SUNY Stony Brook — HEP Seminar

Stony Brook, January 28, 2007

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

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

Arán García-Bellido

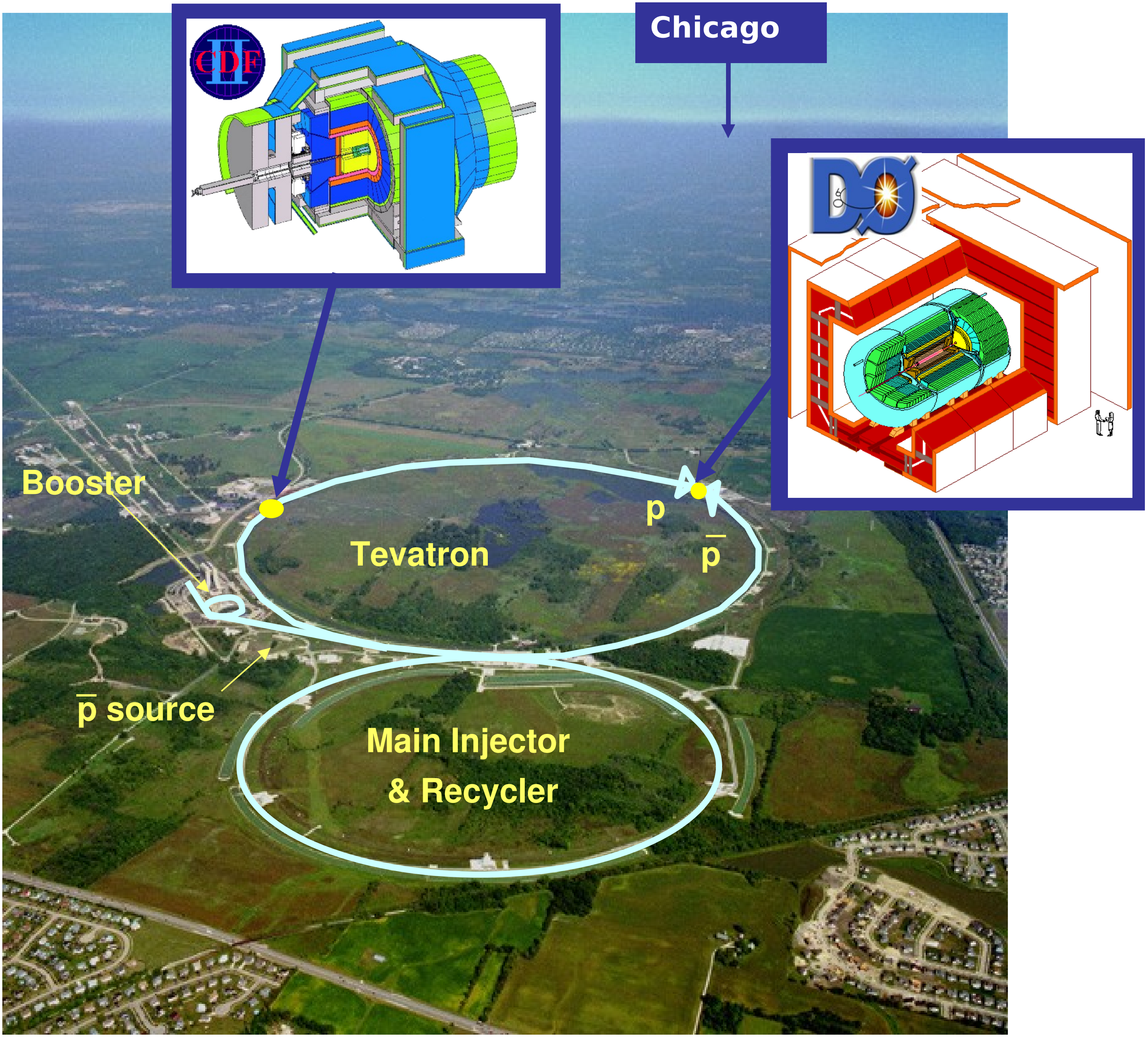
The Tevatron

The highest energy particle accelerator in the world!

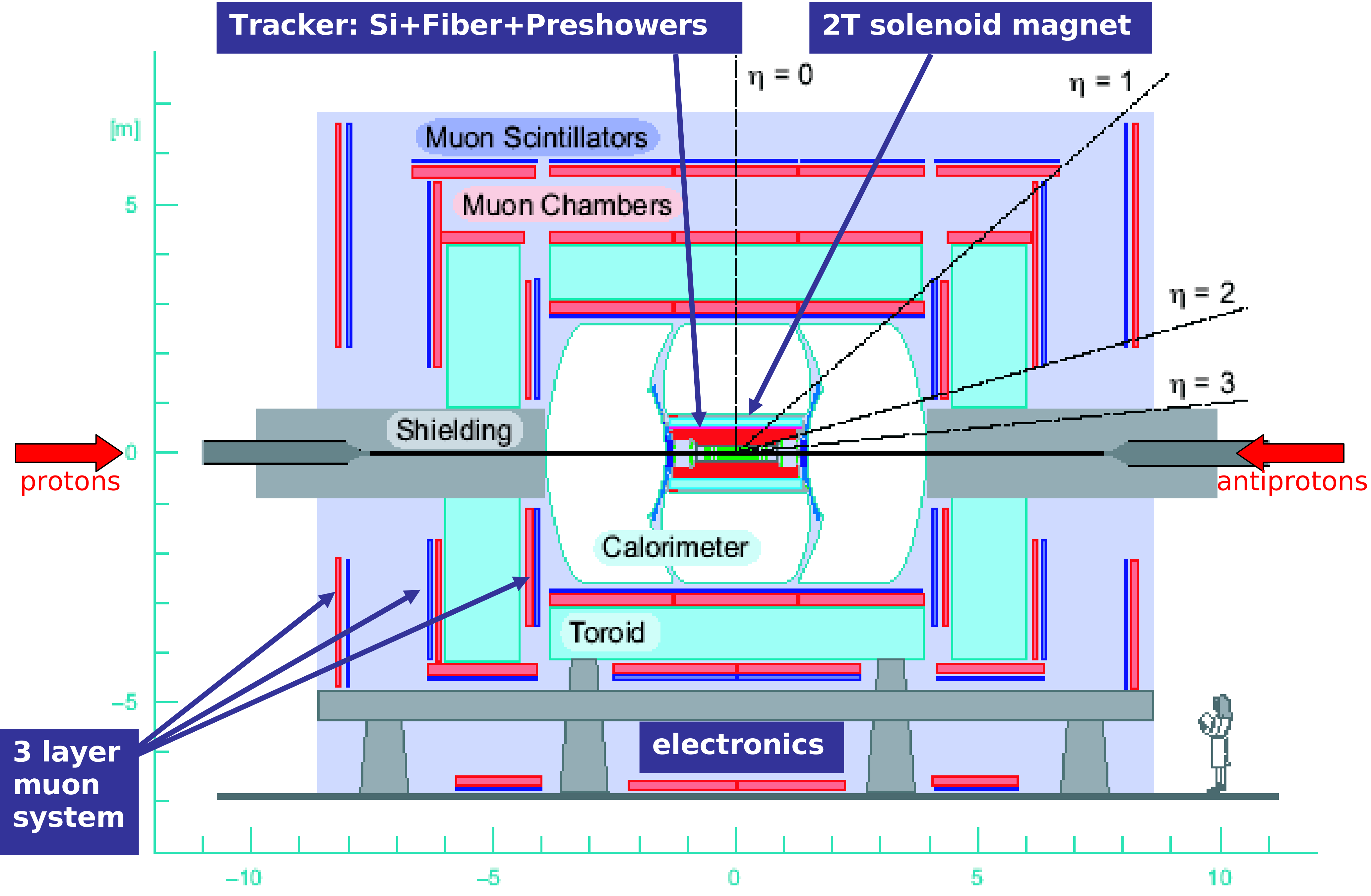
Proton-antiproton collider

Run I 1992-1995
Top quark discovered!

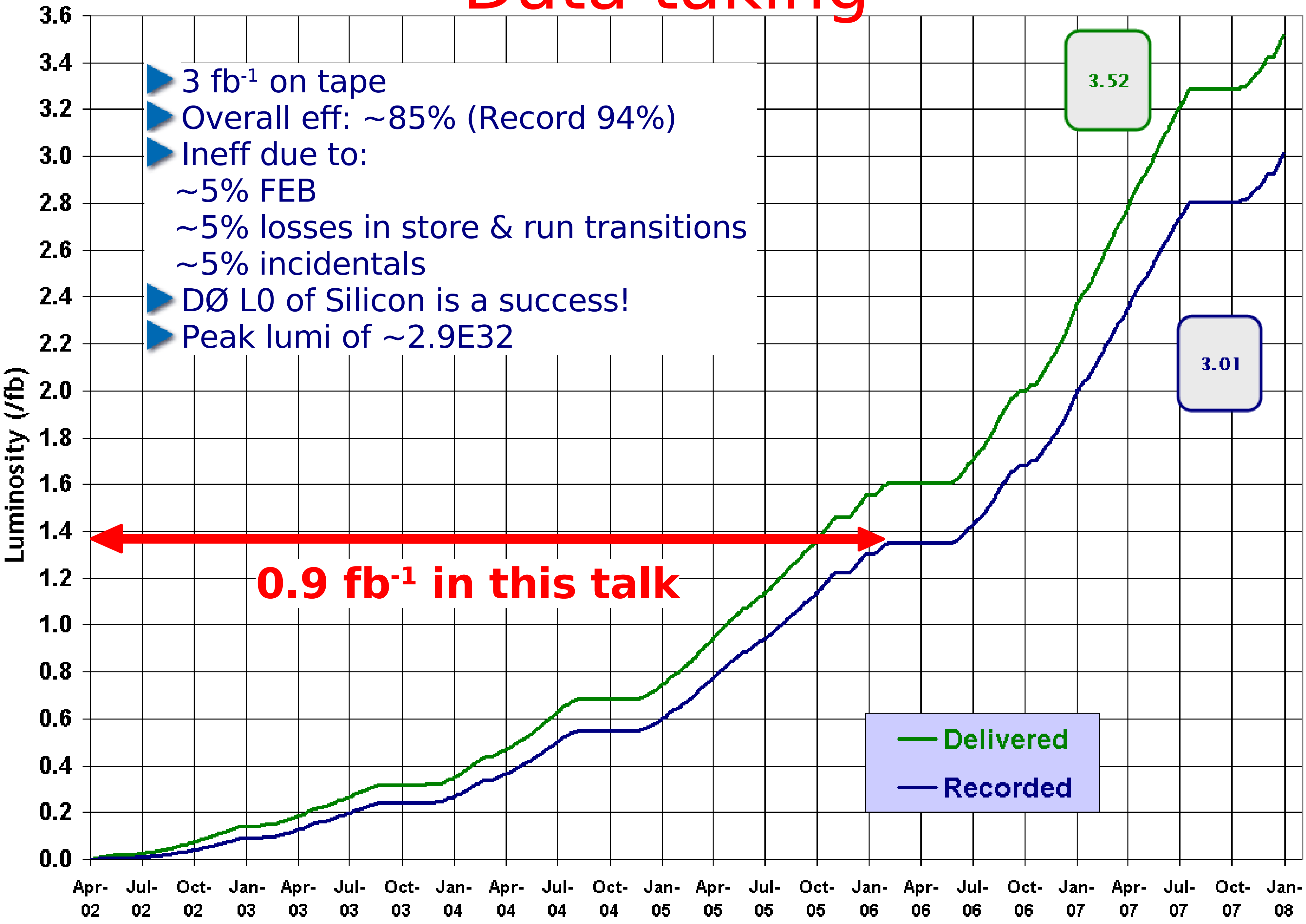
Run II 2001-09
$\sqrt{s} = 1.96$ TeV
$\Delta t = 396$ns
$>3.5\mathrm{fb}^{-1}$ delivered
Peak Lum: $3 \cdot 10^{32}\mathrm{cm}^{-2}\mathrm{s}^{-1}$



DØ for Run II


Tracker: Si+Fiber+Preshowers
2T solenoid magnet
η = 0
η = 1
η = 2
η = 3
[m]
5
0
−5
Muon Scintillators
Muon Chambers
Shielding
Calorimeter
Toroid
electronics
protons
antiprotons
3 layer muon system
−10
−5
0
5
10

Data taking



- 3 fb^{-1} on tape
- Overall eff: ~85% (Record 94%)
- Ineff due to:
  ~5% FEB
  ~5% losses in store & run transitions
  ~5% incidentals
- DØ L0 of Silicon is a success!
- Peak lumi of ~2.9E32

0.9 fb^{-1} in this talk

Top quark physics

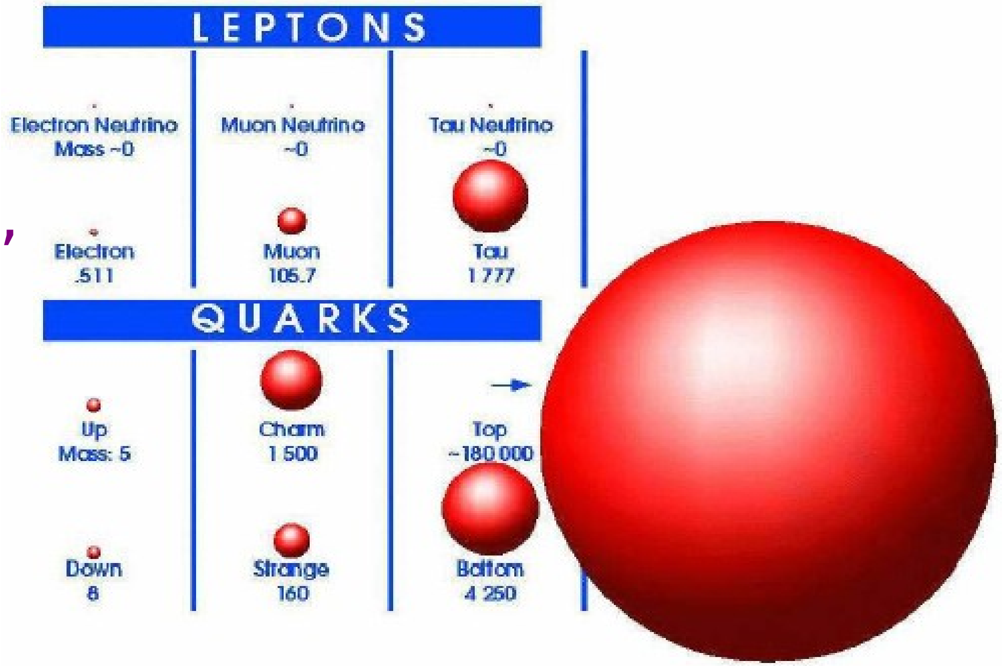
The top quark is a very special fermion:

- Heaviest known particle: 170.9±1.8 GeV
  - $m_t \sim v/\sqrt{2}$, $\lambda_t \sim 1$ → Related to EWSB!
  - Sensitive probe for new physics, FCNCs, ...
- Decays as a free quark: $\tau_t = 5\times10^{-25}$ s $\ll \Lambda_{QCD}^{-1}$
  - Spin information is passed to its decay products
  - Test V-A structure of the SM

We know the mass, cross section, charge and its BR(t→Wb)~1

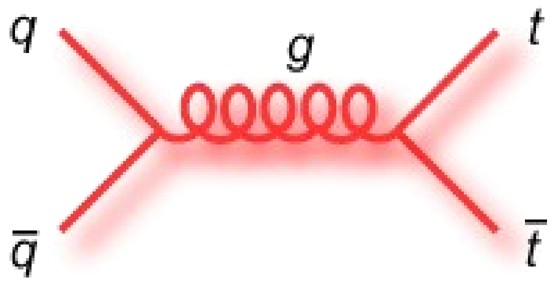
We still don't know: spin, width, lifetime

Plenty of room for new physics



Top quark electroweak production

Dominant top production mode at the Tevatron is in pairs:

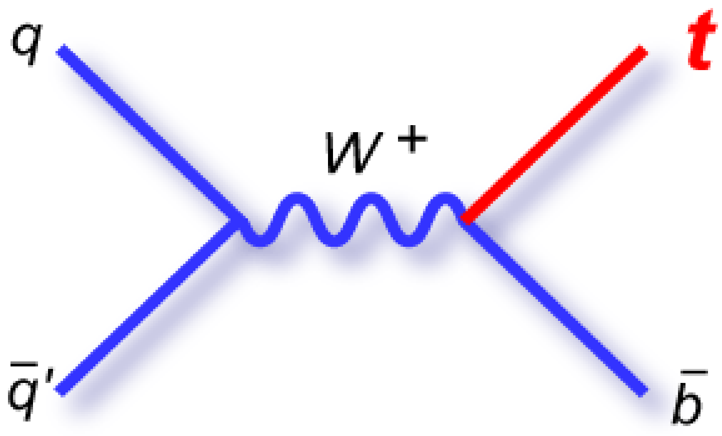


$\sigma_{\mathrm{NNLO}}=6.8\pm1.2$ pb

PRD 68, 114014

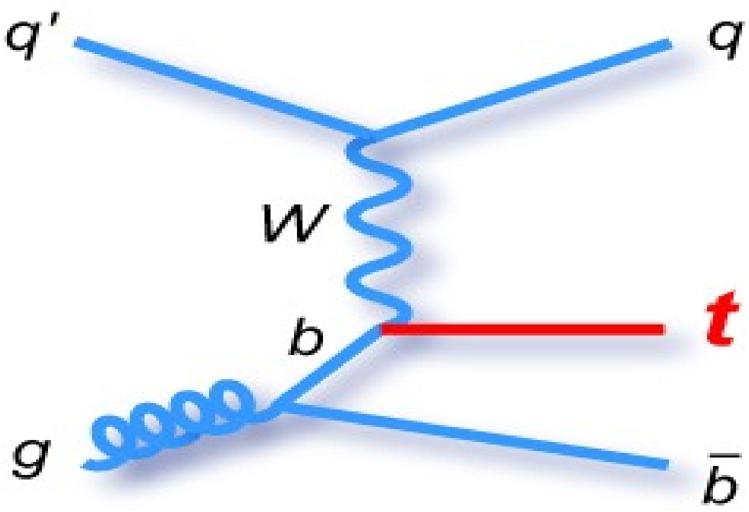
But top quarks can be produced singly via the EW interaction:

PRD 66 (02) 054024
hep-ph/0408049



s-channel (tb)

$\boldsymbol{\sigma_{\mathrm{NLO}}=0.88\pm0.11}$ **pb**

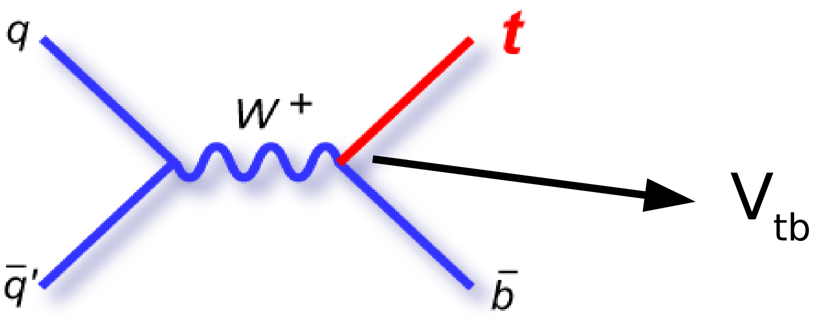


t-channel (tqb)

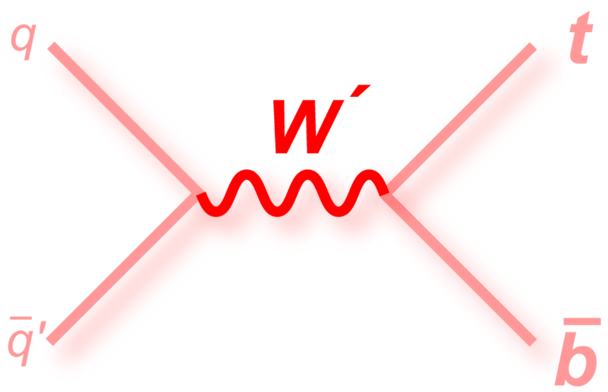
$\boldsymbol{\sigma_{\mathrm{NLO}}=1.98\pm0.25}$ **pb**

Why search for single top?

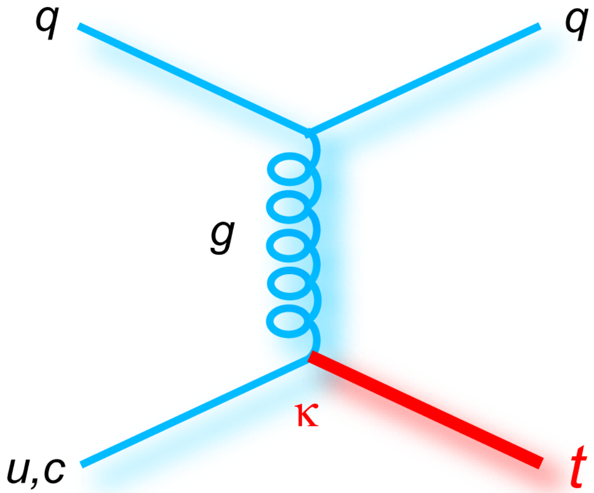
- Access W-t-b coupling
  - measure V_{tb} directly
  - test unitarity of CKM



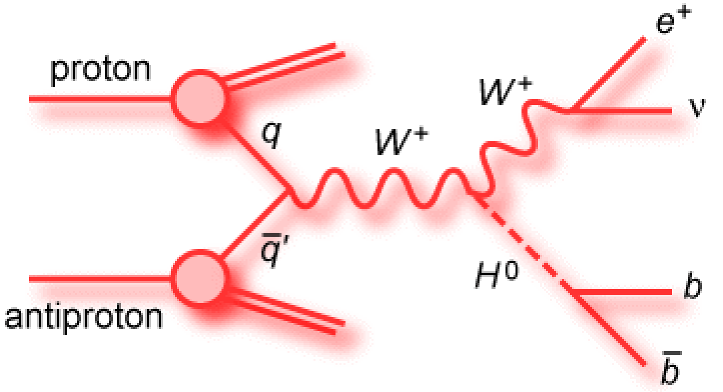
- New physics:
  - s-channel sensitive to resonances: W’, top pions, SUSY, etc…
  - t-channel sensitive to FCNCs, anomalous couplings
- Source of polarized top quarks
- Extract small signal out of a large background



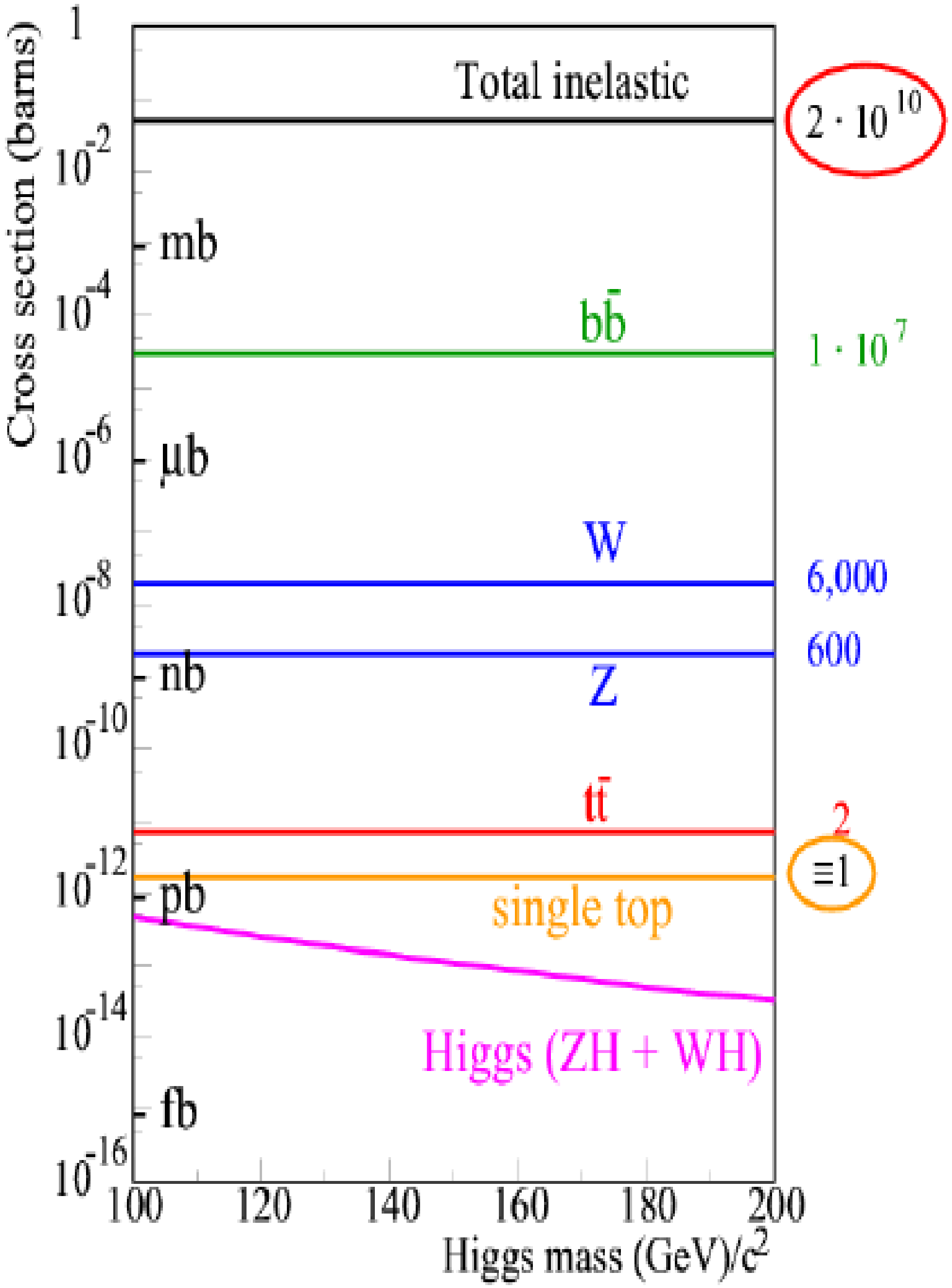
DØ search: hep-ex/0607102



DØ search: hep-ex/0702005



A big challenge!



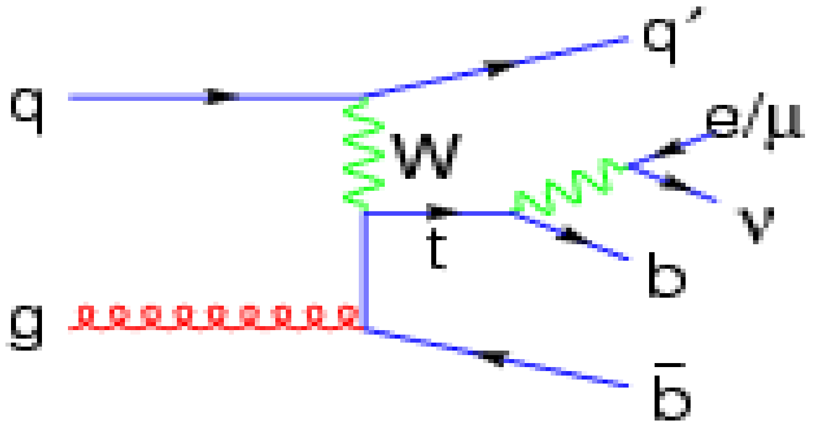
~20 single top events produced per day

But huge backgrounds!

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

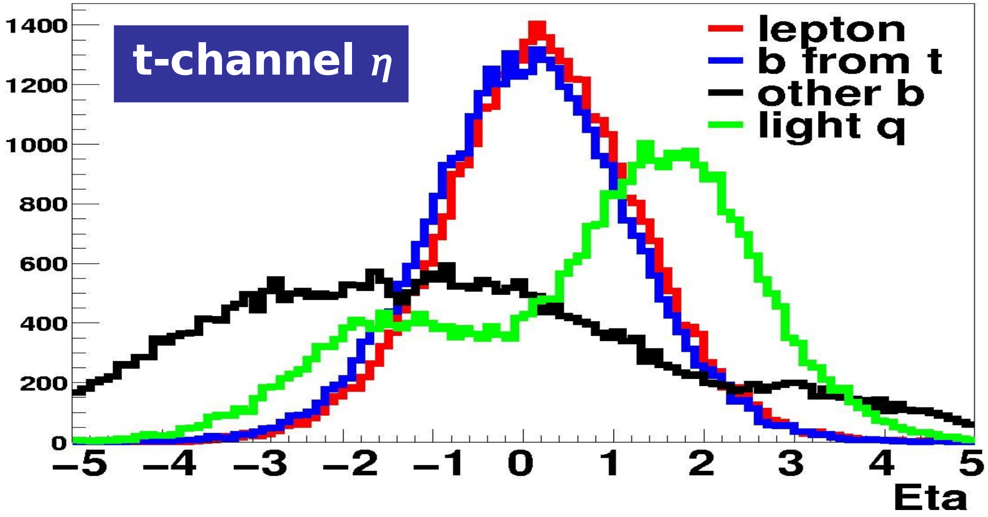
- Background model improvements (PS↔ME matching: MLM)
- Fully reprocessed dataset: new calibrations, jet thresholds, JES,...
- New more efficient NN b-tagger
- Split channels by jet multiplicity
- Combined s+t search added (SM s:t ratio is assumed)

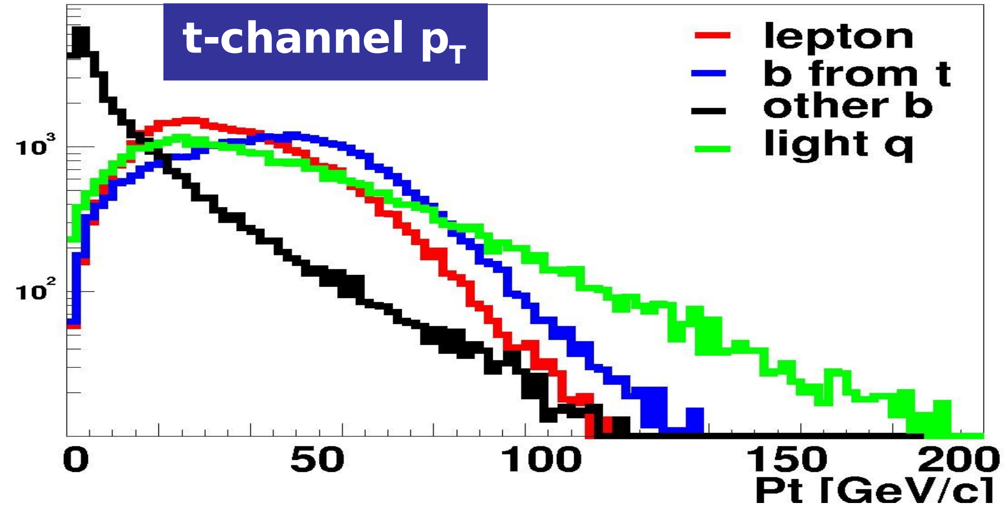
Signal selection



Signature:

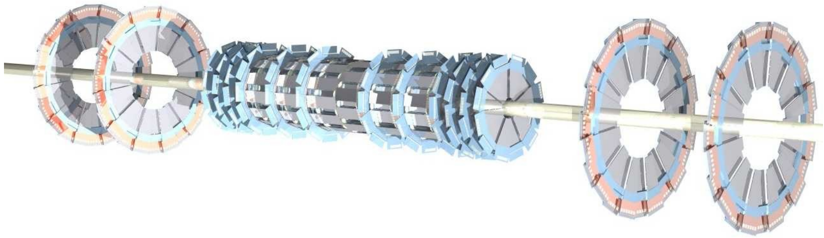
- One high p_T isolated lepton (from W)
- MET (ν from W)
- One b-quark jet (from top)
- A light flavor jet and/or another b-jet



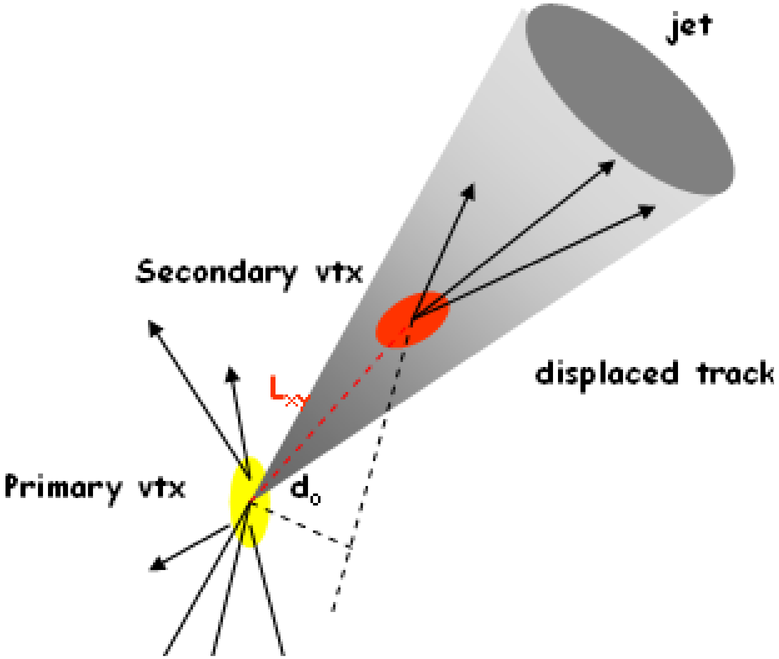


Event selection:

- Only one tight (no loose) lepton:
  - e: p_T >15 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



Tagging b-jets



Three different algorithms for b-jet identification at DØ:

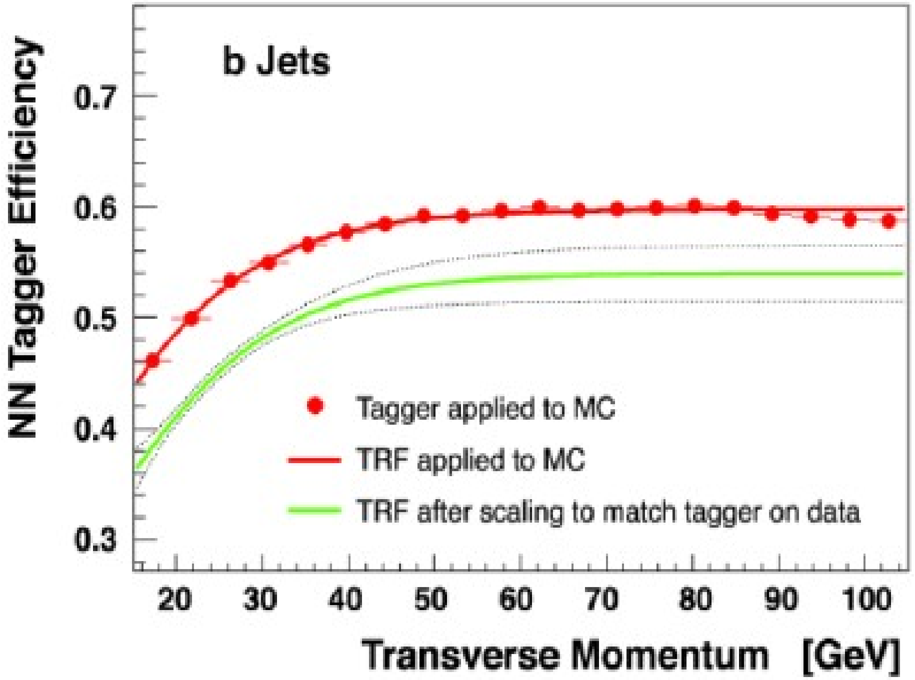
- Two based on tracks with large IP (JLIP, CSIP)
- One based on secondary vertex reconstruction (SVT)
- Combine in NN NEW

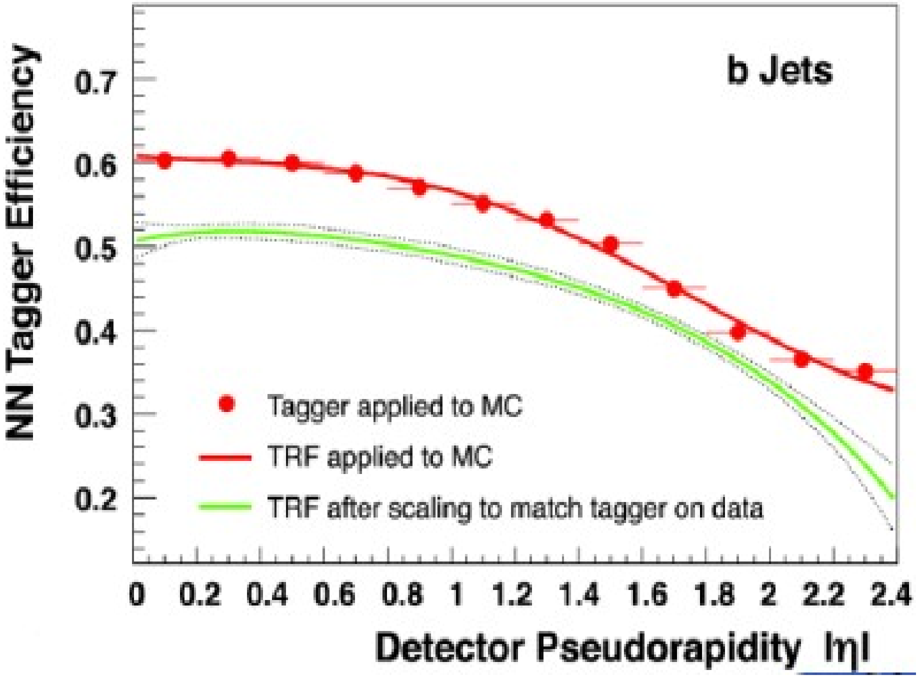
- Neural Net trained on seven variables:

1. Decay length significance SVT
2. Weighted combination of track's IPs
3. JLIP probability
4. χ/dof of the SVT vertex
5. Number of tracks in SVT vertex
6. Mass of the SVT vertex
7. Number of secondary vertices found inside jet

NN b-jet tagger

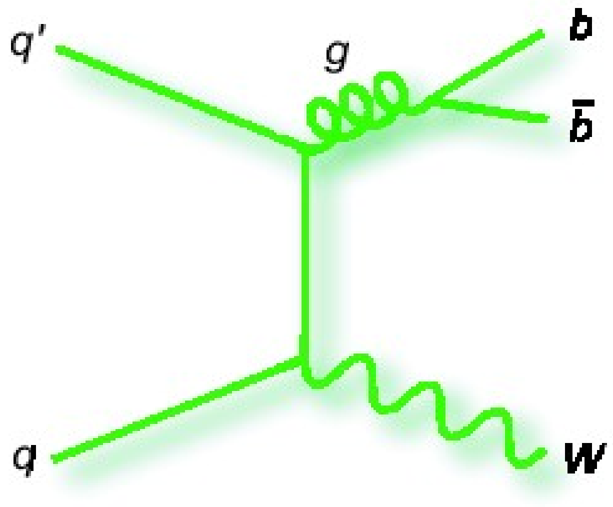
- Much improved performance!
  - Fake rate reduced by 1/3 for same b-efficiency relative to previous tagger
  - Smaller systematic uncertainty
- Tag Rate Functions (TRFs) in η, p_T and z-PV derived in data are applied to MC
- Our operating point:
  - b-jet efficiency: ~50%
  - c-jet efficiency: ~10%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



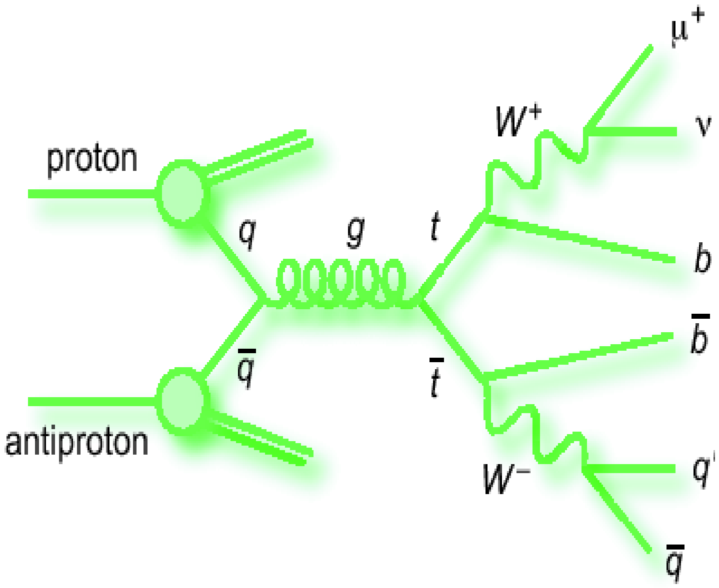


Background modeling

- W+jets: ~o(1000) pb
  - Distributions from Alpgen 2.0
  - Normalization from data
  - Heavy flavor fractions from data



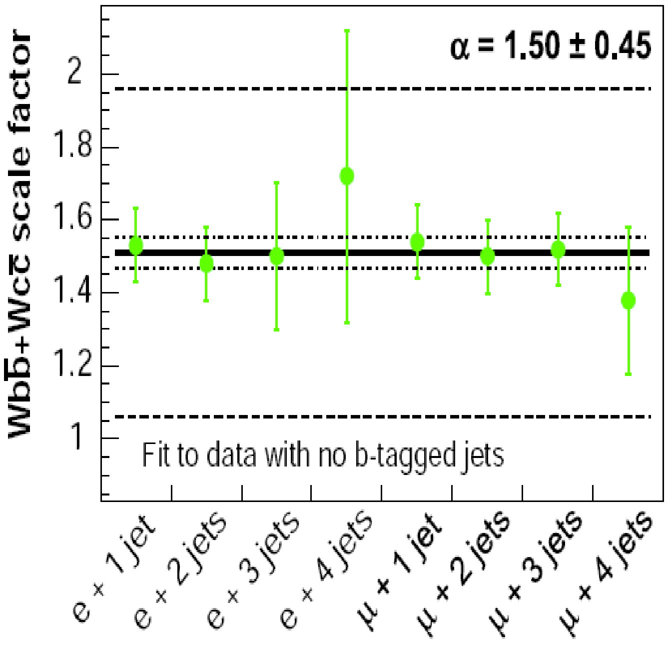
- Top pairs: ~7 pb
  - Topologies: dilepton and ℓ+jets
  - Use Alpgen 2.0 with MLM matching
  - Normalize to NNLO σ

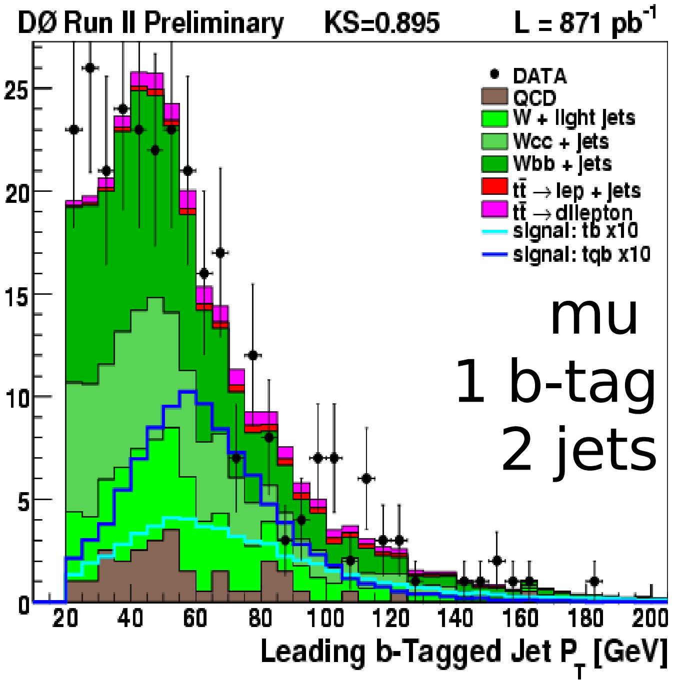


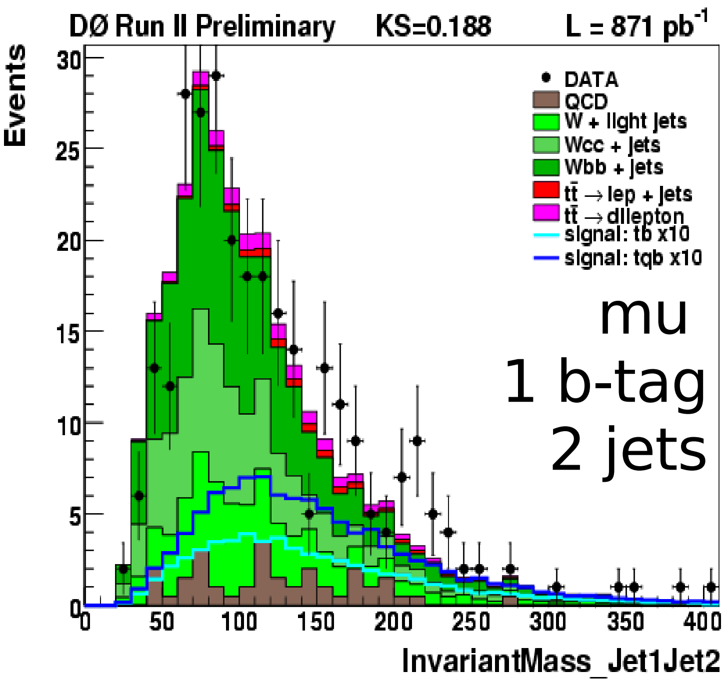
- Multijet events (misidentified lepton)
  - From data

W+jets yield determination

- Normalize W+jets and QCD to data simultaneously before tagging
    - Split data sample in events with real and fake isolated lepton
    - Measure the probability to have an isolated lepton in each sample
- We also know that there are large k-factors for Wbb and Wcc
- 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

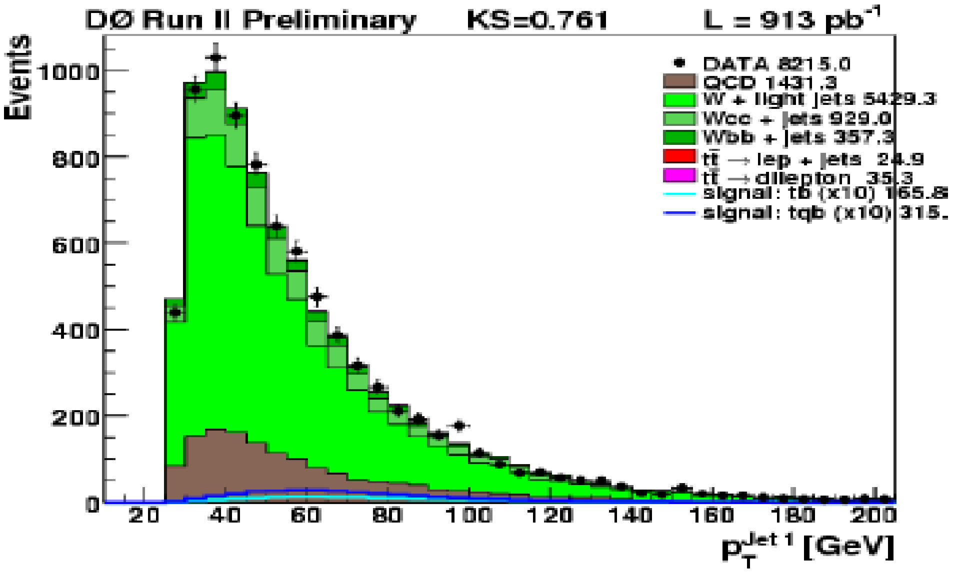


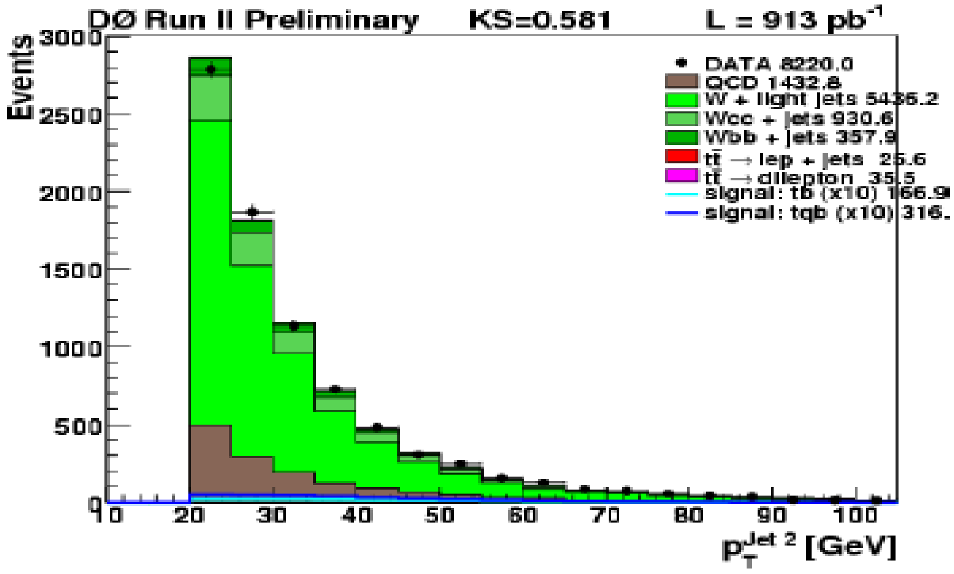


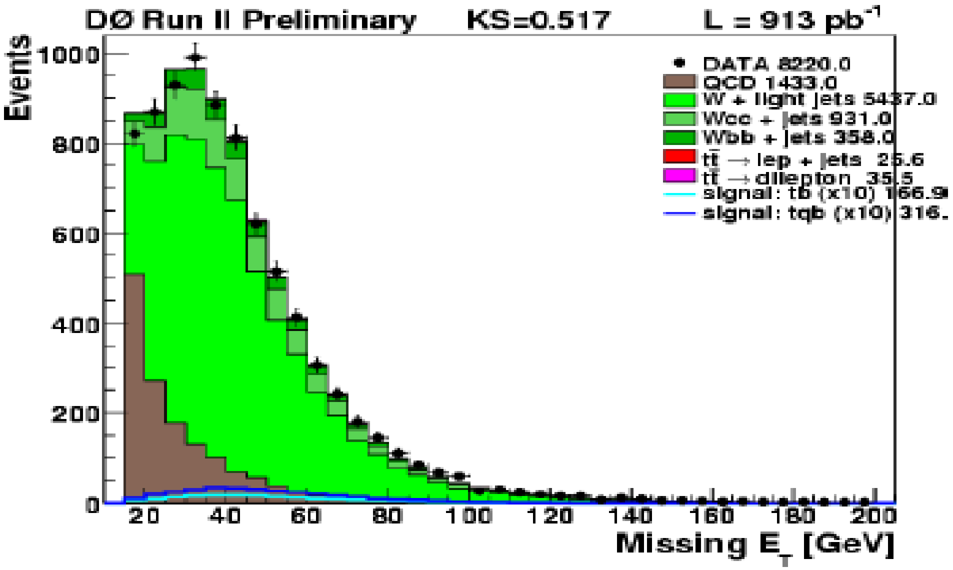


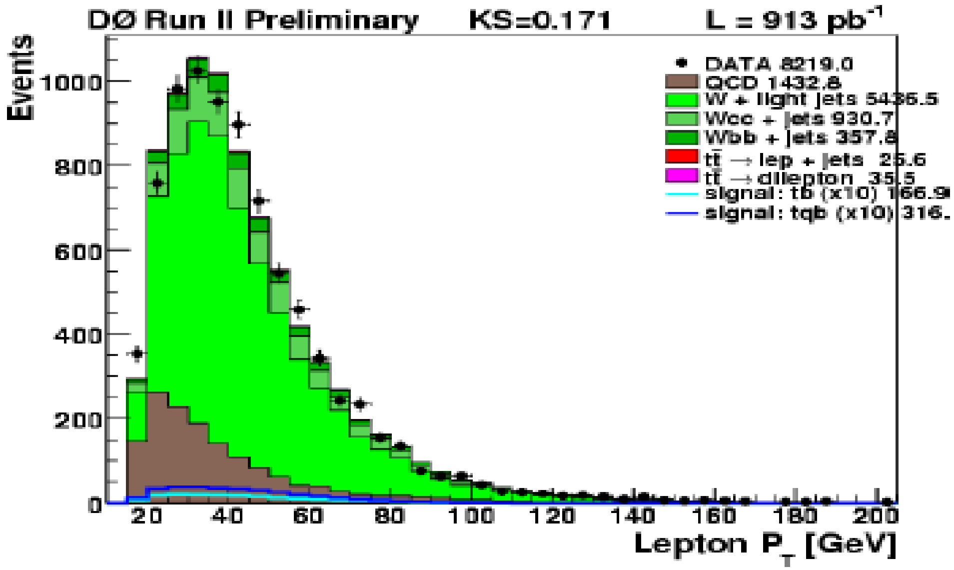
Agreement before tagging

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

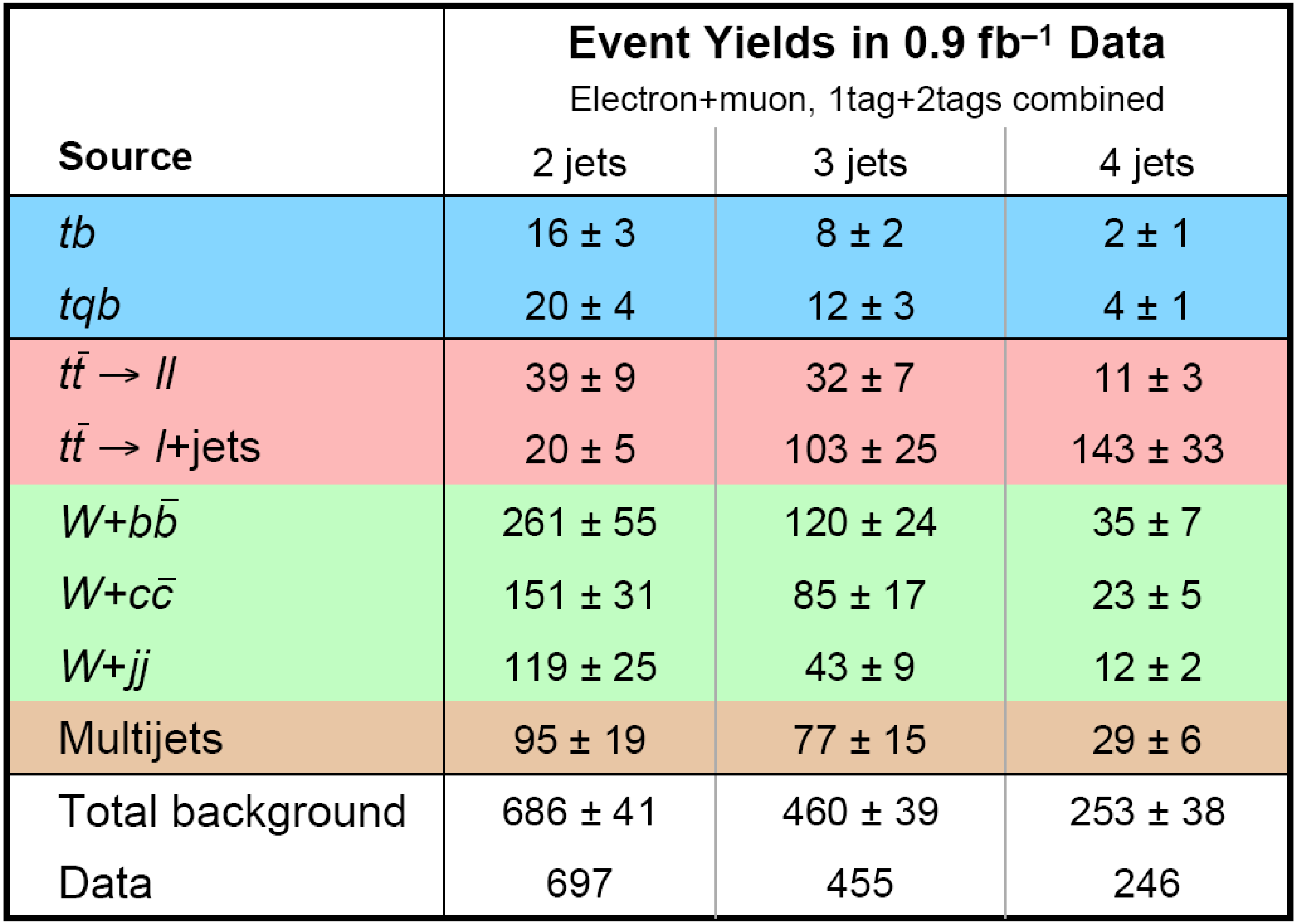








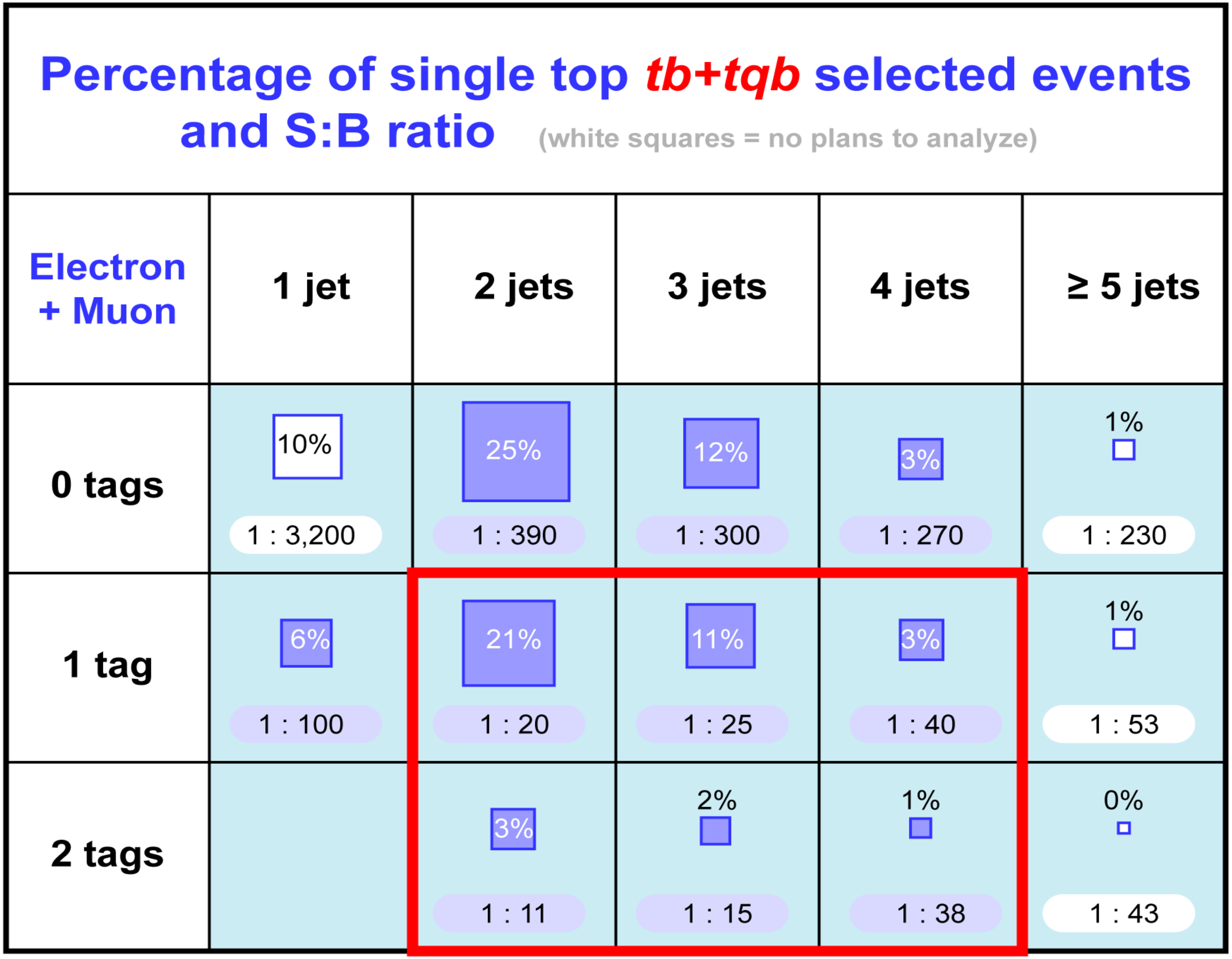
Yields after event selection



Source	Event Yields in $0.9\ \text{fb}^{-1}$ 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

- Optimized the selection to maximize acceptance
  $tb = (3.2 \pm 0.4)\%$ $tqb = (2.1 \pm 0.3)\%$
- Allow a lot of background at this stage!
- Then use multiple distributions to separate signal-background

Event selection and 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

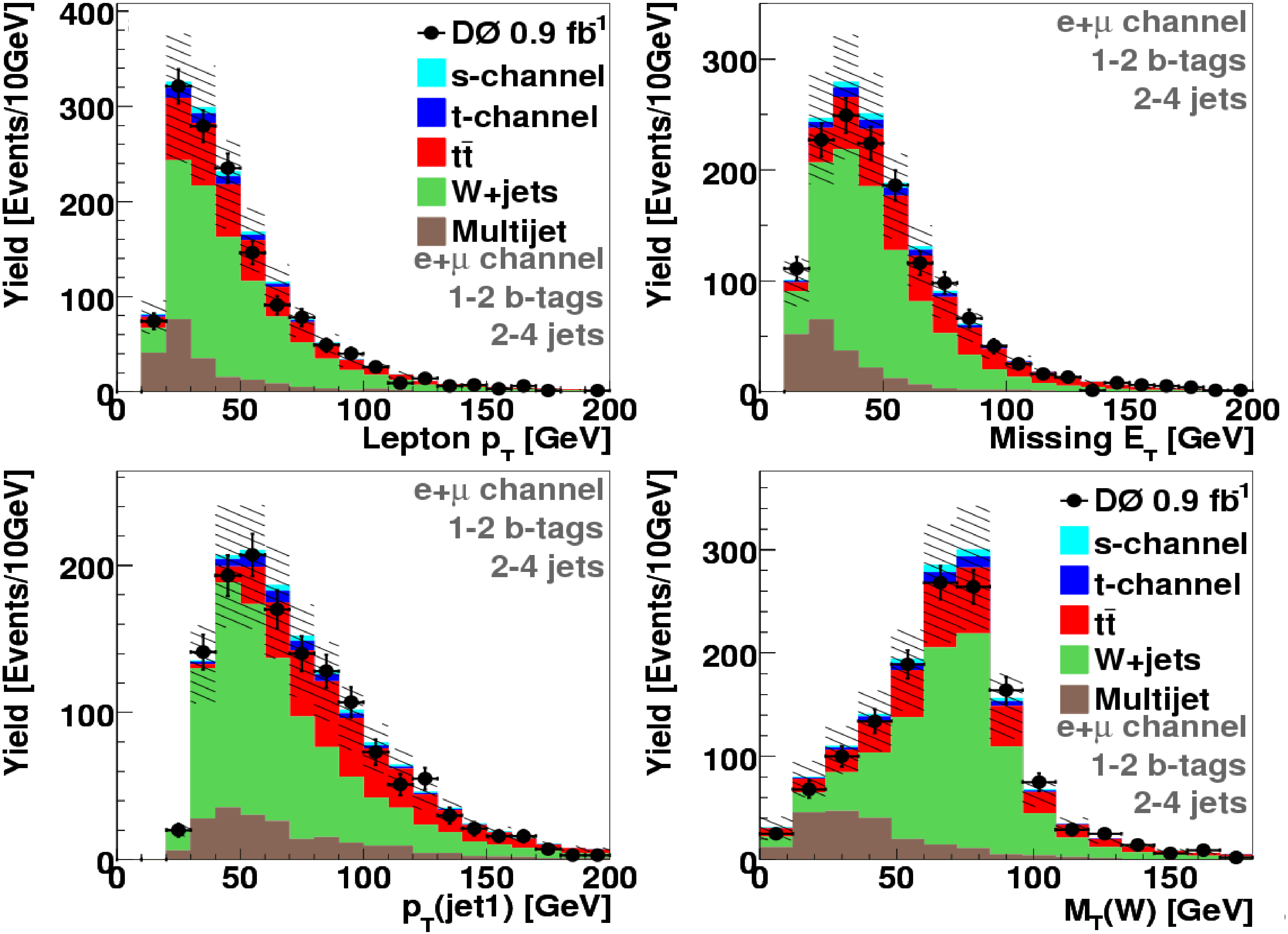
Systematic uncertainties

- Uncertainties are assigned per background, jet multiplicity, lepton channel, and number of tags
- Uncertainties that affect both the **normalization** and the **shapes**: JES and tag rate functions
- Correlations between channels and sources are taken into account

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%

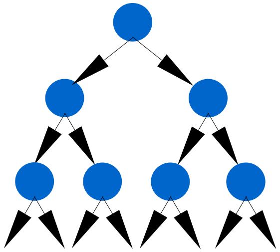
And check 1000s of plots again…



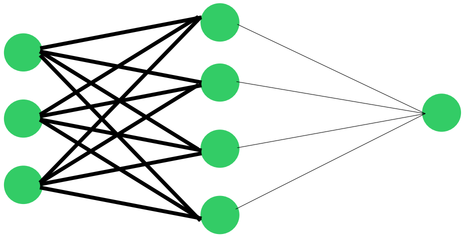
Analysis methods

- Once we understand our data, need to measure the signal
- We cannot use simple cuts to extract the signal: use **multivariate techniques**
- DØ has implemented three analysis methods to extract the signal from the **same dataset**:

Decision Trees



Bayesian NNs



Matrix Elements

$$\int M$$

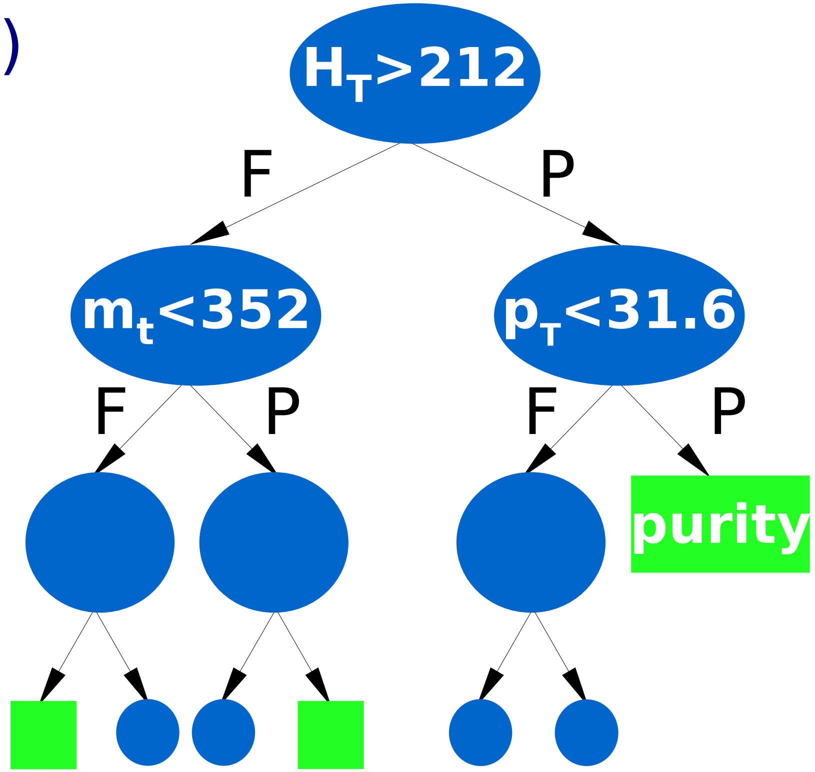
- DT and BNN use same pool of discriminating variables
- ME method uses 4-vectors of reconstructed objects
- Optimized separately for s-channel, t-channel and s+t
- Test response and robustness with ensemble testing

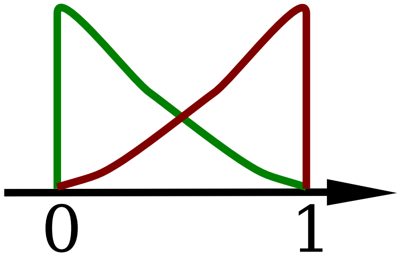
Decision Trees

Machine learning technique widely used in social sciences

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

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





Decision Trees + Boosting

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

AdaBoost algorithm: adaptive boosting

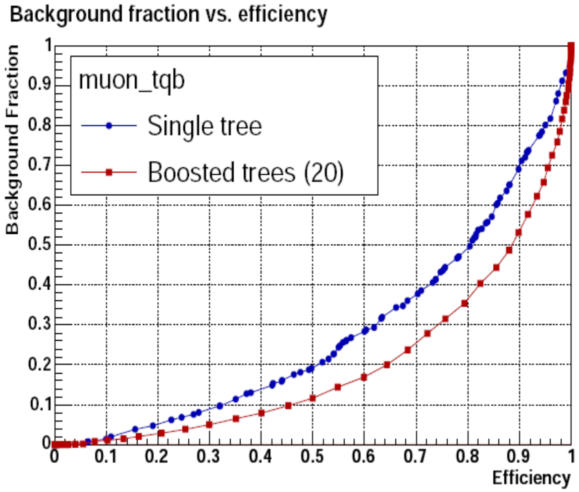
1) Train a tree T_k

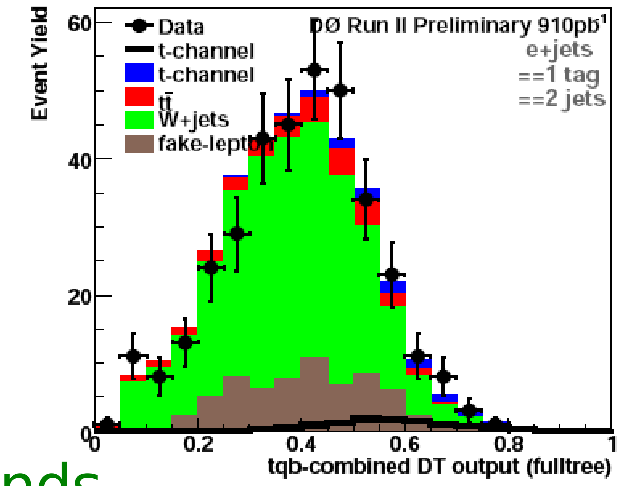
2) Check which events are **misclassified** by T_k

3) Derive tree weight α_k

4) Increase weight of misclassified events

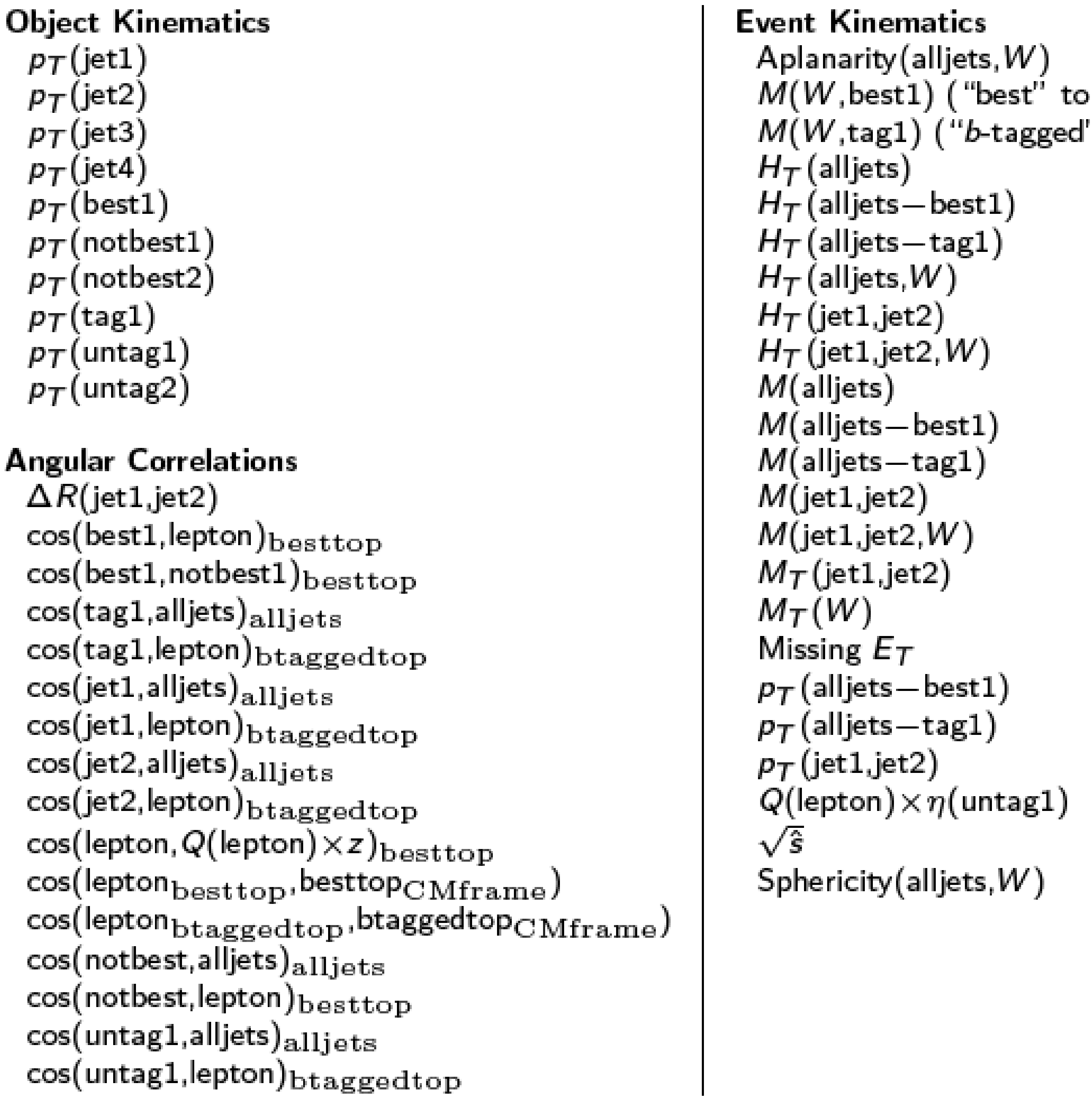
5) Train again to build T_{k+1}





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

Decision Trees: 49 variables



Object Kinematics

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

Angular Correlations

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

Bayesian Neural Networks

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

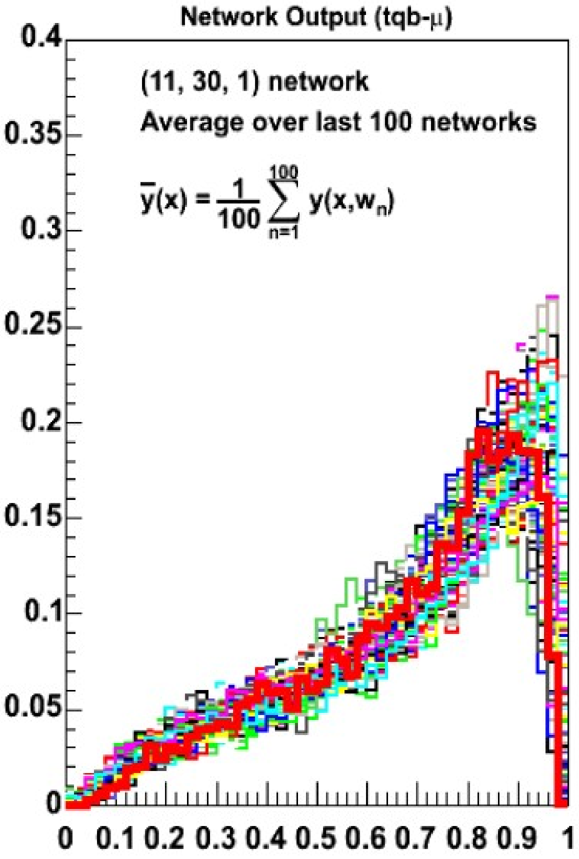
- Instead of choosing one set of weights, find posterior probability density over all possible weights
- Averages over many networks weighted by the probability of each network given the training data
- Use 24 variables (subset of the DT variables) and train against sum of backgrounds

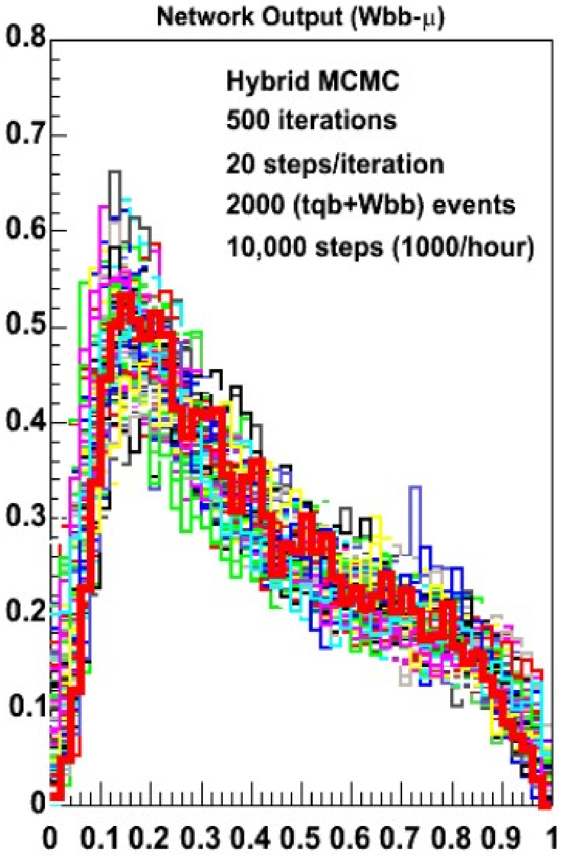
Advantages:

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

Disadvantages:

- Computationally demanding!





Matrix Elements method

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

- Calculate an event probability for signal and background hypothesis

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

Parton distribution functions CTEQ6

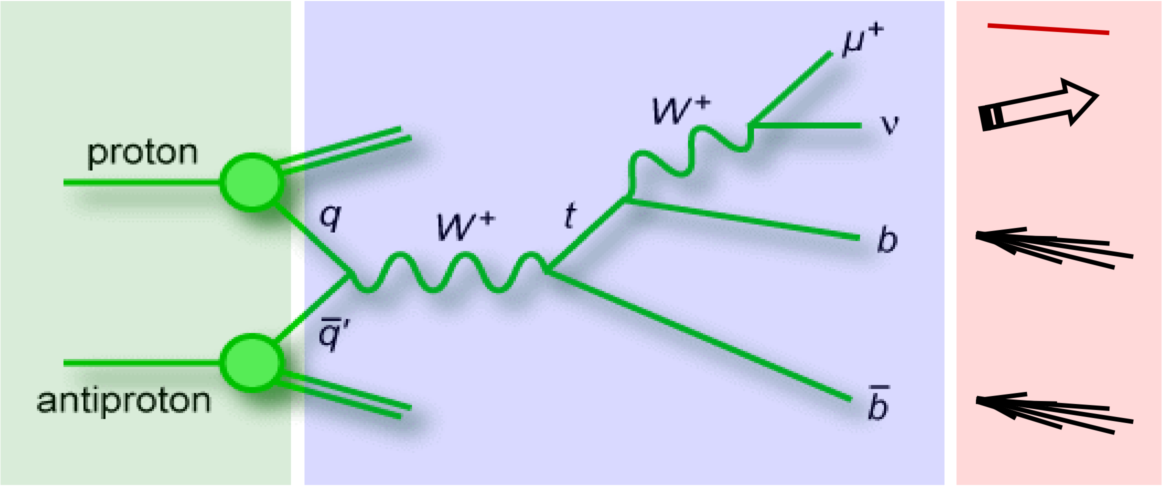
Differential cross section (LO ME from Madgraph)

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

- Uses the 4-vectors of all reconstructed ℓs and jets

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

- Apply b-tagging information



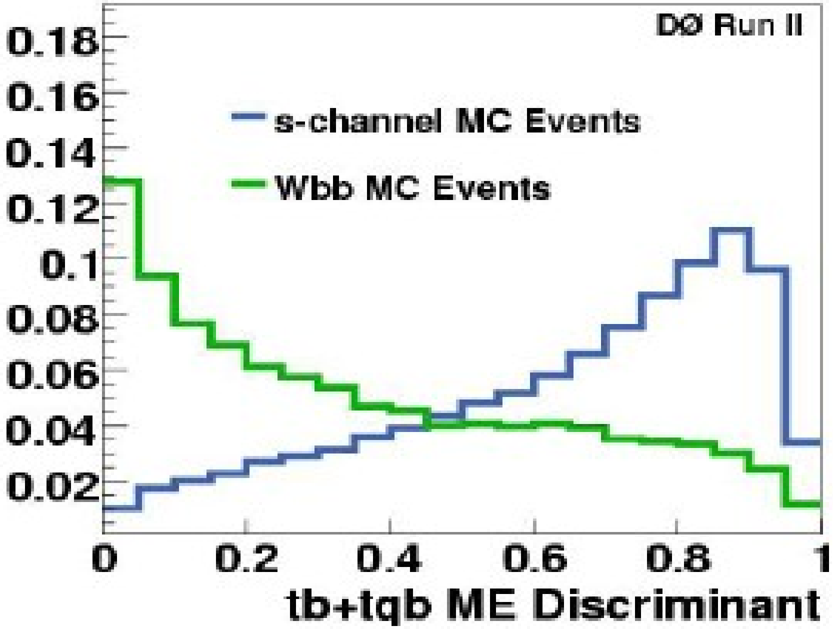
- Integrate over 4 independent variables: assume angles well measured, known masses, momentum and energy conservation

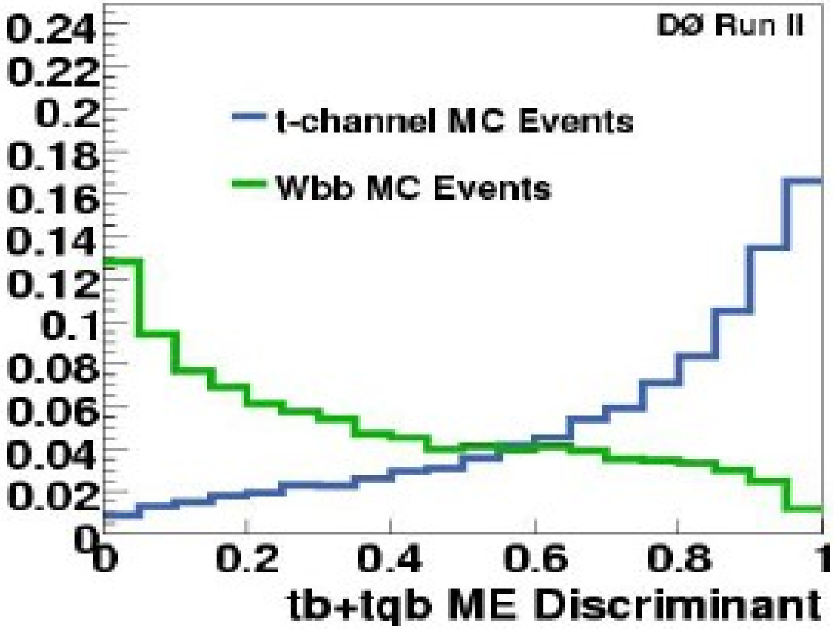
ME discriminant

- Define discriminant based on event probabilities for signal and background

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

- In 2 jet events: use ME for Wbb, Wcg and Wgg backgrounds
- In 3 jet events: use ME for Wbbg, Wggg and tt→ℓ+jets backgrounds
- In tt events, we need to lose one jet: assume one q from W is lost (1.7 times more likely than b) or two jets are merged





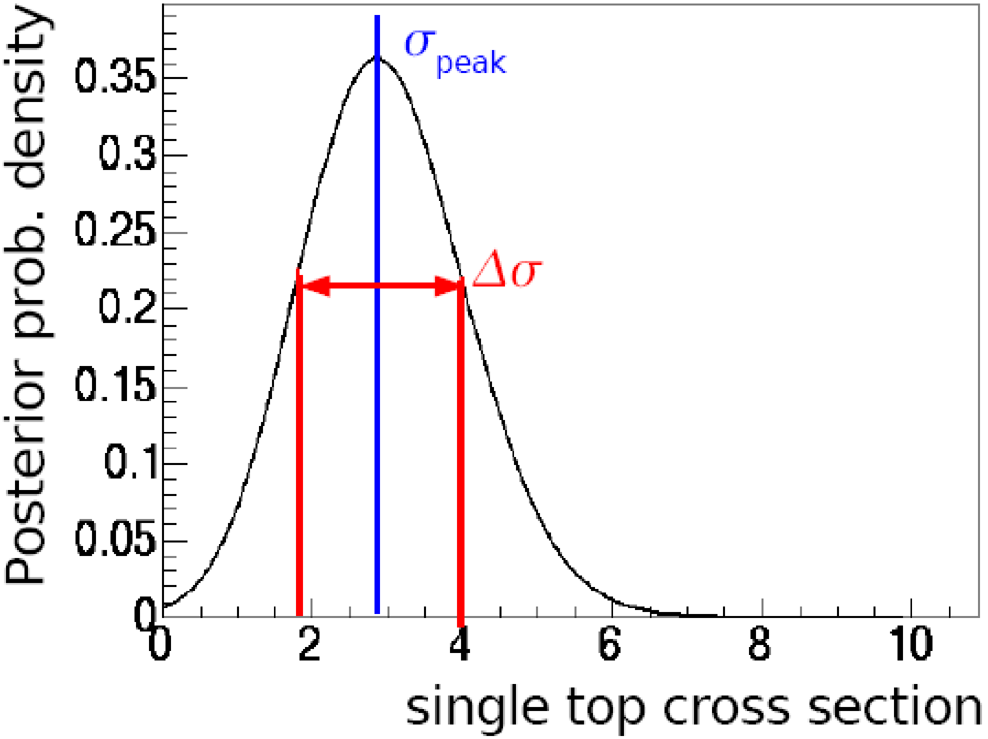
Measuring the cross section

- 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:

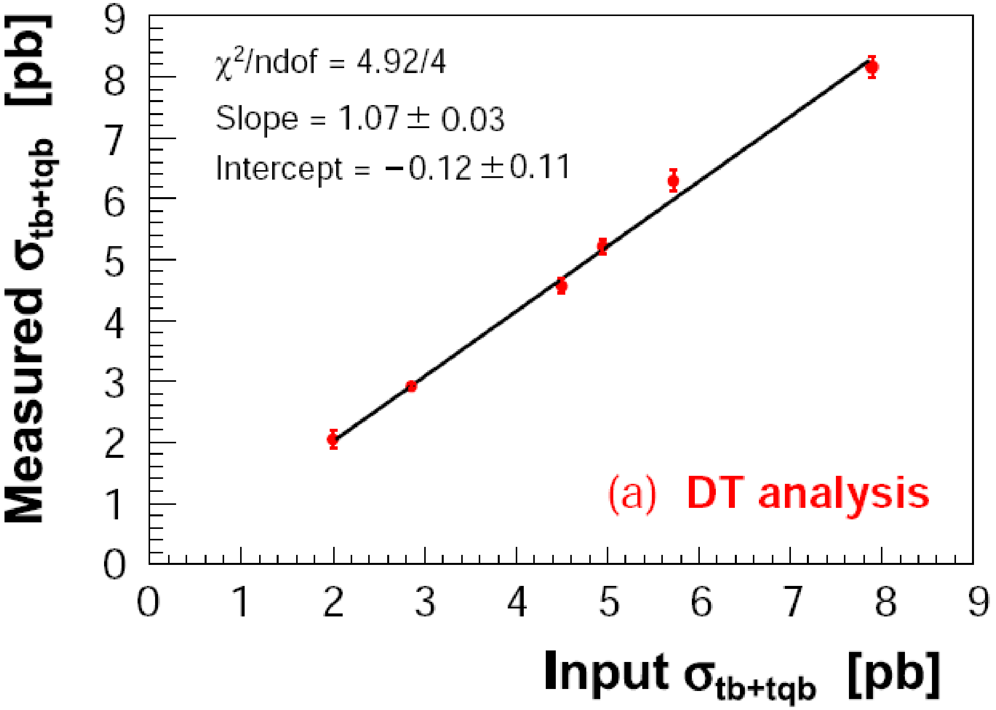
$$Post(\sigma|D) \equiv P(\sigma|D) \propto \int_a \int_b P(D|\sigma, a, b) Prior(\sigma) Prior(a, b)$$

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

Ensemble testing

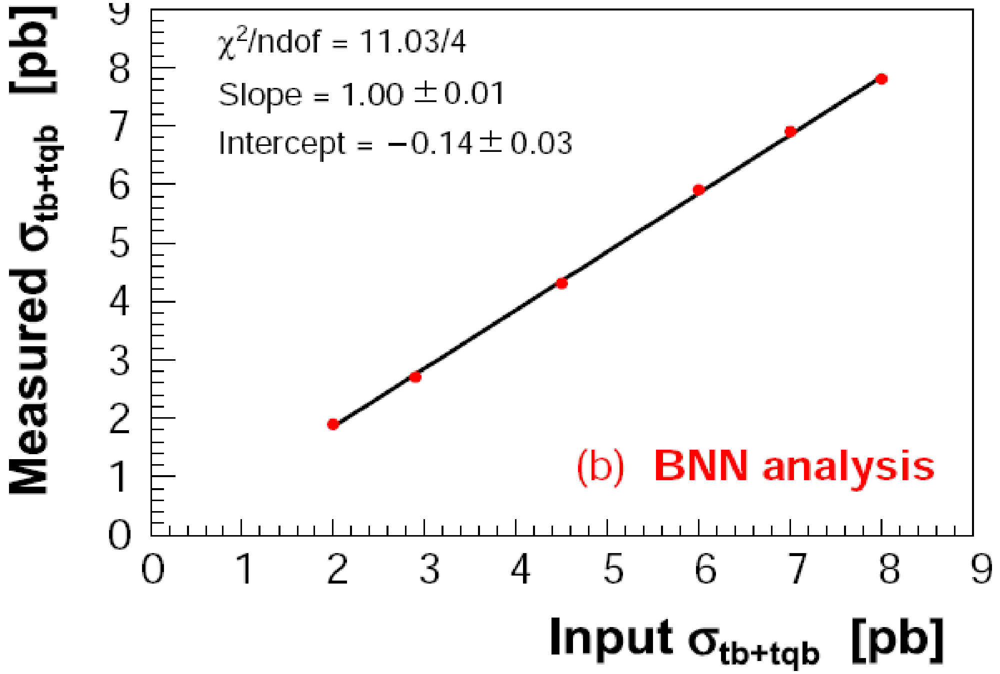
- To verify that all this machinery is working properly, we test with many sets of **pseudo-data**
- Wonderful tool to test analysis methods!
  Run DØ experiment 1000s of times
- Use pool of MC events to draw events with bkgd. yields fluctuated according to **uncertainties**, reproducing the **correlations** between components introduced in the normalization to data
- Randomly sample a Poisson distribution to simulate **statistical** fluctuations
- Generated ensembles include:
  1) 0-signal ensemble ($\sigma_{\mathrm{s+t}}$ = 0 pb)
  2) SM ensemble ($\sigma_{\mathrm{s+t}}$ = 2.9 pb)
  3) “Mystery” ensembles to test analyzers ($\sigma_{\mathrm{s+t}}$ = ?? pb)
  4) Ensemble at measured cross-section ($\sigma_{\mathrm{s+t}} = \sigma_{\mathrm{measured}}$)
  5) A high luminosity ensemble
- Each analysis tests linearity of “response” to single top

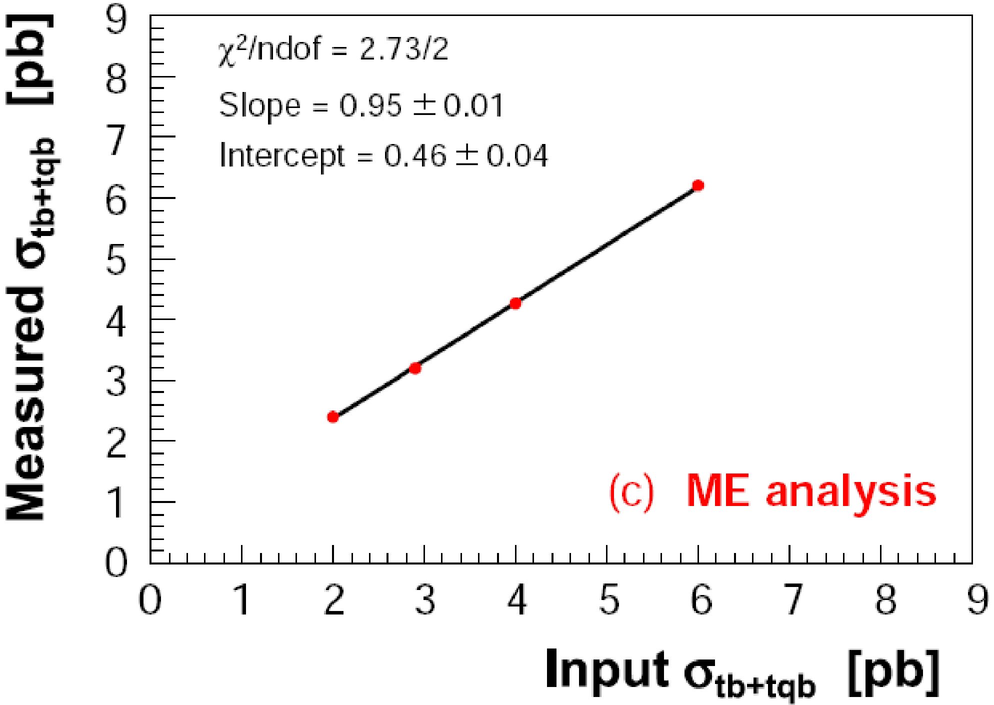
Responses



Using the ensemble tests:

- SM ensemble is returned at the right value
- “Mystery” ensembles are unraveled
- Linear response is achieved





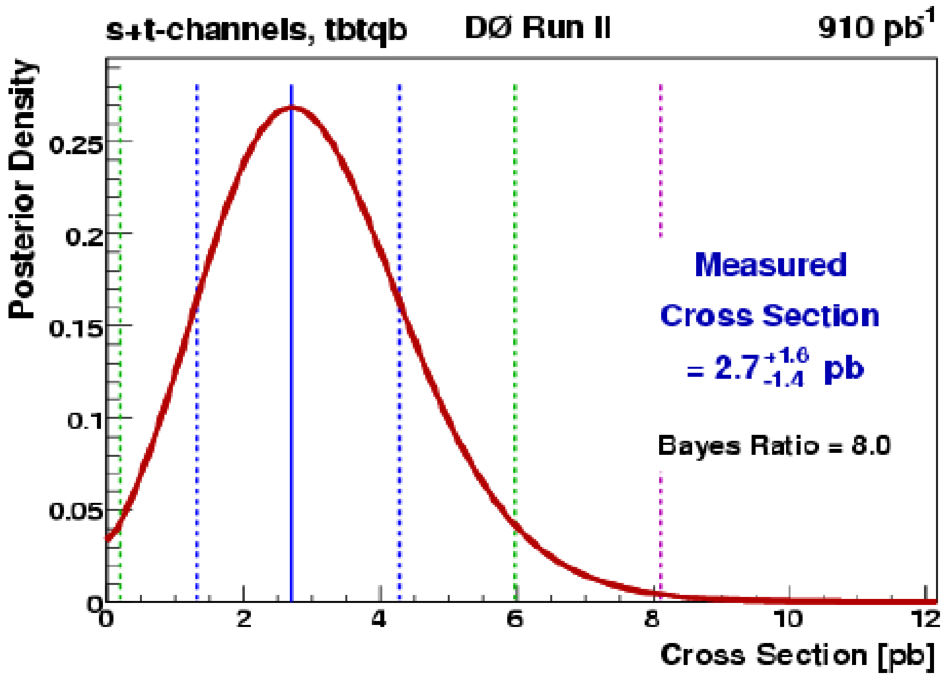
Expected p-values and cross sections

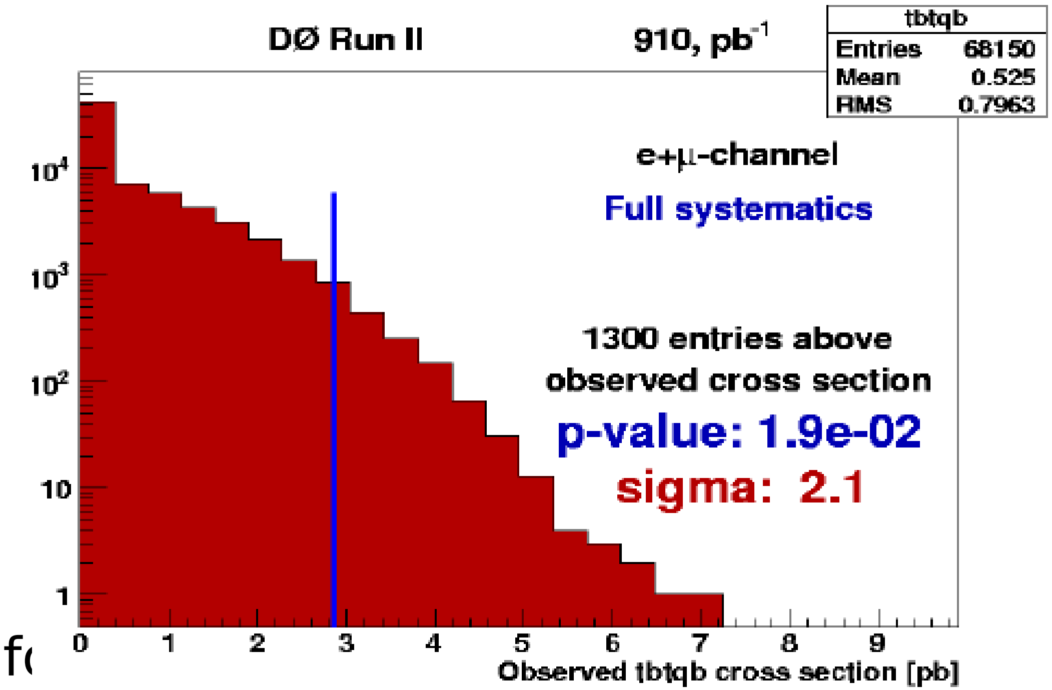
- The expected σ is obtained assuming data=expected SM background
- Use 0-signal ensembles to determine the significance:

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

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

Decision Trees	**Bayesian NN**	**Matrix Elements**
p-value 1.9%	p-value 1.6%	p-value 3.1%
exp. sig. 2.1σ	exp. sig. 2.2σ	exp. sig. 1.9σ

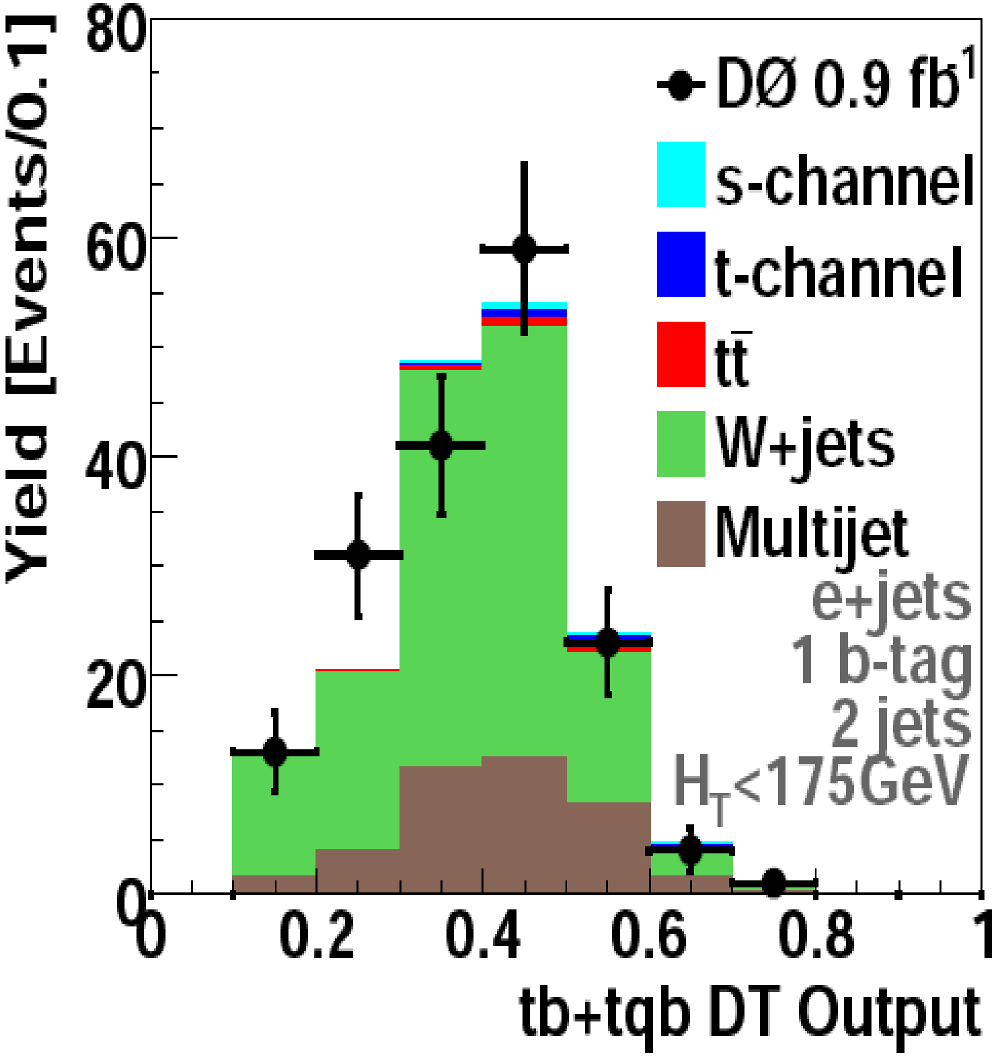


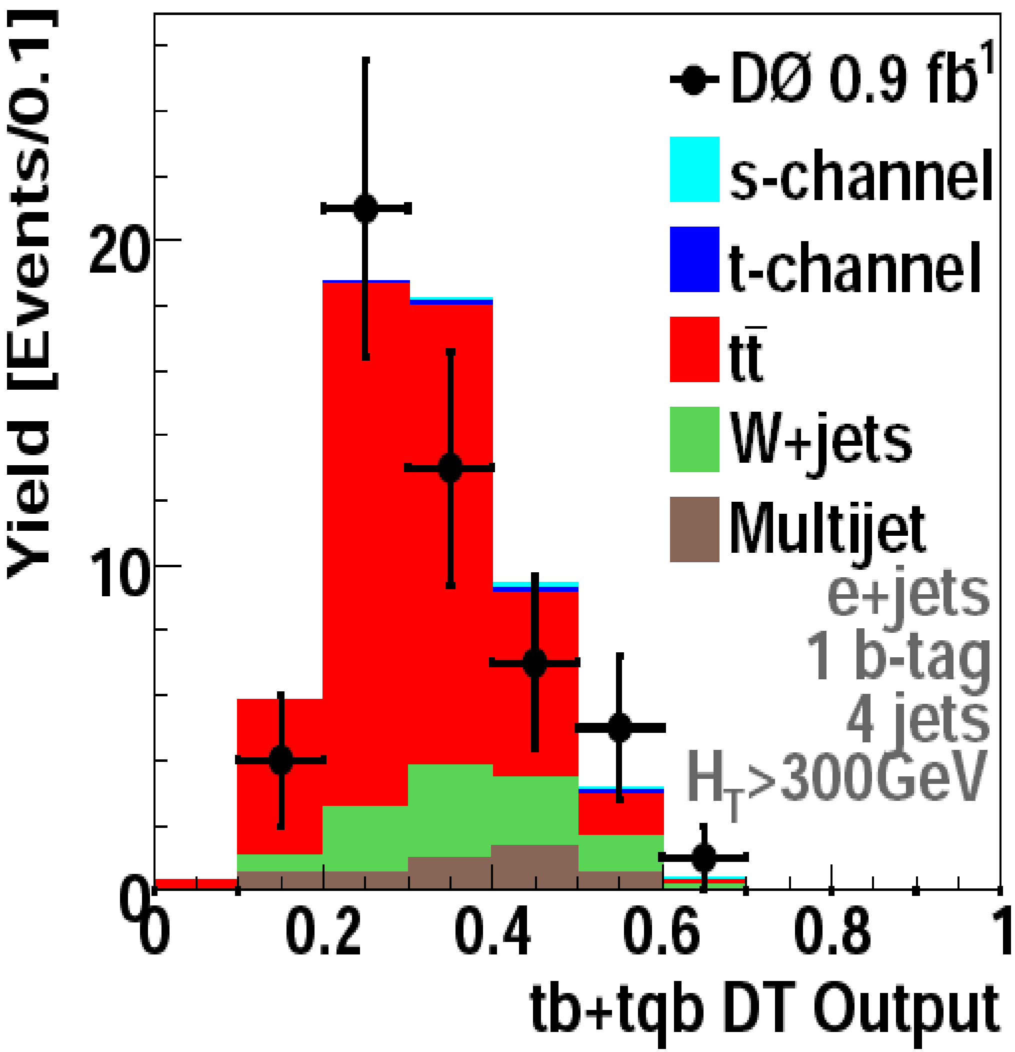


DT cross check samples

Check the description of the data in the DT output

- W+jets: 2 jets and H_T(lepton,MET,alljets) < 175 GeV
- tt: 4 jets and and H_T(lepton,MET,alljets) > 300 GeV

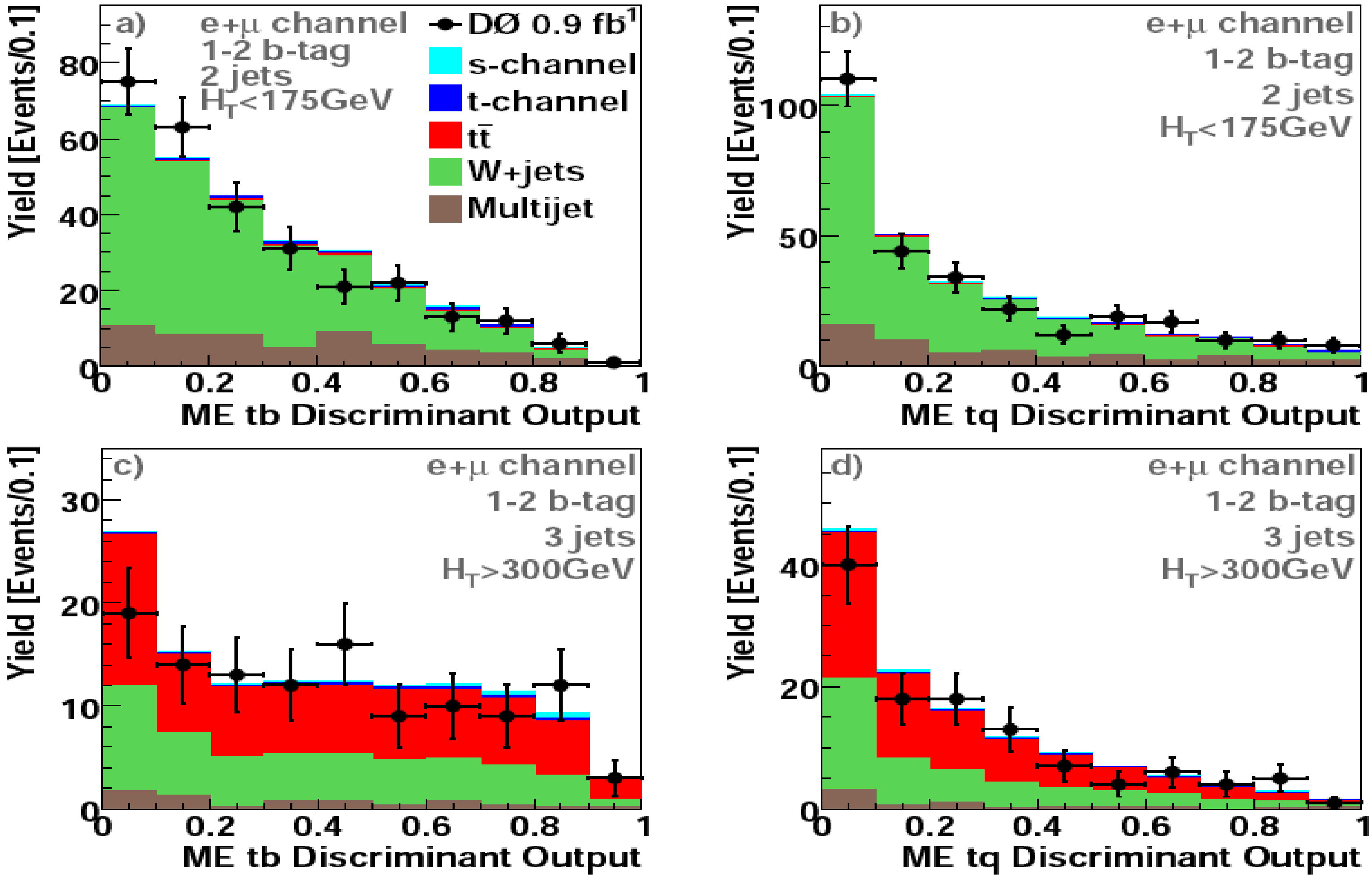




ME cross check samples

Check the description of the data in the ME output

- Soft W+jets: H_T(lepton,MET,alljets) < 175 GeV
- Hard W+jets: H_T(lepton,MET,alljets) > 300 GeV

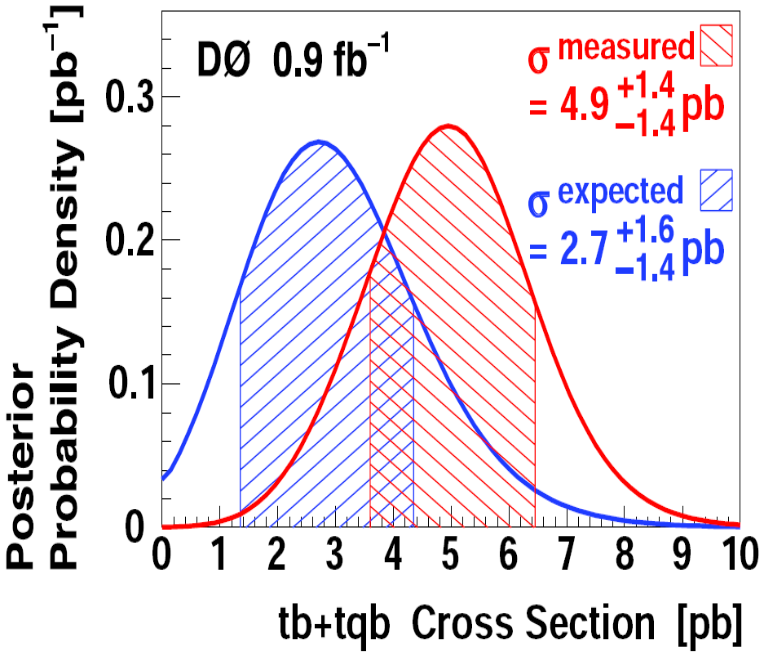


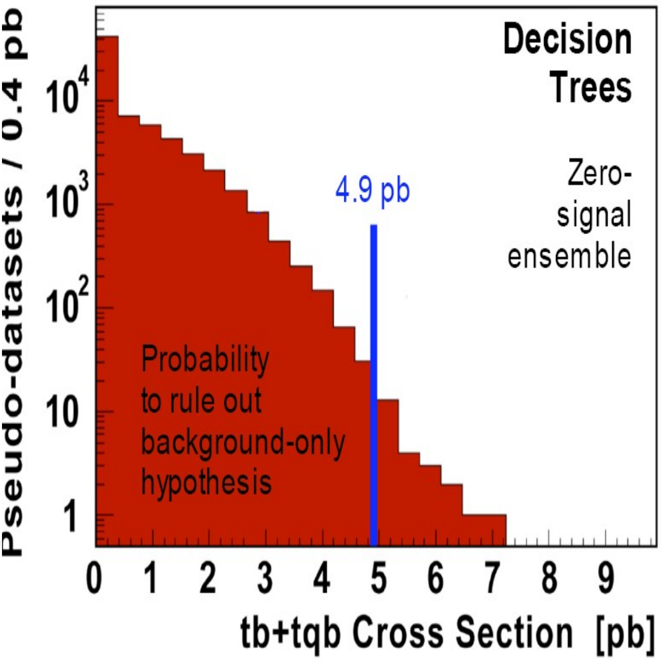
Observed results

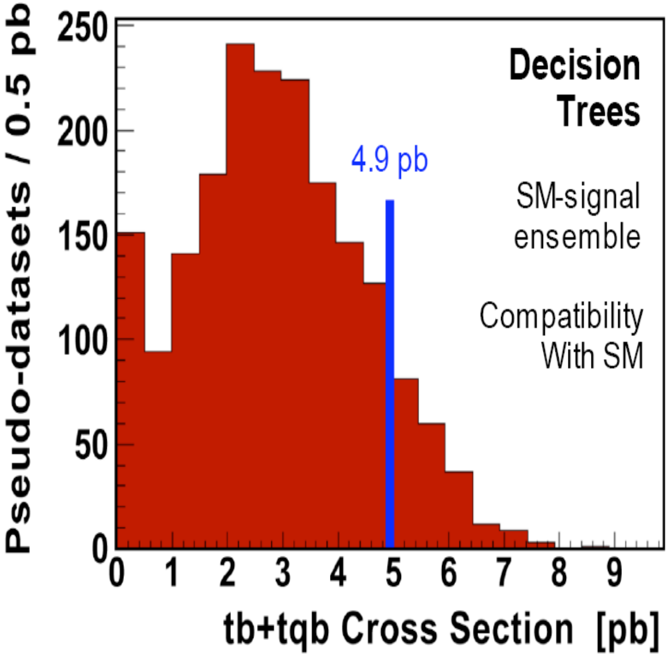
	Decision Trees		Bayesian NN		Matrix Elements	
	Expected	Observed	Expected	Observed	Expected	Observed
σ(tb+qb) [pb]	$2.7^{+1.6}_{-1.4}$	**4.9±1.4**	$2.7^{+1.5}_{-1.5}$	**$4.4^{+1.6}_{-1.4}$**	$3.0^{+1.8}_{-1.5}$	**$4.8^{+1.6}_{-1.4}$**
significance	2.1σ	**3.4σ**	2.2σ	**3.2σ**	1.9σ	**3.2σ**

All three analyses measure >3σ! Evidence for single top production!

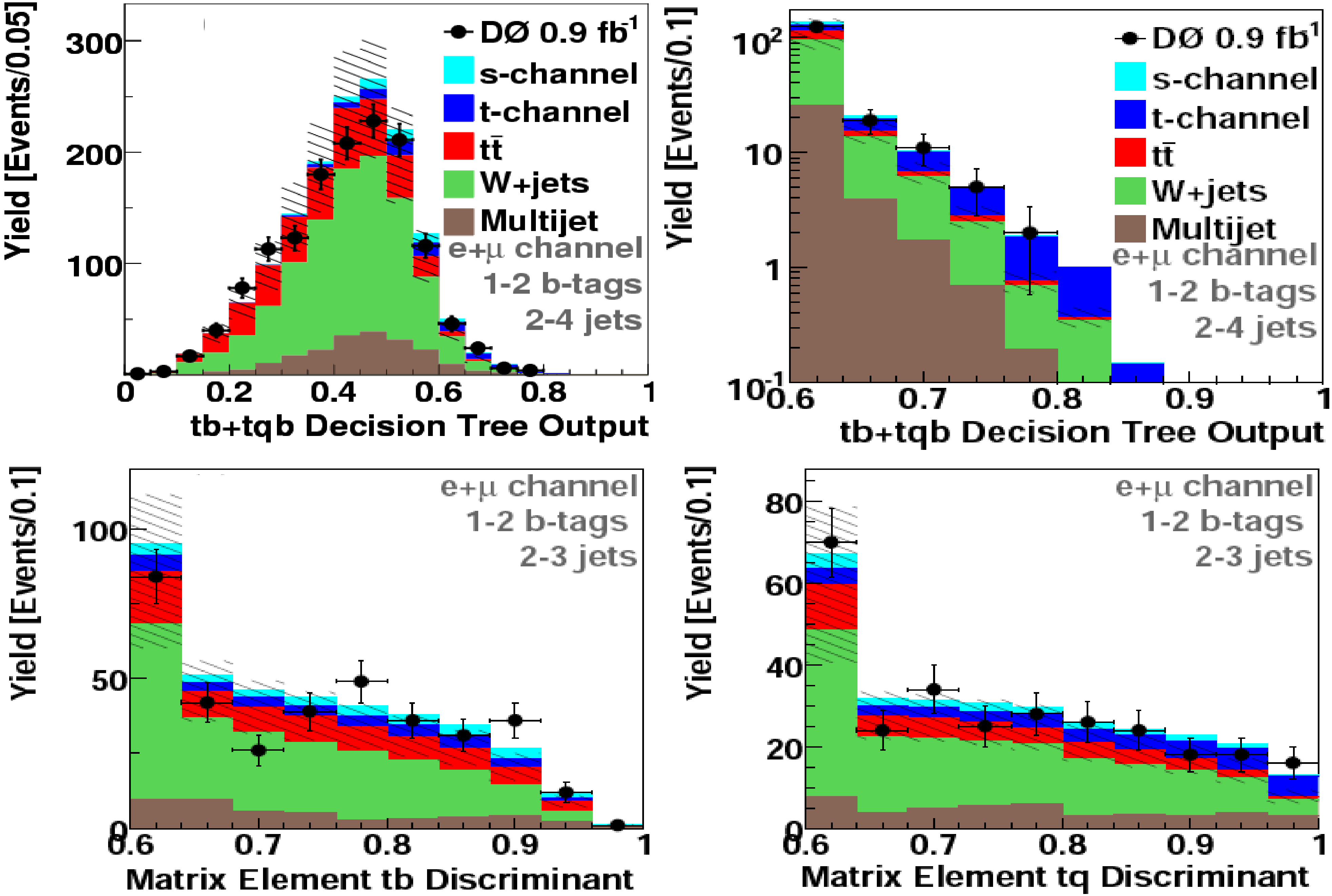
- Results are compatible with the SM at ~1 std. dev.







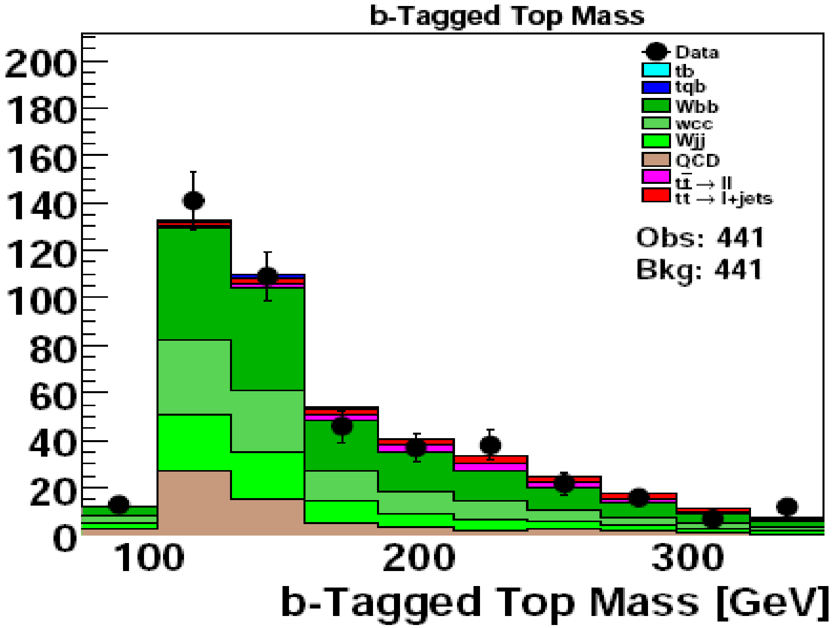
Excess in the high discriminant regions

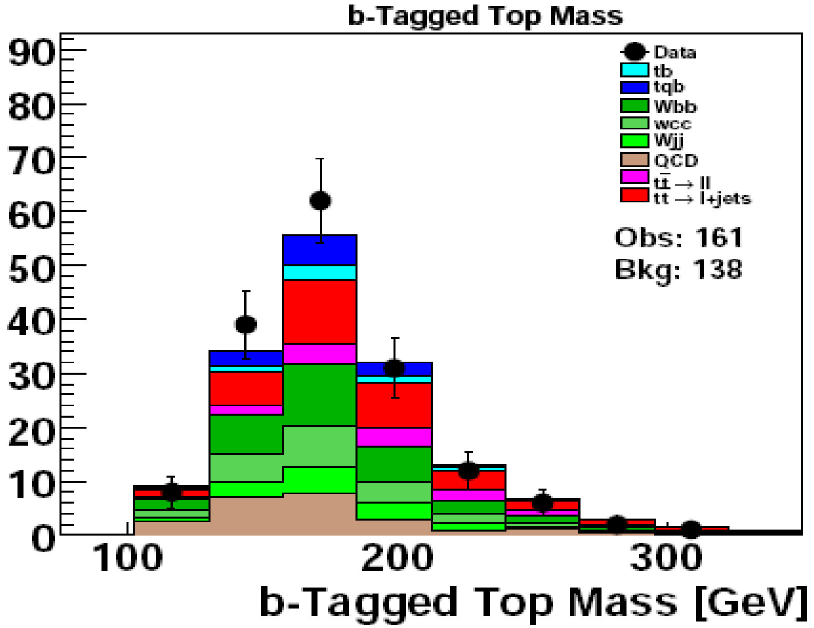


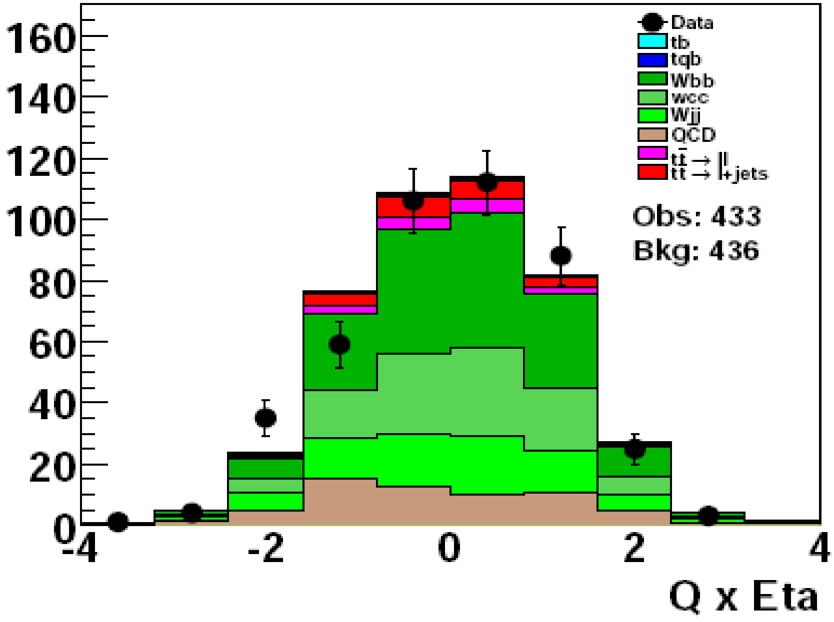
ME event characteristics

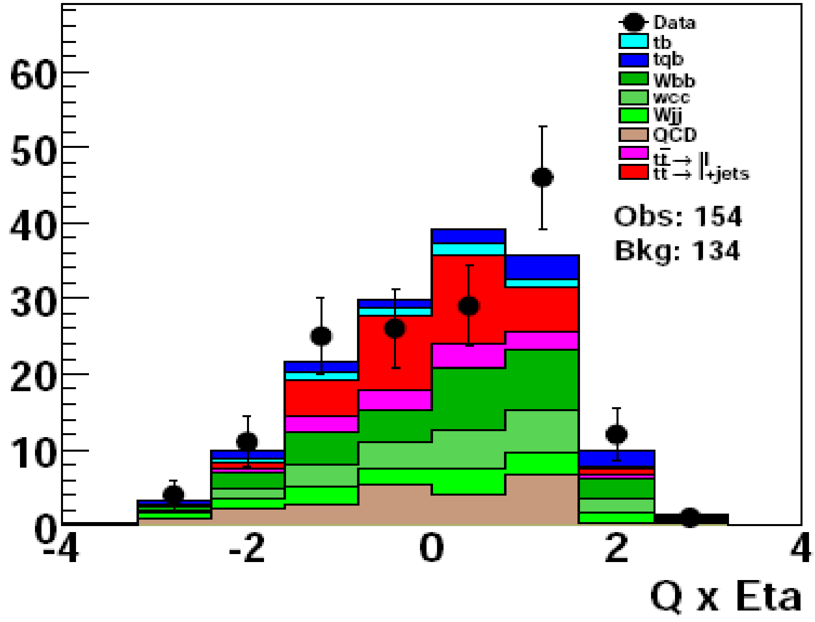
ME Discriminant < 0.4

ME Discriminant > 0.7



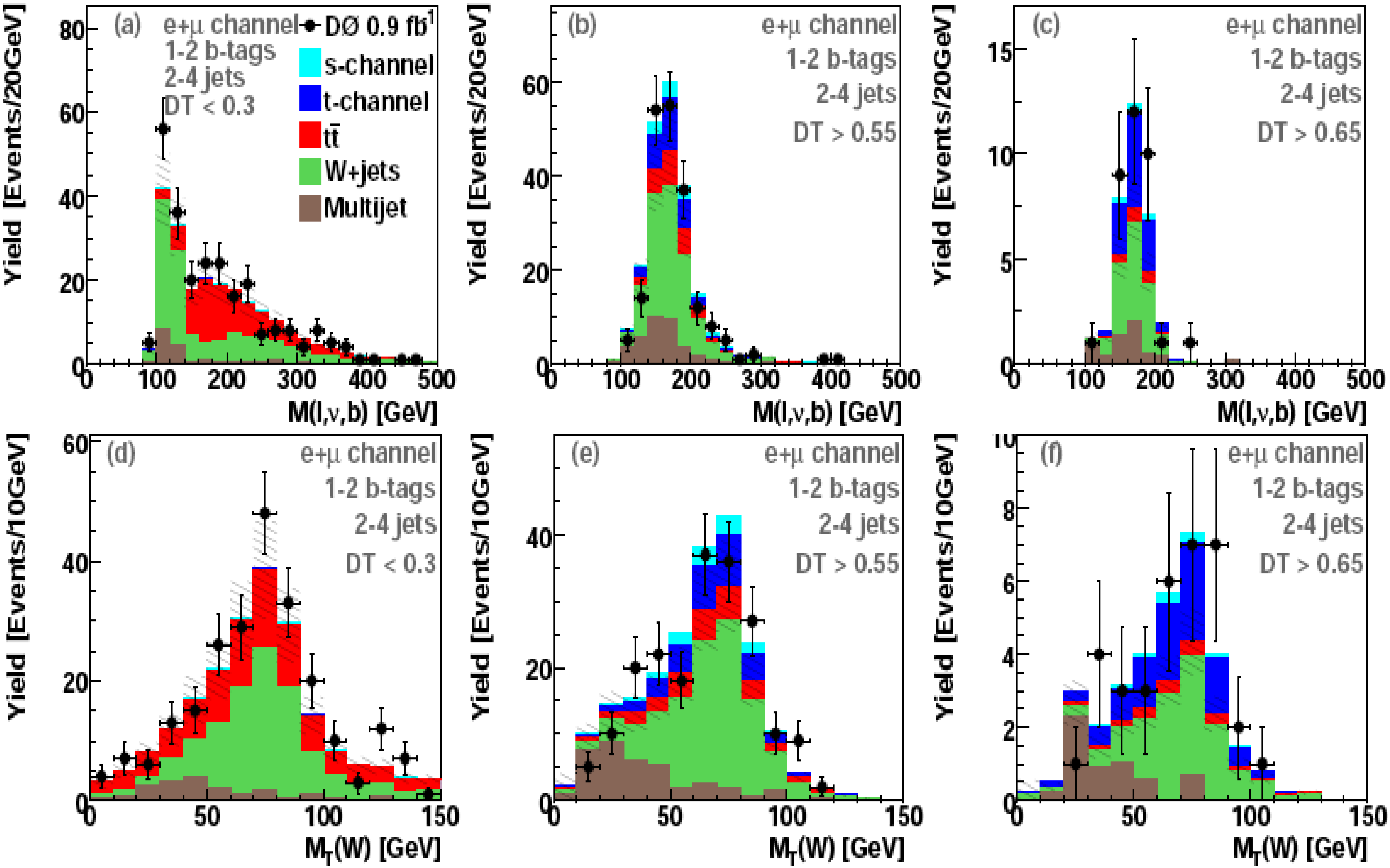




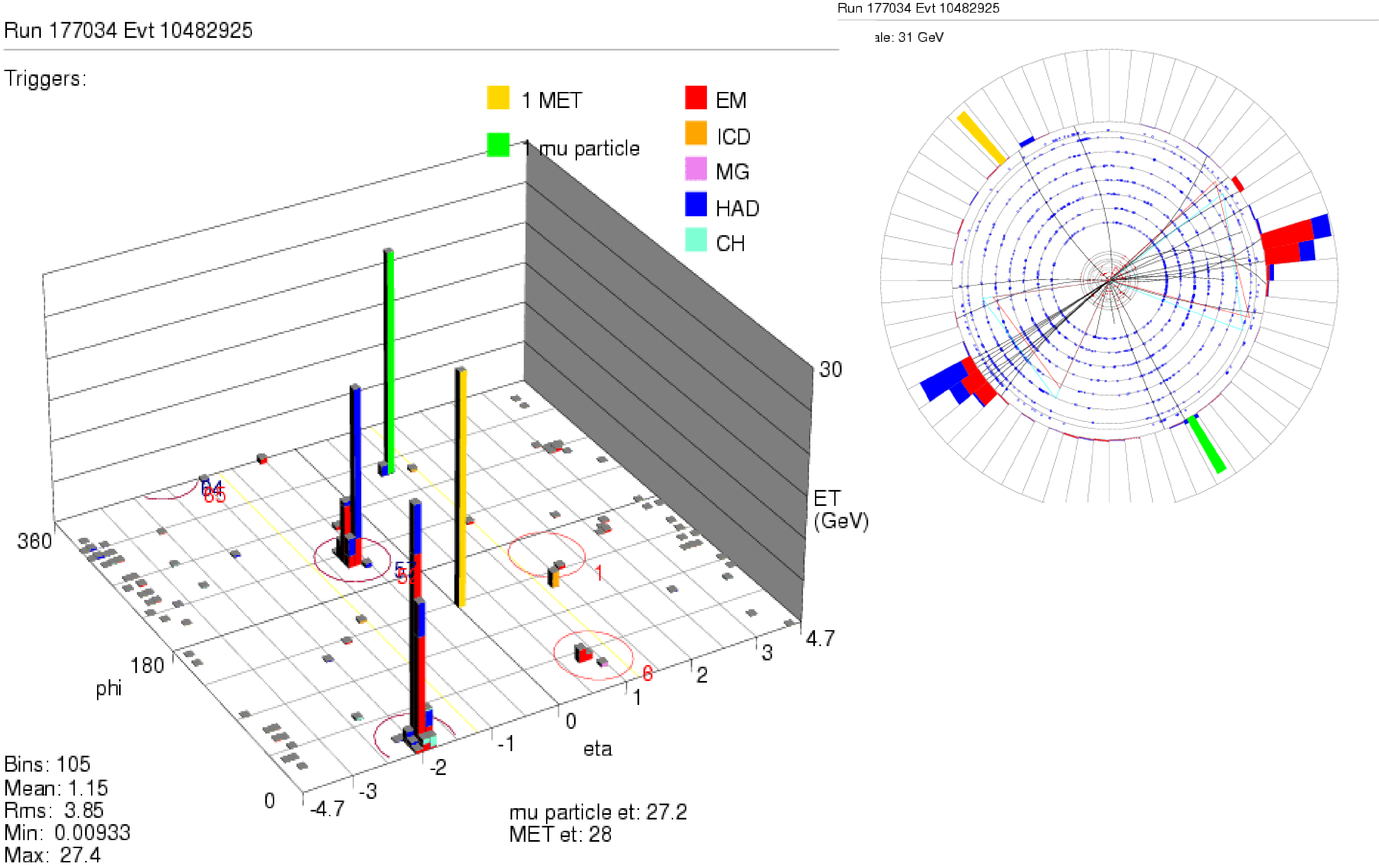


DT event characteristics

DT Discriminant < 0.3 **DT Discriminant > 0.55** **DT Discriminant > 0.65**



A candidate event



Run 177034 Evt 10482925

Triggers:

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

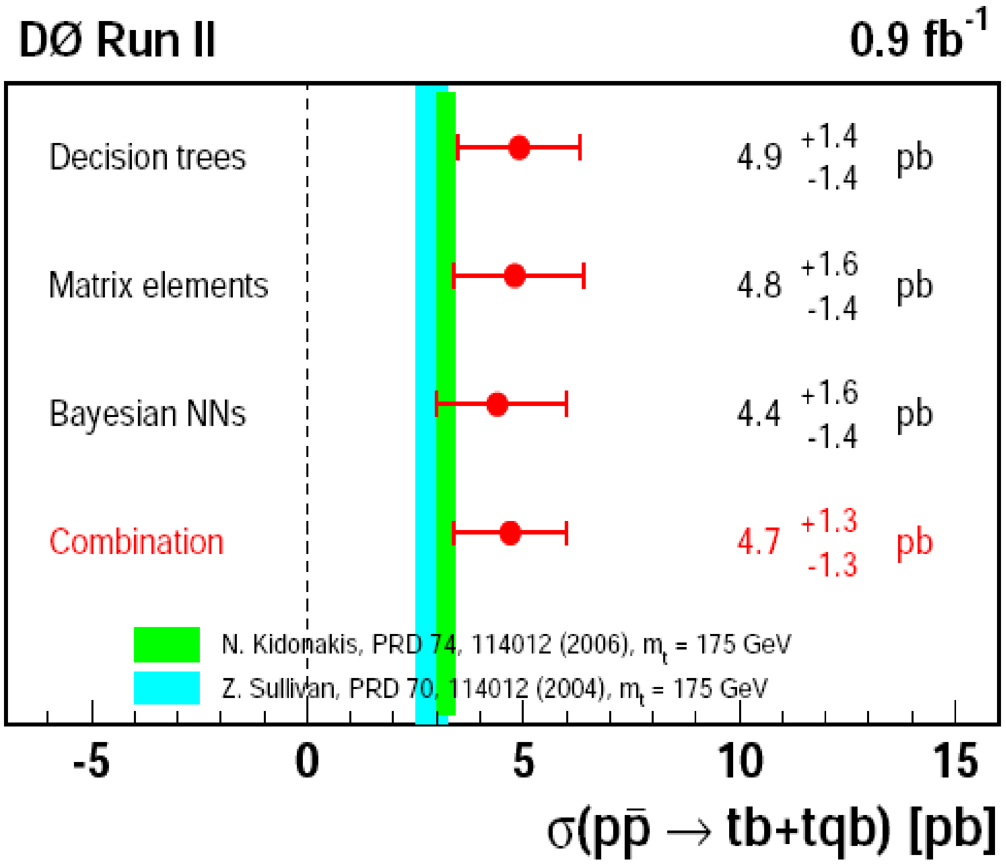
mu particle et: 27.2
MET et: 28

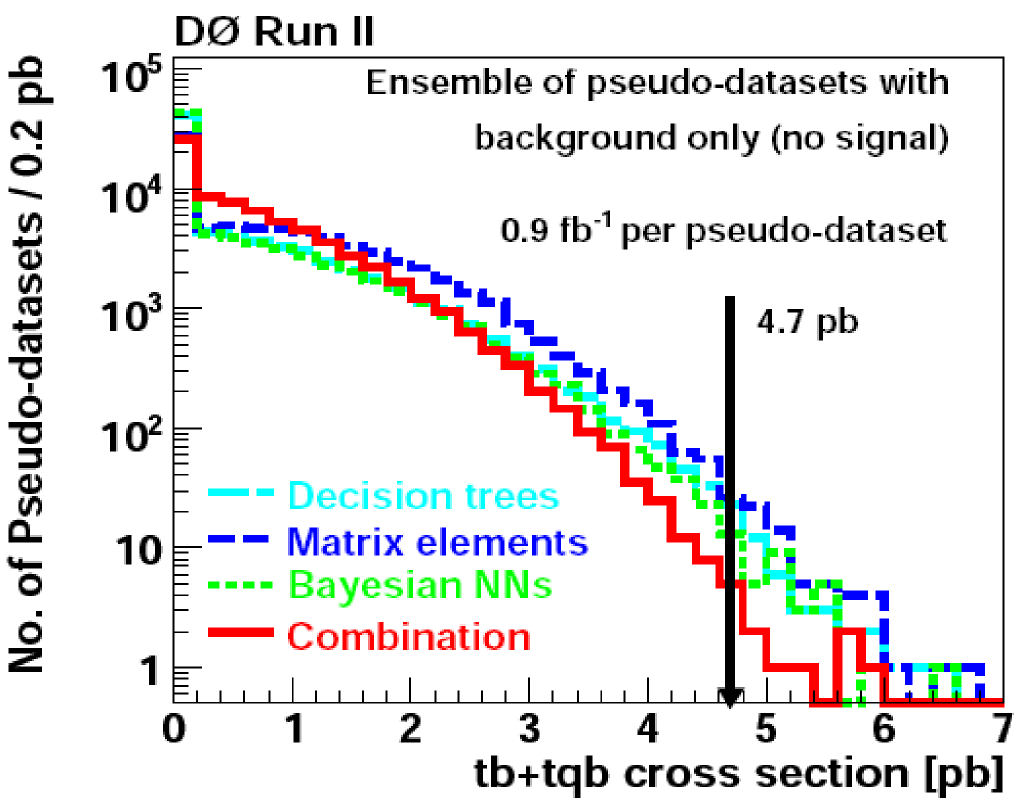
Combination of analyses

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

$$\rho = \begin{array}{cccc} DT & ME & BNN & \\ \left(\begin{array}{c} 1 \\ 0.64 \\ 0.66 \end{array}\right. & \begin{array}{c} 0.64 \\ 1 \\ 0.59 \end{array} & \left.\begin{array}{c} 0.66 \\ 0.59 \\ 1 \end{array}\right) & \begin{array}{c} DT \\ ME \\ BNN \end{array} \end{array}$$

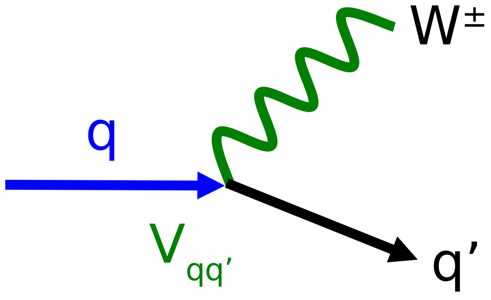
Combined result: 4.7 ± 1.3 pb ➔ Significance of 3.6 std. dev.





CKM matrix element V_{tb}

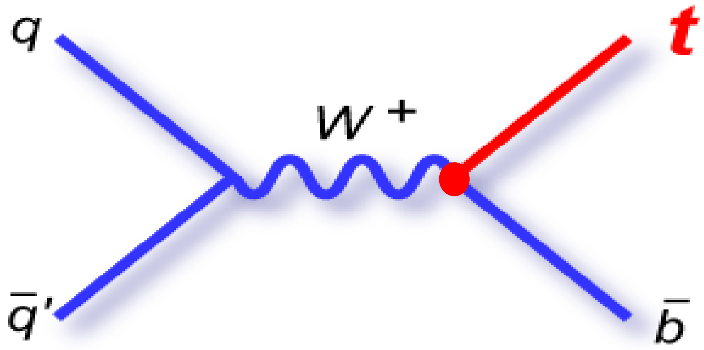
$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & \boxed{V_{tb}} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}$$



- Weak interaction eigenstates and mass eigenstates are not the same: there is **mixing** between quarks → CKM matrix
- In SM: top must decay to W and d, s or b quark
  - $V_{td}^2+V_{ts}^2+V_{tb}^2 = 1$
  - Strong constraints on V_{td} and V_{ts}: **$V_{tb} > 0.998$**
  - Assuming unitarity and 3 generations: B(t→Wb)~100%
- If there is new physics:
  - $V_{td}^2+V_{ts}^2+V_{tb}^2 < 1$
  - No constraint on V_{tb}
  - Interactions between the top quark and weak gauge bosons are extremely interesting!

Measuring $|V_{tb}|$

- Once we have a cross section measurement, we can make the first direct measurement of $|V_{tb}|$
- Calculate posterior in $|V_{tb}|^2$: $\sigma \propto |V_{tb}|^2$



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

- Most general Wtb vertex:

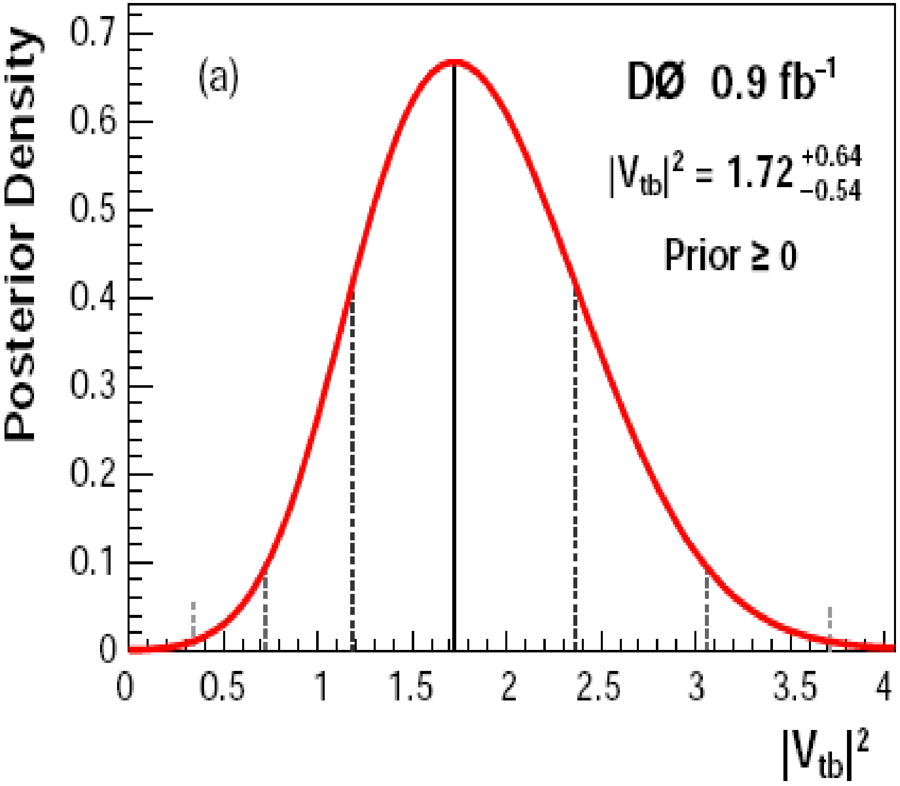
$$\Gamma^{\mu}_{tbW} = -\frac{g}{\sqrt{2}} V_{tb} \left\{ \gamma^{\mu} \left[f_1^L P_L + f_1^R P_R \right] - \frac{i\,\sigma^{\mu\nu}}{M_W} (p_t - p_b)_{\nu} \left[f_2^L P_L + f_2^R P_R \right] \right\}$$

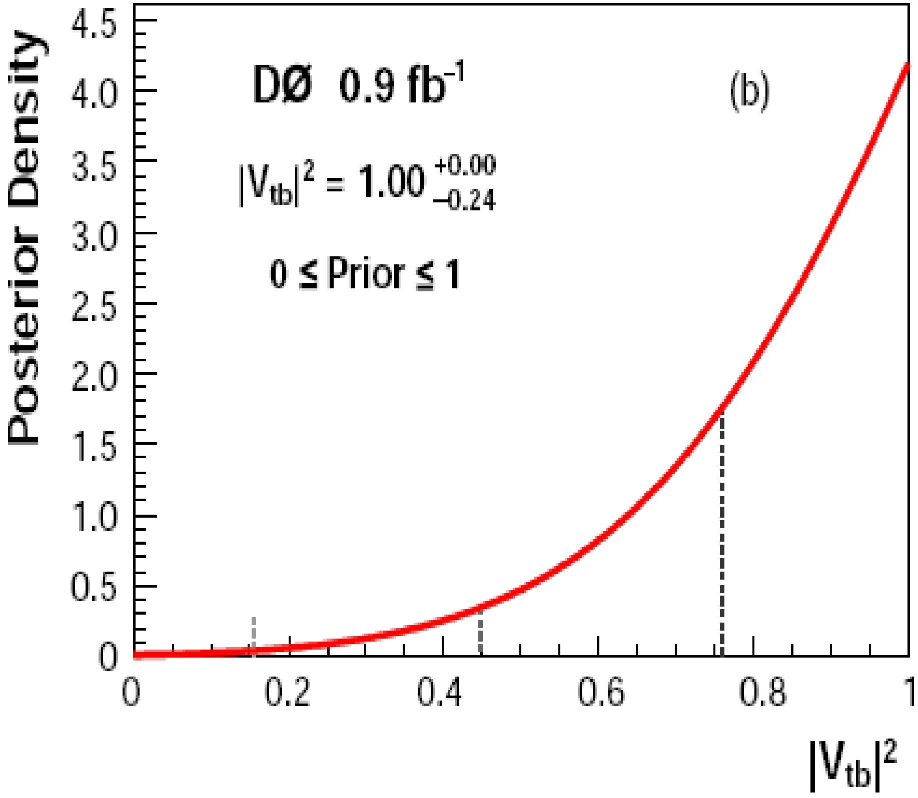
- Assume:
  - **SM top decay**: $V_{td}^2 + V_{ts}^2 \ll V_{tb}^2$
  - Pure V-A interaction: $\mathbf{f_1^R = 0}$
  - CP conservation: $\mathbf{f_2^L = f_2^R = 0}$

No need to assume three quark families or CKM matrix unitarity!

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

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





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

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

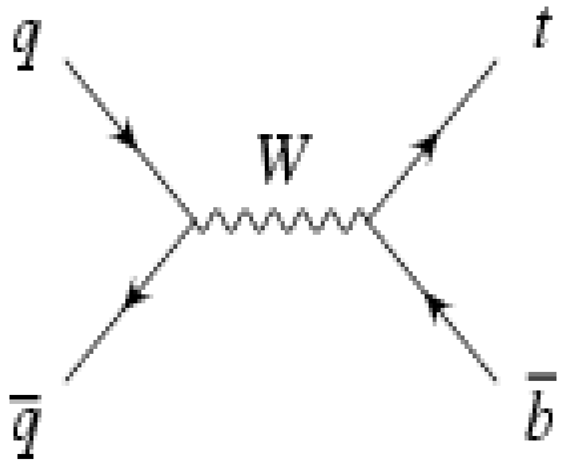
(assuming: $f_1^L = 1$)

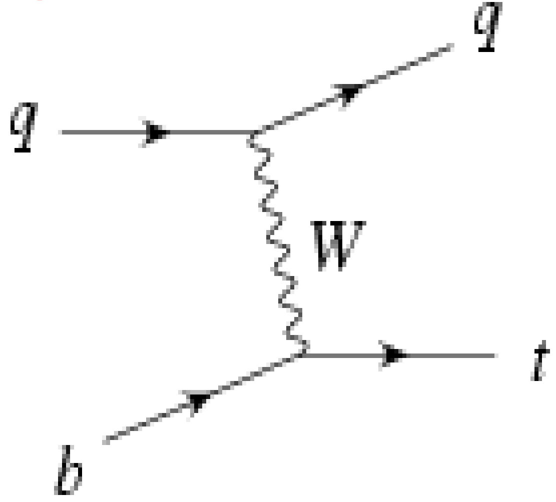
This measurement does not assume 3 generations or unitarity

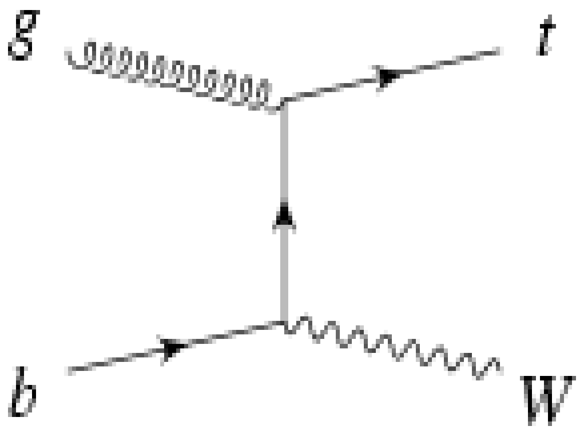
Single top prospects

- In 2008 work on the discovery, possible observation of t-channel alone
- Then the LHC will start with huge production rates:

σ_s=10.6±1.1 pb σ_t=246.6±17 pb $\sigma_{tW}=62.0^{+16.6}_{-3.6}$ pb







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

Conclusions

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

Published in PRL 98, 181802 (2007)

σ(s+t) = 4.7 ± 1.3 pb

3.6σ significance!

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

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

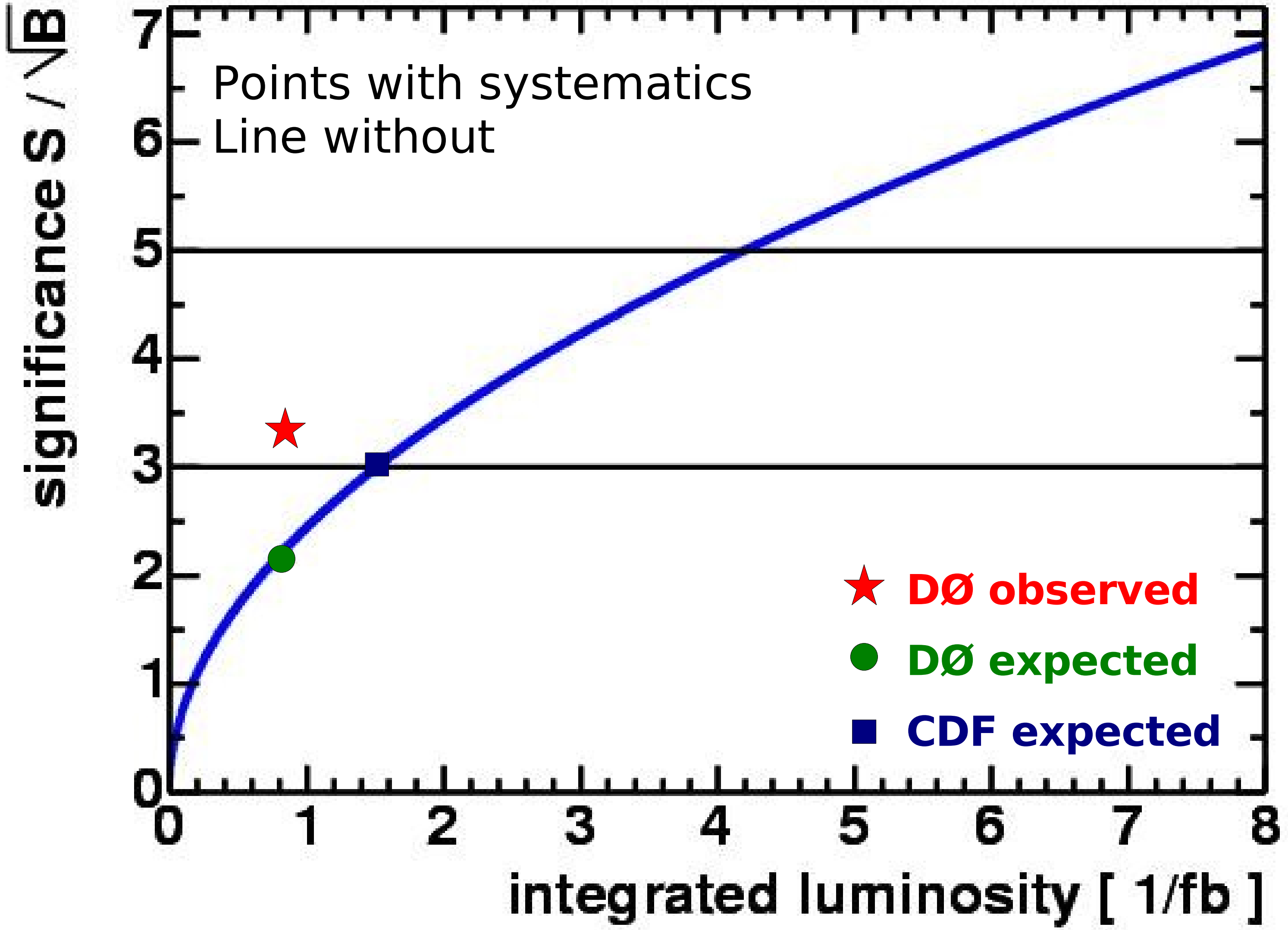
Extra slides

For more information:

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

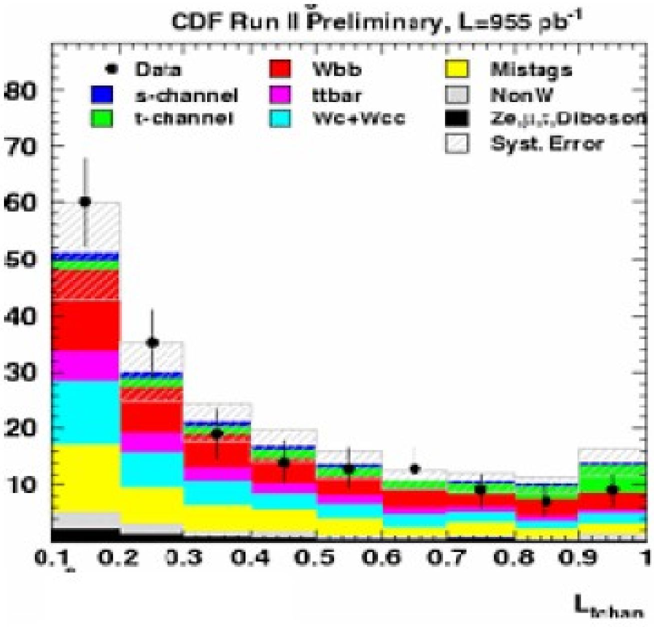
Projections for s+t

Projection by CDF for P5 in 2005



CDF's old results

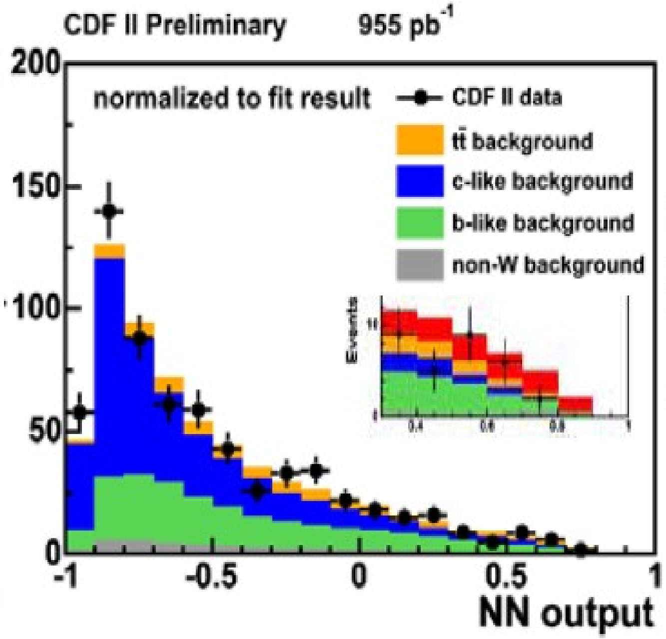
Likelihood



No evidence of signal

σ_{s+t}<2.7pb at 95% C.L.

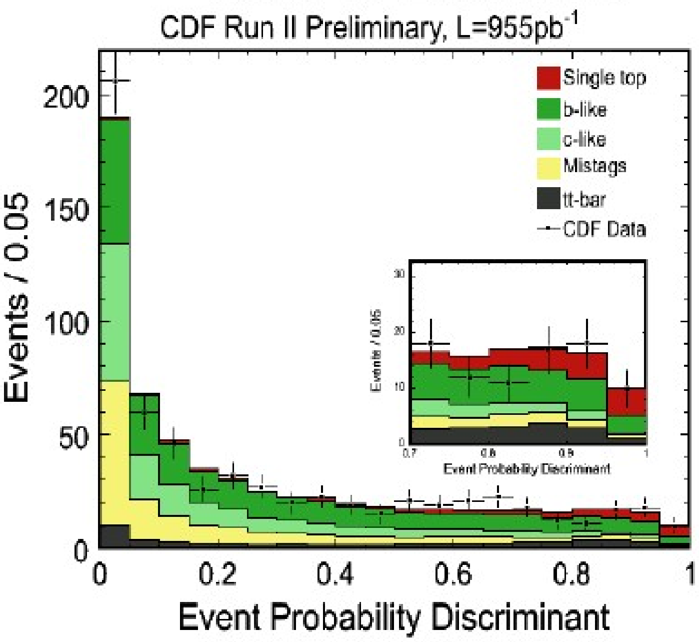
Neural Networks



No evidence of signal

σ_{s+t}<2.6pb at 95% C.L.

Matrix Element

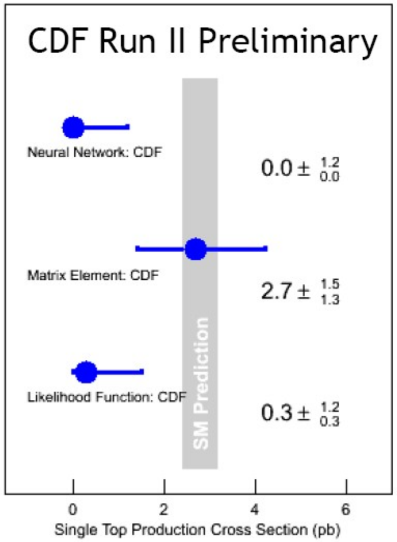


p-value = 1.0% (2.3σ)

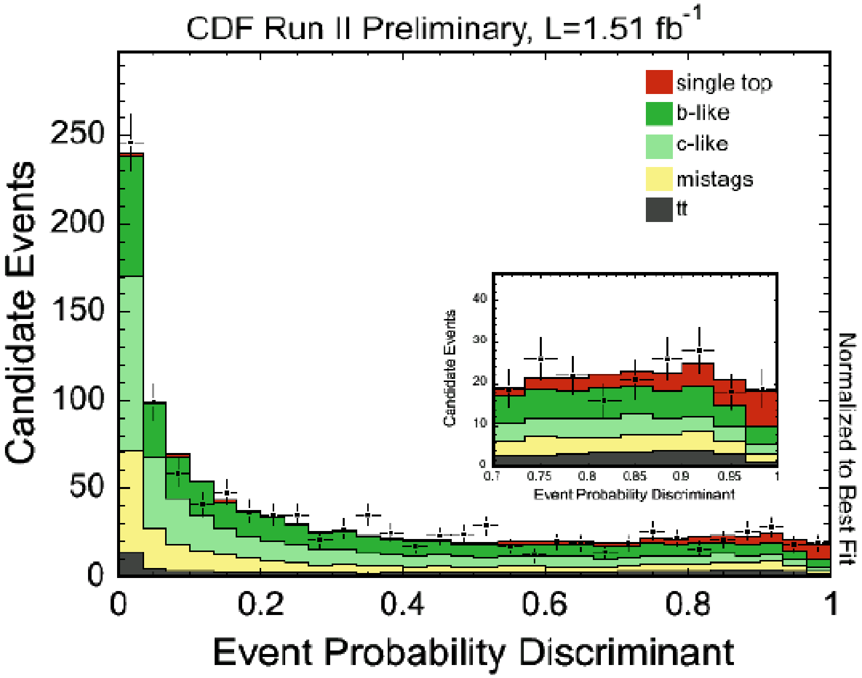
σ_{s+t}=2.7(+1.5/-1.3)pb

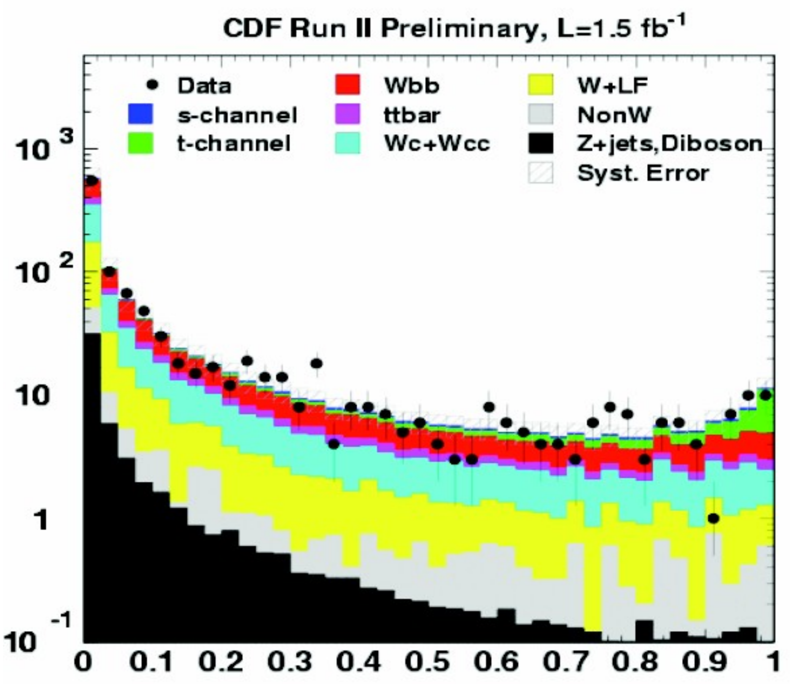
from Bernd Stelzer
@ Moriond EW

- Performed common pseudo-experiments
  - Use identical events
  - ME uses only 4-vectors of lepton, Jet1/Jet2
  - LF/NN uses sensitive event variables
  - Correlation among analyses: ~60-70%
  - 1.2% of the pseudo-experiments had an outcome as different as the one observed in data (using BLUE)



CDF's latest results





$\sigma_{s+t} = 3.0^{+1.2}_{-1.1}$pb
3.0σ expected
3.1σ observed

$\sigma_{s+t} = 2.7^{+1.3}_{-1.1}$pb
2.9σ expected
2.7σ observed

Preparing the way for the LHC

Studies at the Tevatron will help the LHC:

- Wbb measurement (will also help WH search) (DØ: hep-ex/0410062)

  Current limit at 4.6 pb for $p_T(b)>20$GeV

- In general, W+jets background determination techniques

  tt will be main background, but large uncertainties come from W+jets

  Effect of jet vetoes ($N_{jet}=2$), check other methods planned in LHC analyses

- Study charge asymmetries (Bowen, Ellis, Strassler: hep-ph/0412223)

  Signal shows asymmetry in $(Q_\ell \times \eta_j\ ,\ Q_\ell \times \eta_\ell)$ plane at TeV

- Study kinematics of forward jets in t-channel (WW→H at LHC)

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

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

TeV4LHC workshop report: 0705.3251 [hep-ph]

Crash course in Bayesian probability

Bayes' theorem expresses the degree of belief in a hypothesis A, given another B. "Conditional" probability P(A|B):

$$P(A|B)=\frac{P(B\mid A)P(A)}{P(B)}$$

In HEP: $B\rightarrow N_{\text{observed}}$, $A\rightarrow n_{\text{predicted}}=n_{\text{signal}}+n_{\text{bkgd}}$, $n_s=Acc*L*\sigma$

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

P(A): "prior" probability density $\Pi(n_{\text{pred}})=\Pi(Acc*L,n_b)\Pi(\sigma)$

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

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

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

$$P(n_{\text{predicted}}|N_{\text{observed}}) = 1/Z\ L(N_{\text{observed}}|n_{\text{predicted}})\Pi(n_{\text{pred}})$$

Signal modeling

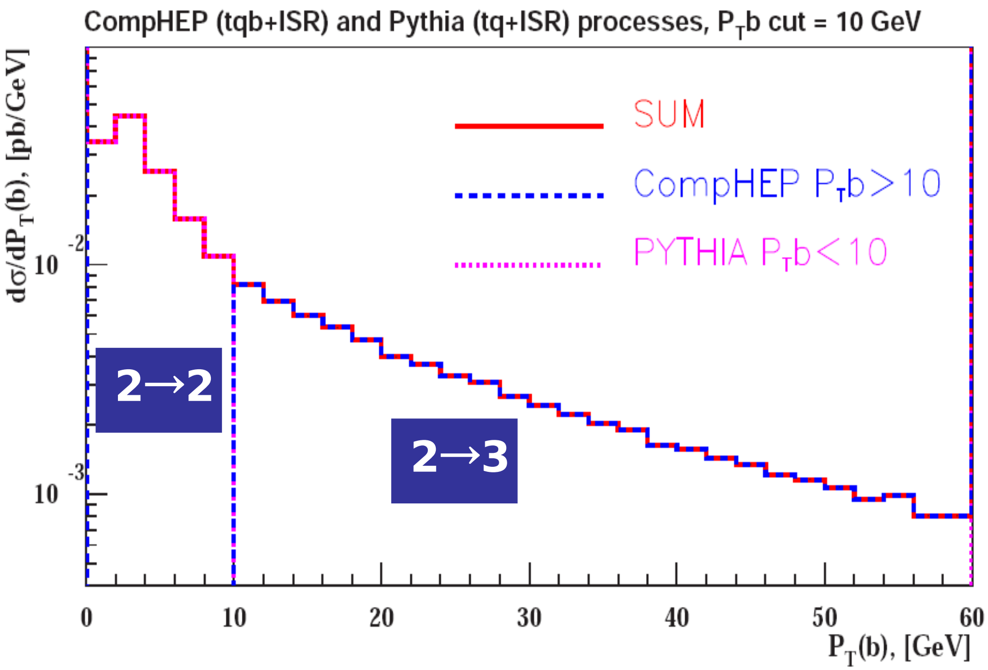
Have to get the t-channel right:
Avoid double counting when different diagrams produce same final states in different kinematic regions
Use ZTOP as NLO benchmark http://home.fnal.gov/~zack/ZTOP

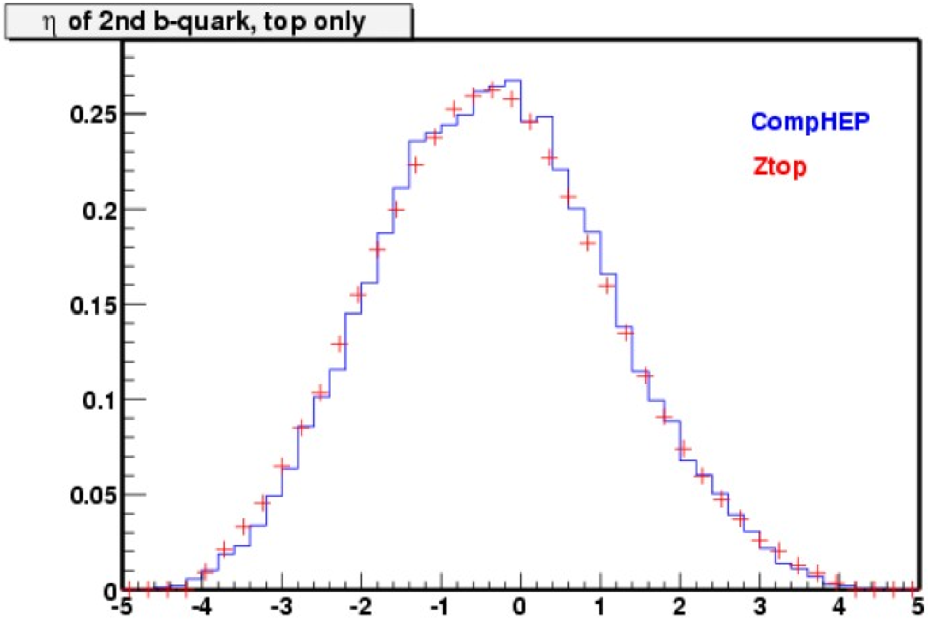
- DØ: “Effective” NLO CompHEP (also used in CMS)

Match 2→2 and 2→3 processes using b p_T for cross over, normalize to NLO

Resulting distributions agree well with ZTOP & MCFM

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





W+jets normalization

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

$$N^{loose}=N^{loose}_{fake}+N^{loose}_{real}$$

$$N^{tight}=\varepsilon_{fake}N^{loose}_{fake}+\varepsilon_{real}N^{loose}_{real}$$

Obtain: N^{loose}_{real} and N^{loose}_{fake}

- Normalize the MC Wjj and Wbb samples to the real ℓ yield found in data, after correcting for the presence of tt events:

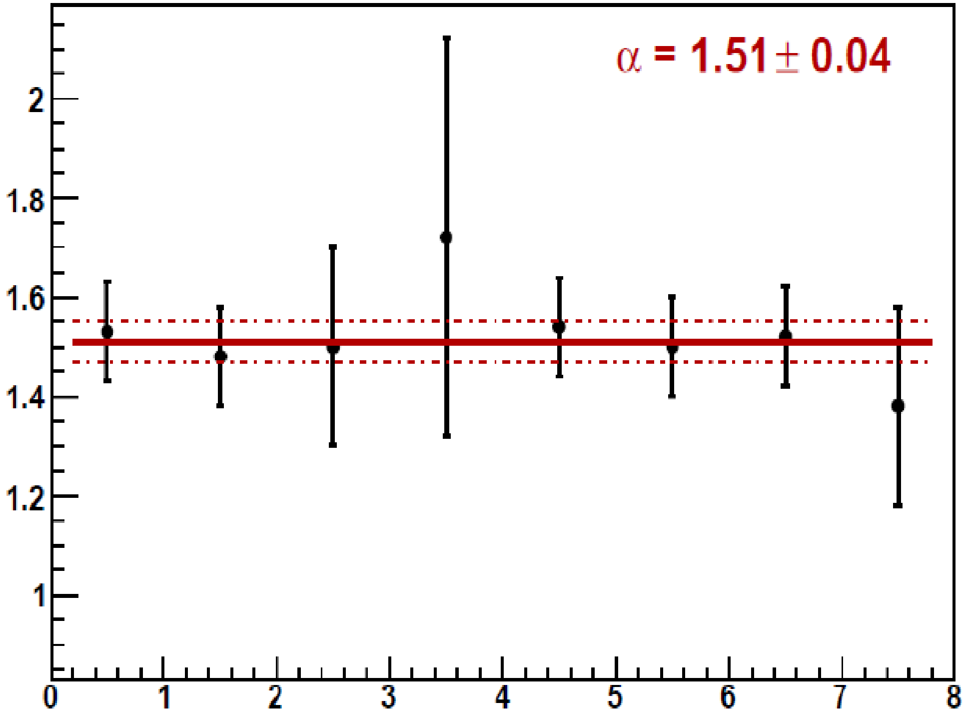
$$\varepsilon_{real}N^{loose}_{real}=SF[Y(Wjj)+Y(Wb\bar{b})+Y(Wc\bar{c})]+Y(t\bar{t})$$

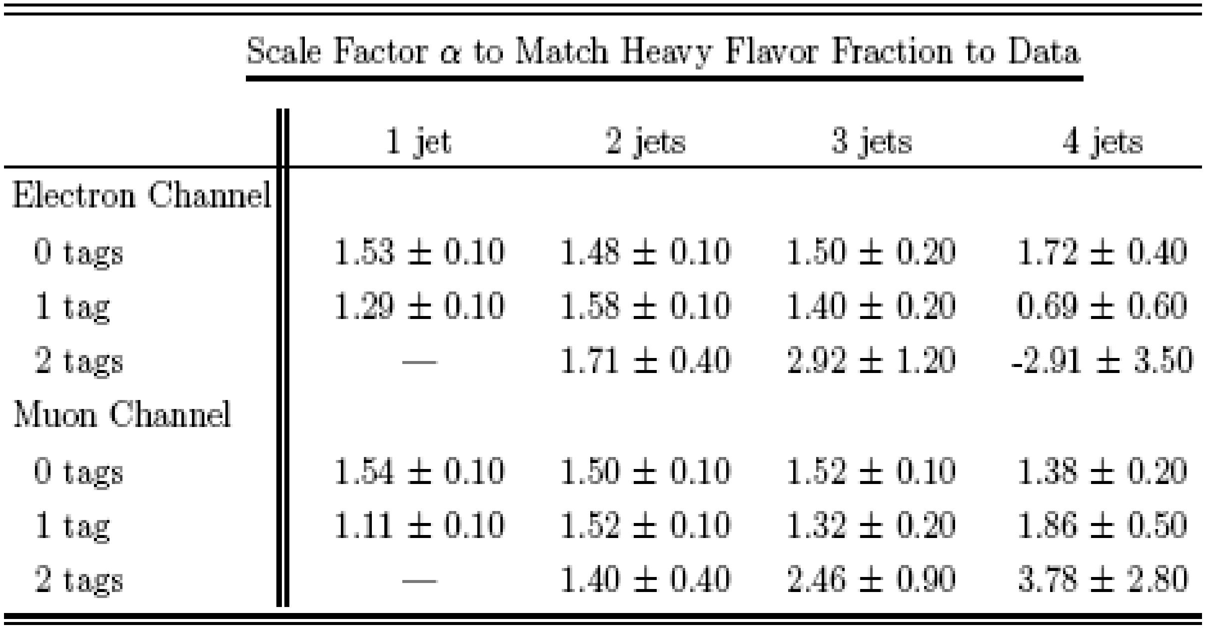
SF=1.4

- The sum Y(Wjj)+Y(Wbb)+Y(Wcc) is done according to the ratio of (Wbb+Wcc)/Wjj found in 0-tag data → 1.5±0.5
- Then apply b-tagging
  - Greatly reduce W+jets background (Wbb ~1% of Wjj)
  - Shift distributions, changes flavor composition

Wbb and Wcc fraction

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



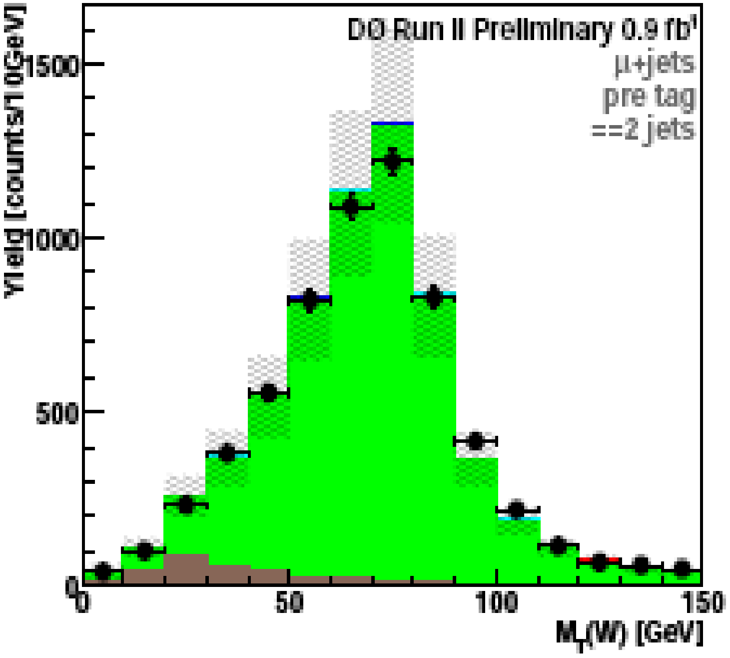


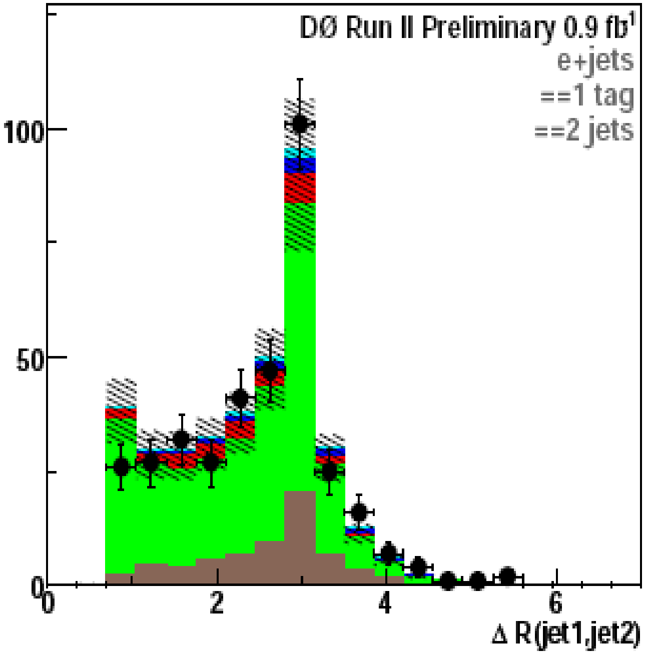
Scale Factor α to Match Heavy Flavor Fraction to Data

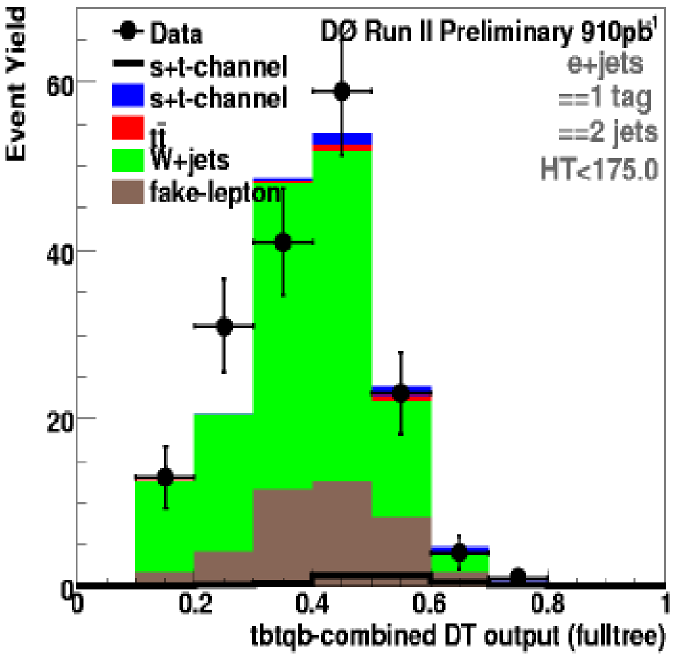
	1 jet	2 jets	3 jets	4 jets
Electron Channel				
0 tags	1.53 ± 0.10	1.48 ± 0.10	1.50 ± 0.20	1.72 ± 0.40
1 tag	1.29 ± 0.10	1.58 ± 0.10	1.40 ± 0.20	0.69 ± 0.60
2 tags	—	1.71 ± 0.40	2.92 ± 1.20	-2.91 ± 3.50
Muon Channel				
0 tags	1.54 ± 0.10	1.50 ± 0.10	1.52 ± 0.10	1.38 ± 0.20
1 tag	1.11 ± 0.10	1.52 ± 0.10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

What about shapes?

- NLO shapes for Wbb are different from Alpgen (LO)
- Specially at low b-jet p_T (<25GeV) and m_{bb} (<25GeV & >80GeV)
  - Until we have a data-based method to extract Wbb or a pT dependent k-factor from MC, we are stuck with a constant
  - Let the data judge. We have found overall good agreement in all kinds of distributions inside our acceptance before and after tagging: angular correlations, pTs, background cross check samples, discriminant outputs…

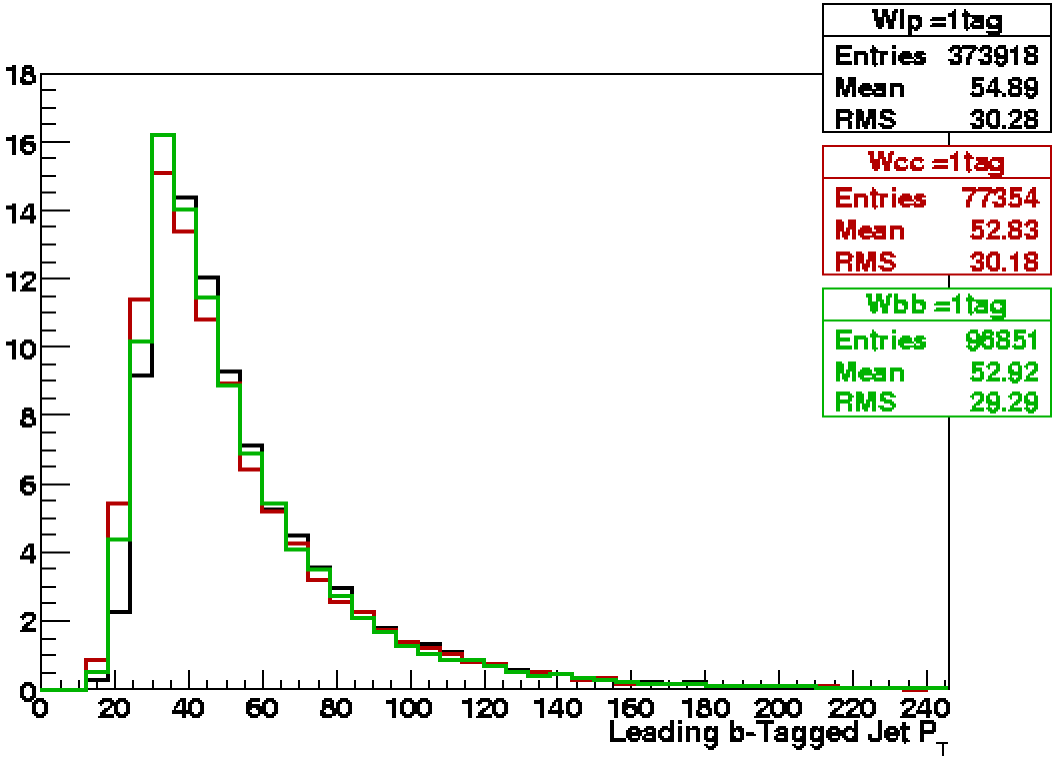


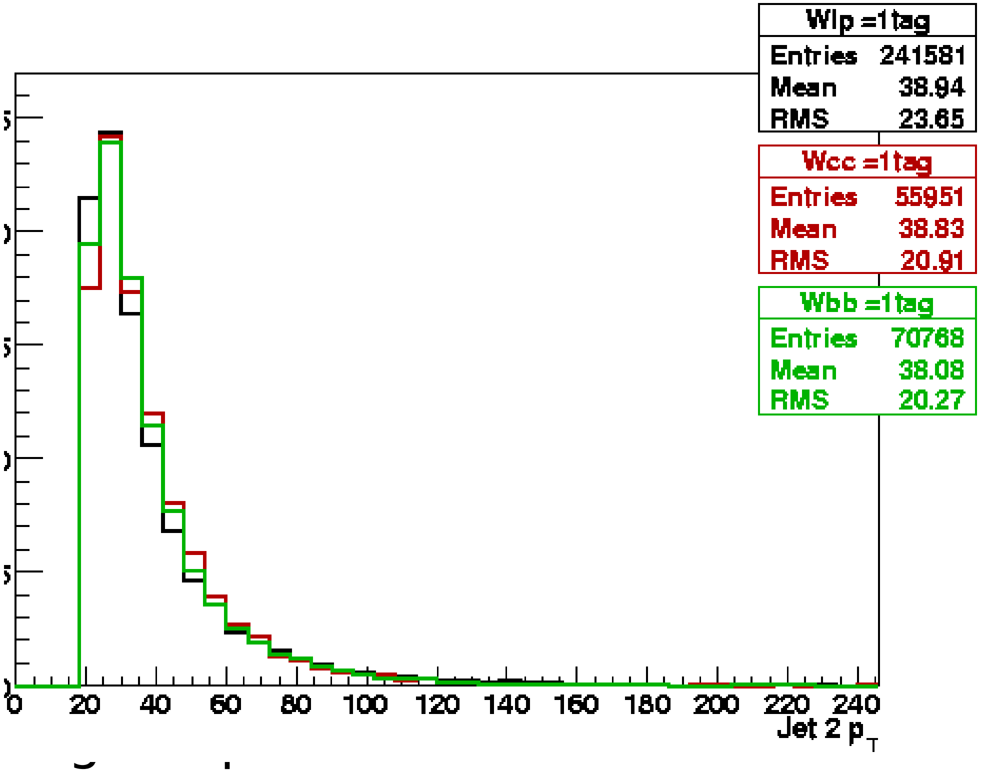




Wbb/Wcc shape difference

- Can you assume that Wbb and Wcc fractions separately can be described by the Wbb+Wcc fraction?
  - We changed the Wbb/Wcc ratio by ±10% and re-calculated the single top cross section:
  - More Wbb, less Wcc: σ(tb+tqb)=4.85±1.4pb
  - Less Wbb, more Wcc: σ(tb+tqb)=4.98±1.5pb
  - Weak dependence based on similarity between Wbb and Wcc shapes





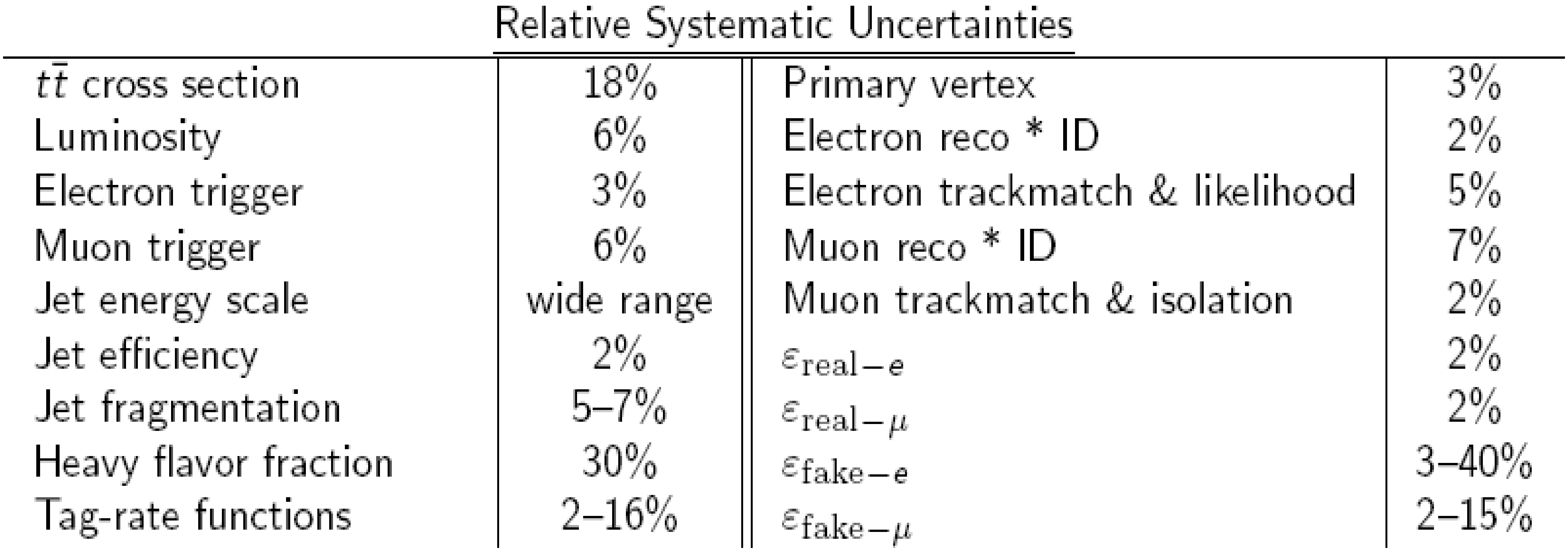
Error on the HF fraction

- How come a 30% error on HF fraction doesn't destroy all sensitivity?
  - This (still) is a statistics limited analysis: 1.2pb out of 1.4pb error comes from stats alone
  - The 30% error (1.5±0.45) covers shape differences in the NLO distributions and between Wbb and Wcc
  - After tagging, the uncertainty on the total W+jets yield is reduced from 30% because:

    a) Not the entire sample is Wbb+Wcc, the uncertainty on the sum is smaller than 30%

    b) The anti-correlation between Wjj and Wbb+Wcc due to the normalization before tagging further reduces the uncertainty
  - This uncertainty is still the largest flat systematic in the end

Systematics

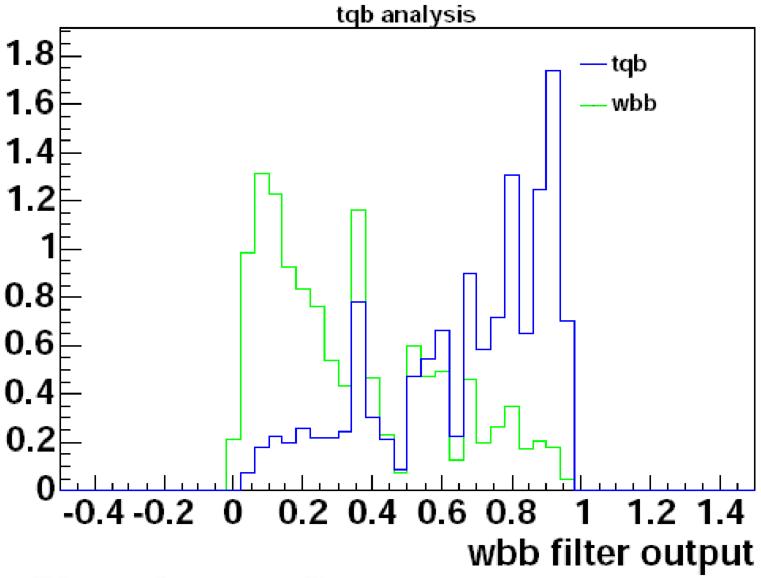


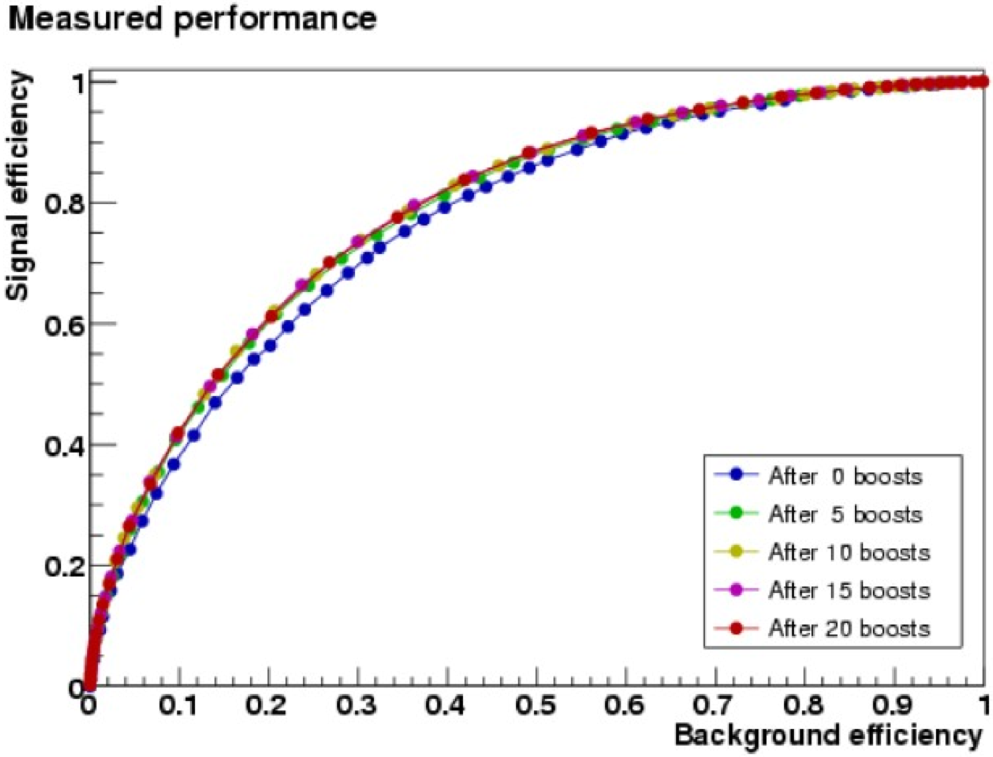
Relative Systematic Uncertainties

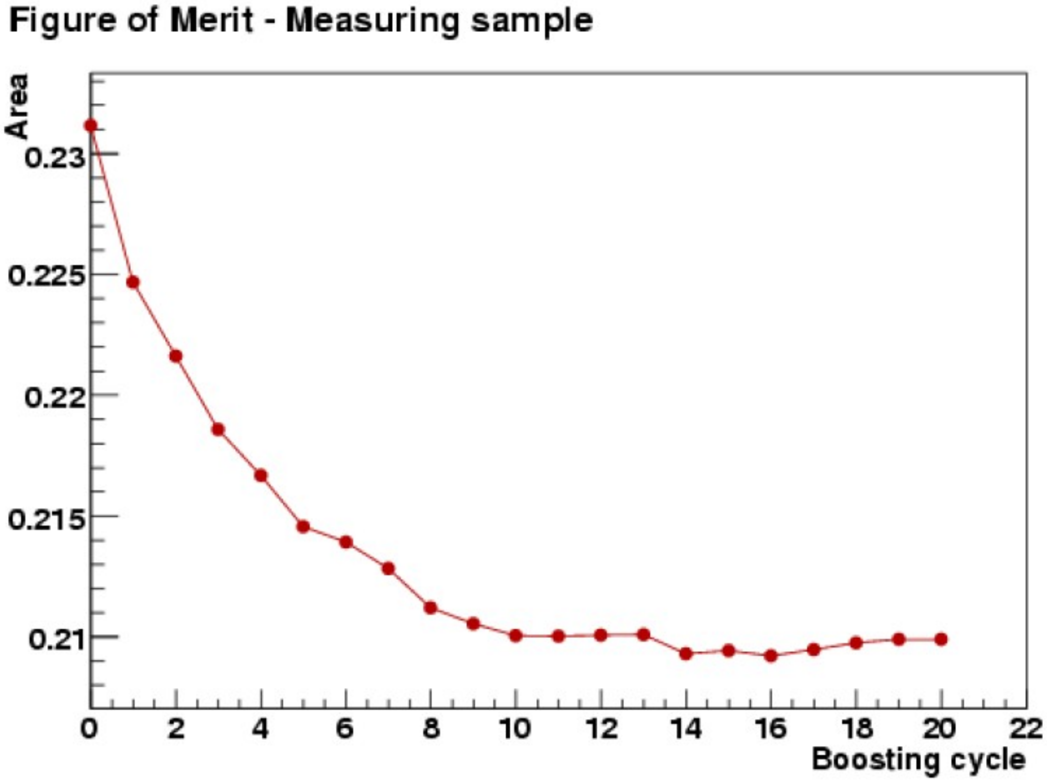
$t\bar{t}$ cross section	18%	Primary vertex	3%
Luminosity	6%	Electron reco * ID	2%
Electron trigger	3%	Electron trackmatch & likelihood	5%
Muon trigger	6%	Muon reco * ID	7%
Jet energy scale	wide range	Muon trackmatch & isolation	2%
Jet efficiency	2%	$\varepsilon_{\mathrm{real}-e}$	2%
Jet fragmentation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

Boosting

- Single trees can have spikes, even with enough statistics of training events
- We use the weighted sum of 20 trees
  - Smoother distributions
  - Better separation
  - More stability







Correlations

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

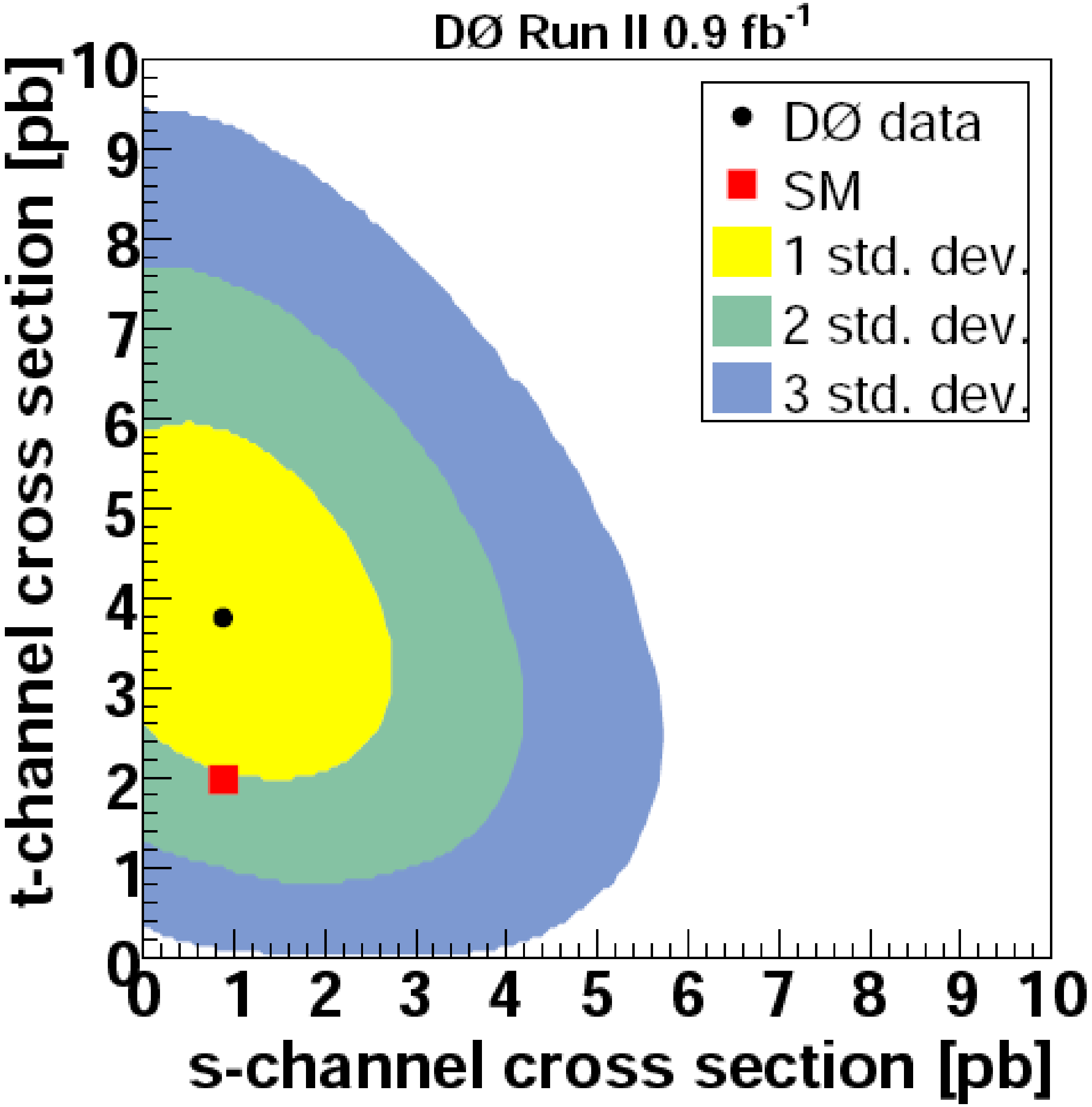
Technique	Electron	Muon
DT vs ME	52%	58%
DT vs BNN	56%	48%
ME vs BNN	46%	52%

- Calculate the linear correlation between the measured cross sections in the same 2000 members of the SM ensemble

	DT	ME	BNN
DT	100%	57%	51%
ME		100%	45%
BNN			100%

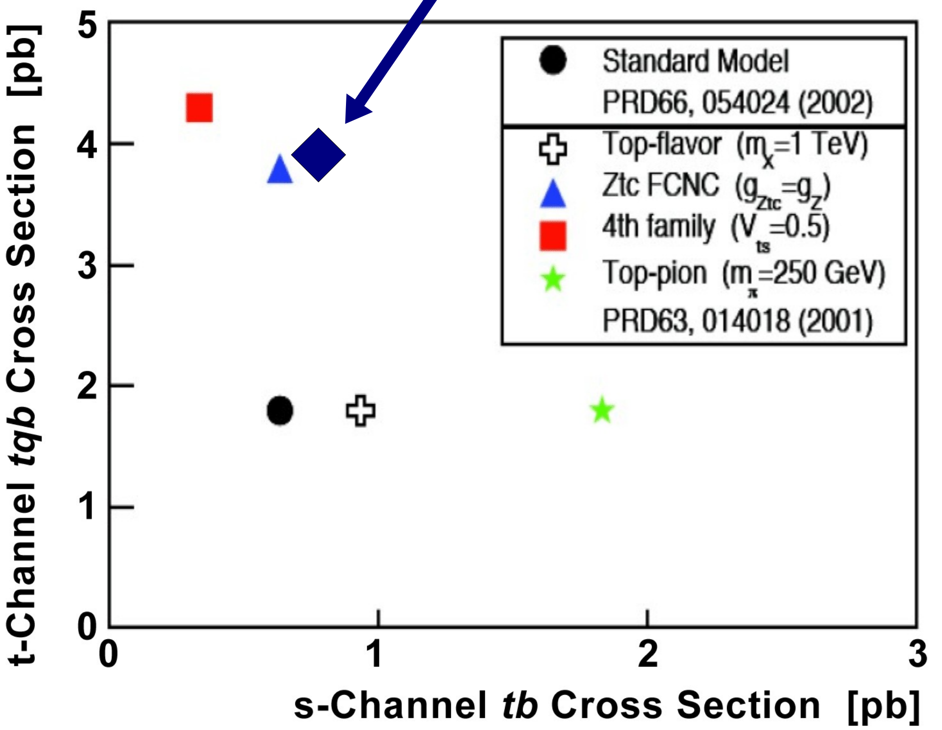
tb and tqb separately

- Remove the constraint of SM s:t ratio
- Measure model independent s- and t-channel cross sections



$\sigma(tqb) = 3.8^{+1.8}_{-1.4}$ pb

$\sigma(tb) = 0.9 \pm 0.9$ pb



Expected p-values and σ

Decision Trees

p-value 1.9%

exp. sig. 2.1σ

Bayesian NN

p-value 1.6%

exp. sig. 2.2σ

Matrix Elements

p-value 3.1%

exp. sig. 1.9σ