Evidence for Single Top Quark Production by the Dzero Experiment

Meenakshi Narain

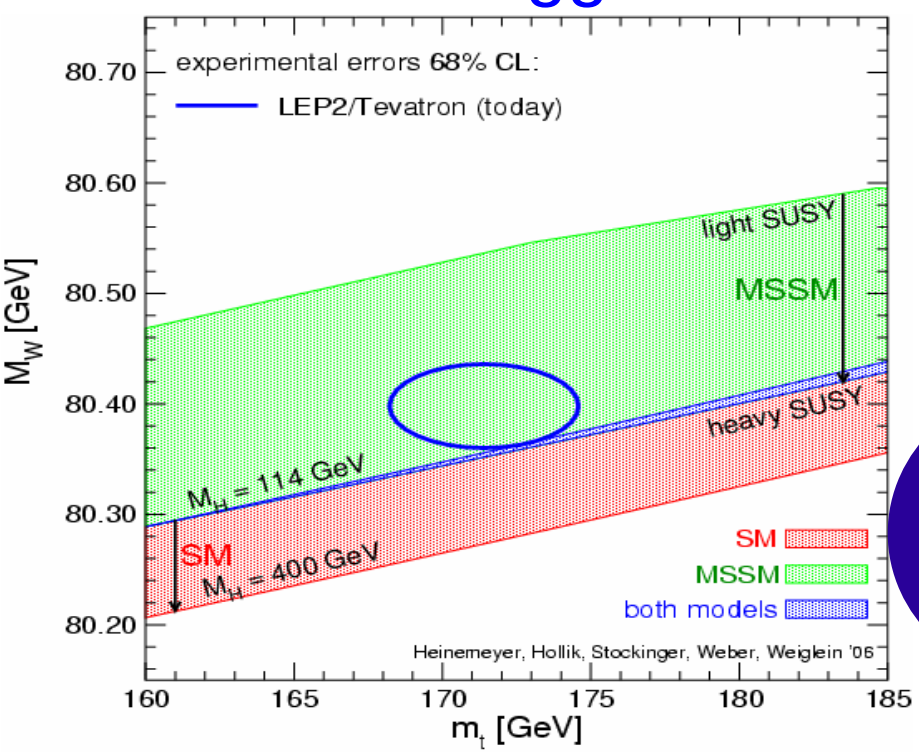
Brown University

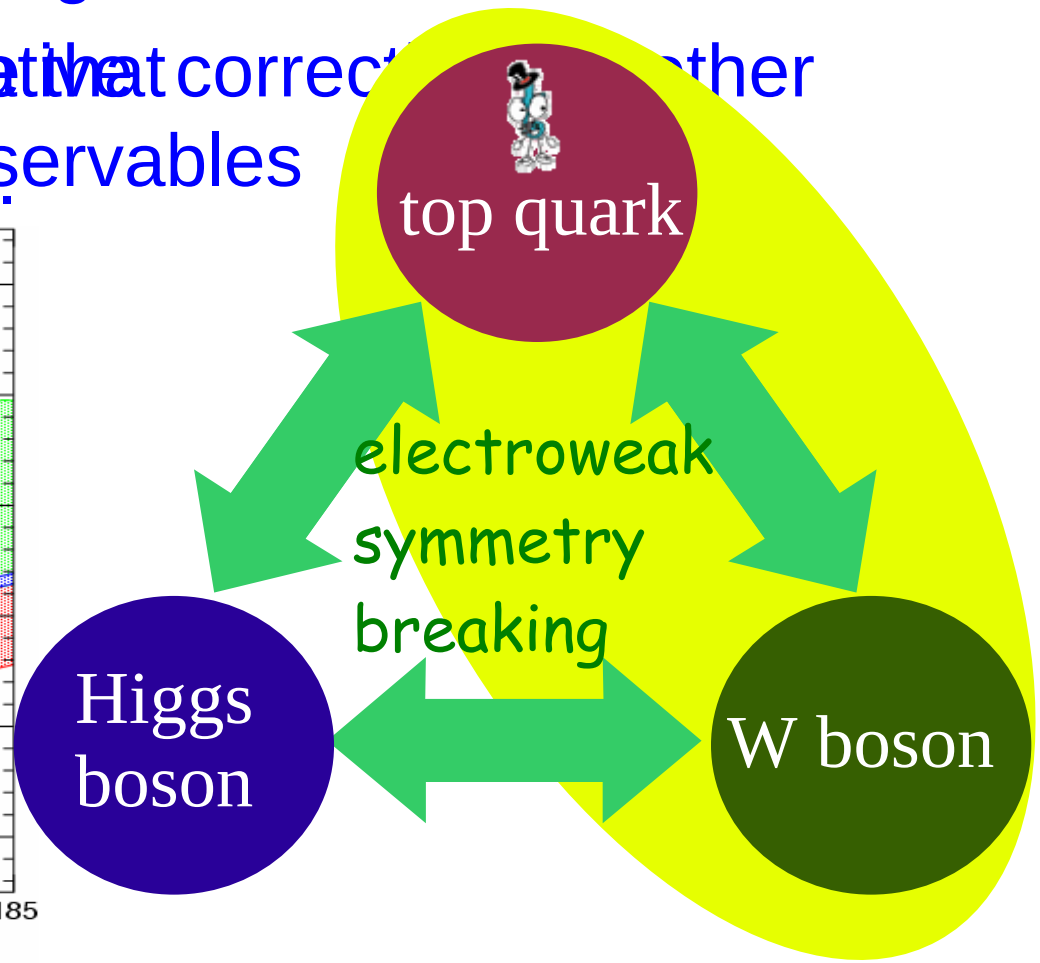
Top Quark Production at the Tevatron

- Discovered more than a decade ago (1995)
- Tevatron is still the only place to produce them!
- Top quarks probe both QCD and Electroweak interactions
  - heaviest known elementary particle
    - It probes physics at much higher energy scale than the other fermions.
  - decays before it hadronizes
    - momentum and spin information is passed to its decay products.
    - No hadron spectroscopy
  - provides unique test of QCD

Top Quark Production at the Tevatron

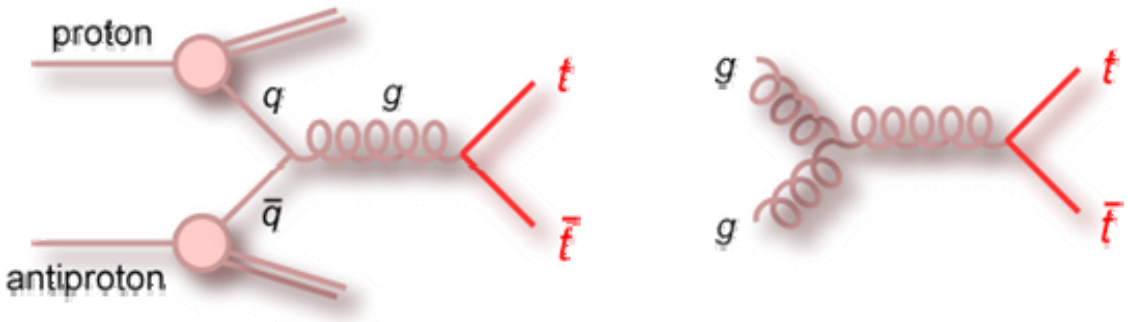
- Constrains mass of the SM Higgs
  - M_{top} is a parameter in the calculations of radiative corrections to other Standard Model observables
  - M_{top} also enters as a parameter in the mass of the Higgs boson.



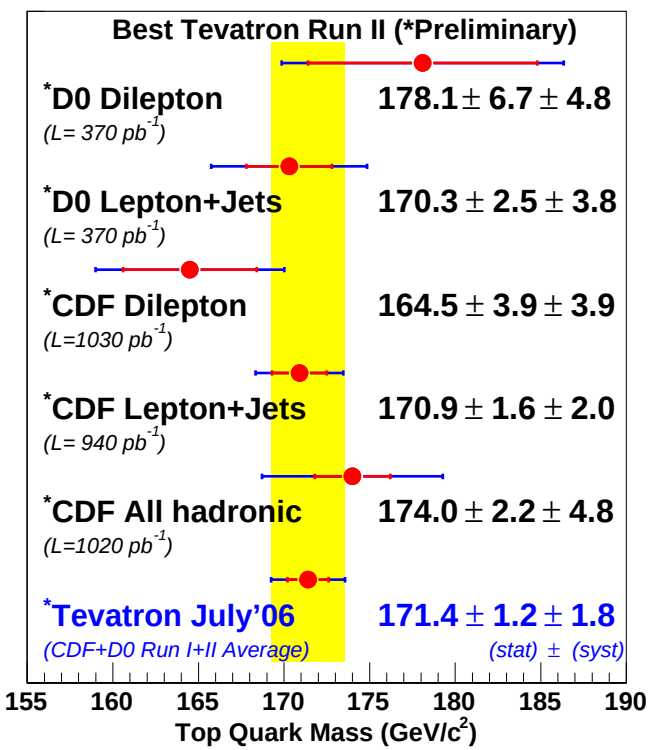


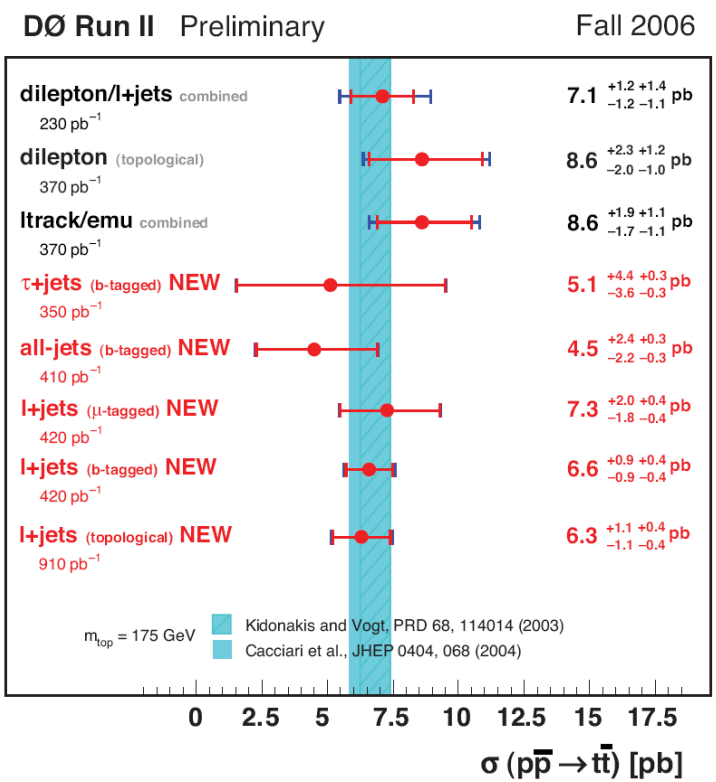
Top Quark Pair Production

- Strong Interaction:



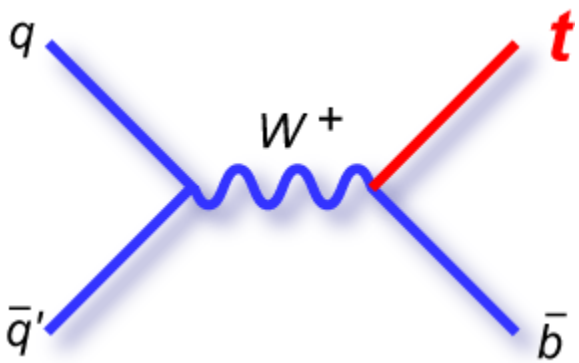
- Various properties have been studied:





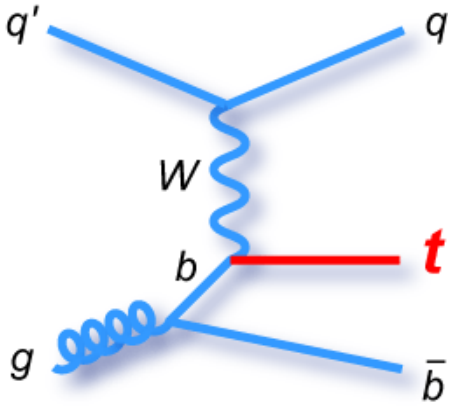
Electroweak Top Quark Production

s-channel (tb)



σ_{NLO}=0.88± 0.11 pb

t-channel (tqb)



σ_{NLO}=1.98± 0.25 pb

- Status until Nov'06:

Dzero Limits

Methodology	Integrated Luminosity	s-channel	t-channel
Cut-based	230 pb^{-1}	10.6 pb	11.3 pb
Neural Network	230 pb^{-1}	6.4 pb	5.0 pb
Likelihood	370 pb^{-1}	5.0 pb	4.4 pb

CDF limits

Methodology	Integrated Luminosity	s+t-channel
Neural Network	695 pb^{-1}	$\sigma < 3.4$ pb @ 95% CL
Likelihood	955 pb^{-1}	$\sigma < 2.7$ pb @ 95% CL
Matrix Element	955 pb^{-1}	$\sigma = 2.7^{+1.5}_{-1.3}$ pb

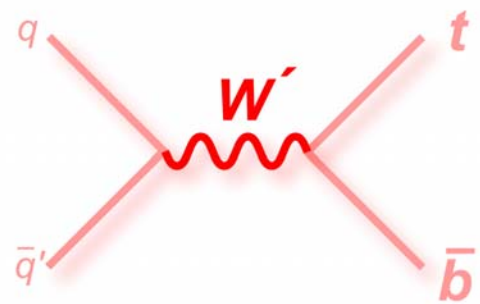
NLO cross sections at m(top) = 175 GeV, Phys.Rev. D70 (2004) 114012

Why are we interested in Single Top?

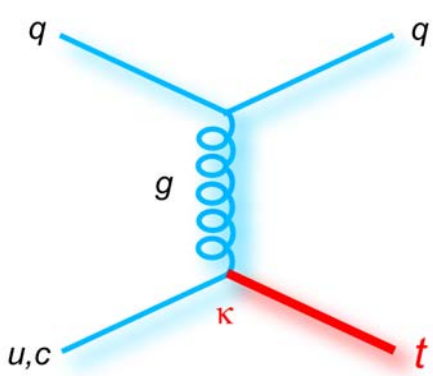
- Test of the SM prediction. Does it exist?
  - Cross section $\propto |V_{tb}|^2$
    - Test unitarity of the CKM matrix, .e.g. $V_{ub}^2 + V_{cb}^2 + V_{tb}^2 \stackrel{?}{=} 1$
      Hints for existence of a 4th generation ?
- Measure the partial width $\Gamma(t \to W\ b)$
  - and hence the top quark lifetime
- Source of polarized top quarks.
  - Spin correlations measurable in decay products.
- Same final state signature as Higgs:
  - WH, $H \to$ bbbar.
    - Understanding single-top backgrounds is prerequisite for Higgs searches at the Tevatron.

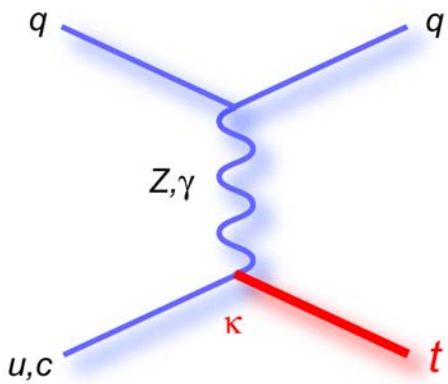
Why are we interested in Single Top Production?

- Test non-SM phenomena
  - Search W' or H+ (s-channel signature)



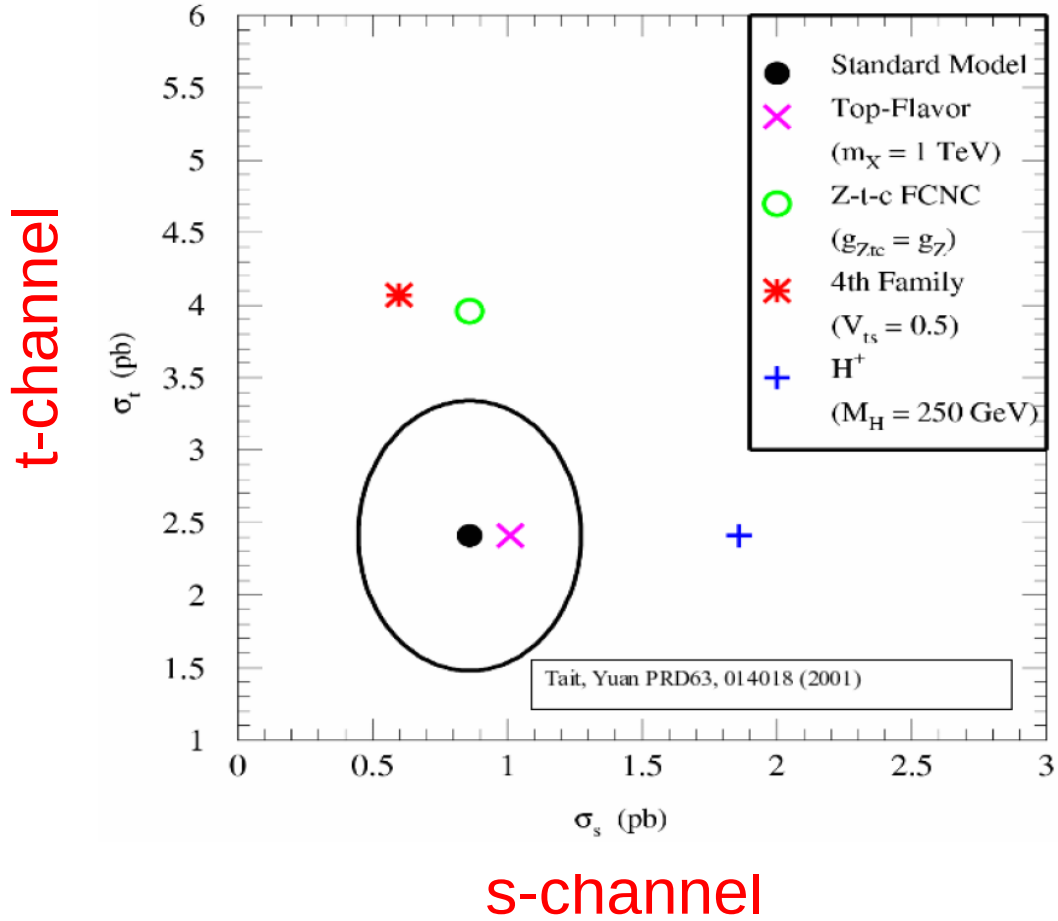
  - Search for FCNC, e.g. ug → t





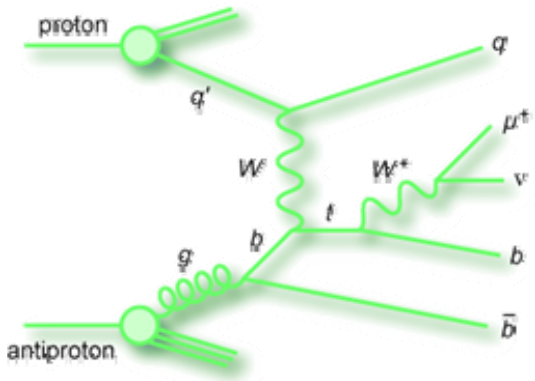
Why are we interested in Single Top Production?

- Test non-SM phenomena

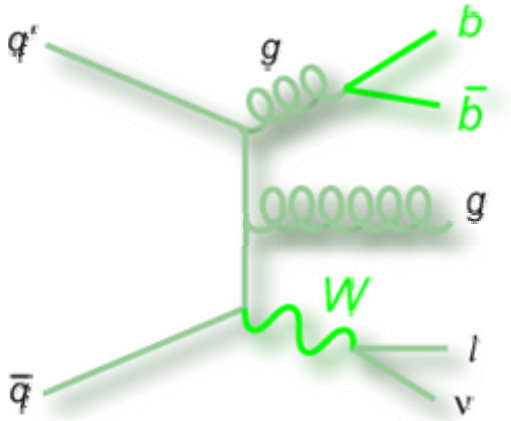


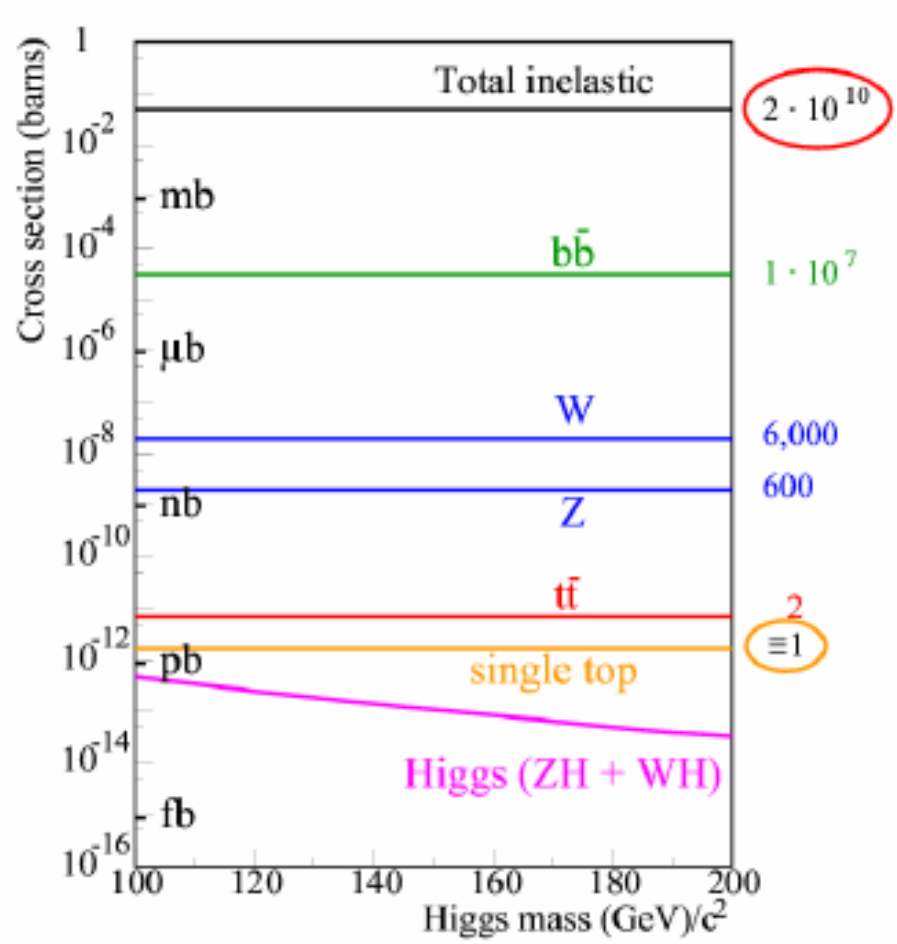
Fishing out the signal is a challenge…

Single top



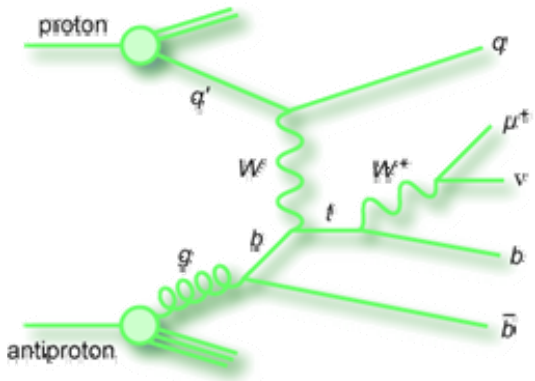
Dominant Background: Wbb, Wjets

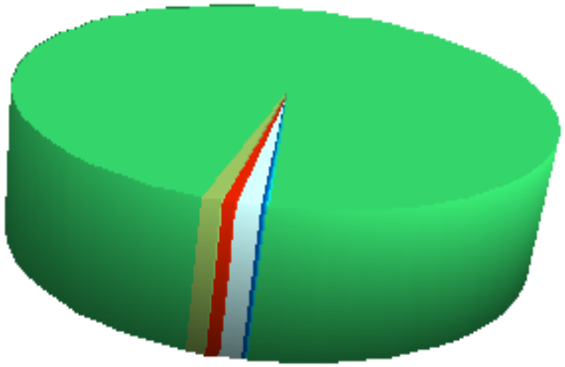




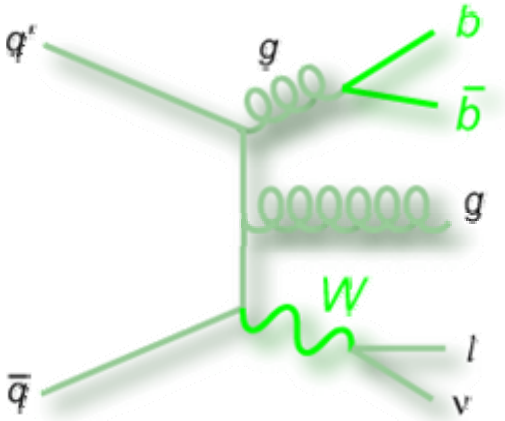
Fishing the signal is a challenge…

Single top





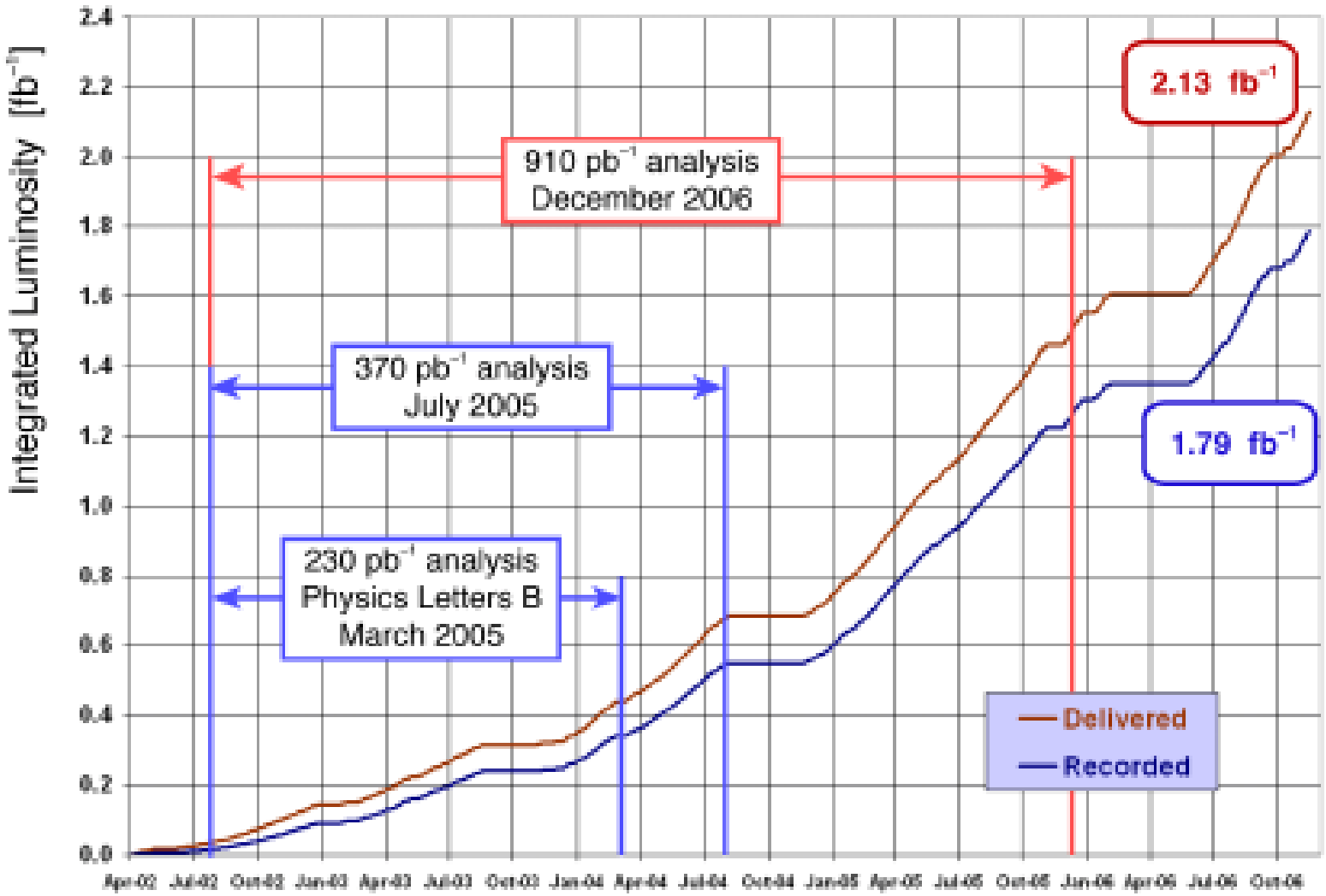
Dominant Background: Wbb, Wjets



- *W+Jets (σ~1000pb)*
- *Wbb (σ~10pb)*
- *ttbar (σ~7pb)*
- *Dibosons (σ~13pb)*
- *t-channel (σ~1.98pb)*
- *s-channel (σ~0.88pb)*

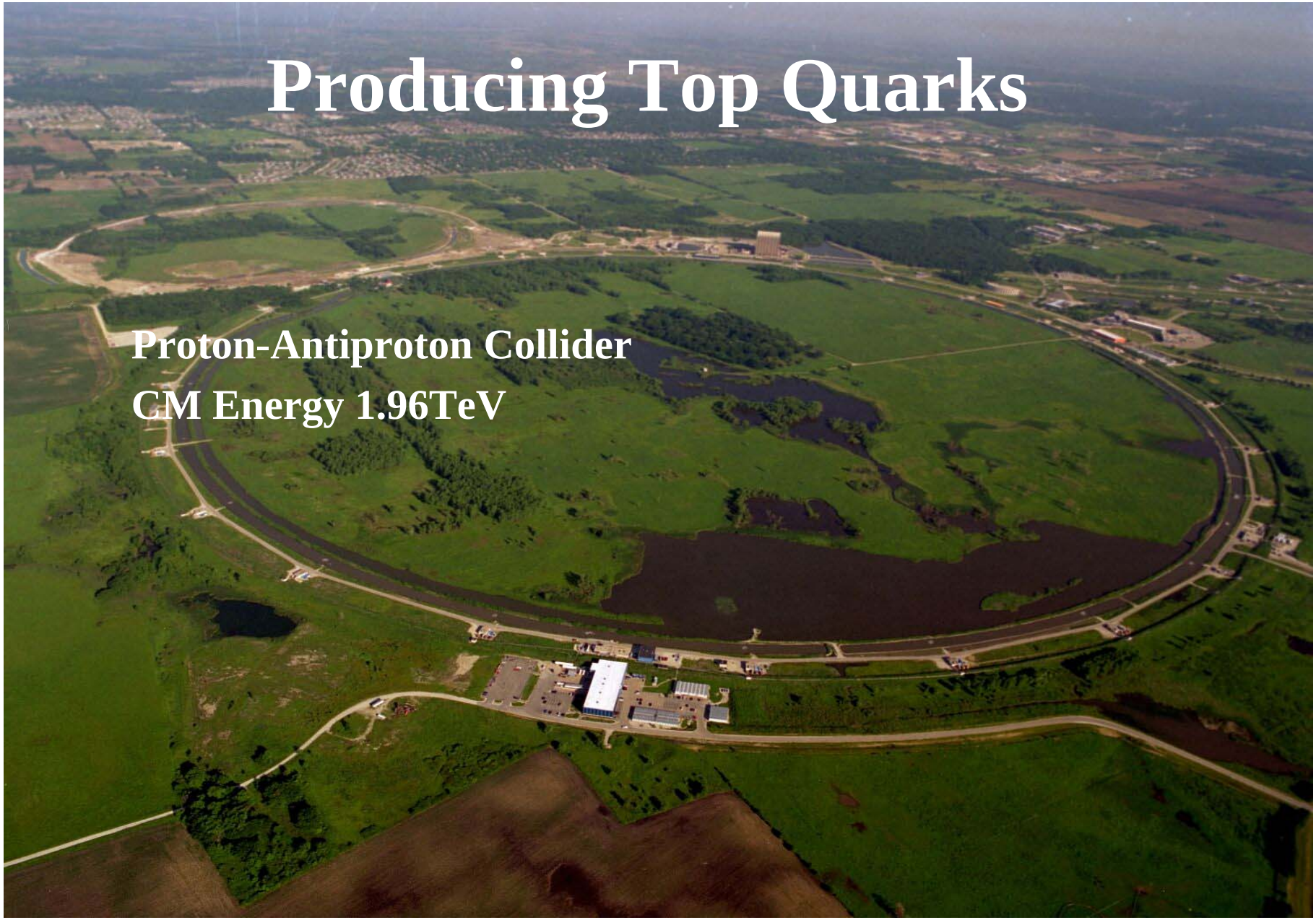
Run II Integrated Luminosity

Apr 2002 – Dec 2006



Many thanks to the Accelerator Division

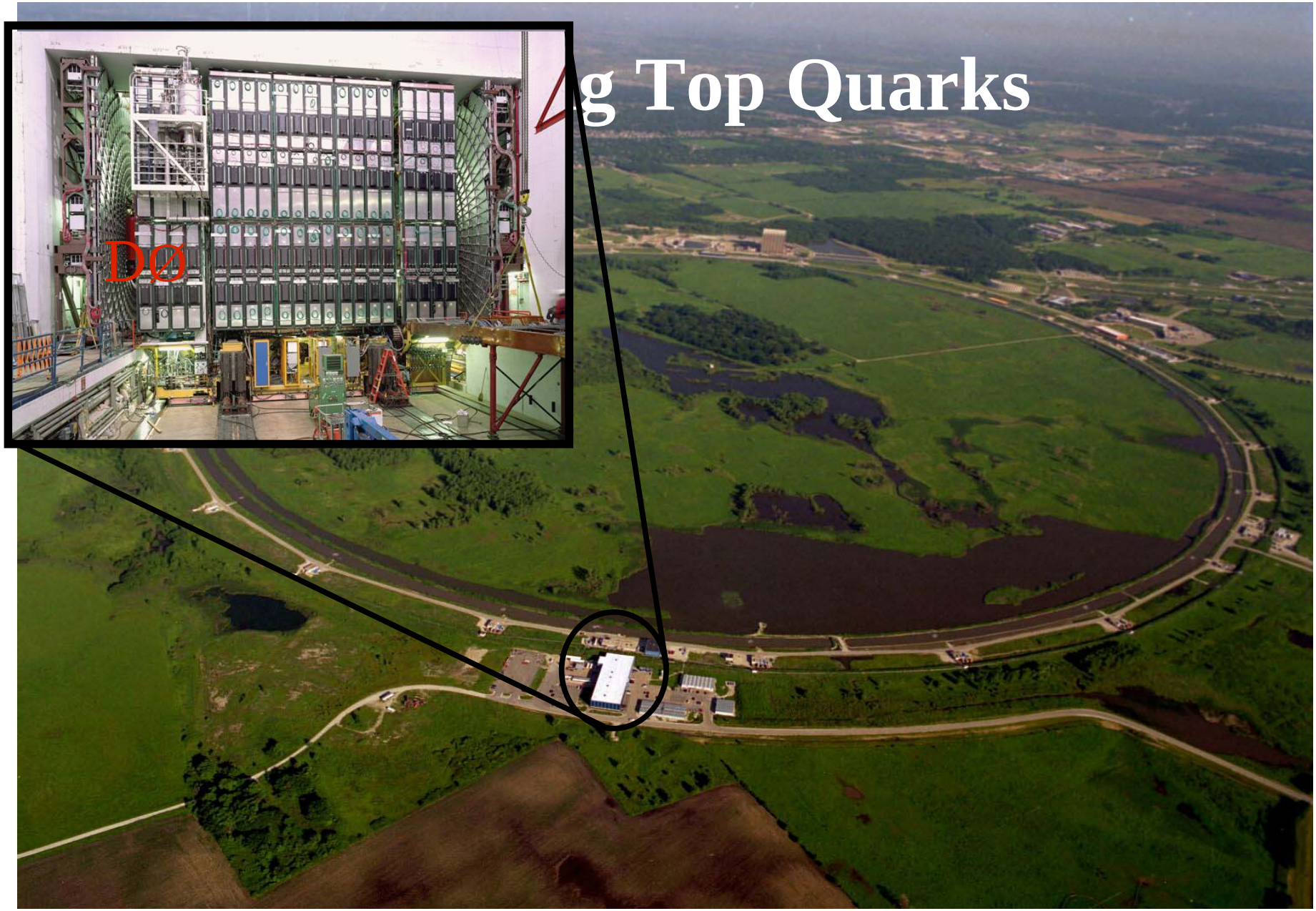
Producing Top Quarks



Proton-Antiproton Collider

CM Energy 1.96TeV

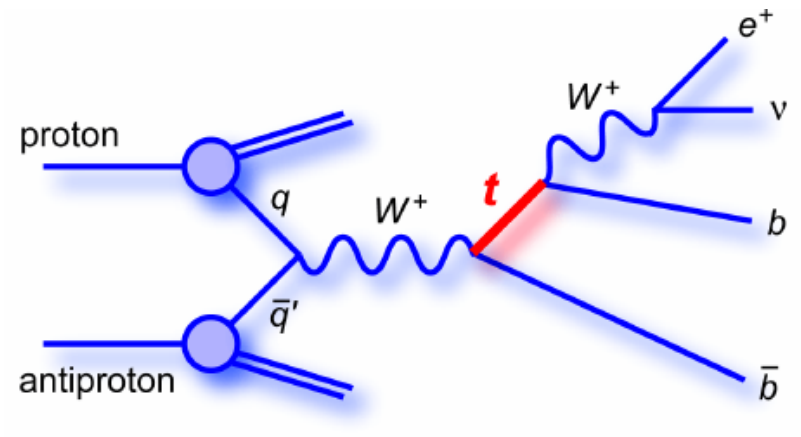
g Top Quarks

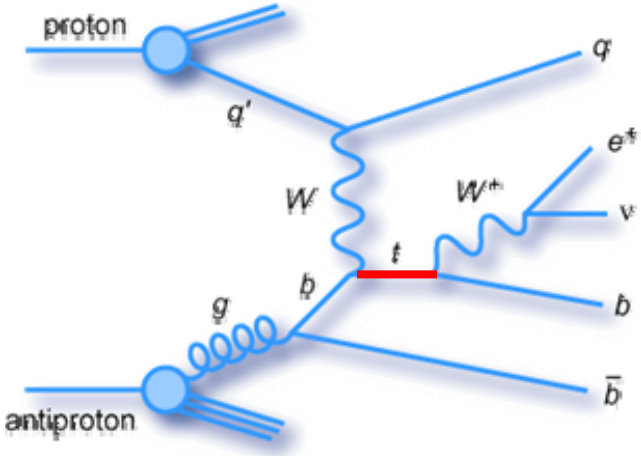


Dzero Collaboration

- 19 countries
- 80 institutions
- 700 physicists

Event Selection



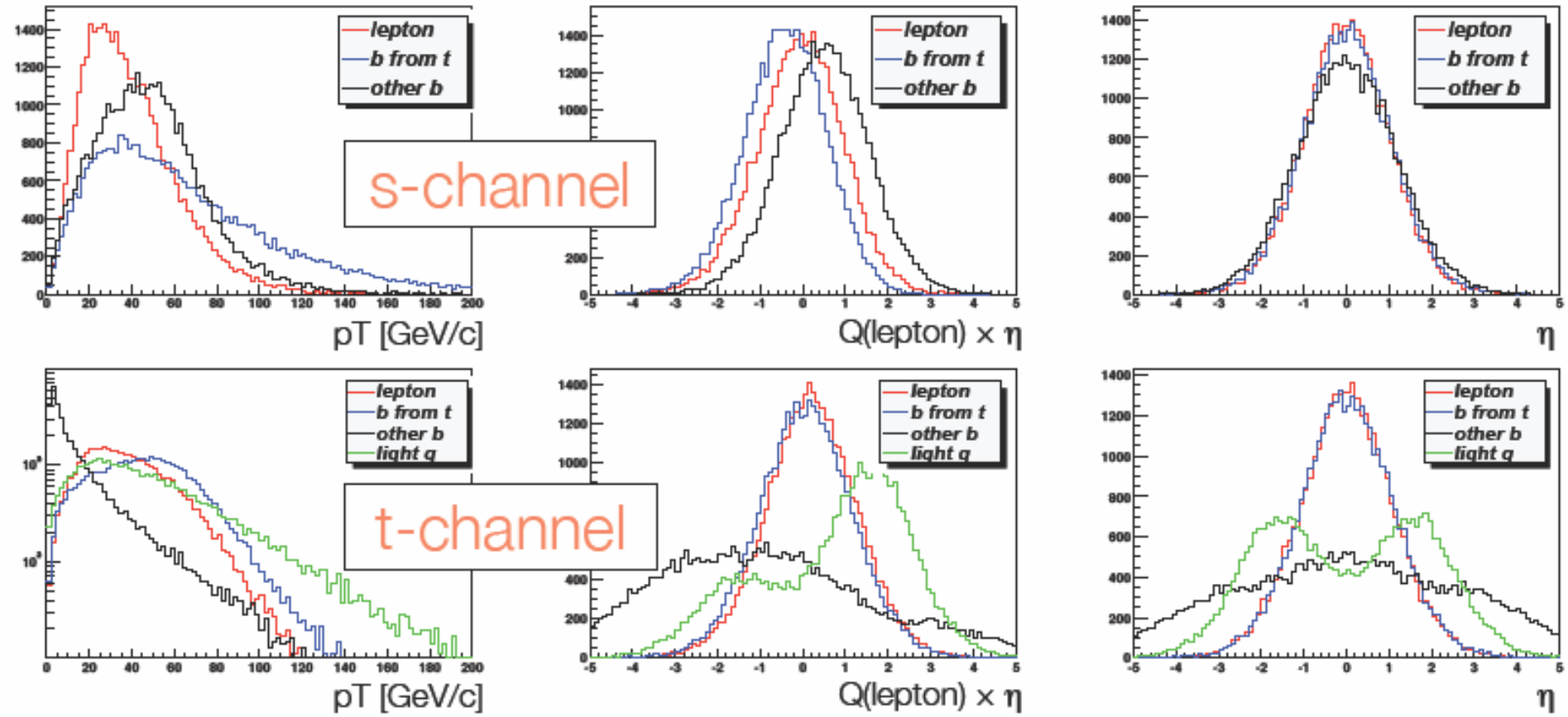


- Final State Signature:
  - charged lepton
  - neutrino (missing ET)
  - bottom quark from the top quark
  - in s-channel, the other bottom quark
  - in t-channel, the bottom & light quark
    - This one is often quite forward.
    - the second bottom quark from gluon splitting is usually not observed.

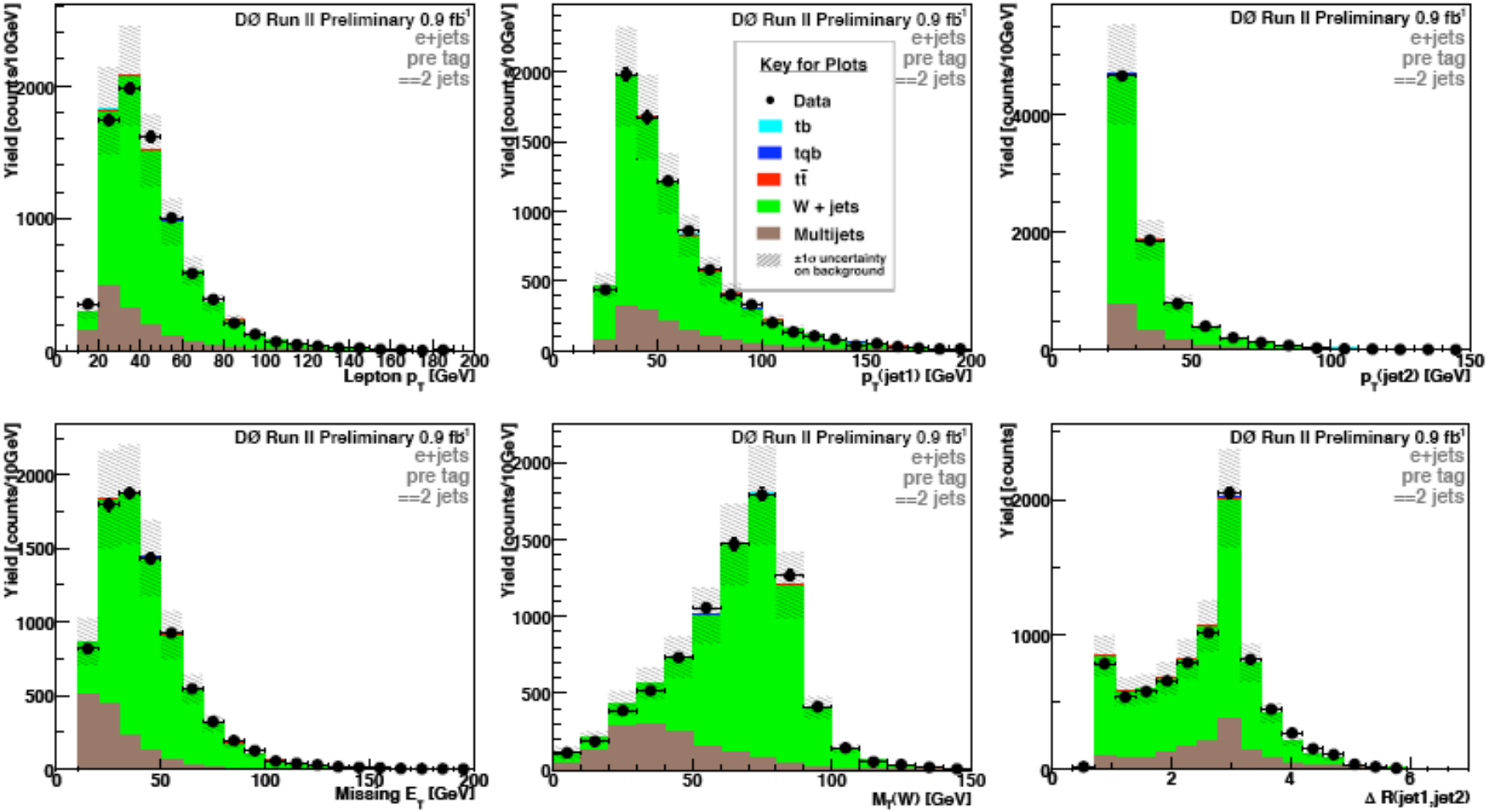
2-4 jets
At least one b-jet

Some distributions

Parton Level

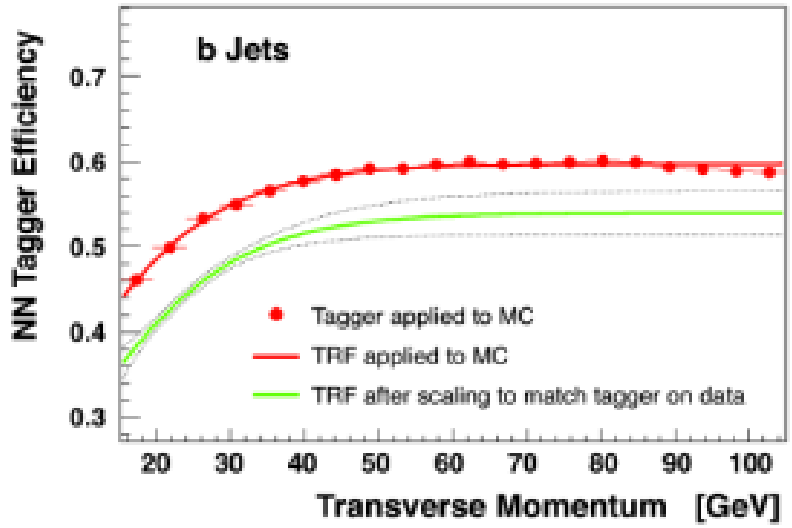


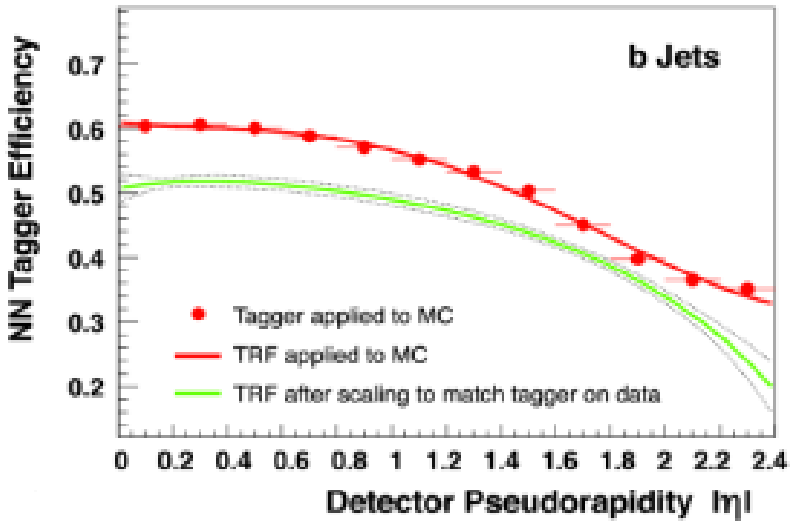
Lepton+2jets:
Data vs MC (no b-tagging)



- **Good description of data.**
- **About 90 variables checked (different jet multiplicities)**

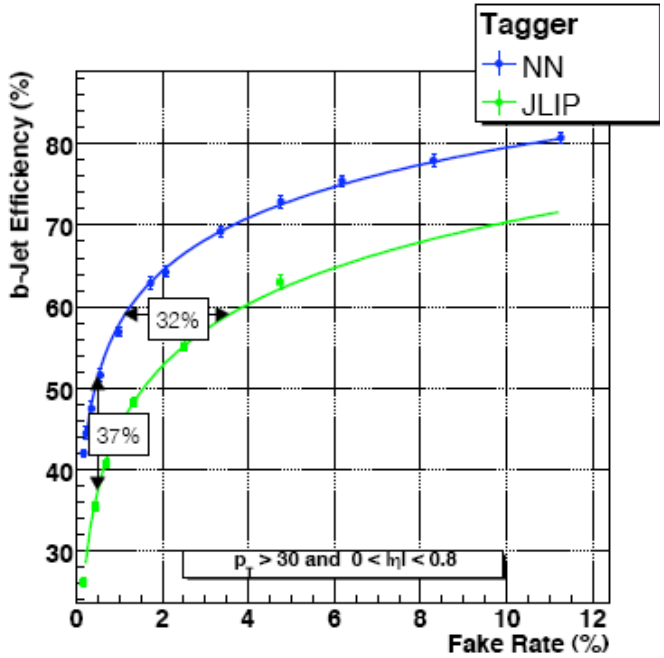
Neural Network b-jet tagging



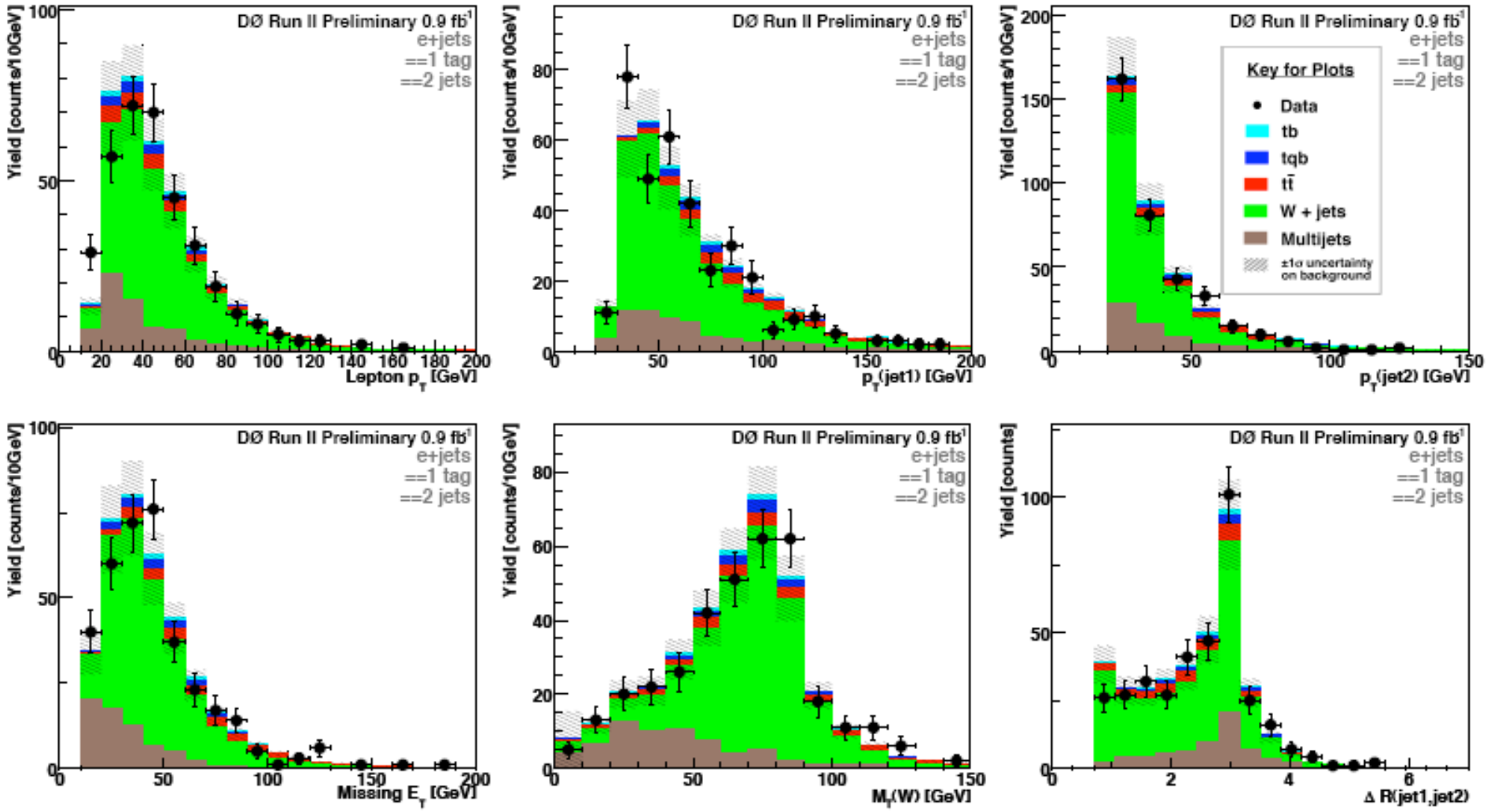


For this analysis:

b-jet efficiency ~ 50%
c-jet efficiency ~ 10%
Light quark efficiency ~ 0.5%



Lepton+2jets:
Data vs MC (w/ b-tagging)



Performance of Event Selection

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

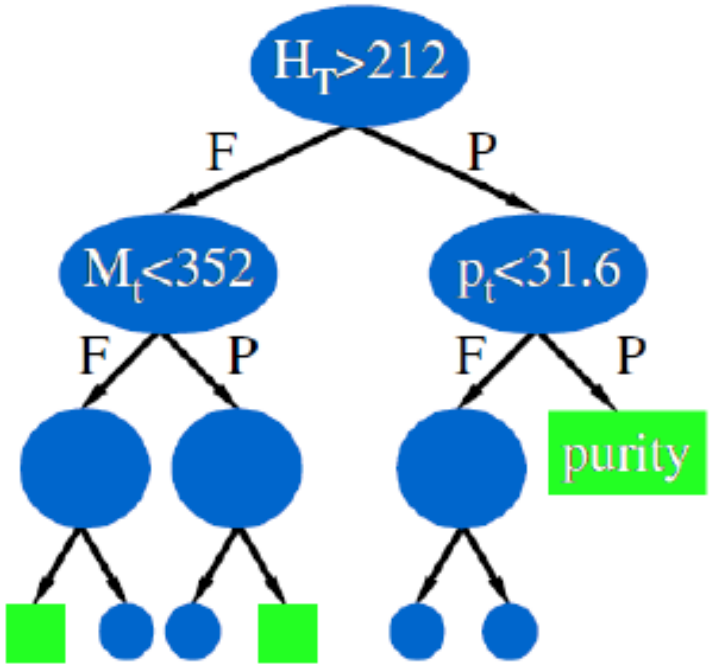
Electron + Muon	1 jet	2 jets	3 jets	4 jets	≥ 5 jets
0 tags	10% 1 : 3,200	25% 1 : 390	12% 1 : 300	3% 1 : 270	1% 1 : 230
1 tag	6% 1 : 100	21% 1 : 20	11% 1 : 25	3% 1 : 40	1% 1 : 53
2 tags		3% 1 : 11	2% 1 : 15	1% 1 : 38	0% 1 : 43

Analysis Techniques

- 3 different methods:
  - Decision Trees
  - Bayesian Neural Networks
  - Matrix Element Method

Decision Trees

- Start with all events (first node)
- For each variable, find the splitting value with best separation betweer children (best cut).
- Select best variable and cut
  - produce Failed and Passed branches



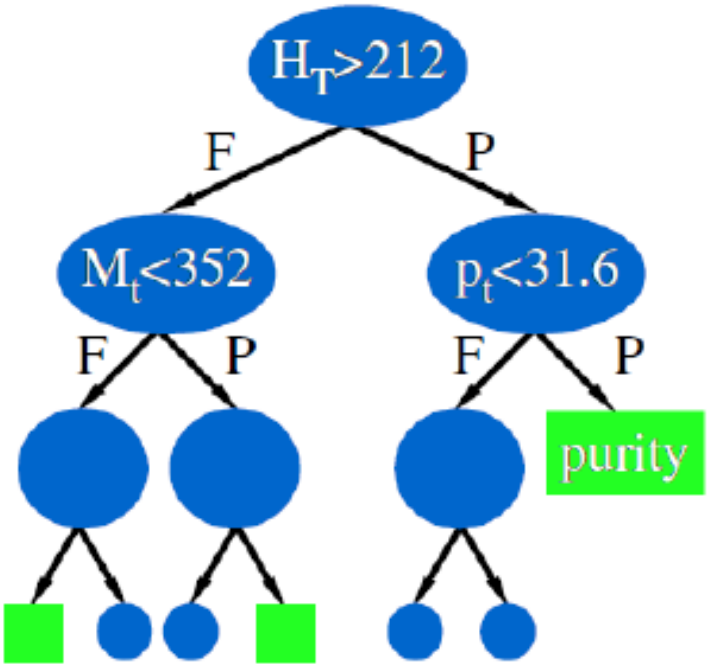
- Repeat recursively on each node
- Stop when improvement stops or when too few events left.

- Terminal node = leaf.

Decision Trees

- Adaptive boosting
  - technique to improve performance of a weak classifier
- 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 :

$$T(i) = \sum_{n=1}^{N_{tree}} \alpha_k \, T_k(i)$$

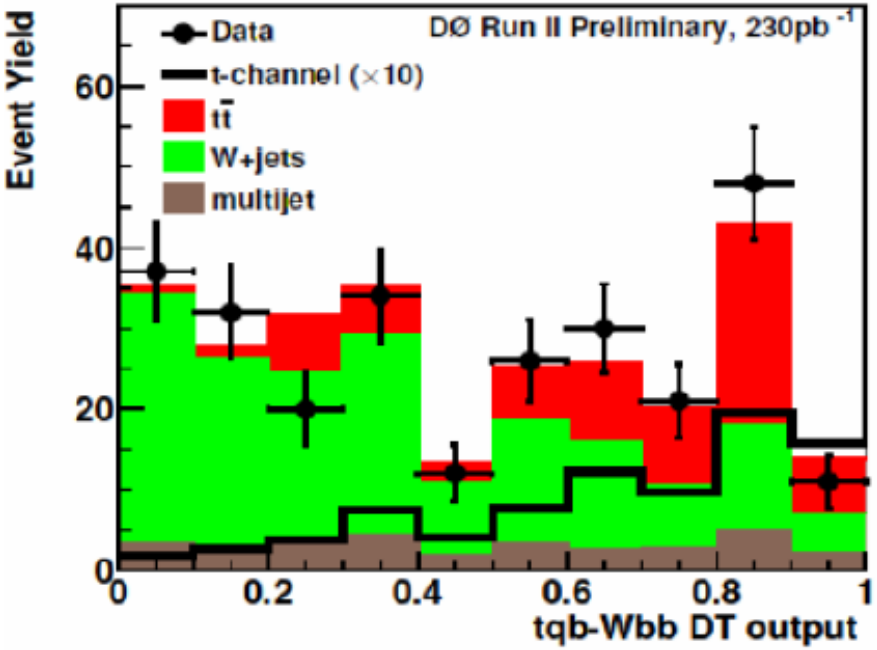


Decision Trees

- Use trained tree
  - run on independent simulated sample, determine purities.
- Apply to Data
  - Signal and background should show separation
  - signal right, background left
- Either
  - cut on output and measure
  - or use whole distribution to measure.

•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 sum of backgrounds



Input Variables

49 of them…

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

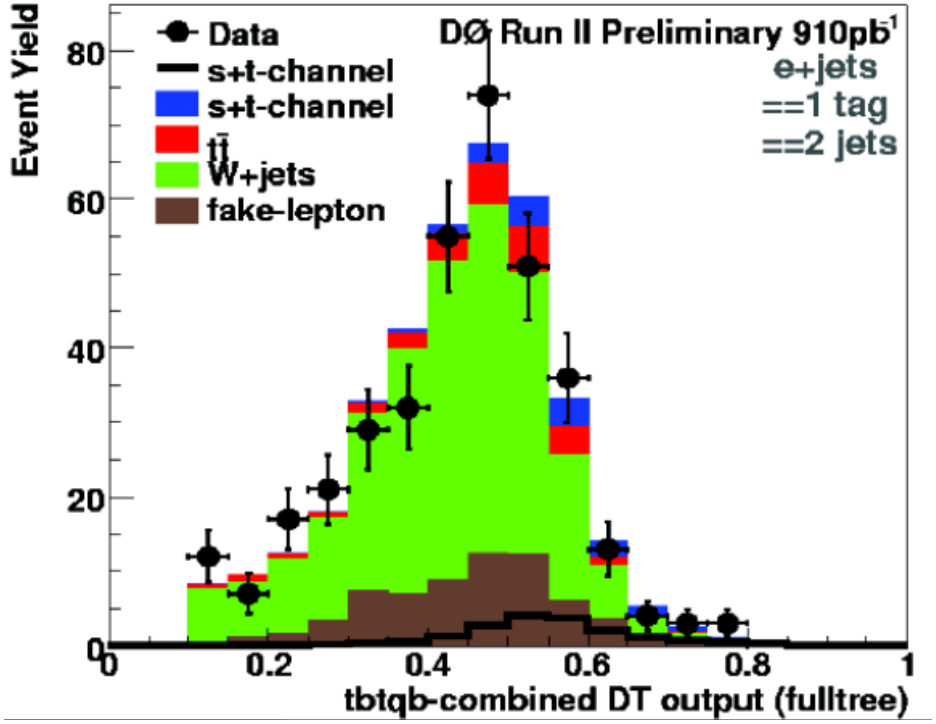
ΔR(jet1,jet2)
cos(best1,lepton)$_{\mathrm{besttop}}$
cos(best1,notbest1)$_{\mathrm{besttop}}$
cos(tag1,alljets)$_{\mathrm{alljets}}$
cos(tag1,lepton)$_{\mathrm{btaggedtop}}$
cos(jet1,alljets)$_{\mathrm{alljets}}$
cos(jet1,lepton)$_{\mathrm{btaggedtop}}$
cos(jet2,alljets)$_{\mathrm{alljets}}$
cos(jet2,lepton)$_{\mathrm{btaggedtop}}$
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

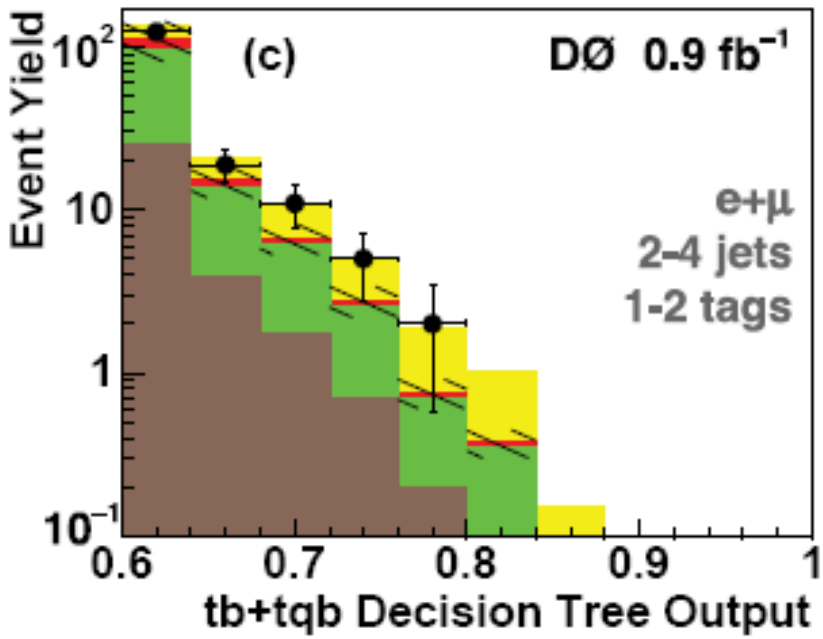
Aplanarity(alljets,W)
M(W,best1) ("best" top mass)
M(W,tag1) ("b-tagged" top mass)
H_T(alljets)
H_T(alljets$-$best1)
H_T(alljets$-$tag1)
H_T(alljets,W)
H_T(jet1,jet2)
H_T(jet1,jet2,W)
M(alljets)
M(alljets$-$best1)
M(alljets$-$tag1)
M(jet1,jet2)
M(jet1,jet2,W)
M_T(jet1,jet2)
M_T(W)
Missing E_T
p_T(alljets$-$best1)
p_T(alljets$-$tag1)
p_T(jet1,jet2)
Q(lepton)$\times\eta$(untag1)
$\sqrt{\hat{s}}$
Sphericity(alljets,W)

Decision Trees applied to data

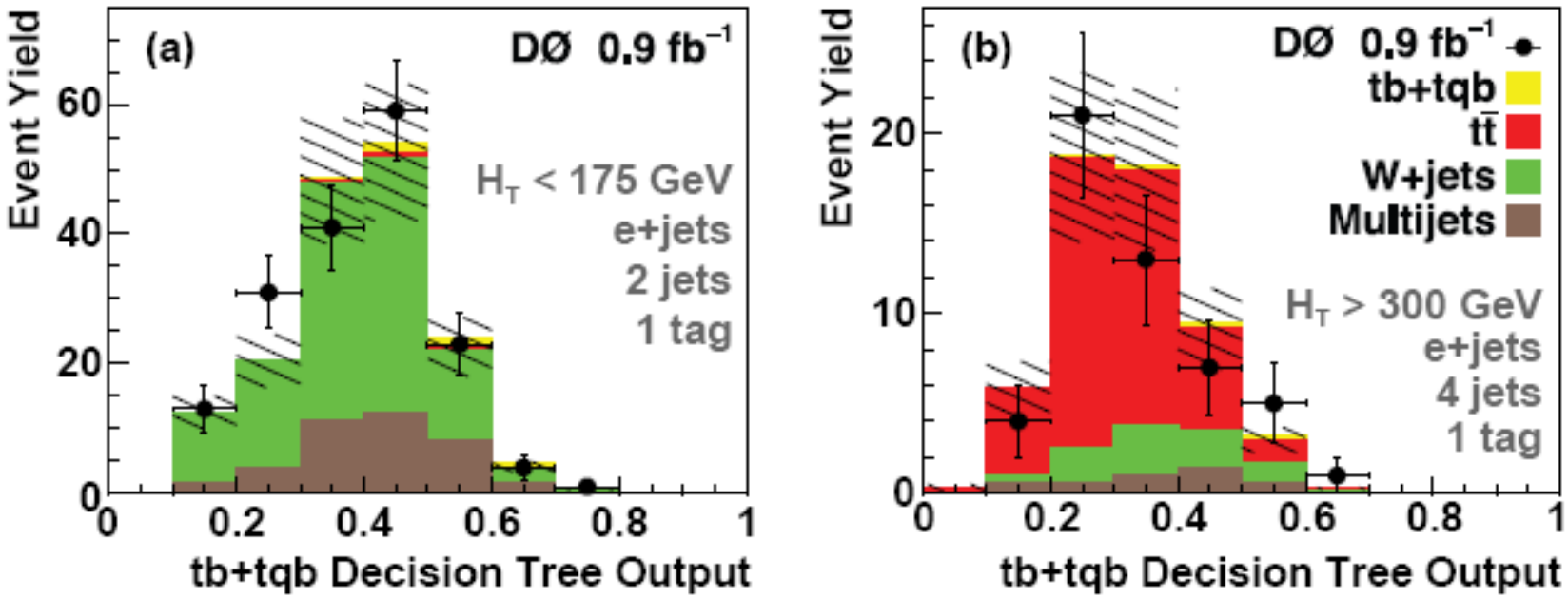
- Electron, 2 jet, 1 tag



All data, zoomed



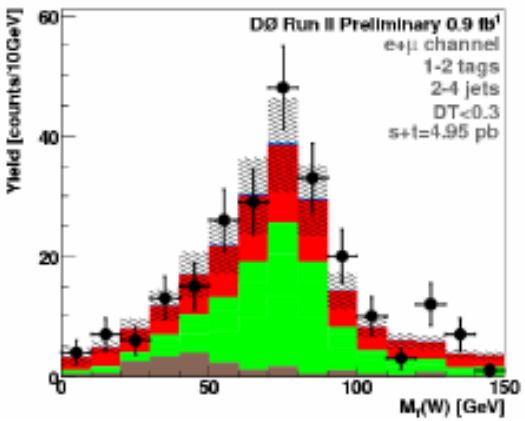
Decision Trees applied to data



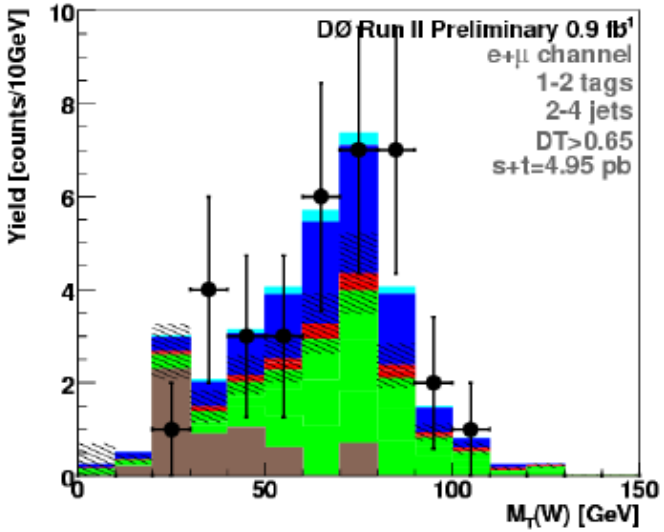
Event Characteristics

$M_T(W)$

$DT < 0.3$

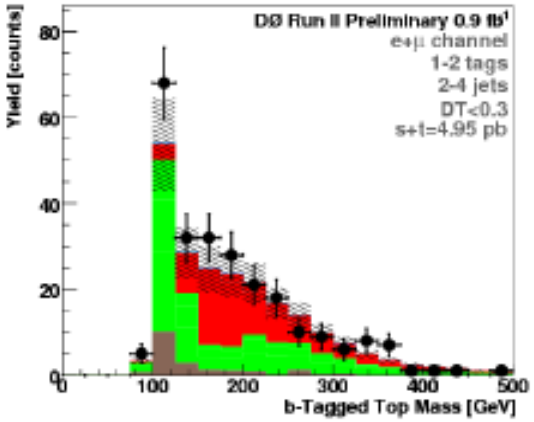


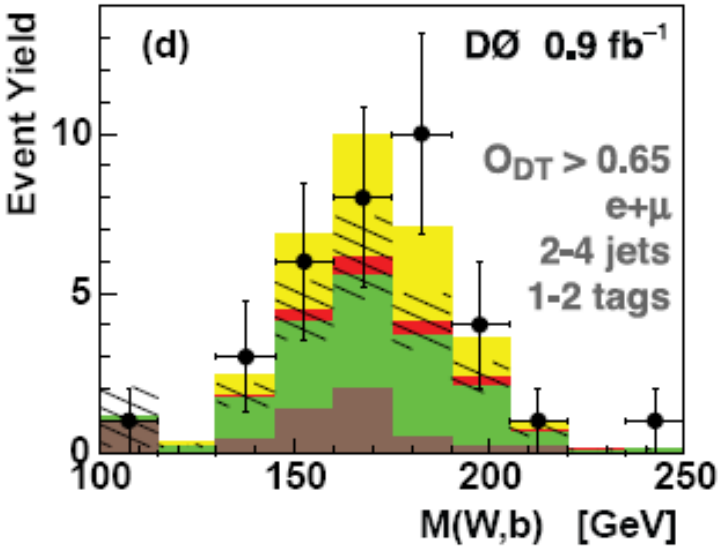
$DT > 0.65$



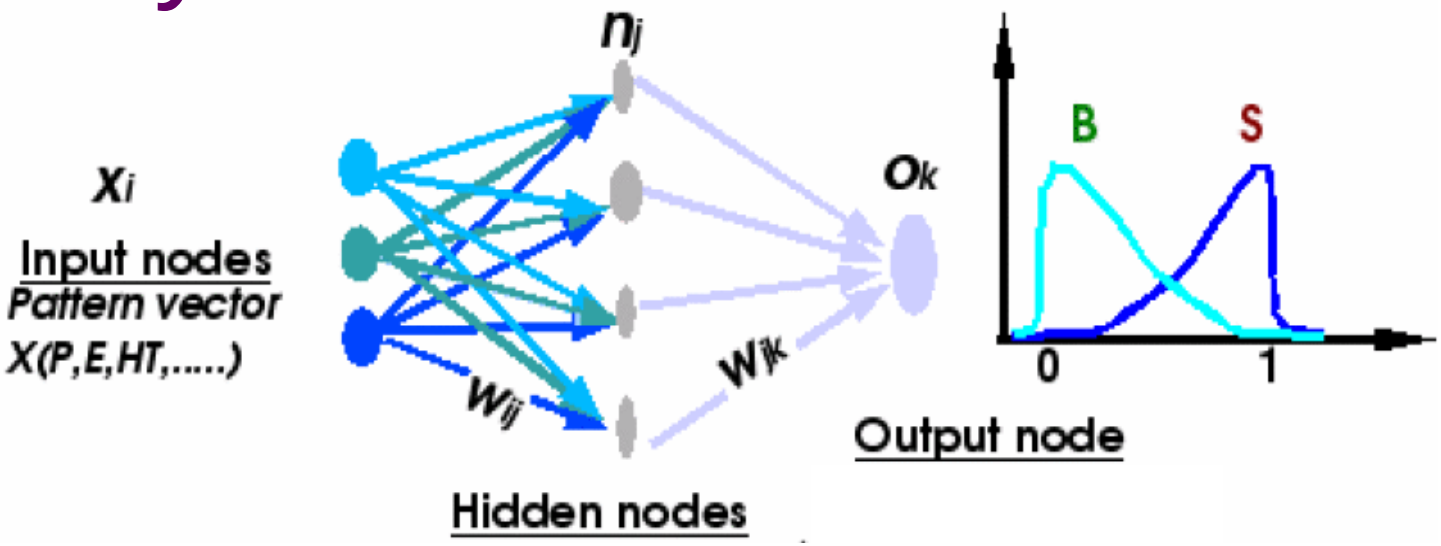
$M(W,b)$

$DT < 0.3$





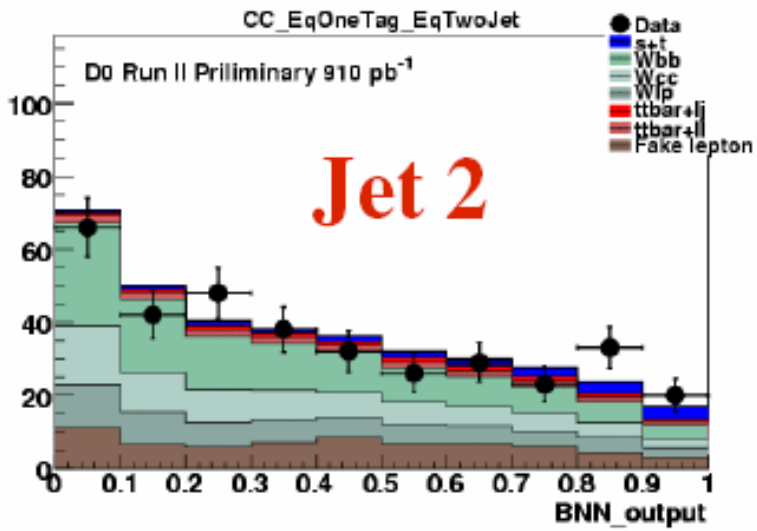
Bayesian Neural Networks

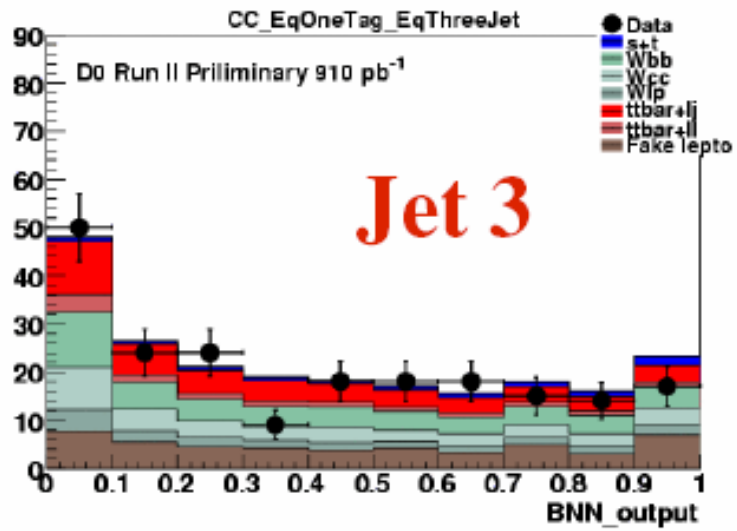


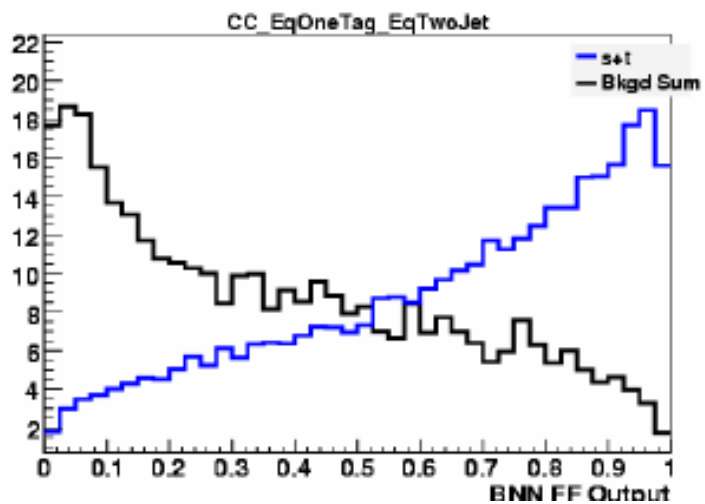
- Instead of choosing one set of weights
  - find posterior probability density over all possible weights
- Average over many networks weighted by the probability of each network given the training data
  - Less prone to overtraining
- For details: http://www.cs.toronto.edu/radford/fbm.software.html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

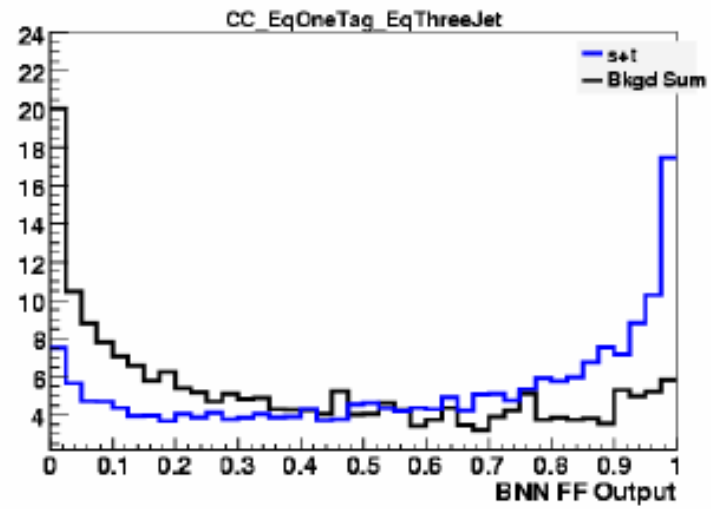
Bayesian Neural Networks

Events with lepton+2 jets.









Matrix Element Method

- Encodes all the kinematic information.
  - The matrix element contains all the properties of the interaction

$$\mathrm{d}\sigma = \frac{|\mathcal{M}|^2}{F}\mathrm{d}Q$$

    - Where *F* is the flux factor and d*Q* is the Lorentz invariant phase space factor.

Matrix Element Method

- Use matrix elements of main signal and background diagrams to compute
  - an event probability density for signal and background hypotheses
- Calculate discriminant:

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

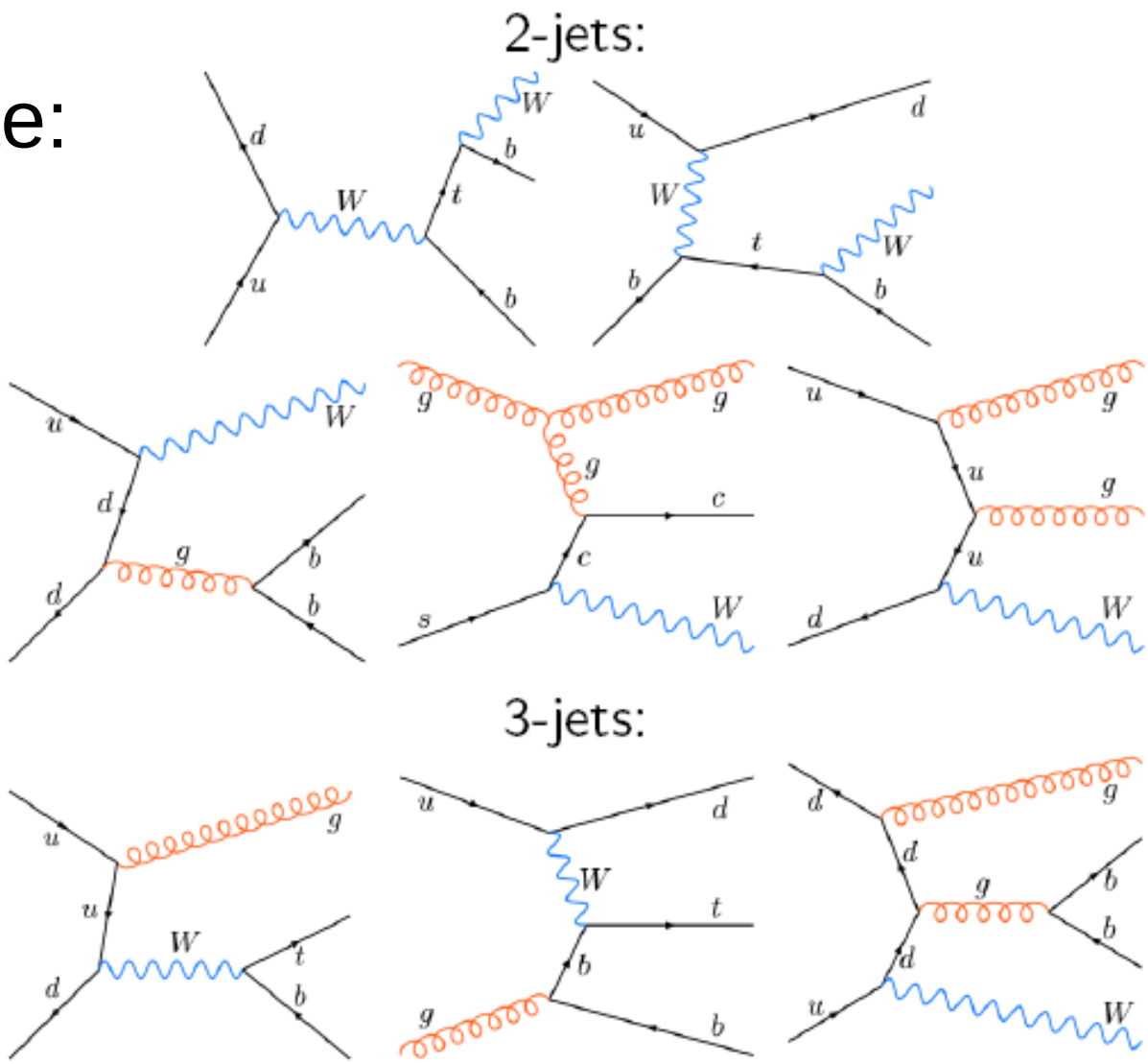
- Define P_{Signal} as properly normalized differential cross section

$$P_{Signal}(\vec{x}) = \frac{1}{\sigma_S} d\sigma_S(\vec{x}) \quad \sigma_S = \int d\sigma_S(\vec{x})$$

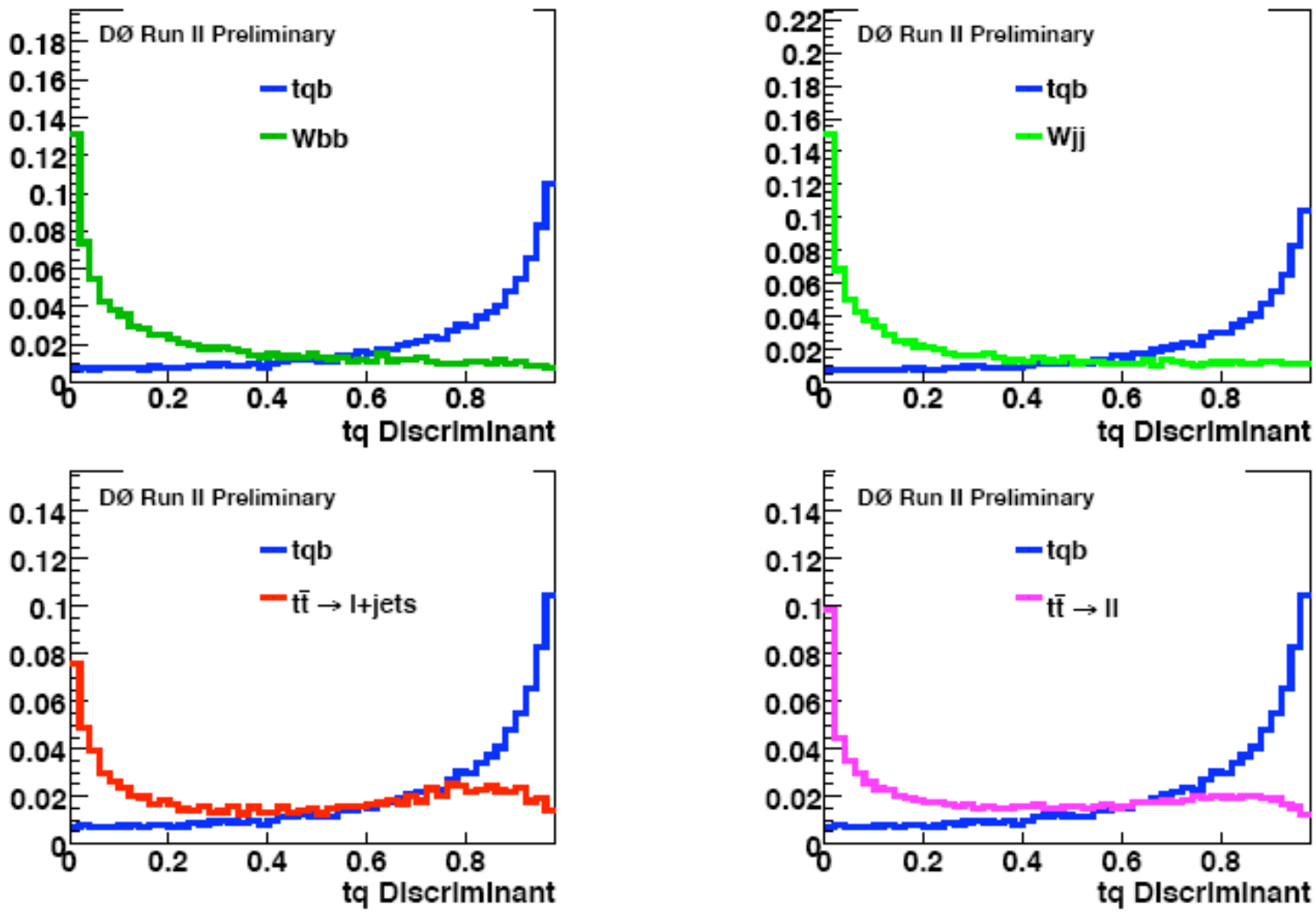
- Technique similar to top quark mass measurement and shares some details (eg. Transfer functions)

Matrix Element Method

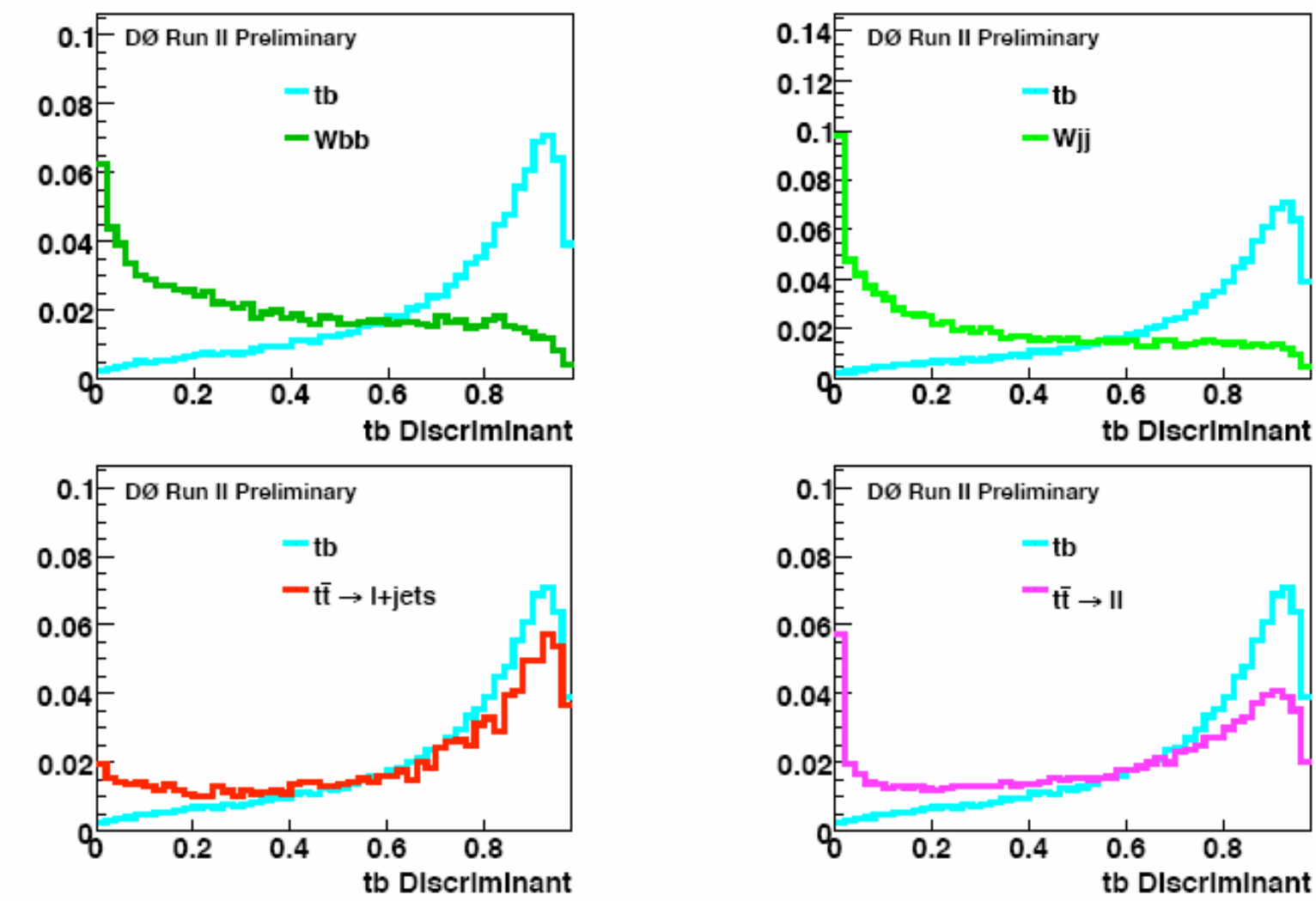
- Compute:



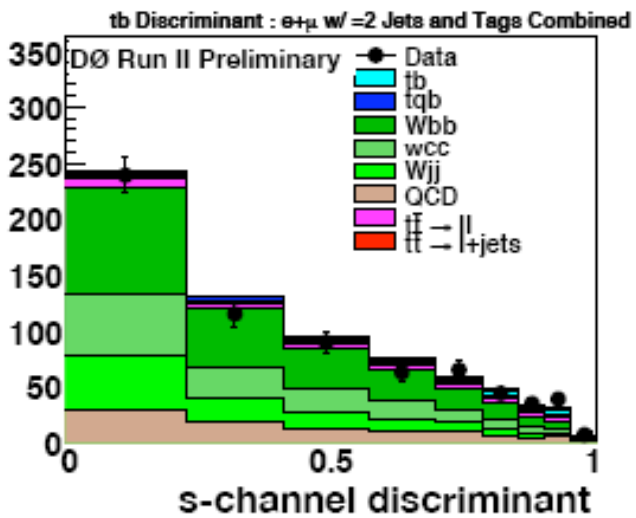
t-channel Discriminant

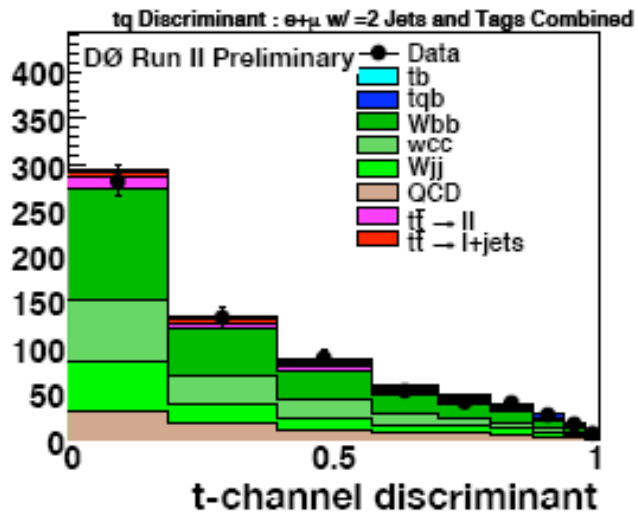


s-channel Discriminant

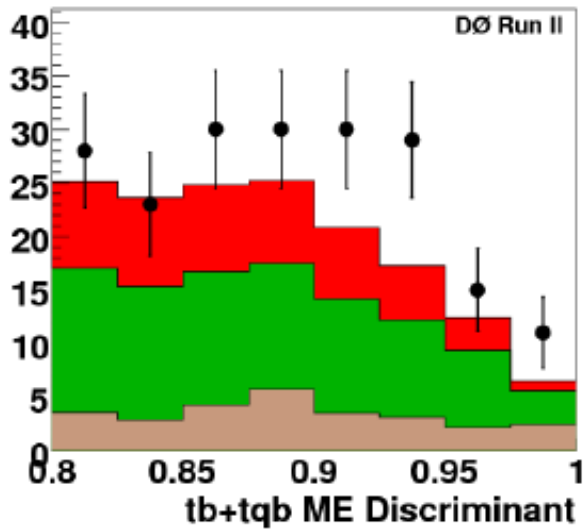


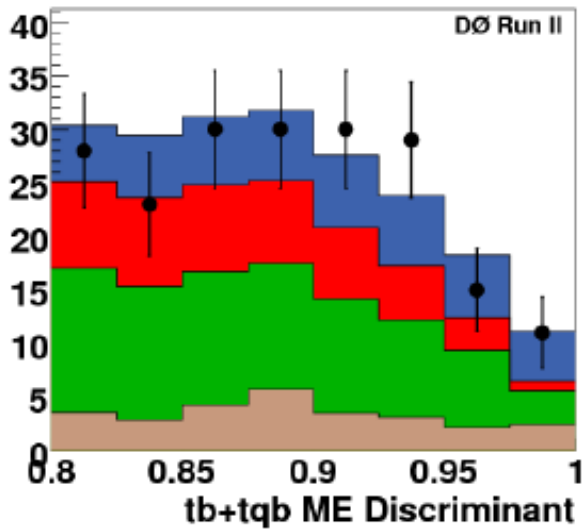
ME applied to data





Discriminant output with and without signal component (all channels combined in 1D to “visualize” excess)





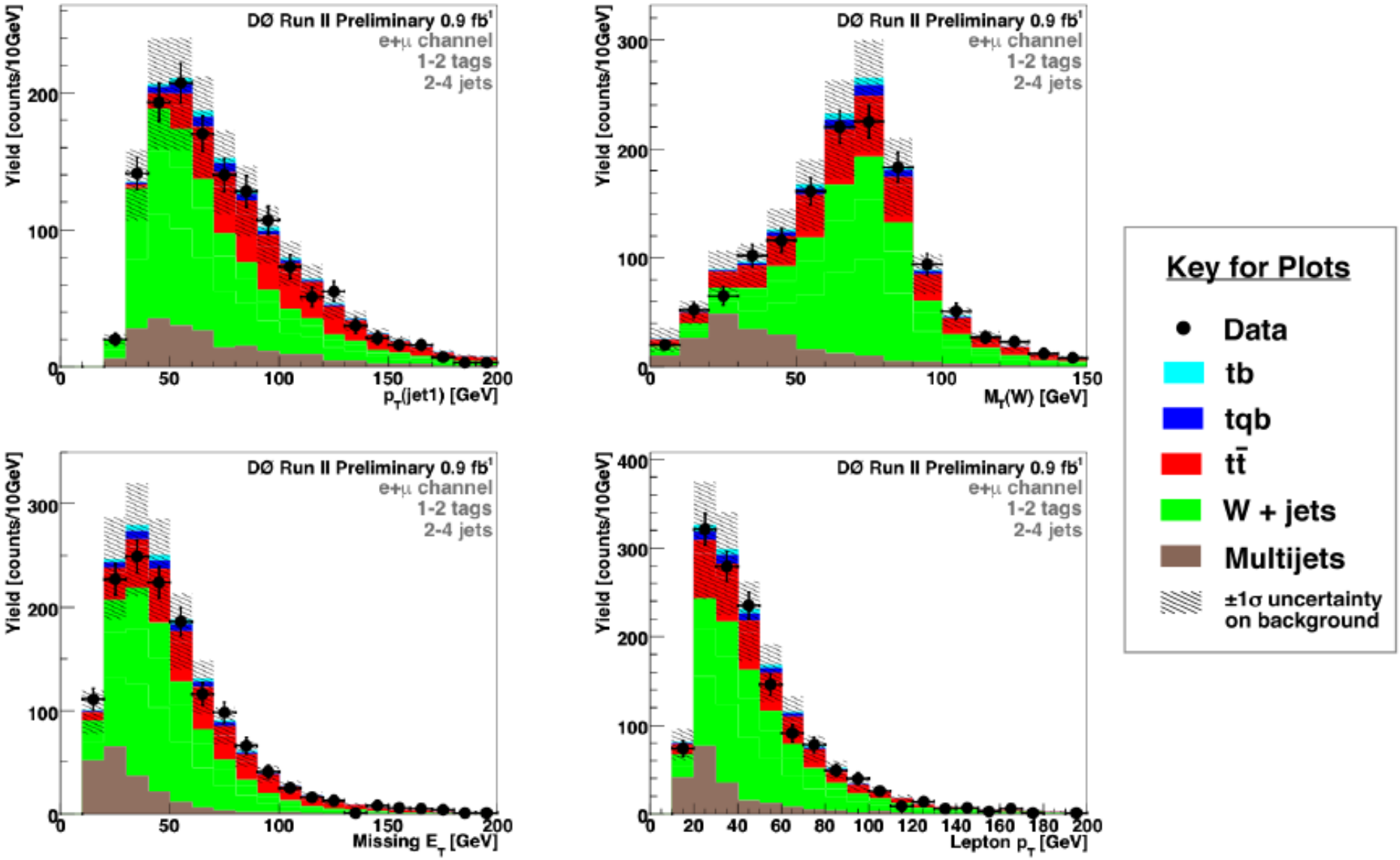
Systematic Errors

- Systematic uncertainties can be either “shaped” (jet energy scale, tag rate functions)
- Shift inputs by ±1, redo analysis or “normalization”
- Uncertainties assigned per background, jet multiplicity, lepton, number of tags

Examples of Relative Systematic Uncertainties

$t\bar{t}$ cross section	18%
Luminosity	6%
Electron trigger	3%
Muon trigger	6%
Jet energy scale	wide range
Jet fragmentation	5–7%
Heavy flavor ratio	30%
Tag-rate functions	2–16%

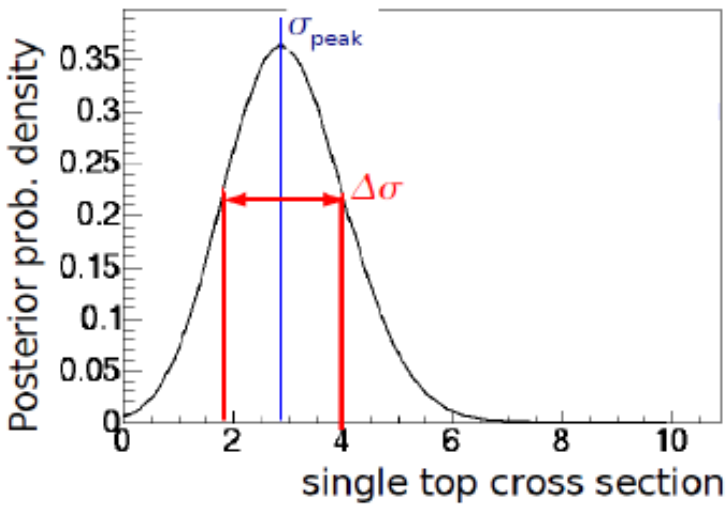
Systematic Errors



Computing Cross Section

- 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

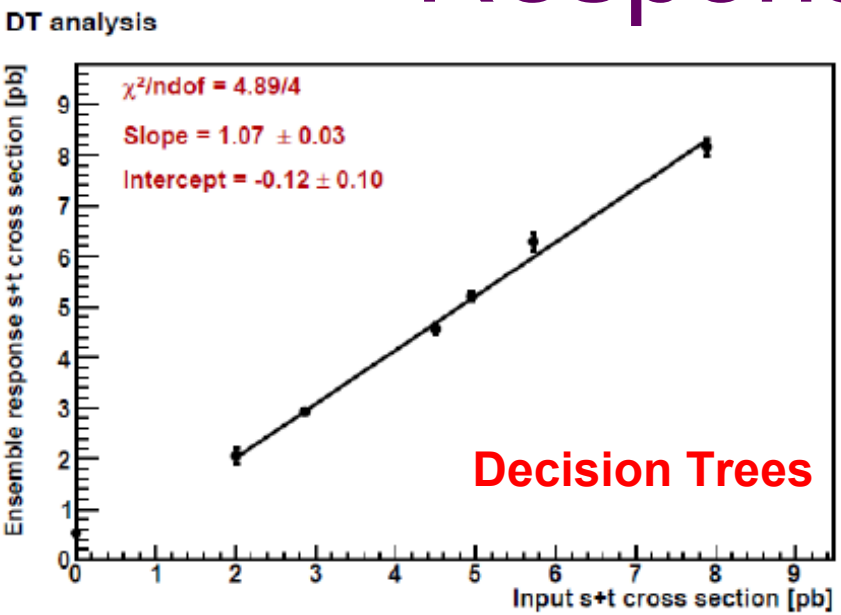
$$Post(\sigma|D) \equiv P(\sigma|D) \propto \int_a \int_b P(D|\sigma, a, b) Prior(\sigma) Prior(a, b)$$

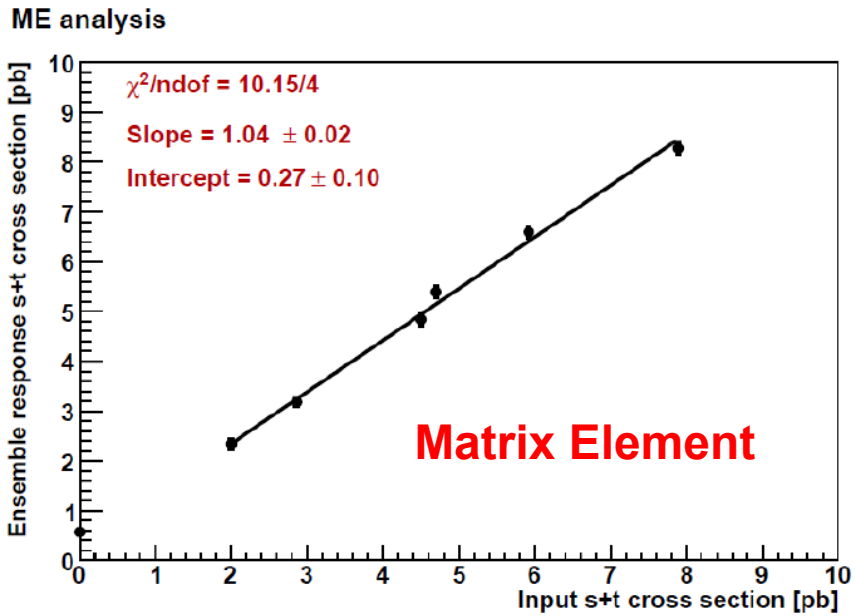
- Bayesian posterior probability density
  - Shape and normalization systematics treated as nuisance parameters
  - Correlations between uncertainties properly accounted for
  - Flat prior in signal cross section

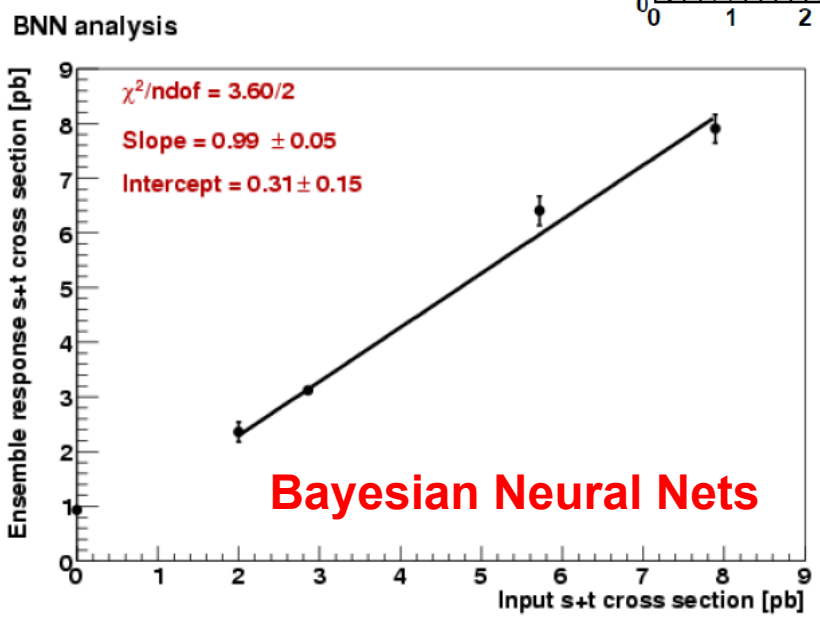
Ensemble Tests

- To verify that all of this 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 include:
  1. 0-signal ensemble (s + t = 0pb)
  2. SM ensemble (s + t = 2.9pb)
  3. “Mystery” ensembles to test analyzers (s + t =??pb)
  4. Ensembles at measured cross section (s + t =measured)
  5. A high luminosity ensemble
- Tests linearity of “response” to single top.

Response Linearity







Sensitivity Determination

- We use our 0-signal ensemble to determine a significance for each measurement.

- Expected p-value
  - The fraction of 0-signal pseudo-datasets in which we measure at least 2.9pb.
- Observed p-value
  - The fraction of 0-signal pseudo-datasets in which we measure at least the measured cross section.

- We also can use the SM ensemble to see how compatible our measured value is with the SM.

Sensitivity Estimate Ensemble

- Use a pool of weighted signal + background events (about 850k in each of electron and muon)

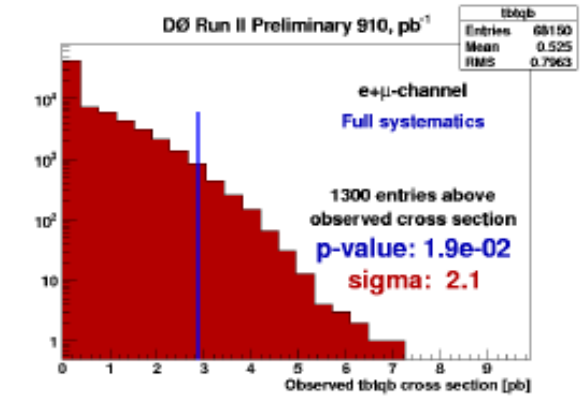
- Fluctuate relative and total yields in proportion to systematic errors
- Randomly sample from a Poisson distribution about the total yield
- Generate a set of pseudo-data (a member of the ensemble)

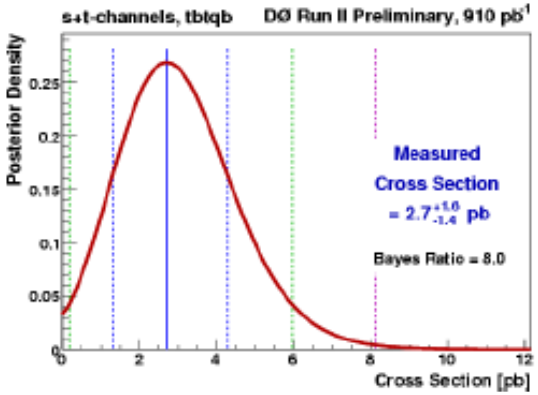
- Pass the pseudo-data through the full analysis chain (including systematic uncertainties)

Expected p-value

Decision Trees

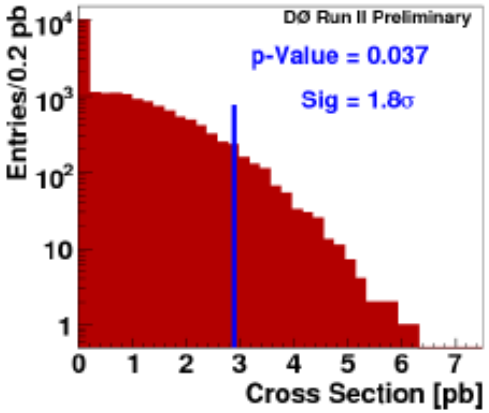
p-value **1.9%**

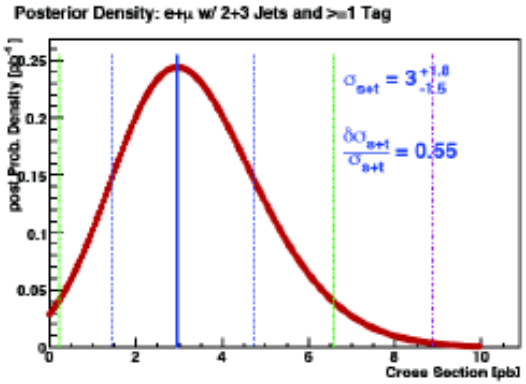




Matrix Elements

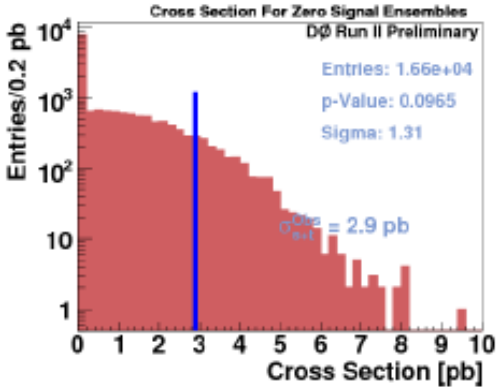
p-value **3.7%**

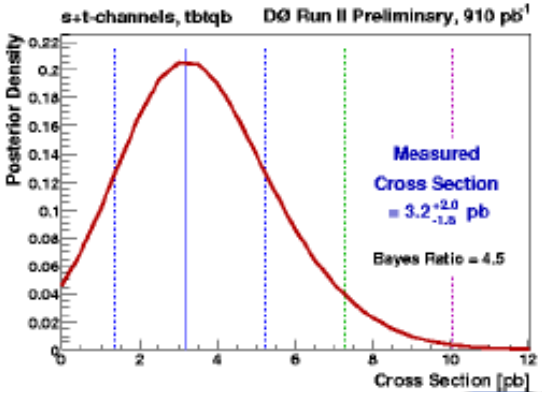




Bayesian NN

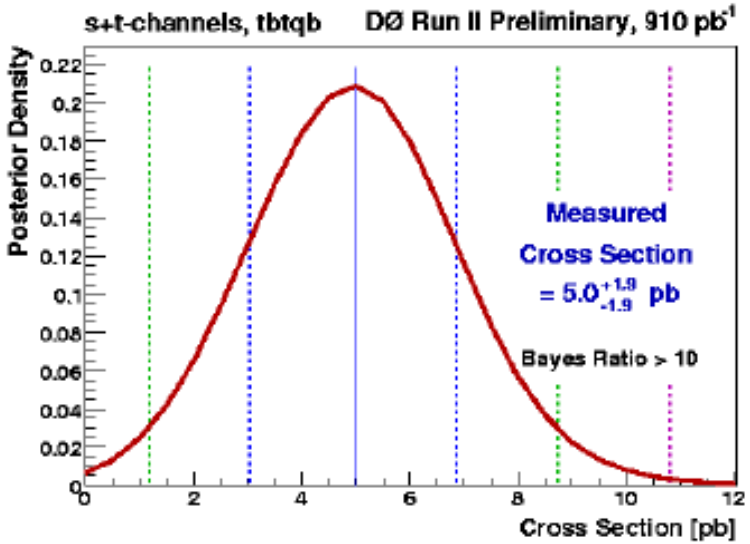
p-value **9.7%**

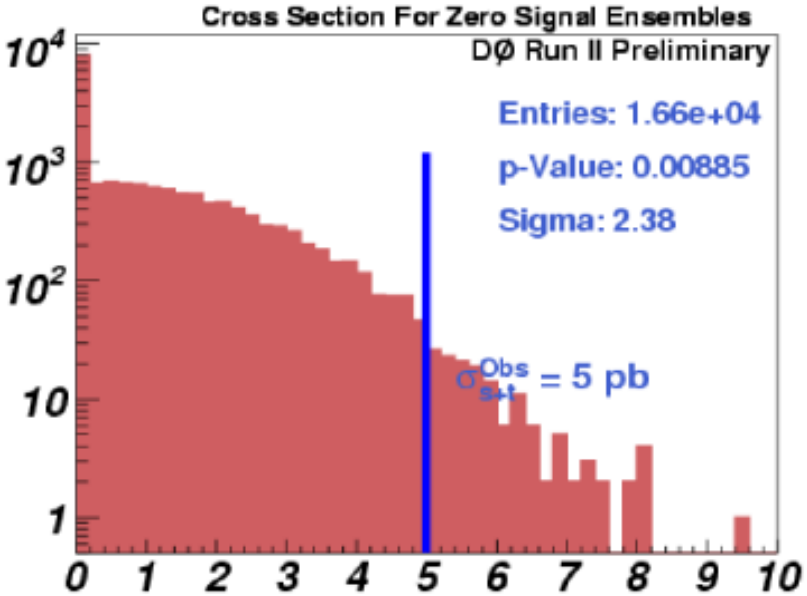




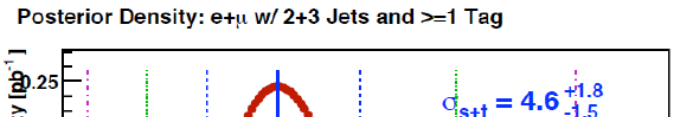
Observed p-value: BNN

- Least sensitive (a-priori) analysis sees 2.4 σ effect!

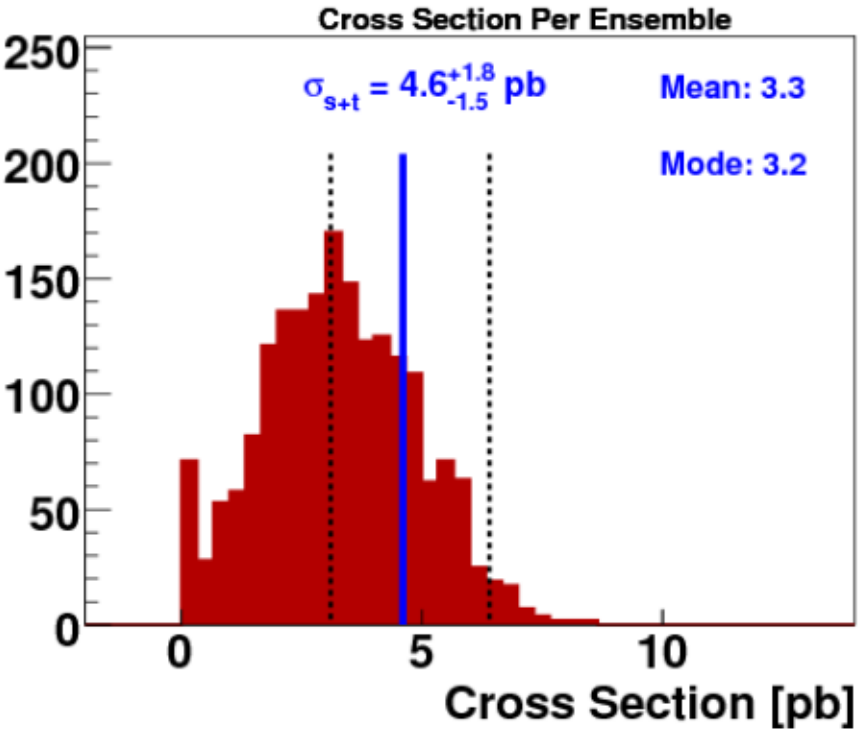


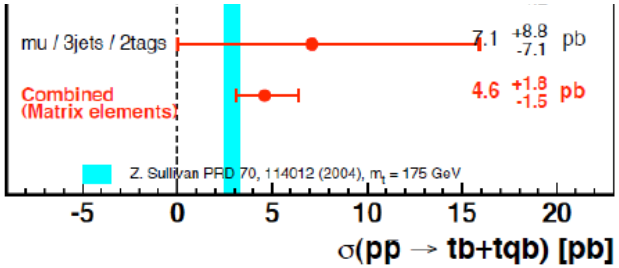


Observed p-value: ME

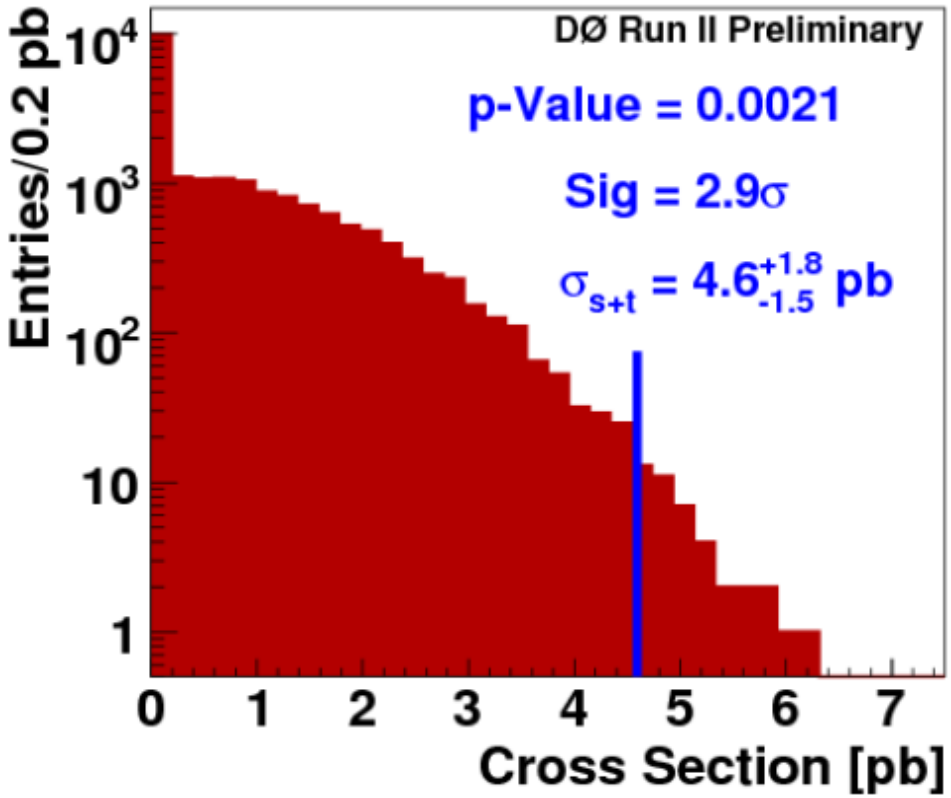


SM compatibility = 21%

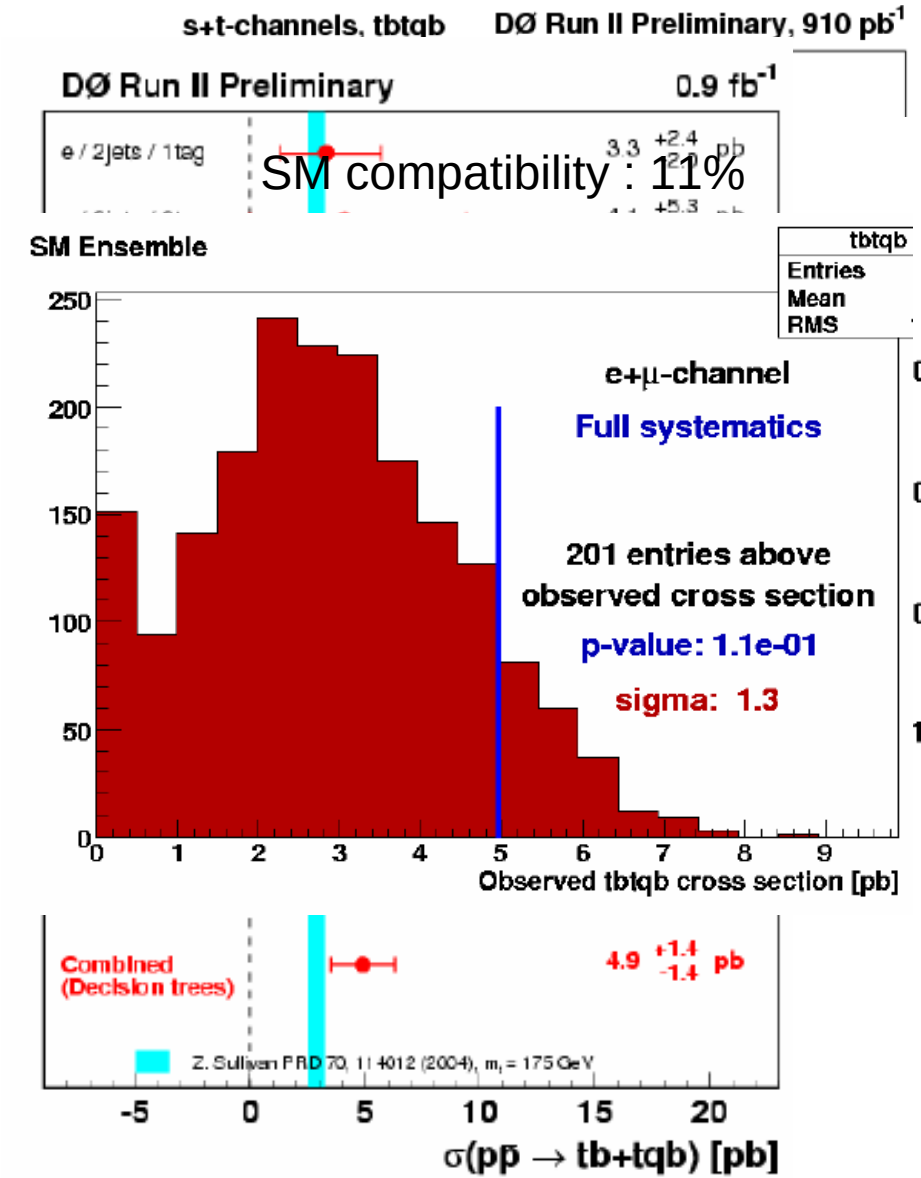




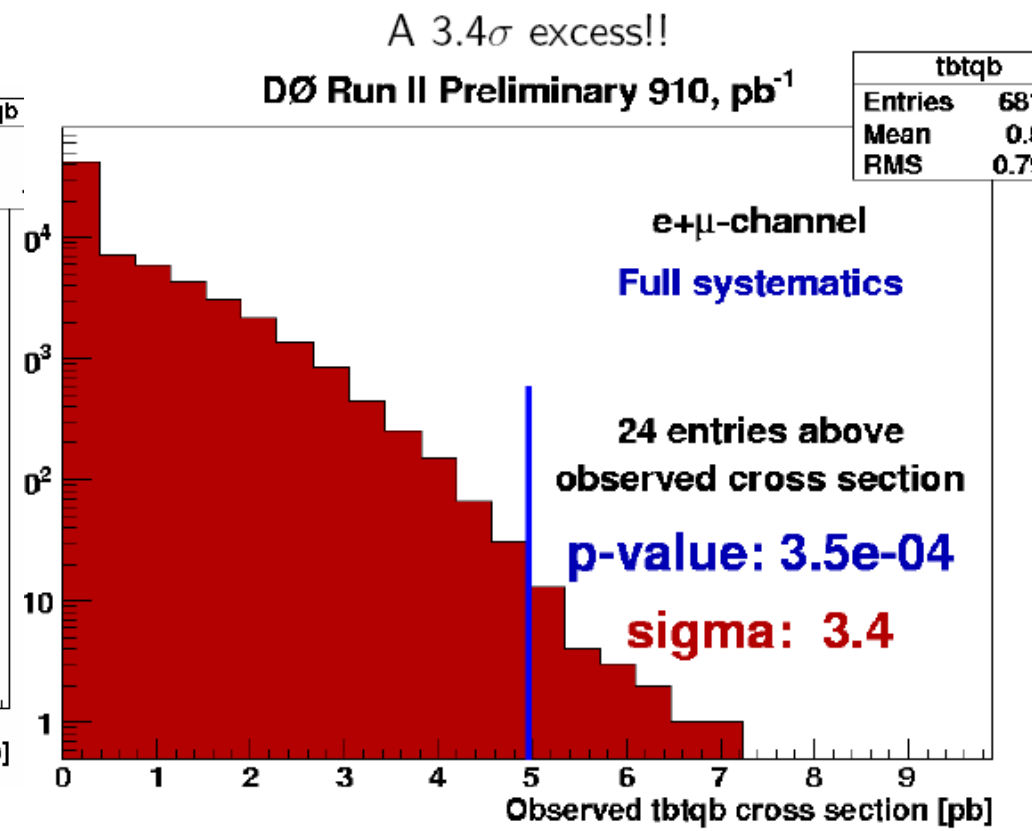
p-value=0.0021, 2.9σ !!



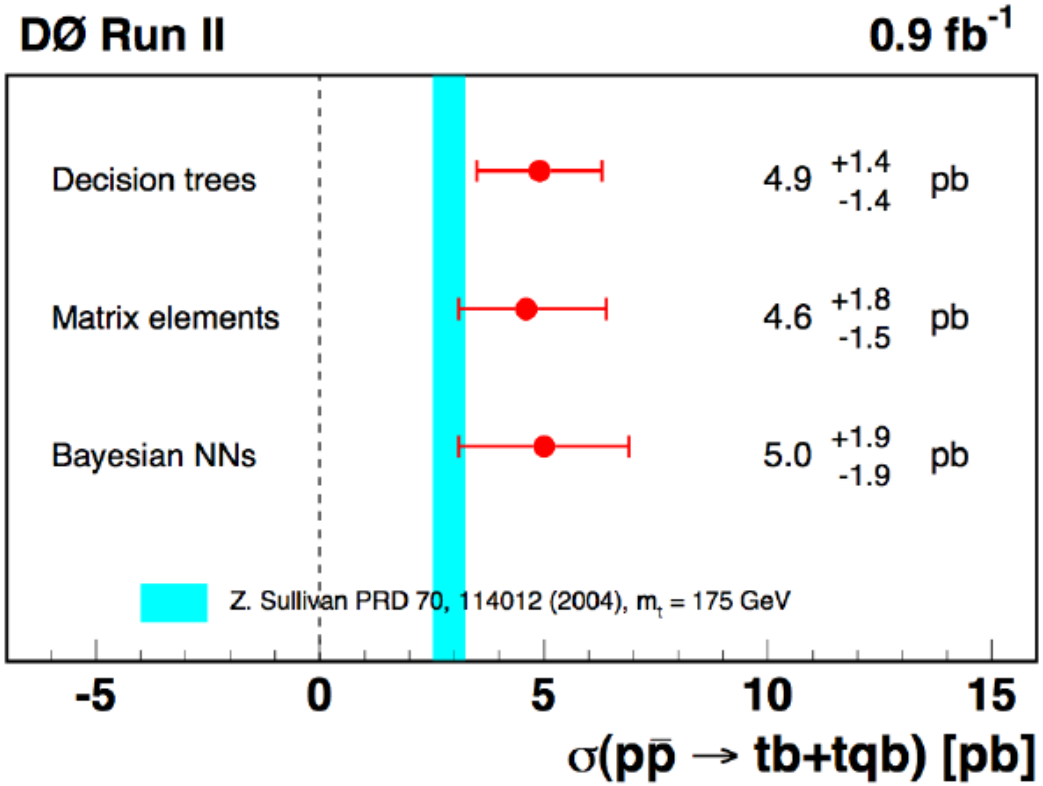
Observed p-value: DT



SM compatibility : 11%



s+t Cross Sections



Correlations:

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

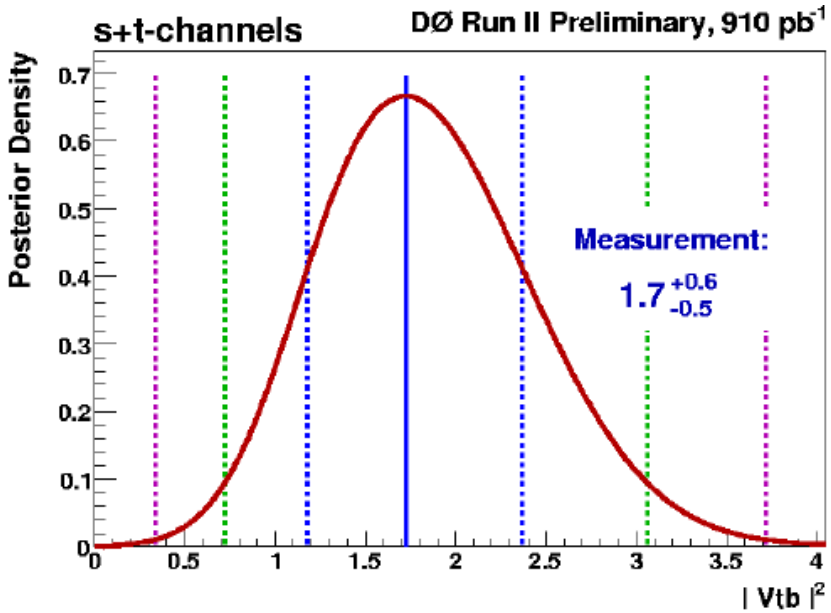
V_{tb}

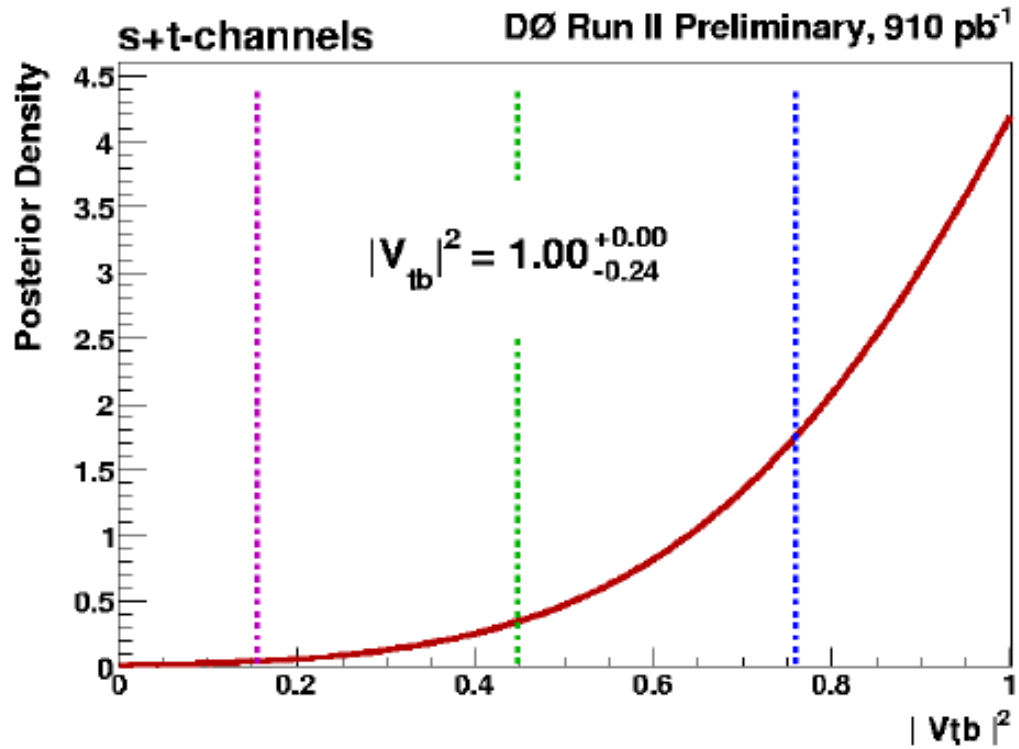
- From single top cross section, we can make the first direct measurement of $|V_{tb}|$.
- Same infrastructure as cross section measurement but make a posterior in $|V_{tb}|^2$.
  - Caveat: assume SM top quark decays.
- Additional theoretical errors are needed

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

- see hep-ph/0408049

V_{tb}





Constrain $|V_{tb}|$ to physical region and integrate:

$$|V_{tb}| = 1.00^{-0.12}_{+0}$$

Conclusion

- First Evidence for Single Top Quark Production
  - s+t cross section: 4.9 ± 1.4pb
  - 3.4 σ significance!

- Three techniques in good agreement.

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

$$|V_{tb}| = 1.00^{-0.12}_{+0}$$