Evidence for single top quark production at the Tevatron

Yann Coadou
on behalf of the CDF and DØ collaborations

Simon Fraser University

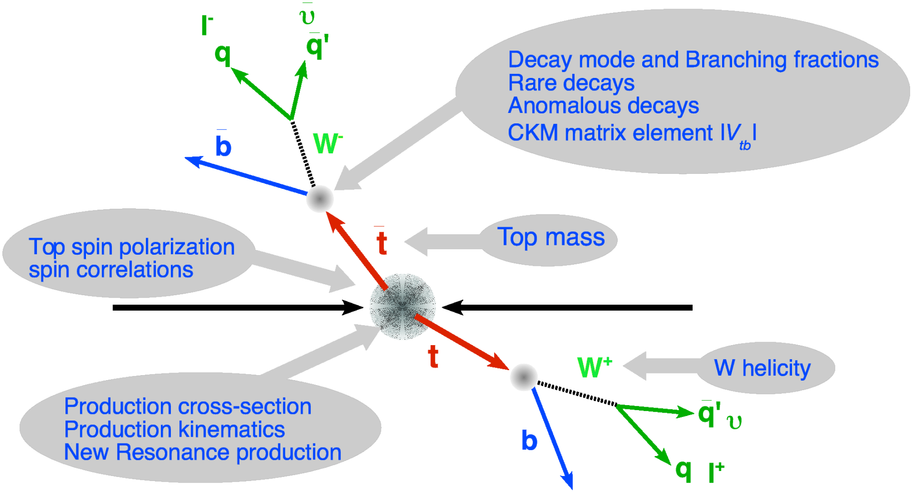
2007 Aspen Winter Conference
Aspen Center for Physics
10 January 2007

Outline

1. **Single top production**
2. **Preparing for the measurement**
   - Event selection
   - Backgrounds
   - b tagging
3. **Multivariate analysis techniques**
4. **Expected sensitivity**
5. **Cross sections and significance**
6. **First direct measurement of $|V_{tb}|$**
7. **Conclusions**

Top quark physics

- The Tevatron is still the only place to make top quarks.

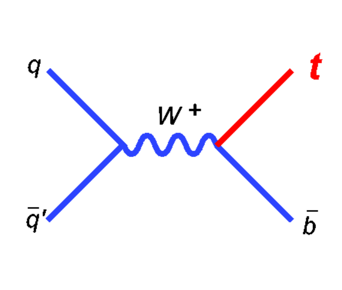


- Other predicted production mode: single top

Single top quark production

- Electroweak production in two main mechanisms at the Tevatron:

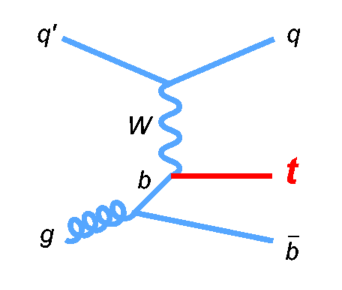
s-channel (tb)



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



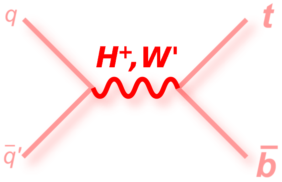
- $\sigma_{NLO} = 1.98 \pm 0.25$ pb(*)
- previous limits (95% C.L.):

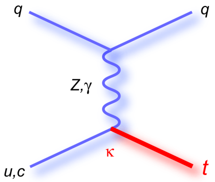
Run II DØ: < 4.4 pb (370 pb^{-1})
Run II CDF: < 3.2 pb (700 pb^{-1})

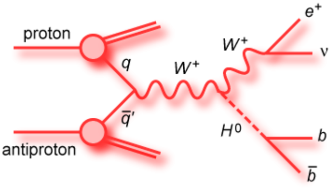
(*) $m_t = 175$ GeV, Phys.Rev. D70 (2004) 114012

Motivation

- Directly measure $|V_{tb}|$ (more later)
- Cross sections sensitive to new physics:
  - s-channel: resonances (heavy W' boson, charged Higgs boson $H^{\pm}$, Kaluza-Klein excited W_{KK}, etc...)
  - t-channel: flavour-changing neutral currents ($t - Z/\gamma/g - c$ couplings)
  - Fourth generation of quarks
- Source of polarized top quarks. Spin correlations measurable in decay products
- Important background to WH associated Higgs production
  - if the tools don't work for single top, forget about the Higgs
- Test of techniques to extract a small signal out of a large background

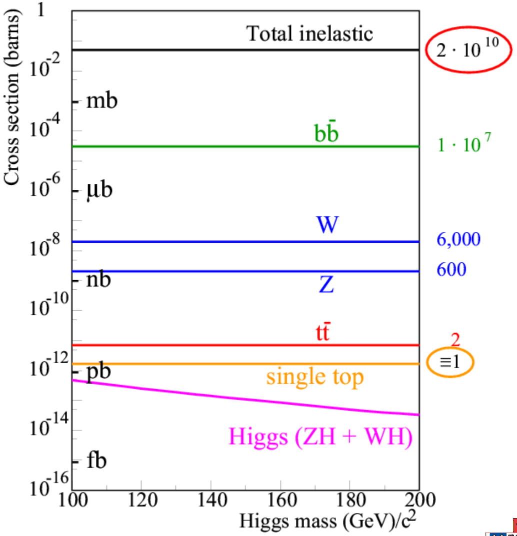




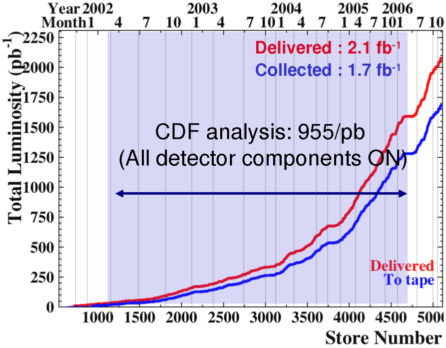


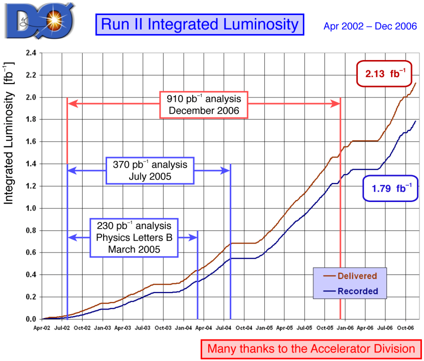
It has been challenging for years...

- 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

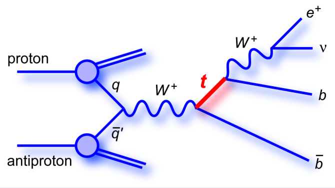


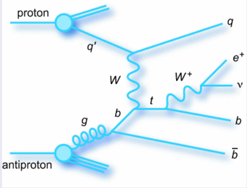
Tevatron luminosity





Event selection





Signature

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

	CDF	DØ
1 lepton	p_T^e>20 GeV, $\|\eta_e\| < 2$ p_T^μ>20 GeV, $\|\eta_\mu\| < 1.1$	p_T^e>15 GeV, $\|\eta_e\|$<1.1 p_T^μ>18 GeV, $\|\eta_\mu\|$<2.0
jets	exactly 2 p_T>15 GeV, $\|\eta\| < 2.8$	2,3,4 p_T>15 GeV, $\|\eta\|$<3.4 leading jet p_T>25 GeV, $\|\eta\|$<2.5 2nd leading jet $p_T > 20$ GeV
MET	$\not{E}_T > 25$ GeV	$15 < \not{E}_T < 200$ GeV
b jet	one or two	

Event selection - S/B

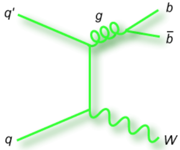
Percentage of single top *tb+tqb* selected events and S:B ratio (white squares = no plans to analyze)

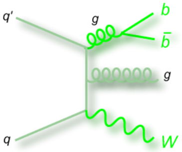
Electron + Muon	1 jet	2 jets	3 jets	4 jets	≥ 5 jets
0 tags	10% 1 : 3,200	25% 1 : 390	12% 1 : 300	3% 1 : 270	1% 1 : 230
1 tag	6% 1 : 100	21% 1 : 20	11% 1 : 25	3% 1 : 40	1% 1 : 53
2 tags		3% 1 : 11	2% 1 : 15	1% 1 : 38	0% 1 : 43

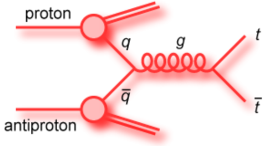
- Slightly different naming conventions and techniques between the two experiments but very similar in the end

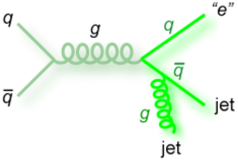
Main backgrounds

- W+jets (Alpgen, normalized to data):
  - W+heavy flavour: Wbb, Wbj, Wcc, Wcj, Wc
  - W+light jets ("mistags")
- $t\bar{t}$ (DØ Alpgen, CDF Herwig, $m_t = 175$ GeV, $\sigma_{NNLO} = 6.8$ pb)
- QCD (a.k.a. multijet, non-W) (from data failing lepton ID)

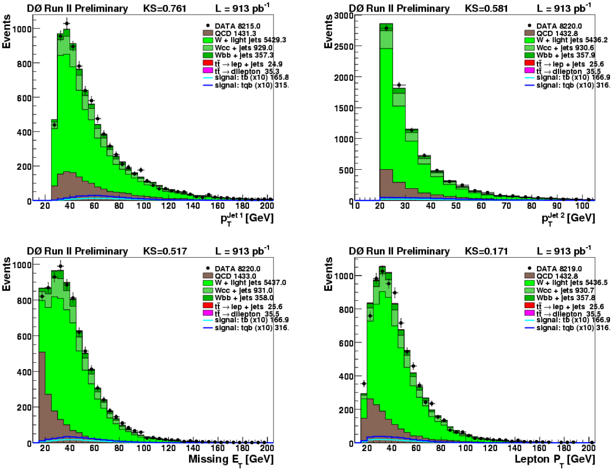








Event selection - Agreement before tagging

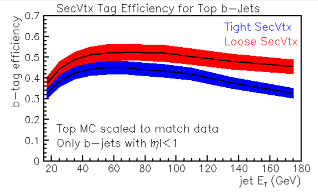


- Normalize W+multijet to data before tagging
- Checked 90 variables, 3 jet multiplicities, 1-2 tags, electron + muon
- Shown: electron, 2 jets, before tagging
- Good description of data

CDF b tagging

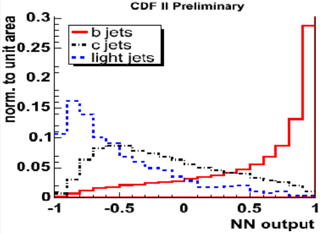
Secondary vertex tagging

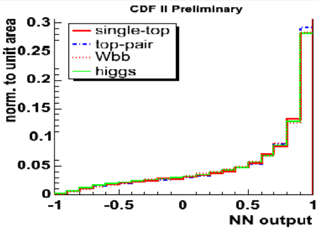
- Long lifetime of B hadrons
- Travel several mm before decaying
- Signature: displaced secondary vertex tagger
- Tagging efficiency per jet $\sim$ 40%



Jet flavour separation

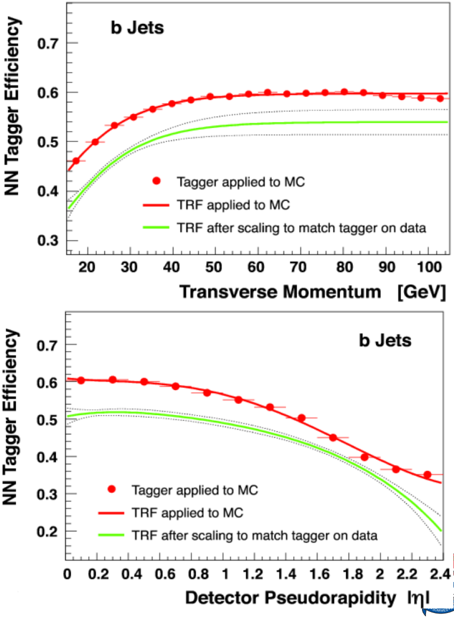
- Second stage: improve separation with 25-input neural network
- Applied on jets b-tagged with secondary vertex
- Good jet flavour separation, independent of b-jet source





DØ b tagging

- NN trained on 7 input variables from existing taggers.
- 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%



Systematic uncertainties - CDF

CDF RunII Preliminary, L=955pb^{-1}

Single Top	Rate Variations	Shape Variations
Jet Energy Scale	✓	✓
Initial State Radiation	✓	✓
Final State Radiation	✓	✓
Parton Dist. Function	✓	✓
Monte Carlo Generator	✓	
Efficiencies / b-tagging SF	✓	
Luminosity	✓	
Total Rate Uncertainty	10.5%	N/A

Backgrounds	Rate Variations	Shape Variations
Jet Energy Scale	✓	✓
Neural Net b-tagger		✓
Mistag Model		✓
Non-W Model		✓
Q^2 Scale in Alpgen MC		✓

Background	Rate Uncertainty
W+bottom	25%
W+charm	28%
Mistag	15%
ttbar	23%

- Rate and shape uncertainties included as nuisance parameters in analyses

Systematic uncertainties - DØ

- 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%

Event Selection - Yields

DØ

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
$t\bar{t} \to ll$	39 ± 9	32 ± 7	11 ± 3
$t\bar{t} \to l$+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

CDF

s-channel	15.4 ± 2.2
t-channel	22.4 ± 3.6
tt	58.4 ±13.5
Diboson	13.7 ± 1.9
Z + jets	11.9 ± 4.4
Wbb	170.9 ± 50.7
Wcc	63.5 ± 19.9
Wc	68.6 ± 19.0
Non-W	26.2 ± 15.9
Mistags	136.1 ± 19.7
Single top	37.8 ± 5.9
Total background	549.3 ± 95.2
Total prediction	587.1 ± 96.6
Observed	644

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

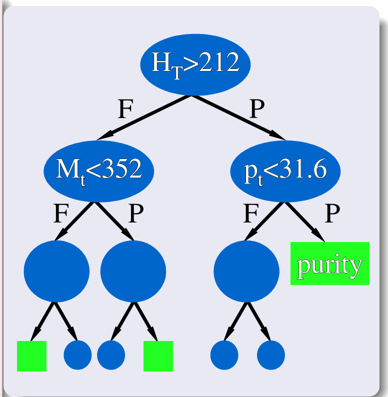
Multivariate analysis techniques

- Likelihood discriminants (CDF)
- Artificial neural network (CDF)
- Matrix element (DØ, CDF)
- Bayesian neural networks (DØ)
- Boosted decision trees (DØ)

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



- Run testing events and data through tree to derive limits
- DT output = leaf purity, close to 1 (0) for signal (bkg)

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

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
- 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 ⇒ 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 trees at DØ

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

Input variables

- Used 49 variables (object and event kinematics, angular correlations)
- Adding variables does not degrade performance
- Tested shorter lists: lost some sensitivity
- Same list used for all channels

Analysis strategy

- Train 36 separate trees: $(s,t,s+t)$ x (e,μ) x (2,3,4 jets) x (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 background diagrams to compute an event probability density for signal and background hypotheses
- Encoded in properly normalized differential cross section for process S:

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

- Only a limited number of Feynman diagrams are used. 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)

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

Likelihood method (CDF)

- Likelihood for a vector of measurements $\vec{x} = x_i$:

$$\mathcal{L}(\vec{x}) = \frac{\mathcal{P}_{signal}(\vec{x})}{\mathcal{P}_{signal}(\vec{x}) + \sum \mathcal{P}_{background}(\vec{x})}, \quad \mathcal{P}(\vec{x}) = \prod_{i}^{N_{variables}} P(x_i)$$

$P(x_i)$ = normalized x_i variable distribution

- Four backgrounds: Wbb, $t\bar{t}$, Wcc/Wc, mistags

t-channel LF Variables:

- total transverse energy: H_T
- $M_{l\nu b}$ (neutrino p_z from kin. fitter)
- $\cos\theta$(lepton,light jet) in top decay frame
- $Q_{lepton} * \eta_{untagged\ jet}$ aka QxEta
- m_{j1j2}
- $\log(ME_{tchan})$ from MADGRAPH
- Neural Network b-tagger
- LF=0.01 for double tagged events

s-channel LF Variables:

- $M_{l\nu b}$
- $\log(H_T{}^* M_{l\nu b})$
- E_T(jet1)
- $\log(ME_{tchan})$
- H_T
- Neural Network b-tagger

Neural network — Bayesian neural networks

Neural network (CDF)

- Three-layer perceptrons using NeuroBayes
- Continuous output between -1 (bkg-like) and +1 (signal-like)
- 26 input variables
- Three networks: tb, tqb and tb+tqb and signal

Bayesian neural networks (DØ)

- 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)

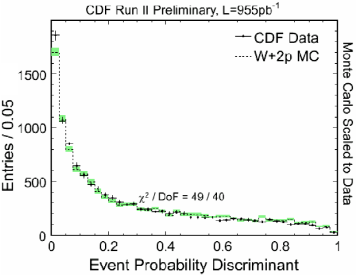
DØ analysis validation

- 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
- All analyses achieved linear response to varying input cross sections

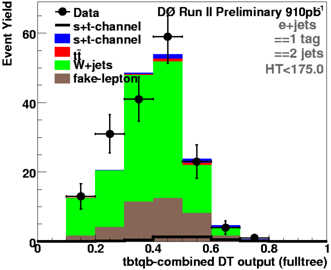
Cross-check samples

- Validate methods using data without looking at signal
- Compare discriminant in model and data
- Good agreement observed

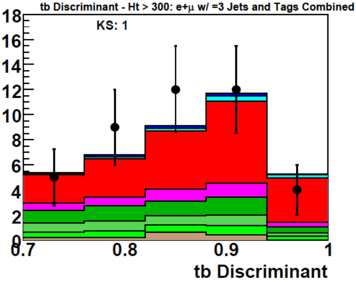
ME W+2jets data (b-jet veto)



DT "W+jets": =2jets, $H_T < 175$ GeV



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

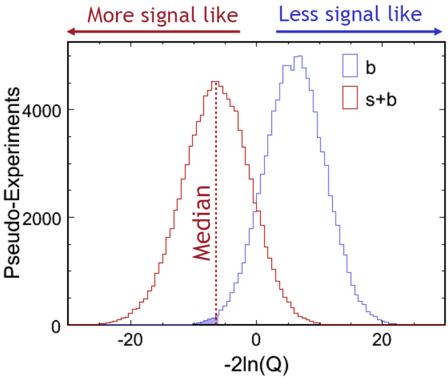


Sensitivity determination at CDF

- 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)

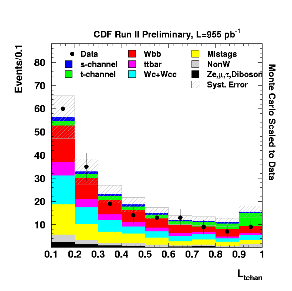
Sensitivity determination at DØ

- Use the 0-signal ensemble:
  - use pool of weighted signal+bkg events
  - fluctuate relative and total yields in proportion to syst. errors
  - randomly sample from a Poisson distribution about total yield
  - generate a set of pseudo data
  - pass the pseudo-data through the full analysis
- 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.

Boosted decision trees	p-value = 1.9%	(2.1σ)
Matrix element	p-value = 3.7%	(1.8σ)
Bayesian neural networks	p-value = 9.7%	(1.3σ)

CDF s+t observed results — Preliminary

Likelihood

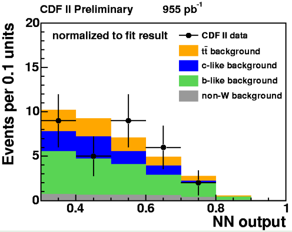


No evidence of signal

$\sigma < 2.7$ pb @ 95% CL

From s and t likelihoods

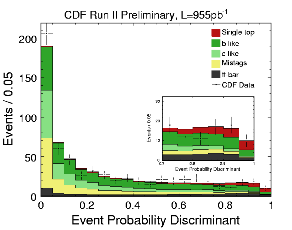
Neural network



no evidence of signal

$\sigma < 2.6$ pb @ 95% CL

Matrix element

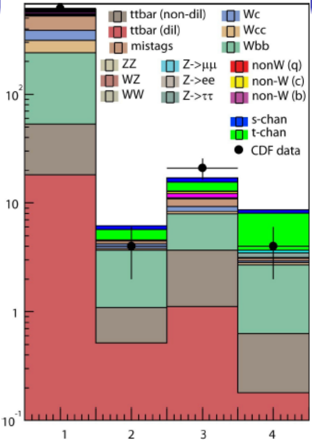


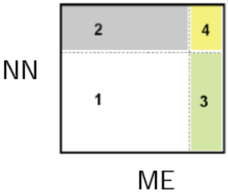
$\sigma = 2.7^{+1.5}_{-1.3}$ pb

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

Bayesian NN

$\sigma = 5.0 \pm 1.9$ pb

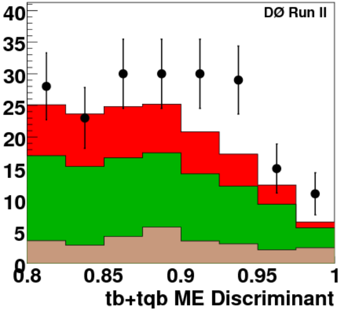
p-value $= 0.89\%$ (2.4σ)

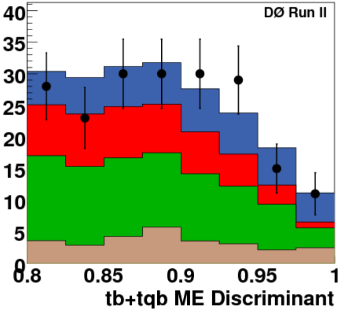
Matrix element

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

p-value $= 0.21\%$ (2.9σ)

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



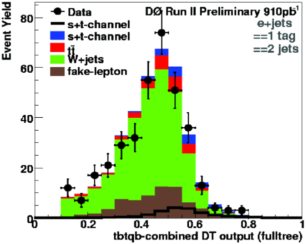


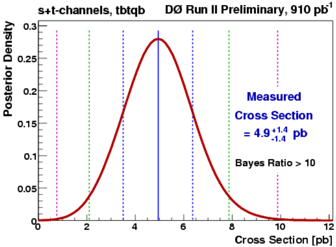
DØ boosted decision tree s+t observed results

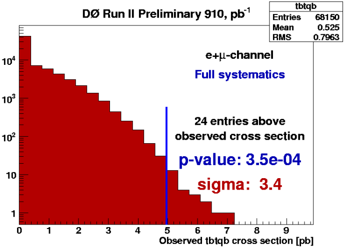
$$\sigma = 4.9 \pm 1.4 \text{ pb}$$

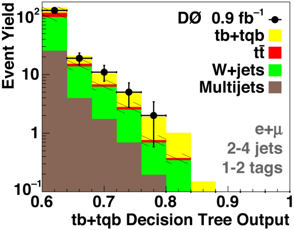
p-value = 0.035% (3.4σ)

SM compatibility: 11% (1.1σ)



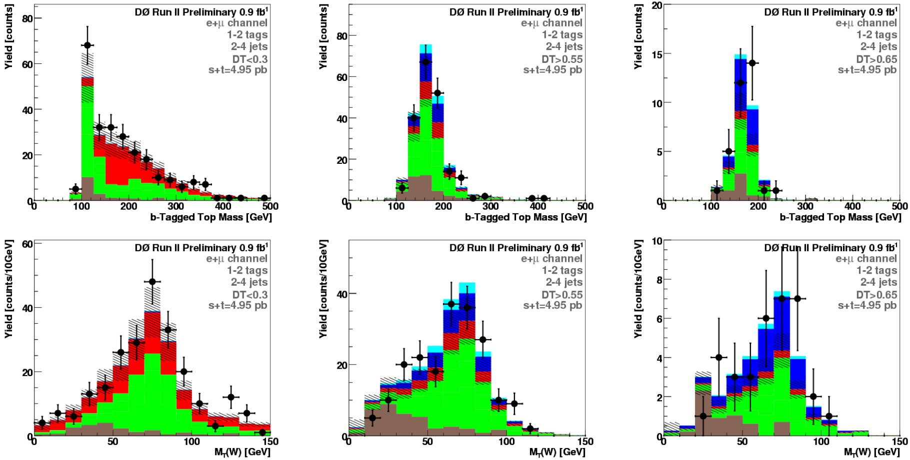




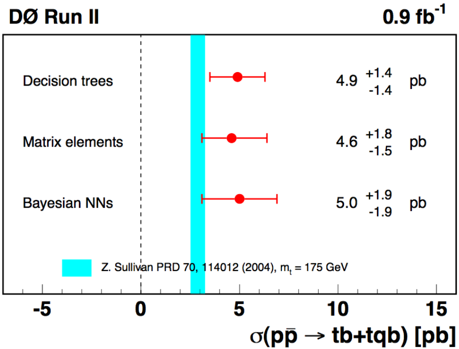


DØ boosted decision tree event characteristics

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



DØ results consistency



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%

First direct 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}: $|V_{tb}| = 0.9991$
- New physics:
  - $V_{td}^2 + V_{ts}^2 + V_{tb}^2 < 1$
  - no constraint on V_{tb}

Result

- Translate tb+tqb cross section into measurement of the strength of $V-A$ coupling $|V_{tb} f_1^L|$ in Wtb vertex (f_1^L: arbitrary left-handed form factor)
- Assume $V_{td}^2 + V_{ts}^2 \ll V_{tb}^2$ and pure $V-A$ and CP-conserving Wtb interaction

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

- Also assuming $f_1^L = 1$:

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

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

Conclusions

- CDF and DØ have been searching for single top signal for years
- A lot of energy invested in the experimental challenges
  - very small signal hidden in enormous background
  - efficient *b*-tagging
  - background modeling (involving data and Monte Carlo)
- Several multivariate techniques being used
- CDF analyses have good sensitivity but got unlucky (2.3σ signal with ME, LF and NN don't see any single top)
- DØ BNN and ME analyses see 2.4σ and 2.9σ signal

Conclusions

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

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

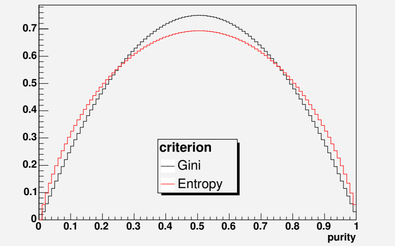
- Decrease of impurity for split s of node t into children t_L and t_R (goodness of split): $\Delta i(s,t) = i(t) - p_L \cdot i(t_L) - p_R \cdot i(t_R)$
- Aim: find split s^* such that:

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

Examples

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

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



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

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

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,besttopframe)$_{\mathrm{besttopCMframe}}$
cos(lepton,btaggedtopframe)$_{\mathrm{btaggedtopCMframe}}$
cos(notbest,alljets)$_{\mathrm{alljets}}$
cos(notbest,lepton)$_{\mathrm{besttop}}$
cos(untag1,alljets)$_{\mathrm{alljets}}$
cos(untag1,lepton)$_{\mathrm{btaggedtop}}$

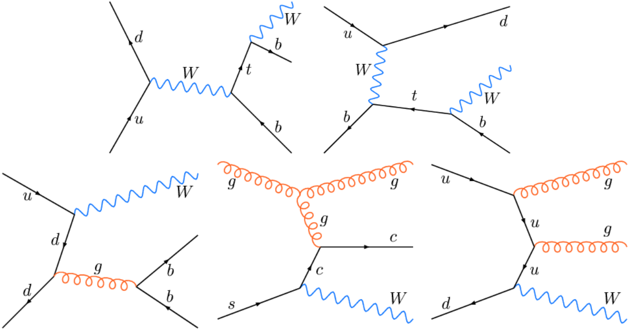
Event Kinematics

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)

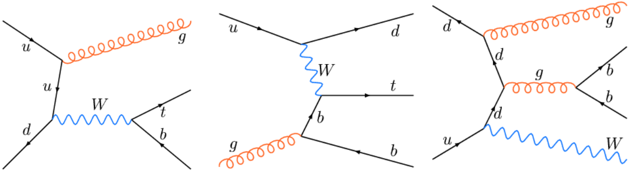
- Adding variables does not degrade performance
- Tested shorter lists, lose some sensitivity
- Same list used for all channels

Matrix element method - D0 diagrams

2-jets:

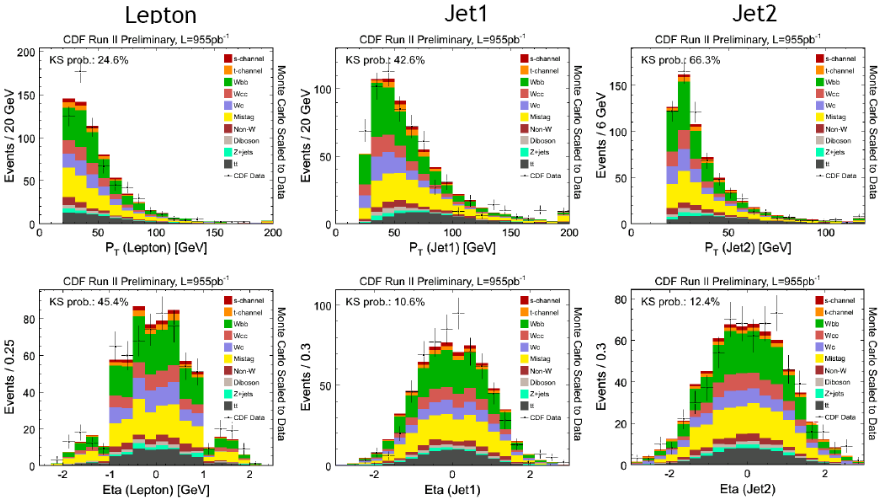


3-jets:

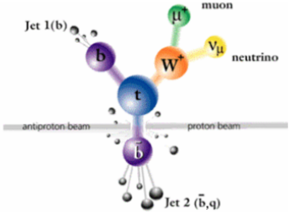


CDF ME inputs (B. Stelzer)

- Input to the Matrix Element Analysis are the measured four-vectors of the Lepton, Jet1 and Jet2 in the W+2jets data (>=1 b-tagged jet)



Matrix element method - Probability (B. Stelzer)



Single Top kinematic quantities:

- 2(*initial*) + 12(*final*) = 14 degrees of freedom
- Assume leptons and angles well measured
- 3(l)+4(angle)+3(P_{in}=P_{fin})+1(E_{in}=E_{fin}) = 11 constraints
- 14 - 11 = 3 integrals => Integrate over neutrino p_z and jet energy of both jets.

Event probability for signal and background hypothesis:

Leading Order matrix element (MadEvent)

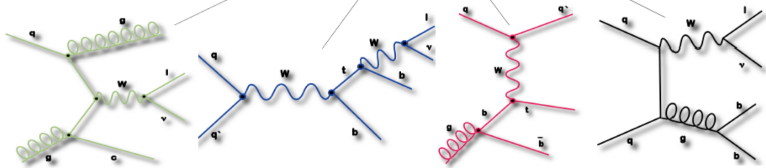
W(E_{jet},E_{part}) is the probability of measuring a jet energy E_{jet} when E_{part} was produced

$$P(p_l^\mu, p_{j1}^\mu, p_{j2}^\mu) = \frac{1}{\sigma}\int d\rho_{j1} d\rho_{j2} dp_\nu^z \sum_{comb} |M(p_i^\mu)|^2 \frac{f(q_1)f(q_2)}{|q_1||q_2|}\phi_4 W_{jet}(E_{jet}, E_{part})$$

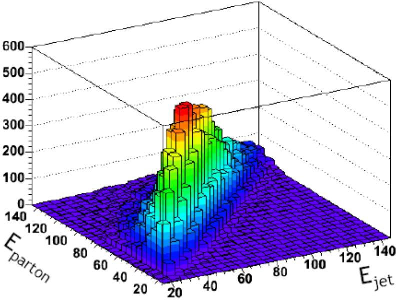
Input only lepton and 2 jets 4-vectors!

Integration over part of the phase space Φ_4

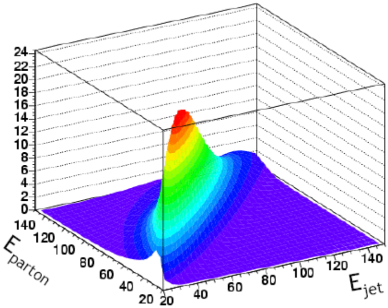
Parton distribution function (CTEQ5)



Full simulation vs parton energy:



Double Gaussian parameterization:



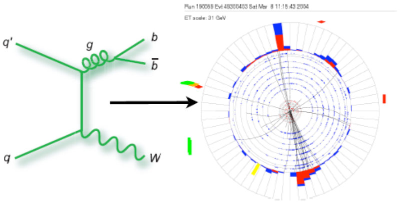
Double Gaussian parameterization:

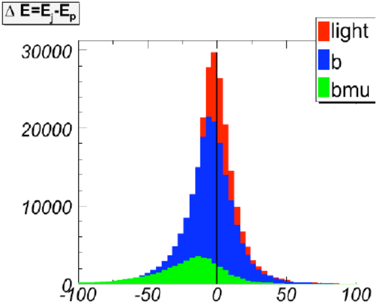
$$W_{jet}(E_{jet},E_{parton})=\frac{1}{\sqrt{2\pi}(p_1+p_2p_5)}[\exp\frac{-(\delta_E-p_1)^2}{2p_2^2}+p_3\exp\frac{-(\delta_E-p_4)^2}{2p_5^2}]$$

where: $p_i = a_i + b_i E_{parton}$ $\quad \delta E = (E_{parton}-E_{jet})$

Matrix element method - DØ transfer functions

- To evaluate $|M|^2$, we must have intial/final state 4-vectors.
  - W(x,y) relates final state **y** to detector state **x**
- Jets
  - Assume angles well measured and Sole dependence on energy difference
  - Calculate for 3 types of jets: light, b, and b w/ mu
- Electrons
  - Assume angles well measured and sole dependence on energy difference
- Muons
  - Dependence on q/P_T, η, and number of SMT hits





W+jets heavy flavour fraction at DØ

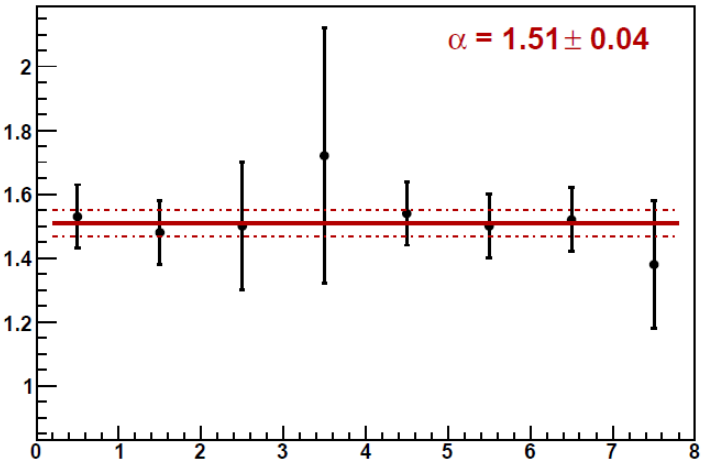
$$\alpha(Wb\bar{b} + Wc\bar{c}) + Wjj + t\bar{t} + \text{QCD} = \text{Data}$$

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

HF Fraction - DØ

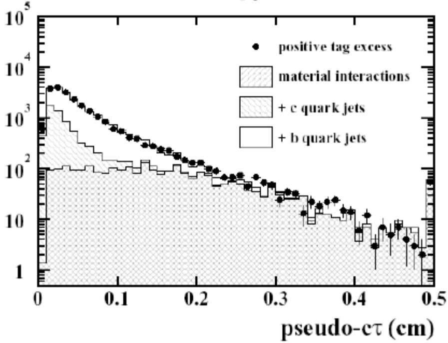
Heavy flavour scale factor α measured in the zero tag bins



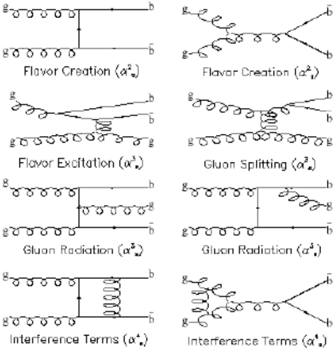
1) Estimate generic jet heavy flavor fraction in ALPGEN Monte Carlo
2) Fit for bottom and charm fraction in generic jet data

Difference between the two outcomes suggests K=1.5±0.4
Result supported by study using MCFM:

$$pseudo-c\tau = L_{2D} \cdot \frac{M^{vtx}}{p_T^{vtx}}$$



J. M. Campbell, J. Houston, Method 2 at NLO, hep-ph/0405276



Binned likelihood fit at CDF (B. Stelzer)

Binned Likelihood Function:

$$\mathcal{L}(\beta_1,\ldots,\beta_5;\delta_1,\ldots,\delta_{10}) = \underbrace{\prod_{k=1}^{B} \frac{e^{-\mu_k}\cdot\mu_k^{n_k}}{n_k!}}_{Poisson\ term} \cdot \underbrace{\prod_{j=2}^{5} G(\beta_j|1,\Delta_j)}_{Gauss\ constraints} \cdot \underbrace{\prod_{i=1}^{10} G(\delta_i,0,1)}_{Systematics}$$

Expected mean in bin k:

$$\mu_k = \sum_{j=1}^{5}\beta_j \cdot \underbrace{\left\{\prod_{i=1}^{10}[1+|\delta_i|\cdot(\epsilon_{ji+}H(\delta_i)+\epsilon_{ji-}H(-\delta_i))]\right\}}_{Normalization\ Uncertainty}$$

$$\underbrace{\cdot\,\alpha_{jk}}_{Shape\ P.} \cdot \underbrace{\left\{\prod_{i=1}^{10}(1+|\delta_i|\cdot(\kappa_{jik+}H(\delta_i)+\kappa_{jik-}H(-\delta_i)))\right\}}_{Shape\ Uncertainty}$$

$B_j = \sigma_j/\sigma_{SM}$ parameter
single top (j=1)
W+bottom (j=2)
W+charm (j=3)
Mistags (j=4)
ttbar (j=5)

k = Bin index

i = Systematic effect

δ_i = Strength of effect

$\varepsilon_{ji\pm}$ = ±1σ norm. shifts

$\kappa_{jik\pm}$ = ±1σ shift in bin k

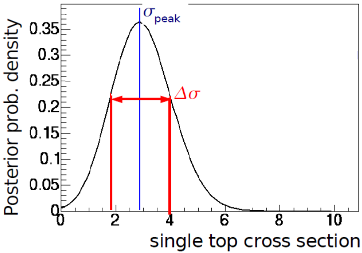
- All sources of systematic uncertainty included as nuisance parameters
- Correlation between Shape/Normalization uncertainty considered (δ_i)

Measuring cross sections at DØ

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

$|V_{tb}|$ determination

- No assumptions on the number of families or unitarity of the CKM matrix
- However, some other model assumptions have been made
- It is assumed that the only existing production mechanism of single top quarks involves the interaction with a W boson (models where single top quark events can be produced e.g. via FCNC interactions or heavy scalar or vector boson exchange, are not considered)
- Assuming $|V_{td}|^2 + |V_{ts}|^2 \ll |V_{tb}|^2$, implying $B(t \to Wb) \simeq 100\%$
- Finally, tbW interaction is CP-conserving and of the $V{-}A$ type, but it is allowed to have an anomalous strength
- Most general tbW vertex:

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

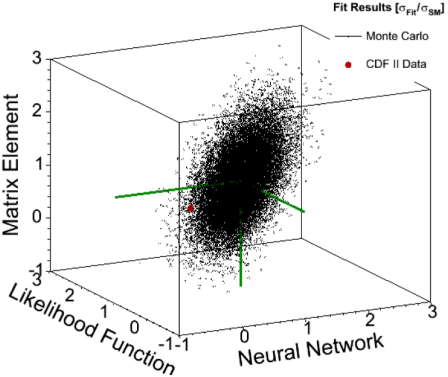
- SM: CP is conserved in the tbW vertex, $f_1^L = 1$ and $f_1^R = f_2^L = f_2^R = 0$

CDF compatibility study

- ME, NN and LF analyses use same input dataset and MC events, but results differ

Potential sources of differences

- ME uses transfer functions
- ME does not use missing ET
- ME integrates over all neutrino p_z, while NN chooses the solution with smaller $|p_z|$
- with two jets in the event, the NN choose the secondary-vertex-tagged jet as the b jet from top quark decay. The ME sums over both possibilities
- NN also allows for soft jets $(8 < E_T(jet) < 15$ GeV$)$



- LF(t), ME(s+t), NN(s+t)
- Coming: all six discriminants: LF(s), LF(t), ME(s+t), NN(s), NN(t), NN(s+t)

CDF compatibility study (cont'd)

- Overlap between 5% highest-ME and 5% highest-NN(s+t) events is 30(43)% for s(t)-channel (left plot)
- Impact of transfer functions (middle plot): NN needs better-measured jets in signal region (close to 0) than ME. Significant effect in t-channel only (black/blue curves)
- Missing ET measurement (right plot): NN needs better-measured MET in signal region (close to 0) than ME