Discovery of Single Top Quark Production

Dag Gillberg
Simon Fraser University (now Carleton University)
For the DØ and CDF Collaborations
March 4, 2009 – both DØ and CDF independently present first observation of single top, **14 years** after discovery of $t\bar{t}$ production
Milestone for small signal in large background!

From R. Wallny’s Wine and Cheese talk
Why single top?

Electroweak single top production:

- **s-channel (tb)**
  - $\sigma_{NLO}^{[†]} = 0.88 \pm 0.11 \text{ pb}$
  - $\sigma_{(N)NLO}^{[‡]} = 1.12 \pm 0.05 \text{ pb}$

- **t-channel (tqb)**
  - $\sigma_{NLO}^{[†]} = 1.98 \pm 0.25 \text{ pb} \ (m_t=175 \text{ GeV})$
  - $\sigma_{(N)NLO}^{[‡]} = 2.34 \pm 0.13 \text{ pb} \ (m_t=170 \text{ GeV})$

- Measurement of top properties, e.g., **polarization**
- **New Physics:** 4th quark generation?
  - $tb$: $W'$, $H^\pm$?
  - $tqb$: FCNC?
- Background for $WH \rightarrow Wb\bar{b}$ — **similar analysis**
- Milestone for small signal in large backgrounds

The Wtb coupling

$V_{CKM} = \begin{pmatrix}
V_{ud} & V_{us} & V_{ub} \\
V_{cd} & V_{cs} & V_{cb} \\
V_{td} & V_{ts} & V_{tb}
\end{pmatrix}$

- Direct $|V_{tb}|$ measurement
- Unitarity test of CKM matrix
- Anomalous Wtb couplings

[†] Z. Sullivan, Phys. Rev. D 70, 114012 (2004)
[‡] N. Kidonakis, Phys. Rev. D 74, 114012 (2006)
Selection and Backgrounds

$t\bar{b}(q) \rightarrow \ell \nu \bar{b}\bar{b}(q)$ selection

- High $p_T$ isolated $e$ or $\mu$ (not for $E_T$+jets analysis)
- $E_T$ from neutrino
- 2 or 3 (or 4 DØ) jets
- At least 1 $b$-tagged jet

Very small signal/background ratio after selection:

CDF Run II Preliminary, L=3.2fb$^{-1}$
Huge background!

- **S : B ≈ 1 : 20**
- Signal acceptance 2–3%
- Counting experiment not possible!
- Sophisticated multivariate methods needed

### Event Yields in 2.3 fb⁻¹ of DØ Data

| Component                  | Yield (± Error) |
|----------------------------|-----------------|
| e,μ, 2,3,4-jets, 1,2-tags combined | 223 ± 30        |
| W+jets                     | 2,647 ± 241     |
| Z+jets, dibosons           | 340 ± 61        |
| t\(\bar{t}\) pairs        | 1,142 ± 168     |
| Multijets                  | 300 ± 52        |
| **Total prediction**       | **4,652 ± 352** |
| **Data**                   | **4,519**       |

### CDF

| Analysis                  | LJ [fb⁻¹] | MJ [fb⁻¹] |
|---------------------------|-----------|-----------|
| tb + t\(\bar{t}\)        | 3.2       | 2.1       |
| W+HF                      | 1551      | 304       |
| t\(\bar{t}\)             | 686       | 185       |
| W+LF, multij.             | 778       | 679       |
| Z+jets, dibos.            | 52        | 129       |
| **Total pred.**           | **3377**  | **1404**  |
| **Observed**              | **3315**  | **1411**  |
### Analysis Strategy

#### 1. Multivariate Techniques
- **Combine** many variables into one powerful discriminant
- **Optimize** different \((\ell, N_{\text{tag}}, N_{\text{jet}})\) sub samples individually

Both DØ and CDF use: Boosted Decision Trees, Matrix Elements, Neural Networks. CDF also uses: Likelihood Functions.

#### 2. Measure Cross Section

Bayesian calculation using data and predicted discriminant distributions → posterior

#### 3. Derive Significance

Significance derived from large ensemble of pseudo-experiments
Boosted Decision Trees

- Sequence of cuts
- Events failing cut continue to be analyzed
- Adding input variables doesn’t degrade performance
- Long list of variables used
  - DØ: 64 variables, CDF: 20
- **Boosting** – create forest of trees, improves performance and stability

**BDT Results**

| $\mathcal{L}$ [fb$^{-1}$] | Significance | $\sigma_{s+t}$ [pb] |
|-----------------------------|--------------|---------------------|
| Data                        |              |                     |
| DØ                           | 2.3          | 4.3$\sigma$         |
|                             | 4.6$\sigma$  |                     |
|                             | 3.7$^{+1.0}_{-0.8}$ |                     |
| CDF Run II Preliminary      | 3.2          | 5.2$\sigma$         |
|                             | 3.5$\sigma$  |                     |
|                             | 2.1$^{+0.7}_{-0.6}$ |                     |
DØ: Bayesian NN (BNN)
- Weighted average of several hundred NNs
- 18-25 input variables

CDF: NeuroBayes Program
- Four separate NNs
- 11-18 input variables incl. jet flavour separator

NN Results

|       | L [fb⁻¹] | Significance | σ_{s+t} [pb] |
|-------|----------|--------------|--------------|
| DØ    | 2.3      | 4.1σ         | 5.2σ, 4.7^{+1.2}_{-0.9} |
| T3    | 3.2      | 5.2σ         | 3.5σ, 1.8^{+0.6}_{-0.6} |
Matrix Element

- Use parton level matrix elements
  \[ [s, t, Wb\bar{b}, Wcg, Wgg, t\bar{t}, Wugg, ggg, ...] \]
- given all reconstructed four-momenta: compute prob. for signal and bkg hypotheses
- To further improve performance:
  DØ: split sample by \( H_T \) to \( t\bar{t} \) & \( W+\text{jets} \) dominated subsets
  CDF: weight events by jet flavour separator

**ME Results**
\[
\begin{array}{llll}
\mathcal{L} & \text{Significance} & \sigma_{s+t} \\
\text{[fb}^{-1}] & \text{Exp.} & \text{Obs.} & \text{[pb]} \\
\text{DØ} & 2.3 & 4.1\sigma & 5.0\sigma & 4.3^{+1.0}_{-1.2} \\
\text{CDF} & 3.2 & 4.9\sigma & 4.3\sigma & 2.5^{+0.7}_{-0.6} \\
\end{array}
\]
Other Analyses (CDF only)

Likelihood Functions
derived using 7-10 variables
Separate likelihood optimized for the s-channel

NEW: $E_T+\text{jets}$ analysis
– Orthogonal to $\ell+\text{jets}$
– Recover events where $\ell$ not reconstructed
– Very large instrumental background!

Separate $s$ and $t$-channel cross section measurement using NN

| $E_T+\text{jets}$ | $L$ [fb$^{-1}$] | Significance | $\sigma_{s+t}$ [pb] |
|-------------------|----------------|-------------|---------------------|
|                   | 2.1            | 1.4$\sigma$ | 2.1$\sigma$         |
|                   | 3.2            | 4.0$\sigma$ | 2.4$\sigma$         |
| LFS [†]           | 3.2            | 1.1$\sigma$ | 2.0$\sigma$         |

† This is for the s-channel only

$E_T+\text{jets}$ NN output
Combination

The individual MVs are combined using:

- **Bayesian Neural Network**
  - Discriminants of the three individual analyses used as input

- **NeuroEvolution of Augmenting Topologies**:
  - NN that optimizes expected $p$-value taking systematics and binning into consideration

All 3 analyses combined:
- $\rightarrow 5\sigma$ significance!
- $\rightarrow$ Observation!

**BLUE combination** used as cross check

**Sensitivity**: $4.3\sigma \rightarrow 4.5\sigma$

All 6 analyses combined:
- $\rightarrow 5\sigma$ significance!
- $\rightarrow$ observation!

**Sensitivity**: $5.2\sigma \rightarrow 5.9\sigma$

---

**Combined Results**

| $\mathcal{L}$ [fb$^{-1}$] | Significance | $\sigma_{s+t}$ [pb] |
|---------------------------|--------------|---------------------|
|                           | Exp.         | Obs.                |
| **D0**                    | 2.3          | 4.5\sigma          | 5.0\sigma | $3.9^{+0.9}_{-0.9}$ |
| **D0**                    | 3.2          | 5.9\sigma          | 5.0\sigma | $2.3^{+0.6}_{-0.5}$ |

---

Dag Gillberg (SFU, Canada)  
Discovery of Single Top – Moriond QCD, Mar 17, 2009
Closer look in the signal region

(BNN comb. output > 0.9)

\[ (\text{BNN comb. output} > 0.9) \]

(Super Discriminant > 0.76)

---

D0 2.3 fb\(^1\)
e+\(\mu\) 2-4 jets
1-2 b-tags

Yield vs. Jet multiplicity

Yield vs. \(Q \times \eta\)

Yield vs. Missing \(E_T\) [GeV]

Yield vs. \(m_{\text{top}}\) [GeV]

---

CDF Run II Preliminary. \(L = 3.2\, \text{fb}^{-1}\)

All channels

Normalized to Prefit

---

\(Q \times \eta\)

---

Dag Gillberg (SFU, Canada)  Discovery of Single Top – Moriond QCD, Mar 17, 2009 13 / 15
Candidate events

DØ Experiment Event Display
Single Top Quark Candidate Event, 2.3 fb^{-1} Analysis

Run 223473 Evt 27278544 Sun Jul 23 19:21:41 2006
ET scale: 28 GeV
Assuming $|V_{td}|^2 + |V_{ts}|^2 \ll |V_{tb}|^2$ and pure $V-A$ and CP-conserving $Wtb$ interaction

- No assumption about number of quark families or CKM unitarity
- Since the single top cross section proportional to $|V_{tb}|^2$, $|V_{tb}|$ essentially is measured as $\sqrt{\sigma_{\text{meas}}/\sigma_{\text{SM}}}$ but more systematic uncertainties need to be considered


\[ |V_{tb}| = 1.07 \pm 0.12, \quad |V_{tb}| > 0.78 \text{ at 95\% CL} \]
Summary

First observation of single top production at the Tevatron by both DØ and CDF!

- Improved direct measurements of $|V_{tb}|$
- Results consistent with the standard model

![DØ logo]

**DØ 2.3 fb⁻¹**

| Method                  | Cross Section (pb) |
|-------------------------|--------------------|
| Decision Trees          | 3.74 ±0.95        |
| Bayesian NNs            | 4.70 ±1.18        |
| Matrix Elements         | 4.30 ±0.99        |
| BLUE Combination        | 4.16 ±0.84        |
| BNN Combination         | 3.94 ±0.88        |

March 2009

N. Kidonakis, PRD 74, 114012 (2006) $m_{top} = 170$ GeV

**CDF Preliminary Single Top Summary**

For $M_{top} = 175$ GeV/c²

| Source                                      | Value ± Error |
|---------------------------------------------|---------------|
| S-Channel Likelihood Function (3.2 fb⁻¹)   | 1.5 ± 0.9     |
| Neural Network (3.2 fb⁻¹)                  | 1.8 ± 0.6     |
| Matrix Element (3.2 fb⁻¹)                  | 2.5 ± 0.7     |
| Likelihood Function (3.2 fb⁻¹)             | 1.6 ± 0.8     |
| Boosted Decision Tree (3.2 fb⁻¹)           | 2.1 ± 0.7     |
| Combination (Lepton+Jets) (3.2 fb⁻¹)       | 2.1 ± 0.6     |
| MET+Jets, (3.2 fb⁻¹)                        | 4.9 ± 2.6     |
| Combination (All Channels) (3.2 fb⁻¹)      | 2.3 ± 0.6     |

Single Top Production Cross Section (pb)

-5 0 5
The CDF Detector

- Central Muon
- Central Calorimeter (EH)
- Wall Calorimeter (H)
- Plug Calorimeter (EH)
- Forward Muon
- Forward Calorimeter (E)
- Luminosity Monitor
- Time of Flight
- Central Outer Tracker
- Silicon Vertex Detector

Dag Gillberg (SFU, Canada)
Discovery of Single Top – Moriond QCD, Mar 17, 2009

19 / 15
### DØ

- One isolated lepton with $p_T > 15$ and $|\eta| < 1.1$ (2.0) for $e$ ($\mu$)
- Veto events with additional leptons
- 2-4 jets, with $p_T > 15$ GeV and $|\eta_{\text{det}}| < 3.4$
- 1-2 $b$-tagged jets
- Leading jet $p_T > 25$ GeV
- Leading $b$-tagged jet $p_T > 20$ GeV
- $\not{E}_T > 20$ (25) for events with 2 (3 or 4) jets
- Remove events with low $H_T(\text{alljets, } \mu, \not{E}_T)$ ($\sim < 120$ GeV) to reduce QCD
- Remove events where $\ell$ aligned/anti-aligned with $\not{E}_T$

### CDF

- One isolated lepton with $p_T > 20$ and $|\eta| < 1.6$ (not for MJ)
- Veto additional leptons
- 2-3 jets, with $p_T > 20$ GeV and $|\eta_{\text{det}}| < 2.8$
- At least one $b$-tagged jet
- $\not{E}_T > 25$ (50) for LJ (MJ)
- MJ only: leading jet $p_T > 35$, second jet $p_T > 25$ GeV
- MJ only: Cut on NN trained to characterize QCD
### Event Yields in 2.3 fb\(^{-1}\) of DØ Data

| Source       | 2 jets     | 3 jets     | 4 jets     |
|--------------|------------|------------|------------|
| s-channel \(tb\) | 62 ± 9    | 24 ± 4     | 7 ± 2      |
| t-channel \(tqb\) | 77 ± 10   | 39 ± 6     | 14 ± 3     |
| \(W + bb\)   | 678 ± 104 | 254 ± 39   | 73 ± 11    |
| \(W + c\bar{c}\) | 303 ± 48  | 130 ± 21   | 42 ± 7     |
| \(W + c\bar{j}\) | 435 ± 27  | 113 ± 7    | 24 ± 2     |
| \(W + j\bar{j}\) | 413 ± 26  | 140 ± 9    | 41 ± 3     |
| \(Z + \text{jets}\) | 141 ± 33  | 54 ± 14    | 17 ± 5     |
| Dibosons     | 89 ± 11   | 32 ± 5     | 9 ± 2      |
| \(t\bar{t} \rightarrow \ell\ell\) | 149 ± 23  | 105 ± 16   | 32 ± 6     |
| \(t\bar{t} \rightarrow \ell + \text{jets}\) | 72 ± 13   | 331 ± 51   | 452 ± 66   |
| Multijets    | 196 ± 50  | 73 ± 17    | 30 ± 6     |
| **Total prediction** | **2,615 ± 192** | **1,294 ± 107** | **742 ± 80** |
| **Data**     | 2,579     | 1,216      | 724        |
## CDF Predicted Yields

| Process        | Number of Events in 3.2 fb$^{-1}$ |
|----------------|-----------------------------------|
|                | $W + 2$ jets                      | $W + 3$ jets                      |
| s-channel      | $58.1 \pm 8.4$                   | $19.2 \pm 2.8$                   |
| t-channel      | $87.6 \pm 13.0$                  | $26.2 \pm 3.9$                   |
| $Wb\bar{b}$    | $656.9 \pm 198.0$                | $201.3 \pm 60.8$                 |
| $Wc\bar{c}$    | $292.2 \pm 90.1$                 | $98.1 \pm 30.2$                  |
| $Wcj$          | $250.4 \pm 77.2$                 | $52.1 \pm 16.0$                  |
| Mistags        | $501.3 \pm 69.6$                 | $151.9 \pm 21.4$                 |
| non-$W$        | $89.6 \pm 35.8$                  | $35.1 \pm 14.0$                  |
| $WW$           | $58.5 \pm 6.6$                   | $21.2 \pm 2.4$                   |
| $WZ$           | $28.9 \pm 2.4$                   | $8.5 \pm 0.7$                    |
| $ZZ$           | $0.9 \pm 0.1$                    | $0.4 \pm 0.0$                    |
| $Z + jets$     | $36.5 \pm 5.6$                   | $15.6 \pm 2.4$                   |
| $t\bar{t}$ dilepton | $69.2 \pm 10.0$                | $60.2 \pm 8.7$                   |
| $t\bar{t}$ non-dilepton | $134.9 \pm 19.6$                | $421.8 \pm 61.1$                 |
| Total signal   | $145.7 \pm 21.4$                 | $45.4 \pm 6.7$                   |
| Total prediction | $2265.0 \pm 375.4$              | $1111.5 \pm 129.5$              |
| Observed in data | $2229$                              | $1086$                             |
## Systematics Uncertainties

| Systematic                                | Rate         | Shape |
|-------------------------------------------|--------------|-------|
| Jet energy scale                          | 0...16%      | ✓     |
| Initial state radiation                   | 0...11%      | ✓     |
| Final state radiation                     | 0...15%      | ✓     |
| Parton distribution functions             | 2...3%       | ✓     |
| Monte Carlo generator                     | 1...5%       | —     |
| Event detection efficiency                | 0...9%       | —     |
| Luminosity                                | 6%           | —     |
| NN flavor separator                       | —            | ✓     |
| Mistag model                              | —            | ✓     |
| Non-W model                               | —            | ✓     |
| ALPGEN Q²                                 | —            | ✓     |
| MC Modeling ($\Delta R, \eta(j_2)$)       | —            | ✓     |
| $W b\bar{b} + W c\bar{c}$ normalization  | 30%          | —     |
| $W c$ normalization                       | 30%          | —     |
| Mistag normalization                      | 17...29%     | —     |
| Top Mass - top-pair normalization         | 23%          | ✓     |

### Components for normalization

- Integrated luminosity: 6.1%
- $tt$ cross section: 12.7%
- $Z+$jets and dibosons cross section: 5.8%
- Branching fractions: 1.5%
- Parton distribution functions (signal only): 3.0%
- Triggers: 5.0%
- Instantaneous luminosity reweighting: 1.0%
- Primary vertex selection: 1.4%
- Lepton identification: 2.5%
- Jet fragmentation: (0.7–4.0)%
- Initial-state and final-state radiation: (0.6–12.6)%
- $b$-jet fragmentation: 2.0%
- Jet reconstruction and identification: 1.0%
- Jet energy resolution: 4.0%
- $W$+$jets$ and $Z$+$jets$ heavy flavor correction: 13.7%
- Multijets normalization to data: (30–54)%
- Monte Carlo and multijets statistics: (0.5–16)%

### Components for normalization and shape

- Jet energy scale for signal: (1.1–13.1)%
- Jet energy scale for total background: (0.1–2.1)%
- $b$ tagging for single-tagged: (2.1–7.0)%
- $b$ tagging for double-tagged: (9.0–11.4)%

### Component for shape only

- ALPGEN reweighting: —
DØ Systematics Uncertainties – all backgrounds

DØ Run II, 2.3 fb⁻¹

Total background uncertainty

Returned by Bayesian binned likelihood fit to data – all systematics and their correlations taken into account and constraints from the data
DØ Systematics Uncertainties – $Wb\bar{b}$

Returned by Bayesian binned likelihood fit to data – all systematics and their correlations taken into account and constraints from the data
Data-background agreement checked in $W+\text{jets}$ and $t\bar{t}$ enriched control samples
DØ Cross-checks and Linearity tests

DØ Run II Prelim. 2.3 fb⁻¹
p17+p20 e+μ channel
1 b-tags
2 jets

DØ Run II Prelim. 2.3 fb⁻¹
p17+p20 e+μ channel
1-2 b-tags
4 jets

Boosted Decision Trees
Slope = 6.994 ± 0.003
Intercept = −4.016 ± 0.018

Input t+b+q Cross Section [pb]

Bayesian Neural Networks
Slope = 0.593 ± 0.003
Intercept = 0.032 ± 0.018

Input t+b+q Cross Section [pb]

Matrix Elements
Slope = 0.086 ± 0.003
Intercept = −0.126 ± 0.018

Input t+b+q Cross Section [pb]
CDF “Golden Event”

Event taken
2007/05/27

light jet $E_T = 52$ GeV

electron $P_T = 66$ GeV/c

b-tagged jet $E_T = 38$ GeV

March 10th, 2009

Rainer Wallny - Observation of Electroweak Single Top Quark Production

Dag Gillberg (SFU, Canada)
CDF – Signal Enriched Region

Signal Features

SD > 0.72

Purity S/B ~ 1.2
Using a large ensemble of pseudo-datasets without any single top content:

**Expected p-value**
Fraction of 0-signal pseudo-datasets in which we measure at least the standard model cross section (2.86/3.46 pb)

**Observed p-value**
Fraction of 0-signal pseudo-datasets in which we measure at least the observed cross section
DØ Significance

**Discovery of Single Top**

- **67.8M pseudo experiments**
- 215 above SM cross section
- p-value: \((3.2 \pm 0.2) \times 10^{-6}\)
- Expected Significance: 
  \(4.51^{+0.01}_{-0.01}\) sigma

---

**DØ Combination**

- **67.8M pseudo-datasets (background-only)**
- 17 above measured cross section
- p-value: \(2.5 \times 10^{-7}\)
- Observed significance: 
  \(= 5.03 \sigma\)

\(\sigma_{\text{meas}} = 3.94 \text{ pb}\)
Measuring a cross section

Probability to observe data distribution $D$, expecting $y$:

$$y = \alpha l \sigma + \sum_{s=1}^{N} b_s \equiv a \sigma + \sum_{s=1}^{N} b_s$$

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

The cross section is obtained

$$Post(\sigma|D) \equiv P(\sigma|D) \propto \int_{a}^{b} \int_{b} P(D|\sigma, a, b) Prior(\sigma) Prior(a, b)$$

- Bayesian posterior probability density
- Shape and normalization systematics treated as nuisance parameters
- Correlations between uncertainties properly accounted for
- Flat prior in signal cross section
Bayesian posterior density
CDF Cross Section Measurement

CDF Run II Preliminary, $L = 3.2 \text{ fb}^{-1}$

$\sigma_{\text{Single Top}} = 2.3^{+0.6}_{-0.5} \text{ pb}$
Direct measurement of $|V_{tb}|$

- **General form of $Wtb$ vertex:**

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

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

- **No need to assume only three quark families or CKM matrix unitarity**
  (unlike for previous measurements using $tt$ decays)

- **Measure the strength of the $V-A$ coupling, $|V_{tb} f_1^L|$, which can be > 1**

**Additional theoretical uncertainties**

|                          | $tb$  | $tqb$ |
|--------------------------|-------|-------|
| Top mass                 | 13 %  | 8.5 % |
| Scale                    | 5.4 % | 4.0 % |
| PDF                      | 4.3 % | 10 %  |
| $\alpha_s$               | 1.4 % | 0.01 %|
DØ | $V_{tb}$ | measurement

DØ 2.3 fb$^{-1}$

$|V_{tb}f_L^1| = 1.07 \pm 0.12$

flat prior $\geq 0$

| $V_{tb}f_L^1|^{2}$

DØ 2.3 fb$^{-1}$

$|V_{tb}| > 0.78$

at 95% CL

$0 \leq$ flat prior $\leq 1$

| $V_{tb}|^{2}$