Facultat de Física

Departament de Física Atòmica, Molecular i Nuclear

VNIVERSITAT DE VALÈNCIA

First measurement of the charge asymmetry in boosted top quark pair production in the ATLAS experiment using the Grid-based Tier-3 facility at IFIC-Valencia

Victoria Sánchez Martínez

Dirigido por

Santiago González de la Hoz y Marcel Vos

Doctorado en Física, Mención Internacional

Mayo 2016

Dr. Santiago González de la Hoz, Profesor Titular del Departamento de Física Atómica, Molecular y Nuclear de la Universidad de Valencia.
Dr. Marcel Vos, Científico Titular del Consejo Superior de Investigaciones Científicas.

CERTIFICAN:

Que la presente memoria **'First measurement of the charge asymmetry in boosted top quark pair production in the ATLAS experiment using the Grid-based Tier-3 facility at IFIC-Valencia'** ha sido realizada bajo nuestra dirección en el Departamento de Física Atómica, Molecular y Nuclear de la Universidad de Valencia por Doña Victoria Sánchez Martínez y constituye su tesis para optar al grado de Doctor en Física.

Y para que así conste, firmamos el presente Certificado:

Firmado
Dr. Santiago González de la Hoz

Firmado
Dr. Marcel Vos

Firmado
Victoria Sánchez Martínez

El trabajo descrito en esta tesis se ha llevado a cabo en el *Instituto de Física Corpuscular* (IFIC) en Valencia, España.

El IFIC es un centro mixto del *Consejo Superior de Investigaciones Científicas* (CSIC) y de la *Universitat de València*. En Julio de 2015 recibió la acreditación **Severo Ochoa** (SEV-2014-0398) que le distingue como Centro de Excelencia. Dicha acreditación reconoce la excelencia y las contribuciones científicas que realizan los centros y unidades a nivel nacional e internacional, su impacto empresarial y social y su capacidad para atraer talento.

A mi family, Los Martínez, por hacerse querer día a día.
A mi nene, por su apoyo moral, paciencia y motivación diaria.
A tod@s los que han creído y confiado en mí.

Acknowledgements

The results explained in this thesis are the outcome of the work of several people collaborating and participating together, and coming from different institutions within the ATLAS Collaboration.

Thanks to the Grid Computing and $t\bar{t}$ resonances groups, especially to José Salt, Samuel Calvet, Dominik Duda, James Ferrando and Jiahang Zhong. They answered every question I had, they supported my work in one way or another and they allowed me to work with them from the proximity and from the distance.

Thanks to the resolved and dilepton charge asymmetry groups: Francesco Rubbo, Richard Hawkings, Clement Helsends, Aurelio Juste, Alison Lister and Roger Naranjo, among others. Their advice, feedback and support in the charge asymmetry meetings were really useful, and I learnt a lot of them and their work.

Thanks to my colleagues of the boosted charge asymmetry group from Michigan: Dan Marley, Allison McCarn, Tom Schwarz and Bennet Magy. We have worked very hard and side by side, regardless of whether it was Friday or Saturday, or whether it was 10pm or 3am. It was very comfortable having coworkers like them.

Thanks to Germán Rodrigo and Juan Antonio Aguilar Saavedra, for their theoretical support and for providing us every "theoretical thing" we needed for our analysis.

Thanks to Alison Lister and Mark Owen. They were very keen on our analysis and gave us the opportunity to publish our paper in the quickest way ever.

Thanks to Helenka Przysiezniak Frey for helping me with the bureaucracy required to make official the MC samples I produced.

Thanks to the international experts for carefully reading this thesis and giving me their feedback.

Thanks to my English teacher, Nora Selmeczi, for helping me improve my English.

Thanks to Elena Romero, Sebastián López, Patricia Fernández, Miguel Villaplana, Adrián Irles and Nacho García for their moral support and for sharing good times with me.

Special mention to my supervisors Santi and Marcel. Thanks to Santi I started my PhD. Thanks to Marcel I didn't give up. And thanks to both of them I've written this thesis. I'll be eternally grateful, for having supported me occupationally and personally. There are still good people in this world. I owe you a big one... many thanks!

And last but not least, thanks to my family and my husband. They have supported me, encouraged me, understood me, adviced me. They are my family and of course, they are the best ones!

Contents

1

The Standard Model and the Top Quark

All ordinary matter is composed by elementary particles. These are supposed to be indivisible and structureless objects. One example of fundamental particle is the well-known electron (e^-), which forms the electrical current. The electron has two heavier brothers, the muon (μ) and the tau (τ). Less well-known examples are the particles that make up the core of the atom, the quarks. There are four fundamental forces in nature that act on the elementary particles. These forces and the known elementary particles are represented in Figure 1.1.

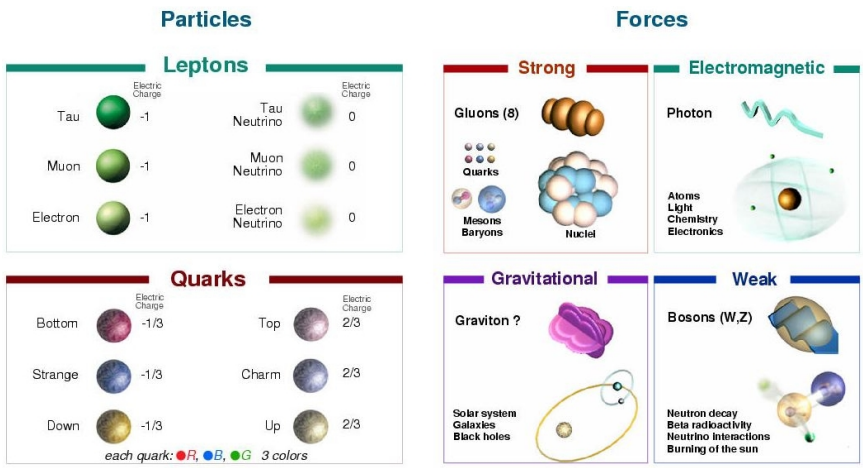


Figure 1.1: Classification of the elementary particles according to the Standard Model and the interactions displayed in nature (Gravity is not included in the Standard Model. It is described by the Einstein general relativity). In the first column are the three different generations of fermions. In the second and third ones are the spin-1 gauge fields. The particle drawings are simple artistic representations.

1.1 Standard Model of Particle Physics

The Standard Model of Particle Physics (SM) is a theory that includes three of the four fundamental forces (with the exception of Gravity) and it classifies the elementary particles in a coherent scheme. According to the SM, elementary particles are classified in two types:

- **Fermions** have half-integer spin and are described by Fermi-Dirac statistics. They follow the Pauli exclusion principle, which establishes that two identical fermions cannot be in the same quantum state at the same time.
  - Quarks: there are three families of quark doublets, which consist of two quarks with the same flavour. Quarks interact via strong, weak and electromagnetic forces.
  - Leptons: are grouped in three families of doublets, in which every lepton has its own associated neutrino. Leptons interact via the electromagnetic and the weak forces, while neutrinos only interact via the weak force.
- **Bosons** have integer spin and are described by the Bose-Einstein statistics. Inside this family are the gauge bosons (associated to the gauge fields), which mediate the fundamental interactions. They are:
  - 8 massless gluons (g) mediate the strong interaction.
  - 1 massless photon (γ) mediates the electromagnetic interaction.
  - 3 massive bosons ($W^{\pm}$ and Z^0) mediate the weak interaction.

The graviton is another hypothetical gauge boson that could mediate the force of Gravity.

According to the SM, every fermion has an associated anti-particle with the same features but with opposite electric charge. A summary of the properties of fermions and bosons of the SM is found in Table 1.1, Table 1.2 and Figure 1.2.

quark	mass (GeV)	charge
u	0.0023	+2/3
d	0.0048	-1/3
c	1.275	+2/3
s	0.095	-1/3
t	173.21	+2/3
b	4.18	-1/3

lepton	mass (GeV)	charge
e	$0.511{\cdot}10^{-3}$	-1
ν_e	~ 0	0
μ	0.106	-1
ν_μ	~ 0	0
τ	1.777	-1
ν_τ	~ 0	0

Table 1.1: Properties of the elementary particles with spin 1/2 according to the SM.

interaction	Theory	gauge boson	mass (GeV)	charge	spin
electromagnetic	QED	γ	0	0	1
weak	EW	Z	91.188 ± 0.002	0	1
		$W^{\pm}$	80.385 ± 0.015	± 1	1
strong	QCD	g	0	0	1
		Higgs	125.7 ± 0.4	0	0

Table 1.2: Properties of the gauge bosons according to the Standard Model. These particles are responsible for the fundamental interactions between fermions.

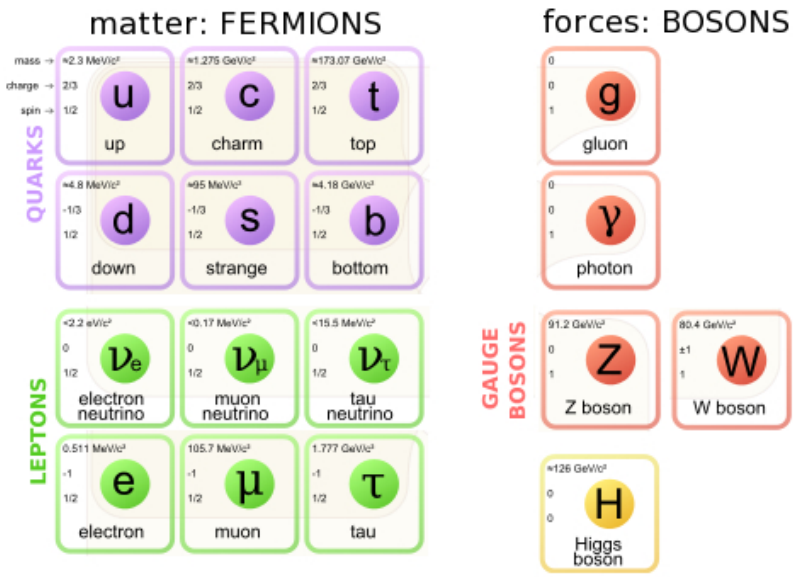


Figure 1.2: Classification of the elementary particles according to the Standard Model. In the first, second and third columns are the three different generations of fermions. In the fourth one are the integer-spin gauge fields. The SM describes the interaction between particles as gauge quantum fields that interact via gauge bosons and it is based on the symmetry group $SU(U)_C \otimes SU(2)_L \otimes U(1)_Y$

All the information about the dynamics of the fundamental particles and their interactions are encrypted in the Lagrangian, $\mathcal{L}$, which is a scalar function of the quantum fields.

1.1.1 Quantum Chromodynamics

Quantum Chromodynamics (QCD) is a quantum field theory that describes the strong interaction, i.e. interaction between quarks and gluons that make up the nucleus of atoms. The gluon is the mediator of this interaction. *Chromos* means "colour" in Greek. This is a suitable name for a theory that introduces a new quantum number called "colour charge" assigned to the quarks. Colour has three degrees of freedom: red, blue and green. The QCD Lagrangian is:

$$\mathcal{L}_{QCD} = \bar{\psi}(\mathrm{i}\gamma^{\mu} D_{\mu} - m)\psi - \frac{1}{4} G^{\alpha}_{\mu\nu} G^{\mu\nu}_{\alpha} \tag{1.1}$$

where ψ is the quark field with mass m, D_{μ} is the gauge covariant derivative associated to the gluonic field A_{μ}, $G_{\mu\nu}$ is the gluonic field tensor and T_a is a matrix associated to the eight Gell-Mann matrices as $T_a = 1/2\lambda_a$ (with $a = 1, 2, ...8$):

$$D_{\mu} = \partial_{\mu} + \mathrm{i} g_s A^a_{\mu} T_a \qquad G_{\mu\nu} = \partial_{\mu} A^a_{\nu} - \partial_{\nu} A^a_{\mu} - \mathrm{i} g_s [A_{\mu}, A_{\nu}] \tag{1.2}$$

defined in this way for a coupling constant $g_s = \sqrt{4\pi\alpha_a}$.

$\mathcal{L}_{QCD}$ is invariant under transformations of the non-Abelian $SU(3)$ group, leading to the conservation of the colour quantum number. A generic colour symmetry transformation, with arbitrary parameter $\theta_a(x)$ and structure constants f_{abc}, would transform the quark field and the gluonic field as:

$$\psi \longrightarrow \mathrm{e}^{-\mathrm{i} g_s T^a \theta_a(x)} \psi \tag{1.3}$$

$$G^a_{\mu\nu} \longrightarrow G^a_{\mu\nu} - \partial_{\mu}\theta_a - g_s f_{abc} \theta_b G^c_{\mu} \tag{1.4}$$

If we define $SU(x)$ as:

$$SU(x) = \mathrm{e}^{-\mathrm{i} g_s T^a \theta_a(x)} \tag{1.5}$$

and we fix $x = 3$ (since colour charge has 3 degrees of freedom), we can say that $SU(x)$ is a non-Abelian $SU(3)$ group which leads to the **conservation of the colour charge**.

The coupling constant g_s is related to the QCD coupling through the equation $g_s = \sqrt{4\pi\alpha_s}$. At small distances between the initial objects (large energies), quarks and gluons behave as quasi-free particles because of the small coupling, hence, QCD can be described perturbatively. At large distances (low energies) the coupling is large, therefore, the quark confinement within subatomic particles is guaranteed. This leads quarks and gluons to form bound states.

1.1.2 Quantum Electrodynamics

Quantum Electrodynamics (QED) is a gauge field theory that describes the electromagnetic force, i.e. describes the interaction between photons and fermions. The photon is the mediator of this interaction. The QED Lagrangian is:

$$\mathcal{L}_{QED} = \bar{\psi}(\mathrm{i}\gamma^{\mu} D_{\mu} - m)\psi - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} \tag{1.6}$$

where ψ is the QED quantum field with mass m, D_{μ} is the gauge covariant derivative, $F_{\mu\nu}$ is the electromagnetic tensor and A_{μ} is the electromagnetic field:

$$D_{\mu} = \partial_{\mu} + \mathrm{i}qA_{\mu} \qquad F_{\mu\nu} = \partial_{\mu} A_{\nu} - \partial_{\nu} A_{\mu} \tag{1.7}$$

defined in this way for a particle with electrical charge q.

$\mathcal{L}_{QED}$ is invariant under local transformations. Therefore, applying a local phase rotation, with arbitrary parameter $\alpha(x)$, over charged fields and over their covariant derivatives must keep the $\mathcal{L}_{QED}$ invariant:

$$\psi \longrightarrow \mathrm{e}^{-\mathrm{i}q\alpha(x)}\psi \tag{1.8}$$

$$D_{\mu}\psi \longrightarrow \mathrm{e}^{-\mathrm{i}q\alpha(x)} D_{\mu}\psi \tag{1.9}$$

Hence, the electromagnetic field will be invariant if it accomplishes that:

$$A_{\mu} \longrightarrow A_{\mu} - \frac{1}{q}\partial_{\mu}\alpha(x) \tag{1.10}$$

If $U(x)$ is definied as:

$$U(x) = \mathrm{e}^{-\mathrm{i}q\alpha(x)} \tag{1.11}$$

and we fix $x = 1$ (the photon is alone in its family, since it is a spin-1 field), we can say that $U(x)$ is an Abelian $U(1)$ group. This is the global symmetry $U(1)$ which leads to the **conservation of the electromagnetic charge**.

The addition of a mass term ($1/2\ m_{\gamma}^2 A_{\mu} A^{\mu}$) in Equation 1.6 would lead to a violation of the gauge symmetry, which would not be compatible with the observed null mass of the photon.

1.1.3 Electroweak Unification

The Electroweak Theory (EW) unifies the two theories that describe electromagnetic and weak interactions. $SU(2)_L \otimes U(1)_Y$ is the gauge symmetry group of the electroweak interaction. $SU(2)_L$ only affects the left-handed fields

and it is defined in the T space (weak isospin charge). $U(1)_Y$ is defined in the Y space (weak hypercharge, constructed from the electric charge and the third component of the weak isospin charge):

$$T_i = \frac{1}{2}\sigma_i \qquad i = 1,2,3 \quad \sigma_i = \text{Pauli matrices} \tag{1.12}$$

$$\frac{Y}{2} = Q - T_3 \tag{1.13}$$

The fields related to quarks and leptons are divided in left-handed and right-handed singlets or doublets as follows:

$$\begin{array}{lll} \begin{pmatrix} u \\ d \end{pmatrix}_L & u_R & d_R \\ \begin{pmatrix} \nu_\ell \\ \ell^- \end{pmatrix}_L & \nu_{\ell R} & \ell^-_R \end{array} \tag{1.14}$$

where ℓ denotes lepton, ν neutrino, u quark up and d quark down fields. The EW Lagrangian is:

$$\mathcal{L}_{EW} = \sum_{k=1}^{3} \left[\mathrm{i}\bar{\psi}_k \gamma^\mu D_\mu \psi_k\right] - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} - \frac{1}{4} W^\alpha_{\mu\nu} W^{\mu\nu}_\alpha + \mathcal{L}_{GF} + \mathcal{L}_{FP} \tag{1.15}$$

where D_μ refers to the covariant derivative, B_μ is the gauge field identified with γ and $\vec{W}_\mu$ is the gauge boson field associated to the $W^\pm$ and Z bosons. $\mathcal{L}_{GF}$ and $\mathcal{L}_{FP}$ are the Gauge Fixing and Faddeev-Popov Lagrangian terms, respectively, needed to quantize the gauge theory. The covariant derivative and the vector fields follow this notation:

$$\begin{aligned} D_\mu &= \partial_\mu - \mathrm{i}g\vec{T}\cdot\vec{W}_\mu - \mathrm{i}g'\frac{Y}{2}B_\mu \\ B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu \\ W^i_{\mu\nu} &= \partial_\mu W^i_\nu - \partial_\nu W^i_\mu + g\vec{W}_\mu \times \vec{W}_\nu \end{aligned} \tag{1.16}$$

The physical gauge bosons are related to the vector fields as follows:

$$\begin{aligned} W^\pm_\mu &= \frac{1}{\sqrt{2}}\left(W^{(1)}_\mu \mp W^{(2)}_\mu\right) && \text{associated to } \mathrm{W}^\pm \\ Z_\mu &= W^{(3)}_\mu \cos\theta_W - B_\mu \sin\theta_W && \text{associated to } \mathrm{Z}^0 \\ A_\mu &= W^{(3)}_\mu \sin\theta_W + B_\mu \cos\theta_W && \text{associated to } \gamma \end{aligned} \tag{1.17}$$

where θ_W is the Weinberg mixing angle. The first identity is a charged current and it involves $W^{\pm}$ bosons. The second and the third are neutral currents involving the Z^0 boson and the γ.

To recover QED, the identification of the A_μ term with the interaction term of QED gives the relation between the constant couplings:

$$q_e = g \sin\theta_W = g' \cos\theta_W \tag{1.18}$$

This identity associates $SU(2)_L \otimes U(1)_Y$ couplings with the electromagnetic one and it provides the unification of the weak and electromagnetic interactions.

1.1.4 Spontaneous Symmetry Breaking

The EW and QCD Lagrangians allow only massless fields, as in the cases of photon and gluon fields. However, $W^{\pm}$ and Z^0 are heavy bosons. In order to generate masses, a mass term must be added to the Lagrangian. This addition introduces a breaking of the gauge symmetry. Moreover, to preserve normalizability, a symmetric Lagrangian is also required. One solution to deal with both dilemmas is the mechanism called Spontaneous Symmetry Breaking (SSB). This method keeps the Lagrangian symmetric but the ground state (the vacuum) is not invariant under gauge transformations. Both are achieved by introducing a new term in the Lagrangian with a potential V:

$$\mathcal{L}_{SSB} = \left(D_\mu \Phi\right)^\dagger \left(D^\mu \Phi\right) - V\left(\Phi\right) \tag{1.19}$$

$$\begin{aligned} V\left(\Phi\right) &= \mu^2 \Phi^\dagger \Phi + \lambda \left(\Phi^\dagger \Phi\right)^2 \qquad \lambda > 0 \ ; \ \mu^2 < 0 \\ \Phi(x) &= \begin{pmatrix} \Phi^\dagger(x) \\ \Phi^0(x) \end{pmatrix} \end{aligned} \tag{1.20}$$

where Φ is a $SU(2)_L$ doublet of complex scalar fields. The $\mathcal{L}_{SSB}$ is invariant under local $SU(2)_L \otimes U(1)_Y$ transformations. This doublet has an infinite set of degenerate states with minimum energy:

$$|\Phi|^2 = \Phi^\dagger \Phi = -\frac{\mu^2}{\lambda} \equiv \frac{\upsilon^2}{2} \tag{1.21}$$

where υ is the vacuum expectation value. All these states keep the $U(1)$ symmetry (this corroborates that the electric charge is conserved) but $SU(2)_L \otimes U(1)_Y$ symmetry is spontaneously broken to the electromagnetic subgroup

$U(1)_{QED}$. Apart from the massless photon, three massless states should appear (although they can be absorbed by the gauge field as we will see) as the Goldstone Theorem suggests:

$$\Phi(x) = \frac{1}{\sqrt{2}} \exp\left\{ \mathrm{i}\frac{\vec{\sigma}}{2}\frac{\vec{\theta}(x)}{\upsilon} \right\} \begin{pmatrix} 0 \\ \upsilon + H(x) \end{pmatrix} \tag{1.22}$$

with $\vec{\theta}$ and $H(x)$ the four real fields. Introducing this field in Equation 1.19, quadratic terms appear for the M_W, M_Z and M_H which means that the gauge bosons have acquired mass:

$$M_W = \frac{g\upsilon}{2} \qquad M_Z = \frac{M_W}{\cos\theta_W} = \frac{\upsilon}{2}\sqrt{g^2 + g'^2} \qquad M_H = \upsilon\sqrt{2\lambda} \tag{1.23}$$

On the contrary, the photon remains massless because $U(1)_{QED}$ is an unbroken symmetry of the SSB model. A neutral boson compatible with the Higgs Boson was announced on 4 July 2012 by ATLAS and CMS Collaborations [1, 2], corroborating a prediction that has been waiting for 50 years to be confirmed.

1.2 Physics beyond the Standard Model

We know that the SM is a successful theory that describes accurately a broad range of phenomena. However, physicists would like to find a deeper explanation for some of the SM parameters. Some unexplained phenomena in astronomy might require extensions of the SM. These shortcomings motivate physicists to keep searching for a more complete theory of particles and fundamental interactions, achieving an "Unified Field Theory" or a "Theory of Everything" able to explain all physical phenomena. The most notable problems are: interpreting the values of the SM parameters, clarifying the Electroweak Symmetry Breaking (EWSB) or explaining the gauge hierarchy problem.

1.2.1 Open issues

The quantum theory used to describe the micro world, and the general theory of relativity used to describe the macro world, are difficult to merge into the same theory. So, **Gravity** is not included in the SM. Since it is much weaker than the other three fundamental forces (this is the **gauge hierarchy problem**), it has a negligible influence on the behaviour of elementary particles. It is the dominant force at the macroscopic scale and it is always attractive.

Dark matter is a kind of matter that can not be detected directly but its existence can be inferred by the gravitational pull that it exerts on the luminous material that we can see. Dark matter is supposed to make up the 85% of the total matter in our galaxy and the 25% of the energy of the universe. Although the SM does not have a good candidate, many aspirants for dark matter have been suggested, such as the following non-baryonic particles: Axion (particles with a tiny amount of mass) and Weakly Interacting Massive Particles (WIMPs). Currently, dark matter is one of the biggest challenges for modern physics.

The expansion of the universe is accelerating. This expansion is attributed to **dark energy**, which is an unknown form of energy that contributes 70% of the total energy in the universe.

Other issue with no explanation in the SM are the **neutrino masses**. The SM describes neutrinos as massless. Many neutrino-oscillation experiments have shown that neutrinos can change their flavour, oscillating as they travel through the space. To make it possible for neutrinos to oscillate, they must have slightly different masses and therefore, at least two neutrinos must have non-zero mass.

Another question with no answer so far is the dominance of matter over antimatter, better know as **baryon asymmetry**. Today, ordinary matter accounts for just 4% of the total energy in the universe. Breaking the CP-symmetry would imply that matter and antimatter behave differently. Several experiments have demonstrated that weak interactions, such as beta decay, are not symmetric with respect to CP, as happens in neutral kaons ($K^0 = s\bar{d}$) and B mesons ($B = b\bar{d}$) decays. One possibility to explain a matter and antimatter imbalance is a CP-violation phase in the CKM matrix, but it does not explain the observed asymmetry.

One of the collider based measurements that has generated a lot of interest is the **$t\bar{t}$ forward-backward asymmetry** (this is the subject of Section 3) because experiments at the Tevatron measured a discrepancy of it with the SM.

1.2.2 Extensions of the Standard Model

Theoretical physicists have proposed theories beyond the Standard Model (BSM) which predict new particles at the TeV scale. These theories aim to improve the stability in the Higgs boson mass (light hierarchy or fine-tuning problem) and aim to develop a unified theory of fundamental forces (gauge coupling unification) or a theory of quantum gravity.

One of the possible extensions of the SM is **supersymmetry** (SUSY). It introduces a new symmetry between fermions and bosons, so that every fermion in the SM has a corresponding bosonic supersymmetric partner (s-particle) and vice-versa. This superpartner has opposite spin-statistics but the same quantum numbers as the SM particle. The lightest superpartner particle is a viable candidate for dark matter.

BSM models [3, 4] with **extra dimensions** were proposed for solving the hierarchy problem, independent of supersymmetry or technicolor. In the first models the SM fields are confined to a 4-dimensional membrane in the extra dimensions while gravitons can freely propagate in the new dimensions. They postulate that the Planck scale $M_{Plank} \sim 1/\sqrt{G_N}$ is no longer a fundamental scale and its enormity is a consequence of the large size of the new dimensions, that are connected to a stronger Gravity. The extra dimensions should be compact with a finite extension. Every particle able to spread in these compact extra dimensions with size R should have an infinite Kaluza-Klein tower of heavy copies, each one with mass $M_n^2 = n^2/R$, separated between them by $1/R$. One of the most famous examples of these extra dimension theories are the Warped or Randal-Sundrum Models (RS) [5, 6], where they propose that the SM is located in a single 3-brane embedded in $n \geq 2$ dimensions.

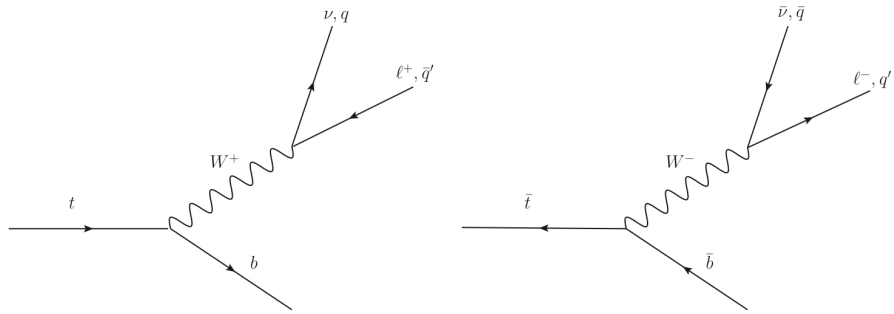
According to **composite Higgs models** (CHM), the gauge hierarchy problem can be solved by making the Higgs boson a composite particle. It postulates that a light Higgs appears as a pseudo-Goldstone boson resulting from the spontaneously breaking of a new chiral symmetry. Then, the quadratic divergence in the Higgs boson mass are cancelled.

Each of these proposals has some attraction for physicists, since all of them are testable in experiments with colliders. So far, however, no new particles have been found.

1.3 Top quark physics

The CDF [7] and D0 [8] collaborations discovered the top quark at the Tevatron $p\bar{p}$ collider in 1995. This quark is the most massive of all elementary particles in the SM, about 40 times heavier than the b quark. Owing to its very short lifetime ($\tau_t \sim 0.5 \cdot 10^{-24}$ s) and high mass, the top quark decays before it can hadronize ($\tau_t < \tau_{QCD} \sim 3 \cdot 10^{-24}$ s). Therefore, it is possible to access its spin states. Due to its mass, the top quark couples to the Higgs boson more strongly, making its physics an important channel to study in Higgs analyses. The top quark plays an important role in several BSM physics scenarios, as it will be described in Section 3.

Because of the top quark decays before hadronization, detectors only measure its decay products. Moreover, its properties are not hidden by hadronization effects and then, it provides a very clean source for fundamental information. Figure 1.3 shows the Feynman diagrams of the decay. The top quark decays in a W boson (both, top quark and W boson have the same sign in their electrical charge) and a b quark. At the same time, the W boson can decay either in a lepton plus its associated neutrino or in a light quark pair.



(a) Top quark decay

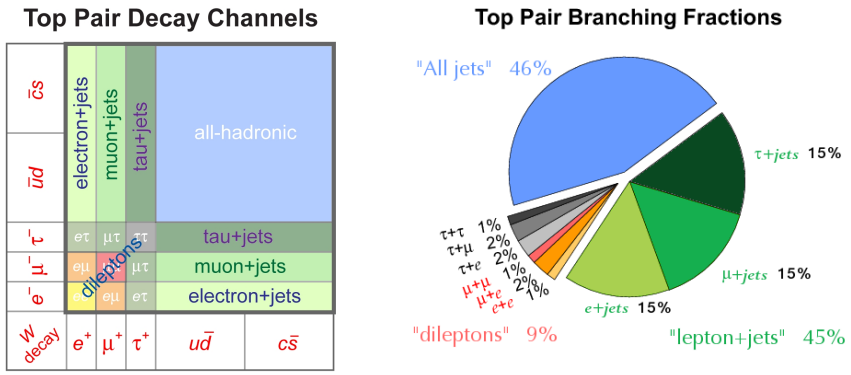
(b) Anti-top quark decay

Figure 1.3: Top and anti-top quarks decay chains. $t \longrightarrow W^+b$ and $\bar{t} \longrightarrow W^-\bar{b}$.

According to the decay of the two W bosons, that are daughters of the top and anti-top quarks, the $t\bar{t}$ processes are classified into three channels:

- **fully hadronic**: both W^+ and W^- decay hadronically.
- **dileptonic**: both W bosons decay leptonically.
- **semileptonic**: one W boson decays leptonically and the other one hadronically.

These channels and their probabilities are summarized in Figure 1.4. A semileptonic decay is depicted in Figure 1.5.



(a) Channels (b) Branching ratios

Figure 1.4: The $t\bar{t}$ pair decay channels and branching ratios.

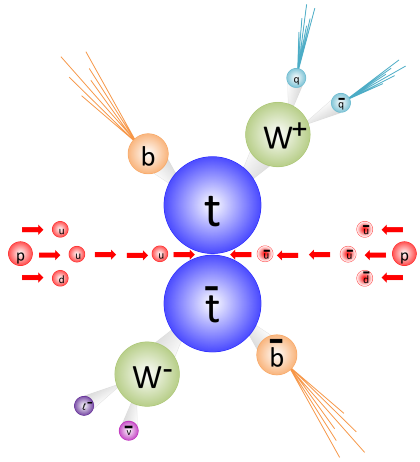


Figure 1.5: Top quark pair production, in semileptonic channel, in which $W^- \longrightarrow \ell^- \bar{\nu}$ and $W^+ \longrightarrow q\bar{q}$.

In conclusion, the top quark is a unique object and there are several reasons that make it an interesting object to look into. Therefore, many experiments have been studying the top quarks for years. Additionally, the SM does not provide answer to why it is alone in having a mass of approximately the EWSB scale. Moreover, it is a challenge to study the complex final state produced by a $t\bar{t}$ pair. Its daughters hadronize, giving place to a final state with missing energy, leptons and jets.

2

Boosted objects and jet substructure

2.1 Jet algorithms

Most hard scattering processes at hadron colliders produce final state particles which carry colour charge, such as quarks or gluons. The outcome of these collisions is the production of highly collimated sprays of energetic hadrons, which are detected in the detector. In order to reconstruct such collisions, several hadrons are clustered in the so called *jets*, which may be defined according to different algorithms [9, 10, 11, 12, 13]. The jet direction $(\eta - \phi)$ and momentum transverse to the beam (p_T) are mapped on the partons. However, jets are composite objects and their internal substructure and their masses contain additional information.

Jet reconstruction algorithms can be split in two categories. On the one hand, **Cone algorithms**, where most of the energy flow of the event is contained in a given geometric conic region. On the other hand, **Sequential recombination algorithms**. At the LHC the experiments rely on the latter, where jet constituents are combined according to the minimal distance between two jet constituents, d_{ij}, and the distance between a constituent and the beam, d_{iB}.

The distance ΔR between two particles i and j is defined as:

$$\Delta R_{i,j} = \sqrt{\Delta y_{i,j}^2 + \Delta \phi_{i,j}^2} \qquad R = \sqrt{y^2 + \phi^2} \tag{2.1}$$

where Δy and $\Delta\phi$ are the differences in rapidities [1] and azimuthal angles respectively. The most commonly used jet algorithms use the following distances:

$$d_{ij} = min(p_{T,i}^{2n}, p_{T,j}^{2n}) \times \frac{\Delta R_{i,j}^2}{R^2} \tag{2.2}$$

$$d_{iB} = p_T^{2n} \tag{2.3}$$

In jet reconstruction, **clustering** refers to the ordering and structure of the two-subjet recombinations made during the reconstruction process. At each step of the clustering process, the two particles with the smallest distance d_{ij} are merged in a single object.

In inclusive jet reconstruction, as in the LHC, clustering continues until the smallest d_{ij} is found to be d_{iB}, in which case the object i is tagged as jet. The value of n in Equation 2.2 gives rise to three different jet algorithms:

- $n = -1 \Rightarrow$ **anti-k_t algorithm** [9]: provides jets that are defined primarily by the highest-p_T constituent, yielding stable and approximately circular jets in the $\eta - \phi$ plane.
- $n = 0 \Rightarrow$ **Cambridge-Aachen algorithm** [10, 11]: provides a purely angular distance.
- $n = 1 \Rightarrow$ **k_t algorithm** [12, 13]: the clustering proceeds in a physically meaningful order, from the softest to the hardest emission.

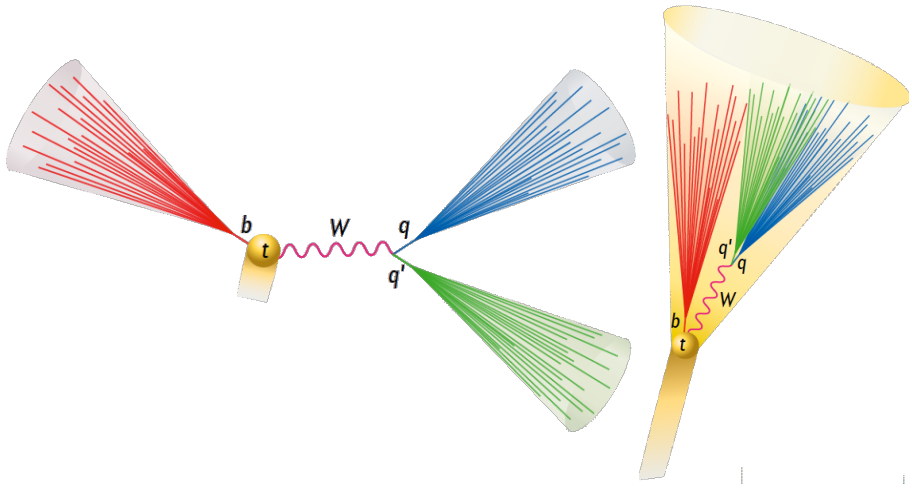
These three algorithms represent the most used infrared and collinear-safe algorithms available for hadron-hadron collider physics so far. They are implemented within the framework of the FastJet software [14, 15].

At the LHC several jet algorithms are used but the anti-k_t algorithm is the default one, because it reduces the pile-up [2], the calibration of the resulting jets is easy, and it has excellent energy resolution. ATLAS uses two default values for R: $R = 0.4$ for small-R jets and $R = 1.0$ for large-R jets.

[1] The rapidity is explained in Section 4.3.
[2] Look at Appendix A.1 for more details.

2.2 Boosted objects

The large centre-of-mass energy at the LHC enables the production of Lorentz-boosted heavy particles, such as W, Z and $Higgs$ bosons, top quarks, and even new particles. If their momentum is much larger than their rest mass, *boosted* objects are reconstructed as a single jet. In Figure 2.1 the topologies of a top quark produced approximately at rest is compared to that of a boosted top quark. Notice that the decay products of the boosted top quark (shown in Figure 2.1(b)) are so close together that they can not be distinguished accurately. In this case, $p_T \gg 2m$, it is imposible to resolve all subjets so, reconstruction as a single jet (large-R jet) is the only option.



(a) Rest (b) $p_T >> m$

Figure 2.1: Decay topologies for top quarks produced (a) at rest and (b) at $p_T >> m$, leading to a boosted $t\bar{t}$ event. Notice that the decay products for the boosted top quark merge and can not be resolved individually.

2.3 Jet substructure

The study of the internal structure of jets goes beyond the 4-momentum description of a single parton and yields new approaches for testing QCD and for searching for new physics in hadronic final states. The study of boosted object production is the main motivation for studying the internal substructure of jets in boosted regime at the LHC.

To characterize jet substructure and distinguish massive boosted objects from gluons or light quarks, a large number of observables is used. In the following, two of the simplest are discussed: jet mass and k_t splitting scales.

Jet mass: it is defined as the invariant mass of the 4-momentum sum of all jet components. Since this sum depends on the algorithm, jet mass is also algorithm-dependent. The boosted jet has $m_{jet} \sim m_{top}$ while QCD jets $m_{jet} \sim \alpha_s p_T R$.

k_t splitting scales: these are defined by reclustering the constituents of a jet with the k_t recombination algorithm, which tends to combine "soft" constituents to form "harder" objects. The k_t splitting scales $\sqrt{d_{ij}}$ are defined as the k_t distance in the clustering process (step that merges j subjets into i), as it is shown in equation 2.4:

$$d_{ij} = min(p_{T,i}^2, p_{T,j}^2) \times \frac{\Delta R_{i,j}^2}{R^2} \tag{2.4}$$

where $\Delta R_{i,j}$ is the distance between the two subjets in $\eta - \phi$ space. Using this definition, the last step of the reclustering in the k_t algorithm is related to $\sqrt{d_{12}}$, where the remaining two subjets are combined into the final jet. Analogously, $\sqrt{d_{23}}$ characterizes the splitting scale in the second-to-last step, where there are three jets and are merged in two, and so on. Both parameters, $\sqrt{d_{12}}$ and $\sqrt{d_{23}}$, are used to distinguish heavy particle decays from light quark or gluon jets. If the jet contains the decay of a boosted heavy object, the expected value of $\sqrt{d_{12}}$ is $m_{particle}/2$, whereas jets from the parton shower of gluons and light quarks (inclusive jets) tend to have smaller values ($\leq 20GeV$).

2.4 Jet grooming algorithms

Jet grooming seeks to get rid of softer components in a jet from an underlying event (UE) or pile-up (PU) and to retain constituents from the hard scatter behind. The jet mass resolution is improved after the application of a grooming technique [16, 17, 18] and it is very useful to look for boosted objects contained in a large-R jets. Three grooming algorithms are widely used:

- **Mass-drop filtering:** this procedure [19] isolates concentrations of energy inside a jet by identifying subjets that are relatively symmetric, each of the two with mass smaller than the original jet mass. This algorithm uses the C/A algorithm and it has two stages:

– **Mass-drop and symmetry.** The last step of the C/A clustering is undone. The jet is split into two subjets, $subj1$ and $subj2$, and they are ordered so that $m_{subj1} > m_{subj2}$. After this splitting, the most massive subjet is required to have smaller mass than the original jet, given by the algorithm parameter μ_{frac}:

$$m_{subj1}/m_{jet} < \mu_{frac} \tag{2.5}$$

Moreover, the splitting must be relatively symmetric following this requirement:

$$\frac{\min(p^2_{T,subj1}\ ,\ p^2_{T,subj2})}{m^2_{jet}} \cdot \Delta R^2_{1,2} > y_{cut} \tag{2.6}$$

where $\Delta R_{1,2}$ is the opening angle between the two subjets, and y_{cut} defines the energy sharing between the two subjets in the original jet. If a jet does not accomplish both, mass-drop and symmetry requirements, it is discarded. This method is illustrated in Figure 2.2(a). Previous $H \to b\bar{b}$ studies [19] found that the optimal value for identifying two-body decays is $y_{cut} = 0.09$ and for discrimitating $H \to b\bar{b}$ from background is $\mu_{frac} = 0.67$.

– **Filtering.** The constituents of $subj1$ and $subj2$ are clustered again using a C/A algorithm with R_{filt} as main parameter, which is defined as:

$$R_{filt} = \min(0.3\ ,\ \Delta R_{1,2}/2) \qquad R_{filt} < \Delta R_{1,2} \tag{2.7}$$

After this, the jet is filtered, i.e., all the constituents outside the three hardest subjets are discarded. This is shown in Figure 2.2(b).

▶ **Trimming:** this method [20, 21] uses k_t algorithms and it is aimed to create subjets of size R_{sub} starting from the initial jet. Subjets with $p_{T,i}/p_{T,jet} < f_{cut}$ are removed (f_{cut} is a parameter of the method and typically it is a few per cent), leading to the so called trimmed jet. This process is shown in Figure 2.3.

This algorithm has a great advantage: the contamination in the reconstructed jet due to PU, multiple parton interactions (MPI) and initial state radiation (ISR) is often much softer than the outgoing partons (associated with the hard-scatter) and their final-state radiation (FSR). Removing the softer components of the final jet results in (a) the suppression of radiation from PU, MPI and ISR and (b) the discarding of a small part of the hard-scatter decay products and FSR.

- **Pruning:** this algorithm [22] is based on the same ideas as trimming (it removes constituents with small p_T) but also modifies the jet substructure to improve heavy particle identification (reducing the systematic effects that obscure their identification), applying an additional requirement over wide-angle radiation. It is invoked in every step of the jet recombination algorithm (either C/A or k_t, since the procedure is independent of the jet finder). It is based on a decision over the kinematic variables to determine whether or not a constituent is considered in the jet reconstruction, but it does not require the reconstruction of subjets. The definition of the pruning procedure (which is illustrated in Figure 2.4) is the following:

  1) Starting with a jet found by any of the jet finding algorithm, either C/A or k_t recombination algorithm is run on the constituents.

  2) Every recombination step, with constituents $subjet1$ and $subjet2$ fulfilling $p_{T,1} > p_{T,2}$, must satisfy the requirement on either z_{cut} or R_{cut}, given in this equation:

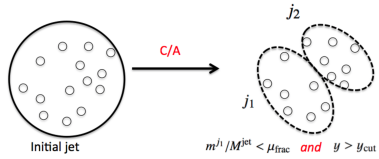
$$\frac{p_{T,1}}{p_{T,1}+p_{T,2}} > z_{cut} \quad ; \quad \Delta R_{1,2} < R_{cut} \cdot \frac{2m_{jet}}{p_{T,jet}} \tag{2.8}$$

  3) $subjet1$ and $subjet2$ are merged if one or both above conditions are met, otherwise, two subjets are not merged and the softer one ($subjet2$) is discarded.

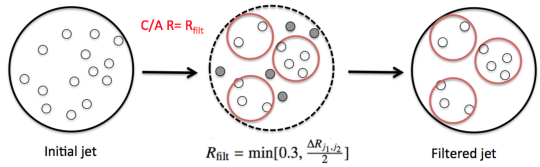
2.5 Summary

The combination of groomed large-R jet reconstruction with a substructure analysis is a powerful tool to identify and reconstruct the most energetic gauge bosons and top quarks produced by the LHC.

A complete description is found in the reports of the BOOST conference [23, 24, 25, 26]. These techniques are used heavily in searches for new physics in $t\bar{t}$ resonances, which is the subject of Chapter 7. In Chapter 8 the first measurement of the charge asymmetry in boosted $t\bar{t}$ production is presented.



(a) Mass-drop and symmetric splitting criteria (first stage).



(b) Filtering (second stage).

Figure 2.2: Diagram depicting the two stages of the **mass-drop filtering** procedure [17].

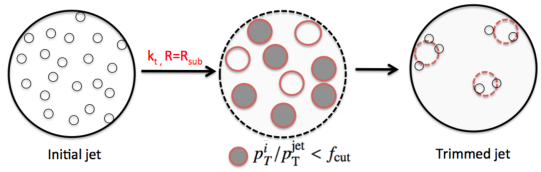


Figure 2.3: Diagram of the jet **trimming** procedure [17].

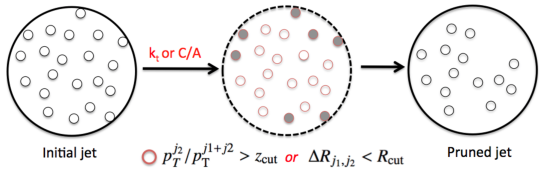


Figure 2.4: Diagram of the jet **pruning** procedure [17].

3

The $t\bar{t}$ charge asymmetry

3.1 Motivation

The top quark is the most massive of the fundamental particles known so far. It was discovered in 1995 at the Tevatron, a $\sqrt{s} = 1.96\ TeV$ $p\bar{p}$ collider built in Fermilab [27], by the CDF [7] and D0 [8] Collaborations. It is the only quark in the SM that decays before it hadronizes, the only fermion whose mass is close to the electroweak scale, and the only quark with large Yukawa coupling to the Higgs boson. Its properties are of great importance in many of the current BSM theories.

Popular BSM models address some of the current open questions, such as the baryon asymmetry of the universe, the nature of dark matter or a mechanism to naturally stabilize the Higgs boson mass at its observed value of approximately 125 GeV. For some of these solutions to work, the new physics scale has to be around the TeV scale, within reach of collider experiments.

Many models that propose alternative mechanisms for electroweak symmetry breaking (EWSB) incorporate new heavy particles that can decay to $t\bar{t}$ states. Some examples include strong EWSB models such as topcolour-assisted technicolour [28] (TC2) and Composite Higgs [29, 30, 31, 32, 33] scenarios. Randall-Sundrum (RS) [34, 35, 6, 36] models, with warped extra dimensions, form an additional class of models that predict heavy particles that decay to $t\bar{t}$ pairs.

Apart from direct searches for new resonances in $t\bar{t}$ production, as it is described in Chapter 7, experiments look for signatures of new physics by performing precise measurements of top quark production and decay.

The Tevatron experiments reported a first measurement of the forward-backward asymmetry (A_{FB}) in top quark pair production that had considerable tension with respect to the SM prediction. At the LHC a related charge asymmetry (A_C) has been measured. In this chapter, an overview is given of the A_{FB} and the A_C at the Tevatron and the LHC respectively.

3.2 The forward-backward asymmetry at the Tevatron

In a hadron collider, top quark pairs are produced by quark-antiquark annihilation ($q\bar{q} \to t\bar{t}X$, with q denoting primarily u, d quarks and X denoting possible additional jets) and gluon-gluon fusion (gg). The Tevatron is a $p\bar{p}$ collider, where quark-antiquark annihilation is the dominant process.

QCD predicts that top quark pair production in $q\bar{q}$ collisions is forward-backward symmetric at leading order (LO). However, a non-zero asymmetry appears at higher orders. The top quark is preferentially emitted in the direction of the incoming light quark and the anti-top quark in the direction of the incoming anti-quark. At the Tevatron, interactions between valence quarks dominate $t\bar{t}$ production, therefore the direction of the incoming quark almost always matches with that of the initial proton.

The forward-backward asymmetry is defined in the $t\bar{t}$ rest-frame as:

$$A_{FB} = \frac{N(\Delta y > 0) - N(\Delta y < 0)}{N(\Delta y > 0) + N(\Delta y < 0)} \tag{3.1}$$

where $\Delta y = y_t - y_{\bar{t}}$, y denotes the rapidity of the top and anti-top quark and N refers to the number of events.

According to the SM, the main contribution to the asymmetry in Equation 3.1 arises at next-to-leading order (NLO) in QCD due to the interference of $\mathcal{O}(\alpha_s^3)$ terms in the cross section, which are odd under the interchange $t \leftrightarrow \bar{t}$ with the initial quarks fixed (from here the denomination of A_{FB} as "charge asymmetry", despite the fact that it does not have any relation with the charge conjugation symmetry C).

The interference of tree-level (Figure 3.1(a)) and one-loop (Figure 3.1(b)) diagrams for $q\bar{q} \to t\bar{t}$ generates a **positive** asymmetry. However, the interference of initial state radiation (Figure 3.1(c)) and final state radiation (Figure 3.1(d)) generates a **negative** asymmetry. The relative size of these contributions depends on the transverse momentum of the $t\bar{t}$ pair.

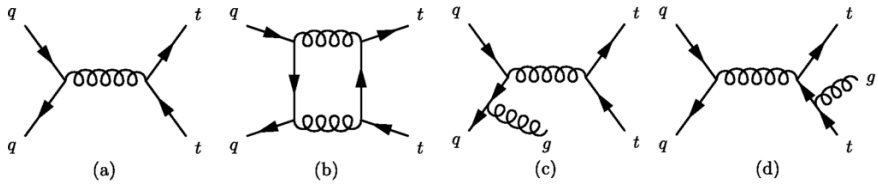


Figure 3.1: Sample Feynman diagrams contributing to $q\bar{q} \to t\bar{t}$ at (a) LO and (b)-(d) NLO in QCD.

The asymmetry generated in $q\bar{q} \to t\bar{t}X$ suffers from dilution due to the $gg \to t\bar{t}X$ subprocess. Other quark subprocesses, such as $s\bar{s}$ and $c\bar{c}$, do not significantly contribute because they have small cross sections and the parton density functions (PDFs) are the same for both, quarks and anti-quarks (both are sea quarks since they come from the initial proton and anti-proton sea, as it is shown in Figure 3.2).

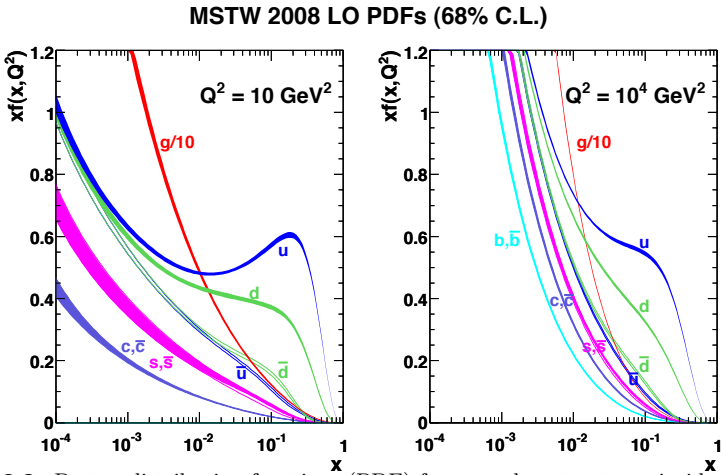


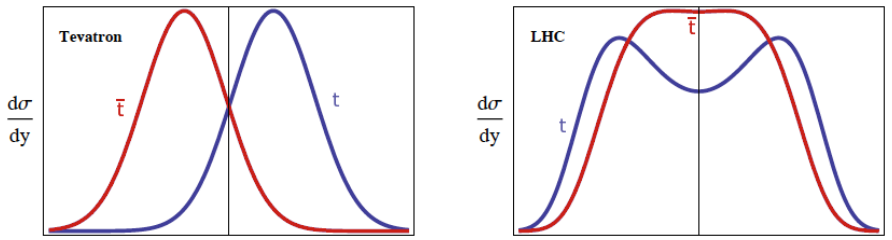
Figure 3.2: Parton distribution functions (PDF) for several parton types inside the proton and two values of the scale of the scattering process Q^2. This figure shows $x \cdot f(x, Q^2)$ versus x, where x is the momentum fraction of the proton.

In 2011 the A_{FB} was measured by D0 and CDF experiments. They found a discrepancy with respect to the SM prediction. The first CDF result was initially reported as a 3σ deviation from the SM. The latest result from CDF is $A_{FB}^{CDF} = 0.164 \pm 0.047$ [37] while D0 obtained $A_{FB}^{D0} = 0.106 \pm 0.030$ [38].

Since the two measurements are uncorrelated, it is possible to combine them as $A_{FB}^{comb} = 0.124 \pm 0.025$ [39]. In recent years the SM prediction has been refined to include weak, mixed QCD-weak and QCD-QED corrections [40]. The SM prediction with NLO EW and NNLO QCD corrections is $A_{FB}^{SM} = 0.095 \pm 0.007$ [41]. This result is compatible with the combined measurement within one standard deviation (1σ).

3.3 The $t\bar{t}$ charge asymmetry at the LHC

The LHC is a pp collider, where the dominant mechanism for $t\bar{t}$ production is expected to be the gg fusion process, while $t\bar{t}$ production via $q\bar{q}$ or qg is small. Since the initial state is symmetric, the A_{FB} is no longer a useful observable. However, due to the asymmetry in the production via $q\bar{q}$ and qg, QCD predicts at the LHC a small excess of centrally produced anti-top quarks while top quarks are produced, on average, at larger absolute rapidities. This is shown in Figure 3.3, where the rapidity distribution of top and anti-top quarks for the LHC and the Tevatron are compared.



(a) Tevatron (b) LHC

Figure 3.3: Rapidity distributions for top quarks (blue line) and anti-top quarks (red line). At the Tevatron (a), tops are in the forward region while anti-tops are in the backward region. At the LHC (b), top quarks are mainly in the forward or backward directions, while anti-top quarks are kept more in the central region. The difference in the rapidity of two particles is independent of Lorentz boosts along the beam axis.

The broader rapidity distribution for top quark can be understood as follows: for $t\bar{t}$ production via $q\bar{q}$ annihilation the valence quark carries, on average, a larger momentum fraction than the sea anti-quark. With top quarks preferentially emitted in the direction of the initial quarks in the $t\bar{t}$ rest frame, the boost into the laboratory frame drives the top quarks mainly in the forward or backward directions, while anti-top quarks are preferentially retained in the central region.

A charge asymmetry can be defined as:

$$A_C = \frac{N(\Delta|y| > 0) - N(\Delta|y| < 0)}{N(\Delta|y| > 0) + N(\Delta|y| < 0)} \tag{3.2}$$

where $\Delta|y| = |y_t| - |y_{\bar{t}}|$ and y denotes the rapidity of the top and anti-top quark.

The contributions of the Feynman diagrams are the same as for the A_{FB}. The main difference is the origin of the initial quarks. While at the Tevatron q and $\bar{q}$ are both valence quarks (both found typically in p and $\bar{p}$ respectively), at the LHC q is a valence quark and $\bar{q}$ is a sea anti-quark (both are originated from initial protons).

The SM prediction for the A_C is one order of magnitude smaller than that for the A_{FB} due to two effects:

- At $\sqrt{s} = 8$ GeV, 80% of the total cross section at the LHC is owing to gg fusion processes. Since it is symmetric, it does not produce any asymmetry but dilutes the one produced in $q\bar{q}$ annihilation.
- The probability that the anti-quark has larger momentum fraction than the quark (in which case a forward top has smaller $|y|$ and contributes negatively to the A_C) is not negligible and leads to a further dilution of the generated asymmetry.

Notice that at the LHC qg processes are not suppressed as they are at the Tevatron, but the asymmetry they generate is small.

Definitely A_{FB} and A_C are different observables. However, both asymmetries reflect the same underlying processes and are correlated. In Figure 3.4) the relation of A_{FB} and A_C is shown for a number of extensions of the SM that affect the asymmetry. Clearly, models with a large deviation at the Tevatron also lead to a large deviation at the LHC.

The SM predictions for the A_C at the LHC, including EW contributions, are $A_C = 0.0115 \pm 0.0006$ and $A_C = 0.0102 \pm 0.0005$ at 7 and 8 TeV respectively [40]. The ATLAS Collaboration measured $A_C^{ATLAS} = 0.006 \pm 0.010(stat) \pm 0.005(syst)$ [42] and the CMS collaboration $A_C^{CMS} = 0.004 \pm 0.010(stat) \pm 0.011(syst)$ [43] at $\sqrt{s} = 7\ TeV$, compatible with the SM prediction. Both measurements are combined giving as a result $A_C = 0.005 \pm 0.007(stat) \pm 0.006(syst)$, which represents an improvement of 40% (18%) with respect to CMS (ATLAS) measurements. All values of the A_C measured at $\sqrt{s} = 7\ TeV$ are summarized in Figure 3.5.

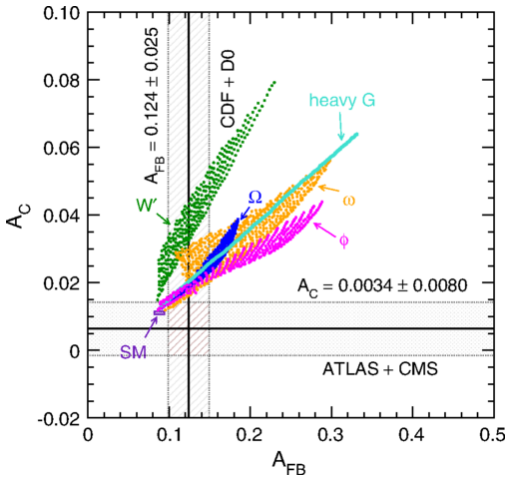


Figure 3.4: Comparison of predictions for the inclusive asymmetries A_{FB} and A_C for several simple models, together with the experimental measurements. Figure from [39].

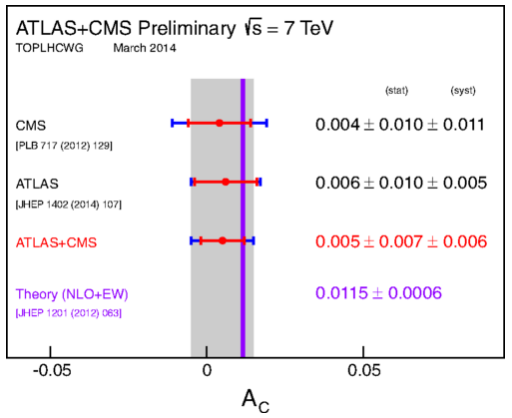


Figure 3.5: Summary of the measurements of the inclusive A_C by the ATLAS and CMS Collaborations in the $\ell + jets$ at 7 TeV, as well as their combination, compared to the theoretical prediction. For each measurement, the outer (inner) error bar indicates the total (statistical) uncertainty. The shaded band illustrates the total uncertainty of the combined result. Figure from [39].

3.4 The charge asymmetry in boosted $t\bar{t}$ pair production

We propose a $t\bar{t}$ charge asymmetry measurement at large invariant mass $m_{t\bar{t}}$ and using techniques developed for boosted topologies [44]. The measurement is motivated by the following arguments.

⋆ The LHC, with a much larger centre-of-mass energy than the Tevatron and its high integrated luminosity, allows to perform the charge asymmetry measurement at much larger invariant mass.

⋆ The sensitivity to BSM contributions increases. This can be seen in Figure 3.6, where the A_C at large invariant mass $m_{t\bar{t}}$ is shown for the major families of the SM extensions that act on the A_C (the same models as in Figure 3.4). The A_C reaches values of 40% for some models.

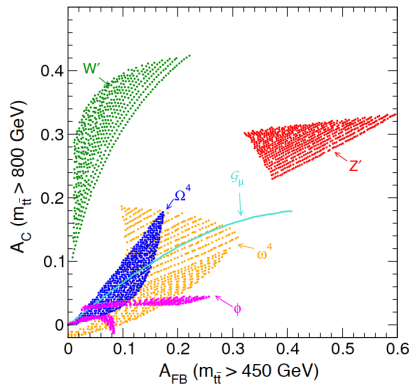


Figure 3.6: A_C prediction at the LHC for a sample with $m_{t\bar{t}} > 800\ GeV$ compared to the A_{FB} for a sample with $m_{t\bar{t}} > 450\ GeV$, for BSM models (Figure from [44]).

⋆ The dilution due to the symmetric gg-initiated process is less pronounced in the boosted regime. This is owing to the gluon PDF, which is much "softer" than that of (valence) quarks. A quantitative result is presented in Figure 3.7.

Moreover, the fraction of $q\bar{q}$-initiated events increases strongly with increasing $m_{t\bar{t}}$, therefore, the dilution due to these symmetric gg-initiated processes decreases. This is shown in Figure 3.8, which shows that the total cross section of gg-initiated processes (dash-dotted red line) is five times larger than that for $q\bar{q}$ initial states (continuous green line). At the $t\bar{t}$ production threshold, the dominance of the gg processes is particularly pronounced. Beyond 1 TeV, $q\bar{q}$ processes regain some terrain: the ratio $gg/q\bar{q}$ is reduced to 3.5.

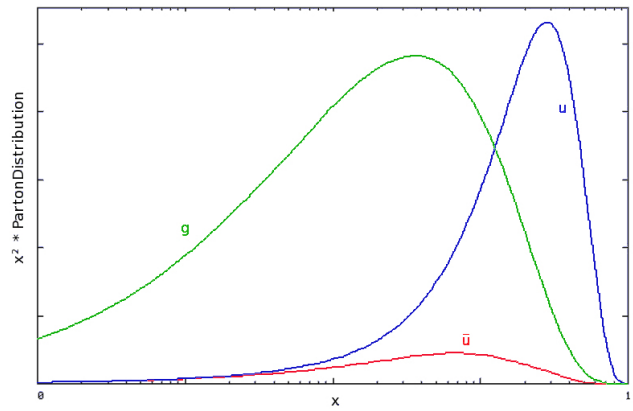


Figure 3.7: The curves "PartonDistribution vs x" corresponding to quark and anti-quark are so similar that are indistinguishable by eye. So "x^2·PartonDistribution vs x" curves have been performed to observe easily the difference in the slopes. Inside the proton, is more probably to find a quark with high fraction momentum (x) than an anti-quark. This can be observed in the figure, since the quark PDF always is above the anti-quark PDF.

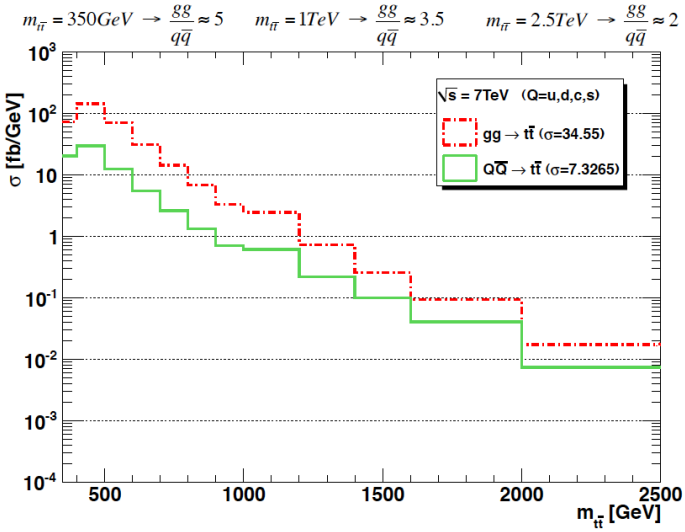


Figure 3.8: gg and $q\bar{q}$ cross sections at the LHC. gg processes dominate the $t\bar{t}$ production. Events in which the $t\bar{t}$ system is produced at high invariant mass suffer less dilution from gg, since this rate decreases.

⋆ Ambiguities in the reconstruction of the top quarks lead to a further dilution of the A_C measurement. For mildly boosted top quarks the ambiguity disappears to a large degree, and a much more robust measurement is possible.

Dilution due to the reconstruction is given by the D factor, defined as $D = 2p - 1$, where p tells you how many times $\Delta|y|^{reco}$ has different sign from $\Delta|y|^{truth}$. The true and reconstructed rapidity difference are related by the D factor as $\Delta|y|^{reco} = D \cdot \Delta|y|^{truth}$. The typical value for the D factor of the boosted reconstruction scheme is 70%, while it is 50% in resolved analyses.

In Figure 3.9, the dilution factor for the boosted analysis improves as $m_{t\bar{t}}$ extends above 1 TeV, while the resolved analysis begins to degrade at just above 0.6 TeV. For small $m_{t\bar{t}}$ values, the boosted analysis consists of very few events, but those that are reconstructed have a much better efficiency for reconstructing the correct $\Delta|y|$ sign. This high efficiency is attributed to the fact that in the boosted topology, there is no ambiguity in the assignment of jets and the top direction is measured very precisely. This is explained in more detail in Section 6.4.2.

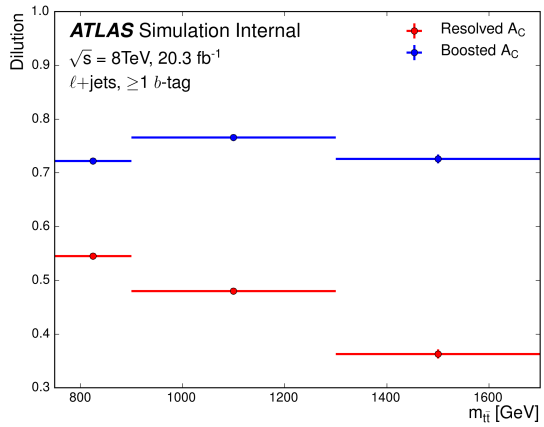


Figure 3.9: The dilution factors for the boosted and resolved analyses as a function of $m_{t\bar{t}}$ for the combined ℓ+jets channel.

⋆ Technique designed specifically for boosted analysis has better acceptance at high $m_{t\bar{t}}$, as it has been explained in the previous paragraph. Furthermore, the boosted algorithm, which is specifically designed for this regime, shows significantly better performance in both lepton channels. This can be seen in Figure 3.10, where the $m_{t\bar{t}}$ resolution, defined as $\left(m_{t\bar{t}}^{reco} - m_{t\bar{t}}^{true}\right)/m_{t\bar{t}}^{true}$, in both algorithms is compared for true $t\bar{t}$ masses between 0.75 and 2 TeV.

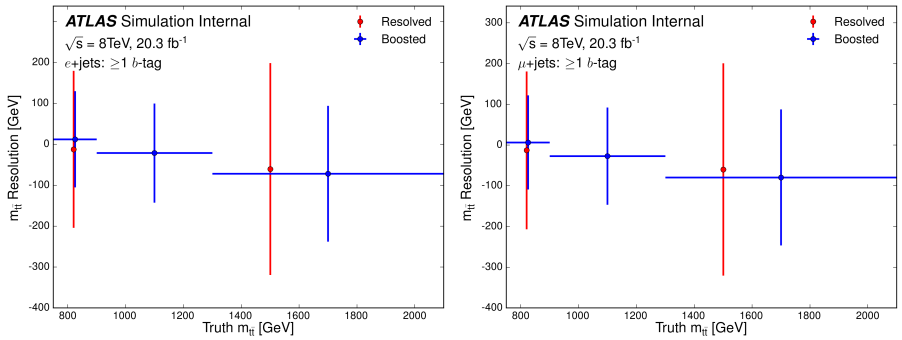


Figure 3.10: The bias and resolution in $m_{t\bar{t}}$ as a function of $m_{t\bar{t}}$ in the boosted and resolved selections. The data points represent the bias, while the error bar indicates the resolution. The e+jets channel is shown in the leftmost panel and the μ+jets channel in the rightmost panel. An additional bin for $m_{t\bar{t}}$>1300 TeV is included for the boosted reconstruction.

All of these arguments motivate the measurement of the boosted A_C at the LHC as high $m_{t\bar{t}}$ range as possible. In Chapter 8 the first result from this measurement, made by the ATLAS Collaboration, is presented [45].

4

The Large Hadron Collider and the ATLAS experiment

4.1 The LHC

The Large Hadron Collider (LHC) [46] is the largest collider ever built. It can accelerate protons or heavy ions. It is located at the *Conseil Européen pour la Reserche Nucléaire* (CERN) [47]. It is assembled in the same 27 km tunnel where Large Electron-Positron Collider (LEP) [48] used to be. This tunnel is buried ~ 100 m underground, near Geneva (Switzerland). The four main interaction points, where experiments are installed, and the LHC tunnel are shown in Figure 4.1.

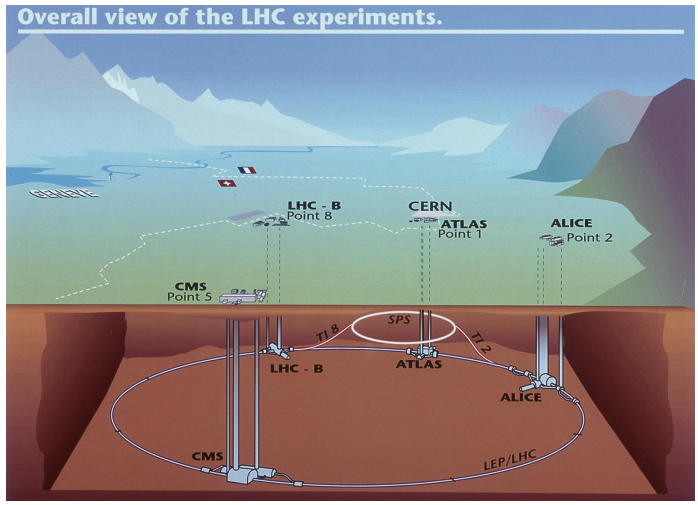


Figure 4.1: Schematic view the LHC and its experiments.

The LHC is **designed** to collide two proton beams, that circulate in opposite directions, each with an energy of up to 7 TeV, reaching a centre-of-mass energy up to 14 TeV. In order to bend such energetic beams and keep them onto their circular trajectory, 1232 superconducting dipole magnets generate a magnetic field of 8.3 T, and 600 superconducting quadrupoles magnets produce a 6.8 T field. The coils are made of niobium-titanium (NbTi) and kept cooled, around 1.9 K, with superfluid liquid helium. The beams are structured in 2808 bunches separated by 25 ns. Both beams cross each other in 4 points of the tunnel, where the detectors are installed. Each crossing can produce over 20 collisions reaching a nominal luminosity [1] of $10^{34} cm^{-2} s^{-1}$.

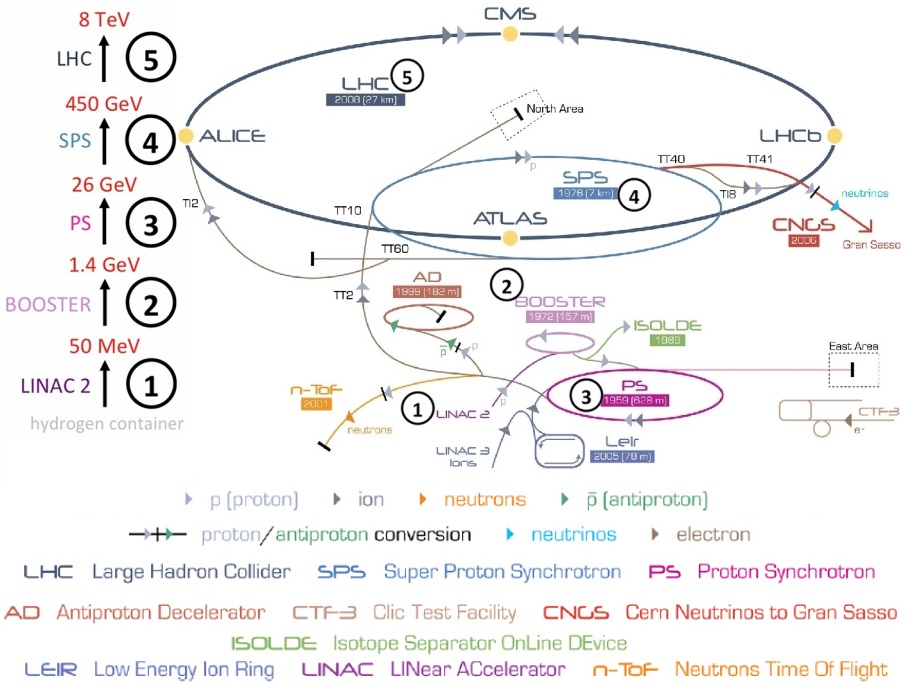
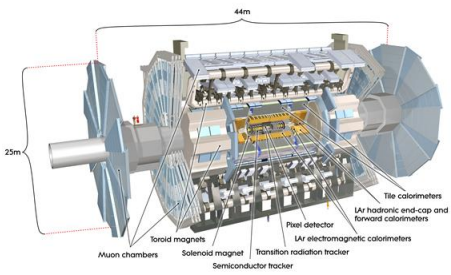


Figure 4.2: Vignette layout of the accelerator complex at CERN, including all elements of the LHC injector chain. In the left side is shown the acceleration process of every beam. The energies quoted here are related to the First LHC Running Period (Run-I).

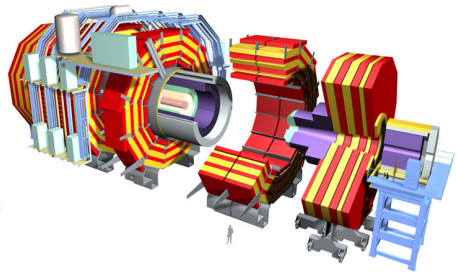
[1] Luminosity is a measure of the number of collisions per cross section and time. More details in Appendix A.1, Equation A.1.

Protons travel through the CERN accelerator complex, starting in a hydrogen container and ending in one of the experiments. The entire voyage is depicted in Figure 4.2. Before entering into the tunnel, protons are extracted from the container by ionization and they are pre-accelerated to 50 MeV in the Linac2 [49]. Then, they enter the Proton Synchrotron Booster (PSB) [50] where they reach an energy of 1.4 GeV each one. From here, the beams are led to the Proton Synchrotron (PS) [51] where their energy is increased up to 25 GeV. Afterwards, they go through the Super Proton Synchrotron (SPS) [52] and reach 450 GeV. Finally, they are injected in the LHC tunnel and accelerated to a final energy of 4 TeV each beam.

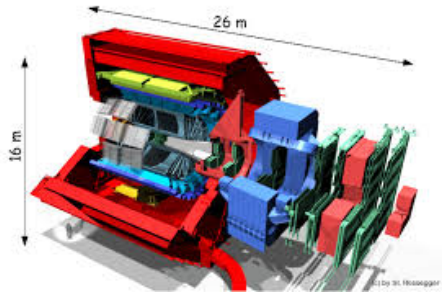
The four interaction points [53] host the experiments shown in Figure 4.3:



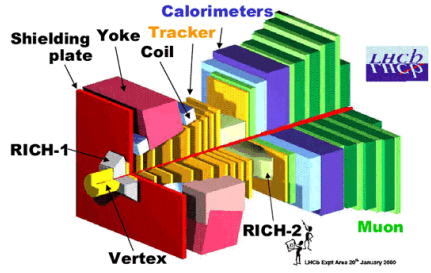
(a) ATLAS



(c) CMS



(b) ALICE



(d) LHCb

Figure 4.3: LHC experiments.

- **ATLAS** (A Toroidal Lhc Apparatus): a multi-purpose experiment built to work mainly at high luminosity and designed to have a high performance for SM precision measurements in addition to Higgs and BSM searches. It is the largest detector hosted at CERN.
- **CMS** (Compact Muon Spectrometer): the other general-purpose experiment, like ATLAS, but both use different and complementary technologies.
- **ALICE** (A Large Ion Collider Experiment): focused on the study of the quark-gluon plasma [2] produced in heavy nucleus collisions (Pb-Pb).
- **LHCb** (LHC Beauty): dedicated to b-quark physics, especially to CP violation measurements in B meson decays.

This chapter is focused on the ATLAS experiment since the analyses explained later are based on data recollected by this detector.

4.2 The ATLAS experiment

The ATLAS experiment [54] is a general-purpose detector designed to exploit the potential of the LHC. Its overall cylindrical design, shown in Figure 4.4, is described in the Technical Design Report [55, 56] in detail.

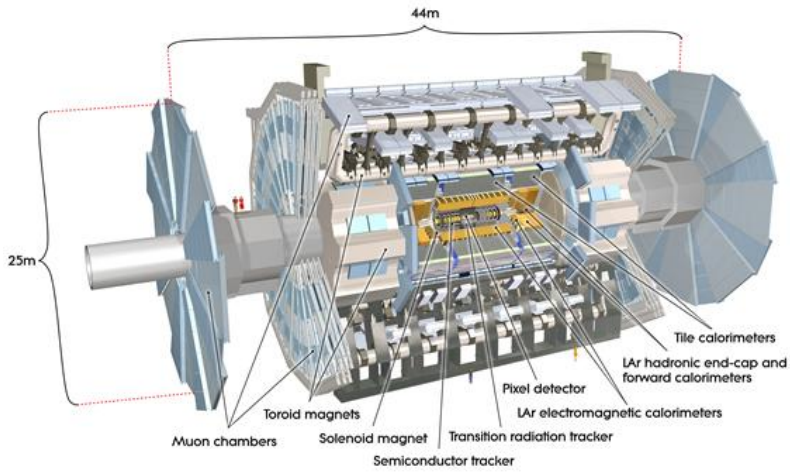


Figure 4.4: The ATLAS experiment.

[2] The quark-gluon plasma is a state of matter composed by asymptotically free quarks and gluons.

With a total weight of 7000 tonnes, ATLAS is the heaviest experiment at the LHC. It has 42 m of total length and 11 m of radius. It is hosted under the ground of the Interaction Point 1 of the LHC. Its components can be grouped in four main systems: the Inner Detector (ID), the Calorimeters, the Magnet system and the Muon Spectrometer. Below these are explained in more detail.

4.2.1 The Inner Detector

The layout of the Inner Detector (ID) is shown in Figure 4.5. It is the closest system to the beam pipe and the interaction point. It is in charge of measuring the momentum of the charged particles. The ID is immersed in a solenoidal magnetic field of 2 T. The coverage extends over the full azimuthal angle and it has acceptance, for physics analyses, up to $|\eta| = 2.5$ [3]. It combines high-resolution detectors with fine-granularity layers in the inner part, and continuous tracking elements in the outer part. Thus, it provides good accuracy in measurements of charged particle tracks at high particle density.

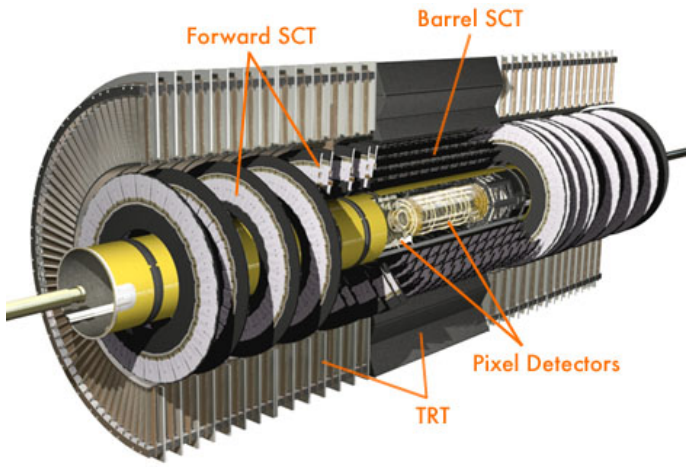


Figure 4.5: The ATLAS Inner Detector.

It is composed of three subsystems: Pixel detector, Semiconductor Tracker and Transition Radiation Tracker, sorted from inner to outer.

[3] η denotes the pseudorapidity. These variables are explained in Section 4.3. $|\eta| = 2.5$ is equivalent to a polar angle $\theta = 10°$.

The **Pixel Detector (PD)** is made of semiconductor pixel detectors, that help to achieve the highest granularity around the interaction point. This sub-detector is based on silicon technologies and occupies the region $5 - 15\ cm$ of radius from the interaction point. It consists of 3 cylindrical barrel layers and 3 disks on each side of the central barrel, covering the forward and backward regions. This design allows for an accurate measurement of decay vertex positions and particle impact parameters.

The **Semiconductor Tracker (SCT)** consists of 4 concentric double layers of silicon microstrip detectors (with radius $30-52\ cm$ from the beam axis) in the barrel and 9 disks in the end-caps (that cover up to $|\eta| < 2.5$). With this layout, the SCT provides track precision measurements and allows the measurement of momentum, vertex position and impact parameter in the intermediate radial range.

The **Transition Radiation Tracker (TRT)** combines $\sim$ 300,000 gaseous radial straws arranged in a barrel and 2 end-caps at the ends of this barrel. Electron identification is ensured by employing xenon gas to detect transition-radiation (it is produced when a relativistic particle crosses an inhomogeneous medium) photons created in a radiator between the straws.

Tracking is performed by combining hits from the three sub-detectors using two different algorithms:

- Inside-out algorithm: it reconstructs most of the primary tracks.
- Outside-in algorithm: it starts from the TRT and goes inwards by adding silicon hits. It reconstructs the vast majority of the secondary tracks coming from conversions, hadronic interactions or V_0 decays.

Primary vertices are reconstructed using an iterative χ^2 fit [57].

4.2.2 The Calorimeters

The ATLAS Calorimeter system can be seen in Figure 4.6. This system measures the direction and the energy of all particles produced in collisions in order to reconstruct photons, τ-leptons, electrons and jets. The energy measurement is based on the formation of a particle-cascade in the material of the calorimeter. These particle showers are produced by pion decays and electromagnetic bremsstrahlung. The ATLAS calorimeter consists of an electromagnetic calorimeter covering the region $|\eta| < 3.2$ and a hadronic calorimeter covering $|\eta| < 4.9$.

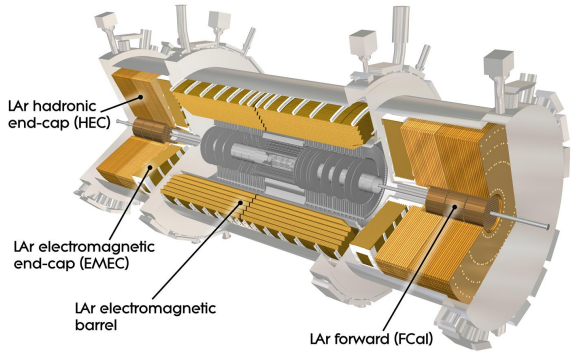


Figure 4.6: The ATLAS Calorimeters.

The **Electromagnetic Calorimeter (ECAL)** is a lead and liquid argon (LAr) sampling calorimeter. The lead ensures the shower development due to its short radiation length and the liquid argon is used as ionization medium to measure the electron ionization. The drift of ionized electrons in the electric field across the gaps produces an inductive signal, which is registered by copper electrodes. The ECAL has an energy resolution of:

$$\frac{\sigma_E}{E} = \frac{11.5\%}{\sqrt{E}} \oplus 0.5\% \text{ (E in GeV)} \tag{4.1}$$

The **Hadronic Calorimeter (HCAL)** is composed by a barrel sampling calorimeter (called TileCal) and end-caps. The barrel uses scintillator plastic plates (tiles) encrusted in iron absorbers. The end-caps use LAr as active material and copper as absorber, and they can measure hadronic and electromagnetic showers. The HCAL has an energy resolution of:

$$\frac{\sigma_E}{E} = \frac{50\%}{\sqrt{E}} \oplus 3\% \text{ (E in GeV)} \tag{4.2}$$

The ATLAS Calorimeter system covers a large geometrical acceptance hermetically. Then, the presence of uncharged particles that do not interact or interact weakly, such as neutrinos, can be deduced or inferred from the measurement of the unbalanced transverse energy, called transverse missing energy (E_T^{miss}).

4.2.3 The Muon Spectrometer

A view of the Muon Spectrometer is presented in Figure 4.7. This spectrometer is the outermost ATLAS sub-detector, which triggers events with high momentum muons ($p_T > 6\ GeV$) and measures accurately the momentum in the pseudo-rapidity region $1 \leq |\eta| \leq 2.7$. This system is very important in the identification of muons since they do not feel the strong interaction and traverse the whole inner ATLAS sub-detectors with hardly any interaction.

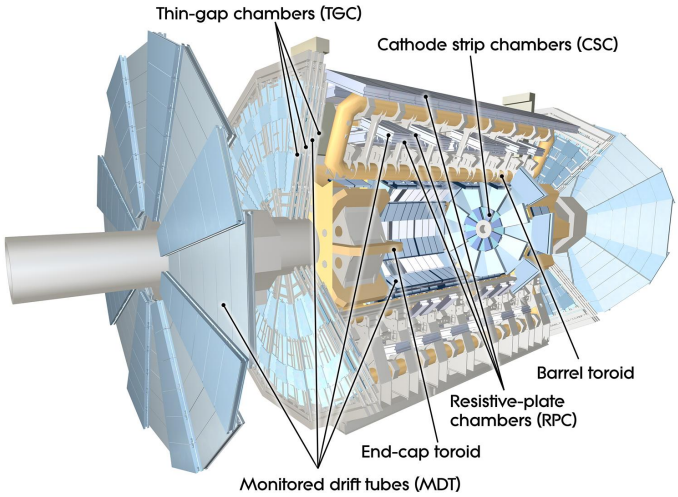


Figure 4.7: The ATLAS Muon Spectrometer and its components.

The muon spectrometer consists of four technologies, grouped in two kind of chambers:

- The precision chambers: the **Monitored Drift Tubes (MDT)** and the **Cathode Strip Chambers (CSC)**. They provide high-precision tracking information by measuring the momentum of the muons.
- The trigger chambers: the **Resistive Plate Chambers (RPC)** and the **Thin Gap Chambers (TGC)**. They are used for triggering with $1.5 - 4\ ns$ of time resolution.

4.2.4 The Magnets

The conceptual layout of the ATLAS Superconducting Magnet System is shown in Figure 4.8. This system is characterized by two different magnetic field subsystems that bend the track of the particles.

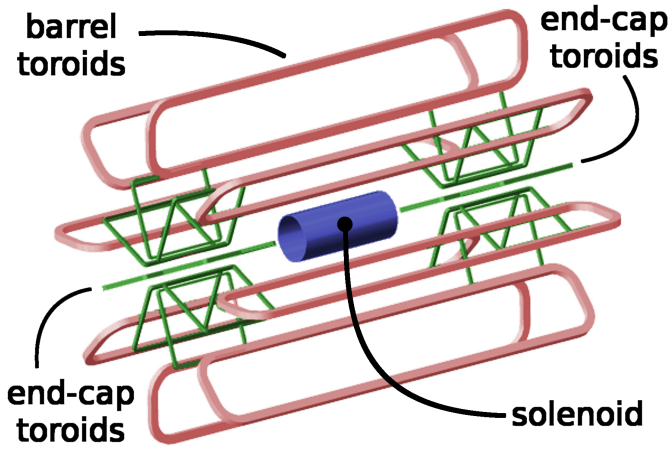


Figure 4.8: The ATLAS Superconducting Magnet System.

The **Central Solenoid (CS)** provides a magnetic field of 2 T. It surrounds the Inner Detector cavity. Its dimensions are 5.3 m of length and 1.2 m of radius. It is optimized so that it minimizes the amount of material in front of the electromagnetic calorimeter.

The super-conducting air-core toroid system has an open structure to minimize the contribution of multiple scattering to the momentum resolution and provides a magnetic field of 4 T. It consists of eight **Barrel Toroids (BT)**, assembled radially and symmetrically around the beam, and two **End-Cap Toroids (ECT)**. Every barrel toroid is 25 m long and fills the range $9.4 - 20.1$ m of radius. Every end-cap toroid has 5 m of length and fills the range $1.65 - 10.7$ m of radius. The toroid provides bending powers of 3 $T \cdot m$ in the BT and 6 $T \cdot m$ in the ECT. The latter provides radial overlap and optimizes the bending power in the transition region $1 < \eta < 1.4$.

4.2.5 The Trigger System

The task of the **Trigger System** is to store information recorded by the ATLAS sub-detectors and to reduce the number of minimum bias events to be stored permanently. Every data stream (jets, muons, electrons, etc) has a different trigger chain based on specific algorithms. The trigger system is organized in three levels called **first level (L1)**, **second level (L2)** and **Event Filter (EF)**, containing increasingly complex algorithms. The schematic view of the trigger decisions is depicted in Figure 4.9.

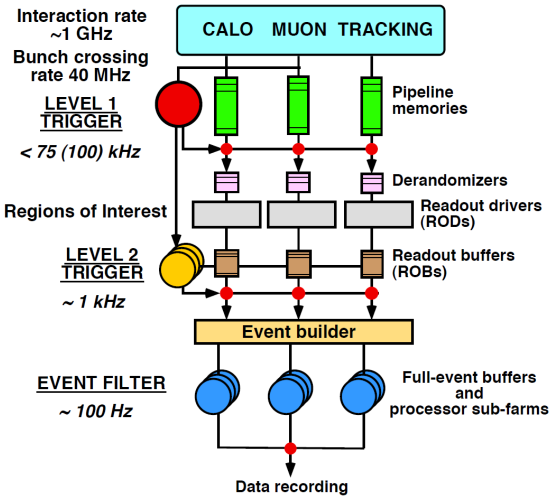


Figure 4.9: ATLAS Trigger Levels.

The **L1** trigger algorithm acts while most data are still stored locally. It discards the major part of non interesting events, with a latency of $\sim 3\ \mu s$. It makes use of a subset of the calorimeter and muon detectors. After L1 the data acquisition rate is reduced to $\sim 100\ kHz$. The **L2** uses fine-grained data from all sub-detectors and is based on reduced Regions of Interest (RoI) defined by L1. Events are processed on farms of PCs. A seed, which consists of a p_T threshold and a $\eta - \phi$ position, is constructed for every event accepted by L1. Then, the L2 algorithms use this seed to construct a RoI window around the seed position. After L2 the data acquisition rate is reduced to $\sim 1 - 2\ kHz$. The **EF** is the last step before recording data permanently. It refines the events that have passed L1 and L2 selections. After EF the data acquisition rate is reduced to few hundred hertz.

4.3 Kinematic variables and Nomenclature

The ATLAS experiment uses a special kind of cylindrical coordinates. It is important to describe this coordinate system to understand the variables used to perform the physics analyses developed in this thesis.

The adron collider coordinate system is shown in Figure 4.10. The beam direction defines the z-axis. The $x-y$ plane (Π_{XY}) is the plane transverse to the beam direction. The positive y-axis is pointing upwards. The positive x-axis is defined as pointing from the interaction point to the centre of the LHC ring. The polar angle θ is the angle from the beam axis. The azimuthal angle ϕ is measured around the beam axis. The transverse momentum p_T and the transverse energy E_T, as well as the missing transverse energy $E_T^{miss}$and other transverse variables, are defined in the $x-y$ plane. The longitudinal momentum p_L or p_z is the component of the momentum along the beam axis.

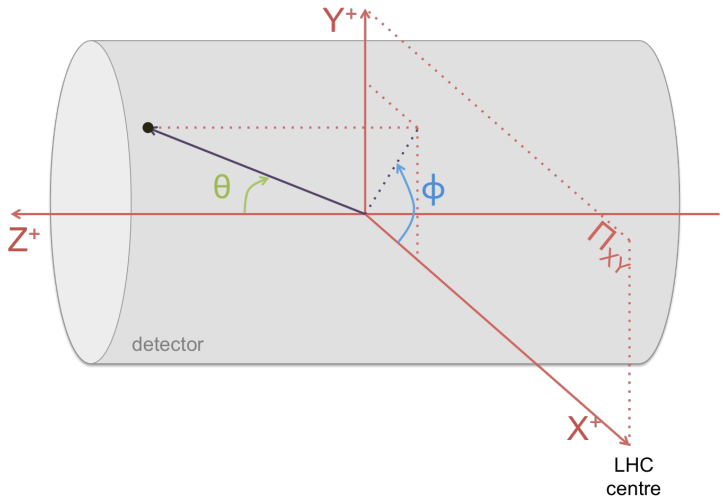


Figure 4.10: Sketch of a hadron collider and its coordinate system.

4.3.1 Rapidity and Pseudorapidity

In hadron collider coordinates, instead of using θ, rapidity y and pseudorapidity η are used. The rapidity y can be calculated from the measured energy and momentum by:

$$y = \frac{1}{2}\ln\left[\frac{E + p_L}{E - p_L}\right] = \tanh^{-1}\left(\frac{P_z}{E}\right) \tag{4.3}$$

and the pseudorapidity η is defined as:

$$\eta = \frac{1}{2}\ln\left[\frac{|\vec{p}| + p_L}{|\vec{p}| - p_L}\right] = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right] \tag{4.4}$$

In the relativistic limit of $p \gg m \longrightarrow \eta = y$.

The difference in rapidity, Δy, is invariant under a Lorentz boost along the beam line (z-axis) to a frame with velocity β:

$$y \longrightarrow y - \tanh^{-1}(\beta) \tag{4.5}$$

Distances between particles can be measured from the distance ΔR, which is the angular separation between two particles in the $\eta - \phi$ space:

$$\Delta R = \sqrt{\Delta^2\eta + \Delta^2\phi} \tag{4.6}$$

The relation between θ and η is shown in Figure 4.11.

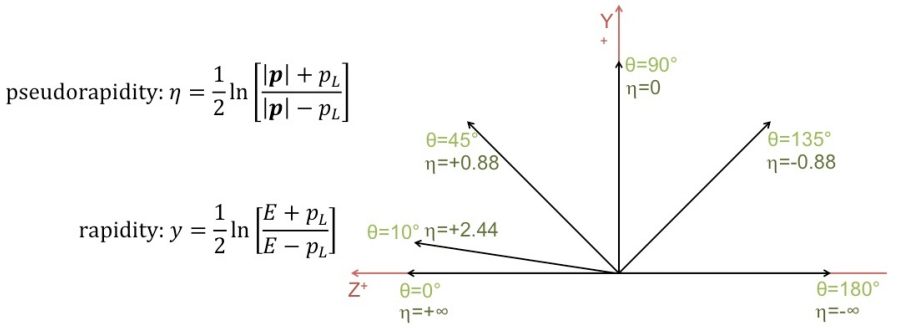


Figure 4.11: Relation between pseudorapidity η and polar angle θ. As θ approaches zero, η tends towards infinity.

This variable y is very important in one of the analyses explained in this thesis, in the measurement of the $t\bar{t}$ charge asymmetry (explained in Chapter 8), since this observable is directly related to the top and anti-top quark rapidities.

In the context of hadron collider experiments, it is common to talk about "forward" and "backward" regions of the detector. The former refers to the positive z-direction and the latter to the negative z-direction. Both regions are related to large y or η absolute values. However, the central region is that close to the centre of the detector and is related to low y or η values.

4.3.2 Charged particle trajectories

Trajectories of charged particles can be described by five helix parameters in an ideal uniform magnetic field. The following helix parameterization is used in ATLAS, with all quantities measured at the point of closest approach to the nominal beam axis $x = 0$ and $y = 0$:

- Parameters in $x - y$ plane:
  - $\mathbf{1/p_T}$ is the reciprocal of the transverse momentum with respect to the beam-axis.
  - $\boldsymbol{\phi}$ is the azimuthal angle, which accomplishes that $\tan\phi = p_y/p_x$.
  - $\mathbf{d_0}$ is the transverse impact parameter, defined as the transverse distance to the beam axis at the point of closest approach.
- Parameters in $R - z$ plane:
  - $\mathbf{\cot\theta}$ is the cotangent of the polar angle, which satisfies $\cot\theta = p_z/p_T$.
  - $\mathbf{z_0}$ is the longitudinal impact parameter, defined as the z position of the track at the point of closest approach.

5

ATLAS Spanish Tier-2 within the ATLAS Computing Model for Run-I

Computing plays an important role in High Energy Physics (HEP). The LHC experiments have to handle data streams of PB/hour. The exploitation of LHC data is impossible without sophisticated computing tools. Since the CERN gave birth to the World Wide Web (www), it has been working very hard to improve the existing technologies and develop new ones. Fifteen years ago, the CERN experiments started using Grid technologies to deal with large amounts of data, allowing them to manage and to analyze those data distributively.

The Worldwide LHC Computing Grid Project (WLCG) [58, 59] is a collaboration composed of more than 170 computing centres distributed in 42 countries, connecting national and international grid infrastructures. The main goal of the WLCG is to build and maintain a distributed computing infrastructure to store, distribute and analyze the data from the LHC experiments. The WLCG is supported by many grid projects, such as The LHC Computing Grid (LCG) [60], The European Grid Initiative (EGI) [61, 62], The NorduGrid - Advanced Resource Connector (ARC) [63, 64] and Open Science Grid (OSG) [65], among others.

Run-I [1] of the LHC is emphasized in this thesis because the analyses of Chapters 7 and 8 are based on Run-I data. During this LHC data-taking period, IFIC was involved in the Spanish Federated Tier-2 (ES-ATLAS-T2) and consequently in the Iberian ATLAS Cloud ([66, 67, 68]), providing computing resources to the ATLAS Collaboration.

[1] Look at Appendix A.1 for more details and definition of periods.

5.1 The ATLAS Computing Model

The ATLAS Computing Model (ACM) [69, 70] defines the requirements on the architecture of the software and its performance. The ACM allows a high degree of decentralization and the possibility to share computing resources owing to its tiered hierarchy based on Grid technologies. It handles everything from the storage of data at CERN (raw events) to physic analyses at home institutes (reprocessed data).

The tiered hierarchy is structured in levels, called Tiers, and is managed centrally by the ATLAS Collaboration. This structure is shown in Figure 5.1.

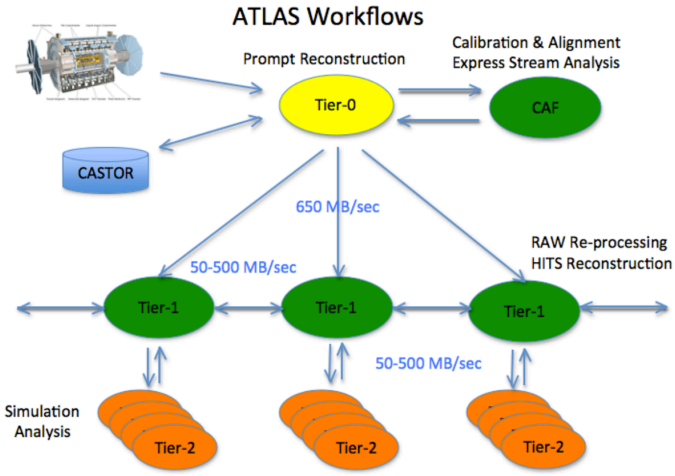


Figure 5.1: The ATLAS Tier structure for the Run-I.

The main roles of every level are listed below:

- First level: **Tier-0**. There is only one and it is hosted at CERN, where raw events are stored. A first reconstruction of event is performed and a copy of it is saved in this centre. Also, a second copy of raw data is distributed in associated Tier-1s.
- Second level: **Tier-1**. There are 11. They manage the permanently stored data (raw, simulated and processed data) and provide computational capacity for the reprocessing and analysis of data.
- Third level: **Tier-2**. There are about 80 centres distributed worldwide. They provide computational and storage capacity suitable for the Monte Carlo (MC) event simulation and the end-user analysis.

- Fourth level: **Tier-3**. There are tens of them. This level is an end-user private analysis facility. Each one is associated with a Tier-2 although they are outside the LCG project.

The set of a Tier-1 and several associated Tier-2s distributed in a given periphery is called **Cloud**.

A problematic consequence of a strictly tiered hierarchy is that when a Tier-1 is in scheduled downtime, the job input files and data transfer of its associated Tier-2 are affected. Therefore, it was proposed to evolve the ACM. In the new structure, Tier-2s with enough network connection can link directly with other Tier-1s or Tier-2s belonging to different clouds. This kind of Tier-2 are tagged as "Tier-2 Directly" (**T2D** [71]) and they have to fulfil two additional requirements:

- The overall transfer rate of big files to/from Tier-1 sites must be above 5 MB/s during the last week and 3 MB/s out of the last 5 weeks.
- The availability of the site must be higher than 90%.

The structural evolution of the ACM, including the T2Ds, is shown in Figure 5.2.

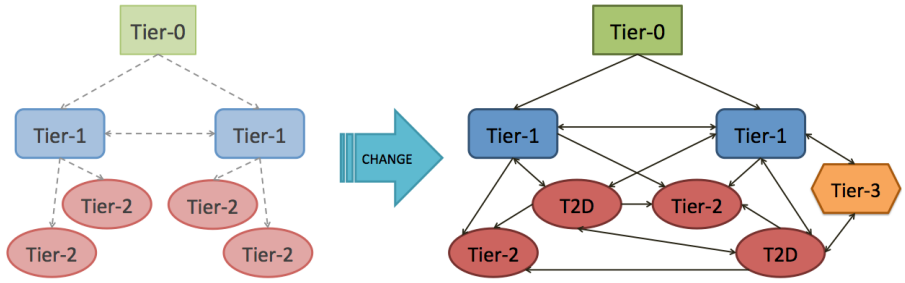


Figure 5.2: The evolution of the ATLAS Computing Model for the Run-I, changing from a tiered structure to a mess model.

The new configuration of the ACM provides a more efficient usage of disk and CPU resources and allows high priority tasks to be done more quickly.

5.2 The ATLAS Event Data Model

The ATLAS Event Data Model (AEDM) [72] describes the evolution of the data format. Raw data undergo several transformations since they are formed in the detector until they are analyzed by physicists. Below are listed the generation sequence of different data types and their main features:

- **Raw Data Object (RDO)**: are the date coming directly from the detector (raw data), which are stored immediately and massively in the Tier-0. These data are generated in *bytestream* format, requiring about 1.6 MB per event.
- **Event Summary Data (ESD)**: contains the information coming from the reconstruction algorithms and data related to the calorimeters and tracks. These data are encoded in POOL/ROOT [2] format. The nominal size is roughly 1 MB per event.
- **Analysis Object Data (AOD)**: is generated from the ESD, leading to the first reconstructed objects (electrons, muons, jets, etc) in POOL/ROOT format. They contain the so-called MC Truth Information related to the reconstructed objects. The desired size is 100 kB per event.
- **Derived Physics Data (DPD)**: is the last file type in the analysis chain. It can be analyzed later inside and outside the ATLAS analysis software and contains very specific information, depending on the needs of each group.
- **TAGs and ntuples**: are databases or ROOT [73, 74] files aimed to make the search for objects easier and to perform physics analyses.

This dataflow, applied to real and MC data, is schematized in Figure 5.3

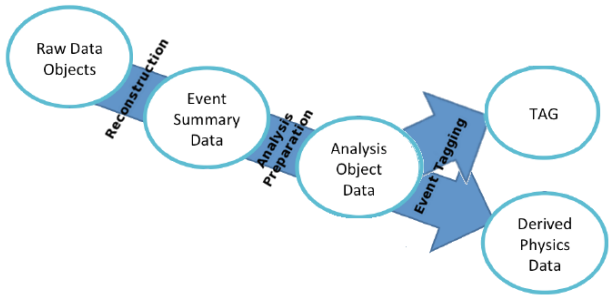


Figure 5.3: Dataflow from the detector to the end-user for the Run-I.

[2] ROOT is a scientific software framework. It provides all the functionalities needed to deal with big data processing, statistical analysis, visualization and storage. It is mainly written in C++ but integrated with other languages such as Python.

5.3 The ATLAS Monte Carlo Production

The simulation data flow [69] shows the processing stages in the production of MC data, which have features and format similar to the real ones. MC data are events simulated in the detector used to foresee the behaviour of real data and to perform realistic approaches by comparing them with the well-known physics backgrounds. The production of these simulated events is an elaborated process, which requires large computing power. Thus, it is managed by the ATLAS Grid infrastructure. To obtain these MC data, a full simulation of the detector is needed, which is called Full Chain and it has the following steps:

- **Generation**: it produces the 4-vectors associated to the particles. In this step Event Generators are used as tools for modelling the complex physics processes.
- **Simulation**: the generated events are passed through a simulation in order to simulate the detector response. The output are hits (energy deposits in the detector). This simulation is performed with the GEANT4 [75, 76] toolkit.
- **Digitization**: in this step hits are translated into digits or Raw Data Objects (RDO), which are mainly times and voltages, as the real detector data. Digitization is tuned by comparing the simulated RDO to real data.
- **Reconstruction**: digits become tracks, clusters and jets in ESD format and several parameters required for the physics analysis are derived. This step is identical for both, MC and real data.
- **AOD production**: the data needed for a given analysis and their further study are chosen.

The Full Chain used during Run-I is shown in Figure 5.4. One can reduce the Full Chain by using Atlfast [77, 78, 79], which provides a fast simulation of the whole chain by taking the generated events and smearing them to produce AOD directly. Atlfast can run on the output of any of these stages.

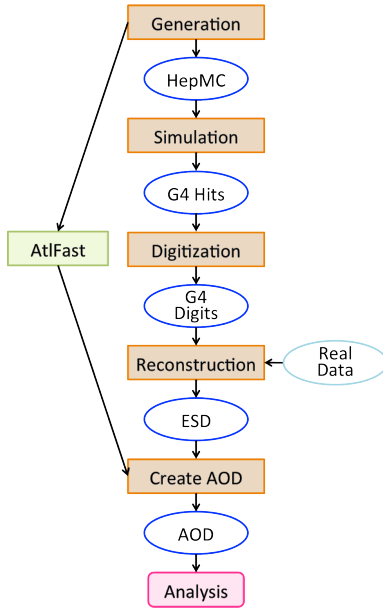


Figure 5.4: Schematic representation of the Simulation Data Flow (Full Chain Monte Carlo production) for the Run-I. Rectangles represent processing stages and ovals represent objects within the ATLAS Event Data Model.

5.4 Distributed Analysis Tools

The end-users are the physicists working daily on physics analyses. Typically, a physics analysis has two parts. In the first stage, physicists run an analysis program that uses a given number of collision events. These events can be stored in different datasets that are usually spread over the different sites. At this step, the Distributed Computing and Data Management Tools, based on Grid Technologies, are used in an exhaustive way. The output of this first step is often a set of ROOT ntuples. In the second stage, the physicists analyze the ntuples interactively in order to get the final plots, to refine the analysis, etc.

The ATLAS Computing Team has developed specific user-friendly tools for end-users to perform the distributed analysis tasks using the Grid infrastructure, and to make their work easier. These tools are aimed to get data information from the experiment, to submit the analysis jobs to the Grid infrastructure and to retrieve the final output. The Distributed Analysis tools used during Run-I were the following:

1. For Data Management:
   - **AMI** (ATLAS Metadata Interface) [80]: it is a web page which has a catalogue of the official metadata, useful for monitoring datasets, releases, number of events, etc.
   - **DaTri** (Data Transfer Request Interface): it is the end-user dataset subscription service. It allows end-users to request a copy of their output files in their local site or institute.
   - **dq2** (Don Quijote 2) [81]: it is applied to acquire information about data, such as dataset name, number of files and sites where dataset is stored. Also it is used to download and register files on Grid. This is the tool most used by the end-user to get their final output result files.

2. For Grid Jobs:
   - **ganga** (Gaudi/Athena [3] and Grid Alliance) [83]: it is a job definition and submission management tool for local, batch system and the Grid. It allows simple switching between testing on a local batch system and large-scale data processing on the Grid. It is implemented in Python [84].
   - **PanDA Client** (Production and Distributed Analysis) [85, 86]: it is an analysis job submission tool. It is the most popular tool used by the end-users because of its easy-to-use handling.
     - **pathena** (Panda Athena): it works in the Athena runtime environment. It is a client tool to submit user-defined jobs to Distributed Analysis systems. It provides a consistent user-interface to Athena users.
     - **prun** (Panda Run): it is a Panda-client software which allows users to submit general jobs to Panda. It is intended to support non-Athena type analysis.
     - **pbook**: it is the next-generation of the bookkeeping application for all Panda analysis jobs.

[3] Athena is a control framework, a concrete implementation of an underlying architecture called Gaudi. Gaudi is a kernel of software common to the ATLAS and LHCb experiments. Athena is the sum of this kernel plus ATLAS-specific enhancements. It is written in C++ and designed with modular component architecture, consisting of core packages such as dernel, services and several tools. External libraries supplement it. More details in [82].

The daily user activity in Distributed Analysis is illustrated with the example below. Simulated and real data input files represent a volume of several TB of information. The analysis activity workflow can be divided as follows:

- Phase 1: Test the analysis locally. Users have to download the input files (that can be either real or simulated data) with DQ2 and then, test their analysis algorithm locally in their Tier-3 infrastructure.
- Phase 2: Submit a first job to the Grid (to the Tier-2 centres) with Ganga and/or PanDA client tool. This job will create an output file with reduced information. The run time for a typical job is, on average, 20 hours.
- Phase 3: Submit a second job to the Grid (normally at the Tier-3) with the objective of doing a refined analysis (reconstruction, application of cuts and selections, etc.). In this case the input is the output of the first job (Phase 2) and the execution time is around 2 hours.

A description of the Phase 2, working within the Ganga Framework, is the following. A python script is created with the purpose of giving the job requirements and its features (such as application address, input and output dataset names, a replica request to our working site, the output splitting requirements, etc.). This script allows to send the job to the Grid. When the job finishes successfully, the output files are transferred to our working site, almost always a Tier-3, where the access is easy for the end-users. This workflow is depicted in Figure 5.5.

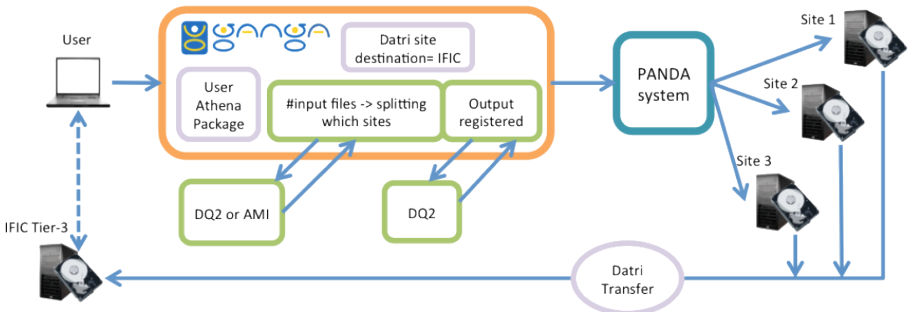


Figure 5.5: Example of an ATLAS workflow for Distributed Analysis for the Run-I.

5.5 Resources and Performance of the Iberian ATLAS Cloud

The Cloud in which the IFIC is involved is the so-called **Iberian ATLAS Cloud**. It is formed by:

- One Tier-1: PIC.
- One Portuguese federated Tier-2s: PT-LIP-LCG-Tier2, which consists of 50% COIMBRA and 50% INGRID.
- One Spanish federated Tier-2s: ES-ATLAS-T2, composed by 50% IFIC, 25% IFAE and 25% UAM.

The Iberian ATLAS Cloud also has one Tier-2 and one Tier-3 associated outside the Europe zone, but they are not included in this section: EELA-UTFSM (Tier-2, Chile) and EELA-UNLP (Tier-3, Argentina).

The performance of the Iberian ATLAS Cloud during the Run-I, in the context of the Grid Computing and Data Distribution Model, is described in the next subsections.

5.5.1 Evolution of the resources

Every year, the resources assigned to the ATLAS Collaboration are fixed in the Memorandum of Understanding (MoU) [87, 88] by the WLCG project. During the Run-I, the Iberian ATLAS Cloud provided the hardware resources fulfilling the ATLAS requirements of the Resource Review Board of the LHCC committee, as Table 5.1 shows. Disk is accounted in Terabytes (TB) and CPU in HEP-SPEC06 [89], which is a HEP-wide benchmark used for measuring CPU performance.

Federation	CPU (HEP-SPEC06)		DISK (TB)	
	Pledges 2013	Current	Pledges 2013	Current
ES-ATLAS-T2	18000	17800	2800	2558.3
PT-LIP-LCG-Tier2	3200	3200	220	183.0
ES-PIC	16269	16269	1785	1812.0

Table 5.1: Hardware resources provided by the Iberian ATLAS Cloud on September 2013. In this table, "current" means September 2013.

This table shows that the pledges in 2013 were carried out roughly. These resources are employed for processing the physics analysis jobs and for producing official MC samples. According to the ATLAS Collaboration, every Tier-2 has to split its resources reserving 50% for each task.

Grid Storage Element (SE) and disk spaces are managed by two distributed storage systems, called **dCache** [90] at PIC, IFAE and UAM, and **Lustre+StoRM** [91, 92] at IFIC, LIP_COIMBRA, LIP_LISBON (at the moment, it no longer exists) and NCG_INGRID_PT. The Worker Nodes have up to 2 *GB* of RAM per CPU core to be able to run the highly demanding ATLAS production jobs.

In addition to the pure Tier-2 resources, each site provides a Tier-3 infrastructure intended for data analysis, which has a part based on Grid architecture and another part being a standard computing cluster. The use of the former or the latter depends on the stage of the analysis.

5.5.2 Data distribution over the Iberian ATLAS Cloud Sites

The storage in ATLAS is organized using SRM [93] Space Tokens [94]. These Space Tokens are controlled through the Distributed Data Management (DDM) [95] system and they are associated to a path to a SE. The main goals of every ATLAS Tier-2 Space Token are listed below:

- SCRATCHDISK: it stores all the output analysis job files. It is a public (all users have permissions) and temporal (one month) space.
- PRODDISK: it is used for the official ATLAS simulation production.
- HOTDISK: it hosts frequent-use dataset, such as database files.
- GROUPDISK: it is reserved for the ATLAS physic groups, that are the only ones with access and permissions.
- DATADISK: it stores the official ATLAS MC and real data. Few users with production role can store inside but every end-user can access it.
- CALIBDISK: it is destined for calibration operations related to the ATLAS detector.
- LOCALGROUPDISK: usually it is a Tier-3, reserved for local users of a given site.

Typically, DATADISK and GROUPDISK are the spaces filled the most.

Data distribution and size in the Space Tokens of the Iberian ATLAS Cloud on September 2013 are shown in Figure 5.6. In every site, the DATADISK fills more than half of the total occupancy, as it was expected.

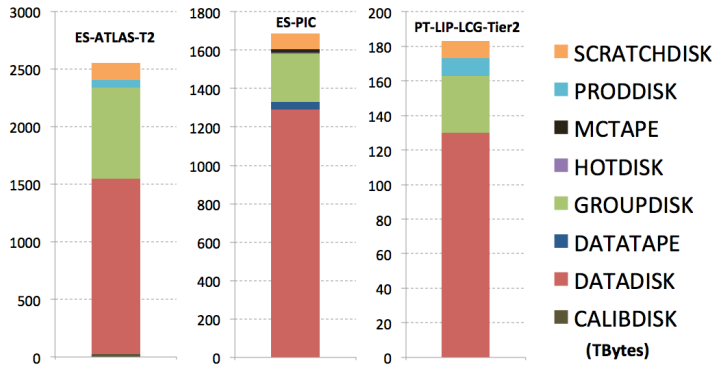


Figure 5.6: Status of ATLAS Space Tokens in the Iberian ATLAS Cloud sites in September 2013.

To get an idea about how ATLAS has worked in order to store this large amount of data, one can think of bytes that have been processed. During Run-I around 36 PB of data, belonging to collision events and MC simulations jobs, have been processed in the Iberian ATLAS Cloud sites, as Figure 5.7 shows.

These 36 PB are similar to the data that Google processed per day in 2009. Taking into account that ATLAS Collaboration has around 3000 members, and Google is spread worldwide and is used by millions of people every second, it is not bad at all the work the ATLAS Collaboration has done in its first stage of operation.

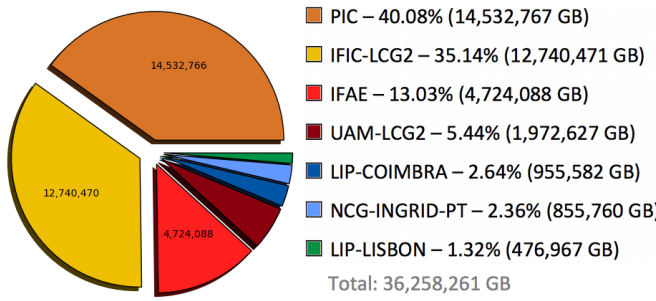


Figure 5.7: Number of physical bytes processed (in GB) in the Iberian ATLAS Cloud sites for the Run-I.

The number of physical bytes processed is translated, in some way, in MC and real data. These data are transferred over the Tiers after the reprocessing and after the ATLAS official production. Figure 5.8 shows the transfer throughput (it is the size of the amount of data transferred as a function of time) from ATLAS sites to Iberian ATLAS Cloud and vice versa during the Run-I. In this figure, it can be seen that there is an increase in data throughput around September 2011, when the changes in the Computing Model were applied. In this period, the transfer throughput from the ATLAS sites to the Iberian ATLAS Cloud sites reached 550 MB/s in September 2012.

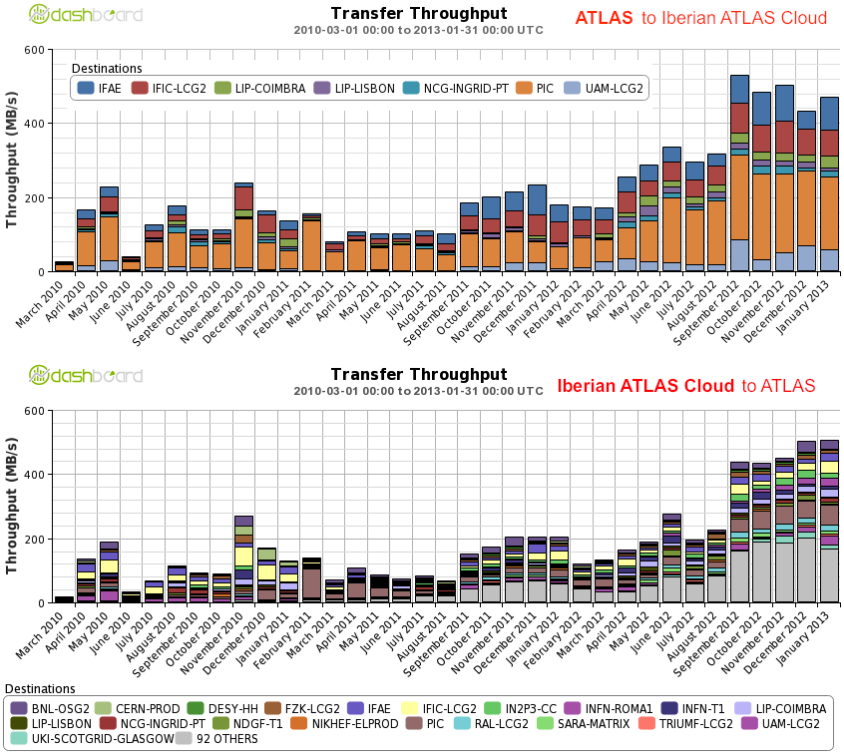


Figure 5.8: Throughput from the ATLAS sites to the Iberian ATLAS Cloud sites (top) and vice versa (bottom) for the Run-I.

In the dataset replication, primary and secondary replicas are distinguished by the new data distribution strategy. If there are many replicas in the Grid, a copy is most likely to be available. Therefore, the bottlenecks are reduced when using Grid jobs. Primary replicas are distributed according to the ACM, at Tier-1s for redundancy and at Tier-2s for analysis. In order to increase analysis opportunities and to use the remaining available disk space, secondary additional replicas of the most popular data are made. There is a tool called Panda Dynamic Data Placement (PD2P) that places secondary replicas at Tier-2s based on an on-demand replication system as well as usage, so that the waiting time of analysis jobs is reduced and a better use of the available storage capacity is made. Besides that, multi-cloud production and direct inter-cloud transfer make Tier-2s less dependent on Tier-1s and therefore, its role has become more important.

The Iberian ATLAS Cloud Tier-2s sites are getting more datasets (around 53%) than Tier-1 (PIC) as it is shown in Figure 5.9. Around 20 PB of data have been transferred to the Iberian ATLAS Cloud sites during the Run-I.

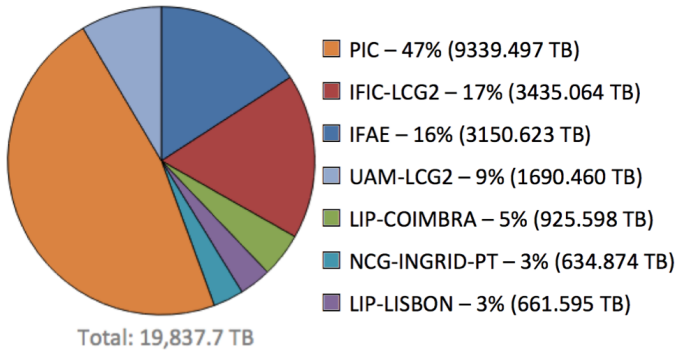


Figure 5.9: Data transfer volume in the Iberian ATLAS Cloud sites for the Run-I.

5.5.3 Distribution of Simulation and Analysis Jobs

In order to optimize our physics analysis outputs and to maximize the use of available CPU and disk resources, production shares (fees) are fixed in order to limit the group production jobs at Tier-1s. Moreover, the share of analysis jobs has been reduced at Tier-1s. Thus, Tier-2s are acquiring more tasks than before. The reconstruction of data is favoured in Tier-1s, whereas simulation and analysis jobs are favoured in Tier-2s.

Figure 5.10 contrasts the amount of completed jobs in the different sites of the Iberian ATLAS Cloud. During the Run I, the Tier-1 has processed around 41% of the total number of jobs. The number of events processed related to these completed jobs during the Run I is shown in Figure 5.11 and is around 163000 million events.

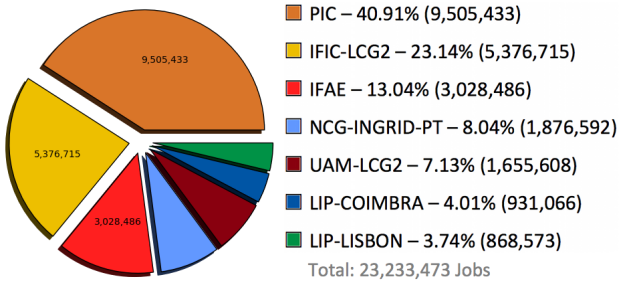


Figure 5.10: Completed jobs (50% analysis and 50% production) in the Iberian ATLAS Cloud sites for the Run-I.

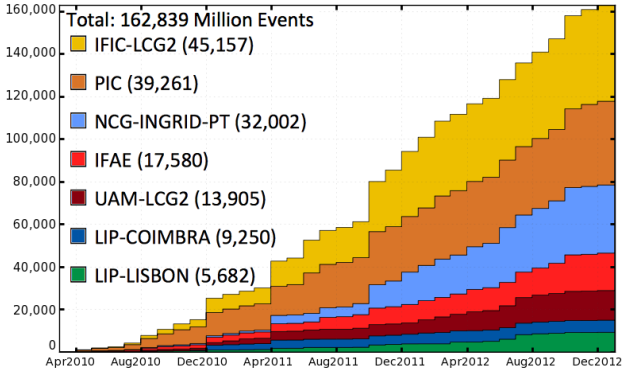


Figure 5.11: Number of events processed belonging to simulation (same as production) plus analysis jobs for the Run-I.

5.5.4 Cloud performance in terms of Availability and Reliability

Controlling the performance and the availability of the ATLAS Tiers is of great importance for the WLCG Collaboration. The site availability metrics

are calculated by the Service and Availability Monitoring system (SAM) [96], which runs different tests at regular intervals throughout the day. A site is supposed to be available if a given set of critical tests is completed successfully. These metrics defines **availability** and **reliability** as follows:

$$availability = \frac{U}{TT - TU} \qquad reliability = \frac{U}{TT - D - TU} \tag{5.1}$$

where U denotes uptime (time during which the site is available), TT refers to the total time, TU is the time during which the status was unknown and D denotes the scheduled downtime.

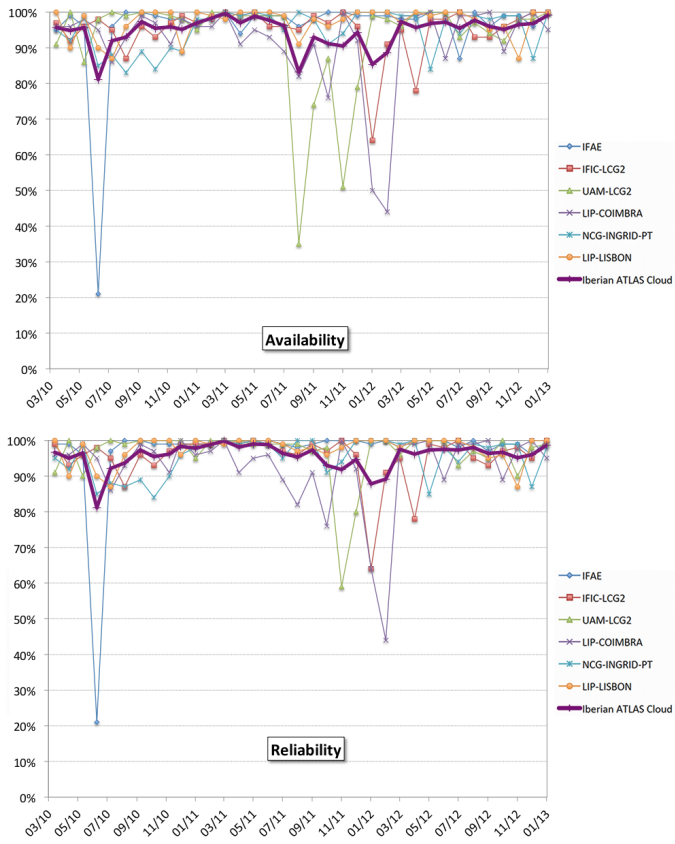


Figure 5.12: Availability and Reliability of Iberian ATLAS Cloud Tier-2s for the Run-I.

Figure 5.12 shows the availability and reliability evolution over the Run-I related to the Iberian ATLAS Cloud Tier-2s sites. As can be observed, both cases show that the average over all Iberian ATLAS Cloud sites is always in the interval $90 - 100\%$.

Another availability metric to take into account is evaluated by the ATLAS Computing Operations. It consists of the overall time that the site spent online considering also the Hammer Cloud (HC) exclusions, which is an automatic system that sends analysis jobs in order to test the proper working of the sites. This metrics determines the Tier-2 qualification. This is called **Hammer Cloud test** [97], and is shown in the Figure 5.13. The HC test average efficiency is greater than 90% during the Run-I for the Iberian ATLAS Cloud sites. IFAE, INGRID and IFIC queues lead the list with 96.1%, 94.24% and 93.39% respectively.

PANDA queue	SITE Name	TIER	CLOUD	History plot From September 2011 to September 2013	offline		brokeroff		online		NoQueue		test	
					%	count	%	count	%	count	%	count	%	count
ANALY_IFAE	ifae	T2D	ES		1.26	15	0.47	19	96.1	133	0	0	2.11	49
ANALY_IFIC	IFIC-LCG2	T2D	ES		1.3	12	1.35	26	93.39	208	0	0	3.91	95
ANALY_LIP-Coimbra	LIP-Coimbra	T2	ES		9.08	24	3.51	61	84.66	135	0	0	2.69	69
ANALY_LIP-Lisbon	LIP-Lisbon	T2D	ES		2.37	14	1.02	19	63.24	139	0	0	2.89	39
ANALY_NCG-INGRID-PT	NCG-INGRID-PT	T2D	ES		1.87	13	1.36	21	94.24	158	0	0	2.47	49
ANALY_PIC	pic	T1	ES		2.6	14	0.85	23	89.39	139	0	0	1.56	47
ANALY_PIC_SL6	pic	T1	ES		0.4	1	0	0	7.66	4	0	0	0.23	3
ANALY_UAM	UAM-LCG2	T2D	ES		4.87	58	10.91	49	77.57	176	0	0	6.34	97

Figure 5.13: ATLAS availability of the Iberian ATLAS Cloud analysis queues for the Run-I. The HC tests were born in Summer 2011.

5.5.5 Software and Computing for Run-II

The LHC took advantage of the First Long Shutdown (LS1, February 2013 to January 2015) in order to maintain and upgrade their installations, aiming to restart their operation in spring 2015 (Run-II) at higher centre-of-mass energies (13 TeV for pp collisions), at less bunch spacing (25 instead of 50 ns) and at higher luminosities, leading to almost double the average number of collisions per bunch-crossing (40 instead of 20). An increase in average data-taking rate up to 1 kHz was also assumed.

Before the start of the LHC operations, the building blocks of the ATLAS Distributed Computing (ADC) architecture were designed and deployed. Although the existing tools worked quite well during the Run-I, they showed certain limitations, i.e., the manpower required for ATLAS operations was too

high. The lessons learned during the Run-I led to redesign the workload management systems and add some other services to help cope with the increment of data volumes and the different kind of computing resources [98].

The DDM system was completely redesigned in 2013 and the new Rucio [99] tool has been deployed progressively during LS1. Certain aspects of the previous DDM implementation (dq2) have been improved in Rucio. It has data discovery based on name and metadata, it does not depend on an external file catalogue, it supports multiple data management protocols in addition to SRM (for instance, WebDAV, XRootD, poxis and gridftp) and it has smarter and more automated data placement tools.

Data access can be a blind alley or a bottleneck for data analysis owing to, for instance, the fact that some datasets are so popular for a short period of time that they are accessed by several analysis groups at the same time, making really difficult and slow their access. A way to avoid this situation during a high request period is to create a "data-federation", in which data on the disk in any site are directly accessible from jobs running in any other federated site. Clearly, data access tools must be clever enough to choose the best data replica to access, depending on the bandwidth and latency between destination and all possible data source sites. Moreover, a data-federation is needed to allow either remote access to data in the case of unavailability of a given file in the local SE or sparse access to single events.

The number and complexity of the production and analysis workflows increased during the Run-I, and they are supposed to increase and diversify even more in the near future. The system for defining and submitting tasks for processing event data had to be redesigned with a new infrastructure layered. The new system, ProdSys2 [100], is made of four core components: the Request Interface lets production managers to define workflows, DEFT translates the user request into a task definition, JEDI generates the job definitions and PanDA executes the jobs in the distributed infrastructure.

The most significant software innovation has been the development of a kind of software able to run in parallel on different CPU cores. Therefore, significantly memory is saved: 1 parallel job running on an 8-core CPU uses 8 times less memory than 8 jobs running on that CPU but having an equivalent processing time. Most or all ATLAS data centres are equipped with multi-core platforms and they already set up a new batch queues for this purpose [101]. Although this software has been tested successfully, the code must still be optimized to take advantage of the additional features of modern compilers.

Many of the specific activities that the Iberian ATLAS Cloud started to do in the LS1 and continue doing in the Run-II are listed below [102]. In addition to these tasks, the infrastructure has continued to provide site support.

Operation of the Data Derivation Framework

The Data Derivation Framework [103, 104] develops a simple mechanism to control the addition of "user data" to the new persistent format in the context of the reduction framework, i.e. targeted samples are produced including common data collections for all physic groups. This framework is the chain of jobs that are executed regularly and it produces either one or several common group data collections for a given input file (real or simulated data). The output size expected for a derived dataset is supposed to be $1 - 10\ TB$, volume that fits properly with the computing resource constraints.

The contribution to the development of the Event Index Project

The Event Index Project [105, 106, 107] consists of a complete catalogue of ATLAS events (all events, real & simulated data at all processing stages). For every permanent output file, only one file identifier and the relevant features of every event are sent to the central catalogue.

The old catalogue (TAGDB) was designed a long time ago, so it is slow, unreliable, intrinsically complex and has scaling problems when the number of events exceeds 10^9. The Event Index Project is an improvement of the TAGDB and its most innovative side is the adaptation and application of the NoSQL tech for cataloguing data of large experiments.

The contents of this new project are mainly three: 1) event identifiers, 2) online trigger pattern & hit counts and 3) references (pointers) to the events at each processing stage (RAW, ESD, AOD, NTUP) in all permanent files on storage.

The implementation of the Federated Data Storage System, FAX

FAX [108, 109] is a data-federation using XRootD redirection technology, in short, it is a way to unify the direct access from the different storage services used by the ATLAS Collaboration. It has two top-level redirectors (one in Europe and one in the US) and up to now, it covers 56% of ATLAS sites, containing 85% of data. This data-federation is used in order to access the data via single entrance, to read a dataset directly from WAN, to carry data from other spaces to a local Tier-3 Xrootd disk (storage cache) and to allow users to share non-DDM data between sites. If more free space is needed, old untouched files are purged.

Progress in Multi-core

Every Iberian ATLAS Cloud site has created a Multi-core queue [110, 111] with several dedicated nodes in order to take advantage of these technologies and to match with the optimized ATLAS Software. These queues are dedicated only for multi-core jobs, thus, Single-core and Multi-core are not mixed in a CPU server. Every multi-core job is supposed to have assigned 8 cores fulfilling the same restrictions as current single core jobs.

5.6 Tier-3 within IFIC-Valencia analysis facility

In our institute, there is a Tier-3 infrastructure attached to the IFIC Tier-2. Currently, the IFIC Tier-3 [112, 113, 114] has around 150 TB of space, of which 80 TB are managed by DDM and 70 TB are controlled by IFIC. The Run-I layout of IFIC-Valencia Tier-3 is shown in Figure 5.14.

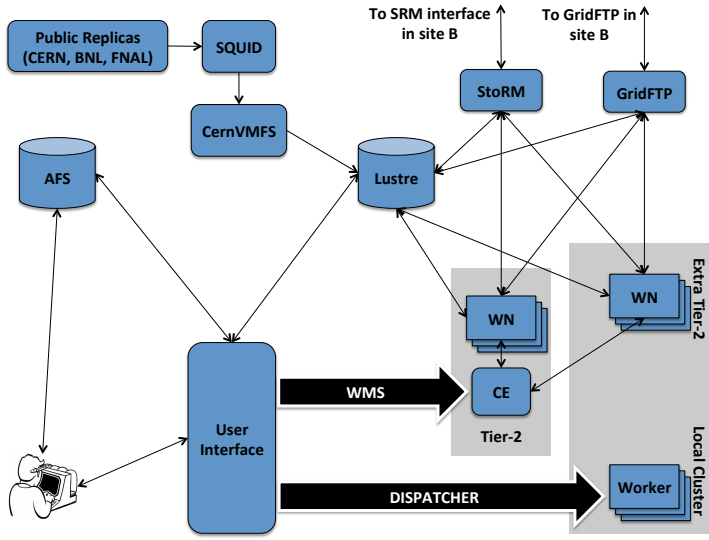


Figure 5.14: Layout of the IFIC-Valencia Tier-3 for the Run-I.

It is important to highlight that IFIC Tier-3 uses the same storage system as the Tier-2, whose central component is the **Lustre** file system, which has been mentioned before. This file system:

- Is a shared file system for clusters.
- Provides a UNIX file system interface, allowing users to access Lustre easily.
- Has a meta-directory Server (MDS) catalogue, which is the only shared resource between IFIC Tier-2 and IFIC Tier-3.
- Has 3 disk servers associated and they are dedicated exclusively to IFIC Tier-3 in order to avoid overlap with IFIC Tier-2.

In order to recover the output dataset generated in the distributed analysis phase, end-users have two options to proceed:

1. Request a subscription to the ATLAS DDM system so that the input dataset is replicated in the LOCALGROUPDISK area. Since this area is allocated at the IFIC Tier-3 facility, users can later access the input and output datasets locally.
2. Download the dataset to the local disk space using some ATLAS DDM client tools.

The choice of either depends on the dataset size. If the dataset has a considerable size (more than 500 GB) the first choice is more efficient. Nevertheless, if the file is smaller, it is worth using the second option.

Owing to the high frequency of analysis jobs that are running on the Tier-3, this computational facility must have low latency and should be highly reliable. Moreover, end-users have additional needs, related to CPU resources, in the last stages of the analysis activity workflow. In the IFIC Tier-3 these resources are:

- Five home-built User Interfaces (UI), that are used to perform interactive analysis on the final datasets produced in the analysis activity workflow phases. In every UI, Scientific Linux CERN 6.7 (64 bits) is installed.
- Two computing elements (CREAM CE): CE03 and CE05. These queues are used to run long and short analysis jobs.

Thanks to these resources, physicists working at IFIC have all they need to perform their physics analysis. Moreover, they can access the ATLAS software (Athena) and all the analysis tools from their own computer. Also, they can store their results locally using Lustre (/lustre/ific.uv.es/grid) as a local file system. In Figure 5.15 there is a schema outlining these steps.

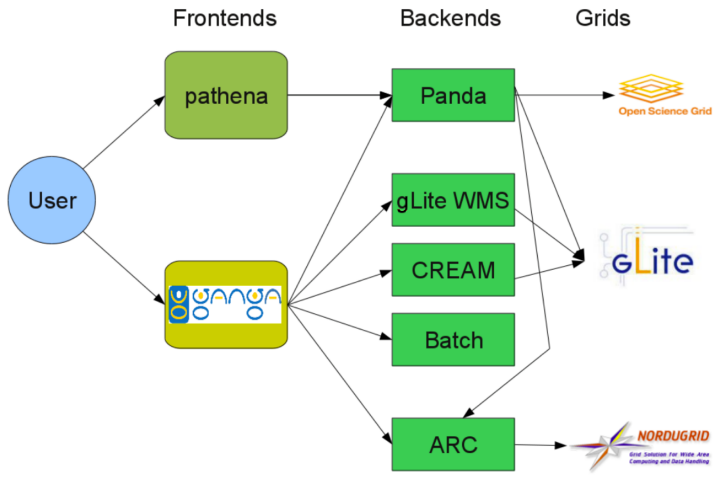


Figure 5.15: Overview of the different Grids and Job submission mechanisms offered to the users in the ATLAS experiment [115] for the Run-I.

5.6.1 Coexistence with other scientific applications running at IFIC-Valencia

The IFIC Tier-3 facility is immersed in the e-Science [116] IFIC environment, which consists of two main infrastructures aiming scientific and technological applications:

- The ATLAS Tier-2: provides storage resources and distributed computing in order to generate MC samples and to analyze real data.
- The Grid-CSIC: allocates storage and computing resources, which are given for different scientific applications, with emphasis on multidisciplinary projects [117]

IFIC Tier-3 experts are exporting their knowledge and experience to other physics groups of the institute, so that nowadays, there are more scientific applications running in Grid mode and using the resources of the Grid-CSIC infrastructure [118]: chemistry, neutrino physics, medical physics (in particular, hadron therapy), medical imaging, lattice QCD calculations, nuclear physics (Agata), etc.

The most common usages from the non-ATLAS physicists are the intensive computing on batch jobs, the computing power obtained in a distributed system

and the interactive analyses of derived data that are the result of applying algorithms to big datasets. Moreover, monitoring tools are the same for the entire infrastructure, for instance, Cacti [119] is used for checking links to data servers in order to monitor data transfers and Ganglia [120] for node monitoring.

In order to prevent interferences between the different parts of the infrastructure, they must be as isolated as possible. For example, disk pools are not the same for the different projects, as Figure 5.16 shows.

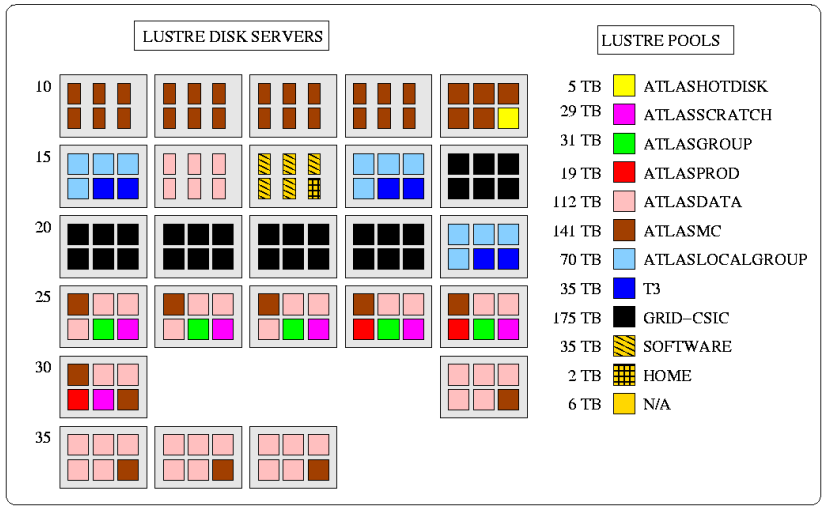


Figure 5.16: Lustre disk pool distribution at IFIC

5.7 Summary

The Iberian ATLAS Cloud has responded very efficiently during the Run-I. The changes in the Computing Model allowed improving the performance of the cloud in terms of connectivity, storage, replication, transfer, etc. All the sites of the cloud provided the resources of CPU, disk and tape needed to fulfil the ATLAS pledge. The required Distributed Analysis tools were provided for the users in order to use/store data and produce experimental results, i.e., the observation of a new particle in the search for the Standard Model Higgs boson.

The LHC underwent a period of maintenance and planned upgrades, the LS1, to prepare the collider for a higher energy and luminosity, and all the implications that this entails. During this period, facilities connected with the experiments were improved, accelerators were upgraded, electronics and computing evolved, etc. This period served to optimize the experiences learned during the Run-I and improve them during the Run-II. Several of the LS1 upgrades were tested for a long period of time in an exercise called Data Challenge (DC14) [98, 121], which claimed to get ATLAS ready for Run-II physics. During DC14 the Integrated Simulation Framework (ISF) was commissioned in the context of physics analyses, large-scale jobs of the updated reconstruction algorithms and the new distributed computing tools were run, and finally, the Run-II Analysis Model was tested, resulting in a gain of experience with the Run-II analysis framework.

Run-II has already started and the ATLAS experiment is facing new challenges and improvements in Computing and Data Management. During this period, the Iberian ATLAS Cloud sites will try to repeat the hard work they have done so far, providing the required services to fulfil the new ATLAS pledge (disk and CPU) and implementing the necessary changes in order to address the new Run-II challenges. These are: applying several improvements in the Distributed analysis and the Derivation Framework, implementing the Federated Data Storage System (ATLAS FAX), providing multi-core queues to serve multi-core jobs, and the most important, developing the ATLAS Event Index Project. The sites of the Iberian ATLAS Cloud performed their designated tasks and successfully meet the challenges.

6

Selection and reconstruction of $t\bar{t}$ events at the LHC

In this chapter the selection and reconstruction of $t\bar{t}$ events is discussed. Two strategies, the **Resolved** and the **Boosted**, are distinguished. The first strategy, the classical method, was developed for top quark production at rest. It resolves all top decay products individually. The second strategy is designed specifically to deal with the collimated topology that arises in the decay of very energetic top quarks. It is the basis of the two analyses presented in Chapters 7 and 8.

6.1 Data and Monte Carlo samples

The two analyses described in this thesis were performed using the pp collisions recorded by the ATLAS detector during the 2012 run at $\sqrt{s} = 8\ TeV$, corresponding to an integrated luminosity of $20.3 \pm 0.6\ fb^{-1}$. The average number of collisions per bunch-crossing was around 21. Data recorded during stable beam conditions and with all sub-detectors systems operational are considered.

Single lepton triggers for electrons and muons were applied. The isolated lepton trigger with a p_T threshold set at 24 GeV was used together with a trigger for non-isolated leptons with 60 GeV of threshold. The p_T threshold for single isolated (non-isolated) muons was 24 GeV (36 GeV).

The response of the detector was simulated with a GEANT4-based [76] simulation of the ATLAS detector. Almost all simulated samples were passed through this full simulation (FullSim). A few samples were passed to a fast simulation (FastSim) using parameterized showers in the calorimeters [78]. Simulated events were reconstructed using the same reconstruction software

used for collision data and weighted so that the average number of pp interactions per bunch-crossing agrees with that observed in data.

The nominal $\mathbf{t\bar{t}}$ sample was generated using the Powheg generator [122] interfaced with the parton shower Pythia6 [123] with the Perugia2011C tune [124]. The CT10 next-to-leading order PDF set [125] was used and the top quark mass was set to $m_{top} = 172.5\ GeV$. The resummation damping factor h_{damp}, a parameter which controls the ME-to-PS matching in Powheg and regulates the high-p_T radiation, was set to the top quark mass ($h_{damp} = m_{top}$). Electroweak corrections were applied on the top quark kinematics to implement the theoretical calculations of References [126, 127, 128].

Alternative $t\bar{t}$ samples, used to evaluate several systematics related to the modelling uncertainty, were simulated using Powheg and MC@NLO v4.1 [129, 130] generators interfaced with Herwig v6.520 [131] for the parton showering and hadronization, and Jimmy v4.31 [132] for the modelling of multiple parton scattering. In all cases, the $t\bar{t}$ samples were simulated in the ℓ+jets channel and normalized to the NNLO+NNLL cross-section $\sigma_{t\bar{t}} = 253^{+13}_{-15}\ pb$.

Production of vector bosons in association with high-p_T jets, such as $\mathbf{W + jets}$ and $\mathbf{Z + jets}$, are also a significant background in this analysis. Samples with these events with up to five additional partons were produced with the Alpgen v2.13 generator [133] interfaced with Pythia6. Configurations with heavy-flavour quarks ($Wb\bar{b}$, $Wc\bar{c}$, Wc, $Zb\bar{b}$ and $Zc\bar{c}$) were included. These samples were normalized to the inclusive NNLO cross-section [134, 135, 136, 137, 138, 139] and, in the case of the $W + jets$ samples, also corrected with an additional scale factor (SF) derived from data.

Production of **single top** quark was generated using several generators. Powheg generator was used to estimate the $s-$ and $Wt-$ channels whereas AcerMC was the one used for the $t-$channel production. Both generators were interfaced with Pythia using the Perugia2011 tune for parton shower modelling and the CTEQ6L1 PDF set. All these samples were normalized to approximate NNLO cross-sections [140, 141, 142].

Diboson samples were modelled using the Sherpa generator [143] with the CT10 PDF set and with up to three extra partons in the matrix element.

Production of heavy gauge bosons in association with $t\bar{t}$, $\mathbf{t\bar{t}V}$, was modelled using MadGraph5 [144, 145, 146] interfaced with Pythia and normalized to NLO cross-section prediction [147].

The data and MC datasets that were used in this analysis are listed in Appendix F in Tables F.1, F.2, F.3, F.4 and F.5.

6.2 Object definitions

In this section, basic final state objects are defined.

Electron candidates are identified in the inner detector as charged particle tracks which have energy deposits associated in the EM calorimeter cluster. The information from the EM calorimeter defines the energy and the ID track gives the angular direction of the electron. The tight electrons must have a transverse energy $E_T > 25\ GeV$ and the pseudorapidity of the energy deposit cluster in the EM calorimeter $|\eta_{cluster}| < 2.47$. Those electron candidates in the transition region between the calorimeter barrel and end-cap ($1.37 < |\eta_{cluster}| < 1.52$) are excluded.

Muon candidates are identified in the muon spectrometer as track segments which have tracks associated in the inner detector. The track segment p_T is determined by a global fit of the track segment which takes into account the energy loss in the calorimeters. The track segment is required to have a longitudinal impact parameter relative to the collision vertex $|z_0| < 2\ mm$ and a transverse impact parameter relative to the collision vertex divided by its uncertainty $|d_0/\sigma(d_0)| < 0.3$, meaning that the track segment is consistent with coming from the hard scatter. Additionally, muon candidates are required to have $p_T > 25\ GeV$ and be in the fiducial region $|\eta| < 2.5$.

Jets are reconstructed using the anti-k_t algorithm, already described in Section 2.1, with radius parameter $R = 0.4$ (small-R jets) or $R = 1.0$ (large-R jets), using as input clusters of calorimeter cells that are topologically connected and calibrated to the hadronic energy scale [148]. The energies of all jets and the masses of all large-R jets have been calibrated to their values at particle level. The jet vertex fraction (JVF) is computed as the summed p_T of all tracks matched to the jet from the primary vertex divided by the summed p_T of all tracks matched to the jet. Small-R jets are required to have $p_T > 25\ GeV$ and be in the fiducial region $|\eta| < 2.5$. If a small-R jet has $|\eta| < 2.4$ and $p_T < 50\ GeV$ is rejected only when $|JVF| < 0.5$, reducing the number of jets generated by PU interactions. To suppress the effect of the PU and to improve background rejection, the jet trimming algorithm, already described in Section 2.4, is applied over the large-R jets. In trimming, the jet constituents are reclustered into subjets using the k_t algorithm with smaller radius parameter ($R_{sub} = 0.3$) and then, subjets with $f_{cut} < 0.05$ are removed. The properties of the trimmed jet are then recalculated using the selected subjets and those considered for the analysis must satisfy $|\eta| < 2.0$ and $p_T > 300\ GeV$.

A **b-tagging** algorithm is applied to the small-R jets. Information from the track impact parameters, the secondary vertex location and the decay topology is combined in a neural network based algorithm (MV1 [149]), whose operating point is chosen such that the b-tagging efficiency in $t\bar{t}$ events is 70 %.

The $\mathbf{E_T^{miss}}$ (the magnitude of missing transverse momentum vector) is calculated from the vector sum of the transverse energy of all calorimeter cells. Their energy is corrected on the basis of the associated physic objects. This means that clusters associated with electrons and small-R jets are replaced by the calibrated energies of these objects and the contribution of muons is added using their momentum obtained from the ID and the muon spectrometer. The E_T^{miss} performance is established from the difference between $Z \to \ell\ell$ and $W \to \ell\nu$ distributions in data and MC events.

Since leptons deposit energy in the calorimeters, they may overlap with jets. To avoid double counting and in order to improve the reconstruction efficiency in the boosted topology, an **overlap removal** procedure [150] is applied. First of all, this method considers electrons and small-R jets if they are close within $\Delta R(e, jet_{R=0.4}) < 0.4$. Then, the electron 4-vector is subtracted from the jet 4-vector and the JVF is recalculated removing the electron track. If the recalculated jet does not pass the original jet criteria, it is rejected. Secondly, electrons that are still close to the recalculated small-R jet within $\Delta R(e, jet_{R=0.4}) < 0.2$ are discarded. Muons are removed if they fall inside a cone of $\Delta R(e, jet_{R=0.4}) < 0.04 + 10\ GeV/p_{T,\mu}$ around a small-R jet axis.

6.3 Event selection

The following common preselection criteria are used to select ℓ+jets events in the algorithms for the boosted and the resolved topologies.

Events are required to have one reconstructed primary vertex with five or more associated tracks. Each event must contain exactly one reconstructed isolated lepton candidate (electron or muon) that matches geometrically to the lepton-trigger object. The sum of the transverse mass of the lepton and E_T^{miss} is identified as:

$$m_T^W = \sqrt{2p_T^{lepton} E_T^{miss}(1 - \cos\Delta\phi)} \tag{6.1}$$

where $\Delta\phi$ is the azimuthal angle between the lepton and E_T^{miss} vectors. Events are rejected if do not satisfy that $E_T^{miss} > 20\ GeV$ and $E_T^{miss} + m_T^W > 60\ GeV$ (this is required in order to suppress QCD multi-jet events).

6.3.1 Event selection for resolved topology

In the classical **resolved-topology** selection, top and anti-top quark candidates are produced at rest. All six final-state objects are reconstructed individually ($t\bar{t} \longrightarrow W^{+}bW^{-}\bar{b} \longrightarrow \ell^{\pm}\nu q\bar{q}b\bar{b}$). This event selection requires at least four small-R jets and at least one of them should be b-tagged. This selection is summarized in Table 6.1.

6.3.2 Event selection for boosted topology

In **boosted-topology events**, the decay products from top and anti-top quarks candidates tend to be back-to-back in the $t\bar{t}$ rest frame. Since the top quarks are boosted, their decay products are collimated in a narrow cone around the direction of the initial candidates. The top quark candidate decaying semileptonically is reconstructed as a W boson plus a (small-R) b-jet close to the isolated lepton from the W boson decay. The top quark candidate decaying hadronically is reconstructed as a single fat jet.

Boosted events must have a small-R jet close to the selected charged lepton within $\Delta R(\ell, jet_{R=0.4}) < 1.5$. If there are more than one small-R jets which satisfy this requirement, the one with highest p_T is selected. Events must have an energetic trimmed large-R jet with $p_T > 350\ GeV$, $m > 100\ GeV$ and a first k_t splitting scale $\sqrt{d_{12}} > 40\ GeV$. This large-R jet is selected if it is well-separated from both, the lepton and the small-R jet associated to the lepton, by requiring $\Delta\phi(\ell, jet_{R=1.0}) > 2.3$ and $\Delta R(jet_{R=1.0}, jet_{R=0.4}) > 1.5$. Finally, at least one of the small-R jets must be b-tagged and matched to either top or anti-top quark candidate and it is required to be within either $\Delta R(\ell, jet_b) < 1.5$ or $\Delta R(jet_{R=1.0}, jet_b) < 1.0$. This selection is summarized in Table 6.2 and sketched in Figure 6.1.

The strategy for selecting boosted ℓ+jets events was developed for the ATLAS $t\bar{t}$ resonance search group [151] and followed by the ATLAS differential cross-section [152] and the ATLAS boosted A_C [153, 154] groups. But there is a difference, the first group used the fat jet trigger in the delayed stream, to recover the muon trigger efficiency, while the others did not use it. Moreover, the "Resolved selection" was used just by the ATLAS $t\bar{t}$ resonance search group.

	electron + jets	muon + jets
Leptons	$\lvert z_0\rvert < 2$ mm $I_{mini} < 0.05$ $\lvert\eta\rvert$ <1.37 or 1.52< $\lvert\eta\rvert$ <2.47 $p_T > 25$ GeV	$\lvert z_0\rvert < 2$ mm & $\lvert d_0/\sigma(d_0)\rvert < 3.$ $I_{mini} < 0.05$ $\lvert\eta\rvert$ <2.5 $p_T > 25$ GeV
Anti-k_t R=0.4 jets	$p_T > 25$ GeV $\lvert\eta\rvert$ <2.5 JVF > 0.5 (if $p_T < 50$ GeV)	

Overlap removal	if $\Delta R(e, \text{jet}_{R=0.4})$ <0.4:
	$\text{jet}'_{R=0.4} = \text{jet}_{R=0.4} - electron$
	
	if $\Delta R(e, \text{jet}'_{R=0.4})$ <0.2:
	e removed
		if $\Delta R(\mu, \text{jet}'_{R=0.4}) < 0.04 + 10$ GeV$/p_T(\mu)$:
		μ removed

jets	1) ≥ 4 small-R jet satisfying $p_T > 25$ GeV ; $\lvert\eta\rvert < 2.5$; $JVF > 0.5$ 2) An event which fails the previous step is selected if it contains 3 small-R jets with one $m_{jet} > 60$ GeV 3) at least one small-R jet is b-tagged	

Table 6.1: Summary of the resolved ℓ+jets event selection.

	electron + jets	muon + jets
Leptons	$\lvert z_0\rvert < 2$ mm $I_{mini} < 0.05$ $\lvert\eta\rvert$ <1.37 or 1.52< $\lvert\eta\rvert$ <2.47 $p_T > 25$ GeV	$\lvert z_0\rvert < 2$ mm & $\lvert d_0/\sigma(d_0)\rvert < 3.$ $I_{mini} < 0.05$ $\lvert\eta\rvert$ <2.5 $p_T > 25$ GeV
Anti-k_t R=0.4 jets	$p_T > 25$ GeV $\lvert\eta\rvert$ <2.5 JVF > 0.5 (if $p_T < 50$ GeV)	

Overlap removal	if $\Delta R(e, \text{jet}_{R=0.4})$ <0.4:
	$\text{jet}'_{R=0.4} = \text{jet}_{R=0.4} - electron$
	
	if $\Delta R(e, \text{jet}'_{R=0.4})$ <0.2:
	e removed
		if $\Delta R(\mu, \text{jet}'_{R=0.4}) < 0.04 + 10$ GeV$/p_T(\mu)$:
		μ removed

Leptonic top	≥ 1 small-R jet satisfying $\Delta R(l, \text{jet}_{R=0.4}) < 1.5$	
Hadronic top	The leading trimmed anti-k_t large-R jet $p_T > 300$ GeV ; $m > 100$ GeV ; $\sqrt{d_{12}} > 40$ GeV $\Delta R(\text{jet}_{R=1.0}, \text{jet}_{R=0.4}) > 1.5$; $\Delta\phi(l\text{jet}_{R=1.0}) > 2.3$	
b-tagging	At least one of: 1) the leading small-R jet satisfying $\Delta R(l, \text{jet}_{R=0.4}) < 1.5$ is b-tagged; 2) at least one small-R jet satisfying $\Delta R(\text{jet}_{R=1.0}, \text{jet}_{R=0.4}) < 1.0$ is b-tagged	
Reconstructed event	$m_{t\bar{t}} > 750$ GeV	

Table 6.2: Summary of the boosted ℓ+jets event selection.

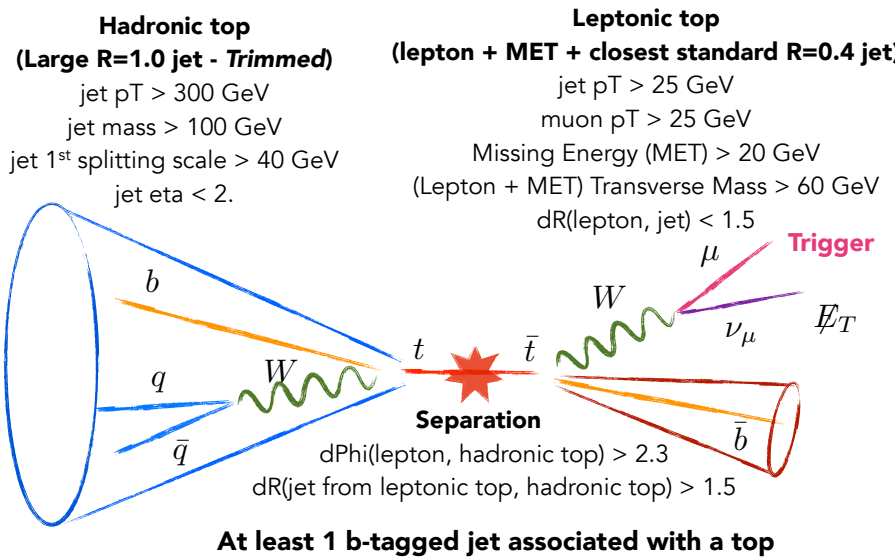


Figure 6.1: Schematic representation of the boosted ℓ+jets event selection.

6.3.3 Event selection for orthogonal and independent analyses

In the search for $t\bar{t}$ resonances, developed in Chapter 7, events that do not pass the boosted selection are tested against the resolved selection. In this way, resolved and boosted samples are kept orthogonal, i.e., without overlap.

Analyses that aim for a selection of all $t\bar{t}$ events (either at rest or boosted) must combine both selections and deal with the overlap. This is the case of the measurement of the $t\bar{t}$ A_C with boosted topologies explained in Chapter 8.

6.4 Reconstruction of the $t\bar{t}$ pair

In the boosted and the resolved reconstruction strategies, the $t\bar{t}$ final state objects in the ℓ+jets channel are combined into top and anti-top quark candidates. The first step, the reconstruction of the neutrino candidates, is common to both strategies. The E_T^{miss} vector is identified with the neutrino transverse momentum ($p_{T,\nu}$). The longitudinal component of the neutrino momentum ($p_{z,\nu}$) is calculated by requiring that the system formed by the lepton plus missing transverse momentum must have the W boson mass (M_W) and solving

the resulting quadratic equation, given in Equation 6.2 [155]. If no real solution exists, the E_T^{miss} vector is varied by the minimal amount required to produce exactly one real solution.

$$p_{z,\nu}^2 - 2 \cdot \frac{\mu\, p_{z,\ell}}{E_\ell^2 - p_{z,\ell}^2} \cdot p_{z,\nu} + \frac{E_\ell^2\, p_{\ell,\nu}^2 - \mu^2}{E_\ell^2 - E_{z,\ell}^2} = 0$$
$$\text{with} \quad \mu = \frac{M_W^2}{2} + p_{T,\ell}\, p_{T,\nu} \cos(\Delta\phi) \tag{6.2}$$

6.4.1 $t\bar{t}$ reconstruction with resolved topologies

After the resolved event selection, the selected objects are:

- One isolated charged lepton.
- Four small-R jets (at least one of them must be b-tagged).
- Missing transverse energy of the event, E_T^{miss}, which is associated to the escaping neutrino. If two $p_{z,\nu}$ real solutions are found in Equation 6.2, both are tested using a χ^2 algorithm (given in Equation 6.3) and the combination that gives the smallest value of χ^2 is used.

$$\chi^2 = \left[\frac{m_{jj} - m_W}{\sigma_W}\right]^2 + \left[\frac{m_{jjb} - m_{jj} - m_{t_h - W}}{\sigma_{t_h - W}}\right]^2 + \\ + \left[\frac{m_{j\ell\nu} - m_{t_\ell}}{\sigma_{t_\ell}}\right]^2 + \left[\frac{(p_{T,jjb} - p_{T,j\ell\nu}) - (p_{T,t_h} - p_{T,t_\ell})}{\sigma_{\text{diff } p_T}}\right]^2 \tag{6.3}$$

where t_h and t_ℓ refer to the hadronic and leptonic top quark candidates respectively. The first term is a constraint using the mass of the W boson that decays hadronically. The second term is a constraint using the mass difference between the hadronic top quark and the hadronic W boson. The third term is a constraint using the mass of the leptonic top quark. And the fourth term is a constraint on the expected p_T balance between the two top quark candidates. The combination of jets, b-jets and neutrino candidates that minimizes the χ^2 is selected.

Once these objects are reconstructed, the **invariant mass of the $t\bar{t}$ system** is computed as the invariant mass of the total 4-vector of the system formed by the final objects (lepton + neutrino + four small-R jets).

6.4.2 $t\bar{t}$ reconstruction with boosted topologies

The reconstruction of the boosted $m_{t\bar{t}}$ system follows the prescription presented by the boosted top quarks decay, where the top and anti-top quark decay products tend to be clearly separated. After the boosted event selection, the selected objects are:

- One single charged lepton.
- At least one small-R jet [1] within $\Delta R(\ell, j) < 1.5$.
- One large-R jet (fat jet).
- One neutrino candidate. If two $p_{z,\nu}$ real solutions are found in Equation 6.2, the one that yields the smallest χ^2 is used.

The top quark candidate decaying semileptonically is formed by the selected small-R jet plus the lepton, while the top quark candidate decaying hadronically is identified with the selected fat jet.

At least one of the small-R jets must be tagged as b-jet. Once these objects are reconstructed, the **invariant mass of the $t\bar{t}$ system** it obtained by adding the 4-vectors of the hadronic and leptonic top quark candidates.

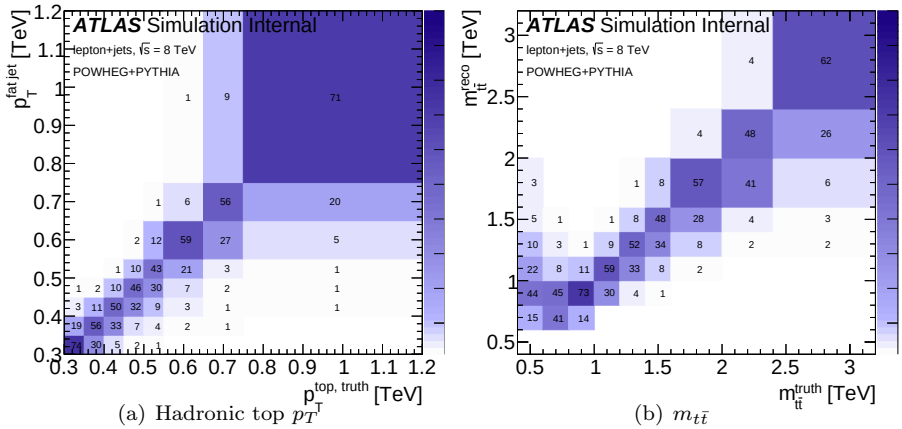
The response matrix tells us how well the reconstructed object map on the true properties. It is represented by drawing the true distribution over its corresponding reconstructed spectrum in several bins.

The mapping of the p_T of the selected hadronic top quark candidate is shown in Figure 6.2(a). The p_T migration matrix is quite diagonal (more than 50% of events are inside the diagonal bins) but slightly asymmetric. This means that the reconstruction tends to underestimate the top quark p_T. This is likely due to the trimming procedure, where the large-R jets are reclustered removing soft radiation and UE. Therefore, trimming removes p_T from the jet, pulling the reconstructed jet p_T below the true jet p_T.

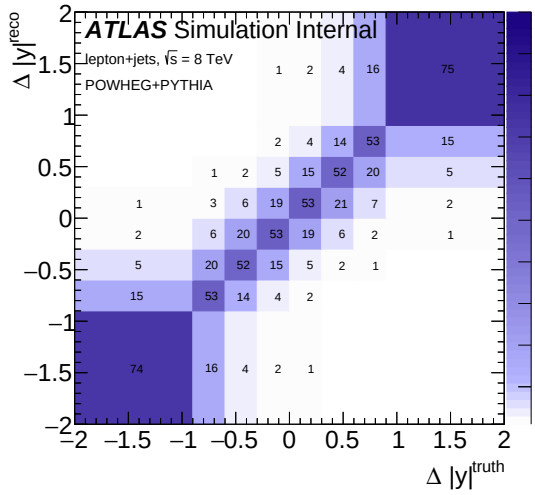
The migration matrices of the mass of the $t\bar{t}$ system are presented in Figure 6.2(b). At low $m_{t\bar{t}}$ we observe events outside the diagonal. This means that the events didn't produce a proper collimated back-to-back topology and soft jets pass the selection.

The response matrices of the rapidity difference $\Delta|y| = |y_t| - |y_{\bar{t}}|$ of the top and anti-top quarks candidates are depicted in Figure 6.2(c). The matrix is fairly diagonal, where less than 10% of events where the $\Delta|y|^{reco}$ differs by more than 0.5 from the $\Delta|y|^{true}$. This matrix is the one used in the unfolding procedure to measure the A_C (this is described in Section 8.3).

[1] if more than one small-R jet candidates exist, the jet with the largest transverse momentum is used.



(a) Hadronic top p_T (b) $m_{t\bar{t}}$



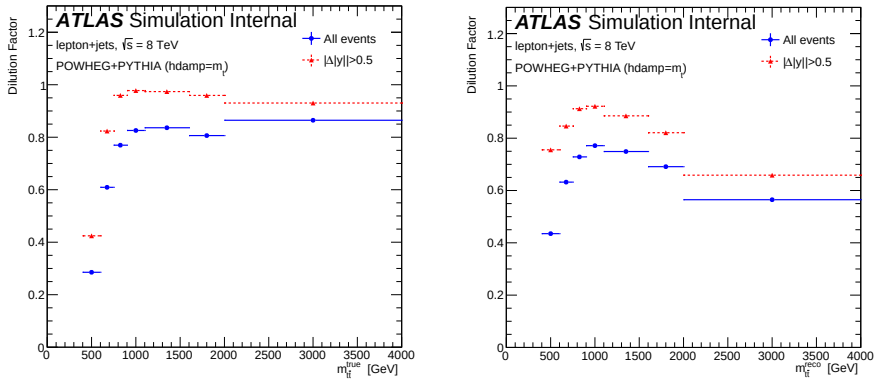
(c) $\Delta|y| = |y_t| - |y_{\bar{t}}|$

Figure 6.2: Migration matrices between reconstructed and true observables. Columns are normalized to 100%.

Migrations tend to reduce the observed asymmetry. A useful quantity in the A_C measurements is the dilution. The reconstructed and true asymmetries are related by the dilution factor D:

$$A_C^{reco} = D \cdot A_C^{truth} \qquad D = 2p - 1 \tag{6.4}$$

where p is the probability to find a $\Delta|y|^{reco}$ and a $\Delta|y|^{truth}$ with the same sign. The evolution of the dilution factor with the $m_{t\bar{t}}$ is plotted in Figures 6.3.



(a) D vs $m_{t\bar{t}}^{true}$ (b) D vs $m_{t\bar{t}}^{reco}$

Figure 6.3: The evolution of the dilution factor with the $m_{t\bar{t}}$. Blue markers indicate the total dilution, including the effect of finite y resolution and of imperfections in the identification of the top and anti-top quark decay products. Red markers represent the dilution calculated with events where $|\Delta|y|| > 0.5$, thus isolating the effect of problems in the reconstruction of the top quark candidates.

There are two contributors to the dilution:

- Bad reconstruction or misidentification of the top quark decay products can lead to a very large difference between $\Delta|y|^{reco}$ and $\Delta|y|^{truth}$.
- The finite resolution of the top quark decay products, propagated to the top quark candidates, can yield the small migrations to the wrong-sign quadrants. This leads to a broader diagonal, i.e., a sign-flip for events with a very small $\Delta|y|$.

The second contributor can be decreased by eliminating events with a small $\Delta|y|^{reco}$, requiring that events have $\Delta|y|^{reco} > 0.5$. In boosted events the typical value for D is 0.7 and 0.9 for events with $\Delta|y|^{reco} > 0.5$. The boosted selection provides significantly better $\Delta|y|$ resolution and dilution than the resolved one.

6.4.3 $t\bar{t}$ candidate events

Events with $m_{t\bar{t}} \sim 1\ TeV$ often produce a mixture of intermediate topologies, which can be reconstructed with both algorithms. This is the so-called Transition Region. Figure 6.4 shows a good example of this event, where both, resolved and boosted reconstructions can be applied.

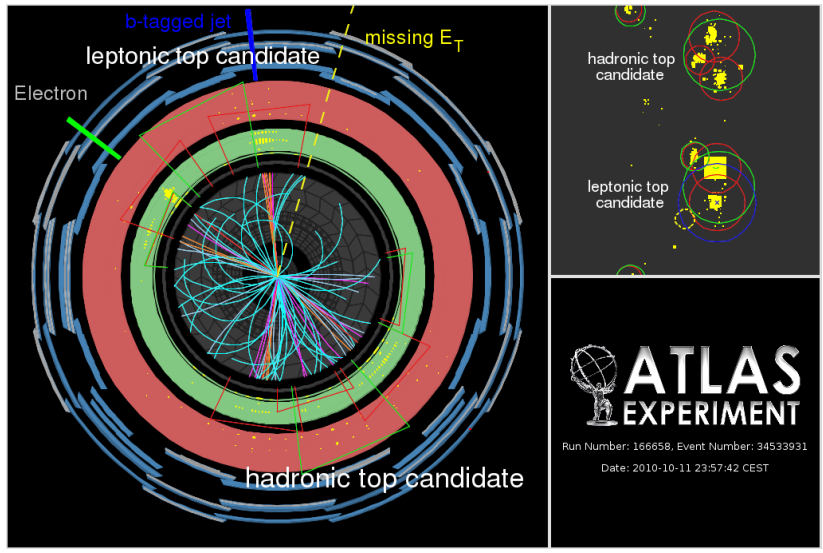


Figure 6.4: Event display for a $t\bar{t}$ candidate event. The leptonic top candidate is formed by a high p_T electron (145 GeV, 11 o'clock), moderate E_T^{miss} (1 o'clock), and a b-tagged jet at 12 o'clock. When reclustered with $R = 1.0$ it acquires a large p_T and mass as it absorbs the electron. Three jets between 4 and 6 o'clock are identified with the hadronic top quark. When reclustered with $R = 1.0$ the three jets merge into a single jet. Jets indicated in red correspond to $R = 0.4$, jets in green to $R = 1.0$.

Figures 6.5 and 6.6 show an example of a $t\bar{t}$ pair candidate with resolved and boosted topologies respectively.

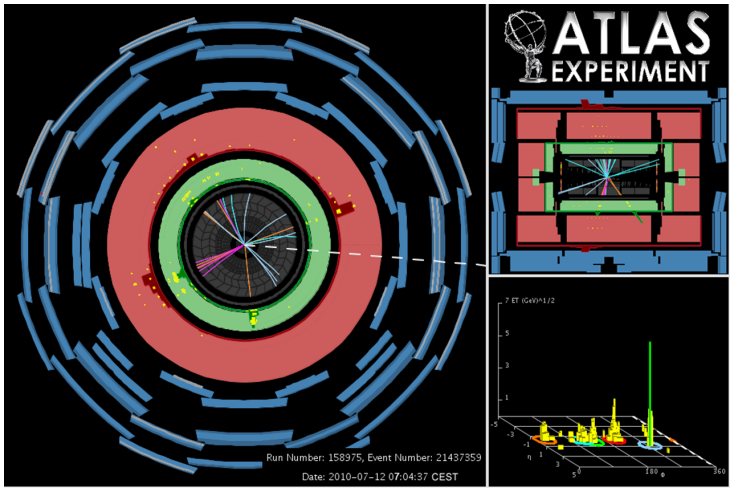


Figure 6.5: Event display of the electron plus jets candidate. The electron is shown as the orange downward-pointing track associated to the green cluster, and as the green tower in the $\eta - \phi$ lego plot. The direction of the missing transverse energy is shown as the dotted line in the $r - \phi$ view.

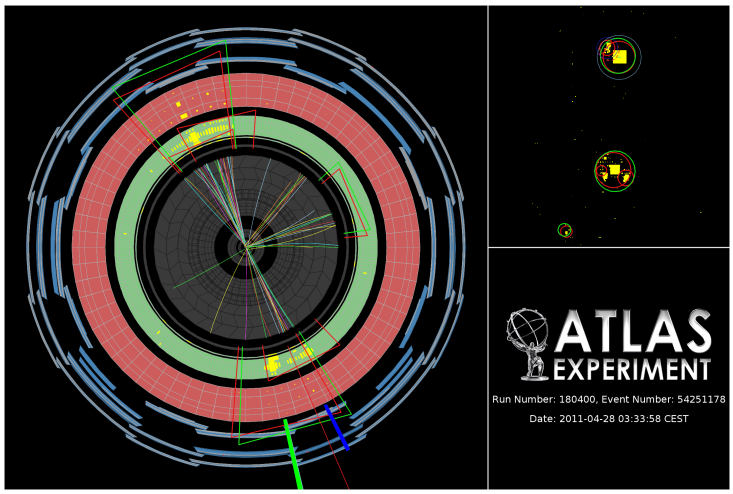


Figure 6.6: Event display for a $t\bar{t}$ candidate event with large mass, $m = 1.6\ TeV$. The left panel displays a transverse view of the charged particle tracks and calorimeter energy deposits. An $\eta - \phi$ view of the same event is shown in the upper right panel. Jets reconstructed with $R = 0.4$ are indicated in red, jets with $R = 1$ in green.

6.5 Background estimation from data

Analyses build a SM template using a combination of MC and data-driven techniques. Data are used to estimate two important backgrounds in the boosted analyses, such as W+jets and multi-jet production.

6.5.1 W+jets production

The W+jets production is the largest non-$t\bar{t}$ contribution in the total MC sample. This background is determined as the original ALPGEN+ PYTHIA W+jets MC samples corrected by the derived SF, which are applied to the normalization and flavour fractions given by the initial MC samples. Data are used to compute these derived SF.

The MC expectation is multiplied by the heavy flavour fraction SF ($f_{flavour}$) to correct for the mismodelling in the generator of the fractions of W production associated with different flavour components (called $K_{flavour}$, which are $W+b\bar{b}$, $W+c\bar{c}$, $W+c$). They are estimated in a sample with the same lepton and E_T^{miss} selections as the signal selection, but with only two small-R jets and no b-tagging requirements. The b-jet multiplicity, in conjunction with the knowledge of the efficiency for b-tagging, c-tagging and mistagging, is used to extract the heavy flavour fraction in this sample. This information is extrapolated to the signal region using the MC simulation, assuming constant relative rates for the signal and control regions.

The overall W+jets normalization can be determined by comparing the W boson charge asymmetry [156, 157] in data with that predicted in the original ALPGEN+ PYTHIAW+jets MC samples. This asymmetry measures the number of W bosons with positive and negative sign and it is well predicted by theory.

The total number of W+jets events in data is estimated as follows:

$$\begin{aligned} N_{W^+} + N_{W^-} &= \left(\frac{r_{MC}+1}{r_{MC}-1}\right)(D_{corr+} - D_{corr-}) \\ r_{\mathrm{MC}} &= \frac{N_{\mathrm{MC},W^+}}{N_{\mathrm{MC},W^-}} \end{aligned} \tag{6.5}$$

where D_{corr+} and D_{corr-} are the number of observed events with positive and negative charged leptons in the data, respectively. The **charge asymmetry method** is extended by adding b-tag information to extract the flavour-dependent SFs, $F_{bb,cc}$, F_c and F_{ll}. The results of this procedure for a standard $t\bar{t}$ selection with four or more jets are presented in Table 6.3, where the fourth

column lists the overall normalization factor for W+jets production. Every SF is applied in these boosted analyses.

channel	F_{bb}, F_{cc}	F_c	F_{ll}	norm.
e+jets	1.355 ± 0.072	0.712 ± 0.028	0.934 ± 0.005	0.89 ± 0.06
μ+jets	1.516 ± 0.080	0.662 ± 0.026	0.873 ± 0.004	0.81 ± 0.05

Table 6.3: The flavour-dependent scale factors $F_{bb,cc}$ for W+jets production with associated bottom and charm, F_c for single charm, and F_{ll} for W+light jets. The numbers are obtained for a sample with four or more jets using the charge asymmetry method extended with b-tag information described in the text. The last column indicates the correction factor for the overall normalization of the W+jets sample.

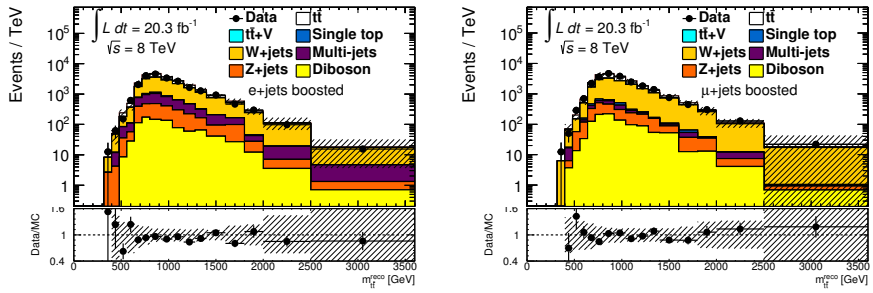
The overall normalization in the kinematic boosted region is determined by applying the flavour-blind charge asymmetry method on a region similar to the final selection. The boosted signal region itself has too few events and is too impoverished in W+jets events to apply Equation 6.5 directly. Instead, an enhanced sample enriched with W+jets events is created by relaxing the requirements on b-tagging, $\Delta\phi(\text{jet}_{R=1.0}, l)$, jet mass and $\sqrt{d_{12}}$. The SF for the e+jets channel is $0.89 \pm 0.06(stat.)$ and for the μ+jets channel is $0.81 \pm 0.05(stat.)$. After correcting by this SF, W+jets events are expected to make up approximately 6% of the total event yield in the signal region.

The shape of the W+jets background in the $m_{t\bar{t}}$ spectrum is studied in a control region (CR). The sample enriched in W+jets events mentioned before is obtained by inverting the b-tagging requirement. The distributions for the $m_{t\bar{t}}$ system candidate and hadronic top quark p_T candidate are shown in Figures 6.7 and 6.8, respectively. The CR is clearly dominated by W+jets production, with a minor contamination from $t\bar{t}$ production and several other SM sources.

6.5.2 QCD multi-jet production

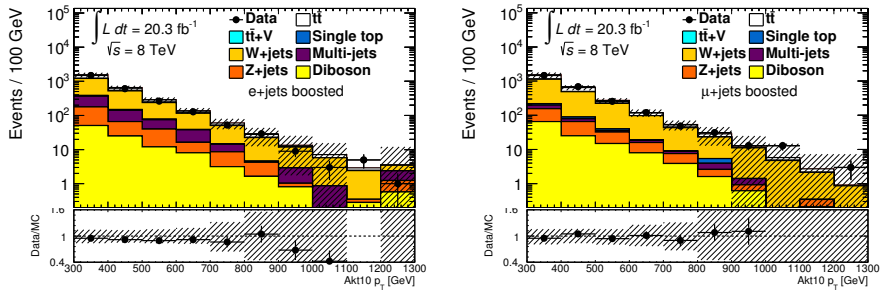
Sometimes QCD multi-jet events imitate the ℓ+jets signature. This background is derived directly from data by using the matrix-method technique (data-driven method [158]). A fake-leptons enriched sample, characterized by non-prompt leptons that satisfy the identification criteria or jets misreconstructed as leptons, is estimated by loosening the lepton identification criteria. The number of events containing fake leptons in the signal region (SR) can be computed as:

$$N_{QCD} = \frac{(\epsilon - 1)\, f}{\epsilon - f} N_T + \frac{\epsilon\, f}{\epsilon - f} N_A \tag{6.6}$$



(a) e+jets (b) μ+jets

Figure 6.7: The $m_{t\bar{t}}$ distribution in a 0-b-tag, W+jets enriched, control region for two channels.



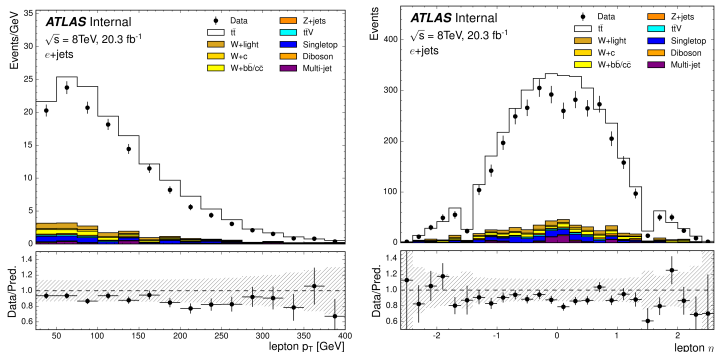
(a) e+jets (b) μ+jets

Figure 6.8: The p_T distribution of the hadronic top quark candidate in a 0-b-tag, W+jets enriched, control region for two channels.

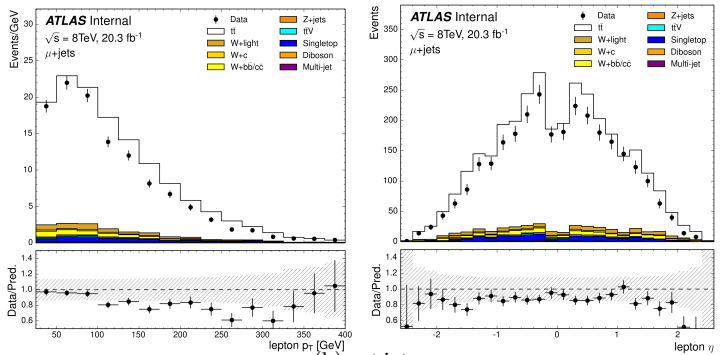
where ϵ and f are the efficiencies for leptons that passed the loose criteria and also the tight one, for real (prompt) and fake (non-prompt) leptons respectively. N_T is the number of events with a tight lepton and N_A the number of events with a loose lepton that failed the tight cuts. The rate ϵ is extracted from MC and validated in data, and the efficiency f is measured using data in fake-enhanced CRs.

6.6 Comparison of Data and MC template

Around 223300 data events pass the resolved selection and 8200 the boosted. The agreement of the data with the prediction in the **boosted** analysis is further tested by studying the distributions of several variables of interest in Figures 6.9, 6.10 and 6.11. Each figure shows different kinematic distributions for events with at least one b-tag. Figures are separated between e+jets and μ+jets. In all figures, the statistical uncertainty on the data is shown as error bars on the markers and the uncertainty on the normalization of the SM expectation is shown as a hatched band.



(a) e+jets



(b) μ+jets

Figure 6.9: Distributions of p_T and η of the lepton candidates in the boosted analysis of Reference [45]. The ratio of data and MC prediction is shown in the insets below the plots. Error bars on the data points indicate the statistical uncertainty. The hashed area, only shown in the ratio portion of each plot, shows the uncertainty of the Standard Model prediction. This includes the statistical uncertainty, the theory uncertainties on the cross-sections and the effect of detector systematics on the expected yield.

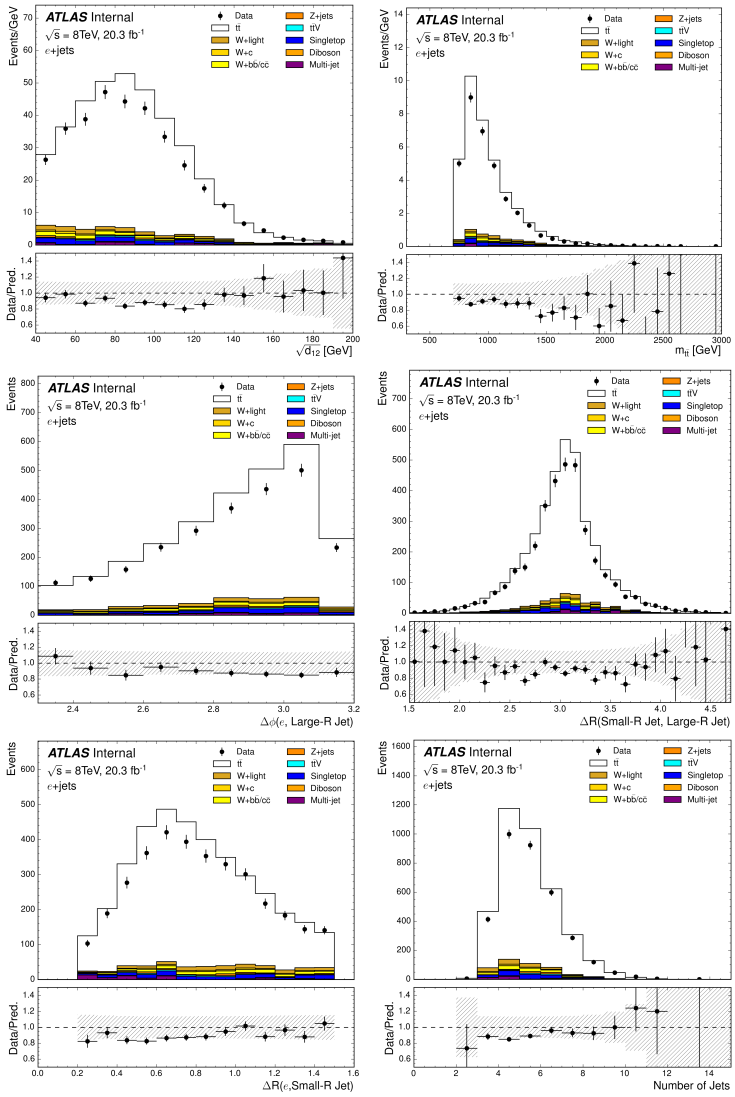


Figure 6.10: The distributions for the e+jets channel in the boosted analysis of Reference [45]. Distributions of (top) the splitting scale, $\sqrt{d_{ij}}$ and $m_{t\bar{t}}$ (middle) $\Delta\phi$(large-R Jet,ℓ) and ΔR(large-R Jet, small-R jet), and (bottom) ΔR(small-R jets, ℓ) and the number of small-R jets. The ratio of data and MC prediction is shown in the insets below the plots. Error bars on the data points indicate the statistical uncertainty. The hashed area, only shown in the ratio portion of each plot, shows the uncertainty of the SM prediction. This includes the statistical uncertainty, the theory uncertainties on the cross-sections and the effect of detector systematics on the expected yield.

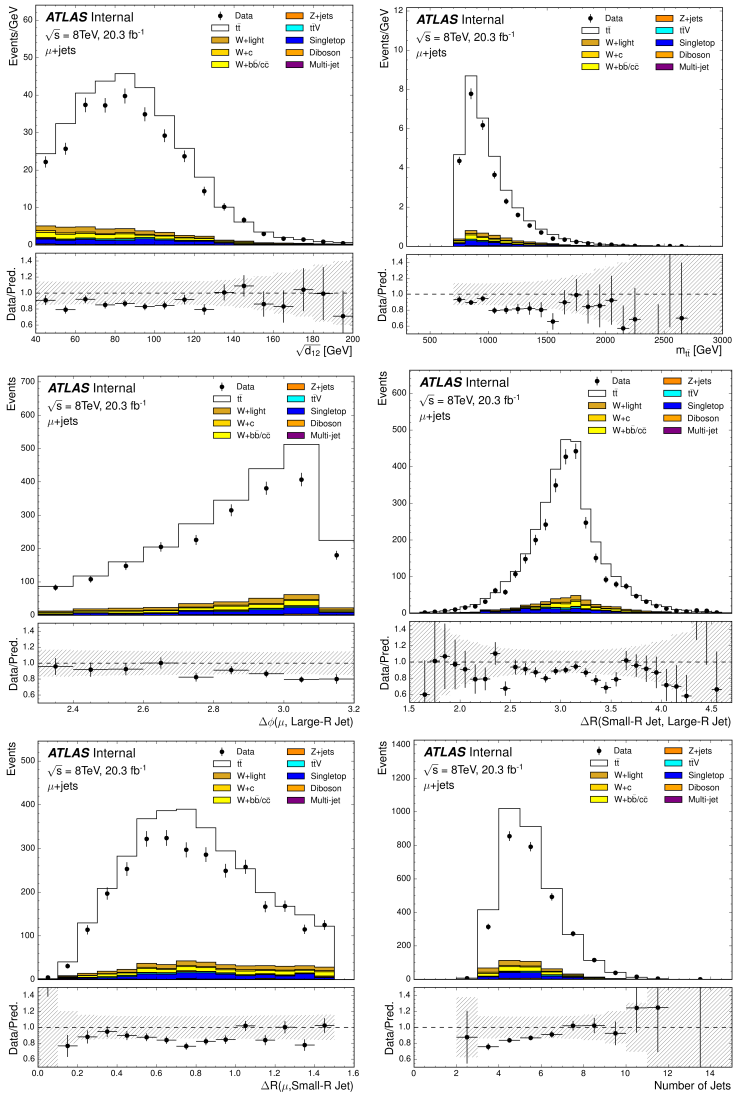


Figure 6.11: The distributions for the μ+jets channel in the boosted analysis of Reference [45]. Distributions of (top) the splitting scale, $\sqrt{d_{ij}}$ and $m_{t\bar{t}}$ (middle) $\Delta\phi$(large-R Jet,ℓ) and ΔR(large-R Jet, small-R jets), and (bottom) ΔR(small-R jets, ℓ) and the number of small-R jets. The ratio of data and MC prediction is shown in the insets below the plots. Error bars on the data points indicate the statistical uncertainty. The hashed area, only shown in the ratio portion of each plot, shows the uncertainty of the SM prediction. This includes the statistical uncertainty, the theory uncertainties on the cross-sections and the effect of detector systematics on the expected yield.

6.7 Summary

In this chapter we have described the selection and reconstruction of top quark pairs at the LHC. Two methods for selecting ℓ+jets events are discussed: the classical, resolved approach, developed at the Tevatron for top quark production at rest, and a new technique designed specifically to deal with boosted top quark production. The latter has superior acceptance and resolution for top quark pairs with a mass above 1 TeV. A SM template is constructed using a combination of MC and data-driven techniques. Data has been used to estimate two important backgrounds in the boosted analysis: W+jets and multi-jet production. Finally, a detailed comparison of data and the MC template is presented for the boosted analysis. Apart from an understood discrepancy in the yield, a reasonable agreement is found. The selection and reconstruction described in this chapter form the basis of the analyses presented in the next chapters.

7

Searches for $t\bar{t}$ resonances

7.1 Introduction

Many BSM scenarios give rise to heavy particles that couple strongly to $t\bar{t}$ pairs. This chapter is based on a search for new heavy particles that decay into $t\bar{t}$ pairs with the ATLAS experiment using an integrated luminosity of 20.3 fb^{-1} of pp collision data collected at a centre-of-mass energy of $\sqrt{s} = 8\ TeV$ [151]. The search focuses on the ℓ+jets final state in which the $t\bar{t}$ pair decays to W^+bW^-b, where one W boson decays hadronically and the other one, leptonically. This leads to a signature with one high transverse momentum lepton, large missing transverse momentum (from the escaping neutrino) and several jets.

The analysis is designed to deal both with *boosted* configurations, where the top decay products overlap in the detector, and with *resolved* configurations, where the top decay products are all well separated in the detector. Generally, the decay products are more boosted when the invariant $t\bar{t}$ mass is larger. The invariant mass spectrum of the $t\bar{t}$ candidate system is reconstructed and examined to search for hints of new physics in the form of bumps or dips in the spectrum.

Data and Monte Carlo samples, Object definitions, Event selection, Reconstruction of the $t\bar{t}$ pair and Background estimation are explained in detail in Chapter 6.

7.2 Benchmark model: The Kaluza Klein Gluon

The main benchmark model adopted for quantifying the sensitivity of high-mass $t\bar{t}$ resonance searches is the heavy Kaluza-Klein gluon (g_{KK}), as produced in Randall-Sundrum models with a single warped extra-dimension [159]. This is a broad and colour-octet vector boson with spin 1. This kind of gluon is interesting in the searches for $t\bar{t}$ resonances because it can not be observed in dileptonic resonances (it does not couple to leptons) and its BR to the top quark is higher than that to others quarks.

In this chapter, the g_{KK} proposed in [159] is chosen. In this model the nominal g_{KK} width Γ is 15.3% M, where M is the resonance mass. We also study those g_{KK} with Γ from 10% M to 40% M.

In previous searches [160, 161, 151], a g_{KK} with $m < 2.1\ TeV$ was excluded by the ATLAS Collaboration. Using a slightly different benchmark model [162], the CMS Collaboration excluded a g_{KK} with $m < 2.5\ TeV$ [163]. The main difference [164] between both benchmark models is that the g_{KK} of the CMS Collaboration has $\Gamma = 20\%\ M$ and larger production cross-section.

The MC samples related to this signal model were produced at IFIC. The interaction $q\bar{q} \longrightarrow g_{KK} \longrightarrow t\bar{t} \longrightarrow \ell^{\pm}\nu_{\ell}q\bar{q}b\bar{b}$ is simulated with MadGraph [144, 145, 146]. The parton shower and hadronization are simulated using Pythia [123]. The MadGraph expectation for the cross-section of the RS g_{KK} is shown in Figure 7.1. The red curve is the central expectation and the shaded green band shows the uncertainty. This band is computed varying independently the factorization and renormalization scales. The nominal values, identified with the mass of the resonance, are varied to twice and half, leading to 9 combinations of the scales for every mass point. The combinations that yields the greatest and the smallest deviation from the nominal value are taken as the boundaries of the error band.

7.3 Systematic uncertainties

The SM expectation for the $t\bar{t}$ spectrum is constructed with MC and data-driven techniques. Here we explain the uncertainties that affect the SM expectation, which can be divided into two main categories: uncertainties that affect reconstructed objects (such as jets) and uncertainties that affect the modelling of certain background or signal processes. Some of the uncertainties affect both

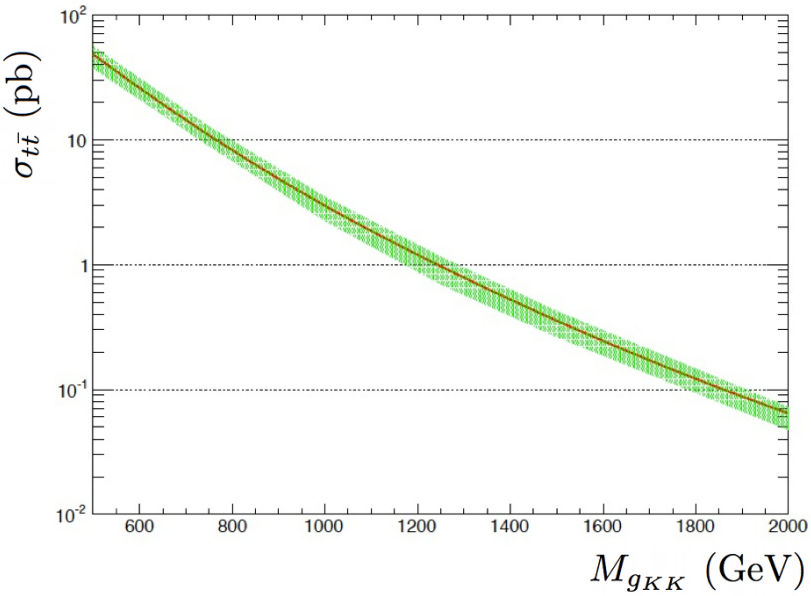


Figure 7.1: The $pp \to g_{KK} \to t\bar{t}$ cross-section returned by MadGraph (LO) versus the invariant mass of the g_{KK}. The results are obtained with the CTEQ6L1 PDF. In order to use these cross-sections in NLO calculations, a $K^{NLO}_{factor} = 1.3$ must be applied.

the shape and the normalization of the $m^{reco}_{t\bar{t}}$ spectra, while others affect the normalization only. Table 7.1 gives an overview of the effects of the dominant systematic uncertainties on the background and signal yields. Only the impact on the overall normalization is shown in the table, but some of the systematic uncertainties have a significant effect on the reconstructed $t\bar{t}$ mass, which is fully taken into account in the analysis.

One of the dominant uncertainties affecting reconstructed objects is the Jet Energy Scale (JES), especially for large-R jets. Uncertainties on the Jet Mass Scale (JMS) and the k_t splitting scales are also important for this analysis, and have an impact of 10% on the overall background yield in the boosted selection. The impact is smaller for the resolved selection since the large-R jets are only used indirectly (via the vetoing of events that satisfy the boosted selection).

Systematic Uncertainties	Resolved selection yield impact [%] on the tot. bckg.	Boosted selection yield impact [%] on the tot. bckg.
Luminosity	2.5	2.6
PDF	2.4	4.7
ISR/FSR	3.7	1.2
Parton Shower and fragmentation	4.8	1.5
$t\bar{t}$ normalization	5.3	5.5
$t\bar{t}$ EW virtual correction	0.2	0.5
$t\bar{t}$ generator	0.3	2.6
$t\bar{t}$ top quark mass	0.6	1.4
W+jets generator	0.3	0.1
Multi-jet normalization, e+jets	0.5	0.2
Multi-jet normalization, μ+jets	0.1	<0.1
JES+JMS, large-R jets	0.1	9.7
JER+JMR, large-R jets	<0.1	1.0
JES, small-R jets	5.6	0.4
JER, small-R jets	1.8	<0.1
Jet Vertex Fraction	0.8	0.2
b-tagging b-jet efficiency	1.1	2.9
b-tagging c-jet efficiency	0.1	0.1
b-tagging light-jet efficiency	<0.1	0.5
Electron efficiency	0.3	0.6
Muon efficiency	0.9	1.0
MC statistical uncertainty	0.4	1.3
ALL systematic uncertainties	10.8	13.4

Table 7.1: Average impact of the dominant systematic uncertainties on the total background yield. The electron and muon channel spectra are added. The shift is given in per cent of the nominal value.

7.4 Comparison of data to the SM template

After all event selection criteria are applied, 223330 data events pass the resolved selection and 8206 the boosted. This means 27 times as many events pass the resolved selection as the boosted, which reflects the falling mass spectrum of the top quark pairs. The event yields from data and from expected background processes are listed in Tables 7.2 and 7.3 together with the associated systematic uncertainties.

	Resolved-topology selection		
	e+jets	μ+jets	SUM
$t\bar{t}$	93000 ± 11000	91000 ± 11000	184000 ± 22000
Single Top	3800 ± 500	3800 ± 500	7600 ± 1000
$t\bar{t}V$	274 ± 40	267 ± 40	541 ± 80
Multi-jet e	5300 ± 1100	–	5300 ± 1100
Multi-jet μ	–	1050 ± 240	1050 ± 240
W+jets	6600 ± 800	7100 ± 800	13700 ± 1500
Z+jets	1400 ± 750	650 ± 340	2000 ± 1080
Dibosons	320 ± 120	310 ± 120	620 ± 240
Prediction	110000 ± 12000	105000 ± 12000	215000 ± 24000
Data	114377	108953	223330

Table 7.2: Data and expected background event yields after the resolved selections. The errors are the sum in quadrature of all systematic uncertainties.

	Boosted-topology selection		
	e+jets	μ+jets	SUM
$t\bar{t}$	4100 ± 600	4000 ± 600	8100 ± 1200
Single Top	138 ± 20	154 ± 20	290 ± 40
$t\bar{t}V$	37 ± 6	38 ± 7	75 ± 13
Multi-jet e	91 ± 18	–	91 ± 18
Multi-jet μ	–	8.6 ± 1.6	8.6 ± 1.6
W+jets	260 ± 50	290 ± 50	550 ± 100
Z+jets	31 ± 16	17 ± 9	48 ± 25
Dibosons	21 ± 8	20 ± 8	41 ± 16
Prediction	4700 ± 600	4500 ± 600	9200 ± 1200
Data	4148	4058	8206

Table 7.3: Data and expected background event yields after the boosted selections. The errors are the sum in quadrature of all systematic uncertainties.

The ATLAS muon trigger system suffers from a 20% inefficiency, relative to the offline event selection used in this analysis, largely due to a lack of geometrical coverage of muon chambers owing to support structures in those regions [165]. To mitigate this loss of efficiency, a large-R jet ($R = 1.0$) trigger was also used to collect muon plus jets events that failed the muon trigger. This

fat jet trigger has recorded 17.4 fb^{-1} data. The chosen trigger threshold yields a uniform efficiency as a function of reconstructed jet transverse momentum, above 99%, for events containing a large-R jet with $p_T > 380$ GeV. For $t\bar{t}$ events with invariant masses above 1.5 TeV, this addition increased the overall trigger efficiency in the muon channel to 96%. Events that pass this fat jet trigger are considered together with those that pass the muon trigger.

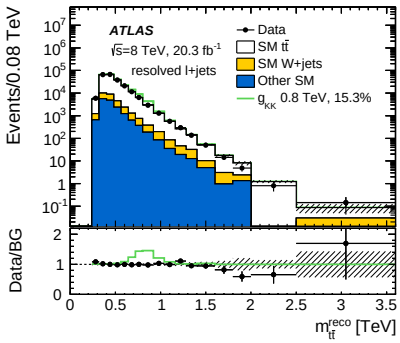
Good agreement is observed between the data and the total expected background. In the boosted-topology selection, the yield is less than the expectation. This is known in several analyses and it is explained by the NNLO calculation [134, 135, 136, 137, 138, 139, 166]. The NLO $m_{t\bar{t}}$ distribution is harder than the NNLO one. If the NNLO $m_{t\bar{t}}$ correction was included, the agreement would improve because the total expectation would decrease.

From now on, three b-tagging categories (one b-tagged jet in the leptonic side, one b-tagged jet in the hadronic side and two b-tagged jets) are taken into account in every channel (e+jets and μ+jets) and in every selection (resolved and boosted). If nothing is said, the sum of the three channels is used.

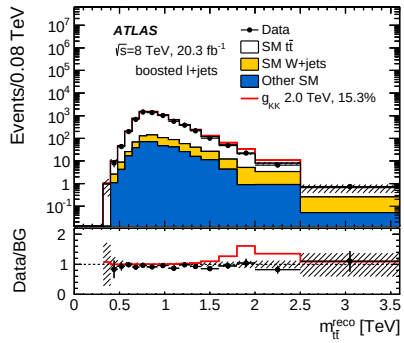
7.5 Results

The observed $m_{t\bar{t}}$ distribution is compared to the SM expectation in Figure 7.2. As explained in Section 7.3, the uncertainty in the background template is indicated with a shaded area. It is evaluated by considering multiple sources of instrumental uncertainties, of which the jet energy scale uncertainty is the most important. Modelling uncertainties are also taken into account for the main backgrounds. The total systematic uncertainties on the yield are 11% in the resolved topology and 13% in the boosted analysis.

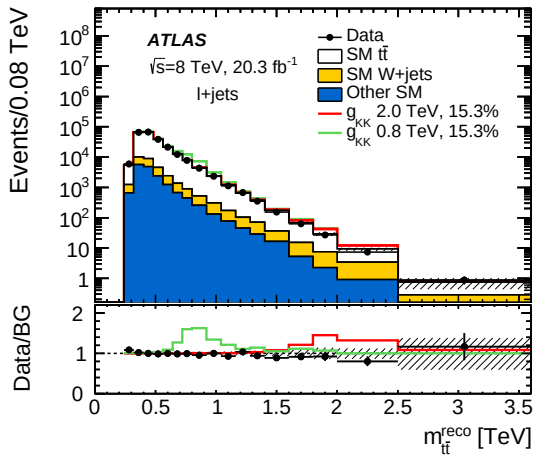
The search is performed on three combinations of the spectra: the six channels of the resolved selections, the six channels of the boosted selections and the twelve channels. In both topologies, all six spectra are fitted simultaneously. After accounting for the systematic uncertainties, no significant deviation from the total expected background is found.



(a) Resolved selection.



(b) Boosted selection.



(c) Both selections.

Figure 7.2: The $m_{t\bar{t}}^{reco}$ distributions, after the nuisance-parameter fit under the background-only hypothesis, summed over (a) all 6 boosted channels, (b) all 6 resolved channels, and (c) all 12 channels compared with data. The SM background components are shown as stacked histograms. The shaded areas indicate the total systematic uncertainties. The red (green) line shows the expected distribution for a hypothetical g_{KK} of mass 2.0 (0.8) TeV, width 15.3%.

7.6 Limits on benchmark models

The observed and expected limits on the production cross-section times $t\bar{t}$ BR are derived for a g_{KK} with nominal 15.3% width, shown in Figure 7.3, and on a g_{KK} with variable width, shown in Figure 7.4. These limits are computed with the **profile likelihood approach** [167] and the **CLs method** [168] for the 95% CL. This method receives as input three parameters: the data $m_{t\bar{t}}$ distribution, the MC $m_{t\bar{t}}$ expectation and the systematic uncertainties. This fit allows every systematic to vary their central value within 1σ. The maximisation of the likelihood can change the central values of the nuisance parameters and their associated uncertainties. The outputs of the method are two: all the nuisance parameters and the signal height, which can be translated in an upper limit in the production cross-section.

The upper cross-section limit is generated using expected data, which is the nominal background estimate built from the nuisance parameters fitted to real data. It excludes a g_{KK} of width 15.3% for masses less than 2.2 TeV, compared to an expected limit of 2.3 TeV using nominal background estimates.

The width dependence of the cross-section limits is also evaluated for the g_{KK} models. The result is shown in Figure 7.4. For a 1 TeV resonance, the limits weaken by approximately a factor of two as the width increases from 10% to 40%. The effect is stronger for 2 TeV and 3 TeV resonances, where the limits weaken by a factor of three over this width range.

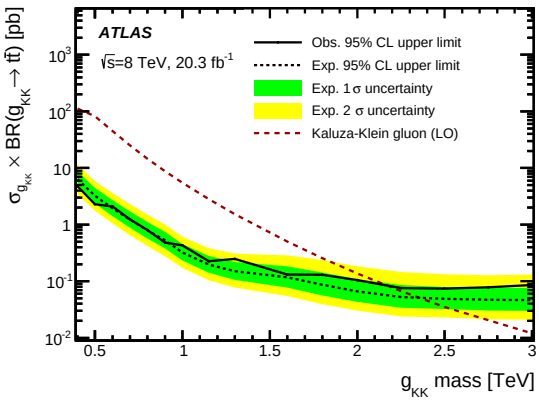


Figure 7.3: Observed and expected upper limits on the production cross-section times BR to $t\bar{t}$ final states as a function of the mass of a Bulk RS Kaluza-Klein gluon. The expected limits are derived from nominal (pre-fit) background estimates. The theoretical predictions for the $\sigma_{product} \times BR$ at the corresponding masses are also shown. Limits on the production cross-sections vary from 4.8 pb for a mass of 0.4 TeV, to 0.09 pb for a mass of 3 TeV.

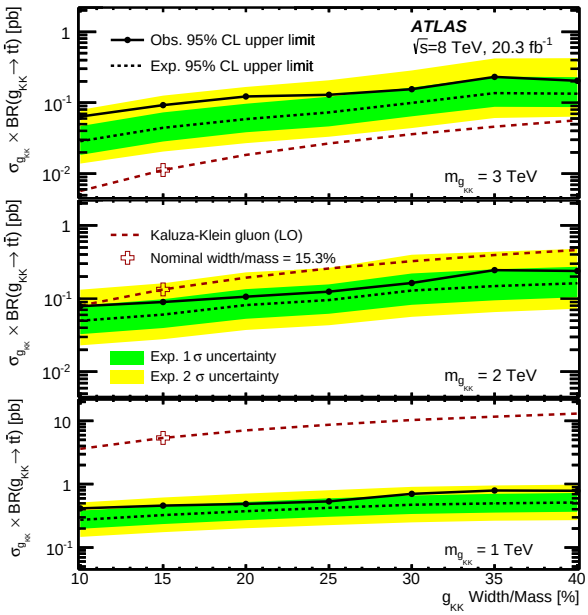


Figure 7.4: Observed and expected upper limits on the production cross-section times BR to $t\bar{t}$ final states as a function of the width of a Bulk RS Kaluza-Klein gluon with three representative mass values. The expected limits are derived from nominal (pre-fit) background estimates. The theoretical predictions for the $\sigma_{product} \times BR$ at the corresponding masses are also shown.

Reference [151] also presents limits on additional benchmark models. A narrow leptophobic topcolour Z' boson (Z'_{TC2} of width 1.2% [28]) with $m < 1.8\ TeV$ is excluded, while $m < 2.0\ TeV$ are expected to be excluded. No mass range is excluded in the Bulk Randall-Sundrum Kaluza-Klein graviton G_{KK} [162].

7.7 Summary

A search for production of new heavy particles decaying to $t\bar{t}$ in the ℓ+jets decay channel was carried out with the ATLAS experiment at the LHC. The search uses data corresponding to an integrated luminosity of 20.3 fb^{-1} of proton-proton collisions at a centre-of-mass energy of 8 TeV. No excess of events beyond the SM predictions is observed in the $t\bar{t}$ invariant mass spectra.

Upper limits on the $\sigma \times BR$ are set for a broad (15.3% width) Randall-Sundrum Kaluza-Klein gluon. Based on these results, the existence of a broad Kaluza-Klein gluon with mass $0.4 < m_{g_{KK}} < 2.2\ TeV$ is excluded at 95% CL, while masses below 2.3 TeV are expected to be excluded. These results probe new physics at higher mass than previous ATLAS searches for the same signature, and the results are applicable to a broader variety of heavy resonances.

8

Measurement of the $t\bar{t}$ charge asymmetry

In this chapter a measurement of the $t\bar{t}$ charge asymmetry in 8 TeV pp collisions is presented [45]. The ATLAS collaboration has been the first to explore the A_C in the regime of highly boosted $t\bar{t}$ production.

8.1 Introduction

The millions of top quark pairs produced at the LHC allow for a measurement of the A_C with unprecedented statistical precision compared to other colliders. With its much larger centre-of-mass energy than any previous collider, the LHC moreover probes a new kinematic regime, even in the first phase of operation. For the first time, a large sample of highly boosted top quarks [1] is available for study [169]. A measurement of the $t\bar{t}$ A_C in the production of boosted top quark pairs is a stringent test of QCD in a regime not explored by previous experiments. Additionally, the measurement is very sensitive to the effect of new massive states.

In this thesis the results are reported of the first measurement of the A_C in top quark pair production based on novel techniques specifically designed to deal with the collimated decay topology of boosted top quarks. The analysis focuses on the lepton+jets (ℓ+jets) final state, where the hadronic top quark decay is reconstructed as a single large-radius (large-R) jet and tagged as such on the basis of a jet substructure analysis. The leptonic top quark decay is reconstructed from a single small-radius (small-R) jet, a single charged lepton (muon or electron[2]), and a neutrino.

[1] Top quarks are referred to as boosted when their transverse momentum exceeds the $t\bar{t}$ rest mass, i.e. when the mass of the $t\bar{t}$ system exceeds approximately 800 GeV, and the top quark decay is collimated in an area of the detector with $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} = 1.0$.

[2] τ-lepton decays to a muon or electron yield an isolated μ or e in the final state and are considered part of the signal.

Event selection and reconstruction are based on the algorithm developed in Reference [170] and have previously been used in a number of $t\bar{t}$ resonance searches [151, 171, 161, 172] and, recently, in a differential cross-section measurement [152]. Compared to previous analyses based on more classical top quark selection and reconstruction schemes, this approach offers an increased acceptance for high mass pairs and more precise reconstruction of the $t\bar{t}$ invariant mass and top quark direction. It is therefore possible to extend the kinematic reach well into the TeV regime.

Data and MC samples, Object definitions, Event selection, Reconstruction of the $t\bar{t}$ pair and Background estimation are explained in detail in Chapter 6.

8.2 Data and MC comparison at detector-level

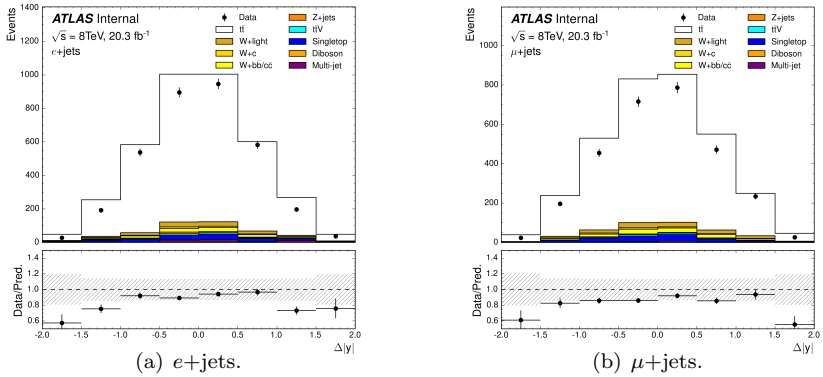
The event yield on the $\sqrt{s} = 8$ GeV data set of the selection in the e+jets and μ+jets data streams is summarized in Table 8.1. The number of events agrees to within a few per mil with the analysis in Reference [152], that is based on a similar, but not exactly identical, selection. Once minor differences in approach are accounted for (e.g., the $t\bar{t}V$ background is not included in Reference [152]), the agreement improves to the per mil level. The results are also in good agreement with those obtained by the 8 TeV $t\bar{t}$ resonance search [151] once the muon stream acceptance is complemented by accepting events that triggered the online fat jet selection in the delayed stream.

The result is compared to an estimate of the contributions of the relevant SM processes. The predicted yield for most processes relies on MC simulation. The W+jets background is normalized using the **charge asymmetry method** [156, 157], explained in Section 6.5.1. The QCD multi-jet contribution is entirely data-driven [158], as explained in Section 6.5.2. The agreement between the predicted signal and background yields and the data is assessed. Table 8.1 gives the number of expected events for each process, where the systematic uncertainties on the background estimates, objects energy scale and reconstruction efficiencies, and MC statistics are taken into account. The total number of observed events is also presented. The prediction is generally found to overestimate the data by approximately one standard deviation, as expected from the $t\bar{t}$ resonance search [151] and differential cross-section measurement [152] using the same algorithm and data set.

	e+jets	μ+jets	ℓ+jets
$t\bar{t}$	4100 ± 600	3600 ± 500	7700 ± 1100
W+jets	263 ± 32	264 ± 32	527 ± 63
Single Top	140 ± 20	138 ± 19	278 ± 39
$t\bar{t}V$	37 ± 19	33 ± 17	70 ± 35
Z+jets	40 ± 27	16 ± 11	56 ± 38
Multi-jet	44 ± 8	4 ± 1	48 ± 9
Dibosons	20 ± 7	18 ± 7	38 ± 14
Prediction	4600 ± 600	4000 ± 500	8600 ± 1100
Data	4145	3600	7730

Table 8.1: Observed and expected number of events in the signal samples. The systematic uncertainties on the SM expectation include those from detector-related uncertainties, uncertainties on the normalization, the luminosity uncertainty and the uncertainty on the cross-section predicted used to normalize the expected yield. More details in Section 8.4.

The distribution for the $\Delta|y| = |y_t| - |y_{\bar{t}}|$ observed in data is shown in Figure 8.1. These distributions are slightly more asymmetric than the SM



(a) e+jets. (b) μ+jets.

Figure 8.1: Detector-level distributions for the difference of absolute rapidities $\Delta|y|$ of top and anti-top quark. The ratio of data and MC prediction is shown in the insets below the plots. Error bars on the data points indicate the statistical uncertainty. The hashed area, only shown in the ratio portion of each plot, shows the uncertainty of the SM prediction. This includes the statistical uncertainty, the theory uncertainties on the cross-sections and the effect of detector systematics on the expected yield.

expectation in both channels. The observed detector-level A_C is 3.7% ± 1.3% (for the ℓ+jets channel) compared to 1.4% ± 1.1%(*stat.*) for the SM prediction.

The A_C is largely driven by $t\bar{t}$ production (nominal MC predicts a detector-level asymmetry of 1.5% for $t\bar{t}$ production). Non-$t\bar{t}$ background processes contribute approximately 11% of the total expected yield. These sources have a very small expected A_C of 0.32%.

8.3 Fully Bayesian Unfolding (FBU)

The unfolding is a procedure that transforms the detector-level result to a parton-level result that can be compared to theoretical predictions. The unfolding technique corrects the observed $\Delta|y|$ distribution for resolution and acceptance effects by inversion of the response matrix, which is constructed with MC events. For this analysis the Fully Bayesian Unfolding [173] package (PyFBU) is used. Uncertainties due to limitations in the understanding of object reconstruction and in the calibration of the experiment are included as nuisance parameters in the unfolding procedure, as well as the normalization of the backgrounds. The result of the unfolding procedure is a posterior distribution for the parton-level A_C and for all nuisance parameters.

After the reconstruction of the top quark and anti-top quark candidates, discussed in Section 6.4.2, events where the pair has an invariant mass of less than 750 GeV are discarded. The distribution is unfolded to the parton-level of top quarks before decay, but after radiation. The event selection limits the measurement to high mass top quark pairs in the central detector. Correcting back to the fully inclusive parton-level would imply a very large extrapolation, and consequently a large modelling uncertainty. We therefore limit the fiducial region for the parton-level top quarks to the measured phase space[3]: $m_{t\bar{t}} > 750\ GeV,\ |\Delta|y|| < 2$. From now on, this will be the so-called *inclusive* region.

The results of e+jets and μ+jets channels are merged before unfolding. The data from both channels are added to form the combined $\Delta|y|$ distribution. Also the background templates and the response matrices used to correct the measurement are simply added up. We have shown in Section 6.4.2 (in particular in Figure 6.2) that the response for both channels is quite similar. The relative weight of both channels in the combination is correctly taken into account in this procedure. More details are found in Appendix C.

[3] A more detailed fiducial region, that includes the exact requirements placed on the charged lepton, neutrino and jets, has the potential to further decrease the modelling uncertainty. However, that approach does not allow a direct comparison to fixed-order calculations.

8.4 Systematic uncertainties

The impact of each of the sources of systematic uncertainty on the A_C measurement is presented in Table 8.2. The largest source of uncertainty is the modelling of the $t\bar{t}$ signal, where alternative choices of matrix element, parton shower, top quark mass and parton density functions may lead to nearly 2% change in the result. The uncertainties on the normalization of non-$t\bar{t}$ SM processes contribute only at the per mil level. Uncertainties in modelling the response of the ATLAS detector are small: 0.4% for the inclusive measurement, dominated by the uncertainty in the fat jet energy and mass response. The unfolding systematic, that includes the effect of limited MC statistics (evaluated by smearing the response matrices) and the bias and non-linearity in the unfolding procedure, is 0.3% for the inclusive measurement.

source of uncertainty	δA_C (%)
signal modelling - PS	$\pm$ 2.0
singal modelling - ME	$\pm$ 1.5
signal modelling - ISR/FSR	$\pm$ 0.1
signal modelling - m_t	$\pm$ 0.1
signal modelling - PDF	$\pm$ 0.4
signal modelling - total	$\pm$ 2.5
background norm.	$\pm$ 0.10
jet energy and resolution - $R = 0.4$ jets	$\pm$ 0.11
jet energy and mass scale - $R = 1.0$ jets	$\pm$ 0.32
b-tag/mis-tag efficiency	$\pm$ 0.18
lepton reco/id/scale	$\pm$ 0.09
missing transverse energy (E_T^{miss})	$\pm$ 0.05
unfolding & MC stat.	$\pm$ 0.56

Table 8.2: The effect on the A_C^{corr} of systematic uncertainties in the signal and background modelling and the description of the detector response. The $\pm 1\sigma$ uncertainties are to be added to the result, i.e., they are not relative errors. More details in Appendix D.

We also present a differential measurement in three bins of $m_{t\bar{t}}$. In Table 8.3 the same breakdown of the systematic uncertainties is presented for the three $m_{t\bar{t}}$ bins. The largest sources of uncertainty in each bin are the modelling of the $t\bar{t}$ signal. The same pattern is observed here. Total uncertainties due to the $t\bar{t}$ signal modelling are at the level of 4.0, 1.4 and 9.3% for the three respective bins in $m_{t\bar{t}}$ while the combined uncertainties modelling the ATLAS detector response are 2.2, 0.6, and 1.5%.

source of uncertainty	δA_C (%) in three $m_{t\bar{t}}$ (TeV) intervals [0.75-0.9]	[0.9-1.3]	[1.3-4.0]
signal modelling - PS	3.2	1.2	6.2
signal modelling - ME	2.4	0.6	5.3
signal modelling - ISR/FSR	0.3	0.1	3.0
signal modelling - PDF	0.4	0.3	3.3
signal modelling - total	4.0	1.4	9.3
background norm.	± 0.16	± 0.25	± 0.37
jet energy and resolution - $R = 0.4$ jets	± 0.45	± 0.27	± 0.43
jet energy and resolution - $R = 1.0$ jets	± 1.61	± 0.59	± 1.04
b-tag/mis-tag efficiency	± 0.23	± 0.15	± 0.67
lepton reco/id/scale	± 0.17	0.10	0.14
E_T^{miss}	± 0.03	± 0.03	± 0.09
unfolding & MC stat.	1.2	0.8	2.1

Table 8.3: The effect on the corrected A_C in the three $m_{t\bar{t}}$ bins of systematic uncertainties in the signal and background modelling and the description of the detector response. The $\pm 1\sigma$ uncertainties are to be added to the result (i.e. they are not relative errors). First bin is $0.75 < m_{t\bar{t}} < 0.9$, second bin is $0.9 < m < 1.3$ and third bin is $1.3 < m < 4.0$, in TeV.

8.5 Results

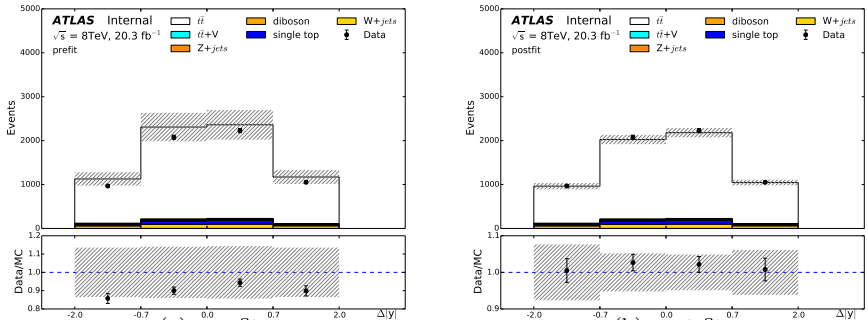
The $\Delta|y|$ distributions before and after unfolding are shown in Figure 8.2. FBU adapts the normalization of the signal and background sources. The $t\bar{t}$ normalization is left floating in the fit, with a flat prior to avoid any bias. The discrepancy between the observed yield and the MC expectation is absorbed in the $t\bar{t}$ normalization: the expected $t\bar{t}$ yield is scaled by 0.86. The uncertainty on the $t\bar{t}$ normalization is ± 0.023 after the fit, leading to a much reduced error band in the rightmost panel of Figure 8.2.

The limited statistics in four $\Delta|y|$ are insufficient to constrain any of the detector systematics (as shown in Figure C.3 in Appendix C).

The result for the A_C on the parton-level fiducial phase space ($0.75\ TeV < m_{t\bar{t}} < 4\ TeV$, $|\Delta|y|| < 2$) is the following:

$$\begin{aligned} A_C &= 4.2\% \pm 2.0\%(stat. + exp.syst.) \pm 2.5\%(mod.) \pm 0.3\%(unf.) \\ &= 4.2\% \pm 3.2\% \end{aligned}$$

where the systematic uncertainties are broken down in signal modelling, and unfolding systematics. Experimental systematics plus the statistical uncertainty are estimated as the RMS of the posterior distribution of the unfolding procedure.



(a) *pre-fit.* (b) *post-fit.*

Figure 8.2: Combined $\Delta|y|$ distributions in data (a) before (*pre-fit*) and (b) after (*post-fit*) profiling the uncertainties, compared to the SM template.

The result without systematics is in good agreement with the variation of the result in pseudo-experiments where data are varied within their statistical uncertainty.

As a cross-check we unfold the $\Delta|y|$ distributions of the e+jets and μ+jets channels separately. The posterior distributions for both channels are shown in Figure 8.3. Superimposed is the combined result. The results for the A_C are $5.2 \pm 2.6\%$ (stat.) for the e+jets channel and $3.5 \pm 2.7\%$ (stat.) for the μ+jets channel. Both results are compatible within the statistical uncertainty.

The differential measurement, correcting for migrations in $\Delta|y|$ and $m_{t\bar{t}}$, yields the following results:

$$\begin{aligned} A_C &= 2.2\% \pm 6.0\%(stat. + exp.syst.) \pm 4.0\%(mod.) \pm 1.2\%(unf.), \\ &= 2.2\% \pm 7.3\%(stat. + syst.), \\ &\text{for the mass interval } (0.75 \text{ TeV} < \mathrm{m_{t\bar{t}}} < 0.90 \text{ TeV}) \end{aligned}$$

$$\begin{aligned} A_C &= 8.6\% \pm 4.1\%(stat. + exp.syst.) \pm 1.4\%(mod.) \pm 0.8\%(unf.), \\ &= 8.6\% \pm 4.4\%(stat. + syst.), \\ &\text{for the mass interval } (0.9 \text{ TeV} < \mathrm{m_{t\bar{t}}} < 1.3 \text{ TeV}) \end{aligned}$$

$$\begin{aligned} A_C &= -2.9\% \pm 11.6\%(stat. + exp.syst.) \pm 9.3\%(mod.) \pm 2.1\%(unf.), \\ &= -2.9\% \pm 15.0\%(stat. + syst.), \\ &\text{for the mass interval } (1.3 \text{ TeV} < \mathrm{m_{t\bar{t}}} < 4.0 \text{ TeV}) \end{aligned}$$

In all cases the fiducial phase space region is limited to $|\Delta|y|| < 2$. The results are summarized in Figure 8.4.

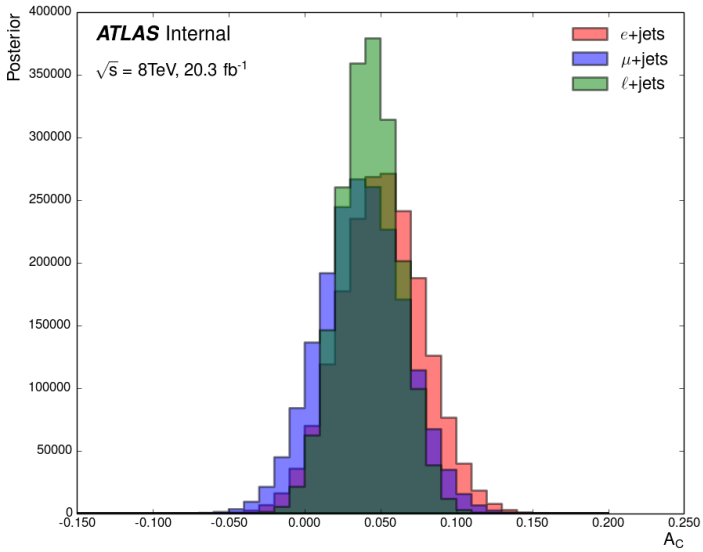


Figure 8.3: The posterior distribution for the inclusive A_C of the e+jets and μ+jets channels, and of the combined result. Only statistical uncertainties are included as most of the systematics are correlated between the two channels.

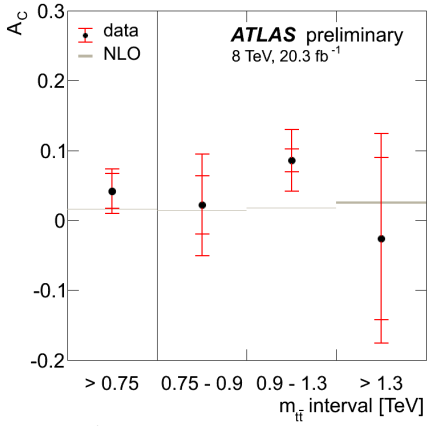
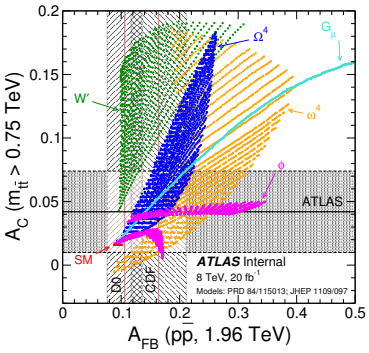


Figure 8.4: Summary of the A_C measurements presented in this note. The asymmetry observed in boosted ℓ+jets events is corrected to the phase space $|\Delta|y|| < 2$ and several $t\bar{t}$ invariant mass intervals, as indicated on the X-axis. The error bars on the data indicate the modelling and unfolding systematics, indicated by the smaller bar, and the total uncertainty, that includes the statistical uncertainty and the experimental systematics.

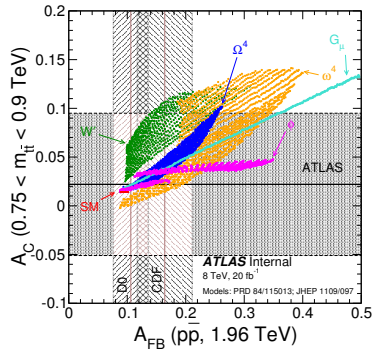
8.6 Impact on BSM scenarios

The A_C measurement at high mass is motivated in part by its ability to confirm or discard extensions of the SM. The strong dilution by gluon-initiated processes at the LHC could wash out the signature of new physics in the inclusive A_C. However, models that predict a large deviation of the A_C measured at the Tevatron typically also predict a significantly enhanced A_C at high mass at the LHC. In Figure 8.5 BSM predictions for the A_C in 8 TeV pp collisions with $m_{t\bar{t}} > 750$ GeV are correlated with the inclusive forward-backward asymmetry in pp collisions at $\sqrt{s} = 1.96$ TeV. The current measurement is indicated as a horizontal line and its error as a band. The inclusive measurements at the Tevatron, $A_{FB} = 0.164 \pm 0.045$ by CDF [37] and $A_{FB} = 0.196 \pm 0.065$ by D0 [174], are indicated as vertical bands.

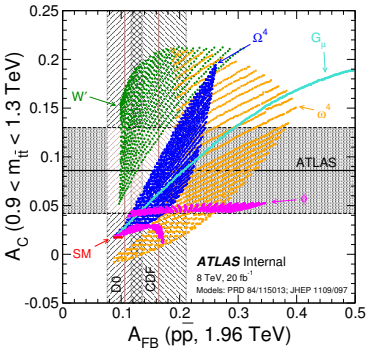
The clouds of points in Figure 8.5 correspond to a number of models in References [175, 176]: a heavy W' boson exchanged in the t-channel, a heavy axigluon (G_μ) exchanged in the s-channel and doublet (ϕ), triplet (ω^4) or sextet (Ω^4) scalars. Each point corresponds to a choice of the new particle mass, in the range between 100 GeV and 10 TeV, and of the couplings, where all values are allowed that yield a total cross-section for top quark pair production at the Tevatron compatible with observations and a high mass $t\bar{t}$ production ($m_{t\bar{t}} > 1$ TeV) at the LHC that is at most three times the SM prediction. The contribution of the new physics to the Tevatron A_{FB} is moreover required to be positive (this last requirement causes the rather sharp vertical edge that is visible in the W' model). The predictions for the Tevatron forward-backward asymmetry and the LHC high mass A_C are calculated using Protos [177], that includes the tree-level SM amplitude and the one(s) from the new particle(s), taking into account the interference between the two contributions. An interpretation of this measurement in terms of an effective field theory is given in Reference [178].



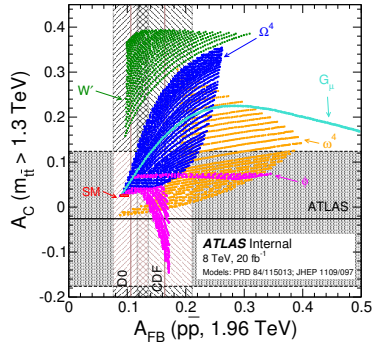
(a) $m_{t\bar{t}} > 0.75\ TeV$



(b) $0.75 < m_{t\bar{t}} < 0.9\ TeV$



(c) $0.9 < m_{t\bar{t}} < 1.3\ TeV$



(d) $m_{t\bar{t}} > 1.3\ TeV$

Figure 8.5: The predictions of a number of extensions of the SM (explained in the text) for the inclusive A_C at the Tevatron and the four high mass asymmetry measurements reported in this note. (a) $m_{t\bar{t}} > 0.75\ TeV$. (c) $0.75 < m_{t\bar{t}} < 0.9\ TeV$. (b) $0.9 < m_{t\bar{t}} < 1.3\ TeV$. (d) $m_{t\bar{t}} > 1.3\ TeV$.

8.7 Conclusions

The charge asymmetry in $t\bar{t}$ production in 8 TeV pp collisions is measured for events with highly boosted top quark pairs. The dataset corresponds to an integrated luminosity of 20.3 fb^{-1} collected by the ATLAS detector in 2012.

Events in the ℓ+jets final state are selected and reconstructed with an algorithm [151, 171, 161, 172, 170] developed specifically for the decay topology of highly boosted top quark pairs. The leptonic top decay is reconstructed, as an electron or muon candidate, a neutrino reconstructed using the W-boson mass constraint, and a nearby small-R jet. The hadronically-decaying top quarks are reconstructed as a large-R jets and tagged using jet substructure techniques. This selection yields good acceptance for events where the top quark pair is produced with large invariant mass ($m_{t\bar{t}} > 0.75\ TeV$) and in the central part of the detector ($|\Delta|y|| < 2$). The agreement between data and MC simulation for several key observables (top quark p_T, $m_{t\bar{t}}$, $\Delta|y|$) is improved compared to more traditional reconstruction scheme.

The observed $\Delta|y|$ distribution, the difference in absolute rapidities of top and anti-top quarks, is corrected to bring the measurement to the parton-level phase space: $m_{t\bar{t}} > 0.75\ TeV$, $|\Delta|y|| < 2$. A matrix unfolding procedure, which uses the FBU [173] algorithm, corrects for the effect of kinematic migrations in mass and rapidity due to reconstruction. The response matrices are based on generated POWHEG MC events with $h_{damp} = m_t$, including PYTHIA parton shower and hadronization, and a detailed model for the detector response. An acceptance correction is applied to remove the selection bias.

The A_C in the high mass, central fiducial region is found to be:

$$A_C = 4.2\% \pm 3.2(tot.) \qquad (0.75TeV < m_{t\bar{t}} < 4TeV, |\Delta|y|| < 2).$$

The uncertainty is dominated by the signal modelling uncertainty followed by the statistical uncertainty of the 8 TeV data set. Uncertainties on the detector response play a minor role.

We compare the measured A_C to the SM prediction. The calculation, that includes the NLO QCD and Electroweak corrections of References [40, 179, 180], predicts a A_C in the same fiducial region of 1.60(4)%[4]. The central value of the measurement is slightly higher than the SM prediction, but the measurement

[4] G. Rodrigo, IFIC (UVEG/CSIC) Valencia, private communication. The SM parameters are the following: $m_t = 173.34\ GeV$ [181], $\alpha_s(m_Z) = 0.11856$ [182]. The proton PDFs used in the calculation are from MMHT2014 [183].

is compatible within one standard deviation. NNLO corrections may further improve the agreement. The MC generators used in ATLAS predict a A_C between 0.6%±0.2%(*stat.*) for the nominal Powheg+Pythia with $h_{damp} = m_t$ and 1.0% ± 0.1% for MC@NLO+Herwig.

	[0.75-0.9]	[0.9-1.3]	[1.3-4.0]
measurement	2.2 ± 7.3 %	8.6 ± 4.4 %	-2.9% ± 15.0 %
SM prediction	1.42 ± 0.04 %	1.75 ± 0.05 %	2.55 ± 0.18 %
Powheg+Pythia	0.5 ± 0.2%	0.8 ± 0.3%	0.1 ± 0.7%
MC@NLO+Herwig	1.1 ± 0.2%	0.9 ± 0.2%	0.5 ± 0.6%

Table 8.4: The measured A_C after the correction to parton-level in three intervals of $m_{t\bar{t}}$, which is the invariant mass of the $t\bar{t}$ system. First bin is $0.75 < m_{t\bar{t}} < 0.9\ TeV$, second bin is $0.9 < m_{t\bar{t}} < 1.3\ TeV$ and third bin $m_{t\bar{t}} > 1.3\ TeV$. The SM prediction from a fixed-order calculation and two event generators are also indicated. The uncertainties correspond to the sum of statistical and systematic uncertainties for the measurement, to the theory uncertainty for the SM prediction and to the statistical uncertainty for MC estimates. In all cases the phase space is limited to $|\Delta|y|| < 2$.

Results of a differential A_C measurement in three $m_{t\bar{t}}$ intervals are reported in Table 8.4. The bin from 0.9 to 1.3 TeV, where the A_C has been measured with a precision of 4.3%, extends the $t\bar{t}$ mass reach of previous analyses into the TeV regime. The SM prediction is also given, as well the expectations in two MC generators. The largest deviation is observed in the intermediate mass bin, where the difference between the measured and predicted A_C corresponds to 1.7 standard deviations.

This measurement agrees with the SM and other measurements at somewhat lower mass [184, 185]. They provide a stringent constraint of the parameter space of several extensions of the SM.

9

Summary

The Large Hadron Collider (LHC) at CERN is the largest collider ever built. It is assembled in the same 27 km tunnel where Large Electron-Positron Collider (LEP) used to be, ~ 100 m underground, near Geneva (Switzerland). The main interaction points have four experiments installed: ATLAS, CMS, ALICE and LHCb. The LHC is designed to collide two proton beams, that circulate in opposite directions, each with an energy of up to 7 TeV, reaching a centre-of-mass energy up to 14 TeV.

The ATLAS experiment is a general-purpose detector designed to exploit the potential of the LHC. With a total weight of 7000 tonnes, ATLAS is the heaviest experiment at the LHC. It has 42 m of total length and 11 m of radius. It is hosted under the ground of the Interaction Point 1 of the LHC. Its main components are four: the Inner Detector (ID), the Calorimeters, the Magnet system and the Muon Spectrometer.

The first LHC running period is emphasized in this thesis because the analyses presented in this thesis are based on Run-I data. During this period, IFIC was involved in the Spanish Federated Tier-2 (ES-ATLAS-T2) and consequently in the Iberian ATLAS Cloud, providing computing resources to the ATLAS Collaboration.

The Iberian ATLAS Cloud has responded very efficiently during the Run-I. The changes in the Computing Model allowed to improve the performance of the cloud in terms of connectivity, storage, replication, transfer, etc. All the sites of the cloud provided the resources of CPU, disk and tape needed to fulfil the ATLAS pledge. The required Distributed Analysis tools were provided to the users in order to use/store data and produce experimental results, i.e., the observation of a new particle in the search for the Standard Model Higgs boson.

The LHC underwent a period of maintenance and planned upgrades, the First Long Shutdown, to prepare the collider for a higher energy and luminosity, and all the implications that this entails. During this period, installations connected with the experiments were improved, accelerators were upgraded, electronics and computing evolved, etc. This period served to optimize the experiences learned during the Run-I and improve them during the Run-II. Several of the LS1 upgrades were tested for a long period of time in an exercise called Data Challenge (DC14), which claimed to get ATLAS ready for Run-II physics. During DC14 the Integrated Simulation Framework (ISF) was commissioned in the context of physics analyses, large-scale jobs of the updated reconstruction algorithms and the new distributed computing tools were run, and finally, the Run-II Analysis Model was tested, resulting in a gain of experience with the Run-II analysis framework.

The top quark is the most massive of the fundamental particles known so far. It was discovered in 1995 at the Tevatron, a $\sqrt{s} = 1.96\ TeV\ p\bar{p}$ collider built in Fermilab, by the CDF and D0 Collaborations. It is the only quark in the SM that decays before it hadronizes, the only fermion whose mass is close to the electroweak scale, and the only quark with large Yukawa coupling to the Higgs boson. Its properties are of great importance in many of the current BSM theories.

Apart from direct searches for new resonances in $t\bar{t}$ production, experiments look for signatures of new physics by performing precise measurements of top quark production and decay.

The Tevatron experiments reported a first measurement of the forward-backward asymmetry (A_{FB}) in top quark pair production that had considerable tension with respect to the SM prediction. At the LHC a related charge asymmetry (A_C) has been measured.

Popular BSM models address some of the current open questions, such as the baryon asymmetry of the universe, the nature of dark matter or a mechanism to naturally stabilize the Higgs boson mass at its observed value of approximately 125 GeV. For some of these solutions to work, the new physics scale has to be around the TeV, within reach of collider experiments.

The LHC is the first machine to directly probe these energies. In the most energetic collisions heavy particles (such as top quarks, W or Z bosons) are produced with a transverse momentum that greatly exceeds their rest mass. The Lorentz boost of these heavy particles changes the topology of their decay products. The classical method to identify and to reconstruct their hadronic decays is not adequate for these "boosted objects". In order to manage this

boosted topology, new techniques have been developed that reconstruct the boosted object as a single *jet* and distinguish it from the background through a substructure analysis.

In order to select and to reconstruct top quark pairs at the LHC, two methods for selecting ℓ+jets events are discussed: the classical, resolved approach, developed at the Tevatron for top quark production at rest, and a new technique designed specifically to deal with boosted top quark production. The latter has superior acceptance and resolution for top quark pairs with a mass above 1 TeV. A SM template is constructed using a combination of MC and data-driven techniques. Data has been used to estimate two important backgrounds in the boosted analysis: W+jets and multi-jet production. Finally, a detailed comparison of data and the MC template is presented for the boosted analysis. Apart from an understood discrepancy in the yield, a reasonable agreement is found.

A search for production of new heavy particles decaying to $t\bar{t}$ in the ℓ+jets decay channel was carried out with the ATLAS experiment at the LHC. The search uses data corresponding to an integrated luminosity of 20.3 fb^{-1} of proton-proton collisions at a centre-of-mass energy of 8 TeV. No excess of events beyond the SM predictions is observed in the $t\bar{t}$ invariant mass spectra.

Upper limits on the $\sigma \times BR$ are set for a broad (15.3% width) Randall-Sundrum Kaluza-Klein gluon. Based on these results, the existence of a broad Kaluza-Klein gluon with mass $0.4 < m_{g_{KK}} < 2.2$ TeV is excluded at 95% CL, while masses below 2.3 TeV are expected to be excluded. These results probe new physics at higher mass than previous ATLAS searches for the same signature, and the results are applicable to a broader variety of heavy resonances.

The charge asymmetry in $t\bar{t}$ production in 8 TeV pp collisions is measured for events with highly boosted top quark pairs. The dataset corresponds to an integrated luminosity of 20.3 fb^{-1} collected by the ATLAS detector in 2012.

The observed $\Delta|y|$ distribution, the difference in absolute rapidities of top and anti-top quarks, is corrected to bring the measurement to the parton-level phase space: $m_{t\bar{t}} > 0.75$ TeV, $|\Delta|y|| < 2$. A matrix unfolding procedure, which uses the FBU algorithm, corrects for the effect of kinematic migrations in mass and rapidity due to reconstruction. An acceptance correction is applied to remove the selection bias.

The A_C in the high mass, central fiducial region is found to be:

$$A_C = 4.2\% \pm 3.2(tot.) \qquad (0.75TeV < m_{t\bar{t}} < 4TeV, |\Delta|y|| < 2).$$

The uncertainty is dominated by the signal modelling uncertainty followed by the statistical uncertainty of the 8 TeV data set. Uncertainties on the detector response play a minor role.

Results of a differential A_C measurement in three $m_{t\bar{t}}$ intervals are reported in Table 9.1. The bin from 0.9 to 1.3 TeV, where the A_C has been measured

	[0.75-0.9]	[0.9-1.3]	[1.3-4.0]
measurement	2.2 ± 7.3 %	8.6 ± 4.4 %	-2.9% ± 15.0 %
SM prediction	1.42 ± 0.04 %	1.75 ± 0.05 %	2.55 ± 0.18 %
Powheg+Pythia	0.5 ± 0.2%	0.8 ± 0.3%	0.1 ± 0.7%
MC@NLO+Herwig	1.1 ± 0.2%	0.9 ± 0.2%	0.5 ± 0.6%

Table 9.1: The measured A_C after the correction to parton-level in three intervals of $m_{t\bar{t}}$, which is the invariant mass of the $t\bar{t}$ system. First bin is $0.75 < m_{t\bar{t}} < 0.9\ TeV$, second bin is $0.9 < m_{t\bar{t}} < 1.3\ TeV$ and third bin $m_{t\bar{t}} > 1.3\ TeV$. The SM prediction from a fixed-order calculation and two event generators are also indicated. The uncertainties correspond to the sum of statistical and systematic uncertainties for the measurement, to the theory uncertainty for the SM prediction and to the statistical uncertainty for MC estimates. In all cases the phase space is limited to $|\Delta|y|| < 2$.

with a precision of 4.3%, extends the $t\bar{t}$ mass reach of previous analyses into the TeV regime. The SM prediction is also given, as well the expectations in two MC generators. The largest deviation is observed in the intermediate mass bin, where the difference between the measured and predicted A_C corresponds to 1.7 standard deviations.

This measurement agrees with the SM and other measurements at somewhat lower mass. They provide a stringent constraint of the parameter space of several extensions of the SM.

Appendices

A

Glossary of terms

A.1 Definitions

↪ **Baryon:** It is a composite subatomic particle made up of three quarks ($q_1q_2q_3$). It is a fermion since it has a fractional spin. It belongs to the hadron family of particles. As quark-based particles, baryons participate in the strong interaction. The most familiar baryons are the protons (uud) and the neutrons (udd).

↪ **Barn:** It is a unit of area, with equivalence 1 $b = 10^{-28}\ m^2$. It is widely used in physics to measure cross-sections. Its inverse is used to measure luminosities. Some useful equivalence are the following:

$$1\ pb = 10^{-12}\ b = 10^{-40}\ m^2 = 10^{3}\ fb \quad \Longrightarrow \quad 1\ pb^{-1} = 10^{-3}\ fb^{-1}$$
$$1\ fb = 10^{-15}\ b = 10^{-43}\ m^2 = 10^{-3}\ pb \quad \Longrightarrow \quad 1\ fb^{-1} = 10^{3}\ pb^{-1}$$

↪ **$\mathcal{CP}$ symmetry:** $\mathcal{CP}$ is the acronym for Charge Conjugation, $\mathcal{C}$, and Parity, $\mathcal{P}$. Charge Conjugation is an operation that changes a matter particle to its corresponding antiparticle. Parity creates a mirror image of a particle or system, reversing left and right. Both, Charge Conjugation and Parity must be flipped to change matter to antimatter with the correct particle helicity (the term that indicates if a particle is left-handed or right-handed).

↪ **Data driven:** Data driven techniques [151, 158] are those that evaluate the background contributions directly from acquired data, taking into account only proper control regions.

↪ **First LHC running period (Run-I):** It is the first operating period of the LHC, which started on March 2010 to January 2013.

↪ **First Long Shut Down (LS1):** It is the first period in which the LHC was stopped to be updated and upgraded. This period was from 14 February 2013 to January 2015.

↪ **JES:** The ATLAS Jet Energy Scale [186]. It is a measurement of the bias introduced when reconstructing jets from real data, because the energy measurement is degraded by all the known effects (from the physics and the detector). This bias is obtained by comparing real and simulated jets.

↪ **Jet:** A jet is a stream of particles emitted when a quark (or a gluon) materializes into "stable" particles (not really stable, but so for detection purposes). We measure the total energy of the particles to try and determine the energy of the original quark, because a knowledge of the latter allows us to make important measurements and new discoveries.

↪ **Luminosity:** The luminosity $\mathcal{L}$ is a parameter related to accelerators that measures the number of interactions per cross section and time. Assuming Gaussian transverse particle bunch profile, it is defined as:

$$\mathcal{L} = \frac{1}{4\pi} \frac{N_1 N_2}{\sigma_x \sigma_y} f N_b \tag{A.1}$$

where $N_{1,2}$ are the number of particles in the two colliding beams respectively, f is the bunch-crossing frequency, N_b is the number of bunches and finally, $\sigma_{x,y}$ are the Gaussian widths in the horizontal and vertical plane of the bunch respectively. The integrated luminosity corresponds to the $\mathcal{L}$ integrated over a given period of time.

↪ **Meson:** It is a subatomic particle composed of one quark and one antiquark ($q_1\bar{q}_2$), bound together by the strong interaction. It is a boson since it has an integer spin. It belongs to the hadron family of particles. The most familiar mesons are the pions ($\pi^+ = u\bar{d}$, $\pi^0 = u\bar{u}$ or $d\bar{d}$, $\pi^- = \bar{u}d$) and the kaons ($K^+ = u\bar{s}$, $K^0 = d\bar{s}$, $\bar{K^0} = \bar{d}s$ and $K^- = s\bar{u}$).

↪ **Pile Up (PU):** It is related to multiple simultaneous proton-proton collisions occuring very close from the collision of interest. When detectors are sensitive to several bunch-crossings, these collisions can affect the signal in the collision of interest.

↪ **Sea and Valence Quarks:** Hadrons, which are made of valence quarks that give rise to their quantum numbers, contain virtual $q\bar{q}$ pairs known as

sea quarks. Sea quarks form when a gluon of the hadron color field splits. This process also works in reverse in that the annihilation of two sea quarks produces a gluon. The result is a constant flux of gluon splits and creations colloquially known as "the sea". Sea quarks are much less stable than their valence counterparts, and they typically annihilate each other within the interior of the hadron. Despite this, sea quarks can hadronize into baryonic or mesonic particles under certain circumstances.

↪ **Second LHC running period (Run-II):** It is the second operating period of the LHC, which started on March 2015 and will finish at the end of 2017.

↪ **Trimming:** It is a jet grooming algorithm. Look at Section 2.4

↪ **Underlying Event (UE):** The Underlying Event [187] is an irreducible background to all processes at hadron colliders such as the LHC. It consists essentially of all of the soft processes accompanying a hard scatter due to additional interacting partons from the same protons. From an experimental point of view, it is impossible to separate these contributions; however, topological properties of the event can be used to define a set of physics observables that are sensitive to different aspects of the UE.

↪ **Unfolding:** It is a procedure that transforms the detector-level result to a parton-level result that can be compared to theoretical predictions. The unfolding technique corrects the observed distribution for resolution and acceptance effects by inversion of the response matrix, which is constructed with MC events.

↪ **User Interface (UI):** At IFIC Tier-3 a User Interface is the component which mediates between end-users and the local file system called Lustre. Currently, there are 5 UI available at ific: ui04.ific.uv.es, ui05.ific.uv.es, ui06.ific.uv.es, sctui01.ific.uv.es and sctui02.ific.uv.es.

A.2 Acronyms

$\mathbf{A_C}$: Charge Asymmetry

ADC: ATLAS Distributed Computing

$\mathbf{A_{FB}}$: Forward-Backward Asymmetry

ALICE: A Large Ion Collider Experiment

ARC: Advanced Resource Connector

ATLAS: A Toroidal LHC Apparatus

C/A: Cambridge-Aachen algorithm

CERN: Conseil Européen pour la Reserche Nucléaire

CMS: Compact Muon Solenoid

CR: Control Region

CSIC: Consejo Superior de Investigaciones Cieníficas

DDM: Distributed Data Management

EGI: European Grid Initiative

EM: electromagnetic

EWSB: Electroweak Symmetry Breaking

FAX: Federated Data Storage System

FCNC: Flavour Changing Neutral Current

FSR: Final State Radiation

GPD: Grid Distributed Production

IFAE: Institut de Física d'Altes Energies, Barcelona

IFIC: Instituto de Física Corpuscular, Valencia

ISR: Initial State Radiation

IT: Informatics

LCG: LHC Computing Grid

LEP: Large Electron-Positron Collider

LHC: Large Hadron Collider

LHCb: Large Hadron Collider Beauty

LL: leading logarithmic

LO: leading order

MC: Monte Carlo

MPI: Multiple Parton Interactions

NLL: next-to-leading-logarithmic

NLO: next-to-leading order

NorduGrid: Nordic Data Grid Facility

OSG: Open Science Grid

PDF: Parton Density Functions

PU: pile-up

PyFBU: Python Fully Bayesian Unfolding

QCD: Quantum Chromodynamics

QED: Quantum Electrodynamics

$\sqrt{\mathbf{s}}$: Centre-of-mass Energy

SAM: Service and Availability Monitoring System

SCR: Signal Control Region

SF: Scale Factor

SM: Standar Model

TSVD: Truncated Singular Value Decomposition

UAM: Universidad Autónoma de Madrid, Madrid

UE: Underlying Event

WLCG: Worldwide LHC Grid Computing

XS or Xsec: Cross Section

B

The Bootstrap Generator Method

The Bootstrap Generator method is applied to each systematic uncertainty in the boosted A_C measurement to remove statistical fluctuations due to limited MC sample size. The Bootstrap Method first defines n identical histograms, each representing a pseudo-dataset for $\Delta|y|$. For each event, n corresponding random weights, w_i, are generated from a Poisson distribution, $P(\lambda = 1)$. Each of the n replica histograms is then filled with the corresponding Poisson-fluctuated events. This process is applied to the nominal sample and all of the systematic samples. For a given systematic, for each pseudo-dataset $i \in [0, n]$ the relative difference with respect to the i-th pseudo-dataset of the nominal is computed for each bin. Finally, the RMS of the relative differences is taken as the statistical uncertainty on that component. Additionally, to properly correlate statistical uncertainty among measurements that share the same events, a seed is used in each event to generate the random weights.

Once we have the statistical uncertainty, a smoothing procedure is applied to eliminate the statistical fluctuations on the relative systematic uncertainty. The $\Delta|y|$ distribution contains only 4 bins, thus the smoothing procedure implemented here differs from that described in [188, 189]. The smoothing procedure is outlined as follows, where a bin is significant if it is inconsistent with 0 ($> 2\sigma$) and has a small relative uncertainty ($< 30\%$):

- Determine whether each bin is significant or not based on the bin content and statistical uncertainty.
- If at least one bin is significant, leave each bin's content unchanged and exit.
- If no bins are significant, combine all 4 bins and check the combined significance.

- ▷ If the combined bins are significant, set each bin's content to the combined values and exit.
- ▷ If the combined bins are insignificant, set each bin's content to 0.0 and exit.

No bias is introduced in the asymmetry by implementing the smoothing procedure this way. Other smoothing procedures combine bins moving from left and right edges of the distributions, which could bias the asymmetry by combining positive $\Delta|y|$ values to be smoothed, while negative $\Delta|y|$ values are left un-smoothed and mis-represent the systematic uncertainty. Further, with only 4 bins in the given distribution, combining bins in an iterative approach is not well-suited for this distribution.

This classification results in three distinct cases of smoothing: No smoothing applied (at least one significant bin), combined significance (combined, the bins are significant), and combined insignificant (combined, the bins are still insignificant). These three cases are illustrated in Figure B.1 for the inclusive and differential measurement. In each plot, the data points each correspond to different bins in each measurement: 4 data points for the inclusive measurement and 12 data points for the measurement in $m_{t\bar{t}}$ (4 bins in $\Delta|y|$ and 3 bins in $m_{t\bar{t}}$). In these figures, there is some asymmetry between the up and down variations for a given systematic. Because of this, in the differential measurement, there bins in $m_{t\bar{t}}$ that do not have equal treatment for the up and down variations, i.e., the up variation may be smoothed, and the down variation is not, or vice-versa.

In the differential measurements, the smoothing is only applied to the 4 $\Delta|y|$ bins in each differential bin, separately, because the asymmetry is calculated in each differential bin.

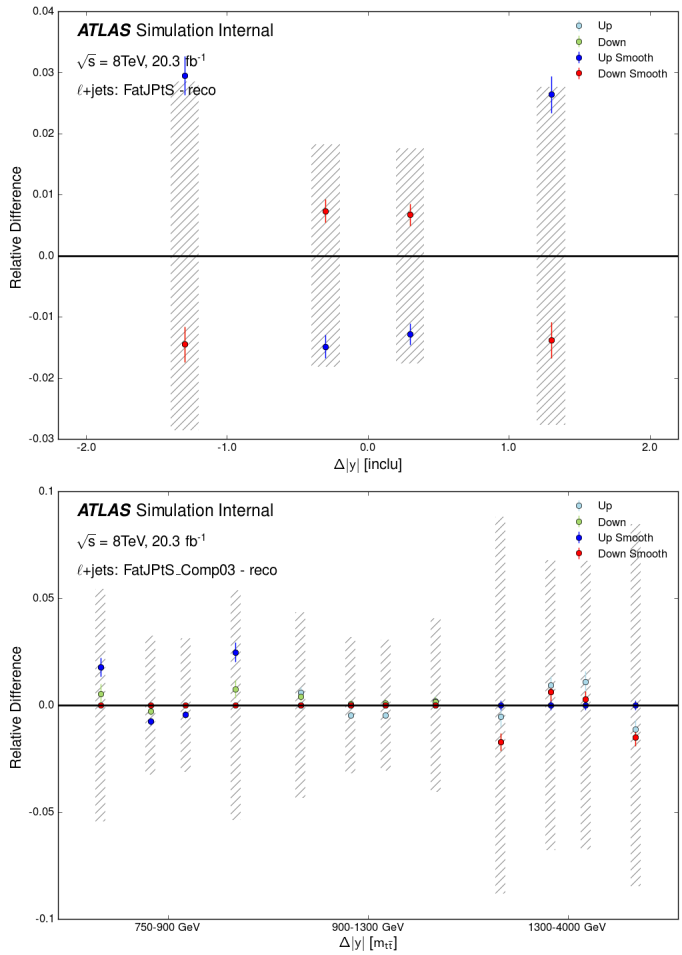


Figure B.1: The systematic uncertainty determined before unfolding and after the bootstrap method has been applied. The plot on the top shows the systematic uncertainty for the large-R jet energy scale in each inclusive bin of $\Delta|y|$ for the $t\bar{t}$ sample. The plot on the bottom shows the differential result for the third p_T component in the large-R jet energy scale for the $t\bar{t}$ sample. The different data points correspond to different bins in $\Delta|y|$, as noted above. The hatched marks show the relative statistical uncertainty due to data in each bin.

C

Unfolding Method

This appendix presents the procedure that transforms the detector-level result (shown in Section 8.2) to a parton-level result that can be compared to theoretical predictions. The unfolding procedure employed here was developed for the measurement of the charge asymmetry in 7 TeV data [190]. The procedure used here is kept for all 8 TeV measurements the procedure in Reference [184]. Four distinct steps are described:

- A final selection criterion to remove badly reconstructed events.
- The combination of the e+jets and μ+jets channels.
- The correction for bin migrations due to the detector resolution.
- The acceptance correction.

After the reconstruction of the top quark and anti-top quark candidates, discussed in Section 6.4.2, events where the pair has an invariant mass of less than 750 GeV are discarded. The distribution is unfolded to the parton-level of top quarks before decay, but after radiation. The event selection limits the measurement to high-mass top quark pairs in the central detector. Correcting back to the fully inclusive parton-level would imply a very large extrapolation, and consequently a large modelling uncertainty. We therefore limit the fiducial region for the parton-level top quarks to the measured phase space[1]:

$$m_{t\bar{t}} > 750\ GeV, \quad |\Delta|y|| < 2 \tag{C.1}$$

The results of e+jets and μ+jets channels are merged before unfolding. The data from both channels are added to form the combined $\Delta|y|$ distribution.

[1] A more detailed fiducial region, that includes the exact requirements placed on the charged lepton, neutrino and jets, has the potential to further decrease the modelling uncertainty. However, that approach does not allow a direct comparison to fixed-order calculations. The full fiducial correction is therefore left for a future analysis.

Also the background templates and the response matrices used to correct the measurement are simply added up. We have shown in Section 6.4.2 (in particular in Figure 6.2) that the response for both channels is quite similar. The relative weight of both channels in the combination, given by the small difference in yield in both channels observed in Section 8.2, is correctly taken into account in this procedure.

The observed $\Delta|y|$ distribution is corrected for resolution effects using a matrix unfolding, in particular the open-source PyFBU implementation of the Fully Bayesian Unfolding (FBU [173]) algorithm. If we define N_t and N_r as the number of truth and reconstructed bins, we can formulate the unfolding procedure as follows. Given an observed spectrum $\boldsymbol{D} \in \mathbb{N}^{N_r}$, a response matrix $\mathcal{M} \in \mathbb{R}^{N_r} \times \mathbb{R}^{N_t}$ and a prior probability distribution π for the true distribution $\boldsymbol{T}$, FBU estimates the posterior probability density for the true distribution:

$$p\left(\boldsymbol{T}|\boldsymbol{D},\mathcal{M}\right) \propto \mathcal{L}\left(\boldsymbol{D}|\boldsymbol{T},\mathcal{M}\right) \cdot \pi\left(\boldsymbol{T}\right) \tag{C.2}$$

where $\mathcal{L}\left(\boldsymbol{D}|\boldsymbol{T},\mathcal{M}\right)$ is the likelihood function of $\boldsymbol{D}$ given $\boldsymbol{T}$ and $\mathcal{M}$ and π is the prior probability density for the true spectrum $\boldsymbol{T}$. As a cross-check, the unfolding is also performed with an alternative method, which is described in Appendix E.

The response matrices are obtained from MC. The nominal result is obtained with the Powheg+Pythia sample, with $h_{damp} = m_t$. The $\Delta|y|$ migrations are corrected with a 4×4 matrix [2].

A bin-by-bin correction deals with the bias of the charge asymmetry introduced by the selection. This acceptance correction multiplies the corrected result by the ratio found in the nominal Powheg MC of the number of events in each bin of the true $\Delta|y|$ distribution after and before the selection[3]. The bin-by-bin factors are shown in Tables C.1 and C.2 for the inclusive and differential cases respectively. The signal acceptance increases strongly with invariant mass of the $t\bar{t}$ system (the acceptance does not reach a plateau until a $m_{t\bar{t}}$ of 1.5 TeV, see Reference [151]). The acceptance is typically higher in the central bins, especially for relatively low-mass events. Importantly, for all bins the acceptance is found to be nearly symmetric: the acceptance found for the first

[2] At reconstructed level the rapidity difference boundaries are not strictly enforced by the selection. Migrations may moreover lead to events with a true rapidity difference outside the range [-2,2]. This small fraction (0.1%) of MC events in the overflow is added to the first or last $|\Delta|y||$ bins of the response matrix.

[3] A cut is applied of $m_{t\bar{t}}^{true} > 750\ GeV$ for the inclusive measurement, mass migrations are handled explicitly in the differential measurement, as explained below

and fourth $\Delta|y|$ bins and for the second and third bins is similar. We thus find only a small bias in the charge asymmetry due to the selection. With this acceptance correction the result is brought back to the phase space in Equation C.1.

$m_{t\bar{t}}^{truth}$ bin	Inclusive	< 750 GeV
First $\Delta\|y\|_{truth}$ bin	2.507 ± 0.018	0.0423 ± 0.0003
Second $\Delta\|y\|_{truth}$ bin	5.759 ± 0.028	0.0423 ± 0.0003
Third $\Delta\|y\|_{truth}$ bin	6.073 ± 0.029	0.0423 ± 0.0003
Fourth $\Delta\|y\|_{truth}$ bin	2.466 ± 0.17	0.0423 ± 0.0003

Table C.1: The bin-by-bin signal acceptance correction factors, in %, for the inclusive response matrix. Uncertainties are statistical only. The $\Delta|y|$ binning for inclusive is [-2.0, -0.7, 0, 0.7, 2.0].

$m_{t\bar{t}}^{truth}$ bin	750-900 GeV	900-1300 GeV	> 1300 GeV
First $\Delta\|y\|_{truth}$ bin	1.589 ± 0.016	4.200 ± 0.035	6.43 ± 0.13
Second $\Delta\|y\|_{truth}$ bin	3.302 ± 0.035	6.49 ± 0.06	7.79 ± 0.14
Third $\Delta\|y\|_{truth}$ bin	3.61 ± 0.04	7.09 ± 0.06	8.37 ± 0.15
Fourth $\Delta\|y\|_{truth}$ bin	1.561± 0.015	4.081 ± 0.034	6.61 ± 0.12

Table C.2: The bin-by-bin signal acceptance correction factors, in %, for the differential response matrices. Uncertainties are statistical only. The $\Delta|y|$ binning for differential is [-2.0, -0.4, 0, 0.4, 2.0], [-2.0, -0.5, 0, 0.5, 2.0], [-2.0, -0.6, 0, 0.6, 2.0] in increasing bins of $m_{t\bar{t}}$.

The prior probability density must be chosen according to what we know about the true distribution before the measurement. The simplest choice is an *uninformative* prior that assigns equal probabilities to all $\boldsymbol{T}$ spectra within a range $[T^{\ulcorner}, T^{\urcorner}]$ wide enough to avoid a bias of the result[4]:

$$\pi(\boldsymbol{T}) \propto \begin{cases} 1 & \text{if } T_t \in [T^{\ulcorner}, T^{\urcorner}], \forall t \in [1, N_t] \\ 0 & \text{otherwise} \end{cases} \tag{C.3}$$

The multi-dimensional space in a broad range around the prior is sampled using the Metropolis-Hastings algorithm [191]. For each point, $\pi(\boldsymbol{T})$ and the

[4]FBU can modify these priors by introducing a regularization parameter α and a regularization function S, but, following previous ATLAS analyses of the charge asymmetry, all results presented in this note correspond to $\alpha = 0$ (no regularization).

likelihood $\mathcal{L}(\boldsymbol{D}|\boldsymbol{T})$ are evaluated. The mean of the posterior distribution yields the unfolded charge asymmetry. The RMS of the posterior distribution is taken as the uncertainty (the statistical uncertainty at this point, as no systematic uncertainties are taken into account yet).

The robustness of this method is evaluated using pseudo-data distributions from MC, where the true asymmetry is known. Several different asymmetries are injected by reweighting the SM $t\bar{t}$ sample according to the true $\Delta|y|$ of the events. The corrected charge asymmetry is shown as a function of the true asymmetry in Figure C.1. The response of the unfolding procedure is found to be quite linear, with the slope of the curve generally close to 1.

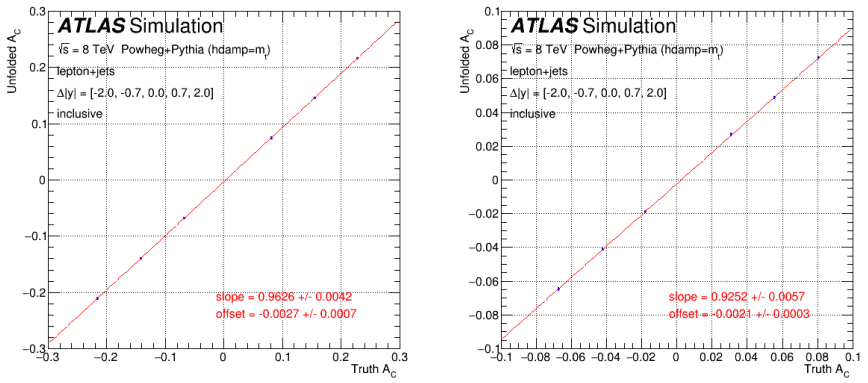


Figure C.1: Linearity tests of the correction procedure with two different reweighting schemes (Linear in the leftmost panel, PROTOS in the rightmost panel). The unfolded result is shown versus the true asymmetry for several MC samples. The response matrices are derived from the nominal Powheg+Pythia MC, with $h_{damp} = m_t$.

As a further stress to the unfolding procedure, the MC events are reweighted both in the true $\Delta|y|$ and the true p_T of the hadronically decaying top quark, to mimic the shape differences between data and MC observed in Section 8.2. Unfolding of the Asimov data set built with the reweighted $t\bar{t}$ sample indeed recovers the true asymmetry after reweighting to within 0.1%. A reweighting of the top quark p_T spectrum to bring the nominal MC in agreement with the observed detector-level spectrum has a negligible impact on the result. Even for a reweighting of the $h_{damp} = \infty$ samples the effect on the charge asymmetry is less than 0.05%.

A number of tests are furthermore performed to test the impact of the resolution on the result. A first cross-check removes the events with small absolute $\Delta|y|$ that are much more likely to flip sign. All events with reconstructed $|\Delta|y|| < 0.3$ are removed from the data sample and from the MC sample used to generate the response matrix. The selection efficiency used in the unfolding is determined with all truth events in the denominator. This yields a result for A_C that is within 0.1% of the result reported in this note. A separate determination of the charge asymmetry in the first and last $\Delta|y|$ and from the two central bins yields good closure for both.

The procedure above is also used to unfold two-dimensional distributions, in which the second dimension corresponds to the bins of a differential variable, $m_{t\bar{t}}$. The spectrum formatting, however, is different in that the two-dimensional histogram is transformed into a one-dimensional histogram. The second dimension of the two-dimensional histogram is wrapped as several consecutive sub-ranges of the first dimension. For four $\Delta|y|$ bins, a response matrix with $N \cdot 4 \times N \cdot 4$ elements is constructed, where N indicates the number of $t\bar{t}$ invariant mass bins. For the differential measurement we measure the charge asymmetry in three mass bins (i.e. $N = 3$). A fourth mass bin, with a single $\Delta|y|$ bin is introduced to keep track of non-negligible migrations across the lower edge of the fiducial region at $m_{t\bar{t}} = 0.75\ TeV$. We thus construct a response matrix with 13×12 elements, as shown in Figure C.2. This matrix describes migrations both across the four $\Delta|y|$ bins in each mass bin and between the mass bins.

The treatment of systematic uncertainties is naturally included in the Bayesian inference approach by extending the likelihood function of the data (given the true distribution and the transfer matrix) with nuisance parameters terms. The *marginal* likelihood is defined as:

$$\mathcal{L}\left(\boldsymbol{D}|\boldsymbol{T}\right) = \int \mathcal{L}\left(\boldsymbol{D}|\boldsymbol{T},\boldsymbol{\theta}\right) \cdot \pi(\boldsymbol{\theta})\ d\boldsymbol{\theta}, \tag{C.4}$$

where $\boldsymbol{\theta}$ are the nuisance parameters, and $\pi(\boldsymbol{\theta})$ their prior probability densities, which are assumed to be Gaussian distributions G with $\mu = 0$ and $\sigma = 1$. A nuisance parameter is associated with each of the uncertainty sources. Two categories of nuisance parameters are considered: the normalizations of the background processes ($\boldsymbol{\theta}_b$), and the uncertainties associated with the object identification, reconstruction and calibration ($\boldsymbol{\theta}_s$). While the first ones only affect the background predictions, the latter, referred to as object systematic uncertainties, affect both the reconstructed distribution for $t\bar{t}$ signal and the total background prediction, referred to as $\boldsymbol{R}(\boldsymbol{T};\boldsymbol{\theta}_s)$ and $\boldsymbol{B}(\boldsymbol{\theta}_s,\boldsymbol{\theta}_b)$, respectively.

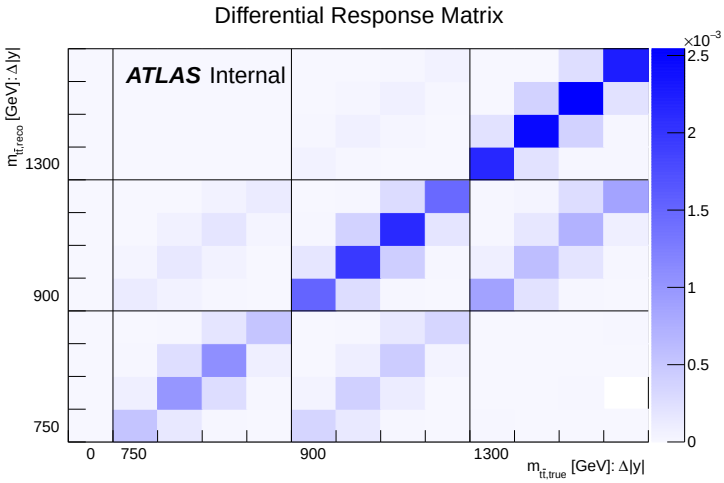


Figure C.2: The response matrix that describes migrations in $\Delta|y|$ and $m_{t\bar{t}}$. All elements of the matrix are normalized such that the sum of all entries of a column sum up to the efficiency for that bin.

The prior probability densities $G(\boldsymbol{\theta}_b)$ are truncated in order to avoid negative background normalizations.

The marginal posterior probability density for $\boldsymbol{T}$ is computed by sampling the $N_t + N_{np}$ parameter space, where N_{np} is the total number of nuisance parameters, and projecting the sample over the $\boldsymbol{T}$ parameter space. The projections over each nuisance parameter give the corresponding posterior probability density, which matches the Gaussian prior for unconstrained nuisance parameters, while it has a narrower shape for nuisance parameters that can be measured in the dataset. The posterior probability density for the charge asymmetry is computed as described before, with the difference that the RMS of the marginal posterior represents the total uncertainty. Analogously, each nuisance parameter is estimated by the mean and RMS of the corresponding projection of the posterior probability density.

Figure C.3 shows the pulls and constraints on the systematic uncertainties relating to detector response and normalization, after the FBU procedure is applied to the inclusive Asimov data and data. In the FBU procedure, a gaussian prior with $mean = 0$ and $sigma = 1$ are assigned to the nuisance parameters. One can see that all of the nuisance parameters are centred near 0

and their bands extend to the 1 sigma line. This shows that with the limited statistics and the reduced number of bins (4) the fit can not pull or constrain any of the systematics significantly.

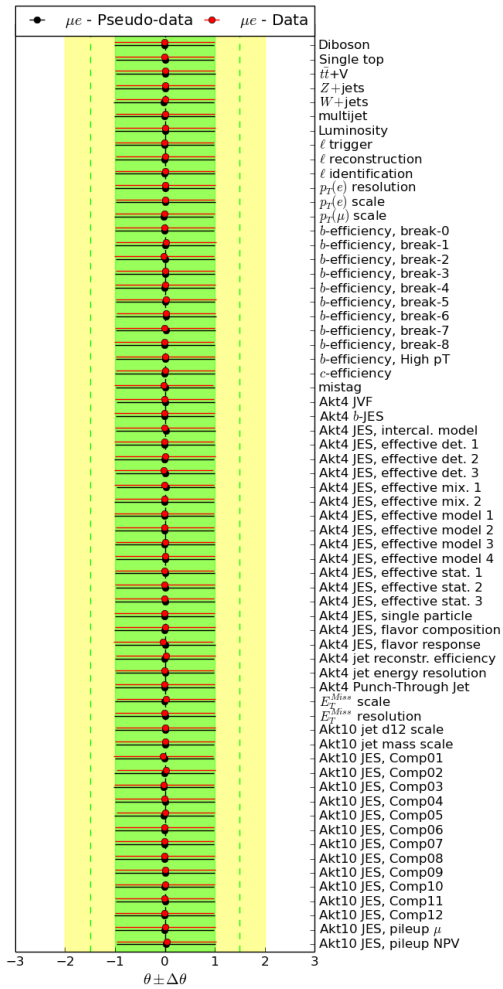


Figure C.3: Pulls and constraints on the nuisance parameters corresponding to detector-related and normalization systematic uncertainties for the Asimov data set (black markers) and the Data (red markers). The green band reflects the 1σ variation, the yellow band the 2σ variation. The dashed green line is just meant to guide the eye to 1.5σ variation.

D

Systematic uncertainties

Systematic uncertainties on the A_C measurement arise from limitations of the understanding of the reconstruction and calibration of physics objects, as well as from imperfections in the MC generator modelling and background estimation. Each of the contributions to the systematics is discussed briefly below. Additionally, to limit the impact of statistical fluctuations due to the limited MC sample size, systematic variations are evaluated using the Bootstrap Generator method. The Bootstrap method is applied to smooth systematic uncertainties that suffer from large statistical uncertainties. To smooth a given systematic, the statistical uncertainty is estimated from a set of n replicas of the nominal spectrum, derived by fluctuating each event in the sample by a Poisson distribution with average one. The resulting statistical uncertainty is applied to the smoothing procedure, outlined in Appendix B.

The dominant experimental effect stems from the modelling of the response of the experiment to jets. The jet energy and mass scales (JES and JMS) are defined using an MC-based calibration [17]. The large-R jet energy scale (JES), jet mass scale (JMS) and k_t splitting scale are constrained using data. For all observables the double ratio of the tracker/calorimeter response and data/MC provides a constraint. For the jet energy scale up to $p_T < 800\ GeV$, the tightest constraint is found in γ+jets events. The peak position of the boosted W boson in $t\bar{t}$ events is used to constrain the jet mass response. Sixteen components of the JES uncertainty are varied separately for large-R jets. An additional MC-based uncertainty, referred to as large-R JES topology uncertainty, is included to reflect the fact that the jets in these calibration samples have a different response (gluon or light quark jets) than those in $t\bar{t}$ events (top-jets). The full difference between the response of these two types of jets is conservatively assumed as the corresponding systematic uncertainty.

The uncertainty on the large-R jet energy resolution (JER) is determined by smearing the jet energy such that the resolution is increased by 20% [18]

and evaluating the effect on the final result. The same smearing procedure is applied to determine the uncertainty due to the large-R jet mass resolution (JMR).

The small-R jet energy scale uncertainty is derived using a combination of simulations, test beam data and *in situ* measurements [192, 148, 193]. Additional contributions from the jet flavour composition, calorimeter response to different jet flavours and pile-up are taken into account[1]. Uncertainties in the jet energy resolution are obtained with an *in situ* measurement of the jet response asymmetry in di-jet events [194].

The efficiency to tag real and fake b-jets is corrected in MC events by applying b-tagging scale factors, extracted in $t\bar{t}$ and dijet samples, that compensate for the residual difference between data and simulation. The associated systematic uncertainty is computed by varying the scale factors within their uncertainty [167, 195, 196]. The b-jet calibration is performed up to $p_T = 300\ GeV$; for larger transverse momenta an additional MC-based extrapolation uncertainty, which ranges from 10% to 30% increasing with b-jet p_T from 300 GeV to 1200 GeV, is applied[2].

The lepton reconstruction efficiency in the simulation is corrected by scale factors derived from measurements of these efficiencies in data using a $Z \to l^+l^-$ enriched control region. The lepton trigger and reconstruction efficiency scale factors, energy scale and resolution are varied within their uncertainties [197, 198].

The uncertainty associated with E_T^{miss} is calculated by propagating the energy scale and resolution systematics on all physics objects to the E_T^{miss} calculation. Additional E_T^{miss} uncertainties arising from energy deposits not associated with any reconstructed objects are also included [199].

An important set of uncertainties on the correction of the data to the parton-level stems from the imperfections in the MC generators, parton shower model and hadronization. The response is found to vary only slightly between different generators. The dilution factor for the inclusive response matrix varies from 0.72 to 0.75. Even for a more exotic production mechanism, a Z' boson decaying to top quark pairs, the dilution factor is similar (0.79-0.83, depending

[1] In July 2015 a bug was uncovered in the application of the in situ calibration constants. It leads to a 2% effect on the JES for high-p_T jets that is not covered by the systematic uncertainty. Large-R jets are not affected as in this case the γ+jet constraint is only used to reduce the uncertainty, not to modify the actual energy scale. The impact of the small-R jet uncertainties on this analysis is negligible (in fact, the corresponding nuisance parameters are pruned by the bootstrap method). Therefore, after discussion with the EB and top group conveners, we decided not to reprocess the data.

[2] The b-jet p_T distribution extends to approximately 600 GeV in this analysis.

on the assumed Z' mass).

The signal modelling uncertainty is evaluated by comparing several alternative MC generators. The uncertainty associated with the Matrix Element generator is estimated by comparing an alternative NLO generator (MC@NLO) with the nominal choice (POWHEG), where both are interfaced to HERWIG for the parton shower and hadronization (samples 105200+AFII and 105860+AFII). The uncertainty related to the parton shower is obtained by comparing POWHEG+HERWIG to POWHEG+ PYTHIA (samples 105860+AFII and 117050+AFII, respectively), and the choice of parton density function (CT10, MSTW, NNPDF and their error sets, following the PDF4LHC prescription) are determined in samples generated with POWHEG+PYTHIA with $h_{damp} = \infty$. The effect of ISR/FSR modelling is estimated with dedicated ACERMC samples with upward and downward variations of the ISR/FSR parameters. In all cases the variation is estimated as the double difference:

$$(A_C^{true,nom} - A_C^{true,alt}) - (A_C^{corr,nom} - A_C^{corr,alt}), \tag{D.1}$$

where A_C^{true} and A_C^{corr} refer to the true and unfolded A_C and *nom* and *alt* identify the two Asimov data sets. The response matrix and acceptance correction are based on the nominal sample in all cases. The uncertainty for PS and ME is the full double difference. We take the total modelling uncertainty as sum in quadrature of all the above components (ME, PS, ISR/FSR, PDF). The impact of different choices for the top quark mass ($m_t = 170\ GeV$ and $m_t = 175\ GeV$) is evaluated and found to be negligible.

The W+jets systematic uncertainties due to the overall normalization and the heavy flavor fraction are obtained by varying the data-driven scale factors within their uncertainty.

The uncertainty on the fake lepton background is included by varying the definition of loose leptons, changing the selection used to form the control region and propagating the statistical uncertainty of parameterizations of the efficiency and the fake rate.

A 48% uncertainty is applied to the normalization of the Z+jets and 38% to the diboson backgrounds. The single-top background normalization is assigned a 7.7% uncertainty associated with the theory calculations [140, 141, 142], and an additional 24%, added in quadrature, for each additional jet. For associated production of a top quark pair and a vector boson ($t\bar{t}V$) the normalization uncertainty is taken to be 50%.

The uncertainty that stems from the unfolding procedure is estimated as the squared sum of two components. The uncertainty due to the limited statistics in the MC sample is estimated using 1000 pseudo-experiments where the elements

of the migration matrix are fluctuated. The bias due to the non-linear reponse is estimated from the linearity tests that inject non-zero A_C using several reweighting functions.The total unfolding uncertainty is 0.5% for the inclusive measurement and reaches 1.7% for the mass bin greater than 1.3 TeV.

The impact of each of these sources of systematic uncertainty on the inclusive A_C measurement is presented in Table D.1. The largest source of uncertainty is the modelling of the $t\bar{t}$ signal, where alternative choices of matrix element, parton shower, top quark mass and parton density functions may lead to nearly 2% change in the result. The uncertainties on the normalization of non-$t\bar{t}$ SM processes contribute only at the per mil level. Uncertainties in modelling the response of the ATLAS detector are small: 0.4% for the inclusive measurement, dominated by the uncertainty in the fat jet energy and mass response. The unfolding systematic, that includes the effect of limited MC statistics (evaluated by smearing the response matrices) and the bias and non-linearity in the unfolding procedure, is 0.3% for the inclusive measurement.

source of uncertainty	δA_C (%)
signal modelling - PS	$\pm$ 2.0
singal modelling - ME	$\pm$ 1.5
signal modelling - ISR/FSR	$\pm$ 0.1
signal modelling - m_t	$\pm$ 0.1
signal modelling - PDF	$\pm$ 0.4
signal modelling - total	$\pm$ 2.5
background norm.	$\pm$ 0.10
jet energy and resolution - $R = 0.4$ jets	$\pm$ 0.11
jet energy and mass scale - $R = 1.0$ jets	$\pm$ 0.32
b-tag/mis-tag efficiency	$\pm$ 0.18
lepton reco/id/scale	$\pm$ 0.09
missing tranverse energy (E_T^{miss})	$\pm$ 0.05
unfolding & MC stat.	$\pm$ 0.56

Table D.1: The effect on the A_C^{corr} of systematic uncertainties in the signal and background modelling and the description of the detector response. The $\pm 1\sigma$ uncertainties are to be added to the result. (i.e. they are not relative errors).

In Table D.2 the same breakdown of the systematic uncertainties is presented for the three $m_{t\bar{t}}$ bins. The largest sources of uncertainty in each bin are the modelling of the $t\bar{t}$ signal. Total uncertainties due to the $t\bar{t}$ signal modelling are

at the level of 4.0, 1.4 and 9.3% for the three respective bins in $m_{t\bar{t}}$, while the combined uncertainties modelling the ATLAS detector response are 2.2, 0.6, and 1.5%.

	δA_C (%) in three $m_{t\bar{t}}$ (TeV) intervals		
source of uncertainty	[0.75-0.9]	[0.9-1.3]	[1.3-4.0]
signal modelling - PS	3.2	1.2	6.2
signal modelling - ME	2.4	0.6	5.3
signal modelling - ISR/FSR	0.3	0.1	3.0
signal modelling - PDF	0.4	0.3	3.3
signal modelling - total	4.0	1.4	9.3
background norm.	± 0.16	± 0.25	± 0.37
jet energy and resolution - $R = 0.4$ jets	± 0.45	± 0.27	± 0.43
jet energy and resolution - $R = 1.0$ jets	± 1.61	± 0.59	± 1.04
b-tag/mis-tag efficiency	± 0.23	± 0.15	± 0.67
lepton reco/id/scale	± 0.17	0.10	0.14
E_T^{miss}	± 0.03	± 0.03	± 0.09
unfolding & MC stat.	1.2	0.8	2.1

Table D.2: The effect on the corrected A_C in the three $m_{t\bar{t}}$ bins of systematic uncertainties in the signal and background modelling and the description of the detector response. The $\pm 1\sigma$ uncertainties are to be added to the result (i.e. they are not relative errors). First bin is $0.75 < m_{t\bar{t}} < 0.9$, second bin is $0.9 < m < 1.3$ and third bin is $1.3 < m < 4.0$, in TeV.

E

Cross-check with an alternative unfolding

This appendix shows a comparison between two different unfolding methods: TSVD vs FBU. The former does regularization while the latter does not.

As a cross-check of the nominal result obtained with Fully Bayesian method described in Section C, we repeat the extraction of the charge asymmetry in the $\Delta|y|$ distribution with an alternative unfolding method. To allow a direct comparison we unfold the three mass bins separately, i.e., without accounting for the migrations between $t\bar{t}$ mass bins. The alternative unfolding is based on regularized matrix unfolding with a method known as Singular Value Decomposition (TSVD [200]). The method works as follows:

- Background subtraction: performed by subtracting the MC expectation for all non-$t\bar{t}$ background from the data.
- Correction for detector smearing: the response matrix is inverted using singular value decomposition, with regularization parameter $kreg = N_{bins} - 1$.
- Acceptance correction: the asymmetry is corrected to the full fiducial region ($m_{t\bar{t}} > 750\ GeV$, $|\Delta|y|| < 2$) using a bin-by-bin correction.

The TSVD result for the charge asymmetry extracted from Asimov data for the inclusive measurement is in good agreement with the nominal result obtained on the same MC with FBU ($A_C = 0.70\%$ for TSVD with $kreg = 3$ or 4, versus 0.73% for FBU).

The uncertainty on the corrected result due to the limited statistics in the data sample and the MC is evaluated using pseudo-experiments, where the data (or MC) event count in each bin (or element of the matrix) is varied according to a Gaussian distribution centred on the nominal result and with a width equal to the statistical uncertainty in that bin (i.e. $\sqrt{N}$).

We limit the fiducial region for the parton-level top quarks to the measured phase space ($m_{t\bar{t}} > 750\ GeV$, $|\Delta|y|| < 2$). We also perform a differential measurement, correcting for migrations in $\Delta|y|$ and $m_{t\bar{t}}$, using the following bins:

- ▷ Differential1: $750 < m_{t\bar{t}} < 900\ GeV$
- ▷ Differential2: $900 < m_{t\bar{t}} < 1300\ GeV$
- ▷ Differential3: $1300 < m_{t\bar{t}} < 4000\ GeV$

Detector level

For reference, the detector-level charge asymmetry of the data sample, and the expectations for SM $t\bar{t}$ production and all SM sources that contribute to the event yield (the Asimov set) are given in Table E.1.

	data A_C (%)	asimov A_C (%)	$t\bar{t}$ A_C (%)
Inclusive	3.7 ± 1.3	1.3 ± 1.2	1.2 ± 1.3
Differential1	3.1 ± 2.0	1.5 ± 1.9	1.0 ± 1.9
Differential2	5.021 ± 1.809	1.2 ± 1.7	1.3 ± 1.8
Differential3	0.451 ± 3.878	1.3 ± 3.6	2.0 ± 4.0

Table E.1: Detector level charge asymmetry for data (leftmost column), the Standard Model expectation based on MC estimates for all source (central column) and the expectation for $t\bar{t}$ production (rightmost column).

The statistical uncertainty quoted in the table is computed from the (expected) number of events with positive $N_+ = N(\Delta|y| > 0)$ and negative sign $N_- = N(\Delta|y| < 0)$, using the following relation:

$$\begin{aligned} \epsilon(A_C) &= \sqrt{\left(\frac{\partial A_C}{\partial N_+} \cdot \epsilon(N_+)\right)^2 + \left(\frac{\partial A_C}{\partial N_-} \cdot \epsilon(N_-)\right)^2} = \\ &= \sqrt{\frac{1 - A_C^2}{N_+ + N_-}} = \frac{2\sqrt{N_+ \cdot N_- \cdot (N_+ + N_-)}}{(N_+ + N_-)^2} \end{aligned} \tag{E.1}$$

Uncertainty due to limited statistics in DATA

To compute the statistical error from data we create 1000 pseudo-experiment, where we smear the input distribution with random numbers from a Gaussian distribution with μ =0 and $\sigma = \sqrt{N}$. Tables E.2 and E.3 show the values for the inclusive and the 3 $m_{t\bar{t}}$ differential bins for data and asimov pseudo-data:

Asimov A_C (%)	PyFBU (α=0)	TSVD (kReg=3)	TSVD (kReg=4)
Inclusive	0.93 ± 1.8	0.79 ± 1.4	0.79 ± 1.4
Differential1	0.76 ± 3.0	0.52 ± 2.6	0.54 ± 2.5
Differential2	1.10 ± 2.5	0.89 ± 1.8	0.89 ± 1.9
Differential3	1.81 ± 6.2	1.03 ± 3.8	1.01 ± 4.2

Table E.2: The average value for the corrected charge asymmetry and the variation in the results obtained with 1000 pseudo-experiments. For each pseudo-experiment the Asimov input spectrum is varied within the statistical uncertainty.

Data A_C (%)	PyFBU (α=0)	TSVD (kReg=3)	TSVD (kReg=4)
Inclusive	4.26 ± 1.9	4.13 ± 1.5	4.20 ± 1.5
Differential1	2.00 ± 3.2	2.92 ± 2.7	2.69 ± 2.6
Differential2	6.74 ± 2.7	6.01 ± 2.0	6.36 ± 2.1
Differential3	0.57 ± 7.1	-0.06 ± 4.4	-0.07 ± 4.8

Table E.3: The average value for the corrected charge asymmetry and the variation in the results obtained with 1000 pseudo-experiments. For each pseudo-experiment the data input spectrum is varied within the statistical uncertainty.

Error due to MC statistics

The error due to the MC statistics can be computed by smearing the response matrix. This means that we have one new matrix every time we smear the original one. Therefore we can check how our matrix affects the Ac measurement. Tables E.4 and E.5 actually tell you how the resolution matrix affects on the measurement:

Asimov A_C (%)	PyFBU (α=0)	TSVD (kReg=3)	TSVD (kReg=4)
Inclusive	0.94 ± 0.4	0.80 ± 0.3	0.80 ± 0.3
Differential1	0.76 ± 0.7	0.71 ± 0.6	0.71 ± 0.6
Differential2	1.09 ± 0.5	0.90 ± 0.7	0.90 ± 0.4
Differential3	1.88 ± 1.2	1.06 ± 1.5	1.06 ± 1.0

Table E.4: The average value for the corrected charge asymmetry and the variation in the results obtained with 1000 pseudo-experiments. For each pseudo-experiment the response matrix used for the correction of the Asimov data is varied within the statistical uncertainty.

data A_C (%)	PyFBU (α=0)	TSVD (kReg=3)	TSVD (kReg=4)
Inclusive	4.26 ± 0.4	4.16 ± 0.3	4.23 ± 0.3
Differential1	2.00 ± 0.7	3.00 ± 0.6	2.77 ± 0.6
Differential2	6.73 ± 0.5	5.88 ± 0.4	6.23 ± 0.5
Differential3	0.66 ± 1.2	-0.15 ± 0.9	-0.19 ± 1.0

Table E.5: The average value for the corrected charge asymmetry and the variation in the results obtained with 1000 pseudo-experiments. For each pseudo-experiment the response matrix used for the correction of the data is varied within the statistical uncertainty.

F

MC samples for ATLAS measurements

Tables F.1, F.2, F.3, F.4, F.5 list all the MC samples that are commonly used in the ATLAS measurements.

Cross-sections used to normalize the MC samples are quoted in Tables F.6 and F.7.

data12_8TeV.periodA.physics_Egamma.PhysCont.NTUP_COMMON.grp14_v01_p1517_p1562/
data12_8TeV.periodB.physics_Egamma.PhysCont.NTUP_COMMON.grp14_v01_p1278_p1562/
data12_8TeV.periodC.physics_Egamma.PhysCont.NTUP_COMMON.grp14_v01_p1278_p1562/
data12_8TeV.periodD.physics_Egamma.PhysCont.NTUP_COMMON.grp14_v01_p1278_p1562/
data12_8TeV.periodE.physics_Egamma.PhysCont.NTUP_COMMON.grp14_v01_p1278_p1562/
data12_8TeV.periodG.physics_Egamma.PhysCont.NTUP_COMMON.grp14_v01_p1278_p1562/
data12_8TeV.periodH.physics_Egamma.PhysCont.NTUP_COMMON.grp14_v01_p1278_p1562/
data12_8TeV.periodI.physics_Egamma.PhysCont.NTUP_COMMON.grp14_v01_p1562/
data12_8TeV.periodJ.physics_Egamma.PhysCont.NTUP_COMMON.grp14_v01_p1562/
data12_8TeV.periodL.physics_Egamma.PhysCont.NTUP_COMMON.grp14_v01_p1562/
data12_8TeV.periodA.physics_Muons.PhysCont.NTUP_COMMON.grp14_v01_p1517_p1562/
data12_8TeV.periodB.physics_Muons.PhysCont.NTUP_COMMON.grp14_v01_p1278_p1562/
data12_8TeV.periodC.physics_Muons.PhysCont.NTUP_COMMON.grp14_v01_p1278_p1562/
data12_8TeV.periodD.physics_Muons.PhysCont.NTUP_COMMON.grp14_v01_p1278_p1562/
data12_8TeV.periodE.physics_Muons.PhysCont.NTUP_COMMON.grp14_v01_p1278_p1562/
data12_8TeV.periodG.physics_Muons.PhysCont.NTUP_COMMON.grp14_v01_p1278_p1562/
data12_8TeV.periodH.physics_Muons.PhysCont.NTUP_COMMON.grp14_v01_p1278_p1562/
data12_8TeV.periodI.physics_Muons.PhysCont.NTUP_COMMON.grp14_v01_p1562/
data12_8TeV.periodJ.physics_Muons.PhysCont.NTUP_COMMON.grp14_v01_p1562/
data12_8TeV.periodL.physics_Muons.PhysCont.NTUP_COMMON.grp14_v01_p1562/

Table F.1: Real data samples.

mc12_8TeV.110404.PowhegPythia_P2011C_ttbar_hdamp172p5_nonallhad.merge.NTUP_COMMON.e3151_s1773_s1776_r4485_r4540_p1575/
mc12_8TeV.110898.Pythia8_AU2MSTW2008LO_EWttbar.merge.NTUP_COMMON.e2512_s1499_s1504_r3658_r3549_p1575/
mc12_8TeV.110404.PowhegPythia_P2011C_ttbar_hdamp172p5_nonallhad.merge.NTUP_COMMON.e3151_a220_a205_r4540_p1575/
mc12_8TeV.117050.PowhegPythia_P2011C_ttbar.merge.NTUP_COMMON.e1728_s1581_s1586_r3658_r3549_p1562/
mc12_8TeV.117050.PowhegPythia_P2011C_ttbar.merge.NTUP_COMMON.e1727_a188_a171_r3549_p1575/
mc12_8TeV.105200.McAtNloJimmy_CT10_ttbar_LeptonFilter.merge.NTUP_COMMON.e1513_s1499_s1504_r3945_r3549_p1562/
mc12_8TeV.105200.McAtNloJimmy_CT10_ttbar_LeptonFilter.merge.NTUP_COMMON.e1513_a159_a171_r3549_p1562/
mc12_8TeV.117840.TTbar_MT1700_nonallhad_PowHeg_Pythia_P2011C.merge.NTUP_COMMON.e2051_s1581_s1586_r3658_r3549_p1562/
mc12_8TeV.117842.TTbar_MT1750_nonallhad_PowHeg_Pythia_P2011C.merge.NTUP_COMMON.e2051_s1581_s1586_r3658_r3549_p1562/
mc12_8TeV.105860.PowhegJimmy_AUET2CT10_ttbar_LeptonFilter.merge.NTUP_COMMON.e1576_a159_a171_r3549_p1575/
mc12_8TeV.105861.PowhegPythia_AUET2BCT10_ttbar_LeptonFilter.merge.NTUP_COMMON.e1317_a159_a165_r3549_p1575/
mc12_8TeV.117209.AcerMCPythia_AUET2BCTEQ6L1_MorePS_ttbar_noallhad.merge.NTUP_COMMON.e1378_a159_a171_r3549_p1575/
mc12_8TeV.117210.AcerMCPythia_AUET2BCTEQ6L1_LessPS_ttbar_noallhad.merge.NTUP_COMMON.e1378_a159_a171_r3549_p1575/
mc12_8TeV.110395.PowhegPythia_P2011C_ttbar_hdamp172p5_nonallhad_mtt_1100_1300.merge.NTUP_COMMON.e3228_s1773_s1776_r4485_r4540_p1575/
mc12_8TeV.110396.PowhegPythia_P2011C_ttbar_hdamp172p5_nonallhad_mtt_1300_1500.merge.NTUP_COMMON.e3228_s1773_s1776_r4485_r4540_p1575/
mc12_8TeV.110397.PowhegPythia_P2011C_ttbar_hdamp172p5_nonallhad_mtt_1500_1700.merge.NTUP_COMMON.e3228_s1773_s1776_r4485_r4540_p1575/
mc12_8TeV.110398.PowhegPythia_P2011C_ttbar_hdamp172p5_nonallhad_mtt_1700_2000.merge.NTUP_COMMON.e3228_s1773_s1776_r4485_r4540_p1575/
mc12_8TeV.110399.PowhegPythia_P2011C_ttbar_hdamp172p5_nonallhad_mtt_2000p.merge.NTUP_COMMON.e3228_s1773_s1776_r4485_r4540_p1575/

Table F.2: $t\bar{t}$ MC samples.

mc12_8TeV.190001.AlpgenPythia_Auto_P2011C_WenuNp1_Akt10Pt250.merge.NTUP_COMMON.e2327_s1581_s1586_r4485_r4540_p1606/
mc12_8TeV.190002.AlpgenPythia_Auto_P2011C_WenuNp2_Akt10Pt250.merge.NTUP_COMMON.e2327_s1581_s1586_r4485_r4540_p1606/
mc12_8TeV.190003.AlpgenPythia_Auto_P2011C_WenuNp3_Akt10Pt250.merge.NTUP_COMMON.e2327_s1581_s1586_r4485_r4540_p1606/
mc12_8TeV.190004.AlpgenPythia_Auto_P2011C_WenuNp4_Akt10Pt250.merge.NTUP_COMMON.e2327_s1581_s1586_r4485_r4540_p1606/
mc12_8TeV.190005.AlpgenPythia_Auto_P2011C_WenuNp5_Akt10Pt250.merge.NTUP_COMMON.e2327_s1581_s1586_r4485_r4540_p1606/
mc12_8TeV.190011.AlpgenPythia_Auto_P2011C_WmunuNp1_Akt10Pt250.merge.NTUP_COMMON.e2328_s1581_s1586_r4485_r4540_p1606/
mc12_8TeV.190012.AlpgenPythia_Auto_P2011C_WmunuNp2_Akt10Pt250.merge.NTUP_COMMON.e2328_s1581_s1586_r4485_r4540_p1606/
mc12_8TeV.190013.AlpgenPythia_Auto_P2011C_WmunuNp3_Akt10Pt250.merge.NTUP_COMMON.e2328_s1581_s1586_r4485_r4540_p1606/
mc12_8TeV.190014.AlpgenPythia_Auto_P2011C_WmunuNp4_Akt10Pt250.merge.NTUP_COMMON.e2328_s1581_s1586_r4485_r4540_p1606/
mc12_8TeV.190015.AlpgenPythia_Auto_P2011C_WmunuNp5_Akt10Pt250.merge.NTUP_COMMON.e2328_s1581_s1586_r4485_r4540_p1606/
mc12_8TeV.190021.AlpgenPythia_Auto_P2011C_WtaunuNp1_Akt10Pt250.merge.NTUP_COMMON.e2329_s1581_s1586_r4485_r4540_p1606/
mc12_8TeV.190022.AlpgenPythia_Auto_P2011C_WtaunuNp2_Akt10Pt250.merge.NTUP_COMMON.e2329_s1581_s1586_r4485_r4540_p1606/
mc12_8TeV.190023.AlpgenPythia_Auto_P2011C_WtaunuNp3_Akt10Pt250.merge.NTUP_COMMON.e2329_s1581_s1586_r4485_r4540_p1606/
mc12_8TeV.190024.AlpgenPythia_Auto_P2011C_WtaunuNp4_Akt10Pt250.merge.NTUP_COMMON.e2329_s1581_s1586_r4485_r4540_p1606/
mc12_8TeV.190025.AlpgenPythia_Auto_P2011C_WtaunuNp5_Akt10Pt250.merge.NTUP_COMMON.e2329_s1581_s1586_r4485_r4540_p1606/
mc12_8TeV.190050.AlpgenPythia_Auto_P2011C_WbbNp0_Akt10Pt250.merge.NTUP_COMMON.e2327_s1581_s1586_r4485_r4540_p1606/
mc12_8TeV.190051.AlpgenPythia_Auto_P2011C_WbbNp1_Akt10Pt250.merge.NTUP_COMMON.e2327_s1581_s1586_r4485_r4540_p1606/
mc12_8TeV.190052.AlpgenPythia_Auto_P2011C_WbbNp2_Akt10Pt250.merge.NTUP_COMMON.e2327_s1581_s1586_r4485_r4540_p1606/
mc12_8TeV.190053.AlpgenPythia_Auto_P2011C_WbbNp3_Akt10Pt250.merge.NTUP_COMMON.e2327_s1581_s1586_r4485_r4540_p1606/
mc12_8TeV.190040.AlpgenPythia_Auto_P2011C_WccNp0_Akt10Pt250.merge.NTUP_COMMON.e2327_s1581_s1586_r4485_r4540_p1606/
mc12_8TeV.190041.AlpgenPythia_Auto_P2011C_WccNp1_Akt10Pt250.merge.NTUP_COMMON.e2327_s1581_s1586_r4485_r4540_p1606/
mc12_8TeV.190042.AlpgenPythia_Auto_P2011C_WccNp2_Akt10Pt250.merge.NTUP_COMMON.e2327_s1581_s1586_r4485_r4540_p1606/
mc12_8TeV.190043.AlpgenPythia_Auto_P2011C_WccNp3_Akt10Pt250.merge.NTUP_COMMON.e2327_s1581_s1586_r4485_r4540_p1606/
mc12_8TeV.190030.AlpgenPythia_Auto_P2011C_WcNp0_Akt10Pt250.merge.NTUP_COMMON.e2327_s1581_s1586_r4485_r4540_p1606/
mc12_8TeV.190031.AlpgenPythia_Auto_P2011C_WcNp1_Akt10Pt250.merge.NTUP_COMMON.e2327_s1581_s1586_r4485_r4540_p1606/
mc12_8TeV.190032.AlpgenPythia_Auto_P2011C_WcNp2_Akt10Pt250.merge.NTUP_COMMON.e2327_s1581_s1586_r4485_r4540_p1606/
mc12_8TeV.190033.AlpgenPythia_Auto_P2011C_WcNp3_Akt10Pt250.merge.NTUP_COMMON.e2327_s1581_s1586_r4485_r4540_p1606/
mc12_8TeV.190034.AlpgenPythia_Auto_P2011C_WcNp4_Akt10Pt250.merge.NTUP_COMMON.e2327_s1581_s1586_r4485_r4540_p1606/

Table F.3: W+jets Akt10 filtered samples.

mc12_8TeV.147105.AlpgenPythia_Auto_P2011C_ZeeNp0.merge.NTUP_COMMON.e1879_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.147106.AlpgenPythia_Auto_P2011C_ZeeNp1.merge.NTUP_COMMON.e1879_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.147107.AlpgenPythia_Auto_P2011C_ZeeNp2.merge.NTUP_COMMON.e1879_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.147108.AlpgenPythia_Auto_P2011C_ZeeNp3.merge.NTUP_COMMON.e1879_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.147109.AlpgenPythia_Auto_P2011C_ZeeNp4.merge.NTUP_COMMON.e1879_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.147110.AlpgenPythia_Auto_P2011C_ZeeNp5incl.merge.NTUP_COMMON.e1879_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.147113.AlpgenPythia_Auto_P2011C_ZmumuNp0.merge.NTUP_COMMON.e1880_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.147114.AlpgenPythia_Auto_P2011C_ZmumuNp1.merge.NTUP_COMMON.e1880_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.147115.AlpgenPythia_Auto_P2011C_ZmumuNp2.merge.NTUP_COMMON.e1880_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.147116.AlpgenPythia_Auto_P2011C_ZmumuNp3.merge.NTUP_COMMON.e1880_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.147117.AlpgenPythia_Auto_P2011C_ZmumuNp4.merge.NTUP_COMMON.e1880_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.147118.AlpgenPythia_Auto_P2011C_ZmumuNp5incl.merge.NTUP_COMMON.e1880_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.147121.AlpgenPythia_Auto_P2011C_ZtautauNp0.merge.NTUP_COMMON.e1881_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.147122.AlpgenPythia_Auto_P2011C_ZtautauNp1.merge.NTUP_COMMON.e1881_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.147123.AlpgenPythia_Auto_P2011C_ZtautauNp2.merge.NTUP_COMMON.e1881_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.147124.AlpgenPythia_Auto_P2011C_ZtautauNp3.merge.NTUP_COMMON.e1881_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.147125.AlpgenPythia_Auto_P2011C_ZtautauNp4.merge.NTUP_COMMON.e1881_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.147126.AlpgenPythia_Auto_P2011C_ZtautauNp5incl.merge.NTUP_COMMON.e1881_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.200332.AlpgenPythia_Auto_P2011C_ZeebbNp0.merge.NTUP_COMMON.e2384_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.200333.AlpgenPythia_Auto_P2011C_ZeebbNp1.merge.NTUP_COMMON.e2384_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.200334.AlpgenPythia_Auto_P2011C_ZeebbNp2.merge.NTUP_COMMON.e2384_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.200335.AlpgenPythia_Auto_P2011C_ZeebbNp3incl.merge.NTUP_COMMON.e2384_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.200340.AlpgenPythia_Auto_P2011C_ZmumubbNp0.merge.NTUP_COMMON.e2385_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.200341.AlpgenPythia_Auto_P2011C_ZmumubbNp1.merge.NTUP_COMMON.e2385_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.200342.AlpgenPythia_Auto_P2011C_ZmumubbNp2.merge.NTUP_COMMON.e2385_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.200343.AlpgenPythia_Auto_P2011C_ZmumubbNp3incl.merge.NTUP_COMMON.e2385_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.200348.AlpgenPythia_Auto_P2011C_ZtautaubbNp0.merge.NTUP_COMMON.e2386_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.200349.AlpgenPythia_Auto_P2011C_ZtautaubbNp1.merge.NTUP_COMMON.e2386_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.200350.AlpgenPythia_Auto_P2011C_ZtautaubbNp2.merge.NTUP_COMMON.e2386_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.200351.AlpgenPythia_Auto_P2011C_ZtautaubbNp3incl.merge.NTUP_COMMON.e2386_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.200432.AlpgenPythia_Auto_P2011C_ZeeccNp0.merge.NTUP_COMMON.e2384_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.200433.AlpgenPythia_Auto_P2011C_ZeeccNp1.merge.NTUP_COMMON.e2384_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.200434.AlpgenPythia_Auto_P2011C_ZeeccNp2.merge.NTUP_COMMON.e2384_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.200435.AlpgenPythia_Auto_P2011C_ZeeccNp3incl.merge.NTUP_COMMON.e2384_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.200440.AlpgenPythia_Auto_P2011C_ZmumuccNp0.merge.NTUP_COMMON.e2385_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.200441.AlpgenPythia_Auto_P2011C_ZmumuccNp1.merge.NTUP_COMMON.e2385_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.200442.AlpgenPythia_Auto_P2011C_ZmumuccNp2.merge.NTUP_COMMON.e2385_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.200443.AlpgenPythia_Auto_P2011C_ZmumuccNp3incl.merge.NTUP_COMMON.e2385_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.200448.AlpgenPythia_Auto_P2011C_ZtautauccNp0.merge.NTUP_COMMON.e2386_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.200449.AlpgenPythia_Auto_P2011C_ZtautauccNp1.merge.NTUP_COMMON.e2386_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.200450.AlpgenPythia_Auto_P2011C_ZtautauccNp2.merge.NTUP_COMMON.e2386_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.200451.AlpgenPythia_Auto_P2011C_ZtautauccNp3incl.merge.NTUP_COMMON.e2386_s1581_s1586_r3658_r3549_p1575/

Table F.4: Z+jets samples.

mc12_8TeV.119353.MadGraphPythia_AUET2BCTEQ6L1_ttbarW.merge.NTUP_COMMON.e1352_s1499_s1504_r3658_r3549_p1575/
mc12_8TeV.119354.MadGraphPythia_AUET2BCTEQ6L1_ttbarWj.merge.NTUP_COMMON.e1352_s1499_s1504_r3658_r3549_p1575/
mc12_8TeV.119355.MadGraphPythia_AUET2BCTEQ6L1_ttbarZ.merge.NTUP_COMMON.e1352_s1499_s1504_r3658_r3549_p1575/
mc12_8TeV.119356.MadGraphPythia_AUET2BCTEQ6L1_ttbarZj.merge.NTUP_COMMON.e1352_s1499_s1504_r3658_r3549_p1575/
mc12_8TeV.110090.PowhegPythia_P2011C_singletop_tchan_lept_top.merge.NTUP_COMMON.e2575_s1773_s1776_r4485_r4540_p1575/
mc12_8TeV.110091.PowhegPythia_P2011C_singletop_tchan_lept_antitop.merge.NTUP_COMMON.e2575_s1773_s1776_r4485_r4540_p1575/
mc12_8TeV.110119.PowhegPythia_P2011C_st_schan_lep.merge.NTUP_COMMON.e1720_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.110140.PowhegPythia_P2011C_st_Wtchan_incl_DR.merge.NTUP_COMMON.e1743_s1581_s1586_r3658_r3549_p1575/
mc12_8TeV.183585.Sherpa_CT10_ZWtoeeqq_MassiveCB.merge.NTUP_COMMON.e2370_s1581_s1586_r4485_r4540_p1575/
mc12_8TeV.183586.Sherpa_CT10_ZZtoeeqq_MassiveCB.merge.NTUP_COMMON.e2370_s1581_s1586_r4485_r4540_p1575/
mc12_8TeV.183587.Sherpa_CT10_ZWtomumuqq_MassiveCB.merge.NTUP_COMMON.e2370_s1581_s1586_r4485_r4540_p1575/
mc12_8TeV.183588.Sherpa_CT10_ZZtomumuqq_MassiveCB.merge.NTUP_COMMON.e2370_s1581_s1586_r4485_r4540_p1575/
mc12_8TeV.183589.Sherpa_CT10_ZWtotautauqq_MassiveCB.merge.NTUP_COMMON.e2370_s1581_s1586_r4485_r4540_p1575/
mc12_8TeV.183590.Sherpa_CT10_ZZtotautauqq_MassiveCB.merge.NTUP_COMMON.e2370_s1581_s1586_r4485_r4540_p1575/
mc12_8TeV.183734.Sherpa_CT10_WWtoenuqq_MassiveCB.merge.NTUP_COMMON.e2347_s1581_s1586_r4485_r4540_p1575/
mc12_8TeV.183735.Sherpa_CT10_WZtoenuqq_MassiveCB.merge.NTUP_COMMON.e2347_s1581_s1586_r4485_r4540_p1575/
mc12_8TeV.183736.Sherpa_CT10_WWtomunuqq_MassiveCB.merge.NTUP_COMMON.e2347_s1581_s1586_r4485_r4540_p1575/
mc12_8TeV.183737.Sherpa_CT10_WZtomunuqq_MassiveCB.merge.NTUP_COMMON.e2347_s1581_s1586_r4485_r4540_p1575/
mc12_8TeV.183738.Sherpa_CT10_WWtotaunuqq_MassiveCB.merge.NTUP_COMMON.e2347_s1581_s1586_r4485_r4540_p1575/
mc12_8TeV.183739.Sherpa_CT10_WZtotaunuqq_MassiveCB.merge.NTUP_COMMON.e2347_s1581_s1586_r4485_r4540_p1575/

Table F.5: $t\bar{t}$+V, Diboson and Single Top samples.

Type	ID sample	$\sigma(pb)$
$t\bar{t}$	110404	114.47
	117050	114.51
	105200	112.94
	105860	115.56
	105861	114.46
	117209	59.624
	117210	59.622
	117840	123.27
	117842	106.46
	110898	0.1823
Diboson	183585	1.4648
	183586	0.2467
	183587	1.4634
	183588	0.2476
	183589	1.4523
	183590	0.2417
	183734	7.2854
	183735	1.9036
	183736	7.2974
	183737	1.9057
	183738	7.2741
	183739	1.9152
$t\bar{t}+V$	119353	0.1041
	119354	0.0933
	119355	0.0677
	119356	0.0873
SingleTop	110090	17.519
	110091	9.3964
	110119	1.6424
	110140	20.461
slices	110395	0.7395
	110396	0.2593
	110397	0.0981
	110398	0.0496
	110399	0.0203

Type	ID sample	$\sigma(pb)$
$W+jets$	190050	0.0125
	190051	0.1198
	190052	0.2825
	190053	0.7321
	190040	0.0133
	190041	0.2244
	190042	0.6919
	190043	1.7859
	190030	0.0875
	190031	0.4722
	190032	0.5699
	190033	0.3791
	190034	0.2991
	190001	0.7157
	190002	1.9920
	190003	2.2144
	190004	1.4867
	190005	1.1185
	190011	0.7064
	190012	1.9221
	190013	2.1249
	190014	1.4169
	190015	1.0612
	190021	0.7047
	190022	1.9309
	190023	2.1416
	190024	1.4297
	190025	1.0705

Table F.6: Cross-sections related to the MC samples employed in this analysis.

Type	ID sample	$\sigma(pb)$
$Z + jets$	147105	718.97
	147106	175.70
	147107	58.875
	147108	15.636
	147109	4.0116
	147110	1.2592
	147113	719.16
	147114	175.74
	147115	58.882
	147116	15.673
	147117	4.0057
	147118	1.2544
	147121	718.87
	147122	175.76
	147123	58.856
	147124	15.667
	147125	4.0121
	147126	1.2560

Type	ID sample	$\sigma(pb)$
$Z + jets$	200332	6.5083
	200333	3.2927
	200334	1.2544
	200335	0.6171
	200340	6.5056
	200341	3.2904
	200342	1.2601
	200343	0.6188
	200348	6.5062
	200349	3.2935
	200350	1.2485
	200351	0.6136
	200432	11.763
	200433	7.1280
	200434	3.3603
	200435	1.7106
	200440	11.795
	200441	7.1123
	200442	3.3708
	200443	1.7059
	200448	11.760
	200449	7.1410
	200450	3.3582
	200451	1.7046

Table F.7: Cross-sections related to the MC samples employed in this analysis.

Física para todos

Para comenzar este pequeño capítulo, vamos a ver cuáles son las definiciones de **física** y de **física de partículas**. Según la Real Academia Española (RAE), la **física** es la *ciencia que estudia las propiedades de la materia y de la energía, y las relaciones entre ambas.* Además, es una ciencia teórica y experimental, y busca que sus conclusiones puedan ser verificables mediante experimentos y que la teoría pueda realizar predicciones de experimentos futuros basados en observaciones previas. Según la Wikipedia, la **física de partículas** es la *rama de la física que estudia los componentes elementales* [1] *de la materia y las interacciones entre ellos*, y también se la conoce como **física de altas energías**.

Importancia de la Física de Partículas en la sociedad

A menudo la gente se pregunta para qué sirve la física. Muchos piensan que los físicos no hacen nada productivo para la sociedad. Para muchas personas es difícil hacerse una idea de cuáles son las tareas de los físicos, y otros piensan que simplemente se divierten haciendo chocar partículas y diciendo que van a descubrir "cosas" nuevas.

Pues bien, la física de partículas ha contribuido (y lo sigue haciendo) a una lista muy larga y creciente de aplicaciones prácticas, beneficiosa para la sociedad. A continuación, vamos a ver algunos de los ejemplos [201] más relevantes.

[1] Componente elemental se refiere a un constituyente de la materia que no está compuesto por partículas más pequeñas ni se conoce que tenga estructura interna.

Medicina

Posiblemente, uno de las creaciones más importantes para la salud sea la del "ciclotrón", que es un acelerador circular de partículas. En 1930 el padre de los aceleradores de partículas, Ernest O. Lawrence, construyó el primer ciclotrón en Berkeley (California) y recibió el Premio Nobel en 1939 por este trabajo y sus aplicaciones. Lo que no sabe mucha gente, es que los primeros ciclotrones se empleaban para producir isótopos médicos para la investigación. En 1938, la madre de Lawrence se convirtió en el primer paciente con cáncer en ser tratado exitosamente con las partículas de ciclotrones. Ella fue la primera de millones de personas que se benefician de este descubrimiento de la física. Gracias a este avance, los médicos actuales utilizan haces de partículas para el diagnóstico y la curación de millones de pacientes con cáncer. Además y gracias a la física de partículas, se ha avanzado mucho en la radioterapia, en la hadrónterapia y en el diagnóstico por imagen médica.

Informática y computación

Una de las creaciones más revolucionarias del siglo pasado y que todos usamos diariamente es la www (World Wide Web), también conocida como La Web o red informática mundial. Allá por el año 1990, Tim Berners-Lee y Robert Cailliau crearon La Web para resolver el problema de la comunicación en una colaboración internacional de miles de físicos de partículas, mientras trabajaban en el CERN. Además, los físicos de partículas han desarrollado tecnologías de computación de vanguardia para poder grabar y analizar el volumen sin precedentes de los datos generados en las colisiones de alta energía, haciendo contribuciones fundamentales a las soluciones en las fronteras de la informática.

Industria

Los físicos de partículas cuentan con la ayuda de la industria para producir y mejorar los millones de componentes que requieren los experimentos, haciendo que las empresas puedan crear/vender nuevos productos y tecnologías que cambian la vida cotidiana.

El LHC usa tecnología puntera para enfriar y conducir partículas cargadas: los sistemas de criogenia y superconductividad más grandes del planeta. Estos sistemas (a menor escala) se emplean a diario en maquinaria pesada, ordenadores o móviles, por ejemplo.

El Modelo Estándar de la Física

Como ya sabemos, el cuerpo humano está hecho de carne y hueso, pero ¿sabemos de qué está formado ese "carne y hueso"? Veamos qué quiere decirnos la Figura 1. Si tuviéramos una lupa más potente que las convencionales, veríamos que nuestro cuerpo está formado por moléculas. Si cogiéramos otra lupa aún más potente observaríamos que las moléculas son conjuntos de átomos. Los átomos están formados por una corteza de electrones y un núcleo de protones y neutrones. Otra lupa súper-potente nos mostraría que los protones y neutrones están formados por quarks. De momento, no tenemos lupas mega-potentes para ver si dentro de los quarks y de los electrones hay más cosas, así que consideremos que estas partículas son elementales, es decir, no compuestas.

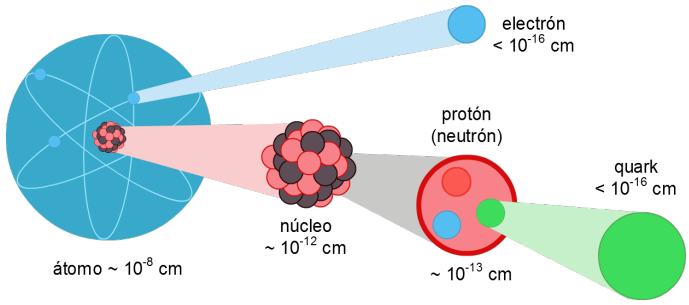


Figura 1: Componentes de la materia común.

Hasta aquí todo está claro, pero ¿cómo se mantienen unidas las moléculas para formar la materia? ¿Y los protones y neutrones para formar los núcleos atómicos? En Física de Partículas, las **fuerzas fundamentales** (llamadas también interacciones fundamentales) desempeñan el mismo papel que el pegamento en nuestra vida cotidiana. Gracias a ellas las partículas se mantienen unidas.

En la naturaleza existen cuatro fuerzas fundamentales que se manifiestan mediante campos de ondas que ejercen efectos sobre las partículas, sobre los cuerpos e incluso sobre la energía. Además, todas las ondas se pueden estudiar como partículas y viceversa. Por tanto las cuatro interacciones también se manifiestan mediante un intercambio de partículas:

- **Gravedad**: aunque es la que más sentimos diariamente, puesto que gracias a ella estamos pegados a la Tierra, es la que menos afecta a las partículas.
- **Electromagnética**: los imanes funcionan gracias a esta fuerza. Además es la responsable de fenómenos presentes en nuestra vida diaria, como

la propagación de la luz, la corriente eléctrica o las señales de radio y televisión.

- **Nuclear débil**: esta fuerza la encontramos en fenómenos radiactivos de tipo beta, que no son otra cosa que desintegraciones de partículas y núcleos atómicos.
- **Nuclear fuerte**: es la fuerza que obliga a los núcleos atómicos a permanecer unidos.

Existe una teoría que clasifica todas las partículas elementales y describe las fuerzas fundamentales casi a la perfección. Se llama el Modelo Estándar (SM). Esta teoría llama **fermiones** a las partículas elementales que constituyen la materia conocida. Además, dice que todas las fuerzas fundamentales, mediante las cuales interactúan las partículas, están relacionadas con **bosones**, que son otro tipo de partículas. Esta agrupación puede observarse en la Figura 2.

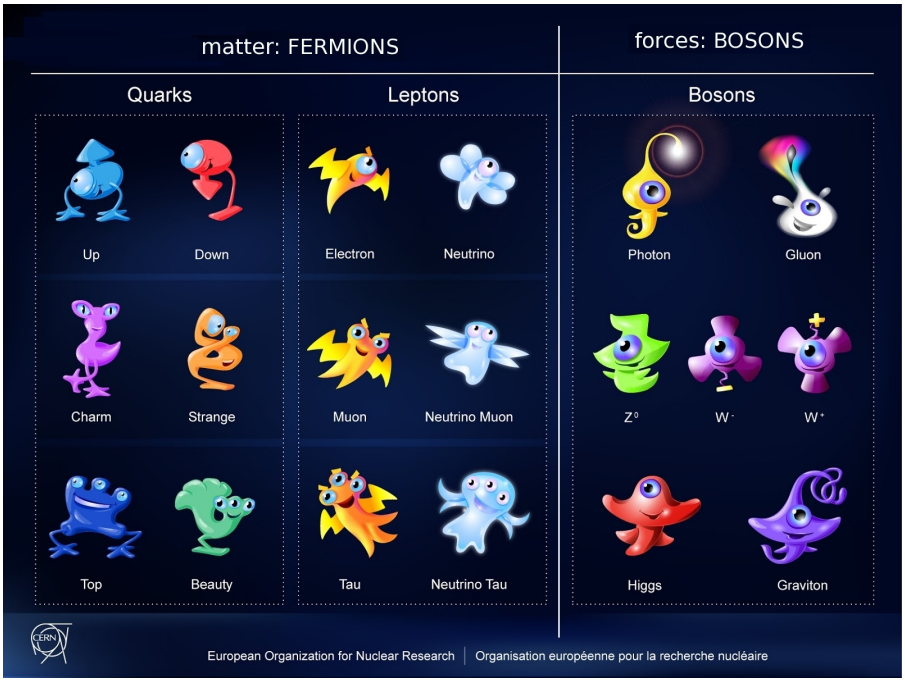


Figura 2: Clasificación de las partículas y fuerzas fundamentales según el Modelo Estándar.

Según el SM, las partículas de materia y las relacionadas con las fuerzas fundamentales, se clasifican según sus propiedades intrínsecas (como la masa, la carga eléctrica o el espín) de la siguiente forma:

- **Fermiones**: Son las partículas materiales y siempre tienen una antipartícula asociada con las mismas características pero con carga eléctrica opuesta (o nula en el caso de partículas neutras). Su espín es siempre 1/2.
  - **Quarks**: tienen carga eléctrica, la cual les permite participar en la interacción electromagnética. También tienen una carga de color (que puede ser roja, azul y verde, definidas por conveniencia) que permite describir cómo interactúan mediante interacción fuerte.
  - **Leptones**: los leptones cargados tienen carga eléctrica y carga de sabor, con lo cual, participan en interacciones electromagnéticas. Los neutrinos sólo interaccionan débilmente (sólo tiene carga de sabor), por eso son las partículas más difíciles de detectar.

- **Bosones**: son partículas que interaccionan con las partículas materiales. Siempre tienen un espín de valor entero.
  - **g (gluón)**: es la partícula portadora de la interacción fuerte. No tiene masa ni carga eléctrica pero sí que posee carga de color, por lo que además de transmitir la interacción fuerte también la sufre.
  - γ **(fotón)**: es la partícula responsable de las manifestaciones electromagnéticas y la portadora de todas las formas de radiación electromagnética. No tiene masa ni carga eléctrica ni carga de color, pero presenta propiedades tanto corpusculares como ondulatorias ("dualidad onda-corpúsculo").
  - $\mathbf{W^{\pm}}$ **y** $\mathbf{Z^0}$: son las partículas mediadoras de la interacción débil (también conocida como interacción nuclear débil). No poseen carga de color pero son muy masivos. El bosón Z^0 no tiene carga eléctrica pero los $W^{\pm}$ sí.
  - **H (higgs)**: no es portador de ninguna interacción pero es esencial para explicar los orígenes de la masa de otras partículas elementales. Tiene espín 0.
  - **G (gravitón)**: es un bosón hipotético, que sería el transmisor de la interacción gravitatoria.

El SM es muy importante para los físicos de partículas, puesto que explica detalladamente, con precisión y casi a la perfección cómo se comportan las partículas elementales cuando interaccionan entre ellas o con una fuerza fundamental.

Física más allá del Modelo Estándar

Como ya hemos dicho anteriormente, el SM es una teoría que describe con precisión muchos fenómenos físicos, pero no es capaz (por ahora) de explicar algunos conceptos ni algunos resultados experimentales. Este hecho hace que los físicos quieran buscar una teoría más completa logrando una "Teoría de Campos Unificada" o una "Teoría del Todo", que explique todos los fenómenos físicos además de incluir la Gravedad y explicar las constantes. Todas las teorías o modelos que son capaces de explicar fenómenos fuera del SM, están dentro de la denominada física más allá del Modelo Estándar (BSM).

Describir en un mismo modelo la teoría cuántica usada para describir el mundo microscópico y la teoría general de la relatividad usada para describir el mundo macroscópico, es muy difícil, por eso la **Gravedad** no está incluida en el SM. Puesto que la Gravedad es mucho más débil que las otras tres fuerzas fundamentales (esto es el denominado **problema de jerarquía gauge**), su influencia en el comportamiento de las partículas elementales es casi despreciable. La Gravedad es la fuerza predominante en el mundo macroscópico y siempre es positiva.

La **materia oscura** es un tipo de materia que no puede detectarse directamente, pero su existencia puede inferirse a partir del empuje gravitacional o los efectos que ésta ejerce en el material luminoso que podemos ver. Se supone que la materia oscura forma el 85 % de toda la materia de nuestra galaxia y el 25 % del universo. Aunque el SM no tiene un buen candidato, muchos aspirantes a materia oscura han sido sugeridos. Actualmente, la materia oscura es uno de los mayores retos para la física moderna.

La expansión del universo se está acelerando. A la presión que causa este fenómeno se la denomina **energía oscura**. Esta forma desconocida de energía contribuye el 70 % del total de la energía en el universo.

Otro fenómeno que no tiene explicación en el SM es la **masa de los neutrinos**. Muchos experimentos de oscilación de neutrinos han mostrado que los neutrinos pueden cambiar su sabor, oscilando a medida que ellos viajan en el espacio. Para que sea posible la oscilación de neutrinos, ellos deben tener masas ligeramente diferentes. Lo que conlleva a que al menos dos de los tres existentes deben tener masa no nula, de forma contraria a lo que predice el SM.

Otra pregunta sin respuesta es por qué la materia domina sobre la antimateria, conocida como la **asimetría bariónica**. Una posible hipótesis que explique esta asimetría es que alguna ley física haya actuado sobre la materia y la antimateria de forma diferente, o sólo existe para una de las dos.

El LHC y ATLAS en el CERN

En física de partículas no tenemos lupas mega-potentes para ver qué hay dentro de las partículas ni el rastro (traza) que dejan. En su lugar tenemos aceleradores de partículas que desempeñan el mismo papel que dichas lupas.

En el CERN (siglas que provienen del francés, *Conseil Européen pour la Reserche Nucléaire*), se encuentra el mayor acelerador y colisionador de partículas construido hasta el momento. Está ensamblado en un túnel circular de 27 km de circunferencia, que está enterrado 100 m bajo la frontera franco-suiza, en Ginebra, como muestra la Figura 3:

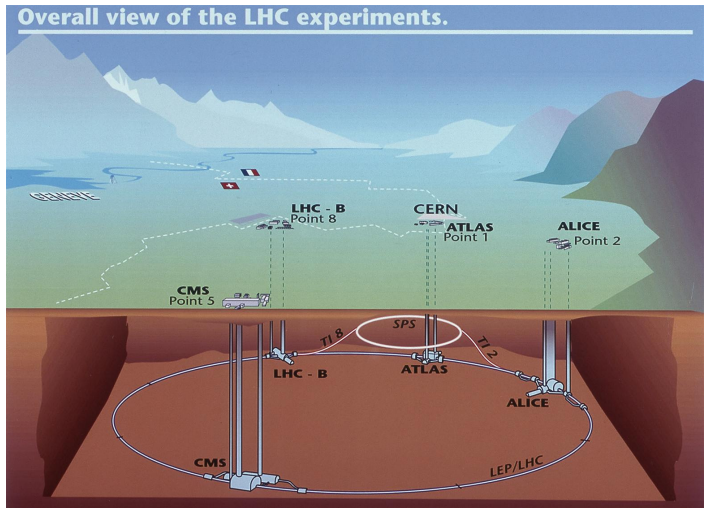


Figura 3: Simulacióin real de dónde están localizados el LHC y sus experimentos.

El LHC está diseñado para colisionar dos haces de protones que circulan en sentidos opuestos, con una energía de 7 TeV [2] cada uno. Los protones viajan a través del túnel casi a la velocidad de la luz y chocan en cuatro puntos diferentes, donde se encuentran cuatro detectores gigantes de partículas.

El LHC, como cualquier acelerador de partículas, tiene tres componentes fundamentales para su buen funcionamiento: los tubos por los que circula

[2] El electronvoltio, eV, es una unidad de energía que representa la variación de energía potencial que experimenta un electrón al moverse entre dos puntos cuya diferencia de potencial es de 1 voltio, y equivale a $1,602176565 \cdot 10^{-19}$ J, valor que se obtiene al multiplicar la carga del electrón $1,602176565 \cdot 10^{-19}$ C por la unidad de potencial eléctrico V. Para hacernos una idea, 1 TeV equivale a la energía cinética de un mosquito volando a una velocidad de 0,4 m/s.

el haz (beam pipes), los elementos de aceleración y el sistema de imanes de desviación. Estos elementos pueden observarse en la Figura 4. Las partículas

Figura 4: Elementos fundamentales de un acelerador de partículas.

van circulando por los tubos y nunca se salen porque los campos magnéticos que proporcionan los imanes de enfoque evitan que esto ocurra. Cada vez que pasan por las cavidades de aceleración, las partículas son aceleradas por un campo eléctrico muy potente. Gracias a los imanes de desviación, las trayectorias de las partículas cargadas se van curvando. Este conjunto de elementos permite a los haces de partículas ir dando vueltas cada vez más rápido.

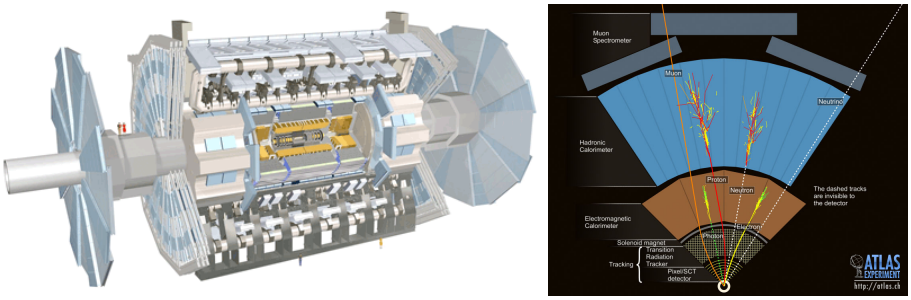
El punto de partida del viaje de los protones a lo largo del túnel (o tubos) es una bombona de hidrógeno gaseoso y el destino final, los detectores. De un átomo de hidrógeno, que está compuesto por un protón y un electrón, se elimina el electrón mediante descargas eléctricas y el protón restante, se envía al sistema de aceleración del LHC mediante campos electromagnéticos. Esto se realiza con todos los átomos de hidrógeno que hay dentro de la bombona, consiguiendo aproximadamente unos 400 millones de colisiones por segundo.

Estas colisiones dan lugar a todas las partículas predichas por el SM y posiblemente, a algunas nuevas. De un choque de dos protones pueden salir, por ejemplo, un quark top y un quark anti-top, o puede formarse un bosón de Higgs. Cada uno de estos procesos tiene una probabilidad determinada. La probabilidad de producir partículas muy pesadas (que son las que resultan más interesantes para los físicos) es muy baja porque muy pocas de las colisiones son suficientemente energéticas para producirlas (aproximadamente 1 entre 1000 millones para las colisiones más interesantes). Por este motivo es importante producir muchísimas colisiones para poder estudiar con precisión las propiedades de las partículas ya conocidas y posiblemente, poder observar las que aún están por descubrir.

Las colisiones y todas sus partículas resultantes (esto se denomina **evento** o suceso) se estudian usando grandes detectores capaces de reconstruir lo que ha sucedido en cada choque. En física de partículas no tenemos cámaras digitales potentes ni tridimensionales, pero en su lugar, tenemos detectores de partículas gigantes que desempeñan la misma función. Estos detectores se construyen en forma de capas, como si fuera una cebolla en forma de cilindro, y cada una de ellas se encarga de recoger información distinta, para poder hacer millones de "fotografías" desde diferentes perspectivas. Luego, esas fotos, se utilizan para reconstruir lo que posiblemente sucedió justo después de la colisión de los dos protones.

Vamos a explicar, en términos generales, cómo es y cómo funciona ATLAS, puesto que es el experimento que se trata en esta tesis. ATLAS, cuyo diseño se representa en la Figura 5(a), es un detector de carácter general construido para trabajar a alta luminosidad [3] y diseñado para rendir al máximo en medidas de precisión del SM además de usarse en búsquedas de nueva física. Junto a CMS, fue uno de los experimentos clave para el descubrimiento del bosón de Higgs, también conocida como la "partícula de Dios".

[3] En física de partículas se define la **luminosidad instantánea** $\mathcal{L}$ como el número de partículas en un haz por unidad de superficie y por unidad de tiempo. Se mide en unidades inversas de sección eficaz por unidad de tiempo, por ejemplo $cm^{-1}s^{-1}$. Al integrar esta cantidad durante un período se obtiene la **luminosidad integrada L**, la cual se mide en unidades inversas de sección eficaz, como por ejemplo fb^{-1}. Cuanto mayor es L, mayor es la probabilidad de que se produzcan sucesos interesantes en un experimento de altas energías.



(a) Diseño (b) Estructura

Figura 5: El experimento ATLAS.

ATLAS es el detector más grande instalado en el CERN, con una longitud de 44 m, un diámetro de 25 m y 7000 toneladas. Tiene un diseño cilíndrico y estructura de capas, como puede observarse en la Figura 5(b). Cada una de esas capas tiene una finalidad diferente:

- **Detector interno**: es el encargado de medir el momento de las partículas cargadas.
- **Los calorímetros**: miden las direcciones angulares y la energía de todas las partículas producidas en cada colisión. Son especialmente importantes para poder reconstruir electrones, fotones y jets, que son chorros de partículas que se forman a partir de quarks y de gluones.
- **Espectrómetro de muones**: es el encargado de identificar muones (éstos atraviesan todo el detector) y después, selecciona eventos con muones de alto momento.
- **Sistema de imanes**: es el encargado de curvar la trayectoria de las partículas cargadas.

Cada componente de ATLAS está conectado a un sistema de lectura de señales eléctricas. En cuanto se detecta un impulso eléctrico, el sistema registra el instante y el lugar exactos donde ha sucedido y toda la información se envía a un super-ordenador compuesto de miles de PCs. De forma casi instantánea y siguiendo unos criterios muy concretos, el ordenador graba sólo los eventos que son interesantes para su posterior análisis.

El modelo de computación de ATLAS

Figura 6: Altura de una torre de CDs.

Para poder recoger toda la información que el detector ATLAS registra, se necesitarían miles de ordenadores como los que tenemos en casa. En lugar de esto, se tienen centros de computación con computadoras muy potentes, capaces de grabar toda esta información y de forma casi instantánea.

Para que nos hagamos una idea de si esta información es "mucha o poca", durante el primer período de funcionamiento del LHC (Run-I) ATLAS recopiló 15 PB (15000 TB) de datos, equivalente a 5000 millones de canciones, más de 20 millones de películas o más de 1500 millones de libros. Este volumen de datos es suficiente para llenar más de tres millones de CDs (formarían una torre de más de 20 km de altura, como muestra la Figura 6) o quince mil discos duros externos, algo impresionante, ¿no crees? Pues el CERN ha desarrollado una infraestructura capaz de gestionar y albergar todos estos datos.

Miles de científicos y físicos de todo el mundo necesitan acceder de forma casi instantánea a esa información. Por ello, el CERN ha creado una infraestructura de computación distribuida y almacenamiento de datos, capaz de manipular esta gran cantidad de datos, conocido como **Grid**. éste incorpora tanto enlaces propios de fibra óptica como partes de Internet de alta velocidad. Para que nos hagamos una idea, el Grid es como el internet que tenemos en casa pero a lo grande, con tanta potencia que nos permitiría descargar cien películas simultáneamente en cuestión de minutos y a la vez que cientos de usuarios.

El proyecto de computación Grid Worldwide del LHC (WLCG) es una colaboración compuesta por más de 170 centros de computación distribuidos por 42 países, y que conecta infraestructuras Grid nacionales e internacionales. El principal objetivo del WLCG es construir y mantener una infraestructura de computación distribuida para almacenar, distribuir y analizar los datos que provienen de los experimentos del LHC.

Actualmente, el modelo de computación de ATLAS (ACM) define los requerimientos necesarios en la estructura del software y su funcionamiento. El ACM permite un alto grado de descentralización y la posibilidad de compartir recursos computacionales debido a su jerarquía de niveles. Estos niveles se denominan **Tiers** y son gestionados centralmente por la colaboración ATLAS. Los niveles de esta estructura son:

- Primer nivel: **Tier-0**. Sólo hay uno y se encuentra en el CERN, donde los datos brutos (los datos que recoge ATLAS justo después de cada colisión) se almacenan. Se realiza una primera reconstrucción de los eventos y se guarda una copia de ellos. Además, una copia secundaria de los datos brutos se distribuye en su Tier-1 asociado.
- Segundo nivel: **Tier-1**. Hay 11. Coordinan los datos almacenados permanentemente (datos brutos, simulados y procesados) y proporcionan capacidad computacional para el reprocesado y análisis de datos.
- Tercer nivel: **Tier-2**. Hay alrededor de 80 centros distribuidos por todo el mundo. Proporcionan capacidad computacional y de almacenamiento adecuado para la simulación de eventos Monte Carlo (MC) y necesario para el análisis final de datos.
- Cuarto nivel: **Tier-3**. Hay decenas. Este nivel es una instalación privada para los usuarios finales. Cada una está asociada con un Tier-2, aunque están fuera del proyecto LCG.

El conjunto de un Tier-1 y varios Tier-2s distribuidos en una periferia dada, se llama **Cloud**. Aunque actualmente todos los Tier-1s y Tier-2s están conectados entre sí y con el Tier-0, permitiendo así un uso más eficiente de los recursos de CPU y de disco, es importante remarcar que cada Cloud se encarga de mantener el buen funcionamiento de sus centros.

Los quarks más veloces

A diferencia del resto de partículas fundamentales, los quarks siempre viven acompañados de otras partículas y es muy difícil verlos solitariamente (este fenómeno se conoce como **libertad asintótica**). Cuando se produce un quark "libre", de inmediato éste busca compañía, generando una avalancha de otros hadrones y formando un chorro de partículas, que se denomina **jet**.

Debido a su elevada masa (equivale a la masa de un átomo de oro), el quark top se desintegra tan rápidamente que no le da tiempo a hadronizar y formar un jet. Esto lo hace especialmente interesante para los físicos de partículas.

El quark top se desintegra casi siempre (el 99 % de las veces) a un bosón W más un quark bottom. A su vez, el bosón W se puede desintegrar a dos jets ó a un leptón más un neutrino. Por lo tanto, los productos de desintegración de un quark top son tres: ó un quark bottom más dos jets (**desintegración hadrónica**) ó un quark bottom más un leptón más un neutrino (**desintegración leptónica**).

Si el quark top se produce casi en reposo, como ocurría en el Tevatrón (es otro acelerador de partículas, situado cerca de Chicago), los productos de su desintegración salen despedidos en direcciones distintas, así que incidirán en diferentes zonas del detector y es más fácil observarlos individualmente. En cambio, si el quark top se produce en una colisión muy energética (topología relativista ó *boosted*), los tres productos de su desintegración salen colimados, incidiendo en una zona muy pequeña del detector y haciendo muy complicado identificarlos. En este último caso, si el quark top se desintegra hadrónicamente, los tres jets se fusionan en uno más gordo (llamado **fat jet**). Por ello se necesitan nuevas técnicas o algoritmos para "reconstruir" lo que ha pasado en estos sucesos relativistas.

Reconstruir un suceso es algo así como averiguar lo que ha pasado en un suceso. No es una tarea fácil puesto que no vemos una sola partícula, sino un montón de partículas a la vez, y de entre todas ellas (usando técnicas avanzadas que no voy a nombrar), debemos identificar cuáles son las que queremos estudiar a fondo. En la Figura 7 podemos ver todas las partículas que se han reconstruido

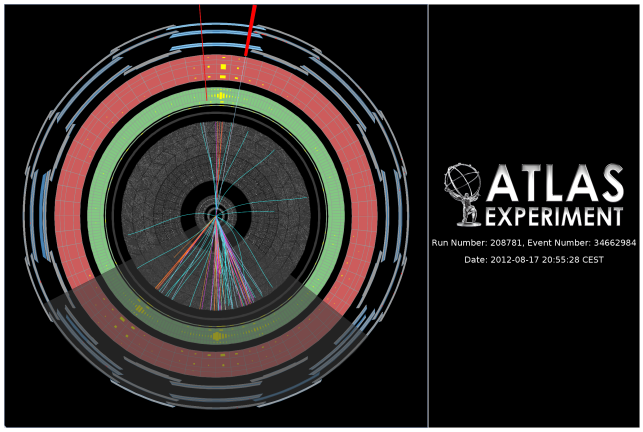


Figura 7: Evento reconstruido en el detector ATLAS.

en un choque de dos protones. Esas líneas de colores son los caminos hipotéticos que han tomado las partículas.

Estudiar este quark parece muy complicado, pero ¿merece la pena? Obviamente la respuesta es sí. Sí porque conocer de forma precisa sus propiedades tiene muchas aplicaciones, como la búsqueda de nuevas partículas masivas (denominadas también **resonancias**) o el estudio de la asimetría de carga, que es la esencia de esta tesis. ¿Y qué significa esto? Veámoslo a continuación.

¿Qué es un observable?

Un **observable** es todo aquello que se puede medir. Los más conocidos en nuestra vida cotidiana son, por ejemplo, posiciones, velocidades, masas o temperaturas. Además tenemos aparatos que los pueden medir directamente, como la regla, la balanza o el termómetro.

Algunos observables, sobre todo los relacionados con la física de partículas, no se pueden medir directamente con aparatos, y se necesita desarrollar técnicas para su medida.

¿Por qué son importantes los observables en física de partículas?

La contestación más directa es que, cuando se pretende descubrir nueva física, se debe **comparar** el observable que se mide con el valor que predice el SM para él. Si se mide el mismo valor, dentro de su error, se corrobora lo que predice el SM. En cambio, si se obtiene un valor diferente al del SM, se puede asegurar que se ha descubierto "algo", ya sea una partícula o una propiedad de ésta.

El ejemplo más famoso es el descubrimiento del Higgs. Se observó que los datos que recogieron los detectores ATLAS y CMS no coincidían del todo con lo que predecía el SM, como muestra la Figura 8. En esta figura, lo que predice el SM se muestra de color rojo y los puntos negros con barras de error es lo que se midió en el LHC. Se puede observar que a la izquierda, los puntos no se parecen a la distribución roja sino que coinciden con la distribución azul, que representa la contribución del bosón de Higgs a esta distribución.

Este estudio que se hizo con el Higgs y permitió descubrirlo, se puede hacer con cualquier otra nueva partícula hipotética. Este tipo de análisis se llama "búsqueda de resonancias", y es uno de los trabajos que más se repite en los experimentos, especialmente en el CERN.

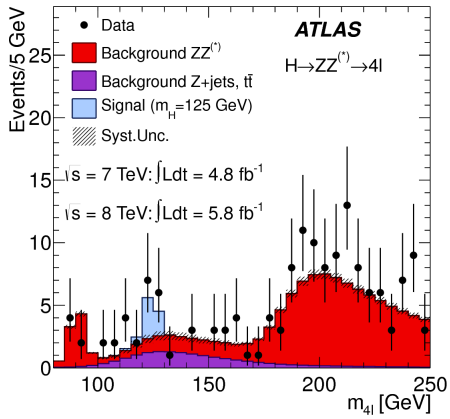


Figura 8: Comparación de la masa del Higgs predicha por el SM (rojo) y medida en el LHC (puntos negros con barras de error).

Dos observables, una tesis

En esta tesis se han estudiado detalladamente dos observables:

- **La masa de los pares $t\bar{t}$ en el LHC**: es la masa con la que se producen un quark top y un quark anti-top juntos.
- **La asimetría de carga $t\bar{t}$ con topologá relativista en el LHC**: está relacionada con el número de quarks top que se producen en los extremos del detector y el número de quarks anti-top que se producen en la zona central del detector.

Estos dos observables se han elegido con la esperanza de que puedan revelar señales de nueva física (desviaciones del SM). En caso negativo, pueden servir para descartar ciertas propuestas y centrarnos en otras extensiones del SM.

Resultados de esta tesis

Cuando hicimos la búsqueda de nuevas resonancias midiendo la masa de los pares $t\bar{t}$ en el LHC, no descubrimos ninguna partícula nueva. En cambio, pudimos decir qué valores de masa no son posibles para varias partículas pesadas hipotéticas que se desintegran a un par $t\bar{t}$. Aunque esto no sea un descubrimiento también es algo bueno, porque de alguna forma, ayudas a los físicos de partículas

a saber dónde deben buscar nueva física y dónde no. Esto es parecido a buscar un spiderman dentro de 1000 huevos sorpresa. Si abres los 100 primeros huevos y no encuentras la figurita, debes seguir buscando dentro de los otros 900. Y debes seguir buscando hasta abrir todos porque no sabes si te encontrarás un spiderman o diez. Está claro que la persona más afortunada será la que encuentre la figurita, pero los demás habrán ayudado a la causa.

Para estudiar la asimetría de carga $t\bar{t}$ con topología relativista en el LHC hemos desarrollado ciertas técnicas para hacer esta medida y que son imprescindibles en los estudios que se están haciendo durante el Run-II del LHC. La medida de la asimetría de carga $t\bar{t}$ con topología relativista (ha sido la primera vez que se ha medido esta asimetría con esta topología) ha mostrado un valor muy parecido al que predecía el SM. Por lo tanto, confirmamos la predicción del SM dentro del error.

Conclusiones

Con este pequeño resumen he querido acercar un poco la física, y en concreto la física de partículas, a todas aquellas personas que no están relacionadas con este campo.

Simplemente quería demostrar que la física sí que es útil para la sociedad, al igual que el resto de ramas de la ciencia e ingeniería, y que si te la explican a tu nivel, no es tan difícil como el mundo cree.

La ciencia y la tecnología de la física de partículas tienen aplicaciones transformadoras que se traducen en beneficios para muchas otras áreas de la sociedad. Con todo esto queda demostrado que la física en general sí que aporta y sigue aportando "cosas" positivas para la sociedad, que benefician a gran parte de la población y que mucha gente usa en su vida cotidiana.

Contributions to the ATLAS Collaboration

The ATLAS Collaboration consists of more than 3000 members. Every article is the result of the work performed by a group of people. In this chapter I list my main contributions to the ATLAS experiment, including internal communications, papers, posters and proceedings.

INTERNAL COMMUNICATIONS

⋙ In the internal note **"A charge asymmetry measurement for high mass $t\bar{t}$ pairs"** I was one of the seven authors. In this note we performed a MC study of the charge asymmetry in $t\bar{t}$ production at large invariant mass ($m_{t\bar{t}>800\ GeV}$). We presented several quantitative arguments that show that this measurement has excellent sensitivity to certain classes of new physics models and s intrinsically robust to modeling errors (these were the prospects for my thesis). Also we discussed the measurement of the charge asymmetry of leptons in $t\bar{t}$ events. (ATL-COM-PHYS-2012-786)

⋙ With two coauthors, we wrote **"Resonance searches in top pair production: a guide to understanding the ATLAS and CMS benchmarks"**. We presented a brief guide to the ATLAS and CMS benchmark models used generally to interpret the results of the $t\bar{t}$ resonance searches. (ATL-COM-PHYS-2013-1040)

COMPUTING

★ I presented the poster **"Lessons learned from the ATLAS performance studies of the Iberian Cloud for the first LHC running period."** in the 20th International Conference on Computing in High Energy and Nuclear Physics 2013. This was a summary of the ATLAS Iberian Cloud performance during the Run-1.

- ▷ Internal communication proceeding: ATL-COM-SOFT-2013-026;
- ▷ Internal communication poster: ATL-COM-SOFT-2013-025;
- ▷ Approved proceeding: ATL-SOFT-PROC-2013-007;
- ▷ Approved poster: ATL-SOFT-SLIDE-2013-765;

★ I was one of the main authors of the poster **"Spanish ATLAS Tier-2 facing up to Run-2 period of LHC."**, presented in the 21st International Conference on Computing in High Energy and Nuclear Physics 2015.

- ▷ Internal communication proceeding: ATL-COM-SOFT-2015-064;
- ▷ Internal communication poster: ATL-COM-SOFT-2015-002;
- ▷ Approved proceeding: ATL-SOFT-PROC-2015-005;
- ▷ Approved poster: ATL-SOFT-SLIDE-2015-071;

PHYSICS

⋙ In **"A search for $t\bar{t}$ resonances in the lepton plus jets final state with ATLAS using $4.7fb^{-1}$ of pp collisions at $\sqrt{s} = 7\ TeV$"** I produced the official Kaluza-Klein gluon (g_{KK}) samples used in this paper to perform the exclusion studies at $\sqrt{s} = 7\ TeV$.

- ▷ 4.7 fb^{-1} Backup note:ATL-COM-PHYS-2012-1753;
- ▷ 4.66 fb^{-1} Internal communication: ATL-COM-PHYS-2012-1356;
- ▷ 5 fb^{-1} Backup note: ATL-COM-PHYS-2012-797;
- ▷ ATLAS Conference Note: ATLAS-CONF-2012-136;
- ▷ Public paper arXiv:1305.2756;
- ▷ Published paper CERN-PH-EP-2013-032, Phys. Rev. D 88 (2013) 012004;

⋙ In **"A search for $t\bar{t}$ resonances using lepton plus jets events in proton-proton collisions at $\sqrt{s} = 8\ TeV$ with the ATLAS detector"** I produced the official Kaluza-Klein gluon (g_{KK}) samples used in this paper to perform the exclusion studies at $\sqrt{s} = 8\ TeV$. One of my supervisors and I extended the interpretation of broad g_{KK}. Also I worked on the *TopRootCore-AnalysisTop* package and I made it compatible with the code used in this note. I cross-checked the table of yield with several members of this author list. In August 2015 it was accepted by the Journal of Hight Energy Physics (JHEP).

- ▷ 14 fb^{-1} Internal communication: ATL-COM-PHYS-2013-207;
- ▷ 14 fb^{-1} Backup note: ATL-COM-PHYS-2013-090;
- ▷ ATLAS Conference Note: ATLAS-CONF-2013-052;
- ▷ 20 fb^{-1} Internal communication: ATL-COM-PHYS-2014-1451;
- ▷ 20 fb^{-1} Backup note: ATL-COM-PHYS-2014-003;
- ▷ ATLAS Conference Note: ATLAS-CONF-2015-009;
- ▷ Public paper: arXiv:1505.07018;
- ▷ Published paper: CERN-PH-EP-2015-090, JHEP 1508 (2015) 148;

⋙ I presented this poster **"ATLAS $t\bar{t}$ resonance searches"** in the 37th International Conference on High Energy Physics 2014, which was a summary of the $\sqrt{s} = 7\ TeV$ results listed before.

- ▷ Internal communication proceeding: ATL-COM-PHYS-2014-1260;
- ▷ Internal communication poster: ATL-COM-PHYS-2014-597;
- ▷ Approved proceeding: ATL-PHYS-PROC-2014-203;
- ▷ Approved poster: ATL-PHYS-SLIDE-2014-632;

⋙ My main contribution to the exploitation of ATLAS data is **"Measurement of the charge asymmetry in boosted top quark pair production on 8 TeVpp collision data collected by the ATLAS experiment."**. I was one of the main authors of this analysis. This has been the first time this measurement has ever been done. In August 2015 it was approved by the ATLAS Collaboration. In September 2015 a preliminary result was made public as *ATLAS Conference Note*. In February 2016 was accepted to be published in *Physics Letters B*.

- ▷ Internal communication and backup note: ATL-COM-PHYS-2015-196;
- ▷ ATLAS Conference Note: ATLAS-CONF-2015-048;
- ▷ Public paper: arXiv:1512.06092;
- ▷ Published paper: Physics Letters B Volume 756, Pages 52-71;

Bibliography

[1] ATLAS Collaboration, 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*, *Phys. Lett.* **B716** (2012) 1–29, [arXiv:1207.7214]. 8

[2] CMS Collaboration, S. Chatrchyan *et. al.*, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, *Phys. Lett.* **B716** (2012) 30–61, [arXiv:1207.7235]. 8

[3] N. Arkani-Hamed, S. Dimopoulos, and G. R. Dvali, *The Hierarchy problem and new dimensions at a millimeter*, *Phys. Lett.* **B429** (1998) 263–272, [hep-ph/9803315]. 10

[4] N. Arkani-Hamed, S. Dimopoulos, and G. R. Dvali, *Phenomenology, astrophysics and cosmology of theories with submillimeter dimensions and TeV scale quantum gravity*, *Phys. Rev.* **D59** (1999) 086004, [hep-ph/9807344]. 10

[5] L. Randall and R. Sundrum, *An Alternative to compactification*, *Phys. Rev. Lett.* **83** (1999) 4690–4693, [hep-th/9906064]. 10

[6] K. Agashe, A. Delgado, M. J. May, and R. Sundrum, *RS1, custodial isospin and precision tests*, *JHEP* **0308** (2003) 050, [hep-ph/0308036]. 10, 23

[7] F. Abe et al., *Observation of top quark production in $\overline{p}p$ collisions with the collider detector at fermilab*, *Phys. Rev. Lett.* **74** (Apr, 1995) 2626–2631. 11, 23

[8] S. Abachi et al., *Observation of the top quark*, *Phys. Rev. Lett.* **74** (Apr, 1995) 2632–2637. 11, 23

[9] M. Cacciari, G. P. Salam, and G. Soyez, *The Anti-k(t) jet clustering algorithm*, *JHEP* **0804** (2008) 063, [arXiv:0802.1189]. 15, 16

[10] Y. L. Dokshitzer, G. Leder, S. Moretti, and B. Webber, *Better jet clustering algorithms*, *JHEP* **9708** (1997) 001, [hep-ph/9707323]. 15, 16

[11] M. Wobisch and T. Wengler, *Hadronization corrections to jet cross-sections in deep inelastic scattering*, [hep-ph/9907280]. 15, 16

[12] S. Catani, Y. L. Dokshitzer, M. Seymour, and B. Webber, *Longitudinally invariant K_t clustering algorithms for hadron hadron collisions*, *Nucl.Phys.* **B406** (1993) 187–224. 15, 16

[13] S. D. Ellis and D. E. Soper, *Successive combination jet algorithm for hadron collisions*, *Phys.Rev.* **D48** (1993) 3160–3166, [hep-ph/9305266]. 15, 16

[14] M. Cacciari and G. P. Salam, *Dispelling the N^3 myth for the k_t jet-finder*, *Phys.Lett.* **B641** (2006) 57–61, [hep-ph/0512210]. 16

[15] M. Cacciari, G. P. Salam, and G. Soyez, *FastJet User Manual*, *Eur.Phys.J.* **C72** (2012) 1896, [arXiv:1111.6097]. 16

[16] J. R. Walsh and S. Zuberi, *Factorization Constraints on Jet Substructure*, [arXiv:1110.5333]. 18

[17] ATLAS Collaboration Collaboration, G. Aad *et. al.*, *Performance of jet substructure techniques for large-R jets in proton-proton collisions at $\sqrt{s}$ = 7 TeV using the ATLAS detector*, *JHEP* **1309** (2013) 076, [arXiv:1306.4945]. 18, 21, 139

[18] ATLAS Collaboration, G. Aad *et. al.*, *Jet mass and substructure of inclusive jets in $\sqrt{s} = 7$ TeV pp collisions with the ATLAS experiment*, *JHEP* **1205** (2012) 128, [arXiv:1203.4606]. 18, 139

[19] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, *Jet substructure as a new Higgs search channel at the LHC*, *Phys. Rev. Lett.* **100** (2008) 242001, [arXiv:0802.2470]. 18, 19

[20] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, *Techniques for improved heavy particle searches with jet substructure*, *Phys. Rev.* **D80** (2009) 051501, [arXiv:0903.5081]. 19

[21] D. Krohn, J. Thaler, and L.-T. Wang, *Jet Trimming*, *JHEP* **02** (2010) 084, [arXiv:0912.1342]. 19

[22] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, *Recombination Algorithms and Jet Substructure: Pruning as a Tool for Heavy Particle Searches*, *Phys. Rev.* **D81** (2010) 094023, [arXiv:0912.0033]. 20

[23] D. Adams *et. al.*, *Towards an Understanding of the Correlations in Jet Substructure*, *Eur. Phys. J.* **C75** (2015), no. 9 409, [arXiv:1504.00679]. 20

[24] A. Altheimer *et. al.*, *Boosted objects and jet substructure at the LHC. Report of BOOST2012, held at IFIC Valencia, 23rd-27th of July 2012*, *Eur. Phys. J.* **C74** (2014), no. 3 2792, [arXiv:1311.2708]. 20

[25] A. Altheimer *et. al.*, *Jet Substructure at the Tevatron and LHC: New results, new tools, new benchmarks*, *J. Phys.* **G39** (2012) 063001, [arXiv:1201.0008]. 20

[26] A. Abdesselam, E. B. Kuutmann, U. Bitenc, G. Brooijmans, J. Butterworth, *et. al.*, *Boosted objects: A Probe of beyond the Standard Model physics*, *Eur.Phys.J.* **C71** (2011) 1661, [arXiv:1012.5412]. 20

[27] Fermilab Laboratory. http://www.fnal.gov/pub/about/. 23

[28] C. T. Hill, *Topcolor assisted technicolor*, *Phys.Lett.* **B345** (1995) 483–489, [hep-ph/9411426]. 23, 101

[29] M. J. Dugan, H. Georgi, and D. B. Kaplan, *Anatomy of a composite higgs model*, *Nuclear Physics B* **254** (1985), no. 0 299 – 326. 23

[30] H. Georgi and D. B. Kaplan, *Composite higgs and custodial su(2)*, *Physics Letters B* **145** (1984), no. 34 216 – 220. 23

[31] H. Georgi, D. B. Kaplan, and P. Galison, *Calculation of the composite higgs mass*, *Physics Letters B* **143** (1984), no. 13 152 – 154. 23

[32] D. B. Kaplan, H. Georgi, and S. Dimopoulos, *Composite higgs scalars*, *Physics Letters B* **136** (1984), no. 3 187 – 190. 23

[33] D. B. Kaplan and H. Georgi, *Su(2) u(1) breaking by vacuum misalignment*, *Physics Letters B* **136** (1984), no. 3 183 – 186. 23

[34] A. Pomarol, *Gauge bosons in a five-dimensional theory with localized gravity*, *Phys.Lett.* **B486** (2000) 153–157, [hep-ph/9911294]. 23

[35] H. Davoudiasl, J. Hewett, and T. Rizzo, *Bulk gauge fields in the Randall-Sundrum model*, *Phys.Lett.* **B473** (2000) 43–49, [hep-ph/9911262]. 23

[36] L. Randall and R. Sundrum, *A Large mass hierarchy from a small extra dimension*, *Phys.Rev.Lett.* **83** (1999) 3370–3373, [hep-ph/9905221]. 23

[37] CDF Collaboration Collaboration, T. Aaltonen *et. al.*, *Measurement of the top quark forward-backward production asymmetry and its dependence on event kinematic properties*, *Phys. Rev. D* **87** (May, 2013) 092002. 25, 111

[38] (The D0 Collaboration) Collaboration, V. M. Abazov *et. al.*, *Measurement of the forward-backward asymmetry in top quark-antiquark production in $p\bar{p}$ collisions using the* lepton+jets *channel*, *Phys. Rev. D* **90** (Oct, 2014) 072011. 25

[39] J. A. Aguilar-Saavedra, D. Amidei, A. Juste, and M. Pérez-Victoria, *Asymmetries in top quark pair production at hadron colliders*, *Rev. Mod. Phys.* **87** (2015) 421–455, [arXiv:1406.1798]. 26, 28

[40] J. H. Kuhn and G. Rodrigo, *Charge asymmetries of top quarks at hadron colliders revisited*, *JHEP* **1201** (2012) 063, [arXiv:1109.6830]. 26, 27, 113

[41] M. Czakon, P. Fiedler, and A. Mitov, *Resolving the Tevatron top quark forward-backward asymmetry puzzle: fully differential Next-to-Next-to-Leading-Order calculation*, [arXiv:1411.3007]. 26

[42] G. Aad, T. Abajyan, and L. Zwalinski, *Measurement of the top quark pair production charge asymmetry in proton-proton collisions at $\sqrt{s} = 7$ tev using the atlas detector*, *Journal of High Energy Physics* **2014** (2014), no. 2. 27

[43] CMS Collaboration, *Inclusive and differential measurements of the $t\bar{t}$ charge asymmetry in proton-proton collisions at $\sqrt{s} = 7$ TeV*, *Phys. Lett. B* **717** (Jun, 2012) 129–150. 31 p. 27

[44] R. Cherkaoui, M. Hyaya, S. Gonzalez de la Hoz, E. Ros, V. Sanchez Martinez, M. Villaplana, and M. Vos, *A charge asymmetry measurement for high mass $t\bar{t}$ pairs*, [ATL-COM-PHYS-2012-786], CERN, Geneva, Jun, 2012. 29

[45] ATLAS Collaboration, V. Sanchez-Martinez *et. al.*, *Measurement of the charge asymmetry in highly boosted top-quark pair production in $\sqrt{s} = 8$ TeV pp collision data collected by the ATLAS experiment, Physics Letters B Volume 756, 10 May 2016, Pages 52-71* , [arXiv:1512.06092]. 32, 88, 89, 90, 103

[46] Bruning, Oliver Sim and Collier, Paul and Lebrun, P and Myers, Stephen and Ostojic, Ranko and Poole, John and Proudlock, Paul, *LHC Design Report*. CERN, Geneva, 2004. 33

[47] European Organization for Nuclear Research, CERN. http://home.web.cern.ch/about. 33

[48] The Large Electron-Positron Collider. http://home.cern/about/accelerators/large-electron-positron-collider. 33

[49] The Linear accelerator 2. http://home.cern/about/accelerators/linear-accelerator-2. 35

[50] The Proton Synchrotron Booster. http://home.cern/about/accelerators/proton-synchrotron-booster. 35

[51] The Proton Synchrotron. http://home.cern/about/accelerators/proton-synchrotron. 35

[52] The Super Proton Synchrotron. http://home.cern/about/accelerators/super-proton-synchrotron. 35

[53] LHC Experiments. http://home.web.cern.ch/about/experiments. 35

[54] The ATLAS Experiment. http://atlas.ch. 36

[55] ATLAS Collaboration, *ATLAS detector and physics performance: Technical Design Report, 1*. Technical Design Report ATLAS. CERN, Geneva, 1999. Electronic version not available. 36

[56] ATLAS Collaboration, *ATLAS detector and physics performance: Technical Design Report, 2*. Technical Design Report ATLAS. CERN, Geneva, 1999. Electronic version not available. 36

[57] ATLAS Collaboration, *ATLAS inner detector: Technical design report. Vol. 1*, . 38

[58] Worldwide LHC Computing Grid. http://wlcg.web.cern.ch/. 47

[59] Worldwide LHC Computing Grid - public web. http://wlcg-public.web.cern.ch/. 47

[60] C. Eck, J. Knobloch, L. Robertson, I. Bird, K. Bos, N. Brook, D. Dllmann, I. Fisk, D. Foster, B. Gibbard, C. Grandi, F. Grey, J. Harvey, A. Heiss, F. Hemmer, S. Jarp, R. Jones, D. Kelsey, M. Lamanna, H. Marten, P. Mato-Vila, F. Ould-Saada, B. Panzer-Steindel, L. Perini, Y. Schutz, U. Schwickerath, J. Shiers, and T. Wenaus, *LHC computing Grid: Technical Design Report. Version 1.06 (20 Jun 2005)*. Technical Design Report LCG. CERN, Geneva, 2005. 47

[61] The European Grid Initiative. http://www.egi.eu/. 47

[62] P. Andreetto *et. al.*, *The glite workload management system*, *Journal of Physics: Conference Series* **119** (2008), no. 6 062007. 47

[63] NorduGrid - Advanced Resource Connector. http://www.nordugrid.org/arc/. 47

[64] M. Ellert *et. al.*, *Advanced resource connector middleware for lightweight computational grids*, *Future Generation Computer Systems* **23** (2007), no. 2 219 – 240. 47

[65] Open Science Grid. http://www.opensciencegrid.org/. 47

[66] M. Villaplana, G. Amorós, G. Borges, C. Borrego, J. Carvalho, M. David, X. Espinal, A. Fernández, J. Gomes, S. González de la Hoz, M. Kaci, A. Lamas, J. Nadal, M. Oliveira, E. Oliver, C. Osuna, A. Pacheco, J. Pardo, J. del Peso, J. Salt, J. Sánchez, and H. Wolters, *Iberian ATLAS Cloud response during the first LHC collisions*, [ATL-SOFT-PROC-2011-018], CERN, Geneva, Jan, 2011. 47

[67] V. Sánchez-Martínez, G. Borges, C. Borrego, J. del Peso, M. Delfino, J. Gomes, S. González de la Hoz, A. Pacheco Pages, J. Salt, A. Sedov, M. Villaplana, and H. Wolters, *Lessons learned from the ATLAS performance studies of the Iberian Cloud for the first LHC running period.*, [ATL-SOFT-PROC-2013-007], CERN, Geneva, Oct, 2013. 47

[68] S. Gonzalez de la Hoz, J. Del Peso, F. Fassi, A. Fernandez Casani, M. Kaci, V. R. Lacort Pellicer, A. D. R. Montiel Gonzalez, E. Oliver Garcia, A. Pacheco Pages, J. Sánchez, V. Sanchez Martinez,

J. Salt, and M. Villaplana Perez, *Spanish ATLAS Tier-2 facing up to Run-2 period of LHC*, [ATL-SOFT-PROC-2015-005], CERN, Geneva, May, 2015. 47

[69] ATLAS Collaboration, G. Duckeck, D. Barberis, R. Hawkings, R. Jones, N. McCubbin, G. Poulard, D. Quarrie, T. Wenaus, and E. Obreshkov, *ATLAS computing: Technical Design Report 2005*, . 48, 51

[70] R. W. L. Jones and D. Barberis, *The evolution of the atlas computing model*, *Journal of Physics: Conference Series* **219** (2010), no. 7 072037. 48

[71] E. Oliver *et. al.*, *Operation of the ATLAS Spahish Tier-2 during the LHC Run I*, in *7th Iberian Grid Infrastructure Conference - IBERGRID*, pp. 227–239, 2013. 49

[72] The ATLAS Event Data Model. https://twiki.cern.ch/twiki/bin/view/AtlasComputing/EventDataModel. 50

[73] ROOT. https://root.cern.ch/. 50

[74] I. Antcheva *et. al.*, *ROOT: A C++ framework for petabyte data storage, statistical analysis and visualization*, *Comput. Phys. Commun.* **180** (2009) 2499–2512, [arXiv:1508.07749]. 50

[75] The GEAnt4. http://geant4.cern.ch/. 51

[76] GEANT4 Collaboration, S. Agostinelli *et. al.*, *GEANT4: A Simulation toolkit*, *Nucl. Instrum. Meth.* **A506** (2003) 250–303. 51, 71

[77] W. Lukas, *Fast Simulation for ATLAS: Atlfast-II and ISF*, [ATL-SOFT-PROC-2012-065], CERN, Geneva, Jun, 2012. 51

[78] *Performance of the Fast ATLAS Tracking Simulation (FATRAS) and the ATLAS Fast Calorimeter Simulation (FastCaloSim) with single particles*, [ATL-SOFT-PUB-2014-001], CERN, Geneva, Mar, 2014. 51, 71

[79] R. W. Jansky, *The ATLAS Fast Monte Carlo Production Chain Project*, [ATL-SOFT-PROC-2015-007], CERN, Geneva, May, 2015. 51

[80] The ATLAS Metadata Interface. https://twiki.cern.ch/twiki/bin/viewauth/AtlasComputing/AtlasMetadataInterface. 53

[81] M. de Oliveira Branco, E. Zaluska, and D. D. Roure, *Managing very-large distributed datasets*, *Lecture Notes in Computer Science* **5331** (November, 2008) 775–792. 53

[82] P. Calafiura, W. Lavrijsen, C. Leggett, M. Marino, and D. Quarrie, *The athena control framework in production, new developments and lessons learned*, in *Computing in high energy physics and nuclear physics. Proceedings, Conference, CHEP'04, Interlaken, Switzerland, September 27-October 1, 2004*, pp. 456–458, 2005. 53

[83] F. Brochu, U. Egede, J. Elmsheuser, K. Harrison, R. W. L. Jones, H. C. Lee, D. Liko, A. Maier, J. T. Moscicki, A. Muraru, G. N. Patrick, K. Pajchel, W. Reece, B. H. Samset, M. W. Slater, A. Soroko, C. L. Tan, and D. C. Vanderster, *Ganga: a tool for computational-task management and easy access to grid resources*, *CoRR* **abs/0902.2685** (2009). 53

[84] "Python programming language." http://www.python.org. 53

[85] Nilsson P, Caballero J, De K, Maeno T, Potekhin M, Wenaus T, "The panda system in the atlas experiment." https://indico.cern.ch/event/34666/session/4/contribution/32/attachments/683811/939296/P_Nilsson_-_PanDA.pdf. 53

[86] T. Maeno, *Panda: distributed production and distributed analysis system for atlas*, *Journal of Physics: Conference Series* **119** (2008), no. 6 062036. 53

[87] "Memorandum of understanding." http://wlcg.web.cern.ch/collaboration/mou. 55

[88] "Memorandum of understanding, 28 april 2015." https://espace.cern.ch/WLCG-document-repository/MoU%20docs/Basics/MoU-blank-example-28APR2015.pdf. 55

[89] "Hep-spec06 benchmark." http://w3.hepix.org/benchmarks/doku.php. 55

[90] The dCache Project. https://www.dcache.org/. 56

[91] "Lustre: A scalable, high-performance file system." http://www.cse.buffalo.edu/faculty/tkosar/cse710/papers/lustre-whitepaper.pdf. 56

[92] "Lustre cluster file system." http://wiki.lustre.org. 56

[93] The Storage Resource Manager.
https://sdm.lbl.gov/srm-wg/doc/SRM.v2.2.html. 56

[94] V. Garonne, G. A. Stewart, M. Lassnig, A. Molfetas, M. Barisits, T. Beermann, A. Nairz, L. Goossens, F. B. Megino, C. Serfon, D. Oleynik, and A. Petrosyan, *The atlas distributed data management project: Past and future*, *Journal of Physics: Conference Series* **396** (2012), no. 3 032045. 56

[95] M. Branco, D. Cameron, B. Gaidioz, V. Garonne, B. Koblitz, M. Lassnig, R. Rocha, P. Salgado, and T. Wenaus, *Managing ATLAS data on a petabyte-scale with DQ2*, *J. Phys.: Conf. Ser.* **119** (2008) 062017. 56

[96] Service and Availability Monitoring.
http://atlas.fis.utfsm.cl/atlas/adcos/AdcMonitoring.htm#Service_and_Availability_Monitor. 61

[97] D. C. van der Ster, J. Elmsheuser, M. Ubeda Garcia, and M. Paladin, *HammerCloud: A Stress Testing System for Distributed Analysis*, [CERN-IT-2011-001], CERN, Geneva, Jan, 2011. 62

[98] D. Barberis, *ATLAS computing challenges before the next LHC run*, [ATL-SOFT-PROC-2014-004], CERN, Geneva, Sep, 2014. 63, 69

[99] V. Garonne, R. Vigne, G. Stewart, M. Barisits, T. B. eermann, M. Lassnig, C. Serfon, L. Goossens, A. Nairz, and the Atlas Collaboration, *Rucio: The next generation of large scale distributed system for atlas data management*, *Journal of Physics: Conference Series* **513** (2014), no. 4 042021. 63

[100] K. De, D. Golubkov, A. Klimentov, M. Potekhin, A. Vaniachine, and the Atlas Collaboration, *Task management in the new atlas production system*, *Journal of Physics: Conference Series* **513** (2014), no. 3 032078. 63

[101] CERN openlab Whitepaper on Future IT Challenges in Scientific Research. http://dx.doi.org/10.5281/zenodo.8765. 63

[102] J. Salt *et. al.*, *New Challenges of the Spanish ATLAS Tier-2 to address the Run-2 period of LHC*, in *8th Iberian Grid Infrastructure Conference - IBERGRID*, pp. 41–52, 2014. 63

[103] J. Catmore, P. Laycock, E. Gramstad, T. Gillam, J. Cranshaw, N. Ozturk, and G. Stewart, *A New Petabyte-scale Data Derivation Framework for ATLAS*, [ATL-SOFT-PROC-2015-041], CERN, Geneva, May, 2015. 64

[104] J. Catmore, *A New Petabyte-scale Data Derivation Framework for ATLAS*, [ATL-SOFT-SLIDE-2015-120] 64

[105] D. Barberis, J. Cranshaw, A. Favareto, A. Fernandez, E. Gallas, S. Gonzales De La Hoz, J. Hrivnac, D. Malon, M. Nowak, F. Prokoshin, J. Salt, J. Sanchez Martinez, R. Yuan, and R. Tobbicke, *The ATLAS EventIndex: Full chain deployment and first operation*, [ATL-SOFT-PROC-2014-007], CERN, Geneva, Oct, 2014. 64

[106] D. Barberis, S. E. Cardenas Zarate, J. Cranshaw, A. Favareto, A. Fernandez Casani, E. Gallas, C. Glasman, S. Gonzalez de la Hoz, J. Hrivnac, D. Malon, F. Prokoshin, J. Salt, J. Sánchez, R. Toebbicke, and R. Yuan, *The ATLAS EventIndex: architecture, design choices, deployment and first operation experience*, [ATL-SOFT-PROC-2015-008], CERN, Geneva, May, 2015. 64

[107] D. Barberis, S. E. Cardenas Zarate, E. Gallas, and F. Prokoshin, *Integration of the EventIndex with other ATLAS systems*, [ATL-SOFT-PROC-2015-053], CERN, Geneva, May, 2015. 64

[108] R. Gardner, S. Campana, G. Duckeck, J. Elmsheuser, A. Hanushevski, F. Hnig, J. Iven, F. Legger, I. Vukotic, and W. Yang, *Data Federation Strategies for ATLAS using XRootD*, [ATL-SOFT-PROC-2013-039], CERN, Geneva, Oct, 2013. 64

[109] R. Gardner, S. Campana, G. Duckeck, J. Elmsheuser, A. Hanushevsky, F. G. Hnig, J. Iven, F. Legger, I. Vukotic, W. Yang, and the Atlas Collaboration, *Data federation strategies for atlas using xrootd*, *Journal of Physics: Conference Series* **513** (2014), no. 4 042049. 64

[110] P. van Gemmeren, S. Binet, P. Calafiura, W. Lavrijsen, D. Malon, and V. Tsulaia, *I/O Strategies for Multicore Processing in ATLAS*, [ATL-SOFT-PROC-2012-014], CERN, Geneva, May, 2012. 65

[111] P. van Gemmeren, S. Binet, P. Calafiura, W. Lavrijsen, D. Malon, and V. Tsulaia, *I/O Strategies for Multicore Processing in ATLAS*, [ATL-SOFT-SLIDE-2012-189] 65

[112] S. Gonzalez de la Hoz *et. al.*, *Analysis facility infrastructure (Tier-3) for ATLAS experiment*, *Eur. Phys. J.* **C54** (2008) 691–697. 65

[113] M. Villaplana *et. al.*, *First tests with Tier-3 facility for the ATLAS experiment at IFIC (Valencia)*, in *4th Iberian Grid Infrastructure Conference - IBERGRID*, pp. 212–220, 2010. 65

[114] M. Villaplana, S. González de la Hoz, A. Fernández, J. Salt, A. Lamas, F. Fassi, M. Kaci, E. Oliver, J. Sánchez, and V. Sánchez-Martínez, *ATLAS Tier-3 within IFIC-Valencia analysis facility*, [ATL-SOFT-PROC-2012-017], CERN, Geneva, May, 2012. 65

[115] J. Elmsheuser and D. van der Ster, *Distributed data analysis in the atlas experiment: Challenges and solutions*, in *International Conference on Computing in High Energy and Nuclear Physics, CHEP 2012*, vol. 396, p. 032035, 2012. 67

[116] Fundacion Epañola para la Ciencia y la Tecnologia, *Libro blanco de e-Ciencia en España*. Ministerio de Educacion y Ciencia, 2004. 67

[117] S. González de La Hoz, E.Oliver, J. Salt, E. Ros, M. Villaplana, C. Escobar, G. Amorós, F. A, M. Kaci, A. Lamas, V. Sánchez, and J. Sánchez., *Aplicaciones de Física utilizando la infraestructura de e-Ciencia del IFIC*, *Resúmenes de las comunicaciones de la XXXIII Bienal de Física* **Vol. I, ISBN: 978-84-86116-8** (September, 2011) 20–21. 67

[118] G. Amoros, J. Ors, A. Fernandez, S. Gonzalez de la Hoz, M. Kaci, A. Lamas, L. March, E. Oliver, J. Salt, J. Sanchez, M. Villaplana, R. Vives, and F. Albiol, *Scientific applications running at ific using the grid technologies within the e-science framework*, in *Advanced Engineering Computing and Applications in Sciences, 2009. ADVCOMP '09. Third International Conference on*, pp. 73 –76, oct., 2009. 67

[119] T. C. Group, "Cacti: The complete rrdtool-based graphing solution." http://www.cacti.net/. 68

[120] "Ganglia monitoring system." http://ganglia.sourceforge.net/. 68

[121] D. Barberis, *ATLAS computing challenges before the next LHC run*, [ATL-SOFT-SLIDE-2014-361] 69

[122] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, *JHEP* **11** (2007) 070, [arXiv:0709.2092]. 72

[123] T. Sjostrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, *JHEP* **05** (2006) 026, [hep-ph/0603175]. 72, 94

[124] P. Z. Skands, *Tuning Monte Carlo Generators: The Perugia Tunes*, *Phys. Rev.* **D82** (2010) 074018, [arXiv:1005.3457]. 72

[125] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky, J. Pumplin, and C. P. Yuan, *New parton distributions for collider physics*, *Phys. Rev.* **D82** (2010) 074024, [arXiv:1007.2241]. 72

[126] J. H. Kuhn, A. Scharf, and P. Uwer, *Electroweak corrections to top-quark pair production in quark-antiquark annihilation*, *Eur. Phys. J.* **C45** (2006) 139–150, [hep-ph/0508092]. 72

[127] J. H. Kuhn, A. Scharf, and P. Uwer, *Electroweak effects in top-quark pair production at hadron colliders*, *Eur. Phys. J.* **C51** (2007) 37–53, [hep-ph/0610335]. 72

[128] J. H. Khn, A. Scharf, and P. Uwer, *Weak Interactions in Top-Quark Pair Production at Hadron Colliders: An Update*, *Phys. Rev.* **D91** (2015), no. 1 014020, [arXiv:1305.5773]. 72

[129] S. Frixione and B. R. Webber, *Matching NLO QCD computations and parton shower simulations*, *JHEP* **06** (2002) 029, [hep-ph/0204244]. 72

[130] S. Frixione, P. Nason, and B. R. Webber, *Matching NLO QCD and parton showers in heavy flavor production*, *JHEP* **08** (2003) 007, [hep-ph/0305252]. 72

[131] G. Corcella, I. G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, P. Richardson, M. H. Seymour, and B. R. Webber, *HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)*, *JHEP* **01** (2001) 010, [hep-ph/0011363]. 72

[132] J. M. Butterworth, J. R. Forshaw, and M. H. Seymour, *Multiparton interactions in photoproduction at HERA*, *Z. Phys.* **C72** (1996) 637–646, [hep-ph/9601371]. 72

[133] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau, and A. D. Polosa, *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, *JHEP* **07** (2003) 001, [hep-ph/0206293]. 72

[134] M. Cacciari, M. Czakon, M. Mangano, A. Mitov, and P. Nason, *Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation*, *Phys. Lett.* **B710** (2012) 612–622, [arXiv:1111.5869]. 72, 98

[135] M. Beneke, P. Falgari, S. Klein, and C. Schwinn, *Hadronic top-quark pair production with NNLL threshold resummation*, *Nucl. Phys.* **B855** (2012) 695–741, [arXiv:1109.1536]. 72, 98

[136] P. Brnreuther, M. Czakon, and A. Mitov, *Percent Level Precision Physics at the Tevatron: First Genuine NNLO QCD Corrections to $q\bar{q} \to t\bar{t} + X$*, *Phys. Rev. Lett.* **109** (2012) 132001, [arXiv:1204.5201]. 72, 98

[137] M. Czakon and A. Mitov, *NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels*, *JHEP* **12** (2012) 054, [arXiv:1207.0236]. 72, 98

[138] M. Czakon and A. Mitov, *NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction*, *JHEP* **01** (2013) 080, [arXiv:1210.6832]. 72, 98

[139] M. Czakon, P. Fiedler, and A. Mitov, *Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $\mathcal{O}(\alpha_S^4)$*, *Phys. Rev. Lett.* **110** (2013) 252004, [arXiv:1303.6254]. 72, 98

[140] N. Kidonakis, *Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production*, *Phys. Rev.* **D83** (2011) 091503, [arXiv:1103.2792]. 72, 141

[141] N. Kidonakis, *Two-loop soft anomalous dimensions for single top quark associated production with a W- or H-*, *Phys. Rev.* **D82** (2010) 054018, [arXiv:1005.4451]. 72, 141

[142] N. Kidonakis, *NNLL resummation for s-channel single top quark production*, *Phys. Rev.* **D81** (2010) 054028, [arXiv:1001.5034]. 72, 141

[143] T. Gleisberg, S. Hoeche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert, and J. Winter, *Event generation with SHERPA 1.1*, *JHEP* **02** (2009) 007, [arXiv:0811.4622]. 72

[144] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, *JHEP* **07** (2014) 079, [arXiv:1405.0301]. 72, 94

[145] J. Alwall, P. Demin, S. de Visscher, R. Frederix, M. Herquet, *et. al.*, *MadGraph/MadEvent v4: The New Web Generation*, *JHEP* **0709** (2007) 028, [arXiv:0706.2334]. 72, 94

[146] F. Maltoni and T. Stelzer, *MadEvent: Automatic event generation with MadGraph*, *JHEP* **0302** (2003) 027, [hep-ph/0208156]. 72, 94

[147] M. V. Garzelli, A. Kardos, C. G. Papadopoulos, and Z. Trocsanyi, $t\,\bar{t}\,W^{+-}$ *and* $t\,\bar{t}\,Z$ *Hadroproduction at NLO accuracy in QCD with Parton Shower and Hadronization effects*, *JHEP* **11** (2012) 056, [arXiv:1208.2665]. 72

[148] ATLAS Collaboration, G. Aad *et. al.*, *Jet energy measurement with the ATLAS detector in proton-proton collisions at* $\sqrt{s} = 7$ *TeV*, *Eur.Phys.J.* **C73** (2013), no. 3 2304, [arXiv:1112.6426]. 73, 140

[149] ATLAS Collaboration, *Measurement of the b-tag Efficiency in a Sample of Jets Containing Muons with 5* $fb-1$ *of Data from the ATLAS Detector*, [ATLAS-CONF-2012-043], CERN, Geneva, Mar, 2012. 74

[150] S. Calvet, C. Pollard, and J. Zhong, *Electron-Jet Overlap Removal for the l + jets* $t\bar{t}$*Resonances Search*, [ATL-COM-PHYS-2014-343], CERN, Geneva, Apr, 2014. 74

[151] ATLAS Collaboration, G. Aad *et. al.*, *A search for* $t\bar{t}$ *resonances using lepton-plus-jets events in proton-proton collisions at* $\sqrt{s} = 8$ *TeV with the ATLAS detector, J. High Energy Phys. 08 (2015) 148* , [arXiv:1505.07018]. 75, 93, 94, 101, 104, 113, 121, 132

[152] *Measurement of the differential cross-section of highly boosted top quarks as a function of their transverse momentum using the ATLAS detector in* $\sqrt{s} = 8$ *TeV proton-proton collisions*, [ATLAS-CONF-2014-057], CERN, Geneva, Sep, 2014. 75, 104

[153] S. Gonzalez de la Hoz, B. A. Magy, D. E. Marley, A. McCarn, D. Melini, V. Sanchez Martinez, T. A. Schwarz, and M. Vos, *Measurement of the charge asymmetry in boosted top quark pair production on 8 TeV pp*

collision data collected by the ATLAS experiment, [ATL-COM-PHYS-2015-196], CERN, Geneva, Mar, 2015. 75

[154] J. A. Aguilar Saavedra, S. Gonzalez de la Hoz, B. A. Magy, D. E. Marley, A. McCarn, D. Melini, V. Sanchez Martinez, T. A. Schwarz, and M. Vos, *Measurement of the charge asymmetry in boosted top quark pair production on 8 TeV pp collision data collected by the ATLAS experiment*, [ATL-COM-PHYS-2015-841], CERN, Geneva, Aug, 2015. 75

[155] T. Chwalek, *Messung der W-Boson-Helizittsanteile in Top-Quark-Zerfllen mit dem CDF II Experiment und Studien zu einer frhen Messung des ttbar-Wirkungsquerschnitts mit dem CMS Experiment.* PhD thesis, Karlsruhe U., Karlsruhe, 2010. Presented 12 Feb 2010. 79

[156] ATLAS Collaboration, G. Aad *et. al.*, *Measurement of the charge asymmetry in top quark pair production in pp collisions at* $\sqrt{s} = 7$ *TeV using the ATLAS detector*, *Eur. Phys. J.* **C72** (2012) 2039, [arXiv:1203.4211]. 85, 104

[157] ATLAS Collaboration, G. Aad *et. al.*, *Measurements of top quark pair relative differential cross-sections with ATLAS in pp collisions at* $\sqrt{s} = 7$ *TeV*, *Eur. Phys. J.* **C73** (2013), no. 1 2261, [arXiv:1207.5644]. 85, 104

[158] ATLAS Collaboration, G. Aad *et. al.*, *Measurement of the top quark pair production cross-section with ATLAS in the single lepton channel*, *Phys. Lett.* **B711** (2012) 244–263, [arXiv:1201.1889]. 86, 104, 121

[159] B. Lillie, L. Randall, and L.-T. Wang, *The Bulk RS KK-gluon at the LHC*, *JHEP* **09** (2007) 074, [hep-ph/0701166]. 94

[160] V. Sanchez-Martinez, *ATLAS ttbar resonance searches*, [ATL-PHYS-PROC-2014-203], CERN, Geneva, Oct, 2014. 94

[161] ATLAS Collaboration, G. Aad *et. al.*, *Search for* $t\bar{t}$ *resonances in the lepton plus jets final state with ATLAS using 4.7* fb^{-1} *of pp collisions at* $\sqrt{s} = 7$ *TeV*, *Phys. Rev.* **D88** (2013), no. 1 012004, [arXiv:1305.2756]. 94, 104, 113

[162] K. Agashe, A. Belyaev, T. Krupovnickas, G. Perez, and J. Virzi, *LHC Signals from Warped Extra Dimensions*, *Phys. Rev.* **D77** (2008) 015003, [hep-ph/0612015]. 94, 101

[163] CMS Collaboration, S. Chatrchyan *et. al.*, *Searches for new physics using the $t\bar{t}$ invariant mass distribution in pp collisions at $\sqrt{s}$=8 TeV*, *Phys. Rev. Lett.* **111** (2013), no. 21 211804, [arXiv:1309.2030]. [Erratum: Phys. Rev. Lett.112,no.11,119903(2014)]. 94

[164] M. Vos, V. Sanchez Martinez, and J. Ferrando, *Resonance searches in top pair production: a guide to understanding the ATLAS and CMS benchmarks*, [ATL-COM-PHYS-2013-1040], CERN, Geneva, Jul, 2013. 94

[165] ATLAS Collaboration, G. Aad *et. al.*, *Muon reconstruction efficiency and momentum resolution of the ATLAS experiment in proton-proton collisions at $\sqrt{s}$ = 7 TeV in 2010*, *Eur. Phys. J.* **C74** (2014), no. 9 3034, [arXiv:1404.4562]. 97

[166] M. Czakon, D. Heymes, and A. Mitov, *High-precision differential predictions for top-quark pairs at the LHC*, [arXiv:1511.00549]. 98

[167] *Calibration of b-tagging using dileptonic top pair events in a combinatorial likelihood approach with the ATLAS experiment*, [ATLAS-CONF-2014-004], CERN, Geneva, Feb, 2014. 100, 140

[168] T. Junk, *Confidence level computation for combining searches with small statistics*, *Nucl. Instrum. Meth.* **A 434** (1999) 435–443. 100

[169] A. Altheimer, A. Arce, L. Asquith, J. Backus Mayes, E. Bergeaas Kuutmann, *et. al.*, *Boosted objects and jet substructure at the LHC. Report of BOOST2012, held at IFIC Valencia, 23rd-27th of July 2012*, *Eur.Phys.J.* **C74** (2014) 2792, [arXiv:1311.2708]. 103

[170] ATLAS Collaboration, *Prospects for top anti-top resonance searches using early ATLAS data.*, [ATL-PHYS-PUB-2010-008], CERN, Geneva, Jul, 2010. 104, 113

[171] *A search for $t\bar{t}$ resonances in the lepton plus jets final state with ATLAS using 14 fb^{-1} of pp collisions at $\sqrt{s} = 8$ TeV*, [ATLAS-CONF-2013-052], CERN, Geneva, May, 2013. 104, 113

[172] ATLAS Collaboration Collaboration, G. Aad *et. al.*, *A search for $t\bar{t}$ resonances in lepton+jets events with highly boosted top quarks collected in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *JHEP* **1209** (2012) 041, [arXiv:1207.2409]. 104, 113

[173] G. Choudalakis, *Fully Bayesian Unfolding*, [arXiv:1201.4612]. 106, 113, 132

[174] D0 Collaboration, V. M. Abazov *et. al.*, *Forward-backward asymmetry in top quark-antiquark production*, *Phys.Rev.* **D84** (2011) 112005, [arXiv:1107.4995]. 111

[175] J. Aguilar-Saavedra and M. Perez-Victoria, *Asymmetries in t $\bar{t}$ production: LHC versus Tevatron*, *Phys.Rev.* **D84** (2011) 115013, [arXiv:1105.4606]. 111

[176] J. Aguilar-Saavedra and M. Perez-Victoria, *Simple models for the top asymmetry: Constraints and predictions*, *JHEP* **1109** (2011) 097, [arXiv:1107.0841]. 111

[177] J. Aguilar-Saavedra, *Single top quark production at LHC with anomalous Wtb couplings*, *Nucl.Phys.* **B804** (2008) 160–192, [arXiv:0803.3810]. 111

[178] M. P. Rosello and M. Vos, *Constraints on four-fermion interactions from the $t\bar{t}$ charge asymmetry at hadron colliders*, [arXiv:1512.07542]. 111

[179] J. H. Kuhn and G. Rodrigo, *Charge asymmetry of heavy quarks at hadron colliders*, *Phys.Rev.* **D59** (1999) 054017, [hep-ph/9807420]. 113

[180] J. H. Kuhn and G. Rodrigo, *Charge asymmetry in hadroproduction of heavy quarks*, *Phys.Rev.Lett.* **81** (1998) 49–52, [hep-ph/9802268]. 113

[181] ATLAS, CDF, CMS, D0 Collaboration, *First combination of Tevatron and LHC measurements of the top-quark mass*, [arXiv:1403.4427]. 113

[182] Particle Data Group Collaboration, K. Olive *et. al.*, *Review of Particle Physics*, *Chin.Phys.* **C38** (2014) 090001. 113

[183] L. Harland-Lang, A. Martin, P. Motylinski, and R. Thorne, *Parton distributions in the LHC era: MMHT 2014 PDFs*, *Eur.Phys.J.* **C75** (2015), no. 5 204, [arXiv:1412.3989]. 113

[184] ATLAS Collaboration, *Measurement of the charge asymmetry in top quark pair production in pp collision data at $\sqrt{s}$ = 8 TeV with the ATLAS detector*, [in circulation]. 114, 131

[185] CMS Collaboration, V. Khachatryan *et. al.*, *Inclusive and differential measurements of the t t-bar charge asymmetry in pp collisions at sqrt(s) = 8 TeV*, [arXiv:1507.03119]. 114

[186] ATLAS Collaboration, D. Schouten, A. Tanasijczuk, and M. Vetterli, *ATLAS Jet Energy Scale*, in *Proceedings, 31st International Conference on Physics in collisions (PIC 2011)*, 2012. arXiv:1201.2429. 122

[187] ATLAS Collaboration, M. Leyton, *Minimum Bias and Underlying Event Measurements with ATLAS*, in *Proceedings, 3rd International Workshop on Multiple Partonic Interactions at the LHC (MPI@LHC 2011)*, pp. 11–19, 2012. arXiv:1202.2090. 123

[188] G. Bohm, *Introduction to statistics and data analysis for physicists*. Verl. Dt. Elektronen-Synchrotron, Hamburg, 2010. 127

[189] G. Aad *et. al.*, *Measurement of dijet cross-sections in pp collisions at 7 tev centre-of-mass energy using the atlas detector*, *Journal of High Energy Physics* **2014** (2014), no. 5. 127

[190] ATLAS Collaboration Collaboration, G. Aad *et. al.*, *Measurement of the top quark pair production charge asymmetry in proton-proton collisions at $\sqrt{s}$ = 7 TeV using the ATLAS detector*, *JHEP* **1402** (2014) 107, [arXiv:1311.6724]. 131

[191] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, *Equation of state calculations by fast computing machines*, *J.Chem.Phys.* **21** (1953) 1087–1092. 133

[192] ATLAS Collaboration, *Jet energy measurement and its systematic uncertainty in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *Submitted to Eur. Phys. J.* (2014) [arXiv:1406.0076]. 140

[193] ATLAS Collaboration, *Single hadron response measurement and calorimeter jet energy scale uncertainty with the ATLAS detector at the LHC*, *Eur. Phys. J.* **C73** (2013) 2305, [arXiv:1203.1302]. 140

[194] ATLAS Collaboration, *Jet energy resolution in proton-proton collisions at $\sqrt{s} = 7$ TeV recorded in 2010 with the ATLAS detector*, *Eur. Phys. J.* **C73** (2013) 2306, [arXiv:1210.6210]. 140

[195] ATLAS Collaboration Collaboration, *Measurement of the b-tag Efficiency in a Sample of Jets Containing Muons with 5 fb^{-1} of Data from the ATLAS Detector*, [ATLAS-CONF-2012-043], CERN, Geneva, Mar, 2012. 140

[196] ATLAS Collaboration Collaboration, *Measurement of the Mistag Rate with 5 fb^{-1} of Data Collected by the ATLAS Detector*, [ATLAS-CONF-2012-040], CERN, Geneva, Mar, 2012. 140

[197] ATLAS Collaboration, *Electron performance measurements with the ATLAS detector using the 2010 LHC proton-proton collision data*, *Eur. Phys. J.* **C72** (2012) 1909, [arXiv:1110.3174]. 140

[198] *Preliminary results on the muon reconstruction efficiency, momentum resolution, and momentum scale in ATLAS 2012 pp collision data*, [ATLAS-CONF-2013-088], CERN, Geneva, Aug, 2013. 140

[199] ATLAS Collaboration, G. Aad *et. al.*, *Performance of Missing Transverse Momentum Reconstruction in Proton-Proton Collisions at 7 TeV with ATLAS*, *Eur.Phys.J.* **C72** (2012) 1844, [arXiv:1108.5602]. 140

[200] A. Hocker and V. Kartvelishvili, *SVD approach to data unfolding*, *Nucl.Instrum.Meth.* **A372** (1996) 469–481, [hep-ph/9509307]. 145

[201] Blog: Universo a la vista. http://universoalavista.blogspot.com.es/2009/07/la-fisica-de-particulas-y-sus.html. 157

Gracias

Gracias a todas aquellas personas que de alguna forma, me han apoyado en esta andadura, especialmente a mi familia y a mi Baby. Han sido 5 años de duro trabajo y muchos sacrificos, de “altibajos”, de frustraciones por no conseguir a la primera lo que quería... pero ha merecido la pena.

Cada etapa de tu vida te enseña cosas nuevas. Esta etapa me ha enseñado que la paciencia es un don y la empatía una virtud. He aprendido a canalizar mis cabreos y controlar mis nervios, y sobre todo he aprendido que todos somos inteligentes, sólo tenemos que descubrir qué es lo que realmente nos gusta y luchar por ello, y demostrarnos a nosotros mismos que somos capaces de conseguir todo aquello que nos propongamos, a pesar de que haya gente que nos dirá lo contrario.

He descubierto que las personas que creen en ti siempre te van a apoyar, en los buenos y en los malos momentos, y siempre te van a animar para seguir adelante caminando a tu lado.

Sólo puedo decir GRACIAS. Gracias a esta etapa y a las personas que me han acompañado, las cuales se sentirán aludidas al leer estas líneas.