Evidence for s-channel single top production at DØ

- Electroweak production of top quarks
- Event selection and background estimation
- Multivariate methods
  - Decision Trees, Bayesian NN, Matrix Elements
- Combination
- Cross sections and significance
- Direct measurement of $|V_{tb}|$

Summary

Arán García-Bellido
Top quark physics

The top quark is a very special fermion:

- Heaviest known particle: $173.2 \pm 0.9$ GeV (arXiv:1305.3929)
- $m_t \sim v/\sqrt{2}$, $\lambda_t \sim 1 \rightarrow$ Related to EWSB!
- Sensitive probe for new physics, FCNCs, ...

Decays as a free quark: $\tau_t = 5 \times 10^{-25}$ s $\ll \Lambda_{\text{QCD}}^{-1}$

- Spin information is passed to its decay products
- Test V-A structure of the SM

Wealth of measurements: Cross section, charge, mass, angular properties, width, lifetime, asymmetries, decays... LHC is a top factory!
Top quark production

- **Dominant mode is through strong interaction:** top pairs
  
  $\bar{q}q$ annihilation
  $\sigma(t\bar{t}) = 7.5\pm0.7$ pb

- **gg fusion**

- **Electroweak interaction:** single top
  
  $s$ channel
  $\sigma(tb) = 1.04\pm0.08$ pb

  $t$ channel
  $\sigma(tqb) = 2.26\pm0.12$ pb

  $tW$ production
  $\sigma(tW) = 0.28\pm0.06$ pb

Kidonakis, $m_t = 172.5$ GeV, Phys. Rev. D 74, 114012
Top quark production

- **Dominant mode is through strong interaction: top pairs**
  - $q\bar{q}$ annihilation
  - $\sigma(t\bar{t}) = 7.5\pm0.7$ pb

- **Electroweak interaction: single top**
  - $s$ channel
  - $\sigma(tb) = 1.04\pm0.08$ pb
  - $t$ channel
  - $\sigma(tqb) = 2.26\pm0.12$ pb

- $tW$ production
  - $\sigma(tW) = 0.28\pm0.06$ pb

Kidonakis, $m_t=172.5$ GeV, Phys. Rev. D 74, 114012
Why do we care?

- Access $W$-$t$-$b$ coupling
- Measure $V_{tb}$ directly
- Test unitarity of CKM

New physics:
- $s$-channel sensitive to resonances: $W'$, $H^+$, top pions, etc...
- $t$-channel sensitive to FCNCs, anomalous couplings

- Source of polarized top quarks
- Extract small signal out of a large background

DØ search: 1101.0806
DØ search: 1006.3575
## Experimental status
(before this analysis)

<table>
<thead>
<tr>
<th></th>
<th>( \sigma ) (NNLO) [pb]</th>
<th>( t_b )</th>
<th>( t_{qb} )</th>
<th>( t_W )</th>
</tr>
</thead>
<tbody>
<tr>
<td>TeV prediction</td>
<td>1.04</td>
<td>2.26</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td>CDF (7.5 fb(^{-1}))</td>
<td>1.8 ± 0.6</td>
<td>1.49 ± 0.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DØ (5.4 fb(^{-1}))</td>
<td>0.98 ± 0.63</td>
<td><strong>2.90 ± 0.59</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>( \sigma ) (NNLO) [pb]</th>
<th>( t_b )</th>
<th>( t_{qb} )</th>
<th>( t_W )</th>
</tr>
</thead>
<tbody>
<tr>
<td>LHC prediction (7 TeV)</td>
<td>4.6</td>
<td></td>
<td>64.6</td>
<td>15.7</td>
</tr>
<tr>
<td>ATLAS (0.7-2.1 fb(^{-1}))</td>
<td>&lt;20.5 (95%CL)</td>
<td></td>
<td><strong>83 ± 20</strong></td>
<td>17 ± 6</td>
</tr>
<tr>
<td>CMS (1.2-4.9 fb(^{-1}))</td>
<td>-</td>
<td></td>
<td><strong>67 ± 6</strong></td>
<td>16 ± 5</td>
</tr>
</tbody>
</table>

**Discovery (> 5 SD)**

**Evidence (>3 SD)**
The Tevatron

- 6.3 km pp collider
- $\sqrt{s} = 1.96$ TeV
- Run I: 1987-1996
- Run II: 2002-2011
- 36x36 bunches
- $10^{11}$ $p$ per bunch
- 396 ns bunch spacing
- 1.8 M crossings/s
- $4.3 \cdot 10^{32}$ cm$^2$s$^{-1}$ peak lumi
- 12 fb$^{-1}$ delivered luminosity
- Detectors recorded data with >90% efficiency
DØ for Run II

Tracker: Si+Fiber+Preshowers

2T solenoid magnet

Muon Scintillators

Muon Chambers

Calorimeter

Toroid

Shielding

3 layer muon system

protons

antiprotons

electronics
Use full data 9.7 fb\(^{-1}\) (with data quality requirement)

Run IIa, 1 fb\(^{-1}\)  
Run IIb, 8.7 fb\(^{-1}\)
A big challenge!

- 32k singletop events produced at the Tevatron
  - → Leptonic decays: 6.8k events
- Compare to huge W+jets background
- We needed 50 times more data to discover singletop (s+t) in 2009 than for t\bar{t} in 1995

New for this analysis:

- Optimized for s-channel
- Inclusive trigger
- Better b-tag algorithm
- Use matrix method for W+jets and multijet normalization
- New discriminants with b-tag info, and improved ME method
- Not assume SM ratio $\sigma_{tb}/\sigma_{tqb}$ for s+t or $V_{tb}$ measurements
- New joint tb and tqb discriminant
Signal selection

Signature:
- One high $p_T$ isolated lepton (from W)
- MET ($\nu$ from W)
- One b-quark jet (from top)
- A light flavor jet and/or another b-jet

Event selection:
- Only one isolated electron or muon, $p_T>$20 GeV:
  - Electron: $|\eta|<1.1$
  - Muon: $|\eta|<2.0$
- MET $>20$ GeV
- 2-3 jets: $p_T>$20 GeV and $|\eta|<2.5$
  - Leading jet: $p_T>$25 GeV
  - Second leading jet: $p_T>$20 GeV
- $H_T(\ell,MET,jets) > 120$ GeV
- One or two b-tagged jets
Background modeling

- **W+jets**: $\sim 1000$ pb
  - Distributions from Alpgen (MLM matching ME$\leftrightarrow$PS) + Pythia
  - Reweigh $\eta$(jet1), $\eta$(jet2) from data
  - Normalization from pre-tag data
  - Heavy flavor fraction from NLO
    - $Wb\bar{b}$ k-factor $\sim 1.9$

- **Top pairs**: $\sim 7$ pb
  - Topologies: dilepton and $\ell+\text{jets}$
  - Distributions from Alpgen
  - Normalize to NNLO $\sigma$

- **Multijet events (misidentified lepton)**
  - From data with non-isolated lepton

- **Z+jets** from Alpgen (scaled to NLO)

- **Diboson (WW, WZ, ZZ)** from Pythia
W+jets yield determination

- Normalize $W$+jets and QCD to data simultaneously before b-tagging (Matrix Method)
  - Split data sample in events with real and fake isolated lepton
  - Measure the probability to have an isolated lepton in each sample

- There are large k-factors for $Wb\bar{b}$, $Wc\bar{c}$ and $Wc\bar{j}$
  - Cross check with 0-tag sample, and by fitting the b-ID output distribution
  - Source of largest single uncertainty: 20% relative error on HF content

![Yield distributions]
Agreement before b-tagging

Normalize W+jets and QCD yields to data before b-tagging
Tagging b-jets

Three different algorithms for b-jet identification at DØ:

- Two based on tracks with large IP (JLIP, CSIP)
- One based on secondary vertex reconstruction (SVT)
- Combine with MVA

![Graph showing b-jet efficiency vs. misidentification rate](image1)

![Graph showing MVA output for b-jets and light jets](image2)
MVA b-jet tagger

- Scale factors derived from data are applied to MC as a function of $\eta$, $p_T$, and z-PV
- 1 tag category: Tight operating point
- 2 tag category: Loose operating point
- Efficiencies in 2 b-tag channel:
  - b-jet efficiency: $\sim$65% per jet
  - c-jet efficiency: $\sim$30% per jet
  - Light efficiency: $\sim$2.9% per jet

![Graphs showing DØ Loose Operating Point and Misidentification Rate vs. $p_T$](image-url)
Yields after event selection

<table>
<thead>
<tr>
<th>Category</th>
<th>Yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>e, μ 2, 3-jets 1, 2 b-tags combined</td>
<td>$257 \pm 31$</td>
</tr>
<tr>
<td>$tb$</td>
<td>$378 \pm 53$</td>
</tr>
<tr>
<td>$W$+jets</td>
<td>$7394 \pm 401$</td>
</tr>
<tr>
<td>diboson, $Z$+jets</td>
<td>$815 \pm 71$</td>
</tr>
<tr>
<td>top pair</td>
<td>$2672 \pm 284$</td>
</tr>
<tr>
<td>multijet</td>
<td>$789 \pm 81$</td>
</tr>
<tr>
<td>Background Sum</td>
<td>$11669 \pm 503$</td>
</tr>
<tr>
<td>Data</td>
<td>$12103$</td>
</tr>
</tbody>
</table>

- Optimized the selection to maximize acceptance $tb = 2.6\%$  $tqb = 1.8\%$
- Allow a lot of background at this stage!
- Then use multiple distributions to separate signal-background
Split analysis in four channels

- Multijet
- Top pair
- Diboson, $Z$+jets
- $W$+jets

**Sensitive to t-channel**

- $tqb$: $B$
- $tb$: 1: 2: 59
- $tb$: 1: 2: 14
- $tb$: 1: 2: 85
- $tb$: 1: 2: 40

**Sensitive to s-channel**

- $tb$: $tqb$: $B$
- 2 b-tags, 2-jets
- 1: 0.3: 14
- 1 b-tag, 3-jets
- 1: 2: 85
- 2 b-tags, 3-jets
- 1: 1: 40

**Constrain backgrounds**
Data-Background comparisons
Cross-check samples

W+jets enriched sample: 2 jets, 1 b-tag, $H_T<175$ GeV

$t\bar{t}$ enriched sample: 3 jets, $\geq 1$ b-tag, $H_T>300$ GeV
Analysis methods

Once we understand our data, need to measure the signal: use multivariate techniques.

DØ has implemented three analysis methods to extract the signal from the same dataset:

- Decision Trees
- Bayesian NNs
- Matrix Elements

- DT, BNN use 1/4 of events for training, 1/2 for measurement
- DT, BNN use well described input variables (KS>0.25)
- DT, BNN are the same used for tqb discovery (5.4 fb⁻¹)
- ME method uses 4-vectors of reconstructed objects
- Optimized separately for s-channel and t-channel
Boosted Decision Trees

- Apply sequential cuts but keep the failing events
  - List of 30 variables optimized for s-channel
  - Same list used for all 4 analysis channels
  - Trained with tb and tqb
- Train another tree produced by enhancing misclassified events (boosting)
- Repeat boosting → smoother, more discriminant output

![Decision Tree Diagram]

- $H_T > 212$
- $m_t < 352$
- $p_T < 31.6$

---

### Background fraction vs. efficiency

- **muon_qtb**
  - Single tree
  - Boosted trees (20)

<table>
<thead>
<tr>
<th>#</th>
<th>BDT input variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$E_T$</td>
</tr>
<tr>
<td>2</td>
<td>$p_T (\ell)$</td>
</tr>
<tr>
<td>3</td>
<td>$\eta (\ell)$</td>
</tr>
<tr>
<td>4</td>
<td>$M (\text{jet1})$</td>
</tr>
<tr>
<td>5</td>
<td>$p_T (\text{untag1})$</td>
</tr>
<tr>
<td>6</td>
<td>$E (\text{untag1})$</td>
</tr>
<tr>
<td>7</td>
<td>$M (\text{untag1})$</td>
</tr>
<tr>
<td>8</td>
<td>$b_{1D} (\text{untag1})$</td>
</tr>
<tr>
<td>9</td>
<td>$p_T (\text{jet2})$</td>
</tr>
<tr>
<td>10</td>
<td>$b_{1D} (\text{tag1})$</td>
</tr>
<tr>
<td>11</td>
<td>$\Delta R (\text{jet1, jet2})$</td>
</tr>
<tr>
<td>12</td>
<td>$\Delta R_{\text{min}} (\ell, \text{jet})$</td>
</tr>
<tr>
<td>13</td>
<td>$\Delta \Phi (\ell, E_T)$</td>
</tr>
<tr>
<td>14</td>
<td>$\Delta \Phi (\text{jet2}, E_T)$</td>
</tr>
<tr>
<td>15</td>
<td>$\Delta \Phi (\text{jet1}, E_T)$</td>
</tr>
<tr>
<td>16</td>
<td>$Q(\ell) \times \eta (\text{untag1})$</td>
</tr>
<tr>
<td>17</td>
<td>$Q(\ell) \times \eta (\text{jet2})$</td>
</tr>
<tr>
<td>18</td>
<td>$Q(\ell) \times \eta (\ell)$</td>
</tr>
<tr>
<td>19</td>
<td>$Q(\ell) \times \eta (\text{tag1})$</td>
</tr>
<tr>
<td>20</td>
<td>$\cos (\ell, \text{jet2})_{\text{lab}}$</td>
</tr>
<tr>
<td>21</td>
<td>$\cos (\ell, \text{jet1})_{\text{lab}}$</td>
</tr>
<tr>
<td>22</td>
<td>$H_T (\text{alljets})$</td>
</tr>
<tr>
<td>23</td>
<td>$H_T (\ell, E_T, \text{alljets})$</td>
</tr>
<tr>
<td>24</td>
<td>$H_T (\ell, E_T)$</td>
</tr>
<tr>
<td>25</td>
<td>$\text{Centrality(alljets)}$</td>
</tr>
<tr>
<td>26</td>
<td>$M_{\text{jet1, jet2}}$</td>
</tr>
<tr>
<td>27</td>
<td>$p_T (\text{jet1, jet2})$</td>
</tr>
<tr>
<td>28</td>
<td>$M_T (W)$</td>
</tr>
<tr>
<td>29</td>
<td>$p_T (W)$</td>
</tr>
</tbody>
</table>
Bayesian Neural Networks

- Uses 4-vectors of objects + $Q(\ell)\eta(\text{untag1}) + \mathcal{M}_T(W) + \text{b-ID output for jets}$
- 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

<table>
<thead>
<tr>
<th>#</th>
<th>BNN input variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$p_T(\text{tag1})$</td>
</tr>
<tr>
<td>2</td>
<td>$\eta(\text{tag1})$</td>
</tr>
<tr>
<td>3</td>
<td>$\Delta \Phi(\ell, \text{tag1})$</td>
</tr>
<tr>
<td>4</td>
<td>$b_{\text{ID}}(\text{tag1})$</td>
</tr>
<tr>
<td>5</td>
<td>$p_T(\text{untag1})$</td>
</tr>
<tr>
<td>6</td>
<td>$\eta(\text{untag1})$</td>
</tr>
<tr>
<td>7</td>
<td>$\Delta \Phi(\ell, \text{untag1})$</td>
</tr>
<tr>
<td>8</td>
<td>$b_{\text{ID}}(\text{untag1})$</td>
</tr>
<tr>
<td>9</td>
<td>$p_T(\ell)$</td>
</tr>
<tr>
<td>10</td>
<td>$\eta(\ell)$</td>
</tr>
<tr>
<td>11</td>
<td>$E_T$</td>
</tr>
<tr>
<td>12</td>
<td>$\Delta \Phi(\ell, E_T)$</td>
</tr>
<tr>
<td>13</td>
<td>$p_T(\text{tag2})$</td>
</tr>
<tr>
<td>14</td>
<td>$\eta(\text{tag2})$</td>
</tr>
<tr>
<td>15</td>
<td>$\Delta \Phi(\ell, \text{tag2})$</td>
</tr>
<tr>
<td>16</td>
<td>$b_{\text{ID}}(\text{tag2})$</td>
</tr>
<tr>
<td>17</td>
<td>$p_T(\text{untag2})$</td>
</tr>
<tr>
<td>18</td>
<td>$\eta(\text{untag2})$</td>
</tr>
<tr>
<td>19</td>
<td>$b_{\text{ID}}(\text{untag2})$</td>
</tr>
<tr>
<td>20</td>
<td>$\mathcal{M}_T(W)$</td>
</tr>
<tr>
<td>21</td>
<td>$Q(\ell) \times \eta(\text{untag1})$</td>
</tr>
</tbody>
</table>
Matrix Elements method

- The idea is to use all available kinematic information from a **fully differential cross-section calculation**

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

- Calculate an event probability for signal and background hypothesis

- Uses the 4-vectors of reconstructed $\ell$ and jets

- Jet-parton assignment: use b-tag information

- TF for e, $\mu$, jets, and jets misreconstructed as e

- Integrate over 4 (2jet) or 5 (3jet) independent variables: assume angles well measured, known masses, momentum and energy conservation
ME discriminants

We use these ME processes

Define discriminant from probabilities for signal and background

\[ D(x) = \frac{P_{\text{signal}}(x)}{P_{\text{signal}}(x) + P_{\text{background}}(x)} \]

New: use b-ID weights in Disc.

<table>
<thead>
<tr>
<th>Single Top</th>
<th>2 Jet</th>
<th>3 Jet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background</td>
<td>Wbb, Wcg, Wgg, top pair, WW, WZ, ggg</td>
<td>Wbbg, Wugg, top pair</td>
</tr>
<tr>
<td>2 jet b-tag</td>
<td>tb, tq</td>
<td>tbg, tqb, tqg</td>
</tr>
</tbody>
</table>

![Graph showing ME discriminant for tb and background]
BNN Combination

Use BNN to combine the 3 methods

Use 1/4 of sample for training
BNN tb combination

- Data
  - tb
  - tqb
  - W+jets
  - Z+jets
  - Diboson
  - tt
  - Multijets

(a) BNNcomb

(b) DØ 9.7 fb⁻¹

Ranked s-channel discriminant

Yield [Events/0.02]

- Data
- tb
- tqb
- W+jets
- Z+jets/Diboson
- tt
- Multijets
BNN tqb combination

- Data
  - $t\bar{b}$
  - tqqb
  - W+jets
  - Z+jets
  - Diboson
  - $t\bar{t}$
  - Multijets

(c) DØ 9.7 fb$^{-1}$ BNNcomb

(d) DØ 9.7 fb$^{-1}$
New discriminant

- Aim to simultaneously measure \(tb\) and \(tqb\) signals without assuming the SM prediction for either
- Need discriminant sensitive to both signals
- Ensure each bin contains enough statistics to have a stable measurement
- Avoid complex binning in 2D
- Split every event based on whether \(D_{tb} > D_{tqb}\) or \(D_{tb} < D_{tqb}\)
If $D_{tb} > D_{tqb}$:
- $tb$ category
- Use $D_{tb}$
- Plot in the range $[0, 1]$

If $D_{tqb} > D_{tb}$:
- $tqb$ category
- Use $D_{tqb}$
- Plot in the range $[1, 2]$
Systematic uncertainties

- Assign to each background and each analysis channel.
- Some affect only the overall scale, and others affect also the discriminant outputs bin-by-bin (shape-changing).

<table>
<thead>
<tr>
<th>Overall Scale</th>
<th>Overall Scale &amp; Shape</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Heavy Flavor Uncertainty</strong></td>
<td><strong>b-tagging Efficiency Uncertainty</strong></td>
</tr>
<tr>
<td>Systematics/Nominal</td>
<td>Systematics/Nominal</td>
</tr>
<tr>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>+1σ_{HF}</td>
<td>+1σ_{BTag}</td>
</tr>
<tr>
<td>nominal</td>
<td>nominal</td>
</tr>
<tr>
<td>-1σ_{HF}</td>
<td>-1σ_{BTag}</td>
</tr>
</tbody>
</table>
Systematic uncertainties

- Assign to each background and each analysis channel
- Some affect only the overall scale, and others affect also the discriminant outputs bin-by-bin (shape-changing)
- Main relative uncertainties are listed here

<table>
<thead>
<tr>
<th>Overall Scale</th>
<th>Overall Scale &amp; Shape</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integrated luminosity</td>
<td>Jet reconstruction up to 1.4%</td>
</tr>
<tr>
<td>Top pair cross section</td>
<td>Jet energy resolution up to 1.1%</td>
</tr>
<tr>
<td>Diboson cross section</td>
<td>Jet energy scale up to 1.2%</td>
</tr>
<tr>
<td>Trigger efficiencies</td>
<td>Flavor-dependent JES up to 1.3%</td>
</tr>
<tr>
<td>Jet fragmentation+higher order</td>
<td>Jet vertex confirmation up to 1.1%</td>
</tr>
<tr>
<td>Initial- and final-state radiation</td>
<td>$b$-ID, 1 $b$-tagged channel up to 6.6%</td>
</tr>
<tr>
<td>Heavy-flavor correction</td>
<td>$b$-ID, 2 $b$-tagged channel up to 8.8%</td>
</tr>
<tr>
<td>Multijet normalization</td>
<td></td>
</tr>
</tbody>
</table>
Extracting the cross section

- Use the BNN combination discriminant in 25 bins
- Use all bins (we don’t cut on the discriminant)
- For each bin, the likelihood $L$ to observe $D$ data events with a known mean $h$ is modeled by the Poisson distribution

$$D \sim h = a_1 \sigma_1 + a_2 \sigma_2 + b$$

Signal acceptance:
- $tb$: 2.6%
- $tqb$: 1.8%

To be measured: signal cross section, backgrounds (known)
Bayesian approach

- Likelihood of observing data distribution $D$, when $h$ is expected:

$$L(D \mid h) \equiv L(D \mid \sigma_1, \sigma_2, a_1, a_2, b) = \prod_{i=1}^{\text{nbins}} L(D_i \mid h_i)$$

- Obtain Bayesian posterior probability as a function of $\sigma_1$, $\sigma_2$:

$$\int L(D \mid \sigma_1, \sigma_2, a_1, a_2, b) \pi(\sigma_1, \sigma_2) \pi(a_1, a_2, b) \, da_1 \, da_2 \, db \propto p(\sigma_1, \sigma_2 \mid D)$$

- Poisson likelihood

- Uniform prior for the signal cross section

- Our state of knowledge, $a$, $b$ with systematic uncertainties

- Shape & normalization systematics treated as nuisance parameters

- Correlations between uncertainties properly accounted for
Two dimensional posterior

![Graph with two-dimensional posterior for t-channel and s-channel cross sections with different error contours and symbols for different models and measurements.]

- **DØ 9.7 fb⁻¹**
- **1 SD**
- **2 SD**
- **3 SD**

- **Measurement**[^1]
- **SM**[^2]
- **Four generations**[^3]
- **Top-flavor**[^3]
- **Top pion**[^3]
- **FCNC**[^4]

[^1]: PRD 74: 114012, 2006
[^3]: PRD 63: 014018, 2001
Measured cross section

\[ \sigma_{\text{expected}} = 2.33^{+0.47}_{-0.44} \text{ pb} \]

\[ \sigma_{\text{observed}} = 3.07^{+0.53}_{-0.49} \text{ pb} \]

No assumption on SM \( \sigma_{tb}/\sigma_{tqb} \)

(Integrate over \( \sigma_{tqb} \))

\[ \sigma_{\text{expected}} = 1.08^{+0.31}_{-0.30} \text{ pb} \]

\[ \sigma_{\text{observed}} = 1.10^{+0.33}_{-0.31} \text{ pb} \]
Asymptotic approximation of the log-likelihood ratio

Tests how likely the data is to fluctuate to the measured $\sigma$ value, in the absence of the signals

- Expected p-values:
  - $t_b$: $1.0 \times 10^{-4}$ (3.7 SD)
  - $t_{qb}$: $9.9 \times 10^{-10}$ (6.0 SD)

- Observed p-values:
  - $t_b$: $1.0 \times 10^{-4}$ (3.7 SD)
  - $t_{qb}$: $6.1 \times 10^{-15}$ (7.7 SD)

All BDT, BNN and ME methods have more than 3 SD significance alone
Individual results

<table>
<thead>
<tr>
<th>Channel</th>
<th>Expected $\sigma$ (pb)</th>
<th>Observed $\sigma$ (pb)</th>
<th>Expected $p$ value</th>
<th>Observed $p$ value</th>
<th>Expected $Z$</th>
<th>Observed $Z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ME$_s$</td>
<td>1.05$^{+0.36}_{-0.34}$</td>
<td>1.12$^{+0.36}_{-0.33}$</td>
<td>$8.1 \times 10^{-4}$</td>
<td>$3.7 \times 10^{-4}$</td>
<td>3.2</td>
<td>3.4</td>
</tr>
<tr>
<td>BNN$_s$</td>
<td>1.06$^{+0.41}_{-0.39}$</td>
<td>1.61$^{+0.43}_{-0.40}$</td>
<td>$3.3 \times 10^{-3}$</td>
<td>$1.5 \times 10^{-5}$</td>
<td>2.7</td>
<td>4.2</td>
</tr>
<tr>
<td>BDT$_s$</td>
<td>1.06$^{+0.35}_{-0.33}$</td>
<td>1.56$^{+0.40}_{-0.37}$</td>
<td>$5.4 \times 10^{-4}$</td>
<td>$2.3 \times 10^{-6}$</td>
<td>3.3</td>
<td>4.6</td>
</tr>
<tr>
<td>$D^s_{\text{comb}}$</td>
<td>1.07$^{+0.32}_{-0.30}$</td>
<td>1.10$^{+0.33}_{-0.31}$</td>
<td>$1.0 \times 10^{-4}$</td>
<td>$1.0 \times 10^{-4}$</td>
<td>3.7</td>
<td>3.7</td>
</tr>
<tr>
<td>ME$_t$</td>
<td>2.27$^{+0.55}_{-0.51}$</td>
<td>2.15$^{+0.54}_{-0.50}$</td>
<td>$6.6 \times 10^{-7}$</td>
<td>$2.8 \times 10^{-6}$</td>
<td>4.8</td>
<td>4.5</td>
</tr>
<tr>
<td>BNN$_t$</td>
<td>2.31$^{+0.54}_{-0.50}$</td>
<td>2.41$^{+0.55}_{-0.51}$</td>
<td>$2.4 \times 10^{-7}$</td>
<td>$1.4 \times 10^{-7}$</td>
<td>5.0</td>
<td>5.1</td>
</tr>
<tr>
<td>BDT$_t$</td>
<td>2.36$^{+0.53}_{-0.50}$</td>
<td>3.70$^{+0.66}_{-0.60}$</td>
<td>$5.4 \times 10^{-8}$</td>
<td>$3.4 \times 10^{-15}$</td>
<td>5.3</td>
<td>7.8</td>
</tr>
<tr>
<td>$D^t_{\text{comb}}$</td>
<td>2.33$^{+0.47}_{-0.44}$</td>
<td>3.07$^{+0.54}_{-0.49}$</td>
<td>$1.0 \times 10^{-9}$</td>
<td>$7.1 \times 10^{-15}$</td>
<td>6.0</td>
<td>7.7</td>
</tr>
<tr>
<td>$D^{s+t}_{\text{comb}}$</td>
<td>3.34$^{+0.53}_{-0.49}$</td>
<td>4.11$^{+0.60}_{-0.55}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Measure combined $\sigma_{tb+tqb}$

- Measure $\sigma_{tb+tqb}$ without assuming the SM $\sigma_{tb}/\sigma_{tqb}$.
- Use 2D posterior p.d.f.
- Integrate over $\sigma_{tqb}$ and obtain 1D posterior p.d.f of $\sigma_{tb+tqb}$.
tb or not tb

- Examine highest BNNcomb bins, with post-fit uncertainty and measured $\sigma_{tb}$

- Integrate from right to left, and plot cumulative signal vs all events
Event characteristics

\( tb \) Category: \( D_{tb} > 0.8 \)

\( tb \) & \( tqb \) Depleted Region

\[ DØ, \ 9.7 \text{ fb}^{-1} \]

Yield [Events/5GeV]

Yield [Events/20GeV]
Event characteristics

*tqb* Category: $D_{tqb} > 0.8$

$tb$ & *tqb* Depleted Region

![Graphs showing event characteristics](image-url)
Run 252918
Event 51093921
Sat. June 13 23:07:10 2009

A tb event candidate

\[ m_t = 171 \text{ GeV} \]
Jet1 b-tag: 0.95
Jet2 b-tag: 0.84

ET scale: 54 GeV
Experimental status
(after this analysis)

<table>
<thead>
<tr>
<th></th>
<th>tb</th>
<th>tqb</th>
<th>tW</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>σ (NNLO) [pb]</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TeV prediction</td>
<td>1.04</td>
<td>2.26</td>
<td>0.28</td>
</tr>
<tr>
<td>CDF (9.4-7.5 fb⁻¹)</td>
<td>1.41 ± 0.44</td>
<td>1.49 ± 0.45</td>
<td>-</td>
</tr>
<tr>
<td>DØ (9.7 fb⁻¹)</td>
<td>1.10 ± 0.33</td>
<td>3.07 ± 0.53</td>
<td>-</td>
</tr>
<tr>
<td><strong>σ (NNLO) [pb]</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LHC prediction (7 TeV)</td>
<td>4.6</td>
<td>64.6</td>
<td>15.7</td>
</tr>
<tr>
<td>ATLAS (0.7-20 fb⁻¹)</td>
<td>&lt;20.5 (95%CL)</td>
<td>83 ± 20</td>
<td>27 ± 6</td>
</tr>
<tr>
<td>CMS (1.2-12.2 fb⁻¹)</td>
<td>-</td>
<td>67 ± 6</td>
<td>23 ± 5</td>
</tr>
</tbody>
</table>

**Discovery (> 5 SD)**

**Evidence (>3 SD)**
Tevatron latest measurements

s-channel Single Top Quark Cross Section

- DØ e/μ+jets 5.4 fb$^{-1}$: $0.68^{+0.38}_{-0.35}$ pb
- CDF MET+jets 9.5 fb$^{-1}$: $1.10^{+0.65}_{-0.66}$ pb
- CDF e/μ+jets 9.4 fb$^{-1}$: $1.41^{+0.44}_{-0.42}$ pb
- DØ e/μ+jets 9.7 fb$^{-1}$: $1.10^{+0.33}_{-0.31}$ pb

$m_t = 172.5$ GeV

N. Kidonakis
PRD 74 114012 (2006)
Measuring $|V_{tb}|$

- Once we have a cross section measurement, we can make a direct measurement of $|V_{tb}|$, since $\sigma_{tb+tqb} \propto |V_{tb}|^2$

\[
\begin{pmatrix}
  d' \\
  s' \\
  b'
\end{pmatrix}
= 
\begin{pmatrix}
  V_{ud} & V_{us} & V_{ub} \\
  V_{cd} & V_{cs} & V_{cb} \\
  V_{td} & V_{ts} & V_{tb}
\end{pmatrix}
\begin{pmatrix}
  d \\
  s \\
  b
\end{pmatrix}
\]

- Most general Wtb vertex [PLB 713, 165 (2012)]:

\[
L = \frac{g}{\sqrt{2}} \bar{b} \gamma^\mu V_{tb} (f^L_1 P_L + f^R_1 P_R) t W^-_\mu - \frac{g}{\sqrt{2}} \bar{b} i \sigma^{\mu\nu} q^\nu V_{tb} \frac{V_{tb}}{M_W} (f^L_2 P_L + f^R_2 P_R) t W^-_\mu
\]

- Assume:
  - SM top decay: $V_{td}^2 + V_{ts}^2 \ll V_{tb}^2$
  - Pure V-A interaction: $f^R_1 = 0$
  - CP conservation: $f^L_2 = f^R_2 = 0$

We are effectively measuring the **strength of the V-A coupling**: $|V_{tb} f^L_1|$, which can be >1

- Do not assume:
  - 3 generations
  - Unitarity of CKM
  - New: $\sigma_{tb}/\sigma_{tqb}$ (NEW)
CKM matrix element $|V_{tb}|$

- Allow $|V_{tb} f_1^L|^2 > 1$
  - $|V_{tb} f_1^L| = 1.12^{+0.09}_{-0.08}$
- Assume $0 \leq |V_{tb}|^2 \leq 1$
  - $|V_{tb}| > 0.92$ @ 95% C.L.
- Additional systematic uncertainties
  - Theoretical uncertainty on single top cross sections
- Complementary to $R_{Wb/Wq}$ measurement in top decays [PRL 107, 121802 (2011)]
- Current limits @ 95% C.L.:
  - CDF (7.5 fb$^{-1}$): $0.78 < |V_{tb}| \leq 1$
  - ATLAS (6 fb$^{-1}$ 8TeV): $0.80 < |V_{tb}| \leq 1$
  - CMS (5 fb$^{-1}$ 8TeV): $0.81 < |V_{tb}| \leq 1$
Conclusions

First evidence of s-channel single top quark production

$$\sigma_{tb} = 1.10 \pm 0.33 \text{ pb}$$


Simultaneously measure $\sigma_{tb}$ and $\sigma_{tqb}$, without assuming the SM prediction for either

Also measure $\sigma_{tb+tqb}$ and $|V_{tb}|$ without assuming the SM ratio of $\sigma_{tb}/\sigma_{tqb}$

$$|V_{tb}| > 0.92 \text{ @ 95% C.L.}$$

Results are consistent with the SM predictions

A legacy measurement at the Tevatron

Looking forward to combination with CDF
Extra slides

For more information:
http://www-d0.fnal.gov/Run2Physics/top/top_public_web_pages/top_public.html
CDF $\ell+$jets result 9.4 fb$^{-1}$

- Isolated e/μ $p_T>$20 GeV
- MET>10 (20) GeV muon (e)
- Two jets: $E_T>$20 GeV, $|\eta|<$2.0, leading jet: $E_T>$30 GeV
- $H_T>$125 GeV, $M_{jj}>30$ GeV
- W+jets normalization: fit to MET in pretag sample
- Train NN with 8 variables for each lepton and tag category

<table>
<thead>
<tr>
<th>Category</th>
<th>TT</th>
<th>TL</th>
<th>T</th>
<th>LL</th>
</tr>
</thead>
<tbody>
<tr>
<td>WW</td>
<td>1.7±0.4</td>
<td>13.2±2.7</td>
<td>184±23</td>
<td>24.8±3.9</td>
</tr>
<tr>
<td>WZ</td>
<td>17.8±2.2</td>
<td>21.2±2.0</td>
<td>52.7±5.4</td>
<td>9.9±0.9</td>
</tr>
<tr>
<td>ZZ</td>
<td>2.4±0.3</td>
<td>2.4±0.2</td>
<td>7.1±0.7</td>
<td>0.96±0.08</td>
</tr>
<tr>
<td>Z + jets</td>
<td>10.9±1.2</td>
<td>20.7±2.3</td>
<td>163±18</td>
<td>27.1±3.1</td>
</tr>
<tr>
<td>$t\bar{t}$</td>
<td>163±21</td>
<td>194±19</td>
<td>502±50</td>
<td>58.1±6.6</td>
</tr>
<tr>
<td>Higgs</td>
<td>6.1±0.6</td>
<td>6.4±0.4</td>
<td>10.3±0.7</td>
<td>1.7±0.2</td>
</tr>
<tr>
<td>$Wbb$</td>
<td>246±99</td>
<td>327±130</td>
<td>1166±468</td>
<td>109±44</td>
</tr>
<tr>
<td>$Wcc$</td>
<td>19.0±7.8</td>
<td>120±49</td>
<td>1158±467</td>
<td>164±67</td>
</tr>
<tr>
<td>$W + Mistag$</td>
<td>4.3±1.3</td>
<td>62±13</td>
<td>978±141</td>
<td>242±34</td>
</tr>
<tr>
<td>Multijet</td>
<td>29±12</td>
<td>47±19</td>
<td>281±112</td>
<td>45±18</td>
</tr>
<tr>
<td>$t$ and $W$-channel</td>
<td>18.1±2.5</td>
<td>35.3±4.2</td>
<td>251±28</td>
<td>13.6±1.5</td>
</tr>
<tr>
<td>s-channel</td>
<td>54.5±6.7</td>
<td>61.2±5.6</td>
<td>109±10</td>
<td>17.8±2.1</td>
</tr>
<tr>
<td>Total Prediction</td>
<td>573±155</td>
<td>911±248</td>
<td>4860±1320</td>
<td>714±181</td>
</tr>
</tbody>
</table>

Observed: 466, 765, 4620, 718

$\sigma_{tb} = 1.41^{+0.44}_{-0.42}$ pb

Significance: 3.8σ (2.9σ expected)
CMS tW observation 12.2 fb$^{-1}$ 8 TeV

- Signal region: 1 tight jet, 1 b-tag
- Control regions, dominated by t$\bar{t}$:
  - 2 tight jets, 1, 2 b-tags
  - Train BDT against t$\bar{t}$ with 13 variables
- Fit done for all channels (ee, e$\mu$, $\mu$)$\mu$ and regions (1j1t, 2j1t, 2j2t) simultaneously

$\sigma_{tW} = 23.4 \pm 5.5 - 5.4$ pb ; $\sigma_{tW}^{SM} = 22.2 \pm 0.6 \pm 1.4$ pb

Significance: 6.0σ (5.4σ expected)

$|V_{tb}| = 1.03 \pm 0.12$ (exp) $\pm 0.04$ (th)

$|V_{tb}| > 0.78$ at 95% C.L.
Discriminant on cross-check samples

s-like

W+jets enriched
1 b-tag
2 jets
$H_T < 175$ GeV

DØ, 9.7 fb$^{-1}$

Data
- $t\bar{b}$
- $tq\bar{b}$
- $W+$jets
- $Z+$jets
- Diboson
- $t\bar{t}$
- Multijets

DØ, 9.7 fb$^{-1}$

tt enriched
1,2 b-tag
3 jets
$H_T > 300$ GeV

DØ, 9.7 fb$^{-1}$

Data
- $t\bar{b}$
- $tq\bar{b}$
- $W+$jets
- $Z+$jets
- Diboson
- $t\bar{t}$
- Multijets
Linearity test

- Generate ensembles of pseudo-data samples
- Each ensemble has a different input signal $\sigma$
- All systematics included
- Extract the signal cross section from each pseudo-data sample
- No calibration needed
<table>
<thead>
<tr>
<th></th>
<th>2</th>
<th>2</th>
<th>3</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of jets</strong></td>
<td>112±23</td>
<td>83±19</td>
<td>33±7</td>
<td>29±7</td>
</tr>
<tr>
<td><strong>Number of b tags</strong></td>
<td>248±50</td>
<td>23±5</td>
<td>75±15</td>
<td>32±7</td>
</tr>
<tr>
<td><em>s channel</em></td>
<td>585±100</td>
<td>275±52</td>
<td>1044±207</td>
<td>767±158</td>
</tr>
<tr>
<td><em>t channel</em></td>
<td>4984±369</td>
<td>715±96</td>
<td>1395±120</td>
<td>300±39</td>
</tr>
<tr>
<td><em>tt</em></td>
<td>544±67</td>
<td>79±10</td>
<td>156±18</td>
<td>36±5</td>
</tr>
<tr>
<td><em>W+jets</em></td>
<td>479±73</td>
<td>65±10</td>
<td>188±33</td>
<td>56±9</td>
</tr>
<tr>
<td><em>Z+jets and diboson</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Background sum</td>
<td>6592±395</td>
<td>1134±110</td>
<td>2784±242</td>
<td>1160±164</td>
</tr>
<tr>
<td>Backgrounds + signals</td>
<td>6952±399</td>
<td>1240±112</td>
<td>2891±243</td>
<td>1220±164</td>
</tr>
<tr>
<td>Data</td>
<td>6859</td>
<td>1286</td>
<td>2725</td>
<td>1233</td>
</tr>
<tr>
<td><em>S(tqb):B</em></td>
<td>1:27</td>
<td>1:52</td>
<td>1:38</td>
<td>1:38</td>
</tr>
</tbody>
</table>
Crash course in Bayesian probability

Bayes’ theorem expresses the degree of belief in a hypothesis A, given another B. “Conditional” probability \( P(A|B) \):

\[
P(A|B) = \frac{P(B|A)P(A)}{P(B)}
\]

In HEP: \( B \rightarrow N_{\text{observed}} \), \( A \rightarrow n_{\text{predicted}} = n_{\text{signal}} + n_{\text{bkgd}} \), \( n_s = \text{Acc} \times \mathcal{L} \times \sigma \)

\( P(B|A) \): “model” density, or likelihood: \( L(N_{\text{observed}}|n_{\text{predicted}}) = n^N e^{-n}/N! \)

\( P(A) \): “prior” probability density \( \prod(n_{\text{pred}}) = \prod(\text{Acc} \times \mathcal{L}, n_b) \prod(\sigma) \)

\( \prod(n_s, n_b) \) multivariate gaussian ; \( \prod(\sigma) \) assumed flat

\( P(B) \): normalization constant \( Z \): \( P(N_{\text{observed}}) \)

\( P(A|B) \): “posterior” probability density \( P(n_{\text{predicted}}|N_{\text{observed}}) \)

\[
P(n_{\text{predicted}}|N_{\text{observed}}) = \frac{1}{Z} L(N_{\text{observed}}|n_{\text{predicted}}) \prod(n_{\text{pred}})
\]
W+jets normalization

Find fractions of real and fake isolated $\ell$ in the data before b-tagging. Split samples in loose and tight isolation:

$$N^{\text{loose}} = N^{\text{fake \_ loose}} + N^{\text{real \_ loose}}$$

$$N^{\text{tight}} = \varepsilon^{\text{fake}} N^{\text{fake \_ loose}} + \varepsilon^{\text{real}} N^{\text{real \_ loose}}$$

Obtain $\varepsilon^{\text{fake}}$ and $\varepsilon^{\text{real}}$ from MC and data samples

Then apply b-tagging

- Greatly reduce W+jets background ($W_{bb} \sim 5\%$ of $W_{jj}$)
- Shift distributions, changes flavor composition

Obtain: $N^{\text{loose \_ real}}$ and $N^{\text{loose \_ fake}}$
Background contribution

DØ

s-channel  t-channel  W+jets  Z+jet, dibosons  tt+tW  Multijets

CMS (arXiv:1209.4533v1)
<table>
<thead>
<tr>
<th>ME processes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Two Jets</strong></td>
</tr>
<tr>
<td><strong>Name</strong></td>
</tr>
<tr>
<td>$tb$</td>
</tr>
<tr>
<td>$tq$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

$W_{bb}$ | $u\bar{d} \rightarrow Wb\bar{b}$ | $W_{bbg}$ | $u\bar{d} \rightarrow Wb\bar{b}g$ |
| $W_{cg}$ | $sg \rightarrow Wc\bar{g}$ | $W_{ugg}$ | $\bar{u}g \rightarrow W\bar{u}gg$ |
| $W_{gg}$ | $u\bar{d} \rightarrow Wg\bar{g}$ | | |
| $W_{W}$ | $u\bar{u} \rightarrow WW$ | | |
| $W_{Z}$ | $u\bar{d} \rightarrow WZ$ | | |
| $ggg$ | $gg \rightarrow ggg$ | | |
| $t\bar{t}$ | $u\bar{u} \rightarrow t\bar{t}$ | $t\bar{t}$ | $u\bar{u} \rightarrow t\bar{t}$ |

The more background diagrams, the better discrimination
ME $t\bar{t}$ modeling

- $t\bar{t} \rightarrow \ell v b q q' b$ (4 jets)
- $t\bar{t}$ yields in 2jet & 3jet channels are comparable to single top
- Light-jets are 1.6 times more likely to be lost than b-jets
- Use simulation to derive a prior of missing jet (3jet) or missing W (2jet)

**Missing W prior**

**Missing Jet**
\[ P = \frac{\sum_j w_j d \sigma_j}{(\sum_j w_j) \sigma_j} \]

\[ w_j = \begin{cases} 
  b_1 b_2, & j=1,2 \\
  (1-b_1)(1-b_2) & \end{cases} \]

**t-channel gets weights for each probability**

\[ D = \frac{b_1 b_2 P_{tq}}{b_1 b_2 P_{tb, Wbb, WZ, tt} + b_1 (1-b_2) P_{tq} + (1-b_1) (1-b_2) P_{Wcg, Wgg, WW, ggg}} \]

**s-channel gets overall weight for event**
Another candidate event

Run 264600 Evt 37760117 Wed Sep 8 07:49:49 2010

$E_\text{scale} = 143 \text{ GeV}$

$m_t = 175 \text{ GeV}$
Jet1 b-tag: 0.32
Jet2 b-tag: 0.39