

FACULTAT DE FÍSICA
DEPARTAMENT DE FÍSICA ATÒMICA, MOLECULAR I
NUCLEAR



VNIVERSITAT
DE VALÈNCIA

**Searches for beyond the Standard Model
physics with boosted topologies in the ATLAS
experiment using the Grid-based Tier-3 facility
at IFIC-Valencia**

Miguel Villaplana Pérez

Dirigido por
Santiago González de la Hoz y Marcel Vos

DOCTORADO EN FÍSICA

Septiembre 2013

Dr. Santiago González de la Hoz.

Contratado Doctor del Departamento de Física Atómica, Molecular y Nuclear de la Universidad de Valencia.

Dr. Marcel Vos.

Ramón y Cajal del Consejo Superior de Investigaciones Científicas.

CERTIFICAN:

Que la presente memoria '**Searches for beyond the Standard Model physics with boosted topologies 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 Don Miguel Villaplana Pérez 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

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

IFIC

**Instituto de Física
CORPUSCULAR**

El IFIC es un centro mixto de la *Universitat de València* y el *Consejo Superior de Investigaciones Científicas* (CSIC)



VNIVERSITAT
DE VALÈNCIA



CSIC

A mis padres.
Por su apoyo y su curiosidad.

Preface

The Large Hadron Collider and the experiments that study its collisions are considered one of the most -if not the most- complex enterprises humanity has ever embarked on. The challenges are diverse: the accelerator, the detectors and the computing infrastructure are examples of state-of-the-art technology in their respective areas, and the physics program of the experiments is unparalleled in its volume, diversity, complexity and impact. The ATLAS Collaboration is formed by nearly 3000 scientist from over 200 institutions around the world. This is the largest scientific collaboration in human history and coordinating it efficiently is a challenge, which should not be underestimated.

The results shown in this thesis are the fruit of the collaboration of several people from different institutions within the ATLAS Collaboration. IFIC has played a leading role in the analyses described here: coordinating people, suggesting new ideas and implementing them, and presenting the results to the scientific community.

Through these lines I would like to outline my role in the work presented in this document.

Grid computing

My primary mission in the Grid computing group of IFIC-Valencia is to test the site's Tier-2 and Tier-3 infrastructure and to give feedback to the system administrators and the technicians. As a member of the ATLAS ES-Cloud squad, I contributed to monitor the behavior of the ATLAS Spanish and Portuguese Tier-2 sites. This work is described in chapter 3. In section 3.2 I describe the Tier-2 activities in the ATLAS ES-Cloud. My work in this area has appeared in several reports: *J. Phys. Conf. Ser. 331, 072068 (2011)*, *J. Phys.: Conf. Ser. 396 032051 (2012)*. In sections 3.3 and 3.4 I present IFIC-Valencia Tier-3 prototype, its design, deployment and performance. As a

Tier-3 attached to a Tier-2 and sharing some of the its qualities, this design is of special interest for the community. I have presented its performance in several conferences (*J. Phys.: Conf. Ser.* 396 042062 (2012)). I won the *Best Student Paper Award* in the 4th IBERGRID Conference held in Braga (Portugal) in 2010 (*ISBN 978-84-9745-549-7, pages 212-220*).

Jet substructure and boosted $t\bar{t}$ resonance searches

In chapters 4 and 5 of this thesis, I study jet substructure and its application to searches of new physics with highly boosted top quarks in ATLAS. My work in this subjects is included in several publications and ATLAS notes.

In “*Prospects for top anti-top resonance searches using early ATLAS data*” (*ATL-PHYS-PUB-2010-008, Jul, 2010*) I studied how fat anti- k_t jets and jet substructure observables could be used in resonance searches.

In “*Jet mass and substructure of inclusive jets in $\sqrt{s} = 7$ TeV pp collisions with the ATLAS experiment*” (*JHEP 1205 (2012) 128*) I measured the invariant mass and k_t splitting scales of fat anti- k_t jets and used them to identify boosted top quarks. This was the first time these quantities were measured.

I presented the first boosted top quarks ever seen in the BOOST2011 conference in Princeton University (*ATL-COM-PHYS-2011-485*) and contributed to the paper “*Jet Substructure at the Tevatron and LHC: New results, new tools, new benchmarks*”, *Journal of Physics G: Nuclear and Particle Physics*, 39 (2012), no. 6 063001.

The commissioning of fat jet substructure was crucial for the search presented in chapter 5. I performed many of the supporting studies on jet-related systematics for this analysis. These results were published by the ATLAS Collaboration in *Journal of High Energy Physics (JHEP)*: “*A search for resonances in lepton+jets $t\bar{t}$ events with highly boosted top quarks in 2 fb^{-1} of pp collisions at $\sqrt{s} = 7\text{ tev}$.” (*JHEP 1209 (2012) 041*).*

Contents

Preface	vii
1 The Standard Model and beyond	1
1.1 Beauty, symmetries and physics	1
1.2 Standard Model of particle physics	2
1.3 Symmetry breaking and mass generation	4
1.4 Experimental foundations of the Standard Model	6
1.5 Physics beyond the Standard Model	9
1.6 BSM physics with top quarks	12
2 The discovery machine	15
2.1 The Large Hadron Collider	15
2.2 The experiments	17
2.3 Hadron collider coordinate system	19
2.4 The ATLAS detector	20
2.4.1 The Inner Detector	20
2.4.2 The calorimeters	23
2.4.3 The muon spectrometer	25
2.4.4 The magnets	25
2.5 Performance	26
3 IFIC-Valencia Tier-3 within the ATLAS computing model	29
3.1 ATLAS computing and data distribution model	29
3.1.1 The tiered hierarchy	30
3.1.2 Data management	32
3.1.3 Distributed analysis	34
3.1.4 Computing model evolution	36
3.2 Tier-2 activities in the ATLAS ES-Cloud (Spain and Portugal)	39

3.2.1	Data placement	39
3.2.2	Dataset replication and data access	41
3.2.3	Job distribution	41
3.2.4	Infrastructure and operation	41
3.3	Analysis at local facilities. Tier-3 at IFIC-Valencia	44
3.3.1	Coexistence with other scientific applications running at IFIC-Valencia	47
3.4	First performance tests of a PROOF farm prototype for the Tier-3 at IFIC-Valencia	48
4	Jet substructure	55
4.1	Jets in high energy physics	55
4.2	Jet algorithms	56
4.3	Boosted objects and jet substructure	60
4.3.1	$t\bar{t}$ reconstruction	61
4.3.2	Jet substructure	63
4.4	Commissioning of jet mass and k_t splitting scales	66
4.4.1	Event selection	66
4.4.2	Monte Carlo samples	67
4.4.3	Jet calibration	67
4.4.4	Systematic uncertainties	69
4.4.5	Prospects for multiple pp interactions	70
4.4.6	Jet substructure distributions	72
4.5	Conclusions	74
5	New physics searches in $t\bar{t}$ pairs with boosted topologies	77
5.1	Benchmark models	77
5.1.1	The leptophobic Z' boson	78
5.1.2	The Kaluza–Klein gluon	78
5.2	Data and Monte Carlo samples	81
5.3	Object definition and event selection	83
5.3.1	Physics object selection criteria	84
5.3.2	Event selection	86
5.4	Data-driven background estimates	89
5.4.1	QCD background	89
5.4.2	W +Jets background	90
5.5	$t\bar{t}$ mass reconstruction	91
5.6	Data-MC comparison	93
5.7	Systematic uncertainties	95

5.8	Interpretation. Compatibility with the null hypothesis	99
5.8.1	Limits on benchmark models	99
5.9	Summary	102
Appendices		105
A	First observation of high p_T top quarks with a boosted topology	107
A.1	Introduction	107
A.2	Data and simulated samples	108
A.3	Trigger and selection	108
A.4	Reconstruction of fat jets	112
A.5	Candidate events	113
A.6	Conclusions	114
Bibliography		119
Acknowledgements		135
Resumen		137
	Belleza, simetría y física de partículas	137
	El Modelo Estándar de física de partículas	139
	Fundamentos experimentales del Modelo Estándar	141
	Física más allá del Modelo Estándar	143
	CERN, LHC y ATLAS	146
	El modelo de computación de ATLAS	150
	El Tier-3 del IFIC	152
	Jets en física de partículas	155
	<i>Boosted objects</i> en física de partículas	156
	Subestructura de jets	158
	Búsquedas de nueva física en pares $t\bar{t}$ de alto momento.	159
Un juego de niños		165
Agradecimientos		167

1 The Standard Model and beyond

1.1 Beauty, symmetries and physics

Beauty is not a word people would use to describe physics. An outsider's view of physics consists of a blackboard full of unintelligible equations, Greek symbols and strange diagrams. Physicists, on the other hand, defend themselves by saying that, to the trained eye, those equations, symbols and diagrams become beautiful. Although these two positions may seem irreconcilable, they are not because, in both cases, the ideas of beauty and elegance are often based on those of simplicity and symmetry. Using words of M. Gell-Mann "a theory appears to be beautiful or elegant when it can be expressed concisely in terms of mathematics we already have". A good example of this idea are Maxwell's equations of electromagnetism. It would take a whole page to write down these equations that represent the unification of electricity and magnetism. Using vector calculus they can be written in a more compact form as follows,

$$\begin{aligned}\nabla \cdot \vec{E} &= 4\pi\rho \\ \nabla \cdot \vec{B} &= 0 \\ \nabla \times \vec{E} &= -\frac{1}{c} \frac{\partial \vec{B}}{\partial t} \\ \nabla \times \vec{B} &= \frac{4\pi}{c} \vec{J} + \frac{1}{c} \frac{\partial \vec{E}}{\partial t}\end{aligned}\tag{1.1}$$

We know that these equations are symmetrical under rotations in space. If we rotate the whole space by some angle a given phenomenon of electricity or magnetism won't change. In special relativity, Einstein looked at a whole set of symmetries for electromagnetism called Lorentz's transformations. Using these symmetries and tensor calculus Maxwell's equations became simpler and

therefore more elegant and beautiful.

$$\partial_\nu F^{\mu\nu} = \frac{4\pi}{c} J^\mu \quad (1.2)$$

$$\epsilon^{\alpha\beta\mu\nu} \partial_\beta F_{\mu\nu} = 0 \quad (1.3)$$

But the role of symmetries in physics goes far beyond that of an elegant tool. Symmetries are intrinsically linked to conservation laws through Noether's theorem [1]. In fact, one of the most important results of the past century in particle physics is the description of strong, weak and electromagnetic interactions in terms of gauge field theories. And these field theories emerge from an underlying symmetry called gauge invariance.

One ingredient in the formulation of equations 1.2 and 1.3 is the electromagnetic field strength $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$. The vector potential $A^\mu(x)$ is the quantum field creating and annihilating photons. Both the field strength and Maxwell's equations are invariant under gauge transformations.

$$A^\mu(x) \longrightarrow A^\mu(x) + \partial^\mu \Lambda(x) \quad (1.4)$$

Electromagnetism is a long range force with a massless messenger, the photon. All its properties and the derivation of Maxwell's equations can be written in the form of a simple Lagrangian describing free photons,

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} \quad (1.5)$$

Equation 1.5 is the perfect example of how powerful the concept of symmetry can be in physics. A symmetry such as gauge invariance combined with the use of advanced mathematical tools like tensor calculus allowed not only for an even more elegant representation of the laws governing electromagnetism but also for a deeper understanding of the theory itself.

1.2 Standard Model of particle physics

Schrödinger's equation of a particle of charge q with an electromagnetic field A^μ reads,

$$\frac{1}{2m} \left(-i\vec{\nabla} + q\vec{A} \right)^2 \psi = \left(\frac{i\partial}{\partial t} + qV \right) \psi \quad (1.6)$$

where V and \vec{A} are the scalar and vector potentials that represent the electromagnetic field ($A^\mu = (V, \vec{A})$). Equation 1.6 can be extended to the relativistic case by replacing all derivatives by covariant derivatives $D_\mu = \partial_\mu - iqA_\mu$. Charged fields and covariant derivatives of charged fields have the same local transformations,

$$\psi(x) \longrightarrow U(x)\psi(x), \quad D_\mu\psi(x) \longrightarrow U(x)D_\mu\psi(x) \quad (1.7)$$

As this remains true no matter the order in which we apply successive transformations of the form $U(x) = e^{iq\Lambda(x)}$, we say that $U(x)$ is an Abelian $U(1)$ group.

Weak interactions are different as they give rise to processes like $e^- \longrightarrow \nu_e$ where the identity of matter is changed. In addition to this, only left-handed spin-1/2 fields are involved. The coupling of electrons to photons, the neutral current, is written as $\bar{e}\gamma^\mu A_\mu e$. The charged current can be written as $\bar{E}_L\gamma^\mu W_\mu^\pm \tau^\pm E_L$, where $E_L = \begin{pmatrix} \nu_e \\ e_L \end{pmatrix}$ forms an isospin doublet and γ^μ and τ^\pm are Dirac's spin and Pauli's raising and lowering matrices respectively. The non-Abelian $SU(2)_L$ is the smallest group of gauge transformations that can generalise equation 1.7 acting on E_L . Non-Abelian gauge groups have gauge fields that are self-interacting. This can be seen in the field strength of $SU(2)$,

$$W_{\mu\nu}^i = \partial_\mu W_\nu^i - \partial_\nu W_\mu^i + g\epsilon^{ijk}W_\mu^j W_\nu^k \quad (1.8)$$

A third generator, τ^3 , appears naturally from $SU(2)_L$. Therefore, $SU(2)_L$ predicts 3 gauge fields, W_μ^\pm and W_μ^3 . The gauge field W_μ^3 can only stand for the left-handed part of the electromagnetic current. Another gauge field needs to be included, B_μ , which couples both to $\begin{pmatrix} \nu_e \\ e_L \end{pmatrix}$ and to e_R . B_μ is associated to $U(1)_Y$, where Y is called hypercharge. A superposition of the fields W_μ^3 and B_μ creates the photon and the Z .

For the strong interaction the gauge group is $SU(3)$ as we are counting two quarks, u and d , and each quark has one out of three possible colors. Similarly to the case of leptons, the quarks can be represented with a doublet of $SU(2)_L$, $\begin{pmatrix} u \\ d \end{pmatrix}$ and two singlets u_R and d_R . This group of $SU(3)$ has 8 generators represented by the Gell-Mann matrices T^a and so there are 8 messengers of the strong "color" force, the gluons.

Thus, the three fundamental interactions can be unified under the gauge groups $SU(2)_L, U(1)_Y$ and $SU(3)_c$. To help generalise equation 1.7, three coupling constants are needed, g, g' and g_s . With them, the gauge transformations

and the covariant derivative become,

$$\begin{aligned} \psi(x) &\longrightarrow U(x)\psi(x) = e^{i\theta_3^a(x)T^a} e^{i\theta_2^b(x)\frac{\tau^b}{2}} e^{i\theta_1(x)Y} \psi(x) \\ D_\mu &= \partial_\mu - ig_s T^a A_\mu^a - ig \frac{\tau^b}{2} W_\mu^b - ig' Y B_\mu \end{aligned} \quad (1.9)$$

The information about the interactions of fermions and bosons is encoded in the covariant derivative. The kinetic term of the Lagrangian then reads,

$$\mathcal{L} = \sum_{j=Q_L^k, u_R^k, d_R^k, E_L^k, e_R^k, \nu_R^k} \bar{\psi}_j i\gamma^\mu D_\mu \psi_j \quad (1.10)$$

where $k=1,2,3$ stands for the 3 generations of quarks and leptons.

1.3 Symmetry breaking and mass generation

We know that weak interactions describe short range forces. This means that their associated bosons are massive. The introduction of a term such as Dirac's mass term for fermions, equation 1.11, breaks $SU(2)_L$ because it acts in the same way on both the left- and the right-handed components of the field.

$$-m\bar{\psi}\psi = -m(\bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L) \quad (1.11)$$

One method to generate mass is called hidden or spontaneous symmetry breaking (SSB) [2, 3, 4]. In this mechanism the vacuum state is not invariant under the gauge transformation, while the symmetry is kept in the Lagrangian. This is accomplished by introducing a scalar field with a potential $V = \lambda(|\phi|^2 - v/2)^2, \lambda > 0$. Here, the vacuum expectation value of the scalar field $\langle 0|\phi|0 \rangle = \frac{v}{\sqrt{2}}$ is not 0. In this kind of potentials the most stable state, the one with minimum energy, is not the state with maximum symmetry. When the system goes to the state with minimum energy the symmetry is broken.

In the case of electroweak interactions, we want to break the symmetry so that $SU(2)_L \times U(1)_Y \longrightarrow U(1)_{em}$. To do this, we need three massive bosons and a massless photon. For this, we introduce the following scalar doublet,

$$\Phi = \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}}(v + H) \end{pmatrix} e^{iw^j \frac{\tau^j}{2v}}, \quad j = 1, 2, 3 \quad (1.12)$$

where v is the vacuum expectation value, w^j are Goldstone bosons related to the longitudinal components of W^\pm and Z^0 and H is a Higgs boson. In order

for the physical particle spectra and their interactions to become apparent, we need to rewrite the Lagrangian in the unitary gauge,

$$\mathcal{L}_H = (D^\mu \Phi)^\dagger (D_\mu \Phi) - V(\Phi^\dagger \Phi) \quad (1.13)$$

The masses of the W^\pm and the Z^0 can be derived from the kinetic term that appears in the Lagrangian. W_μ^3 and B_μ combine to generate the massless photon and the Z^0 .

$$\begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} \cos \theta_W & -\sin \theta_W \\ \sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix} \quad (1.14)$$

The angle θ_W is known as weak mixing angle or Weinberg angle and it measures the relation between the masses of Z^0 and W^\pm as follows,

$$M_W = \frac{gv}{2}, \quad M_Z = \frac{v}{2} \sqrt{g^2 + g'^2} = \frac{M_W}{\cos \theta_W} \quad (1.15)$$

From the potential term in 1.13 we can see that the mass of the physical Higgs boson can be identified with $M_H = \sqrt{2\lambda}v^2$. Given that λ is a free parameter, the mass of the Higgs boson can't be predicted from basic principles.

Fermion masses are also introduced in the SM with the Higgs mechanism. The mass terms are introduced via Yukawa couplings y_l . In the case of leptons,

$$-\mathcal{L}_m^{leptons} = \sum_{j=e,\mu,\tau} y_l^j (\bar{L}_j \Phi l_{R,j} + \bar{l}_{R,j} \Phi^\dagger L_j) \quad (1.16)$$

which in unitary gauge writes,

$$-\mathcal{L}_m^{leptons} = \sum_{j=e,\mu,\tau} \frac{y_l^j v}{\sqrt{2}} \left(1 + \frac{H}{v}\right) \bar{l}_j l_j, \quad m_l^j = \frac{y_l^j v}{\sqrt{2}} \quad (1.17)$$

It is remarkable that the couplings of the fermions to the Higgs are proportional to the fermion mass. This happens as well to the quarks with the difference that, unlike with leptons, we have two right-handed singlets. Therefore, the Lagrangian must contain additional terms for u_R and d_R . This fact causes a mixing between the families in the charged current and it is one cause of CP violation in the SM. This mixing is specified in the Cabibbo-Kobayashi-Maskawa (CKM) matrix.

The power of the gauge invariance principle is shown beautifully in the SM. The SM can describe a large number of processes from just a few parameters.

The most used set is formed by the Fermi constant G_F , the electromagnetic coupling α_{em} and the masses of the Z^0 and the Higgs.

$$\begin{aligned} G_F &= \frac{1}{v^2\sqrt{2}}; & \alpha_{em} &= \frac{g^2g'^2}{4\pi(g^2 + g'^2)} \\ m_Z^2 &= \frac{1}{4}(g^2 + g'^2)v^2; & m_H^2 &= 2\lambda v^2 \end{aligned} \tag{1.18}$$

See [5] for a full explanation and [6, 7] for comprehensive reviews on the SM.

1.4 Experimental foundations of the Standard Model

Of course the SM was not born out of symmetry and elegance from the very beginning. The construction of the SM has been a constant interplay between theory and experiment. Often theory has been able to guide experimenters and successfully predict the results of experiments. The discovery of the Higgs boson is a good example of this category of ‘announced discoveries’. On other occasions, however, unexpected discoveries have indicated new roads in particle physics. The classical example of this second category is the discovery of the muon in 1937. Anderson and Neddermeyer were studying cosmic radiation at Caltech when they saw particles that had the same charge of the electron but their curvature in a magnetic field indicated a mass between those of the electron and the proton¹. The confusion the muon produced in the physics community during the decade of the 1930’s is embodied perfectly in I. I. Rabi’s exclamation regarding the unanticipated particle: ‘Who ordered that?’.

Table 1.1 captures the highlights of particle physics in the last century. The column labeled ‘experiment’ lists important observations in experiments. Not shown in the table but equally important are the different techniques and technologies used by the experiments. From the bubble chamber, the diffusion chamber, and the photographic plates to the modern semiconductor detectors and calorimeters, as well as the accelerator technology: magnets, cooling ... all of them played a role without which such observations would have never been possible. The second column in table 1.1 shows major milestones of theory progress.

In the table the story of particle physics starts with the discovery of nuclear β decay by C.D. Ellis and others in 1927. An explanation would not be provided

¹This new particle with the mass of a meson was first thought to be Yukawa’s pion, proposed to explain nuclear binding. The first pion would actually be discovered in 1947.

1.4. Experimental foundations of the Standard Model

Table 1.1: Particle physics time line [8].

Year	Experiment	Theory
1927	β decay discovered	
1928		Dirac: Wave equation for electron
1930		Pauli suggests existence of neutrino
1931	Positron discovered. Chadwick discovers neutron	Dirac realizes that positrons are part of his equation
1933/4		Fermi introduces theory for β decay
1933/4		Yukawa discusses nuclear binding in terms of pions
1937	μ discovered in cosmic rays	
1938	Baryon number conservation	
1946		μ is not Yukawa's particle
1947	π^+ discovered in cosmic rays	
1946-50		Tomonaga, Schwinger and Feynman develop QED
1948-52	First artificial π . K^+ discovered. $\pi^0 \rightarrow \gamma\gamma$. 'V-particles' Λ^0 and K^0 . Δ : excited state of nucleon	
1954		Yang and Mills: Gauge theories
1956		Lee and Yang: Weak force might break parity
1956	CS Wu and Ambles: Yes it does	
1961		Eightfold way as organizing principle
1962	ν_μ and ν_e	
1964		Quarks (Gell-man and Zweig) u, d, s.
1964		Fourth quark suggested
1965		Higgs boson prediction
1967		Colour charge: all particles are color neutral
1967		Glashow, Salam and Weinberg: Unification of electromagnetic and weak interactions.
1968	Homestake experiment: Solar neutrino problem.	
1968/9	DIS at SLAC: constituents of protons seen	
1974		Wess-Zumino: SUSY
1973	Neutral weak current seen at Gargamelle bubble chamber at CERN	QCD as the theory of colour interactions. Gluons
1973		Asymptotic freedom
1974	J/ψ ($c\bar{c}$) meson	
1976	D^0 meson ($\bar{u}c$) confirms theory	
1976	τ lepton	
1977	b quark	
1979	Gluon signature at PETRA	
1981		MSSM
1983	W^\pm and Z^0 seen at UA1	
1984		Composite Higgs scalars
1989	SLAC suggests only 3 generations of light neutrinos	
1995	Tevatron: t quark at 175 GeV mass	
1998	Super-Kamiokande confirms neutrino oscillations	ADD extra dimensions
1999		Randall-Sundrum models
2012	LHC: new neutral boson at 126 GeV mass. Higgs boson?	

until 1934. Fermi's idea of weak interactions as a four fermion interaction evolved during the following two decades. In 1957, Sudarshan and Marshak realized that the results of several electron-neutrino angular correlation experiments on β decays were internally inconsistent and came up with the V-A theory as a solution. Fermi's theory with the weak force being transmitted by a heavy, charged particle could successfully describe all weak interaction data available but was still incomplete. The predicted neutrino cross sections violated unitarity for center of mass energies above 300 GeV and the theory could not be calculated to higher order. On top of that, no W boson with a mass up to 20 GeV had been found at that time. A complete explanation came with the unification of electromagnetic and weak interactions in 1967 [9]. The theory of Glashow, Weinberg and Salam was based on massless, Yang-Mills particles as carriers of the weak interaction. The carriers acquired mass through the Higgs mechanism. Compared to Fermi's theory where only charged weak bosons were exchanged, this theory also predicted the existence of processes where no charge was transferred. The neutral weak current was discovered in 1973 using the Gargamelle bubble chamber at CERN.

The following decades would bring the discovery of a whole zoo of particles. This chaotic atmosphere favored brilliant ideas such as Gell-Mann's quark model and the development of QCD. Strangeness suggested the idea of a new family of quarks that would be confirmed by the discoveries of the J/ψ meson, in SLAC and BNL in 1974, and the D^0 meson in 1976. Later that year, the discovery of the τ lepton suggested the existence of yet another family of particles. The b quark would be seen in Fermilab one year later, in 1977.

Since then experimental results continued to support theoretical predictions. The discovery of the W^\pm and the Z^0 required a huge effort. The Super Proton Synchrotron was originally installed at CERN to accelerate protons and it was latter converted into a $p\bar{p}$ collider. The weak force carriers were finally seen at the UA1 experiment in 1983. Twelve years later, in 1995, the last and heaviest quark was finally seen. The Tevatron, a $\sqrt{s} = 1.96 \text{ TeV } p\bar{p}$ collider built in Fermilab, was the first to discover the top quark and measure its mass with high precision.

Only one boson was left unseen and 30 years would pass until the first glimpse of the Higgs boson appeared. The discovery of a $\sim 125 \text{ GeV}$ neutral boson was announced on 4 July 2012 by the LHC experiments [10, 11]. If it is the Higgs boson, this discovery will corroborate a prediction that has waited around 50 years to be confirmed.

Even though nearly all the particles predicted by the SM have been discovered and its predictions have been accurately confirmed, there is still room for surprise.

More extensive expositions of the experimental foundations of the SM can be found in references [12, 13].

1.5 Physics beyond the Standard Model

Even if the Standard Model is highly successful in its description - to very good accuracy - of a broad range of phenomena, physicists continue to search for a better, more complete model of particles and their interactions.

One motivation is the search for an explanation for the values of the parameters of the theory. In the fermionic sector the free parameters include the masses of the $3n$ fermions and the $(n - 1)^2$ parameters in the CKM matrix. Assuming there are 3 generations of fermions ($n=3$) and adding the strong coupling, g_s , the number of free parameters in the SM totals 18. Ideally, one would want all these parameters to be derived from first principles. A ‘Theory of Everything’ able to explain all physical phenomena is for many too far away, an impossible utopia. The best answer to those was given by the Argentinian film maker Fernando Biri [14].

“What is the use of utopia? What is the use of a goal so far away that no matter how much one walks it will never be reached? It is just for that, utopia is useful because it makes us walk”

The first challenge to mention can not be other than clarifying the *electroweak symmetry breaking* (EWSB). Even though the newly discovered state is compatible with the Higgs boson there is still room for surprise. The new boson does not have to be exactly the one that Higgs and others predicted. It can be the lightest of several Higgs-like bosons, it can be composite, etc.

The *gauge hierarchy problem* refers to the fact that gravity is extremely weak compared to the other forces. Another way to look at it is through the fact that the Higgs boson is so much lighter than the Planck mass ($M_{Planck} = 2.4 \times 10^{18} GeV$). This is so because radiative corrections to the Higgs boson mass are at the scale at which new physics (ie. gravitation) appear, and therefore the Higgs boson mass is expected to be comparable to this scale. Gravitational interactions become relevant at energies near the Planck scale. If we want to extrapolate the SM to such energies, while keeping the electroweak scale ($v = 246 GeV$) compatible with the observed values and a light Higgs boson, we need an extreme fine-tuning of the model parameters. Although fine-tuning is theoretically consistent, many people would rather have a more elegant solution.

There are many questions waiting for an answer. Dark matter and neutrino oscillations are experimentally firmly established, but have no explanation in the SM. The cosmological constant or quantum gravity are challenges we have still to find an answer for. Here we will focus on the questions that will most likely be answered with the data from the Large Hadron Collider (LHC) (Sec. 2.1). For a full discussion of physics beyond the SM (BSM) in the LHC era see [15, 16, 17].

If *dark matter* [18] particles are the source of the anomalies observed in large astrophysical systems [19] quite possibly they can be produced at the LHC and discovered in its experiments. The other possible explanation to these anomalies would be to assume a deviation from the known laws of gravitation and the theory of general relativity [20].

Explaining the dominance of matter over anti-matter, *baryon asymmetry* [21, 22], is another of the challenges waiting for an answer. Although any CP violation would work, it is also known as the *strong CP problem* because the CP-violating phase in the CKM matrix is the only place where there is a distinction between matter and anti-matter in the SM. Unfortunately, this phase does not seem to be enough to explain the observed asymmetry.

One of the most notorious anomalies at previous colliders is the $t\bar{t}$ forward-backward asymmetry measurement at the Tevatron. With an observed value of $A_{FB}^{exp} = 0.20 \pm 0.07_{stat} \pm 0.02_{sys}$, it is much larger than it should be within the SM, $A_{FB}^{SM} = 0.087 \pm 0.010$ [23]. At the LHC, with a symmetric initial state, no $t\bar{t}$ forward-backward asymmetry can be generated. However, top quarks are usually emitted in the forward and backward detectors following the incoming valence quark direction while anti-top quarks remain more central. The last public measurements of the charge asymmetry are compatible with the SM prediction [24]. This results as well as direct resonance searches disfavour a BSM explanation. However, more data are needed to give a conclusive answer.

In view of the experimental and theoretical shortcomings of the SM, there has been a steady flow of proposals for new TeV-scale physics over the last six decades.

One of the most well known BSM candidates is supersymmetry (SUSY). SUSY relates fermions to bosons in such a way that for every boson in the SM there is a corresponding fermionic superpartner and vice-versa [25]. It is a renormalizable perturbative theory, which means it allows for accurate predictions. The fine-tuning problem is avoided due to the fact that loops of superpartners cancel quadratic corrections to the Higgs boson mass. SUSY can explain the underlying cause of EWSB and offers viable candidates for dark matter and the source of baryon asymmetry. On top of that, SUSY suggests

the unification of all gauge interactions at high energies. All these qualities have made SUSY the focus of both theorists and experimentalists since the LEP era. However, no sign of it has been found so far, not even in the latest searches at the LHC.

BSM models with extra dimensions were introduced to explain the hierarchy between gravity and the other interactions. Quadratic divergences pull m_{Higgs} to a higher scale. The SM starts failing at M_{Planck} since at this scale gravitation becomes a quantum theory. Instead of fine-tuning the SM parameters in order to keep the Higgs boson light, the hierarchy problem can be solved by bringing down the Planck scale itself. Since $G_N \sim 1/M_{Planck}^2$, the size of the Planck scale might be connected to a too weak gravitation. This can be fixed by adding to the geometrical setup of the theory more dimensions where gravitation is stronger [26]. The fact that the extra dimensions have not been detected so far implies that such dimensions should be compact and have a finite extent. Particles allowed to propagate in such compact extra dimensions will show a pattern of Kaluza-Klein towers of heavy copies with masses separated by the inverse of the size of the extra dimension.

These theories were first introduced in order to explain the mismatch between gravity and the rest of the fundamental interactions. For instance, theories with large extra dimensions try to minimize the changes in the SM by allowing only gravity to propagate in the extra dimensions. The initial motive lost presence later and variations appeared, like universal extra dimensions theories, which allow all SM particles to propagate in the extra dimensions. Warped or Randall-Sundrum (RS) models provide an explanation of the large hierarchy between the electroweak scale and the Planck scale by introducing extra dimensions with an intrinsic curvature [27, 28].

In composite Higgs models the hierarchy problem is solved by assuming that the Higgs is not fundamental but a composite particle. In the same way that light pions appear as (pseudo-)Goldstone bosons of the spontaneously broken chiral symmetry, a light Higgs is assumed to appear as a (pseudo-)Goldstone boson of a new spontaneously broken chiral symmetry. In this scenario, the new gauge prevents the quadratic divergence of the Higgs mass [29, 30]. Although this approach has been strongly constrained experimentally, it has not been completely ruled out.

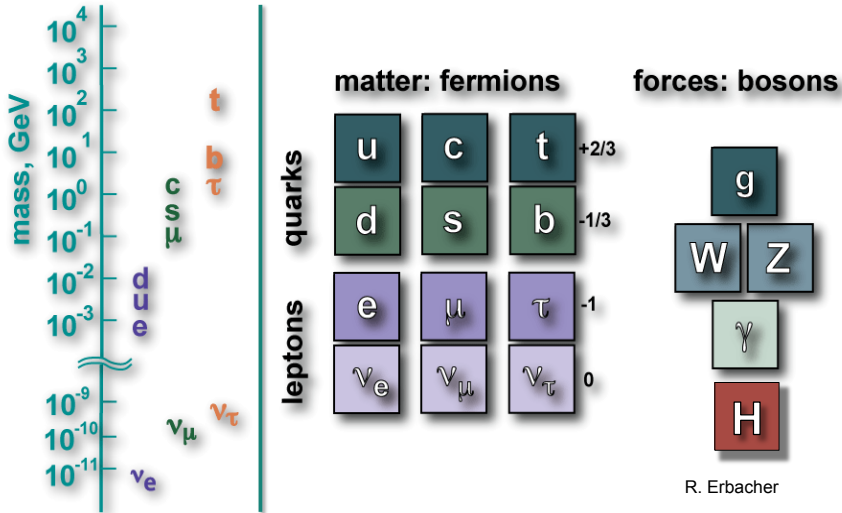


Figure 1.1: Mass hierarchy and periodic table of SM fundamental particles.

1.6 BSM physics with top quarks

All the fundamental particles of the SM are shown in figure 1.1. The top quark is the most massive of them by far and should therefore couple most strongly to the Higgs boson. Even when the top quark is virtual, and it only appears in loop corrections, its presence still produces large effects. It plays a special role in EWSB, quadratic divergences to Higgs mass are nearly completely due to top quark.

The properties of the heaviest quark are of great importance in many of the proposed theories for BSM physics as well. Top quark loop corrections to Higgs mass are largely canceled by the stop loop corrections in SUSY models. The top quark is present in corrections without which the lightest Higgs could not be heavier than a Z boson in MSSM models. The top quark drives radiative EWSB in SUSY and, as heavy higgses may decay to top quarks, it can also be used to determine their properties.

Some RS scenarios allow SM fermions and gauge bosons to propagate into the extra dimensions. Resonant production of the KK excitations of the SM particles should be the first signal of such scenarios. Among them KK gluons

have the largest production rate and they decay primarily to $t\bar{t}$ resonances [31]. Different scenarios can be distinguished by looking at angular distributions of the leptons from the top quark decay.

In some models based on compositeness such as top-color assisted technicolor the Higgs boson appears as a $t\bar{t}$ bound state. A large class of these models has already been excluded by LHC data as a 125 GeV Higgs boson is not generally predicted by these theories.

Quarks benefit from the strong coupling and have a large production cross-section at the LHC compared with colourless particles. Its large mass causes the top quark to decay before it can hadronize. It is therefore possible to access its spin states. This implies that a detailed study of top quark properties could bring new phenomena to light.

Top quarks can appear alone via EW (single-top) production or they can appear in pairs via strong production. At the LHC, top quark pairs are mainly produced through gluon fusion and, with less probability ($< 15\%$, depending on \sqrt{s}), through $q\bar{q}$ fusion (see figure. 1.2).

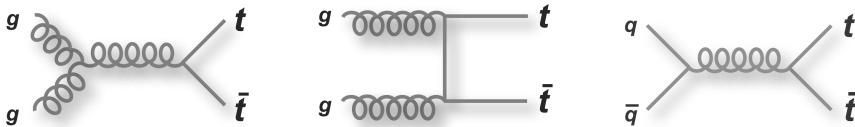


Figure 1.2: Feynman diagrams representing the top quark pair production at LO [32].

The top quark can only decay through the weak interaction into a W boson and a down-type quark, which is a b quark nearly 100% of the times. Therefore, $t\bar{t}$ final states can be classified depending on how the W boson decays as seen in figure 1.3.

The study of the top quark has been a challenge since its start and has brought about the usage of new tools and techniques that are now common in the high energy physics community. The decay of top quark pairs leads to a complex final state. Its decay products are reconstructed as jets, missing energy and leptons. New techniques had to be developed in order to deal with complex final states with significant missing energy and taggable particles. By tagging the b-initiated jets and the lepton from the W decay, top quarks can be easily identified and distinguished from anti-top quarks.

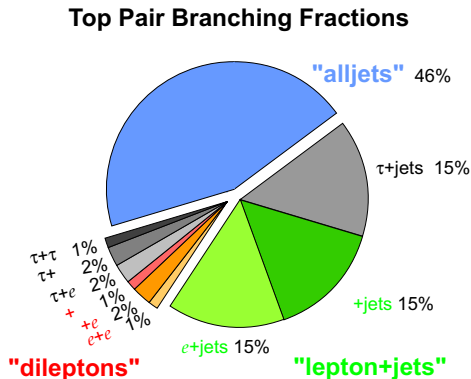


Figure 1.3: Top quark pair branching fractions [32].

It is the quark which reveals more information about its production and decay processes because it decays before it can hadronize. We can look for anomalies in production cross sections such as top quark pairs produced via unknown mediators (resonance search). We can measure its mass, width, lifetime, polarization and charge with precision. Anomalous couplings, CP violation or non-SM decays/branching ratios might appear in the $t \rightarrow Wb$ process. The lepton carries information about the top quark production kinematics and can give away information on polarization, spin correlations, W helicity, flavour changing neutral currents (FCNC), etc.

Due to its important role in most of the ‘best-seller’ theories in the market and the amount of information that can be accessed through its study, the top quark might be the most promising handle to look for BSM at the LHC. It might lead to the discovery of new particles, its properties might turn out to be anomalous or it can even happen that we find new physics that mimics top quark signatures. The SM can be tested thoroughly through the study of top quark properties. It is the quark which is most strongly coupled to the Higgs boson and its study will most probably shed more light into the EWSB mechanism. All these qualities make the top quark a key piece in the physics program of the LHC experiments.

2 The discovery machine

2.1 The Large Hadron Collider

Since its foundation in 1954, the history of the *Conseil Européen pour la Recherche Nucléaire* (CERN) [33] has been dedicated to particle accelerators. The first accelerator hosted by what is now the largest particle physics laboratory of the world was a 600 MeV SynchroCyclotron (SC) in 1957. In 1971, the Intersecting Storage Rings (ISR) would produce worlds first protonproton collisions. Years later, in 1983, the W and Z bosons would be discovered at the Super Proton Synchrotron ($Spp\bar{p}S$) confirming the unification of electromagnetic and weak theories. In the 90's, the most powerful lepton collider ever built, the Large Electron-Positron Collider (LEP), allowed precision measurements of the masses of the W and Z bosons, among other important SM values.

Assembled in the 27 km tunnel where LEP used to be, the LHC [34] is the most powerful particle accelerator ever built. At full energy operation, the LHC will collide proton beams at 7 TeV each. This is scheduled to happen by the beginning of 2015 after a 20 month shutdown. In 2010 and 2011, the accelerator has been working at 3.5 TeV per beam, and at 4 TeV per beam in 2012. Heavy ion collisions, typically of lead nuclei, were produced as well.

The LHC project has constituted an enormous engineering challenge. In a typical run, the two beams circulating in opposite directions carry 2808 bunches of 10^{11} protons each. The beams cross each other in 4 points where different detectors are placed producing, on average, over 20 collisions every 50 ns (figure 2.1). When working at full power, the proton beams carry 724 MJ of energy. To bend such energetic beams, 1232 superconducting dipoles (figure 2.2) are deployed along the LHC tunnel and another 600 quadrupoles and sextupoles maintain the beams focused and stable. A 8.3 Tesla field is needed to achieve the required bending power. The coils are made of NiTi superconducting cable, kept at 1.9 K by superfluid liquid He. They are 15 m

long, weigh 35 tonnes and store in their magnetic field a total of about 10 GJ in the full ring.

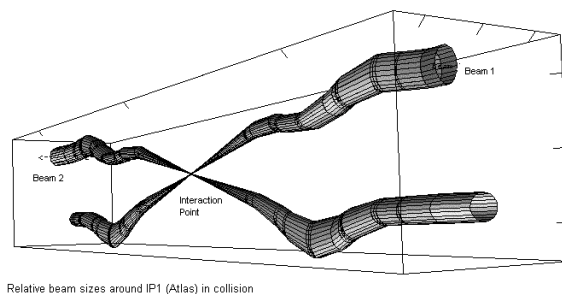


Figure 2.1: Relative beam sizes at ATLAS interaction point (IP1) [34].

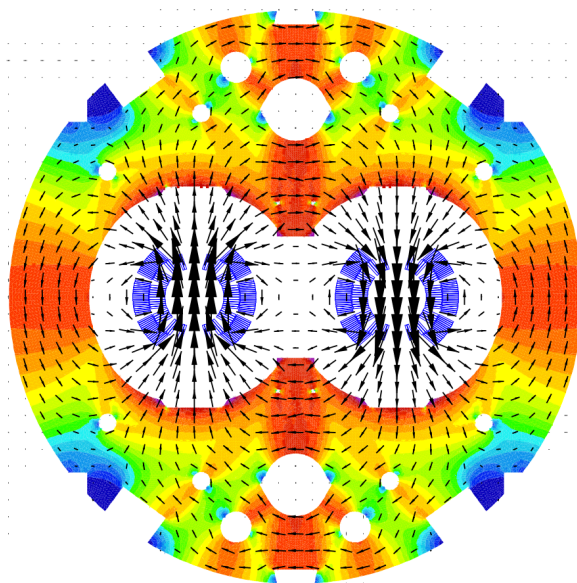


Figure 2.2: LHC superconducting dipole 2-in-1 configuration [34].

About 1 nanogram of protons circulates per day in a typical LHC fill. In their trip from the hydrogen container to the LHC the protons travel not only through the accelerator complex shown in figure 2.3, but also through CERN's history. The protons extracted by ionizing hydrogen are accelerated to 50 MeV in the Linac2. From there, they are led to the Proton-Booster Synchrotron (PSB) where its energy is increased to 1.4 GeV. Afterwards, the beam reaches 25 GeV in the 54 year old PS and 450 GeV in the 37 year old SPS before being injected in the LHC where they will finally reach up to 7 TeV.

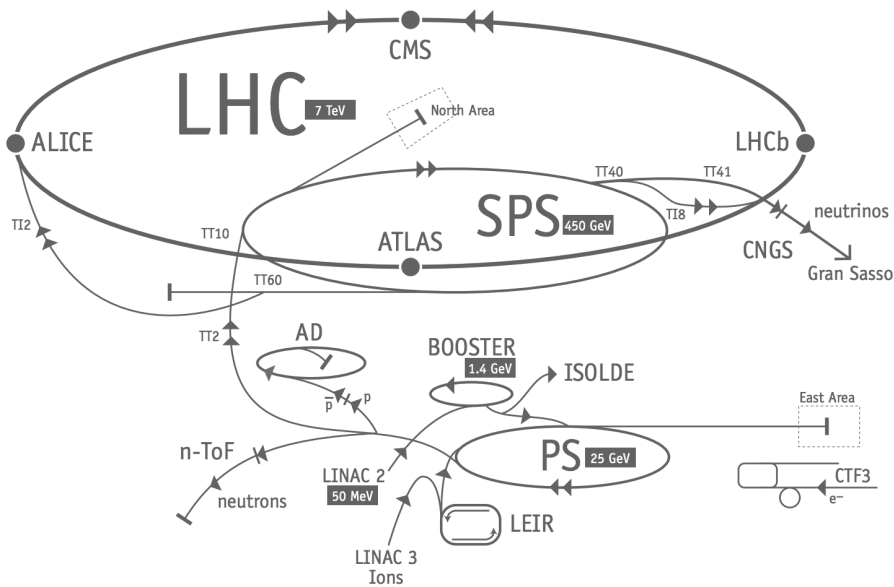


Figure 2.3: Layout of the full CERN accelerator complex, including all elements of the LHC injector chain [34]. The four interaction regions hosting the main LHC experiments, ALICE, ATLAS, CMS and LHCb, are also shown.

2.2 The experiments

Four large experiments are installed in the four interaction points (figure 2.4). In addition to this, other smaller experiments are placed along the ring [35].

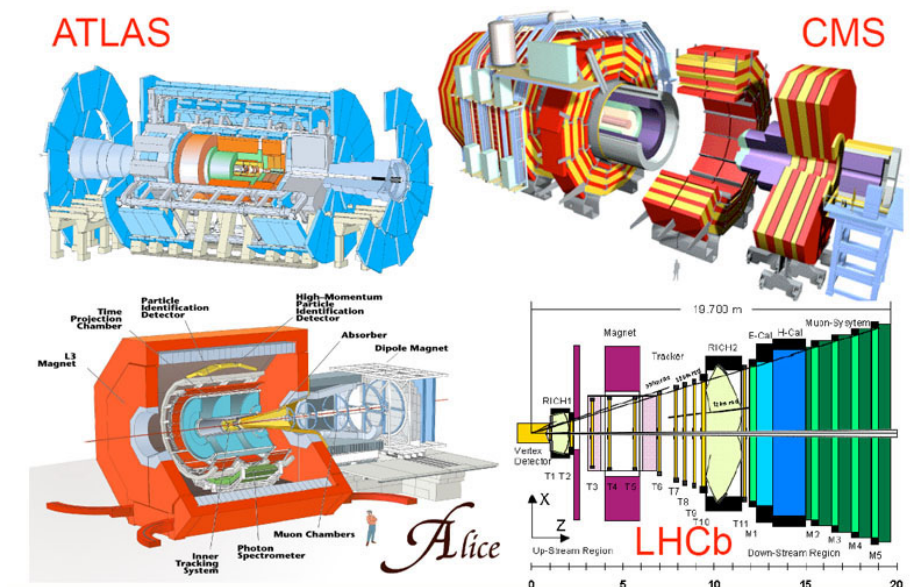


Figure 2.4: Detectors deployed at the four LHC interaction points [35].

ATLAS (section 2.4) and CMS are general purpose experiments. They both have an inner tracker, electromagnetic and hadronic calorimeters and a muon spectrometer. The biggest difference lies in the magnet system. While ATLAS has a 2 T solenoid covering the inner detector and a 2-8 T/m toroid for the muon spectrometer, CMS uses a single 4 T non-linear magnetic field.

LHCb is interested in CP violation and rare decays in hadrons with heavy flavors such as B hadrons. It is designed to study particles that emerge close to the beam direction. For this it incorporates a vertex detector for tracking and a RICH detector for particle identification.

ALICE is designed for the heavy ion collision program. It has an inner tracker and a TPC in the central part and a muon spectrometer in the forward direction. With this setup ALICE studies the physics of the strong interacting matter at extreme densities.

The aim of the TOTEM experiment is to measure the total pp cross-section and study elastic and diffractive scattering at the LHC. It is installed close to the CMS collision point and has silicon sensors to measure the protons 200 m

before and after CMS.

LHCf studies neutral-particle production cross sections in the very forward region of both pp and heavy ion collisions. Combined with the results from TOTEM, these measurements will increase our understanding of the development of atmospheric showers induced by ultra high energy cosmic rays.

Approved by the CERN Research Board in December 2009, the MoEDAL experiment consist of layers of plastic covering the vertex detector of LHCb. Its purpose is to detect traces of stable particles such as a magnetic monopoles or a massive stable supersymmetric particles.

2.3 Hadron collider coordinate system

It is natural to use cylindrical coordinates to describe the detectors. The beam direction is usually defined as the z axis, the x axis pointing to the center of the LHC ring and the y axis pointing upwards. Then, the *rapidity* y is defined as follows,

$$y = \frac{1}{2} \ln \frac{E + P_z}{E - P_z} = \tanh^{-1} \frac{P_z}{E} \quad (2.1)$$

Differences of rapidity are invariant under a Lorentz boost in the z -direction to a frame with velocity β ,

$$y \longrightarrow y - \tanh^{-1} \beta \quad (2.2)$$

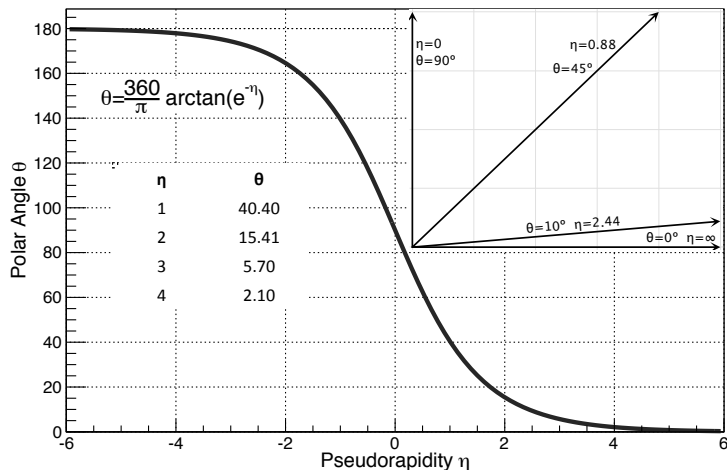
When $p \gg m$ the definition can be simplified (equation 2.3) using $\cos \theta = P_z/P$,

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

This is the so called *pseudorapidity* whose relation with the polar angle θ is shown in figure 2.5.

Another common practice in the hadron collider community is the measurement of distances in the $\eta - \phi$ space. This is usually represented as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

In every collision, an unknown part of the energy of the incoming protons escapes through the beam pipe. Due to this fact, momentum conservation

Figure 2.5: Relation between pseudorapidity η and polar angle θ .

criteria can only be applied to the plane transverse to the beam direction (x-y plane). Quantities constrained to the transverse plane appear very often in LHC analyses. Well known examples are the *missing transverse energy* (E_T^{miss}) or the *transverse mass* (M_T).

2.4 The ATLAS detector

ATLAS [36] (A Toroidal LHC Apparatus) is one of the two general purpose detectors at the LHC. Situated at Point1, directly opposite to CERN main entrance, the detector is approximately a cylinder with a total length of 42 m and a diameter of 25 m and weights approximately 7000 tonnes. The four major components of the detector appear in figure 2.6: the inner tracker, the calorimeters, the muon spectrometer and the magnet system.

2.4.1 The Inner Detector

Immersed in a solenoidal magnetic field of 2 T, the Inner Detector (ID) measures the momentum of the charged particles. It covers the full range in ϕ and the range $|\eta| < 2.5$ and it is composed by three sub-detectors shown in figure 2.7. The barrel part of the Pixel Detector consists of 3 layers. The pixels are

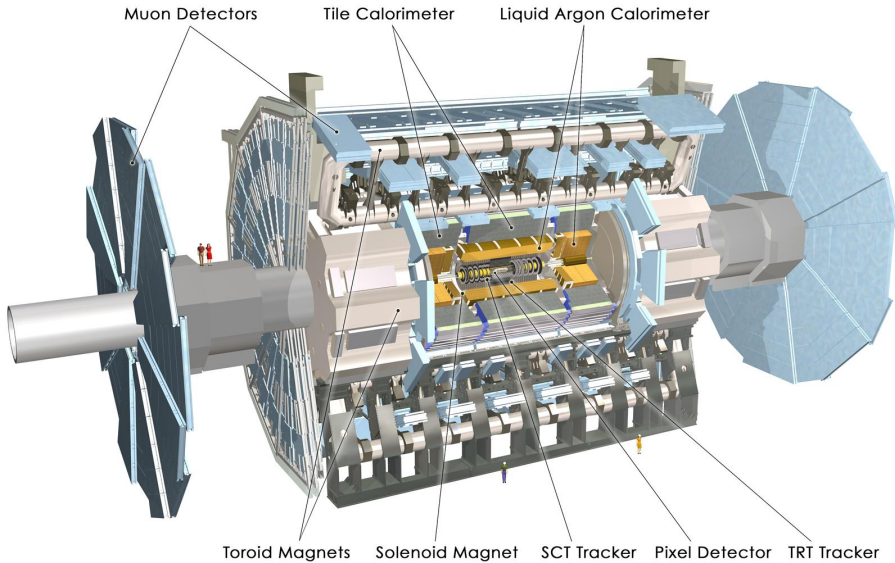


Figure 2.6: A detailed computer generated image of the ATLAS detector and it's systems [37].

$50 \times 400 \mu\text{m}$. Each silicon wafer has 46,080 pixels totalling around 80 million readout channels in the Pixel Detector. The forward and backward regions are covered by disks equipped with wedge-shaped sensors. This design allows for a precise measurement of impact parameter and vertex position.

The Semiconductor Tracker (SCT) has 4 double layers of silicon microstrip detectors in the barrel and 9 disks in the endcaps. The $80 \mu\text{m} \times 12 \text{ cm}$ strips are more practical in order to cover a larger area. In the outer part of the ID, the Transition Radiation Tracker (TRT) is a combined straw tracker and transition radiation detector. The TRT provides two-dimensional measurement points with 0.170 mm resolution for charged particle tracks with $p_T > 0.5 \text{ GeV}$. It contains 52544 axial straws of about 150 cm length and 4 mm diameter in the barrel part and a total of 319488 radial straws in the endcaps.

Tracking is performed by combining hits from the three sub-detectors using two different algorithms. An inside-out algorithm reconstructs most primary tracks. An outside-in algorithm starts from the TRT and goes inwards by

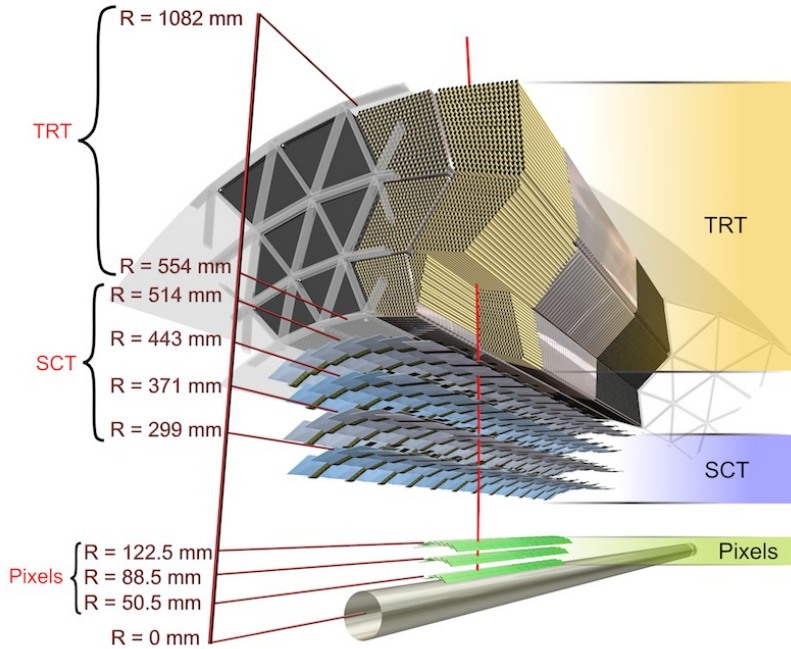


Figure 2.7: Computer generated image of the ATLAS Inner Detector [37].

adding silicon hits. The outside-in algorithm reconstructs most secondary tracks coming from conversions, hadronic interactions or V_0 decays. Primary vertices are reconstructed using an iterative χ^2 fit.

Table 2.1: Intrinsic resolution of the three ID sub-detectors [38].

Pixel Detector ($R\phi/z$)	SCT ($R\phi/z$)	TRT ($R\phi$)
10 μm / 115 μm	17 μm / 580 μm	130 μm

The intrinsic resolutions of the three sub-detectors are shown in table 2.1. The track momentum resolution is affected by irregularities in the features of the

tracking system and the magnetic field. In reference [38] the track momentum resolution in the barrel region is given for muons with p_T up to 1 TeV in a solenoidal field, without a beam constraint by the following approximation,

$$\sigma\left(\frac{1}{p_T}\right) = 0.36 \oplus \frac{13}{p_T \sin^{1/2} \theta} (\text{TeV}^{-1}) \quad (2.4)$$

where the second term is due to multiple scattering. Therefore, $\sigma(\frac{1}{p_T})$ is assumed to be constant in absence of multiple scattering or when its effects are negligible.

The transverse impact parameter (d_0) of a given track is the closest distance from the track to the primary vertex in the transverse plane. The resolution of d_0 is limited by two main factors. The intrinsic resolution (and misalignment) of the tracker is the dominant factor at high p_T while the effect of multiple-scattering in the detector material dominates at low p_T . The d_0 resolution can be approximated in the barrel region of the ID by equation 2.5, where b is a constant.

$$\begin{aligned} \sigma_{d_0}^{track} &= \sigma_{intrinsic} \oplus \sigma_{MS} \\ &= \sigma_{intrinsic} \oplus \frac{b}{p_T \sin^{1/2} \theta} \\ &\sim 11 \oplus \frac{73}{p_T \sin^{1/2} \theta} (\mu\text{m}) \end{aligned} \quad (2.5)$$

2.4.2 The calorimeters

ATLAS calorimeters [38] are shown in figure 2.8. Energy measurements for electrons, positrons and photons in the barrel region, $|\eta| \lesssim 1.5$, are provided by a sampling electromagnetic (EM) calorimeter with an accordion-shaped structure of lead as absorber and an active medium of liquid argon (LAr) that gives its name to the system. The LAr granularity is $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$.

Hadronic calorimetry is provided by a scintillating-tile detector with steel as absorber in the central region ($|\eta| < 1.7$) where it reaches a granularity of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$.

In the endcaps ($1.5 \lesssim |\eta| < 3.2$) EM calorimetry is provided by a second LAr calorimeter with an accordion-shaped lead absorber. Forward hadronic calorimetry uses LAr as well but with flat copper layers as absorber.

In the very forward region, closer to the beam pipe ($3.1 \lesssim |\eta| < 4.9$), another LAr system is deployed. For EM calorimetry copper is used as absorber while tungsten is the choice for hadronic calorimetry.

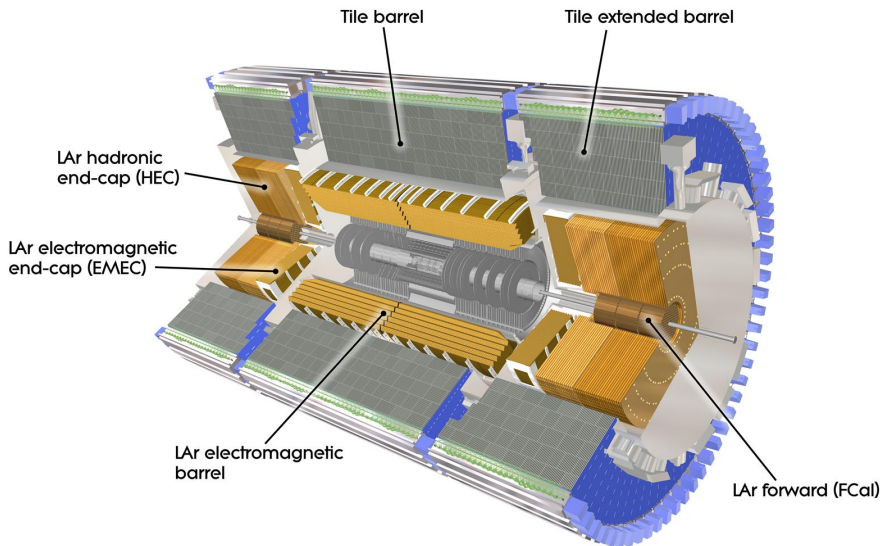


Figure 2.8: Computer generated image of the ATLAS calorimeter [37].

The energy resolution of a calorimeter [39] is typically parametrized as

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus b \oplus \frac{c}{E}, \quad (E \text{ in } GeV) \quad (2.6)$$

The constant a is usually referred to as the stochastic term and represents statistics-related fluctuations such as intrinsic shower fluctuations, photoelectron statistics, dead material in front of the calorimeter and sampling fluctuations. The systematic term b accounts for detector non-uniformities, calibration uncertainty and radiation damage of the active medium. The term c is due to electronic and pile-up noise. The energy resolution of the ATLAS EM calorimeter in the barrel region is,

$$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus 0.5\%, \quad (E \text{ in } GeV) \quad (2.7)$$

The energy resolution in equation 2.7 was measured after electronic noise suppression [36], hence the absence of the third term.

The combined energy resolution of the barrel LAr and tile calorimeters was measured using charged pions in the energy range between 10 and 300 GeV [36].

$$\frac{\sigma_E}{E} = \frac{52\%}{\sqrt{E}} \oplus 3\% \oplus \frac{1.6\%}{E}, \quad (E \text{ in GeV}) \quad (2.8)$$

The position resolution depends on the width of the EM showers generated by the incident particles and on the transverse granularity of calorimeter. The Molière radius, $\sim 2.5 \text{ cm}$ for the ATLAS EM calorimeter, gives a good scale of the width of the EM showers. If two particles enter the calorimeter separated by a distance of order the effective Molière radius of the material their showers overlap and it is not possible to identify them separately no matter how good the granularity is. In the EM calorimeter the position resolution has been tested using electron and photon and pion beams. At a beam energy of 245 GeV, the measured values along η lie between 1.5×10^{-4} and 3.3×10^{-4} (in units of η), in ϕ the resolution is in the range $50 - 60(\text{mrad})/\sqrt{E(\text{GeV})}$ over the whole coverage (barrel and end-caps).

2.4.3 The muon spectrometer

The outermost layer of ATLAS is occupied by the muon spectrometer (figure 2.9) [38]. The system achieves a resolution of 10% for the determination of the momentum of 1 TeV muons. This is accomplished by using drift tubes and an optical alignment system that can correct possible chamber deformations within $30 \mu\text{m}$. Hence the name of the detector is Monitored Drift Tubes (MDT). Cathode Strip Chambers (CSC) are used in the parts where a more radiation tolerant technology is required.

2.4.4 The magnets

Two superconducting magnet systems are used in the ATLAS detector [38]. The solenoid surrounding the Inner Detector is designed to generate a 2 T magnetic field in order to curve the high energy charged particles produced in the collisions. The toroid magnet system consists of eight large superconducting loops in the barrel and two toroidal magnets in the endcaps. Situated within the muon spectrometer, the toroidal field varies from 2 to 8 T/m and it stores up to 1.6 GJ of energy.

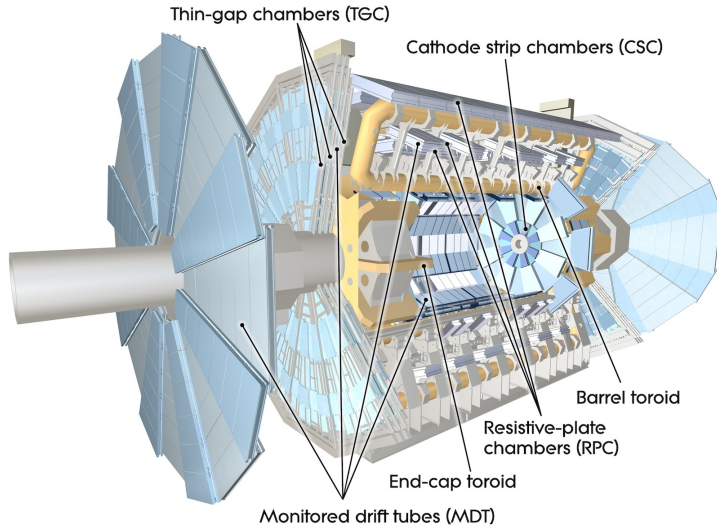


Figure 2.9: Computer generated image of the ATLAS muon spectrometer components [37].

2.5 Performance

Both the LHC and ATLAS have been performing well beyond expectation since the start of the data taking by the end of 2009. Since then, the LHC has delivered 27 fb^{-1} , where 23 fb^{-1} of 8 TeV data were recorded in 2012 (figure 2.10). With a data taking efficiency higher than 95% and more than 99% of its channels working, ATLAS supplies data with an unmatched quality. In order to analyse the data, the ATLAS collaboration has designed a distributed computing model based on GRID technologies. The ATLAS computing model and its evolution since the start of the LHC is discussed in section 3.1.

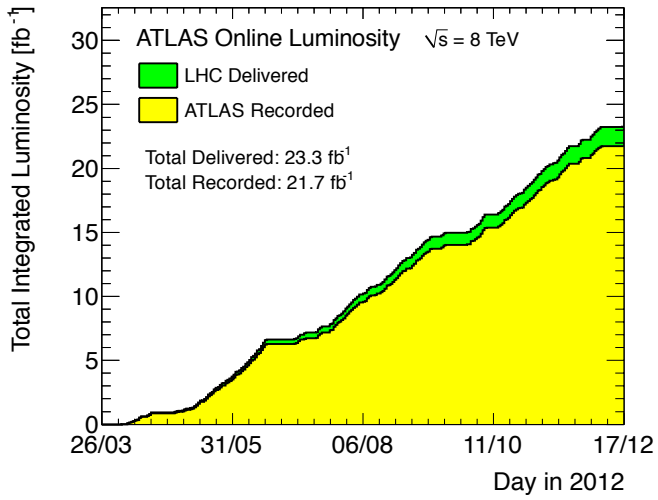


Figure 2.10: Cumulative luminosity versus day delivered to (green), and recorded by ATLAS (yellow) during stable beams and for pp collisions at 8 TeV centre-of-mass energy in 2012 [40].

3 IFIC-Valencia Tier-3 within the ATLAS computing model

3.1 ATLAS computing and data distribution model

Since the LHC data taking started in November 2009, several thousands of millions of collision events have been recorded by the ATLAS experiment. This translates, as it can be seen in figure 3.1, into around 140 PetaBytes of ATLAS data and MC simulations being stored and analyzed by nearly 3000 physicists from 174 institutions spread all over the world.

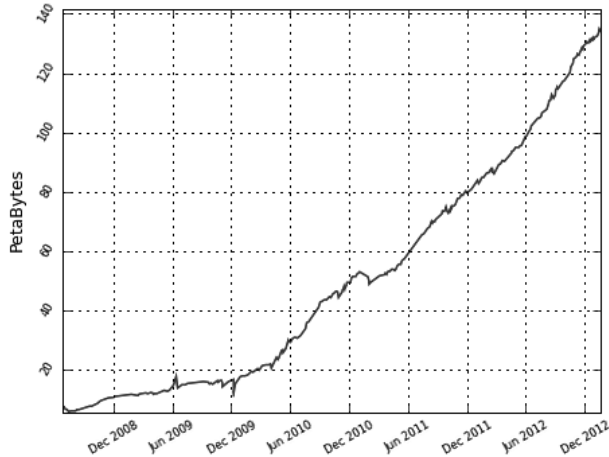


Figure 3.1: Total disk space usage according to DQ2.

Faced with the difficulty that entails analysing such amount of data in a single place or even with the combined effort of several computing centres, the LHC experiments chose a model with a high degree of decentralization and the possibility of sharing resources, a model based on Grid technologies.

The term “Grid computing” originated in the early 1990s. In the same way electric power was easy to access thanks to the electric power grid, computing power would be as accessible thanks to the computing grid. The power grid metaphor spread after Ian Foster and Carl Kesselman published their seminal work, “The Grid: Blueprint for a new computing infrastructure” [41] in 2004. A Grid can be defined as an infrastructure that enables sharing, selection and aggregation of geographically distributed ‘autonomous’ resources dynamically at runtime depending on their availability, capability, performance, cost and user’s requirements.

Based on the Grid paradigm, the WLCG project (Worldwide LHC Computing Grid) [42] gives support to the computing models of the LHC experiments. Strictly following the MONARC project proposal [43], it groups the different types of computing centers of the ATLAS collaboration in a tiered hierarchy that ranges from Tier-0 at CERN, down to the 11 Tier-1 centers and the nearly 80 Tier-2 centers distributed world wide.

The ATLAS Computing Model [44, 45] handles everything from the storage of raw events at CERN to physics analysis on refined data at home institutes. The ATLAS Collaboration has established the responsibilities of the various centres of the different countries for the reliable operation of the computing facilities. A proper sizing and organization of the resources, an efficient mechanism to access the data or robust algorithm development are examples of key items in the so-called solution for the steady-state period during the data taking.

3.1.1 The tiered hierarchy

The tier structure shown in figure 3.2 is managed centrally by the ATLAS Collaboration. Even though they are out of the ATLAS pledge, the end-user private analysis facilities (Tier-3) also play an important role in the model.

Located at CERN, the Tier-0 is in charge of the first data processing. The initial design was found to underestimate the disk capacity needed by the Tier-0 to move the data in from the experiment and out to the tiers. For this reason, extra resources were placed at CERN in order to perform the initial calibration and alignment, as well as monitoring of the detector. For the preliminary understanding of the data, large scale access to raw data had to be provided as well. The Calibration and Alignment Facility (CAF) also hosts build machines,

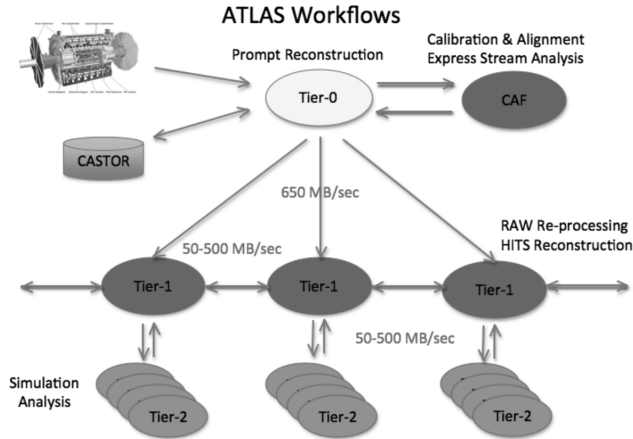


Figure 3.2: The ATLAS tier structure.

machines for the automated software testing, etc. which are needed for the basic operation of the experiment offline computing. In May 2012 CERN signed a contract for an extension to its data centre with the Wigner Research Centre for Physics in Budapest. The Wigner Research Centre will host CERN equipment that will substantially extend the capabilities of the LHC Computing Grid Tier-0 activities. The Tier-0 will be able to transfer critical functions to the Wigner Centre, mitigating the risk of having all of Tier-0 in one location.

A portion of the raw data is sent to the 12 Tier-1 facilities which are responsible for its long-term organization and safety and for reprocessing it to derived formats. The reprocessing takes place either when better calibrations are available or at the end of data taking, when algorithmic improvements can be applied. In normal conditions, this is expected to happen twice a year although the frequency was higher, around 5 times, for the initial data taking. In addition to this, physics and detector performance groups have scheduled access to large quantities of data in Tier-1 sites. Tier-1s provide services for a cloud of associated Tier 2 facilities and host file catalogues and file transfer services. These sites are linked to each other and to CERN by a dedicated Optical Private Network (OPN) and they play an important role in the transfer of data around the world. The data transfers from a given site to another site of a different cloud go via its associated Tier-1. This role has been the object of discussion recently and will be commented in section 3.1.4. Tier-1

production operations are separated from on-demand user access, with the exception of a small group of users that has limited read access to the data and some processing power for early data studies.

There are 38 ATLAS Tier-2 centers spread over the world. A Tier-2 center often is a federation of multiple sites and, in practice, these sites are usually treated as independent by the Collaboration. Therefore, the number of Tier-2 sites grows to around 80 when counting site-by-site. These centers take care of analyzing the data, which are stored locally on disk, providing resources for on-demand usage and producing Monte Carlo simulations. The data management middleware¹ does not yet support storage accounting and quotas on a per user basis so there is no long-term user space provided in the Tier-2s. The model offers a substantial scratch space for user jobs and a centrally managed group space. The physics and performance groups decide centrally what to store in the latter. In many clouds, this is solved with private additional space for uses from a geographical community. These different storage areas are defined using the space token mechanism in the Storage Resource Manager [46], and access rights are controlled by role (e.g. production, user) and group (e.g. top-physics, ES-atlas).

Tier-0 at CERN and Tiers 1 and 2 around the world are well defined within the computing model and have been thoroughly tested in the last years. Tier-3 centres, on the other hand, are institution-level non-ATLAS funded or controlled centres that participate presumably most frequently in support of the particular interests of local physicists (users at the local facility decide how these resources are used). Tier-3 facilities must provide the software tools to access data and to perform local analysis. These local resources can vary widely in size, they must be highly reliable and they must have low latency. Within the ATLAS model such sites are mostly used for interactive or batch analysis of physics data sets. It is up to the different institutions to propose possible Tier-3 configurations and software setups that match the requirements of their users' analyses.

3.1.2 Data management

In the trigger system, the data is organized in several streams according to the different trigger types. Duplication of events, that is inevitable in this model, is kept below 10%. Improvements in detector understanding or better calibrations relevant to certain streams may require its reprocessing with high priority. In

¹*Middleware* is computer software that provides services to software applications beyond those available from the operating system.

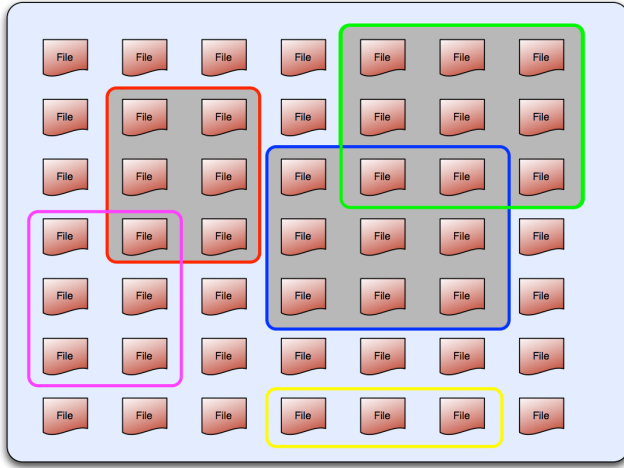


Figure 3.3: ATLAS Data Management core concepts. Individual files are collected into datasets (coloured outlines). Datasets may overlap and contain the same files (e.g., green and blue). Datasets themselves may be aggregated into containers, e.g., the red, blue and green datasets form the grey container.

order to accomplish this, the streams definition won't change under reprocessing. Data are also divided into blocks of fixed integrated luminosity. This way, the integrated luminosity of a given sample can be calculated assuming the analysis was applied on whole luminosity blocks.

In order to optimize the management and access of the data, collections of files of similar data are grouped into datasets as shown in figure 3.3. Subsequently, datasets can be grouped into containers. The containers are especially useful as they allow similar data processed at different times to be accessed without affecting the different datasets definition.

One copy of the RAW data that come from the detector is kept at CERN. Another copy is distributed among the Tier-1 sites. The different data formats defined are stored at Tier-1 sites as well.

- Event Summary Data (ESD) contains the full output of reconstruction but are deleted after 6 weeks for space reasons as they are usually not needed for physics analysis.

- Analysis Object Data (AOD) are a summary of the events reconstruction with physics objects such as electrons, muons, jets, etc. Smaller than the ESD, the AOD was the format used by the physicists in their analyses during the detector installation and commissioning.
- Derived Physics Data (DPD) are more refined formats created to satisfy the needs of the community. They usually contain a subset of the data plus some user-created information of special interest for the analysis. Some are just a selection of interesting events, others have all the events stored but have had some information removed in order to lighten them. The information removed varies from whole unneeded containers to individual objects that don't satisfy some quality criteria or even parts of individual objects that won't be used in the analysis.

Initially, the data format meant to be used in physics analyses was the AOD. However, once the data taking started, AOD files were found to be too large to be downloaded to local machines. In addition to this, many users wanted to work in an alternative framework such as ROOT [47]. In this situation different more refined versions of the ESD and AOD started to proliferate. Easier to handle and more customized, DPD has become the preferred format for physics analysis. Dedicated tools make easier for any user to create customized DPDs both from ESD and AOD. On top of that group production runs at Tier-1 and Tier-2 sites to create group-customized DPDs.

3.1.3 Distributed analysis

The large scale and complexity of the ATLAS experiment is inherited by its software structure. A large amount of CPUs is required in order to process the huge volume of real and simulated data. On top of that, several different analysis can run on this data simultaneously. Therefore, a high connectivity to the data is extremely important. The distributed tools used to analyse the data should be kept as simple and efficient as possible in order not to add more complexity to the already challenging situations physicists have to face.

In order to understand the distributed analysis model it is important to know its software. The ATLAS framework, Athena [45], is based on Gaudi [48]. Written in C++ [49] and Python [50], the Athena design separates algorithms from data and transient data (in memory) from persistent data (on disk). Its use cases include from high-level trigger to data simulation or reconstruction as well as physics analysis.

The distributed analysis model of the ATLAS experiment has the following key characteristics:

- The detector and simulation data are distributed worldwide by the data management system DQ2 [51].
- The different releases of Athena used to be installed at the different sites worldwide by a central installation system. This changed and now ATLAS software releases and the smaller file-based database are downloaded and cached at the sites and on the worker nodes thanks to CernVM-FS [52]. This is discussed in section 3.1.4.
- The user analysis code and tasks are sent to the sites where the data to be analyzed are located. The distributed analysis tools for handling all the user jobs management are Ganga [53] and the Panda client tools [54].
- The output and results of the user jobs are written to the scratch disk of the site where the job was executed but can be retrieved back to the user's local computer using the DQ2 command line tools or through a Data Transfer Request Interface (DaTRI). This transfer requests are restricted and have to be approved by ATLAS Computing Management.

ATLAS users can choose among several software solutions and data formats to do their analyses. Nevertheless, a typical analysis work-flow runs as follows:

- The first step is to find the real or simulated data on which the analysis is going to run. This can be done either using DQ2 tools or through the ATLAS Metadata Interface (AMI) [55]. The typical data volume is usually of the order of the TeraByte.
- In the case of MC samples, they can be produced using Athena tools but they must be officially validated by the ATLAS Collaboration if the results are expected to be made public.
- Athena code defined by a user or a physics group processes the data on the Grid in parallelized jobs. The output can vary from one scenario to another but it often is a set of ROOT ntuples and, in the case of physics analysis groups, they are usually D^3PD ² files.

²In the old naming convention D^1PD and D^2PD had the same format as an AOD (Athena compliant). The tertiary DPD type, D^3PD , refers to flat ROOT ntuples.

- D^3PDs can be further processed locally or on the Grid. On non-Grid resources, analysis can be performed via batch systems on local clusters or on PROOF [56] farms.

Monitoring and testing are important parts of the system. Expert shifters provide user support and act as a link communicating the problems users find to the sites involved. In addition to this, the HammerCloud stress testing system [57] has helped to improve the overall job efficiency. Even though the distributed analysis system is working well, future improvements can still be achieved by introducing new test cases that represent better the real usage of the infrastructure, a better error categorization or a better compatibility of the system with ROOT.

3.1.4 Computing model evolution

The ATLAS computing model has been changing continuously since the very beginning of the experiment. A model that intends to satisfy the needs of such a huge and diverse community needs to adapt to developments in concordance with its users' needs and computing technologies. This ability to adapt and improve is realized thanks to the rigorous tests that the model and the infrastructure have gone through all these years. These tests exposed the model weaknesses and bottlenecks well before the LHC data taking period started in 2009 and are responsible for the excellent performance shown so far.

The original model was designed assuming that only Tier-1 sites would have good network connection both among themselves and with their associated Tier-2 sites. A Tier-1 plus its associated Tier-2 form an ATLAS cloud. In this scenario, Tier-2 centres of a given cloud could only receive data from another cloud through their associated Tier-1. Nowadays, the network connections among sites are much better than what was foreseen in the computing model design. In order to utilize the improved networking capacity to its fullest, Tier-2 sites considered to be well connected (Tier-2Ds) are made multi-cloud sites [58]. A site is considered well connected if it can reach an overall transfer rate of at least 5 MB/s to at least 10 of the 12 Tier-0 plus Tier-1s in ATLAS for files heavier than 1 GB. These sites can connect directly to other Tier-1 or Tier-2 sites from a different cloud enabling a more efficient usage of disk and CPU resources and allowing high priority tasks to be done more quickly. This is shown in figure 3.4.

To identify well connected sites, regular transfers are generated to provide measurements which are displayed in the Site Status Board framework (SSB) [59].

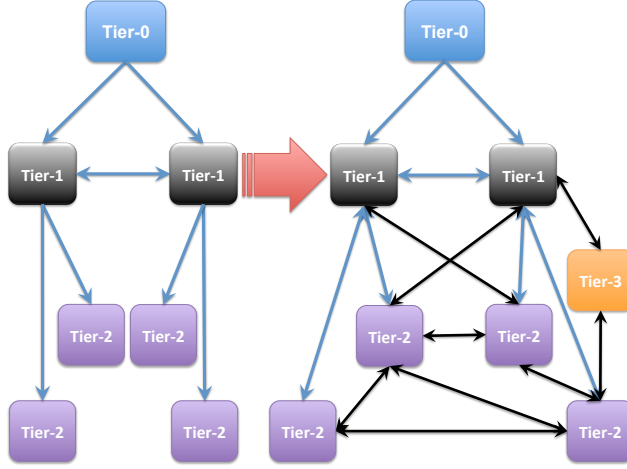


Figure 3.4: Evolution of the ATLAS computing model

Half of the Tier-2 sites are considered to be well connected. In parallel to direct transfers, such Tier-2s can now be assigned simultaneously to many Tier-1s for the production activity. It allows assigning the computing capacity to the Tier-1 hosting the highest priority task. A shortcoming of this model is that the sum of the CPU capacity at the Tier-1 and Tier-2 sites to produce data may not be in balance against the disk capacity at the Tier-1 to host the aggregated output. As an extension, it is also possible to associate even a Tier-1 site with other Tier-1 sites so that it can contribute to the task assigned to the other Tier-1.

Sites availability definition for analysis jobs is based on the panda queue status,

$$A = \frac{t_{up}}{t_{tot} - t_{unk}} \quad (3.1)$$

where t_{up} is the uptime, t_{tot} is the period of time evaluated and t_{unk} is the time when the status is unknown. In terms of availability, at the end of the month every Tier-2 is analysed and classified in one of the following four categories:

- Alpha: *availability* > 90% and site is a T2D.
- Bravo: *availability* > 90% but site is not a T2D.

- Charlie: $90\% > \textit{availability} > 80\%$.
- Delta: $\textit{availability} < 80\%$.

Following this categories sites with better availability and network connections will receive more input data according to the data distribution system pre-defined in the model. In order to reduce the waiting time of analysis jobs, there is a dynamic placement [60] of extra data replicas based on usage at Tier-1s for redundancy and at Tier-2s for analysis. For special requests with the approval of the Collaboration, there is also an on-demand replication system.

But the changes in the computing model do not end with the introduction of Tier-2Ds. To make a more flexible usage of distributed computing resources a quasi real time utilization of the best possible network connections and routing between sites is necessary. For this, the current weekly sonar tests are not enough and the global deployment of a more fine-granulated testing such as PerfSONAR will be very useful. From the infrastructure point of view, major improvements are foreseen in the next few years thanks to the LHCONe initiative [61], which should provide a more controlled environment for connecting Tier-2 sites. Ultimately, utilization of Tier-2s and Tier-1s should be fully transparent from the network perspective. Only at this point ATLAS will be able to utilize all storage and computing resources in the most flexible way.

On the other hand, evolutions have come with CernVM-FS [52] and Frontier/Squid [62]. CernVM-FS is a network file system based on HTTP, with which files are downloaded and cached at the sites and on the worker nodes. The ATLAS software releases and the smaller file-based database are now installed on the server at CERN, and there is no more need to install them at the site where CernVM-FS is used. This has removed the workload in software installation and the bottlenecks with the shared file systems. Frontier/Squid is an HTTP-based system to access databases with caching, avoiding a high load on the database and latency in its access from remote sites. The introduction of this system has removed limits related to the database access, making possible to run jobs at Tier-2 sites while accessing the database at Tier-1 sites. With this, any type of job can now run at any Grid site.

The changes in the ATLAS computing model indirectly affect Tier-3 operations as well. In order to satisfy the new Tier-2 quality requirements, the availability and the connectivity of the site need to be constantly evaluated. This permanent monitoring of the infrastructure meant an increase from the original 8h per working day and the inclusion of weekends.

The infrastructure had to be updated in order to fulfill the new requirements, specially the availability and the transfer requirements imposed by the project. Even though funding of the Tier-3 infrastructure is not included in the experiment pledges as it shares some of the Tier-2 computing and storage resources, it also benefits from these updates.

3.2 Tier-2 activities in the ATLAS ES-Cloud (Spain and Portugal)

In Spain, there is a federated Tier-2 [63, 64] made up of IFIC (Instituto de Física Corpuscular de Valencia), IFAE (Instituto de Física de Altas Energías de Barcelona) and UAM (Universidad Autónoma de Madrid). IFIC represents 50% of the ATLAS Spanish resources and has the responsibility to coordinate the activities of the Spanish Tier-2 federation. The members of the ATLAS user community in Portugal belong all to one research institute (LIP), with branches in Lisbon, Coimbra and Braga. The Portuguese federated Tier-2 consists of three sites, one at LIP in Coimbra, one at LIP in Lisbon and one at the National Grid Center in Lisbon. Each of the ES Cloud sites has an associated Tier-3 that shares the infrastructure of the Tier-2 at the same site, using the same queuing system. Additional resources for computing and storage for the Tier-3 have been added to the clusters. Whereas the storage for Tier-2 and Tier-3 are logically separated, for running the jobs a fair share queuing system was set up that guarantees that on a medium time scale the Tier-2 activities get the share that corresponds to the capacity of the Tier-2 site established for the ATLAS computing Grid. On a short time scale, whenever the activities of the local user group on the Tier-3 part of the site do not occupy the complete slot of their share, additional jobs can be run for the Tier-2 activities, and vice versa. This way we can optimize the occupation of the installed infrastructure.

3.2.1 Data placement

The storage in ATLAS is organized using space tokens [65]. These space tokens are controlled through the Distribution Data Management (DDM) system and they are associated to a path to a Storage Element (SE). The data distribution and size in the ES Cloud space token DATADISK and the sum of all space tokens on January 2013 are shown in figure 3.5. Table 3.1 shows the usage of the complete set of space tokens at IFIC-Valencia.

ATLASDATADISK is dedicated to data from the experiment (cosmic and collisions) as well as Monte Carlo production. Further space tokens are reserved

3. IFIC-VALENCIA TIER-3 WITHIN THE ATLAS COMPUTING MODEL

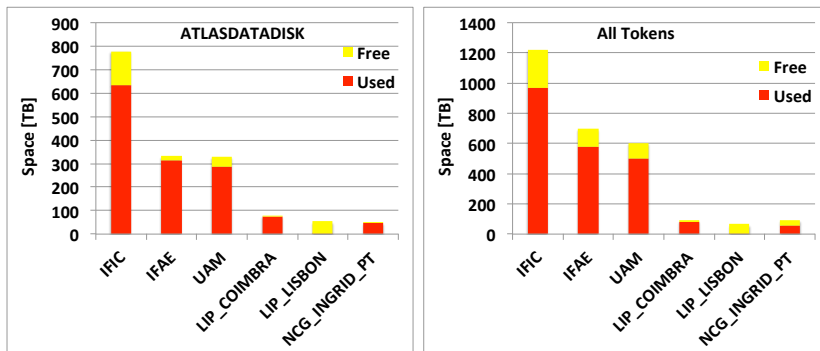


Figure 3.5: Status of ATLAS space tokens in the Iberian Cloud sites on January 2013.

Table 3.1: Usage of storage resources at IFIC-Valencia, January 2013.

Token	Assigned/TB	Used/TB	Free/TB	%
CALIB	22.0	18.3	3.7	83
DATA	778.1	632.9	145.1	81
PROD	31.9	11.3	20.6	35
GROUP	349.3	296.1	53.2	84
SCRATCH	51.2	25.2	25.9	49
HOT	7.0	1.0	5.9	15
LOCAL	76.8	66.6	10.2	86
sum	1,239.4	984.9	254.5	79

for physics analysis and performance groups. The LOCALGROUPDISK token is the connection between the ATLAS DDM and the private resources allocated at the Tier-3 facilities of the institutes. The number of space tokens is foreseen to decrease in the near future. DATADISK and GROUPDISK will merge in order to improve the space assignment flexibility and to prevent dataset duplication. HOTDISK will disappear as now conditions data and database releases are installed in CernVM-FS.

3.2.2 Dataset replication and data access

The new data distribution strategy distinguishes between primary and secondary replicas. In order to increase analysis opportunities, additional secondary replicas of popular data are made using the remaining available disk space. Primary replicas are distributed according to the computing model, at Tier-1s for redundancy and at Tier-2s for analysis. There is a dynamic placement of secondary replicas at Tier-2s based on usage as well as an on-demand replication system. On top of that, multi-cloud production and direct inter-cloud transfer make Tier-2s less dependent to Tier-1s and thus, its role has become more important. ES Cloud Tier-2 sites are getting more datasets than Tier-1 (PIC) as shown in figure 3.6.

3.2.3 Job distribution

In order to optimize our physics output and make maximal use of available CPU and disk resources production shares are fixed to limit group production jobs at Tier-1. Analysis share at Tier-1s has been reduced as well. Therefore, a large part of the analysis and the MC production is done at Tier-2s. Figure 3.7 shows the amount of jobs completed in the different sites of the ES Cloud during the beginning of the years 2011 and 2012. From the comparison of the two pie charts it is clear that there is now less weight at Tier-1 (PIC). The fraction of ES Cloud jobs completed at the Tier-1 dropped considerably, $\sim 15\%$, once these changes took effect.

3.2.4 Infrastructure and operation

The use of automatic tools for system management tasks is essential. To install and configure the operating system, Grid middleware and the storage system, QUATTOR [66] or PUPPET [67] are used. The ATLAS software is available for the collaboration via a network file system based on HTTP, CVMFS [68], where files and file metadata are cached on demand. For the cache, a SQUID [69] server is used. The database access by ATLAS jobs is done through SQUID plus FRONTIER [62] servers. There are two levels of monitoring: one from the global LHC Grid (the Service Availability Monitoring is one of its modules), and another one internal to the site, which uses tools like Nagios [70], Cacti [71] or Ganglia [72]. To avoid jeopardizing the centre availability, every site has a pre-production computing cluster where software updates are tested before they are put in place. Pre-production machines can be virtualised, via systems like

3. IFIC-VALENCIA TIER-3 WITHIN THE ATLAS COMPUTING MODEL

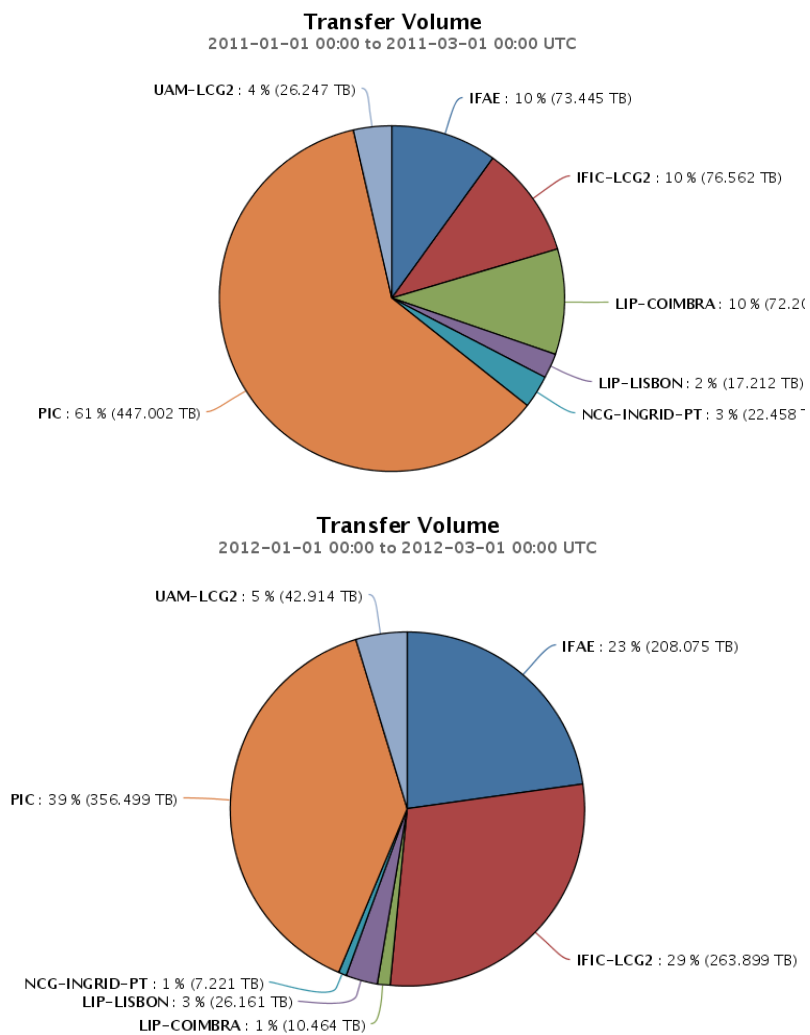


Figure 3.6: Data transfer volume in the ES Cloud sites during January and February 2011(top) and February 2012 (bottom).

3.2. Tier-2 activities in the ATLAS ES-Cloud (Spain and Portugal)

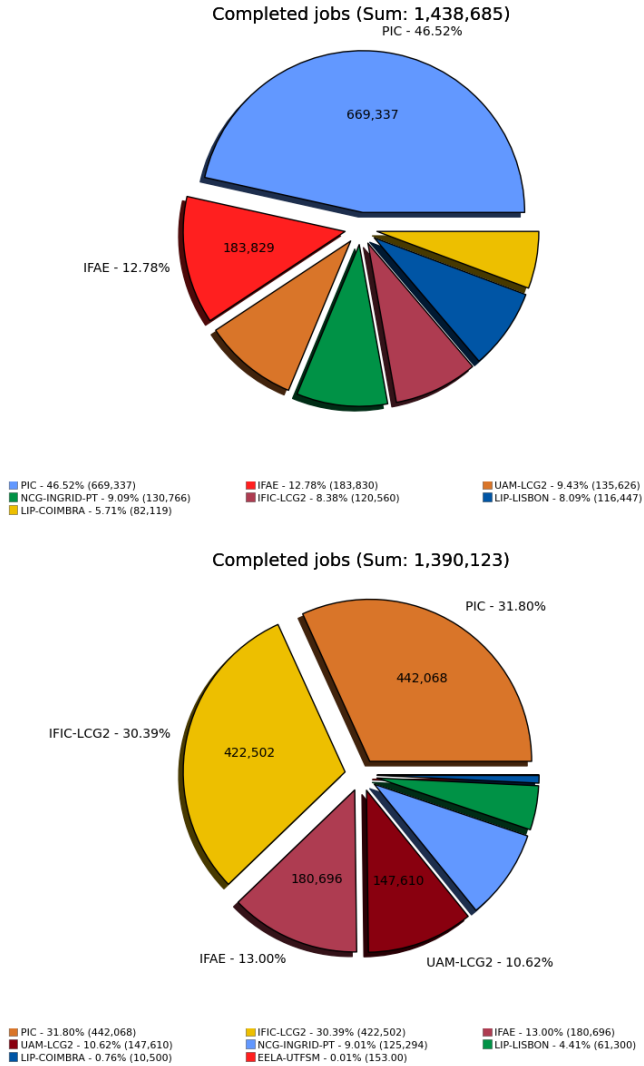


Figure 3.7: Completed job volume in the ES Cloud sites during January and February of 2011 (top) and February of 2012 (bottom).

xen or openvz, in order to economize resources. The CPU usage must be shared fairly between Monte Carlo production and user analysis jobs. The percentage assigned to each role is configured in the scheduler (Maui [73]), which is used by the batch queue system (Torque [74]) to submit jobs according to their priority.

3.3 Analysis at local facilities. Tier-3 at IFIC-Valencia

The design of IFIC-Valencia Tier-3 is shown in figure 3.8. We have the computing scheme which has been adopted by ATLAS for a Tier-3 centre in its minimal version. IFIC Tier-3 [75, 76] is attached to IFIC Tier-2. It currently has around 100 TB (60 TB under ATLAS Distributed Data Management (DDM tool) control plus 40 TB under IFIC control). An important feature of IFIC Tier-3 is that it uses the same storage system as the Tier-2. Its central component is the Lustre file system [77], a shared file system for clusters. The Lustre file system is available for Linux and provides a POSIX-compliant UNIX file system interface. This interface allows users to access the file system easily. Another important component of Lustre is the meta-directory Server (MDS), a catalogue

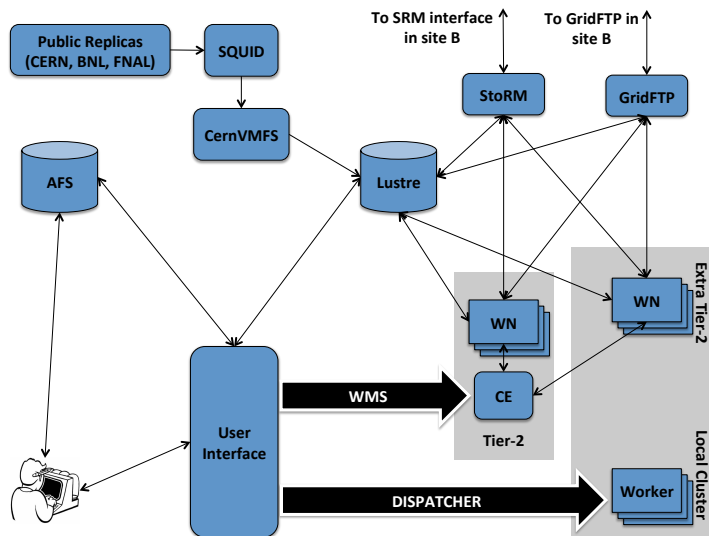


Figure 3.8: Design of IFIC-Valencia Tier-3.

that, in the case of IFIC, is the only shared resource between Tier-2 and Tier-3. In this case 3 disk servers are dedicated exclusively to Tier-3 to avoid overlap with Tier-2. PROOF [56] enables interactive analysis of large sets of ROOT files in parallel on clusters of computers or many-core machines.

The user can follow two different procedures in order to retrieve datasets generated in the distributed analysis phase:

- Request a subscription to the ATLAS DDM system to replicate the dataset in the LOCALGROUPDISK area. This disk space is allocated at the Tier-3 facility of the institute and it is connected to a storage element on the Grid through which the ATLAS DDM can perform the replication. This way the data can later be accessed locally. The replication is managed automatically by the ATLAS DDM.
- Use the ATLAS DDM client tools to download the dataset to the local disk space.

The computational facility must be highly reliable and must have low latency. These properties are particularly important in the last phase of the analysis due to the high frequency of the jobs that are running on the Tier-3 to obtain the final results. The CPU resources are organized in different architectures. This way the Tier-3 adapts better to the various needs the users might have:

- Some home-built User Interfaces (UI) are used to perform interactive analysis on the final datasets produced in the distributed analysis phase.
- A local batch farm to provide additional computing power for analysis that need to run on local resources.
- A PROOF farm for parallel processing of ROOT analysis jobs.

The three UIs have Scientific Linux CERN 5.5 (64 bits) installed. They are available through a dispatcher which makes them appear as one (ui00.ific.uv.es) to the users. Two computing elements (CREAM CE) are currently available at IFIC, CE03 and CE05. Thanks to this the user can:

- Access the AFS cell of IFIC. AFS is the common file system in the institute.
- Use Lustre (/lustre/ific.uv.es/grid) as a local file system. Lustre is a high performance file system that is dedicated only to our SE.

3. IFIC-VALENCIA TIER-3 WITHIN THE ATLAS COMPUTING MODEL

- Access ATLAS software and Grid tools (production releases, Ganga and DDM tools) at Lustre thanks to CERN-VM-FS.
- Perform local checks to develop analysis code before submitting larger jobs via Grid.
- Search data in the Grid and copy them locally.
- Send jobs to the Grid.

IFIC uses StoRM as the SRM storage element. StoRM is the interface of IFIC storage services with the Grid and it provides information about the location of the data available at IFIC.

With these tools being operational, physicists at IFIC have access to the ATLAS software (Athena) and all analysis tools (Ganga, DDM, ROOT) from their desktop or laptop. IFIC ATLAS users can perform analysis on simulated events or data and they can store their results locally following various workflow paths. Figure 3.9 summarizes them and shows that users can run on the different Grids (LCG/EGEE/EGI [78] , OSG [54] and NDGF/Nordu-Grid [79]) using the tools installed at IFIC.

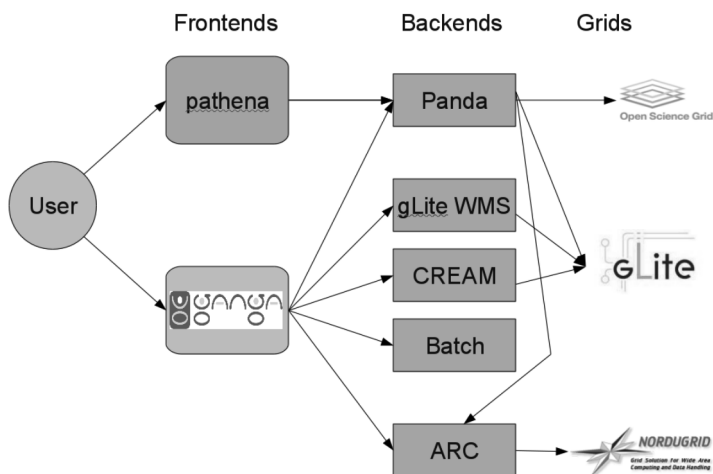


Figure 3.9: Overview of the different Grids and Job submission mechanisms offered to the users in the ATLAS experiment [60].

3.3.1 Coexistence with other scientific applications running at IFIC-Valencia

The Tier-3 facility is part of the e-Science [80] environment of IFIC consisting of two main infrastructures targeting both scientific and technological applications: the ATLAS Tier-2 and the GRID-CSIC. The Spanish ATLAS Tier-2 project provides distributed computing and storage resources to generate Monte Carlo events and to analyse real data collected by the ATLAS experiment. GRID-CSIC distributed computing and storage resources, on the other hand, are provided for different scientific applications and with emphasis on multidisciplinary projects [81].

The experience acquired by the physicists and technicians working in the Grid computing group and specially those working directly in the ATLAS Tier-3 is being exported to other physics groups of the institute. At IFIC there are now more scientific applications running in Grid mode and/or using the resources of the GRID-CSIC infrastructure (figure 3.10): neutrino physics, medical physics (in particular, hadron therapy), medical imaging, lattice QCD calculations, nuclear physics, etc. (see reference [82]). The most frequent usage from the non-ATLAS physicists is either the computing power obtained in a distributed system or the intensive computing of batch jobs but there is room for interactive analysis of derived data coming from the application of algorithms to big datasets as well.

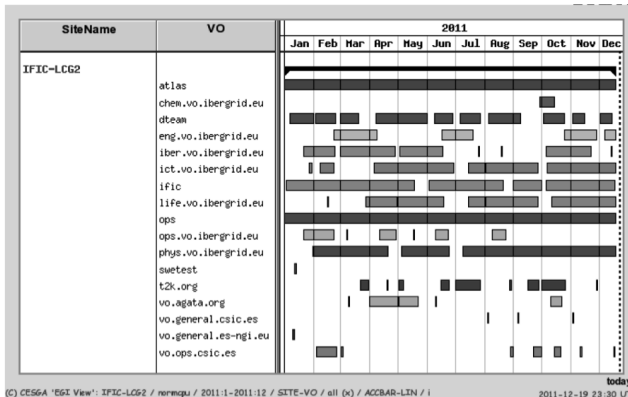


Figure 3.10: Virtual Organizations supported at IFIC.

3. IFIC-VALENCIA TIER-3 WITHIN THE ATLAS COMPUTING MODEL

Tools used for monitoring are the same for the entire infrastructure, including Tier-2 and Tier-3 resources, and in general all the scientific computing at IFIC. To monitor data transfers we are using Cacti [71] as a tool for checking links to data servers. We can check Tier-3 servers, which are isolated from others, enhancing performance and not disturbing Tier-2 transfers. The most demanding parts of the infrastructure are as isolated as possible to prevent interferences between them. For example, disk pools are different for the different projects (figure 3.11).

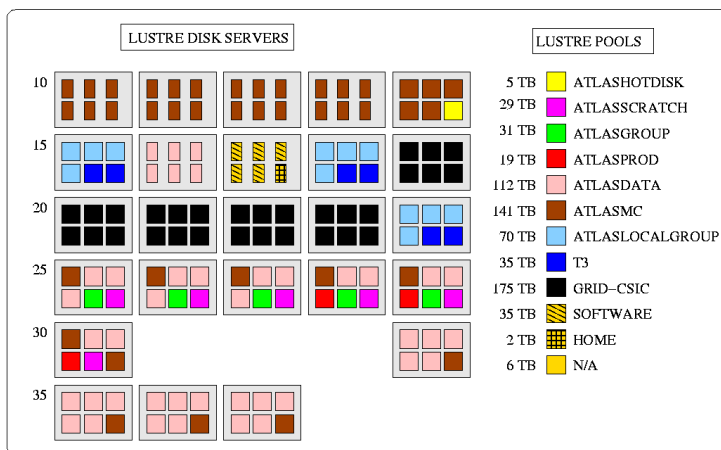


Figure 3.11: Lustrre disk pool distribution at IFIC

For node monitoring, including computing and storage, we are using Ganglia [72].

3.4 First performance tests of a PROOF farm prototype for the Tier-3 at IFIC-Valencia

In addition to the Grid resources of the Tier-3, a PROOF farm prototype has been deployed for interactive analysis of n-tuples. The use of PROOF is motivated because it provides an alternative and dynamic approach to run high energy physics analysis on distributed systems. A typical analysis workflow involves several steps. Once a first version of the analysis algorithm has been implemented it is run over the data available. The improvements derived from

the first run need to be implemented in the analysis algorithm, which will need to run over the data again forming a continuous refinement cycle. These analyses can be both I/O and CPU bound. The computing infrastructure needs many disks, to improve the I/O rate, many CPUs for processing and as much memory as possible to cache. Exploiting intrinsic parallelism is the best way to satisfy these requirements and be able to analyze the data in a reasonable time.

PROOF is a system for the interactive analysis of very large sets of ROOT data files on a cluster of computers. It speeds up the query processing by employing inherent parallelism in event data. Direct access to ESD or AOD is not required, but just direct access to where these n-tuples are generated. It is clear that a Tier-3 infrastructure must be partly inside the Grid, in order to get the data (ESD, AOD, dESD, etc), and partly outside the Grid for interactive analysis of n-tuples (dAOD, TAG, etc). In any case, this infrastructure, Tier-3 Grid and non-Grid resources, is using the same storage element in order to access to the data. Logically, The PROOF farm must be well connected to the storage element in order to get a fast access to the data.

Test using one machine with 8 cores (PROOF-Lite)

The PROOF farm prototype was subjected to several performance tests. This test consisted on reading some values from a dataset and filling a histogram. For this test we used a dataset with 3684500 events (7675.24 MB) composed of 372 files (22MB per file). The data was stored locally and on Lustre file system. The data rates shown in figure 3.12 are values given by PROOF and were calculated by simply dividing the amount of data processed by the time it took to process them.

As in this test CPU was more important than I/O, figure 3.12 shows that the Lustre file system had a nearly equivalent behavior as the local storage. Only when the 8 cores were used and there was no CPU left for I/O, the test reading from Lustre started to slightly deviate from linearity.

Test on a cluster of machines

The Lustre configuration at IFIC used 20 disk servers. Three of them are dedicated exclusively to our Tier-3 (see figure 3.13) in order to avoid overlap with our Tier-2 disk servers. In this way, the only shared resource between our Tier-2 and Tier-3 is the Lustre metadirectory server (MDS).

This test was undertaken with 1440 files (32GB) running on 128 cores (16 nodes):

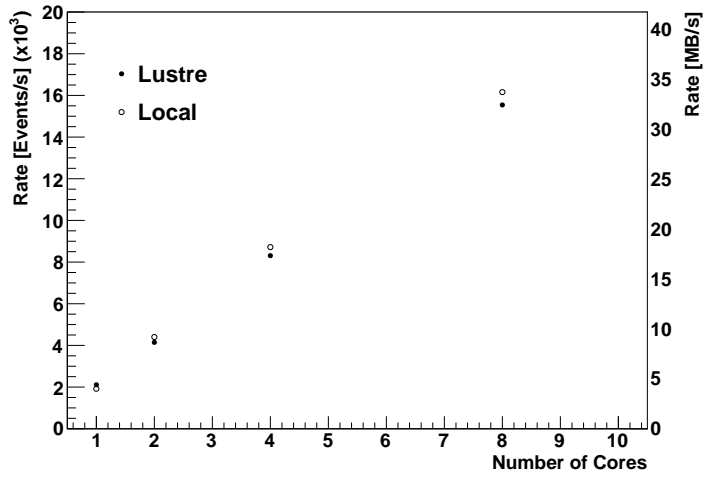


Figure 3.12: PROOF-Lite performance. Processed data rate as a function of the number of cores used.

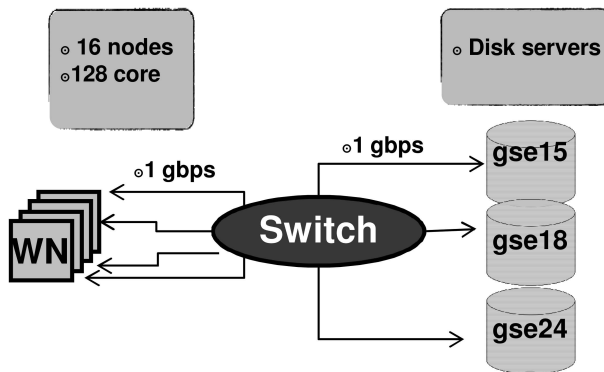


Figure 3.13: IFIC's Tier-3 PROOF farm.

3.4. First performance tests of a PROOF farm prototype for the Tier-3 at IFIC-Valencia

- 16 x HP BL460c, 8 cores, 2 x Intel Xeon E5420@2.5 GHz, 16 GB RAM
- 2 HD SAS 146 GB (15000 rpm)

Data was accessed via Lustre using its direct file protocol. With 128 cores we start losing linearity (figure 3.14). The following tests showed that we are limited by our disk server interface.

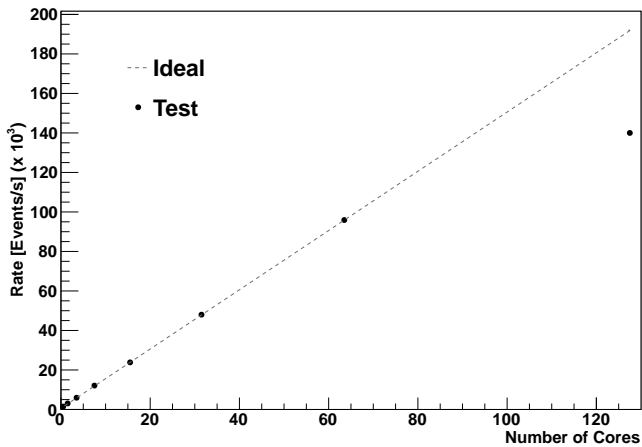


Figure 3.14: PROOF farm scalability.

Sequential read test

We made each of the 128 cores (16 nodes) read 100 random files with `dd` (`bs=32 KB`)³. A total of 10995 files (225 GB) were used. Test showed a bandwidth = 357 MB/s and as it is showed in figure 3.15, the disk server interfaces were saturated.

Bandwidth values were obtained from the switch CISCO X6509 counters (5 minute intervals) using `CACTI` [71].

³The `dd` utility copies the standard input to the standard output. The `bs=n` operand sets both input and output block size to `n` bytes.

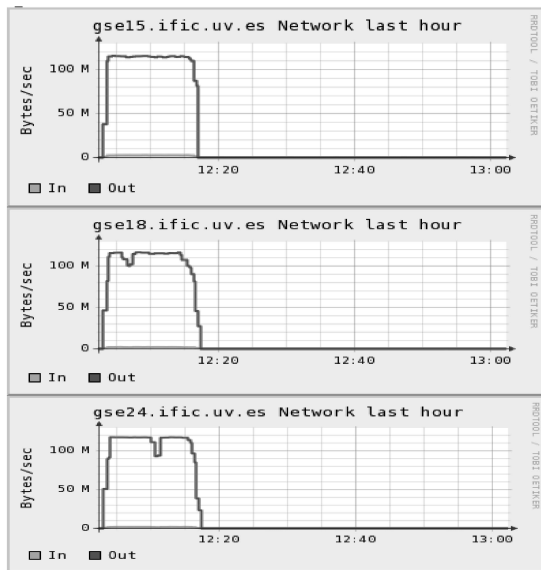


Figure 3.15: IFIC Tier-3 disk servers I/O read performance.

Test using 4 simultaneous PROOF sessions

With this test we wanted to see whether running several simultaneous PROOF sessions would show some performance degradation. In table 3.2 we show the result of running over 3684500 events (372 files, 7 GB) in a PROOF farm composed of 128 cores with only one PROOF session.

Table 3.2: Farm performance with only 1 PROOF session running.

N	Init(s)	Elapsed(s)	Rate(evt/s)	Rate(MB/s)
128	2.5	36	101634.4	228.3

Results of running the same test with 4 simultaneous PROOF sessions are shown in table 3.3. For this test we made 4 copies of the same data and made each PROOF session run the same analysis but reading from a different copy. As this time the farm was being shared, times and rates in table 3.3 change accordingly but we do not observe any added degradation. This supports the

3.4. First performance tests of a PROOF farm prototype for the Tier-3 at IFIC-Valencia

fact that PROOF has a good scalability.

Table 3.3: Farm performance with 4 simultaneous PROOF sessions running.

N	Init(s)	Elapsed(s)	Rate(evts/s)	Rate(MB/s)
128	6.0	2:38	23234.3	53.8
128	8.1	2:39	23133.0	53.8
128	8.1	2:36	23530.9	54.4
128	7.3	2:37	23362.0	54.7
Total			93260.2	216.7

Conclusions

Performance tests show good PROOF behaviour. The farm prototype exhibits correct scalability and concurrent use is possible without added degradation. The tests identified a possible bottleneck in the prototype. Lustre performance is limited by the ethernet interface of the disk server. Close monitoring of the interface is needed to prevent saturation. In case of a farm upgrade, this can be solved by aggregating a second interface (channel bonding).

4 Jet substructure

4.1 Jets in high energy physics

At short distances, asymptotic freedom allows quarks and gluons to move as quasi-free particles. When they are accelerated, they produce bremsstrahlung cascades of gluons and $q\bar{q}$ pairs, which then turn into jets of tightly collimated hadrons. The measurement of the jet momentum provides information on the quark or gluon that originated it. Therefore, jets played an essential role in the establishment and testing of the theory of strong interactions, quantum chromodynamics (QCD).

The first experimental evidence of jets was found in SPEAR, a $\sqrt{s} = 8$ GeV e^+e^- collider at SLAC [83, 84]. Predictions expected the process shown in equation 4.1,

$$e^+e^- \longrightarrow q\bar{q} \longrightarrow 2 \text{ jets} \tag{4.1}$$

where the $q\bar{q}$ pair hadronises evolving into two jets. The result announced in 1975 needed of a refined analysis that showed that there was indeed a jet structure in the events and that jet angular distributions were consistent with those of a pair of spin-1/2 particles.

The first jets visible to the naked eye would appear in 1979 in another e^+e^- collider, PETRA at DESY. The center-of-mass energy of 46 GeV allowed for clear 2-jet events. The 3-jet events, such as the one shown in figure 4.1, were interpreted as $e^+e^- \longrightarrow q\bar{q}g$, where the hard gluon is radiated off the emitted quark.

The SM was further tested in the following decades. The experimental and theoretical progress as well as the development of Monte Carlo tools consolidated jet physics. At the energies of the LEP era¹ it was already possible to study

¹LEP started at a center-of-mass energy of 90 GeV, enough to produce the Z^0 boson. In

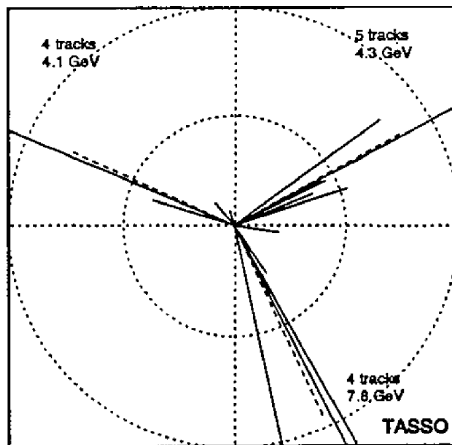


Figure 4.1: Observation of $e^+e^- \rightarrow q\bar{q}g \rightarrow 3 \text{ jets}$ [85] in the TASSO detector.

gluon self-interactions in processes like $e^+e^- \rightarrow q\bar{q}gg$. A beautiful example of a four-jet event interpreted as a $Z^0 Z^0 \rightarrow b\bar{b}q\bar{q}$ event taken by the DELPHI experiment is shown in the left-side frame on figure 4.2. Techniques like the reconstruction of secondary vertices inside jets allowed for the identification of b-quark-generated jets from the electroweak boson decay.

In hadronic machines such as Tevatron or the LHC jet reconstruction becomes a crucial tool. Jets played an essential role in the discovery of the top quark in Tevatron [86]. A good example is the event in the right-side frame on figure 4.2, where a $t\bar{t}$ pair is shown leaving a ‘e+4 jets’ signal in the CDF detector of the Tevatron. Apart from testing the SM at the highest possible energy, determining the gluon distribution in the proton or measuring QCD couplings, jets can be used to search for BSM phenomena (see chapter 5).

4.2 Jet algorithms

In the process of jet reconstruction the essential question that needs to be addressed is to decide which particles to cluster together. Jet algorithms have

a second phase, LEP2 reached ~ 200 GeV, where pair production of the electroweak bosons was studied in detail.

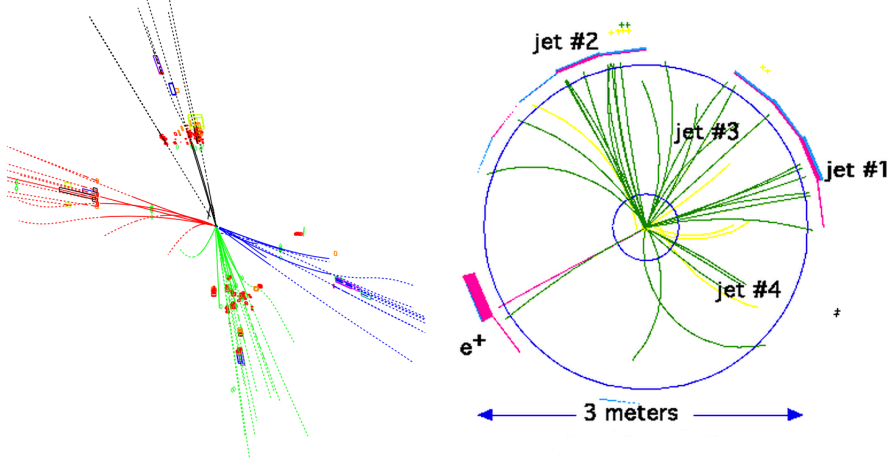


Figure 4.2: (Left) $Z^0 Z^0 \rightarrow b\bar{b}q\bar{q}$ event taken by the DELPHI experiment of the LEP accelerator. Secondary vertices are reconstructed in the b-jets (black and red). (Right) Observation of a $t\bar{t}$ pair by the CDF detector at Tevatron. The \bar{t} quark decays hadronically to \bar{u} (jet 2), d (jet 3) and \bar{b} (jet 1). The t quark decays to b (jet 4), e^+ and a ν_e estimated to be at 6 o'clock.

given different answers to this question over the last 30 years.

A good jet algorithm should provide a good matching between jet and parton properties even though, in practice, we work with particle-level jets, which are not directly comparable to partons. Calculability moreover requires infra-red and collinear safe algorithms, that have little sensitivity to soft and collinear emissions. It is also desirable for the algorithms to be easy to implement both in theory calculations and in experiment analyses. Calibration must be possible and, of course, it is also desirable for them to be fast.

Jet algorithms can be categorised in two main groups. Cone algorithms were most widely used at the Tevatron. They assume that most of the energy flow of the event is contained in a given geometric region (i.e.. a cone). The fact that better defined shape made them simpler to calibrate was the main argument for cone algorithms. Sequential recombination algorithms, on the other hand, have been most popular in LEP and HERA. The clustering inverts the parton shower by combining the constituents of the jet according to subsequent ‘distance’

criteria. Larger objects are built this way, which are then combined again until the jet is formed. The final output of a sequential recombination algorithm can vary depending on how the ‘distance’ used to pair jet constituents is defined. In general the distance between two jet constituents, d_{ij} , and the distance between a constituent and the beam, d_{iB} , are defined as,

$$\begin{aligned} d_{ij} &= \min(p_{Ti}^{2n}, p_{Tj}^{2n}) \times \frac{\Delta R_{i,j}^2}{R^2} \\ d_{iB} &= p_T^{2n} \end{aligned} \quad (4.2)$$

In contrast to previous experiments, where the number of jets is specified to the algorithm, jet reconstruction at the LHC is inclusive. Clustering continues until the minimum distance is found to be d_{iB} , in which case the object i is considered a jet. This produces a variable number of jets, governed by R . In equation 4.2, the value n can take the values 1 , 0 and -1 giving rise to three different jet algorithms.

- The k_t algorithm [87, 88] takes $n = 1$. It clusters soft collinear radiation first sort of inverting PYTHIA shower.
- Cambridge-Aachen corresponds to $n = 0$ [89, 90]. It is based on an angular ordering so it clusters the closest radiation first. This is related to HERWIG showers.
- The value $n = -1$ defines the anti- k_t algorithm [91]. It clusters hard collinear radiation first. This has no parton-shower interpretation but it provides a collinear-safe way of producing round jets, which are easier to calibrate.

The first sequential recombination algorithms were computationally prohibitive but this is no longer a problem thanks to the FASTJET package [92]. Nowadays, sequential recombination algorithms are simple and fast. In addition to this, assigning a clustering sequence to the events made room for the study of new observables based on the substructure of the jet.

Figure 4.3 shows the behaviour of different jet algorithms. The hard jets of the event are all circular jets with radius R for the anti- k_t algorithm. The choice of an optimal value of R carries some subtleties. According to the following rule from reference [93],

$$\frac{\langle p_{T,jet} - p_{T,parton} \rangle}{p_T} \propto \alpha_s \ln \frac{1}{R} + O(\alpha_s) \quad (4.3)$$

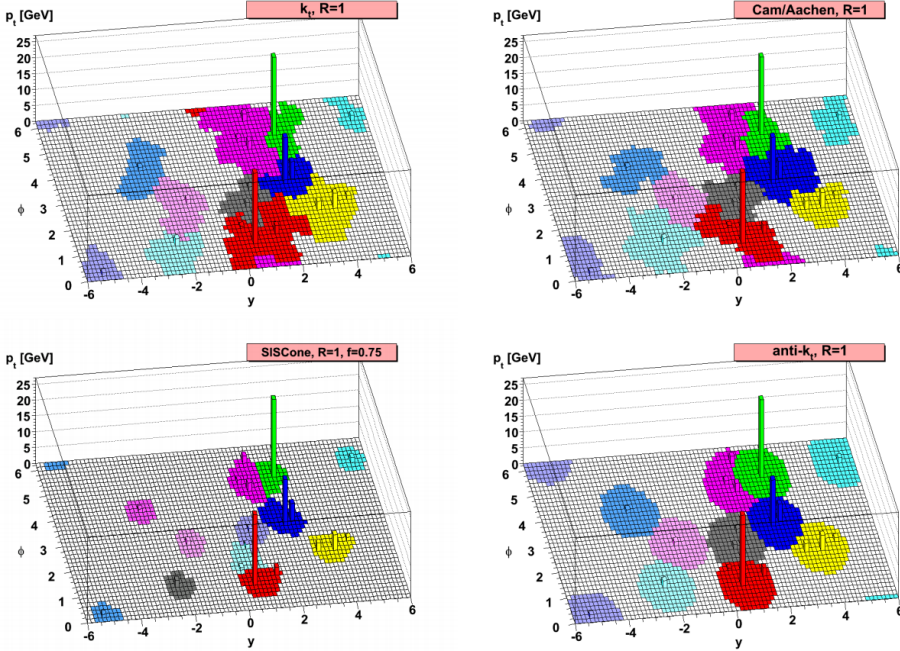


Figure 4.3: Example of different jet algorithms at work on a sample parton-level event generated with HERWIG mixed with $\sim 10^4$ random soft particles [93].

a big R , which implies better containment, is preferred. This is an approximation that depends on whether the jet is quark- or gluon-originated among other things. Following this rule, a gluon-induced jet with $R = 0.4$ is expected to have a $p_T \sim 8 - 10\%$ smaller than the original parton while a quark-induced jet is expected to have a p_T underestimated $\sim 4 - 5\%$.

On the other hand, a small R leads to a lesser presence of additional pp collisions (pile-up), hadronisation and underlying event (UE) effects.

The LHC experiments use different jet algorithms but have opted for the anti- k_t algorithm as a default for several reasons: it shows a reduced sensitivity to pile-up and noise, the regular area of the resulting jets makes them easy to calibrate, and it has an excellent energy resolution. ATLAS uses default R -values of 0.4 and 0.6.

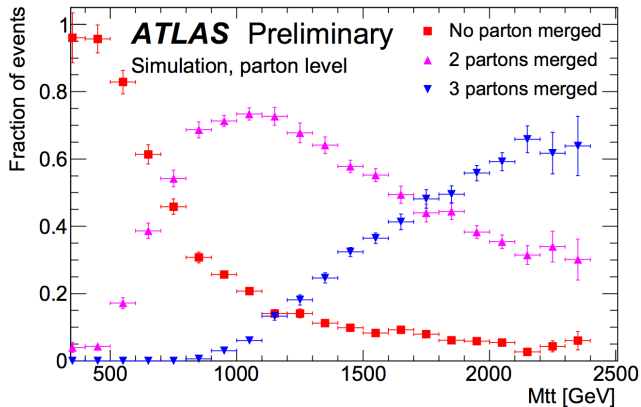


Figure 4.4: Probability that the three partons from a hadronic top decay are found within a ΔR distance of 0.8. No parton is merged (red), two of the three partons are merged (purple) and all three partons are merged (blue) [95].

4.3 Boosted objects and jet substructure

As the LHC explores a new energy regime, heavy SM particles like the W^\pm and Z^0 gauge bosons and the top quark are frequently produced with a momentum that considerably exceeds their mass. Only a handful of $t\bar{t}$ pairs with a mass over 1 TeV were created at the Tevatron. In the $\sim 20 \text{ fb}^{-1}$ of proton-proton collisions at 8 TeV of the first run of the LHC there are tens of thousands of $t\bar{t}$ pairs, and they will be counted by the millions when the LHC reaches its $\sim 14 \text{ TeV}$ design energy after the 2013-2014 shutdown [94]. These objects are already a crucial part of the physics programme of the LHC and their importance is bound to increase in the future.

The Lorentz boost of these particles alters the observed topology in an important way. Highly boosted objects thus represent a challenge to the conventional object identification and isolation criteria, that were developed primarily for particles approximately at rest in the laboratory frame. As it is shown in figure 4.4, the partons from the decay of the boosted particle are collimated into a smaller and smaller area and standard jet algorithms no longer resolve the resulting partons individually. The decay products are merged within a single jet. This strongly affects classical $t\bar{t}$ reconstruction algorithms,

which rely on the individual resolution of jets, and a new approach is needed to deal with the highly boosted top quark sample.

4.3.1 $t\bar{t}$ reconstruction

The decay of top quark pairs can produce several different final state topologies. This section is focused on the reconstruction techniques which apply to the ‘lepton+jets’ decay channel described in section 1.6, where one of the two W bosons emerging from the decay of the top quark pair decays into quarks, while the other W boson decays into a charged lepton (electron or muon) and a neutrino.

Resolved $t\bar{t}$ reconstruction

This method assumes that the $t\bar{t}$ pair decays at rest. The reconstruction in the $l + jets$ channel starts by identifying a well isolated, good quality lepton. The missing transverse energy of the event, E_T^{miss} , is associated to the escaping neutrino. The longitudinal momentum of the neutrino, p_z , is reconstructed imposing a W boson mass constraint to the lepton- E_T^{miss} system. This leads to a quadratic equation and, therefore, to several possible outcomes. If the discriminant of the quadratic equation is negative, the E_T^{miss} is varied until a null discriminant is achieved. If the quadratic equation leads to two different solutions, the one with the smallest p_z is chosen [96].

In the simplest approach, the mass of the $t\bar{t}$ system is reconstructed as the four-vector sum of the lepton, the neutrino, and the leading four jets. Restrictions on jet’s p_T and η are imposed to reduce the contribution of jets from initial- and final-state radiation, the jets considered must have $p_T > 20 \text{ GeV}$ and $|\eta| < 2.5$. Jets that are clearly separated, in angular distance ΔR , from the rest of the activity of the event are removed as well. One of the jets must be tagged as a b-jet. An example of a $t\bar{t}$ pair candidate decaying at rest is shown in figure 4.5.

A second approach uses a χ^2 to improve the efficiency of selecting jets from the top quark decay. A template is constructed imposing a constraint on the W and the top quark masses. In the cases where two solutions are found for the neutrino p_z , both are tested and the combination that leads to the smallest χ^2 is used.

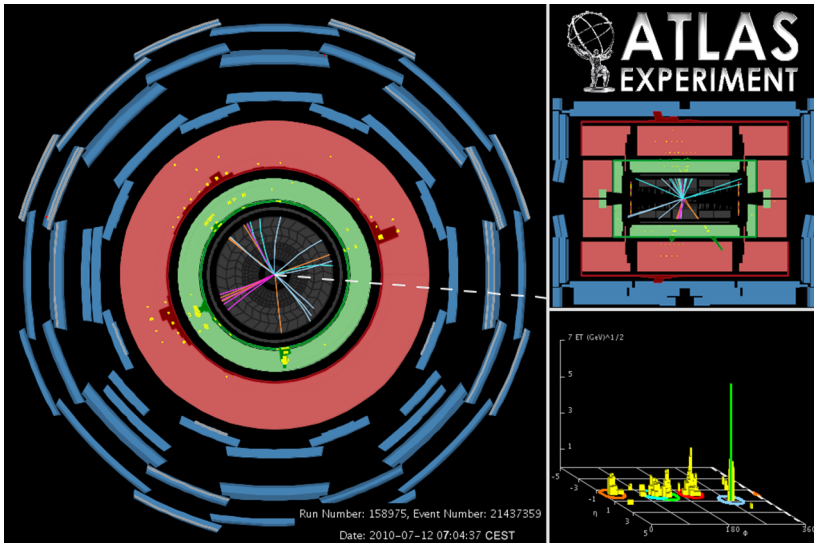


Figure 4.5: Event display of the electron plus jets candidate. The electron is shown as the orange downward-pointing track associated to the green cluster, and as the green tower in the $\eta - \phi$ lego plot. The direction of the missing transverse energy is shown as the dotted line in the $r - \phi$ view.

Transition region

As it is shown in figure 4.4, the efficiency of reconstructing separately the three partons from the top quark decay drops as $M_{t\bar{t}}$ increases, laying under 20% from $M_{t\bar{t}} = 1$ TeV on. Events in the mass range around the TeV often produce a mixture of topologies. They are no longer resolved events neither all decay products are within a single jet. A good example is the event shown in figure 4.6, where both resolved and boosted reconstruction algorithms can be applied.

Boosted $t\bar{t}$ reconstruction

An alternative approach to the reconstruction and identification of the decay of boosted objects like the top quarks shown in figure 4.7 is described in [97]. Rather than trying to resolve the jets individually, the complete decay is

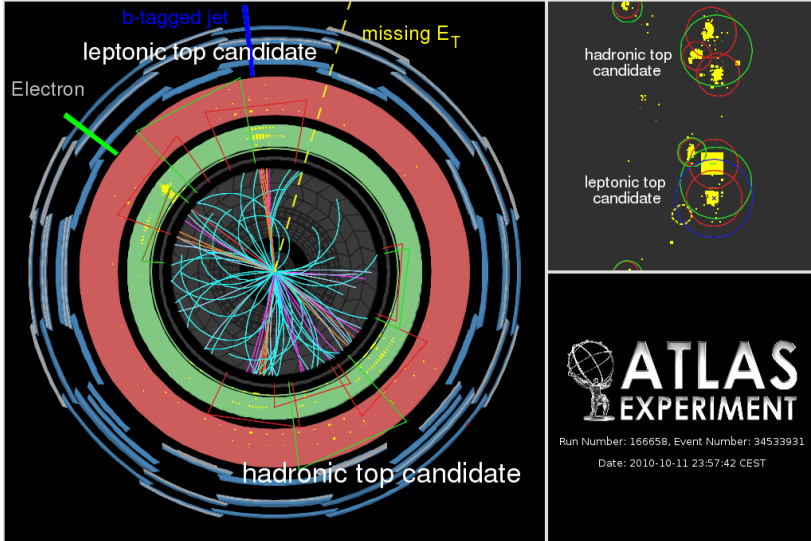


Figure 4.6: Event display for a $t\bar{t}$ candidate event. The leptonic top candidate is formed by a high p_T electron (145 GeV, 11 o'clock), moderate E_T^{miss} (1 o'clock), and a b-tagged jet at 12 o'clock. When reclustered with $R = 1.0$ it acquires a large p_T and mass as it absorbs the electron. 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 $R = 1.0$.

reconstructed as a single *fat* jet. The composite nature of the jet is revealed by an analysis of the substructure of the jet.

Many Monte Carlo studies [95, 98, 99] have shown that an analysis of boosted final states can greatly enhance the ATLAS potential in many areas of its physics programme.

4.3.2 Jet substructure

An analysis of jet substructure can successfully identify jets which contain a boosted object decay. Some techniques focus on cleaning away unwanted particles as their presence decreases jet mass resolution. Others look for

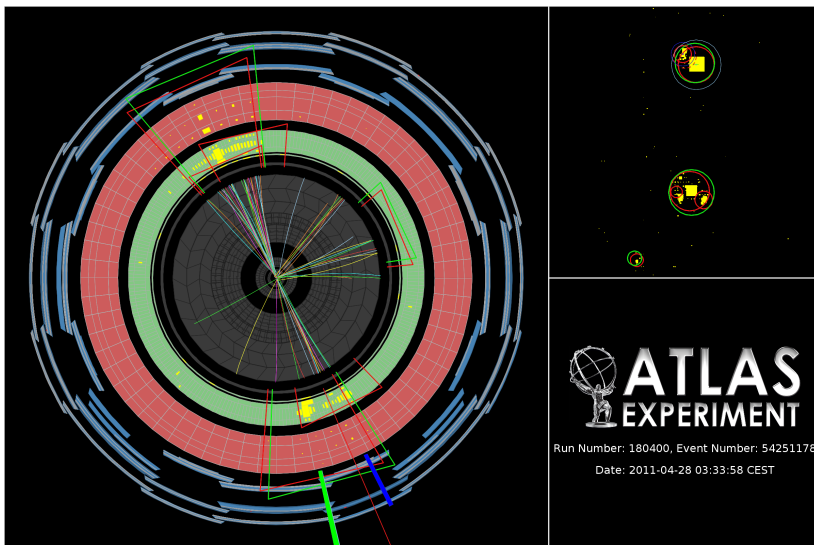


Figure 4.7: Event display for a $t\bar{t}$ candidate event with large mass, $m = 1.6$ TeV. The left panel displays a transverse view of the charged particle tracks and calorimeter energy deposits. An $\eta - \phi$ view of the same event is shown in the upper right panel. Jets reconstructed with $R = 0.4$ are indicated in red, jets with $R = 1$ in green.

substructure by identifying subjects within a given jet. There are also methods based on the energy flow within the jet. A more detailed discussion can be found in the proceedings of the BOOST series of workshops [100, 101, 102]. This section will focus on the observables needed for the analysis of chapter 5: jet mass and the k_t splitting scales.

Jet mass

It is interesting to discuss to what extent jet properties such as mass or p_T can be related to those of the original partons. Both jet p_T and mass are the result of the 4-vector-addition of the jet components. Therefore, perturbative radiation has a non-trivial effect on these properties. Soft radiation can escape the jet area diminishing the measured energy but, on the other hand, similar

soft particles coming from UE, pile-up or nearby jets can have a non-negligible contribution to mass specially when found in the outer area of the jet.

More precise theoretical calculations are still needed, as these quantities depend on whether the jet is quark- or gluon-generated. They also depend on the jet algorithm used and on the global environment surrounding the parton. Nevertheless, qualitative rules of thumb have been derived assuming $R \ll 1$,

$$\langle M^2 \rangle \propto \alpha_s p_T^2 R^2 \quad (4.4)$$

A case of particular interest in this thesis is the study of highly boosted top quarks. When measuring the top quark momentum, it is convenient to use a large R in order to include the whole decay of the heavy quark inside the jet. When looking for the decay structure, on the other hand, a too wide jet can degrade mass resolution. Adjusting the R -value depending on the p_T of the jet (i.e. $R \sim \frac{2m_t}{p_T}$) would be the optimal solution. In this thesis, anti- k_t jets with $R = 1$ are used to reconstruct the top quark decay.

k_t splitting scales, $\sqrt{d_{ij}}$

The k_t algorithm reconstructs jets by aggregating ‘soft’ constituents first, following equation 4.5, and forms ‘harder’ objects in every iteration. The k_t splitting scales, $\sqrt{d_{ij}}$, are defined as the k_t -distances clustering process. The splitting scale variable $\sqrt{d_{12}}$ corresponds to the final clustering step, where the last two sub-jets are combined into the final jet [98]. In the same way, in the step where there are three sub-jets, $\sqrt{d_{23}}$ is the k_t -distance at which two of them merge leaving two sub-jets, and so on.

$$d_{ij} = \min(p_{Ti}^2, p_{Tj}^2) \times \frac{\Delta R_{i,j}^2}{R^2} \quad (4.5)$$

When the jet has its origin in the decay of a heavy particle, and there are more than one parton inside the jet, the distribution of jet components is different to that of light quark- or gluon-generated jets where components are mainly concentrated in the center of the jet. The k_t splitting scales provide good estimations of this structure. The value of $\sqrt{d_{12}}$, for instance, is expected to be $m/2$ in the case that the jet contains the decay of a boosted heavy resonance of mass m , while the inclusive jet distribution is concentrated at lower values, $\lesssim 20 \text{ GeV}$, although with a tail extending to high values.

4.4 Commissioning of jet mass and k_t splitting scales

In this section, the normalised cross-section as a function of jet mass is measured for anti- k_t jets with an R -parameter of 1.0. In addition to the mass, k_t splitting scales [98] are measured. An inclusive sample of high p_T jets from the 2010 $\sqrt{s} = 7 \text{ TeV } pp$ dataset was used for the measurements. The selected data set corresponds to an integrated luminosity of $35.0 \pm 1.1 \text{ pb}^{-1}$ [103]. This is a natural continuation of the studies in previous experiments [104, 105, 106, 107, 108, 109]. It also complements previous ATLAS studies [110] probing the shape of jets reconstructed with the anti- k_t algorithm [91] with a radius parameter $R = 0.4$ and 0.6.

The measurements presented in this section are for jets with $|y| < 2$ and in four 100 GeV p_T bins ranging from 200 to 600 GeV. These requirements select jets that are not biased by trigger effects and are contained entirely within the barrel and end-cap subsystems of the ATLAS detector [111].

4.4.1 Event selection

Events first go through a first level hardware-based calorimeter trigger system whose efficiency was evaluated in data and found to contain no significant biases for the selection used here.

In order to reject detector noise and non-collision backgrounds, events are required to contain a primary vertex consistent with the LHC beam spot and with at least 5 tracks with $p_T > 150 \text{ MeV}$. In addition to this, events are discarded if any bad quality [112] anti- k_t jet with $R = 0.6$ and $p_T > 30 \text{ GeV}$ is found. This selection removes approximately 3% of events in this dataset.

The presence of pile-up has a significant impact on jet substructure [100]. Only events with exactly 1 reconstructed primary vertex ($N_{\text{PV}} = 1$) were used in this study. This requirement selects approximately 22% of the events. Vertex finding is highly efficient, thus no additional systematic uncertainties are considered.

Jets are built from three-dimensional topological clusters, topoclusters, of calorimeter cells. The algorithm starts at seed cells with significant energies compared to the expected noise. The cluster grows by adding neighbouring cells given their energy signal-to-noise ratio lays under a low threshold. An additional medium threshold is used to decide whether the newly added cells can become additional seeds. Higher thresholds are used for seeds and neighbours in order to suppress both electronics and pile-up noise, while the low threshold for

the cluster growth ensures that tails of showers are not discarded [113]. These clusters are made massless.

4.4.2 Monte Carlo samples

Several MC generators, including PYTHIA 6.423 [114] and HERWIG++ 2.4 [115], were used to produce samples of inclusive jet events. These programs implement leading-order (LO) perturbative QCD (pQCD) matrix elements for $2 \rightarrow 2$ processes. Additionally, ALPGEN 2.13 [116] and SHERPA 1.2.3 [117] are used for some cross-checks. SHERPA and ALPGEN implement $2 \rightarrow n$ processes such as explicit QCD multijet production. Parton-showers are calculated in leading-logarithmic approximation. Showers are p_T ordered in PYTHIA and angular ordered in HERWIG++. Fragmentation into particles is implemented in PYTHIA following the string [118] and in HERWIG++ the cluster [119] model. ALPGEN is interfaced with HERWIG [120, 121] for parton-shower and fragmentation. Unless otherwise specified, PYTHIA samples use the AMBT1 tune [122]. In some figures the Perugia2010 PYTHIA tune is used [123], which has been found to describe jet shapes more accurately at ATLAS [110]. Leading-order parton density functions are taken from the MRST2007 LO* set [124, 125], unless stated otherwise. No pile-up was included in any of these samples.

The MC generated samples are passed through a full simulation [126] of the ATLAS detector and trigger, based on GEANT4 [127]. The Quark Gluon String Precompound (QGSP) model is used for the fragmentation of nuclei, and the Bertini cascade (BERT) model for the description of the interactions of the hadrons in the medium of the nucleus [128].

4.4.3 Jet calibration

The calibration of energies measured by the calorimeter is of crucial importance for jet measurements. The non-uniformity and non-compensation of the calorimeter, the presence of dead material and magnetic field effects are examples of issues that need to be accounted for in the calibration. Jet constituents are topoclusters at electromagnetic energy scale (EM scale). The standard ATLAS approach in 2010 used a jet-level correction to calibrate the whole p_T of the jet. This global jet energy scale (JES) correction is derived from MC and is then applied directly to the EM scale. The so called EM+JES scheme estimates the resulting uncertainty on the energy scale for anti- k_t jets with $R = 0.4$ and 0.6 at the level of 3-4%. However, this jet-level correction grows to 20-30% in the case of large- R jets.

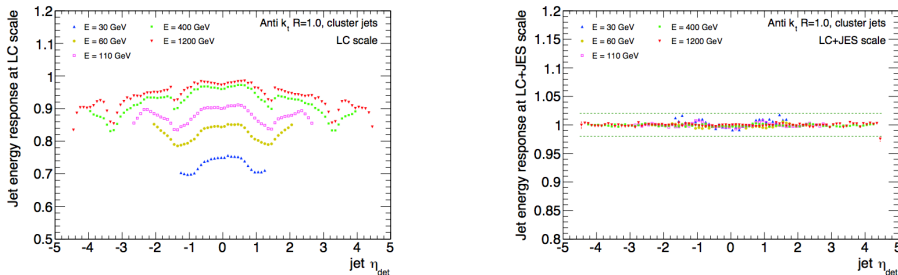


Figure 4.8: Jet energy response before (left) and after calibration (right) for anti- k_t jets with $R=1.0$.

For the study of the jet substructure of large- R jets, another calibration procedure is applied to the topoclusters. In order to account for calorimeter non-compensation, the calibration varies depending on whether the topoclusters have an hadronic or electromagnetic origin. This so-called local calibration (LC) represents the best available approximation of individual particle reconstruction in the calorimeter and has been shown to improve the energy resolution of jets [129]².

A global jet-level correction is applied to the large- R jets. The correction is done after a numerical inversion procedure, where the jet response, p_T^{reco}/p_T^{truth} , measured as a function of truth transverse momentum is transformed into jet response that is a function of the reconstructed jet transverse momentum. The difference lies in the fact that the correction is applied on LC-topoclusters whereas the standard acts on EM-topoclusters. Therefore, the jet-level energy correction is of order $\lesssim 10\%$ in the energy range of interest because the dominant sources of miss-calibration have already been dealt with in the LC-topoclusters. The usage of LC-topoclusters reduces the need for jet-level corrections. Most substructure observables are used without jet-level corrections. An exception is jet mass. Corrections in this case are a bit larger but still of order 15-20%. The jet energy and mass response is shown in figures 4.8 and 4.9 before and after the calibration. Finally, the correction on η is of the order 0.01.

²The usage of LC-topoclusters became the standard approach in subsequent analyses.

4.4. Commissioning of jet mass and k_t splitting scales

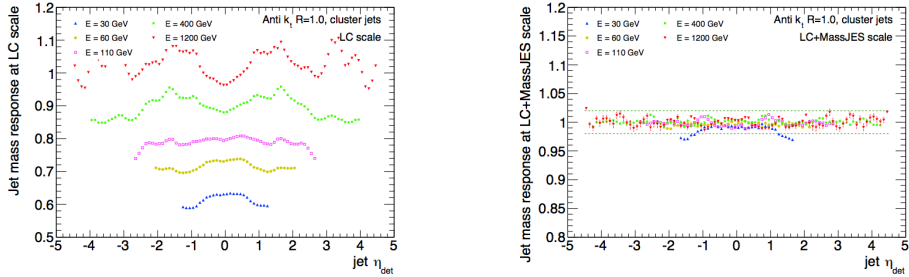


Figure 4.9: Jet mass response before (left) and after calibration (right) for anti- k_t jets with $R = 1.0$.

4.4.4 Systematic uncertainties

The dominant systematic uncertainty for this study comes from the modeling of the calorimeter response. Therefore, scale- and resolution-related systematic uncertainties have been considered for each substructure variable.

The scale uncertainties are constrained by comparing calorimeter-jets and jets built from ID tracks, track-jets. This comparison allows for some separation of detector and physics effects because systematic effects in the ID and in the calorimeter are highly uncorrelated. The tracks used belong to the selected sample of events and fulfill the criteria $p_T > 500 \text{ MeV}$ and $|z_0| < 5 \text{ mm}$, where z_0 is the z -coordinate of the track at closest approach to the z -axis. This technique is limited by inner-detector systematics and confidence in Monte Carlo modelling of the relative behaviour of the charged and neutral components of jets.

Calorimeter- and track-jets are considered to be matched if they are found within a $\Delta R < 0.3$ cone. Ratios, r^X , and double ratios, ρ^X , are defined for the substructure variables of the matched pairs of jets.

$$\begin{aligned}
 r^X &= \frac{X_{\text{calorimeter-jet}}}{X_{\text{track-jet}}} \\
 \rho^X &= \frac{r_{\text{data}}^X}{r_{\text{MC}}^X}
 \end{aligned}
 \tag{4.6}$$

where X stands for the variables of interest (p_T , mass, $\sqrt{d_{12}}$ and $\sqrt{d_{23}}$). The ratios in figure 4.10 show that data and MC are in good agreement. The double

ratios are used to isolate possible physics effects in the MC modeling. The double ratios are expected to be consistent to unity as long as detector effects are well described in the MC samples. Deviations from unity indicate the response of either tracker or calorimeter is mismodelled. The tracker response is assumed to be well modelled so the full deviation is assigned to the calorimeter. Thus, the jet energy scale uncertainty and that of other observables can be constrained using data. Below each plot in figure 4.10, the corresponding double ratio is shown as well. In order to account for possible uncertainties due to different fragmentation and hadronisation models, these double ratios are also calculated with a variety of MC programs.

Scale uncertainties are then the result of adding in quadrature the uncertainty estimated on the ID and the deviation from unity observed in the double ratios. This result, that varies in the range of 3-6%, is the dominating uncertainty on the final measurements.

As an additional cross-check, MC-based tests were used to determine the dependence of the detector response on a number of different variables. These included samples with modified detector geometry, different low-energy physics models and MC packages implementing different high-energy physics models. These tests indicated variations of a similar order of magnitude to those observed in the in-situ studies. The in-situ track-jet study is limited by ID acceptance and only extends as far as $|\eta| < 1.0$, which corresponds to $\simeq 75\%$ of the jets in the measured distributions. However, the MC-based tests also indicate no strong η -dependence from any of the different possible types of mismodelling examined. Based on this, the systematic uncertainty is applied to the entire sample.

Resolution uncertainties are taken from MC tests like the ones described above because the mass and substructure variable resolutions are difficult to validate in situ with this dataset. In-situ tests of the JER [130] for anti- k_t jets with $R = 0.4$ and 0.6 indicate that the jet p_T resolution predicted by simulation is in good agreement with that observed in the data. From studying the variations in resolution created by varying detector geometry, low-energy physics and high-energy physics models, resolution uncertainties of around 20% are conservatively estimated.

4.4.5 Prospects for multiple pp interactions

Multiple simultaneous pp interactions (pile-up) are contained in the ATLAS 2010 data set [131]. Pile-up constitutes a background of soft, diffuse radiation that offsets the energy measurement of jets and has a non-negligible impact on

4.4. Commissioning of jet mass and k_t splitting scales

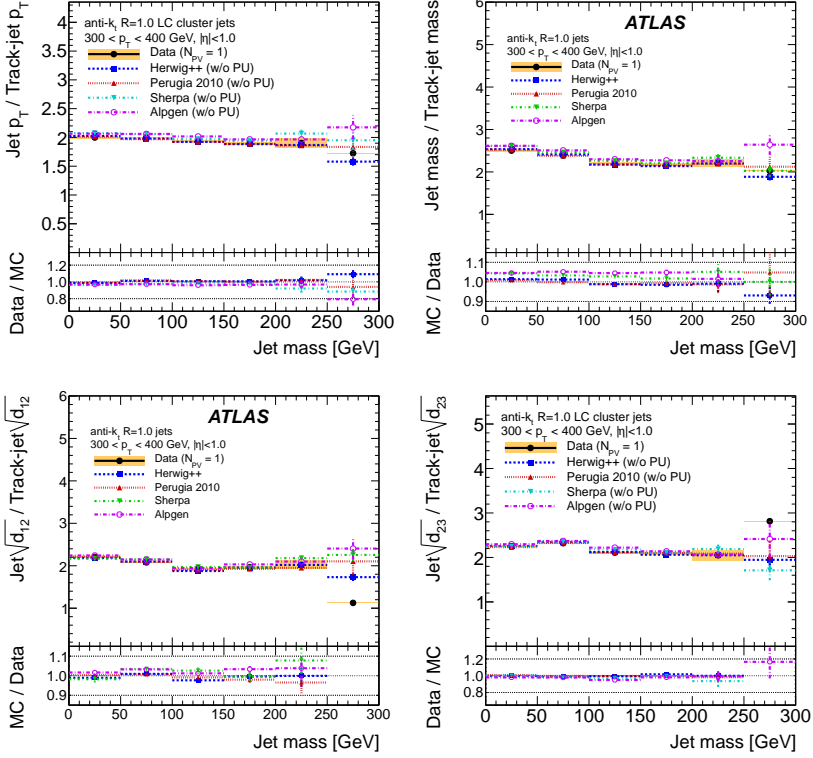


Figure 4.10: The single ratio r^X for jet p_T , jet mass, $\sqrt{d_{12}}$ and $\sqrt{d_{23}}$ in the 300-400 GeV p_T bin. The double ratio ρ^X is represented by the bottom frame of each plot.

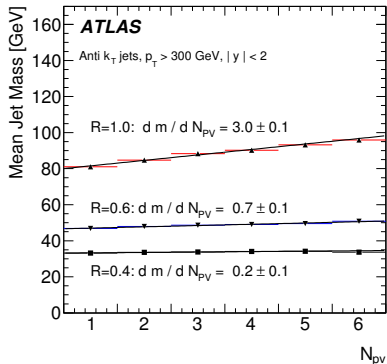


Figure 4.11: The mean mass for jets with $p_T > 300\text{GeV}$ as a function of the number of primary vertices identified in the event. Comparisons show the effect for anti- k_t jets with different R -parameters. Each set of points is fitted with a straight line.

jet substructure observables [100]. The growth of jet mass with the number of primary vertices present in the event is shown in figure 4.11 for different jet sizes and it is a clear prove that pile-up has a large impact on the invariant mass of fat jets. The results presented in this chapter have been elaborated using events containing only one pp interaction. Nevertheless, reference [132] also demonstrates that jet substructure techniques which reduce the area of jets are extremely promising for suppressing the effects of pile-up.

4.4.6 Jet substructure distributions

Detector-level distributions for jet p_T , η , mass and $\sqrt{d_{12}}$ are shown in figures 4.12 and 4.13. The statistical uncertainty represented in ratios, $MC/Data$, is the addition in quadrature of MC and data. Representative distributions of the substructure variables are shown for the 300-400 GeV bin only. MC is normalised to the data separately in each plot. The properties of these jets are observed to be reasonably well modelled by leading-order parton-shower MC.

4.4. Commissioning of jet mass and k_t splitting scales

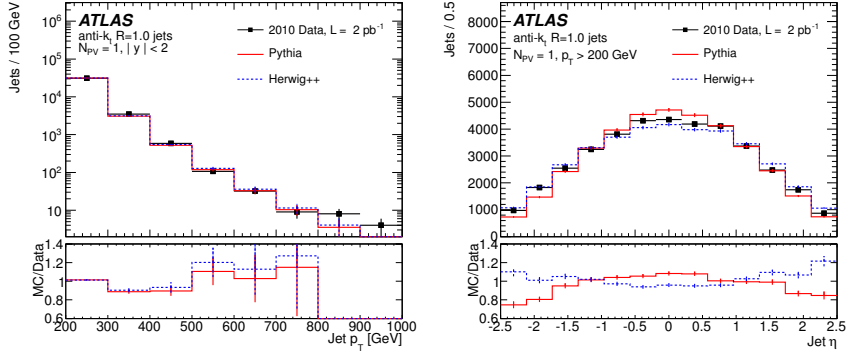


Figure 4.12: p_T (left) and η distribution (right) of anti- k_t 1.0 jets with $p_T > 200$ GeV.

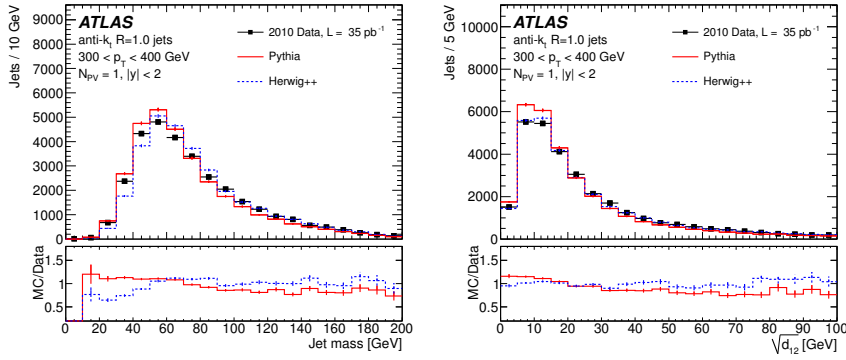


Figure 4.13: Distributions for mass (left) and $\sqrt{d_{12}}$ (right) of anti- k_t 1.0 jets with $|y| < 2.0$ in the 300-400 GeV p_T bin.

4.5 Conclusions

The understanding of jet substructure variables, their commissioning with the first data from ATLAS and the estimation of their uncertainties are essential for physics analyses involving boosted objects like the one described in chapter 5. The results described in this chapter are part of a more extensive study published by the ATLAS Collaboration in *Journal of High Energy Physics* [132]. This was the first particle-level measurement of these variables at the LHC and the first study of large R jets at any experiment. There is broad agreement between data and leading-order parton-shower MC predictions from PYTHIA and HERWIG++, although there would seem to be some scope to improve this. The systematic uncertainties estimated for jet p_T , mass, and substructure variables anti- k_t jets are shown in table 4.1.

Table 4.1: Final systematic uncertainties for anti- k_t jets.

Uncertainty	200–300 GeV	300–400 GeV	400–500 GeV	500–600 GeV
JES	4.0%	5.2%	6.0%	3.9%
JMS	4.5%	4.5%	6.0%	6.0%
JER	20.0%	20.0%	20.0%	20.0%
JMR	20.0%	20.0%	20.0%	20.0%
$\sqrt{d_{12}}$ scale	4.4%	3.8%	6.0%	6.8%
$\sqrt{d_{23}}$ scale	4.4%	3.4%	5.1%	5.1%
$\sqrt{d_{12}}$ res.	21.0%	22.0%	28.0%	31.0%
$\sqrt{d_{23}}$ res.	20.0%	20.0%	21.0%	25.0%

Even though jet mass exhibits disagreements with MC, reference [132] shows that the splitting and filtering procedure [99] can solve the problem. The substructure variables $\sqrt{d_{12}}$ and $\sqrt{d_{23}}$ are reasonably well reproduced by MC predictions.

The presence of pile-up has a non-negligible effect on jet mass and substructure observables. Therefore, the results presented in this chapter have been elaborated using events containing only one pp interaction.

Generally it seems that jet mass and substructure quantities can be successfully reproduced by leading-order parton-shower MC. This result bodes well for future analyses aiming to make use of substructure techniques to search for new physics.

A comprehensive review on jet finding at hadron colliders is found in [93] and good overviews of the status of algorithms for the reconstruction of boosted

objects at the time of the LHC start-up is found in references [100, 101, 102].

5 New physics searches in $t\bar{t}$ pairs with boosted topologies

The ATLAS experiment has performed searches for a wide range of signatures of BSM physics. Among these signatures, there are those which manifest through resonant production of pairs of SM particles. This chapter describes a search for new heavy particles decaying to $t\bar{t}$ pairs.

Previous searches for $t\bar{t}$ resonances were most carried out by the CDF [133, 134, 135, 136, 137], and D0 [138, 139] collaborations at Run II of the Fermilab Tevatron Collider.

In this chapter a novel $t\bar{t}$ reconstruction technique that is designed for boosted top quarks is discussed. The decay products of boosted top quarks are so collimated that standard $t\bar{t}$ selection criteria such as lepton isolation cease to work efficiently and these events end up being discarded. The ATLAS top-tagging algorithm [140, 95] exploits the potential of the jet substructure observables presented in chapter 4. By recovering these otherwise discarded $t\bar{t}$ events, this study can extend the mass reach of $t\bar{t}$ resonance searches. The topologies considered are those where one of the top quarks decays hadronically ($t \rightarrow Wb \rightarrow q\bar{q}b$), while the W boson from the other top quark decays to a neutrino and, either an electron ($e + jets$) or a muon ($\mu + jets$).

5.1 Benchmark models

In order to quantify the sensitivity of $t\bar{t}$ resonance searches two different signal models are used. These are well known proxies to extensions of the SM. For comparison with previous searches a generic narrow resonance, a Z' boson, is used. A heavy gluon, which generates a broad resonance, is used to complement the search. In addition to this, it is also important to distinguish between colour-singlet and colour-octet benchmarks as argued in reference [141].

5.1.1 The leptophobic Z' boson

Many models assume that the mass of the top quark implies that the third family plays a special role in the dynamics of EWSB. In the case of topcolor, the introduction of a new strong gauge force, which couples preferentially to the third family, gives rise to the formation of $b\bar{b}$ and $t\bar{t}$ condensates, which bestow the top and the bottom quarks with a large mass. There are several ways to bring fermion masses down to the correct scale. In ‘topcolor assisted technicolor’ [142], this is accomplished through the introduction of additional strong dynamics.

At this point, a mechanism, which favours $t\bar{t}$ while it blocks $b\bar{b}$, is needed in order to make the top quark heavy but keep the bottom quark relatively light. The SM Z^0 boson does precisely this but the SM coupling, g_1 , is so small that, in topcolor schemes, where the couplings of the new strong force to the third family are large, it would require of fine-tuning. Therefore, a new Z' boson is introduced. These models produce a narrow resonance and the experimental resolution dominates the width of the reconstructed peak for masses well over the TeV. By rescaling the cross-section times Branching Ratio the limits can be interpreted in other models. Events for this model are generated with PYTHIA [114].

In this study, the choice of the couplings corresponds to the sequential Z' boson [143], where all the couplings are identical to those of the SM Z^0 boson. A specific scenario is considered to rescale the cross-section, where the Z' couples only to the first and the third generations of quarks and it has negligible couplings to leptons. This model appears as ‘model IV’ in reference [144] with the coupling to the right-handed component of the u -type quarks, $f_1 = 1$, the coupling to the right-handed component of the d -type quarks, $f_2 = 0$, and a width of 1.2% of the Z' boson mass. This resonance is somewhat narrower than a sequential Z' boson but has a substantially larger production cross-section times branching ratio to $t\bar{t}$. This same set of parameters was used by the D0 Collaboration. Table 5.1 summarizes the different mass points generated and corresponding signal cross-sections.

5.1.2 The Kaluza–Klein gluon

Models with extra dimensions provide an interesting solution to the hierarchy problem by allowing the graviton to propagate into the additional dimensions. If the SM gauge bosons are allowed to propagate into the extra dimensions, new heavier Kaluza–Klein (KK) partners for the gluon and electro-weak bosons are

Table 5.1: Cross-sections for the leptophobic Z' boson.

$m_{Z'}$ (GeV)	$\sigma(Z') \times \text{BR}(Z' \rightarrow t\bar{t})$ (pb)
600	10.3
700	5.6
800	3.2
900	1.9
1000	1.2
1200	0.46
1400	0.19
1600	0.068
1800	0.039
2000	0.018

expected with a mass separation inversely proportional to the size of the extra dimension. These gluons would escape detection in leptonic decay channels and can only be observed in experimentally challenging hadronic final states.

The first and lightest of the KK gluons from Randall–Sundrum (RS) models is considered here [31]. This heavy gluon has a natural predilection for heavy quarks. The benchmark scenario used is based on the basic RS setup. In this case, the couplings to light quarks are $-0.2g_s$, $0.2g_s$ for left-handed bottom quarks, g_s for right-handed bottom quarks and left-handed top quarks and, $4g_s$ for right-handed top quarks. The resulting Branching Ratio, $BR(g_{KK}) = 92.5\%$, offers a strong incentive to explore the $t\bar{t}$ final state. Phenomenologically, one of the challenges that such models present is that the width of such states is typically non-negligible with respect to the experimental resolution. In reference [145] many different parameter sets for the KK gluon, each with a quite different phenomenology, are discussed.

The setup of the authors of reference [31] is not viable for masses of the first KK state below approximately 2 TeV. While experimental bounds can be softened by invoking custodial symmetries [146, 147, 148] a KK gluon mass below 2 TeV is very difficult to reconcile with electroweak precision measurements from LEP. Nevertheless, the model has become a benchmark also for resonance searches with sensitivity below that mark. A complete discussion of the experimental bounds on warped 5D models is found in [149] and references therein.

The interaction $q\bar{q} \rightarrow g_{KK} \rightarrow t\bar{t} \rightarrow b\bar{b}q\bar{q}l^\pm\nu_l$ is generated with Madgraph [150, 151]. The parton shower and hadronization are simulated using PYTHIA.

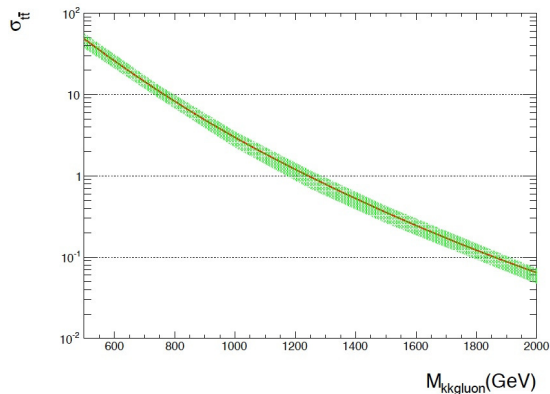


Figure 5.1: The $pp \rightarrow g_{KK} \rightarrow t\bar{t}$ cross section returned by MadGraph versus the invariant mass of the KK gluon.

Table 5.2 summarizes the different mass points generated and corresponding signal cross-sections.

Table 5.2: Cross-sections for the resonant KK gluon.

$m_{g_{KK}}$ (GeV)	$\sigma(g_{KK}) \times \text{BR}(g_{KK} \rightarrow t\bar{t})$ (pb)
600	39.4
700	20.8
800	11.6
900	6.8
1000	4.1
1200	1.7
1400	0.73
1600	0.35
1800	0.18
2000	0.095

The MadGraph prediction for the cross-section of the basic RS KK gluon times the Branching Ratio for the decay into a $t\bar{t}$ is presented in figure 5.1. This strict tree-level result is obtained with the CTEQ6L1 Parton Density Function (PDF).

The uncertainty on the production rate is estimated as follows:

- The renormalisation and factorisation scale are varied independently from the nominal value identified with the mass of the resonance, to twice and half the central value. The combinations that yield the largest upward and downward deviation from the nominal result are taken as the scale error, that should account for contributions from higher order diagrams. In figure 5.1 the scale uncertainty is indicated as a shaded band.
- The maximum upward and downward variation of the value of the cross section among results calculated with the error PDF sets provided by the CTEQ collaboration are taken as the systematic error due to the uncertainty in the PDFs. For a KK gluon mass of 1 TeV the CTEQ6 error sets yield a maximal variation in the cross section of 3%.
- The statistical uncertainty is negligible with respect to the systematic uncertainties and is ignored here.

Next-to-leading order effects can be partially accounted for by using the PYTHIA prediction with MRSTLO* PDF set. This result is found to be 10% larger than the tree-level prediction from MadGraph with CTEQ6L1. A recent NLO calculation of the cross-section for massive colour octet states [152] finds K-factors of that order. NLO effects have not been taken into account in this study.

The interference between diagrams involving the new coloured resonance with gluon-mediated SM production leads to non negligible effects on the $t\bar{t}$ mass distribution [153]. For instance, for the choice of parameters used here, figure 5.2 shows that the interference causes a sharpening of the resonant peak. This effect introduces a model-dependence that can form a severe practical obstacle to interpretation of the search results in different models: the exact interference pattern depends on choices of the couplings that yield otherwise compatible phenomenology. Most searches, including the one reported here, do not take into account the interference.

5.2 Data and Monte Carlo samples

The data used in this search were collected by the ATLAS detector at the LHC in 2011. Only data recorded under stable beam conditions and with all sub-detectors operational are used. Single muon and single electron triggers are applied with transverse momentum thresholds set at 18 GeV for muons and 20 GeV or 22 GeV for electrons. This corresponds to an integrated luminosity

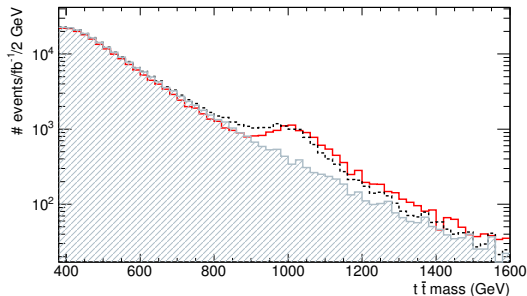


Figure 5.2: The $t\bar{t}$ invariant mass distribution. The shaded histogram corresponds to the Standard Model continuum, the dashed line to the sum of SM and g_{KK} resonant production and the continuous line to the full interference of SM and g_{KK} resonant production.

of $2.05 \pm 0.08 \text{ fb}^{-1}$ [103]. The object requirements used in the offline selection are more stringent than those used in the trigger, and the offline p_T thresholds are chosen on the efficiency plateau for the trigger.

After event generation, all simulated samples were run through a GEANT-based [127] simulation of the ATLAS detector and reconstructed using the same reconstruction software used for data. The trigger response is emulated in the offline software and simulated events were required to pass the corresponding simulated single electron or muon triggers as well.

The following SM background processes were simulated using the Monte Carlo technique:

- The irreducible ‘continuum’ $t\bar{t}$ background was simulated using the generator MC@NLO [154, 155] and showered using HERWIG [120] in association with JIMMY [156] to model effects due to the underlying event and multiple parton scattering. Only events in which at least one of the W bosons decays leptonically were produced, corresponding to a cross-section of 80.2 pb^{-1} to which a K-factor of 1.11 was applied to account for higher-order corrections. The K-factor normalises the sample to the approximate NNLO calculation from Hather [157], corresponding to a total $t\bar{t}$ cross-section of 164.6 pb^{-1} [158].
- Electroweak single top quark production was simulated using the same

programs, with leptonic W -boson decays required for the s - and t -channel processes leading to cross-sections of 1.4 pb and 21.5 pb respectively. All decays were produced for the Wt process using a cross-section of 14.6 pb. No K-factors were applied.

- W and Z plus jets production samples with leptonic vector boson decays were simulated with the ALPGEN [116] generator in exclusive bins of parton multiplicity for multiplicities lower than five, and inclusively above that. Wb and $Wb\bar{b}$ events were generated separately with ALPGEN. Double counting was avoided by removing events with b -quarks from the W +light jet samples. The Z plus jets events included photon interference and were required to have a dilepton invariant mass $40 < m_{\ell\ell} < 2000$ GeV. The events were showered with HERWIG and JIMMY and matching was performed with the MLM [159] method. The cross-sections and number of events for the different subsamples are given in table 5.3. Flat K-factors equal to 1.20 (1.25) are applied to the W +jets (Z +jets) samples. These K-factors reproduce the inclusive cross-section at NNLO in QCD [160] as calculated with FEWZ [161].
- Diboson samples have been produced using HERWIG with JIMMY, with a filter requiring the presence of one lepton with $p_T > 10$ GeV and $|\eta| < 2.8$. The cross-sections (K-factors) used for these filtered samples are: 11.75 pb (1.52) for WW production, 3.43 pb (1.58) for WZ production, and 0.98 pb (1.20) for ZZ production. The K-factors are such that the cross-sections agree with NLO QCD results [160] obtained using the MCFM [162, 163] generator.

All samples were simulated including the effects due to multiple pp interactions per bunch-crossing.

5.3 Object definition and event selection

The reconstruction of boosted $t\bar{t}$ events starts by the selection of one isolated light lepton. The leptonic top candidate emerges as the combination of the aforementioned lepton, the missing transverse energy and an anti- k_t , $R = 0.4$ jet. The hadronic top decay is reconstructed in one single anti- k_t , $R = 1.0$ jet in the hemisphere opposite to the lepton. The criteria applied in the reconstruction of the key objects for this analysis and the event selection scheme are described in the following sections.

Table 5.3: Cross-sections for the various $W \rightarrow \ell\nu$ plus jets and $Z \rightarrow \ell\ell$ plus jets subsamples.

Subsample	cross-sec. (pb)	Subsample	cross-sec. (pb)
$W \rightarrow e\nu + 0\text{lp}$	6913	$Z \rightarrow ee + 0\text{lp}$	662
$W \rightarrow e\nu + 1\text{lp}$	1293	$Z \rightarrow ee + 1\text{lp}$	133
$W \rightarrow e\nu + 2\text{lp}$	377	$Z \rightarrow ee + 2\text{lp}$	40.3
$W \rightarrow e\nu + 3\text{lp}$	101	$Z \rightarrow ee + 3\text{lp}$	11.2
$W \rightarrow e\nu + 4\text{lp}$	25.3	$Z \rightarrow ee + 4\text{lp}$	2.7
$W \rightarrow e\nu + 5\text{lp}$	6.9	$Z \rightarrow ee + 5\text{lp}$	0.8
$W \rightarrow \mu\nu + 0\text{lp}$	6935	$Z \rightarrow \mu\mu + 0\text{lp}$	658
$W \rightarrow \mu\nu + 1\text{lp}$	1281	$Z \rightarrow \mu\mu + 1\text{lp}$	133
$W \rightarrow \mu\nu + 2\text{lp}$	375	$Z \rightarrow \mu\mu + 2\text{lp}$	39.6
$W \rightarrow \mu\nu + 3\text{lp}$	101	$Z \rightarrow \mu\mu + 3\text{lp}$	11.1
$W \rightarrow \mu\nu + 4\text{lp}$	25.7	$Z \rightarrow \mu\mu + 4\text{lp}$	2.8
$W \rightarrow \mu\nu + 5\text{lp}$	7.0	$Z \rightarrow \mu\mu + 5\text{lp}$	0.8
$W \rightarrow \tau\nu + 0\text{lp}$	6836	$Z \rightarrow \tau\tau + 0\text{lp}$	657
$W \rightarrow \tau\nu + 1\text{lp}$	1277	$Z \rightarrow \tau\tau + 1\text{lp}$	133
$W \rightarrow \tau\nu + 2\text{lp}$	377	$Z \rightarrow \tau\tau + 2\text{lp}$	40.4
$W \rightarrow \tau\nu + 3\text{lp}$	101	$Z \rightarrow \tau\tau + 3\text{lp}$	11.0
$W \rightarrow \tau\nu + 4\text{lp}$	25.7	$Z \rightarrow \tau\tau + 4\text{lp}$	2.9
$W \rightarrow \tau\nu + 5\text{lp}$	7.0	$Z \rightarrow \tau\tau + 5\text{lp}$	0.7

5.3.1 Physics object selection criteria

Physics objects are required to satisfy the following requirements:

- Electrons must satisfy the standard quality criteria. That is, their EM shower must have a matching track in the ID [164] and EM shower's shape must be consistent with the expectations. The EM energy clusters are required to be within the EM calorimeter kinematic acceptance region, $E_T > 25$ GeV and $|\eta| < 2.47$, and excluding the calorimeter barrel-endcap crack region, $1.37 < |\eta| < 1.52$. Isolation is based on the energy deposits found in a cone of $\Delta R = 0.2$ around the electron. After subtracting the energy of the electron, an additional correction is applied to account for the energy deposited by particles from additional pp interactions. The total transverse energy found in the cone must be less than 3.5 GeV. This isolation requirement reduces the contamination by non-isolated electrons due to decays of hadrons (including heavy flavour) in jets.

- Muons are reconstructed by combining tracks from the Muon Chambers with tracks from the Inner Detector. The candidate trajectories are required to satisfy $|\eta| < 2.5$. The isolation criteria use both the energy deposition in the calorimeter, e_T , and the sum of the tracks transverse momenta, p_T , in a cone of $\Delta R = 0.3$ around the muon candidate. Non-isolated muons are rejected if both e_T and p_T are less than 4 GeV after subtracting the contribution of the muon to these quantities.
- Two anti- k_t jet collections are used in this study. Standard anti- k_t , $R = 0.4$ jets are reconstructed from topoclusters and calibrated using the EM+JES scheme. The reconstruction and calibration of anti- k_t , $R = 1.0$ jets using LC-topoclusters as well as the additional correction applied to the jet invariant mass are explained in section 4.4. The calibration factors derived for anti- k_t , $R = 1.0$ jets are only valid for $|\eta| < 2$, thus only jets within this region are considered.
- Missing transverse energy, E_T^{miss} , is the four-vector sum of the topoclusters which are not associated to any physics object of the event. Such association takes place in a chosen order: electrons, anti- k_t , $R = 0.4$ jets, and muons. The energy scale of the topoclusters is calibrated according to the physics object they are associated to. In order to avoid double counting, calorimeter cells associated to electrons are corrected omitting the out-of-cluster correction. After removing the clusters related to electrons and jets at their corrected energy scale, the contribution from the muons is subtracted. The remaining clusters are included in the computation of the E_T^{miss} at the EM energy scale.

Overlap between the different object categories is avoided by the following procedure. Jets within $\Delta R = 0.2$ of a good quality electron are removed. Events where the selected electron is separated by less than $\Delta R = 0.4$ of any $R = 0.4$ jet and with $p_T > 20$ GeV are rejected. Muons within $\Delta R = 0.4$ of any $R = 0.4$ jet with $p_T > 20$ GeV are rejected as well.

Several scale factors are applied to the reconstructed objects depending on their η and p_T , due to differences in identification, reconstruction and trigger efficiencies. The scale factors were obtained from simulation and the corresponding efficiencies were measured in collision data. The uncertainties on these scale factors are used to determine the corresponding systematic uncertainties.

5.3.2 Event selection

A number of selection criteria were applied to enhance the $t\bar{t}$ fraction of the sample:

1. Events must pass either the single-electron or single-muon trigger defined in section 5.2.
2. Events must have at least one reconstructed primary vertex with five or more associated tracks with $p_T > 400$ MeV.
3. Events are rejected if they contain a bad quality jet with $p_T > 20$ GeV. Bad quality jets usually appear due to out-of-time activity or calorimeter noise, or they are jets located in a problematic area of the calorimeter [112].
4. Events are required to have a good quality single isolated lepton as defined in section 5.3: one electron with $E_T > 25$ GeV or one muon with $p_T > 20$ GeV. The selected lepton must be the one that passed the corresponding trigger. Events are discarded if the selected electron shares an Inner Detector track with a non-isolated muon (muon close to a jet).
5. Different selections are applied to the electron and muon channels in order to suppress multi-jet events based on the fact that the E_T^{miss} in the event is dominated by the ν from the W -boson decay. The variables used are E_T^{miss} and the transverse mass defined as $M_T = \sqrt{2p_T^\ell E_T^{miss}(1 - \cos \Delta\phi)}$, where p_T^ℓ is the charged lepton p_T and $\Delta\phi$ is the azimuthal angle between the lepton and E_T^{miss} .
 - In the $e + \text{jets}$ channel, $E_T^{miss} > 35$ GeV and $M_T > 25$ GeV.
 - In the $\mu + \text{jets}$ channel, $E_T^{miss} > 20$ GeV and $E_T^{miss} + M_T > 60$ GeV.

Even assuming that the ν from the W -boson decay is the sole responsible for all the E_T^{miss} of the event, its z-component remains still undefined. The momentum of the neutrino can be reconstructed by imposing a W mass constraint on the lepton- E_T^{miss} system. Different scenarios can rise from the resulting quadratic equation.

- If there are two real solutions, the solution with the smallest $|p_z|$ is chosen.
- If the discriminant of the quadratic equation is negative, then E_T^{miss} is rescaled until the discriminant becomes null.

The selection steps listed so far, based on the reconstruction of the leptonic W boson candidate from the charged lepton and the signature of the ν , are similar to those of a previous analysis of the same final state [165].

The two tops are emitted in opposite directions in the transverse plane. In this study, the top-quarks decay products are assumed to approximately retain the direction of their parents. Therefore, the anti- k_t , $R = 1.0$ jet from the hadronic top quark is expected to be found approximately *back-to-back* in ϕ to the semi-leptonic top decay.

At this point, requirements on jets used in previous analysis, such as jet multiplicity or b-tagging, are substituted by an analysis of anti- k_t , $R = 1.0$ jets substructure, in particular, jet mass and $\sqrt{d_{12}}$ (see section 4.3.2).

The four-vector momentum associated with the $t\bar{t}$ system is reconstructed by adding the four-momenta of the semi-leptonically decaying top quark candidate and the hadronically decaying top quark candidate.

6. Events must contain at least one anti- k_t , $R = 0.4$ jet with $p_T > 30$ GeV within the region $0.4 < \Delta R(l, j) < 1.5$ defined around the direction of the charged lepton. When more than one anti- k_t , $R = 0.4$ jets are found, the one closest to the lepton is retained. The semi-leptonic top candidate is the four-vector sum of the lepton, the reconstructed neutrino candidate and the selected jet.
7. In order to prevent double counting of calorimeter cells from the two different jet collections, events are accepted only if they have at least one anti- k_t , $R = 0.4$ jet at a minimum distance $\Delta R(j, j) > 1.5$ from the anti- k_t , $R = 0.4$ jet associated with the leptonic top candidate.
8. The anti- k_t , $R = 1.0$ jet is required to have $p_T > 250$ GeV, $m_j > 100$ GeV and $\sqrt{d_{12}} > 40$ GeV. A large mass is expected if the jet generates from a top quark. The value of $\sqrt{d_{12}}$ should be of the order of the W mass. If several anti- k_t , $R = 1.0$ candidates are found, the one with the highest p_T is selected as the hadronic top candidate.
9. For a subset of the data period, a part of the EM calorimeter Front End was non-functional creating a dead region, a hole, in the LAr calorimeter. For this period, events are rejected if a jet with $p_T > 20$ GeV is closer than 0.1 in R to the problematic region. This procedure has been found to work well for anti- k_t , $R = 1.0$ jets too. In MC, the effect of the hole is modelled by randomly sampling a subset of the simulated sample, proportional to

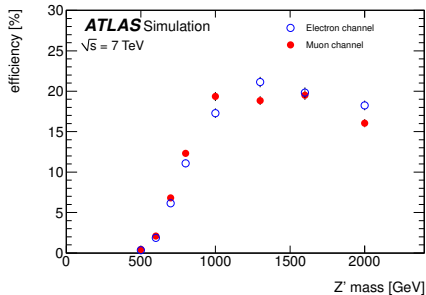


Figure 5.3: Total selection efficiency as a function of signal mass [166].

the integrated luminosity of the affected data period, for an artificial hole introduction.

The acceptance of this selection is approximately 15% for a 1 TeV Kaluza Klein gluon, while it is 20% in the classical analysis [165]. As the mass of the resonance increases, the classical algorithm decreases its acceptance, while that of this selection increases. At around 1.6 TeV both approaches cross with an acceptance of 17%. Even though the resulting sample of resonant $t\bar{t}$ events shows a strong overlap with the one obtained with the classical analysis, one third of the signal events selected here is not found in the classical selection. This strong dependence of the acceptance with the resonance mass is shown for $Z' \rightarrow t\bar{t}$ production in figure 5.3, where the acceptance of the selection increases for resonance masses well beyond the TeV. From 2 TeV on, however, the acceptance is degraded. This is mainly caused because, in the highly boosted regime, requirements such as lepton isolation or minimum distance between lepton and nearest jet cease to be applicable efficiently.

After this selection, the dominant background is SM $t\bar{t}$ production. In addition to this, QCD multi-jet production and W boson production in association with jets are expected to have an important contribution and their estimations are data-driven (see section 5.4). Further contributions from SM processes are expected to be small, and their prediction is based on MC estimates.

5.4 Data-driven background estimates

Apart from SM $t\bar{t}$ production, the event count in the signal region incorporates contributions from several SM processes, whose MC prediction suffers from large uncertainties. This is the case of contributions such as that of W boson production in association with jets or the contribution of QCD multi-jet production. In the later, the isolated lepton signature is faked by leptons from heavy-flavour decays such as muons from π^\pm and K^\pm decays in flight, electrons from photon conversions, or misidentified hadrons.

The Monte Carlo prediction for the contribution of several Standard Model processes to the event count in the signal region suffers from large uncertainties. In some cases, Monte Carlo statistics is moreover insufficient to make precise predictions. For the QCD multi-jet background and Standard Model W +jet production the background prediction is therefore based on measurements on data in carefully defined signal-free regions (control regions). In the following sections the results from this data-driven approach are presented.

5.4.1 QCD background

The QCD multi-jet simulation suffers from large systematic and statistical uncertainties. Therefore this background must be estimated from data directly. This is achieved by exploiting the phase space regions with leptons of less reconstruction quality. Such regions are generally more abundant with QCD multi-jet events. The event topology and kinematic criteria are chosen to resemble the definition of our signal region, in order to reduce potential systematic uncertainties. The so-called matrix method [167] was used to disentangle the mixture of non-prompt leptons from QCD, and prompt leptons from the W/Z bosons.

Applying a looser lepton definition, a sample of N_L events is created. Of these N_L loose events, a fraction, N_T , also passes the tight lepton definition used in this analysis. The remaining loose events, which do not pass the tight lepton selection, are called anti-tight, N_A , thus $N_L = N_T + N_A$.

N_A is a reasonably pure sample of QCD multi-jet events. N_T , on the other hand, contains contributions from prompt leptons from W or Z bosons produced in QCD or new physics processes. The components of N_T can, therefore, be divided into the fraction of events with prompt leptons that pass the tight selection, $\epsilon \times N_{prompt}$, and the fraction of events from QCD that pass the tight selection, $f \times N_{multi-jets}$, where the efficiency, ϵ , is the probability of a prompt lepton passing tighter signal selection criteria, and the fake-rate, f , is the

corresponding probability for leptons from multi-jet events. These parameters are measured in data.

For determination of the fake-rate f a control sample rich in multi-jet events is used, where the contamination from prompt leptons is estimated from MC simulation. The same procedure is applied on simulated data to ensure that the measured value of f in the control region is consistent with that of the signal region of interest. The systematic uncertainties on f include contributions from the uncertainty in the yield of the subtracted background sources and from the differences in the definition of the signal and multi-jet control regions.

The efficiency ϵ is determined using a tag-and-probe technique in a control sample of leptons from $Z^0 \rightarrow \ell\ell^-$, where at least one of the leptons must satisfy the tight quality criteria. In order to create a $Z \rightarrow e^+e^-$ -dominated sample, the invariant mass of the two leptons is required to be between 86 GeV and 96 GeV. MC samples were also studied to verify that the measured value of ϵ is applicable to the events of our interest. The systematic uncertainties on ϵ cover the differences between the efficiency for $Z \rightarrow ee$ events and other sources of electrons, i.e. the background and signal processes considered in this analysis.

This way, the multi-jet contribution to the signal region can be estimated as,

$$f \times N_{multi-jets} = \frac{(\epsilon - 1) f}{\epsilon - f} \times N_T + \frac{\epsilon f}{\epsilon - f} \times N_A \quad (5.1)$$

The weights that appear in equation 5.1 can be used to construct kinematic distributions for the multi-jet background of events with a tight lepton, with $(N_T, N_A) = (1, 0)$, and for events with a loose lepton that fails the tight cuts (anti-tight), with $(N_T, N_A) = (0, 1)$. Both efficiency and fake-rate are found to be largely independent of the kinematic variables of interest. Variations of the cuts on jet p_T , mass or $\sqrt{d_{12}}$ have no impact on f within the statistical and systematic uncertainties. The efficiency ϵ is constant in the $t\bar{t}$ invariant mass range studied.

5.4.2 W +Jets background

The W +jets background is simulated using the ALPGEN generator, showered with HERWIG and JIMMY as described in section 5.2. In order to reduce the uncertainty in the normalization, the total contribution of W +jets to the sample is estimated from the observed charge asymmetry in W -boson production in

the LHC [168],

$$N_{W+jets} = (N_{W^+} + N_{W^-})^{\text{pred}} = \left(\frac{r_{MC} + 1}{r_{MC} - 1} \right) (N_{W^+} - N_{W^-})^{\text{data}}, \quad (5.2)$$

The ratio $r_{MC} = N_{W^+}/N_{W^-}$ in equation 5.2 is derived from MC simulation.

The W+jets-enriched control sample is obtained by removing jet mass and $\sqrt{d_{12}}$ requirements and by loosening the cut on the transverse momentum of the hadronically decaying top quark candidate ($p_T > 150$ GeV). The contribution of all SM backgrounds but $t\bar{t}$ and W+jets is small in this control sample and is removed using the MC prediction. Standard Model $t\bar{t}$ production is charge-symmetric to good approximation, thus it does not contribute to the global charge asymmetry studied here. The resulting scale factors, 0.77 and 0.75, for the e +jets and μ +jets channel respectively, are compatible with unity within the uncertainty. The uncertainty in the normalization and shape of the W+jets background are discussed in Section 5.7.

5.5 $t\bar{t}$ mass reconstruction

The top quarks are produced approximately back-to-back and the boost confines their decay products into a small area of the detector. Thanks to this fact, the components of the top decay can be selected on the basis of vicinity (or indeed reconstructed as a single object) as described in section 4.3.1.

Leptonic top candidate

The selection described in section 4.3.1 provides a good quality and well isolated lepton that matches the corresponding trigger. All the E_T^{miss} in the event is associated to the transverse momentum of the escaping neutrino from the W boson decay. The longitudinal component of the momentum of the neutrino is obtained by imposing a W mass constraint on the lepton- E_T^{miss} system. Several scenarios can arise when solving the resulting quadratic equation. The solution with the smallest p_z is used when two real solutions exist. If there is no real solution, then the E_T^{miss} is rescaled until the discriminant in the quadratic equation becomes null.

The leptonic top quark is formed by adding an anti- k_t , $R = 0.4$ to the reconstructed W boson on the basis of proximity to the lepton. The jet with $p_T > 20$ GeV and within $\Delta R_{lj} < 1.5$ which is closer to the lepton is selected as the b-jet. Jets falling within a $\Delta R_{lj} < 0.4$ cone around the lepton are excluded.

This last requirement is the result of practical constraints. Measurements on data of the electron trigger and reconstruction efficiency for this particular region were not available at the time of this study. This requirement reduces the acceptance because well isolated leptons are less frequent in the busy environment of the boosted regime. Alternative lepton selection methods had been already tested in simulation [95] and were used in subsequent analyses [169].

Hadronic top candidate

The products of the decay of the hadronic top quark are expected to merge into a single anti- k_t , $R = 1.0$ jet. The ‘fat’ jet must fulfil: $p_T > 250$ GeV, jet mass $m_j > 100$ GeV, first k_T splitting scale $\sqrt{d_{12}} > 40$ GeV, and they must lay on the opposite hemisphere to the anti- k_t , $R = 0.4$ jet selected in the leptonic top reconstruction. The one with highest p_T is finally chosen as the hadronic top quark.

$t\bar{t}$ mass

The $t\bar{t}$ system is the result of the four-vector sum of the leptonic and hadronic top candidates. Radiation of the top quarks before they decay introduces additional jets in the event. Events where these jets are used in the reconstruction usually appear in the high-mass tail of the $t\bar{t}$ mass spectrum and can worsen the resolution. This is intrinsically addressed in the reconstruction procedure presented here. Jets from initial state radiation are typically far from the boosted top decay, and do not fulfil the vicinity criteria on which this reconstruction technique is based [95]. The truth-level and reconstructed $t\bar{t}$ mass distributions are shown in figure 5.4 for different KK gluon masses. The low-mass tail observed for large mass resonances is the result of the convolution of KK gluon line shape with the steeply falling parton luminosity function.

The $t\bar{t}$ mass resolution is presented in the left panel of figure 5.5 for the same KK gluon mass points. In the rightmost panel the relative deviation, $(m_{reco} - m_{truth})/m_{truth}$, is plotted. Only events where the truth-level $t\bar{t}$ mass is within 30% of the nominal resonance were used in the plots. The mass resolution is found to be approximately 10% for resonance masses between 1 TeV and 1.6 TeV. Similar results are obtained for the Z' samples.

In order to quantify the impact of pile-up on the mass resolution 10 additional interactions per bunch crossing have been considered. This nearly doubles the conditions in the data set used for this study. The expected resolution is degraded to 10-13%.

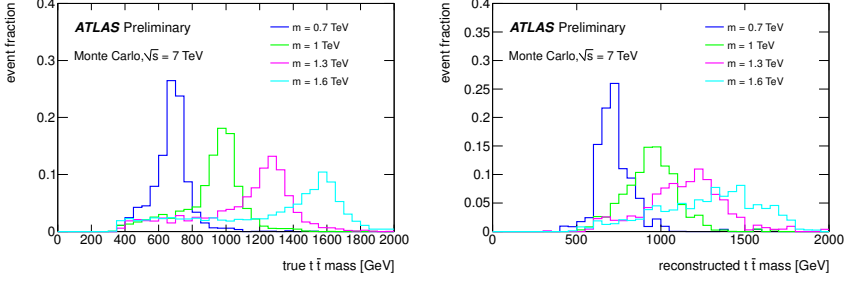


Figure 5.4: The truth-level $t\bar{t}$ mass distribution (left), and the reconstructed $t\bar{t}$ mass distribution after the event selection of section 5.3 (right) for KK gluon resonances in the 0.7-1.6 TeV mass range.

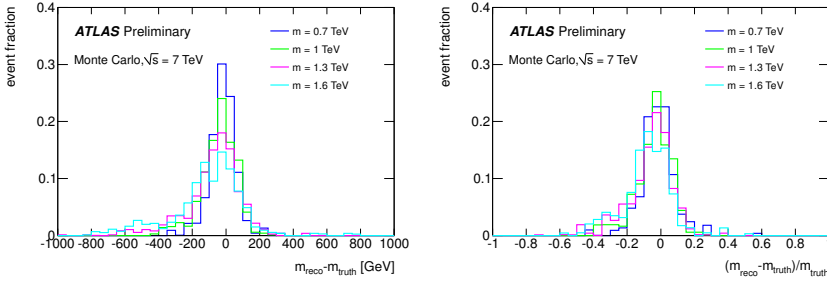


Figure 5.5: The $t\bar{t}$ mass resolution (left), and the relative deviation (right) for four different KK gluon masses.

5.6 Data-MC comparison

Table 5.4 lists the number of events selected with the recipe described in section 5.3 and the expected benchmark yields. The number of data events left after selection is in excellent agreement with the MC prediction. Systematic uncertainties on the expected yields are discussed in section 5.7.

The sum of the reconstructed $t\bar{t}$ mass spectra for the e +jets and μ +jets channel is presented in figure 5.6. The W +jets contribution has been scaled according to the description in section 5.4.2. Distributions for jet mass and $\sqrt{d_{12}}$ are shown in figure 5.7. In figure 5.8 the p_T distributions for leptons in

Table 5.4: Selected data events and expected background yields after full selection.

Type	e +jets	μ +jets	Sum
$t\bar{t}$	510 ± 40	620 ± 50	1130 ± 90
W +jets	202 ± 34	300 ± 50	500 ± 80
Multijets	45 ± 23	30 ± 15	75 ± 38
Z +jets	41 ± 20	34 ± 16	75 ± 36
Single top	21 ± 2	27 ± 3	48 ± 5
Dibosons	3.4 ± 0.2	4.5 ± 0.2	7.9 ± 0.4
Total	830 ± 60	1010 ± 70	1840 ± 130
Data	803	1034	1837

the signal region are presented.

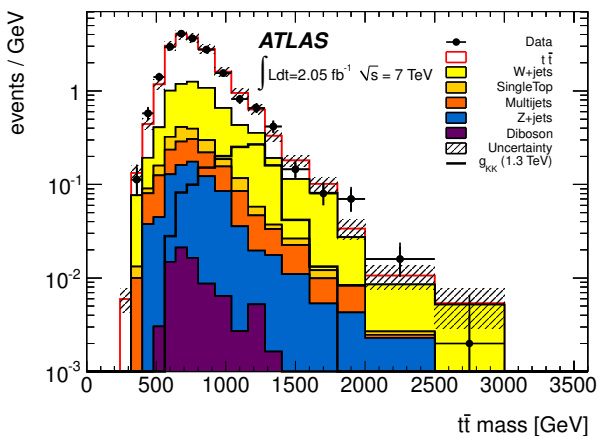


Figure 5.6: Reconstructed invariant mass distribution of the $t\bar{t}$ candidates after signal selection. The shaded band indicates the uncertainty in the normalization of the SM prediction, but does not include the shape uncertainty or the impact of uncertainties on reconstructed objects. The variable bin size is chosen to match the mass resolution for a resonant signal.

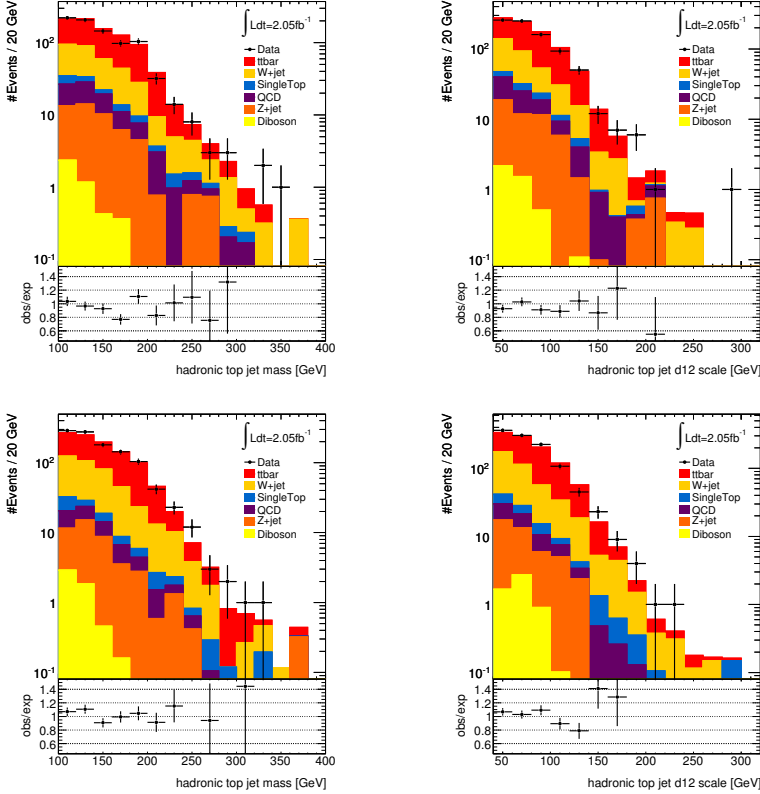


Figure 5.7: Mass and $\sqrt{d_{12}}$ of the hadronic top jet candidate (the anti- k_t , $R = 1.0$ jet) in the signal region for the e +jets channel (top panels) and for the μ +jets channel (bottom panels).

5.7 Systematic uncertainties

An overview of the systematic uncertainties associated to several aspects of this analysis is given in table 5.5. Such uncertainties include energy scales, reconstruction efficiencies and resolutions of the key objects of this study as well as normalisation and shape uncertainties associated to the simulated samples

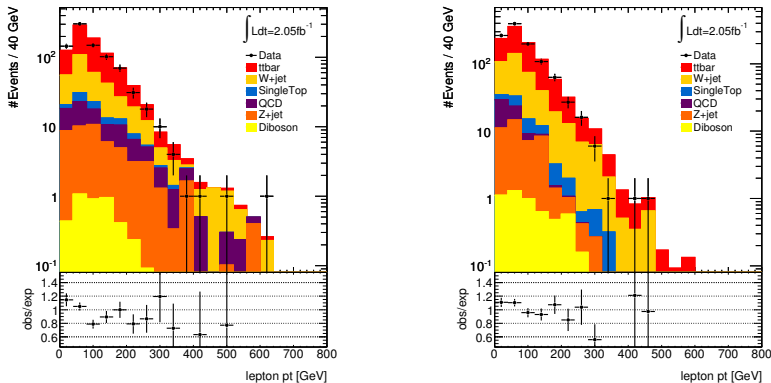


Figure 5.8: p_T distributions in the signal region for muons in the μ +jets channel (right) and electrons in the e +jets channel (left)

and the data-driven backgrounds. A total of 30 sources of systematic uncertainty are taken into account in the interpretation of the result in section 5.8. The relative impact of every single systematic effect to the expected sensitivity shown in the third column of table 5.5 is estimated by running the limit setting machinery using the nominal expected background as pseudo-data and the corresponding shifted background and a $1.3 \text{ TeV } Z'$ signal for comparison. Sources with negligible impact on the sensitivity are omitted and groups of related uncertainties have been combined in table 5.5. All uncertainties except ‘luminosity’ and those labelled ‘normalization’ affect the yield and the shape of the reconstructed mass distribution.

The total uncertainty on the luminosity measurement is 3.7% [103]. This uncertainty affects all MC-derived samples. Its impact on the background appears reduced in table 5.5 because two of the backgrounds, W +jets and multi-jets, are determined from data.

The effect of the uncertainty of the PDF on the shape is estimated using different PDF sets as recommended in the *PDF4LHC* manual [170]. A scale uncertainty is applied on the nominal $t\bar{t}$ spectrum to account for the effect of the uncertainty of the PDF on the normalisation. In addition, the $t\bar{t}$ normalization is affected by the uncertainty in the theoretical prediction of the $t\bar{t}$ production cross-section.

Table 5.5: Systematic uncertainties and their impact on the sensitivity. In the first two columns the relative impact (in percent) is shown on the total expected background yield (nominally 1840 events) and on the number of selected signal events (a Z' with a mass 1.3 TeV is chosen as the benchmark). The third column lists the relative variation for this benchmark of the expected limit on the production cross section times branching fraction if the corresponding systematic effect is ignored.

Systematic effect	Impact on yield [%]		Impact on sensitivity [%]
	background	Z' 1.3 TeV	
Luminosity	2.5	3.7	0.4
PDF uncertainty	3.1	1.0	0.2
$t\bar{t}$ normalization	4.9	—	0.7
$t\bar{t}$ ISR, FSR	6.3	—	0.7
$t\bar{t}$ fragmentation & parton shower	3.4	—	0.9
$t\bar{t}$ generator dependence	2.8	—	2.2
W + jets normalization	4.3	—	1.4
W + jets shape	<i>norm.</i>	—	0.1
Multi-jets normalization	2.1	—	0.2
Multi-jets shape	<i>norm.</i>	—	1.1
Z + jets normalization	2.0	—	0.5
Electron ID and reconstruction	1.1	1.3	1.0
Muon ID and reconstruction	2.2	2.1	4.8
Jet energy and mass scale	6.7	2.0	5.2
Jet energy and mass resolution	4.7	4.0	1.2

The imperfections in the modelling of the SM $t\bar{t}$ background are accounted for by using three different systematic uncertainties as described in [167]. The initial- and final-state radiation (ISR/FSR) systematic uncertainty is determined by considering $t\bar{t}$ samples generated with ACERMC each with different amounts of ISR and FSR, separately and in combination. ISR was found to be the largest effect and is taken as the overall systematic uncertainty. The uncertainty due to parton shower and fragmentation imperfections is gauged through the variation of POWHEG samples when the parton showering is done with PYTHIA or HERWIG. A similar method was used to study the possible impact of the generator. In this case, MC@NLO and POWHEG were compared, both using HERWIG for the parton showering. These uncertainties individually lead to 3–6% variations in the total background yield.

The statistical uncertainty on the yield of W + jets amounts to less than

10%. A systematic uncertainty of 14% is introduced as well to account for a variety of effects on the normalisation and shape of the $W + jets$ contribution to the $m_{t\bar{t}}$ distribution. The $W + jets$ PDF uncertainty is treated by varying the different PDF sets, and then normalising them to the nominal yield. This effect is completely absorbed by the overall W normalisation uncertainty. Possible effects due to the extrapolation from the control region to the signal region are accounted for in the jet scale and resolution uncertainties. As a cross-check, the normalization procedure described in section 5.4.2 was repeated on several $W + jets$ validation regions and the results were found to be consistent.

The shape uncertainty includes the impact of the systematic uncertainties on jets and a $W + jets$ modelling uncertainty. The latter is obtained using ALPGEN samples generated with different values for the functional form of the factorisation scale and the minimum p_T of the partons. The resulting spectra are normalised to the nominal yield times the data-driven scale factor so that only shape effects are considered.

A constant normalization uncertainty of 50% is applied to the data-driven multi-jet background estimation. To determine the shape uncertainty, the procedure described in section 5.4.1 to derive from data the multi-jet contribution to the signal region was applied on samples made with two different loose lepton selection criteria and the result was compared.

The systematic uncertainties on backgrounds with minor impact in this study such as $Z + jets$, single top and dibosons, are estimated with flat scalings of the spectra following the recommendations. These have a small impact on the final limits.

The reconstruction of charged leptons is affected by several sources of systematic uncertainty. The trigger scale factor and efficiency, the reconstruction efficiency, and the resolution have been taken into account. The values associated to muons, specially the impact on the sensitivity, were overestimated at the time of the publication of this study. Subsequent analyses show values consistent with those of the electrons [169].

The jet energy scale uncertainty is an important source of systematic uncertainty on the acceptance and produces the most significant impact on the sensitivity. It is less than 3% for anti- k_t , $R = 0.4$ jets in the energy range of interest to this study. A slightly larger energy scale is associated to anti- k_t , $R = 1.0$ jets together with an additional uncertainty on the jet mass of 4-5%. An additional uncertainty of 1% is added to the anti- k_t , $R = 1.0$ jet energy scale due to the pile-up modelling imperfections and non-closure of the MC reweighting procedure. This has been described in section 4.4.4 and in reference [132]. The two jet collections are treated as fully correlated for the limit combination. The

uncertainty on the pile-up modelling is propagated separately.

5.8 Interpretation. Compatibility with the null hypothesis

After the reconstruction of the $t\bar{t}$ mass spectrum, the compatibility of the data with the SM-only (null) hypothesis is evaluated to search for hints of new physics in the form of bumps or dips in the spectrum.

The combination of the results from the two channels, e +jets and μ +jets, is done as follows. Instead of performing the search on the combined spectrum resulting from the addition of the two spectra, which entails losing information regarding the intrinsic differences between the two channels (acceptance, shape of the signal, etc.), the option chosen is to search for overlaps between the spectra from the two channels. This is based on the fact that, if a $t\bar{t}$ resonance exists, bumps will arise in both the e +jets and μ +jets spectra at approximately the same mass point.

The search procedure is done systematically with the BumpHunter [171], a hypothesis testing tool that searches for local data excesses or deficits of varying width compared to the expected background. The p -value represents the probability for an observed excess or deficit to be found under the assumption of the null hypothesis. The p -value of the most interesting bump or dip is found by comparing the test statistic from data with the test statistics found in N pseudo-experiments. In this case, ten thousand pseudo-experiments were generated by Poisson fluctuations of the expected background. A p -value of 1 means that the bump or dip is not significant, as no pseudo-experiment gave a BumpHunter test statistic smaller than the one in data. A small p -value, on the other hand, is indicative of a significant deviation, as the test statistic found in data was large compared to the pseudo-experiment results.

The most significant excess is found in the $t\bar{t}$ mass region between 1.8 and 2.5 TeV. It is most pronounced in the electron channel. When the systematic uncertainties are accounted for, the p -value is 0.08 (1.4σ), including the look-elsewhere effect, evaluated over the full mass range. No other deviations with respect to the SM prediction with a significance beyond 1σ are found. Thus we can conclude that there is no indication of $t\bar{t}$ resonances in this data set.

5.8.1 Limits on benchmark models

Upper limits are set at 95% CL on the benchmark models production cross-section times branching fraction using Bayesian techniques [172]. The prior

probability distribution used in this method, which is flat in the cross-section, is a good approximation of the reference prior [173], and the likelihood is calculated using a Poisson function. The systematic uncertainties described in section 5.7 are found to have a significant impact on the sensitivity: in the 1.0–1.5 TeV mass range the limit on the rate including all systematic uncertainties is typically a factor two weaker than the limit that would be derived with statistical uncertainties only. The systematic uncertainties are accounted for by assuming they are normally distributed and convolving a Gaussian with the posterior probability distribution for each one. They are only weakly constrained by the data. Two exceptions are the jet energy scale and the $t\bar{t}$ generator dependence, which are constrained to about a half or a third, respectively, of their prior uncertainty. The result has been cross-checked with the so-called CL_s method [174, 175] and is in good agreement with it. Not allowing the data to constrain the systematic uncertainties in the CL_s method reduces the sensitivity on the signal cross-section by no more than 20%, which is less than the expected 1σ variation.

The resulting limits for the narrow Z' and the broad coloured KK gluon are presented in figure 5.9 and in table 5.6.

In the case of the narrow resonance, upper limits on $\sigma_{Z'} \times BR(t\bar{t})$ range from approximately 8 pb for $M_{Z'} = 0.6$ TeV to 610 fb at 1 TeV and 220 fb at 1.6 TeV. These are in good agreement with the expected limits and lead to the exclusion of the mass range between 0.6 TeV and 1.15 TeV for the leptophobic topcolor Z' model considered here [144].

The observed limits on the broad ($\Gamma/m = 15.3\%$) KK gluon are 650 fb at 1 TeV and 370 fb at 1.6 TeV. They are slightly weaker than those on the Z' boson due to the impact of the resonance's width, impact that is observed also in the expected limits and is most pronounced for large resonance masses. At 1 TeV the expected limit on the Z' boson is a 30% stronger than the KK gluon limit and the difference reaches a factor two at 2 TeV. The KK gluon model of Lillie et al. [31] is excluded for a resonance mass below 1.5 TeV, again in good agreement with the expectation.

The observed upper cross-section limits for a Z' and a KK gluon at 2 TeV are about 1.5σ higher than the expected value, reflecting the small data excess at 1.8–2.5 TeV, seen in figure 5.6.

5.8. Interpretation. Compatibility with the null hypothesis

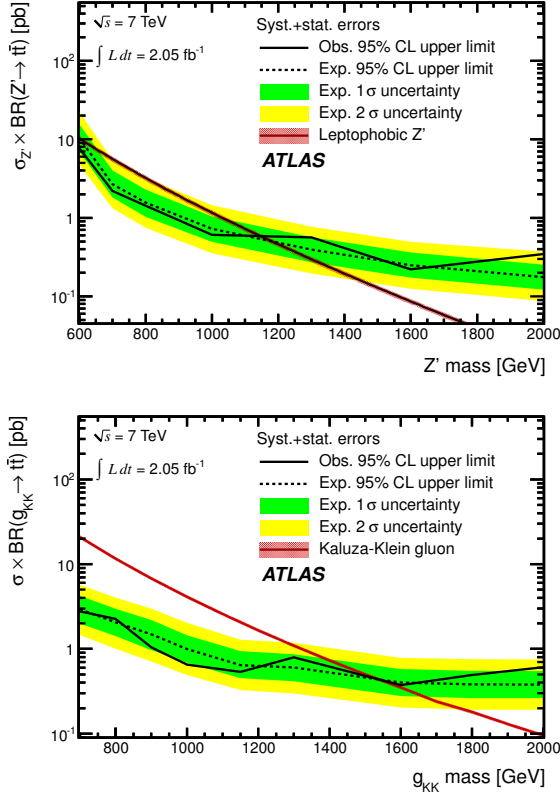


Figure 5.9: Expected (dashed line) and observed (solid line) upper limits on the production cross-section times the $t\bar{t}$ branching fraction of (top) Z' and (bottom) Kaluza–Klein gluons. The dark (green) and light (yellow) bands show the range in which the limit is expected to lie in 68% and 95% of pseudo-experiments, respectively, and the smooth solid (red) lines correspond to the predicted production cross-section times branching fraction for the model. The band around the signal cross section curve is based on the effect of the PDF uncertainty on the prediction [166].

Table 5.6: Observed and expected upper limits on the production cross section times branching fraction for $Z' \rightarrow t\bar{t}$ and $g_{KK} \rightarrow t\bar{t}$ respectively, including systematic and statistical uncertainties [166]. The expected limit $\pm 1\sigma$ variation is also given.

$Z' \rightarrow t\bar{t}$ limits				
Z' Mass [GeV]	Observed [pb]	Expected [pb]	-1σ [pb]	$+1\sigma$ [pb]
600	7.7	10.4	7.0	15.6
700	2.2	2.7	1.8	4.0
800	1.4	1.6	1.0	2.3
1000	0.61	0.72	0.49	1.0
1300	0.56	0.39	0.27	0.57
1600	0.22	0.25	0.17	0.36
2000	0.34	0.18	0.12	0.25
3000	0.27	0.27	0.19	0.41
$g_{KK} \rightarrow t\bar{t}$ limits				
g_{KK} Mass [GeV]	Observed [pb]	Expected [pb]	-1σ [pb]	$+1\sigma$ [pb]
700	2.8	2.9	2.0	4.2
800	2.3	2.1	1.4	3.0
900	1.0	1.5	0.97	2.2
1000	0.65	0.99	0.69	1.4
1150	0.53	0.64	0.45	0.94
1300	0.80	0.60	0.42	0.87
1600	0.37	0.40	0.28	0.58
1800	0.49	0.38	0.26	0.55
2000	0.61	0.38	0.26	0.55

5.9 Summary

The results reported in this chapter are based on an analysis of 2.05 fb^{-1} of $\sqrt{s} = 7 \text{ TeV}$ ATLAS data collected during 2011 and have been published in *Journal of High Energy Physics* [166]. The selection and reconstruction of $t\bar{t}$ pairs in the $l + jets$ decay channel using a novel technique aimed at events with boosted topologies is described. The reconstruction scheme relies on vicinity criteria and the jet substructure observables presented in chapter 4 and in references [95] and [132]: jet mass and the first k_t splitting scale $\sqrt{d_{12}}$. It should be noted that the current result represents a partial implementation of the algorithm designed for highly boosted top quarks. The classical isolation requirement applied in the lepton selection causes a drop in the acceptance

from 2 TeV on.

The data set is compared to a template for the expected SM $t\bar{t}$ mass spectrum. The template is a combination of fully data-driven QCD inclusive jet production estimates, a MC simulation of W +jets production with a data-driven normalization, and MC simulations for several other SM processes as described in sections 5.2 and 5.4.

Systematic uncertainties in the normalization and shape of the SM prediction are evaluated based on MC studies and in-situ techniques in section 5.7. The dominant uncertainties are found to be the jet energy and mass scale and the normalization of the W +jets background.

No significant bump has been found in the reconstructed $t\bar{t}$ spectrum once the systematic uncertainties have been taken into account.

Upper limits on the cross-section times branching ratio of the narrow Z' resonance range from approximately 30 pb for a mass of 500 GeV to 100 fb for masses beyond 2 TeV, in good agreement with the expected limits. For the leptophobic Z' of the topcolor model of reference [142] a mass smaller than 1.2 TeV is excluded by this analysis.

Limits are also derived for broader resonances that are typical of models with colored heavy objects. The expected upper limit on the cross-section times branching ratio of the KK gluon ($\Gamma = 0.15M$) of reference [31] is approximately a factor two larger than that of a Z' boson of the same mass. A lower limit on the KK gluon mass of 1.6 TeV is found.

This is the first application of the boosted paradigm in ATLAS data. This algorithm performs specially well in the mass range from 800 GeV compared to other coetaneous resonance searches [165, 176]. The reconstruction is moreover intrinsically robust against initial state radiation and pile-up. All this proves that reconstruction techniques aimed at boosted objects can greatly enhance the potential of searches for new physics beyond the SM in the LHC.

A summary of the evolution of the searches for a narrow Z' resonance is shown in figure 5.10. The results on the 4.7 fb^{-1} of 7 TeV data from ATLAS that appear in the figure have been produced with a complete implementation of the algorithm. This implementation includes dedicated lepton isolation and selection criteria, and the use of new substructure techniques designed to reduce the impact of pile-up. Preliminary results on the 8 TeV data set have pushed the limits further up.

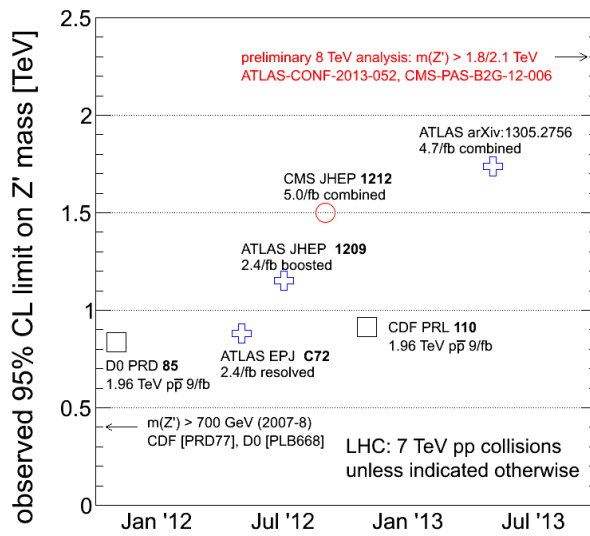


Figure 5.10: Evolution of the searches for a narrow Z' resonance decaying into a $t\bar{t}$ pair [102]. The analysis described in this thesis appears as *ATLAS JHEP 1209 2.4/fb boosted*.

Appendices

A First observation of high p_T top quarks with a boosted topology

A.1 Introduction

In the 2010 data set, with an integrated luminosity of 35 pb^{-1} , a few thousands of $t\bar{t}$ candidate events were recorded. The production cross-section was measured in the same data set to be $\sigma_{t\bar{t}} = 145 \pm 31_{-27}^{+42} \text{ pb}$ [177]. The candidate $t\bar{t}$ pair with the highest reconstructed invariant mass had a mass well beyond 1 TeV. The events in the high-mass tail are expected to yield boosted topologies. A detailed study of such events was an important step in the commissioning of reconstruction algorithms for boosted objects.

In this appendix we proceed to isolate $t\bar{t}$ events where at least one of the top candidates is reconstructed as a single jet. We rely on the standard selection used by the ATLAS Collaboration to form a clean sample of $t\bar{t}$ events in the lepton+jets final state [177]. We then reconstruct top and anti-top quark candidates following a simple algorithm [178]. The $t\bar{t}$ mass is determined by adding the four-vectors of the four hardest jets, the charged lepton and the neutrino. The sample of events with large invariant mass of the $t\bar{t}$ pair ($m_{t\bar{t}} > 700 \text{ GeV}$) is the starting point for our search. We re-reconstruct these events using an algorithm specifically designed for the boosted topology that arises in the decay of high p_T top quarks [95] and determine the values of the corresponding top-tagging observables.

This appendix is organized as follows. We define the data and Monte Carlo samples used in Section A.2. The basic selection strategy for $t\bar{t}$ events and that of high mass candidates is discussed in Section A.3. In Section A.4 we provide an outline of the boosted approach to top reconstruction and define a number of observables. The candidates are presented in Section A.5, where event displays and summary tables with more quantitative information on

selected jet substructure and embedded lepton observables are presented for the selected candidates. Finally, in Section A.6 we summarize our findings.

A.2 Data and simulated samples

The data set was recorded during 2010. Only events recorded under stable beam conditions and with all subsystems of the detector fully operational are used. The integrated luminosity is 35 pb^{-1} .

We use the same Monte Carlo simulations used to reproduce the data composition in the top cross section measurement [177]. For the samples of top quark events the next-to-leading order (NLO) generator MC@NLO v3.41 [154, 155, 179], is used with an assumed top-quark mass of 172.5 GeV and with the NLO parton density function (PDF) set CTEQ66 [180]. In the case of the small background of single-top production, the ‘diagram removal scheme’ [181] is used to remove overlaps between the single-top and the $t\bar{t}$ final states. Alpgen v2.13 [116] and the LO PDF set CTEQ6L1 [182], are used to generate QCD multi-jet events and W/Z boson production in association with multiple jets. Events are hadronised with Herwig, using Jimmy for the underlying event model. After generation and full simulation of the ATLAS detector response through GEANT, simulated events are reconstructed with the same tools used for the data.

A.3 Trigger and selection

The object definition, trigger and selection used were developed for the $t\bar{t}$ cross section measurement [177]. Here, we present only a summary:

- The electron cluster must have $p_T > 20 \text{ GeV}$ and pseudo-rapidity $|\eta_{cluster}| < 2.47$ (without being in the transition region $1.37 < |\eta_{cluster}| < 1.52$). In addition requirements are imposed on the ratio of the electron cluster energy to the track momentum, the presence of a hit in the innermost pixel layer and on the calorimeter cluster isolation as a function on the cluster p_T .
- Muons are derived from track segments in the muon chambers. Then the tracks from the muon chambers are matched to tracks found in the inner detector. The muon candidates must have $p_T > 20 \text{ GeV}$ and $|\eta_{cluster}| < 2.5$. In addition, isolation is required both in the calorimeter and in the

tracker and muons are required to have an angular separation $\Delta R > 0.4$ ¹ from any jet with $p_T > 20$.

- Jets are reconstructed from topological clusters of energy deposits in the calorimeter using the anti- k_T algorithm with a resolution parameter $R = 0.4$. Jets are first calibrated at the electromagnetic energy scale (EM). Afterwards, jets are calibrated at the hadronic energy scale using a simulation-derived correction. Double counting of electrons as jets is avoided by removing the closest jet to an electron candidate if $\Delta R < 0.2$. Jets are used only if they have $p_T > 25$ GeV, $\eta_{jet} < 2.5$ and they meet the standard jet/ E_T^{miss} quality requirements.
- Jets are b-tagged using the SV0 algorithm [183]. The SV0 algorithm is based on the fact that the b-hadrons have a typical flight path of a few millimeters that can be observed in the detector. A jet is considered as b-tagged if it contains a secondary vertex with $L/\sigma(L) > 5.72$, where L is the decay length and σ its uncertainty.
- Missing transverse energy, E_T^{miss} , is obtained from the vector sum of the transverse energy of the calorimeter cells associated to electrons (at the EM scale) and jets (at the calibrated scale) plus the transverse momentum of the selected muons.

A single lepton trigger (electron or muon) is used. The trigger requirements need to change through the data taking period in order to match the rapid increase in LHC luminosity and to incorporate the last results from the commissioning of the trigger system itself. Despite the changes, the thresholds are chosen low enough to ensure that leptons with $p_T > 20$ GeV lie in the efficiency plateau.

Events are discarded if no jet with $p_T > 10$ GeV at the EM scale does pass the standard quality criteria or if a primary vertex with more than four good tracks is not present.

The $t\bar{t}$ pair reconstruction applied both to data and simulated samples is summarised in table A.1. After b -tagging, this selection is expected to yield a high purity lepton+jets $t\bar{t}$ sample. The data-driven estimates of QCD and W+jets backgrounds follow exactly the same procedures described in [178].

The MC expectation for the contributions from different processes yields an estimate of the signal and background contributions to the selected sample.

¹ ΔR is the distance in the η, ϕ plane, $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$

Table A.1: Event selection requirements.

Cut	Description
C0	Single lepton trigger (electron or muon) fired
C1	Only 1 reconstructed lepton with $p_T > 20$ GeV
C2	Lepton matches corresponding trigger
C3	Non-collision background rejection: Require a primary vertex with number of good tracks > 4
C4	Jet cleaning: no bad jets with EM scale $p_T > 10$ GeV
C5	Transverse missing energy > 20 GeV
C6	Transverse missing energy + W -boson transverse mass > 60 GeV
C7	At least 4 jets with $p_T > 25$ GeV and $ \eta < 2.5$
C8	≥ 1 good jet ($p_T > 25$ GeV and $ \eta < 2.5$) with b-tagging SV0 weight $L/\sigma(L) > 5.72$

For the electron sample we expect 2.7 signal ($t\bar{t}$) events for each event due to background (the sum of contributions due to single-top production, vector boson + jets production, di-boson production and QCD jet production). For the muon sample the S/B ratio is 4.2. The various contributions of data and simulated events are shown in table A.2.

Table A.2: Number of expected and observed events for the electron and muon channels after application of all selection cuts described in this section.

	Electron channel		Muon channel	
	all	$m_{t\bar{t}} > 700$ GeV	all	$m_{t\bar{t}} > 700$ GeV
$t\bar{t}$	183.4	33.7	196.9	36.2
Single top	9.0	2.3	9.5	3.1
W +jets	17.8	6.1	23.5	7.0
Z +jets	2.1	0.5	1.2	0.5
Diboson	0.1	0.04	0.1	0.04
QCD	39.0	9.1	12.2	2.8
Total expected	251.0	51.7	239.6	49.6
Data observed	241	44	235	38

The initial $t\bar{t}$ pair reconstruction is based on a robust algorithm labeled *four hardest jet* in reference [178]. Exactly the same algorithm is used in reference [95], where it is labeled *minimal approach*. The $t\bar{t}$ mass is formed by adding the four highest p_T jets to the estimated neutrino four momentum. The

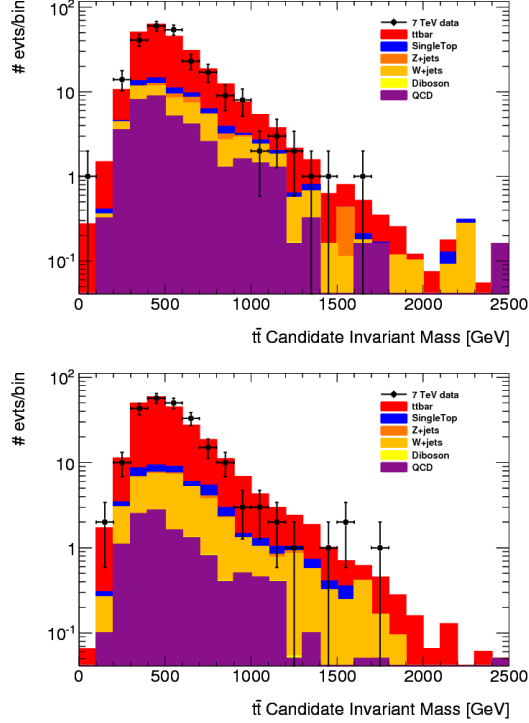


Figure A.1: Reconstructed $t\bar{t}$ mass distribution for the 2010 data set. (top) electron channel and (bottom) muon channel. The MC expectations for contributions from SM processes are indicated with different fill colours.

neutrino's longitudinal momentum (ν_{p_z}) is determined by imposing a W-boson mass constraint. If the discriminant of the quadratic equation is negative, the missing transverse energy is re-scaled until a null discriminant is obtained. The reconstructed $t\bar{t}$ mass spectra for events with electrons and muons are shown in figure A.1.

For the purpose of this study only those events that populate the high mass tail of the $m_{t\bar{t}}$ spectrum are retained. We find 82 events with $m_{t\bar{t}} > 700$ GeV. Again, MC is used to estimate the purity of this sample. The S/B ratio is expected to be of 1.8 and 2.7 for the electron and muon samples, respectively.

Even though the purity is slightly deteriorated with respect to that of the full sample, we can be confident that the majority of selected events indeed correspond to SM $t\bar{t}$ production.

A.4 Reconstruction of fat jets

We re-reconstruct the 82 events with large $m_{t\bar{t}}$ following the boosted approach. This approach was developed primarily for highly boosted top quarks [140], but has since been shown to be competitive already at rather modest top quark p_T [95].

The boosted approach is based on jets reconstructed with the anti- k_T algorithm with a resolution parameter $R = 1.0$ on locally calibrated topological calorimeter clusters. The calibration for these non-standard jets was taken from reference [132]. Jet-level corrections are derived using a MC-based procedure employed also to the standard anti- k_T jets with $R = 0.4$ and 0.6 . In the standard procedure jet-level corrections are applied only to jet energy and pseudo-rapidity. For the anti- k_T jets with $R = 1.0$, a correction is also applied to the jet invariant mass.

The baseline selection for the boosted approach is based on a requirement on jet invariant mass, $m_j > 100$ GeV [95]. In these MC studies, the requirement on jet invariant mass is combined with a cut on the energy sharing variable $z_{12} > 0.08$ ($z_{ij} = d_{ij}/(d_{ij} + m_j^2)$) and on the reconstructed W -boson mass (Q_W). At this stage, we will not reconstruct these complex observables with uncertain calibration, but concentrate on the splitting scales d_{ij} instead.

The isolated lepton from the W -decay in the lepton+jets final state is a powerful handle to select the signal. In reference [95] a number of observables were used as discriminants between the $t\bar{t}$ signal and the QCD (and W + jets) background:

$$z_l = \frac{E_l}{E_l + E_j}, \quad x_l = \frac{2p_l \cdot p_j}{(p_l + p_j)^2}, \quad y_l = p_{l\perp j} \times \Delta R_{lj}$$

where E_j and E_l are the jet and lepton energies, p_j and p_l their momenta, and $p_{l\perp j}$ the lepton transverse momentum with respect to the candidate b -jet. The variable y_l yields a measure similar to (the square root of) the k_T distance used by the k_T clustering algorithm [140]. The following cuts were applied for all leptons:

The most striking difference between the selection and initial reconstruction of section A.3 and the boosted approach lies in the resolution parameter (jet

Table A.3: Isolated lepton-based observables used to discriminate between $t\bar{t}$ signal and the QCD background.

	All leptons	Electron channel	Muon channel
ΔR_{lj}	< 1	> 0.25	> 0.15
x_l	< 1.2	> 0.4	> 0.35
z_l	< 0.8	> 0	> 0.15

size) used for jet reconstruction. Both rely on the anti- k_T algorithm to cluster the topological clusters in the calorimeter, but the jets of the initial selection are reconstructed with $R=0.4$, while for the boosted approach $R=1.0$ is used.

A.5 Candidate events

For each of the 82 selected events we carefully compared the basic objects used in the event selection of section A.3 with the result of the re-reconstruction of section A.4 with $R=1.0$. We concentrated particularly on events where several of the jets reconstructed with $R=0.4$ ‘merge’ into a single jet when the event is reclustered with $R=1.0$.

Our search so far has selected several solid candidates for top fat jets where the hadronic top quark decay is reconstructed as a single jet. We present the event displays and a summary of the quantitative results of two of them in the following pages.

The leptonic top candidate in the event displayed in table A.4 is formed by an isolated muon, E_T^{miss} and a b-tagged $R=0.4$ jet. A second muon reconstructed inside this jet corroborates the heavy flavour hypothesis. In the opposite ϕ hemisphere, three jets are reconstructed with $R=0.4$, which combined mass, $m_{jet1+jet3+jet4}$, totals 176.5 GeV. The picture does change with $R=1.0$, where the three jets merge to form a single jet with a mass of 225 GeV, $\sqrt{d_{12}}=105$ GeV and $\sqrt{d_{23}}=44$ GeV.

In the event shown in table A.5 the three $R=0.4$ jets that form the hadronic top have a combined mass, $m_{jet2+jet3+jet4}$, of 179.1 GeV. When reclustered with $R=1.0$ the three jets merge into a single fat jet with $m_j=197.1$ GeV, $\sqrt{d_{12}}=110$ GeV and $\sqrt{d_{23}}=40$ GeV. The measured values of the substructure of the fat jet, suggest that it has indeed been generated by a boosted top quark. The leptonic top products are located in the opposite ϕ hemisphere, a feature

expected in boosted topologies as well. It is formed by a high p_T electron, moderate E_T^{miss} and a b-tagged jet.

In figure A.2 the measured values for jet mass and splitting scales are indicated as arrows. They are superposed on the expected distributions from Monte Carlo for QCD di-jet production and SM $t\bar{t}$ production with $m_{t\bar{t}} < 700$ GeV and $m_{t\bar{t}} > 700$ GeV. These distributions correspond to jets reconstructed from simulated topological calorimeter clusters and calibrated according to the MC-based procedure for jets in an inclusive QCD sample described before. No event selection is applied and all jets with $p_T > 100$ GeV are plotted.

The observed values for the jet mass of the boosted top candidates are compatible with the peak that forms around the top mass in the $m_{t\bar{t}} > 700$ GeV distribution.

A.6 Conclusions

We have isolated a clean sample of $t\bar{t}$ candidate events in the data set collected in 2010. The invariant mass of the candidate $t\bar{t}$ pairs reaches 1.6 TeV. After a close inspection of the 82 events with greatest $m_{t\bar{t}}$ interesting topologies were selected. In these events the three anti- k_T jets (with $R = 0.4$) corresponding to the hadronically decaying top candidate merge into a single jet when the event is reclustered with a larger jet size ($R = 1.0$). Several jet substructure observables, such as the invariant mass of the ($R = 1.0$) anti- k_T jet and its k_T splitting scales, corroborate that these events represent the first top quarks with a boosted topology detected in the ATLAS experiment. These events were applauded at the BOOST2011 conference as the world's first boosted heavy particles ever seen [184].

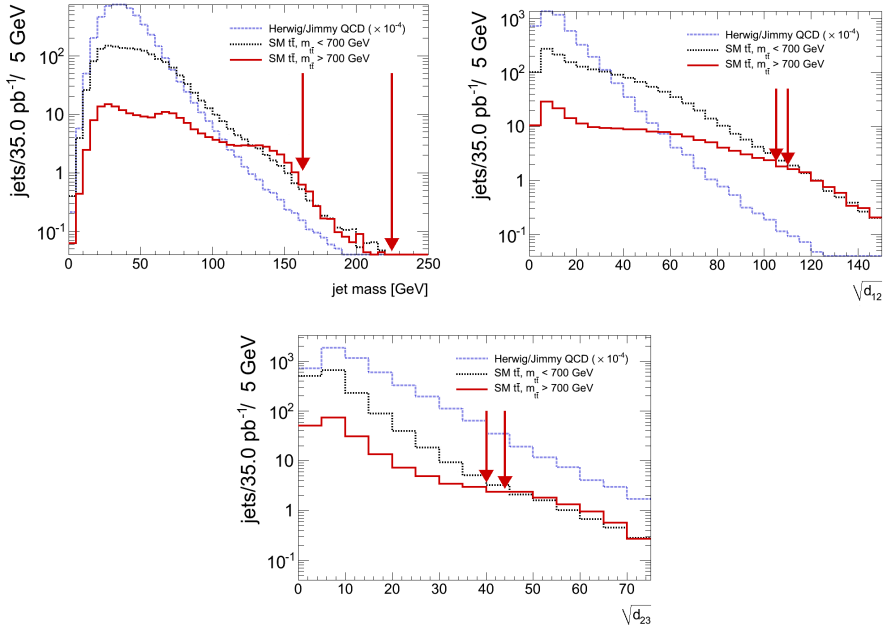
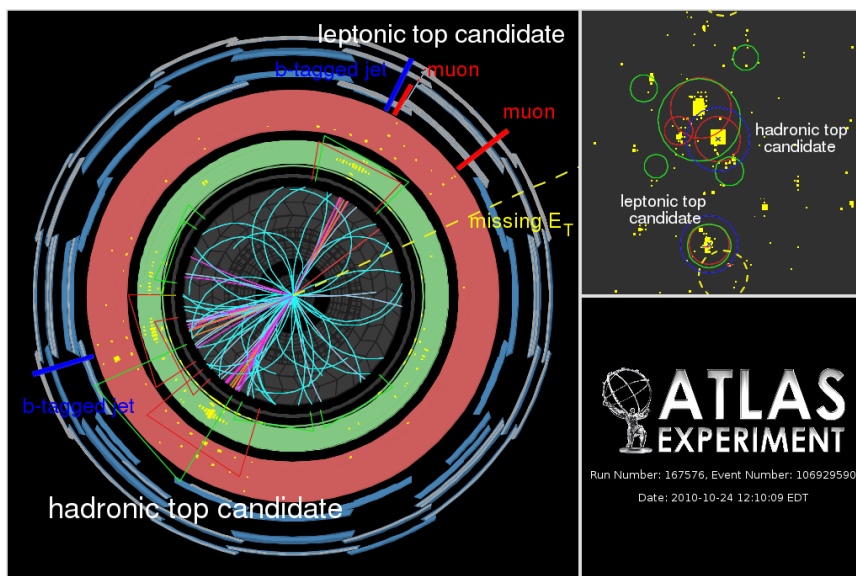


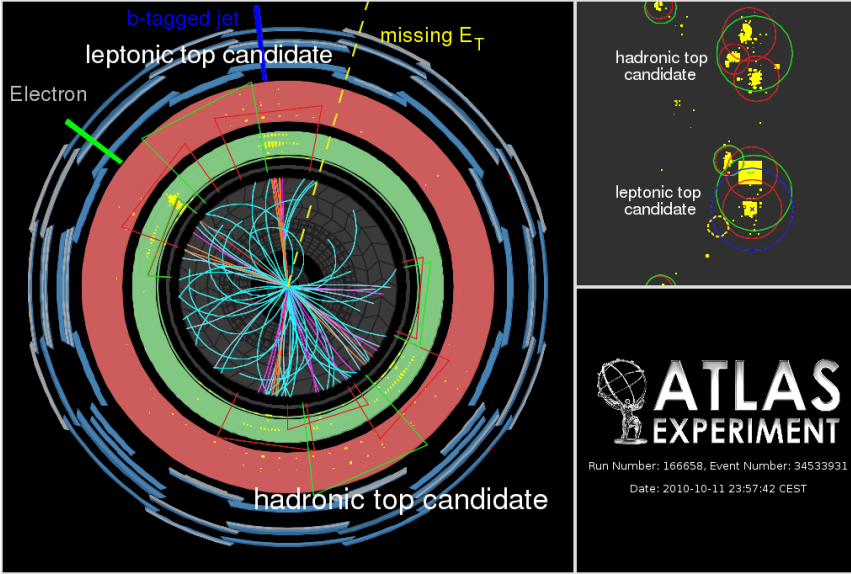
Figure A.2: Calibrated jet mass and $1 \rightarrow 2$ and $2 \rightarrow 3$ splitting scales for jets with $p_T > 100$ GeV. The distributions include the detector resolution, but no event selection has been applied. The three curves correspond to QCD di-jet production (J3-J7, blue dotted line), SM $t\bar{t}$ production with $m_{t\bar{t}} < 700$ GeV (black dashed line) and $m_{t\bar{t}} > 700$ GeV (red continuous line). The measured mass and splitting scales of the two top *monojet* candidates are indicated with continuous arrows.

Table A.4: Summary table for event 106929590 of run 167576. Brief description: The leptonic top candidate is formed by the isolated muon at 2 o'clock and the E_T^{miss} separated by $\Delta\phi = 0.27$ rad and the b-tagged jet at 1 o'clock. A second muon reconstructed inside this jet corroborates the heavy flavour hypothesis. In the opposite ϕ hemisphere, between 7 and 8 o'clock, three jets are reconstructed with $R = 0.4$. The picture does change with $R = 1.0$, where the three jets merge to form a single jet with a mass of 225 GeV. Legend: jets indicated in red correspond to $R = 0.4$, jets in green to $R = 1.0$.



Leptonic top	E_T^{miss} : $E_T = 159$ GeV, $\phi = 0.4$ muon: $p_T = 114$ GeV, $\eta = 0.21$, $\phi = 0.66$ $\Delta R_{lj} = 0.83$, $x_l = 0.87$, $y_l = 70.5$, $z_l = 0.48$ jet 3, $E_T = 90$ GeV, $\eta = -0.5$, $\phi = 1.1$, $m_j = 11.3$ GeV
Hadronic top ($R=0.4$ jets)	jet 1, $E_T = 205$ GeV, $\eta = -0.8$, $\phi = -2.2$ rad, $m_j = 18.3$ GeV + jet 2, $E_T = 115$ GeV, $\eta = -0.2$, $\phi = -2.8$ rad, $m_j = 9.7$ GeV + jet 4, $E_T = 49$ GeV, $\eta = -1.3$, $\phi = -2.7$ rad, $m_j = 10.6$ GeV $m_{jet1+jet3+jet4} = 176.5$ GeV
Hadronic top ($R=1.0$ jet)	jet 1, $E_T = 418.2$ GeV, $\eta = -0.8$, $\phi = -2.4$ rad, $m_j = 224.8$ GeV $\sqrt{d_{12}} = 105$ GeV, $\sqrt{d_{23}} = 44$ GeV

Table A.5: Summary table for event 34533931 of run 166658. Brief description: leptonic top candidate formed by high p_T electron (145 GeV, 11 o'clock), moderate E_T^{miss} (1 o'clock), and the b-tagged jet at 12 o'clock. When reclustered with $R = 1.0$ it acquires a large p_T , mass and $1 \rightarrow 2$ splitting scale 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 with $m_j = 197$ GeV, $\sqrt{d_{12}} = 110$, $\sqrt{d_{23}} = 40$. Legend: jets indicated in red correspond to $R = 0.4$, jets in green to $R = 1.0$.



Leptonic top	E_T^{miss} : $E_T = 36$ GeV, $\phi = -1.5$ electron: $p_T = 145$ GeV, $\eta = 1.1$, $\phi = 2.5$ $\Delta R_{lj} = 0.5$, $x_l = 0.85$ ($X' = 27$), $y_l = 76$, $z_l = 0.29$ jet 1, $E_T = 194$ GeV, $\eta = 1.2$, $\phi = 1.7$, $m_j = 16.6$ GeV
Hadronic top ($R=0.4$ jets)	jet 2, $E_T = 155$ GeV, $\eta = 1.1$, $\phi = -0.7$ rad, $m_j = 22.7$ GeV + jet 3, $E_T = 113$ GeV, $\eta = 1.3$, $\phi = -1.7$ rad, $m_j = 14.0$ GeV + jet 4, $E_T = 54$ GeV, $\eta = 0.6$, $\phi = -1.7$ rad, $m_j = 8.1$ GeV $m_{jet2+jet3+jet4} = 179.1$ GeV
Hadronic top ($R=1.0$ jets)	jet 1, $E_T = 355.5$ GeV, $\eta = 1.3$, $\phi = -1.1$ rad, $m_j = 197.1$ GeV $\sqrt{d_{12}} = 110$ GeV, $\sqrt{d_{23}} = 40$ GeV

Bibliography

- [1] E. Noether, *Invariante variationsprobleme*, *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* **1918** (1918) 235–257. [2](#), [138](#)
- [2] P. W. Higgs, *Broken symmetries and the masses of gauge bosons*, *Phys. Rev. Lett.* **13** (Oct, 1964) 508–509. [4](#), [139](#)
- [3] F. Englert and R. Brout, *Broken symmetry and the mass of gauge vector mesons*, *Phys. Rev. Lett.* **13** (Aug, 1964) 321–323. [4](#), [139](#)
- [4] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global conservation laws and massless particles*, *Phys. Rev. Lett.* **13** (Nov, 1964) 585–587. [4](#), [139](#)
- [5] D. H. Perkins, *Introduction to High Energy Physics*. Cambridge University Press, Apr., 2000. [6](#)
- [6] F. Boudjema and D. Zeppenfeld, *The standard model of particle physics: an introduction to the theory*, *Comptes Rendus Physique* **3** (2002), no. 9 1097 – 1106. [6](#)
- [7] M. K. Gaillard, P. D. Grannis, and F. J. Sciulli, *The standard model of particle physics*, *Rev. Mod. Phys.* **71** (Mar, 1999) S96–S111. [6](#)
- [8] N. Walet, *Nuclear and particle physics*, February, 2003. [7](#), [142](#)
- [9] S. Weinberg, *A Model of Leptons*, *Phys.Rev.Lett.* **19** (1967) 1264–1266. [8](#), [141](#)
- [10] The 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*, [[arXiv:1207.7214](#)]. [8](#)

- [11] The CMS Collaboration, S. Chatrchyan *et. al.*, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, [[arXiv:1207.7235](#)]. 8
- [12] R. Fernow, *Introduction to experimental particle physics*. Cambridge University Press, 1986. 9, 143
- [13] R. N. Cahn and G. Goldhaber, *The Experimental Foundations of Particle Physics*. Cambridge University Press, 2 ed., Aug., 2009. 9, 143
- [14] F. Biri. <http://www.imdb.com/name/nm0083767/>. 9, 144
- [15] T. Plehn, *New physics searches for the LHC*, [[hep-ph/1102.5178](#)]. 10
- [16] J. D. Lykken, *Beyond the standard model*, [[hep-ph/1005.1676](#)]. 10
- [17] D. E. Morrissey, T. Plehn, and T. M. P. Tait, *Physics searches at the LHC*, [[hep-ph/0912.3259](#)]. 10
- [18] G. Bertone, D. Hooper, and J. Silk, *Particle dark matter: evidence, candidates and constraints*, *Physics Reports* **405** (2005), no. 56 279 – 390. 10, 144
- [19] N. Jarosik *et. al.*, *Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Sky Maps, Systematic Errors, and Basic Results*, [[arXiv:1001.4744](#)]. 10, 144
- [20] P. Kroupa, M. Pawlowski, and M. Milgrom, *The failures of the standard model of cosmology require a new paradigm*, [[arXiv:1301.3907](#)]. 10, 145
- [21] J. M. Cline, *Baryogenesis*, 2006. 10, 145
- [22] A. D. Sakharov, *Violation of cp in variance, c asymmetry, and baryon asymmetry of the universe*, *Soviet Physics Uspekhi* **34** (1991), no. 5 392. 10, 145
- [23] J. H. Kuhn and G. Rodrigo, *Charge asymmetries of top quarks at hadron colliders revisited*, *JHEP* **1201** (2012) 063, [[arXiv:1109.6830](#)]. 10
- [24] The ATLAS Collaboration, G. Aad *et. al.*, *Measurement of the charge asymmetry in top quark pair production in $p\bar{p}$ collisions at $\sqrt{s} = 7\text{TeV}$ using the ATLAS detector*, *The European Physical Journal C - Particles and Fields* **72** (2012) 1–27. 10.1140/epjc/s10052-012-2039-5. 10

-
- [25] S. P. Martin, *A Supersymmetry Primer*, [[hep-ph/9709356](https://arxiv.org/abs/hep-ph/9709356)]. 10, 145
- [26] N. Arkani-Hamed, S. Dimopoulos, and G. Dvali, *The hierarchy problem and new dimensions at a millimeter*, *Physics Letters B* **429** (1998), no. 3-4 263 – 272. 11, 145
- [27] L. Randall and R. Sundrum, *A Large mass hierarchy from a small extra dimension*, *Phys.Rev.Lett.* **83** (1999) 3370–3373, [[hep-ph/9905221](https://arxiv.org/abs/hep-ph/9905221)]. 11
- [28] L. Randall and R. Sundrum, *An Alternative to compactification*, *Phys.Rev.Lett.* **83** (1999) 4690–4693, [[hep-th/9906064](https://arxiv.org/abs/hep-th/9906064)]. 11
- [29] H. Georgi and A. Pais, *CP violation as a quantum effect*, *Phys. Rev. D* **10** (Aug, 1974) 1246–1250. 11, 145
- [30] D. B. Kaplan, H. Georgi, and S. Dimopoulos, *Composite higgs scalars*, *Physics Letters B* **136** (1984), no. 3 187 – 190. 11, 145
- [31] B. Lillie, L. Randall, and L.-T. Wang, *The Bulk RS KK-gluon at the LHC*, *JHEP* **0709** (2007) 074, [[hep-ph/0701166](https://arxiv.org/abs/hep-ph/0701166)]. 13, 79, 100, 103, 162
- [32] The D0 Collaboration. <http://www-d0.fnal.gov>. 13, 14
- [33] European Organization for Nuclear Research, CERN. <http://home.web.cern.ch/about>. 15, 146
- [34] 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. 15, 16, 17, 146, 147
- [35] LHC Experiments. <http://home.web.cern.ch/about/experiments>. 17, 18, 148
- [36] The ATLAS Collaboration, G. Aad *et. al.*, *The ATLAS Experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008) S08003. 20, 24, 25, 148
- [37] The ATLAS Experiment. <http://atlas.ch>. 21, 22, 24, 26, 149
- [38] The ATLAS Collaboration, ed., *ATLAS detector and physics performance: Technical Design Report, 1*. Technical Design Report ATLAS. CERN, 1999. 22, 23, 25
- [39] Particle Data Group Collaboration, J. Beringer *et. al.*, *Review of Particle Physics (RPP)*, *Phys.Rev.* **D86** (2012) 010001. 24

- [40] The ATLAS Experiment - Public Results.
<https://twiki.cern.ch/twiki/bin/view/AtlasPublic>. 27
- [41] I. Foster and C. Kesselman, eds., *The grid: blueprint for a new computing infrastructure*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1999. 30
- [42] “WLCG.” <http://lcg.web.cern.ch/lcg>. 30
- [43] “The monarc project.” <http://MONARC.web.cern.ch/MONARC>. 30
- [44] R. W. L. Jones and D. Barberis, *The evolution of the atlas computing model*, *Journal of Physics: Conference Series* **219** (2010), no. 7 072037. 30
- [45] The ATLAS, *Atlas Computing: technical design report*. CERN, Geneva, 2005. 30, 34, 150
- [46] “The storage resource manager working group.”
<http://sdm.lbl.gov/srm-wg/index.html>. 32
- [47] I. Antcheva, M. Ballintijn, B. Bellenot, M. Biskup, R. Brun, N. Buncic, P. Canal, D. Casadei, O. Couet, and V. Fine, *ROOT — A C++ framework for petabyte data storage, statistical analysis and visualization*, *Computer Physics Communications* **180** (Dec., 2009) 2499–2512. 34
- [48] G. Barrand *et. al.*, *Gaudi : The software architecture and framework for building lhc data processing applications*, in *International Conference on Computing in High Energy and Nuclear Physics, CHEP 2000*, pp. 92–95, Feb., 2000. 34
- [49] B. Stroustrup, *The C++ Programming Language: Special Edition*. Addison-Wesley Professional, 3 ed., Feb., 2000. 34
- [50] “Python programming language.” <http://www.python.org>. 34
- [51] 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*, *Journal of Physics: Conference Series* **119** (2008), no. 6 062017. 35

-
- [52] A. D. Salvo, A. D. Silva, D. Benjamin, J. Blomer, P. Buncic, A. Harutyunyan, A. Undrus, and Y. Yao, *Software installation and condition data distribution via cernvm file system in atlas*, in *International Conference on Computing in High Energy and Nuclear Physics, CHEP 2012*, vol. 396, p. 032030, 2012. 35, 38
- [53] J. Moscicki *et. al.*, *Ganga: A tool for computational-task management and easy access to grid resources*, *Computer Physics Communications* **180** (2009), no. 11 2303 – 2316. 35
- [54] T. Maeno, *Panda: distributed production and distributed analysis system for atlas*, *Journal of Physics: Conference Series* **119** (2008), no. 6 062036. 35, 46
- [55] S. Albrand, T. Doherty, J. Fulachier, and F. Lambert, *The atlas metadata interface*, *Journal of Physics: Conference Series* **119** (2008), no. 7 072003. 35
- [56] M. Ballintijn, R. Brun, F. Rademakers, and G. Roland, *The proof distributed parallel analysis framework based on root*, 2003. 36, 45
- [57] D. C. van der Ster, J. Elmsheuser, M. Ú. García, and M. Paladin, *Hammercloud: A stress testing system for distributed analysis*, *Journal of Physics: Conference Series* **331** (2011), no. 7 072036. 36
- [58] S. G. de la Hoz, *The evolving role of tier2s in atlas with the new computing and data distribution model*, *Journal of Physics: Conference Series* **396** (2012) 032050. 36
- [59] A. J *et. al.*, *Automating atlas computing operations using the site status board*, in *International Conference on Computing in High Energy and Nuclear Physics, CHEP 2012*, vol. 396, p. 032072, 2012. 36
- [60] 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. 38, 46
- [61] “LHCONE.” <http://lhcone.web.cern.ch>. 38
- [62] D. Barberis, F. Bujor, J. de Stefano, A. L. Dewhurst, D. Dykstra, D. Front, E. Gallas, C. F. Gamboa, F. Luehring, and R. Walker,

- Evolution of grid-wide access to database resident information in atlas using frontier*, in *International Conference on Computing in High Energy and Nuclear Physics, CHEP 2012*, vol. 396, p. 052025, 2012. 38, 41
- [63] The ATLAS Collaboration, M. Villaplana *et al.* for the ATLAS Collaboration, *Iberian atlas cloud response during the first lhc collisions*, in *International Conference on Computing in High Energy and Nuclear Physics, CHEP 2012*, vol. 331, p. 072068, 2011. 39
- [64] S. González de la Hoz *et al.*, *Atlas spanish tier2 experiences during the step09 period, EGEE09 Conference*. (September, 2009). 39
- [65] V. Garonne *et al.*, *The atlas distributed data management project: Past and future*, in *International Conference on Computing in High Energy and Nuclear Physics, CHEP 2012*, vol. 396, p. 032045, 2012. 39
- [66] <http://www.quattor.org>. 41
- [67] <http://puppetlabs.com/solutions/configuration-management>. 41
- [68] J. Blomer, P. Buncic, and T. Fuhrmann, *Cernvm-fs: delivering scientific software to globally distributed computing resources*, in *Proceedings of the first international workshop on Network-aware data management, NDM '11*, (New York, NY, USA), pp. 49–56, ACM, 2011. 41
- [69] <http://www.squid-cache.org>. 41
- [70] “Nagios.” <http://www.nagios.org/>. 41
- [71] T. C. Group, “Cacti: The complete rrdtool-based graphing solution.” <http://www.cacti.net/>. 41, 48, 51
- [72] “Ganglia monitoring system.” <http://ganglia.sourceforge.net/>. 41, 48
- [73] www.clusterresources.com/products/maui. 44
- [74] <http://www.clusterresources.com/pages/products/torque-resource-manager.php>. 44
- [75] S. Gonzalez de la Hoz *et al.*, *Analysis facility infrastructure (Tier-3) for ATLAS experiment, Eur. Phys. J.* **C54** (2008) 691–697. 44, 152

-
- [76] 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. 44, 152
- [77] I. Cluster File System, “Lustre: A scalable, high performance file system.” <http://www.lustre.org/>. 44
- [78] P. Andreetto *et. al.*, *The glite workload management system*, *Journal of Physics: Conference Series* **119** (2008), no. 6 062007. 46
- [79] M. Ellert *et. al.*, *Advanced resource connector middleware for lightweight computational grids*, *Future Generation Computer Systems* **23** (2007), no. 2 219 – 240. 46
- [80] Fundacion Española para la Ciencia y la Tecnología, *Libro blanco de e-Ciencia en España*. Ministerio de Educacion y Ciencia, 2004. 47
- [81] 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. 47
- [82] 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. 47
- [83] G. Hanson *et. al.*, *Evidence for jet structure in hadron production by e^+e^- annihilation*, *Phys. Rev. Lett.* **35** (Dec, 1975) 1609–1612. 55, 155
- [84] J. D. Bjorken and S. J. Brodsky, *Statistical model for electron-positron annihilation into hadrons*, *Phys. Rev. D* **1** (Mar, 1970) 1416–1420. 55, 155
- [85] B. Wiik, *First Results from PETRA*, in *Int. Conf. on Neutrinos, Weak Interactions and Cosmology*, vol. 1, p. 113, June, 1979. 56

- [86] The CDF Collaboration, F. Abe *et. al.*, *Observation of top quark production in $\bar{p}p$ collisions*, *Phys.Rev.Lett.* **74** (1995) 2626–2631, [[hep-ex/9503002](#)]. 56, 156
- [87] S. Catani, Y. L. Dokshitzer, M. H. Seymour, and B. R. Webber, *Longitudinally invariant K_t clustering algorithms for hadron hadron collisions*, *Nucl. Phys.* **B406** (1993) 187–224. 58
- [88] S. D. Ellis and D. E. Soper, *Successive combination jet algorithm for hadron collisions*, *Phys.Rev.* **D48** (1993) 3160–3166, [[hep-ph/9305266](#)]. 58
- [89] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, *Better Jet Clustering Algorithms*, *JHEP* **08** (1997) 001, [[hep-ph/9707323](#)]. 58
- [90] M. Wobisch and T. Wengler, *Hadronization corrections to jet cross sections in deep- inelastic scattering*, *PITHA* **16** (1999) [[hep-ph/9907280](#)]. 58
- [91] M. Cacciari, G. P. Salam, and G. Soyez, *The anti- k_t jet clustering algorithm*, *JHEP* **04** (2008) 063, [[arXiv:0802.1189](#)]. 58, 66
- [92] 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](#)]. 58
- [93] G. P. Salam, *Towards Jetography*, *Eur. Phys. J.* **C67** (2010) 637–686, [[arXiv:0906.1833](#)]. 58, 59, 74, 158
- [94] M. Vos, *Boosting sensitivity to new physics*, *Cern Courier* (Sep, 2012). <http://cerncourier.com/cws/article/cern/50799>. 60
- [95] The ATLAS Collaboration, “Prospects for top anti-top resonance searches using early atlas data..” ATLAS refereed public note, ATL-PHYS-PUB-2010-008, Jul, 2010. 60, 63, 77, 92, 102, 107, 110, 112
- [96] The ATLAS Collaboration, G. Aad *et. al.*, *Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 7$ TeV with ATLAS*, *The European Physical Journal C* **72** (2012), no. 1 1–35. 61
- [97] M. H. Seymour, *Searches for new particles using cone and cluster jet algorithms: A Comparative study*, *Z. Phys.* **C62** (1994) 127–138. 62

-
- [98] J. M. Butterworth, B. E. Cox, and J. R. Forshaw, *WW scattering at the CERN LHC*, *Phys. Rev.* **D65** (2002) 096014, [[hep-ph/0201098](#)]. [63](#), [65](#), [66](#)
- [99] 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](#)]. [63](#), [74](#)
- [100] A. Abdesselam *et al.*, *Boosted objects: a probe of beyond the Standard Model physics*, *Eur. Phys. J.* **C71** (2011) 1661, [[arXiv:1012.5412](#)]. [64](#), [66](#), [72](#), [75](#), [158](#)
- [101] A. Altheimer *et al.*, *Jet substructure at the Tevatron and LHC: new results, new tools, new benchmarks*, *Journal of Physics G: Nuclear and Particle Physics* **39** (2012), no. 6 063001. [64](#), [75](#), [158](#)
- [102] A. Altheimer *et al.*, *Proceedings of the BOOST 2012 conference, to be submitted to Eur. Phys. J.*, . [64](#), [75](#), [104](#), [158](#)
- [103] The ATLAS, *Updated luminosity determination in pp collisions at $\sqrt{s} = 7$ tev using the atlas detector*, *ATLAS-CONF-2011-011* (2011). [66](#), [82](#), [96](#)
- [104] The CDF Collaboration, T. Aaltonen *et al.*, *Preliminary Results of a Search for Boosted Top Quarks by CDF II*, *CDF Note* **10234** (2010). [66](#)
- [105] The CDF Collaboration, T. Aaltonen *et al.*, *The Substructure of High Transverse Momentum Jets Observed by CDF II*, *CDF Note* **10199** (2010). [66](#)
- [106] ZEUS Collaboration, S. Chekanov *et al.*, *Subjet Distributions in Deep Inelastic Scattering at HERA*, *Eur. Phys. J* **63** (2009) 527. [66](#)
- [107] The CDF Collaboration, D. Acosta *et al.*, *Study of Jet Shapes in Inclusive Jet Production in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV*, *Phys. Rev. D* **71** (2005) 112002. [66](#)
- [108] ZEUS Collaboration, S. Chekanov *et al.*, *Substructure dependence of jet cross sections at HERA and determination of α_s* , *Nucl. Phys. B* **700** (2004) 3. [66](#)

- [109] ZEUS Collaboration, S. Chekanov *et. al.*, *Measurement of subjet multiplicities in neutral current deep inelastic scattering at HERA and determination of α_s* , *Phys. Lett. B* **558** (2003) 41. 66
- [110] The ATLAS Collaboration, G. Aad *et. al.*, *Study of jet shapes in inclusive jet production in pp collisions at $\sqrt{s} = 7$ tev using the atlas detector*, *Phys. Rev. D* **83** (2011) 052003. 66, 67
- [111] D. W. Miller, S. Dong, and A. Schwartzman, *Measurement of Hadronic Event Shapes and Jet Substructure in Proton-Proton Collisions at 7.0 TeV Center-of-Mass Energy with the ATLAS Detector at the Large Hadron Collider*. PhD thesis, Stanford U., Menlo Park, CA, 2011. Presented 03 Jun 2011. 66
- [112] The ATLAS Collaboration, G. Aad *et. al.*, *Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 7$ TeV, submitted to Eur. Phys. J. C* (2011) [[arXiv:1112.6426](#)]. 66, 86
- [113] W. Lampl *et. al.*, *Calorimeter clustering algorithms : Description and performance*, Tech. Rep. ATL-LARG-PUB-2008-002, CERN, 2008. 67
- [114] T. Sjostrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, *JHEP* **05** (2006) 026, [[hep-ph/0603175](#)]. for older versions, see <http://home.thep.lu.se/~torbjorn/Pythia.html>. 67, 78
- [115] M. Bahr *et. al.*, *Herwig++ Physics and Manual*, *Eur. Phys. J.* **C58** (2008) 639–707, [[arXiv:0803.0883](#)]. 67
- [116] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau, and A. D. Polosa, *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, *JHEP* **0307** (2003) 001, [[hep-ph/0206293](#)]. 67, 83, 108
- [117] T. Gleisberg *et. al.*, *Event generation with SHERPA 1.1*, *JHEP* **02** (2009) 007, [[arXiv:0811.4622](#)]. 67
- [118] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand, *Parton Fragmentation and String Dynamics*, *Phys. Rept.* **97** (1983) 31–145. 67
- [119] B. R. Webber, *A QCD Model for Jet Fragmentation Including Soft Gluon Interference*, *Nucl. Phys.* **B238** (1984) 492. 67
- [120] G. Corcella *et. al.*, *HERWIG 6.5 release note*, [[hep-ph/0210213](#)]. 67, 82

-
- [121] G. Corcella *et. al.*, *Herwig 6: An event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)*, *JHEP* **01** (2001) 010, [[hep-ph/0011363](#)]. 67
- [122] The ATLAS Collaboration, G. Aad *et. al.*, *Charged-particle multiplicities in pp interactions at $\sqrt{s} = 900$ GeV measured with the ATLAS detector at the LHC*, *Phys.Lett.* **B688** (2010) 21–42, [[arXiv:1003.3124](#)]. 67
- [123] P. Z. Skands, *Tuning Monte Carlo Generators: The Perugia Tunes*, [[arXiv:1005.3457](#)]. 67
- [124] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, *Parton distributions for the LHC*, *Eur. Phys. J.* **C63** (2009) 189–285, [[arXiv:0901.0002](#)]. 67
- [125] A. Sherstnev and R. S. Thorne, *Parton Distributions for LO Generators*, *Eur. Phys. J.* **C55** (2008) 553–575, [[arXiv:0711.2473](#)]. 67
- [126] The ATLAS Collaboration, G. Aad *et. al.*, *The ATLAS Simulation Infrastructure*, *Eur. Phys. J.* **C70** (2010) 823–874, [[arXiv:1005.4568](#)]. 67
- [127] GEANT4 Collaboration, S. Agostinelli *et. al.*, *GEANT4: A simulation toolkit*, *Nucl. Instrum. Meth.* **A506** (2003) 250–303. 67, 82
- [128] A. Ribon *et. al.*, *Status of Geant4 hadronic physics for the simulation of LHC experiments at the start of LHC physics program*, CERN-LCGAPP 2010-02, CERN, 2010. 67
- [129] The ATLAS Collaboration, G. Aad *et. al.*, *Properties of jets and inputs to jet reconstruction and calibration with the atlas detector using proton-proton collisions at $\sqrt{s} = 7$ tev*, Tech. Rep. ATLAS-CONF-2010-053, CERN, Geneva, Jul, 2010. 68
- [130] The ATLAS Collaboration, G. Aad *et. al.*, *Jet energy resolution and reconstruction efficiencies from in-situ techniques with the atlas detector using proton-proton collisions at a center of mass energy $\sqrt{s} = 7$ tev*, Tech. Rep. ATLAS-CONF-2010-054, CERN, Geneva, 2010. 70
- [131] The ATLAS Collaboration, “In-situ jet energy scale and jet shape corrections for multiple interactions in the first atlas data at the $\sqrt{s} = 7$ TeV.” ATLAS-CONF-2011-030, 2011. 70

- [132] The 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](#)]. 72, 74, 98, 102, 112, 159
- [133] The CDF Collaboration, T. Aaltonen *et. al.*, *Search for resonant $t\bar{t}$ production in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ -TeV*, *Phys. Rev. Lett.* **100** (2008) 231801, [[arXiv:0709.0705](#)]. 77
- [134] The CDF Collaboration, T. Aaltonen *et. al.*, *Limits on the production of narrow $t\bar{t}$ resonances in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV*, *Phys. Rev.* **D77** (2008) 051102, [[arXiv:0710.5335](#)]. 77
- [135] The CDF Collaboration, T. Aaltonen *et. al.*, *Search for new color-octet vector particle decaying to $t\bar{t}$ in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ tev*, *Submitted to Phys. Lett. B.* (2009) [[0911.3112](#)]. 77
- [136] The CDF Collaboration, T. Aaltonen *et. al.*, *A search for resonant production of $t\bar{t}$ pairs in 4.8 fb^{-1} of integrated luminosity of $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV*, *Phys.Rev.* **D84** (2011) 072004, [[arXiv:1107.5063](#)]. 77
- [137] The CDF Collaboration, T. Aaltonen *et. al.*, *Search for resonant production of $t\bar{t}$ decaying to jets in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV*, *Phys.Rev.* **D84** (2011) 072003, [[arXiv:1108.4755](#)]. 77
- [138] The D0 Collaboration, V. Abazov *et. al.*, *Search for $t\bar{t}$ resonances in the lepton plus jets final state in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV*, *Phys. Lett.* **B668** (2008) 98–104, [[arXiv:0804.3664](#)]. 77
- [139] The D0 Collaboration, V. M. Abazov *et. al.*, *Search for a Narrow $t\bar{t}$ Resonance in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV*, [[arXiv:1111.1271](#)]. Long author list - awaiting processing. 77
- [140] The ATLAS Collaboration, G. Aad *et. al.*, *Reconstruction of high mass $t\bar{t}$ resonances in the lepton+jets channel*, Tech. Rep. ATL-PHYS-PUB-2009-081. ATL-COM-PHYS-2009-255, CERN, Geneva, May, 2009. 77, 112
- [141] K. Joshi, A. D. Pilkington, and M. Spannowsky, *The dependency of boosted tagging algorithms on the event colour structure*, *Phys.Rev.* **D86** (2012) 114016, [[arXiv:1207.6066](#)]. 77

-
- [142] C. T. Hill, *Topcolor assisted technicolor*, *Phys.Lett.* **B345** (1995) 483–489, [[hep-ph/9411426](#)]. 78, 103, 162
- [143] Langacker, Paul, *The physics of heavy Z' gauge bosons*, *Rev. Mod. Phys.* **81** (Aug, 2009) 1199–1228. 78
- [144] R. M. Harris, C. T. Hill, and S. J. Parke, *Cross section for topcolor $Z'(t)$ decaying to t anti- t* , [[hep-ph/9911288](#)]. 78, 100
- [145] U. Baur and L. H. Orr, *Searching for $t\bar{t}$ Resonances at the Large Hadron Collider*, *Phys. Rev.* **D77** (2008) 114001, [[arXiv:0803.1160](#)]. 79
- [146] K. Agashe, R. Contino, L. Da Rold, and A. Pomarol, *A custodial symmetry for $Z b$ anti- b* , *Phys. Lett.* **B641** (2006) 62–66, [[hep-ph/0605341](#)]. 79
- [147] G. Cacciapaglia, C. Csaki, G. Marandella, and J. Terning, *A New Custodian for a Realistic Higgsless Model*, *Phys. Rev.* **D75** (2007) 015003, [[hep-ph/0607146](#)]. 79
- [148] B. Lillie, J. Shu, and T. M. P. Tait, *Kaluza-Klein Gluons as a Diagnostic of Warped Models*, *Phys. Rev.* **D76** (2007) 115016, [[arXiv:0706.3960](#)]. 79
- [149] H. Davoudiasl, S. Gopalakrishna, E. Ponton, and J. Santiago, *Warped 5-Dimensional Models: Phenomenological Status and Experimental Prospects*, [[arXiv:0908.1968](#)]. 79
- [150] F. Maltoni and T. Stelzer, *MadEvent: Automatic event generation with MadGraph*, *JHEP* **02** (2003) 027, [[hep-ph/0208156](#)]. 79
- [151] 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](#)]. 79
- [152] R. S. Chivukula, A. Farzinnia, R. Foadi, and E. H. Simmons, *Production of Massive Color-Octet Vector Bosons at Next- to-Leading Order*, [[arXiv:1111.7261](#)]. 81
- [153] New Physics Working Group Collaboration, G. Brooijmans *et. al.*, *New Physics at the LHC. A Les Houches Report: Physics at TeV Colliders 2009 - New Physics Working Group*, [[arXiv:1005.1229](#)]. 81

- [154] S. Frixione and B. R. Webber, *Matching NLO QCD computations and parton shower simulations*, *JHEP* **0206** (2002) 029, [[hep-ph/0204244](#)]. [82](#), [108](#)
- [155] S. Frixione, P. Nason, and B. R. Webber, *Matching nlo qcd and parton showers in heavy flavour production*, *Journal of High Energy Physics* (2003), no. 08 007. [82](#), [108](#)
- [156] 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](#)]. [82](#)
- [157] M. Aliev, H. Lacker, U. Langenfeld, S. Moch, P. Uwer, *et. al.*, *HATHOR: HAdronic Top and Heavy quarks crOss section calculatoR*, *Comput.Phys.Commun.* **182** (2011) 1034–1046, [[arXiv:1007.1327](#)]. [82](#)
- [158] Allwood-Spiers *et. al.*, *Monte carlo samples used for top physics*, Tech. Rep. ATL-PHYS-INT-2010-132, CERN, Geneva, Dec, 2010. [82](#)
- [159] S. Hoeche *et. al.*, *Matching parton showers and matrix elements*, [[hep-ph/0602031](#)]. [83](#)
- [160] J. Butterworth, E. Dobson, U. Klein, B. Mellado Garcia, T. Nunnemann, J. Qian, D. Rebuzzi, and R. Tanaka, *Single boson and diboson production cross sections in pp collisions at sqrts=7 tev*, Tech. Rep. ATL-COM-PHYS-2010-695, CERN, Geneva, Aug, 2010. [83](#)
- [161] C. Anastasiou, L. J. Dixon, K. Melnikov, and F. Petriello, *High precision QCD at hadron colliders: Electroweak gauge boson rapidity distributions at NNLO*, *Phys. Rev.* **D69** (2004) 094008, [[hep-ph/0312266](#)]. [83](#)
- [162] J. M. Campbell and R. Ellis, *An Update on vector boson pair production at hadron colliders*, *Phys.Rev.* **D60** (1999) 113006, [[hep-ph/9905386](#)]. [83](#)
- [163] J. M. Campbell, R. Ellis, and C. Williams, *Vector boson pair production at the LHC*, *JHEP* **1107** (2011) 018, [[arXiv:1105.0020](#)]. [83](#)
- [164] The ATLAS Collaboration, G. Aad *et. al.*, *Electron performance measurements with the ATLAS detector using the 2010 LHC proton-proton collision data*, *Eur. Phys. J.* **C 72** (2012) 1909, [[arXiv:1110.3174](#)]. [84](#)

-
- [165] The ATLAS Collaboration, G. Aad *et. al.*, *A search for $t\bar{t}$ resonances with the ATLAS detector in 2.05 fb^{-1} of proton-proton collisions at $\sqrt{s} = 7\text{ TeV}$* , *The European Physical Journal C* **72** (2012), no. 7 1–23. [87](#), [88](#), [103](#), [162](#)
- [166] The ATLAS 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 $\sqrt{s} = 7\text{ TeV}$ with the ATLAS detector*, *JHEP* **1209** (2012) 041, [[arXiv:1207.2409](#)]. [88](#), [101](#), [102](#), [159](#)
- [167] The ATLAS Collaboration, G. Aad *et. al.*, *Measurement of the top quark pair production cross-section with ATLAS in the single lepton channel*, *Phys. Lett. B* **711** (2012) 244–263. [89](#), [97](#)
- [168] The ATLAS Collaboration, G. Aad *et. al.*, *Measurement of the charge asymmetry in top quark pair production in pp collisions at $\sqrt{s} = 7\text{ TeV}$ using the ATLAS detector*, [[arXiv:1203.4211](#)]. Submitted to Eur. Phys. J. C. [91](#)
- [169] The ATLAS Collaboration, G. Aad *et. al.*, *A search for $t\bar{t}$ resonances in the lepton plus jets final state using 4.66 fb^{-1} of pp collisions at $\sqrt{s} = 7\text{ TeV}$* , Tech. Rep. ATLAS-CONF-2012-136, CERN, Geneva, Sep, 2012. [92](#), [98](#)
- [170] M. Botje, J. Butterworth, A. Cooper-Sarkar, A. de Roeck, J. Feltesse, *et. al.*, *The PDF4LHC Working Group Interim Recommendations*, [[arXiv:1101.0538](#)]. [96](#)
- [171] G. Choudalakis, *On hypothesis testing, trials factor, hypertests and the BumpHunter*, [[arXiv:1101.0390](#)]. [99](#)
- [172] The D0 Collaboration, I. Bertram *et. al.*, “A recipe for the construction of confidence limits.” FERMILAB-TM-2104, Fermilab, 2000. [99](#)
- [173] D. Casadei, *Reference analysis of the signal + background model in counting experiments*, *JINST* **7** (2012), no. 01 P01012. [100](#)
- [174] T. Junk, *Confidence level computation for combining searches with small statistics*, *Nucl. Instrum. Meth. A* **434** (1999) 435–443. [100](#)
- [175] A. L. Read, *Presentation of search results: The $CL(s)$ technique*, *J.Phys. G* **28** (2002) 2693–2704. [100](#)

- [176] The CMS Collaboration, S. Chatrchyan *et. al.*, *Search for resonant $t\bar{t}$ production in lepton+jets events in pp collisions at $\sqrt{s}=7$ tev*, *J. High Energy Phys.* **12** (Sep, 2012) 015. 37 p. [103](#), [162](#)
- [177] The ATLAS Collaboration, G. Aad *et. al.*, *Measurement of the top quark-pair production cross section with ATLAS in pp collisions at $\sqrt{s} = 7\text{TeV}$* , *Eur. Phys. J.* **C71** (2011) 1577, [[arXiv:1012.1792](#)]. [107](#), [108](#)
- [178] The ATLAS Collaboration, M. Barisonzi *et. al.*, *A search for new high-mass phenomena producing top quarks with the atlas experiment*, Tech. Rep. ATL-CONF-2011-070, CERN, Geneva, 2011. [107](#), [109](#), [110](#)
- [179] S. Frixione, E. Laenen, P. Motylinski, and B. R. Webber, *Single-top production in mc@nlo*, *Journal of High Energy Physics* (2006), no. 03 092. [108](#)
- [180] P. M. Nadolsky, H.-L. Lai, Q.-H. Cao, J. Huston, J. Pumplin, *et. al.*, *Implications of CTEQ global analysis for collider observables*, *Phys.Rev.* **D78** (2008) 013004, [[arXiv:0802.0007](#)]. [108](#)
- [181] S. Frixione, E. Laenen, P. Motylinski, C. White, and B. R. Webber, *Single-top hadroproduction in association with a w boson*, *Journal of High Energy Physics* (2008), no. 07 029. [108](#)
- [182] J. Pumplin *et. al.*, *New generation of parton distributions with uncertainties from global qcd analysis*, *JHEP* **07** (2002) 012, [[hep-ph/0201195](#)]. [108](#)
- [183] The ATLAS Collaboration, G. Aad *et. al.*, *Calibrating the b-tag and mistag efficiencies of the sv0 b-tagging algorithm in 3 pb^{-1} of data with the atlas detector*, Tech. Rep. ATLAS-CONF-2010-099, CERN, Geneva, Dec, 2010. [109](#)
- [184] M. Villaplana for the ATLAS Collaboration, *Boosted tops at atlas*, in *BOOST2011*, (Princeton, USA), ATLAS refereed conference note, ATL-PHYS-SLIDE-2011-267, May, 2011. [114](#)

Acknowledgements

The results shown in this thesis are the fruit of the collaboration of several people from different institutions within the ATLAS Collaboration.

The part of this thesis related to Grid computing is described in chapter 3. In section 3.2 the activities of the Tier-2 sites of Spain and Portugal are reported. IFIC has played a leading role in the coordination of these activities. Other contributions come from IFAE and PIC in Barcelona, from UAM in Madrid, and from the Portuguese sites, often represented by Helmut Wolters. Sections 3.3 and 3.4 describe the Tier-3 infrastructure deployed at IFIC. This part of my work was done in close collaboration with the system administrators Javier Sánchez and Álvaro Fernández.

I would like to thank Dario Barberis, Andreu Pacheco, José del Peso, José Salt and Javier Sánchez for their comments, suggestions and useful discussions regarding the computing part of this thesis.

The main contribution to the work presented in chapter 4 belongs to the institutes IFIC, University College London (UCL) and SLAC National Accelerator Laboratory. Other contributions come from Deutsches Elektronen-Synchrotron (DESY), McGill University and University of Oxford.

IFIC, DESY and University of Oxford provided the main contributions to the results shown in chapter 5. Other contributions came from Arizona University, Chicago University, Columbia University, Duke University, Glasgow University, McGill University, NIKHEF and Wuppertal University.

I would like to thank Jon Butterworth, María José Costa, Cigdem Issever, Tilman Plehn, Gavin Salam, Ariel Schwartzman and Michael Spannowsky for their comments, suggestions and useful discussions regarding the physics described in this thesis.

Last but not least, I would like to thank Santiago González de la Hoz and Marcel Vos for their patience and support.

Resumen

Belleza, simetría y física de partículas

Belleza no es una palabra que la gente suele usar para referirse a la física. Esta ciencia básica es comúnmente representada por pizarras llenas de ecuaciones, símbolos y extraños diagramas. Los físicos, por otro lado, nos defendemos diciendo que, una vez se entienden, todas estas ecuaciones, estos símbolos y diagramas son también hermosos. La verdad es que las dos posturas no son, para nada, irreconciliables ya que conceptos como belleza o elegancia están, en ambos casos, comúnmente relacionados con otros como simplicidad o simetría.

En palabras de M. Gell-Mann ‘una teoría es bella o elegante cuando puede expresarse de forma concisa en términos matemáticos bien entendidos’. Un buen ejemplo de esta idea son las ecuaciones de Maxwell para describir el electromagnetismo. Hace falta una página entera para escribir estas ecuaciones que representan la unificación de la electricidad y el magnetismo.

$$\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} = 4\pi\rho \quad (\text{R.1})$$

$$\frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z} = 0 \quad (\text{R.2})$$

$$\begin{aligned} \frac{\partial E_x}{\partial y} - \frac{\partial E_y}{\partial x} + \frac{1}{c} \frac{\partial B_z}{\partial t} &= 0 \\ \frac{\partial E_y}{\partial z} - \frac{\partial E_z}{\partial y} + \frac{1}{c} \frac{\partial B_x}{\partial t} &= 0 \\ \frac{\partial E_z}{\partial x} - \frac{\partial E_x}{\partial z} + \frac{1}{c} \frac{\partial B_y}{\partial t} &= 0 \end{aligned} \quad (\text{R.3})$$

$$\begin{aligned}
\frac{\partial B_x}{\partial y} - \frac{\partial B_y}{\partial x} - \frac{1}{c} \frac{\partial E_z}{\partial t} &= \frac{4\pi}{c} \\
\frac{\partial B_y}{\partial z} - \frac{\partial B_z}{\partial y} - \frac{1}{c} \frac{\partial E_x}{\partial t} &= \frac{4\pi}{c} \\
\frac{\partial B_z}{\partial x} - \frac{\partial B_x}{\partial z} - \frac{1}{c} \frac{\partial E_y}{\partial t} &= \frac{4\pi}{c}
\end{aligned} \tag{R.4}$$

Sin embargo, usando cálculo vectorial pueden escribirse de forma más compacta así,

$$\begin{aligned}
\nabla \cdot \vec{E} &= 4\pi\rho \\
\nabla \cdot \vec{B} &= 0 \\
\nabla \times \vec{E} &= -\frac{1}{c} \frac{\partial \vec{B}}{\partial t} \\
\nabla \times \vec{B} &= \frac{4\pi}{c} \vec{J} + \frac{1}{c} \frac{\partial \vec{E}}{\partial t}
\end{aligned} \tag{R.5}$$

Sabemos que estas ecuaciones son simétricas bajo rotaciones en el espacio. Si hacemos girar el espacio entero un ángulo cualquiera, fenómenos relacionados con electricidad o magnetismo seguirán sucediendo de la misma forma. En su teoría de la relatividad especial, A. Einstein observó un conjunto nuevo de simetrías para el electromagnetismo llamadas transformaciones de Lorentz. Usando estas simetrías y cálculo tensorial las ecuaciones de Maxwell se pueden escribir de forma más sencilla y, por tanto, más bella y elegante.

$$\partial_\nu F^{\mu\nu} = \frac{4\pi}{c} J^\mu \tag{R.6}$$

$$\epsilon^{\alpha\beta\mu\nu} \partial_\beta F_{\mu\nu} = 0 \tag{R.7}$$

El papel de las simetrías en física va más allá del de meras herramientas elegantes. El teorema de Noether [1] demuestra que las simetrías están intrínsecamente ligadas a leyes de conservación. De hecho, uno de los resultados más importantes del siglo pasado en física de partículas es la descripción de las interacciones fuerte, débil y electromagnética en términos de las llamadas teorías de campos *gauge*. Las leyes que gobiernan estas teorías son simétricas bajo las llamadas transformaciones *gauge*.

Uno de los ingredientes en la formulación de las ecuaciones R.6 y R.7 es, $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$, el tensor que representa al campo electromagnético. El

electromagnetismo es una fuerza de largo alcance con un mensajero sin masa, el fotón. El potencial vector, A , que aparece en la definición de $F^{\mu\nu}$, es el campo cuántico que crea y destruye fotones. Tanto $F^{\mu\nu}$ como las ecuaciones R.6 y R.7 son simétricos bajo transformaciones *gauge*. Todas las propiedades del electromagnetismo así como la derivación de las ecuaciones de Maxwell se pueden escribir en la forma de un simple Lagrangiano que describe fotones libres así,

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} \quad (\text{R.8})$$

La ecuación R.8 es un ejemplo perfecto de lo poderoso que puede ser en física el concepto de simetría. Una simetría como la *gauge*, combinada con el uso de matemáticas avanzadas como el cálculo tensorial, ha dado lugar, no sólo a una representación más elegante de las leyes que gobiernan el electromagnetismo, sino a una comprensión más profunda de la teoría en sí.

El Modelo Estándar de física de partículas

El poder de la simetría *gauge* se muestra excelentemente en el Modelo Estándar, en inglés *Standard Model* (SM). La tres interacciones fundamentales, fuerte, débil y electromagnética, se pueden unificar bajo los grupos *gauge* $SU(2)_L$, $U(1)_Y$ y $SU(3)_c$. Toda la información sobre los fermiones y bosones que aparecen en la figura 1 y sus interacciones se encuentra codificada en las transformaciones *gauge* y en la definición de las derivadas covariantes como sigue,

$$\begin{aligned} \psi(x) \longrightarrow U(x)\psi(x) &= e^{i\theta_3^a(x)T^a} e^{i\theta_2^b(x)\frac{\tau^b}{2}} e^{i\theta_1(x)Y} \psi(x) \\ D_\mu &= \partial_\mu - ig_s T^a A_\mu^a - ig \frac{T^b}{2} W_\mu^b - ig' Y B_\mu \end{aligned} \quad (\text{R.9})$$

donde g_s , g y g' son las constantes de acoplamiento. El SM es capaz de describir un grandísimo número de procesos con una precisión sin precedentes. Sin embargo, la partículas predichas por este modelo no han de tener masa, ya que la introducción de términos de masa rompe la simetría. Una forma de generar masa consiste en romper la simetría sólo de forma local, manteniendo la simetría global del sistema. Es la llamada rotura espontánea de simetría, en inglés *spontaneous symmetry breaking* (SSB) [2, 3, 4], que introduce un potencial del tipo, $V = \lambda(|\phi|^2 - v/2)^2$, $\lambda > 0$, con un valor esperado en el vacío distinto de 0. En este tipo de potenciales el valor más estable, el de menor energía, no es simétrico. Así, la simetría sólo se rompe cuando el sistema baja al estado de

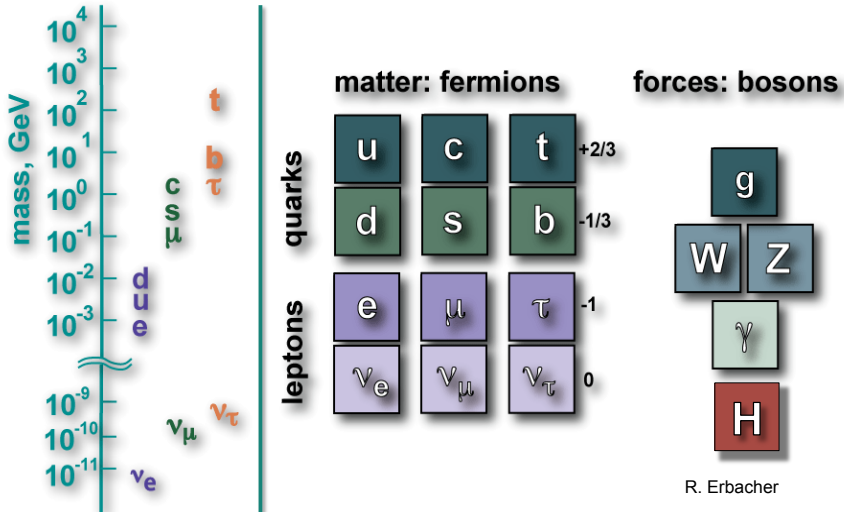


Figura 1: Jerarquía de masa y tabla periódica de las partículas fundamentales del Modelo Estándar.

mínima energía . La existencia del bosón de Higgs es una de las consecuencias de la aplicación de este método para dotar de masa a las partículas de la figura 1. Dicha masa es proporcional a los acoplamientos de cada partícula con el bosón de Higgs.

El SM es capaz de describir un grandísimo número de procesos con una precisión sin precedentes a partir de un pequeño número de parámetros. El conjunto más usado lo forman la constante de Fermi G_F , el acoplamiento electromagnético α_{em} y las masas de los bosones Z^0 y Higgs.

$$G_F = \frac{1}{v^2 \sqrt{2}}; \quad \alpha_{em} = \frac{g^2 g'^2}{4\pi(g^2 + g'^2)}$$

$$m_Z^2 = \frac{1}{4}(g^2 + g'^2)v^2; \quad m_H^2 = 2\lambda v^2 \quad (R.10)$$

Fundamentos experimentales del Modelo Estándar

El SM no nació de repente a partir de simetrías y elegancia. La construcción del SM ha sido una interacción constante entre teoría y experimentos. A menudo la teoría ha sido capaz de guiar a los científicos y ha predicho satisfactoriamente los resultados de los experimentos. El descubrimiento del bosón de Higgs es un buen ejemplo de esta categoría de ‘descubrimientos anunciados’. En otras ocasiones, sin embargo, descubrimientos inesperados han abierto nuevos caminos en física de partículas. El ejemplo clásico de esta segunda categoría es el descubrimiento del muón en 1937. Anderson y Neddermeyer estaban estudiando radiación cósmica en Caltech cuando vieron partículas que tenían la misma carga que el electrón pero su curvatura en un campo magnético indicaba una masa entre la del electrón y la del protón. La confusión que produjo el muón en la comunidad científica durante la década de los 1930 se representa perfectamente en la conocida exclamación de I. Rabi respecto a la inesperada partícula: ‘*Who ordered that?*’, ¿Quién ha pedido eso?’.

La tabla 1 resume los momentos más importantes del último siglo en física de partículas. La columna ‘*experiment*’ lista observaciones importantes en experimentos. La segunda columna muestra los mayores hitos en el progreso de la teoría.

La historia de la física de partículas empieza en la tabla 1 con el descubrimiento de la desintegración nuclear β por C.D. Ellis y otros en 1927. No se encontraría una explicación plausible hasta 1934. La idea de Fermi para explicar las interacciones débiles como la interacción de cuatro fermiones evolucionó durante las siguientes décadas. En 1957, Sudarshan y Marshak se dieron cuenta de que los resultados de las correlaciones angulares entre electrón-neutrino publicados por varios experimentos sobre desintegraciones β eran inconsistentes in la teoría V-A como solución. Una teoría como la de Fermi con la fuerza débil transmitida por una partícula cargada pesada podía describir acertadamente todos los resultados experimentales de que se disponía entonces sobre interacciones débiles pero era aún incompleta. Las secciones eficaces predichas para neutrinos violaban unitariedad para energías en centro de masas por encima de 300 GeV y la teoría no podía calcularse a orden mayor. Además, no se había podido encontrar un bosón W con una masa de hasta 20 GeV.

La explicación completa vino con la unificación del electromagnetismo y las interacciones débiles en 1967 [9]. La teoría de Glashow, Weinberg y Salam estaba basada en partículas de Yang-Mills sin masa como portadores de la interacción débil. Estas partículas adquirirían masa a través del mecanismo de rotura espontánea de simetría propuesto por Higgs y otros. En comparación con

Tabla 1: Particle physics time line [8].

Year	Experiment	Theory
1927	β decay discovered	
1928		Dirac: Wave equation for electron
1930		Pauli suggests existence of neutrino
1931	Positron discovered. Chadwick discovers neutron	Dirac realizes that positrons are part of his equation
1933/4		Fermi introduces theory for β decay
1933/4		Yukawa discusses nuclear binding in terms of pions
1937	μ discovered in cosmic rays	
1938	Baryon number conservation	
1946		μ is not Yukawa's particle
1947	π^+ discovered in cosmic rays	
1946-50		Tomonaga, Schwinger and Feynman develop QED
1948-52	First artificial π . K^+ discovered. $\pi^0 \rightarrow \gamma\gamma$. 'V-particles' Λ^0 and K^0 . Δ : excited state of nucleon	
1954		Yang and Mills: Gauge theories
1956		Lee and Yang: Weak force might break parity
1956	CS Wu and Ambles: Yes it does	
1961		Eightfold way as organizing principle
1962	ν_μ and ν_e	
1964		Quarks (Gell-man and Zweig) u, d, s.
1964		Fourth quark suggested
1965		Higgs boson prediction
1965		Colour charge: all particles are color neutral
1967		Glashow, Salam and Weinberg: Unification of electromagnetic and weak interactions.
1968	Homestake experiment: Solar neutrino problem.	
1968/9	DIS at SLAC: constituents of protons seen	
1974		Wess-Zumino: SUSY
1973	Neutral weak current seen at Gargamelle bubble chamber at CERN	QCD as the theory of colour interactions. Gluons
1973		Asymptotic freedom
1974	J/ψ ($c\bar{c}$) meson	
1976	D^0 meson ($\bar{u}c$) confirms theory	
1976	τ lepton	
1977	b quark	
1979	Gluon signature at PETRA	
1981		MSSM
1983	W^\pm and Z^0 seen at UA1	
1984		Composite Higgs scalars
1989	SLAC suggests only 3 generations of light neutrinos	
1995	Tevatron: t quark at 175 GeV mass	
1998	Super-Kamiokande confirms neutrino oscillations	ADD extra dimensions
1999		Randall-Sundrum models
2012	LHC: new neutral boson at 126 GeV mass. Higgs boson?	

la teoría de Fermi, donde sólo se intercambiaban bosones débiles cargados, esta teoría también predecía la existencia de procesos donde no se transfería carga alguna. La llamada corriente neutra débil fue descubierta en 1973 usando la cámara de burbujas Gargamelle en el CERN.

Las décadas siguientes traerían el descubrimiento de un gran número de partículas. Esta atmósfera tan caótica favoreció la aparición de ideas brillantes como el modelo quark de Gell-Mann y el desarrollo de QCD. La aparición de la extrañeza sugirió la existencia de una nueva familia de partículas que sería confirmada por los descubrimientos del mesón J/ψ , en SLAC y BNL en 1974, y el mesón D^0 en 1976. Más tarde ese mismo año, el descubrimiento del leptón τ sugirió la existencia de otra nueva familia de partículas. El quark b se descubriría en Fermilab un año después, en 1977.

Desde entonces los experimentos continuaron corroborando las predicciones teóricas. Para el descubrimiento de los bosones W^\pm y Z^0 fue necesario un gran esfuerzo. El *Super Proton Synchrotron* se había instalado en el CERN para acelerar protones y fue convertido después en un colisionador de protones y antiprotones. Los portadores de la fuerza débil fueron finalmente descubiertos en el experimento UA1 en 1983. Doce años después, en 1995, el último y más pesado de los quarks fue descubierto. El Tevatron, un colisionador de protones a $\sqrt{s} = 1,96 \text{ TeV}$ construido en Fermilab, fue el primero capaz de ver el quark top y medir su masa con precisión.

Ya sólo quedaba un bosón por descubrir y el bosón de Higgs tardaría 30 años en dejarse entrever. El descubrimiento de un bosón sin carga y con una masa de $\sim 125 \text{ GeV}$ fue anunciado el 4 de julio de 2012 por los experimentos del LHC. Este descubrimiento corrobora una predicción que ha esperado alrededor de 50 años a ser confirmada.

Aunque las partículas predichas por el SM se han descubierto y sus predicciones se han confirmado con precisión aun hay espacio para la sorpresa. Se puede encontrar exposiciones ms extensas sobre los fundamentos experimentales del Modelo Estándar en las referencias [12, 13].

Física más allá del Modelo Estándar

El Modelo Estándar ha demostrado ser capaz de describir un amplio abanico de procesos con una precisión sin precedentes. Aun así, los físicos seguimos buscando un modelo más completo de las partículas y sus interacciones.

Una de nuestras motivaciones es la búsqueda de una explicación para los valores que toman los parámetros de la teoría. En el sector fermiónico los

parámetros libres incluyen las masas de los $3n$ fermiones y los $(n-1)^2$ parámetros en la matriz CKM. Asumiendo que hay 3 generaciones de fermiones ($n=3$) y añadiendo la constante de acoplamiento fuerte, g_s , el número total de parámetros libres en el Modelo Estándar suma 18. Idealmente, nos gustaría que todos estos parámetros pudieran derivarse de principios básicos. Una ‘Teoría del Todo’ capaz de explicar todos los fenómenos físicos está para muchos demasiado lejos, una utopía imposible. La mejor respuesta a éstos la dio el cineasta argentino Fernando Biri [14].

“¿Para qué sirve la utopía? ¿Cual es la utilidad de una meta tan lejana que, no importa cuanto se ande, nunca será alcanzada? Precisamente para eso, utopía es útil porque nos hace andar”

Hay muchas preguntas sin respuesta. La materia oscura y las oscilaciones de neutrinos están firmemente establecidas experimentalmente, pero no tienen explicación dentro del Modelo Estándar. La constante cosmológica o gravedad cuántica son también retos a los que aún tenemos que responder. Aquí nos centraremos en aquellas cuestiones que pueden encontrar su respuesta en los datos del Gran Colisionador de Hadrones (LHC) del que hablaremos más adelante.

El primer reto a mencionar no puede ser otro que clarificar el mecanismo de rotura de la simetría electrodébil. Aunque la partícula recientemente descubierta es compatible con el bosón de Higgs aún puede haber alguna sorpresa. El nuevo bosón no tiene por qué ser exactamente el que predijeron Higgs *et al.* Puede no venir sólo sino ser el más ligero de un grupo de bosones parecidos al de Higgs, puede no ser fundamental sino estar compuesto por otras partículas, etc.

El problema de la jerarquía *gauge* se refiere a que la gravedad es extremadamente débil comparada con el resto de fuerzas. Otra forma de ver el problema tiene que ver con el hecho de que el bosón de Higgs es mucho más ligero que la masa de Planck ($M_{Planck} = 2,4 \times 10^{18} GeV$). Sabemos que las correcciones a la masa del bosón de Higgs son del orden de la escala a la que aparece nueva física y, por tanto, la masa del bosón de Higgs debería ser del mismo orden. Si asumimos que la nueva física es gravitación, entonces la escala de nueva física es la escala de Planck. Si queremos extrapolar el Modelo Estándar a esas energías y a la vez mantener ligero el bosón de Higgs, es necesaria una elección muy precisa de los parámetros libres del Modelo Estándar. Aunque dicha elección es consistente con la teoría, muchos prefieren una solución más elegante.

Si las partículas que componen la materia oscura [18] son la fuente de las anomalías observadas en el cosmos [19], es muy probable que estas partículas se

produzcan en el LHC y sean descubiertas en sus detectores. La otra explicación posible a dichas anomalías pasa por modificar las leyes de la gravitación y de la relatividad [20].

Explicar el dominio de la materia sobre la antimateria [21, 22] es otro reto que espera su respuesta. El único sitio en el Modelo Estándar en el que se da un trato diferente a materia y antimateria es la fase en la matriz CKM en la que se viola simetría CP. Desafortunadamente, esta fase no parece ser suficiente para explicar la simetría observada.

En vista de que el Modelo Estándar no es capaz de dar explicación a esta lista de retos, tanto experimentales como teóricos, ha habido un flujo constante de propuestas para extender el modelo y que predicen nueva física en la escala del TeV durante las últimas 6 décadas.

Una de los candidatos más conocidos es supersimetría (SUSY). Esta propuesta relaciona fermiones y bosones de forma que por cada bosón del Modelo Estándar hay un supercompañero fermiónico y vice versa [25]. Esta teoría es renormalizable, lo que quiere decir que sus predicciones son muy precisas. El problema de la jerarquía *gauge* se evita debido a la presencia de estas nuevas partículas supersimétricas cancela las correcciones a la masa del bosón de Higgs. SUSY puede explicar el mecanismo de rotura de la simetría electrodébil y ofrece candidatos viables a materia oscura y al origen de la diferencia entre materia y antimateria. Además, SUSY sugiere la unificación de todas las fuerzas *gauge* a altas energías. Todas estas cualidades han hecho que SUSY sea el centro de atención tanto de físicos teóricos como experimentales. Sin embargo, hasta la fecha no se ha encontrado ningún signo de su existencia.

Las extensiones del Modelo Estándar con dimensiones extra se introdujeron para explicar la diferencia entre gravedad y el resto de interacciones. En lugar de ajustar los parámetros del Modelo Estándar para mantener ligero al bosón de Higgs, lo que se hace es añadir nuevas dimensiones a la teoría en las que la gravitación es la más fuerte [26]. El hecho de que las nuevas dimensiones no se han visto implica que éstas deben ser compactas y de tamaño finito. Cuando se permite que una partícula del Modelo Estándar se propague por las dimensiones extra aparece una serie de copias pesadas (modos de Kaluza-Klein) de dicha partícula cuyas masas están relacionadas con el tamaño de la dimensión que podrían ser detectadas por los experimentos del LHC.

En modelos de Higgs compuesto el problema de jerarquía se soluciona asumiendo que el bosón de Higgs es una partícula compuesta. Su existencia se explica de forma similar a la de los piones ligeros, que aparecen como bosones (pseudo-)Goldstone al romperse espontáneamente la simetría chiral. En estos escenarios la nueva simetría mantiene ligero al bosón de Higgs [29, 30]. Aunque

estas propuestas han sido acotadas fuertemente por los experimentos, aún no se han podido descartar completamente.

CERN, LHC y ATLAS

Desde su fundación en 1954 a propuesta de los premios Nobel Louis de Broglie e Isidore Rabi, el *Conseil Européen pour la Recherche Nucléaire* (CERN) [33] es uno de los primeros ejemplos de la Europa unida que hoy conocemos. La historia de este laboratorio situado en Ginebra, en la frontera entre Francia y Suiza, ha estado dedicada a los aceleradores de partículas. El primer acelerador acogido por el que está considerado actualmente como el mayor laboratorio de física de partículas del mundo fue un *SynchroCyclotron* (SC) de 600 MeV en 1957. En 1971, tuvieron lugar las primeras colisiones de protones de la historia en los *Intersecting Storage Rings* (ISR). Años después, en 1983, las colisiones protón-antiprotón del *Super Proton Synchrotron* (*SppS*) permitieron el descubrimiento de los bosones W^\pm y Z^0 confirmando así la unificación del electromagnetismo y la fuerza débil. En los 90, el colisionador de leptones más potente jamás construido, el *Large Electron-Positron Collider* (LEP), permitió medir las propiedades de W y Z y otros procesos del Modelo Estándar con precisión.

Instalado en el túnel de 27 km donde anteriormente estuvo LEP, el Gran Colisionador de Hadrones, en inglés *Large Hadron Collider* (LHC) [34], es el acelerador de partículas más potente de la historia. Alrededor de 1 nanogramo de protones circulan cada día en el LHC. En su viaje desde el contenedor de hidrógeno hasta el LHC, los protones, no sólo viajan a través del complejo de aceleradores mostrado en la figura 2, sino que también viajan a través de la historia del propio laboratorio. Los protones extraídos ionizando hidrógeno se aceleran hasta los 50 MeV en el *Linac2*. Desde ahí, se dirigen al *Proton-Booster Synchrotron* (PSB), donde su energía aumenta hasta los 1.4 GeV. Después, los haces de protones alcanzan los 25 GeV en el PS, que tiene 54 años de edad, y suben hasta los 450 GeV en el SPS, de 37 años de edad, antes de ser inyectados en el LHC. Operando a toda potencia, el LHC será capaz de hacer chocar haces de protones de 7 TeV. Esto está programado para comienzos de 2015. Desde su puesta en marcha a finales de 2009, el acelerador ha trabajado con haces de 3.5 TeV hasta 2011, y con haces de 4 TeV en 2012. En este tiempo también se han colisionado iones pesados, típicamente de plomo.

Cuatro grandes experimentos aprovechan las colisiones producidas por el LHC (figura 3). Además, hay otros experimentos de menor tamaño instalados

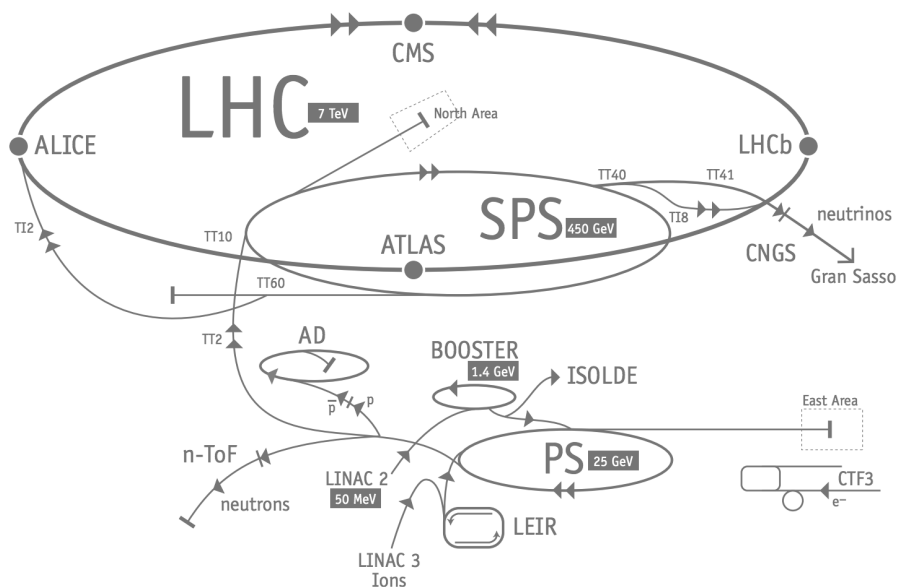


Figura 2: Esquema del complejo de aceleradores del CERN [34].

en el acelerador.

ATLAS y CMS son experimentos de propósito general. Ambos tienen un detector de trazas, calorímetros electromagnéticos y hadrónicos, y un espectrómetro de muones. La mayor diferencia está en el sistema de imanes. Mientras ATLAS tiene un solenoide de 2 T cubriendo el detector de trazas y toroides de 2-8 T/m en el espectrómetro de muones, CMS usa un único campo magnético no lineal de 4 T.

LHCb se ocupa de estudiar violación de CP y desintegraciones raras de hadrones con sabor como los B. Está diseñado para detectar partículas que emergen cerca de la dirección de haz. Con este propósito, incorpora un detector de vértices para ver las trazas de las partículas y un detector RICH para identificarlas.

ALICE se ocupa de las colisiones de iones pesados. Tiene un detector de trazas y una TPC en la parte central, y espectrómetros de muones a los lados. Con este diseño, ALICE estudia el comportamiento de la materia a densidades extremas.

El propósito del experimento TOTEM es medir la sección eficaz total de

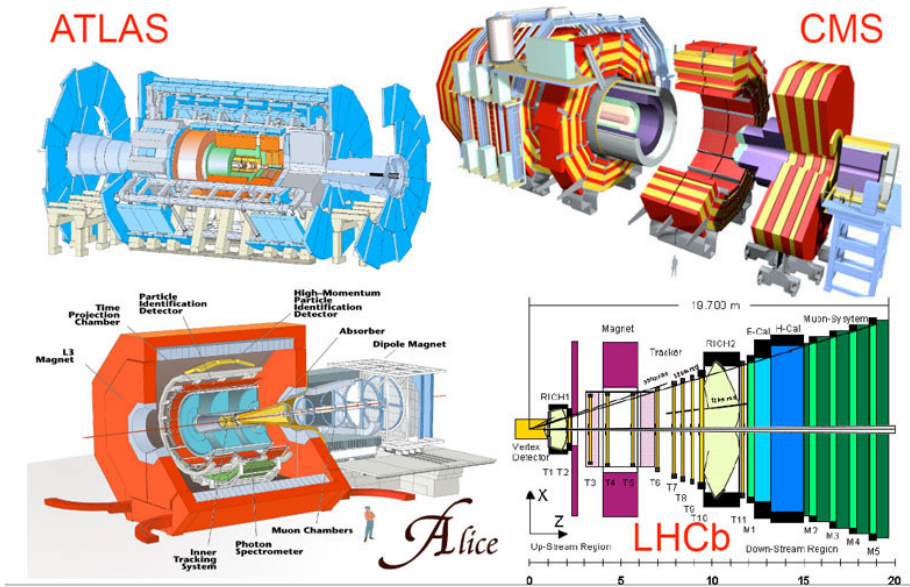


Figura 3: Detectores instalados en los cuatro puntos de interacción del LHC [35].

interacción protón-protón y estudiar dispersiones elástica y difractiva en el LHC. Está instalado cerca del punto de interacción de CMS y tiene sensores de silicio que miden los protones 200 m antes y después de dicho punto.

LHCf estudia la producción de partículas neutras en direcciones muy próximas al haz de protones o de iones pesados. Combinado con los resultados de TOTEM, estas medidas mejorarán nuestra comprensión del desarrollo de las cascadas atmosféricas que aparecen cuando rayos cósmicos de muy alta energía chocan contra la atmósfera.

Aprobado por el CERN en diciembre de 2009, el experimento MoEDAL consiste en unas capas de plástico centelleador que cubren al detector de vértices de LHCb. Su propósito es detectar trazas de partículas estables como pueden ser los monopolos magnéticos o partículas supersimétricas masivas.

De todos los experimentos que alberga el LHC, ATLAS [36] es el de mayor tamaño. El detector tiene forma cilíndrica, mide 42 metros de largo por 25 de alto y pesa unas 7000 toneladas. Está formado por varias capas concéntricas de detectores, como muestra la figura 4, que permiten detectar y diferenciar las

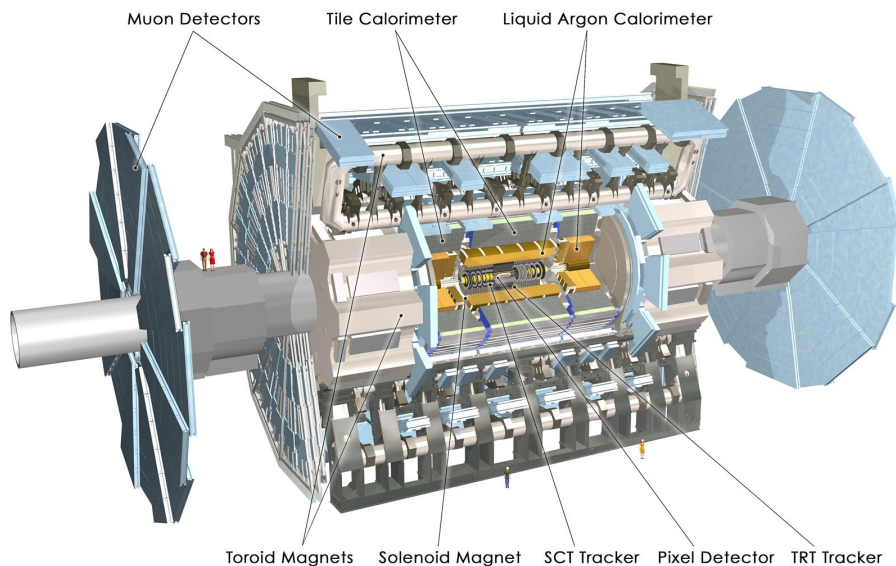


Figura 4: Experimento ATLAS [37].

partículas resultantes de las colisiones de protones.

Inmerso en un campo magnético solenoidal de 2 T, el detector interno de ATLAS mide el momento de las partículas cargadas. Está formado por tres subdetectores que, de dentro hacia fuera, son: el detector de píxeles, el detector de trazas semiconductor (SCT) y el detector de trazas de radiación de transición (TRT). La reconstrucción de las trazas se lleva a cabo combinando información de los tres subdetectores usando dos algoritmos. Un algoritmo que trabaja de dentro hacia fuera y reconstruye las trazas primarias. Otro algoritmo empieza en el TRT y continúa hacia dentro añadiendo información de los detectores de silicio. Este segundo algoritmo reconstruye la mayor parte de las trazas secundarias, que vienen de conversiones, interacciones hadrónicas o desintegraciones de V_0 . Los vértices primarios se reconstruyen usando un ajuste iterativo.

Las medidas de energía de electrones, positrones y fotones las proporciona un calorímetro electromagnético de muestreo con una estructura en forma de acordeón. Tiene plomo como material absorbente y argón líquido como medio

activo. Para la calorimetría hadrónica se usa un detector con una estructura de tejas intercaladas de plástico centelleador como medio activo y acero como absorbente.

La capa más externa de ATLAS la ocupa el espectrómetro de muones. Se usan tubos de deriva y un sistema de alineamiento óptico que es capaz de corregir las posibles deformaciones de la cámara. En las partes donde es necesaria una tecnología más resistente a la radiación se usan cámaras de tiras catódicas.

En ATLAS se usan dos sistemas de imanes superconductores. Rodeando al detector interno se ha instalado un solenoide diseñado para generar un campo magnético de 2 T que curve la trayectoria de las partículas cargadas. También hay un sistema de imanes que genera un campo magnético toroidal de entre 2 y 8 T/m situado dentro del espectrómetro de muones.

Tanto el LHC como ATLAS han funcionado por encima de las expectativas desde el comienzo de la toma de datos a finales de 2009. Desde entonces, el LHC ha producido 27 fb^{-1} de datos, de los cuales 23 fb^{-1} son a 8 TeV se produjeron en 2012. Con una eficiencia en la toma de datos superior al 95% y más del 99% de sus canales funcionando, ATLAS proporciona datos de una calidad sin igual. Para poder analizar estos datos, la colaboración ha diseñado un sistema de computación distribuido basado en tecnologías Grid.

El modelo de computación de ATLAS

Tanto el LHC como ATLAS han mostrado un rendimiento sin precedentes desde que empezaran a funcionar en diciembre de 2009. Con una eficiencia en la toma de datos superior al 95% y más del 99% de sus canales de lectura operativos, ATLAS suministra datos de una calidad sin par. Desde su inicio, casi 27 fb^{-1} de colisiones han sido almacenados por el experimento ATLAS de los cuales 23 fb^{-1} se obtuvieron en 2012. Esto se traduce en más de 140 PetaBytes (PB) de datos y simulaciones que están siendo analizados por casi 3000 físicos de 174 instituciones alrededor del mundo. Siguiendo un modelo basado en tecnologías Grid [45], los diferentes centros de computación del experimento ATLAS se agrupan según una jerarquía que va desde el Tier-0 en el CERN pasando por los 11 Tier-1 y los casi 80 centros Tier-2 repartidos por todo el mundo. En España hay un Tier-2 federado formado por IFIC (Instituto de Física Corpuscular de Valencia), IFAE (Instituto de Física de Altas Energías de Barcelona) y UAM (Universidad Autónoma de Madrid). IFIC representa el 50% de los recursos españoles y es el responsable de la coordinación de las actividades de la federación. Cada uno de los centros de la federación española

tiene asociado un Tier-3 que comparte su infraestructura con el Tier-2. Los Tier-3 disponen de recursos adicionales de CPU y disco que, cuando están desocupados, pueden servir de apoyo al Tier-2.

El modelo de distribución de datos de ATLAS distingue entre réplicas primarias y secundarias. Réplicas primarias se distribuyen a los Tier-1 por redundancia y a los Tier-2 para análisis. El espacio en disco restante se llena con réplicas secundarias de los datos más populares. Para maximizar el uso de los recursos, desde 2012 unas cuotas limitan el número de trabajos tanto de análisis como de producción de simulaciones que se pueden procesar en el Tier-1. Como resultado, gran parte de los análisis y de la producción de simulaciones tiene lugar en los Tier-2. Como puede verse en la figura 5, de los más de 45 mil millones de colisiones que fueron procesadas en España en 2012 más del 70 % tuvo lugar en los Tier-2. En términos de espacio en disco esto equivale a más de 2 PB de datos almacenados en el Tier-2.

El software de ATLAS es accesible para la colaboración gracias a un sistema de ficheros en red basado en HTTP, CVMFS, en el que las versiones se descargan bajo demanda. Existen dos niveles de monitorización: una interna y otra global vía LHC Grid.

El modelo de computación distribuida de ATLAS ha evolucionado durante los primeros años de actividad del LHC. Ejemplos son una distribución más dinámica de los datos y de los trabajos, tanto de análisis como de producción de simulaciones, o mejoras en la monitorización de los centros y de la red. El funcionamiento del Tier-2 federado español ha sido excelente ofreciendo una disponibilidad² y una fiabilidad³ combinada que roza 100 % durante este periodo como se muestra en la figura 6. Debido a su excelente rendimiento, los tres centros del Tier-2 federado español han sido calificados como centros de máxima calidad, T2D. Esto ha generado un aumento considerable tanto en el número de trabajos como en la cantidad de datos recibidos en España haciendo del Tier-2 español una pieza fundamental en un modelo de computación que ha permitido analizar los datos de ATLAS a un ritmo sin precedentes.

²La disponibilidad (availability) se define como la fracción de tiempo que el centro está disponible respecto al tiempo total.

³La fiabilidad (reliability) se define como la fracción de tiempo que el centro está disponible respecto al tiempo total pero sin tener en cuenta las paradas programadas.

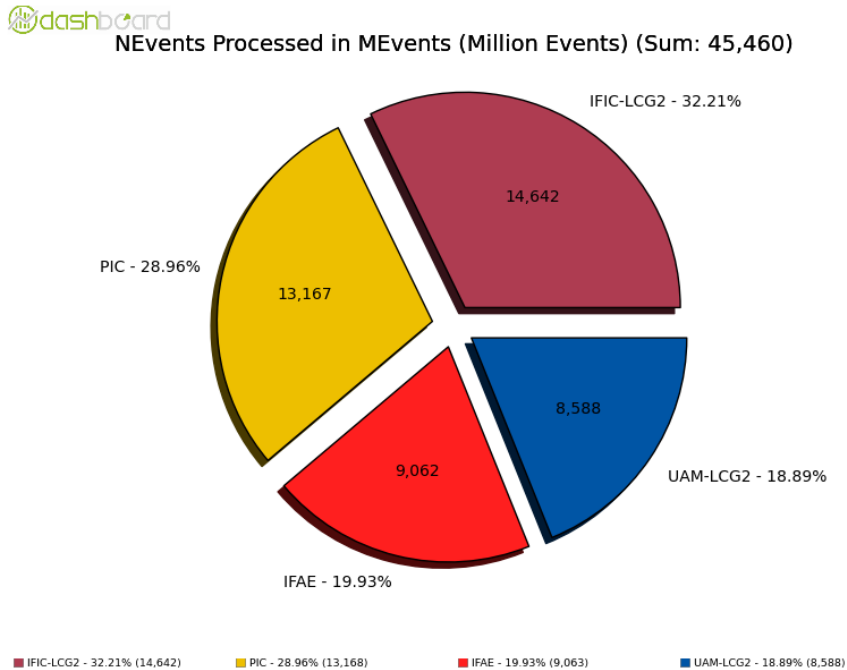


Figura 5: Volumen de colisiones procesadas en 2012 en el Tier-1 (PIC) y los centros del Tier-2 federado español.

El Tier-3 del IFIC

Los Tier-3 son centros que no están controlados por la colaboración ATLAS y que sirven de apoyo a los intereses particulares de sus usuarios. Suelen ofrecer la posibilidad de analizar datos de forma lineal (*batch*) o en paralelo (*interactive*).

Según el modelo de computación de ATLAS, los trabajos deben ejecutarse en aquellos centros donde los datos requeridos estén disponibles. Sin embargo, en un Tier-3 los físicos pueden llevar a cabo sus análisis *in situ* además de disponer de los recursos Grid de ATLAS.

El Tier-3 del IFIC [75, 76] está ligado a su Tier-2, que tiene el 50% de los recursos del Tier-2 federado español, y comparte con éste algunos puntos de su diseño. Como se muestra en la figura 7, una parte de los recursos del Tier-3 está acoplada a los del Tier-2 en un entorno Grid. Estos recursos son usados

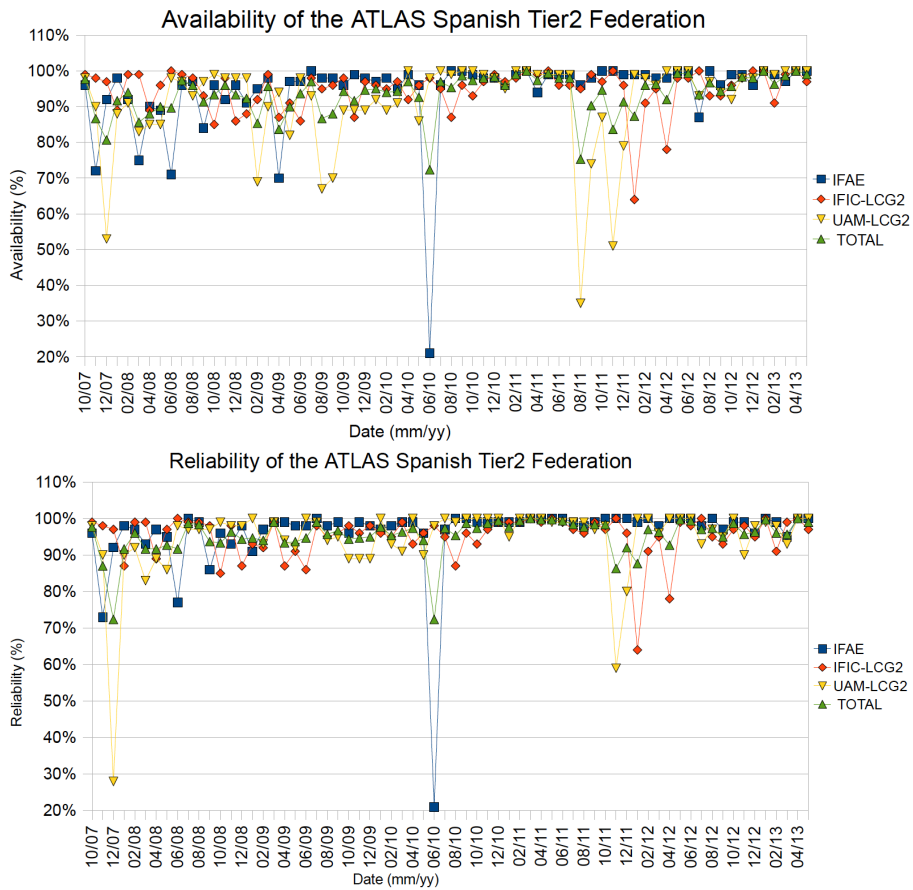


Figura 6: (Arriba) Disponibilidad (availability) y (abajo) fiabilidad (reliability) de los centros del Tier-2 federado español.

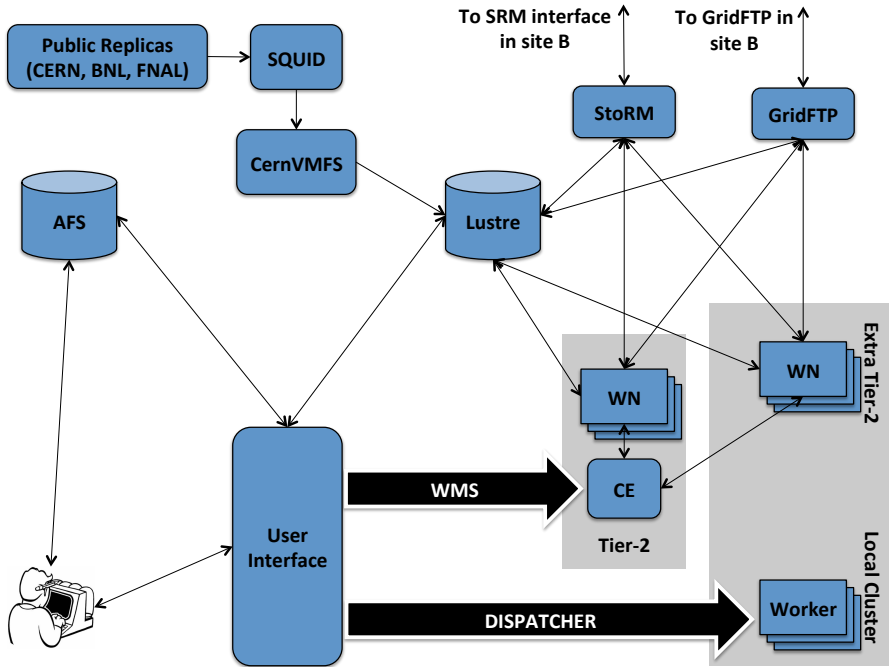


Figura 7: Esquema del diseño del Tier-3 del IFIC.

preferentemente por los usuarios del IFIC pero pueden servir de apoyo al Tier-2 cuando están desocupados.

Además, se han hecho las primeras pruebas para la instalación de una granja PROOF para el análisis en paralelo de datos.

Una característica importante del Tier-3 del IFIC es que usa el mismo sistema de almacenamiento que el Tier-2 (Lustre + StoRM). Su componente principal es el sistema de ficheros distribuido para clústers Lustre. Este sistema de ficheros está disponible para Linux y ofrece un interfaz compatible con POSIX. Gracias a este interfaz los usuarios pueden acceder al sistema usando los comandos típicos de Linux.

En ATLAS, la distribución de datos se organiza usando unos identificadores llamados *space tokens*. El Tier-3 del IFIC tiene asociado un *space token* que le permite gestionar los datos de los usuarios, ATLASLOCALGROUPDISK.

Actualmente, el Tier-3 tiene alrededor de 100 TB dedicados que están divididos en dos partes. Una parte bajo control de ATLAS, 60 TB, en la que sólo se puede escribir usando las herramientas Grid. Y otra independiente de 40 TB.

Jets en física de partículas

A distancias cortas, la libertad asintótica permite a los quarks y gluones moverse como si de partículas libres se tratara. Cuando se aceleran, estas partículas producen radiación de frenado, *bremssstrahlung*, que en este caso se compone de gluones y pares $q\bar{q}$ que, a su vez, se convierten en corros de hadrones muy colimados que llamamos jets. La determinación del momento del jet revela información sobre el quark o gluón que lo originó. Es por eso que los jets jugaron un papel esencial en el establecimiento de la teoría de las interacciones fuertes, la cromodinámica cuántica.

La primera evidencia experimental de la existencia de jets se encontró en SPEAR, un colisionador e^+e^- , de $\sqrt{s} = 8$ GeV instalado en SLAC [83, 84]. Las predicciones esperaban el proceso que aparece en la ecuación R.11,

$$e^+e^- \longrightarrow q\bar{q} \longrightarrow 2 \text{ jets} \quad (\text{R.11})$$

donde el par $q\bar{q}$ hadroniza hasta convertirse en dos jets. El resultado anunciado en 1975 necesitó un análisis refinado que demostró que verdaderamente se trataba de la estructura de unos jets y que las distribuciones angulares de dichos jets eran consistentes con aquellas de un par de partículas con espín 1/2.

Los primeros jets vistos a simple vista aparecerían en 1979 en otro colisionador e^+e^- , PETRA en DESY. Los 46 GeV de energía en centro de masas permitieron producir sucesos con dos jets limpios. Sucesos con tres jets, como el que se muestra en la figura 8, se interpretaron como $e^+e^- \longrightarrow q\bar{q}g$, donde uno de los quarks emitidos radiaba un gluón muy energético.

El Modelo Estándar se puso a prueba concienzudamente durante las siguientes décadas. El progreso teórico y experimental, así como el desarrollo de herramientas Monte Carlo, contribuyeron a consolidar la física de jets. A las energías de la era LEP⁴ ya era posible estudiar las interacciones del gluón consigo mismo en procesos como $e^+e^- \longrightarrow q\bar{q}gg$. Un buen ejemplo es el suceso con cuatro jets que se muestra en la figura 9. Este suceso tomado en el

⁴LEP empezó con una energía en centro de masas de 90 GeV, suficiente para producir el bosón Z^0 . En una segunda fase, LEP2 alcanzó los 200 GeV, donde se estudió con detalle la producción en pares de bosones electrodébiles.

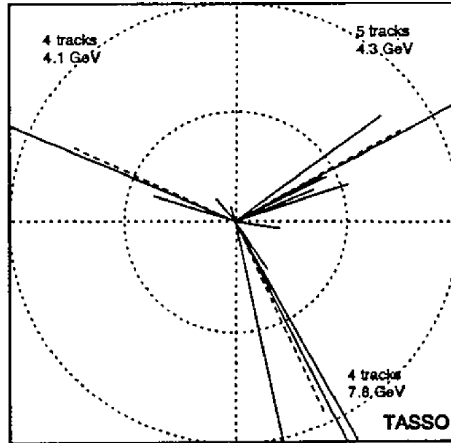


Figura 8: Observación de $e^+e^- \rightarrow q\bar{q}g \rightarrow 3 \text{ jets}$ en el detector TASSO.

experimento DELPHI es interpretado como $Z^0 Z^0 \rightarrow b\bar{b}q\bar{q}$. Técnicas como la reconstrucción de vértices secundarios dentro de los jets permitieron la identificación de jets generados por quarks b provenientes de la desintegración de bosones electrodébiles.

En máquinas hadrónicas como el Tevatron o el LHC la reconstrucción de jets se convierte en una herramienta esencial. Los jets fueron imprescindibles en el descubrimiento del quark top en Tevatron [86]. Un buen ejemplo es el suceso de la derecha en la figura 9, donde un par $t\bar{t}$ aparece dejando una señal ‘e + 4 jets’ en el detector CDF de Tevatron. Aparte de poner a prueba el Modelo Estándar a la mayor energía posible, determinar la distribución de gluones en el interior del protón o medir con precisión los valores de las constantes de acoplamiento de QCD, los jets también pueden usarse para buscar física más allá del Modelo Estándar.

Boosted objects en física de partículas

En el nuevo régimen de energías al que nos permite llegar el LHC se producen partículas pesadas, como los bosones *gauge* W^\pm y Z^0 o el quark top, con un momento que excede considerablemente su masa. En Tevatron, sólo se produjo un puñado de pares $t\bar{t}$ con una masa del orden de 1 TeV. En los $\sim 20 \text{ fb}^{-1}$ de

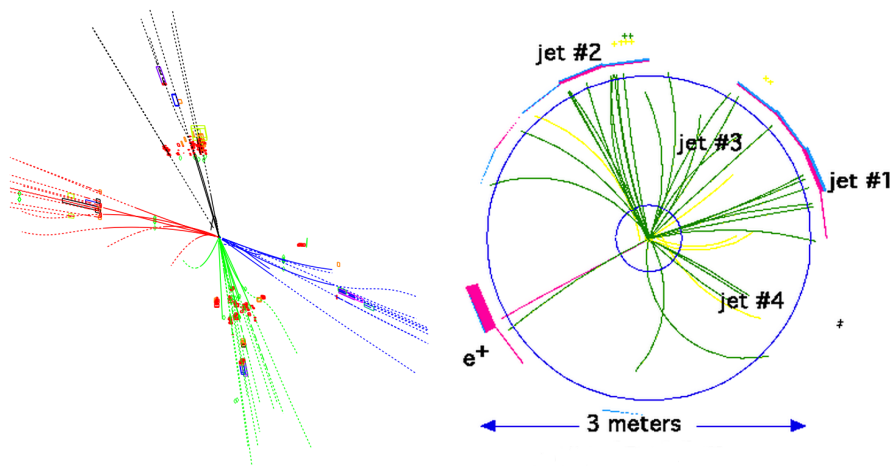


Figura 9: (Izquierda) Evento $Z^0 Z^0 \rightarrow b\bar{b}q\bar{q}$ recogido por el detector DELPHI del acelerador LEP. Aparecen vértices secundarios reconstruidos en los jets b (negro y rojo). (Derecha) Observación de un par $t\bar{t}$ por el detector CDF de Tevatron. El quark \bar{t} se desintegra hadrónicamente a \bar{u} (jet 2), d (jet 3) y \bar{b} (jet 1). el quark t se desintegra en un quark b (jet 4), e^+ y un ν_e estimado en la dirección de las 6 en punto.

colisiones protón-protón a 8 TeV producidos en el LHC ya hay miles y serán millones cuando el acelerador alcance su energía de diseño, ~ 14 TeV, en 2015. Estos *boosted objects* son ya una parte esencial del programa de física del LHC y su importancia está llamada a crecer en el futuro.

El *boost* de Lorentz de estas partículas altera drásticamente la topología observada haciendo inservibles los criterios de identificación y aislamiento convencionales, que fueron desarrollados para partículas aproximadamente en reposo. A medida que los partones de la desintegración de la partícula se coliman en un área del detector más y más pequeña, según se muestra en la figura 10, los algoritmos de jets dejan de ser capaces de reconstruir los jets resultantes por separado. En su lugar, toda la desintegración es reconstruida como un único jet.

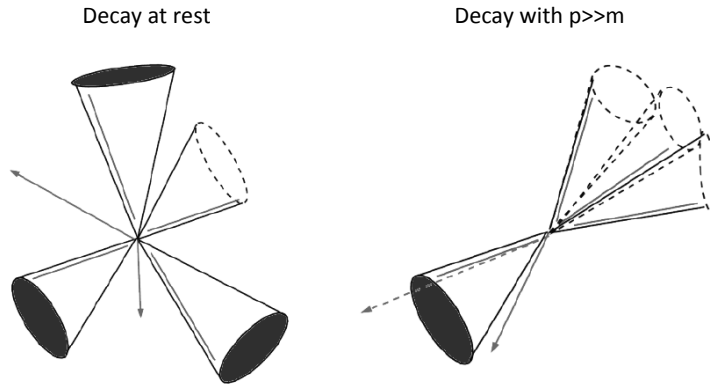


Figura 10: Esquema de la desintegración de un par $t\bar{t}$ (izquierda) en reposo y (derecha) con una energía mucho mayor que la masa del par produciendo así la topología típica de la desintegración de un *boosted object*.

Subestructura de jets

Es posible identificar a los jets que contienen la desintegración de un *boosted object* mediante el estudio de su estructura interna. La comprensión de variables y técnicas relacionadas con la subestructura de jets, su puesta a punto con datos de ATLAS, y la estimación de sus incertidumbres son esenciales en los análisis que involucran *boosted objects*. Algunas técnicas se centran en limpiar los jets de partículas no deseadas cuya presencia disminuye la resolución de masa. Otras buscan subestructura identificando subjets dentro de un jet dado. También hay métodos basados en el flujo de energía dentro del jet. Puede encontrarse una discusión más detallada sobre estas técnicas en las referencias [93, 100, 101] y [102]. Este texto se va a centrar sólo en los observables que serán usados más adelante: masa de jets y las escalas de fraccionamiento del algoritmo k_t , $\sqrt{d_{ij}}$.

El observable más sencillo conceptualmente hablando es la masa invariante del jet, m_j . Si la desintegración de un objeto pesado como el quark top está con-

tenida completamente en un solo jet, m_j de dicho jet será mayor que la de aquellos jets que procedan de quarks ligeros o de gluones.

El algoritmo de reconstrucción de jets k_i empieza agrupando componentes de baja energía y cercanos entre sí formando en cada iteración subjets más y más pesados. Según describe la ecuación R.12, $\sqrt{d_{ij}}$, indica la escala a la que el algoritmo realiza las sucesivas iteraciones hasta formar el jet. Así, $\sqrt{d_{12}}$ involucra al último paso, donde los dos últimos subjets se unen formando el jet final. De la misma forma, $\sqrt{d_{23}}$ indica la escala a la que se pasa de tres a dos subjets, etc.

$$\begin{aligned} d_{ij} &= \min(p_{Ti}^2, p_{Tj}^2) \times \frac{\Delta R_{i,j}^2}{R^2} \\ \Delta R_{ij} &= \sqrt{\Delta\eta^2 + \Delta\phi^2} \end{aligned} \quad (\text{R.12})$$

La distribución de los componentes de un jet varía según tenga éste su origen en quarks ligeros o en gluones, en cuyo caso los componentes se concentran en su mayoría en el centro del jet, o bien en la desintegración de una partícula pesada. Las escalas de fraccionamiento, $\sqrt{d_{ij}}$, son sensibles a esta estructura interna. El valor de $\sqrt{d_{12}}$, por ejemplo, suele ser del orden de $m/2$ en el caso de jets que contienen la desintegración de una partícula pesada de masa m , mientras que la distribución inclusiva se centra en valores menores.

Los resultados descritos aquí son parte de una discusión más extensa publicada por la Colaboración ATLAS en *Journal of High Energy Physics* [132]. Esta es la primera medida de este tipo de variables en el LHC y el primer estudio de jets de gran tamaño (gran R) en cualquier experimento. Los datos se muestran compatibles con las predicciones obtenidas con los generadores Monte Carlo PYTHIA y HERWIG++. Las incertidumbres sistemáticas estimadas para el momento de jets, su masa y las dos primeras $\sqrt{d_{ij}}$ se muestran en la tabla 2.

Búsquedas de nueva física en pares $t\bar{t}$ de alto momento.

Los resultados que se reportan en esta sección están basados en un análisis de los 2.05 fb^{-1} de datos de ATLAS a $\sqrt{s} = 7 \text{ TeV}$ que se recogieron en 2011 y han sido publicados en *Journal of High Energy Physics* [166]. El estudio se centra en el canal de desintegración de pares $t\bar{t}$ conocido como ‘ $l + jets$ ’, en el que uno de los quarks top se desintegra hadrónicamente, $t \rightarrow b qq$, y el otro quark top deja un leptón en el estado final, $t \rightarrow b l\nu$. Los pares $t\bar{t}$ se reconstruyen y seleccionan utilizando una técnica novedosa que se centra en sucesos con la

Tabla 2: Incertidumbres sistemáticas finales para jets anti- k_t .

Uncertainty	200–300 GeV	300–400 GeV	400–500 GeV	500–600 GeV
JES	4.0 %	5.2 %	6.0 %	3.9 %
JMS	4.5 %	4.5 %	6.0 %	6.0 %
JER	20.0 %	20.0 %	20.0 %	20.0 %
JMR	20.0 %	20.0 %	20.0 %	20.0 %
$\sqrt{d_{12}}$ scale	4.4 %	3.8 %	6.0 %	6.8 %
$\sqrt{d_{23}}$ scale	4.4 %	3.4 %	5.1 %	5.1 %
$\sqrt{d_{12}}$ res.	21.0 %	22.0 %	48.0 %	31.0 %
$\sqrt{d_{23}}$ res.	20.0 %	20.0 %	21.0 %	25.0 %

topología típica de la desintegración de *boosted objects* discutida en secciones anteriores.

El *boost* confina los productos de la desintegración de los quarks top en un área pequeña del detector. Esto hace posible la utilización de criterios de cercanía entre los productos de la desintegración del quark top o incluso permite reconstruir dichos productos como un único jet de gran tamaño. Es por eso que en este análisis a la colección de jets recomendada por ATLAS, anti- k_t , $R = 0,4$ se añade una colección de jets de mayor tamaño, anti- k_t , $R = 1,0$.

El proceso empieza seleccionando un leptón aislado de buena calidad. El ν se reconstruye a partir de una estimación de la energía faltante en el suceso (en inglés *missing transverse energy*, E_T^{Miss}). La componente longitudinal del momento del neutrino, p_z^ν , se obtiene imponiendo a la suma de leptón y ν la masa del bosón W .

El top leptónico, $t \rightarrow b l \nu$, se forma seleccionando como jet b el jet anti- k_t , $R = 0,4$ más cercano al leptón con $p_T > 20 \text{ GeV}$ y $\Delta R_{lj} < 1,5$.

Los productos de la desintegración del top hadrónico, $t \rightarrow b q q$, se reconstruyen como un único jet anti- k_t , $R = 1,0$. Este jet debe cumplir: $p_T > 250 \text{ GeV}$, masa $m_j > 100 \text{ GeV}$, $\sqrt{d_{12}} > 40 \text{ GeV}$, además, debe encontrarse en el hemisferio opuesto al jet anti- k_t , $R = 0,4$ usado en la reconstrucción del top leptónico. El jet anti- k_t , $R = 1,0$ con mayor p_T de entre los que cumplen estos requisitos es el elegido como top hadrónico.

El sistema $t\bar{t}$ es el resultado de sumar los cuadrivectores de los tops hadrónico y leptónico. Este método de reconstrucción es intrínsecamente robusto frente a los efectos de la presencia de jets adicionales en el suceso procedentes de la radiación de los quarks top andes de desintegrarse. Este tipo de jets no cumple los criterios de cercanía impuestos. Los sucesos en que se usan estos jets

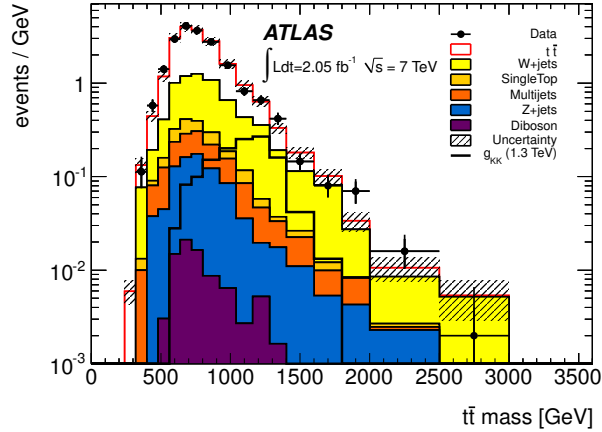


Figura 11: Distribución de la masa invariante reconstruida de los candidatos a $t\bar{t}$ después del proceso de selección de la señal. El área rayada indica la incertidumbre en la normalización de la predicción del SM, pero no incluye la incertidumbre en la forma o el impacto de las incertidumbres en los objetos reconstruidos. El tamaño de los bins se ha elegido acorde a la resolución de masa para la señal.

suelen caer en la cola a alta masa de la distribución y contribuyen, por tanto, a empeorar la resolución de masa.

Es importante mencionar que esta es una implementación parcial del algoritmo diseñado para identificar *boosted top quarks*. El requisito de aislamiento aplicado al leptón es el que se usa comúnmente en las técnicas clásicas y causa una caída en la aceptación a partir de 2 TeV.

Los datos de ATLAS se comparan con un patrón del espectro de masa de los pares $t\bar{t}$ predicho por el Modelo Estándar como se puede ver en la figura 11. El patrón es una combinación de estimaciones de la producción inclusiva de jets en QCD obtenidas a partir de los datos, una simulación de la producción de $W + jets$ con una normalización derivada de datos, y simulaciones Monte Carlo de varios procesos del Modelo Estándar que pueden confundirse con la señal de los pares $t\bar{t}$.

La evaluación de las incertidumbres sistemáticas en la normalización y la forma de las predicciones del Modelo Estándar se basa en simulaciones Monte

Carlo y en técnicas aplicadas *in situ*. Las incertidumbres dominantes son la asociada a la determinación de la escala de energía y la masa de los jets, y la asociada a la normalización del fondo $W + jets$.

No se ha encontrado ninguna desviación significativa entre datos y patrón del Modelo Estándar una vez incluidas las incertidumbres. En ausencia de discrepancias, se procede a poner límites a la presencia de nueva física. En el caso de la resonancia estrecha que produce el Z' leptofóbico de los modelos topcolor de la referencia [142], los límites en $cross - section \times branching ratio$ van desde los 30 pb para una masa de 500 GeV hasta 100 fb para masas superiores a 2 TeV, en acuerdo con los límites esperados. Estos modelos han sido excluidos para masas menores de 1.2 TeV por este análisis.

También se han derivado límites para el casos de resonancias más anchas típicas de modelos con objetos pesados con carga de color. El límite superior en $cross - section \times branching ratio$ para el gluón de Kaluza-Klein ($\Gamma = 0,15M$) de la referencia [31] es aproximadamente un factor dos mayor que el correspondiente a un Z' de la misma masa. El límite inferior en la masa del gluón de Kaluza-Klein se ha fijado en 1.6 TeV. Estos límites aparecen gráficamente en la figura 12.

Esta es la primera aplicación de estas técnicas en ATLAS. El método funciona especialmente bien para masas superiores a 800 GeV comparado con otras búsquedas de resonancias coetáneas [165, 176] como se muestra en la figura 13. Además, el método de reconstrucción se muestra robusto frente a los efectos de la radiación del estado inicial y del *pile-up*. Todo esto prueba que las técnicas de reconstrucción dirigidas a *boosted objects* pueden mejorar el potencial de las búsquedas de nueva física más allá del Modelo Estándar en el LHC.

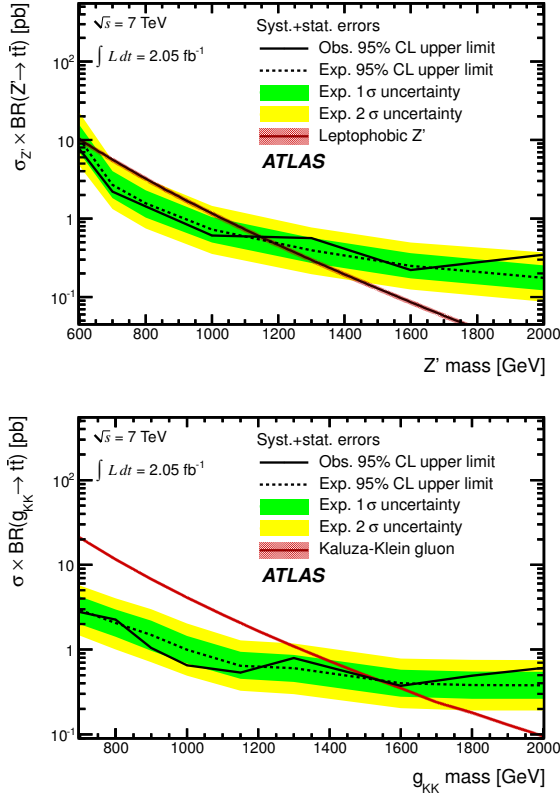


Figura 12: Límites superiores esperados (línea discontinua) y observados (línea sólida) en la sección eficaz de producción de Z' (arriba) y gluones de Kaluza–Klein (abajo). Las bandas muestran el rango en que el límite se espera que se encuentre en el 68 % (verde) y el 95 % (amarilla) de los pseudo-experimentos, y las líneas rojas corresponden a la sección eficaz predicha por los modelos. La banda alrededor de la curva de la señal esta basada en el efecto de la incertidumbre de las PDF usadas en la predicción. f

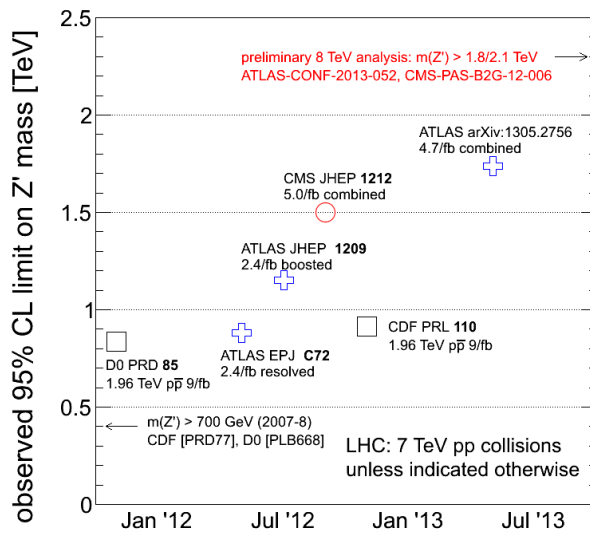


Figura 13: Evolución de las búsquedas de una resonancia Z' estrecha que se desintegra en un par $t\bar{t}$.

Un juego de niños

Artículo ganador de los II Premios Opinión Innovadora convocado por la Càtedra de Divulgació de la Ciència UCC+i de la Universitat de València con el apoyo del diario Las Provincias.

Si hay una cualidad que es imprescindible en un científico esa es la curiosidad y, de eso, los niños tienen de sobra. Todos los padres han sufrido alguna vez la curiosidad de sus hijos.

- ¿Qué hay en esa bolsa?
- Manzanas, hijo.
- Ah... y ¿qué hay dentro de las manzanas?
- Si la partes por la mitad verás que dentro hay un hueso, hijo.
- ¿Y dentro del hueso? ¿qué hay?
- Mmm... no sé (ahí me ha pillado)... ah, sí, ¡átomos!, hijo, todo está hecho de átomos.
- ¿Y dentro de los átomos no hay nada?
- ... eso preguntáselo a tu madre...

Tenemos paciencia con los niños porque sabemos que todos nos hemos hecho estas preguntas alguna vez. La curiosidad ha llevado a algunos a seguir preguntando. Han mirado dentro del hueso de manzana y han visto moléculas. Han visto también que las moléculas están hechas de átomos y los átomos de electrones, protones y neutrones, y no se han parado ahí. Gracias a esa curiosidad hoy podemos explicar todo lo que sucede a nuestro alrededor en términos de partículas elementales y sus interacciones. Como si de un juego de LEGO se tratara, el universo que nos rodea está formado por partículas que actúan como bloques, los fermiones, y partículas que actúan como pegamento, los bosones.

A diferencia de las manzanas, un protón no se puede partir con un cuchillo y las piezas que lo forman no se pueden ver con los ojos. Entonces, ¿cómo podemos estudiar cosas tan pequeñas? Pues jugando con ellas, por ejemplo, al billar. Jugando al billar con los átomos, lanzando pequeñas partículas contra ellos y viendo la dirección en que éstas rebotan, descubrimos que los átomos no son a-tomos (indivisibles) sino que están formados por electrones, protones y neutrones. En el billar real si golpeamos con suficiente fuerza las bolas pueden romperse y lo mismo pasa en física de partículas. Esta es otra forma de estudiar el interior de los átomos, quizá la más intuitiva, romperlos.

El problema es que los protones que hay en el interior de los átomos se mantienen unidos gracias a un pegamento muy fuerte. Este bosón, llamado gluón, es 10^{38} veces más fuerte que la gravedad. ¡Esto es un uno seguido de 38 ceros! De hecho, es la cosa más fuerte que se conoce. Están los protones tan bien pegados que para romperlos hace falta mucha energía así que tenemos que darles un buen golpe. Lo que hacemos es acelerarlos todo lo posible y hacer que choquen entre sí. Esto es lo que se hace en aceleradores de partículas como el LHC.

Romper protones puede ser muy divertido pero no hay que olvidarse de recoger los trozos. No queremos que escape nada así que cubrimos la zona del choque lo mejor posible con varias capas de detectores. Cada capa está pensada para reaccionar a diferentes propiedades de las partículas. Así podemos identificar los restos de la colisión viendo qué capas se activan y cuales no.

La cantidad de información que se genera es tan grande que para analizarla se ha desarrollado el Grid, un sistema que permite que los datos se analicen por científicos de todo el mundo de forma distribuida.

Con todos estos datos jugamos a '¿quién es quién?'. Este juego consiste en encontrar a un personaje preguntando sobre sus características, ¿lleva gafas?, ¿tiene bigote?, ... de forma que vamos descartando los candidatos que no se ajustan a la descripción hasta que sólo queda uno. Jugando a este juego, preguntando y descartando, conseguimos encontrar indicios del famoso bosón de Higgs. Pero a diferencia del juego original en el que se elige un personaje de 24 candidatos, en el caso del escurridizo bosón encontramos un Higgs por cada billón de candidatos.

Aun no estamos totalmente seguros pero, si se confirma que el nuevo bosón es el bosón de Higgs, habremos respondido a una pregunta que llevamos haciéndonos 50 años. ¡Olvidémonos de la bata blanca y el pelo alborotado! La imagen que mejor representa a un científico es la de un niño curioso que nunca deja de preguntar ¿por qué? ¿por qué? ¿por qué? y es por eso que la ciencia es, en realidad, un juego de niños.

Agradecimientos

Estas son las últimas líneas de un texto que resume muchos años de trabajo y cierra una etapa de mi vida que no olvidaré. Por eso, me gustaría terminar esta tesis como termino esta etapa, diciendo que estoy agradecido.

Agradecido a mis padres por su apoyo incondicional y su insaciable curiosidad. Las dos pero, sobre todo, la segunda, la curiosidad, sin ella y sin vosotros esta tesis no habría sido posible. A mi hermana, por defenderme cuando no estoy, por los recados y por el ‘¿a dónde vas vestido así?’.

Agradecido también a Pepe Salt por no venderme la moto, por ayudarme a empezar y animarme a seguir. A Santi y a Marcel por su paciencia y su confianza, y por demostrarme que el trabajo duro, con amigos, es menos duro.

Agradecido al grupo de computing del IFIC por tratar de convertirme, unos al islam, otros a UNIX, todos con la mejor de las sonrisas. Al grupo de ATLAS en Valencia por sus preguntas y sus respuestas, por el plot del mes y el ‘ola k ase’. A la oficina más caliente del mundo mundial, a mis chicas, por aguantarme, que no es poco.

Agradecido a la banda de física de neutrinos, los que fueron y los que son, por dejarse espiar, por los motes, por las inauguraciones y las despedidas, por las cenas, por la cervecita ‘rápida’ después del curro, por las cosas que ‘no han pasado’ y los lugares en los que ‘no hemos estado’.

Agradecido a los locos de la piscina por compartir su locura, por tocarme los pies, por el ‘estírate’, el ‘sube el culo’ y el ‘saca el hombro’, y también por els pastissets de boniato, la música y la poesía.

Agradecido a mis amigos, los de verdad, que, a parte de los del pueblo, son Justo, Nacho y Alberto, por hacerme dos veces grande, por las conversaciones, las maldades y las bodas secretas, por las historias de Mordor y la araña negra y, por encima de todo, por las risas.

Por todo esto y muchas otras cosas, a todos vosotros y especialmente a mis padres, gracias.