Evidence for single top at DØ

- Electroweak production of top quarks
- Event selection and background estimation
- Multivariate methods
  - Decision Trees, Matrix Elements, Bayesian NN
- Cross checks. Expected sensitivity
- Cross sections and significance
- First direct measurement of $|V_{tb}|$
- Combination
- Summary
Event selection designed to be as loose as possible:

- Only one tight (no loose) lepton:
  - $e$: $p_T > 15$ GeV and $|\eta^{\text{det}}| < 1.1$
  - $\mu$: $p_T > 18$ GeV and $|\eta^{\text{det}}| < 2.0$
- MET > 15 GeV
- 2-4 jets: $p_T > 15$ GeV and $|\eta^{\text{det}}| < 3.4$
- Leading jet: $p_T > 25$ GeV ; $|\eta^{\text{det}}| < 2.5$
- Second leading jet: $p_T > 20$ GeV
- One or two b-tagged jets

Acceptance: $tb = (3.2 \pm 0.4)\%$
$tqb = (2.1 \pm 0.3)\%$
Signal and Background modeling

- Signal is modeled with CompHEP (effective NLO) + Pythia
- W+jets and ttbar shapes from Alpgen with MLM matching + Pythia
  - Jet-parton matching avoids double counting → better model
- ttbar normalized to NNLO $\sigma = 6.8\pm1.2$ pb (Kidonakis, PRD 68, 114014)
- QCD from our selected data with non-isolated lepton
- Normalize W+jets and QCD to data before tagging
- Determine Wbb and Wcc factor in W+jets from zero-tagged data
  - Constant factor describes heavy flavor kinematics well
  - Largest single uncertainty: 30% relative error on Wbb+Wcc composition
Yields and systematic uncertainties

- Expect some 62 signal and 1400 background events
- Uncertainties are assigned per background, jet multiplicity, lepton channel, and number of tags
- Jet energy scale and b-tag eff. affect the shapes of distributions
- Statistics dominated analysis: systematics contribution to the uncertainty is small

### Event Yields in 0.9 fb⁻¹ Data

<table>
<thead>
<tr>
<th>Source</th>
<th>Event Yields</th>
<th>2 jets</th>
<th>3 jets</th>
<th>4 jets</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Electron+muon, 1tag+2tags combined</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$tb$</td>
<td></td>
<td>16 ± 3</td>
<td>8 ± 2</td>
<td>2 ± 1</td>
</tr>
<tr>
<td>$tqb$</td>
<td></td>
<td>20 ± 4</td>
<td>12 ± 3</td>
<td>4 ± 1</td>
</tr>
<tr>
<td>$t\bar{t} \rightarrow ll$</td>
<td></td>
<td>39 ± 9</td>
<td>32 ± 7</td>
<td>11 ± 3</td>
</tr>
<tr>
<td>$t\bar{t} \rightarrow l\bar{l}+jets$</td>
<td></td>
<td>20 ± 5</td>
<td>103 ± 25</td>
<td>143 ± 33</td>
</tr>
<tr>
<td>$W^+b\bar{b}$</td>
<td></td>
<td>261 ± 55</td>
<td>120 ± 24</td>
<td>35 ± 7</td>
</tr>
<tr>
<td>$W^+c\bar{c}$</td>
<td></td>
<td>151 ± 31</td>
<td>85 ± 17</td>
<td>23 ± 5</td>
</tr>
<tr>
<td>$W^+j\bar{j}$</td>
<td></td>
<td>119 ± 25</td>
<td>43 ± 9</td>
<td>12 ± 2</td>
</tr>
<tr>
<td>Multijets</td>
<td></td>
<td>95 ± 19</td>
<td>77 ± 15</td>
<td>29 ± 6</td>
</tr>
<tr>
<td>Total background</td>
<td></td>
<td>686 ± 41</td>
<td>460 ± 39</td>
<td>253 ± 38</td>
</tr>
<tr>
<td>Data</td>
<td></td>
<td>697</td>
<td>455</td>
<td>246</td>
</tr>
</tbody>
</table>

### Relative systematic uncertainties

- $W$+jets&QCD normalization: 18 – 28%
- top pair normalization: 18%
- Tag rate functions (shape): 2 – 16%
- Jet energy scale (shape): 1 – 20%
- Luminosity: 6%
- Trigger modeling: 3 – 6%
- Lepton ID: 2 – 7%
- Jet modeling: 2 – 7%
- Other small components: few%
Check distributions

![Graphs showing distributions](image)

**Key for Plots**
- Data
- $tb$
- $tqb$
- $t\bar{t}$
- $W + \text{jets}$
- Multijets
- $\pm 1\sigma$ uncertainty on background

Arán García-Bellido

Evidence for single top at DØ
Boosted Decision Trees

Idea: recover events that fail criteria in cut-based analysis

- Find best simple cut in each node looking at 49 physics motivated variables
- Output: purity $N_s/(N_s+N_B)$ for each event. Signal is $tb+tqb$.
- Boosting: retrain 20 times to learn from misclassified events
- Most discriminant: $M(\text{alljets})$, $M(W,b_1)$, $\cos(b,\ell)_{\text{top}}$, $Q(\ell)\eta(\text{light-jet})$

Background fraction vs. efficiency

**Evidence for single top at DØ**

Arán García-Bellido
Matrix Elements method

- Use all available kinematic information from a **fully differential cross-section calculation** ➔ See T. Gadfort talk in YSF session

- Calculate an event probability for signal and background hypothesis

\[
P(\vec{x}) = \frac{1}{\sigma} \int \frac{d\sigma}{dq_1} f(q_1; Q) dq_1 \frac{d\sigma}{dq_2} f(q_2; Q) dq_2 \times |M(\vec{y})|^2 \phi(\vec{y}) dy \times W(\vec{x}, \vec{y})
\]

- Integrate over 4 independent variables: assume angles well measured, known masses, momentum and energy conservation

\[
D_s(\vec{x}) = P(S|\vec{x}) = \frac{P_{\text{Signal}}(\vec{x})}{P_{\text{Signal}}(\vec{x}) + P_{\text{Background}}(\vec{x})}
\]

- Analysis only uses 2&3 jet bins
  - Wbg, Wcg, Wgg and Wbbg in P

Parton distribution functions CTEQ6  
Differential cross section (LO ME from Madgraph)  
Transfer Function: maps parton level (y) to reconstructed variables (x)

Arán García-Bellido  
Evidence for single top at DØ
Bayesian Neural Networks

A different sort of NN (http://www.cs.toronto.edu/radford/fbm.software.html):

- Instead of choosing one set of weights, find posterior probability density over all possible weights
- Averages over many networks weighted by the probability of each network given the training data
- Use 24 variables (subset of the DT variables) and train against sum of backgrounds

Advantages:
- Less prone to overfitting, because of Bayesian averaging
- Network structure less important: can use large networks!
- Optimized performance

Disadvantages:
- Computationally demanding!
Linear response

- Use ensemble testing to show analysis calibration
- Use pool of MC events to draw events with bkgd. yields fluctuated according to uncertainties, reproducing the correlations between components introduced in the normalization to data
- Randomly sample a Poisson distribution to simulate statistical fluctuations
- Linear response, negligible bias
Cross check samples

Check description of the two main backgrounds

- “Soft” $W$+jets: 2 jets and $H_T$(lepton,MET,alljets) < 175 GeV
- “Hard” $W$+jets: 3,4 jets and $H_T$(lepton,MET,alljets) > 300 GeV

Arán García-Bellido  Evidence for single top at DØ
(summed) Discriminants output

DØ Run II preliminary

Event Yield

tb+tqb Decision Tree Output

0.9 fb⁻¹

tb+tqb blue

W+jets green

Multijets brown

±1σ uncertainty on background

DØ Run II preliminary

Event Yield

tb+tqb Decision Tree Output

0.9 fb⁻¹

tb+tqb blue

W+jets green

Multijets brown

±1σ uncertainty on background

DØ Run II Preliminary, L=0.9 fb⁻¹

Event Yield

tb+tqb ME Discriminant

DØ Run II Preliminary, L=0.9 fb⁻¹

Event Yield

tb+tqb ME Discriminant

Arán García-Bellido

Evidence for single top at DØ
**Expected and observed results**

<table>
<thead>
<tr>
<th></th>
<th>Decision Trees</th>
<th>Matrix Elements</th>
<th>Bayesian NN</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma(tb+qb)$ [pb] Expected</td>
<td>2.7$^{+1.6}_{-1.4}$</td>
<td>3.0$^{+1.8}_{-1.5}$</td>
<td>3.2$^{+2.0}_{-1.8}$</td>
</tr>
<tr>
<td>$\sigma(tb+qb)$ [pb] Observed</td>
<td>4.9$\pm$1.4</td>
<td>4.6$^{+1.8}_{-1.5}$</td>
<td>5.0$\pm$1.9</td>
</tr>
<tr>
<td>significance Expected</td>
<td>2.1$\sigma$</td>
<td>1.8$\sigma$</td>
<td>1.3$\sigma$</td>
</tr>
<tr>
<td>significance Observed</td>
<td>3.4$\sigma$</td>
<td>2.9$\sigma$</td>
<td>2.4$\sigma$</td>
</tr>
</tbody>
</table>

DT measures $3.4\sigma$ excess! Evidence for single top production!

Results are compatible with the SM at ~1 std. dev.
First direct measurement of $|V_{tb}|$

Once we have a cross section measurement, we can make the first direct measurement of $|V_{tb}|$.

Calculate posterior in $|V_{tb}|^2$: $\sigma \propto |V_{tb}|^2$

Assume:
- **SM top decay**: $V_{td}^2 + V_{ts}^2 \ll V_{tb}^2$
- Pure V-A and CP conserving interaction

Additional theoretical errors are needed

This measurement does not assume 3 generations or unitarity
Combination of analyses

- Combine the three measurements with BLUE method
- Method requires to measure the correlations
- Used SM pseudo-datasets with systematics

\[ \rho = \begin{pmatrix}
DT & ME & BNN \\
1 & 0.57 & 0.51 \\
0.57 & 1 & 0.45 \\
0.51 & 0.45 & 1
\end{pmatrix} \]

Combined result: \( 4.8 \pm 1.3 \text{ pb} \) ➔ Significance of 3.5 std. dev.

**DØ Run II**

- Decision trees: 4.9 \(+1.4\ -1.4\) pb
- Matrix elements: 4.6 \(+1.8\ -1.5\) pb
- Bayesian NNs: 5.0 \(+1.9\ -1.9\) pb
- Combination: 4.8 \(+1.3\ -1.3\) pb

**Graphical Representation**

- Ensemble of pseudo-datasets with background only (no signal)
- 0.9 fb\(^{-1}\) per pseudo-dataset
- Combined result: 4.8 pb

\[ \sigma(p\bar{p} \rightarrow tb+tq\bar{b}) \text{ [pb]} \]

**Evidence for single top at DØ**
Conclusions

First evidence for single top quark production and direct measurement of $|V_{tb}|$

(hep-ex/0612052 submitted to PRL)

$$\sigma(s+t) = 4.8 \pm 1.3 \text{ pb}$$

3.5$\sigma$ significance!

$$|V_{tb}| > 0.68 \text{ @ 95\%C.L.}$$

- Challenging analysis: small signal hidden in huge complex background
- Expand to searches of new phenomena
- We now have double the data to analyze!
Signal and Background modeling

- Signal is modeled with CompHEP (effective NLO) + Pythia
- $W + \text{jets}$ and $ttbar$ shapes from Alpgen with MLM matching + Pythia
- $ttbar$ normalized to NNLO $\sigma = 6.8 \pm 1.2$ pb
- QCD from our selected data with non-isolated lepton
- Normalize $W + \text{jets}$ and QCD to data before tagging (SF $\sim 1.4$)
- Determine $Wbb$ and $Wcc$ fractions in $W + \text{jets}$ from zero-tagged data

$Wbb + Wcc$ factor $1.50 \pm 0.45$ makes all distributions match data
We form a binned likelihood from the discriminant outputs

Probability to observe data distribution D, expecting y:

\[ y = \alpha D \sigma + \sum_{s=1}^{N} b_s = a \sigma + \sum_{s=1}^{N} b_s \]

\[ P(D|y) = P(D|\sigma, a, b) = \prod_{i=1}^{nbins} P(D_i|y_i) \]

And obtain a Bayesian posterior probability density as a function of the cross section:

\[ Post(\sigma|D) = P(\sigma|D) \propto \int_a^b P(D|\sigma, a, b)Prior(\sigma)Prior(a, b) \]

- Shape and normalization systematics treated as nuisance parameters
- Correlations between uncertainties properly accounted for
- Flat prior in signal cross section
NN b-jet tagger

- NN trained on 7 input variables from SVT, JLIP and CSIP taggers
- Much improved performance!
  - Fake rate reduced by 1/3 for same b-efficiency relative to previous tagger
  - Smaller systematic uncertainty
- Tag Rate Functions (TRFs) in $\eta$, $p_T$ and z-PV derived in data are applied to MC
- Our operating point:
  - b-jet efficiency: $\sim$50%
  - c-jet efficiency: $\sim$10%
  - Light-jet efficiency: $\sim$0.5%
Decision Trees: 49 variables

**Object Kinematics**
- $p_T^{(jet1)}$
- $p_T^{(jet2)}$
- $p_T^{(jet3)}$
- $p_T^{(jet4)}$
- $p_T^{(best1)}$
- $p_T^{(notbest1)}$
- $p_T^{(tag1)}$
- $p_T^{(untag1)}$
- $p_T^{(untag2)}$

**Angular Correlations**
- $\Delta R^{(jet1,jet2)}$
- $\cos^{(best1,lepton)}^{besttop}$
- $\cos^{(best1,notbest1)}^{besttop}$
- $\cos^{(tag1,alljets)}^{alljets}$
- $\cos^{(tag1,lepton)}^{btaggedtop}$
- $\cos^{(jet1,alljets)}^{alljets}$
- $\cos^{(jet1,lepton)}^{btaggedtop}$
- $\cos^{(jet2,alljets)}^{alljets}$
- $\cos^{(jet2,lepton)}^{btaggedtop}$
- $\cos^{(lepton,Q(\text{lepton})\times z)}^{besttop}$
- $\cos^{(lepton,btaggedtop,btaggedtop,CMframe)}$
- $\cos^{(notbest,alljets)}^{alljets}$
- $\cos^{(notbest,lepton)}^{besttop}$
- $\cos^{(untag1,alljets)}^{alljets}$
- $\cos^{(untag1,lepton)}^{btaggedtop}$

**Event Kinematics**
- Aplanarity$(\text{alljets}, W)$
- $M(W,\text{best1})$ ("best" top mass)
- $M(W,\text{tag1})$ ("b-tagged" top mass)
- $H_T^{(alljets)}$
- $H_T^{(alljets—best1)}$
- $H_T^{(alljets—tag1)}$
- $H_T^{(alljets,W)}$
- $H_T^{(jet1,jet2)}$
- $H_T^{(jet1,jet2,W)}$
- $M(\text{alljets})$
- $M(\text{alljets—best1})$
- $M(\text{alljets—tag1})$
- $M(\text{jet1,jet2})$
- $M_T^{(jet1,jet2)}$
- $M_T^{(W)}$
- Missing $E_T$
- $p_T^{(alljets—best1)}$
- $p_T^{(alljets—tag1)}$
- $p_T^{(jet1,jet2)}$
- $Q(\text{lepton})\times \eta(\text{untag1})$
- $\sqrt{\hat{s}}$
- Sphericity$(\text{alljets}, W)$

**Most discrimination:**
- $M(\text{alljets})$
- $M(W,\text{tag1})$
- $\cos^{(tag1,\text{lepton})^{btaggedtop}}$
- $Q(\text{lepton})\times \eta(\text{untag1})$

- Adding variables does not degrade performance
- Tested shorter lists, lose some sensitivity
- Same list used for all channels
\[
\begin{pmatrix}
  d' \\
  s' \\
  b'
\end{pmatrix} = 
\begin{pmatrix}
  V_{ud} & V_{us} & V_{ub} \\
  V_{cd} & V_{cs} & V_{cb} \\
  V_{td} & V_{ts} & V_{tb}
\end{pmatrix}
\begin{pmatrix}
  d \\
  s \\
  b
\end{pmatrix}
\]

Most general Wtb vertex:

\[
\Gamma_{tbW}^{\mu} = -\frac{g}{\sqrt{2}} V_{tb} \left\{ \gamma^{\mu} \left[ f_1^L P_L + f_1^R P_R \right] - \frac{i \sigma^{\mu\nu}}{M_W} (p_t - p_b)_\nu \left[ f_2^L P_L + f_2^R P_R \right] \right\}
\]

Assume:

- **SM top decay**: \( V_{td}^2 + V_{ts}^2 \ll V_{tb}^2 \)
- Pure V-A interaction: \( f_1^R = 0 \)
- CP conservation: \( f_2^L = f_2^R = 0 \)

We are effectively measuring the **strength of the V-A coupling**: \( |V_{tb} f_1^L| \), which can be >1