$ event topology adequate cuts (section 4.6) a reasonable rejection is achieved, but because

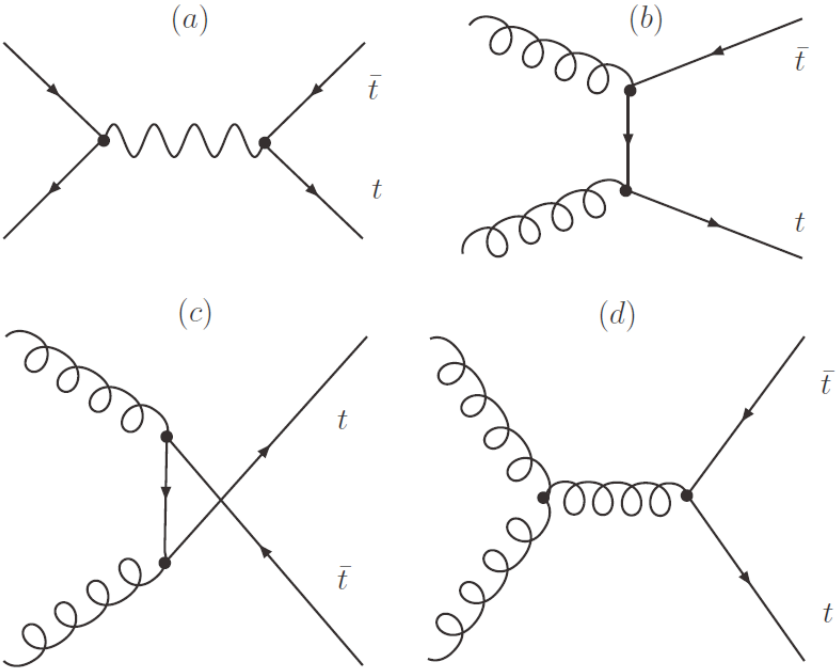


Figure 4.7: *Feynman diagrams at tree level of* $t\bar{t}$ *background.Both* $q\bar{q} \rightarrow t\bar{t}$ *and* $gg \rightarrow t\bar{t}$ *are taken into consideration.*

of its high production cross-section (833±100 pb[1]), $t\bar{t}$ remains one of the main background sources.

W(W) + jets

An other potential background can originate from the production of one or two W bosons with additional jets, contributing mainly to lh-channel. An example of two characteristic Feynman diagrams is given in figure 4.8 where in the W+jets case (figure 4.8,left) a third jet is produced via radiative corrections and in the WW+jets case (figure 4.8,right), a third originates from one of the W boson decay products. Considering then, a lepton W decay, a signature similar to the signal is being produced. A very effective cut against this kind of background is provided by the transverse mass m_T (section 4.6) of the lepton and $\not{E}_T$.

As shown in figure 4.8, the production of a WW pair can also provide a signature which is quite similar to the signal, considering one leptonic decay and one hadronic decay respectively for the two W bosons. However, the cross-section of this process is ∼20 times lower than the one for the Z boson production and this background can thus be neglected for this study.

QCD

In a proton-proton collider such as LHC, the cross-section for multijet production is

---

[1] calculated up to NLO order including NLL soft gluon resummation [62]. Uncertainty reflects the theoretical error obtained from varying the renormalisation scale by a factor of two [19] [63].

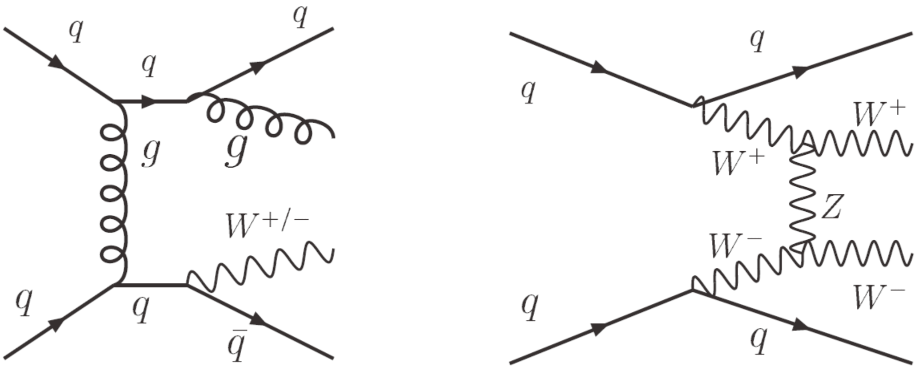


Figure 4.8: *Feynman diagrams of W + jets production by W radiation (left) and WW + jets production by triple electroweak boson interaction (right).*

very high[1]. Nevertheless, since no isolated leptons are present in these events, they can be suppressed requiring the two τ-decay products in the final state: either two isolated leptons (ee, $\mu\mu$ or $e\mu$), either one isolated lepton (e or μ) and a hadronic τ. In this category, potential background is caused by events containing misidentified leptons or leptons from b meson decays,as a result, an excellent performance of electrons, muons and τ-jets identification algorithms could efficiently suppress this background. The remaining $b\bar{b}$ containing real leptons from b decays can be reasonably suppressed by requiring that the leptons are isolated. In the scope of the present analysis, QCD background was neglected.

4.3 Monte Carlo samples - Detector simulation

In the following section, a description of all different Monte Carlo (MC) samples used in VBF analysis, is given. Different event generators have been used to simulate the signal and background processes incorporating all state of the art updates and NLO corrections. The simulated signal and background data samples and the corresponding cross-sections are summarized in tables 4.3, 4.4, 4.5 and 4.6. All datasets have been produced in the frame of the central ATLAS Monte Carlo production with the Athena software framework[2] [64].

The events generated by the Monte Carlo generators are then passed through the detailed simulation of the ATLAS detector based on the GEANT4 [65] package for the description of the detector response. In case of the $t\bar{t}$ background, the detailed detector simulation has been replaced by the parameterized and thus faster simulation of the detector response using ATLFAST [66] package. This was the only possibility to obtain a sufficient number of events needed for statistically significant analysis results.

4.3.1 Monte Carlo generators

A brief description of each of the used Monte Carlo generators is presented in the following:

---

[1] at the order of mb, see figure 1.21

[2] Athena release 12.0.6. All MC samples production was performed in the frame of ATLAS CSC report [19]

	k-Factor	Lepton Filter Efficiency (%)	Cross Section (fb)	Events	Integrated Luminosity (fb^{-1})
$H \to \tau\tau \to ll$					
$m_H = 105$ GeV	1.05	53	26	26 000	1000
$m_H = 110$ GeV	1.04	54	25	27 250	1090
$m_H = 115$ GeV	1.00	57	24	28 250	1180
$m_H = 120$ GeV	1.05	57	22	86 750	3940
$m_H = 125$ GeV	1.04	58	19	25 250	1330
$m_H = 130$ GeV	1.05	57	16	25 750	1610
$m_H = 135$ GeV	1.05	58	13	27 000	2080
$H \to \tau\tau \to lh$					
$m_H = 105$ GeV	1.05	46	180	67 500	380
$m_H = 110$ GeV	1.04	47	173	84 750	490
$m_H = 115$ GeV	1.00	47	160	28 250	180
$m_H = 120$ GeV	1.05	47	145	132 250	910
$m_H = 125$ GeV	1.04	48	127	86 500	680
$m_H = 130$ GeV	1.05	48	108	26 500	250
$m_H = 135$ GeV	1.04	48	87	85 250	980

Table 4.3: *Signal Monte Carlo data samples used for this study considering the ll-channel and the lh-channel. All signal datasets include a lepton filter (see text for details). The cross-sections include the filter efficiency and are scaled to the next-to-leading order with the specified k-factor described in second column (branching ratios also included). The last column shows the integrated luminosity the generated numbers of events correspond to.*

- Pythia [67]

  A general-purpose generator for hadronic interactions in leading order with parton showering for initial and final state QCD radiation, which leads to relatively soft jets. No spin correlations are taken into account.

- Herwig [68]

  A leading order general-purpose generator with a different parton showering compared to Pythia. Herwig is particularly sophisticated in treating the decays of unstable particles with full spin correlation. The Jimmy program [69] is used to simulate the underlying event. Most of the signal samples have been simulated with Herwig.

- ALPGEN [70]

  A leading order generator with a different method to match the jets from the parton-showering model to the ones from matrix-element calculations, leading to harder jets.

	VBF Filter Efficiency (%)	Cross Section (fb)	Events	Integrated Luminosity (fb^{-1})
ALPGEN QCD samples:				
$Z \to \tau\tau \to lh$				
+ 0 jets	0.3	2 750	11 250	4.1
+ 1 jets	1.0	2 450	92 250	37.7
+ 2 jets	4.0	3 100	515 500	166.0
+ 3 jets	10.6	2 500	356 000	142.0
+ 4 jets	20.7	1 380	218 750	158.0
+ ≥5 jets	29.6	730	90 000	122.0
$Z \to \tau\tau \to ll$				
+ 0 jets	0.04	320	9 850	31.0
+ 1 jets	0.13	280	161 800	574.0
+ 2 jets	0.43	330	439 900	1340.0
+ 3 jets	1.12	260	156 000	590.0
+ 4 jets	2.08	140	172 100	1230.0
+ ≥5 jets	3.30	80	48 450	599.0
$Z \to \tau\tau \to ee$				
+ 0 jets	1.0	9 060	13 750	1.5
+ 1 jets	2.8	6 450	7 500	1.2
+ 2 jets	9.9	7 420	252 250	34.0
+ 3 jets	25.8	6 110	103 250	16.9
+ 4 jets	44.8	2 990	23 250	7.8
+ ≥5 jets	64.0	1 580	14 000	8.9
$Z \to \tau\tau \to \mu\mu$				
+ 0 jets	0.6	4 990	13 400	2.7
+ 1 jets	2.4	5 410	11 250	2.1
+ 2 jets	8.2	6 260	192 250	30.7
+ 3 jets	23.7	5 610	86 500	15.4
+ 4 jets	42.6	2 830	21 500	7.6
+ ≥5 jets	61.9	1 520	15 250	10.0
Sherpa EW samples:				
$Z \to \tau\tau \to lh$ + <4 jets				
No filter	—	1 690	198 500	117.0
VBF filter	12.6	214	49 500	230.0

Table 4.4: *Z + jets, MC data samples used for this study considering QCD and EW production. The second column shows the VBF filter efficiency (see text for details). The cross-sections include the filter efficiency and are scaled to the next-to-leading order with a k-factor of 1.24 (branching ratios also included). The last column shows the integrated luminosity the generated numbers of events correspond to. The MC generators used are ALPGEN and Sherpa, explained in the text.*

ALPGEN QCD samples	VBF Filter Efficiency (%)	Cross Section (fb)	Events	Integrated Luminosity (fb^{-1})
$W \to \tau\tau \to e\nu$				
+ 2 jets	8.2	77 500	94 450	1.22
+ 3 jets	20.7	56 600	95 000	1.68
+ 4 jets	28.5	29 700	98 750	3.32
+ $\geq$5 jets	55.1	15 300	93 500	6.11
$W \to \tau\tau \to \mu\nu$				
+ 2 jets	7.5	71 000	97 950	1.38
+ 3 jets	19.4	53 100	77 750	1.46
+ 4 jets	36.7	28 000	58 500	2.09
+ $\geq$5 jets	54.7	15 300	96 950	6.35

Table 4.5: *W + jets, MC data samples used for this study considering QCD production using ALPGEN as a MC generator. The second column shows the VBF filter efficiency (see text for details). The cross-sections include the filter efficiency and are scaled to the next-to-leading order with a k-factor of 1.15 (branching ratios also included). The last column shows the integrated luminosity the generated numbers of events correspond to.*

Special emphasis is given to final states with large jet multiplicities, based on exact leading order evaluation of partonic matrix elements. It is also interfaced to Herwig/Jimmy for the hadronization and underlying event simulation. ALPGEN was used to simulate the QCD Z+jets and W+jets background processes. In practice, Z+0,1 jets and W+0,1 jets samples have not been used. In most of the cases, the background contribution originating from these specific samples is negligible with the respect to the contribution of the samples Z+n jets and W+n jets, where $n \geq 2$.

- Sherpa [71]

  A recent multipurpose generator for hadronic events with the matrix element generator AMEGIC++ [72]. Sherpa also includes electroweak processes, therefore it was used to simulate the electroweak $Z \to \tau\tau$ background.

- MC@NLO [73]

  A next-to-leading order generator for QCD processes. Like for ALPGEN, MC@NLO was also interfaced to Herwig/Jimmy for the hadronization and underlying event simulation. Since higher order effects are particularly important for $t\bar{t}$ production, MC@NLO was used to generate this background.

ATLAS Detector Simulation	Filter	Cross Section (fb)	Events ($\times 10^3$)	Integrated Luminosity (fb^{-1})
Full Simulation	1 lepton	461×10^3	930	1.5
Fast Simulation	—	833×10^3	94 300	83.0

Table 4.6: *$t\bar{t}$ MC data samples used for this study considering. MC@NLO was the MC generator used for this process and therefore the cross-section is at next-to-leading order accuracy. No filter has been applied in case of the fast detector simulation and for the detailed simulation a lepton filter has been applied (see text for details). The last column shows the integrated luminosity the generated numbers of events correspond to, where the event weights from MC@NLO have been taken into consideration.*

All τ-lepton decays have been simulated by Tauola Monte Carlo package [74], which is interfaced with the generators described above.

Event Filter

In order to increase signal and background statistics in the phase space of interest, event filters have been applied at the generator level to save CPU time by eliminating events that would not survive the particle identification cuts (for signal) and the analysis cuts (for background) or in general the detector acceptance limitations.

Lepton filter for signal samples:

- If N_l stands for the number of electrons plus muons with $p_T \geq$5 GeV and $|\eta| \leq$2.7:
  $N_l \geq 2$ for the samples $H \rightarrow \tau\tau \rightarrow ll$
  $N_l \geq 1$ for the samples $H \rightarrow \tau\tau \rightarrow lh$

VBF filter for background Z + jets and W + jets event samples:

- If N_l stands for the number of electrons plus muons with $p_T \geq$10 GeV and $|\eta| \leq$2.7:
  $N_l \geq 2$ for the samples $Z \rightarrow ee, \mu\mu$ and $Z \rightarrow \tau\tau \rightarrow ll$
  $N_l \geq 1$ for the samples $Z \rightarrow \tau\tau \rightarrow lh$
- At least two jets (cone algorithm with size $\Delta R \leq$0.4) within $|\eta| \leq$5.0 and
  $p_T \geq$20 GeV for the jet with the highest p_T
  $p_T \geq$15 GeV for the jet with the second highest p_T
- A pair of jets with an invariant mass of $m_{jj} \geq$300 GeV and a pseudorapidity gap of $\Delta\eta_{jj} \geq$2

Lepton filter for $t\bar{t}$:

- At least one lepton (e, μ or τ) within $|\eta| \leq 5.0$

All filter efficiencies are shown in tables 4.3, 4.4, 4.5 and 4.6 for the totality of the used samples. They are thus taken into account when measuring the final cross sections after the application of the analysis cuts (see section 4.9).

Pile-up Events

As described in the first chapter, at the LHC design luminosity of L = 10^{34} cm^{-2}s^{-1}, 23 proton-proton collisions are expected to occur every 25 ns. These collisions are inelastic scattering events (minimum bias), which are also recorded together with the interesting interaction in the same bunch crossing, forming the pile-up events. The additional particles from pile-up events potentially cause difficulties in the reconstruction of the hard parton collision processes especially to the jet reconstruction. For instance, the energy determined for a jet from the interesting event could be higher due to energy deposits from pile-up particles or even, the pile-up events could result in additional (low p_T usually) jets in the event.

If the pile-up originates from the same bunch crossing as the main interaction, it is called in-time pile-up. Contrary to the in-time pile-up, the contribution originates from adjoint bunch crossings, is called out-of-time pile-up. The latter can be the case for several detector components with a response and read out time longer than 25 ns (LAr calorimeter or the drift tube detectors of the muon system). Other pile-up contributions may occur from the cavern background or from showers induced by particles from cosmic rays especially effecting muon reconstruction and in some cases jets reconstruction.

For the VBF $H \to \tau^+\tau^-$ analysis both categories of pile-up may affect the results. A way to eliminate out-of-time pile-up is provided by time measurements in the calorimeter. It is though rather a complicated study, not sufficiently validated yet. In-time pile-up can cause a higher effect considering that adds a contribution to the central calorimeter sampling and it is visible from all ATLAS sub-detectors.

A special care was shown to pile-up studies and therefore several pile-up samples were generated as described in table 4.7 corresponding to the initial LHC luminosity performance L = 10^{33} cm^{-2}s^{-1}.

4.3.2 Detector Simulation

The detector response to the generated particles can be simulated by either a detailed detector simulation (so called Full Simulation)[64] or the fast detector simulation ATLFAST[66].

4.3.2.1 Full Simulation

The full simulation of the ATLAS detector response is based on a detailed and realistic description of particle interactions with the detector material as provided by the GEANT4 simulation package[65]. A detailed description of the detector geometry and material distribution is used for the simulation, as well as for the exact map of the magnetic field, the

	Pile-up Luminosity ($cm^{-2}s^{-1}$)	Cavern Background Simulation	Cross Section (fb)	Events
$H \rightarrow \tau\tau \rightarrow ll$				
$m_H = 120$ GeV	10^{33}	yes	22	66 250
$m_H = 120$ GeV	10^{33}	no	22	42 250
$m_H = 120$ GeV	2×10^{33}	no	22	44 000
$H \rightarrow \tau\tau \rightarrow lh$				
$m_H = 120$ GeV	10^{33}	yes	145	89 500
$m_H = 120$ GeV	10^{33}	no	145	49 500
$m_H = 120$ GeV	2×10^{33}	no	145	49 500
$Z \rightarrow \tau\tau$ (VBF filter applied)				
+ 2 jets	10^{33}	no	2 500	98 250
+ 3 jets	10^{33}	no	2 010	85 250
+ 4 jets	10^{33}	no	1 100	36 250
+ ≥5 jets	10^{33}	no	592	8 000

Table 4.7: *Systematic samples produced for pile-up studies for signal and the most important background Z + jets. As in simple samples, lepton and VBF filters are applied to signal and background samples respectively.*

realistic response of the read out electronics and the trigger behavior. The simulation output is then processed by the same reconstruction algorithms which will be used for the real collision data. Full simulation processes results in the best possible predictions but it is consuming a lot of time and CPU. The full simulation of one single event, typically needs approximately 20 minutes on a standard computer, therefore due to computing resources limitations, only a limited number of events can be fully simulated.

4.3.2.2 ATLFAST

In order to simulate a large number of events needed for some of the background processes (for example 100 millions of $t\bar{t}$ events), the much faster detector simulation software ATLFAST has been used. It is based on a parameterized description of the detector performance. The momentum vectors of the generated particles are smeared and the identification efficiency is parameterized according to studies with the full detector simulation performed for the most important physics particle objects. A homogeneous magnetic field is assumed in the inner detector volume and a very simplified calorimeter model is used to simulate the calorimeter response. No difference between electromagnetic and hadronic calorimeter and no longitudinal segmentation of the calorimeters is taken into account. Calorimeter granularity is approximated as $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ for $|\eta| < 3.2$ and $\Delta\eta \times \Delta\phi = 0.2 \times 0.2$ for $3.2 < |\eta| < 5.0$. The energy of a generated shower particle is deposited in only one calorimeter cell and a cone cluster algorithm with a cone size of $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} = 0.4$ is applied to all cells with a deposited energy above the threshold of 1.5 GeV. The resulting clusters are associated with the originally generated electrons, photons, jets or τ-jets.

The fast simulation of the detector takes approximately ~0.1 s per event using only 1% of the CPU Full Simulation is consuming. Clearly, the results obtained by the fast simulation depend on the accuracy of the parameterization of the detector response and are not as realistic as the ones obtained from full simulation. Moreover, as mentioned already, ATLFAST is the only way of ATLAS simulation for statistics which exceed a few million events. Fast simulation has been used in this analysis for the top sample, after normalisation to full simulation.

4.4 Analysis software framework: EventView

The performed analysis was made within the EventView package developped for the scope of the VBF $H \rightarrow \tau^+\tau^-$analysis called "HiggsToTauTau". EventView[30] is an ATHENA (section 1.2.5) physics analysis framework, designed within ATLAS community the years 2004-2005. The conceptual definition of EventView was formulated as described by its developers [30] is summarised as follows:

An EventView is a collection of physics objects, which are coherent, exhaustive and mutually exclusive. EventViews are not unique; for each event a user may wish to consider the event with multiple different views. From this view, a user may wish to calculate several quantities for instance, such as jets with a cone ΔR=0.4 and associate it with another view of the same event where jets are reconstructed using jet algorithms of ΔR=0.7

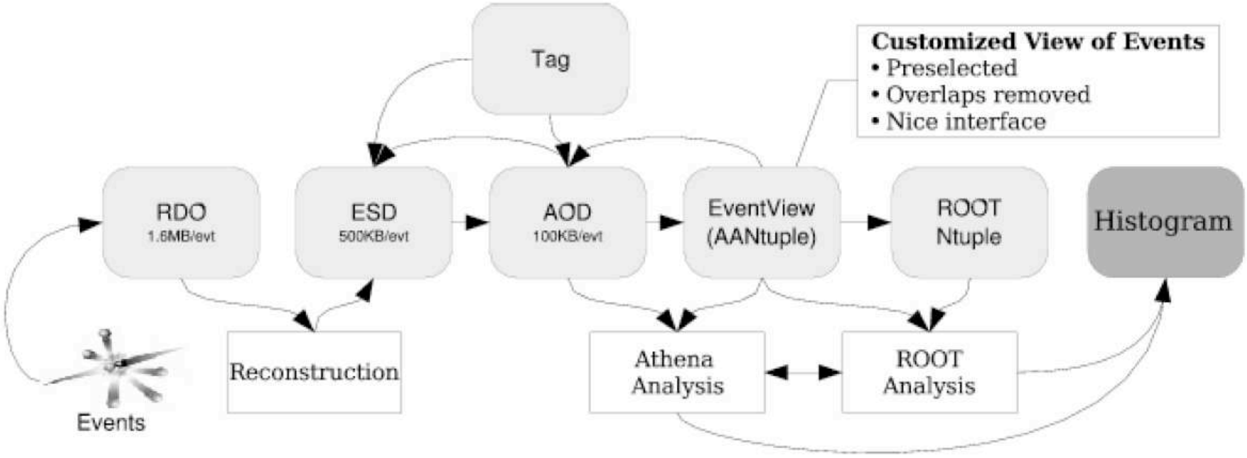


Figure 4.9: *EventView placed in the events flow from their production until the final histogram. The notation of RDO, ESD, AOD, Tag is referring to different types of event collections depending on the information they contain.*

A diagrammatic representation of events flow from their production until the final histogram is given in figure 4.9. As illustrated, EventView is the link between the particle collections (RDO,ESD,AOD,Tag) and the ROOT analysis. The main idea is that the framework provides general tools for common analysis tasks like particle selection and overlap removal, observable calculation, combinatorics, recalibration, systematics evaluation, and generating ntuples. The result is that users can perform the biggest part of their analyses in Athena by chaining and configuring a set of these tools and producing an ntuple for further analysis in ROOT. Furthermore, the EventView framework gives the opportunity to users to easily develop and mix their own C++ tools with the common EventView tools and share their configurations and tools with other collaborators.

One of the most important aspects of EventView is the overlap removal. The issue is that a single physical object might be reconstructed by several algorithms and included in multiple particle containers. For instance, an electron can be also reconstructed as a jet in the jets collection. This is what is meant by "overlap": One must take care to remove the overlap before continuing with the analysis. The difficulty is that there is no general purpose view of the event and therefore the correct procedure is analysis-dependent. EventView is giving a general way of introducing the overlap removal in every physics analysis. The overlap removal is mainly done by the so-called "Particle Inserters", which are basic EventView components whose use is summarized as follows:

i. retrieve a collection of particles (electron collection, muon collection, etc)

ii. loop through the collection of particles

iii. apply particle-type-specific preselection

iv. check if the preselected particles match any other particle in the EventView in some ΔR cone

v. if preselected and no overlap, insert the particle into the EventView

This procedure allows one to specify the priority of the different particle types (e.g. if an electron is also reconstructed as a tau and a jet, one can choose to insert either the electron, tau, or jet into the EventView). It also guarantees that particles overlap is removed in sense of ΔR and therefore a particle that is reconstructed as different types of particles is only entered once at the EventView.

Following the ΔR size of every reconstructed object defined within reconstruction algorithms, the ΔR thresholds used in the present VBF $H \to \tau^+\tau^-$ analysis are set as the followings:

i. electron: No ΔR cut. All electron objects contained in the electron collection enter the EventView.

ii. muon: ΔR=0.1. For all electrons inserted in previous step a ΔR defined as $\Delta R = \sqrt{(\eta_\mu - \eta_{e_i})^2 + (\phi_\mu - \phi_{e_i})^2}$ is calculated for every muon candidate found in the muon collection, and if ΔR <0.1 is found an overlap between electron and muon is assumed and thus the muon is not inserted in the EventView.

iii. hadronic τ: ΔR=0.2 (for the lh-channel).

iv. jets: ΔR=0.4, which in coherence with the size of the jet cone chosen for the present analysis (see following section).

Depending on the analysis, the order of entering the objects in the EventView may change according to specific physics needs. For VBF $H \rightarrow \tau^+\tau^-$, the order presented above was chosen favoring the insertion firstly of electrons, then muons and then hadronic τ's and QCD jets.

4.4.1 Acceptance Challenge within EventView

One of the first goals of the ATLAS group VBF $H \rightarrow \tau^+\tau^-$was to converge in the same numbers of cut flow cross sections, within the different analysis tools used by the collaborators. This was a very important exercise in order to validate all the different ATLAS analysis tools available at that time, comprehend the different particle identification algorithms and event selection cuts and finally, verify that all collaborators, having used the same event datasets for signal and background are finally measuring the same number of events after the cuts application. The validation achieved within the scope of this PhD study, was related with the analysis software framework used which was the EventView package. It was a long procedure demanding high interaction with all members of VBF $H \rightarrow \tau^+\tau^-$group who were using different analysis software tools, which nevertheless, resulted to a very good understanding of the cuts performance (section 4.6).

4.5 Reconstruction of VBF tagging jets

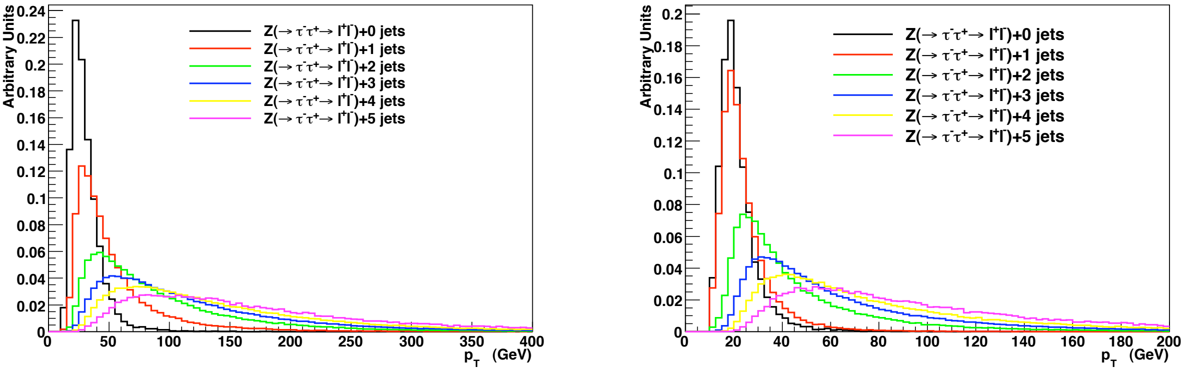
As mentioned in the introduction, the jet activity of the vector boson fusion process is unique in several ways, providing many handles to suppress backgrounds and isolate a sample of signal events with high purity. One of the most characteristic aspects of VBF processes is the presence of two quark-initiated jets (figure 4.2 “forward jets” or “VBF tagging jets”) which due to the process topology, have the tendency to be produced in the forward regions of the detector, resulting to high $l\eta|$ values introducing a pseudorapidity gap in between.

Moreover, another very important jet feature of VBF process is the central jet veto. Due to the electroweak character of Vector Boson Fusion (colorless exchange), additional QCD radiation activity is expected to be low. As a consequence, no other jets apart from the tagging ones, are expected[1], and thus the application of a central jet veto could result to a significant rejection of background. More details are given in section 4.10

A general remark concerning the jet multiplicities of the background processes W+jets and Z+jets is illustrated in figure 4.10. The p_T of the highest p_T jet and second highest p_T jet is shown for $Z+n$jets, for n≥0. Contributions originating from the +0 jet and +1 jet show a low p_T peak and considering that the basic requirement of VBF tagging jets is p_T ≥40 GeV and p_T ≥20 GeV for the highest and second highest p_T jet of the event (discussed later), these two contributions may be neglected. Similar results are obtained for the $W + 0, 1$ jets and therefore $+0, 1$ jets contributions for W and Z background are from now on neglected.

---

[1] not taking into consideration potential jets that might originate from the tau decay products such as the hadronic tau jets



(a) (b)

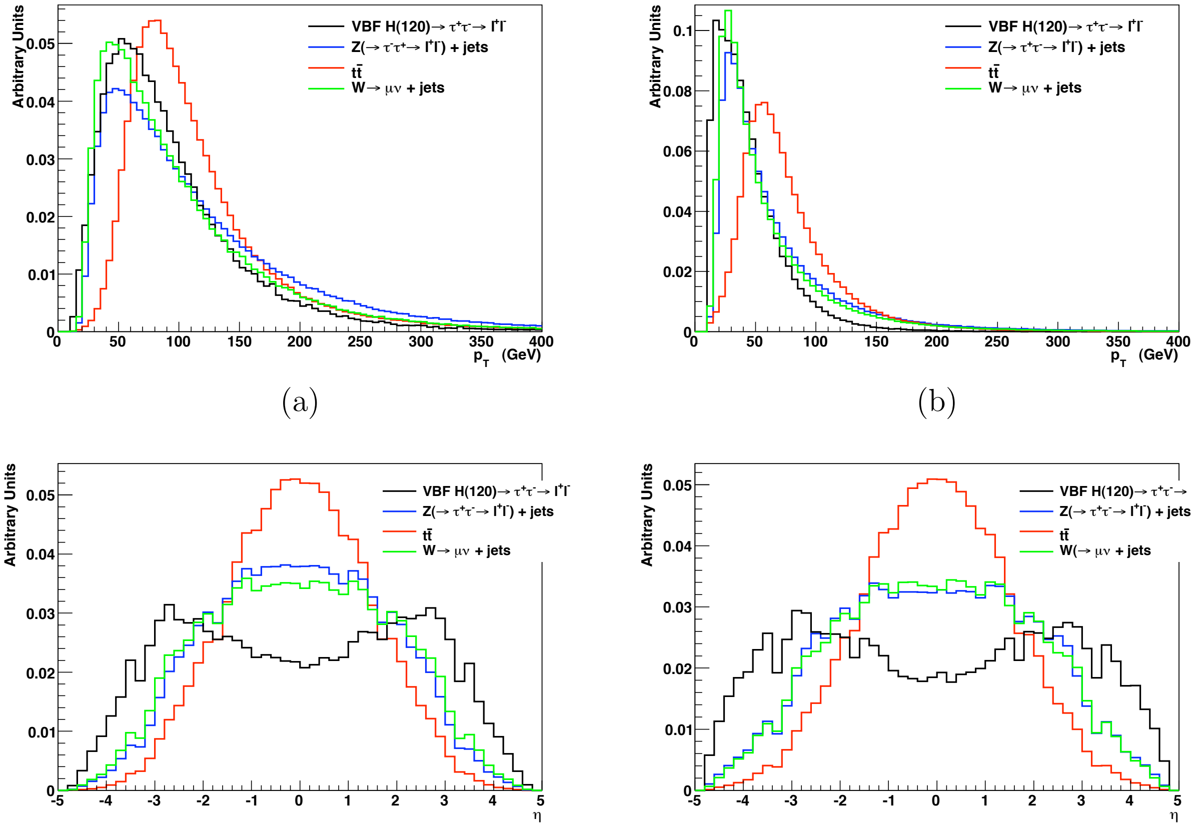
Figure 4.10: *Jet p_T spectra of highest (a) and second highest (b) p_T jets of different samples $Z(\rightarrow \tau\tau \rightarrow ll) + n$ jets, where $n \geq 0$.*

The p_T and η spectra of the highest and second highest p_T jets in signal for ll-channel (figure 4.11), signal for lh-channel (figure 4.12) superimposed with various background samples is shown. For the p_T distributions (figures 4.11a,b and 4.12a,b) no p_T or η cuts are applied. Comparing these two figures, one could observe that VBF jet kinematics are almost identical in both channels. A peak at $\sim$ 60 GeV ($\sim$ 25 GeV) is observed for the jet with the highest (second highest) p_T . In both channels, in the scope of identifying the VBF tagging jets two p_T cuts are applied:

- $p_T \geq$ 40 GeV for the jet with the highest p_T
- $p_T \geq$ 20 GeV for the jet with the second highest p_T

These two p_T cuts are applied when plotting the η distributions of the two VBF tagging jets (figures 4.11)c,d and 4.12c,d). VBF jets have the tendency to be spread in all eta values showing a preference in pseudorapidity values: $|\eta|$ $\sim$3, while background tagging jets are more central (especially in $t\bar{t}$ events). Comparing VBF jet η distributions for the two different signal samples, a small difference is observed in the central region, where lh sample gives slightly more VBF jets. The reason is that in lh-sample, some of the jets, which originate from the τ, are misidentified as normal QCD jets. As explained in the VBF topology, these τ jets are expected to be central.

In section 1.3.5, the main topics of jet reconstruction in ATLAS were presented. A VBF analysis requires a jet finding efficiency in the forward region of the calorimeter, which leads to different challenges for jet reconstruction since efficiency drops in these regions. After number of studies within the VBF Higgs to tau tau ATLAS group [19], the most adequate jet algorithm was found to be the TopoCluster jets with a cone size of 0.4. TopoCluster jets compared to the Tower jets show a better performance at the forward region of the calorimeter (figure 4.13) since Tower jets require a seed in E_T whereas Topoclusters as explained in section 1.3.4.2 require a seed in E. This is possible thanks to the better noise suppression of topocluster algorithm with respect to Tower algorithm. A p_T cut is applied for all jets at the level of EventView and it is equal to 15 GeV.



(a) (b)

(c) (d)

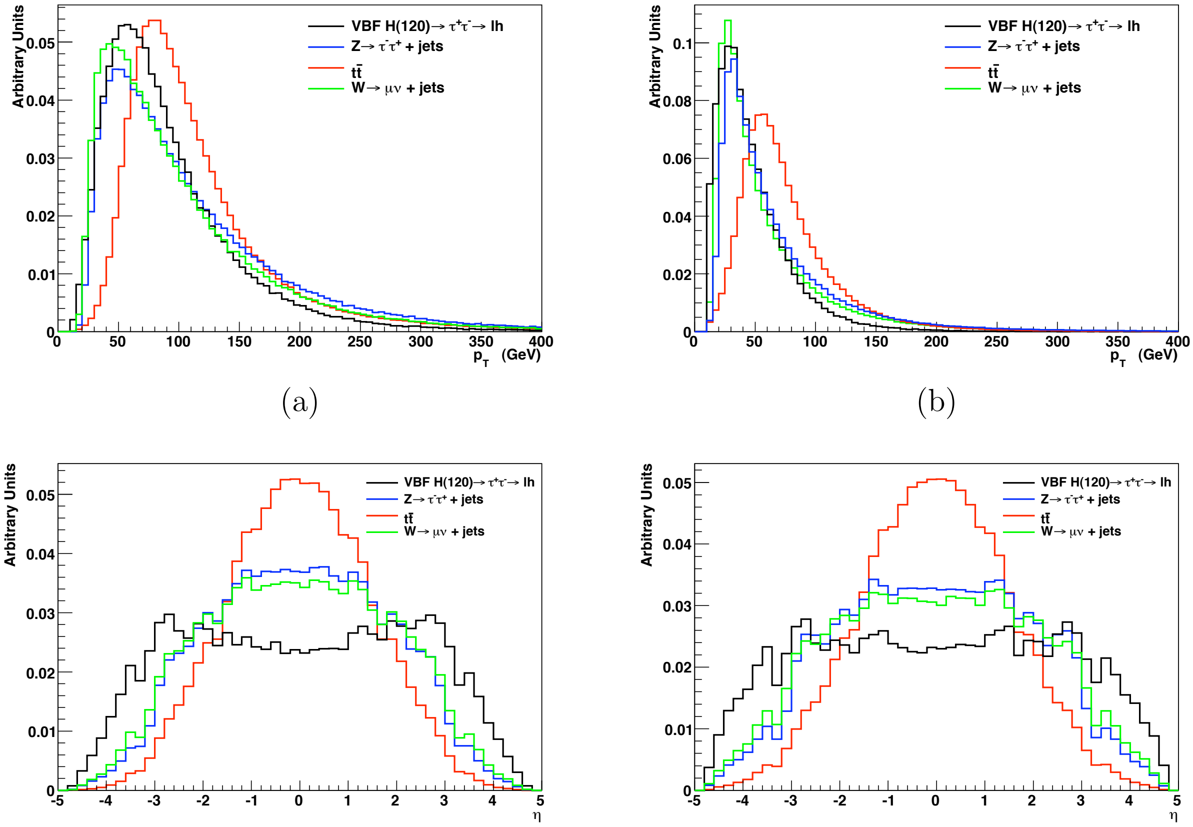
Figure 4.11: *Jet p_T spectra of highest (a) and second highest (b) p_T jets in signal (ll-channel) and various background processes, no p_T or η cuts are applied. The respective plots of jet η distribution of the highest (c) and second highest (d) p_T jets are also presented. In this case additional p_T cuts are applied: $p_T > 40$ GeV is required for the highest p_T jet and $p_T > 20$ GeV is required for the second highest p_T jet. For W+jets and Z+jets background, the number of jets varies from 2 to 5.*

4.5.1 Identification of VBF tagging jets

As shown already, successfully identifying the quark-initiated tagging jets from the VBF process is very important for the VBF $H \rightarrow \tau^+\tau^-$study. Typically, the VBF tagging jets are found in opposite hemispheres, but there are two approaches to incorporating this requirement in the analysis. One option is to define the tagging jets as the two highest p_T jets in the event (figures 4.11a,b and 4.12a,b), and reject the event from the signal candidates if they are in the same hemisphere[1]. A second possibility is to define the first tagging jet to be the highest p_T jet in the event and the second tagging jet to be the highest p_T jet in the opposite hemisphere. In this second approach it is not required that the second tagging jet is the second highest p_T jet in the event. These two strategies were compared, and it was

---

[1] therefore require $\eta_{j_1} \times \eta_{j_2}$ <0



(a) (b)

(c) (d)

Figure 4.12: *Jet p_T spectra of highest (a) and second highest (b) p_T jets in signal (lh-hannel) and various background processes, no p_T or η cuts are applied. The respective plots of jet η distribution of the highest (c) and second highest (d) p_T jets are also presented. In this case additional p_T cuts are applied: $p_T > 40$ GeV is required for the highest p_T jet and $p_T > 20$ GeV is required for the second highest p_T jet. For W+jets and Z+jets background, the number of jets varies from 2 to 5.*

found that the first method matched the quark-initiated tagging jets from the hard process in a more realistic way.

After having identifying the VBF jets with the method described in the previous section, some useful variables are plotted in following figure 4.14. Figure 4.14,a illustrates the pseudorapidity gap ($\Delta\eta_{jj}$) between the two tagging jets. The tendency of VBF jets in signal implied in figures 4.11c,d and 4.12c,d is confirmed with a peak of $\Delta\eta_{jj}$ at ∼4.8, while the respective gap in background samples is smaller. For background events, the VBF filter (section 4.3) is applied and it explains the spike at $\Delta\eta_{jj}$ =2.2. $\Delta\eta_{jj}$ is one of the discriminating variables between signal and background with a threshold equals to 4.4. Similarly, figure 4.14,b shows the invariant mass of the di-jet system (M_{jj}) defined by the two tagging jets. Similarly to $\Delta\eta_{jj}$ distribution, the VBF filter applied in background is responsible for the peak at m_{jj} =300 GeV. A separation signal versus background is observed for the value

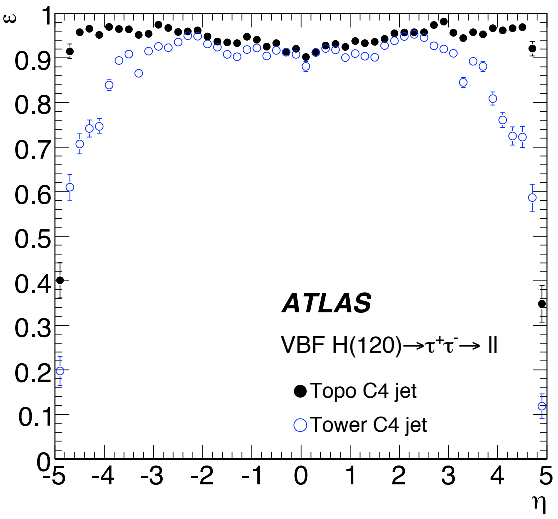


Figure 4.13: *Jet reconstruction efficiency for the Cone jet algorithm with* ΔR= *0.4 as a function of the generator-level jet* η *for the jets based on TopoClusters and Towers [19].*

M$_{jj}$=600 GeV which indicates another cut will be used further in the event selection. Signal sample used for these figure is a m_H=120 GeV decaying in the ll-channel. Results are similar in the lh-channel giving the same rejection thresholds.

4.5.2 VBF tagging jets (η, ϕ, p_T) distributions

An interesting point regarding the VBF tagging jets is to explore their kinematics in terms of η, ϕ, and p_T. Results are obtained using a dataset of m_H=120 GeV, where $H \rightarrow \tau^+\tau^- \rightarrow ll$.

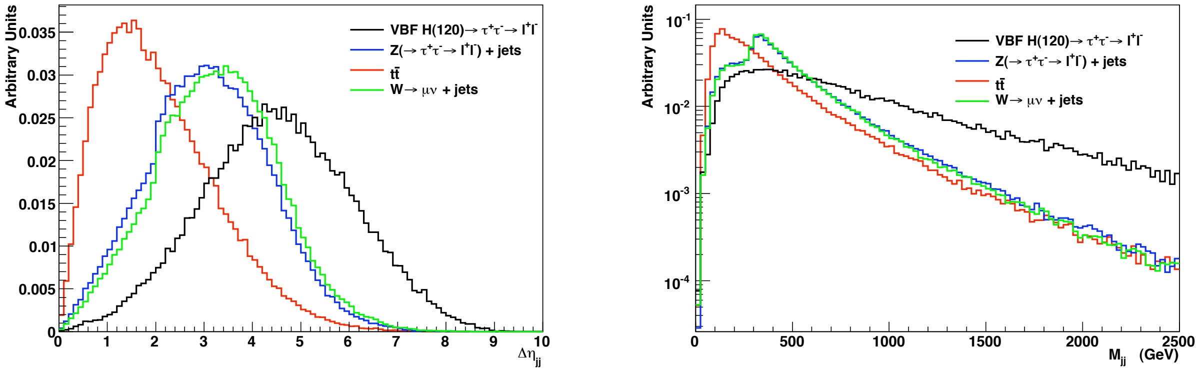
Figures 4.15 and 4.16 show the highest p_T jet distribution in an (η,p_T) plane for signal and background samples. One can observe immediately the tendency of $t\bar{t}$ to give central jets, while signal sample shows a preference within the forward regions of the detector. On the other hand, investigating the Z+jets, it is shown that the jet kinematics, in the topology which contributes with 2 jets (figures 4.16a), are highly similar with the ones of the signal. For the +3,+4, or +5 jets contributions, a clear difference is observed between background and signal, resulting in the presence of the majority of jets in the central region of the detector for the background samples. The not shown W+jets distributions are similar to the Z+jets one.

Figure 4.17, show the respective distributions of the azimuthal separation between the two leading jets. The two jets tend to be back to back for p_T balancing reason, for signal as for backgrounds.

4.6 Event Selection

4.6.1 Trigger

As described in section 1.2.4, the trigger in ATLAS consists of three levels, with the third level (event filter) performing the final decision. For this analysis two different triggers have



(a) (b)

Figure 4.14: *Pseudorapidity gap between tagging jets (a) and invariant mass of di-jet system composed by the tagging jets (b).p_T cuts applied: p_T > 40 GeV for the highest p_T jet and p_T > 20 GeV for the second highest p_T jet and also $\eta_{j_1} \cdot \eta_{j_2}$ <0 is required. For W+jets and Z+jets background, the number of jets varies from 2 to 5. The VBF filter is applied.*

been used:

- Isolated electrons with p_T ≥22 GeV ("e22i").
- Isolated muons with p_T ≥20 GeV ("mu20").

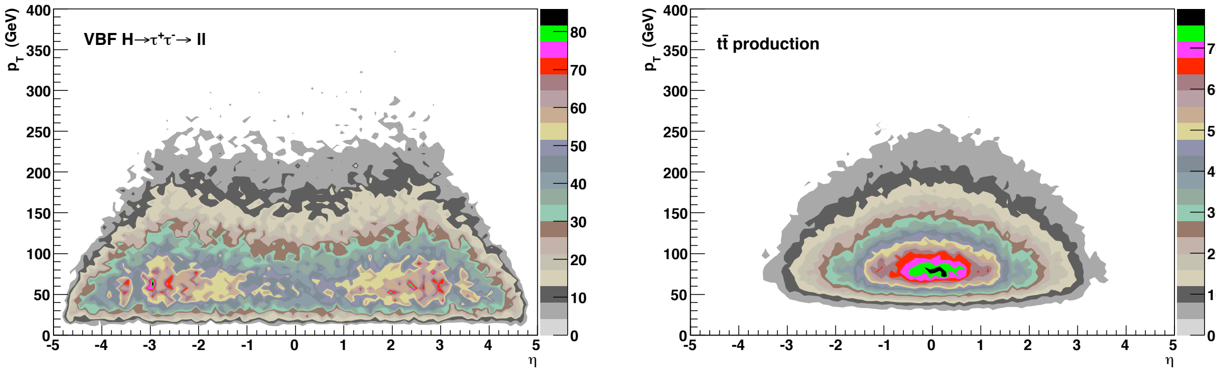
The trigger efficiency for VBF $H \rightarrow \tau^+\tau^-$(with m_H=120 GeV) is 9.0% for events selected by the electron trigger and 9.9% in the case of muons [19]. The trigger efficiencies include detector acceptance and are normalized with respect to the production cross-section for VBF $H \rightarrow \tau^+\tau^-$.

4.6.2 Electron reconstruction and identification

Electron candidates are formed from a cluster of cells in the electromagnetic calorimeter together with a matched track as described in details, in section 1.3.1. In this analysis, the level of electron quality "Medium cuts" is used, since it provides sufficient fake rejection and provides a higher signal efficiency. In addition to the standard electron identification, the energy in an isolation cone of radius ΔR=0.2 around the electron, is required to contain less than 10% of the electron's E_T . The reason for applying the isolation cut is to reject the contamination from hadronic jets. Finally, the p_T of the offline reconstructed lepton is set to satisfy the p_T threshold of the corresponding trigger, which is not strictly enforced due to slight differences between the offline reconstruction and the trigger algorithms. A summary of electron identification requirements is given in table 4.8.

4.6.3 Muon reconstruction and identification

As described in section 1.3.2, muon candidates can be seeded from either tracks in the inner detector or in the standalone muon spectrometer. In VBF $H \rightarrow \tau^+\tau^-$the highest quality



(a) (b)

Figure 4.15: *(η,p_T) distribution of the jet with the highest p_T in the event for a Higgs boson (m_H=120 GeV) sample, where $H \rightarrow \tau^+\tau^- \rightarrow ll$ (a) and a $t\bar{t}$ sample (b). No extra cuts are applied.*

Electron ID & requirements
Electron ID: Medium isolation E_T (ΔR=0.2)/p_T $\leq$0.1
p_T $\geq$25 GeV for trigger electron (“e22i”)
p_T $\geq$15 GeV for other electrons

Table 4.8: *Summary of the identification requirements for electrons in VBF $H \rightarrow \tau^+\tau^-$ analysis.*

muon candidate is required (“Staco Muon”):

i. by extrapolating the track in the muon spectrometer to the interaction point

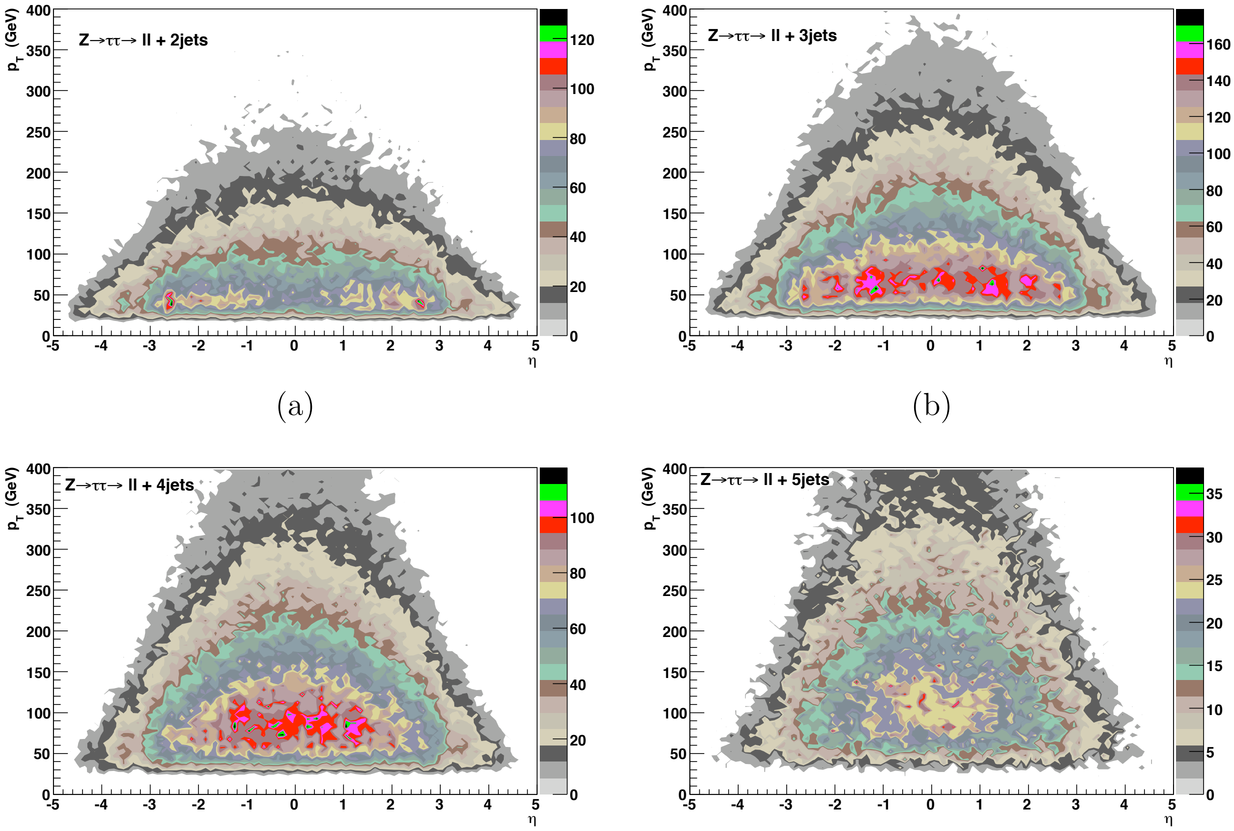
ii. finding a matching inner detector track, and

iii. forming a combined track if the two tracks satisfy various quality requirements.

The muon identification is composed of requirements on track quality and hit multiplicity in several muon stations. Similarly to the electrons, we require an isolation condition that the summed E_T within a radius ΔR of 0.2 is less than 10% of the muon p_T in order to reject the contamination from jets. Finally, the p_T of the offline reconstructed muon is set to satisfy the p_T threshold of the corresponding trigger. A summary of muon identification requirements is given in table 4.9.

4.6.4 τ-jet reconstruction and identification

Since, approximately 65% of tau lepton decays produce hadrons, τ-jet reconstruction and identification plays an important role within the analysis in the di-tau decay of Higgs boson.



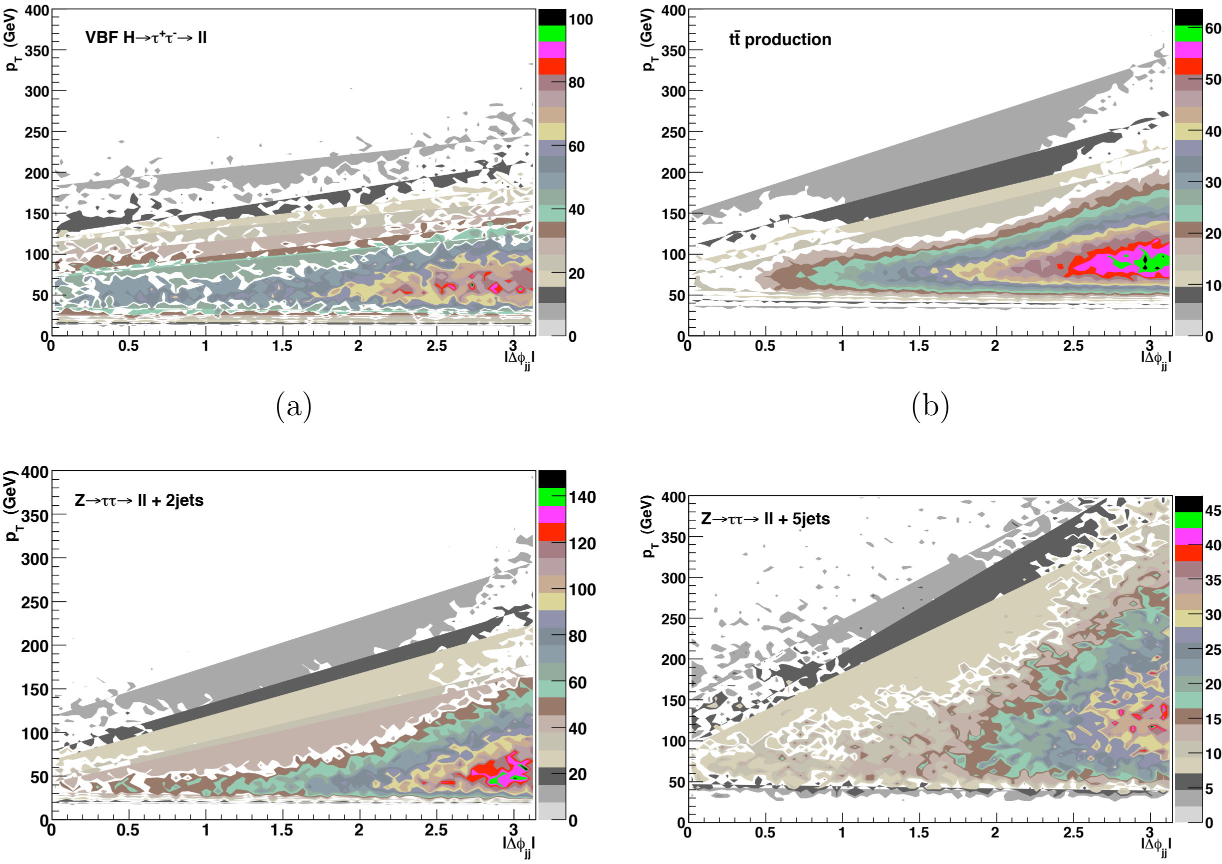
(a) (b)

(c) (d)

Figure 4.16: *(η,p_T) distribution of the jet with the highest p_T in the event for $Z \to \tau\tau \to ll$+2jets (a), $Z \to \tau\tau \to ll$+3jets (b), $Z \to \tau\tau \to ll$+4jets (c), $Z \to \tau\tau \to ll$+5jets (d). No extra cuts are applied.*

ATLAS hadronic τ ID is described in details in section 1.3.3. As already mentioned, the majority of hadronic tau decays are composed of single-prong candidates with one charged pion, which provides a track and a hadronic shower, and potentially associated neutral pions that provide an additional electromagnetic sub-cluster. In addition, three-prong tau decays are also reconstructed, but with a higher rate of fakes from QCD jets. Due to the high momentum of the taus produced in this process, the decay products are collimated into a narrow region.

In this analysis, the calorimeter-based algorithm was used. The calorimeter-seeded algorithm uses a log-likelihood ratio that provides discriminating power from a variety of track quality and shower shape information to discriminate between taus and jets. The discriminating variable is designed to maintain a high tau efficiency while rejecting fake tau candidates from jets, leaving the precise working point to be optimized in the context of a specific analysis. The cuts on the discriminating variable and p_T of the tau candidates were optimized within the VBF $H \to \tau^+\tau^-$analysis group [19], with respect to a simple $s/\sqrt{s+b}$ performance measure. The background sample included Z+jets, W+jets, and $t\bar{t}$ +jets, which



(a) (b)

(c) (d)

Figure 4.17: *($|\Delta\phi_{jj}|$,p_T) distribution between the two highest p_T jets in the event for a m_H=120 GeV, where $H \rightarrow \tau^+\tau^- \rightarrow ll$ (a), $t\bar{t}$ (b), $Z \rightarrow \tau\tau \rightarrow ll$+2jets (c), $Z \rightarrow \tau\tau \rightarrow ll$+5jets (d). No extra cuts are applied.*

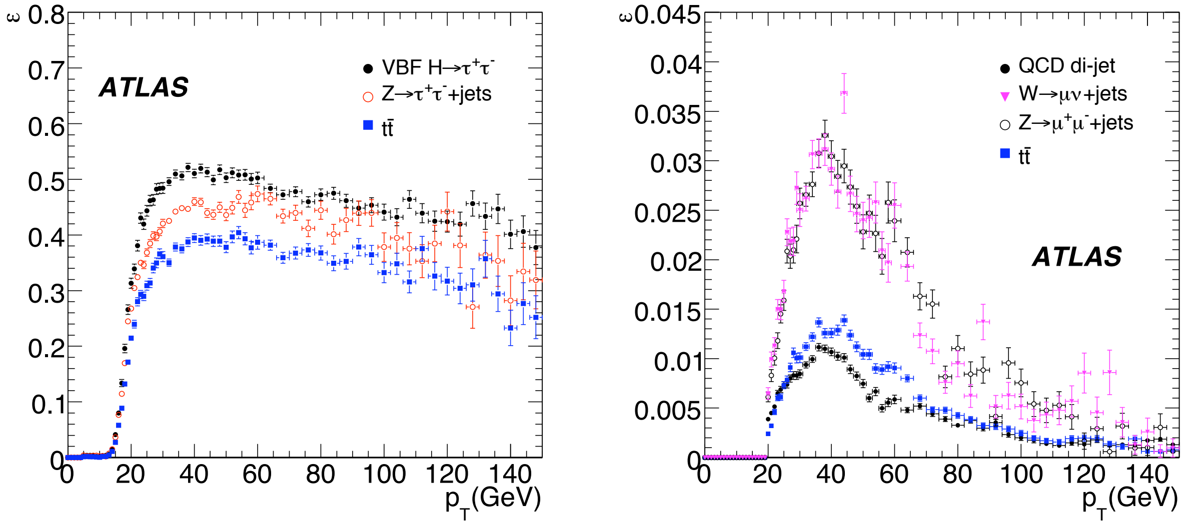
comprises a background sample with a representative mixture of real and fake taus. The jet fragmentation modeling indicates that quark-initiated jets (such as the VBF tagging jet, for instance) are more collimated and have a 6-8 times higher fake rate than gluon-initiated jets.

Several subsets of the final event selection criterion were evaluated, and the final optimization was found to be reasonably stable and nearly independent of p_T . After the optimization, the calorimeter-seeded algorithm's log-likelihood ratio was required to be greater than 4, corresponding to an identification efficiency of 50.0±0.2% and a fake jet selection efficiency of 1% for gluon-initiated jets and 2.5% for quark-initiated jets.

Finally, we present the hadronic tau reconstruction and identification performance in figure 4.18(a) and the fake-jet tagging rate (b) as a function of p_T , respectively. The reason why $t\bar{t}$ shows a worse performance is due to the fact that there are plenty central jets in the final state topology faking τ's. A summary of the hadronic τ identification requirements is given in table 4.10.

Muon ID & requirements
Muon ID: Combined Muon isolation E_T (ΔR=0.2)/p_T $\leq$0.1
p_T $\geq$20 GeV for trigger muon ("mu20i")
p_T $\geq$10 GeV for other muons

Table 4.9: *Summary of the identification requirements for muons in VBF* $H \to \tau^+\tau^-$ *analysis.*



(a) (b)

Figure 4.18: *Reconstruction and identification efficiency of the hadronic tau (a) and the jet-fake rejection efficiency (b) as a function of* p_T*, respectively [19].*

4.6.5 *b*-jet veto

In the *ll*-channel, the largest background contribution comes from $t\bar{t}$ (+jets)$\longrightarrow$ *lvblvb* (+jets). By introducing a veto on b-tagged jets (section 1.3.6) it is possible to reduce this background. A likelihood ratio approach is used to estimate the probability that a jet originated from a *b*-quark and in this analysis, a cut of >1.0 on the output value of this *b*-jet identification algorithm (*b*-tag weight) is applied to separate *b*-jets from light quark jets.

This cut is reasonably efficient and the $t\bar{t}$ background can be reduced by a factor ∼2-3. Figure 4.19 demonstrates the efficiency of the *b*-jet veto as a function of the leading VBF tagging jet p_T for the signal and $t\bar{t}$ background. A significant rejection (∼70%) is achieved for the $t\bar{t}$ background, while approximately 90% of the signal events survive the *b*-jet veto. To be noted that the *b*-jet veto is only applied in the *ll*-channel.

Hadronic τ ID & requirements
Muon ID: Calorimeter-seeded
p_T $\geq$30 GeV
Track multiplicity : 1 or 3 tracks
$\lvert$charge$\rvert$=1
Log Likelihood Ratio ≥ 4
Electron Veto:
minimum TRT HT/LT $\leq$0.2 if $\lvert\eta_\tau\rvert$ $\leq$1.7 and LT $\geq$10
E_T^{HAD}/p_T $\geq$0.002 in matched electron object

Table 4.10: *Selection criteria and requirements for the hadronic τ identification from the calorimeter-based reconstruction algorithm for VBF $H \rightarrow \tau^+\tau^-$analysis.*

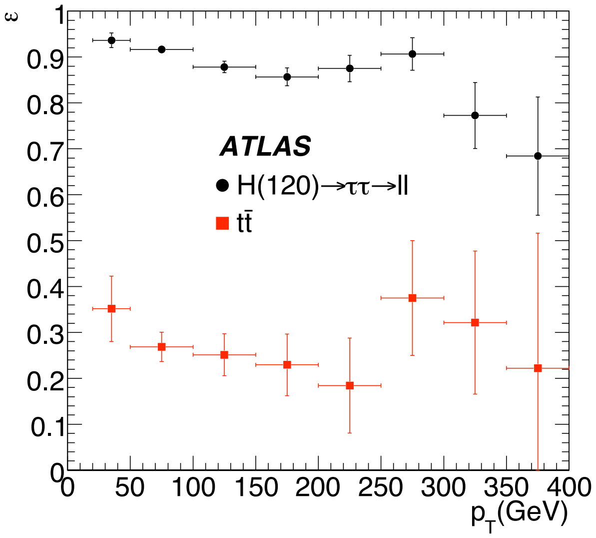


Figure 4.19: *Efficiency of the b-jet veto as a function of the leading VBF tagging jet p_T for the signal and $t\bar{t}$ background [19].*

4.6.6 Missing transverse energy ($\not{E}_T$) reconstruction

Due to the presence of leptons τ, significant missing transverse energy ($\not{E}_T$) is expected in $H \rightarrow \tau^+\tau^-$events because neutrinos are always associated with the τ decays. There is one neutrino found in hadronic τ decay final state and two in the leptonic τ decay. Details regarding $\not{E}_T$ reconstruction are given in section 1.3.4 and in particular in $H \rightarrow \tau^+\tau^-$analysis, the performance of the $\not{E}_T$ algorithm plays an important role in this analysis because $\not{E}_T$ is used in the mass reconstruction of the τ pair (section 4.7). Finallly, the $\not{E}_T$ resolution is what limits the $m_{\tau\tau}$ resolution, which implies the need of a perfect understanding of $\not{E}_T$ resolution. By requiring a large $\not{E}_T$, it is possible to reduce many backgrounds that do not contain neutrinos. The rejection power of $\not{E}_T$ variable is illustrated in figure 4.20, where $\not{E}_T$ variable is plotted for signal (m_H=120 GeV, $H \rightarrow \tau^+\tau^- \rightarrow ll$), and background ($t\bar{t}$, $Z \rightarrow ll$ + jets) samples. The only requirements asked for this plot are the existence of VBF

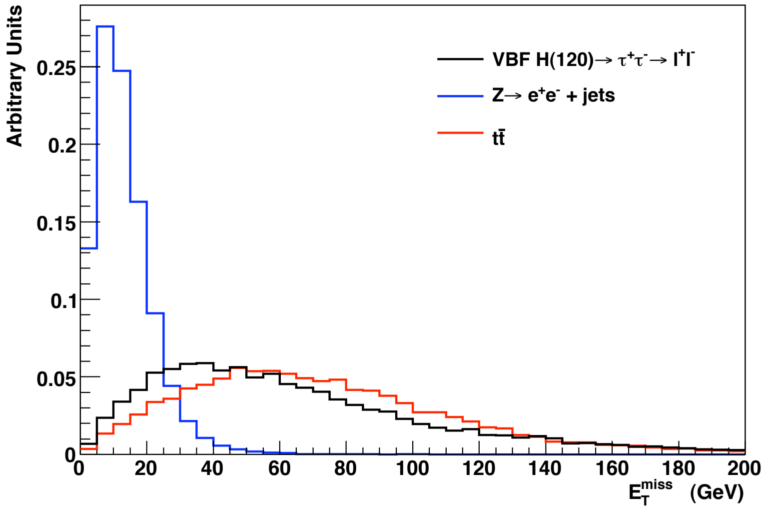


Figure 4.20: *$\not\!\!E_T$ distribution for signal (m_H=120 GeV, $H \rightarrow \tau^+\tau^- \rightarrow ll$), and background ($t\bar{t}$, $Z \rightarrow e^+e^-$ + jets) samples. The existence of two VBF tagging jets in opposite hemispheres satisfying the respective p_T cuts and the existence of two exactly isolated leptons (e or μ) of opposite charge, are required.*

tagging jets in opposite hemispheres satisfying the respective p_T cuts, plus the existence of two exactly isolated leptons (e or μ) of opposite charge. As expected, $\not\!\!E_T$ variable shows a significant rejection power against $Z \rightarrow l^+l^-$ since no neutrinos are expected, while the respective rejection power is poor against $t\bar{t}$. The reason is the lepton filter applied at the level of $t\bar{t}$ generation.

$\not\!\!E_T \geq$ 30 GeV for the lh-channel and $\not\!\!E_T \geq$ 40 GeV for the ll-channel, is finally required.

4.7 Reconstruction of the Higgs Mass

Although there are several neutrinos in the event, it is possible to reconstruct the $\tau^+\tau^-$ invariant mass ($m_{\tau\tau}$) by making the approximation that the decay products of the τ are collinear with the τ in the laboratory frame. This is a good approximation since $m_H >> 2m_\tau (m_\tau$=1.777 GeV [45]) and hence the τ's are highly boosted. Therefore, the direction of the emitted neutrinos can be assumed to be the same as the direction of the visible τ-decay products (electrons, muons or τ-jets). The splitting up of the total missing transverse energy into the two neutrino contributions from the two τ decays and the collinearity of the neutrinos with the visible τ decay producs, is illustrated in figure 4.21 for the ll-channel (μ^-e^+ decay).

The momenta fractions χ_1 and χ_2 carried by the visible τ-lepton decay products with respect to the initial τ-lepton momenta are defined as:

$$\chi_1 = \frac{\overrightarrow{p}^1}{\overrightarrow{p}^{\tau_1}} \quad , \quad \chi_2 = \frac{\overrightarrow{p}^2}{\overrightarrow{p}^{\tau_2}} \tag{4.3}$$

where $\overrightarrow{p}^{\tau_{1(2)}}$ represents the momenta of the τ-leptons and $\overrightarrow{p}^{1,(2)}$ the momenta of the visible

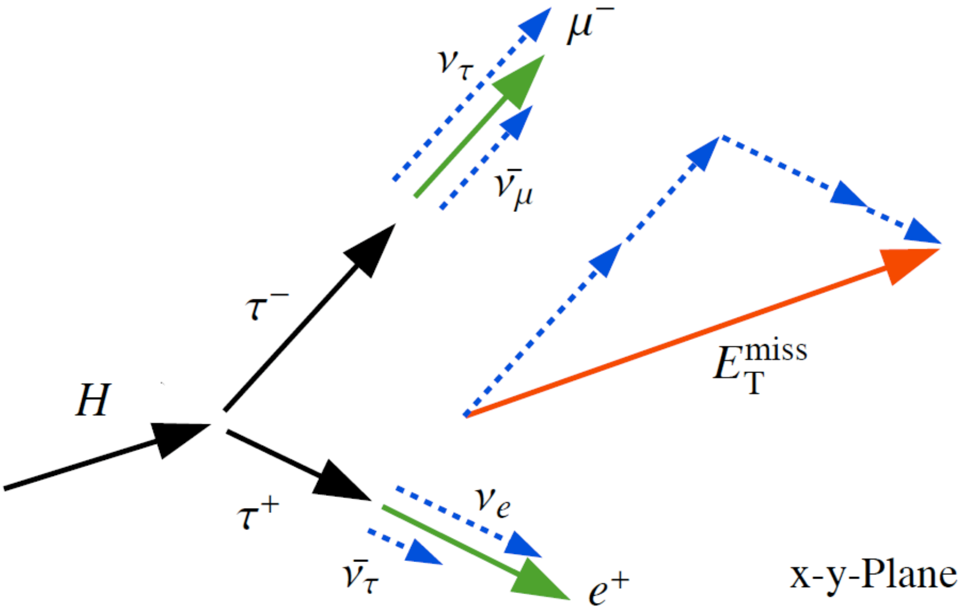


Figure 4.21: *Sketch of the collinear approximation and the reconstruction of the invariant mass of the di-tau system ($m_{\tau\tau}$). Due to the strong boost of τ leptons, it can be assumed that visible and non-visible (neutrinos) τ-decay products are emitted in the flight direction of the decaying τ. This direction can be extracted by the direction of the visible τ-decay products (in this example: μ^- and e^+). The measured missing transverse energy ($\not{E}_T$) is then separated into the two neutrino direction contributions ($\nu_\tau + \bar{\nu}_\mu$ and $\bar{\nu}_\tau + \nu_e$) from the two τ decays. This allows for the complete reconstruction of the τ-lepton four-momenta and thus of the Higgs boson mass. The $x - y$ plane is perpendicular to the beam direction z.*

τ-lepton decay products (electron, muon or τ-jet). Similarly moving to transverse (x,y) plane, τ-lepton momenta can be written as follows:

$$p_x^{\tau_1} = p_x^1 + p_x^{\nu_1} \quad , \quad p_x^{\tau_2} = p_x^2 + p_x^{\nu_2} \tag{4.4}$$

$$p_y^{\tau_1} = p_y^1 + p_y^{\nu_1} \quad , \quad p_y^{\tau_2} = p_y^2 + p_y^{\nu_2} \tag{4.5}$$

where $p_{x,(y)}^{\nu_{1,(2)}}$ are the x and y momenta components of the neutrinos originating from the $\tau_{1(2)}$ lepton.

Since the missing transverse energy expected to originate from the neutrinos, one can write for the x,y components of E_T^{miss}:

$$E_x^{miss} = p_x^{\nu_1} + p_x^{\nu_2} \tag{4.6}$$

$$E_y^{miss} = p_y^{\nu_1} + p_y^{\nu_2} \tag{4.7}$$

and using equations 4.3, 4.4 and 4.5, the x,y components of E_T^{miss} can be written as:

$$E_x^{miss} = p_x^1 \frac{1 - \chi_1}{\chi_1} + p_x^2 \frac{1 - \chi_2}{\chi_2} \tag{4.8}$$

$$E_y^{miss} = p_y^1 \frac{1-\chi_1}{\chi_1} + p_y^2 \frac{1-\chi_2}{\chi_2} \tag{4.9}$$

Solving this system of equations one obtains:

$$\chi_1 = \frac{p_x^1 p_y^2 - p_x^2 p_y^1}{p_y^2(p_x^1 + E_x^{miss}) - p_x^2(p_y^1 + E_y^{miss})} \tag{4.10}$$

$$\chi_2 = \frac{p_x^1 p_y^2 - p_x^2 p_y^1}{p_x^1(p_y^2 + E_y^{miss}) - p_y^1(p_x^2 + E_x^{miss})} \tag{4.11}$$

Therefore, one can easily calculate the momentum fractions χ_1 and χ_2 of the visible τ-decay products using equations 4.10 and 4.11.

Moving to $m_{\tau\tau}$ calculation, if the τ rest mass is neglected and the collinear approximation is imposed, one can write:

$$m_{\tau\tau} = \sqrt{2p_{\tau_1}p_{\tau_2}(1 - cos\Delta\phi_{\tau\tau})} \tag{4.12}$$

and incorporating the χ_1, χ_2 variables:

$$m_{\tau\tau} = \frac{m_{12}}{\sqrt{\chi_1\chi_2}} \tag{4.13}$$

where m_{12} is the invariant mass of the system of the visible τ decay products.

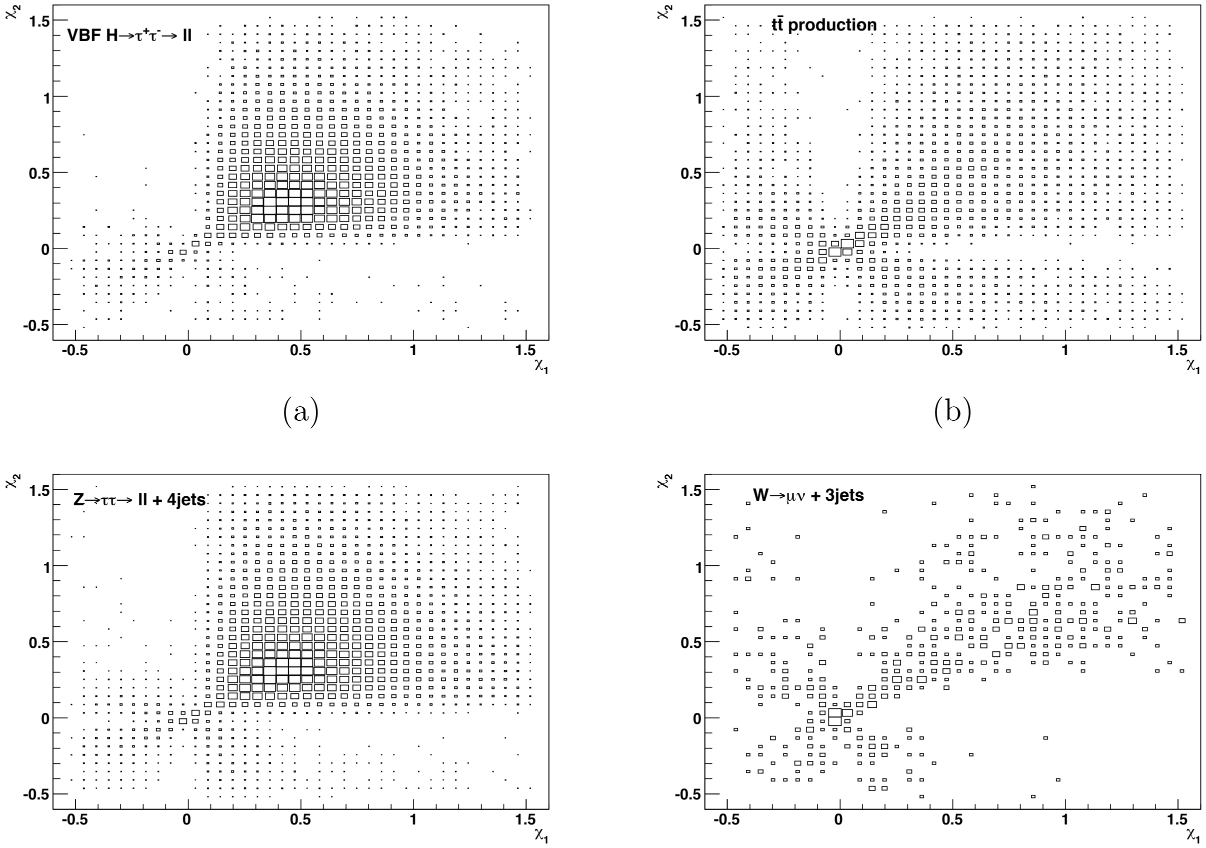
Figure 4.22 illustrates the χ_1, versus χ_2 distributions for several samples of signal and background. To be noted that resolution effects in $\not{E}_T$ may lead to unphysical solutions with either $\chi_{1(2)} < 0$ or $\chi_{1(2)} > 1$. Only p_T - η cuts are applied for VBF tagging jets and exactely 2 leptons are required for 4.22(a), 4.22(b) and 4.22(c) while exactly one lepton is required for 4.22(d), to reproduce these distributions. It is important to notice, that distributions are highly similar for signal (ll-channel) and $Z \to \tau\tau \to ll$ case. The same number of neutrinos in the event final state(4 neutrinos are expected: 2 ν_τ and 2 ν_l) results to similar $\not{E}_T$ distribution in both samples (see figure 4.20 as well) and as a consequence similar $\chi_{1(2)}$ momenta transfer to the visible τ decay products is expected. To conclude, a cut of 0≤ $\chi_{l1(l2)}$ ≤0.75 is demanded for the ll-channel and a respective cut of 0≤ χ_l ≤0.75 and 0≤ χ_h ≤1.0 is required for the lh-channel. The tighter cut on 0≤ $\chi_{l1(l2)}$ ≤0.75 has been found to provide a better background rejection in the ll-channel [19].

If the two τ's are back-to-back, then the equations 4.8 and 4.9 are linearly-dependent and one cannot solve to find the $\chi_{1(2)}$. This is the reason why an extra cut of cos$\Delta\phi_{\tau\tau}$ ≥-0.9 is required. Equation 4.13 shows explicitly that potential cuts on $\chi_{1(2)}$ will impose constraints on the reconstructed mass for a given event, which results in an asymmetric distribution for $m_{\tau\tau}$.

4.8 Cut factorization method

Due to the lack of sufficiently large Monte Carlo generated samples for the background a cut factorization method was used to estimate the background rate at the end of the analysis. For this method, cuts are grouped into four categories:

- Categories I. - II. : cuts that are approximately uncorrelated:



(a) (b)

(c) (d)

Figure 4.22: *Distributions for χ_1, versus χ_2 for: signal events (m_H=120 GeV $H \to \tau^+\tau^- \to$ ll)(a), $t\bar{t}$ events (b), $Z \to \tau\tau \to ll$+4jets events (c) and $W \to \mu\nu$+3jets events (d). p_T - η cuts are applied for VBF tagging jets and exactely 2 leptons are required for (a), (b) and (c) while exactly one lepton is required for (d).*

  - Category I : cuts related to the τ decays from the Higgs boson candidate (trigger, lepton ID, hadronic tau ID, $\not{E}_T$, the collinear approximation, and the transverse mass). The rejection power of this group of cuts is mainly dominated by detector performance issues.
  - Category II : cuts related to the VBF tagging jets (forward jets, jet separation, and dijet mass). In this case the rejection is dominated by the jet kinematic properties.

- Category III : contains cuts related to both the τ decay products and the VBF tagging jets (centrality, central jet veto, and mass window cut).

- Category IV : consists of the fundamental selection cuts, which are related to leptons and jets (dilepton veto, jet multiplicity and forward jets without centrality).

First, the background cut efficiency is determined for each of the two first uncorrelated categories (I and II) individually. Then, the efficiencies of each cut of category III is determined separately with respect to the set of cuts of category IV. Finally, the product of all efficiencies gives the total background cut efficiency.

4.9 Summary of cut flow

4.9.1 *ll*-channel

The event selection for *ll*-channel is summarized below, including some kinematic requirements specific to the *ll*-channel, which result to the respective cuts:

- Trigger: electron trigger "e22i" or muon trigger "mu20".
- Trigger lepton: at least one lepton must have a reconstructed p_T greater or equal to the corresponding trigger requirement.
- Dilepton: exactly two identified leptons with opposite charge.
- Missing E_T : E_T^{miss} $\geq$40 GeV.
- Collinear approximation: 0$\leq$ $\chi_{l1,l2}$ $\leq$0.75 and cos$\Delta\phi_{\tau\tau}$ $\geq$-0.9.
- Jet multiplicity: at least one jet with p_T $\geq$40 GeV and at least one additional jet with p_T $\geq$20 GeV.
- VBF tagging jets: in opposite hemispheres $\eta_{j_1} \times \eta_{j_2}$ $\leq$0, with τ centrality min$(\eta_{j_1}, \eta_{j_2})\leq$ $\eta_{lep_{1,2}}$ $\leq$max(η_{j_1}, η_{j_2}) for the two highest p_T jets.
- *b*-jet veto: the event is rejected if either tag jet has *b*-tag weight greater than 1.
- Jet kinematics: $\Delta\eta_{jj}$ $\geq$4.4 and dijet mass m_{jj} $\geq$700 GeV for two VBF tagging jets.
- Central jet veto: the event is rejected if there are any additional jets with p_T $\geq$20 GeV in $|\eta|$ $\leq$3.2. Further details are given in following section 4.10.
- Mass window: m_H−15 GeV$\leq m_{\tau\tau} \leq m_H$+15 GeV around the test mass m_H.

Table 4.11 summarizes the cross-section for signal and background events after applying each of the cuts described above.

4.9.2 *lh*-channel

The respective event selection for *lh*-channel is summarized below, including some kinematic requirements specific to the *lh*-channel, which result to the following cuts:

- Trigger: electron trigger "e22i" or muon trigger "mu20".
- Trigger lepton: at least one lepton must have a reconstructed p_T greater or equal to the corresponding trigger requirement.

- Dilepton veto: exactly one identified lepton (ensures this sample is disjoint from the ll-channel).
- Hadronic τ: exactly one identified hadronic τ with opposite charge of the lepton (e or μ).
- Missing E_T : E_T^{miss} $\geq$30 GeV.
- Collinear approximation: 0$\leq$ χ_l $\leq$0.75, 0$\leq$ χ_h $\leq$1 and cos$\Delta\phi_{\tau\tau}$ $\geq$-0.9.

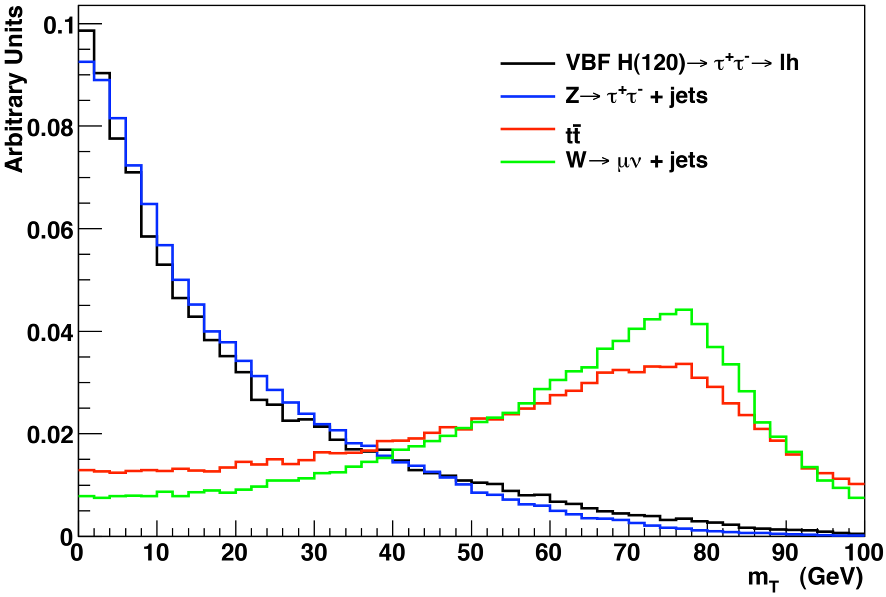


Figure 4.23: *m_T distribution in lh-channel for signal events (m_H=120 GeV, $H \rightarrow \tau^+\tau^- \rightarrow$ lh), $t\bar{t}$ events, $Z \rightarrow \tau\tau$+jets events and $W \rightarrow \mu\nu$+jets events. p_T - η cuts are applied for VBF tagging jets and the presence of exactly one isolated leptons is also required.*

- Transverse mass (m_T): in order to further suppress the W+ jets and $t\bar{t}$ backgrounds (see figure 4.23), the following cut on the transverse mass of the lepton and E_T^{miss} is required: m_T $\leq$30 GeV,
  where m_T, is defined as follows:

$$m_T = \sqrt{2p_T^{lep}E_T^{miss}(1 - cos\Delta\phi)} \quad . \tag{4.14}$$

  p_T^{lep} denotes the transverse momentum of the lepton in the lh-channel and $\Delta\phi$ is the angle between the lepton and E_T^{miss} in the transverse plane.

- Jet multiplicity: at least one jet with p_T $\geq$40 GeV and at least one additional jet with p_T $\geq$20 GeV.
- VBF tagging jets: in opposite hemispheres $\eta_{j_1} \times \eta_{j_2}$ $\leq$0, with τ centrality min(η_{j_1}, η_{j_2})$\leq$ $\eta_{lep_{1,2}}$ $\leq$max(η_{j_1}, η_{j_2}) for the two highest p_T jets.

- b-jet veto: the event is rejected if either tag jet has b-tag weight greater than 1.
- Jet kinematics: $\Delta\eta_{jj} \geq 4.4$ and dijet mass $m_{jj} \geq 700$ GeV for two VBF tagging jets.
- Central jet veto: the event is rejected if there are any additional jets with $p_T \geq 20$ GeV in $|\eta| \leq 3.2$. Further details are given in following section 4.10.
- Mass window: $m_H - 15\ \mathrm{GeV} \leq m_{\tau\tau} \leq m_H + 15$ GeV around the test mass m_H.

Table 4.12 summarizes the cross-section for signal and background events after applying each of the cuts described above.

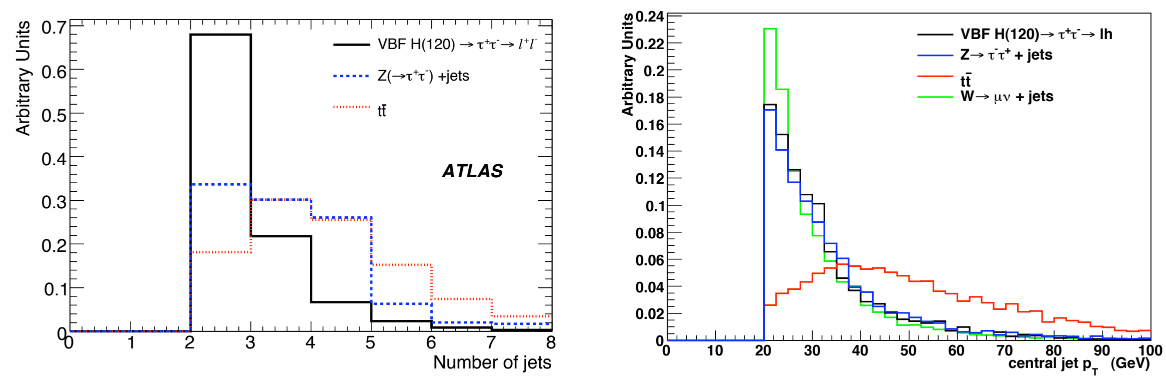
Cut	signal ll-chan. (fb)	$Z \to \tau\tau$ +n jets QCD(n≥1) (fb)	$Z \to \tau\tau$ +n jets EW(n≥1) (fb)	$t\bar{t}$ Full (fb)	$t\bar{t}$ Fast (fb)	$Z \to ll$ +n jets (n≥1) (fb)	$W \to l\nu$ +n jets (n≥1) (fb)
None	309.1	168.4×10^3	1693	833×10^3	-	768.6×10^3	8649×10^3
Trig.	57.2(1)	$51.5(1)\times10^3$	230(1)	$209.8(2)\times10^3$	-	$633.8(4)\times10^3$	$4411(9)\times10^3$
1 p_T	49.5(1)	$42.7(1)\times10^3$	190(1)	$179.1(2)\times10^3$	-	$588.0(4)\times10^3$	$3815(9)\times10^3$
N lep.	5.46(3)	$42.5(5)\times10^3$	19.2(4)	$21.7(1)\times10^3$	-	$369.9(5)\times10^3$	$2.5(2)\times10^3$
$\not{E}_T$	3.17(3)	744(18)	9.9(3)	16847(99)	-	2683(67)	1148(176)
Coll.	2.15(2)	454(14)	6.2(2)	1817(33)	Atlfast	104(12)	46(21)
N jets	1.77(2)	262(8)	5.8(2)	1722(32)	1699(4)	73(8)	14(6)
VBF	1.34(2)	39(2)	2.0(1)	294(13)	324(1)	10(3)	1.2(2)*
b-jet	1.16(2)	30(2)	1.5(1)	89(7)	90.3(9)	9(3)	1.0(2)*
Jets	0.63(1)	2.71(5)	0.57(5)	11.8(3)*	26.7(5)	0.66(3)*	0.19(4)*
CJV	0.56(1)	1.24(3)	0.43(4)	1.9(1)*	2.6(1)	0.27(1)*	0.10(2)*
m win.	0.45(1)	0.23(1)	0.04(1)	0.10(2)*	0.06(2)	0.058(3)*	0.01(1)*

Table 4.11: *Signal (m_H=120 GeV) and background cross sections for for the ll-channel. An asterisk is used to indicate cross sections estimated from the cut factorization method described in section 4.8. Fast simulation data is used for $t\bar{t}$ after collinear approximation cut.*

Cut	signal lh-chan. (fb)	$Z \rightarrow \tau\tau$ +n jets QCD(n≥1) (fb)	$Z \rightarrow \tau\tau$ +n jets EW(n≥1) (fb)	$t\bar{t}$ Full (fb)	$t\bar{t}$ Fast (fb)	$Z \rightarrow ll$ +n jets (n≥1) (fb)	$W \rightarrow l\nu$ +n jets (n≥1) (fb)
None	309.1	168.4×10^3	1693	833×10^3	-	768.6×10^3	8649×10^3
Trig.	57.2(1)	$51.5(1)\times10^3$	230(1)	$209.8(2)\times10^3$	-	$633.8(4)\times10^3$	$4411(9)\times10^3$
1 p_T	49.5(1)	$42.7(1)\times10^3$	190(1)	$179.1(2)\times10^3$	-	$588.0(4)\times10^3$	$3815(9)\times10^3$
N lep	43.4(1)	$38.4(1)\times10^3$	171(1)	$156.4(2)\times10^3$	-	$216.5(4)\times10^3$	$3811(9)\times10^3$
Hadr.τ	8.02(7)	3062(42)	19.3(4)	5224(56)	-	20250(156)	32537(1012)
$\not{E}_T$	4.96(5)	850(20)	12.1(3)	4251(50)	-	468(26)	21001(801)
Coll.	3.34(5)	514(15)	7.8(2)	606(19)	-	17(3)	324(46)
m_T	2.46(4)	415(13)	6.5(2)	176(10)	Atlfast	11(2)	67(18)
N jets	2.02(4)	235(7)	6.0(2)	162(9)	167(1)	8(1)	49(11)
VBF	1.52(3)	40(3)	2.3(1)	32(4)	26.1(4)	1.3(6)	2.9(3)*
Jets	0.82(2)	2.7(1)	0.72(6)	1.8(1)*	3.6(1)	0.10(1)*	0.7(1)*
CJV	0.72(2)	1.2(1)	0.49(5)	0.25(4)*	0.43(5)	0.047(6)*	0.43(6)*
m win.	0.61(2)	0.11(2)	0.04(1)	0.012(5)*	0.03(1)	0.08(1)*	0.020(6)*

Table 4.12: *Signal (m_H=120 GeV) and background cross sections for for the lh-channel. An asterisk is used to indicate cross sections estimated from the cut factorization method described in section 4.8. Fast simulation data is used for $t\bar{t}$ after m_T cut.*

4.10 Central jet veto cut study



(a) (b)

Figure 4.24: *(a) Jet multiplicity distribution for the signal, Z+jets, and $t\bar{t}$ background after requiring the cuts up to the N-jets level in the list of cuts for the ll-channel (see table 4.11). (b) p_T distribution of central jet which vetos the event due to CJV for signal and background in lh-channel. Jet and lepton cuts are applied (see table 4.12) apart from the hadronic τ one. Only jets with p_T ≥20 GeV are taken into consideration.*

4.10.1 Definition

The central jet veto is an important cut of VBF $H \rightarrow \tau^+\tau^-$analysis and an optimization study is presented in the following section. The idea of rejecting an event when extra jets are found, summarizes the description of this cut and apart from VBF $H \rightarrow \tau^+\tau^-$analysis, central jet veto cut is also used in the $Z \rightarrow ll$ analysis, in other VBF H decay channels such as W^+W^-, $\gamma\gamma$ or in some SUSY studies, which have clean leptonic signatures without no jet contamination. In a $p-p$ collider like LHC, QCD jets will contaminate every physical signature, and considering as well, the additional jets caused by the pileup from minimum bias interactions occurring close in time to the triggered event, one realizes how challenging is the application of such a cut.

As already mentioned, the central jet veto cut (CVJ) rejects an event if a third jet is found[1] within an η range of $|\eta|$ ≤3.2, having a p_T ≥20 GeV. If only two jets (the VBF tagging jets) are found in the event CJV performs no rejection. Figure 4.24(a) shows the jet multiplicity distribution for the signal and backgrounds after requiring two VBF tagging jets in opposite hemispheres and after having applied all the lepton related cuts (ll-channel). As expected, a clear peak for a number of two jets is observed for signal, while the fraction of signal events with three or more jets is small. On the other hand, backgrounds such as $t\bar{t}$ and Z+jets show a higher jet multiplicity implying a higher possibility to be rejected by CJV. Figure 4.24(b) shows the p_T distribution of the 3^{rd} jet of the event if p_T ≥20 GeV,

---

[1] apart from the two VBF tagging jets

which rejects the event. As expected the p_T of the 3^{rd} in signal, Z and W show similar behavior while in $t\bar{t}$, it is more energetic.

4.10.2 CJV alternatives

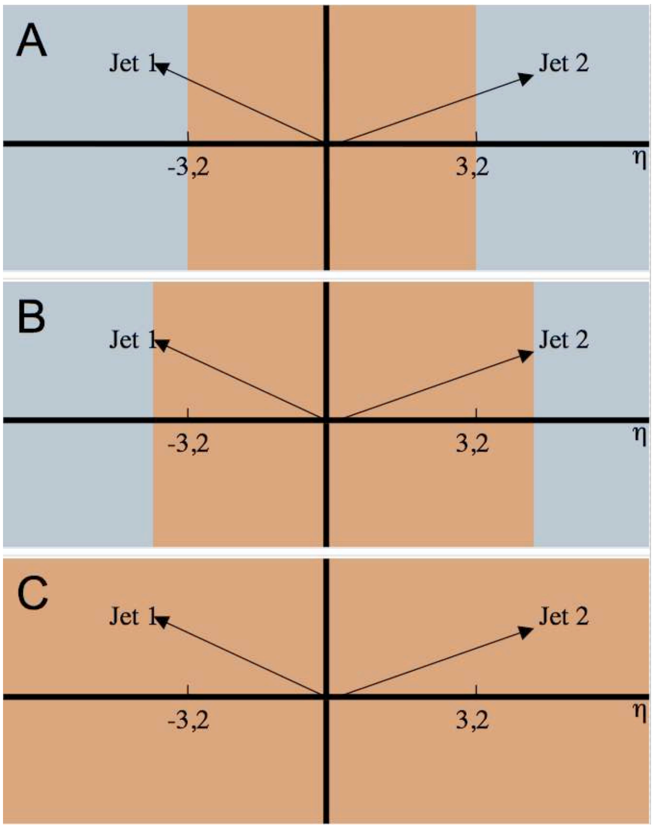


Figure 4.25: *Three different $\eta_{3^{rd}jet}$ examined regions for the CJV application. Brown colored region denotes the range of interest.*

The first goal of the CJV study was to optimize the p_T cut threshold of 20 GeV in addition to the optimization of the η range of searching the third jet. As illustrated in figure 4.25, three different choices of η range were studied:

A. $|\eta_{3^{rd}jet}| \leq 3.2$, which was the default region used within the VBF $H \rightarrow \tau^+\tau^-$group[1]

B. $\min(\eta_{j_1}, \eta_{j_2}) \leq \eta_{3^{rd}jet} \leq \max(\eta_{j_1}, \eta_{j_2})$. Similarly to the centrality cut defined in section 4.6, the η range of finding a 3^{rd} jet is not fixed like in the previous case, but it varies depending on the $\Delta\eta_{jj}$ rapidity gap of the two VBF tagging jets[2].

C. no constraints in $\eta_{3^{rd}jet}$. 3^{rd} jet can be found anywhere in the detector under the ATLAS hardware limitation: $|\eta| < 5$.

---

[1] so-called as “default”

[2] Since a $\Delta\eta_{jj} > 4.4$ is applied, η range cannot be lower than 4.4.

The jet identification p_T threshold was changed to 10 GeV and for each one of the three cases defined above, the CJV cut efficiency is calculated for signal and background using different cut values in the p_T of the 3^{rd} jet. In the following, the term "efficiency" is used with the interpretation as the number of events passing the CJV cut divided by the total number of events before the cut. Since it is the most dominant background, $Z \to \tau\tau$+jets was used to evaluate $\varepsilon_{background}$ and results to be presented are for lh-channel. CJV behavior is quite similar in ll-channel and therefore, it is not presented in this document.All the cuts are applied up to CJV (table 4.12) modifying some of the respective thresholds in order to gain statistics. In more details, these more loose cuts are defined as follows:

- VBF tagging jets: at least one jet with p_T ≥30 GeV and at least one additional jet with p_T ≥15 GeV.
- Missing E_T : E_T^{miss} ≥20 GeV.
- Transverse mass: m_T ≤50 GeV,

All the other cuts are applied with the same thresholds as described in section 4.9.2. Results are then validated with the default cut flow values and it was found that the CJV cut efficiencies were behaving the same.

4.10.3 CJV cut efficiency results

Figure 4.26 shows the results are obtained for the CJV cut efficiency obtained by considering the three different $\eta_{3^{rd}jet}$ constraints described in previous section. Each group of the three colored points corresponds to the same p_T of $_{3^{rd}jet}$ cut starting from 10 GeV increasing with a step of 5 GeV each time. Table 4.13 contains the detailed results of the CJV cut efficiency for signal and background until the p_T cut of 35 GeV. A higher p_T cut would be meaningless since CJV cut becomes inefficient. Values contained in the table are represented in the figure 4.26 and the goal was to manage to increase ε_{signal} while decreasing or keeping at the same low level the $\varepsilon_{background}$. As a result, a better performance of CJV cut is achieved if the $\eta_{3^{rd}jet}$ lies between the η of the two VBF tagging jets, for every p_T cut, while searching a 3^{rd} with no η restrictions makes the cut less effective. Focusing on the p_T cut of 20 GeV, a 5% of improvement is achieved for ε_{signal} while $\varepsilon_{background}$ stays practically stable. Following the optimized p_T cut of VBF $H \to \tau^+\tau^-$group, which is 20 GeV, a reference to this specific p_T threshold will be done from now on, when comparing all the possible CJV cut methods.

Several $|\eta_{3^{rd}jet}|$ thresholds were tested to compete the default one of $|\eta_{3^{rd}jet}|$ ≤3.2. Figure 4.27 shows the η distribution of central jet for signal and background in lh-channel. Only jets with p_T ≥20 GeV within $|\eta|$ ≤3.2 are taken into consideration since these are the jets, which cause the rejection of the event considering the default CJV cut. As observed, the central jet, which is the third jet in the event, shows a preference towards the central region of the detector and the effect is even more visible for the $t\bar{t}$ background which is highly rejected by the CJV cut. On the other hand, a very similar distribution is observed for signal and $Z \to \tau\tau$+jets background.

The $|\eta_{3^{rd}jet}|$ range, which was found to give the most optimal results was $|\eta_{3^{rd}jet}|$ ≤2.8. A higher value than 3.2 was having a poor performance since the CJV was approximately

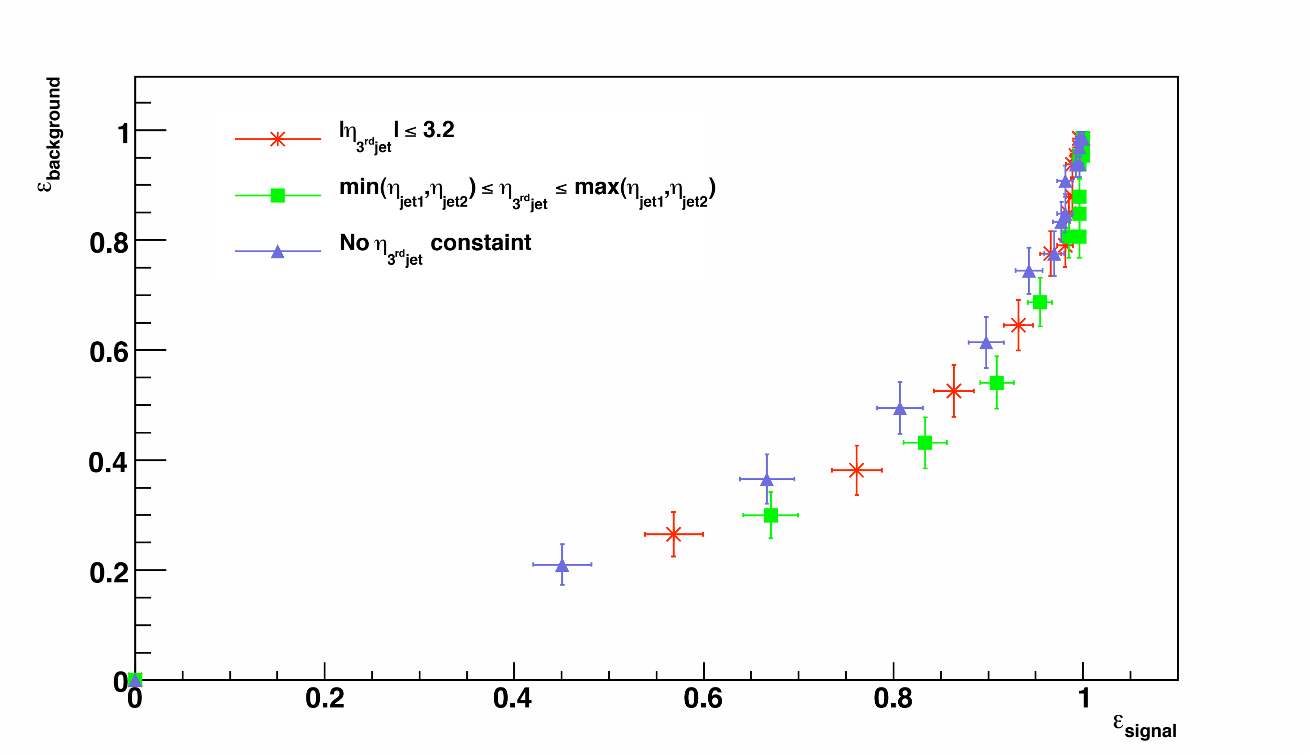


Figure 4.26: *Central Jet Veto cut efficiency for three different cases of* $\eta_{3^{rd}jet}$ *constraints (see text for details).* $H \rightarrow \tau^+\tau^- \rightarrow lh$ *is the signal sample used, while* $Z \rightarrow \tau\tau$*+jets is used for background.*

similar to the case already tested of no constraints in η (figure 4.26, blue triangle markers). Results obtained by the method $|\eta_{3^{rd}jet}| \leq 2.8$ are presented in figure 4.28, superimposed with the results obtained by the default method and the $\min(\eta_{j_1}, \eta_{j_2}) \leq \eta_{3^{rd}jet} \leq \max(\eta_{j_1}, \eta_{j_2})$ method. As observed, even if $|\eta_{3^{rd}jet}| \leq 2.8$ presents a better performance for the $|\eta_{3^{rd}jet}| \leq \eta_{fix}$ case it is slightly less efficient with the respect to the $\min(\eta_{j_1}, \eta_{j_2}) \leq \eta_{3^{rd}jet} \leq \max(\eta_{j_1}, \eta_{j_2})$ case. The difference is more visible in the p_T cut thresholds of 10 GeV and 15 GeV, while for $p_T = 20$ GeV, CJV cut efficiency for signal and background is 0.905 ± 0.018 and 0.526 ± 0.045, while 0.91 ± 0.02 and 0.54 ± 0.05 are the respective efficiencies for $\min(\eta_{j_1}, \eta_{j_2}) \leq \eta_{3^{rd}jet} \leq \max(\eta_{j_1}, \eta_{j_2})$ method. As a conclusion, one could say a clear improvement is seen if modifying the $|\eta_{3^{rd}jet}| \leq 3.2$ to $|\eta_{3^{rd}jet}| \leq 2.8$.

signal

p_T cut (GeV)	$\lvert\eta_{3^{rd}jet}\rvert \leq 3.2$	$\min(\eta_{j_1},\eta_{j_2}) \leq \eta_{3^{rd}jet} \leq \max(\eta_{j_1},\eta_{j_2})$	$\lvert\eta_{3^{rd}jet}\rvert < 5$
10	0.56 ± 0.03	0.67 ± 0.03	0.45 ± 0.03
15	0.76 ± 0.03	0.83 ± 0.02	0.67 ± 0.03
20	0.86 ± 0.02	0.91 ± 0.02	0.81 ± 0.02
25	0.93 ± 0.02	0.95 ± 0.02	0.90 ± 0.02
30	0.97 ± 0.02	0.98 ± 0.01	0.94 ± 0.01
35	0.98 ± 0.01	0.996 ± 0.004	0.97 ± 0.01

background

p_T cut (GeV)	$\lvert\eta_{3^{rd}jet}\rvert \leq 3.2$	$\min(\eta_{j_1},\eta_{j_2}) \leq \eta_{3^{rd}jet} \leq \max(\eta_{j_1},\eta_{j_2})$	$\lvert\eta_{3^{rd}jet}\rvert < 5$
10	0.26 ± 0.03	0.30 ± 0.04	0.21 ± 0.04
15	0.38 ± 0.04	0.43 ± 0.05	0.37 ± 0.05
20	0.53 ± 0.05	0.54 ± 0.05	0.49 ± 0.05
25	0.65 ± 0.05	0.69 ± 0.04	0.61 ± 0.05
30	0.78 ± 0.02	0.81 ± 0.04	0.74 ± 0.04
35	0.79 ± 0.04	0.81 ± 0.04	0.78 ± 0.04

Table 4.13: *Central Jet Veto cut efficiency for* $H \rightarrow \tau^+\tau^- \rightarrow lh$ *signal and* $Z \rightarrow \tau\tau$*+jets samples. Three different cases of* $\eta_{3^{rd}jet}$ *constraints (see text for details) are considered.*

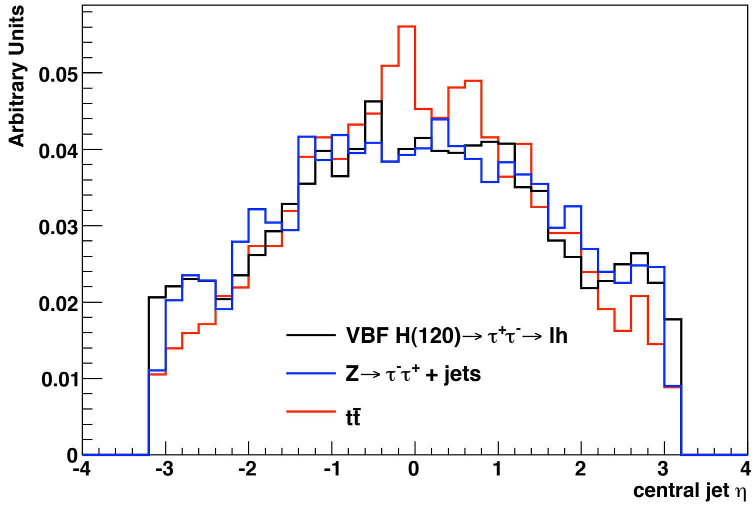


Figure 4.27: η *distribution of central jet which vetos the event due to CJV for signal and background in lh-channel. Jet and lepton cuts are applied (see table 4.12) apart from the hadronic* τ *one. Only jets with* $p_T \geq 20$ *GeV are taken into consideration with* $|\eta| \leq 3.2$*.*

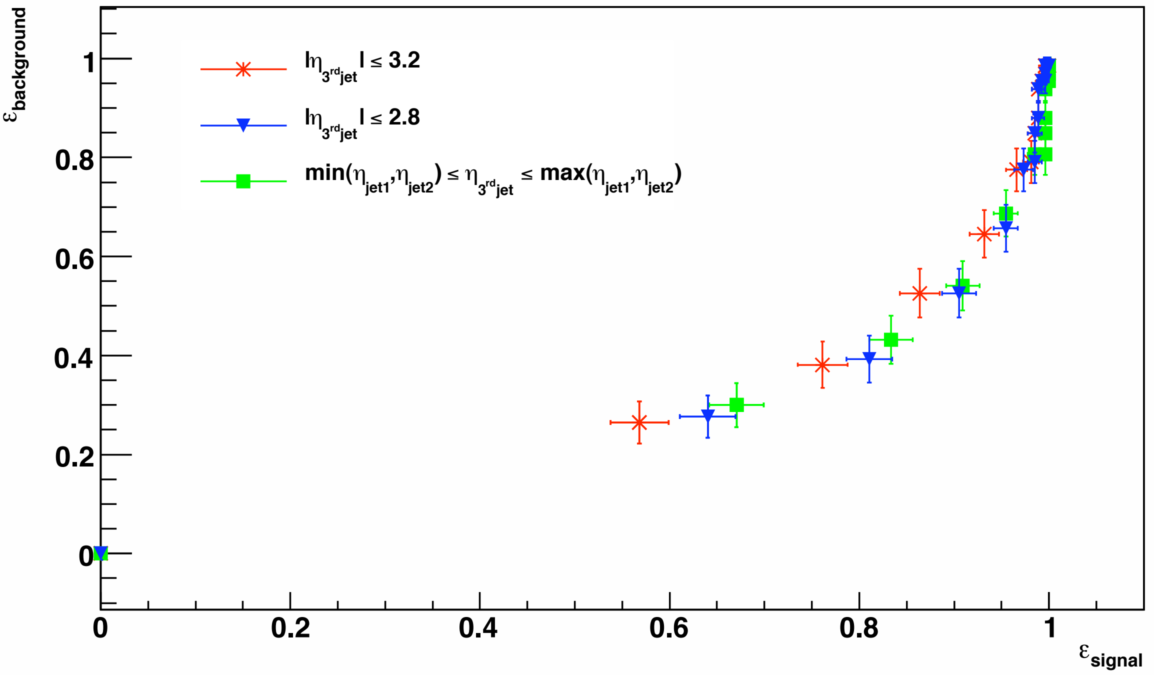


Figure 4.28: *Central Jet Veto cut efficiency for the default* η *scenario (red markers),* $|\eta_{3^{rd}jet}|$ $\leq$*2.8 (blue markers) and* $min(\eta_{j1}, \eta_{j2}) \leq$ $\eta_{3^{rd}jet}$ $\leq max(\eta_{j1}, \eta_{j2})$ *(green markers).* $H \rightarrow \tau^+\tau^- \rightarrow lh$ *is the signal sample used, while* $Z \rightarrow \tau\tau$*+jets is used for background.*

4.10.4 Splitting effect of jets

Observing the CJV cut efficiency plots and the tables 4.13 one could see a ratio of ∼15% of signal events rejected by the CJV cut, while a minor effect was expected. Investigating this feature an (η,ϕ) event display was produced (figure 4.29) illustrating a signal event rejected by the CJV cut. An effect of the presence of the 3^{rd} vetoing jet, very "close" to the VBF tagging jet is observed and performing the similar study for all the rejected signal events, the conclusion of approximatively 60% of signal events were presenting a similar behavior. This was the motivation for investigating the ΔR[1] variable, defining the (η,ϕ) distance, between the VBF tagging jet and the 3^{rd} potential vetoing jet (figure 4.30). Since the cone size of the jet is ΔR=0.4 the distribution starts from this value and shows a peak at the value

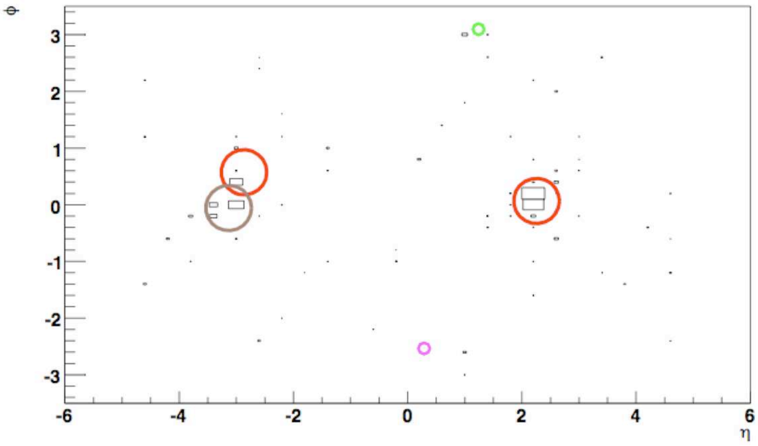


Figure 4.29: *(η,ϕ) event display illustrating a $H \rightarrow \tau^+\tau^- \rightarrow \tau(hadron) + e$ event rejected by the CJV cut. The red circles correspond to the VBF tagging jets, the brown circle to a 3^{rd} jet with a p_T greater than 20 GeV, the green circle to a hadronic τ and the purple circle to an electron. Open boxes correspond to the E_T distribution of the topological clusters*

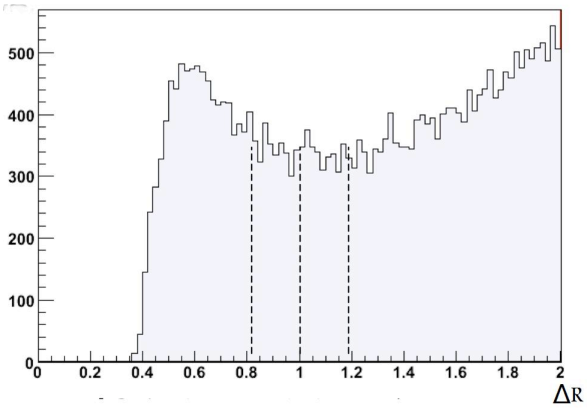


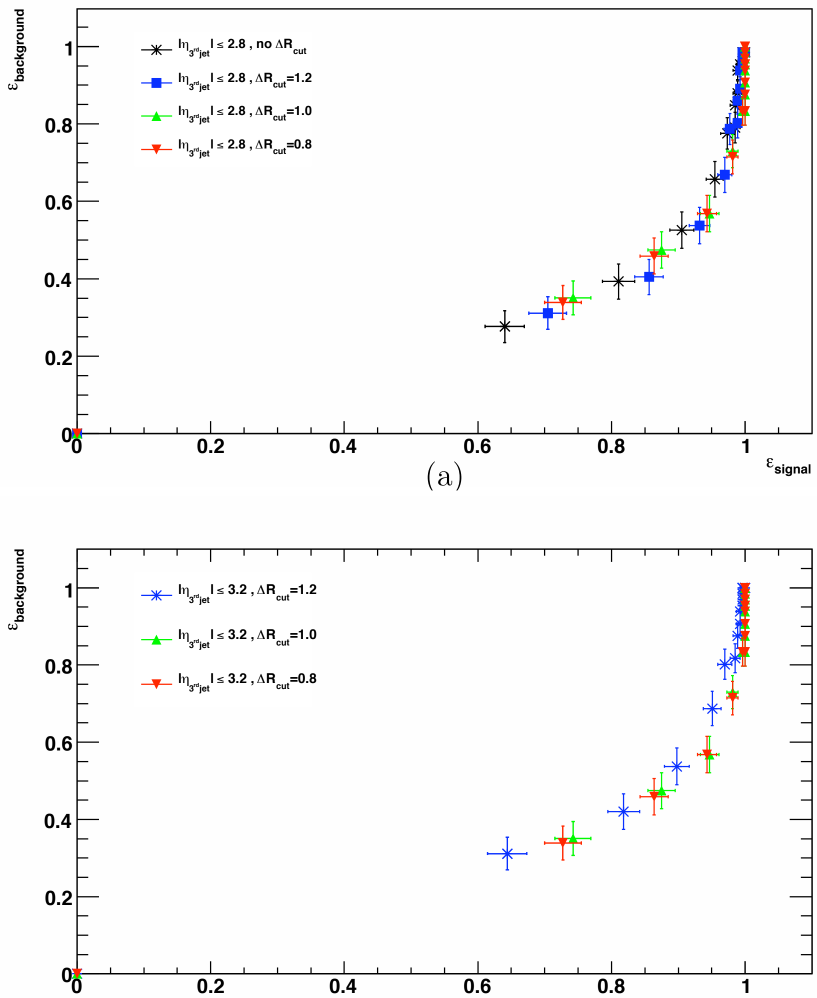
Figure 4.30: *ΔR distribution between the VBF tagging jet and the 3^{rd} potential vetoing jet for a signal event. Vertical lines correspond to potential extra cuts: ΔR=0.8, ΔR=1.0 and ΔR=1.2.*

---

[1] Reminder: $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$

ΔR ~0.55. A reasonable amount of central jets are found thus close to the VBF tagging ones and affect the CJV cut signal efficiency. This is explained due to the fact that the VBF jets can be highly energetic and thus expand outside a jet cone of ΔR=0.4. Therefore, the jet algorithm "splits" the one jet into two, producing a secondary jet which might cause the rejection of a signal event when applying the CJV cut.

A potential solution to the splitting effect of jets could be an extra condition within the CJV cut to exclude these jets. Three possible conditions corresponding to ΔR >0.8, ΔR >1.0 and ΔR >1.2 (see vertical lines of figure 4.30) for $|\eta_{3^{rd}jet}| \leq 2.8$ and $|\eta_{3^{rd}jet}| \leq 3.2$ are examined and results are presented in figure 4.31. A slightly better performance of CJV cut efficiency is achieved with the additional condition of ΔR >1.0 compared to the ΔR >0.8 and ΔR >1.2 and thus it is chosen to be used for further optimisation.



(a)

(b)

Figure 4.31: *CJV cut efficiency for* $|\eta_{3^{rd}jet}| \leq 2.8$ *(a) and* $|\eta_{3^{rd}jet}| \leq 3.2$ *(b). The event rejection is performed only if central jet fulfills the following conditions:* ΔR *>1.2 (blue markers),* ΔR *>1.0 (green markers) and* ΔR *>0.8 (red markers). The default cut efficiency (without the* ΔR *cut) is provided for* $|\eta_{3^{rd}jet}| \leq 2.8$ *(black markers), as a reference.*

Figure 4.32 shows the comparison between the default CJV cut efficiency and the CJV cut efficiency adding the ΔR cut presented below. For the p_T cut of 20 GeV, an increase of 5% in CJV cut ε_{signal} is presented, while $\varepsilon_{background}$ remains stable (at 0.53 ± 0.05, see table 4.13).

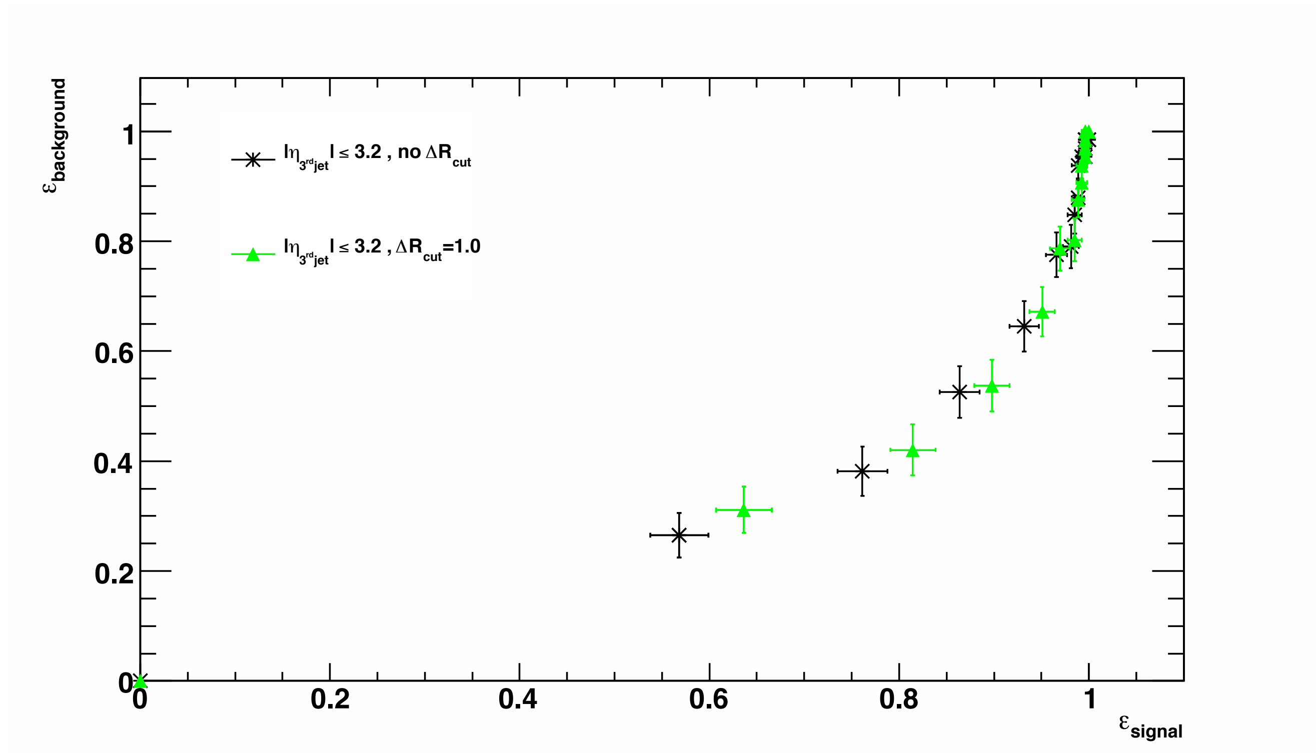
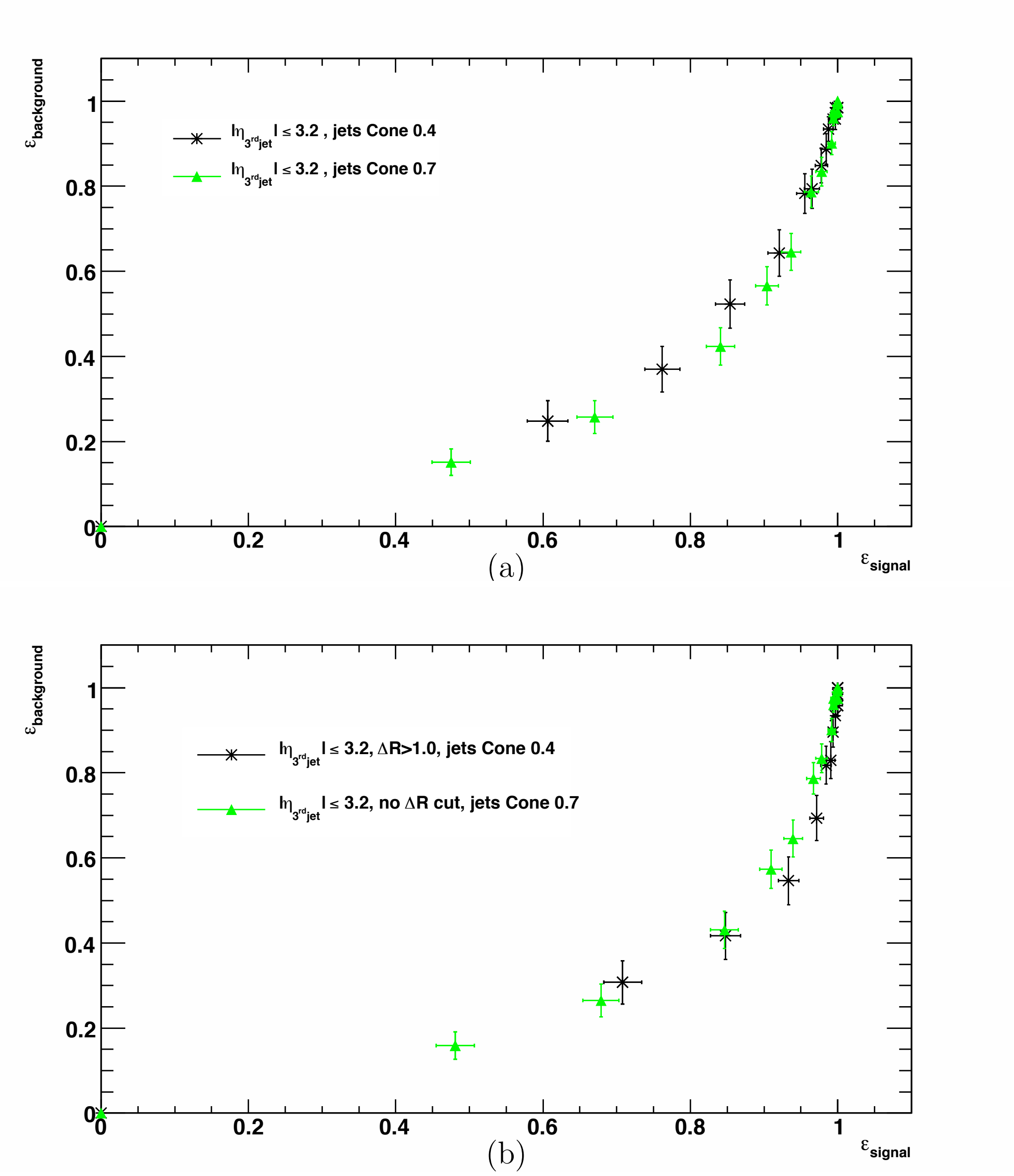


Figure 4.32: *CJV cut efficiency for the default η scenario (black markers) and for the default CJV cut in addition to a ΔR=1.0 condition (green markers).*

A possible solution, apart from incorporating the ΔR >1.0 condition in the CJV cut, would be the use of the other jet cone size used in ATLAS, which is ΔR_{cone}=0.7. In that case, splitting effects illustrated in figure 4.29 would be probably less visible. Figure 4.33(a) shows the comparison of the default CJV cut efficiency for the two different sizes of ΔR_{cone} jets. For the p_T cut of 20 GeV, a very similar ε_{signal} is observed, while a 30% of better background rejection is shown for jets ΔR_{cone}=0.7. However, overall analysis optimisation has shown that ΔR_{cone}=0.4 was better, since changing the jet size also has an impact on VBF tagging variable, like m_{jj}, which more than compensate the change in CJV efficiency. To validate the additional ΔR >1.0 cut for jets with a ΔR_{cone}=0.4, figure 4.33(b) was reproduced illustrating the default CJV cut efficiency for ΔR_{cone}=0.7 jets superimposed with the CJV cut efficiency for ΔR_{cone}=0.4 adding the condition of ΔR >1.0. The additional ΔR >1.0 cut has a slightly better performance than increasing the jet cone size, and in addition it has no impact on the VBF jet taggging efficiency.



(a)

(b)

Figure 4.33: *(a) CJV cut efficiency using cone 0.4 jets (black markers) cone 0.7 jets (green markers). (b) CJV cut efficiency for the default η scenario plus a ΔR cut equal to 0.1, using cone 0.4 jets (black markers) and using cone 0.7 jets without applying ΔR cut (green markers).*

Including both improvements in the baseline CJV cut 4.34 was produced. The cut efficiency is presented for the baseline method (open triangles) the baseline method adding the ΔR cut (green triangles) and the min(η_{j_1}, η_{j_2})$\leq$ $\eta_{3^{rd}jet}$ $\leq$max(η_{j_1}) method adding the ΔR cut (red squares). A 12% of cut efficiency in signal is achieved in total for a similar background rejection. Considering the low cross sections presented in previous section, such an improvement can contribute to a better signal significance.

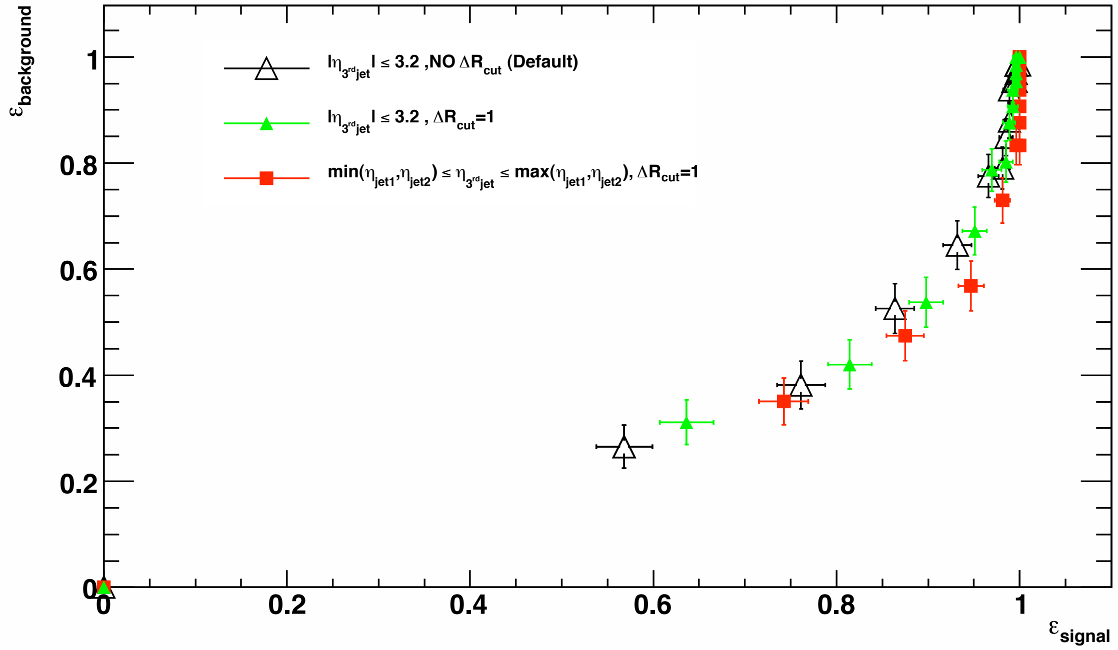


Figure 4.34: *Central Jet Veto cut efficiency for the default η scenario (open markers), $|\eta_{3^{rd}jet}|$ $\leq$3.2 plus the ΔR cut (green markers) and min(η_{j_1}, η_{j_2})$\leq$ $\eta_{3^{rd}jet}$ $\leq$max(η_{j_1}, η_{j_2}) plus the ΔR cut (red markers).*

To conclude, figure 4.35 is shown. It refers to a default CJV cut efficiency graph, which summarizes three different event cases:

- signal and $Z \rightarrow \tau\tau$+jets as background
- signal and $t\bar{t}$ as background
- signal and $Z \rightarrow \tau\tau$+jets as background, considering pile-up contribution for a L=10^{33}cm^{-2}s^{-1}

The same "construction" rules were kept for construction this graph and each point corresponds to a p_T cut starting from the value of 10 GeV with a step of 5 GeV. To help the reader, lines are connecting the points which belong the the same data groups. As a result, 3 main of curves are produced which consist of two lines corresponding to the min(η_{j_1}, η_{j_2})$\leq$ $\eta_{3^{rd}jet}$ $\leq$max(η_{j_1}) method (red line) and the default method (black line). The first group of points (right-handed side) corresponds to the CJV cut efficiency where background is represented by $t\bar{t}$ (triangle markers). As expected, a very good cut efficiency

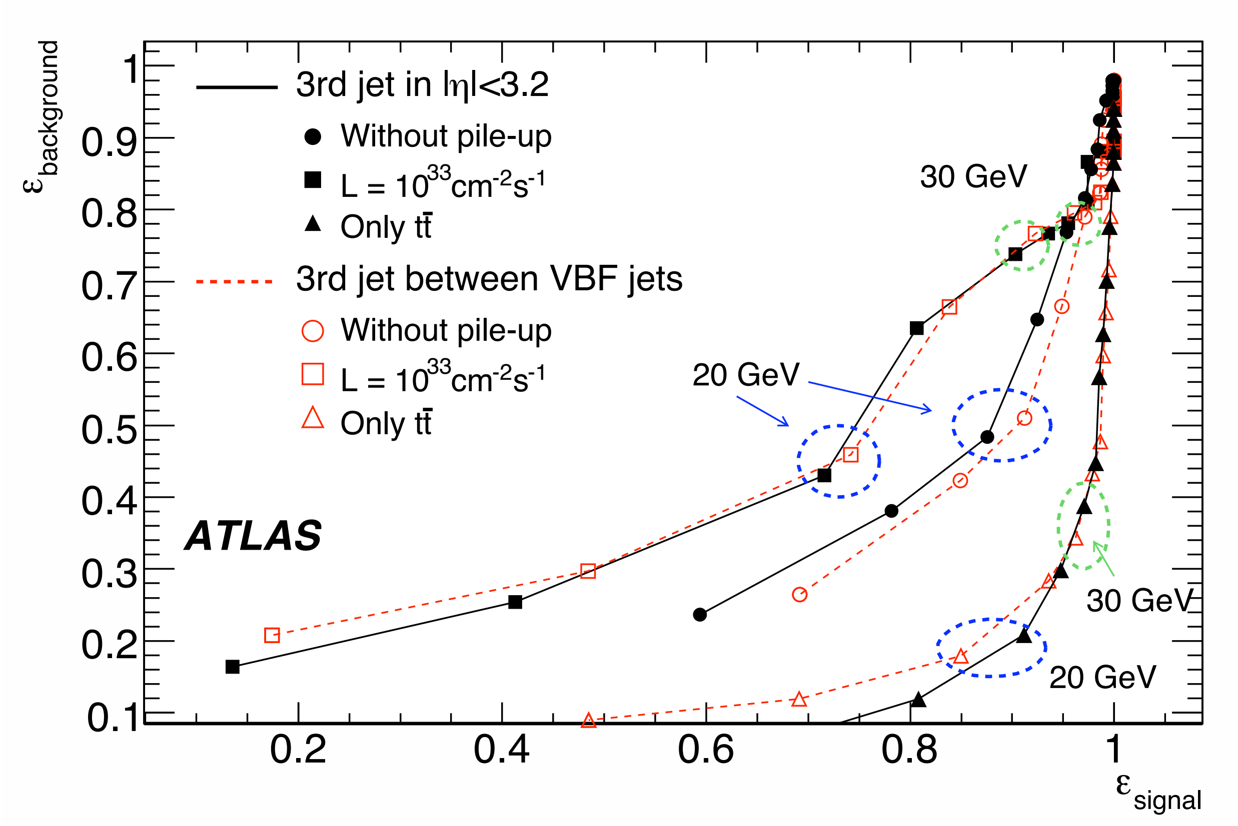


Figure 4.35: *Central Jet Veto cut efficiency for signal and background with and without pileup. Also shown is the case for $t\bar{t}$ only background. Two methods are illustrated: CJV default method (black markers), min*(η_{j_1}, η_{j_2})$\leq \eta_{3^{rd}jet} \leq$*max*($\eta_{j_1}, \eta_{j_2}$) *(red markers).*

is achieved against this background, rejecting a significant part of $t\bar{t}$ events, while ε_{signal} remains at the level of ∼90%.

The curve in the middle corresponds to the CJV cut efficiency considering $Z \rightarrow \tau\tau$+jets as background. It is obvious that compared to the $t\bar{t}$ events CJV becomes less efficient, since $Z \rightarrow \tau\tau$+jets topology is very similar to the VBF $H \rightarrow \tau^+\tau^-$signal topology. Moving to the final curve at the left-handed part of the graph, efficiency becomes even poorer due to the pile-up events. Pile-up (section 4.3) contribution results in a presence of additional low p_T jets. As a result, this jets will cause the drop of ε_{signal} vetoing a significant number of signal events: from ∼0.85, ε_{signal} drops to ∼0.71 for a p_T cut of 20 GeV. A special attention should be thus drawn to the pile-up events when using CJV cut and even in more general jet related cuts. A possible solution to minimize this effect, would be the validation of jets through the tracks.

4.10.5 Cluster veto cut

Cluster veto cut is a similar approach to the CJV cut, described in previous section and it is based on a similar idea. Instead of searching for a reconstructed 3^{rd} jet which might veto the event, an E_T sum of topoclusters is proposed as a replacement of the 3^{rd} p_T cut. The idea is to find the optimal E_T cut which will allow a better cut efficiency. Figure 4.36 presents an (η,ϕ) event display illustrating a $H \rightarrow \tau^+\tau^- \rightarrow \tau(hadron) + e$ event rejected by the CJV cut (figure 4.36 (a)) and a background event of the sample $Z \rightarrow \tau\tau \rightarrow \tau(hadron) + \mu$+2jets, which was passing the CJV cut(figure 4.36 (b)). It is clear that for the background event, there is a higher cluster activity, which gives the motivation of performing a cluster veto study.

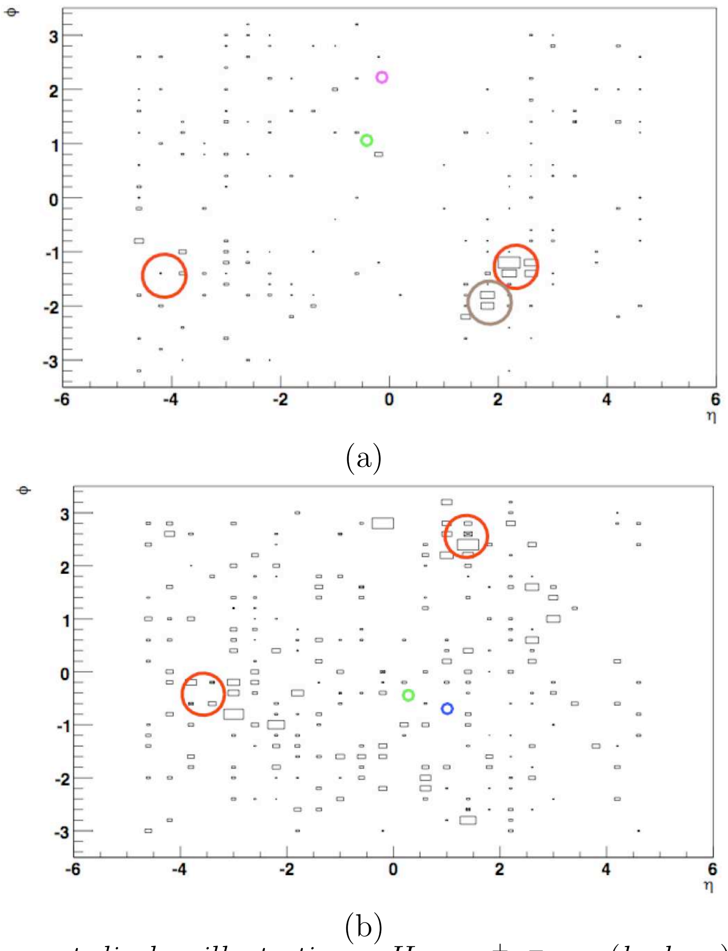


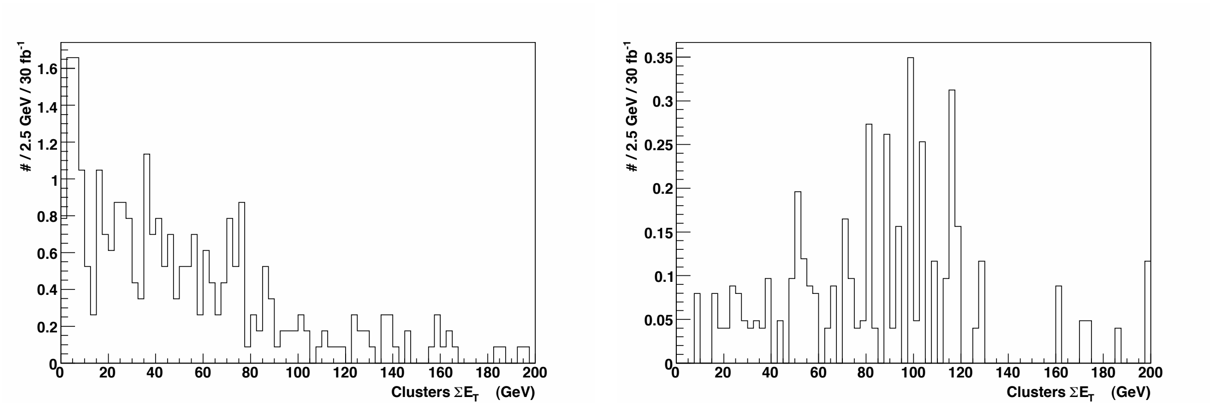
Figure 4.36: *(η,ϕ) event display illustrating a $H \rightarrow \tau^+\tau^- \rightarrow \tau(hadron) + e$ event rejected by the CJV cut (a) and a respective event display of the background sample $Z \rightarrow \tau\tau \rightarrow \tau(hadron) + \mu$+2 jets, which passed the CJV cut (b). The red circles correspond to the VBF tagging jets, the brown circle to a 3^{rd} jet with a p_T greater than 20 GeV, the green circle to a hadronic τ, the purple circle to an electron and the blue to a muon. Open boxes correspond to the E_T distribution of the topological clusters. Only clusters with $E_T \geq 500$ MeV are selected.*

A first step is to remove the clusters associated to the expected final objects of the event.

Considering the size of the reconstructed objects and the splitting effect of VBF tagging jets for a 0.4 jet cone, presented in previous section, the ΔR thresholds presented in table 4.14 were chosen to ensure the full removal of the clusters which correspond either to the Higgs boson decay products, either to the VBF tagging jets.

particle object	ΔR for removal
electron	0.1
muon	0.2
hadronic τ	0.4
VBF tagging jets	1.0

Table 4.14: *ΔR thresholds for removing clusters which belong to reconstructed objects: e, μ, τ and VBF tagging jets.*



(a) (b)

Figure 4.37: *clusters E_T for signal (a) and background (b). Only clusters with $E_T \geq$ 500 MeV are selected.*

The new tested variable consists of the sum of E_T of remaining clusters per event. An additional cut of cluster E_T ≥500 MeV is asked to avoid a possible bias from noise fluctuations. The sum of E_T for signal and background is illustrated in figure 4.37, after applying one the following η requirements:

- $\min(\eta_{j_1}, \eta_{j_2}) \leq \eta_{clusters} \leq \max(\eta_{j_1}, \eta_{j_2})$
- $\eta_{clusters} \leq 3.2$
- $\eta_{clusters} \leq 2.8$
- and finally, sum up all available clusters without any η restriction.

The requirement $\eta_{clusters} \leq 2.8$ has the best performance, which is however worse than the standard central jet veto (figure 4.35). This is most likely due to the impact of the underlying event.

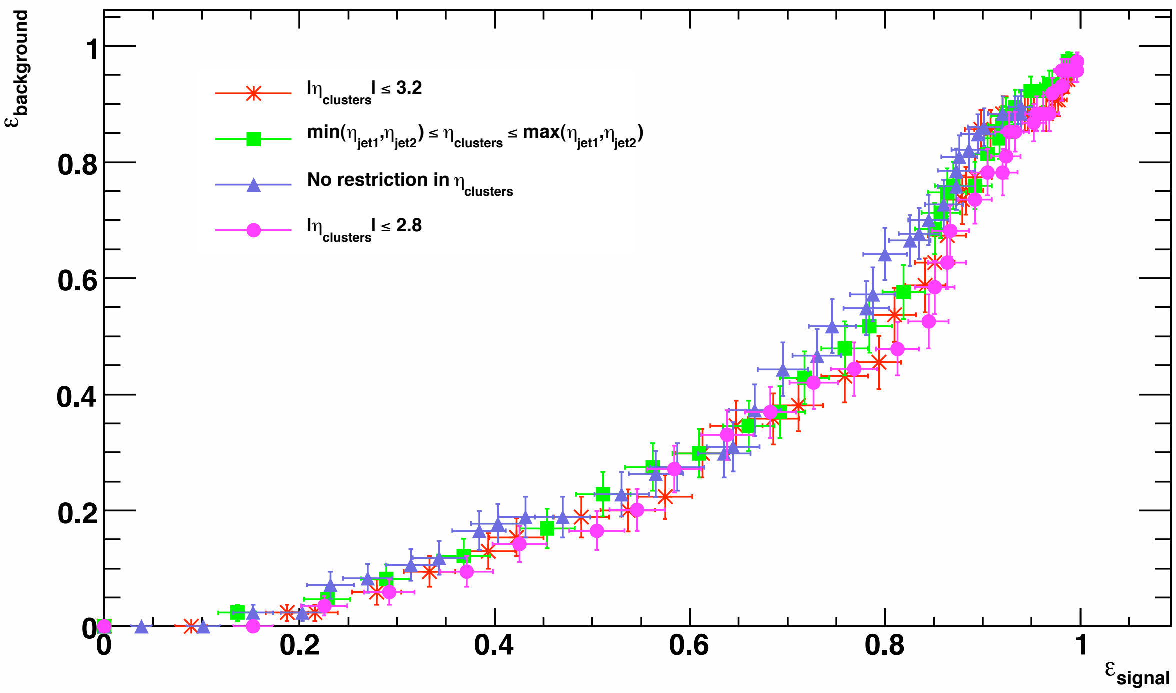


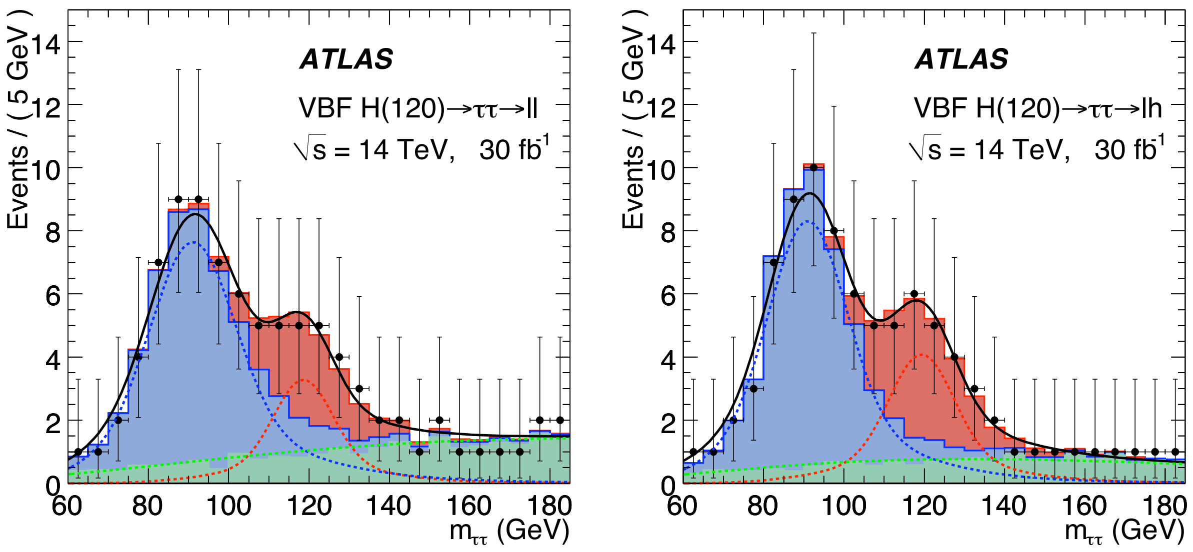
Figure 4.38: *Cluster E_T Veto cut efficiency for signal and background. Four different methods are represented (see text for details).*

4.11 Mass determination and signal significance

4.11.1 Higgs boson mass

For a Higgs boson mass of m_H=120 GeV figure 4.39 was reproduced showing the expected number of events for a luminosity of 30 fb^{-1} after applying all the cuts presented in section 4.6 for ll-channel (figure 4.39(a)) and lh-channel (figure 4.39(b)). The different colors correspond to the different event data sets. A clear signal excess is observed for a luminosity of 30 fb^{-1} but is is to be noted that pile-up contribution has not been taken into consideration. Due to the Z resonance at ∼90 GeV, the m_H resolution is an essential ingredient of the analysis. Its correct calibration on data will be a key part of the future analysis, the commissioning study on random trigger event shown in chapter 2 being a first valuable step.

For the different mass bins presented in section 4.3, m_H the resolution was extracted and presented in figure 4.40 for both ll- and lh-channel. For the default mass value of m_H=120 GeV, a resolution of 11.5 GeV is achieved for the ll-channel and 13.2 GeV is the respective one for the lh-channel. In general, ll-channel shows a slightly worse performance with the respect to the lh because of the higher amount of $\not{E}_T$ expected in the full leptonic decay mode. The same argument can justify the fact that moving to higher masses of m_H the resolution becomes worse.



(a) (b)

Figure 4.39: *$m_{\tau\tau}$ distribution after applying all the described cuts for ll-channel (a) and lh-channel (b). Red color represents the signal, blue stands for Z background and green for W, $t\bar{t}$ and QCD background.These results do not include the impact of pileup [19].*

4.11.2 Signal significance

In order to investigate the existence of a new particle such as the Higgs boson, a statistical hypothesis test is performed. Let B be the notation for the background-only hypothesis with no Higgs boson and S the hypothesis which assumes the presence of the Higgs boson as well. Discovery of Higgs boson means that the B hypothesis has to be rejected at a high level of confidence. Typically, a 5σ significance is required to claim discovery.

The probability p for rejecting the background-only hypothesis B, it is the probability of observing $n \geq N$ events in the case that only background is present:

$$p = \int_N^\infty \rho_B(n)dn \tag{4.15}$$

where where N is the number of selected events in the experiment and $\rho_B(n)$ is the probability density function (*pdf*) for the background-only hypothesis given by the Poisson distribution:

$$\rho_B(n) = \frac{e^{-I_b} I_b^n}{n!} \tag{4.16}$$

with I_b being the expected number of background events. The signal significance Z is defined as the number of standard deviations at which a Gaussian distribution centered to zero (μ=0) and standard deviation $\sigma = 1$ gives an one-sided tail area equal to the p probability. Therefore, the significance Z is presented in the integral limit of the probability p:

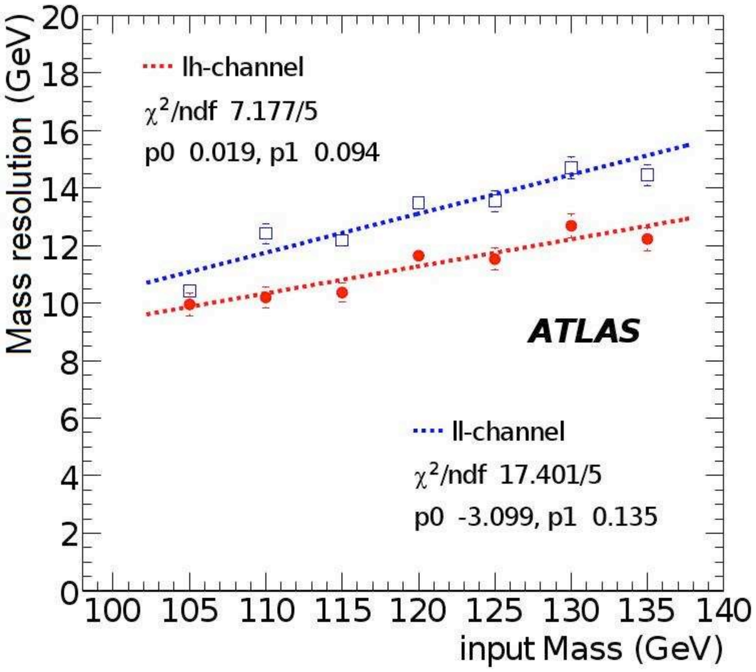


Figure 4.40: *Mass resolution as a function of m_H. These results do not include the impact of pileup [19].*

$$p = \int_Z^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 1 - A(Z) \tag{4.17}$$

where $A(Z)$ is the cumulative distribution of the normalized Gaussian function. A signal significance of Z = 5 corresponds to a p-value of $2.87 \cdot 10^{-7}$. In the limit of relatively large background event numbers such as the case of the present study, the PDF $\rho_B(n)$ can be approximated by a Gaussian *pdf* with mean $\mu = I_b$ and a standard deviation of $\sigma = \sqrt{I_b}$. So the new $\rho'_B(n)$ can be written:

$$\rho'_B(n) = \frac{1}{\sqrt{2\pi I_b}} e^{-\frac{(n-I_b)^2}{2I_b}} \tag{4.18}$$

resulting to the relation:

$$Z = \frac{I_s}{\sqrt{I_b}} \quad , \tag{4.19}$$

where I_s is the number of signal events.

This is the formula used to calculate the signal significance in present study, counting the remaining events for signal and background after the application of all cuts.

This calculation was done for all the different possible m_H bins available and the results are presented in table 4.15 and figure 4.41.

Last column of table 4.15 contains the expected signal significance after the combination of *ll*- and *lh*-channel, which is the the quadratic sum of the two decay channels signal

m_H	ll-channel	lh-channel	combined
105	2.20	2.85	3.80
110	2.46	3.45	4.46
115	2.86	4.18	5.32
120	2.80	4.23	5.36
125	2.67	3.97	5.08
130	2.42	3.54	4.62
135	2.17	3.38	4.35
140	2.14	2.66	3.55

Table 4.15: *Expected signal significance for ll- , lh- and combined channel. The event counting method was used.*

significance. An excess above 5σ is found for the mass range m_H ~(115,125) GeV, which implies a discovery of a Higgs boson for $L = 30\ fb^{-1}$. The significance decreases at lower mass due to the larger contribution of the Z background, and at larger mass due to the lower signal cross section. However one should note that the analysis has not been re-optimised in other mass points, hence some improvement would still be possible.

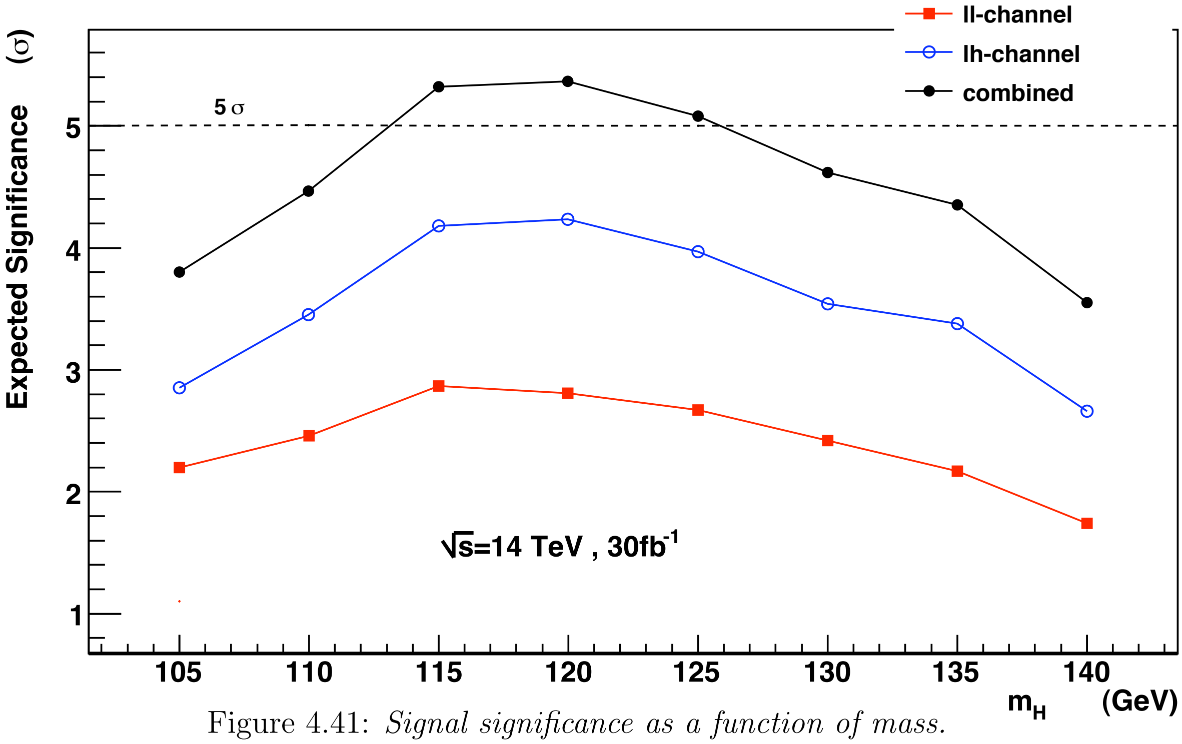


Figure 4.41: *Signal significance as a function of mass.*

Chapter 5

Conclusion

The ATLAS experiment at the Large Hadron Collider is designed to explore a wide range of new phenomena in particle physics with proton-proton collisions at energies up to 14 TeV. One of the ATLAS primary objectives is the discovery of the Higgs boson. Experimental and theoretical bounds constrain the Standard Model (SM) Higgs boson mass to the range 114.4 GeV$\leq m_H \leq$185 GeV.

One of the most promising discovery channels, especially in the lower mass range, is the vector-boson fusion (VBF) production of the Higgs boson with the subsequent decay into a τ-lepton pair ($H \rightarrow \tau^+\tau^-$). The decay modes where both τ leptons decay leptonically (e or μ) and where one τ lepton decays leptonically and the other one hadronically have been studied. The VBF production mechanism provides a set of special characteristics which allow the rejection of large background expected in LHC, such as $t\bar{t}$ production and $Z \rightarrow \tau\tau$, processes, which are characterized by a cross section ten orders of magnitude greater than the one of the VBF Higgs boson. As a result, very detailed and accurate studies should be performed in order to understand and reject the background.

The VBF topology consists of two highly energetic jets in the forward regions of the detector providing a characteristic signature for Higgs detection in this production process, with two jets in high $|\eta|$ values. A cut-based analysis was used focusing on the central jet veto (CJV) cut optimization: Since VBF is a purely electroweak process, little QCD activity is expected and thus no central jets are expected for signal events. This allows the CJV cut application to further suppress background processes.

Combining the results of the ll-channel and lh-channel, an excess above 5σ of signal significance is found for the mass range $m_H\sim$ (115,125) GeV, which implies a discovery of a Higgs boson for L = 30 fb^{-1}. A maximum signal significance of 5.4σ is achieved for a Higgs mass of 120 GeV. Nevertheless, the effect of pile-up is not taken into consideration.

Figure 5.1 shows the SM Higgs boson signal significance for the different decay channels studied in ATLAS and their combination, for a luminosity of L=10 fb^{-1}, which corresponds to the first 2 years of LHC operation. A 5σ discovery is achieved for a m_H above $m_H \geq\sim$127 GeV with the biggest contribution originating from the $H \rightarrow \gamma\gamma$ decay channel. The studied decay channel $H \rightarrow \tau^+\tau^-$, even though it contributes slightly in the whole mass range $m_H\sim$(110,200) GeV, it provides the most important contribution at the low mass range $m_H\sim$(110,125) GeV.

Figure 5.2 shows the signal significance as a function of luminosity (L) and as a function

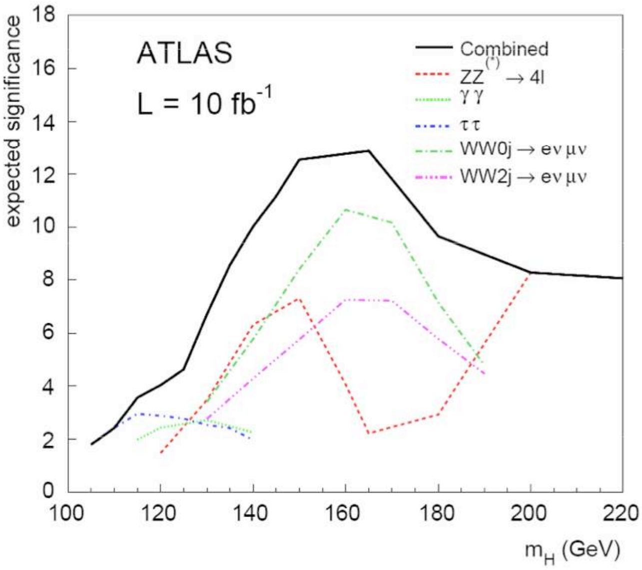


Figure 5.1: *SM Higgs boson signal significance for different decay channels and combined, for a luminosity of* L *= 10 fb*$^{-1}$ *[75].*

of m_H, with 0 fb^{-1} $<L<$10 fb^{-1} corresponding to the first year of a full LHC operation. One can see that a 5σ discovery is expected at the mass region of 140 GeV$<m_H<$180 GeV

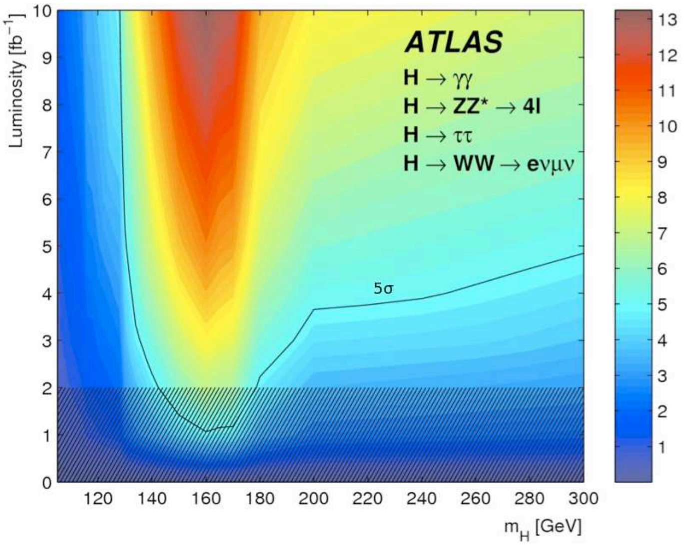


Figure 5.2: *SM Higgs boson signal significance (in color) as a function of luminosity and the* m_H*, combining different decay channels [76].*

with a luminosity of L=2 fb^{-1}, if Higgs boson exists. Concerning the $H \to \tau^+\tau^-$similarly to the conclusion of figure 5.1, its contribution becomes significant at the low mass range but a luminosity greater than $L > 10$ fb^{-1} is needed to reach the 5σ threshold postponing a possible discovery in this mass region, to the late years of LHC operation.

Finally, an important issue of the present VBF $H \rightarrow \tau^+\tau^-$ analysis is the presence of missing transverse energy ($\not{E}_T$) due to the neutrinos originating from the τ decays. $\not{E}_T$ resolution is of high importance since it affects directly the resolution of the Higgs boson mass in the $H \rightarrow \tau^+\tau^-$decay (a final Higgs boson mass resolution of ~10 GeV is achieved). The reason is that $\not{E}_T$ participates in the the reconstruction of the $m_{\tau\tau}$, using the approximation that all τ decay products (visible and non-visible) are collinear to the direction of the primary τ. This was the initial boost for performing a commissioning analysis of ATLAS calorimeters studying $\not{E}_T$ variables.

The most dominant contribution of $\not{E}_T$ originates from calorimeter and thus an understanding and a control of calorimeter cells behavior will result to a more realistic $\not{E}_T$ reconstruction. The first step towards an understanding of $\not{E}_T$ reconstruction was done within the scope of the present analysis by measuring $\not{E}_T$ contribution of electronics noise in random triggered data obtained during the single-beam/cosmic data-taking period of fall 2008.

A Gaussian behavior of noise in all LAr calorimeters was found, whereas a region of non-Gaussian tails in Tile calorimeters was high-lighted. Several aspects which may affect the final $\not{E}_T$ reconstruction were addressed and understood, such as the influence of calibration parameters in energy reconstruction, the influence of bad cells ("noisy") and the influence of non-Gaussian tails of Tile calorimeters, especially when measuring $\not{E}_T$ using topoclusters. For the latter, a new noise model was tested using a double Gaussian parameterization and it resulted to more realistic results closer to the real data. This double Gaussian parameterization is now implemented in the official ATLAS software releases.

To conclude, this analysis resulted to a better understanding and to a more realistic modeling of noise in ATLAS calorimeters, providing at the same time a useful calorimeter commissioning tool, which allowed the observation and the correction of several unexpected features, preparing the experiment for the real data. Part of the results obtained by this study, are already officially approved by the ATLAS collaboration [10].

Bibliography

[1] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. 13 (1964) 508-509.

[2] Y. Nambu and G. Jona-Lasinio, Dynamical model of elementary particles based on an analogy with superconductivity. I, Phys. Rev. 122 (1961) 345-358.

[3] J. Goldstone, A. Salam, and S. Weinberg, *Broken Symmetries*, Phys. Rev. 127 (1962) 965-970.

[4] LEP Working Group for Higgs Boson Searches, R. Barate et al., *Search for the standard model Higgs boson at LEP*, Phys. Lett. B565 (2003) 61-75, (arXiv:hep-ex/0306033).

[5] Tevatron New-Phenomena and Higgs working group, G. Bernardi et al.,*Combined CDF and D0 Upper Limits on Standard Model Higgs Boson Production at High Mass (155-200GeV/c^2) with 3 fb^{-1} of data*, FERMILAB-PUB-08-270-E, August, 2008. (arXiv:0808.0534 [hep-ex]).

[6] ALEPH collaboration et al., *Precision Electroweak Measurements and Constraints on the Standard Model*, CERN-PH-EP-2008-020, CERN, Geneva, December, 2008. http://cdsweb.cern.ch/record/1142930/.

[7] A. Djouadi,"The Higgs bosons in the minimal supersymmetric model," Phys. Rept. **459** (2008) 1 [arXiv:hep-ph/0503173].

[8] G. Bhattacharyya, A. Datta, S. K. Majee and A. Raychaudhuri, "Exploring the Universal Extra Dimension at the LHC," arXiv:0904.0937 [hep-ph].

[9] P. Meade and L. Randall, "Black Holes and Quantum Gravity at the LHC," JHEP **0805** (2008) 003 [arXiv:0708.3017 [hep-ph]].

[10] https://twiki.cern.ch/twiki/bin/view/Atlas/ApprovedNoisePlotsJetEtMiss

[11] The LHC Study Group CERN/AC/DI/FA/90-06 (October 1990).

[12] E. Nappi [ALICE Collaboration], "Design and performance of the ALICE experiment at LHC," *Prepared for 5th Italy-Japan Symposium on Recent Achievements and Perspectives in Nuclear Physics, Naples, Italy, 3-7 Nov 2004*

[13] [LHCb Collaboration], "LHCb TDR computing technical design report," CERN-LHCC-2005-019

[14] G. L. Bayatian *et al.* [CMS Collaboration], "CMS physics: Technical design report," CMS-TDR-008-1

[15] "ATLAS: Detector and physics performance technical design report. Volume 1 - 2,"
ATLAS-TDR-14; (Volume 1)
ATLAS-TDR-15; (Volume 2)

[16] G. Aad *et al.* [ATLAS Collaboration], "The ATLAS Experiment at the CERN Large Hadron Collider," JINST **3** (2008) S08003.

[17] A. Djouadi,"The Anatomy of electro-weak symmetry breaking." Phys. Rept. **459** (2008) 1 [arXiv:hep-ph/0503173].

[18] ATLAS collaboration, "ATLAS Inner detector: technical design report. 1", CERN-LHCC-97-016, http://cdsweb.cern.ch/record/331063;
"ATLAS Inner detector: technical design report. 2", CERN-LHCC-97-017, http://cdsweb.cern.ch/record/331064.

[19] G. Aad *et al.* [The ATLAS Collaboration], "Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics," arXiv:0901.0512.

[20] Experimental Techniques in High Energy Physics, Frontiers in Physics, ed. T. Ferbel (1987)

[21] B. Aubert et al., Development and construction of large size signal electrodes for the ATLAS electromagnetic calorimeter, Nucl. Instrum. Meth. A 539 (2005) 558.

[22] ATLAS collaboration, Liquid argon calorimeter technical design report, CERN-LHCC-96-041, http://cdsweb.cern.ch/record/331061.

[23] D.M. Gingrich et al., Construction, assembly and testing of the ATLAS hadronic end-cap calorimeter, 2007 JINST 2 P05005.

[24] W.E. Cleland, E.G. Stern, Signal processing considerations for liquid ionisation calorimeters in a high rate environment, Nuclear Instruments and Methods in Physics Research A 338 (1994) 467-497

[25] https://twiki.cern.ch/twiki/bin/view/Atlas/ApprovedPlotsLAr

[26] https://twiki.cern.ch/twiki/bin/view/Atlas/ApprovedPlotsTile

[27] ATLAS PhD thesis, V. Tisserand, "Optimisation du détecteur ATLAS pour la recherche du boson de Higgs se désintegrant en deux photons au LHC", CERN-THESIS-99-032 LAL-97-01.

[28] ATLAS Collaboration. ATHENA, The ATLAS Common Framework - Developer Guide.

[29] G. Barrand et al. GAUDI - a software architecture and framework for building LHCb data processing applications. In CHEP 2000, 2000.

[30] Cranmer,K; Farbin,A ; Shibata,A, EventView - The Design Behind an Analysis Framework, ATL-SOFT-PUB-2007-008

[31] D. Quarrie, A. Shibata, et al. Taming the beast: Using python to control ATLAS software. In Europython 2006, 2006.

[32] W. Lampl et al., Calorimeter Clustering Algorithms: Description and Performance, ATLASLARG- PUB-2008-002 (2008).

[33] S. Hassini, et al., NIM A572 (2007) 77-79.

[34] Th. Lagouri, et al., IEEE Trans. Nucl. Sci. 51 (2004) 3030-3033.

[35] T. Cornelissen, et al., ATL-SOFT-PUB-2007-007 (2007).

[36] A. Ventura, et al., "The Muon High Level Trigger of the ATLAS experiment", ATL-DAQ-PROC-2009-010 , ATL-COM-DAQ-2009-031.

[37] http://press.web.cern.ch/press/PressReleases/
Releases2008/attachments/CERN081205bLHCrestart.pdf

[38] Hubaut F., Petit E. , Pralavorio P., Rousseau D., Varouchas D.,Olariu A., *Testing calorimetric Missing Transverse Energy and jet reconstruction with random and cosmic data*, ATLAS note ATL-COM-PHYS-2009-227.

[39] F. Spano, *Simulation of the full noise pattern in Tile Calorimeter Front End electronics: a phenomenological approach to coherent effects*, ATLAS note ATL-TILECAL-PUB-2008-011.

[40] B. Salvachua et al, *Online energy and phase reconstruction during commissioning phase of the ATLAS Tile Calorimeter*, ATLAS note ATL-COM_TILECAL-2008_004.

[41] https://twiki.cern.ch/twiki/bin/view/Atlas/LArP3CommissioningUnhappyChannels

[42] https://twiki.cern.ch/twiki/bin/view/Atlas/TileChannelStatusFlags

[43] [ATLAS Collaboration], "ATLAS liquid argon calorimeter: Technical design report," CERN-LHCC-96-41;

[44] [ATLAS Collaboration], "ATLAS tile calorimeter: Technical design report," CERN-LHCC-96-42;

[45] S. Eidelman *et al.* [Particle Data Group], "Review of particle physics," Phys. Lett. B **592** (2004) 1.

[46] S.L.Glashow, Nucl.Phys 22 (1961) 579. S.Weinberg, Phys. Rev. Lett. 19 (1967) 1264. A.Salam, Proceedings of the 8th Nobel Symposium, N.Svartholm editor, 367, NY 1968, Wiley Intescience.

[47] A. Pich, The Standard Model of Electroweak Interactions, (arXiv:0705.4264).

[48] UA1 Collaboration, G. Arnison et al., Experimental observation of lepton pairs of invariant mass around 95 GeV/c^2 at the CERN SPS collider, Phys. Lett. B126 (1983) 398-410.

[49] UA2 Collaboration, M. Banner et al., Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN pp collider, Phys. Lett. B122 (1983) 476-485.

[50] UA2 Collaboration, P. Bagnaia et al., Evidence for $Z^0 \rightarrow e^+e^-$ at the CERN $\bar{p}p$ collider, Phys. Lett. B129 (1983) 130-140.

[51] K. Riesselmann, Limitations of a standard model Higgs boson, DESY-97-222 (1997), (arXiv:hep-ph/9711456).

[52] LEP Working Group for Higgs Boson Searches, R. Barate et al., Search for the standard model Higgs boson at LEP, Phys. Lett. B565 (2003) 61-75, (arXiv:hep-ex/0306033).

[53] ALEPH collaboration et al., Precision Electroweak Measurements and Constraints on the Standard Model, CERN-PH-EP-2008-020, CERN, Geneva, December, 2008. http://cdsweb.cern.ch/record/1142930/.

[54] CDF and D0 Collaborations, *Combined CDF and D0 Upper Limits on Standard Model Higgs-Boson Production with up to 4.2 fb^{-1} of Data*,FERMILAB-PUB-09-060-E, CDF Note 9713, D0 Note 5889.

[55] For a compilation of the most accurate theoretical results for SM and MSSM Higgs cross sections at the Tevatron and the LHC see the Webpage: *http://maltoni.home.cern.ch/maltoni/TeV4LHC/* .

[56] M. S. Carena and H. E. Haber, Higgs boson theory and phenomenology, Prog. Part. Nucl. Phys. 50 (2003) 63-152, (arXiv:hep-ph/0208209).

[57] M. Duehrssen et al., “Extracting Higgs boson couplings from LHC data”, Phys. Rev. D 70 (2004) 113009.

[58] T. Plehn, D. L. Rainwater and D. Zeppenfeld, Determining the structure of Higgs couplings at the LHC, Phys. Rev. Lett. 88 (2002) 051801.

[59] C. Ruwiedel, M. Schumacher, and N. Wermes, Prospects for the Measurement of the Structure of the Coupling of a Higgs Boson to Weak Gauge Bosons in Weak Boson Fusion with the ATLAS Detector, Eur. Phys. J. C 51 (2007) 385.

[60] T. Plehn, D. L. Rainwater and D. Zeppenfeld, “A method for identifying $H \rightarrow \tau^+\tau^- \rightarrow e^\pm\mu^\pm$ missing p_T at the CERN LHC”, Phys. Rev. D 61 (2000) 093005.

[61] M. Schumacher, “Investigation of the discovery potential for Higgs bosons of the minimal supersymmetric extension of the standard model (MSSM) with ATLAS”, hep-ph/0410112.

[62] J. M. Campbell, J. W. Huston and W. J. Stirling, “Hard Interactions of Quarks and Gluons: A Primer for LHC Physics,” Rept. Prog. Phys. **70** (2007) 89 [arXiv:hep-ph/0611148].

[63] R. Bonciani et al., Nucl., Phys. B529, (1998) 424.

[64] ATLAS Computing Group, ATLAS Computing: Technical Design Report, CERN-LHCC-2005-022, CERN, Geneva, 2005. http://cdsweb.cern.ch/record/837738.

[65] GEANT4 Collaboration, S. Agostinelli et al., GEANT4: A simulation toolkit, Nucl. Instrum. Meth. A506 (2003) 250-303.

[66] *Atlfast homepage* http://www.hep.ucl.ac.uk/atlas/atlfast/.

[67] T. Sjostrand, S. Mrenna, and P. Skands, PYTHIA 6.4 physics and manual, JHEP 05 (2006) 026, (arXiv:hep-ph/0603175).

[68] [G. Corcella et al., HERWIG 6: An event generator for hadron emission reactions with interfering gluons (including supersymmetric processes), JHEP 01 (2001) 010, (arXiv:hep-ph/0011363).

[69] J. M. Butterworth, J. R. Forshaw, and M. H. Seymour, Multiparton interactions in photoproduction at HERA, Z. Phys. C72 (1996) 637-646, (arXiv:hep-ph/9601371).

[70] M. L. Mangano et al., ALPGEN, a generator for hard multiparton processes in hadronic collisions, JHEP 07 (2003) 001, (arXiv:hep-ph/0206293).

[71] T. Gleisberg et al., SHERPA 1.alpha, a proof-of-concept version, JHEP 02 (2004) 056, (arXiv:hep-ph/0311263).

[72] F. Krauss, R. Kuhn, and G. Soff, AMEGIC++ 1.0: A matrix element generator in C++, JHEP 02 (2002) 044, (arXiv:hep-ph/0109036).

[73] S. Frixione and B. R. Webber, Matching NLO QCD computations and parton shower simulations, JHEP 06 (2002) 029, (arXiv:hep-ph/0204244).

[74] S. Jadach et al., The tau decay library TAUOLA: Version 2.4, Comput. Phys. Commun. 76 (1993) 361-380.

[75] E. Richter-Was [Atlas Collaboration], “Prospect for the Higgs searches with the ATLAS detector”, arXiv:0903.4198 [hep-ex].

[76] G. Cowan [Atlas Collaboration], “Discovery Potential for the Standard Model Higgs at ATLAS” Les Rencontres de Physique de la Vallee d’Aoste, La Thuile, March 2009, ATL-PHYS-SLIDE-2009-040

Abstract

The subject of this thesis is the evaluation of the discovery potential of the ATLAS detector at the Large Hadron Collider for the Standard Model Higgs boson in vector boson fusion (VBF) production and a subsequent decay into a τ-lepton pair ($H \rightarrow \tau^+\tau^-$). This is one of the most promising discovery channels in the low mass range, which is the mass range favored from precision measurements of the electroweak interaction. The decay modes where both τ leptons decay leptonically and where one τ lepton decays leptonically and the other one hadronically were studied.

The characteristic vector boson fusion topology, which consists of two jets in the forward regions of the detector and the Higgs boson decay products in the central region, provides a unique signature allowing the suppression of background. In addition, since vector boson fusion is a purely electroweak process, no QCD activity is expected and thus no central jets are expected for signal events. This allows the central jet veto cut application to further reject background processes. A cut-oriented analysis was used, focusing on the central jet veto cut optimization.

The main objective was to investigate the Higgs boson ATLAS discovery potential with an integrated luminosity of 30 fb^{-1} in the mass range 115 GeV$\leq m_H \leq$140 GeV, for an LHC energy at the center of mass $\sqrt{s}$=14 TeV. After the application of all cuts, an excess above 5σ of signal significance was found for the mass range 115 GeV$\leq m_H \leq$125 GeV.

In $H \rightarrow \tau^+\tau^-$decay channel, transverse missing energy ($\not{E}_T$) resolution is of high importance since it affects directly the resolution of the Higgs boson mass. This was the initial motivation for performing and presenting in this thesis a calorimetric commissioning study based on $\not{E}_T$ quantities, focused on the electronics noise $\not{E}_T$ contribution, measured with cosmics data of the period fall 2008. A Gaussian behavior of noise in all liquid argon calorimeters was found, whereas a region of non-Gaussian tails in scintillating tile calorimeter was high-lighted. For the latter, a new noise model was tested using a double Gaussian parameterization resulting in a more realistic description. Finally, this analysis provided a useful calorimeter commissioning tool, which allowed the observation and the correction of several features in the ATLAS calorimeter behavior.

Key words: LHC, ATLAS, VBF, Higgs Boson, $H \rightarrow \tau^+\tau^-$, jets, calorimeter, commissioning, $\not{E}_T$, noise.

Résumé

Le but de cette thèse est d'évaluer le potentiel de découverte par le détecteur ATLAS au Large Hadron Collider (LHC) du boson de Higgs du modèle standard. Le mode de production étudié est défini par la fusion de bosons vecteurs (VBF) suivie de la désintégration en une paire de leptons τ ($H \rightarrow \tau^+\tau^-$). C'est l'un des canaux de découverte les plus prometteurs pour un boson de Higgs de basse masse, qui est favorisé par les mesures de précision dans le secteur électro-faible. Deux modes de désintégration ont été étudiés: celui où les deux leptons τ se désintègrent en lepton et celui où un lepton τ se désintègre en lepton et l'autre en hadron.

La topologie caractéristique de fusion de bosons vecteurs, qui est constituée des deux jets vers l'avant du détecteur et des produits de désintégration de boson de Higgs dans la région centrale, fournit une signature unique permettant la suppression du bruit de fond. En outre, puisque la fusion des bosons vecteurs est un processus purement électro-faible, aucune activité de QCD n'est prévue et donc aucun jet central n'est attendu pour des événements de signal. Ceci permet l'application de la coupure de "central jet veto" pour encore réduire le bruit de fond. Une analyse basée sur des coupures a été utilisée, se concentrant sur l'optimisation de la coupure de "central jet veto".

L'objectif principal était d'étudier le potentiel de découverte du boson de Higgs avec ATLAS, sous une luminosité intégrée de 30 fb^{-1} dans la gamme de masse 115 GeV$\leq m_H \leq$140 GeV, pour une énergie au centre de la masse $\sqrt{s}$=14 TeV. Après l'application des coupures, un excès au-dessus de 5σ de signal significance a été trouvé pour la gamme de masse 115 GeV$\leq m_H \leq$125 GeV.

Dans le canal de désintégration $H \rightarrow \tau^+\tau^-$, la résolution de l'énergie transverse manquante ($\not{E}_T$) est importante puisqu'elle affecte directement la résolution sur la masse du boson de Higgs. Cela a été la motivation initiale pour réaliser et présenter dans cette thèse une étude de "commissioning" calorimétrique basé sur les quantités $\not{E}_T$, plus particulièrement sur la contribution du bruit de l'électronique, mesurée dans des événements cosmiques enregistrés pendant l'automne 2008. Un comportement gaussien de bruit dans tous les calorimètres de l'argon liquide a été trouvé, tandis qu'un comportement non-gaussien dans une partie du calorimètre à tuile scintillante a été observé. Pour ce dernier, un nouveau modèle de bruit a été examiné utilisant une paramétrisation de double gaussienne ayant pour résultat une description plus réaliste. En conclusion, cette analyse a fourni un outil de "commissioning" du calorimètre, qui a permis l'observation et la correction de plusieurs effets dans le comportement des calorimètres d'ATLAS.

Mots clés: LHC, ATLAS, VBF, Higgs Boson, $H \rightarrow \tau^+\tau^-$, jets, calorimeter, commissioning,

$\not{E}_T$, noise.

Σύνοψη

Το θέμα αυτής της διατριβής είναι η αξιολόγηση της δυνατότητας ανακάλυψης του μποζονίου Higgs, το οποίο παράγεται με το μηχανισμό σύντηξης διανυσματικών μποζονίων (Vector Boson Fusion, VBF), στον ανιχνευτή ATLAS του μεγάλου επιταχυντή αδρονίων (LHC), που βρίσκεται στη Γενεύη. Το κανάλι διάσπασης του μποζονίου Higgs που μελετήθηκε είναι αυτό που αποτελείται από δύο λεπτόνια τ ($H \rightarrow \tau^+\tau^-$). Πρόκειται για ένα από τα πιο ελπιδοφόρα κανάλια ανακάλυψης στο φάσμα των χαμηλών μαζών, που ευνοείται από τις μετρήσεις ακριβείας στην ηχλεκτροασθενή κλίμακα. Οι τρόποι διάσπασης όπου και τα δύο λεπτόνια τ διασπώνται λεπτονικά ή όπου το ένα λεπτόνιο τ διασπάται λεπτονικά και το άλλο αδρονικά μελετήθηκαν στην παρούσα διατριβή.

Η χαρακτηριστική τοπολογία της σύντηξης διανυσματικών μποζονίων, που αποτελείται από δύο jets στις εμπρόσθιες περιοχές του ανιχνευτή και απο τα προϊόντα διάσπασης του μποζονίου Higgs στο κεντρικό τμήμα του ανιχνευτή, παρέχει μία ιδιαίτερη μορφολογία σήματος που επιτρέπει την απόρριψη του υποβάθρου. Επίσης, η συγκεκριμένη διεργασία έχει χαρακτήρα ηλεκτροασθενή και συνεπώς δεν αναμένεται QCD δραστηριότητα, που θα είχε ως συνέπεια τη δημιουργία κεντρικών jet. Αυτό δίνει τη δυνατότητα στην εφαρμογή ενός cut που απορρίπτει ενα γεγονός φυσικής αν ανιχνευθεί κεντρικό jet (central jet veto cut). Μία ανάλυση βασισμένη σε cuts χρησιμοποιήθηκε, εστιάζοντας στη βελτιστοποίηση του central jet veto cut.

Ο κύριος στόχος ήταν να διερευνηθεί αν είναι ευφικτή μία ανακάλυψη του μποζονίου Higgs σε φωτεινότητα 30 fb^{-1}, για ενα φάσμα μαζών 115 GeV$\leq m_H \leq$140 GeV και ενέργεια κέντρου μάζας $\sqrt{s}$=14 TeV. Μετά από την εφαρμογή όλων των cuts, βρεθηκε μια περιοχή επάνω απο τα 5σ στη καμπύλη του signal significance για μάζες 115 GeV$\leq m_H \leq$125 GeV, υποδηλώνοντας τη δυνατότητα ανακάλυψης, εφόσων η μάζα του ανήκει σε αυτό το φάσμα μαζών.

Στο κανάλι διασπασης $H \rightarrow \tau^+\tau^-$, η ανάλυση του εγκάρσιου ελλείματος ενέργειας (missing transverse energy, $\not\!\!E_T$) είναι ιδιαίτερα σημαντική γιατί αντανακλά την ανάλυση της μάζας του μποζονίου Higgs. Αυτό ήταν το αρχικό κίνητρο για την εκτέλεση μίας μελέτης λειτουργίας του θερμιδόμετρου (calorimeter commissioning) του ATLAS εστιάζοντας σε ποσότητες $\not\!\!E_T$. Πιο συγκεκριμένα, η ανάλυση εφαρμόστηκε στο θόρυβο των ηλεκτρονικών, που μετρήθηκε σε πραγματικά δεδομένα, προερχόμενα από την κοσμική ακτινοβολία (φθινόπωρο 2008). Τα θερμιδομετρα υγρού αργού βρέθηκε να παρουσιάζουν μια συμπεριφορά γκαουσσιανή, ενω τα TILE (σπινθηρηστές) θερμιδόμετρα έδειξαν μία μη γκαουσσιανή συμπεριφορά. Για το τελευταίο, ένα νέο μοντέλο παραμετροποίησης του θορύβου μελετήθηκε, βασισμένο σε μία κατανόμη διπλής γκαουσσιανής, το οποίο έδειξε μία πιο ρεαλιστική περιγραφή του θορύβου. Κλείνοντας, αυτή η ανάλυση εδωσε τη δυνατότητα να παρατηρηθούν και να διορθωθούν μία σειρά προβλημάτων του θερμιδόμετρου του ATLAS.

Λέξεις κλειδιά: LHC, ATLAS, VBF, Higgs Boson, $H \rightarrow \tau^+\tau^-$, jets, calorimeter, commissioning, $\not{E}_T$, noise.

Remerciements, Ευχαριστίες, Acknowledgements

Je remercie le directeur du Laboratoire de l'Accélérateur Linéaire Guy Wormser, de m'avoir accueilli au sein de ce laboratoire.

Mes remerciements vont aussi à tous les membres du jury: Charalampos Anastasiou, Gregorio Bernardi, Kyle Cranmer, David Rousseau, Isabelle Wingerter-Seez et Guy Wormser. Un merci particulier à Gregorio Bernardi et Isabelle Wingerter-Seez d'avoir été mes rapporteurs.

Je tiens particulièrement à remercier mon directeur de thèse David Rousseau pour son aide, son soutien et l'orientation qu'il a donné à mes travaux de thèse. Au-delà de son savoir-faire, j'ai vraiment apprécié sa très grande disponibilité, sa patience, son aide et ses explications sur tous les aspects du travail (même quand je devais le faire répéter plusieurs fois). Je le remercie de m'avoir fait découvrir et aimer la recherche en physique des particules et de m'inspirer encore plus sur la voie de la recherche.

Je voudrais remercier tous les gens du LAL qui m'ont aidé pendant ces trois ans dans ce laboratoire. Je suis très reconnaissant de l'accueil dont j'ai bénéficié au sein du ce groupe. Je remercie le directeur du groupe, Daniel Fournier. Merci également à Lydia Iconomidou-Fayard de son accueil chaleureux tout le long de mon séjour. Son aide et son soutien étaient incomparables: Lydia, merci! Merci à Louis Fayard pout tout ce qu'il a fait pour moi, les explications sur la physique du boson de Higgs et tout le reste. Je remercie tous les membres du groupe (dans l'ordre des bureaux au LAL): Geneviève, Patrick Puzo (merci pour ton idée après mon problème de powerpoint), Sébastian Binet (merci pour tes explications du software et les "hot tips" sur le groupe de Berkeley), Daniel Fournier (merci pour tes conseils et merci de m'avoir fait des commentaires sur le manuscrit de ma thèse), Lydia Iconomidou-Fayard (Ευχαριστώ), Laurent Serin (merci pour tes explications sur le calorimètre et le billet pour aller à Barcelone), R.D. Schaffer (merci d'avoir reinitialisé mon mot de passe au centre de calcul n fois parce que je l'oubliais), Louis Fayard (merci pour ton soutien énorme), Dirk Zerwas (merci pour ta volonté de répondre à toutes mes questions), Caroline Collard (merci de m'avoir attendu aux pistes de La Thuile), David Rousseau (merci, merci, merci encore), Marc Escalier (merci pour tes questions qui m'ont obligées à m'interroger encore plus sur la physique), Luc Poggioli (My office-mate! : merci pour tous ces moments de joie et de rire et des explications sur la physique et comment faire des replicas de dataset), Jean Baptiste De Vivie (merci pour toutes tes réponses qui m'ont énormément aidé même si peut-être tu ne l'as pas réalisé), Marumi Kado (également merci pour tes réponses et tes suggestions), Sophie Henrot-Versille, Nikola Makovec (merci pour les explications sur les jets et la musique que tu avais préparé au road-trip à Toscane), Reisaburo Tanaka (merci pour tes explications sur le particle-flow), Abdenour Lounis (merci beaucoup Abdenour pour les discussions qu'on a eu pendant ces trois ans. C'était toujours très agréable et utile de parler avec toi, d' écouter tes conseils et de partager ton expérience), Laurent Duflot (merci pour avoir répondu à toutes mes questions et merci pour tes remarques), Jean-Francois Grivaz (merci pour les questions (très pointues) que tu m'as posé, merci aussi pour les réponses, car souvent je les ignorais, merci pour tes conseils et pour tes remarques sur ce mansucrit). Je tiens particulièrement à remercier Claire Adam-Bourdarios pour son courage de lire en détail toute ma thèse et m'avoir fait beaucoup de corrections et de remarques. Je veux aussi la remercier pour les discussions qu'on a eu les jours où j'étais dans le même bureau qu'elle au CERN concernant ma culture générale (cinéma, voyages, religion, etc).

Je tiens particulièrement à remercier Achille Stocchi, David Rousseau, Lydia Iconomidou-Fayard et Louis Fayard pour leur aide et leur soutien pendant une période assez difficile que j'ai eu pendant ma thèse. Grâce à eux, j'ai compris que l'on n'était pas uniquement des collègues de bureau et que les relation humaines, la compréhension, et la personnalité sont aussi des valeurs fondamentales pour un physicien.

Merci également à tous les jeunes du group ATLAS au LAL. Je commence d'abord avec ceux qui sont déjà partis: merci à Iro Koletsou (merci pour tous tes explications sur le pile-up et la physique du Higss en gamma-gamma, merci pour ton soutien surtout à mon arrivée au LAL, tes blagues. Ta bonne humeur a toujours été un soutien surtout au début de ma thèse. Merci d'avoir voyagé pour être présente à ma soutenance. Σε ευχαριστώ για όλα, σου εύχομαι επίσης πολλά συγχαρητήρια και κάθε ευτυχία (ξέρεις γιατί..)), Mathieu Plamondon (merci d'etre là aujourd'hui), Tatiana Cervero (merci pour ton amitié!). Merci encore à: Henso Abreu Aguilar (gracias el loco), à Manu Turlay (merci la rock-star du groupe), Jean-Baptiste Blanchard (merci pour les moments musicaux que tu as partagé avec moi quand tu chantais/sifflais au LAL), Fany Dudziak (merci pour les discussions et tes gâteaux, courage pour ta thèse), Pauline Bernat (merci pour tes remarques, conversations, questions et toi aussi courage pour ta thèse), Ilija Vukotic (thanks for the discussions, your remarks and for playing tennis and basketball with me), Mathieu Heller (merci Mathieu pour ton soutien, tes blagues, ta bonne humeur presque constante, c'était pas mal tes skills en djembe à la piazza de Santa Croce), Mathieu Benoit (merci pour les discussions et pour le chip du pixel que tu nous a montré). Many thanks to Sara Diglio (scusami, I wasn't very present these last few months you arrived at LAL. I wish you "bonne continuation" avec ton post-doc). Une pensée aussi pour les nouveaux thésards du groupe Marthe, Jean-Baptiste, Adrien, Nancy et la nouvelle post-doc Yu. Je vous souhaite bon courage. Je voudrais particulièrement remercier mon "Super office-mate" ces 2 dernières années François Niedercorn. J'ai trouvé grâce à lui une aide et un soutien remarquable. Merci pour m'avoir supporté dans le bureau et d'avoir répondu à toutes mes questions (du q-factor (que d'ailleurs je n'ai toujours pas compris) jusqu' à comment dit-on "The honour is all mine" en français. J'espère qu'un jour avant la fin de ta thèse, tu réussiras de te déplacer sur un autre coin du bureau et ne pas rester à côté de la porte). Merci encore François pour avoir corrigé mes remerciements et oui, je pense que tu es un jeune physicien très promising!

Un grand merci à Francesco Polci. Francesco, merci pour toutes ces discussions qu'on a eu pendant notre séjour au LAL, tes remarques, tes conseils, tes mails, tes blagues, ta bonne humeur, tes *olive ascolane*, ton ecran LCD avec les images de Mona Lisa, ta cuisine italienne, ta grande famille, les chambres des hôtels qu'on a partagé au youth hostel de Genève, à Moriond, à Toscane, à ... Tu as été un exemple pour moi, j'espère qu'on fera bientôt ce voyage à l'Amérique du Sud et bonne chance avec ton nouveau poste!

I would also to thank all the people from CERN who helped me a lot with their suggestions, comments and answers, especially Kyle Cranmer and Soshi Tsuno concerning the Higgs physics analysis. Je remercie également Guillaume Unal, Irène Vichou, Luca Fiorini, Tancredi Cari, Isabelle Wingerter-Seez, Laurent Serin, Pascal Pralavorio et Elisabeth Petit pour leur disponibilité, remarques et coopération pendant mon travail sur le commissioning du calorimètre. Je remercie particulièrement Elisabeth Petit (une jeune physicienne très prometteuse avec une personalité formidable), pour son aide, son soutien, sa coopération et son amitié. Elisabeth, merci d'être venue au LAL pour ma soutenance.

Je remercie tous mes amis qui ont rendu possible fait possible la réalisation de cette these avec leur soutien. Merci à Marcel Reinhard (ok, in english for you, just to be coherent) for sharing an appartment for almost three years with me, thanks for the beers and all the alcoholic liquids, your lasagna version, the trips to Germany, your high support and help while we were students in NPAC and during my PhD. Thanks for reparing my bike, for helping me with the satellite antenna and many many other things. Thanks for your advices, your remarks and your point of view, which was very helpful to me.

Je remercie également mes amis du bâtiment 208, Yasmine Amhis, João Firmino da Costa, Viola Sordini et Justine Serrano. Yasmine merci pour tes cours de français dans le RER au début de ma thèse, merci pour tes blagues, merci pour m'avoir fait connaître toute ta famille, merci pour ton esprit méditerranéen ensolleilé, merci pour les voyages à Moriond, à Toscane, merci d'être là pour ma thèse. João, merci pour ton aide et soutien immense pendant le master. Je ne sais pas comment j'aurais réussi sans toi et Marcel. Merci pour les dîners chez toi, pour le Portugal et la bouffe délicieuse qu' on a gouté chez ta grand-mère pendant que tu regardais le foot à la télé. Merci également pour tes remarques et tes conseils. Grazie mille à Viola (ciao tesoro!), merci pour toutes nos conversations (débiles ou pas), merci pour les danses et pour "my dream is to fly", tes conseils, ton amitié. J'ai trop hâte de te voir à Genève à partir d'octobre. Merci aussi à Justine (la force tranquille de la physique du *b*), les repas chez elle et les soirées qu'on a fait ensemble à son bar préféré ("Point éphémère")

Merci beaucoup à Clement Helsens pour tout ce qu'il a fait pour moi. Merci pour les voyages qu'on a fait ensemble (Trest, Grèce, Biarritz, Toscane, Barcelone, Moriond, plus à venir), merci pour tes conseils que je n'oublierai jamais, tes remarques et toutes les conversations générales qu'on a eu sur la physique, la société d'aujourd'hui et l'humanité en général. Merci de m'avoir fait découvrir la cuisine diététique et ton monde très spécial et intéressant), merci d'être toujours là aux moments de crise (et j'ai en eu pas mal), merci pour la communication, ton aide et ton soutien.

Un grand merci à Andreas Goudelis. Un ami très précieux qui m'a beaucoup soutenu aux moment difficiles. Sa présence à Paris a été pour moi toujours une source de soutien, d'écoute, des bons conseils, des blagues (Αντώνη, ου ου ου ου !), a lot of fun, des discussion politiques,etc. Je lui souhaite une bonne continuation avec sa thèse. Merci également pour le voyage en Corse, les nombreux repas qu'on a fait ensemble et la notion du meta-post-guitar rock qu'il m'a fait découvrir. Andrea, je n'oublierai jamais ton meilleur conseil: Σκατά σ' όλα τα πρέπει, ζήσε μόνο για τα ϑέλω.

Merci aussi à Dimitra pour son soutien, les discussions et son γαλακτομπούρεκο, merci à Spyros (le cousin), merci également à mon cher ami Leonidas qui est venu de Londres pour ma soutenance. Leonidas, merci pour tous ces voyages aux îles grecques (surtout à Corfou).

Merci beaucoup à Despina pour son soutien et ses conseils. Son aide et sa présence à Paris ont joué un rôle important pour réaliser et comprendre beaucoup de choses dans ma vie professionnelle et personnelle. Δέσποινα, σε ευχαριστώ πολύ.

Merci beaucoup à tout le groupe de batucada (Molica, Inès, Rémi, Laurent, Loic, Marc, Valentine, Sophie, Sylvie, Jacques, Rejane, Hélène, Marie, Dorothée, Nawfel, René, Damien, ...) qui a coloré percussivement ma dernière année de thèse. Je vous promets que je continuerai de taper sur ma caisse sans hésitations et toujours très fort.

I would like also to thank all of my professors and tutors I have since my first year at the university in Athens and especially professor George Zoupanos. Under his guidance, I found myself at LAL with the chance to work with such a competent and strong group in physics. His advice has been (and still is) extremely precious to me and I always consider it with great attention. Merci également à Achille de m'avoir accepté au sein du master NPAC, à l' école doctorale "Constituants élémentaires - Systèmes complexes" et d' avoir accepté d'être mon parrain pendant ces trois ans. Son aide a été cruciale au moment où j'en ai eu besoin.

Le plus grand merci va à Mary. Son aide et sont soutien ont été incomparables. Avec sa manière unique elle m'a aidé (et elle m'aide toujours, heureusement) à obtenir une autre thèse[1] qui est la thèse de la vie. Μαίρη, ένα μεγάλο ευχαριστώ για όλα όσα μου έχεις προσφέρει, για όσα μου έχεις πει, για όσα μου έχεις δείξει. Σε ευχαριστώ για τους αμέτρητους 'φάρους' που μου έχεις προσφέρει απλώχερα (κάποιες φορές χωρίς καν να το συνειδητοποιείς). Μερίδιο αυτού του μεγάλου φάρου, που συνοψίζεται στις σελίδες αυτής της διατριβής, είναι και δικό σου. Σε ευχαριστώ.

Un grand merci aussi à ma famille pour tout le soutien qu'ils m'offrent continuement et pour tout ce qu'ils ont fait pour moi. Αγαπημένοι μου γονείς, Σπύρο: μέρος αυτής της επιτυχίας, που σχετίζεται με την επίτευξη ενός πολύ σημαντικού στόχου που επετεύχθη σε προσωπικό και επαγγελματικό επίπεδο με την ολοκλήρωση των σπουδών μου στη Γαλλία, σας ανήκει. Το κερδίσατε με τη στήριξη και την καθοδήγηση που μου παρείχατε όλα αυτά τα χρόνια. Θα σας είμαι για πάντα ευγνώμων.

In summary, this PhD thesis was an intense, stressing, hard and easy, interesting, happy and sad, learning and fun experience for me, and this was thanks to many people. I am unable to name them all, since I would probably need a few more pages and I guess, David would not appreciate this. To finish, the most exciting and creative part of these three years at LAL was my productive evolution in the particle physics research field, learning a lot from all the physicsts with whom I had the chance to work. And this was combined at the same time with an improving and deeper understanding of myself. Thanks a lot to everyone who contributed towards this achievement.

---

[1] mot d'origine grecque "θέση", qui veut dire *position, opinion, avis*