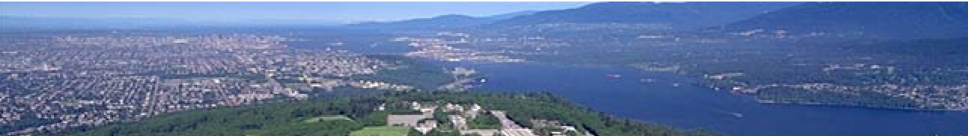
Evidence for single top quark production at DØ

Yann Coadou

Simon Fraser University

McGill University, Montréal
26 January 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. **Conclusion**

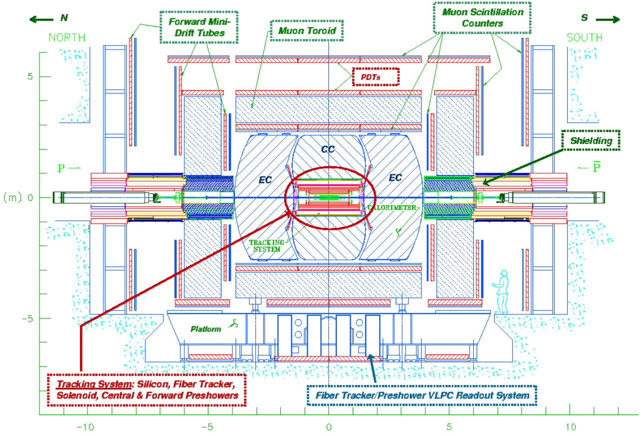
The Tevatron at Fermilab

- 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

- 670+ physicists, 90 institutes, 19 countries

The collaboration

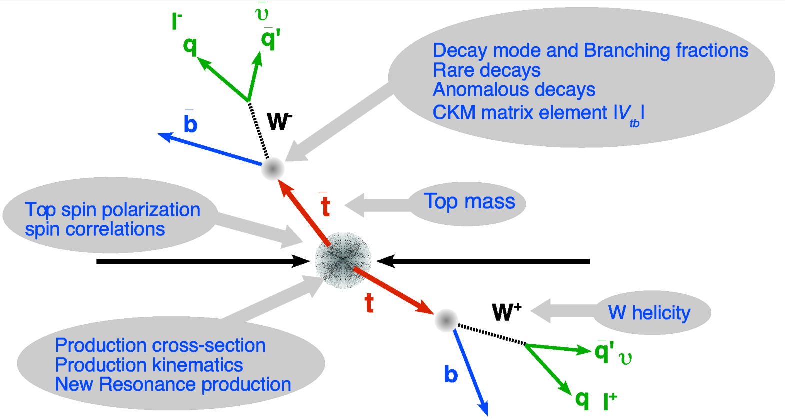
- 670+ physicists, 90 institutes, 19 countries

DØ McGill

- Brigitte Vachon
- Chris Potter
- Gustavo Kertzscher
- Camille Belanger-Champagne
- Philipe Vachon-Rivard, Bertrand Chapleau

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

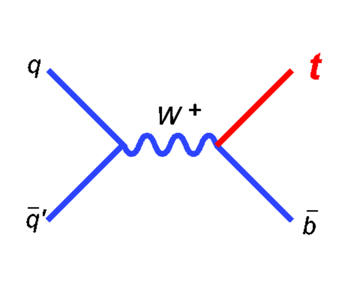


©B.Vachon

Single top quark production

- Never observed before: electroweak production

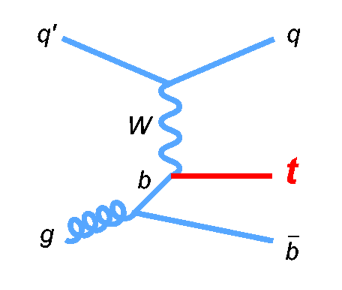
s-channel (tb)



- $\sigma_{NLO} = 0.88 \pm 0.11$ 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)



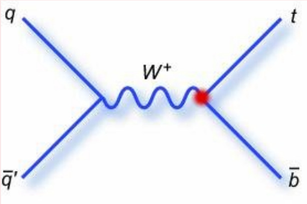
- $\sigma_{NLO} = 1.98 \pm 0.25$ pb(*)
- previous limits (95% C.L.):

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

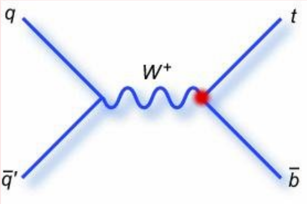
Why do we care? — $|V_{tb}|$

- Has never been observed before!
- It should happen (if SM is right)
- The value of the cross section is a SM test and the first measurement of $|V_{tb}|$



Why do we care? — $|V_{tb}|$

- Has never been observed before!
- It should happen (if SM is right)
- The value of the cross section is a SM test and the first measurement of $|V_{tb}|$



Direct access to $|V_{tb}|$

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

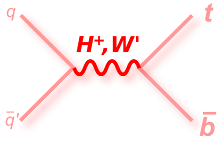
- Weak interaction eigenstates are not mass eigenstates
- In SM: top must decay to a W and d, s or b quark
  - $V_{td}^2 + V_{ts}^2 + V_{tb}^2 = 1$
  - constraints on V_{td} and V_{ts}: $|V_{tb}| = 0.9991^{+0.000034}_{-0.000004}$
- New physics:
  - $V_{td}^2 + V_{ts}^2 + V_{tb}^2 < 1$
  - no constraint 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

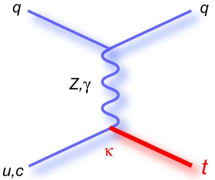
s-channel: charged resonances

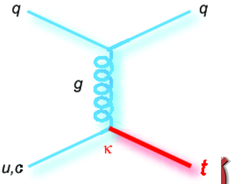
- heavy W' boson in topflavour model (separate interaction for 3rd family)
- charged Higgs boson $H^{\pm}$ in models with extra Higgs doublets (e.g. MSSM)
- charged top pion in topcolor-assisted technicolor
- 4th generation (reduced cross section from $|V_{tb}| < 1$)
- Kaluza-Klein excited W_{KK}, etc...



t-channel: new interactions

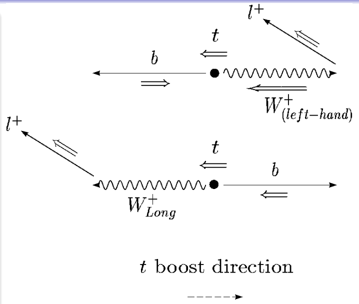
- flavour-changing neutral currents ($t - Z/\gamma/g - c$ and/or $t - Z/\gamma/g - u$ couplings)
- 4th generation (potentially strong enhancement from large V_{ts})





Why do we care? — 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



Helicity basis

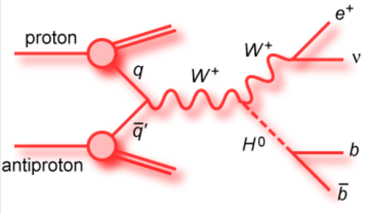
- $\hat{z}$-axis along top quark direction of motion
- ~80(97)% polarization for s(t) channel at parton level
- ~68(88)% polarization for s(t) channel after reconstruction

Optimal basis

- $\hat{z}$-axis along direction of d-type quark: mostly antiproton beamline for s-channel, spectator jet for t-channel
- ~96(98)% polarization for s(t) channel at parton level
- ~71(89)% polarization for s(t) channel after reconstruction

Why do we care? — Backgrounds, analysis techniques

Higgs searches



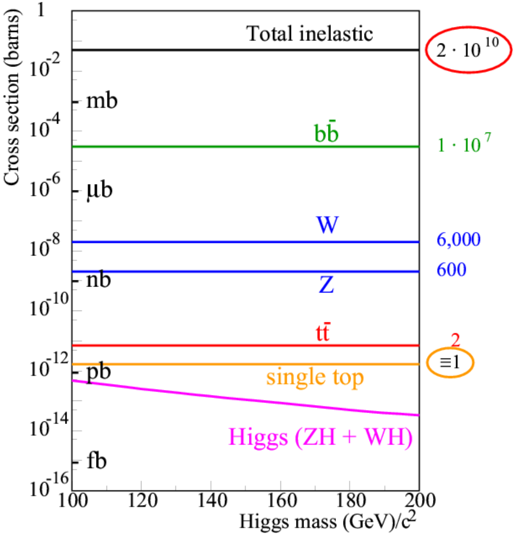
- Important background to *WH* associated Higgs production
- As soon as we discover it, somebody will try to get rid of it....

Advanced analysis techniques

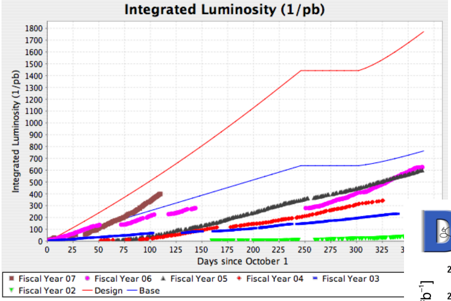
- Test of techniques to extract small signal out of large background
- If tools don't work for single top, forget about the Higgs and other small signals
- If tools don't work at Tevatron, not much hope for LHC

It has been challenging for years...

- Several publications since Run I by DØ and CDF
- 7 DØ and 6 CDF PhDs
- $\sigma_{t\bar{t}}$ only $\sim 2 \times \sigma_{singletop}$, but has striking signature

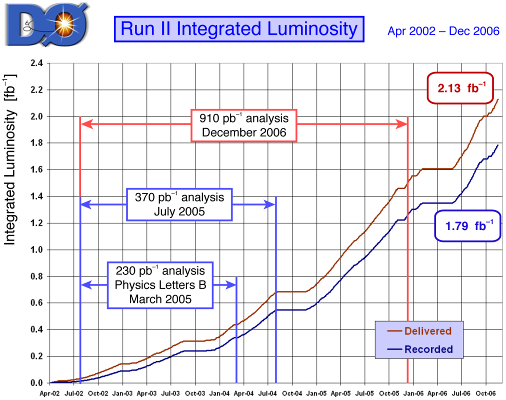


Tevatron and DØ luminosities



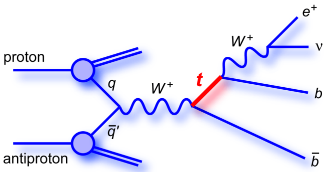
- Interesting physics only a tiny fraction of collisions
  $\Rightarrow$ increase the number of collisions

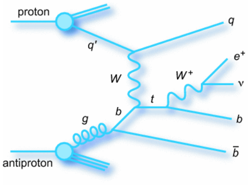
- Trigger required to select events at 50 Hz from 2.5 MHz collisions
- Large amount of data, reprocessed on computing grid
- Lots of Monte Carlo events produced on the grid



Many thanks to the Accelerator Division

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

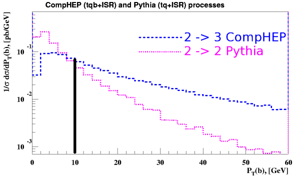
- Most difficult background
- Alpgen+Pythia (MLM matching between matrix elements and parton shower)
- Heavy flavour fraction and normalization from data

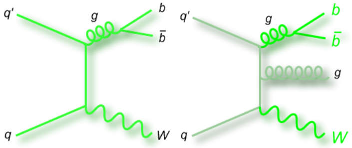
$t\bar{t}$ ($m_t = 175$ GeV)

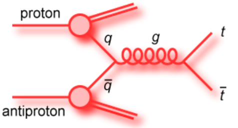
- Alpgen+Pythia (MLM)
- Normalized to $\sigma_{NNLO} = 6.8$ pb

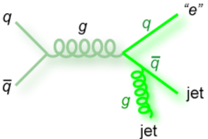
Multijet events

- misidentified lepton, from data



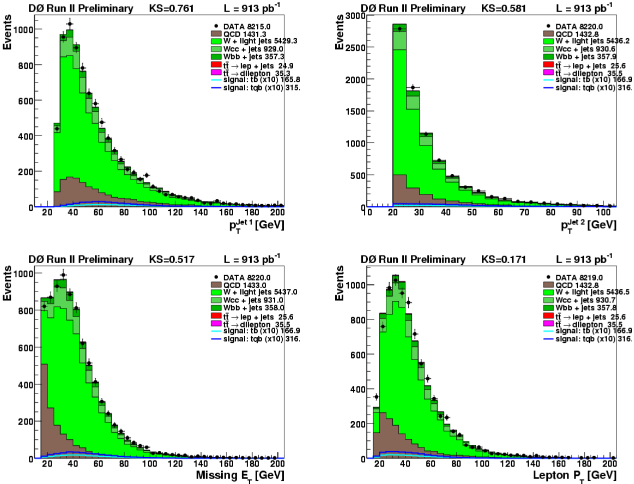






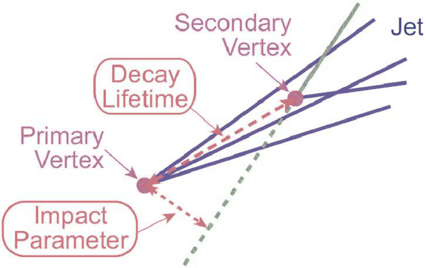
Event selection — Agreement before tagging

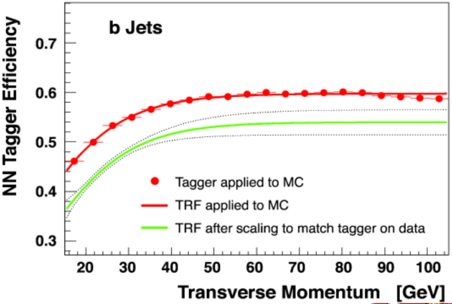
- Normalize W+jets and multijet to data before tagging
- Checked 90 variables, 4 jet multiplicities, electron + muon
- Good description of data



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

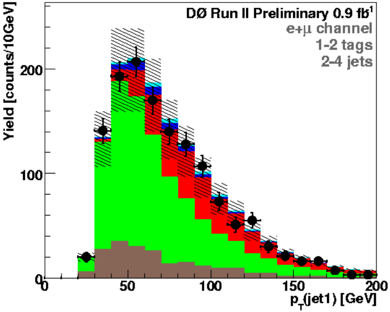
Systematic uncertainties

- 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

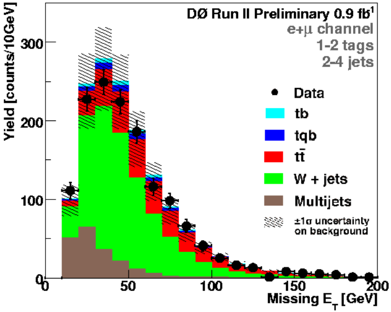
$t\bar{t}$ cross section	18%	Primary vertex	3%
Luminosity	6%	e reco * ID	2%
Electron trigger	3%	e trackmatch & likelihood	5%
Muon trigger	6%	μ reco * ID	7%
Jet energy scale	wide range	μ trackmatch & isolation	2%
Jet efficiency	2%	$\varepsilon_{\mathrm{real}-e}$	2%
Jet fragmentation	5–7%	$\varepsilon_{\mathrm{real}-\mu}$	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



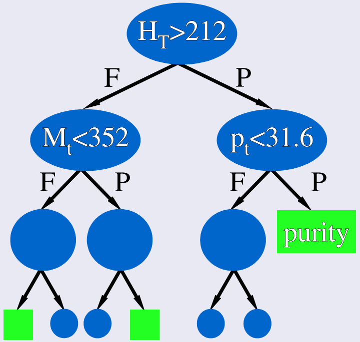
Multivariate analysis techniques

- 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:

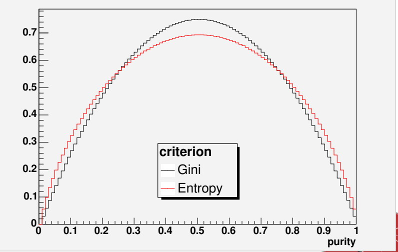
$$\Delta i(s^*, t) = \max_{s \in \{\text{splits}\}} \Delta i(s,t)$$

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

Examples

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

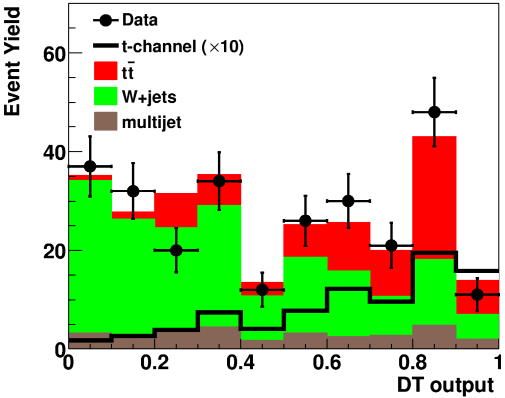
$$entropy = - \sum_{i=s,b} p_i \log p_i$$



Decision 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



Boosting a decision tree

Boosting

- Recent technique to improve performance of a weak classifier
- Recently used on decision trees by GLAST and MiniBooNE
- Basic principal 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: $T(i) = \sum_{n=1}^{N_{\text{tree}}} \alpha_k T_k(i)$

- Averaging $\Rightarrow$ 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

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

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

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 properly 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)

Analysis validation — Ensemble testing

- To verify that all of this machinery is working properly we test with many sets of pseudo-data
- Wonderful tool to test analysis methods! Run DØ experiment 1000s of times!
- Generated ensembles:
  - 0-signal ensemble ($s + t\ \sigma = 0$ pb)
  - SM ensemble ($s + t\ \sigma = 2.9$ pb)
  - "Mystery" ensembles to test analyzers ($s + t\ \sigma =$?? pb)
  - Ensembles at measured cross section ($s + t\ \sigma =$ measured)
  - A high luminosity ensemble

Ensemble Testing - Details

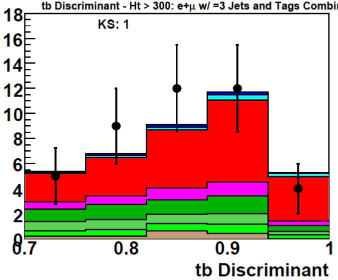
- Use a pool of weighted signal + background events (about 850k in each of electron and muon)
- Fluctuate relative and total yields in proportion to systematic errors, reproducing correlations
- Randomly sample from a Poisson distribution about the total yield to simulate statistical fluctuations
- Generate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

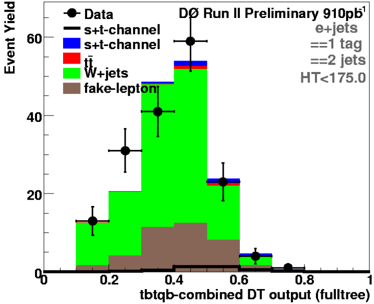
All analyses achieved linear response to varying input cross sections

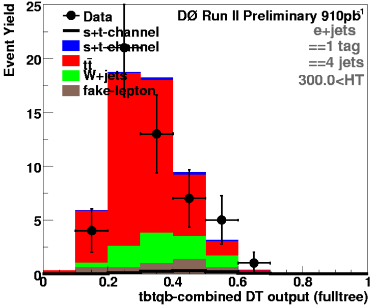
Cross-check samples

- 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





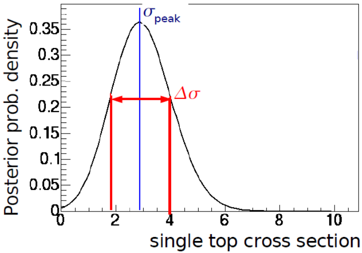


Measuring cross sections

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

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

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

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

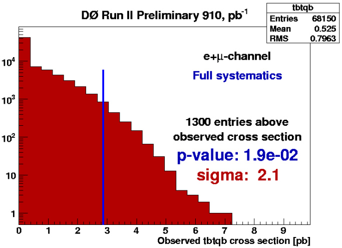
Fraction of 0-signal pseudo-datasets in which we measure at least the observed cross section.

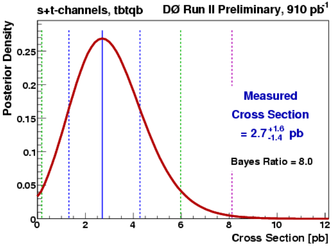
- Also use the SM ensemble to check compatibility of observed result with SM prediction

Expected sensitivity s+t

Decision Trees

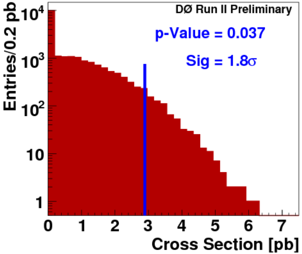
p-value **1.9%** (2.1σ)

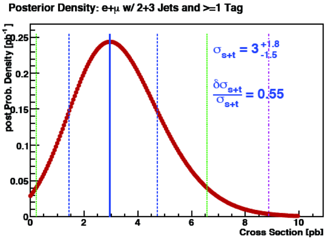




Matrix Elements

p-value **3.7%** (1.8σ)





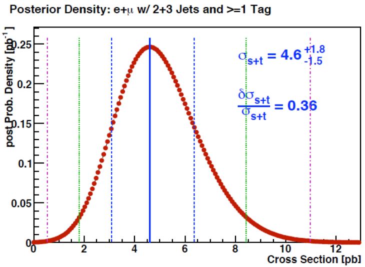
Matrix element s+t observed results

Matrix element

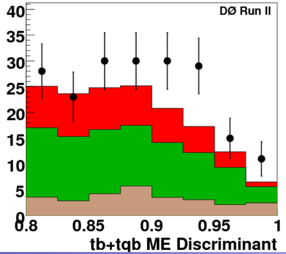
$\sigma = 4.6^{+1.8}_{-1.5}$ pb

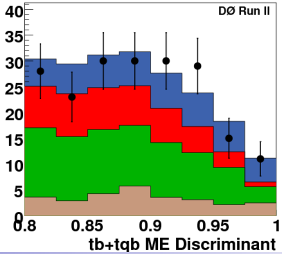
p-value = 0.21% (2.9σ)

SM compatibility 21%



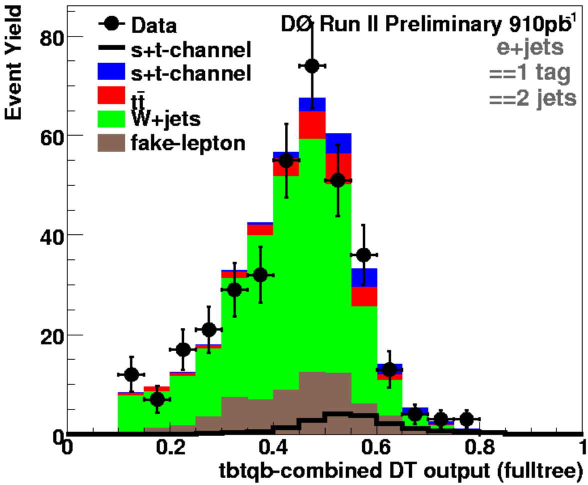
- ME discriminant output, with and without signal content (all channels combined)





Decision trees on data

- We have 36 different Decision Trees
- Example: electron, 2 jet, 1 tag

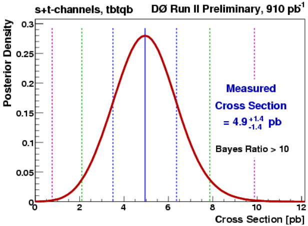


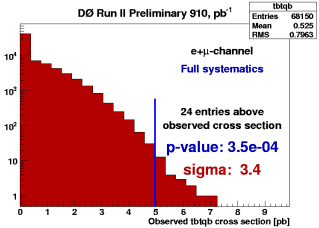
Boosted decision tree observed results

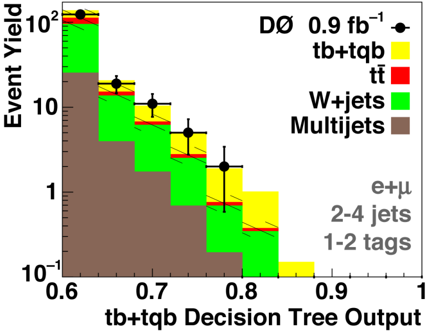
$$\sigma_{s+t} = 4.9 \pm 1.4 \text{ pb}$$

p-value $= 0.035\%$ (3.4σ)

SM compatibility: 11% (1.3σ)

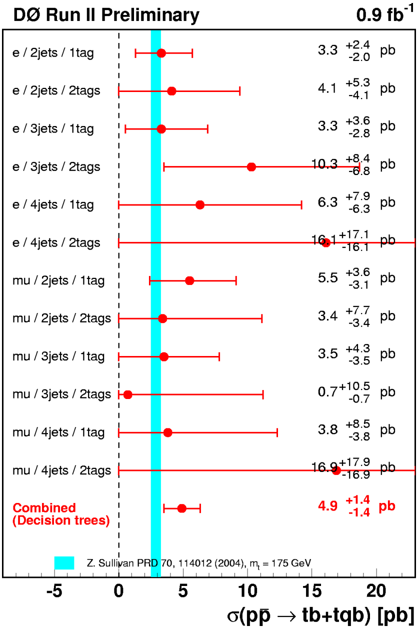






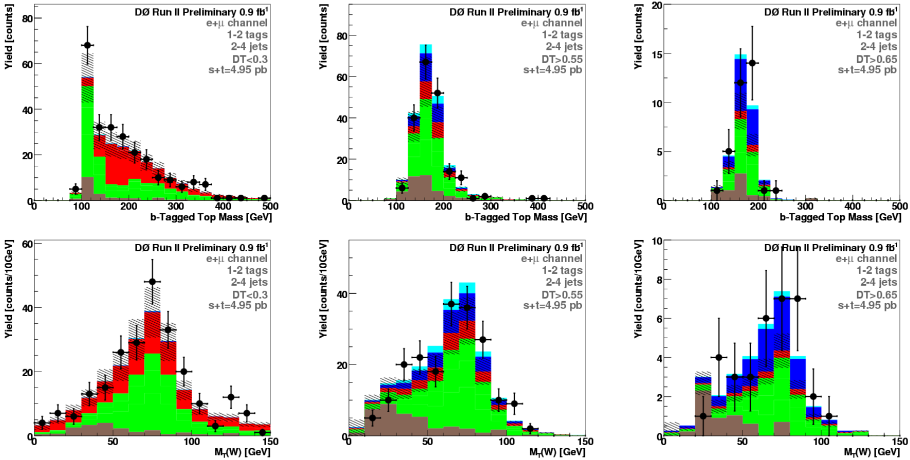
$$\sigma_s = 1.0 \pm 0.9 \text{ pb}$$
$$\sigma_t = 4.2^{+1.8}_{-1.4} \text{ pb}$$

Decision trees — Summary

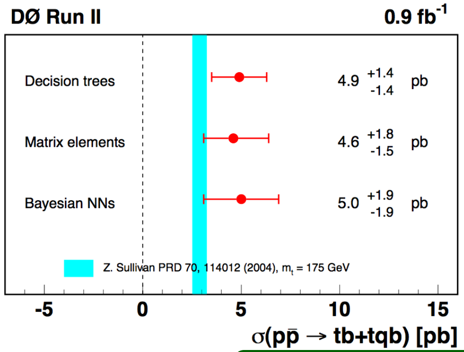


Boosted decision tree event characteristics

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



s+t summary — Correlations



High discriminant correlation

Choose the 50 highest events in each discriminant and look for overlap

	Electron	Muon
DT vs ME	52%	58%
DT vs BNN	56%	48%
ME vs BNN	46%	52%

Linear correlation

Measured cross section in 400 members of SM ensemble with all three techniques and calculated the linear correlation between each pair

	DT	ME	BNN
DT	100%	39%	57%
ME		100%	29%
BNN			100%

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%

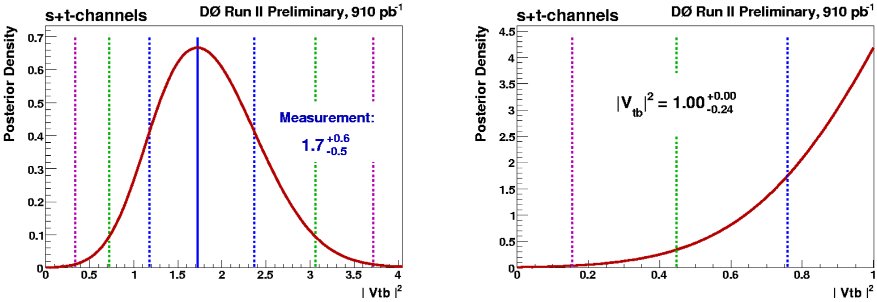
- Most general *Wtb* coupling ($P_{L,R} = (1 \mp \gamma_5)/2$):

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

- SM: $f_1^L = 1$, $f_1^R = f_2^L = f_2^R = 0$
- 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



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

$\mathbf{0.68 < |V_{tb}| \leq 1}$ **@ 95% CL**
(assuming $f_1^L = 1$, flat prior in [0,1])

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

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

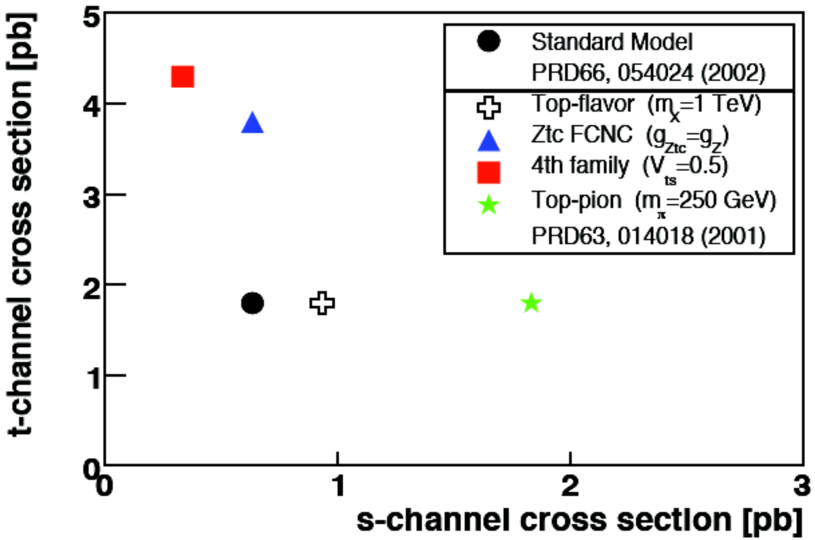
(Always assuming $V_{td}^2 + V_{ts}^2 \ll V_{tb}^2$ and pure $V-A$ and CP-conserving Wtb interaction)

hep-ex/0612052, submitted to PRL

- Working on understanding correlations and on combinations
- A lot more data already at hand

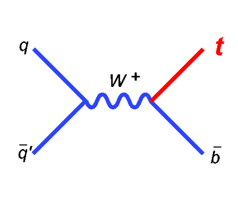
Backup slides

Motivation - New Physics



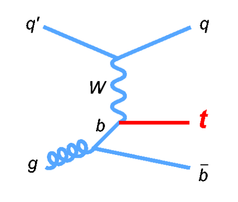
Top quark spin

s-channel: $u\bar{d} \to t\bar{b}$



p	$\bar{p}$	fraction
u	$\bar{d}$	98%
$\bar{d}$	u	2%

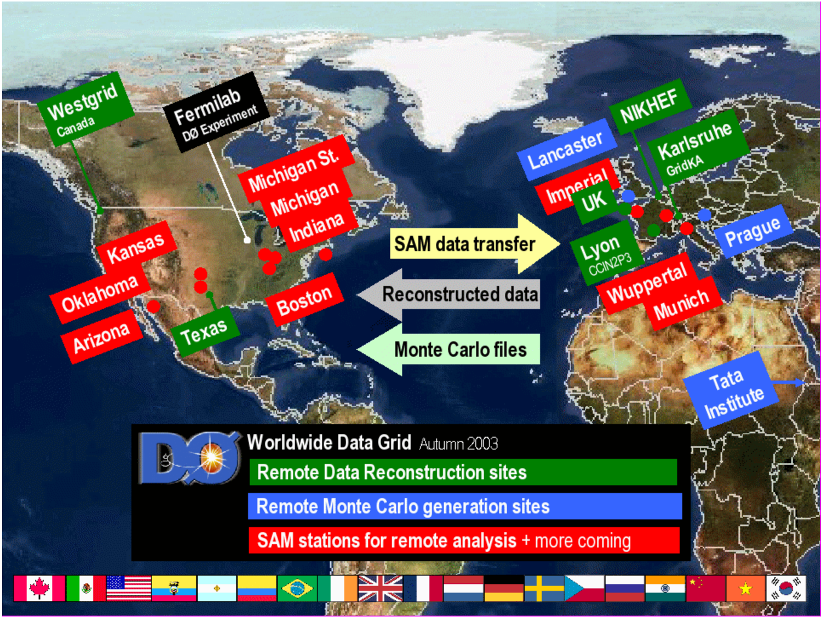
t-channel: $ug \to t\bar{b}d$, $\bar{d}g \to t\bar{b}\bar{u}$



p	$\bar{p}$	fraction
u	g	74%
g	$\bar{d}$	20%
g	u	3%
$\bar{d}$	g	3%

G. Mahlon, S. Parke, PRD55(1997)7249-7254

Data reprocessing - Monte Carlo production



W+jets normalization

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

$$\begin{aligned} N^{loose} &= N^{loose}_{fake} + N^{loose}_{real} \\ N^{tight} &= \epsilon_{fake} N^{loose}_{fake} + \epsilon_{real} N^{loose}_{real} \end{aligned}$$

Solve for N^{loose}_{fake} and N^{loose}_{real}

- Normalize MC W+jets samples to real lepton yield found in data, accounting for presence of $t\bar{t}$:

$$\epsilon_{real} N^{loose}_{real} = SF \times (Wjj + Wcc + Wbb) + t\bar{t}, \quad SF \sim 1.4$$

- $(Wbb + Wcc)/Wjj$ found in zero-tag sample: 1.5±0.5
- Then apply b-tagging
  - reduces W+jets background
  - changes flavour composition and kinematic distributions

W+jets heavy flavour fraction

$$\alpha(Wb\bar{b} + Wc\bar{c}) + Wjj + t\bar{t} + \text{QCD} = \text{Data}$$

Scale Factor α to Match Heavy Flavor Fraction to Data

	1 jet	2 jets	3 jets	4 jets
Electron Channel				
0 tags	1.53 ± 0.10	1.48 ± 0.10	1.50 ± 0.20	1.72 ± 0.40
1 tag	1.29 ± 0.10	1.58 ± 0.10	1.40 ± 0.20	0.69 ± 0.60
2 tags	—	1.71 ± 0.40	2.92 ± 1.20	-2.91 ± 3.50
Muon Channel				
0 tags	1.54 ± 0.10	1.50 ± 0.10	1.52 ± 0.10	1.38 ± 0.20
1 tag	1.11 ± 0.10	1.52 ± 0.10	1.32 ± 0.20	1.86 ± 0.50
2 tags	—	1.40 ± 0.40	2.46 ± 0.90	3.78 ± 2.80

Event Selection - Yields

Source	Event Yields in 0.9 fb^{-1} Data Electron+muon, 1tag+2tags combined		
	2 jets	3 jets	4 jets
tb	16 ± 3	8 ± 2	2 ± 1
tqb	20 ± 4	12 ± 3	4 ± 1
$t\bar{t} \rightarrow ll$	39 ± 9	32 ± 7	11 ± 3
$t\bar{t} \rightarrow l$+jets	20 ± 5	103 ± 25	143 ± 33
$W+b\bar{b}$	261 ± 55	120 ± 24	35 ± 7
$W+c\bar{c}$	151 ± 31	85 ± 17	23 ± 5
$W+jj$	119 ± 25	43 ± 9	12 ± 2
Multijets	95 ± 19	77 ± 15	29 ± 6
Total background	686 ± 41	460 ± 39	253 ± 38
Data	697	455	246

- Expected single top signal is smaller than background uncertainty!
  $\Rightarrow$ No counting experiment, requires advanced analysis techniques

Matrix element discriminants

DØ discriminants

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

$$P_{bkg}^{2jets}(\vec{x}) = c_{Wbb} P_{Wbb}(\vec{x}) + c_{Wcg} P_{Wcg}(\vec{x}) + c_{Wgg} P_{Wgg}(\vec{x})$$

$$P_{bkg}^{3jets}(\vec{x}) = P_{Wbbg}(\vec{x})$$

- c_{Wbb}, c_{Wcg} and c_{Wgg} are in principle the relative fractions of each background
- optimized for each channel to increase sensitivity

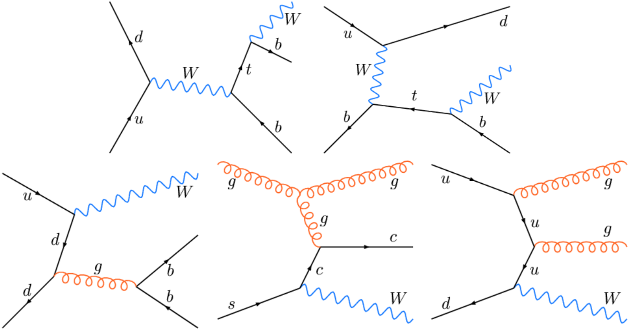
CDF discriminant

$$EPD = \frac{b \cdot P_{signal}}{b \cdot P_{signal} + b \cdot P_{Wbb} + (1-b) P_{Wcc} + (1-b) P_{Wcj}}$$

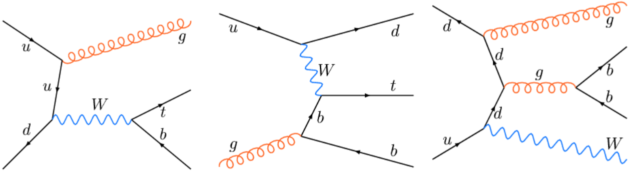
- b is the neural network b-tagger output converted to probability

Matrix element method - D0 diagrams

2-jets:



3-jets:



Bayesian neural networks

A different sort of neural network

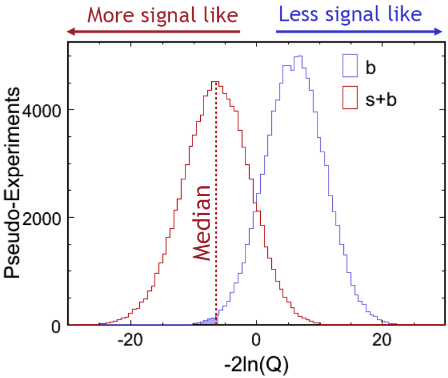
- Instead of choosing one set of weights, find posterior probability density over all possible weights
- Averaging over many networks weighted by the probability of each network given the training data
- Less prone to overtraining
- For details see: http://www.cs.toronto.edu/∼radford/fbm.software.html
- Use 24 variables (subset of DT variables)

Sensitivity determination at CDF

- Using the CLs method developed at LEP
- Compare two models at a time
- Test statistic:

$$Q = \frac{L(data|s+b)}{L(data|b)}$$

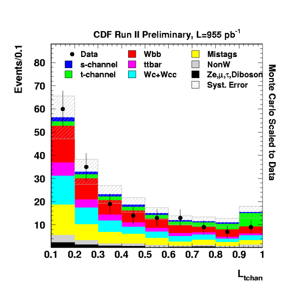
- Systematic uncertainties included in pseudo-experiments
- Expected sensitivity: median p-value



Likelihood	median p-value = 2.3%	(2.0σ)
Matrix element	median p-value = 0.6%	(2.5σ)
Neural network	median p-value = 0.5%	(2.6σ)

CDF s+t observed results — Preliminary

Likelihood

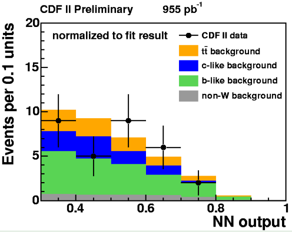


No evidence of signal

$\sigma < 2.7$ pb @ 95% CL

From s and t likelihoods

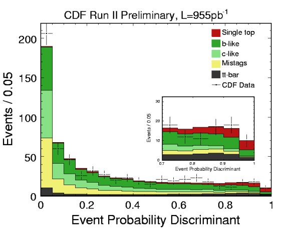
Neural network



No evidence of signal

$\sigma < 2.6$ pb @ 95% CL

Matrix element

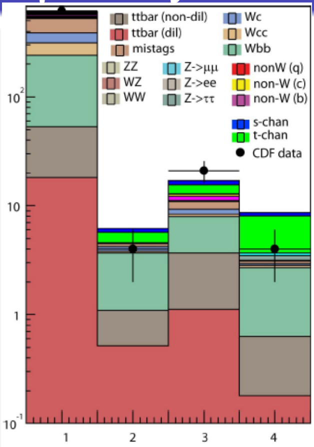


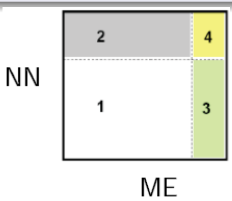
$\sigma = 2.7^{+1.5}_{-1.3}$ pb

p-value = 1.0% (2.3σ)

CDF observed results — Compatibility

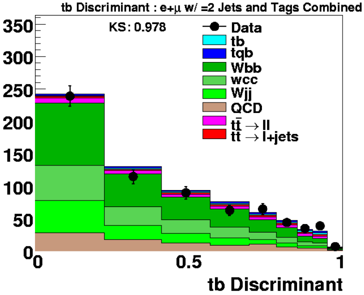
- CDF spent great deal of time (6 months) and effort understanding if the different results are something more than a statistical fluctuation.
- Eliminated possibility of obvious and even subtle bugs
- 6-discriminant compatibility coming soon
- Now investigating if features of the MC modeling affect one analysis more than the other.
- Analysing more data should shed some light

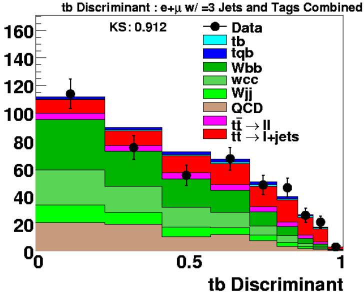


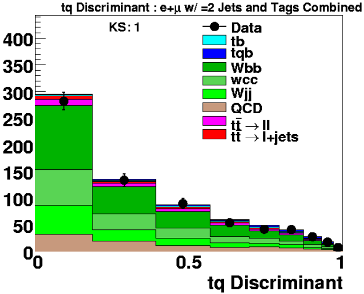


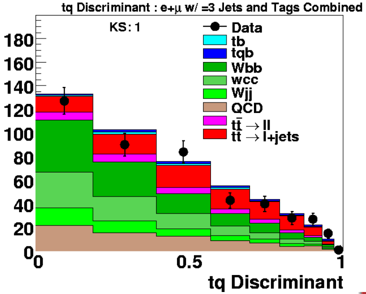
Bin 1: NN<0.8 && EPD<0.9
Bin 2: NN>0.8 && EPD<0.9
Bin 3: NN<0.8 && EPD>0.9
Bin 4: NN>0.8 && EPD>0.9

Matrix element outputs

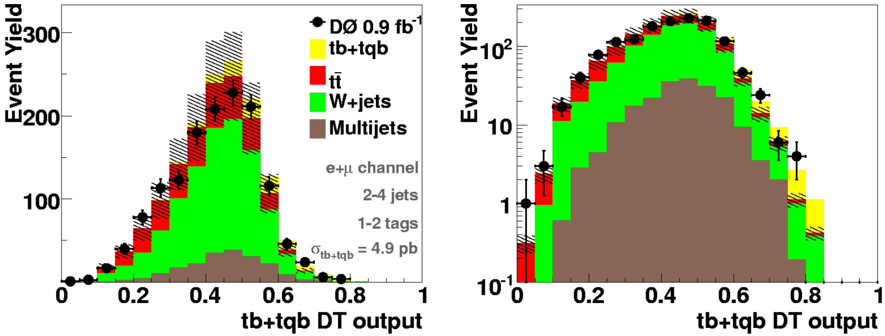






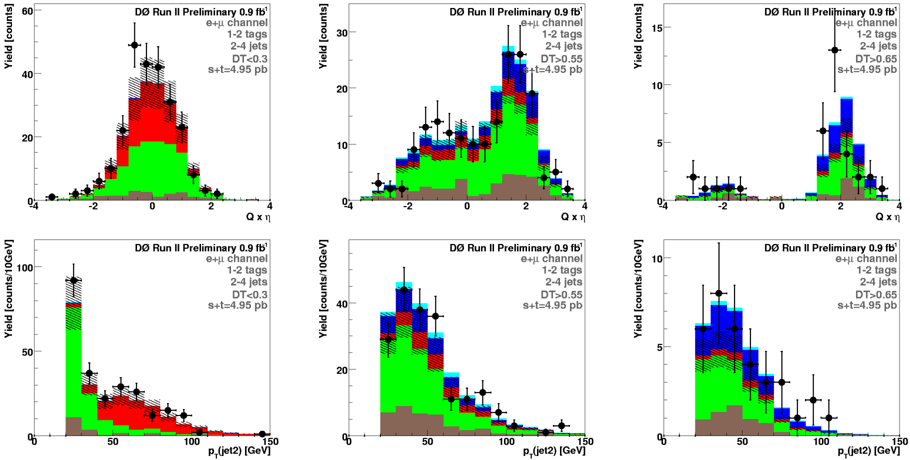


Decision tree combined outputs



Boosted decision tree event characteristics

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



Single top prospects — Tevatron and LHC

Tevatron

- By 2008 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:

$$\sigma_s^{t/\bar{t}} = 6.6/4.1 \text{ pb } (\pm 10\%)$$

$$\sigma_t^{t/\bar{t}} = 156/91 \text{ pb } (\pm 5\%)$$

$$\sigma_{tW}^{t/\bar{t}} = 34/34 \text{ pb } (\pm 10\%)$$

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