Evidence for production of single top quarks at DØ and a first direct measurement of $|V_{tb}|$

- Electroweak production of top quarks at DØ
- 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 $|V_{tb}|$

Summary

Arán García-Bellido
The highest energy particle accelerator in the world!

Proton-antiproton collider

**Run I 1992-1995**
Top quark discovered!

**Run II 2001-09(?)**
$\sqrt{s} = 1.96$ TeV
$\Delta t = 396$ns
$>2fb^{-1}$ delivered
Peak Lum: $2 \cdot 10^{32} \text{cm}^{-2}\text{s}^{-1}$
DØ for Run II

Tracker: Si+Fiber+Preshowers

2T solenoid magnet

γ = 0
γ = 1
γ = 2
γ = 3

Muon Scintillators

Muon Chambers

Shielding

Calorimeter

Toroid

3 layer muon system

electronics

protons

antiprotons

Arán García-Bellido

First evidence for single top
Data taking

1.80 fb⁻¹ on tape
Overall eff: ~85%
Ineff due to:
~5% FEB
~5% losses in store & run transitions
~5% incidentals
DØ L0 of Silicon is a success!
Electron cooling working nicely
Peak lumi of ~2.5E32

0.9 fb⁻¹ in this talk

Delivered
Recorded
Top quark physics

The top quark is a very special fermion:

- Heaviest known particle: \(171.4 \pm 2.1\) GeV
- \(m_t \sim v/\sqrt{2}, \lambda_t \sim 1\) \(\rightarrow\) 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

We still don't know: spin, width, lifetime
We know the mass, cross section, charge and its \(\text{BR}(t \rightarrow Wb) \sim 1\)
Plenty of room for new physics
Top quark electroweak production

s-channel (tb)

\[ \sigma_{\text{NLO}} = 0.88 \pm 0.11 \text{ pb} \]

Current limits @ 95% C.L.:

DØ (370pb\(^{-1}\)) \( \sigma_{tb} < 5.0 \text{ pb} \)

CDF (700pb\(^{-1}\)) \( \sigma_{tb} < 3.1 \text{ pb} \)


\text{t}-channel (tqb)

\[ \sigma_{\text{NLO}} = 1.98 \pm 0.25 \text{ pb} \]

Current limits @ 95% C.L.:

DØ (370pb\(^{-1}\)) \( \sigma_{tqb} < 4.4 \text{ pb} \)

CDF (700pb\(^{-1}\)) \( \sigma_{tqb} < 3.2 \text{ pb} \)
Why search for single top?

- Access $W$-t-$b$ coupling
- measure $V_{tb}$ directly $\rightarrow$ more on this later
- test unitarity of CKM

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

Source of polarized top quarks

Extract small signal out of a large background

DØ search: hep-ex/0607102
Arán García-Bellido

DØ search: t.b.s. PRL
First evidence for single top
A big challenge!

~20 single top events produced per day

But huge backgrounds!

We have benefited greatly from the following improvements for this analysis:

- Background model improvements (PS↔ME matching: MLM)
- Fully reprocessed dataset: new calibrations, jet thresholds, JES,...
- New more efficient NN b-tagger
- Split channels by jet multiplicity
- Combined s+t search added (SM s:t ratio is assumed)
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 tight (no loose) lepton:
  - $e$: $p_T > 15 \text{ GeV}$ and $|\eta^{\text{det}}| < 1.1$
  - $\mu$: $p_T > 18 \text{ GeV}$ and $|\eta^{\text{det}}| < 2.0$
- MET $> 15 \text{ GeV}$
- 2-4 jets: $p_T > 15 \text{ GeV}$ and $|\eta^{\text{det}}| < 3.4$
  - Leading jet: $p_T > 25 \text{ GeV}$ ; $|\eta^{\text{det}}| < 2.5$
  - Second leading jet: $p_T > 20 \text{ GeV}$
- One or two b-tagged jets
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: $\sim 0.5\%$
Background modeling

- **W+jets**: $\sim o(1000)$ pb
  - Distributions from Alpgen 2.0
  - Normalization from data
  - Heavy flavor fractions from data

- **Top pairs**: $\sim 7$ pb
  - Topologies: dilepton and $\ell +$jets
  - Use Alpgen 2.0 with MLM matching
  - Normalize to NNLO $\sigma$

- **Multijet events (misidentified lepton)**
  - From data
Agreement before tagging

- Normalize $W$+jets and QCD yields to data before tagging
- Check 90 variables (in $e$, $\mu$ x 2, 3, 4 jets)
- Good description of data
Yields after event selection

<table>
<thead>
<tr>
<th>Source</th>
<th>2 jets</th>
<th>3 jets</th>
<th>4 jets</th>
</tr>
</thead>
<tbody>
<tr>
<td>$tb$</td>
<td>16 ± 3</td>
<td>8 ± 2</td>
<td>2 ± 1</td>
</tr>
<tr>
<td>$tqb$</td>
<td>20 ± 4</td>
<td>12 ± 3</td>
<td>4 ± 1</td>
</tr>
<tr>
<td>$t\bar{t} \rightarrow ll$</td>
<td>39 ± 9</td>
<td>32 ± 7</td>
<td>11 ± 3</td>
</tr>
<tr>
<td>$t\bar{t} \rightarrow l+\text{jets}$</td>
<td>20 ± 5</td>
<td>103 ± 25</td>
<td>143 ± 33</td>
</tr>
<tr>
<td>$W+b\bar{b}$</td>
<td>261 ± 55</td>
<td>120 ± 24</td>
<td>35 ± 7</td>
</tr>
<tr>
<td>$W+c\bar{c}$</td>
<td>151 ± 31</td>
<td>85 ± 17</td>
<td>23 ± 5</td>
</tr>
<tr>
<td>$W+jj$</td>
<td>119 ± 25</td>
<td>43 ± 9</td>
<td>12 ± 2</td>
</tr>
<tr>
<td>Multijets</td>
<td>95 ± 19</td>
<td>77 ± 15</td>
<td>29 ± 6</td>
</tr>
<tr>
<td>Total background</td>
<td>686 ± 131</td>
<td>460 ± 75</td>
<td>253 ± 42</td>
</tr>
<tr>
<td>Data</td>
<td>697</td>
<td>455</td>
<td>246</td>
</tr>
</tbody>
</table>

- Optimized the selection to maximize acceptance
- Allow a lot of background at this stage!
- Then use multiple distributions to separate signal-background

Arán García-Bellido
First evidence for single top
Event selection and S:B

<table>
<thead>
<tr>
<th>Electron + Muon</th>
<th>1 jet</th>
<th>2 jets</th>
<th>3 jets</th>
<th>4 jets</th>
<th>≥ 5 jets</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 tags</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10%</td>
<td>25%</td>
<td>12%</td>
<td>3%</td>
<td>1%</td>
<td></td>
</tr>
<tr>
<td>1:3,200</td>
<td>1:390</td>
<td>1:300</td>
<td>1:270</td>
<td>1:230</td>
<td></td>
</tr>
<tr>
<td>1 tag</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5%</td>
<td>21%</td>
<td>11%</td>
<td>3%</td>
<td>1%</td>
<td></td>
</tr>
<tr>
<td>1:100</td>
<td>1:20</td>
<td>1:25</td>
<td>1:40</td>
<td>1:53</td>
<td></td>
</tr>
<tr>
<td>2 tags</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3%</td>
<td>2%</td>
<td>1%</td>
<td>0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1:11</td>
<td>1:15</td>
<td>1:38</td>
<td>1:43</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Percentage of single top $tb+tqb$ selected events and S:B ratio (white squares = no plans to analyze)
Systematic uncertainties

- Uncertainties are assigned per background, jet multiplicity, lepton channel, and number of tags

- Uncertainties that affect both the **normalization** and the **shapes**: JES and tag rate functions

- Correlations between channels and sources are taken into account

---

**Examples of Relative Systematic Uncertainties**

<table>
<thead>
<tr>
<th>Source</th>
<th>Uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t\bar{t}$ cross section</td>
<td>18%</td>
</tr>
<tr>
<td>Luminosity</td>
<td>6%</td>
</tr>
<tr>
<td>Electron trigger</td>
<td>3%</td>
</tr>
<tr>
<td>Muon trigger</td>
<td>6%</td>
</tr>
<tr>
<td>Jet energy scale</td>
<td>wide range</td>
</tr>
<tr>
<td>Jet fragmentation</td>
<td>5–7%</td>
</tr>
<tr>
<td>Heavy flavor ratio</td>
<td>30%</td>
</tr>
<tr>
<td>Tag-rate functions</td>
<td>2–16%</td>
</tr>
</tbody>
</table>
And check 1000s of plots again...
Analysis methods

Once we understand our data, need to measure the signal.

We cannot use simple cuts to extract the signal:

use **multivariate techniques**

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

- Decision Trees
- Matrix Elements
- Bayesian NNs

- DT and BNN use same pool of discriminating variables
- ME method uses 4-vectors of reconstructed objects
- Optimized separately for s-channel, t-channel and s+t
- Test response and robustness with ensemble testing
Decision Trees
Machine learning technique widely used in social sciences
Idea: recover events that fail criteria in cut-based analysis

- Start with all events (first node)
- For each variable, find the splitting value with best separation between children
- Select best variable and cut: produce Pass and Failed branches
- Repeat recursively on each node
- Stop when improvement stops or when too few events left
- Terminal node: leaf with purity = $N_s/(N_s+N_B)$
- Output: purity for each event

Machine learning technique widely used in social sciences
Idea: recover events that fail criteria in cut-based analysis

- Start with all events (first node)
- For each variable, find the splitting value with best separation between children
- Select best variable and cut: produce Pass and Failed branches
- Repeat recursively on each node
- Stop when improvement stops or when too few events left
- Terminal node: leaf with purity = $N_s/(N_s+N_B)$
- Output: purity for each event
Decision Trees + Boosting

Boosting is a recent technique to improve the performance of any weak classifier: recently used in DTs by GLAST and MiniBooNE

**AdaBoost algorithm**: adaptive boosting

1) Train a tree $T_k$
2) Check which events are **misclassified** by $T_k$
3) Derive tree weight $\alpha_k$
4) Increase weight of misclassified events
5) Train again to build $T_{k+1}$

- We have trained 36 separate trees: $(s, t, s+t)x(e,\mu)x(2,3,4\text{ jets})x(1,2\text{ tags})$
- Use 1/3 of MC events for training
- For each signal, train against sum of backgrounds
- Signal leaf if purity $>0.5$; Minimum leaf size $=100$ events; Goodness of split: Gini factor; Adaboost $\beta = 0.2$; boosting cycles $= 20$
Decision Trees: 49 variables

Object Kinematics
$p_T(jet1)$
$p_T(jet2)$
$p_T(jet3)$
$p_T(jet4)$
$p_T(best1)$
$p_T(notbest1)$
$p_T(notbest2)$
$p_T(tag1)$
$p_T(untag1)$
$p_T(untag2)$

Angular Correlations
$\Delta R(jet1,jet2)$
$\cos(best1,lepton)_{besttop}$
$\cos(best1,notbest1)_{besttop}$
$\cos(tag1,alljets)_{alljets}$
$\cos(tag1,lepton)_{btaggedtop}$
$\cos(jet1,alljets)_{alljets}$
$\cos(jet1,lepton)_{btaggedtop}$
$\cos(jet2,alljets)_{alljets}$
$\cos(jet2,lepton)_{btaggedtop}$
$\cos(lepton,Q(lepton) \times z)_{besttop}$
$\cos(lepton,besttopframe)_{besttopCMframe}$
$\cos(lepton,btaggedtopframe)_{btaggedtopCMframe}$
$\cos(notbest,alljets)_{alljets}$
$\cos(notbest,lepton)_{besttop}$
$\cos(untag1,alljets)_{alljets}$
$\cos(untag1,lepton)_{btaggedtop}$

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

- Adding variables does not degrade performance
- Tested shorter lists, lose some sensitivity
- Same list used for all channels
Matrix Elements method

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

Calculate an event probability for signal and background hypothesis

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

**Parton distribution functions CTEQ6**

**Differential cross section (LO ME from Madgraph)**

**Transfer Function**: maps parton level (y) to reconstructed variables (x)

Uses the 4-vectors of all reconstructed \(\ell\)s and jets

This analysis: 2&3 jet events only, match partons to jets

Apply b-tagging information

Need to integrate over 4 independent variables: assume angles well measured, known masses, momentum and energy conservation
ME discriminant

Define discriminant based on event probabilities for signal and background

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

- In 2 jet events: use ME for Wbg, Wcg and Wgg backgrounds
- In 3 jet events: use ME for Wbbg background
- No ttbar ME used thus far: no separation in the 3\(^{rd}\) jet bin!
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!
Measuring the cross section

- We form a binned likelihood from the discriminant outputs.

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

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

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

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

$$Post(\sigma|D) \equiv P(\sigma|D) \propto \int_a^b \int_P(D|\sigma, a, b) Prior(\sigma) Prior(a, b)$$

- Shape and normalization systematics treated as nuisance parameters.
- Correlations between uncertainties properly accounted for.
- Flat prior in signal cross section.
Ensemble testing

To verify that all this machinery is working properly, we test with many sets of **pseudo-data**

Wonderful tool to test analysis methods! Run DØ experiment 100s of times

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

Generated ensembles include:

1) 0-signal ensemble \( \sigma_{s+t} = 0 \) pb
2) SM ensemble \( \sigma_{s+t} = 2.9 \) pb
3) “Mystery” ensembles to test analyzers \( \sigma_{s+t} = ?? \) pb
4) Ensemble at measured cross-section \( \sigma_{s+t} = \sigma_{\text{measured}} \)
5) A high luminosity ensemble

Each analysis tests linearity of “response” to single top
Responses

Using the ensemble tests:
- SM ensemble is returned at the right value
- “Mystery” ensembles are unraveled
- Linear response is achieved
Sensitivity and Significance

We have used our 0-signal ensembles to determine a significance for each measurement.

**Expected p-value:** the fraction of 0-signal pseudo-datasets in which we measure at least 2.9 pb.

**Observed p-value:** the fraction of 0-signal pseudo-datasets in which we measure at least the measured cross section.

We also use the SM ensembles to see how compatible our measured value is with the SM.
Expected p-values

**Decision Trees**
- p-value 1.9%

**Matrix Elements**
- p-value 3.7%

**Bayesian NN**
- p-value 9.7%

Arán García-Bellido
First evidence for single top
DT cross check samples

Check the description of the data in the DT output

- W+jets: 2 jets and $H_T(\text{lepton,MET,alljets}) < 175$ GeV
- tt: 4 jets and $H_T(\text{lepton,MET,alljets}) > 300$ GeV
ME cross check samples

Check the description of the data in the ME output

- Soft W+jets: $H_T(\text{lepton},\text{MET},\text{alljets}) < 175$ GeV
- Hard W+jets: $H_T(\text{lepton},\text{MET},\text{alljets}) > 300$ GeV
First evidence for single top

Least sensitive (a-priori) analysis sees a $2.4\sigma$ effect!
Matrix Elements observed results

2.9σ excess!
SM compatibility = 21%
Decision Trees observed results

\( \sigma = 4.9 \pm 1.4 \text{ pb} \)

3.4\( \sigma \) excess!

SM compatibility = 11%
Arán García-Bellido 

First evidence for single top
Excess in the high discriminant regions
ME event characteristics

**ME Discriminant < 0.4**

![Graph showing b-Tagged Top Mass distribution with discriminant < 0.4]

- Data: Various processes labeled as $tb$, $tq$, $qg$, $Wb$, $Wc$, $Wj$, $QCD$, $H \rightarrow \tau \tau$, $H \rightarrow \ell\ell$, $H \rightarrow \ell\ell$.
- Obs: 441
- Bkg: 441

**ME Discriminant > 0.7**

![Graph showing b-Tagged Top Mass distribution with discriminant > 0.7]

- Data: Various processes labeled as $tb$, $tq$, $qg$, $Wb$, $Wc$, $Wj$, $QCD$, $H \rightarrow \tau \tau$, $H \rightarrow \ell\ell$, $H \rightarrow \ell\ell$.
- Obs: 161
- Bkg: 138

Arán García-Bellido

First evidence for single top
DT event characteristics

DT Discriminant < 0.3

DT Discriminant > 0.55

DT Discriminant > 0.65

Arán García-Bellido

First evidence for single top
A candidate event

Run 177034 Evt 10482925

Triggers:

- 1 MET
- 1 mu particle

ET (GeV)

Bins: 105
Mean: 1.15
Rms: 3.85
Min: 0.00933
Max: 27.4

mu particle et: 27.2
MET et: 28

Arán García-Bellido

First evidence for single top
s+t summary: all methods

DØ Run II

0.9 fb⁻¹

Decision trees

Matrix elements

Bayesian NNs

Z. Sullivan PRD 70, 114012 (2004), m_t = 175 GeV

σ(pp → tb+tqb) [pb]

4.9 ± 1.4 pb

4.6 ± 1.8 pb

5.0 ± 1.9 pb

Arán García-Bellido

First evidence for single top
Correlations

- Take the 50 highest ranked data events in each method and look for overlap:

<table>
<thead>
<tr>
<th>Technique</th>
<th>Electron</th>
<th>Muon</th>
</tr>
</thead>
<tbody>
<tr>
<td>DT vs ME</td>
<td>52%</td>
<td>58%</td>
</tr>
<tr>
<td>DT vs BNN</td>
<td>56%</td>
<td>48%</td>
</tr>
<tr>
<td>ME vs BNN</td>
<td>46%</td>
<td>52%</td>
</tr>
</tbody>
</table>

- Calculate the linear correlation between the measured cross sections in the same 400 members of the SM ensemble:

<table>
<thead>
<tr>
<th></th>
<th>DT</th>
<th>ME</th>
<th>BNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>DT</td>
<td>100%</td>
<td>39%</td>
<td>57%</td>
</tr>
<tr>
<td>ME</td>
<td>100%</td>
<td>100%</td>
<td>29%</td>
</tr>
<tr>
<td>BNN</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>
Weak interaction eigenstates and mass eigenstates are not the same: there is **mixing** between quarks $\rightarrow$ CKM matrix

In SM: top must decay to $W$ and $d$, $s$ or $b$ quark

- $V_{td}^2 + V_{ts}^2 + V_{tb}^2 = 1$
- Strong constraints on $V_{td}$ and $V_{ts}$: $V_{tb} > 0.998$
- Assuming unitarity and 3 generations: $B(t \rightarrow Wb) \sim 100\%$

If there is new physics:

- $V_{td}^2 + V_{ts}^2 + V_{tb}^2 < 1$
- No constraint on $V_{tb}$
- Interactions between the top quark and weak gauge bosons are extremely interesting!
Measuring $|V_{tb}|$

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

- Use the same infrastructure as for the cross section measurement, but make a posterior in $|V_{tb}|^2$

Additional theoretical errors are needed

Caveat: assume SM decays

Most general $Wtb$ vertex:

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

SM: $f_1^L = 1$ and $f_1^R = f_2^L = f_2^R = 0 \Rightarrow$ CP is conserved

We are effectively measuring the strength of the V-A coupling: $|V_{tb} f_1^L|$, which can be $>1$
First direct measurement of $|V_{tb}|$

$|V_{tb}| = 1.3 \pm 0.2$

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

(assuming: $f_1^L = 1$)

This measurement does not assume 3 generations or unitarity

Arán García-Bellido  First evidence for single top
Conclusions

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

(hep-ex/0612052 submitted to PRL)

$\sigma(s+t) = 4.9 \pm 1.4$ pb

$3.4\sigma$ significance!

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

- Working on the combination and more!
- Expand to searches of new phenomena
- We now have double the data to analyze!
Extra slides
CDF's latest results

Single Top Production Cross Section

Matrix Element
\( (L = 955 \, \text{pb}^{-1}) \)

2.7 ± \( 1.5 \) \( \frac{1.3}{1.3} \)

Likelihood Function
\( (L = 955 \, \text{pb}^{-1}) \)

0.3 ± \( 1.2 \) \( \frac{0.3}{0.3} \)

Neural Network
\( (L = 695 \, \text{pb}^{-1}) \)

0.8 ± \( 1.3 \) \( \frac{0.9}{0.9} \)
Single top in a couple of years

- By 2007 we will have observed single top and measured its cross section to \( \sim 10\% \) at the Tevatron

- Then the LHC will start with huge production rates:
  \[
  \sigma_s = 10.6 \pm 1.1 \text{ pb} \quad \sigma_t = 246.6 \pm 0.25 \text{ pb} \quad \sigma_{tW} = 62.0^{+16.6}_{-3.6} \text{ pb}
  \]

- Observe all three channels (s-channel will be tough)
- tW mode offers new window into top physics
- Measure \( V_{tb} \) to a few %
- Large samples: study properties
Preparing the way for the LHC

Studies at the Tevatron will help the LHC:

- **Wbb measurement** (will also help WH search) \( (DØ: \text{hep-ex/0410062}) \)
  
  Current limit at 4.6 pb for \( p_T(b) > 20 \text{GeV} \)

- **In general, W+jets background determination techniques**
  
  \( tt \) will be main background, but large uncertainties come from \( W+jets \)
  
  Effect of jet vetoes \( (N_{jet} = 2) \), check other methods planned in LHC analyses

- **Study charge asymmetries** \( (\text{Bowen, Ellis, Strassler: hep-ph/0412223}) \)
  
  Signal shows asymmetry in \( (Q_\ell \times \eta_j, Q_\ell \times \eta_\ell) \) plane at TeV

- **Study kinematics of forward jets in t-channel** (\( WW \rightarrow H \) at LHC)

- **Even measure asymmetry in production rate** \( (\text{Yuan: hep-ph/9412214}) \)
  
  (probe CP-violation in the top sector):

  \[
  A_t = \frac{\sigma(p \bar{p} \rightarrow tX) - \sigma(p \bar{p} \rightarrow \bar{t} X)}{\sigma(p \bar{p} \rightarrow tX) + \sigma(p \bar{p} \rightarrow \bar{t} X)}
  \]

TeV4LHC workshop report to appear soon
Arán García-Bellido

First evidence for single top

Tevatron luminosity prospects

<table>
<thead>
<tr>
<th>Fiscal Year</th>
<th>30mA/hr (fb⁻¹)</th>
<th>25mA/hr (fb⁻¹)</th>
<th>20mA/hr (fb⁻¹)</th>
<th>15mA/hr (fb⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FY03</td>
<td>0.33</td>
<td>0.33</td>
<td>0.33</td>
<td>0.33</td>
</tr>
<tr>
<td>FY04</td>
<td>0.34</td>
<td>0.34</td>
<td>0.34</td>
<td>0.34</td>
</tr>
<tr>
<td>FY05</td>
<td>0.61</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>FY06</td>
<td>0.87</td>
<td>0.70</td>
<td>0.67</td>
<td>0.63</td>
</tr>
<tr>
<td>FY07</td>
<td>1.81</td>
<td>1.25</td>
<td>0.98</td>
<td>0.74</td>
</tr>
<tr>
<td>FY08</td>
<td>1.95</td>
<td>1.52</td>
<td>1.03</td>
<td>0.72</td>
</tr>
<tr>
<td>FY09</td>
<td>2.14</td>
<td>1.80</td>
<td>1.24</td>
<td>0.86</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: $B \rightarrow N_{\text{observed}}, A \rightarrow n_{\text{predicted}} = n_{\text{signal}} + n_{\text{bkgd}}, n_s = \text{Acc} \times L \times \sigma$

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

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

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

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

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

$$P(n_{\text{predicted}}|N_{\text{observed}}) = \frac{1}{Z} L(N_{\text{observed}}|n_{\text{predicted}}) \prod(n_{\text{pred}})$$
The goal is to obtain $\sigma_s$, $\sigma_t$, and $\sigma_{s+t}$, without any SM assumption. 
Previously we have used $\sigma_s^{SM}$ to derive $\sigma_t$ and vice versa. 
As before, use likelihood from 2D discriminant output. 
Float $\sigma_s$ and $\sigma_t$ and consider flat priors.

Build Likelihood $\equiv L(\sigma_s, \sigma_t)$
Obtain Posterior $\equiv P(\sigma_s, \sigma_t; n_{\text{obs}})$

For the combined limit: 
replace: $s \rightarrow z-t$ where $z=s+t$ at the Likelihood level
Additional constraint on priors: $t \leq z \rightarrow$ the prior for $t$ depends on $z$

$\sigma_{x}^{95} = \int_{0}^{\sigma_{x}^{95}} P(\sigma_x; n_{\text{obs}}) d\sigma_x$
Non-SM couplings

Top is a good place to look for deviations from SM:
- $\sigma$ under control, one dominant decay $t \rightarrow Wb$, no top hadrons,...

Generalized Lagrangian for the Wtb interaction \((\text{hep-ph/0503040})\):

$$ \mathcal{L}_{tbW} = \frac{g}{\sqrt{2}} W^-_\mu \bar{b} \gamma^\mu \left( f_1^L P_L + f_1^R P_R \right) t $$

$$ - \frac{g}{\sqrt{2} M_W} \partial_\nu W^-_\mu \bar{b} \sigma^{\mu\nu} \left( f_2^L P_L + f_2^R P_R \right) t + \text{h.c.} $$

f_1: “vector”-like
f_2: “tensor”-like

$P_{R(L)} = (1 \pm \gamma_5)/2$

In SM: $f_1^L = V_{tb} \sim 1$;

$\begin{align*}
    f_1^R &= f_2^L = f_2^R = 0
\end{align*}$

Effective single top production cross section:

There are strong bounds on tensor couplings:
- from unitarity $|f_2| < 0.6$, and from $b \rightarrow s\gamma$: $|f_2^L| < 0.004$

But Tevatron can set direct limits. The goal is:
- Set limits simultaneously on all four couplings
- Set individual limits
Non-SM couplings strategy

$f_1^L$ and $f_1^R$ have same $p_T$ distributions
Angular variables and spin are different

Separate data into $s$-channel (2 tags) and $t$-channel (1 tag + 1 untag) samples based on NN output

Top quark spin correlations separate between $L$ and $R$ couplings

$tb$: Helicity basis $\theta$ (lepton, top direction)

tqb: Optimal basis $\theta$ (lepton, pbar)

Use flat prior for four square terms:

$|f_1^L|^2$, $|f_1^R|^2$, $|c_1 f_1^L + f_2^R|^2$, $|c_1 f_1^R + f_2^L|^2$

$c_1$ is a fixed constant

Obtain limits on these four terms

Arán García-Bellido

First evidence for single top
Signal modeling

Have to get the t-channel right:
Avoid double counting when different diagrams produce same final states in different kinematic regions
Use ZTOP as NLO benchmark http://home.fnal.gov/~zack/ZTOP

▶ DØ: “Effective” NLO CompHEP (also used in CMS)
Match $2 \to 2$ and $2 \to 3$ processes using $b$ $p_T$ for cross over, normalize to NLO $\sigma$

Resulting distributions agree well with ZTOP & MCFM

▶ Recently available: MC@NLO, Alpgen 2, C.-P. Yuan et al.

Aràn García-Bellido

First evidence for single top
**W+jets normalization**

- Find fractions of real and fake isolated $\ell$ in the data before b-tagging. Split samples in loose and tight isolation:
  \[ N_{\text{loose}} = N_{\text{fake}} + N_{\text{real}} \]
  \[ N_{\text{tight}} = \varepsilon_{\text{fake}} N_{\text{fake}} + \varepsilon_{\text{real}} N_{\text{real}} \]

Obtain: $N_{\text{real}}$ and $N_{\text{fake}}$

- Normalize the MC Wjj and Wbb samples to the real $\ell$ yield found in data, after correcting for the presence of $t\bar{t}$ events:
  \[ \varepsilon_{\text{real}} N_{\text{real}}^{\text{loose}} = SF \left[ Y(Wjj) + Y(Wb\bar{b}) + Y(Wc\bar{c}) \right] + Y(t\bar{t}) \]

  \[ SF = 1.4 \]

- The sum $Y(Wjj)+Y(Wbb)+Y(Wcc)$ is done according to the ratio of $(Wbb+Wcc)/Wjj$ found in 0-tag data $\rightarrow 1.5\pm0.5$

- Then apply b-tagging
  - Greatly reduce $W+$jets background ($Wbb \sim 1\%$ of $Wjj$)
  - Shift distributions, changes flavor composition
Wbb and Wcc fraction

- We use our own data to derive the Wbb+Wcc fraction: something very close to 1.5 is needed to describe our data.
- This is not a measurement of Wbb, but a fraction determination. The full W+jets yield is scaled to data.
- Until we have our own measurement, this is the best we can do.

![Graph](image)

\[ \alpha = 1.51 \pm 0.04 \]

<table>
<thead>
<tr>
<th>Scale Factor $\alpha$ to Match Heavy Flavor Fraction to Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electron Channel</td>
</tr>
<tr>
<td>0 tags</td>
</tr>
<tr>
<td>1 tag</td>
</tr>
<tr>
<td>2 tags</td>
</tr>
<tr>
<td>1 jet 1.53 ± 0.10 1.48 ± 0.10 1.50 ± 0.20 1.72 ± 0.40</td>
</tr>
<tr>
<td>2 jets 1.29 ± 0.10 1.58 ± 0.10 1.40 ± 0.20 0.69 ± 0.60</td>
</tr>
<tr>
<td>3 jets 1.50 ± 0.20</td>
</tr>
<tr>
<td>4 jets 1.72 ± 0.40</td>
</tr>
<tr>
<td>Muon Channel</td>
</tr>
<tr>
<td>0 tags</td>
</tr>
<tr>
<td>1 tag</td>
</tr>
<tr>
<td>2 tags</td>
</tr>
<tr>
<td>1 jet 1.54 ± 0.10 1.50 ± 0.10 1.52 ± 0.10 1.38 ± 0.20</td>
</tr>
<tr>
<td>2 jets 1.11 ± 0.10 1.52 ± 0.10 1.32 ± 0.20 1.86 ± 0.50</td>
</tr>
<tr>
<td>3 jets 1.40 ± 0.40</td>
</tr>
<tr>
<td>4 jets 2.46 ± 0.90 3.78 ± 2.80</td>
</tr>
</tbody>
</table>

Arán García-Bellido

First evidence for single top
Wbb shapes

The kinematic distributions of the Wbb samples used are in gross disagreement with LO samples using conventional scale choices and with NLO calculations:

- Alpgen v2.05 with MLM matching disagrees with NLO shapes.
- The data should be the judge. We have found overall good agreement in all kinds of distributions inside our acceptance before and after tagging: angular correlations, pTs, background cross check samples, discriminant outputs...

Arán García-Bellido
First evidence for single top
Can you assume that Wbb and Wcc fractions separately can be described by the Wbb+Wcc fraction?

- We changed the Wbb/Wcc ratio by ±10% and re-calculated the single top cross section:
- More Wbb, less Wcc: $\sigma(tb+tqb)=4.85\pm1.4\text{pb}$
- Less Wbb, more Wcc: $\sigma(tb+tqb)=4.98\pm1.5\text{pb}$
- Weak dependence based on similarity between Wbb and Wcc shapes
Error on the HF fraction

How come a 30% error on HF fraction doesn't destroy all sensitivity?

- This (still) is a statistics limited analysis: 1.2pb out of 1.4pb error comes from stats alone
- After tagging, the uncertainty on the total W+jets yield is reduced from 30% because:
  - a) Not the entire sample is Wbb+Wcc, the uncertainty on the sum is smaller than 30%
  - b) The anti-correlation between Wjj and Wbb+Wcc due to the normalization before tagging further reduces the uncertainty
- This uncertainty is still one of the largest in the end
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 in NN
Ensemble testing details

- Use a pool of weighted signal+background events (about 850k in each of electron and muon)
- Fluctuate relative and total yields in proportion to systematic errors
  - reproducing the **correlations** between backgrounds imposed by our normalization to data
- Randomly sample from a Poisson distribution about the total yield to simulate **statistical fluctuations**
- Generate a set of pseudo-data (a member of the ensemble)
- Pass the pseudo-data through the full analysis chain (including systematic uncertainties)
Neural Networks

Input Nodes: One for each variable $x_i$

- $M_T$ (jet1,jet2)
- $M$ (alljets)
- $p_T$ (jet1,jet2)
- $p_T$ (notbest2)
- $p_T$ (notbest1)
- $\cos(l,Q(1)x)_{bestop}$
- $M (W,best)$
- $M (W,tag1)$
- $\Delta R (jet1,jet2)$
- $\sqrt{s}$
- $p_T$ (tag1)

Output Node: linear combination of hidden nodes

$$n_k(x, \hat{W}_k) = \frac{1}{1+e^{-w_{ik}x_i}}$$

Hidden Nodes: Each is a sigmoid dependent on the input variables

- Neural Networks
- Input Nodes: One for each variable $x_i$
- Hidden Nodes: Each is a sigmoid dependent on the input variables
- Output Node: linear combination of hidden nodes
- $n_k(x, \hat{W}_k) = \frac{1}{1+e^{-w_{ik}x_i}}$
Training method and optimization

1) Initialize weights
2) Minimize error function on training sample
3) Update weights. This is the first epoch.
4) Repeat procedure. After each epoch, apply NN filter on independent testing sample. Stop training when testing error increases

- Used 60% of events for training, 40% for testing
- Optimize number of training epochs and number of hidden nodes
- MLPFit implementation, many others in the market
### Systematics

<table>
<thead>
<tr>
<th>Relative Systematic Uncertainties</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$t\bar{t}$ cross section</td>
<td>18%</td>
</tr>
<tr>
<td>Luminosity</td>
<td>6%</td>
</tr>
<tr>
<td>Electron trigger</td>
<td>3%</td>
</tr>
<tr>
<td>Muon trigger</td>
<td>6%</td>
</tr>
<tr>
<td>Jet energy scale</td>
<td>wide range</td>
</tr>
<tr>
<td>Jet efficiency</td>
<td>2%</td>
</tr>
<tr>
<td>Jet fragmentation</td>
<td>5–7%</td>
</tr>
<tr>
<td>Heavy flavor fraction</td>
<td>30%</td>
</tr>
<tr>
<td>Tag-rate functions</td>
<td>2–16%</td>
</tr>
</tbody>
</table>
New physics

First evidence for single top