

Vrije Universiteit Brussel



Faculteit Wetenschappen
Departement Natuurkunde

Usage of the ParticleFlow method at the CMS detector

Michael Maes

Promotor: Prof. Dr. Jorgen D'Hondt

Academiejaar: 2008 – 2009

Proefschrift ingediend met het oog op het behalen van
de academische graad Master in de Fysica

Introduction

The Standard Model of particle physics is a theory which describes the elementary particles in our universe and how they interact. This theory has been tested with a very high precision by numerous experiments.

At the end of 2009, the Large Hadron Collider at CERN should start colliding proton beams. This hadron collider, located near Geneva in Switzerland, will be the biggest and most powerful particle collider in the world. It aims to collide proton beams of 7TeV at a design-luminosity $10^{34} \text{cm}^{-2} \text{s}^{-1}$. This means that it has a collision energy which is about 7 times higher than the Tevatron $p\bar{p}$ collider, currently the most powerful collider in the world. As there are a number of fundamental problems in the Standard Model of particle physics, the search for experimental proof for theories extending the Standard Model has never been more important. Due to its high collision energy and luminosity, the Large Hadron Collider is the ideal candidate to shed some light on the physics beyond the Standard Model. One of the four big experiments located on the LHC collider is the Compact Muon Solenoid experiment (CMS). This compact but nevertheless large general purpose detector is designed to search for various signals of new physics. To analyse the collisions in the detector, all the stable particles produced in interactions should be identified and reconstructed. If a quark appears in the final state of the collision, it will give rise to a collection of collimated hadrons. Seen from the interaction point this will lead to a narrow cone of particles called a jet. As many signals for new physics contain these jets, it is of key importance to have a good method to reconstruct them with the highest precision possible. Traditionally, jets are reconstructed using a cone algorithm with calorimeter cells as input. A new method was implemented in CMS, the Particle Flow method. Using a combination of all the CMS sub-detectors, it tries to identify and reconstruct all particles in the event. This can be used as input to the cone algorithm to build ParticleFlow jets.

The first two chapters of this thesis present an introduction of the Standard Model, the Large Hadron Collider and the CMS experiment. In the third chapter the jet reconstruction in CMS is introduced. The fourth and fifth chapter compare the traditional Calorimeter jets with ParticleFlow jets using simulated top-quark events. Finally, in the sixth chapter the conclusions of the previous chapters are presented.

Dankwoord

Onder de vele mensen die een woordje van dank verdienen, wil ik eerst en vooral mijn promotor Prof. Dr. Jorgen D'Hondt bedanken. Bedankt voor de begeleiding en de vele uren die je in mijn thesis geïnvesteerd hebt. Het was een zeer leerrijke periode voor mij en dit was niet mogelijk geweest zonder jouw advies en expertise.

Ook wil ik de doctoraatsstudenten en de post-doctorale onderzoekers van de top quark groep bedanken die altijd bereid waren om uitleg en advies te geven. Vele problemen op het vlak van computing zijn ondanks jullie hulp opgelost!

Uiteraard wil ik ook mijn thesissenoot Stijn bedanken met wie ik gedurende dit jaar vele interessante discussies gevoerd heb binnen en buiten de Fysica.

Verder wil ik mijn vriendin Freya, mijn ouders en mijn vrienden en familie bedanken. Ik dank jullie voor de steun en begrip die ik van jullie in deze stressvolle periode kreeg. In het bijzonder wil ik ook Freya en mijn vader bedanken voor het nalezen van mijn thesis hoewel het onderwerp hun niet echt lag.

Finally, I want to thank Patrick Janot, Colin Bernet and Simone Gennai for their help in validating the ParticleFlow jets in my samples and their feedback on some of the work I've done.

Contents

Introduction	i
Dankwoord	ii
1 The Standard Model	1
1.1 The particles and their interactions	1
1.2 Mathematical formalism	2
1.3 Problems in the Standard Model	4
1.4 Possible extensions of the Standard Model	7
1.5 Top Quark Physics	7
2 Physics at the LHC	10
2.1 The Large Hadron Collider	10
2.1.1 Motivation for a proton collider	10
2.1.2 Design and operation of the LHC	12
2.1.3 Detectors at the LHC	15
2.2 The Compact Muon Solenoid	16
2.2.1 CMS coordinate conventions	17
2.2.2 The Tracker	18
2.2.3 The Electromagnetic Calorimeter	19
2.2.4 The Hadronic Calorimeter	20
2.2.5 The Muon System	21
2.3 CMS trigger system	21
3 Jet Reconstruction	24
3.1 Cone Algorithms	25
3.2 ParticleFlow reconstruction of Jets	27
3.3 Jet Energy Scale Corrections	32
4 Jet Performance in $t\bar{t}$ events	34
4.1 Data generation and detector simulation	34

4.2	Comparison between CaloJets and PFJets	35
4.2.1	Jet Selection	35
4.2.2	Jet Energy Corrections	36
4.2.3	Transverse momentum resolution and response	37
4.2.4	Angular resolution	47
4.3	Reconstructing m_{Top} with CaloJets and PFJets	50
4.3.1	The m_{Top} reconstruction	51
4.3.2	The m_{Top} resolution and bias	56
5	Using ParticleFlow in $Z' \rightarrow t\bar{t}$ events	58
5.1	The Z' boson	58
5.2	Reconstruction of $m_{t\bar{t}}$ in $Z' \rightarrow t\bar{t}$ events	60
5.3	Identifying boosted top quarks	64
6	Conclusions	73
6.1	Performance of ParticleFlow with Jets	73
6.1.1	Jet resolution and response	73
6.1.2	Reconstruction of m_{Top}	74
6.2	Using ParticleFlow in analysis	75
6.2.1	Reconstruction of the $t\bar{t}$ invariant mass in Z' decays	75
6.2.2	Identification of boosted top-quarks in Z' decays	75
6.3	Stability of the algorithm	76
6.4	Outlook	76
	Summary	78
	Samenvatting	79
	A Flavour Corrections for PFJets	80
	B Gaussian fit for resolution plots	82
	Bibliography	84

Chapter 1

The Standard Model

The Standard Model of Particle Physics [1, 2] is a quantum field theory that describes the building blocks of our universe and their interactions (except gravity) and was established in the early 70's. The construction of this model was strongly motivated by the quest for a unified theory of all the forces in the universe. All the theories we presently know would then be a special case of this unified theory. The Standard Model is actually not a real unification theory as all the forces it describes have a different coupling constant. This model is basically built up by two theories, the Electroweak theory and Quantumchromodynamics.

1.1 The particles and their interactions

The Standard Model distinguishes two different groups of particles. The first group are the matter particles or the so-called *fermions*. These are all spin-1/2 particles, which means they are bounded to the Pauli exclusion principle. In this group we can distinguish three generations of leptons and quarks which are summarised in Table 1.1. The neutrinos are massless¹ and

	Generation 1	Generation 2	Generation 3
Leptons	e (electron)	μ (muon)	τ (tau)
	ν_e (electron neutrino)	ν_μ (muon neutrino)	ν_τ (tau neutrino)
Quarks	u (up)	c (charm)	t (top)
	d (down)	s (strange)	b (bottom)

Table 1.1: Fermions of the Standard Model

¹Approximation in the theory, as they have a small experimentally determined mass

	Boson	Mass (GeV/c ²)
Electromagnetic force	Photon (γ)	Massless
Weak force	W^\pm and Z^0	80.389 ± 0.025 and 91.1876 ± 0.0021
Strong force	Gluon	Massless

Table 1.2: Bosons of the Standard Model[4]

neutral. The quarks have a charge which is a fraction of the elementary charge (the electron charge). This fraction is $2/3$ for up-type quarks and $1/3$ for down-type quarks. Each of the 12 particles in Table 1.1 also has its anti-particle. These anti-particles have the same mass as their corresponding particle, but an opposite charge.

From the three generations of particles presented, only the first contains the primary building blocks of all the matter in the universe. With the up and down quarks (and their anti-quarks) protons are formed. Together with electrons, they give rise to atoms. The atoms finally give rise to molecules and matter. The two other generations are actually identical copies of the first, but with higher mass.

The second group of particles in the Standard Model are the ones which carry the forces. These spin-1 particles are called *bosons*. These bosons and the forces they mediate are summarized together with their mass in Table 1.2. The first two forces were unified by Weinberg and Salam to the Electroweak theory. Finally the strong force is described in Quantum Chromodynamics which predicts eight gluons that mediate the force between the quarks. This theory introduces a color charge for the quarks (red, green or blue for quarks and the anti-color for anti-quarks). It is this color charge that allows the gluons, which are all colored, to interact amongst each other. In nature, we only see colorless particles which means that bare quarks cannot be observed experimentally. The quarks are always grouped in hadrons, which are then color neutral. If one tries to break up for instance a quark-anti-quark pair, the increased energy-density in the intermediate gluon-field will give rise to a quark and an anti-quark resulting in two color neutral pairs.

1.2 Mathematical formalism

The particles and interactions in the Standard Model are described as a Quantum Field Theory [1]. This means that the theory is based on the minimisation of the action according to the least action principle of Hamilton.

$$\delta S = \delta(\mathcal{L}) \quad (1.1)$$

Where the Lagrange density \mathcal{L} is a function of the particle wave function and the first derivative. In terms of the fermions introduced earlier, the Dirac Lagrange density is given by (putting c and \hbar to unity):

$$\mathcal{L} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi \quad (1.2)$$

In this equation, γ^μ represents the 4x4 Dirac matrices². This Lagrangian density only describes free fermions, so no interactions between fermions and bosons are incorporated in the theory. To include interactions in this framework, it is required that the Lagrangian stays invariant under local phase transformations [1, 2]. The invariance is obtained by replacing the partial derivative in eq. 1.2 by a covariant derivative.

$$\mathcal{D}_\mu = \partial_\mu - ig\frac{\vec{\tau}}{2}\vec{A}_\mu \quad (1.3)$$

In this covariant derivative, $\vec{\tau}$ are the mutually commuting generators of a certain Lie group. Here a new vector field \vec{A}_μ is introduced which brings a new term into the Lagrangian density (eq. 1.2). This term describes the coupling of fermions with this vector field resulting in the following Lagrangian density.

$$\mathcal{L} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi - ig\bar{\psi}\gamma^\mu\frac{\vec{\tau}}{2}\vec{A}_\mu\psi \quad (1.4)$$

The factor g in this equation gives the strength of the coupling of the fermions to the vector field. Consequently the gauge boson interaction term in the electroweak theory is introduced by making the Lagrangian invariant under $SU(2)_L$ for the weak and $U(1)_Y$ for the electromagnetic interaction. The subscript L of the $SU(2)$ group indicates that the weak interaction favors left-handed³ rather than right-handed fermions. The strong interaction is described by the $SU(3)$ group. Taking all this together, the Standard Model is a theory mathematically described by a $SU(3) \times SU(2)_L \times U(1)_Y$ group. Although the Standard Model Lagrangian is nice and simple, it does not allow to easily derive predictions from it. To ease the calculations in the Standard Model, a schematic approach was introduced by R. Feynman. These diagrams, called Feynman Diagrams, visually present the interactions between particles. In this way it is possible to draw the diagram for a certain process and to derive the matrix element from this diagram, using the Feynman rules. Perturbative calculations are then performed by putting the Feynman diagrams in a series with a rising number of vertices (rising power of the coupling constant).

² $\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu}$

³Left-handed spinors transform as $\psi_L = \frac{1}{2}(1 - \gamma^5)$ while right-handed spinors transform as $\psi_R = \frac{1}{2}(1 + \gamma^5)$. ($\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$)

1.3 Problems in the Standard Model

Even though the Standard Model survived numerous of experiments and provides an accurate theoretical frame which can be used to calculate processes and derive predictions, it has some flaws. Because left- and right-handed spinors transform differently under the $SU(2)_L \times U(1)_Y$ group, a mass-term like $m\psi\bar{\psi}$ in eq. (1.4) is not invariant under this transformation. To preserve the invariance of the Lagrangian, all the masses need to be set to zero. This means that the Standard Model is unable to make any statements about the mass of the particles it describes, while the fermions and some gauge bosons do have a non-zero mass.

This problem was addressed by R. Brout, F. Englert and P.W. Higgs resulting in the Higgs-mechanism [3, 5]. This mechanism ensures the mass of the W and Z bosons by introducing a *spontaneous symmetry breaking* in the $SU(2)_L \times U(1)$ gauge. The easiest way to break this gauge symmetry is to introduce a complex scalar doublet (the Higgs-field) into the Lagrangian:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \quad (1.5)$$

Doing this, the Lagrangian takes the following form:

$$\mathcal{L}_{Higgs} = (\mathcal{D}^\mu \phi)^\dagger \mathcal{D}_\mu \phi - \underbrace{\mu^2 \phi^\dagger \phi - \lambda^4 (\phi^\dagger \phi)^2}_{V(\phi)} \quad (1.6)$$

In eq. (1.6), \mathcal{D} is the covariant derivative for the $SU(2)_L \times U(1)$ gauge, μ represents the mass parameter introduced by the mechanism while λ is the strength of the Higgs boson field self interaction. By putting μ^2 to be smaller than zero, the symmetry is spontaneously broken as the minimum of the potential $V(\phi)$ is no longer unique. Using a unitary gauge the doublet can be rewritten as

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \quad (1.7)$$

In this gauge, v represents the vacuum expectation value and $h(x)$ is the actual Higgs-field expressed as a fluctuation of v . Now out of the four degrees of freedom only one remains because the other three are absorbed by the vector bosons which become massive. Thus this mechanism gives mass to the gauge bosons and to the fermions by couplings to the Higgs-field⁴ and predicts a new neutral massive particle, the *Higgs boson*.

The Higgs mechanism is a good candidate to predict the mass of all the particles in the Standard Model, but the Higgs Boson has not yet been observed

⁴Which unfortunately adds extra parameters to the model

experimentally. The only experimental information on the Higgs Boson is provided by previous experiments, like LEP. These were able to put limits on the Higgs mass. The current lower limit for m_H is $114.4\text{GeV}/c^2$ (95% CL)[6] (see Figure 1.2).

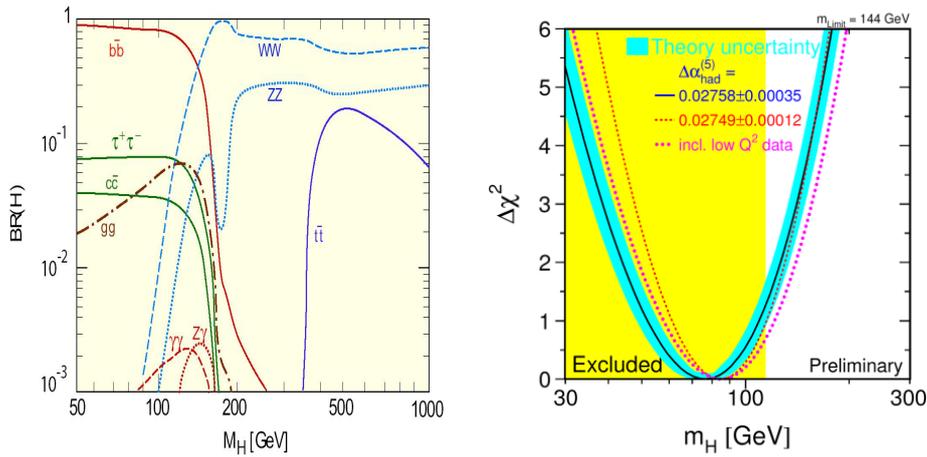


Figure 1.1: Higgs boson decay modes. Figure 1.2: Higgs boson mass limits.

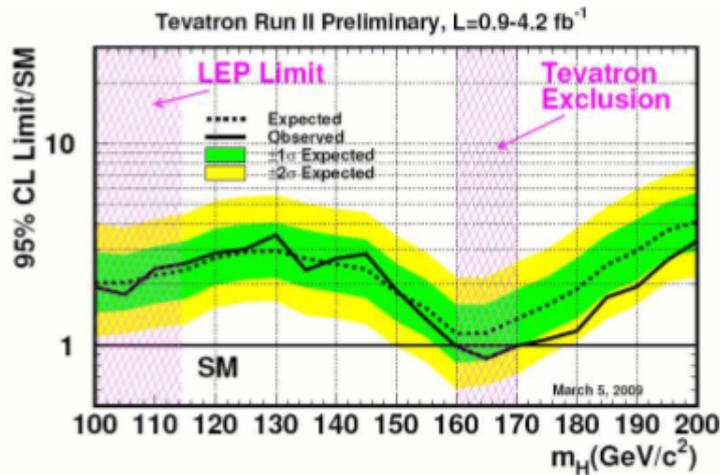


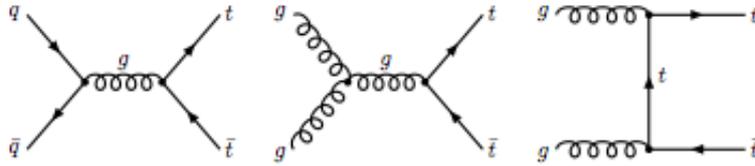
Figure 1.3: 95% CL limits on the ratios to the SM cross section as function of the Higgs mass for the combined CDF and DØ data [7].

Figure 1.3 shows the recent exclusion of the $160\text{-}170\text{ GeV}/c^2$ mass region for the Standard Model Higgs boson mass [7]. This range was excluded by combining the analyses from both the CDF and DØ experiment at the Tevatron

collider. Both analyses were *Higgs Mass Higgs Searches* ($m_H > 135\text{GeV}/c^2$) using the $H \rightarrow W^+W^-$ decay channel.

Other problems within the Standard Model also arise:

- Gravity is not incorporated in this model.
- There is a clear asymmetry between matter and anti-matter in the universe. This is not incorporated in the theory.
- Why does the model include 19 parameters and why do the coupling constants take the values they do?
- Why are there three generations of fermions?
- What is Dark Matter? Experiments indicate that Dark Matter is not constructed with Standard Model particles.

Figure 1.4: Leading Order Feynman diagrams for $t\bar{t}$ -production

1.4 Possible extensions of the Standard Model

Among many extensions (e.g. extra dimensions) to the Standard Model, the most popular one is SuperSymmetry. This theory predicts a new symmetry between fermions and bosons stating that for each boson in nature you have a fermionic partner and vice versa. Also SUSY manages to take down the Planck scale to the TeV-scale which brings it much closer to the electroweak symmetry breaking scale and manages to produce a stable Higgs field. Among many new particles introduced by this extension is the so-called Lightest SuperSymmetric Particle (LSP) which might be the key in understanding Dark Matter. There are high expectations for SUSY particles to appear at TeV-colliders.

1.5 Top Quark Physics

The top-quark is the heaviest quark in the Standard Model. This quark was only discovered in 1995 at the Tevatron $p\bar{p}$ collider because up to then, no collider could produce enough collision-energy (1.8TeV during Run-I and 1.96TeV during Run-II) to create a particle of such large mass. The next collider that will be able to produce this particle will be the Large Hadron Collider (see section 2.1.1). All the experimental knowledge about the top-quark thus originates from the CDF and D0 experiments. The current "World average" of the top-quark mass, obtained from both CDF and D0 data, has a value of $170.9 \pm 1.8 \text{ GeV}/c^2$ [8]. The top-quark is produced in strong interactions, via $q\bar{q}$ -annihilation or gluon fusion. The Leading Order Feynman diagrams for $t\bar{t}$ -production are shown in Figure 1.4. Each top-quark will then decay to a W boson and a b quark. The decay of a top-quark into a W boson and a down-type quark is also allowed by the Standard Model, but this decay is rather unlikely[9].

$$R = \frac{\mathcal{B}(t \rightarrow W^+b)}{\mathcal{B}(t \rightarrow W^+q)} = 0.97_{-0.08}^{+0.09} \quad (1.8)$$

	Cross-Section (pb)
Tevatron 1.96 TeV	$7.01^{+0.58}_{-0.83}$
LHC 10 TeV	415^{+39}_{-25}
LHC 14 TeV	$854^{+105.5}_{-104}$

Table 1.3: NLO cross-section for top-pair production [23] calculated for a top-mass of $172\text{GeV}/c^2$

Each W boson will then decay into a lepton and a neutrino ($\mathcal{B} = 1/3$) or in a quark and an anti-quark ($\mathcal{B} = 2/3$). In the decay of a top-quark pair, three main decay channels can be distinguished based on the decay mode of the both quarks. If both top-quarks decay into a quark and an anti-quark, we call this decay *fully hadronic*. If instead one top-quark decays leptonically, the decay is called *semi-leptonic*. Finally, the both top-quarks can decay into a lepton and a neutrino which is called a *di-leptonic* decay. The branching ratios for these three decays are obtained by multiplying the two contributions, e.g. $\frac{2}{3} \times \frac{2}{3} = \frac{4}{9}$ for fully hadronic decays. The branching ratio of the di-leptonic and semi-leptonic decay equals $\frac{1}{9}$ and $\frac{2}{9}$ respectively. The decay modes are summarised in Figure 1.5. In this figure, it is also shown that in the semi- and di-leptonic decay, the probability is the same for the three generations of leptons. The top-quark is an important particle at the LHC because it will be produced in large numbers due to its large cross-section. A theoretical estimation of the NLO⁵ $t\bar{t}$ -production cross-section at the LHC is given in Table 1.3. As you can see, the cross-section at the LHC is about two orders of magnitude bigger than at the Tevatron, therefore resulting in a much higher production rate of top-quarks. This allows for a more precise measurement of its mass and other properties. Even at startup, when the LHC collision-energy will be around 10TeV, the production cross-section will still be high enough to collect a large number of top-quarks.

There are plenty of reasons to look at the top-quark and some of these aspects will be briefly discussed here. One of the nice things about the top-quark is that due to its very short lifetime (about 10^{-25}s [10]) it will decay before it can form a jet. This makes the top-quark the only quark which can be studied in a detector as a "free" particle. Another aspect is the indirect measurement of the SM Higgs boson mass. Here the top-quark can be useful because the strength of coupling of a fermion to the Higgs boson field is proportional to the fermion mass. As the mass of the top-quark is quite

⁵Next to Leading Order re-summation calculation

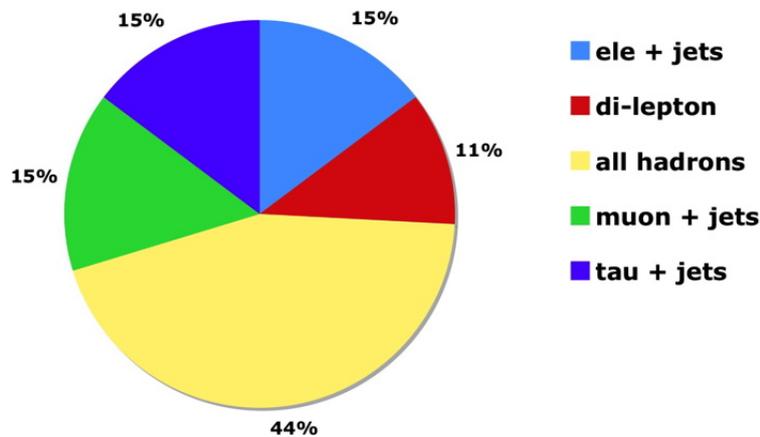


Figure 1.5: Branching ratios of the $t\bar{t}$ -decay

large, the Yukawa coupling is very close to unity which makes the prediction for the most probable Higgs mass strongly dependent on the mass of the top-quark. Furthermore the mass-dependent couplings of various extensions of the Standard Model make the heavy top-quark very useful for searches beyond the Standard Model at TeV hadron colliders. As some resonances will decay into a top-pair⁶, resonant production mechanisms can for example be found by looking at the invariant mass distribution of the top-pairs. On the other hand the top-quark can form a potential background for searches for new physics and thus make some analyses, like the search of the Higgs boson in the $gg \rightarrow t\bar{t}H$ channel, more difficult.

There is a certain potential for the experiments at the Large Hadron Collider (section 2.1.1) to use the top-quark to calibrate the detector. This is now possible because of the large statistics of top-quarks that will be collected as a result of the large collision energy.

⁶e.g.: $Z' \rightarrow t\bar{t}$

Chapter 2

Physics at the LHC

2.1 The Large Hadron Collider

Because of the various problems encountered in the Standard Model and the quest for experimental validation of the extensions to it, it is necessary to explore particle physics on an ever increasing energy-scale. Up to now, the Tevatron collider at Fermilab has been the world's most powerful accelerator with its collision energy of almost 2TeV. Unfortunately the D0 and CDF experiments were not yet able to find the Higgs boson or see SUSY particles. This is why an experiment even deeper in the TeV-scale was needed and this led to the construction of the Large Hadron Collider (LHC) [11, 12].

2.1.1 Motivation for a proton collider

The Large Hadron Collider has been constructed by CERN (Organisation Européenne de la Recherche Nucleaire) near Geneva, Switzerland. This large



Figure 2.1: View of the LHC trajectory from the sky.

collider is located in the 27km long tunnel, previously belonging to LEP¹ which crosses the French-Swiss border.

In this collider two proton beams will be accelerated to almost the speed of light and then brought to collision at the interaction points. The proton collisions will be produced at about 14TeV, which is seven times higher than the Tevatron energy. Besides the high collision energy, the LHC is also designed to reach the very high luminosity of $10^{34} \text{cm}^{-2} \text{s}^{-1}$. This makes the LHC the biggest and most energetic accelerator in the world once it starts circulating and colliding beams.

The choice to collide protons with protons is based on synchrotron radiation [21]. The synchrotron radiation is formed by particles that radiate off energy when they pass through a bending magnet. The amount of energy-loss is proportional to their energy and mass in the following way:

$$\Delta E \sim \frac{E^4}{m^4 R} \quad (2.1)$$

This means that in order to minimize the energy-loss ΔE , you either need to use more massive particles or a drastic increase of the bending radius in the circular accelerator (R). This effect only becomes important at high beam energies which excluded the possibility to build an e^+e^- -collider. Due to the very large mass of the proton, compared to that of the electron, the protons have a negligible energy-loss as a result of synchrotron radiation even at the TeV-scale. Another question that might arise is: why collide protons with protons and not protons with anti-protons, like it is the case at Tevatron?

This choice was made because it is not trivial to make a \bar{p} beam. In order to make this beam, you need to fire a proton-beam on a fixed target and then filter the secondary particles for anti-protons and store those. This would not be possible to do for the LHC because the \bar{p} production rate would never be high enough to reach the design luminosity.

The consequence of colliding same-charge particles is that they can't be accelerated in the same beam-pipe as was the case for accelerators like LEP and Tevatron. Hence, the LHC uses two separate beam-pipes crossing each other in the interaction points. Another issue on working with protons in a collider is the fact that these are not elementary particles and the momentum of the proton is distributed among the quarks and gluons (also referred to as partons) it contains. So in reality a proton-proton collision will rather be a parton-parton collision and the center-of-mass energy that will be available to create new particles will be much lower than the proton collision energy of 14TeV (depending on the momentum-distribution of partons inside the

¹Previous e^+e^- -collider

proton). The main characteristic of the collider (from the physics point of view) is the *luminosity*, \mathcal{L} [21].

This property is related to the number of collisions per second and to the number of events we see for a certain process:

$$N_{process} = \sigma_{process} \int \mathcal{L} dt \quad (2.2)$$

Where N is the number of events recorded and σ is the cross-section of the process under consideration. From this formula it is clear that for the discovery of new physics, the luminosity should be as large as possible.

Due to the large luminosity and collision energy, the LHC accelerator opens up a whole new window for discoveries. A few items of the physics potential of the LHC is listed below.

- The Minimal SuperSymmetric Standard Model (MSSM) predicts particles which are in the LHC acceptance.
- The discovery of the Higgs boson (if it exists).
- There is a possibility for the discovery of extra dimensions.
- Study of CP violation in B-physics.
- Study of the quark-gluon plasma.

2.1.2 Design and operation of the LHC

Before the protons can start circulating in the Large Hadron Collider, they go through a number of pre-accelerators in the "accelerator complex" which is shown in Figure 2.2. The protons are produced from hydrogen in the proton source. These protons are then accelerated in the first linear pre-accelerator LINAC2. In the next step the protons are injected in the proton booster. This is an accelerator which has four beampipes on top of each other. From the booster the protons are carried through a transfer line to the Proton Synchrotron (PS). This is one of the earliest accelerators of CERN (put in place in the 50's) which is now used to pre-accelerate the protons. This accelerator brings the protons up to 25GeV and provides the final bunch structure needed for LHC operation. Finally the protons are accelerated in the Super Proton Synchrotron. When the protons are extracted from the SPS and injected in the LHC, they have an energy of about 450GeV. The final acceleration to the nominal energy of 7TeV is done by the RadioFrequency Cavity of the LHC accelerator.

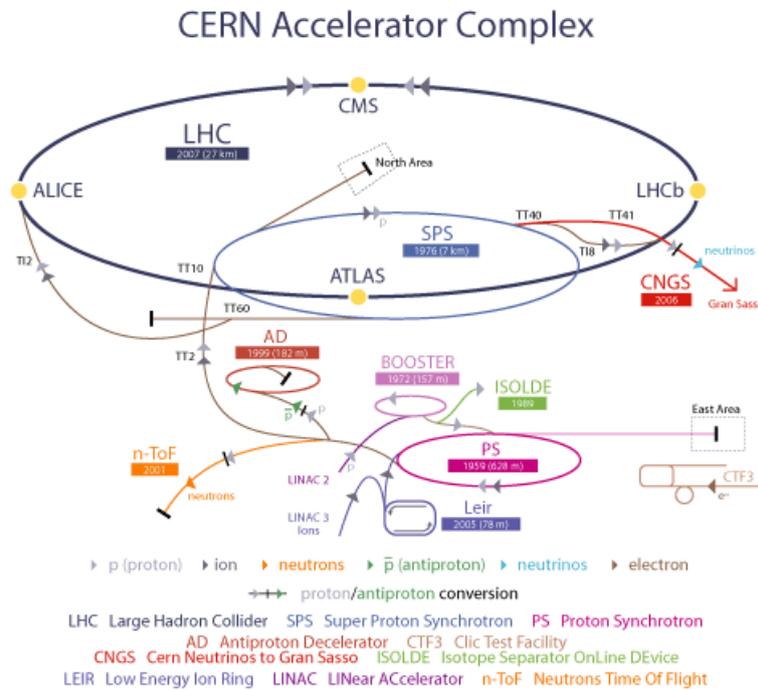


Figure 2.2: The CERN accelerator complex

The proton beam is subdivided into bunches. Each bunch has a minimal diameter of $16 \mu\text{m}$ and contains about 10^{11} protons. In the 27km circumference ring 2808 bunches of protons, separated by 25 ns, will be injected although there are 3546 available bunches. The empty bunches are needed to guarantee safe dumping of the beam in case of problems.

The proton beams are guided through the ring by bending them using dipole magnets (see Figure 2.3). As mentioned before two separate beampipes are needed because both beams contain particles of the same charge. This leads to the development of a single cryostat to house both beampipes and their dipole magnet. The 1232 dipole magnets of the LHC are superconducting because they have to produce a magnetic field of 8.4T to bend these 7TeV particles. These magnets are 15m long and weigh about 35 tons. The cryostat contains a cooling system with liquid Helium which cools the magnet down to 1.9K. This "extreme" cooling is needed to keep the magnets superconducting, with therefore no resistance for the high electric currents. Aside from the dipole, the ring also contains quadrupole magnets. These magnets are used to focus the beam in both the horizontal and vertical plane which results in a beam-diameter of $16 \mu\text{m}$ for the LHC beams. In the vicinity

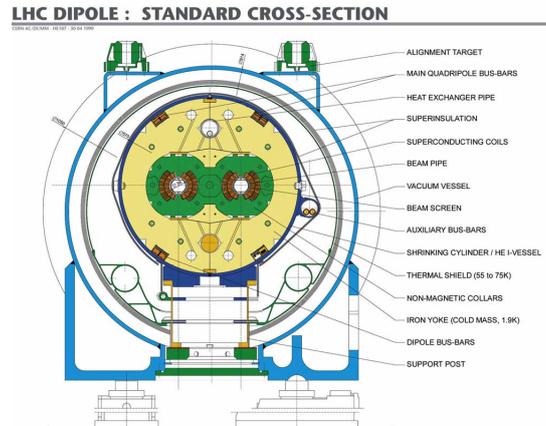


Figure 2.3: A cross-sectional view of the LHC dipole magnet

of the interaction points (IP's), quadrupoles are used to focus the beams to even smaller dimensions. Because of this extra focusing, the particle density in the beam gets bigger, resulting in a higher collision-probability when the beams cross each other. On top of dipoles and quadrupoles, higher order correcting magnets are installed on the accelerator.

On the 10th of September 2008, the first proton beams were circulated through the Large Hadron Collider. This was a historical event as the LHC was ready for duty. Unfortunately, during the commissioning of sector 3-4 of the accelerator (between interaction points 3 and 4) for operation at 5TeV, a faulty electrical connection between the magnets caused a leak in the cooling system. Because of this leak, the magnets started to warm up which increased their resistivity resulting in severe mechanical damage. A total of 53 magnets in this sector needed to be replaced which was completed on the 30th April 2009.

The LHC is set to start colliding beams by the end of 2009. For these collisions, the beams will be accelerated to an energy of only about 5GeV per beam. Afterwards the beam energy will be increased to the nominal value. The first years of LHC operation will be in "low-luminosity" mode where the luminosity will be around $10^{32} \text{cm}^{-2} \text{s}^{-1}$. "high-luminosity" runs are planned afterwards. These runs will be at the design luminosity of $10^{34} \text{cm}^{-2} \text{s}^{-1}$, two orders of magnitude higher than with the low-luminosity run.

2.1.3 Detectors at the LHC

At modern accelerators, there are multiple interaction points around the ring. At these interaction points a particle detector is placed around the beam-pipe to record the particles created by the proton-proton collisions. In the case of the LHC, there are four major detectors installed: CMS, ATLAS, LHCb and ALICE. Below, the physics goals for these four detectors are briefly described along with some technical requirements a detector at the LHC should fulfill.

- CMS [13, 14] (Compact Muon Solenoid): This is one of the two big general purpose detectors. This detector is described in detail in section 2.2. The aim of this general purpose detector is to be sensitive to signals from a wide range of Beyond the Standard Model theories.
- ATLAS [15, 16] (A Toroidal LHC ApparatuS): This second general purpose detector is, with a diameter of 25m and a length of 46m, the biggest detector installed on the LHC. This detector has the same physics programme as CMS. The benefit of having two such detectors is to be able to double-check discovery of new signals. The main difference between CMS and ATLAS is the magnet. While CMS uses a very powerful solenoid field, ATLAS uses an external toroid field accompanied by a smaller inner solenoid.
- LHCb [19, 20] (Large Hadron Collider beauty): The LHCb detector is specialised in B-physics. This detector will, amongst other things, look for the reason of CP-violation in the B-sector.
- ALICE [17, 18] (A Large Ion Collider Experiment): ALICE will look at heavy-ion collisions and try to get some answers about the matter state in the early universe.

The design of the LHC detectors is not straightforward, because some strict requirements must be followed. The requirements on the detector-design originate from the physics programme and from the accelerator properties.

- As mentioned before, the luminosity during LHC operation will (hopefully) reach about $10^{34} \text{cm}^{-2} \text{s}^{-1}$. This means that there are on average 22 collisions per bunch crossing². This high interaction rate demands very fast read-out electronics and a very fast and efficient trigger system. In general the time needed to extract the data from the detector is some orders of magnitude larger than the time between two collisions.

²collisions/bunch = $\sigma_{inelastic} \times \mathcal{L} \times \text{Bunch Separation} \times \text{Fraction of filled bunches}$

- Due to the very high collision rate and the high center-of-mass energy, a lot of radiation is produced inside the detector. This results in radiation damage in the detector components. The aim is to construct the detector with materials and electronics which are the least affected by the high radiation doses which is a difficult task.
- Because of the bunch-structure of the proton-beams, multiple collisions in one bunch-crossing are possible. This effect is called *in-time pile-up*. There is also a pile-up effect rising from the fact that it takes longer to read out the detector than the time between two bunch-crossings. This is called *out-of-time pile-up*. To remove this pile-up the detector needs to be highly granular.
- Given the fact that only 1 out of 10^{12} collisions contains a possible signal for new physics, the detectors must have a very good resolution and be hermetic. Without a hermetic detector and a good resolution for all sub-detectors, it will not be possible to isolate the signal from the much more abundant background processes (for instance QCD events in some analyses). The hermicity of the detector is crucial as some of the theories that will be tested contain MET (missing transverse energy) as a result of the production of non-interacting particles like the neutrino.

2.2 The Compact Muon Solenoid

The CMS detector is one of the general purpose detectors at the LHC. This means that this large machine will be used for various new physics searches and the hermicity of it is of crucial importance. In this way a good resolution for the reconstruction of missing-energy caused by non-interacting particles can be achieved. Furthermore, the CMS detector has a state-of-the-art Electromagnetic Calorimeter necessary for example to detect the Higgs boson decaying into two photons.

A schematic view of the detector is presented in Figure 2.4. Here you can clearly see the cylindrical multi-layer structure that is common for high energy particle detectors. The central part of the cylinder is often referred to as *the barrel*. The discs placed on both sides of the barrel, to have coverage in the more forward region are referred to as *the endcaps*. The detector consists of a tracker, a calorimeter, a solenoid magnet and a muon system. The very powerful superconducting solenoid magnet of CMS is the largest magnet ever built and provides a field of 3.8T. This field is used to bend the particles in the detector in order to measure their momentum. First a short overview of

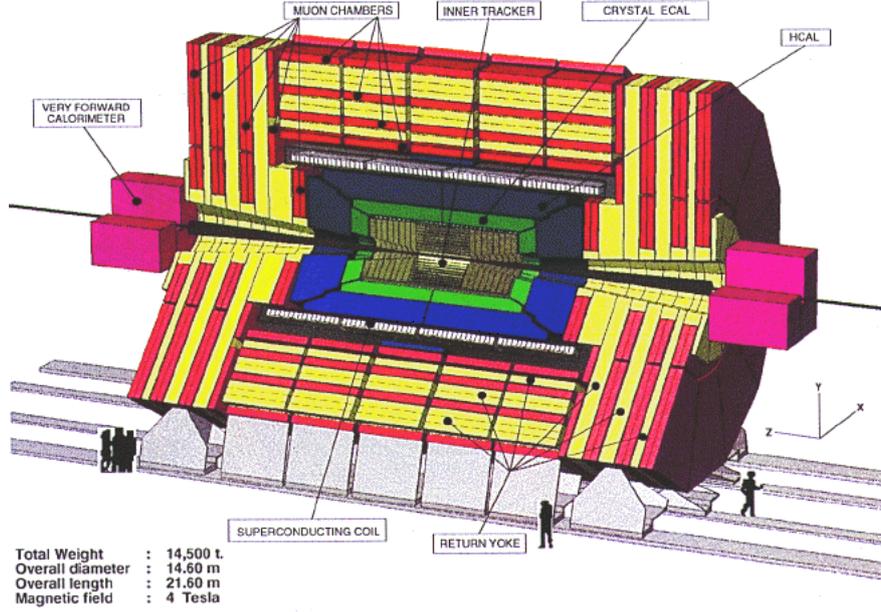


Figure 2.4: Schematic view of the CMS detector.

the coordinate system will be presented. The sub-detectors of CMS will be described in the following sections. The last section of this chapter covers the read-out and trigger.

2.2.1 CMS coordinate conventions

The coordinate system[22] used in the CMS-detector has its origin in the interaction point of the detector. The z-axis is taken along the beam towards the Jura mountains, the y-axis is the vertical direction and the x-axis is taken as the horizontal direction which is pointing inwards to the center of the circle. In general the position of all the parts of the detector is given in terms of η and ϕ . The azimuthal angle ϕ is measured from the x-axis in the x-y plane. The angle η is given by

$$\eta = -\ln \left[\tan \left(\frac{\theta}{2} \right) \right] \quad (2.3)$$

The polar angle θ is measured from the z-axis. As the pseudo-rapidity η is a good estimate of the rapidity, which is invariant under Lorentz boosts, η is preferred over the polar angle θ .

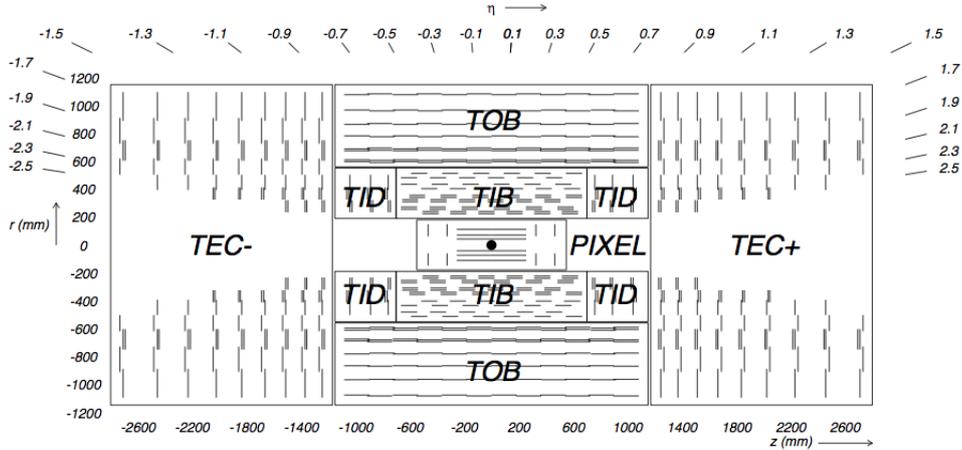


Figure 2.5: Schematic overview of the tracker.

2.2.2 The Tracker

The tracking system is the detector-layer closest to the beam. This sub-detector is responsible for the measurement of tracking points. In this way, the track of a charged particle can be reconstructed later on. Due to the magnetic field and the curvature of the track, the momentum can be measured from the radius of curvature of the particle in the magnetic field.

Because of the large expected particle flux in the proton-proton collisions, the tracker will have to deal with a high overall occupancy. To ensure an effective pattern recognition, the tracker has to be designed so that the occupancy in each channel stays low enough. For operation, the tracker will be cooled down to -10°C to improve the radiation hardness of the sub-detector. The tracker has a diameter of 2.4m and a length of 5.4m and is completely built up with semi-conducting material (silicon). The characteristics depend on the distance to the interaction point. Closest to the interaction point, the pixel detector is installed. This system contains 3 barrel³ layers (TIB, TOB and TID) and 2 endcap disks (TEC+ and TEC-) on each side of the barrel as shown in Figure 2.5. The pixel detector consists of about 1400 modules with 16000 detector chips each. The pixels have a square shape with a dimension of $100 \times 150 \mu\text{m}^2$. This part of the tracker is used for the identification of long-lived particles such as b and c-hadrons and the measurement of their impact parameter. The pixel detector will also be used to reconstruct the primary vertex and secondary vertices, if they appear in the event. The pixel detector has a r - ϕ resolution of $10 \mu\text{m}$ and a z resolution of $20 \mu\text{m}$.

³The barrel covers an η -range $|\eta| \leq 1.4$ and the Endcaps cover the region $1.4 \leq |\eta| \leq 2.5$

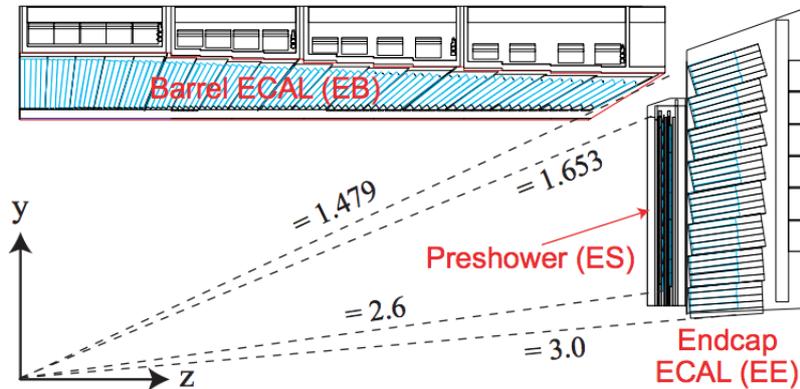


Figure 2.6: The Electromagnetic Calorimeter

In the regions further away from the beam the occupancy is reduced which allows the use of silicon strips. The schematic overview of the tracker modules in this region are shown in Figure 2.5. The barrel layers of the tracker consist of the Tracker Inner Barrel (TIB) and Tracker Outer Barrel (TOB) and the dimensions of the barrel modules and the number of silicon strips they contain depend on the position of the module, which means that the amount of tracker material crossed by the charge particle depends on the direction of the particle. This complex multi-layer design of the tracker assures that the occupancy in all channels stays low enough.

2.2.3 The Electromagnetic Calorimeter

It is shown in Figure 1.5 that, given the current limit on the Higgs boson mass, the Higgs boson di-photon decay is a probable channel for discovery. To be able to detect the two photons in this decay, a very precise Electromagnetic Calorimeter is required. This is why CMS opted to construct a very homogeneous calorimeter using PbWO_4 crystals. This type of crystals was chosen because they are radiation hard and they have a short light-decay time. In general when a particle strikes the crystal, 80% of the light produced, will be collected at the photodiode after 25ns.

The ECAL is crucial for the identification of electrons and photons and a precise measurement of their energy and direction. The ECAL had a resolution of 0.5% for particles of 50GeV. In Figure 2.6, a schematic view of the CMS ECAL is shown. Here you see that again this sub-detector is split in a Barrel part covering the range $|\eta| < 1.5$ and Endcap discs (called Dees) covering $1.5 < |\eta| < 3$. The ECAL Barrel (EB) consists of 61200 crystals grouped into modules. As you can see, the crystals are mounted at an angle of about

3 degrees with respect to the z-axis so that they "look" at the interaction point. This ensures that particles do not escape through the space between neighbouring crystals which would lead to energy-loss. The crystals in the barrel have a dimension of $0.0174 \times 0.0174^\circ$ in $(\eta - \phi)$ -space with a depth of 25.8 radiation lengths.

In the ECAL Endcap (EE), the granularity is lower and the crystals are somewhat larger. This makes the distinction between $H \rightarrow \gamma\gamma$ and $\pi^0 \rightarrow \gamma\gamma$ difficult. This is why a pre-shower detector (ES) is installed, using lead radiators to initiate an electromagnetic shower. After each layer of radiators, silicon strips are used to determine the hit position of the shower.

2.2.4 The Hadronic Calorimeter

The next layer of the CMS detector is the Hadronic Calorimeter (HCAL). This second layer of calorimetry is designed to measure the energy of hadrons through their nuclear interaction in the calorimeter material. This interaction results in hadronic as well as electromagnetic showers. The problem is that even though the energy deposit in HCAL is quite large, the nuclear interaction probability is low. This makes the HCAL less precise than the ECAL.

The HCAL is divided into three parts (see Figure 2.7): the Hadron Barrel (HB), the Hadron Endcap (HE) and the HCAL Outer (H0). The Barrel and Endcap HCAL cover a pseudorapidity-range of $|\eta| < 3$. Both these calorimeters are constructed as sampling calorimeters. This means that these have a series of absorbing layers (iron in this case) interleaved with a layer of scintillating material (plastic scintillator). The HCAL Outer complements this system with an extra layer of plastic scintillators just outside of the magnet coil. Taken together, the HCAL has a thickness close to 10 interaction lengths, good enough to stop most hadrons. For the range $|\eta| < 2$, the HCAL cells have a dimension of 0.087×0.087 in $(\eta - \phi)$ -space. This is roughly equal to the dimension of a 5 by 5 ECAL cluster. Outside this region, the dimension of the cells grow with increasing pseudo-rapidity [22].

In the very forward region the HCAL Forward is positioned extending the coverage to $|\eta| < 5$. These two calorimeters (on both sides of the detector) complement the HCAL in order to ensure that most of the transverse energy is collected. This is crucial for a precise MET measurement.

The scintillation light from the plastic scintillators in the HB and HE is collected by photodiodes. Another technology is needed for HF. For HF, a steel absorber is used. Due to the high radiation dose in this area, quartz fibers are used that emit Cherenkov light.

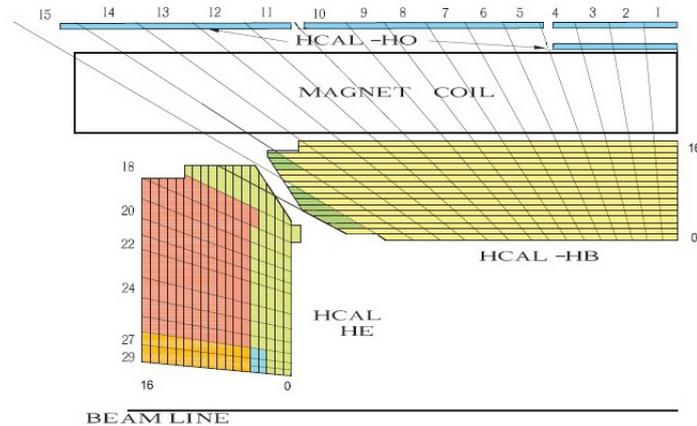


Figure 2.7: Schematic view of the HCAL.

2.2.5 The Muon System

The muon is an important particle for the physics programme of CMS. This particle is able to fly through most of the detector while depositing very few energy. This is why a Muon system is installed outside the solenoid coil as the outermost layer of CMS. The muon stations in Barrel and Endcaps are interleaved with the iron return yokes. This provides the stopping power. These muon stations record tracking hits which are combined as track segments. In a final step, the muon track segments are combined with the track from the central tracker to form a *global muon track*.

As shown in Figure 2.8, the total Muon System in the Barrel has a coverage of $|\eta| < 1.3$ and the Endcaps extend it to $|\eta| < 2.4$. Due to the inhomogeneous magnetic field in the forward region, different technologies are used for Barrel and Endcaps. In the Barrel, Drift Tubes (DT) are used. In the Endcaps it was chosen to use Cathode Strip Chambers (CSC). Both detector types are complemented with Resistive Plate Chambers (RPC), but in the endcaps they cover only up to $|\eta| < 1.6$. Outside this region, the Endcap track segments are only built up by the CSC-hits. In the Barrel the muon system reaches a tracking resolution of $100\mu\text{m}$ and in the Endcaps $80\text{-}450\mu\text{m}$ depending on the station.

2.3 CMS trigger system

The trigger system is a part of the CMS data-acquisition which selects the events that are useful to be stored and the ones that should be dumped. This

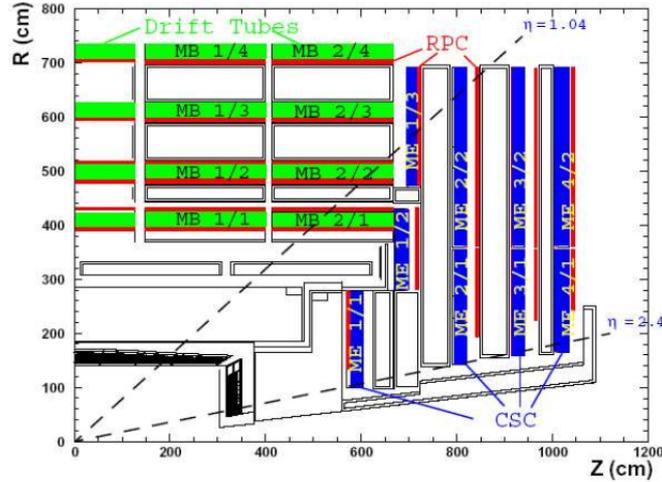


Figure 2.8: Schematic view of the Muon System.

system is necessary in this experiment because a raw event in CMS uses 1MB of storage space and with an interaction rate of 40MHz it is not possible to read all events. This is why CMS implemented two physical trigger levels: The Level-1-trigger (L1) and the High Level Trigger (HLT).

After every collision, the signals from all electronic channels of the detector are grouped into the Front End Drivers (FEDs). These are electronics-crates which contain pipeline memories able to store data from 128 bunch crossings. From the FEDs, the data are carried to the Level-1 trigger and a decision is sent back. As the pipelines only have room for 128 bunch crossings and the interval between them is 25ns, the total time available for the L1 trigger is about $3 \mu\text{s}$. This time includes the time to transport the data and the decision leaving only $1 \mu\text{s}$ for the real decision. Due to this short time-frame, the L1-trigger was constructed with custom-built trigger hardware to enable such fast decision. The Level-1-trigger algorithms have access to data from ECAL, HCAL and the muon system to make a decision. The following objects are used to do a first event selection:

- electrons, photons, muons
- total energy, Missing Transverse Energy
- number of jets (with a user-definable threshold)

In proton-proton collisions, most of the events produced are "minimum-bias events". These events are characterized by low- p_T hadrons while signal-events (like a Higgs boson decay or $t\bar{t}$ -production) often include high- p_T

leptons. These leptons can be used to define a trigger criterion. With the L1-trigger the rate is reduced from 40MHz to 100kHz.

The second trigger step is the HLT. Because the rate has dropped significantly, the HLT uses commercial CPU's. The use of commercial CPU's lowers the cost and improves the flexibility of the system as for these systems, the software environment and C++ algorithms from the offline analysis can be used. This enables HLT to use more advanced algorithms which include the reconstruction of the tracks from the central tracker. After the HLT, the rate is lowered to 100Hz which is acceptable from the storage point of view.

Chapter 3

Jet Reconstruction

In most analyses at the CMS detector, final state quarks are expected. For these quarks, we need to be able to reconstruct their four-momentum. As presented in the first chapter, these quarks are coloured particles which requires them to fragment in color-neutral hadrons (QCD confinement) before they can be detected. This process is called hadronisation and produces a measurable, narrow flow of hadrons which we call a *jet*. The only problem is that in determining parton properties through jets, a lot of detector effects need to be taken into account such as: dead material, cracks between modules, out-of-cone showering, loss of low momentum tracks in the jet due to the magnetic field and detector noise.

In the first section of this chapter, the standard cone algorithm [24] will be introduced. This algorithm is used in this thesis to cluster objects into jets in the calorimeter and at generator level. In Section 3.2, the ParticleFlow method is presented. This is a new¹ algorithm that uses all subdetectors of CMS to cluster objects into jets. These two algorithms will be compared more thoroughly in the next chapters.

In the last section of this chapter, the Jet Energy Scale Corrections will be discussed.

In the next chapters, the CaloJets and the ParticleFlow jets, clustered according to the cone algorithm, will be thoroughly compared in their performance using top-quark events.

¹It has been successfully used in other experiments but it has just been implemented into the CMS software

3.1 Cone Algorithms

The iterative cone algorithm is very straightforward and probably one of the fastest jet clustering algorithms possible. In this section, the algorithm is described, followed by a description of two possible types of input objects. This cone algorithm is called iterative because it consists of a number of steps. These steps are performed in a loop until all the jets are identified. To check whether an input object belongs to a cone or not, the cone algorithm uses the distance in (η, ϕ) -space between the object and the cone axis which should be smaller than the opening angle R of the cone.

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \leq R \quad (3.1)$$

The procedure goes as follows:

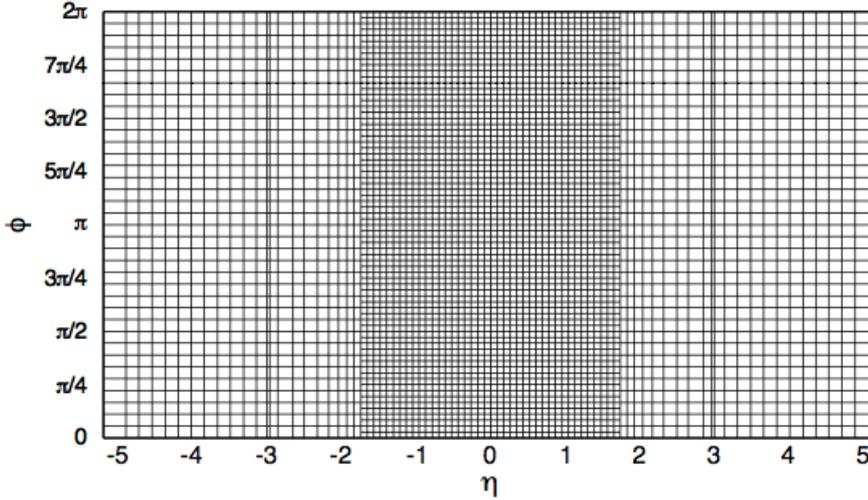
- The input objects are sorted according to their transverse energy (E_T).
- A trial-cone is created with an opening angle R in (η, ϕ) -space around the object with the highest E_T , if it exceeds $E_{T,min}^{seed}$
- The kinematic properties of the proto-jet are calculated using a recombination scheme.
- The E_T -weighted centroid of the proto-jet is calculated.
- If the direction of the trial-cone and the proto-jet is the same, the proto-jet is labelled as a final jet and the contributing input objects are removed from the list
- If this is not the case, this proto-jet is used as a seed for a new iteration.
- The loop continues until there are no input objects with $E_T > E_{T,min}^{seed}$

The recombination scheme used for the jets in CMS is called the Energy-scheme or E-scheme, which combines the jet-constituents as four-vector.

$$p^J = (E^J, \mathbf{p}^J) = \sum_{i \in Jet} (E^i, p_x^i, p_y^i, p_z^i) \quad (3.2)$$

$$p_T^J = \sqrt{(p_x^J)^2 + (p_y^J)^2} \quad (3.3)$$

This algorithm thus takes two parameters: the cone opening angle R and the seed threshold $E_{T,min}^{seed}$. In this thesis two Jet-types are used which are clustered using the cone algorithm, where the opening angle is taken to be 0.5 and the seed threshold 1GeV. The cone algorithm with these parameters is further referred to as the "*iterativeCone5*" algorithm.

Figure 3.1: Granularity of the CaloTowers in (η, ϕ) -space

Calorimeter Jets

The first type of jets used, are the Calorimeter Jets or CaloJets. These jets are clustered from CaloTowers. These CaloTowers are formed by combining towers from ECAL and HCAL. The CMS detector has about 4320 CaloTowers which are shown in Figure 3.1. These CaloTowers are distributed symmetrically in η and ϕ . We see that in the barrel, the granularity is the highest where the towers have the dimension of one HCAL cell (0.087×0.087 in (η, ϕ) -space which equals a 5 by 5 cluster of crystals for the ECAL. If we go towards the Endcaps and forward regions, the HCAL cells get bigger, which lowers the granularity and eventually the quality of the jets. In the vicinity of $|\eta| = 3$ the overlap between HF and the Endcap calorimeters is clearly visible in Figure 3.1. This line represents In each CaloTower the energies of all contributing cells are summed to form the total tower-energy. The only problem is that the detector-noise can produce CaloTowers too. Therefore a threshold is applied on all calorimeter cells in order to remove part of the noise. This threshold is obviously larger for the HCAL than for the ECAL. For HCAL, the threshold is set to 0.9 GeV for HB, 1.4 GeV for HE and 1.1 GeV for HO. For ECAL, the threshold is applied to the sum of the energies of all cells contributing to the tower. This threshold is set to 0.2 GeV for EB and 0.45 GeV for EE.

Generator Level Jets

Ideally the properties of the Jet should exactly match these of the quark, but unfortunately this is usually not the case due to detector effects². This is why we use Generator Level Jets or GenJets as a benchmark for studies on Jet Performance. In this thesis, these jets are clustered according to the iterativeCone5 algorithm. In this case, the algorithm clusters the final-state Monte-Carlo particles³ at the vertex. This means that the magnetic field inside CMS has no effect at all on the GenJets. Thus within a given clustering algorithm, a GenJet provides the best possible approximation of the original parton. The fact that these jets are clustered at the vertex, makes them independent of the detector itself.

3.2 ParticleFlow reconstruction of Jets

The ParticleFlow method [25] is currently under development for the CMS experiment. This method aims to identifying all stable particles in the event by combining information of all the CMS subdetectors. The algorithm produces a list of particles which will be used in this case to build Jets in the same way as Generator Level Jets. The benefit of such method is that the particles are clustered into jets at the vertex when their momentum and direction is not affected by the strong magnetic field of the superconducting solenoid. This is why the ParticleFlow Jets, from now on referred to as PFJets, are expected to have a better energy resolution as well as a better angular resolution compared to the Calorimeter Jets. This will be checked in the next chapter.

The ParticleFlow method is not only useful to provide a good input for jet algorithms, it can also assist in determination of E_T^{miss} as well as b-tagging and τ -identification.

On all fronts, the performance of the CMS detector suits the use of ParticleFlow.

- The large silicon tracker can be used to reconstruct particle-tracks with a good precision and low fake rate down to very low momenta.
- The homogeneous ECAL has a high granularity which allows to distinguish photons from charged particles.
- The HCAL has a much lower granularity than the ECAL which causes the neutral hadrons and the charged hadrons to end up in the same

²e.g. due to electronic noise and cracks between calorimeter modules

³Where you have to remove the undetectable particles!

cluster. Due to its good energy resolution it can still be used to isolate neutral hadrons by measuring the energy-excess of a charged-hadron cluster compared to its associated tracks.

The ParticleFlow algorithm consists of three large steps. The first step is to acquire the *building bricks* from all the subdetectors: tracks, calorimeter clusters and muon track segments. When all this information is gathered, a link algorithm will combine the information from the subdetectors which belong together into *PFBlocks*. These blocks are sent to the particle identification and reconstruction step which will produce the final list of particles in the event.

To be able to deliver the building bricks with high efficiency and low fake rate, an additional iterative track finding and cluster algorithm was developed. These will be briefly discussed here.

Iterative Tracking

The iterative tracking algorithm consisting of five steps, was implemented for the ParticleFlow method. This algorithm is designed to provide high tracking efficiency at the lowest possible fake rate. In the first step very tight criteria are used to seed and reconstruct the tracks, which provide very low fake rate but also a quite low track finding efficiency. This is solved in the next iteration step where the hits that are unambiguously related to the tracks from the first step are removed from the list. The track finding then continues with loosened criteria which enhances the efficiency while keeping the fake rate at the same level.

The fourth and fifth step relax the constraint on the vertex position, allowing secondary vertex reconstruction due to photon emission, nuclear interactions in the tracker material or the decay of long lived particles. This iterative approach allows to reconstruct particle tracks with a p_T as low as 150MeV/c and with vertices more than 50cm away from the primary vertex. During the first three iterations, the tracking efficiency of hadrons in jets is as high as 90% even rising to 99.5% for isolated muons.

It is clear that the central tracker is the center-piece of the ParticleFlow reconstruction because of the superior momentum-resolution compared to the calorimetry and the ability to provide the direction of the particle without interference from the magnetic field.

Calorimeter Clustering

The clustering at the calorimeter has multiple purposes.

- Measure the energy and position of neutral hadrons and photons.
- Separate neutral hadrons from charged ones.
- Complement energy measurement for charged hadrons which have poorly reconstructed tracks (due to low-quality or very high p_T).
- Identify electrons and Bremsstrahlung photons they emit while crossing the tracker material.

Therefore a specific clustering procedure was developed for Particle Flow which acts on HCAL, ECAL and the PreShower (PS) detector separately, but not on HF. The clustering is performed in the following steps. First the calorimeter cells with an energy above a certain threshold are taken as Cluster Seeds. With these seeds, topological clusters are formed and calorimeter cells are added to the cluster when they have at least one side in common with a crystal already in it. As an extra requirement, each cell, candidate for addition to the cluster, must have a signal which is at least two standard deviations larger than the expected electronic noise. When all the topological clusters are formed, each cluster Seed is assigned to a PFCluster. This means that every topological cluster contains as much PFClusters as it contains Cluster seeds. The energy in the topological cluster is shared among its PFClusters, proportional to $e^{-\frac{d_{ij}^2}{R^2}}$, where d_{ij} is the distance between PFCluster i and cell j . Then the position of the PFClusters is recomputed as the center-of-gravity of the five or nine central cells. This process is iterated until the position of the PFCluster is stable. For the ECAL, an extra correction to the PFCluster position is added because the crystals are tilted.

In general, a particle can have a track, PFCluster. In the case of a muon also a track-segment from the muon stations is obtained. Finally, these "building bricks" need to be combined by a *link algorithm* to be able to reconstruct particles.

Link algorithm

The link algorithm now produces blocks by linking elements. The quality of each link is then defined by the distance between the linked elements. Thanks to the high granularity of CMS, the blocks typically contain only a few elements. The advantage of the method is that, with very complex events, the number of blocks increases. In general the number of elements in the block will remain the same. So even with complex events, the performance of the algorithm doesn't change much.

Track - Cluster link

To make the link between a central track and the calorimeter clusters, the track is propagated from the outermost hit in the tracker to the calorimeters. The propagation is carried out to a depth corresponding to the maximum of a shower profile in the ECAL and the HCAL. In the Endcaps, the track is also propagated to the two layers of the PS. The track and the calorimeter cluster are linked if the propagated track is within the cluster boundaries. The cluster envelope can be enlarged by one cell in each direction due to the uncertainty on the position of the shower maximum, multiple scattering and energy leaks. The link distance is then defined by the distance in (η, ϕ) -space between the extrapolated track and the cluster. In addition, it is possible to collect Bremsstrahlung photons coming from electrons in this way. At each intersection between tracker layers, the tangent to the electron-track is propagated to the ECAL with the same linking as in the previous case.

Cluster - Cluster link

The link between clusters of HCAL and ECAL or PS and ECAL is present if the cluster from the most granular one is within the envelope of the less granular one. This means that the ECAL cluster must fit into the HCAL cluster or the PS cluster has to fit into the ECAL cluster. Like the track-cluster link, the envelope is allowed to grow a bit in size. The link distance is again the distance (η, ϕ) -space between the elements.

Track - Muon Track-Segment link

Finally, the linking is done between tracks from the central tracker and track segments from the muon system. When the global fit between the two tracks results in an acceptable χ^2 , the tracks are linked and we speak of a global muon. When a muon track can be fit to multiple tracker tracks, this leads to multiple global muons. Only the global muon with the lowest χ^2 is retained. In this case, the link distance is defined by the χ^2 value rather than the distance in (η, ϕ) -space as before.

Particle reconstruction and identification

Now that the Particle Flow blocks are built, they are fed to a reconstruction and identification algorithm. This step will build the list of particles and provides a global description of the event. The ParticleFlow Jets are built by feeding this list of reconstructed particles into a jet algorithm. In this last

part, the reconstruction and identification step is briefly outlined.

Electrons tend to give rise to short tracks and lose energy in the tracker as a result of the emission of Bremsstrahlung photons. The tracks are sent to a pre-identification step where the tracks are refit to try to follow the electron path to the calorimeter. The electron is then finally identified by a combination of track and calorimeter variables and labelled as a *ParticleFlow electron*.

If global muons appear in a certain block, their momentum is compared to the momentum determined by the central tracker. If the deviation is less than three standard deviations, the global muon is called a *ParticleFlow muon*.

The remaining tracks in a block now belong to charged hadrons. It is possible that some of these tracks are of poor quality. When such a track is detected, it will be removed from the block only when the relative p_T uncertainty is smaller than the calorimetric energy resolution. About 0.2% of the tracks are removed in this way, but about 10% of these removed tracks actually came from a charged particle. This is not a big deal in ParticleFlow because the particle is not lost, as it is still detectable in the calorimeter.

At this point, the block contains mainly hadrons (charged or neutral) and photons. To identify neutral hadrons, the comparison is made between the momentum of the track (or the sum of the momenta of the tracks) linked to the PFCcluster and the cluster energy. The neutral hadron is then identified as an excess on top of the charged hadron energy. To be able to do this, the cluster energies need to be calibrated. It is possible for a certain track to get linked to multiple ECAL or HCAL clusters. In this case only the closest link is preserved for HCAL. For ECAL, if the extra links come from photons, the links should be dropped to allow photon detection. If the links are caused by fluctuations in the hadronic shower, the links should be kept to avoid double-counting. In general, the links to ECAL clusters are ordered according to the link distance. The links are removed once the total calibrated calorimetric energy (HCAL+ECAL) is larger than the momentum of the track. If the total tracker momentum is larger than the calibrated calorimeter energy by more than three standard deviations, a relaxed muon and fake track search is performed. This is followed by an ordering of the tracks according to the uncertainty on their momentum. Finally, the tracks are removed one by one until the total momentum is equal to the calorimeter energy or there are no tracks left with a p_T -uncertainty above 1GeV/c. The remaining tracks in the block give rise to *ParticleFlow charged hadrons*. In the opposite case, when the calibrated calorimeter energy exceeds the total momentum of the tracks with a relative difference bigger than the calorimeter resolution, the excess can be labelled as *ParticleFlow Photons* or *ParticleFlow Neutral Hadrons*. In general if the excess is bigger than the total ECAL energy, a photon is

identified with this energy, leaving the rest of the excess to a neutral hadron. This favouring of photons in ECAL is justified by the fact that for example in a jet roughly 25% of the energy is coming from photons while only 3% is coming from neutral hadrons. The remaining ECAL and HCAL clusters in the block, which are not linked to any track, are identified as respectively *ParticleFlow Photons* and *ParticleFlow Neutral Hadrons*.

3.3 Jet Energy Scale Corrections

Ideally, a RecoJet would exactly match the GenJet's energy. This is unfortunately not the case as the average energy of a RecoJet is deteriorated by for example detector-noise, out-of-cone showering, cracks in the calorimeter modules or dead detector material. The purpose of Jet Energy Corrections [26] (JEC) is then to correct the four-vector of the RecoJet so it matches on average the GenJet's energy. Although for CaloJets a calibration is already applied on the cells, this is in general not enough. For CMS, a multi-layer correction chain was developed containing seven different levels of correction that can be applied to the jets.

- Level 1: An offset correction is applied to account for electronic noise and pile-up.
- Level 2: A relative correction for variation in jet response with pseudo-rapidity relative to a control region.
- Level 3: An absolute p_T correction for the jet p_T with respect to the particle level in the control region.
- Level 4: An optional correction for the variation of jet-response with electromagnetic energy-fraction. (EMF)
- Level 5: An optional flavour correction towards the particle level for different types of jets (b-jet or light-quark jets).
- Level 5: An optional correction for the underlying event due to soft interactions.
- Level 7: An extra optional correction to parton level.

At this point, these corrections are obtained from Monte-Carlo data by correcting RecoJets with respect to GenJets. A large number of QCD events are simulated with $0 < \hat{p}_T < 4000$ GeV/c. These QCD events are divided into $(E_T^{GenJet}, \eta^{GenJet})$ -bins. In each of these bins, the ratio between $E_T^{RecoJet}$ and (E_T^{GenJet}) is fitted with a Gaussian. Based on the mean value of the fit, a correction factor is derived for each of these bins.

When collision data will arrive, a data-driven method to calculate the energy corrections is available. For more information on this method, please refer to [26].

Chapter 4

Jet Performance in $t\bar{t}$ events

In this chapter, the Calorimeter Jets are compared with the jets clustered from particles provided by the new Particle Flow method (PFJets). The PF-Jets are expected by construction to perform better than the CaloJets which will be checked. In the first section, a jet comparison in terms of resolution and response of the jets is made. In the second section, the performance of both jet types is further checked by combining 3 jets in order to reconstruct the mass of the top-quark. For both studies, the full hadronic $t\bar{t}$ -decay is used mainly to prevent one-particle jets to be clustered from electrons or muons.

4.1 Data generation and detector simulation

All the data used in this thesis results from simulation as no real data is available up to present. The event generation and detector simulation is performed in the following steps.

- A Monte-Carlo generator like Madgraph [27] generates random events for the requested process at parton-level.
- These parton-level events are passed to PYTHIA [29] which performs the fragmentation and hadronisation.
- The last step is a full detector simulation using the GEANT4 software [28]. This software simulates the data-flow leaving the detector as will be the case for real data-taking.

When the raw-data is produced with the full detector simulation, the last step is to perform the reconstruction of this data into physics objects. This reconstruction is performed by the CMS Software Framework (CMSSW).

CMSSW is further used to perform the jet study for this thesis. The data samples used in this chapter contain no background events and no pile-up. These events were simulated using a perfect detector.

4.2 Comparison between CaloJets and PF-Jets

4.2.1 Jet Selection

To be able to make a good comparison between the performance of CaloJets and PFJets, a careful jet-selection must be performed. In this thesis the jets are selected, following the scheme in Figure 4.1. First the six quarks produced in the full hadronic $t\bar{t}$ -decay are selected at Monte-Carlo level. For

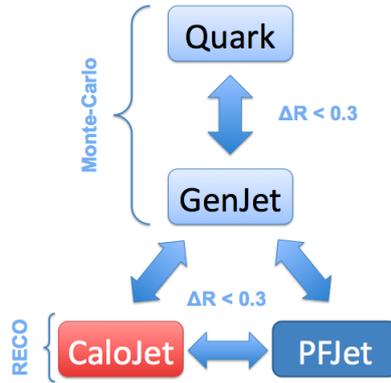


Figure 4.1: Jet selection scheme

these quarks, the corresponding GenJet is sought with the requirement that the distance between the quark and the GenJet is less than 0.3 in (η, ϕ) -space (as defined by eq. 3.1). For each of the six resulting GenJets, a CaloJet and a PFJet match is sought again using a maximum distance of 0.3 in (η, ϕ) -space. If the GenJet cannot be matched to both a CaloJet and a PFJet, the GenJet is discarded. This last requirement ensures that a comparison is made between jets that represent the same quark. Furthermore a final selection is made from the matched jets according to their p_T and η . We did not look at the jet reconstruction efficiency, therefore the motivation for this tripple matching.

The cut on p_T is set to 20 GeV/c as jets not surviving this cut cannot be regarded as jets but rather as artifacts of the algorithm. This cut is applied

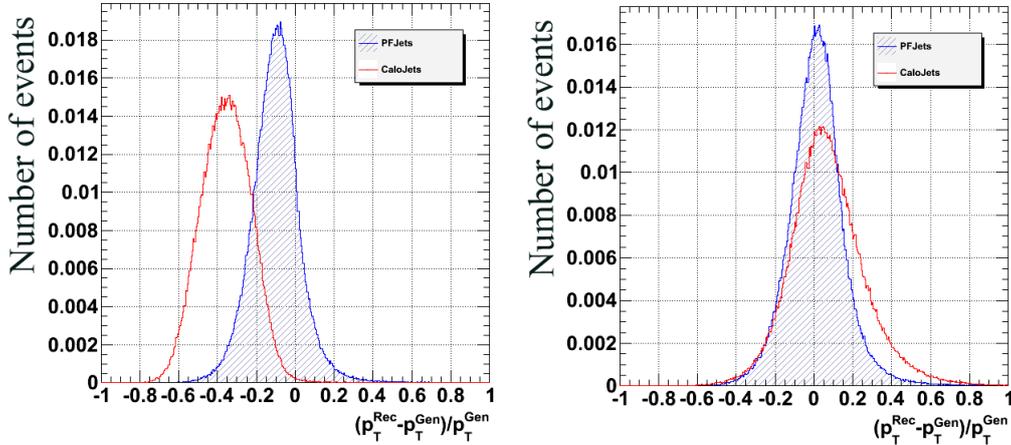


Figure 4.2: The normalized distribution of $\Delta p_T/p_T$ for Uncorrected RecoJets (Left) and L2L3Corrected RecoJets (Right)

after L2L3 corrections have been applied. Finally, depending on the plot in question, a cut on the position in the detector is used:

- Barrel: $|\eta| < 1.4$
- Endcaps: $1.4 < |\eta| < 2.4$
- FullEta (Barrel + Endcap): $|\eta| < 2.4$

To have the maximum (expected) gain from the Particle Flow method, the jets used in this section are restricted to the tracker region ($|\eta| < 2.4$). The tracker actually extends to $|\eta| < 2.5$, but reducing this to $|\eta| < 2.4$ ensures that the jets are fully contained in the tracker region. Consequently, the Endcap region ($|\eta| < 3$) is also restricted to $|\eta| < 2.4$.

4.2.2 Jet Energy Corrections

Before we go into jet resolution and response, it is useful to first look at the jet corrections. In Figure 4.2, the relative difference between the transverse momentum of RecoJets (CaloJets or PFJets) and GenJets is shown which should be as close to zero as possible.

$$\frac{\Delta p_T}{p_T} = \frac{p_T^{\text{RecoJet}} - p_T^{\text{GenJet}}}{p_T^{\text{GenJet}}} \quad (4.1)$$

The distribution for uncorrected CaloJets and PFJets shows that the relative deviation for the CaloJets is roughly 35% on average compared to only 10%

for PFJets. This plot confirms the expectation that in general, the ParticleFlow jets require smaller correction than CaloJets. The plot on the right shows the same distribution for CaloJets and PFJets with L2 and L3 corrections applied (see section 3.3), which strongly shift the CaloJet distribution towards zero. The average $\Delta p_T/p_T$ for CaloJets and PFJets now differs by almost a factor 5 (this is shown in Table 4.1) and we see that the distribution for PFJets is centered closer to zero than for the CaloJets.

In Table 4.1, also the RMS for the $\frac{\Delta p_T}{p_T}$ -distribution before and after JEC corrections is printed. The RMS for the PFJets improves only about 2% with respect to the RMS of the CaloJets when they are uncorrected. After JEC corrections, the relative improvement in the RMS is roughly 25%!

	CaloJets		PFJets	
	RMS	Mean	RMS	Mean
Uncorrected	0.131 ± 0.092	-0.355 ± 0.131	0.129 ± 0.091	-0.099 ± 0.129
L2L3 Corrected	0.195 ± 0.138	0.071 ± 0.195	0.146 ± 0.103	0.015 ± 0.146

Table 4.1: Mean value and RMS of the $\frac{\Delta p_T}{p_T}$ -distribution before and after JEC corrections.

These results show that the PFJets indeed perform better with respect to the CaloJets. Although the $\frac{\Delta p_T}{p_T}$ -distribution is very helpful in making a first comparison between PFJets and CaloJets, it can't tell us anything about the evolution of the mean and RMS of the distribution, for example with the jet p_T .

Finally, it might seem odd that only L2 and L3 corrections are applied. In general L2, L3, L5 and L7 corrections are applied to CaloJets from simulation-data. This is not the case for PFJets, where no L5 or L7 corrections should be applied (see Appendix A). For this reason, only L2 and L3 corrections are applied to treat both Jet types on an equal footing.

4.2.3 Transverse momentum resolution and response

To make a thorough comparison of CaloJets and PFJets, the p_T -distribution of the GenJets is binned. In these bins, the $\frac{\Delta p_T}{p_T}$ -distribution is determined for both types of jets and the mean value (response) and the RMS (resolution) of the distribution is calculated and plotted. This plot allows to see the evolution of the resolution and response with increasing GenJet p_T . The only problem in this method is that the calculated mean and RMS are influenced statistical fluctuations. This is why a fit on the $\frac{\Delta p_T}{p_T}$ -distribution in

each p_T^{GenJet} bin was introduced which will provide an estimate of these two parameters. In Figures 4.3 and 4.4, the distribution in two Barrel and End-

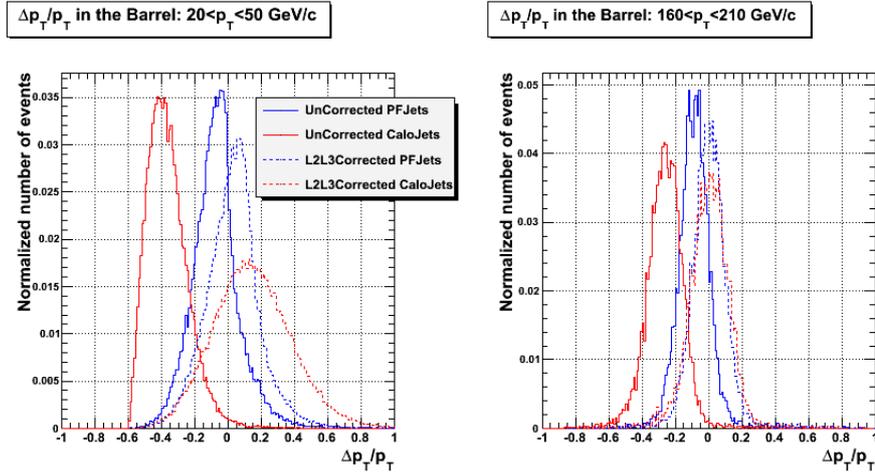


Figure 4.3: The normalized distribution of $\frac{\Delta p_T}{p_T}$ for two different bins of p_T^{Gen} in the Barrel

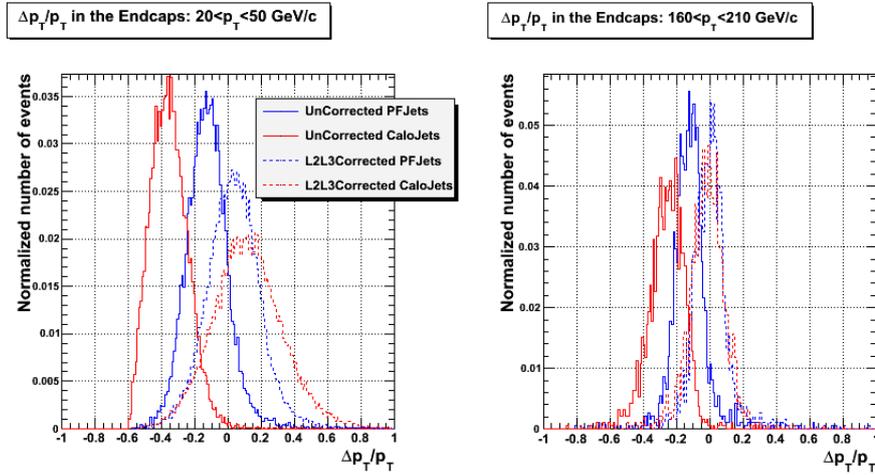


Figure 4.4: The normalized distribution of $\frac{\Delta p_T}{p_T}$ for two different bins of p_T^{Gen} in the Endcaps

cap bins is shown. From this figure, it is clear that a Gaussian fit-function is a good choice. Two Gaussian are fitted in each bin, one from -1 to +1 (including the tails of the distribution) and one only covering 20 bins around

the maximum of the histogram (discarding the tails of the distribution). In general the results obtained by these two fits are in agreement, which supports the choice of a Gaussian fit (see Appendix B). In general, we will use the following terms for the future plots (with obs representing the jet observable under study):

- Jet resolution: σ of $\frac{obs^{Rec}-obs^{Gen}}{obs^{Gen}}$ (Gaussian fit)
- Jet response: μ of $\frac{obs^{Rec}-obs^{Gen}}{obs^{Gen}}$ (Gaussian fit)

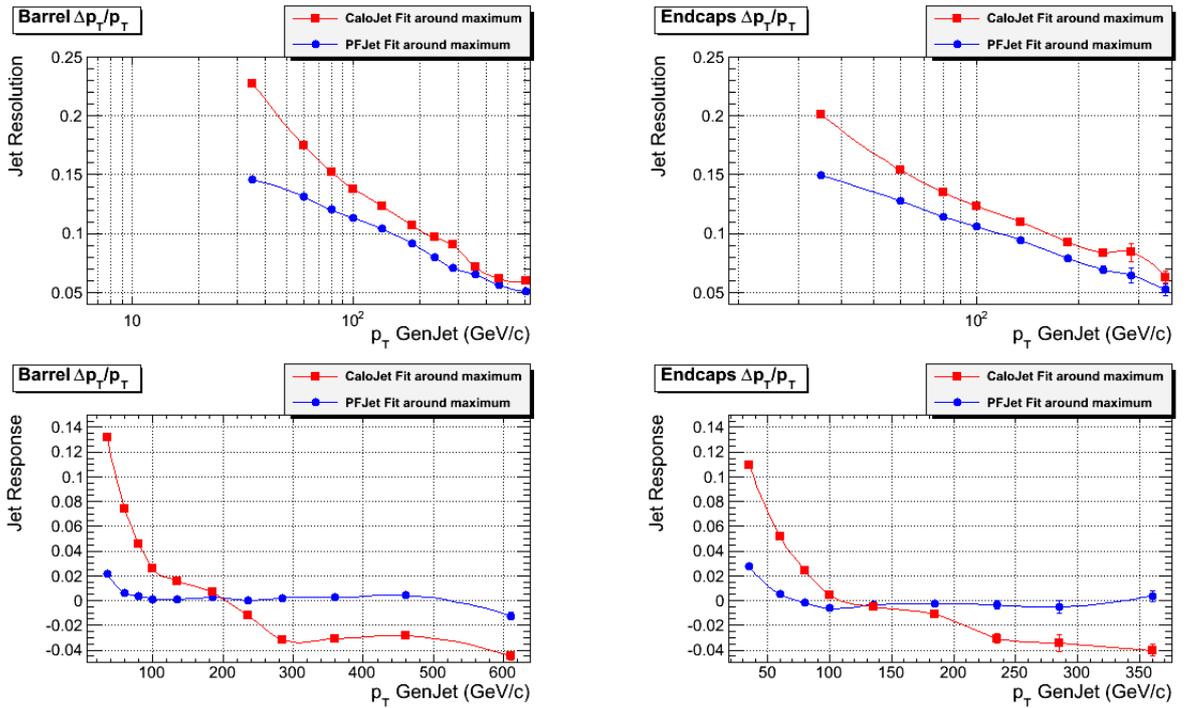
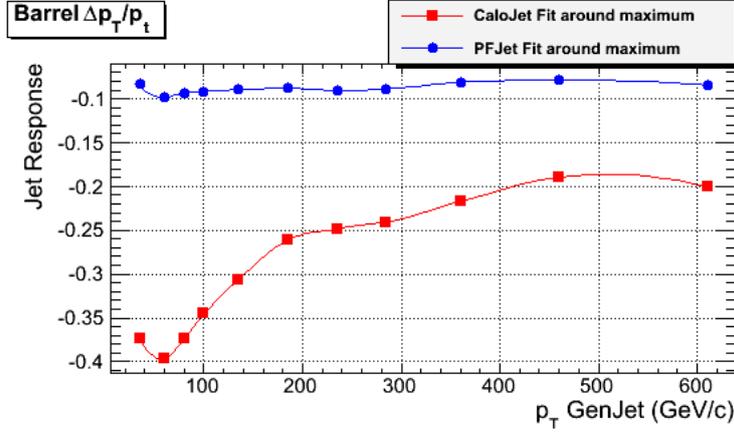


Figure 4.5: p_T resolution and response for corrected jets in Barrel and Endcaps

Figure 4.5 shows the resolution and response in function of p_T^{Gen} corrected Jets in the Barrel and the Endcaps. First of all, it is clear from the jet response plots that the PFJets benefit from the particle identification and reconstruction because only little energy is lost over the whole acceptance. This leads to a nearly flat p_T -response for the corrected PFJets. The CaloJets on the other hand, show a stronger dependence of the response on the jet- p_T especially for a low p_T where the absolute difference in energy-bias between

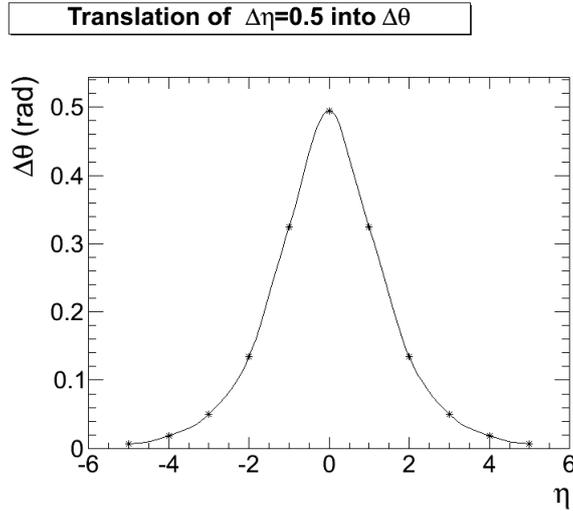
Figure 4.6: p_T response for uncorrected jets in the Barrel

CaloJets and PFJets is roughly 10% in Barrel and Endcaps. In this region, the difference in response becomes quite large because a low- p_T CaloJet can lose some energy as low momentum charged particles will be bent out of the cone by the strong magnetic field. The fact that the response changes sign around 200 GeV/c is just an artifact of the corrections applied to the jets. This is obvious when comparing the response in the Barrel for corrected jets with the same plot for the uncorrected jets (see Figure 4.6). In the latter plot, the largest difference again appears at low p_T and gradually decreasing with increasing p_T . Along with the p_T response, the relative resolution improves with increasing jet p_T . If a quark with a high momentum is produced, the particles resulting from the hadronisation will be very collimated resulting in a narrow jet. This means that particles in jets with a high p_T are closer to each other and usual result in an improved energy resolution. For low momentum jets, the energy resolution is deteriorated because of instrumental effects (e.g.: noise, pedestals,...). This results in a worse p_T resolution for CaloJets and also to a smaller extent for PFJets. At low p_T , the gain in the resolution with PFJets is about 20-40% in the Barrel and 10-25% in the Endcaps. Again, the PFJets benefit from the combination of all subdetectors to reconstruct and identify all particles in the event.

It might seem odd that the range of p_T^{Gen} in the Endcaps is nearly half the range in the Barrel. This choice originates from the definition of transverse momentum.

$$p_T = p \sin(\theta) \quad (4.2)$$

This equation shows that a certain energy deposit ($E=|\vec{p}|$) will result in a lower p_T in the Endcaps than in the Barrel.

Figure 4.7: Translation of $\Delta\eta$ into $\Delta\theta$

Resolution and response as a function of η

In exactly the same way, we can have a look at the evolution of the p_T resolution and response with η . The interesting thing about the η dependence is that we know what to expect. For an uncorrected jet, the resolution and response will be quite steady in the Barrel while improving a bit in the Endcaps. The improvement in the Endcaps can be explained by the fact that the jets are clustered in (η, ϕ) space and η is no linear scale. This results in narrower cones towards the Endcaps as shown in Figure 4.7 where an interval of length 0.5 in η is translated to a θ interval for different values of η . Furthermore we expect to see a small deterioration of the resolution and the (uncorrected) response in the transition region between the Barrel and the Endcaps because of dead material.

Figure 4.8 shows the resolution and response for both corrected and uncorrected jets. For the uncorrected jets, the plots are in good agreement with the predictions. The PFJets have a better resolution over practically the whole η region and a vastly superior response. The corrected jets, show almost the same resolution except for the resolution of the PFJets. This has improved relative to the CaloJet resolution over the whole region. Finally the response for corrected jets is shown in the bottom-right plot in Figure 4.8. Keeping in mind the Jet Energy Scale corrections introduced in Section 3.3, the response is expected to be independent of η . Looking at this plot, it is

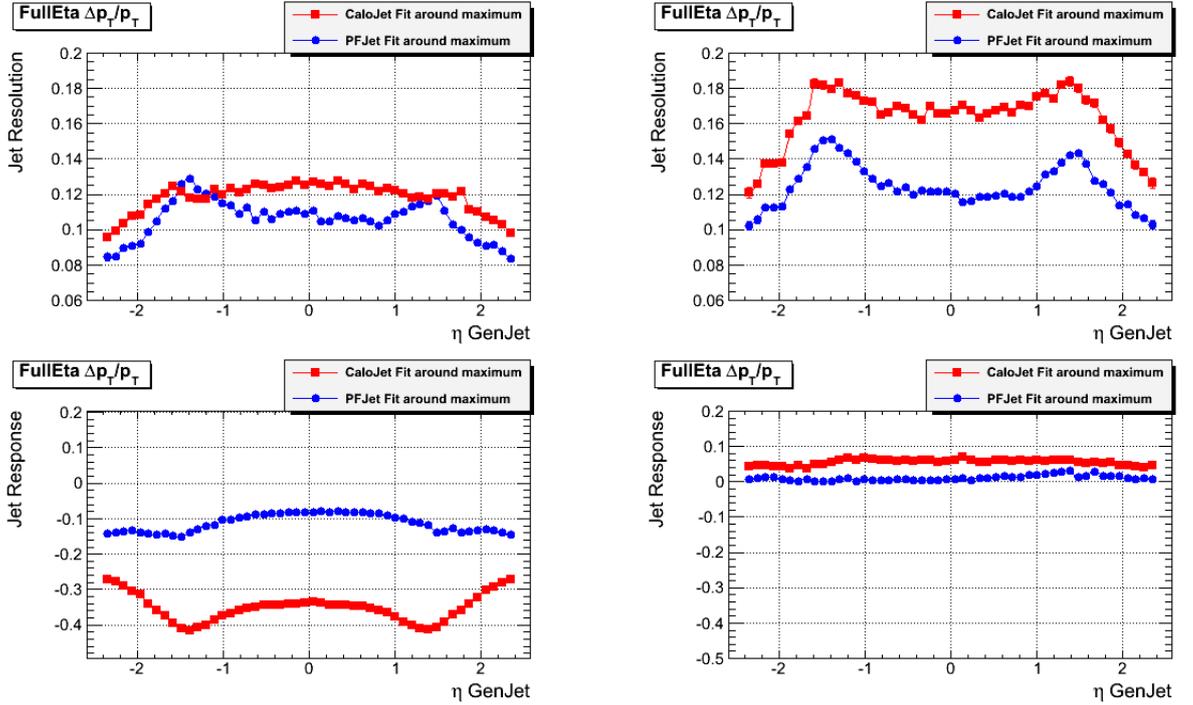


Figure 4.8: p_T resolution and response in function of η for Uncorrected jets (left) and Corrected jets (right)

clear that the response is not flat for either CaloJets or PFJets. To calculate the Level2 JEC correction, a large QCD multi-jet sample is binned according to the jet p_T . For each of these bins, the correction is derived. Because in this plot all p_T bins are used, it is not troublesome that the response is not entirely flat. More problematic is the asymmetry in the response between the positive and negative η regions of the detector, which seems to only occur for the PFJets. To conclude on the origin of this effect, a new plot was created, from the previously shown response. In this plot, the points from $\eta < 0$ are mirrored to $\eta > 0$ which will be called the $\eta^- p_T$ response. The points already in $\eta > 0$ are called the $\eta^+ p_T$ response. This puts both response curves in the same region, which makes comparison easier. The difference between η^- and η^+ should be as close to zero as possible. This is why the subtraction of both curves is shown in Figure 4.9. In this figure it is clear that for CaloJets, the scale of variation is so small that the p_T response can be regarded as symmetric in η . For the PFJets however, we see that the asymmetry not only appears in the response of corrected jets, but also in the response of uncorrected jets. This means that this asymmetry is not an

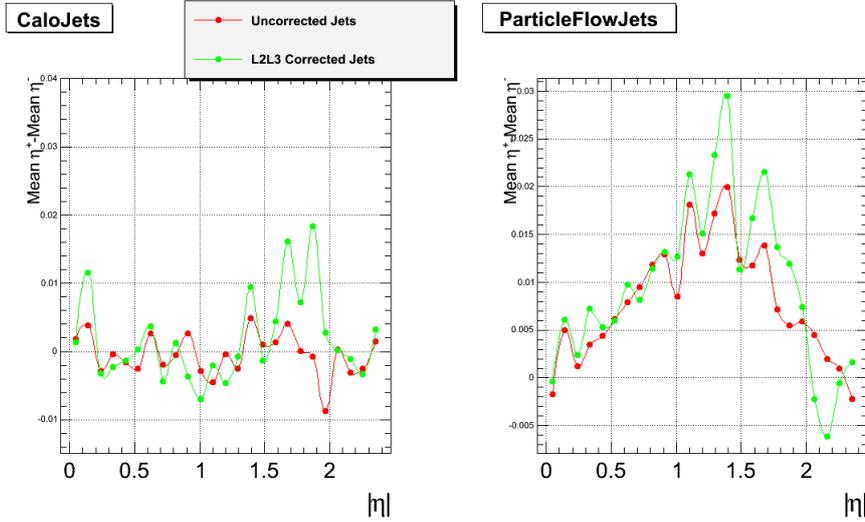


Figure 4.9: Check of the asymmetry in the η dependence of the p_T response.

artifact of the corrections but rather a ”bug” in the algorithm¹. At the time of writing, a new version of the ParticleFlow algorithm was put into place in a pre-release of the CMS software framework². Unfortunately no Jet Energy Scale Corrections are calculated before the new software version is released. This means that with this version, we could only check if the asymmetry is still present for the uncorrected jets. The resolution and response obtained for uncorrected jets with this new version is shown in Figure 4.10. At first sight, the asymmetry seems to have disappeared but a similar plot to the one in Figure 4.9 was made for the ”new” PFJets and is shown in Figure 4.11.

By comparing Figures 4.9 and 4.11, we conclude that the asymmetry in the response of p_T as function of η disappeared in the new software version. Thus we can conclude that the observed feature was only a bug in the code. Unfortunately the origin of this bug is unknown as it disappeared in the new version of the algorithm.

Resolution and response as a function of ϕ

Because we know that there should be no dependence on ϕ of the p_T resolution and response versus ϕ , this is another straightforward benchmark for the jets. Unless some parts of the detector are malfunctioning, we thus expect a flat line for resolution and response. In Figure 4.12, the resolution and

¹Under CMSSW_2_2_7

²Pre-release version: CMSSW_3_1_0_pre4

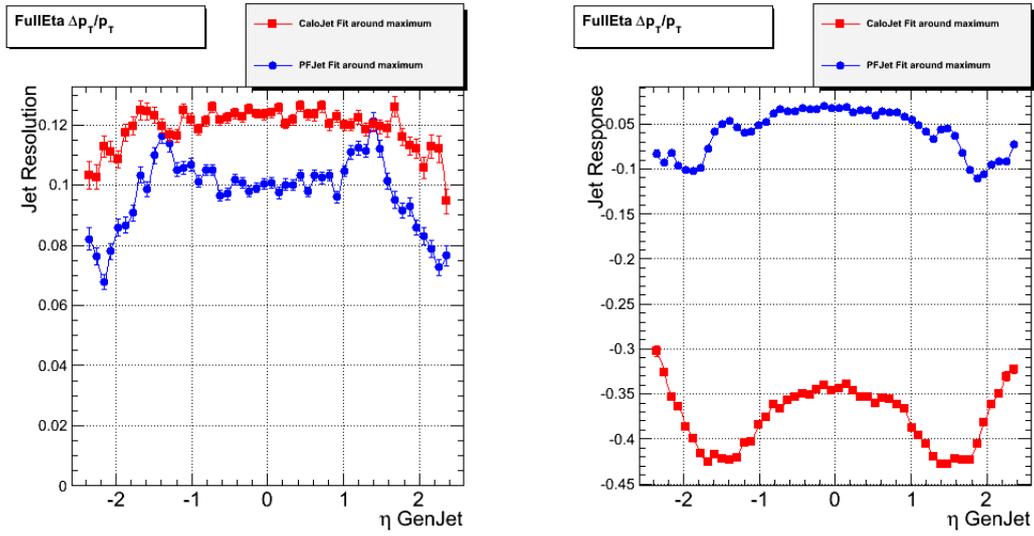


Figure 4.10: p_T resolution and response for uncorrected jets using the pre-released CMS software (CMSSW_3_1_0_pre4).

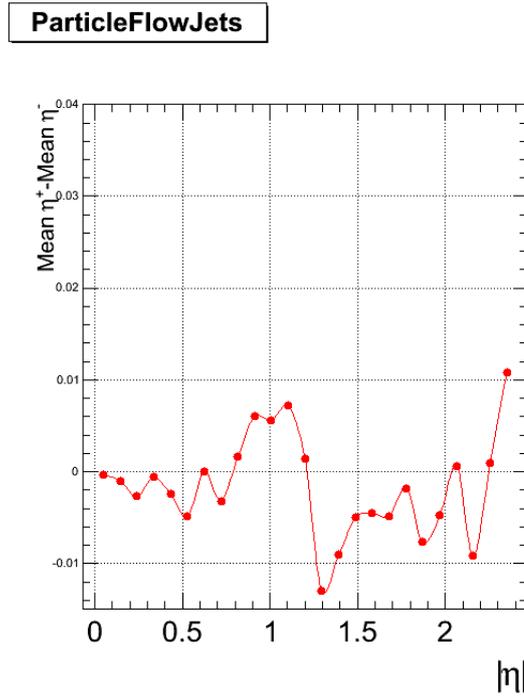


Figure 4.11: Check of the asymmetry in the η dependence of the p_T response using the pre-released CMS software (CMSSW_3_1_0_pre4).

response are shown. We see that for CaloJets, the behaviour is as expected while for PFJets two "bumps" are appearing at $+\frac{\pi}{2}$ and $-\frac{\pi}{2}$. These bumps are not expected and any explanation is lacking. Keeping in mind the pre-

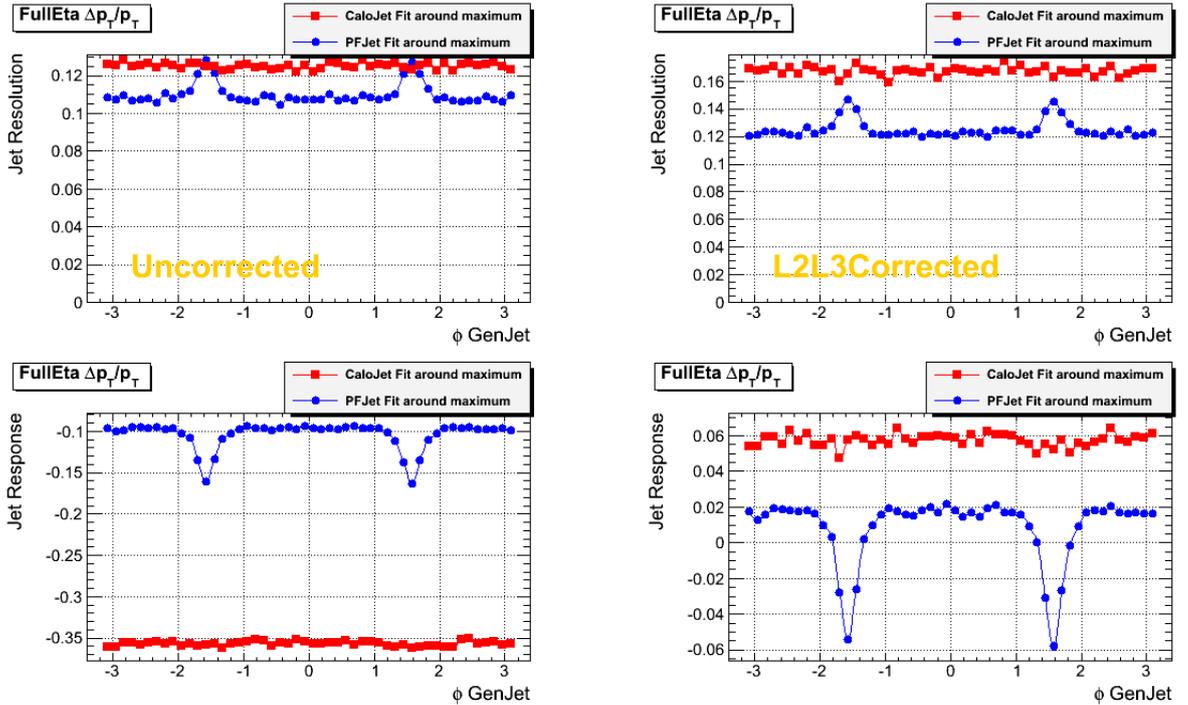


Figure 4.12: p_T resolution and response for corrected jets as a function of ϕ

vious bug in the p_T response as function of η , this resolution and response was also checked with the new pre-released software. Figure 4.13 shows that the bumps observed in Figure 4.12 are again caused by a bug in the ParticleFlow algorithm in the older software. Hence this study had contributed in debugging the reconstruction software deployed by the CMS collaboration.

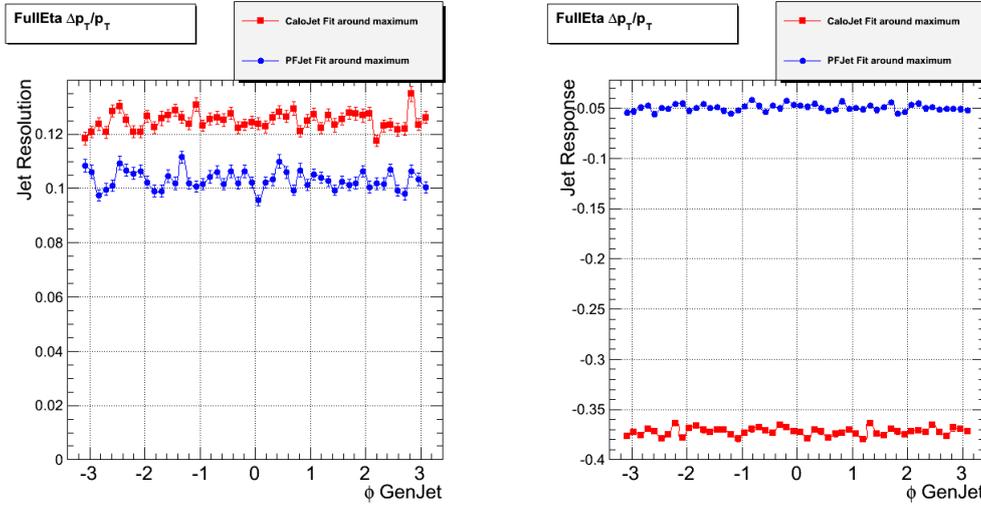


Figure 4.13: p_T resolution and response as a function of ϕ for uncorrected jets using the pre-released CMS software (CMSSW_3.1.0_pre4).

Two other variables were found that either show improved p_T response or resolution for the PFJets: the number of constituents and the charge multiplicity.

Resolution and response as a function of the number of constituents

The same procedure is used as before to show the p_T resolution and response as a function of the number of jet constituents. Because the number of constituents is not defined in the same way for CaloJets (number of towers) and for PFJets (number of particles), we use the number of the GenJet constituents on the x-axis of the plot shown in Figure 4.14 for corrected jets. The resolution for both types somewhat improves with a rising number of constituents, but the improvement of the PFJets relative to the CaloJets stays nearly constant. More spectacular is the response. For jets with a small number of constituents the reconstruction of the ParticleFlow algorithm results in much less energy-loss which improves the response.

Resolution and response as a function of the charge multiplicity

The charge multiplicity of a jet is defined as the number of charged particles inside a jet. For both CaloJets and PFJets this property is obtained by

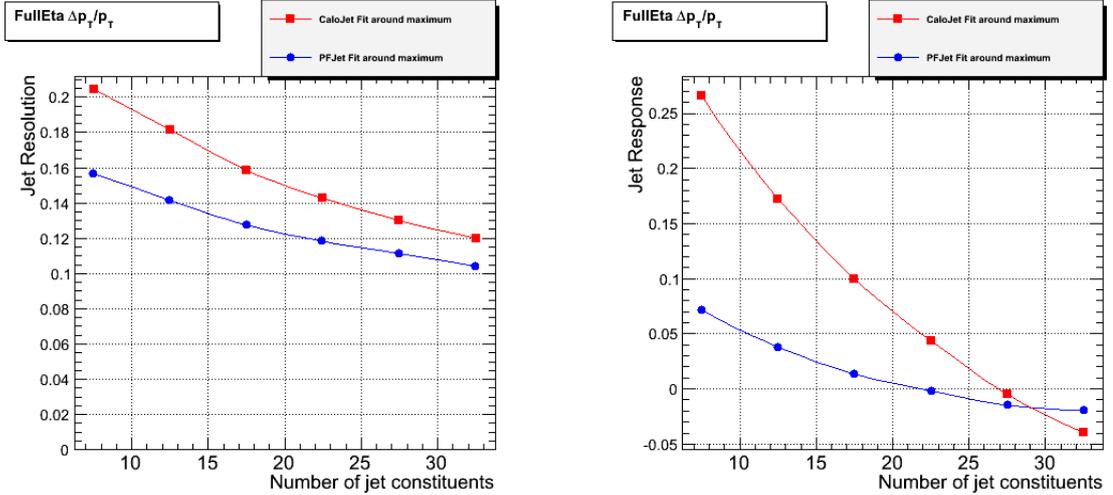


Figure 4.14: p_T resolution and response as a function of the number of jet constituents.

counting the number of tracks associated with the jet. A cut on the p_T of the track is set to $1\text{GeV}/c$ to reject fake tracks. In Figure 4.15 the p_T resolution and response as function of the charge multiplicity is shown. The relative improvement of the PFJet resolution, compared to the CaloJet resolution, is quite constant with the charge multiplicity. This is not the case for the response. The response for PFJets is almost flat while the CaloJet response strongly varies with this observable. If we look at Figure 4.16 where the response is shown for uncorrected jets, we see a much better response for PFJets. In this plot both CaloJets and PFJets are quasi flat. It seems that the jet corrections introduce a dependence on the charge multiplicity for CaloJets. For the PFJets, the corrections are much smaller which might explain why this dependence is not present.

4.2.4 Angular resolution

In the previous part, the p_T resolution and response for PFJets and CaloJets were compared. Because ParticleFlow puts all particles at the vertex, their position should not be influenced by the magnetic field. Hence, a better angular resolution is expected for PFJets with respect to CaloJets. Using the same method as presented in the previous part, it is possible to look at the η and ϕ resolution in function of the jet p_T .

Figures 4.17 and 4.19, show the evolution of respectively the η and ϕ resolution with the GenJet p_T . In both cases it is clear that again the PF-

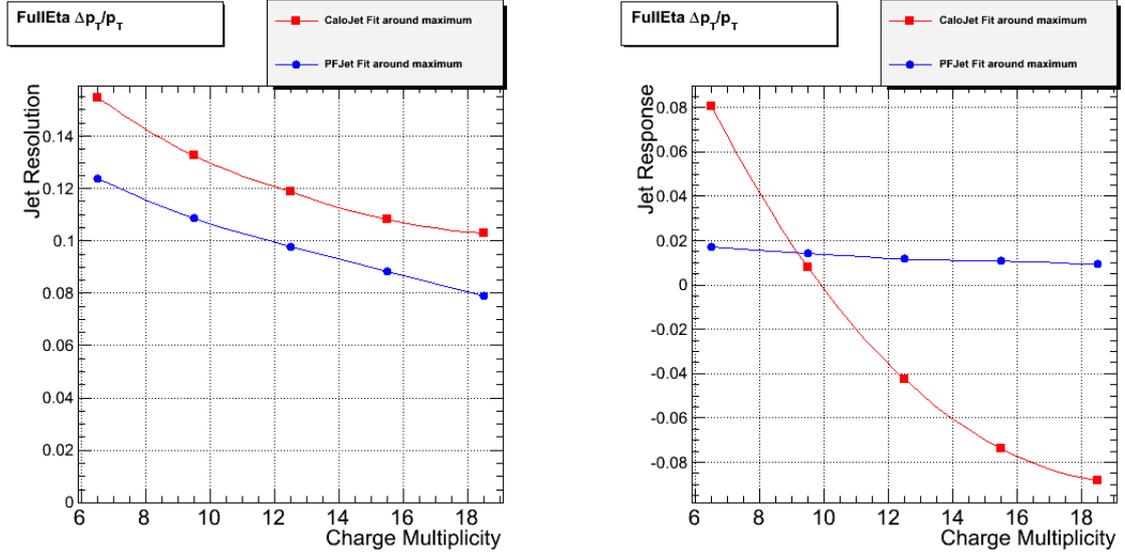


Figure 4.15: p_T resolution and response as a function charge multiplicity.

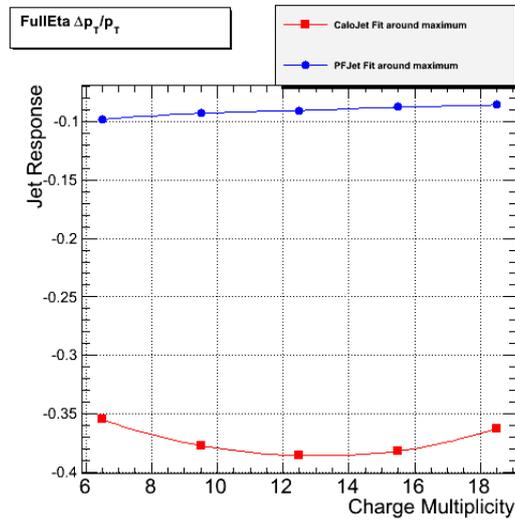
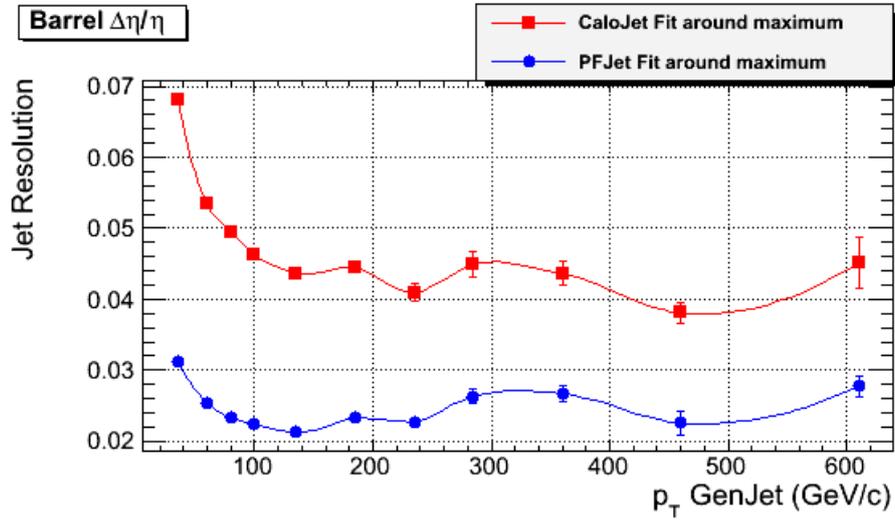
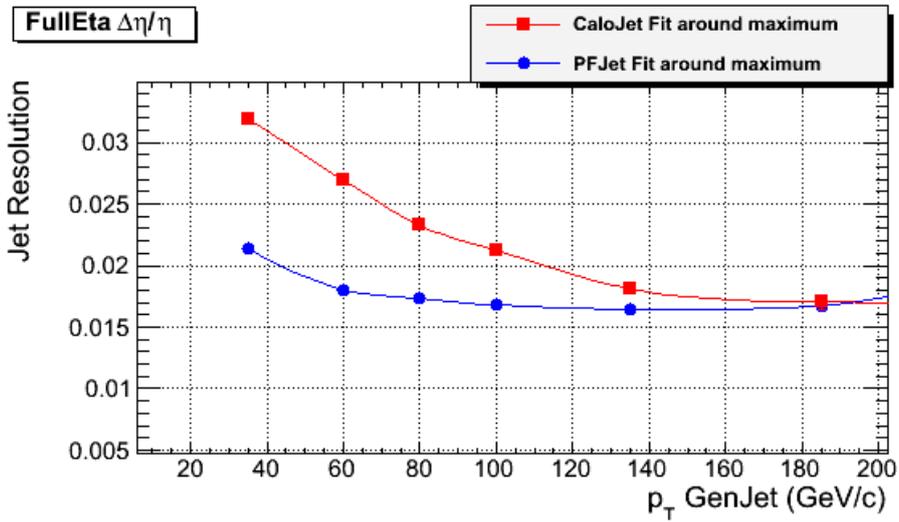


Figure 4.16: p_T response as a function charge multiplicity for uncorrected jets.

Figure 4.17: η -resolution for jets in the Barrel using CMSSW_2_2_7.Figure 4.18: η -resolution for uncorrected jets in the tracker region using CMSSW_3_1_0_pre4.

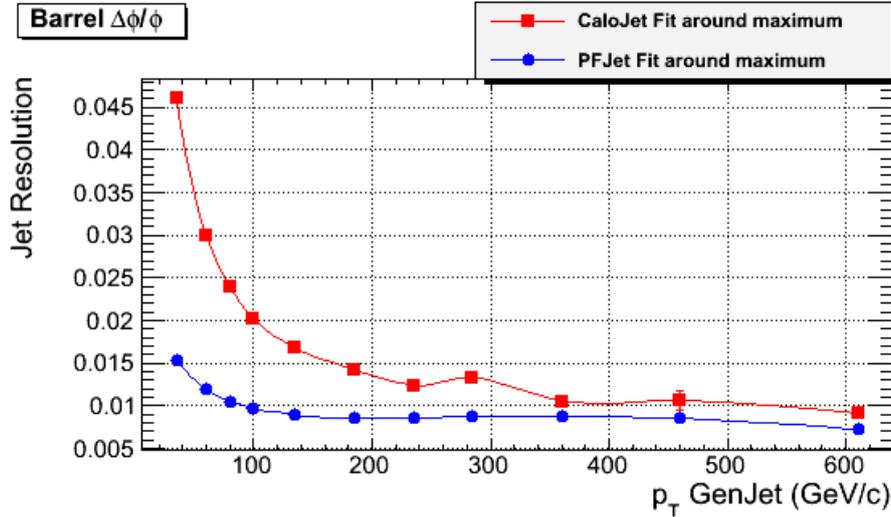


Figure 4.19: ϕ -resolution for jets in the Barrel using CMSSW_2_2_7.

Jets improve the resolution at low p_T . At high p_T , where the jets are very collimated, the resolution for PFJets and CaloJets becomes comparable. Unfortunately, this last remark cannot be united with what is shown in Figure 4.17. Keeping in mind the bugs presented earlier, this resolution was checked with the pre-released version of the CMS software resulting in Figure 4.18. Although this plot only shows the resolution up to a GenJet p_T of 200 GeV/c (due to the low statistics in the $t\bar{t}$ sample used in this version of the software), it shows the expected evolution with p_T . The same dependence on jet p_T is seen in Figure 4.19, which shows the ϕ resolution.

Both the η and ϕ resolutions clearly benefit from the particle identification and reconstruction of the ParticleFlow algorithm. At low jet p_T , the improvement for PFJets ϕ resolution is even bigger than the improvement in η because the CaloJets ϕ resolution is further deteriorated by the magnetic field. The η resolution is better for PFJets up to a factor of 2, the ϕ resolution up to a factor of 3.

4.3 Reconstructing m_{Top} with CaloJets and PFJets

In the last section, a comparison between PFJets and CaloJets is presented using single jets. In this section jets will be combined to construct a physics

object, more precise the mass of the top-quark. In this thesis only the fully hadronic channel is used for the $t\bar{t}$ -decay resulting in a 6-jets topology³. The jet matching explained in section 4.2.1 and shown in Figure 4.1 is used here. The only extra requirement is that at least 3 jets originating from the (anti-)top-quark are requested to be able to calculate the mass of the top-quark.

4.3.1 The m_{Top} reconstruction

Once the three jets from the hadronic (anti-)top-decay are identified via matching with the generated partons of the $t\bar{t}$ -decay, the mass of this quark can be calculated as:

$$m_t = \sqrt{(E_b + E_q + E_{\bar{q}})^2 - (p_x^b + p_x^q + p_x^{\bar{q}})^2 - (p_y^b + p_y^q + p_y^{\bar{q}})^2 - (p_z^b + p_z^q + p_z^{\bar{q}})^2} \quad (4.3)$$

The same formula can be used to calculate the mass of the anti-top-quark. This mass is equal to that of the top-quark thanks to the CPT-symmetry in the Standard Model.

In Figure 4.20, the mass of the top-quark reconstructed with GenJets and RecoJets is shown. Ideally, the mass of the top-quark would be distributed according to a Breit-Wigner distribution (see Figure 4.21) but due to detector effects a Gaussian smearing term is introduced. Again the GenJets are used as a benchmark for the RecoJets because they represent the best possible reconstruction.

Given the better momentum (and energy) resolution for PFJets, shown in the previous section, it was expected to see a better mass resolution for top-quarks reconstructed with PFJets. This is not very clear from the distributions shown in Figure 4.20. But what is obvious is that the mass-distribution from PFJets is much better aligned with the GenJet distribution than is the case for the CaloJets. For the same level of Jet Energy Corrections, the PF-Jets thus give a better reconstruction (on average) of the mass of this quark than CaloJets.

³ $t(\bar{t}) \rightarrow W^+(W^-) + b(\bar{b})$ and $W^\pm \rightarrow q + \bar{q}$ resulting in 6 jets

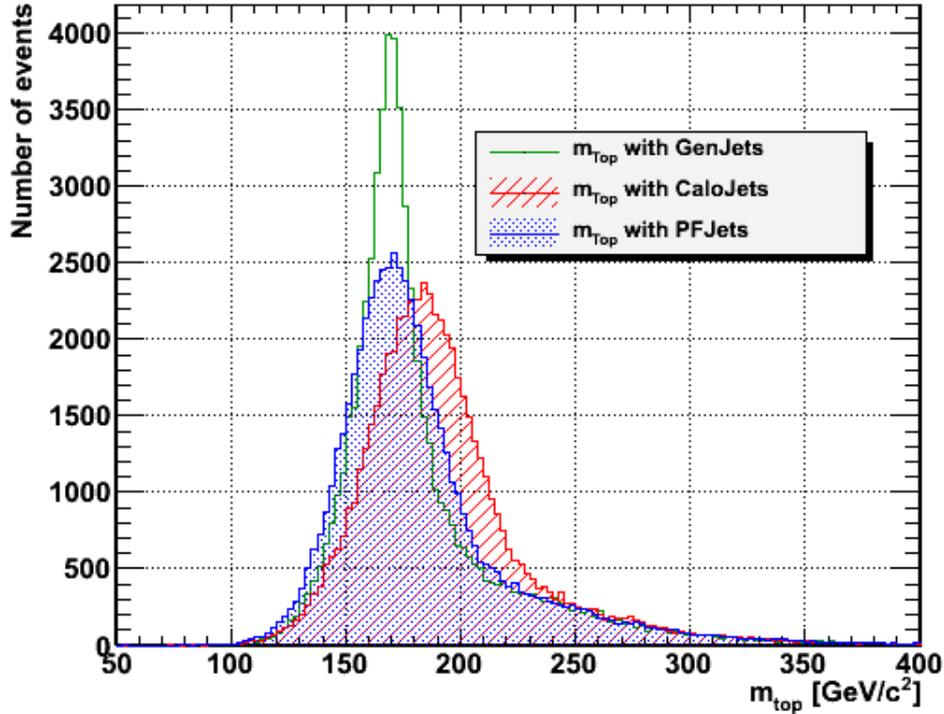


Figure 4.20: m_{Top} reconstructed with jets at generator and reconstruction level.

To conclude on this, a fit was made on the central Gaussian part of the distribution to exclude the large tail. This fit is shown for PFJets and CaloJets in Figure 4.22 and the fitted parameters are summarised in Table 4.2.

	Mass (GeV/c^2)	Width (GeV/c^2)
CaloJets	182.9 ± 0.2	22.29 ± 0.27
PFJets	169.9 ± 0.1	19.8 ± 0.2
GenJets	169.2 ± 0.1	10.21 ± 0.13

Table 4.2: m_{Top} fit results from a Gaussian Fit

From Table 4.2, it is clear that the PFJets mass bias is much smaller than for CaloJets (respectively 0.4% and 8% relative to the mass via GenJets). When using PFJets, the resolution (width of the peak) is about 10% smaller than with CaloJets. Thus, the PFJets provide a better mass resolution and

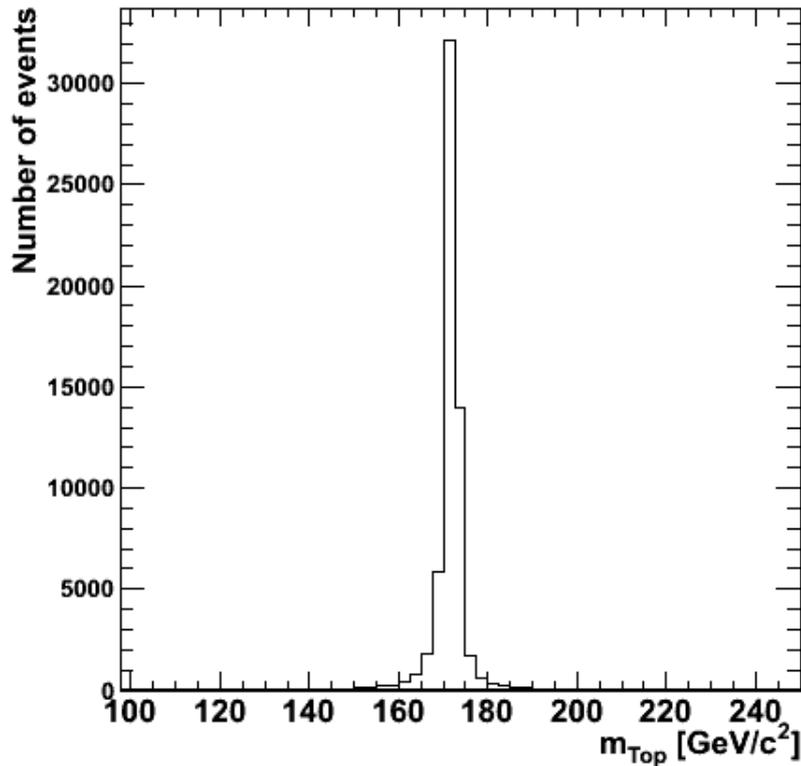


Figure 4.21: m_{Top} reconstructed using quarks.

response.

Finally it is interesting to look at the efficiency of matching the three jets originating from the (anti-)top-quark and consequently the efficiency in matching all six jets. Due to the improvement in angular resolution, the PFJets might improve this efficiency. To perform this check, the same jet selection scheme was used but without the requirement that a GenJet matches both a CaloJet and a PFJet. In this way it is possible to match the three jets of a top-quark for the PFJets while this is not the case for the CaloJets and vice versa.

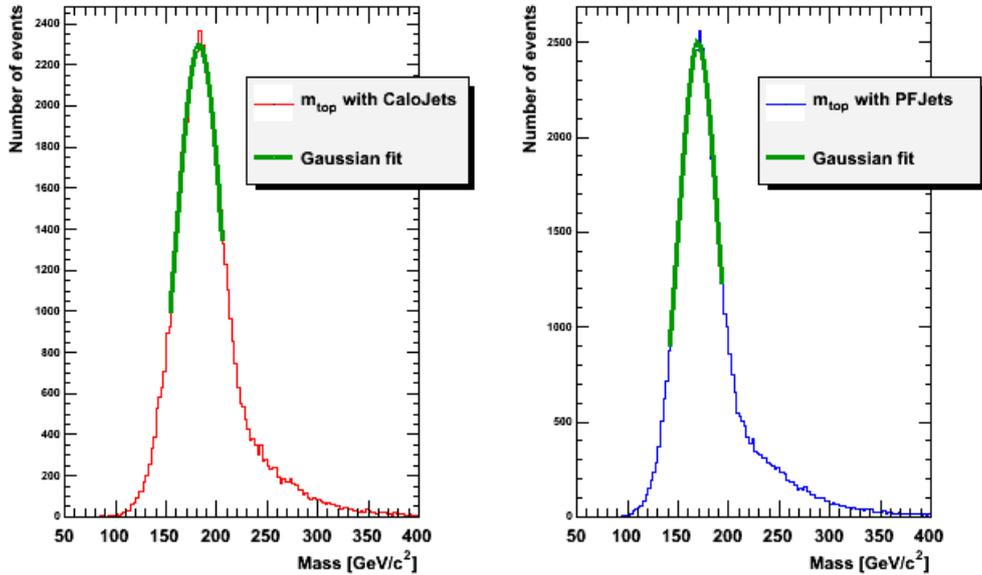


Figure 4.22: Fit of $m_{T\bar{t}}$ reconstructed with CaloJets (left) and PFJets (right).

Figure 4.23 shows the efficiency of matching both the six jets originating from the top-quark and the anti-top-quark. The efficiency is plotted in function of the cut on the GenJet p_T . As all six GenJets have to pass this cut, the efficiency is defined as the number of events where the 6 jets were matched and the corresponding GenJets pass the p_T cut, divided by the total number of processed events. This cut on the GenJet p_T instead of on the RecoJet p_T was chosen again to treat both CaloJets and PFJets on an equal footing. This is again explained by the fact that PFJets have a smaller energy correction. Looking at this plot, it is clear that for low p_T cuts, the efficiency is higher for PFJets.

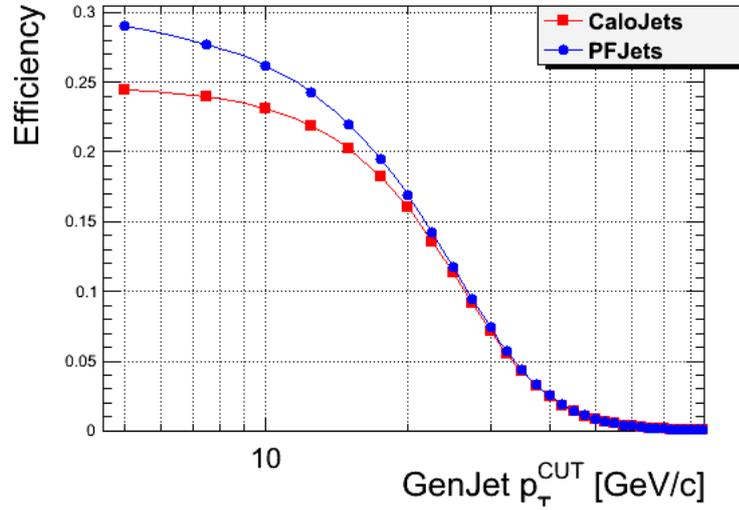
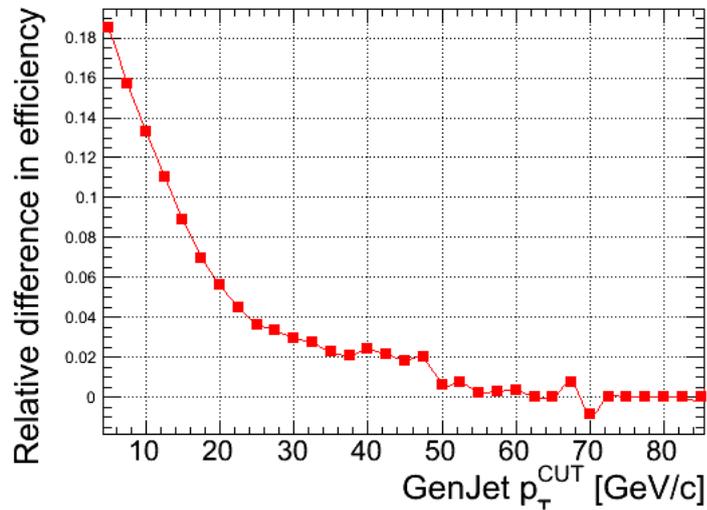
Figure 4.23: Matching efficiency for $t\bar{t}$.

Figure 4.24: Relative difference in matching efficiency between CaloJets and PFJets.

To have an even better view of the possible improvement with PFJets, the relative difference in efficiency between PFJets and CaloJets ($\frac{\epsilon_{PF} - \epsilon_{Calo}}{\epsilon_{Calo}}$) is

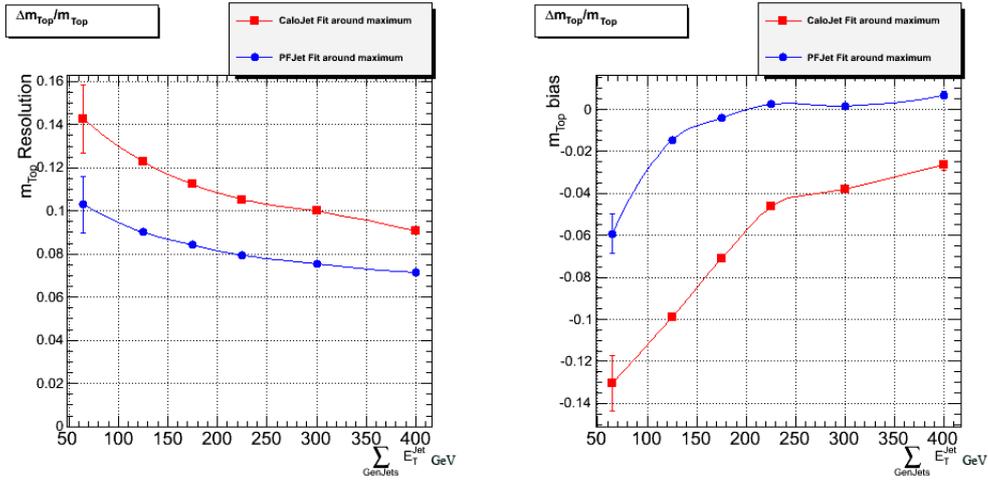


Figure 4.25: m_{Top} resolution and bias as function of $\sum_{\text{Jets}} E_T^{\text{Jet}}$.

shown in Figure 4.24. This plot clearly shows that only for p_T cuts lower than 20-25 GeV/c, the relative improvement is significant. However, in general no analysis will use jets with a p_T lower than 30 GeV/c because these jets are mainly fake jets. This would then lead to a conclusion that there is no real improvement in the matching efficiency when using PFJets. However, one needs to keep in mind the p_T resolution and response for PFJets shown in the previous section. There it was shown that PFJets have a significantly improved resolution and response for low p_T . This suggests that a PFJet with a p_T as low as 15-20 GeV/c can still represent a real jet. This is not the case for CaloJets. Thus it would not be impossible to use PFJets with such low p_T resulting in a higher matching efficiency for $t\bar{t}$.

4.3.2 The m_{Top} resolution and bias

As is done for the comparison between individual CaloJets and PFJets, the tri-jet mass resolution and bias can be evaluated in bins of a certain observable. The mass resolution is defined as the width of the Gaussian fit of $\frac{m_{\text{Top}}^{\text{Rec}} - m_{\text{Top}}^{\text{Gen}}}{m_{\text{Top}}^{\text{Gen}}}$ in each bin. The bias is then defined as the mean of this fit. The difference is that now, observables should combine the three jets from the (anti-)top-quark. The mass resolution and response will now be shown in bins of the scalar sum of the transverse energy of the three jets and the total number of constituents for the three jets (Figures 4.25 and 4.26).

In both cases, the PFJets have a better mass resolution over the whole range of the observable, but the improvement is quasi constant and the resolution

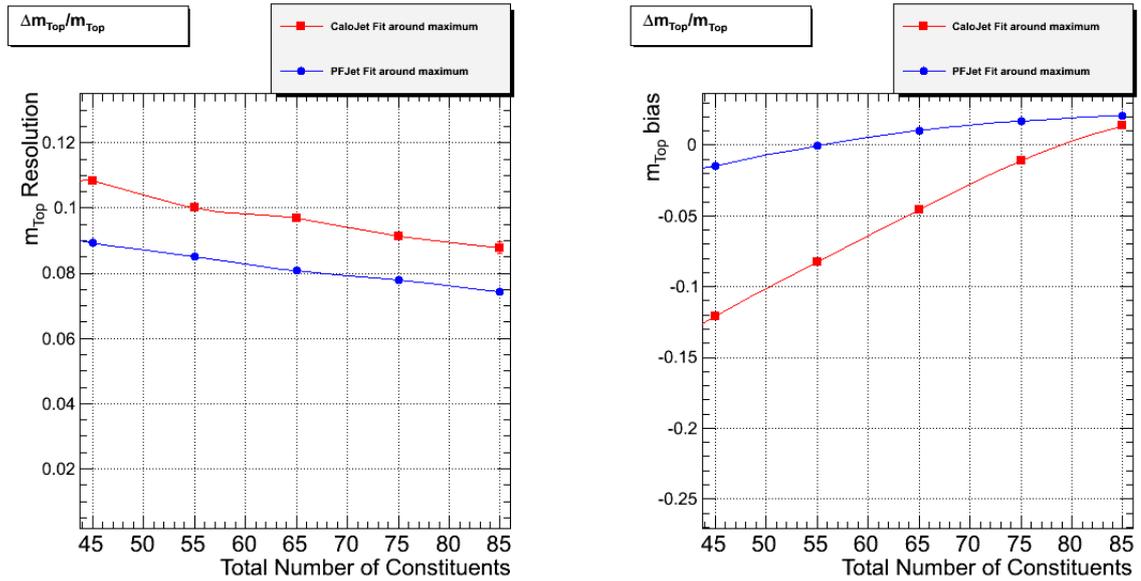


Figure 4.26: m_{Top} resolution and bias as function of the total number of constituents.

is almost independent from the observable. For the mass bias, a clear improvement is seen for the PFJets with respect to the CaloJets is seen when the total number of jet constituents drops. This is expected because the momentum response for CaloJets is worse than for PFJets when jets have a lower number of constituents. For tri-jet system with a low $\sum_{Jets} E_T^{Jet}$, the bias with CaloJets is again bigger than with PFJets because again the PFJets benefit from the particle reconstruction.

Chapter 5

Using ParticleFlow in $Z' \rightarrow t\bar{t}$ events

In this last chapter, it is checked the the search for physics beyond the Standard Model can be improved by using ParticleFlow techniques. More precisely, the PFJets will be used to look into the Z' -boson that decays into $t\bar{t}$. In the first section an overview of the theoretical background of the Z' boson will be given. This will be rather short because this goes beyond the scope of this thesis. Section 2 will address the reconstruction of the $t\bar{t}$ invariant mass using Z' events. Finally, in Section 3 will look into the possibility of selecting boosted-top events and how ParticleFlow might improve this selection.

5.1 The Z' boson

One of the big flaws in the Standard Model, is the inability to assign masses to the particles. As presented in Chapter 1, the mechanism of spontaneous symmetry breaking can solve this problem. There is in fact another way to introduce mass, without requiring an elementary scalar boson, called dynamical symmetry breaking. One of these dynamical symmetry breaking models is *Technicolor* [30]. In technicolor, additional massless fermions are added, called technifermions. These technifermions form fermion condensates which cause breaking of the global chiral symmetry. When choosing a $SU(N)_{TC}$ technicolor gauge group to describe the technifermions, left-handed technifermions are weak doublets and right-handed technifermions weak singlets. In this way the electroweak symmetry is broken to electromagnetism giving rise to Goldstone bosons, called techni-pions. Three of these techni-pions become longitudinal components of the W and Z bosons. These bosons get a mass that is proportional to the pion decay constant. Thus, technicolor can

explain the weak boson masses.

To explain the masses of the Standard Model particles, extra interactions should be introduced like *Extended Technicolor* gauge interactions. In Extended Technicolor (ETC), a larger gauge group embeds technicolor, color and flavour altogether. This gauge group then breaks down to technicolor and color at energies high in the TeV-scale. This breaking predicts the existence of a massive gauge boson which is responsible for the interactions between fermions and technifermions. This gives rise to the couplings which generate fermion masses.

Due to the high mass of the top-quark, it is expected that the third generation of fermions couples stronger to the symmetry-breaking sector than the lighter two. This might lead to the assumption that the top-quark plays a special role in the electroweak symmetry breaking. This leads to a model called *Topcolor* where the symmetry breaking is obtained by condensation of top-quarks at a certain energy scale. The model proposes a gauge group: $SU(3)_h \times SU(3)_l \times SU(2) \times U(1)$ where $SU(3)_h$ couples to the third generation, with coupling constant g_h , and $SU(3)_l$ couples to the two others, with coupling constant g_l ($g_l \ll g_h$). In this model, new strong $U(1)$ interactions are introduced to ensure top-quark condensation and prevent bottom-quark condensation. Unfortunately this model cannot reconcile with the high mass of the top-quark. As a result the Topcolor is combined with Technicolor in *Topcolor-assisted Technicolor* (TC2). Here the weak hyper-charge is also separated between the third and the first two fermion generations leading to the following gauge group: $SU(3)_h \times SU(3)_l \times SU(2) \times U(1)_h \times U(1)_l$. The third generation now transforms under the Topcolor group $SU(3)_h \times U(1)_h$ and the others under $SU(3)_l \times U(1)_l$. At an energy scale of about 1TeV this symmetry breaks dynamically into $SU(3)_{QCD} \times U(1)$. At a scale above the electroweak scale¹, the groups $U(1)_h$ and $U(1)_l$ break into $U(1)$ giving rise to a gauge boson. This gauge boson, called Z' , is a linear combination of the two original hyper-charge bosons which couples to left- and right-handed fermions.

The Z' boson almost exclusively decays into a $t\bar{t}$ pair resulting in the same event topology as was used in the previous chapter. Again, the top-quarks decay almost 100% into a W boson and a b quark and the decay in the fully hadronic channel is chosen.

¹About 246GeV

5.2 Reconstruction of $m_{t\bar{t}}$ in $Z' \rightarrow t\bar{t}$ events

In the previous chapter, the mass of the Standard Model top-quark was reconstructed with both CaloJets and PFJets using a Monte-Carlo based jet matching. In the same way the mass of the Z' boson is reconstructed as the invariant mass of the $t\bar{t}$ -system. As the Z' boson decays into a $t\bar{t}$ -pair, it results in the same event topology as Standard Model $t\bar{t}$ events, the matching presented in Section 4.2.1 and Figure 4.1 is used. The only difference is that it is no longer required to have both a CaloJet and a PFJet matched to the GenJet. This relaxation of the matching is applied to benefit from the higher $t\bar{t}$ matching rate with PFJets (see Figures 4.23 and 4.24). The $t\bar{t}$ invariant mass is calculated in the same way as the mass of the single top-quark by extending eq. (4.3) to six jets. The distribution of the invariant mass, constructed using GenJets and GenParticles, is shown in Figure 5.1. These fully hadronic Z' events were generated using Madgraph followed by

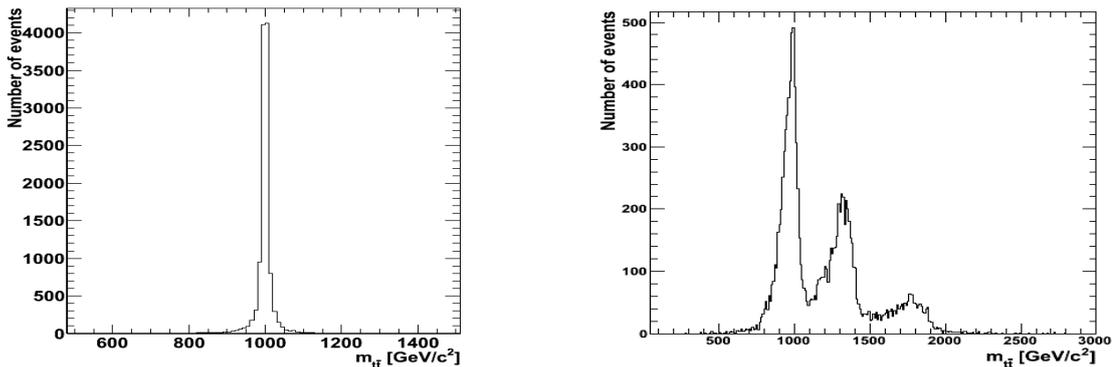


Figure 5.1: Invariant mass of the $t\bar{t}$ -system reconstructed with GenParticles (left) and GenJets (right)

a full detector simulation. In the generation of the events, the Z' mass was fixed to $1\text{TeV}/c^2$. This reflects in an invariant mass for the $t\bar{t}$ pair narrowly distributed around this value when using the GenParticles. Looking at the same distribution for GenJets one expects to see the same peak but with a Gaussian smearing. In fact, three peaks show up. These two peaks clearly don't belong to the tail of the peak at $1\text{TeV}/c^2$. So what is causing these peaks? It is shown in Figure 5.2 that with increasing mass of the Z' boson, the angle between the direction of the generated top-quark and the direction of the jets gets smaller. This is caused by a boost of the top-quark. For $m_{Z'}$ equal to $1\text{TeV}/c^2$, the angle ranges between 0.4 and 1.4 rad (RMS spread), or equivalently 23 and 80 degrees. Due to the boost, the top-quark decays into

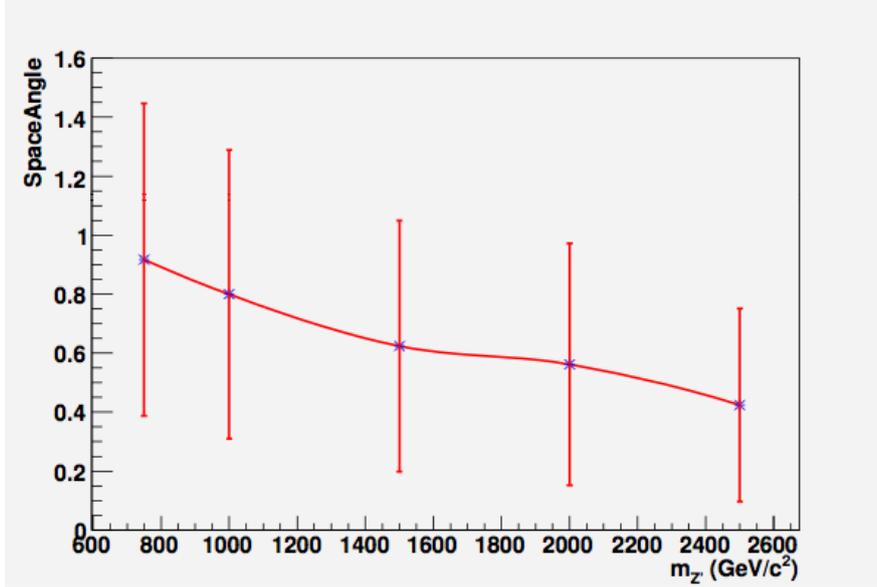


Figure 5.2: Largest space angle between the direction of the generated top quark and the direction of the jets in function of the Z' mass

three final-state quarks very close to each other. When the distance between two quarks becomes very small, it can happen that these two quarks fragment into one jet. The second peak in the right plot in Figure 5.1 is thus caused by merging of two jets into one. The third peak corresponds to a similar effect where all three quarks end up in the same jet. This means that the $t\bar{t}$ fully hadronic channel can have a 6-, 5-, 4-, 3-, 2-jet topology in the Z' decay.

Using the jet matching scheme outlined in Section 4.2.1, the same GenJet (and RecoJet) could be matched to multiple quarks. When the mass of the $t\bar{t}$ system is then reconstructed, one gets these peaks because of double-, triple-counting of the same jet. To solve this, the following procedure is applied. Starting from the six matched Quark-GenJet-RecoJet triplets, the distance in (η, ϕ) -space between two quarks is checked. If it is smaller than 0.8 and the distance between their GenJets is smaller than 0.1, it is assumed that the quarks merged into the same jet. One of the two jets is then removed from the list. This procedure is repeated until all jet-duplicates are removed. Then the invariant mass of the $t\bar{t}$ -system was recalculated producing the distributions in Figure 5.3. The only peak left is the one at $1\text{TeV}/c^2$ which is expected given the simulated $m_{Z'}$ mass of $1\text{TeV}/c^2$.

Comparing the distribution for CaloJets and PFJets, it is clear that the improvement obtained with PFJets compared to CaloJets is much smaller than was the case with the reconstruction of the Standard Model m_{Top} . To

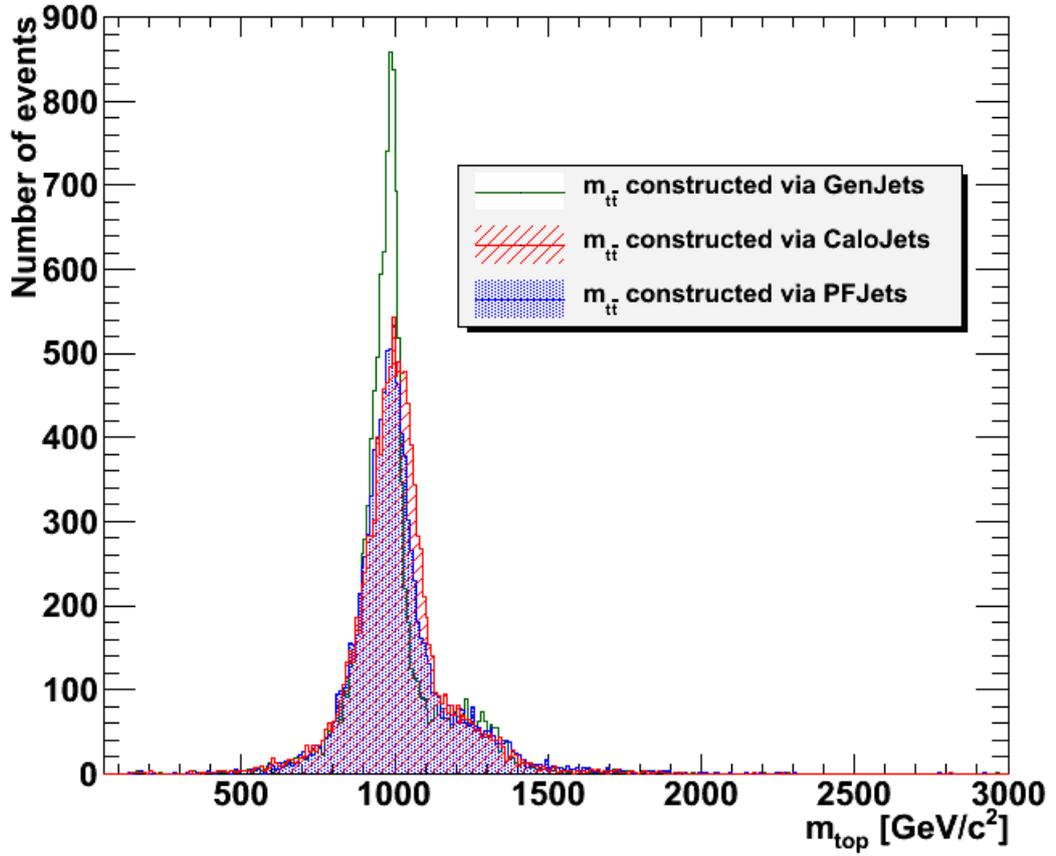
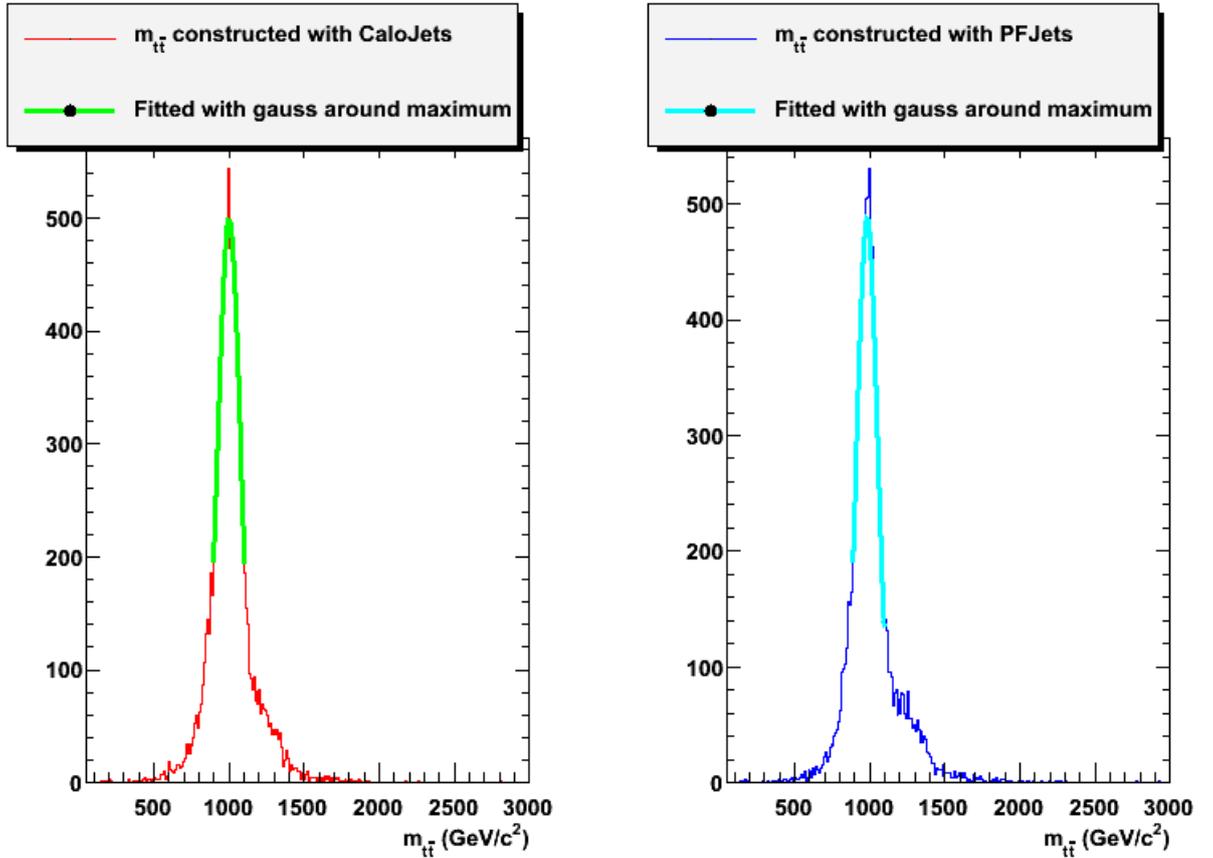


Figure 5.3: Invariant mass of the $t\bar{t}$ system, reconstructed with GenJets and RecoJets. The double- and tripple-counted jets, due to the fragmentation of multiple quarks into one jet, are removed using the method presented in the test.

Figure 5.4: Gaussian fit of the $m_{t\bar{t}}$ distribution

check this more carefully, the peak is fitted with a Gaussian. The fits of the invariant mass of the $t\bar{t}$ -system using CaloJets and PFJets are shown in Figure 5.4 and the fitted parameters are given in Table 5.1. The results

	Mass (GeV/c^2)	Width (GeV/c^2)
CaloJets	999.8 ± 1.2	75.53 ± 1.56
PFJets	982 ± 1.1	69.84 ± 1.35
GenJets	976.7 ± 0.7	41.02 ± 0.82

Table 5.1: The $m_{t\bar{t}}$ fit results

in Table 5.1 show an improvement of about 8% in the mass resolution with PFJets. The $t\bar{t}$ mass reconstructed with PFJets has a bias of only 0.5% with respect to the mass reconstructed with GenJets. This value is again much

smaller than the 2.3% bias which is obtained reconstructing the mass with CaloJets. To conclude, the PFJets again lead to a better mass resolution and lower bias when reconstructing the $t\bar{t}$ invariant mass.

5.3 Identifying boosted top quarks

It was explained in the previous section of this chapter that top-quarks originating from the decay of the Z' boson can be boosted. This can result in a fragmentation of two or three quarks into one jet. Such a top-quark is called a *boosted top*. The boosted top quark is a very useful probe to detect the Z' boson if it exists. The only problem is that the 2-jet topology is overtaken by a dominant background of QCD di-jet events and from Standard Model $t\bar{t}$ events. These are not the only backgrounds possible, though the most abundant ones. The Z' production cross-section is orders of magnitude smaller than the cross sections for QCD di-jet and $t\bar{t}$, which enforces a careful event selection. It was proposed to construct a cone with opening angle $R=1.0$ in (η, ϕ) -space around the two leading E_T CaloJets². These cones are called *SuperJets*. The mass inside a SuperJet is then calculated using the CaloTowers inside. Finally a cut on the SuperJet mass, which is an estimate for the top-quark mass, could distinguish signal and background.

Given the better overall performance of the PFJets compared to CaloJets, why not use ParticleFlow? It might improve the Z' event-selection. So instead of using the leading E_T CaloJets, one can use the leading E_T PFJets and replace the CaloTowers by PFCandidates³.

The samples which are used as background are summarised in Table 5.2. The sample used as signal⁴ is given in Table 5.3. All these samples were simulated using a collision energy of 10TeV.

To meet the single-jet trigger threshold⁵ at 10TeV proton collision energy⁶, a cut of 110GeV/c is applied on the p_T of both leading jets. This cut is applied to CaloJets and PFJets separately on each sample. For each accepted signal or background event, two SuperJets are constructed around the leading jets. The one with the largest mass out of two is put into the plot Figure 5.5 and the lightest one into the plot in Figure 5.6. To make these plots, all the samples are rescaled to an integrated luminosity of 1000 pb^{-1} . This value is

²Leading jets after L2L3 Corrections.

³These are the reconstructed particles produced by the ParticleFlow algorithm.

⁴The calculation of the cross-section of this process is presented in [31], but the actual value at 10TeV collision energy was obtained in a private conversation with the author.

⁵CMS Trigger Menu: https://twiki.cern.ch/twiki/bin/view/CMS/TSG_03_V_09

⁶This is expected to be the startup condition of LHC

Sample	Events	σ (pb)	L (pb^{-1})
QCD di-jet $80GeV/c < \widehat{p}_T < 120GeV/c$	48600	1617240	0.031
QCD di-jet $120GeV/c < \widehat{p}_T < 170GeV/c$	54135	255987	0.21
QCD di-jet $170GeV/c < \widehat{p}_T < 230GeV/c$	48600	48325	1.01
QCD di-jet $230GeV/c < \widehat{p}_T < 300GeV/c$	59400	10623.3	5.59
QCD di-jet $300GeV/c < \widehat{p}_T < 380GeV/c$	68688	2634.94	26.07
QCD di-jet $380GeV/c < \widehat{p}_T < 470GeV/c$	46656	733.099	75.29
Standard Model $t\bar{t}$ (fully hadronic)	82868	184	2.92

Table 5.2: Background samples generated with Pythia at a collision energy of 10TeV.

Sample	Events	σ (pb)	L (pb^{-1})
$Z' \rightarrow t\bar{t}$ (fully hadronic)	70000	3.28	21341.46

Table 5.3: Signal sample generated with Madgraph at a collision energy of 10TeV.

chosen because at lower values, not much Z' events are left.

The number of signal and background events at this integrated luminosity is given in Table 5.4.

	Sample	Events (Calo)	Events (ParticleFlow)
Signal	$Z' \rightarrow t\bar{t}$	2906	2891
Background	pp $\rightarrow t\bar{t}$ (Standard Model)	48136	45800
	QCD di-jet (Combined)	$\sim 2 \times 10^{14}$	$\sim 2 \times 10^{14}$

Table 5.4: Number of signal and background events at $1000pb^{-1}$. The $t\bar{t}$ decay is taken leptophobic.

It is clear from both figures that the Z' signal events are buried in the background. An event selection only based on the p_T cut on the leading jets and on the SuperJet mass is clearly not enough. Not even when using ParticleFlow to construct the SuperJet and its mass. This is not striking when looking at the number of events for signal and background with $1000 pb^{-1}$ data (Table 5.4). Although a p_T cut on the leading jets can't dig the signal out of its background, it can suppress the Standard Model $t\bar{t}$ background significantly. This can be seen in the p_T distribution of the leading jets shown

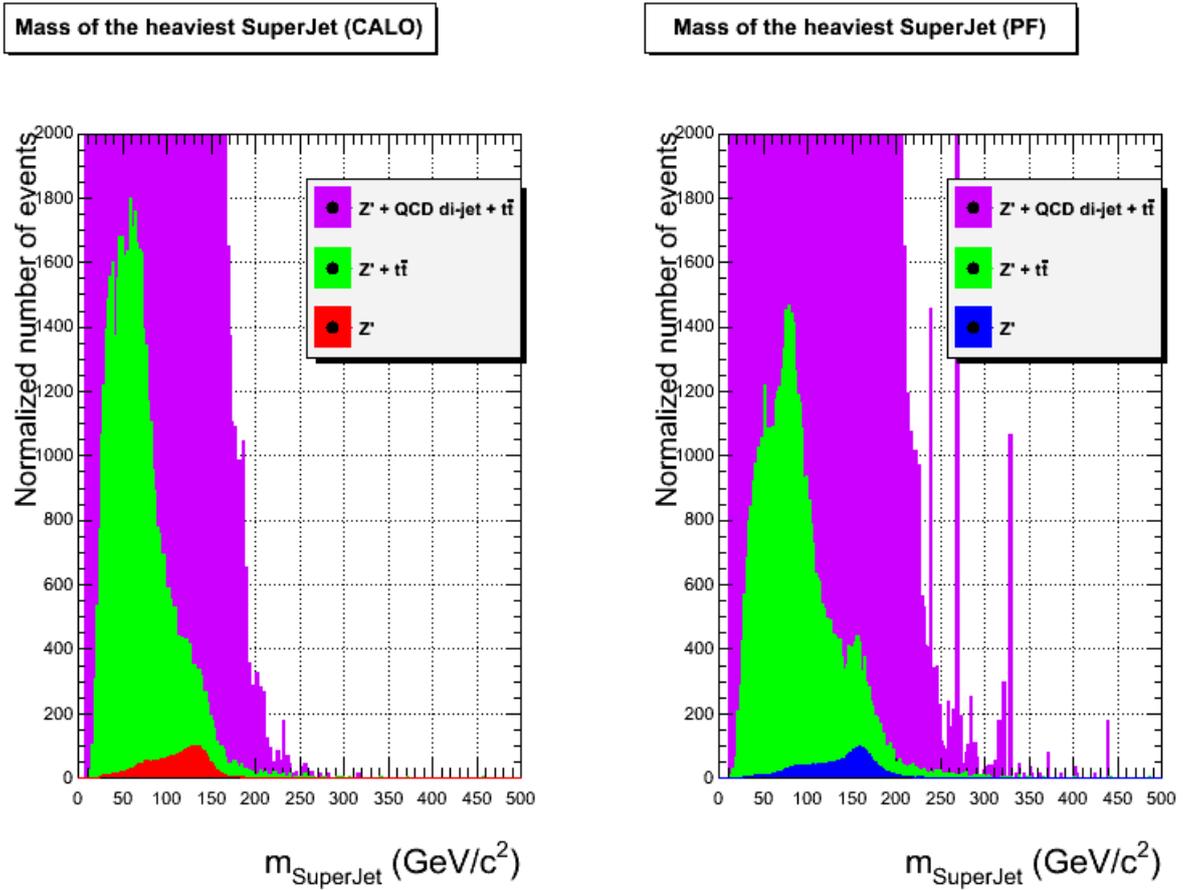


Figure 5.5: Mass of the heaviest SuperJet in the event. The SuperJet is constructed around one of the two (corrected) leading CaloJets (left) or PFJets (right) with $p_T > 110\text{GeV}/c$. All the contributions to this plot were rescaled to an integrated luminosity of 1000pb^{-1} .

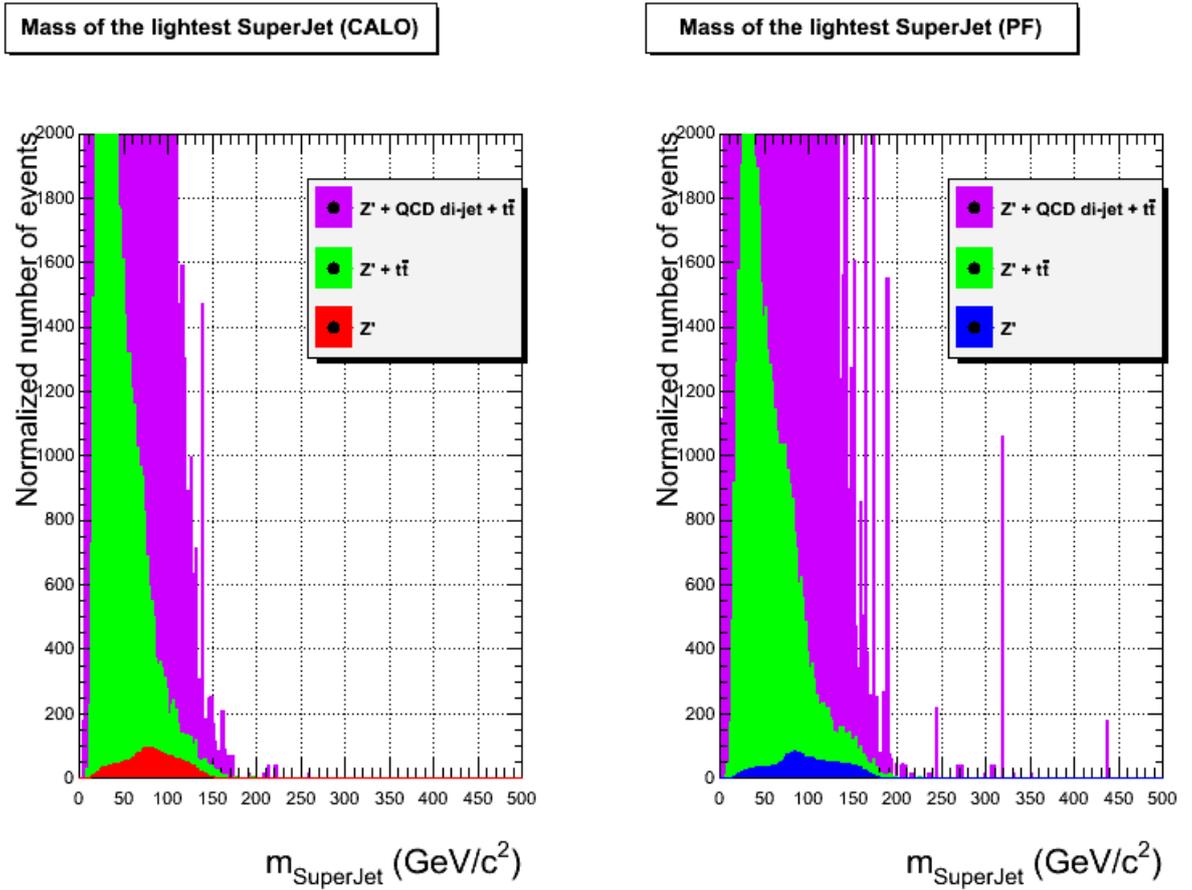


Figure 5.6: Mass of the lightest SuperJet in the event. The SuperJet is constructed around one of the two (corrected) leading CaloJets (left) or PFJets (right) with $p_T > 110\text{GeV}/c$. All the contributions to this plot were rescaled to an integrated luminosity of 1000pb^{-1} .

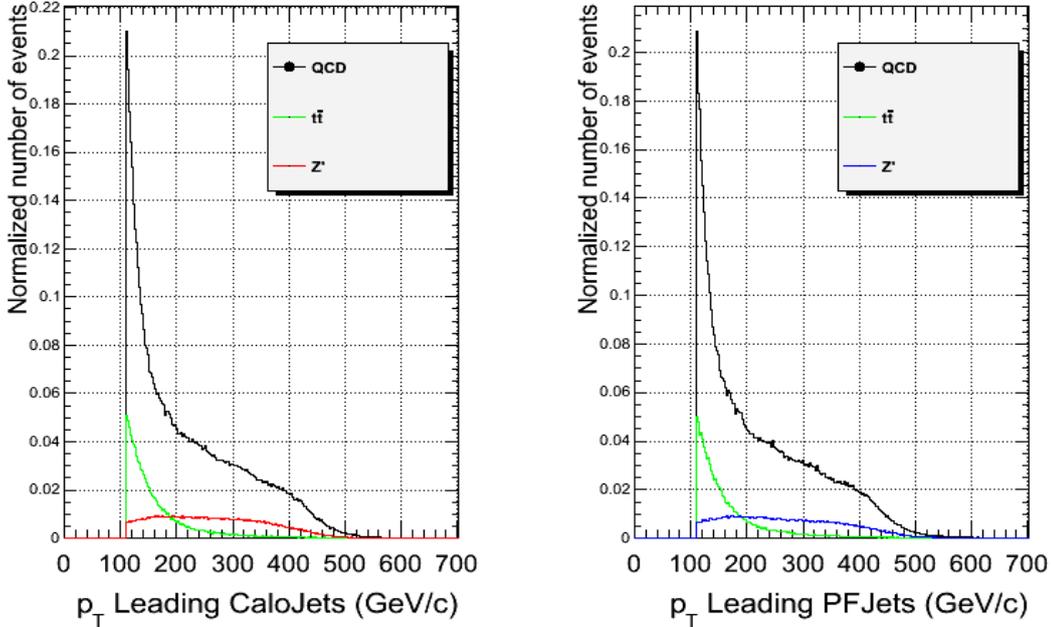


Figure 5.7: p_T distribution for the two leading jets in signal (Z') and background ($t\bar{t}$ and QCD di-jet).

in Figure 5.7. When raising the p_T cut from 110 GeV/c to values above 200 GeV/c , the number of background $t\bar{t}$ events will drop, but the QCD di-jet background still dominates. The separation of the Z' signal events and the dominant QCD di-jet background deserves a study on its own.

Although it seems that ParticleFlow cannot do better than the Calorimeter approach, this is not entirely correct. Looking at the plot of the largest SuperJet mass (Figure 5.5) it can be seen that ParticleFlow produces a more narrow peak which is centered at a mass $> 150 \text{GeV}/c^2$. For the SuperJet mass calculated with CaloTowers, the peak is not only broader but it is also centered at lower mass.

This last remark can be studied into more detail by looking at Figure 5.8. In the two top plots, the mass of the lightest and heaviest SuperJet are shown for CaloTowers and PFCandidates in signal events. All background was omitted to be able to compare the signal shape. It is clear that the heaviest SuperJet mass represents the mass of the boosted top-quark, which is better reconstructed using PFCandidates as mentioned before. The plot for the lightest SuperJet mass shows a peak at roughly 75 to 80 GeV/c^2 which

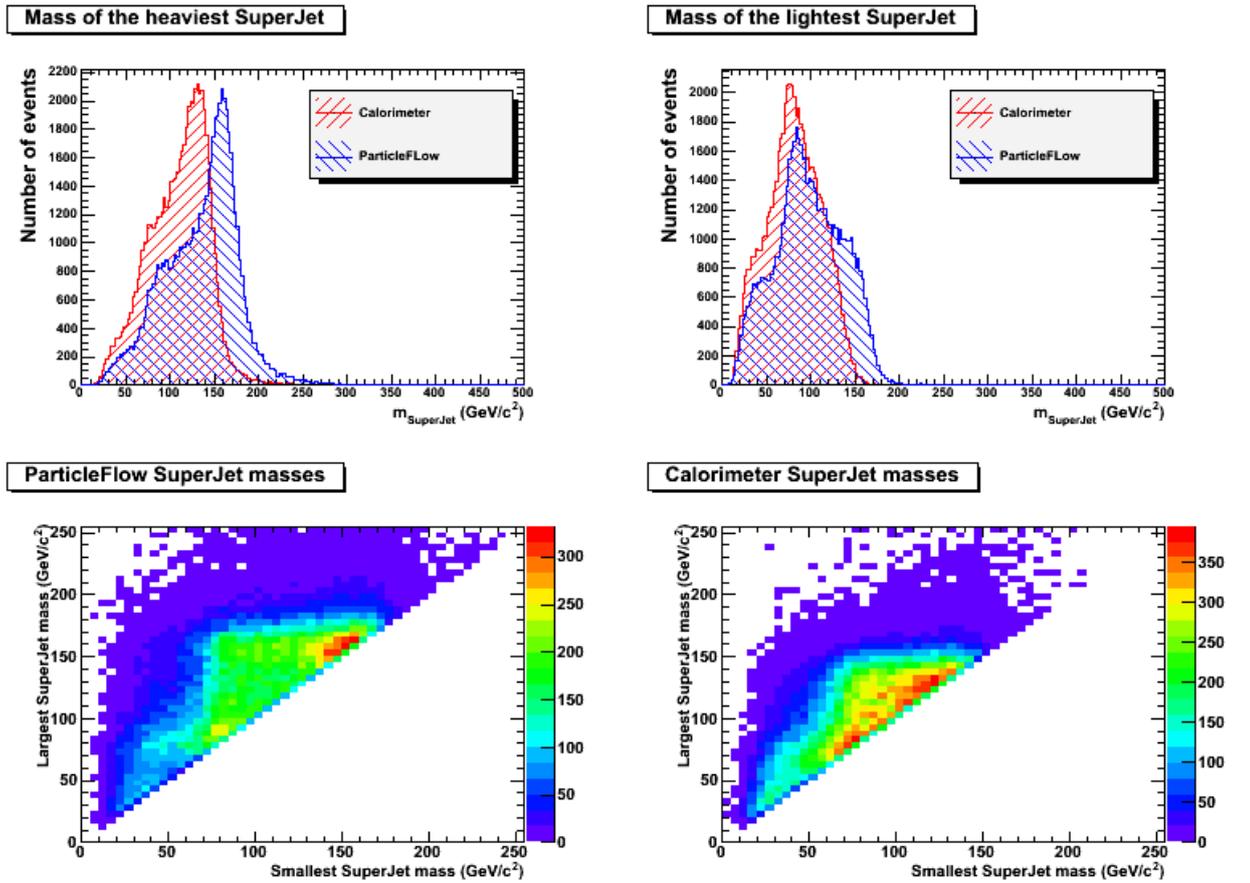


Figure 5.8: Mass of the SuperJet in the signal events with $p_T^{\text{Leading}} > 110 \text{ GeV}/c$. No rescaling to 1000 pb^{-1} is done as only the signal events are considered.

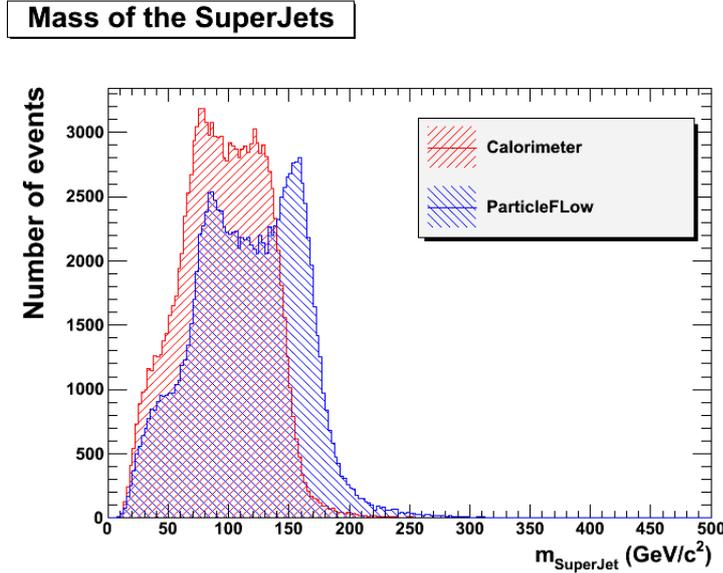


Figure 5.9: Mass of both SuperJets in the signal events with $p_T^{Leading} > 110 \text{ GeV}/c$. No rescaling to 1000 pb^{-1} is done as only the signal events are considered.

corresponds to the mass of the W-boson. The events in this peak are the ones where only the two quarks from the W-boson lie close to each other, which leads to one jet from the W-boson and one separate b-jet. The difference between the CaloTower and PFCandidate approaches is even more clear in the bottom plots of Figure 5.8 where the mass of the heaviest SuperJet is plotted as a function of the mass of the lightest. For ParticleFlow, two distinct peaks arise. The highest peak (red) represents the signal events where both SuperJets represent the boosted top-quark. The second (lower) peak arises when both SuperJets represents the W-boson. The green intermediate region represents combinations of the previous cases. The upper left part of this region represents the case where one SuperJet in the event represents the boosted top-quark and the other only the W-boson. For the CaloTowers on the other hand, no such distinct peaks are observed. In this case there distribution consists of one nearly flat peak.

This is also shown in Figure 5.9 where the the mass-distribution of the both SuperJets is plotted. Taking all these plots into account it is safe to conclude that when new variables will be introduced to suppress the QCD di-jet background, ParticleFlow will certainly improve the selection of boosted top-quark events.

Finally, it might be derived from Table 5.4 that using ParticleFlow leads to

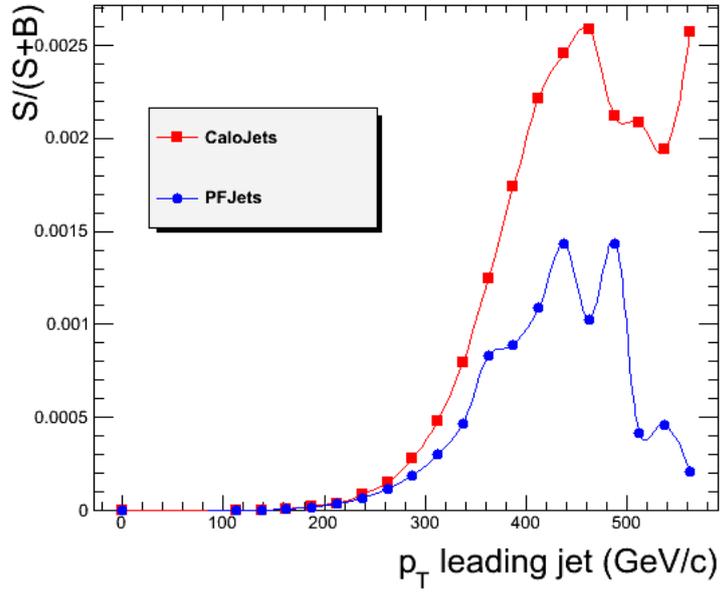


Figure 5.10: Signal over signal+background in function of the p_T of the leading jet at 1000 pb^{-1} with $p_T^{\text{Leading}} > 110 \text{ GeV}/c$.

less signal events. While the $t\bar{t}$ background is also reduced in the case of ParticleFlow, the QCD di-jet background remains quasi the same. This results in a signal over signal+background ratio shown in Figure 5.10. So using only the calorimeters, more signal and $t\bar{t}$ background events at high leading jet p_T are found. This is also seen from Figure 5.11 where the signal over signal+background ratio is plotted with only $t\bar{t}$ events as background. This plot shows that this ratio is equivalent for both. This might lead to a conclusion that maybe the CaloJets/CaloTowers approach will perform slightly better. This is probably not the case as Figure 5.9 clearly shows that a narrower peak is obtained with ParticleFlow. The ParticleFlow peak is also centered at higher mass which possibly allows a larger cut on the SuperJet mass and a higher selection purity.

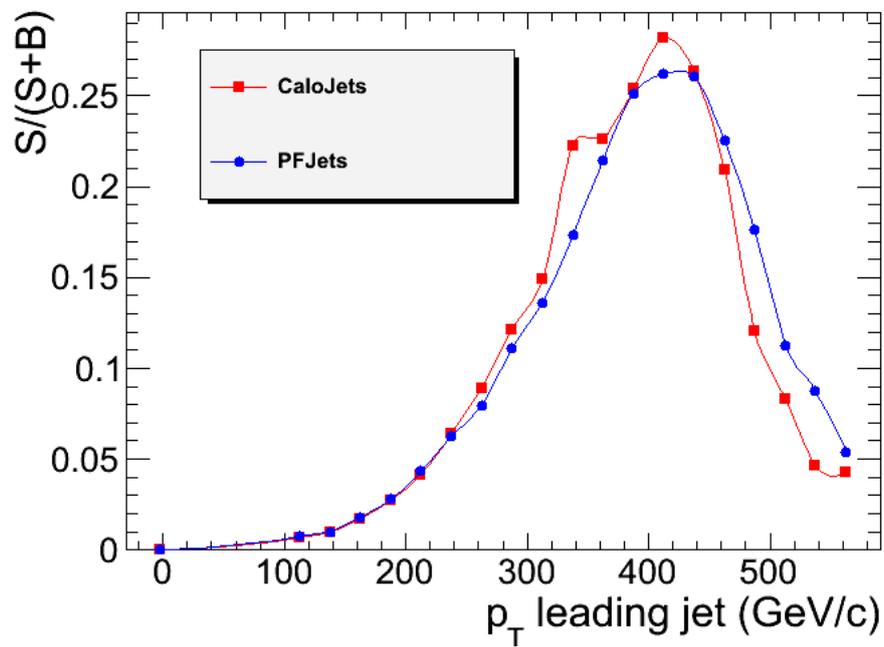


Figure 5.11: Signal over signal+background in function of the p_T of the leading jet at 1000 pb^{-1} with $p_T^{\text{Leading}} > 110 \text{ GeV}/c$. In this plot only $t\bar{t}$ events are counted as background.

Chapter 6

Conclusions

In the third chapter of this thesis, the ParticleFlow algorithm was introduced. This algorithm uses all the CMS subdetectors to identify the final-state particles in each event. From these particles, a jet can be clustered. The cone algorithm was also introduced which produces these ParticleFlow jets as well as the traditional Calorimeter jets. The aim of this thesis was to study the difference in performance between these two types of jets.

6.1 Performance of ParticleFlow with Jets

In Chapter 4, a comparison was made between the jets clustered using ParticleFlow and the Calorimeter Jets (PFJets and CaloJets). This was done by comparing the transverse momentum resolution and response within top-quark events for both types of jets. Secondly the jets were used to reconstruct the mass of the top-quark. Although the conclusions presented in this section are based on the performance in $t\bar{t}$ events, they are in general valid for all events with similar jets.

6.1.1 Jet resolution and response

To check the performance of PFJets compared to CaloJets, the distribution of $\frac{p_T^{Rec} - p_T^{Gen}}{p_T^{Gen}}$ was studied where Gen denotes the generator-level jet (GenJet). For uncorrected jets, we saw a shift away from zero of 10% for PFJets compared to 35% for CaloJets which indicates that the response for PFJets is better. The distribution of $\frac{p_T^{Rec} - p_T^{Gen}}{p_T^{Gen}}$ for L2L3 corrected jets, showed a shift of the distribution, which is 5 times larger for CaloJets (around 7%) than for PFJets (around 1.5%). An additional improvement of 25% in the RMS using PFJets compared to using CaloJets was observed. To study the resolution

and response in more detail, the distribution was binned in terms of p_T^{Gen} . In each of these bins the resolution and response was determined. This revealed a relative improvement of the PFJets resolution compared to the CaloJets resolution of 20% to 40% in the Barrel and 10% to 25% in the Endcaps over the whole plotted p_T -range. Furthermore, the PFJets shown an absolute improvement of the response which is almost 12% (10%) at low p_T in the Barrel (Endcaps). The same method was applied on bins in η and ϕ . This led to the discovery of two bugs in the ParticleFlow algorithm within the current stable CMS Software Framework (CMSSW_2.2.X). First of all the p_T -response in function of η is not symmetric and there is no physical reason why this should occur. The p_T resolution and response in function of ϕ also showed an unexpected feature, namely two "bumps" at $\phi = \pm\pi/2$ while this should be flat. In both cases, the response was determined again using a new version of the CMS Software Framework (CMSSW_3.1.0_preX). Fortunately, the resolution and response behave as expected in this release which confirms that the unexpected behaviour seen before is due to bugs in the code.

Although it is asymmetric, the response in function of η for corrected jets shows an energy-bias which is a factor of 3 smaller for PFJets than for CaloJets. What is even more promising is that, using the new CMS Software, this energy-bias (for uncorrected jets) in the Barrel is almost a factor of 10 smaller for PFJets than for CaloJets compared to a factor of 4 when using the current CMS Software version.

Furthermore the angular resolution was checked for both jet-types. It was shown that the η -resolution improved by a factor of 2 for PFJets compared to CaloJets. For the ϕ -resolution a factor of 3 is obtained because of the influence of the magnetic field on the CaloJets.

6.1.2 Reconstruction of m_{Top}

Now that it is shown that the PFJets have a better momentum resolution and response, both types are used to reconstruct the mass of the top-quark. The full hadronic channel in the Standard Model $t\bar{t}$ -decay was chosen because there is no Missing Transverse Energy and the benefit from ParticleFlow is enlarged. While for the reconstruction with PFJets, the mass bias is less than 1% it is about 8% when using CaloJets. This shows that the ParticleFlow Jets better estimate the mass of the top-quark. Furthermore, it is shown that the mass resolution when using PFJets is about 10% better than when using CaloJets. The mass resolution and response was also checked as a function of the number of jet constituents of the tri-jet system and the scalar sum of their E_T . This revealed again a better mass resolution and response for the

PFJets, but the improvement compared to CaloJets is independent of the observables.

Finally, the efficiency to match a $t\bar{t}$ -pair was calculated. It turned out that for a p_T cut on the GenJets higher than 20GeV/c the matching efficiency with PFJets is 6-18% bigger than with CaloJets.

As jets are very important in the physics programme of CMS, it is very promising to see the improvement of jet performance obtained by the ParticleFlow technique.

6.2 Using ParticleFlow in analysis

Because of the importance of jets in CMS, the ParticleFlow method will most certainly be deployed in future analyses. To check the potential of ParticleFlow in this area, the PFJets were deployed in a study of a theory beyond The Standard Model: Topcolor-assisted Technicolor. This model predicts a massive gauge boson, the Z' boson, which decays almost exclusively into $t\bar{t}$.

6.2.1 Reconstruction of the $t\bar{t}$ invariant mass in Z' decays

The invariant mass of the $t\bar{t}$ system was reconstructed in practically the same way as the mass of the top-quark. Again the fully hadronic decay-mode was chosen for convenience. It was expected, from the improvement on m_{Top} , to see an improvement in the $t\bar{t}$ mass resolution. The $t\bar{t}$ mass resolution is improved by 8% for PFJets compared to CaloJets. What is peculiar in this case is that two distinct peaks appeared in the $t\bar{t}$ invariant mass distribution. These peaks seem to correspond to events where two or three quarks from the top-decay lead to only one jet. Such a top-quark is called a boosted-top. Thus, in these events these events, two or three quarks matched the same GenJet. These double- and tripple-counted jets were removed to reconstruct the mass. Again the mass bias when using PFJets is smaller than when using CaloJets (0.5% and 2.4% respectively).

6.2.2 Identification of boosted top-quarks in Z' decays

As these boosted top-quarks are a good probe to detect the Z' boson, it might be useful to check how ParticleFlow can assist their selection. As the three quarks from the top quark decay can fragment in only one jet, the $t\bar{t}$ decay can match a two-jet topology. A dominant background of QCD di-jet events

then needs to be rejected. A possibility to do this is to use SuperJets around the two leading jets. If a top-quark in the Z' event is boosted, the mass inside the SuperJet will be of the order of the top-mass. It was shown that when using PFJets and PFCandidates rather than CaloJets and CaloTowers, the SuperJet mass distribution contains a sharper peak which is centered more closely to the mass of the top-quark. This shows that once the dominant QCD di-jet background can be reduced, it is likely that the boosted top-quark selection will be improved by using ParticleFlow.

6.3 Stability of the algorithm

As ParticleFlow uses all CMS subdetectors to reconstruct the final-state particles in the event, one expects a change in performance when a certain sub-detector fails duty. This question is not yet addressed in this thesis. According to the ParticleFlow group [25], when for example a layer of the tracker fails, the calorimeters will take over at a slightly lower p_T with no large effect on the resolution and jet energy scale. In an other case, where a module of the HCAL is turned off, there would be no CaloJets in this direction. In this case, the jets will still be reconstructed by ParticleFlow although with a reduction on the energy scale of 10-20%.

6.4 Outlook

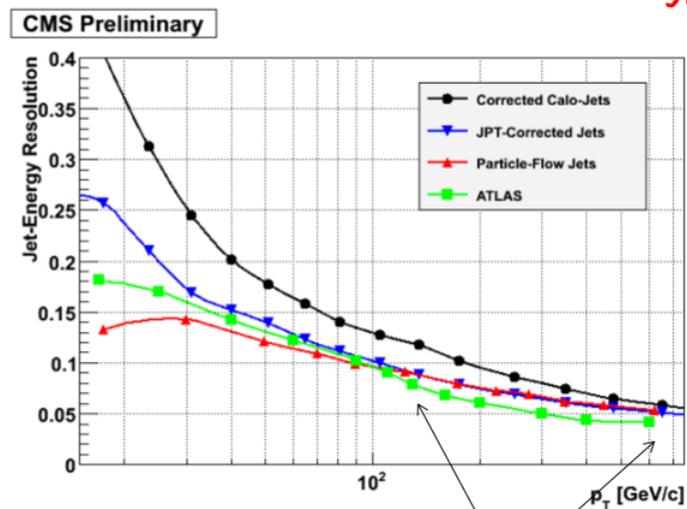
Taking into consideration all the previous conclusions, ParticleFlow seems to be a very promising method to provide the input for the jet-reconstruction¹. This method will most certainly be used for all future analyses in CMS as it can't be stressed enough that jets are of key importance in the quest for physics beyond the Standard Model.

To end this thesis, a plot is shown where the latest² ParticleFlow jets (uncorrected) are compared to Calorimeter jets (corrected) and to the jets in the ATLAS experiment. This plot was shown in the CMS Week Plenary Meeting (March '09). This plot shows that the ParticleFlow jets not only beat the Calorimeter jets in performance but for jet p_T below 100GeV/c they are even better than the jets in the ATLAS experiment. ATLAS might be the biggest detector, but it's only a matter of time before CMS has the "biggest" jets!

¹ParticleFlow will also reconstruct Missing Transverse Energy which is of key importance.

²Using CMSSW_3.1.0_preX

□ Comparison with ATLAS (CERN-OPEN-2008-020): Jet Energy Resolution



• There are still areas for improvements in CMS : our challenge for oct. 09 !

Figure 6.1: Comparison of ParticleFlow jets with Calorimeter jets and jets in ATLAS [32].

Summary

By the end of 2009, the Large Hadron Collider will start colliding proton beams. One of the large experiments that will collect the collision data is the Compact Muon Solenoid. The study of top-quarks through their decays is a key part of the physics programme of the CMS Collaboration. As most of the top-quark decays contain jets, the jet reconstruction is a crucial part of the CMS reconstruction software. Thanks to the ParticleFlow method, which combines information of all the subdetectors to reconstruct all stable particles in each event, jets can be reconstructed with higher precision compared to the use of only calorimeter information. In the first part, individual PFJets were compared to their corresponding CaloJet. This revealed that the p_T resolution is better for PFJets mainly for jets with a low p_T where the gain can be as high as 40% in the Barrel and 25% in the Endcaps. It was also shown that the p_T response in the Barrel and Endcaps is a factor 3 better for PFJets than for CaloJets. Also the angular resolution was checked. It was observed that PFJets improve the η resolution by a factor of 2 while improving the ϕ resolution by even a factor 3. Furthermore the mass of the top-quark was reconstructed with CaloJets and PFJets. A comparison between the two reconstructions was made using GenJets as a benchmark. It was observed that a gain of 10% on the width of the mass distribution is obtained with PFJets compared to CaloJets and that the mass bias is smaller. Finally a check of the $t\bar{t}$ matching efficiency revealed that for low cuts on the GenJet p_T , an improvement of around 6% in efficiency can be obtained with PFJets. In the last part of this thesis, the ParticleFlow method was used in the reconstruction and selection of $Z' \rightarrow t\bar{t}$ events. Firstly, the mass of the Z' boson was reconstructed via the invariant mass of the $t\bar{t}$ system. Again a smaller bias was obtained with PFJets and the resolution improved by roughly 10% compared to the CaloJet reconstruction. Another interesting aspect of these events is that the top-quarks can be boosted leading to a merging of two or three quarks into one jet. These boosted top quarks are a good probe for the discovery of the Z' boson using the SuperJet mass to select them. When using ParticleFlow to produce the mass distribution of the two SuperJets rather than the CaloTowers, the peak representing the mass of the top-quark is much more narrow and closer centered to the actual top-mass.

Samenvatting

Eind 2009 zal de Large Hadron Collider opgestart worden. Een van de grote experimenten dat de gegevens van deze botsingen zal verzamelen is de Compact Muon Solenoid, of CMS detector. De studie van top-quarks via hun verval is een belangrijk deel van het onderzoeksprogramma in de CMS Collaboratie. Omdat de meeste van de top quark vervallen quarks bevatten in de eindtoestand, is de jet reconstructie van cruciaal belang in de reconstructie software van CMS. Dankzij de ParticleFlow methode, die de informatie van alle sub-detectors van CMS combineert om zo alle stabiele deeltjes in de eindtoestand te reconstrueren, kunnen jets met hogere precisie gereconstrueerd worden dan wanneer alléén de informatie van de calorimeter gebruikt wordt. De vergelijking tussen individuele PFJets en CaloJets onthulde dat, vooral bij jets met een lage p_T , de resolutie van de PFJets beter is dan die van de CaloJets. Deze verbetering kan tot 40% zijn in de Barrel en tot 25% in de Endcaps. Verder werd er aangetoond dat de p_T respons van de PFJets in het algemeen een factor 3 beter is dan die van CaloJets. Tot slot werd ook de hoekresolutie bepaald en hier werd getoond dat de verbetering met de PFJets een factor 2 bedraagt voor de η resolutie en een factor 3 voor de ϕ resolutie. Vervolgens werd de massa van de top quark gereconstrueerd met CaloJets en PFJets die dan vergeleken werden met de reconstructie via GenJets waaruit blijkt dat de PFJets de breedte van de distributie verbeteren met 10%. De verschuiving van de massa distributie met PFJets ten opzichte van de distributie met GenJets is ook kleiner dan bij CaloJets. Tenslotte werd ook de efficiëntie bepaald om een $t\bar{t}$ paar te identificeren. Het blijkt dat voor een lage snede op de p_T van de GenJets een verbetering van 6% in de efficiëntie bereikt wordt met PFJets ten opzichte van CaloJets. De ParticleFlow techniek werd ook gebruikt in de reconstructie en de selectie van $Z' \rightarrow t\bar{t}$ gebeurtenissen. Eerst en vooral werd de massa van het Z' boson bepaald via de invariante massa van het $t\bar{t}$, waar opnieuw een kleinere verschuiving werd waargenomen voor de reconstructie met PFJets. Hier werd ook aangetoond de massa distributie in het geval van PFJets 10% smaller is dan in het geval van CaloJets. Een ander interessant aspect van deze gebeurtenissen is dat de top quarks een boost kunnen hebben. Dit zorgt ervoor dat in sommige gevallen, twee of drie quarks uit het top quark verval fragmenteren in één enkele jet. Deze boosted top quarks kunnen gebruikt worden in de zoektocht naar het Z' boson. Om de boosted top quarks te selecteren, kan men gebruik maken van de SuperJet massa. Wanneer men gebruik maakt van ParticleFlow om deze massa te reconstrueren, bevat de distributie van de SuperJet massa scherpere piek die véél dichter bij de massa van de top quark gecentreerd is.

Appendix A

Flavour Corrections for PFJets

In this thesis, all jets are corrected using L2 and L3 corrections. No L5 corrections were applied because they are not suited for Particle Flow. This can be seen in the plots in Figure A.1 where it is clear that the L5 correction shifts the $\frac{\Delta p_T}{p_T}$ -distribution away from zero for both b-quark and light-quark jets. This is expected because the ParticleFlow algorithm reconstructs and identifies charged hadrons, neutral hadrons and photons individually with their own calibration. In Figure A.3, the Jet Energy Scale for different flavours of ParticleFlow jets is shown which only changes by a few %. For the CaloJets, this is the opposite. Figure A.2 shows that for all flavours, the response is improved when applying the L5 correction on top of the L2 and L3 corrections.

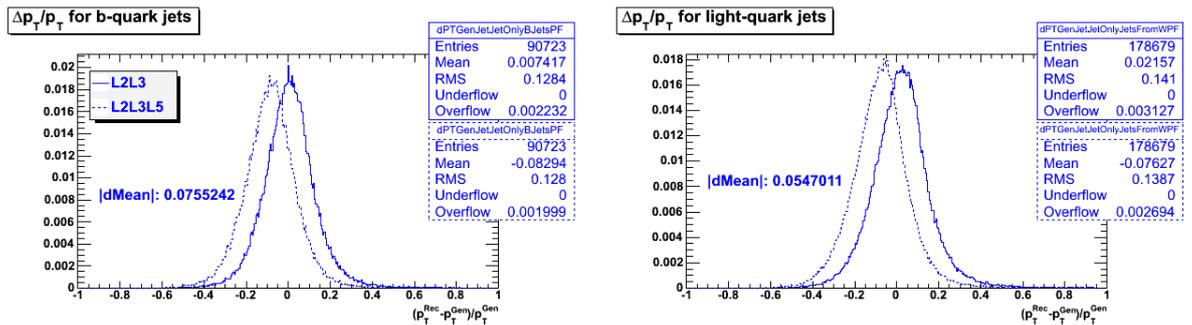


Figure A.1: L5 corrections applied on L2L3-Corrected PFJets.

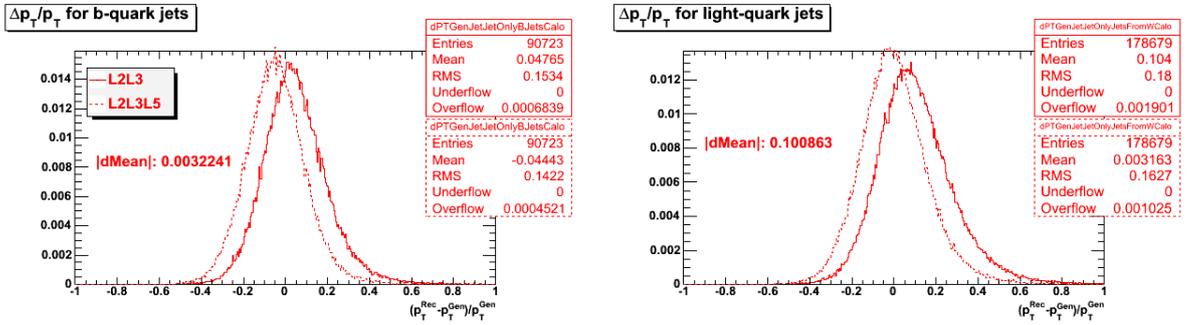


Figure A.2: L5 corrections applied on L2L3-Corrected CaloJets.

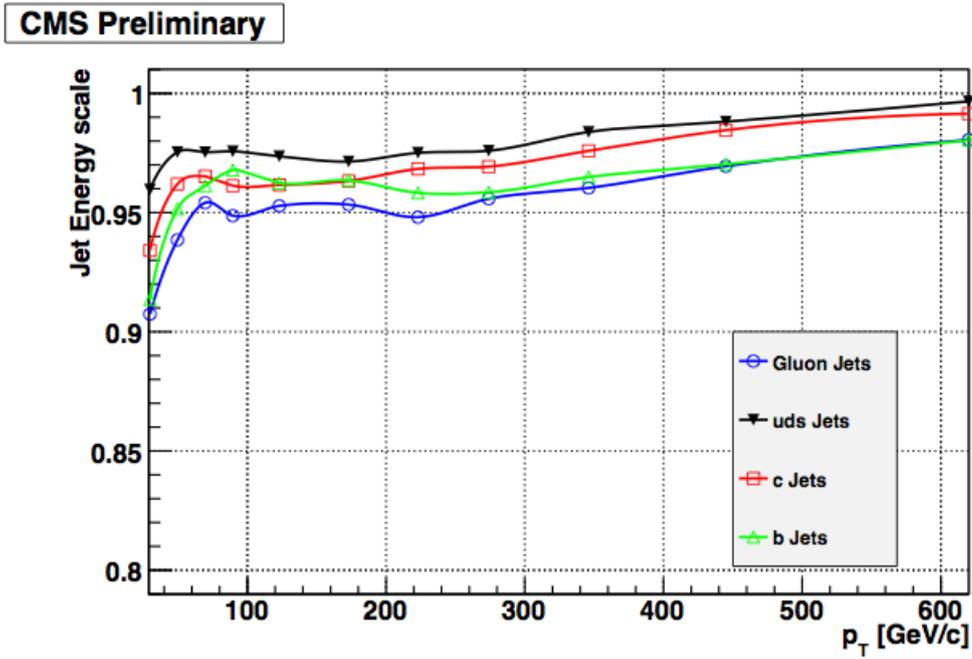


Figure A.3: Jet Energy Scale for different flavours of PFJets[25].

Appendix B

Gaussian fit for resolution plots

In Chapter 4, a resolution plot was constructed using a gaussian fit in each bin of the observable on the x-axis. The choice for a fit was motivated by the fact that the RMS (and the mean in smaller extent) of the histogram in each bin is strongly dependent on the statistics in that bin and a Gaussian fitfunction was chosen based on the distribution shown for two bins in Barrel and Endcaps (Figures 4.3 and 4.4). If the Gaussian is indeed the proper function to use, a fit over the whole histogram range ($-1 \rightarrow +1$) should produce the same results as a fit around the maximum bin (10 bins on the left and 10 on the right). This is indeed the case as is shown in Figure B.1.

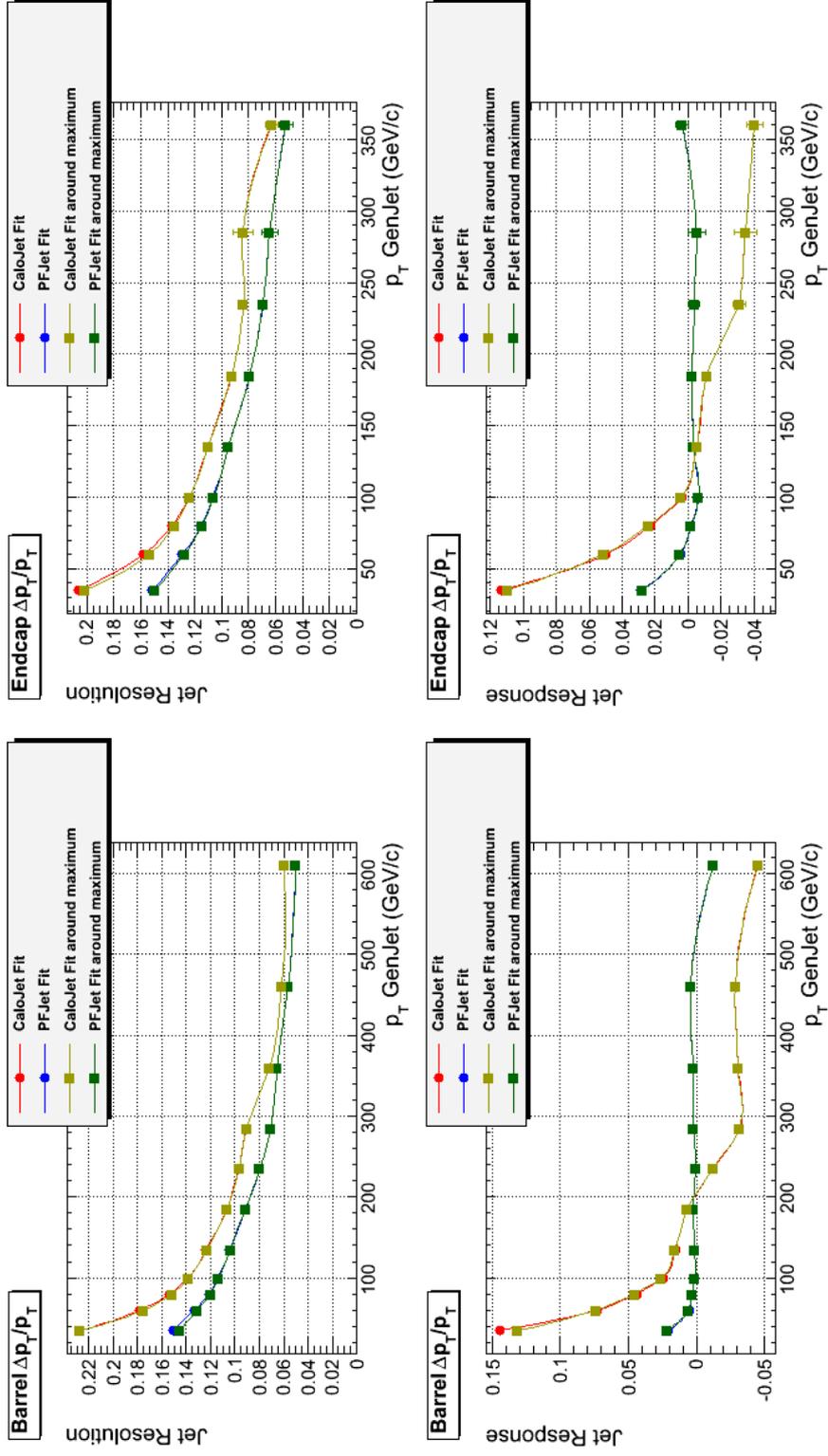


Figure B.1: p_T resolution and response in Barrel and Endcaps using both fits.

Bibliography

- [1] F. Mandl and G. Shaw, *Quantum Field Theory*, (1984) 358 p.
- [2] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory*, (1995) 815 p.
- [3] F. Englert and R. Brout, *Phys. Rev. Lett.* **13** (1964) 321-322.
- [4] *Particle Data Group*,
<http://pdg.lbl.gov>.
- [5] P.W. Higgs, *Phys. Rev. Lett.* **13** (1964) 508-509.
- [6] Higgs mass limits,
<http://pdg.lbl.gov/2008/tables/rpp2008-sum-gauge-higgs-bosons.pdf>.
- [7] Searches for a high mass Standard Model Higgs boson at The Tevatron,
[arXiv:0905.2090v1](https://arxiv.org/abs/0905.2090v1) [[hep-ph](#)].
- [8] CDF Collaboration, A combination of CDF and DO results on the mass of the top quark, [hep-ex/0703034](#).
- [9] A. Ivanov, Top Quark Properties from Top Pair Events and Decays, 34th International Conference on High Energy Physics, Philadelphia, 2008.
- [10] A. H. Hoang, Top Threshold Physics, [arXiv:0604185v1](https://arxiv.org/abs/0604185v1) [[hep-ph](#)].
- [11] L. Evans and P. Bryant, The LHC machine, 2008 JINST 3 S08001.
- [12] The LHC machine:
<http://lhc.web.cern.ch>.
- [13] The CMS Collaboration, S. Chatrchyan et. al., The CMS experiment at the CERN LHC, 2008 JINST 3 S08004.

- [14] The CMS experiment:
<http://cms.cern.ch>.
- [15] The ATLAS Collaboration, G. Aad et. al., The ATLAS Experiment at the CERN Large Hadron Collider, 2008 JINST 3 S08003.
- [16] The ATLAS experiment,
<http://atlas.web.cern.ch>.
- [17] The ALICE Collaboration, K. Aamodt et. al., The ALICE experiment at the CERN LHC, 2008 JINST 3 S08002.
- [18] The ALICE experiment:
<http://aliceinfo.cern.ch/Collaboration/>.
- [19] The LHCb Collaboration, A. Augusto Alves Jr et. al., The LHCb Detector at the LHC, 2008 JINST 3 S08005.
- [20] The lhc-b experiment:
<http://lhcb.web.cern.ch>.
- [21] K. Wille, The Physics of Particle Accelerators: an introduction, Oxford Press 2005.
- [22] The CMS Collaboration, Detector performance and software, Physics Technical Design Report, Volume I (CMS TDR 8.1).
- [23] The theoretical top cross section at the Tevatron and the LHC, [arXiv:0805.3844v1](https://arxiv.org/abs/0805.3844v1) [hep-ph].
- [24] Run-II Jet Physics, [arXiv:hep-ex/0005012v2](https://arxiv.org/abs/hep-ex/0005012v2).
- [25] The Particle Flow Physics Object Group, ParticleFlow Reconstruction of Jets, MET and Taus, **CMS-AN2009/039**.
- [26] The CMS Collaboration, Plans for Jet Energy Corrections at CMS, **CMS PAS JME-07-002**.
- [27] MadGraph/MadEvent v4: The New Web Generation, Johan Alwall, Pavel Demin, Simon de Visscher, Rikkert Frederix, Michel Herquet, Fabio Maltoni, Tilman Plehn, David L. Rainwater and Tim Stelzer **JHEP 0709:028,2007**, [arXiv:0706.2334](https://arxiv.org/abs/0706.2334).
- [28] Geant4 - A Simulation Toolkit, S. Agostinelli et. al., Nuclear Instruments and Methods A 506 (2003) 250-303.

- [29] A Brief Introduction to PYTHIA 8.1, **arXiv:0710.3820v1** [**hep-ph**].
- [30] S. Eidelman et al. (Particle Data Group), Phys. Lett. B **592**, 1 (2004) (URL: <http://pdg.lbl.gov>).⁴
- [31] R. M. Harris, C. T. Hill and S. J. Parke, Cross-section for top-color Z -prime(t) decaying to t anti- t , **FERMILAB-FN-0687**, **hep-ph/9911288**.
- [32] P. Janot, CMS Week Plenary Meeting, ParticleFlow Event Reconstruction with CMS, 19 March 09.