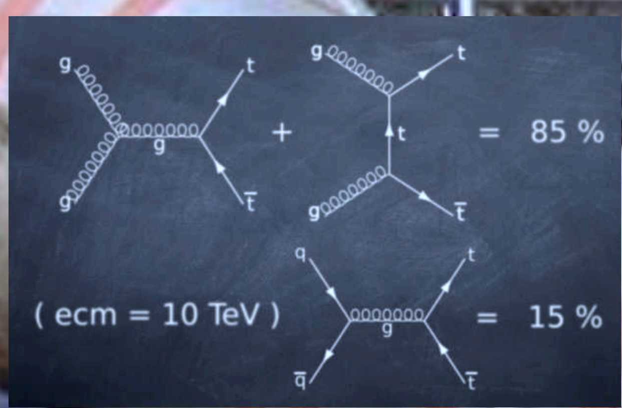
PROSPECTS FOR EARLY TOP PAIR CROSS-SECTION MEASUREMENTS AT ATLAS

Dustin Urbaniec

Columbia University, New York, NY, U.S.A.

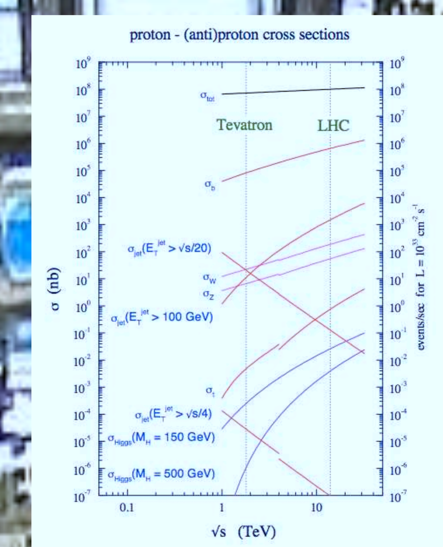
On behalf of the ATLAS Collaboration



INTRODUCTION

After QCD jets, W, and Z-bosons, the production of top quarks will be the dominant process at the LHC. The measurment of the cross-section of top quark pair production is important for many reasons. It will provide an important test of the Standard Model as theoretical predictions are now at the level of around 10%. An abundant $t\bar{t}$ sample will also serve as a useful calibration tool for reconstructed objects such as jets and missing E_T. In addition, $t\bar{t}$ events will be a significant background for many new physics searches. Thus, a well understood $t\bar{t}$ cross-section measurement is essential at the LHC.

At the 10 TeV pp collisions, the dominant production mechanism is gluon-gluon fusion. Top pair events are characterized by the number of W-bosons that decay leptonically. Considered here are the *dilepton channel* [1], where both W bosons decay leptonically, as well as the *single lepton channel* [2], where one W decays leptonically and the other hadronically. All distributions have been simulated at 10 TeV and are normalized to 200 pb^{-1}.

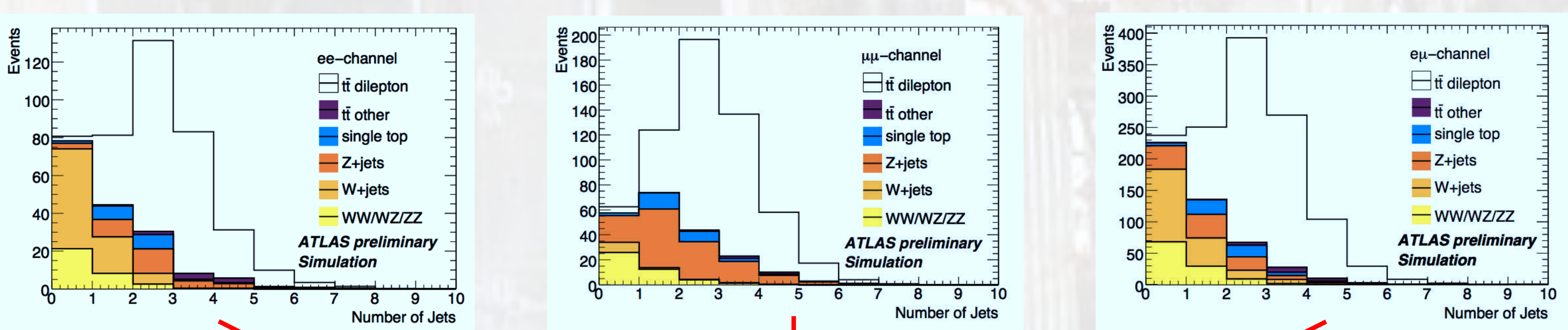


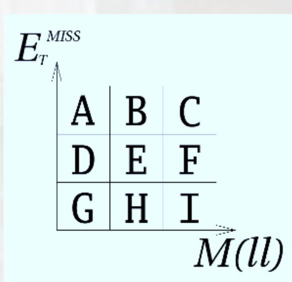
DILEPTON ANALYSIS STRATEGIES

EVENT SELECTION: Dilepton $t\bar{t}$ events are selected by requiring:

- exactly 2 tight leptons (electron or muon) with $p_T > 20$ GeV and $|\eta| > 2.5$
- $\not{E}_T > 35$ GeV for ee and $\mu\mu$ events or $\not{E}_T > 20$ GeV for $e\mu$ events
- at least 2 jets with $p_T > 20$ GeV and $|\eta| > 2.5$
- For the ee and $\mu\mu$ channels, events with a dilepton invariant mass within 5 GeV of the Z mass (91 GeV) are vetoed to minimize the Drell-Yan background

FIGURE 1: JET MULTIPLICITY DISTRIBUTION AFTER ALL CUTS EXCEPT $N_{jets} \geq 2$





BACKGROUND ESTIMATION: To estimate the Drell-Yan contribution to the signal region, the Monte Carlo prediction is scaled to match the observed number of events in the sideband regions of data (high missing E_T and invariant mass far from 91 GeV):

$$A_{Est} = G_{Data}(\frac{A_{MC}}{G_{MC}})(\frac{B_{Data}}{H_{Data}})(\frac{H_{MC}}{B_{MC}}), \quad C_{Est} = I_{Data}(\frac{C_{MC}}{I_{MC}})(\frac{B_{Data}}{H_{Data}})(\frac{H_{MC}}{B_{MC}}) \quad (1)$$

The Drell-Yan background contribution is then estimated as $A_{Est} + C_{Est}$. The systematic uncertainty is obtained by varying the boundaries of the Z mass and missing E_T windows.

Another important background in the dilepton channel comes from fakes in W+jets and QCD. To measure the fake rate f, two samples dominated by fakes will be used. A tag and probe method on Z→ ll events will be used to determine the efficiency ε of the selection on real leptons. A loose (L) and tight (T) lepton selection is defined. For two leptons, the selections relate to the truth in an efficiency matrix:

$$\begin{bmatrix} N_{TT} \\ N_{TL} \\ N_{LT} \end{bmatrix} = \begin{bmatrix} \epsilon_1\epsilon_2 & \epsilon_1 f_2 & f_1\epsilon_2 \\ \epsilon_1(1-\epsilon_2) & \epsilon_1(1-f_2) & f_1(1-\epsilon_2) \\ (1-\epsilon_1)\epsilon_2 & (1-\epsilon_1)f_2 & (1-f_1)\epsilon_2 \end{bmatrix} \begin{bmatrix} N_{RR} \\ N_{RF} \\ N_{FR} \end{bmatrix} \quad (2)$$

where R (F) stands for real (fake). All events reconstructed with two tight leptons that do not come from two real leptons are considered to be fakes:

$$N_{\text{Fake}} = \left[\frac{f_2(\epsilon_2-1)}{\epsilon_2-f_2} + \frac{f_1(\epsilon_1-1)}{\epsilon_1-f_1}\right] N_{TT} + \frac{f_2\epsilon_2}{\epsilon_2-f_2}N_{TL} + \frac{f_1\epsilon_1}{\epsilon_1-f_1}N_{LT} \quad (3)$$

LIKELIHOOD: The cross-section is measured by maximizing the following likelihood function:

$$L(\sigma_{\text{sig}}, \mathcal{L}, \alpha_j) = \prod_{l\in\{ee,\mu\mu,e\mu\}} \left\{ \prod_{i\in\text{bins}} \left[Pois(N_i^{\text{obs}}|N_{i,\text{tot}}^{\text{exp}}) Gaus(\tilde{\mathcal{L}}|\mathcal{L}, \sigma_{\mathcal{L}}) \prod_j Gaus(\tilde{\alpha}_j = 0|\,\alpha_j,\, \Delta_{\alpha_j} = 1) \right] \right\} \quad (4)$$

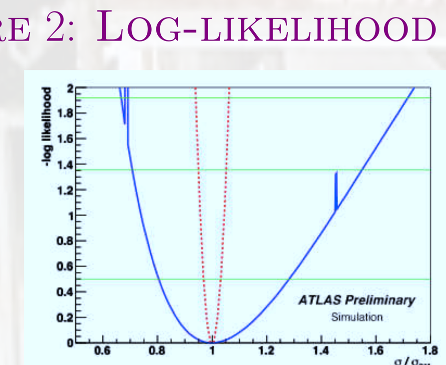
- N_i^{obs} is the number of observed events in the i^{th} jet multiplicity bin and $N_{i,\text{tot}}^{\text{exp}}$ is the number of expected events in that bin
- $\mathcal{L}$ is the luminosity (with $\tilde{\mathcal{L}}$ being the nominal value and $\sigma_{\mathcal{L}}$ the uncertaintiy)
- α_j represent the various systematic uncertainties grouped such that the corresponding variations in the efficiencies ε are expected to be uncorrelated

SYSTEMATICS: The likelihood can be maximized to estimate the parameters $\hat{\sigma}_{\text{sig}}$, $\hat{\mathcal{L}}$, and $\hat{\alpha}_j$. The final systematic uncertainty is estimated from the likelihood profile. The likelihood ratio r and profile λ are defined as:

$$r(\sigma_{\text{sig}}) = \frac{L(\sigma_{\text{sig}}, \hat{\mathcal{L}}, \hat{\alpha}_j)}{L(\hat{\sigma}_{\text{sig}}, \hat{\mathcal{L}}, \hat{\alpha}_j)}, \quad \lambda(\sigma_{\text{sig}}) = \frac{L(\sigma_{\text{sig}}, \hat{\hat{\mathcal{L}}}, \hat{\hat{\alpha}}_j)}{L(\hat{\sigma}_{\text{sig}}, \hat{\mathcal{L}}, \hat{\alpha}_j)} \quad (5)$$

where $\hat{\hat{\mathcal{L}}}$ and $\hat{\hat{\alpha}}_j$ are found from maximizing the likelihood while holding σ_{sig} fixed. The distribution $-2\log\lambda(\sigma_{\text{sig}}^{\text{true}})$ is then used to establish a confidence interval. The figure below shows the log-likelihood curve for all channels combined. The blue curve is the logarithm of the likelihood profile, while the dotted red curve is the logarithm of the likelihood ratio. The table shows the largest contributions to the relative uncertainty (68% confidence level) on the cross-section.

FIGURE 2: LOG-LIKELIHOOD CURVES



$\Delta\sigma/\sigma$ (%)	ee channel	$\mu\mu$ channel	$e\mu$ channel	combined
Stat only	-7.5 / 7.8	-6.0 / 6.2	-4.0 / 4.1	-3.1 / 3.1
Luminosity	-17.3 / 26.3	-17.4 / 26.2	-17.4 / 26.2	-17.4 / 26.2
Electron Efficiency	-4.5 / 5.0	0.0 / 0.0	-2.2 / 2.4	-1.9 / 1.9
Muon Efficiency	0.0 / 0.0	-4.6 / 5.2	-2.1 / 2.2	-2.2 / 2.3
Jet Energy Scale	-3.4 / 3.2	-3.0 / 4.5	-2.5 / 2.5	-2.8 / 3.0
ISR FSR	-4.0 / 4.2	-3.6 / 3.7	-3.5 / 3.5	-3.6 / 3.7
Signal Generator	-4.7 / 5.4	-4.6 / 5.4	-4.7 / 5.3	-4.7 / 5.3
Drell Yan	-1.4 / 1.3	-2.2 / 2.2	-0.5 / 0.5	-0.8 / 0.9
Fake Rate	-9.7 / 9.5	-1.1 / 1.1	-6.2 / 6.2	-4.0 / 4.0
All syst but Lum.	-12.7 / 13.9	-8.9 / 10.2	-9.4 / 10.2	-8.7 / 9.6
All systematics	-21.0 / 30.3	-19.3 / 28.3	-19.5 / 28.5	-19.3 / 28.1
Stat + Syst	-22.3 / 31.3	-20.2 / 29.0	-19.9 / 28.8	-19.5 / 28.3

RESULTS: The prospects for measuring the $t\bar{t}$ cross-section in the dilepton channel at ATLAS have been studied. A simple set of selections was defined and likelihood method to extract the cross-section measurement and its uncertainty has been developed. The expected uncertainties in the dilepton channel for 200 pb^{-1} of 10 TeV data are shown to the right.

$$\Delta\sigma_{ee}/\sigma_{ee} = (^{+7.8}_{-7.5}(\text{stat})^{+13.9}_{-12.7}(\text{syst})^{+27.0}_{-17.8}(\text{lumi}))\% \quad (6)$$

$$\Delta\sigma_{\mu\mu}/\sigma_{\mu\mu} = (^{+6.2}_{-6.0}(\text{stat})^{+10.3}_{-8.8}(\text{syst})^{+27.0}_{-17.9}(\text{lumi}))\% \quad (7)$$

$$\Delta\sigma_{e\mu}/\sigma_{e\mu} = (^{+4.1}_{-4.0}(\text{stat})^{+10.2}_{-9.4}(\text{syst})^{+27.1}_{-18.0}(\text{lumi}))\% \quad (8)$$

SINGLE LEPTON ANALYSIS STRATEGIES

EVENT SELECTION AND RECONSTRUCTION: Single lepton $t\bar{t}$ events are selected by requiring:

- exactly 1 tight lepton (electron or muon) with $p_T > 20$ GeV and $|\eta| > 2.5$
- $\not{E}_T > 20$ GeV (Not performed in the HT2 analysis. See below)
- at least 4 jets with $|\eta| > 2.5$ and $p_T > 20$ GeV, 3 of which with $p_T > 40$ GeV
- The hadronic top quark candidate is defined as the 3 jet combination whose total p_T is the greatest among all 3 jet combinations

There are some variations in the analysis at this point. One possibility is an M_W-cut by requiring the invariant mass of at least one of the three 2 jet combinations to be within 10 GeV of the W mass. Another possibility is the HT2 analysis described below. The arrows show the analysis paths that have been studied.

DEFAULT SELECTIONS: The top plot shows the reconstructed top mass without the M_W-cut and the bottom is with the cut applied.

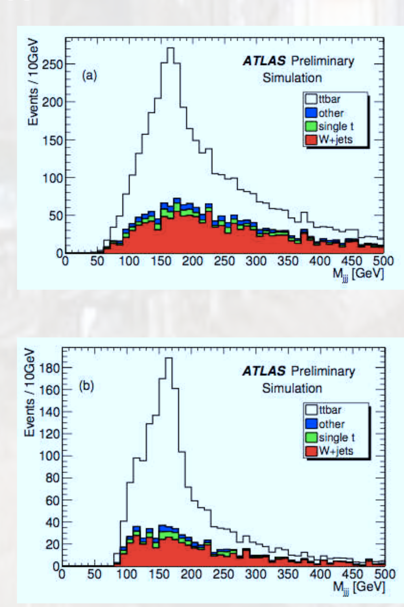


FIGURE 3: TOP MASS

HT2 ANALYSIS: It is possible that an accurate understanding of the missing E_T measurement will require substantial commissioning. One variation in the selections is to perform the measurement without a missing E_T cut in favor of a cut on HT2 (scalar sum of the p_T of the lepton and the 2nd, 3rd, and 4th jet). This is motivated by the fact that in QCD, the energy tends to be dominated by the two leading jets, in contrast to $t\bar{t}$ events which share the energy more equally. The lepton p_T and η cuts are also tightened.

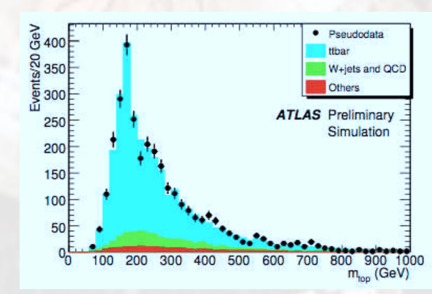


FIGURE 4: TOP MASS FOR HT2 ANALYSIS

CUT AND COUNT: The total number of events that pass the event selection are counted and all expected backgrounds are subtracted:

$$\sigma = \frac{N_{\text{sig}}}{\mathcal{L}\times\varepsilon} = \frac{N_{\text{obs}} - N_{\text{bkg}}}{\mathcal{L}\times\varepsilon} \quad (9)$$

Where $\mathcal{L}$ is the luminosity and ε is the signal efficiency. One of the most important backgrounds in this analysis is W-boson production in association with jets. The Monte Carlo predictions for the fraction of W events which are produced with 4 or more jets have a large uncertainty. Since the W to Z ratio uncertainty is smaller, we can estimate the W contribution to the signal region by measuring the ratio of W to Z events in a control region in data and extrapolating to the signal region:

$$(\frac{\text{W}^{\text{SR}}}{\text{W}^{\text{CR}}})_{\text{data}} = (\frac{\text{Z}^{\text{SR}}}{\text{Z}^{\text{CR}}})_{\text{data}} \cdot C_{\text{MC}}, \quad C_{\text{MC}} = \frac{(\text{W}^{\text{SR}}/\text{W}^{\text{CR}})_{\text{MC}}}{(\text{Z}^{\text{SR}}/\text{Z}^{\text{CR}})_{\text{MC}}} \quad (10)$$

TOP MASS FIT: Another way to extract the cross-section is to do a fit to the 3 jet invariant mass to determine the signal size. After the M_W-cut, a binned maximum likelihood fit is performed with a Gaussian for the 3 jet invariant mass and a 6th order Chebyshev polynomial for the background. The integral of the fitted Gaussian is then used as an estimator for the number of correctly reconstructed $t\bar{t}$ events.

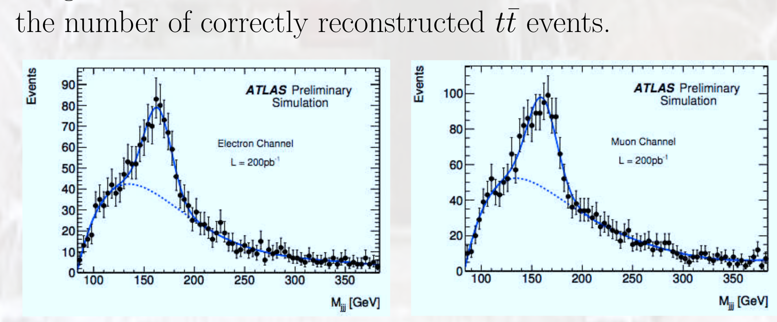


FIGURE 5: FITS TO m_{jjj}

SYSTEMATICS: The dominant expected systematic uncertainties in the each method are shown in the chart below.

Source	Cut and Count method				Fit method		HT2 Analysis			
	e-analysis		μ-analysis		e-analysis	μ-analysis	e-analysis		μ-analysis	
	default	+M$_W$ cut	default	+M$_W$ cut	+M$_W$ cut	+M$_W$ cut	Counting	Fit	Counting	Fit
Stat.	±2.5	±3.4	±2.3	±3.1	±14.1	±15.2	±3.2	±5.7	±3.0	±5.6
50% W+jets	±25.1	±17.4	±28.1	±19.8	±3.3	±5.6	±22	-	± 21	-
20% W+jets	±10.0	**±7.0**	±11.2	**±7.9**	**± 1.5**	**± 2.6**	**±9**	-	**±8**	-
100% QCD	-	-	-	-	-	-	±8	-	±1	-
JES (10%,-10%)	+24.8-23.4	+15.9-19.1	+20.5-22.3	+11.9-17.9	-14.4	-15.4	+15-23	-21-14	+23-24	-27-18
JES (5%,-5%)	+12.3-11.9	**+8.6-9.3**	+10.4-10.9	**+6.1-8.4**	**-3.7**	**-3.9**	**+8-11**	**-11-4**	**±12**	**-12-7**
ISR/FSR	±9.1	**+7.6-8.2**	±8.2	**+5.2-8.3**	**-12.9**	**-12.9**	**±8**	**±7**	**±13**	**±5**
Fitting Model	-	-	-	-	**±3.3**	**±4.7**	-	**±2.9**	-	**±2.3**
10% Lum.	±11.6	±11.2	±11.4	±11.1	±10	±10	±13	±10	±12	±10
20% Lum.	±23.2	**±22.3**	±22.8	**±22.2**	**±20**	**±20**	±26	±20	±24	±20
Tot. without Lum.	+18.8-18.5	**+14.4-15.2**	+17.5-17.7	**+11.9-14.7**	**+6.4-14.9**	**+6.0-14.7**	**+19.2-20.6**	**+10.5-15.2**	**+20.2-20.2**	**+9.4-25.3**

RESULTS: Several simple analyses have been developed and studied. The choice of analysis path can best be determined when data arrives. The expected uncertainties on the $t\bar{t}$ cross-section measurement in the single lepton channel for the default selections in the cut and count method are shown to the right. These results are derived for 200 pb^{-1} of data produced with 10 TeV collisions.

$$\frac{\Delta\sigma_e}{\sigma_e} = (3(\text{stat})^{+14}_{-15}(\text{syst}) \pm 22(\text{lumi}))\% \quad (11)$$

$$\frac{\Delta\sigma_\mu}{\sigma_\mu} = (3(\text{stat})^{+12}_{-15}(\text{syst}) \pm 22(\text{lumi}))\% \quad (12)$$

References

[1] The ATLAS Collaboration. Prospects for measuring top pair production in the dilepton channel with early ATLAS data at $\sqrt{s} = 10$ TeV. *ATL-PHYS-PUB-2009-086*, 2009.

[2] The ATLAS Collaboration. Prospects for the top pair production cross-section at $\sqrt{s} = 10$ TeV in the single lepton channel in ATLAS. *ATL-PHYS-PUB-2009-087*, 2009.