Rencontres de Moriond
Electroweak Interactions and Unified Theories

La Thuile, March 12, 2007

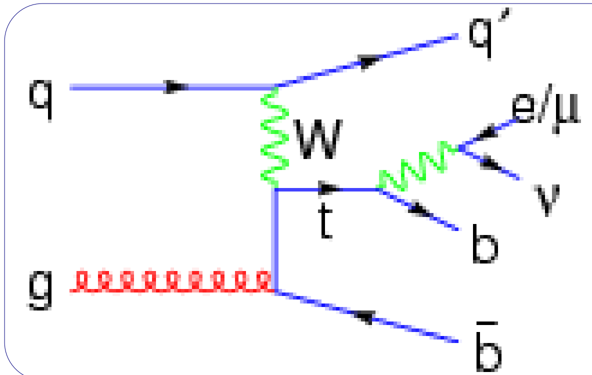
Evidence for single top at DØ

- Electroweak production of top quarks
- Event selection and background estimation
- Multivariate methods
  - Decision Trees, Matrix Elements, Bayesian NN
- Cross checks. Expected sensitivity
- Cross sections and significance
- First direct measurement of $|V_{tb}|$
- Combination
- Summary

UNIVERSITY OF WASHINGTON

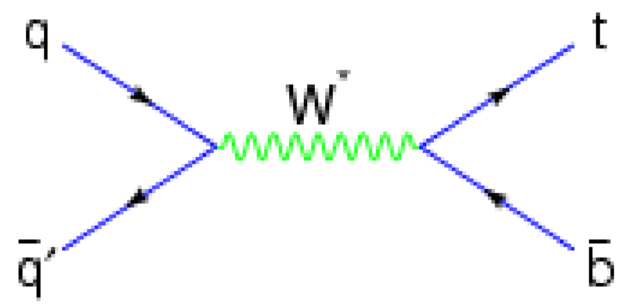
Arán García-Bellido
for the DØ collaboration

Signal selection



t-channel

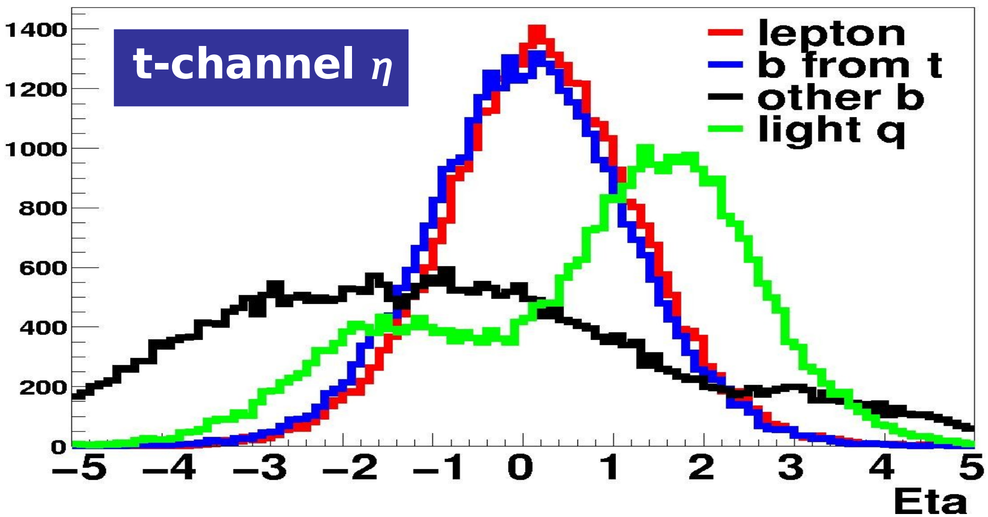
1.98±0.25 pb

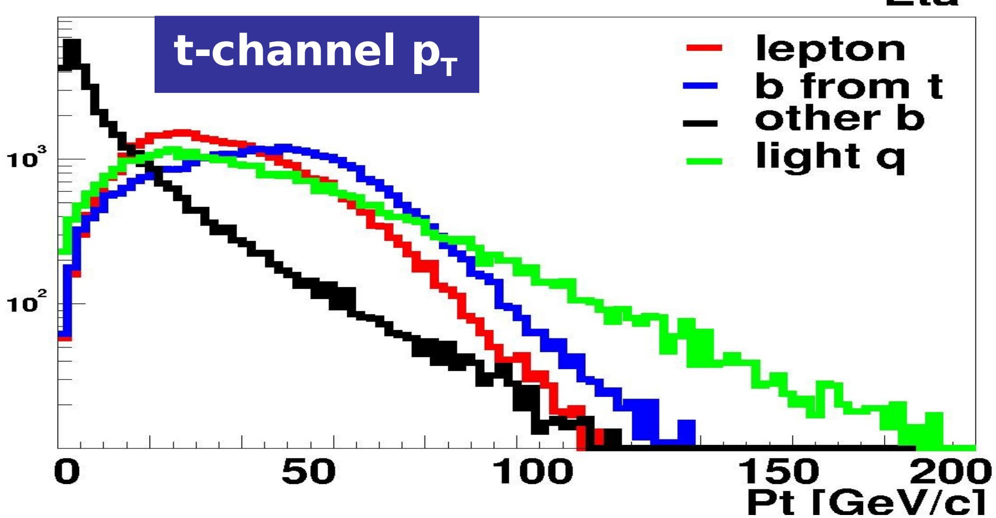


s-channel

0.88±0.11 pb

PRD 66 (02) 054024
hep-ph/0408049





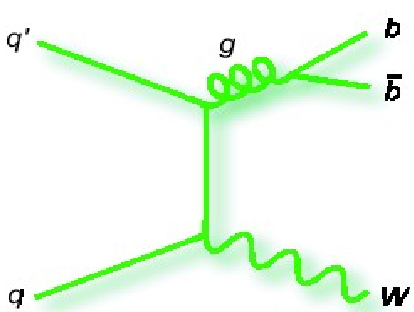
Event selection designed to be as loose as possible:

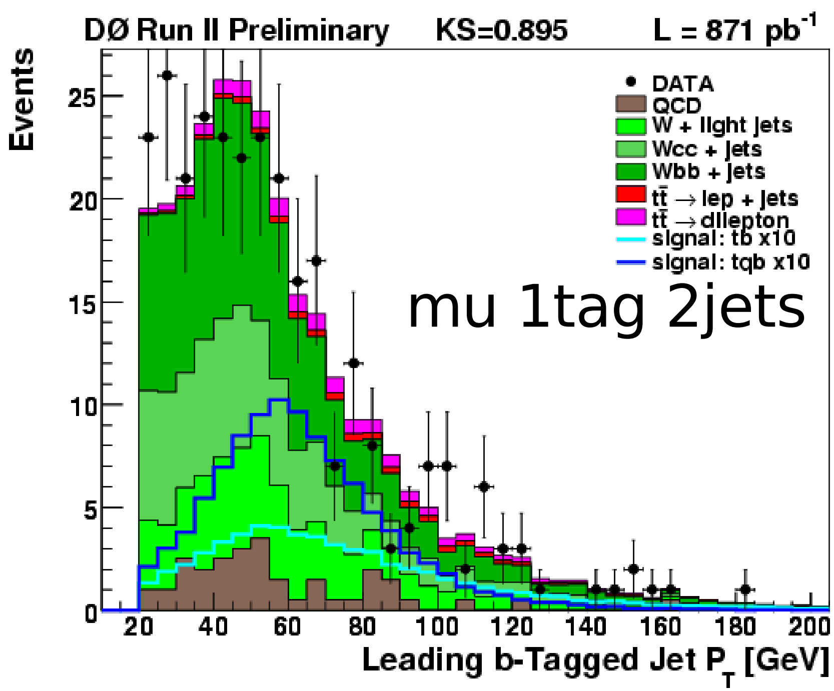
- Only one tight (no loose) lepton:
  - e: p_T >15 GeV and $|\eta^{det}|$<1.1
  - μ: p_T >18 GeV and $|\eta^{det}|$<2.0
- MET > 15 GeV
- 2-4 jets: p_T >15 GeV and $|\eta^{det}|$<3.4
  - Leading jet: p_T>25GeV ; $|\eta^{det}|$<2.5
  - Second leading jet: p_T >20 GeV
- One or two b-tagged jets

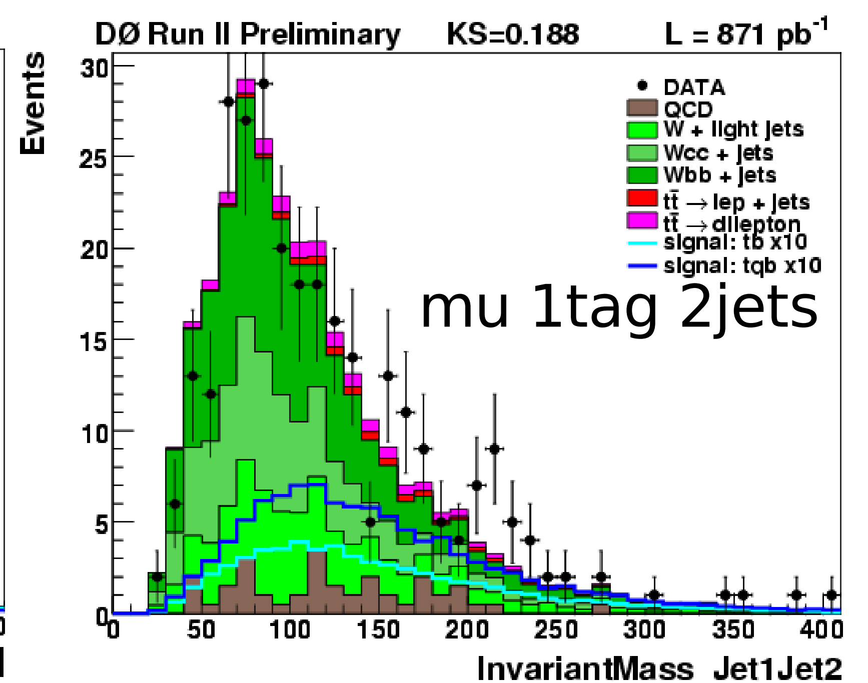
Acceptance: tb=(3.2 ± 0.4)%
tqb=(2.1 ± 0.3)%

Signal and Background modeling

- Signal is modeled with CompHEP (effective NLO) + Pythia
- W+jets and ttbar shapes from Alpgen with MLM matching + Pythia
  - Jet-parton matching avoids double counting → better model
- ttbar normalized to NNLO σ = 6.8±1.2 pb (Kidonakis, PRD 68, 114014)
- QCD from our selected data with non-isolated lepton
- Normalize W+jets and QCD to data before tagging
- Determine Wbb and Wcc factor in W+jets from zero-tagged data
  - Constant factor describes heavy flavor kinematics well
  - Largest single uncertainty: 30% relative error on Wbb+Wcc composition







Arán García-B

Yields and systematic uncertainties

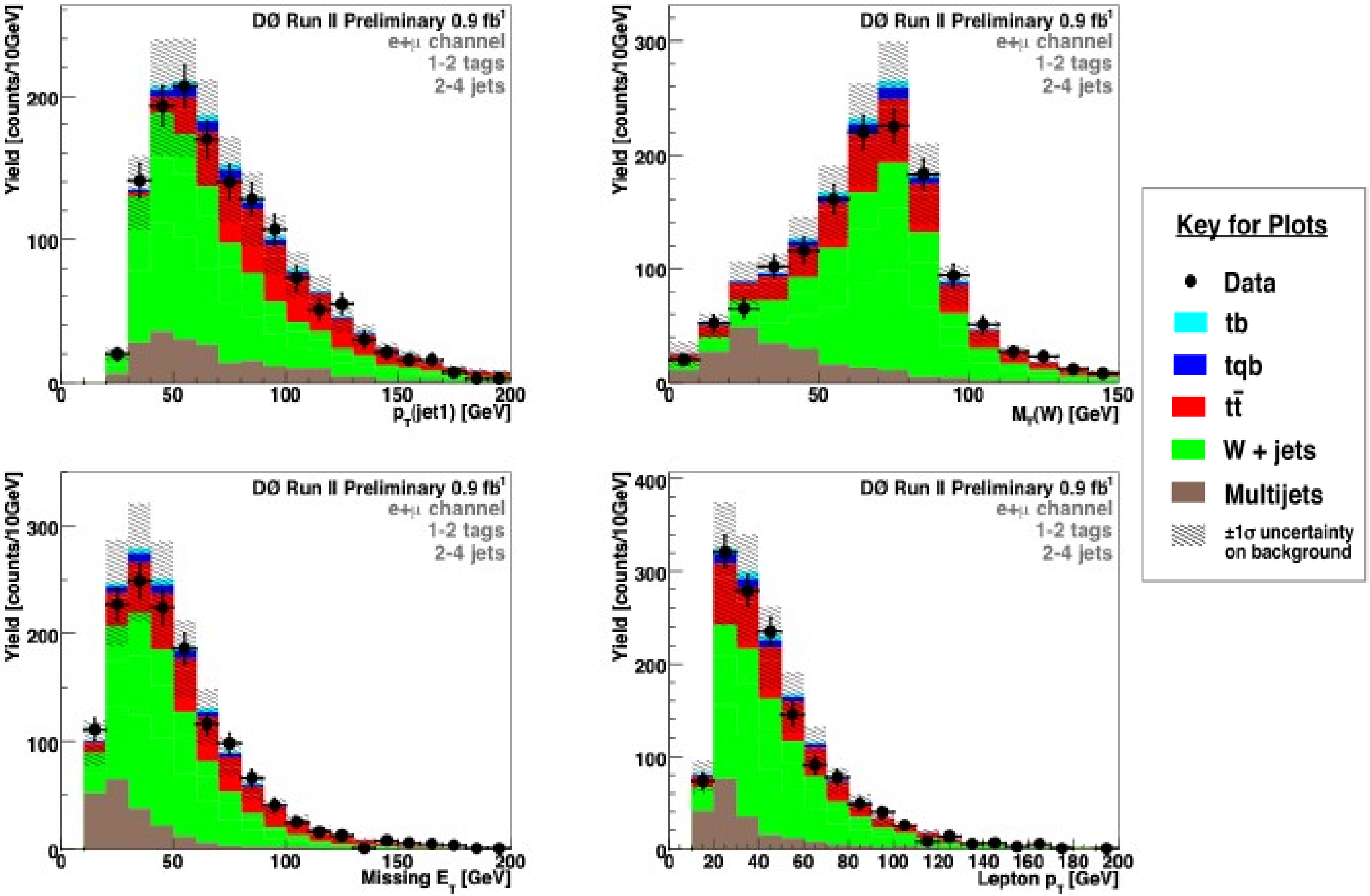
- Expect some 62 signal and 1400 background events
- Uncertainties are assigned per background, jet multiplicity, lepton channel, and number of tags
- Jet energy scale and b-tag eff. affect the shapes of distributions
- Statistics dominated analysis: systematics contribution to the uncertainty is small

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

Relative systematic uncertainties

Component	Size
W+jets&QCD normalization	18 – 28%
top pair normalization	18%
Tag rate functions (shape)	2 – 16%
Jet energy scale (shape)	1 – 20%
Luminosity	6%
Trigger modeling	3 – 6%
Lepton ID	2 – 7%
Jet modeling	2 – 7%
Other small components	few%

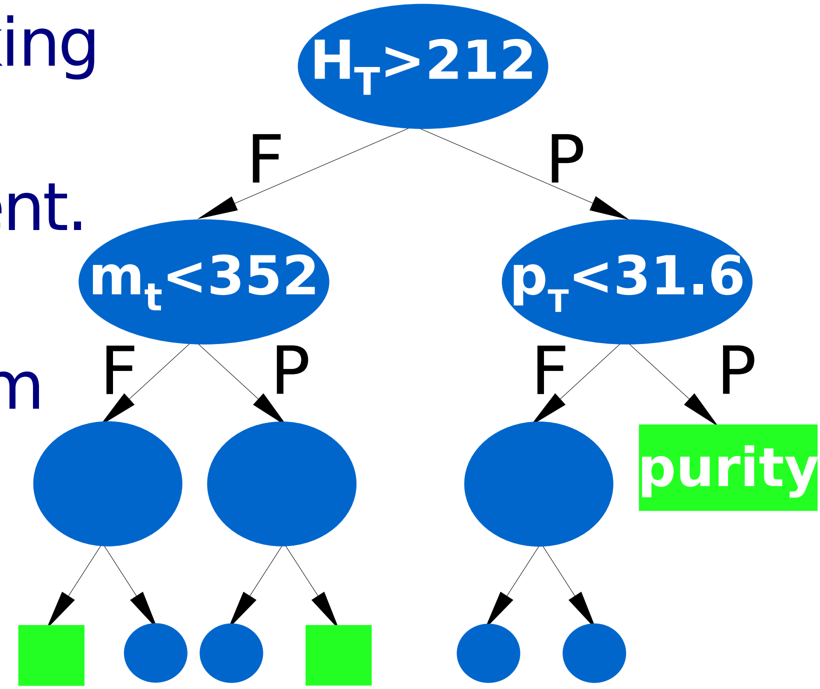
Check distributions

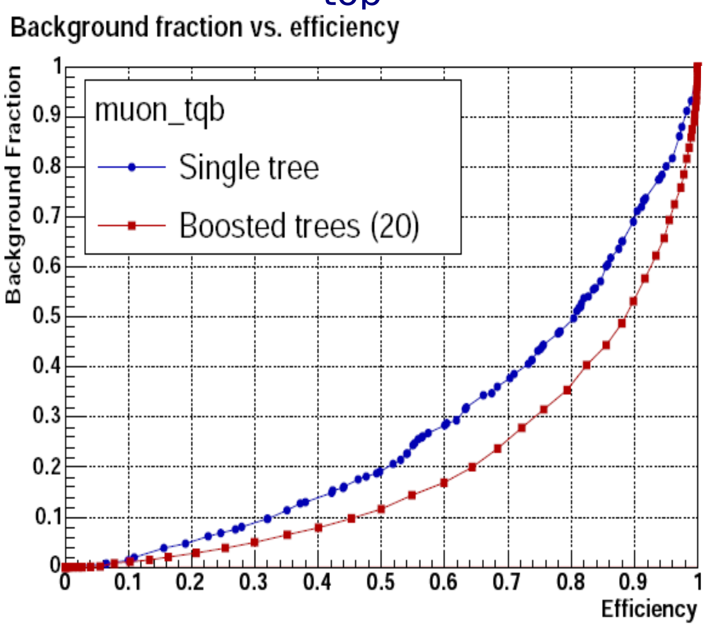


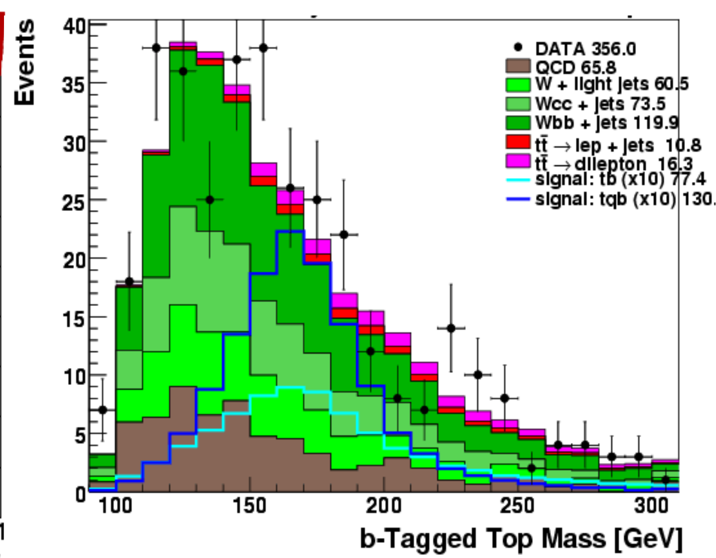
Boosted Decision Trees

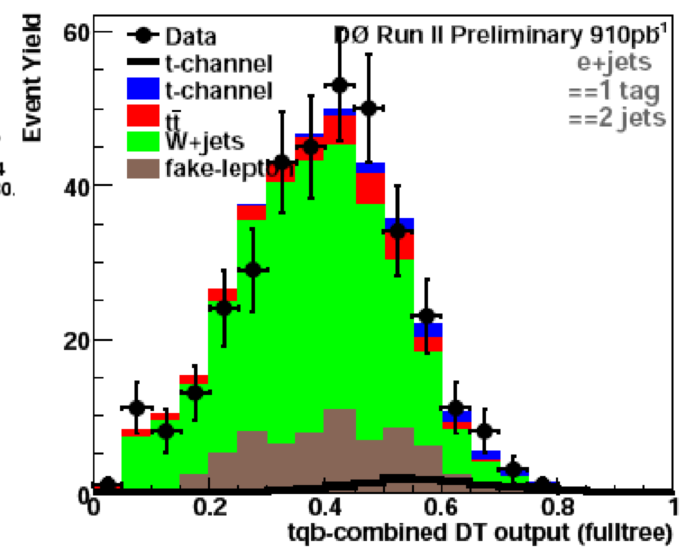
Idea: recover events that fail criteria in cut-based analysis

- Find best simple cut in each node looking at 49 physics motivated variables
- Output: purity $N_S/(N_S+N_B)$ for each event. Signal is tb+tqb.
- Boosting: retrain 20 times to learn from misclassified events
- Most discriminant: M(alljets), M(W,b_1), $\cos(b,\ell)_{top}$, $Q(\ell)\eta$(light-jet)









Matrix Elements method

- Use all available kinematic information from a **fully differential cross-section calculation** → See T. Gadfort talk in YSF session
- Calculate an event probability for signal and background hypothesis

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

Parton distribution functions CTEQ6

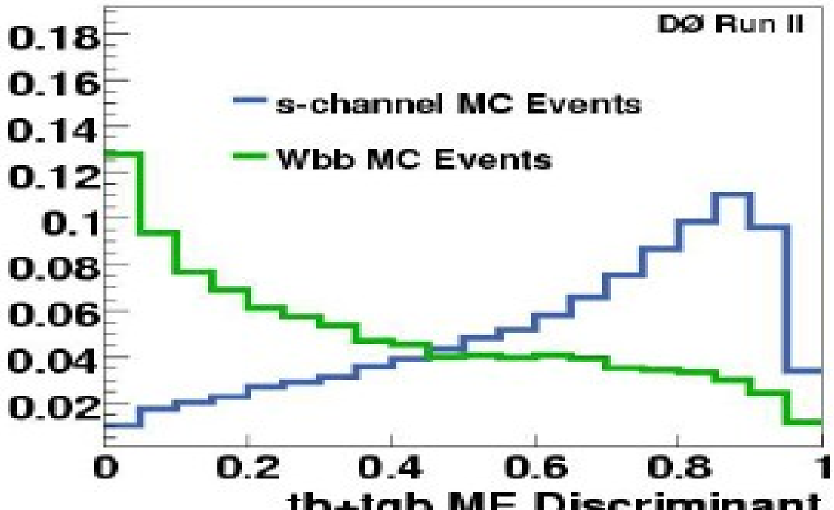
Differential cross section (LO ME from Madgraph)

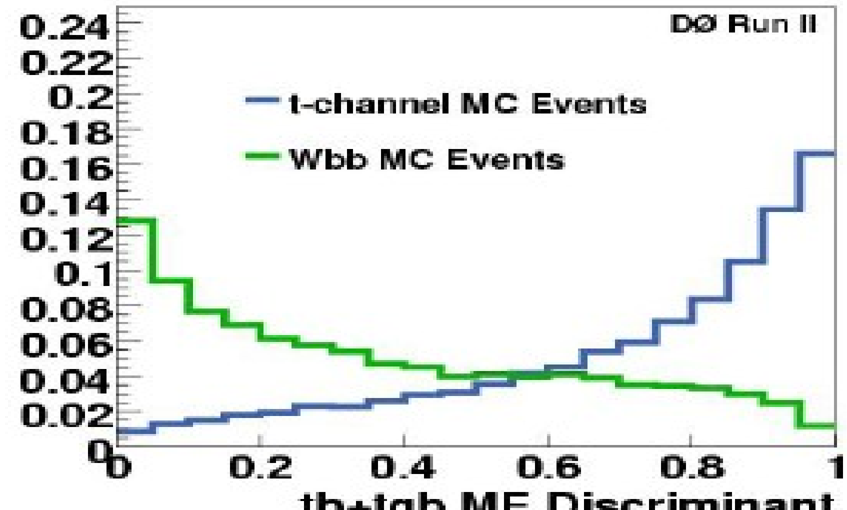
Transfer Function: maps parton level (y) to reconstructed variables (x)

- Integrate over 4 independent variables: assume angles well measured, known masses, momentum and energy conservation

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

- Analysis only uses 2&3 jet bins
- Wbg, Wcg, Wgg and Wbbg in $P_{Background}$





Bayesian Neural Networks

A different sort of NN (http://www.cs.toronto.edu/radford/fbm.software.html):

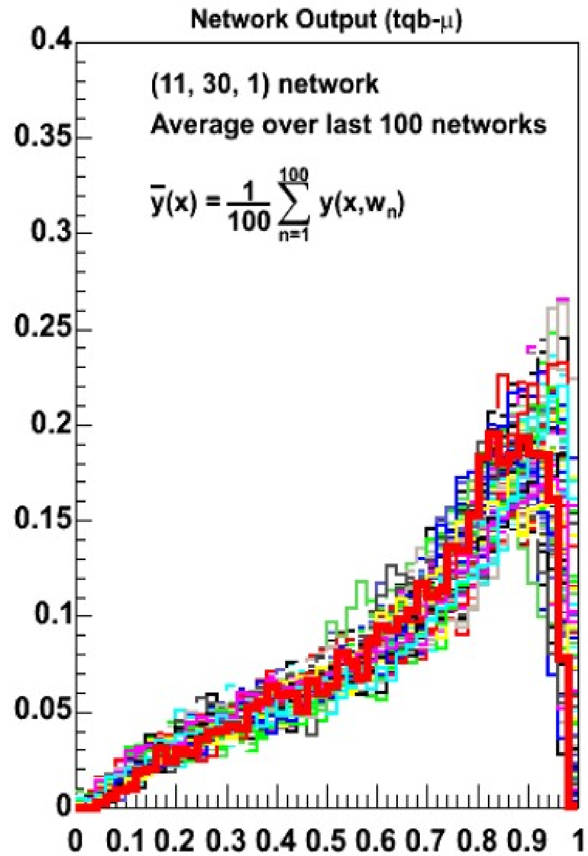
- Instead of choosing one set of weights, find posterior probability density over all possible weights
- Averages over many networks weighted by the probability of each network given the training data
- Use 24 variables (subset of the DT variables) and train against sum of backgrounds

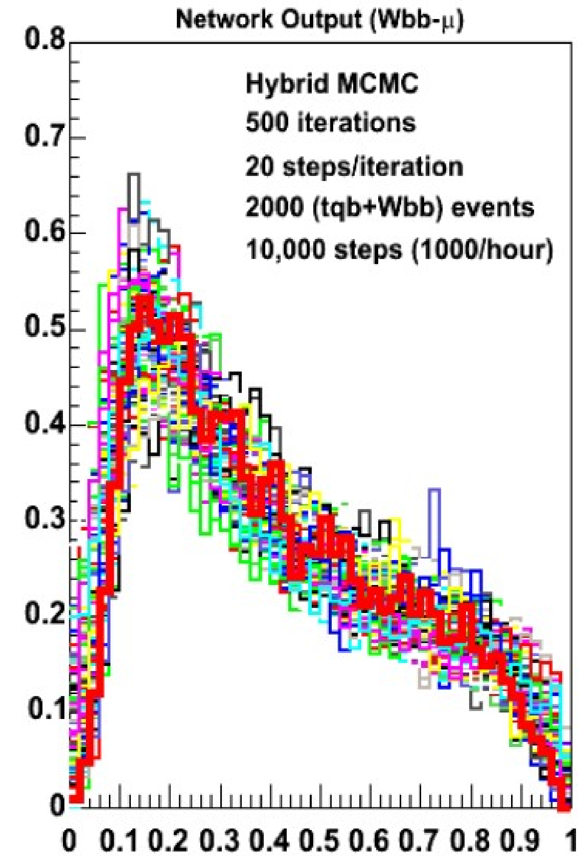
Advantages:

- Less prone to overfitting, because of Bayesian averaging
- Network structure less important: can use large networks!
- Optimized performance

Disadvantages:

- Computationally demanding!

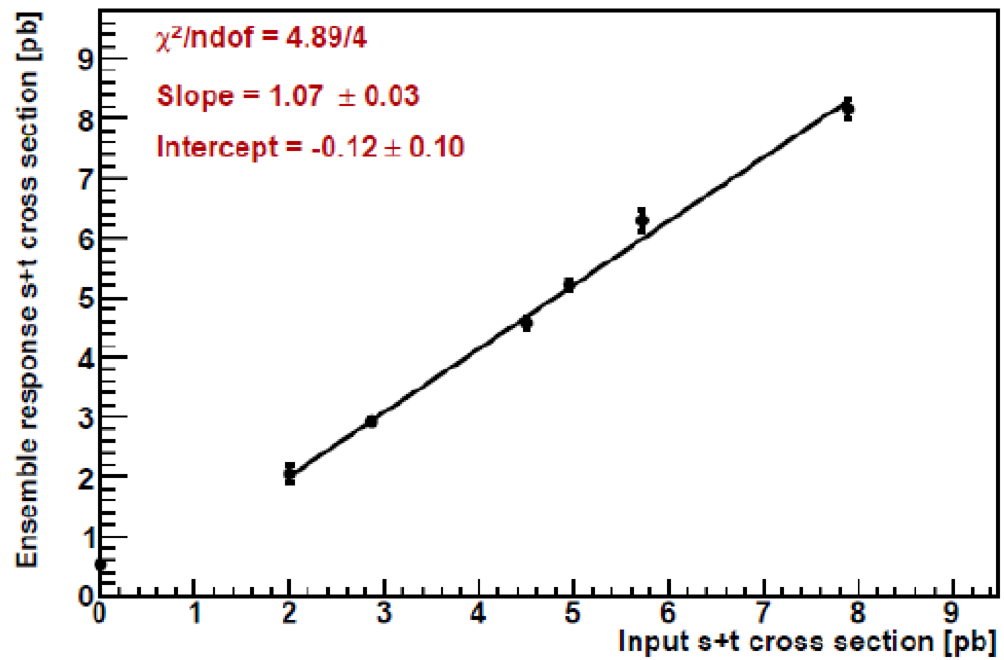




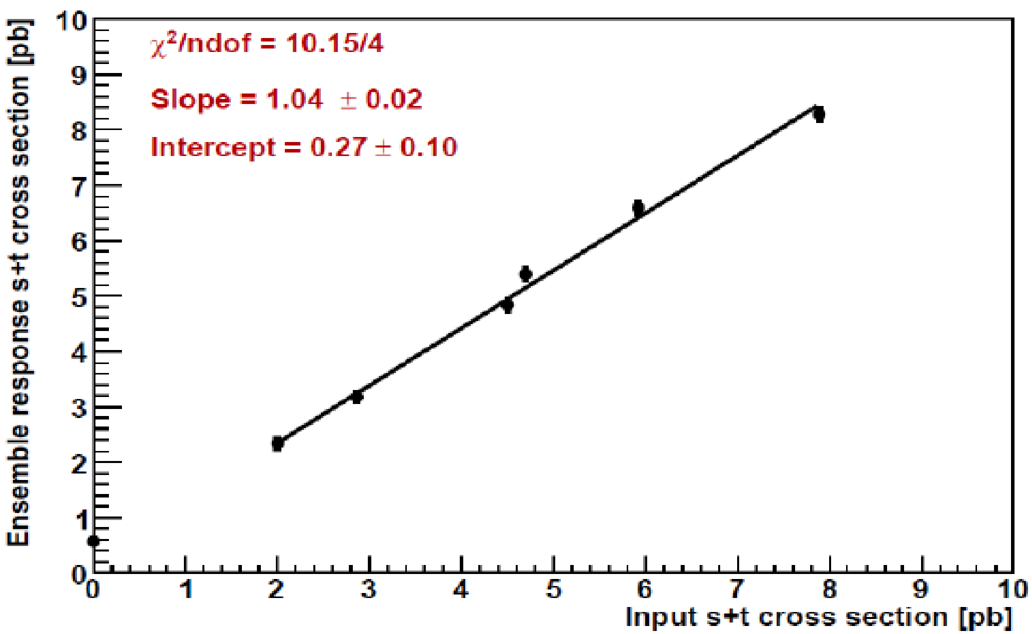
Linear response

- Use ensemble testing to show analysis calibration
- Use pool of MC events to draw events with bkgd. yields fluctuated according to uncertainties, repproducing the correlations between components introduced in the normalization to data
- Randomly sample a Poisson distribution to simulate statistical fluctuations
- Linear response, negligible bias

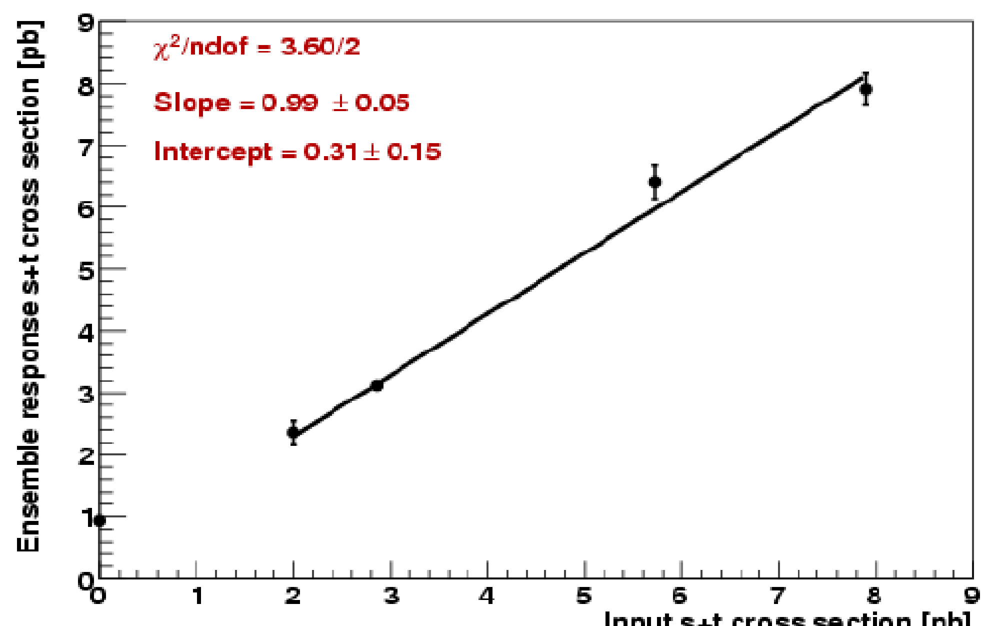
DT analysis



ME analysis

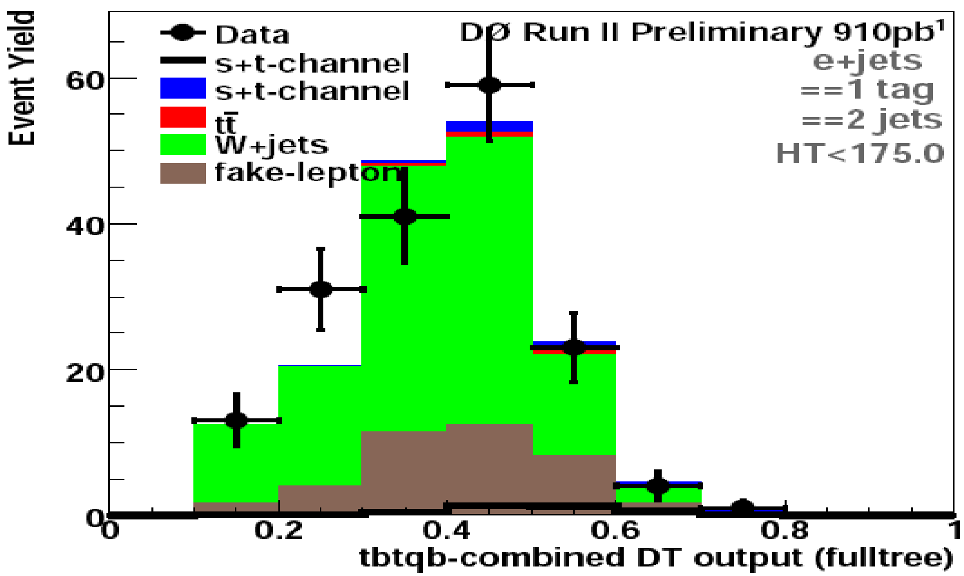


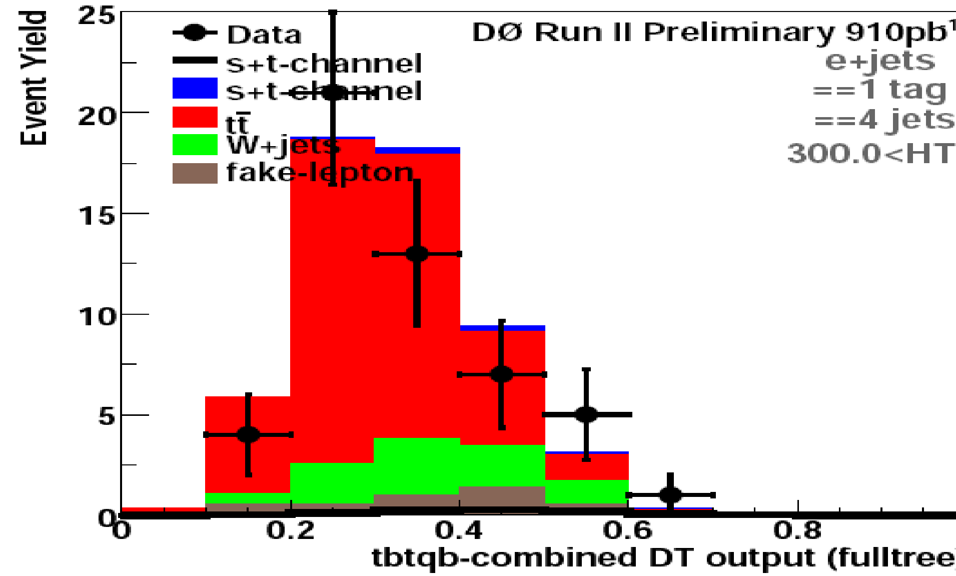
BNN analysis

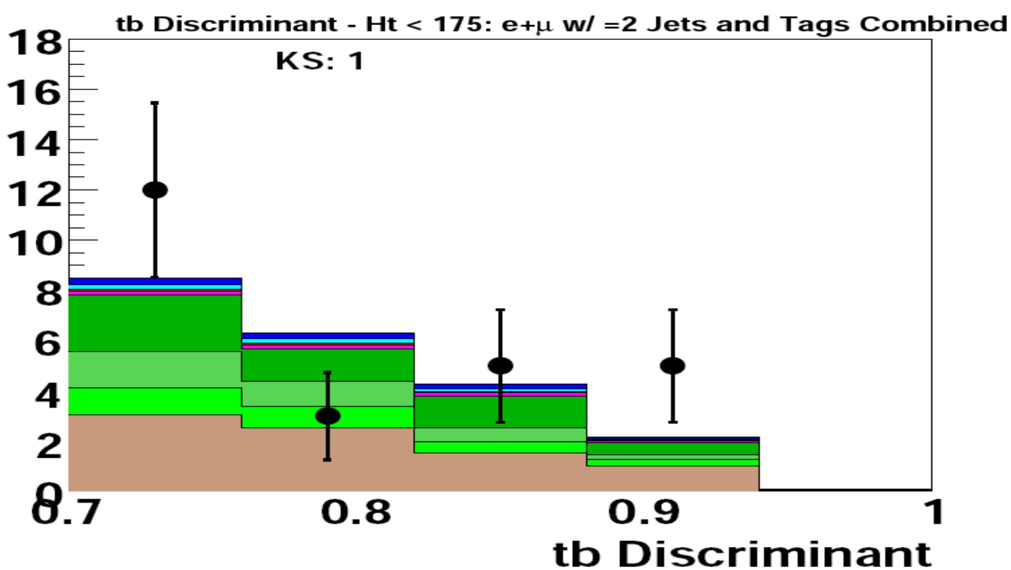


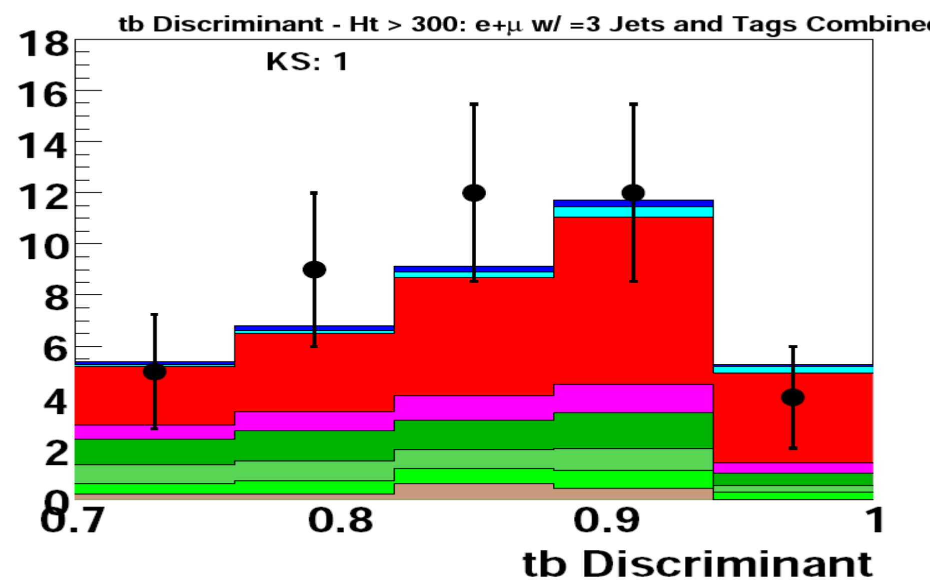
Cross check samples

- Check description of the two main backgrounds
  - "Soft" W+jets: 2 jets and H_T(lepton,MET,alljets) < 175 GeV
  - "Hard" W+jets: 3,4 jets and H_T(lepton,MET,alljets) > 300 GeV

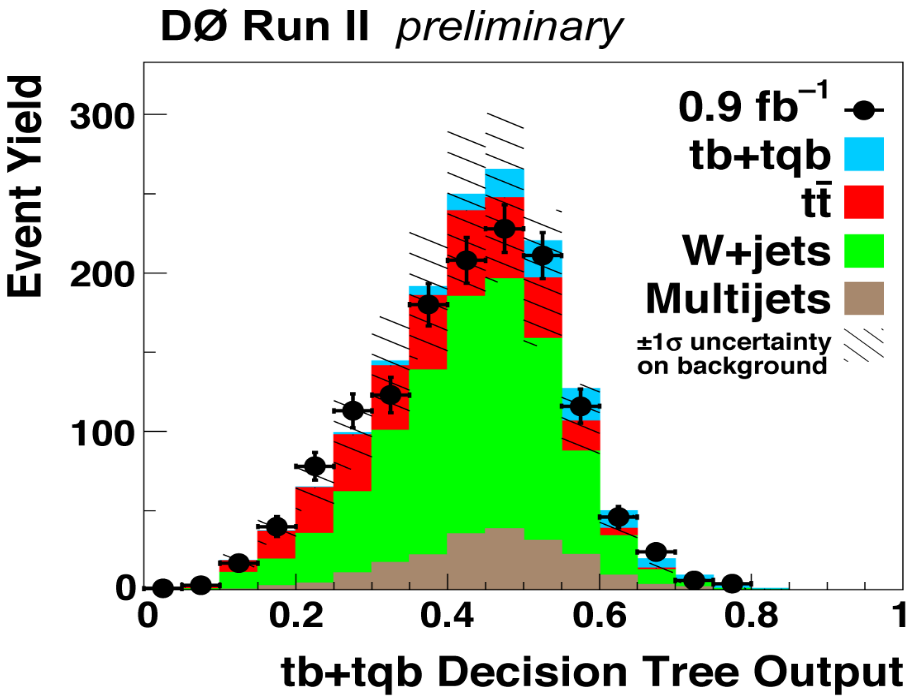


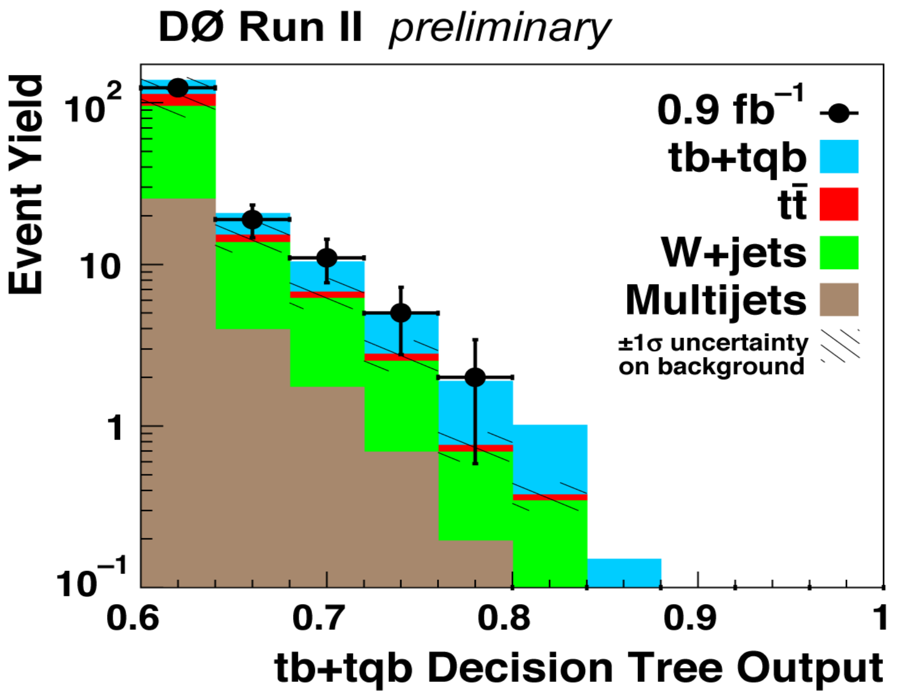


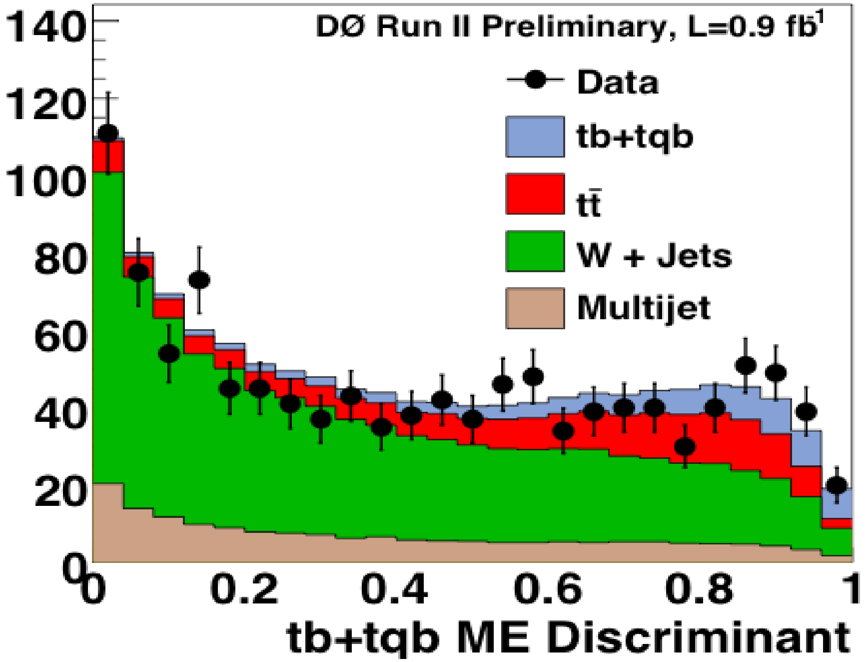


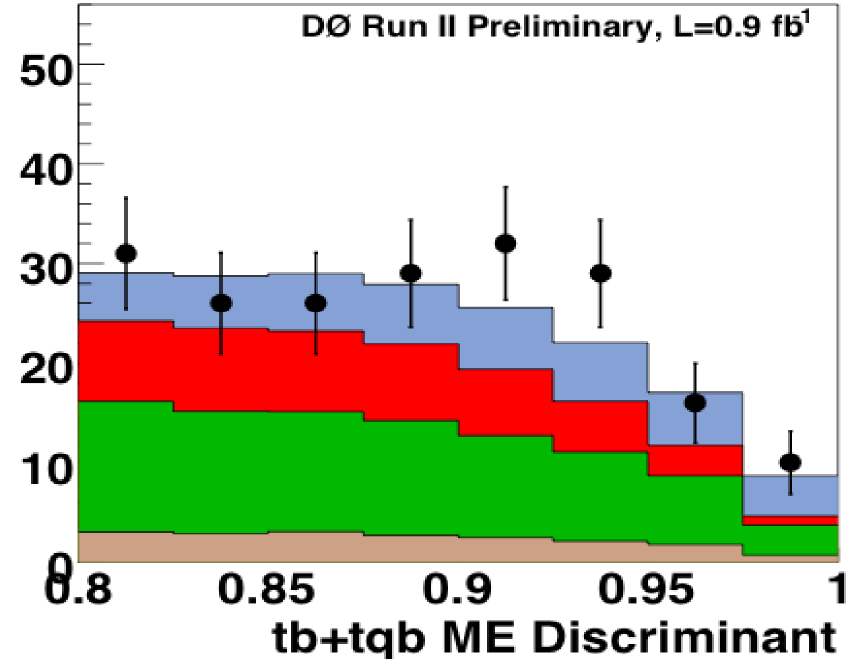


(summed) Discriminants output







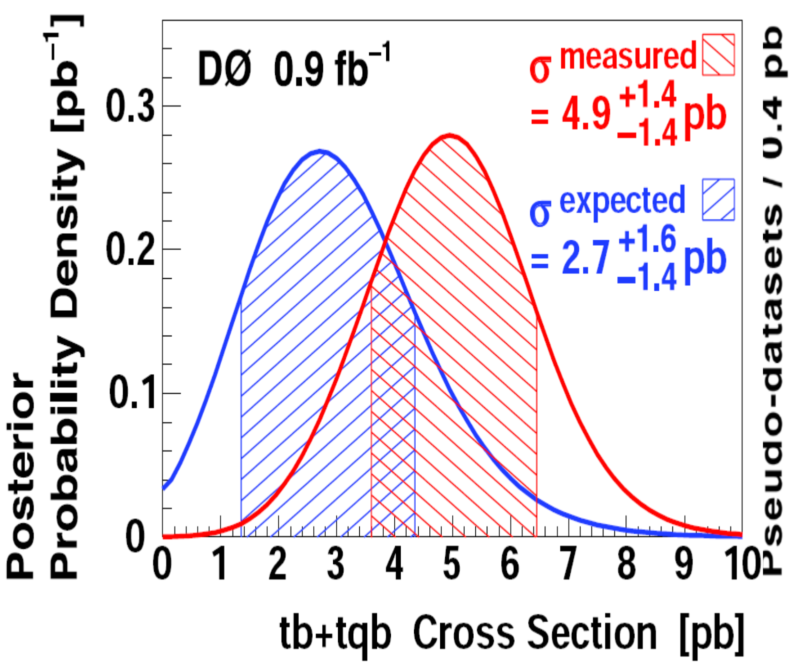


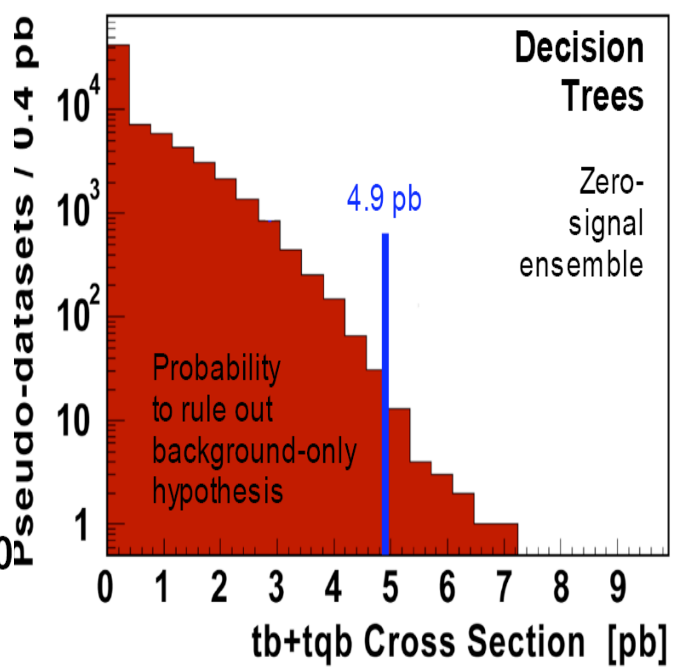
Expected and observed results

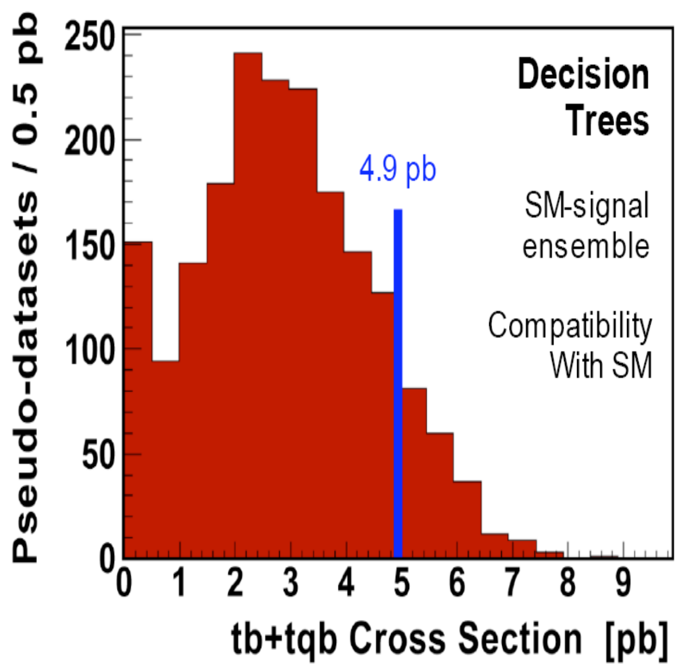
	Decision Trees		Matrix Elements		Bayesian NN	
	Expected	Observed	Expected	Observed	Expected	Observed
σ(tb+qb) [pb]	$2.7^{+1.6}_{-1.4}$	**4.9±1.4**	$3.0^{+1.8}_{-1.5}$	$\mathbf{4.6^{+1.8}_{-1.5}}$	$3.2^{+2.0}_{-1.8}$	**5.0±1.9**
significance	2.1σ	$\mathbf{3.4\sigma}$	1.8σ	$\mathbf{2.9\sigma}$	1.3σ	$\mathbf{2.4\sigma}$

DT measures 3.4σ excess! Evidence for single top production!

- Results are compatible with the SM at ~1 std. dev.







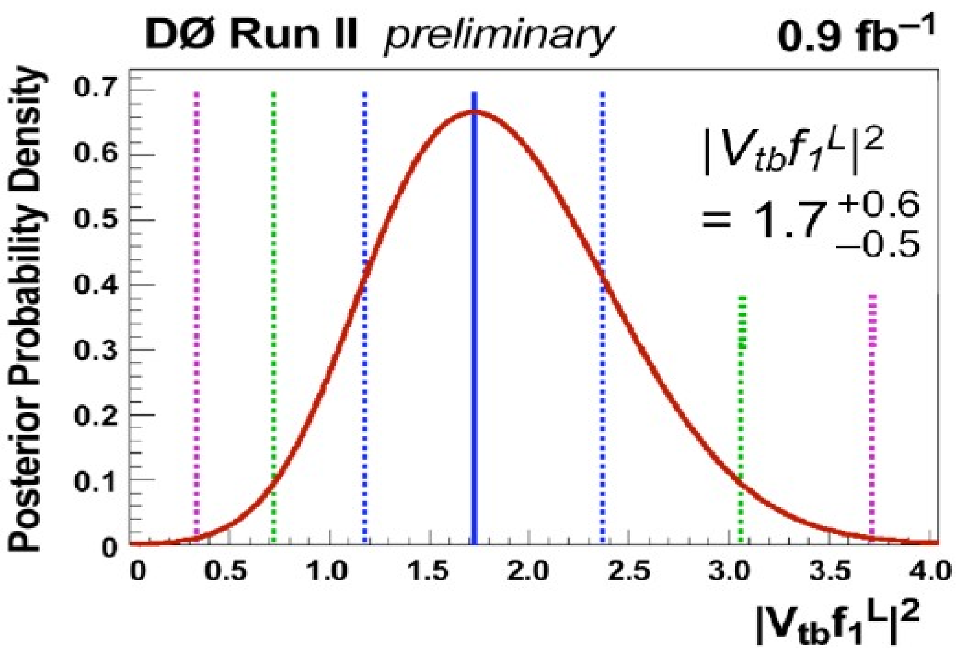
First direct measurement of $|V_{tb}|$

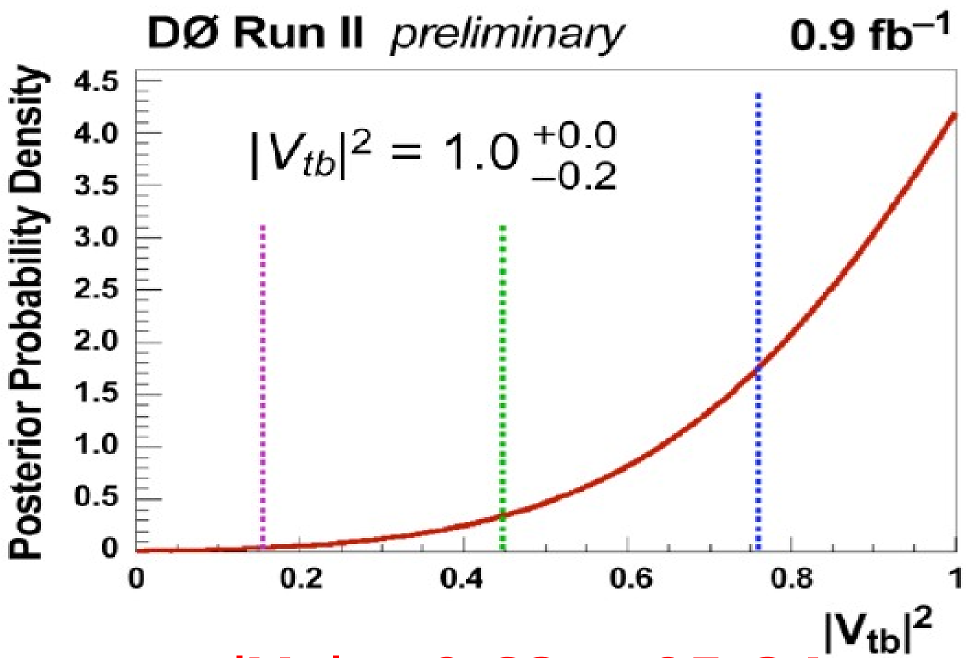
- Once we have a cross section measurement, we can make the first direct measurement of $|V_{tb}|$
- Calculate posterior in $|V_{tb}|^2$: $\sigma \propto |V_{tb}|^2$
- Assume:
  - **SM top decay**: $V_{td}^2 + V_{ts}^2 \ll V_{tb}^2$
  - Pure V-A and CP conserving interaction

Additional theoretical errors are needed

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

hep-ph/0408049





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

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

assuming $f_1^L = 1$ (SM)

This measurement does not assume 3 generations or unitarity

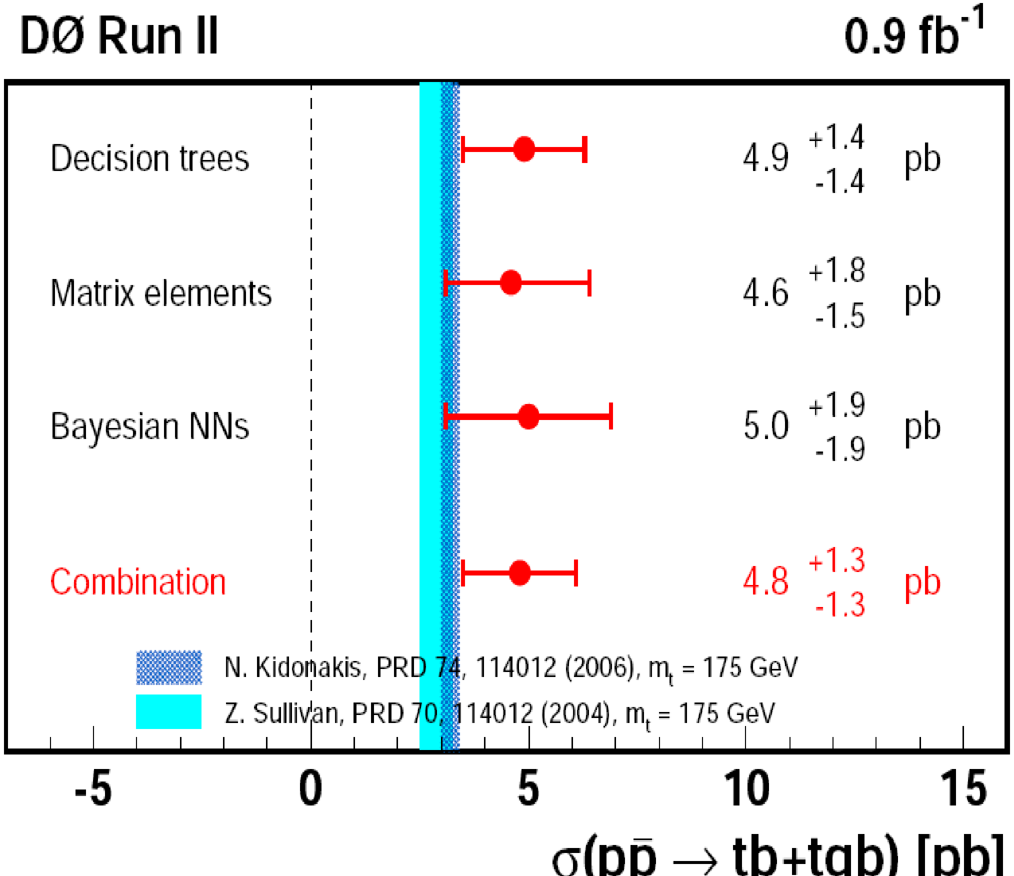
Combination of analyses

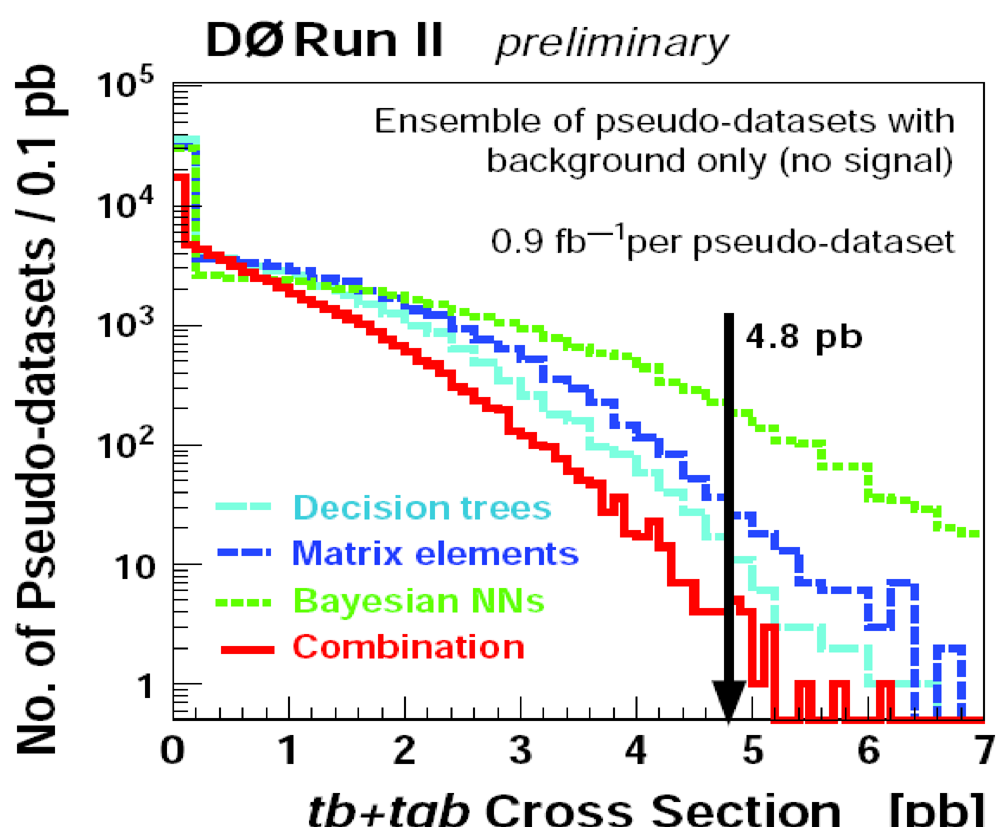
new!

- Combine the three measurements with BLUE method
- Method requires to measure the correlations
- Used SM pseudo-datasets with systematics

$$\rho = \begin{array}{cccc} DT & ME & BNN & \\ \end{array}$$
$$\rho = \left(\begin{array}{cccc} 1 & 0.57 & 0.51 & DT \\ 0.57 & 1 & 0.45 & ME \\ 0.51 & 0.45 & 1 & BNN \end{array} \right)$$

Combined result: 4.8 ± 1.3 pb ➔ Significance of 3.5 std. dev.





Conclusions

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

(hep-ex/0612052 submitted to PRL)

$$\sigma(s+t) = 4.8 \pm 1.3 \text{ pb}$$

$$3.5\sigma \text{ significance!}$$

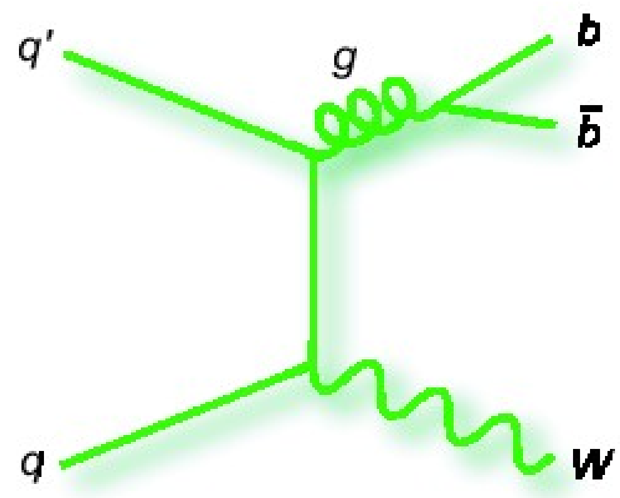
$$|V_{tb}| > 0.68 \text{ @ } 95\% \text{C.L.}$$

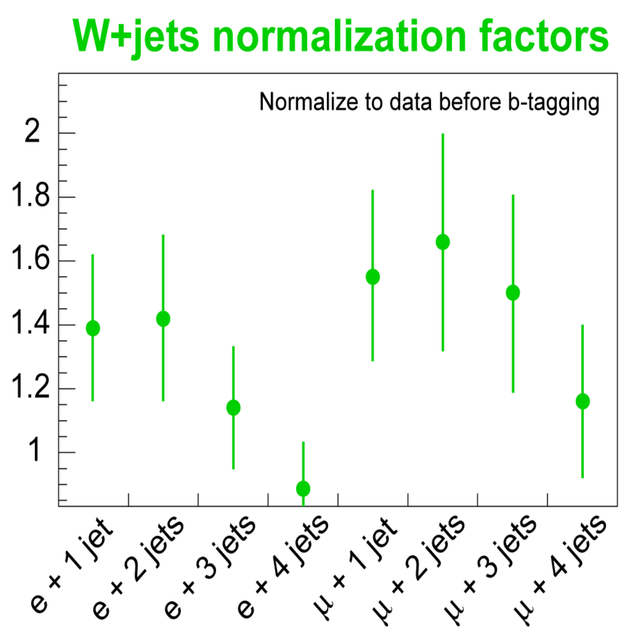
- Challenging analysis: small signal hidden in huge complex background
- Expand to searches of new phenomena
- We now have double the data to analyze!

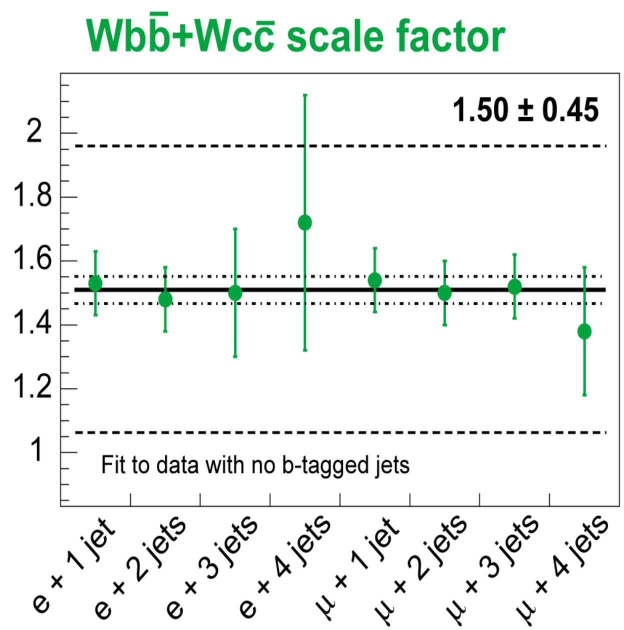
Extra Slides

Signal and Background modeling

- Signal is modeled with CompHEP (effective NLO) + Pythia
- W+jets and ttbar shapes from Alpgen with MLM matching + Pythia
- ttbar normalized to NNLO σ = 6.8±1.2 pb
- QCD from our selected data with non-isolated lepton
- Normalize W+jets and QCD to data before tagging (SF ~ 1.4)
- Determine Wbb and Wcc fractions in W+jets from zero-tagged data
  - Wbb+Wcc factor 1.50±0.45 makes all distributions match data







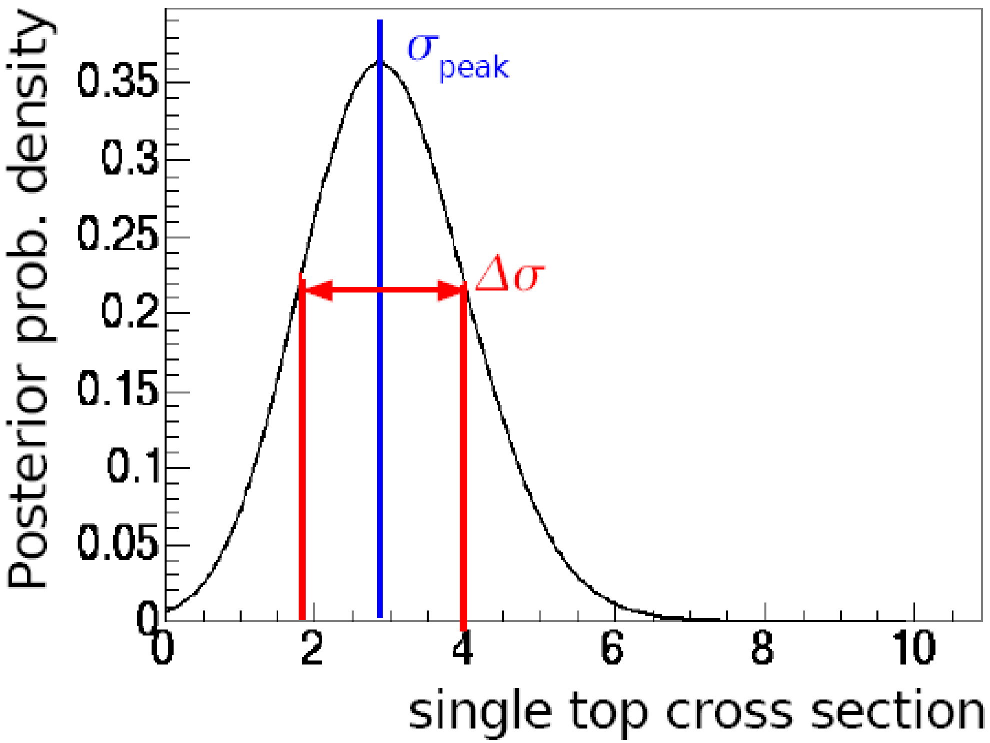
Measuring the cross section

- We form a binned likelihood from the discriminant outputs

- Probability to observe data distribution D, expecting y:

$$y = \underbrace{\alpha \mathscr{L} \sigma}_{\text{signal}} + \underbrace{\sum_{s=1}^{N} b_s}_{\text{bkgd.}} = a\sigma + \sum_{s=1}^{N} b_s$$

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



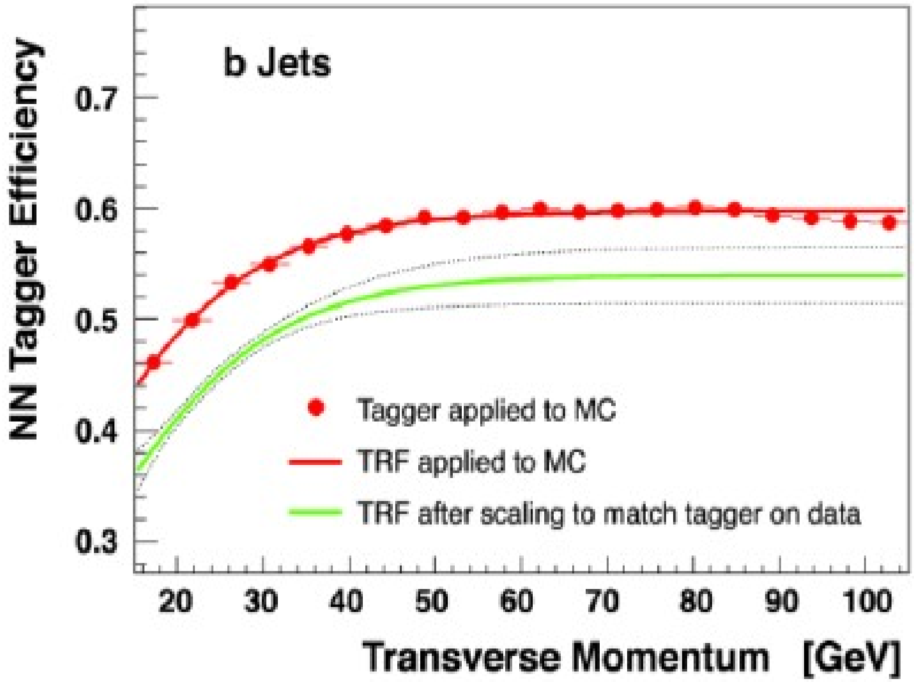
- And obtain a Bayesian posterior probability density as a function of the cross section:

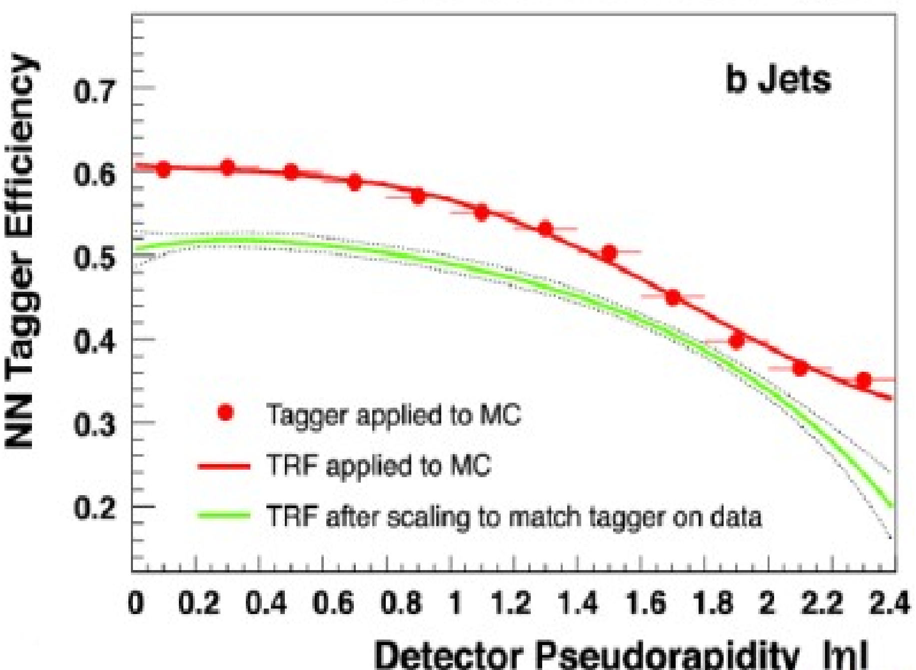
$$Post(\sigma|D) \equiv P(\sigma|D) \propto \int_a \int_b P(D|\sigma, a, b) Prior(\sigma) Prior(a, b)$$

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

NN b-jet tagger

- NN trained on 7 input variables from SVT, JLIP and CSIP taggers
- Much improved performance!
  - Fake rate reduced by 1/3 for same b-efficiency relative to previous tagger
  - Smaller systematic uncertainty
- Tag Rate Functions (TRFs) in η, p_T and z-PV derived in data are applied to MC
- Our operating point:
  - b-jet efficiency: ~50%
  - c-jet efficiency: ~10%
  - Light-jet efficiency: ~0.5%





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$_{\mathrm{besttop}}$,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)

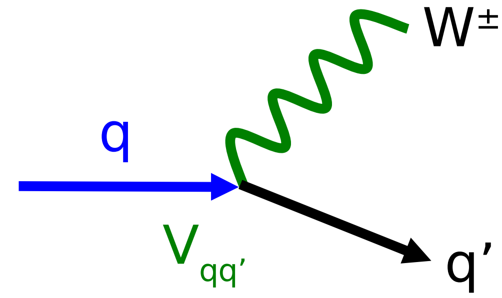
Most discrimination:

M(alljets)
M(W,tag1)
cos(tag1,lepton)$_{\mathrm{btaggedtop}}$
Q(lepton) x η(untag1)

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

$|V_{tb}|$

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}$$



- Most general Wtb vertex:

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

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

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