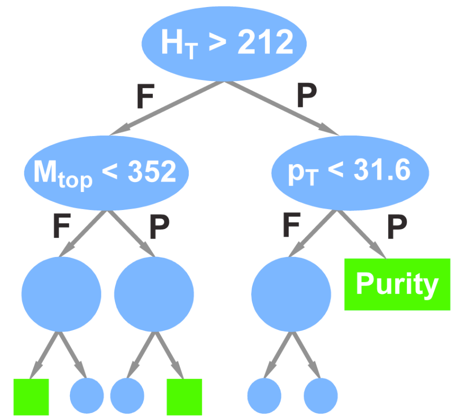
Using Boosted Decision Trees

to find

Single Top Quark Events

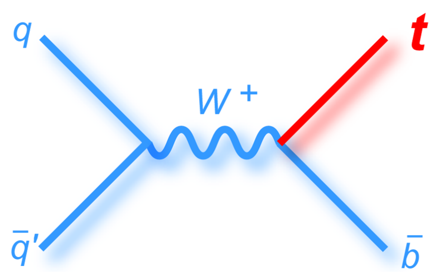
at the

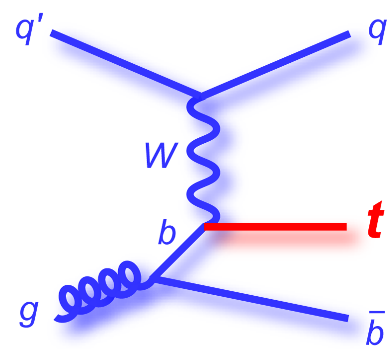
DØ Experiment



Ann Heinson
University of California, Riverside
for the DØ Collaboration

American Physical Society Meeting
Sunday May 3, 2009





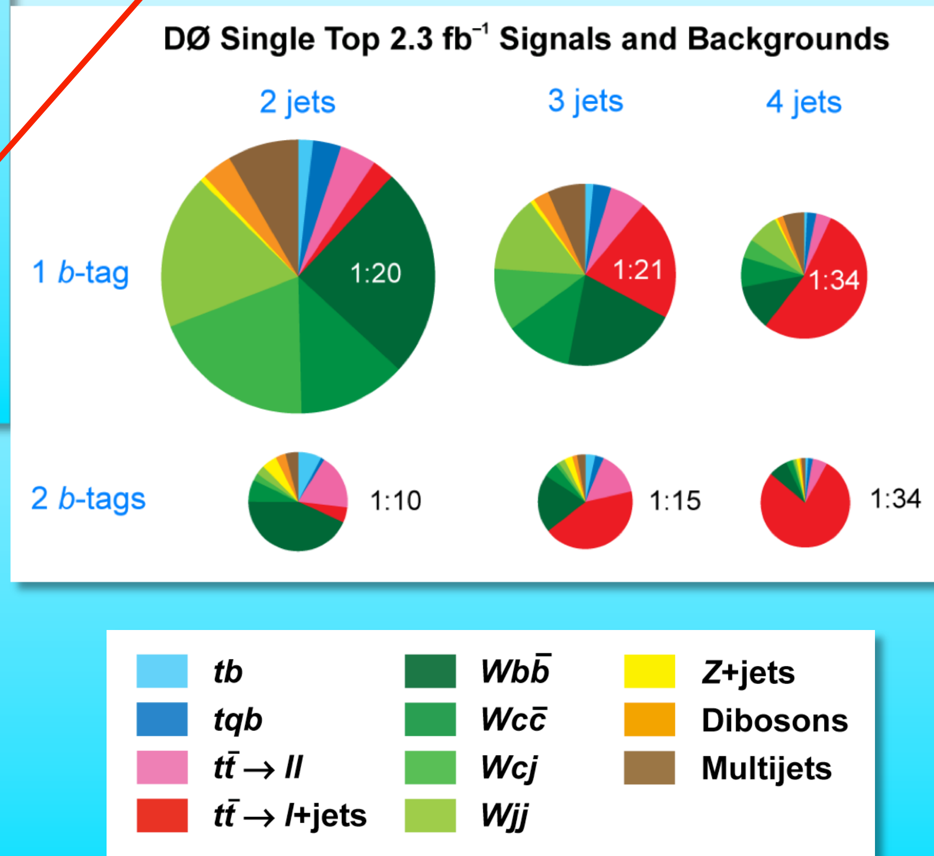
Event Yields After Selection

Event Yields in 2.3 fb^{-1} of DØ Data

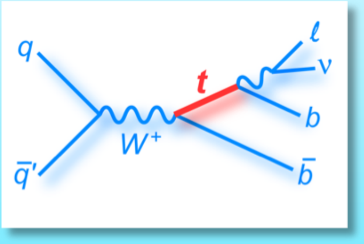
e,μ, 2,3,4-jets, 1,2-tags combined

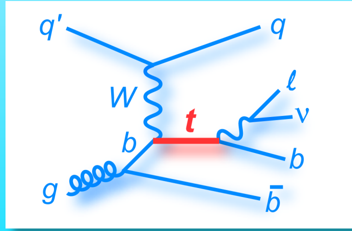
tb + *tqb*	223 ± 30
W+jets	2,647 ± 241
Z+jets, dibosons	340 ± 61
$t\bar{t}$ pairs	1,142 ± 168
Multijets	300 ± 52
Total prediction	4,652 ± 352
Data	4,519

Need to find
223 predicted signal events
in 20x larger background

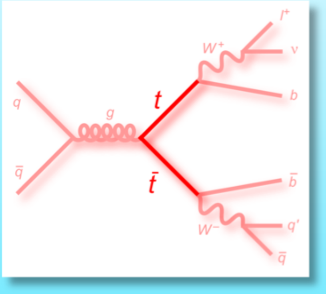


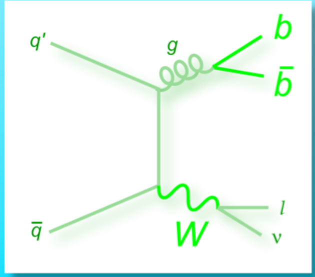
Separating Signal from Background





- Counting events to separate signal from background is useless when S:B = 1:20
- Use the shapes of many variables to add extra information
- Combine the variables in a multivariate discriminant, many choices are available



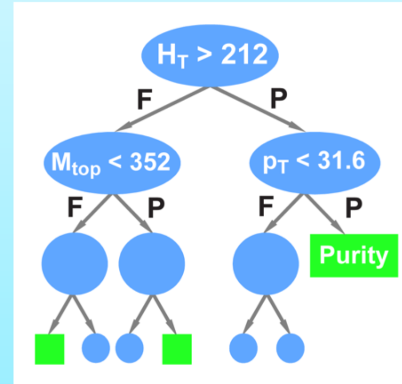


- **Advantages of Boosted Decision Trees**
  - Fast to train
  - Can combine all backgrounds during training and measurement, no loss of sensitivity compared to keeping them separate (unlike for traditional neural networks)
  - Not degraded by the addition of more input variables (unlike neural networks), so no need to optimize the choice, just use all sensitive variables with good agreement to data

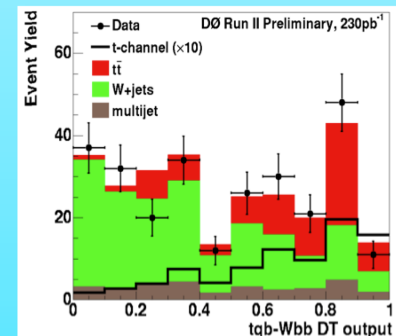
How Decision Trees Work

- Idea: recover events that fail criteria in cut-based analyses
- Start at first "node " with "training sample" of signal and background MC events
  - Test each variable, find splitting value with best separation
  - Select variable and splitting value with best overall separation to produce two "branches ⟶ " with corresponding events, Failed and Passed cut
- Repeat recursively on each node
- Stop when improvement ends or too few (100) events are left
- Terminal node is called a "leaf " with

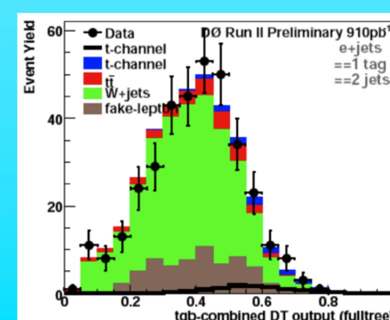
  Purity = $N_{signal}/(N_{signal}+N_{background})$
- Decision tree output for each event = leaf purity value (closer to 0 for background, closer to 1 for signal)
- Boosting averages the results of many trees, dilutes the discrete nature of the output, improves performance by ~20%
- Adaptive boosting algorithm used, 50 boosting cycles
- Trained 24 sets of trees: (e,μ) × (2,3,4 jets) × (1,2 b-tags) × (Run IIa, Run IIb)
- Run independent MC and data through tree to derive results



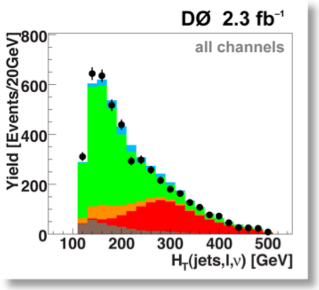
Before boosting

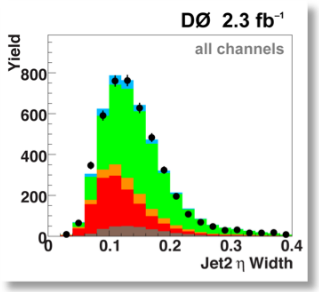


After boosting

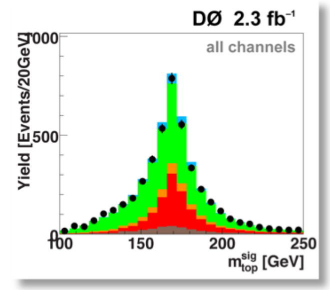


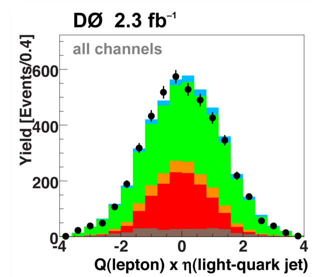
Choice of Input Variables





- Start with 600 variables expected to differ between signal and at least one of the background components
- Remove variables from list with low KS-test value between data and background
- Remove variables with not much discrimination power to reduce computation time later (rank them after decision tree training)
- Use 64 remaining variables in all analysis channels





BDT – Object Kinematics

- p_T(jet2)
- p_T(jet3)
- p_T(jet4)
- p_T(tag1)
- p_T(light2)
- p_T(notbest2)
- p_T(lepton)
- $\not\!\!E_T$
- Q(lepton)$\times\eta$(jet1)
- Q(lepton)$\times\eta$(jet2)
- Q(lepton)$\times\eta$(best)
- Q(lepton)$\times\eta$(light1)
- Q(lepton)$\times\eta$(light2)

BDT – Event Kinematics

- Centrality(alljets)
- H_T(alljets)
- H_T(alljets$-$tag1)
- H_T(alljets$-$best)
- H_T(jet1,jet2)
- H_T(jet1,jet2,lepton,$\not\!\!E_T$)
- H_T(alljets,lepton,$\not\!\!E_T$)
- H_T($\not\!\!E_T$,lepton)
- H(alljets$-$tag1)
- M(alljets)
- M(alljets$-$best)
- M(alljets$-$tag1)
- M(jet1,jet2)
- M(jet1,jet2,W)
- M(jet3,jet4)
- M_T(jet1,jet2)
- p_T(jet1,jet2)
- $\sqrt{\hat{s}}$
- $M_T(W)$

BDT – Jet Reconstruction

- Width$_\eta$(jet2)
- Width$_\eta$(jet4)
- Width$_\phi$(jet4)
- Width$_\eta$(tag1)
- Width$_\eta$(light2)
- Width$_\phi$(light2)

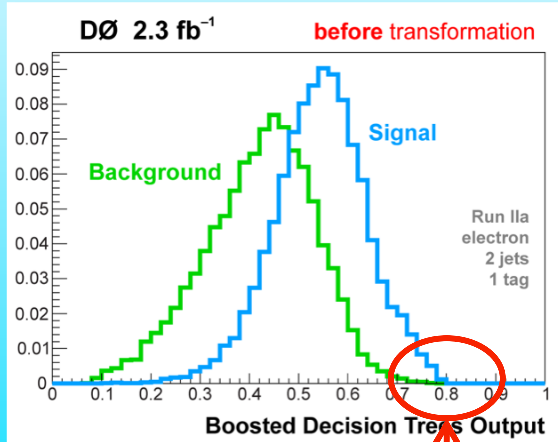
BDT – Top Quark Reconstruction

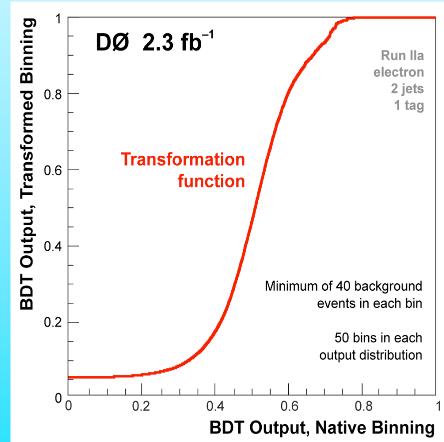
- $M(W,$best1) ("best" top mass)
- $M(W,$tag1) ("b-tagged" top mass)
- $M(W,$tag1$, S2)$ (with 2nd ν solution)
- $M(W,$jet1)
- $M(W,$jet1$, S2)$
- $M(W,$jet2)
- $M(W,$jet2$, S2)$
- $M(W,$notbest2)
- $M(W,$notbest2$, S2)$
- $M_{\rm top}^{\Delta M^{\rm min}}$
- $M_{\rm top}^{\rm sig}$
- $\Delta M_{\rm top}^{\rm min}$
- Significance$_{\rm min}(M_{\rm top})$

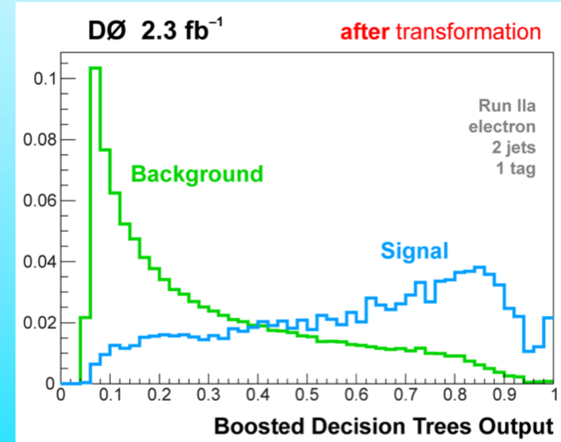
BDT – Angular Correlations

- ΔR(jet1,jet2)
- ΔR(jet1,lepton)
- ΔR(tag1,lepton)
- ΔR(light1,lepton)
- $\Delta\phi$(lepton,$\not\!\!E_T$)
- cos(best,lepton)$_{\rm besttop}$
- cos(best,notbest)$_{\rm besttop}$
- cos(jet1,lepton)$_{\rm btaggedtop}$
- cos(tag1,lepton)$_{\rm btaggedtop}$
- cos(lepton$_{\rm besttop}$,besttop$_{\rm CMframe}$)
- cos(lepton$_{\rm btaggedtop}$,btaggedtop$_{\rm CMframe}$)
- cos(tag1,lepton)$_{\rm btaggedtop}$
- cos(lepton,Q(lepton)$\times z$)$_{\rm besttop}$

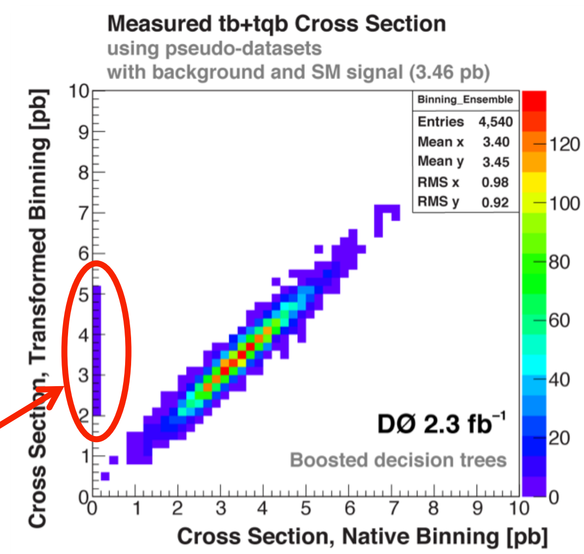
Output Distribution Transformation







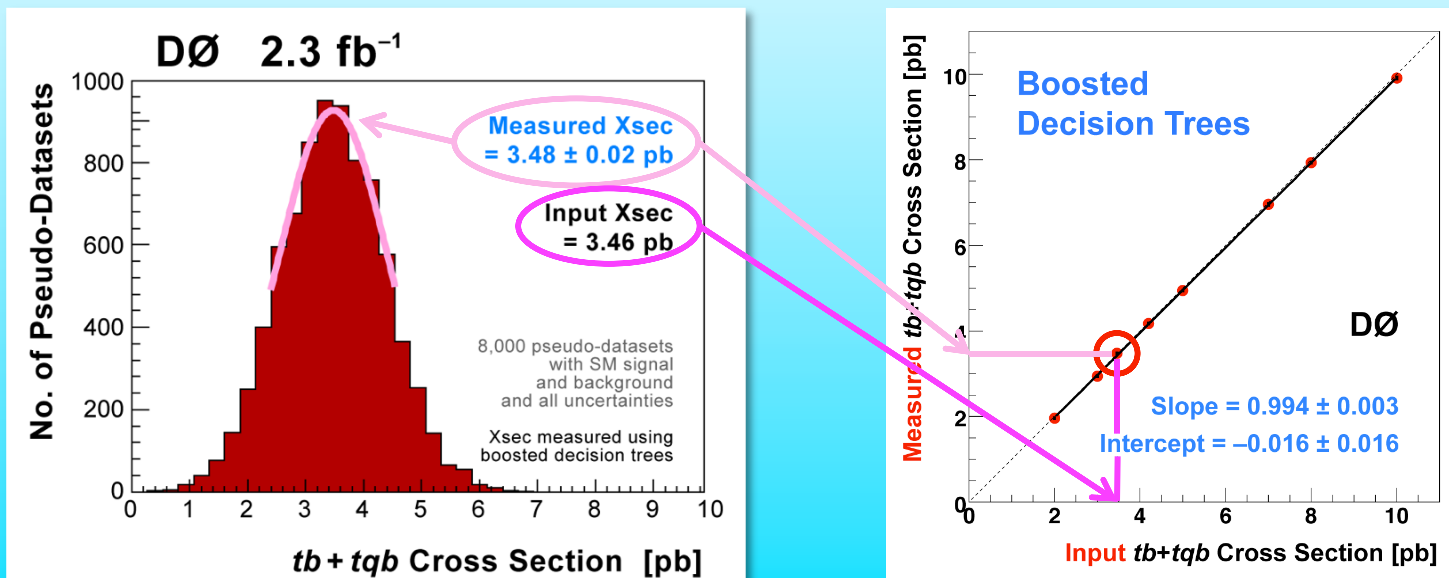
- After boosting, the decision tree outputs are pushed to the center
- Some bins have predicted signal and data but very few background events
- Transform outputs so that all bins have at least 40 background events in them



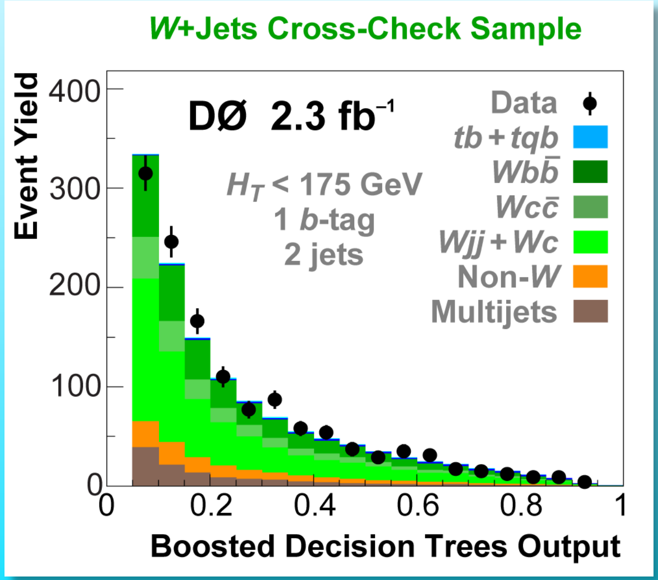
Avoids the problematic situation seen here

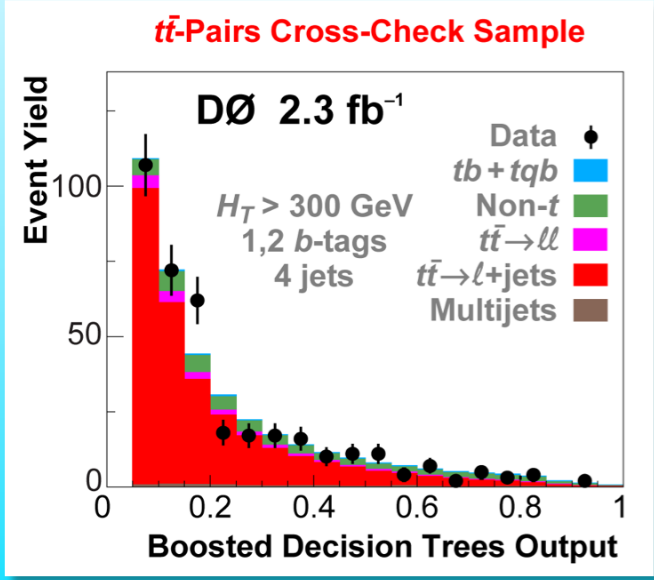
Boosted Decision Trees Performance

- Test whether the BDTs can reproduce a given signal cross section
- Use ensembles of pseudo-datasets containing:
  - Fully simulated background Poisson-sampled from the model
  - Single top quark signals at different input cross sections
  - All systematic uncertainties
- Highly linear response, almost no offset



Background Model Checks





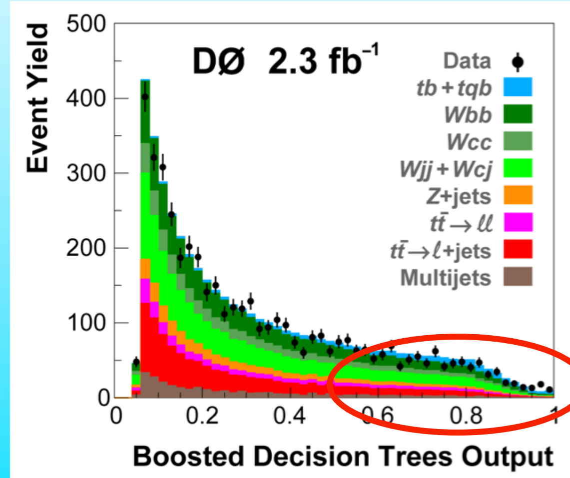
- We test the shapes and normalization of the two main components of the background using two cross-check samples:
  - "W+jets" – exactly 2 jets, exactly 1 b-tag, $H_T < 175$ GeV
  - "$t\bar{t}$ pairs" – exactly 4 jets, 1 or 2 b-tags, $H_T > 300$ GeV
- Good agreement between background and data is seen in all input variables and in the BDT output distributions shown above

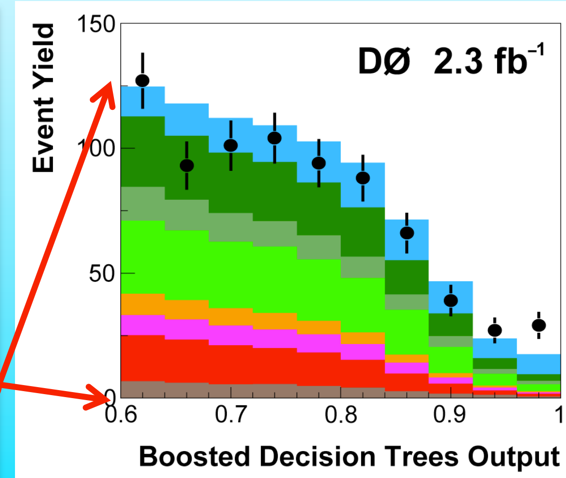
Boosted Decision Tree Results

- Final discriminant output distribution

  (all 24 analysis channels summed in the plots, for illustration only)

- Expected significance = 4.3 σ

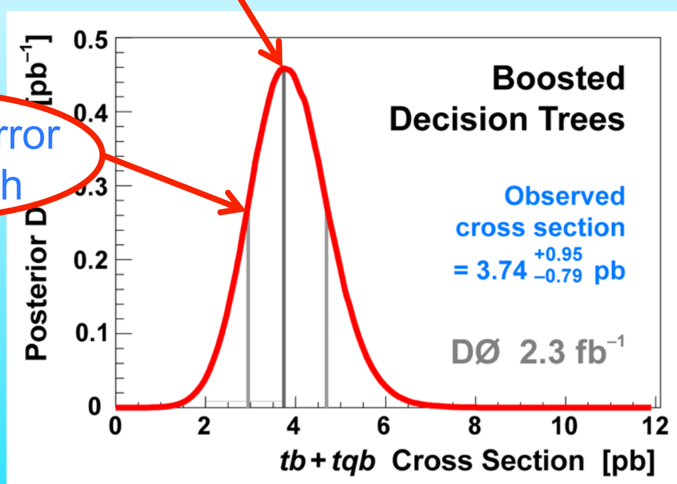




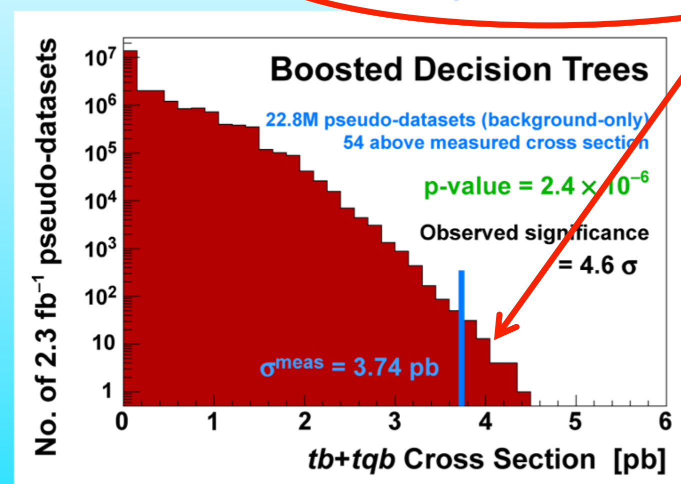
Measure cross section from peak position

Measure error from width

Measure significance by counting pseudo-datasets



- Posterior density distribution



- Significance measurement

Summary

- Boosted decision trees are a powerful tool for separating signal from background

- We measure

$$\sigma_{\mathrm{BDT}}\left(p\bar{p} \rightarrow tb + X,\ tqb + X\right) = 3.74\ ^{+0.95}_{-0.79}\ \text{pb}$$

for m_{top} = 170 GeV

- The probability that the background fluctuated up to fake the signal is 1 in 455,000, giving a signal significance of 4.6 σ