



## Measuring the Top Pair Background to the Higgs Search in the MET+b-jets Channel with $L=5.7 \text{ fb}^{-1}$

The CDF Collaboration  
URL <http://www-cdf.fnal.gov>  
(Dated: July 19, 2010)

We present the first measurement of the top pair production cross section in events with large missing transverse energy, two or three high- $p_T$  jets, where at least one is identified as a b-jet. A veto on loosely identified electrons and muons is applied. We reduce the dominant QCD multijet background using neural network techniques, and then use another neural network to isolate the top pair signal from the remaining backgrounds. Analyzing  $5.7 \text{ fb}^{-1}$  of data, we measure top pair production cross section with  $\sigma_{t\bar{t}} = 7.12_{-1.12}^{+1.20} (\text{stat.} + \text{syst.}) \text{ pb}$ .

*Preliminary Results for Summer 2010 Conferences*

## I. INTRODUCTION

Top quark has been discovered in 1995 at Tevatron by CDF and D0 experiments. The top pair production cross section has been measured in lepton+jets, dilepton and all-hadronic decay modes. We measure the top pair production cross section in events with large missing transverse energy ( $\cancel{E}_T$ ) and 2 or 3 high- $p_T$  jets, where at least one is identified as a b-jet. A veto on loosely identified electrons and muons is applied. This is the first measurement for top pair production with this signature, in which  $\sim 66\%$  events decay in lepton+jets mode, and the other 34% decays in dilepton mode. As a complementary to existing measurements, it can be combined with those to achieve greater precision and test more stringently QCD NLO predictions.

MET+b-jets channel is also a very interesting channel to searches for new physics. For example, it is one of the most sensitive channel for low mass Higgs search at Tevatron [1], as well as other searches like SUSY/leptoquark analyses. And top pair is a significant background in these physics searches. In this analysis, we employed many state-of-the-art techniques, and followed the strategies which have been used in low mass Higgs search analysis.

To improve the signal-to-background ratio, we select jets identified as originating from b quarks using b-tagging algorithms. Even after these requirements, the ratio is still too low to achieve sensitivity to top pair production. We further exploit the kinematic and topological characteristics of top pair events using neural networks to isolate the signal from dominant QCD background and subsequently from the remaining backgrounds.

## II. EVENT SELECTION

Without identified lepton information, the largest background to MET+jets analyses is the QCD multijet background. QCD multijet has very high production rate at a hadron collider. Although these processes generally do not produce neutrinos, mismeasured jet energy do result in a significant imbalance of transverse energy. Furthermore, QCD b quark production yields neutrinos whenever one b-hadron decays semi-leptonically, thus giving additional missing transverse energy. Due to the mismeasurement, most such events will have MET align with their 2<sup>nd</sup> or 3<sup>rd</sup> jet. Thus we are using these topological characteristics ( $\Delta\phi(\cancel{E}_T, j_{2,3}) > 0.4$  and  $\Delta\phi(\cancel{E}_T, j_1) > 1.5$ ) to reduce a large part of QCD multijet at the first stage. The event yields are shown in Table I.

The b-tagging algorithms we are using in this analysis are SECVTX [2] and JETPROB [3]. SECVTX is a b-tagging algorithm based on secondary vertex reconstruction, and JETPROB however tags a jet depending on the probability that all tracks associated with a jet come from the primary interaction vertex.

We accept events with exactly one SECVTX-tagged jet (1S), two SECVTX-tagged jets (SS), and one SECVTX-tagged and one JETPROB-tagged jet (SJ).

| <i>CDF Run II Preliminary <math>L = 5.7fb^{-1}</math></i> |                     |                  |                  |
|---|---------------------|------------------|------------------|
|   | 1S                  | SS               | SJ               |
| Single Top  | 236.5 $\pm$ 40.7    | 41.8 $\pm$ 8.1   | 34.1 $\pm$ 6.7   |
| Diboson   | 128.8 $\pm$ 17.1    | 15.3 $\pm$ 2.7   | 13.8 $\pm$ 2.4   |
| W+LF  | 869.3 $\pm$ 267.3   | 6.5 $\pm$ 2.4    | 26.4 $\pm$ 8.8   |
| W+HF  | 610.1 $\pm$ 187.5   | 50.1 $\pm$ 16.9  | 59.4 $\pm$ 19.5  |
| Z+LF  | 241.6 $\pm$ 74.3    | 3.1 $\pm$ 1.1    | 8.2 $\pm$ 2.7    |
| Z+HF  | 278.9 $\pm$ 85.7    | 33.2 $\pm$ 10.9  | 32.9 $\pm$ 10.7  |
| Multijet  | 7819.6 $\pm$ 469.2  | 215.3 $\pm$ 35.1 | 586.8 $\pm$ 75.7 |
| Top Pair  | 536.2 $\pm$ 35.6    | 120.7 $\pm$ 12.1 | 107.4 $\pm$ 11.5 |
| Tot. Exp.   | 10721.0 $\pm$ 585.5 | 486.0 $\pm$ 43.2 | 869.0 $\pm$ 80.6 |
| Data  | 10721               | 486              | 869              |

TABLE I: Acceptance table in Preselection.

After preselection, the multijet production is still the dominant background. We thus use a neural network approach to further reduce the QCD multijet. Neural network technique exploits the correlations among the many observables which provide discrimination between signal and backgrounds. In this analysis, we use 15 variables as the QCD Neural Network inputs, which inherit from the previous Single Top analysis [4]. The distribution for these 15 variables after preselection is shown in Fig 2. The NN output is shown in Fig. 3.

We choose  $NN_{QCD} > -0.5$  to define the signal region. Table III shows the event yields in Signal Region.

| Variable   | Description  |
|--|--|
| $\cancel{E}_T$                                   | Absolute amount of the missing transverse energy   |
| $\cancel{p}_T$                                   | Absolute amount of the missing transverse momentum   |
| $\cancel{E}_T / \sqrt{\sum E_T}$                 | Missing $E_T$ significance   |
| $\cancel{E}_T / H_T$                             | Ratio of $\cancel{E}_T$ to $H_T$   |
| $\cancel{H}_T / \cancel{E}_T$                    | Ratio of $\cancel{H}_T$ to $\cancel{E}_T$  |
| $M(\vec{\cancel{E}}_T, \vec{j}_1, \vec{j}_2)$    | Invariant mass of $\vec{\cancel{E}}_T$ , $\vec{j}_1$ and $\vec{j}_2$                         |
| $\Delta\phi(\vec{\cancel{E}}_T, \cancel{p}_T)$   | Azimuthal difference between $\vec{\cancel{E}}_T$ and $\cancel{p}_T$                         |
| $Max(\Delta\phi(\vec{j}_i, \vec{j}_k))$          | Maximum of $\Delta\phi$ between any two jets $\vec{j}_i, \vec{j}_k$                          |
| $Max(\Delta R(\vec{j}_i, \vec{j}_k))$            | Maximum of $\Delta R$ between any two jets $\vec{j}_i, \vec{j}_k$                            |
| $Min(\Delta\phi(\vec{\cancel{E}}_T, \vec{j}_i))$ | Minimum of $\Delta\phi$ between $\vec{\cancel{E}}_T$ and any jet $\vec{j}_i$                 |
| $Min(\Delta\phi(\cancel{p}_T, \vec{j}_i))$       | Minimum of $\Delta\phi$ between $\cancel{p}_T$ and any jet $\vec{j}_i$                       |
| $\phi^*$   | $\Delta\phi$ of $(j^1, j^2)$ axis in their rest frame, and their vector sum in the lab frame |
| Sphericity                                       | $S = \frac{3}{2}(\lambda_2 + \lambda_3)^a$   |
| $\sum p_T^{chg d} / p_T^{j1}$                    | Fraction of $p_T^{j1}$ carried by charged particles displaced from the primary vertex        |
| $\sum p_T^{chg d} / p_T^{j2}$                    | Fraction of $p_T^{j2}$ carried by charged particles displaced from the primary vertex        |

<sup>a</sup>A momentum tensor is defined as  $M_{lm} = \frac{\sum_o j_l^o j_m^o}{\sum_o |j^o|}$ , where  $j^o$  is the momentum of a reconstructed jet, and  $l$  and  $m$  are Cartesian coordinates. The index  $o$  runs over the number of jets in the event. The sphericity in an event is defined as  $S = \frac{3}{2}(\lambda_2 + \lambda_3)$ , where  $\lambda_2$  and  $\lambda_3$  are the smallest two eigenvalues of the normalized momentum tensor.

TABLE II: Input variables to the neural network devised to suppress the multijet background.

| CDF Run II Preliminary $L = 5.7\text{fb}^{-1}$ |                |              |              |
|--|----------------|--------------|--------------|
|  | 1S             | SS           | SJ           |
| Single Top                                     | 195.7 ± 33.7   | 39.0 ± 7.5   | 30.4 ± 6.0   |
| Diboson  | 95.7 ± 12.7    | 13.8 ± 2.4   | 11.5 ± 2.0   |
| W+LF   | 494.7 ± 152.2  | 5.1 ± 1.9    | 17.9 ± 6.1   |
| W+HF   | 405.0 ± 124.5  | 41.7 ± 14.1  | 44.6 ± 14.6  |
| Z+LF   | 155.9 ± 47.9   | 2.6 ± 0.9    | 6.1 ± 2.0    |
| Z+HF   | 184.2 ± 56.6   | 27.6 ± 9.1   | 25.1 ± 8.2   |
| Multijet                                       | 1732.0 ± 103.9 | 93.6 ± 15.3  | 145.1 ± 18.7 |
| Top Pair                                       | 503.4 ± 33.4   | 117.5 ± 11.8 | 102.9 ± 11.0 |
| Tot. Exp.                                      | 3766.6 ± 239.5 | 340.9 ± 26.8 | 383.6 ± 28.9 |
| Data   | 3814           | 290          | 401          |

TABLE III: Acceptance table in Signal Region.

### III. FINAL DISCRIMINANT

Though a good signal/background ratio has been achieved by using the  $NN_{QCD}$ , we are still interested to develop a final discriminant to separate signal bins and background bins as much as possible, since we know backgrounds like W/Z+jets have very large uncertainties. These uncertainties can worsen the measurement, especially when we are using a binned likelihood technique to measure its cross section.

We use Neural Network again to develop the final discriminant. The 5 NN input variables are shown in Fig. 4, and the final discriminant ( $NN_{sig}$ ) is shown in Fig. 5.

### IV. SYSTEMATIC UNCERTAINTIES

Systematic uncertainties are split in normalization uncertainties and shape uncertainties. The normalization uncertainty reflects changes to the event yield due to the systematic effect while the shape uncertainty reflects changes to the template histograms. Both of these effects can be included, depending on the source of the systematic uncertainty.

The normalization uncertainties are summarized in Table V VI VII. We also assign a systematic shape uncertainty to the multijet model due to possible signal contamination.

| Variable                                      | Description  |
|---|--|
| $H_T$   | Scalar sum of the jet energies                                 |
| $\cancel{E}_T$                                | Missing transverse energy                                      |
| $M(\vec{\cancel{E}}_T, \vec{j}_1, \vec{j}_2)$ | Invariant mass of $\cancel{E}_T$ , $\vec{j}_1$ and $\vec{j}_2$ |
| $M(j_1, j_2)$                                 | Invariant mass of two leading jets                             |
| $NN_{QCD}$                                    | output of the QCD removal NN after the cut $NN_{QCD} > -0.5$   |

TABLE IV: Input variables to the neural network aimed at discriminating  $t\bar{t}$  production from the backgrounds remaining after the  $NN_{QCD} > -0.5$  requirement.

| <i>CDF Run II Preliminary <math>L = 5.7\text{fb}^{-1}</math></i> |             |                               |              |              |             |             |
|--|-------------|-------------------------------|--------------|--------------|-------------|-------------|
| Syst.  | $t\bar{t}$  | Single Top                    | Diboson      | W+jets       | Z+jets      | Multijet    |
| Luminosity   | $\pm 6\%$   | $\pm 6\%$                     | $\pm 6\%$    | $\pm 6\%$    | $\pm 6\%$   | -           |
| PDF  | $\pm 2\%$   | $\pm 2\%$                     | $\pm 2\%$    | $\pm 2\%$    | $\pm 2\%$   | -           |
| Lepton veto  | $\pm 2\%$   | $\pm 2\%$                     | $\pm 2\%$    | $\pm 2\%$    | $\pm 2\%$   | -           |
| Cross-section  | -           | $\pm 15.9\%(s) \pm 15.2\%(t)$ | $\pm 11.5\%$ | $\pm 30\%$   | $\pm 30\%$  | -           |
| JES  | $\pm 0.8\%$ | $\pm 4.6\%$                   | $\pm 7.0\%$  | $\pm 12.7\%$ | $\pm 8.3\%$ | $\pm 5.6\%$ |
| Trigger Eff.   | $\pm 0.5\%$ | $\pm 0.8\%$                   | $\pm 0.7\%$  | $\pm 0.9\%$  | $\pm 0.7\%$ | $\pm 1.1\%$ |

TABLE V: Systematics uncertainties.

| <i>CDF Run II Preliminary <math>L = 5.7\text{fb}^{-1}</math></i> |           |              |              |
|--|-----------|--------------|--------------|
| Multijet SF  | 1S        | SS           | SJ           |
| Multijet   | $\pm 6\%$ | $\pm 16.3\%$ | $\pm 12.9\%$ |

TABLE VI: Multijet scale factor uncertainties.

| <i>CDF Run II Preliminary <math>L = 5.7\text{fb}^{-1}</math></i> |             |             |             |
|--|-------------|-------------|-------------|
| $t\bar{t}$   | 1S          | SS          | SJ          |
| Color Reconnection   | $\pm 1.6\%$ | $\pm 1.9\%$ | $\pm 1.5\%$ |
| ISR/FSR  | $\pm 2.7\%$ | $\pm 4.0\%$ | $\pm 1.8\%$ |
| Herwig Hadronization   | $\pm 3.9\%$ | $\pm 0.3\%$ | $\pm 1.8\%$ |

TABLE VII: Additional uncertainties for  $t\bar{t}$ .

## V. CROSS SECTION MEASUREMENT AND RESULTS

We use a binned likelihood technique to measure the cross section. The likelihood function  $L$  is given by the product of the likelihood for each of the different sub-tagging categories  $L_c$ , where  $L_c$  is defined as,

$$L_c = \prod_{i=1}^{n_{bins}} P(n_i | \mu_i) = \prod_{i=1}^{n_{bins}} \frac{\mu_i^{n_i} e^{-\mu_i}}{n_i!}. \quad (1)$$

where  $n_i$  is the data count in that particular bin and  $n_{bins}$  is the number of bins in the distribution which is scanned to look for an excess of signal-like events. The prediction in each bin is a sum over signal and background contributions [4]:

$$\mu_i = \sum_{k=1}^{n_{bkg}} b_{ik} + s_i \quad (2)$$

In this analysis, using  $5.7\text{fb}^{-1}$  data collected by CDF at Tevatron, we measure  $\sigma_{t\bar{t}} = 7.12_{-1.12}^{+1.20}$  (*stat.* + *syst.*) pb.

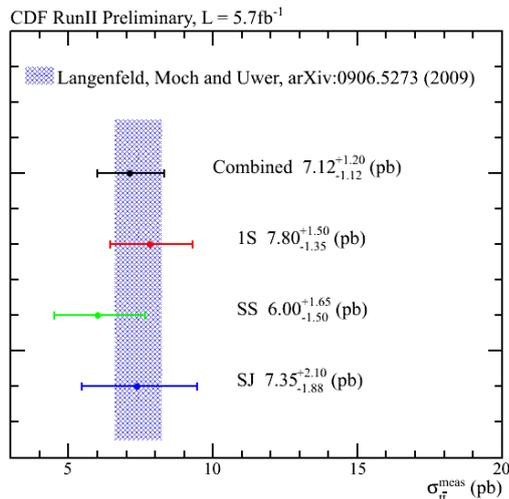


FIG. 1: The result of top pair cross section measurement.

## VI. SUMMARY

In conclusion we report the first measurement of the top pair production cross section of  $\sigma_{t\bar{t}} = 7.12^{+1.20}_{-1.12}$  (*stat.+syst.*) pb, using an exclusive selection of MET+2/3jets decays, analyzing  $\int L dt = 5.7 \text{ fb}^{-1}$  data. The result is complementary to other top pair cross section analyses at CDF experiment. It maintains high sensitivity with respect to  $W \rightarrow \tau\nu$  decays, and is in good agreement with SM calculations and previous measurements.

## Acknowledgments

We thank the Fermilab staff and the technical staffs of the participating institutions for their vital contributions. This work was supported by the U.S. Department of Energy and National Science Foundation; the Italian Istituto Nazionale di Fisica Nucleare; the Ministry of Education, Culture, Sports, Science and Technology of Japan; the Natural Sciences and Engineering Research Council of Canada; the National Science Council of the Republic of China; the Swiss National Science Foundation; the A.P. Sloan Foundation; the Bundesministerium fuer Bildung und Forschung, Germany; the Korean Science and Engineering Foundation and the Korean Research Foundation; the Particle Physics and Astronomy Research Council and the Royal Society, UK; the Russian Foundation for Basic Research; the Comision Interministerial de Ciencia y Tecnologia, Spain; and in part by the European Community's Human Potential Programme under contract HPRN-CT-20002, Probe for New Physics.

- 
- [1] T. Aaltonen et al. (CDF Collaboration), Phys. Rev. Lett. 104, 141801 (2010)
  - [2] A. Acosta et al. (CDF Collaboration), Phys. Rev. D 71, 052003 (2005)
  - [3] A. Abulencia et al. (CDF Collaboration), Phys. Rev. D 74, 072006 (2006)
  - [4] T. Aaltonen et al. (CDF Collaboration), Phys. Rev. D 81, 072003 (2010)

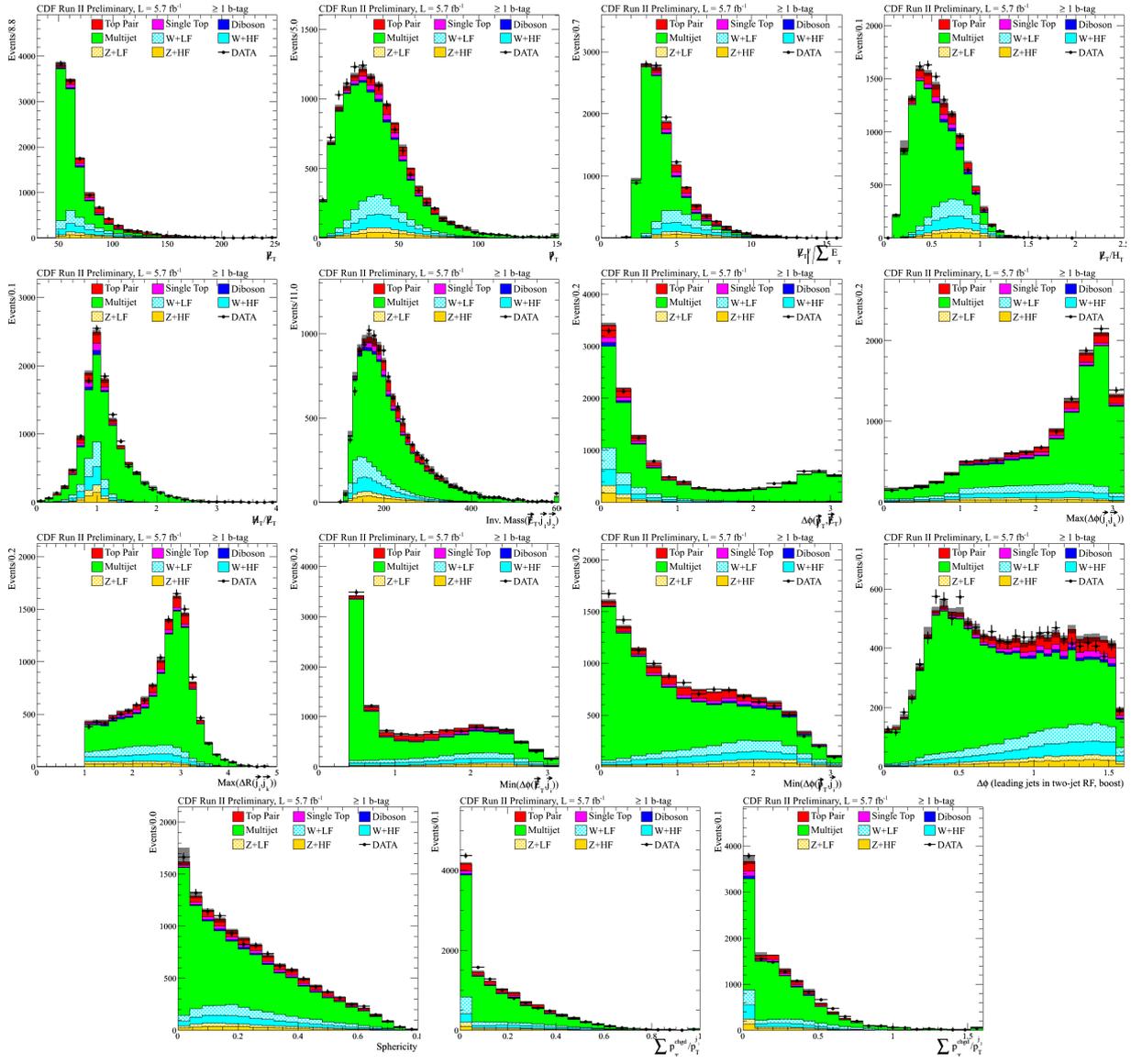


FIG. 2: The distribution of 15 input variables for the  $NN_{QCD}$  after preselection.

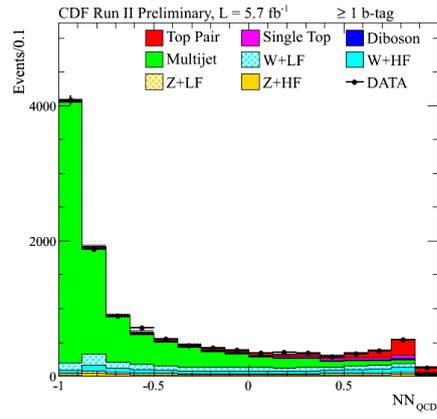


FIG. 3: The  $NN_{QCD}$  distribution after preselection.

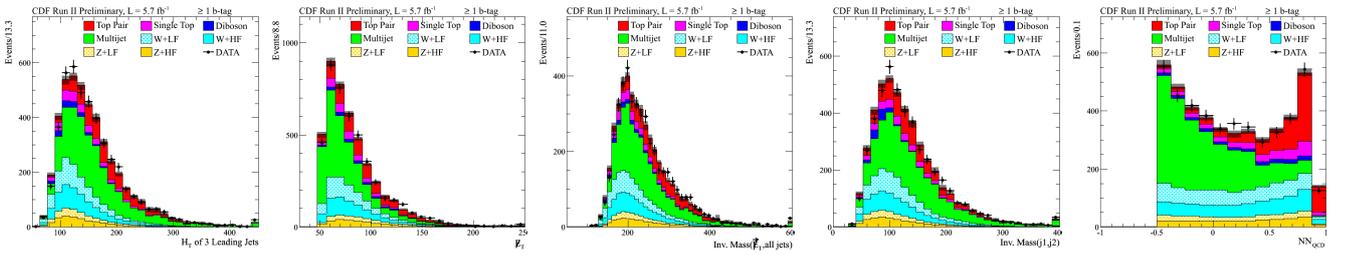


FIG. 4: The distribution of 5 NN inputs for the final discriminant in Signal Region.

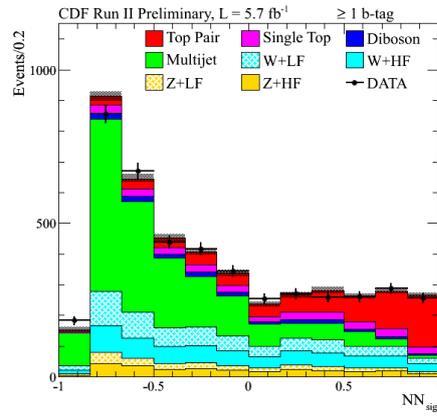


FIG. 5: The  $NN_{sig}$  distribution in Signal Region.