Search for Single Top Quark Production Using Bayesian Neural Networks

Daekwang Kau

Florida State University

PHENO May 7, 2007

Outline

- Single top quark production
- Backgrounds
- Event selection
- Analysis strategy
- Bayesian neural networks (BNN)
- Cross section measurement
- BNN optimization
- Significance of the signal
- Conclusions

Single top quark production

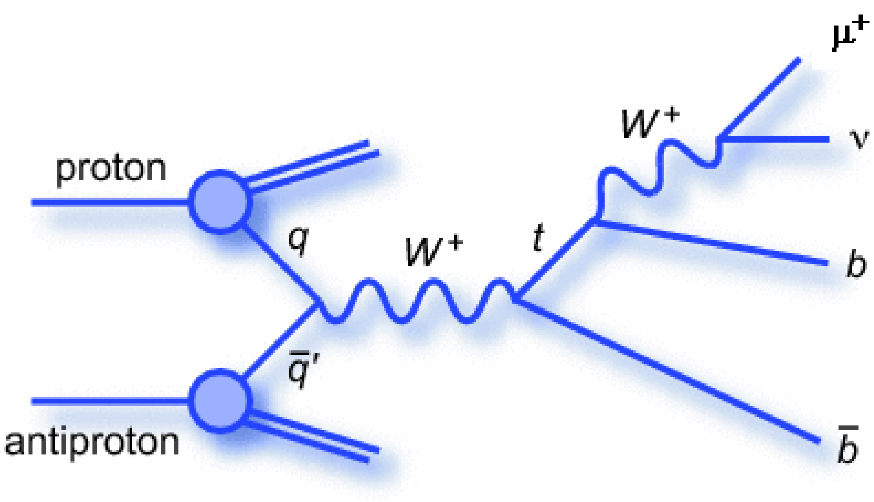
Final states of signal include:

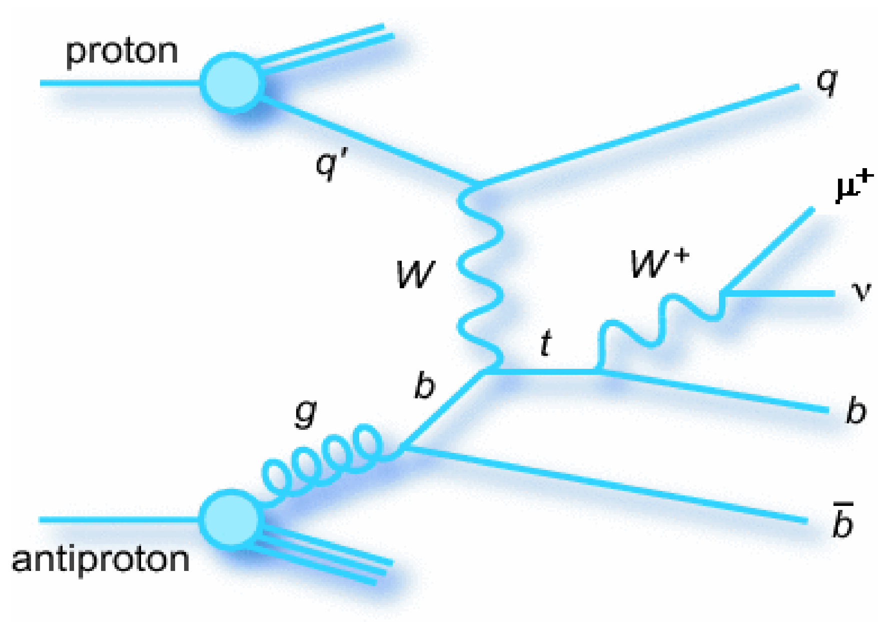
s-channel

One high p_T lepton, $\not{E}_T$,
2 high p_T b-jets

t-channel

One high p_T lepton, $\not{E}_T$,
1 high p_T b-jet, high p_T light jet





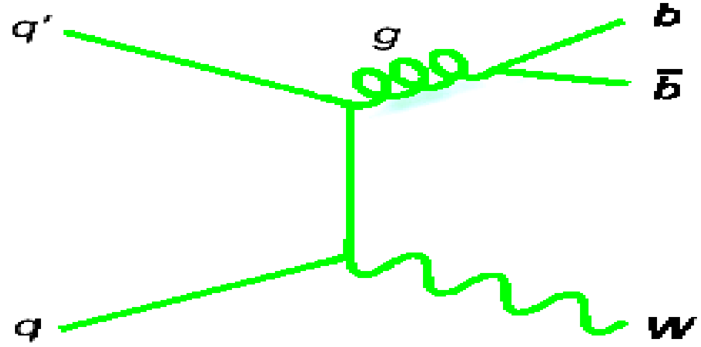
$\sigma_{s-channel}$ 0.88pb ± 8%

$\sigma_{t-channel}$ 1.98pb ± 25%

Zack Sullivan et al. PRD 66 (02) 054024

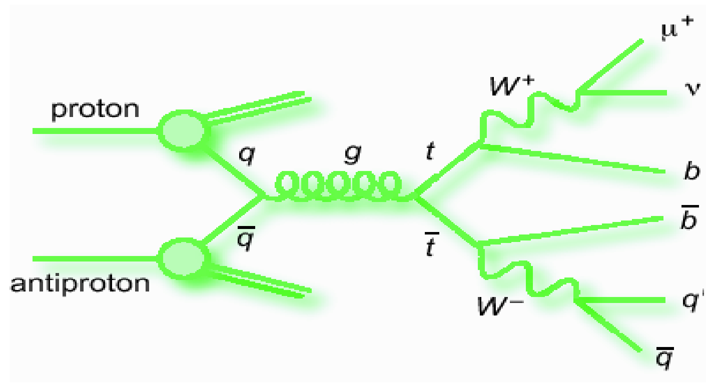
Signal MC samples are generated with CompHEP-SingleTop generator
Partons are hadronized using Pythia

Backgrounds



W+Jets

Contain a lepton from W decay
Estimate shapes from MC
Normalize to data

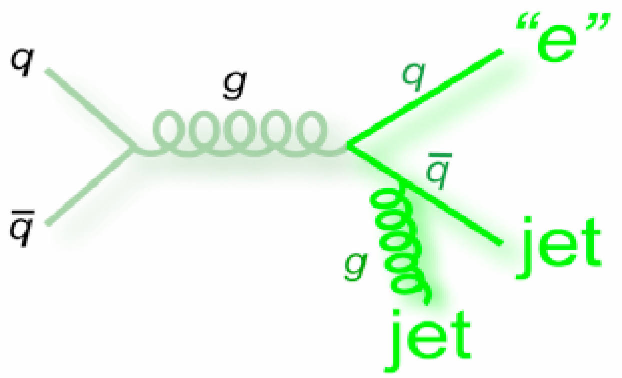


Top-pair production

Lepton + jets
- One W decays hadronically
- One W decays leptonically

Dilepton
- Both Ws decay leptonically

Estimate from MC



Multijet events

Contain fake isolated lepton
Estimate from data

Event selection

- **Purpose**
  - **Select W events**
  - **Maximize acceptance for the signal**

- **Requirements**
  - **1 isolated muon p_T > 18 GeV or electron p_T > 15 GeV**
  - **$\not{E}_T$ > 15 GeV**
  - **Jets: p_T(jet1) > 25 GeV, $|\eta| < 2.5$, pT(jet2) > 20 GeV, $|\eta| < 3.4$, p_T(others) > 15 GeV, $|\eta| < 3.4$**

- **B-tagging**
  - **One or two jets tagged as b-jets**

Event selection

Divide the dataset into independent channels

Fractions of expected signal and S:B ratios are significantly different

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

S:B ratio is at most 9%

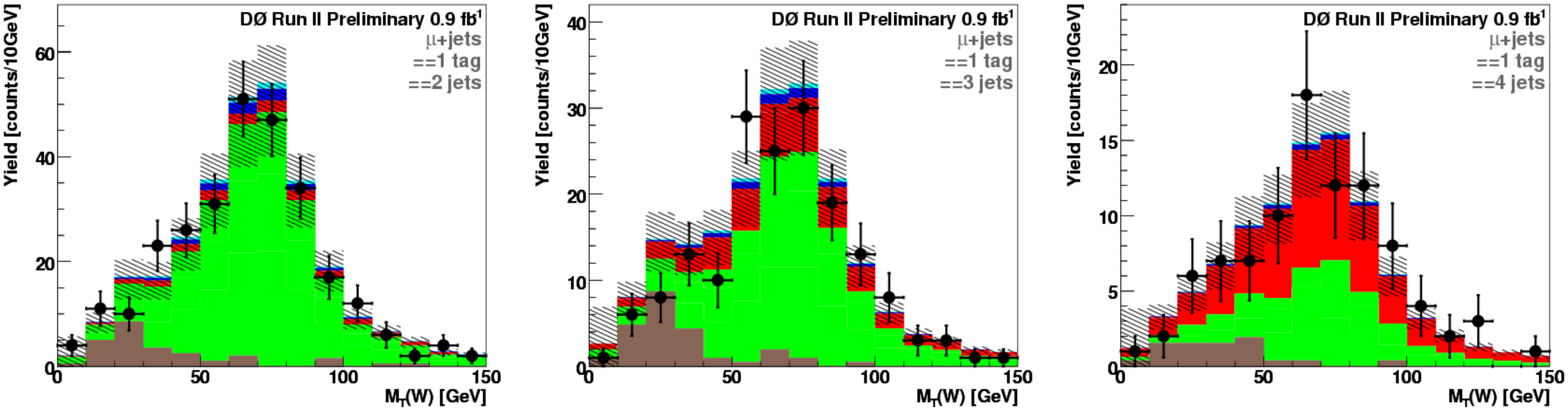
More sophisticated techniques are required for further signal and background discrimination

Event selection

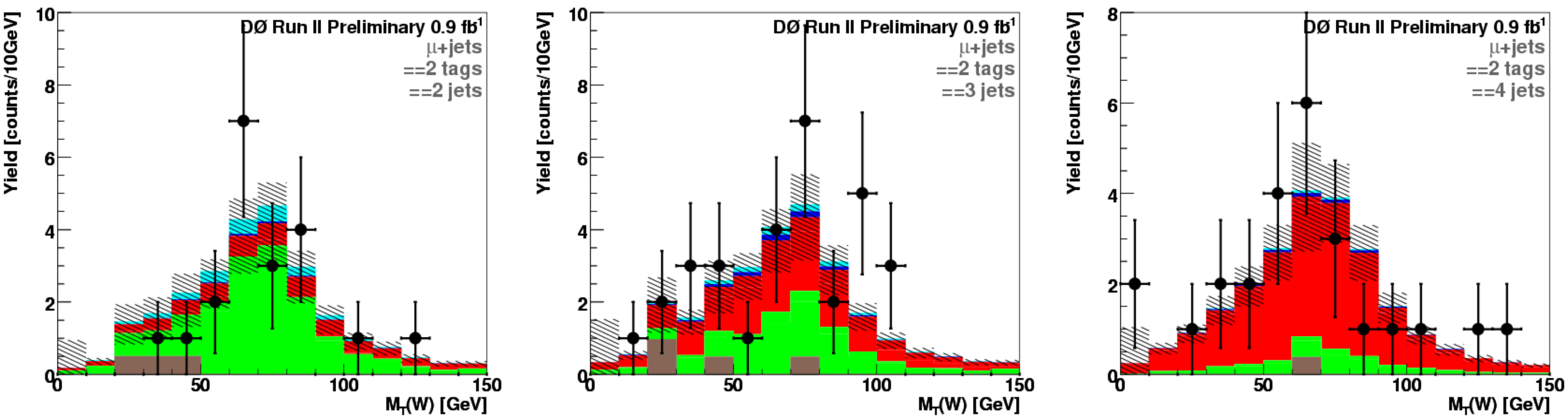
(light) blue : signal
red: ttbar
green: W+jets
gray: multijet

W transverse mass (muon)

1 tag



2 tags



2 jets 3 jets 4 jets

Analysis strategy

Event selection (12 channels)

↓ ↓ ↓

Decision trees (DT) | **Bayesian neural networks (BNN)** | **Matrix element (ME)**

↓

Selecting variables

↓

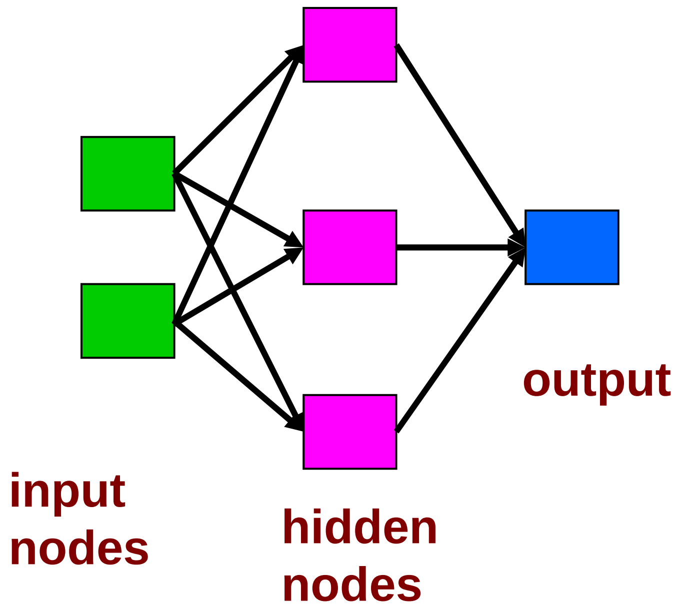
Training BNN

↓

Cross section measurement

Bayesian neural networks

schematic of
neural networks



Neural networks are non-linear functions that can map a vector of n real-valued inputs

$$X = (x_1, x_2, ..., x_n), t$$

into one output $y(X; W)$

Train networks with N events

$$(X_1, t_1), (X_2, t_2), ..., (X_N, t_N)$$

Make predictions for new events

$$(X_{N+1}, X_{N+2}, ...)$$

Target t is a binary classification label (1,0)

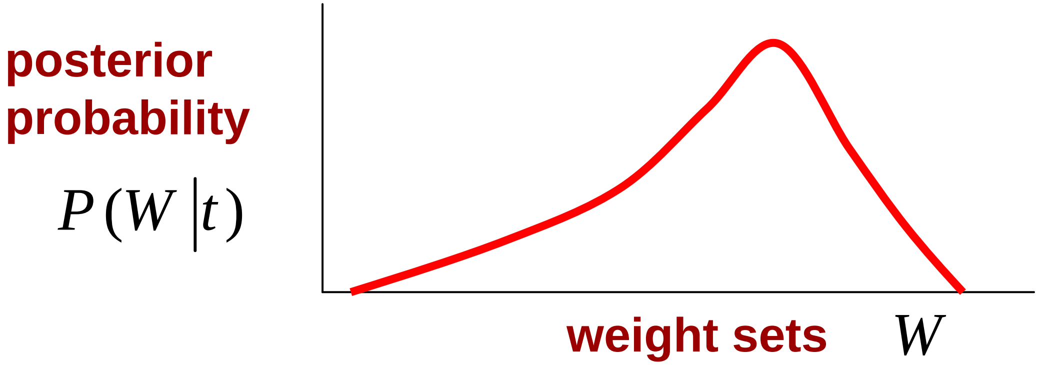
1 : signal

0 : background

Bayesian neural networks

Find posterior probability density
over all possible weight sets

$$P(W|t) = \frac{P(t|W)P(W)}{P(t)}$$



To make prediction for new events

$$\bar{y}(X) = \int y(X;W)P(W|t)\,dW$$

$$\cong \frac{1}{M}\sum_{i=1}^{M} y(X;W_i)$$

Bayesian neural networks

posterior **likelihood** **prior** ← **> 400 dimensional Gaussian distribution**

$$P(W|t) = \frac{P(t|W)P(W)}{P(t)}$$

Not possible to calculate the posterior density analytically

Draw a sample of points W_i from the posterior density P(*W*|t) using a Markov Chain Monte Carlo method

BNN training

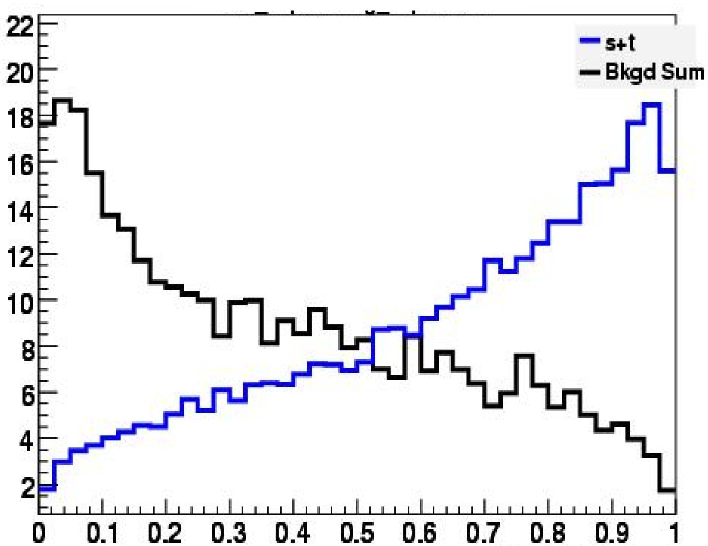
- **24 input variables in three categories**

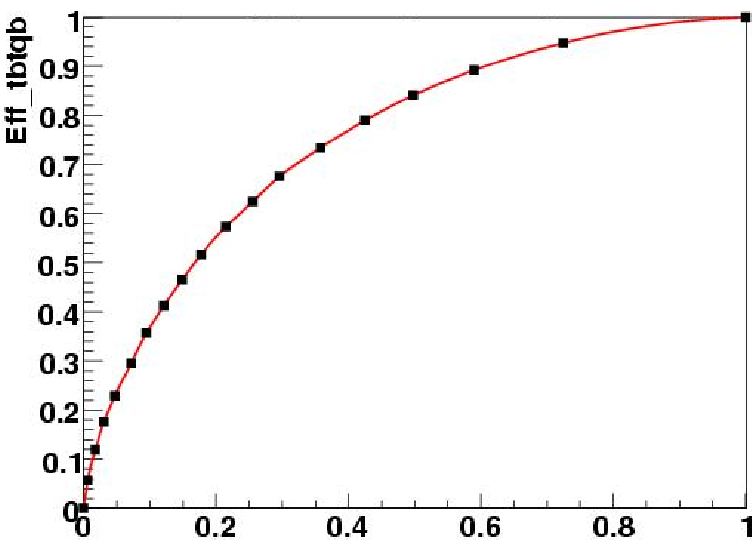
  Individual object kinematics

  Global event kinematics

  Angular variables

- **The number of hidden nodes = 40**

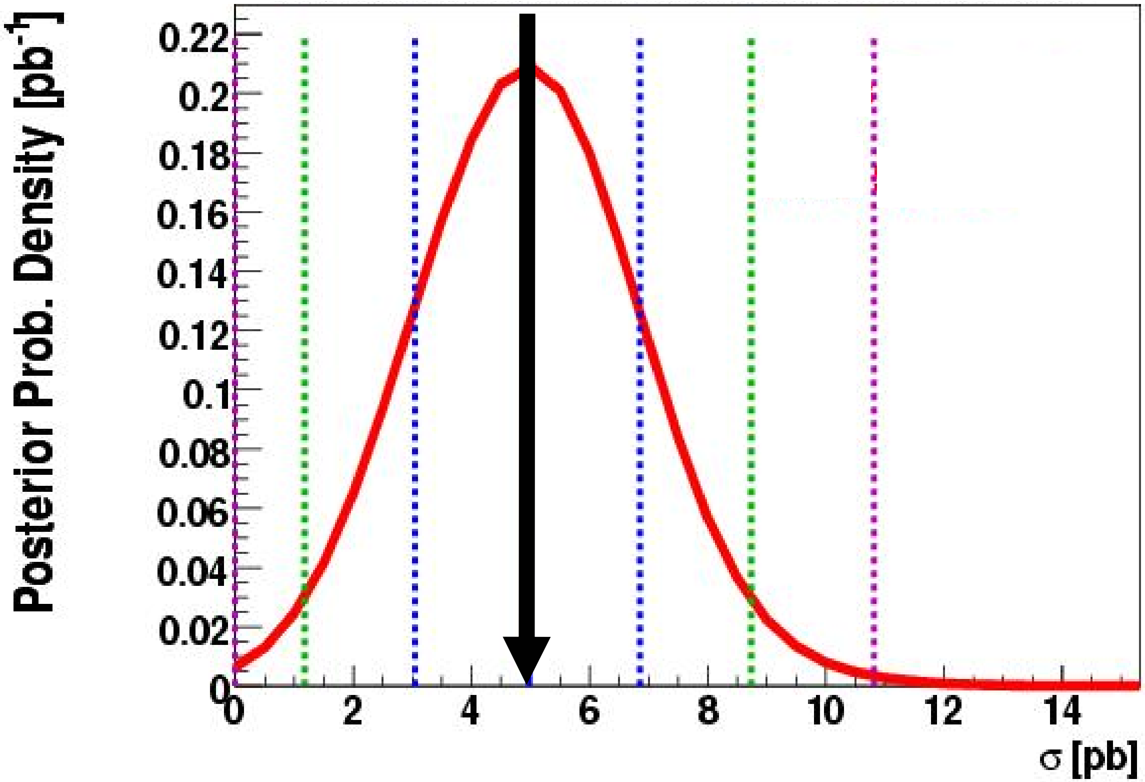




Cross section measurement

$$Posterior(d \mid D) \equiv P(\sigma, a, b \mid D) \propto likelihood(D \mid d)\, prior(d)$$

$$d = \varepsilon L \sigma + \sum_{i=1}^{n} b_i \equiv a\sigma + \sum_{i=1}^{n} b_i$$



Final result

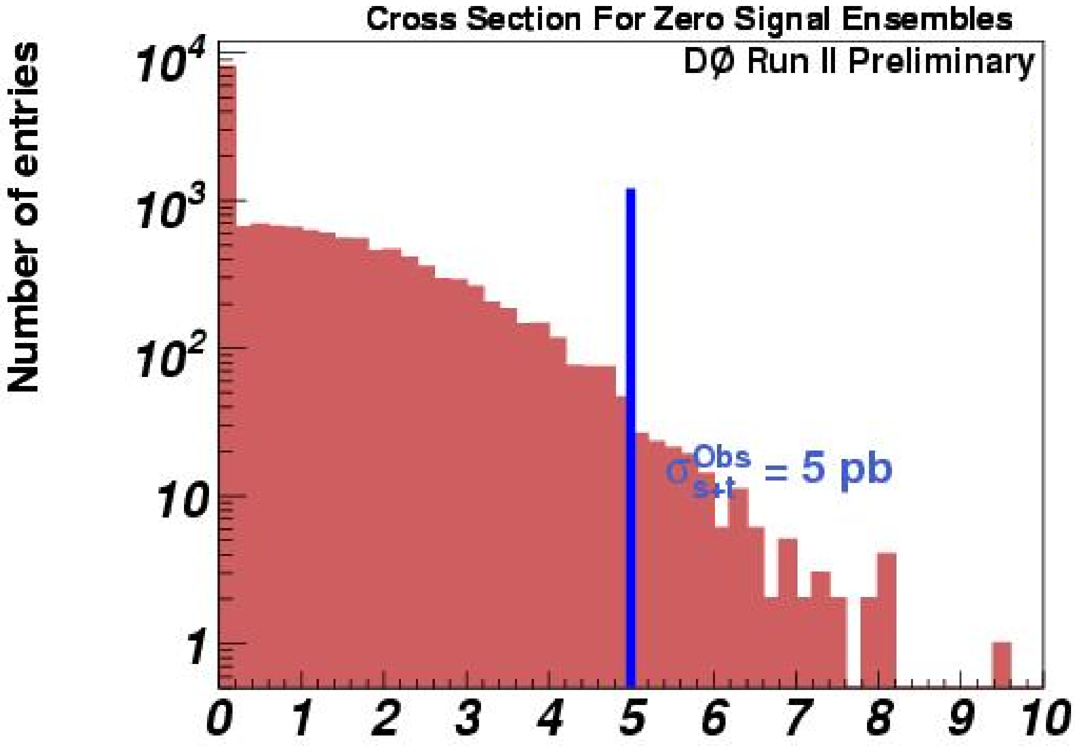
$\sigma_{s+t} = 5.0 \pm 1.9$pb

Significance of signal

Generate 80,000 background only ensemble

Test static : cross section

p-value : probability to measure a cross section equal to or higher than reference value



expected

p-value: 0.0097

Standard deviation: 1.3

observed

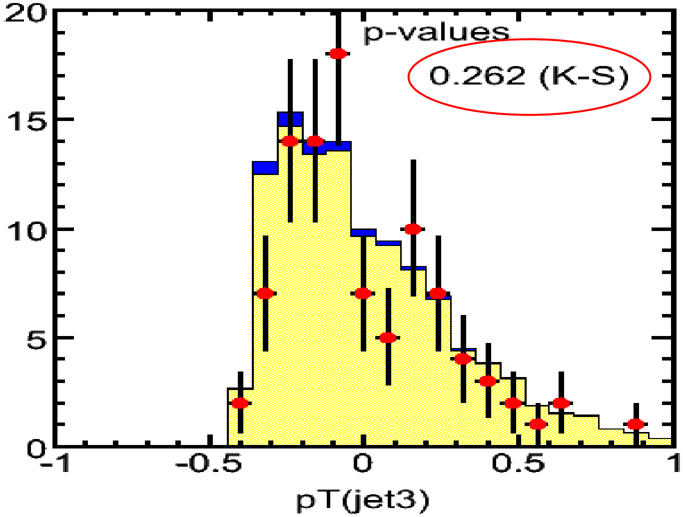
p-value: 0.009

Standard deviation: 2.4

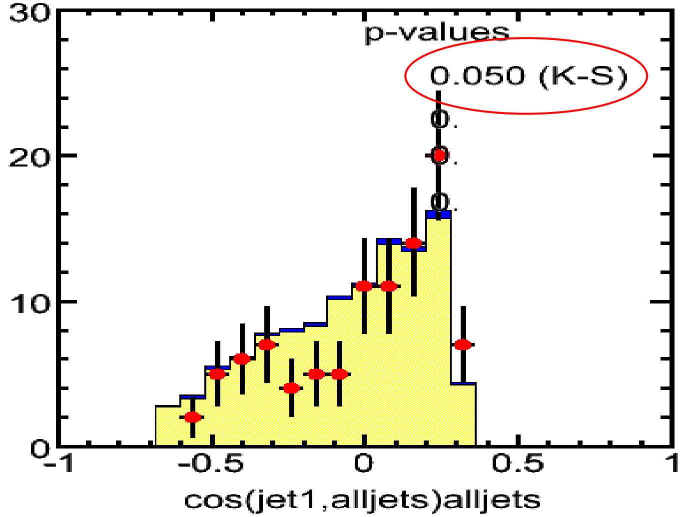
BNN optimization

Check modelling of variables

Compute distribution of a discrepancy measure similar to K-S test



selected



rejected

Order the variables according to their importance

Build discrimination rules using sequence of if statements and measure how often a particular variable appears

Select best ~ 20 variables out of ~ 50

BNN optimization

Long tail in event weight introduces noise in training sample

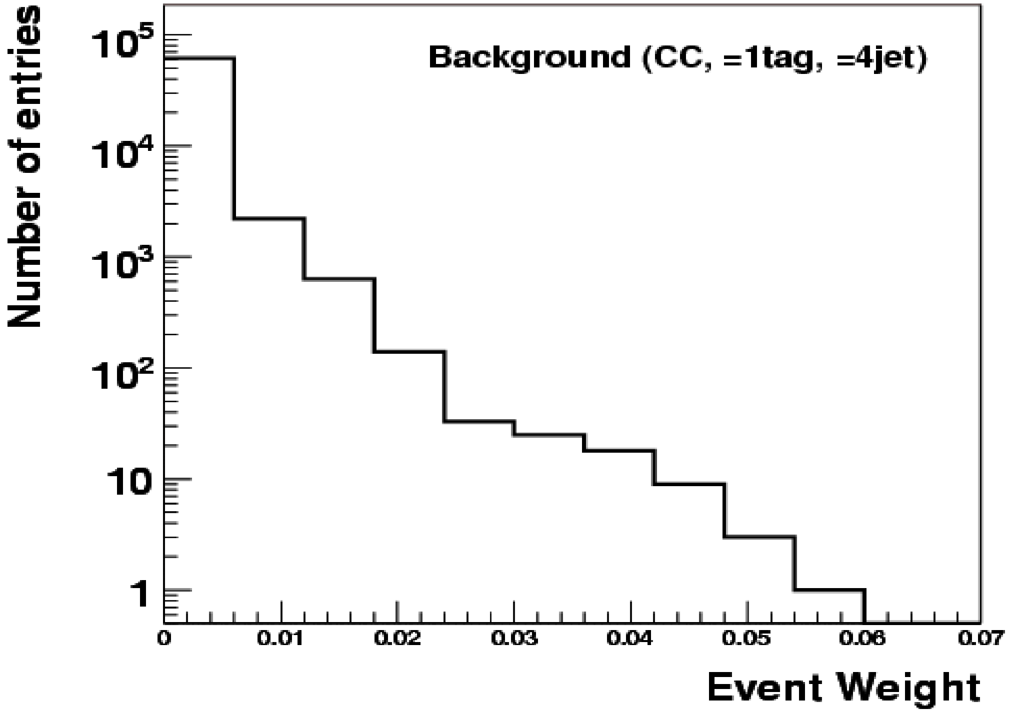
- **Old BNN**

Width of priors allowed to adapt to noise level in training sample

But large event weights caused prior widths to become too large and led to poorer BNN performance than expected

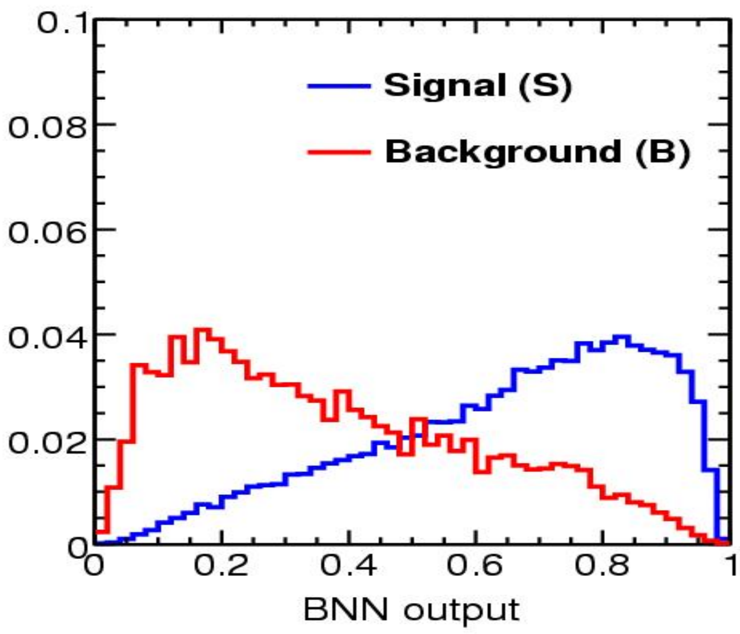
- **New BNN**

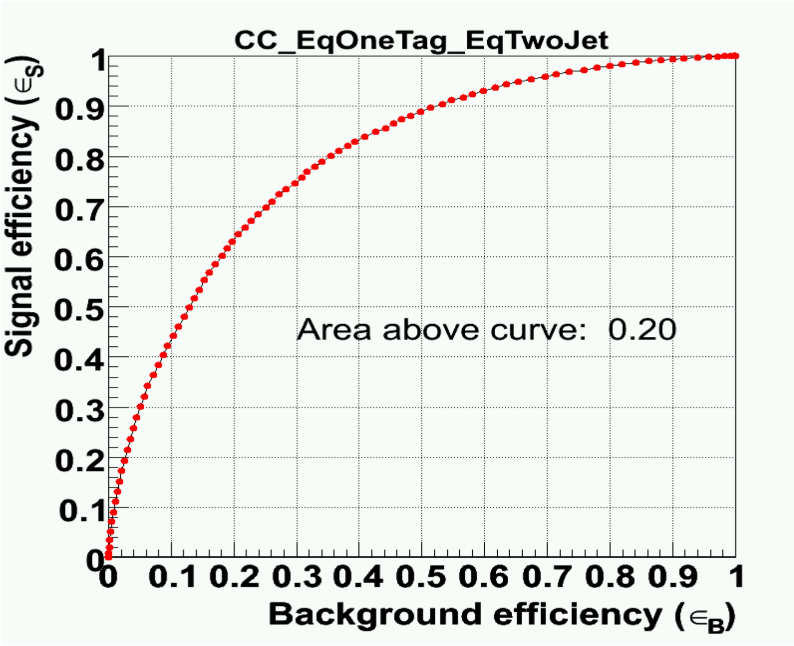
Set the width of the prior on each neural network parameter to a small fixed value



BNN optimization

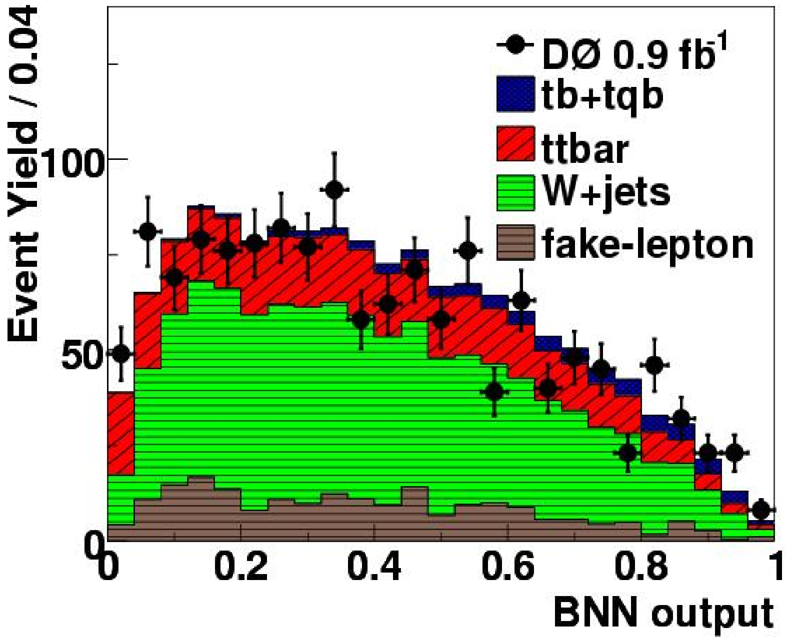
- Number of inputs ~ 20 and the number of hidden nodes = 20
- Training sample: 10,000 signal + 10,000 background



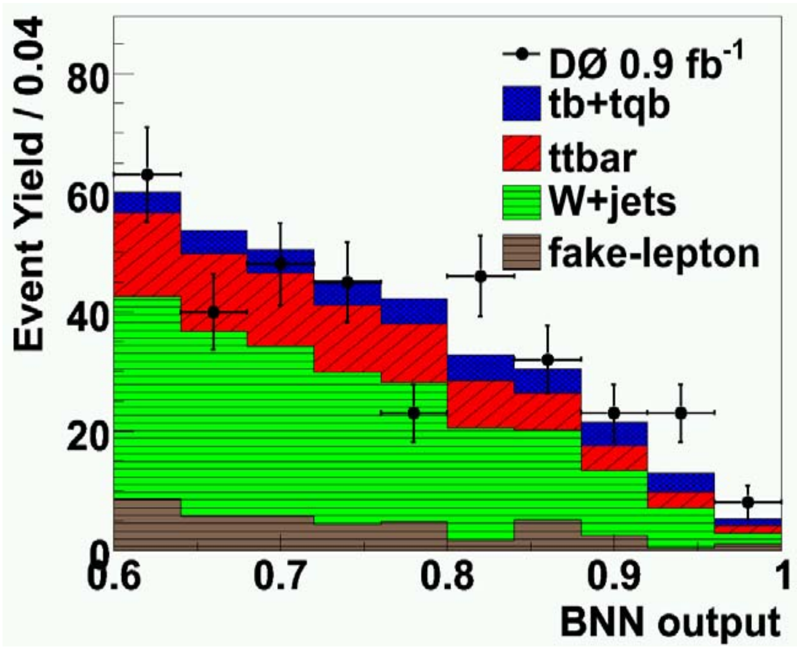


BNN optimization

Output of optimized BNN



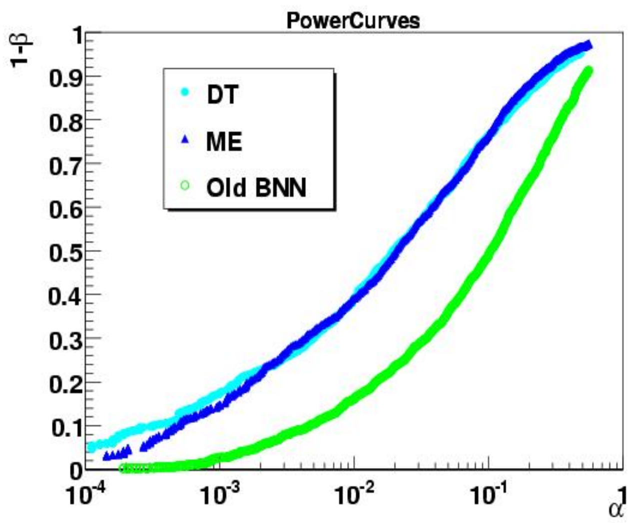
Sum of all 12 channels

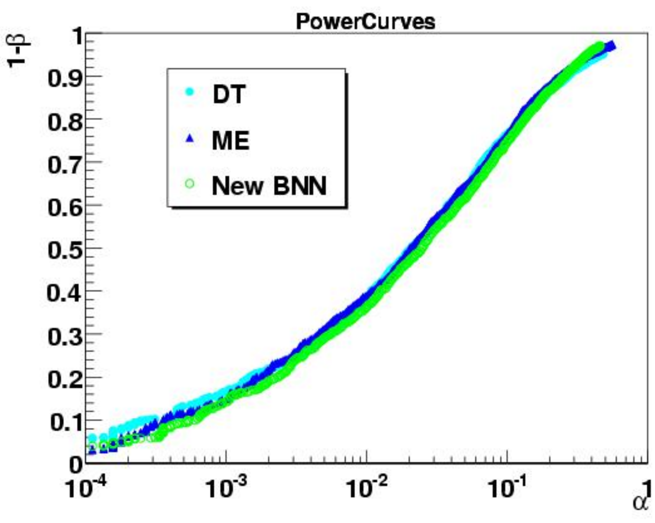


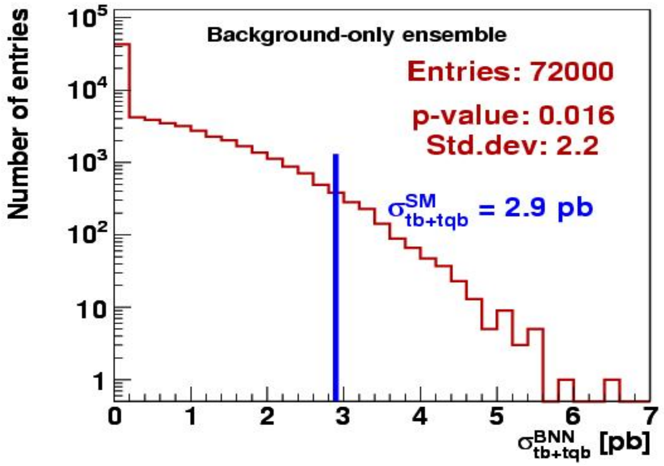
Near a BNN output of 1

Significance of signal

The p-value computed from the SM signal + background ensemble (y-axis) versus the p-value computed from the background-only ensemble (x-axis)







Significance of

expected signal

(new BNN)

Conclusions

We measure a single top quark production cross section of 5.0 ± 1.9pb

This analysis results in a p-value of 0.9%, corresponding to a 2.4 standard deviation significance

Improved result is in the process of approval

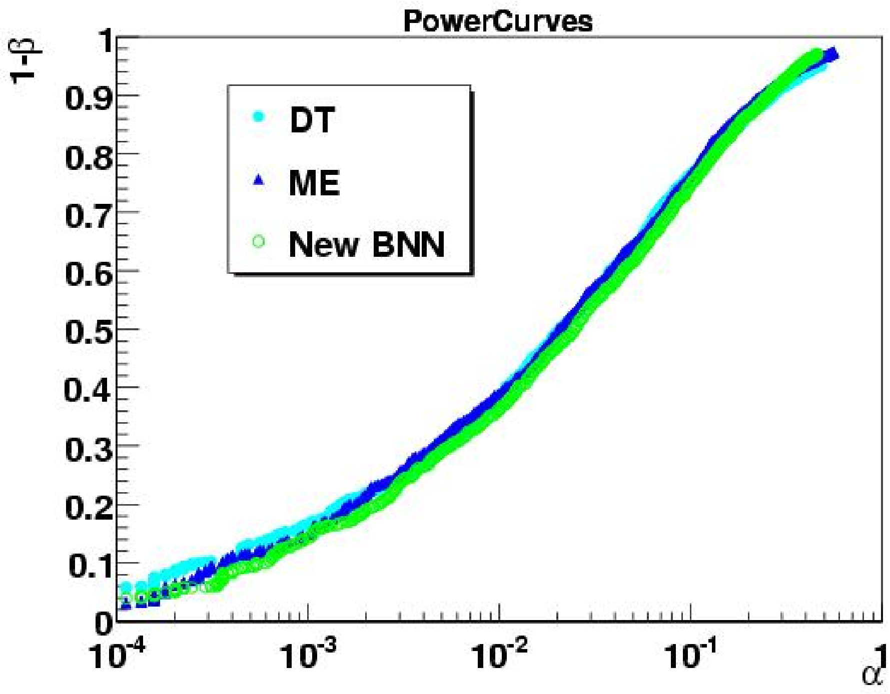
$$\sigma_{s+t} = 5.0 \pm 1.9\text{pb}$$

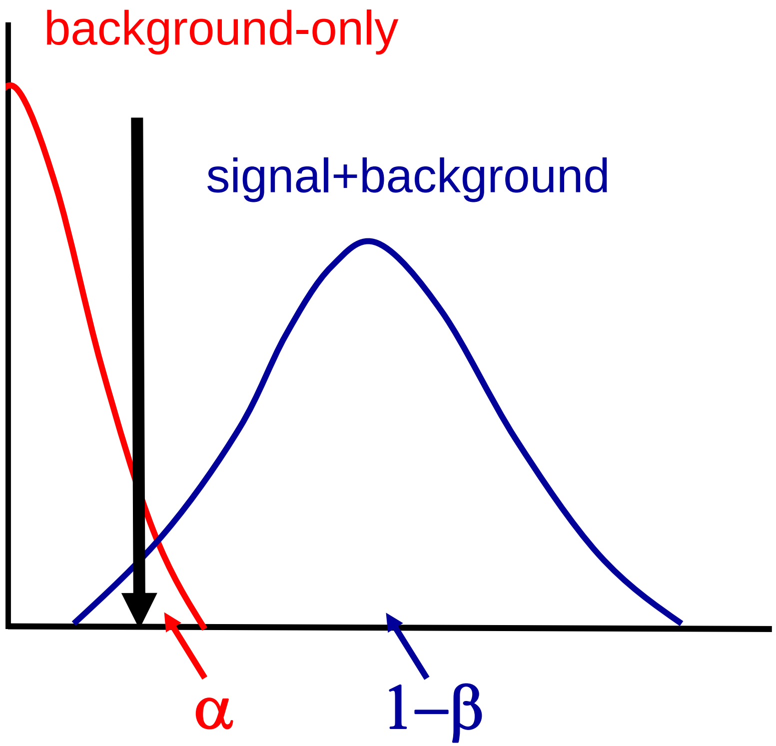
2.4 standard deviation significance

Back Up

Significance of signal

The p-value computed from the SM signal + background ensemble (y-axis) versus the p-value computed from the background-only ensemble (x-axis)





0 ~ 10 pb

Cross section measurement

Bayesian posterior probability density

$$Posterior(d \mid D) \equiv P(\sigma, a, b \mid D) \propto likelihood(D \mid d)\, prior(d)$$

$$d = \varepsilon L \sigma + \sum_{i=1}^{n} b_i \equiv a\sigma + \sum_{i=1}^{n} b_i$$

$$P(D \mid d) = P(D \mid \sigma, a, b) = \prod_{j=1}^{bins} \frac{\exp(-d_j)\, d_j^{D_j}}{\Gamma(D_j + 1)}$$

$$P(\sigma \mid D) \propto \iint likelihood(D \mid d)\, prior(d)\, da\, db$$