Status and Prospects for observation of top quark pair production in ATLAS

Borut Paul Kerševan, Ljubljana Group

Faculty of Mathematics and Physics
University of Ljubljana
Eksperimental Particle Phys. Dept.
Jozef Stefan Institute

On behalf of the ATLAS Collaboration

QCD 10, Montpellier 2010
The LHC Operation in 2010/2011

- High energy proton collisions at 7 TeV are ongoing.
- Target $L_{\text{int}} \sim 1 \text{ fb}^{-1}$ by end of 2011.

LHC program for 2010-11 focusses on:

- Detector commisioning.
- Standard Model Physics.
- Physics beyond TeV scale:
  - Possible surprises (Susy, Higgs, Exotics..)
ATLAS Performance so far

- Expect a few pb\(^{-1}\) by the end of summer 2010

Cumulative integrated luminosity versus day delivered to (green), and recorded by ATLAS (yellow) during stable beams and for 7 TeV centre-of-mass energy. Given is the luminosity as determined online using counting rates measured by the luminosity detectors, and by assuming a total inelastic cross section of 71.5 mb.
Roadmap to Top quark Re-discovery

- We are heavily involved in understanding our detector and basic reconstruction objects:
  - leptons, jets, b-tagging, Missing transverse energy (MET)
- understanding QCD processes
- First steps towards Z and W + jets observation.
- Need to understand the above items when proceeding to top quark re-discovery.
Top Quark Physics @ ATLAS

- Top physics @ ATLAS:
  - **First:**
    - Top re-discovery: top-antitop production
    - Detector calibrations
      Jet energy scale (JES)
      b-tagging efficiency
  - **Next:**
    - Precision tests of SM
    - Search beyond SM:
      top-antitop resonances
      anomalous couplings
      non-SM top production

QCD: top-antitop production:
1 fb\(^{-1}\) @ 7 TeV: 160k tt pairs produced
Top re-discovery process

\[ \sigma_{tt} = 160 \text{ pb} @ 7 \text{ TeV} \sim 20 \times \text{Tevatron} \]

EW: single top production:
Sensitive to beyond SM physics

Single top (t-channel)
\[ \sigma_t \sim 1/3 \sigma_{tt} \]
Top Pair Decay Channels

- **Top-antitop decay modes:**
  - **Di-leptonic channel:**
    - low statistics (9%)
    - clean signature S/B 4.5-6.5
  - **Semi-leptonic channel (e/µ):**
    - 45% of total
    - S/B ~ 1 (without b-tag)
    - Visible top and/or W invariant mass peaks helps
  - **Fully hadronic channel:**
    - 46% of total:
    - high QCD backgrounds.
    - Not easy to trigger on (no leptons, need b-tag trigger?)
    - Not for early data!
The first top measurement to be done with early data is the $t\bar{t}$ cross-section determination.

- **Dilepton channel**
  - 2 opp. charged leptons
  - 2 jets or more
  - Missing $E_T$
- **Backgrounds:**
  - 2 real leptons
    - Drell-Yan
    - Dibosons
    - Single-top ($Wt$)
  - fake leptons
    - $W+$jets, QCD multijet [$>>1000\times\sigma_{t\bar{t}}$]
- **Strategy:**
  - Veto Z window and require MET

- **Lepton+jets channel**
  - 1 lepton
  - 4 jets or more
  - Missing $E_T$
- **Backgrounds:**
  - combinatorial
  - $W+$jets [$\sim 40\times\sigma_{t\bar{t}}$]
  - single-top
  - fake leptons from QCD multijet [$>>1000\times\sigma_{t\bar{t}}$]
- **Strategy:**
  - reconstruct hadronic top
  - cut or fit methods
Re-discovery Methods

- **Simple methods first:**
  - "Cut and count" (C&C): \( \sigma_{\text{signal}} = \frac{(N_{\text{data}} - N_{\text{Background}})}{\text{eff.} \times \text{acc.} \times L} \)
  - Template fit to a discriminating variable \( \rightarrow \sigma_{\text{signal}} \)

- **Cautious with event selection:**
  - Very simple/safe definition of measured objects.
  - Missing \( E_T \) poorly known at startup.
  - Consider b-tag but use cautiously: Commissioning of b-tagging algorithms will take some time..

- **Data-driven background measurements:**
  - Do not rely (too much) on simulation:
    - Theoretical uncertainties (QCD-multijet, W+jets about 80-100% ...)
    - Acceptance uncertainties (lepton fake rate, b-tagging efficiency ...)
  - Preferable to measure (main) backgrounds using data itself.
A note on b-tagging

- b-tagging commissioning is ongoing:
  - not envisaged in the first top re-discovery analyses presented here.
  - as soon as the commissioning is completed it can be included: strongly suppresses the W+jets, QCD and Drell-Yan backgrounds.
ATLAS Lepton+jets @10 TeV

- Basic event selection:
  - high pT isolated lepton
  - 4 high pT jets, 3 of them > 40 GeV

- In addition:
  - Missing $E_T > 20$GeV
  - Combine 3 jets to form hadronic Top (selection that maximizes $p_T$ top)
  - 2 jets satisfy W mass constraint $(m_W \pm 10 \text{ GeV}) \varepsilon=10\%$.

<table>
<thead>
<tr>
<th>200 pb-1 @ 10 TeV</th>
<th>Signal: el. channel</th>
<th>backgrounds</th>
<th>S/B</th>
</tr>
</thead>
<tbody>
<tr>
<td>After selection</td>
<td>1286</td>
<td>598</td>
<td>2.1</td>
</tr>
</tbody>
</table>

ATL-PHYS-PUB-2009-087
Atlas W+jets Data Driven Estimate

- **Event selection:**
  - require 2 leptons $80 < m_{ll} < 100$ GeV for Z+jets and 1 lepton +MET for W+jets

- **Use extrapolation:**
  - low jet multiplicity region as control region (CR) for Z+jets (W+jets) events:
  - Extrapolate ratio in the top signal region (SR): 4 jets or more
  - Use ratio $W/Z$ to extrapolate $W$ into signal region:
    \[
    \left(\frac{W_{SR}}{W_{CR}}\right)_{data} = \left(\frac{Z_{SR}}{Z_{CR}}\right)_{data} \cdot C_{MC}
    \]
    \[
    C_{MC} = \frac{\left(\frac{W_{SR}}{W_{CR}}\right)_{MC}}{\left(\frac{Z_{SR}}{Z_{CR}}\right)_{MC}}
    \]

- **Overall uncertainty:**
  - 20 % @ 10 TeV with 200 pb$^{-1}$
  - 50 % @ 7 TeV with 10 pb$^{-1}$
  - 20 % @ 7 TeV with 100 pb$^{-1}$

- **Largest contributions:**
  - 50 % uncertainty on QCD
  - $C_{MC}: \sim 12 \%$ Pythia - Alpgen discrepancy
Two Main Methods to Extract cross-section

- **Cut and count method (C&C):**
  - **Advantage:** simple method
  - **Disadvantage:** Depends on background subtraction. Evaluate backgrounds from data!

- **The fit or template fit to hadronic top mass:**
  - **Advantage:** much less sensitive to the Jet energy scale and background normalisation
  - **Disadvantage:** relies on shape of distributions

<table>
<thead>
<tr>
<th>Method</th>
<th>$\Delta\sigma/\sigma$</th>
<th>Dominant systematics</th>
</tr>
</thead>
<tbody>
<tr>
<td>C&amp;C</td>
<td>$3.4^{+18.2}_{-21.1}$ (stat) + 29.3(lumi)%</td>
<td>JES, ISR/FSR, W+jets</td>
</tr>
<tr>
<td>Fit</td>
<td>$14^{+6}_{-15}$ (stat) + 20(lumi)%</td>
<td>JES, ISR/FSR</td>
</tr>
<tr>
<td>No MET</td>
<td>$3.4^{+23}_{-25}$ (stat) + 34(lumi)%</td>
<td>JES, ISR/FSR, W+jets</td>
</tr>
</tbody>
</table>

200 pb$^{-1}$ @ 10 TeV, electron channel
Top Re-discovery at 7 TeV - lepton + jets

- With ~10 pb-1 notable signal:
  - ATLAS will have ~60 signal events with an expected background of ~40 events per lepton flavor.
  - Results before and after the additional $M_W$ cut are shown:
    2 jets satisfying $W$ mass constraint ($m_W \pm 10$ GeV)
  - Note that in addition b-tagging might be used:
    S/B enhanced up to ~5 times!

<table>
<thead>
<tr>
<th>Channel</th>
<th>el. analysis</th>
<th>$\mu$ analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t\bar{t}$~</td>
<td>53</td>
<td>64</td>
</tr>
<tr>
<td>W+jets</td>
<td>29</td>
<td>40</td>
</tr>
<tr>
<td>single top</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>Other</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

**Expected sensitivity for 10 pb$^{-1}$ @7 TeV**
Scaled down from ATL-PHYS-PUB-2009-087 results

**Cross-section scaling:**

$$\sigma(t\bar{t})_7 \approx 40\% \sigma(t\bar{t})_{10}$$

$$\sigma(W + \text{jets})_7 \approx 45\% \sigma(W + \text{jets})_{10}$$

**Additional $M_W$ cut: $M_{jj}$ in $M_W \pm 10$ GeV**
ATLAS di-lepton Analysis

200 pb^{-1} @10 TeV

- Basic event selection:
  - 2 high pT (>20 GeV) isolated leptons
  - ≥2 jets pT >20 GeV
  - MET>20 (e\mu)/35 (ee/\mu\mu) GeV
  - veto Z mass window

Cross-section estimation:
- Simple cut and count method
- Profile likelihood ratio
- Final sensitivity (e\mu,ee,\mu\mu combined):

\[ \frac{\Delta \sigma}{\sigma} = 3.1(\text{stat})^{+9.6}_{-8.7}(\text{syst})^{+26.2}_{-17.4}(\text{lumi})\%

Main syst: signal model, fake rate, Luminosity not included.
Drell-Yan and Fake Leptons

• **Drell-Yan estimate from data:** Drell-Yan mainly in G,H,I regions - top-antitop in A,C
  
  ▶ Use Z Monte-Carlo and data (same for $C_{Est}$)

  \[
  A_{Est} = \frac{G_{Data}}{G_{MC}} \left( \frac{B_{Data}}{H_{Data}} \right) \left( \frac{H_{MC}}{B_{MC}} \right)
  \]

• **Systematic uncertainty (stat. dependent!):**
  
  ▶ Shift boundaries of the grid:
  
  ~15% for both channels @ 10 TeV and 200 pb$^{-1}$

• **Fake leptons in QCD jets:** from W+jets (1 fake lepton) and QCD (2 fake leptons):
  
  ▶ **muons:**
    
    • Decays in flight of π or K mesons in light jets
    • Punch through muons
    • Semileptonic decay in a heavy flavor (b/c) jet
  
  ▶ **electrons:**
    
    • Jets with high EM fraction
    • Photon mis-associated with a track, conversions
    • Semileptonic decay in a heavy flavor (b/c) jet
Fake leptons in detail

- **Fake leptons from QCD jets:** from W+jets (1 fake lepton) and QCD (2 fake leptons)

- Define a “tight(T)” and a “loose(L)” lepton, the latter enriched with fake leptons.

- Efficiencies to select a “tight(T)” lepton for real (R) $\varepsilon$ and fake(F) $f$ leptons from data:
  
  - $\varepsilon_{1|2} = \frac{N_{T,R}}{N_{T,R} + N_{L,R}}$: Tag-and-probe on two leptons in Z mass window, low MET required.
  
  - $f_{1|2} = \frac{N_{T,F}}{N_{T,F} + N_{L,F}}$: one (tight or loose) lepton and low MET required.
  
  - $\varepsilon_{1|2}$ and $f_{1|2}$ for first|second lepton can exhibit $p_T$ or $\eta$ dependence.

- Neglecting the possibility of two fake leptons being reconstructed as tight leptons:
  
  $$
  \begin{bmatrix}
  N_{TT} \\
  N_{TL} \\
  N_{LT}
  \end{bmatrix}
  =
  \begin{bmatrix}
  \varepsilon_1 \varepsilon_2 & \varepsilon_1 f_2 & f_1 \varepsilon_2 \\
  \varepsilon_1 (1-\varepsilon_2) & \varepsilon_1 (1-f_2) & f_1 (1-\varepsilon_2) \\
  (1-\varepsilon_1)\varepsilon_2 & (1-\varepsilon_1)f_2 & (1-f_1)\varepsilon_2
  \end{bmatrix}
  \begin{bmatrix}
  N_{RR} \\
  N_{RF} \\
  N_{FR}
  \end{bmatrix}
  $$

- Inverting and defining as fake all events with two tight leptons not coming from 2 real ones:
  
  $$
  N_{Fake} = \left[ \frac{f_2(\varepsilon_2 - 1)}{\varepsilon_2 - f_2} + \frac{f_1(\varepsilon_1 - 1)}{\varepsilon_1 - f_1} \right] N_{TT} + \frac{f_2\varepsilon_2}{\varepsilon_2 - f_2} N_{TL} + \frac{f_1\varepsilon_1}{\varepsilon_1 - f_1} N_{LT}
  $$

- Rate uncertainty on MC 50-100% (need data for estimation).
Top Re-discovery at 7 TeV - dilepton

- With ~10 pb-1 convincing signal
  - ATLAS will have ~30 events with an expected background of 5 or 6.
  - Even with 5 pb-1, signal already plausible.
  - At 1 pb-1, interesting event displays will start to appear at conferences
  - Note that in addition b-tagging might be used.

Expected sensitivity for 10 pb⁻¹ @7 TeV

Scaled down from ATL-PHYS-PUB-2009-086 results

<table>
<thead>
<tr>
<th>Channel</th>
<th>N(signal)</th>
<th>N(background)</th>
</tr>
</thead>
<tbody>
<tr>
<td>e-μ</td>
<td>14</td>
<td>2.5</td>
</tr>
<tr>
<td>e-e</td>
<td>4.3</td>
<td>1.1</td>
</tr>
<tr>
<td>μ-μ</td>
<td>6.6</td>
<td>1.9</td>
</tr>
<tr>
<td>Total</td>
<td>25</td>
<td>5.5</td>
</tr>
</tbody>
</table>

Cross-section scaling:

\[ \sigma(t\bar{t})_7 \approx 40\% \sigma(t\bar{t})_{10} \quad \sigma(W + jets)_7 \approx 45\% \sigma(W + jets)_{10} \]
Conclusions

- Signatures of top pair signal will come fast
  - Significant signal for “rediscovery” will require $O(10\text{pb}^{-1})$
- By the end of 2010 LHC might have statistics comparable to TeVatron.
- By the end of 2011 the samples will be significantly larger.
  - Precision Standard Model physics: top quark properties.
  - Hints of new Physics?
- Still a lot of work ahead:
  - Trigger and lepton ID
  - b-tagging
  - Light jet and b jet energy scale
  - Missing $E_T$
  - Estimate backgrounds from data: QCD, W/Z + jets.