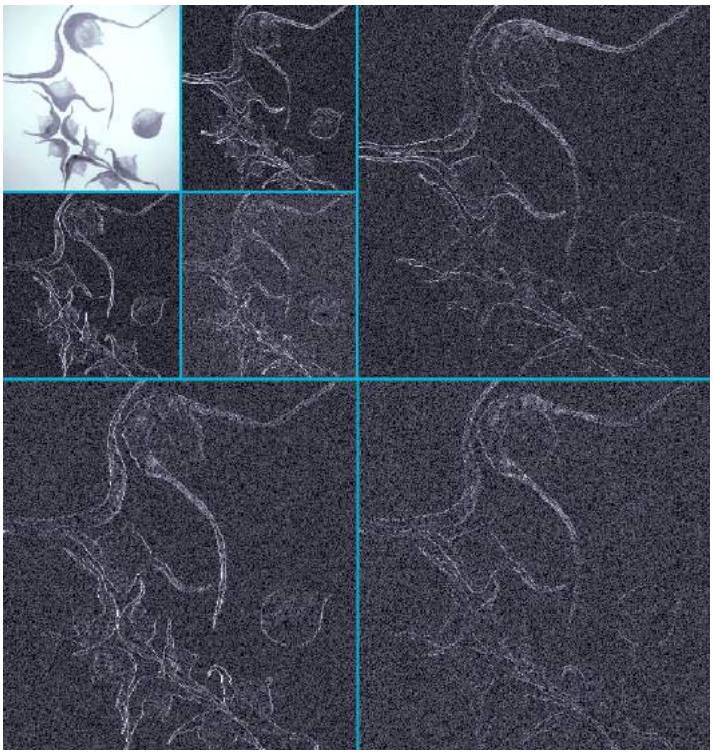
FREJA THORESEN

WAVELETS & INFORMATION THEORY FOR PILE-UP REMOVAL



MSC THESIS
NIELS BOHR INSTITUTE, UNIVERSITY OF COPENHAGEN

WAVELETS & INFORMATION THEORY FOR PILE-UP REMOVAL

FREJA THORESEN

NIELS BOHR INSTITUTE,
UNIVERSITY OF COPENHAGEN

Supervisor	Dr. T. Petersen
Second Supervisor	Dr. J. Monk
Referee	Dr. P. Christiansen

Wavelets & Information Theory for Pile-up Removal
MSc Thesis, University of Copenhagen
xi + 129 pages; illustrated, with bibliographic references

Set in 10/14 pt Palatino Linotype using pdfLaTeX with the Tufte-LaTeX package
Cover art: Tine Thoresen

Thesis submitted on 6 September 2016 and defended on 12 October 2016
for the completion of the degree of Master of Science (MSc) in Physics
at the Niels Bohr Institute, University of Copenhagen.

First printing, October 2016

UNIVERSITY OF
COPENHAGEN

Contents

Abstract

At the Large Hadron Collider (LHC) at CERN the luminosity and momentum of the colliding protons are increasing in the Run 2 of the LHC compared to Run 1. When the luminosity increases it leads to several protons colliding at the same time, which we call pile-up. When several protons collide at the same time, we need to able to distinguish these events from each other, since we only perform analysis for one event at a time. The consequence of pile-up can be smearing of signals in the detectors, and it can also have a big effect of the reconstruction of objects and particles. This has especially a huge effect on the reconstruction of missing transverse energy and jets.

In this thesis a method named wavelet denoising is suggested as a method for removal of pile-up. This method was tested by reconstruction of missing transverse energy. Missing transverse energy is the vector sum of all reconstructed objects in an event. Since this measurement is a combination of all particles in an event, it is useful for testing how well the wavelet methods perform.

The method of wavelets is used for denoising of $t\bar{t}$ datasets with pile-up ranging from 70 to 210. The wavelet transformation is performed on particle flow objects, which are objects containing all charged and neutral particles of an event. The particle flow objects will also contain pile-up signals, and we want to remove these from our analysis.

Several methods of wavelets have been tested for removal of pile-up. The simplest one is the flat denoising method. Here we wavelet transform the particle flow object and make a simple flat cut in their energies. Then we wavelet inverse transform the image to see the difference. Depending on the change in the image, the corresponding pile-up particles are removed. There are also wavelet methods using models. These models can contain information of a clean event, and the wavelet cut will be made as a function of this model. This results into a better removal of pile-up. Lastly maximum entropy and information theory was used in combination with wavelets. Unlike the other methods, the maximum entropy method also uses a model for the noise in the data.

The results include the improvement of missing transverse energy with the methods of wavelets. The particle flow algorithm already does well in removal of pile-up, but not perfectly. Using wavelets the removal of pile-up in particle flow objects is increased even further.

Acknowledgement

Firstly I would like to thank my supervisor Troels Petersen for giving me the opportunity to work on this thesis in the ATLAS group at the Niels Bohr Institute. Troels has been a great support and ever enthusiastic. I would also like to thank my second supervisor James Monk for in depth conversations on wavelets and particle physics and for his great patience in introducing me to these subjects.

During this last year of working on my thesis, I had the chance to have a lot of helpful conversations with the people at the ATLAS group at the Niels Bohr Institute. Especially I would like to thank Milena Bajic for helping me with the usage of SUSY tools and Geert-Jan Besjes for helping me with RooFit. Also a thanks to Lars Egholm Pedersen for discussions on general physics and introducing me to the work of the group. A special thanks to Andreas Soegaard for insightful skype conversations and a general helpful attitude.

My fellow students Mike Lauge and Daniel Nielsen have been great to discuss the basics of particle physics and programming. We shared an office together for the last year, and I really appreciate the opportunity to discuss with someone at the same level as me.

A special thanks to my friends and family.

Profile

I have a Bachelor degree in Natural Sciences and IT, which is an education that focuses on computational science. My Master degree is in Physics, and it is focused on computational physics. I was searching for a Master thesis with a lot of computational work, hence I sought out particle physics with big data handling. Although my background includes a lot of computer science, I don't have any experience in theoretical particle physics. Therefore this thesis will lightly outline the basics of particle physics and the Standard Model, but it will not be the main focus. The main focus is to perform a computation using wavelets to minimize the noise in the signal of missing transverse energy. The project started out by examining fat jets and 3D fits in RooFit of the decay channels of the Higgs particle, but soon I found my interest in wavelets instead. Wavelets have before been used for cleaning of fat jets, and I was supposed to continue this work. But when I realized the potential of using wavelets for reconstructing missing transverse energy, I chose this path instead. More than half way through my thesis I read about using the maximum entropy principle in cleaning of astrophysical images and wondered if I could do the same for cleaning the particle images. When I found that the method seemed reasonable, I decided to include this as one of the main methods used in this thesis. It has been overall a great experience to have the freedom to research my interests.

Introduction

The LHC at CERN has during Run 1 been successful in recording data for general physics purposes. During this run the famous Higgs particle was discovered. Now the Run 2 has begun with a greater luminosity and a higher energy of the colliding protons. When two protons collide they will shatter and the remaining quarks and gluons from the protons can pair up to create new particles. These new particles decay, and the ATLAS detector tries to detect the particles. At high luminosity more protons separately collide at the same time.

The problem of pile-up rises when multiple proton-proton collisions happen at during the same bunch crossing, and also from remnants of electronic signals from previous bunch crossings in the detectors. The reconstruction of kinematic variables, such as missing transverse energy, must be corrected for the contribution of pile-up. The aim is to ensure a physics reconstruction performance comparable for the one achieved with low luminosity, i.e. with no pile-up events. The data amount extracted at CERN is greatly increasing. With the increased amount of data taking in one go, the noise in the data is also increasing. The noise rising from pile-up has especially high impact on measurables such as jets and missing transverse energy. This is because both of the objects sum over a bunch of particles, and since some of these might be from pile-up, the jets or missing transverse energy can be overestimated. In this thesis wavelet methods are described to effectively remove the noise in missing transverse energy coming from pile-up.

Motivation

Missing transverse energy can be used as a measurement of neutrinos. But it is not for certain that missing transverse energy is *only* for neutrinos. The missing transverse energy measurement could also be used in searches for physics beyond the standard model or exotic physics. For example, the SUSY model depicts a Lightest Symmetric Particle (LSP). This particle is the lightest SUSY particle, and therefore also the particle that other SUSY particles decay to. A candidate for the LSP could eg. be a neutralino, which is a neutral particle from the SUSY model. Missing transverse energy could also be used in searches for Dark Matter. Dark Matter is non-interactive, or nearly non-interactive, which is also why we can't detect Dark Matter (so far).

Outline of Thesis

This thesis will first outline the basics of the Standard Model and interactions of elementary particles. In Chapter 2 the parts of the ATLAS detector is explained to discuss how the particles get detected and how well we can reconstruct them. There are many algorithms to reconstruct particles such as muons, electrons and jets. In Chapter 2 I will mainly be focusing on how ATLAS reconstructs missing transverse energy and how it correct for pile-up measurements of missing transverse energy. Chapter 2 will also explain the Particle Flow Objects (PFO) which will be used throughout this thesis.

In Chapter 3 the basics of wavelet transformations and information theory is described. Here we look into how to perform a wavelet transform and use this for denoising of images of particles. For denoising several methods are suggested, where one of them include information theory and the maximum entropy principle.

Chapter 4 concerns event selection of datasets. The main datasets used are $t\bar{t} \rightarrow bbqqlv$, since these datasets are at pile-up at $70-90$, $170-180$ and $190-210$, which make them ideal for testing the removal of pile-up using wavelets. Although there are still some problems with these datasets since they are not complete, they are good enough for analysing the missing transverse energy as a function of pile-up. The problems with the datasets are also described in this Chapter.

Lastly the Chapter 5 will show the results of wavelet cleaning. Methods included are flat denoising, track filtering, track scaling and maximum entropy.

Theory

1.1 Particle Physics

Particle Physics is a field of physics which seek to understand the fundamental particles of our universe. The Standard Model is the most successful theory, although it is not complete. The Standard Model was developed throughout the last half of the 20th century, and is therefore a fairly new theory. In the 20th century most of the elementary particles were discovered, except for the tau neutrino (2000) and the Higgs boson (2012). And even though some of the elementary particles are already mapped out, there are still some a lot of discoveries yet to be uncovered.

1.1.1 Forces

There exist four forces: Strong, electromagnetic, weak and gravitational forces. The strength of the forces is shown in Figure 1.1. The Standard Model is an attempt of describing all of the four forces in one theory. So far, three of the four forces are explained in the Standard Model, since physicists have not yet succeeded in explaining the gravitational force with the Standard Model. One reason for this is that the gravitational force is much weaker than the other forces, and it is therefore difficult at the moment to measure at the particle level. One theory of the gravitational force involves the graviton, which has not been found yet, and it is not known if it even exists.

Table 1.1: The strength and particles of the four forces.

Force	Strength	Theory	Mediator
Strong	10	Chromodynamics	Gluon
Electromagnetic	10^{-2}	Electrodynamics	Photon
Weak	10^{-13}	Flavordynamics	W and Z
Gravitational	10^{-42}	Geometrodynamics	Graviton

1.1.2 The Standard Model

The Standard Model combines three of the four forces of the universe by the interactions of the elementary particles.

Elementary particles are divided into two main groups: Fermions (quarks and leptons) and bosons. Fermions are particles with spin 1/2 and are known as matter particles. They obey the Pauli exclusion principle [1]. The fermions make up the matter of the universe. The quarks and leptons are fermions, since they have spin 1/2. The quarks can make up mesons and hadrons by paring up, and the charged leptons can interact with the mesons or hadrons. Mesons are particles made up of one quark and one anti-quark, and the hadrons are made from three quarks.

[1] Paulis exclusion principle states that two particles can not co-exist in the same state at the same location at the same time.

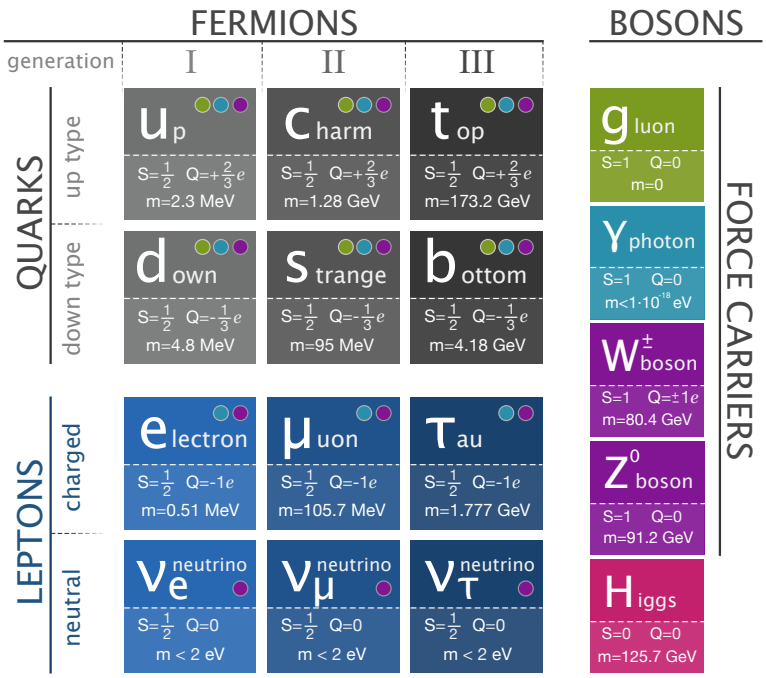


Figure 1.1: The Standard model particles are divided into 3 catagories: Quarks, leptons and bosons [1].

The bosons are known as force carrier particles and they have spin 1 or 0. The bosons each represent one of the forces. The gluon is a part of the strong force, which is the force binding quarks together and also manages the interactions between quarks. The photon governs the electromagnetic force, where the interactions between light and matter is explained. The Z and W bosons are used in the weak interactions, which is describing the decays of particles.

All particles in the Standard Model has an anti-particle with opposite charge. If a particle and anti-particle is combined, they can annihilate. Some particle's anti-particle also correspond to the particle itself, as the Z boson or the photon since their charge is zero.

The theories describing the three of the forces are Quantum Electrodynamics (QED), Quantum Chromodynamics (QCD) and the weak theory.

Quantum electrodynamics (QED) describes the interactions of *charged* particles. These interactions are mediated by *photons*, which means in an electrodynamic interaction, the photon is always involved. An example of an electromagnetic interaction is shown in Figure 1.2.

Quantum Chromodynamics (QCD) is the theory of the strong force. QCD describes the interactions of *colored* particles. These interactions are mediated by *gluons*, see for example Figure 1.3.

In QCD, color play the role of charge, and the fundamental process is $q \rightarrow q + g$. Since leptons do not carry color, they do not participate in the strong interactions. Unlike the photon, the gluon carries a color charge. This means that the gluons can also interact with themselves. QED and weak interactions does not affect color. At a strong vertex,

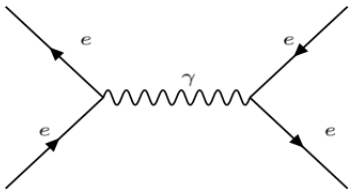


Figure 1.2: Example of a electromagnetic interaction. An electron and a positron annihilate to form a photon, which then produces a new electron and positron. Here the photon is a *virtual* particle.

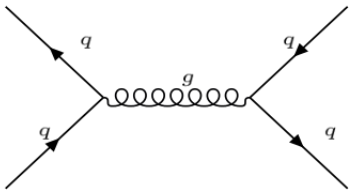


Figure 1.3: Example of a strong interaction. A quark and an anti-quark annihilate to form a gluon, which then creates a new quark anti-quark pair.

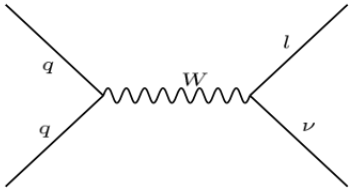


Figure 1.4: Example of a weak interaction. A quark anti-quark pair annihilates to create a W boson, which then can for example decay into a neutrino and a lepton.

the quark color changes but the difference is carried off by the gluon.

The mediators in the weak interactions are $W^{\pm}$ and Z bosons. In contrast to the other mediators, the photon and the gluon, these are very massive particles and the resulting interactions are therefore of very short range. An example of a weak interaction is in Figure 1.4.

1.1.3 Conservation Laws

In physics the two most important laws of conservation are,

- Conservation of Charge
- Conservation of Energy

In particle physics there are more additional important conservation laws to consider,

- Color Conservation
- Baryon Number Conservation
- Lepton Number Conservation
- Flavor Conservation (if not a weak interaction)

These conservation laws are valid in the Standard Model so far, but the search for new physics includes looking for interaction that violates some of these laws. This is known as CP-violation (CP for Charge Parity). CP violation is seen for example in the CKM matrix explaining quark mixing, and also the PMNS matrix describing neutrino mixing. This means that there is a small probability that quarks or neutrinos can change their flavour. CP violation is necessary for describing baryogenesis. One of the big mysteries in particle physics, is explaining why we have more particles than anti-particles. The answer could rely on CP-violation.

1.1.4 Uncertainty

Heisenberg's uncertainty principle allows a particle to 'borrow' an amount of energy ΔE as long as the energy is returned in time Δt satisfying,

$$\Delta E \Delta t \approx mc^2 \Delta t > \frac{\hbar}{2} \tag{1.1}$$

The particle 'borrowing' the energy is named a 'virtual' particle, which can violate energy conservation for a brief time.

1.2 Elementary Particles

The elementary particles are building blocks of our universe. An elementary particle can decay into a lighter elementary particle. This is flavor changing. The quarks and neutrinos can also change their flavor. Some of the main qualities of the elementary particles is discussed in this section.

1.2.1 Quarks

In 1964 Gell-Mann and Zweig independently suggested that all hadrons are in fact composed of even more elementary particles, which Gell-Mann called *quarks*. The theory of quarks was quite successful, except for one embarrassment: in spite of the most diligent search, no one has ever detected an individual quark. This means that the quark is never free, which is called quark confinement. Closely related to the quark confinement is the other property of asymptotic freedom. Asymptotic freedom means that quarks interact weakly at short distances. If two quarks are connected by a gluon field, and move away from each other, a new quark anti-quark pair will arise from the gluon field. This is illustrated in figure 1.5.

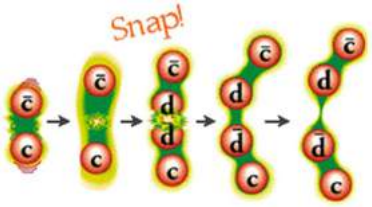


Figure 1.5: Gluon field "snapping" into a new quark-antiquark pair.

There are 3 generations of quarks which consist of 6 quarks and 6 anti-quarks.

First generation	$up(u), down(d)$
Second generation	$charm(c), strange(s)$
Third generation	$top(t), bottom(b)$

To make hadrons from these quarks, there are two rules to be followed,

1. Every baryon is composed of three quarks, and every anti-baryon is composed of three anti-quarks.

2. Every meson is composed one quark and one anti-quark.

The quarks normally stay within their own generation. The mixing between the generations is described by,

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{1.2}$$

which is the *Cabibbo-Kobayashi-Maskawa matrix* (CKM). If this matrix was a unit matrix, then d', u', b' would be the same as d, s, b and there would be no cross-generational transitions. But as it turns out, this matrix is *not* unitary, as you can see below,

$$V = \begin{pmatrix} 0.974 & 0.227 & 0.004 \\ 0.227 & 0.973 & 0.042 \\ 0.08 & 0.042 & 0.999 \end{pmatrix} \tag{1.3}$$

This means that the quarks most likely stays within their own generation, but it is possible that a quark can decay into a different generation. This fact is also what permits strangeness-changing processes. This is a process such as

$$\Lambda(uds) \to \Omega^-(sss) \tag{1.4}$$

Properties of the top quark The top quark is the heaviest quark and therefore the properties of the top quark differs compared to the other quarks. In particular, it is also much heaver than the W and Z bosons. Hence it can decay by the first-order weak interaction

$$t \to q + W^+ \tag{1.5}$$

where by conservation of charge, the quark is either d, s or b. In most cases it will be the b quark, by the CKM matrix. The top quark has a lifetime of

$$\tau = \Gamma^{-1} \approx 4 \times 10^{-25}\ \text{s} \tag{1.6}$$

where the other quarks have a lifetime of 10^{-12} s or more. The $t\bar{t}$ decay modes is shown in Figure 1.6

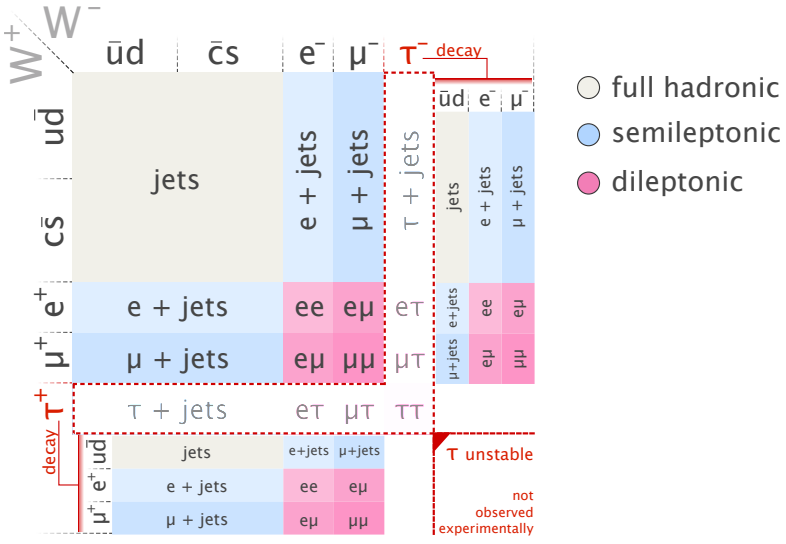


Figure 1.6: $t\bar{t}$ decay modes [2].

1.2.2 Gluons

There are eight types of gluons in QCD. The theory describing the states of the 8 gluons is group theory. Here the group SU(3) describes the gluons as the basis states of the Lie Algebra. This means that gluons transform in the adjoint representation of SU(3), which is 8-dimensional. The SU(3) group is the group of 3×3 unitary matrices with determinant 1. Quarks transform under this representation of SU(3) where quarks can come in one of 3 colours: red, green, blue.

Each gluon carries one unit of color and one unit of anti-color. In terms of color SU(3) symmetry, these nine states constitute a color octet, [2]

[2] The color octet is defined from the Gell-Mann matrices.

$$|1\rangle = (r\bar{b} + b\bar{r})/\sqrt{2} \qquad |5\rangle = -i(r\bar{g} - g\bar{r})/\sqrt{2} \tag{1.7}$$

$$|2\rangle = -i(r\bar{b} - b\bar{r})/\sqrt{2} \qquad |6\rangle = (b\bar{g} - g\bar{b})/\sqrt{2} \tag{1.8}$$

$$|3\rangle = (r\bar{r} - b\bar{b})/\sqrt{2} \qquad |7\rangle = -i(b\bar{g} - g\bar{b})/\sqrt{2} \tag{1.9}$$

$$|4\rangle = (r\bar{g} + g\bar{r})/\sqrt{2} \qquad |8\rangle = (r\bar{r} + b\bar{b} - 2g\bar{g})/\sqrt{6} \tag{1.10}$$

You could image a color singlet

$$|9\rangle = (r\bar{r} + b\bar{b} + g\bar{g})/\sqrt{3} \tag{1.11}$$

If the singlet gluon existed, it would be as common as the photon. Also because of the property of the SU(3), which is that it is 8-dimensional, it means that the color singlet gluon can not exist.

1.2.3 Neutrinos

Neutrinos was first postulated because of difficulties with explaining the lack of energies in decays. Neutrinos are very hard to detect [3]. One of the unsolved mysteries in particle physics is neutrino oscillations. It has been found that neutrinos can oscillate between different flavours, ie. for example a ν_e can become a ν_μ. This is where a neutrino is created with one flavour, but can later be measured to have a different flavour. At first the neutrino was thought to have zero mass, but is has later been found that they indeed do have a non-zero mass. Just as the quarks, the neutrinos also have a mixing matrix.

[3] At the IceCube project, they are trying to detect cosmological neutrinos and neutrinos transversing the earth.

1.2.4 W and Z bosons

By conservation of lepton number and charge, the possible leptonic decays of W are,

$$W^- \to \tau\bar{\nu}_\tau \tag{1.12}$$

$$W^- \to \mu\bar{\nu}_\mu \tag{1.13}$$

$$W^- \to e\bar{\nu}_e \tag{1.14}$$

Table 1.2: W decay rates [3].

Mode	Fraction (Γ_i/Γ)
$W \to e\nu$	$(10.75 \pm 0.13)\%$
$W \to \mu\nu$	$(10.57 \pm 0.15)\%$
$W \to \tau\nu$	$(11.25 \pm 0.20)\%$
$W \to q\bar{q}$	$(67.60 \pm 0.27)\%$

Table 1.3: Z decay rates [3].

Mode	Fraction (Γ_i/Γ)
$Z \to \nu_e, \nu_\mu, \nu_\tau$	$(20.00 \pm 0.06)\%$
$Z \to e, \mu, \tau$	$(10.2 \pm 0.003)\%$
$Z \to hadrons$	$(69.91 \pm 0.06)\%$

The decays of W into quark pairs are,

$$W^- \to d\bar{u} \tag{1.15}$$

$$W^- \to s\bar{c} \tag{1.16}$$

The pair $t\bar{b}$ is far to heavy a pair to come from W, and is therefore not possible. The colours of the quarks should add up to zero, ie. *red + anti-red*, *blue + anti-blue* and *green + anti-green*. This means that the W boson can decay into either 9 combinations of quarks. The decay modes of the W boson is shown in Table 1.2.

Unlike the W boson, the Z boson has zero charge, and therefore the possible decays are,

$$Z \to l\bar{l} \tag{1.17}$$

$$Z \to q\bar{q} \tag{1.18}$$

1.2.5 Standard Model Higgs

The Higgs particle is an essential part of the Standard Model. It provides the theory with mass for the gauge bosons of the weak interaction and for the fermions.

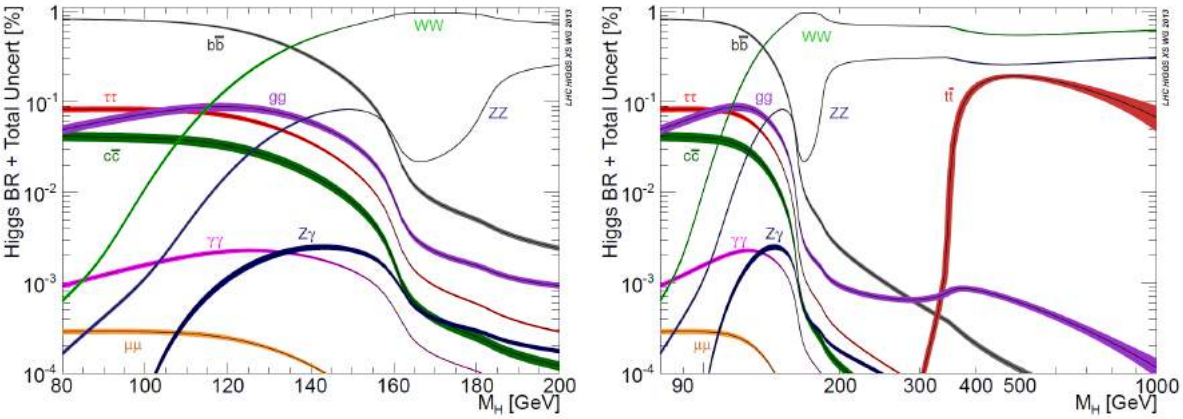


Figure 1.7: Branching ratios for the Higgs decay, as a function of the Higgs mass [4].

In 2012 the Higgs particle was discovered at CERN with a mass of 125 GeV. The branching ratios in the limits of a Higgs mass of 100 - 200 GeV and 90 - 1000 GeV is shown in Figure 1.7.

The 125 GeV Higgs particle mostly decays into b-quarks. In analysis it can be difficult to distinguish the b-quarks from the other particles, and for this reason one mostly looks into the Higgs decaying to bosons. This is because the Z and W bosons can decay into leptons, and leptons are much easier to detect. Figure 1.8 shows the percentwise distribution of the decays of the 125 GeV Higgs boson.

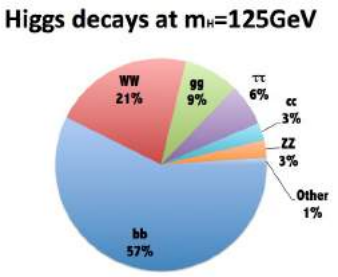


Figure 1.8: Decays of the 125 GeV Higgs particle [5].

1.2.6 *Dark Matter*

Dark Matter is not explained by the Standard Model yet. Dark matter is an unidentified type of matter which is about 27 % of the mass and energy of the observable universe. We do not know much about dark matter yet, but we expect the existence of dark matter particles. These particles should be non-interacting with photons, since if they interacted with photons we would probably already have seen these particles. But we can not detect these particles, hence the name "dark" matter particles. The dark matter particles have mass, since we can see gravitational effect of dark matter particles in the observable universe.

Experiment

2.1 The Large Hadron Collider

CERN is the largest research institution in the world. At CERN the largest accelerator is called the Large Hadron Collider (LHC), which has a 27 km circumference. In 2012 the Higgs particle was discovered during Run 1 of proton collisions with an energy of 8 TeV. In 2015 the Run 2 began, with an energy of 13 TeV. The basics of the CERN accelerator complex will be explained in this section, including the luminosity of the LHC and distribution of pile-up. The increased luminosity at the LHC can enable high precision tests of QCD, electroweak interactions, flavour physics and new physics. In this section we discuss the kinematics of the colliding protons. This includes the parton distributions of protons at different momentum transfers, and the cross sections of the particles emerging from a proton-proton collision.

2.2 Particle Creation

Particles are created at the LHC when the protons collide, as illustrated in Figure 2.1 where to protons collide and the remaining quarks and gluon pair up to create the Higgs particle and a $t\bar{t}$ pair. The Higgs particle can then decay into eg. WW which can decay further into eg. $lvqq$. The final state $lvqq$ can then be measured at the detectors at the LHC.

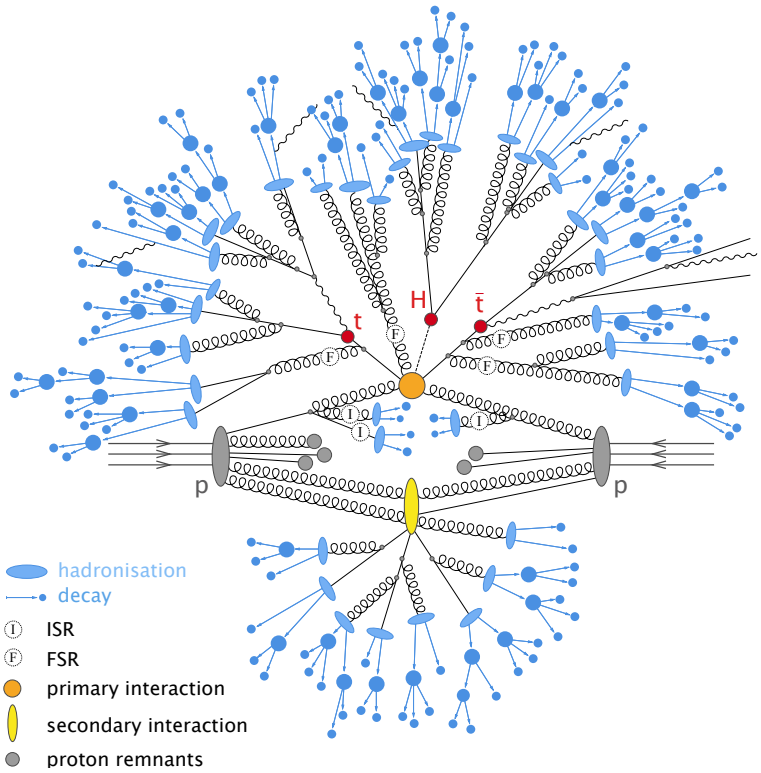


Figure 2.1: pp collision creating $t\bar{t}H$. The two protons collides and the quarks from the protons will then pair up to create new particles. These particles can decay into other particles or elementary particles such as leptons. The particles that decays into quarks will be seen in the detector as jets. [6].

To achieve a large amount of energy in a collision event, both protons in the event are accelerated. If a accelerated proton instead hit a target of protons, the energy in the collision would be smaller.

We say that the proton consists of two up-quarks and one down-quark. But this is actually not entirely correct. The proton consists of many quarks and gluons, but for every quark there are roughly two times more up-quarks and one time more down-quarks. This is illustrated in Figure 2.2, where a parton distribution is shown.

The parton distribution function $f(x_i, Q^2)$ is the effective density of partons, as a function of the momentum fraction $x_i = p_i/P$ [7]. Here P is the momentum of the proton and p_i is the momentum of constituent i. Note that at low momentum transfer, Figure 2.2 (a) it is clear that the proton consist mostly of the u and d quarks, but at higher momentum transfer the density of the other quarks increases.

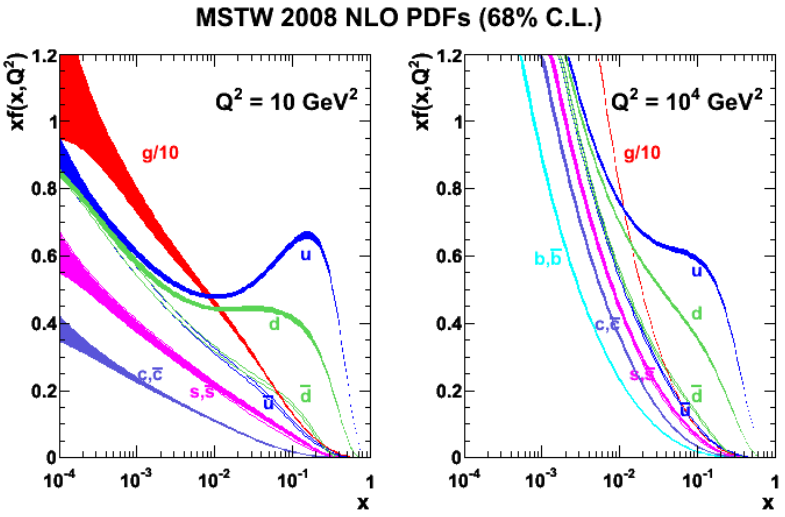
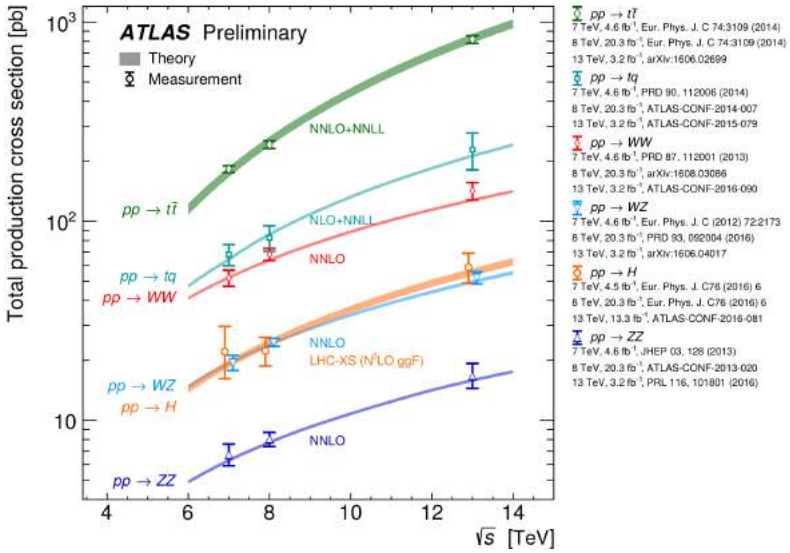
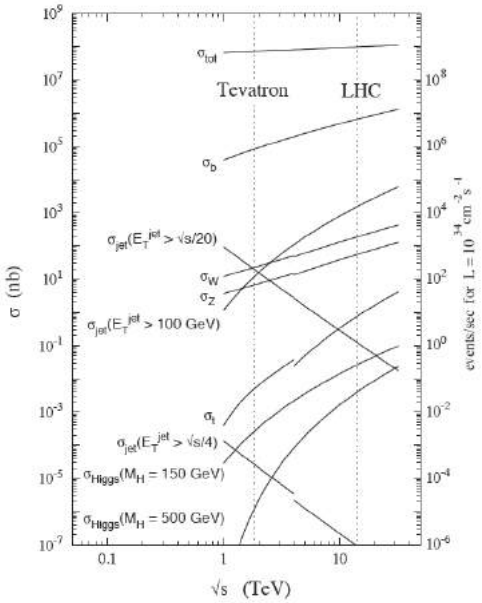


Figure 2.2: Two distributions of partons in a proton at different energy scales [8]. The x-axis is the fraction of energy of the proton the parton has, and the y-axis is the likelihood of observing a parton.

The center-of-mass energy of an event should be large enough, to be able to create heavy particles. The recipe to create heavy particles is to have enough energy and many repeated events to increase the possibility of seeing a heavy particle. Cross sections are a measure of how often an event occurs. The cross sections measured at ATLAS at center-of-mass energies of 7,8 and 13 TeV is shown in Figure **??** (a). Notice the cross sections measured at 13 TeV is generally higher than the cross sections at 7 or 8 TeV. The cross sections for center-of-mass energies of 1 to 10 TeV is in Figure 2.3 (b).



(a) Cross section measured by ATLAS.



(b) Cross section for a center-of-mass energy between 1 and 10 TeV.

Figure 2.3: Summary of total production cross-section as a function of center-of-mass energy $\sqrt{s}$. The cross-sections increase as the center-of-mass increases. In Figure (a) the cross sections for events involving top quarks and/or bosons are shown. In Figure (b) the cross sections for jets and the Higgs particle is shown. As the center-of-mass energy increases, the cross sections of high p_T jets increases and the cross sections of low p_T jets decreases.

2.2.1 The CERN Accelerator Complex

At the LHC several detectors and projects are placed as seen in Figure 2.4. In the beginning the beam is made of hydrogen nucleus, which consist of one proton and one electron. Electric fields switch from positive to negative to pull the electron and proton from each other. The electrons are then dumped, and the protons will continue through the beam line. The accelerators increase the energy of the protons.

The beam starts in the Booster, from which it is injected into the Proton Synchrotron (PS) accelerator. When the particles leave the PS they have an energy of about 25 GeV. Then the particles are injected into the Super Proton Synchrotron (SPS), where the energy of the particles are boosted up to a maximum of 450 GeV. At last the particles are injected into the LHC where they can so far get an energy of 6 TeV. The LHC might later be able to accelerate the protons up to a center-of-mass energy of 14 TeV. Several experiments extract the beam at locations at PS, SPS and LHC, depending on what energy is needed for the experiments. The main detectors placed at the LHC are ATLAS, ALICE, CMS and LHCb. The detectors CMS and ATLAS are general purpose experiments, meaning that they can be used for research in several subjects of physics.

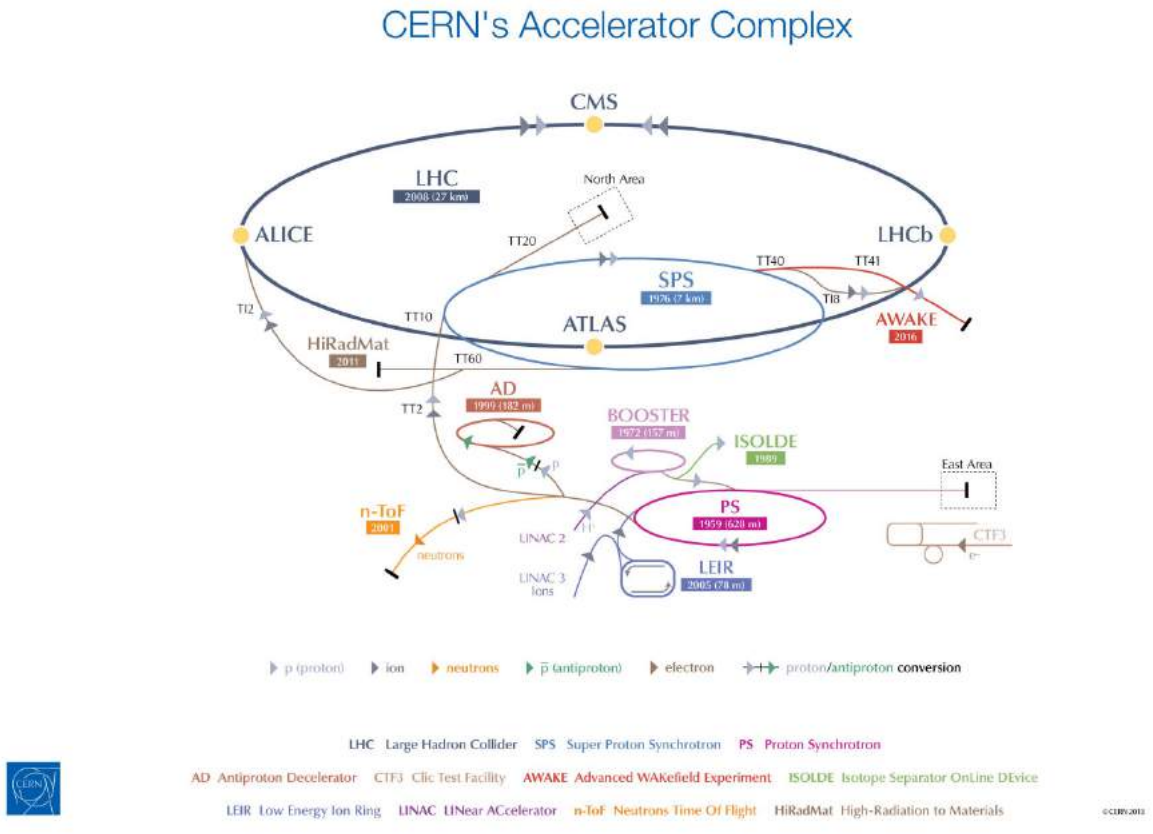


Figure 2.4: CERN accelerator complex.

2.2.2 Luminosity

Instantaneous luminosity in particle physics is the rate by which the protons collide. If two bunches containing N_1 and N_2 particles collide head-on with frequency f, the instantaneous luminosity is

$$\mathcal{L} = \frac{N_1 N_2 f}{4\pi\sigma_x\sigma_y} = \frac{N^2 f}{4\pi\sigma_x\sigma_y} \tag{2.1}$$

if the two bunches have equal size. The total number of events is then the cross section of interest, σ, times the time integral over the instantaneous luminosity ,$\mathcal{L}$

$$N_{total} = \sigma \times \int \mathcal{L}(t)dt \tag{2.2}$$

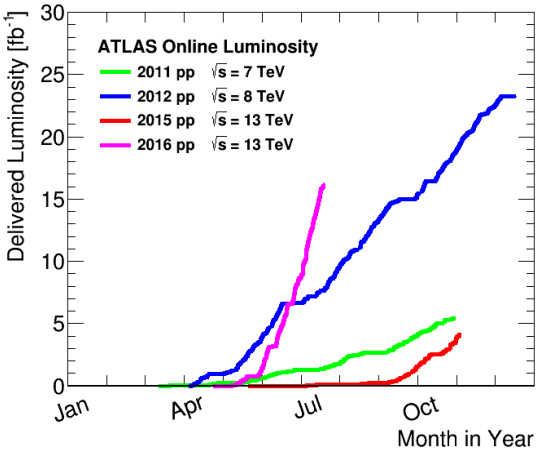


Figure 2.5: Instantaneous luminosity measurements [9].

It is important to have a large luminosity, since some of the processes have a very low cross section, see Figure **??**. E.g. the probability of a Higgs event occurring is about 4 orders of magnitude less than an event with a W boson occurring. Figure 2.5 shows the instantaneous luminosity during Run 1 and 2.

2.2.3 Pile-up

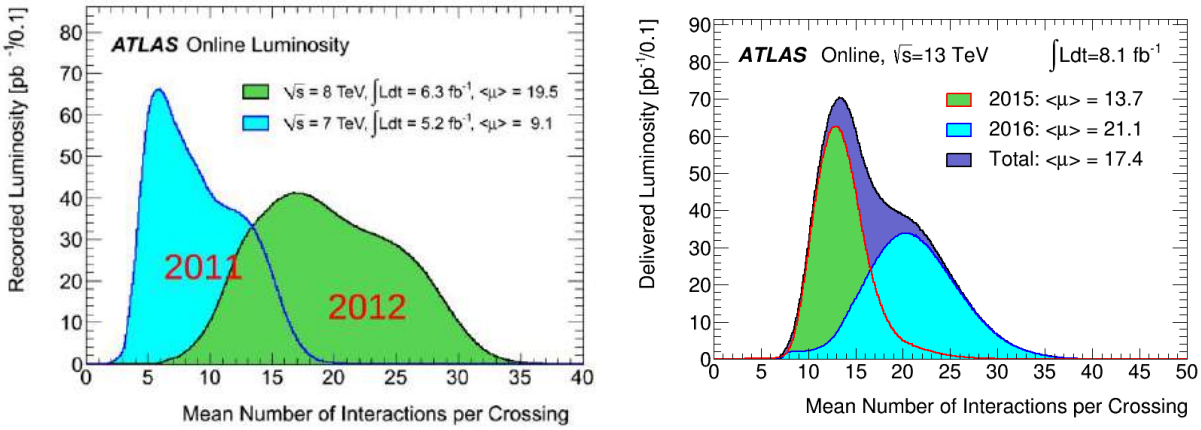
When the protons are colliding at the LHC, several collisions can happen at the same time. When several protons collide at the same time, the secondary particles from the events travel through the detector. To distinguish between the different events, we need to able to tell which events the particles come from. The pixel detector in the inner detector helps to distinguish this, as described in Chapter 2.

The average number of collisions μ is a sensitive observable variable, given by the out-of-time pile-up

$$\mu = \frac{L \times \sigma_{incl}}{N_{bunch} \times f_{LHC}} \tag{2.3}$$

where L is the measured instantaneous luminosity, σ_{incl} is the total inelastic proton-proton cross section, N_{bunch} is the number of bunches in the LHC, and f_{LHC} is the LHC revolution frequency [10].

In Figure 2.6 pile-up is shown for the run in 2011 and 2012. Run 2 at CERN is running right now, where the pile-up is currently around x, but this could change.



(a) Pile-up in Run 1 from 2011-2012 [11] .

(b) Pile-up in Run 2 from 2015-2016 [11].

Figure 2.6: Distribution of pile-up.

2.2.4 *Coordinate System at CERN*

The interaction point is where the protons collide, and it is defined as the origin of the coordinate system. The beam direction defines the z-axis and the x-axis contains the left and right coordinates with reference to the beam, and lastly the y-axis is up and down with reference to the beam. The azimuthal angle ϕ is determined in usual way around the beam axis, and the polar angle θ is the angle from the beam axis.

Pseudorapidity is defined as

$$\eta = -\ln\tan(\theta/2) \tag{2.4}$$

The particles will usually be analysed in the coordinate system $(\theta, \eta, z$, but massive objects such as jets will be measured according to the rapidity

$$y = 1/2\ln[(E + p_z)/(E - p_z)] \tag{2.5}$$

But for massless particles pseudorapidity is equal to rapidity. It is hard to measure the longitudinal part of momentum of the particles, but easy to measure the angle. The distance ΔR in the pseudorapidity-azimuthal angle space is defined as

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

The differences in pseudorapidity are invariant under boosts along the z-axis, and therefore the pseudorapidity is the desired choice of describing the position of a particle.

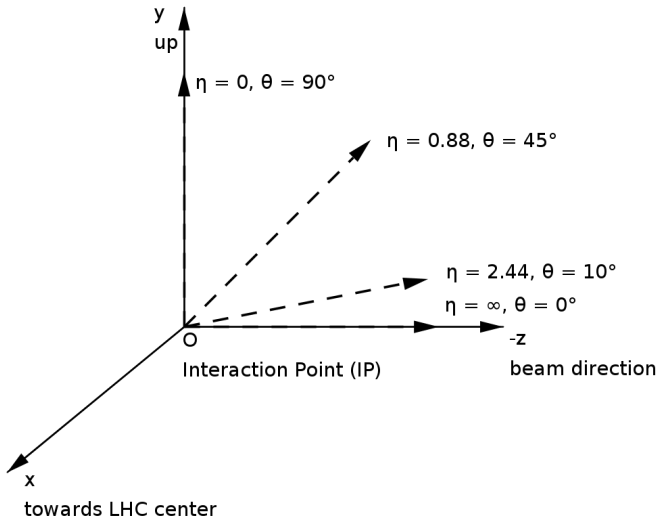


Figure 2.7: Coordinate system. The ATLAS detector is a cylinder centred around the interaction point.

2.3 The ATLAS detector

The ATLAS detector is the largest detector at the LHC at CERN. The ATLAS detector was build to study a wide variety of physics. It consist of three major detectors: The Inner Detector (ID), the calorimeters (ECAL and HCAL) and the Muon Spectrometer (MS). These detectors can detect charged particles, neutral particles and muons. The charged particles can be tracked in the inner detector. The neutral particles can not be tracked, but an energy deposit of neutral particles in the calorimeter can be measured. The muons are heavier than other charged particles, and therefore need a higher magnetic field to be bend to measure their track and momentum. This is carried out by the muon spectrometer.

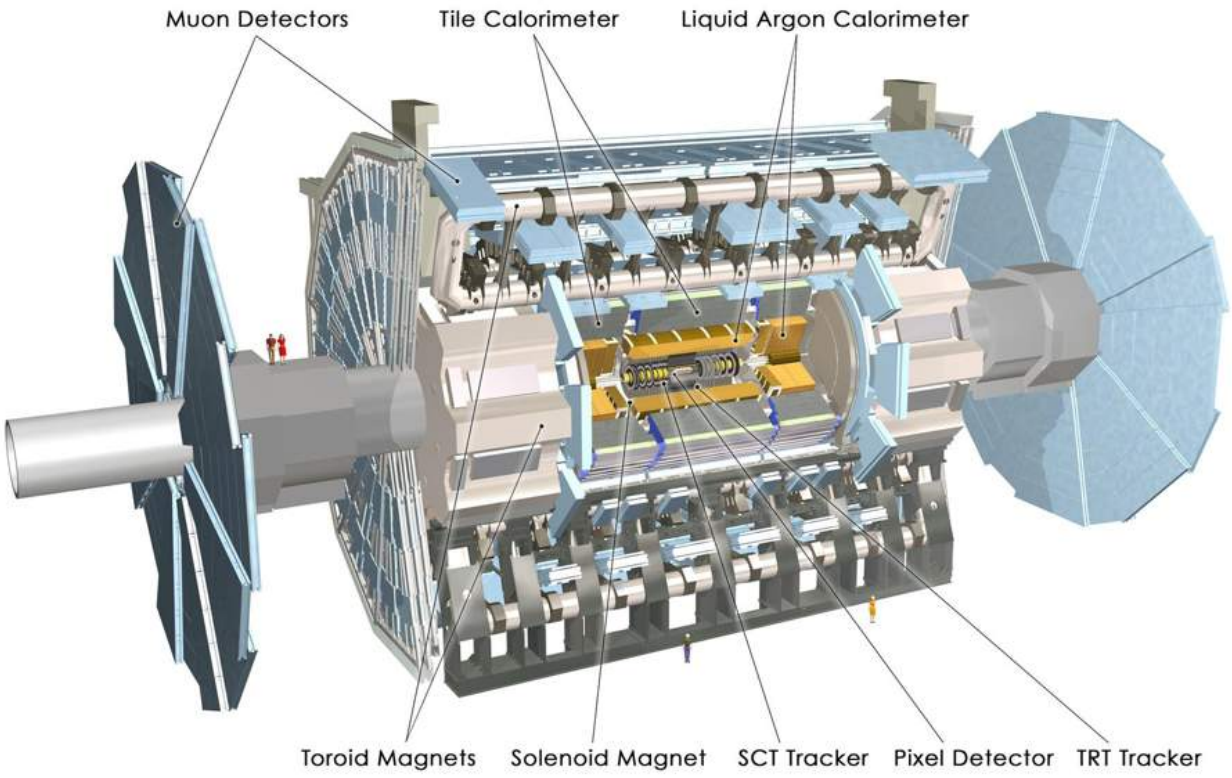


Figure 2.7: The ATLAS detector. The SCT Tracker, Pixel Detector and TRT Tracker are all elements of the inner detector. After the inner detector come the calorimeters, and lastly the muon spectrometer.

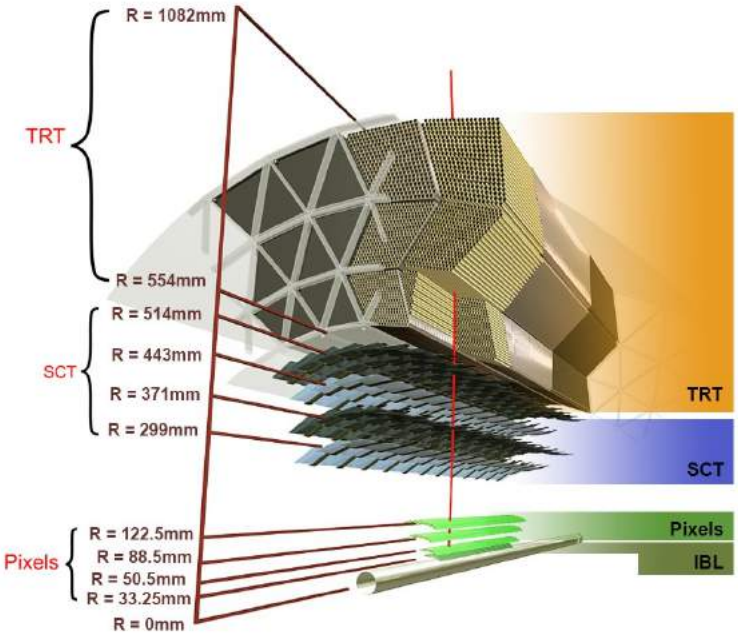
2.3.1 The Inner Detector

The inner detector at ATLAS is closest to the interaction point. The main purpose of the inner detector is to determine the tracks of charged particles, and measure their momentum. The ATLAS Inner Detector consists of three detectors: Pixel detector, SCT (Semiconductor Tracker) and TRT (Transition Radiation Tracker). To detect charged particles, the inner detector bends the particles with a magnetic field of 2 T. The Inner Detector is designed to measure the momentum of charged particles within the pseudorapidity range $|\eta| < 2.5$.

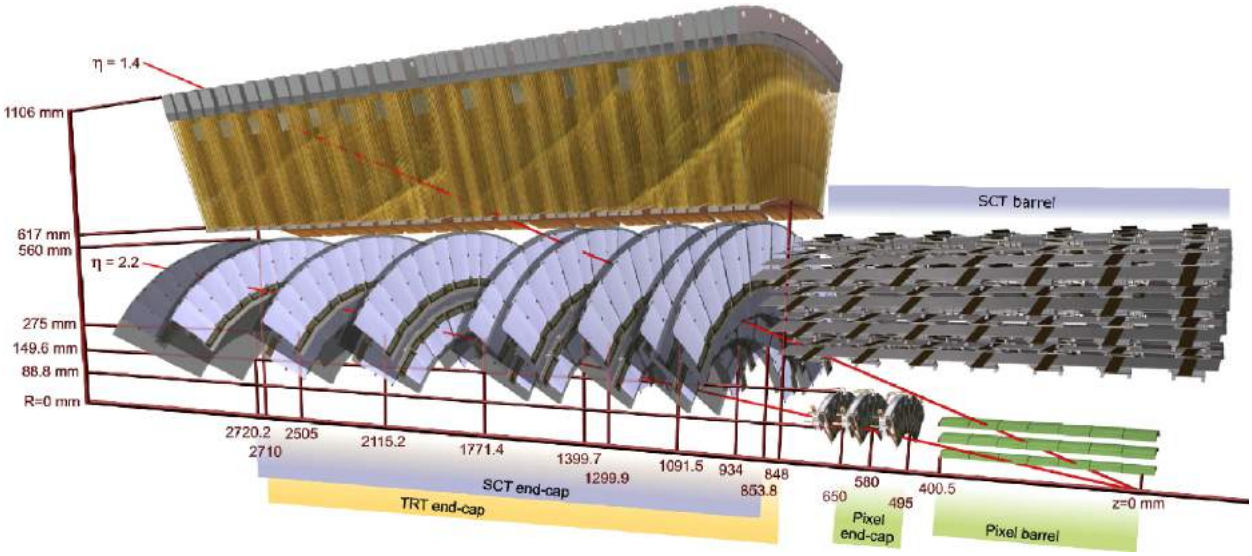
The inner detector is divided into barrel and end-caps, both are shown in Figure 2.8. The barrel region is formed like a cylinder and has several detector layers. The barrel region has a length of $\pm$ 3512 mm and a radius of 1150 mm. The end-caps are disks in the transverse plane.

The Pixel and the SCT are semiconductor detectors. This means that when a charged particle passes through, it will create electron-hole pairs in the material. Then the electrons can be collected by an external electric field.

Between run 1 and run 2 a series of upgrades were made to the inner detector. This included adding a new fourth layer, the insertable B-layer (IBL).



(a) The ATLAS inner detector, including the new insertable B-layer (IBL) [12].



(b) The end caps of the ATLAS inner detector [13].

Figure 2.8: Views of the ATLAS inner detector.

Detector Component	Required resolution	$\lvert\eta\rvert$ coverage	
		Measurement	Trigger
Tracking	$\sigma_{p_T}/p_T = 0.05\% p_t \oplus 0.1\%$	± 2.5	
EM calorimetry	$\sigma_E/E = 10\%/\sqrt{E} \oplus 0.7\%$	± 3.2	± 2.5
Hadronic calorimetry barrel and end-cap forward	$\sigma_E/E = 50\%/\sqrt{E} \oplus 3\%$ $\sigma_E/E = 100\%/\sqrt{E} \oplus 10\%$	± 3.2 $3.1 < \lvert\eta\rvert < 4.9$	± 3.2 $3.1 < \lvert\eta\rvert < 4.9$
Muon spectrometer	$\sigma_{p_T}/p_T = 10\%$ at $p_T = 1$ TeV	± 2.7	± 2.4

Figure 2.8: Resolution and η coverage of the different detector components at ATLAS [13].

The pixel and SCT detector can measure the bending of particles in the inner detector and the measurements are also used for determining the primary vertex in the event and the momentum of charged particles. The pixel detector consist of 1744 sensors with 47232 silicon pixels each [13]. The sensors are 250 μm thick and the dimensions are 19×63 mm^2.

When a particle hits a layer of silicon that makes up a pixel it knocks several electrons out of their place in the silicon. The electrons are then attracted to one direction by an electric field. The charge collected by each pixel is proportional to the number of particles that hit it.

The SCT detector consists of 4088 modules. The modules cover a surface of 60 m^2 of silicon [13]. The sensors has a thickness of 285 ± 15 μm. Each layer of the SCT consist of two strips which form a small angle between them. With this, the detector can provide both z and ϕ measurements of a particle. As the angle between SCT modules is small, the resolution is better in the ϕ-direction.

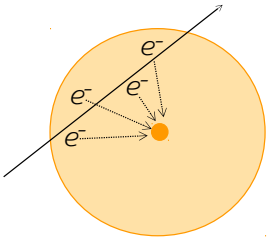


Figure 2.9: Radiation in the TRT. A particle passes through the TRT tube. The particle with a longer path through the tube will produce more electrons. The electrons then drift to the anode which is in the centre of the tube.

The TRT detector consists of gas filled straws which will be ionized by the particles passing through. The TRT straws are 4 mm in diameter. Each straw has an anode wire in the centre. When a particle ionizes the gas in a straw, the electrons will drift towards the anode, Figure 2.9. The charge absorbed at the anode is the signal from the TRT detector. The number of TRT readout channels is approximately 361000.

TRT does not provide a good resolution of the z component of particles. For a good resolution of the tracks of particles, all the information obtained in the inner detector is combined.

The TRT is important for electron identification. The TRT has different material between the tubes, which means that if an electron moving from one material to another with different dielectric properties, the electron will radiate a photon. This photon is seen as a

high threshold peak in the detector, and will have more energy than the ionization electrons. The energy of the radiation is dependent on the γ-factor of the particle. An electron has a γ-factor of around 200 times more than a pion with the same momentum, which is why the γ radiation and electron production is distinguishable in the readout of the detector. This is distinguishable by looking at the high and low threshold of the TRT. If the intensity if above the high threshold an electron is identified, and if below, another particle passed through. The TRT is therefore very good at identifying electrons. The signal received from a TRT straw is shown in Figure 2.10.

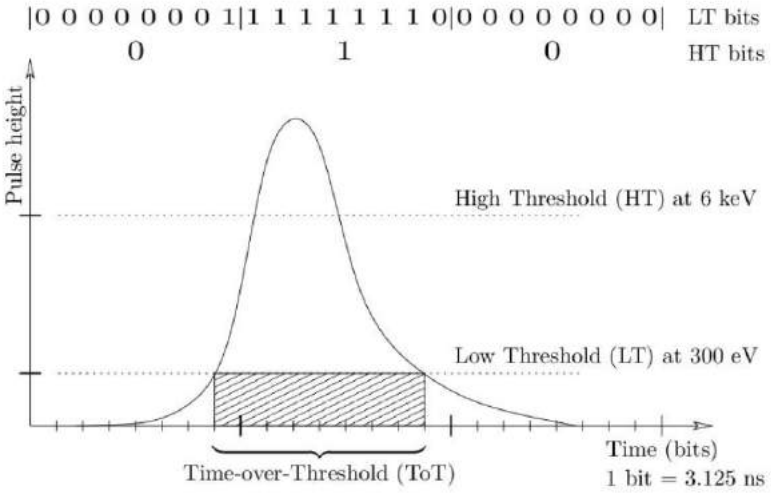


Figure 2.10: The high- and low-threshold peaks from a TRT signal. If the signal is above the high threshold, the particle is identified as an electron. If the signal is below the high threshold and above the low threshold it is another particle.

2.3.2 *Calorimeter*

Calorimeters can measure the energy of electrons, photons and hadrons. The only particle the calorimeters can not measure are neutrinos and it can not fully measure muons. At the ATLAS detector there is both an electromagnetic and hadronic calorimeter. The design of the calorimeter is shown in Figure 2.12. The calorimeter must provide a good measurement of energy of particles and hadron showers. The calorimeter should also limit the number of particles which are not muons passing through to the muon system. With the large η coverage and thickness of the calorimeter system, it ensures a good measurement of missing transverse energy in an event.

The electromagnetic calorimeter is divided into4 regions: presampler, 1^{st} sampling, 2^{nd} sampling and 3^{rd} sampling. The presampler is a thin layer of argon and will correct for the energy loss in the other materials (mostly the solonoid magnet) the particles passed through. The geometry of the 1^{st}, 2^{nd} and 3^{rd} sampling is in Figure 2.11. The first sampler has a good resolution in η, which is used to account for photon π^0 separation[1]. Most of the energy of the particles is deposited in the 2nd layer, and only the high energy particles will

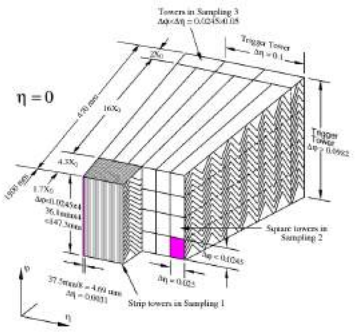


Figure 2.11: Sketch of the EM calorimeter system [14]. The three layers of the EM module is shown.

[1] Since a normal decay is $\pi^0 \to \gamma\gamma$, you need to be able to tell where the photons come from, since they might be from a different process, eg. $H \to \gamma\gamma$.

deposit energy in the 3^{rd} layer. The 1st layer has a radiation length of 4.3, the 2^{nd} of 16 and the 3^{rd} of 2.

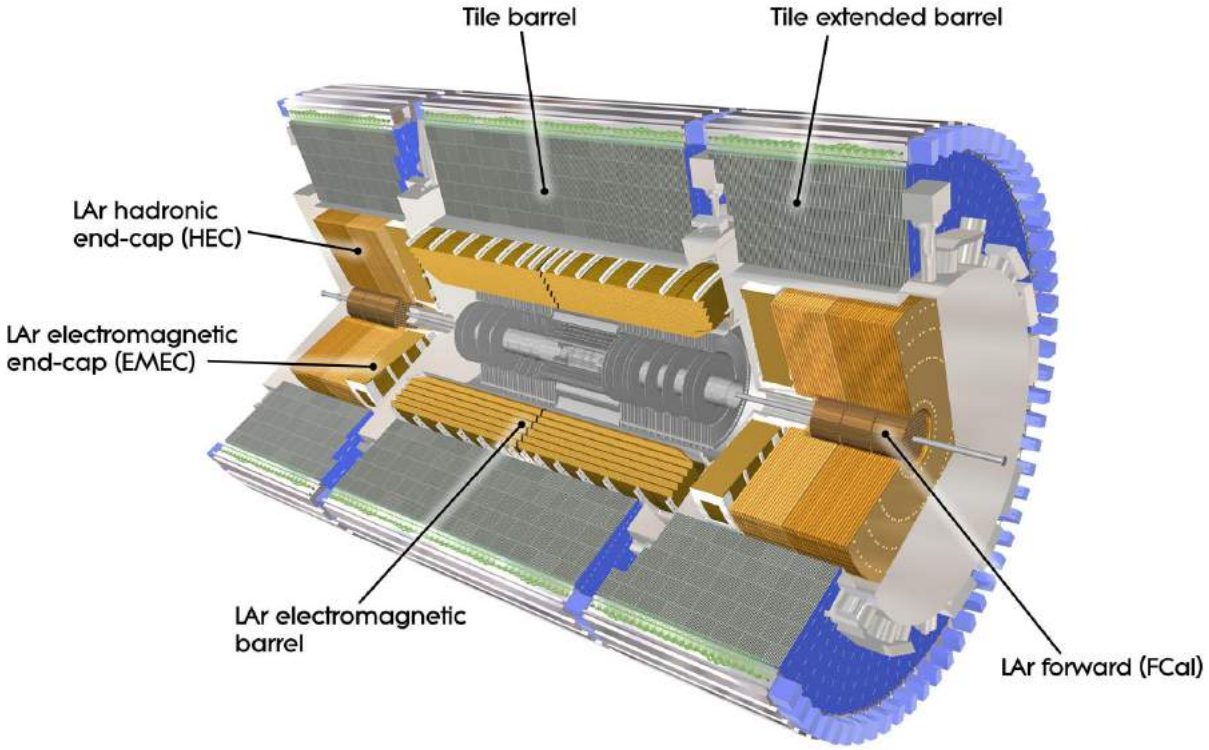


Figure 2.12: View of the ATLAS calorimeter system [13].

The electromagnetic calorimeter is made of lead and liquid argon. It is divided into a central barrel and end-cap regions on either side of the detector. The lead gives the shower development with its short radiation length of 0.56 cm and the secondary electrons create ionisation in the narrow gaps of liquid argon.

A relativistic electron looses energy to Bremsstrahlung in proportion to its current energy,

$$\frac{dE}{dx} = -\frac{1}{L_R}E \tag{2.7}$$

where L_R is the radiation length of the traversed material. L_R is 0.56 cm for Pb, 1.8 cm for Fe, 8.9 cm for Al etc. Thus Pb is chosen for the electromagnetic calorimeter at ATLAS, since it has the shortest radiation length. Energy carried away by a Bremsstrahlung photon is converted into a e^+e^- pair after a mean flight distance of $\frac{9}{7}L_R$. This effect leads to an electromagnetic shower where an incoming electron or photon cascades into more and more electrons, positrons and photons.

The hadronic calorimeter is needed to detect particles which can not be fully contained in the electromagnetic calorimeter. Hadrons

will usually not deposit all their energy in the EM, but perhaps a minor ionizing signal. Therefore the hadronic calorimeter is needed with more material behind the EM.

2.3.3 *Muon Spectrometer*

The muon spectrometer is the outermost part of the ATLAS detector. The muon system in shown in Figure 2.12. The muon system is based on the magnetic deflection of muon tracks in the large superconducting toroid magnets. The toroid magnets provide a magnetic field of $1.5 - 5.5$ T. The muon spectrometer is large enough to measure the bending of the muons. The detector covers $|\eta| < 2.7$ and can identify muons with a momentum above 3 GeV and provide precise measurements of momentum up to 1 TeV.

The MS is comprised of both trigger and tracking detectors. The MS is constructed by three separate cylindrical stations, with a distance of about 2 m between them (radius of 5 m, 7.5 m and 10 m). The end-caps have their own magnetic field, and consist of two disks on either side of the interaction point.

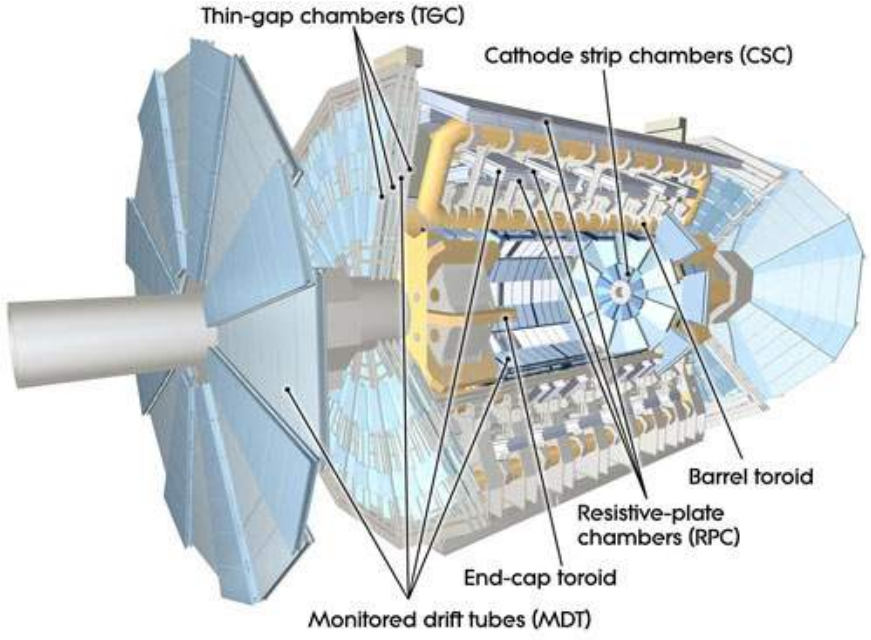


Figure 2.12: View of the ATLAS muon system. [13].

The detection of muons is mainly in the Monitored Drift Tubes (MDT), see Figure 2.13. The MDT is made of 2×3 drift tubes. These drift tubes are made by aluminium and have a diameter of 6 cm and a length in range of 0.9 to 6.3 m. The MDT covers $|\eta| < 2.7$

In the barrel region, Resistive Plate Chambers (RPC) are used. The RPC stations consist of two units each with two layers each.

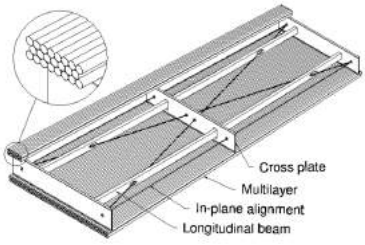


Figure 2.13: Scheme of a Monitored Drift Tube chamber.

2.3.4 Trigger System and Data Acquisition

The trigger system will determine which events that should be saved. Many event occur with e.g. no particular high energy jet or lepton, and these kind of events are therefore not especially wanted for physics analysis. The trigger system is divided into the Level 1 trigger (L1) and the High Level Trigger (HLT). L1 is hardware and HLT is software.

The Level 1 (L1) muon trigger looks for large transverse momentum muons, by reconstructing tracks that point to the interaction point. They are designed to provide a fast but spatially coarse trigger response for the muons. Two layers of RPC chambers are installed in the middle station, and a third layer is installed on the outer chamber station. Thin Gap Chambers (TGC) are installed at the end-caps and are also a part of the L1 trigger. Tests have provided that the L1 trigger has a trigger efficiency of 99.6 %. The L1 trigger also gets a signal from the electromagnetic calorimeter.

The HLT receives all information from an event, and then makes the decision to keep or throw an event, depending on the energies of the entities of the event.

In Run 1 the L1 trigger would trigger on events with more than 80 GeV of missing transverse energy. In Run 2 the pile-up is increased, and this effects the reconstruction of missing transverse energy. Pile-up will make the estimation at trigger level of missing transverse energy higher, than it actually is. When both triggers agree on a event, the event is recorded. The definition of an event is therefore every time a bunch has activated a trigger. The events are recorded and sorted by the following,

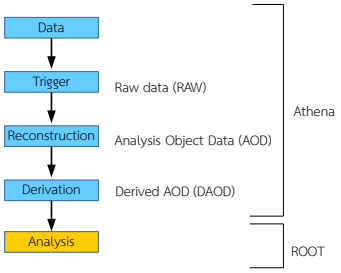


Figure 2.14: Process from data to analysis.

- Lumninosity (about 1 min recording)
- Run (never reset)
- sub-period (group of runs which are similar)
- period (selection of similar periods)
- LHC Run

When the data is recorded, it also contains information of whether it is bad data. For example one of the detectors could not be recording, but if a detector is not crucial for a specific analysis, the data can be still used.

When analysing the obtained data, we work through several steps. After the trigger the objects of the event needs to be reconstructed.

When all objects are reconstructed, you can make a selection of events you may find interesting. For example if one is interested in $WW \rightarrow lvqq$ events, you can check if the reconstruction objects contains two high p_T jets and one high p_T lepton. When the events are selected, one can carry out the main analysis. In this paper I am mostly working on the reconstruction and derivation levels, since I will be attempting to reconstruct the missing transverse energy of an event, but not really using this for a specific analysis.

A summary of Run 1 and Run 2 is shown in Table 2.1.

Period: Year	Bunch spacing	$\sqrt{s}$	Peak luminosity	Peak number of collisions per bunch
Run-1: 2012	50 ns	8 TeV	8×10^{33} cm^{-2} s^{-1}	40 (at 8×10^{33} cm^{-2} s^{-1})
Run-2: 2015-2018	25 ns	13 TeV	$1 - 2 \times 10^{34}$ cm^{-2} s^{-1}	25-50 at ($1 - 2 \times 10^{34}$ cm^{-2} s^{-1})

Table 2.1: Summary of Run-1 and Run-2 [15]

2.4 Reconstruction of Particles

There are advanced and well tested algorithms for reconstruction of particles at the ATLAS detector. The tracks of all charged particles are reconstructed and traced back to a vertex in the beam line. These tracks can be from e.g. muons, electrons or jets. The neutrinos can not be reconstructed by tracks, and instead a calculation of missing transverse energy is used. The particle flow algorithm is an alternative way of reconstructing all particles, and this is the reconstruction used in this paper.

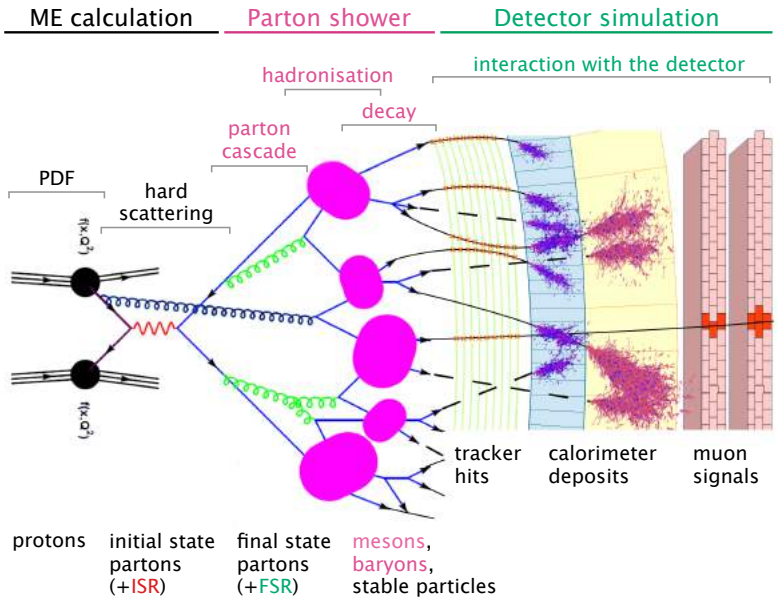


Figure 2.15: Schematic of an event in a detector [16]. When two protons collide, the quarks can pair up into new particles. These particles decay and their final states can be detected. When the charged particles travel trough the inner detector, they leave tracks which can be reconstructed. In the calorimeters the particles deposit energy. The muon spectrometer detects the high energy muons. Note that this is not a schematic of ATLAS, but a general schematic.

2.4.1 Track and Vertex Reconstruction

Tracks are reconstructed from signals in the inner detector. This is done by matching three-dimensional hits in each sub-detector. The tracks represent a charged particle bending in the magnetic field of the inner detector. There are several algorithms to reconstruct the tracks of particles [17]. When the tracks are reconstructed, the vertexes can be determined. The path of the particles will be traced back to the beam line, and the origin of the particle can be determined. The tracks are not entirely smooth, and will not all be tracked back to one specific point in the beam line. Thus an iterative Chi^2 fit is made using the seed and nearby tracks. If tracks are displaced more than 7σ of a vertex, the tracks will seed a new vertex. All vertices are required to contain at least two tracks, and the vertex with the highest sum of p_T is the primary vertex, and all other vertices are named secondary vertices or pile-up vertices. The efficiency of vertex reconstruction was found to be decreasing as a function of pile-up, Figure 2.16, in the paper [17].

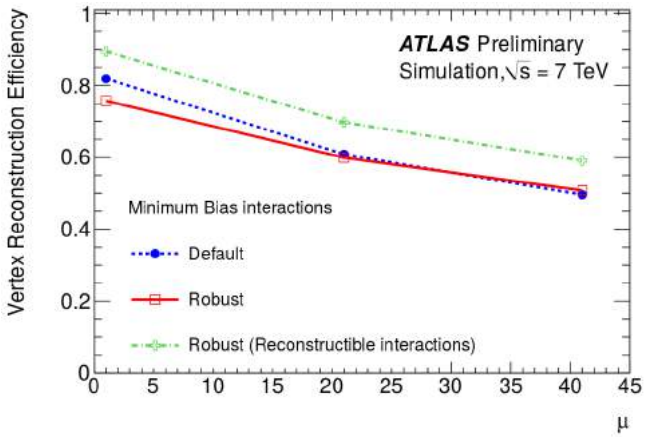


Figure 2.16: Vertex reconstruction efficiency as a function of pile-up [17].

2.4.2 Electrons

Electrons are reconstructed from clusters in the electromagnetic calorimeter (EM) [18], and are required to have a track in the inner detector (ID) and must be associated with the primary vertex. The electrons are summarized at three different levels: loose, medium and tight.

When using the *loose cuts* for electrons, the reconstruction of electrons relies only on a limited information from the calorimeters. Here cuts are applied to shower-shapes, derived from only the middle layer of the electromagnetic calorimeter. The problem with these cuts is that this has a low background rejection, thus it is sometimes better to use medium or tight electrons.

The *medium cuts* is an improvement by adding cuts on the strips of the first layer of the EM calorimeter and on the tracking variables. The cuts on the strips are effective for rejecting the $\pi^0 \to \gamma\gamma$ decays, see [13] for details of this cut. The tracking variables include cuts in the number of hits in the pixel and SCT detector. Using the medium cut, the jet rejection is increased by a factor of 3-4 with respect to the loose cuts, but the medium cuts reduce the identification efficiency by about 10 %.

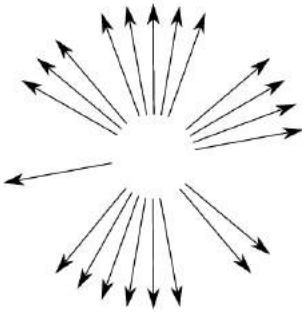
The *tight cuts* make use of all the electron identification tools currently available. Some of the tight cuts include rejecting more electrons from conversions, number of hits in the TRT and on the ratio of high-threshold hits in the TRT. These cuts result in the highest isolated electron identification and the highest jet rejection.

2.4.3 Muons

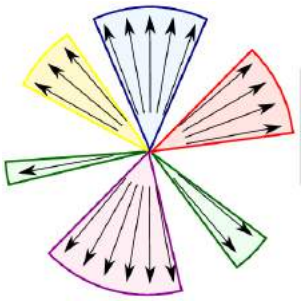
Muons are reconstructed using both the inner detector and the muon system, and must also be a part of the primary event. Just as the electrons, the muons can be reconstructed in different ways which are: standalone muons, inner detector muons, combined muons, tagged muons and merging muons. The standalone muons is the simplest

way of reconstructing muons. This is where a track is found in the muon spectrometer and then extrapolated back to the beam line. *Inner detector muons* are formed using the primary track reconstruction described in [19]. *Combined muons* are reconstructed by matching the standalone muons to the inner detector. *Tagged muons* propagate all inner detector tracks with a high enough momentum out to the first station of the muon spectrometer. If sufficient nearby segments are found by the muon spectrometer, a tagged muon is found. *Merged muons* includes using several of the muon reconstruction algorithms.

2.4.4 Jets



(a) Input: List of particles.



(b) Output: Groups of particles → *jets*.

Figure 2.17: Jet algorithm.

A high-energy quark, anti-quark or gluon can be transformed into a spray of hadrons, i.e. jets. Jets are produced in the QCD hard scattering process. The most used reconstruction algorithm for jets is the $anti - k_T$ algorithm.

The variables used for a jet reconstruction algorithm are [20],

$$d_{ij} = \min(k_{T,i}^{2p}, k_{T,j}^{2p}) \frac{\Delta_{ij}^2}{R^2} \tag{2.8}$$

$$d_{iB} = k_{T,i}^{2p} \tag{2.9}$$

where d_{ij} is the distances between entities, such as particles and pseudojets, and d_{iB} is the distance between an entity, i, and the beam, B. $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ and $k_{T,i}$, y_i and ϕ_i are respectively the transverse momentum, rapidity and azimuth of particle i. p is a parameter to govern the relative power of the energy versus geometrical scales, and R is the usual radius. With $p = 1$ the algorithm is the *inclusive* k_T algorithm and with $p = -1$ it corresponds to the anti-k_T algorithm. The algorithms cluster particles into jets according to the minimum distances between the particles and the momentum of the particles. A schematic of the inputs and outputs of a jet clustering algorithm is in Figure 2.4.4.

2.4.5 TopoClusters

The signals from the calorimeters are collected into variable sized clusters, named Topo-clusters. Topo-clusters are formed by an algorithm starting from a calorimeter cell with a high seed signal. The seeding, growth and boundary of a topo-cluster is dependent on several variables with different thresholds [21] . The topological clustering algorithm groups neighbouring clusters with an energy above a given threshold, to the seed signal. The topological clustering algorithm tries to minimize the noise in the surrounding cells. The effect of noise reduction is largest for smaller clusters.

The Topological Cluster algorithm starts from the calorimeter cells, of which the ATLAS calorimeters have about 100000, and identifies

any cells with an energy significantly above the expected energy due to electronic noise. These cells will provide as a seed for the topological cluster. Any other cells with an energy greater than some secondary threshold are then added to the cluster. This process continues until there are no further cells adjacent to any cells already added to a cluster, that has sufficient energy to pass the secondary threshold. The energy thresholds are defined in terms of the expected amount of electronic noise

$$\Gamma = |\frac{E_{cell}}{\sigma_{noise,cell}}| \tag{2.10}$$

The thresholds used at ATLAS are 4 for the seed, 2 for the secondary threshold and 0 for the tertiary threshold [22]. The energy here is measured at the electromagnetic scale. The ATLAS calorimeters are non-compensating, meaning that the measured energy in the calorimeter does not actually correspond to the energy of a particle. This is because of the secondary particles that can not be measured in the calorimeter, such as non-interacting neutrinos. Therefore a calibration must be made after the topocluster algorithm to correct for this effect. This is normally corrected for in the jet algorithms which make use of the topoclusters.

2.4.6 Missing Transverse Energy

Neutrinos can not be directly detected in the ATLAS detector. Therefore the energy of the neutrinos passing through the detector is not deposited anywhere. But the neutrinos can be detected indirectly by measuring missing transverse energy. By conservation of energy, the total energy of the system must be zero, since the protons collide from opposite directions.

The $\vec{E_T^{miss}}$ is reconstructed from the negative vector sum of the transverse momenta of all detected particles.

Missing transverse energy at ATLAS is a measurement with two constituents: A soft term and a hard term. The hard represents the fully reconstructed objects, such as jets and leptons, and the soft term is reconstructed from the soft event signal contribution. The E_T^{miss} is calculated using the components along the x and y axis [23],

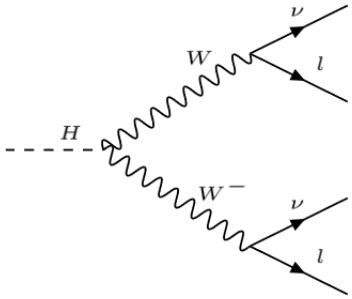


Figure 2.18: Higgs decay to WW where the W decay into a lepton and a neutrino.

$$E_{x(y)}^{miss} = E_{x(y)}^{miss,e} + E_{x(y)}^{miss,\gamma} + E_{x(y)}^{miss,\tau} + E_{x(y)}^{miss,jets} + E_{x(y)}^{miss,\mu} + E_{x(y)}^{miss,SoftTerm} \tag{2.11}$$

where the terms are calculated by the negative vectorial sum of the momentum the reconstructed objects.

The E_T^{miss} and the azimuthal angle are calculated as

$$E_T^{miss} = \sqrt{(E_x^{miss})^2 + (E_y^{miss})^2} \tag{2.12}$$

$$\phi^{miss} = \arctan(E_y^{miss}/E_x^{miss}) \tag{2.13}$$

The variable sum E_T is a scalar sum of the transverse momentum, and it is a measure for the *total activity* of a event.

$$\sum E_T = \sum E_T^{HardTerm} + \sum E_T^{SoftTerm} = \sum_{HardObjects} p_T + \sum_{SoftSignals} p_T \tag{2.14}$$

Note that the components in the hard term is already cleaned for pile-up in their reconstruction algorithms, but the soft term is not as such corrected for pile-up contributions [10]. The soft term consists of energy from tracks and calorimeter cell cluster signals which are not associated with a hard object, and it is therefore difficult to clean this mixed object.

2.4.7 Methods for removal of pile-up in missing transverse energy

The hard signal contribution to missing transverse energy consists of elements such as fully reconstructed and calibrated photons, leptons and jets. But the soft term contributing to the missing transverse energy is not easily corrected for pile-up. The challenge in correcting events for pile-up is developing suppression methods for the soft-event contribution.

The *Soft-Term Vertex-Fraction* uses the scalar sum of the soft terms, $\sum E_T^{SoftTerm}$ associated with the hard scatter vertex, and combines this with the sum of all soft-event track-p_T in the event, $E^{miss}_{T,TrackSoftTerm}$. These variables are used as a ratio, and this ratio is then used to scale all soft-event contributions in a event. The ratio is given by,

$$STVF = \frac{\sum_{i=1}^{N_{track}(V_{HS})} p_{T,i}^{track}(V_{HS})}{\sum_{k=1}^{N_{PV}} \sum^{N_{track}(V_k)} p_{T,i}^{track}(V_k)} \tag{2.15}$$

where $p_{T,i}^{track}(V_k)$ is the p_T of the soft-event track i coming from vertex V_k, and $N_{track}(V_k)$ is the total number of reconstructed tracks not associated with any hard object from this vertex. A track is associated with a hard object if it is the source of a reconstructed particle, used for particle identification, or generally overlaps with calorimeter signals representing the particle. The correction for pile-up is then made by scaling the different components of missing transverse energy by $STVF$,

$$E^{miss,SoftTerm}_{x(y),corrected} = STVF \cdot E^{miss,SoftTerm}_{x(y)} \tag{2.16}$$

$$E^{miss,SoftTerm}_{T,corrected} = STVF \cdot E^{miss,SoftTerm}_{T} \tag{2.17}$$

$$\sum E^{SoftTerm}_{T,corrected} = STVF \cdot \sum E^{SoftTerm}_{T} \tag{2.18}$$

The *Jet Vertex-Fraction* is used to filter jets contributing to the hard term in the missing transverse energy E_T^{miss}, and the scalar sum of transverse energies $\sum E_T$. This method also uses a ratio, as the STVF

method, but it is restricted to tracks associated with a jet.

$$JVF = \frac{\sum_{i=1}^{N_{track}^{jet}(V_{HS})} p_{T,i}^{jet}(V_{HS})}{\sum_{k=1}^{N_{PV}} \sum^{N_{jet}(V_k)} p_{T,track,i}^{jet}(V_k)} \tag{2.19}$$

where $N_{track}^{jet}(V_k)$ is the number of tracks from vertex V_k associated with the jet, and $p_{T,track,i}^{jet}(V_k)$ is the p_T of the track i associated with the vertex V_k and associated with the jet. Each jet will then have an assigned ratio JVF. This can only be calculated for jets within the inner detector acceptance, ie. $|\eta| < 2.4$. The ratio JVF is,

$$JVF = \begin{cases} -1 \text{ if no tracks associated with jet} \\ 0..1 \text{ if central jet with tracks} \end{cases} \tag{2.20}$$

Jets within the inner detector acceptance and with 20 GeV $< p_T <$ 50 GeV are only accepted if the corresponding JVF ratio is not zero. If $JVF = 0$ then the jet will be completely excluded from the E_T^{miss} calculations. Other jets with $|\eta| > 2.4$ and $p_T > 50$ GeV will always contribute to the E_T^{miss} calculations.

The *jet-area-based pile-up suppression* works by firstly clustering all particles in an event into jets. These jets can have a momentum of zero, describing a void, or higher to describe a collection of particles. The areas of these jets are determined using so called *ghost particles*, which is basically a monte carlo method, this is described in [10].

The soft-term contribution of E_T^{miss} from the jet-area based filter is,

$$E_{x(y)}^{miss,softterm} = -\sum_{i=1}^{N_{filterjet}} p_{x(y),i}^{jet} \tag{2.21}$$

and the following filter applied to the momentum of the jet is,

$$p_{T,i}^{jet} = \begin{cases} 0 \text{ if } p_{T,i}^{filterjet} < \rho_{evt}^{med}(\eta_i^{filterjet}) \cdot A_i^{filterjet} \\ p_{T,i}^{filterjet} - \rho_{event}^{median}(\eta_i^{filterjet}) \cdot A_i^{filterjet} \text{ if } p_{T,i}^{filterjet} \geq \rho_{evt}^{med}(\eta_i^{filterjet}) \cdot A_i^{filterjet} \end{cases} \tag{2.22}$$

where the median transverse momentum density ρ_{event}^{median} is determined from ρ-jets. The reconstruction of ρ-jets in described in detail in [10].

The *Extrapolated Jet Area* (EJA) method uses the momenta-density using the soft events from only the central part of ATLAS, ie. $|\eta| < 2$. This central density is then extrapolated to the forward region.

Extrapolated Jet Area with Filter (EJAF) uses the same technique as EJA, but this method also includes a JVF-based selection applied to the filter-jets.

The *Jet Area Filtered* (JAF) method uses the momenta-density calculated from soft events in the region $|\eta| < 4.9$. It also applies JVF-based selection on the filter jets.

2.4.8 Particle Flow Objects (PFOs)

Particle Flow objects are objects where the topoclusters and ID track information is combined. Using the track from ID the algorithm then searches for the nearest topocluster. The particle flow objects records energy and charge of particles.

The particle flow approach aims to improve the jet resolution by tracing the paths of individual particles throughout the detector, collecting together the energy deposits left in each subdetector system.

At CMS the usage of the particle flow algorithm improved the resolution of the measured particle energies. This was basically done by replacing the energy deposited by electrically charged particles measured in the calorimeter detectors with the momentum measured in the charged particle tracking detectors. This method makes use of the fact that the resolution in the charged particle tracking systems is better than the energy measurements in the calorimeter at low energies. At ATLAS the calorimeter has a superior resolution compared to the tracking system at energies above 150 GeV [24], see Figure 2.19

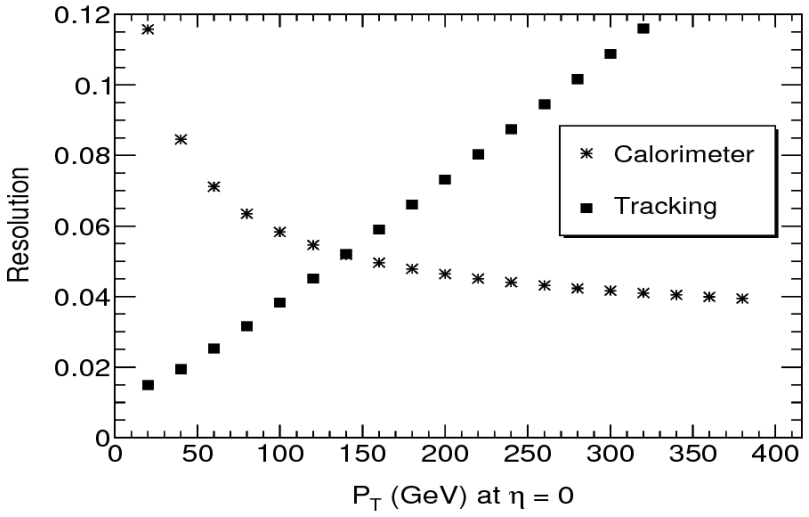


Figure 2.19: Resolution of Single Pions at $\eta = 0$ in the calorimeter and charged particle tracking system [24].

Particle Flow is an attempt to reconstruct all stable particles in an event, ie. photons, charged and neutral hadrons, electrons and muons. The particle flow algorithm tries to avoid double counting of energy from the same particle, and also tries to separate energy deposits from different particles. The reasons for using the particle flow algorithm are the following:

- The tracker has better energy resolution than the calorimeter at low p_T.
- In a jet soft charged particles are swept away by the field, but can be recovered by the tracker
- The tracker can pick up charged particles that would be below calorimeter noise threshold

- The tracker has better η, ϕ resolution
- The tracker can tell which vertex charged particles comes from → Charged Hadron Subtraction (CHS).

The particle flow algorithm aims to make use of all detector information to optimize the measurement of particles. An example of the input and output of the particle flow algorithm is shown in Figure 2.20.

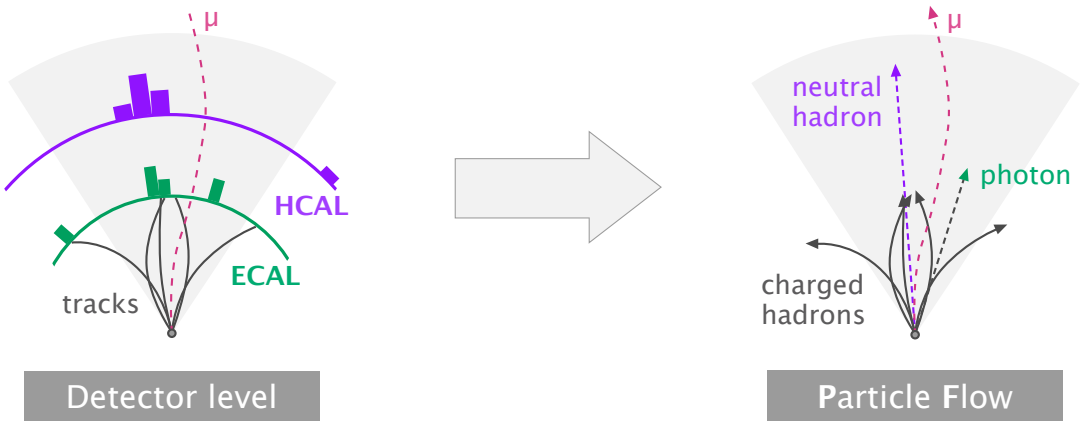


Figure 2.20: Event showing the energy deposits and tracks. The Particle Flow Algortihm reconstructs the objects using both tracker and cluster information. [25]

Particle flow objects are a representation of particles divided into two groups: *charged* and *neutral*. The charged particle flow objects have information from the Inner Detector, ie. track information. The neutral particle flow objects only have information from the hits in the calorimeter. The output of the particle flow algorithm is a single list of reconstructed particles. This list can be used for higher level algorithms, such as jet clustering, missing transverse energy estimations and particle isolation. The main data members of a particle flow object are

- Particle type
- 4-momentum
- Energy deposit in ECAL and in HCAL
- Various flags, which may be useful in other analyses

The Particle Flow algorithm in ATLAS is named eflowRec. Tracks with $|\eta| < 2.5$ considered, and it uses all topological clusters. The idea of this particle flow algorithm is to measure little energy in the calorimeter, and as much energy in the trackers as possible.

The eflowRec algorithm takes as input all topological clusters and reconstructed tracks to produce a list of eflowObjects, in this paper usually referred to as particle flow objects (PFOs). The PFOs can then correspond to an isolated track, an isolated topological cluster,

a modified topocluster that has had the charged energy deposit subtracted, or a track that was matched to the topocluster. The algorithm works by firstly finding a track close to a topocluster, ie. finding the minimum value of

$$\sqrt{\frac{(\eta_{track} - \eta_{clus})^2}{\sigma_\eta^2} + \frac{(\phi_{track} - \phi_{clus})^2}{\sigma_\phi^2}} \tag{2.23}$$

where η_{track} and ϕ_{track} are the extrapolated track coordinates at the second layer of the electromagnetic calorimeter and η_{clus} and ϕ_{clus} are the topocluster coordinates in the calorimeter. σ_η and σ_ϕ are the usual standard deviations. The cluster nearest to the track is required to pass the condition

$$E_{clus} > E_{expected} - k_2 \times \sigma_{expected} \tag{2.24}$$

where E_{clus} is the energy of the cluster at the electromagnetic scale, $E_{expected}$ is the expected energy deposit at the electromagnetic scale from a charged pion with the energy and η the track is measured to have, $\sigma_{expected}$ is the width of $E_{expected}$ and k_2 is a free parameter. This condition is required to ensure the cluster does not have significantly less energy than what the charged particle is expected to deposit.

If the condition is satisfied then the expected energy deposit is removed from the cluster. If the condition is not met, we use the cluster energy and not the track energy. After this, another condition is tested,

$$E_{clus} < k_1 \times \sigma_{expected} \tag{2.25}$$

where k_1 is a free parameter. If the condition is true, then the cluster is discarded. Otherwise the cluster is kept.

Wavelets and Information Theory

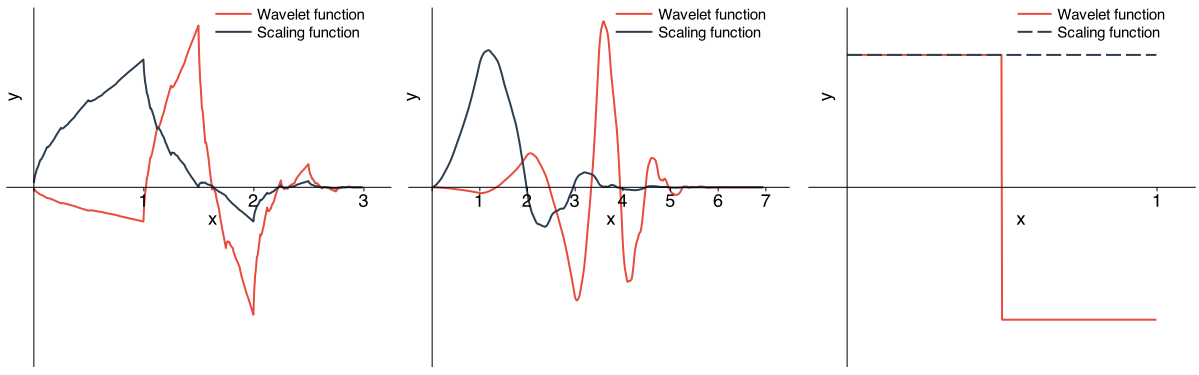
3.1 Introduction to Wavelets

The theory of wavelets is a mathematical tool developed in the 1980s and 90s. The wavelet transform is a transform of an object into the frequency and space (or time) domain. Wavelet transforms can be used for denoising of images. This is done by transforming an image into the frequency domain, where the noise in the image is more distinguishable, and can therefore be eliminated from the signal. On the front page of this thesis you can see a painting which is wavelet transformed. The resulting transformation contains information of the edges and noise of the image. The noise is much easier to recognize in the transformed image than in the original image. When the noise is easier to determine, we can make cuts in the coefficients of the wavelet transformation, resulting in an image with a lower amount of information. The same method is what we want to use in particle physics. We can wavelet transform the particles and then make a cut in their momenta, such that the pile-up particles momenta are removed from our image of particles.
In this section we first have a look at wavelets as a mathematical tool, and shortly after an example wavelet transform is shown.

3.1.1 Multiresolution Analysis

Multiresolution analysis is decomposition of an image into several resolutions. The image will be fully represented by a combination of several transformed images at different resolutions. The transformed images encodes features of different scales, each contribution is sensitive to a different resolution. We use two families of functions to represent an image at different resolutions: The *scaling* and *wavelet* functions. The functions are manipulated through two parameters: Scaling and translating.

In wavelet analysis one mother wavelet is used, which determines the shape of all the other wavelet functions used in the representation of the signal. The mother wavelet can come in different forms. In general the mother wavelet is a normalized function, and it does not have to be differentiable. Some examples of mother wavelets are the Haar and Daubechies wavelet functions, Figure 3.1.



(a) Haar (D_2) (b) Daubechies 4 (D_4) (c) Daubechies 8 (D_8)

Figure 3.1: Examples of mother wavelet and scaling functions [26].

The *scaling functions* are defined by,

$$\phi_{j,k} = 2^{j/2}\phi(2^j x - k) \tag{3.1}$$

where j is the dilation parameter and k is the position parameter. If we fix $j = j_0$, then we can generate a set of functions translated by k,

$$\{\phi_{j_0,k}(x)\} \tag{3.2}$$

which will together cover some subspace V_0. The scaling function can only be translated by integer values, and it is orthogonal to its integer translated scaling functions. Therefore the integer translation makes sure that the scaling functions do not overlap with the original scaling function. A scaling function at scale j will span some subspace V_j,

$$V_j \quad = \quad \text{Span}\{\phi_{j,k}(x)\} \tag{3.3}$$

The scaling function $\phi_{j_1,k}(x)$ will cover the subspace V_1 and the scaling function $\phi_{j_2,k}$ will cover the subspace V_2 and so on. In general the subspaces will make up the entire space,

$$V_{-\infty} \subset ... \subset V_{-1} \subset V_0 \subset V_1 \subset ... \subset V_\infty \tag{3.4}$$

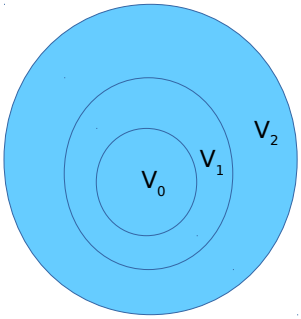


Figure 3.2: Example of the subspaces of V.

The *wavelet functions* are defined similarly to the scaling functions

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j - k) \tag{3.5}$$

i.e. the wavelet function can also be dilated by j and translated by k. A set of wavelet functions can be made from a mother wavelet. The mother wavelet is a function which is limited in the space domain, meaning it has values in a range, og zeros otherwise. We define the mother wavelet as

$$\psi(x) = \psi_{0,0}(x) \tag{3.6}$$

The mother wavelet has zero mean,

$$\int_{-\infty}^{\infty} \psi(x)dx = 0 \tag{3.7}$$

and is normalized

$$||\psi(x)||^2 = \int_{-\infty}^{\infty} \psi(x)\psi^* dx = 1 \tag{3.8}$$

Using the mother wavelet we can form a set of wavelet functions,

$$\{\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)\} \tag{3.9}$$

where the set is orthonormal. A wavelet function $\psi_{j,k}(x)$ will cover the subspace W_j, i.e.

$$W_j = \text{Span}\{\psi_{j,k}(t)\} \tag{3.10}$$

where W_j is defined as the difference of the subspace V_{j+1} and V_j. This means that the wavelet functions cover the spaces in between the V spaces, such that

$$V_1 = V_0 \oplus W_0 \tag{3.11}$$
$$V_2 = V_1 \oplus W_1 = V_0 \oplus W_0 \oplus w_1 \tag{3.12}$$
$$V_n = V_{n-1} \oplus W_{n-1} \oplus V_{n-2} \oplus W_{n-2} \oplus ... \oplus V_0 \oplus W_0 \tag{3.13}$$
$$= V_0 \oplus W_0 \oplus W_1 \oplus ... \oplus W_{n-1} \tag{3.14}$$

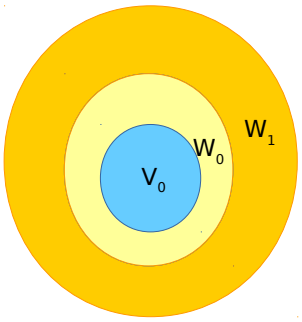


Figure 3.3: Example of the subspaces of the scaling functions (subspaces V) and wavelet functions (subspaces W).

Using the scaling and wavelet functions we want to represent a signal for some resolution j. The wavelet transform can be used to approximate a signal $f(x)$ using the scaling functions $\phi_{j_0,k}$, and the wavelet functions $\psi_{j,k}$ with $j \geq j_0$.

The wavelet transform is then represented by scaling and wavelet coefficients,

$$\omega_\phi(j_0,k) = \int f(x)\phi_{j_0,k}(x)dx \tag{3.15}$$
$$\omega_\psi(j,k) = \int f(x)\psi_{j,k}(x)dx \tag{3.16}$$

and the inverse wavelet transform can be obtained by using the wavelet coefficients,

$$f(x) = \frac{1}{C_\psi}\int_0^\infty \int_{-\infty}^\infty \omega^*_{\psi j,k}(x)\psi(j,k)djdk \tag{3.17}$$

where C_ψ is defined as,

$$C_\psi = \int_0^\infty \frac{|\Psi(u)|^2}{u}du < \infty \tag{3.18}$$

where $\Psi(u)$ is the Fourier transform of the mother wavelet, $\psi(x)$. Equation 3.18 is named the admissibility condition.

3.1.2 Discrete Wavelet transform in One Dimension

The subspaces spanned by the scaling function at low scales are nested with those at higher scales. The equation describing this is called the multiresolution equation, or dilation equation

$$\phi(x) = \sum_k h_\phi[k]\sqrt{2}\phi(2x-k) \tag{3.19}$$

where h_ϕ is known as a *low pass filter*. The meaning of the dilation equation is that the function $\phi(2x)$ is a function with higher frequency components than $\phi(x)$. The functions are designed such that we can apply the low pass filter $h_\phi[n]$ to the function $\phi(2x)$ to obtain the

function with a lower frequency $\phi(x)$. There is a similar relation for the wavelet functions,

$$\psi(x) = \sum_k h_\psi[k]\sqrt{2}\phi(2x-k) \tag{3.20}$$

where h_ψ is known as a *high pass filter*. These two filters are related by

$$h_\psi[k] = (-1)^k h_\phi[1-k] \tag{3.21}$$

Using the low and high pass filters the scaling and wavelet functions can be made at some resolution j, see Figure 3.4.

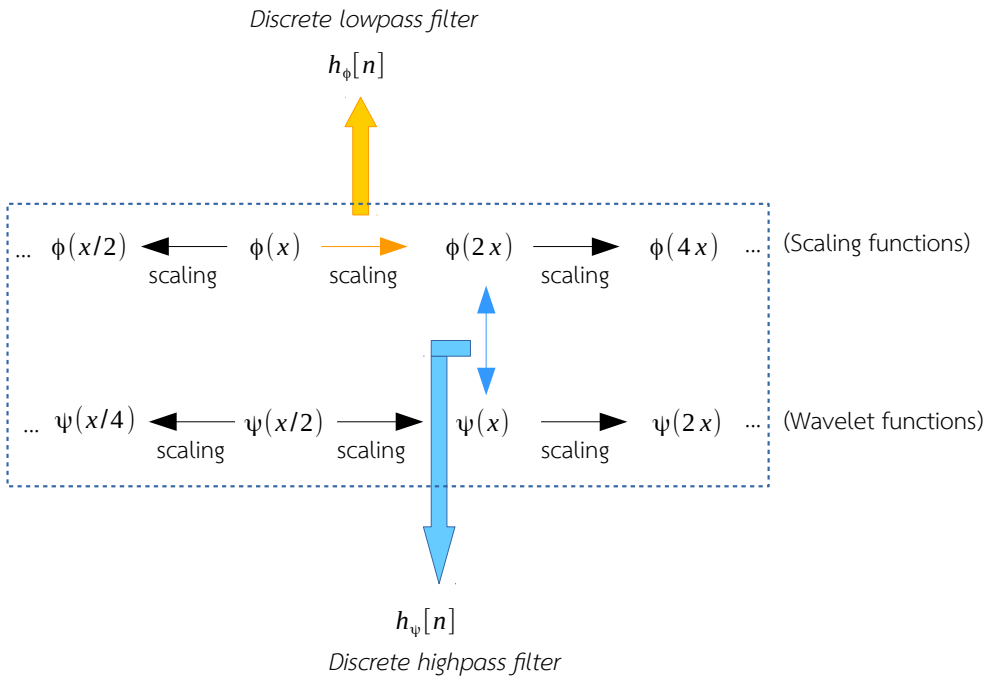


Figure 3.4: When applying the low pass filter to a scaling function, the scaling function changes to a higher or lower scale. When applying a high pass filter you can determine a wavelet function from a scaling function.

When the scaling and wavelet functions are given, we can represent a discrete signal $s[n]$ by,

$$s[n] = \frac{1}{\sqrt{M}}\sum_k \omega_\phi[j_0,k]\phi_{j_0,k}[n] + \frac{1}{\sqrt{M}}\sum_{j=j_0}\sum_k \omega_\psi[j_0,k]\psi_{j,k}[n] \tag{3.22}$$

Since the scaling functions are orthogonal to each other and the same with the wavelet functions, we can simply take the inner product to obtain the coefficients. The scaling coefficients are then,

$$\omega_\phi[j_0,k] = \frac{1}{\sqrt{M}}\sum_n s[n]\phi_{j_0,k}[n] \tag{3.23}$$

and the wavelet coefficients are

$$\omega_\psi[j,k] = \frac{1}{\sqrt{M}}\sum_n s[n]\psi_{j,k}[n] \tag{3.24}$$

The scaling coefficients are also named *approximation coefficients* and the wavelet coefficients are also named *detail coefficients*. The *approximation coefficients* represents the lowest level of of decomposition.

3.1.3 Example: Haar Wavelet in One Dimension

The Haar wavelet is a piecewise constant function. The Haar scaling function is,

$$\phi(x) = \begin{cases} 1 & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Using Equation 3.1 we can define for $j = 1$ the scaling functions,

$$\phi_{1,0} = 2^{1/2}\phi(2x) = 2^{1/2} \cdot \begin{cases} 1 & \text{if } 0 \leq x \leq 1/2 \\ 0 & \text{otherwise} \end{cases} \tag{3.26}$$

$$\phi_{1,1} = 2^{1/2}\phi(2x-1) = 2^{j/2} \cdot \begin{cases} 1 & \text{if } 1/2 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{3.27}$$

Then we have,

$$\phi(x) = h_\phi[0] \cdot 2^{j/2}\phi(2x) + h_\phi[1] \cdot 2^{j/2}\phi(2x-1) \tag{3.28}$$

and by inserting $x = 0$ and $x = 1$ the low pass filter values are then,

$$h_\phi[0] = \frac{1}{\sqrt{2}} \tag{3.29}$$

$$h_\phi[1] = \frac{1}{\sqrt{2}} \tag{3.30}$$

When the low pass filter coefficients are determined, it is easy to determine the high pass filters by equation 3.16,

$$h_\psi[0] = (-1)^0 h_\phi[1-0] = \frac{1}{\sqrt{2}} \tag{3.31}$$

$$h_\psi[1] = (-1)^1 h_\phi[1-1] = -\frac{1}{\sqrt{2}} \tag{3.32}$$

and the wavelet function is then,

$$\psi_{0,0} = h_\psi(0)\sqrt{2}\phi(2x) + h_\psi(1)\sqrt{2}\phi(2x-1) \tag{3.33}$$

$$= \phi(2x) - \phi(2x-1) \tag{3.34}$$

$$= \begin{cases} 1 & \text{if } 0 \leq x \geq 1/2 \\ -1 & \text{if } 1/2 \leq x < 1 \\ 0 & \text{otherwise} \end{cases} \tag{3.35}$$

This is just the scaling and wavelet function at scale $j = 0$. The wavelet and scaling functions at higher scales can be made in the

same way. When the wavelet and scaling functions are obtained, they can be used for the wavelet transformation, and the scaling and wavelet coefficients are calculated from equation 3.21 and 3.22. A full example of a 1D Haar wavelet transform is shown in Figure 3.5 and the inverse transformation is in Figure 3.6.

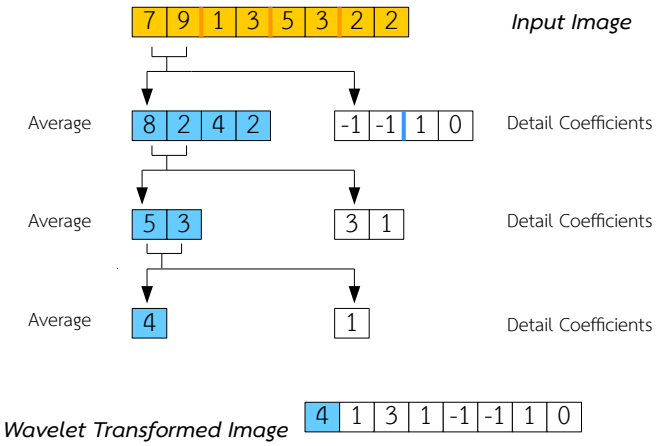


Figure 3.5: The blue numbers represent averages, the white numbers represent wavelet or difference coefficients and the orange numbers are the original signal.
This is an example of the Haar wavelet transform. The algorithm will take the average of elements two by two, and subtract the neighbouring element. By subtracting the neighbouring element from the average, the difference is stored. The difference is also know as detail coefficients or wavelet coefficients. The only coefficients we need to keep are the first average and all of the detail coefficients, to be able to fully restore our image.

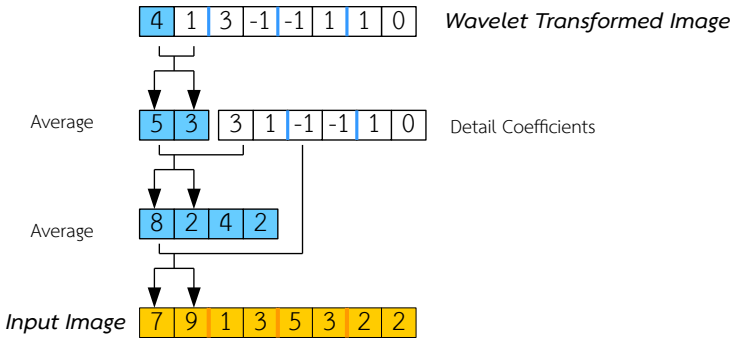


Figure 3.6: In the transformed image, the only average left is in the first position. In this case it is 6. The other averages will be known again by adding and subtracting the difference coefficients. Here 0 is added and subtracted in the first iteration. In the second iteration 2 is added and subtracted to 6, and so forth. When the first elements of the original signal are known, the next average will be treated.

In terms of the high and low pass filters the wavelet transformation is shown in a schematic in Figure 3.6. The schematic corresponds exactly to what we did in Figure 3.5, but here it is more general, i.e. it is the procedure for all 1D wavelet transforms, independent of the mother wavelet. The input image will pass through a high and a low pass filter. The image through the high pass filter corresponds to the wavelet coefficients at the first level. To get the wavelet coefficients at the second level, the input image goes through a low pass filter and then a high pass filter. This continues until after using the last high pass filter, and only one wavelet coefficient remains at the last level.

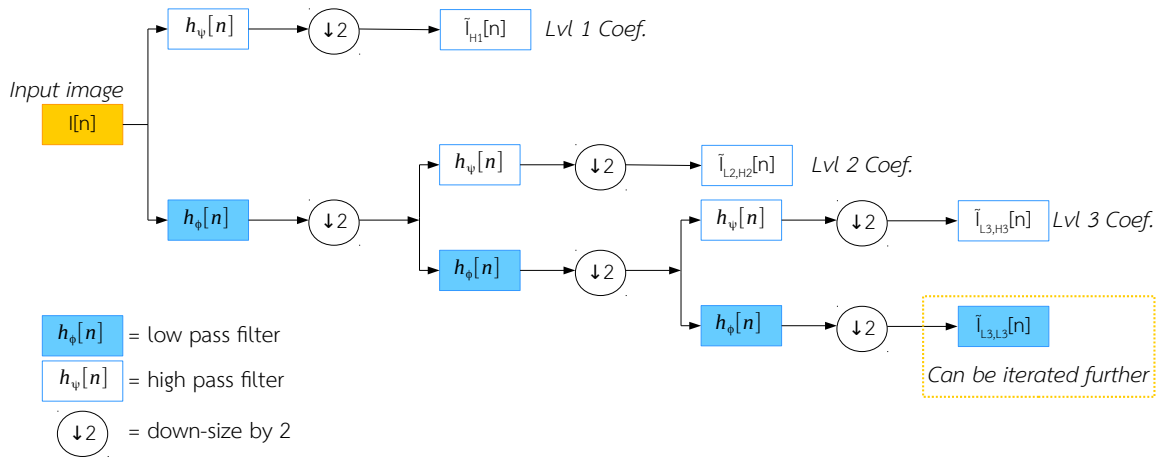


Figure 3.6: Schematic of a 1D wavelet transform. The by 2 means down-sizing, meaning that the number of wavelet coefficients for the next level is divided by two.

Matrix representation The transformation performed in Figure 3.5 can also represented in matrix form as shown in Figure 3.7.

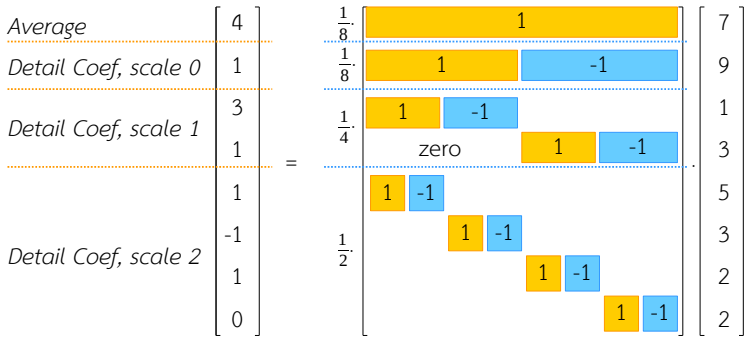


Figure 3.7: A matrix representation of the Haar wavelet transform.

Notice that the transformation in Figure 3.7 and 3.5 is *not* a normalized wavelet transform. It is just an example of a wavelet transform. If the wavelet coefficients were normalized, the total information of the wavelet transformed image should have a value close to the information of the original image. For example, if the sum of the image is 48, then at the first scale, the sum of the coefficients should also be 48. This all depends on the choice of normalization of the wavelet coefficients. One popular option is to multiply the coefficients by $2^{j/2}$.

If we use the normalized wavelet transformation matrix for our transformation, we can obtain wavelet coefficients as a linear combination of the wavelet functions, Figure 3.8.

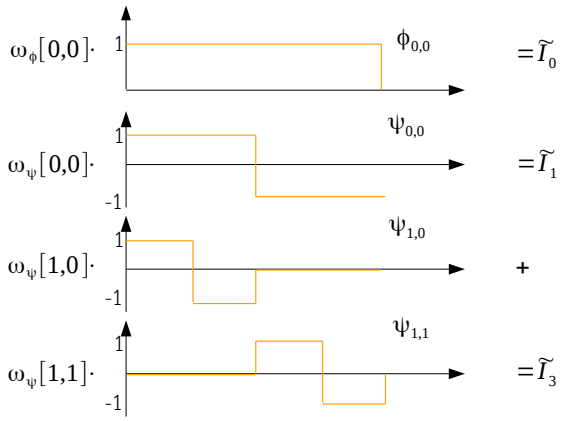


Figure 3.8: The linear dependency of the wavelet/detail coefficients, the wavelet functions and the image. I_j is the image at scale j. This is for a normalized wavelet transformation.

3.1.4 Discrete Wavelet Transform in Two Dimensions

The discrete wavlet transform in two dimensions can be performed by

- Computing the 1-D transform for each column in the image, and placing the resulting vectors into a new matrix.

- Computing the 1-D transform of each row

This also correspond to using the high- and low-pass filters. To get the first transformation in 1D we simply use the high- and low-pass filters to get the results. After using the filters, the resulting transformation is at scale 0, and we say that we downsized the input "by two". To do the wavelet transformation for scale two in 2D we apply the high- and low-pass filters again. This is illustrated in Figure 3.8.

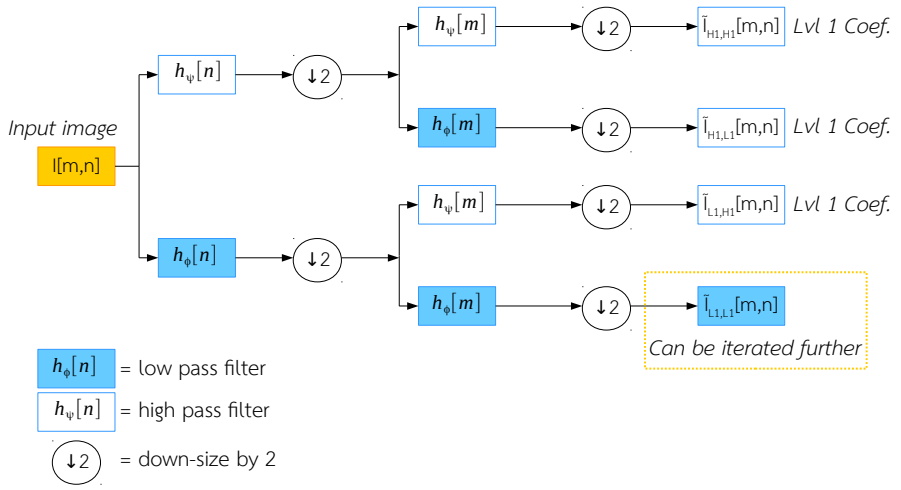


Figure 3.8: Schematic of a 2D wavelet transform. The first step is to do a basic wavelet transform in one direction. Then we can do a wavelet transform in the other direction. The by 2 means scaling, such that we have the image at different resoultions. The result is 4 different images at different resolutions.

Matrix representation Again we can make a matrix representation of the wavelet transform, but this time in 2 dimensions. Here there will be several matrices made of wavelet bases, which will be multiplied by the image of interest. The matrix wavelet bases for the Haar wavelet is shown in Figure 3.9 for scales $i, j = 0..4$.

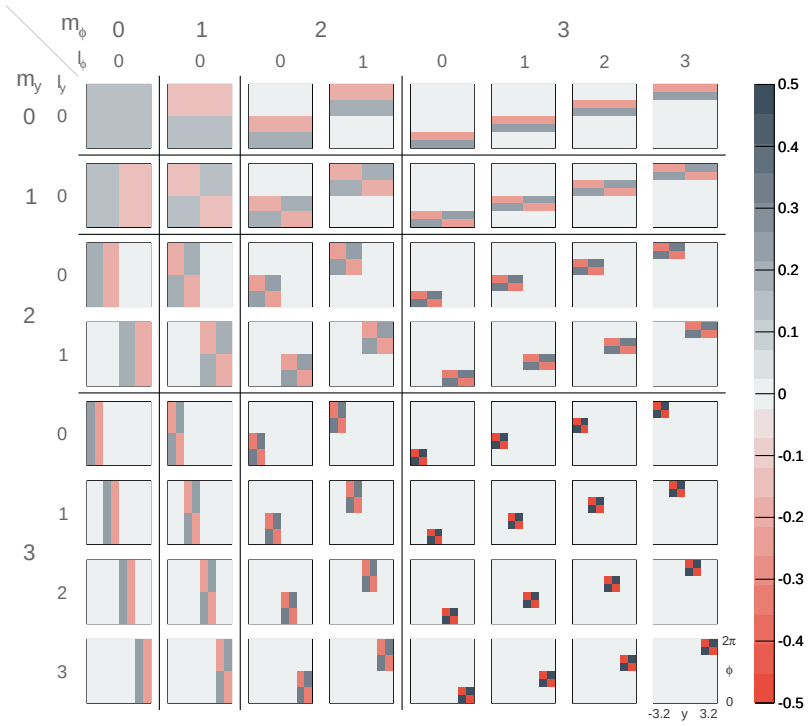


Figure 3.9: The Haar wavelet basis functions two dimensions, from [26].

To get a wavelet transformation of an image, we multiply the image with the wavelet bases for each scale. The result is an image for each scale. For example see Figure 3.10, where a wavelet transformation was performed on a simple dataset. A full vizualization of a wavelet transform up to scale 6 is in Figure 4.18. This is just a demonstration of how the wavelet coefficients could look like for the Haar wavelet algorithm. In further analysis in this thesis the transformations are done at a much larger scale of 64 for a good resolution.

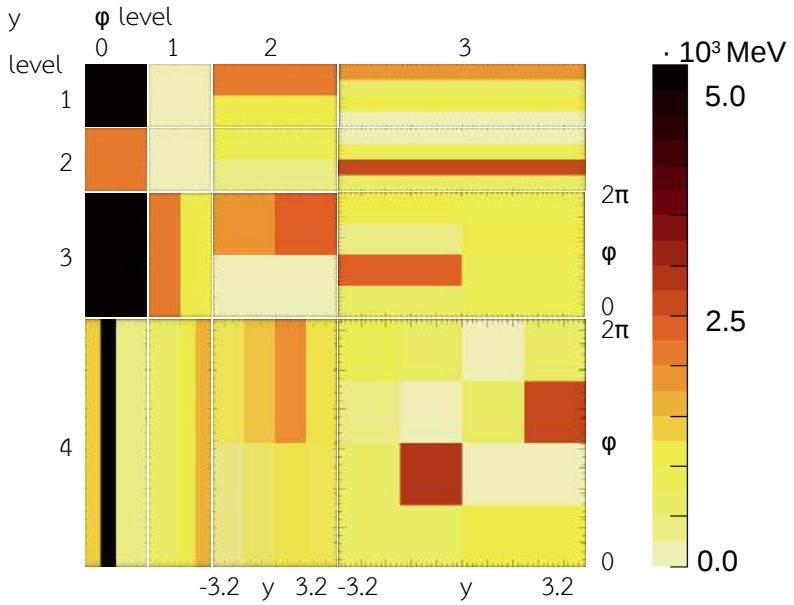


Figure 3.10: Wavelet coefficients made from the wavelet basis functions from Figure 3.9.

3.1.5 *Computational Complexity of Discrete Wavelet Transform*

In the wavelet transform, each filtered output is decimated by a factor of 2. For example, consider a input of size N. For the first level, the low-pass filter does a total of $N/2$ computations, and the high-pass filter also does $N/2$ computations. Therefore the total number of computations at the first level is N. At the second level each filter performs a total of $N/4$ computations, and the total number of computations at level two is $N/2$. In an m-level wavelet transform, the total number of computations is then

$$N + \frac{N}{2} + \frac{N}{4} + ... + 2 = 2(N-1) \tag{3.36}$$

This correspond to a run time of $\mathcal{O}(N)$. For a 2 dimensional wavelet analysis we do the same in two directions and the resulting run time is $\mathcal{O}(N+M)$ for an image of size $N \times M$.

3.1.6 *Comparison with Fourier*

The Fourier transform is used in many fields of physics, as an approximation of some signal. The Fourier transform is powerful when used on stationary signals, meaning that the is no large change in the properties of the signal. This is especially good when a signal is composed of sine and cosine signals. But the Fourier transform is not as good at approximating non-stationary signals. This is where the Wavelet transform can be more useful. Since the Wavelet functions does not have to be differentiable, they can contain large variations, and can

be able to approximate spiky signals without too many computations. The main difference of Fourier and wavelets is that wavelets are well localized in both time and frequency domain, where the standard Fourier transform is only localized in the frequency. I.e. the fourier coefficients does not contain information of the position of time of signal. Although the short-time Fourier transform (STFT) is localized in time and frequency, there are difficulties with the frequency time resolution and wavelets often give a better signal representation.

3.2 Filtering and Wavelet Coefficient Thresholding

The image of the data contains noise, hence the wavelet coefficients will also contain noise. It is therefore necessary to determine which coefficients contain noise, and which coefficients contain signal, or a mixture of these. There are several ways to remove the noise in the wavelet coefficients. Normally these approaches are divided into two groups: Hard thresholding and soft thresholding. The methods to remove noise in images using the coefficients of wavelet transforms are described in the this section.

3.2.1 Coefficient Thresholding in Particle Physics

At the LHC there is pile-up in the data samples collected. Pile-up is when there are several protons colliding at the same time. This problem occurs when we are unable to determine which vertex a particle is coming from. By being unable to determine the origin of the particle, we need to determine to keep or throw this particle from our dataset. If we were to throw away each particle we were unsure about, the reconstruction of the event would eventually be wrong. Wavelets can be a very useful way of giving a qualified guess for which particles we should remove from our event. The way to use wavelets in particle physics is to give the algorithm a list of particles with their energy and position, and the wavelet algorithm will then remove the pile-up particles.

An wavelet image at level j has coefficients given by $\omega_{j,k}$ and we want to introduce a statistical significance test for each coefficient. When the method is chosen for determining the significance of a coefficient, we have removed or reduced some of the wavelet coefficients. The inverse wavelet transform can then be performed, and the resulting image should be of less information than the original image. In particle physics we have a discrete image of particles. Therefore when the inverse wavelet transform is performed, the energies of the particles should not be altered, but we should rather *remove* the particles we suspect for being pile-up particles. In [27] such a method was suggested. In [27] the choice of removing a particle depended on the ratio between the pixel in the original image, and the wavelet cleaned image.

$$r_{pixel} = \frac{p_{T,clean}^{pixel}}{p_{T,original}^{pixel}} \tag{3.37}$$

With this ratio the particles energy is scaled by

$$\tilde{p}_T = \begin{cases} p_T & \text{if } r > cut \\ 0 & \text{otherwise} \end{cases} \tag{3.38}$$

The idea is illustrated in Figure 3.11,

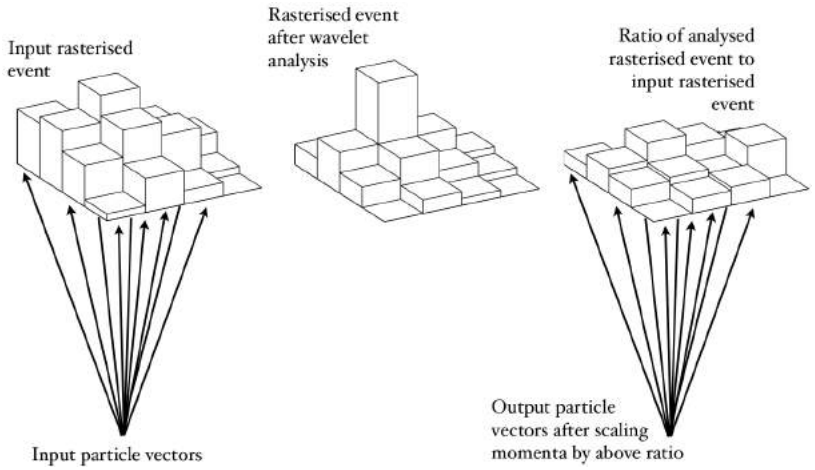


Figure 3.11: The process of removal of pile-up using wavelets, from [27]. After using the wavelet analysis the image has changed. The ratio of the input image and the wavelet filtered image decide which particles should be kept in the event and which should be thrown.

3.2.2 *Filtering*

Two main filtering methods are used: Hard thresholding and soft thresholding. Hard thresholding is where a coefficient is set to zero, if it is less than some threshold. Hard thresholding can be done using flat denoising or track filtering. In soft thresholding for some threshold a coefficient is given a new value, which is not necessarily zero. Soft thresholding can be done using the methods: Track scaling, maximum entropy or a user definition. If necessary one can also combine using hard and soft thresholding. Soft thresholding is when the coefficients are not removed but rather rescaled accordingly to some function.

Flat Denoising Flat denoising is making a constant cut in the wavelet coefficients. Any coefficient with a value below this threshold is set to zero.

$$\tilde{\omega}_{j,k} = \begin{cases} \omega_{j,k} & \text{if } |\omega_{j,k}| \geq T_j \\ 0 & \text{otherwise} \end{cases} \tag{3.39}$$

where T can be dependent on the scale j or constant.

Algorithm 1 Flat Denoising

for all $\omega_{j,k}$ **do**
  if $\omega_{j,k} < T_j$ **then**
    $\tilde{\omega}_{j,k} = 0$
  else
    $\tilde{\omega}_{j,k} = \omega_{j,k}$
  end if
end for

3.2.3 Using a Model

The wavelet coefficients can also be adjusted by using a model. The model can represent how the data should most likely look like. A good model involves using the information from track particles. In ATLAS the origin of track particles are estimated. The track particles come from the primary vertex or some secondary vertex. The pixel detector at ATLAS is good at determining the origin of charged particles, but can not determine the origin of neutral particles. Since the charged particle flow objects do not contain information of origin, we can match the charged particle flow objects to the track particles, and the track particles to the primary vertex container, as illustrated in Figure 3.7. Now the model can contain the charged particle flow objects matched to a primary vertex. This model can be used to set a probability function, such that if a wavelet coefficient almost matches a wavelet coefficient from the primary vertex charged particles, the coefficient should be kept at a high value, and not cut. You can also make the assumption, that if a neutral particle with a wavelet coefficient close to the charged particles from the primary vertex, it should be kept. There might be a correlation between the charged and neutral particles. Although the correlation has not been precisely determined in this paper, the assumption of some correlation is useful for filtering the wavelet coefficients. The methods *track filtering* and *track scaling* make use of track particles. The maximum entropy method makes use of both a model for the signal and noise, as explained in Section 3.3.

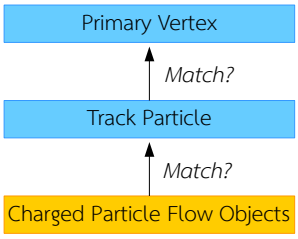


Figure 3.12: Flow of finding the primary vertex of a charged particle flow object.

Track filtering Track filtering is where the track information is used for making a cut in the wavelet coefficients. The track information is whether a *charged* particle flow object comes from a primary vertex or a secondary vertex. The idea emerges from assuming the following,

$$\frac{\omega_{j,k}^{\text{all tracks from PV}}}{\omega_{j,k}^{\text{all tracks}}} \approx \frac{\omega_{j,k}^{\text{all PFOs from PV}}}{\omega_{j,k}^{\text{all PFOs}}} \tag{3.40}$$

i.e. that the wavelet coefficients of pile-up particles among track particles, is somehow related to the wavelet coefficients of pile-up particles among particle flow objects. The scaling of the coefficients will by determined by the ratio of the wavelet coefficients for all tracks and the wavelet coefficients from tracks originating from the primary vertex. The ratio tells you something about the amount of pile-up for one particular wavelet coefficient. The wavelet coefficient will be set as follows,

$$\tilde{\omega}_{j,k} = \begin{cases} \omega_{j,k} & \text{if } \frac{|\omega_{j,k}^{\text{all tracks from PV}}|}{|\omega_{j,k}^{\text{all tracks}}|} \geq T_{ratio} \\ 0 & \text{otherwise} \end{cases} \tag{3.41}$$

The track filtering algorithm is in Algorithm 2.

Algorithm 2 Track Filtering

$\omega_{j,k}$ = wavelet coefficients from the image to denoise
$\omega_{j,k}^{\text{all tracks}}$ = wavelet coefficients from the track particles
$\omega_{j,k}^{\text{all tracks from PV}}$ = wavelet coefficients from the primary vertex track particles
for all $\omega_{j,k}$ **do**
  if $\omega_{j,k} = 0.0$ **then**
    do nothing
  else if $\omega_{j,k}^{\text{all tracks from PV}} / \omega_{j,k}^{\text{all tracks}} < T_{ratio}$ **then**
    $\tilde{\omega}_{j,k} = 0$
  end if
end for

Track Scaling In this method the information of the tracks are used. The wavelet coefficients are scaled by the ratio of the wavelet coefficients from all tracks and tracks from the primary vertex. Other choices of scaling could also be considered, e.g. scaling the wavelet coefficients by a ratio of the charged particle flow objects from the primary vertex.

Algorithm 3 Track Scaling

$\omega_{j,k}$ = wavelet coefficients from the image to denoise
$\omega_{j,k}^{\text{all tracks}}$ = wavelet coefficients from the track particles
$\omega_{j,k}^{\text{all tracks from PV}}$ = wavelet coefficients from the primary vertex track particles
for all $\omega_{j,k}$ **do**
  if $\omega_{j,k} = 0.0$ **then**
    do nothing
  else
    $\tilde{\omega}_{j,k} = \omega_{j,k}^{\text{all tracks}} / \omega_{j,k}^{\text{all tracks from PV}}$
  end if
end for

Dynamic Flat Denoising dependent on a ratio These methods depends on assuming that the fraction of neutral and charged particles which come from pile-up can be used. This means that e.g. if 20 % of all charged particles are from pile-up[1], $20\% \times \beta$ of all neutral particles should also be from pile-up. Since the ratio is known for charged particles, we could use this for cutting in the neutral particles. The neutral particles origin can not be determined otherwise. The idea here is to make a dynamic flat denoising, such that we could make a more harsh cuts in the events with a high ratio of charged pile-up particles.

[1] where the charged particles coming from pile-up are determined as the track particles matched to a primary vertex.

The ratio denoising algorithm will make cuts in the wavelet coefficients until the *number of* neutral particles remaining is comparable

to the number of charged particles from the primary vertex, see Algorithm 4.

Algorithm 4 Ratio Denoising

pfo_{ch}^{PV} = charged PFOs from primary vertex
pfo_{ch} = charged PFOs
pfo_{neu}^{PV} = neutral PFOs from primary vertex
pfo_{neu} = all neutral PFOs
cut = 1000 MeV
β = const.
$ratio = \frac{|pfo_{ch}^{PV}|}{|pfo_{ch}|}$
while $\frac{|pfo_{neu}^{PV}|}{|pfo_{neu}|} < ratio \cdot \beta$ **do**
  $pfo_{neu}^{PV} = FlatDenoise(cut, pfo_{neu})$
  $cut+ = 100$ MeV
end while

Dynamic Flat Denoising dependent on a p_T-ratio In the p_T ratio denoising algorithm, the cut does not depends on the number of particles, but rather the $\sum p_T$ of the particles. The idea here is to continue to make cuts in the neutral wavelet coefficients, until the total energy of the neutral particles is comparable to the total energy of the charged particles coming from the primary event.

Algorithm 5 p_T-Ratio Denoising

pfo_{ch}^{PV} = charged PFOs from primary vertex
pfo_{ch} = charged PFOs
pfo_{neu}^{PV} = neutral PFOs from primary vertex
pfo_{neu} = all neutral PFOs
cut = 1000 MeV
β = const.
$ratio_{p_T} = (\sum_{E_T} pfo_{ch}^{PV})/(\sum_{p_T} pfo_{ch})$
while $(\sum_{p_T} pfo_{neu}^{PV})/(\sum_{p_T} pfo_{neu}) < ratio \cdot \beta$ **do**
  $pfo_{neu}^{PV} = FlatDenoise(cut, pfo_{neu})$
  $cut+ = 100$ MeV
end while

A combination The different choices for cuts in coefficients can be combined. The track denoising method for the charged particle flow objects seems promising. The neutral particle flow objects do not have any information of whether it comes from a primary or a secondary vertex, and is therefore a more 'random' object than the charged particle flow objects. Therefore a combination of the algorithms can be used, such that one uses e.g. track scaling for the charged particle flow objects and ratio denoising for the neutral particle flow objects.

3.3 Information Theory and Entropy

The concept of entropy was first introduced by Boltzmann to statistical mechanics to measure the number of ways a macroscopic state can be realized. In 1948 Shannon founded the mathematical theory of communication. Shannon suggested that the information gained in a measurement depends on the number of possible outcomes from which one is realized.

In this section we look into how to use the maximum entropy principle to make efficient cuts in the wavelet coefficients. The wavelet transform is carried out as usual, and the maximum entropy principle is used on the transformed image. This method will make use of the information of an image can be estimated as entropy of the image. In physics we use entropy as a measurement of the chaos of a system. I.e. if an image has a high entropy, it means that it is noisy. By assuming our pile-up particles appear as Gaussian noise in an image, we can denoise our image by the method of maximum entropy (MEM).

3.3.1 Information Theory

The data at LHC is often corrupted by noise or pile-up particles. In general, noise can be described as following a Gaussian or Poisson distribution or a combination of these two. Gaussian noise is similar to what you can find on old televisions showing flickering of black and white dots, and Poisson noise is more a periodic kind of noise. In this analysis using information theory and entropy, I will be assuming the pile-up noise is Gaussian, although as we will see in Section 4.2.3 that this is a bit of a rough estimation, since the pile-up particles are not evenly distributed in η. Let us call the observed data Y and the data without noise X, meaning we want to estimate X given Y. Bayes theorem states that

$$p(X|Y) = \frac{p(Y|X) \cdot p(X)}{p(Y)} \tag{3.42}$$

here $p(Y|X)$ is the probability distribution of getting the data Y given the original signal X, i.e. it represents the distribution of noise. In the case with uncorrelated Gaussian noise it is,

$$p(Y|X) = exp(-\sum_{pixels} \frac{(Y-X)^2}{2\sigma^2}) \tag{3.43}$$

$p(Y)$ is independent of X and is therefore considered as stationary noise, i.e. constant. $p(X)$ is the distribution of the solution X. It is possible to derive the probability of X from its entropy. The main idea of information theory is to get a relation between the received information and the probability of the observed event.

Consider,

$$I(E) = f(p) \tag{3.44}$$

where $I(E)$ is the information gained given an event, and p is the probability of the event occurring, and f is some function to be defined[2]. The information is a decreasing function of probability. This means that the more information we have, the less will be the probability of an event. One choice for the information measure is,

$$I(E) = k \cdot \ln(p) \tag{3.47}$$

where k is a constant. k is generally chosen to be -1 to make the information a positive measure. The mean information is denoted as

$$H = -\sum_i p_i \ln(p_i) \tag{3.48}$$

which a physicist might recognize as the entropy of a system. This entropy was established by Shannon in 1948. Entropy is maximal when all events have the same probability, which basically means that the system is most undefined. The entropy is minimal when one event is sure and the system is perfectly known. The entropy can be used to describe the probability of the solution,

$$p(X) = \exp(-\alpha H(X)) \tag{3.49}$$

So if we know the entropy of the solution, we can estimate $p(X)$. Let us insert these expressions into Bayes theorem,

$$p(X|Y) = \frac{exp(-\sum_{pixels} \frac{(Y-X)^2}{2\sigma^2}) \cdot \exp(-\alpha H(X))}{p(Y)} \tag{3.50}$$

and taking the natural logarithm,

$$\ln(p(X|Y)) = -\sum_{pixels} \frac{(Y-X)^2}{2\sigma^2} - \alpha H(X)) - \ln(p(Y)) \tag{3.51}$$

Since $\ln(p(Y))$ is constant we can omit this term. The most probable solution can then be found by minimizing the function,

$$J(X) = \sum_{pixels} \frac{(Y-X)^2}{2\sigma^2} + \alpha H(X)) = \frac{\chi^2}{2} + \alpha H(X) \tag{3.52}$$

This function is a linear combination of the entropy and a quantity corresponding to χ^2 in statistics. χ^2 is a measure of the disagreement between the data and the predictions of a model. α is a constant parameter and can be viewed as a value fixing the relative weight between the goodness-of-fit and the entropy.

The definition of the entropy is fundamental, since from its definition our solution will depend on it.

[2] We assume that the information has the property of additivity. So if we have only two events we have the information,

$$I(E) = I(E_1) + I(E_2) \tag{3.45}$$

Since E_1 and E_2 are independent, the probability of both occurring is equal to the product of the probabilities associated with the events occurring p_1 and p_2,

$$f(p_1 p_2) = f(p_1) + f(p_2) \tag{3.46}$$

3.3.2 Estimation of Entropy

We want to estimate the unknown probability density of the data, $p(X)$. Shannon defined the entropy of an image by,

$$H_{Shannon}(X) = -\sum_{k}^{N_b} p_k \log p_k \tag{3.53}$$

where X is an image containing integer values, N_b is the number of possible values of a given pixel, and p_k is the usual probability. The problem with Shannons entropy is that it can create an essential error when estimating the entropy with a good precision, this problem is described by Frieden in 1978. Therefore other entropies are normally used. A few possible entropies are,

- Burg (1978)

$$H_{Burg}(X) = -\sum_{k=1}^{pixels} \ln(X_k) \tag{3.54}$$

- Frieden (1978)

$$H_{Frieden}(X) = -\sum_{k=1}^{pixels} X_k \ln(X_k) \tag{3.55}$$

- Gull and Skilling (1991)

$$H_{Gull}(X) = \sum_{k=1}^{pixels} X_k - M_k - X_k \ln \frac{X_k}{M_k} \tag{3.56}$$

where M is a given model which is usually some flat image. The entropy from Gull and Skilling has the vulnerability that it is difficult to estimate a model, and the solution can depend on the choice of a model (Bontekoe et al. 1994).

The entropy function has the property of being maximal when the image is flat, and decreasing when we introduce some information. Therefore minimizing the information is the same as maximizing the entropy, which leads to the famous Maximum Entropy Method (MEM).[3]

One difficulty with the method of maximum entropy is to estimate the α constant in the probability function of X. Later in this chapter, the estimation of α using noise information of the image of particle flow objects is discussed.

When determining a good physical measurement of entropy, we should keep in mind what our signal consist of,

$$Y = X + B + N \tag{3.57}$$

where Y is our received signal, X is the signal of interest, B is the flat background and N is the noise. We set the following criteria upon our problem:

[3] The concept of maximizing or minimizing entropy can be rather confusing, since it depends on the definition of the entropy. E.g. for the Shannon entropy, the information is minimal when the entropy is minimal, but for other common definitions of entropy, the information is minimal when the entropy is maximal. So for the Shannon entropy it should be named Minimal Entropy Method.

- The information in a flat signal is zero. ($X = N = 0, B = const.$)

- The information in a signal is independent of the background noise, i.e. $H(X)$ is independent of B.

- The information in a signal is dependent of the noise, i.e. $H(X)$ is dependent of N.

- The amount of information is dependent on the correlation in the signal. This means that if X contains large peaks above N, the information is high. If you then distribute the individual high peaks among the signal, the information should decrease. So even though the sum of values in X is the same, the information should still decrease.

These points are desirable properties, and should be contained in the choice of entropy function. In [28] the Shannon entropy is chosen as the only one for having all of these properties.

3.3.3 Multiscale Entropy

How can the method of information theory and entropy be introduced in wavelet denoising? Let's start by introducing the entropy as the sum of information contained in our wavelet coefficients,

$$H(X) = \sum_{j=1}^{l} \sum_{k=1}^{N_j} h(\omega_{j,k}) \tag{3.58}$$

where h the information of a wavelet coefficient $\omega_{j,k}$, and we have the same property related to information and entropy as before

$$h(\omega_{j,k}) = -\ln p(\omega_{j,k}) \tag{3.59}$$

For Gaussian noise, this is given by [28]

$$h(\omega_{j,k}) = \frac{\omega_{j,k}^2}{2\sigma_j^2} + const. \tag{3.60}$$

where σ_j is the noise at scale j. In later analysis of minimising the information, the constant term has no effect and will be omitted. Since this information is dependent on the noise, the solution will also be dependent on this modelling of the noise at different scales σ_j. Notice that the information is proportional to the value of the wavelet coefficients. This means that the higher value a wavelet coefficient has, the less probable it is that it is noise. This entropy fulfils the requirements set in prior section, just as Shannon's entropy. Also equivalent to Shannon's entropy, here information increases with entropy, meaning that it is a minimum entropy method.

So far we have defined the total entropy, now let's have a look at an entropy divided into the contribution to entropy from noise, H_n and the contribution of entropy from signal, H_s.

$$H(X) = H_s(X) + H_n(X) \tag{3.61}$$

where X is the input image containing both signal and noise. We can express this in form of wavelet coefficients, where we divide the information coming from a wavelet coefficient into that from signal $h_s(\omega_{j,k})$ and from noise $h_n(\omega_{j,k})$, and we have

$$H_s = \sum_j \sum_k h_s(\omega_{j,k}) \tag{3.62}$$

$$H_n = \sum_j \sum_k h_n(\omega_{j,k}) \tag{3.63}$$

If a wavelet coefficient value is small, it should contribute to h_n as noise, and if it is large compared to the noise standard deviation it should contribute to h_s as signal.

The information functions for the wavelet coefficients in the case of Gaussian noise is [28],

$$h_n(\omega_{j,k}) = \frac{1}{\sigma_j^2} \int_0^{|\omega_{j,k}|} u \cdot \text{erfc}(\frac{|\omega_{j,k}| - u}{\sqrt{2}\sigma_j}) du \tag{3.68}$$

$$h_s(\omega_{j,k}) = \frac{1}{\sigma_j^2} \int_0^{|\omega_{j,k}|} u \cdot \text{erf}(\frac{|\omega_{j,k}| - u}{\sqrt{2}\sigma_j}) du \tag{3.69}$$

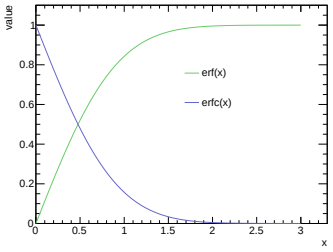


Figure 3.13: The error function is

$$\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt \tag{3.66}$$

and the complementary error function is

$$\text{erfc}(z) = 1 - \text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_z^\infty e^{-t^2} dt \tag{3.67}$$

3.3.4 Multiscale Entropy Filtering

Now we want to remove the noise in the data. We search for a solution, $\tilde{D}$, such that the difference between D and $\tilde{D}$ minimizes the information due to the signal, and such that $\tilde{D}$ minimizes the information due to the noise. So we want to search for a solution $\tilde{D}$, which minimizes the function J,

$$J(\tilde{D}) = H_s(D - \tilde{D}) + H_n(\tilde{D}) \tag{3.70}$$

The solution can be controlled by adding a constant α,

$$J(\tilde{D}) = H_s(D - \tilde{D}) + \alpha \cdot H_n(\tilde{D}) \tag{3.71}$$

In practise we minimize for each coefficient $\omega_{j,k}$ [29],

$$j(\tilde{\omega}_{j,k}) = h_s(\omega_{j,k} - \tilde{\omega}_{j,k}) + \alpha \cdot h_n(\tilde{\omega}_{j,k}) \tag{3.72}$$

The noise standard deviation and α need to be estimated. α can be set to constant, but in this thesis α was chosen such that the standard deviation of the residue of the wavelet coefficients is close to the noise standard deviation. The α constant will start at some minimum value

and then increase until it reaches its maximum value, set by the user, or until the standard deviation of the residue is almost equal to the standard deviation of the noise. The pseudocode for the multiscale entropy algorithm is in Algorithm 6.

3.3.5 *The Multiscale Entropy Filtering Algorithm*

We can now to make an algorithm:

Algorithm 6 Multiscale Entropy Filtering

Estimate the noise σ_j in the data at each scale j
Wavelet transform the data
$\alpha_j^{min} = 0$
$\alpha_j^{max} = 1$
$\epsilon = 0.0001$
for all scale j **do**
  while $\alpha_j^{max} - \alpha_j^{min} > \epsilon$ **do**
    $\alpha_j = \frac{\alpha_j^{min} + \alpha_j^{max}}{2}$
    for all wavelet $\omega_{j,k}$ **do**
      $\tilde{\omega}_{j,k} = \min\left(j(\omega_{j,k}, \tilde{\omega}_{j,k})\right)$
    end for
    $\sigma_j^r = \sqrt{\frac{1}{N_j}\sum_{k=1}^{N} {}_j(\omega_{j,k} - \tilde{\omega}_{j,k})^2}$
    if $\sigma_j^r > \sigma_j$ **then**
      $\alpha_j^{max} = \alpha_j$
    else
      $\alpha_j^{min} = \alpha_j$
    end if
  end while
end for
Reconstruct the filtered image from $\tilde{\omega}_{j,k}$ by the inversed wavelet transform

A figure of the impact of the α constant is in Figure 3.14.

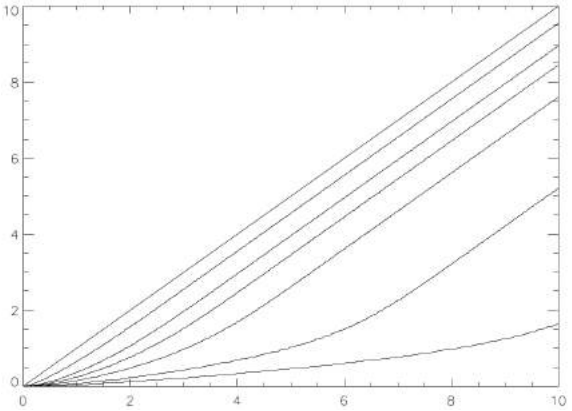


Figure 3.14: Filtered wavelet coeffcients versus the wavelet coefficients with different α values, figure from [28]. From the top curve to the bottom one, α is respectively equal to 0, 0.1, 0.5, 1, 2, 5 and 10. This plot shows that the wavelet coefficient is decreasing when the α constant is increasing.

If the α constant was set prior to computation, it would save a lot

of computing time. This is because, as seen in 6, there is a while-loop ensuring a good α parameter is chosen. The conclusion of whether or not the α parameter was a good choice for some scale j, depends on the standard deviation of the residual of the coefficients at scale j, σ_j. The goal is to have a standard deviation close to the model noise standard deviation. The model noise standard deviation was estimated for each scale using track particles with no association to the primary vertex. The track particles were wavelet transformed, such that the model of the coefficients for noise was known. For this analysis, we simply took the averages of the pile-up standard deviations for all scales j. The result is in Figure 5.9.

3.3.6 Using a Model

If we have a model D_m for the data, the model can also be inserted into the filtering equation,

$$J_m(\tilde{D}) = H_s(D - \tilde{D}) + \alpha H_n(\tilde{D} - D_m) \tag{3.73}$$

and for each wavelet this becomes,

$$j_m(\tilde{\omega}_{j,k}) = h_s(\omega_{j,k} - \tilde{\omega}_{j,k}) + \alpha_j h_n(\tilde{\omega}_{j,k} - \omega^m_{j,k}) \tag{3.74}$$

where $\omega^m_{j,k}$ is the wavelet coefficient from the model. The model can contain different kinds of information. For example it can make sure to keep coefficients with a high value, such as

$$\omega^m_{j,k} = p_s(\omega_{j,k})\omega_{j,k} \tag{3.75}$$

so if $p_s = 1$, then $\omega^m_{j,k} = \omega_{j,k}$ and the coefficient will be not be changed. In the case of removal of pile-up, the model function was made of track particles coming from the primary vertex. Then the image of these particles was wavelet transformed and used in the entropy analysis.

$$\omega^m_{j,k} = \omega^{\text{trk. particles from PV}}_{j,k} \tag{3.76}$$

The results of this can be seen in Section 5.1. The point of using this model is to make sure that if a wavelet coefficient at the same position as a track particle wavelet coefficient, is kept, since it will have a high probability of being from the primary vertex.

Another improvement to the multiscale maximum entropy method is to introduce another constant parameter. If the α_j constants is over- or under-smoothing all the α_j constants can be down- or up-scaled by a constant α_s.

$$j_m(\tilde{\omega}_{j,k}) = h_s(\omega_{j,k} - \tilde{\omega}_{j,k}) + \alpha_j \alpha_s h_n(\tilde{\omega}_{j,k} - \omega^m_{j,k}) \tag{3.77}$$

If $\alpha_s > 1$ the result will be over-smoothing of the image and if $\alpha_s < 1$ it will be under-smoothing. When the multiscale filtering

algorithm was used in this thesis, the user constant α_s was set to be less than 1, since the algorithm smoothing was too strong.

Monte Carlo Datasets for Analysis

4.1 Event Selection

For analysis of missing transverse energy a number of $t\bar{t}$ was used. The $t\bar{t}$ decay into one lepton, one neutrino and four jets. These datasets were chosen because they contain high amount of pile-up. This section describes the kinematics and problems with the samples.

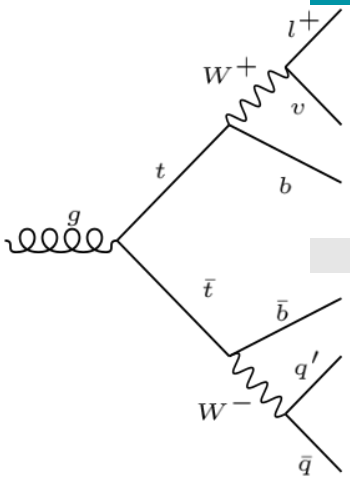


Figure 4.1: $t\bar{t} \to l\nu b\bar{b}q\bar{q}$

4.1.1 Event Selection for $t\bar{t}$

A selection of datasets with the event $t\bar{t} \to l\nu b\bar{b}q\bar{q}$ was chosen. These datasets contain a wide range of physics objects, since there are a total of 4 jets and also very high pile-up (ranges from 70-210 mean interactions per crossing). These datasets will used as a *test* case for the method of wavelet denoising. The performance of wavelet denoising using these samples should be representable of a wide range of signals. There are several datasets of these $t\bar{t}$ events, each with a different pile-up, see Table 4.1. A full summary of the datasets used in this paper is in Appendix A.1. The energy of the neutrino of the event will always have an energy of about 100 GeV or higher, see Figure 4.2 (a). So we can reconstruct the same E_T^{miss} but at different rates of pile-up, which is exactly why this dataset is great for testing of the wavelet methods since it has the same physics event but at different values of pile-up. Originally we searched for a high pile-up sample containing with a clear missing transverse energy signal, such as $W \to l\nu$, but did not find a sufficient one since the pile-up in such samples is mostly about 20-40.

Table 4.1: Summary of the $t\bar{t}$ datasets.

Pileup	**Name**	**Tag**
70 - 90	ttbar_MET100	r7709
70 - 90	ttbar_MET100	r7702
70 - 90	ttbar_MET100	r7699
130 - 150	ttbar_MET100	r7700
130 - 150	ttbar_MET100	r7768
190 - 210	ttbar_MET100	r7769
190 - 210	ttbar_MET100	r7701

The event selection used is the same as the event selection for $t\bar{t}$ in [30], see Table 4.2. This event selection requires events with one isolated high p_T lepton. It also requires 4 jets, and at least one of them should be a b-tagged jet.

Cut	**Event Selection**
Lepton	Electrons (isolated): $p_T > 60(24)$ GeV Muons (isolated): $p_T > 36(24)$ GeV
Exactly one isolated lepton	Muons: $p_T > 25$ GeV, $\|\eta\| < 2.5$ Electrons: $p_T > 25$ GeV
Jets	$\|\eta\| < 2.47$, excluding $1.37 < \|\eta\| < 1.52$ ≥ 4 jets $p_T > 25$ GeV, $\|\eta\| < 1.52$ ≥ 1 b-tagged jets at $\epsilon_b = 70\%$

Table 4.2: Selection of $t\bar{t}$ events.

4.1.2 The problems of the $t\bar{t}$ samples

A problem with the $t\bar{t}$ sample was the reconstruction of particles. With a high number of jets and pile-up in the samples, the jets can sometimes fake a lepton. This means that the signal from a jets in interpreted as a lepton, and therefore we get more counts of high p_T leptons. So when we try to select the good events, not many events survives. This is especially because of the restriction of only having one good lepton. Since the leptons gets an overly estimation of momentum, several leptons survives the cuts, and then the event will fail when the cut of having *only one* good lepton is used. The event selection for three different datasets is in Table 4.3.

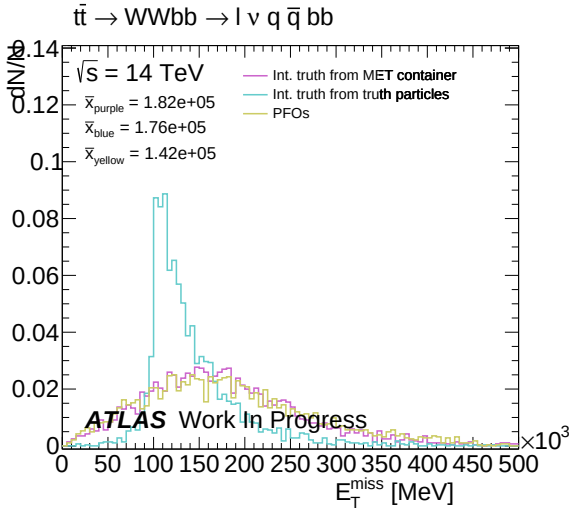
r7768	Initial	50000
	Single lepton trigger	50000
	Exactly one high or iso lepton	24025
	Jets	10825
	b-jets	9432
r7769	Initial	50000
	Single lepton trigger	50000
	Exactly one high or iso lepton	24239
	Jets	11117
	b-jets	10847
r7699	Initial	50000
	Single Lepton Trigger	50000
	Exactly one high or iso lepton	23797
	jets	10654
	b-jets	7135

Table 4.3: Results of the event selection for the $t\bar{t}$ datasets.

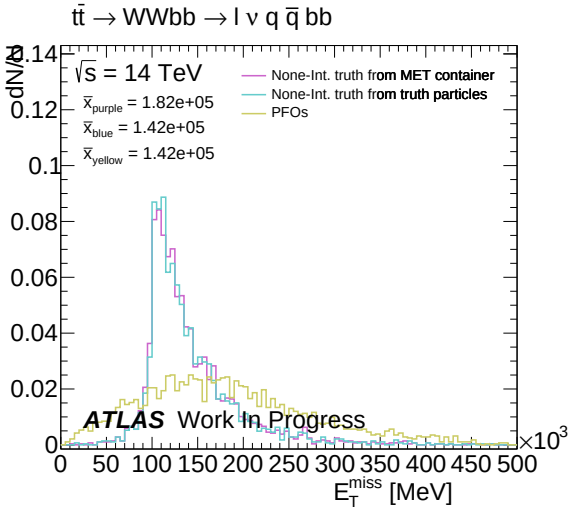
Since the $t\bar{t}$ samples are at *recon* level, and not *merge*, the datasets do not contain the metadata container "Trigger Menu". The "Trigger Menu" is needed to reconstruct missing transverse energy using the ATLAS reconstruction.

It was found that the truth MET containers in the $t\bar{t}$ datasets did not correspond to the true momenta of the neutrino in the events, see the analysis in Figure 4.2. Therefore the estimate of the true missing transverse energy was determined from the vector sum of all truth particles in the $t\bar{t}$ events[1].

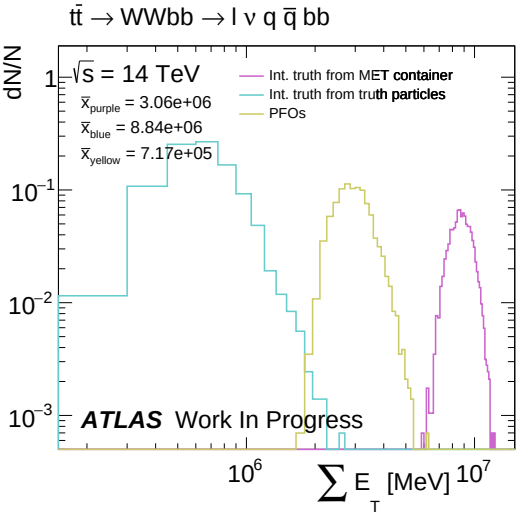
[1] The truth particles from the primary event can be found by looping over all truth particles and selecting the ones with the variable barcode < 100000, which means that they are from the primary event. Note that selecting the truth particles with $status = 0$ will result in a selection of all final state particles, including the ones interactions from materials.



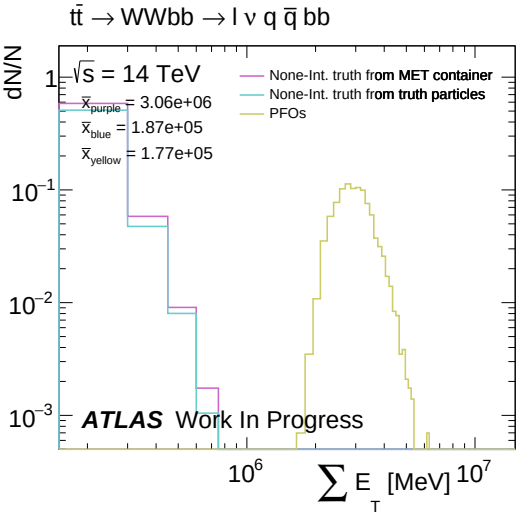
(a) E_T^{miss} from interactive particles, ie. all particles excluding neutrinos.



(b) E_T^{miss} from non-interactive particles, ie. all neutrinos.



(c) $\sum E_T$ from interactive particles, ie. all particles excluding neutrinos.



(d) $\sum E_T$ from non-interactive particles, ie. all neutrinos.

Figure 4.2: Comparison of the truth MET containers, MET made from truth particles and lastly particle flow objects as reference. In the E_T^{miss} and $\sum E_T$ results in (b) and (d), the truth MET containers and truth particles have nearly the same result. In the results in (a) and (c) the truth MET containers and truth particles have a large difference. The E_T^{miss} for the truth container is much more like the particle flow objects. This seems strange, since the true value of E_T^{miss} should not be that similar to the (simulated) experimental value. The same case is shown in (c) where the truth MET containers has a *higher* value of $\sum E_T$ than the particle flow objects. For consistency, we use the MET obtained from truth particles, which does no include the effect of any ATLAS processing.

To summarize,

- Test the method of wavelet on $t\bar{t}$ samples, and compare with truth containers.
- It is better to make the MET truth containers than to use the existing ones in the dataset.

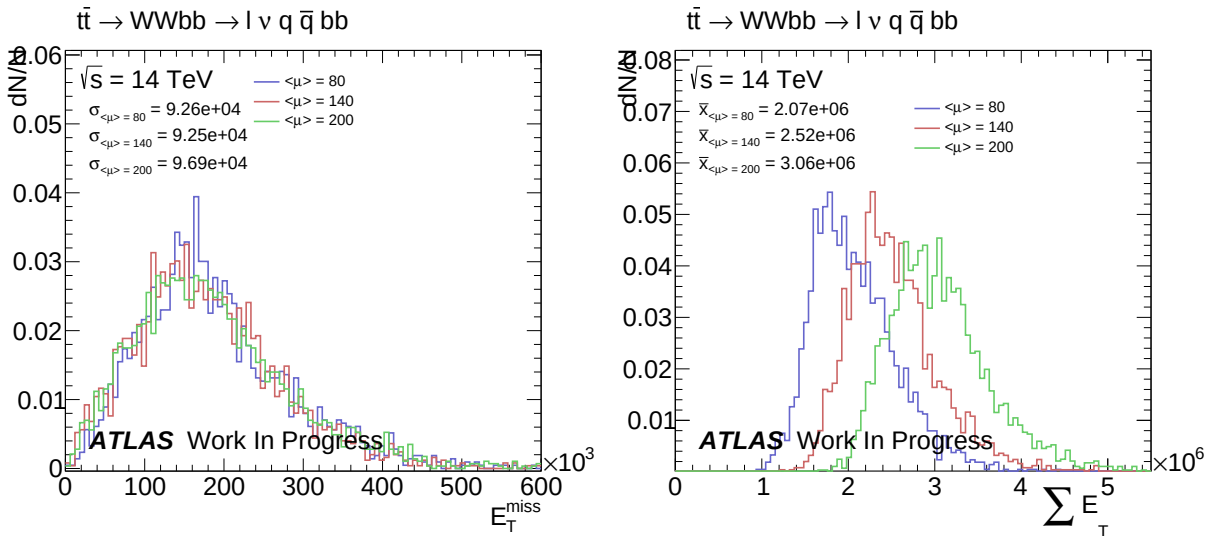
The results for reconstruction of MET in the $t\bar{t}$ samples can not be directly compared to the ATLAS reconstruction.

4.2 Kinematics of the $t\bar{t}$ datasets

In this section the kinematics of the $t\bar{t}$ datasets are shown. The missing transverse energy is calculated without the use of wavelets, and this is the starting point of our analysis. We want to be able to improve the measurements of missing transverse energy in the $t\bar{t}$ samples. The $t\bar{t}$ datasets have 3 stages of pile-up: 70-90, 140-170 and 190-210, and the missing transverse energy will be estimated for each of these stages. Lastly we look into the kinematics of particle flow objects and truth particles.

4.2.1 Missing Transverse Energy

E_T^{miss} is the measure of the missing energy in the event, and is normally assigned to a neutrino. $\sum E_T$ is a measurement of how much energy there is in the system, ie. the absolute sum of the transverse momenta in the event. $\sum E_T$ is a good value to use to see how much energy you can remove from an event, where as the MET can not always show how much energy that was really removed, since it is a vector sum of the momenta of particles. In Figure 4.2 you can see the true values of E_T^{miss} and $\sum E_T$ for the different $t\bar{t}$ datasets. With the different values of pile-up in datasets, the measurement of E_T^{miss} and $\sum E_T$ should also vary. An example of these quantities for three of the datasamples is shown in Figure 4.3.

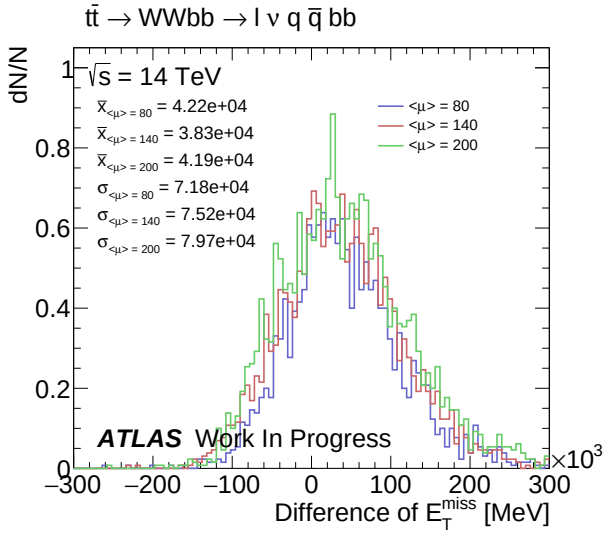


(a) MET from all $t\bar{t}$ files.

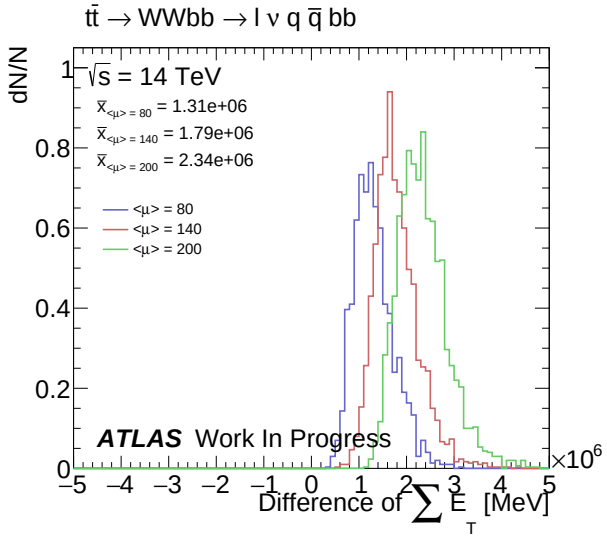
(b) Sumet from all $t\bar{t}$ files.

Figure 4.3: The distribution of the MET and sumet from three of the $t\bar{t}$ datasets reconstructed from particle flow objects. Notice the sumet in (b) increases as a function of increasing pile-up.

As mentioned earlier, $\sum E_T$ can tell you about how much energy needs to be removed in the event. Using the quantaties obtained from truth particles, we can see how much energy we need to remove on average from our events in Figure 4.4 by taking the difference of the reconstructed and truth values of E_T^{miss} and $\sum E_T$.



(a) Difference of E_T^{miss} and truth E_T^{miss} from all $t\bar{t}$ files.



(b) Difference of $\sum E_T$ and truth $\sum E_T$ from all $t\bar{t}$ files.

Figure 4.4: The distribution of the difference of reconstructed E_T^{miss} and $\sum E_T$ and truth values for three of the $t\bar{t}$ datasets. For E_T^{miss} in Figure (a) the difference gets worse when pile-up is high. In Figure (b) the difference increases as a function of increasing pile-up. This shows how much energy we need to remove from our event, eg. for r7799 we need to remove about 2.3 TeV!

The two quantities E_T^{miss} and $\sum E_T$ are nice measurements of goodness. But they do not tell you anything about how well the direction of E_T^{miss} is reconstructed. To determine how well the direction of E_T^{miss} is reconstructed, we use the parallel and perpendicular components. See Figure 4.5 for the definition. We want our E_T^{miss} vector to be as close to the truth E_T^{miss} vector as possible. Therefore the difference of the parallel component should be zero, and the perpendicular component should also be zero.

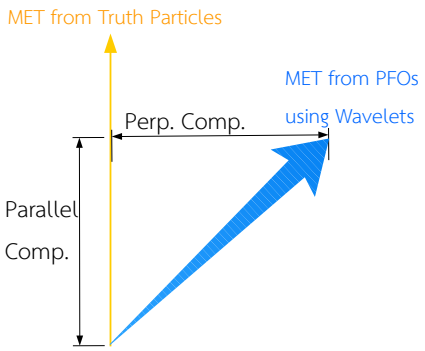
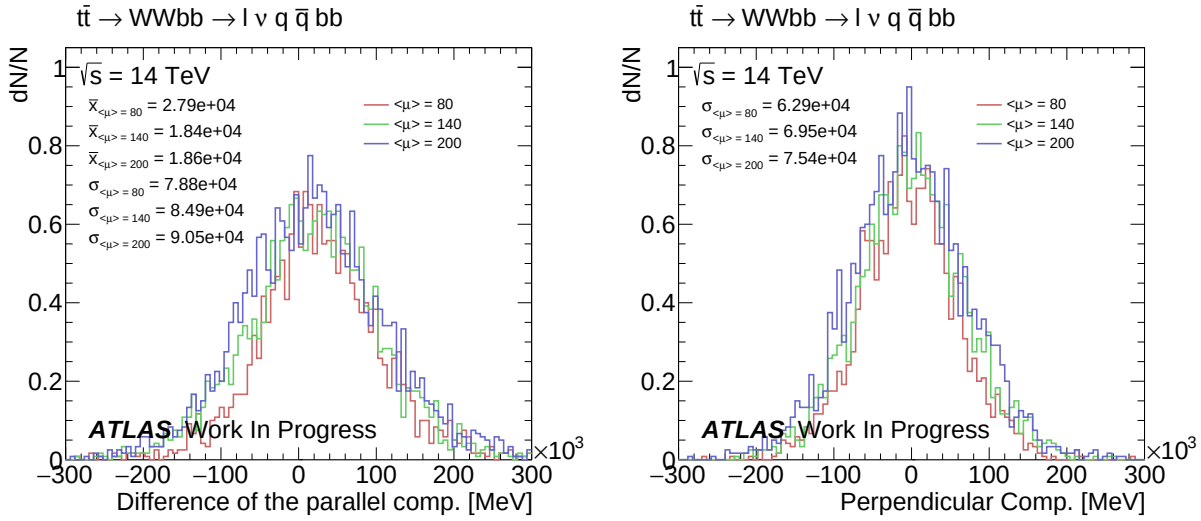


Figure 4.5: Illustration of the E_T^{miss} vectors. The difference of the truth MET and reconstructed MET can be measured in terms of a perpendicular and parallel component.

The equations used for calculating the parallel and perpendicular components to the truth MET vector are given by normal vector calculus and is shown in Appendix A.2. The measurements of the parallel and perpendicular components are in Figure 4.6. The difference in the perpendicular components tends to be influenced by pile-up. The parallel component can be influenced by pile-up, but it can also include the effect of calorimeter response. The effect of the calorimeter response would be constant for different values of pile-up.



(a) Difference of the parallel component to truth.

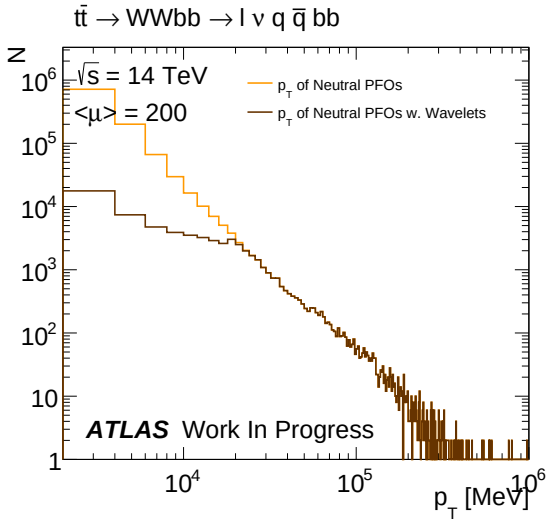
(b) Perpendicular component of MET.

Figure 4.6: The goodness measurements for three $t\bar{t}$ datasets.

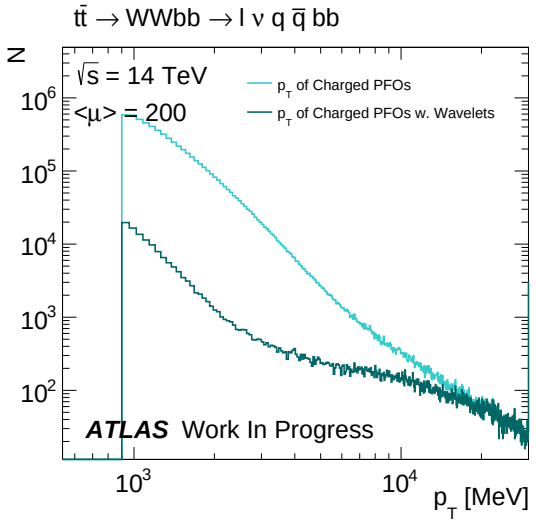
Now we have examined the MET measurements for the $t\bar{t}$ datasets using *only* particle flow objects. But there is definitely a need for improvements, remember there was a total of 2.3 TeV too much energy in the dataset with a mean pile-up of 200! In Section 5.1 the results are shown for the same quantities using wavelet denoising. But before doing the full analysis, we should have a look at the particle flow objects themselves, which need to be cleaned for pile-up.

4.2.2 Distribution of Particle Flow Ojects (PFOs)

In this section the distribution of momentum in the particle flow objects (PFOs) are reviewed. The result of the particle flow algorithm is a list of reconstructed charged and neutral particles. An example run of wavelet flat denoising with a cut of 4 GeV was run on the particle flow objects, just to show that after using wavelets, the total momentum of both the charged and neutral PFOs are decreased, see Figure 4.7. Note that the particles with high transverse momenta are not removed by the methods of wavelets. This is because it is more likely that high p_T particles are from the primary event, than from some secondary event.



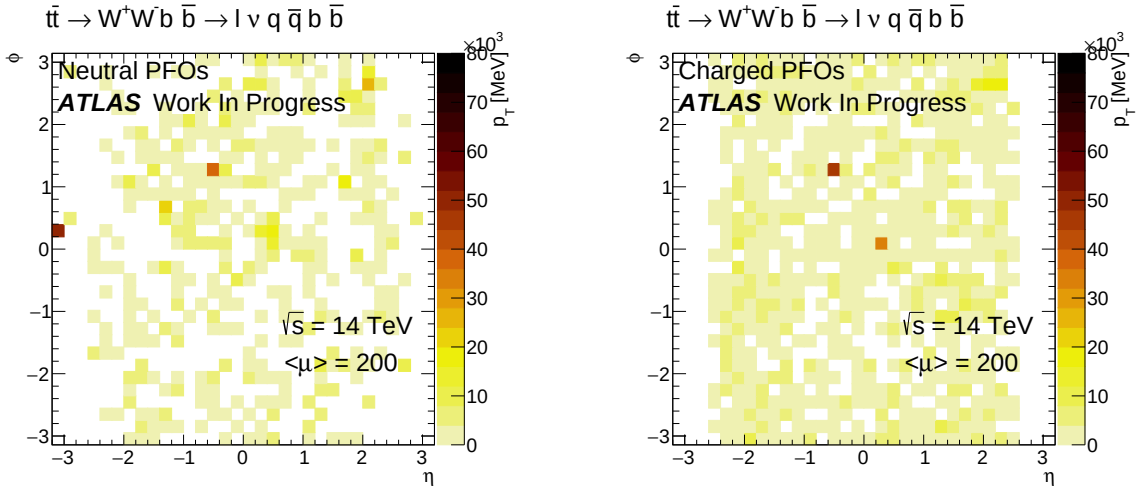
(a) p_T of neutral particle flow objects with and without wavelet flat denoising with a cut of 4.5 GeV.



(b) p_T of charged particle flow objects with and without wavelet flat denoising with a cut of 4.5 GeV.

Figure 4.7: The distribution of p_T of particle flow objects in a $t\bar{t}$ dataset with mean pile-up of 200. The charged PFOs are only reconstructed down to 400 MeV.

In Figure 4.8 the position of the particle flow objects are shown. The charged PFOs does not exceed a position of $\eta = 2.5$, because the charged PFOs are reconstructed from both the calorimeter and inner detector, where the inner detector can measure particles with $|\eta| \leq 2.5$. The neutral particle flow objects can be collected from only the calorimeter. In Figure 4.8 it is also shown that some particles leave a signal in both the charged and neutral particles, where sometimes the particle will e.g. only leave a signal in the charge particle container.



(a) p_T and positions of neutral particle flow objects.

(b) p_T and positions of charged particle flow objects.

Figure 4.8: The (ϕ, η) distribution of the momenta of the particle flow objects in one $t\bar{t}$ event with a pile-up of 200.

4.2.3 *Truth Particles*

The truth particles are used for reconstructing the true missing transverse energy. The momenta of the truth particles is shown in Figure 4.9.

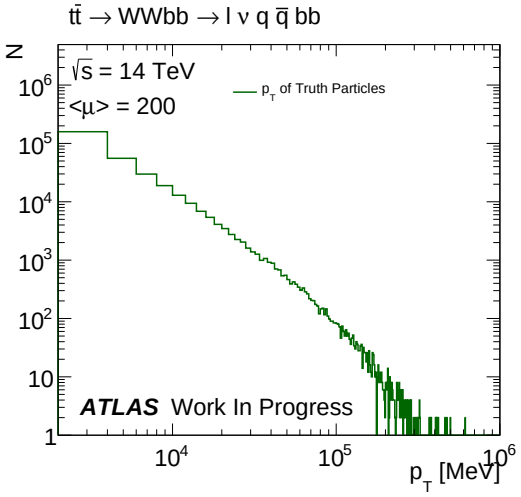
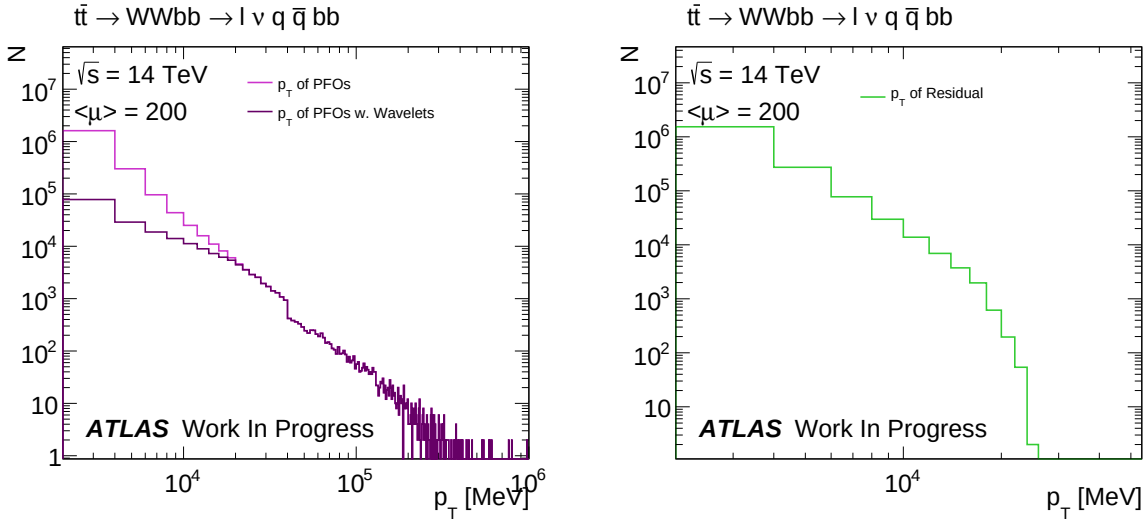


Figure 4.9: p_T of truth particles in the $t\bar{t}$ dataset with a mean pile-up of 200.

The goal is to remove particle flow objects, such that the remaining particle flow objects more closely correspond to the truth particles.

The transverse momenta of the particle flow objects with and without usage of the method of wavelets is shown in Figure 4.10. Again you can see that the low p_T particles are preferably removed, and the high p_T particles remain. The momenta of the removed particles is shown in Figure 4.10. The removed particles do not have p_T above 12 GeV.

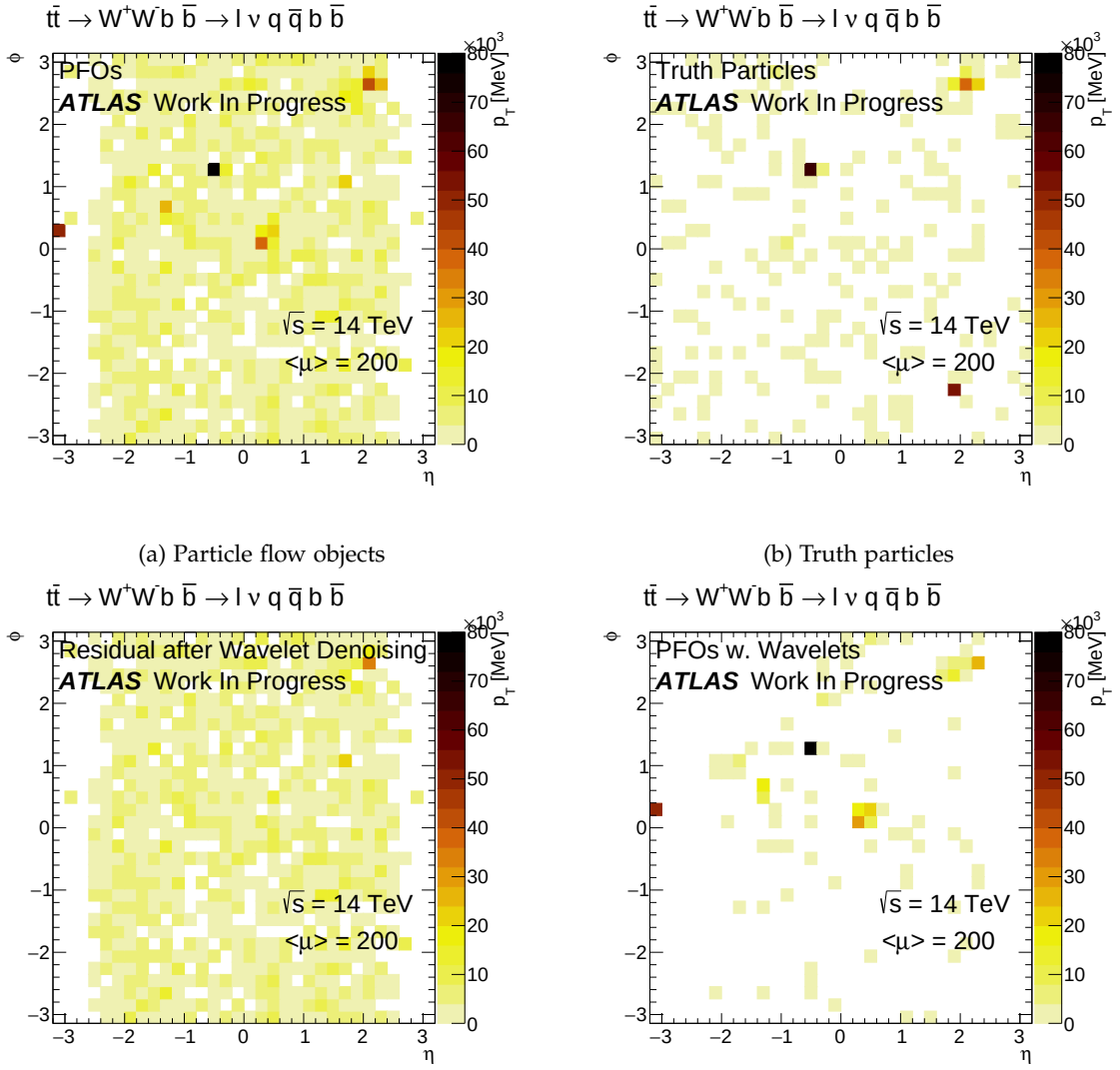


(a) p_T of every PFO in 1 event with and without wavelet flat denoising with a cut of 4.5 GeV.

(b) p_T of every PFO in 1 event and the residual of the PFOs after wavelet flat denoising with a cut of 4.5 GeV.

Figure 4.10: The distribution of p_T of particle flow objects in a single $t\bar{t}$ event with a pile-up of 200.

When removing particles from the events, it is also important that it is the right particles we remove, such that the positioning of the remaining momenta is correct. The positioning for the particle flow objects for a single event is in Figure 4.11 (a), and the corresponding truth event is in Figure 4.11 (b). Using wavelets, some of the particle flow objects will be removed, and the remaining particles should then correspond to the truth image.



(a) Particle flow objects

(b) Truth particles

(c) Residual

(d) Cleaned particle flow objects

Figure 4.11: These 2D histograms shows the distribution of particles in different cases. These plots give an overview of the problem at hand. In Figure (a) the position and energies of the particle flow objects are shown. The pile-up particles are also included, and the picture has not been denoised yet. In Figure (b) the truth particles are shown. In Figure (c) you can see the removed pile-up particles from the PFOs. These particles were removed using a simple example wavelet method named Flat Denoising, with a cut of 4.5 GeV, see Section 5.1. In Figure (d) you can see the final result. This is the PFOs with wavelet cleaning, and hopefully without pile-up particles.

4.2.4 Wavelet Coefficients

The dataset used for this analysis is the $t\bar{t}$ dataset with a mean pile-up of 200. The distribution of the wavelet coefficients for a transformation of particle flow objects is shown in Figure 4.12. In the next few plots you will encounter the term *frequency band*, which corresponds to the set of coefficients that all have the same angular scale. The new naming is shown in Table 4.4.

Table 4.4: Frequency Bands.

Image at (m_ϕ, m_η)	Frequency Band
(0,0)	0
(0,1)	1
(0,2)	2
(0,3)	3
(0,4)	4
(0,5)	5
(0,6)	6
(1,0)	7
...	...
(6,6)	48

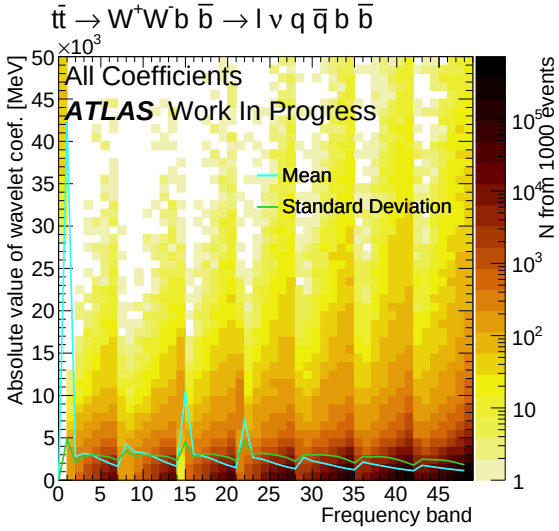


Figure 4.12: Values of the wavelet coefficients for 100 $t\bar{t}$ events with mean pile-up of 200. The mean and standard deviations for each frequency band is shown.

In Figure 4.12 that mean and standard deviation at the images with frequency bands of 14 ($m_y = 2, m_\phi = 0$) and 21 ($m_y = 3, m_\phi = 0$) are much larger than the other images. If you look at the images of the wavelet coefficients, (full image in Figure 4.18), you can see that this is in the differences over η. Frequency bands 14 and 21 show that there are more particles centred in η than at $|\eta| > 2$. This pattern will still be in the coefficients even if they are scaled down or up. Note that for frequency band 0 in Figure 4.13 the values are always over $\tilde{3}0$ GeV. This is the mean value of the input image. This mean value will be scaled down by a factor of $1/ < \mu >$ to estimate the new mean of the

cleaned image. The scaling of the mean by $1/ < \mu >$ is used for all wavelet methods. In Figure 4.13 all the values of the coefficients are shown.

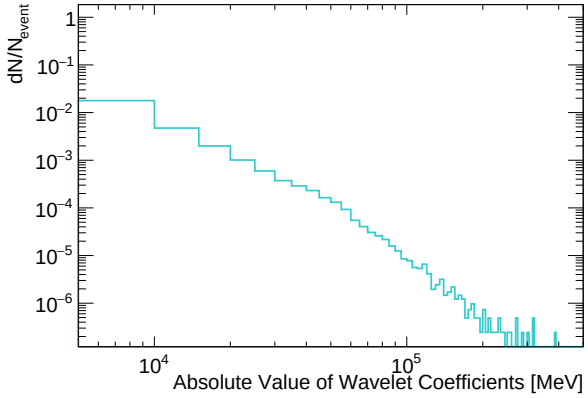


Figure 4.13: Values of the wavelet coefficients for 1000 $t\bar{t}$ events with mean pile-up of 200.

To determine a reasonable choice of cut of the wavelet coefficients, the coefficient values are considered, as shown in Figure 4.13. Most of the coefficients has a value less than 10 GeV. These are candidates for coefficients for pile-up. If a wavelet coefficient has a high value, it means that it is significant. If it has a low value, it can be interpreted as noise. The problem now is to determine at which value do we separate the noise from the signal. Experiments for this is in Section 5.1.

The track scaling method scales the wavelet coefficients according to the ratio of coefficients from all tracks to tracks from the primary vertex. The smoothing coefficient is scaled by $1/ < \mu >$ and all other wavelet coefficients are scaled by the track ratio. Since the coefficients from tracks are approximate to the coefficients of the full event, this method aims to preferably reduce the values of coefficients in regions without signal.

The idea is to scale down the coefficients with high probability of being from pile-up, and scale up the coefficients with high probability of being from the signal from the primary vertex. The result from running the track scaling method on 1000 $t\bar{t}$ events with a mean pile-up of 200 is in Figure 4.14.

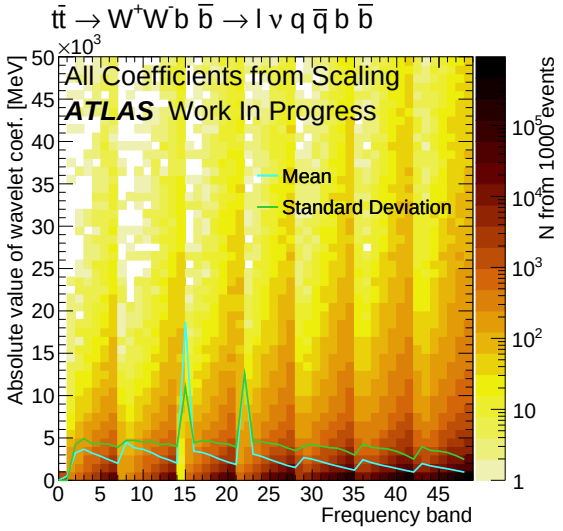


Figure 4.14: Values of the wavelet coefficients after using track scaling for 1000 $t\bar{t}$ events with mean pile-up of 200.

The distribution of coefficients in Figure 4.14 from track scaling looks almost no different than the original coefficients in Figure 4.12. But this is only because that the changes are hard to see in those figures. The important point is, that the track scaling method did not severely change the high value wavelet coefficients. In Figure 4.15 the changes are easier to see. When comparing Figure 4.15 and 4.13 it is clear that some coefficients has been up- or down-scaled.

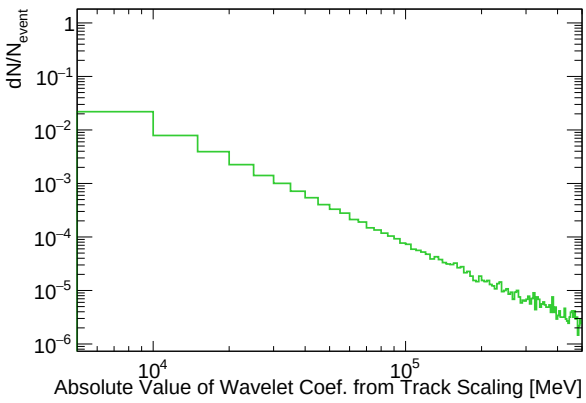


Figure 4.15: Values of the wavelet coefficients for 100 $t\bar{t}$ events with mean pile-up of 200.

The last scaling method used in this paper is the maximum entropy method. The maximum entropy method scales the coefficients according to a model coefficient value and an estimation of the standard deviation of the wavelet coefficients of the noise in the image. The effect of the model value can be scaled by a constant α. This constant takes different values for different coefficient numbers, as described in Algorithm 6. The resulting α constant for 1000 $t\bar{t}$ events is shown in Figure 4.16. The α constant was set to be between 0 and 1. A high value of α will scale the wavelet coefficients largely based on the model coefficient, and a low value of α will either not change the

wavelet coefficient much, or it will down-scale it. The down-scaling depends on the value of the coefficient, eg. a very low coefficient with a low α value will likely be set to zero. But if it is high enough, it will only be down-scaled by a little fraction. The resulting wavelet coefficients scaled using the maximum entropy method is in Figure 4.17.

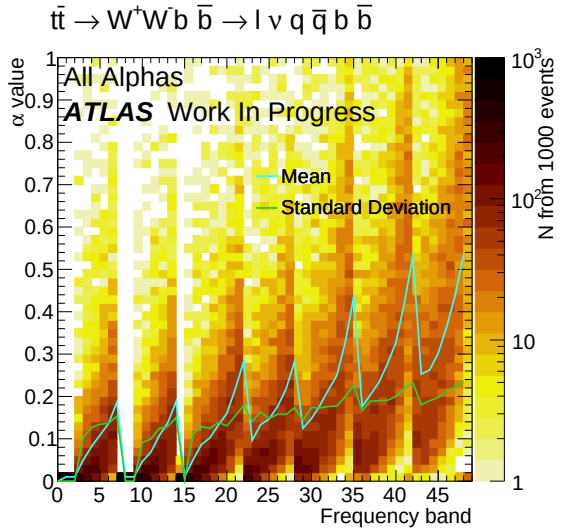


Figure 4.16: Values of α constants used for the MEM method, using the $t\bar{t}$ dataset with mean pile-up of 200.

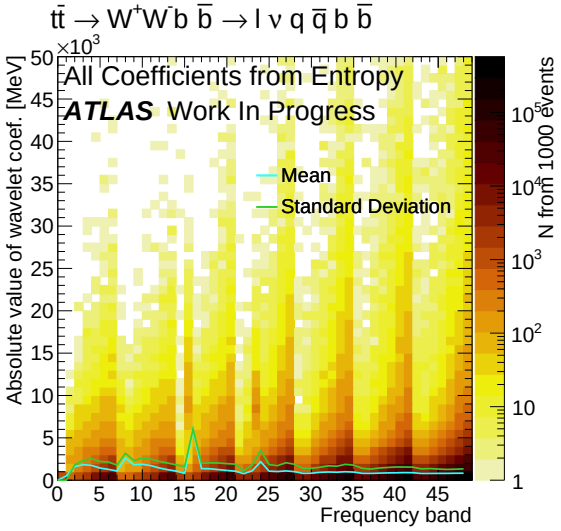


Figure 4.17: Values of the wavelet coefficients after using the MEM method for the $t\bar{t}$ dataset with mean pile-up of 200.

The resulting coefficients from the maximum entropy method is also shown in Figure 4.18. A maximum value for a wavelet coefficient to get was set to 5000 GeV, to make the computing time a bit smaller.

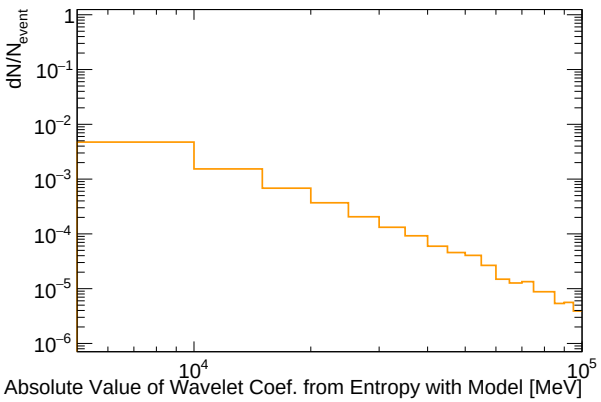


Figure 4.18: Values of the wavelet coefficients from the MEM method for the $t\bar{t}$ dataset with mean pile-up of 200. The peak at 10^5 MeV was user set. This was set as a maximum value of which a wavelet coefficient could have, to limit computations. This user value was only set for this plot, and for further analysis no maximum is set.

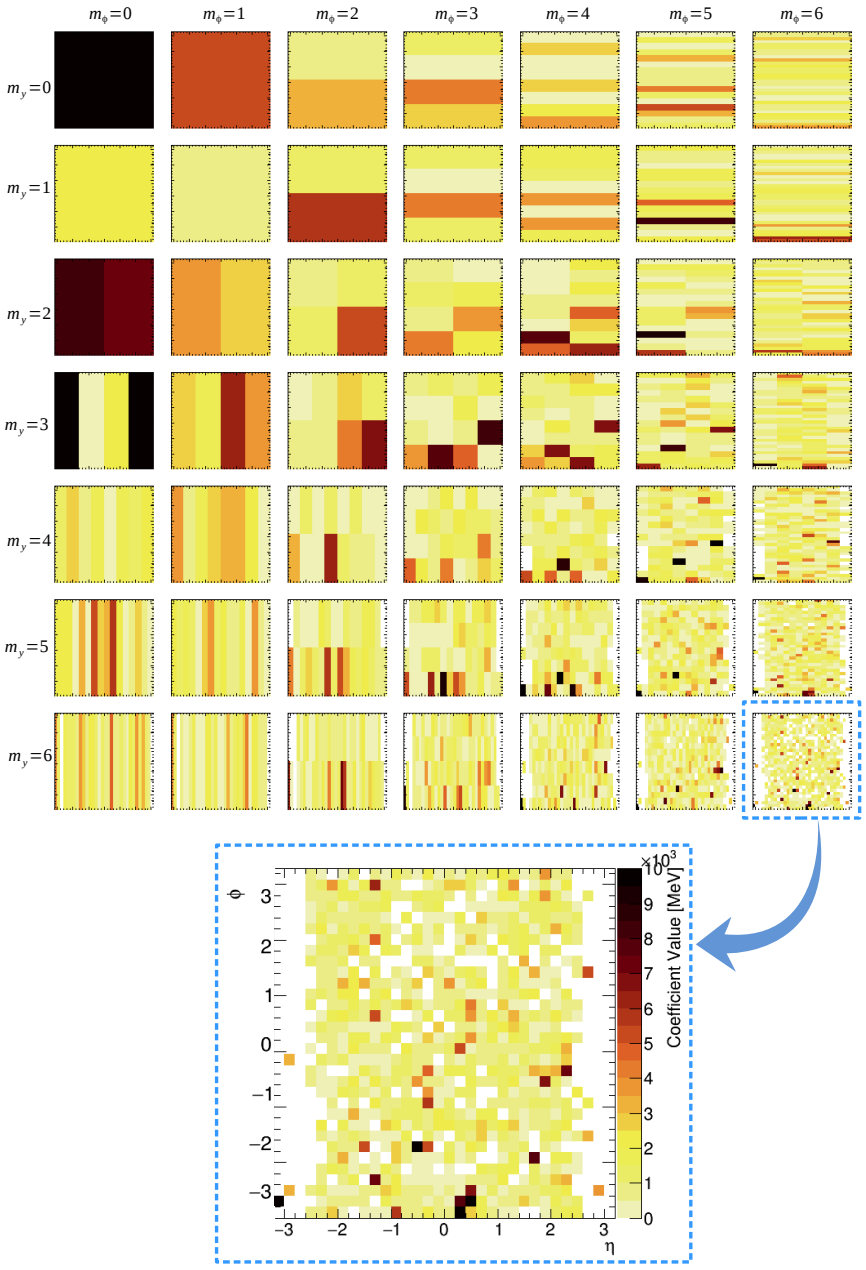


Figure 4.18: Visualization of all coefficients in a wavelet transform. At coefficient at level $(0,0)$ is called the smoothing coefficient. At higher level of m_y and m_ϕ there are more wavelet coefficients which result in a better resolution of the transformed image. The axis of the images are all the same, and can be seen in the enhanced image.

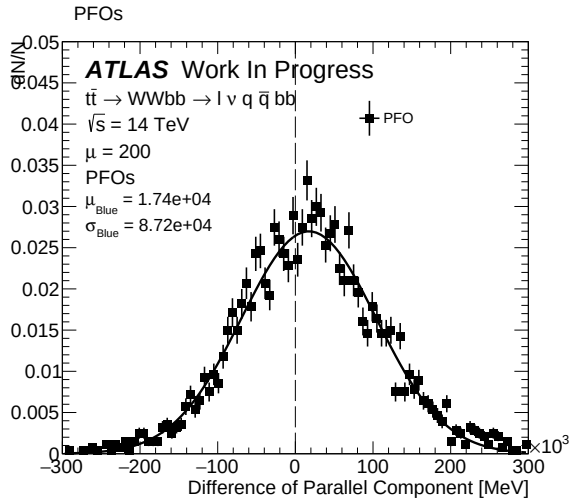
Results

5.1 Using Wavelet Methods

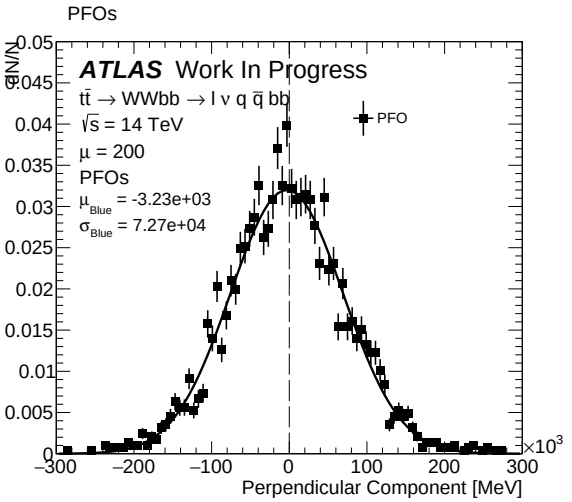
This section describes the different experiments with wavelets. The wavelet methods were tested on the $t\bar{t}$ datasets, using different settings. The methods used are flat denoising, track filtering, track scaling and maximum entropy.

5.1.1 Setup for Wavelet Tests

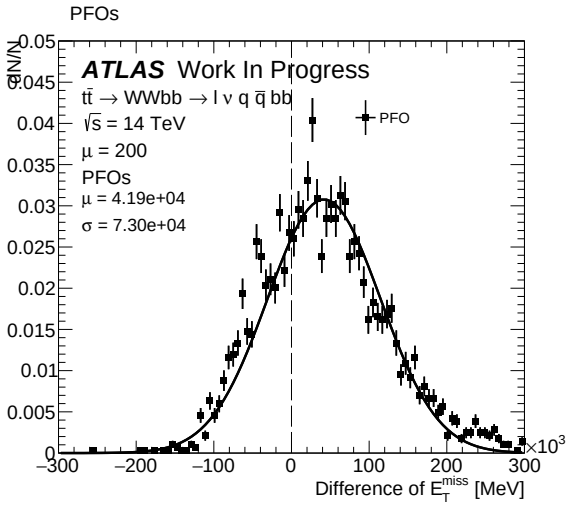
Before we test the methods of wavelet cuts, lets have a look at the initial problem. For this analysis the $t\bar{t}$ dataset with a mean pile-up of 200 is used. This dataset is the one with the highest pile-up used in this thesis. If we can remove the pile-up for this worst case, we should also be able to do it at lower (but still high) pile-up. The initial case is shown in Figure 5.1. As described in section 4.1 we use the difference of the parallel component of the MET vectors and the perpendicular component of the MET vector as metrics. We expect the perpendicular component to by symmetrical around zero, but the parallel components difference to truth may be biased due to over- or under-estimating of the missing transverse energy. Moreover the difference of $\sum E_T$ should be centred at zero, i.e. close to the true values. For some of the tests performed several parameters were tested, and they will be summarized. The measurables we need to improve on are the general width and means of the fits in Figure 5.1. If we can get the widths and means closer to zero, it means that the measurements of missing transverse energy has been improved.



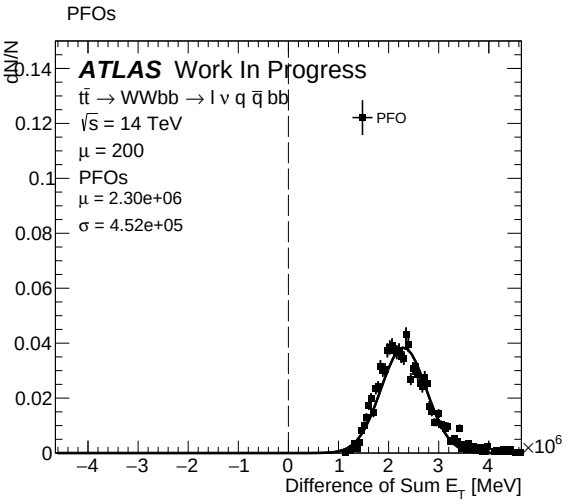
(a) Difference of the truth parallel component and the reconstructed parallel component.



(b) Difference of the truth parallel component and the reconstructed perpendicular component.



(c) Difference of the truth E_T^{miss} component and the reconstructed E_T^{miss}.



(d) Difference of the truth sum E_T component and the reconstructed sum E_T.

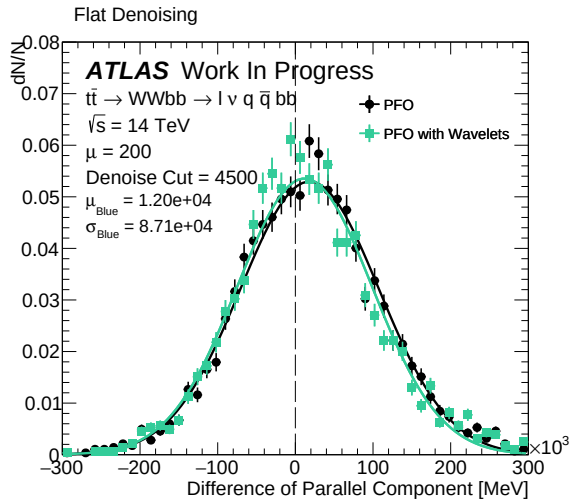
Figure 5.1: The reconstructed missing transverse energy components using all particle flow objects without removal of pile-up. In Figure (a) the difference of the parallel component of the reconstructed and true MET is shown. In Figure (b) the difference of the perpendicular component of the reconstructed and true MET is shown. In Figure (c) the difference of the true E_T^{miss} and the reconstructed E_T^{miss} is shown. In Figure (d) the difference of sum E_T truth and reconstructed sum E_T is shown.

5.1.2 Flat Denoising

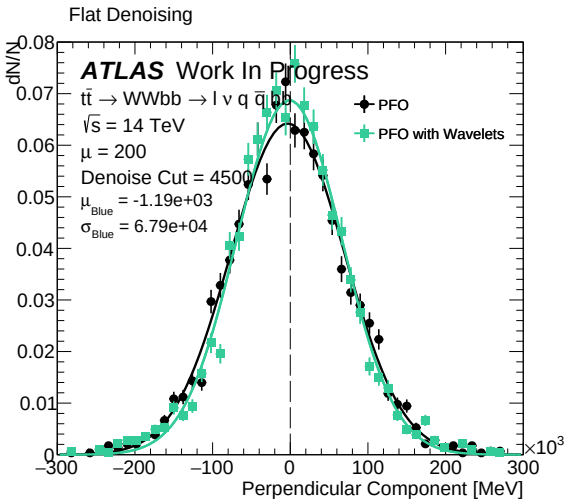
We need to determine the scale of which we need to make the cut, such that we can remove some of the pile-up particles, and also keep other particles of the event. Here different values for the cut was used, and the result for one of the cuts is in Figure 5.2. A summary of these experiments is in Figure 5.3. The values tested for the flat cut are,

$$Cut = 0.0, 0.5, 1.0, ..., 10.0 \text{ GeV} \tag{5.1}$$

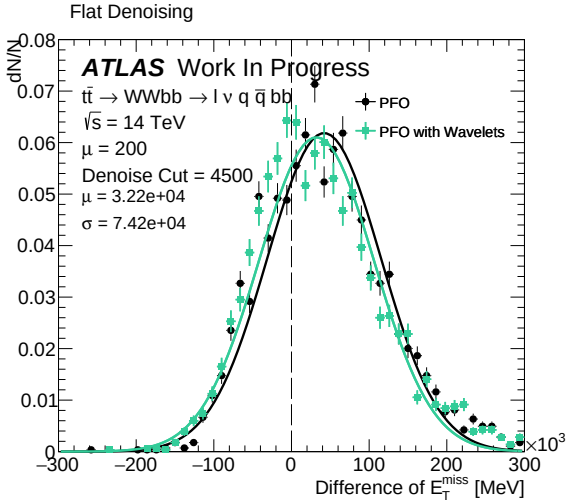
and the results were made for all of the metrics for missing transverse energy. In Figure 5.3 (a) it is shown that for a higher cut the mean of the parallel component of the difference of the parallel component decreases. The best value is when the mean is close to zero, here it would be 5.0 - 6.0 GeV. The most noticeable measurable in the flat denoising case is the fit mean of $\sum E_T$. It is clear that the preferable cut value here is 4.0 - 5.0 GeV.



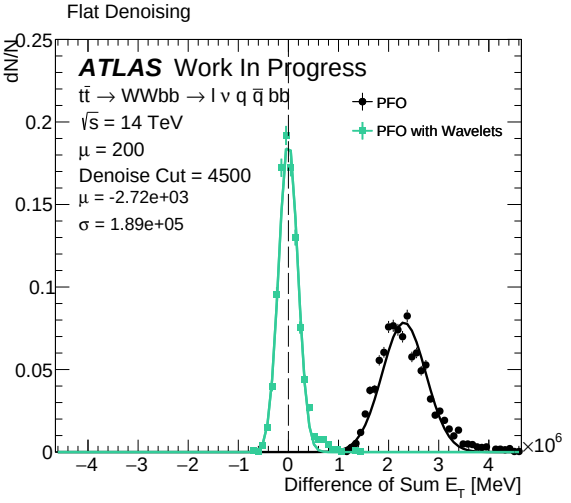
(a) Difference of the truth parallel component and the reconstructed parallel component.



(b) Difference of the truth parallel component and the reconstructed perpendicular component.



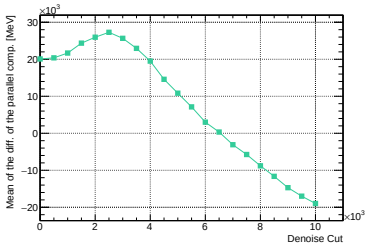
(c) Difference of the truth E_T^{miss} component and the reconstructed E_T^{miss}.



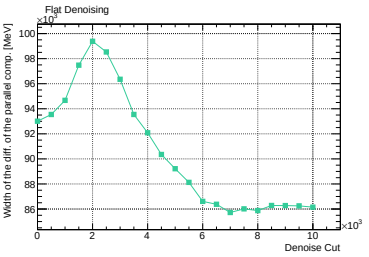
(d) Difference of the truth sum E_T component and the reconstructed sum E_T.

Figure 5.2: Results of a flat denoising cut of 4.5 GeV in the wavelet coefficients. In Figure (a) the difference of the parallel component of the reconstructed and true MET is shown. In Figure (b) the difference of the perpendicular component of the reconstructed and true MET is shown. In Figure (c) the difference of the true E_T^{miss} and the reconstructed E_T^{miss} is shown. In Figure (d) the difference of sum E_T truth and reconstructed sum E_T is shown.

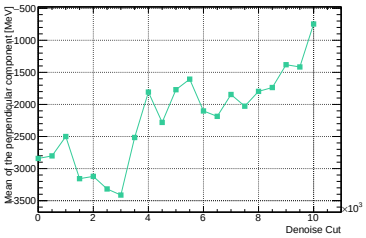
Figure 5.3: Fit values for the experiments made for the flat denoising of particle flow objects. A consistently good denoise cut seems to be of 5 GeV in this case, but eg. 4 GeV is not a bad choice either.



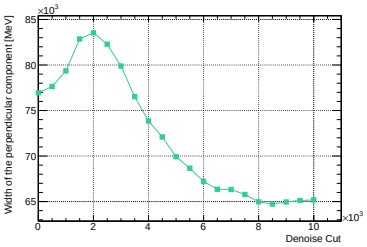
(a) Fit mean of parallel.



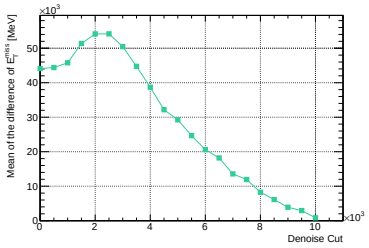
(b) Fit width of parallel.



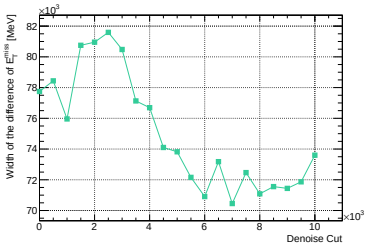
(c) Fit mean of perp.



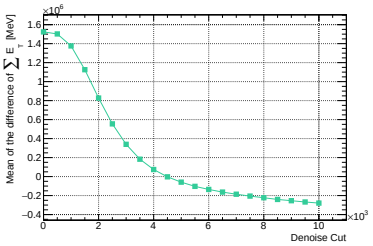
(d) Fit width of perp.



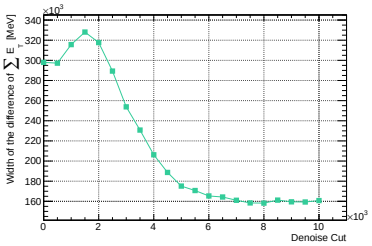
(e) Fit mean of E_T^{miss}.



(f) Fit width of E_T^{miss}.



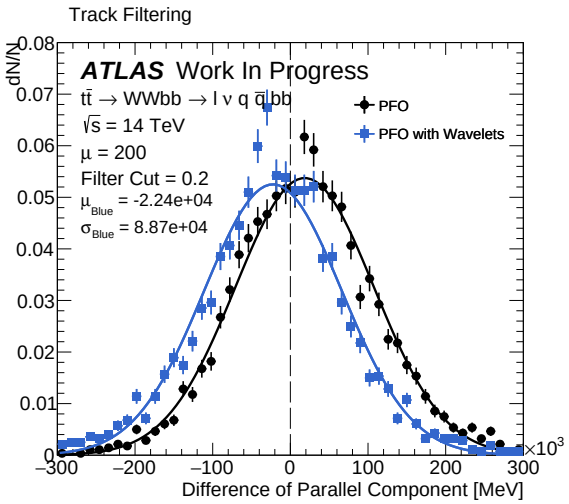
(g) Fit mean of sum E_T.



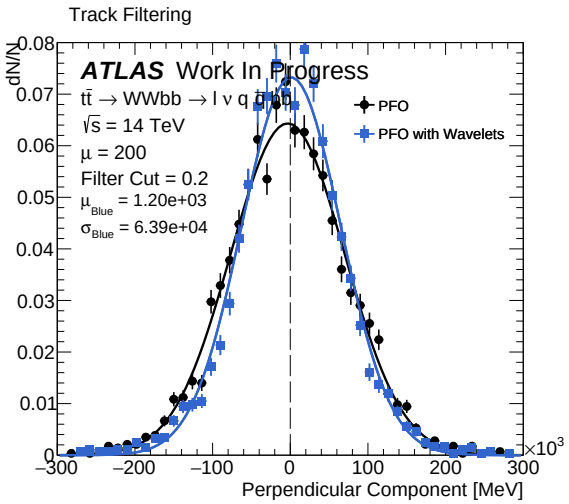
(h) Fit width of sum E_T

5.1.3 Track Filtering

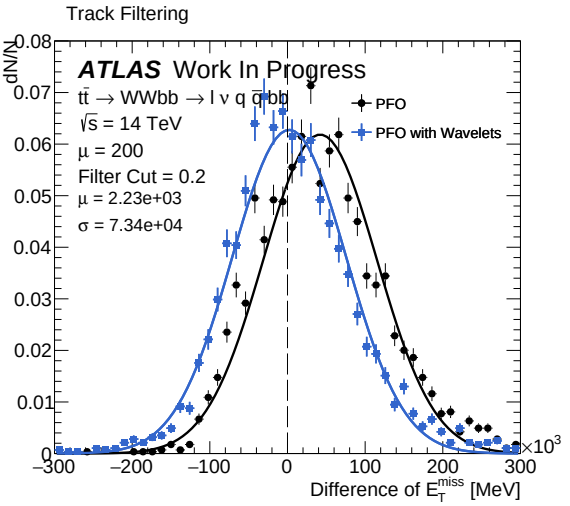
In the wavelet method track filtering a ratio needs to be determined by the user. This is to make a cut corresponding to Algorithm 2. The variable ratio is set to be between 0 and 1.0. The best choice is in Figure 5.4 and a summary of the choices is in Figure 5.5.



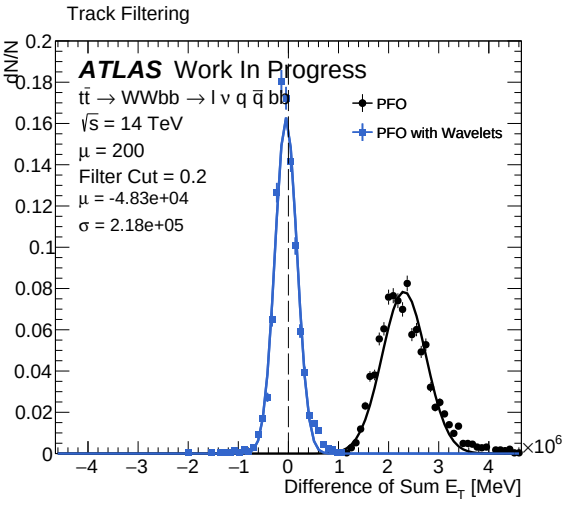
(a) Difference of the truth parallel component and the reconstructed parallel component.



(b) Difference of the truth perpendicular component and the reconstructed perpendicular component.



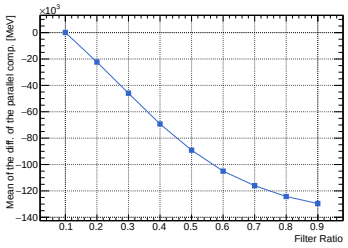
(c) Difference of the truth E_T^{miss} and the reconstructed E_T^{miss} component.



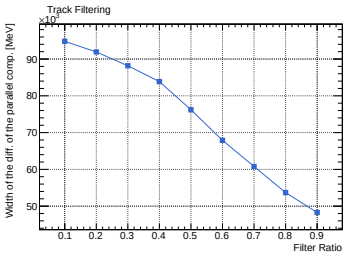
(d) Difference of the truth sum E_T and the reconstructed sum E_T.

Figure 5.4: Results of a track cut of 0.2 in the wavelet coefficients. In Figure (a) the difference of the parallel component of the reconstructed and true MET is shown.In Figure (b) the difference of the perpendicular component of the reconstructed and true MET is shown. In Figure (c) the difference of the true E_T^{miss} and the reconstructed E_T^{miss} is shown. In Figure (d) the difference of sum E_T truth and reconstructed sum E_T is shown.

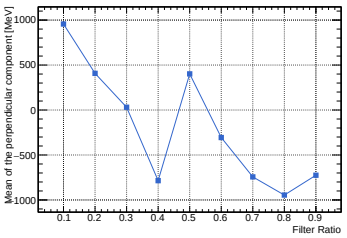
Figure 5.5: Fit results for track filtering. The variable is the filter ratio, and it varies from 0.0 to 1.0 by 0.1.



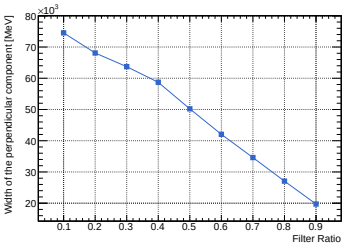
(a) Fit mean of parallel.



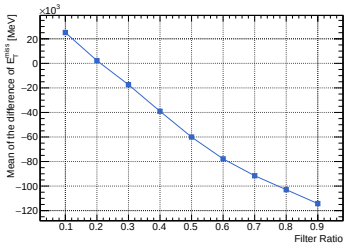
(b) Fit width of parallel.



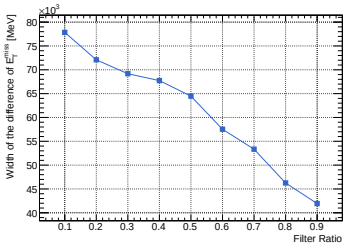
(c) Fit mean of perp.



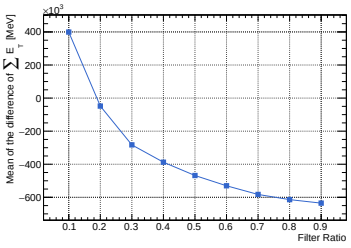
(d) Fit width of perp.



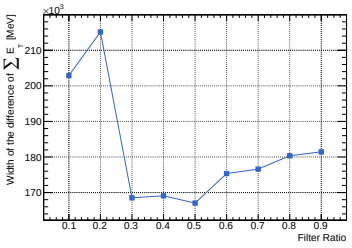
(e) Fit mean of E_T^{miss}.



(f) Fit width of E_T^{miss}.



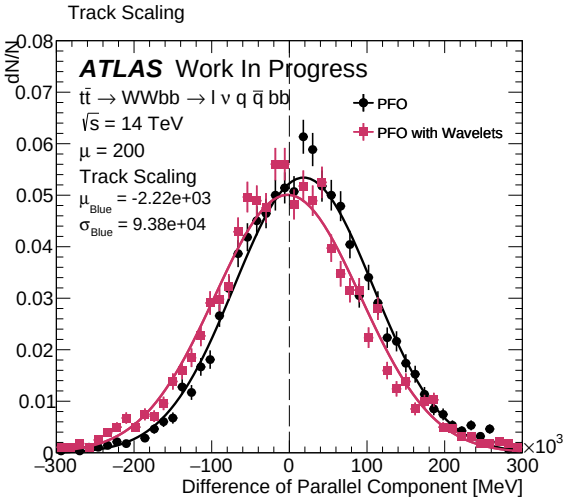
(g) Fit mean of E_T.



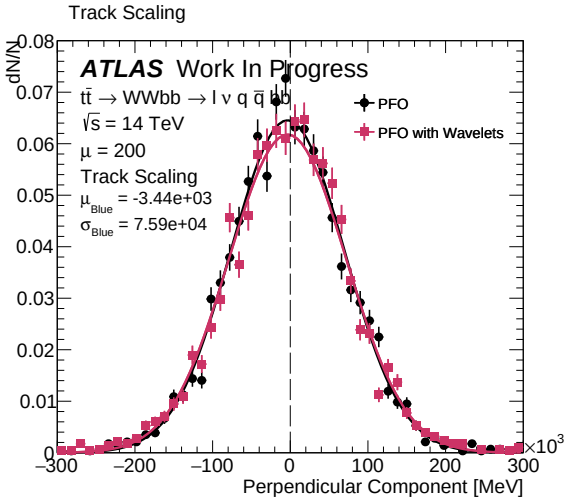
(h) Fit width of sum E_T.

5.1.4 Track Scaling

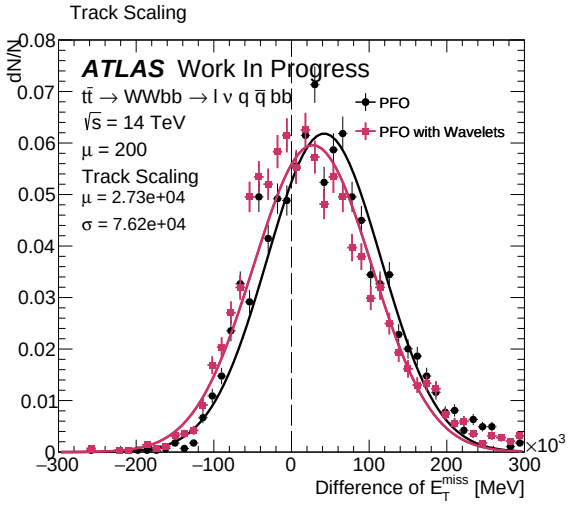
The method of track scaling has no need for any user set values. Therefore there is only one case of track scaling. The track scaling method scales the wavelet coefficients by a ratio of the wavelet coefficients of all tracks and tracks from the primary vertex. The track scaling method removed particles such that the resulting difference in $\sum E_T$ is centred at 404 GeV.



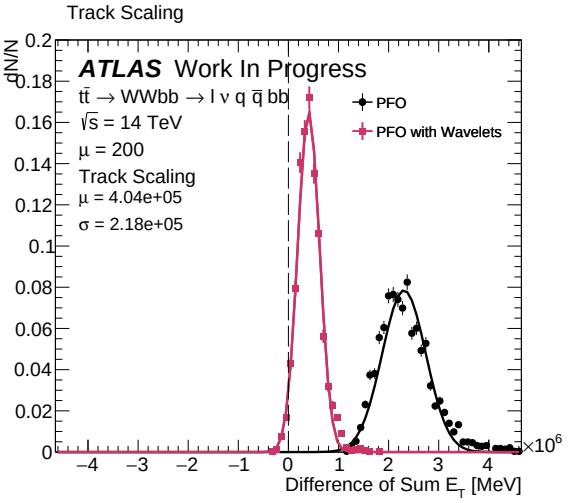
(a) Difference of the truth parallel component and the reconstructed parallel component.



(b) Difference of the truth perpendicular component and the reconstructed perpendicular component.



(c) Difference of the truth E_T^{miss} and the reconstructed E_T^{miss} component.

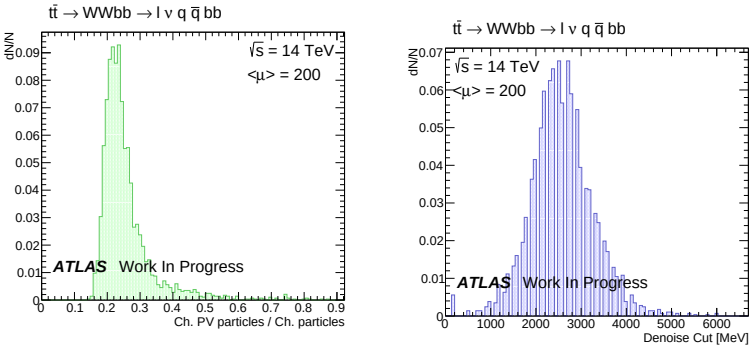


(d) Difference of the truth sum E_T and the reconstructed sum E_T.

Figure 5.6: Results of the track scaling method. In Figure (a) the difference of the parallel component of the reconstructed and true MET is shown.In Figure (b) the difference of the perpendicular component of the reconstructed and true MET is shown. In Figure (c) the difference of the true E_T^{miss} and the reconstructed E_T^{miss} is shown. In Figure (d) the difference of sum E_T truth and reconstructed sum E_T is shown.

5.1.5 Ratio Denoising

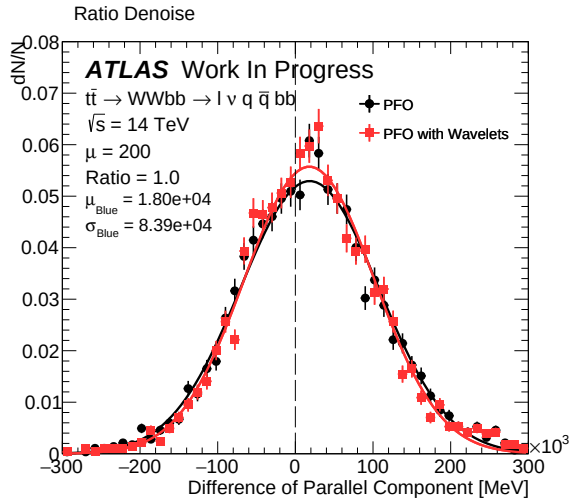
When using the ratio denoising method we make cuts in the wavelet coefficients until the ratio of removed neutral particles is comparable to the ratio of the charged particles from primary vertex and from pile-up. In Figure 5.7 (a) the ratio of the number of charged particles from primary vertex and secondary vertices is shown. This shows that accordingly to the information from the track particles, on average only 20 % of the charged particles are from the primary vertex. Using the ratio denoise method, the neutral particles will therefore also be removed until there are 20 % neutral particles left. In Figure 5.7 (b) the resulting flat cuts in the coefficients of the neutral particles is shown. A maximum value was set, such that we can not perform a cut higher than 10 GeV, but most of the time a cut of 2.5 GeV seems sufficient. Since the cuts are not as hard as the 4.5 GeV flat cut in the flat denoising method, the amount of energy removed from the events are not as much as in the flat denoising case. The cuts therefore need to be higher and further experiments for this is in Appendix A.3.



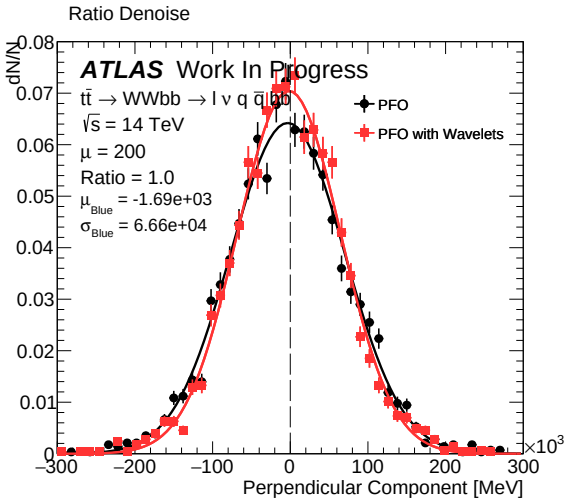
(a) Ratio of number of tracks from PV and all tracks.

(b) Ratio Denoise Cut

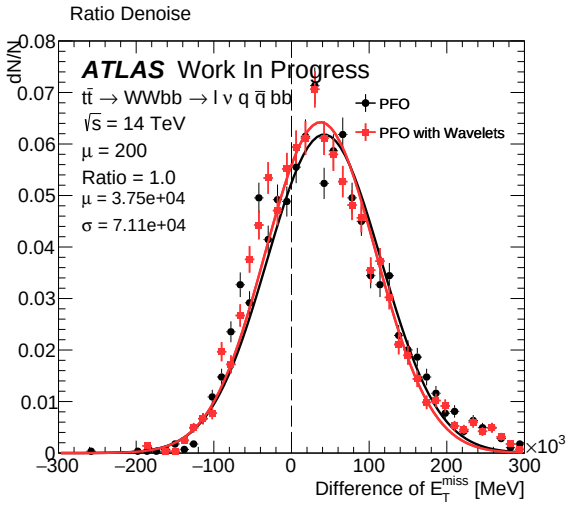
Figure 5.7: The ratio of the track particles from the primary vertex and all track particles and the resulting denoise cuts using these ratios.



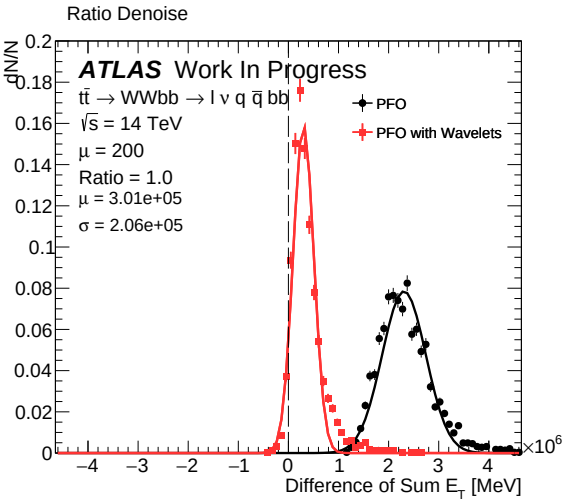
(a) Difference of the truth parallel component and the reconstructed parallel component.



(b) Difference of the truth perpendicular component and the reconstructed perpendicular component.



(c) Difference of the truth E_T^{miss} and the reconstructed E_T^{miss} component.



(d) Difference of the truth sum E_T and the reconstructed sum E_T.

Figure 5.8: Results of a ratio denoise cut in the wavelet coefficients. In Figure (a) the difference of the parallel component of the reconstructed and true MET is shown.In Figure (b) the difference of the perpendicular component of the reconstructed and true MET is shown. In Figure (c) the difference of the true E_T^{miss} and the reconstructed E_T^{miss} is shown. In Figure (d) the difference of sum E_T truth and reconstructed sum E_T is shown.

5.1.6 Maximum Entropy Method

The maximum entropy method uses the information of the input wavelet coefficient, a model coefficient and standard deviations of wavelet coefficients of pile-up particles to make a probability function of the best value of the input wavelet coefficient. The function to be minimized in the MEM function will be examined in this section.

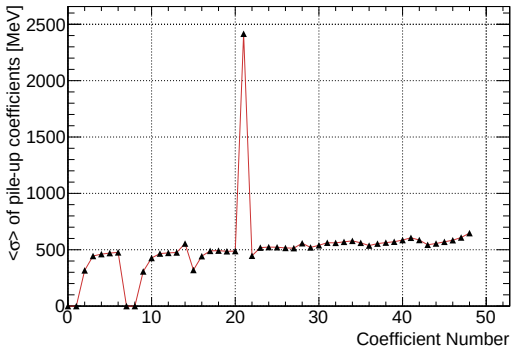
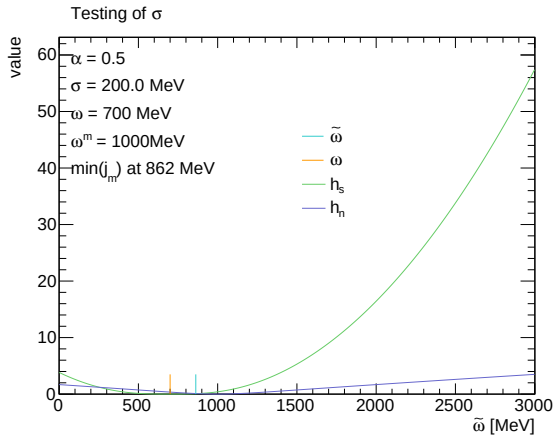


Figure 5.9: Standard Deviations of the pile-up particle coefficients at all scales.

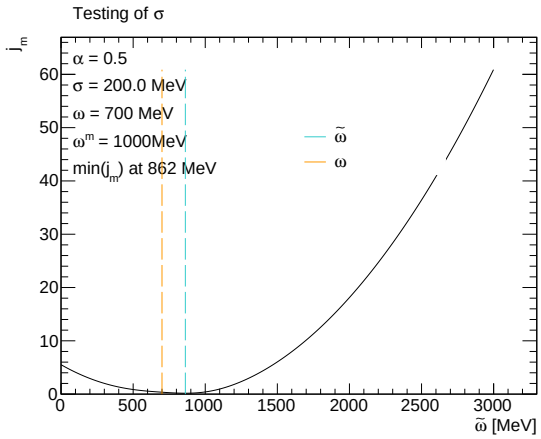
The method includes a model for noise of the signal, and the more you know about the noise, the better the denoising could be. In the track scaling and track filtering methods we used models for the signal and *not* for the noise. If the standard deviations of the the wavelet coefficients of pile-up for each frequency band is given, the MEM method can evaluate if there need to be removed more or less of the values of the wavelet coefficients. The estimated standard deviations of the wavelet coefficients of the pile-up particles for each frequency band is in Figure 5.9. Frequency band 21 ($m_y = 3, m_\phi = 0$) has a much higher standard deviation than the other frequency bands, which is caused by the difference in η at this level, as explained in Section 4.2.4.

The method needs to compute a integral and minimize a function for each coefficient.

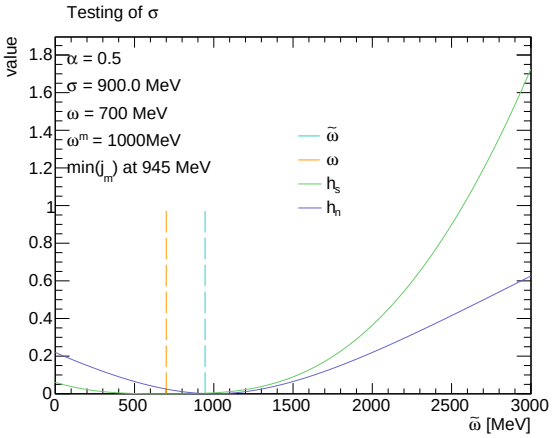
The wavelet coefficient, $\tilde{\omega}$ which minimizes the function $j_m(\tilde{\omega})$ is chosen as the optimal wavelet coefficient in the MEM method. The function j_m has several dependencies which include the initial wavelet coefficient ω, the model coefficient ω_m, σ of the wavelet coefficients for each frequency band and the constant α. The only actual free parameter is the constant α. The standard deviations of the wavelet coefficients at different frequency bands will influence the j_m function by scaling the width of the entropy contribution from the noise in the signal h_n. A standard deviation is calculated for each frequency band of the wavelet coefficient in the input image. With a low value of the standard deviation the h_n function will appear wider than for a high value of the standard deviation, see Figure 5.10. This means that for a high value of the standard deviation of a frequency band, the model coefficient will have less influence on choosing an optimal new wavelet coefficient.



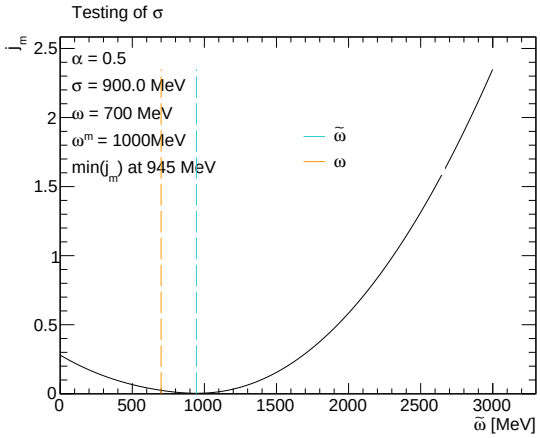
(a) The h_s and h_n functions for $\sigma = 200$ MeV



(b) The j_m function for $\sigma = 200$ MeV



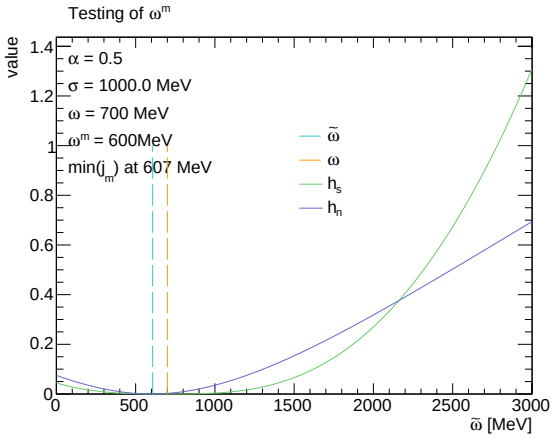
(c) The h_s and h_n functions for $\sigma = 900$ MeV



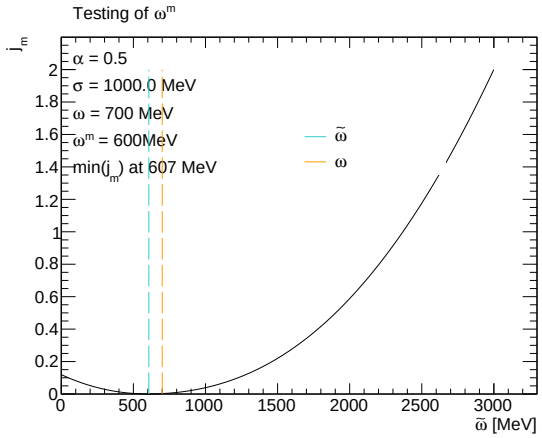
(d) The j_m function for $\sigma = 900$ MeV

Figure 5.10: The influence of a high and low value of a standard deviation for a single wavelet coefficient. The input coefficient ω has a value of 0.7 GeV and the model coefficient has a value of 1.0 GeV. In Figure (a) and (c) the functions h_n and h_s are shown for two different values of the standard deviation. In Figure (b) and (d) the resulting j_m function is shown, where the minimum of this function will determine the denoised wavelet coefficient. For a standard deviation of 200 MeV the new calculated wavelet coefficient is 862 MeV and for a standard deviation of 900 MeV the calculated wavelet coefficient is 945 MeV.

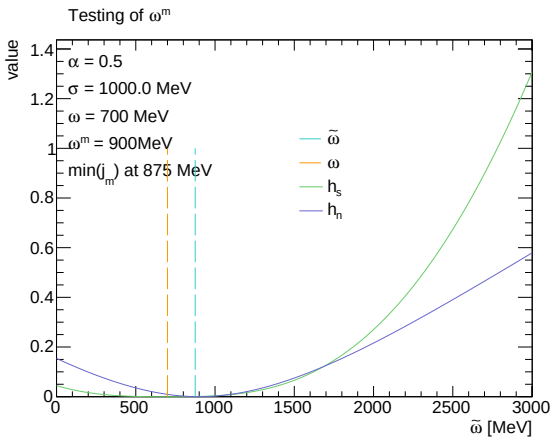
The model coefficient ω^m will influence the function j_m such that the optimal wavelet coefficient will be drawn closer to the model coefficient. Two experiments for the model are in Figure 5.11. The higher a value a model coefficient has, the higher the value of the new estimated coefficient will have.



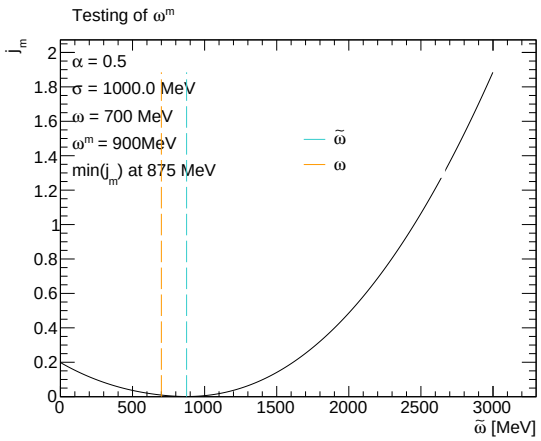
(a) The h_s and h_n functions for $\omega^m = 600$ MeV



(b) The j_m function for $\omega^m = 600$ MeV



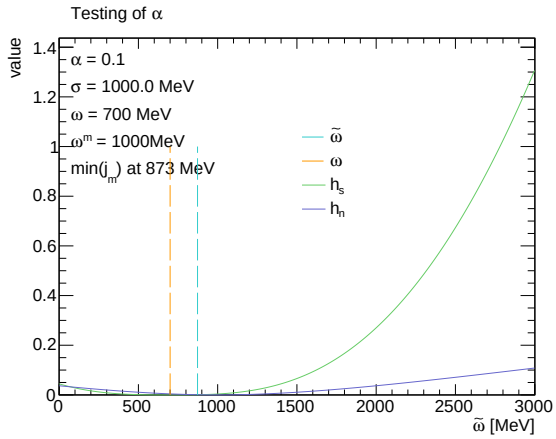
(c) The h_s and h_n functions for $\omega^m = 900$ MeV



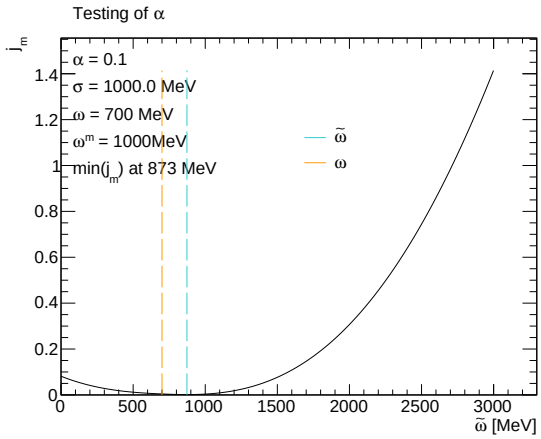
(d) The j_m function for $\omega^m = 900$ MeV

Figure 5.11: The model coefficient determines the minimum value of the h_s function. In Figure (a) and (b) the h_s and h_n functions are shown for a the model coefficient values of 0.5 GeV and 0.9 GeV. In Figure (b) and (d) the corresponding j_m functions are shown. When the model coefficient $\omega^m = 0.6$ GeV the minimum value of j_m is found at 607 MeV, and for the model coefficient $\omega^m = 0.9$ GeV the resulting j_m function has a minimum at 875 MeV.

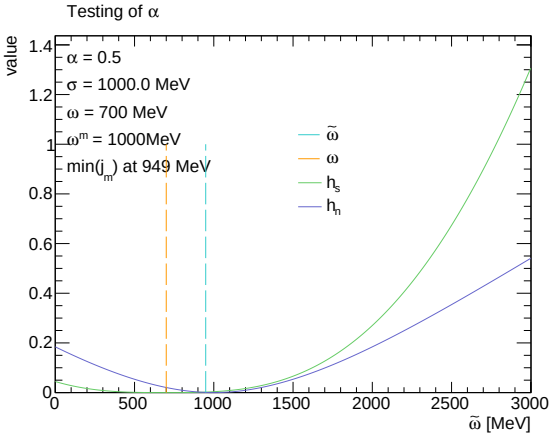
Lastly there is the α constant, which is the only free parameter for the function j_m. Tests of the α constant is in Figure 5.12. If the α constant is high, the model coefficient can influence the minimum of the j_m function more, than for a low value of α.



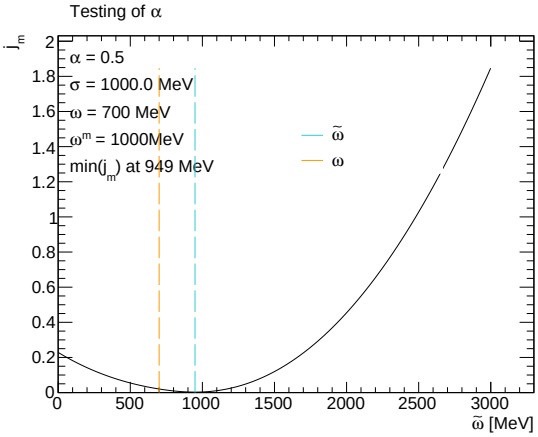
(a) The h_s and h_n functions for $\alpha = 0.1$ MeV



(b) The j_m function for $\alpha = 0.1$ MeV



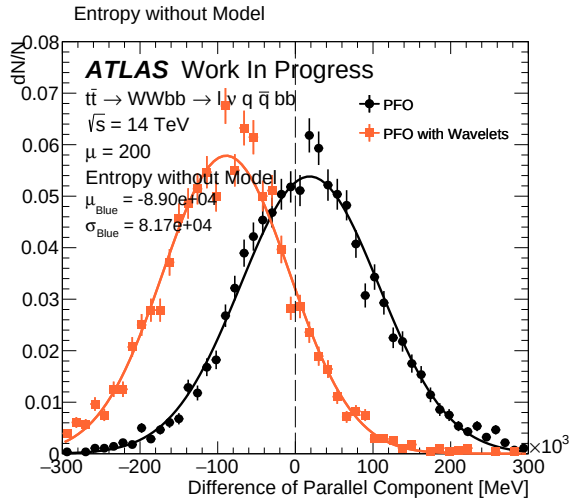
(c) The h_s and h_n functions for $\alpha = 0.5$ MeV



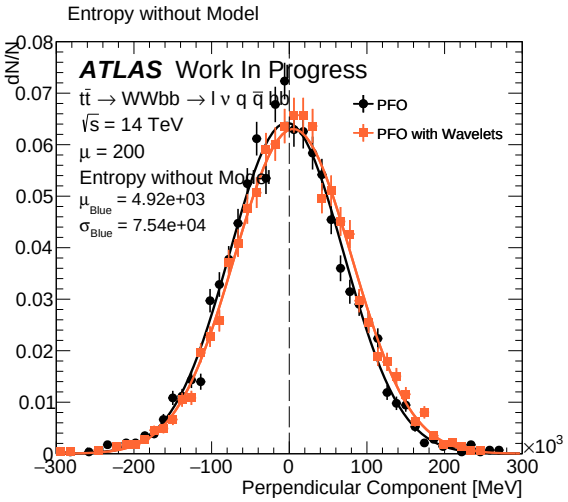
(d) The j_m function for $\alpha = 0.5$ MeV

Figure 5.12: Result of the h_s, h_n and j_m function for $\alpha = 0.1$ and $\alpha = 0.5$. In Figure (a) and (c) the h_s and h_n function are shown. The j_m functions are in Figure (b) and (d). For $\alpha = 0.1$ the resulting new estimate of the wavelet coefficient is 875 MeV and for $\alpha = 0.5$ the wavelet coefficient is 949 MeV. This means that a higher value of α will make a heavier weight of the model coefficient.

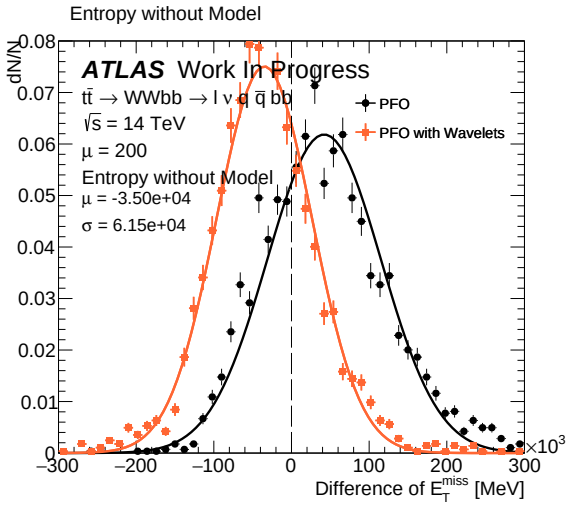
The results for using the MEM method without a model is in Figure 5.13 and with a model is in Figure 5.14. The MEM method does not remove enough particles, since the $\sum E_T$ differences is not centred at zero. The parallel component of E_T^{miss} differs greatly from the true value. This could mean that the MEM method is removing some of the high p_T particles and keeping some of the lower p_T particles. A reason for this could be because of the α constant. The α constant determines how great the noise contribution to the entropy is for a coefficient. This α constant was set between 0 and 1, i.e. the model coefficient will be less significant than the actual coefficient when scaling the coefficients. This α value might not have been set low enough, since the coefficients are being scaled down too much. This results in the high p_T particles being removed from the event. There is also almost no difference in using the MEM method with and without a model. This could mean that the model contains a lot of zeros, which would be the same as not having a model. A new model could be considered where it could rely on both the charged and neutral particles, since at the moment it only relies on the charged particles.



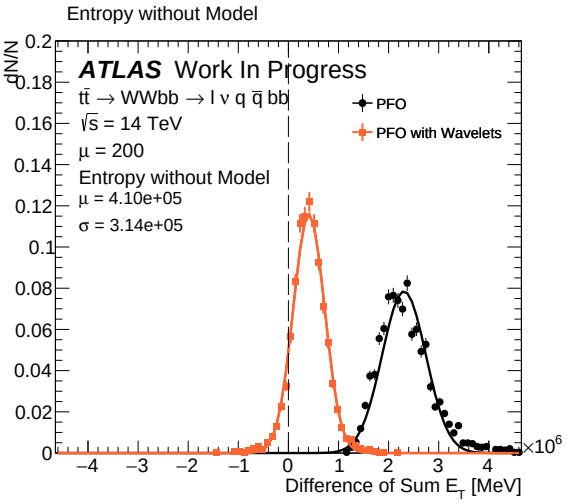
(a) Difference of the truth parallel component and the reconstructed parallel component.



(b) Difference of the truth perpendicular component and the reconstructed perpendicular component.

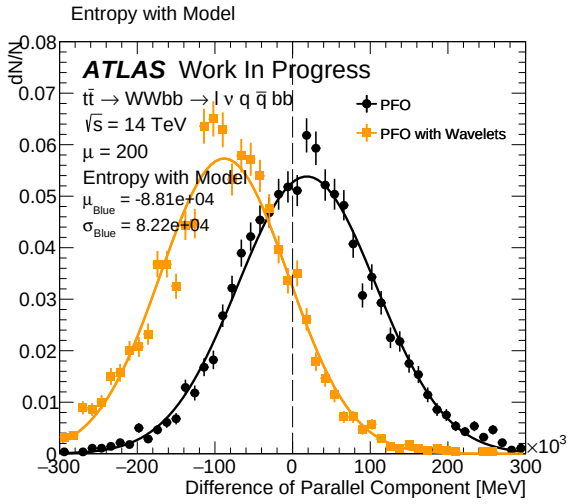


(c) Difference of the truth E_T^{miss} and the reconstructed E_T^{miss} component.

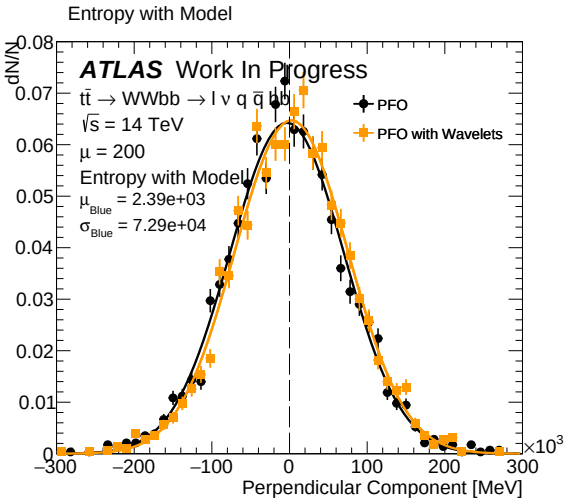


(d) Difference of the truth sum E_T and the reconstructed sum E_T.

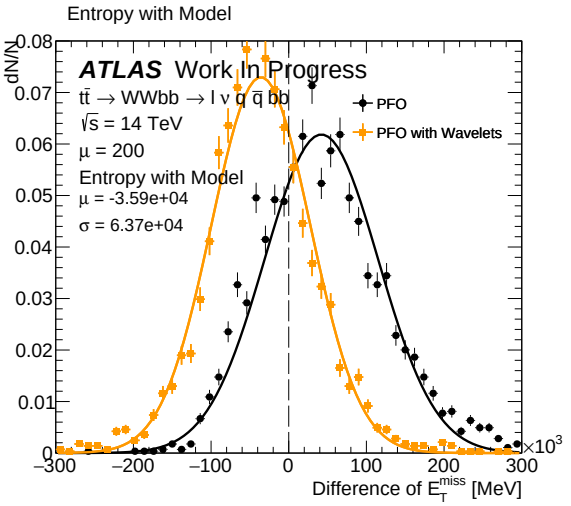
Figure 5.13: Results of an entropy scaling of the wavelet coefficients without a model. In Figure (a) the difference of the parallel component of the reconstructed and true MET is shown. A gaussian centered at zero with a small width is the desired result. In Figure (b) the difference of the perpendicular component of the reconstructed and true MET is shown. Figure (a) and (b) demonstrates a good choice of cuts in wavelet coefficients, since both the correctness of the parallel and perpendicular component is improved. In Figure (c) the difference of the true E_T^{miss} and the reconstructed E_T^{miss} is shown. In Figure (d) the difference of sum E_T truth and reconstructed sum E_T is shown.



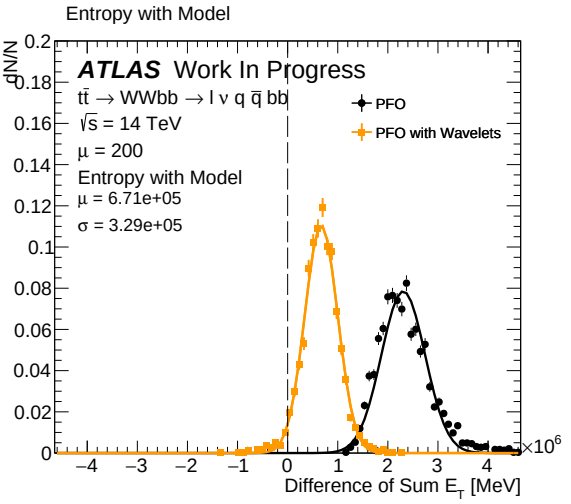
(a) Difference of the truth parallel component and the reconstructed parallel component.



(b) Difference of the truth perpendicular component and the reconstructed perpendicular component.



(c) Difference of the truth E_T^{miss} and the reconstructed E_T^{miss} component.

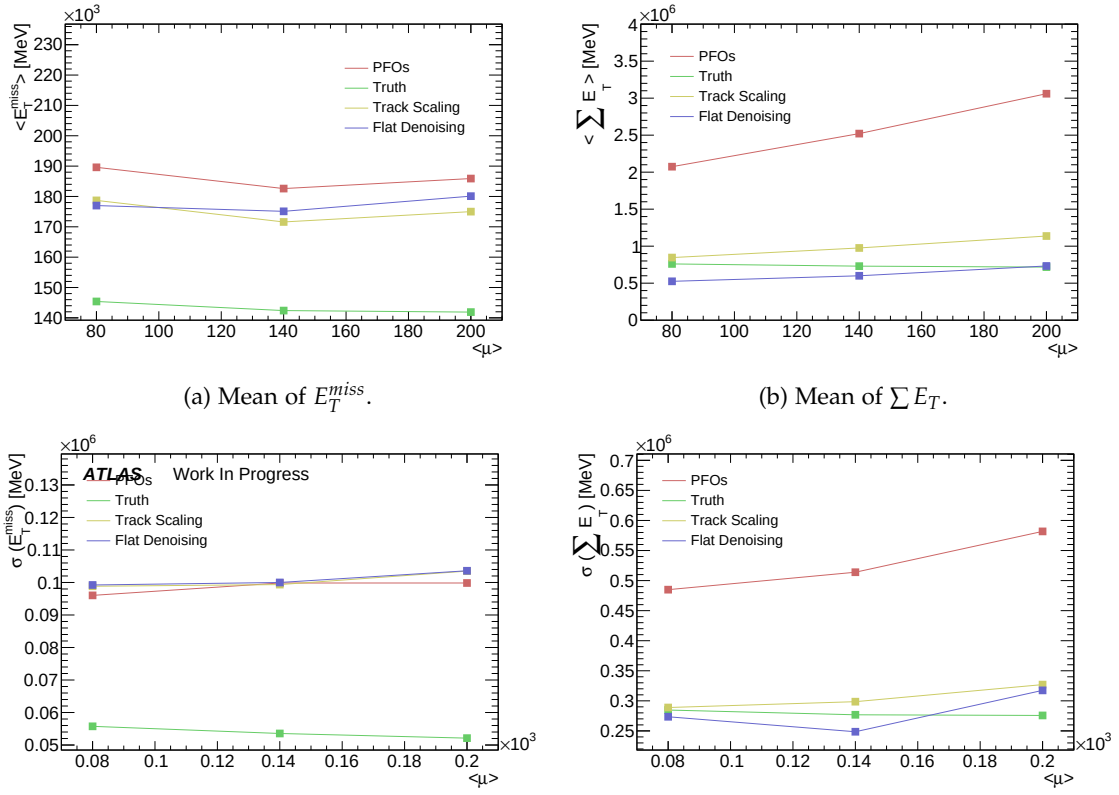


(d) Difference of the truth sum E_T and the reconstructed sum E_T.

Figure 5.14: Results of an entropy scaling of the wavelet coefficients using a model. In Figure (a) the difference of the parallel component of the reconstructed and true MET is shown. A gaussian centered at zero with a small width is the desired result. In Figure (b) the difference of the perpendicular component of the reconstructed and true MET is shown. Figure (a) and (b) demonstrates a good choice of cuts in wavelet coefficients, since both the correctness of the parallel and perpendicular component is improved. In Figure (c) the difference of the true E_T^{miss} and the reconstructed E_T^{miss} is shown. In Figure (d) the difference of sum E_T truth and reconstructed sum E_T is shown.

5.2 Summary of Results

The result for the flat denoising case was that a cut of about 4.5 GeV was reasonable. In the track filtering case the ratio cut was good around 0.2. A summary of the mean and standard deviations by the best methods for the $t\bar{t}$ case at several pile-up stages is shown in Figure 5.15.



(a) Mean of E_T^{miss}.

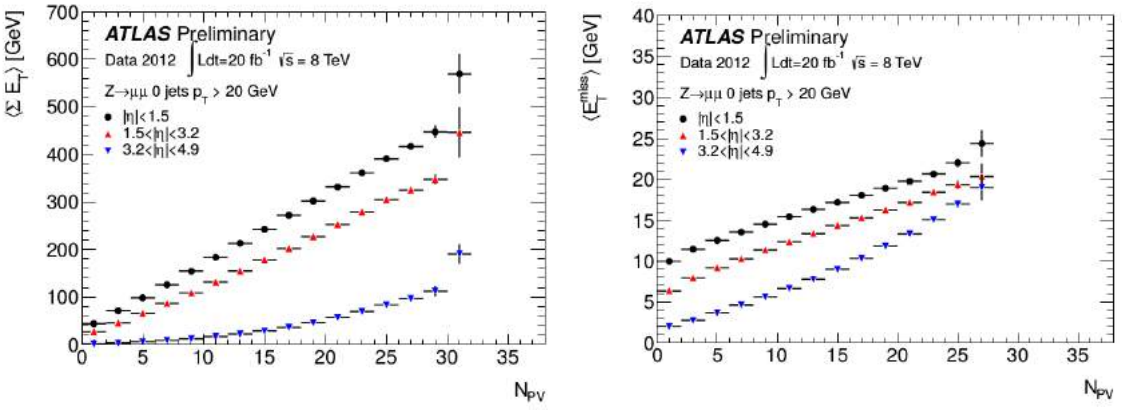
(b) Mean of $\sum E_T$.

(c) Standard deviation of E_T^{miss}.

(d) Standard deviation of $\sum E_T$.

Figure 5.15: E_T^{miss} and $\sum E_T$ as a function of pile-up, using different wavelet methods and the $t\bar{t}$ datasets. In general the mean of the $\sum E_T$ is decreased when using the wavelet methods, and the value is closer to the true value. The mean of the E_T^{miss} decreases slightly when using wavelet methods and is therefore closer to the true value. The standard deviation of E_T^{miss} does not change and the standard deviation of $\sum E_T$ decreases with use of wavelet methods.

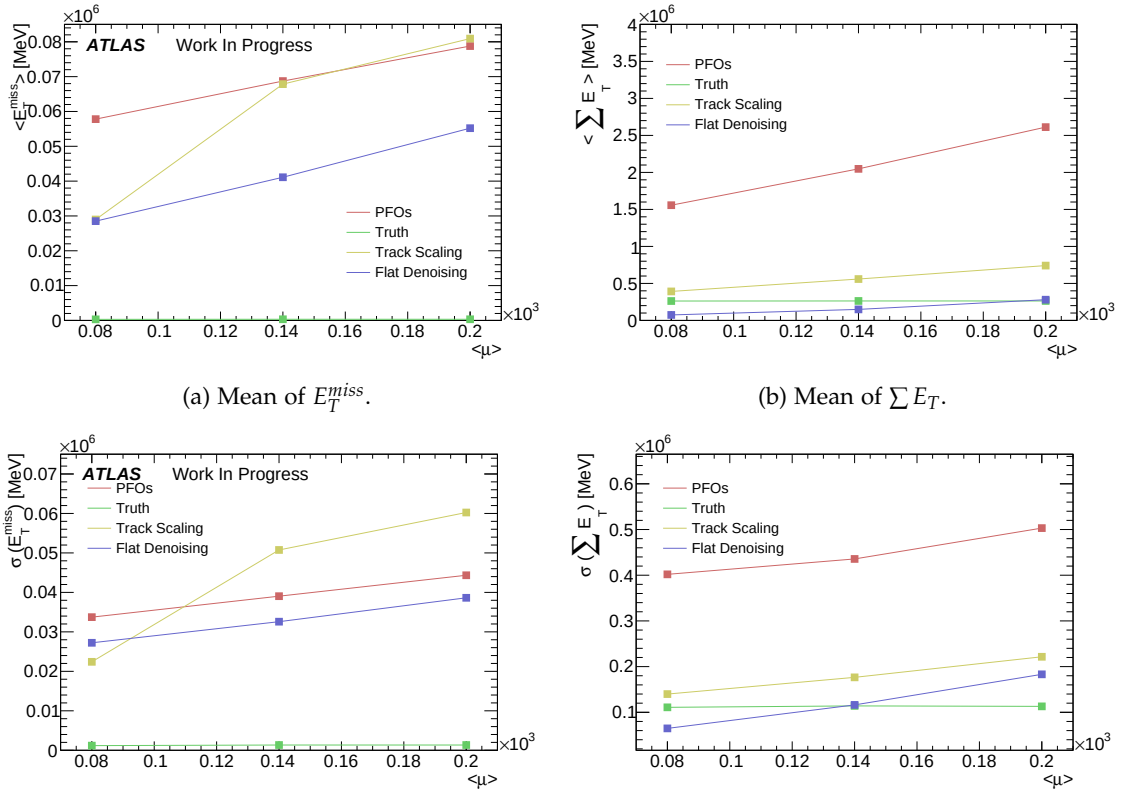
With the best methods of wavelet denoising for reconstructing missing transverse energy, we can compare this to other methods. In [10] the reconstruction of missing transverse energy was tested at different pile-up, see Figure 5.16. I performed similar experiments for $Z \to \mu\mu$ at pile-up ranging from $70 - 210$, to see the development of missing transverse energy. The results are in Figure 5.17.



(a) $\sum E_T$ as a function of pile-up.

(b) E_T^{miss} as a function of pile-up.

Figure 5.16: In [10] experiments were made of reconstructing missing transverse energy as different pile-up. In the $Z \to \mu\mu$ sample the missing transverse energy should be zero, but as you can see the measurements of the missing transverse energy ranges from $0 - 23$ GeV.



(a) Mean of E_T^{miss}.

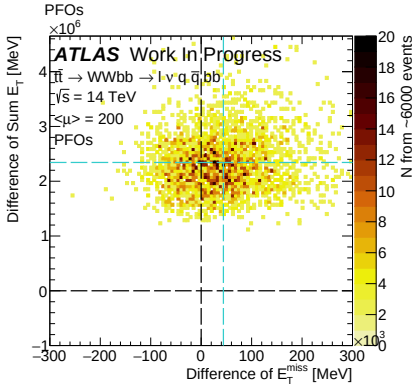
(b) Mean of $\sum E_T$.

(c) Standard deviation of E_T^{miss}.

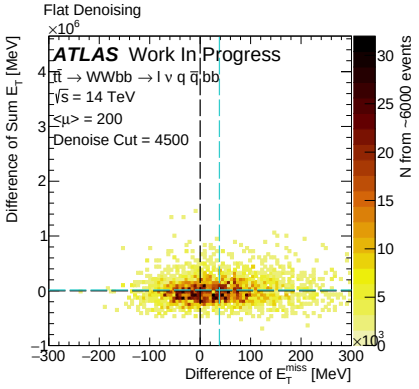
(d) Standard deviation of $\sum E_T$.

Figure 5.17: E_T^{miss} and $\sum E_T$ as a function of pile-up, using different wavelet methods and the $Z \to \mu\mu$ datasets. In general the mean of the $\sum E_T$ is decreased when using the wavelet methods, and the value is closer to the true value. The standard deviation of $\sum E_T$ decreases with use of wavelet methods.

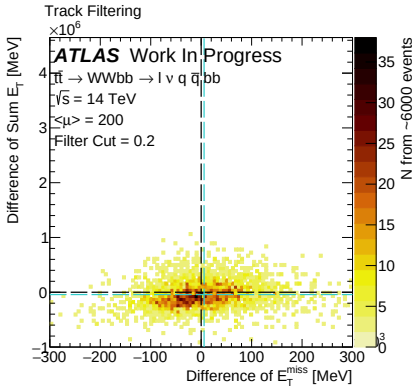
In Figure 5.18 2D scatter plots of the difference from truth of E_T^{miss} and $\sum E_T$ is shown using different wavelet methods. These scatter plots correspond to the reconstruction performed in Section 4.2. The mean and standard deviation of E_T^{miss} does not change significantly. The mean and standard deviation of $\sum E_T$ does however change. There are no clear correlation between E_T^{miss} and $\sum E_T$ in these experiments.



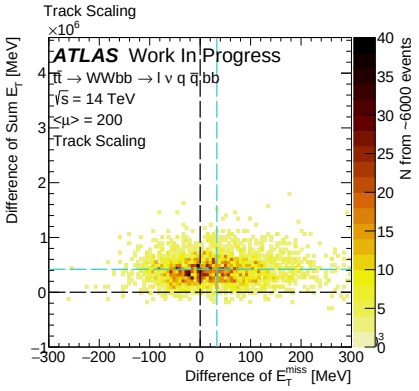
(a) PFOs.



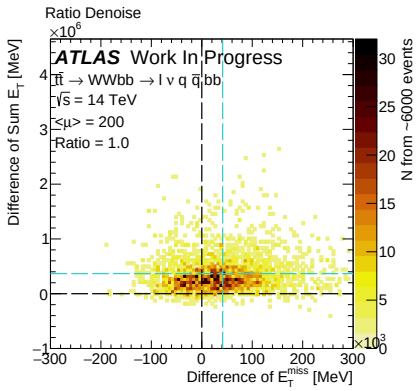
(b) Flat denoising with a cut of 4.5 GeV.



(c) Track filtering with a ratio cut of 0.2.



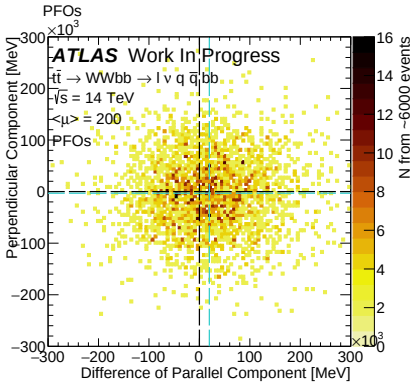
(d) Track scaling.



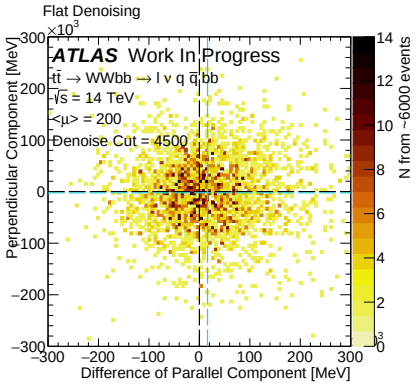
(e) Ratio denoising.

Figure 5.18: 2D plot of the difference of reconstructed and true $\sum E_T$ and the difference of reconstructed and true E_T^{miss} using different kind of wavelet methods.

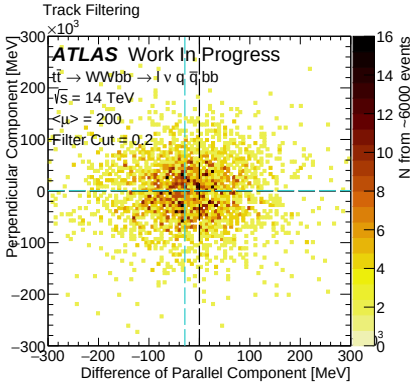
In Figure 5.19 2D scatter plots of the difference from truth of the parallel and perpendicular components of E_T^{miss} is shown using different wavelet methods. The mean and standard deviations of the difference of the components of E_T^{miss} does not change significantly.



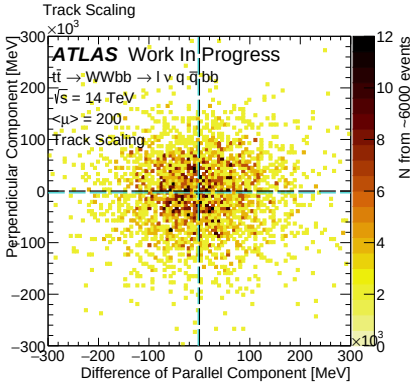
(a) PFOs.



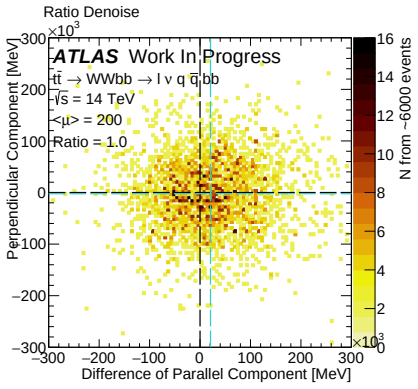
(b) Flat denoising with a cut of 4.5 GeV.



(c) Track filtering with a ratio cut of 0.2.



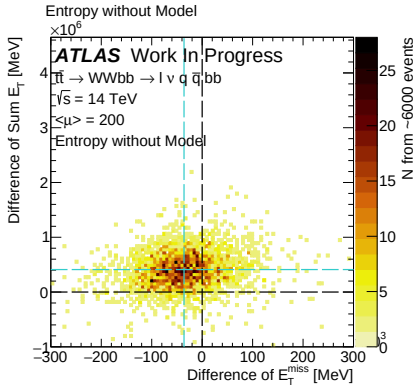
(d) Track scaling.



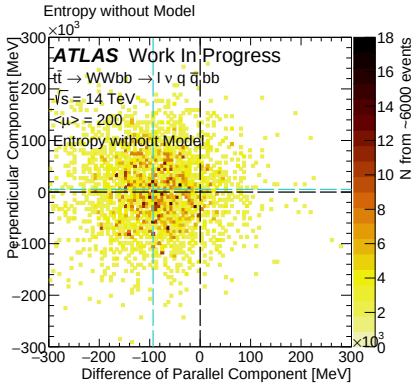
(e) Ratio denoising.

Figure 5.19: 2D plot of the perpendicular component of E_T^{miss} and the difference of reconstructed and true parallel component of E_T^{miss} using different kind of wavelet methods.

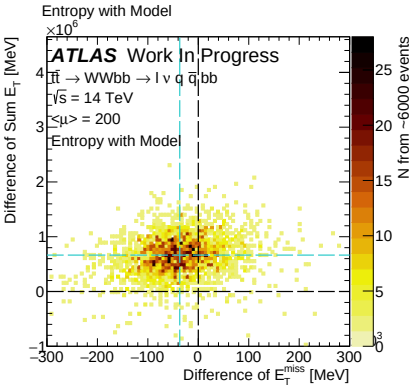
Lastly, the 2D scatter plots of the true difference of E_T^{miss}, $\sum E_T$ and the components of E_T^{miss} using the MEM methods are in Figure 5.20. The reconstruction of $\sum E_T$ improves, but there are no other significant improvements in the other reconstructions. The reconstruction of the parallel component gets slightly worse than the reconstruction obtained without using MEM. In general it seems that the MEM method is down-scaling the high coefficients too much, which might explain this difference in the parallel component. The MEM method will down-scale coefficients with a large difference to the model coefficient more, than if there was a small difference between the coefficients and model coefficients. For further improvements, more testing of the α constants should be performed when using MEM, since it might correct for this problem.



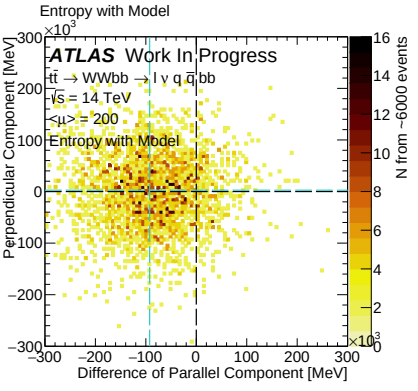
(a) Track filtering with a ratio cut of 0.2.



(b) Track scaling.



(c) Track filtering with a ratio cut of 0.2.



(d) Track scaling.

Figure 5.20: 2D plot of the perpendicular component of E_T^{miss} and the difference of reconstructed and true parallel component of E_T^{miss} using the maximum entropy method with and without a model.

Concluding remarks

5.2.1 Summary

In this thesis the methods of wavelets were introduced for pile-up removal for datasets with a mean pile-up of 70-210 at $\sqrt{s} = 14$ TeV.

The flat denoising method proves good for pile-up cleaning. In this method is that the user has to chose a cut in the energies of the coefficients. This cut has a great impact of the results of reconstructing of MET, and should therefore be chosen with care. To avoid making a user cut, the method of ratio denoise was introduced. This method did better than the flat denoising method, but it is much slower to run. The track filtering method does not require a cut in energies but a cut in the ratio of information from a model to the wavelet coefficients. This has proved to be a good method when using a ratio of 0.1. The track scaling method does not need any user parameters and performs well but not as good as the flat denoising and track filtering methods. Lastly the MEM method proves promising and gives some reasonable results. The MEM method seemed to be scaling the coefficients too harshly, but this can be adjusted using the α constants. Summary plots for the different methods for pile-up at 80, 140 and 200 was shown in Figure 5.15 and 5.17. Overall the general method of wavelets provided good results for pile-up cleaning.

5.2.2 Outlook

The next step could be to test these wavelets on real data instead of Monte Carlo datasets. The center-of-mass energy at the LHC is currently at 13 TeV, but the methods here were tested at 14 TeV since there were no other datasets available for this analysis. If more datasets of high pile-up are produced, we can put the methods of wavelets to further tests. The wavelet cleaning method could also be introduced earlier in the reconstruction process, such that the pile-up particles are removed before the ATLAS reconstruction algorithms are used. This could also be beneficial when reconstructing missing transverse energy, since wavelets could improve the soft-term of missing transverse energy.

There is also room for improvements of the wavelet methods. The MEM method can be optimized using computing efficient algorithms for estimating the integrals and minimization methods. In this thesis the MEM method was not fully optimized and is therefore much slower than all of the other methods. But since the MEM method is an iterative method, it will always be slower than the other wavelet methods.

With the improvement of reconstruction of missing transverse energy, it is now ready for other physics analysis. This can for example be for the decay $H \to WW \to l\nu q\bar{q}$ to detect the already discovered

Higgs boson or the undiscovered Graviton. Other physics analyses include exotic decays or the search for SUSY particles. There are also currently searches for a heavier Higgs boson, to see if there exist one Higgs boson ore several at different masses. Missing transverse energy is interesting for New Physics, since the missing transverse energy can come from neutrinos or from invisible particles beyond the Standard Model.

Appendices

A.1 Datasets

The datasets for this thesis was chosen from the list at REF(LINK). This is a list of high pile-up samples at 14 TeV. The two high pile-up samples used were the physics case of Z decaying into two muons, and the $t\bar{t}$ data sample where the top quarks decay into 4 quarks, 1 neutrino and 1 lepton.

A.1.1 $Z\mu\mu$

The $Z \to \mu\mu$ datasets were used for testing of the reconstruction of missing transverse energy, which should be zero. Details of the $Z \to \mu\mu$ datasets used is in Table A.1.

Pileup	Name	Generator	Type	Tag
70-90	Zmumu	Powheg + Pythia8	recon + AOD	r7699
130-150	Zmumu	Powheg + Pythia8	recon + AOD	r7768
190-210	Zmumu	Powheg + Pythia8	recon + AOD	r7769

Table A.1: The $Z \to \mu\mu$datasets

A.1.2 $t\bar{t}$

The $t\bar{t}$ datasets have one neutrino in each event with an momenta of at least 100 GeV. Information of the $t\bar{t}$ datasets can be found in Table A.2.

Pileup	Name	Generator	Type	Tag
70-90	ttbar_hdamp175p5_MET100	Powheg + Pythia8 + EvtGen	recon + AOD	r7699
70-90	ttbar_hdamp175p5_MET100	Powheg + Pythia8 + EvtGen	recon + AOD	r7702
70-90	ttbar_hdamp175p5_MET100	Powheg + Pythia8 + EvtGen	recon + AOD	r7709
130-150	ttbar_hdamp175p5_MET100	Powheg + Pythia8 + EvtGen	recon + AOD	r7768
130-150	ttbar_hdamp175p5_MET100	Powheg + Pythia8 + EvtGen	recon + AOD	r7700
190-210	ttbar_hdamp175p5_MET100	Powheg + Pythia8 + EvtGen	recon + AOD	r7701
190-210	ttbar_hdamp175p5_MET100	Powheg + Pythia8 + EvtGen	recon + AOD	r7769

Table A.2: The $t\bar{t}$ datasets

Table A.3: Event Selection for the $t\bar{t}$ datasets.

	Cut	**Events**
r7709	Initial	50000
	Single Lepton Trigger	50000
	Exactly one high or iso lepton	23832
	jets	10613
	b-jets	7322
r7768	Initial	50000
	Single lepton trigger	50000
	Exactly one high or iso lepton	24025
	Jets	10825
	b-jets	9432
r7769	Initial	50000
	Single lepton trigger	50000
	Exactly one high or iso lepton	24239
	Jets	11117
	b-jets	10847
r7699	Initial	50000
	Single Lepton Trigger	50000
	Exactly one high or iso lepton	23797
	jets	10654
	b-jets	7135
r7700	Initial	49950
	Single lepton trigger	49950
	Exactly one high or iso lepton	23937
	Jets	10834
	b-jets	9422
r7701	Initial	45000
	Single lepton trigger	45000
	Exactly one high or iso lepton	21742
	Jets	10020
	b-jets	9784
r7702	Initial	33100
	Single lepton trigger	33100
	Exactly one high or iso lepton	15721
	Jets	7060
	b-jets	4648

A.2 Components of Missing Transverse Energy

The goodness parameters used in this thesis include the parallel and perpendicular components of the reconstructed missing transverse energy vector projected onto the truth missing transverse energy vector. The equations used for this are,

$$E_{\parallel}^{miss} = \frac{p_x^{miss} \cdot p_x^{truth,miss} + p_y^{miss} \cdot p_y^{truth,miss}}{\sqrt{|p_T^{truth,miss}|}} \tag{A.2}$$

$$E_{x,\perp}^{miss} = p_x^{miss} - E_{\parallel}^{miss} \cdot \frac{p_x^{truth,miss}}{\sqrt{|p_T^{truth,miss}|}} \tag{A.3}$$

$$E_{y,\perp}^{miss} = p_y^{miss} - E_{\parallel}^{miss} \cdot \frac{p_y^{truth,miss}}{\sqrt{|p_T^{truth,miss}|}} \tag{A.4}$$

$$E_{\perp}^{miss} = \text{sign}(E_{x,\perp}^{miss})\sqrt{(E_{x,\perp}^{miss})^2 + (E_{y,\perp}^{miss})^2)} \tag{A.5}$$

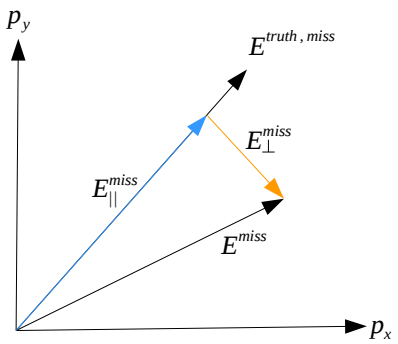
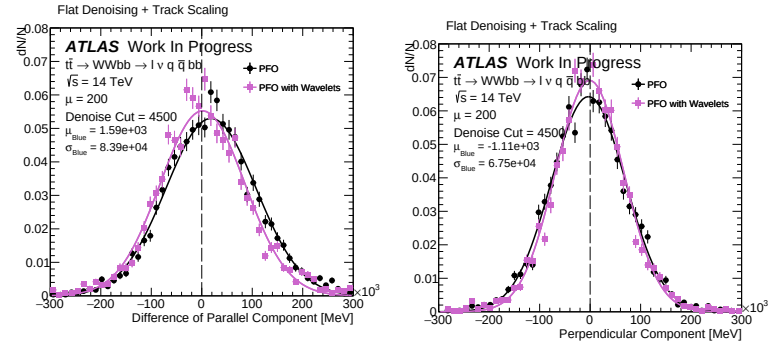


Figure A.1: Visualization of the parallel and perpendicular components of E_T^{miss}

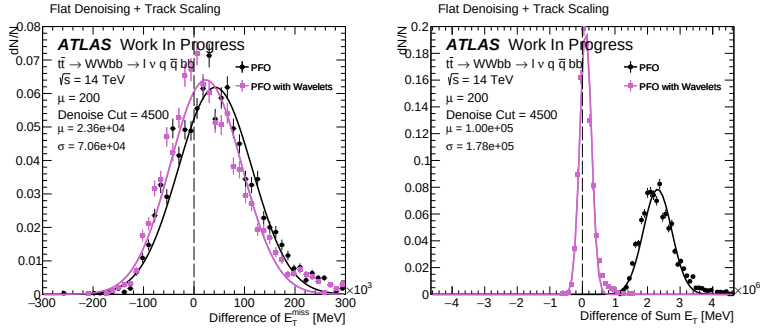
A.3 Additional Wavelet Results

In this section a combination of wavelet methods were used. The first case is using track scaling on the charged particle flow objects and flat denoising on the neutral particles flow objects. The track scaling method uses information from the inner detector. The inner detector can not see the neutral particles, and therefore a flat denoise cut is performed on the neutral particle flow objects. The flat denoise cut was set to 4.5 GeV, since from tests in Section 5.1 it proved to be a good choice. The results from the combined methods of flat denoising and track scaling is in Figure A.2.



(a) Difference of the parallel component.

(b) The perpendicular component.



(c) Difference of E_T^{miss}

(d) Difference of $\sum E_T$

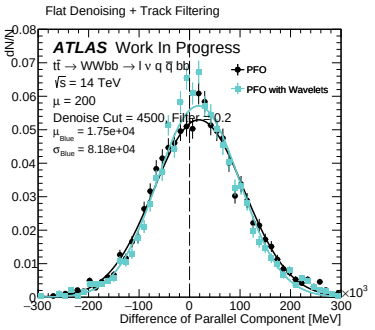
Figure A.2: Results of track scaling of the wavelet coefficients of the charged particles, and flat denoise cut of the wavelet coefficients of the neutral particles. In Figure (a) the difference of the parallel component of the reconstructed and true MET is shown. In Figure (b) the difference of the perpendicular component of the reconstructed and true MET is shown. In Figure (c) the difference of the true E_T^{miss} and the reconstructed E_T^{miss} is shown. In Figure (d) the difference of sum E_T truth and reconstructed sum E_T is shown.

The next method tested is using filter denoising on the charged particle flow objects and flat denoising on the neutral particle flow objects. The filter denoising uses information from the charged particles passing through the inner detector, and is therefore only used on the charged particle flow objects. The filter ratio was set to 0.2 and the denoise cut was at 4.5 GeV since these choices proved well in Section 5.1. Results are shown in Figure A.3.

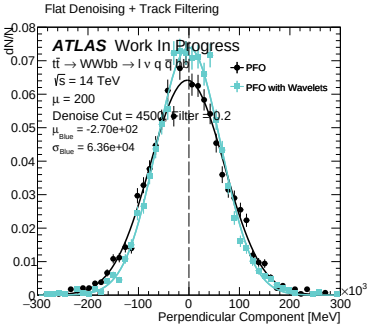
More tests were performed using the ratio denoise method than shown in Section 5.1. In these tests the ratio that the neutral particles must fullfill was scaled by a user set constant β

$$ratio = \frac{|pfo_{ch}^{PV}|}{|pfo_{ch}|} \times \beta \tag{A.6}$$

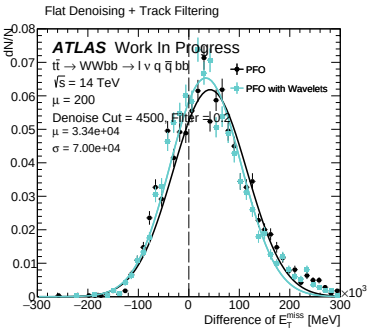
Figure A.3: Results of a filter denoising in the wavelet coefficients of the charged particle flow objects and a flat denoise cut in the neutral particles flow objects. In Figure (a) the difference of the parallel component of the reconstructed and true MET is shown. In Figure (b) the difference of the perpendicular component of the reconstructed and true MET is shown. In Figure (c) the difference of the true E_T^{miss} and the reconstructed E_T^{miss} is shown. In Figure (d) the difference of sum E_T truth and reconstructed sum E_T is shown.



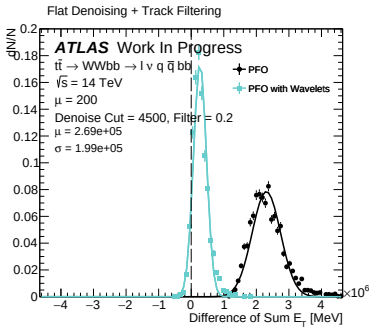
(a) Difference of the parallel component.



(b) The perpendicular component.



(c) Difference of E_T^{miss}



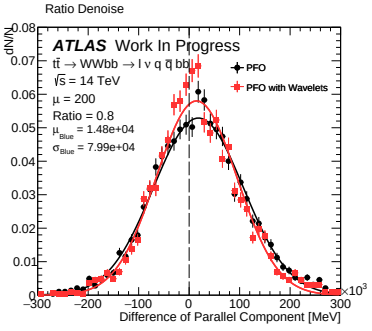
(d) Difference of $\sum E_T$

where we want to cut in the neutral particles flow objects until the following equation is valid,

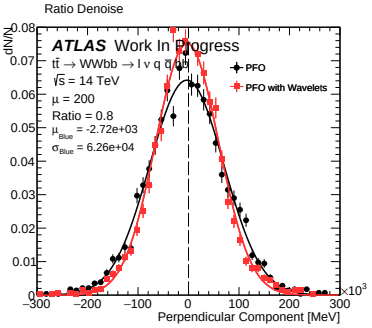
$$\frac{|pfo_{neu}^{PV}|}{|pfo_{neu}|} \geq ratio \tag{A.7}$$

By scaling the ratio by a constant, we can make a harder or softer cut in the neutral coefficients. If $\beta < 1$ then the neutral particles flow objects has a smaller cut than if $\beta \geq 1$. The best value for β was determined to be 0.8. The result the components of missing transverse energy for $\beta = 0.8$ is in Figure A.4 and a summary of using other β values is in Figure A.5.

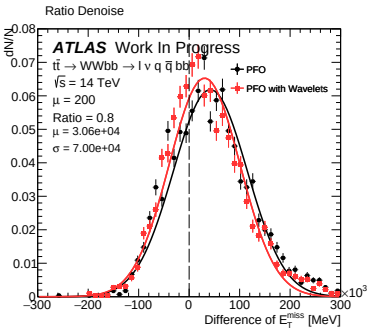
Figure A.4: Results of a ratio denoise cut in the wavelet coefficients. In Figure (a) the difference of the parallel component of the reconstructed and true MET is shown. A gaussian centered at zero with a small width is the desired result. In Figure (b) the difference of the perpendicular component of the reconstructed and true MET is shown. Figure (a) and (b) demonstrates a good choice of cuts in wavelet coefficients, since both the correctness of the parallel and perpendicular component is improved. In Figure (c) the difference of the true E_T^{miss} and the reconstructed E_T^{miss} is shown. In Figure (d) the difference of sum E_T truth and reconstructed sum E_T is shown.



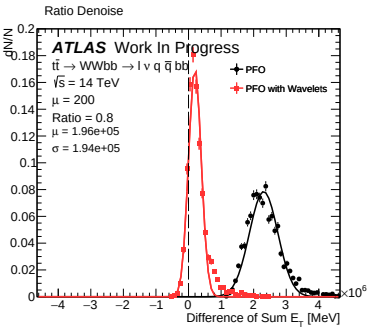
(a) Difference of the parallel component.



(b) The perpendicular component

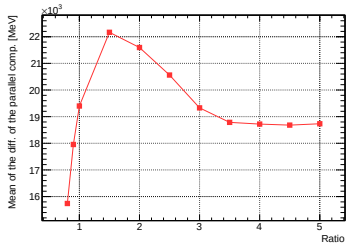


(c) Difference of E_T^{miss}

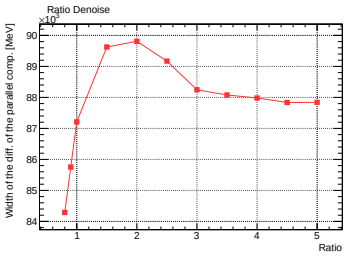


(d) Difference of $\sum E_T$

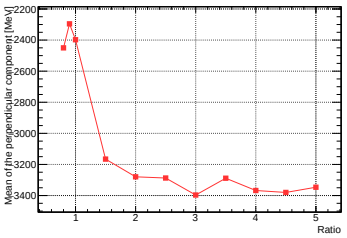
Figure A.5: Fit results for ratio denoise. The variable is the ratio, and it varies from 0.8 to 2.5 by 0.1.



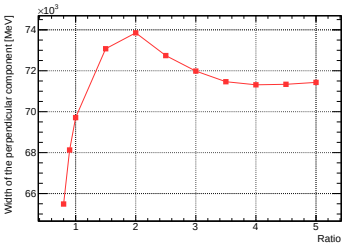
(a) Fit mean of parallel.



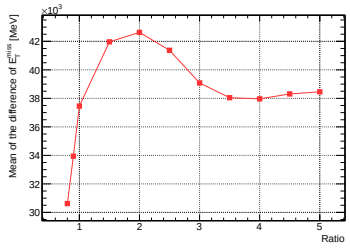
(b) Fit width of parallel.



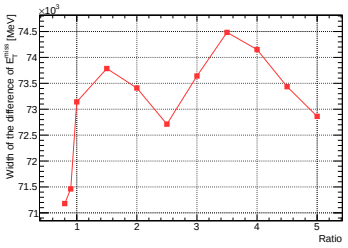
(c) Fit mean of perp.



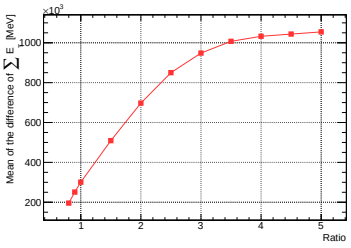
(d) Fit width of perp.



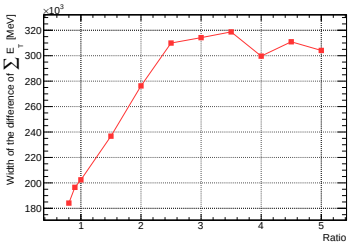
(e) Fit mean of E_T^{miss}.



(f) Fit width of E_T^{miss}.

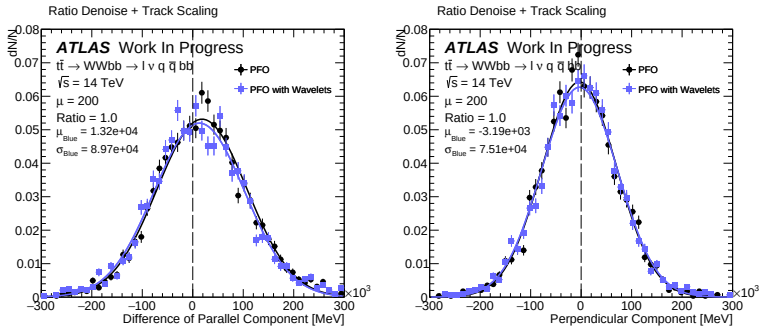


(g) Fit mean of E_T.

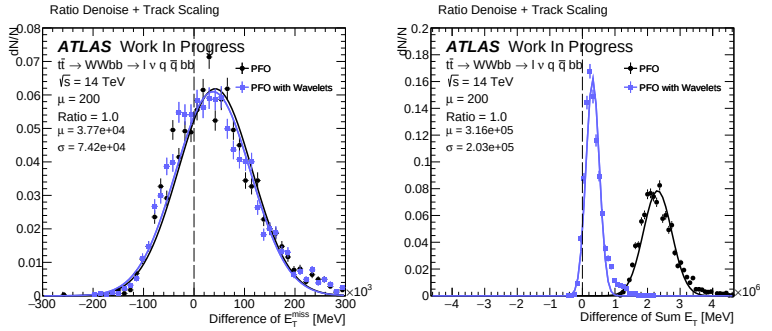


(h) Fit width of sum E_T.

The ratio denoise method can also be combined with the track filtering or track scaling method. This is because the ratio denoise method makes cuts in the neutral particle flow objects, but not in the charged particle flow objects. For scaling or cutting in the charged particle flow objects, the methods of track scaling and filtering was used. The result of the ratio denoise combined with track scaling is in Figure A.6 and the result of the ratio denoise method combined with track filtering is in Figure A.7.



(a) Difference of parallel component. (b) The perpendicular component.

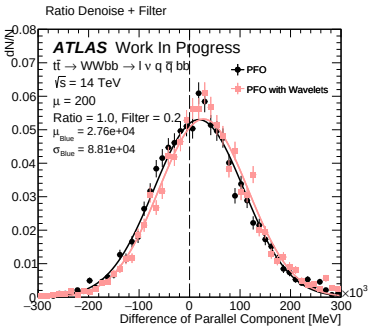


(c) Difference of E_T^{miss} (d) Difference of $\sum E_T$

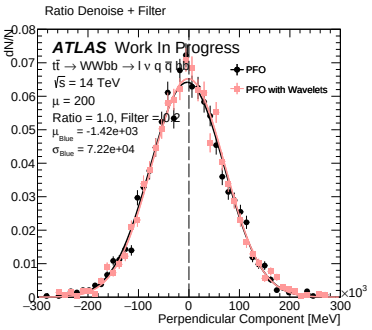
Figure A.6: Results of a track scaling cut in the wavelet coefficients of the charged PFOs and a ratio denoise of the coefficients of the neutral PFOs. In Figure (a) the difference of the parallel component of the reconstructed and true MET is shown. A gaussian centered at zero with a small width is the desired result. In Figure (b) the difference of the perpendicular component of the reconstructed and true MET is shown. Figure (a) and (b) demonstrates a good choice of cuts in wavelet coefficients, since both the correctness of the parallel and perpendicular component is improved. In Figure (c) the difference of the true E_T^{miss} and the reconstructed E_T^{miss} is shown. In Figure (d) the difference of sum E_T truth and reconstructed sum E_T is shown.

The ratio denoise using the p_T measurements instead of the number of neutral particle flow objects was also combined with the track filtering and scaling methods. The track filtering result is in Figure A.8 and the track filtering result is in Figure A.9.

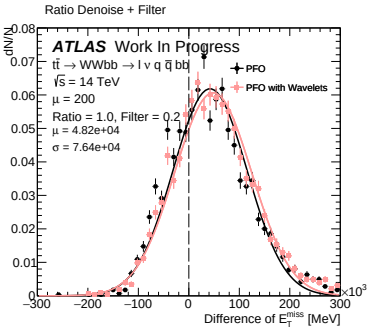
Figure A.7: Results of a track filtering cut of the charged PFOs and a ratio denoise cut of the neutral PFOs. In Figure (a) the difference of the parallel component of the reconstructed and true MET is shown. A gaussian centered at zero with a small width is the desired result. In Figure (b) the difference of the perpendicular component of the reconstructed and true MET is shown. Figure (a) and (b) demonstrates a good choice of cuts in wavelet coefficients, since both the correctness of the parallel and perpendicular component is improved. In Figure (c) the difference of the true E_T^{miss} and the reconstructed E_T^{miss} is shown. In Figure (d) the difference of sum E_T truth and reconstructed sum E_T is shown.



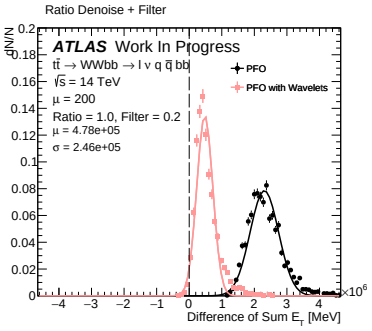
(a) Difference of parallel component.



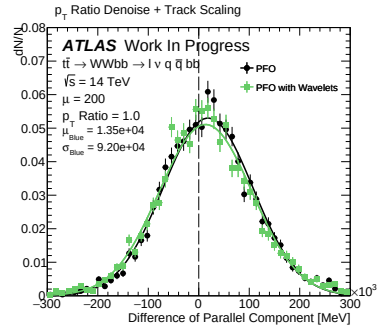
(b) Perpendicular component.



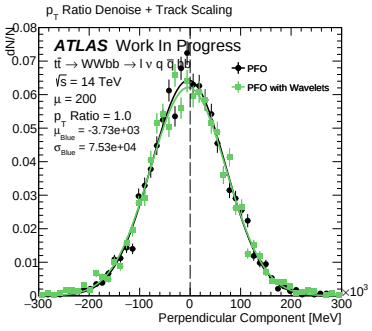
(c) Difference of E_T^{miss}



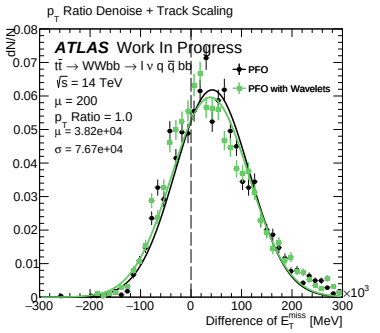
(d) Difference of $\sum E_T$



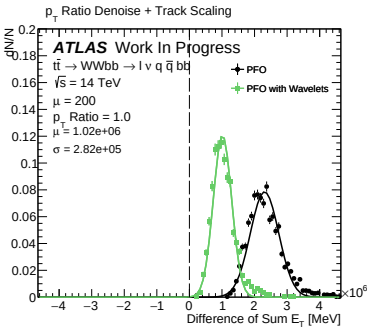
(a) Difference of parallel component.



(b) Perpendicular component.



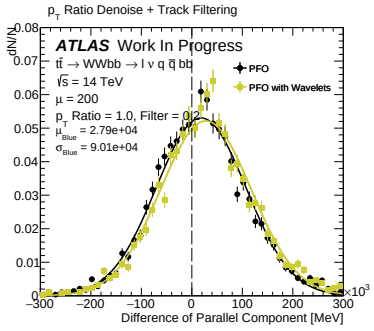
(c) Difference of E_T^{miss}



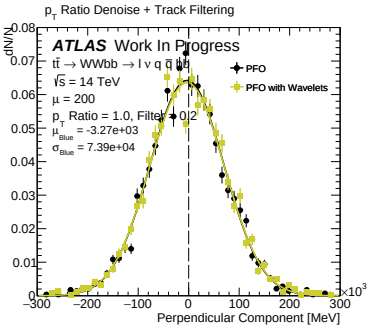
(d) Difference of $\sum E_T$

Figure A.8: Results of a track scaling of the charged particle flow objects and a p_T ratio denoise of the neutral particle flow objects. In Figure (a) the difference of the parallel component of the reconstructed and true MET is shown. A gaussian centered at zero with a small width is the desired result. In Figure (b) the difference of the perpendicular component of the reconstructed and true MET is shown. Figure (a) and (b) demonstrates a good choice of cuts in wavelet coefficients, since both the correctness of the parallel and perpendicular component is improved. In Figure (c) the difference of the true E_T^{miss} and the reconstructed E_T^{miss} is shown. In Figure (d) the difference of sum E_T truth and reconstructed sum E_T is shown.

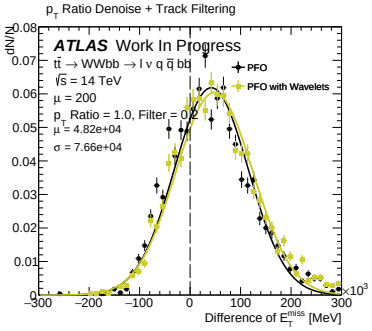
Figure A.9: Results of a track filtering of the charged particle flow objects and a p_T ratio denoise of the neutral particle flow objects. In Figure (a) the difference of the parallel component of the reconstructed and true MET is shown. A gaussian centered at zero with a small width is the desired result. In Figure (b) the difference of the perpendicular component of the reconstructed and true MET is shown. Figure (a) and (b) demonstrates a good choice of cuts in wavelet coefficients, since both the correctness of the parallel and perpendicular component is improved. In Figure (c) the difference of the true E_T^{miss} and the reconstructed E_T^{miss} is shown. In Figure (d) the difference of sum E_T truth and reconstructed sum E_T is shown.



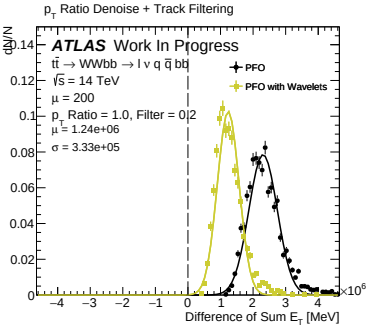
(a) Difference of the parallel component.



(b) The perpendcular component.



(c) Difference of E_T^{miss}



(d) Difference of $\sum E_T$

List of Figures

List of Algorithms

List of Tables

References

[1] Nazar Bartosik. Sm-particles table. [http://bartosik.pp.ua/hep_sketches/sm_particles], Jun 2016.

[2] Nazar Bartosik. tt decay channels. [http://bartosik.pp.ua/hep_sketches/tt_decay_channels], Jun 2016.

[3] K. A. Olive et al. Review of Particle Physics. *Chin. Phys.*, C38:090001, 2014.

[4] A. Denner, S. Heinemeyer, I. Puljak, D. Rebuzzi, and M. Spira. Standard Model Higgs-Boson Branching Ratios with Uncertainties. *Eur. Phys. J.*, C71:1753, 2011.

[5] Quantum Diaries. Multiple Year Collision Plots, 2016. [http://www.quantumdiaries.org/2012/06/09/what-next-for-the-higgs/].

[6] Nazar Bartosik. Mc simulation of a pp collision. [http://bartosik.pp.ua/hep_sketches/pp_collision_ttH], Jul 2016.

[7] P. Skands. Introduction to QCD. Lecture notes for TASI 2012. [arXiv:1207.2389].

[8] A. D. Martin A, W. J. Stirling B, R. S. Thorne C, and G. Watt C. Parton distributions for the lhc, 901.

[9] ATLAS Collaboration. Branching fractions of the Higgs boson at 125GeV, 2012. [https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2].

[10] Pile-up Suppression in Missing Transverse Momentum Reconstruction in the ATLAS Experiment in Proton-Proton Collisions at sqrts = 8 TeV. Technical Report ATLAS-CONF-2014-019, CERN, Geneva, May 2014.

[11] ATLAS Collaboration. Pileup Interactions and Data Taking Efficiency , 2016. [https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2].

[12] Track Reconstruction Performance of the ATLAS Inner Detector at $\sqrt{s} = 13$ TeV. Technical Report ATL-PHYS-PUB-2015-018, CERN, Geneva, Jul 2015.

[13] ATLAS Collaboration (G. Aad *et al.*). The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3(S08003), 2008. [inSPIRE:796888].

[14] Nikiforos Nikiforou. Performance of the ATLAS Liquid Argon Calorimeter after three years of LHC operation and plans for a future upgrade. In *Proceedings, 3rd International Conference on Advancements in Nuclear Instrumentation Measurement Methods and their Applications (ANIMMA 2013)*, 2013.

[15] Yu Nakahama on behalf of the ATLAS Collaboration. The atlas trigger system: Ready for run-2, 2015.

[16] Nazar Bartosik. Event simulation in hep experiments. [http://bartosik.pp.ua/hep_sketches/event_simulation], Jul 2016.

[17] V Lacuesta. Track and vertex reconstruction in the atlas experiment. *Journal of Instrumentation*, 8(02):C02035, 2013.

[18] Electron efficiency measurements with the ATLAS detector using the 2012 LHC proton-proton collision data. Technical Report ATLAS-CONF-2014-032, CERN, Geneva, Jun 2014.

[19] T. Cornelissen *et al.* Concepts, Design and Implementation of the ATLAS New Tracking (NEWT). Technical Report ATL-SOFT-PUB-2007-007. ATL-COM-SOFT-2007-002, CERN, Geneva, Mar 2007. [CDS:1020106].

[20] Matteo Cacciari, Gavin P. Salam, Gregory Soyez. The anti-k_t jet clustering algorithm. *JHEP 0804:063,2008*, 2008. [arXiv:0802.1189].

[21] ATLAS Collaboration. Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1. 2016. [arXiv:1603.02934].

[22] G. Aad et al. Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics. 2009.

[23] Performance of missing transverse momentum reconstruction for the ATLAS detector in the first proton-proton collisions at at $\sqrt{s}$= 13 TeV. Technical Report ATL-PHYS-PUB-2015-027, CERN, Geneva, Jul 2015.

[24] Mark Hodgkinson, Dan Tovey, Rob Duxfield. Energy Flow Reconstruction with the eflowRec Combined Reconstruction Software in Athena 15.6.9.8. *JINST*, (588), 2010.

[25] Nazar Bartosik. Pf concept in cms. [http://bartosik.pp.ua/hep_sketches/cms_particle_flow], Jun 2016.

[26] Andreas Soegaard. Boosted Bosons and Wavelets. Master Thesis, University of Copenhagen, 2015.

[27] J. W. Monk. Wavelet Analysis: Event De-noising, Shower Evolution and Jet Substructure Without Jets. 2014.

[28] J.-L. Starck and F. Murtagh. *Astronomical Image and Data Analysis*. Springer, 2006.

[29] Chambolle et. al. Nonlinear Wavelet Image Processing: Variational Problems, Compression, and Noise Removal through Wavelet Shrinkage. *IEEE Trans. Image Processing, Vol. 7, No. 3*, 1998.

[30] Georges Aad et al. Measurements of top-quark pair differential cross-sections in the lepton+jets channel in pp collisions at $\sqrt{s} =$ 8 TeV using the ATLAS detector. 2015.

WAVELETS & INFORMATION THEORY FOR PILE-UP REMOVAL