PHYSICAL REVIEW D VOLUME 46, NUMBER 11 1 DECEMBER 1992

Snagging the top quark with a neural net

Howard Baer
Physics Department, Florida State University, Tallahassee, Florida 32306

Debra Dzialo Karatas
Center for Particle Physics, The University of Texas, Austin, Texas 78712

Gian F. Giudice
Theory Group, Department of Physics, The University of Texas, Austin, Texas 78712
(Received 20 December 1991)

The search for the top quark at $p\bar{p}$ colliders in the one-lepton-plus-jets channel is plagued by an irremovable background from W-boson-plus-multijet production. In this paper, we show how the top-quark signal can be distinguished from background in the distribution of neural network output. By making a cut on the network output, we maximize the ratio of signal to background in a final event sample, and compare our results with those obtained by making kinematical cuts on the data sample. We also demonstrate the robustness of the neural network method by training the neural network on signal events of one top mass and testing upon another.

PACS number(s): 13.85.Qk, 14.80.Dq

The top quark t is a necessary ingredient of the standard model (SM), as it is required to ensure the cancellation of triangle anomalies [1], to account for the $e^+e^- \to b\bar{b}$ forward–backward jet asymmetry [2], and for the suppression of neutral current B decays [3]. Radiative corrections to SM parameters measured at the CERN e^+e^- collider LEP constrain the top-quark mass to be 89 GeV $< m_t <$ 191 GeV [4]. The lack of a top-quark signal at the Fermilab Tevatron to date furthermore restricts $m_t > 91$ GeV [5]. The search for top quarks of mass up to $\sim 120-140$ GeV at the Fermilab Tevatron is expected to resume in 1992, with the planned collection of 20 pb^{-1} of data.

The top quark at the Fermilab Tevatron collider is expected to be produced dominantly via the $gg \to t\bar{t}$ and $q\bar{q} \to t\bar{t}$ pair production channels. Since $m_t > M_W + m_b$, a SM top quark decays via $t \to bW$ where the W is on shell; hence, a $t\bar{t}$ event results in a $b\bar{b}W^+W^-$ final state. The cleanest signal for $t\bar{t}$ occurs in the two-lepton channel; the two leptons arise from the leptonic decay of each W boson. The principal backgrounds from W-pair production and $\gamma, Z \to \tau^+\tau^-$ are expected to be controllable [6]. Event rates in the two-lepton channel will be low because of the two leptonic branching ratios; moreover, this channel does not provide a direct determination of the top-quark mass because the two neutrinos from $W \to l\nu$ carry off an unknown amount of p_T. The best one can do is to either match the total event rate to QCD predictions (which depend on m_t), or to measure the relative rates, $R_{1/0}$ and $R_{2/1}$, of multijet production in events with two isolated leptons [6]. Here, $R_{i/j}$ stands for the ratio of events with i jets to events with j jets.

The single-lepton channel of $t\bar{t}$ decay, on the other hand, in which one W decays leptonically and the other hadronically, offers the possibility of a direct determination of the top-quark mass, since the presence of only a single neutrino permits a kinematical reconstruction of the event. By computing the invariant mass of the three leading jets for each event, the first two of which are expected to come from the hadronic W and the third from one of two b quarks, one may be able to determine the top-quark mass. This channel will be sought by the experimentalists both for its confirmation of the existence of the top quark, as well as for its value in providing a direct reconstruction of the top-quark mass, but it is not considered nearly as clean as the two-lepton channel because now the principal background is "W-plus-jets" production, which occurs with a large relative rate even with well-formulated kinematical cuts. In addition to this, when one constructs the invariant mass of the three leading jets, one includes the correct b quark jet less than half the time when one takes into account the presence of initial- and final-state radiation. This fact, plus the irrepressible presence of background contamination in any final event sample, tends to smear out the peak at the top-quark mass, and so one may fail to obtain a precise determination of m_t after all. Indirect top-quark mass measurements may also be made by comparing experiment with theoretical expectations for $R_{3/2}$ and $R_{4/3}$ in this channel [7].

In this paper, we report on our use of a neural network to distinguish signal from background events in the single-lepton channel of top-quark production at the Tevatron. Neural networks have proven to be an efficient technique for pattern recognition with important applications in a variety of high-energy physics problems [8]; our application is another in the area of signal-versus-background discrimination.

The problem of identifying signal over background processes with much larger total rates is a common one in

high-energy physics, especially in hadron collider experiments. The usual procedure of imposing linear cuts on the kinematical variables of each event is often not sufficient in providing the reduction of the background/signal ratio necessary to assess the experimental observability of the process. This problem, on the other hand, is well-suited for the neural network. The network builds a nonlinear functional representation of the signal and background process that permits a classification scheme more efficient than the conventional cuts, which simply correspond to a hyperplane separation in the kinematical variable space. The neural network can achieve an efficiency in signal-versus-background discrimination very close to the theoretical maximum determined by the Bayes decision rule [9].

We employ the feed-forward model of neural networks based on an input layer, a number of hidden layers, and an output layer. Each layer L ($L = 1, ..., M$) consists of $N^{(L)}$ nodes. The input layer corresponds to $L = 1$, and the output layer to $L = M$ (we take $N^{(M)} = 1$, i.e., a single output node). During the learning phase of the procedure, the net is confronted with a set of signal and background events, called the training set, which has been generated by Monte Carlo simulation. For each event, the kinematical variables of the process, $x_i^{(1)}$, for $i = 1, ..., N^{(1)}$, are presented as input to the neural network. From these inputs, the output of the next layer of nodes is computed using the recursive rule:

$$x_j^{(L)} = \frac{1}{2}\left[1 + \tanh\left(\frac{1}{T}\sum_{i=1}^{N^{(L-1)}} \omega_{ji}^{(L)} x_i^{(L-1)}\right)\right], \quad j = 1, ..., N^{(L)}, \quad L = 2, ..., M. \tag{1}$$

In Eq. (1), T is the "temperature," a free parameter of the algorithm, and the $\omega_{ji}^{(L)}$ are the connectivity weights.

The learning algorithm of the neural network consists in computing a set of weights $\omega_{ji}^{(L)}$, connecting nodes of adjacent layers, such that the total quadratic error in the identification of a given training set of events is minimized. Minimization by means of the gradient descent method leads to the back-propagation learning rule [10] employed here. For each event of the training set, the neural network output $x^{(M)}$ is compared with the target value t ($t = 0$ for background events, $t = 1$ for signal events) and the connectivity weights, which at the beginning are chosen at random, are modified accordingly by

$$\Delta\omega_{ji}^{(L)} = -\eta\delta_j^{(L)} x_i^{(L-1)} + \alpha\Delta\omega_{ji}^{(L)\text{old}}, \qquad i = 1, ..., N^{(L-1)}, \quad j = 1, ..., N^{(L)} \tag{2}$$

where the $\Delta\omega_{ji}^{(L)\text{old}}$ are the changes in the weights of the previous iteration, and the $\delta_j^{(L)}$ is defined by recursion:

$$\delta^{(M)} = x^{(M)}\left(x^{(M)} - t\right)\left(1 - x^{(M)}\right), \tag{3}$$

$$\delta_j^{(L)} = \sum_{k=1}^{N^{(L+1)}} \delta_k^{(L+1)} \omega_{kj}^{(L+1)} x_j^{(L)}\left(1 - x_j^{(L)}\right). \tag{4}$$

Here η and α are two free parameters of the learning algorithm. During this learning phase, the neural network is fed each event of the entire training set N_{cycle} times, and the weights are updated according to Eqs. (2)–(4). Then the weights are held fixed and the neural network is ready to be used as a signal-background classifier. At this point the network performance can be tested on a sample of new events that it has not previously encountered, the testing set. With such a neural network, therefore, we attempt to separate the one-lepton top-quark signal from background, and compare our results with the method of conventional kinematical cuts.

The training and testing event sets for top-quark production and the relevant background are simulated by the Monte Carlo event generator PYTHIA [11], at $p\bar{p}$ center of mass energy $\sqrt{s}$ =1.8 TeV. The total $t\bar{t}$ production cross section generated by PYTHIA using Eichten-Hinchliffe-Lane-Quigg (EHLQ) set 1 parton distributions [12] is used for normalization. We have checked that these results agree within theoretical uncertainty with QCD leading-log cross sections, even though certain interference terms are neglected in the string picture. We do not include any next-to-leading-log effects in our total $t\bar{t}$ cross section; this is a conservative choice, since these effects generally enhance the total cross section. Spin correlations are, of course, neglected in PYTHIA; this is appropriate for top-quark masses less than $\sim$ 140 GeV.

The W+multijet background is generated by PYTHIA from $q\bar{q} \to Wg$ and $qg \to Wq$ subprocesses, where the divergences are regulated by requiring p_T of the final-state parton to be greater than 10 GeV; additional jets are then generated by initial- and final-state QCD shower corrections. The specific lepton-plus-multijet cross section normalization is not well determined by this procedure [13], so we fix the normalization at values obtained from the next-to-leading-log plus shower approach of Ref. [14], which agrees well with matrix element predictions from Refs. [7] and [13]. We invoke a simple detector simulation, with calorimeter cells of size $\delta\eta \times \delta\phi = 0.1 \times 0.1$, with cells extending between $-3.5 < \eta < 3.5$ in pseudorapidity η. We assume hadronic energy resolution $\Delta E/E = 50\%/\sqrt{E}$.

We require that each generated event contain: (a) one and only one electron or muon with $p_T > 20$ GeV and pseudorapidity $|\eta_l| < 3.0$, and (b) total missing en-

ergy $\not{E}_T > 20$ GeV. We require that each event satisfy (c) lepton isolation defined by limiting the sum of hadronic energy $\sum_c E_T < 3$ GeV inside a cone of size $\Delta r \equiv \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} = 0.4$ centered about the electron momentum, and finally that each event contain (d) at least 3 hadronic jets, each of energy $E_{\rm jet} > 15$ GeV and pseudorapidity $|\eta_{\rm jet}| < 2.5$. The cuts in η correspond approximately to the acceptance of the collider detector at Fermilab (CDF), as does the jet definition for which we have taken the cone size to be $\Delta r = 0.7$. The combination of the severe cut on lepton and neutrino p_T, along with the lepton isolation requirement, results in a large suppression of backgrounds due to $b\bar{b}$ and $c\bar{c}$ pairs. These cuts (a), (b), and (c), however, do little to suppress the W background; on the contrary, these are precisely the kind of cuts usually invoked to select a leptonically decaying W signal. It is in fact the deep similarity between signal and background topology which makes it so difficult to distinguish between them. In order to reduce the background to the same order of magnitude as the signal cross section, therefore, we must also require (d). The resulting signal cross section, for the three cases $m_t = 100, 120, 140$ GeV, and the background rate from W-plus-multijet production, with the so-called acceptance cuts (a)–(d), are presented in the first three columns of Table I; we have checked that these rates are consistent with those of Refs. [6], [7], and [14]. Notice that the rate for background from W production is still often considerably larger than that of the signal, especially for the larger values of m_t.

We begin by training the neural network to distinguish between the top-quark signal and the W-plus-multijets background. As inputs for the neural network we choose the following 10 kinematical variables to describe each event: the p_T and pseudorapidity η_l of the hardest lepton, the number of jets $N_{\rm jet}$ satisfying requirement (d) above, the missing transverse energy of the event $\not{E}_T$, the energy $E^i_{\rm jet}$, and angular separation with respect to the lepton momentum $\Delta r^i_{\rm jet}$ for the three hardest jets $i = 1, 2, 3$. We choose then a network architecture based on ten input nodes, one hidden layer of 12 nodes, and a single output node ($M = 3$, $N^{(1)} = 10$, $N^{(2)} = 12$, $N^{(3)} = 1$) and train the neural network on 10 000 events generated by PYTHIA. The weights are updated every time after the net has been presented with a group of 20 events, consisting of mixed signal and background events arranged at random, but in a proportion to their predicted relative cross sections. For $m_t = 100$ GeV, the proportion is 9 signal and 11 background events, for $m_t = 120$ GeV, we use 6 signal and 14 background, and for $m_t = 140$ GeV, 3 signal and 17 background. The entire training set of events is presented to the net $N_{\rm cycle} = 10\,000$ times. The neural network is simulated by the program JETNET [15] with the following choice of parameters: temperature $T = 1.0$ [see Eq. (1)], $\alpha = 0.5$, and learning parameter $\eta = 0.0001$ [see Eq. (2)]. We have checked that modifications in this choice of parameters do not largely affect the neural network performance.

We next test the network performance by holding the weights fixed and confront the neural network with a fresh set of 2500 + 2500 signal and background events. The performance of the neural network is illustrated in Fig. 1 where the 5000 events are distributed according to their network output $x^{(3)}$ [see Eq. (1)]. For each top-quark mass there are two distributions, one for the 2500 true signal events and one for the 2500 true background events. A strict midpoint rule is employed here so that an event with output $x^{(3)} > 0.5$ ($x^{(3)} < 0.5$) is interpreted by the neural network as signal (background). Correct signal classifications, therefore, are regarded as only those true signal events to the right of one-half. One may notice the decided lack of events in the first and last bins of neural network output for $m_t = 100$ GeV. This dearth of events is due to the fact that, for m_t close to the W mass, the signal and background events are so similar that the neural network identifies few events that are so clearly signal-like or backgroundlike that they trigger the extrema values of output $x^{(3)}$. As the top-quark mass increases and the signal and background events become more dissimilar, these first and last bins gradually fill. One also notices that, as the top-quark mass increases with respect to the W mass, the identification accuracy improves, since the top quark decay products then have larger transverse momentum than in the case of the background. The overall neural network performance, measured in terms of total correct event classifications, is summarized in Table II.

Figure 1 is shown to illustrate clearly the neural network performance in absolute terms, but what is more important, of course, is how the neural network performs in an experimental situation, that is, a situation in which the testing set of events is presented to the network in a proportion normalized by their respective cross sections. Once the weights ω_{ji} are fixed, however, the network response to any given event is fixed, so we multiply the

TABLE I. Signal and background cross sections with acceptance cuts (a)–(d) ($\sigma^{S,B}_{\rm acc}$), with conventional severe set of cuts (a)–(g) ($\sigma^{S,B}_{\rm sev}$), as described in the text, and with cut on last three bins of neural network output ($\sigma^{S,B}_{\rm 13b}$).

m_t (GeV)	$\sigma^S_{\rm acc}$ (pb)	$\sigma^B_{\rm acc}$ (pb)	$\frac{\sigma^S_{\rm acc}}{\sigma^B_{\rm acc}}$	$\sigma^S_{\rm sev}$ (pb)	$\sigma^B_{\rm sev}$ (pb)	$\frac{\sigma^S_{\rm sev}}{\sigma^B_{\rm sev}}$	$\sigma^S_{\rm 13b}$ (pb)	$\sigma^B_{\rm 13b}$ (pb)	$\frac{\sigma^S_{\rm 13b}}{\sigma^B_{\rm 13b}}$
100	13.3	17.2	0.77	6.2	3.2	1.9	4.3	1.2	3.6
120	6.6	17.2	0.38	3.9	3.2	1.2	2.5	0.6	4.2
140	3.2	17.2	0.19	2.1	3.2	0.65	1.0	0.2	5.0

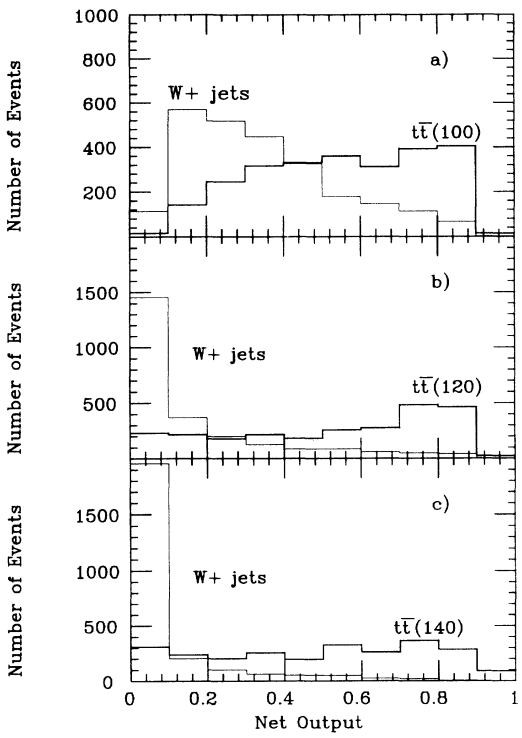

FIG. 1. Distribution of neural network output for test sets composed of equal numbers of signal and background events. For (*a*) $m_t = 100$ GeV, (*b*) $m_t = 120$ GeV, and (*c*) $m_t = 140$ GeV.

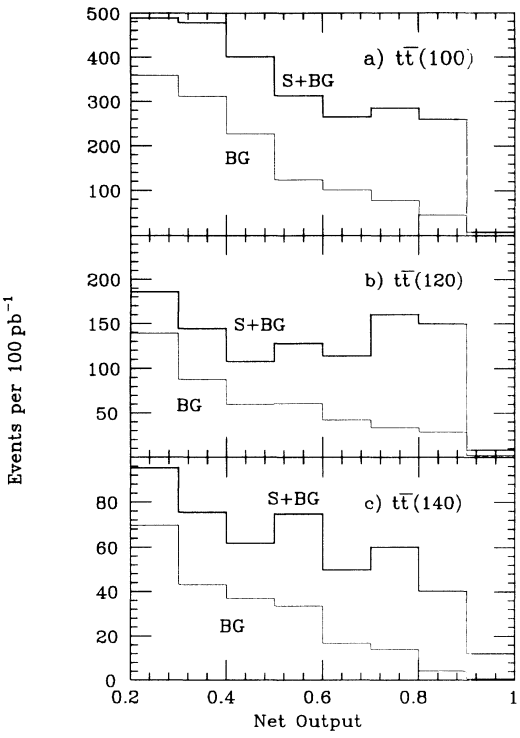

FIG. 2. Distribution of neural network output for the same events of Fig. 1, but for which the number of events is normalized according to the respective cross sections, and corresponds to the number expected per year assuming a luminosity $\mathcal{L} = 100$ pb^{-1}/yr. In (*a*) is shown separately the background events and the sum of $m_t = 100$ GeV signal-plus-background events, in (*b*), the same for $m_t = 120$ GeV, and in (*c*), for same for $m_t = 140$ GeV.

number of signal and background events in each bin of Fig. 1 by their respective cross sections and divide by the total number of events of that type. The results are shown in Fig. 2 wherein the background distribution and the sum of signal and background distributions are normalized according to the number of events expected per year at the Tevatron, assuming an integrated luminosity of 100 pb^{-1}/yr.

To evaluate the performance of the neural network, we compare the results in Table II with the Bayes decision rule (BDR), which is known [9] to minimize the average probability of error. In our case, the efficiency of the BDR for signal (background) recognition corresponds to the integration of the signal (background) probability distributions Σ_S (Σ_B) over the phase space where $\Sigma_S > \Sigma_B$ ($\Sigma_B > \Sigma_S$). In the last row of Table II we present the results of a numerical integration done by discretizing the 10-dimensional space of the kinematical variables used by the neural network as input. We have not tried to estimate the errors on the BDR results in Table II, which are certainly large. Nevertheless, one can see that neural network performance mimics the trend, at least, of the estimate of the BDR in that it increases with increasing

TABLE II. Neural network performance on test set of equal number of signal and background events.

m_t	100 GeV	120 GeV	140 GeV
Correct signal classifications	$\frac{1481}{2528} = 58.6\ \%$	$\frac{1504}{2528} = 59.5\ \%$	$\frac{1327}{2528} = 52.5\ \%$
Correct background classifications	$\frac{1970}{2472} = 79.7\ \%$	$\frac{2233}{2472} = 90.3\ \%$	$\frac{2373}{2472} = 96.0\ \%$
Correct total classifications	$\frac{3451}{5000} = 69.0\ \%$	$\frac{3737}{5000} = 74.7\ \%$	$\frac{3700}{5000} = 74.0\ \%$
Bayesian limit	70%	78%	80%

values of top-quark mass.

In order to obtain a final event sample that is most rich in signal events, we make a final cut upon the data of Fig. 2 by collecting those events in the last three bins of network output near one, i.e., $x^{(3)} > 0.7$. These bins contain the events most likely, to the trained eye of the neural network, to be most signal-like, and where one indeed finds the highest concentration of true signal events and the fewest background. In columns 7–9 of Table I we list the effective cross sections σ_{13b} for signal and background and the signal/background ratio that one obtains by cutting on the last three bins of neural network output. We compare these to the cross sections σ_{sev} corresponding to the best set of conventional kinematical cuts shown in columns 4–6 of Table I. These cuts are used to limit the presence of W-plus-multijets background [6] and will be known as the "severe cuts" in what follows. This set of cuts consists of all of our "acceptance cuts" listed above, and in addition include that: (*e*) the sum of E_T for each event exceeds 210 GeV, (*f*) the invariant mass $m(j_a j_b)$ of any two jets a and b satisfy $|m(j_a j_b) - M_W| < 15$ GeV, and (*g*) the transverse mass of the lepton and missing p_T, $m_T(l, p_T)$, satisfy 25 GeV$< m_T(l, p_T) <$100 GeV. The rates for σ_{sev} also were obtained from PYTHIA and the normalization has been altered in the same way described above as for σ_{acc}.

The cut on the last three bins of neural network output is found to be more efficient than the set of "severe cuts" as the ratio of signal/background that one obtains from the neural network is more favorable than the severe cuts case by a factor of 1.9 for m_t =100 GeV, 3.5 for m_t =120 GeV, and the neural network is 7.7 more effective for m_t =140 GeV. There is a certain trade-off at play here in the favor of the neural network in that, as the cross section for the signal falls with increasing top-quark mass, the neural network efficiency grows, whereas the ratio of signal/background for the severe cuts only suffers by the falling cross section for increasing top-quark mass.

In an effort to demonstrate top-quark mass reconstruction, we have plotted (but do not show here) the invariant mass of the three leading jets for signal and background events with $N_{\text{jet}} \geq 3$ and 4 subject to both the conventional set of cuts as well as subject to the acceptance plus neural network cut. Rather than obtaining a sharp peak at m_t as in the parton-level calculations of Ref. [6], we obtained a substantially broader distribution that did not adequately reflect the top-quark mass, at least for the range of masses which are considered here. This is due to a number of effects: (*i*) PYTHIA includes multiple parton radiations, which adds extra hard jets from the initial state, and broadens the jets from real W decay, (*ii*) the associated b jet can be misidentified or can be energetically too soft, and (*iii*) we have included the previously referred to hadronic energy smearing.

Finally it must be determined how well the neural network can succeed in identifying the top-quark signal under realistic experimental conditions in which the top-quark mass is not reliably known. For this purpose we train the neural network on the same 6:14 mix of signal and background events passing the "acceptance cuts" for m_t = 120 GeV as before and fix the weights, but then test the neural network on a test set of 5000 m_t = 100 GeV signal-plus-background events, as well as on a test set of m_t = 140 GeV signal and background events. The results are summarized in Table III. Upon comparing with the results of Table II, we find that the overall network performance is degraded by no more than 5% for the m_t = 100 GeV events, while overall network performance actually increases for the m_t = 140 GeV case. This is not a surprising result when one recalls that the m_t = 140 weights were designed to identify background events with the greatest efficiency (because training took place on a 3:17 mix of signal and background events), and were therefore the poorest at identifying the signal. It can then be expected that the signal identification will be dramatically increased with the use of the m_t = 120 weights, and so boost the overall performance measure.

TABLE III. Neural network performance on test sets composed of signal events for m_t = 100 and 140 GeV using the weights of training on m_t = 120 GeV signal.

m_t	100 GeV	140 GeV
Correct signal classifications	$\frac{1014}{2555}$ = 39.9 %	$\frac{1905}{2555}$ = 74.6 %
Correct background classifications	$\frac{2206}{2445}$ = 90.2 %	$\frac{2206}{2445}$ = 90.2 %
Correct total classifications	$\frac{3220}{5000}$ = 64.4 %	$\frac{4111}{5000}$ = 82.2 %

To illustrate these cross-testing results more clearly in an experimental situation, we plot in Fig. 3 the neural network output for the two cases above such that the

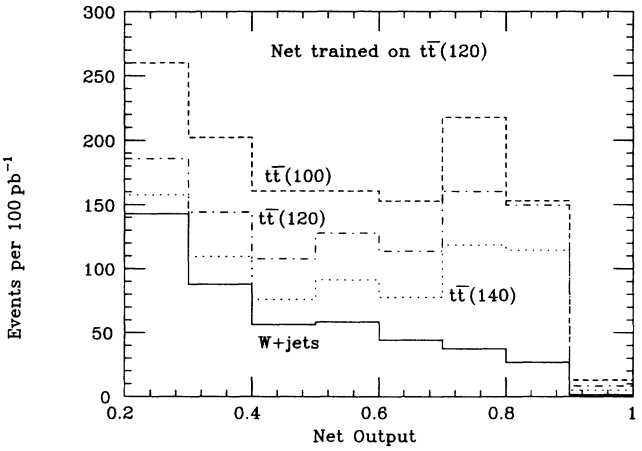

FIG. 3. Distribution of neural network output for test sets of m_t = 100 GeV (dashed line) and m_t = 140 GeV (dotted line) using the weights obtained by training on the m_t = 120 GeV events. Numbers of signal and background events are normalized according to respective cross sections and correspond to the number expected per year assuming $\mathcal{L}$ = 100 pb^{-1}/yr. For comparison we also show the background distribution (solid line) and the results for the m_t = 120 test set (dot-dash line).

number of signal and background events are normalized according to their respective cross sections. We show the network output distribution for signal-plus-background events for $m_t = 100$ and 140 GeV, as well as the background events alone. For comparison we also include the signal-plus-background distribution for $m_t = 120$ GeV, i.e., the test set which has events of the same top-quark mass as the training set. We show only the distribution in network output $x^{(3)} > 0.2$ in order to cut out the large background peak and to see more clearly the network response to the various signals. For all top-quark mass values one sees a clear increase in occupation of the bins for $x^{(3)} > 0.6$ over the case for which there are only background events. We propose that, by comparing such distributions with experimental data, one may not only confirm the existence of top-quark events in the event sample, but also extract a value for the top-quark mass.

The authors wish to thank Duane Dicus for participation, advice, and many helpful discussions of this work. They have also benefited from discussions with David Chao. Computing resources were provided in part by the University of Texas Center for High Performance Computing. This research has been supported in part by the U.S. Department of Energy Grant No. DOE-FG05-85ER40200 (D.D.K.) and by the Robert A. Welch Foundation and NSF Grant No. PHY9009850 (G.F.G.).

[1] For a review, see J. C. Taylor, *Gauge Theories of Weak Interactions* (Cambridge University Press, Cambridge, England, 1976).

[2] JADE Collaboration, W. Bartel *et al.*, Phys. Lett. **146B**, 437 (1984).

[3] V. Barger and S. Pakvasa, Phys. Lett. **81B**, 195 (1979); G. Kane and M. Peskin, Nucl. Phys. **B195**, 29 (1982); CLEO Collaboration, A. Bean *et al.*, Phys. Rev. D **35**, 3533 (1987).

[4] ALEPH Collaboration, D. Decamp *et al.*, Z. Phys. C **53**, 1 (1992).

[5] CDF Collaboration, F. Abe *et al.*, Phys. Rev. Lett. **64**, 142 (1990); Phys. Rev. D **43**, 664 (1991); L. Nodulman (private communication).

[6] H. Baer, V. Barger, and R.J.N. Phillips, Phys. Rev. D **39**, 3310 (1989).

[7] W. Giele and W. Stirling, Nucl. Phys. **B343**, 14 (1990).

[8] L. Lönnblad, C. Peterson, and T. Rögnvaldsson, Phys. Rev. Lett. **65**, 1321 (1990); B. Denby, Report No. FERMILAB-CONF-90-94, 1990 (unpublished); T.D. Gottschalk and R. Nolty, Report No. CALT-68-1680, 1990 (unpublished); G. Stimpfl-Abele and L. Garrido, Comput. Phys. Commun. **64**, 46 (1991); D. Cutts *et al.*, Report No. DOE-ER-03130-52C, 1990 (unpublished); P. Bhat *et al.*, Report No. DESY-90-144, 1990 (unpublished); C.S. Lindsey and B. Denby, Nucl. Instrum. Methods A **302**, 217 (1991); I. Csabai, F. Czako, and Z. Fodor, Phys. Rev. D **44**, 1905 (1991); Nucl. Phys. **B374**, 288 (1992); L. Bellantoni *et al.*, Nucl. Instrum. Methods A **310**, 618 (1991); R.A. Vazquez, F. Halzen, and E. Zas, Phys. Rev. D **45** 356 (1992); L. Garrido and V. Gaitan, Report No. UAB-LFAE-91-04, 1991 (unpublished); C. Bortolotto, A. De Angelis, and L. Lanceri, Nucl. Instrum. Methods A **306** 459 (1991); L. Gupta *et al.*, Report No. FERMILAB-PUB-91-117, 1991 (unpublished); L. Lönnblad, C. Peterson, and T. Rögnvaldsson, Phys. Lett. B **278**, 181 (1992); C. Peterson, Report No. LU-TP-92-5, 1992 (unpublished); A. Cherubini and R. Odorico, Report No. DFUB-91-11, 1991 (unpublished); Z. Phys. C **53**, 139 (1992); G. Barbagli, G. D'Agostini, and D. Monaldi, Report No. ROME-N-992-1992, 1992 (unpublished); G. Bahan and R. Barlow, Report No. MAN-HEP-92-1, 1992 (unpublished); S. Bianchin *et al.*, Report No. INFN-AE-92-14C, 1992 (unpublished); G. Marchesini, G. Nardulli, and G. Pasquariello, Report No. BARI-TH-92-98, 1992 (unpublished).

[9] R.O. Duda and P.E. Hart, *Pattern Classification and Scene Analysis* (Wiley, New York, 1973).

[10] D.E. Rumelhart, G.E. Hinton, and R.J. Williams, in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, edited by D.E. Rumelhart and J.L. McClelland (MIT Press, Cambridge, MA, 1986), Vol. 1.

[11] H.-U. Bengtsson and T. Sjöstrand, PYTHIA 5.5 program and manual; see H.-U. Bengtsson and T. Sjöstrand, Comput. Phys. Commun. **46**, 43 (1987).

[12] E. Eichten, I. Hinchliffe, K. Lane, and C. Quigg, Rev. Mod. Phys. **56**, 579 (1984).

[13] W. Giele, T. Matsuura, M. Seymour, and B. Webber, in *Research Directions for the Decade*, Proceedings of Snowmass Summer Study on High Energy Physics, Snowmass, Colorado, 1990, edited by E.L. Berger and I. Butler (World Scientific, Singapore, 1991).

[14] H. Baer and M. H. Reno, Phys. Rev. D **45**, 1503 (1992).

[15] L. Lönnblad, C. Peterson, and T. Rögnvaldsson, Nucl. Phys. **B349**, 675 (1991).