CDF note 9251

Combination of CDF Single Top Quark Searches with 2.2 fb^{-1} of Data

The CDF Collaboration
URL http://www-cdf.fnal.gov
(Dated: March 25, 2008)

There are currently three separate CDF searches for single top production: an analysis using neural networks, an analysis using a multivariate likelihood function technique, and an analysis using matrix element discriminants. We combine these analyses using neural networks that take the discriminant outputs from the three analyses as inputs to form a single, more powerful discriminant. The weights for the combination neural networks are chosen using genetic algorithms to provide optimal sensitivity. This analysis uses 2.2 fb^{-1} of CDF Run II data collected between February 2002 and August 2007 at the Tevatron in proton-antiproton collisions at a center-of-mass energy of 1.96 TeV. We measure a combined single top s- and t-channel cross section of 2.2 ± 0.7 pb. The observed signal has a significance of 3.7σ, while the median significance in psuedo-experiments is 5.1σ. These sensitivities represent an improvement of approximately 9% for the observed significance and 13% for the expected significance over the best single analysis. From the cross section measurement we extract a value for $|V_{tb}|$ of 0.88 ± 0.14(exp.)± 0.07(theory). As a cross check, we combine the results of the individual analyses using a modified version of the Best Linear Unbiased Estimator (BLUE) technique to obtain a result of $2.1^{+0.7}_{-0.6}$ pb. BLUE also finds that the observed signal has a significance of 3.7σ while its expected significance is 4.7σ.

Preliminary Results for Winter 2008 Conferences

I. INTRODUCTION

Although top quarks are predominantly produced at the Tevatron in pairs through the strong interaction, there is also an electroweak production mechanism that produces single top quarks. The top pair production process allowed the joint CDF-DØ discovery of the top quark in Run 1, and is currently used in Run II for a variety of precision top quark measurements and searches. On the other hand, although both CDF and DØ have evidence for electroweak single top production, neither experiment has observed this process. Measuring the cross section for electroweak single top production places direct experimental constraints on the value of the magnitude of the CKM matrix element $|V_{tb}|$. In addition, several models with physics beyond the Standard Model predict modifications to the observed rate of single top production. Therefore, conclusive observation of electroweak single top production and precise measurement of its cross section remains a priority at the Tevatron.

CDF currently performs three separate searches for electroweak single top production. Each search uses a common event selection, but implements a different multivariate analysis techniques: a multivariate likelihood function [1], neural networks [2], and a matrix element discriminant [3]. In each analysis, multiple variables are combined into a single variable with improved discrimination between the single top signal and the background. The discriminant distribution for the data is fit using templates created from signal and background models to extract a single top cross section and to calculate the significance of the observed signal.

The goal of the single top search combination is to use the maximum amount of information from the three individual analyses to create a single, more sensitive analysis. The strategy taken here is to combine the individual single top searches at the analysis level. We use the discriminants from the individual analyses as inputs to a new super-discriminant. We choose a neural network to serve as the super discriminant and optimize for sensitivity using a technique called neuro-evolution. Neuro-evolution employs genetic algorithms to choose the neural network weights and topology that maximizes a given figure of merit, in this case, the expected p-value. The final evolved neural network provides a single discriminant that is more sensitive than any of the individual analyses.

As a cross check to the evolved neural network super-discriminant combination, we also perform a more conventional combination of the individual single top results using a modified version of the Best Linear Unbiased Estimator (BLUE) technique. The BLUE approach is to combine the results of the individual analyses rather than combining their discriminants. In addition to providing a cross check to the evolved neural network combination, BLUE also provides information on the compatibility of the individual single top search results with each other and with the Standard Model hypothesis.

II. COMBINATION USING EVOLVED NEURAL NETWORKS

Although the discriminants from the individual analyses are certainly correlated–with correlation coefficients ranging from 0.42 to 0.79–the fact that the discriminant outputs are not completely correlated suggests that each analysis does not use identical information to discriminate signal from background. Furthermore, because the correlations between pairs of discriminants are different for different physics processes, it may even be possible to use these correlations themselves to further improve separation between signal and background. For these reasons, we pursue a multivariate technique to combine the individual analysis discriminants into a single super-discriminant, using a neuro-evolution technique. A similar technique was previously employed to improve the event selection in a dilepton top mass measurement [4, 5]. Here we apply neuro-evolution to obtain a neural network that combines the discriminant values of the individual analyses into a discriminant with improved sensitivity.

A. Neuro-Evolution

A typical approach to neural network training involves using a gradient descent method, such as backpropagation, to minimize the classification error, defined by $\sum (o_i - t_i)^2$, where o_i is the output of the neural network and t_i is the desired output, usually zero for background and one for signal. Although backpropagation is a powerful and fast technique for training neural networks, it is not necessarily true that a neural network that minimizes the classification error will also provide the greatest sensitivity in a search. Therefore, it is interesting to explore additional training methods capable of optimizing quantities directly related to the problem of interest. One such technique known as neuro-evolution involves using genetic algorithms, rather than backpropagation, to search the space of possible neural network weights. Because this search proceeds stochastically, it can be used to optimize an arbitrary figure of merit, in contrast to gradient descent methods which require a figure of merit with well-behaved with calculable derivatives. The neuro-evolution package used for this analysis is Neuro-Evolution of Augmenting Topologies (NEAT) [6]. NEAT

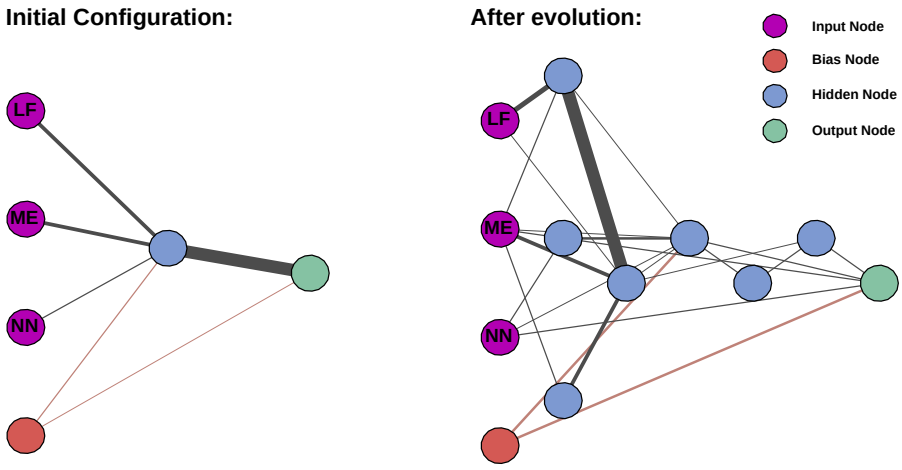

FIG. 1: The diagram on the right indicates the initial neural network configuration from which neuro-evolution proceeds. Hidden nodes use a sigmoid activation function while the output node uses a linear activation function. In the initial configuration, the weight between the hidden node and the output node provides the freedom to scale the neural network output while the weight between the bias node and the output allows control over the offset. The diagram on the left shows an example of an evolved neural network. The thickness of the lines indicates the relative magnitude of the weights.

has the advantage that in addition to optimizing the neural network weights, it also varies the network topology, adding complexity as needed to improve performance.

Neuro-evolution with NEAT begins from a population of neural networks generated from a seed network by randomly varying the network weights. Evolution then proceeds in generations. In each generation, the following steps are completed:

1. The fitness of each neural network is evaluated by calculating the networks performance using a figure of merit.
2. Networks with poor fitness are removed from the population.
3. The remaining networks are allowed to replenish the population through mutation and breeding. Possible mutations include randomly changing one or more the ANN weights, randomly adding a link between nodes, and randomly adding new nodes. Breeding involves blending randomly selected features from two networks.

The population of networks remaining at the end of this process for one generation becomes the initial population for the next generation.

For this problem, the optimal neural network we are searching for is the one that would give us the largest signal significance or p-value. Unfortunately, calculating the expected p-value for a given neural network is computationally expensive. As an alternative, we histogram the neural network output separately for the expected signal and background contributions and then calculate quadrature sum of $S/\sqrt{B}$ for the histogram bins, where S is the expected signal contribution and B is the expected background contribution for each bin. This figure of merit is correlated to the expected p-value but can be calculated reasonably quickly. At the end of the evolution, we collect the networks with the highest fitness values and do a more detailed estimation of the expected p-value that includes the full range of systematic uncertainties in order to select a final neural network that has the best sensitivity, including the effects of systematic uncertainties.

Another challenge in optimizing the sensitivity in a search like this is to select an optimal binning for the neural network output histograms. In principal, each neural network considered during evolution could require a different binning for optimal sensitivity. Because it is not possible manually to choose the best binning for the large number of candidate neural networks considered during evolution, an automated binning scheme must be used. We choose to solve the problem by incorporating the choice of optimal binning in the evolution process itself. During the fitness calculation, we fill histograms with a fixed number of bins in an output range from zero to one. However, we choose an initial neural network configuration as shown in Figure 1, which provides NEAT with the freedom to scale and translate the neural network output within the histogram range to make the best use of the available bins.

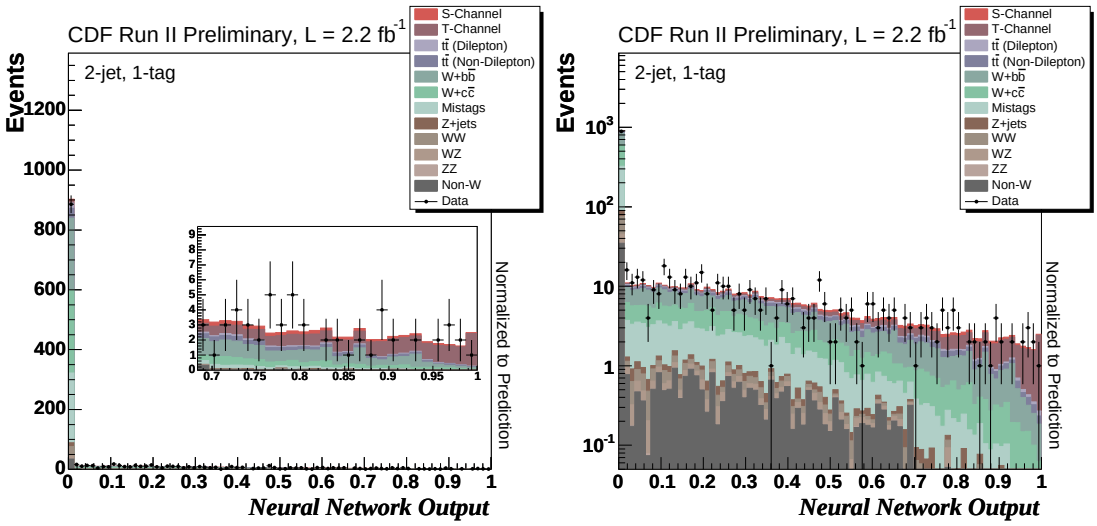

FIG. 2: This figure shows the evolved neural network output for events in the two-jet, one-tag channel. The binning used for this plot is the one chosen during neuro-evolution.

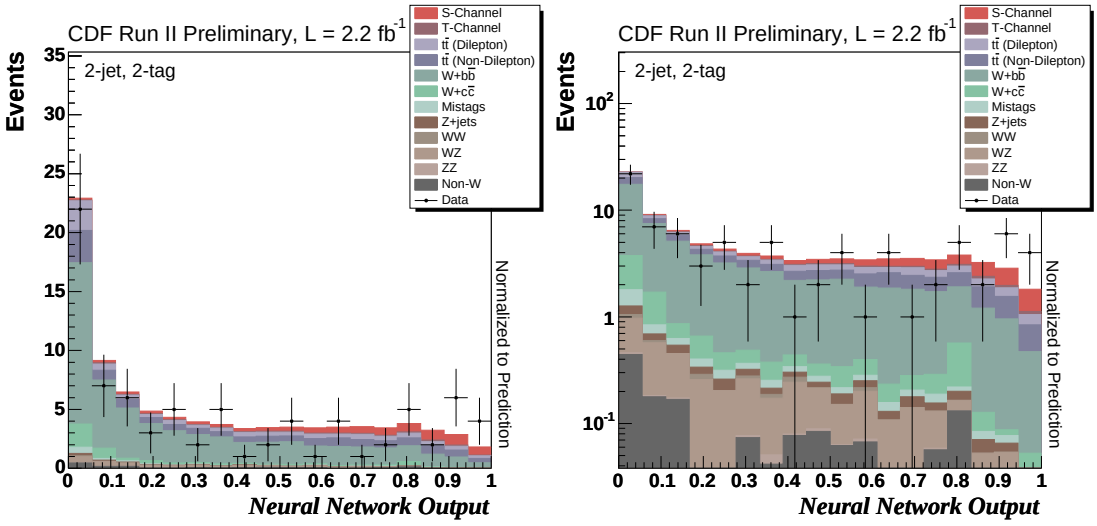

FIG. 3: This figure shows the evolved neural network output for events in the two-jet, two-tag channel. The binning used for this plot is the one chosen during neuro-evolution.

B. Evolved Neural Network Output

Because the individual analyses optimize separate discriminants for the two- and three-jet events and the single- and double-b-tag events, we evolve separate neural networks for each of the following subsets of the data: two-jet, single-b-tag; two-jet, double-b-tag; three-jet, single-b-tag; and three-jet, double-b-tag. The final evolved neural networks are shown in Figures 2 through 5. The binning shown in these plots is that chosen by the evolution as the optimal binning. In addition, the matrix element analysis uses extended muon coverage data collected on a separate trigger from the data mentioned above. Because only the matrix element analysis uses this data, the discriminant used in the combination for those events is just the matrix element discriminant, shown in Ref. [3].

For visualization purposes, we also plot the combined output for all data (including the extended muon events) in Figure 6. The binning shown here is arbitrary.

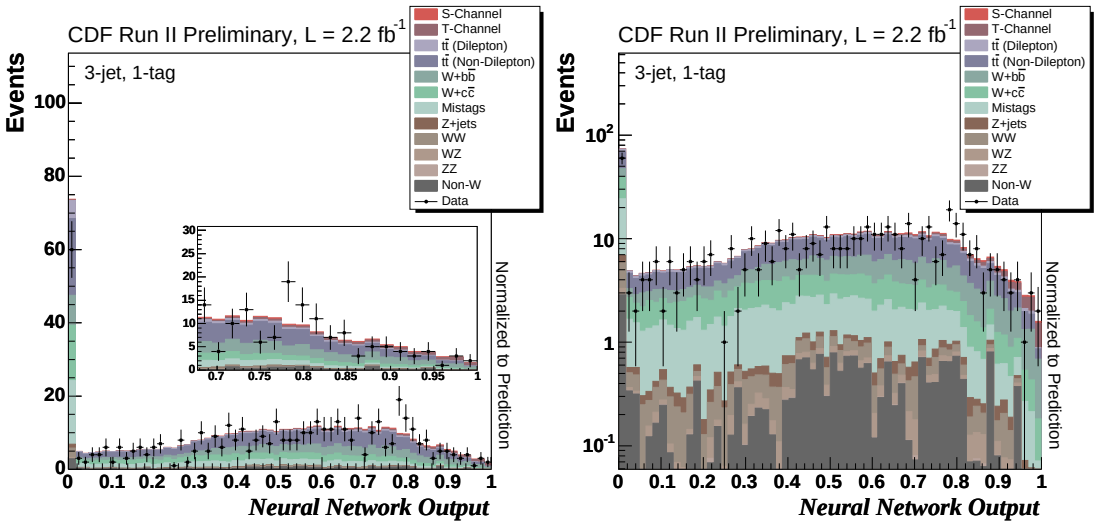

FIG. 4: This figure shows the evolved neural network output for events in the three-jet, one-tag channel. The binning used for this plot is the one chosen during neuro-evolution.

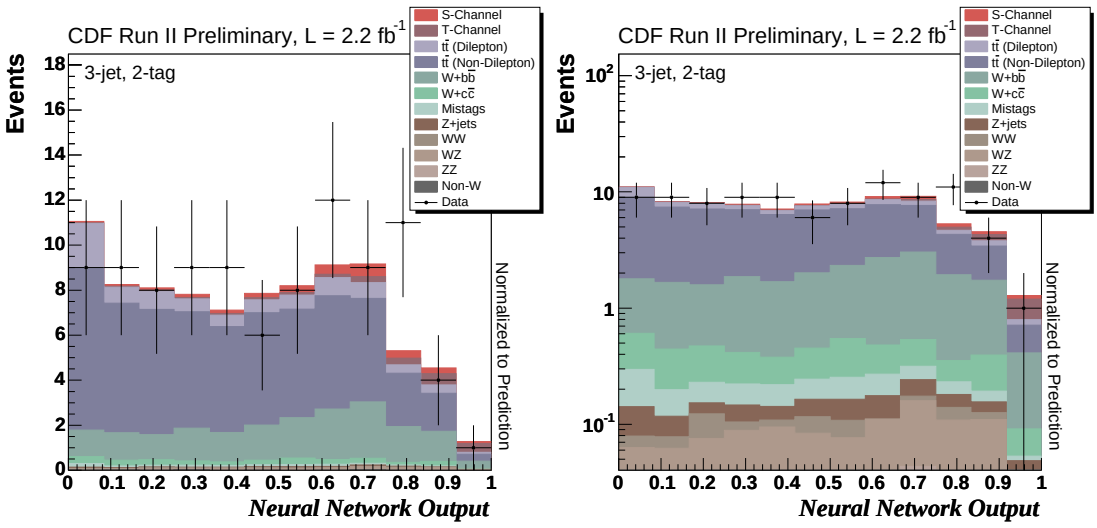

FIG. 5: This figure shows the evolved neural network output for events in the three-jet, two-tag channel. The binning used for this plot is the one chosen during neuro-evolution.

C. Cross Section Fit and $|V_{tb}|$

The cross section is extracted using a binned likelihood fit in which nuisance parameters are treated with a Bayesian marginalization technique. The likelihood curve from the fit is shown in Figure 7. The measured cross section, assuming s-channel plus t-channel single top production in the Standard Model proportions, and a top mass of 175 GeV/c^2, is 2.2 ± 0.7 pb.

The cross section measurement is also converted directly into a measurement of $|V_{tb}|$ based on the assumption that $|V_{tb}|^2 \gg |V_{ts}|^2 + |V_{td}|^2$ as follows:

$$|V_{tb,Measured}|^2 = \frac{\sigma_{Measured}}{\sigma_{SM}} \times |V_{tb,SM}|^2 \tag{1}$$

We obtain $|V_{tb}| = 0.88 \pm 0.14$ (experimental) ± 0.07(theory).

A 95% confidence level lower limit on $|V_{tb}|$ is extracted using a Bayesian technique assuming a flat prior on $|V_{tb}|^2$

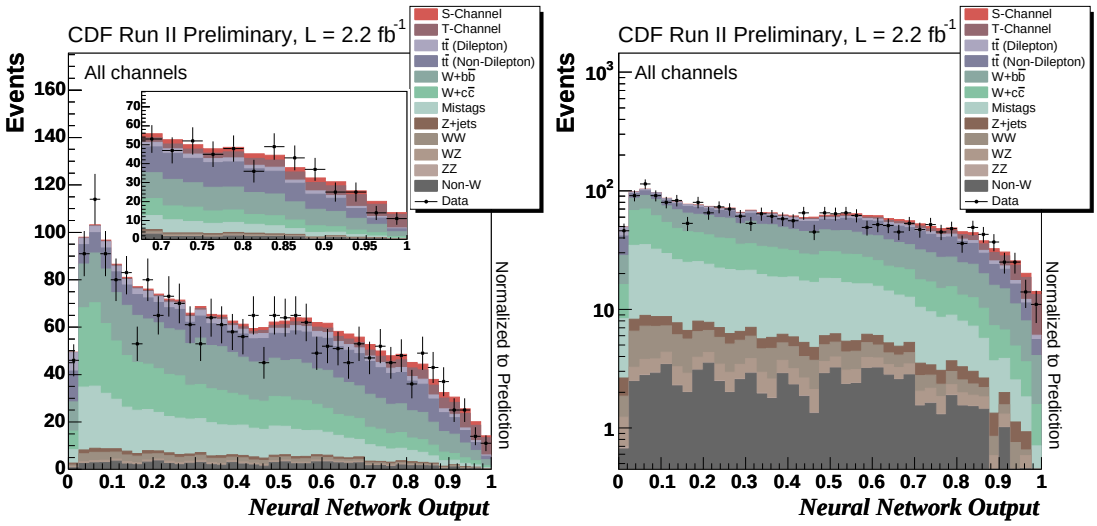

FIG. 6: This figure shows the evolved neural network output for events in all channels, including the extended muon coverage events. The binning used for this plot is arbitrary.

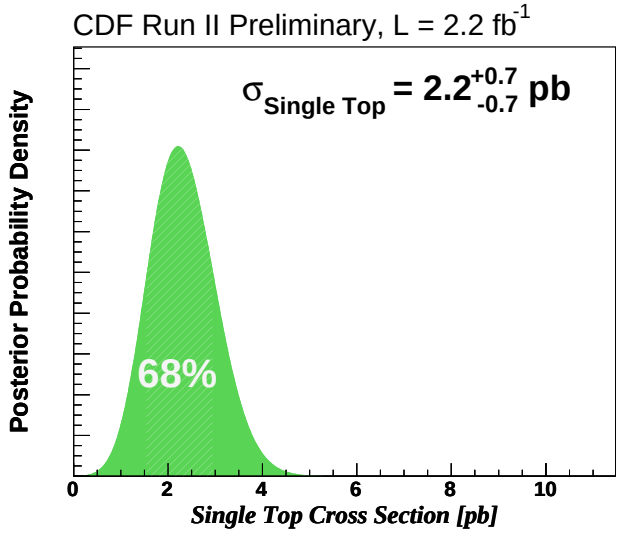

FIG. 7: This plot shows the posteriori probability density used to extract the cross section. The hatched region shows the shortest interval around the maximum value that contains 68% of the area, which is used to quote the uncertainty on the measured cross section.

between zero and one. The posterior probability density is shown in Figure 8, corresponding to $|V_{tb}| > 0.66$ at 95% confidence.

D. p-Value

To calculate the significance of the observed signal by comparing the result obtained in data to a distribution of results from pseudo-experiments drawn assuming the absence of single-top production (background-only or "b" hypothesis). The test statistic we use for this calculation is the standard likelihood ratio test statistic:

$$Q = -2 \ln \frac{P(data|s+b)}{P(data|b)}, \tag{2}$$

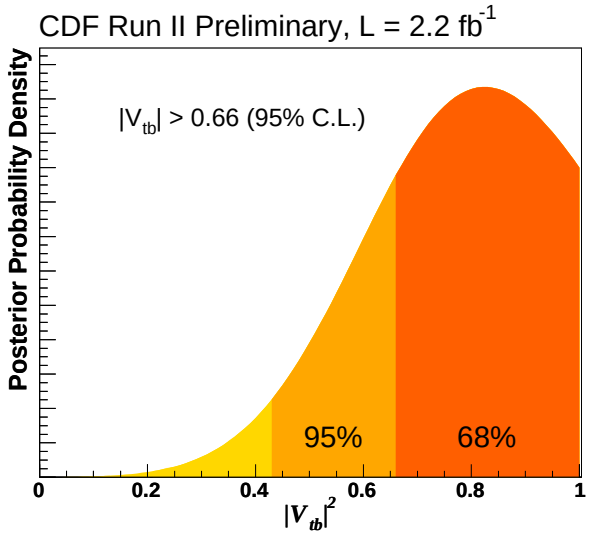

FIG. 8: This plot shows the posteriori probability density used to extract the limit on $|V_{tb}|$. A flat prior in $|V_{tb}|^2$ in the physical region (between 0 and 1) is assumed.

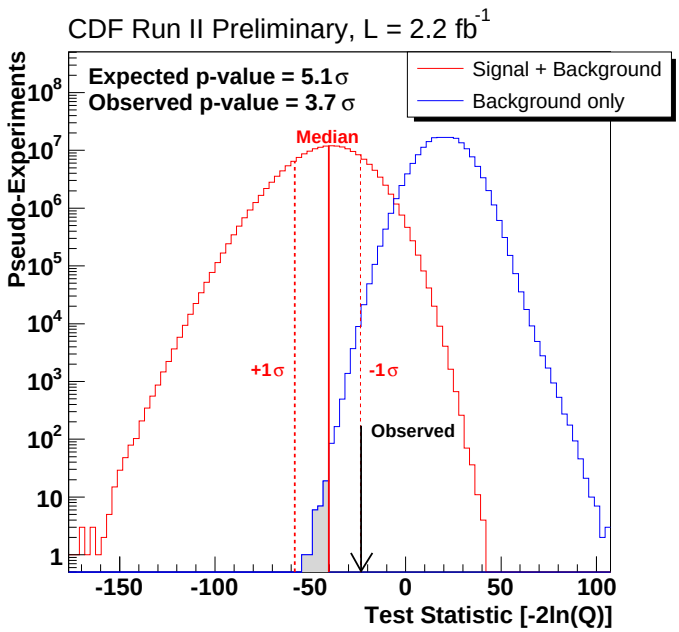

FIG. 9: This figure shows the test statistic distribution for pseudo-experiments drawn from the "background-only" (no single top production) and "signal + background" (single top production according to the Standard Model) hypotheses. The median of the signal plus background distribution determines the expected p-value. The test statistic value observed in data is indicated by the arrow.

where "s+b" refers to the hypothesis with the Standard Model amount of single-top signal and "b" refers to the hypothesis that single-top is absent. The p-value, or probability for the data to come from the background-only hypothesis is 9.4×10^{-5}, which corresponds to 3.7σ. We calculate the expected p-value assuming the Standard Model amount of single-top production by generating pseudo-experiments from the "s+b" hypothesis and calculating the p-value for the median test statistic from the distribution of those pseudo-experiments, shown in Figure 9. This expected p-value is 2.0×10^{-7}, which corresponds to 5.1σ. Assuming Standard Model single top production, the test-statistic value observed in data represents a one sigma fluctuation.

III. CROSS CHECK AND COMPATIBILITY STUDIES USING ASYMMETRIC ITERATIVE BLUE

A. Introduction to BLUE

As a cross check to the super-discriminant result using NEAT, the method of forming a Best Linear Unbiased Estimate, BLUE, is used to make a weighted average of the three single top cross section measurements [7, 8]. This method is based upon minimizing a χ^2 while taking into account different types of errors and different correlations between these errors.

In the simple case of combining N measurements of the same parameter μ that are completely uncorrelated with each other one can find the average of these measurements by finding the value of μ that is the minimum of the χ^2 expression:

$$\chi^2(\mu) = \sum_i^N \left(\frac{m_i - \mu}{\sigma_i} \right)^2 \tag{3}$$

where m_i and σ_i are the measured value and error of the ith analysis, and μ is the average of the measurements. Equation 3 may be re-written as

$$\chi^2(\mu) = \delta^T(\mu) \cdot \mathbf{S}^{-1} \cdot \delta(\mu) \tag{4}$$

where $\delta(\mu)$ is the column vector where $\delta_i = m_i - \mu$ and $\mathbf{S}$ is the covariance matrix. In this case, $\mathbf{S}$ is diagonal with $\mathbf{S}_{(i,i)} = \sigma_i^2$.

BLUE is based upon the same formula as equation 4. In order to combine N measurements which each have E different types of errors, $\mathbf{S}$ is defined as

$$\mathbf{S} = \sum_e^E \mathbf{S}_e \tag{5}$$

and

$$\mathbf{S}_{e(i,j)} = \sigma_{ei} \cdot \sigma_{ej} \cdot \rho_{e(i,j)} \tag{6}$$

where σ_{ei} is the eth type of error on the ith measurement and $\rho_{e(i,j)}$ is the correlation of the eth error between the ith and jth measurement.

The beauty of BLUE lies in the fact that you don't have to run a minimization routine on this χ^2 to get the mean. Defining $\mathbf{H}$ to be the inverse of the covariance matrix and sum_H as below.

$$\mathbf{H} \equiv \mathbf{S}^{-1} \tag{7}$$

and

$$sum_H \equiv \sum_{i=1}^N \sum_{j=1}^N \mathbf{H}_{i,j} \tag{8}$$

a weight for each measurement, w_i, can then be calculated as

$$w_i \equiv \sum_{j=1}^N \frac{\mathbf{H}_{i,j}}{sum_H} \tag{9}$$

Note that weights can be negative, but

$$\sum_{i=1}^N w_i = 1 \tag{10}$$

BLUE predicts

$$\mu_{\text{best}} = \sum_{i=1}^{N} w_i \cdot m_i \tag{11}$$

is the minimum of the χ^2. The total error σ is calculated as below.

$$\sigma = \frac{1}{\sqrt{sum_H}} \tag{12}$$

As with all weighted averaging techniques, the weights assigned to each analysis in BLUE are calculated with knowledge of each analyses uncertainties. Many uncertainties vary with the result of the analysis (*e.g.,* a 10% acceptance uncertainty will be a 0.2 pb uncertainty if 2 pb is measured and a 0.3 pb uncertainty if a 3 pb result is measured). If ignored, this effect can lead to a bias in combining different measurements.

To take this effect into account, we first parametrize all uncertainties (statistical and systematic alike) as a function of the single top cross section. Next, we iterate BLUE as follows. We first take a simple weighted average and have all cross section measurements reevaluate their uncertainties at the weighted average. Finally, we then run BLUE and reevaluate the uncertainties again. This is repeated until the starting combination and final combination cross sections are within a small uncertainty (*i.e.,* smaller than a relative difference of 0.1%).

B. Correlations

As mentioned above, the three analyses share a common event selection (except for ME which adds additional muon types). In order to assess the level of correlation between the three Single-Top measurements, correlated pseudo-experiments (PEs) are thrown from the pool of Monte Carlo events for each signal and background sample according to their expected contributions and uncertainties. For each PE the global discriminant is constructed for each analysis: LF, ME, and NN. A binned-likelihood fitter is then used to obtain the fit values for each analysis by fitting against reference histograms of signal and background (templates). The pull ($\frac{\beta_{\text{Fit}}-\beta_{\text{true}}}{\sigma_{\text{Fit}}}$) distributions of PE's for each analysis were checked to guarantee that the throwing and fitting of PEs are unbiased.

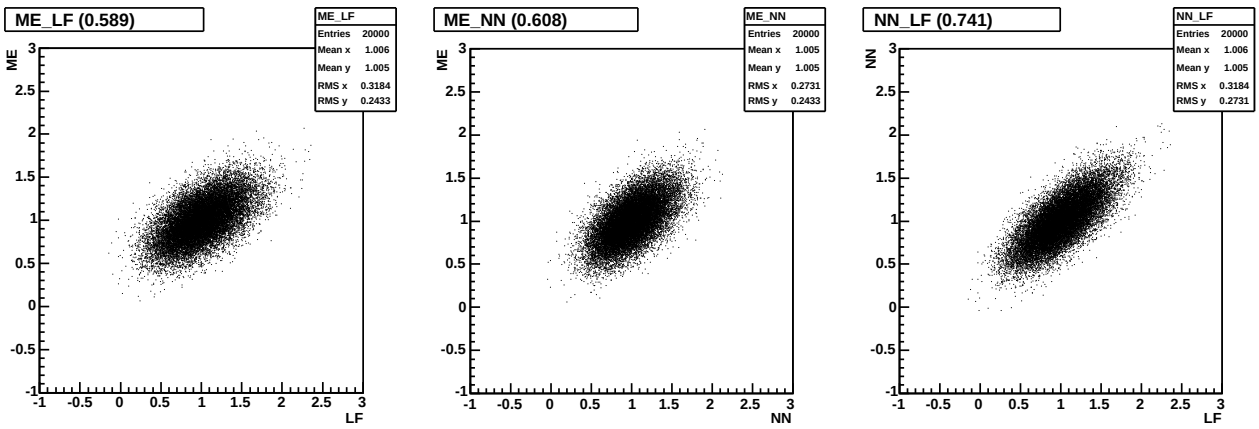

FIG. 10: Scatter plot of the β values for each analysis pair from the correlated PEs. The correlation coefficients are included in the figure titles.

The two-dimensional scatter plots of the fitted cross section values for each analysis pair shown in Figure 10 are used to derive correlation coefficients. These coefficients quantify the level of correlation between each analysis pair needed for the BLUE combination and are summarized in Table I.

C. Asymmetric Uncertainties and Asymmetric Iterative BLUE

In many analyses, it is not uncommon to have asymmetric uncertainties; this analysis is no exception. The issue with BLUE is that it is based on Gaussian χ^2 distributions and is not able to directly treat asymmetric uncertainties.

	LF	ME	NN
LF	1.0	0.599	0.741
ME	—	1.0	0.609
NN	—	—	1.0

TABLE I: Correlations coefficients between pairs of cross section values.

To get around this issue, we have developed Asymmetric Iterative BLUE (AIB).

The structure of an AIB combination is simply three BLUE combinations at one time. The *center* BLUE combination uses the average value of all uncertainties. It is this combination that determines both the central value and the total length of the uncertainty. This combination is the same as the standard BLUE combination.

There are also *upper* and *lower* BLUE combinations that are given the upper and lower uncertainties respectively. We calculate the total uncertainties for the *upper* and *lower* BLUE combinations using the central value of the *center* BLUE combination and apply the ratio of the *upper* BLUE and *lower* BLUE to the total uncertainty of the *center* combination.

$$\mathcal{R}_{\text{upper}} = \frac{\sigma_{\text{upper BLUE}}}{\sigma_{\text{upper BLUE}} + \sigma_{\text{lower BLUE}}} \tag{13}$$

$$\sigma_{\text{upper}} = 2 \cdot \mathcal{R}_{\text{upper}} \cdot \sigma_{\text{center BLUE}} \tag{14}$$

$$\sigma_{\text{lower}} = 2 \cdot (1 - \mathcal{R}_{\text{upper}}) \cdot \sigma_{\text{center BLUE}} \tag{15}$$

It has been shown that AIB covers slightly better than BLUE when dealing with asymmetric uncertainties. Note that by construction, the final value and average uncertainties are involved. Second, by construction, the average of AIB upper and lower uncertainty is always equal to the final value and uncertainty from BLUE. Specifically,

$$\sigma_{\text{center BLUE}} \equiv \frac{\sigma_{\text{upper}} + \sigma_{\text{lower}}}{2} \tag{16}$$

D. Measuring Systematic Uncertainties

For individual analyses, it is easy to generate PEs with complicated systematic uncertainties. For each PE, simply all nuisance parameters within their uncertainties. For parameters that describe shape uncertainties, one can easily use a linear extrapolation. For example, if we have a JES shifted sample of $+1\ \sigma$, we can estimate what the effect of a JES shifted sample of $+0.8\ \sigma$.

In the BLUE combination, it is not trivial to use the same techniques since we always need to throw correlated PEs[9]. To measure the magnitude of various systematic uncertainties, we generate correlated PEs using systematic-varied sets of events (*e.g.,* JES shifted up 1 σ), and fit the resulting distributions with the nominal templates for all three analyses, measuring both upper and lower uncertainties. This procedure is repeated for many different true values of the single top cross section. These parametrized distributions are given to the BLUE combination code. See Figure 11 for the parametrization of the largest uncertainties for each of the three analyses.

E. AIB Pull Distributions with Systematic Uncertainties

When checking the pull distribution of the combination with AIB, we want to generate the PE with the same assumptions that are used to measure the single top cross section. As mentioned above, generating correlated PEs with all systematic uncertainties is difficult. In its place, we generate PEs with only rate systematic uncertainties and then we coherently *smear* the three analyses results to account for the systematic uncertainties. For each systematic uncertainty, we generate a single random number (shift) from a unit Gaussian distribution. If the number is positive (negative), we use the upper (lower) uncertainties for the three analyses and add the product of the shift and the uncertainty to the measured cross section.

In Figure 12 we find that the pull distribution of AIB is well described by an unbiased unit Gaussian. Also included in Figure 12 is the pull distribution for a standard BLUE combination. It is clear that the AIB pull distributions are better described by a unit Gaussian than those of BLUE (even in the case where only statistical errors are considered).

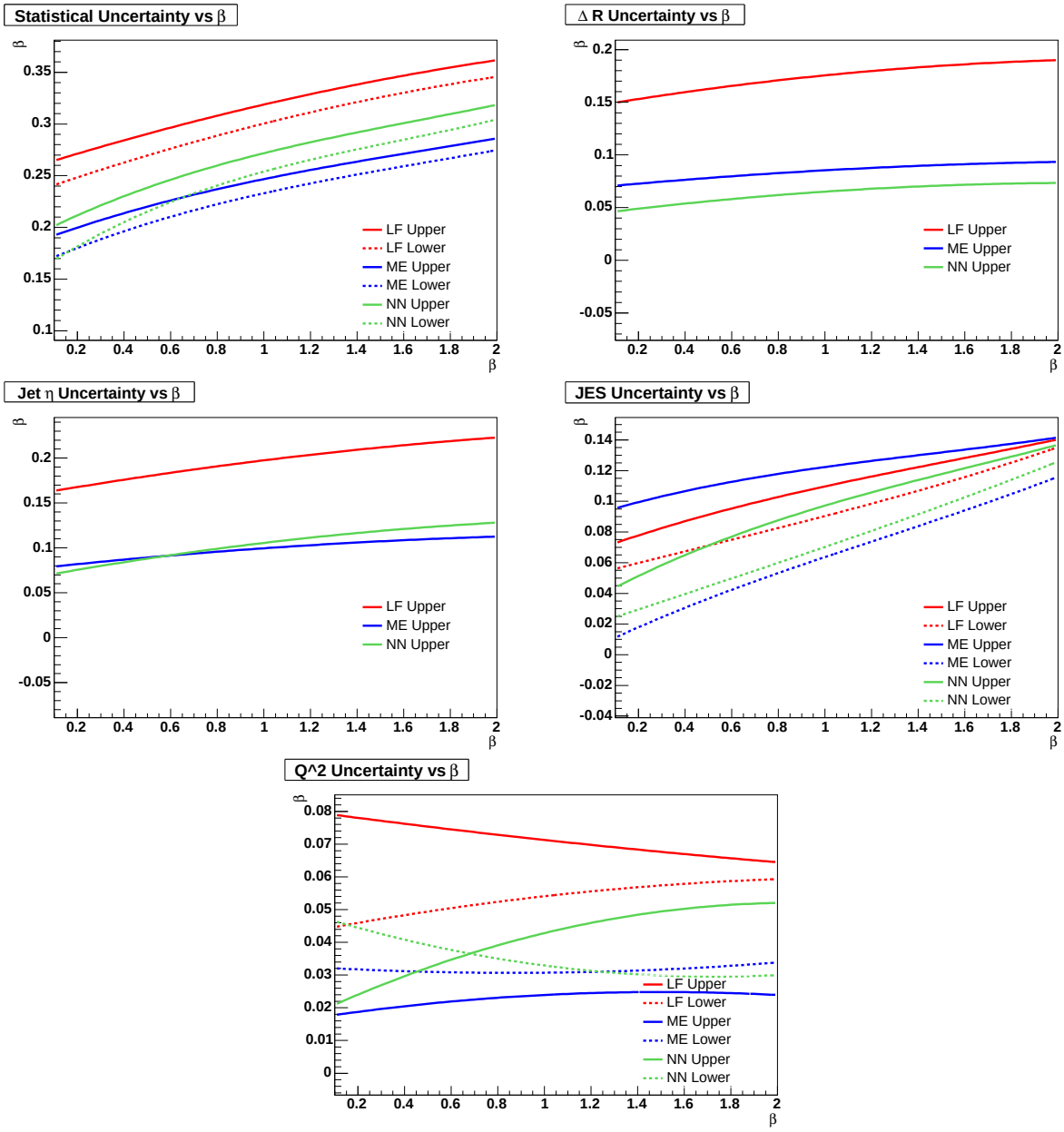

FIG. 11: Parametrization of statistical uncertainties, ΔR systematic uncertainties, jet η systematic uncertainties, JES systematic uncertainties, and Q^2 systematic uncertainties.

F. p-Value

The BLUE method calculates the p-value similarly to most other methods: We generate a large ensemble of PEs generated with all systematic uncertainties, but with no signal present. We then ask what percentage of them have a signal-like figure of merit larger than what we see in data. In the BLUE case, we use the combined single top cross section as our figure of merit. The β distributions for PEs with no signal present are shown in Figure 13. Looking at pseudo-experiments with no signal present, AIB calculates an observed sensitivity of 3.7 σ with a median expected sensitivity of 4.7 σ. Based on expected p-value, the AIB combination is about 7 % more sensitive than any individual analysis.

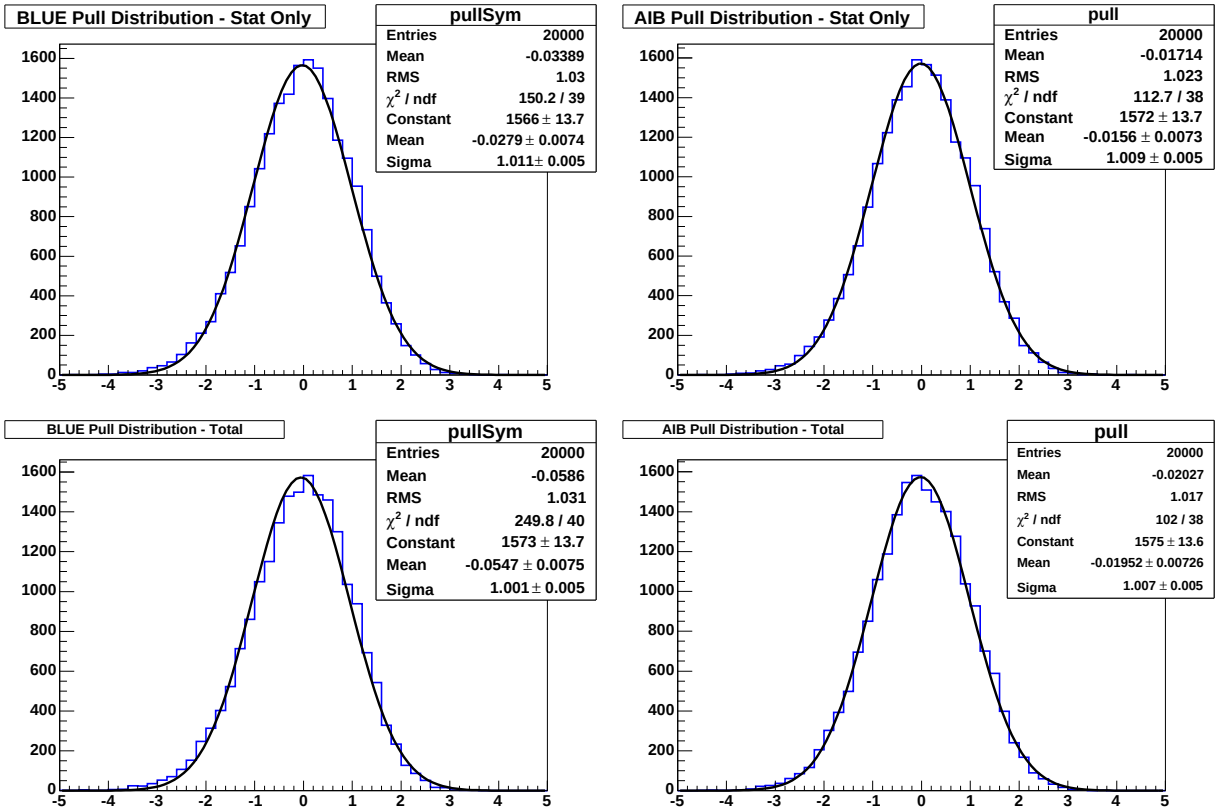

FIG. 12: Pull distribution for BLUE (left) and AIB (right) run on PEs with statistical only (top) and systematic uncertainty smearing (bottom) after un-biasing the individual analyses. Note that the AIB pull distributions are better described by a unit Gaussian than the BLUE distributions in both cases.

G. BLUE: Results

BLUE needs the measured cross section from each analysis and the correlation coefficients (Table I) as inputs for the combination. The cross section values used in the following BLUE result are given in Table II.

Analysis	σ
LF	1.79
ME	2.17
NN	1.97

TABLE II: Single top production cross section values measured by each analysis and used as the input for the BLUE combination.

Applying the AIB procedure defined above yields a *combined cross section measurement is* $2.1^{+0.7}_{-0.6}$ *pb*. From PE's with ($\sigma_{singleTop} = 2$ *pb*), the three-analysis compatibility including all systematics was found to be 87 %, that is to say that 87 % of the time the χ^2 from the PE was higher than that observed in data. In addition the compatibility with the standard model was checked with PE's ($\sigma_{singleTop} = 2.86$ *pb*). 14.8 % of the time the PE had a combined cross section of less than that observed in data. This represents a discrepancy with respect to the standard model of less than 1.1 σ.

To summarize, the AIB combined cross section is $2.1^{+0.7}_{-0.6}$ *pb*. Compatibility is found to be very good between the three analyses, and better than 1.1 σ with the standard model expectation. AIB has an expected p-value of 4.7 σ and an observed p-value of 3.7 σ. This combined cross section result using AIB is in almost perfect agreement with and serves as an important cross check of the result from the super-discriminant method using NEAT discussed in the early sections of this note.

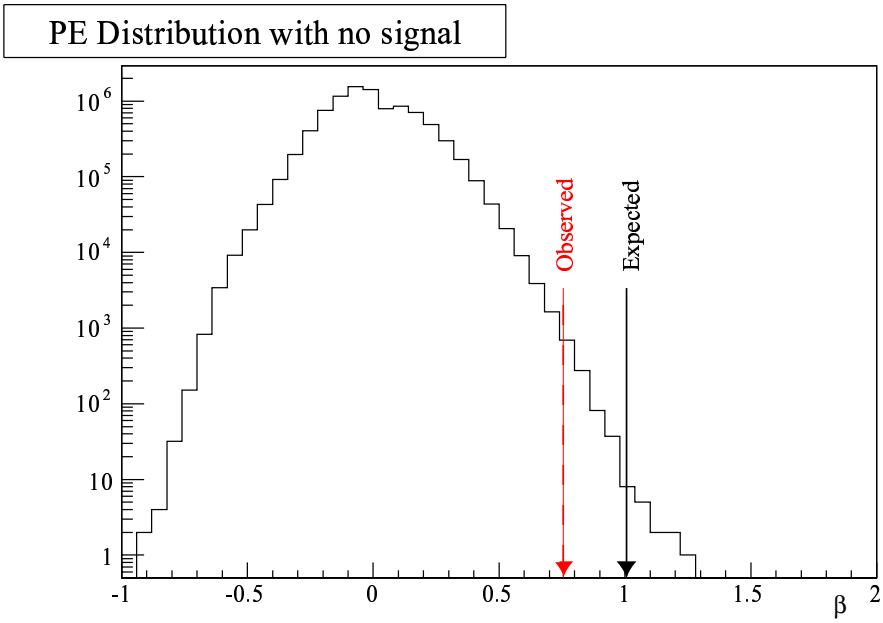

FIG. 13: The distribution of fit β values from PE's with zero signal thrown ($\beta = 0$). The observed β represents a 3.69 σ significance of signal top signal over a background only hypothesis.

IV. CONCLUSION

By combing the three separate CDF single-top quark searches using evolved neural networks, we obtain an expected signal significance of 5.1σ, which is an improvement of 13% over the best individual analysis. The p-value in observed in data corresponds to 3.7σ which is 9% better than the best individual analysis. The cross section measured is 2.2 ± 0.7 pb and we set a limit of $|V_{tb}| > 0.66$ at 95% confidence level. This combination is cross checked using AIB, a variant of the standard BLUE technique. The AIB result agrees well with the evolved neural network result. Furthermore, usig AIB, we determine that the results from the three separate single-top searches are very consistent with one another.

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 für Bildung und Forschung, Germany; the Korean Science and Engineering Foundation and the Korean Research Foundation; the Science and Technology Facilities Council and the Royal Society, UK; the Institut National de Physique Nucleaire et Physique des Particules/CNRS; the Russian Foundation for Basic Research; the Comisión Interministerial de Ciencia y Tecnología, Spain; the European Community's Human Potential Programme; the Slovak R&D Agency; and the Academy of Finland.

[1] S. Budd, *et al.*, "Multivariate Likelihood Search for Single-Top-Quark Production with 2.2 fb^{-1}," CDF Note 9221 (public); `http://www-cdf.fnal.gov/physics/new/top/confNotes/cdf9221_LF2.2.pdf`.

[2] D. Hirschbühl, *et al.*, "Search for Electroweak Single Top-Quark Production using Neural Networks with 2.2 fb^{-1} of CDF II data," CDF Note 9217 (public); `http://www-cdf.fnal.gov/physics/new/top/2008/singletop/NN/plots/Conference_Note_NN2fb.pdf`.

[3] F. Canelli, *et al.*, "Measurement of Single Top Quark Production in 2.2 fb^{-1} of CDF II Data using the Matrix Element Technique," CDF Note 9223 (public); `http://www-cdf.fnal.gov/physics/new/top/confNotes/cdf9223_stME_WC2008_v2.pdf`.
[4] S. Whiteson and D. Whiteson. "Stochastic optimization for collision selection in high energy physics." In *Proceedings of the Nineteenth Annual Innovative Applications of Arti ficial Intelligence Conference*, 1819-1825, 2007.
[5] B. Jayatilaka, *et al.*, "Measurement of the Top Quark Mass in the Dilepton Channel using a Matrix Element Method and Neuroevolution Selection with 2.0 fb^{-1}," CDF note 9130; `http://www-cdf.fnal.gov/physics/new/top/confNotes/cdf9130_me_dil_nn_pub.pdf`.
[6] Kenneth O. Stanley and Risto Miikkulainen, Evolutionary Computation **10 (2)** 99-127,2002.
[7] L.Lyons, D. Gibaut, and P. Clifford, NIM **A270**, 110-117 (1988)
[8] L.Lyons, A. Martin, and D. Saxon, Phys. Rev. D **41**, 3 (1990)
[9] It is technically possible to generate three correlated PEs with continuous systematic uncertainties. One would need both a lot more statistics as well as 3-D template morphing and this is currently beyond the scope of this combination effort.