



DIFFERENTIAL CROSS-SECTION
MEASUREMENT OF
TOP-QUARK-PAIR EVENTS AS A
FUNCTION OF THE NUMBER OF
JETS WITH THE CMS DETECTOR

Zur Erlangung des akademischen Grades eines
DOKTORS DER NATURWISSENSCHAFTEN
von der Fakultät für Physik
des Karlsruher Instituts für Technologie (KIT)
genehmigte

DISSERTATION

von

Dipl.-Phys. Alexis Lucas Descroix

aus Grenoble (Frankreich)

Referent: Prof. Dr. Ulrich Husemann
Korreferent: Prof. Dr. Thomas Müller
Tag der mündlichen Prüfung: 3. Juli 2015

DOI: 10.5445/IR/1000049462

Declaration of authorship

I declare that I have developed and written the enclosed thesis completely by myself, and have not used sources or means without declaration in the text.

Selbstständigkeitserklärung

Ich versichere wahrheitsgemäß, die Arbeit selbstständig angefertigt, alle benutzten Hilfsmittel vollständig und genau angegeben und alles kenntlich gemacht zu haben, was aus Arbeiten anderer unverändert oder mit Abänderungen entnommen wurde.

Alexis Lucas Descroix

Karlsruhe, den 9. September 2015

Introduction

Today's discovery is tomorrow's
background.

R. Shrock, "Neutrinos and implications
for physics beyond the standard model".
Proceedings, Conference, Stony Brook,
USA, October 11-13, 2002.

Research in the field of elementary particle physics attempts to understand the behavior of Nature at the smallest scales. Nowadays, decades of experimental and theoretical work and technical progress have brought our knowledge to a high level of precision. Therefore, current research in particle physics focuses on production of rare particles and unusual properties.

Rare findings mean at the same time that they only have a low impact on human beings' everyday life, mostly limited to our better understanding. This rises a defensible question: why should we pursue research in this field if today we cannot make anything out of our findings?

Knowing the building blocks of our universe and their interactions better leads to a deeper understanding of the world we live in. The urge to address such a fundamental question is, in my opinion, what makes us special within Nature. We *want* to know how it works, why it works, etc. Inheritors of this knowledge, rather in a long time, might find applications. Today, we are grateful to elementary particle physics of the nineteenth century for discovering the electron, nowadays used everywhere.

In 1995, the first observation of the top quark was a great achievement for particle physics because it completed the picture of elementary components of matter and rewarded its successful predictions. Since the top quark has a huge mass, the first measurement of the top quark has been extremely challenging, requiring enormous human and material investments. Twenty years later, the top quark is established as the heaviest known elementary particle. The accurate measurement of its properties has become possible although it remains a challenging task. However, it is also appealing because

its extreme mass, the second most massive quark is approximately 40 times lighter, makes it unique:

- The particle discovered in Summer 2012, the Higgs boson, is also very heavy but still lighter than the top quark. It interacts preferentially with massive particles, and therefore the top quark is its favorite partner. This coupling is a fundamental property whose measurement requires accurate knowledge about the top quark.
- As the most massive particle, the top quark is situated at the frontier to the unknown physics. New physics, potentially made of particles carrying even larger masses might couple strongly to the top quark.
- Is the theory of particle physics valid at much higher energies, like at the big bang? Or is our current understanding of particle physics only a simplification at low energy of a theory which unifies interactions at high energy? If deviations from our expectations should appear, then it is expected to take place at higher energy scales, and hence rather with massive particles, like the top quark.

Thanks to the large number of proton-proton collisions provided by the largest and most powerful particle-collision facility in the world, the Large Hadron Collider, whose results are detected by the Compact Muon Solenoid experiment, many top-quark production events are available. Hence, the accurate investigation of the top quark can be addressed with fantastic precision. This takes place via a large number of studies, measuring a large spectrum of properties related to the top quark, its production and decay features. One of these properties is the capacity of top quarks to be produced together with additional emissions of particles which are observed via particle jets. The probability for producing such additional jets in top-quark pair events is measured in this thesis, presented as a differential cross-section measurement.

Due to the complexity of the theory of particle physics, numerical tools have to include various approximations in order to simulate processes like the top-quark pair production with additional emissions. The results of the measurement are compared to predictions from several simulation tools, which achieves the confrontation of these approximations to the reality. Moreover, results are presented in a way to enable their reproduction by other experiments and the theory community.

In the first part of this thesis, all required theoretical, technical, and experimental fundamentals are presented. The theory basics are given in Chapter 1. Thanks to the theory of particle physics, the simulation of the signal

and the background events expected in proton-proton collisions is possible. This is presented in Chapter 2. The measurement itself is performed with data delivered by the LHC. This and the detector required in order to record the collision results analyzed in this thesis are presented in Chapter 3. The content of the recorded data needs to be converted into physics objects like particles or jets. This is described in Chapter 4. Lastly, the analysis tools used in this thesis are explained in Chapter 5.

In the second part, a summary of the publication achieved during the PhD period of the author of these lines, a measurement of the number of jets in top-quark pair events, is given in Chapter 6. Chapter 7 gives an overview of the analysis performed in the next chapters. The analysis setup is given in Chapter 8, which is made of the description of the event selection, the MC-simulation samples used, and corrections applied. A data-based estimation of background contributions is detailed in Chapter 9 and the resulting yield of events in data and the expected signal and background yields given by the modeling are shown in Chapter 10 together with control distributions. Finally, the actual measurement is presented in Chapter 11. This contains the calculation of the results, the presentation of the results, and the comparison of the results to various predictions.

Contents

I	Theoretical and Technical Foundations	1
1	Theory - The Standard Model of Elementary Particle Physics	3
1.1	The Elementary Particles	4
1.1.1	Leptons and Quarks	4
1.1.2	Bosons	5
1.2	Particle Properties	6
1.3	Theoretical Foundations of the Standard Model	8
1.3.1	From Classical Point Mechanics to a Quantum Field Theory for the Standard Model	9
1.3.2	Calculations and Feynman Diagrams	11
1.3.3	Quantum Chromodynamics	14
1.3.4	Further Interactions	17
1.4	The Top Quark	19
1.4.1	Top-Quark Production	19
1.4.2	Top-Quark Decay and Signature	20
1.4.3	Main Properties of the Top Quark	22
1.4.4	State of the Art for Cross Section of Top-Quark Pair Events with Jets	24
1.4.5	Top-Quark Pair Events with Jets – Beyond the Standard Model	26
2	Simulation of Data	27
2.1	Parton Distribution Function	27
2.2	Matrix-Element Event Generators	29
2.2.1	Matrix-Element Computation	30
2.2.2	Examples of Matrix-Element Generators	31
2.3	General-Purpose MC Event Generation	32
2.3.1	Parton Showering and Hadronization	33
2.3.2	Examples of General-Purpose MC Event Generators	34

2.3.3	Matching between Matrix-Element Generation and Parton Showering and Hadronization	34
2.4	Detector Simulation	35
3	Experiment - Particle Collider and Detector	37
3.1	Large Hadron Collider at CERN	37
3.2	The Compact Muon Solenoid Experiment	41
3.2.1	Tracking system	43
3.2.2	Calorimeters	47
3.2.3	Superconducting Solenoid	50
3.2.4	Muon System	51
3.2.5	Trigger and Data-Acquisition System	51
3.3	Pile-up	52
4	Reconstruction of Physics Object	55
4.1	Global Event Description: Particle Flow	55
4.2	Tracks	56
4.3	Vertices	57
4.4	Electrons and Photons	57
4.5	Muons	58
4.6	Jets	60
4.7	Missing Transverse Energy	62
5	Analysis Techniques	65
5.1	Maximum-Likelihood Method	65
5.2	Data Unfolding	67
II Study of the Jet Production in Top-Quark Pair Events with CMS		71
6	Summary of a Jet-Multiplicity Measurement with Data from 2011	73
6.1	Measurement of the Jet Multiplicity	74
6.2	Multiplicity of Additional Jets	76
6.3	Conclusion and Outlook	79
7	Analysis Strategy	81
8	Analysis Setup	83
8.1	Event Selection	83
8.1.1	Selection of Electrons	85

8.1.2	Selection of Muons	88
8.1.3	Selection of Jets	89
8.1.4	Detector-Level Selection	90
8.1.5	Particle-Level Selection	91
8.2	Data and MC-Simulation Samples	94
8.3	Corrections of Data and MC Simulation	96
8.3.1	Jet Energy	96
8.3.2	Pile Up	97
8.3.3	Top- p_T Spectrum	98
8.3.4	Lepton Reconstruction, Selection, and Trigger Efficiencies	100
8.3.5	Modeling of the b -Tagging Efficiency in MC	100
9	Data-Based Background Estimation	103
9.1	Extraction of the QCD-Multijet Model from Data	104
9.1.1	Lepton Within a Jet	104
9.1.2	Fake Lepton	105
9.1.3	Choice of the QCD-Multijet Model	105
9.1.4	Validation of the QCD-Multijet Model	105
9.2	Categorization of W +Jets Events	106
9.3	Simultaneous Template Fit in Sideband	106
9.4	Results and Discussion	111
10	Event Yields and Control Plots	115
11	Jet Multiplicity in Top-Quark Pair Production	121
11.1	Conversion of the Jet-Multiplicity Spectrum to the Particle Level in a Fiducial Phase Space	123
11.2	Measurement of the Top-Quark Pair Differential Cross Section as a Function of the Number of Jets	128
11.3	Systematic Uncertainties	129
11.3.1	Detector-Related Systematic Uncertainties	130
11.3.2	Modeling-related systematic uncertainties	131
11.3.3	Background-Related Systematic Uncertainties	133
11.4	Results and Discussion	135
11.5	Comparison with Predictions from Further MC Tools	140
11.6	Comparison with Earlier Measurements	143
12	Conclusion and Outlook	147

III Appendix and Bibliography	151
A Technical List of Samples	153
B Control plots including systematic uncertainties	157
C Control Plots	159
Bibliography	159

Part I

Theoretical and Technical Foundations

Chapter 1

Theory - The Standard Model of Elementary Particle Physics

Elementary particle physics enables us to understand Nature at the smallest scales. Nevertheless, it aims at answering questions at much larger scales, like: What is the universe made of? What is its origin? What makes matter behave the way it is observed? This is possible because elementary particles are the constituents of all known matter and some of them are the carriers of fundamental forces.

A theory makes use of these particles for the description of many observed phenomena: the Standard Model (SM) of elementary particle physics. It is the result of almost a century of predictions and experimental tests. Nowadays it is tested by measurements with great precision. This chapter presents the theoretical foundations of elementary particle physics, following the approach of an experimental particle physicist: start in Sec. 1.1 with the description of the elementary particles which have been observed so far and then mention their properties in Sec. 1.2. The underlying theory describing the dynamics and interactions of these particles (with a focus on the context of this thesis) is presented in Sec. 1.3. Finally, more details are given about the top quark in Sec. 1.4.

1.1 The Elementary Particles

Elementary particles are fundamental building blocks, treated as point-like objects in their theoretical description. Furthermore, some of them have an extremely short lifetime. Hence, their observation can be very challenging. Certain particles can only be observed indirectly; therefore, their existence needs to be postulated¹. They are called elementary because they are assumed to have no substructure, and therefore to constitute the “bricks” of matter and interactions. Only collective phenomena of these bricks can be observed at larger length scales. However, gravitation could not yet be included in the SM. Thus, since large scale phenomena are strongly influenced by gravitation, the application of the SM is so far limited to the microscopic world.

The SM is made of different types of particles: electrically-charged and neutral leptons, quarks, and bosons. They are presented in the rest of this section.

1.1.1 Leptons and Quarks

Six lepton types exist, ordered in three pair or generations. Each generation consists of an electrically-charged particle and its neutrino (electrically neutral): the electron and the electron neutrino, the muon and the muon neutrino, and the tau and the tau neutrino. This organization is due to special relations between both members of a pair when interacting with the weak interaction, see Sec. 1.3.4.

Due to different lifetimes and masses, the three electrically-charged leptons leave a very different signature when they are detected in a high energy physics experiment like CMS. Electrons are stable particles but because of their very small mass, they produce synchrotron radiation when accelerated (for example by a magnetic field) and lose energy through bremsstrahlung in dense media like a particle detector. Muons are on the one hand more massive than electrons, and therefore they produce less synchrotron radiation and interact less in the detector. On the other hand, they are unstable but live long enough ($\approx 2 \mu\text{s}$ [1]) to traverse a whole detector when produced with relativistic energies e.g. in a particle collider. Taus are even more massive than muons so they have a shorter lifetime ($\approx 3 \cdot 10^{-13}\text{s}$ [1]). Thus,

¹This epistemological precaution stems from the dilemma: “Do I see an elementary particle (direct), or the phenomena following the presence of something I like to call elementary particle (indirect)?”. Both possibilities are compatible with this thesis.

they decay right after their production. Finally, neutrinos are very special particles because their interaction with material is extremely faint. As a consequence they are mostly “invisible”. Dedicated experiments (e.g. large water Čerenkov detectors) can detect them via their decay products but this is beyond the scope of this thesis. Table 1.1 summarizes the properties of leptons.

Six types of quarks are known: the down quark, the up quark, the strange quark, the charm quark, the bottom quark, and the top quark, corresponding to six quark flavors. Table 1.1 summarizes their properties. The two first quarks, the up and down quarks, are the constituents of protons and neutrons. Therefore, they are very abundant in the universe. More massive quarks are unstable and decay very fast into lighter elementary particles. As a consequence they are rare and their observation is only possible when they are produced artificially in large amounts at high energy facilities, like the LHC (see Sec. 3.1). The most massive known elementary particle, the top quark, has been discovered in 1995 [2, 3]. Its observation is very challenging but also appealing since its mass is at the frontier of the unknown of high energy physics. The investigation of specific properties of the top-quark production at the LHC is the main topic of this thesis. This particle is introduced with more details in Sec. 1.4.

1.1.2 Bosons

Gauge bosons are elementary particles mediating interactions between other elementary particles, they are summarized in Table 1.2. The photon mediates the electromagnetic interaction, e.g. the repulsion between two electrons is seen in elementary particle physics as an exchange of a virtual photon between them. Similarly, the strong interaction, for instance holding together quarks in a proton nucleus, is mediated by gluons. The W and Z bosons mediate the weak interaction, for example the flavor changing of quarks inside of a neutron, necessary for its decay into a proton. The W and Z bosons together with the photon are the observed bosons of the electroweak (EW) interaction which unifies the electromagnetic and the weak interactions. Finally, the Higgs boson, certainly corresponding to the new boson discovered in summer 2012 [4, 5], should couple to the mass of particles.

The organization of the particles as shown in this section follows a logical structure based on properties of elementary particles. These properties are explained in Sec. 1.2.

Table 1.1: The three quark generations and the three lepton generations, each consisting of an up-type quark with a down-type quark, and of an electrically-charged lepton with the corresponding neutrino. Electric charges (Q) are given in units of the elementary electric charge $e = 1.602176487(40) \times 10^{-19}$ C [1]. As the lepton and quark masses are not predicted by the SM, the given values are obtained from experimental measurements summarized in [1].

	Generation	Flavor	Symbol	Q [e]	Mass [MeV/ c^2]
Leptons	1	electron	e	-1	0.511
		electron neutrino	ν_e	0	$< 2 \cdot 10^{-6}$ *
	2	muon	μ	-1	105.7
		muon neutrino	ν_μ	0	< 0.19 *
	3	tau	τ	-1	1776.82 ± 0.16
		tau neutrino	ν_τ	0	< 18.2 *
Quarks	1	down	d	$-1/3$	$4.8_{-0.3}^{+0.5}$
		up	u	$2/3$	$2.3_{-0.5}^{+0.7}$
	2	strange	s	$-1/3$	95 ± 5
		charm	c	$2/3$	$(1.275 \pm 0.025) \cdot 10^3$
	3	bottom	b	$-1/3$	$(4.18 \pm 0.03) \cdot 10^3$
		top	t	$2/3$	$(173.07 \pm 0.52 \pm 0.72) \cdot 10^3$

*One should keep in mind that neutrinos are assumed to be massless in the SM.

1.2 Particle Properties

A convention in particle physics is to use natural units. As a consequence the speed of light, the reduced Planck constant, and the Boltzmann constant are equal to one: $c = \hbar = k_B = 1$. This convention is applied the remainder of this thesis.

Mass The concept of the rest mass of a particle seems quite natural. However, in an attempt to describe how this property is given to each particle, the successful Higgs theory introduces complex mechanisms in which the mass of elementary particles arises from their coupling with a new field, the Higgs field. The more a particle couples to this field, the more massive it is. More on this can be found in Sec. 1.3.

Electric Charge The electric charge of particles is also a very basic concept. It happens that all observed particles carry an integer multiple of the electric charge of the electron e (the electron is taken as a reference since it

Table 1.2: The gauge bosons of the main fundamental interactions and the Higgs boson. The electric charge is symbolized by Q . The mass values are taken from [1] except the Higgs boson mass, which is from [6].

Name	Symbol	Mediated interaction	Q [e]	Spin	Mass [GeV/c^2]
gluon	g	strong	0	1	0
photon	γ	electroweak	0	1	0
W bosons	W^\pm	electroweak	± 1	1	80.398 ± 0.025
Z boson	Z^0	electroweak	0	1	91.188 ± 0.002
Higgs boson	H	-	0	0	125.09 ± 0.24

is the first observed elementary particle). Quarks carry indeed integer multiples of one third of e but they are always observed in bound states which in turn always carry an integer multiple of e .

Each type of electrically-charged particle has a copy, an anti-particle, with the opposite electric charge. For example the electron carries a negative electric charge ($-e$) and the anti-electron, called positron, carries the opposite charge (e).

If not specified differently, “charge” refers to the electric charge in this thesis.

Spin Elementary particles are seen as point-like so they cannot rotate or deform. However, a measured property of elementary particles behaves like a quantized angular momentum. Therefore, this property is called the spin S . The square of the spin S^2 is an intrinsic property of an elementary particle, leading to fixed values allowed for its projection on a given quantization axis. This projection (conventionally on the z axis S_z) can only take values separated by a quantum, e.g. an electron can only take the values $S_z = -\hbar/2$ or $\hbar/2$, where \hbar is the reduced Planck constant: $\hbar \equiv h/(2\pi) \approx 1.055 \cdot 10^{-34}$ Js. By the measurement of the spin of a particle on the z axis, two types of particles are observed: Fermions show half integer values of \hbar (e.g. $-\frac{3}{2}\hbar$, $-\frac{1}{2}\hbar$, $\frac{1}{2}\hbar$, $\frac{3}{2}\hbar$, etc ...), whereas bosons show integer values of \hbar (e.g. $-2\hbar$, $-1\hbar$, 0 , $1\hbar$, $2\hbar$, etc ...).

Weak Isospin In general, isospin is the name given to a quantum number carried by particles showing pair-like properties. The name isospin is not related to any angular momentum but to the behavior of certain particle pairs which is similar to spin-1/2 systems. In the scope of this thesis it is

the weak isospin, related to the weak interactions. All left-handed² fermions and all right-handed anti-fermions carry a weak isospin charge which is the projection T_3 of the isospin on a given axis. All up-type quarks (up, charm, and top) carry the isospin charge $+1/2$ and it is $-1/2$ for all down-type quarks (down, strange, and bottom). Moreover, neutrinos carry an isospin charge $+1/2$ and charged leptons (e^- , μ^- , and τ^-) carry an isospin charge $-1/2$. Finally, the W^+ and W^- boson carries respectively the isospin charge $T_3 = +1$ and -1 .

Through flavor changing, as briefly presented in Sec. 1.3.4, a quark can transform to another one, or a charged lepton into a neutrino, or vice versa. This kind of interaction is controlled by the conservation of the weak isospin charge whose value can change (between $+1/2$ and $-1/2$) when emitting a W^\pm boson.

Color Color is a quantum number carried by quarks and gluons, the color-charged particles. It was introduced in order to describe some of their properties due to the strong interaction. More on this topic can be found in Sec. 1.3.3. Three colors are possible (by convention: blue, green, and red) and quarks carry one of these colors, anti-quarks carry an anti-color whereas gluons carry a combination of colors and anti-colors.

The color charge enables an intuitive understanding of a complex property of the strong interaction: the confinement. In elementary particle physics only color neutral objects can be observed. Therefore, quarks cannot be observed alone but only in bound states. Indeed in bound states, e.g. a proton, the three quarks can carry the three colors, so the sum of the colors is white, or neutral. Three-quark bound states are called baryons, whereas mesons are states made of a quark and an anti-quark carrying a color and the corresponding anti-color respectively.

1.3 Theoretical Foundations of the Standard Model

The underlying mathematical structure of the SM is a relativistic quantum field theory (QFT). The following short introduction to this kind of theories and the subsequent exemplary explanations on its applications for the SM

²This is the chirality, an intrinsic particle property. Massless particles are left or right handed if the projection of their spin on their momentum is positive or negative.

of particle physics are based on the lectures given by M. Dasgupta [7] and T. Ohl [8].

1.3.1 From Classical Point Mechanics to a Quantum Field Theory for the Standard Model

The goal is to describe the mechanics of elementary particles. At first glance, classical mechanics seems to address this elegantly: a particle is described like a mathematical point whose mass m and acceleration $\ddot{x}(t)$ at instant t are related to F , the sum of all forces applied to this particle, following Newton's law:

$$m\ddot{x}(t) = F,$$

where for simplicity, only one dimension is shown. For convenience, one prefers the equivalent description given by the Lagrangian formalism in the derivation of a field theory. The Lagrange function, the difference between the kinetic and the potential energy:

$$L(x, \dot{x}) = T - V = \frac{1}{2}m\dot{x}^2 - V,$$

is used in order to define the action:

$$S = \int_{t_0}^{t_1} L(x, \dot{x}) dt,$$

the integral of the Lagrange function over a given time interval t_0 to t_1 . The action is minimized for particle trajectories satisfying Newton's law, following the principle of least action.

The last quantity to be introduced is the Hamiltonian H , corresponding to the total energy of the particle:

$$H(x, p) = p\dot{x} - L(x, \dot{x}) = T + V,$$

providing an equivalent description of particle dynamics if it fulfills Hamilton's equations:

$$\frac{\partial H}{\partial x} = -\dot{p}, \quad \frac{\partial H}{\partial p} = \dot{x}.$$

Since these particles are extremely small, classical mechanics needs to be replaced by quantum mechanics. Performing the canonical quantization, one transforms dynamical variables into operators (e.g. $p_i \rightarrow \hat{p}_i = -i\hbar\frac{\partial}{\partial x_i}$) followed with commutation relations.

The description of the state of a system as a particle is then given by a state vector: $|\psi\rangle$ and its projection on the coordinate space gives: $\langle x|\psi\rangle = \psi(x)$ the wave function of the particle. Its time evolution is obtained by applying the Hamiltonian operator \hat{H} , following Schrödinger's equation:

$$i\hbar\frac{\partial}{\partial t}\psi(x,t) = \hat{H}\psi(x,t).$$

A solution for a given initial wave function $\psi(x,t_0)$ at time t_0 and with a constant Hamiltonian operator is:

$$\psi(x,t) = e^{-\frac{i}{\hbar}\hat{H}(t-t_0)}\psi(x,t_0).$$

The next step of the construction of a QFT is the expansion of the single-particle description to the field description. Instead of a single particle in motion, a field $\phi(x,t)$ describes the state of the system at any time-space point. Elementary particle physics is pursued at high energy, making relativistic effects non-negligible, considering the relatively low mass of particles. After it is included in the QFT as well, the Lagrange density is rewritten as a function of the field ϕ and its four-vector derivative $\partial^\mu\phi$ as follows: $\mathcal{L}(\phi, \partial^\mu\phi)$ which is given by the function $L(\phi, \dot{\phi}) = \int d^3x\mathcal{L}(\phi, \partial^\mu\phi)$ and the action is now: $S = \int d^4x\mathcal{L}(\phi, \partial^\mu\phi)$. Minimizing again this action leads to the Klein-Gordon equation:

$$(\square + m^2)\phi(x) = 0.$$

While this is valid for a scalar field, i.e. particles without spin, one has to introduce Dirac's formalism in order to describe both spin states of spin-1/2 particles as well as their antiparticle. Dirac's equation is:

$$(i\gamma^\mu\partial_\mu - m)\psi(x) = 0,$$

where γ^μ are Dirac's matrices. In Dirac's representation they are:

$$\gamma^0 = \begin{pmatrix} \mathbf{1} & 0 \\ 0 & -\mathbf{1} \end{pmatrix}, \quad \gamma^i = \begin{pmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{pmatrix},$$

using Pauli's matrices σ^i .

The extension of the Klein-Gordon equation to spin-1 particles like photons and W and Z bosons is done by defining the four-potential $A^\mu = (\phi, \vec{A})$, where ϕ is the electric field potential and \vec{A} is the magnetic vector potential. Both fields are related via Maxwell's equations. Proca's equation:

$$\partial_\mu(\partial^\mu A^\nu - \partial^\nu A^\mu) + m^2 A^\nu = 0$$

describes the dynamic behavior of massive spin-1 particles.

Finally, the introduction of interactions succeeds with new symmetries under a certain kind of gauge transformation, in other words by making the Lagrange function invariant under a gauge transformation which depends on the point in space and time where it is applied (local gauge transformation). Each interaction results from a certain symmetry group and is mediated by means of the corresponding gauge bosons. The symmetry leading to the introduction of the color is presented in Sec. 1.3.3.

The description of particle behavior by means of the calculation of the probability for a certain phenomenon to happen, e.g. a simple free particle propagation or an interaction between particles (radiation, absorption, decay, or annihilation), is described in the next section.

1.3.2 Calculations and Feynman Diagrams

In the scope of this thesis, the probability for a given particle interaction to happen is described by the cross section. More on this quantity is given in Chapter 3. A rather general formula for its calculation for the interaction of two particles (index a and b) producing n particles is obtained with Fermi's golden rule, taken from [7]:

$$d\sigma = \frac{|T|^2}{4\sqrt{(p_a p_b)^2 - m_a^2 m_b^2}} \frac{\widetilde{dp}_1 \dots \widetilde{dp}_n}{\prod_i n_i!} (2\pi)^4 \delta^4(p_a + p_b - p_1 - \dots - p_n),$$

with:

$$\widetilde{dp}_i = \frac{d^3 \vec{p}_i}{(2\pi)^3 2p_0} \Big|_{p_0 = \sqrt{\vec{p}_i^2 + m_i^2}$$

which is the infinitesimal phase-space element for the final-state particle i . The term $(2\pi)^4 \delta^4(\dots)$ keeps the overall momentum conserved. The term n_i is the number of final-state particle of the same type as the particle i . The denominator with n_i takes care of cases in which particles of the same type appear in the final state, since they are indistinguishable their permutation cannot count as additional probability. Finally, the denominator $1/(4\sqrt{(p_a p_b)^2 - m_a^2 m_b^2})$ is the flux factor which is required for the normalization of the initial-state phase space. In the case of proton-proton collisions at the LHC the initial-state particles a and b are unknown because they arise from each proton (therefore, they are commonly called partons). Only the probability to have a certain kind of parton, with certain properties like a momentum p_a , engaged in the interaction are given by parton distribution

functions, which can be measured in deep-inelastic scattering experiments. More on this topic is presented in Sec. 2.1. The observable $d\sigma$ is called the differential cross section because it describes the cross section of the process as a function of the final-state configuration. It has to be integrated over the phase space of the final-state particles in order to get an absolute cross section σ .

A challenge lies in calculating the interaction amplitude T , making use of the SM laws. Including interactions into the theory requires to use perturbation theory which leads to infinite series of terms in the calculation. These terms were successfully interpreted in the middle of the 20th century by Richard Feynman as all the possible ways for the interaction studied to take place, relating an initial state to a final state. Each of these ways is represented as a diagram called Feynman diagram. Particles propagating in space and time are represented by lines of different type, depending on their nature. Their interactions with other particles are symbolized by line crossings called vertices. An illustration can be found in Fig. 1.1.

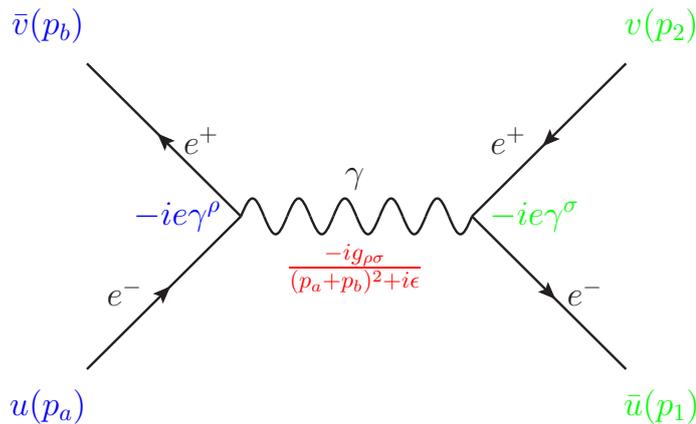


Figure 1.1: Feynman diagram of the e^+e^- scattering exchanging a photon as propagator. The colored mathematical terms are used in the exemplary calculation of the interaction amplitude iT (Eq. 1.1). Reproduction of an illustration from [8].

All Feynman diagrams have to be identified and calculated for the complete calculation of T . All possible transitions means to include all topological combinations (geometrical combinatorial and loops within the Feynman diagrams) and possible emission of particles in addition to the interaction have to be taken into account. Thus, the inclusion of all possibilities would require an infinite number of Feynman diagrams, which is impossible to calculate. However, an approximation is possible because not all the terms of the perturbation-theory expansion are significant. Hence, for a calculation one

has to decide which precision has to be achieved and then the corresponding orders of the perturbation theory need to be included in the calculation.

The leading order (LO) of perturbation theory corresponds to the contribution from Feynman diagrams with as few loops and additional emissions as possible, i.e. these are the diagrams requiring as few interaction vertices as possible. For top-quark pair production at leading order no loop needs to be included, the diagram is then called a tree-level diagram. Orders of the perturbation theory correspond to expansions in the coupling constant. Therefore, the leading-order Feynman diagrams require a minimum number of coupling constants in the calculation. Thus, if the coupling constant is small, these diagrams dominate the overall transition probability.

Diagrams containing additional particles emitted or loops must have lower probabilities because they require more coupling constants in the calculation than the leading-order diagrams. The calculation of an interaction including not only the leading-order diagrams but also those with one more loop or one more additional emission reaches the next-to-leading order (NLO) of the perturbation theory. Such a prediction performs a higher precision than the LO calculation; however, the calculation of many more Feynman diagrams including loop diagrams is required, which is challenging. Further orders of the perturbation theory can be reached, i.e. going to next-to-next-to-leading order (NNLO), etc.

Fig. 1.1 shows a typical tree-level Feynman diagram representing one of the possible ways for an electron to interact with a positron via a photon in an elastic scattering. The colored mathematical elements labeling the diagram give the terms that are taken into account for the amplitude calculation of this interaction:

$$iT = \bar{v}(p_b)(-ie\gamma^\rho)u(p_a) \frac{-ig_{\rho\sigma}}{(p_a + p_b)^2 + i\epsilon} \bar{u}(p_1)(-ie\gamma^\sigma)v(p_2), \quad (1.1)$$

which simplifies to:

$$i \frac{e^2}{(p_1 + p_2)^2} [\bar{v}(p_b)\gamma^\rho u(p_a)][\bar{u}(p_1)\gamma_\rho v(p_2)].$$

For the calculation of the cross section for this scattering, TT^\dagger is calculated taking into account all possible spin initial and final states. The result can be simplified with Mandelstam's variables:

$$s = (p_a + p_b)^2, \quad t = (p_a - p_1)^2, \quad \text{and} \quad u = (p_a - p_2)^2,$$

and, under the assumption that the mass of electrons is negligible compared to the total energy \sqrt{s} , to:

$$|T|^2 = \frac{t^2 + u^2}{s^2}.$$

1.3.3 Quantum Chromodynamics

One interaction of the SM is especially relevant for the topic of this thesis: the strong interaction, which is described by Quantum Chromodynamics (QCD). It describes the behavior of quarks and gluons, producing most jets in high-energy experiments. The present summary is based on [9].

The discovery of the Δ^{++} , a baryon with the unusual charge $+2e$ in 1951 can be seen as the first hint for something new at that time. The flavor and spin of this particle were determined as: $|\Delta^{++}\rangle = |u_\uparrow u_\uparrow u_\uparrow\rangle$, where all quantum numbers known at that time are the same for all constituents. However, since the constituents of the $|\Delta^{++}\rangle$ are fermions, they cannot be in exactly the same state. In 1965 an additional degree of freedom was introduced: the color, which allows the existence of the Δ^{++} because this degree of freedom can take three different values. The concept of color originates from a new symmetry utilized for QCD. This symmetry is described within the special unitary group of degree 3, $SU(3)$, meaning that the group operations are unitary matrices of dimension 3×3 with determinant 1. The color is a new quantum number and the charge of QCD. It can take three values, corresponding to the three degrees of the symmetry group. Conventionally the color values are blue, green, and red. This number of colors has been measured for instance with the ratio: $\sigma(e^+e^- \rightarrow q\bar{q})/\sigma(e^+e^- \rightarrow \mu^+\mu^-)$, which is proportional to the number of color-anticolor pairs among which the quark pair is allowed to choose: blue-antiblue, green-antigreen, or red-antired. The adjoint representation of $SU(3)$ is made of eight 3×3 generation matrices, called Gell-Mann matrices. They correspond to the eight gauge bosons of QCD, the gluons. These particles carry linear combinations of colors and anticolors. The conservation of color charge can be illustrated within Feynman diagrams as shown on Fig. 1.2.

The Lagrangian of QCD is:

$$\mathcal{L} = \bar{\psi}_q^i (i\gamma^\mu) (D_\mu)_{ij} \psi_q^j - m_q \bar{\psi}_q^i \psi_{qi} - \frac{1}{4} F_{\mu\nu}^a F^{a\mu\nu}.$$

The quark field of flavor q and color i is written ψ_q^i , m_q is the quark mass and $F_{\mu\nu}^a$ is the gluon field strength tensor with the gluon index a . The covariant derivative of QCD is:

$$(D_\mu)_{ij} = \delta_{ij} \partial_\mu - ig_s (t^a)_{ij} A_\mu^a,$$

where g_s is the strong coupling, A_μ^a is the gluon field, and each $(t^a)_{ij}$ is proportional to one of the eight Gell-Mann matrices $(\lambda^a)_{ij}$.

The gluon field strength tensor:

$$F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + g_s f^{abc} A_\mu^b A_\nu^c$$

shows, in addition to the two first “standard” terms responsible for gluon propagation, a last term with factors f^{abc} , the structure constants of SU(3), and the product of two gluon fields. This term enables the self interaction of the gluon field, thus leading to three and four-gluon vertices. Gluons can interact with each other because the SU(3) group is non-abelian. This has fundamental consequences for QCD.

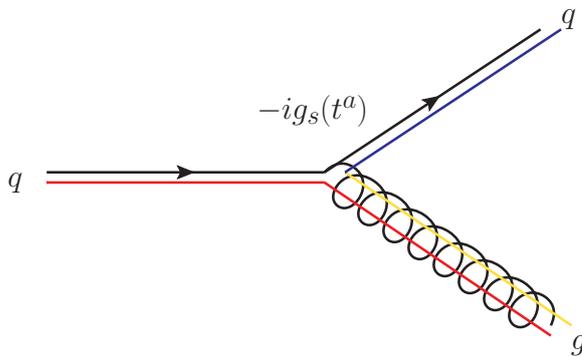


Figure 1.2: A quark carrying the red color charge has radiated a gluon and given its color this gluon. The outgoing quark carries a new color charge, blue, created at the vertex. For the conservation of the color charge, also the opposite color (yellow, symbolizing anti-blue) is carried by the gluon. The mathematical term at the vertex illustrates the contribution of the gluon radiation by a quark in calculations, described in this section.

Running of the coupling Any interaction within a QFT shows a running coupling. When zooming further, for instance into a shorter time interval, around a strong interaction like the one represented in Fig. 1.2, more virtual interactions accompanying the main one appear, as it is allowed by Heisenberg’s uncertainty principle $\Delta E \Delta t \geq \hbar$. This infinite or fractal behavior is solved in the SM by renormalizing the interactions. A consequence of this is the dependency of the strong interaction coupling on the energy scale Q at which the interaction takes place. For QCD the running is negative, which means that lower (higher) energy interactions occur with a stronger (weaker) coupling g_s , which is often used under the form $\alpha_s = g_s^2/4\pi$. The running of α_s is described by the renormalization group equation (RGE):

$$Q^2 \frac{\partial \alpha_s}{\partial Q^2} = \beta(\alpha_s),$$

where the beta function is $\beta(\alpha_s) = -\alpha_s^2(b_0 + b_1\alpha_s + b_2\alpha_s^2 + \dots)$. An approximation of α_s dependency is:

$$\alpha_s(Q^2) = \alpha_s(M_Z) \frac{1}{a + b_0\alpha_s(M_Z)\ln\frac{Q^2}{M_Z^2} + \mathcal{O}(\alpha_s^2)},$$

where b_0 is a calculable constant proportional to theoretical variables like the number of colors and the number of quark flavors. $\alpha_s(M_Z) \approx 0.12$ is the strong-coupling strength at the scale of the Z-boson mass. Measurements of α_s are shown on Fig. 1.3.

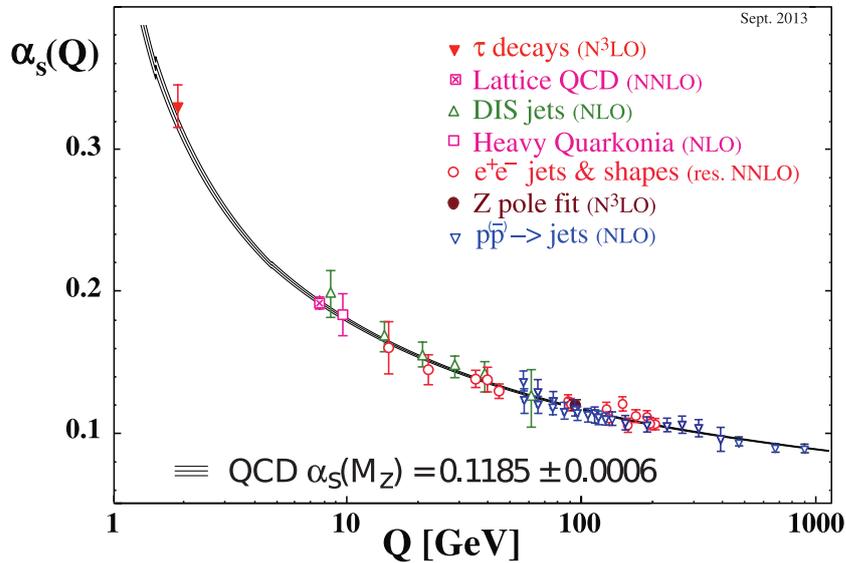


Figure 1.3: The running of α_s as a function of the energy scale Q , from [1].

Confinement and Asymptotic Freedom A consequence of the running of α_s is a very strong coupling at low energy scales. Therefore, no low energy quark or gluon can exist in a free state, they live strongly bound in hadrons. When trying to separate the valence quarks from a hadron the field strength increases like for a spring, until it is energetically more favorable to create a new hadron instead of further increasing the binding strength.

Because of the high value of α_s at low energy scales, the perturbation theory cannot be utilized for calculation of QCD under these conditions. Still, predictions are highly necessary for the modeling of processes following the production of high energy quarks and gluons. For this purpose empirical models are applied, as presented in Chapter 2.

At high energy scales α_s gets weaker, so that perturbation theory remains

valid. As a consequence, it is possible to calculate the transition amplitudes for processes including the strong interaction, as described in Sec. 1.3.2.

A quark produced at a high energy, e.g. following a proton-proton collision at CERN, is approximately free and boosted in a given direction. It emits gluons as shown in Fig. 1.2 which reduces its energy more and more. The emitted gluons can also radiate further gluons or decay into further quarks which can emit more gluons. This process is called parton showering and it leads to a large number of color-charged daughter particles sharing the energy of the original mother quark. At some point their energies become so small that α_s diverges rapidly and all color-charged particles are forced to be bound into hadrons: they hadronize. For Q at the value of $\Lambda_{\text{QCD}} \approx 200$ MeV, the strong coupling is tending to infinity, meaning that calculations of QCD, based on perturbation theory, are not valid anymore.

Particles generated by the parton shower of the original quark followed by the hadronization of these particles are mostly hadrons and decay products of short-living hadrons (mostly hadrons or sometimes leptons). Their direction is on average close to the original quark, forming a cone-shaped particle cascade. This object is called a jet. Experimentally jets were observed the first time with the SPEAR storage ring at SLAC in 1975 [10], colliding electrons against positrons at center-of-mass energies between 3 and 7.4 GeV. Only relatively stable hadrons of a jet or the various decay products can be observed in a detector. The robust reconstruction of jets by means of clustering of the energy depositions in the detector is performed via clustering algorithms, as presented in Sec. 4.6.

1.3.4 Further Interactions

The remaining interactions of the SM are briefly presented here.

Quantum Electrodynamics Charged particles can interact among each other thanks to the electromagnetic interaction. This happens by means of the exchange of a photon, as shown in Fig. 1.1. This interaction is described by Quantum Electrodynamics (QED). This is, like QCD, a QFT but based on the unitary group $U(1)$, which is abelian. As a consequence there is only one mediator of the force (the photon) which cannot interact with other photons. Another consequence is that QED is not confined: the attraction strength between an electron and a positron is proportional to the inverse square of their distance.

Weak Interaction The β^- decay of the neutron into a proton is due to the decay of one of the neutron's down valence quarks into an up valence quark. This decay is due to the weak interaction: a mediator of the weak interaction, the W^- gauge boson, is radiated by the original down valence quark. Flavor changing is possible between each quark flavor when conserving the charge (for instance a charm quark can change to a down or strange quark by emitting a W^+ boson) but changes within one generation of mass eigenstate: (u,d) , (c,s) , and (t,b) are strongly preferred and flavor changing between flavors from different generations of mass eigenstate are suppressed. Moreover, exchanges among leptons are possible, but only within lepton-flavor pairs: (e,ν_e) , (μ,ν_μ) , and (τ,ν_τ) . The W boson has a high masses, hence the weak interaction is weak (up to the scale of ~ 100 GeV), because a lot of energy is needed for its creation. This property gives free neutrons a relative long mean lifetime (~ 15 minutes [1]).

The weak interaction is described by the Quantum Flavordynamics (QFD) which is a QFT based on the special unitary group of degree two $SU(2)$.

The QFD and QED are unified into the EW theory [1]. It is based on $SU(2) \times U(1)$ whose generators correspond to four massless gauge bosons: W^1 , W^2 , W^3 , and B . The spontaneous symmetry breaking of the EW theory leads to a mixing of the massless gauge bosons, creating the observed photon and W boson: $A = B \cos \theta_W + W^3 \sin \theta_W$ and $W^\pm = (W^1 \mp iW^2)/\sqrt{2}$. A further gauge boson, the Z boson, arises from the combination: $Z = -B \sin \theta_W + W^3 \cos \theta_W$. While the W boson enables flavor changing for leptons and quarks, flavor-conserving interactions are made possible by the Z boson. Such an interaction can be illustrated when replacing the photon propagator by a Z boson in Fig. 1.1. The weak mixing angle θ_W is defined with $\cos \theta_W = \frac{m_W}{m_Z}$.

Higgs Mechanism A main enigma of the SM has been the high mass of the W and Z bosons, causing the gauge symmetry to be broken in the electroweak interaction. Moreover, calculations were showing divergences due to four-boson vertices (with W or Z), where a new scalar boson could help to solve these divergences. A theory developed by Englert, Brout, Higgs, Guralnik, Hagen, and Kibble [11–16] and nowadays known as the Higgs theory predicts an additional interaction between particles and a field, the Higgs field. This theory enables the spontaneous symmetry breaking of the electroweak theory, required in order to allow massive gauge bosons for the electroweak interaction. An excitation of the Higgs field, the Higgs boson, is also predicted and recent observations of a boson (mass: ~ 125 GeV) at

CERN in Summer 2012 [5, 17] indicate confidently the validity of this theory. The mass of fermions is included into the SM by an additional term, the Yukawa-coupling term, describing the coupling between the Higgs field and all fermions in order to predict their masses, as a function of dedicated coupling constants. This coupling is proportional to the mass of the interacting fermion.

Gravitation The gravitation, which is responsible for the attraction between massive objects, is not included in the SM because so far no attempt to quantize this interaction was successful in describing observations. This is an open questions of the SM.

1.4 The Top Quark

The confinement of quarks makes their observation as free particles impossible, contrary to e.g. electrons. They can only be studied within their bound states. The top quark is an exception, which is one of the many reasons why it is worth to be investigated. It is so massive that it decays before producing a bound state or a jet of hadrons. This unique property enables the investigation of quark properties without any bias from the hadronization. Moreover, the mass of the top quark is the highest one ever observed for an elementary particle, so the top quark might be related to unknown processes taking place at so far unreached energies. Finally, its very high mass makes it an important partner of the Higgs boson, because the latter couples through Yukawa couplings to the mass of fermions, so the top quark is the particle known today with the strongest coupling to the Higgs boson.

1.4.1 Top-Quark Production

The top quark can be produced as a top quark-antiquark pair or alone as a single top quark, where the first production channel is more likely than the second one, because in the single case the weak interaction is required to create the top flavor from another flavor. The single top quark has three main production channels called s, t, and tW , as shown with Feynman diagrams at the leading order of the perturbation theory in Fig. 1.4.

The present thesis investigates top quarks produced in pairs at the LHC. In this case, the dominant pair-production channel features gluons (gg) in the

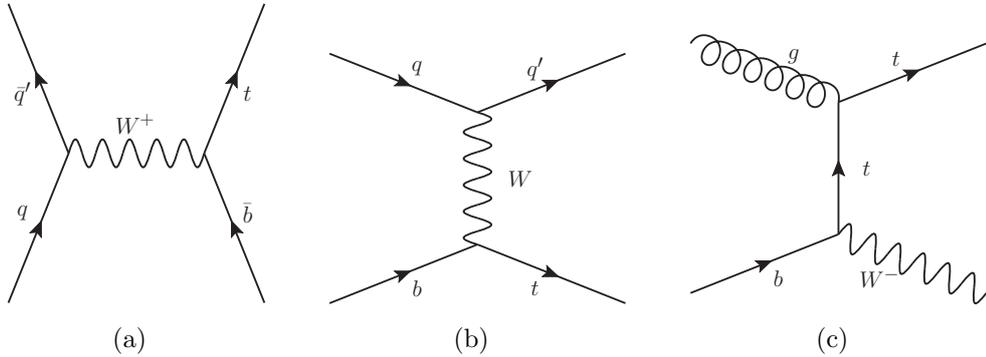


Figure 1.4: Feynman diagrams of single top-quark production in the s channel (a), the t channel (b), and the tW channel (c).

initial state, meaning that the colliding protons deliver gluons for the collision. More pair-production channels including quarks in the initial state ($q\bar{q}$ LO and gq at NLO) are possible. Illustrations of top-quark pair production are given with Feynman diagrams at the leading order of the perturbation theory in Fig. 1.5. Because of the very large center-of-mass energy provided by the LHC, quarks are less likely emitted by the colliding protons than gluons, as explained in Sec. 2.1, suppressing channels containing quarks in the initial state.

No distinction is made between the various pair-production channels in this thesis. Therefore, the pair-production cross sections given in Sec. 1.4.3 are inclusive.

1.4.2 Top-Quark Decay and Signature

Thanks to the large mass of the top quark, its decay via an on-shell W boson is possible and according to [1], the decay into its partner, the bottom quark, is strongly dominating. The large top-quark mass provides an immense decay phase-space, so its decay is very fast, even before a hadron can be formed via confinement mechanisms, as explained in Sec. 1.3.3. Based on measurements of the top-quark decay width $\Gamma_t = 2.0 \pm 0.5$ GeV from [1], the top-quark mean lifetime is $\tau_t = \hbar/\Gamma_t = (3.3 \pm 0.9) \times 10^{-25}$ s.

Hence, the decay of a top-quark pair produces in almost 100% of the cases a bottom quark, a bottom antiquark, a W^+ , and a W^- boson. While the bottom-flavored quarks produce jets, the W bosons (here as an example the W^+ boson) can decay into two quarks, mostly into the isospin pairs:

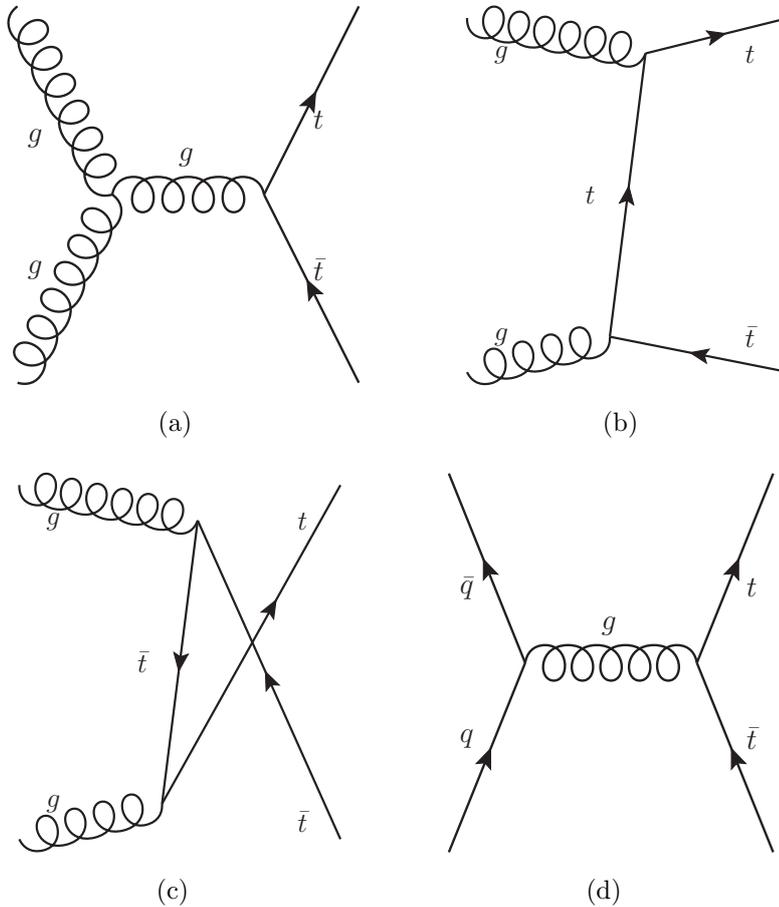


Figure 1.5: Feynman diagrams of top-quark pair production with gluons in the s channel (a), in the t channel (b), and in the u channel (c) and with quarks in the initial state annihilating to a gluon in the s channel (d).

(u, \bar{d}) and $(c, \bar{s})^3$ or into a charged and a neutral lepton: (e^+, ν_e) , (μ^+, ν_μ) , or (τ^+, ν_τ) . Since the masses of the possible decay products, quarks and leptons, are negligible compared to the W -boson mass and their couplings to the W boson are equal, each possible decay has the same weight within the decay phase space, thus the same probability to be produced. Taking into account the three possibilities from the three colors of the quark pairs produced in the W -boson decays, one can represent the decay phase space of a top quark pair schematically as in Fig. 1.6. If both W bosons decay into leptons, the decay channel is called dileptonic. If both W bosons decay into quarks, the decay

³Since the mass of the top quark is larger than the mass of the W boson, the latter cannot decay into the last quark generation (t, \bar{b}) .

channel is called all-hadronic. The last possibility is that one of the W bosons decays into leptons and the other one decays into quarks, this decay channel is called semi-leptonic (electron+jets, muon+jets, and tau+jets). This latter decay channel, including electron+jets and muon+jets channels, is the one analyzed in this thesis.

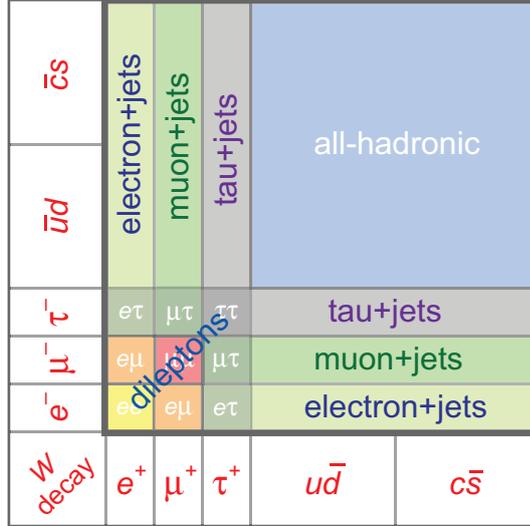


Figure 1.6: All possible decay channels of a top-quark pair. The areas are proportional to the approximative branching ratios of the respective channels. [18]

1.4.3 Main Properties of the Top Quark

Results from TeVatron and LHC for the measurements of the top-quark mass and cross section are given in this section.

Top-Quark Mass The mass of the most massive elementary particle needs to be measured precisely as it is a parameter of the SM. Combinations of measurements from TeVatron and LHC experiments are:

$$m_t^{\text{TeVatron}} = 174.34 \pm 0.64 \text{ GeV [19]}, \quad m_t^{\text{LHC}} = 173.29 \pm 0.95 \text{ GeV [20]}$$

which are consistent with the average calculated by [1]. In addition, a summary of top-quark mass measurements is given Fig. 1.7 with a focus on CMS measurements and comparisons with the TeVatron combination and the world combination. A slight discrepancy is observed between TeVatron and LHC measurements.

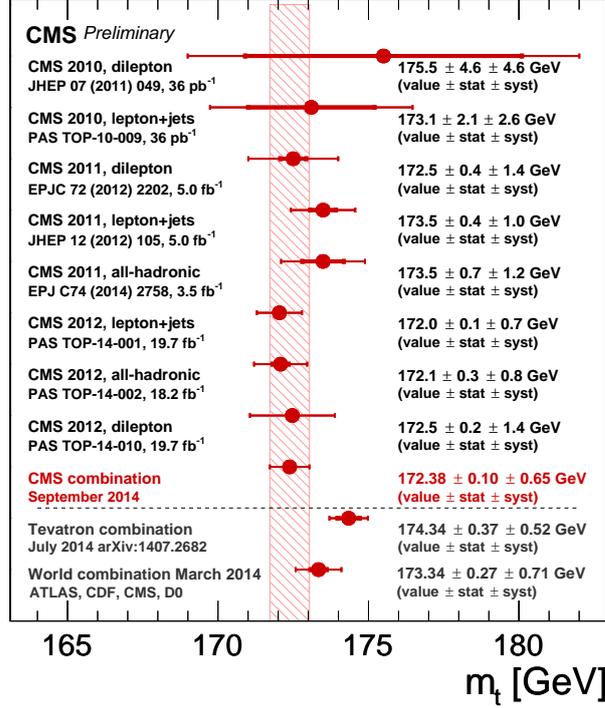


Figure 1.7: Summary of top-quark mass measurements from the CMS collaboration from pp collisions at a center-of-mass energy of 7 and 8 TeV at the LHC and the CMS combination. The red-hashed band corresponds to the uncertainties on this latter measurement. This is compared to the TeVatron mass combination and the world combination including results from ATLAS, CDF, CMS, and DØ. Taken from [21].

Top-Quark Cross Section Measurements of the cross section provide knowledge about the coupling of the top quark to other particles of the SM and potential BSM particles. Therefore, those measurements are essential. Combinations of cross section measurement for top-quark pair events ($t\bar{t}$ events) from TeVatron (CDF and DØ experiments separately) and LHC experiments are:

$$\begin{aligned}\sigma_{t\bar{t}}^{\text{CDF}}(\sqrt{s} = 1.96 \text{ TeV}) &= 7.46^{+0.66}_{-0.80} \text{ pb} [22], \\ \sigma_{t\bar{t}}^{\text{DØ}}(\sqrt{s} = 1.96 \text{ TeV}) &= 8.18^{+0.98}_{-0.87} \text{ pb} [22], \\ \sigma_{t\bar{t}}^{\text{LHC}}(\sqrt{s} = 8 \text{ TeV}) &= 241.4 \pm 8.5 \text{ pb} [20].\end{aligned}$$

The variable \sqrt{s} represents the center-of-mass energy of the collisions. A summary of the recent measurements, including preliminary results, showing

a remarkable agreement with predictions calculated with the latest NNLO+NNLL⁴ tool for inclusive $t\bar{t}$ cross section, top++ [24], is given in Fig. 1.8.

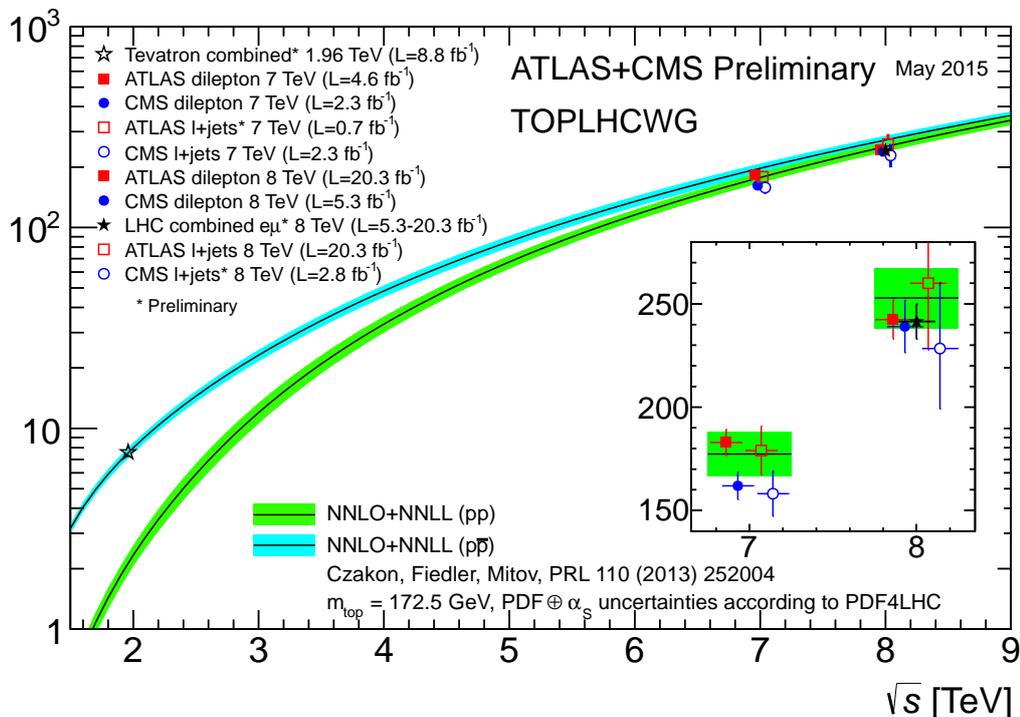


Figure 1.8: Summary of measurements of the top-quark pair cross section for center-of-mass energies of $\sqrt{s} = 7$ and 8 TeV from the LHC and $\sqrt{s} = 1.96$ TeV from the Tevatron. They are compared to NNLO+NNLL predictions calculated by [25] with top++ [24]. Figure provided by the TOPLHCWG [26, 27].

1.4.4 State of the Art for Cross Section of Top-Quark Pair Events with Jets

The present thesis aims at a precision measurement of top-quark and QCD properties related to the production of jets in $t\bar{t}$ events. Calculations of the $t\bar{t}$ cross section with one jet emitted in addition to the $t\bar{t}$ system have been performed at NLO. A first publication [28] was further refined in [29], where the latter includes the simulation of parton showering. Requiring a trans-

⁴The leading logarithm (LL) term stands for the approximation made in the parton showering, as explained in [23]. NNLL means next-to-next-to-leading logarithm, achieving a higher precision.

verse momentum⁵ of the additional jet to be > 50 GeV, in pp collisions at the LHC with $\sqrt{s} = 7$ and 14 TeV the calculated cross section is $53.1^{+4.1}_{-8.9}$ and $376.1^{+20.1}_{-45.4}$ pb respectively. A calculation with two additional jets at NLO has been published as well [30]. Requiring that the additional jets show a transverse momentum > 50 GeV, an absolute value of the rapidity < 2.5 , and an angular distance between both additional jets $\Delta R > 0.5$ ⁶, the cross section for pp collisions at $\sqrt{s} = 7$ TeV is calculated to be 9.82 ± 0.02 pb. It is relevant to note that in [30], the additional jets are radiated before the top quarks decay. Hence, radiation from decay products is not considered. When identifying additional-jet properties in experimental data, this distinction is non-trivial. Another point to raise is the identification of the additional jets in $t\bar{t}$ events. In the aforementioned calculations, additional jets are identified as such making use of the top quarks (i.e. before they decay). The same procedure is obviously impossible with real data, thus a comparison of a measurement with theory has to rely on simulation information about the top quarks, the so-called parton level. Using the latter information is, however, debatable since no unique definition of this simulation level exists. This leads to model-dependent measurements. This is further discussed in Sec. 6 and the measurement presented extensively in the second part, summarized in Sec. 7, attempts to get rid of such model dependencies.

A pioneering measurement of additional jets in $t\bar{t}$ events was performed in 2009 by CDF with a conference note [31]. Their result yields $\sigma_{t\bar{t}+j} = 1.6 \pm 0.5$ pb which is in agreement with the prediction $\sigma_{t\bar{t}+j} = 1.79^{+0.16}_{-0.31}$ pb from [32] at the time of publication. Several public results have been provided by the ATLAS collaboration: based on simulation only [33] and with data from the LHC run at $\sqrt{s} = 7$ TeV [34]. Comparable measurements have been published by the CMS collaboration [35]. In both papers the differential cross-section measurement as a function of the number of jets is reported and compared to predictions from LO and NLO generators, as well as predictions from LO generators with shifted generation parameters (representing the theory uncertainties).

A large contribution to the analysis from [35] (in the semi-leptonic channel) and the elaboration of this paper has been performed within the PhD work of the author of these lines. Therefore, the results of this paper are presented and discussed in Sec. 6, as a reinforcement of the main results which are making use of more recent data from pp collisions at $\sqrt{s} = 8$ TeV.

A top-quark-mass measurement for $t\bar{t}$ events with additional jets has been

⁵Transverse means here projected on the plane transverse to the proton-beam axis.

⁶See Sec. 3 for the definition of the rapidity and the angular distance ΔR between particles/objects.

published by ATLAS [36] making use of the $t\bar{t} + 1$ jet cross-section measurement.

1.4.5 Top-Quark Pair Events with Jets – Beyond the Standard Model

Any deviations observed from SM predictions are signs of new physics. In the frame of this thesis, following models can be considered:

- Pairs of spin-3/2 excited top quark can decay into standard top-quark pairs via the radiation of two additional gluons. A search is presented in [37].
- A minimal supersymmetric model including the violation of R -parity predicts supersymmetric partners of bottom quarks (sbottoms) as the lightest supersymmetric particle. The sbottom decays into a top quark and a strange quark. The sbottom-pair production leads to the production of a top-quark pair accompanied by two strange quarks. It has been searched in the dileptonic channel in [37].
- Baryon-number violating top-quark decays are predicted for instance in supersymmetry models. This can lead to a decay like: $t \rightarrow \bar{b}\bar{c}\mu^+$ where a quark replaces the expected muon neutrino. In case of a $t\bar{t}$ pair where one of the top quark decays by such a baryon-number violating process and the other top quark decays in a standard way, the resulting topology is very similar to a standard $t\bar{t}$ decay (except the neutrino, which is anyway not required in the measurement presented in this thesis) and it shows an additional jet from the \bar{c} of the latter example. This is presented in [38].
- Anomalous coupling of the top quark originating from deviations from point-like behavior would show up in the jet multiplicity in $t\bar{t}$ events. This is mentioned in [31].

Chapter 2

Simulation of Data

The SM presented in the previous chapter provides predictions for any physical process, e.g. the production of $t\bar{t}$ events. In order to make sure that these predictions are correct, the simulation of the processes according to the predictions is very useful. The signal as well as the expected background processes can be simulated and their summed contributions compared to measurement data. Basically if all simulation variables agree with those from data within the data uncertainties, the theory can be validated. This is the common way to perform data analysis in particle physics.

The simulation is the process of artificially producing a sample which can be compared to the data. In this thesis, these MC-simulation samples are dataset of events. An event is the record of a proton-proton collision and the subsequent decays of the particles, like the $t\bar{t}$ pair, generated by the collision. This includes the propagation of the produced particles until their detection or absorption in the detector.

The simulation is based on the Monte-Carlo method (MC) [39] which makes use of the law of large numbers and the generation of random numbers. As shown for instance in Chapter 1, multi-dimensional integrals appear in the SM. The MC integration is well suited to such calculations.

2.1 Parton Distribution Function

Protons consist of three valence quarks (two up quarks and one down quark). Moreover, gluons and quarks can appear from the vacuum, violating four-momentum conservation. This is allowed for a short enough time, according to Heisenberg's uncertainty principle $\Delta t \Delta E \geq \hbar/2$. Hence, the valence

quarks of protons are seen in high-energy experiments, like proton-proton collisions of the LHC, presented in Sec. 3, as swimming in a sea of quarks (therefore called sea quarks) and gluons in the protons. Since the strong interaction holds all these proton constituents together at an energy scale which is too low to allow calculations using the perturbation theory, no prediction with perturbation theory is possible. This concerns the proton constituents, and therefore the partons emerging from the colliding protons, e.g. at the LHC.

The parton model [40] provides probabilistic predictions for the partons, but the parameters of this model rely on empirical measurements of the proton structure. These predictions can be included as explained in [9] thanks to the factorization theorem. It allows to consider separately processes taking place within the proton (at low energy scale) from the high-energy reaction between the partons emerging from the colliding protons. This theorem is made valid by the fundamentally different scales involved since the reactions inside a proton take place at much larger scales than reactions between partons in LHC collisions. As a consequence the cross section $d\sigma_{h_1 h_2}$ of a process taking place between hadrons h_1 and h_2 (protons at the LHC) can be written as the convolution of the hard-process cross section $d\sigma$, as calculated in Sec. 1.3.2, and of the probability f to find the partons i and j emerging from the hadrons h_1 and h_2 [9]:

$$d\sigma_{h_1 h_2} = \sum_{i,j} \int_0^1 dx_i \int_0^1 dx_j f_{i/h_1}(x_i, \mu_F^2) f_{j/h_2}(x_j, \mu_F^2) \frac{d\sigma}{dx_i dx_j}.$$

The probability $f_{i/h_1}(x_i, \mu_F^2)$ for a parton of type i to emerge from the hadron h_1 with the momentum $x_i \cdot p_{h_1}$ is called the parton distribution function (PDF). The Bjorken scaling variable x_i is the fraction of the hadron momentum p_{h_1} carried by the parton i . Finally, μ_F^2 is the factorization scale. It represents the energy scale at which the partons are entering the high-energy reaction.

The determination of the PDF can only succeed via probes of the proton structure at a given energy scale. This is achieved via fits to data typically originating from deep-inelastic scattering experiments: electrons colliding with protons like at SLAC [41] and HERA (from the H1 collaboration [42] and from the ZEUS collaboration [43]); muons colliding with protons like at NA4 [44] at CERN; or with neutrinos colliding with protons like with the neutrino beam of Fermilab [45]. These probes of the proton take place at lower-energy scales (≤ 1 TeV) so the PDFs need to be extrapolated to the LHC energy (with the DGLAP equations [46–48]). Information from jets measured for instance at HERA [49] are also used for the elaboration of the

PDFs. An illustration of the PDF set utilized for the simulation of the LHC collisions for this thesis is given in Fig. 2.1 for a scale of approximately the top-quark mass.

In a proton-proton collision at the LHC, e.g. with $\sqrt{s}=8$ TeV, approximately 1% only of the energy of each proton is required for the $t\bar{t}$ production. As shown in Fig. 2.1, for $x = 1\%$, gluons have the highest probability to be emitted. This explains why the $t\bar{t}$ production at the LHC is dominated by gluons in the initial state.

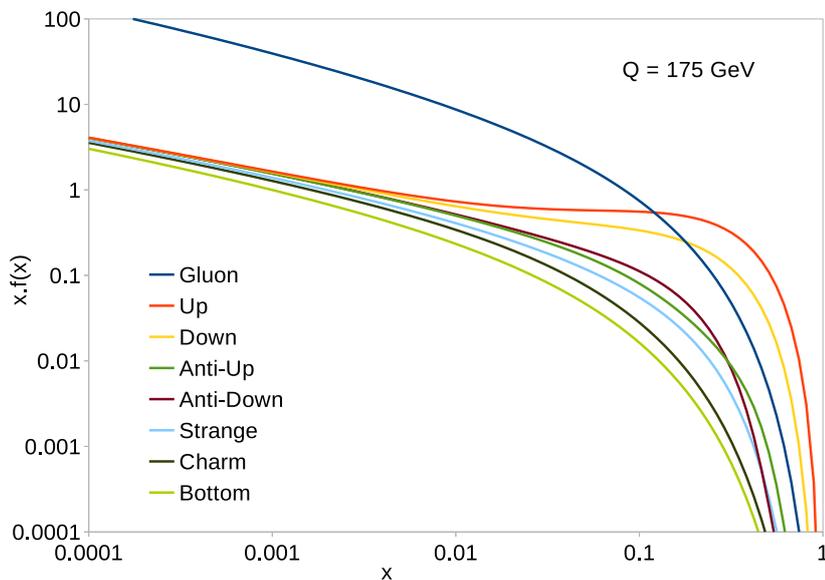


Figure 2.1: Parton distribution function times Bjorken scaling variable ($f(x, Q) \cdot x$) from the CTEQ6L [50] PDF set for gluons and quarks emerging from a proton as a function of x at a high Q scale. Calculated with the online tool from the Durham HepData Project [51].

2.2 Matrix-Element Event Generators

The generation of proton-proton collision events with the MC method is basically separated in two steps because of the change of energy domain. The high-energy processes, for instance the creation of the $t\bar{t}$ pair, where the perturbation theory can be applied, are calculated by a matrix-element generator. Once this is done, the resulting color-charged particles need to undergo a specific treatment at a lower energy scale, presented in Sec. 2.3.

2.2.1 Matrix-Element Computation

Automated tools exist for the generation of events at the hard-process level with different levels of precision, i.e. orders of the perturbation theory. With partons from the colliding protons as input, a large range of processes can be calculated. Among others, the user has to specify the process to be generated, at which center-of-mass energy it should take place, and how many events are required. In a first step, the differential cross sections of all Feynman diagrams of the configured process (and according to the order of the perturbation theory) are calculated. Then the MC method is utilized for the generation of the events. The probability for each event to take place depends on the momentum and the type of both incoming partons as well as on the corresponding Feynman diagram. Afterwards, a point in the allowed final-state phase space is taken randomly which determines the momenta of the output particles.

The more precision is required the more calculation intensive the MC generation gets since for higher orders of the perturbation theory more Feynman diagrams need to be calculated. At LO, the $t\bar{t}$ simulation requires no loop and no additional quark or gluon, as shown in the previous chapter in Fig. 1.6. However, a substantial fraction of $t\bar{t}$ events produced at the LHC is accompanied by additional jets, as studied by the analysis presented in this thesis. Their simulation can succeed with limited precision during the parton-showering, presented in Sec. 2.3.1. Nevertheless, additional jets which are measurable by the CMS detector are very likely to originate from high-energy radiation $p_T^{\text{rad}} \gg \Lambda_{\text{QCD}}$, where the perturbation theory is valid, so their occurrence can be precisely calculated taking into account processes with additional emissions. This is possible with matrix-element generators at LO like MADGRAPH 5, presented in Sec. 2.2.2, by explicitly configuring the generator to produce for instance $t\bar{t}$ events with zero, one, two, etc. additional quarks or gluons. This leads to an approximation of the higher orders of the perturbation theory because a part of the higher-order contributions is taken into account (the additional radiation) while the loops are ignored. Such a generation technique is extensively utilized in this thesis for the simulation of the $t\bar{t}$ MC simulation.

NLO tools exist too, providing the further precision of one additional radiation or one additional loop in the diagrams considered. The automation of these tools represents a challenge mainly because of the complex treatment of loops. Furthermore, in the case of combining exclusive NLO MC-simulation samples with different emission multiplicities, like $t\bar{t} + 0$ and $t\bar{t} + 1$, the merging needed in order to suppress the overlap between the real-emission

matrix element of the lower multiplicity ($t\bar{t} + 0$ emission + 1 NLO emission) and the born-level matrix element of the higher multiplicity ($t\bar{t} + 1$ emission + 0 NLO emission) represents an additional level of complexity.

One should mention that reaching a higher precision via higher order of the perturbation theory provides not only a better description of the process by the inclusion of more Feynman diagrams. It also reduces the dependency of the calculated cross sections on the choice of the renormalization scale Q^2 . As mentioned in Sec. 1.3.3 the strong interaction needs to be renormalized, leading to an assumption, the artificial renormalization scale. Calculating all orders of the perturbation theory for a given process would lead to a result which is completely independent from the choice of Q^2 . Depending on the calculated process, LO calculations can show significant dependencies on Q^2 which is an uncertainty of the simulation. This is strongly reduced by going to NLO calculations.

For cases like the $t\bar{t}$ production whose output particles, the top quarks, need to be decayed before the parton showering, this can be done by the matrix-element generator.

2.2.2 Examples of Matrix-Element Generators

MadGraph 5 is a LO generator utilized in CMS for the hard-process generation of lots of signal and background processes. It enables to generate among others $t\bar{t}$ events with a limited number of additional radiation for an approximation of higher orders. Recently MADGRAPH¹ has been merged with an automated NLO tool, aMC@NLO an automated NLO generator, into the **MadGraph aMC@NLO** package [52]. This enables the production for instance of $t\bar{t}$ events with the NLO precision and by means of elaborated matching/merging techniques [53] to generate also $t\bar{t} + X$ additional emissions. This new tool was not ready in time for the CMS-official MC production in 2011-2012 runs. Therefore, no very large MC-simulation sample from **MadGraph aMC@NLO** could be provided for this analysis. However, in the last months of the elaboration of this thesis, this tool was used by the CMS collaboration for the production of a new $t\bar{t}$ MC simulation. Predictions from this MC simulation are compared to the results of the measurement performed in this thesis in Sec. 11.5.

¹“Mad” stands for Madison of the University of Wisconsin-Madison

MC@NLO is a matrix-element generator at NLO [54, 55]. The gain in precision is achieved by the calculation of the additional Feynman diagrams with a further radiation or a further loop with respect to the LO generation. A specific subtraction technique [56] needs to be applied in order to take into account the cases where parts of the phase space covered by MC@NLO are also filled by the subsequent parton-showering tool. Such cases would lead to a wrong double counting, and hence a weight which can be positive or negative is associated by MC@NLO to each generated event, for the purpose of subtraction of the faulty overlap.

POWHEG is also a NLO matrix-element generator [57], which name means POSitive-Weight Hardest Emission Generator. The main difference to MC@NLO is the treatment of the events in prediction of the subsequent parton showering, as presented for MC@NLO in the previous paragraph. The method used in POWHEG [58–60] takes care of the hardest radiation of color-charged particles, as the name says. This first radiation needs therefore to be vetoed in the parton showering. The method induces negative event weights only for a negligible fraction of the events, compared to MC@NLO, but it requires an adaptation of the parton-showering tool.

2.3 General-Purpose MC Event Generation

The previous matrix-element generation step has produced particles, carrying possibly high momenta. In the output of this generation step, all color-charged particles are recorded as free particles which is not possible due to confinement, so further parton-showering and hadronization steps are needed in order to generate observable particles. Most hadrons produced by the hadronization are unstable, their subsequent decays need to be simulated as well. Since each proton involved in a proton-proton collision loses a parton in the hard interaction, the remnant material is highly unstable and produces low-energy showers in the vicinity of the proton beam. This activity is called the underlying event. It is simulated and added to the event actually simulated.

All these simulation steps are performed for the MC simulation utilized in this thesis with general-purpose MC event generators (GPMC) presented in the following.

In addition, these tools are able to perform the full simulation of a large set of processes, replacing the matrix-element generators. However, this succeeds

by means of approximations, thus, this feature is only utilized in this thesis for the simulation of processes whose accurate description is allowed to be limited.

2.3.1 Parton Showering and Hadronization

Color-charged particles produced by the matrix-element generator, like the quarks produced by the decay of a $t\bar{t}$ pair, need a special treatment. While their momentum is large enough, quarks and gluons can radiate further gluons and gluons can split into quark pairs. This process repeats in a so-called parton showering. This leads to the production of large numbers of particles. Therefore, the simulation of this step has to rely on statistical models based on approximations. Sudakov factors provide for a showering particle the probability not to radiate a gluon or split into a quark pair, as a function of Q and t respectively the energy and the evolution scale of the shower. The evolution scale t can e.g. be the squared transverse momentum p_T^2 or the squared angle θ^2 of the radiation with respect to the radiating parton. The showering needs to be performed in an increasing or decreasing evolution-scale order, i.e. ordered in p_T or angle of the showered particles. This method is only valid for small radiation angles so it is not correct for the full event simulation. It is a good approximation once the hard process, including the radiation of additional quarks and gluons, has been produced by the matrix-element generator, i.e. if no hard radiation needs to be simulated by the parton showering.

The more radiation and gluon splitting have happened, the smaller the interaction scale Q gets. This leads to an increase of α_s . At the point of the showering where Q gets close to QCD scale Λ_{QCD} , α_s becomes large enough to stop the parton showering and to force particles to bind into hadrons in a way that colors annihilate with anti-colors within mesons and triplets of colors combine into baryons. This latter step is called hadronization. The hadronization of color-charged particles takes place at limited values of Q , where no deterministic calculation is possible within the perturbation theory. Different empirical models, mentioned later on, exist for the simulation of this step.

2.3.2 Examples of General-Purpose MC Event Generators

PYTHIA is extensively used for the parton-showering and hadronization steps of the MC simulations in this thesis. It is also utilized for the complete event generation of several processes where no high accuracy is necessary. In PYTHIA 6 [61] the parton showering is p_T -ordered. In PYTHIA 8 [62] the parton showering is improved to a so-called dipole shower [63]. The hadronization is performed with the Lund string model.

HERWIG stands for Hadron Emission Reactions With Interfering Gluons. It is also a well-known general-purpose MC event generator [64], used in this thesis as an alternative to PYTHIA. The parton showering is angular ordered and the hadronization utilizes the cluster model.

SHERPA means the Simulation of High-Energy Reactions of PArticles and it is a more recent tool including matrix-element generators [65]. As PYTHIA 8, the parton showering is based on p_T -ordering and improved to a dipole shower. The hadronization is achieved with the cluster model. A MC-simulation sample of $t\bar{t}$ events has been generated with SHERPA by Ralf Farkas with help of Marco Harrendorf at KIT. The collaborative work results in further comparisons of the measurement with NLO predictions, this can be found in Sec. 11.5.

2.3.3 Matching between Matrix-Element Generation and Parton Showering and Hadronization

For the $t\bar{t}$ MC simulation in this thesis, the matrix-element tools are utilized, making use of their accuracy for the hard process. The result of the matrix-element generation is then given as input to a GPMC tool in order to complete the simulation in the lower-momentum regime. The latter step can lead to the emission of further jets. In the case of a matrix-element generation with various emission multiplicities, like at NLO or at LO with explicit further emissions required, this means a potential double counting of events with the same emission multiplicity. For example, a $t\bar{t}$ event is accompanied by two jets which were generated by MADGRAPH. Then the event is passed to PYTHIA which performs the parton-showering and hadronization steps without further jet emission. In another event, MADGRAPH generates only

one further jet and then PYTHIA provokes the emission of one further jet. The same resulting emission multiplicity is found in these two events, so they cause a double counting. This has to be corrected at NLO: MC@NLO and POWHEG have their own treatment, mentioned previously. At LO, for instance in this thesis, the signal simulation is achieved with MADGRAPH interfaced with PYTHIA and the configuration of MADGRAPH requires several further emissions (zero to three). Hence, a matching technique is required in order to cancel any double counting. This is basically achieved by a cut on the momenta of further emissions from PYTHIA. As a consequence, PYTHIA is only allowed to generate further emissions below a threshold if the resulting emission multiplicity would overlap with a multiplicity already provided by the matrix-element generator. The CKKW [66] and MLM [67] matching are well-known techniques.

2.4 Detector Simulation

Lastly, remaining particles with a relatively large mean lifetime can propagate on longer trajectories and enter the detector and interact with it. Hence, the detection has to be simulated in order to take into account all effects of the detector, including the uncertainties, and reproduce exactly the events measured with the CMS detector. The detector's geometry, the efficiency of sensitive material, the interactions with passive material, and the creation of the electric signal are taken into account by a full simulation (FullSim) of the CMS detector [68]. Since this step requires a long computation time, a faster simulation using approximations is available (FastSim) [69], nonetheless, the detector simulation of all MC-simulation samples used in this thesis was performed with the full simulation.

Chapter 3

Experiment - Particle Collider and Detector

The measurements presented in this thesis rely on data from proton-proton collisions delivered by the Large Hadron Collider at CERN¹ in Geneva. This chapter presents the facilities for the proton acceleration and the detection of proton-collision products which provided the data. Some parts of the text originate from the diploma thesis of the author of this thesis [70].

3.1 Large Hadron Collider at CERN

The Large Hadron Collider (LHC) is the powerful particle collider of CERN situated near Geneva over the French-Swiss border. It consists of a ring constructed approximately 100 meters underground (see Fig. 3.1). All along the 26.7 km of the ring circumference, two parallel beams made of bunches of particles (protons or heavy ions) circulate almost at the speed of

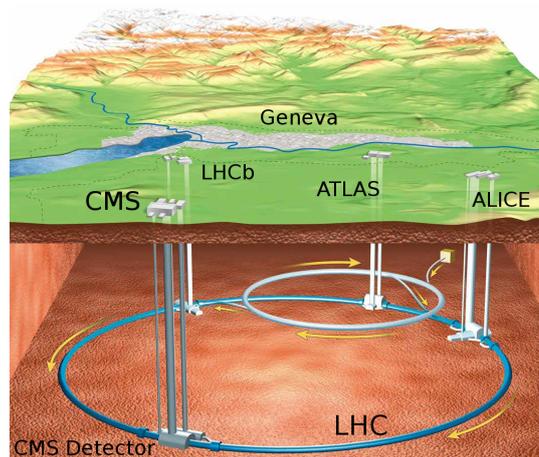


Figure 3.1: Drawing of the LHC, [71].

¹European Organization for Nuclear Research (The CERN acronym stems from the former name: Conseil Européen pour la Recherche Nucléaire)

light in opposite directions. At four points of the ring, the two beams are deflected to a collision point. The very high-energetic collisions taking place here create varieties of particles which can be observed by detectors built around each collision point.

Between the proton source and the final ring of the LHC, a pre-acceleration system prepares the protons for the injection (see Fig. 3.2). First, the protons are extracted from the source, a simple bottle of hydrogen, and split from their electrons. Then in ascending energy order, the LINear ACcelerator (LINAC2), the Proton Synchrotron Booster (PSB or Booster), the Proton Synchrotron (PS), and finally the Super Proton Synchrotron (SPS) accelerate the protons up to the energy of 450 GeV. At this point, the beam is ready for the injection in the LHC where it is accelerated up to the final energy. Several injection cycles are needed for a complete fill of the LHC.

While the LHC is able to provide heavy-ion collisions, this is not the topic of this thesis, thus only the proton-proton collision system is described.

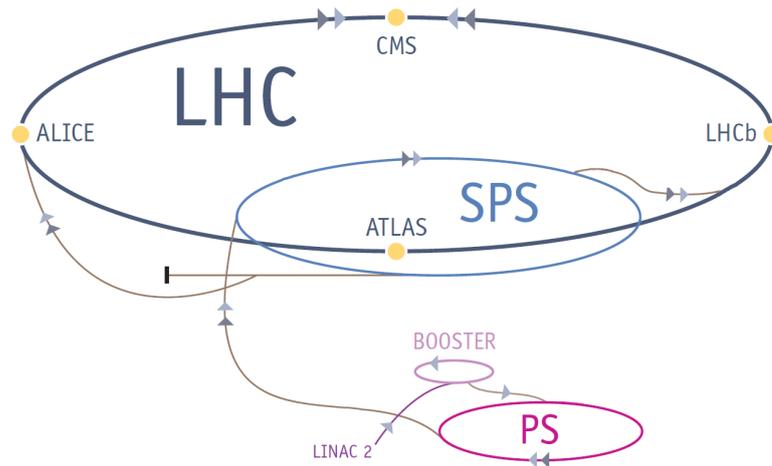


Figure 3.2: Sketch of the LHC pre-acceleration chain: the LINAC2, then the Booster, the PS, and finally the SPS, from [72].

The LHC [73] is designed for a final beam energy of 7 TeV, corresponding to a center-of-mass energy $\sqrt{s} = 14$ TeV. A first long run with a beam energy of 3.5 TeV ($\sqrt{s} = 7$ TeV) started in March 2011 and ended in October 2011. A second run started in April 2012 with an increased beam energy of 4 TeV ($\sqrt{s} = 8$ TeV) and provided the large dataset for the analysis presented in this thesis.

The longitudinal acceleration of the protons up to the collision energy is applied with a 400 MHz superconducting cavity system situated in octant 4 of the LHC (see Fig. 3.3).

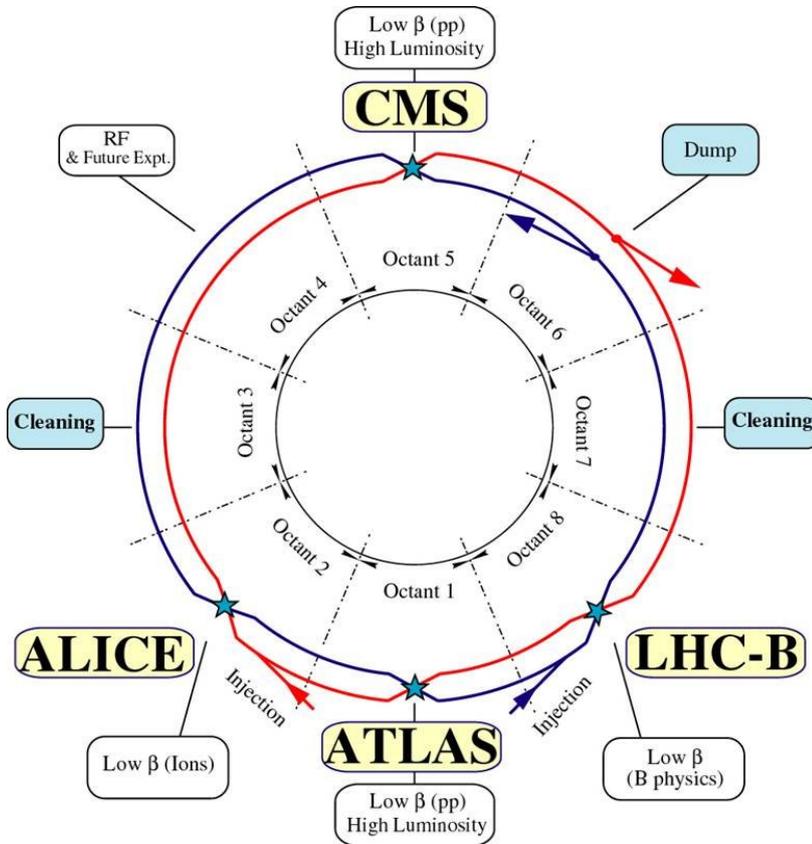


Figure 3.3: Sketch of the LHC ring. The blue and the red ring represent the two parallel proton beams collimated at collision points depicted by the blue stars for the four main experiments. “RF” stands for the radio-frequency cavities dedicated to the acceleration of the protons in the LHC ring and “low β ” means this is a point of high focus of the beams, from [74].

The beam is bent with 1,232 dipole magnets distributed along the LHC ring. The magnetic field strength required in order to keep the proton bunches in the orbit is 8.33 Tesla for an energy of 7 TeV per beam. This very high magnetic-field intensity requires superconducting dipole magnets. The superconductivity is achieved with a cooling system based on superfluid Helium, hence at the temperature of 1.9 K. A total of 8,400 additional multipole magnets focus the beams and correct their orbits.

A fundamental particle-collider parameter, the instantaneous luminosity \mathcal{L} , describes the flux density of particles at the collision point. With n_i being the interaction rate of a physical process i taking place per time unit and σ_i the cross section of this process, the basic particle-collider formula is:

$$n_i = \mathcal{L} \cdot \sigma_i \text{ with } \mathcal{L} = f \cdot N_B \frac{N_1 N_2}{A}$$

where f is the revolution frequency, N_B the number of colliding bunches, N_1 and N_2 are the numbers of particle in each bunch and A is the effective cross section of the beam. An integration of n_i over a given time, for instance over the whole run time of the LHC at 8 TeV provides the number of events produced by all proton-proton collisions in this run. Similarly, the integration of \mathcal{L} provides the integrated luminosity which is a measure of the number of collisions. For instance the measurement presented in this thesis makes use of data delivered by the LHC (and recorded by the CMS detector, presented in Sec. 3.2) in 2012 which corresponds to an integrated luminosity of approximately 20 fb^{-1} .

The design luminosity of the LHC of $\mathcal{L} = 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ has been planned to be achieved by means of 2808 bunches of 1.5×10^{11} protons for each beam at a beam energy of 7 TeV. The LHC research run started in 2010 with a beam energy of 3.5 TeV. With the same energy the 2011 LHC run has reached a maximum instantaneous luminosity of $3.6 \cdot 10^{33} \text{ cm}^{-2}\text{s}^{-1}$ [75]. This quantity was increased up to an instantaneous luminosity of about $7.7 \cdot 10^{33} \text{ cm}^{-2}\text{s}^{-1}$ in the 2012 run period with an energy per beam of 4 TeV [76]. The gain of instantaneous luminosity was made possible by improving all beam parameters, as summarized in Table 3.1. Fig. 3.4 shows the integrated luminosity for the three first run periods of LHC, corresponding to the data volume delivered by the LHC at the collision point of the CMS detector.

Table 3.1: Summary of the LHC main parameter values reached in 2010, 2011, and 2012. The beta variable (β^*) represents how much the beam is squeezed at the collision point. Taken from [75] and [76].

LHC beam parameters	2010	2011	2012
Bunch spacing [ns]	150	50	50
Number of bunches per beam	368	1380	1374
Average number of protons per bunch [10^{11}]	1.2	1.5	1.6-1.7
β^* at the collision point of CMS [m]	3.5	1	0.6

Four main experiments along the ring collect the results of these high energy collisions in order to observe the properties of nature under such extreme conditions (see Fig. 3.3). Two of them, ATLAS (A Toroidal LHC Apparatus) [78] and CMS (Compact Muon Solenoid) [79], are general-purpose detectors while the two others, ALICE (A Large Ion Collider Experiment) [80] and the LHCb (LHC beauty) [81], are specialized detectors. ALICE studies

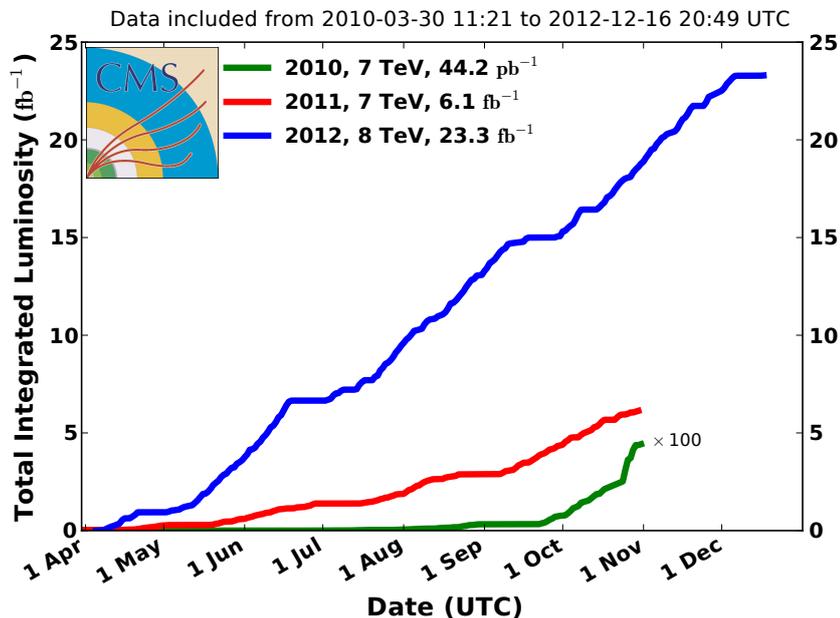


Figure 3.4: Integrated luminosity delivered to CMS during the three first data-taking periods (2010-2012), from [77].

lead-ion collisions which could produce quark-gluon plasmas. LHCb studies the physics of bottom quarks, for instance in the search for CP-violating processes, in order to explain the asymmetry between matter and antimatter in the universe.

3.2 The Compact Muon Solenoid Experiment

The CMS detector (Compact Muon Solenoid) is a multi-purpose detector for particle physics. The sub-detectors of CMS are arranged onion-like in layers around the collision point and immersed in a solenoidal magnetic field (3.8 Tesla for the inner part and 2 Tesla for the muon system) produced by a superconducting magnet. The purpose of the magnetic field is to bend the trajectories of charged particles. This bending allows to measure the momentum of these particles and the sign of their electric charge.

The detection of top-quark events delivered by the LHC requires the information from all sub-detectors of CMS, therefore they are all briefly described in this section. An overview of the CMS experiment is provided by Fig. 3.5.

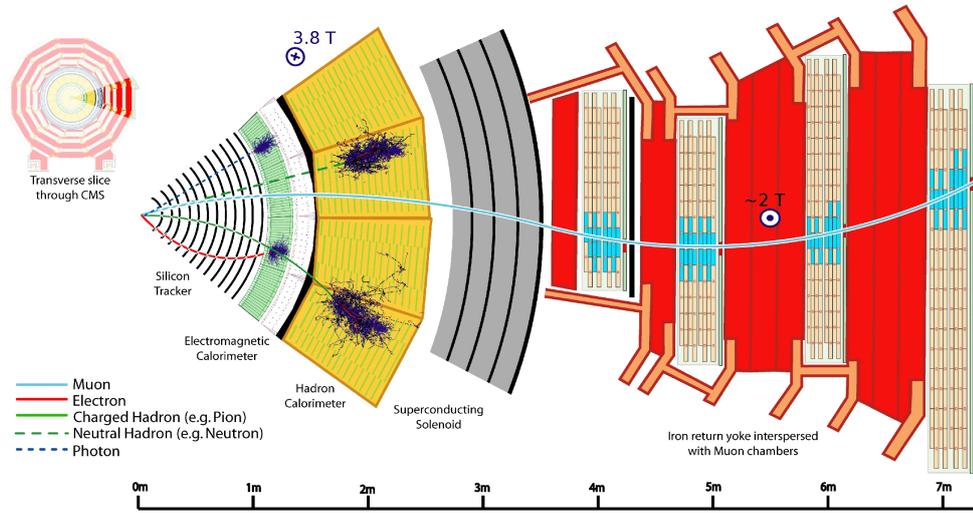


Figure 3.5: A slice of the CMS detector with the traces of particles flying through it, taken from [82] and modified.

The CMS detector is made of the following parts: (inner to outer)

- Silicon pixel detector
 - Silicon strip detector
 - Electromagnetic calorimeter
 - Preshower detector
 - Hadron calorimeter
 - Superconducting solenoid
 - Muon system
- } Tracking system
- } Calorimetry

For the description of the geometry of the different parts of the detector, the coordinate system has to be defined. CMS uses a cylindrical coordinate system with the z direction along the beam circulating in the anticlockwise direction, r the distance to the interaction point in the center of the detector, and ϕ the angle in the x - y -plane (orthogonal to the beam axis, x axis points radially towards the center of the LHC and y axis points vertically upwards). In particle physics the rapidity y of a particle with respect to the beam axis is used additionally. This variable is defined with the longitudinal momentum p_z and the energy E of a particle as $y = 1/2 \cdot \ln\left(\frac{E+p_z}{E-p_z}\right)$. In case of a massless particle, this is equal to the pseudo-rapidity η which is directly related to the angle θ between the beam axis and the flight direction of a

particle: $\eta = -\ln[\tan(\theta/2)]$.

These two last functions are utilized in order to reduce the dependency of the density of measured particle on θ . The reason for this is that most of the particles resulting from an interaction receive low transverse momenta (p_T , projection of the momentum on the transverse plane) thus their direction is likely to be close to the beam while only a much smaller number of particles fly orthogonally to the beam. When expressed as a function of the rapidity and pseudo-rapidity, the distribution of the particle density gets flatter. In the CMS detector, both regions of high $|\eta|$ are mainly covered by detectors oriented orthogonally to the z -axis. Both sides are referred as the end caps. They close the barrel part which is the region of low $|\eta|$ where detectors are oriented coaxially to the beam line.

Finally, the angular distance between two objects is noted $\Delta R = \sqrt{(\phi_1 - \phi_2)^2 + (\eta_1 - \eta_2)^2}$, where either the rapidity or the pseudo-rapidity can be utilized.

3.2.1 Tracking system

Nearly all top quarks decay via the weak interaction into a bottom quark and a W boson. Bottom quarks hadronize and produce therefore jets. These jets are a very important part of the signature of a top-quark pair event because bottom quarks are almost always produced in top-quark decays, while they are much rarer in the relevant background processes². Therefore, an efficient way to select top-quark pair events in data is to require that two jets originate from a bottom quark. It is not possible to find out which kind of quark (or a gluon) has produced a jet by just measuring its energy depositions. All jets look mostly the same, thrusts of particles boosted in the same direction, and their inner structure cannot be decomposed easily. Fortunately bottom quarks are an exception to this. They hadronize into B hadrons whose mean lifetime ($\approx 10^{-12}$ s [1]) enables it to fly several millimeters before decaying. This decay has to happen via a flavor changing, i.e. via the weak interaction, because of the bottom flavor. This flavor changing cannot happen within the flavor generation because the weak-isospin partner of the bottom quark, the top quark, is much more massive. Hence, the flavor can only be changed into either charm or up, across flavor generations which is strongly suppressed in the SM. This results in a relatively long B -hadron mean lifetime.

A consequence is that particles produced by the B -hadron decay do not

²Because of their non-negligible mass, the bottom quarks are less abundantly produced in typical processes taking place in the collisions of the LHC.

originate exactly from the proton-proton interaction point (primary vertex). Thanks to CMS's very accurate tracking system, the displacement of these decay products with respect to the primary vertex can be measured and their point of origin, where the B hadron has decayed (secondary vertex) can be reconstructed. This information (track displacement and secondary vertices) is utilized for the identification of jets originating from bottom quarks (b-tagging) which is essential for the event selection applied in the analysis presented in this thesis.

In CMS, the tracks are mainly reconstructed with the signal coming from the pixel and strip sub-detectors (tracking system) in the vicinity of the interaction point. This can be combined with measurements from the muon chambers surrounding the whole detector for muon tracks.

The description of the tracking system can be found in this section and the description of the muon chambers is described in Sec. 3.2.4.

The track displacement with respect to the interaction point can be measured accurately with a detector installed as close to the interaction point as possible. Nevertheless, the closer to the interaction point the more the particle density is increased. This environment demands the tracking detector to have a very fine segmentation in order to measure the track of each charged particle separately and with sufficient precision.

Moreover, the tracking detector has to cope with a high proton-bunch crossing rate combined with ≈ 20 inelastic interactions per bunch crossing, corresponding to $\approx 10^9$ collision events per second in the detector. This large number implies for the tracking detector to stand such conditions without serious damage and to have a fast response. For these purposes, semi-conductor detectors have been chosen. A large number of small semi-conducting modules (pixels in the inner part and strips in the outer part) are arranged in order to achieve a high spatial resolution for the reconstruction of the tracks while keeping a maximal coverage of the solid angle around the collision point.

Principle, Material, and Geometry

The silicon detector consists of a large number of modules. Fig. 3.6 shows a cut of the whole tracking system along the beam axis, where each line represents one module. All of these modules are based on pn-junctions of doped silicon.

Semi-conducting sensors like the modules used in CMS are based on the

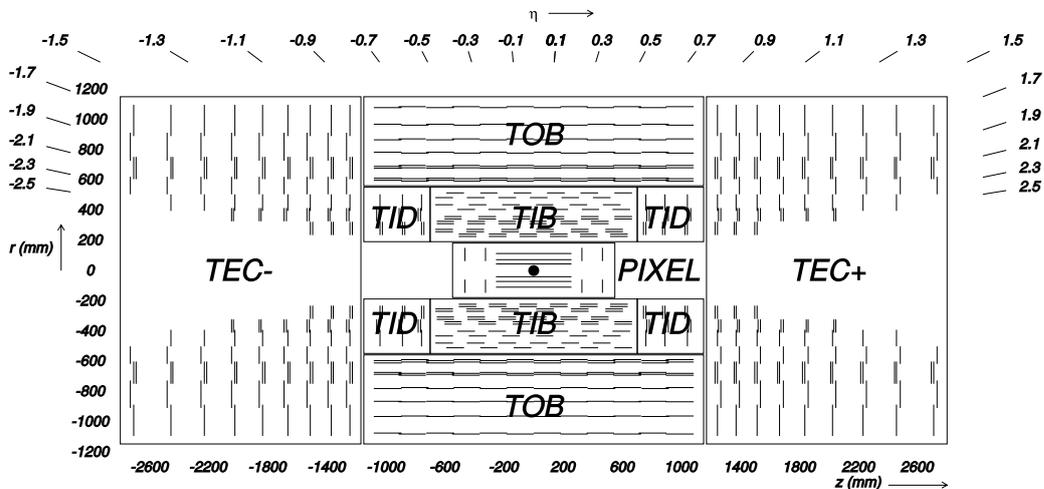


Figure 3.6: Schematic view of the silicon detector of CMS in $r - z$ view. The corresponding angular (η) range is represented on the top of the figure with the radial ticks. One can see the collision point depicted with the black dot in the middle. The modules of the pixel detector (PIXEL), the Tracker Inner Barrel (TIB), and the Tracker Outer Barrel (TOB) are approximately parallel to the beam axis while the modules of the Tracker Inner Discs (TID) and the Tracker End-Caps (TEC) are approximately perpendicular to the beam axis. Taken from [79].

creation of electron-hole pairs in the material, as a consequence of the crossing of a charged particle. These free charge carriers are accelerated by the electric field of the pn-junction and the resulting current is amplified in order to produce the signal. The combination of all module measurements along the trajectory of the charged particle enables to reconstruct its track.

Starting very close to the collision point, the first subpart of the tracking system is the silicon pixel detector (PIXEL). Its sensors are segmented into 66 million of pixels (size $100 \times 150 \mu\text{m}$) arranged concentrically around the collision point. Then come the Tracker Inner Barrel (TIB) and the Tracker Outer Barrel (TOB). These are silicon strip sensors arranged concentrically around the pixel detector. Finally, the Tracker Inner Discs (TID) and the Tracker End-Caps (TEC) are juxtapositions of disks dedicated to particles flying at large $|\eta|$. They are also equipped with silicon strips. See Table 3.2 for a summary of the geometry and find more details on the CMS silicon detector in [79]. The η coverage of the tracking system is $|\eta| < 2.5$, as can be seen in Fig. 3.6.

Table 3.2: Summary of the geometrical coverage of CMS tracking system [79].

Subpart	Radial coverage [cm]	z -axis coverage [cm]
Pixel	$4.4 < r < 15$	$ z < 46.5$
TIB	$25.5 < r < 49.8$	$ z < 70$
TID	$20 < r < 50$	$80 < z < 90$
TOB	$55 < r < 116$	$ z < 118$
TEC	$22.5 < r < 113.5$	$124 < z < 282$

Performance

The single point resolution of the pixel detector is very high ($\approx 10 \mu\text{m}$), which is a major strength of the CMS detector. However, several points are required for an accurate track reconstruction. This is presented in Sec. 4.2. The track p_T resolution evaluated with single muon tracks achieved with the whole silicon detector is depicted in Fig. 3.7. For typical muons with an approximate p_T of 100 GeV, the relative p_T resolution is between 1.5% and 2% for $|\eta| < 1.5$ and degrades gradually to about 7% for $|\eta| \approx 2.4$. Lower- p_T muons (e.g. $p_T = 1$ or 10 GeV) have a p_T resolution better than 2% for $|\eta| < 2.4$.

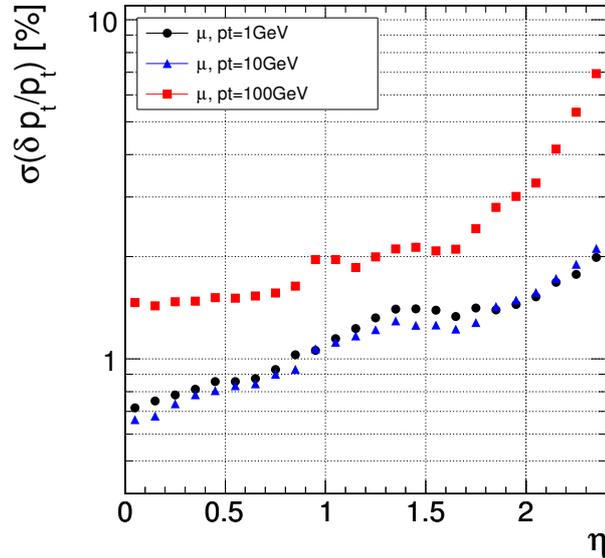


Figure 3.7: Distribution of the p_T resolution of muon tracks as a function of muon η , for various muon- p_T values, achieved when using the whole silicon detector [79].

3.2.2 Calorimeters

The CMS calorimeter system consists of an electromagnetic and a hadron calorimeter, dedicated to the energy measurement of different particle types via their complete absorption in the detector.

Electromagnetic Calorimeter

The electromagnetic calorimeter (ECAL) is dedicated to the energy measurement of photons, electrons, and positrons. It is a homogeneous calorimeter made of 75,848 lead tungstate (PbWO_4) crystals. In the barrel part ($|\eta| < 1.479$) 61,200 crystals have a front-face cross section of $22 \times 22 \text{ mm}^2$ and a length of 230 mm. At both end-cap parts ($1.479 < |\eta| < 3.0$) 7,324 crystals have a front-face cross section of $28.6 \times 28.6 \text{ mm}^2$ and a length of 220 mm. An illustration of the electromagnetic calorimeter is shown in Fig. 3.8.

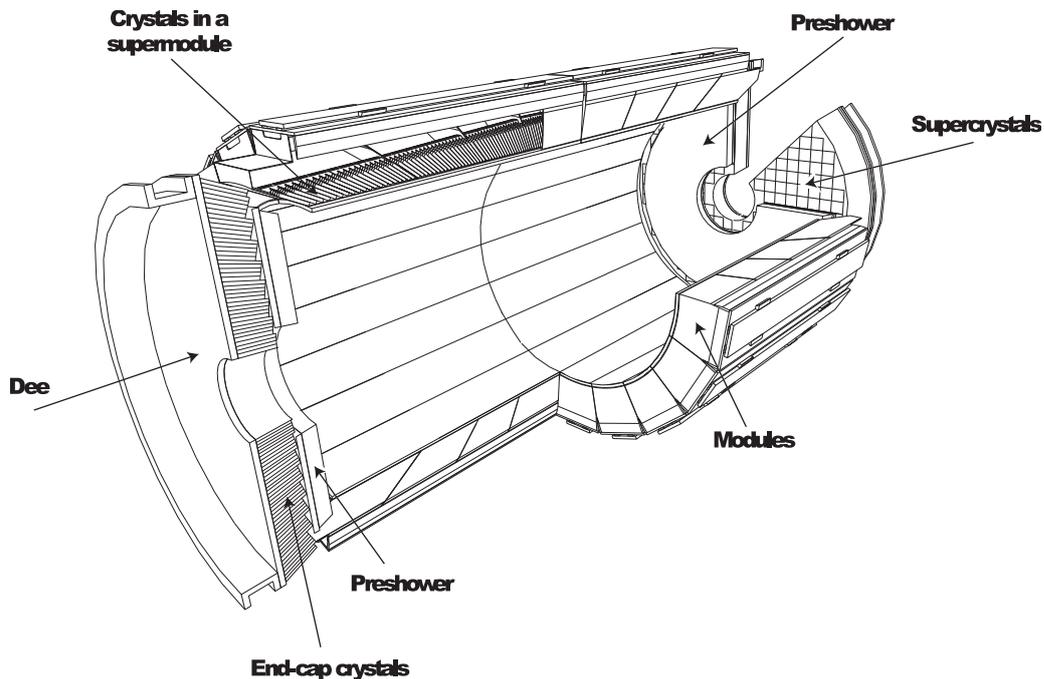


Figure 3.8: A sketch of the electromagnetic calorimeter of CMS [79].

In the electromagnetic calorimeter, photons convert to electron-positron pairs whereas electrons and positrons emit photons via bremsstrahlung. In this

repeated process (bremsstrahlung produces photons, photons produce electrons and positrons) the number of particles grows exponentially resulting in a shower of electrons, positrons and photons (see Fig. 3.5). When the energy of these particles falls below a certain energy limit, the bremsstrahlung emission and photon conversion are no longer possible. Then the particles only lose their energy through ionization of the detector material.

During this process, the atoms of the crystal get excited. These excited atoms relax by radiating visible light which is measured with avalanche photodiodes in the barrel part and with vacuum phototriodes in the end-cap parts. The intensity of the light signal is related to the number of shower particles in the calorimeter, which is related to the energy of the particle which has initiated the shower (e^\pm/γ).

One important parameter of the material used for the calorimeter is the radiation length X_0 . After an electron has radiated energy (typically via bremsstrahlung) until it has only $1/e$ of its initial energy remaining, it has covered on average the distance X_0 in the calorimeter material. Given the radiation length over density of PbWO_4 crystals of $\frac{X_0}{\rho} = 0.89$ cm and the absolute length of the crystals, the electromagnetic calorimeter has in the barrel region a length of $25.8 \frac{X_0}{\rho}$, which is enough to absorb any electron or photon expected in CMS.

As can be seen in Fig. 3.8, a part of this calorimeter is called the preshower. This sub-detector covers the region of $1.653 < |\eta| < 2.6$ in the end-cap in order to assist the ECAL in this high-activity region. The preshower detector is dedicated to the identification of neutral pions and it also helps with the identification of electrons and the position determination of initial electrons or photons. This detector is a sampling calorimeter made of a layer of lead radiators and a layer of silicon sensors.

An energy-dependent energy resolution of $\frac{\Delta E}{E} = 2.8\%/\sqrt{E} \oplus 0.12\%/E \oplus 0.3\%$ is obtained for the whole ECAL including the preshower detector, where the first term is related to stochastic effects, the second one to noise effects, and the last one represents a constant term. More details can be found in [79].

Hadron Calorimeter

The detector built further outward, the hadron calorimeter (HCAL), measures the energy of charged or neutral hadrons like mesons (pions or kaons) and baryons (neutrons or protons). These are typical constituents of jets.

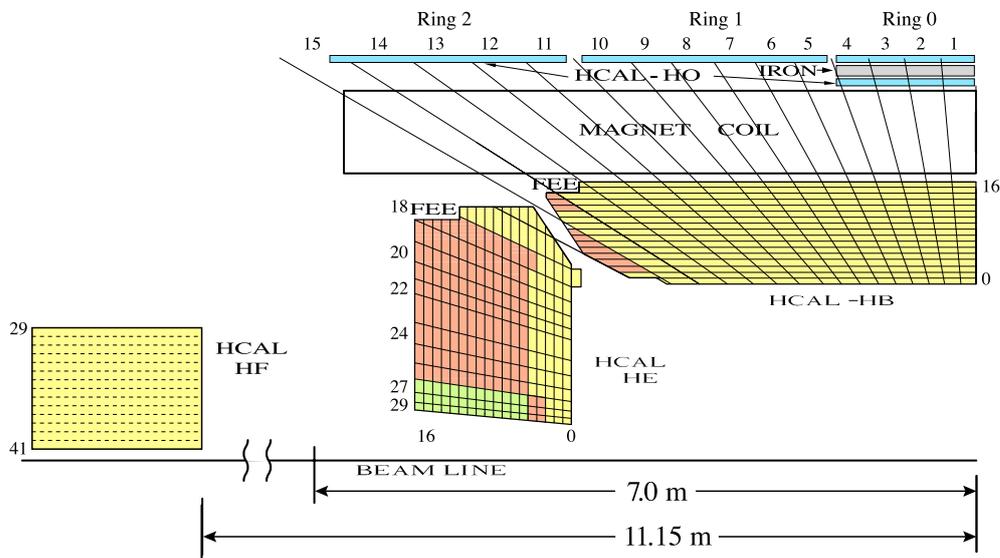


Figure 3.9: Transverse section through a quarter of the hadron calorimeter of CMS along the beam axis, showing the geometrical configuration [83]. The towers depicted with the same color have a common signal read-out to provide a longitudinal segmentation of the measurement.

The CMS hadron calorimeter is a sampling calorimeter made of alternating slices of passive material (non-magnetic brass) and slices of active material (plastic scintillator), see Fig. 3.9 for an illustration.

The incoming hadrons interact with the passive material through strong interaction and produce lighter hadrons. The produced particles excite the active material while traversing it and penetrate the next layer of passive material, where the process repeats etc. This also produces a shower (shown in Fig. 3.5). The excited atoms of the active material emit visible light which is transported by light fibers and measured by hybrid photodiodes (photomultiplier tubes with silicon targets).

The hadron calorimeter consists of several sub-detectors. The hadron barrel (HB) calorimeter covers the central region until $|\eta| < 1.3$ and the hadron endcap (HE) calorimeter is in charge of the $|\eta|$ region between 1.3 and 3.0. In the central region the full absorption of high-energy hadrons which were not absorbed in the HB is achieved by a sub-detector located beyond the magnet coil (HO, for hadron outer calorimeter). The material of the magnet provides additional absorption. Finally, the hadron forward (HF) calorimeter absorbs hadrons with large η , $3.0 < |\eta| < 5.191$, and has a configuration dedicated to the extremely high flux expected in this region. The HF calorimeter has been

designed to cope with an average of 760 GeV energy deposition per bunch crossing, whereas 100 GeV is expected in the rest of the detector.

The passive material chosen for the hadron calorimeter shows an interaction length of $\lambda_I = 16.42$ cm. The thickness of the hadron calorimeter seen from the collision point increases with $|\eta|$. The overall thickness of the hadron calorimeter is $11.8\lambda_I$, except in the barrel-endcap transition region. Finally, the hadron calorimeter segmentation in angular space is $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$, which is coarser than in the ECAL ($\Delta\eta \times \Delta\phi = 0.0174 \times 0.0174$). More details can be found in [79, 83].

The design energy resolution is worse compared to the energy resolution of the ECAL, but this is expected: the HB, the HE, and the HO calorimeters have $100\%/\sqrt{E} \oplus 5\%$ [84], while the HF calorimeter has $198\%/\sqrt{E} \oplus 9\%$ [85].

The Centauro And Strange Object Research (CASTOR) and the Zero Degree Calorimeter (ZDC) are two additional detectors dedicated to measurements in very-forward regions, covering $-6.86 < \eta < -5.32$ and $\eta < -8.5$ respectively. While CASTOR is embedded in the CMS detector behind the HF, ZDC is built around the beam pipe approximately 140 m away from CMS, more details in [79].

3.2.3 Superconducting Solenoid

A magnetic field in the detector is necessary in order to determine the sign of the charge and the p_T of charged particles and a high-intensity field is required for high-momentum particles. The tracking and calorimetry systems are embedded in the solenoid. It is a set of wire coils of 5.9 m diameter and 12.9 m length which produces a uniform 3.8 Tesla field oriented along the beam axis in the inner part of the magnet solenoid and 2 Tesla in the outer vicinity of the solenoid. For the outer part of the magnetic field, a 10,000-ton steel yoke surrounding the magnet coils contains and guides the outer field through the muon system.

The wires are made of superconducting material and co-extruded with pure aluminum. The superconduction requires a cooling system and the aluminum acts as a thermal stabilizer. More details can be found in [86].

3.2.4 Muon System

Many important high energy physics reactions produce muons. As they are minimum ionizing particles, their interaction rate in the calorimeters is very low. Except muons, almost all measurable particles are stopped by the material of the detector parts which were described until here. Therefore, a dedicated detector can be constructed around the hadron calorimeter for the identification of muons and an accurate measurement of their momenta.

The CMS muon system is made of Drift Tubes chambers (DT) in the barrel region ($|\eta| < 1.2$) and Cathode Strip Chambers (CSC) in the end-cap regions ($0.9 < |\eta| < 2.4$). Both are dedicated to the measurement of the muon momentum. See Fig. 3.10(a) for illustration.

The DT are tubes filled with a gas mixture and a thin wire is located in the middle of the tube. A muon traversing the detector ionizes the gas. The resulting electrons are shifted by a high voltage applied between the tube and the wire. The accelerated electrons ionize further the gas leading to a multiplication of electrons collected at the anode. The same mechanism but in a different geometry applies for the CSC with a chamber of strips orthogonal to wires. The idea of the CSC is to have a faster response and a higher position resolution, because this detector is in a zone with a larger particle rate and density. Moreover, Resistive Plate Chambers (RPC) placed in the barrel and end-cap regions provide a very fast response, which is required for the first-level trigger (< 25 ns). The design η coverage of this latter sub-detector is $|\eta| < 2.1$.

The evaluation of the p_T resolution of average muons ($10 < p_T < 1000$ GeV) measured with the whole muon system shows a decrease of resolution from 7% to 25% respectively for $|\eta| = 0$ and 2.25. This is illustrated in Fig. 3.10(b) as a function of the muon η . More details can be found in [79, 86, 87].

3.2.5 Trigger and Data-Acquisition System

With the design bunch-crossing rate of 40 MHz, it is impossible to record the measurements of the CMS detector for every collision event. A solution is to only record events showing interesting objects, which are identified by fast-responding sub-detectors. Only events showing basic physics objects (photons, electrons, muons and jets) are recorded. This is the task of the trigger.

The full signal of each event is stored in a pipelined buffer, which provides the time required by the first trigger step, called Level-1 trigger, to decide whether the event is worth being further considered. If this hardware-based

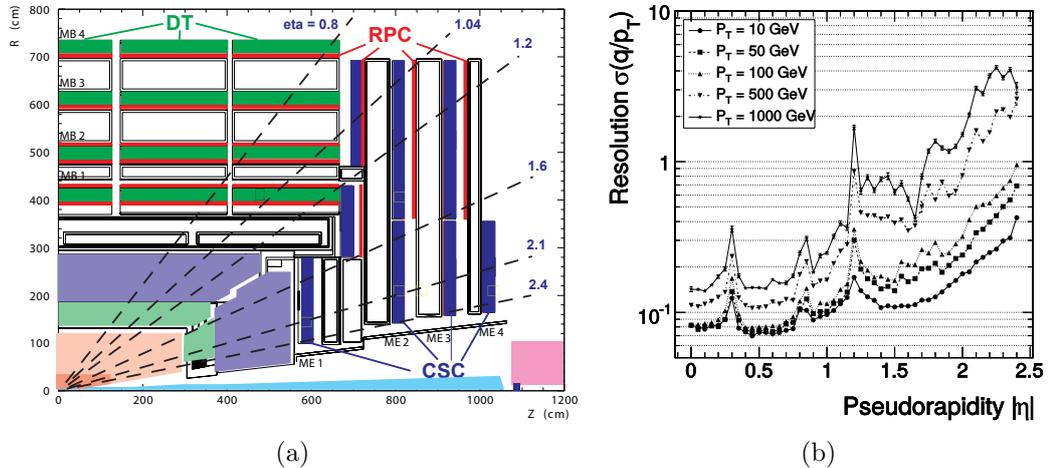


Figure 3.10: (a) Layout of one quarter of the CMS muon system along the beam axis, taken from [86]. The design coverage of the RPC was originally $|\eta| < 2.1$; however, due to budget limitations, it was reduced to 1.6 in the initial detector, as illustrated here. Starting in the bottom-left corner, the tracking system (in dark and light orange), the electromagnetic calorimeter (in light green), and the hadron calorimeter (in violet) are also shown. (b) Distribution of the q/p_T resolution (thus taking into account the right identification of the muon charge q) for simulated muon tracks with various values of p_T as a function of η using only hits from the muon system (with a vertex constraint), taken from [87].

trigger could identify hints for interesting physics objects from the calorimeters or the muon system, it is transferred to the next triggering step. After the Level-1 trigger, the event rate is reduced to less than 100 kHz. The second and last trigger level is called the High-Level Trigger (HLT) and is software-based. Using speed-optimized version of offline-reconstruction algorithms, it reconstructs physics objects and applies a more precise filtering on them. This leads to a reduced event rate of a few 100 Hz which can be stored on disk. More information can be found in [79, 88].

3.3 Pile-up

The physics program of the LHC is based on gathering an extremely large quantity of proton-proton collision-event records. This is required for precision measurements of known processes, as e.g. the measurement presented in this thesis, and for processes investigated at the LHC which generally have very low predicted cross sections. The production of a lot of collision events

in a finite time requires a high instantaneous luminosity, as shown in Sec. 3.1, meaning a high collision frequency.

In order to achieve a high collision frequency the bunches are built with a very large number of protons ($\approx 10^{11}$). As a consequence multiple collision per bunch crossing can happen. If one of the collision leads to the production of a triggering object like a muon, all signals are stored, i.e. not only the collision of interest is present in the record but also the overlapping measurements from simultaneous collisions. These additional measurements are called in-time pile-up (PU).

Out-of-time PU occurs because a small spacing between neighboring bunches is left in the beam (50 ns in 2011 and 2012), which is also done in order to have a high collision frequency. However, since some of the sub-detectors have long signal-integration times, larger than the bunch spacing, measurements from several bunch crossing can be recorded within a single event record.

PU represents a main challenge for the physics program of the LHC because it adds foreign information into the events of interest. This worsens strongly the quality of the detection: a larger multiplicity of charged particles in the tracking system makes the track reconstruction more difficult and less accurate; more vertices can be reconstructed, which makes the position of the true collision vertex more uncertain. PU-mitigation techniques have been developed within the CMS collaboration: PU-subtraction techniques e.g. in the jet energy, track selection (e.g. for b -tagging). Moreover, the simulation of PU turns out to be describing data with sufficient quality.

Chapter 4

Reconstruction of Physics Object

Going from the electronic signal, delivered by the detector-readout system, to physics objects which can be used for data analysis purpose, requires a sophisticated reconstruction step. This is described in this chapter. Some parts of the text originate from the diploma thesis of the author of this thesis [70].

4.1 Global Event Description: Particle Flow

Although sub-parts of the CMS detector are dedicated to the detection of given particles, the measurement of all particles produced by the proton-proton collisions can be improved by making use of the additional information provided by all detectors. This is done in CMS with the Particle-Flow technique (PF) which combines all relevant inputs to achieve the best possible detection. The aim is to reconstruct all stable particles entering the detector (electrons, muons, photons, and neutral and charged hadrons) by means of combination algorithms. As a result, each energy deposition of the detector is assigned to a single physical particle which is likely to have provoked this deposition. See more details in [89–92]. The integration of the complete detector information into an optimal reconstruction via PF techniques is assumed along this chapter whereas representative improvements due to PF are emphasized.

4.2 Tracks

Tracks play a key role in most physics studies. Their reconstruction is possible for all charged particles in CMS thanks to the accurate tracking system [93] and with the support of the muon system [87] for muon tracks. The reconstruction of tracks, including muon tracks, is presented in the following.

Hit Reconstruction The electric signal from pixel and strip modules in the tracking system, and from the muon-system subparts (DT, CSC, and RPC) is clustered into hits, where only signals fulfilling a given threshold are considered, in order to eliminate background noise. Subsequently, the position of the hit as well as its uncertainty is determined.

Seed Generation Starting points for the tracking are created, making use of compatible hit pairs or triplets. These track seeds provide first trajectories. Each trajectory is extrapolated up to the outer detector layers in order to find where hits are expected in the next layers.

Pattern Recognition This algorithm uses the Kalman filter method [94] in order to associate the hits corresponding to each seed extrapolation. The hits of the successive layers are included iteratively, building in parallel all possible tracks.

Track Fitting In order to achieve the best precision, each track is refitted and smoothed. This can be done several times with different parameters (e.g. track-quality threshold, outward/inward iterations, use information from previous refitting iteration) depending on the detector subpart.

Several types of tracks are reconstructed, adapted to different physical purposes. For instance, for triggering purpose parameters are adapted for faster computation. Most offline analyses depend on a highly efficient track reconstruction. For this purpose, the parameters of the reconstruction are tuned for a higher track finding efficiency, which requires more CPU resources.

4.3 Vertices

A precise reconstruction of the origin of the products of an event, the Primary Vertex (PV), is essential for different high level triggers as well as for offline analyses [93]. Tracks fulfilling quality criteria and compatible with an origin within the beam spot (where previous interactions took place on average) are clustered. This clustering is done based on the z coordinate of their point of closest approach to the center of the beam spot. The algorithm has to cope with vertices close to each other due to high PU, as presented in Sec. 3.3, so a statistical method is utilized, generating groups of tracks made of those which are likely to originate from a common vertex. Finally, a fit is performed for each vertex candidate in order to estimate its parameters out of the associated tracks. Each track gets a weight w_i related to the probability of the track to belong to the vertex. A number of degrees of freedom of the vertex is calculated $n_{\text{dof}} = -3 + 2 \cdot \sum_i w_i$ which is a measure for the number of tracks associated to the vertex and hence for its significance. Finally, the hard-process PV is chosen as the one showing the largest sum of p_T over the tracks associated to it.

4.4 Electrons and Photons

An electron produced in the interaction region follows a curved trajectory in the detector because of the magnetic field. The curve runs in principle with a constant η and a constantly varying ϕ . The electron is detected by a track in the silicon detector and, if it has enough momentum, by an energy deposit in the electromagnetic calorimeter where it is absorbed. The correlation of these measurements provides a good identification of electrons.

Since the electrons traverse an orthogonal magnetic field, they emit bremsstrahlung photons. This energy loss leads to an increase of the curvature of the electron trajectory during the flight and the bremsstrahlung photons deposit energy in the electromagnetic calorimeter in the vicinity of the electron. Further bremsstrahlung photons are emitted due to interactions with the detector material. The increase of the curvature identified by the tracker as well as the presence of the bremsstrahlung deposits associated to the electron deposits in the electromagnetic calorimeter are taken into account for the identification of electrons and the precise measurement of their energy. Using the calorimeter measurement only, the electron momentum resolution is about 5% and 12% at very low p_T (≈ 10 GeV) in the barrel and the endcap region. It is reduced to 2% and 11% when using the information

from the tracker in addition [95]. More details on the electron reconstruction algorithm in CMS can be found in [96].

In case no track originating from the PV is found to correspond to an energy cluster in the electromagnetic calorimeter, a photon can be reconstructed. Photons are likely¹ to annihilate into an e^+e^- pair before entering the electromagnetic calorimeter because of the detector material traversed beforehand, this is called photon conversion. Finding oppositely-charged tracks with a possible common vertex and compatible with a zero invariant mass enables to identify such cases. This identification is also essential for the elimination of electrons from photon conversion, when only electrons from the hard process are searched. The energy resolution is found to be between 1 and 4% depending on η , on the detector coverage, and on the shape of the electromagnetic shower [98]. More on the photon reconstruction can be found in [99].

4.5 Muons

A muon produced in the interaction region is measured with a curved track in the silicon detector and in case it has enough momentum it is also measured in the muon chambers. These two tracks should have opposite curvatures because of the structure of the CMS magnetic field. The association of both tracks provides a robust identification of muons, since these particles are supposed to be the only ones able to traverse the whole detector up to the muon chambers. However, interactions with the detector material can produce hadrons, mostly π^\pm , which can reach the muon system. Thanks to low-energy deposits from muons traversing the calorimeters, their identification and reconstruction can be improved.

The tracks from the tracking system and from the muon chambers are associated in two different ways:

- Global-muon - For every track found in the muon system (called stand-alone muon) a matching track from the tracking system is searched. The two matching tracks then constitute a global-muon. The track of this global-muon is re-fitted combining hits from the track of the silicon detector and from the track of the muon chamber.
- Tracker-muon - Every track from the silicon detector with a sufficient momentum is considered as a muon and is extrapolated to the muon

¹The photon conversion probability is approximately 27% at the center of the barrel ($\eta = 0$) and increases to 62% at the end of the ECAL barrel ($|\eta| = 1.4$) [97].

chambers. If at least one segment in one of the subparts of the muon chambers matches the extrapolation of the tracker track, the candidate is defined as a tracker-muon track.

If both associations fail, the stand-alone muon is called a stand-alone muon track only.

An evaluation of the muon-track p_T resolution measured with the tracking system, the muon system only, and with the combination of both measurements is depicted in Fig. 4.1. For low-momentum muons ($p_T < 10$ GeV) the accuracy is dominated by the tracking system. It reaches 0.5% in the central region of the detector ($|\eta| < 0.2$) and 1.5% in the forward region ($2.0 < |\eta| < 2.4$). At high momentum (2 TeV) the accuracy degrades strongly for the tracking system. The combination with the muon system measurement, however, improves the overall resolution to $\approx 5\%$ in the central region and 30% in the forward region. More information about the muon reconstruction can be found in [100].

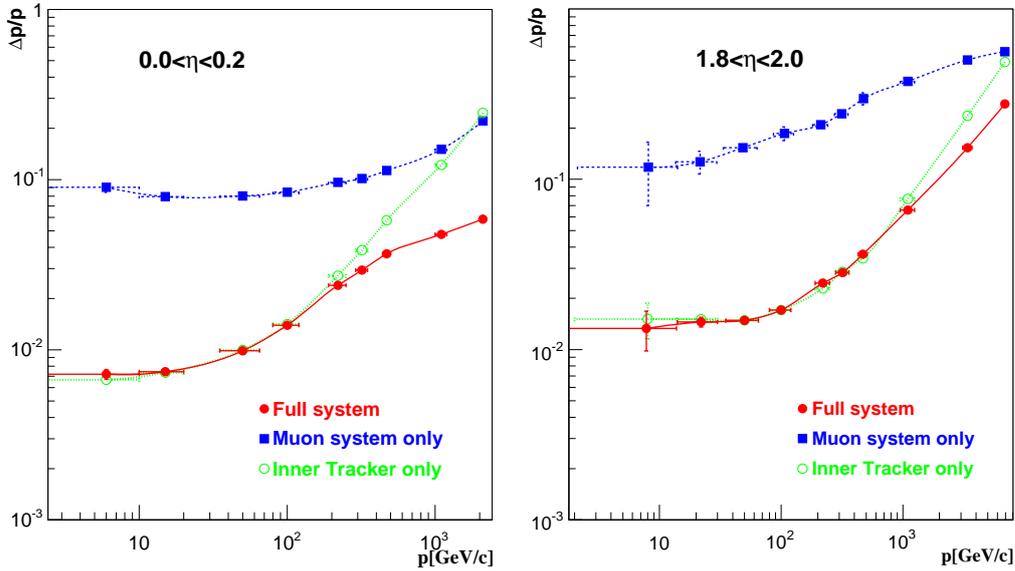


Figure 4.1: Momentum resolution for muon tracks measured at low η (left) and high η (right) using the tracking system only, the muon system only, or the combination of both [79].

4.6 Jets

Several kinds of jet-reconstruction algorithms exist, performing the clustering of energy deposits in the detector. An overview of the present development in the field of jet reconstruction algorithms can be found in [101]. Two main requirements, the collinear and infrared safety, have to be fulfilled by a jet clustering algorithm in order to reconstruct the energy of the particle which has generated the jet without significant influence from non-perturbative effects (parton showering and hadronization). The collinear safety ensures that, during a parton shower, collinear gluon splittings (splitting with small angle between produced particles) will not lead to changes in the final jet. The infrared safety ensures stability of the clustering in case of additional soft-parton radiation, which could be taken as seeds for the clustering. The use of unsafe algorithms leads to divergences in the calculations of QCD processes because cancellation of diverging terms tends to fail in case of varying number of clustered jets. Elaborated algorithms fulfill these safety requirements. One of these, used in this work for all jet reconstructions, called anti- k_t , is presented in the following.

The anti- k_t clustering makes use of the whole information provided by the PF technique, i.e. photons, electron, muons, and charged and neutral hadrons. If required it is possible to exclude energy deposition from leptons. This is relevant in the analysis presented in this thesis where the measurement of the high-energy leptons from $t\bar{t}$ decay products do not belong within jets. This requires not only to eliminate the lepton track from the clustering input, also the photons possibly radiated by the lepton need to be excluded. The jet-clustering algorithm recognizes, by means of the p_T of the input particles, which of them belong together in jets. A weighted distance d_{ij} between two clustering objects i and j is used for this purpose. The weighting depends on the p_T of the clustering input:

$$d_{ij} = \min\left(\frac{1}{p_{T,i}^2}, \frac{1}{p_{T,j}^2}\right) \frac{\Delta R_{ij}^2}{R^2} \quad \text{with} \quad \Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$$

The anti- k_t algorithm consists of the following steps:

1. Work out all the d_{ij} and find the smallest one.
2. Find the minimum between d_{ij} and $\frac{1}{p_{T,i}^2}$.
3. If it is d_{ij} , recombine i and j into a single new clustering object and return to step 1.

4. Otherwise, declare i to be a final-state jet and remove it from the list of clustering objects.
5. Stop when there is no clustering object remaining.

In CMS the parameter R , related to the radius of the created jets, has been typically chosen to be 0.5 for LHC run in 2011 and 2012. The reconstruction of k_t jets is very similar but instead of $\frac{1}{p_{T,j}}$ and $\frac{1}{p_{T,j}}$, the inverses $p_{T,j}$ and $p_{T,j}$ are used. A more detailed description of this algorithm can be found in [101].

The measurement of the energy and the direction of jets is influenced by a large number of detector and physical effects. Hence, robust jet measurements require the application of corrections on the reconstructed jets. The corrections applied in this thesis are elaborated and documented by the CMS collaboration, they are presented in Sec. 8.3. Fig. 4.2 shows that the energy resolution of anti- k_t PF jets is approximately 15% at low p_T and improved at high p_T . A discrepancy is observed between the estimation from data and from MC simulation. This is corrected as illustrated in Fig. 4.2 too. The procedure is explained in Sec. 8.3.1.

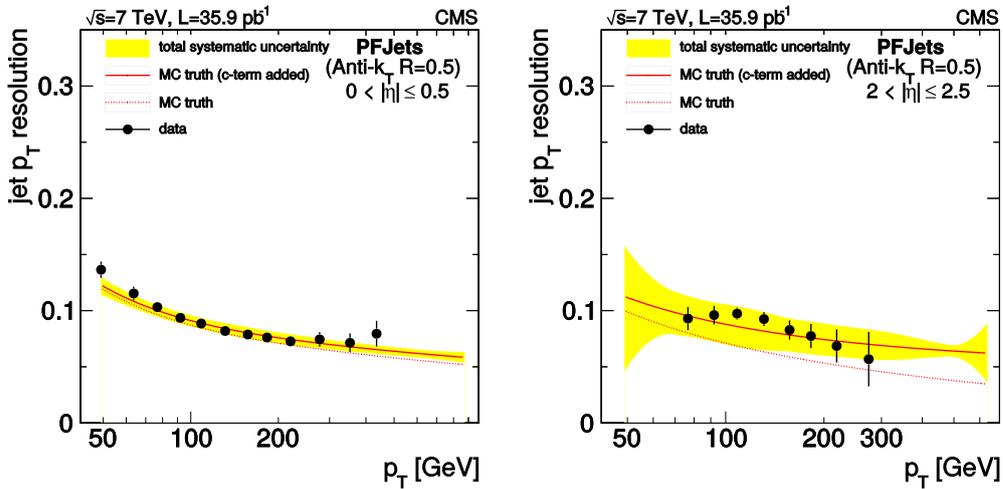


Figure 4.2: Momentum resolution of PF jets from the anti- k_t algorithm at low η (left) and high η (right) estimated in data and MC simulation after the energy-scale corrections have been applied. The yellow band represents the jet-energy scale uncertainty, the red-dashed (red-full) line is the jet-energy resolution in MC before (after) its correction. [102].

Identification of b -Jets As mentioned in Sec. 3.2.1, a jet originating from the parton showering and hadronization of a bottom quark, called a b -jet, can be distinguished from jets originating from other quark flavors or gluons because long-living B -hadrons are produced. Their decay vertices are displaced with respect to the PVs, leading to so-called Secondary Vertices (SV) and the tracks originating from these SVs show a shift with respect to the PVs too. By means of algorithms, called b -tagging algorithms, which make use of b -jet properties (SV, shifted tracks, etc.) it is possible to estimate whether a jet originates from a bottom quark or not.

In CMS, the best b -tagging efficiency is achieved with the CSV (Combined Secondary Vertex) algorithm which combines all available b -jet information of a jet into a single discriminator [103]. Beyond the straightforward variables (mass of SV, distance between PV and SV), the number of SV, the number of tracks associated to them, and the fraction of the jet energy carried by the B -hadron are used for instance. Moreover, even jets where no clean SV could be reconstructed are used, because they still might contain b -tagging information like the track multiplicity and their shift with respect to the PV. Fig. 4.3 shows the non- b -jet misidentification probability as a function of the b -jet efficiency for the CSV algorithm and further b -tagging algorithms. The operating point utilized in this thesis is defined with a misidentification probability of udsg-jets (jets originating from gluons or from up, down, and strange quarks) of 1%, this corresponds on Fig. 4.3 to an identification efficiency of b -jets of $\approx 67\%$ for the CSV tagger. Moreover, the histograms show that for most working points, the CSV algorithm achieves the best identification efficiency, compared to algorithms using only the information from the track shift (TCHE and TCHP) or from reconstructed secondary vertices (SSVHE and SSVHP).

4.7 Missing Transverse Energy

The initial state of proton-proton collisions at the LHC has an overall p_T (\vec{p}_T^{all}) which equals zero. As a consequence of the momentum conservation no \vec{p}_T^{all} should be seen in the final state of a collision, too. However, particles like neutrinos and unknown particles, e.g. neutralinos from SUSY, cannot be detected by the CMS detector because they almost do not interact with matter. Since the detector is built almost hermetically around the collision point, any non-zero \vec{p}_T^{all} detected is a sign of such invisible particles. To be more precise, the negative of the \vec{p}_T^{all} can be assumed to be the resulting p_T of the invisible particles. The energy corresponding to the p_T which is missing to

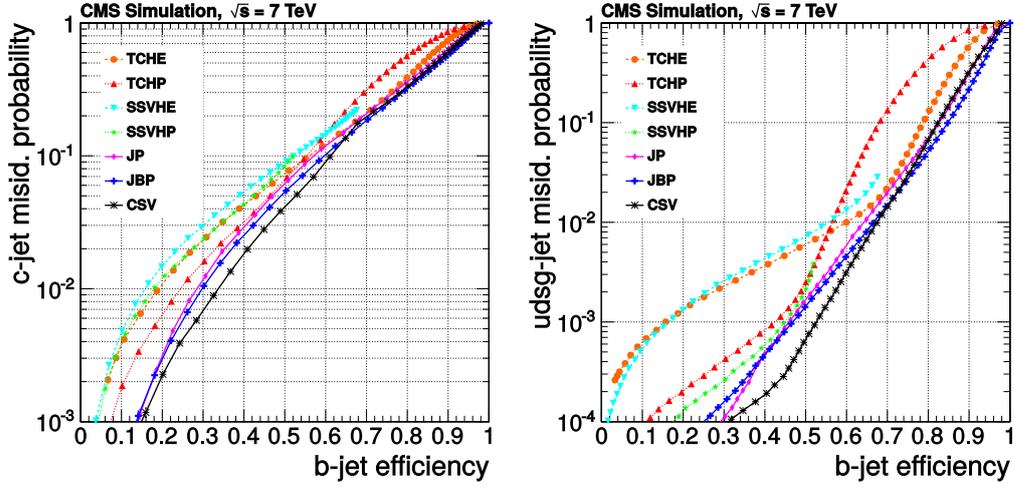


Figure 4.3: Performance curves of various b -tagging algorithms with MC simulation [103]. On the y -axis the misidentification probability of jets originating from charm quarks (c -jets, left) or from gluons and up, down, and strang quarks ($udsg$ -jets, right) is shown as a function of the identification efficiency of b -jets on the x -axis.

eliminate the \vec{p}_T^{all} is called the missing transverse energy (\cancel{E}_T). While an equivalence is valid between the \cancel{E}_T and e.g. the neutrino in case exactly one neutrino has been produced, cases with more than one neutrino are more complex, but this is not required in this thesis. The performance of the \cancel{E}_T reconstruction is improved by the use of the PF technique with respect to its reconstruction with calorimeter only. This is due to the identification of each particle, where those not originating from the hard process can be eliminated. A measurement of the performance of \cancel{E}_T reconstruction with CMS can be found in [104].

Chapter 5

Analysis Techniques

Several analysis techniques are employed in the measurement presented in this thesis. A rough description is given for the maximum likelihood in Sec. 5.1 and the data-unfolding mechanism in Sec. 5.2.

5.1 Maximum-Likelihood Method

A common task in data analysis is the estimation of the parameters of a model, which is assumed to describe data. In this thesis, a template fit to data is performed, and a likelihood is calculated in order to decide what is the best fit. This example is explained here.

In a template fit, the probability distribution of a given variable, called template in this context, is created for each part of the model, i.e. for the $t\bar{t}$ and the background processes, using the MC simulation. The fitting procedure consists in varying the cross section of each template in order to achieve the best agreement between the sum of the templates and the data distribution. The task of the likelihood is to find which set of cross sections is the best one. The variable chosen for the fit has to provide a sufficient discrimination power, i.e. the shape of the templates from the different processes have to be different, otherwise the templates will act redundantly. For instance, one of two very similar templates is chosen to represent both, which is likely to give wrong results.

The probability of the fitting variable to take the value x is given by the probability density function $f(x|\sigma)$, obtained from the various models with the normalization parameters σ . The likelihood function L , as explained e.g. in [105], gives the probability of the model to describe the set of N data

events, showing the values $y_0, y_1, \dots, y_i, \dots, y_N$ of the chosen variable:

$$L(\sigma) = \prod_{i=0}^N f(y_i|\sigma). \quad (5.1)$$

The likelihood function reaches its maximum if σ take the best values, i.e. the sum of the models assuming the parameter values σ has the highest possible agreement with data. Hence, the set of σ providing the highest likelihood has to be searched by the fit. This is explained in the last paragraph of this section.

For a binned probability density function, as it is the case for the fit performed in this thesis, the probability to find n_j data events in the bin j with an expected number of events μ_j given by the model, is a Poisson probability

$$P(n_j|\mu_j) = \frac{\mu_j^{n_j} e^{-\mu_j}}{n_j!}.$$

The expected number μ_i is equal to $f(x|\sigma)$ integrated in the interval of the bin j and n_j is the number of data events falling into this bin. This probability can be inserted in Eq. 5.1 in the place of $f(y_i|\sigma)$. Moreover and for convenience, the likelihood function is transformed into the negative log-likelihood function:

$$-\ln [L(\sigma)] = -\ln \left[\prod_{j=0}^{N_b} \frac{\mu_j^{n_j} e^{-\mu_j}}{n_j!} \right] = -\sum_{j=0}^{N_b} n_j \ln \mu_j + \sum_{j=0}^{N_b} \mu_j + \sum_{j=0}^{N_b} \ln n_j!,$$

where N_b is the number of bins considered. The $-\ln [L(\sigma)]$ needs to be minimized during the fit. In this context the last term can be ignored, since it does not depend on σ .

Constraints on the parameter values can be added to the likelihood function in order to support some values rather than others, based on prejudices on the parameters. A typical constraint is the limitation of the deviation of a parameter a_i from its initial value $a_{i,0}$, here $\beta_i = a_i/a_{i,0}$ is constrained to 1. A Gaussian-shaped constraint is a priori a natural choice. The likelihood from Eq. 5.1 is then extended to:

$$\prod_{i=0}^N f(y_i|\sigma) \cdot c(\beta_i, \varepsilon_i), \quad \text{with} \quad c(\beta_i, \varepsilon_i) = \frac{1}{\varepsilon_i \sqrt{2\pi}} e^{-\frac{(\beta_i - 1)^2}{2\varepsilon_i^2}},$$

where μ_i is the value to which the fit parameter β_i is constrained, i.e. typically one. The relative standard deviation, ε_i , constrains the fit parameter β_i .

However, if the model is not accurate or if any process included in the model has a low discrimination power, the minimization of $-\ln [L(\sigma)]$ can lead to negative parameter values, which is unphysical and not excluded by a Gaussian-shaped constraint, especially in case of soft constraints. Then it makes sense to use a log-normal constraint, as it is done in this thesis:

$$c(\beta_i, \varepsilon_i) = \frac{1}{\beta_i \varepsilon_i \sqrt{2\pi}} e^{-\frac{(\ln \beta_i)^2}{2\varepsilon_i^2}},$$

because if $\beta_i \rightarrow 0$, then $c(\beta_i, \varepsilon_i) \rightarrow 0$, and hence $-\ln [c(\beta_i, \varepsilon_i)] \rightarrow +\infty$.

Finally, the minimization of $-\ln [L(\sigma)]$ is required in order to find the best set of σ . Depending on the complexity of the function to minimize, different techniques are recommended, see for instance [105]. The minima of $-\ln [L(\sigma)]$ correspond to positions in parameter space where all gradients of $-\ln [L(\sigma)]$ are zero. Algorithms exist in order to optimize the search. In this thesis, the minimization is performed with the MIGRAD tool included in the MINUIT package [106], using a variable-metric method [107].

5.2 Data Unfolding

Instruments used in experimental physics can have a significant impact on the measurements. Not only statistical uncertainties but also systematic bias can occur, constituting the response of the instrument. For instance, in the frame of the jet-multiplicity measurement presented in this thesis, the number of jets produced by a proton-proton collision, e.g. in a $t\bar{t}$ event, can be different from the number of jets measured by the detector because of several effects, such as non-sensitive regions of the detector and detection resolution and efficiency. A statistical method, called unfolding, is used in this thesis in order to retrieve the measurement which would have been obtained with a totally unbiased instrument. The example of the jet multiplicity is utilized in this section to describe the unfolding. More information can be found e.g. in [105].

The real spectrum of the jet multiplicity can be represented as a vector \mathbf{f} of dimension n , where the element f_j is the number of $t\bar{t}$ events produced with j jets. The vector \mathbf{f} is related to \mathbf{g} , the vector of the detected jets (dimension m) by:

$$\mathbf{g} = \mathbf{A} \cdot \mathbf{f}, \quad (5.2)$$

where A_{ij} is an element of the migration matrix \mathbf{A} of dimension $m \times n$, corresponding to the probability to observe i jets when actually j jets were

produced in the event. The migrations are preferably limited, hence \mathbf{A} is almost diagonal. Thanks to the full detector simulation applied on the MC simulation, an estimation of the response matrix is available.

The naive way to unfold the measured jet-multiplicity spectrum would be to invert \mathbf{A} , providing directly the desired jet-multiplicity vector \mathbf{f} . In this case, a fit to the measured distribution is performed, making use of the Eq. 5.2, translated into m equations of the n unknown components of the desired vector \mathbf{f} :

$$\hat{g}_i(\mathbf{f}) = \sum_{j=0}^n A_{ij} \cdot f_j, \quad (5.3)$$

where $\hat{g}_i(\mathbf{f})$ are the estimators of the components of \mathbf{g} . A negative log-likelihood is calculated in order to estimate how likely it is for the estimated vector \mathbf{f} to lead to the measured jet-multiplicity distribution. The expectations are given by the Poisson distributions of mean values $\hat{g}_i(\mathbf{f})$, and hence the variable terms of the likelihood function are:

$$S(\mathbf{f}) = \sum_{i=0}^m [\hat{g}_i(\mathbf{f}) - g_i \cdot \ln \hat{g}_i(\mathbf{f})], \quad (5.4)$$

which is the variable to minimize.

However, this technique can lead to a high sensitivity of the result to the statistical fluctuations in \mathbf{g} , leading to large anticorrelations between neighbouring members of the result, \mathbf{f} . A regularization is utilized in order to mitigate these fluctuations, i.e. an assumption of smoothness of the resulting distribution is introduced in the function to minimize. Concretely, the curvature of \mathbf{f} is calculated and added to the likelihood from Eq. 5.4 so the function to minimize looks as following:

$$R(\mathbf{f}) = S(\mathbf{f}) + \tau \cdot C(\mathbf{f}), \quad (5.5)$$

where $C(\mathbf{f})$ represents the curvature of \mathbf{f} and τ is the regularization parameter. The latter has to be optimized in order to effectively reduce the unwanted fluctuations without distorting \mathbf{f} . For this purpose, the averaged square global correlation of \mathbf{f} is calculated [108]:

$$\bar{\rho}(\tau) = \frac{1}{n} \sqrt{\sum_{i=0}^n \rho_i(\tau)^2}, \quad \text{with } \rho_i(\tau) = \max \left\{ \text{Corr}(f_i, \sum_{i \neq j} \alpha_j f_j) \right\}. \quad (5.6)$$

The averaged term $\rho_i(\tau)^2$ is the maximum of the correlation between f_i and any linear combination of all other components of \mathbf{f} , written $\sum_{i \neq j} \alpha_j f_j$ with

$\alpha_j \in \mathbb{R}$ and $i \neq j$, as written in Eq. 5.6. The term f_i depends on the chosen value of τ , and hence $\bar{\rho}$ is written as a function of this parameter. In case of a too weak or a too strong regularization, $\bar{\rho}(\tau)$ increases [109]. Therefore, the optimized value of τ is chosen as the one leading to the minimal $\bar{\rho}(\tau)$.

Part II

Study of the Jet Production in Top-Quark Pair Events with CMS

Chapter 6

Summary of a Jet-Multiplicity Measurement with Data from 2011

In this chapter, measurements published in the European Physics Journal C, “Measurement of jet multiplicity distributions in $t\bar{t}$ production in pp collisions at $\sqrt{s} = 7$ TeV”, from August 2014 [35], are summarized. This paper presents analyses of $t\bar{t}$ events selected from data recorded by the CMS experiment in 2011 in proton-proton collisions at $\sqrt{s} = 7$ TeV and corresponding to an integrated luminosity of $\mathcal{L} = 5 \text{ fb}^{-1}$. The $t\bar{t}$ production has been studied in this paper with a focus on jets accompanying these events in the dileptonic and the ℓ +jets channel. On the one hand, jets are expected from $t\bar{t}$ decay products, but this is expected to be very well modeled. On the other hand, the occurrence of emissions leading to additional jets is the property tested in this paper. This measurement enables to probe perturbative QCD at the top-quark scale and it has some sensitivity to new physics which is likely to enhance the production of $t\bar{t}$ + jets events. The measurement presented later in this thesis (called here “8 TeV analysis”) improves one of those performed for [35] in the ℓ +jets channel, making use of data from proton-proton collisions at $\sqrt{s} = 8$ TeV. The author of the present thesis has strongly contributed to [35] in the ℓ +jets channel¹. This paper was presented at the TOP2014 conference in Cannes (France), see the proceedings article in [110]. The general setup of the measurements presented in this chapter is mostly

¹More precisely, the measurements in the ℓ +jets channels were performed and documented by the KIT group, namely the author of these lines as main contributor, and Olaf Böcker, Prof. Ulrich Husemann, Dr. Patricia Lobelle, Hannes Mildner, and Shawn Williamson.

compatible with the one of the 8 TeV analysis. Therefore, the steps prior to the measurement: event selection, background estimation, and MC-simulation and data samples and their corrections are not described in detail here. The measurement of the jet multiplicity in $t\bar{t}$ events has been performed with two alternative methods. The $t\bar{t}$ differential cross section as a function of the number of jets has been measured at a detector-independent level (via unfolding). This is presented in section Sec. 6.1. A template fit to data has been performed, in which the $t\bar{t}$ MC simulation has been categorized as a function of the number of additional jets. This procedure enables the measurement of the $t\bar{t}$ differential cross section as a function of the number of *additional* jets, which is presented in Sec. 6.2. Some discussion, motivating the improvements for the 8 TeV analysis, follows in Sec. 6.3.

6.1 Measurement of the Jet Multiplicity

The first measurement presented in [35] focuses on the number of jets found in $t\bar{t}$ events. The data and MC-simulation events are selected for the ℓ +jets channel (e +jets or μ +jets) and then the background contributions (from MC simulation or data-driven methods) are subtracted from the jet-multiplicity spectrum found in data. The main background contaminations arise from single top events and from $t\bar{t}$ events decaying into channels other than ℓ +jets. The identification of the decay channel in the MC simulation succeeds thanks to parton-level information. Further background events are those from pure QCD processes (QCD multijet) and those containing a W or a Z/γ^* boson, decaying into leptons. The modeling of the QCD-multijet and W +jets background processes is improved by means of data-driven techniques, whereas the modeling of all other processes fully relies on MC simulation.

The resulting distribution of the number of selected jets is corrected by means of unfolding in order to invert migration effects in the jet-multiplicity distribution which originate from efficiency and resolution effects of the detector. The resulting event yield is extrapolated to a fiducial phase-space by a last correction applied bin-by-bin. These corrections are required in order to provide a measurement which is detector independent, hence it can be reproduced by groups outside the CMS collaboration. The same correction techniques are performed in the 8 TeV analysis, described in Sec. 11.1.

Finally, the jet-multiplicity distributions in the e +jets and μ +jets channels are converted to normalized differential fiducial cross sections as a function of the number of jets. The published results are obtained by the combination of both channels via an average weighted by the uncertainties. This is illus-

trated in Fig. 6.1, where the results are also compared to predictions from various MC simulation tools for a measure of the modeling uncertainties. The same tools with similar configurations are utilized in the 8 TeV analysis, see Sec. 8.2.

The agreement is very good between the results and the predictions from MADGRAPH +PYTHIA and POWHEG+PYTHIA up to 7 jets, while the prediction for 8 jets is slightly outside the measurement uncertainties. MC@NLO+HERWIG predicts too few jets, this is significant for 6 jets and more. The factorization and renormalization scale (Q^2) as well as the jet-energy threshold for the interface between matrix-element and parton-showering steps (matching threshold) are varied up and down in the MADGRAPH +PYTHIA configuration for alternative predictions. The agreement between these predictions and the results is modified, but not significantly, with regards to the measurement uncertainties. The best description is provided by POWHEG+PYTHIA.

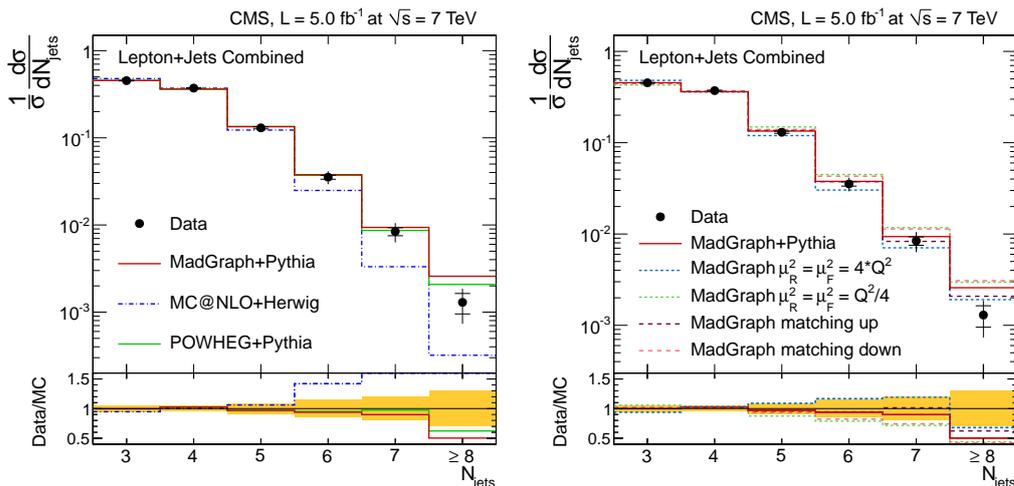


Figure 6.1: Combined normalized differential fiducial $t\bar{t}$ cross section as a function of the number of detector-independent jets in the ℓ +jets channel, measured with 7 TeV data and published in [35]. The results are compared to predictions from various generators (left) or from MADGRAPH with various parameter sets (right). The vertical error bars represent the total uncertainties and the intersecting horizontal bars represent the statistical uncertainties alone.

In [35], the same measurement is performed with $t\bar{t}$ events decaying into dileptonic² channels (ee , $e\mu$, and $\mu\mu$) for two jets p_T requirements: 30 and > 60 GeV. In both cases, compatible results are obtained.

²The measurement in the dileptonic channel has been performed by a partner group from DESY.

6.2 Multiplicity of Additional Jets

A second measurement in [35] comes a step closer to the property of interest, namely the occurrence of additional jets in $t\bar{t}$ events. For this purpose, additional jets have to be identified. In principle, this is not possible in data in the ℓ +jets channel because the two jets expected from the hadronically decaying W boson cannot be distinguished from additional jets. However, it is possible to distinguish events containing different numbers of additional jets, if one can find a variable sensitive to the presence of additional jets in an event, i.e. a variable distributed differently for events containing different numbers of additional jets.

Performing a full event reconstruction, jets are assigned to the hypothetical $t\bar{t}$ decay products. Several combinations are possible, especially if more than four jets are found in the event evaluated. A way to choose the best combination of jets is to calculate, for each combination, a test statistic χ^2 which compares the masses of the reconstructed top quarks and the hadronically decaying W boson with the theoretical values. The combination with the lowest χ^2 is the one providing the most correct masses, so it is likely to be the right reconstruction. The χ^2 shows a sensitivity to additional jets for events with the same number of jets. For instance, in an event with five jets, there should be one additional jet. However, it is possible that this event has two additional jets if one of the jets expected from the $t\bar{t}$ decay is not measured (e.g. too low p_T), then the absence of this expected jet inhibits a correct full event reconstruction, leading to larger χ^2 values.

The additional jets are identified in the $t\bar{t}$ MC simulation by means of a matching of non-additional jets to $t\bar{t}$ decay products. In the ℓ +jets decay channel, the decay products are the two bottom quarks, the two quarks from the hadronically decaying W boson, and the lepton from the decay of the other W boson. A jet is identified as additional if its ΔR to each decay product is larger than 0.5. This enables to build categories of $t\bar{t}$ events: $t\bar{t} + 0$, $t\bar{t} + 1$, and $t\bar{t} + \geq 2$ additional jets. Only three categories can be built because of the limited number of events available for ≥ 3 additional jets. One should note here that the jets used for the categorization of the events are those obtained before the detector simulation. The quarks utilized for the matching originate from the parton-level information of the MC simulation. Lastly, the selected $t\bar{t}$ MC-simulation events from the three separate categories are fitted to data in a template fit. The fit variable is square root of the test statistic χ^2 for the jet combination with the lowest χ^2 in each event. The fit variable is $\sqrt{\chi^2}$ because it shows higher sensitivity than χ^2 since it stretches the χ^2 distribution in the low- χ^2 region, where most discrimination

power is located. As explained previously, the χ^2 test statistic provides discrimination between events with the same number of jets. Therefore, the fit templates are filled with the $\sqrt{\chi^2}$ of selected events in three separate regions: events with 4, 5, and ≥ 6 jets, separately in the e +jets and μ +jets channel since different background contaminations are expected. The fit of all six resulting regions is performed simultaneously. An illustration of the fit is given in Fig. 6.2.

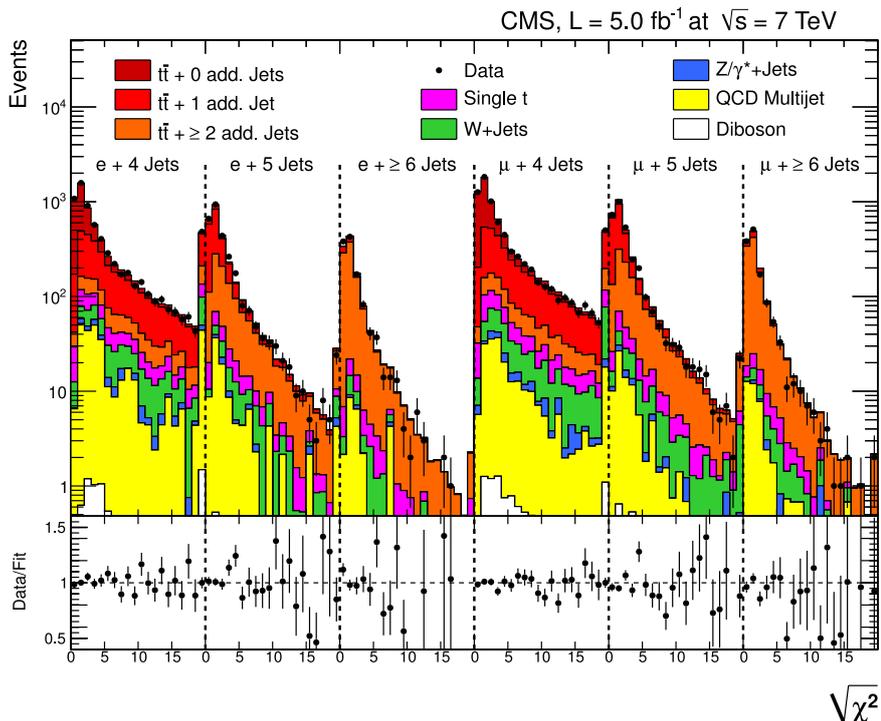


Figure 6.2: Result of the template fit performed simultaneously in both lepton channels and the three jet-multiplicity regions (4, 5, and ≥ 6 jets) with 7 TeV data and MC simulation, as published in [35].

Beyond the differences of event yields in the split jet-multiplicity regions (more events with zero additional jets than with one or at least two additional jets in the four-jet region), which is expected, some discrimination is observed within each jet-multiplicity region where the templates of different additional-jet multiplicity show different shapes.

The fit results are translated into a normalized differential fiducial cross section as a function of the number of additional jets. For this purpose, the differential cross section is calculated from $t\bar{t}$ MC simulation, making use of the categorization described before, with events selected with detector-

independent variables. The latter selection defines the fiducial phase space, in the same spirit as explained in Sec. 8.1.5. The normalization of the three $t\bar{t}$ categories obtained in the fit are applied to the predictions. Finally, the distribution is normalized to unity, providing the distribution given in Fig. 6.3.

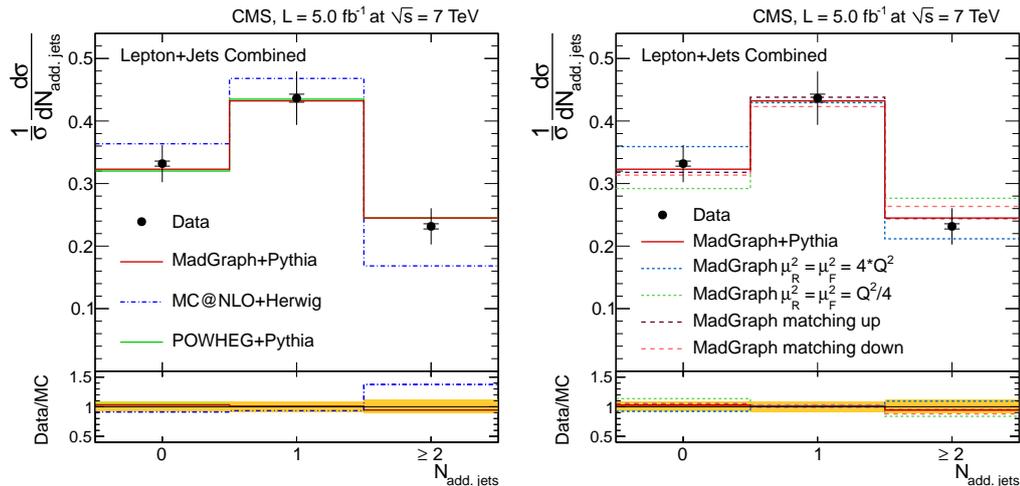


Figure 6.3: Combined normalized differential fiducial $t\bar{t}$ cross section as a function of the number of detector-independent additional jets in the ℓ +jets channel, measured with 7 TeV data and published in [35]. The results are compared to predictions from various generators (left) or from MADGRAPH with various parameter sets (right). The vertical error bars represent the total uncertainties and the intersecting horizontal bars represent the statistical uncertainties alone.

Similarly to the conclusions from the previous section, the predictions from MADGRAPH+PYTHIA and POWHEG+PYTHIA agree well with the measurement. MC@NLO+HERWIG seems to produce too few jets. A decreased Q^2 -scale value leads, as expected, to more additional jets, and vice versa. The amplitude of the Q^2 -scale uncertainty recommended by the theory community turns into large effects on the results shown here. The measurement uncertainties are partly smaller than the effect of the Q^2 -scale uncertainty on the predictions. The predictions from MADGRAPH+PYTHIA with varied matching threshold are closer to the nominal predictions and agree with the measurement, according to the uncertainties.

In the dileptonic channel, the identification of additional jets is less laborious since any non- b -tagged jet should be additional. Therefore, a kinematic study of jets identified as additional could be performed in [35]. The rate of events showing no additional jet with p_T larger than a given threshold has been

calculated in dileptonic events. This rate, called gap fraction, is measured in a large jet- p_T range and comparing the results to predictions from various generators provides a study of the kinematics of additional jets. Conclusions on the quality of MC descriptions are different from the two previous jet-multiplicity measurements. MC@NLO+HERWIG provides a better kinematic description of the first additional jet than MADGRAPH+PYTHIA and POWHEG+PYTHIA. Increasing the Q^2 -scale value also improves the description of this property by MADGRAPH+PYTHIA.

6.3 Conclusion and Outlook

The normalized differential fiducial cross section of $t\bar{t}$ production as a function of the number of jets and of additional jets has been summarized in this chapter, more details can be found in [35]. The measurement allows to test the predictive power of the MC simulation for heavy-flavor production with high emission multiplicities. The predictions from MADGRAPH+PYTHIA and POWHEG+PYTHIA are in very good agreement with the measurement whereas predictions from MC@NLO+HERWIG show too few jets. The statistical uncertainties become significant at the largest jet multiplicities. Therefore, using the larger set of data provided in 2012 by the LHC (four times more data) at $\sqrt{s} = 8$ TeV should enable to achieve a better precision. It should be noted that systematic uncertainties originating from modeling uncertainties (Q^2 -scale and matching threshold uncertainties, explained in Sec. 11.3) show large values for the high jet multiplicities where this is partly due to the low number of events available in this region in the MC-simulation samples used for their estimation. With larger MC-simulation samples, these systematic uncertainties would probably have been smaller.

The results have been published as a normalized cross section in order to cancel systematic uncertainties which cause normalization variations. However, this affects the comparison with the predictions because the cross sections with the largest yields (3-4 jets and one additional jet) set an arbitrary offset to the curves. This means that MC models show good agreement for high-yield multiplicities whereas the agreement is worse in the low-yield multiplicities. In Fig. 6.3 for instance, the prediction from MC@NLO+HERWIG seems to be wrong for zero additional jets and ≥ 2 additional jets. However, without a normalization to unity, the picture could be different showing that the prediction from MC@NLO+HERWIG is right for zero and one additional jet, as it is actually expected. The 8 TeV analysis, described from the

next chapter onwards, presents non-normalized results, so the comparisons are not biased.

The use of parton-level information from the $t\bar{t}$ MC simulation in this analysis is significant: identification of b -jets before the detector simulation and identification of additional jets via a geometrical matching between quarks and jets. This is a drawback when attempting to compare different predictions because their internal technical treatment of color-charged particles can differ, partly based on empirical assumptions. Only hadrons, leptons, except taus, and photons are stable particles (mean lifetime $> 10^{-10}$ s) corresponding to observable objects. During the TOP2014 conference, such statements were expressed by the theory community, see for instance [111]. A recommended solution is utilized in the 8 TeV analysis, using no parton-level information for the definition of the b -tagging for jets in the MC simulation before the detector simulation. The adaptation of the measurement of additional jets in order not to use the parton-level information happened to be more complex. Therefore, this latter measurement could not be reproduced with 8 TeV data in time for this thesis.

In order to increase the potential application of the 8 TeV measurement for a broader audience, the RIVET tool has been utilized. This tool enables all groups of the particle-physics community to perform the same analysis on MC simulation by means of a compatible software. For this purpose, only particle-level information, i.e. no parton-level information (quarks or gluons) and only particles before the simulation of the detector effects, is allowed. An implementation of the jet-multiplicity analysis in RIVET is written for the 8 TeV analysis, as presented in Sec. 8.1.5.

A CMS measurement of the jet flavors in W +jets events is used in [35] in order to include rough corrections of the rates of $W+b$, $W+c$, and W +light (i.e. the remaining flavors and gluons) in the MC simulation. Moreover, a data-driven estimation of the normalization of the W +jets MC simulation is performed, making use of the lepton charge asymmetry expected at the LHC. A method for an independent estimation of these background processes in a centralized way is proposed in the 8 TeV analysis, see Sec. 9.

The improvements mentioned above are included in the 8 TeV analysis, as described in the following chapters.

Chapter 7

Analysis Strategy

In the previous chapter, a measurement of the jet multiplicity with 7 TeV data has been presented. It has motivated improvements for the reproduction of this analysis with the larger number of data events delivered in proton-proton collisions at 8 TeV. This chapter gives an overview of the steps required for the execution of this analysis, which is described in details in the following chapters. Some parts of the text in all following chapters of the present document originate from the CMS-internal analysis note [112], written by the author of this thesis in order to document this measurement within the CMS collaboration.

Data and MC Simulation Data events from proton-proton collisions recorded by the CMS detector and MC simulation of all signal and background processes are processed in order to be stored locally in a light format, conserving only the information required for the analysis.

Selection Events in data and MC-simulation samples are skipped if they do not fulfill one of the selection criteria roughly described in the following:

- The events utilized for the measurement itself have to contain exactly one charged lepton (electron or muon) and at least four jets. At least two of these jets have to be b -tagged. See Sec. 8.1.4 for more details.
- Events in an independent control sample are also kept. They have to show exactly one charged lepton and two or three jets. At least one of these jets has to be b -tagged. These events are utilized for the data-driven estimation of background processes: QCD multijet and W +jets.

- In order to extract a model for the QCD-multijet background from data, data events are also kept if they fulfill one of the two previous selections, but applying loosened lepton-selection criteria.
- In the MC simulation, particle-level objects, obtained after the parton-showering and hadronization steps but before the detector simulation, are used for the definition of a detector-independent, so-called fiducial, phase space. The selection of this phase space, in which the results are defined, is described in Sec. 8.1.5.

Corrections Corrections are applied in order to account for any bad modeling like the b -tagging and the lepton-identification efficiencies in the MC simulation. Jets in data events are also corrected in order to take into account the detector response. All corrections applied are presented in Sec. 8.3.

Data-Driven Background Estimation Selected events containing two or three jets are utilized for the estimation of the normalization of background processes for which the background modeling cannot rely on MC simulation. A template fit to a distribution in data events is performed. This is described in Chapter 9. The resulting event yields are shown in Chapter 10 together with comparisons of variable distributions.

Correction of Detector Effects The measured number of jets is distorted by detection and reconstruction effects specific to CMS. In order to produce detector-independent results, these modifications have to be corrected. This procedure is made possible by the detector simulation which is applied on the MC-simulation events. This step is applied on selected data events after all non- $t\bar{t}$ background contributions have been subtracted. These corrections are presented in Sec. 11.1.

Calculation of the Results Finally, the number of corrected data events is converted to a cross section, dividing it by the integrated luminosity, $L \approx 20 \text{ fb}^{-1}$. Each differential cross-section value, as a function of the number of jets, is scaled to the fiducial phase space, making use of the rate of $t\bar{t}$ events from MC simulation fulfilling the particle-level selection, over those fulfilling the detector-level selection (for each jet multiplicity). The resulting cross sections from the e +jets and the μ +jets channels are added and compared to predictions from various MC tools, at LO and NLO. These final steps are described in Chapter 11.

Chapter 8

Analysis Setup

The previous chapter has provided an overview of the structure of the analysis. The first steps consist in selecting the data and the MC simulation and applying corrections required for a good modeling of the data.

8.1 Event Selection

This analysis focuses on $t\bar{t}$ events decaying in the semi-leptonic channel with an electron or a muon, summarized as:

$$t\bar{t} \rightarrow W^+W^-b\bar{b} \rightarrow \mu\nu_\mu(e\nu_e)qq'\bar{b}\bar{b}.$$

The final state is characterized by the presence of one high- p_T isolated electron or muon (originating from a W -boson decay) associated with \cancel{E}_T which is caused by the undetected neutrino. Another attribute is the production of four quarks at leading order of the perturbation theory, leading to the appearance of four jets on average, two light jets coming from the hadronization of the light quarks, originating from the hadronic decay of a W boson, and two heavy-flavor jets, deriving from the bottom quarks hadronization. The first challenge of the events selection is to eliminate the huge QCD-multijet background, showing the largest cross section at the LHC, ≈ 60 mb. These events are those involving no heavy particle like W/Z boson or heavier particles in the hard process, but basically gluons and quarks (except top quarks). A trigger cut requiring high-energy charged leptons eliminates most of the QCD-multijet events since they have no high-energy lepton production channel, unlike the signal process. The high-energy charged leptons are well

adapted for triggering because they can be identified with excellent efficiencies.

High-energy leptons can be produced in W or Z/γ^* -boson¹ production events, whose cross sections, ≈ 38 and ≈ 4 nb is also much larger than the signal. See the Feynman diagrams in Fig. 8.1.

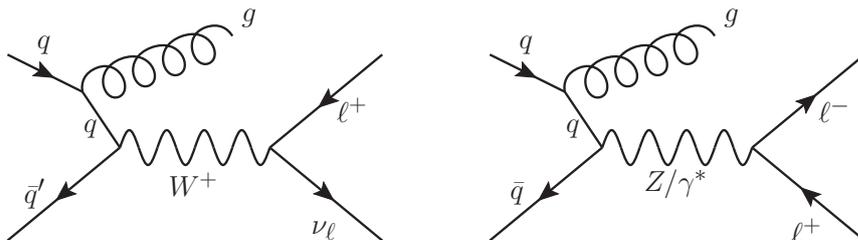


Figure 8.1: Feynman diagrams for W +jets and Z/γ^* +jets production at the leading order of the perturbation theory. On the left: the background process of the W -boson production associated with a radiated gluon. The W boson is produced by the annihilation of a quark q and an anti-quark \bar{q}' of a different flavor and it decays into an anti-lepton ℓ^+ and the neutrino of the corresponding flavor ν_ℓ . On the right: the background process of the production of Z boson or photon associated with a radiated gluon. The Z boson or photon is produced by the annihilation of a quark q and an anti-quark \bar{q} of the same flavor and it decays into a lepton-anti-lepton pair. In both diagrams, the additional gluon which is radiated by a quark of the initial state shows how additional jets can be produced in such processes.

While Z/γ^* -boson events can be well eliminated by forbidding additional leptons, W -boson events show only one high-energy charged lepton and its neutrino, like in $t\bar{t}$ events. The selection of high-energy charged leptons is described in Sec. 8.1.1 and 8.1.2. The rate of W -boson events found in data is strongly reduced by requiring high-energy jets, since in these events, additional jets only appear through initial or final-state radiation, as illustrated in Fig. 8.1, which is roughly reduced by factors of α_s whereas four jets in average are expected in signal events. For the same reasons, QCD-multijet events are also reduced by this jet requirement. The jet selection is given in Sec. 8.1.3

Signal events always produce two bottom quarks which can be identified thanks to the accurate CMS tracking system. Fortunately, b -jets are very unlikely in QCD-multijet, $W \rightarrow \ell$ +jets, and $Z/\gamma^* \rightarrow \ell$ +jets events. The reason is that bottom quarks are mainly produced by the splitting of additional

¹The high virtuality, required for the photon, if decaying into massive particles, is symbolized by the asterisk on γ .

gluons into bottom quarks. Hence, this production is not only canceled because of a factor of α_s appearing, but also due to the large bottom-quark mass, leading to a limited phase space for the gluon splitting. This property is used for the event selection given in Sec. 8.1.4 by applying b -tagging requirements.

Although tremendous fractions of QCD-multijet and $W/Z/\gamma^*$ +jets events are canceled by the trigger and event selection, a part of it is still able to fulfill the event selection, due to their large cross sections. These non-negligible contributions to the selected data are modeled in this analysis by means of dedicated data-based techniques, presented in Chapter 9.

Further background processes are modeled with MC simulation because they show a relatively high acceptance through the event selection, but their cross sections are smaller than the $t\bar{t}$ cross section. The main background process surviving the full selection is the single-top production because it shows high-energy charged leptons, jets, and b -jets, as the signal. See the Feynman diagrams shown in Fig. 1.4. Lastly the diboson production (WW , WZ , and ZZ) is the smallest background contribution considered. See Feynman diagrams in Fig. 8.2. WW events have a substantial cross section (≈ 50 pb), but bottom quarks required by b -tagging need to be produced via gluon splitting, as in QCD-multijet processes. Bottom quarks can be produced in WZ events in the Z -boson decay, but additional jets need to be radiated and the actual production cross section is almost one order of magnitude smaller than the signal. Finally, ZZ events have a very small cross section and leptons can only be produced in lepton pairs, which is efficiently canceled by the addition-lepton veto presented in the event selection.

The reconstruction of the different objects (charged leptons and jets) used for the event selection is based on the Particle-Flow (PF) techniques after subtracting the pile-up (PU) component of charged hadrons, as mentioned in [113]. See Sec. 3.3 for the introduction to PU. The PF techniques and the subtraction of charged hadrons were presented in Chapter 4.

8.1.1 Selection of Electrons

Electron candidates originate from the PF reconstruction. They are required to pass the following identification criteria and kinematic cuts in order to select high-quality electrons from W -boson decays [114]:

- $p_T > 30$ GeV, because electrons from W -boson decays are expected with a high p_T . Moreover, electron events are triggered for electrons with $p_T > 27$ GeV, see Sec. 8.1.4;

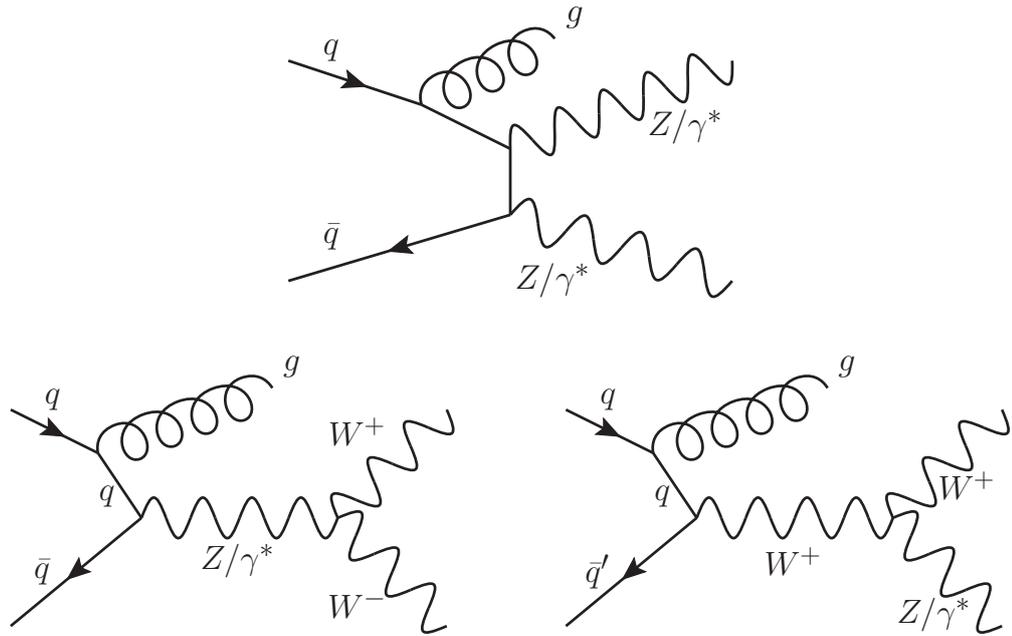


Figure 8.2: Feynman diagrams of the ZZ , WW , and WZ boson production with jets at the leading order of the perturbation theory. Top: the background process of the ZZ diboson production associated with a radiated gluon (photons can be produced instead of the Z bosons). The annihilation of a quark anti-quark pair of the same flavor leads to the production of two photons or Z bosons. Bottom left: the background process of the WW diboson production associated with a radiated gluon. A Z boson or a photon, originating from the annihilation of a quark q and an anti-quark \bar{q} of the same flavor, decays into a W -boson pair. Bottom right: the background process of the WZ diboson production associated with a radiated gluon. A W boson, produced by the annihilation of a quark q and an anti-quark \bar{q}' of a different flavor, radiates a Z boson or a photon. In all diagrams, the additional gluon which is radiated by a quark of the initial state shows how additional jets can be produced in such processes.

- $|\eta| < 2.5$, this is the limit of the tracking system coverage;
- reject electrons located within the interface region between the barrel and the end-cap of the electromagnetic calorimeter: reject if $1.4442 < |\eta| < 1.5660$;
- electrons or positrons originating from photons interacting within the detector material (called conversion photons) represent a background since they do not originate from a top-quark decay. They are rejected by looking for conversion partners in the silicon tracker. A partner candidate is another electron with opposite charge sign. If the electron and its partner candidate build a realistic secondary vertex (SV), defined by following requirements:
 - probability of the vertex $> 10^{-6}$;
 - no hit on the line between the primary vertex (PV) and the SV;
 - PV-SV distance (projected on the normalized SV momentum) > 2 cm,

then the electron is likely to stem from a photon conversion, so it is eliminated. Electrons are also rejected if any tracker hit is missing in the track, in order to further eliminate electrons originating from photon interactions in the detector;

- a multivariate technique (MVA) based on the combination of several variables providing discrimination between electrons from W or Z/γ^* -boson decays and jets misidentified as electron or conversion electrons is employed. The electron is required to have an MVA discriminator value > 0.9 .
- the minimal distance between the PV and the extrapolated electron track, projected on the transverse plane (transverse impact parameter) < 0.02 cm, applied to the track from the inner tracker. This rejects PU contamination;
- relative isolation (RelIso) < 0.1 . The RelIso indicates how much activity is found in the closest vicinity of the electron. Non-isolated electrons are likely to originate from a hadron decay within a jet, rather than from the W -boson decay. The RelIso calculation is based on isolation contributions identified by the PF technique and effects from neutral PU particles are corrected. This is calculated in the following way:

$$\text{RelIso} = \frac{\text{PFChargedHadIso} + \max(0, \text{PFIso}(\text{Photon} + \text{NeutralHad}) - \rho A_{\text{eff}})}{p_{\text{T}}(e)},$$

where charged (PFChargedHadIso) and neutral hadrons (NeutralHad) as well as photons identified by the PF technique and found within a cone of $R = 0.3$ around the electron direction are considered. The neutral component of PU expected in the cone is corrected by the product of an effective area correction (A_{eff}), determined from data, and the p_{T} density of the event (ρ), calculated with k_t jets with a distance parameter of 0.6. $p_{\text{T}}(e)$ represents the electron p_{T} ;

8.1.2 Selection of Muons

Muon candidates are PF muons. The following identification criteria are applied in order to select high-quality muons from W -boson decays [115]:

- $p_{\text{T}} > 30$ GeV, with the same justification as for electrons;
- $|\eta| < 2.1$, according to the trigger cut, given in Sec. 8.1.4;
- is a GlobalMuon, hence information from the muon system and from the tracking system are compatible and combined;
- at least one hit in the pixel tracker and at least six tracker layers for the track reconstruction. This guarantees a good track quality and makes sure that the muon originates from the collision point;
- at least two matched segments, for the sake of the muon-system measurement quality. A segment corresponds to the track resulting from the signal in a single station of the muon-system;
- $\chi^2/\text{ndof} < 10$ for the global-muon fit. The requirement on χ^2 divided by the number of degrees of freedom, ndof , ensures a sufficient quality of the full muon track;
- transverse impact parameter of the muon < 0.2 cm, applied on the track from the inner tracker. This rejects PU contamination;
- absolute difference in z coordinate between the PV and the muon track vertex < 0.5 cm, which further eliminates PU contamination;

- relative isolation (RelIso) < 0.12 , for the same reason as in Sec. 8.1.1. The RelIso calculation is based on isolation contributions identified by the PF technique and corrected for PU effects:

$$\text{RelIso} = \frac{(\text{PFChargedHadIso} + \max(0, \text{NeutralHadIso} + \text{PhotonIso} - 0.5 * \text{sumPUPt}))}{p_T(\mu)}.$$

Each numerator component is computed from PF isolation deposits in a cone of $R = 0.4$ around the muon direction. `sumPUPt` is the sum of the p_T of the charged particles found in the cone of interest but not originating from the PV. $p_T(\mu)$ represents the muon p_T .

8.1.3 Selection of Jets

Jets are reconstructed using PF candidates (without considering the charged PF particles coming from PU) reconstructed with the anti- k_t algorithm with a size parameter of $R = 0.5$. Input objects to the clustering are filtered by a so-called top-projection [116] in order to eliminate objects used for the electron or muon reconstruction. The jet selection is applied after the jet-energy and jet-energy resolution corrections, described in Sec. 8.3.1. The selection criteria are [117]:

- $p_T > 30$ GeV. Although the PF technique allows to use jets with $p_T > 10$ GeV, as it is stated in [102], a tighter p_T requirement is applied because low-momentum jets are very likely in background events, e.g. from QCD-multijet processes, and have large energy scale uncertainties;
- $|\eta| < 2.5$, corresponding to the tracking-system acceptance;
- loose jet identification, making sure that the jet has enough constituents, charged or neutral, and is not purely electromagnetic or hadronic. A jet made solely of energy deposition of a single calorimeter cell indicates a misidentified jet, due to calorimeter noise. Following selection is required:
 - at least two constituents;
 - a fraction of neutral hadronic energy < 0.99 ;
 - a fraction of neutral electromagnetic energy < 0.99 ;
 - and if the jet has $|\eta| < 2.4$:

- a fraction of charged electromagnetic energy < 0.99 ;
- a fraction of charged hadronic energy > 0 ;
- at least one charged constituent.

8.1.4 Detector-Level Selection

Online Selection

The real time selection of events is achieved with the Level-1 Trigger and the High-Level Trigger (HLT), presented in Sec. 3.2.5. In order to collect electron+jets events, single-electron triggers with loose identification cuts are used (CMS name: HLT_Ele27_WP80_v*). Similarly, single isolated muon triggers (CMS name: HLT_IsoMu24_eta2p1_v*) are used to collect muon+jets events. The trigger selections are applied as well on MC simulation, where the response of the trigger is simulated.

Offline Selection

Events are selected following the requirements agreed upon in the top-quark group of the CMS collaboration [118].

An event cleaning procedure is applied. Beam scrapping events are vetoed by requiring more than 25% of high purity tracks with respect to the total number of tracks when the event has at least 10 tracks. Events with anomalous HCAL noise are also rejected.

The presence of at least one well reconstructed PV is required:

- not tagged as a fake vertex;
- position of the PV on the z axis $|PV_z| < 24$ cm;
- radial position of the PV has to lie within the beam pipe: $|PV_{xy}| < 2$ cm;
- at least four tracks assigned to it by the vertex fit.

Events are kept if they contain either exactly one muon or exactly one electron fulfilling the requirements described in Sec. 8.1.1 and 8.1.2. In order to remove Z + jets and dileptonic $t\bar{t}$ events, any event containing extra muons or electrons passing the following loose-lepton cuts are rejected:

Loose-electron definition:

- $p_T > 20$ GeV and $|\eta| < 2.5$;
- $\text{RelIso} < 0.15$, as defined before;
- MVA discriminator > 0.5 , with the same MVA definition as in Sec. 8.1.1.

Loose-muon definition:

- is a GlobalMuon;
- $p_T > 10$ GeV and $|\eta| < 2.5$;
- $\text{RelIso} < 0.2$, as defined before.

In summary, in order to remove as many background events as possible while keeping a good signal efficiency, the following sequential cuts are applied, defining the signal region:

- event trigger;
- event cleaning;
- PV selection;
- exactly one identified and isolated charged lepton, either an electron (Sec. 8.1.1) or a muon (Sec. 8.1.2). This lepton is called the main lepton;
- no loose muons and no loose electrons should be present in the event in addition to the one accepted in the previous step;
- at least four selected jets (8.1.3) are required. Any jet found within $\Delta R < 0.5$ of the main lepton is discarded;
- at least two of the selected jets must fulfill the medium requirement of the CSV algorithm, Sec. 4.6. This corresponds to a light-jet misidentification probability of 1% and a discriminator cut at 0.679.

8.1.5 Particle-Level Selection

As discussed in Sec. 6.3, in the previous measurement of the jet multiplicity with the CMS detector [35], the definition of the measurement phase space was partly based on parton-level information (e.g. quarks) of MC-simulation events which is provided by PYTHIA or HERWIG. Since this is simulation-model dependent, this is changed for this analysis, namely giving

up completely any parton-level information for the definition of the measurement phase space. The latter is defined by means of a selection applied on particle-level objects for a high compatibility of the results.

As also mentioned in Sec. 6.3, a modern tool for the implementation of analyses to be applied on MC simulation, the Robust Independent Validation of Experiment and Theory (RIVET v1.8.2) [119], has been employed for the definition of the phase-space as well as of the jets at particle level. By this means, the measurement presented in this thesis can bring consistent inputs to the theory community since the RIVET-analysis code to this measurement can be provided to any physicist willing to test his or her model. They can apply the RIVET analysis on their model and compare the results to those achieved by the present measurement.

In the following, the definition of the particle-level objects and their use in the particle-level selection are described. These definitions are chosen to be close to the detector-level analysis, in order to have compatible phase spaces. However, a full agreement is not possible. In the $t\bar{t}$ MC simulation, 81 and 87% of events fulfilling the particle-level selection also fulfill the detector-level selection in the e +jets and the μ +jets channels. The other way around, 75 and 74% of events fulfilling the detector-level selection also fulfill the particle-level selection (statistical uncertainties on these values represent 0.05).

Particle-Level Objects

High-energy leptons, especially electrons, are likely to be accompanied by photons radiated after the lepton creation. This means that the leptons actually recorded in the MC simulation, called here bare leptons, have lost a fraction of their energy. Adding the four-momenta of a bare lepton and of its corresponding neutrino, e.g. when both originate from a W -boson decay, can therefore lead to a wrong four momentum since it does not provide exactly the W -boson mass. This is assumed to be a small error, but needs to be corrected by associating each bare lepton to the photons it has radiated. In this thesis, particle-level leptons are the products of a jet clustering performed with the FastJet [120] implementation of the anti- k_t algorithm with a small size parameter, $R = 0.1$. The inputs of the clustering are all charged leptons and all photons found in the event. They correspond to particles with particle-type number (particle ID, e.g. given by [1]) equal to ± 11 for electrons, ± 13 for muons, ± 15 for taus, and 22 for photons. These particles are taken from the final-state record of RIVET, corresponding to particles exiting the MC

generator². Then, each resulting object, called “dressed lepton” is filtered. Only dressed leptons containing a bare lepton whose p_T is larger than half of the dressed-lepton p_T are kept. The lepton flavor of a dressed lepton is taken from the bare lepton found in it and carrying the highest p_T .

Jets at particle-level are clustered with the FastJet implementation of the anti- k_t algorithm with 0.5 as size parameter. All particles found in the final-state record of RIVET are included in the clustering input, except the particles contained in all identified dressed leptons and all neutrinos. Furthermore, all B hadrons with $p_T > 5$ GeV are also included in the clustering input for b -tagging purpose. These particles need to be taken from the unstable-particle record of the MC simulation, since their lifetime is short so they have already been decayed by the MC generator. This means that these particles should not constitute the event final state, since this is fulfilled by their decay products. Thus, B hadrons are added to the jet-clustering input, but in order to cancel energy double counting, their four momenta are scaled down to have the arbitrary low energy of 10^{-20} GeV. Once the jet clustering has been applied on the input particles, jets found to contain a B hadron are tagged as particle-level b -jets.

Event Phase-Space Definition

The main lepton of the selection is defined as a dressed electron or muon with $p_T > 30$ GeV and $|\eta| < 2.4$. Loose leptons are dressed electrons and muons fulfilling looser kinematic requirements: $p_T > 15$ GeV and $|\eta| < 2.5$. Jets must fulfill the following kinematic requirements: $p_T > 30$ GeV and $|\eta| < 2.5$. Moreover, a jet is eliminated in case it is too close to the main lepton, i.e. if $\Delta R(\text{jet}, \text{main lepton}) < 0.5$.

The event selection applied for the definition of the phase space is defined as follows:

- require exactly one main lepton;
- allow no loose lepton in addition to the previous main lepton;
- require at least four selected jets;
- at least two of the jets counted in the previous step have to be b -tagged.

²These are the particles given as input for instance to the detector simulation.

8.2 Data and MC-Simulation Samples

As presented in Sec. 8.1.4, the analysis selection requires exactly one high-energy lepton and at least four high-energy jets, where at least two of them have to be identified as b -jets. This event selection achieves a very good signal purity. The remaining background contributions arise from single-top, W/Z +jets, QCD multijet, and $WW/WZ/ZZ$ production. These processes as well as the signal are modeled using MC simulation. The corresponding samples are listed in Table 8.1. Their production was done in summer 2012 with the CMS Software version 5.3.12: CMSSW_5_3_12, including a full simulation of the CMS detector. This list with the full CMS-names and technical information can be found in Table A.1 of Appendix A

The $t\bar{t}$ MC-simulation sample originates from MADGRAPH 5 interfaced via MLM matching [67] with PYTHIA 6. The top-quark mass is set to $m_t = 172.5$ GeV and the PDF set is taken from CTEQ6L1 [50]. For the calculation of the theoretical systematic uncertainties, dedicated MC-simulation samples, shown in Table 8.2, are used, the list with the full CMS-names and technical information can be found in Table A.2 of Appendix A. The renormalization and factorization scale in MADGRAPH is the default definition for the simulation of pairs of heavy particles: $Q^2 = M(\text{top})^2 + \Sigma p_T^2(\text{jet})$. It is varied by factors 4 and 0.25 in the $t\bar{t}$ MC simulation in order to estimate the Q^2 -scale uncertainty. The jet- p_T threshold for the MLM matching between matrix-element jets from MADGRAPH and parton-showering jets from PYTHIA is varied between an upper and a lower value around the nominal one by a factor 2 and 0.5 in $t\bar{t}$ events. The nominal value is set to 20 GeV. The $t\bar{t}$ total cross section given in Table 8.1 was calculated within the TOPLHCWG as described in [122] with top++ v2.0 [24], assuming a top-quark mass $m_t = 172.5$ GeV.

The simulation of W +jets processes is the sum of the MC simulation of $W+2$, $W+3$, and $W+4$ -jets processes after their normalization according to their respective cross sections. This sum is referred from this point on as the MC simulation of W +jets events.

CMS data used for the analysis have been taken from the single muon and single electron data paths, whose technical path names are listed in Table A.3 and A.4 of Appendix A. The data taking has been achieved in December 2012 but a reprocessing took place in January 2013 in order to take into account the latest improvements of the data reconstruction. This corresponds to the full 2012 data set delivered by the LHC at 8 TeV, corresponding to a total integrated luminosity of 19.7 fb^{-1} with 2.6% uncertainty.

Table 8.1: Summary of MC-simulation samples used for the present measurement. All events contain PU simulation assuming a peak instantaneous luminosity of $7.4 \cdot 10^{33}$ of the LHC [121]. MC-simulation samples generated with MADGRAPH and PYTHIA have LO precision whereas single-top is simulated at NLO precision thanks to POWHEG. The generator + parton-showering/hadronization tools are mentioned, the cross section (XS) assumed are given, as well as the publications (or tools) they originate from (Ref.). Symbols: $\ell = e$ or μ , $m_{\ell\ell}$ is the dileptonic invariant mass, \hat{p}_T is the p_T of the hard scattering. In the QCD-multijet samples, requirements are applied in order to focus on those events likely to fulfill the signal selection. For the μ +jets channel, a muon with a minimal p_T is required: “with μ of $p_T > 15$ GeV”. In the e +jets channel, an heavy-flavor-quark decay into electron should be found: “ $b/c \rightarrow e$ ” and jets with a dominant energy contribution from electrons and photons should be found: “ e/γ enriched”.

Physical Process	Generator+PS/Hadro.	XS [pb]	Ref.
$t\bar{t} + 0/1/2/3$ jets (LO)	MADGRAPH+PYTHIA	252.89	[122]
t, t channel	POWHEG+PYTHIA	56.4	[123]
\bar{t}, t channel	POWHEG+PYTHIA	30.7	[123]
t, tW channel	POWHEG+PYTHIA	11.1	[123]
\bar{t}, tW channel	POWHEG+PYTHIA	11.1	[123]
t, s channel	POWHEG+PYTHIA	3.79	[123]
\bar{t}, s channel	POWHEG+PYTHIA	1.76	[123]
$W \rightarrow \ell\nu_\ell + 2$ jets	MADGRAPH+PYTHIA	2159.20	[124]
$W \rightarrow \ell\nu_\ell + 3$ jets	MADGRAPH+PYTHIA	640.40	[124]
$W \rightarrow \ell\nu_\ell + 4$ jets	MADGRAPH+PYTHIA	264.00	[124]
$Z/\gamma^* \rightarrow \ell\ell +$ jets ($m_{\ell\ell} > 50$ GeV)	MADGRAPH+PYTHIA	3503.71	[124]
WW	PYTHIA	54.83	[125]
WZ	PYTHIA	33.21	[125]
ZZ	PYTHIA	8.05	[125]
QCD, with μ of $p_T > 15$ GeV	PYTHIA	134680.0	Pythia6
QCD, $b/c \rightarrow e$ ($20 < \hat{p}_T < 30$ GeV)	PYTHIA	167388.0	Pythia6
QCD, $b/c \rightarrow e$ ($30 < \hat{p}_T < 80$ GeV)	PYTHIA	167040.0	Pythia6
QCD, $b/c \rightarrow e$ ($80 < \hat{p}_T < 170$ GeV)	PYTHIA	12981.9	Pythia6
QCD, $b/c \rightarrow e$ ($170 < \hat{p}_T < 250$ GeV)	PYTHIA	620	Pythia6
QCD, $b/c \rightarrow e$ ($250 < \hat{p}_T < 350$ GeV)	PYTHIA	103.2	Pythia6
QCD, $b/c \rightarrow e$ ($\hat{p}_T > 350$ GeV)	PYTHIA	23.49	Pythia6
QCD, e/γ enriched ($20 < \hat{p}_T < 30$ GeV)	PYTHIA	29148.6	Pythia6
QCD, e/γ enriched ($30 < \hat{p}_T < 80$ GeV)	PYTHIA	4615893.0	Pythia6
QCD, e/γ enriched ($80 < \hat{p}_T < 170$ GeV)	PYTHIA	183294.9	Pythia6
QCD, e/γ enriched ($170 < \hat{p}_T < 250$ GeV)	PYTHIA	4650	Pythia6
QCD, e/γ enriched ($250 < \hat{p}_T < 350$ GeV)	PYTHIA	559	Pythia6
QCD, e/γ enriched ($\hat{p}_T < 350$ GeV)	PYTHIA	89.1	Pythia6

These data samples are cleaned by requiring that luminosity sections³ were validated by the data quality and validation groups of CMS [126]. This means among others that all parts of the detector are fully in service when data are being recorded.

³A luminosity section is a time interval of a run in which the luminosity is approximately constant (1-2 minutes).

Table 8.2: MC-simulation samples used for the estimation of the model systematic uncertainties. The generator + parton-showering/hadronization tools are mentioned. The same cross section as for the standard $t\bar{t}$ MC simulation sample, given in Table 8.1, is valid.

Physical Process	Parameters	Generator+PS/Hadro.
$t\bar{t} + 0/1/2/3$ jets (LO)	Q^2 scale up	MADGRAPH+PYTHIA
	Q^2 scale down	MADGRAPH+PYTHIA
	jet matching up	MADGRAPH+PYTHIA
	jet matching down	MADGRAPH+PYTHIA
$t\bar{t}$ (NLO)	Nominal	POWHEG+PYTHIA
		POWHEG+HERWIG
		MC@NLO+HERWIG

8.3 Corrections of Data and MC Simulation

Several corrections are applied to the MC simulation in order to correctly describe the data. These corrections have uncertainties which are taken into account as systematic uncertainties. The figures shown in this section make use of the modeling and the normalization of the QCD-multijet and W +jets contributions presented in Chapter 9.

8.3.1 Jet Energy

Corrections have to be applied on the jet energy in MC simulation and data because the jet energy measured by the detector (or its simulation) does not match well the corresponding particle-level energy, i.e. the total energy of the particles produced by the parton showering and the hadronization of quarks or gluons, for instance from an additional gluon radiated in a $t\bar{t}$ event. The jet-energy corrections aim at an agreement between the detector-level and particle-level values of the jet energy and energy resolution. These corrections need to be applied before the jet selection. A full description can be found in [102].

Jet-Energy Scale Jet-energy scale corrections are estimated by means of data-driven methods: making use of the jet-energy balance in di-jet or Z/γ^* +jets events, or exploiting the absence of \cancel{E}_T in Z/γ^* +jets events [102]. The following corrections are applied iteratively on jets for this analysis, according to recommendations [127] from the jet-energy resolution and correction subgroup of the CMS collaboration [128]:

- **Pile-Up Correction** A first step removes energy depositions due to pile-up activity (applied on jets from data and MC simulation).
- **Relative Correction** A correction of the jet is applied in order to achieve a flat response of the jet energy as a function of $|\eta|$ (applied on jets from data and MC simulation).
- **Absolute Correction** An offset correction is applied in order to bring the energy scale to the particle level (applied on jets from data and MC simulation).
- **Residual Correction** In the MC simulation, the jet energy is in good agreement between detector level and particle level after the first three steps have been applied. However, some limited discrepancies remain between data and MC simulation. They are accounted for in this last step, which is applied only on jets from data.

The actual correction values are taken from [129], they represent 5 to 20% of the original jet energy, respectively for central ($\eta = 0$) and forward jets ($|\eta| = 2.0$) for a jet- p_T of 50 GeV. It is worth to remark here that the uncertainties on these corrections are the dominant source of systematic uncertainties of the measurement presented in this thesis.

Jet-Energy Resolution Once the correction of the jet-energy scale is applied, it is observed that the jet-energy resolution in data is worse than in MC simulation, as shown previously in Sec. 4.6. Therefore, for each reconstructed jet in MC simulation, the difference between its p_T (p_T^{reco}) and the p_T of its associated particle-level jet (p_T^{gen}) is corrected with a factor c , depending on $|\eta|$ of the jet, using the following formula:

$$p_T^{\text{reco}} \rightarrow p_T^{\text{gen}} + c(p_T^{\text{reco}} - p_T^{\text{gen}})$$

where c takes the values given in Table 8.3. More explanations on the jet-energy resolution corrections can be found in [102].

8.3.2 Pile Up

The MC simulation is completed by mixing pile-up events to the originally simulated events, as presented in Sec. 3.3. The instantaneous luminosity of the LHC, as a main parameter for the multiplicity of pile-up events, is constantly changing during data taking and cannot be predicted at the earlier time of the MC production. Therefore, the MC simulation is mixed with

Table 8.3: Values of the factor applied in order to fit the jet-energy resolution of MC simulation to the resolution observed in data [130].

$ \eta $ of the corrected jet	c value
$ \eta < 0.5$	1.079
$0.5 < \eta < 1.1$	1.099
$1.1 < \eta < 1.7$	1.121
$1.7 < \eta < 2.3$	1.208
$2.3 < \eta < 2.8$	1.254
$2.8 < \eta < 3.2$	1.395
$3.2 < \eta < 5.0$	1.056

pile-up events from a realistic luminosity scenario and once data has been delivered by LHC, the MC simulation must be corrected to mimic the actual LHC conditions.

The number of pile-up interactions per bunch crossing in data is estimated using the instantaneous luminosity provided by the pixel detector of CMS, for each luminosity section. The instantaneous luminosity is multiplied by the total inelastic proton-proton cross section of $69.4 \text{ mb} \pm 5\%$ (recommended value for 2012, see [131]), hence providing the number of pile-up interactions. The distribution of the number of pile-up interactions in data is compared to its distribution in MC simulation and the differences are canceled by a reweighting of the MC simulation [132]. Pile up in the MC simulation has been produced with PYTHIA assuming a scenario with a peak instantaneous luminosity of $7.4 \cdot 10^{33}$ of the LHC [121].

Fig. 8.3 shows the very good agreement between MC simulation and data for the distribution of the number of reconstructed PVs after the pile-up correction has been applied.

8.3.3 Top- p_T Spectrum

Differential $t\bar{t}$ cross-section measurements as a function of the p_T of reconstructed top quarks (top- p_T) at 7 and 8 TeV, [133] and [134], have observed that the distribution of this variable with for instance MADGRAPH +PYTHIA is harder than in data. However, higher precision calculations from [135] show an improved description. Relying on this, a correction is applied in order to improve the simulation of the shape of the top-quark p_T by modifying the weight of $t\bar{t}$ events in MC simulation, as documented in [136]. As an illustration of the correction, the p_T spectrum of selected jets is shown on fig. 8.4. The shape agreement is improved, as can be mainly seen in the ratio plot. On the other hand, an overall reduction of the event yield in the

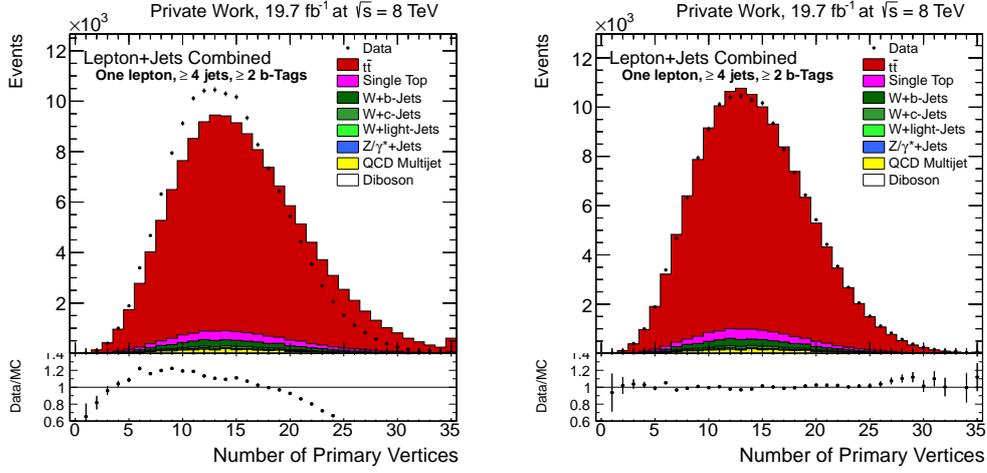


Figure 8.3: Distribution of the number of reconstructed PV before (left) and after (right) the pile-up correction in the combined ℓ +jets channel.

$t\bar{t}$ MC simulation is provoked by this correction (reduction of 3% of the sum of event weights in the $t\bar{t}$ MC simulation).

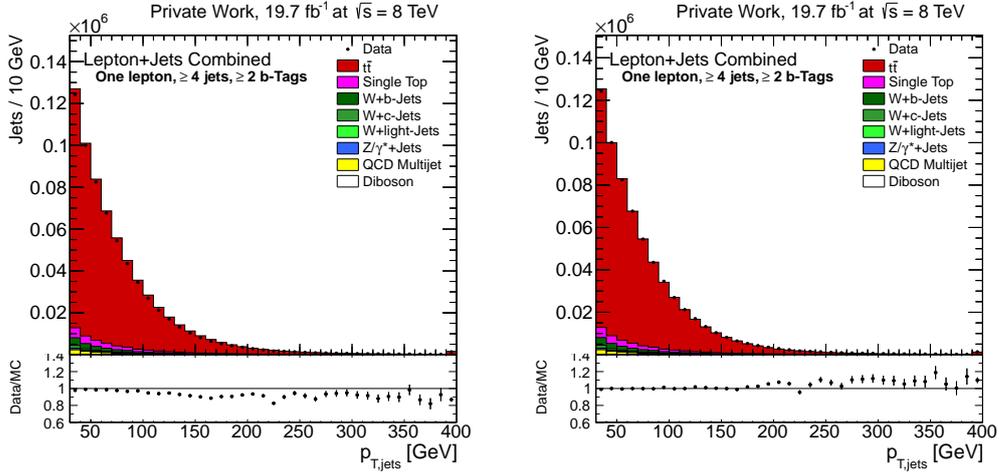


Figure 8.4: Distribution of jet p_T in the combined ℓ +jets channel before (left) and after (right) the top- p_T correction has been applied.

8.3.4 Lepton Reconstruction, Selection, and Trigger Efficiencies

The efficiency of the reconstruction, the selection, and the triggering of electrons and muons is not the same in data and MC simulation. In order to correct the latter, these efficiencies have to be estimated precisely in data [137, 138]. This is performed in the CMS collaboration making use of tag&probe techniques. In the present case, Z -boson events from data are used where the Z boson decays into a pair of electrons or muons. A tag lepton is selected with very tight requirements. If another lepton compatible with the tag lepton (same flavor and opposite sign) is found in the event, this second lepton can be a probe lepton. The dileptonic invariant mass of the tag-probe lepton pair is required to be close to the Z boson mass to reduce background contamination. Then a test of the probe lepton is performed, requiring this lepton to fulfill further cuts (kinematic cut, isolation, trigger). This provides an efficiency $\varepsilon = \frac{N_{\text{pass}}}{N_{\text{all}}}$ as a function of the cut. Typical values are above 80% for the triggering, reconstruction, and selection of electrons or muons with $p_{\text{T}} > 30$ GeV. This is measured also in MC and the comparison between the efficiencies provides scale factors $\text{SF} = \frac{\varepsilon^{\text{data}}}{\varepsilon^{\text{MC}}}$. The resulting efficiency modifications ($|1 - \text{SF}|$) for electrons (muons) are below 5% (3%) in the central region and for $p_{\text{T}} > 50$ GeV. They increase up to $\approx 10\%$ in the end-cap for $30 < p_{\text{T}} < 40$ GeV. They are utilized for the correction of the lepton efficiencies in MC simulation, depending on the p_{T} and η of the corresponding lepton. Recipes and SF values are CMS-internal, taken from [139] and [140].

8.3.5 Modeling of the b -Tagging Efficiency in MC

As the MC simulation does not reproduce exactly the b -tagging performance observed in data, SFs have to be applied. Their values for the tagging of b -jets are given by Eq. 8.1 with uncertainties between 2% and 7%.

$$\text{SF}_b = 0.939158 + 0.000158694 \cdot p_{\text{T}} - 2.53962 \cdot 10^{-7} \cdot p_{\text{T}}^2. \quad (8.1)$$

These recommendations were estimated by the b -tagging group of the CMS collaboration [141] using jets containing soft muons. They apply only for jets with p_{T} between 20 and 800 GeV. For jets beyond these limits the boundary SF with twice the boundary uncertainty are taken. The same values are used for c -jets but with twice the uncertainties. The SF as well as

the corresponding uncertainties for light-jets are computed in different p_T and η bins, following the recipe and values given in [141, 142]. Fig. 8.5 shows the SF and their uncertainties for b -tagging efficiency and misidentification.

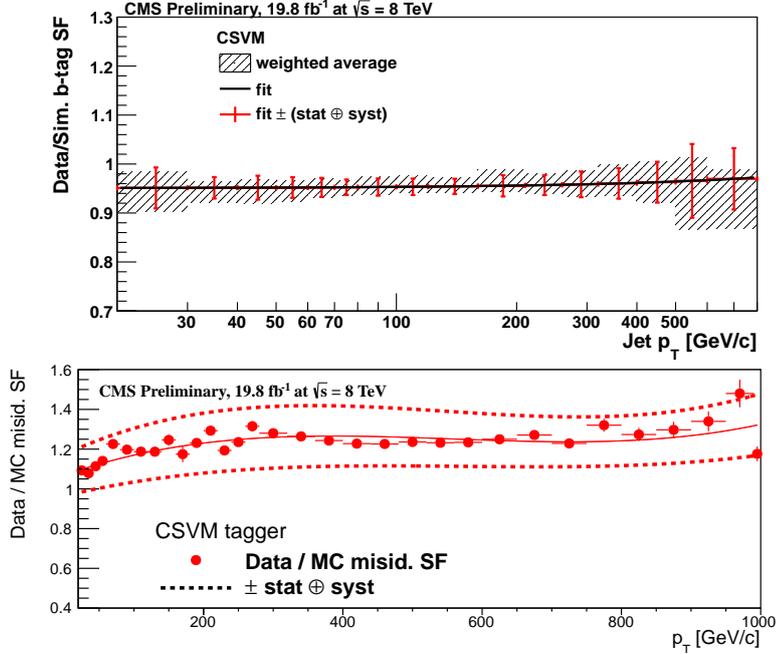


Figure 8.5: SF of the b -tagging efficiency (top) and of the misidentification probability (bottom) as a function of the jet p_T , measured for the CSV tagger at the medium working point. Taken from [143].

The probability for a jet with flavor q ($q = b, c$, or light, where light = gluons or u, d , or s quarks), to be tagged as originating from a bottom quark in data or MC can be described as in Eq. 8.2.

$$\begin{aligned} P_q^{\text{MC}} &= \text{Eff}_q^{\text{MC}}, \\ P_q^{\text{data}} &= \text{SF}_q \cdot \text{Eff}_q^{\text{MC}}, \end{aligned} \quad (8.2)$$

where SF_q is the tagging-efficiency scale factor and Eff_q^{MC} is the tagging efficiency for a jet of flavor q in the MC simulation. The efficiencies Eff_q^{MC} are estimated with MC events fulfilling the event selection but before the b -tagging requirement. The probability to tag zero jets in an event is given by Eq. 8.3.

$$P(0 \text{ } b\text{-tag}) = \prod_i (1 - P_{q(i)}), \quad (8.3)$$

where the product \prod_i is running on all the selected jets i of the event, each jet carrying $q(i)$, its flavor. In order to be able to plot the distribution of the b -tag jet multiplicity, each exclusive probability, the two first are shown in Eq. 8.4, are required.

$$P(1 \text{ } b\text{-tag}) = \sum_i [P_{q(i)} \cdot \prod_j (1 - P_{q(j)})] \text{ with } j \neq i, \quad (8.4)$$

$$P(2 \text{ } b\text{-tags}) = \sum_i \sum_j [P_{q(i)} \cdot P_{q(j)} \cdot \prod_k (1 - P_{q(k)})] \text{ with } j > i \text{ and } k \neq i, j.$$

Knowing this, the probability to tag at least two jets in the event can easily be computed as in Eq. 8.5.

$$P(\geq 2 \text{ } b\text{-tags}) = 1 - P(0 \text{ } b\text{-tag}) - P(1 \text{ } b\text{-tag}). \quad (8.5)$$

As a consequence, for every event passing the 2- b -tag selection, a weight w , given in Eq. 8.6, is calculated to account for the difference of probabilities for such an event in MC and in data.

$$w = \frac{P^{\text{data}}(\geq 2 \text{ } b\text{-tags})}{P^{\text{MC}}(\geq 2 \text{ } b\text{-tags})} \quad (8.6)$$

Chapter 9

Data-Based Background Estimation

While the modeling of several background processes; Z +jets, single top, and diboson, fully relies on the MC simulation and the dedicated cross-section calculations, the modeling of QCD-multijet and W +jets background processes is achieved in this analysis with a special treatment.

Due to the huge cross section for QCD multijet-events and at the same time the very small rate of these events fulfilling the full event selection, the use of MC simulation would require to generate an enormous number of events. With too small MC-simulation samples, too few events are available for the analysis, which leads to large statistical fluctuations which propagate to the result uncertainties. Therefore, a QCD-multijet model is extracted from data and normalized through a template fit. Furthermore, in the MC simulation of W +jets events the heavy-flavor fraction, defined as the rate of W +jets events containing bottom or charm quarks with respect to those events containing only light quarks and gluons, has been observed not to be describing data perfectly. See e.g. [144]. As a consequence the normalizations of the separate W +jets flavor parts is estimated through the same template fit as well. A template fit has already been performed for the previous publication [35], summarized in Chapter 6. Here an improved procedure is used where both background estimations (for the QCD-multijet and the W +jets events) are achieved simultaneously together with the heavy-flavor fraction in W +jets events for an estimate which is dedicated and adapted to the present measurement.

As the background estimation contains a fit to data, it should not be performed with the same data and MC-simulation events as those used for the

measurement, in order to avoid bias on the results. Thus, selected events showing either two or three jets are utilized for the background estimation, which represent an independent event region. This is further referred as the sideband region. Furthermore, data and MC events showing exactly one b -tagged jet (called 1-tag events) and at least two b -tagged jets (called ≥ 2 -tag events) have been used for this estimation.

9.1 Extraction of the QCD-Multijet Model from Data

QCD-multijet events are able to fulfill the event selection if they contain an isolated lepton. This is either possible when a real lepton has been produced, typically by a hadron decay within a jet or when a jet has been misidentified as a lepton. Two different techniques for the extraction of the QCD-multijet model from data are presented in Sec. 9.1.1 and 9.1.2. Their general idea consists in loosening certain lepton-selection criteria. This provides models with similar properties to QCD-multijet events from data fulfilling the full event selection.

9.1.1 Lepton Within a Jet

In the case in which QCD-multijet events fulfill the event selection with a real lepton, the latter should originate from a decay of a particle within a jet. The lepton can fulfill the selection if it carries a large fraction of the jet energy, or if it is accidentally measured without much jet activity in its vicinity. In both cases it is measured as isolated (see definition of the isolation in Sec. 8.1). The QCD-multijet events in which the real lepton is located within the jet and is hence non-isolated are more frequent. They are assumed to be similar to those QCD-multijet events containing an isolated lepton, which contaminate the selected data events. Here, the required similarity is limited to an agreement between shapes of variables relevant for the present analysis. By selecting data events with the standard selection steps but with the exception of the high-energy lepton which has to be non-isolated, a QCD-multijet model is extracted. The cut on the RelIso variable is set to be

$$0.3 < \text{RelIso} < 1.0.$$

The lower bound is set to 0.3, relatively distant from the “isolated”-selection value ($< 0.1/0.12$ for e/μ), in order to limit the contamination from non-QCD-

multijet processes which are likely to show isolated leptons. The higher bound is set at 1.0 because too large values of the R_{ellso} are correlated to small lepton- p_{T} values, according to the R_{ellso} definition. This can lead to strong bias of the lepton- p_{T} spectrum in the QCD-multijet model.

This modeling of the QCD-multijet events is called “antiRel”. It is extracted both in the signal region (at least four jets) and in the sideband region (two or three jets).

9.1.2 Fake Lepton

In the case of a jet looking like a lepton, a “fake lepton”, the lepton-quality cuts (“ID cuts”) failed to discard it. Therefore, a way to extract from data QCD-multijet events showing “fake leptons” is to require events to contain a lepton for which at least one of the lepton-ID cuts is not fulfilled. All lepton-selection cuts except p_{T} , η , R_{ellso} , and $\Delta R(\text{lepton-jet})$ are concerned.

This alternative QCD-multijet model is called “antiID”. Similarly to the antiRel model, it is extracted in the signal region (≥ 4 jets) and in the sideband region (2 or 3 jets)

9.1.3 Choice of the QCD-Multijet Model

In the e +jets channel the antiID model is chosen because it is observed that it provides a better description of leptons at high η than the antiRel modeling. In contrary, in the μ +jets channel, the antiID modeling provides too few events and shows a larger contamination from other processes. Therefore, the antiRel modeling is chosen in this channel, showing a satisfactory description of the high- η leptons. These different behaviors with respect to the QCD-multijet extraction techniques are due to the different trigger selections applied and the different source of QCD-multijet events in the e +jets and the μ +jets channels.

9.1.4 Validation of the QCD-Multijet Model

Fig. 9.1 and Fig. 9.2 show the distribution of data and MC events selected for the QCD-multijet extraction in the sideband and the signal region, i.e. loosening the ID cuts or inverting the R_{ellso} cut, respectively, in the e +jets and μ +jets channel. The lepton- p_{T} (R_{ellso}) variable is shown in the e +jets (μ +jets) channel for the MC simulation and data, whereas data correspond

to the QCD-multijet model used in the analysis after subtraction of contaminations. In the sideband region, the QCD-multijet model in both lepton channels is, according to the simulation, very pure. More contaminations are observed in the signal region, but they are not dominating the QCD-multijet model. There is a good shape agreement between the QCD-multijet model from data and the description from the MC simulation, within its statistical uncertainties. The MC simulation QCD-multijet events shows as expected large statistical fluctuations, this motivates the extraction of QCD-multijet events from data.

One more check of the QCD-multijet model is given in Fig. 9.3 showing the jet-multiplicity distribution in the MC simulation and in the data model. The reference distribution is assumed here to be the MC distribution where the standard lepton selection (called “MC QCD” in the legend) is applied. Within the fairly large statistical uncertainties, good agreement between this reference distribution and the QCD-multijet model, is observed.

9.2 Categorization of W +Jets Events

The MC simulation of the W +jets process is split into two parts for the background estimation. All events containing at least one selected jet reconstructed in the vicinity ($\Delta R < 0.5$) of a bottom or a charm quark are categorized as heavy flavor (W +HF) and all remaining events are categorized as light flavor (W +light). The ratio of W + b -jets to W + c -jets events is taken from the simulation. The effect of this assumption on the result is included as a systematic uncertainty, where the b/c ratio is varied by factors of 2 and 0.5 before repeating the whole analysis.

9.3 Simultaneous Template Fit in Sideband

The variable chosen for the template fit is the missing transverse energy (\cancel{E}_T) because it provides good discrimination between QCD-multijet events, where no \cancel{E}_T is expected, and $t\bar{t}$, single top, and W +jets events, where high-energy neutrinos are produced. Z +jets events do not contain neutrinos from the hard process, and therefore a low discrimination between the Z +jets and QCD-multijet templates is expected. The \cancel{E}_T does not provide discrimination between W +jets categories (W +HF and W +light). The sensitivity of the fit to the heavy-flavor content of the templates is provided by a categorization

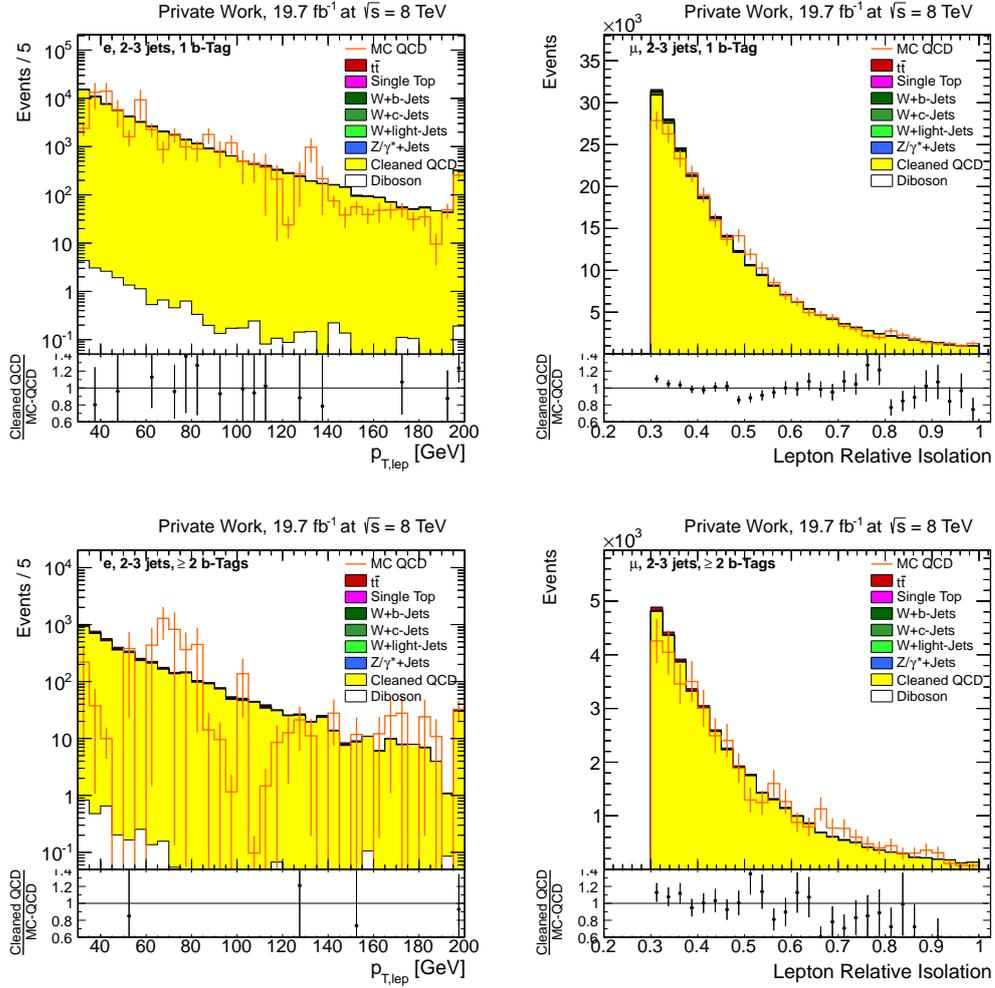


Figure 9.1: Distribution of the lepton- p_T (RelIso) variable in the e +jets (μ +jets) channel when applying the antiID (antiRel) selection on MC simulation and on data. The QCD-multijet model used further in this analysis is the yellow histogram, after the subtraction of all non-QCD-multijet simulation but without the fit normalization. The latter are represented by the remaining colored histograms. The MC simulation for QCD-multijet production is represented scaled to the QCD-multijet model for a shape comparison. The distributions are shown for the e +jets (left) and μ +jets (right) channel, and for 1-tag events (top) and ≥ 2 -tag events (bottom) in the sideband region (containing two or three jets).

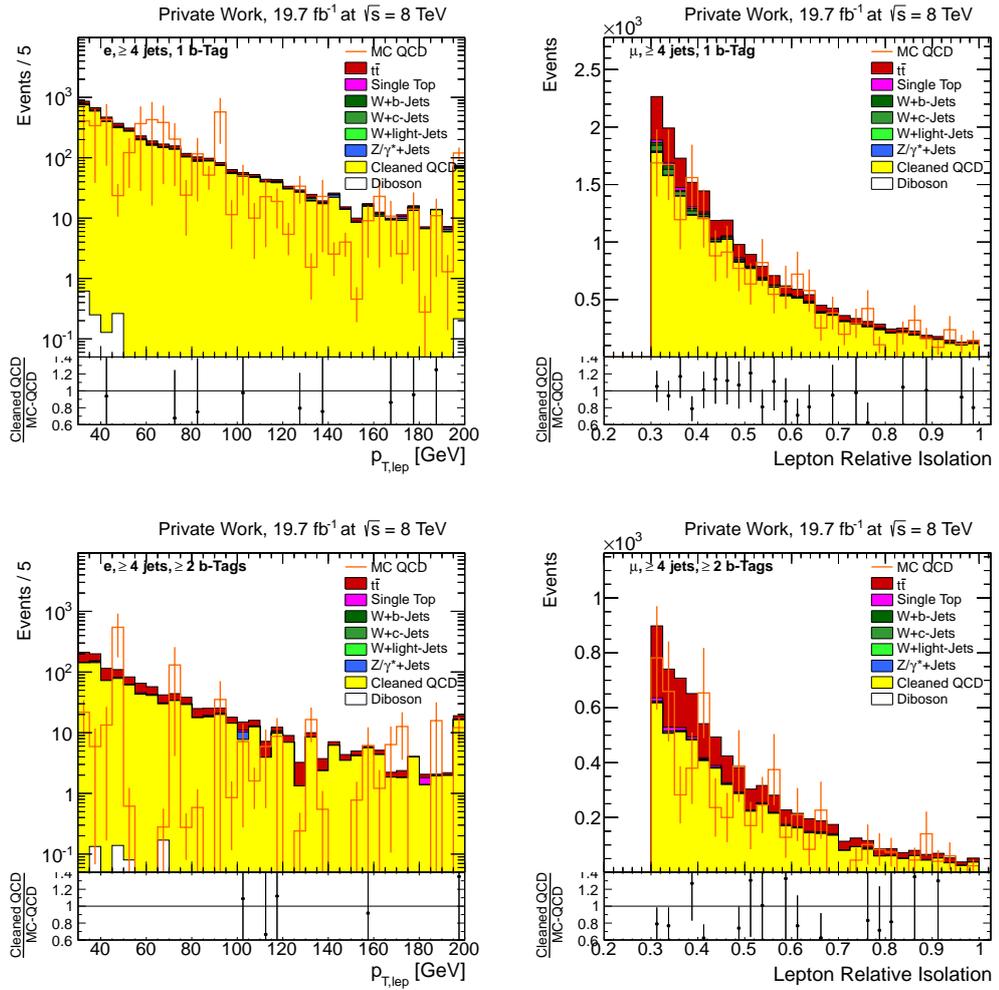


Figure 9.2: Distribution of the lepton- p_T (RelIso) variable in the e +jets (μ +jets) channel when applying the antiID (antiRel) selection on MC simulation and on data as in Fig. 9.1 but in the signal region (containing at least four jets).

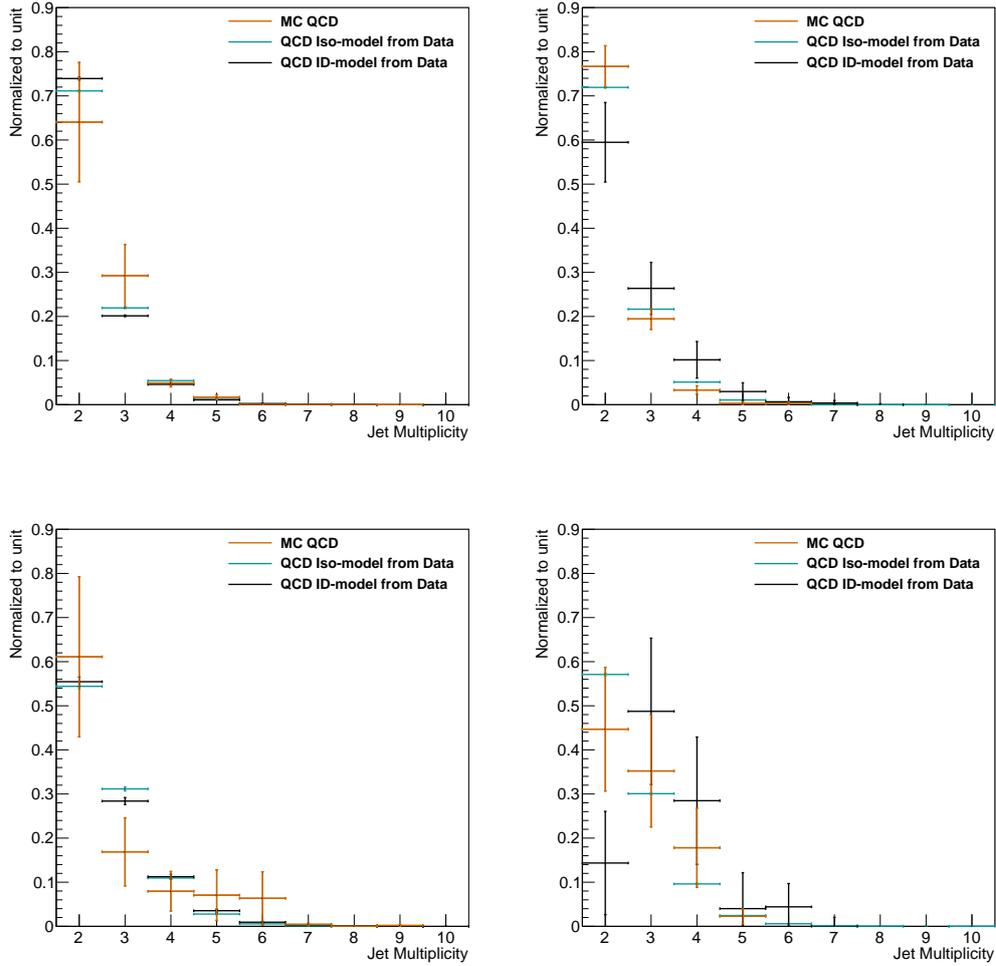


Figure 9.3: Jet multiplicity found in QCD-multijet events described by MC and by the data model in the signal region, normalized to unity. Events in the e +jets (μ +jets) channel are shown on the left (right) and events with only one b -tag (at least two b -tags) are shown on the top (bottom).

of the events depending on the number of b -tagged jets they contain. This categorization is described in the following.

The \cancel{E}_T distributions in the QCD-multijet model and the MC simulation of remaining processes ($t\bar{t}$, single top, both W +jets parts, Z +jets, and diboson) are fitted to the \cancel{E}_T distribution in data within the sideband region. The fitting method is a negative log-likelihood, as presented in Sec. 5.1, with constraints described in the following. The sideband region is split beforehand into four categories in order to increase the discrimination power (1-tag and ≥ 2 -tag events and for the e +jets and μ +jets channel).

For the QCD-multijet template, each of the four parts constitute an independent template. A first reason is that QCD-multijet models are not expected to require the same normalization by the fit in the e +jets and μ +jets channels, since their extraction method from data is different. This is not the case for all other processes for which the MC simulation is consistent for both channels. A second reason is that a correlation is expected between the number of b -tagged jets and non-isolated leptons in the antiRel model, since a large fraction of non-isolated leptons selected by this method originate from heavy-hadron decays within jets. This causes the antiRel extraction from data to overestimate significantly the yield of QCD-multijet model in the ≥ 2 -tag region with respect to the 1-tag region. The fit is able to correct this effect if QCD-multijet models in 1-tag and ≥ 2 -tag parts are independent templates. For all remaining processes, the templates of the four parts are linked, so only one template per process is fitted. An illustration of the fit configuration is given in Fig. 9.4 by the fit result. The discrimination between both W +jets categories is provided by the fact that the fit is performed simultaneously in separate regions (1-tag and ≥ 2 -tag regions) since W +HF events show more b -tagged jets than W +light events.

The Z +jets template is constrained to its predicted normalization with a log-normal constraint of 3%, according to the uncertainty on NNLO cross-section calculation from [145] because it shows a similar \cancel{E}_T distribution as the QCD-multijet template.

A low discrimination between $t\bar{t}$ and single top templates was experienced; therefore, they have been merged into one single template (called “Top”) for this fit. Finally, the diboson template is too small and has to be constrained with a 5% log-normal constraint, according to the uncertainties from NLO cross-section calculations [146]. No more constraints are applied.

9.4 Results and Discussion

The results of the fit are illustrated in Fig. 9.4. A good agreement between the scaled templates and data is observed, except in the tails where fluctuation-like discrepancies up to 10% are visible in the ratio plot. The numerical results for the fit factors with the corresponding estimated number of events in the MC simulations and the QCD-multijet model are given in Table 9.1. The fit factor of the “Top” template containing the signal events is 0.996 ± 0.004 , which is consistent with one. This means, that the fit results are consistent with the signal predictions. As a consequence, no bias on the later measurement is expected from the present fit procedure. The fit increases the scale of $W+HF$ events by a factor of 1.83 ± 0.05 and decreases the scale of $W+light$ by a factor of 0.3 ± 0.1 , which is comparable with the correction applied in [35]. The uncertainties given for both latter factors are statistical only. As expected the fit corrects the bias previously mentioned, due to the correlation between the number of b -tagged jets and non-isolated leptons: in the μ +jets channel, the fit factor of the QCD-multijet model in the ≥ 2 -tag region is much smaller ($\approx 50\%$) than in the 1-tag region. This is not the case in the e +jets channel, where such a correlation is not expected. The QCD-multijet fit factors in the 1-tag region, named α and β in the table for clarity, are non-relevant scales reflecting the extrapolation from the antiID (antiRel) electron (muon) phase space to the signal and sideband phase space. They are not expected to be consistent with 1.0 or with each other. The measured values are approximately $\alpha = 1.79$ and $\beta = 0.148$.

It is important to stress here that only the fit results for the QCD-multijet models (e +jets and μ +jets channels) in the ≥ 2 -tag region, for the $W+HF$, and for the $W+light$ are further utilized in the measurement presented in this thesis. The extrapolation of the fit results from the sideband region to the measurement region leads to uncertainties which are discussed in Sec. 11.3.

Shape uncertainties are considered for the probed templates. Their effects are estimated by changing the templates accordingly and performing the fit again. For the $W+HF$ template, the ratio of $W+b$ -jets and $W+c$ -jets is varied by factors 2 and 0.5 and the results for W +jets and QCD-multijet events are propagated to the results, as described in Sec. 11.3. For the QCD-multijet model in the e +jets channel, the shape is varied by employing the QCD-multijet model obtained from data with the antiRel technique. In the μ +jets channel, the shape uncertainty is estimated by shifting the RelIso-cut window for the extraction of the QCD-multijet model from data, the shifted window

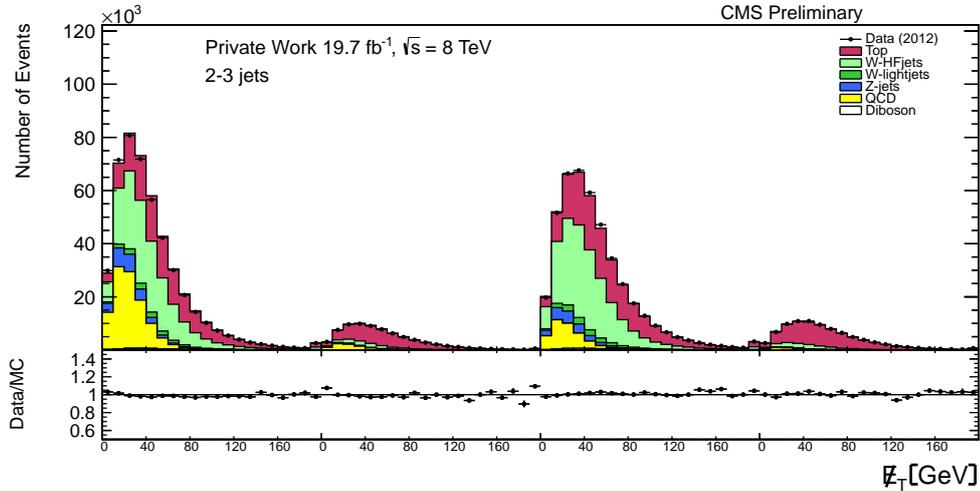


Figure 9.4: Distribution of \cancel{E}_T in the four fit regions (from left to right, each starting on the x-axis at zero: e +jets – 1 tag ; e +jets – ≥ 2 tags ; μ +jets – 1 tag ; μ +jets – ≥ 2 tags). All templates are scaled according to the fit results.

Table 9.1: Results of the template fit providing the normalization of the QCD-multijet model and the W +HF and W +light MC simulation for the measurement presented in this thesis. The value of the QCD-multijet factors α and β , respectively of the e +jets and μ +jets channels, are only written in the text because they are not relevant for the sanity of the fit in this table.

Template	Fit factor \pm unc. (const.)	Res. event yield
Top	$0.996 \pm 0.4\%$ (-)	$4.17 \cdot 10^5$
QCD-multijet (e +jets - 1 tag)	$\alpha \pm 0.9\%$ (-)	$1.09 \cdot 10^5$
QCD-multijet (e +jets - ≥ 2 tags)	$(\alpha \cdot 1.05) \pm 3\%$ (-)	$8.2 \cdot 10^3$
QCD-multijet (μ +jets - 1 tag)	$\beta \pm 2\%$ (-)	$3.61 \cdot 10^4$
QCD-multijet (μ +jets - ≥ 2 tags)	$(\beta \cdot 0.50) \pm 8\%$ (-)	$2.8 \cdot 10^3$
W +HF	$1.83 \pm 3\%$ (-)	$3.9 \cdot 10^5$
W +light	$0.3 \pm 40\%$ (-)	$2.8 \cdot 10^4$
Z +jets	$1.07 \pm 3\%$ (3%)	$4.8 \cdot 10^4$
Diboson	$1.00 \pm 5\%$ (5%)	$8.9 \cdot 10^3$

ranges are:

$$0.2 < \text{RelIso} < 0.7 \quad \text{and} \quad 0.5 < \text{RelIso} < 2.0.$$

In the case of the QCD-multijet shape uncertainty, switching from the nominal QCD-multijet model to the alternative model in the e +jets channel affects also significantly the resulting fit factor of the QCD-multijet template in the μ +jets, et vice versa. This should not be neglected, hence the variation observed for the QCD-multijet e +jets template, when evaluating the QCD-multijet shape uncertainty in the μ +jets channel, is added in quadrature as

additional cross-section uncertainty to the original uncertainties from the fit. The same is done when evaluating the QCD-multijet shape uncertainty in the e +jets channel, the other way around.

The resulting normalization uncertainties for the QCD-multijet estimation are given in Table 9.2, together with the normalization uncertainties for both W +jets templates resulting from the fit.

Table 9.2: Uncertainties on the estimation of the QCD-multijet normalization resulting from quadratic sum of the fit uncertainty and the fit-factor variation resulting from correlations between the channels when estimating the QCD-multijet shape uncertainty in each channel. The uncertainties on the estimated cross-section for W +HF and W +light background processes are also shown.

Template	Uncertainty
QCD-multijet (e +jets channel - 1 tag)	$\pm 2\%$
QCD-multijet (e +jets channel - ≥ 2 tags)	$\pm 3\%$
QCD-multijet (μ +jets channel - 1 tag)	$\pm 13\%$
QCD-multijet (μ +jets channel - ≥ 2 tags)	$\pm 14\%$
W +HF	$\pm 3\%$
W +light	$\pm 40\%$

Significantly larger uncertainties arise in the μ +jets channel than in the e +jets channel for the QCD-multijet model. This is due to a strong impact of the correlations between the QCD-multijet templates of the e +jets and μ +jets. They lead to a significant variation of the fit result in the μ +jets channel (12%), when the shape uncertainty of the QCD-multijet model in the e +jets channel is being estimated. Moreover, the fit results for the QCD-multijet templates in the μ +jets channel show originally larger uncertainties (8%) than in the e +jets channel (3%).

The data-driven method presented in this chapter can successfully estimate the contaminations from QCD-multijet and W +HF/light background events in the selected data events simultaneously. The fit result for the signal contribution is consistent with the predictions so no unwanted bias is expected on the result.

Chapter 10

Event Yields and Control Plots

In the two previous chapters the analysis has been prepared by means of selection of data and MC simulation, event corrections, and estimation of background contributions. The resulting event yields are given in Table 10.1 for each MC simulation sample and for data.

Table 10.1: Number of events of MC-simulation samples and data fulfilling the full event selection (documented in Sec. 8.1) in the 1-tag region and the ≥ 2 -tag region with statistical uncertainties. All corrections described in Sec. 8.3 and background estimations from Chapter 9 are included.

	e +jets		μ +jets	
	1 tag	≥ 2 tags	1 tag	≥ 2 tags
$t\bar{t}$	74,830 \pm 80	59,500 \pm 70	84,830 \pm 80	68,260 \pm 80
Single top	4,660 \pm 40	2,550 \pm 30	5,320 \pm 50	2,860 \pm 40
W +b-jets	5,220 \pm 70	1,720 \pm 40	5,870 \pm 70	1,920 \pm 40
W +c-jets	8,880 \pm 90	560 \pm 20	9,960 \pm 90	670 \pm 23
W +light-jets	980 \pm 10	26 \pm 2	1,110 \pm 11	34 \pm 2
Z +jets	2,070 \pm 70	360 \pm 30	1,700 \pm 60	260 \pm 30
QCD multijet (fit)	5,830 \pm 50	1,080 \pm 30	2,000 \pm 50	300 \pm 20
Diboson	343 \pm 6	70 \pm 2	370 \pm 6	80 \pm 3
Expected sum	102,800 \pm 200	65,900 \pm 100	111,200 \pm 200	74,400 \pm 100
Data	98,900 \pm 300	65,600 \pm 300	106,400 \pm 300	75,300 \pm 300

The sum of the expectations from MC simulation and from background modeling overestimates slightly the yields for 1-tag events. While this discrepancy is canceled in the ≥ 2 -tag region of the e +jets channel, in the μ +jets channel the number of events estimated is significantly below the data value, accord-

ing to the quoted uncertainties. Including relevant systematic uncertainties, as illustrated for instance in the figures in Appendix B, it can be seen that these systematic uncertainties fully cover the discrepancies.

The expected number of $t\bar{t}$ events in the signal region (≥ 2 tags) represents approximately 90% of the total expected number of events. This good signal purity expected in data events allows a precise measurement of $t\bar{t}$ events, as performed in the next chapter.

As a control of the signal and background modeling, distributions are shown before the detector-effect corrections, described in Chapter 11, are applied. Fig. 10.1 and 10.2, show selected events but requiring exactly one b -tagged jet, while Fig. 10.3 and 10.4 represent the events fulfilling the full selection (i.e. requiring at least two b -tagged jets). Further distributions are shown in Appendix C.

In general, good agreement ($\leq 10\%$ disagreement) is observed in both 1-tag and ≥ 2 -tag regions and in both e +jets and μ +jets channels. In the 1-tag region, the offset observed in Table 10.1 is visible and it is strongly reduced in the ≥ 2 -tag region. However, some significant discrepancies are discussed here, while it has to be kept in mind that only statistical uncertainties are shown on the data points:

- Fig. 10.1 and 10.3: the jet-multiplicity is the measured variable. A good agreement is observed up to 7 or 8 jets. See Sec. 11.4 for the discussion of the results including all systematic uncertainties. The jet- η distributions show slight discrepancies between data and the MC simulation. The most significant effects are visible for high η , where detection is more challenging due to higher activity and PU. In Appendix B, the same distributions include the uncertainties from the Q^2 -scale uncertainty in the $t\bar{t}$ sample and from the jet-energy scale uncertainties, showing that all discrepancies are covered.
- Fig. 10.2 and 10.4: the electron- p_T distribution for 1-tag events shows a discrepancy at low p_T . This shows a slight overestimate of the QCD-multijet cross section. The discrepancy is solved when requiring at least two b -tags.

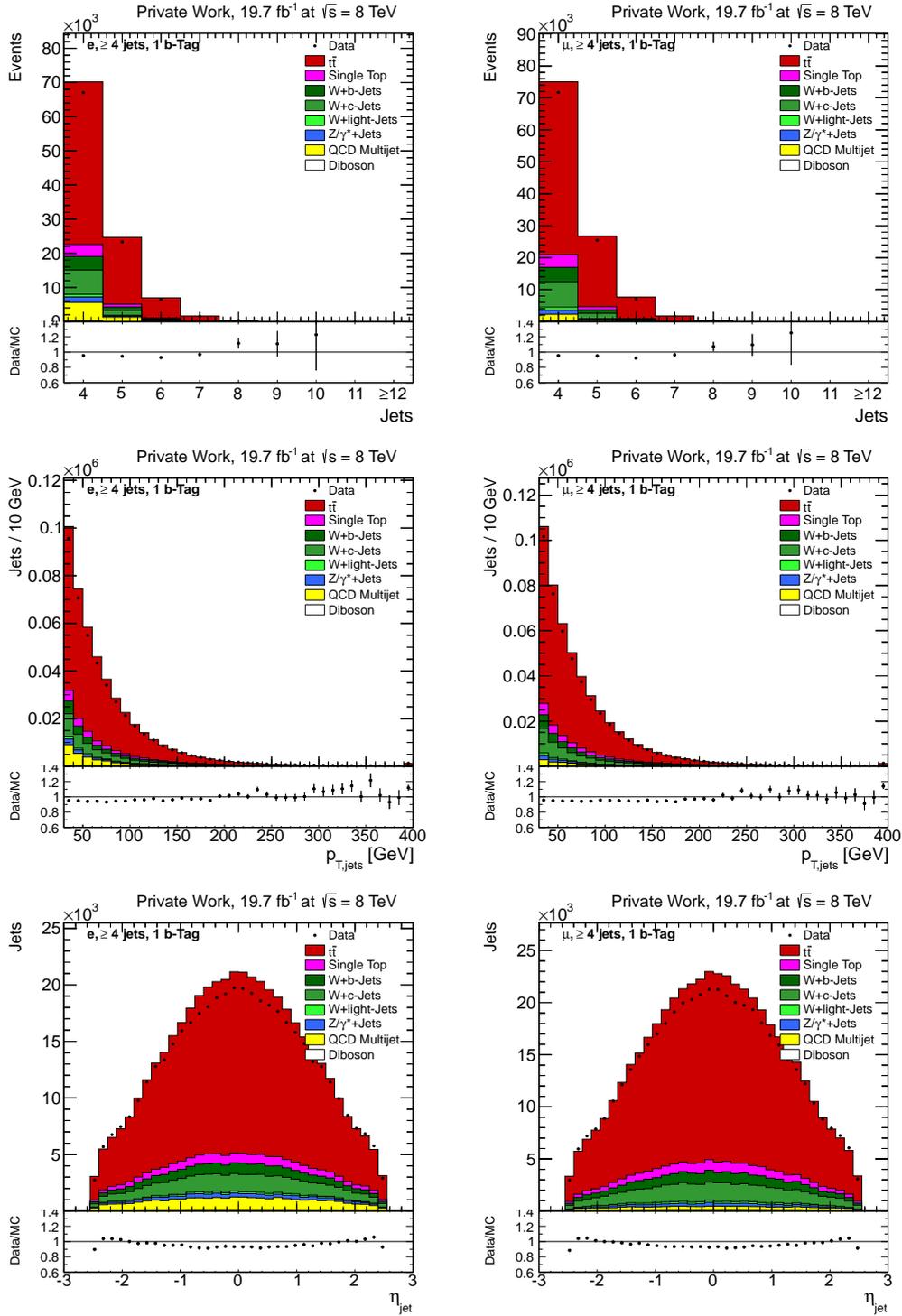


Figure 10.1: Distribution of the number of selected jets (top), their p_T (middle), and their η (bottom), in the e +jets channel (left) and μ +jets channel (right) for 1-tag events.

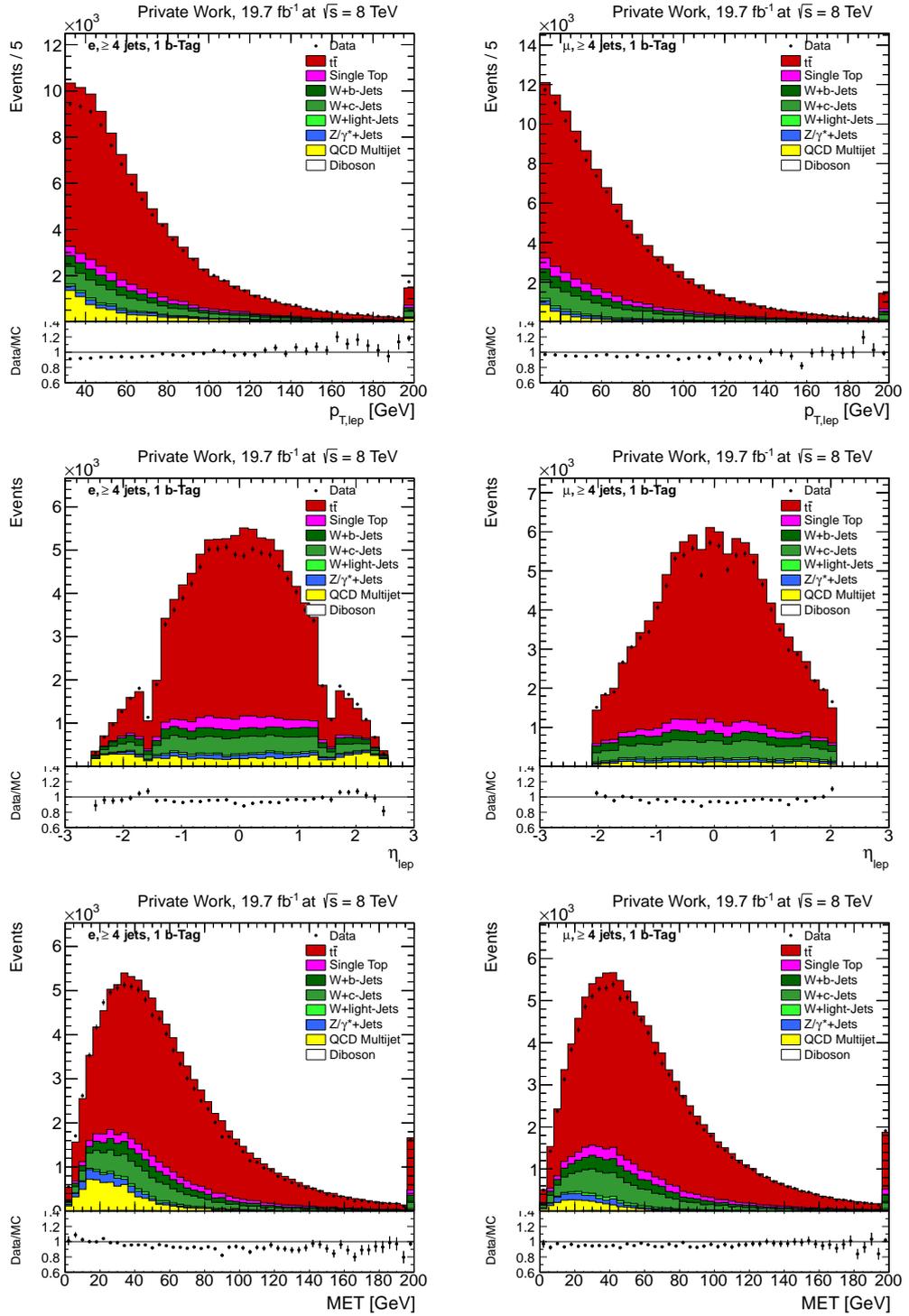


Figure 10.2: Distribution of the p_T of the selected lepton (top), η of this lepton (middle), and of the \cancel{E}_T (bottom), in the $e+$ jets channel (left) and $\mu+$ jets channel (right) for each 1-tag event.

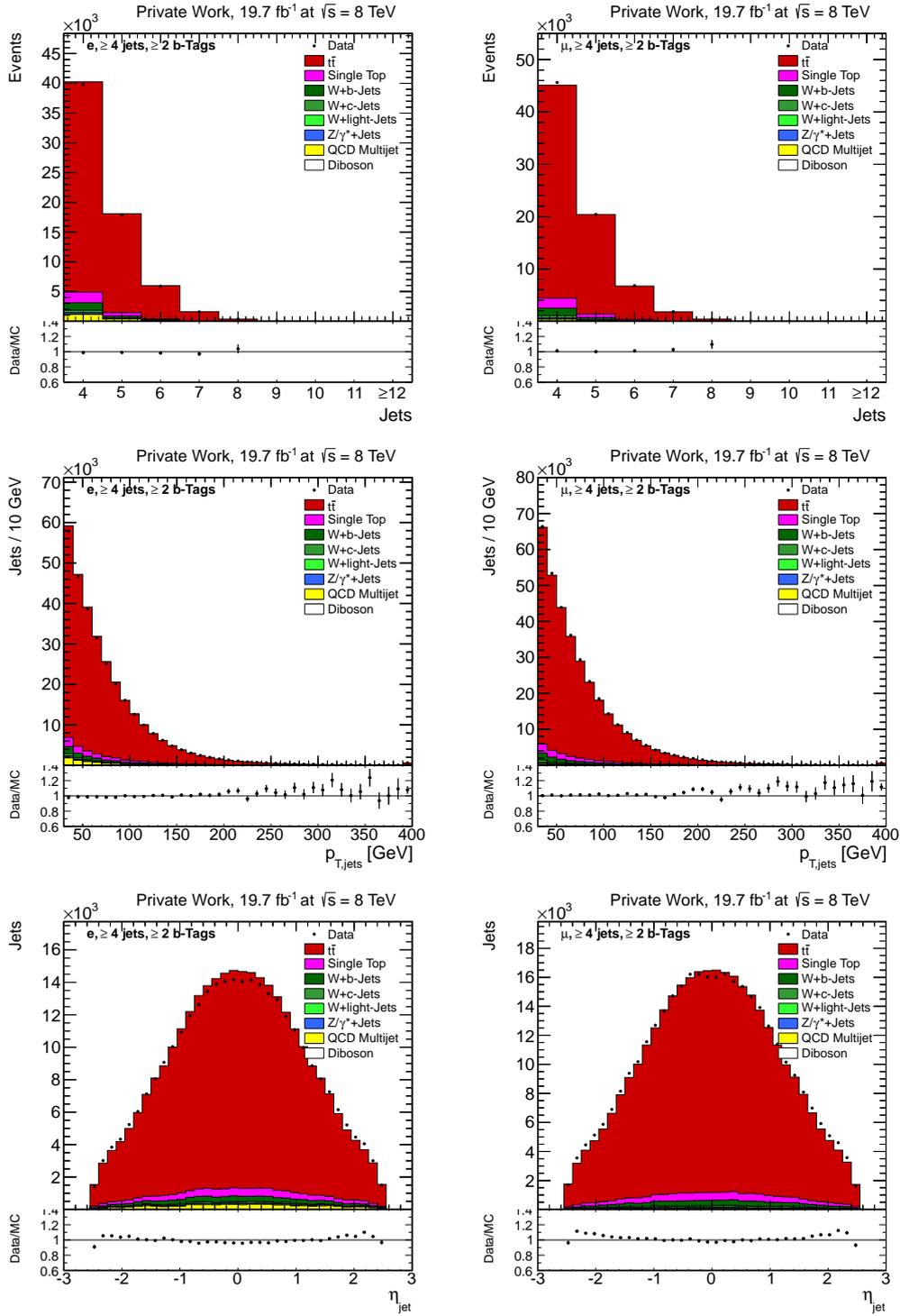


Figure 10.3: Distribution of the number of selected jets (top), their p_T (middle), and their η (bottom), in the e +jets channel (left) and μ +jets channel (right) for ≥ 2 -tag events.

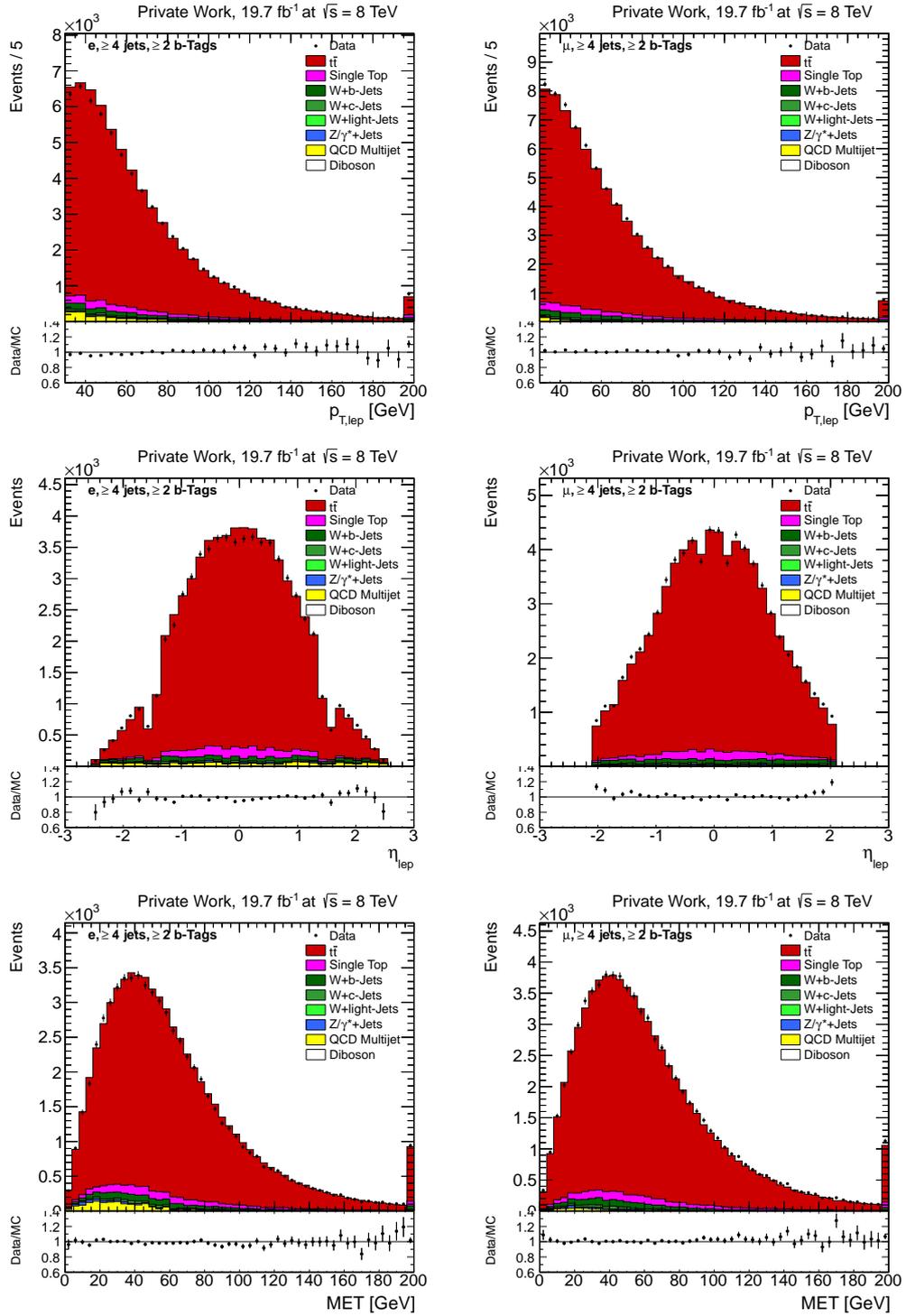


Figure 10.4: Distribution of the p_T of the selected lepton (top), η of this lepton (middle), and of the \cancel{E}_T (bottom), in the e +jets channel (left) and μ +jets channel (right) for each ≥ 2 -tag event.

Chapter 11

Jet Multiplicity in Top-Quark Pair Production

In the previous Chapters 8 and 9, the events to be analyzed have been selected and corrected. Background contributions remaining in the selected data events have been estimated. This allows at this point to present the measurement itself. Starting with the number of jets found in the selected events, the measurement of the differential fiducial cross section of $t\bar{t}$ events as a function of the number of jets is described in this chapter.

All non- $t\bar{t}$ contributions estimated in Chapter 9 or from MC simulation, listed in Sec. 8.2, are subtracted from the jet-multiplicity spectrum in data, separately in the e +jets and μ +jets channels. The two resulting distributions of the number of jets could be converted to a differential cross section by a trivial division by the integrated luminosity of the data. However, this measurement could not be compared to measurements from other experiments or from theory calculations because the measured jet multiplicity is affected by the detection process. While this is taken into account in the MC simulation provided by the CMS collaboration through the detector simulation, it is not available to any external communities. Therefore, the results without corrections for the detector effects are not “universal”, because they cannot not be reproduced outside CMS.

In Sec. 11.1, corrections are presented in order to produce a particle-level result which is defined in a given phase space, called fiducial phase space. The dedicated selection describing this phase space has been presented in Sec. 8.1.5, making use of particle-level objects. The corrections presented enable to convert the jet multiplicity detected by CMS into the jet multiplicity which would have been seen within the fiducial phase space if CMS was

a “perfect” detector. Detector effects which have to be corrected are often called migrations, as they change the number of jets and events counted for the results. For comprehension, migrations can be split into three parts:

- Incoming migrations represent cases in which a $t\bar{t}$ event has been detected by the CMS detector but it did not fulfill the fiducial-phase-space selection. Such migrations can occur e.g. through energy-resolution effects causing the detector to measure the p_T of a jet above the selection threshold while the jet p_T is below the threshold on particle level. Moreover, b -tagging and charged lepton selections can also cause incoming migrations, also taken into account.
- Internal migrations occur when a $t\bar{t}$ event fulfills both selections (detector and particle level) but it shows a different number of jets before and after detection. In other words, it has different numbers of jets at particle and detector level (but the number of jets is always at least larger than the minimal number of four). For instance, if a jet with sufficient p_T cannot be fully measured because it traverses a non-sensitive part of the detector, this jet might carry less p_T and therefore be discarded, reducing the overall number of jets detected. Through e.g. energy resolution effects, as presented for the incoming migration, the opposite effects can occur as well, increasing the number of jets detected. For each event, the probability that it is measured with more or fewer jets than at particle level is roughly equal, as illustrated in Fig. 11.1, when reading the matrix along individual rows, i.e. for a fixed number of particle-level jets. However, the jet-multiplicity spectrum is falling strongly, hence the absolute exchanges between two neighboring bins is not equal, it is stronger in the direction of increasing multiplicity (“feed down”). This statistical effect needs to be corrected as well.
- Outgoing migrations represent cases in which a $t\bar{t}$ event decaying within the fiducial phase space is seen by the CMS detector as deficient, so it was canceled. In this case, opposite to the incoming migrations, a jet e.g. had a p_T large enough to fulfill the selection when entering the detector, but some energy deposition have not been detected, so the jet energy is lower, hence it is not selected. The same effects can take place for b -tagging or the charged lepton selection.

In practice, these migrations are addressed by two distinct correction steps presented in the next section. The first one, called here unfolding, takes care of internal migrations and incoming migrations which are concerning only the jets. The second correction step addresses all outgoing migrations as well as the remaining incoming migrations (i.e. those not related to jets).

This second step is called here acceptance correction.

11.1 Conversion of the Jet-Multiplicity Spectrum to the Particle Level in a Fiducial Phase Space

Assuming a correct simulation of all detector effects in the $t\bar{t}$ MC simulation, the migrations can be corrected. The unfolding, whose technical description was given in Sec. 5.2, corrects internal migrations (e.g. five particle-level jets migrating to four detector-level jets) and incoming migrations concerning the number of jets (e.g. three particle-level jets migrating to four detector-level jets). The migration matrix A_{ij} is built with the number of jets detected as a function of number of jets before the detection process has affected the events. The events filling the matrix are $t\bar{t}$ MC events fulfilling the detector-level selection, hence the matrix provides the migration information for the events at the detector level, which are those to be corrected. The migration matrices of the e +jets and μ +jets channels are represented in Fig. 11.1. In both channels, one can see that the diagonal elements dominate. As mentioned previously, for n particle-level jets, the numbers of events with $n + 1$ and $n - 1$ detected jets are approximately the same. While the latter observation succeeds when looking at fixed rows of the matrices, a different conclusion is found when looking at fixed columns: for m detector-level jets, the number of events containing $m - 1$ particle-level jets is generally larger than with $m + 1$ particle-level jets. This is a consequence of the feed-down effect mentioned previously. For instance, among the events in which five jets are detected, a larger fraction of these events contain four jets at particle level than six. This is due to the much larger number of events containing four particle-level jets than six, which leads to an overall stronger event migration from events with four particle-level jets than from those with six particle-level jets.

Based on the Singular Value Decomposition (SVD) method [108, 147, 148], whose basics are given in Sec. 5.2, the inverse migration matrix A_{ij} is calculated. The value of the regularization parameter for the SVD method, τ , is chosen in order to minimize the averaged square global correlation between bins of the unfolded jet-multiplicity spectrum.

The purity and stability of the migration matrix are defined with the number of events $N_{x,y}$ fulfilling the detector-level selection and containing x detector-

level jets and y particle-level jets:

$$\text{Purity}(i) = \frac{N_{i,i}}{\sum_j N_{i,j}} \quad \text{Stability}(i) = \frac{N_{i,i}}{\sum_j N_{j,i}}.$$

The purity and the stability found in the migration matrices of both channels can be seen in Fig. 11.2. High values of the purity and stability translate into limited corrections required, and vice versa. Both variables drop to lower values for higher jet multiplicities, which is expected since the more jets are produced at particle-level, the more probable it is that one of them migrates, i.e. at least one jet more or one jet less is found at the detector level. While the lowest value of the stability is at $\approx 45\%$ for both channels, the purity descends down to $\approx 35\%$ ($\approx 30\%$) in the e +jets (μ +jets) channel. The lower value for the purity is a consequence of the feed-down effect described previously and observed in the migration matrices. The discrepancies observed between both channels are covered by the statistical uncertainties which are shown in Fig. 11.2.

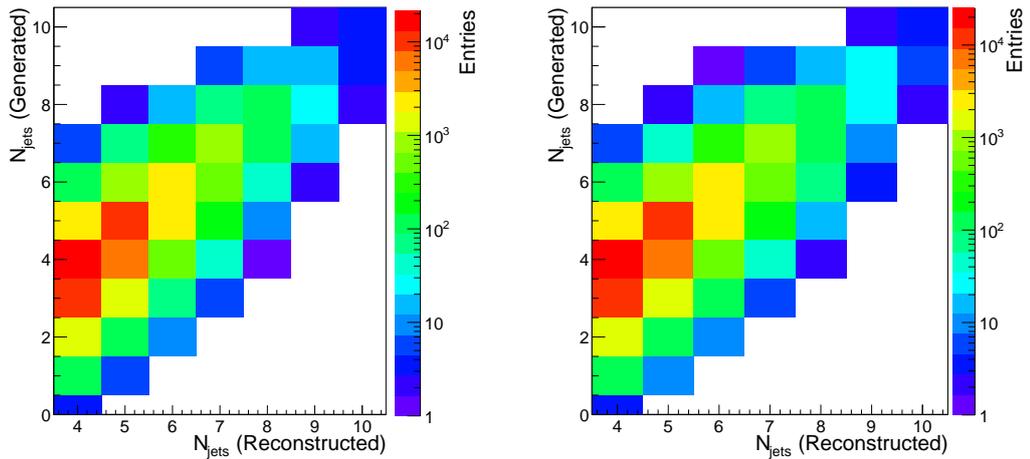


Figure 11.1: Migration matrix simulated in $t\bar{t}$ events after applying the full detector-level selection in the e +jets (left) and μ +jets (right) channel. The x -axis represents the number of detector-level jets and the y -axis represents the number of particle-level jets. The z dimension represents the number of $t\bar{t}$ MC simulation events showing the (x, y) jet configuration. Each of these matrices is used for the unfolding in the corresponding channel.

The optimal value of the regulation factor τ found in the e +jets and μ +jets channel are 0.39 and 0.41, respectively. Fig. 11.3 shows the values of the averaged square global correlation and the curvature of the unfolded jet-multiplicity spectrum as a function of τ . It can be seen that the curvature is larger in the μ +jets channel (≈ 5) than in the e +jets channel (≈ 2).

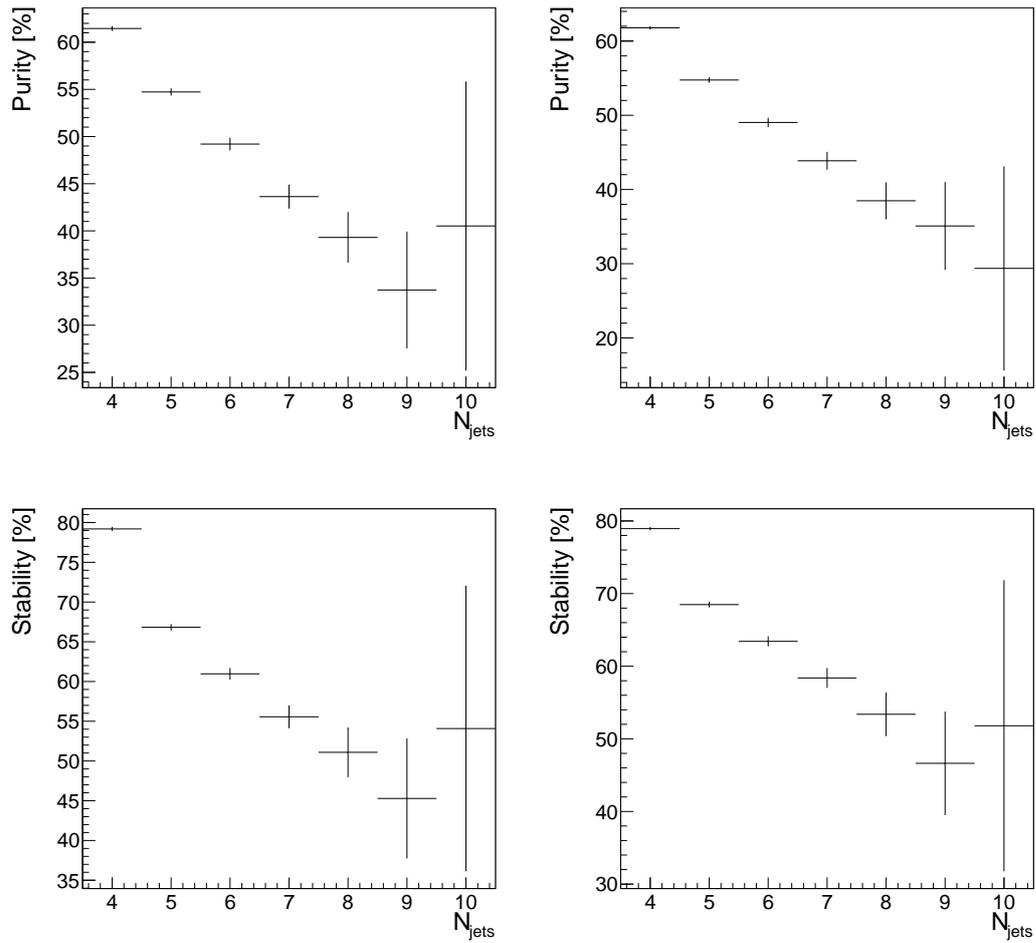


Figure 11.2: Purity and stability in percent (top and bottom) estimated in the e +jets and μ +jets channels (left and right) with the $t\bar{t}$ MC simulation.

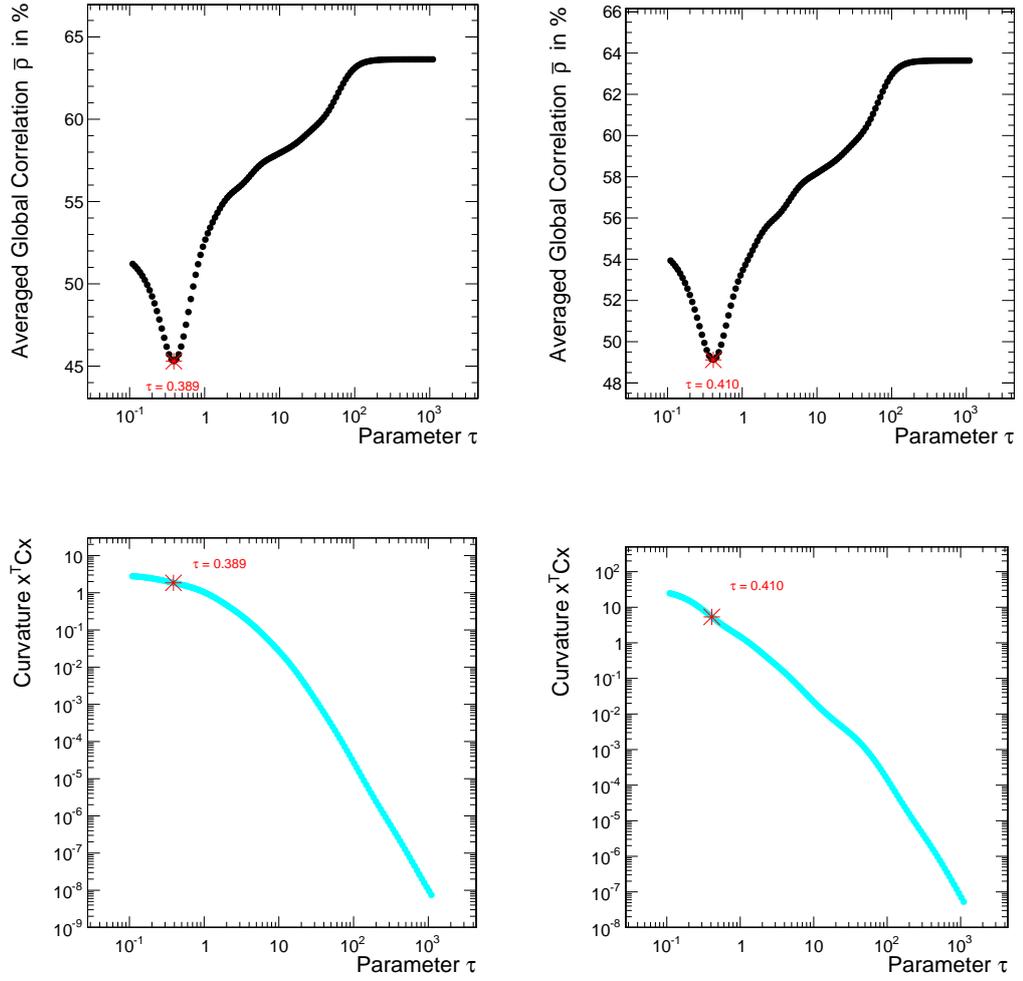


Figure 11.3: Scan of τ parameter in the e +jets (left) and μ +jets (right) channel. In the two top plots, the minimum of the averaged global correlation (y -axis) is searched and the corresponding τ (the red star marker) is taken for the actual unfolding utilized for the results quoted in Sec. 11.4. In the two bottom plots, the curvature of the resulting jet-multiplicity spectrum is drawn as a function of τ .

The statistical uncertainties arising from the unfolding are estimated with pseudo-experiments. The unfolding is repeated a thousand times and for each experiment the migration matrix is replaced by a pseudo matrix which reproduces the original one, but containing values varied randomly within statistical uncertainties. For this, each position (x, y) of a pseudo matrix is filled with a random value taken from the Poisson probability distribution whose expectation value is the value for (x, y) in the original migration matrix.

Finally, a closure test is performed. The unfolding of the non-unfolded $t\bar{t}$ events from MC simulation fulfilling the detector-level selection should in principle result exactly in the particle-level spectrum, obtained with the particle-level selection. Because of the statistical tools employed for the unfolding, an exact match is not expected. The test is very well passed in the e +jets channel. In the μ +jets channel, discrepancies are spotted for the highest jet multiplicities but they are very limited: from four up to eight jets, the discrepancies represent less than 1% of the statistical uncertainties on the result. For nine and at least 10 jets the discrepancies represent 3 and 15% of the statistical uncertainties. It is interesting to observe that the increasing discrepancies observed for the largest jet multiplicities can be related to the larger curvature of the unfolded jet-multiplicity vector in the μ +jets channel, as observed in the curvature plots of Fig. 11.3. Anyway, the unfolding process passes the closure test in both channels since the bias observed in the μ +jets channel is not seen as significant with respect to the total uncertainties.

Remaining migrations - Acceptance A final correction is applied on the unfolded distribution of the number of jets in order to address migrations remaining after the unfolding step. The ratio $\varepsilon_i = n_i^{\text{detec}}/n_i^{\text{part}}$ is calculated, where n_i^{detec} (n_i^{part}) is the number of events fulfilling the detector-level (particle-level) selection and showing i particle-level jets. This step converts the phase space from detector level to particle level.

11.2 Measurement of the Top-Quark Pair Differential Cross Section as a Function of the Number of Jets

Calculation in Separate Channels The schematic formula utilized for the calculation of the differential cross section with i particle-level is:

$$\frac{d\sigma_i}{dN_{\text{jets}}} = \frac{\sum_j \text{unf}_{ij} (n_j^{\text{data}} - n_j^{\text{non-}t\bar{t}})}{\mathcal{L} \cdot \varepsilon_i}.$$

The term $n_j^{\text{data}} - n_j^{\text{non-}t\bar{t}}$ shows the subtraction of non- $t\bar{t}$ contributions from the data for events showing j detector-level jets. The factor unf_{ij} symbolizes the correction of migration effects which have brought events with i particle-level jets to be detected with j jets. A sum over the contributions from all multiplicities of detector-level jets is performed. The factor ε_i provides the acceptance correction and the conversion of the number of events to a cross section is done dividing by \mathcal{L} , the integrated luminosity of the data.

Calculation of Predictions The corresponding predictions are calculated after the particle-level selection has been applied on the $t\bar{t}$ MC simulation, as given in Sec. 8.1.5. The number of remaining MC $t\bar{t}$ events showing i particle-level jets is scaled as follows in order to produce the predicted differential cross section:

$$\frac{d\sigma_i^{\text{pred}}}{dN_{\text{jets}}} = W_i \frac{\sigma_{t\bar{t}} \cdot c_{\text{BR}}}{N^{\text{MC}}},$$

where $\sigma_{t\bar{t}} = 252.89$ pb is the total $t\bar{t}$ NNLO+NNLL cross section calculated by the TOPLHCWG [26], using top++ [24], given in Sec. 8.2. The branching ratio of $t\bar{t}$ events into the ℓ +jets channel taken in MADGRAPH is slightly wrong. Therefore, in this case, it is corrected by applying $c_{\text{BR}} = (0.108 \cdot 9) \cdot (0.676 \cdot 1.5) = 0.985608$ on the total $t\bar{t}$ cross section (only when using MADGRAPH). This is an approximation since not only ℓ +jets events, according to parton-level information, are selected by the particle-level selection. The number of events found in the MC-simulation samples before any selection is N^{MC} .

When using MC simulation from NLO generators like POWHEG, SHERPA, MC@NLO, or aMC@NLO, events carry weights not equal to unity which has to be taken into account. Therefore, “the number of events” should always be understood as “the sum of the event weights”.

Result Combination The combined results are the sum of the results from the separate calculations in the e +jets and μ +jets channels. The corresponding statistical uncertainties are added in quadrature since they are fully uncorrelated. The combined systematic uncertainties are calculated as described in the following.

11.3 Systematic Uncertainties

For the present measurement, the principle of the systematic uncertainty calculation is to repeat the whole analysis with varied configuration according to the uncertainty sources. This includes the variation of the migration matrix for any systematic uncertainty influencing the $t\bar{t}$ MC simulation. The background estimation is the exception, it is not repeated for each systematic uncertainty because it has dedicated systematic sources (model-shape uncertainties) whose resulting fit factors are propagated consistently through the analysis.

The shifts between the resulting cross-section values and the nominal values (i.e. without any varied configuration) are the systematic uncertainties.

For simplification, each systematic uncertainty is symmetrized: in general up and down variations of the parameters occur. The largest shift of the result for each cross-section value is chosen as representative of this systematic. In some cases the systematic uncertainty is by definition made of only one variation. Then, the uncertainty is also made symmetric with the same procedure.

Moreover, for most systematic uncertainties a full correlation is expected between the e +jets and the μ +jets channel. Therefore, for the combined ℓ +jets channel, each systematic uncertainty is the linear sum of the systematic uncertainty found in each channel. Lastly, all individual sources of systematic uncertainty are expected to be uncorrelated with all others. Thus, the total systematic uncertainties, as well as the total uncertainties are calculated for the combined ℓ +jets channel by summing all components in quadrature.

Systematic uncertainties originate from detector effects, theoretical assumption, and modeling uncertainties. They are described in the following.

11.3.1 Detector-Related Systematic Uncertainties

Significant sources of systematic uncertainties originate from the detection of data events. This section also contains systematic uncertainties which originate from modeling uncertainties but they are estimated centrally by the CMS collaboration, e.g. b -tagging efficiency and PU.

Jet-Energy Scale (JES) The jet-energy corrections applied are affected by uncertainties. They are taken into account repeating the analysis with the jet-energy correction shifted up and down by its uncertainty in all MC-simulation samples. The effect on the result is the “JES” uncertainty.

Jet-Energy Resolution (JER) A correction of the jet-energy resolution is applied in MC simulation, see in Sec. 8.3.1. This correction has an uncertainty which is taken into account repeating the analysis with the resolution correction increased or decreased by the uncertainty on this correction. The effect on the result is the “JER” uncertainty.

b -Tagging Scale Factor The b -tagging efficiency correction applied in MC simulation, as presented in Sec. 8.3.5, has uncertainties. The uncertainties are fully correlated for b -jet and c -jet efficiencies. Therefore, the b -tagging efficiency uncertainty is estimated independently in the case of b and c -jets from the case of light-jets. For both the procedure is similar: the analysis is repeated varying the corresponding SF up and down within their uncertainties. The effects on the result are the “ b -Tagging (b,c)” and “ b -Tagging (light)” uncertainties.

Pile-Up Correction For the estimation of the pile-up intensity in data, an inelastic scattering cross section is assumed in order to translate the instantaneous luminosity into a rate of interactions taking place at the CMS collision point. This is required for the correction of the pile-up intensity in the MC simulation, as described in Sec. 8.3.2. The 5% uncertainty on this assumed cross section (69.4 mb) is taken into account by repeating the analysis with the total inelastic cross section varied up and down within its uncertainty. The effect on the result is the “Pile-up” uncertainty.

Correction of the Lepton Reconstruction, Selection, and Trigger Efficiency The various lepton-efficiency corrections applied in MC simulation (see in Sec. 8.3.4) contain uncertainties. Varying simultaneously all of these corrections up and down within their uncertainty, the analysis is repeated. The effect on the result is the “Lepton Efficiency” uncertainty.

Luminosity The uncertainty on the luminosity measurement is $\pm 2.6\%$ [149]. This is taken into account by repeating the analysis with the value of the luminosity varied up and down, within its uncertainty. This affects only the normalization of the MC simulation and the denominator of the cross-section calculation. The effect on the result is the “Luminosity” uncertainty.

11.3.2 Modeling-related systematic uncertainties

Most model dependencies of the present measurement arise because of the migration corrections (unfolding and acceptance corrections).

Factorization and Renormalization Scale The $t\bar{t}$ MC simulation requires assumption on the factorization and renormalization scale, leading to uncertainties on these values. These uncertainties are here especially important because the generator of the standard $t\bar{t}$ MC simulation utilizes the leading order of the perturbation theory, which shows a significant dependency on the scale values. Varying at the same time the value of the factorization and renormalization scale with a factor of 2 and 0.5 with respect to the value of the standard $t\bar{t}$ MC simulation, two alternative $t\bar{t}$ MC-simulation samples have been produced. The analysis is repeated with these two alternative $t\bar{t}$ MC-simulation samples and the effect on the result is the “ $Q^2(t\bar{t})$ ” uncertainty.

Jet-Energy Matching Threshold Between Matrix-Element and Parton-Showering Similarly to the previous uncertainty source, an assumption has to be made on the jet-energy threshold applied when interfacing the matrix-element (here MADGRAPH) to the parton-showering (PYTHIA) tool. Two alternative $t\bar{t}$ MC-simulation samples have been produced using a jet-energy threshold varied by a factor of 2 and 0.5 with respect to the value of the standard $t\bar{t}$ MC simulation. The analysis is repeated with these two alternative $t\bar{t}$ MC-simulation samples and the effect on the result is the “Matching ($t\bar{t}$)” uncertainty.

Top- p_T Reweighting Because of the bad modeling of the top- p_T spectrum in the $t\bar{t}$ MC simulation, an event-weight based correction is applied, see Sec. 8.3.3. An estimate of the uncertainties from this correction is obtained by repeating the analysis without this correction. The effect on the result is the “Top- p_T Rew.” uncertainty.

LO versus NLO The generation of the $t\bar{t}$ MC simulation is achieved with a LO generator tool (MADGRAPH). This means neglecting Feynman diagrams, i.e. contributions to the SM description of the $t\bar{t}$ production. A systematic uncertainty takes into account this simplification by repeating the analysis when the $t\bar{t}$ MC simulation from POWHEG+PYTHIA is used instead of the standard one. The effect on the result is the “LO vs. NLO ” uncertainty.

Parton-Showering and Hadronization Tool The effect of the parton-showering and hadronization step on the MC simulation used in the measurement should be tested. This could be done by repeating the analysis when using the POWHEG+PYTHIA and the POWHEG+HERWIG and comparing the results. However, the MC simulation from the latter tool does not describe the data well¹, due to the bad modeling of b -tagging. In the previous measurement at $\sqrt{s} = 7$ TeV [35], the same problem happened, and therefore POWHEG+PYTHIA and MC@NLO+HERWIG have been used. Since the comparison of these tools includes also effects from the matrix-element generators, the resulting shifts of the results have been arbitrarily averaged. In the present measurement, see the next section, it is shown that MC@NLO+HERWIG does not describe data well for the high jet multiplicity. For at least ten jets, the predicted cross section takes a negative value, because of the negative event weights of MC@NLO, combined with a too small number of available events. This discards MC@NLO for the estimation of the systematic uncertainty. As a consequence, no reliable MC simulation is ready on time for the estimation of this systematic uncertainty, so it has to be neglected. In the 7 TeV measurement, this systematic uncertainty turned out to be negligible for low jet multiplicities and it increases up to $\approx 10\%$ for at least eight jets. While this is non-negligible compared to the total systematic uncertainty found in the present analysis for eight jets ($\approx 30\%$), it would increase the total systematic-uncertainty value from 30%

¹This can be related to outdated tools or wrong settings. At the time of the MC production, HERWIG 6 has been used, whereas this is superseded by HERWIG++.

to 32%. This does not influence the statement of this analysis, and therefore it is acceptable to neglect the present systematic uncertainty.

Parton-Distribution Function and Color Reconnection Modeling

The effects of the uncertainties on the PDF and the color reconnection model have not been estimated for the present measurement. However, they have both been estimated in the previous measurement at $\sqrt{s} = 7$ TeV [35] and have shown to have a very limited effect on the normalized differential fiducial cross section as a function of the number of jets ($\leq 2\%$ for PDF, between 1 and 5%, for the lowest and the highest jet multiplicities, for the color reconnection). This is documented in the CMS internal note [150]. As a consequence, these uncertainties can be neglected for this analysis.

11.3.3 Background-Related Systematic Uncertainties

Background contributions estimated with MC simulation and data-driven techniques are subtracted from the data events before the migration corrections are applied. Therefore, the uncertainties from the background estimations have an impact on the result. However, their influence is small due to the limited background contributions in the signal region.

Cross Section for Single-Top, Z +jets, and Diboson Processes The cross sections assumed for the normalization of the MC simulation of the background processes are taken from calculations, as referenced in Sec. 8.2. They are affected by uncertainties which are taken here into account. The analysis is repeated for the single-top, Z +jets, and diboson MC simulation, varying their cross section up and down within their uncertainties:

- Single top cross sections for all channels (t , s , and tW channels) and for top quark and anti-quark production are varied simultaneously. The largest uncertainty, arising in the calculation in the tW channel at NNLO approximation, is taken: $\pm 7\%$ [123].
- Z +jets: the theory uncertainty on inclusive Z/γ^* +jets at NNLO of $\pm 3\%$ is utilized, as performed in [145].
- Diboson in WW , WZ , and ZZ production channels. The uncertainties are taken from NLO calculations [146]: $\pm 5\%$.

These variations are done separately for each of these three processes. The effects on the results due to the cross-section uncertainties are called the “Single Top CS”, “Z+Jets CS”, and “Diboson CS” uncertainties.

Cross Section for W +jets and QCD-Multijet Estimations The normalizations of the W +jets MC simulation and the QCD-multijet model are estimated by means of a fit to data, as presented in Sec. 9. The results of the fit contain uncertainties: the W +HF and W +light categories have respectively 3% and 35% uncertainties and the QCD-multijet model in the e +jets and μ +jets channels have respectively 5% and 16% uncertainties, as given in Table 9.2. As no systematic uncertainty is considered for the MC simulation of W +jets events², the normalization uncertainty of 3% provided by the fit for the W +HF part is probably underestimated. Since an uncertainty of 50% has been applied on W + b -jets in the previous measurement [35], smaller uncertainties are assumed here since a dedicated measurement has been performed. Taking 30% for the cross section uncertainty of the W +HF leads to systematic uncertainties up to 1%, which is anyway not significant compared to other systematic uncertainties. This is taken as a rough estimate, while larger values like 50% would also not lead to significant effects on the total systematic uncertainties. The analysis is repeated for each process: the QCD-multijet model in the e +jets and μ +jets channel, the W +HF, and the W +light, by varying the cross sections up and down within their uncertainties. The respective effects on the result are called “ W +HF-Jets CS”, “ W +Light-Jets CS”, and “QCD CS”.

Background-Estimation Shape Uncertainties In the fit for the background estimation, the ratio of W +jets events showing b -jets and c -jets is fixed to the value given by the MC simulation. This assumption gives rise to a shape uncertainty as this ratio is not known to describe the data correctly. This shape uncertainty is evaluated by repeating the whole analysis (including the background estimation) varying the b/c ratio in the W +jets MC simulation with factors 2 and 0.5. The effect on the result is the “ B/C in W +jets” uncertainty.

The shape uncertainties of the QCD-multijet models are estimated by replacing the QCD-multijet models with alternative ones, as described in Sec. 9.4. The analysis is repeated with each of these alternative QCD-multijet models,

²Variations of Q^2 scale and jet-matching threshold in the W +jets MC simulation, as done for the $t\bar{t}$ MC simulation, would provide such an estimate. This could not be performed in time for this thesis.

including the fit results. The effects on the results are called ‘‘QCD Shape’’ and ‘‘QCD Rellso’’ uncertainty in the e +jets and μ +jets channels.

11.4 Results and Discussion

The results before combination are given in Table 11.1 and 11.2, all systematic uncertainties are listed in Table 11.3 and 11.4.

Table 11.1: Differential fiducial $t\bar{t}$ cross section as a function of the number of particle-level jets in the e +jets channel. The relative statistical, systematic, and total uncertainties are also shown.

Category	$d\sigma/dN_{\text{jets}}$ [pb]	Stat. unc.	Syst. unc.	Total unc.
$t\bar{t} \rightarrow e + 4$ jets	4.3	1%	8%	8%
$t\bar{t} \rightarrow e + 5$ jets	2.0	2%	12%	12%
$t\bar{t} \rightarrow e + 6$ jets	0.66	4%	13%	14%
$t\bar{t} \rightarrow e + 7$ jets	0.17	8%	18%	20%
$t\bar{t} \rightarrow e + 8$ jets	0.041	18%	25%	31%
$t\bar{t} \rightarrow e + 9$ jets	0.015	20%	37%	42%
$t\bar{t} \rightarrow e + \geq 10$ jets	0.0046	25%	32%	40%

Table 11.2: Differential fiducial $t\bar{t}$ cross section as a function of the number of particle-level jets in the μ +jets channel. The relative statistical, systematic, and total uncertainties are also shown.

Category	$d\sigma/dN_{\text{jets}}$ [pb]	Stat. unc.	Syst. unc.	Total unc.
$t\bar{t} \rightarrow \mu + 4$ jets	4.4	1%	8%	8%
$t\bar{t} \rightarrow \mu + 5$ jets	2.0	1%	11%	11%
$t\bar{t} \rightarrow \mu + 6$ jets	0.70	3%	13%	13%
$t\bar{t} \rightarrow \mu + 7$ jets	0.19	7%	22%	23%
$t\bar{t} \rightarrow \mu + 8$ jets	0.042	17%	33%	37%
$t\bar{t} \rightarrow \mu + 9$ jets	0.015	17%	45%	49%
$t\bar{t} \rightarrow \mu + \geq 10$ jets	0.0055	21%	52%	56%

The results in both lepton channels show good agreement with each other within the uncertainties. The total uncertainties are dominated by systematic uncertainties over the whole jet-multiplicity range.

The combined results are shown in Table 11.5. They are compared to predictions from MC simulations produced with the nominal MADGRAPH+PYTHIA at LO, and from the alternative NLO generators POWHEG+

Table 11.3: Summary of all relative systematic uncertainties (in %) for the differential fiducial $t\bar{t}$ cross section as a function of the number of particle-level jets in the e +jets channel. The uncertainties from: b-tagging (light), lepton efficiency, and cross section of single top, W+Light-Jets, Z+Jets, and Diboson are all $\leq 1\%$. Therefore, they are not shown in the table.

$e + N$ jets:	4 jets	5 jets	6 jets	7 jets	8 jets	9 jets	≥ 10 jets
$d\sigma/dN_{\text{jets}}$ [pb]	4.3	2.0	0.66	0.17	0.041	0.015	0.0046
JES	3	6	8	12	14	12	17
JER	0.2	0.6	0.4	2	3	2	4
b-Tagging (b,c)	4	4	3	3	2	2	2
Pile-up	0.7	0.4	0.2	0.2	0.9	2	3
Luminosity	3	3	3	3	3	3	3
Q^2 ($t\bar{t}$)	2	5	4	6	4	9	14
Matching ($t\bar{t}$)	0.7	2	2	6	18	17	21
LO vs. NLO	2	5	6	6	5	27	2
Top- p_T Rew.	3	5	7	8	8	10	10
W+HF-Jets CS	1	0.4	0.6	0.4	0.08	0.5	0.5
B/C in W+jets	1	0.5	0.5	0.6	0.03	0.6	1
QCD Shape	0.3	0.6	0.4	3	1	3	0.08
Syst. unc.	8	12	13	18	25	37	32
Stat. unc.	1	2	4	8	18	20	25
Total unc.	8	12	14	20	31	42	40

Table 11.4: Summary of all relative systematic uncertainties (in %) for the differential fiducial $t\bar{t}$ cross section as a function of the number of particle-level jets in the μ +jets channel. The uncertainties from: b-tagging (light), lepton efficiency, and cross section of single top, W+Light-Jets, Z+Jets, and Diboson are all $\leq 1\%$. Therefore, they are not shown in the table.

$\mu + N$ jets:	4 jets	5 jets	6 jets	7 jets	8 jets	9 jets	≥ 10 jets
$d\sigma/dN_{\text{jets}}$ [pb]	4.4	2.0	0.70	0.19	0.042	0.015	0.0055
JES	4	6	8	12	15	9	17
JER	0.3	0.2	0.7	1	1	1	6
b-Tagging (b,c)	4	4	3	3	2	2	2
Pile-up	0.9	0.2	0.4	0.3	3	0.6	3
Luminosity	3	3	3	3	3	3	3
Q^2 ($t\bar{t}$)	2	4	4	5	9	14	18
Matching ($t\bar{t}$)	0.8	2	1	5	22	31	42
LO vs. NLO	2	4	5	15	15	27	16
Top- p_T Rew.	3	5	6	7	10	9	9
W+HF-Jets CS	1	0.6	0.3	0.7	0.4	0.4	0.09
B/C in W+jets	1	0.7	0.5	0.8	0.7	0.3	0.1
QCD RelIso	0.08	0.09	0.1	0.4	1	0.6	0.9
Syst. unc.	8	11	13	22	33	45	52
Stat. unc.	1	1	3	7	17	17	21
Total unc.	8	11	13	23	37	49	56

Table 11.5: Combined differential fiducial $t\bar{t}$ cross section as a function of the number of particle-level jets in the ℓ +jets channel (summed channel). The relative statistical, systematic, and total uncertainties are also shown.

Category	$d\sigma/dN_{\text{jets}}$ [pb]	Stat. unc.	Syst. unc.	Total unc.
$t\bar{t} \rightarrow \ell + 4$ jets	8.7	0.8%	8%	8%
$t\bar{t} \rightarrow \ell + 5$ jets	4.0	1%	11%	11%
$t\bar{t} \rightarrow \ell + 6$ jets	1.4	2%	13%	13%
$t\bar{t} \rightarrow \ell + 7$ jets	0.36	5%	20%	20%
$t\bar{t} \rightarrow \ell + 8$ jets	0.083	12%	29%	31%
$t\bar{t} \rightarrow \ell + 9$ jets	0.030	13%	41%	43%
$t\bar{t} \rightarrow \ell + \geq 10$ jets	0.010	16%	42%	45%

PYTHIA, POWHEG+HERWIG, and MC@NLO+HERWIG in Fig. 11.4.

The statistical uncertainties on the various predictions are approximately equal to 0.3; 0.5; 0.9; 2; 5; 10; 30% for four to ≥ 10 jets. For clarity they are not represented in Fig. 11.4, except for illustration in the ratio plot for the predictions from MADGRAPH+PYTHIA with varied Q^2 scale and from MC@NLO³.

The comparison on the top of Fig. 11.4 shows a very good agreement between data and the nominal MADGRAPH+PYTHIA predictions between four and seven jets. This can be explained by the explicit requirement of zero to three additional emissions at the matrix-element level, which exactly sums up to four to seven jets in the ℓ +jets channel if all decay products of the $t\bar{t}$ pair have produced well measured jets. Some disagreement appears for eight jets, while this is not significant with respect to the measurement uncertainties. This corresponds to the first additional emission which is purely simulated by the parton showering. For nine and more jets the discrepancies are significant, too low cross section are predicted there.

An even better agreement is observed between the results and predictions from NLO generators POWHEG+PYTHIA and POWHEG+HERWIG. At the highest jet multiplicities, too few events are predicted by all generators, whereas POWHEG shows the smallest disagreement until ≥ 10 jets. The NLO prediction from MC@NLO+HERWIG shows too few jets. For six jets and more, the prediction from MC@NLO+HERWIG is more than

³The usage of event weights in MC@NLO makes it necessary to analyze a larger number of events in order to achieve a comparable precision on its predictions as with other generators. Still, in the two latest jet-multiplicity bins, the errors are 30 and 40%.

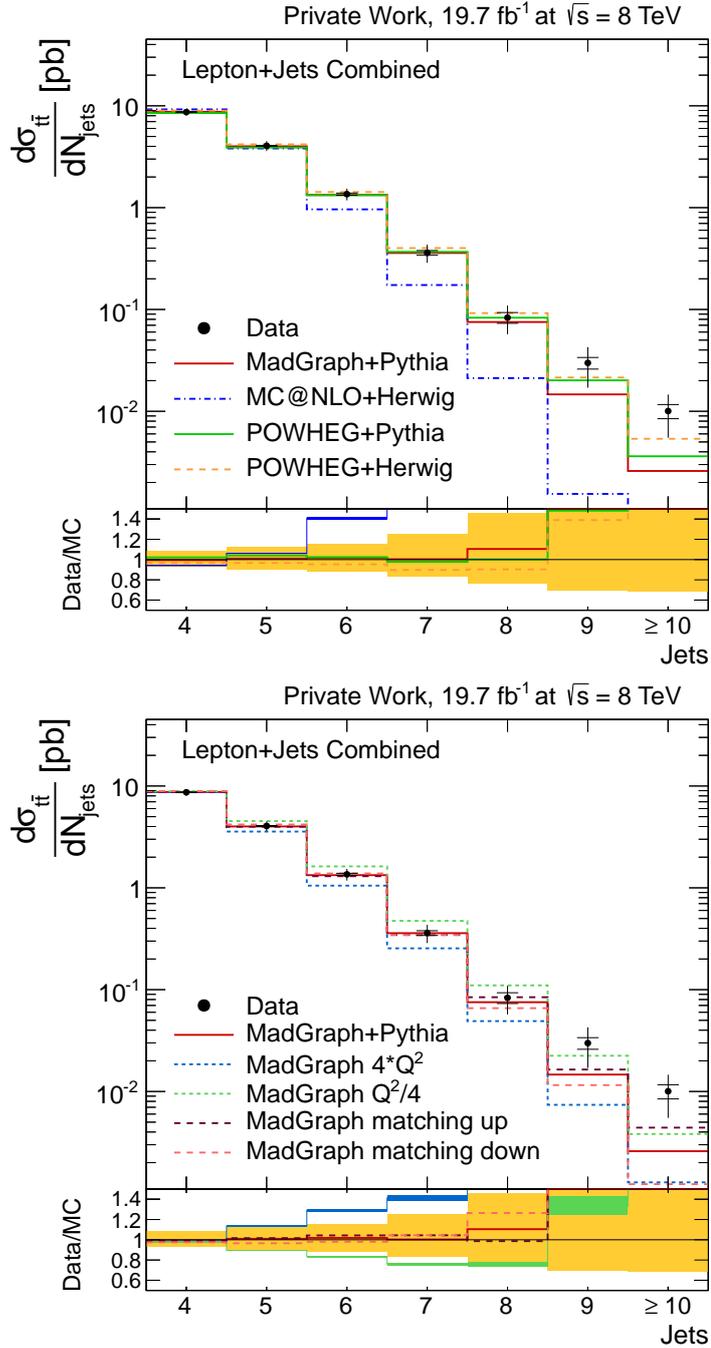


Figure 11.4: Combined differential fiducial $t\bar{t}$ cross section as a function of the number of particle-level jets in the ℓ +jets channel. The results are compared to predictions from various generators (top) or from MADGRAPH with various parameter sets (bottom). The vertical error bars represent the total uncertainties and the intersecting horizontal bars represent the statistical uncertainties alone. In the ratio plots, the yellow band represents the total measurement uncertainties. The statistical uncertainties on the predictions from MADGRAPH+PYTHIA with varied Q^2 scale and from MC@NLO are represented by the widths of the error bands.

3σ lower than the measurement. A similar agreement between the measurement and the NLO predictions would be expected, but while MC@NLO generates pure NLO including one additional emission and the virtual corrections from loops, POWHEG performs also corrections of the first emission [58–60], which seems to strongly improve the prediction of the jet multiplicity. As mentioned in Sec. 11.6, a parameter of POWHEG, h_{damp} , plays a role in the hardest emission at matrix-element and might be related to the better predictions from POWHEG than MC@NLO.

The results are also compared in Fig. 11.4 to predictions from MADGRAPH+PYTHIA with varied parameters: the factorization and renormalization scale (Q^2) is varied with factors 4 and 0.25 (equivalent to factors 2 and 0.5 applied on Q of Sec. 11.3), the jet-energy matching threshold between the matrix-element and the parton showering (matching) is varied up and down (with factor 2 and 0.5). The variation of the matching threshold has a significant impact on the predictions from MADGRAPH+PYTHIA in events with at least eight jets, but the variation of the Q^2 scale has a larger effect starting at 5-jet events. For both parameters, a modification of the shape of the predicted jet-multiplicity spectrum is observed. Increasing the matching threshold improves the agreement with the measurement at largest jet multiplicity (≥ 8 jets). Decreasing the matching threshold makes the agreement with data worse than in the nominal case at high jet multiplicity (≥ 8 jets). Increasing the Q^2 scale reduces the probability of additional emissions. Hence, the predicted spectrum is shifted to lower jet multiplicities: the prediction disagrees with the measurement for events with at least five jets. Decreasing the Q^2 scale shows as expected the opposite effect for events with four to seven jets: a decreased Q^2 -scale enables more additional emissions. Nevertheless, the effect is too strong since too many events with five to eight jets are observed. Then, the trend is reversed for events with ≥ 9 jets, where the additional-jet production is too weak, so decreasing the Q^2 scale does not support enough the radiation of further jets by the parton showering. To summarize the discussions about the predictions with decrease Q^2 scale, one could be tempted to state that a lower Q^2 scale in MADGRAPH+PYTHIA can improve the description of the jet-multiplicity spectrum because further jet emissions are supported. However, this variation only affects the lower jet-multiplicity events where the matrix-element emissions play a role. At larger jet multiplicities, where further emissions are only generated by the parton showering, the further-emission rate is too low, as with the nominal Q^2 scale value. Moreover, looking only at the jet-multiplicity description, one cannot judge the general description of $t\bar{t}$ events for a particular choice of Q^2 scale.

Finally, the difference between the predicted jet-multiplicity spectrum from MADGRAPH+PYTHIA with $4 \cdot Q^2$ and $Q^2/4$ scale variation is significantly larger than the measurement uncertainties. This cannot bring any statement on the uncertainty on the theoretical Q^2 scale⁴ since this scale is not an observable that can be measured in any process. Nevertheless, the uncertainty on the number of jets produced in $t\bar{t}$ events at the LHC does not have to be taken from the theoretical Q^2 -scale uncertainty since, as it is shown with the present measurement, a dedicated measurement can provide better precision. This can be useful for further analyses or for calibration purposes in the field of top-quark physics.

11.5 Comparison with Predictions from Further MC Tools

The implementation of this analysis in RIVET enables comparison of the present measurement to further MC tools. The particle-level selection and the calculation of the differential fiducial cross section described previously in this chapter can be performed on any MC simulation. The condition is that $t\bar{t}$ production is simulated in proton-proton collisions at $\sqrt{s} = 8$ TeV and the event record is stored in the HepMC format [151]. MC simulation samples of $t\bar{t}$ events have been generated with automated tools which are intended to be utilized for the Run 2 of the LHC. Their predictions are compared in Fig. 11.5 to the results presented in the previous section.

- CMS-Official MC simulation from aMC@NLO. The generation of $t\bar{t}$ events with zero, one, and two additional emissions has been performed exclusively and with NLO precision, hence leading to up to three additional emissions. These events have been merged using the FxFx prescription [53]. The parton-showering and hadronization steps have been performed with PYTHIA 8.
- Local SHERPA v2.1.1 production by Ralf Farkas and Marco Harrendorf. Two different configurations have been utilized:
 - $t\bar{t}$ with zero and one additional emission at NLO. Two and three additional emissions are also included but with LO precision (called here Sherpa0/1NLO).

⁴The experimental community would likewise be surprised by a statement from the theory community like: “Your jet-energy scale uncertainty is over-estimated”.

- $t\bar{t}$ with zero additional emission at NLO. One, two, and three additional emissions are also included but with LO precision (called here Sherpa0NLO).

The interface between the matrix-element and the parton-showering steps is achieved in SHERPA with an improved version of the CKKW algorithm [152]. The merging between the various multiplicities of additional emissions is done with the MEPS@NLO technique [153] in SHERPA.

The precision of the MC simulation with both configurations is also tested with variations of the Q^2 scale with factors of 4 and 1/4, where a Q^2 -scale dependency lower than at LO is expected.

Two million events have been generated for each SHERPA production.

First, the predictions from aMC@NLO+PYTHIA 8 show a jet-multiplicity spectrum with quite good agreement with data up to nine jets. In general, slightly too many jets are produced, while the effect is mostly significant for seven jets, corresponding to the highest jet multiplicity in which the matrix-element simulation is contributing. While a lower Q^2 -scale dependency is expected for aMC@NLO+PYTHIA 8 compared to LO tools, this cannot be tested since no MC-simulation samples including Q^2 -scale variations have been delivered in time for this thesis.

Predictions from both SHERPA configurations without Q^2 -scale variations (nominal), shown in the top of Fig. 11.5, are in good agreement with each other and with the data according to the statistical uncertainties on the predictions and the total measurement uncertainties, illustrated with the various error bands in the ratio plot. Due to the large event weights found in events generated by SHERPA, large statistical uncertainties on the predictions from SHERPA appear. Nevertheless, it is possible to conclude that the Sherpa0NLO configuration achieves a better prediction of the jet multiplicity than the Sherpa0/1NLO configuration. This judgment is based on the comparison for seven jets, in which the prediction from Sherpa0/1NLO drops with 1σ significance with respect to the measurement, while the predictions from the Sherpa0NLO configuration are very good up to eight jets.

Observing the bottom part of Fig. 11.5 where predictions from SHERPA with varied Q^2 scale are displayed, the same large statistical uncertainties arise, as for the nominal SHERPA predictions. These uncertainties might cause statistical fluctuations, explaining why the nominal predictions from SHERPA are not always located between those with up and downward-varied Q^2 scale, as would be expected. Still, it is possible to observe that the predictions up to six jets from the Sherpa0/1NLO configuration and up to seven jets from

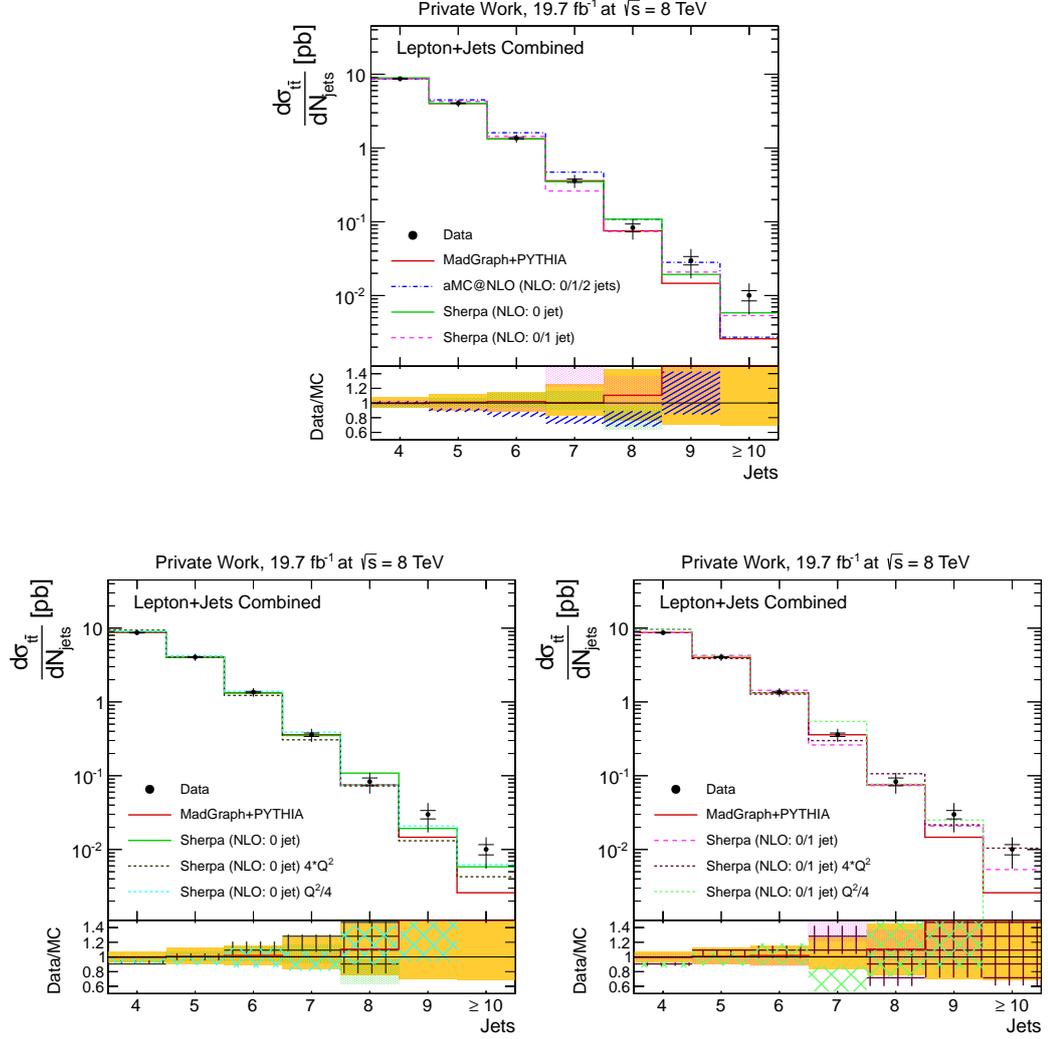


Figure 11.5: Combined differential fiducial $t\bar{t}$ cross section as a function of the number of particle-level jets in the ℓ +jets channel. The vertical error bars represent the total uncertainties and the intersecting horizontal bars represent the statistical uncertainties alone. Top: the results are compared to predictions from various generators: MADGRAPH, aMC@NLO, and Sherpa with 0 and with 0/1 additional emission at NLO. Bottom: the results are compared to predictions from Sherpa without additional emission at NLO (left) and with zero and one additional emission at NLO (right) with varied Q^2 scale. In the ratio frames, the dotted, straight-crossed, and oblique-crossed areas represent statistical uncertainties on the prediction from the nominal, the up, and down variations of the Q^2 scale for the SHERPA predictions. The statistical uncertainties on predictions from aMC@NLO are represented with oblique lines.

the Sherpa0NLO configuration are almost fully compatible with the data, according to the measurement uncertainties. In Fig 11.4, the predictions from MADGRAPH with Q^2 -scale variations are not compatible with measurements for six and seven jets. This shows that predictions from SHERPA, thanks to NLO precision, are less sensitive to the Q^2 -scale uncertainty than LO predictions from MADGRAPH, as expected.

Interestingly, the discrepancy observed for seven jets between the measurement and the prediction from the nominal Sherpa0/1NLO configuration (top of Fig 11.5) is reproduced with increased Q^2 scale (bottom right of the same figure), while the opposite effect happens in the case of the decreased Q^2 scale. This could indicate a wrong setting of the Sherpa0/1NLO configuration. This could also be due to a misbehavior of SHERPA when dealing with the interface between the highest matrix-element emission multiplicity, six jets⁵, and the first further emission origination solely from the parton showering, taking place for seven jets.

11.6 Comparison with Earlier Measurements

Comparing the results of this chapter to the previous results using proton-proton collision data at $\sqrt{s} = 7$ TeV [35], summarized in Sec. 6, the agreement between data and predictions from MADGRAPH+PYTHIA and POWHEG+PYTHIA up to seven jets is compatible. For eight-jet events, the predicted cross section at $\sqrt{s} = 7$ TeV is larger than measured. This is the opposite tendency to the observations at $\sqrt{s} = 8$ TeV, given in Sec. 11.4. A possible explanation is the statistical uncertainty of the predictions for eight-jet events, since smaller MC simulation samples were produced at $\sqrt{s} = 7$ TeV data. The conclusion about MC@NLO+HERWIG is compatible: too few events with large jet multiplicities are produced by this generator. The statistical uncertainties are smaller in the 8 TeV measurement because more data was recorded. Accordingly, larger MC-simulation samples were generated for the 8 TeV analysis. This reduces the uncertainty from the correction of detector effects and the impact of statistical fluctuations on the estimation of systematic uncertainties relying on dedicated $t\bar{t}$ MC simulation (Q^2 scale and matching threshold).

The results from ATLAS at $\sqrt{s} = 7$ TeV [34] cannot be exactly compared with the results from the present measurement because several details of

⁵Four jets from the $t\bar{t}$ decay products + one real additional emission from NLO + one exclusive emission from the $t\bar{t}+1$ jet process = six jets.

the analysis setup are different (e.g. normalization of the result, jet clustering parameter R , and jet p_T cut) but the agreement between predictions and data can be compared. In Fig. 11.6, the differential fiducial cross section as a function of the number of particle-level jets measured by ATLAS is shown. The results from the e +jets and μ +jets channels are averaged in a single result. This is compared to predictions from NLO generators: POWHEG+PYTHIA, POWHEG+HERWIG, and MC@NLO+HERWIG, and LO generator ALPGEN interfaced with PYTHIA or HERWIG, and with varied parameters of these generators. For an alternative prediction from POWHEG+PYTHIA, the h_{damp} parameter is modified to 172.5 GeV instead of ∞ , the default parameter value⁶. This parameter “[...] effectively regulates the high- p_T radiation in POWHEG [...]” [34]. See [58–60] for technical precisions. The MC simulation sample from ALPGEN was configured to produce up to five additional hard partons. For the ALPGEN+HERWIG MC simulation sample, separate simulations of $t\bar{t}+b\bar{b}$ and $t\bar{t}+c\bar{c}$ events were produced and merged to the nominal $t\bar{t}$ simulation in order to have enough statistical precision in the heavy-flavor range. The Q^2 -scale up and down variations used by CMS correspond to the opposite variations of the α_s parameter utilized by ATLAS for the predictions from ALPGEN+PYTHIA.

As also observed with data from CMS at $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV, MC@NLO +HERWIG produces too few jets in the ATLAS analysis.

A better agreement between data and the prediction from ALPGEN+HERWIG than from ALPGEN+PYTHIA is observed for seven and eight jets. This can be due to the special treatment of the heavy-flavor events in the MC simulation sample from ALPGEN+HERWIG, mentioned previously. However, an effect from using PYTHIA or HERWIG cannot be excluded.

The “tuning” applied in POWHEG+PYTHIA improves the description of the jet multiplicity significantly because its predictions agree with data for the whole measurement range, whereas the prediction from the nominally configured POWHEG+PYTHIA shows discrepancies for eight jets. While the latter observation corresponds to those from CMS at $\sqrt{s} = 7$ TeV, this slightly contradicts the observations from this thesis. As shown in Fig. 11.4, the predictions from POWHEG+PYTHIA for eight jets is in very good agreement with the measurement at 8 TeV, whereas for higher jet multiplicities the predictions are lower than the data. A higher precision is achieved with 8 TeV data and in a wider range.

⁶Infinity is the value assumed to have been used for the MC simulation from POWHEG+PYTHIA provided by the CMS collaboration, and hence used in this thesis.

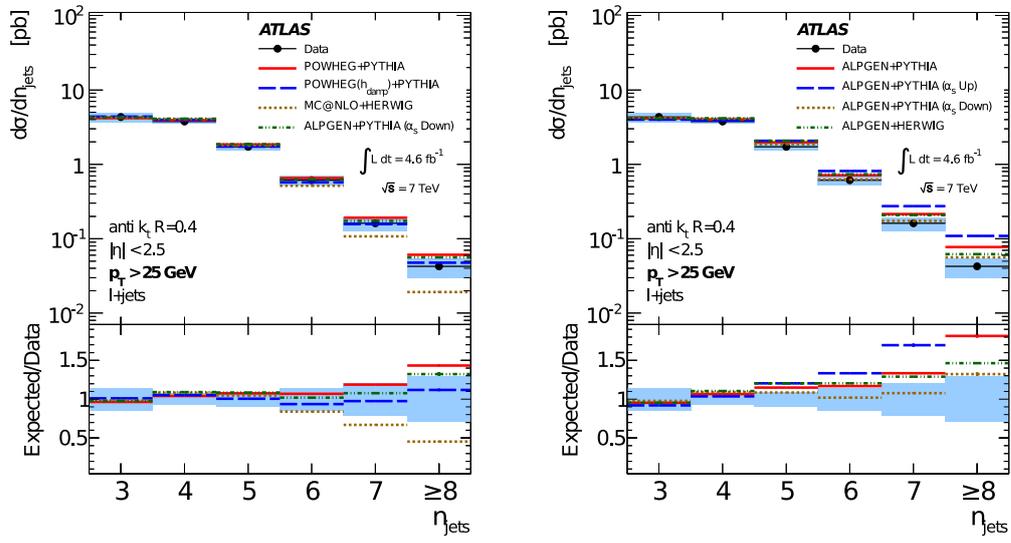


Figure 11.6: Differential fiducial $t\bar{t}$ cross section as a function of the number of particle-level jets measured with data from proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector, taken from [34]. The results are compared to predictions from NLO generators or ALPGEN (LO) with decreased α_s (left) and LO generator ALPGEN with various values of α_s and parton-showering/hadronization tools used (right).

Chapter 12

Conclusion and Outlook

A precision measurement of top-quark production has been performed with data delivered by the LHC and recorded by the CMS experiment in 2012. Top-quark pair events decaying in the ℓ +jets channel have been analyzed and the results have been compared to predictions.

A summary of a former publication, presenting the measurement of the normalized differential fiducial cross section of the $t\bar{t}$ production as a function of the number of jets with data delivered in 2011 by the LHC at $\sqrt{s} = 7$ TeV, has been given. In order to disentangle the measurement from jets expected from the decay products of $t\bar{t}$ pairs, a more sophisticated differential measurement has also been included in this publication, and has been summarized in this thesis too: the $t\bar{t}$ -production differential cross section as a function of the number of additional jets. Several wishes for improvements of this analysis have been raised:

- The normalization of the differential cross section has been applied since it reduces systematic uncertainties related to normalization uncertainties. However, this also biases comparisons to predictions because it artificially improves the agreement for those jet multiplicities with the highest cross sections.
- Making use of parton-level information, i.e. color-charged particles recorded in the MC simulation before they were showered and hadronized, is very convenient when performing a measurement, because it gives access to a theoretical interpretation of the results. However, the definition of such particles is not physical because they cannot be observed and their implementation in MC simulations depends on models.

These points have been taken into account for the differential fiducial cross-section measurement of the top-quark pair production as a function of the number of jets with data delivered in 2012 by the LHC at a center-of-mass energy of $\sqrt{s} = 8$ TeV, which is the main focus of this thesis. Further improvements with respect to the measurement at 7 TeV have been realized: the RIVET analysis tool has been utilized in order to facilitate the use of the results for further investigations in groups of the particle-physics community and the data-driven background estimation has been improved. The main results of this thesis are represented in Fig. 12.1.

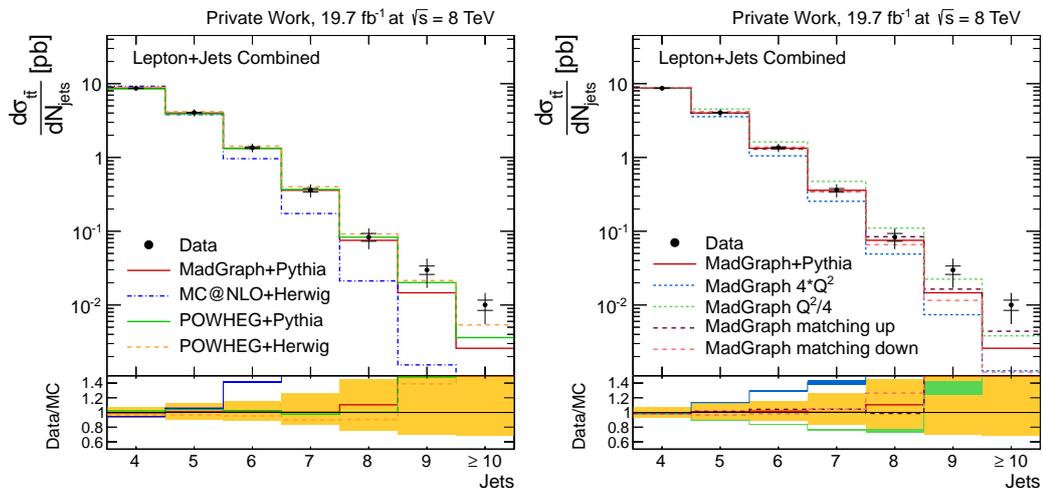


Figure 12.1: Combined differential fiducial $t\bar{t}$ cross section as a function of the number of jets in the ℓ +jets channel. The results are compared to predictions from various generators (left) or from MADGRAPH with various parameter sets (right).

Uncertainties have decreased between 7 and 8 TeV analyses for the highest jet multiplicities (≥ 6 jets). This is not only due to the smaller statistical uncertainties, but also due to better estimation of systematic effects, thanks to larger amounts of data and MC-simulation events.

The predictions from the standard $t\bar{t}$ MC simulation, produced with MADGRAPH+PYTHIA achieves a very good description of the jet-multiplicity spectrum up to eight jets. This is made possible, in spite of LO precision only, by the simulation of additional emissions (up to three in this case) in MADGRAPH, achieving an approximation of higher orders. For further jet multiplicities, the predicted cross section does not agree well with the measurement, showing that the parton showering is not well suited for the generation of additional jets. A good description is achieved by POWHEG+PYTHIA and POWHEG+HERWIG, while it is better than

MADGRAPH+PYTHIA for the highest jet multiplicities. MC@NLO+HERWIG predicts significantly too few jets. The disagreement is larger than 3σ for six jets and gets even worse for higher jet multiplicities.

The results are also compared to predictions from MADGRAPH+PYTHIA where the Q^2 scale and the matching threshold have been varied in both directions. The effects of the Q^2 -scale uncertainty are the largest, even larger than the measurement uncertainties. This is a useful observation for the experimental community, since it shows that $t\bar{t}$ events from data can provide precision for future analysis for which the $t\bar{t}$ production is a background and for detector-calibration purpose. However, this scale is not intended to be measured in data.

The results have also been compared to predictions from recent NLO tools featuring the simulation of multiple emission multiplicities. A first MC-simulation sample has been produced by the CMS collaboration making use of aMC@NLO+PYTHIA 8. The resulting predictions agree quite well with the results up to high jet multiplicities.

Two further $t\bar{t}$ MC-simulation samples have been produced at KIT with SHERPA, simulating $t\bar{t}$ with zero or zero and one additional emission at NLO precision, completed with up to three emissions at LO precision. Due to large statistical uncertainties, conclusions can be drawn for low jet multiplicities only, where good agreement with the measured jet multiplicity is observed. The configuration of SHERPA with zero additional emissions at NLO provides a better description of the jet-multiplicity spectrum than having one additional emission also simulated at NLO. Moreover, the Q^2 -scale dependency of the NLO MC simulation by SHERPA has been tested, varying the Q^2 scale. As expected, NLO MC simulation is less sensitive to Q^2 -scale uncertainties than LO MC simulation provided by MADGRAPH.

The conclusions from the jet-multiplicity measurements with data at 7 and 8 TeV agree at low jet multiplicities. The comparisons to predictions in the highest jet multiplicities (eight jets and more) are different but this could be explained by statistical fluctuations in predictions in the 7 TeV comparisons. The measurement of the differential cross section as a function of the number of additional jets has only been performed with 7 TeV data because the challenging conversion of this analysis to particle-level would have required more time. However, the conclusions from this analysis with 7 TeV data were compatible with the jet-multiplicity measurement. No better precision could be achieved by the additional-jet measurement, despite the higher sophistication of the measurement. Therefore, this was not seen as a priority for the 8 TeV analysis.

Comparisons of the results to those obtained by the ATLAS collaboration with 7 TeV data have been discussed. Conclusions arise which are compatible with those provided by the 7 TeV analysis summarized in this thesis. ATLAS has also performed a $t\bar{t}$ MC simulation using POWHEG with a modification of a parameter related to the additional emissions. This seems to improve the predictions from POWHEG significantly.

As a conclusion to this thesis, the investigation of $t\bar{t}$ events can be performed very accurately with the data delivered by the LHC in 2012 at 8 TeV because a large amount of $t\bar{t}$ events could be selected with a minimal contribution from background processes. The $t\bar{t}$ events could be investigated up to a very high multiplicity of jets, which drives the test of SM predictions at a high level of precision. It has been shown that LO MC tools, now in principle superseded by NLO tools, can perform a very good description of high-multiplicity processes. A drawback of the LO is its sensitivity to scale uncertainty. Modern MC-simulation tools with NLO precision have been tested for the high-multiplicity region of the $t\bar{t}$ processes and fairly good modeling could be observed. This work brings confidence that searches for high-multiplicity processes, likely to be contaminated by $t\bar{t}$ events, can be performed at LHC since it has shown that the SM predictions can describe data from the LHC at 8 TeV very accurately.

The next steps of this analysis would be the development of the measurement of the differential cross section as a function of the number of additional jets. The identification of the additional jets in the MC simulation without any help from the parton-level information should be possible by means of a full event reconstruction with assignment of jets to $t\bar{t}$ decay products. The analysis of jets in $t\bar{t}$ production could be complemented with an interpretation of the results making use of parton-level information. A publication is in preparation in collaboration with a group from the Demokritos Institute in Athens, Greece. It is planned to publish the results obtained at 8 TeV, presented in this thesis, since they achieve a better precision than the previous publication.

By the consistent use of model-independent definitions for the measurement, the results are intended to provide useful and reliable knowledge for future investigations of the top-quark properties, e.g. with data taking at 13 TeV¹.

¹The 13 TeV run started at the LHC on the day when these lines were written.

Part III

Appendix and Bibliography

Appendix A

Technical List of Samples

Tables A.1 and A.2 show the MC-simulation samples with their names within the CMS collaboration. The corresponding data sets are given in Tables A.3 and A.4.

Table A.1: List of MC-simulation samples used, all processed with PU assuming a peak instantaneous luminosity of $7.4 \cdot 10^{33}$ [121]. Almost all samples come from Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM. $t\bar{t}$ comes from Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM. QCD multijet muon enriched comes from both v1 and v3 and QCD multijet for BcToE processes comes from v2.

Physical Process	Primary Dataset Name	Cross-Section [pb]	Ref.
$t\bar{t}$	/TTJets_MSDecays_central_TuneZ2star_8TeV-madgraph-tauola/	252.89	[24]
t in t channel	/T_t-channel_TuneZ2star_8TeV-powheg-tauola/	56.4	[123]
\bar{t} in t channel	/Tbar_t-channel_TuneZ2star_8TeV-powheg-tauola/	30.7	[123]
t in tW channel	/T_tW-channel-DR_TuneZ2star_8TeV-powheg-tauola/	11.1	[123]
\bar{t} in tW channel	/Tbar_tW-channel-DR_TuneZ2star_8TeV-powheg-tauola/	11.1	[123]
t in s channel	/T_s-channel_TuneZ2star_8TeV-powheg-tauola/	3.79	[123]
\bar{t} in s channel	/Tbar_s-channel_TuneZ2star_8TeV-powheg-tauola/	1.76	[123]
$W + 2$ jets	/W2JetsToLNu_TuneZ2star_8TeV-madgraph-tauola/	2159.20	[124]
$W + 3$ jets	/W3JetsToLNu_TuneZ2star_8TeV-madgraph-tauola/	640.40	[124]
$W + 4$ jets	/W4JetsToLNu_TuneZ2star_8TeV-madgraph-tauola/	264.00	[124]
$Z/\gamma^* +$ jets	/DYJetsToLL_M-50_TuneZ2star_8TeV-madgraph-tauola/	3503.71	[124]
WW	/WW_TuneZ2star_8TeV_pythia6_tauola/	54.83	[125]
WZ	/WZ_TuneZ2star_8TeV_pythia6_tauola/	33.21	[125]
ZZ	/ZZ_TuneZ2star_8TeV_pythia6_tauola/	8.05	[125]
QCD muon enriched	/QCD_Pt_20_MuonEnrichedPt_15_TuneZ2star_8TeV_pythia6/	134680.0	Pythia6
QCD BC to E (20 to 30 GeV)	/QCD_Pt_20_30_BCtoE_TuneZ2star_8TeV_pythia6/	167388.0	Pythia6
QCD BC to E (30 to 80 GeV)	/QCD_Pt_30_80_BCtoE_TuneZ2star_8TeV_pythia6/	167040.0	Pythia6
QCD BC to E (80 to 170 GeV)	/QCD_Pt_80_170_BCtoE_TuneZ2star_8TeV_pythia6/	12981.9	Pythia6
QCD BC to E (170 to 250 GeV)	/QCD_Pt_170_250_BCtoE_TuneZ2star_8TeV_pythia6/	620	Pythia6
QCD BC to E (250 to 350 GeV)	/QCD_Pt_250_350_BCtoE_TuneZ2star_8TeV_pythia6/	103.2	Pythia6
QCD BC to E (≥ 350 GeV)	/QCD_Pt_350_BCtoE_TuneZ2star_8TeV_pythia6/	23.49	Pythia6
QCD EM enriched (20 to 30 GeV)	/QCD_Pt_20_30_EMEnriched_TuneZ2star_8TeV_pythia6/	29148.6	Pythia6
QCD EM enriched (30 to 80 GeV)	/QCD_Pt_30_80_EMEnriched_TuneZ2star_8TeV_pythia6/	4615893.0	Pythia6
QCD EM enriched (80 to 170 GeV)	/QCD_Pt_80_170_EMEnriched_TuneZ2star_8TeV_pythia6/	183294.9	Pythia6
QCD EM enriched (170 to 250 GeV)	/QCD_Pt_170_250_EMEnriched_TuneZ2star_8TeV_pythia6/	4650	Pythia6
QCD EM enriched (250 to 350 GeV)	/QCD_Pt_250_350_EMEnriched_TuneZ2star_8TeV_pythia6/	559	Pythia6
QCD EM enriched (≥ 350 GeV)	/QCD_Pt_350_EMEnriched_TuneZ2star_8TeV_pythia6/	89.1	Pythia6

Table A.2: MC-simulation samples used for the estimation of the the modeling uncertainties. “*” located in the primary dataset name replaces “Summer12_DR53X-PU_S10_START53_V19”. “**” located in the primary dataset name replaces “Summer12_DR53X-PU_S10_START53_V7A”

Dataset Description		Primary Dataset Name
$t\bar{t}$ Q^2 scale	up	/TTJets_MSDecays_scaleup_TuneZ2star_8TeV-madgraph-tauola/*-v1
	down	/TTJets_MSDecays_scaledown_TuneZ2star_8TeV-madgraph-tauola/*-v1
$t\bar{t}$ jet matching	up	/TTJets_MSDecays_matchingup_TuneZ2star_8TeV-madgraph-tauola/*-v2
	down	/TTJets_MSDecays_matchingdown_TuneZ2star_8TeV-madgraph-tauola/*-v2
$t\bar{t}$ POWHEG+PYTHIA		/TT_CT10_TuneZ2star_8TeV-powheg-tauola/**-v2
$t\bar{t}$ POWHEG+HERWIG		/TT_CT10_AUET2_8TeV-powheg-herwig/*-v1
$t\bar{t}$ MC@NLO+HERWIG		/TT_8TeV-mcatnlo/**-v1

Table A.3: Summary of data datasets used in the electron+jets analysis.

Dataset Description	Dataset Name
Run2012A	/SingleElectron/Run2012A-22Jan2013-v1
Run2011B	/SingleElectron/Run2012B-22Jan2013-v1
Run2011C	/SingleElectron/Run2012C-22Jan2013-v1
Run2011D	/SingleElectron/Run2012D-22Jan2013-v1

Table A.4: Summary of data datasets used in the muon+jets analysis.

Dataset Description	Dataset Name
Run2012A	/SingleMu/Run2012A-22Jan2013-v1
Run2011B	/SingleMu/Run2012B-22Jan2013-v1
Run2011C	/SingleMu/Run2012C-22Jan2013-v1
Run2011D	/SingleMu/Run2012D-22Jan2013-v1

Appendix B

Control plots including systematic uncertainties

The p_T distribution of all selected jets in selected events in the ≥ 2 -tag region is shown in Fig. B.1. It includes the Q^2 scale uncertainties in the $t\bar{t}$ MC-simulation sample and the jet-energy scale uncertainties in all MC simulations. The discrepancies are covered by these systematic uncertainties.

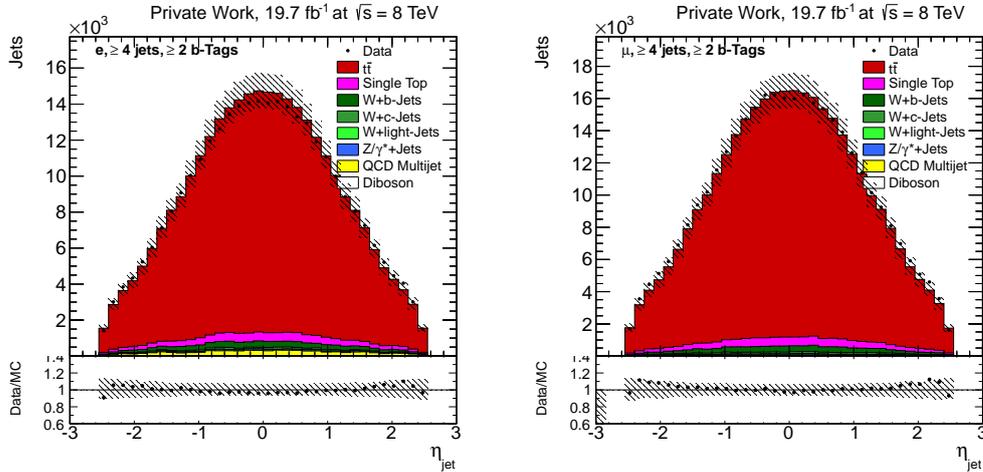


Figure B.1: Distribution of the p_T of all selected jets in the e +jets channel (left) and μ +jets channel (right). The error bars on the data points represent the statistical uncertainties from the data, whereas the hashed areas in the histogram and in the ratio plot represent the sum of the systematic uncertainties on the MC simulation (Q^2 scale in the $t\bar{t}$ simulation and jet-energy scale in all MC simulations).

Appendix C

Control Plots

Discussion of observed discrepancies:

- Fig. C.1 and C.6: the b -tagging discriminator distribution shows an expected jump at ≈ 0.68 due to the b -tagging cut applied on jets. The agreement is bad because a single-working-point correction is applied, instead of a shape correction. This is allowed since no correct shape description is required in this analysis, only the event yield counts. By construction, the QCD-multijet model in the μ +jets channel has a R_{eff} between 0.3 and 1.0 in order to be extracted from data. This is visible on these histograms.
- Fig. C.2 and C.7: in both channels, M3, the invariant mass distribution of the jet triplet providing the highest p_T shows small discrepancies for 1-tag events. Especially, absolute discrepancies are visible around the top-quark mass peak. The latter is resolved in ≥ 2 -tag events whereas a slope appears between 100 and 600 GeV. This is a collateral effect of the top- p_T reweighting applied on the $t\bar{t}$ MC simulation, as described in Sec. 8.3. The top- p_T is correlated to the top mass and M3, hence this correction makes the M3 distribution softer in the $t\bar{t}$ MC simulation.
- Fig. C.3 to C.5 and C.8 to C.10: discrepancies are observed, similarly to those discussed in the main text, in Sec. 10. The same discussion is valid: systematic uncertainties cover the discrepancies, as shown in Fig. B.1 from Appendix B.

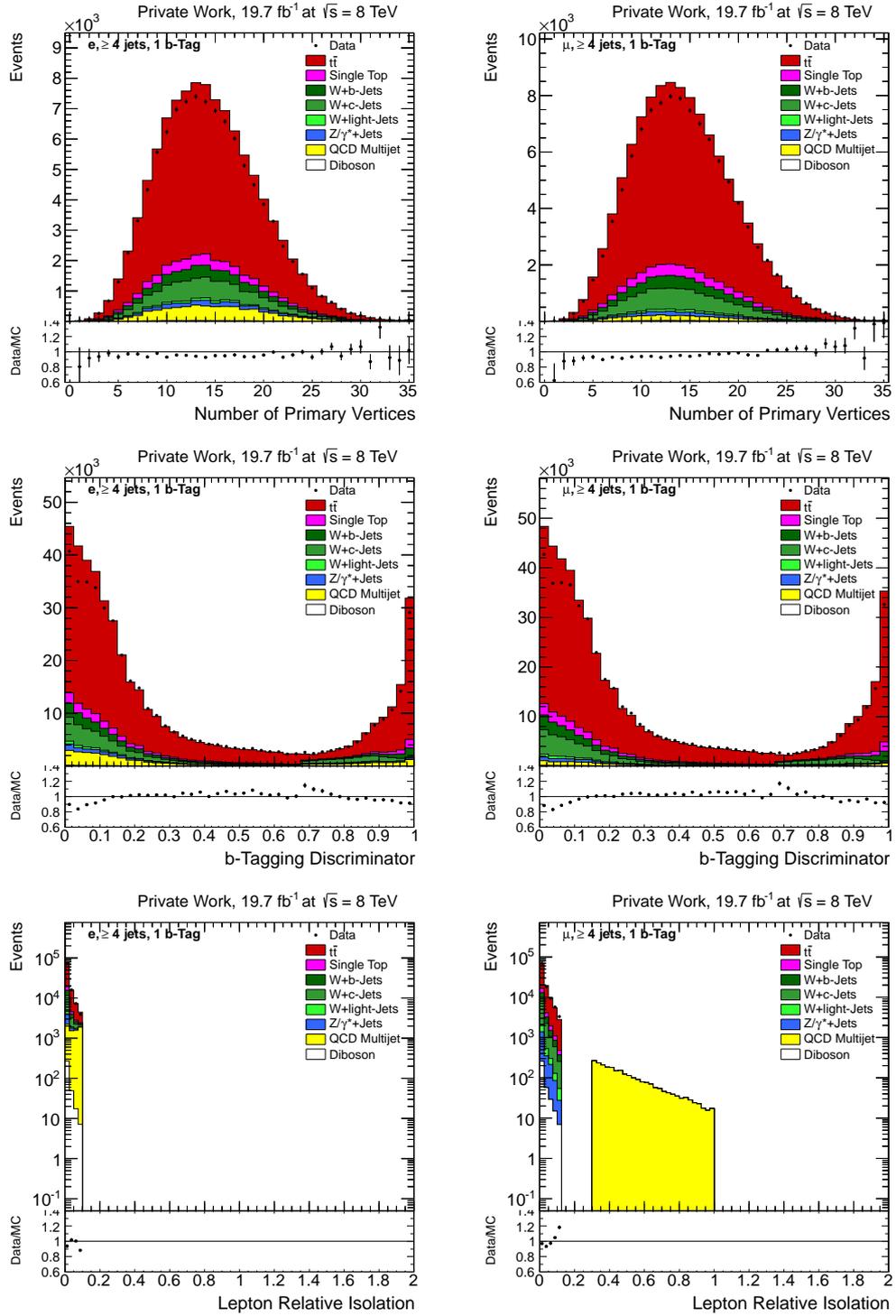


Figure C.1: Distribution of the number of reconstructed primary vertices (top), the b -tagging discriminator for all selected jets (middle), and the relative isolation of this lepton (bottom), in the e +jets channel (left) and μ +jets channel (right) for 1-tag events.

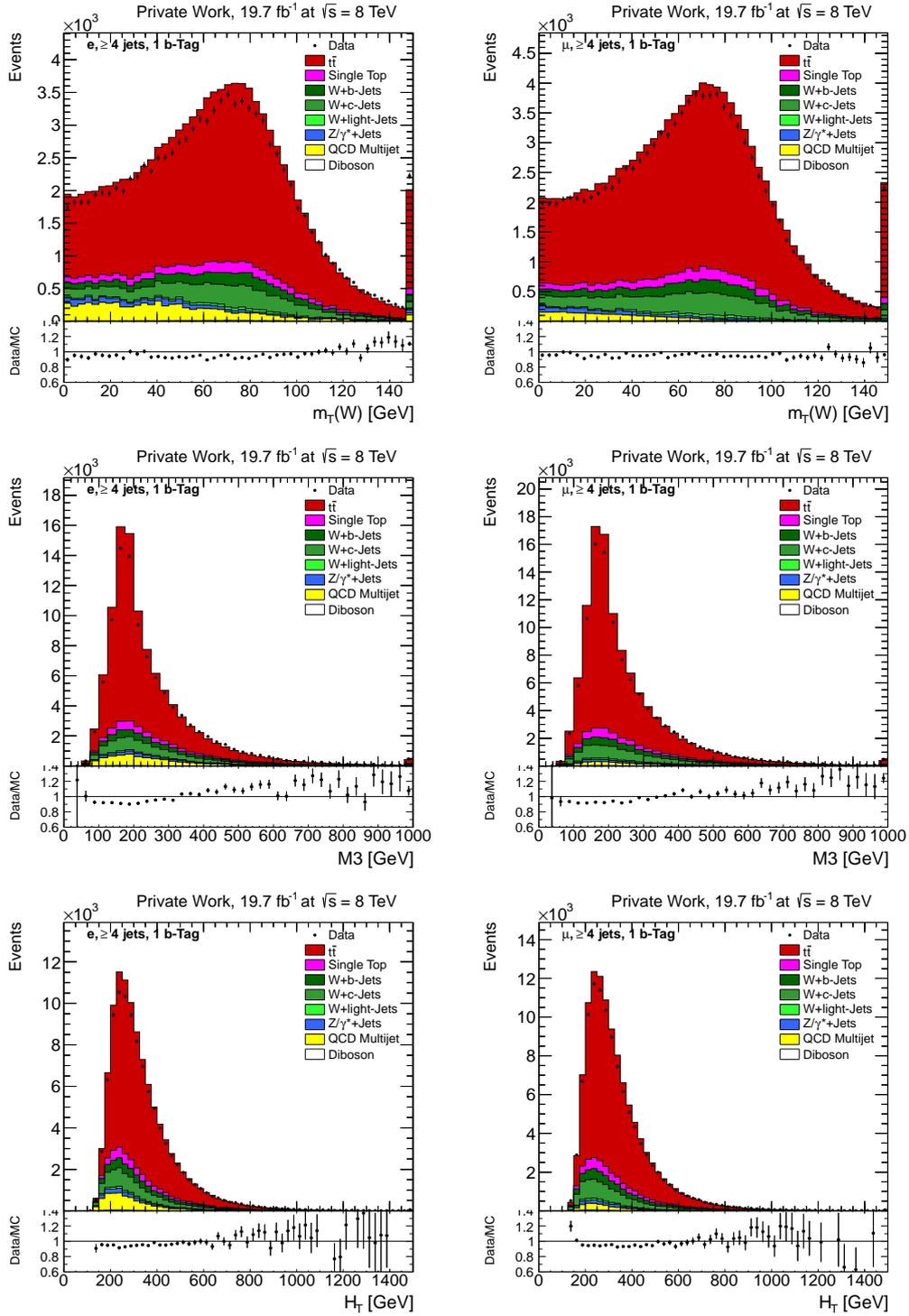


Figure C.2: Distribution of $M_T(W)$ the W -boson transverse invariant mass using the p_T of the selected lepton and the measured \cancel{E}_T (top), $M3$, the invariant mass of the hardest three-jet vectorial sum (middle), and H_T , the p_T scalar sum of all selected jets (bottom) in the e +jets channel (left) and μ +jets channel (right) for 1-tag events.

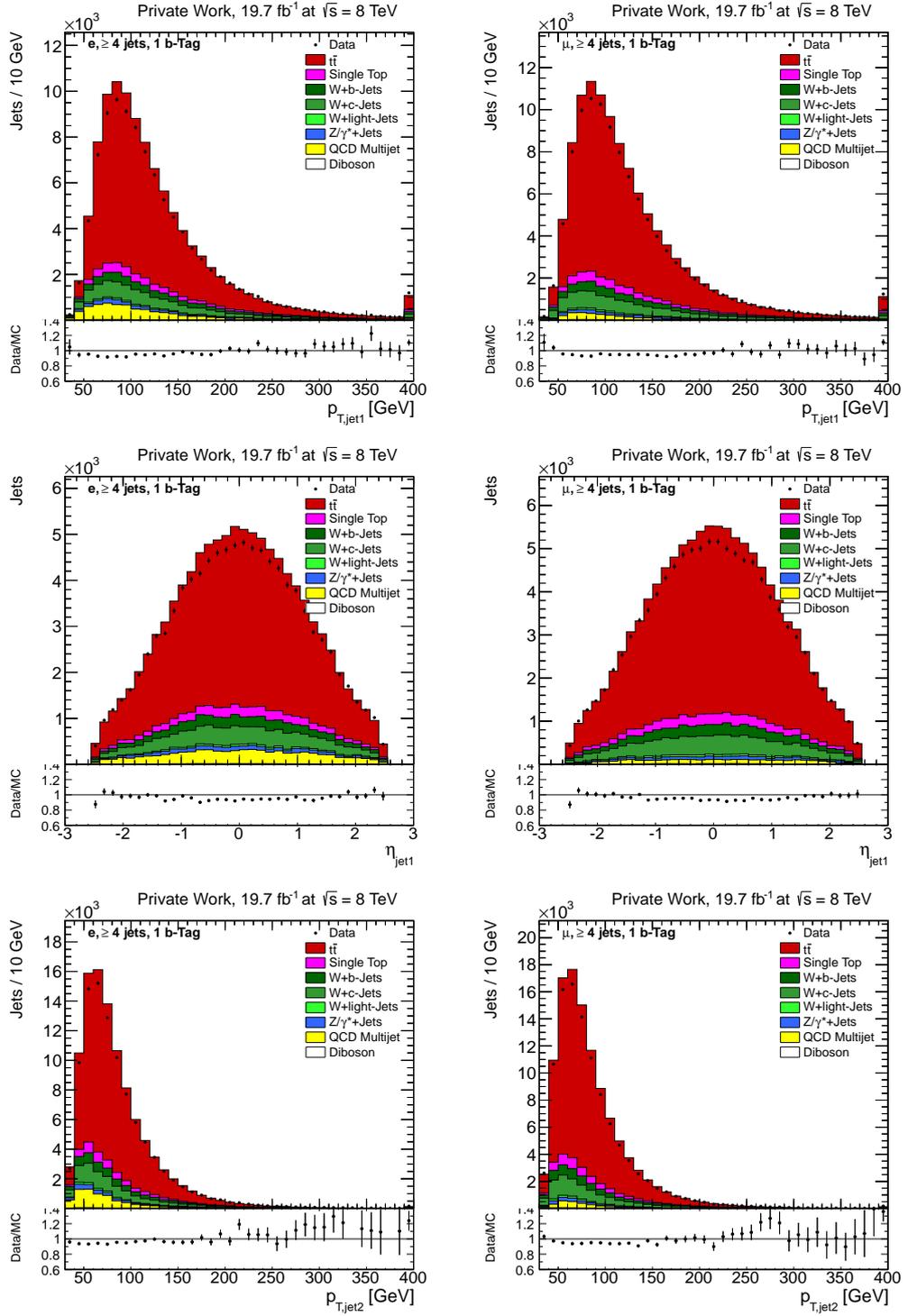


Figure C.3: Distribution of the p_T the hardest selected jet (top), η of this jet (middle), and the p_T of the second-hardest selected jet (bottom), in the $e + \text{jets}$ channel (left) and $\mu + \text{jets}$ channel (right) for 1-tag events.

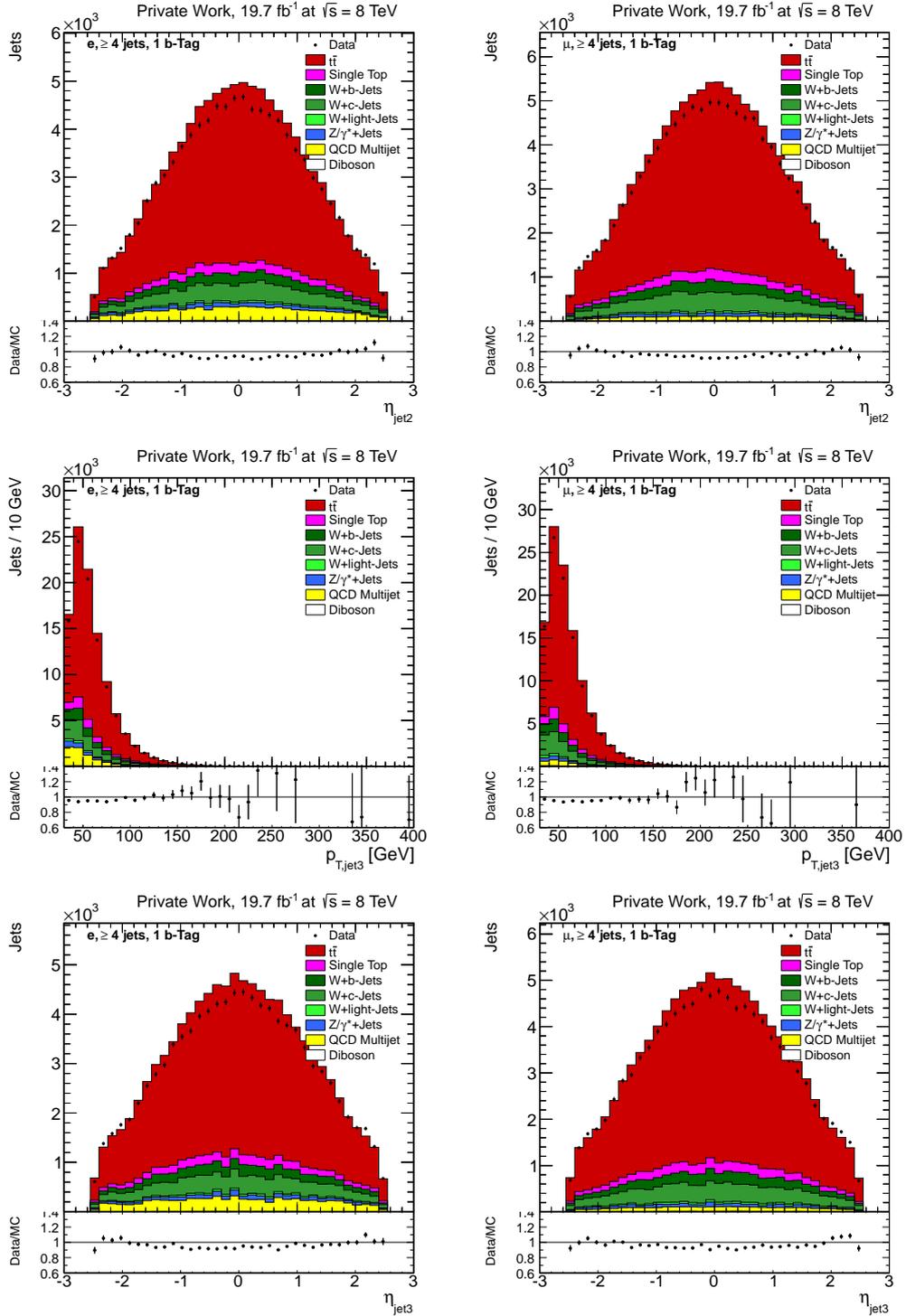


Figure C.4: Distribution of η of the second-hardest selected jet (top), the p_T of the third-hardest selected jet (middle), and η of this jet (bottom), in the e +jets channel (left) and μ +jets channel (right) for 1-tag events.

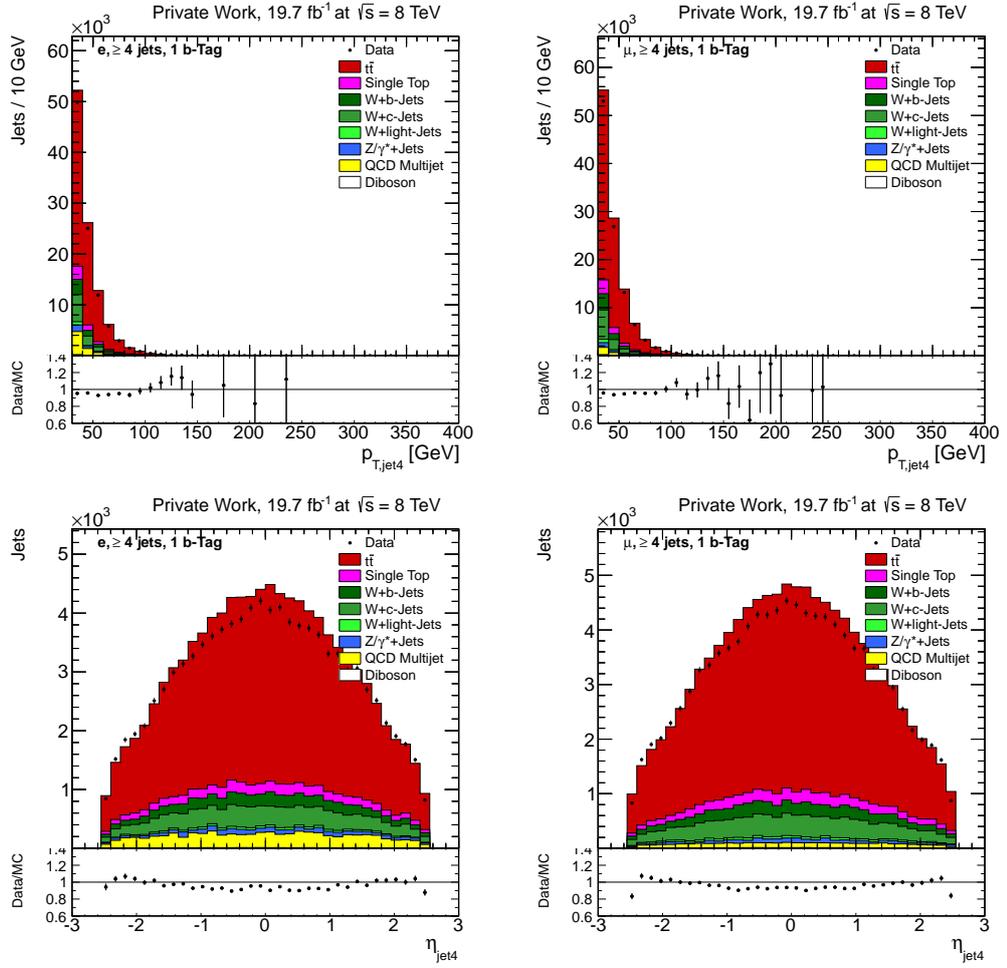


Figure C.5: Distribution of the p_T the fourth hardest selected jet (top) and η of this jet (bottom), in the e +jets channel (left) and μ +jets channel (right) for 1-tag events.

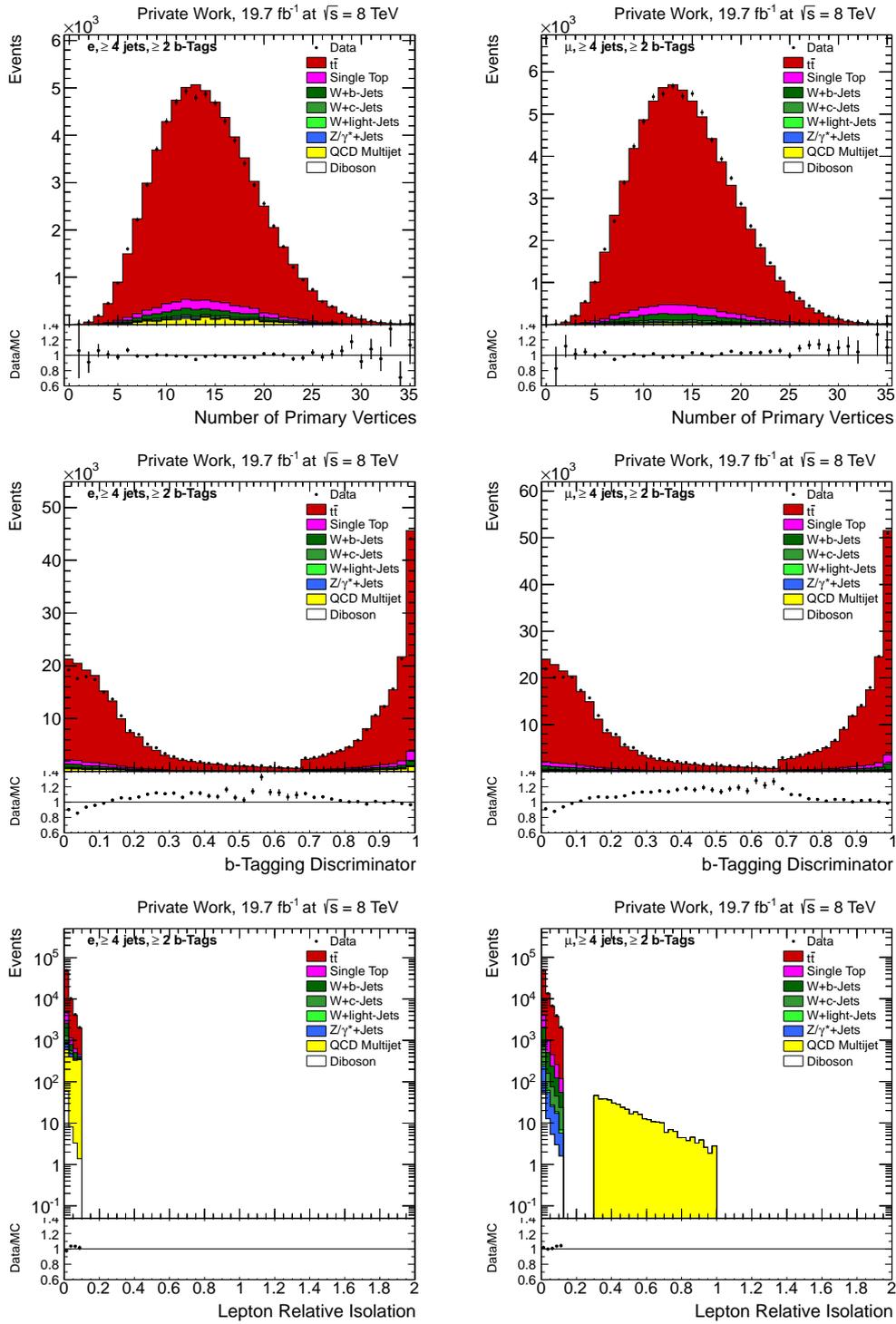


Figure C.6: Distribution of the number of reconstructed primary vertices (top), the b -tagging discriminator for all selected jets (middle), and the relative isolation of this lepton (bottom), in the e +jets channel (left) and μ +jets channel (right) for ≥ 2 -tag events.

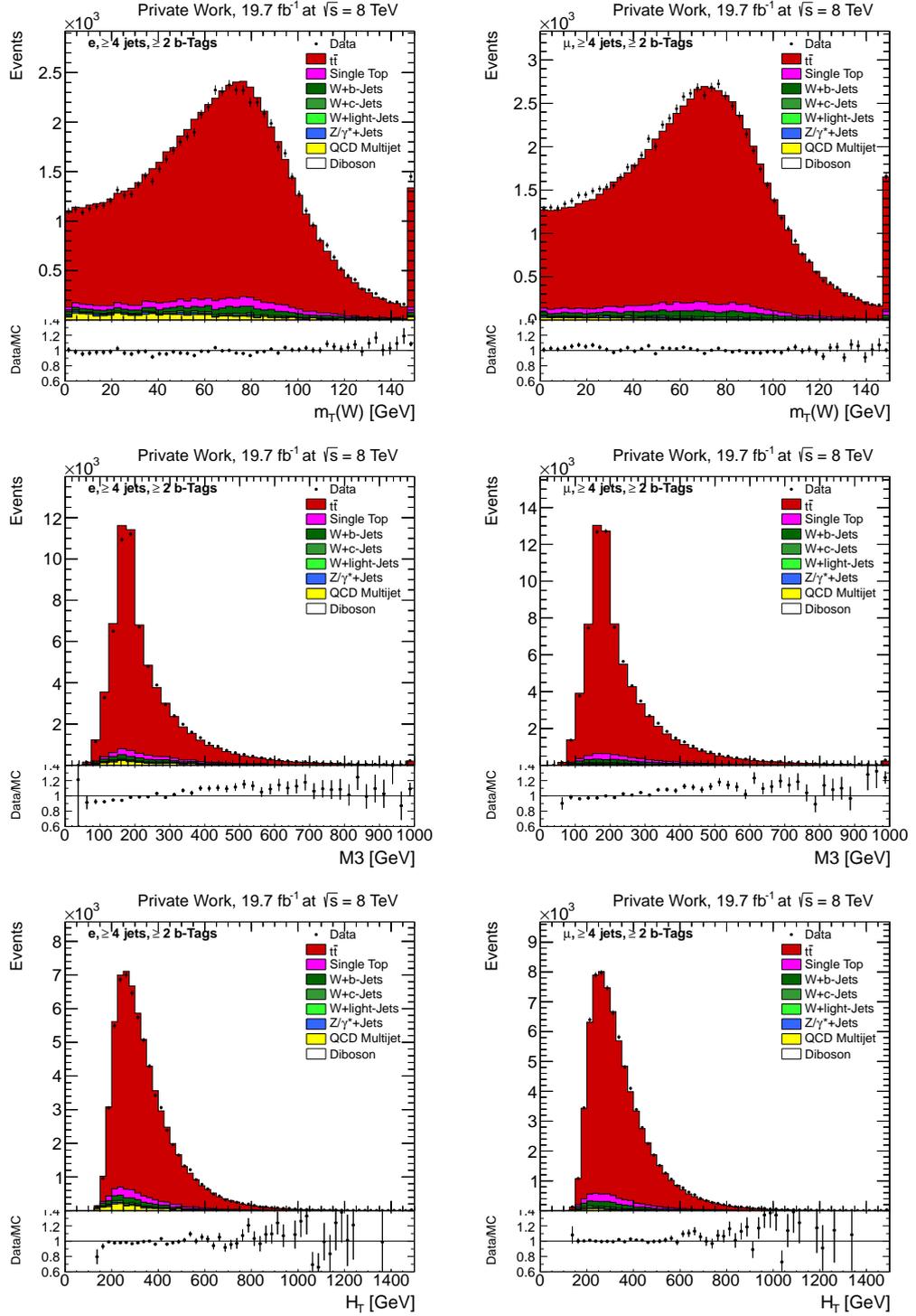


Figure C.7: Distribution of $M_T(W)$ the W -boson transverse invariant mass using the p_T of the selected lepton and the measured \cancel{E}_T (top), $M3$, the invariant mass of the hardest three-jet vectorial sum (middle), H_T , and the p_T scalar sum of all selected jets (bottom) in the e +jets channel (left) and μ +jets channel (right) for ≥ 2 -tag events.

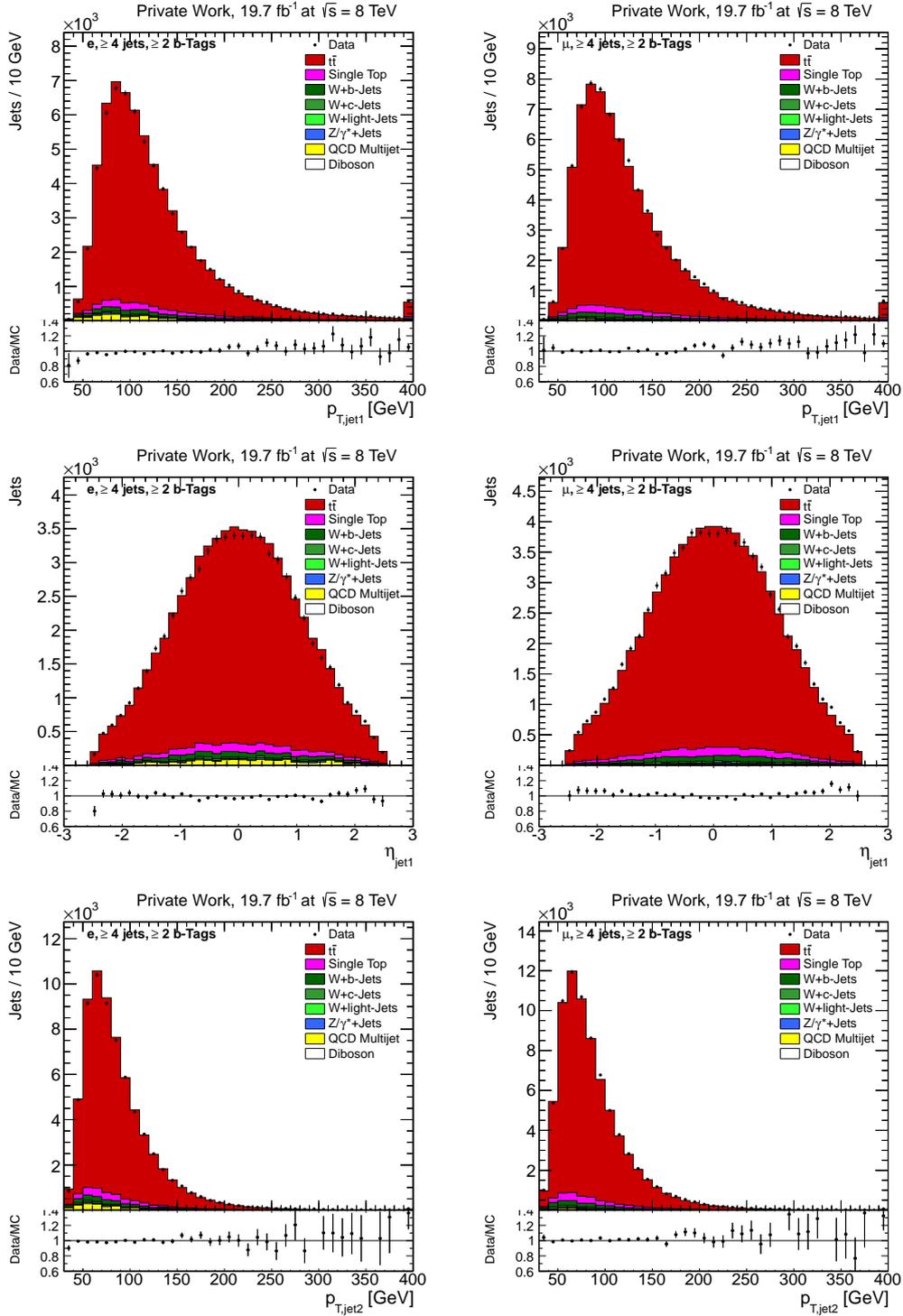


Figure C.8: Distribution of the p_T the hardest selected jet (top), η of this selected jet (middle), and the p_T of the second hardest selected jet (bottom), in the e +jets channel (left) and μ +jets channel (right) for ≥ 2 -tag events.

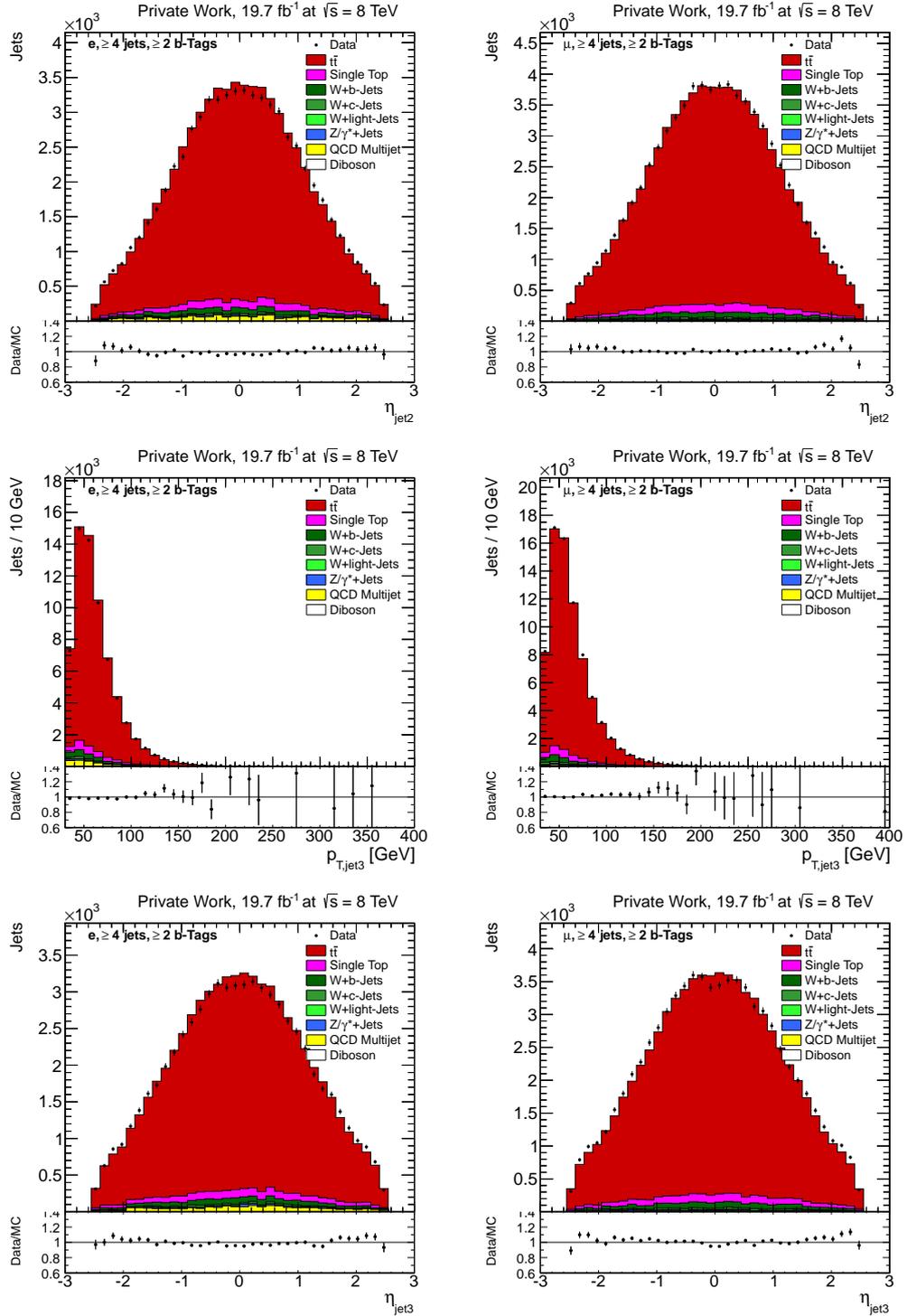


Figure C.9: Distribution of η of the second hardest selected jet (top), the p_T of the third hardest selected jet (middle), and η of this selected jet (bottom), in the $e+jets$ channel (left) and $\mu+jets$ channel (right) for ≥ 2 -tag events.

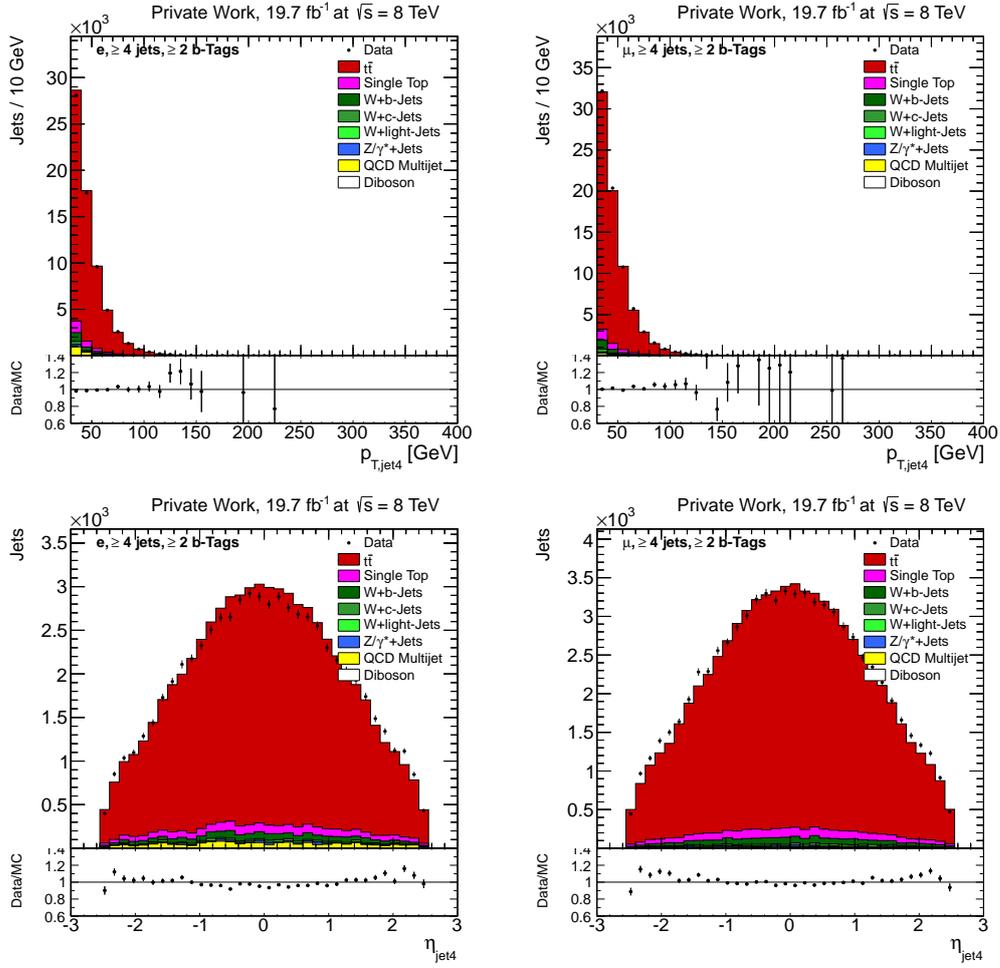


Figure C.10: Distribution of the p_T the fourth hardest selected jet (top) and η of this selected jet (bottom), in the e +jets channel (left) and μ +jets channel (right) for ≥ 2 -tag events.

Bibliography

- [1] K. Olive *et al.*, “Review of Particle Physics,” *Chin.Phys.*, vol. C38, p. 090001, 2014.
- [2] F. Abe *et al.*, “Observation of top quark production in $\bar{p}p$ collisions,” *Phys.Rev.Lett.*, vol. 74, pp. 2626–2631, 1995.
- [3] S. Abachi *et al.*, “Observation of the top quark,” *Phys.Rev.Lett.*, vol. 74, pp. 2632–2637, 1995.
- [4] G. Aad *et al.*, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC,” *Physics Letters B*, vol. 716, no. 1, pp. 1 – 29, 2012.
- [5] S. Chatrchyan *et al.*, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC,” *Physics Letters B*, vol. 716, no. 1, pp. 30 – 61, 2012.
- [6] G. Aad *et al.*, “Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments,” 2015. <http://inspirehep.net/record/1356276> retrieved on April 10, 2015.
- [7] M. Dasgupta, “An Introduction to Quantum Field Theory,” lectures given at the RAL School for High Energy Physicists, Somerville College, Oxford, September 2008. <http://www.hep.manchester.ac.uk/u/dasgupta/teaching/Dasgupta-08-Intro-to-QFT.pdf>, retrieved on April 10, 2015.
- [8] T. Ohl, “Feynman Diagrams for Pedestrians,” lectures given at the 45th Maria Laach School for High Energy Physics, September 2013, retrieved on April 10, 2015.
- [9] P. Skands, “Introduction to QCD,” pp. 341–420, 2013. <http://inspirehep.net/record/1121892>, retrieved on April 10, 2015.

- [10] G. Hanson *et al.*, “Evidence for Jet Structure in Hadron Production by e^+e^- Annihilation,” *Phys. Rev. Lett.*, vol. 35, pp. 1609–1612, Dec 1975.
- [11] F. Englert and R. Brout, “Broken Symmetry and the Mass of Gauge Vector Mesons,” *Phys. Rev. Lett.*, vol. 13, pp. 321–323, Aug 1964.
- [12] P. W. Higgs, “Broken symmetries, massless particles and gauge fields,” *Physics Letters*, vol. 12, no. 2, pp. 132 – 133, 1964.
- [13] P. W. Higgs, “Broken Symmetries and the Masses of Gauge Bosons,” *Phys. Rev. Lett.*, vol. 13, pp. 508–509, Oct 1964.
- [14] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, “Global Conservation Laws and Massless Particles,” *Phys. Rev. Lett.*, vol. 13, pp. 585–587, Nov 1964.
- [15] P. W. Higgs, “Spontaneous Symmetry Breakdown without Massless Bosons,” *Phys. Rev.*, vol. 145, pp. 1156–1163, May 1966.
- [16] T. W. B. Kibble, “Symmetry Breaking in Non-Abelian Gauge Theories,” *Phys. Rev.*, vol. 155, pp. 1554–1561, Mar 1967.
- [17] S. Chatrchyan, H. Heath, and G. Heath, “Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s} = 7$ and 8 TeV,” *JHEP*, vol. 1306, p. 081, 2013.
- [18] DØ Collaboration, “Useful Diagrams of Top Signals and Backgrounds,” http://www-d0.fnal.gov/Run2Physics/top/top_public_web_pages/top_feynman_diagrams.html, retrieved on April 10, 2015.
- [19] Tevatron Electroweak Working Group, “Combination of CDF and D0 results on the mass of the top quark using up to 9.7 fb^{-1} at the Tevatron,” 2014. <http://inspirehep.net/record/1305618>, retrieved on April 10, 2015.
- [20] G. Cortiana on behalf of the ATLAS and CMS collaborations within the TOPLHCWG, “Review of recent top-quark LHC combinations,” 2014. <http://inspirehep.net/record/1331394>, retrieved on April 10, 2015.
- [21] The CMS Collaboration, “Combination of the CMS top-quark mass measurements from Run 1 of the LHC,” 2014. <http://inspirehep.net/record/1319377>, retrieved on April 10, 2015.
- [22] F. Deliot and D. A. Glenzinski, “Top Quark Physics at the Tevatron,” *Rev.Mod.Phys.*, vol. 84, p. 211, 2012.

- [23] L. Lipatov, “Leading logarithmic approximation in QCD,” *Physics of Atomic Nuclei*, vol. 67, no. 1, pp. 83–89, 2004.
- [24] M. Czakon and A. Mitov, “Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders,” *Comput.Phys.Commun.*, vol. 185, p. 2930, 2014.
- [25] M. Czakon, P. Fiedler, and A. Mitov, “Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\frac{4}{s})$,” *Phys.Rev.Lett.*, vol. 110, p. 252004, 2013.
- [26] TOPLHCWG - Top Physics LHC Working Group, <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TopLHCWG>, retrieved on April 10, 2015.
- [27] TOPLHCWG - TOPLHC WG Summary Plots <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TopLHCWGSummaryPlots>, retrieved on June 12, 2015.
- [28] S. Dittmaier, P. Uwer, and S. Weinzierl, “NLO QCD corrections to t anti- t + jet production at hadron colliders,” *Phys.Rev.Lett.*, vol. 98, p. 262002, 2007.
- [29] S. Alioli, S.-O. Moch, and P. Uwer, “Hadronic top-quark pair-production with one jet and parton showering,” *JHEP*, vol. 1201, p. 137, 2012.
- [30] G. Bevilacqua, M. Czakon, C. Papadopoulos, and M. Worek, “Hadronic top-quark pair production in association with two jets at Next-to-Leading Order QCD,” *Phys.Rev.*, vol. D84, p. 114017, 2011.
- [31] The CDF Collaboration, “Measurement of the $t\bar{t}$ + jets Cross Section with 4.1 fb^{-1} ,” CDF Note 9850, http://www-cdf.fnal.gov/physics/new/top/2009/xsection/ttj_4.1invfb/, retrieved on April 10, 2015.
- [32] S. Dittmaier, P. Uwer, and S. Weinzierl, “Hadronic top-quark pair production in association with a hard jet at next-to-leading order QCD: Phenomenological studies for the Tevatron and the LHC,” *Eur.Phys.J.*, vol. C59, pp. 625–646, 2009.
- [33] “Comparison of Monte Carlo generator predictions for gap fraction and jet multiplicity observables in top-antitop events,” Tech. Rep. ATL-PHYS-PUB-2014-005, CERN, Geneva, May 2014. <http://cds.cern.ch/record/1703034>, retrieved on April 10, 2015.

- [34] G. Aad *et al.*, “Measurement of the $t\bar{t}$ production cross-section as a function of jet multiplicity and jet transverse momentum in 7 TeV proton-proton collisions with the ATLAS detector,” *JHEP*, vol. 1501, p. 020, 2015.
- [35] S. Chatrchyan *et al.*, “Measurement of jet multiplicity distributions in $t\bar{t}$ production in pp collisions at $\sqrt{s} = 7$ TeV,” *Eur.Phys.J.*, vol. C74, no. 8, p. 3014, 2014.
- [36] The ATLAS Collaboration, “Determination of the top-quark pole mass using $t\bar{t}+1$ -jet events collected with the ATLAS experiment in 7 TeV pp collisions,” 2014. <http://cds.cern.ch/record/1951319>, retrieved on April 10, 2015.
- [37] The CMS Collaboration, “Search for pair production of resonances decaying to a top quark plus a jet in final states with two leptons,” 2013. <http://cds.cern.ch/record/1630845>, retrieved on May 20, 2015.
- [38] The CMS Collaboration, “Search for baryon number violating top quark decays in pp collisions at 8 TeV,” 2013. <http://inspirehep.net/record/1260833>, retrieved on May 20, 2015.
- [39] N. Metropolis and U. Stanislaw, “The Monte Carlo Method,” *Journal of the American Statistical Association*, vol. 44, pp. 335–341.
- [40] R. Feynman, “The Behavior of Hadron Collisions at Extreme Energies,” in *Special Relativity and Quantum Theory* (M. Noz and Y. Kim, eds.), vol. 33 of *Fundamental Theories of Physics*, pp. 289–304, Springer Netherlands, 1988.
- [41] E. D. Bloom, D. H. Coward, H. DeStaabler, J. Drees, G. Miller, L. W. Mo, R. E. Taylor, M. Breidenbach, J. I. Friedman, G. C. Hartmann, and H. W. Kendall, “High-Energy Inelastic $e - p$ Scattering at 6° and 10° ,” *Phys. Rev. Lett.*, vol. 23, pp. 930–934, Oct 1969.
- [42] I. Abt *et al.*, “Measurement of the proton structure function $\{F_2\}$ (x , Q^2) in the low- x region at $\{\text{HERA}\}$,” *Nuclear Physics B*, vol. 407, no. 3, pp. 515 – 535, 1993.
- [43] M. Derrick *et al.*, “Measurement of the proton structure function $\{F_2\}$ in ep scattering at $\{\text{HERA}\}$,” *Physics Letters B*, vol. 316, no. 23, pp. 412 – 426, 1993.

- [44] A. Argento *et al.*, “Measurement of the interference structure function $xG_3(x)$ in muon-nucleon scattering,” *Physics Letters B*, vol. 140, no. 12, pp. 142 – 144, 1984.
- [45] W. G. Seligman *et al.*, “Improved Determination of α_s From Neutrino-Nucleon Scattering,” *Phys. Rev. Lett.*, vol. 79, pp. 1213–1216, Aug 1997.
- [46] Y. L. Dokshitzer, “Calculation of the Structure Functions for Deep Inelastic Scattering and $e^+ e^-$ Annihilation by Perturbation Theory in Quantum Chromodynamics.,” *Sov.Phys.JETP*, vol. 46, pp. 641–653, 1977.
- [47] V. Gribov and L. Lipatov, “Deep inelastic $e p$ scattering in perturbation theory,” *Sov.J.Nucl.Phys.*, vol. 15, pp. 438–450, 1972.
- [48] G. Altarelli and G. Parisi, “Asymptotic Freedom in Parton Language,” *Nucl.Phys.*, vol. B126, p. 298, 1977.
- [49] G. Grindhammer, H1 Collaboration and ZEUS Collaboration, “Determination of PDFs and $\alpha_S(M_Z)$ from inclusive and jet measurements in DIS @ HERA,” *AIP Conference Proceedings*, vol. 1441, no. 1, pp. 225–228, 2012.
- [50] J. Pumplin, D. Stump, J. Huston, H. Lai, P. M. Nadolsky, *et al.*, “New generation of parton distributions with uncertainties from global QCD analysis,” *JHEP*, vol. 0207, p. 012, 2002.
- [51] The Durham HepData Project, “Online PDF plotting and calculation,” <http://hepdata.cedar.ac.uk/pdf/pdf3.html>, retrieved on June 1st, 2015.
- [52] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, *et al.*, “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations,” *JHEP*, vol. 1407, p. 079, 2014.
- [53] R. Frederix and S. Frixione, “Merging meets matching in MC@NLO,” *JHEP*, vol. 1212, p. 061, 2012.
- [54] S. Frixione and B. R. Webber, “Matching NLO QCD computations and parton shower simulations,” *JHEP*, vol. 0206, p. 029, 2002.
- [55] S. Frixione, P. Nason, and B. R. Webber, “Matching NLO QCD and parton showers in heavy flavor production,” *JHEP*, vol. 0308, p. 007, 2003.

- [56] S. Frixione, F. Stöckli, P. Torrielli, and B. R. Webber, “NLO QCD corrections in Herwig++ with MC@NLO,” *JHEP*, vol. 1101, p. 053, 2011.
- [57] S. Frixione, P. Nason, and G. Ridolfi, “A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction,” *JHEP*, vol. 0709, p. 126, 2007.
- [58] P. Nason, “A New method for combining NLO QCD with shower Monte Carlo algorithms,” *JHEP*, vol. 0411, p. 040, 2004.
- [59] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with Parton Shower simulations: the POWHEG method,” *JHEP*, vol. 0711, p. 070, 2007.
- [60] S. Alioli, P. Nason, C. Oleari, and M. Re, “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX,” *JHEP*, vol. 1006, p. 043, 2010.
- [61] T. Sjöstrand, S. Mrenna, and P. Skands, “PYTHIA 6.4 physics and manual,” *Journal of High Energy Physics*, vol. 2006, no. 05, p. 026, 2006.
- [62] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, *et al.*, “An Introduction to PYTHIA 8.2,” *Comput.Phys.Commun.*, vol. 191, pp. 159–177, 2015.
- [63] T. Sjöstrand and P. Z. Skands, “Transverse-momentum-ordered showers and interleaved multiple interactions,” *Eur.Phys.J.*, vol. C39, pp. 129–154, 2005.
- [64] G. Corcella, I. Knowles, G. Marchesini, S. Moretti, K. Odagiri, *et al.*, “HERWIG 6.5 release note,” 2002.
- [65] T. Gleisberg, S. Höche, F. Krauss, M. Schönherr, S. Schumann, *et al.*, “Event generation with SHERPA 1.1,” *JHEP*, vol. 0902, p. 007, 2009.
- [66] S. Catani, F. Krauss, B. R. Webber, and R. Kuhn, “QCD Matrix Elements + Parton Showers,” *Journal of High Energy Physics*, vol. 2001, no. 11, p. 063, 2001.
- [67] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, “Matching matrix elements and shower evolution for top-pair production in hadronic collisions,” *Journal of High Energy Physics*, vol. 2007, no. 01, p. 013, 2007.

- [68] S. Banerjee, “CMS Simulation Software,” Tech. Rep. CMS-CR-2012-146, CERN, Geneva, Jun 2012. <http://cds.cern.ch/record/1457820>, retrieved on April 23, 2015.
- [69] R. Rahmat, R. Kroeger, and A. Giammanco, “The Fast Simulation of The CMS Experiment,” *Journal of Physics: Conference Series*, vol. 396, no. 6, p. 062016, 2012. <http://stacks.iop.org/1742-6596/396/i=6/a=062016>, retrieved on June 1st, 2015.
- [70] A. Descroix, “Study of b-tagging algorithms for top-quark physics at cms,” 2011. <https://ekp-invenio.physik.uni-karlsruhe.de/record/45233>, retrieved on June 10, 2015.
- [71] ABB Group, “ABB robots help recreate the Big Bang,” <http://www.abb.com/cawp/seitp202/c659a4d6ad5d52acc12574c1004bdbe2.aspx>, retrieved on April 10, 2015.
- [72] C. Lefevre, “LHC: the guide,” <http://cdsweb.cern.ch/record/1165534/>, retrieved on April 10, 2015.
- [73] L. Evans and P. Bryant, “LHC Machine,” *Journal of Instrumentation*, vol. 3, no. 08, p. S08001, 2008.
- [74] CERN, “LHC images,” http://lhc-machine-outreach.web.cern.ch/lhc-machine-outreach/lhc_in_pictures.htm, retrieved on April 10, 2015.
- [75] D. Fournier, “Performance of the LHC, ATLAS and CMS in 2011,” *EPJ Web Conf.*, vol. 28, p. 01003, 2012.
- [76] M. Lamont, “Status of the LHC,” *Journal of Physics: Conference Series*, vol. 455, no. 1, p. 012001, 2013.
- [77] The CMS Collaboration, “CMS Luminosity - Public Results,” <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>, retrieved on April 10, 2015.
- [78] The ATLAS Collaboration, “The ATLAS Experiment at the CERN Large Hadron Collider,” *Journal of Instrumentation*, vol. 3, no. 08, p. S08003, 2008.
- [79] The CMS Collaboration, “The CMS experiment at the CERN LHC,” *Journal of Instrumentation*, vol. 3, no. 08, p. S08004, 2008.
- [80] The ALICE Collaboration, “The ALICE experiment at the CERN LHC,” *Journal of Instrumentation*, vol. 3, no. 08, p. S08002, 2008.

- [81] The LHCb Collaboration, “The LHCb Detector at the LHC,” *Journal of Instrumentation*, vol. 3, no. 08, p. S08005, 2008.
- [82] The CMS Collaboration, “Interactive Slice of the CMS detector,” https://cms-docdb.cern.ch/cgi-bin/PublicDocDB/RetrieveFile?docid=4172&version=2&filename=CMS_Slice.gif, retrieved on April 10, 2015. From parent site: <http://cms.web.cern.ch/>.
- [83] S. Chatrchyan *et al.*, “Performance of the CMS hadron calorimeter with cosmic ray muons and LHC beam data,” *Journal of Instrumentation*, vol. 5, p. 3012, Mar. 2010.
- [84] S. Abdullin *et al.*, “Design, performance, and calibration of CMS hadron-barrel calorimeter wedges,” *Eur.Phys.J.*, vol. C55, pp. 159–171, 2008.
- [85] G. Bayatian *et al.*, “Design, performance and calibration of the CMS forward calorimeter wedges,” *Eur.Phys.J.*, vol. C53, pp. 139–166, 2008.
- [86] G. L. Bayatian *et al.*, *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. Technical Design Report CMS, Geneva: CERN, 2006. <http://cds.cern.ch/record/922757>, retrieved on April 10, 2015.
- [87] G. L. Bayatian *et al.*, *The CMS muon project: Technical Design Report*. Technical Design Report CMS, Geneva: CERN, 1997. <http://cds.cern.ch/record/343814>, retrieved on April 10, 2015.
- [88] S. Cittolin, A. Rcz, and P. Spiccas, *CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project*. Technical Design Report CMS, Geneva: CERN, 2002. <http://cds.cern.ch/record/578006>, retrieved on April 10, 2015.
- [89] “Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET,” 2009. <https://inspirehep.net/record/925379>, retrieved on April 24, 2015.
- [90] “Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector,” 2010. <http://inspirehep.net/record/925318>, retrieved on April 24, 2015.
- [91] “Commissioning of the Particle-Flow reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV,” 2010. <http://inspirehep.net/record/925292>, retrieved on April 24, 2015.

- [92] “Particle-flow commissioning with muons and electrons from J/Psi and W events at 7 TeV,” 2010. <http://inspirehep.net/record/925288>, retrieved on April 24, 2015.
- [93] The CMS Collaboration, “Description and performance of track and primary-vertex reconstruction with the CMS tracker,” *Journal of Instrumentation*, vol. 9, no. 10, p. P10009, 2014.
- [94] R. Frühwirth, “Application of Kalman filtering to track and vertex fitting,” *Nuclear Instruments and Methods in Physics Research Section A*, vol. 262, no. 23, pp. 444 – 450, 1987.
- [95] V. Khachatryan *et al.*, “Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV,” 2015. <http://inspirehep.net/record/1343791>, retrieved on April 24, 2015.
- [96] The CMS Collaboration, “Electron reconstruction and identification at $\sqrt{s} = 7$ TeV,” no. CMS-PAS-EGM-10-004, 2010. <http://cds.cern.ch/record/1299116>, retrieved on April 10, 2015.
- [97] J. Nysten, “Photon Reconstruction in CMS,” Tech. Rep. CMS-CR-2004-004, CERN, Geneva, Feb 2004. <http://cds.cern.ch/record/787484>, retrieved on June 8, 2015.
- [98] V. Khachatryan *et al.*, “Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV,” 2015. <http://inspirehep.net/record/1343792>, retrieved on April 24, 2015.
- [99] The CMS Collaboration, “Photon reconstruction and identification at $\sqrt{s} = 7$ TeV,” no. CMS-PAS-EGM-10-005, 2010. <http://cds.cern.ch/record/1279143>, retrieved on April 10, 2015.
- [100] The CMS Collaboration, “Performance of Muon Identification in pp Collisions at $\sqrt{s} = 7$ TeV,” *CMS-PAS-MUO-10-002*, 2010. <http://cds.cern.ch/record/1279140>, retrieved on April 10, 2015.
- [101] G. P. Salam, “Towards Jetography,” *Eur.Phys.J.*, vol. C67, pp. 637–686, 2010.
- [102] The CMS Collaboration, “Determination of jet energy calibration and transverse momentum resolution in CMS,” *Journal of Instrumentation*, vol. 6, no. 11, p. P11002, 2011.

- [103] S. Chatrchyan *et al.*, “Identification of b-quark jets with the CMS experiment,” *JINST*, vol. 8, p. P04013, 2013.
- [104] “Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET,” Tech. Rep. CMS-PAS-PFT-09-001, CERN, 2009. Geneva, Apr 2009. <https://cds.cern.ch/record/1194487>, retrieved on June 8, 2015.
- [105] V. Blobel and E. Lohrmann, *Statistische und numerische Methoden der Datenanalyse*. Teubner-Studienbücher : Physik, Stuttgart: Teubner, 1998. <http://www.desy.de/~blobel/ebuch.html>, retrieved on May 13, 2015.
- [106] F. James and M. Winkler, “MINUIT User’s Guide,” 2004. <https://inspirehep.net/record/1258345>, retrieved on May 14, 2015.
- [107] W. C. Davidon, “Variable Metric Method for Minimization,” 1966. <https://inspirehep.net/record/50075>, more recent document online: <http://www.cs.berkeley.edu/~jduchi/papers/Davidon91.pdf>, both retrieved on May 14, 2015.
- [108] F. James, *Statistical methods in experimental physics*. New Jersey [u.a.]: World Scientific, 2. ed., repr. ed., 2008. 1. ed.: Statistical methods in experimental physics by W. T. Eadie...
- [109] V. Blobel, “Data unfolding, lecture given in March 2010 at the “Terascale Statistics Tools School”, DESY-Hamburg, Germany,” www.desy.de/~blobel/school_march10.pdf, retrieved on May 13, 2015.
- [110] A. Descroix, “Jet Multiplicity in Top-Quark Pair Events at CMS - Poster Proceedings,” 2014. <http://inspirehep.net/record/1327225>, retrieved on May 28, 2015.
- [111] F. Krauss, “A theory perspective on Top2014,” 2014. <http://arxiv.org/abs/1412.7343>, retrieved on June 10, 2015.
- [112] “Measurement of jet multiplicity in ttbar events in the lepton+jets channel at 8 TeV,” Tech. Rep. AN-14-087, CERN, Geneva, 2014. CMS internal: <http://cms.cern.ch/iCMS/user/noteinfo?cmsnoteid=CMS%20AN-2014/087>, retrieved on June 10, 2015.
- [113] A. Perloff, on behalf of the CMS collaboration, “Pileup measurement and mitigation techniques in CMS,” *Journal of Physics: Conference Series*, vol. 404, no. 1, p. 012045, 2012.

- [114] The CMS Collaboration, “CMS Top EGamma Coordination.” CMS internal: <https://twiki.cern.ch/twiki/bin/view/CMS/TopEGMRun1>, retrieved on June 1st, 2015.
- [115] The CMS Collaboration, “Top Muon Information for Analysis.” CMS internal: <https://twiki.cern.ch/twiki/bin/view/CMS/TopMUORun1>, retrieved on June 1st, 2015.
- [116] The CMS Collaboration, “Particle Flow in PAT (PF2PAT): Top-Projection or avoiding double-counting,” https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuidePF2PAT#Top_projection_or_avoiding_doubl, retrieved on May 2, 2015.
- [117] The CMS Collaboration, “Top Jet/MET Analysis.” CMS internal: <https://twiki.cern.ch/twiki/bin/view/CMS/TopJMERun1>, retrieved on June 1st, 2015.
- [118] The CMS Collaboration, “Top reference selections and recommendations.” CMS internal: <https://twiki.cern.ch/twiki/bin/view/CMS/TopEventSelectionRun1>, retrieved on June 1st, 2015.
- [119] Web site of the Robust Independent Validation of Experiment and Theory (RIVET), <http://rivet.hepforge.org/>, retrieved on May 21, 2015
- [120] Web site of the FastJet package, <http://fastjet.fr/>, retrieved on May 21, 2015
- [121] The CMS Collaboration, “TWiki of the Physics Data/MC Validation - Production: Gathering vital information about the Summer12 Monte-Carlo production,” CMS internal: https://twiki.cern.ch/twiki/bin/view/CMS/PdmVProductionSummer12#Summer12_R53X. More details in presentation (CMS internal): <https://indico.cern.ch/event/169299/contribution/7/material/slides/1.pdf>. All retrieved on April 30, 2015.
- [122] TOPLHCWG, “NNLO+NNLL top-quark-pair cross sections – ATLAS-CMS recommended predictions for top-quark-pair cross sections using the Top++v2.0 program,” <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TtbarNNLO>, retrieved on June 15, 2015.
- [123] N. Kidonakis, “Differential and total cross sections for top pair and single top production,” pp. 831–834, 2012. <http://arxiv.org/abs/1205.3453>, retrieved on June 10, 2015.

- [124] K. Melnikov and F. Petriello, “Electroweak gauge boson production at hadron colliders through $O(\alpha_S^2)$,” *Phys.Rev.*, vol. D74, p. 114017, 2006.
- [125] J. M. Campbell and R. K. Ellis, “Radiative corrections to Z b anti-b production,” *Phys.Rev.*, vol. D62, p. 114012, 2000.
- [126] The CMS Collaboration, “2012 Analysis,” CMS internal: <https://twiki.cern.ch/twiki/bin/viewauth/CMS/PdmV2012Analysis>, retrieved on June 1st, 2015.
- [127] The CMS Collaboration, “Introduction to Jet Energy Corrections at CMS.” CMS internal: <https://twiki.cern.ch/twiki/bin/viewauth/CMS/IntroToJEC>, retrieved on May 24, 2015.
- [128] The CMS Collaboration, “Jet Energy Resolution and Corrections (JERC) Subgroup.” CMS internal: <https://twiki.cern.ch/twiki/bin/view/CMS/JetEnergyScale>, retrieved on May 24, 2015.
- [129] The CMS Collaboration, “Recommended Jet Energy Corrections and Uncertainties For Data and MC.” CMS internal: <https://twiki.cern.ch/twiki/bin/view/CMS/JECDataMC>, retrieved on May 24, 2015.
- [130] The CMS Collaboration, “Jet Energy Resolution.” CMS internal: <https://twiki.cern.ch/twiki/bin/view/CMS/JetResolution>, retrieved on May 24, 2015.
- [131] The CMS Collaboration, “Official recommendations for Pile Up with 2012 data.” CMS internal: <https://twiki.cern.ch/twiki/bin/viewauth/CMS/PileupSystematicErrors>, retrieved on June 1st, 2015.
- [132] The CMS Collaboration, “Utilities for Accessing Pileup Information for Data,” CMS internal: <https://twiki.cern.ch/twiki/bin/view/CMS/PileupJSONFileforData>, retrieved on May 4, 2015.
- [133] The CMS Collaboration, “Measurement of differential top-quark-pair production cross sections in pp collisions at $\sqrt{s} = 7$ TeV,” *The European Physical Journal C*, vol. 73, no. 3, 2013.
- [134] The CMS Collaboration, “Measurement of differential top-quark pair production cross sections in the lepton+jets channel in pp collisions at 8 TeV,” Tech. Rep. CMS-PAS-TOP-12-027, CERN, Geneva, 2013.
- [135] N. Kidonakis, “NNLL threshold resummation for top-pair and single-top production,” *Phys.Part.Nucl.*, vol. 45, no. 4, pp. 714–722, 2014.

- [136] The CMS Collaboration, “Studies on top-pt reweighting.” CMS internal: <https://twiki.cern.ch/twiki/bin/viewauth/CMS/TopPtRewighting>, retrieved on June 1st, 2015.
- [137] V. Khachatryan *et al.*, “Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV,” Tech. Rep. arXiv:1502.02701. CERN-PH-EP-2015-004. CMS-EGM-13-001, CERN, Geneva, Feb 2015. Comments: Submitted to JINST.
- [138] S. Chatrchyan *et al.*, “The performance of the CMS muon detector in proton-proton collisions at $\sqrt{s} = 7$ TeV at the LHC,” *J. Instrum.*, vol. 8, p. P11002, Jun 2013. Comments: Submitted to JINST.
- [139] The CMS Collaboration, “Official efficiency-correction recommendations for electrons in 2012 run.” CMS internal: <https://twiki.cern.ch/twiki/bin/viewauth/CMS/KoPFAElectronTagAndProbe>, retrieved on June 1st, 2015.
- [140] The CMS Collaboration, “Official efficiency-correction recommendations for muons in 2012 run.” CMS internal: <https://twiki.cern.ch/twiki/bin/viewauth/CMS/MuonReferenceEffs>, retrieved on June 1st, 2015.
- [141] The CMS Collaboration, “Usage of b Tag Objects for 8 TeV Data with the 53X Re-reco.” CMS internal: <https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation53XReReco>, retrieved on June 1st, 2015.
- [142] The CMS Collaboration, “Mistag Scale Factors for b-tagging in 2012 data.” CMS internal: http://cms.cern.ch/iCMS/jsp/openfile.jsp?tp=draft&files=AN2013_150_v2.pdf, retrieved on June 1st, 2015.
- [143] “Performance of b tagging at $\sqrt{s} = 8$ TeV in multijet, ttbar and boosted topology events,” Tech. Rep. CMS-PAS-BTV-13-001, CERN, Geneva, 2013. <http://cds.cern.ch/record/1581306>, retrieved on June 1st, 2015.
- [144] The CMS Collaboration, “Measurement of the $t\bar{t}$ production cross section in pp collisions at $\sqrt{s} = 7$ TeV with lepton+jets final states,” *Physics Letters B*, vol. 720, no. 120133, pp. 83 – 104, 2013.

- [145] S. Chatrchyan *et al.*, “Measurement of Inclusive W and Z Boson Production Cross Sections in pp Collisions at $\sqrt{s} = 8$ TeV,” *Phys. Rev. Lett.*, vol. 112, p. 191802, May 2014.
- [146] J. M. Campbell and R. Ellis, “MCFM for the Tevatron and the LHC,” *Nucl.Phys.Proc.Suppl.*, vol. 205-206, pp. 10–15, 2010.
- [147] V. Blobel, “An Unfolding method for high-energy physics experiments,” pp. 258–267, 2002. <http://inspirehep.net/record/593047>, retrieved on June 10, 2015.
- [148] A. Hocker and V. Kartvelishvili, “SVD approach to data unfolding,” *Nucl.Instrum.Meth.*, vol. A372, pp. 469–481, 1996.
- [149] The CMS Collaboration, “CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update,” 2013. <http://inspirehep.net/record/1260876>, retrieved on May 6, 2015.
- [150] “Measurement of jet multiplicity in $t\bar{t}$ events in the lepton+jets channel at 8 TeV,” Tech. Rep. AN-14-087, CERN, Geneva, 2014. CMS internal: <http://cms.cern.ch/iCMS/user/noteinfo?cmsnoteid=CMS%20AN-2012/239>, retrieved on June 10, 2015.
- [151] “HepMC - a C++ Event Record for Monte Carlo Generators,” <http://lcgapp.cern.ch/project/simu/HepMC/>, retrieved on May 20, 2015.
- [152] S. Höche, F. Krauss, S. Schumann, and F. Siegert, “QCD matrix elements and truncated showers,” *Journal of High Energy Physics*, vol. 2009, no. 05, p. 053, 2009.
- [153] S. Höche, F. Krauss, M. Schönherr, and F. Siegert, “QCD matrix elements + parton showers. The NLO case,” *Journal of High Energy Physics*, vol. 2013, no. 4, 2013.

Acknowledgment

I would like to thank Prof. Dr. Ulrich Husemann for allowing me to join his group at KIT as a PhD student. During these four years under his supervision, I have benefited from his wide and huge knowledge and experience, his impressive dynamism and motivation, and his availability and friendly nature. Many thanks for the good management, the constructive criticisms, and the thorough review of my thesis. Many thanks also to Prof. Dr. Thomas Müller for accepting to be my co-referent, the hints for preparing my doctor oral exam and in general for setting a pleasant, convivial, and still studious atmosphere in the institute, as its leader. Thanks a lot to Prof. Dr. Johann Kühn, who has accepted to be my mentor for my PhD period. I appreciated our constructive meetings, which I found useful for standing back with respect to my work.

Dr. Patricia Lobelle has been post-doc in my group almost the whole time of my PhD and shared the office with me. Her thoughtful and patient disposition has helped me not to rush on my first idea, but to think twice, as it is often justified. She has coordinated exchanges and studies for the publication we have achieved in these years, always available for a quick and conclusive check or for chats anytime I was questioning my work or just having a lack of motivation (“¡No quiero trabajar!”). Thank you Patricia for all of this. Moreover, I would like to thank you very much for the intense review of my thesis in the last weeks, for your help, and your support.

Many thanks to Marco Harrendorf, Hannes Mildner, and Shawn Williamson, currently PhD students in my group. I have been delighted by the open-minded, funny, and enriching team spirit taking place with you. Thanks for the constructive exchanges in physics and programming. Thanks also for your contributions to the publication and for the review of my thesis. Thanks a lot to Ralf Farkas, bachelor student in my group, for working hard to produce in the last days the SHERPA samples, with help from Marco Harrendorf. It is a very nice contribution to my thesis. I would like to thank also Olaf Böcker, Frederik Schaaf, and Judith Schwab for their work with

me. Your contributions are very valuable for the past publication as well as for the coming one. Thank you Karim El Morabit and Tobias Pfozner, currently the two Master students in my group, for the review of my thesis and the nice work together. Thanks a lot Alexander Wlotzka and Genessis Perez, who gently accepted to review my theory chapter.

I would like to thank all the colleagues from the IEKP institute in campus north and south of KIT, in the software and the hardware groups as well as from the administration and technical department for the nice atmosphere, the constructive exchanges, and your help. I would also like to thank all the current and former admin team of the IEKP for spending so much of your time to make our IT-resources work very well.

Finally, I would like to thank my family for being patient (“Mais ça n’en fini jamais cette thèse!”) and supporting me all over these years away from you. Thank you all my friends for your supporting and sustaining friendships. Last but not least, thank you Julian for being who you are to me, for being there for me, and for spanning with me our own dimension. Du bist mein Lieblingsteilchen!