IL NUOVO CIMENTO Vol. 33 C, N. 4 Luglio-Agosto 2010
DOI 10.1393/ncc/i2010-10675-0

Colloquia: TOP2010

The short-term strategy and plan for $t\bar{t}$ and single-top at the LHC

P. Ferrari(1), B. Hegner(2) for the ATLAS and CMS Collaborations

(1) *Nikhef National Institute for Subatomic Physics and University of Amsterdam*
Kruislaan 409, P.O. Box 41882, NL - 1009 DB Amsterdam, Netherlands
(2) *CERN - 1211 Geneve 23, Switzerland*

(ricevuto il 22 Luglio 2010; approvato il 22 Luglio 2010; pubblicato online il 18 Ottobre 2010)

Summary. — This paper is intended to give an overview of first $t\bar{t}$ and single-top physics results that the CMS and ATLAS Collaborations are going to provide with the early data. Particular attention is given to the determination of the $t\bar{t}$ cross-section and to the observation of the single top process at LHC. Early results on top quark mass determination are also discussed.

PACS `14.65.Ha` – Top quarks.

1. – Introduction

The top quark completes the three-generation structure of the Standard Model (SM). After QCD jets, W and Z bosons, the production of top quarks is the dominant process in pp collisions at multi-TeV energies: already in the first phase of LHC running at a centre-of-mass energy ($\sqrt{s}$) of 7 TeV with $10\,\text{pb}^{-1}$ of integrated luminosity, hundreds of $t\bar{t}$ pairs will be produced. Soon after, the electroweak single top process (given that the t-channel production has a cross-section of about 1/3 of $t\bar{t}$) will also provide an abundant sample of events. Studying these events may answer many questions on the standard model and beyond.

2. – The $t\bar{t}$ production cross-section determination in the semi-leptonic channel at LHC

Several techniques have been developed by both ATLAS and CMS to measure the $t\bar{t}$ cross section in the single lepton channel. The studies reported here are focused on 20 (200) pb^{-1} of CMS (ATLAS) data assuming 10 TeV pp collisions. The analyses have been developed without the use of b-tagging, since it was not expected to have reached the required level of performance with early data. On the contrary, present results by CMS and ATLAS on the first data show that the b-tagging algorithms designed for early data taking are well understood and well modelled by the MC.

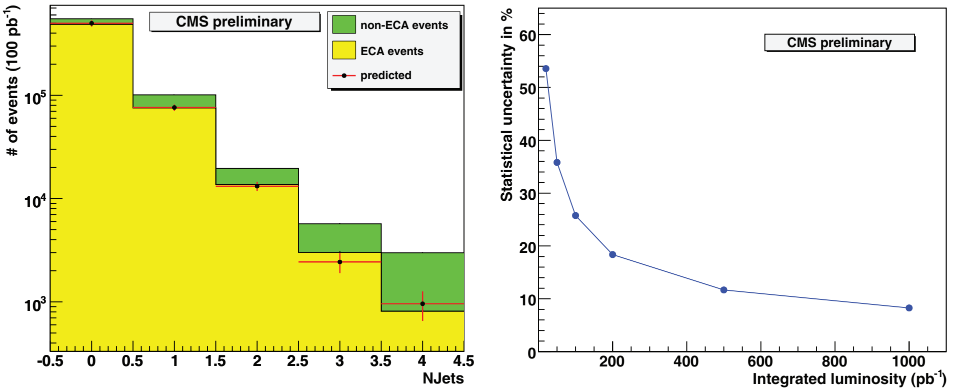

Fig. 1. – Left: true ECA and non-ECA events as a function of the jet multiplicity for an integrated luminosity of 100 pb^{-1}. The errors bars correspond to statistical and systematical uncertainties summed in quadrature. Right: statistic uncertainty on the estimate of the ECA contribution as a function of the integrated luminosity.

Both collaborations are using very simple and robust selections to select $t\bar{t}$ events in the semi-leptonic channel. The relevant backgrounds have been evaluated; they are W/Z+jets, single-top, di-bosons, QCD multi-jets.

CMS requires to have exactly one high-p_T isolated muon ($p_T > 20$ GeV) or electron ($p_T > 30$ GeV) and 4 high-p_T jets ($p_T > 30$ GeV). In the electron channel a couple of different options to reject photon conversions and Z+jets are considered. No missing transverse energy (MET) cut is applied. This leads to a S/B of about 1.9 and 1 in the muon and the electron channel, respectively [1, 2].

ATLAS requires to have exactly one high-p_T isolated muon or electron ($p_T > 20$ GeV), 4 high-p_T jets ($p_T > 20$ GeV), three out of which should have at least $p_T > 40$ GeV. MET is required to be > 20 GeV. In addition the invariant mass of two of the jets is required to be within 10 GeV from the reconstructed W boson mass, this leads to a S/B ~ 2 both in the muon and the electron channel [3].

While those estimations are based purely on MC, there are large theoretical uncertainties on the cross-sections for W+jets and QCD backgrounds, ranging from 80 to 100%. The event yields are therefore extracted with better precision from the data themselves, even with a small amount of integrated luminosity.

2˙1. *W+jets data-driven background estimation.* – A number of techniques have been developed to estimate the W+jets background. One of CMS methods uses a novel technique based on the charge asymmetry of such events in pp collisions [1]: the total number of events with charge asymmetry can be obtained by counting the difference between the number of events with an anti-lepton N_+ and with a lepton N_-, which is assumed to be dominated by W+jets events according to $(N_+ + N_-)_{\rm data} = R_\pm(W) \times (N_+ - N_-)_{\rm data}$ where $(N_+ - N_-)_{\rm data}$ is measured in data. The factor $R_\pm(W) = \frac{(N_{W^+} + N_{W^-})}{(N_{W^+} - N_{W^-})}$ is obtained from simulation and can be used to extract the $W^{+(-)}$ cross-section. This method suffers from a large statistical uncertainty related to the term $(N_+ - N_-)_{\rm data}$ and is suitable to be used when an integrated luminosity of about 100 pb^{-1} is achieved. Figure 1 shows the distribution of predicted, true Events with Charge Asymmetry (ECA) and non-ECA events as a function of the jet multiplicity.

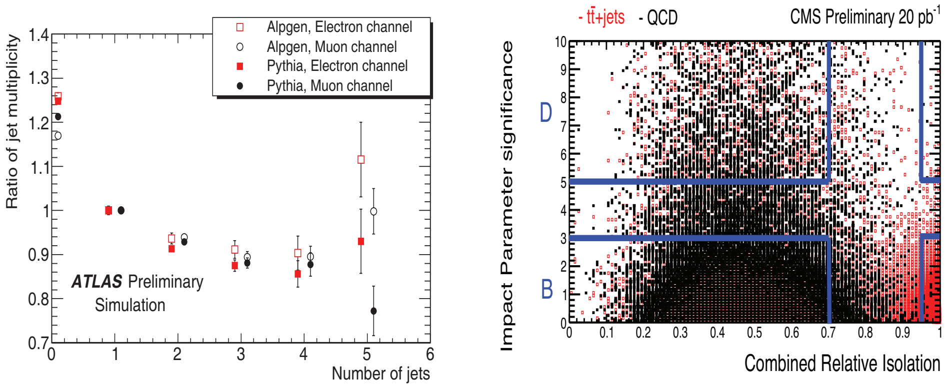

Fig. 2. – Left: ratio of reconstructed jet multiplicity for $W \to l\nu + jets$ over $Z \to ll + jets$ events. The ratio is taken after event selection cuts and normalizing the ratio to the 1 jet bin. Statistical errors are shown. Right: two-dimensional distribution of the muon impact parameter significance sig(d0) and RelIso variables.

ATLAS exploits another data-driven method that can be used with smaller integrated luminosities. Both Z+jets and W+jets events are selected in a low jet multiplicity region that is used as a control region (CR). The ratio of W+jets and Z+jets is extrapolated into the $t\bar{t}$ signal region (SR) with 4 or more jets. The extrapolation is possible since the ratio $(W^{\rm SR}/W^{\rm CR})_{\rm data} = (Z^{\rm SR}/Z^{\rm CR})_{\rm data} \times C_{\rm MC}$ is well known theoretically [3]: the factor $C_{\rm MC}$ is obtained from Monte Carlo. The ratio of reconstructed jet multiplicity for $W \to l\nu + jets$ over $Z \to ll + jets$ events is shown in fig. 2. The overall systematic and statistical uncertainties are about 20% at $\sqrt{s} = 10\,{\rm TeV}$ with $200\,{\rm pb}^{-1}$, and are expected to be about 50% (20%) at $\sqrt{s} = 7\,{\rm TeV}$ with 10 (100) $\rm pb^{-1}$. The largest contributions to the systematic uncertainty are coming from an assumed 50% uncertainty on the amount of QCD background and the knowledge of the $C_{\rm MC}$ factor.

2·2. *QCD data-driven background estimation.* – Both collaborations plan to measure the QCD background from the data. The so-called ABCD method is going to be adopted. In this method two independent discriminating variables are used in order to define a control region enriched in background and to extrapolate the amount of background in the signal region. CMS is choosing as discriminating variables the lepton isolation variable (RelIso) defined as the lepton energy relative to the lepton p_T and the lepton impact parameter significance [1]. Four independent areas A, B, C and D are chosen in the two-dimensional plane defined by the distribution of the two discriminating variables (see fig. 2 right) and the background in the signal region is estimated as: $N_A = (N_B N_C)/N_D$.

In addition CMS is exploiting a method that is based on extrapolating (via a fit) the RelIso variable from the control region where the energy deposition in a cone around the lepton is high to the signal region with low energy deposition (RelIso < 0.05). In fact W-like events, which contain an isolated lepton, are strongly peaked towards very small values of the RelIso variable, while QCD jet events are broadly distributed. Table I shows the number of QCD events estimated by the ABCD and the RelIso fit method in the muon channel. Systematic effects are calculated by changing the boundaries for the ABCD areas and by changing the binning and fitting range for the RelIso method: the systematic uncertainty associated with those methods is conservatively assigned to be 50%.

Table I. – *Number of QCD events estimated by the ABCD and the RelIso fit methods as a function of jet multiplicity in the muon channel. The results are in agreement with the expected number of events within errors for the signal region N-jets* ≥ 4*, and in the control regions N-jets* ≤ 3.

Jets	N(QCD) Predicted	N(QCD) Estimated ABCD	N(QCD) Estimated RelIso
2	327	325 ± 26	378 ± 82
3	53	48 ± 9	47 ± 24
≥ 4	7	12 ± 5	13 ± 7

2·3. *Results.* – Both ATLAS and CMS are determining the $t\bar{t}$ cross-section by using a simple counting method or by fitting the hadronic top mass reconstructed using the 3 jets that maximise the sum of their transverse momenta (see fig. 3 left).

The $t\bar{t}$ cross-section in the electron channel is obtained by ATLAS with the precision of $3(\text{stat})^{+14}_{-15}(\text{syst}) \pm 22(\text{lumi})\%$ and $14(\text{stat})^{+6}_{-15}(\text{syst}) \pm 20(\text{lumi})\%$ using the counting and the fit method, respectively, with an integrated luminosity of $200\,\text{pb}^{-1}$ at $10\,\text{TeV}$; a counting analysis not using any constraint on the missing traverse momentum obtains $3.4(\text{stat})^{+23}_{-25}(\text{syst}) \pm 34(\text{lumi})\%$. Figure 3 right shows the statistical uncertainty *versus* integrated luminosity obtained in electron and the muon channel, respectively by CMS at $\sqrt{s} = 10\,\text{TeV}$. In addition CMS is also performing a template fit to the η of the muon. The systematic errors (excluding the contribution from luminosity) are of about 19.2% and 23.8% for the muon η template fit, and of 23.8% for the hadronic top mass fit in the muon channel and of 20% in the electron channel.

By scaling the cross-sections to $\sqrt{s} = 7\,\text{TeV}$, already few pb^{-1} of integrated luminosity lead us to interesting results: CMS results obtained at $\sqrt{s} = 10\,\text{TeV}$ with $20\,\text{pb}^{-1}$, directly translate in what can be expected for $50\,\text{pb}^{-1}$ at $\sqrt{s} = 7\,\text{TeV}$. Already with $10\,\text{pb}^{-1}$ at $7\,\text{TeV}$, 60 top events per lepton flavour per experiment over a background of about 40, are expected: the di-jet mass will show a peak near the W, providing a further confirmation of the observation of a top signal.

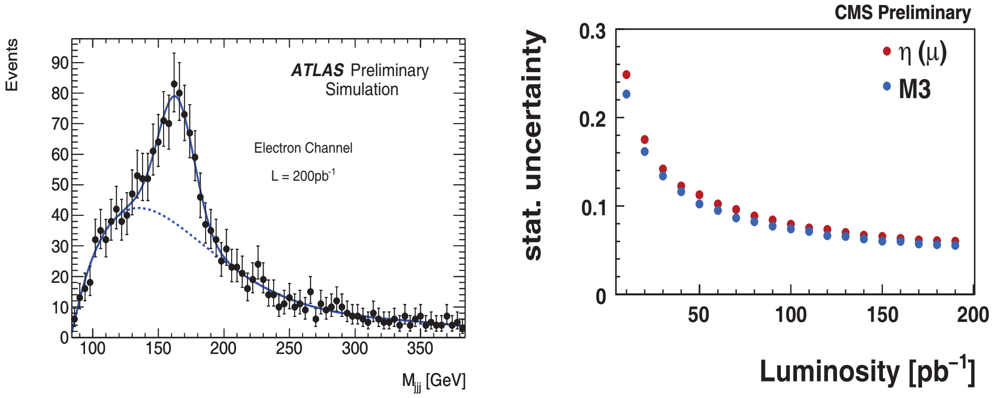

Fig. 3. – Left: fit to the hadronic top mass in the electron channel by ATLAS. Right: the statistical error *vs.* integrated luminosity on the muon channel by CMS.

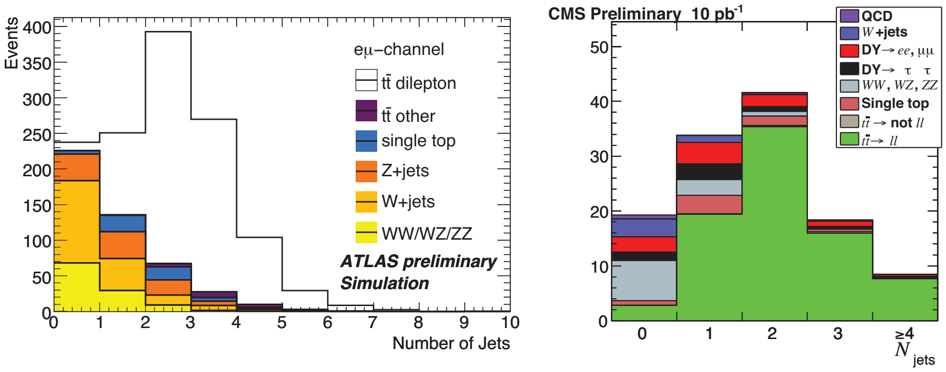

Fig. 4. – Jet multiplicity for dileptonic top events and the corresponding background after selection. Left: ATLAS. Right: CMS.

3. – The $t\bar{t}$ cross-section determination in the di-leptonic channel

Both collaborations are using very simple and robust selections to select $t\bar{t}$ events in the di-leptonic channel. CMS ($10\,\mathrm{pb}^{-1}$) requires two high-p_T isolated leptons ($p_T > 20\,\mathrm{GeV}$), at least 2 high-p_T jets of ($p_T > 30\,\mathrm{GeV}$) and a missing transverse momentum of $20\,\mathrm{GeV}$ ($e\mu$) or $30\,\mathrm{GeV}$ ($ee/\mu\mu$). An additional veto in the mass window of $15\,\mathrm{GeV}$ around the Z is used in the ee and $\mu\mu$ channel. This leads to a S/B of about 4 to 1 in all channels combined and about 9 to 1 in the $e\mu$ channel alone [4]. ATLAS ($200\,\mathrm{pb}^{-1}$) requires to have two high-p_T isolated leptons ($p_T > 20\,\mathrm{GeV}$), at least 2 high-p_T jets of ($p_T > 20\,\mathrm{GeV}$) and a missing transverse momentum of $20\,\mathrm{GeV}$ ($e\mu$) or $35\,\mathrm{GeV}$ ($ee/\mu\mu$). An additional veto in the mass window of $5\,\mathrm{GeV}$ around the Z is used in the ee and $\mu\mu$ channel. This leads to a S/B of about 4 to 1 [5].

The resulting signal and background contributions in the jet multiplicity for the ATLAS and CMS studies are shown in fig. 4.

3·1. *Data-driven Drell-Yan background estimation.* – In the e^+e^- and the $\mu^+\mu^-$ final states, the contribution from events with Drell-Yan (DY)+jets can be almost as large as the signal itself. Requiring a large transverse momentum and a mass outside the Z window significantly reduces the DY background. Still mis-measurements let DY events in the tail of the MET distribution pass the selections. Both ATLAS and CMS estimate the DY contribution in the signal region by scaling the MC prediction to match the observed numbers in the sidebands. The ATLAS Collaboration uses a grid as shown in fig. 5 with the columns defined by a Z mass window between $86\,\mathrm{GeV}$ and $96\,\mathrm{GeV}$. The rows are defined by MET values of $15\,\mathrm{GeV}$ and $35\,\mathrm{GeV}$. Using data from regions B, G, H and I plus MC estimates for the other regions, the estimated DY contributions in the signal regions A and C can be obtained by $A_{\mathrm{Est}} = G_{\mathrm{Data}}(\frac{A_{\mathrm{MC}}}{G_{\mathrm{MC}}})(\frac{B_{\mathrm{Data}}}{H_{\mathrm{Data}}})(\frac{H_{\mathrm{MC}}}{B_{\mathrm{MC}}})$ and $C_{\mathrm{Est}} = I_{\mathrm{Data}}(\frac{C_{\mathrm{MC}}}{I_{\mathrm{MC}}})(\frac{B_{\mathrm{Data}}}{H_{\mathrm{Data}}})(\frac{H_{\mathrm{MC}}}{B_{\mathrm{MC}}})$. In the study [5] the systematic uncertainty of applying this method in ATLAS was obtained by varying the grid boundaries. It is estimated to be around 15%.

In contrast to this study, CMS chooses to base its DY estimation on only one axis with a dilepton mass window ($76\,\mathrm{GeV}$–$106\,\mathrm{GeV}$). This reduces the dependence on a good jet and MET simulation. The systematic uncertainty of this method is estimated to be

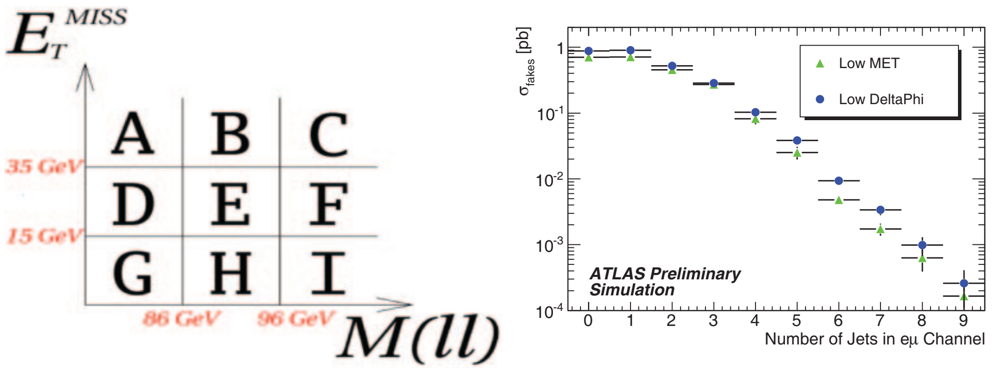

Fig. 5. – Left: diagram of MET *vs.* $M_{l^+l^-}$ regions used for Drell-Yan data-driven background estimates. Sectors A and C mostly contain signal, while the sectors B, G, H, I contain mainly DY events. Right: fake predictions for eμ as a function of the number of jets for ATLAS.

30% [4]. Contributions from QCD multijets and W+jets with fake isolated and prompt leptons are also determined from the data themselves at CMS.

3˙2. *Results*. – Both ATLAS and CMS are determining the $t\bar{t}$ cross-section by using a simple counting method. In addition ATLAS also exploits the usage of a profile likelihood ratio [5] and CMS has developed an analysis requiring b-tagging and using a tighter cut on MET that is designed for higher luminosities (about 100 pb^{-1}).

The cross-section measurement at $\sqrt{s} = 10$ TeV for the ee, eμ, $\mu\mu$ channels combined is obtained with a precision of $\pm 3.1(\text{stat})^{+9.6}_{-8.7}(\text{syst})\%$ and $\pm 15(\text{stat}) \pm 10(\text{syst})\%$ by ATLAS (200 pb^{-1}) and CMS (10 pb^{-1}), respectively.

By scaling the cross-sections to $\sqrt{s} = 7$ TeV, already with 10 pb^{-1} of integrated luminosity, a convincing signal can be collected. Each experiment will have at this point about 30 events with an expected background of about 5. Even with $L_{\text{int}} = 5$ pb^{-1}, the signal will look plausible, with about 15 $t\bar{t}$ and 3 background events. Table II shows the prediction of the number of events in each channel based on the results of ATLAS scaled to 7 TeV and 10 pb^{-1}.

4. – Single top production in the t-channel

The $t\bar{t}$ production is not the only way to produce top quarks. Single top is an interesting alternative, it tests directly the electroweak production mechanism. It is produced via three different production mechanisms that probe different kinematics and

TABLE II. – *Number of events expected in the eμ, ee and $\mu\mu$ channel by ATLAS at* 7 TeV *with an integrated luminosity of* 10 pb^{-1}.

Channel	N(signal) Predicted	N(Background)
ee	14	2.5
eμ	4.3	1.1
$\mu\mu$	6.6	1.9
Total	25	5.5

are sensitive to different beyond the Standard Model scenarios. The t-channel production process is also characterized by a large cross section at the LHC: $\sigma(t\text{-chan}) \sim 1/3\sigma_{tt}$. ATLAS has recently published a new analysis in the t-channel using both a cut based and a likelihood method. The selection is requiring 1 isolated lepton with $p_T > 20\,\text{GeV}$, MET $> 20\,\text{GeV}$, 2 to 4 jets with $p_T > 30\,\text{GeV}$. The requirement that the transverse mass of the reconstructed W is larger than 30 GeV helps against the QCD background. In addition at least one jet needs to be b-tagged. Only events in the 2 jet multiplicity bin are used for the cross-section evaluation. The observables entering the likelihood are angular variables chosen for their insensitivity to jet energy scale systematic uncertainties [6]. The signal-over-background is of 0.64 and 0.89 for the cut based and the likelihood analysis, respectively. The total uncertainty on the cross-section determination at $\sqrt{s} = 10\,\text{TeV}$ with an integrated luminosity of $200\,\text{pb}^{-1}$ is 15(stat) ± 34.7(syst) ± 11(lumi) and 14(stat) ± 32.1(syst) ± 11(lumi), respectively. The largest systematic contribution is due to the b-tagging and amounts to $\sim 26\%$ for the cut method and $\sim 22\%$ for the likelihood method.

This analysis also profits of a data-driven determination of the backgrounds: the idea is to construct a discriminant that allows for a simultaneous determination of the W+jets and $t\bar{t}$ rates which are the main background processes for single top-quark production. The method first selects a control region (the 3-jet sample before applying b-tagging) and then extrapolates the background in the signal region (the 2-jet bin). *A priori* this sample will be composed of approximately 68% W+jets events and 24% $t\bar{t}$ events. Thus, a sufficient rate of both processes is present, which allows for a simultaneous determination of the two rates. A Neural Network (NN) is built using kinematic variables and a simultaneous maximum likelihood fit of the NN outputs for the 2 backgrounds is performed. The overall statistical and systematics uncertainty on the measurement is of ±14.1% and ±6.9% on the W+jets and $t\bar{t}$ backgrounds, respectively.

Similarly, CMS has performed an analysis in the t-channel. The selection is very similar with respect to the ATLAS one, the only differences being a harder b-tag cut, no cut on MET and a harder cut on $M_T(W) > 50\,\text{GeV}$. The analysis is then profiting of the fact that tops are almost 100% left-handed polarised by using polarisation templates to measure the cross-section [7]. The overall uncertainty on the measurement is $\sim 39\%$, including a 35% statistical error contribution. As a perspective for $\sqrt{s} = 7\,\text{TeV}$, both collaborations expect to observe a 3σ excess with $500\,\text{pb}^{-1}$ and to get an observation of the single top in the t-channel with $\sim 1\,\text{fb}^{-1}$.

5. – Early top mass measurement

ATLAS has recently produced new results for the top mass determination at 10 TeV with $200\,\text{pb}^{-1}$ [8]. Two different methods have been developed, a one-dimentional template fit based on minimal event information that exploits the top stabilised mass estimator given by $m_{\text{top}}^{\text{stab}} = \frac{m_{\text{top}}^{\text{reco}}}{m_W^{\text{reco}}} m_W^{\text{PDG}}$, which is sensitive to the changes of m_{top}, but is stable with respect to the jet energy scale (JES). In addition a two-dimentional template fit that simultaneously determines m_{top} and the JES from the data has been exploited. Since this method is requiring 2 b-tags and imposing a kinematic fit to obtain the p_z component of the neutrino by using M_W as a constraint, it requires a better understanding of the detector and is aimed at larger integrated luminosities. Figure 6 shows the results of the 1D and 2D fits on $100\,\text{pb}^{-1}$ and 1 fb^{-1} of data, respectively. The resulting precision of the measurement is of ±2(stat) ± 4(syst)% and ±0.6(stat) ± 2(syst)%, respectively.

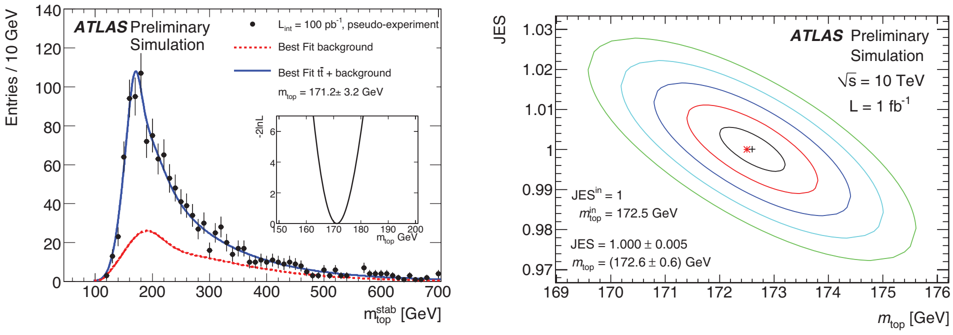

Fig. 6. – Left: 1D fit to the stabilised hadronic top mass in the muon channel by ATLAS. Right: 2D fit of m_{top} and JES in the muon channel, only statistical errors are shown.

6. – Conclusions

We have given an overview of the expectations for the early top quark measurements by the ATLAS and CMS Collaborations at the LHC. Robust and simple analyses have been developed to observe the first $t\bar{t}$ signal with a not yet fully understood detector: data-driven methods have been studied to extract the main backgrounds, namely the QCD, W+jets and Drell-Yann contributions. Interesting results can be expected already with integrated luminosities as low as $5\,\text{pb}^{-1}$ at $\sqrt{s} = 7\,\text{TeV}$, including the observation of the top at LHC.

REFERENCES

[1] CMS Collaboration, CMS PAS TOP-09-003 (2009).
[2] CMS Collaboration, CMS PAS TOP-09-004 (2009).
[3] ATLAS Collaboration, ATL-PHYS-PUB-2009-087 (2009).
[4] CMS Collaboration, CMS PAS TOP-09-002 (2009).
[5] ATLAS Collaboration, ATL-PHYS-PUB-2009-086 (2009).
[6] ATLAS Collaboration, ATL-PHYS-PUB-2010-003 (2010).
[7] CMS Collaboration, CMS PAS TOP-09-005 (2009).
[8] ATLAS Collaboration, CMS PAS TOP-09-005 (2009).