Top quark mass measurement in the electron+jets decay channel using the Matrix Element Method

Guillaume Garcia

Departement Fysica en Sterrenkunde

Universiteit Gent

Acknowledgements

Scientific discovery and scientific knowledge have been achieved only by those who have gone in pursuit of it without any practical purpose whatsoever in view.

-Max Planck (1848-1947)

Samenvatting

Moderne deeltjesfysica wordt beschreven door het Standaard Model. Dit is ontwikkeld sinds de jaren 1970 en beschrijft de theorie achter drie van de in het heelal aanwezige vier fundamentele krachten, het elektro-magnetisme, de zwakke interactie en de sterke interactie. De eerste twee zijn succesvol verenigd in de elektro-zwakke theorie. Het Standaard Model beschrijft de interacties tussen materie (quarks en leptonen) en de fundamentele krachtendragers (ijkbosonen). De recente ontdekking van een nieuw deeltje met de eigenschappen van het Standaard Model Higgs boson, vertegenwoordigt een bijkomend succes voor dit theoretisch model. Een nauwkeurige test van de Standaard Model parameters is niettemin noodzakelijk om de limieten van de theorie af te tasten en kan uiteindelijk leiden tot de ontdekking van nieuwe natuurkundige processen. Eén van deze vrije parameters is de top quark massa van $173.34\,\mathrm{GeV}/c^2$, gemeten met een nauwkeurigheid van plus of minus $0,76\,\mathrm{GeV}/c^2$.

De top is het zwaarste van alle bekende fundamentele deeltjes. Hij werd ontdekt in 1995 door twee experimenten (CDF en DØ) aan de Fermilab versneller in de buurt van Chicago, USA. De top quark is zeer interessant om te bestuderen vanwege enkele unieke kenmerken. Dit kleine deeltje is bijna even zwaar als een goudatoom en heeft een zeer korte levensduur. Zijn imposante massa in vergelijking met die van andere quarks betekent dat het een zeer sterke heeft koppeling met het Higgs boson. Zijn zeer korte levensduur laat het niet toe gebonden toestanden te vormen, waardoor de studie van naakte quark eigenschappen mogelijk. Het is ook betrokken in vele zoektochten naar nieuwe fysica

als vervalproduct of als achtergrondbijdrage.

Het onderzoeksprogramma van de Large Hadron Collider (LHC), de opvolger van de nu afgeschakelde Fermilab verscneller, is gericht op het creëren van de voorwaarden om hoog-energetische deeltjes te produceren. Dit maakt het onderzoek mogelijk van een groot aantal fysische processen, maar in het bijzonder, met zijn hoge massacentrum energie en hoge luminositeit, is de versneller in staat om duizenden top quarks maken. De LHC versneller faciliteit in CERN (Genève, CH) bevindt zich in de tunnel eerder gebruikt door de Large Electron-Positron Collider. De LHC is een bijna-cirkelvormige proton-proton versneller met een omtrek van 27 kilometers onder de Zwitsers-Franse border. Langs deze perimeter zijn vier grote detectoren geïnstalleerd. De Compact Muon Solenoid (CMS) detector is één van hen, en is ontworpen voor het identificeren en detecteren van de deeltjes geproduceerd in de botsingen die plaatsvinden in het midden van de detector. Analyses worden vervolgens uitgevoerd op de botsingsgegevens geregistreerd door de detector en deze helpen bij het verfijnen van de Standaard Model voorspellingen.

Het in dit proefschrift beschreven werk betreft de analyse van proton-proton botsingen die zich hebben voorgedaan in de loop van het jaar 2012 bij een massacentrum energie van 8 TeV. Gegevens die door de detector werden geregistreerd, werden gefilterd om botsingen te isoleren die aanleiding gaven tot een top-anti-top quark paar. Deze quarks vervallen in een elektron en vier jets. Voor een opgenomen luminositeit van 19,8 fb-1, zijn 8898 gebeurtenissen geanalyseerd na een zeer strikte selectie. De analyse werd uitgevoerd met de Matrix Element Methode, één van de meest accurate methodes voor een gegeven set van evenementen. Hierbij worden geselecteerde gegevens onderzocht op een event-by-event basis en wordt maximaal gebruik gemaakt van de kinematica van het evenement. De methode is zeer flexibel in de zin

dat ze om het even welk theoretisch model kan testen, terwijl rekening gehouden wordt met alle mogelijke effecten geïntroduceerd door experimentele onzekerheden. Deze analyse leidde uiteindelijk tot een top quark massa meting van $170.94 \pm 2.22\,\mathrm{GeV}/c^2$

Contents

List of Figures

Introduction

Curiosity has always pushed Mankind to understand and interpret its surroundings. Over the last century, the development of Quantum Physics allowed a better comprehension of the infinitesimal world and the origin of the universe as we know it.

Since the mid-twentieth century, the Standard Model of Particle Physics (SM) - a quantum field theory - has proven to be successful at predicting the existence of fundamental particles and their interactions. The heaviest of these twelve particles (six quarks, six leptons) is the top quark. This tiny particle is as heavy as a gold atom, but has a very short lifetime, of about $5 \times 10^{-25}s$. Those unique properties used to make it very difficult to produce and observe. Its experimental discovery has been achieved at the Tevatron accelerator facility in 1994.

The Large Hadron Collider (LHC), a proton-proton collider with a higher centre-of-mass energy, is considered to be a "top quark factory". We use data collected during 2012 to measure the mass of the top quark in the semi-electronic decay channel. For this purpose, we apply the Matrix Element Method on rigorously selected collision events. This method, based on an event-by-event analysis is the most precise for a given set of events. Achieving a precision measurement on the top quark properties is very interesting given its uniqueness compared to other fundamental particles, and also helps in performing precision tests of the Standard Model.

This thesis is decomposed into 5 parts. The first chapter outlines both theoretical and experimental frames in two distinctive chapters. The former contains

an introduction to the Standard Model, its fundamental particles - more specifically the top quark - and its application to proton-proton collisions. The latter presents the accelerator facility at CERN and a close look at the Compact Muon Solenoid (CMS) detector which provides data used in this analysis. The second part is dedicated to the detection and the selection of top quark pair events. First, an introduction on how physics processes are simulated using Monte Carlo generators is presented. Then, an overview on how the signals from the various detector elements are reconstructed into physical objects is detailed. Finally, the selection criteria that are applied on reconstructed objects to ensure the highest purity in the final sample are explained. The third part shows the principle of the Matrix Element Method (MEM), the theory behind it and all its important elements that come into play. The fourth part focuses on the application of the MEM. It begins with the validation of the method on simulated events, starting from the simplest case (events generated at the parton level), complicating more and more the simulation to reflect effects coming from the detector. Then, the calibration of the method is discussed, before moving to the actual measurement and the estimation of the statistic and different systematic uncertainties. The last part describes a combination of the result from this analysis and other top quark mass measurements from the collaboration, using the Best Linear Unbiased Estimator (BLUE) method. The final result is discussed and also compared to individual results on the top quark mass from CMS and to the latest combination from CMS, ATLAS, CDF and DØ.

Part I

Theoretical and Experimental Introduction

Chapter 1

The Standard Model and the Top Quark

The fundamental particles and their interactions are described by the Standard Model of particle physics. It groups the quantum theoretical models of the electromagnetic, weak and strong interactions. This model does not yet include gravitation, the fourth major interaction. In addition, due to the infinitesimal size and high energy of the interacting objects, in particle accelerators physics, gravitation effects can be neglected. In this thesis, the Heaviside-Lorentz convention is adopted, thus $\hbar = c = 1$.

In this chapter, the theoretical frame of the analysis is introduced. The first half of this chapter includes an overview of the fundamental Standard Model particles and interactions, while the second brings more details on the top quark study.

1.1 Fundamental Particles

Table 1.1 shows the fundamental particles of the Standard Model and their basic properties. Each particle listed here has its associated antiparticle, which has the same mass and opposite charge. Electrical charges are expressed as multiples of the elementary charge $e = 1.602 \times 10^{-19}$ C.

	Interaction	Name	Charge	Mass
Leptons spin = 1/2	Weak	electron neutrino (ν_e)	0	< 22 eV
	EM, Weak	electron (e)	-1	0.511 MeV
	Weak	muon neutrino (ν_μ)	0	< 0.17 eV
	EM, Weak	muon (μ)	-1	105.7 MeV
	Weak	tau neutrino (ν_τ)	0	< 15.5 eV
	EM, Weak	tau (τ)	-1	1.777 GeV
Quarks spin = 1/2	EM, Weak, Strong	up (u)	2/3	2.4 MeV
		down (d)	-1/3	4.8 MeV
		charm (c)	2/3	1.27 GeV
		strange (s)	-1/3	104 MeV
		top (t)	2/3	173.4 GeV
		bottom (b)	-1/3	4.2 GeV
Gauge bosons spin = 1	EM	photon (γ)	0	0
	Weak	W boson (W^+, W^-)	±1	80.4 GeV
	Weak	Z boson (Z)	0	91.2 GeV
	Strong	gluon (g)	0	0
Higgs boson spin = 0	EM, Weak, Strong	H	0	125.0 GeV

Table 1.1: Table listing the Standard Model fundamental particles and their properties.

1.1.1 Fermions

Fermions are the components of matter found in the universe. They are half integer spin particles that obey the Pauli exclusion principle, which states that no fermions with the same quantum state can simultaneously exist.

Fermions are divided in two categories of quarks and leptons. There are three generations of quarks and leptons, each generation is made of two particle doublets, giving in total 6 quarks and 6 leptons. The first generation is constituted of the lightest and stable fermions. The next generations are composed of heavier and unstable particles. Consequently, fermions from a the second or third generations decay into fermion from a lower generation.
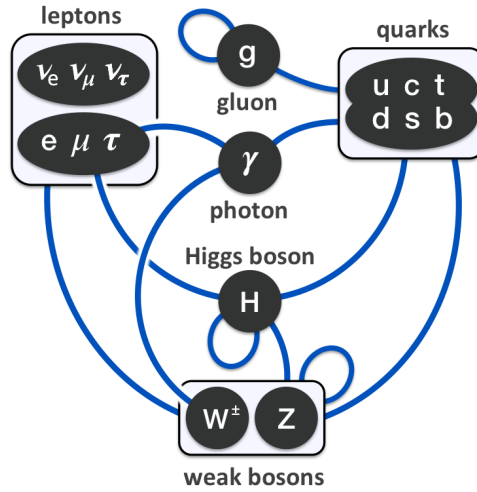
Figure 1.1: Diagram showing the possible interactions between Standard Model fundamental particles

1.1.1.1 Leptons

Each lepton generation is composed of a particle doublet. Each doublet is made of a charged lepton, carrying the fundamental charge $-e$, and an associated neutral lepton called *neutrino*. The lepton from the first generation is the *electron* (e)), which is stable. Leptons from second and third generations, the *muon* (μ) and the *tau* (τ) are produced in high energy particle interactions. Charged leptons are subjects to the electroweak interaction.

Neutrinos do not have electrical charge and have a very small mass compared to charged leptons. They only interact through the weak force. Consequently, they are very difficult to observe in particle collider detectors. Experimentally, this implies missing energy when reconstructing the final state kinematics.

1.1.1.2 Quarks

Quarks have a fractional electric charge, worth $+2/3e$ for up-flavour quarks (u,c,t) and -1/3e for down-flavour quarks (d,s,b). In addition, they also carry a colour charge. Those properties allow them to interact via both the electroweak

and the strong forces.

Quarks cannot exist in an isolated state. Colour confinement constrain them to hadronise and form composite particles where quarks are bound together due to the strong force. Such particle multiplets are called hadrons. Amongst hadrons, one can several categories. Baryons, made of 3 quarks, like the proton (uud) and the neutron (udd), and mesons, made of a quark and an antiquark. There are also exotic hadrons made of four or more quarks. The LHCb collaboration recently observed a tetra-quark $Z(4430)$ [20] and has been able to confirm the existence of pentaquarks [21]. The top quark though, is not subject to hadronisation. It has a very short lifetime ($\sim 5 \times 10^{-25}$ s) so it decays before having the time to hadronise. Further discussions on top quark physics can be found in section 1.2.

1.1.2 Gauge Bosons and Interactions

1.1.2.1 Electromagnetic and Weak Interactions

Photons are mediators of the electromagnetic (EM) force, described by quantum electrodynamics (QED). Photons are massless, charge-less, bosons with unitary spin. QED is a field theory with the symmetry group $U(1)$, describing interactions between photons and fermions involving photon exchanges.

The weak force (group $SU(2)$) is carried by W and Z bosons and affects all the Standard Model fermions. Gauge bosons associated to the weak interaction are massive and have a short lifetime ($\sim 10^{-25}$ s), therefore the weak force can act only on very short distances.

W$^\pm$ bosons are electrically charged and have a mass of 80.385 ± 0.015 GeV [22]. Fermions can, by emitting/absorbing a W boson, change from one flavour group to another. The probabilities from a quark to move from the up-flavour group to the down-flavour are given by the Cabibbo-Kobayashi-Maskawa (CKM) matrix [22]:

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \tag{1.1}$$

The probability for a quark i to transform into a quark j is proportional to $|V_{ij}|^2$. All the V_{CKM} matrix have been independently and experimentally derived, as:

$$|V_{CKM}| = \begin{pmatrix} 0.97427 \pm 0.00015 & 0.22534 \pm 0.0006 & 0.00351^{+0.00015}_{-0.00014} \\ 0.22520 \pm 0.00065 & 0.97344 \pm 0.00016 & 0.0412^{+0.0011}_{-0.0005} \\ 0.00867^{+0.00029}_{-0.00031} & 0.0404^{+0.0011}_{-0.0005} & 0.999146^{+0.000021}_{-0.000046} \end{pmatrix} \tag{1.2}$$

Z bosons have no electrical charge and a mass of 91.1876 ± 0.0021 GeV [22]. In a fundamental interaction involving the exchange of a Z boson, apart from a momentum transfer, particles properties such as charge, flavour and colour remain unchanged.

W$^{\pm}$ bosons decay into fermions, either into a lepton and a neutrino of a given flavour group (10.86%), or in up-type quark and a down-type quark (67.41%). Z boson decay into a fermion and its associated antiparticle, so either a l^+l^- pair (3.36%) or a $q\bar{q}$ pair (69.91%). The numbers in parenthesis correspond to the branching ratio of the bosons decay modes and are extracted from [22].

1.1.2.2 The Higgs Mechanism

The Higgs mechanism represents the spontaneous symmetry breaking of the electroweak theory that unifies both EM and weak interactions. At low energies the two forces are decoupled, but at high energies such as existed in a very early Universe, they merge into a combined electroweak force.

The idea behind the Higgs mechanism is the existence of a complex doublet of scalar fields (Higgs doublet). When the neutral component of the Higgs dou-

blet reaches a vacuum expectation value the spontaneous electroweak symmetry breaking takes place. In the process, massless Goldstone bosons are produced and absorbed by W and Z bosons which acquire masses. Fermions gain masses when interacting with the Higgs field through Yukawa couplings.

1.1.2.3 Strong Force

Gluons (g) are the carriers of the strong interaction. They only interact with object having a colour charge, *i.e.* quarks and gluons. The strong interaction is described by quantum chromodynamics (QCD), a field theory of symmetry group $SU(3)$. It is responsible, for example, for the cohesion of protons and neutrons and also the binding of the atomic nucleus.

The strong coupling increases when the energy of the interaction decreases or when the interaction distance increases. Those features imply that a quark cannot exist as a free particle, and must form bound states (pairs, triplets...) as discussed in section 1.1.1.

It also explains the hadronisation process, in which a high energy quark will lose its energy by radiating a virtual gluon that will decay into a quark-anti-quark pair. This process will continue until the quark compounds created in the process reach a stable state. Consequently, a quark produced in high energy collisions will be observed as a shower of particles, called a *jet*, in the detector.

1.2 Top Quark Physics at LHC

1.2.1 Proton-proton Collisions

The proton is a compound particle and therefore its structure and possible interactions are more complex to describe than for fundamental particles. The proton is made of three valence quarks (two quarks u and one quark d). Gluons are constantly absorbed and emitted within the proton and can also lead to a quark-anti-quark pair creation. During proton-proton collisions, see figure 1.2, several interactions take place. The hard scattering is the interaction between

quarks and gluons in the protons. Remnants from the original protons form the underlying event. The partons that participate in the hard scattering process can also radiate gluons before and after the collision yielding initial and final state radiations respectively. Photons can also be radiated by final state particles.

Protons are also grouped together (in bunches) to form the colliding beam. For a given collision, many proton-proton interactions occur. In the end, only the event with hard interaction is kept for study, and extra collisions contribute to the event as what is called *pile-up*.
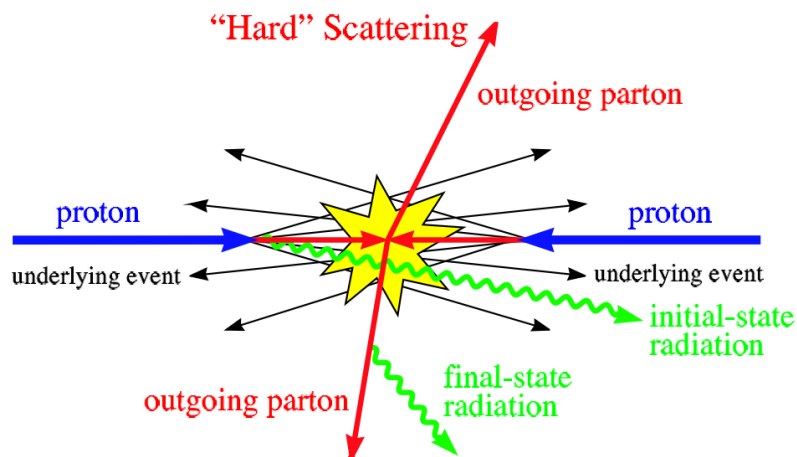
Figure 1.2: Schematic of a proton-proton collision, displaying the main hard scattering, the underlying event activity, and the initial/final state radiations [1].

In hadron colliders, the available centre-of-mass energy can be written as:

$$\sqrt{s} = \sqrt{(p_1 + p_2)^2} = 2E_{beam}, \qquad (1.3)$$

where p_1 and p_2 are the four-momentum of the colliding objects. For this analysis which processes data recorded in 2012, the centre-of-mass energy was $\sqrt{s} = 8\,\mathrm{TeV}$.

1.2.2   The Top Quark

The top quark is the heaviest of all fermions. It has an electrical charge of $+\frac{2}{3}e$. In 1973, Kobayashi and Maskawa postulated the existence of a third generation doublet. Their thoughts were rapidly confirmed in 1977, with the bottom quark discovery. Nonetheless, given its large mass compared to its generation partner, the bottom quark, the top quark requires a much higher centre-of-mass energy collision to be produced. It was first observed in 1995 by the CDF and DØ collaborations [23]. The detectors analysed proton-anti-proton collisions from the *Tevatron* accelerator located at Fermilab, in the vicinity of Chicago.

Several reasons make the top quark important to study. First, its mass is one of the free parameters of the Standard Model, therefore an accurate measurement helps in refining theoretical models. As introduced in section 1.1.1, it is by far the heaviest known quark, with a mass of $m_t = 173.34 \pm 0.27 (\mathrm{stat.}) \pm 0.71 (\mathrm{syst.}) \mathrm{GeV}$ [24]. However, the top quark mass measured in high energy physics experiments is subject to conventions. Here and generally in direct measurements, it is reconstructed as the invariant mass of the top quark decay products and compared to simulation using leading-order Monte Carlo generators. On the theory side, the top quark mass is generally described using the concept of pole mass from perturbation theory, as the top quark can be considered as a bare unstable fermion which does not hadronise [25, 26].

The large mass of the top quark makes it so short-lived that it does not hadronise before decaying. This makes studies of isolated quarks possible, like *e.g.* charge asymmetry and spin correlations. [27, 28]. Top quarks are also present in interactions involving Higgs bosons or hypothetical super-symmetric ($SUSY$) particles, as in direct decay products or in background processes. Also an accurate measurement on the top quark mass, which is close to the electroweak scale, would help in refining Standard Model predictions through loop corrections. State of the art physics and SM extensions would greatly benefit from a better knowledge of the top quark and its properties.

1.2.3 Creation

Top quarks pairs (top-anti-top) can be produced through gluon fusion or quark-anti-quark annihilation, see figure 1.3. At LHC, about 90% of the top quark pair production is ensured via gluon fusion. Single top quarks can also be produced through EWK processes, see figure 1.4. This gives the top quark an important role in testing two fundamental interactions of the SM. The relevant processes giving rise to top quarks are detailed in this section.
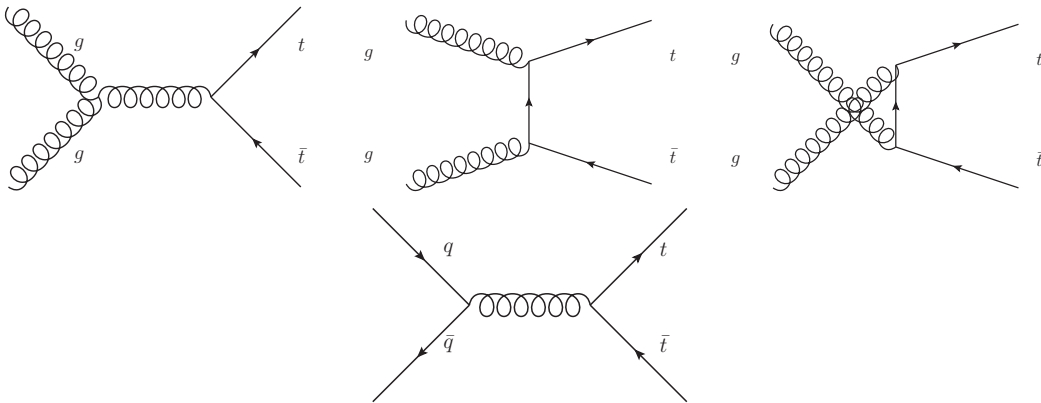
Figure 1.3: Feynman diagrams for $t\bar{t}$ pair production via gluon fusion (top diagrams) and quark annihilation (bottom diagram)

Because of the large mass of the top quark, high centre-of-mass energy collisions are required to produce top quark pairs. At the *Tevatron* the top pair production cross-section has been measured as $\sigma_{t\bar{t}} = 7.60 \pm 0.41$pb [29] for a centre of mass energy of $\sqrt{s} = 1.96$ TeV. With the shutdown of *Tevatron* in 2011, studies on the top quark now continue at the Large Hadron Collider (LHC). The LHC accelerates and collides protons with protons, rather than anti-protons. Due to higher centre-of-mass energy collisions, the production cross-section for a centre-of-mass energy of $\sqrt{s} = 8$ TeV is measured to be 227 ± 3(stat.)± 11(syst.)± 10(lumi.)pb [30].

Top quark pairs are produced via the strong interaction, and therefore are described by QCD. The theoretical $t\bar{t}$ pair production cross-section depends on the
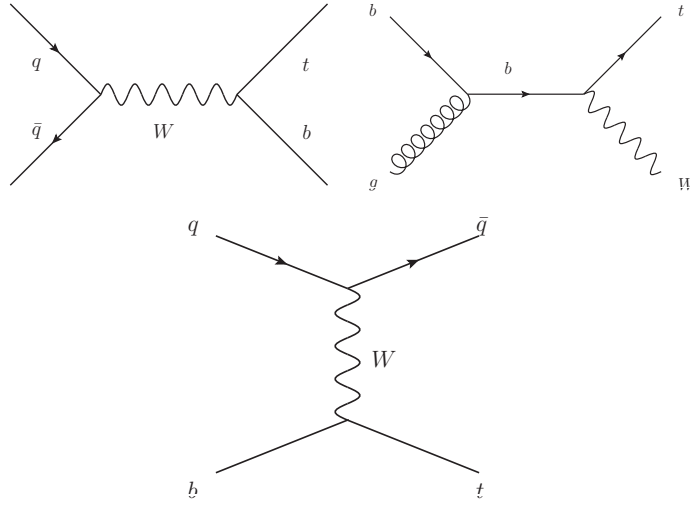
Figure 1.4: Feynman diagrams for EWK single-top production

top quark mass (m_t), the centre-of-mass energy of the collision squared $(\sqrt{s})$ and can be written as:

$$\sigma_{pp \to t\bar{t}}(m_t, s) = \sum_{i,j=q,\bar{q},g} \int dx_i dx_j f_i(x_i, \mu_f^2) f_j(x_j, \mu_f^2) \sigma_{ij \to t\bar{t}}(\hat{s}, m_t, \mu_f, \alpha_s) \qquad (1.4)$$

A sum over all the quarks and gluons (i, j) contributing to the collision is made. x represents the proton momentum fraction carried by individual partons, $f(x, \mu_f)$ are the proton Parton Distribution Functions (PDF) with the energy scale parameter μ_f, see chapter 3. $\hat{s} = x_i x_j s$ is the effective centre of mass energy, α_s is the strong coupling constant and $\sigma_{ij \to t\bar{t}}$ is the partonic cross-section.

Figure 1.5 shows the evolution of the production cross section of several physics processes as a function of the centre of mass energy. Top-anti-top pair production cross section is labelled as σ_t. It increases for higher values of $\sqrt{s}$. This effect, coupled with the high luminosity delivered by the collider, makes the LHC a true *top quark factory*.
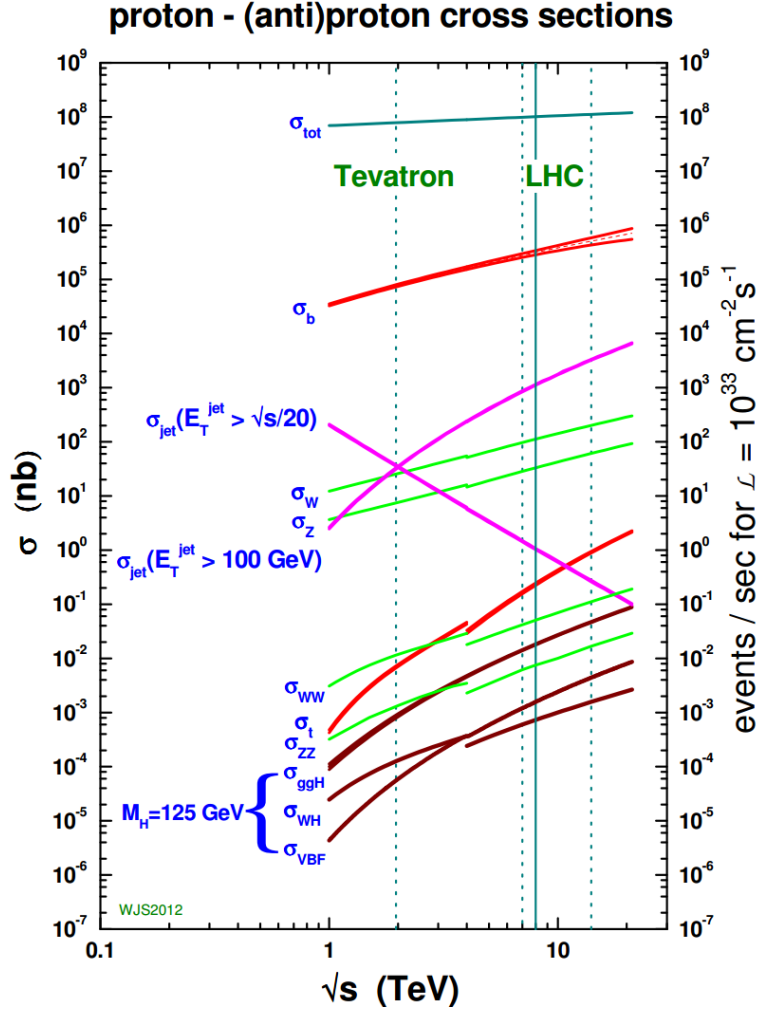
Figure 1.5: Estimated production cross section as a function of centre of mass energy. Vertical lines show centre of mass energies for Tevatron (1.96 TeV) and LHC (7, 8 and 14 TeV). The discontinuity represents the change of $p\bar{p}$ collisions (Tevatron) to pp collisions (LHC) [2]

1.2.4 Decay

The top quark decays almost exclusively into a bottom quark and a W boson given the large value of $|V_{tb}|$ in the CKM matrix. Consequently, the decay products of a top quark pair depend on the decay of the W bosons. The W boson either decays into a pair of quarks or a lepton-neutrino pair. If both W bosons

decay hadronically, the top pair decay channel is labelled as the fully-hadronic, has a probability of 45.7% and the following signature

$$t\bar{t} \rightarrow W^+ W^- b\bar{b} \rightarrow q\bar{q}q\bar{q}b\bar{b} \tag{1.5}$$

If both the W bosons decay leptonically, the top quark pair decay channel is labelled as fully-leptonic, has a probability of 10.5% and the following signature

$$t\bar{t} \rightarrow W^+ W^- b\bar{b} \rightarrow l\nu l\nu b\bar{b} \tag{1.6}$$

If one of the W bosons decays leptonically and the other ones decays hadronicaly, the top quark pair decay channel is labelled as the semi-leptonic, has a probability of 43.8% and the following signature

$$t\bar{t} \rightarrow W^+ W^- b\bar{b} \rightarrow l\nu q\bar{q}b\bar{b} \tag{1.7}$$
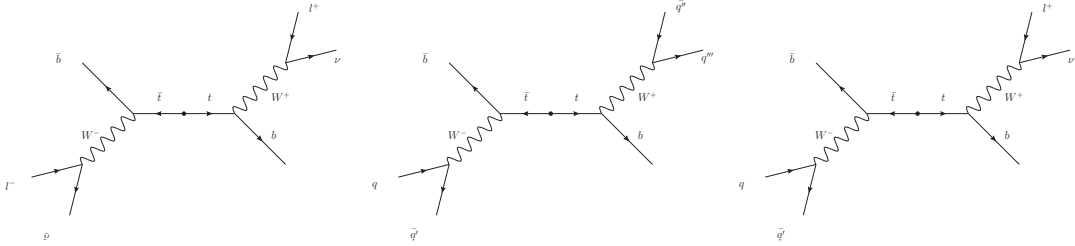
Figure 1.6: Feynman diagrams of the different $t\bar{t}$ decay modes: fully-leptonic (left), fully-hadronic (centre) and semi-leptonic (right)

The different top quark pair decay modes and their experimental signatures, shown in figure 1.6 present different features for the measurement of the top quark properties. The fully-hadronic channel, while being the most common, implies six quarks as decay products that will hadronise and give rise to six jets. This makes difficult to identify top quark events among the large multi-jet background at LHC. On the opposite, the fully-leptonic channel has a very clear experimental

signature, with two high energy isolated leptons. But the presence of two neutrinos which travel through the detector unnoticed and the small branching fraction make the top quarks kinematics reconstruction challenging. The semi-leptonic channel is considered as a good compromise between the two decay channels mentioned above. It has a branching fraction relatively large compared to the fully-leptonic decay mode and its experimental signature is easier to identify and separate from background contributions compared to the fully-hadronic channel.

Chapter 2

The LHC and CMS

The Large Hadron Collider (LHC) is a circular particle accelerator located at the European Organisation for Nuclear Research (CERN) below the French-Swiss border in the surroundings of Geneva, in Switzerland. The LHC is the largest particle collider up to date. Operations began in 2010 with collisions of a centre-of-mass energy of $7\,\mathrm{TeV}$. In 2012, the centre-of-mass energy ramped up to $8\,\mathrm{TeV}$. The LHC has been designed to allow in the future collisions at a centre-of-mass energy of $14\,\mathrm{TeV}$, with an instantaneous luminosity of $10^{34}\,\mathrm{cm}^{-1}\,\mathrm{s}^{-1}$. Most of the beams circulating into the accelerator are made of protons, but the LHC can also be used as a heavy-ion collider by replacing one or both proton beams by lead ion beams.

In this chapter, the general structure of the LHC accelerator facility is detailed. A brief description of experiments placed alongside the LHC tunnel is also given. Finally, a general description of the Compact Muon Solenoid (CMS) detector can be found in a dedicated section. It emphasizes the working principle and the goals of each sub-detector present in CMS, based on the Technical Design Report (TDR) [3].

2.1 The LHC

The LHC was built during the last decade, in the tunnel that used to host the Large Electron-Positron collider (LEP) on the CERN site. The tunnel follows a quasi-circular shape and lies underground at a depth between 45 m and 170 m. It has been excavated in the second half of the 1980's and has a total length of 26.7 km. The LEP was shut down in 2000 and the construction of the LHC began afterwards. After a commissioning phase in 2009, first collisions at a centre-of-mass energy of 7 TeV occurred in March 2010.

2.1.1 Design and Operation

The LHC has been designed to probe the TeV scale as a successor of the Tevatron proton-anti-proton collider at Fermilab. The LHC aims towards colliding particles at higher energy, but also at a higher rate. This rate, or number of events per second for any process is given by

$$\frac{dN}{dt} = \mathscr{L}\sigma \tag{2.1}$$

where σ is the cross-section for the studied process, and $\mathscr{L}$ is the instantaneous luminosity, depending on the beam parameters:

$$\mathscr{L} = \frac{N_b^2 n_b f_{rev} \gamma}{4\pi \epsilon_n \beta^*} F \tag{2.2}$$

N_b is the number of accelerated particles per bunch, n_b is the number of bunches per beam and f_{rev} is the revolution frequency. The beam profile is given by γ the relativistic factor, ϵ_n the normalised beam emittance, the β^* function at the collision point representing the transversal size of the beam, and F the geometric reduction introduced to correct the beam directions non-perfect alignment. F writes as

$$F = \frac{1}{\sqrt{1 + \left(\frac{\theta_c \sigma_z}{2\sigma^*}\right)^2}} \tag{2.3}$$

where θ_c is the beams crossing angle, σ_z the root mean square bunch length and $\sigma*$ the root mean square beam size. From the equations listed above, a high rate implies a high luminosity and therefore requires beams with high intensity and energy. Figure 2.1 shows the evolutions of both the integrated luminosity delivered by LHC and integrated luminosity recorded by CMS during 2012 data taking period and for proton-proton collisions. The integrated luminosity is simply the integration of instantaneous luminosity over time.

$$\mathscr{L}_{int} = \int \mathscr{L} dt \tag{2.4}$$

CMS Integrated Luminosity, pp, 2012, $\sqrt{s} = 8$ TeV
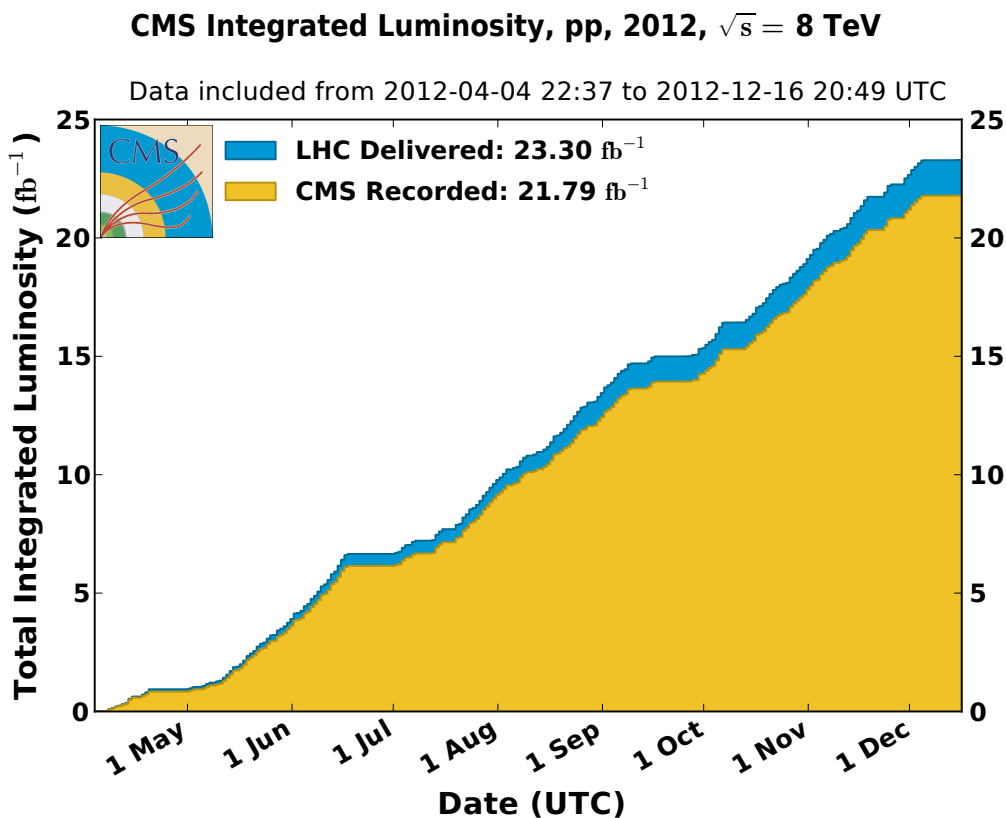
Figure 2.1: Integrated luminosity delivered by the LHC (blue) and recorded by the CMS detector (yellow) as a function of time during the 2012 data taking period

Considering the luminosity requirements the LHC has to achieve, *i.e.* beams

with a high intensity and a high energy, it was decided the LHC would accelerate protons. Indeed, proton beam bunches are directly obtained from hydrogen ionisation, which is rather cheap and simple when compared to the production of anti-proton bunches. Moreover, electron and positrons, due to having a smaller mass, tend to loose a lot of energy through radiation (bremsstrahlung).

Figure 2.2 shows the LHC accelerator complex at CERN. The proton acceleration proceeds in different steps. First, a proton stream is accelerated up to 50 MeV in the *Linac*2 then injected into the *Proton Synchrotron* (PS). Proton bunches are formed inside the PS and their energy ramps up to 26 GeV. From the PS, proton bunches are injected into the *Super − Proton Synchrotron* (SPS) where they are accelerated up to an energy of 450 GeV. Finally, proton bunches are injected in the LHC, spaced in time by 50 ns, and are further accelerated to 4 TeV.

In the LHC, the proton beams circulate in two separate vacuum pipes, represented on figure 2.3. The circular, 27 km long collider is made of eight octants. Each octant can be divided into one arced and one straight section. The beams trajectories are bended with the use of superconducting Nb-Ti dipole magnets delivering magnetic fields up to 8.3 T and kept focused using superconducting quadrupole magnets. Proton bunches are injected at interaction points (IP) 2 and 8. They are accelerated in radio frequency (RF) cavities located in the fourth octant. The two beams cross from one magnet bore to the other at only four IP (1,2,5 and 8), where the detectors are located and where collisions take place. Sections 2 and 8 each contain in addition the injection systems (TI2,TI8). The sixth octant contains the sections used to dump the beams out of the collider with the help of horizontally deflecting magnets. Sections 3 and 7 each contain two collimation systems.

During the 2012 data taking period, the LHC accelerated and collided beams with an energy of 4 TeV. The beams were made of 1374 bunches spaced by 50 ns, each bunch containing $\sim 1.7 \times 10^{11}$ protons. With this configuration, a peak luminosity of $7.7 \times 10^{33}\,\mathrm{cm^{-2}\,s^{-1}}$ has been achieved, allowing CMS to record
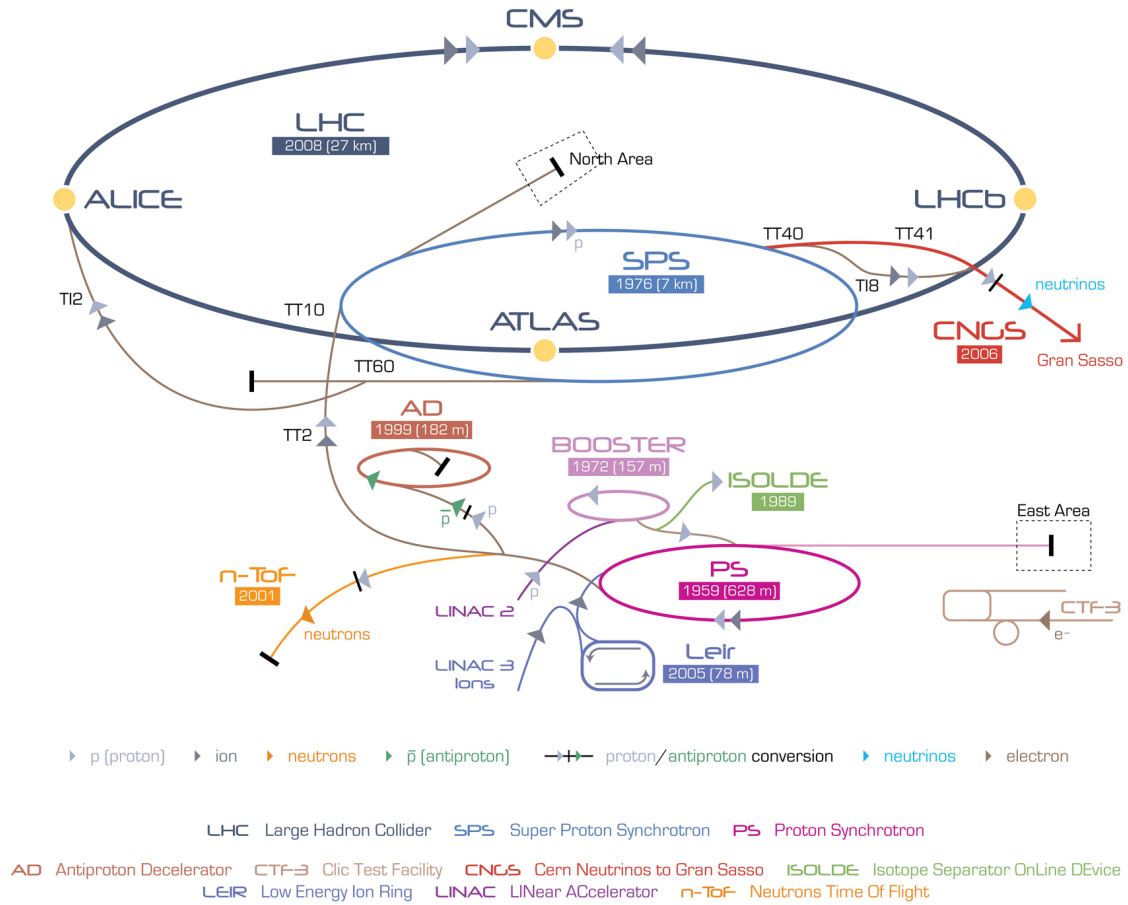
Figure 2.2: Schematic of the accelerator facility at CERN. The main accelerator is depicted along with the various accelerators used to feed it

an integrated luminosity of $21.79\,\mathrm{fb}^{-1} \pm 2.5\%(\mathrm{syst.}) \pm 0.5\%(\mathrm{stat.})$ [31].

2.1.2 Experiments at the LHC

In this section, a brief description of the detectors located along the LHC accelerator is given. The CMS detector is described in more details in the dedicated section 2.2. There are in total four massive detectors situated at LHC. Two multipurpose detectors CMS and ATLAS (A Toroidal LHC ApparatuS) were designed to probe high energy scale physics for a wide variety of physics measurements and searches. ALICE (A Large Ion Collider Experiment) is dedicated to the study
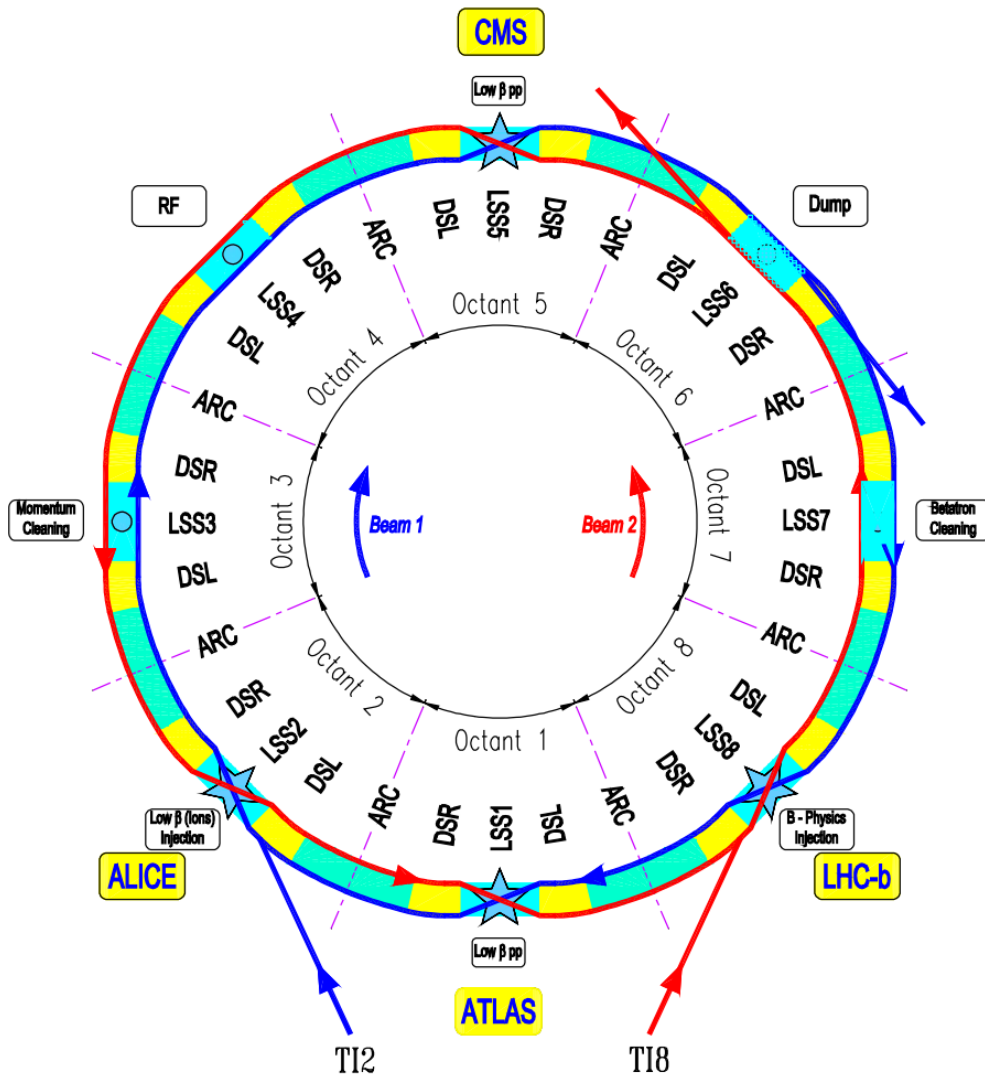
Figure 2.3: Schematic of the LHC displaying the path taken by the two beams. The eight different sectors, the four interaction points (IP) as well as the injection and dump pipes.

of heavy-ion collisions, probing the quark-gluon plasma medium rising from those high temperature and energy density conditions. LHCb (LHC beauty) on its side, is devoted to the study of bottom-quark physics and CP violation.

2.1.2.1 ALICE

ALICE [32] has been developed in the continuity of experiments conducted at SPS (CERN, CH) and RHIC (Brookhaven, US) in the 1980's dedicated to ion-ion collisions. The LHC is able to accelerate and collide lead-ion beams at a centre-of-mass energy per nucleon of 2.76 TeV. It pushes further the limit reached by previous experiments by one order of magnitude, allowing to study QCD processes in extreme energy density and temperature conditions.

ALICE is a massive detector ($16 \times 16 \times 26\,\mathrm{m}^3$) composed of 18 sub-detectors. Its main feature is its ability to track precisely and identify particles in a very busy environment. This is ensured by the presence of three-layers tracker system and a large variety of sub-detectors, each of them being dedicated to the identification of a specific particle type.

2.1.2.2 LHCb

LHCb [33] is dedicated to heavy flavour physics, aiming to look for new physics processes in CP violation, and rare decays of heavy flavour *charm* and *bottom* quarks. It benefits from the large production cross-section at LHC for these processes. The heavy flavour hadrons are predominantly produced in the forward cone region around the beam axis, rather than in the beam transverse plane.

LHCb is a single-arm spectrometer, covering the forward region from $\sim$ 15 mrad to 300 mrad. It relies on a high resolution tracker to look for displaced tracks originating from b-quarks and on a high momentum and transverse mass resolution to accurately identify and measure b decay products of interest.

2.1.2.3 ATLAS

ATLAS [34] is one of the two general-purpose detectors (along with CMS) present at LHC. It is the largest detector located at CERN ($22 \times 22 \times 46\,\mathrm{m}^3$). Like CMS, it investigates a large range of physics. Both experiments confirmed the observation of a particle with properties consistent with the Standard Model

Higgs boson, and continue searches for physics beyond the SM.

ATLAS detection principle is similar to what is done within CMS. They have the same sub-structure, using the same concentric sub-detector elements. From the inside to the outside of the detector:

- Vertex detector and inner tracker

- Electromagnetic calorimeter (ECAL)

- Hadronic calorimeter (HCAL)

- Muon system

Both CMS and ATLAS were developed independently and simultaneously, aiming for the same goal. The main differences between two detectors are on the technical solutions (design, materials) chosen by the two collaborations. For example, on the design side, the magnetic system for ATLAS is made of a central solenoid (2 T) in addition to 8 barrel and 2 end-cap toroids, whereas CMS uses a single solenoid (4 T) coupled with a return iron yoke.

2.2   The Compact Muon Solenoid

The Compact Muon Solenoid is a general purpose detector and aims at a wide range of particle physics analysis. The detector has a cylindrical shape being 14.6 m long and with a diameter of 12.6 m. Its name comes from its relative small size and massive weight (14,000 tons), its ability to detect and measure muons, and its feature being a large solenoidal magnet.

CMS has been developed in the early 1990's following several requirements. The benchmark used for the detector design was the detection of the Standard Model Higgs boson. The preferred processes for the search of the Higgs boson are the $H \to \gamma\gamma$, $H \to WW$ and $H \to ZZ$ channels. The ability to detect the final-state particles for those processes relies on several criteria:

- a precise resolution and identification for muons with an energy of up to 1 TeV

- a electromagnetic calorimeter with low response time and high resolution

- a tracker with good charged particle resolution and reconstruction efficiency

CMS can be decomposed into four major components placed concentrically around the interaction point (IP) where the collisions take place. As shown of figure 2.4, from the inside to the outside of the detector can be found the inner tracking system, followed by the electromagnetic and hadronic calorimeters. Those sub-detectors are contained inside the magnetic coil. Outside the coil, the muon system can be found. CMS can be decomposed into three geometrical parts, a central cylindrical shaped part (barrel) and two discs located on each side of the barrel (end-caps).
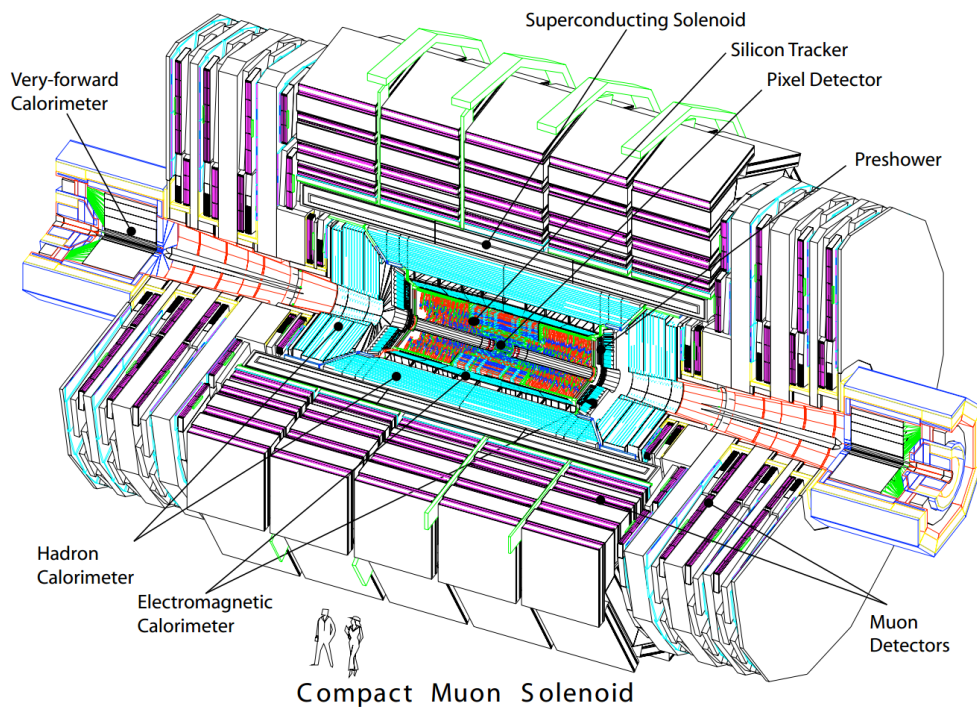
Figure 2.4: Open view of the CMS detector [3].

23

2.2.1 Geometry

Before presenting and discussing the different parts of the CMS detector, the coordinate conventions chosen by the collaboration are presented. The CMS detector has a cylindrical shape, aligned on the LHC beam axis. The origin of the CMS coordinate system is on the interaction point, at the centre of the detector. The x-axis lies in the horizontal plane, pointing at the centre of the LHC. The y-axis lies in the vertical plane, pointing to the top of the detector. The z-axis is the cylinder axis, its direction going counter-clockwise when looking at the LHC from above. Given the detector shape and the collisions' spherical nature, cylindrical and spherical coordinates are typically used. The azimuthal angle ϕ is defined in the (x, y) plane, the polar angle θ and the radial transverse distance ρ are defined with respect to the z-axis.

In hadron colliders, given the relativistic nature of the collisions, the polar angle θ is usually replaced by the pseudo-rapidity

$$\eta = -\log \tan \frac{\theta}{2} \tag{2.5}$$

Differences in pseudo-rapidity $\Delta\eta$ are Lorentz invariant under boost.

2.2.2 Inner Tracking System

The inner tracking system [4,5] is situated at the centre of CMS, and therefore is the sub-detector the closest from the interaction point. It has a cylindrical shape, being 5.8 m long and 2.4 m wide.

At the very centre of the tracking system can be found the pixel detector, covering a cylindrical volume of radius 20 cm around the beam axis and a length of 53 cm. The pixel detector is made of three layers in the barrel region, and two discs in each of the end-cap regions, as displayed on figure 2.5. The size of a pixel is $\simeq 100 \times 100\,\mu\text{m}$ featuring in total 66 million pixels. This very high granularity ensures the average occupancy to be of the order of 10^{-4} per LHC crossing.
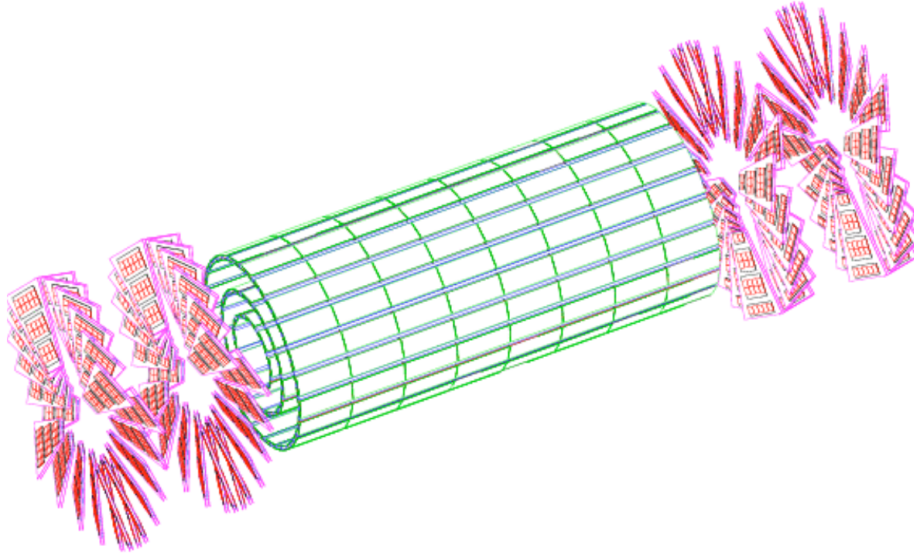
Figure 2.5: Schematic of the CMS pixel detector [4].

The silicon strip tracker envelops the pixel detector, extending the total volume of the inner tracker to a cylinder with a radius of 110 cm and a length of 5.4 m. It comprises in total 10 million read-out channels, and covering about 200 m². As shown on figure 2.6, the silicon strip detector, located around the pixel detector, can be decomposed into different parts depending on the region they cover. In the barrel, the tracker is divided into two parts:

- Tracker Inner Barrel (TIB) featuring 4 layers

- Tracker Outer Barrel (TOB) featuring 6 layers

The end-cap region is also divided into two parts:

- Tracker End Cap (TEC) made of nine rings centred on the beam axis

- Tracker Inner Disk (TID) made of three rings centred on the beam axis

Figure 2.6 shows that some layers are doubled and slightly shifted, ensuring a simultaneous measurement in both the (r, θ) and (r, z) coordinates, therefore

improving the spatial resolution. The CMS inner tracker system has been developed for high precision measurements on primary vertex position (see figure 2.7), and high charged particles tracks reconstruction efficiency (see figure 2.8). Such a feature is required for the determination of secondary vertices used in the identification of *bottom*-flavoured jets, see section 4.2.3.3. Track reconstruction is also a vital component of particle flow algorithms, see section 4.1.1.
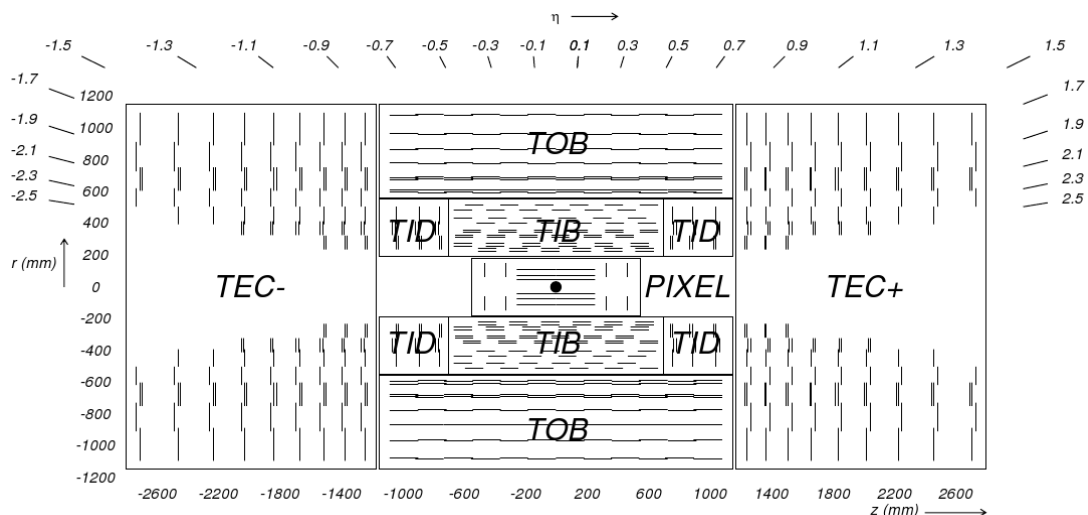
Figure 2.6: Longitudinal view of the CMS inner tracker. The pixel detector is shown in the central section, around the interaction point. Enveloping the pixel detector, the silicon strip detector is displayed [5].

2.2.3   Electro-magnetic Calorimeter

The CMS electromagnetic calorimeter (ECAL) [6, 35, 36] is a homogeneous, hermetic detector. It is made of lead tungstate ($PbWO_4$) scintillating crystals. The hermeticity and high granularity of the detector are ensured by the low radiation length ($X_0 = 0.89\,\text{cm}$) and low Moliere radius ($2.2\,\text{cm}$) of the chosen construction material, allowing the detector to completely stop high energy electrons and photons while remaining relatively compact. In addition, in lead tungstate crystals, 80% of the scintillating light is emitted within $25\,\text{ns}$, leading to a fast detector response time.
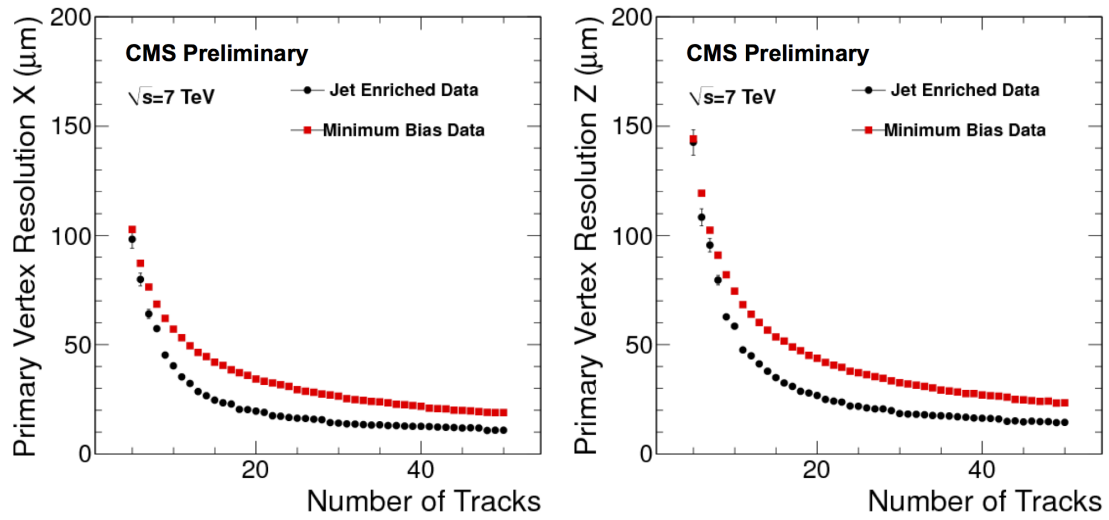
Figure 2.7: Primary vertex resolution in the transverse plane (left) and along the beam axis (right) as a function of the number of tracks associated to the vertex [4].
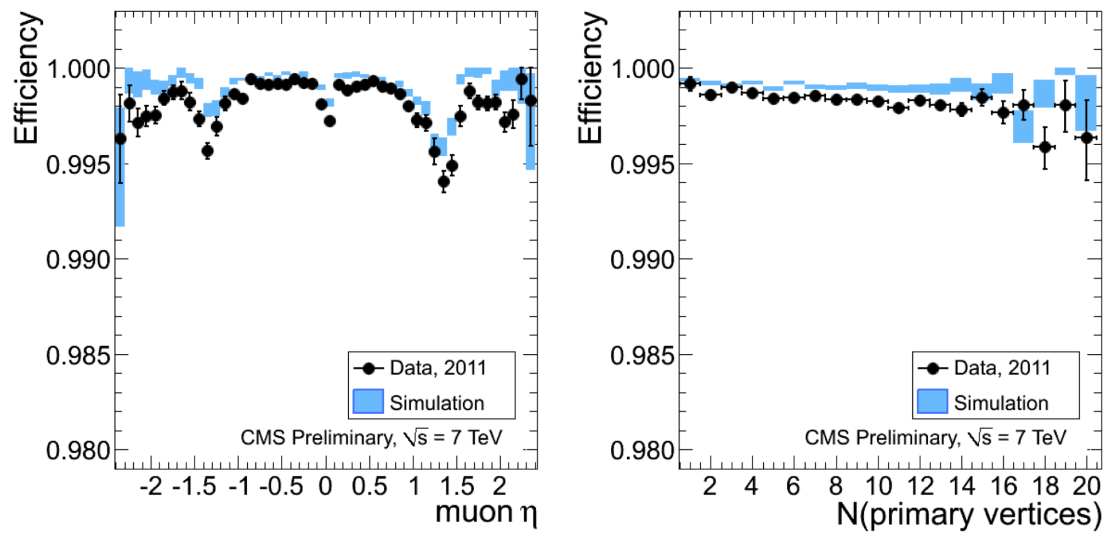
Figure 2.8: Muon reconstruction efficiency in the inner tracker as a function of pseudo-rapidity (left) and number of primary vertices (right) [4].

Scintillating light is converted to electrical signal using photodetectors. In

the barrel region ($0 < |\eta| < 1.479$), two avalanche photo diodes (APD) are used per crystal, while in the endcaps ($1.479 < |\eta| < 3.0$) vacuum photo-triodes (VPT) are used. VPT are more radiation resistant than the APD. As APD and scintillating crystals have a temperature dependent gain, ECAL operates at a temperature regulated within ± 0.05 °C.

Figure 2.9 displays the layout of the CMS electromagnetic calorimeter. In the barrel, scintillating crystals are assembled in 36 super-modules, each one made of 1700 crystals. The crystal have a frontal area of about 2.2×2.2 cm for a length of 23 cm $= 25.8 X_0$. In the endcaps, crystals are grouped in super-crystals. Each crystal has a frontal area of 2.68×2.68 cm for a length of 22 cm $= 24.7 X_0$. The super-crystals are assembled into four half-disks *'dees'*, each comprising 138 5×5 super-crystals and 18 special shaped super-crystals along the inner and outer radii. A pre-shower detector is also located in the barrel in front of the ECAL to allow a better π^0/γ separation. It consists of two planes of silicon detectors, separated by a layer of lead absorber. Figure 2.10 shows the ability of ECAL to reconstruct the Z boson mass with a resolution smaller than 2% in the barrel and smaller than 3% in the endcaps.

2.2.4   Hadronic Calorimeter

CMS hadronic calorimeter (HCAL) [7, 37], surrounding ECAL, is a plastic scintillator/brass absorber sampling calorimeter covering the pseudo-rapidity range $0 < |\eta| < 3.0$. It is located between the ECAL and the superconducting solenoid. However, this volume is not sufficient to ensure the containment of hadronic showers. Therefore an additional layer of scintillators is located on the outside of the coil.

Scintillating light is propagated through wavelength shifting fibres, enclosed in the scintillators, towards multi-channel hybrid photodiodes ensuring the photodetection. HCAL is divided into four parts: the hadron barrel (HB), the hadron endcaps (HE), the hadron outer (HO) and the hardon forward (HF) calorimeters.
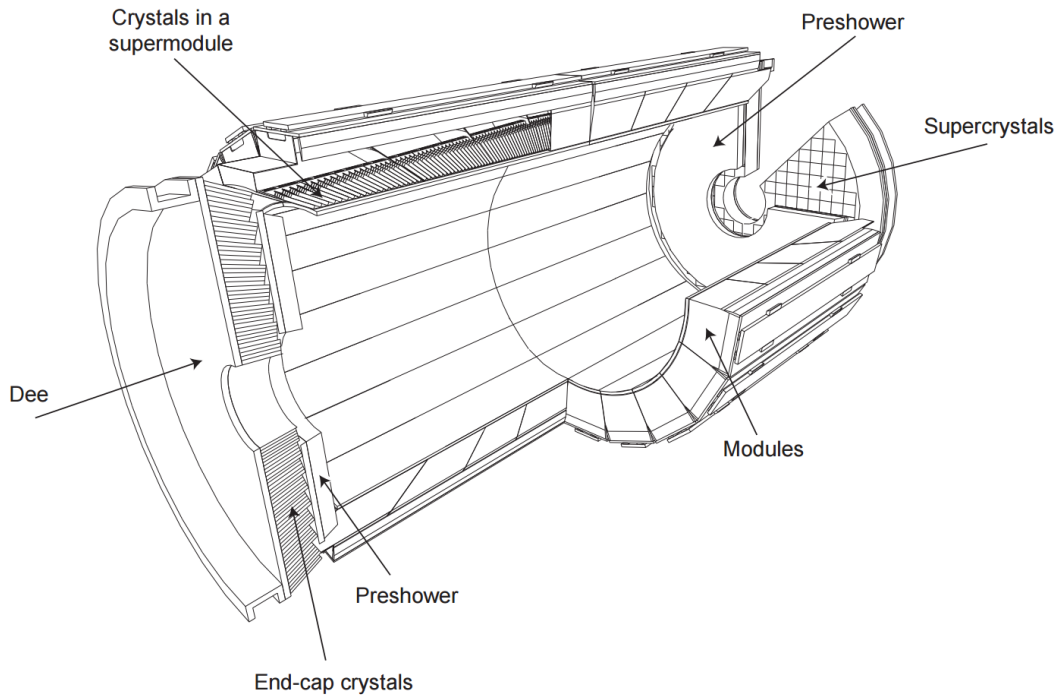
Figure 2.9: CMS ECAL layout showing the barrel super-modules, the two end-cap *'dees'* and the two pre-shower detectors [6].
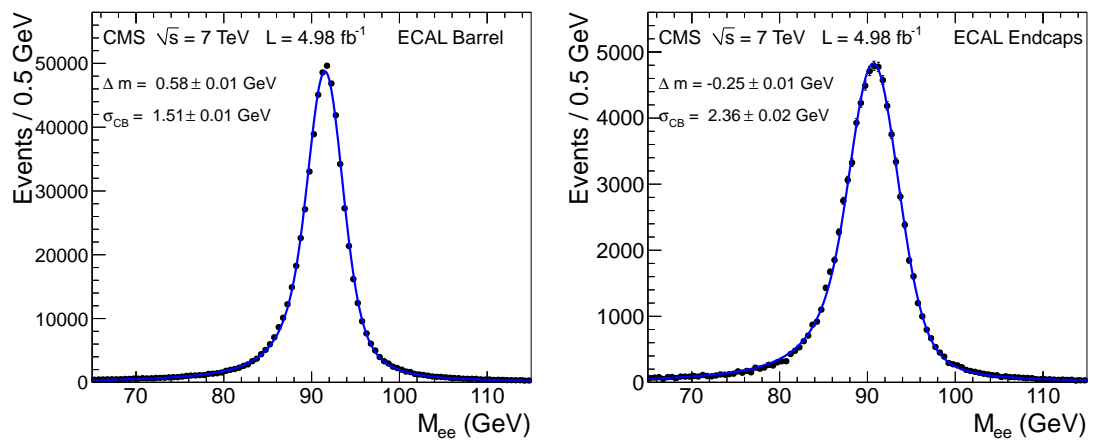
Figure 2.10: Di-electron invariant mass for $Z \to ee$ events in the barrel ECAL (left) and endcap ECAL (right) [6].

The hadron barrel (HB) consists of 2304 towers covering the pseudo rapidity range $0 < |\eta| < 1.4$. Each tower has a segmentation $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$. HO corresponds to the extra calorimeter layers outside the CMS coil. It covers the pseudo-rapidity region $0 < |\eta| < 1.26$ and allows to increase the effective thickness of HCAL to about 10 interaction lengths, thus ensuring a better energy resolution but also prevents hadronic showering in the muon system by absorbing the shower tails. The HE covers the pseudo rapidity range $1.3 < |\eta| < 3.0$. Tower size in HE is the same as in HB for $|\eta| < 1.74$. Beyond this region, tower size regularly increases to a maximal size of $\Delta\eta \times \Delta\phi = 0.350 \times 0.174$. Finally, the HF covers the pseudo rapidity range $3.0 < |\eta| < 5.0$ and is located at $11.2\,\mathrm{m}$ from the interaction point, on each side of the detector. It is a steel/quartz fibre calorimeter. Figure 2.11 shows the layout of the calorimeter towers for HB, HE and HO. Towers readouts are segmented, represented by different colours on figure 2.11, providing a longitudinal segmentation of the HCAL readout system.

2.2.5 Magnet System

In order to accurately measure the momentum of charged particles, and more particularly high energy muons, a high magnetic field is required. It is delivered by a $12.9\,\mathrm{m}$ long and $5.9\,\mathrm{m}$ wide superconducting solenoid enveloping the tracking system and both calorimeters. The CMS magnet has been designed to deliver a $4\,\mathrm{T}$ field nominally, but after ageing studies, the value has been lowered to $3.8\,\mathrm{T}$ [8, 38]. Figure 2.12 shows the intensity and lines of the magnetic field within CMS.

The solenoid is coupled with a steel yoke displayed on figure 2.4 as the white layers between the purple muon system's active layers. It serves several purposes [8]. First, it improves the magnetic field homogeneity in the tracker volume by returning the solenoid's magnetic flux. It also acts as a mechanical support structure for the detector. Finally, it acts as an absorber, allowing only muons and neutrinos to go through the muon system.
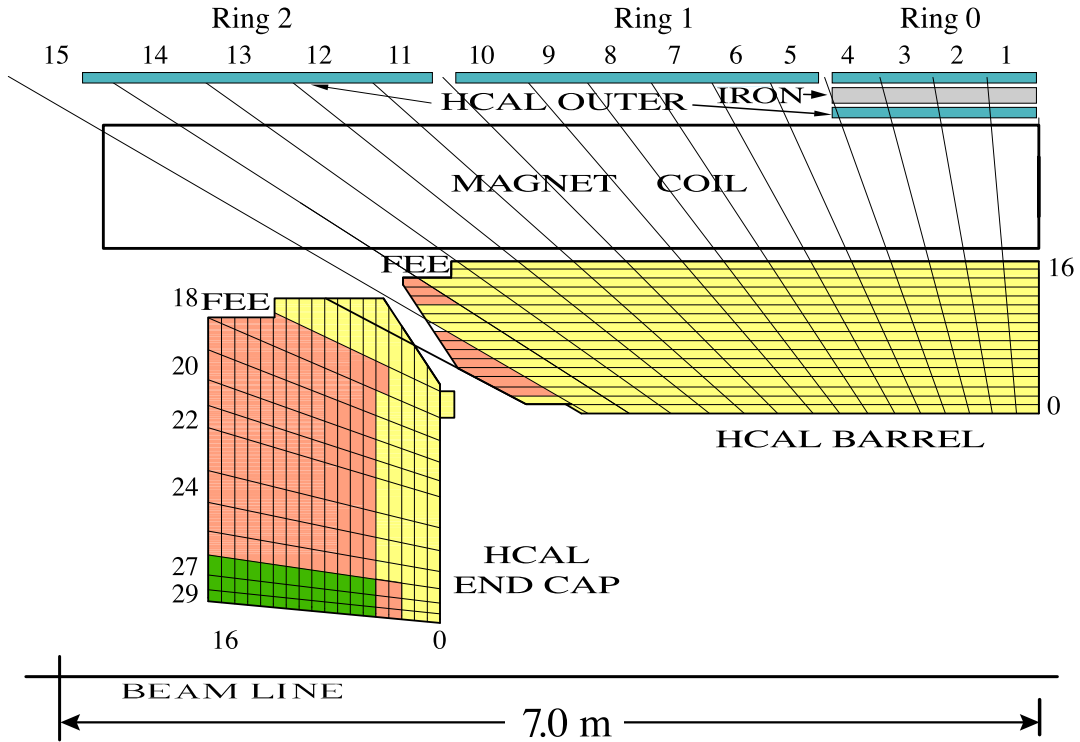
Figure 2.11: (r, z) view of the HCAL towers layout for HB, HE and HO. HF is not displayed on this figure. Numbers on top and left represent the segmentation in η. Numbers on the right refer to scintillator layers. Signals coming from tower segments with the same colour are added optically [7].

2.2.6 Muon System

The muon system [9,38,39] aims at identifying muons and measuring their momentum. It is also used as a crucial element for the triggering of events containing muons, therefore it requires a fast response time. The muon system is combination of three different types of gaseous detectors integrated in the iron yoke of the magnet system. In the barrel region, it consists of five wheels which are segmented into 12 sectors in ϕ. Each wheel consist of four muon station layers. In each of the end-caps, the muon system is composed of four discs, with the trapezoidal shaped muon detectors arranged in rings. The first disk consists of three rings, while the remaining disks are composed of two rings. The layout

31
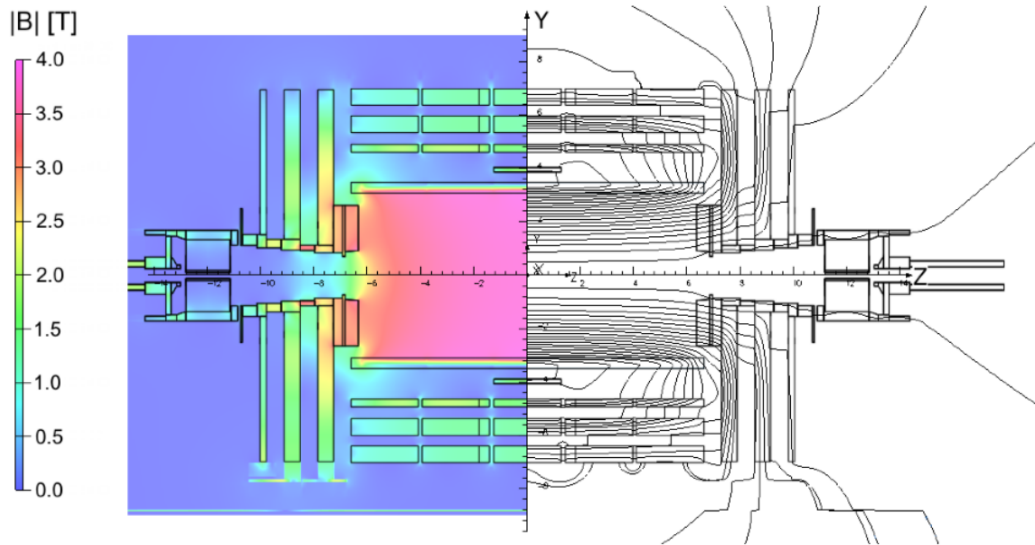
Figure 2.12: Expected magnetic field intensity (left) and field lines (right) displayed for a longitudinal section of the CMS detector for a central magnetic flux density of 3.8 T. Each field line represent an magnetic flux increment of 6 Wb [8].

of the muon system can been seen on figure 2.13. The choice on the technical solution mostly depends on the radiation environment of the considered region.

In the barrel region ($|\eta| < 1.4$) where the muon rate and the magnetic field are low, the muon system is composed of drift tube (DT) chambers. DT are assembled in wheels with a size varying between $5966 \times 290 \times 2536$ mm and $1990 \times 290 \times 2536$ mm depending on their radial distance to the interaction point. In the end-cap region, up to $|\eta| = 2.4$, where the muon rate and the magnetic field are much higher, cathode strip chambers (CSC) are used. They are 25 cm and up to 2 m long. In addition to DT and CSC, the CMS muon system features resistive plate chambers (RPC) up to $|\eta| = 1.6$. RPCs present a fast response time in addition to a good time resolution making them ideal for a high rate trigger system.

From figure 2.14 the muon system presents a high resolution, being able to separate resonances in the di-muon spectrum. The resolution of ~ 100 MeV in the spectrum region around 10 GeV allows to distinguish the three $\Upsilon(nS)$ peaks. The
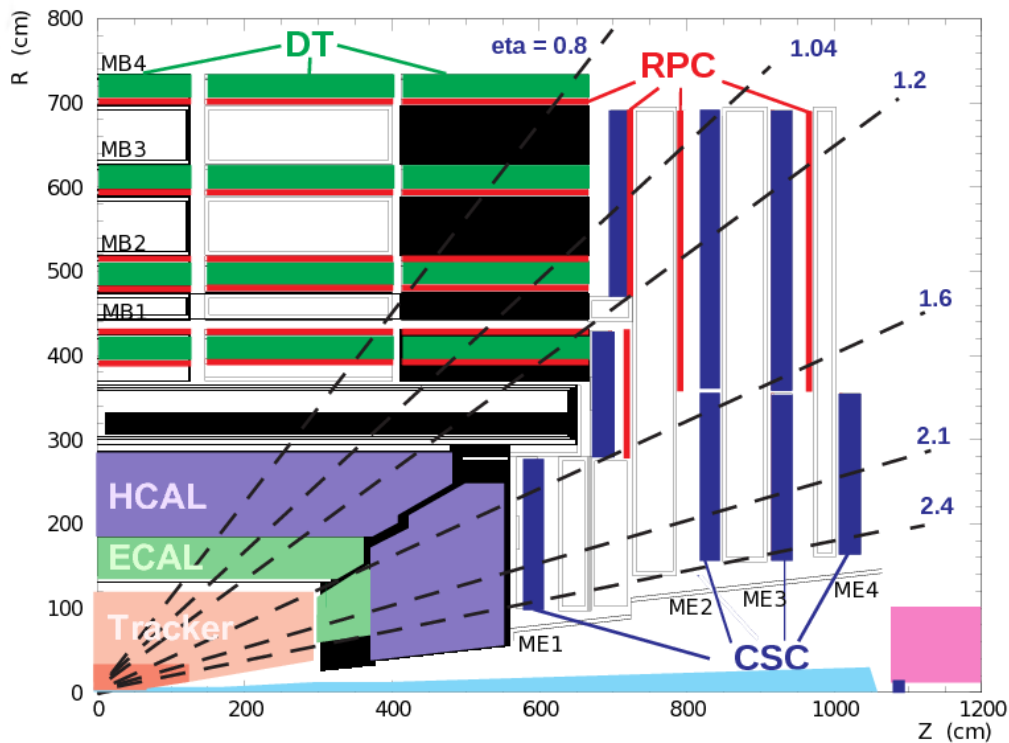
Figure 2.13: Layout of the muon system in CMS [9].

resolution can be further improved to $70\,\text{MeV}$ in this region if both muons have a pseudo-rapidity $|\eta| < 1.0$ [9].

2.2.7 Trigger System

During the 2012 data taking period, protons bunches are time-separated by $50\,\text{ns}$ corresponding to a rate of $20\,\text{MHz}$. Computing and storage limitations, as well as physics interests, imply to filter this massive quantity of collisions information. Indeed, technical restrictions allow a storage of about $300\,\text{Hz}$. Moreover, within the millions of collisions taking place every second, most of them represent low energy interactions and have no interests with respect to the physics programme at LHC.
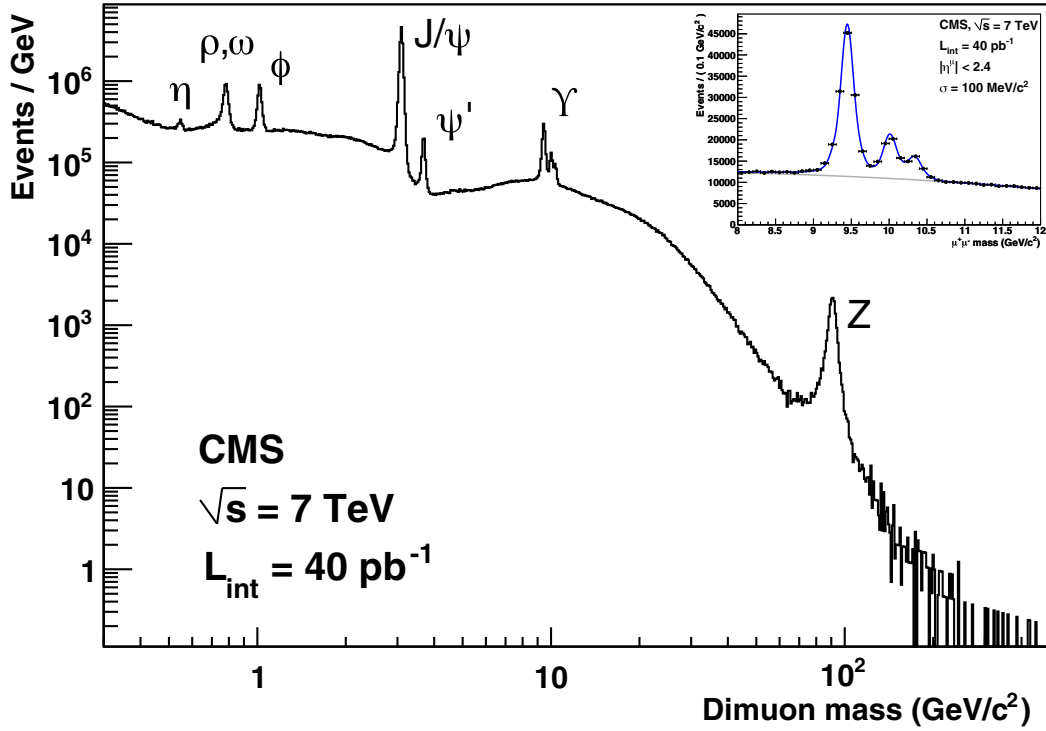
Figure 2.14: Di-muon invariant mass spectrum. The inset emphasises the CMS muon system ability to distinguish the three $\Upsilon(nS)$ due to its 100 MeV resolution for this energy range [9].

The role of the trigger system is to perform a fast analysis on every event recorded by the detector and decide if the event will be stored for further analysis. In CMS, the trigger system works at two levels. First, the level-1 (L1) hardware-based trigger system quickly identifies the physics objects of interest in the event. Then, the event is eventually passed to the software-based high-level trigger (HLT), which performs a more advanced reconstruction and analysis and determines if the event is sent to the mass-storage. In the L1 trigger system the event rate is reduced to $\sim 100\,\text{kHz}$. The HLT is able to reduce the rate down to $300\,\text{Hz}$.

2.2.7.1 Level-1 Trigger System

The level-1 trigger system is organised around three detector-based subsystems: the L1 muon trigger, the L1 calorimeter trigger and the L1 global trigger. The L1 muon trigger is based on the different sub-detectors used in the muon system: the drift tubes, the cathode strip chambers and the resistive plate chambers. The L1 calorimeter trigger uses information from both ECAL and HCAL and provides triggering information for electrons, photons and jets. The L1 global trigger combines information from both L1 muon and L1 calorimeter trigger systems.

Triggering in the level-1 trigger is fast as it only uses local detector information. No correlations are considered between the sub-detectors and no time-consuming reconstructions in the inner tracker are used. Figure 2.15 shows a diagram summarising the structure of the L1 trigger system.
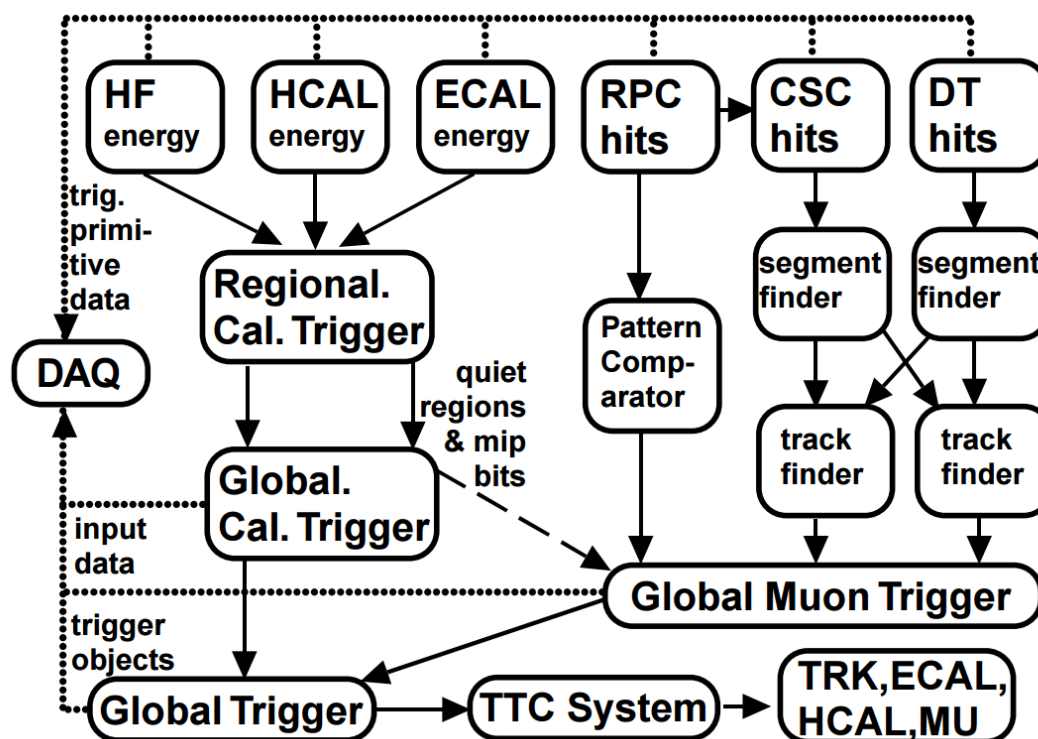
Figure 2.15: Diagram presenting the L1 trigger working principle [10].

2.2.7.2 High-level Trigger System

Once the event rate is reduced to $100\,\mathrm{kHz}$ by the L1 trigger, it becomes possible to transfer the events kept by the L1 trigger for further analysis by the HLT system. A given trigger path consists in a succession of trigger sequences. Each sequence has a specific role in performing a partial analysis of the event based on the information received from all the relevant sub-detectors. The event is selected by the trigger path if it satisfies every trigger sequence.

During data taking, a given list of trigger paths is defined and arranged in a trigger menu, specifying which software elements are used for the event reconstruction. The HLT performs an optimised version of the full CMS reconstruction process as depicted on figure 2.16. Readout Units (RU) store the event fragments selected by the L1 trigger which are subsequently reconstructed by Builder Units (BU) into complete events. The full event content is then passed to one of the HLT Filter Units (FU) responsible for reconstructing and analysing the event. Events satisfying the HLT filtering are forwarded to the Storage Manager (SM). This process is relatively time-consuming compared to the output rate of the L1 trigger system. Consequently, this process is parallelised with the help of about 1000 FUs. The HLT is then able to reduce further the event rate to $300\,\mathrm{Hz}$. A more complete description of the specific trigger objects used by this analysis can be found in section 5.1.
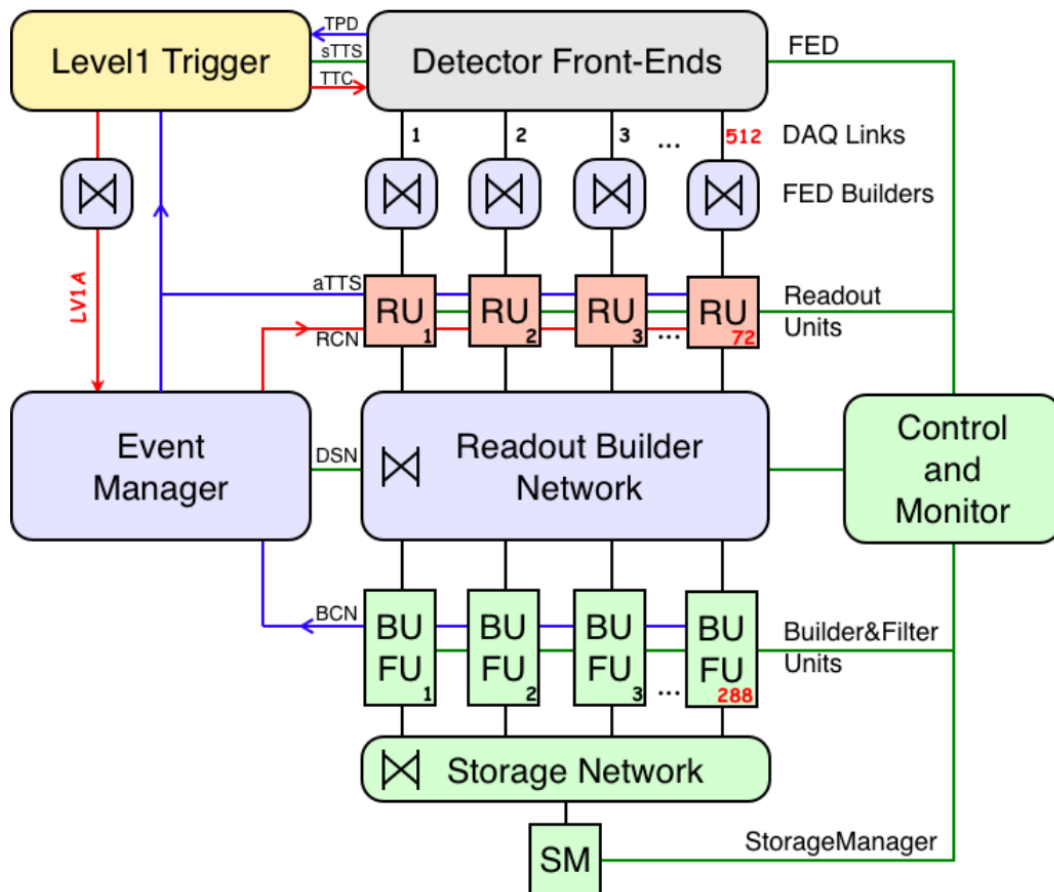
Figure 2.16: Diagram presenting the CMS Data Acquisition System working principle. It represents one of the eight slices of the event builder [11].

Part II

Event simulation, reconstruction and selection

Chapter 3

Event Simulation

It is essential to construct simulated events (Monte Carlo, MC) as precisely as possible in order to test the analysis method before applying it to data recorded by the detector. In this chapter, the different steps leading to fully simulated events are detailed, as depicted in figure 3.1.

The event simulation can be decomposed in several steps. The two partons (quarks and gluons), coming from the colliding protons are described with the parton density functions (PDF) derived from recorded data. At production stage, CMS makes use of the CTEQ6.L libraries to simulate parton momentum transfer and parton-parton interaction probability. Next, the matrix element simulates the hard process, giving rise to the final state partons as a result of the interaction between the initial-state partons. This is typically ensured by a tree-level matrix element generator like MADGRAPH [40] or a higher order generator like POWHEG [41] or MC@NLO [42]. During this step, all tree-level Feynman diagrams are generated for a given process, for example particle decays or $2 \rightarrow n$ scattering processes, while considering user-specific requests in terms of initial and final state particles or kinematic constraints. Then, the fragmentation process takes place, simulated by the parton showering and hadronisation event generator. Coloured partons are assembled into jets made colour-neutral hadrons, which will in turn decay into lower energy particles. This step is typically ensured by PYTHIA [43] or HERWIG [44]. Then, generated particles are passed to GEANT4 [45]
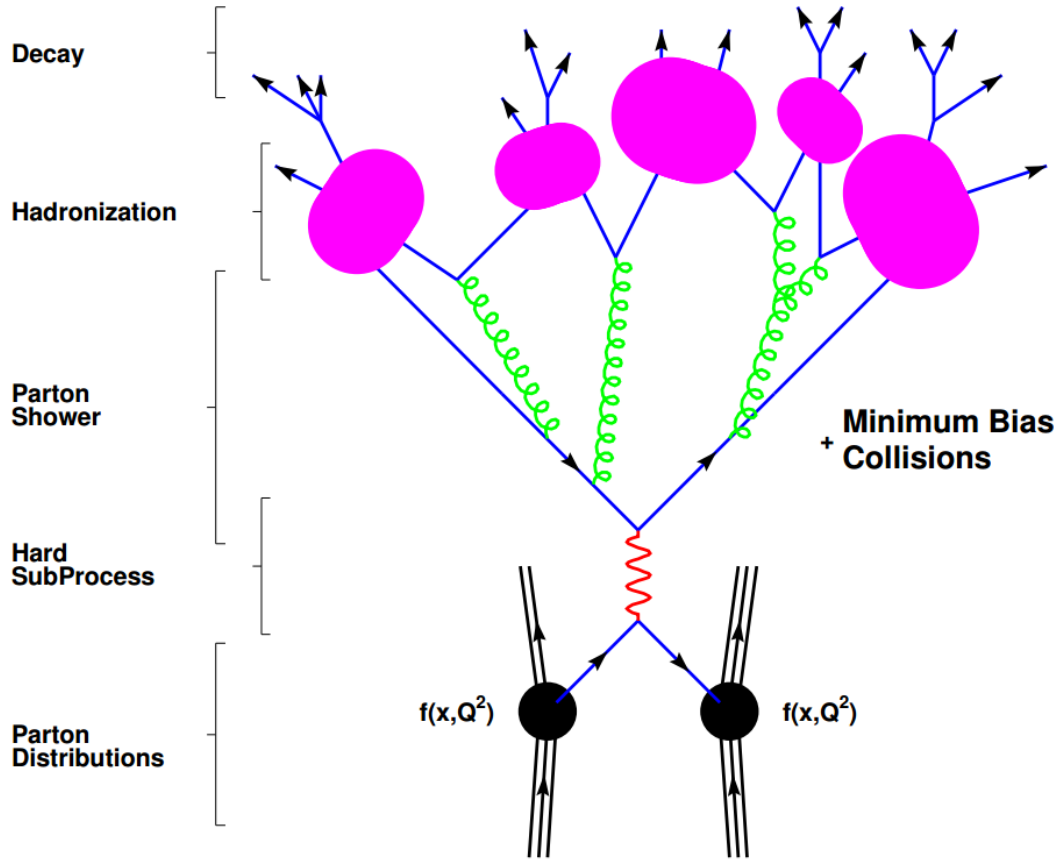
Figure 3.1: Schematic representing the different steps in the modelling of a proton-proton collision [2].

to simulate their interaction with the detector medium. GEANT4 is a simulation tool dedicated to the modelling of particle-matter interactions. It can handle a wide range of geometries, which are defined by the user.

The parton density functions [46, 47] describe how proton constituents acquire energy from it through a momentum transfer. PDF are derived experimentally and represent the probability density to find in the proton a parton carrying a momentum fraction x at a squared energy scale $Q^2 = \mu_f^2$ (see equation 1.2.3). Figure 3.2 shows an example of the Parton Distribution Function for two different

values of the factorisation scale.
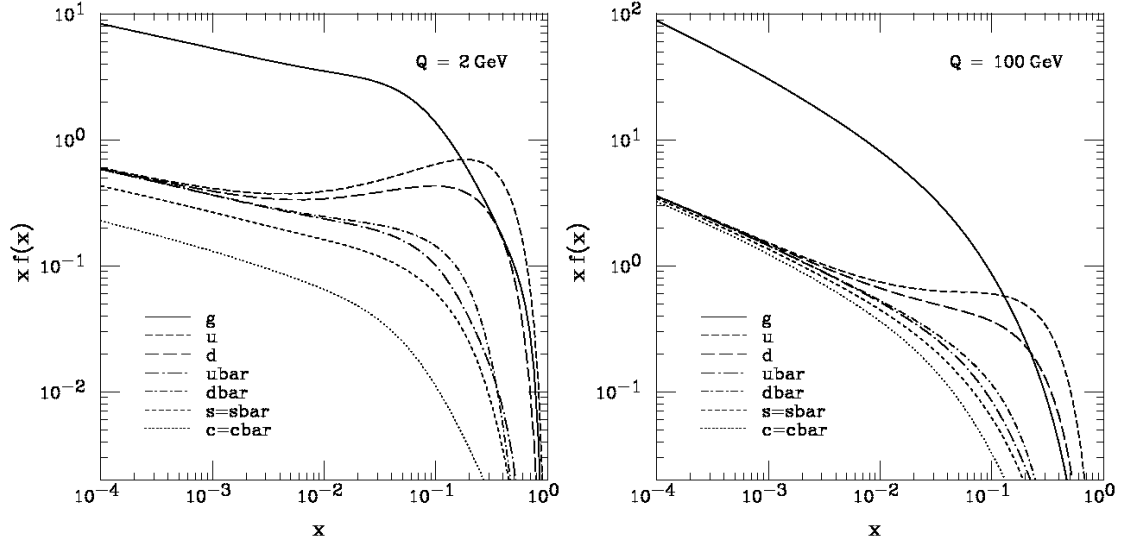
Figure 3.2: CTEQ6M Parton density functions at factorisation scale $Q = 2\,\text{GeV}$ (left) and $Q = 100\,\text{GeV}$ (right) [12].

In this analysis, the $t\bar{t}$ samples are generated with MADGRAPH event generator coupled with MADSPIN [48]. MADGRAPH is interfaced with PYTHIA with the help of the MLM algorithm [49]. Signal samples are generated for seven different top quark mass input values (166.5, 169.5, 171.5, 172.5, 173.5, 175.5, 178.5 GeV)

The W+jets and Z+jets background samples are also generated using MADGRAPH interfaced with PYTHIA, whereas the single-top samples are generated with POWHEG coupled with PYTHIA.

For the sake of systematic uncertainty estimation, $t\bar{t}$ samples are also generated with POWHEG interfaced with PYTHIA on one hand, and with POWHEG coupled with HERWIG, another parton shower generator. Those samples are used to estimate the systematic effect due to the choice of the matrix element and parton showering generators.

PYTHIA parton shower generator can be tuned differently to account for

41

several effects [50]. Accordingly, 5 different tunes are used:

- Tune$Z2*$: tune used for all the nominal samples

- TuneP11 and TuneP11noCR : tunes used to evaluate the systematic effect due to colour reconnection

- TuneP11TeV and TuneP11mpiHi : tunes used to evaluate the systematic effect due to underlying event activity

Table 3.1 lists the signal and main background processes associated to the simulated samples and their cross section. The multijet, W+jets and Z+jets are the dominant processes arising from proton-proton collisions at LHC. As this analysis relies on an extensive use of b-tagging, the contribution from multijet processes is shown to be negligible after the complete selection. Consequently, the samples associated to these processes are not used in this analysis. All the signal, background and systematic samples used in this analysis are listed in appendix A1.

sample	cross section (pb)
$t\bar{t}$	252.9 (NNLO)
single-top s-channel	5.2 (NLO)
single-top t-channel	84.7 (NLO)
single-top tW	22.4 (NNLO)
W+1jet	6662.8 (NNLO)
W+2jets	2159.2 (NNLO)
W+3jets	640.4 (NNLO)
W+4jets	264.0 (NNLO)
Z+jets	3503.7 (NNLO)
multijet, e/m enriched, $20\text{GeV} < \hat{p}_T < 30\text{GeV}$	29148.6 (NNLO)
multijet, e/m enriched, $30\text{GeV} < \hat{p}_T < 80\text{GeV}$	4615893.0 (NNLO)
multijet, e/m enriched, $80\text{GeV} < \hat{p}_T < 170\text{GeV}$	183294.9 (NNLO)
multijet, $b/c \rightarrow e$, $20\text{GeV} < \hat{p}_T < 30\text{GeV}$	167388.0 (NNLO)
multijet, $b/c \rightarrow e$, $30\text{GeV} < \hat{p}_T < 80\text{GeV}$	167040.0 (NNLO)
multijet, $b/c \rightarrow e$, $80\text{GeV} < \hat{p}_T < 170\text{GeV}$	12981.9 (NNLO)

Table 3.1: Table listing signal and background processes and their cross section taken from [51], [52] and [53]. Cross-sections have been evaluated at the highest order available.

Chapter 4

Event Reconstruction

The reconstruction of data recorded by CMS is a vital step to translate signals from the different components of the detector to physical objects that can be used in analysis. In this chapter, the Particle Flow algorithm is presented, and its application in reconstructing various physical objects is detailed.

4.1 Particle Flow Algorithm

CMS relies on the Particle Flow (PF) algorithm [54] to reconstruct and identify particles. This method uses the full potential of the detector by combining information from each of the sub-detectors, leading to a more precise estimation of particles energy, direction and type. For a given event, reconstruction is achieved in two main steps. First, "elements" such as tracker tracks, calorimeter clusters, muon tracks, are grouped into "blocks". Signals from the same block are assumed to originate from a single particle. As a second step, blocks are fed to an identification algorithm that will establish a listing of all the particles that have been detected.

4.1.1 Iterative Tracking

The silicon tracker from CMS plays a dominant role in the PF algorithm. It has much better resolution than the calorimeters for low transverse momentum

objects (up to a few GeV). It can provide with high precision the direction of charged particles coming from the primary vertex and its close surroundings.

Iterative tracking is used to ensure the highest possible tracker performance. This strategy starts by reconstructing the tracks that obey a very tight selection criteria. The tracker hits that are found to unambiguously belong to these tracks are removed from the tracker hit collection. The process is repeated several times on the remaining hits, while loosening the selection criteria between each iteration.

As the number of iterations increases, so does the tracking efficiency, as more and more hits are successfully reconstructed, and the fake rate remains low as less hits are present, reducing the number of possible tracks.

4.1.2   Calorimeter Clustering

The calorimeter clustering aims at different goals: [54]

- Detect and measure the energy and direction of neutral particles (neutral hadrons, photons)

- Provide information on charged particles complementing measurements from the tracker

- Separate neutral particles from charged particles when combining information from calorimeters and the tacker

- Reconstruct and identify electrons and associated Bremsstrahlung photons

- Help the charged hadron energy measurement in conjunction with the tracker, or if the tracker was not successful

The clustering is performed separately in each calorimeter sub-detectors: HCAL barrel, ECAL barrel, HCAL end-caps, ECAL end-caps, and the two first PS layers. No clustering is performed in HF.

The clustering algorithm can be decomposed into three steps. First, it looks for "cluster seeds", calorimeter cells where the energy deposit is a local maximum. Second, it constructs "topological clusters" by grouping cells that are direct cluster neighbours and with an energy larger than a given threshold. In ECAL this threshold is twice the standard deviation of the electronics noise (80 MeV in the barrel, 300 MeV in the end-caps), whereas in HCAL the threshold is set to 800 MeV. Finally, when all the *topological clusters* are formed, each *cluster seed* is associated to a *particle flow cluster*. For cells belonging to several *topological clusters* their energy is shared as function of the cell-cluster distance.

4.1.3   Link Algorithm

A particle going through the detector usually gives rise to several particle-flow elements charged particle tracks and/or particle flow clusters. The link algorithm, as its name indicates, connects those elements together to fully reconstruct each particle. It also needs to remove any double-counting from the different sub-detectors. The algorithm considers any possible pair of PF elements and returns "blocks" of linked elements. The quality of the link is quantified by the distance in the (η, ϕ) plane between the two members of a given element pair.

A link between a charge particle track and a calorimeter cluster is constructed as follows. The track is extrapolated in steps following the detector geometry. From its last hit in the tracker, to the two layers of the PS, then to inside the ECAL, and finally inside the HCAL. The track is successfully linked to a cluster if the extrapolation is within the cluster boundaries. Two calorimeter clusters can also be linked (PS and ECAL or ECAL and HCAL) if the cluster in the calorimeter with a higher granularity is within the envelope of the cluster in the less granular calorimeter.

4.2 Particle Reconstruction and Identification

This is the last step performed by the particle flow algorithm. It considers the "blocks" of PF elements obtained out of the link algorithm and returns a list of reconstructed and identified particles usable in physics analysis. In this section, the identification of the various objects used in this analysis is explained.

4.2.1 Muons

Muon reconstruction [39] is performed using track information from the silicon tracker (*tracker tracks*) and the muon system (*standalone muons tracks*). Two reconstruction methods used the two collections.

- **Global muons:** reconstructed from a standalone muon track extrapolated to the tracker and succesfully matched to a tracker track. This method is better suited for large transverse momentum muons.

- **Tracker muons:** reconstructed from a the tracker track extrapolated to the muon system and succesfully matched to a muon segment. If several tracker tracks are matched to the same segment, only the one leading to the best match is considered. This approach is more efficient for low transverse momentum, complementing the global muon approach.

It can also occur that the two approaches fail and that only a standalone muon track remains. Due to a high tracker efficiency, this happens only for 1% of collisions. The majority of standalone muons being cosmic-ray muons.

The three algorithms provide the muon candidates collection and their combination allows a high performance muon reconstruction. Different levels of identification efficiency and purity can be achieved, depending on the selection applied on the muon candidates.

- A **soft muon** is a muon from the tracker muon collection. The track must match a segment in the outermost station, and that segment must not form

a better match with another track.

- A **global muon** is a muon from the global muon collection. It has no further requirements

- A **tight muon** is a muon from both tracker and global collections. It must satisfies additional requirements:

  - $p_T > 3\,\mathrm{GeV}$
  - normalised χ^2 of the global track fit < 10
  - at least one hit in the muon chambers
  - matched to segments in at least two stations
  - its tracker track must have an impact parameter (radial distance between the track and the IP) $d_{xy} < 2\,\mathrm{mm}$ from the primary vertex

4.2.2 Electrons

Two complementary track seeding techniques are used in the electron reconstruction [55], based on two different sub-detectors: the tracker and the ECAL.

The tracker driven seeding is suitable for low momentum electrons or electrons within a jet. Candidate tracks are first selected using a boosted decision tree to reject fakes from light hadrons as much as possible. Then they are extrapolated from the tracker to the ECAL.

The calorimeter driven seeding is optimised for isolated electrons with a $p_T > 5$ GeV. It forms *super-clusters* out of PF clusters reconstructed in the ECAL. Super clusters of transverse energy $E_T > 4\,\mathrm{GeV}$ are matched to tracker tracks.

When extrapolating electron tracks, a dedicated modeling of the electron energy loss is used and trajectories are fitted using a *gaussian sum filter* (GSF). Tangents to the tracker tracks are also extrapolated to the ECAL to account for

any Bremsstrahlung photons emitted by electrons.

In addition to the reconstruction, a preselection is applied. For electrons only found with the tracker driven seeding, the preselection is based on a multivariate analysis [56]. For electrons found with the calorimeter driven technique, the selection criterion is based on the geometrical matching between the GSF track and the supercluster. Electrons failing this selection but passing the multivariate analysis are kept.

4.2.3 Jets

As seen in section 1.1.2, quarks and gluons originating from the hard scattering and any other gluon radiation hadronise to form colour-neutral objects, giving rise to hadronic showers. The particles within that cascade are clustered into a jet during reconstruction stage. In this analysis, this is performed using the anti-k_T algorithm.

4.2.3.1 The anti-k_T Algorithm

The anti-k_T algorithm [13] is a fast jet finding algorithm. Unlike algorithms used in the past like the *iterative cone* [57], it is infrared and collinear safe. It means the algorithm is stable if soft particles are added, and if the energy of a particle is split among two collinear particles. It considers as input the four-momentum of the reconstructed PF particles, distance parameters d_{ij} between particles i and j, and d_{iB} between particle i and the beam. They are defined as follows:

$$d_{ij} = min(k_{Ti}^{2p}, k_{Tj}^{2p})\frac{\Delta_{ij}^2}{R^2}, \tag{4.1}$$

$$d_{iB} = k_{Ti}^{2p} \tag{4.2}$$

where k_T represents the particle four-momenta, $\Delta_{ij} = \sqrt{\Delta\eta_{ij}^2 + \Delta\phi_{ij}^2}$ is the distance between particles i and j in the (η,ϕ) plane, R is a radius parameter and p a power parameter governing the relative power of energy versus geometrical

scales. For the k_T algorithm, $p = 1$. In the case of the anti-k_T algorithm $p = -1$. The algorithm identifies the smallest of the distances. If it is d_{ij}, it combines entities i and j, while if it is d_{iB} it removes entity i from the list and calls it a jet. Distances are then re-evaluated in case of a recombination ($E_T = E_{T,i} + E_{T,j}$, $\eta = \frac{\eta_i E_{T,i} + \eta_j E_{T,j}}{E_T}$, $\phi = \frac{\phi_i E_{T,i} + \phi_j E_{T,j}}{E_T}$), and the method is repeated until there are no entities left.

An event with a few separated hard particles and many soft particles will be reconstructed as follows. The distance between a hard (i) and a soft (j) particle $d_{ij} = min(\frac{1}{k_{Ti}^2}, \frac{1}{k_{Ti}^2})\frac{\Delta_{ij}}{R^2}$ only depends on the hard particle transverse momentum and its separation with the soft particle. It is much smaller in this case compared to the distance between two particles. Consequently, soft particles tend to cluster with hard particles rather than clustering with other soft particles. If another hard particle is present in the vicinity of the first hard particle, such as $R < \Delta_{ij} < 2R$, the algorithm finds two hard jets. If $\Delta_{ij} < R$ the algorithm finds a single jet. To summarise, the algorithm output is not impacted by the presence of soft particles but is sensitive to hard particles. This analysis uses a radius parameter of $R = 0.5$ which presents a good compromise between jets separation capability and amount of particles left non-clustered. Figure 4.1 shows an example of the reconstruction of several jets using different algorithm. Jets that have been clustered by the k_T and Cambridge/Aachen [58] algorithms present irregular shapes. The SiS cone-based [59] and the anti-k_T algorithm give rise to jets with a uniform profile. The anti-k_T algorithm is preferred to cone based algorithm as its execution is less time consuming.

4.2.3.2 Jet Energy Scale and Corrections

Jet energy scale represents the detector's ability to accurately reconstruct jet energies. Particles energy reconstruction is not perfect and a significant proportion is lost via different effects: dead or non-instrumented regions in the calorimeter, electronics noise. As a consequence, CMS has developed jet energy corrections [60] following a factorised approach as shown on figure 4.2. Each of the four
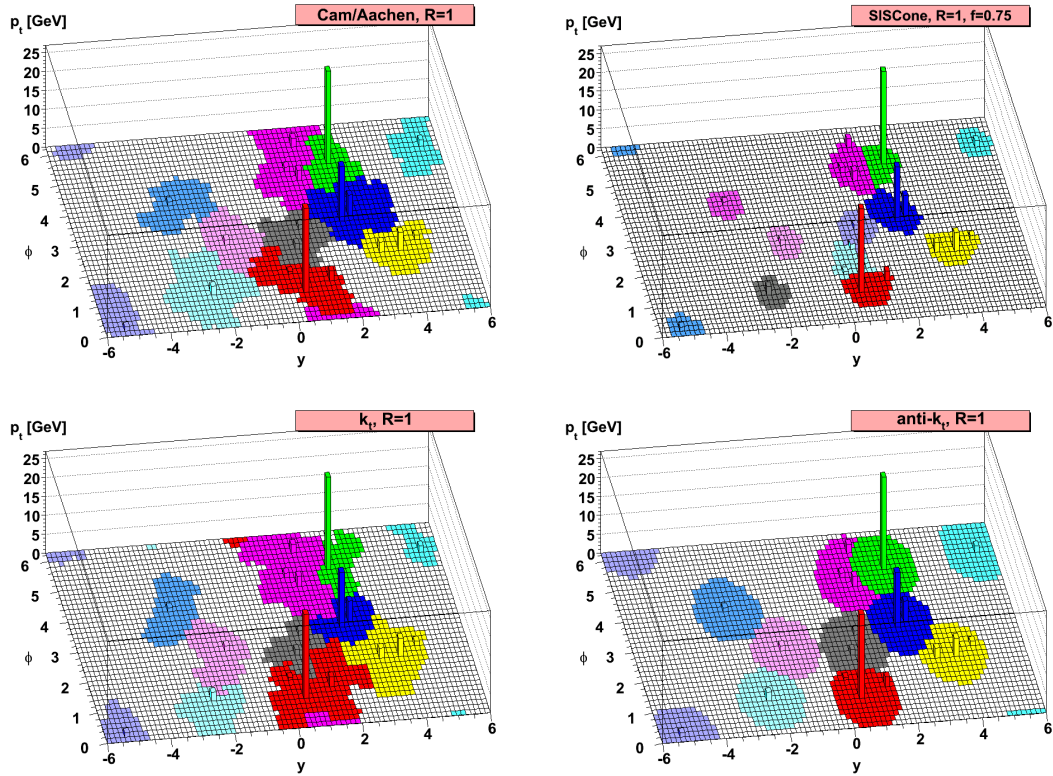
Figure 4.1: Comparison of jet shapes using different clustering algorithms [13]. On top, for Cambridge/Aachen (left), SiSCone (right). Below, for the k_T (left) and anti-k_T (right)

level corrects for a different effect, by scaling the jet four-momentum by a scaling factor which depends on the jet physical and geometrical properties. There are extra corrections that exist but they are not applied in this analysis. The corrections used are labelled as Level-1 (L1), Level-2 (L2) Relative, Level-3 (L3) Absolute and Level-2/Level-3 (L2L3) Residual.

L1 corrections remove in the jet energies the contributions from neutral hadrons coming from pile-up interactions. Next, the L2 and L3 corrections are applied which uniform the detector response with respect to jet η and p_T respectively. The residual corrections (L2L3 not shown on figure 4.3) are only applied to data recorded by the detector to account for any discrepancies that exist be-

Figure 4.2: Overview of the factorised approach for applying Jet Energy Corrections. The three mandatory first steps are applied in this analysis. They are responsible for rendering the jet energy response uniform across the phase-space covered by the detector [2]. Calorimeter jet technically refers to a jet that has been reconstructed using calorimeter but also tracker information.

tween recorded and simulated events. These corrections have first been evaluated at $\sqrt{s} = 7\,\text{TeV}$ for a luminosity of $36\,\text{pb}^{-1}$. They have been updated in the frame of the TOPLHC working group for $\sqrt{s} = 8\,\text{TeV}$ collisions with a luminosity of $19.6\,\text{pb}^{-1}$. The different uncertainties on the Jet Energy Corrections are shown on figure 4.3. Level-4, 5 and 6 are not applied in this analysis. They show advantages for specific analysis, for example if the jets need to be calibrated back to parton level.
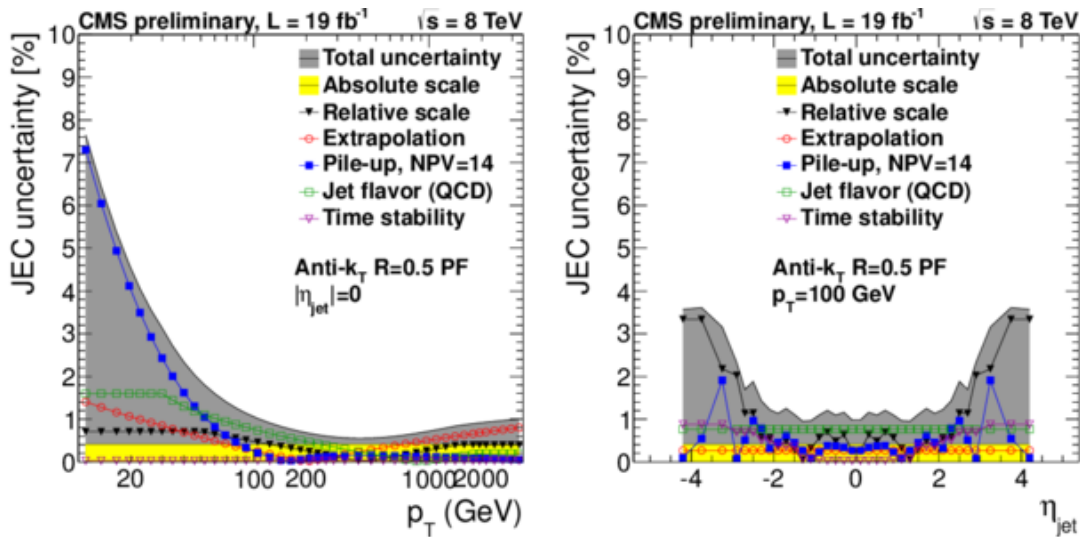
Figure 4.3: Jet Energy Corrections uncertainties for jets with pseudo-rapidity $\eta = 0$ as a function of their transverse momentum (left) and for jets with a transverse momentum $p_T = 100\text{GeV}$ as a function of their pseudo-rapidity (right) [14].

4.2.3.3   *b*-tagging Algorithm

   The proper identification of jets originating from a *b*-quark is crucial for this analysis, as they represent the direct decay products of the top quark pair elements. Identifying *b*-jets can therefore help in reducing the contributions from background events. *b*-jets present specific features compared to non-*b*-jets. The hadronisation of the *b*-quark gives rise to *B*-hadrons that have a large mass and a reasonably long lifetime. Several algorithms are used in CMS to identify *b*-jets [61]. Figure 4.4 shows the mis-identification probability versus the efficiency of the *b*-tagging algorithms used in CMS. The analysis uses the Combined Secondary Vertex (CSV) approach as it presents the best compromise between efficiency and purity, especially for mis-identification rates lesser than 1%.
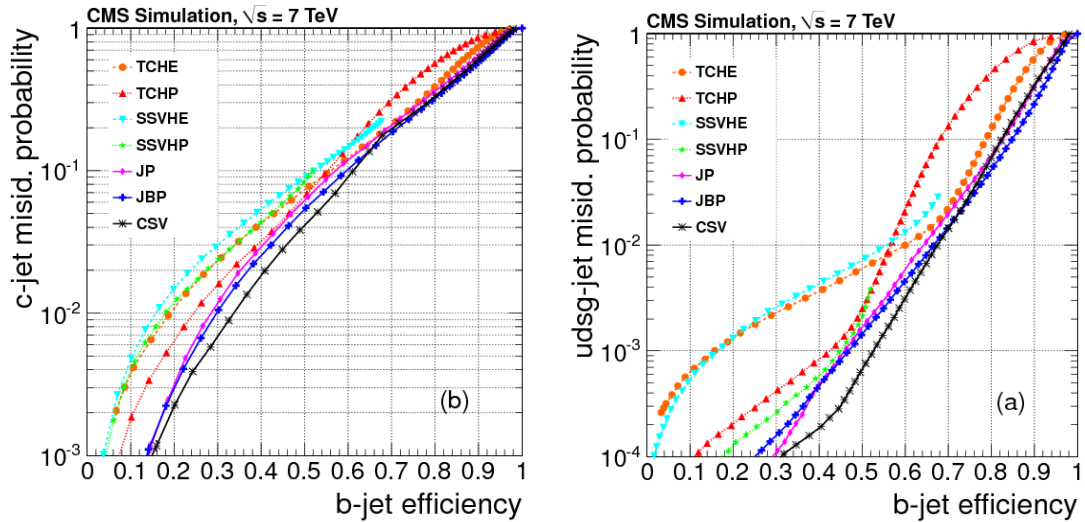
Figure 4.4: *b*-tagging efficiency versus charm (left) and light jet mis-identification rates for the different *b*-tagging algorithms used in CMS [15].

   The CSV algorithm relies on the presence of a secondary vertex in the event under consideration. Indeed, the *b*-hadron originating from the *b*-quark will travel a significant distance in the detector before decaying into lighter products. Experimentally this is observed as the presence of displaced tracks originating from

a vertex that is not the event's primary vertex.

The algorithm proceeds in distinct steps. First, the identification of the secondary vertex is made using charged-particle tracks from the jet under consideration. Only the tracks within a cone of $\Delta R = 0.3$ around the jet axis are considered for this purpose. If a matching vertex is found, the tracks are removed, and the process is repeated until no vertex candidates are found. Following the first step, jets are classified in three different categories.

- **RecoVertex:** a secondary vertex is successfully fitted and at least two tracks are matching.

- **PseudoVertex:** the fit of the secondary vertex fails, but at least two tracks are close to the jet (with an impact parameter significance $S = \frac{IP}{\sigma_{IP}}$ larger than 2).

- **NoVertex:** no vertices are found.

This categorisation along with many other variables described below are used by the algorithm. Some variables are not defined for some jet categories.

- flight distance projected on the ϕ plane (RecoVertex)

- jet-vertex flight direction angle (RecoVertex)
  This translates the energy carried by the b-hadron, and correlates the vertex position with respect to the jet

- vertex mass (RecoVertex, PseudoVertex)

- number of tracks coming from the vertex (RecoVertex, PseudoVertex)

- vertex energy ratio (RecoVertex, PseudoVertex)
  Energy ratio of charged tracks associated to the vertex with respect to all the tracks in the jet.

- η_{rel} of all tracks (RecoVertex, PseudoVertex)
  distribution of tracks pseudo-rapidity compared the jet pseudo-rapidity

- number of selected tracks in the jet

- impact parameter significance of the highest quality tracks.

Depending on the jet category, the variables are combined in a likelihood ratio that will discriminate between b-jets versus non-b-jets. Further discussion on the choice of the discriminant in this analysis is found in section 5.4.1.

Chapter 5

Event Selection

The single-electron decay channel of a top quark pair, as seen in section 1.2, implies the detection of two b-jets coming from the top quarks decay, in addition to two jets and a lepton-neutrino pair coming from the decay of the two W bosons. An experimental signature example can be seen on figure 5.1. There are many physical processes mimic the experimental signature of a top quark pair decaying to an electron and four jets. For example an event where jets are produced along with a W boson can be identified as a top quark pair decay. Similarly, an event with jets produced in association with a Z boson can be misidentified as a top quark pair decay signature if one of the two leptons from the Z boson decay is not reconstructed or does not satisfy the selection criteria associated to the lepton definition.

To ensure a high purity final sample, tight selection criteria are applied to data recorded by CMS. Events from the competing processes are also generated as described in chapter 3. The impact of selection on purity can then be evaluated and controlled. In this chapter, the details of the object selection criteria are first detailed. Then, the concept of reweighing is discussed, presenting the different weights used in the analysis to compensate simulation imperfections. Finally, the impact of the selection on the sample composition is scrutinised, emphasising the criteria specific to this analysis and that are applied on top of the reference event selection [62].
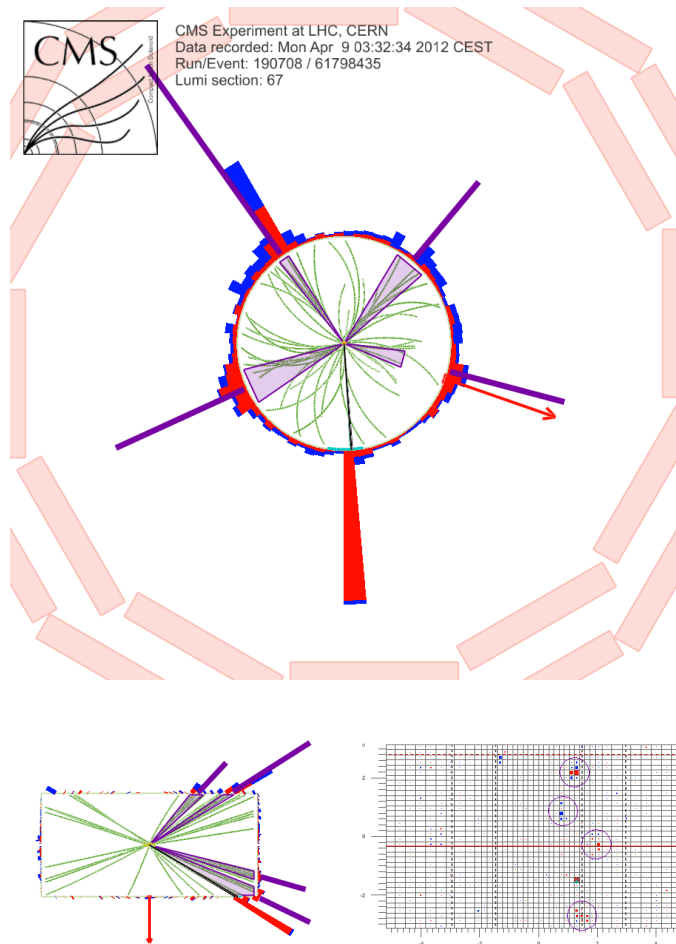
Figure 5.1: Event 61798435 from run 190708 recorded by CMS during 2012. It presents all the signatures of a $t\bar{t} \to e$+jets event. The upper figure is a (ρ, ϕ) view of the detector. The lower figures are a (ρ, z) view of the detector (left) and two dimensional projection of the signals recorded by the calorimeters in the (η, ϕ) plane. Red and blue *towers* depict the energy deposits measured by the ECAL and HCAL respectively. The clear blue curved segment associated to the largest deposit in ECAL implies the presence of an isolated electron with high energy. The purple segments and cones describe the four jets. The missing transverse energy is represent by the red arrow. Figure realised with the help of FireWorks CMS event display [16].

5.1 Trigger, Vertex and Event Cleaning

A first set of technical filters is applied on events recorded by the detector. Their purpose is to filter events slightly affected by technical issues during the data taking. Either because there was a high noise level in the hadronic calorimeter (*HBHO noise filter*) or because the events recorded by the detector do not actually correspond to beam collision data but are rather arising from beam remnants nonetheless detected (*beam scraping noise filter*).

Events are required to fire the trigger path labelled HLT_Ele27_WP80. It implies the presence of an online electron with a transverse momentum of at least $27\,\mathrm{GeV}$. In addition, it must satisfy the tight electron identification at the working point which ensures a 80% reconstruction efficiency. Trigger efficiency is derived from data and simulated Monte Carlo events using the Tag&Probe method [63]. Performance differences between data and Monte Carlo events are then corrected by introducing a scaling factor estimated using $Z \rightarrow ee$ events. The resulting systematic effect is then handled through reweighing, as detailed in section 5.5.1. This trigger is chosen because it has the loosest selection possible, while being un-pre-scaled and available over the complete 2012 data taking period.

The event must also contain at least one non-fake primary vertex. It means that the vertex position must be correlated to the beam direction. In a frame following the cylindrical structure of the detector, it must have a ρ coordinate (*i.e.* impact parameter) smaller than $2\,\mathrm{cm}$ and a z coordinate smaller than $24\,\mathrm{cm}$.

5.2 Signal Electron

The experimental signature of the physical process $t\bar{t} \rightarrow e\nu_e b\bar{b}q\bar{q}$ implies the presence of a high energy electron, coming directly from one of the W bosons decay. A first kinematic selection criteria is applied. The electron must have a transverse momentum $p_T > 30\,\mathrm{GeV}$ and a pseudo-rapidity $\eta < 2.5$. It implies a similar trigger response in recorded and simulated collisions, while ensuring a trigger efficiency larger than 80%, as shown on figure 5.5. Electrons in the region corresponding

to the transition between barrel and endcap ECAL ($1.4442 < |\eta| < 1.5660$) are excluded due to the presence of readout cables and consequently a gap in the sensitive material, leading to a degraded electron energy resolution.

As seen in section 4.2.2, electrons are identified using a Multi Variate Analysis (MVA) discriminator. During the event selection, the following criteria are applied. The electron, chosen among the *GsfElectron* collection, must have a MVA discriminator large than 0.9. In addition, the *GsfElectron* track impact parameter must not be larger than 0.02 cm with respect to the primary vertex.

Electron isolation is introduced as the ratio between the tracks from hadrons and photons transverse momentum and the electron transverse momentum:

$$I_{rel}^e = \frac{\sum p_T^{charged\,hadrons} + \max\left(0, \sum p_T^{neutral\,hadrons} + \sum p_T^{photons} - \rho A_{eff}\right)}{p_T^{electron}} \quad (5.1)$$

The sum of the tracks transverse momentum consider all tracks within a $\Delta R = 0.3$ cone around the electron trajectory. An effective area correction [64] ρA_{eff} term is included to correct for effects due to pile-up interactions. ρ is the event energy density that represents the soft jets activity in the event. It is evaluated as the median of the p_{Ti}/A_i distribution where i loops over all the jets in the event and A_i stands for the jet area, that can be approximated as $\sim \pi r^2$, r being the radius of the cone used in the jet clustering algorithm. For the determination of ρ, the k_T algorithm is used with a cone size of $R = 0.5$. Effective areas A_{eff} are evaluated from data events, for different $|\eta|$ bins and given in table 5.1. They are determined in the following way. Non corrected relative isolation and event energy density show a linear dependence with respect to the number of primary vertices [65]. Both distributions are fitted with linear functions ($ax + b$) and the effective area is taken as the ratio of the two functions' slopes:

$$A_{eff} = \frac{a_{iso}}{a_\rho} \quad (5.2)$$

The electron relative isolation has to be smaller than $I_{rel} < 0.1$ and in addition no

jets must be present in a ΔR cone of 0.3 around the electron.

	A_{eff}
$|\eta| \leq 1.0$	0.130 ± 0.001
$1.0 < |\eta| \leq 1.479$	0.140 ± 0.002
$1.479 < |\eta| \leq 2.0$	0.070 ± 0.001
$2.0 < |\eta| \leq 2.2$	0.090 ± 0.001
$2.2 < |\eta| \leq 2.3$	0.110 ± 0.002
$2.3 < |\eta| \leq 2.4$	0.110 ± 0.003
$2.4 < |\eta|$	0.140 ± 0.004

Table 5.1: Table showing the Effective Areas for the considered pseudo-rapidity regions [64]

5.3 Di-lepton Veto

To reject any events coming from a di-leptonic decay of the top quark pair, any event containing an additional isolated lepton are discarded. The second lepton is defined with looser selection cuts (loose lepton) compared to the signal one (tight lepton). Below the criteria defining a loose lepton are detailed.

A loose electron is an *GsfElectron* with a transverse momentum $p_T > 20\,\text{GeV}$ and pseudo-rapidity $|\eta| < 2.5$. Its multivariate discriminator had to be larger than 0 and its relative isolation must be smaller than 0.15.

Loose muons are muons from the *PFMuon* collection and have to be identified as global or tracker muons as described in section 4.2.1. They have a transverse momentum $p_T > 10\,\text{GeV}$ and a pseudo-rapidity $|\eta| > 2.5$. The muon relative isolation must also be smaller than 0.2. Muon relative isolation is slightly different from electron relative isolation.

$$I_{rel}^{\mu} = \frac{\sum p_T^{charged\,hadrons} + \max\left(0, \sum p_T^{neutral\,hadrons} + \sum p_T^{photons} - 0.5 \sum p_T^{charged\,particles}\right)}{p_T^{muon}}$$

(5.3)

The correction term for pile-up effects $\sum p_T^{charged\,particles}$ is the sum of the transverse momentum of all the charged particles not coming from the primary vertex and detected around the muon (in a cone of $\Delta R = 0.4$). This deposit is divided by 2 to account for the average charged to neutral particles ratio, as measured in [66]

5.4 Jets

As described in section 4.1, jets are reconstructed using the anti-k_T clustering jet algorithm with a distance parameter of $R = 0.5$. On this particle flow jets collection, a first set of quality criteria is applied. Those requirements are meant to ensure that the jets are reconstructed using several sub-detector elements, therefore leading to an improved jet quality and reducing the jet misidentification probability. For each jet, the charged electro-magnetic fraction has to be smaller than 0.99 as well as the neutral hadronic and neutral electro-magnetic fractions. The charged hadronic fraction has to be larger than zero and the number of charged hadrons in the the jet has to be larger than zero.

In addition, jets are required to have a transverse momentum $p_T > 30\,\text{GeV}$ and a pseudo-rapidity $|\eta| < 2.4$. The number of jets per event is required to be exactly 4 to ensure minimal contamination from initial-state and final-state radiations. It also reduces the number of possible jet-parton permutations when calculating event likelihoods, leading to a reduced computation time.

5.4.1 b-tagging Algorithm

To reduce further the background contamination, amongst the four selected jets, only events with two jets tagged as b-jets are kept. This is achieved by using the *Combined Secondary Vertex* (CSV) b-tagging algorithm. A description can be found in section 4.2.3.3.

The b-tagging algorithm is set to operate at the tight working point ($CSVT$) corresponding to a discriminator value of 0.898. It means a given jet is consid-

ered as a *b*-jet if its *b*-tagging discriminator is larger than 0.898. This choice on the working point is made to ensure a reasonable *b*-tagging efficiency (55%) while keeping the misidentification probability as low as possible ($\sim 3\%$), see figures 5.2 and 4.4.
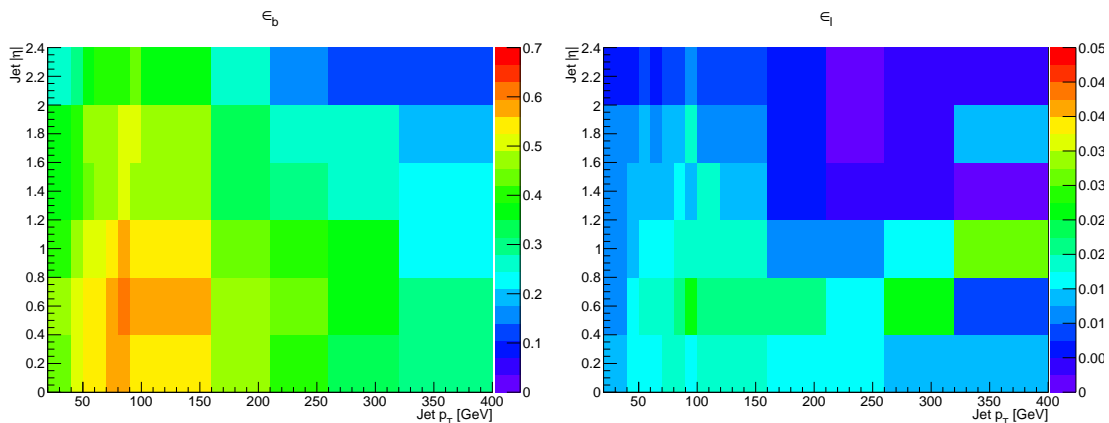
Figure 5.2: *b*-tagging efficiency (left) and light jets mis-tagging rate (right) as a function of the jet p_T and $|\eta|$. Values are expressed as probabilities.

5.4.2   HitFit

HITFIT is a kinematic fitting module, developed by S. Snyder in 1995. It has been used in a top quark mass measurement in the frame of the DØ collaboration [67]. Its principal purpose was to fully reconstruct the event kinematics of a top-anti-top pair decaying in the lepton+jets channel. It has been adapted to be used within the CMS collaboration software by S. Sumowidagdo [68]. In this section, the algorithm working principle and its application to his analysis are presented. Further discussions on its performance can be found in [69]. HITFIT uses as inputs the 4-vectors and the resolutions of the measured final-state particle. The kinematic fitter then builds a χ^2 function as

$$\chi^2 = (\mathbf{x}^f - \mathbf{x}^m)^T \mathbf{R} (\mathbf{x}^f - \mathbf{x}^m) \tag{5.4}$$

$\mathbf{x}^f$ and $\mathbf{x}^m$ represent respectively the fitted and measured 4-vectors of the final-state particles. $\mathbf{R}$ is the inverse covariance matrix containing the resolutions. The χ^2 function is then minimised, via the Lagrange multipliers method and accounting for the following constraints

$$
\begin{aligned}
m_{l\nu} &= m_W \\
m_{q\bar{q}} &= m_W \\
m_t &= m_{\bar{t}}
\end{aligned} \tag{5.5}
$$

The minimisation considers that the constituent mass of both the hadronic and leptonic legs should be equal to the known mass of the W boson and that no mass asymmetry exists between the top and the anti-top quarks. Resolution on the objects are introduced to limit the change in kinematics, so that a given object kinematics returned by HITFIT are still compatible with the kinematics measured by the detector given their resolution. The resolution for an object with a transverse momentum p_T writes as

$$
\sigma = \sqrt{C^2 p_T^2 + R^2 p_T + N^2} \tag{5.6}
$$

where C is a constant, R is the resolution and N is the noise parameter. The parameters are evaluated for different bins in η and separately for each type of particles (b-jets, light jets and electrons). They can be found in appendix A3. Figure 5.3 shows the resolution for a few $|\eta|$ bins. More details on resolution functions can be found in [69].

To further increase the final sample $t\bar{t}$ purity, only events where at least one combination has a χ^2 smaller than 5 are kept. To avoid perturbations that could arise from jet misidentification, the output of b-tagging algorithm is not passed to HITFIT and all possible permutations ($4! = 24$) are considered. Figure 5.11 shows that this value for the discriminant represents a good compromise between the final sample size and the contamination from non $t\bar{t}$ processes.
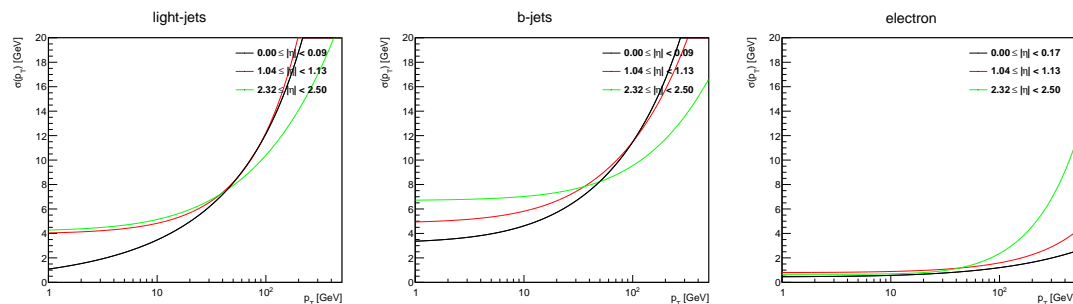
Figure 5.3: Resolution functions analytic form for light jets (left), b-jets (centre) and electrons (right) for several $|\eta|$ bins. The parameters for all the considered $|\eta|$ bins are shown in appendix A3.

5.5   Event Tuning and Reweighing

In this section the various weights used to tune simulated events are presented. Generally, they aim to correct for any discrepancy that can been observed between the analysis of simulated events and recorded events. These discrepancies come from different effects. For example the detector configuration evolves regularly during the data taking period (technical issues, better understanding of the detector). Therefore, a simulated sample generated at a given time may need to be corrected to account for the evolution of the detector settings. Discrepancies can also come from an imperfection in the physics process modelling. This section lists all the sources responsible for reweighing simulated events and how the weights are derived.

5.5.1   Trigger and Lepton Identification weights

All the simulated events selected and used in this analysis are reweighed to account for any discrepancy in the trigger and lepton identification efficiencies. In this case, the weights are derived as scale factors, *i.e.* the efficiency ratio between recorded events and simulated events.

The trigger and electron lepton identification scale factors are evaluated

by a dedicated Physics Object Group in the CMS collaboration [17, 65] and are summarised in table 5.2 and 5.3. They are derived for the identification/isolation and trigger efficiencies shown respectively on figures 5.4 and 5.5
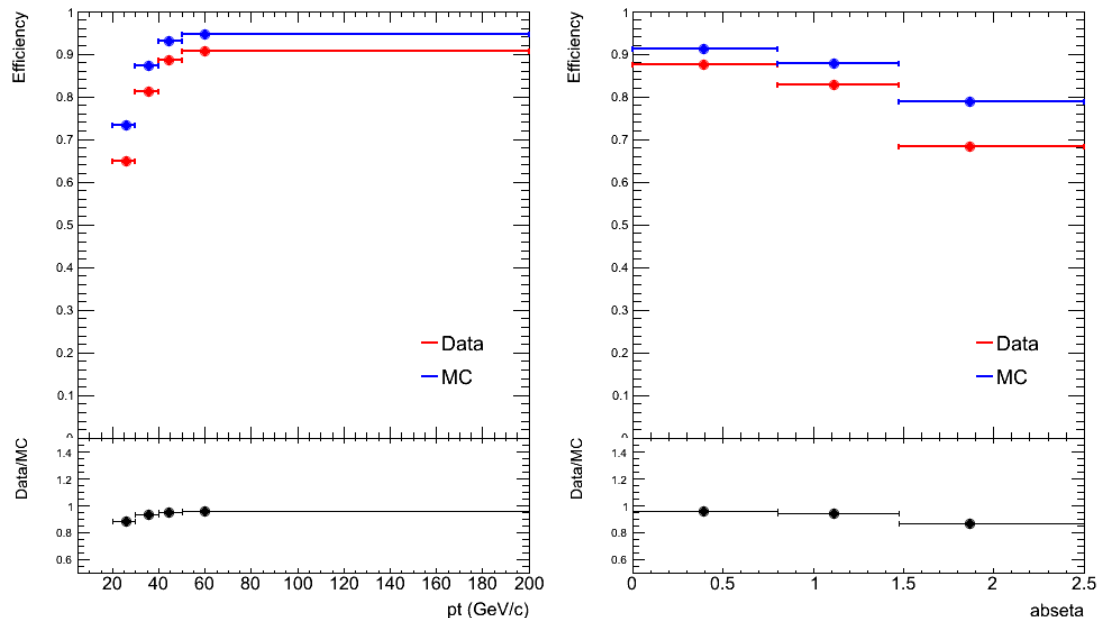
Figure 5.4: ID/ISO efficiency for electron with an MVA discriminator > 0.9 as a function of the electron transverse momentum (left) and pseudo-rapidity (right) [17].

	$30 < p_T < 40$	$40 < p_T < 50$	$50 < p_T < 200$
$0 < |\eta| < 0.8$	$0.987^{+0.012}_{-0.017}$	0.997 ± 0.001	0.998 ± 0.002
$0.8 < |\eta| < 1.478$	$0.964^{+0.002}_{-0.001}$	0.980 ± 0.001	0.988 ± 0.002
$1.478 < |\eta| < 2.5$	$1.004^{+0.006}_{-0.006}$	1.033 ± 0.007	$0.976^{+0.015}_{-0.012}$

Table 5.2: Scale factors for the single electron trigger path HLT_Ele27_WP80. Scale factors are provided for several bins of the electron transverse momentum and pseudo-rapidity

5.5.2 Pile-up Interaction weights

Simulated events are generated considering a given pile-up distribution. During data taking period, the instantaneous luminosity delivered by the LHC
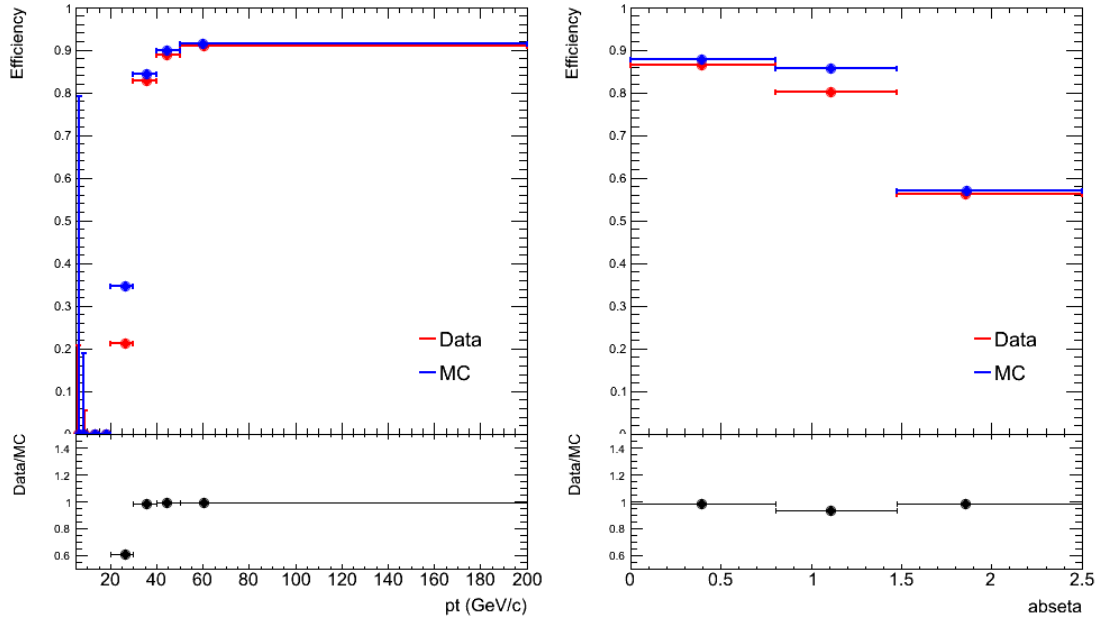
Figure 5.5: Trigger efficiency for the trigger path HLT_Ele27_WP80 as a function of the electron transverse momentum (left) and pseudo-rapidity (right) [17].

accelerator increases, therefore shifting the pile-up distribution towards a higher number of pile-up interactions. This analysis uses the recommendations provided by the collaboration [70]. Figure 5.6 shows the normalised number of pile-up interactions for the simulated and for recorded events. Each simulated event, depending on its number of true pile-up interactions, gets a weight, defined as the ratio of the distributions shown on figure 5.6. The dashed line represent the uncertainty on the pile-up distribution by shifting the average number of pile-up interactions by $\pm 5\%$ and are used to estimate the systematic uncertainty on the top quark mass due to pile-up interactions. Further discussion can be found in section 12.3.3

5.5.3 Parton Distribution Functions

The simulated events are generated using the parton distribution function CTEQ6L1 library. However, no uncertainties are provided for this set of PDF libraries. Uncertainties on the PDF are provided for the more recent CT10NNLO

66

	$30 < p_T < 40$	$40 < p_T < 50$	$50 < p_T < 200$
$0 < |\eta| < 0.8$	0.939 ± 0.003	0.950 ± 0.001	0.957 ± 0.001
$0.8 < |\eta| < 1.478$	0.920 ± 0.002	0.949 ± 0.002	0.959 ± 0.003
$1.478 < |\eta| < 2.5$	0.907 ± 0.005	0.937 ± 0.008	$0.954^{+0.011}_{-0.010}$

Table 5.3: Scale factors for the electron identification and isolation. Scale factors are provided for several bins of the electron transverse momentum and pseudorapidity
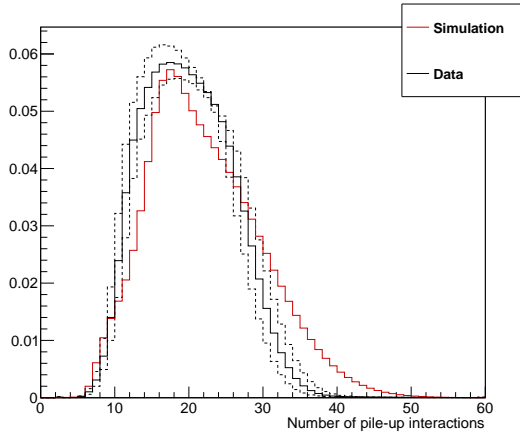
Figure 5.6: Number of pile-up interactions for simulated events in red and recorded events in black. The dashed lines represent the uncertainty on the pile-up distribution estimated on recorded events

libraries. Therefore, during reconstruction, simulated events are reweighed to the envelope of CT10NNLO. Uncertainties on this set of PDF will be accounted for later on as a systematic uncertainty (see section 12.2.8). For each simulated event the partons momentum fraction is evaluated, and the event is attributed a weight equals to the ratio of the two PDF at these given momentum fraction values.

5.5.4 b-tagging Reweighing

To account for performance difference in simulated and recorded events, a reweighing technique is used so that b-tagging efficiency (ϵ_b), charm misidentification rate (ϵ_c) and light jet misidentification rate (ϵ_l) in simulated events are corrected to reflect the performance of b-tagging in recorded events. It is relatively

easy to determine ϵ_b and ϵ_l on simulated events, knowing the true flavour on the reconstructed jets. This information is obviously not available on data recorded by the detector, therefore it is more complicated to estimate b-tagging performance on those events.

The behaviour of the algorithm also depends on the event topology, therefore depends on the selection criteria used in a given analysis. Nevertheless it is possible to evaluate performance on recorded events by estimating performance on simulated events and correcting it by using scale-factors. The scale factors do not depend on the event topology and are centrally provided by the CMS collaboration dedicated working group [15]. Scale factors are provided for b-jets, c-jets and light (u, d, s, g) jets. First, $b-$tagging performance is evaluated on simulated events for different (p_T, η) bins by considering the following ratios:

$$
\begin{aligned}
\epsilon_b &= \frac{n_{b-jets,tagged}}{n_{b-jets}} \\
\epsilon_c &= \frac{n_{c-jets,tagged}}{n_{c-jets}} \\
\epsilon_l &= \frac{n_{l-jets,tagged}}{n_{l-jets}}
\end{aligned}
\tag{5.7}
$$

ϵ_b, ϵ_c, and ϵ_l are shown on figure 5.7. To reweigh simulated events accordingly, the probabilities to find a given jet configuration in a given simulated or recorded event are introduced

$$
\begin{aligned}
P_{MC} &= \prod_{i=tagged} \epsilon_i \prod_{j=not\ tagged} (1 - \epsilon_j) \\
P_{DATA} &= \prod_{i=tagged} \epsilon_i SF_i \prod_{j=not\ tagged} (1 - \epsilon_j SF_j)
\end{aligned}
\tag{5.8}
$$

Those probabilities are evaluated for each events, looping on all the selected jets, and replacing (ϵ_i, SF_i) in the above equation by $(\epsilon_{b,c,l}, SF_{b,c,l})$ depending on the true jet flavour. The weight associated to the simulated event is then given by

$$
\omega = \frac{P_{DATA}}{P_{MC}}
\tag{5.9}
$$
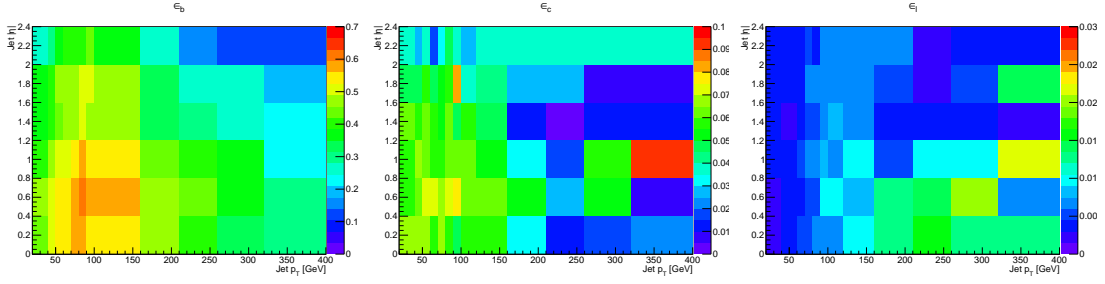
Figure 5.7: *b*-tagging efficiency (left), charm jets (right) and light jets mis-tagging rate (down) as a function of the jet p_T and $|\eta|$. Values are expressed as probabilities.

5.5.5 Top p_T reweighing

It was found in several top-anti-top differential cross-sections analysis [71–73] that the transverse momentum distributions of the top quarks presented a different behaviour in simulated and recorded events. Indeed, the spectrum in recorded events tends to be softer than predictions from simulation. This discrepancy can be reduced by considering NNLO predictions [74]. Consequently, event scale factors have been derived to correct NLO simulated events. The weight is given by the following:

$$\omega = \sqrt{SF(p_{Tt})SF(p_{T\bar{t}})}$$
$$SF(p_T) = e^{a+bp_T} \tag{5.10}$$

where p_{Tt} and $p_{T\bar{t}}$ stand for the top and anti-top transverse momenta respectively, and $SF(p_T)$ is given by the coefficients listed in table 5.4 depending on the decay channel of the top quark pair. The weight distribution for $t\bar{t} \to e+$jets events is shown on figure 5.8

channel	a	b
semi-leptonic	0.159	-0.00141
di-leptonic	0.148	-0.00129
combined	0.156	-0.00137

Table 5.4: Coefficients used in the determination of the top p_T event weight [75].
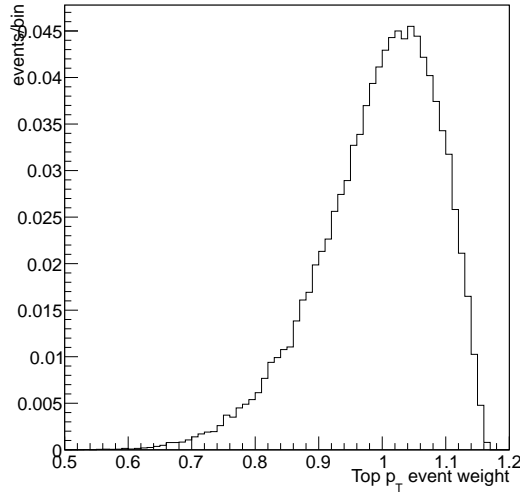
Figure 5.8: Top p_T event weight distribution for $t\bar{t} \rightarrow e+$jets events

5.6 Selection Summary

In this section, the selection criteria impact on the final sample is detailed. More specifically, all the cuts added after the requirement $n_{jets} \geq 4$ are discussed. This step corresponds to a standard selection recommended by the Top Physics dedicated Physics Analysis Group [62]. This set of criteria will be labelled as "reference selection" in the following. As the matrix element method performs well on a sample with limited size, and to reduce likelihood calculation time, additional selection criteria are applied on top of the reference selection.

Several physical processes mimic the signature of a top quark pair decaying into an electron and four jets. The reference selection sample still contain a sensible amount of background that could bias the measurement during the likelihood calculation stage. Table 5.5 shows the sample composition after various steps in the analysis selection, and lists each process contribution fraction once the full selection has been applied.

The samples associated to the simulated physics processes are listed in table

	$t\bar{t} \to e$	$t\bar{t} \to other$	$Z+jets$	$SingleTop$	$Wl\nu+jets$	$TotalMC$	$Data$	$\epsilon_{t\bar{t}}$	π
primary vertex	334684 ± 223	178828 ± 157	4187663 ± 3399	97224 ± 415	10266065 ± 10121	15064466 ± 7515	203411893	0.4808	0.0222
isolated electron	285995 ± 203	138335 ± 138	2141135 ± 2413	81109 ± 379	7825891 ± 8751	10472467 ± 6205	66423945	0.4108	0.0273
muon veto	285496 ± 203	94119 ± 114	2134477 ± 2410	78391 ± 371	7825286 ± 8751	10417770 ± 6203	66334133	0.4101	0.0274
di-electron veto	285270 ± 203	83482 ± 107	1407140 ± 1954	77740 ± 370	7823859 ± 8750	9677493 ± 6040	64025881	0.4098	0.0295
$n_{jets} \geq 4$ @30GeV	138969 ± 140	20939 ± 53	20080 ± 233	9834 ± 129	100713 ± 440	290537 ± 371	330583	0.1996	0.4783
$n_{jets} = 4$	83630 ± 108	13713 ± 43	16335 ± 210	7238 ± 111	82618 ± 412	203536 ± 330	236424	0.1201	0.4109
$2\times$ CSVT	16354 ± 47	2766 ± 19	134 ± 19	960 ± 41	378 ± 24	20595 ± 60	20806	0.0235	0.7941
HitFit	7967 ± 33	791 ± 10	34 ± 9	280 ± 21	64 ± 14	9138 ± 38	8898	0.0114	0.8718
fraction	0.8718	0.0866	0.0037	0.0306	0.0070				

Table 5.5: Table detailing all the cut-flow steps used in the analysis. The simulated events yields are normalised to the measured luminosity $\mathcal{L} = 19.762\,\text{fb}^{-1}$. The last two columns represent respectively the $t\bar{t} \to e+jets$ signal selection efficiency and purity. The last row represents the event fraction after the complete selection.

3.1 and in appendix A1. The event yields for simulated events have been scaled to correspond to the measured luminosity of $\mathscr{L} = 19.6\,\mathrm{fb}^{-1}$ and represent the number of expected events for each process after selection criteria are applied. For each step, the agreement between the number of expected events in simulation and the number of recorded events is discussed. The $t\bar{t} \to e + jets$ selection efficiency ($\epsilon_{t\bar{t}}$) and purity (π) defined below can also be found in the last two columns of table 5.5. The selection efficiency for a physical process is given by the ratio between the number of events passing the selection and the total number of events generated for this given process. The purity is given by the ratio between the number of selected events associated to a process and the total number of events passing the selection.

$$
\begin{aligned}
\epsilon_{t\bar{t}} &= \frac{N_{t\bar{t}\to e}^{sel}}{N_{t\bar{t}\to e}^{tot}} \\
\pi &= \frac{N_{t\bar{t}\to e}^{sel}}{N^{sel,\,tot}}
\end{aligned}
\tag{5.11}
$$

Figure 5.9 shows the distribution of the jets transverse momentum and the jet multiplicity after the reference selection. At this stage of the selection, the main background contribution comes from W+jets. A W boson decaying leptonically and produced in association with jets can mimic the top quark semi-leptonic decay signature. An other major background contribution comes from the decay of Z bosons. If one of the lepton from the Z boson decay is not correctly identified, the event will not be discarded by the reference selection. An excess of data can be observed between simulation and data yields. This is due to the presence of a multi-jet background, not used here. This process is mostly composed of low energy, light flavoured-jets. If one jet is misidentified as an electron (fake) or if an electron from a hadron decay is misidentified as an isolated electron, the event signature is similar to the top quark decay signature.

Figure 5.10 shows the b-tagged jets multiplicity and the jet transverse momentum distribution after requiring exactly four jets per event. Jets are considered b-tagged if the discriminator related to the Combined Secondary Vertex is larger

than 0.898, corresponding to the *Tight* working point (CSVT). Requiring exactly two jets tagged among the four selected reduces significantly the amount of background, doubles the signal purity, and totally suppresses the multi-jet background. The *b*-tagging algorithm output is only used to discriminate between signal and background events. It is not used by HITFIT either MADWEIGHT has discussed in sections 5.4.2 and 6.3 respectively.

Figure 5.11 shows the HITFIT χ^2 distribution of the best jet-parton combination in an event after requiring exactly two *b*-tagged jets. As depicted in section 5.4.2, events are kept only if the χ^2 of the best combination is smaller than five. This helps increasing further the sample signal purity to about 87%.

Figures 5.12 to 5.16 show the distributions of the selected objects once the complete selection has been applied. Distributions of simulated events show a good agreement with distributions from data events. The content of selected data events can be properly estimated in terms of physical processes contributions, allowing to perform tests and calibration of the analysis method before applying it to data recorded by the detector.
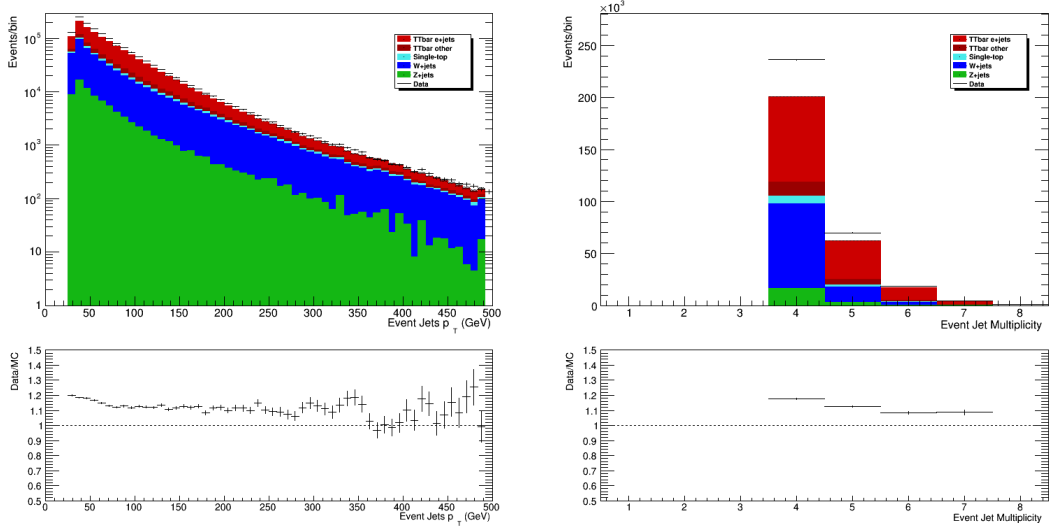
Figure 5.9: Transverse momentum (left) and jet multiplicity (right) distributions of all the jets after the reference selection. The excess of data is due to the non-modelled multi-jet background.
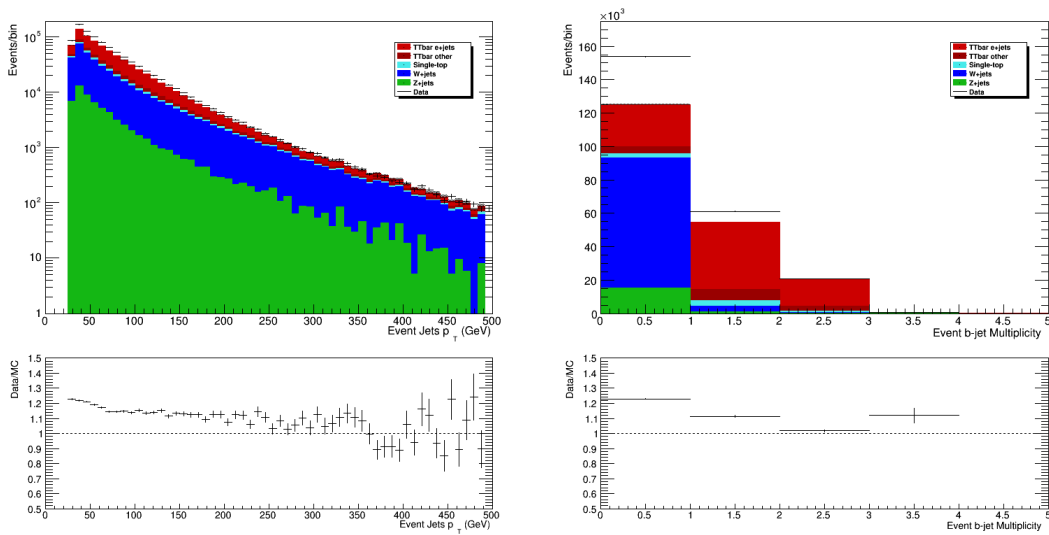
Figure 5.10: Transverse momentum (left) and b-jet multiplicity (right) distributions of all the jets after the requirement $N_{jets} = 4$.
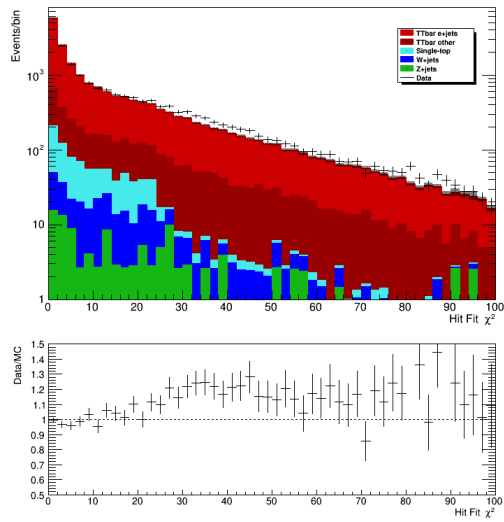
Figure 5.11: Minimum HITFIT χ^2 distribution after requiring exactly two b-tagged jets
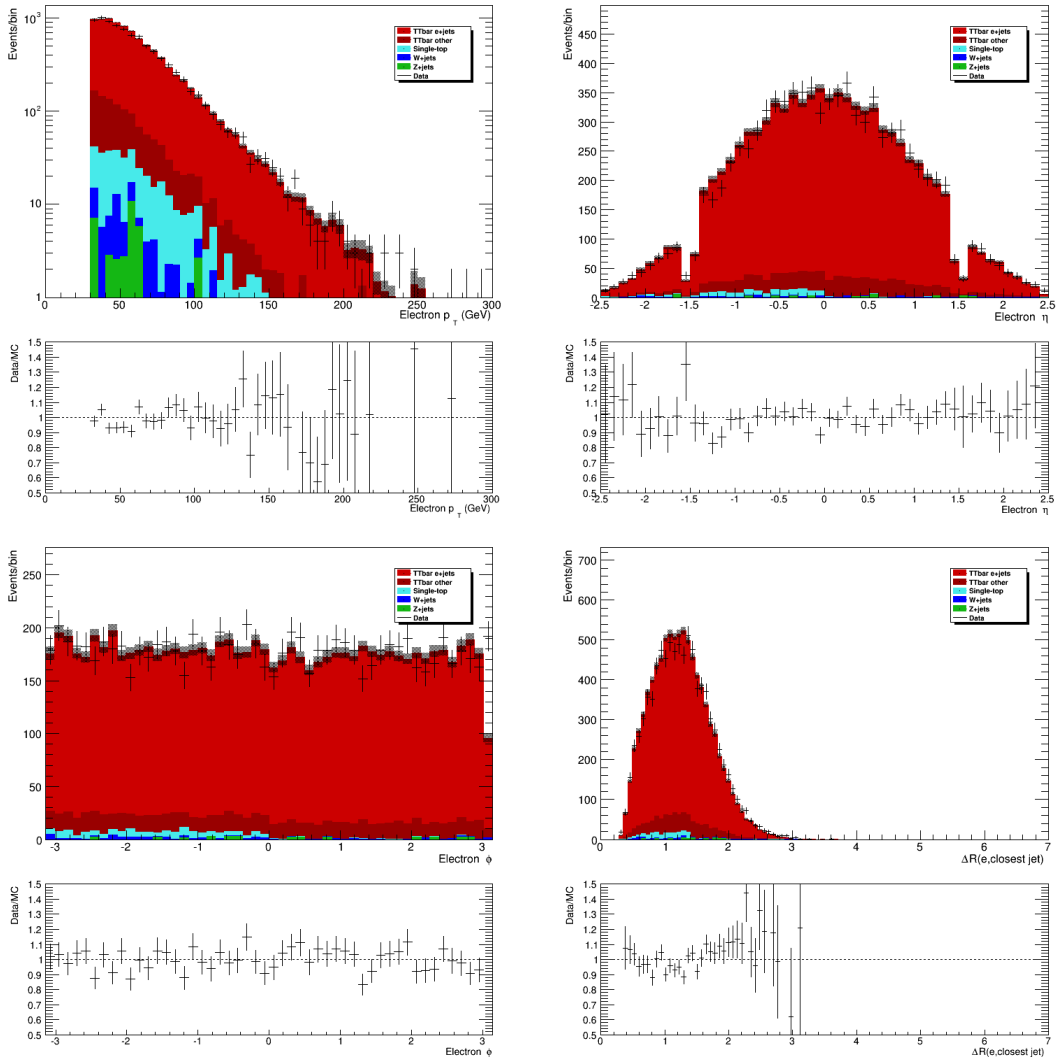
Figure 5.12: Transverse momentum (upper left) pseudo-rapidity (upper right) azimuthal angle (lower left) distributions of the selected electron after the complete selection has been applied. The distribution of the ΔR between the electron and the closest jet is shown on the lower left plot.
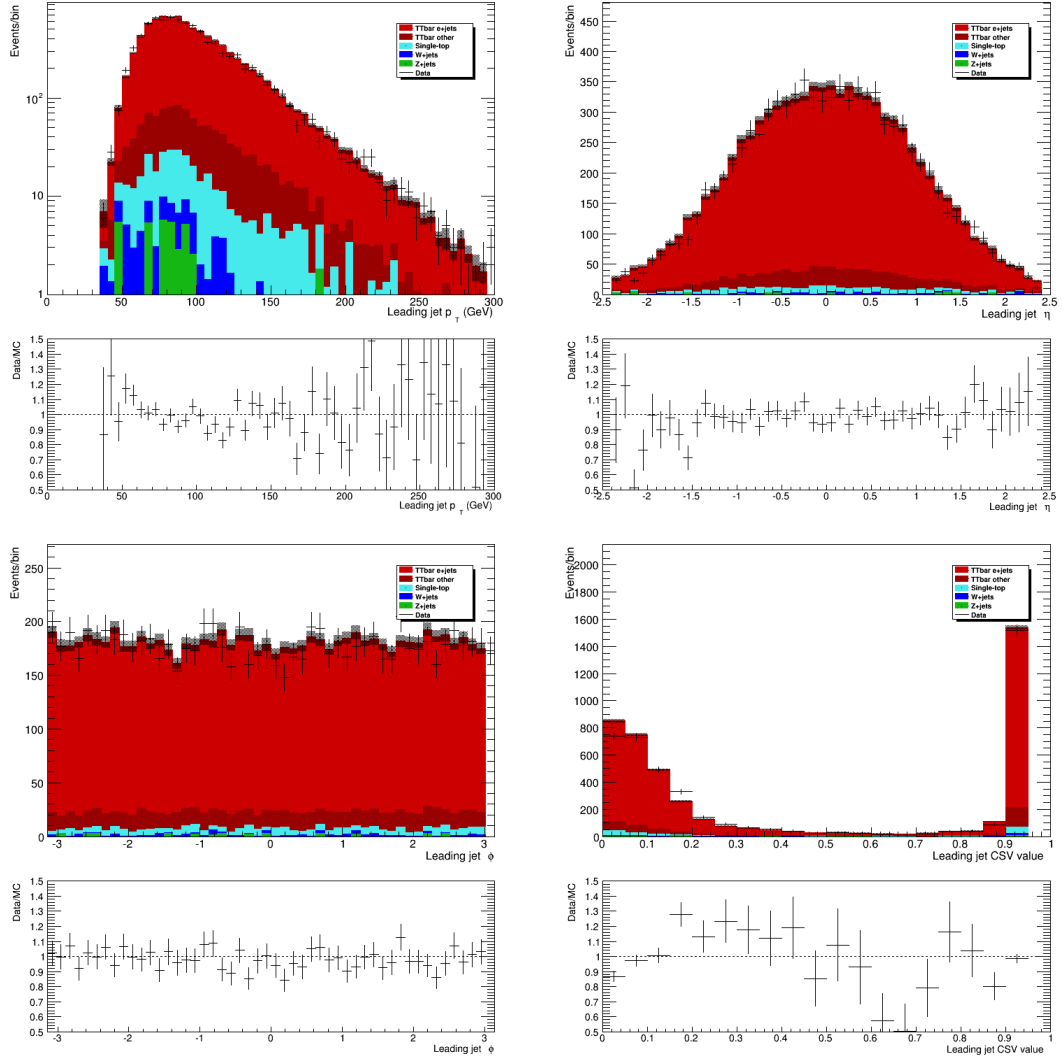
Figure 5.13: Transverse momentum (upper left), pseudo-rapidity (upper right), azimuthal angle (lower left), and b-tagging discriminant (lower left) distributions of the selected leading jet after the complete selection has been applied.

77
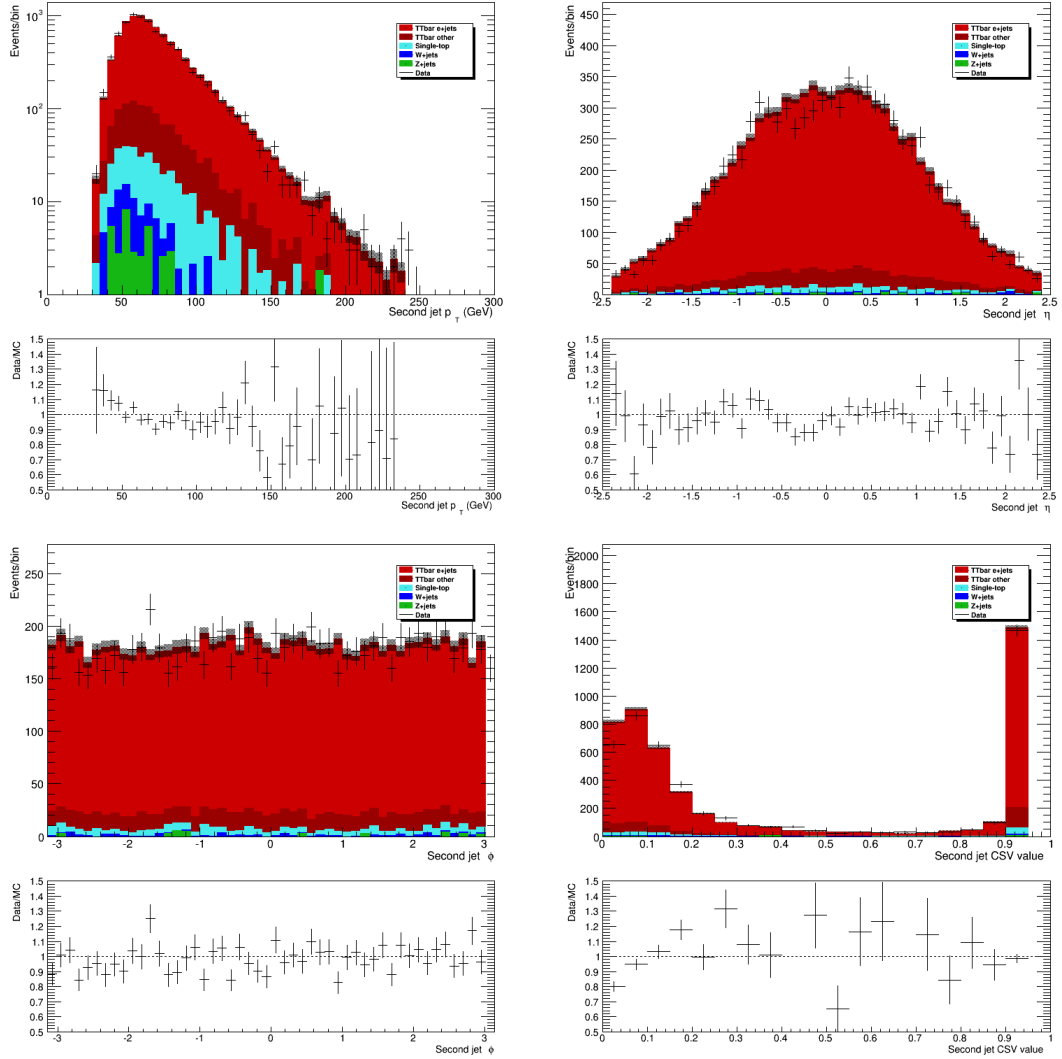
Figure 5.14: Transverse momentum (upper left), pseudo-rapidity (upper right), azimuthal angle (lower left), and *b*-tagging discriminant (lower left) distributions of the selected second leading jet after the complete selection has been applied.

78
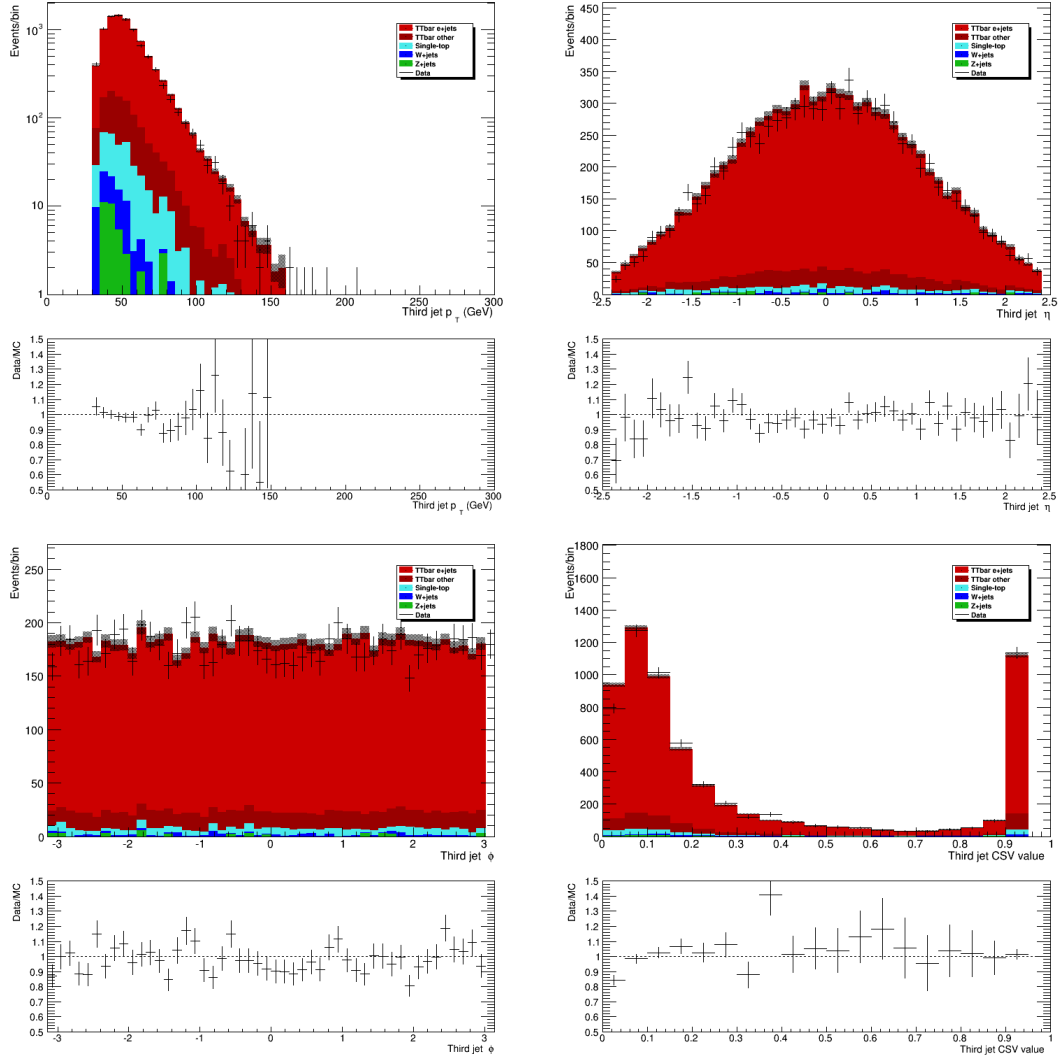
Figure 5.15: Transverse momentum (upper left), pseudo-rapidity (upper right), azimuthal angle (lower left), and *b*-tagging discriminant (lower left) distributions of the selected third leading jet after the complete selection has been applied.
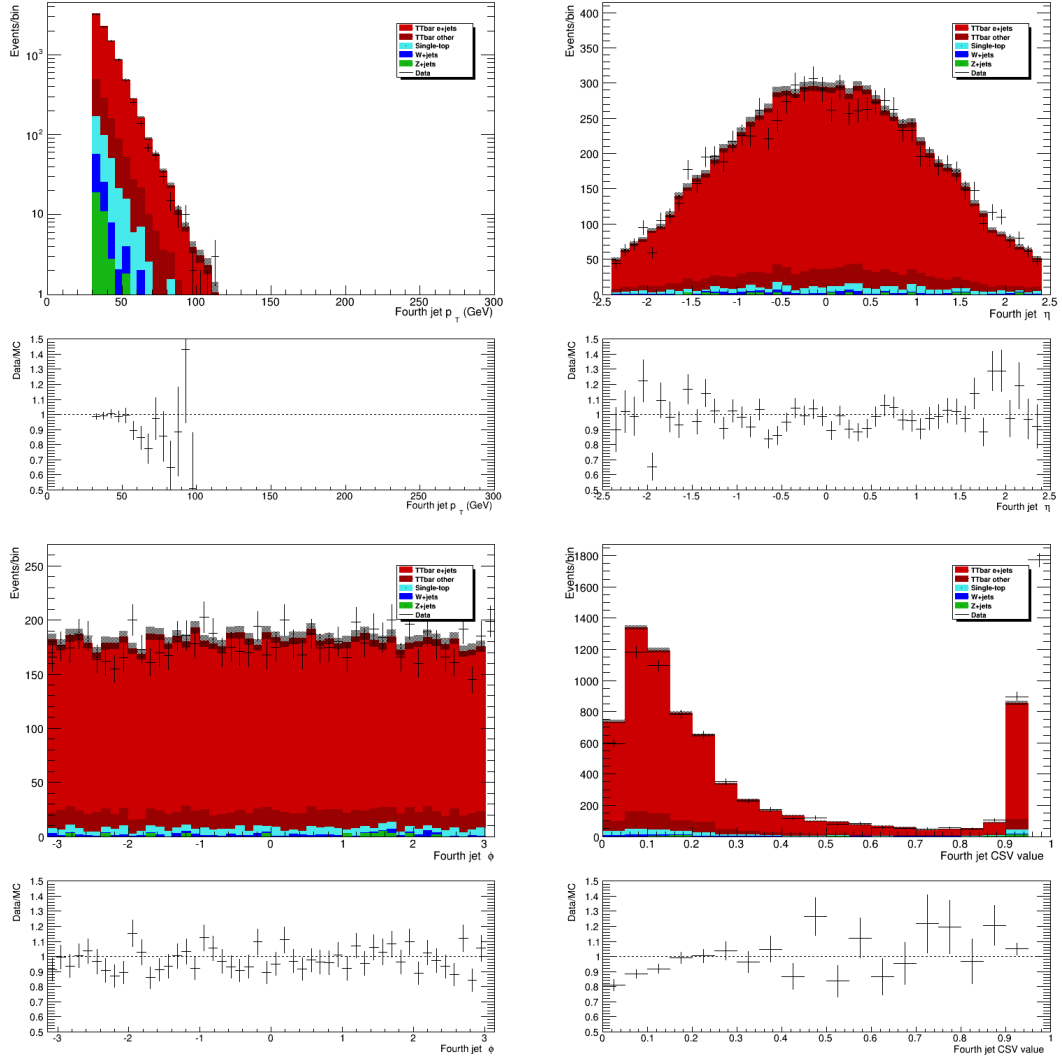
Figure 5.16: Transverse momentum (upper left), pseudo-rapidity (upper right), azimuthal angle (lower left), and *b*-tagging discriminant (lower left) distributions of the selected fourth leading jet after the complete selection has been applied.

Part III

The Matrix Element Method

Chapter 6

Likelihood calculation

The Matrix Element Method [76] was developed shortly after the discovery of the top quark, with the goal to reduce as much as possible the uncertainty on the measurement of the top quark's properties. It is a sophisticated and accurate method, based on likelihood estimation on an event-by-event basis. Its name comes from its ability to compare experimental signatures with event topologies predicted by the matrix element. It is also very flexible in the sense it can evaluate any unknown parameters from the theory modelling and/or from the detector response. The sample likelihood is evaluated for different hypothesis on the parameter(s) to be estimated, and its maximisation leads to the measured value. In this chapter, a general definition of likelihood is first presented. Then, its application to top pairs decaying following the single-electron decay channel is detailed. Finally, the details of the computation using MADWEIGHT software is introduced.

6.1   Definition

The sample likelihood $\mathscr{L}_{sample}$ is defined as the product of all the events likelihoods as :

$$\mathscr{L}_{sample}(\mathbf{x}; \mathbf{s}, \mathbf{f}) = \prod_{i=1}^{n} \mathscr{L}_{evt}(x_i; \mathbf{s}, \mathbf{f}) \tag{6.1}$$

where x_i stands for the kinematic properties of the ith event, $\mathbf{s}$ represents the parameters to measure and $\mathbf{f}$ the event fraction of a given physical process con-

tributing to the sample, with $\sum_P f_P = 1$.

A given event can be produced by several processes. The event likelihood is therefore a linear combination of the likelihoods to observe this event via a given process P:

$$\mathscr{L}_{evt}(x_i; \mathbf{s}, \mathbf{f}) = \sum_P f_P \mathscr{L}_P(x_i; \mathbf{s}) \tag{6.2}$$

In this analysis, in order to reduce the calculation time and because of a high signal purity as seen in chapter 5, the background processes are neglected as they do not have a significant impact on the measurement. Therefore,

$$\mathscr{L}_{evt}(x_i; \mathbf{s}, \mathbf{f}) = \mathscr{L}_{t\bar{t}}(x_i; \mathbf{s}) \tag{6.3}$$

Selected background events contribution is assumed to be $t\bar{t}$-like. Any bias introduced by this assumption is absorbed in the calibration curve.

6.2   Process Likelihood

In this section the determination of $\mathscr{L}_{t\bar{t}}$ is detailed. It represents the likelihood of an event produced by proton-proton hard-scattering producing a $t\bar{t}$ quark pair and decaying to a lepton, a neutrino, two b-jets and two light jets. It is proportional to the differential cross-section $d\sigma_{t\bar{t}}(pp \rightarrow b\bar{b}q\bar{q}l\nu; \mathbf{s})$. In the following, the $b\bar{b}q\bar{q}l\nu$ final state is denoted y.

The differential cross-section is constructed as the convolution of the probability for the incoming protons to produce the initial-state partons a_1 and a_2 via parton density functions, the differential cross-section for the partonic process $(a_1 a_2 \rightarrow y)$, and the transfer functions as the probability for the detector to reconstruct the observed event given the final-state partons $q_1...q_6$. This convolution is pictured in figure 6.1.

Starting from the differential cross-section $d\sigma_{t\bar{t}}$ associated to the process
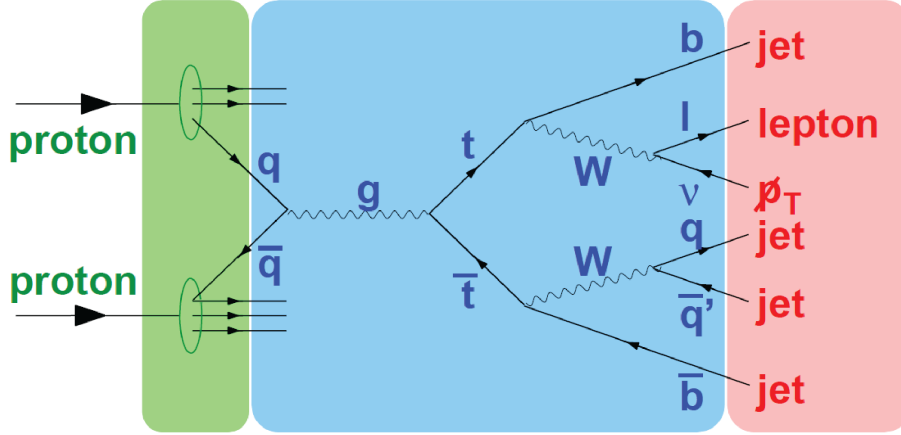
Figure 6.1: Schematic of the convolution performed by the Matrix Element Method. The green block represents the integration on the initial state (PDF), the blue block pictures the integration over the theory model considered (Matrix Element) and the red block shows the integration over the final state products and the detector response (transfer functions)

$a_1 a_2 \rightarrow y$, written in [22] as:

$$d\sigma_{t\bar{t}}(a_1 a_2 \rightarrow y; \mathbf{s}) = \frac{(2\pi)^4 |\mathcal{M}_{t\bar{t}}(a_1 a_2 \rightarrow y; \mathbf{s})|^2}{4\sqrt{(p_1 \cdot p_2)^2 - m_1^2 m_2^2}} d\Phi_6(p_1 + p_2; q_1...q_6) \qquad (6.4)$$

where p_i and m_i are respectively the four-momenta and the mass of the initial state partons (a_i) and $\mathcal{M}_{t\bar{t}}$ is the Lorentz-invariant matrix element for the $a_1 a_2 \rightarrow y$ process, which gives its name to the method. $d\Phi_6$ denotes the phase space element for the 6 final state objects four-momenta (q_i), defined as:

$$d\Phi_6(p_1 + p_2; q_1...q_6) = \delta^4 \left(p_1 + p_2 - \sum_{i=1}^{6} q_i \right) \prod_{i=1}^{6} \frac{d^3 q_i}{(2\pi)^3 2E_i} \qquad (6.5)$$

To reflect the proton structure, equation 6.4 is convoluted with parton density functions and summed over all possible flavour compositions of the incoming par-

tons,

$$d\sigma_{t\bar{t}}(pp \to y; \mathbf{s}) = \int_{\xi_1, \xi_2} \sum_{a_1, a_2} d\xi_1 d\xi_2 f^{a_1}(\xi_1) f^{a_2}(\xi_2) d\sigma_{t\bar{t}}(a_1 a_2 \to y; \mathbf{s}) \qquad (6.6)$$

where $f^{a_i}(\xi_i)$ are the parton density functions, describing the probability of finding a parton with flavour a_i and momentum fraction ξ_i in the ith proton.

The last step is meant to introduce the detector response, including resolution effects so that the event partonic final state y is reconstructed as x. The differential cross-section then becomes:

$$d\sigma_{t\bar{t}}(pp \to x; \mathbf{s}) = \int_y d\sigma_{t\bar{t}}(pp \to y; \mathbf{s}) W(x, y; \mathbf{s}) \qquad (6.7)$$

where $W(x, y; \mathbf{s})$ is the transfer function, representing how the detector maps partons to reconstructed objects, see chapter 7. The integrated cross-section is introduced as:

$$\sigma_{t\bar{t}}^{tot}(\mathbf{s}) = \int_{x,y} d\sigma_{t\bar{t}}(pp \to y; \mathbf{s}) W(x, y; \mathbf{s}) dx \qquad (6.8)$$

From equations 6.7 and 6.8, the normalised likelihood is thus written as:

$$\mathscr{L}_{t\bar{t}}(x; \mathbf{s}) = \frac{d\sigma_{t\bar{t}}(pp \to x; \mathbf{s})}{\sigma_{t\bar{t}}^{tot}(\mathbf{s})} \qquad (6.9)$$

The integrated cross-section $\sigma_{t\bar{t}}^{tot}(\mathbf{s})$ is evaluated and returned by MAD-WEIGHT, using the same matrix element as for the likelihood estimation. It is calculated for each hypothesis $\mathbf{s}$.

It remains to normalise the likelihood for acceptance effects. Because of the detector does not cover the entire phase-space and background contribution need to be rejected as much as possible, not all the events will be used for the sample likelihood estimation. Therefore, kinematic selection and the geometrical limitations of the detector need to be accounted for. These phase-space restrictions are considered by including an acceptance function ($f_{acc}(x)$) when normalising the

likelihood. The transfer functions (see chapter 7) have been chosen such that $\int W(x, y; \mathbf{s})dx = 1$ by construction. So, to ensure a proper normalised likelihood, $\mathscr{L}_{t\bar{t}}(x; \mathbf{s}) = \mathscr{L}_{evt}(x; \mathbf{s})$ is substituted by

$$\bar{\mathscr{L}}_{evt}(x; \mathbf{s}) = \frac{1}{C}\mathscr{L}_{evt}(x; \mathbf{s})f_{acc}(x) \tag{6.10}$$

$\bar{\mathscr{L}}_{evt}(x; \mathbf{s})$ is normalised if C is defined as

$$C = \int \mathscr{L}_{evt}(x; \mathbf{s})f_{acc}(x; \mathbf{s})dx \tag{6.11}$$

with $f_{acc}(x) = 1$ if the event is accepted and $f_{acc}(x) = 0$ otherwise. Combining equations 6.1, 6.9 and 6.10 the logarithm of the sample likelihood reads

$$
\begin{aligned}
\ln \mathscr{L}_s(\mathbf{x}; \mathbf{s}) &= \ln \left(\prod_{i=1}^{n} \bar{\mathscr{L}}_{evt}(x; \mathbf{s}) \right) \\
&= \sum_{i=1}^{n} \ln \bar{\mathscr{L}}_{evt}(x; \mathbf{s}) \\
&= \sum_{i=1}^{n} \ln \mathscr{L}_{evt}(x; \mathbf{s}) + n \ln f_{acc}(x) - n \ln \int f_{acc}(x)\mathscr{L}_{evt}(x; \mathbf{s})dx \\
&\simeq \sum_{i=1}^{n} \ln \mathscr{L}_{evt}(x; \mathbf{s}) - n \ln \int f_{acc}(x)\mathscr{L}_{evt}(x; \mathbf{s})dx
\end{aligned}
\tag{6.12}
$$

The term $n \ln f_{acc}(x)$ can be dropped as it is a constant factor and therefore has no effect on the likelihood maximisation. The correction applied to $\mathscr{L}_{evt}(x; \mathbf{s})$ describes the acceptance term. The term

$$\int f_{acc}(x)\mathscr{L}_{evt}(x; \mathbf{s})dx = \frac{n_{t\bar{t}, accepted}}{n_{t\bar{t}, generated}} = \bar{f}_{acc}(\mathbf{s}) \tag{6.13}$$

represents the mean acceptance function, derived from [77], and can be easily estimated on Monte Carlo samples generated for different $\mathbf{s}$. Figure 10.1 shows an example of the acceptance curve, and in this case has been evaluated on fully simulated $t\bar{t} \to e+$jets Monte Carlo events.

When moving out the integrated cross-section out of the integral and including corrections for acceptance effects, the normalised sample likelihood becomes:

$$
\begin{aligned}
\ln \mathscr{L}_s(\mathbf{x}; \mathbf{s}) &\simeq \sum_{i=1}^{n} \ln \mathscr{L}_{evt}(x; \mathbf{s}) - n \ln \bar{f}_{acc}(\mathbf{s}) \\
&= \sum_{i=1}^{n} \ln \left(\frac{1}{\sigma_{t\bar{t}}^{tot}(\mathbf{s})} \int_{x,y} d\sigma_{t\bar{t}}(pp \to y; \mathbf{s}) W(x, y; \mathbf{s}) dx \right) - n \ln \bar{f}_{acc}(\mathbf{s}) \\
&= \sum_{i=1}^{n} \ln \int_{x,y} d\sigma_{t\bar{t}}(pp \to y; \mathbf{s}) W(x, y; \mathbf{s}) dx - n \ln \bar{f}_{acc}(\mathbf{s}) - n \ln \sigma_{t\bar{t}}^{tot}(\mathbf{s}) \\
&\equiv \sum_{i=1}^{n} \ln \mathscr{L}_{MW} - n \ln \bar{f}_{acc}(\mathbf{s}) - n \ln \sigma_{t\bar{t}}^{tot}(\mathbf{s})
\end{aligned}
$$

(6.14)

$\mathscr{L}_{MW}$ is introduced to reflect the part of the phase space integration done by MADWEIGHT.

6.3   Monte Carlo Integration

While the concepts behind the method are rather simple, the likelihood computation requires calculating a sophisticated multi-dimensional integral. Indeed, the integrand ($\mathscr{L}_{MW}$) has a highly fluctuating shape and depends on many variables. This step is performed using MADWEIGHT [78], a phase-space generator, developed as a MADGRAPH module. MADWEIGHT is based on the VEGAS adaptive Monte Carlo integration algorithm. It sweeps the original phase-space and iteratively samples it, focusing on regions of the phase-space where the contributions to the integral are the largest. This ensures the phase space is segmented enough where the integrand is maximum and allows an efficient integration.

Likelihood estimation is entirely done with MADWEIGHT. It has to be provided with parton density functions, transfer functions and the physical processes

it has to consider in the computation. Its link with MADGRAPH ensures it has direct access to Feynman diagrams contributing to the process and therefore it has access to the matrix element.

Evaluating the likelihood integral requires a lot of computation, therefore several optimisations and assumptions are made. As introduced in the previous section, likelihoods are only evaluated for the signal process. The matrix element for the single-electron decay channel of a top quark pair includes then 30 Feynman diagrams. Including the W+jets background in the likelihood estimation would have brought in the integral more than 3000 additional diagrams. A high signal purity obtained with a tight selection as described in chapter 5 makes the "signal-only matrix element" assumption valid. MADWEIGHT also gives priority to jet-parton combinations that are more likely to give rise to the reconstructed event, considerably reducing the integration time. This last optimisation has been proved to be quite effective, as no major gain in computing time has been observed when reducing the number of possible permutations from 24 to 4 (by using b-tagging information). However, even with the considered optimisations, the processing of a single fully reconstructed event, with the use of realistic transfer functions can take up to one hour.

6.4   MadWeight Input

MADWEIGHT uses LHC Olympics files as data input for the likelihood calculations. They contain the event kinematics and basic information of the final-state particles four-vectors. A description of the LHCO files can be found in [79]. The analysis selection code has been adapted to provide such text files using the information contained in the coming from the standard CMS software framework.

Below lies the LHCO entries corresponding to the event displayed on figure 5.1. The very first line is a label describing the different columns. The line labelled 0 describes the run and event number of the event. Lines labelled 1 to 4 correspond to the 4 selected jets kinematics and b-tagging information (type = 4). The line labelled 5 corresponds to the electron (type =1) kinematics and the

btag column contains the jet index which is the closest to the electron. The last line corresponds to the neutrino (type =6).

```
#       type    eta     phi     pt      jmass   ntrk    btag    had/em  ev_weight       hel
0       190708.61798435 0
1       4       1.224   2.186   98.36   7.42    8.0     2       0.00    1.00            0
2       4       1.473   3.554   56.33   14.04   17.0    0       0.00    1.00            0
3       4       1.953   6.023   42.92   8.80    7.0     2       0.00    1.00            0
4       4       0.812   0.860   39.38   6.62    7.0     0       0.00    1.00            0
5       1       1.304   4.784   64.01   0.04    1.0     2       0.00    1.00            0
6       6       0.000   6.125   53.50   0.00    1.0     0       0.00    1.00            0
```

In addition, an example of user-defined parameters set used during the MADWEIGHT integration can be found in appendix A2.

Chapter 7

Transfer Functions

Transfer functions are crucial elements for the likelihood calculation. They map expectations from the theory model to objects measured in the detector. In MADWEIGHT, transfer functions are considered as a product of individual resolution functions associated to each type of particle.

Transfer functions are evaluated as probability densities for deviation of the measured quantities from their true value. For this analysis, it is assumed that objects are well reconstructed in the (η, ϕ) plane and that their mass remains unchanged. The resolution functions only have an impact on particle momentum reconstruction. As the detector measures the energy deposited by particles, a parametrisation based on $\Delta E = E_{parton} - E_{reco}$, as the difference in energy between the generated and the reconstructed objects, is chosen. In this chapter, the method to determine jet and electron transfer functions are detailed.

7.1 Jet Transfer Functions

In this analysis, transfer functions have been determined by Volker Adler in the frame of his post-doctoral research. The method used for the determination of jet transfer functions is described here.

Jet transfer functions are expressed as the sum of two Gaussian distribu-

tions, following a similar approach as in previous DØ measurement [80]:

$$W(E_{parton}, E_{reco}) = \frac{1}{(p_2 + p_3 p_5)\sqrt{2\pi}} \left(e^{\frac{-(\Delta E - p_1)^2}{2p_2^2}} + p_3 e^{\frac{-(\Delta E - p_4)^2}{2p_5^2}} \right) \qquad (7.1)$$

This transfer function analytic form ensures $\int W(x, y)dx = 1$. (p_1, p_4) are the two means of the two Gaussian functions and (p_2, p_5) their standard deviations. p_3 is the mixing term, defining the amplitude of the second Gaussian with respect to the first. Parameters p_i are considered to be phase-space-dependent and therefore are evaluated for different bins in E_{parton} and $|\eta_{parton}|$, the energy and pseudo-rapidity of the parton-level particle.

To achieve parameters estimation, the selection presented in chapter 5 is applied on simulated events, but without the criterion based on the kinematic fitter output and with a looser cut on the number of jets ($n_{jets} \geq 4$). On top of the selection, partons and jets are required to be matched unambiguously. It means a parton will be matched to a jet, only if exactly one jet can be found in a ΔR cone of radius 0.3. ΔE distributions are then fitted by a double Gaussian function defined in equation 7.1. It must be noted that due to differences between the event selection used to derive transfer functions parameters and the event selection used in the event analysis, a bias on the extracted top quark mass is introduced and will be corrected at calibration stage.

ΔE distributions are fitted in each of the $(E_{parton}, |\eta_{parton}|)$ slices by the double Gaussian function introduced in this section. Figure 7.1 shows an example of the ΔE distribution from simulation fitted to a double Gaussian. b-jets and light jets transfer functions are treated independently. Figures 7.2 and 7.3 show the fit results for the considered $(E_{parton}, \eta_{parton})$ bins. Fit parameters are extracted for each phase-space's sub-region. Appendix 5 shows explicitly the values for all the different jet transfer functions parameters. Once the parameters are estimated, they are handed on in MADWEIGHT allowing realistic transfer functions to be used in the likelihood calculations.
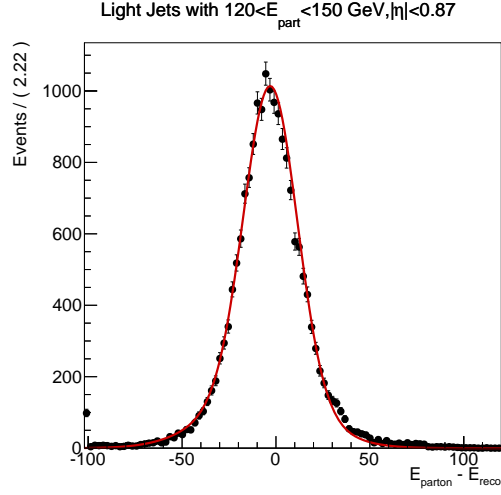
Figure 7.1: Example of $\Delta E = E_{parton} - E_{reco}$ fitted to a double Gaussian.
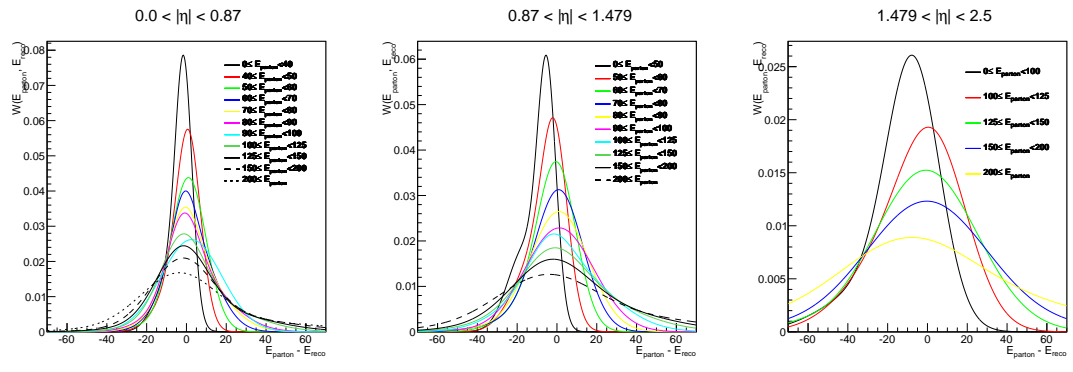
Figure 7.2: Transfer functions for central (left), intermediate (centre) and forward (right) b-jets.

7.2 Electron Transfer Functions

The approach presented in the frame of jet transfer functions (section 7.1) is repeated for electrons objects. A double Gaussian (see equation 7.1) is fitted to the $\Delta E = E_{parton} - E_{reco}$ distributions where ΔE is the energy difference between the generated and the reconstructed electron. Electron pseudo-rapidity and azimuthal angle are assumed to be well-reconstructed.
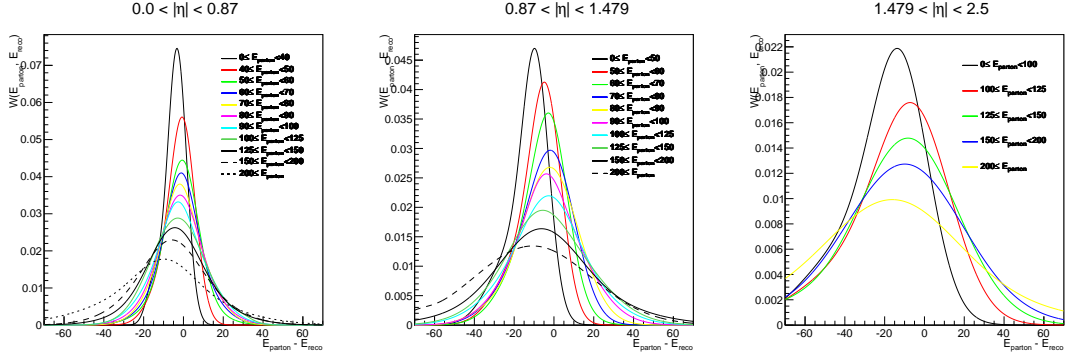
Figure 7.3: Transfer functions for central (left), intermediate (centre) and forward (right) light jets.

Similarly to jet transfer functions, electron transfer functions parameters are estimated for different $(E_{parton}, \eta_{parton})$ slices. Figure 7.4 shows the transfer functions for electrons, and their parameters are summarised in appendix 5.
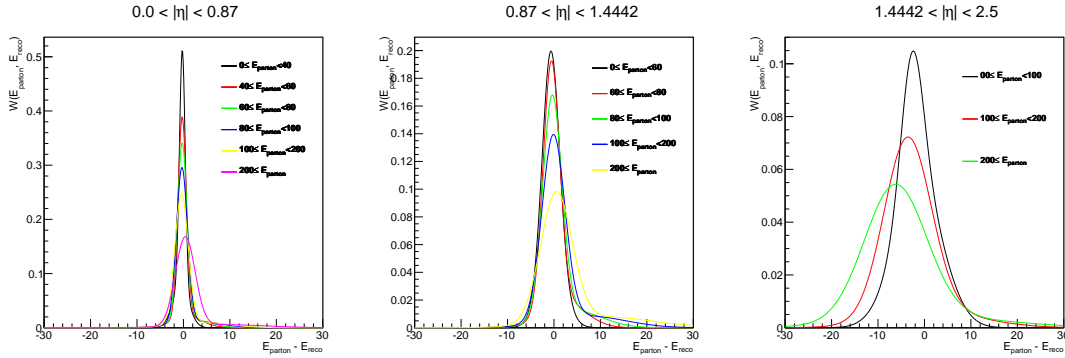
Figure 7.4: Transfer functions for central (left), intermediate (centre) and forward (right) electrons.

Electron transfer functions are not used in MADWEIGHT, but are used in validation studies to estimate the impact it can have on the top quark mass extraction when they are neglected during likelihood calculation.

Chapter 8

Ensemble Testing

The Matrix Element Method is introduced and described in chapter 6. In this chapter, its application to a top quark mass measurement is detailed. The concepts behind ensemble tests as well as the method to extract the measured quantities are explained.

8.1 Pseudo-tests

As explained in chapter 6, the estimation of the sample likelihood for a given set of events yields the measurement on the top quark mass. When performed on simulated events, pseudo-tests (or pseudo-experiments) are constructed by randomly picking events from a large pool. The procedure is detailed step-by-step.

First, $\mathscr{L}_{evt}(x; \mathbf{s})$ is evaluated for all the events within the given sample, generated with a specific value for the top quark mass. Those events constitute the pool. Then, a pseudo-experiment is constructed by randomly picking events within the pool, based on their individual weights. Events with a larger weight have a higher probability to be selected in the ensemble. The likelihood of a single event is not representative of the final measurement, as seen on figure 8.1. Grouping event likelihood by multiplying them allows to suppress the tails and tend to symmetrise the parabola around the most favoured top quark mass value. At the

end of this stage the sample likelihood is obtained. $\mathcal{L}_s(x; \mathbf{s})$. The parameters of the parabolic fit are stored, so their statistical study can be performed in later stages.
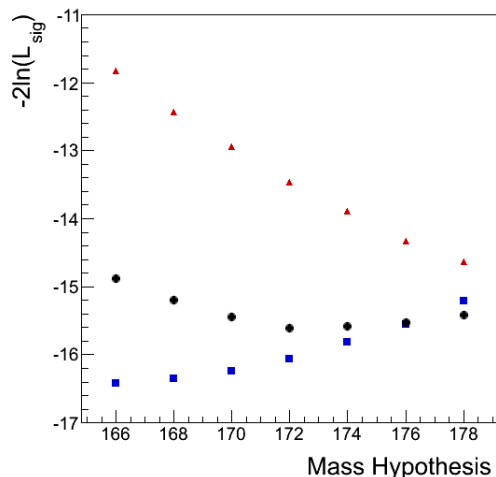
Figure 8.1: Likelihood values for three different generator-level events (each one with a different colour) generated with $m_t = 172.5\,\text{GeV}$

The first step is repeated n_{pE} times, allowing the study the parabolic fit parameters' distributions. The measured top quark mass (abscissa of the minimum of the parabola) and its uncertainty (width of the parabola) are assumed to be Gaussian distributed. The pull (g) is evaluated for each pseudo-experiment and its distribution [81] is constructed as:

$$g_i = \frac{x_i - \bar{x}}{\sigma_i} \tag{8.1}$$

where x_i and σ_i are respectively the measured top quark mass and its uncertainty for a given pseudo-experiment and $\bar{x}$ is the mean of the measured top quark mass distribution across all the pseudo-experiments. The pull distribution is compared to the ideal case where the pull distribution is Gaussian distributed, centred around $\bar{g} = 0$ and has a width $\omega = 1$. The width of the pull distribution represents the agreement between the fitted uncertainties and the uncertainty on the measure-

ment.

The two first steps are repeated for samples generated with different input top quark mass values, so the method behaviour can be checked across a wide range of hypotheses as detailed in chapters 9 and 10

8.2 Fitting Procedure and Top Quark Mass Extraction

8.2.1 Resampling Corrections

The present analysis maximises the likelihood information available through resampling. Indeed, likelihood estimation is computationally intensive and therefore performed on limited number of events. When constructing a given pseudo-experiment, events can be picked even if they are already present in a previous pseudo-experiment. Therefore resampling corrections must be considered when estimating statistical uncertainty for a given sample [82]. The corrected uncertainties on the quark top mass and on the pull width are corrected as

$$
\sigma'_{m_t} = \frac{\sigma_{m_t}}{\sqrt{n_{ind}}}
$$
$$
\sigma'_{\omega} = \omega \sqrt{\frac{1}{2} \left(\frac{1}{n_{pool}} + \frac{1}{n_{ens} - 1} \right)}
$$

(8.2)

where n_{ens} and n_{pool} are respectively the number of built pseudo-experiments and the total number of events in the pool. The number of independent ensembles is defined as the ratio between the number of events in the pool and the number of events per pseudo-experiment: $n_{ind} = \frac{n_{pool}}{n_{pE}}$.

8.2.2 Top Quark Mass Extraction

Following the equations presented in chapter 6, the negative logarithm of the normalised sample likelihood (see eq. 6.14) for a top quark mass measurement

for a set of reconstructed events $\mathbf{x}$ is expressed as

$$-\ln \bar{\mathscr{L}}_s(\mathbf{x}; m_t) = -\ln \mathscr{L}_{MW}(\mathbf{x}; m_t) + n \ln \sigma_{t\bar{t}}^{tot}(m_t) + n \ln \left(\bar{f}_{acc}(m_t)\right) \qquad (8.3)$$

The uncertainty on the normalised sample likelihood $\sigma\left(\bar{\mathscr{L}}_s(\mathbf{x}; m_t)\right)$ is given below, $\mathscr{L}_{MW evt_i}(\mathbf{x}; m_t)$ being the raw likelihood value returned by MADWEIGHT for the event i,

$$\sigma\left(\bar{\mathscr{L}}_s(\mathbf{x}; m_t)\right) = \sigma\left(\mathscr{L}_{MW}(\mathbf{x}; m_t)\right) + n \frac{\sigma\left(\sigma_{t\bar{t}}^{tot}(m_t)\right)}{\sigma_{t\bar{t}}^{tot}(m_t)} + n \frac{\sigma\left(\bar{f}_{acc}(m_t)\right)}{\bar{f}_{acc}(m_t)}$$

$$\Leftrightarrow \sigma\left(\bar{\mathscr{L}}_s(\mathbf{x}; m_t)\right) = \sum_{i=1}^{n_{pE}} \frac{\sigma\left(\mathscr{L}_{MW evt_i}(\mathbf{x}; m_t)\right)}{\mathscr{L}_{MW evt_i}(\mathbf{x}; m_t)} + n \frac{\sigma\left(\sigma_{t\bar{t}}^{tot}(m_t)\right)}{\sigma_{t\bar{t}}^{tot}(m_t)} + n \frac{\sigma\left(\bar{f}_{acc}(m_t)\right)}{\bar{f}_{acc}(m_t)}$$

$$(8.4)$$

The sample likelihood negative logarithm is then fitted by a parabola of the following form:

$$\frac{\left(m_t^{fit} - m_t^{gen}\right)^2}{2\sigma^2(m_t)} + k \qquad (8.5)$$

where $\sigma(m_t)$ is the width of the parabola and corresponds to the uncertainty on the measurement.

Part IV

Top quark mass extraction

Chapter 9

Validation

In this chapter, the validation of the Matrix Element Method applied to a top quark mass measurement is presented. The feasibility of the method is checked, starting with parton level information. The process is then refined to move towards a realistic case. In steps, selection and acceptance term are introduced, jet energies are smeared and accounted for using realistic transfer functions during likelihood integration. A last step consists of smearing the electron energies to reflect the perturbation introduced by the detector, while ignoring it during likelihood calculation, to estimate the bias on the extracted top mass introduced by neglecting electron transfer functions. For every step, the agreement between the input top quark mass value and the top quark mass value returned by the Matrix Element Method is quantified with the help of pull distributions introduced in chapter 8.

9.1   Parton-level Validation

Generator-level samples are generated using MADGRAPH for different values of the top quark mass. They do not include hadronisation or parton showering simulation, nor any reconstruction modelling from the detector. Therefore, transfer functions used for the events integration in MADWEIGHT are modelled as δ-functions. A closure test of the likelihood calculation and its subsequent analysis is possible.

Likelihoods are calculated with MADWEIGHT for five samples generated with processes corresponding to the decay of a top quark pair into an electron-neutrino pair and four quarks, for different values of the top quark mass (169.5, 171.5, 172.5, 173.5, 175.5)GeV. Each sample contains about three thousands events. Likelihood calculation is made for hypothesis spaced by 100 MeV, centred on the top quark mass value used to generate the sample. A total of 500 ensembles is made, for each of the five samples, each ensemble being composed of 200 events. No selection is used for this step of the validation, therefore no acceptance term is required. Figure 9.1 shows the result for a single pseudo-experiment. The negative logarithm of the sample likelihood is plotted for each of the top quark mass hypothesis value, and then fitted by a parabola as discussed in section 8.2. The width of the parabola, its associated uncertainty and the uncertainty on the parabola minimum abscissa are added in quadrature to yield the uncertainty on the top quark mass extraction [83]. The three distributions are depicted on figure 9.2.

Figure 9.3 to 9.7 show the outcome of all the pseudo experiments. Extracted top quark mass distributions are fitted with a Gaussian function. The width of the parabolic fit, the uncertainty on the width and the uncertainty on the minimum of the parabola are added in quadrature to evaluate the uncertainty on the extracted top quark mass. The pull distributions are built as explained in chapter 8. The text in red corresponds to the mean of the distributions (extracted top mass, statistical error and pull mean). The pull width is extracted by considering the root mean square of the pull distribution. Those values are corrected for resampling.

Summary: The results from all the pseudo-experiments are summarised on figure 9.8. An agreement within two standard deviations between the extracted and generated top mass can be observed. The pull width however, is not compatible with the unity (0.92 ± 0.01). It represents an over-estimation of the uncertainty on the extracted top quark mass and could possibly be explained by the conservative approach considered to evaluate the uncertainty on the measurement *i.e.* by adding in quadrature the three error components. As a cross-check to look for any pull width dependency on the generated top quark mass, a linear fit has been
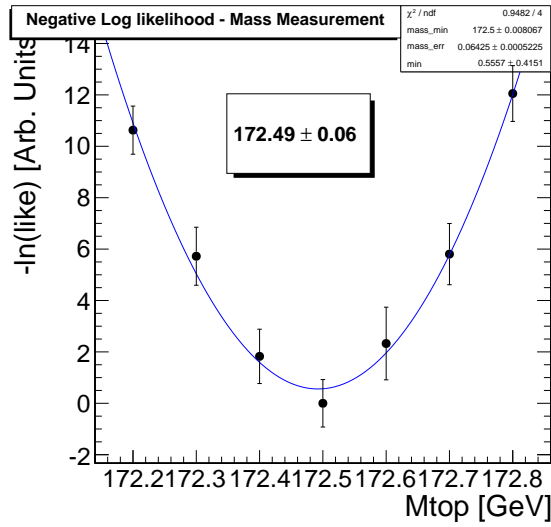
Figure 9.1: Negative logarithm of a single pseudo-experiments. Events have been generated with a top quark mass of 172.5 GeV. A parabolic fit has been applied to the distribution, and its minimum corresponds to the extracted top mass for this pseudo-experiment. The width parameter of the parabola corresponds to the uncertainty on the extracted top quark mass for that given pseudo-experiment
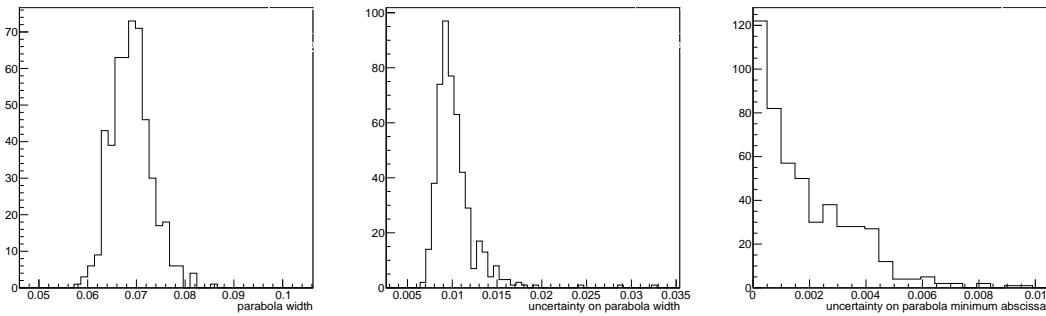
Figure 9.2: The three components of the uncertainty on the top quark mass extraction are shown. The parabolic fit width (left), its associated uncertainty (centre) and the uncertainty on the parabola minimum abscissa (right). The three components are added in quadrature.

performed. It shows a slope compatible with zero within two standard deviations, as can be seen on figure 9.9.
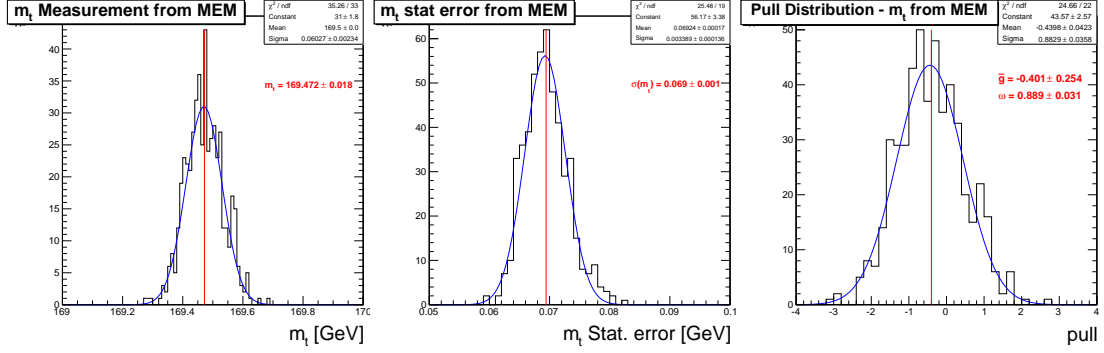
Figure 9.3: Top quark mass (left), error on the extracted top quark mass (centre) and pull distributions (right) for pseudo-experiments performed on the sample generated with $m_t = 169.5\,\text{GeV}$
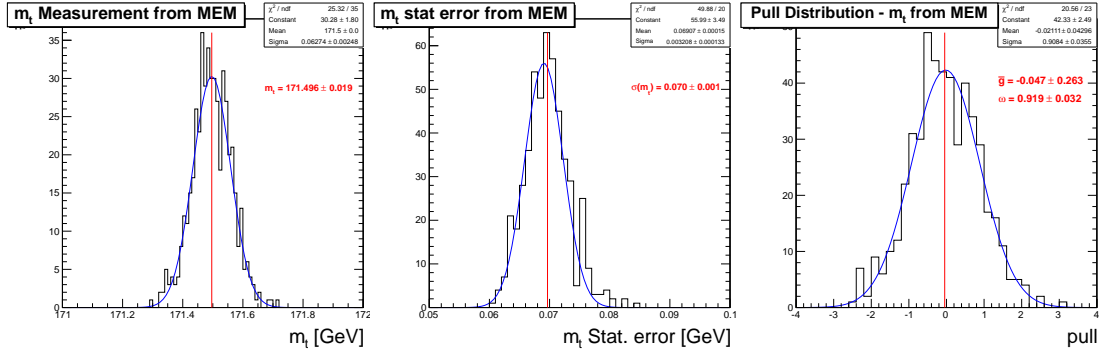
Figure 9.4: Top quark mass (left), error on the extracted top quark mass (centre) and pull distributions (right) for pseudo-experiments performed on the sample generated with $m_t = 171.5\,\text{GeV}$

9.2 Jet Energy Smearing

The parton-level samples generated during the first step of the validation (see section 9.1) are now smeared to reflect the use of realistic jet transfer functions. The jet transfer functions presented in section 7.1 are used as probability density functions to smear the partons energies. On top of the smearing, selection
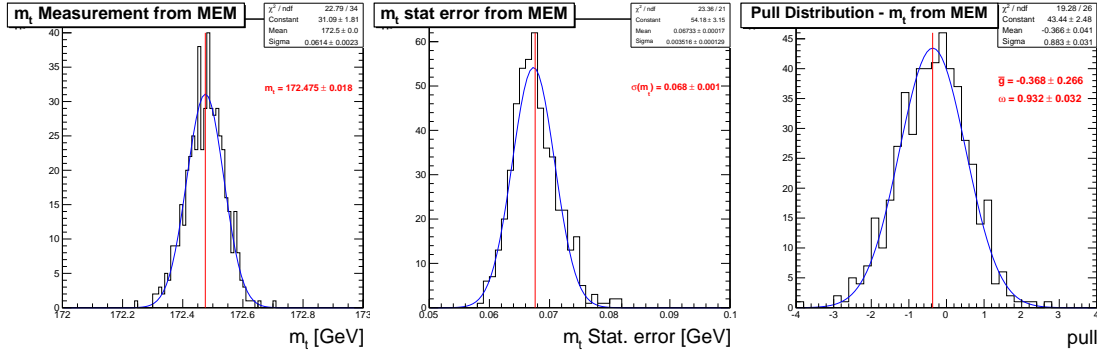
Figure 9.5: Top quark mass (left), error on the extracted top quark mass (centre) and pull distributions (right) for pseudo-experiments performed on the sample generated with $m_t = 172.5\,\text{GeV}$
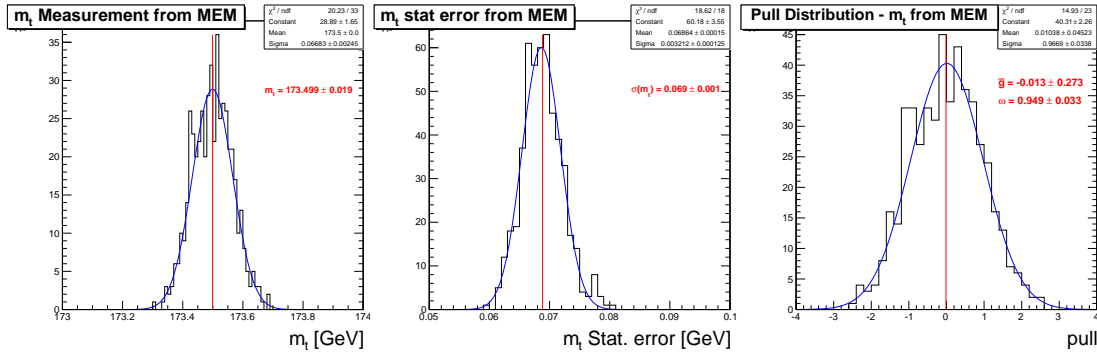
Figure 9.6: Top quark mass (left), error on the extracted top quark mass (centre) and pull distributions (right) for pseudo-experiments performed on the sample generated with $m_t = 173.5\,\text{GeV}$

criteria listed in table 9.2 are applied.

Jet transfer functions and selected events are then handed to MADWEIGHT for likelihood calculation. To account for the uncertainty due to the smearing of parton energies, mass hypothesis are spaced by $1\,\text{GeV}$ and centred on the top mass value used in the generation. Two additional hypothesis are considered outside of this range, spaced by $2\,\text{GeV}$. Practically, while considering the samples generated with a top mass of $172.5\,\text{GeV}$, it results in 11 different mass hypothesis: 165.5, 167.5, 169.5, 170.5, 171.5, 172.5, 173.5, 174.5, 175.5, 177.5, 179.5 GeV. Samples
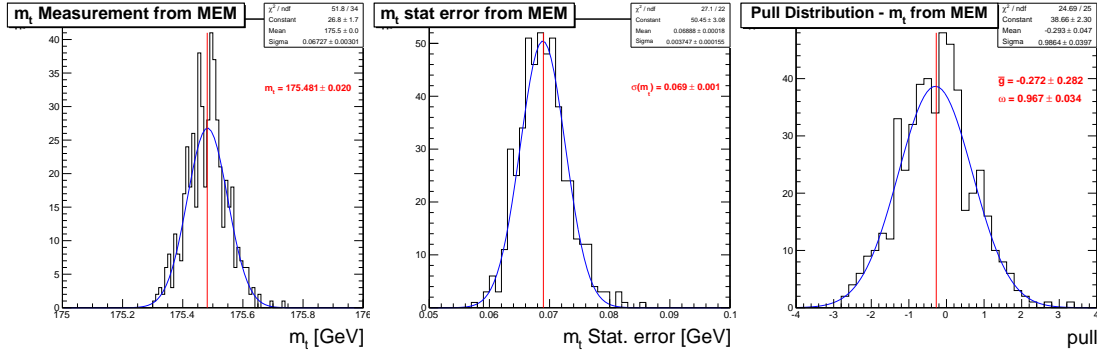
Figure 9.7: Top quark mass (left), error on the extracted top quark mass (centre) and pull distributions (right) for pseudo-experiments performed on the sample generated with $m_t = 175.5\,\text{GeV}$
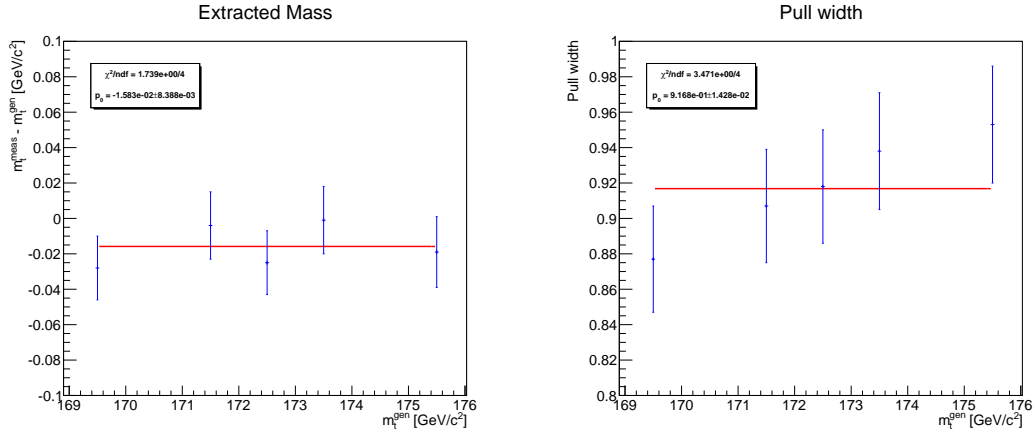
Figure 9.8: Difference between extracted and generated top quark masses (left) and pull widths (right) as a function of the generated top quark mass.

likelihoods are derived on 500 pseudo-experiments, each containing 1000 events. The acceptance curve used for this step of the validation can be seen on figure 9.10. After normalisation, the samples likelihoods are fitted to a parabola and the top mass is extracted for each of the samples. This step is meant to validate the use of realistic transfer function within MADWEIGHT.
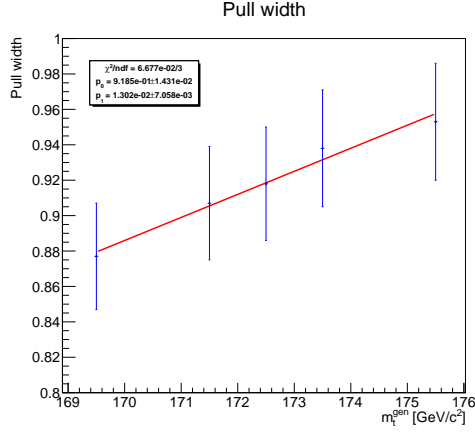
Figure 9.9: As a cross-check, the pull width values from figure 9.8 are fitted to a first order polynomial. The slope (p_1) is compatible with zero within two standard deviations.

Object	Kinematic selection
electron	$p_T > 30\,\mathrm{GeV}$
	$|\eta| < 2.5$
jet	$p_T > 30\,\mathrm{GeV}$
	$|\eta| < 2.4$

Table 9.1: Kinematic selection criteria applied on parton-level objects during validation

Summary: The extracted top quark mass is in general in agreement within one standard deviation with the value used during the generation. However, the sample generated with $173.5\,\mathrm{GeV}$ shows an agreement within 1.9 standard deviation. The uncertainty on the extracted top quark mass is stable across all samples. The pull distribution widths are slightly large than one, showing an average value of 1.07 as can be observed on figure 9.16.

105
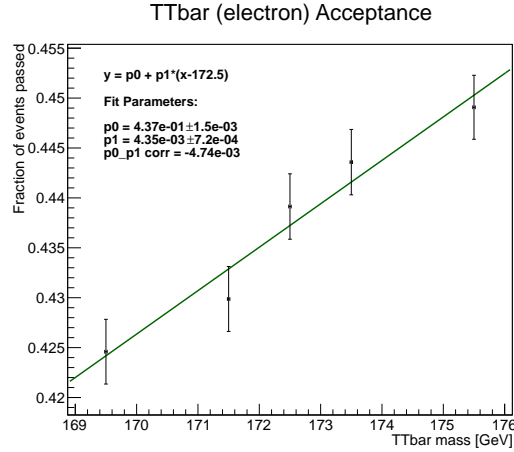
Figure 9.10: Acceptance curve used in the validation of the use of realistic jet transfer functions.
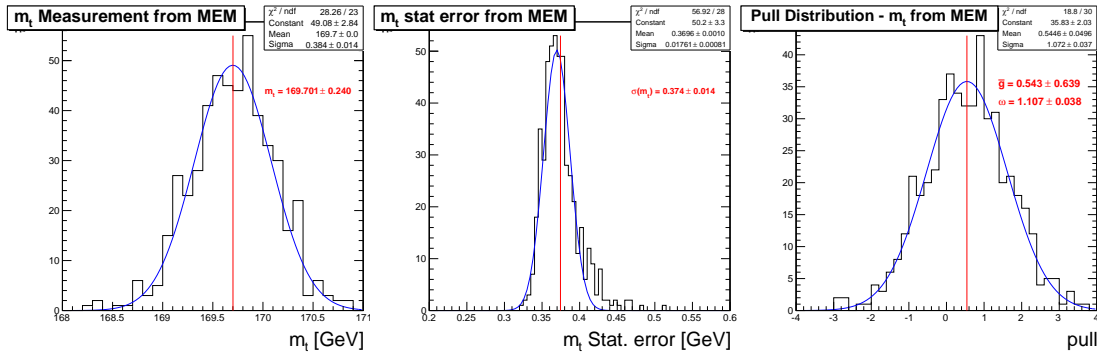
Figure 9.11: Top quark mass (left) and pull distributions (right) for pseudo-experiments performed on the sample generated with $m_t = 169.5\,\text{GeV}$ where the jet energies have been smeared according to the jet transfer functions.

9.3 Electron Energy Smearing

Once the jets are smeared and the use of jet transfer function is validated, the impact of neglecting electron transfer function is scrutinised. Similarly to the jet smearing, electron energies are smeared using the transfer functions presented in section 7.2. During likelihood calculation, electron transfer function are still defined as δ-functions within MADWEIGHT. The result on the top quark mass is compared to the case where only jets are smeared. Consequently, MADWEIGHT
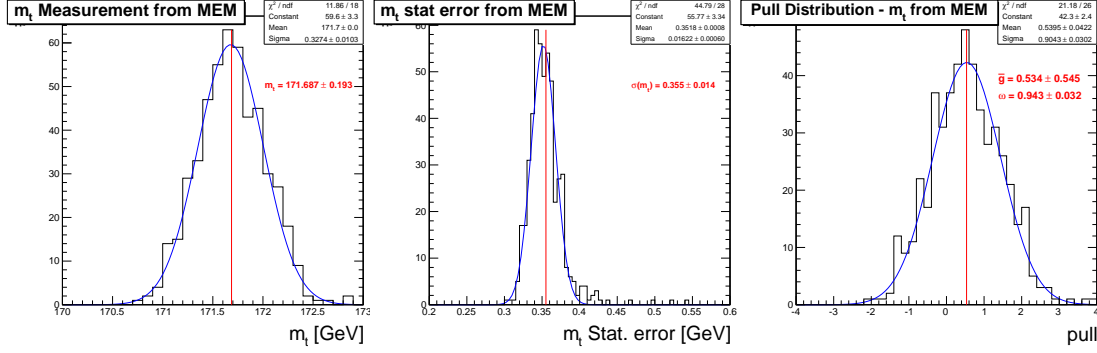
Figure 9.12: Top quark mass (left) and pull distributions (right) for pseudo-experiments performed on the sample generated with $m_t = 171.5\,\text{GeV}$ where the jet energies have been smeared according to the jet transfer functions.
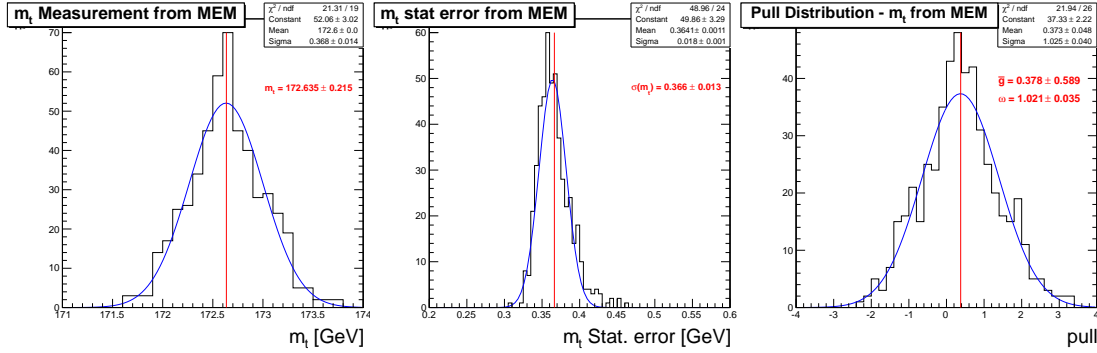
Figure 9.13: Top quark mass (left) and pull distributions (right) for pseudo-experiments performed on the sample generated with $m_t = 172.5\,\text{GeV}$ where the jet energies have been smeared according to the jet transfer functions.

configuration for this validation step is exactly the same as in the validation of the use of jet transfer functions. Figures 9.18 to 9.22 show the results of the ensemble tests for all the different samples.

Summary: Figure 9.23 shows that the results from this step present a 2.5 standard deviation on the top quark mass with respect to the input values, resulting from the choice of neglecting electron transfer functions when calculating likelihoods. However this bias does not show a top quark mass dependence. It will
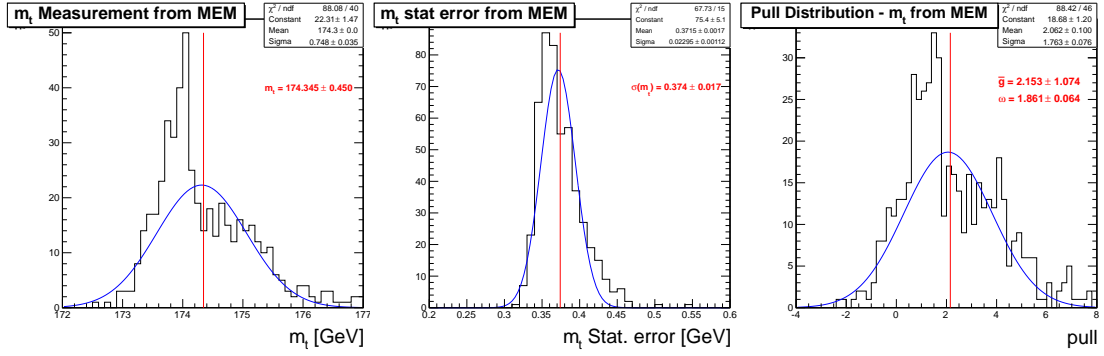
Figure 9.14: Top quark mass (left) and pull distributions (right) for pseudo-experiments performed on the sample generated with $m_t = 173.5\,\text{GeV}$ where the jet energies have been smeared according to the jet transfer functions.
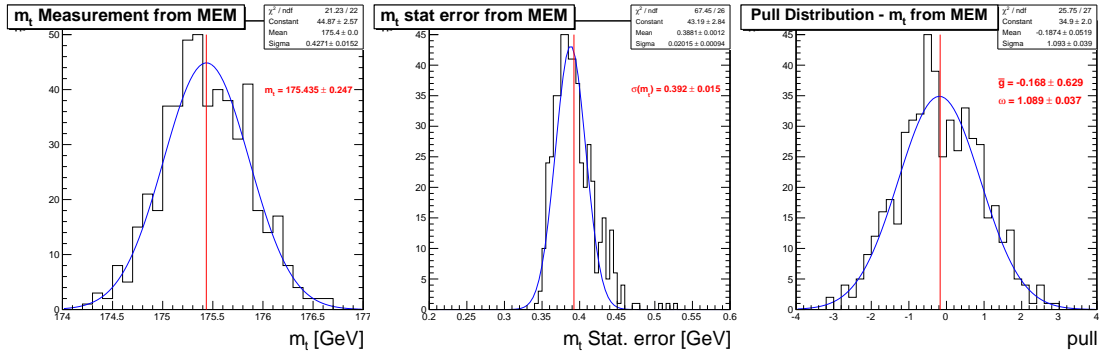
Figure 9.15: Top quark mass (left) and pull distributions (right) for pseudo-experiments performed on the sample generated with $m_t = 175.5\,\text{GeV}$ where the jet energies have been smeared according to the jet transfer functions.

be absorbed in the calibration curve when running the method on fully simulated events.
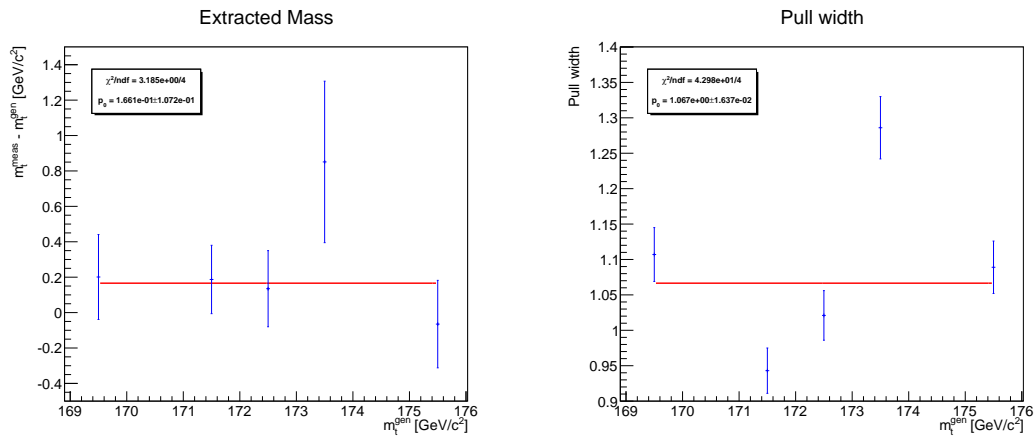
Figure 9.16: Difference between extracted and generated top quark mass (left) and pull width (right) as a function of the generated top quark mass for samples where the jet energies have been smeared.
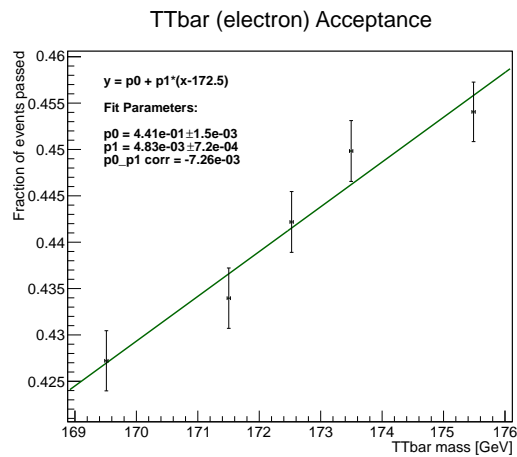
Figure 9.17: Acceptance curve used for samples where jets and electrons energies have been smeared.
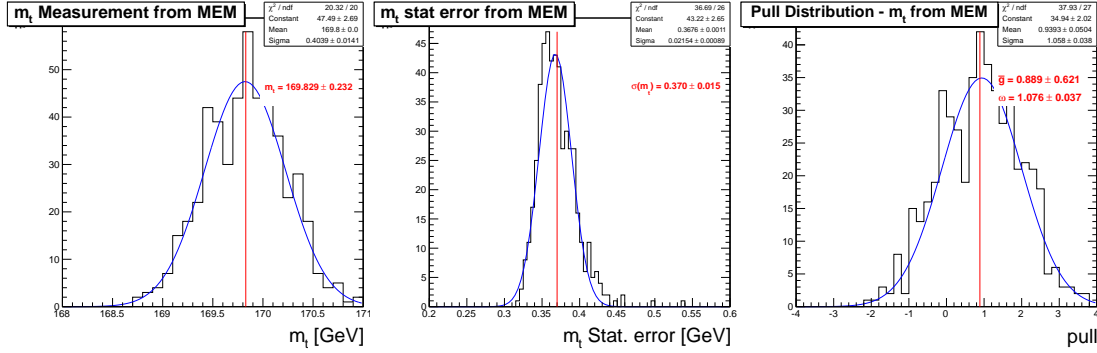
Figure 9.18: Top quark mass (left) and pull distributions (right) for pseudo-experiments performed on the sample generated with $m_t = 169.5 \, \text{GeV}$ where the jet and electron energies have been smeared.
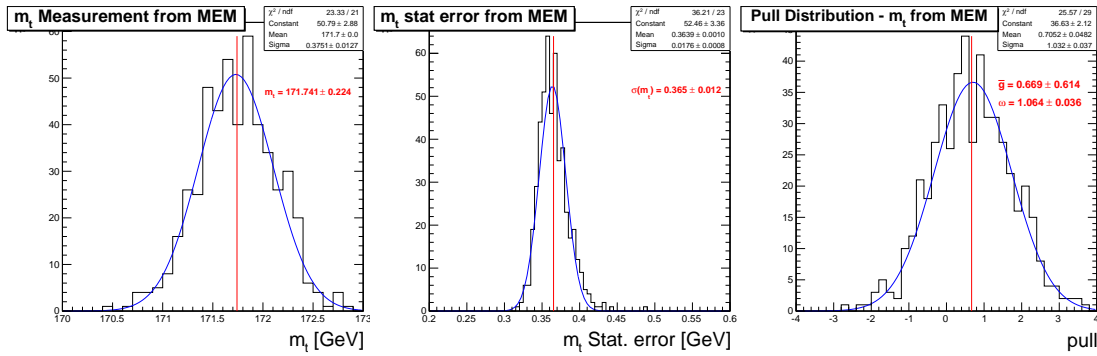
Figure 9.19: Top quark mass (left) and pull distributions (right) for pseudo-experiments performed on the sample generated with $m_t = 171.5 \, \text{GeV}$ where the jet and electron energies have been smeared.
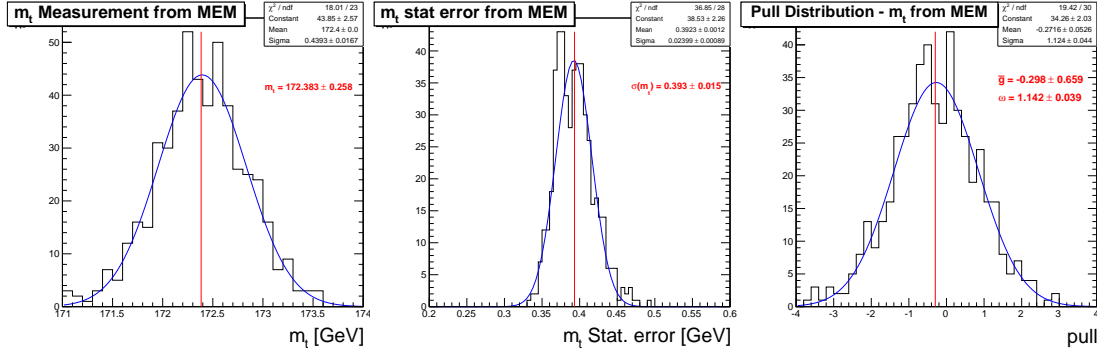
Figure 9.20: Top quark mass (left) and pull distributions (right) for pseudo-experiments performed on the sample generated with $m_t = 172.5\,\text{GeV}$ where the jet and electron energies have been smeared.
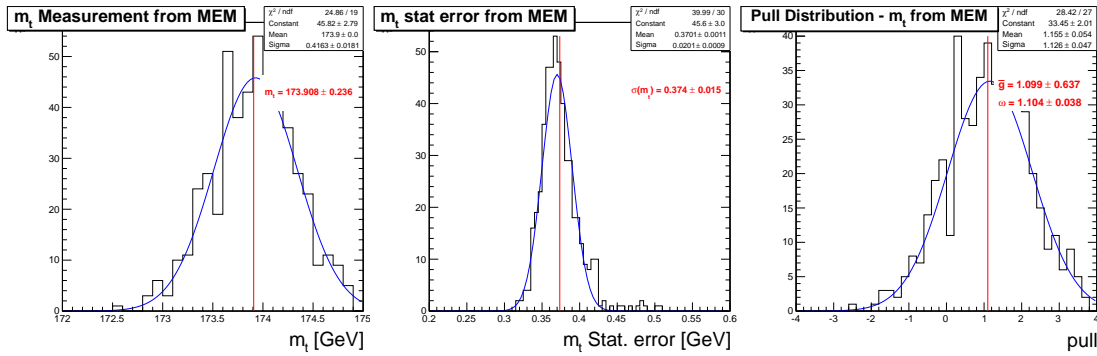
Figure 9.21: Top quark mass (left) and pull distributions (right) for pseudo-experiments performed on the sample generated with $m_t = 173.5\,\text{GeV}$ where the jet and electron energies have been smeared.
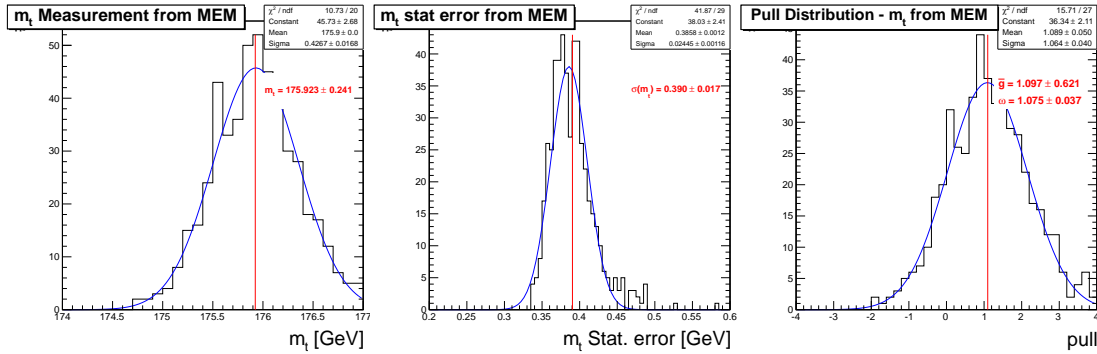
Figure 9.22: Top quark mass (left) and pull distributions (right) for pseudo-experiments performed on the sample generated with $m_t = 175.5 \,\mathrm{GeV}$ where the jet and electron energies have been smeared.
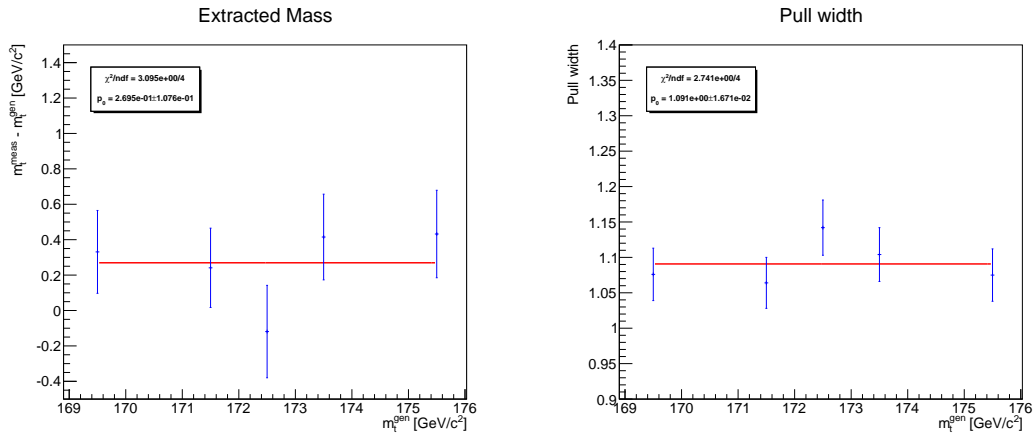
Figure 9.23: Difference between extracted and generated top quark mass (left) and pull width (right) as a function of the generated top quark mass for samples where the jet and electron energies have been smeared.

Chapter 10

Calibration

The calibration is a necessary step to account and correct for all the discrepancies introduced by the various assumptions made in the previous steps. Indeed biases have been introduced, for example, by considering only signal likelihoods, neglecting electron transfer functions and using binned jet transfer functions assumed to have a double Gaussian functional form.

For the calibration process seven different samples are used. They consist of fully simulated Monte Carlo samples, generated with different top quark mass values: $m_t = 166.5, 169.5, 171.5, 172.5, 173.5, 175.5, 178.5$ GeV. For each sample, 2000 ensembles are constructed, containing each 8898 events, corresponding to the number of events selected in recorded data over the 2012 data taking period. The likelihoods are calculated for hypothesis intervals of $4\,\mathrm{GeV}$. Likelihoods are normalised for the acceptance curve shown on figure 10.1. Top quark mass is extracted for each of the samples and the difference between the extracted top mass and the top mass value used in the generation of the samples. Calibration points are fitted to a linear curve of the form $f(m_t^{meas} - 172.5) = p_0 + p_1(m_t^{meas} - 172.5)$.

Two calibration curves are presented in this chapter. A first calibration curve is derived from ensembles entirely made of $t\bar{t}$ events, the fraction of signal events is set as the fraction expected from table 5.5, the remainder is made of $t\bar{t} \rightarrow$ other events. This first calibration curve is shown on figure 10.2. As the estimation

of systematic uncertainties is performed using a mixture of $t\bar{t}$ events (see 12), this calibration curve is used to estimate the top quark mass nominal value.
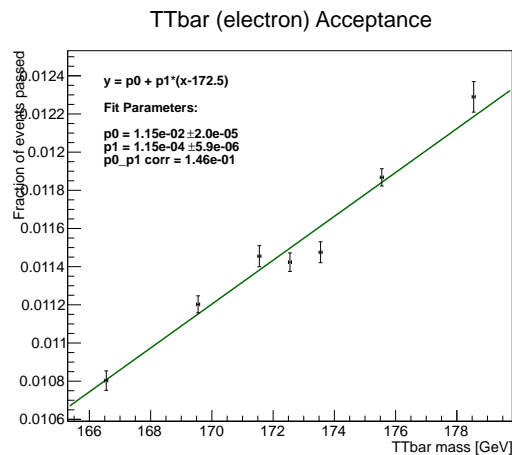
Figure 10.1: Acceptance curve for fully simulated Monte Carlo events. The fraction of $t\bar{t} \to e+$jets accepted events is evaluated for each of the input samples.
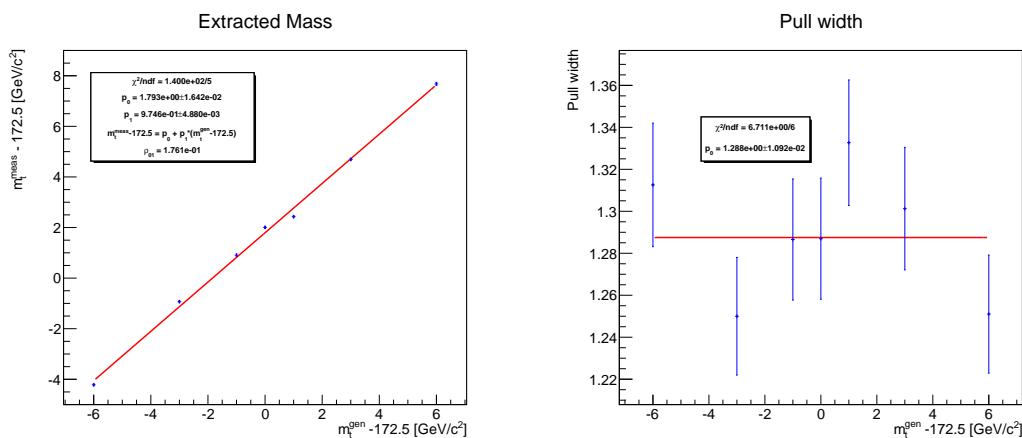
Figure 10.2: Calibration curve showing the extracted top mass (left) and the associated pull distributions width for pseudo-experiments containing a mix of $t\bar{t} \to e+$jets and $t\bar{t} \to$ other events.

A second calibration curve is derived from ensembles made of $t\bar{t}$ mixed with background events from the $W+$jets and $Z+$jets backgrounds. They are mixed

according to the proportions given by table 5.5. Due to the lack of mass dependent single-top samples, the effect of including single-top events on the measurement is evaluated at 172.5 GeV. The bias introduced by including single-top events in pseudo-experiments induces a bias of 134 MeV on the extracted top quark mass (see figure 10.3). This impact on the top quark mass is then propagated to the calibration curve. The resulting calibration curve is shown on figure 10.4 and is used in chapter 11 to calibrate the result obtained with recorded data events. Calibration points contributing to the final calibration curve are shown in figures 10.5 to 10.11.
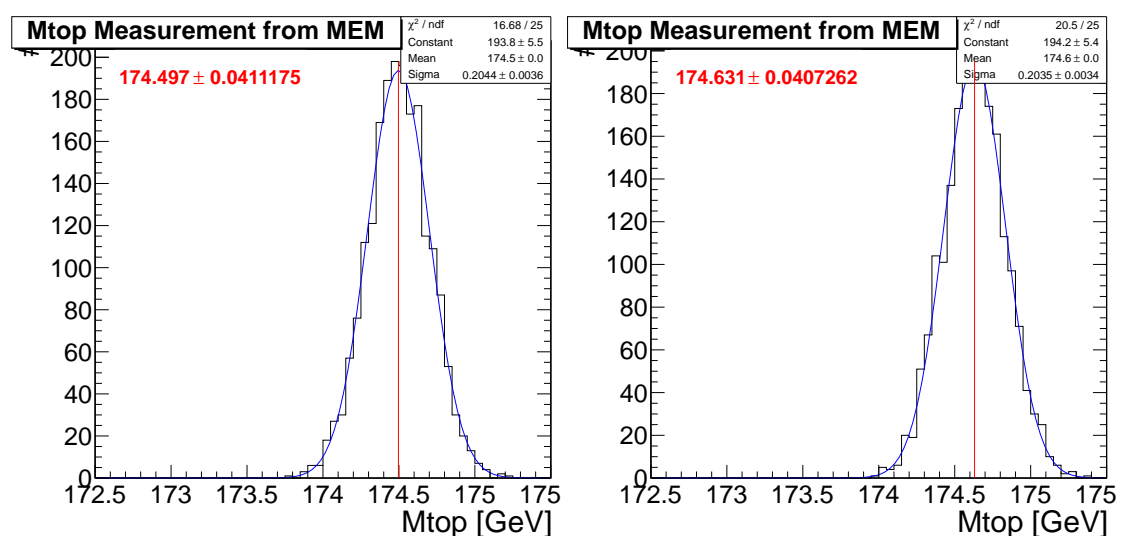
Figure 10.3: Top quark mass distributions for pseudo-experiments containing a mixture of $t\bar{t}$ and non $t\bar{t}$ events events. The left plot does not include contribution from single-top events, while the right one does. The shift in the resulting top quark mass is propagated to the calibration curve.

As shown on the calibration curve, a significant bias on the extracted to quark mass of 1.923 GeV needs to be accounted for. The bias can be explained by two major facts. As explained in chapter 7, the selection used in the derivation of jet transfer functions is slightly different from the selection used by the analysis. MADWEIGHT integration is very sensitive to the parton energies, and an eventual mis-modelling of the detector response leads to biased parton-level kinematics estimation. The bias could be reduced by deriving jet transfer functions on events
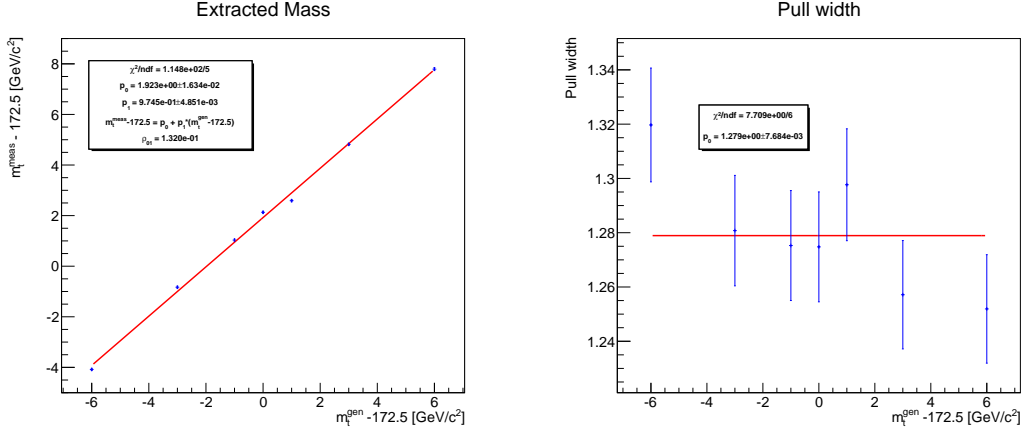
Figure 10.4: Calibration curve showing the extracted top mass (left) and the associated pull distributions width for pseudo-experiments containing a mix of $t\bar{t}$ and non $t\bar{t}$ events events. The shift due to the consideration of single-top events has been propagated.

where the analysis selection criteria has been applied. Another possible bias source can be explained by the choice of neglecting electron transfer functions. This choice has been made in order to reduce the computation time. The fitted average pull width ($\bar{g} = 1.28$) implies that the uncertainty on the top quark mass measurement is underestimated by 28%.
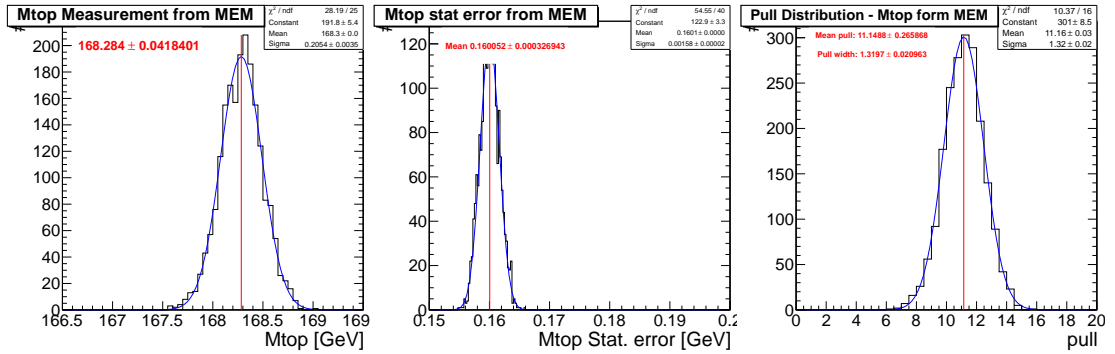
Figure 10.5: Top quark mass (left), uncertainty (centre) and pull (right) distributions for pseudo-experiments containing a mixture of $t\bar{t}$ and non-$t\bar{t}$ events. Background contribution from single-top is not included in those ensembles. $t\bar{t}$ events have been generated for $m_t = 166.5\,\mathrm{GeV}$
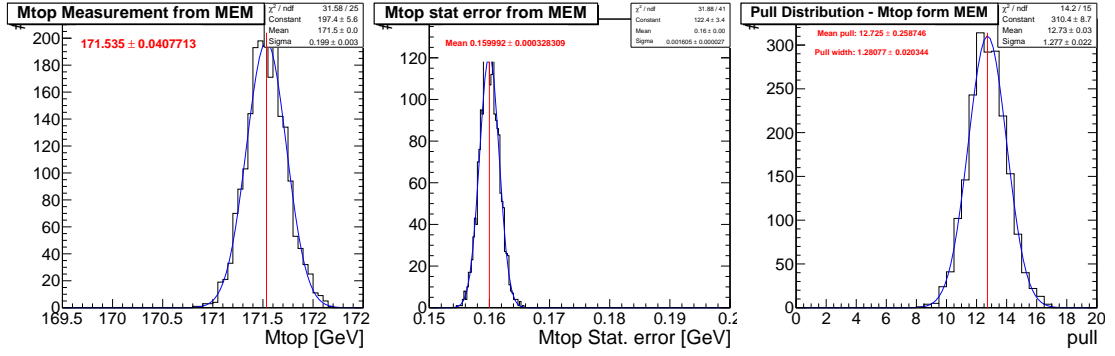
Figure 10.6: Top quark mass (left), uncertainty (centre) and pull (right) distributions for pseudo-experiments containing a mixture of $t\bar{t}$ and non $t\bar{t}$ events events. Background contribution from single-top is not included in those ensembles. $t\bar{t}$ events have been generated for $m_t = 169.5\,\text{GeV}$
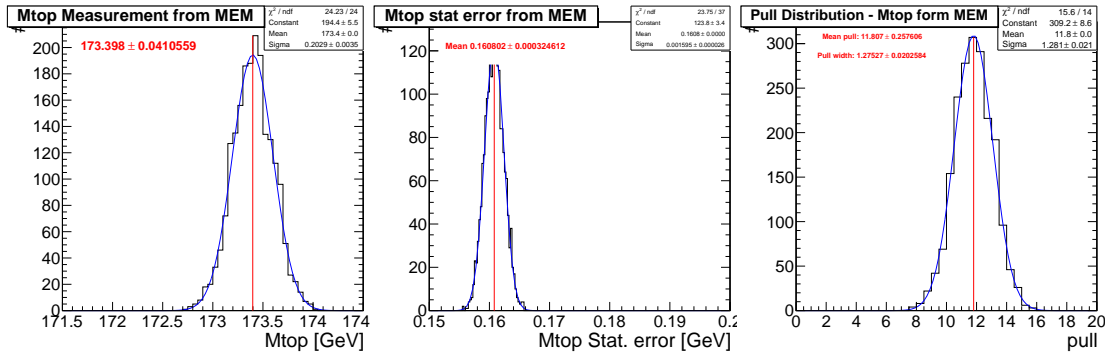
Figure 10.7: Top quark mass (left), uncertainty (centre) and pull (right) distributions for pseudo-experiments containing a mixture of $t\bar{t}$ and non $t\bar{t}$ events events. Background contribution from single-top is not included in those ensembles. $t\bar{t}$ events have been generated for $m_t = 171.5\,\text{GeV}$
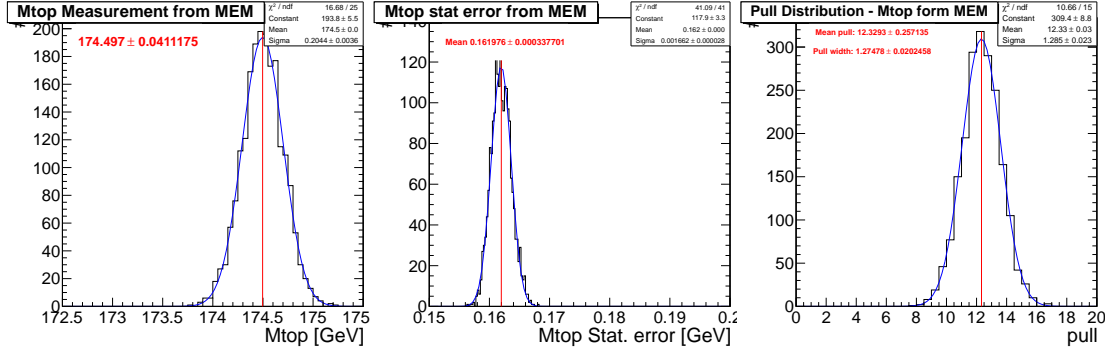
Figure 10.8: Top quark mass (left), uncertainty (centre) and pull (right) distributions for pseudo-experiments containing a mixture of $t\bar{t}$ and non $t\bar{t}$ events events. Background contribution from single-top is not included in those ensembles. $t\bar{t}$ events have been generated for $m_t = 172.5\,\mathrm{GeV}$
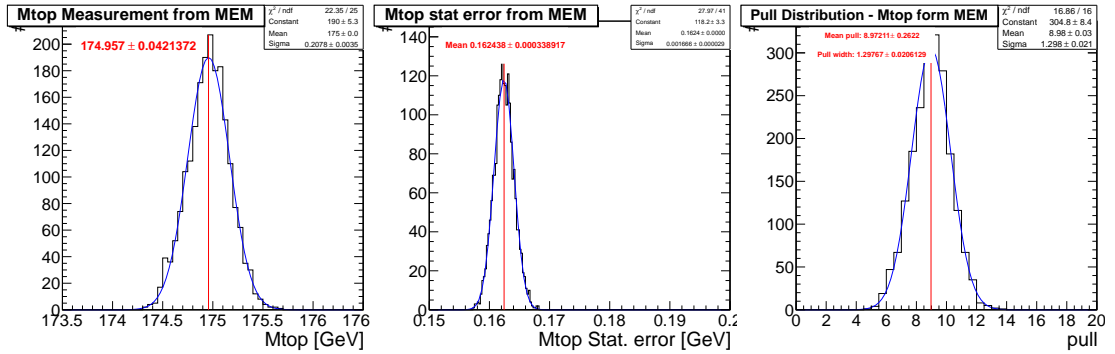
Figure 10.9: Top quark mass (left), uncertainty (centre) and pull (right) distributions for pseudo-experiments containing a mixture of $t\bar{t}$ and non $t\bar{t}$ events events. Background contribution from single-top is not included in those ensembles. $t\bar{t}$ events have been generated for $m_t = 173.5\,\mathrm{GeV}$
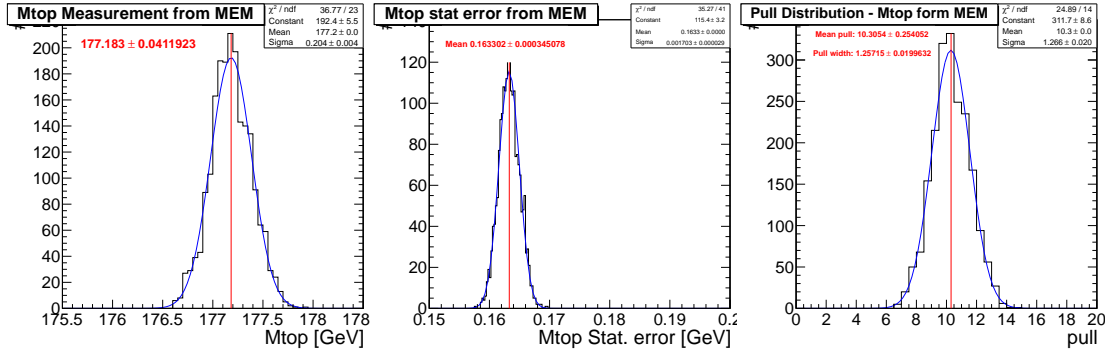
Figure 10.10: Top quark mass (left), uncertainty (centre) and pull (right) distributions for pseudo-experiments containing a mixture of $t\bar{t}$ and non $t\bar{t}$ events events. Background contribution from single-top is not included in those ensembles. $t\bar{t}$ events have been generated for $m_t = 175.5\,\text{GeV}$
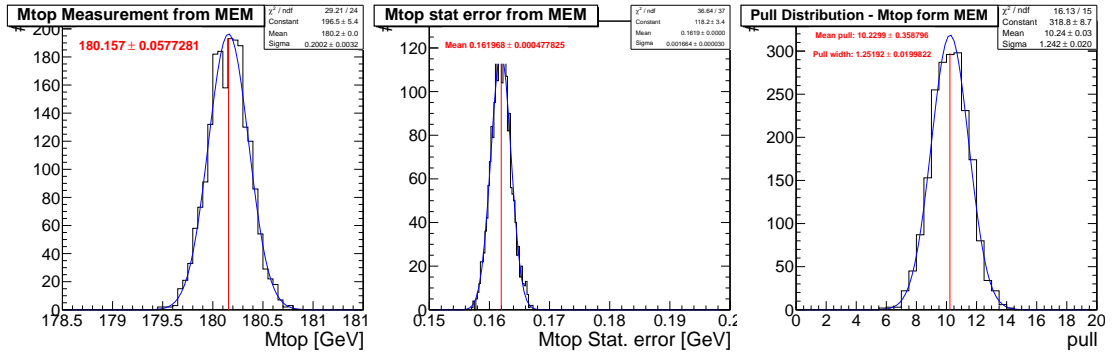
Figure 10.11: Top quark mass (left), uncertainty (centre) and pull (right) distributions for pseudo-experiments containing a mixture of $t\bar{t}$ and non $t\bar{t}$ events events. Background contribution from single-top is not included in those ensembles. $t\bar{t}$ events have been generated for $m_t = 178.5\,\text{GeV}$

119

Chapter 11

Mass Extraction

The actual measurement is performed on the data ensemble comprising the 8898 events selected from the data recorded by the detector during the year 2012. In figure 11.1 for each top quark mass hypothesis, the raw likelihood distributions from those event are compared to the ones obtained on $t\bar{t} \to e$+jets events from the simulated sample generated with $m_t = 172.5\,\text{GeV}$. The distributions show a good agreement. The result from the data events ensemble is shown of figure 11.2. Likelihoods have been normalised to the acceptance shown on figure 10.1. The parabolic fit behaves well and properly describe the data points within their uncertainties. The top mass extraction yields:

$$m_t = 172.91 \pm 0.17\,\text{GeV} \tag{11.1}$$

After calibration, as shown on figure 11.3 the resulting top quark mass is $m_t^{cal} = 170.944 \pm 0.175\,\text{GeV}$. The uncertainty on the measurement needs to be corrected by 28% to account for the average pull width shown on figure 10.4, yielding a final result of:

$$m_t^{corr} = 170.94 \pm 0.22\,\text{GeV} \tag{11.2}$$
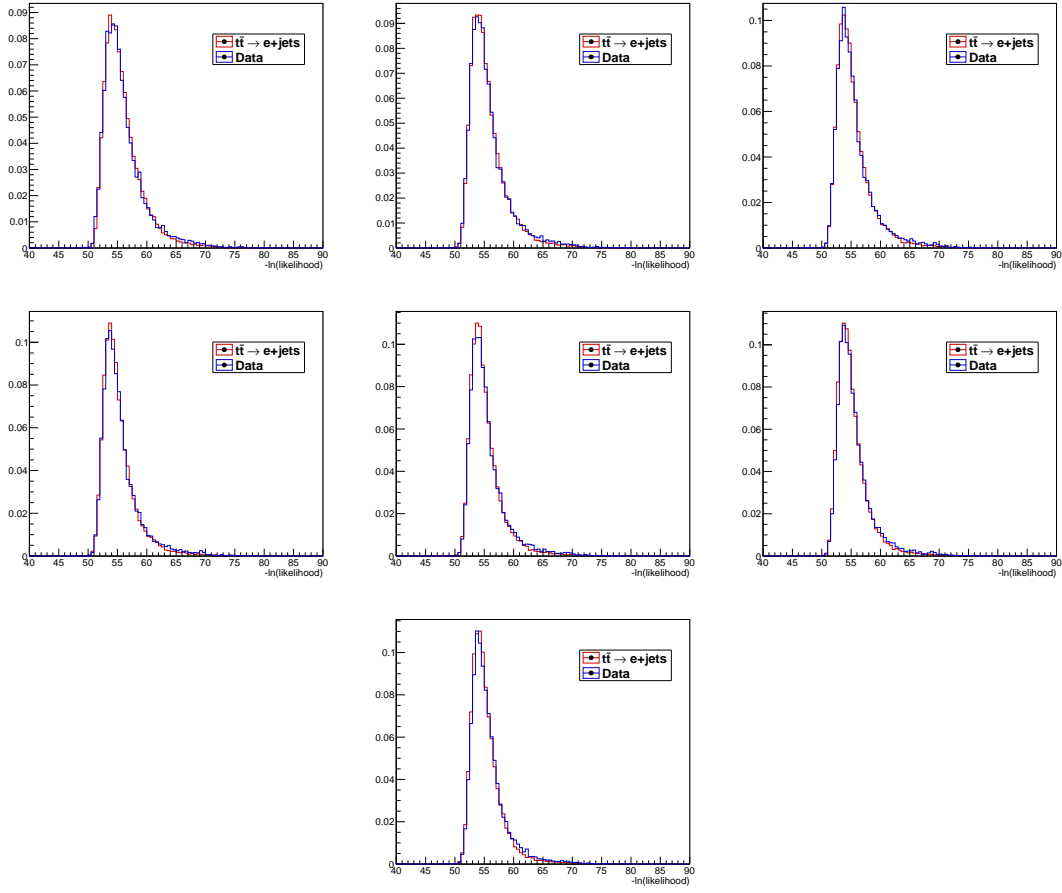
Figure 11.1: Comparison of the likelihoods obtained with selected events from recorded data and events from the simulation sample generated with $m_t = 172.5\,\text{GeV}$. The comparison is made for each of the seven mass hypotheses, from left to right and up to down: 162, 166, 170, 174, 178, 182 and 186 GeV

121
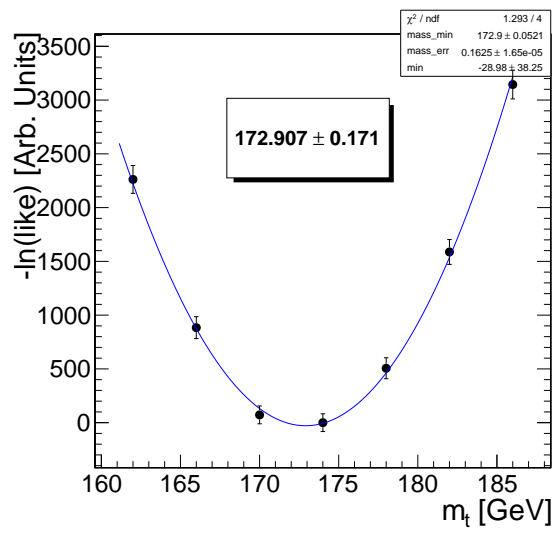
Figure 11.2: Negative logarithm of the sample likelihood distribution as a function of the top quark mass hypotheses for the pseudo-experiment performed on selected data events
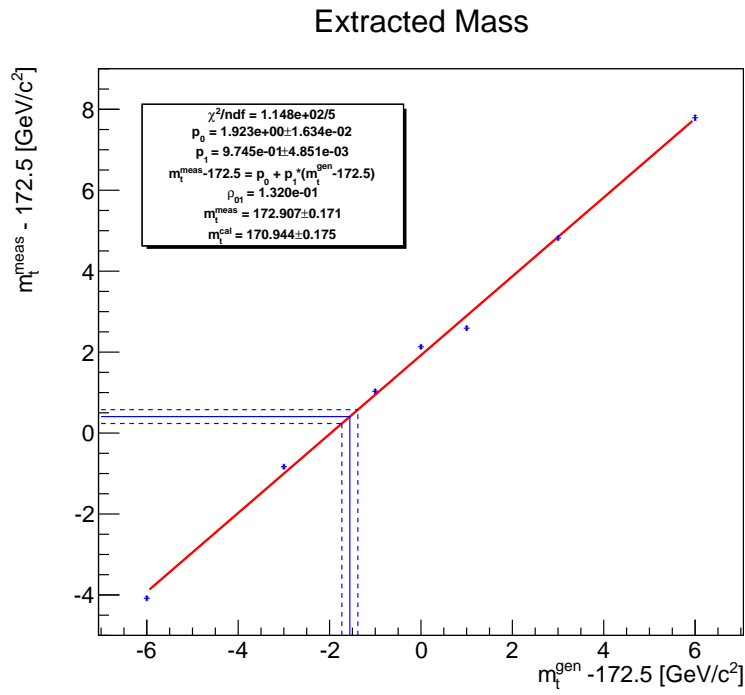
Figure 11.3: The extracted top quark mass and its associated uncertainty are projected on the calibration curve. The calibrated measurement yields $m_t^{cal} = 170.944 \pm 0.175\,\mathrm{GeV}$

Chapter 12

Systematic Uncertainties

In this chapter, the various sources for systematic uncertainties are detailed. They follow a categorisation prescribed in the frame of the TOPLHC [84] working group, a group dedicated to harmonise uncertainty sources across ATLAS and CMS and establish correlations among them. The goal is to be able to combine measurements on the top quark from different analysis and collaborations. A similar process took place at Tevatron, combining results from CDF and DØ. A combined result benefits from the combined statistics and leads to a general improvement on the measurement's precision.

The uncertainties affecting the measurement can be distinguished in three categories. There are uncertainties directly coming from the method itself, uncertainties due to the modelling of physics processes, and uncertainties arising from the detector response. Their evaluation is done using fully simulated samples. For uncertainties arising from reweighing events as described in section 5.5, the event weights are varied by one standard deviation in both directions. Ensembles are then rebuilt accounting for the change in the event weight. For uncertainties affecting the event kinematics or the modelling of physics processes, likelihoods are recalculated for all the events before redrawing ensembles. The acceptance term is also updated when the systematic effect changes the event kinematic if mass dependent samples are available.

The quoted systematic uncertainty is then the difference between the top quark mass measured in the nominal case, and the top quark mass measured with ensembles where the parameter responsible for the systematic effect has been varied. Both variations are quoted, but the largest shift is considered as the final systematic following a conservative approach. Systematic uncertainties are evaluated for samples generated with $m_t = 172.5\,\text{GeV}$ and ensembles are constructed using a mixture of $t\bar{t} \rightarrow e+$jets and $t\bar{t} \rightarrow$ other events. The only exception being logically the estimation of systematic uncertainty due to the sample composition where all the different background are included.

12.1 Systematic Uncertainties from the Method

In this section, the uncertainties arising directly from the Matrix Element Method it self are discussed. This group is composed of the uncertainties coming from the calibration, the normalisation, the acceptance and the sample composition when constructing ensembles. They are estimated by changing the method parameters when constructing the ensembles or when normalising the sample likelihoods.

12.1.1 Calibration

The actual measurement is calibrated as described in chapter 10 using the linear curve shown on figure 10.4. To account for systematic effect due to the calibration, the linear fit parameters are simultaneously varied up and down by their uncertainties. The final measurement is then recalibrated using the linear curve obtained with the varied parameters. The systematic uncertainty arising from calibration is then the difference between the results obtained with the varied curves and the nominal curve. Varying the calibration curve up and down induces a $\pm 0.016\,\text{GeV}$ shift on the extracted top quark mass.

12.1.2 Normalisation

The sample likelihood is normalised to the cross section of the mass hypothesis as discussed in section 6.2. The cross section is evaluated by MADGRAPH when generating a sample given the top quark mass assumption and carry an intrinsic uncertainty. When normalising the sample likelihoods, the cross section values for each mass hypothesis are varied up and down by their uncertainties. The resulting shift on the calibrated top quark mass is then quoted as a systematic uncertainty. The up and down variation yield respectively -0.010 and $-0.005\,\mathrm{GeV}$.

12.1.3 Acceptance

The acceptance term derived in section 6.2 from equation 6.13 is explicitly estimated for each of the samples generated with a different top mass. The fraction of selected events is then fitted with a linear curve. The linear fit parameters are then varied by their uncertainty and a new acceptance curve is obtained. Similarly to what is done for the cross section normalisation, when normalising the sample likelihoods, the calibrated top quark mass is re-evaluated using the varied acceptance curves. The resulting shift of the top quark mass is then quoted as a systematic. Varying the acceptance function up and down induces a shift of -0.031 and $+0.016\,\mathrm{GeV}$ respectively.

12.1.4 Background Composition

When constructing ensembles, the proportion of background events is varied up and down. The amount of $t\bar{t}$ background events is varied by $\pm 10\%$. The proportion of other backgrounds are varied by a factor 2 up/down. As this analysis heavily relies on the final sample purity, this uncertainty is of particular importance. Ensembles are redrawn for the samples generated with $m_t = 172.5\,\mathrm{GeV}$ with the background proportions varied, the remainder being composed of $t\bar{t} \to e$ signal events and the resulting shift on the top quark mass is taken as a systematic. Varying the background fractions up and down induces a shift of -0.138 and $-0.081\,\mathrm{GeV}$ respectively.

12.2 Systematic Uncertainties from Theory Modelling

In this section all the systematic uncertainties arising from the modelling of physics processes are discussed. Most of them are evaluated using dedicated simulated samples, with the exception of B-hadron fragmentation and hadronisation, modelling of top quark p_T and parton distribution functions that are handled via reweighing.

12.2.1 Matrix Element and Parton Showering Generators

The choice on the matrix element and parton showering generators needs to be accounted for as a systematic uncertainty. Nominal samples have been generated using the matrix element generator MADGRAPH coupled the parton shower models from PYTHIA as explained in chapter 3. The top quark mass is evaluated using samples generated with POWHEG/PYTHIA and samples generated with POWHEG/HERWIG. Results obtained with POWHEG/PYTHIA are compared to the nominal results to assess the systematic uncertainty due to the matrix element generator. Results obtained with POWHEG/HERWIG are compared to results obtained with POWHEG/PYTHIA to estimate the systematic uncertainty due to the modelling of parton showers. The uncertainty related to the use of a different Matrix Element generator yields a variation on the top quark mass of $+0.132$ GeV. The uncertainty related to the use of a different parton showering generator yields a variation on the top quark mass of -0.733 GeV.

12.2.2 Matrix Element-Parton Showering Matching

As introduced in chapter 3, the MADGRAPH is interfaced with PYTHIA using the MLM algorithm. The algorithm matches partons to jets that are above a given threshold (20 GeV by default). In dedicated systematics samples, this threshold is changed to 10/40 GeV for the down/up variations respectively. Results on the top quark mass obtained with those systematically varied samples

are compared to the nominal result and the difference is quoted as a systematic uncertainty. Varying the matching threshold up and down induces a shift on the top quark mass of $+0.321$ and $+0.232\,\text{GeV}$ respectively.

12.2.3 Factorisation Scale

The factorisation scale corresponds to the amount of squared transverse momentum used in the evolution of the parton showering process. It is defined as

$$Q^2 = m_t^2 + \sum_{jets} p_T^2 \tag{12.1}$$

Q is varied by factor 2 up and down during the generation of dedicated samples, at both the matrix element and parton showering steps. Results obtained with the dedicated samples are compared to the nominal result and the difference on the top quark mass is quoted as a systematic uncertainty. Varying the factorisation scale up and down induces a shift on the top quark mass of -0.370 and $+0.552\text{GeV}$ respectively.

12.2.4 Underlying Event

To account for uncertainties on the modelling of the underlying event activity, samples are generated with different Perugia 2011 tunes. The tunes used for the evaluation of the underlying event systematic are the mpiHi and Tevatron parametrisations [50] and represent the up and down variations respectively. Results on the top quark mass obtained with these two samples are compared to the result obtained with the nominal Perugia 2011 tune, and the difference is quoted as a systematic uncertainty. Accounting for the underlying event modelling uncertainty yields an up and down variation on the top quark mass of -0.222 and $-0.172\,\text{GeV}$ respectively

12.2.5 Colour Reconnection

Similarly to the evaluation of the systematic arising from the modelling of the underlying event activity, the colour reconnection is turned off in the Perugia

2011 tune. The result on the top quark mass obtained with the sample where colour reconnection has been turned off is compared to the result obtained with the nominal Perugia 2011 tune. The difference between the two results is taken as a systematic uncertainty. Turning off the colour reconnection in the generator yields a variation on the top quark mass of $-0.287\,\mathrm{GeV}$

12.2.6 B-fragmentation

It was noticed that the B-fragmentation modelled in PYTHIA does not correctly describe data recorded by the LEP experiments ALEPH [85] and DELPHI [86]. To account for this systematic effects x_B distributions are evaluated for two different PYTHIA tunes, namely $Z2^*$ and $Z2^*rbLEP$ [18]. x_B is defined as the transverse momentum ratio of generated weakly decaying B-hadron and its matched jet (see figure 12.1). Events are then reweighed to account for the difference in the x_B distributions. The weights are then used when constructing ensembles to evaluate the systematic uncertainty. The result on the top quark mass obtained with the reweighing enabled is compared to the nominal result without reweighing, and the difference is quoted as a systematic uncertainty. Enabling the reweighing related to the B-fragmentation induces a $-1.525\,\mathrm{GeV}$ shift on the top quark mass.

According to figure 12.1, events with x_B between 0.8 and 0.9 tend to get a larger weight. Such events containing less energetic b-jets, when processed by MADWEIGHT would lead to a smaller reconstructed top mass. When ensembles are redrawn with the reweighing enable, those will represent a larger fraction of the final sample, leading to a smaller extracted top quark mass.

12.2.7 Semi-leptonic B-hadron Decay

The semi-leptonic B decay branching ratio in PYTHIA is the same for both processes $B^+ \rightarrow l\nu X$ and $B^0 \rightarrow l\nu X$. To account for uncertainties on the measurement of the two branching ratio [22] the nominal 0.25 branching ratio is varied by -0.45% and +0.77%. Evens are then reweighed to account for this change in the branching ratio, before redrawing ensembles. Results on the top quark mass
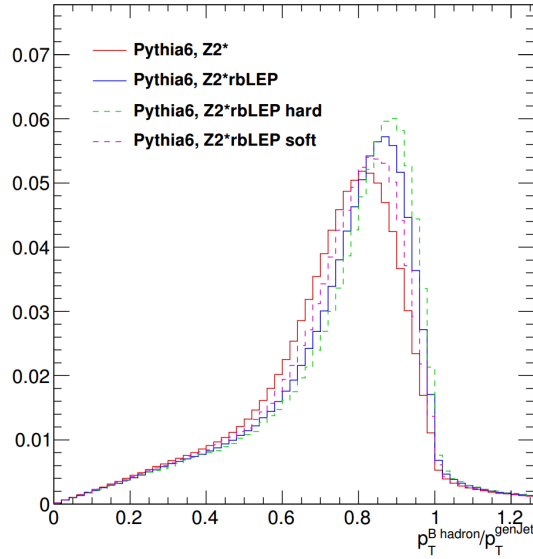
Figure 12.1: $x_B = p_T^{Bhad}/p_T^{Bjet}$ distributions for various PYTHIA tunes. Events are reweighed according to the ratio between the blue and red distributions [18].

obtained with the reweighing enabled are compared to the nominal result and the difference is quoted as a systematic uncertainty. Varying the semi-leptonic B decay branching ratio up and down yields a variation on the top quark mass of -0.211 and $+0.030$ GeV respectively.

12.2.8 Parton Distribution Functions

The Hessian method is used [87] to evaluate the uncertainty due to the choice of parton distribution functions by constructing a N eigenvector basis [88]. PDFs and their uncertainties are estimated with the help of a χ^2 fit to collision data using N free parameters. The N parameter values leading to fit χ^2 minimisation are used as central values. The Hessian error matrix is then diagonalised giving rise to N eigenvectors. Each eigenvector is then varied within its uncertainties leading to $2N$ new parameter sets. This analysis makes use of the CTEQ6L1 during the event generation and reweighs to the CT10 NNLO PDF library containing in total $N = 25$ eigenvectors. Events are reweighed for each variation before proceeding to ensemble testing. The overall uncertainty due to PDF is then evaluated according

to the *Master equation*

$$\Delta m_t^+ = \sqrt{\sum_{i=1}^{N} max(X_i^+ - X_0, X_i^- - X_0, 0)^2}$$

$$\Delta m_t^- = \sqrt{\sum_{i=1}^{N} max(X_0 - X_i^+, X_0 - X_i^-, 0)^2}$$

(12.2)

where X_0 is the top quark mass obtained with the nominal PDF reweighing and $X_i^{\pm}$ is the top quark mass obtained with the up/down variation of the ith eigenvector. The nominal PDF event weight distribution is shown on figure 12.2. The up and down variation of the PDF weights yield an uncertainty of $+0.028$ and $-0.069 \, \mathrm{GeV}$ respectively.
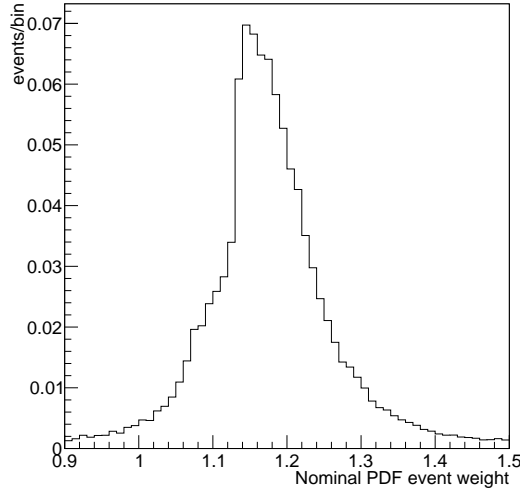
Figure 12.2: Nominal PDF event weight distribution. The central peak shift towards higher values is due to reweighing from CTEQ6L1 to CT10 NNLO

12.2.9   Top p_T Modelling

To estimate the uncertainty due to the modelling of the top p_T, the reweighing presented in section 5.5.5 is disabled before constructing the ensembles. The

result obtained without the reweighing is compared to the nominal case and the difference is quoted as a systematic uncertainty. Disabling the top p_T reweighing yields a shift of -0.084 GeV on the top quark mass.

12.3 Systematic Uncertainties from the detector Response

12.3.1 Trigger, Lepton Identification and Lepton Isolation Efficiencies

To account for trigger performance and lepton identification difference between recorded data and simulation, events in simulation are reweighed with the help of scale factors provided by the dedicated Physics Object Group, as explained in section 5.5.1. The scale factors are varied by their uncertainties and new weights are derived. Ensembles are then redrawn with the varied weights. The shift induced on the top quark mass is then quoted as a systematic. As the reconstruction of the top quark pair mostly depend on the jet rather than the electron kinematics, this systematic uncertainty is expected to be very small. Up and down variation of the trigger efficiency scale factor yield a -0.010 and -0.005 GeV variation on the top quark mass.

12.3.2 Jet Energy Resolution

The determination of the Jet Energy Resolution presented in section 4.2.3.2 are subject to various uncertainty sources. The nominal scaling factors used to smear the jets is varied within its uncertainty. This leads to change in the event kinematics. Likelihood calculation is performed on those systematically varied samples before redrawing ensembles. The resultant shift on the top quark mass is then quoted as a systematic uncertainty. Varying the jet energy resolution correction factor up and down yields a shift on the top quark mass of $+0.349$ and $+0.058$ GeV respectively.

12.3.3 Pile-Up

The number of average pile-up interactions is varied within $\pm 5\%$, as shown of figure 5.6. This implies a new pile-up event weight distribution. The systematically varied weights are used when redrawing ensembles, and the resulting top quark mass is compared to the nominal case. The difference is then quoted as a systematic. Varying the pile-up weights up and down yields a shift on the top quark mass of -0.041 and -0.071 GeV respectively.

12.3.4 b-tagging

The scale factors provided by the dedicated Physics Object Group to account for the b-tagging algorithm performance difference in data and simulation have associated uncertainties. These uncertainties are used to vary up and down the three different scale factor and new event weights are derived. The systematically varied eights are considered when drawing ensembles. The results obtained with the systematically varied weights are compared to the nominal case, and the difference is quoted as a systematic uncertainty. Figure 12.3 shows the nominal and systematically varied weight distributions for both the $b$tagging efficiency and light mis-tagging rate. Varying b-tagging scale factor up and down yields a shift on the top quark mass of -0.001 and -0.016 GeV respectively. The variation of the light mis-identification rate yields a variation on the top quark mass of -0.002 and -0.016 GeV.

12.3.5 Jet Energy Scale

The top quark mass is reconstructed using the jet energies, particularly through the use of jet transfer functions. Consequently, this systematic is expected to have a major impact on the top quark mass measurement. In the frame of the TOPLHC working group, Jet Energy Scale uncertainty sources are decoupled in correlation groups [89] in order to ease future combinations within the CMS collaboration, but also with other experiments at the LHC (ATLAS) and Tevatron (CDF and DØ).
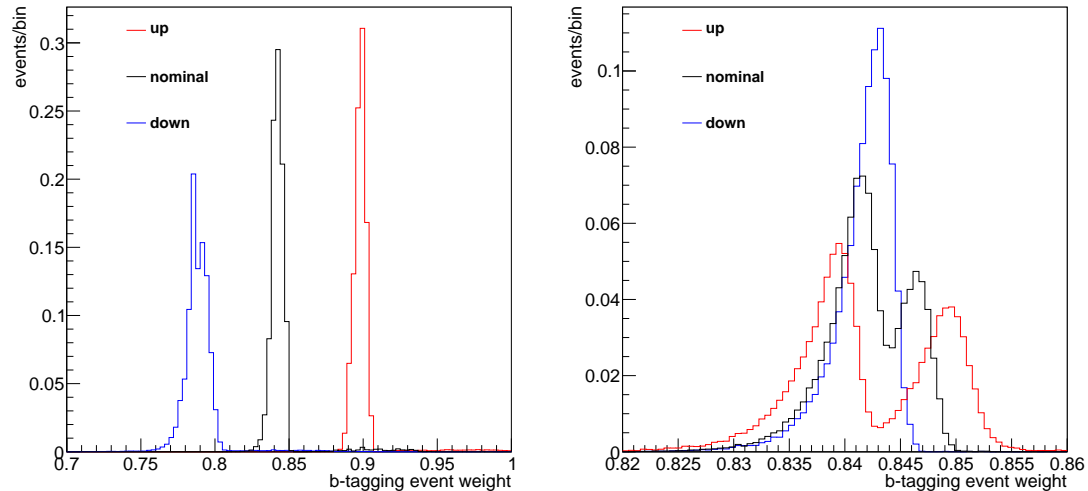
Figure 12.3: Nominal and varied weights for *b*-tagging efficiency systematic effect (left). Nominal and varied weights for mis-tagging rate systematic (right). The observed double-peak structure is explained by the use of two different scale factor for light jet mis-tagging rates (SF_b, SF_c)

Five groups can be distinguished; the *In-situ* correlation group, the *Flavour* correlation group, the *Inter-calibration* correlation group, the *Pile-up p_T* correlation group and the *b-JES* correlation group. Additional uncertainty sources are considered uncorrelated from these groups and between each other and constitute a sixth group, labelled as *Uncorrelated* group. Technically, each component of this group should be treated separately, but given the large number of sources, the fact that no correlation exist between them and the time-consuming nature of likelihood calculations, they are added in quadrature before the smearing is performed.

Each uncertainty source has its own up and down variation (within $\pm 1\sigma$) leading to an up and down smearing of the jet energies. The smearing is applied before selection and gives rise to samples with systematically varied jet kinematics. Each sample is then passed to MADWEIGHT for likelihood calculation. As the smearing is assumed to be top quark mass-dependent, the acceptance curve is also updated to account for the change in selection efficiency. The different

groups composition, a brief explanation from [90] and the resulting uncertainties (up/down) on the top quark mass follow:

- *In-situ* correlation group: *AbsoluteMPFBias* ($+0.362/-0.755$)
  This constant term arises from neutrinos and ISR activity outside of detector acceptance.

- *Flavour* correlation group: *FlavourPureCharm* ($-0.184/-0.182$), *Falvour-PureQuark* ($-0.284/-0.144$), *FlavourPureGluon* ($-0.204/-0.401$)
  This component describes the difference in jet energy response for different jet flavours.
  Likelihoods are calculated independently for the three components, and the variation on the top quark mass are added linearly after ensemble testing yielding an overall uncertainty of ($-0.672/-0.727$).

- *Inter-calibration* correlation group: *RelativeFSR* ($-0.383/-0.207$)
  This term corresponds to the p_T and η-dependent JES component.

- *Pile-up* p_T correlation group: *PileUpPtBB* ($+0.036/-0.499$), *PileUpPtEC* ($-0.012/-0.384$), *PileUpPtHF* ($-0.148/-0.342$)
  Additional uncertainty term introduced in 2012 due to the increase in luminosity compared to 2011 data taking period.
  Likelihoods are calculated independently for the three components, and the variations on the top quark mass are added in quadrature after ensemble testing yielding ($0.153/0.716$). The resulting term is added in quadrature to the pile-up uncertainty term derived from event reweighing, resulting in the total pile-up systematic quoted in chapter 13 for the combination.

- *b-JES* correlation group: *FlavourPureBottom* ($-0.029/-0.589$)

- *Uncorrelated* group ($+0.465/-1.175$): *AbsoluteStat, AbsoluteScale High-PtExtra, SinglePionECAL, SinglePionHCAL, Time, RelativeJEREC1, RelativeJEREC2, RelativeJERHF, RelativePtBB, RelativePtEC1, RelativePtEC2, RelativePtHF, RelativeStatEC2, RelativeStatHF, PileUpDataMC, PileUpBias*
  This category regroups uncertainties on the jet energy scale determination

coming from various effects are that are not considered correlated. A brief description follows, taken from [14, 89]:

- *AbsoluteStat*: p_T dependent statistical uncertainty from the global fit
- *RelativeStatEC2, RelativeStatHF*: η-dependent statistical uncertainty
- *AbsoluteScale*: uncertainty from the lepton scale ($\pm 0.11\%$)
- *HighPtExtra*: uncertainty from the high p_T extrapolation
- *SinglePionECAL, SinglePionHCAL*: uncertainty on the single pion response ($\pm 1.35\%$ in HCAL, $\pm 4.2\%$ in ECAL)
- *Time*: time dependence of the jet energy response due to radiation damage
- *RelativeJEREC1, RelativeJEREC2, RelativeJERHF*: η-dependent uncertainty from the jet energy resolution
- *RelativePtBB, RelativePtEC1, RelativePtEC2, RelativePtHF*: (η,p_T) dependent uncertainty related to the JES parametrisation uncertainty.
- *PileUpDataMC, PileUpBias*: residual uncertainty due to pile-up

12.4 Summary of the Systematic Uncertainties

The different systematic uncertainties are summarised in table 12.4. It lists all the different categories presented in sections 12.1, 12.2 12.3 and the associated up and down variations induced on the top quark mass. To account for a future combination, uncertainties are symmetries following a conservative approach. The symmetric uncertainty value is taken as the maximum absolute value of the two up and down variations. When only one variation is available (matrix element and parton shower generators, top p_T reweighing, *B*-fragmentation and colour reconnection) the symmetric uncertainty is quoted as plus or minus half of that uncertainty. The calibrated final result yields:

$$m_t = 170.94 \pm 0.22 (\text{stat.}) \pm 2.21 (\text{syst.}) \text{ GeV} \qquad (12.3)$$

	Down [GeV]	Up [GeV]	Symmetric [GeV]
Calibration	-0.016	+0.016	0.016
Normalisation	-0.005	-0.010	0.010
Acceptance	+0.016	-0.031	0.030
Background	-0.081	-0.138	0.136
Sub-total			**0.141**
b-tagging	-0.016	+0.001	0.016
mis-tagging	-0.016	+0.002	0.016
JER	+0.058	+0.349	0.349
Lepton ID	-0.008	-0.010	0.010
Pile-up	-0.071	-0.041	0.070
Trigger	+0.005	-0.010	0.010
Sub-total			**0.357**
Colour reconnection	-0.287		0.144
Underlying event	-0.172	-0.222	0.222
ME-PS matching	+0.232	+0.321	0.321
Q^2	+0.552	-0.370	0.552
PDF	-0.069	+0.028	0.069
semi lep B BR	+0.030	-0.211	0.211
B fragmentation	+1.525		0.762
Sub-total			**1.052**
Matrix element	+0.132		0.066
Parton showering	-0.733		0.366
Top p_T	-0.084		0.042
Sub-total			**0.374**
in-situ	-0.755	+0.362	0.755
Flavor group	-0.727	-0.672	0.727
inter calibration	-0.207	-0.383	0.383
bJES	-0.589	-0.029	0.589
Pile-up group	+0.716	+0.153	0.716
Uncorrelated group	-1.175	+0.465	1.175
Sub-total			**1.867**
Total			**2.209**

Table 12.1: Table summarising the systematic uncertainties of the top quark mass. The symmetric uncertainty estimated following the conservative approach depicted in section 12.4.

Part V

Combination, Summary and Outlook

Chapter 13

Combination

In this chapter the concepts behind the combination of physics results is presented, and more specifically, its application to the combination of top quark mass results from CMS. The first section is dedicated to the presentation of the Best Unbiased Linear Estimator (BLUE) method. In the second section, the combination of CMS results for 2010, 2011 and 2012 measurements is first presented. Then, the results from the 2012 lepton+jets measurement is replaced by the results from this analysis to have an overview on how it impacts the combination.

13.1 The BLUE Method

The BLUE method [91] is commonly used in physics experiments to combine measurements on the same physical quantity. The method determines a linear combination of the input measurements and their respective statistical and systematic uncertainties that minimises the overall uncertainty on the combined result, while taking into account the correlations that can exist between the input measurements. The method relies on an appropriate choice of correlation coefficients for the different systematic uncertainty source categories. The categorisation used here is sufficiently segmented so that the correlation coefficients are assumed to be either one or zero. A similar approach has been followed by previous Tevatron [92] and LHC combinations [93].

We consider a combination $\hat{y}$ based on y_i inputs. $\hat{y}$ writes

$$\hat{y} = \sum_{i=1}^{n} \alpha_i y_i \tag{13.1}$$

where α_i is the weight associated to the ith measurement with the constraint

$$\sum_{i=1}^{n} \alpha_i = 1 \tag{13.2}$$

The variance σ on the combination $\hat{y}$ writes as

$$\sigma^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} \rho_{ij} \sigma_i \sigma_j \alpha_i \alpha_j \tag{13.3}$$

where ρ_{ij} denotes the correlation between measurements i and j. The BLUE method consists then in finding the n values of α_i that minimise σ.

13.2 Combination of CMS Top Quark Mass Measurements

During my doctoral studies, I have been involved in the CMS top quark mass combination group, for the publication of a combination using CMS results from 2010 and 2011. It includes five results from CMS. Two measurements from 2010, namely one published result in the lepton+jets channel, and one preliminary result in the di-lepton channel, in addition to three measurements from 2011, namely one preliminary result in the all-jets channel, and two public results, one in the lepton+jets channel and one in the di-lepton channel. This combination has been published in [19] and has been written in collaboration with Steven Wimpenny. It represents an update of a first round of combination [94] which did not include the latest recommended Jet Energy Scale categorisation and in which I also have been involved as author, along with Kelly Beernaert and Martijn Mulders.

As an additional step, the measurement from this analysis is added to the

combination from 2011, and is shown in the column labelled *e+jets 2012* in table 13.1. The table shows input measurements central values, their associated uncertainties and the considered correlation coefficients across channels and data-taking periods. The correlations between input measurements can be found in table 13.2. The combination of the measurements on the top quark mass yields:

$$m_t = 173.40 \pm 0.25(\text{stat.}) \pm 0.93(\text{syst.}) \tag{13.4}$$

It represents an improvement of 20 MeV on the uncertainty with respect to the previous combination [19]. This can be explained by the fact that the measurement from 2011 in the lepton+jets channel still carries most of the weight as it is still very competitive in terms of systematic uncertainties when compared to other measurements.

As discussed in [19] and [94], the correlation between two inputs i and j is limited to σ_i/σ_j (with $\sigma_i < \sigma_j$) to avoid an overestimation of the correlation between significantly different uncertainties [95]. Removing this constraint induces a shift of +225 MeV on the combined result while increasing the overall uncertainty by 15 MeV. Figure 13.1 compares the input measurements to the combined result.

A more recent CMS combination has been published in [90] but it is only used in this dissertation for comparison with the final measurement. The comparison is shown in chapter 14

	di-l 2010	l+jets 2010	di-l 2011	l+jets 2011	all-jets 2011	e+jets 2012	correlation year	correlation channel	combination
Measured m_t	175.50	173.10	172.50	173.49	173.49	170.94			173.40
Statistical Uncertainty	4.60	2.10	0.43	0.27	0.69	0.22	0	0	0.25
in-situ JES factor	n/a	n/a	n/a	0.33	n/a	n/a	0	0	0.22
relative b-jet scale	0.90	0.90	0.76	0.61	0.49	0.89	1	1	0.56
flavour group	1.21	0.87	0.58	0.11	0.58	0.73	1	1	0.22
inter calibration group	0.17	0.08	0.08	0.01	0.08	0.38	1	1	0.03
MPF In-situ group	0.76	0.16	0.35	0.02	0.35	0.76	1	1	0.12
uncorrelated group	1.48	1.90	0.69	0.24	0.69	1.18	1	1	0.32
other uncorrelated JES	3.28	n/a	n/a	n/a	n/a	n/a	0	0	0.01
Lepton energy scale	0.30	n/a	0.14	0.02	n/a	0.01	1	1	0.03
MC generator	0.50	n/e	0.04	0.02	0.19	0.37			0.05
ISR/FSR	0.20	0.20	n/a	n/a	n/a	n/a			0.01
PDF	0.50	0.10	0.09	0.07	0.06	0.07			0.07
Factorization scale	0.60	1.10	0.55	0.24	0.22	0.55			0.24
ME-PS matching threshold	0.70	0.40	0.19	0.18	0.24	0.32			0.19
B decay BR	n/a	n/a	n/a	n/a	n/a	0.21			0.01
B-fragmentation	n/a	n/a	n/a	n/a	n/a	0.76			0.01
Top p_T	n/a	n/a	n/a	n/a	n/a	0.04			0.01
Signal							1	1	
Jet energy resolution	0.50	0.10	0.14	0.23	0.15	0.35			0.19
b-tagging	0.40	0.10	0.09	0.12	0.06	0.02			0.09
MET scale	0.10	0.40	0.12	0.06	n/a	n/a			0.05
Detector Modelling							1	1	
Underlying event	1.30	0.20	0.05	0.15	0.20	0.22	1	1	0.14
Background MC	0.10	0.20	0.05	0.13	n/a	0.14	1	1	0.09
Background Data	n/a	0.40	n/a	n/a	0.13	n/a	0	0	0.03
Fit calibration and MC statistics	0.20	0.10	0.40	0.06	0.13	0.04	0	0	0.06
Pile-up	1.00	0.10	0.11	0.07	0.06	0.72	1	0	0.07
Colour reconnection	n/e	n/e	0.13	0.54	0.15	0.14	1	1	0.37
Trigger	n/a	n/a	n/a	n/a	0.24	0.01	1	0	0.06
Total Systematic Uncertainty	4.53	2.66	1.46	1.03	1.23	2.21			0.93
Total Uncertainty	6.45	3.39	1.52	1.06	1.41	2.22			0.96
Combination weight (%)	-0.09	0.59	8.03	66.43	24.88	0.11			

Table 13.1: Table summarising the results used in the CMS 2011 combination of top quark mass published measurements in addition to the measurement performed by this analysis. The categorisation corresponds to the set of recommendations from the TOPLHC working group. Table taken from [19].

	2010 dilepton	2010 l+jets	2011 dilepton	2011 ljets	2011 all-jets	2012 e+jets
2010 dilepton	1.0000000	0.2035361	0.1968034	0.0935085	0.1531304	0.2502698
2010 l+jets	0.2035361	1.0000000	0.3504150	0.1666358	0.2607003	0.3678536
2011 dilepton	0.1968034	0.3504150	1.0000000	0.3691676	0.6150345	0.4966169
2011 ljets	0.0935085	0.1666358	0.3691676	1.0000000	0.3135831	0.2652699
2011 all-jets	0.1531304	0.2607003	0.6150345	0.3135831	1.0000000	0.4506804
2012 e+jets	0.2502698	0.3678536	0.4966169	0.2652699	0.4506804	1.0000000

Table 13.2: Table showing the correlation coefficients between the input measurements.

143

√s = 7 TeV and 8 TeV

2010 dilepton	175.50 ± 4.60 ± 4.53
JHEP 07 (2011) 049 (L=36/pb)	(val. ± stat. ± syst.)
2010 lepton+jets	173.10 ± 2.10 ± 2.66
CMS PAS TOP-10-009 (L=36/pb)	(val. ± stat. ± syst.)
2011 dilepton	172.50 ± 0.43 ± 1.46
EPJC 72 (2012) 2202 (L=5.0/fb)	(val. ± stat. ± syst.)
2011 lepton+jets	173.49 ± 0.27 ± 1.03
JHEP 12 (2012) 105 (L=5.0/fb)	(val. ± stat. ± syst.)
2011 all-jets	173.49 ± 0.69 ± 1.23
arXiv:1307.4617 (L=3.54/fb)	(val. ± stat. ± syst.)
2012 e+jets	170.94 ± 0.22 ± 2.21
This analysis (L=19.6/fb)	(val. ± stat. ± syst.)
Combined result	173.40 ± 0.25 ± 0.93
up to L=19.6/fb	(val. ± stat. ± syst.)

Combined result

m_t [GeV]

Figure 13.1: Summary plot showing the input measurements used in the combination. This combination represents an update on the combination published in [19] by including the measurement from this analysis.

Chapter 14

Summary and outlook

This analysis has proceeded to a measurement of the top quark mass in the single-electron decay channel of top-anti-top quark pair. It uses proton-proton collisions data at a centre-of-mass energy $\sqrt{s} = 8\,\text{TeV}$ recorded by the Compact Muon Solenoid detector at the Large Hadron Collider during the year 2012, corresponding to an integrated luminosity of $19.6\,\text{fb}^{-1}$. The analysis is based on the Matrix Element Method, a powerful analysis tool developed about two decades ago at Fermilab. The method relies on the construction of sample likelihoods on an event-by-event basis. Likelihoods are obtained for a range of top quark mass hypotheses as the convolution of both experimental and theoretical user-defined quantities. The maximisation of the sample likelihood allows to extract the most probable hypothesis corresponding to the measured top quark mass.

The method is first tested on generator-level objects for signal events. Then experimental effects are introduced via the use of realistic transfer functions and the application of selection criteria. Effects from the electron reconstruction are also investigated to justify the neglect of realistic electron transfer functions. For all these steps the measurement returned by the analysis method is compared to the true value used to generate the input samples. The measurement on collisions data yields a top quark mass of:

$$m_t = 170.94 \pm 0.22(\text{stat.}) \pm 2.21(\text{syst.})\text{GeV} \qquad (14.1)$$

This result is compared with other results from CMS, their combination and the latest world combination regrouping results from CMS, ATLAS, DØ and CDF as of March 2014, as seen on figure 14.1. The final result is compatible with the CMS combination and with the world combination [24] as of March 2014. The world combination does not contain yet latest results from LHC collaborations, combining analysis from 7 TeV collisions only.

The uncertainty of the measurement presented here is dominated by jet energy scale systematic uncertainties when compared to other results from the same data-taking period. The likelihood integration is very sensitive to small jet energy variations and this explains why the jet energy scale is the main source of systematic uncertainty. A higher precision on the final measurement could be achieved by extending the method to an in-situ measurement of the jet energy scale. Either by constraining the W boson mass and comparing it to the invariant mass of the light jets, or by estimating likelihoods in a two-dimensional hypothesis grid (m_t, JES). Given the time scale allocated to this analysis and the high computing power required by the method this approach has not been possible, but definitely represents a challenging and attractive extension of the analysis method. A better comprehension of the B fragmentation uncertainty source could possibly help in reducing the overall uncertainty on the measurement. This uncertainty has been introduced recently and the collaboration has already dedicated a working group to address this issue [96].

The uncertainty related to the uncorrelated jet energy scale uncertainty source has also a large impact on the final measurement. This category is composed of 17 different sources that have been added in quadrature before proceeding the likelihood integration. The bias could possibly be reduced by independently applying jet energy scale variation from each category, giving rise to 2×17 systematically varied samples that will be used for likelihood calculation. It would result in 17 up and 17 down variations on the top quark mass that would be added in quadrature to estimate the total systematic uncertainty coming from the uncorrelated JES group.

√s = 7 TeV and 8 TeV

2010 dilepton JHEP 07 (2011) 049 (L=36/pb)	175.50 ± 4.60 ± 4.52 (val. ± stat. ± syst.)
2010 lepton+jets CMS PAS TOP-10-009 (L=36/pb)	173.10 ± 2.10 ± 2.66 (val. ± stat. ± syst.)
2011 dilepton EPJC 72 (2012) 2202 (L=5.0/fb)	172.50 ± 0.43 ± 1.46 (val. ± stat. ± syst.)
2011 lepton+jets JHEP 12 (2012) 105 (L=5.0/fb)	173.49 ± 0.27 ± 1.03 (val. ± stat. ± syst.)
2011 all-jets arXiv:1307.4617 (L=3.54/fb)	173.49 ± 0.69 ± 1.23 (val. ± stat. ± syst.)
2012 all-jets CMS PAS TOP-14-002 (L=18.2/fb)	172.10 ± 0.27 ± 0.84 (val. ± stat. ± syst.)
2012 dilepton CMS PAS TOP-14-010 (L=19.6/fb)	172.50 ± 0.17 ± 1.40 (val. ± stat. ± syst.)
2012 lepton+jets CMS PAS TOP-14-001 (L=19.6/fb)	172.00 ± 0.11 ± 0.74 (val. ± stat. ± syst.)
CMS combination Sept. 2014 CMS PAS TOP-14-015 up to L=19.6/fb	172.38 ± 0.10 ± 0.65 (val. ± stat. ± syst.)
2012 e+jets This analysis (L=19.6/fb)	170.94 ± 0.22 ± 2.21 (val. ± stat. ± syst.)

World combination March 2014

m_t [GeV]

Figure 14.1: Comparison of the result from this analysis to the CMS combination from September 2014 and the latest world combination from March 2014.

As the main inconvenient from this analysis is the CPU power required to perform integration and likelihood estimation, the optimisation of MADWEIGHT represents a crucial feature for the improvement of this analysis method. MAD-WEIGHT developers are constantly working to improve the running performance of this software. In parallel of the development of this analysis, sensible progress has been achieved concerning MADWEIGHT performance. Another solution also

147

lies and the analysts' side. By reducing the number of possible jet-parton permutations through the use of b-tagging, the computing time could also be significantly reduced.

Appendices

A1 - List of Monte Carlo samples

Data samples

```
/SingleElectron/Run2012A-22Jan2013-v1/AOD
/SingleElectron/Run2012B-22Jan2013-v1/AOD
/SingleElectron/Run2012C-22Jan2013-v1/AOD
/SingleElectron/Run2012D-22Jan2013-v1/AOD
```

Signal samples

```
/TTJets_MSDecays_central_TuneZ2star_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM
/TTJets_MSDecays_mass166_5_TuneZ2star_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM
/TTJets_MSDecays_mass169_5_TuneZ2star_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM
/TTJets_MSDecays_mass171_5_TuneZ2star_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM
/TTJets_MSDecays_mass173_5_TuneZ2star_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM
/TTJets_MSDecays_mass175_5_TuneZ2star_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM
/TTJets_MSDecays_mass178_5_TuneZ2star_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM
```

Background Samples

```
/DYJetsToLL_M-50_TuneZ2Star_8TeV-madgraph-tarball/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
/T_s-channel_TuneZ2star_8TeV-powheg-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
/Tbar_s-channel_TuneZ2star_8TeV-powheg-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
/T_t-channel_TuneZ2star_8TeV-powheg-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
/Tbar_t-channel_TuneZ2star_8TeV-powheg-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
/T_tW-channel-DR_TuneZ2star_8TeV-powheg-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
/Tbar_tW-channel-DR_TuneZ2star_8TeV-powheg-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
/W1JetsToLNu_TuneZ2Star_8TeV-madgraph/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
/W2JetsToLNu_TuneZ2Star_8TeV-madgraph/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
/W3JetsToLNu_TuneZ2Star_8TeV-madgraph/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
/W4JetsToLNu_TuneZ2Star_8TeV-madgraph/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
```

Systematics samples

```
/TT_CT10_AUET2_8TeV-powheg-herwig/Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM
/TT_CT10_TuneZ2star_8TeV-powheg-tauola/Summer12_DR53X-PU_S10_START53_V7A-v2/AODSIM
/TTJets_FullLeptMGDecays_TuneP11TeV_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM
/TTJets_HadronicMGDecays_TuneP11TeV_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM
/TTJets_SemiLeptMGDecays_TuneP11TeV_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM
```

```
/TTJets_FullLeptMGDecays_TuneP11_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM
/TTJets_HadronicMGDecays_TuneP11_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM
/TTJets_SemiLeptMGDecays_TuneP11_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM
/TTJets_FullLeptMGDecays_TuneP11mpiHi_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM
/TTJets_HadronicMGDecays_TuneP11mpiHi_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM
/TTJets_SemiLeptMGDecays_TuneP11mpiHi_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM
/TTJets_FullLeptMGDecays_TuneP11noCR_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM
/TTJets_HadronicMGDecays_TuneP11noCR_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM
/TTJets_SemiLeptMGDecays_TuneP11noCR_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM
/TTJets_MSDecays_scaledown_TuneZ2star_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM
/TTJets_MSDecays_scaleup_TuneZ2star_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM
/TTJets_MSDecays_matchingdown_TuneZ2star_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V19-v2/AODSIM
/TTJets_MSDecays_matchingup_TuneZ2star_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V19-v2/AODSIM
```

A2 - MadWeight settings

Here are listed all the parameters used during the MADWEIGHT integration.

```
##################################################################
##                                                              ##
##                        MadWeight                             ##
##                        =============                         ##
##                                                              ##
##                        Run control                          ##
##                        -----------                          ##
##                                                              ##
##                                                              ##
##    Author: Shannon Crucy (UGent)                            ##
##                                                              ##
##    Version:     5.0.0                                        ##
##    Last change: 29/04/14                                     ##
##                                                              ##
##################################################################
##                                                              ##
##  This Card defines all specific parameters of Madweight      ##
##                                                              ##
##################################################################
#****************************************************************
##                  select run options                         ##
#****************************************************************
Block MW_Run
#   TAG               VALUE                UTILITY
    name              Results   # name for the run
    nb_exp_events     NEVENTS   # number of experimental events to consider
    MW_int_points     5000      # number of points (by permutation) in MadWeight integration for survey
    MW_int_refine     25000     # number of points (by permutation) in MadWeight integration for refine
    precision         0.005     # stops computation if precision is reached.
    nb_event_by_node  1         # one job submission compute the weight for N events
    log_level         weight    # from low level of log to extensive log:
                                # weight, permutation, channel, full
    use_cut           F         # use the cut defined in run_card.dat
    bw_cut            F         # use the BW cut
    nwa               0.1       # width below narrow width approximation is used.
    isr               0         # isr=0 : ignore ISR effect (except if all FS particles are visible)
                                # isr=1 : correct kinematic based on reconstructed Pt(isr)
    inputfile         './Events/input.lhco' # path to the input file (in lhco format)
    pretrained            F                     #TENTATIVE: turns off pretraining of TF


#****************************************************************
##                  define the different param_card's          ##
#****************************************************************
Block MW_parameter
#   TAG      VALUE               UTILITY
    mode        1        #    type of input
                         #    0 : inputs are read from the cards: param_card_1.dat, param_card_2.dat,...
                         #    1 : redefines some values from param_card.dat according to the form below
```

```
                            #   2 : same but the value for different parameters are modified simultaneously

        11      mass         # Block of the parameter to change
        12      6            # id of the parameter to change
        13           162.0                     #Values for top mass hypothesis
        13           166.0
        13           170.0
        13           174.0
        13           178.0
        13           182.0
        13           186.0




#
#                          #  second parameter #
#  use same syntax for parameters 3,4,...
#**********************************************************************
##                          Permutations                    ##
#**********************************************************************
Block MW_perm
#   TAG           VALUE              UTILITY
    permutation  T          # make permutation
    bjet_is_jet  T          # consider permutation between b-jets and light jets
    montecarlo   T          # Monte-Carlo over permutation (Huge speed up if many permutation)
    preselect    'default'  # How to pre-select the correct permutation set.
                            # put 'None' if no pre-selection to perform.
                            # You can set the path to a fortran file defining the require function
                            # See file SubProcesses/permutation_weight_default.dat for
                            # instructions.
    min_perm_cut 5e-4       # Cut for discarding permutation on the preselected method
#**********************************************************************
##                     Phase-Space Integration mapping          ##
#**********************************************************************
Block MW_gen
    force_nwa    2          # Only consider the change of variable alligning particles
                            # with width smaller than this value. This speed up the code
                            # but can lead to zero weight for background event where the
                            # kinematic doesn't agree with the associated mass.
                            # if "mw_run nwa" parameter is bigger than this value, that
                            # value is used for this parameter automatically.
```

Below are shown the model parameters from MADGRAPH5 that are passed to MADWEIGHT

```
######################################################################
## PARAM_CARD AUTOMATICALY GENERATED BY MG5 FOLLOWING UFO MODEL   ####
######################################################################
##                                                                  ##
##  Width set on Auto will be computed following the information    ##
##        present in the decay.py files of the model.               ##
##        See  arXiv:1402.1178 for more details.                    ##
##                                                                  ##
######################################################################


###################################
## INFORMATION FOR MASS
###################################
Block mass
     5 4.700000e+00 # MB
     6 1.730000e+02 # MT
    15 1.777000e+00 # MTA
    23 9.118800e+01 # MZ
    25 1.250000e+02 # MH
## Dependent parameters, given by model restrictions.
```

```
## Those values should be edited following the
## analytical expression. MG5 ignores those values
## but they are important for interfacing the output of MG5
## to external program such as Pythia.
  1 0.000000 # d : 0.0
  2 0.000000 # u : 0.0
  3 0.000000 # s : 0.0
  4 0.000000 # c : 0.0
 11 0.000000 # e- : 0.0
 12 0.000000 # ve : 0.0
 13 0.000000 # mu- : 0.0
 14 0.000000 # vm : 0.0
 16 0.000000 # vt : 0.0
 21 0.000000 # g : 0.0
 22 0.000000 # a : 0.0
 24 80.419002 # w+ : cmath.sqrt(MZ__exp__2/2. + cmath.sqrt(MZ__exp__4/4. - (aEW*cmath.pi*MZ__exp__2)/(Gf*sqrt__2)))


###################################
## INFORMATION FOR SMINPUTS
###################################
Block sminputs
    1 1.325070e+02 # aEWM1
    2 1.166390e-05 # Gf
    3 1.180000e-01 # aS


###################################
## INFORMATION FOR YUKAWA
###################################
Block yukawa
    5 4.700000e+00 # ymb
    6 1.730000e+02 # ymt
   15 1.777000e+00 # ymtau


###################################
## INFORMATION FOR DECAY
###################################
DECAY   6 1.491500e+00 # WT
DECAY  23 2.441404e+00 # WZ
DECAY  24 2.047600e+00 # WW
DECAY  25 6.382339e-03 # WH
## Dependent parameters, given by model restrictions.
## Those values should be edited following the
## analytical expression. MG5 ignores those values
## but they are important for interfacing the output of MG5
## to external program such as Pythia.
DECAY  1 0.000000 # d : 0.0
DECAY  2 0.000000 # u : 0.0
DECAY  3 0.000000 # s : 0.0
DECAY  4 0.000000 # c : 0.0
DECAY  5 0.000000 # b : 0.0
DECAY 11 0.000000 # e- : 0.0
DECAY 12 0.000000 # ve : 0.0
DECAY 13 0.000000 # mu- : 0.0
DECAY 14 0.000000 # vm : 0.0
DECAY 15 0.000000 # ta- : 0.0
DECAY 16 0.000000 # vt : 0.0
DECAY 21 0.000000 # g : 0.0
DECAY 22 0.000000 # a : 0.0
```

This last set of parameters corresponds to the experimental setup in terms of beam energy, detector physical limitations and parton distribution functions.

```
#*****************************************************************
#                    MadGraph/MadEvent                          *
#              http://madgraph.hep.uiuc.edu                     *
```

```
#                                                                    *
#                          run_card.dat                              *
#                                                                    *
#  This file is used to set the parameters of the run.               *
#                                                                    *
#  Some notation/conventions:                                        *
#                                                                    *
#   Lines starting with a '# ' are info or comments                  *
#                                                                    *
#   mind the format:   value   = variable    ! comment               *
#*********************************************************************
#
#********************
# Running parameters
#********************
#
#*********************************************************************
# Tag name for the run (one word)                                    *
#*********************************************************************
    ''      = run_tag ! name of the run. overwritten by the MW card
#*********************************************************************
# Run to generate the grid pack                                      *
#*********************************************************************
   .false.    = gridpack  !True = setting up the grid pack
#*********************************************************************
# Number of events and rnd seed                                      *
#*********************************************************************
       0      = iseed   ! rnd seed
#*********************************************************************
# Collider type and energy                                           *
#*********************************************************************
       1      = lpp1  ! beam 1 type (0=NO PDF)
       1      = lpp2  ! beam 2 type (0=NO PDF)
    4000      = ebeam1  ! beam 1 energy in GeV
    4000      = ebeam2  ! beam 2 energy in GeV
#*********************************************************************
# Beam polarization from -100 (left-handed) to 100 (right-handed)    *
#*********************************************************************
       0      = polbeam1 ! beam polarization for beam 1
       0      = polbeam2 ! beam polarization for beam 2
#*********************************************************************
# PDF CHOICE: this automatically fixes also alpha_s and its evol.    *
#*********************************************************************
 'cteq6l1'    = pdlabel     ! PDF set
#*********************************************************************
# Renormalization and factorization scales                           *
#*********************************************************************
 T        = fixed_ren_scale ! if .true. use fixed ren scale (false is beta)
 T        = fixed_fac_scale ! if .true. use fixed fac scale (false is beta)
 91.1880 = scale            ! fixed ren scale
 91.1880 = dsqrt_q2fact1    ! fixed fact scale for pdf1
 91.1880 = dsqrt_q2fact2    ! fixed fact scale for pdf2
 1        = scalefact        ! scale factor for event-by-event scales
#*********************************************************************
# Matching - Warning! ickkw > 1 is still beta
#*********************************************************************
 0        = ickkw            ! 0 no matching, 1 MLM, 2 CKKW matching
 1        = highestmult      ! for ickkw=2, highest mult group
 1        = ktscheme         ! for ickkw=1, 1 Durham kT, 2 Pythia pTE
 1        = alpsfact         ! scale factor for QCD emission vx
 F        = chcluster        ! cluster only according to channel diag
 F        = pdfwgt           ! for ickkw=1, perform pdf reweighting
#*********************************************************************
#
#**********************************
#*********************************************************************
#
```

```
#********************************
# BW cutoff (M+/-bwcutoff*Gamma)
#********************************
  40  = bwcutoff   ! desactivate by default in the MadWeight_card
#*******************
# Standard Cuts   ! desactivate by default in the MadWeight_card
#*******************
#
#******************************************************************
# Minimum and maximum pt's                                        *
#******************************************************************
  0   = ptj      ! minimum pt for the jets
  0   = ptb      ! minimum pt for the b
  0   = pta      ! minimum pt for the photons
  0   = ptl      ! minimum pt for the charged leptons
  0   = misset   ! minimum missing Et (sum of neutrino's momenta)
  0   = ptheavy  ! minimum pt for one heavy final state
 1.0  = ptonium  ! minimum pt for the quarkonium states
 1d5  = ptjmax    ! maximum pt for the jets
 1d5  = ptbmax    ! maximum pt for the b
 1d5  = ptamax    ! maximum pt for the photons
 1d5  = ptlmax    ! maximum pt for the charged leptons
 1d5  = missetmax ! maximum missing Et (sum of neutrino's momenta)
#******************************************************************
# Minimum and maximum E's (in the lab frame)                      *
#******************************************************************
  0   = ej     ! minimum E for the jets
  0   = eb     ! minimum E for the b
  0   = ea     ! minimum E for the photons
  0   = el     ! minimum E for the charged leptons
 1d5  = ejmax ! maximum E for the jets
 1d5  = ebmax ! maximum E for the b
 1d5  = eamax ! maximum E for the photons
 1d5  = elmax ! maximum E for the charged leptons
#******************************************************************
# Maximum and minimum rapidity                                    *
#******************************************************************
 1d2  = etaj    ! max rap for the jets
 1d2  = etab    ! max rap for the b
 1d2  = etaa    ! max rap for the photons
 1d2  = etal    ! max rap for the charged leptons
 1d2  = etaonium ! max rap for the quarkonium states
 0d0  = etajmin ! min rap for the jets
 0d0  = etabmin ! min rap for the b
 0d0  = etaamin ! min rap for the photons
 0d0  = etalmin ! main rap for the charged leptons
#******************************************************************
# Minimum and maximum DeltaR distance                             *
#******************************************************************
 0.   = drjj    ! min distance between jets
 0    = drbb    ! min distance between b's
 0.   = drll    ! min distance between leptons
 0.   = draa    ! min distance between gammas
 0    = drbj    ! min distance between b and jet
 0.   = draj    ! min distance between gamma and jet
 0.   = drjl    ! min distance between jet and lepton
 0    = drab    ! min distance between gamma and b
 0    = drbl    ! min distance between b and lepton
 0.   = dral    ! min distance between gamma and lepton
 1d2  = drjjmax ! max distance between jets
 1d2  = drbbmax ! max distance between b's
 1d2  = drllmax ! max distance between leptons
 1d2  = draamax ! max distance between gammas
 1d2  = drbjmax ! max distance between b and jet
 1d2  = drajmax ! max distance between gamma and jet
 1d2  = drjlmax ! max distance between jet and lepton
 1d2  = drabmax ! max distance between gamma and b
```

156

```
 1d2 = drblmax ! max distance between b and lepton
 1d2 = dralmax ! maxdistance between gamma and lepton
#*******************************************************************
# Minimum and maximum invariant mass for pairs                    *
#*******************************************************************
 0   = mmjj    ! min invariant mass of a jet pair
 0   = mmbb    ! min invariant mass of a b pair
 0   = mmaa    ! min invariant mass of gamma gamma pair
 0   = mmll    ! min invariant mass of l+l- (same flavour) lepton pair
 1d5 = mmjjmax ! max invariant mass of a jet pair
 1d5 = mmbbmax ! max invariant mass of a b pair
 1d5 = mmaamax ! max invariant mass of gamma gamma pair
 1d5 = mmllmax ! max invariant mass of l+l- (same flavour) lepton pair
#*******************************************************************
# Minimum and maximum invariant mass for all letpons             *
#*******************************************************************
 0   = mmnl    ! min invariant mass for all letpons (l+- and vl)
 1d5 = mmnlmax ! max invariant mass for all letpons (l+- and vl)
#*******************************************************************
# Inclusive cuts                                                  *
#*******************************************************************
 0 = xptj ! minimum pt for at least one jet
 0 = xptb ! minimum pt for at least one b
 0 = xpta ! minimum pt for at least one photon
 0 = xptl ! minimum pt for at least one charged lepton
#*******************************************************************
# Control the pt's of the jets sorted by pt                       *
#*******************************************************************
 0   = ptj1min ! minimum pt for the leading jet in pt
 0   = ptj2min ! minimum pt for the second jet in pt
 0   = ptj3min ! minimum pt for the third jet in pt
 0   = ptj4min ! minimum pt for the fourth jet in pt
 1d5 = ptj1max ! maximum pt for the leading jet in pt
 1d5 = ptj2max ! maximum pt for the second jet in pt
 1d5 = ptj3max ! maximum pt for the third jet in pt
 1d5 = ptj4max ! maximum pt for the fourth jet in pt
 0   = cutuse  ! reject event if fails any (0) / all (1) jet pt cuts
#*******************************************************************
# Control the Ht(k)=Sum of k leading jets                         *
#*******************************************************************
 0   = htjmin ! minimum jet HT=Sum(jet pt)
 1d5 = htjmax ! maximum jet HT=Sum(jet pt)
 0   = ht2min ! minimum Ht for the two leading jets
 0   = ht3min ! minimum Ht for the three leading jets
 0   = ht4min ! minimum Ht for the four leading jets
 1d5 = ht2max ! maximum Ht for the two leading jets
 1d5 = ht3max ! maximum Ht for the three leading jets
 1d5 = ht4max ! maximum Ht for the four leading jets
#*******************************************************************
# WBF cuts                                                        *
#*******************************************************************
 0   = xetamin ! minimum rapidity for two jets in the WBF case
 0   = deltaeta ! minimum rapidity for two jets in the WBF case
#*******************************************************************
# maximal pdg code for quark to be considered as a jet            *
# otherwise b cuts are applied                                    *
#*******************************************************************
 4 = maxjetflavor
#*******************************************************************
# Jet measure cuts                                                *
#*******************************************************************
 0   = xqcut   ! minimum kt jet measure between partons
#*******************************************************************
```

A3 - Resolution Functions

Here are presented the resolution function parameters (equation 5.6) used by HITFIT.

bin	C	R	N
$0.00 \leq \|\eta\| < 0.17$	0.0000001	0.1116958	0.4465630
$0.17 \leq \|\eta\| < 0.26$	0.0043426	0.0868324	0.5067006
$0.26 \leq \|\eta\| < 0.35$	0.0009738	0.0972946	0.5108543
$0.35 \leq \|\eta\| < 0.43$	0.0000003	0.1078232	0.5268910
$0.43 \leq \|\eta\| < 0.52$	0.0007470	0.1152864	0.4787032
$0.52 \leq \|\eta\| < 0.61$	0.0057875	0.0635109	0.7129480
$0.61 \leq \|\eta\| < 0.70$	0.0059804	0.0430556	0.7994914
$0.70 \leq \|\eta\| < 0.78$	0.0000015	0.1124020	0.5358127
$0.78 \leq \|\eta\| < 0.87$	0.0080070	0.0000071	0.9554642
$0.87 \leq \|\eta\| < 0.96$	0.0000000	0.0955331	0.7628243
$0.96 \leq \|\eta\| < 1.04$	0.0000001	0.1328288	0.5243128
$1.04 \leq \|\eta\| < 1.13$	0.0057991	0.1256884	0.7911570
$1.13 \leq \|\eta\| < 1.22$	0.0000003	0.1398852	1.0961863
$1.22 \leq \|\eta\| < 1.30$	0.0000037	0.1362100	1.0745044
$1.30 \leq \|\eta\| < 1.39$	0.0000002	0.1448911	1.1525609
$1.39 \leq \|\eta\| < 1.48$	0.0000000	0.0758274	1.5690025
$1.48 \leq \|\eta\| < 1.65$	0.0141146	0.1874686	1.2555640
$1.65 \leq \|\eta\| < 1.74$	0.0000002	0.2070571	0.6964012
$1.74 \leq \|\eta\| < 1.83$	0.0000007	0.1620558	0.8152754
$1.83 \leq \|\eta\| < 1.93$	0.0107941	0.1503900	0.5303855
$1.93 \leq \|\eta\| < 2.04$	0.0000003	0.1638662	0.5608603
$2.04 \leq \|\eta\| < 2.17$	0.0165470	0.0000079	1.0451610
$2.17 \leq \|\eta\| < 2.32$	0.0147569	0.1484052	0.0002876
$2.32 \leq \|\eta\| < 2.50$	0.0226895	0.0000294	0.6589472

Table 1: Table showing resolution function parameters for electrons.

bin	C	R	N
$0.00 \leq |\eta| < 0.09$	0.0339143	1.0498081	3.1933158
$0.09 \leq |\eta| < 0.17$	0.0301536	1.0400983	4.0859623
$0.17 \leq |\eta| < 0.26$	0.0351717	0.9344085	5.1758787
$0.26 \leq |\eta| < 0.35$	0.0000002	1.1115271	3.1918092
$0.35 \leq |\eta| < 0.43$	0.0109333	1.0665496	3.8119158
$0.43 \leq |\eta| < 0.52$	0.0000012	1.0785656	3.7008919
$0.52 \leq |\eta| < 0.61$	0.0000087	1.0332245	4.4985712
$0.61 \leq |\eta| < 0.70$	0.0000008	1.0698196	3.4332194
$0.70 \leq |\eta| < 0.78$	0.0000089	1.0418117	3.7582456
$0.78 \leq |\eta| < 0.87$	0.0258158	1.0129058	4.6143325
$0.87 \leq |\eta| < 0.96$	0.0000002	1.0569491	4.3990954
$0.96 \leq |\eta| < 1.04$	0.0000007	1.0467933	4.3327277
$1.04 \leq |\eta| < 1.13$	0.0200445	1.0227807	4.8311554
$1.13 \leq |\eta| < 1.22$	0.0000072	1.0618716	4.7889984
$1.22 \leq |\eta| < 1.30$	0.0000131	1.1599805	3.2867124
$1.30 \leq |\eta| < 1.39$	0.0423873	0.9131824	5.7965311
$1.39 \leq |\eta| < 1.48$	0.0175476	1.1942401	2.7735863
$1.48 \leq |\eta| < 1.57$	0.0350734	1.0536675	5.2394280
$1.57 \leq |\eta| < 1.65$	0.0000001	0.8733962	6.5263116
$1.65 \leq |\eta| < 1.74$	0.0000010	0.8499133	6.2791884
$1.74 \leq |\eta| < 1.83$	0.0000674	0.8516685	5.9834905
$1.83 \leq |\eta| < 1.93$	0.0629578	0.5186734	6.2659769
$1.93 \leq |\eta| < 2.04$	0.0000041	0.8317480	5.5201662
$2.04 \leq |\eta| < 2.17$	0.0000030	0.7404611	6.2077052
$2.17 \leq |\eta| < 2.32$	0.0429159	0.6705800	6.0748948
$2.32 \leq |\eta| < 2.50$	0.0000044	0.6808852	6.6872208

Table 2: Table showing resolution function parameters for b-jets.

bin	C	R	N
$0.00 \leq |\eta| < 0.09$	0.0550557	1.0824954	0.0005458
$0.09 \leq |\eta| < 0.17$	0.0522282	1.0963008	0.0003938
$0.17 \leq |\eta| < 0.26$	0.0500649	1.0943606	0.0011667
$0.26 \leq |\eta| < 0.35$	0.0468126	1.0926515	0.0003195
$0.35 \leq |\eta| < 0.43$	0.0422185	1.0934590	0.0001993
$0.43 \leq |\eta| < 0.52$	0.0457314	1.0957947	0.0000889
$0.52 \leq |\eta| < 0.61$	0.0390785	1.0958564	0.0002972
$0.61 \leq |\eta| < 0.70$	0.0463172	1.0873420	0.0000330
$0.70 \leq |\eta| < 0.78$	0.0432268	1.0951669	0.0000921
$0.78 \leq |\eta| < 0.87$	0.0486335	1.0604886	0.6047445
$0.87 \leq |\eta| < 0.96$	0.0432786	1.1116478	0.0003794
$0.96 \leq |\eta| < 1.04$	0.0520726	1.0927815	0.0003268
$1.04 \leq |\eta| < 1.13$	0.0793913	0.8422948	3.9503718
$1.13 \leq |\eta| < 1.22$	0.0572457	1.1359796	0.0000206
$1.22 \leq |\eta| < 1.30$	0.0335641	1.1861008	0.0001887
$1.30 \leq |\eta| < 1.39$	0.0199005	1.2441185	0.0004854
$1.39 \leq |\eta| < 1.48$	0.0000002	1.1595740	2.9860835
$1.48 \leq |\eta| < 1.57$	0.0548260	1.1721283	0.0004322
$1.57 \leq |\eta| < 1.65$	0.0000103	1.0260876	3.7080667
$1.65 \leq |\eta| < 1.74$	0.0000078	1.0786299	2.8293796
$1.74 \leq |\eta| < 1.83$	0.0232382	0.8419282	4.6769467
$1.83 \leq |\eta| < 1.93$	0.0428637	0.9699944	0.0002350
$1.93 \leq |\eta| < 2.04$	0.0493282	0.9707288	2.2383767
$2.04 \leq |\eta| < 2.17$	0.0000107	1.0272089	1.3342771
$2.17 \leq |\eta| < 2.32$	0.0841025	0.5483953	4.2945710
$2.32 \leq |\eta| < 2.50$	0.0000123	0.9520769	4.1680710

Table 3: Table showing resolution function parameters for light jets.

A4 - HitFit settings

Below is written the jet specific configuration for HITFIT

```
[mass_studies]
# ------------------------------------------------
# maximum number of jets to be considered in the
# jet combinatorics (has to be >= 4, can be set to
# -1 if you want to take all)
# ------------------------------------------------
maxNJets = 4
#-------------------------------------------------
# maximum number of jet combinations finally
# written into the event, starting from the "best"
# (has to be >= 1, can be set to -1 if you want to
# take all)
#-------------------------------------------------
maxNComb = 24
# ------------------------------------------------
# option to use b-tagging
# ------------------------------------------------
bTagAlgo = combinedSecondaryVertexBJetTags
minBDiscBJets = 0.898 #CSVT working point
maxBDiscLightJets = 0.898
useBTagging = true
# ------------------------------------------------
# set mass values used in the constraints
# set mass to 0 for no constraint
# ------------------------------------------------
mW = 80.4
mTop = 0
# ------------------------------------------------
# specify jet correction level
# ------------------------------------------------
jetCorrectionLevel = "L3Absolute"
# ------------------------------------------------
# rescale jets
# ------------------------------------------------
jes = 1.0
jesB = 1.0
```

Below is the main configuration for HITFIT, controlling the fit constraints and other various parameters.

```
#
#   @file RunHitFitConfiguration.txt
#
#   @brief Example of input file to be used with RunHitFit class.
#
```

```
#    @author Haryo Sumowidagdo <Suharyo.Sumowidagdo@cern.ch>
#
#    @date Mon Aug 30 14:39:33 CEST 2010
#
#    $Id: RunHitFitConfiguration.txt,v 1.1 2011/05/26 09:46:52 mseidel Exp $
#######################################################################
# Top_Fit parameters
#

# Minimum hadronic W mass allowed before fit
mwhad_min_cut  =  0

# Maximum hadronic W mass allowed before fit
mwhad_max_cut  = 10000

# Maximum difference in mass between leptonic and hadronic top quark mass
# allowed before fit
mtdiff_max_cut = 10000

# Maximum jet invariant mass cut allowed
jet_mass_cut   =  10000

# Maximum number of solution to be keep
nkeep = 1680

# If true, solve neutrino pz by requiring the leptonic side and
# hadronic side to have equal mass
# If false, solve neutrino pz by requiring m(lepton,neutrino) = mW
solve_nu_tmass = true

# Fit ttH -> l+jets+bb
# Should always be set to FALSE
# Reason: Event with the same ttbar jet assignment but different
# Higgs-jet/b-bbar jet assignment ARE not weighed differently by HitFit.
# So it is just a waste of CPU time and storage area with no gain in
# information.
# For ttH analysis, it is better to fit ttbar first, and then perform the dijet
# invariant mass analysis later.
do_higgs_flag = false

# If true, print event after fit in Top_Fit fitter.
# If false, do not print event after fit in Top_Fit fitter.
# Should always be set to FALSE
print_event_flag = false

#######################################################################
# Constrained_Top parms
#
# Mass of the b-quark.
bmass = 4.7

# Require that the leptonic top and hadronic top to have equal mass after
# the fit.

equal_side = true

#######################################################################
# Fourvec_Constrainer parms.
#
# All the objects are fixed to constant masses for the fit.
# (These masses are attributes of the objects in the Fourvec_Event.)
# This is done by scaling either the 4-vector's 3-momentum or energy,
# depending on the setting of this parameter.
# If TRUE: Keep E and rescale three-momentum.
# If FALSE: Keep three-momentum and rescale E.
use_e = true
```

162

```
# Center-of-mass energy.  Used to force a step cut
# if the fit goes into an unphysical region.
# Tevatron Run 1 : e_com = 1800
# Tevatron Run 2 : e_com = 1960
# LHC Run        : e_com = 7000, 8000, 10000, 12000, 13000, 14000
e_com = 7000

# If this is true and the event does not have a neutrino,
# then the fit will be done without the overall transverse
# momentum constraint (and thus the missing Et information
# will be ignored).  If the event does have a neutrino, this parameter
# is ignored.
ignore_met = false

######################################################################
# Chisq_Constrainer parms.
#

# If true, print a trace of the fit to cout.
printfit = false

# If true, check the chisq formula by computing it directly from G.
# This requires that G_i be invertible.
use_G = false

# Convergence threshold for sum of constraints.
constraint_sum_eps = 0.01

# Convergence threshold for change in chisq.
chisq_diff_eps = 0.01

# Maximum number of iterations permitted.
maxit = 10000

# Maximum number of cut steps permitted.
maxcut = 20

# Fraction by which to cut steps.
cutsize = 0.5

# Smallest fractional cut step permitted.
min_tot_cutsize = 1e-10

# When use_G is true, the maximum relative difference permitted between
# the two chisq calculations.
chisq_test_eps = 1e-5

######################################################################
# Base_Constrainer parms.
#
# If true, check the constraint gradient calculations by also
# doing them numerically.
test_gradient = false

# When test_gradient is true, step size to use for numeric differentiation.
test_step = 0.002

# When test_gradient is true, maximum relative difference permitted
# between returned and numerically calculated gradients.
test_eps = 0.035
```

A5 - Transfer Functions parameters

		p_1	p_2	p_3	p_4	p_5
$0.0 < |\eta| < 0.87$	$0 \leq E_p < 40$	-1.319	4.153	0.221	-8.404	7.465
	$40 \leq E_p < 50$	1.084	6.076	0.180	-8.393	8.728
	$50 \leq E_p < 60$	1.003	8.881	0.027	-23.874	8.428
	$60 \leq E_p < 70$	-1.230	5.771	1.671	1.990	12.232
	$70 \leq E_p < 80$	4.282	15.052	1.266	-1.487	7.737
	$80 \leq E_p < 90$	-1.636	7.982	0.656	5.161	16.841
	$90 \leq E_p < 100$	2.558	15.214	0.003	-43.712	4.043
	$100 \leq E_p < 125$	-2.145	10.691	0.376	9.925	22.164
	$125 \leq E_p < 150$	-2.182	12.569	0.285	12.416	27.419
	$150 \leq E_p < 200$	-2.446	15.073	0.228	18.580	32.954
	$200 \leq E_p$	-4.168	18.757	0.200	22.540	45.058
$0.87 < |\eta| < 1.479$	$0 \leq E_p < 50$	-5.027	4.623	0.341	-16.294	8.471
	$50 \leq E_p < 60$	-1.187	6.469	0.378	-12.097	10.418
	$60 \leq E_p < 70$	0.059	9.588	0.140	-15.301	13.014
	$70 \leq E_p < 80$	-18.390	17.073	12.755	1.241	11.937
	$80 \leq E_p < 90$	1.167	14.642	0.026	-31.706	20.543
	$90 \leq E_p < 100$	1.213	17.312	0.012	-56.639	10.986
	$100 \leq E_p < 125$	-3.330	10.236	1.777	3.553	22.548
	$125 \leq E_p < 150$	-2.660	15.134	0.773	7.748	28.772
	$150 \leq E_p < 200$	-3.915	18.439	0.528	12.885	34.872
	$200 \leq E_p$	-6.001	23.442	0.419	17.856	47.386
$1.479 < |\eta| < 2.5$	$0 \leq E_p < 100$	-31.486	21.647	6.786	-7.164	13.328
	$100 \leq E_p < 125$	-20.152	21.990	2.606	3.493	17.028
	$125 \leq E_p < 150$	-65.657	12.150	32.928	-0.739	25.816
	$150 \leq E_p < 200$	-0.309	29.803	0.087	1.291	61.919
	$200 \leq E_p$	6.325	61.754	1.345	-10.387	31.241

Table 4: Table summarising b-jets transfer function parameters

		p_1	p_2	p_3	p_4	p_5
$0.0 < |\eta| < 0.87$	$0 \leq E_p < 40$	-3.247	4.807	0.149	-14.930	3.673
	$40 \leq E_p < 50$	-0.138	6.187	0.204	-10.323	7.615
	$50 \leq E_p < 60$	-0.497	7.570	0.276	-2.428	13.988
	$60 \leq E_p < 70$	-1.216	7.146	0.688	-0.668	13.476
	$70 \leq E_p < 80$	-2.421	6.379	1.866	-0.660	12.636
	$80 \leq E_p < 90$	-1.893	8.498	0.887	-0.914	14.629
	$90 \leq E_p < 100$	-3.208	6.630	2.059	-1.465	14.572
	$100 \leq E_p < 125$	-2.864	13.516	0.055	28.533	5.771
	$125 \leq E_p < 150$	-4.679	12.154	0.472	-1.933	21.532
	$150 \leq E_p < 200$	-5.080	21.410	0.830	-6.095	12.518
	$200 \leq E_p$	-12.007	30.723	0.982	-9.319	13.936
$0.87 < |\eta| < 1.479$	$0 \leq E_p < 50$	-9.074	6.600	0.295	-21.308	11.250
	$50 \leq E_p < 60$	-3.929	7.755	0.296	-17.119	11.938
	$60 \leq E_p < 70$	-1.506	8.602	0.539	-10.172	13.992
	$70 \leq E_p < 80$	-0.048	11.283	0.349	-14.831	14.825
	$80 \leq E_p < 90$	-35.032	9.934	16.591	-2.054	14.293
	$90 \leq E_p < 100$	-4.492	8.261	2.641	-2.524	18.227
	$100 \leq E_p < 125$	-3.168	21.842	0.583	-2.418	11.852
	$125 \leq E_p < 150$	-8.183	12.292	2.964	-2.661	22.899
	$150 \leq E_p < 200$	-1.800	34.109	1.855	-7.023	18.999
	$200 \leq E_p$	-69.021	6.535	16.762	-10.703	29.300
$1.479 < |\eta| < 2.5$	$0 \leq E_p < 100$	-37.934	22.261	3.129	-11.963	14.207
	$100 \leq E_p < 125$	-36.597	27.351	3.651	-5.162	18.533
	$125 \leq E_p < 150$	-68.799	22.489	11.165	-8.079	25.037
	$150 \leq E_p < 200$	-35.647	53.234	9.678	-9.181	28.709
	$200 \leq E_p$	-15.608	31.650	0.369	-22.291	63.261

Table 5: Table summarising light jets transfer function parameters

		p_1	p_2	p_3	p_4	p_5
$0.0 < |\eta| < 0.87$	$0 \leq E_{gen} < 40$	-0.241	0.616	0.163	-0.579	1.763
	$40 \leq E_{gen} < 60$	-0.279	0.901	0.054	0.761	3.300
	$60 \leq E_{gen} < 80$	-0.269	1.028	0.048	1.505	3.989
	$80 \leq E_{gen} < 100$	-0.327	1.185	0.040	2.890	5.168
	$100 \leq E_{gen} < 200$	-0.418	1.420	0.042	3.900	5.835
	$200 \leq E_{gen}$	0.402	2.106	0.035	8.580	9.210
$0.0 < |\eta| < 0.87$	$0 \leq E_{gen} < 60$	-0.664	1.810	0.246	-0.411	2.757
	$60 \leq E_{gen} < 80$	2.230	4.537	9.400	-0.561	1.771
	$80 \leq E_{gen} < 100$	-0.371	2.051	0.077	4.320	5.980
	$100 \leq E_{gen} < 200$	7.280	8.135	15.970	-0.122	2.469
	$200 \leq E_{gen}$	10.860	10.560	11.980	0.511	3.389
$0.0 < |\eta| < 0.87$	$00 \leq E_{gen} < 100$	-1.312	5.130	0.980	-2.634	2.350
	$100 \leq E_{gen} < 200$	2.430	10.550	12.400	-3.628	5.050
	$200 \leq E_{gen}$	-6.275	6.640	0.075	2.860	15.630

Table 6: Table summarising electron transfer function parameters

References

[1] E. Halkiadakis. Proceedings for TASI 2009 Summer School on 'Physics of the Large and the Small': Introduction to the LHC experiments. 2010. xii, 8

[2] B. Klein. Measurement of the top quark pair production cross section in the electron and jets channel at sqrt(s) = 7 TeV with the CMS detector at the LHC. 2012. xii, xiii, xiv, 12, 40, 52

[3] The CMS Collaboration. *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software.* Technical Design Report CMS. CERN, Geneva, 2006. xii, 15, 23

[4] V. Veszpremi. Operation and performance of the CMS tracker. *JINST*, 9:C03005, 2014. xii, xiii, 24, 25, 27

[5] V. Khachatryan, A. M. Sirunyan, A. Tumasyan, W. Adam, T. Bergauer, M. Dragicevic, J. Erö, C. Fabjan, M. Friedl, R. Frühwirth, and et al. CMS tracking performance results from early LHC operation. *European Physical Journal C*, 70:1165–1192, December 2010. xiii, 24, 26

[6] Serguei Chatrchyan et al. Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in pp Collisions at $\sqrt{s}$ = 7 TeV. *JINST*, 8:P09009, 2013. xiii, 26, 29

[7] CMS Collaboration. Performance of the CMS hadron calorimeter with cosmic ray muons and LHC beam data. *Journal of Instrumentation*, 5:3012, March 2010. xiii, 28, 31

[8] CMS Collaboration. Precise mapping of the magnetic field in the CMS barrel yoke using cosmic rays. *Journal of Instrumentation*, 5:3021, March 2010. xiii, 30, 32

[9] The CMS Collaboration. Performance of CMS muon reconstruction in pp collision events at sqrt(s) = 7 TeV. *Journal of Instrumentation*, 7:2P, October 2012. xiii, 31, 33, 34

[10] The CMS Collaboration. *CMS TriDAS project: Technical Design Report, Volume 1: The Trigger Systems.* Technical Design Report CMS. xiii, 35

[11] A. Afaq, W. Badgett, G. Bauer, K. Biery, V. Boyer, et al. The CMS High Level Trigger System. *IEEE Trans.Nucl.Sci.*, 55:172–175, 2008. xiii, 37

[12] Jonathan Pumplin, Daniel Robert Stump, Joey Huston, Hung-Liang Lai, Pavel Nadolsky, and Wu-Ki Tung. New generation of parton distributions with uncertainties from global qcd analysis. *Journal of High Energy Physics*, 2002(07):012, 2002. xiii, 41

[13] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The Anti-k(t) jet clustering algorithm. *JHEP*, 0804:063, 2008. xiv, 49, 51

[14] Jet energy scale global fit and uncertainties. Technical Report CMS AN-2015/023. xiv, 52, 136

[15] Serguei Chatrchyan et al. Identification of b-quark jets with the CMS experiment. *JINST*, 8:P04013, 2013. xiv, 53, 68

[16] Fireworks workbook twiki page. https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookFireworks. xiv, 57

[17] Tag & probe studies twiki. https://twiki.cern.ch/twiki/bin/viewauth/CMS/KoPFAElectronTagAndProbe. xv, 65, 66

[18] M. Seidel. Slides from cms top mass meeting, january 17th, 2014. xxi, 129, 130

[19] Top quark mass combinations using the 2010, 2011 and 2012 data. Technical Report CMS-PAS-TOP-13-002, CERN, Geneva, 2013. xxi, 140, 141, 142, 144

[20] Roel Aaij et al. Observation of the resonant character of the $Z(4430)^-$ state. *Phys.Rev.Lett.*, 112(22):222002, 2014. 5

[21] Roel Aaij et al. Observation of $J/\psi p$ resonances consistent with pentaquark states in $\Lambda_b^0 \to J/\psi K^- p$ decays. 2015. 5

[22] K.A. Olive and Particle Data Group. Review of particle physics. *Chinese Physics C*, 38(9):090001, 2014. 5, 6, 84, 129

[23] Claudio Campagnari and Melissa Franklin. The Discovery of the top quark. *Rev.Mod.Phys.*, 69:137–212, 1997. 9

[24] First combination of Tevatron and LHC measurements of the top-quark mass. 2014. 9, 146

[25] S. Alekhin, A. Djouadi, and S. Moch. The top quark and Higgs boson masses and the stability of the electroweak vacuum. *Phys.Lett.*, B716:214–219, 2012. 9

[26] Serguei Chatrchyan et al. Determination of the top-quark pole mass and strong coupling constant from the t t-bar production cross section in pp collisions at $\sqrt{s} = 7$ TeV. *Phys.Lett.*, B728:496–517, 2014. 9

[27] U. De Sanctis. Measurements of the top quark charge asymmetry at the LHC. 2012. 9

[28] Serguei Chatrchyan et al. Measurements of $t\bar{t}$ spin correlations and top-quark polarization using dilepton final states in pp collisions at $\sqrt{s} = 7$ TeV. *Phys.Rev.Lett.*, 112(18):182001, 2014. 9

[29] Timo Antero Aaltonen et al. Combination of measurements of the top-quark pair production cross section from the Tevatron Collider. *Phys.Rev.*, D89(7):072001, 2014. 10

[30] J. Andrea. Precise top quark cross-section results at LHC. *J.Phys.Conf.Ser.*, 452(1):012025, 2013. 10

[31] CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update. Technical Report CMS-PAS-LUM-13-001, CERN, Geneva, 2013. 19

[32] Betty Bezverkhny Abelev et al. Performance of the ALICE Experiment at the CERN LHC. *Int.J.Mod.Phys.*, A29:1430044, 2014. 21

[33] Roel Aaij et al. LHCb Detector Performance. *Int.J.Mod.Phys.*, A30(07):1530022, 2015. 21

[34] G. Aad et al. Performance of the ATLAS Detector using First Collision Data. *JHEP*, 1009:056, 2010. 21

[35] D. J. A. Cockerill. The CMS Electromagnetic Calorimeter at the LHC. *ArXiv e-prints*, October 2008. 26

[36] CMS Electromagnetic Calorimeter Group, P. Adzic, N. Almeida, D. Andelin, I. Anicin, Z. Antunovic, R. Arcidiacono, M. W. Arenton, E. Auffray, S. Argiro, and et al. Radiation hardness qualification of $PbWO_4$ scintillation crystals for the CMS Electromagnetic Calorimeter. *Journal of Instrumentation*, 5:3010, March 2010. 26

[37] S. Abdullin et al. Design, performance, and calibration of CMS hadron-barrel calorimeter wedges. *Eur.Phys.J.*, C55:159–171, 2008. 28

[38] The CMS Collaboration. The performance of the CMS muon detector in proton-proton collisions at sqrt(s) = 7 TeV at the LHC. *JINST*, 8:P11002, 2013. 30, 31

[39] Performance of muon identification in pp collisions at s**0.5 = 7 TeV. Technical Report CMS-PAS-MUO-10-002, CERN, 2010. Geneva, 2010. 31, 47

[40] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, and Tim Stelzer. MadGraph 5 : Going Beyond. *JHEP*, 1106:128, 2011. 39

[41] Carlo Oleari. The POWHEG-BOX. *Nucl.Phys.Proc.Suppl.*, 205-206:36–41, 2010. 39

[42] Stefano Frixione and Bryan R. Webber. Matching nlo qcd computations and parton shower simulations. *Journal of High Energy Physics*, 2002(06):029, 2002. 39

[43] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 0605:026, 2006. 39

[44] G. Corcella, I.G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, et al. HER-WIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes). *JHEP*, 0101:010, 2001. 39

[45] S. Agostinelli et al. GEANT4: A Simulation toolkit. *Nucl.Instrum.Meth.*, A506:250–303, 2003. 39

[46] Ringaile Placakyte. Parton Distribution Functions. 2011. 40

[47] J. Pumplin, D.R. Stump, J. Huston, H.L. Lai, Pavel M. Nadolsky, et al. New generation of parton distributions with uncertainties from global QCD analysis. *JHEP*, 0207:012, 2002. 40

[48] Pierre Artoisenet, Rikkert Frederix, Olivier Mattelaer, and Robbert Rietkerk. Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations. *JHEP*, 1303:015, 2013. 41

[49] Johan Alwall, Stefan Hoche, F. Krauss, N. Lavesson, L. Lonnblad, et al. Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions. *Eur.Phys.J.*, C53:473–500, 2008. 41

[50] Peter Zeiler Skands. Tuning Monte Carlo Generators: The Perugia Tunes. *Phys.Rev.*, D82:074018, 2010. 42, 128

[51] Ttbar cross section twiki page. 43

[52] Single top cross section twiki page. 43

[53] Standard model cross section twiki page. 43

[54] Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET. Technical Report CMS-PAS-PFT-09-001, CERN, 2009. Geneva, Apr 2009. 44, 45

[55] Electron reconstruction and identification at sqrt(s) = 7 TeV. Technical Report CMS-PAS-EGM-10-004, CERN, Geneva, 2010. 48

[56] Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector. Technical Report CMS-PAS-PFT-10-001, 2010. 49

[57] G. Soyez. *The SISCone and anti-k_t jet algorithms.* Apr 2008. 49

[58] Yuri L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber. Better jet clustering algorithms. *JHEP*, 08:001, 1997. 50

[59] Gerald C. Blazey et al. Run II jet physics. In *QCD and weak boson physics in Run II. Proceedings, Batavia, USA, March 4-6, June 3-4, November 4-6, 1999*, pages 47–77, 2000. 50

[60] CMS Collaboration. Determination of jet energy calibration and transverse momentum resolution in CMS. *Journal of Instrumentation*, 6:11002, November 2011. 50

[61] Performance of b tagging at sqrt(s)=8 TeV in multijet, ttbar and boosted topology events. Technical Report CMS-PAS-BTV-13-001, CERN, Geneva, 2013. 53

[62] Top reference event selection twiki. https://twiki.cern.ch/twiki/bin/view/CMS/TopEventSelectionRun1. 56, 70

[63] J. Goh et al. Electron efficiency measurement for top quark physics at sqrt(s) = 8 tev. 2012. 58

[64] Effective area corrections twiki. https://twiki.cern.ch/twiki/bin/view/CMS/EgammaEARhoCorrection. 59, 60

[65] Vardan Khachatryan et al. Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV. 2015. 59, 65

[66] Commissioning of the Particle-Flow reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV. Technical Report CMS-PAS-PFT-10-002, CERN, Geneva, 2010. 61

[67] Scott Stuart Snyder. Measurement of the top quark mass at D0. 1995. 62

[68] S. Sumowidagdo. Programming documentation of HitFit: A kinematic fitter for top quark-antiquark pair lepton+jets events. 2011. 62

[69] S. Dildick. Application of kinematic fitting to top quark mass reconstruction in the μ+jets channel at sqrt(s) = 7 TeV with the CMS detector. 2012. 62, 63

[70] Pile-up reweighing twiki. https://twiki.cern.ch/twiki/bin/view/CMS/PileupMCReweightingUtilities. 66

[71] Serguei Chatrchyan et al. Measurement of differential top-quark pair production cross sections in pp colisions at $\sqrt{s} = 7$ TeV. *Eur.Phys.J.*, C73(3):2339, 2013. 69

[72] Vardan Khachatryan et al. Measurement of the differential cross section for top quark pair production in pp collisions at sqrt(s) = 8 TeV. 2015. 69

[73] Measurement of differential top-quark pair production cross sections in the lepton+jets channel in pp collisions at 8 TeV. Technical Report CMS-PAS-TOP-12-027, CERN, Geneva, 2013. 69

[74] Nikolaos Kidonakis. NNLL threshold resummation for top-pair and single-top production. *Phys.Part.Nucl.*, 45(4):714–722, 2014. 69

[75] Top p_t reweighing twiki. https://twiki.cern.ch/twiki/bin/viewauth/CMS/TopPtReweighting. 69

[76] Frank Fiedler, Alexander Grohsjean, Petra Haefner, and Philipp Schiefer-decker. The Matrix Element Method and its Application in Measurements of the Top Quark Mass. *Nucl.Instrum.Meth.*, A624:203–218, 2010. 82

[77] M. F. Canelli. Helicity of the W Boson in single lepton $t\bar{t}$ events. 2003. 86

[78] Pierre Artoisenet, Vincent Lemaitre, Fabio Maltoni, and Olivier Mattelaer. Automation of the matrix element reweighting method. *JHEP*, 1012:068, 2010. 87

[79] Lhco description webpage. http://madgraph.phys.ucl.ac.be/Manual/lhco.html. 88

[80] P. Haefner. Measurement of the top quark mass with the matrix element method in the semileptonic decay channel at dø. 91

[81] L Demortier and L Lyons. Everything you always wanted to know about pulls. Technical Report CDF/ANAL/PUBLIC/5776, CDF, February 2002. 95

[82] R. Barlow. Application of the bootstrap resampling technique to particle physics experiments. *Preprint MAN-HEP-99-4*, 2000. 96

[83] First application of matrix element method for top quark mass measurement in di-muon channel, with $36pb^{-1}$ at cms. Technical Report CMS AN-2011/120. 100

[84] Toplhc working group twiki. https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TopLHCWG. 124

[85] A. Heister et al. Study of the fragmentation of b quarks into B mesons at the Z peak. *Phys.Lett.*, B512:30–48, 2001. 129

[86] J. Abdallah et al. A study of the b-quark fragmentation function with the DELPHI detector at LEP I and an averaged distribution obtained at the Z Pole. *Eur.Phys.J.*, C71:1557, 2011. 129

[87] J. Pumplin, D. Stump, R. Brock, D. Casey, J. Huston, et al. Uncertainties of predictions from parton distribution functions. 2. The Hessian method. *Phys.Rev.*, D65:014013, 2001. 130

[88] D Bourilkov, R C Group, and M R Whalley. LHAPDF: PDF use from the Tevatron to the LHC. 2006. 130

[89] Jet energy correction uncertainty sources twiki. https://twiki.cern.ch/twiki/bin/viewauth/CMS/JECUncertaintySources#2012_JEC_uncertainty_correlation. 133, 136

[90] Combination of the CMS top-quark mass measurements from Run 1 of the LHC. Technical Report CMS-PAS-TOP-14-015, CERN, Geneva, 2014. 135, 141

[91] Louis Lyons, Duncan Gibaut, and Peter Clifford. How to Combine Correlated Estimates of a Single Physical Quantity. *Nucl.Instrum.Meth.*, A270:110, 1988. 139

[92] T. Aaltonen et al. Combination of the top-quark mass measurements from the Tevatron collider. *Phys.Rev.*, D86:092003, 2012. 139

[93] LHC Combination: Top mass. Technical Report CMS-PAS-TOP-12-001, CERN, Geneva, 2012. 139

[94] Top mass combination. Technical Report CMS-PAS-TOP-11-018, CERN, Geneva, 2012. 140, 141

[95] Andrea Valassi and Roberto Chierici. Information and treatment of unknown correlations in the combination of measurements using the BLUE method. *Eur.Phys.J.*, C74:2717, 2014. 141

[96] Fraghad task force twiki page. 146