Evidence for single top quark production at DØ

Yann Coadou

CERN
(formerly at Simon Fraser University)

EP Seminar, CERN
30 July 2007

Outline

1. **Tevatron accelerator and DØ detector**
2. **Single top quark production — Should you care?**
3. **Preparing for the measurement**
   - Event selection
   - Signal and background samples
   - *b* tagging
4. **Multivariate analysis techniques**
5. **Expected sensitivity**
6. **Cross sections and significance**
7. **First direct measurement of $|V_{tb}|$**
8. **New combination of analyses**
9. **Conclusion**

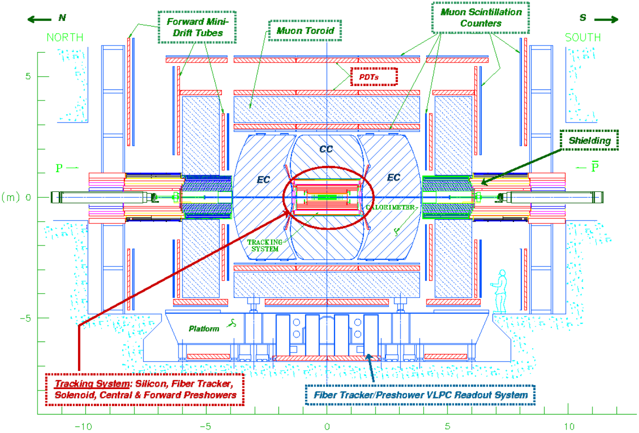
The Tevatron at Fermilab

- Located outside Chicago, Illinois
- The world's highest-energy accelerator
- $p\bar{p}$ collider, centre-of-mass energy 1.96 TeV
- Run I: 1992-1996 at 1.8 TeV
- Started operating for Run II in March 2001
- Upgraded for Run II
  - 396 ns bunch spacing
  - new Main Injector and Recycler
    $\Rightarrow$ increased antiproton intensity

- Peak luminosity $> 2.5 \cdot 10^{32}$ cm^{-2}s^{-1}

The DØ detector upgrade

- 2 T superconducting solenoid
- silicon detector
- fiber tracker
- preshower detector
- upgraded muon system
- new calorimeter electronics
- upgraded trigger and DAQ

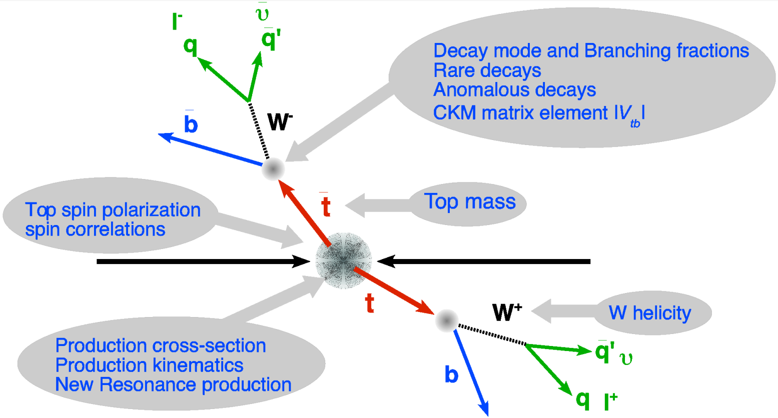


The collaboration

- 600+ physicists, 89 institutes, 18 countries

Top quark physics

- top quark discovered in 1995 by CDF and DØ at the Tevatron
- Heaviest of all fermions
- Couples strongly to Higgs boson
- So far only observed in pairs, only at the Tevatron

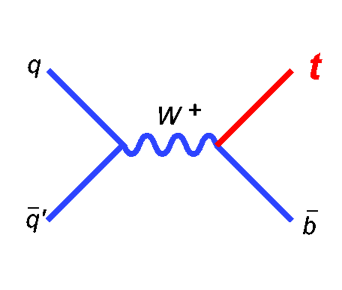


©B.Vachon

Single top quark production

- Never observed before: electroweak production

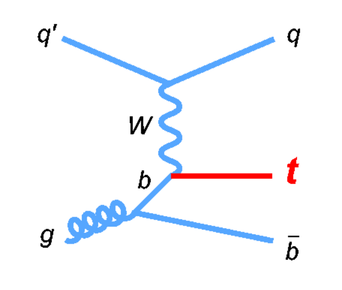
s-channel (tb)



- $\sigma_{NLO} = 0.88 \pm 0.11$ pb (*)
- previous limits (95% C.L.):

Run II DØ: < 5.0 pb (370 pb^{-1})
Run II CDF: < 3.1 pb (700 pb^{-1})

t-channel (tqb)



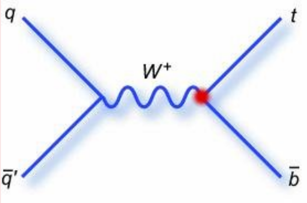
- $\sigma_{NLO} = 1.98 \pm 0.25$ pb(*)
- previous limits (95% C.L.):

Run II DØ: < 4.4 pb (370 pb^{-1})
Run II CDF: < 3.2 pb (700 pb^{-1})

(*) $m_t = 175$ GeV, Phys.Rev. D70 (2004) 114012

Why do we care? — $|V_{tb}|$

- Has never been observed before!
- Should happen in SM
- The value of the cross section is a SM test and the first measurement of $|V_{tb}|$



Direct access to $|V_{tb}|$

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

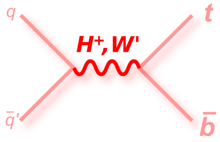
- In SM: top must decay to a W and d, s or b quark
  - $V_{td}^2 + V_{ts}^2 + V_{tb}^2 = 1$
  - constraints on V_{td} and V_{ts}: $|V_{tb}| = 0.9991^{+0.000034}_{-0.000004}$
- New physics:
  - $V_{td}^2 + V_{ts}^2 + V_{tb}^2 < 1$
  - no constraint on V_{tb}
  - e.g. 4th generation: $0.07 < |V_{tb}| < 0.9993$

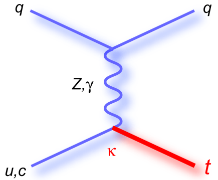
Why do we care? — New physics

- s and t cross sections differently sensitive to new physics

s-channel: charged resonances

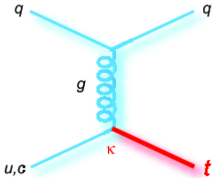
- heavy W' boson in topflavour model (separate interaction for 3rd family)
- charged Higgs boson $H^{\pm}$ in models with extra Higgs doublets (e.g. MSSM)
- charged top pion in topcolor-assisted technicolor
- 4th generation (reduced cross section from $|V_{tb}| < 1$)
- Kaluza-Klein excited W_{KK}, etc...





t-channel: new interactions

- flavour-changing neutral currents ($t\!-\!Z/\gamma/g\!-\!c$ and/or $t\!-\!Z/\gamma/g\!-\!u$ couplings)
- 4th generation (potentially strong enhancement from large V_{ts})

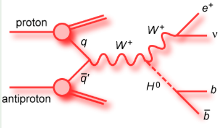


Why do we care? — Spin, Higgs, analysis techniques

Top quark spin

- Large mass ⇒ top quark decays before it can hadronize (no top jets)
- First chance to study a bare quark!
- Top polarization reflected in angular distributions of decay products
- SM predicts high degree of left-handed tops ⇒ possible sign of new physics, or help pin down what new physics

Higgs searches



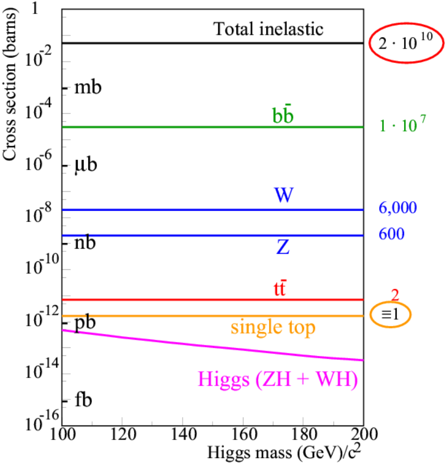
- Important background to *WH* associated Higgs production
- As soon as we discover it, somebody will try to get rid of it....

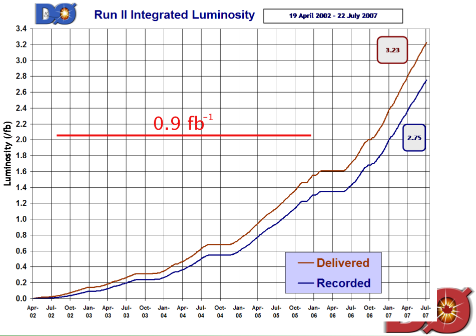
Advanced analysis techniques

- Test of techniques to extract small signal out of large background
- If tools don't work for single top, forget about the Higgs and other small signals
- If tools don't work at Tevatron, not much hope for LHC

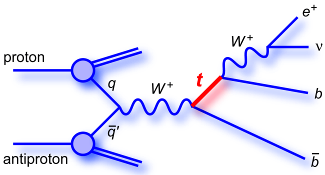
It has been challenging for years...

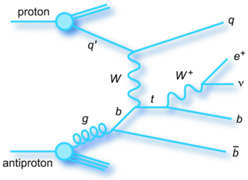
- Several publications since Run I by DØ and CDF
- 7 DØ and 6 CDF PhDs (Dec '06)
- $\sigma_{t\bar{t}}$ only $\sim 2 \times \sigma_{singletop}$, but has striking signature





Event selection





Signature

- isolated lepton
- $\not{E}_T$
- jets
- at least 1 b-jet

Event selection

- Only one tight (no loose) lepton
  - electron: $p_T > 15$ GeV, $|\eta_{det}| < 1.1$
  - muon: $p_T > 18$ GeV, $|\eta_{det}| < 2$
- $15 < \not{E}_T < 200$ GeV
- 2-4 jets: $p_T > 15$ GeV, $|\eta| < 3.4$
  - Leading jet: $p_T > 25$ GeV, $|\eta_{det}| < 2.5$
  - Second leading jet: $p_T > 20$ GeV
- Mis-reconstructed events: require $\not{E}_T$ direction not aligned or anti-aligned in azimuth with lepton or jet
- One or two *b*-tagged jets

Signal and backgrounds

Single top signal ($m_t = 175$ GeV)

- CompHEP-SingleTop + Pythia

W+jets

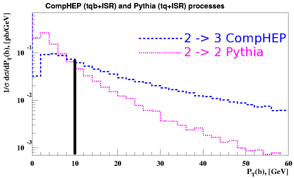
- Most difficult background
- Alpgen+Pythia (MLM matching between matrix elements and parton shower)
- Heavy flavour fraction and normalization from data

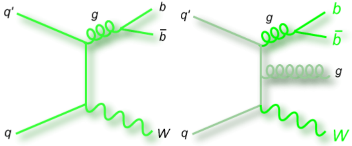
$t\bar{t}$ ($m_t = 175$ GeV)

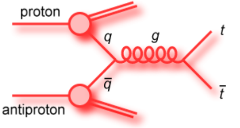
- Alpgen+Pythia (MLM)
- Normalized to $\sigma_{NNLO} = 6.8$ pb

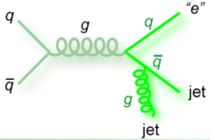
Multijet events

- misidentified lepton, from data



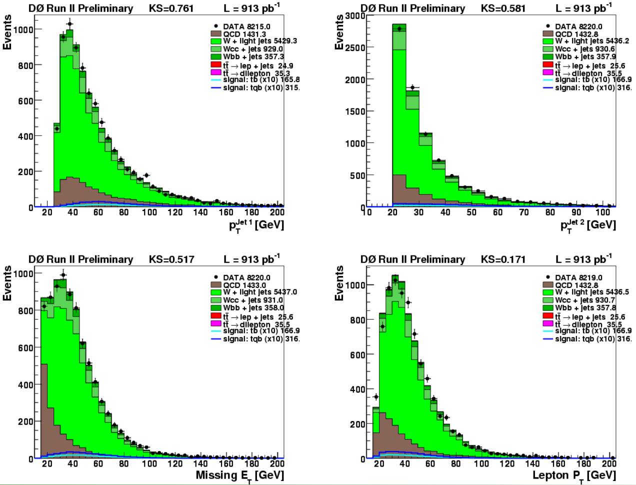






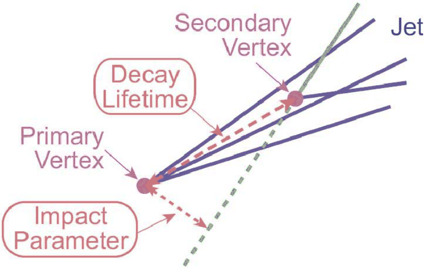
Event selection — Agreement before *b* tagging

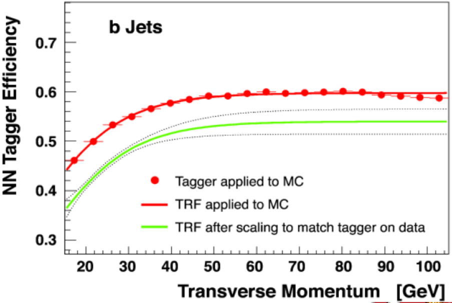
- Normalize W+jets and multijet to data before *b* tagging
- Checked 90 variables, 4 jet multiplicities, electron + muon
- Good description of data



b-jet tagger

- NN trained on 7 input variables from existing taggers.
  - secondary vertices
  - impact parameter
- Much improved performance:
  - fake rate reduced by 1/3 for same *b* efficiency relative to previous tagger
  - smaller systematic uncertainties
- Tag Rate Functions (TRFs) in η, p_T, *z*-PV applied to MC
- Operating point:
  - *b*-jet efficiency $\sim$ 50%
  - *c*-jet efficiency $\sim$ 10%
  - light jet efficiency $\sim$ 0.5%





Event selection — Splitting by S:B

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

Electron + Muon	1 jet	2 jets	3 jets	4 jets	≥ 5 jets
0 tags	10% 1 : 3,200	25% 1 : 390	12% 1 : 300	3% 1 : 270	1% 1 : 230
1 tag	6% 1 : 100	21% 1 : 20	11% 1 : 25	3% 1 : 40	1% 1 : 53
2 tags		3% 1 : 11	2% 1 : 15	1% 1 : 38	0% 1 : 43

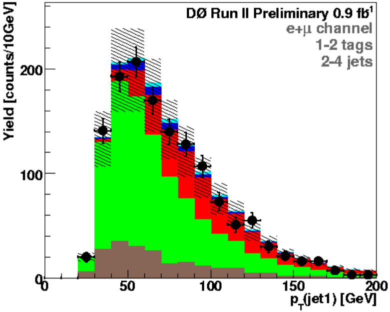
Systematic uncertainties

- Assigned per background, jet multiplicity, lepton flavour and number of tags
- Uncertainties that affect both normalisation and shapes: jet energy scale and tag rate functions (b-tagging parameterisation)
- All uncertainties sampled during limit-setting phase

Relative systematic uncertainties

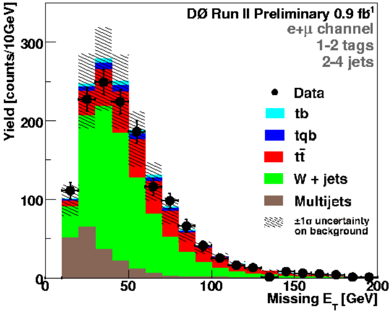
$t\bar{t}$ cross section	18%	Primary vertex	3%
Luminosity	6%	e reco * ID	2%
Electron trigger	3%	e trackmatch & likelihood	5%
Muon trigger	6%	μ reco * ID	7%
Jet energy scale	wide range	μ trackmatch & isolation	2%
Jet efficiency	2%	$\varepsilon_{\mathrm{real}-e}$	2%
Jet fragmentation	5–7%	$\varepsilon_{\mathrm{real}-\mu}$	2%
Heavy flavor ratio	30%	$\varepsilon_{\mathrm{fake}-e}$	3–40%
Tag-rate functions	2–16%	$\varepsilon_{\mathrm{fake}-\mu}$	2–15%

Agreement after tagging



Sample	# of Events
s&t-channel Signal	62
Wjj	174
tt→l+jets	266
Wbb & Wcc	675
Mis-ID's leptons	201
Diboson,tt→ dileptons	82

Totals	2 Jets	3 Jets	4 Jets
Data	697	455	246
Total Background	685	460	253
Signal	36	20	6



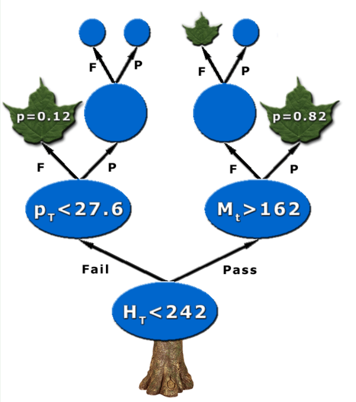
Multivariate analysis techniques

- Boosted decision trees
- Matrix element
- Bayesian neural networks

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
  - sort all events by each variable
  - for each variable, find splitting value with best separation between two children (mostly signal in one, mostly background in the other)
  - select variable and splitting value with best separation, produce two branches with corresponding events ((F)ailed and (P)assed cut)
- Repeat recursively on each node
- Splitting stops: terminal node = leaf



- DT output = leaf purity, close to 1 (0) for signal (bkg)

Ref: Breiman *et al*, "Classification and Regression Trees", Wadsworth (1984)

Splitting a node

Impurity $i(t)$

- maximum for equal mix of signal and background
- symmetric in p_{signal} and $p_{background}$
- minimal for node with either signal only or background only
- strictly concave $\Rightarrow$ reward purer nodes

- Decrease of impurity for split s of node t into children t_L and t_R (goodness of split): $\Delta i(s,t) = i(t) - p_L \cdot i(t_L) - p_R \cdot i(t_R)$
- Aim: find split s^* such that:

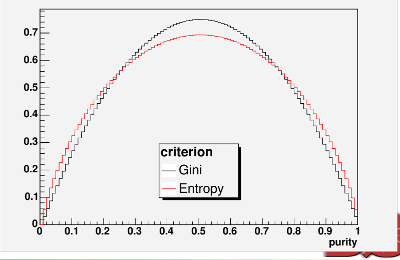
$$\Delta i(s^*, t) = \max_{s \in \{\text{splits}\}} \Delta i(s,t)$$

- Maximizing $\Delta i(s,t) \equiv$ minimizing overall tree impurity

Examples

$$Gini = 1 - \sum_{i=s,b} p_i^2 = \frac{2sb}{(s+b)^2}$$

$$entropy = -\sum_{i=s,b} p_i \log p_i$$



Decision trees — 49 input 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

ΔR(jet1,jet2)
cos(best1,lepton)$_{\text{besttop}}$
cos(best1,notbest1)$_{\text{besttop}}$
cos(tag1,alljets)$_{\text{alljets}}$
cos(tag1,lepton)$_{\text{btaggedtop}}$
cos(jet1,alljets)$_{\text{alljets}}$
cos(jet1,lepton)$_{\text{btaggedtop}}$
cos(jet2,alljets)$_{\text{alljets}}$
cos(jet2,lepton)$_{\text{btaggedtop}}$
cos(lepton,Q(lepton)$\times z$)$_{\text{besttop}}$
cos(lepton$_{\text{besttop}}$,besttop$_{\text{CMframe}}$)
cos(lepton$_{\text{btaggedtop}}$,btaggedtop$_{\text{CMframe}}$)
cos(notbest,alljets)$_{\text{alljets}}$
cos(notbest,lepton)$_{\text{besttop}}$
cos(untag1,alljets)$_{\text{alljets}}$
cos(untag1,lepton)$_{\text{btaggedtop}}$

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)
Missing E_T
p_T(alljets$-$best1)
p_T(alljets$-$tag1)
p_T(jet1,jet2)
Q(lepton)$\times\eta$(untag1)
$\sqrt{\hat{s}}$
Sphericity(alljets,W)

- Adding variables does not degrade performance
- Tested shorter lists, lost some sensitivity
- Same list used for all channels

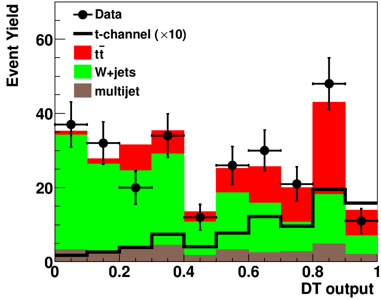
Decision tree output

Measure and apply

- Take trained tree and run on independent pseudo-data sample, determine purities
- Apply to data
- Should see enhanced separation (signal right, background left)
- Could cut on output and measure, or use whole distribution to measure

Limitations

- Instability of tree structure
- Piecewise nature of output



Advantages

- DT has human readable structure (no black box)
- Training is fast
- Deals with discrete variables
- No need to transform inputs
- Resistant to irrelevant variables

Boosting a decision tree

Boosting

- Recent technique to improve performance of a weak classifier
- Recently used on decision trees by GLAST and MiniBooNE
- Basic principle on DT:
  - train a tree T_k
  - $T_{k+1} = \text{modify}(T_k)$

AdaBoost algorithm

- Adaptive boosting
- Check which events are misclassified by T_k
- Derive tree weight α_k
- Increase weight of misclassified events by e^{α_k}
- Train again to build T_{k+1}
- Boosted result of event i: $T(i) = \sum_{k=1}^{N_\text{tree}} \alpha_k T_k(i)$

- Averaging ⇒ dilutes piecewise nature of DT
- Usually improves performance

Ref: Freund and Schapire, "Experiments with a new boosting algorithm", in *Machine Learning: Proceedings of the Thirteenth International Conference*, pp 148-156 (1996)

Decision tree parameters

DT choices

- 1/3 of MC for training
- AdaBoost parameter $\beta = 0.2$
- 20 boosting cycles
- Signal leaf if purity > 0.5
- Minimum leaf size $= 100$ events
- Same total weight to signal and background to start
- Goodness of split - Gini factor

Analysis strategy

- Train 36 separate trees:
  - 3 signals (s,t,$s+t$)
  - 2 leptons (e,μ)
  - 3 jet multiplicities (2,3,4 jets)
  - 2 b-tag multiplicities (1,2 tags)
- For each signal train against the sum of backgrounds

Matrix element method

- Pioneered by DØ top mass analysis. Now used in search
- Use the 4-vectors of all reconstructed leptons and jets
- Use matrix elements of main signal and bkgd diagrams to compute event probability density for signal and bkgd hypotheses
- Goal: calculate a discriminant:

$$D_s(\vec{x}) = P(S|\vec{x}) = \frac{P_{signal}(\vec{x})}{P_{signal}(\vec{x}) + P_{bkg}(\vec{x})}$$

- Encoded in normalized differential cross section for process S:

$$P_S(\vec{x}) = \frac{1}{\sigma_S} d\sigma_S(\vec{x}), \quad \sigma_S = \int d\sigma_S(\vec{x})$$

Used only limited number of Feynman diagrams

- Sensitivity would increase (but so does computation time) if more diagrams were included. In particular, no $t\bar{t}$ diagrams are computed (serious limitation for >2 jets)

Bayesian neural networks

A different sort of neural network

- Instead of choosing one set of weights, find posterior probability density over all possible weights
- Averaging over many networks weighted by the probability of each network given the training data
- Used 25 variables (subset of DT variables)
- Same strategy as DT: 36 different BNN

Advantages

- Less prone to overtraining
- Details of each network not important

Limitation

- Darker black box
- Computationally demanding

Implementation: Flexible Bayesian Modeling (FBM) package

http://www.cs.toronto.edu/∼radford/fbm.software.html

Analysis validation

Ensemble testing

- Test the whole machinery with many sets of pseudo-data
- Like running DØ experiment 1000s of times
- Generated ensembles with different signal contents (no signal, SM, other cross sections, higher luminosity)

Ensemble generation

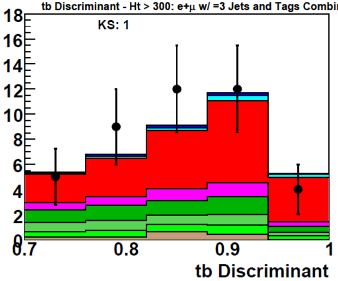
- Pool of weighted signal + background events
- Fluctuate relative and total yields in proportion to systematic errors, reproducing correlations
- Randomly sample from a Poisson distribution about the total yield to simulate statistical fluctuations
- Generate pseudo-data set, pass through full analysis chain (including systematic uncertainties)

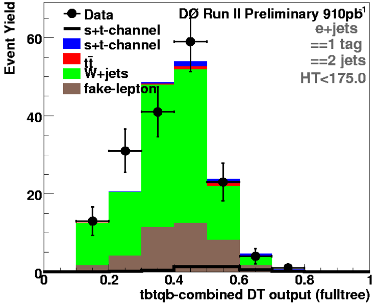
All analyses achieved linear response to varying input cross sections and negligible bias

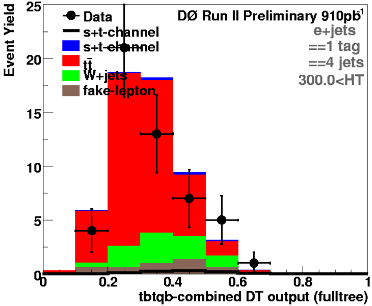
Cross-check samples

- Validate methods on data in no-signal region
- **"W+jets":** $=$2jets, H_T(lepton,$\not\!\!E_T$,alljets) $<$ 175 GeV
- **"ttbar":** $=$4jets, H_T(lepton,$\not\!\!E_T$,alljets) $>$ 300 GeV
- Good agreement

ME "hard W+jets": $=$3jets, $H_T > 300$ GeV







Sensitivity determination

- Use the 0-signal ensemble

Expected p-value

Fraction of 0-signal pseudo-datasets in which we measure at least 2.9 pb (SM single top cross section)

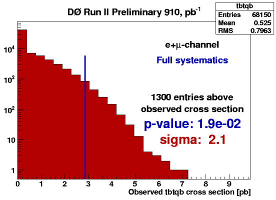
Observed p-value

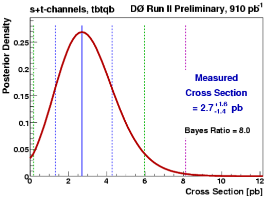
Fraction of 0-signal pseudo-datasets in which we measure at least the observed cross section.

- Also use the SM ensemble to check compatibility of observed result with SM prediction

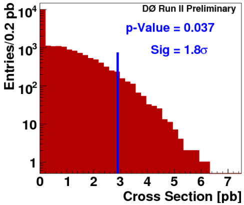
Expected sensitivity s+t

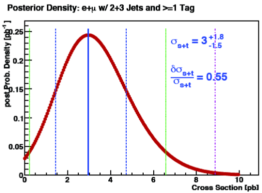
Decision trees
p-value **1.9%** (2.1σ)



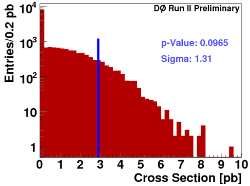


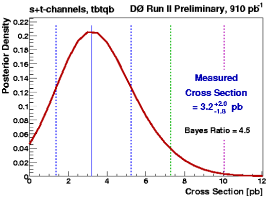
Matrix elements
p-value **3.7%** (1.8σ)





Bayesian NN
p-value **9.7%** (1.3σ)





ME and BNN s+t observed results

Matrix element

$\sigma = 4.6^{+1.8}_{-1.5}$ pb
p-value = 0.21% (2.9σ)
SM compatibility 21%

Bayesian NN

$\sigma = 5.0 \pm 1.9$ pb
p-value = 0.89% (2.4σ)
SM compatibility 18%

New preliminary ME result

- Included $t\bar{t} \rightarrow \ell$+jets ME in 3-jet discriminant

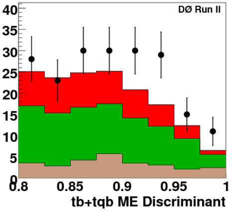
$\sigma = 4.8^{+1.6}_{-1.4}$ pb
exp. p-value = 3.1% (1.9σ)
obs. p-value = 0.082% (3.2σ)

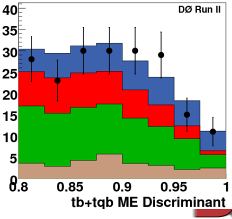
New preliminary BNN result

- Better treatment of noisy training data

$\sigma = 4.4^{+1.6}_{-1.4}$ pb
exp. p-value = 1.6% (2.2σ)
obs. p-value = 0.083% (3.1σ)

- ME discriminant output, with and without signal content (all channels combined)

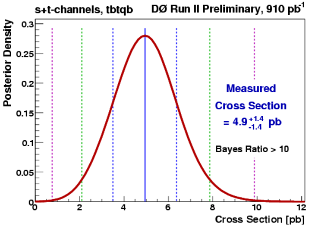


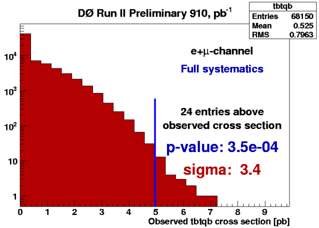


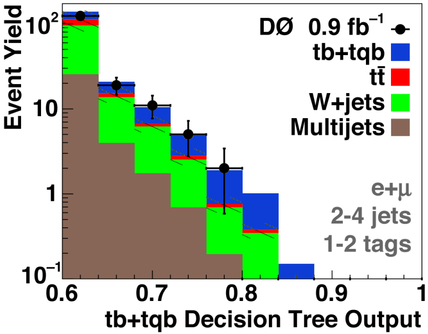
Boosted decision tree observed results

$\sigma_{s+t} = 4.9 \pm 1.4$ pb
p-value $= 0.035\%$ (3.4σ)
SM compatibility: 11% (1.3σ)

Evidence for single top production!



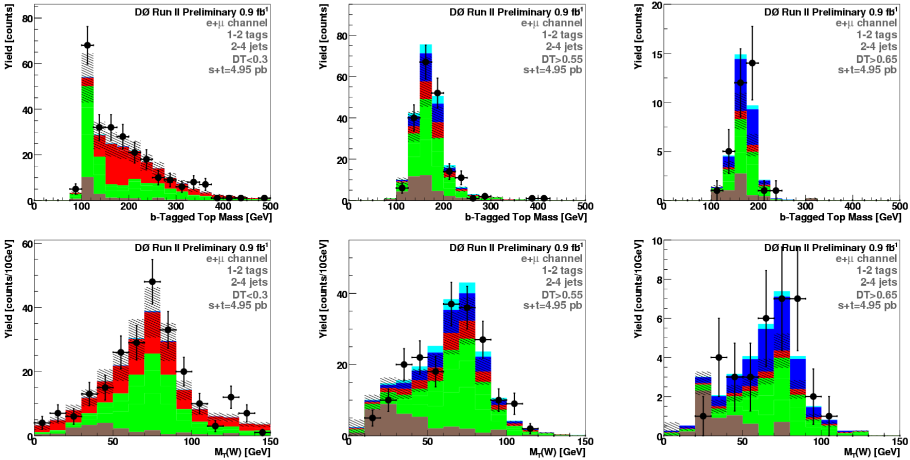




$$\sigma_s = 1.0 \pm 0.9 \text{ pb}$$
$$\sigma_t = 4.2^{+1.8}_{-1.4} \text{ pb}$$

Boosted decision tree event characteristics

$DT < 0.3$ | $DT > 0.55$ | $DT > 0.65$



Measuring $|V_{tb}|$

- Now that we have a cross section measurement, we can make the first direct measurement of $|V_{tb}|$
- Use the same infrastructure as for cross section measurement but make a posterior in $|V_{tb}|^2$

Additional theoretical errors (hep-ph/0408049)

	s	t
top mass	13%	8.5%
scale	5.4%	4.0%
PDF	4.3%	10.0%
α_s	1.4%	0.01%

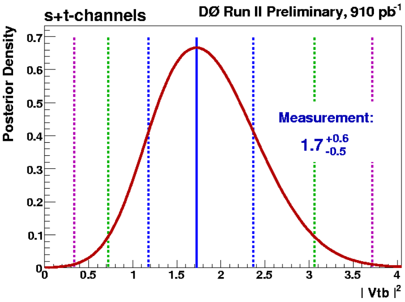
- Most general *Wtb* coupling ($P_{L,R} = (1 \mp \gamma_5)/2$):

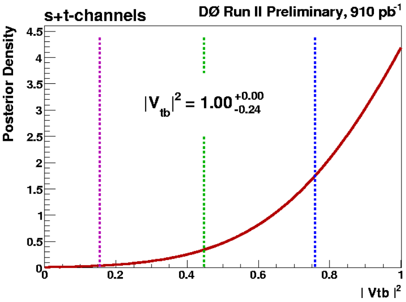
$$\Gamma^{\mu}_{tbW} = -\frac{g}{\sqrt{2}} V_{tb} \bar{u}(p_b) \left[\gamma^{\mu} (f_1^L P_L + f_1^R P_R) - \frac{i\sigma^{\mu\nu}}{M_W} (f_2^L P_L + f_2^R P_R) \right] u(p_t)$$

- SM: $f_1^L = 1$, $f_1^R = 0$ (pure $V-A$), $f_2^L = f_2^R = 0$ (CP conservation)
- Effectively measuring strength of $V-A$ coupling $|V_{tb} f_1^L|$, can be > 1

First direct measurement of $|V_{tb}|$

- Assuming $V_{td}^2 + V_{ts}^2 \ll V_{tb}^2$ and pure $V-A$ and CP-conserving Wtb interaction





$$|\mathbf{V_{tb}f_1^L}| = \mathbf{1.3 \pm 0.2}$$

$$\mathbf{0.68 < |V_{tb}| \leq 1 \text{ @ } 95\% \text{ CL}}$$

(assuming $f_1^L = 1$, flat prior in [0,1])

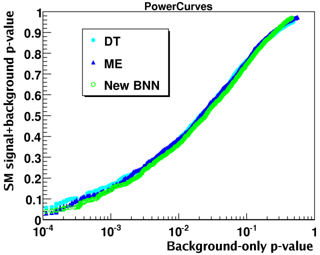
- No assumption about number of quark families or CKM matrix unitarity

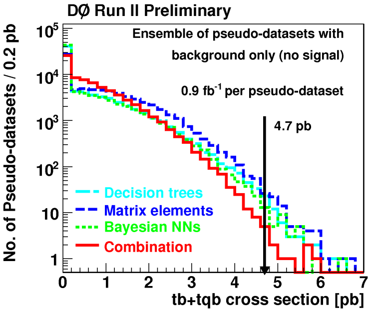
New: combination of s+t analyses

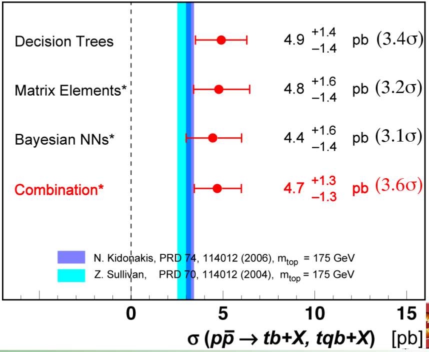
Correlations

- 3 analyses with similar performance on same dataset
- Combined using BLUE method

	DT	ME	BNN
DT	100%	64%	66%
ME		100%	59%
BNN			100%







Conclusion

First evidence for single top quark production (DØ decision trees)

$$\sigma(p\bar{p} \rightarrow tb + X, tqb + X) = 4.9 \pm 1.4 \text{ pb}$$

3.4σ significance

First direct measurement of $|V_{tb}|$ (DØ decision trees)

$$|V_{tb} f_1^L| = 1.3 \pm 0.2$$

assuming $f_1^L = 1$: $0.68 < |V_{tb}| \leq 1$ @ 95% CL

(Always assuming $V_{td}^2 + V_{ts}^2 \ll V_{tb}^2$ and pure $V{-}A$ and CP-conserving Wtb interaction)

Published in Phys. Rev. Lett. 98, 181802 (2007) (hep-ex/0612052)

New preliminary combination of DT, ME and BNN

$$\sigma(p\bar{p} \rightarrow tb + X, tqb + X) = 4.7 \pm 1.3 \text{ pb}$$

3.6σ significance

- A lot more data already at hand

Single top prospects — Tevatron and LHC

Tevatron

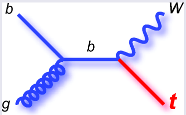
- By 2009 we should have observed single top production and measured its cross section to 15-20%
- $|V_{tb}|$ is then known to $\sim$10%

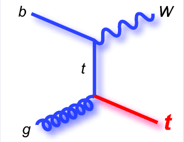
LHC

- Much larger production rates:

$$\begin{aligned}\sigma_s^{t/\bar{t}} &= 6.6/4.1 \text{ pb } (\pm 10\%)\\ \sigma_t^{t/\bar{t}} &= 156/91 \text{ pb } (\pm 5\%)\\ \sigma_{tW}^{t/\bar{t}} &= 34/34 \text{ pb } (\pm 10\%)\end{aligned}$$

- Try to observe all three channels (s-channel challenging)
- $|V_{tb}|$ measured to percent level
- Large samples ⇒ study properties





Backup slides

- More information:
  http://www-d0.fnal.gov/Run2Physics/top/public/fall06/singletop