Outline

- Introduction of QCD Multijet Background in Top Physics
- Ways to Reduce QCD Multijet background
- Methods of QCD Multijet Background Estimation
- QCD Multijet Estimation in Top Public Results
- Conclusion
Introduction
Top Events and QCD Multijet Backgrounds

- Top events signatures comparing to QCD multijet backgrounds:
  - isolated lepton with large transverse momentum from W decay
  - large missing transverse momentum (MET) if W decays leptonically
  - reconstructed W and top mass spectrum ($m_W, m_T$) from final state particles

Examples:

- Top pair process
- Single top process
- QCD-multijet process
How QCD multijet events become background?

- QCD multijet events with misidentified and non-prompt leptons (collectively called fake leptons) could pass top event selection
  - also referred as fake lepton background in many cases
  - negligible after dilepton top event selections, but fake lepton backgrounds could also come from W/Z+jets, tt l+jets events

sources of fake leptons

**fake electrons**
- b-quarks and c-quarks decay semileptonically
- jets misidentified as electrons
  - few charged tracks
  - little energy in hadronic compartments
- photon conversion

**fake muons**
- b-quarks and c-quarks decay semileptonically
- pions or kaons which decay in flight within tracking region
- punch-through hadrons
Ways to Reduce QCD Multijet Background

- Good lepton identification and isolation
- Reasonable large MET requirement
- Selections on $m_W$, $m_T$ or so-called triangular cut
  - triangular cut: cut on transverse mass of lepton and MET system ($MTW + MET$).

Example of electron+jets selection in top pair analysis:
- exactly one electron with $p_T>20\text{GeV}$
- $\text{Etcone}20<4\text{GeV}$
- MET $>35\text{GeV}$
- $MTW>25\text{GeV}$
- $\geq 3\text{jets with } p_T>25\text{GeV}, |\eta|<2.5$

use $MTW + MET > 60\text{GeV}$ to reduce multijet background

1. $\text{Etcone}20$ is the transverse momentum deposit in cone $\Delta R<0.2$ around electron position, removing electron energy leakage
Top and QCD Multijet Cross Sections

- Top pair cross section for proton–proton collisions at a $\sqrt{s}=7(8)$ TeV is $167^{1(2382)}\text{pb}$ for $m_T=172.5\text{GeV}$
- QCD multijet cross section ($\sim\mu\text{b/mb}$) is many orders of magnitude higher than top cross section
  - high QCD multijet rejection factors from top event selections
  - still usually contributes as one of the most important non-top backgrounds after V+jets
- Good estimation of QCD multijet background is important for precision measurements and for searches in top analysis

Methods of QCD Multijet Background Estimation
Estimate QCD multijet from Monte Carlo or data?

- Well modeled fake lepton events from MC could be possible, but
  - not all analysis have fake lepton MC samples available
  - hard to generate enough statistics in the tails of QCD multijet production
  - much easier in dileptons, where the fakes are coming from EW and tt processes that are very well known and available with high statistics
    - example: ATLAS-CONF-2013-097

- Data-driven QCD multijet estimation are a lot cheaper and widely used
  - available as long as data is ready

- Two major approaches:
  - matrix method
  - template fit method where QCD multijet templates from
    - jet-electron model
    - anti-electron/muon selection
Matrix Method

- **real leptons**: leptons from W (Z) decay
- **tight leptons**: standard selection in analysis
- **loose leptons**: looser identification and/or no isolation requirement

Solve Equation to Obtain $N_{\text{fake}}^{\text{tight}}$

\[
N_{\text{loose}}^{\text{tight}} = N_{\text{loose}}^{\text{real}} + N_{\text{fake}}^{\text{real}}
\]
\[
N_{\text{tight}}^{\text{fake}} = \frac{\varepsilon_{\text{fake}}}{\varepsilon_{\text{real}} - \varepsilon_{\text{fake}}} (N_{\text{loose}}^{\text{real}} - N_{\text{tight}}^{\text{real}})
\]

- $N_{\text{loose}}^{\text{real}}$: Number of events with one loose lepton
- $N_{\text{tight}}^{\text{real}}$: Number of events with one tight lepton
- Using data to obtain $\varepsilon_{\text{real}} = \frac{N_{\text{tight}}^{\text{real}}}{N_{\text{loose}}^{\text{real}}}$ and $\varepsilon_{\text{fake}} = \frac{N_{\text{fake}}^{\text{real}}}{N_{\text{fake}}^{\text{real}}}$.
- $N_{\text{fake}}^{\text{tight}}$: Estimated multijet events after tight selection

basic form in l+jets – extension to 4x4 matrix in dilepton
Matrix Method: $\varepsilon_{\text{real}}$

- Select control sample: $Z \rightarrow ll$
  - Two Loose leptons with opposite signs
  - $m_{ll} \sim m_Z$

Tag and probe method:
- One lepton must be "tight"
- Calculate the other lepton tight rate: $\varepsilon_{\text{real}}$

- $\varepsilon_{\text{real}}$ obtained with leptons from Z decay could be different from the ones from top
  - use corrections from $\varepsilon_{\text{top}}/\varepsilon_Z$ using MC samples
**Matrix Method: $\varepsilon_{\text{fake}}$**

- $\varepsilon_{\text{fake}}$ is measured from data control regions dominated by the contributions of fake leptons
  - example selections/combinations: low MET, low $m_W$ or $m_T$, high $d_0$ significance
  - subtract $W$+jets and $Z$+jets backgrounds using Monte Carlo simulation

---

**QCD multijet as a function of $m_T$**

**Major event selection**
- 1 isolated lepton with $p_T>20\text{GeV}$
- $\geq 4$ jets with $p_T>20\text{GeV}$, $|\eta|<2.5$
- $\text{MET}>20\text{GeV}$

---

**ATLAS-CONF-2010-087**
Parameterization and Reweighting

- QCD multijet background shape as a function of a given variable
  - obtained by applying the matrix method in bins of the variable
  - parameterized $\varepsilon_{\text{real}}$ and $\varepsilon_{\text{fake}}$ -> parameterized $N_{\text{tight}}$

- Reweight data to QCD multijet background model
  - re-write $N_{\text{tight}}$ as a sum of all loose events:

$$N_{\text{fake}}^{\text{tight}} = \sum_{\text{loose}} \frac{\varepsilon_{\text{fake}}}{\varepsilon_{\text{real}} - \varepsilon_{\text{fake}}} (\varepsilon_{\text{real}} - w_i^{\text{tight}}) = \sum_{\text{loose}} W_i$$

if event is tight $w_i^{\text{tight}} = 1$, else $w_i^{\text{tight}} = 0$
then if event is tight $W_i = \frac{\varepsilon_{\text{fake}} (\varepsilon_{\text{real}} - 1)}{\varepsilon_{\text{real}} - \varepsilon_{\text{fake}}}$
if event is loose but not tight $W_i = \frac{\varepsilon_{\text{fake}} \varepsilon_{\text{real}}}{\varepsilon_{\text{real}} - \varepsilon_{\text{fake}}}$

- loose data sample reweighted with $W_i$ $\rightarrow$ QCD multijet background
Systematics in Matrix Method

- Uncertainties of $\varepsilon_{\text{real}}$ estimation could come from:
  - statistical uncertainties
  - differences between $\varepsilon_{\text{real}}$ in Z and top events
  - uncertainties on the selection of events for Tag&Probe method

- Systematics of $\varepsilon_{\text{fake}}$:
  - used $W+jets$ and $Z+jets$ backgrounds uncertainties
  - differences between the fake sample compositions
  - differences between parametrizations
Template Fit Method

- Get QCD background template from data-driven method
  - jet-electron model
  - anti-lepton selection
- Get other processes templates from MC
  - top events, V+jets backgrounds, diboson backgrounds
- Fit templates to data
  - fit with sensitive variable in control region, extrapolate QCD estimation in signal region
  - or directly get QCD normalization from final measurement fit
Template Fit: Jet-electron Model

- Idea: use data samples with “electron-like” jet to simulate QCD multijet with fake leptons
  - the origin of the method from CDF
Template Fit: Anti-lepton Model

- Get QCD multijet background template from a QCD multijet enriched data sample where leptons failing some of the lepton cuts
  - reverse lepton identification (ID)/isolation

- Similar fit idea as with jet-electron model

an example from CDF
PRD 84, 031101 (R) (2011)

events with lepton candidate failing any two identification cuts are used to model QCD background
Systematics of Template Fit Method

- Cross check the fit with another variable, like the transverse $W$ mass.
- Vary the QCD multijet templates with different selections of control region.
- Evaluate the effects from pile-up.
  - divide the jet-electron data sample into a high pile-up sample and a low pile-up sample.
    - usually based on the number of primary vertices in events.
  - compare estimated QCD multijet of two samples.
QCD Multijet Estimation in Top Public Results
One Example of QCD Multijet Estimation in ATLAS

- Measurements of top quark pair relative differential cross-sections with ATLAS
  - used 2.05 fb^{-1} data collected by ATLAS at \( \sqrt{s} = 7 \text{TeV} \)

- Used matrix method for QCD multijet background estimation
  - estimated QCD multijet fraction\(^1\) is ~4.4%

\[\begin{array}{c|cc}
\text{Channel} & \mu + \text{jets} & e + \text{jets} \\
\hline
\bar{t}t & 11100 \pm 700 & 7400 \pm 500 \\
W + \text{jets} & 1700 \pm 700 & 1300 \pm 500 \\
\text{Single top} & 490 \pm 50 & 338 \pm 32 \\
Z + \text{jets} & 192 \pm 20 & 154 \pm 26 \\
\text{Diboson} & 34 \pm 4 & 21 \pm 3 \\
\text{Fake-leptons} & 800 \pm 800 & 250 \pm 250 \\
\text{Signal+bkg} & 14400 \pm 1700 & 9500 \pm 1100 \\
\text{Observed} & 14416 & 9187 \\
\end{array}\]

- muon+jets channel:
  - loose sample: discarding the isolation requirements
  - control region for \( \epsilon_{\text{fake}} \cdot \text{MTW} < 20 \text{GeV} \), with an additional requirement MET+MTW < 60 GeV
  - parameterization: parameterized as a function of muon \(|\eta|\) and of the leading jet pT

- electron+jets channel:
  - loose sample: looser identification criteria
  - control region for \( \epsilon_{\text{fake}} \cdot 5 \text{GeV} < \text{MET} < 20 \text{GeV} \)
  - parameterization: parameterized as a function of electron \(|\eta|\)

- 100% uncertainty on QCD background estimation
  - varying the control region requirements (MTW, MET)
  - the statistical uncertainty and MC background corrections

\(^1\) QCD fraction is expected \(N_{\text{QCD}}/N_{\text{signal+background}}\)
One Example of QCD Multijet Estimation in CMS

- Measurement of the $t\bar{t}$-bar production cross section in pp collisions at $\sqrt{s} = 7$ TeV with lepton + jets final states
  - used 2.3 fb$^{-1}$ data collected by CMS at $\sqrt{s}=7$TeV
  - cross section measurement is performed using a maximum profile likelihood fit
    - to the number of reconstructed jets ($N_{\text{jet}}$), the number of $b$-tagged jets ($N_{\text{tag}}$), and the secondary vertex mass (SVM) distribution in the data

- The QCD multijet estimation is obtained by template fit method (fitting MET distribution in data)
  - electron + jets channel: QCD distribution from MC
  - muon + jets channel: QCD distribution from muon data sample failing muon isolation requirement in region $\text{MET}<20\text{GeV}$
    - normalization enters as initial number in the final fit with 100% uncertainty

- Use alternative QCD shapes with different selection for shape systematics
  - different lepton ID/isolation/MET selection

The combined fit for the electron + jets channel, for $\geq 2$-b-tag events
### QCD multijet Estimation Methods in ATLAS/CMS/D0/CDF

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Channel</th>
<th>Goal</th>
<th>Luminosity (fb⁻¹)</th>
<th>Reference</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMS</td>
<td>lepton+jets dilepton</td>
<td></td>
<td></td>
<td>CMS-PAS-TOP-12-030</td>
<td>anti-lepton</td>
</tr>
<tr>
<td>CMS</td>
<td>dilepton</td>
<td>σ(ttbar)</td>
<td>2.3</td>
<td>arXiv:1208.2671</td>
<td>matrix method</td>
</tr>
<tr>
<td>D0</td>
<td>lepton+jets</td>
<td>σ(ttbar)</td>
<td>0.9</td>
<td>PRL 100, 192004 (2008)</td>
<td>matrix method</td>
</tr>
<tr>
<td>D0</td>
<td>dilepton</td>
<td>σ(ttbar), mₜ</td>
<td>1.0</td>
<td>PLB 679, 177 (2009)</td>
<td>anti-lepton</td>
</tr>
<tr>
<td>CDF</td>
<td>lepton+jets</td>
<td>σ(ttbar)</td>
<td>2.7</td>
<td>Phys.Rev.D84:031101,2011</td>
<td>anti-lepton</td>
</tr>
<tr>
<td>CDF</td>
<td>single top</td>
<td>σ(t)</td>
<td>3.2</td>
<td>Phys.Rev.D82:112005,2010</td>
<td>jet-electron anti-lepton</td>
</tr>
</tbody>
</table>
Conclusion
Conclusion

- QCD multijet background estimation is very important for precision top measurements and new physics searches.
- In top analysis the data-driven methods are widely used for QCD multijet background estimation:
  - Matrix method, templates fit method with jet-electron QCD model or anti-lepton QCD model.
- Systematics estimation of QCD multijet background usually comes from cross checks of alternative approaches in the method.
- QCD multijet background estimation methods well fit in current top physics analysis.
- Lots of dedicated efforts on QCD multijet background reduction, modeling and estimation.
Cross Check Method in Matrix Method

### Iteration Method for $\varepsilon_{\text{fake}}$

- **Use multijet enriched control sample** $E_T^{\text{miss}} < 10$ GeV

1. $\varepsilon_0^{\text{fake}} = \left( \frac{N_{\text{tight}}}{N_{\text{loose}}} \right) E_T^{\text{miss}} < 10$

2. $N_{\text{fake0}}^{\text{tight}} = \frac{\varepsilon_0^{\text{fake}}}{\varepsilon_{\text{real}} - \varepsilon_0^{\text{fake}}} (N_{\text{loose}}^{\text{real}} - N_{\text{tight}}^{\text{real}})$

3. $k_0^{W/Z} = \frac{N_{\text{tight}}^{\text{W/Z}} - N_{\text{fake0}}^{\text{tight}} - N_{\text{t\bar{t}},MC}^{\text{tight}} - N_{\text{t},MC}^{\text{tight}}}{N_{\text{W+jets},MC}^{\text{tight}} + N_{\text{Z+jets},MC}^{\text{tight}}}$

4. $\varepsilon_i^{W/Z} = \left( \frac{N_{\text{tight}}^{\text{W/Z}} - k_i^{W/Z} (N_{\text{W+jets},MC}^{\text{tight}} + N_{\text{Z+jets},MC}^{\text{tight}}) - N_{\text{t\bar{t}},MC}^{\text{tight}}}{N_{\text{loose}}^{\text{W/Z}} - k_i^{W/Z} (N_{\text{W+jets},MC}^{\text{loose}} + N_{\text{Z+jets},MC}^{\text{loose}}) - N_{\text{t\bar{t}},MC}^{\text{loose}}} \right) E_T^{\text{miss}} < 10$
Example of Control Region for $\epsilon_{\text{fake}}$

QCD in $\mu$+jets channel dominated by heavy flavor jets $\Rightarrow$ larger impact parameter ($d0$) respecting to primary vertex

$$\epsilon_{\text{fake}} = \frac{N_{\text{tight, } d0sig>5}}{N_{\text{loose, } d0sig>5}}$$
m +jets channel
- the loose data sample is defined by discarding the isolation requirements in the default muon selection
- the fake-muon efficiencies are derived from a low MTW control region, $MTW < 20$ GeV, with an additional requirement $MET + MTW < 60$ GeV.
- the efficiencies for real and fake muons are parameterized as a function of muon $|\eta|$ and of the leading jet $p_T$.

e+jets channel
- the loose data sample is defined by selecting events with electrons meeting looser identification criteria. The 3.5 GeV electron isolation requirement is loosened to 6 GeV.
- the fake-electron efficiencies are determined using a low MET control region (5 GeV < $MET < 20$ GeV). The efficiencies for real and fake-electrons are parameterized as a function of electron $|\eta|$.
- The uncertainty on the fake-lepton background is estimated by varying the requirements on the low MTW and low MET control regions, taking into account the statistical uncertainty and background corrections. The total uncertainty is estimated to be 100%.
- the differential analysis has results in each of their bins, so uncertainties tend to be a bit larger

- electron + jets channel
  - the QCD multijet background distributions are obtained from MC
- muon + jets channel
  - obtained from a background-enriched data sample defined as \( I_{rel} > 0.125 \) and \( MET < 20 \text{ GeV} \).
  - The QCD normalizations are evaluated individually in each \( N_{jet} \) and \( N_{tag} \) sub-sample.
    - the initial values that enter the profile likelihood fit.
  - QCD multijet is conservatively constrained with Gaussian uncertainties of 100% of its expected event yields,
- electron+jets channel
  - default QCD shape of electron channel: from events with relaxed requirements on the electron isolation and identification, and no \( MET \) requirement imposed.
  - the alternative shape is obtained from the region corresponding to the event selection used in the \( tt \) cross section measurement.
- muon + jets channel
  - the statistical fluctuations in the normalization for the \( MET \) distributions obtained from muon non-isolated \( (I_{rel} > 0.125) \) and isolated \( (I_{rel} < 0.125) \) regions are taken as the systematic uncertainty.

The SVM is defined as the mass of the sum of four-vectors of the tracks associated to the secondary vertex with an assumption that all particles have the pion mass.

The relative isolation is defined as \( I_{rel} = (ET_{\text{charged}} + ET_{\text{photon}} + ET_{\text{neutral}})/p_T \), where \( p_T \) is the lepton transverse momentum, and \( ET_{\text{charged}} \), \( ET_{\text{photon}} \), and \( ET_{\text{neutral}} \) are transverse energies of the charged particles, the reconstructed photons, and the neutral particles not identified as photons. The sum of the transverse energies is computed in a cone of size \( \Delta R = 0.3 \) around the lepton direction, excluding the lepton candidate itself. We require \( I_{rel} \) to be less than 0.125 for muons and 0.10 for electrons.
# More Results in Latest ATLAS Publications

<table>
<thead>
<tr>
<th>Channel</th>
<th>Goal</th>
<th>Luminosity (fb⁻¹)</th>
<th>Reference</th>
<th>Method</th>
</tr>
</thead>
</table>

QCD background estimation systematic uncertainties

1. use jet-electron method to estimate the systematics
2. 50% uncertainties on QCD background estimation from fake rate calculation
3. 80% uncertainties: limited data sample (64%) + jet misidentification rate uncertainty (50%)
4. 50% uncertainites: obtained from pile-up impact studies and alternative fit with m_T(W)
To get good prediction of QCD background from MC, we need:

- High precision on fragmentation and hadronization models.
- High precision on description of interactions with matter and in shower model.
- Good description of hadronic activity.
- Precise estimate of the cross section.
Event Selection in PRD 84, 031101 (R) (2011)

- at least one jet
- exactly one lepton candidate both with transverse energy ET > 20 GeV and pseudorapidity |η| < 2.0.
- at least one jet is btagged,
- at least 20 GeV of missing transverse energy in the event.
- to reduce QCD backgrounds, we require the transverse mass of the W, MTW to be at least 10 GeV/c² for muons, and at least 20 GeV/c² for electrons.
The multijets template is modeled using a sideband region where the $S_{\text{MET}}$ requirement is lowered to $3 < S_{\text{MET}} < 4$.

- This selection greatly enhances the contribution from multijet events, reducing other contributions (e.g. from $t\bar{t}$ events) to less than 1%.

- The assumption that the $n_{\text{track}}$ distribution for multijets events is independent of the $S_{\text{MET}}$ has been verified both with MC simulation over the entire $S_{\text{MET}}$ range, and with data, using the $2 < S_{\text{MET}} < 3$ and $4 < S_{\text{MET}} < 5$ sideband regions.
Multijet Systematics in tau+jet channel

<table>
<thead>
<tr>
<th>multijets template</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Closure to $2 &lt; S_{_{ETmiss}} &lt; 3$ sideband</td>
<td>1%</td>
</tr>
<tr>
<td>Closure to $4 &lt; S_{_{ETmiss}} &lt; 5$ sideband</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>Statistical uncertainties ($2 &lt; S_{_{ETmiss}} &lt; 3$)</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>Statistical uncertainties ($3 &lt; S_{_{ETmiss}} &lt; 4$)</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>Statistical uncertainties ($4 &lt; S_{_{ETmiss}} &lt; 5$)</td>
<td>1%</td>
</tr>
</tbody>
</table>
Multijet Modeling in All Jets ttbar Events

- Reference: CMS-TOP-11-017
- The multijet background is estimated using an event mixing technique.
  - use all events after the b-tagging selection
  - the jets are mixed between the different events based on their position in a $p_T$-ordered list in the event
  - every jet in the events in the multijet background model originates from a different event in the data, with the $p_T$-ordered position preserved.
  - no duplicate jets, in terms of their $p_T$-ordering, are allowed.
  - at least two b-tagged jets are found in every new event.
  - the kinematic fit to a $tt$ hypothesis is performed on each mixed event
QCD multijet Estimation with Fully Hadronic Decays

- Reference: JHEP 1301 (2013) 116
- Estimated using control regions dened by loosening the selection requirements for top-quark candidates and for associated b-tagged jets.
  - the multijet background is estimated in a manner similar to the HEPTopTagger analysis.
  - various control regions are used in order to reduce biases resulting from the observed correlations between the recoil and leading jet.