Strategy for discovering a low-mass Higgs boson at the Fermilab Tevatron

Pushpalatha C. Bhat
Fermi National Accelerator Laboratory, Batavia, Illinois 60510

Russell Gilmartin and Harrison B. Prosper
Department of Physics, Florida State University, Tallahassee, Florida 32306

(Received 14 January 2000; published 12 September 2000)

We have studied the potential of the CDF and DØ experiments to discover a low-mass standard model Higgs boson, during run II, via the processes $p\bar{p}\to WH\to l\nu b\bar{b}$, $p\bar{p}\to ZH\to l^+l^-b\bar{b}$ and $p\bar{p}\to ZH\to \nu\bar{\nu}b\bar{b}$. We show that a multivariate analysis using neural networks that exploits all the information contained within a set of event variables leads to a significant reduction, with respect to *any* equivalent conventional analysis, in the integrated luminosity required to find a standard model Higgs boson in the mass range 90 GeV/$c^2<M_H<130$ GeV/c^2. The luminosity reduction is sufficient to bring the discovery of the Higgs boson within reach of the Fermilab Tevatron experiments, given the anticipated integrated luminosities of run II, whose scope has recently been expanded.

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

I. INTRODUCTION

The success of the standard model (SM) of particle physics, which provides an accurate description of almost all particle phenomena observed so far [1–3], has been spectacular. However, one crucial aspect of it remains mysterious: the fundamental mechanism that underlies electroweak symmetry breaking (EWSB) and the origin of fermion mass. Elucidating the nature of EWSB is the next major challenge of particle physics and will be the focus of upcoming experiments at the Fermilab Tevatron and the CERN Large Hadron Collider (LHC) during the early years of the twenty-first century.

In many theories, EWSB occurs through the interaction of one or more doublets of scalar (Higgs) fields with the initially massless fields of the theory. An important goal over the next decade is to determine whether or not, in broad outline, this picture of EWSB is correct. In the standard model there is a single scalar doublet. The EWSB endows the weak bosons ($W^{\pm}$,Z) with masses and gives rise to a single physical neutral scalar particle called the Higgs boson (H_{SM}). In minimal supersymmetric (SUSY) extensions of the SM, two Higgs doublets are required resulting in five physical Higgs bosons: two neutral CP-even scalars (h,H), a neutral CP-odd pseudoscalar (A) and two charged scalars ($H^{\pm}$). Non-minimal SUSY theories generally posit more than two scalar doublets.

Given this picture of EWSB, the direct and indirect measurements of the top quark and W boson masses constrain the mass of the SM Higgs boson ($M_{H_{SM}}$), as indicated in Fig 1. A global fit to all electroweak precision data, including the top quark mass, gives a central value of $M_{H_{SM}}=107^{+67}_{-45}$ GeV/c^2 and a 95% confidence level upper limit of 225 GeV/c^2 [1]. In broad classes of SUSY theories the mass M_h of the lightest CP-even neutral Higgs boson h is constrained to be less than 150 GeV/c^2 [5]. In the minimal supersymmetric SM (MSSM), the upper bound on M_h is lowered to about 130 GeV/c^2 [6,7]. This bound is reasonably robust with respect to changes in the parameters of the theory. Furthermore, in the limit of large pseudoscalar Higgs boson mass, $M_A\gg M_Z$, where M_Z is the mass of the Z boson, the properties of the lightest MSSM Higgs boson h are indistinguishable from those of the SM Higgs boson, H_{SM}. These intriguing indications of a low-mass Higgs boson motivate the study of strategies that maximize the potential for its discovery at the upgraded Tevatron [8]. This paper describes a strategy that achieves this goal.

The current 95% C.L. lower limit on the Higgs boson mass, from the CERN e^+e^- collider LEP, is 107.9 GeV/c^2 [9] and is expected to reach close to 114 GeV/c^2 [7] in the near future. We have therefore studied the mass range 90 GeV/$c^2<M_H<130$ GeV/c^2, where H, hereafter, denotes the SM Higgs boson, H_{SM}. The cross sections for SM Higgs boson production at the Fermilab Tevatron are shown in Fig 2. At $\sqrt{s}=2$ TeV, the dominant process for the production of Higgs bosons in $p\bar{p}$ collisions is $gg\to H$. The

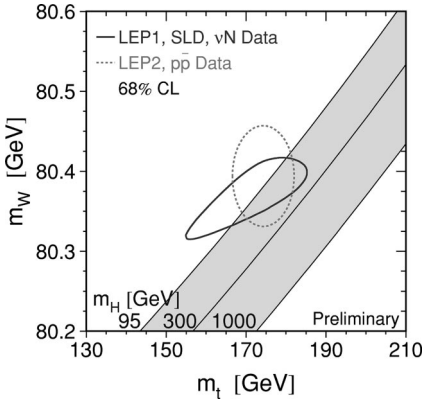

FIG. 1. The correlation between the W boson mass and the top quark mass as predicted by the standard model, for various possible values of the Higgs boson mass. (Each line corresponds to the mass value shown.) Also shown are the 68% C.L. contours from direct (dashed contour) and indirect (solid contour) measurements of the W boson and top quark mass. From Ref. [4].

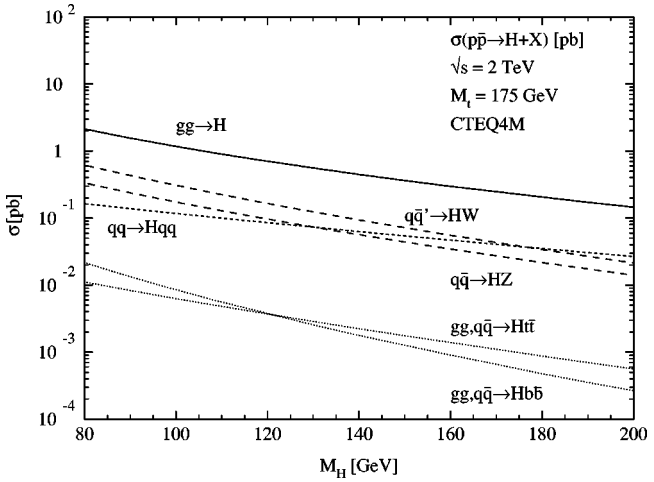

FIG. 2. Cross sections for various Higgs production processes in $p\bar{p}$ collisions at $\sqrt{s}=2$ TeV as a function of Higgs boson mass. From Ref. [10].

Higgs boson decays to a $b\bar{b}$ pair about 85% of the time. Unfortunately, even with maximally efficient *b*-tagging this channel is swamped by QCD di-jet production. The more promising channels are $p\bar{p}\to WH\to l\nu b\bar{b}$, $p\bar{p}\to ZH\to l^+l^-b\bar{b}$ and $p\bar{p}\to ZH\to\nu\bar{\nu}b\bar{b}$, which are the ones we have studied.

In *WH* events the lepton can be lost because of deficiencies in the detector or the event reconstruction or the lepton energy being below the selection threshold. For such events the *reconstructed* final state would be indistinguishable from that arising from the process $p\bar{p}\to ZH\to\nu\bar{\nu}b\bar{b}$. We have therefore studied these processes in terms of the channels: *single lepton* ($l+\not{E}_T+b\bar{b}$ from *WH*), *di-lepton* ($l^+l^-b\bar{b}$ from *ZH*) and *missing transverse energy* ($\not{E}_T+b\bar{b}$ from *ZH* and *WH*), where $\not{E}_T$ denotes the missing transverse energy from all sources, including neutrinos. For each of these channels, we have carried out a comparative study of multivariate and conventional analyses of these channels in which we compare signal significance and the integrated luminosity needed for discovery.

The paper is organized as follows: In Sec. II we describe our strategy in general terms. Sections III, IV and V, respectively, describe our analyses of the single lepton, di-lepton and missing transverse energy channels. Our conclusions are given in Sec. VI.

II. OPTIMAL EVENT SELECTION

In conventional analyses a cut is applied to each event variable, usually one variable at a time, after a visual examination of the signal and background distributions. Although analyses done this way are sometimes described as ''optimized,'' in practice, unless the signal and background distributions are well separated, the traditional procedure for choosing cuts is rarely optimal in the sense of *minimizing the probability to mis-classify events.* Since we wish to maximize the chance of discovering the Higgs boson we need to achieve the optimal separation between signal and background, while maximizing the signal significance. Given any set of event variables, optimal separation can always be achieved if one treats the variables in a fully *multivariate* manner.

Given a set of event variables, it is useful to construct the discriminant function *D* given by

$$D=\frac{s(\mathbf{x})}{s(\mathbf{x})+b(\mathbf{x})}, \tag{2.1}$$

where $\mathbf{x}$ is the vector of variables that characterize the events and $s(\mathbf{x})$ and $b(\mathbf{x})$, respectively, are the *n*-dimensional probability densities describing the signal and background distributions. The discriminant function $D=r/(1+r)$ is related to the *Bayes discriminant function* which is proportional to the likelihood ratio $r\equiv s(\mathbf{x})/b(\mathbf{x})$. Working with *D*, instead of

TABLE I. Cross section times branching ratio for the *WH* and *ZH* processes we have studied, for various M_H [10] and for the various backgrounds. Note: For tb, $t\bar{t}$ and ZZ processes we give the total cross section.

$WH\to l\nu b\bar{b}$		$ZH\to l^+l^- b\bar{b}$		$ZH\to\nu\bar{\nu}b\bar{b}$	
M_H (GeV/c^2)	$\sigma\times BR$(fb)	M_H (GeV/c^2)	$\sigma\times BR$(fb)	M_H (GeV/c^2)	$\sigma\times BR$(fb)
90	119.0	90	20.3	90	40.6
100	85.4	100	14.8	100	29.6
110	62.3	110	10.9	110	21.8
120	45.3	120	8.22	120	16.4
130	34.1	130	6.25	130	12.5
Backgrounds					
$Wb\bar{b}$	3500.0	$Zb\bar{b}$	350.0	$Zb\bar{b}$	700.0
WZ	164.8				
tbq	800.0	tbq	800.0	tbq	800.0
	σ (fb)		σ (fb)		σ (fb)
		ZZ	1235.0	ZZ	1235.0
tb	1000.0	tb	1000.0	tb	1000.0
$t\bar{t}$	7500.0	$t\bar{t}$	7500.0	$t\bar{t}$	7500.0

TABLE II. Single lepton channel. Variables used in training the neural networks for signals against specific backgrounds.

$Wb\bar{b}$	WZ	$t\bar{t}$
E_T^{b1}	E_T^{b1}	E_T^{b1}
E_T^{b2}	E_T^{b2}	E_T^{b2}
$M_{b\bar{b}}$	$M_{b\bar{b}}$	$M_{b\bar{b}}$
H_T	H_T	H_T
E_T^l	E_T^l	$\not{E}_T$
S	S	$\Delta R(b_1,l)$
$\Delta R(b_1,b_2)$	η_l	$\Delta R(b_1,b_2)$

directly with $\mathbf{x}$, brings two important advantages: (1) it reduces a difficult n-dimensional optimization problem to a trivial one in a single dimension and (2) a cut on D can be shown to be optimal in the sense defined above.

There is, however, a practical difficulty in calculating the discriminant D. We usually do not have analytical expressions for the distributions $s(\mathbf{x})$ and $b(\mathbf{x})$. What is normally available are large discrete sets of points $\mathbf{x}_i$, generated by Monte Carlo simulations. Fortunately, however, there are several methods available to approximate the discriminant D from a set of points $\mathbf{x}_i$, the most convenient of which uses feed-forward neural networks. Neural networks are ideal in this regard because they approximate D directly [11,12].

Many neural network packages are available, any one of which can be used to calculate D. We have used the JETNET package [13] to train three-layer (that is, input, hidden and output) feed-forward neural networks (NN). The training was done using the back-propagation algorithm, with the target output for the signal set to one and that for the background set to zero. In this paper we use the terms ''neural network output'' and ''discriminant'' interchangeably. However, the

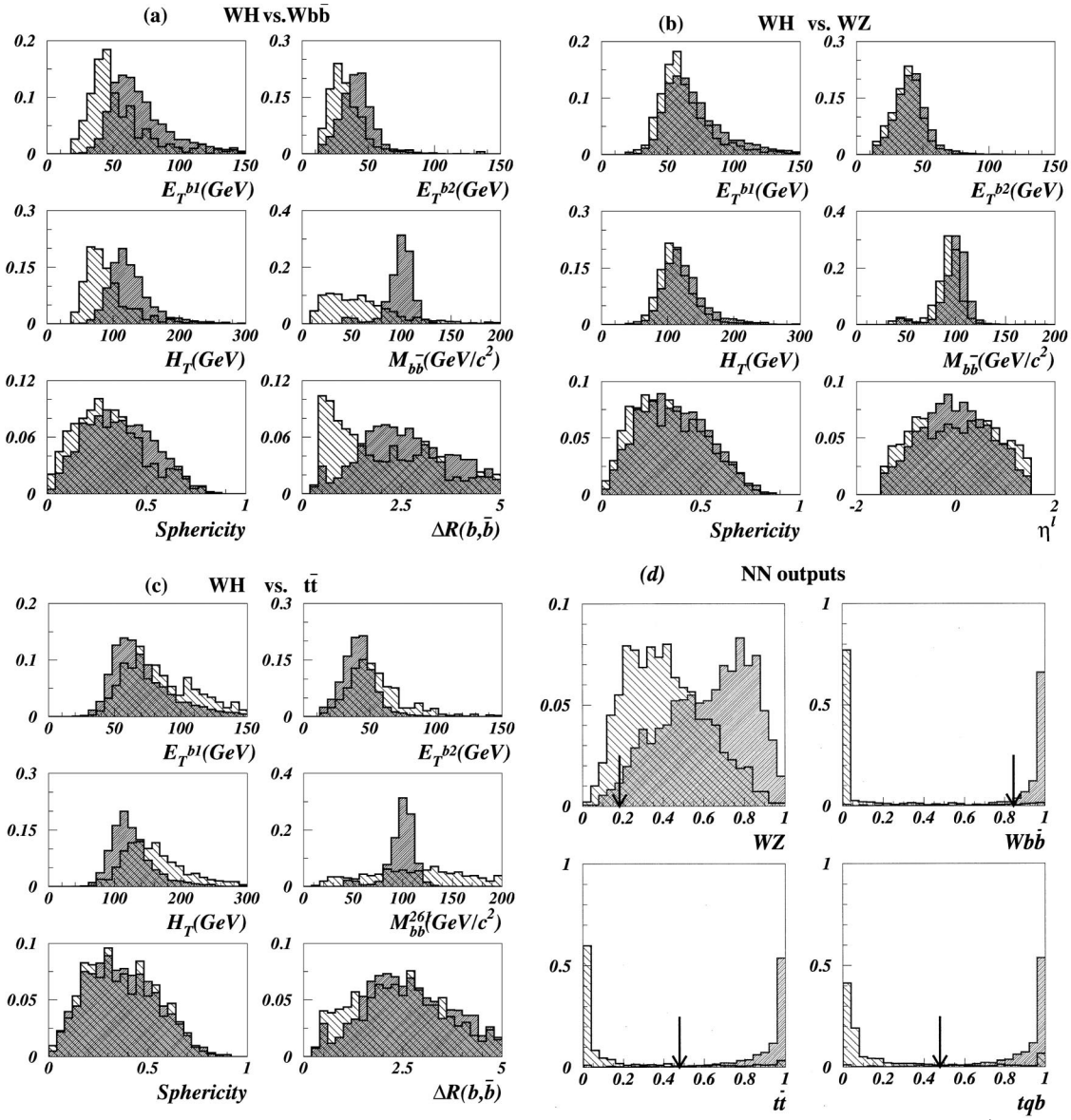

FIG. 3. Distributions of some of the variables used in the NN analysis for WH ($M_H = 100$ GeV/c) signal (heavily shaded) and backgrounds (lightly shaded) (a) WH vs $Wb\bar{b}$, (b) WH vs WZ, and (c) WH vs $t\bar{t}$. In (d) we compare the neural network output distributions for signal and various backgrounds. The arrows indicate the cuts.

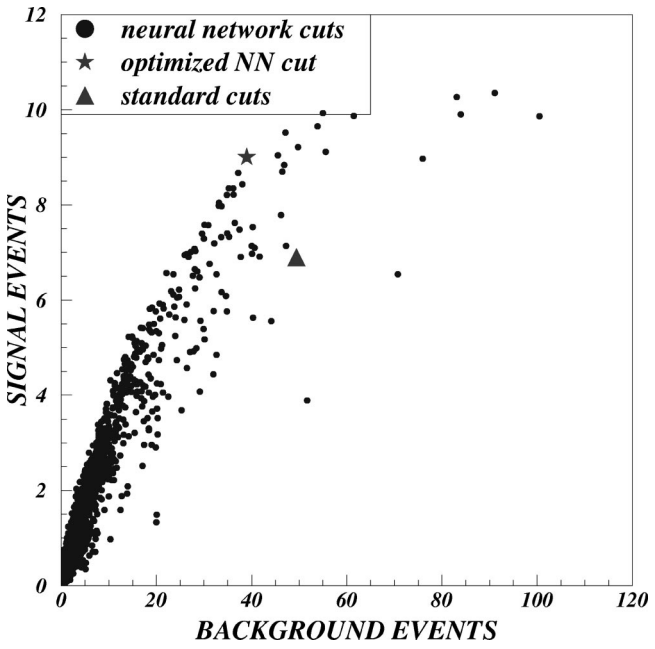

FIG. 4. Single lepton channel. The number of signal events vs number of background events for 1 fb^{-1} using various combination of cuts on the three neural network outputs. The standard cuts are optimized based on studies done in the Higgs working group using conventional methods.

distinction between the exact discriminant D, as we have defined it above, and the network output, which provides an estimate of D, should be borne in mind.

III. SINGLE LEPTON CHANNEL

We have considered final states with a high p_T electron (e) or muon (μ) and a neutrino from W decay and a $b\bar{b}$ pair from the decay of the Higgs boson. The WH events were simulated using the PYTHIA program [14] for Higgs boson masses of $M_H = 90$, 100, 110, 120 and 130 GeV/c^2. In Table I we list the cross section $\times$ branching ratio (BR) we have used for the process $p\bar{p} \to WH \to l\nu b\bar{b}$ where $l = e, \mu, \tau$.

The processes $p\bar{p} \to Wb\bar{b}$, $p\bar{p} \to WZ$, $p\bar{p} \to t\bar{t}$, single top production—$p\bar{p} \to W^* \to t\bar{b}$ and $p\bar{p} \to Wg \to tqb$, which have the same signature, $l\nu bb$, as the signal, are the most important sources of background. They have all been included in our study. The Wbb sample was generated using COMPHEP [15], a parton level Monte Carlo program based on exact leading order (LO) matrix elements. The parton fragmentation was done using PYTHIA. The single top, $t\bar{t}$ and WZ events were simulated using PYTHIA. To generate the s-channel process, $W^* \to tb$, we forced the W to be produced off-shell, with $\sqrt{\hat{s}} > m_t + m_b$, and then selected the final state in which $W \to tb$. The cross sections used for the background processes are given in Table I.

To model the expected response of the Collider Detector at Fermilab (CDF) and DØ run II detectors we used the SHW program [16], which provides a fast (approximate) simulation of the trigger, tracking, calorimeter clustering, event reconstruction and b-tagging. The SHW simulation predicts a di-jet mass resolution of about 14% at $M_H = 100$ GeV/c^2, varying only slightly over the mass range of interest. However, to allow for comparisons with the other WH and ZH studies at the Physics at Run II SUSY/Higgs workshop [8], some of which do not use SHW, we have re-scaled the di-jet mass variables for all signal and background events so that the resolution is 10% at each Higgs boson mass. The consensus of Run II workshop is that such a mass resolution can be achieved, albeit with considerable effort.

TABLE III. Single lepton channel. Results for the number of signal and background events (top portion of the table) for 1 fb^{-1} of integrated luminosity. The cuts on the network outputs were chosen to yield maximum significance for each Higgs boson mass, leading to different background counts at each mass.

M_H GeV/c^2	90	100	110	120	130
Number of events (1 fb^{-1})					
WH	8.65	8.97	4.81	4.41	3.71
$Wb\bar{b}$	12.28	12.48	5.84	9.66	20.12
WZ	7.52	10.32	1.72	1.00	0.97
tqb	0.51	0.95	0.58	0.71	0.96
tb	2.46	5.40	3.44	5.89	9.33
$t\bar{t}$	5.63	9.89	7.24	8.39	14.49
Total background	28.40	39.04	18.81	25.67	45.87
Signal significance					
S/B	0.31	0.23	0.26	0.17	0.081
$S/\sqrt{B}$ (1 fb^{-1})	1.62	1.44	1.11	0.87	0.55
$S/\sqrt{B}$ (2 fb^{-1})	2.29	2.04	1.57	1.23	0.78
$S/\sqrt{B}$ (30 fb^{-1})	8.87	7.89	6.08	4.77	3.01
Required luminosity (fb^{-1})					
5σ	9.5	12.1	20.3	33.0	82.6
3σ	3.4	4.3	7.3	11.9	29.8
1.96σ (95% C.L.)	1.5	1.9	3.1	5.1	12.7

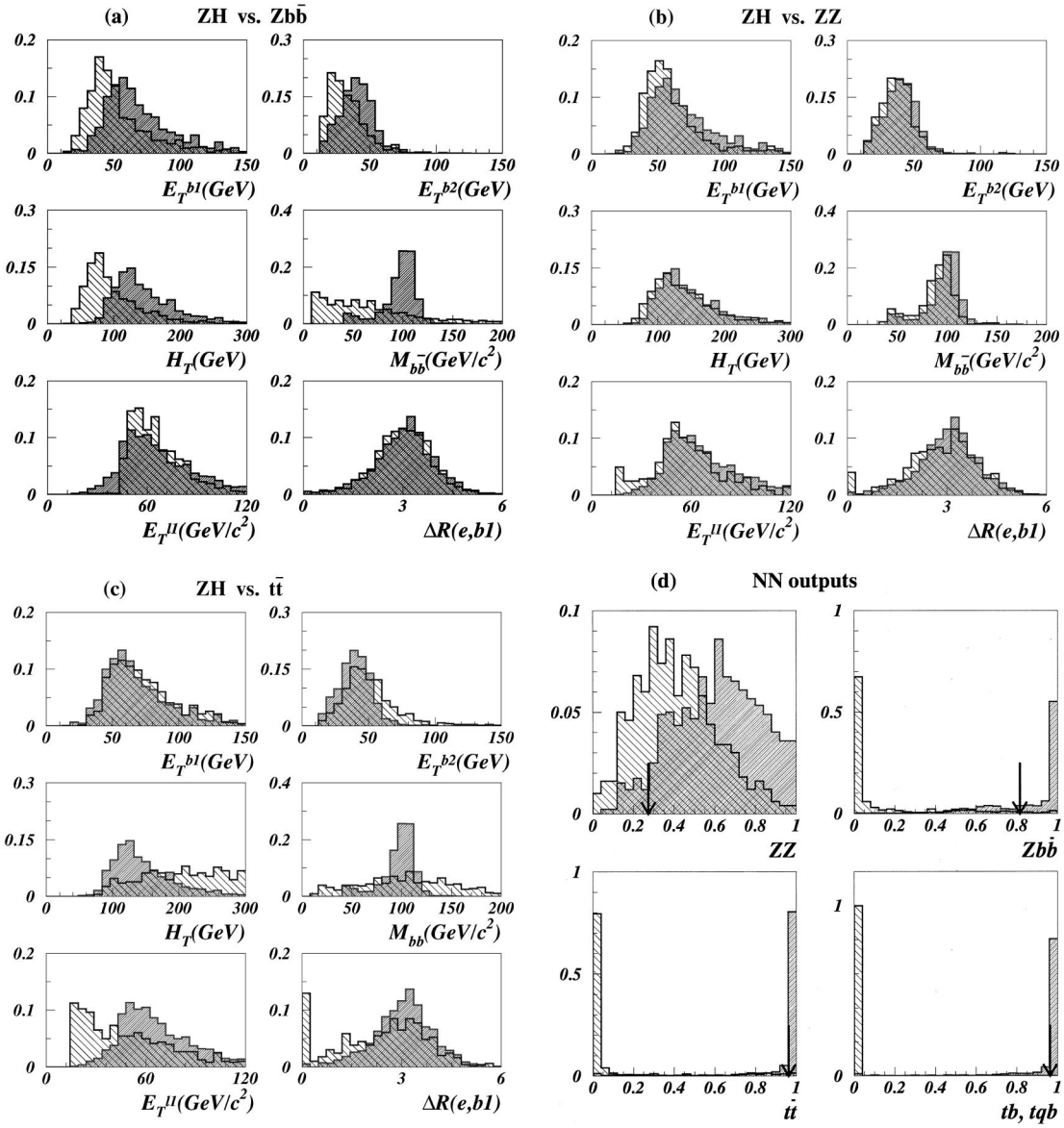

FIG. 5. Di-lepton channel. Distributions of variables used in training the neural networks for signal (with $M_H = 100$ GeV/c) and different backgrounds and the results of the trained networks. (a) Signal vs $Zb\bar{b}$ background; (b) signal vs ZZ background; (c) signal vs $t\bar{t}$ background and (d) distributions of neural network outputs for networks trained using signal vs the backgrounds ZZ, $Zb\bar{b}$ and $t\bar{t}$. The signal histograms are heavily shaded. The arrows indicate the cuts.

In principle, multivariate methods can be applied at all stages of an analysis. However, in practice, experimental considerations, such as trigger thresholds and the need to restrict data to the phase space in which the detector response is well understood, dictate a set of loose cuts on the event variables. These cuts define a *base* sample of events. In our case, the base sample was determined by the following cuts:

(i) the transverse momentum of the isolated lepton $P_T^l > 15$ GeV/c

(ii) the pseudo-rapidity of the lepton $|\eta_l| < 2$

(iii) the missing transverse energy in the event $\not{E}_T > 20$ GeV

(iv) two or more jets in the event with $E_T^{jet} > 10$ GeV and $|\eta_{jet}| < 2$.

Since the Higgs decays into a $b\bar{b}$ pair we impose the requirement that two jets be b-tagged. This of course does little to reduce the dominant Wbb background, due to the presence of the bb pair, but it becomes powerful when the invariant mass, $M_{b\bar{b}}$, of the b-tagged jets is used as an event variable. The di-jet mass distributions for the signal is expected to peak at the Higgs boson mass, whereas one expects a broad distribution for the background, with the exception of the WZ background which peaks at the Z boson mass.

One of the b-tags was required to be *tight* and the other *loose* [16]. A tight b-tag is defined by an algorithm that uses the silicon vertex detector, while a loose b-tag is defined by the same algorithm with looser cuts or by a soft lepton tag [16]. The mean double b-tagging efficiency in SHW is about 45%.

TABLE IV. Di-lepton channel. Results for 1 fb^{-1}.

M_H (GeV/c^2)	90	100	110	120	130
Number of events					
ZH	1.26	0.87	0.79	0.80	0.58
$Zb\bar{b}$	0.61	0.45	0.61	1.50	1.42
ZZ	2.04	1.44	1.42	0.83	0.31
$t\bar{t}$	0.28	0.05	0.23	0.44	0.18
Total background	2.93	1.94	2.26	2.77	1.91
S/B	0.43	0.45	0.35	0.29	0.31
$S/\sqrt{B}$	0.74	0.63	0.54	0.48	0.42

We searched for variables that discriminate between the signal and the backgrounds and arrived at the following set:

(i) E_T^{b1},E_T^{b2}–transverse energies of the b-tagged jets
(ii) $M_{b\bar{b}}$–invariant mass of the b-tagged jets
(iii) H_T–sum of the transverse energies of all selected jets
(iv) E_T^l–transverse energy of the lepton
(v) η_l–pseudo-rapidity of the lepton
(vi) $\not{E}_T$–missing transverse energy
(vii) S–sphericity $[S=\frac{3}{2}(Q_1+Q_2)]$ where Q_i are the eigenvalues obtained by diagonalizing the normalized momentum tensor $M_{ab}=\Sigma_i p_{ia}p_{ib}/\Sigma_i|p_i|^2$ where the sums are over the final state particle momenta and the subscripts a and b refer to the spatial components of the momenta p_i

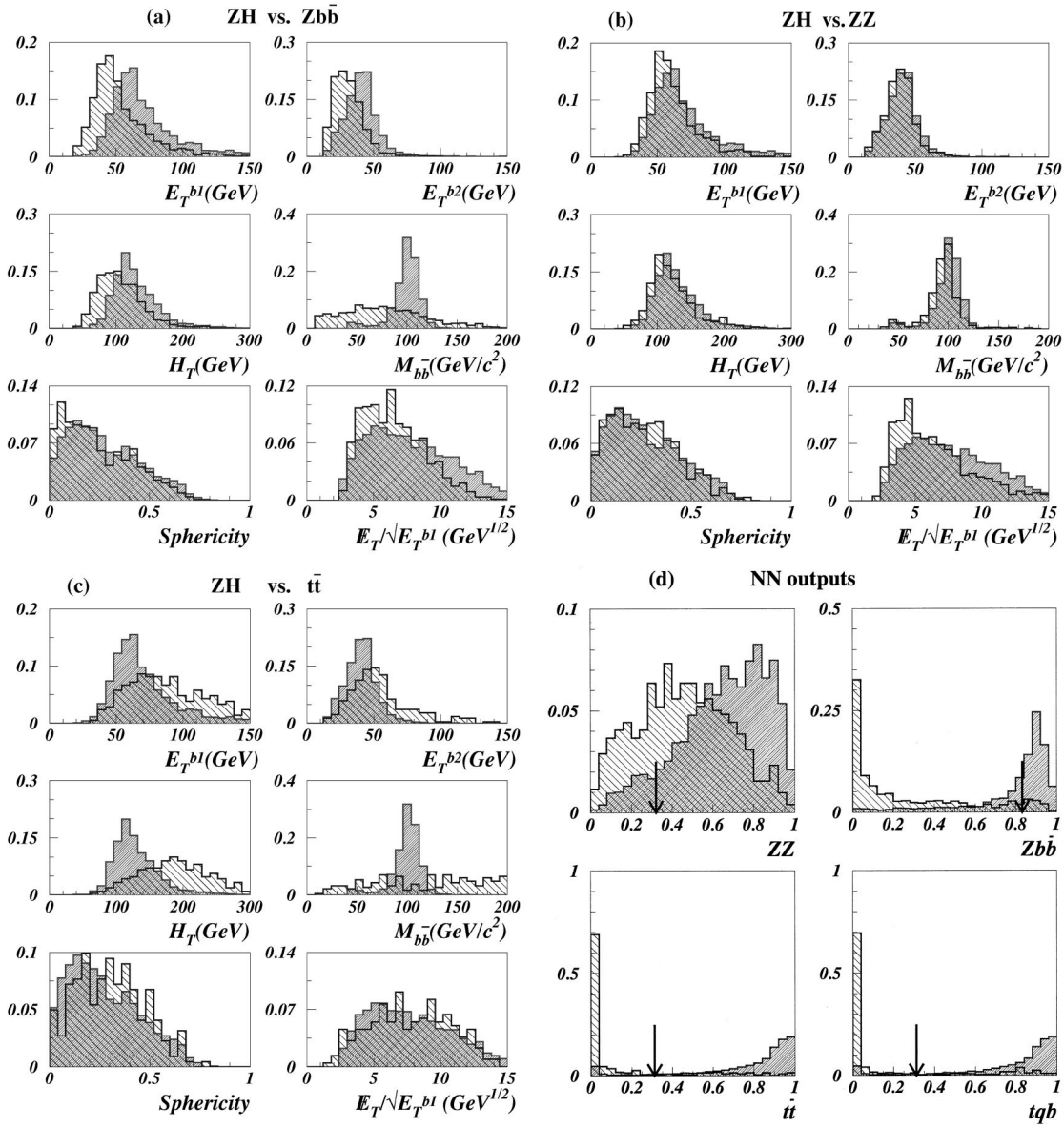

FIG. 6. Missing transverse energy channel. Distributions of variables used in training the neural networks for signal (with $M_H=100$ GeV/c^2) and different backgrounds, together with distributions of network outputs. (a) Signal vs $Zb\bar{b}$; (b) signal vs ZZ; (c) signal vs $t\bar{t}$ and (d) distributions of neural network outputs for networks trained using signal vs the backgrounds ZZ, $Zb\bar{b}$ and $t\bar{t}$. The signal histograms are heavily shaded. The arrows indicate the cuts.

TABLE V. Missing transverse energy channel. Results for 1 fb^{-1}.

M_H (GeV/c^2)	90	100	110	120	130
Number of events					
ZH	6.66	4.37	3.53	2.76	2.16
WH	5.59	3.75	2.79	1.98	1.70
Total signal	12.25	8.12	6.32	4.74	3.86
$Zb\bar{b}$	8.12	4.97	4.83	3.85	3.92
Wbb	21.70	13.12	10.68	8.22	7.53
ZZ	11.24	6.14	2.59	1.05	0.59
WZ	7.95	4.49	1.99	0.90	0.54
tqb	0.63	0.27	0.37	0.24	0.29
tb	6.83	2.99	4.27	5.12	6.40
$t\bar{t}$	5.10	2.70	3.00	3.00	4.35
Total background	61.57	34.8	27.73	22.38	23.62
S/B	0.20	0.23	0.23	0.21	0.16
$S/\sqrt{B}$	1.56	1.38	1.20	1.00	0.79

(viii) $\Delta R(b_1,b_2)$–the distance, in the (η,ϕ)-plane, between the two b-tagged jets, where $\Delta R = \sqrt{\Delta\eta^2+\Delta\phi^2}$ and ϕ is the azimuthal angle

(ix) $\Delta R(b_1,l)$–the ΔR distance between the lepton and the first b-tagged jet.

Most of the variables used are directly measured (reconstructed) kinematic quantities while some are deduced variables. The choice of $M_{b\bar{b}}$ as a discriminating variable is obvious, as discussed earlier. The variable H_T is a measure of the ''temperature'' of the interaction; a large H_T is a sign of the decay of massive objects. For example, WH events would have larger H_T (increasing with M_H) than the Wbb background, but smaller H_T than the $t\bar{t}$ background. The WH events are also more spherical than the Wbb events and have larger values of sphericity. The $\Delta R(b,b)$ is smaller for $Wb\bar{b}$ background where the b-jets come mainly from $g\to b\bar{b}$ than in WH events where the b-jets come from the heavy object decay $H\to b\bar{b}$.

TABLE VI. Comparison of $S/\sqrt{B}$ achievable with conventional and neural networks cuts. Shown in the last column are the ratios of integrated luminosity required in the multivariate analysis to that required in the conventional analysis for a 5σ observation.

Channel	Mass (GeV)	Standard cuts	Neural net	L^{NN}/L^{std} (for 5σ obsv.)
$l+\not{E}_T+b\bar{b}$	100	0.98	1.44	0.46
	110	0.69	1.11	0.39
	120	0.58	0.87	0.44
	130	0.44	0.55	0.64
$\not{E}_T+b\bar{b}$	100	1.09	1.38	0.62
	110	0.85	1.20	0.50
	120	0.67	1.00	0.49
	130	0.54	0.78	0.47
$l^+l^-b\bar{b}$	100	0.48	0.63	0.58
	110	0.40	0.52	0.59
	120	0.40	0.48	0.69
	130	0.33	0.42	0.61

TABLE VII. Combined results of all three channels. We have simply added the signal counts and background counts from all three channels to get the total expected signal and background, respectively.

M_H (GeV/c^2)	90	100	110	120	130
$S/\sqrt{B}$ (1 fb^{-1})	2.4	2.1	1.7	1.4	1.0
$S/\sqrt{B}$ (2 fb^{-1})	3.3	3.0	2.4	2.0	1.5
$S/\sqrt{B}$ (30 fb^{-1})	12.9	11.5	9.4	7.7	5.7
Required luminosity					
5σ (Conventional)	7.5	10.5	18.3	26.6	42.2
5σ (NN)	4.5	5.7	8.5	12.6	22.7
3σ (NN)	1.6	2.1	3.0	4.5	8.2
95% C.L. (NN)	0.7	0.9	1.3	1.9	3.5

For each Higgs boson mass we trained three networks to discriminate against the main backgrounds $Wb\bar{b}$, WZ and $t\bar{t}$. The subsets of variables used to train the networks are listed in Table II while in Figs. 3(a–c) we show the distributions of some of these variables. Each network has 7 input variables, 9 hidden nodes and one output node. We calcuated three discriminants D for every signal and background event and for every Higgs boson mass. Figure 3(d) shows the distributions of the discriminants for signal and background calculated using the network trained to discriminate between signal events, with $M_H=100$ GeV/c^2, and the specified background. We note that all backgrounds, with the excep-

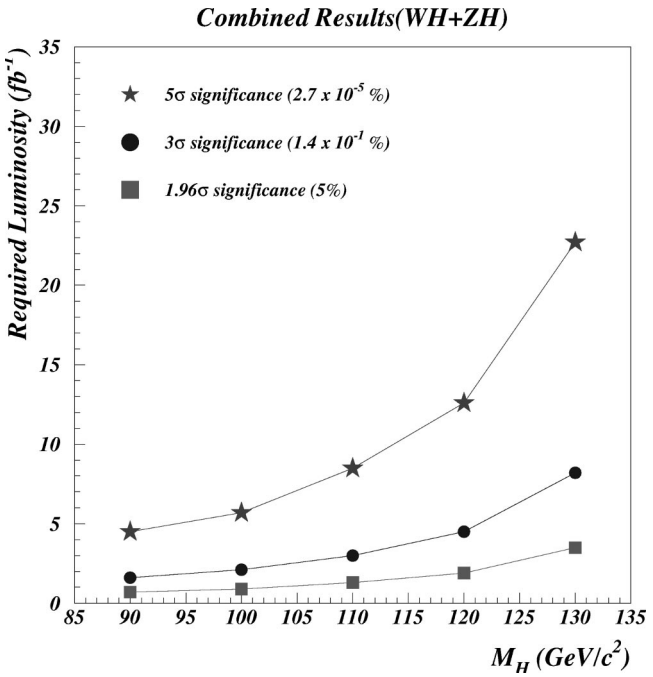

FIG. 7. Required integrated luminosity, with all channels combined, at 5σ, 3σ and 1.96σ (95% C.L.) significance, for NN analysis.

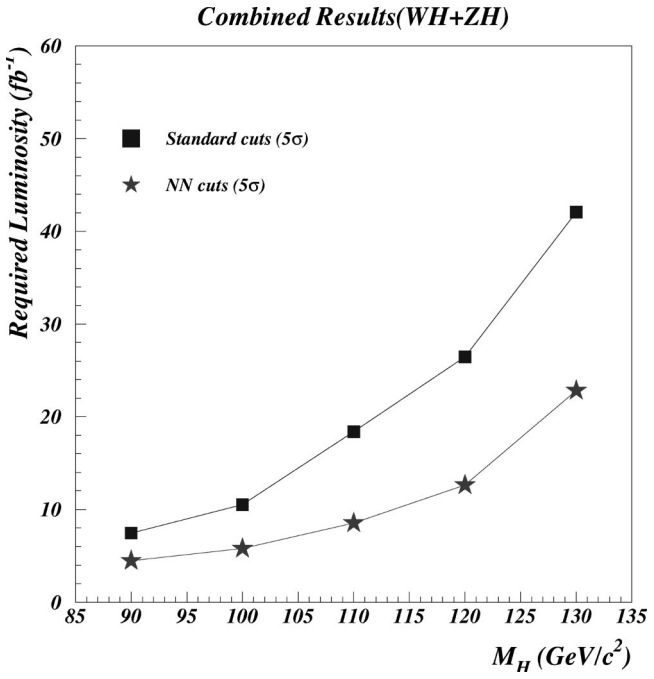

FIG. 8. Comparison of required integrated luminosity for a 5σ observation with all channels combined for NN and standard cuts. The luminosities given are for a *single* Tevatron experiment, as in the previous plots. For a given integrated luminosity the NN analysis provides a much higher discovery reach in mass.

tion of WZ, are well separated from the signal. For Higgs boson masses close to the Z mass the WZ background is kinematically identical to the signal and therefore difficult to deal with. But for Higgs boson masses well above the Z mass the discrimination between WH and WZ improves, as does that between WH and the other backgrounds. (In all figures, the signal histograms are shaded dark while the background histograms are shaded light.) The arrows in Fig. 3(d) indicate the cuts applied to the discriminants. The cuts were chosen to maximize $S/\sqrt{B}$, where S and B are the signal and background counts, respectively. The cuts to suppress the WZ background vary from 0.18 to 0.80, increasing for higher Higgs boson masses; the cuts to suppress $Wb\bar{b}$ are generally about 0.8, while those for top events are in the range 0.35 to 0.75.

At this stage it is instructive to compare the conventional and multivariate approaches, to assess what has been gained by using the latter approach. In Fig. 4 we compare the signal efficiency vs background efficiency (given in terms of the number of events for 1 fb^{-1}) for an ensemble of possible cuts on the three discriminants (using the random grid search technique [17]) with the efficiencies obtained using the standard cuts defined by the Run II Higgs Workshop [8]. Each dot corresponds to a particular set of cuts on the three discriminants; the triangular marker indicates what is achieved using the standard cuts, while the star indicates the results obtained from an optimal choice of cuts (which maximizes $S/\sqrt{B}$) on the three network outputs. Table III shows results for the WH channel.

IV. DI-LEPTON CHANNEL

For the di-lepton channel we followed a strategy similar to that described for the single lepton channel. The final state signature considered is: two high P_T same flavor leptons ($e\bar{e}$ or $\mu\bar{\mu}$) from Z boson decay and two b-jets (from $H\rightarrow b\bar{b}$).

The ZH events were generated using PYTHIA for Higgs boson masses of 90, 100, 110, 120 and 130 GeV/c^2. The principal backgrounds are due to ZZ, $Zb\bar{b}$, single top and $t\bar{t}$ production. The $Zb\bar{b}$ background sample was generated using COMPHEP, with fragmentation done using PYTHIA, while all other samples were generated using PYTHIA. As before, the SHW program was used to simulate the detector response and we assumed that two jets are b-tagged (one tight and one loose). The cross sections for signal and background are shown in Table I. The base sample was determined by the following cuts:

(i) $P_T^l > 10$ GeV/c
(ii) $|\eta_l| < 2$
(iii) $\not\!\!E_T < 10$ GeV
(iv) at least two jets with $E_T^{jet} > 8$ GeV and $|\eta_{jet}| < 2$.

A network was trained for each Higgs boson mass and for each of the three backgrounds with the following variables:

(i) E_T^{b1}, E_T^{b2}
(ii) P_T of the two leptons
(iii) $M_{b\bar{b}}$
(iv) $M_{l\bar{l}}$–invariant mass of the leptons
(v) H_T
(vi) $\Delta R(b_1, l)$ between the first lepton and the first b-tagged jet.

Distributions of these variables, as well as those of the network output, are shown in Figs. 5(a-d). The signal distributions are for $M_H = 100$ GeV/c^2. Our results after applying cuts on the three network outputs, for the di-lepton channels are summarized in Table IV.

V. MISSING TRANSVERSE ENERGY CHANNEL

This channel has contributions from both $ZH\rightarrow \nu\bar{\nu}b\bar{b}$ and $WH\rightarrow (l)\nu b\bar{b}$ where (l) denotes the lepton that is lost. The event generation and detector simulation were carried out as described in the single lepton and di-lepton channel studies. The base sample was defined by the cuts

(i) $|\eta_l| < 2$
(ii) $\not\!\!E_T > 10$ GeV/c
(iii) no isolated lepton with $P_T^l > 10$ GeV/c
(iv) $E_T^{jet3} < 30$ GeV
(v) at least two jets with $E_T^{jet} > 8$ GeV and $|\eta_{jet}| < 2$.

The three networks were trained with $ZH\rightarrow \nu\bar{\nu}b\bar{b}$ events as signal and $Zb\bar{b}$, ZZ and $t\bar{t}$ as the three backgrounds, respectively. The same networks were used to evaluate contributions from WH and the relevant backgrounds. We used the following variables to train the networks:

(i) E_T^{b1}, E_T^{b2}
(ii) $M_{b\bar{b}}$
(iii) H_T
(iv) $\not\!\!E_T$
(v) S

(vi) $\mathcal{C}$–centrality ($\Sigma_{jets}E_T/\Sigma_{jets}E$, with $E_T^{jet}>15$ GeV)
(vii) $\not\!\!E_T/\sqrt{E_T^{b1}}$
(viii) minimum $\Delta\phi(\text{jet},\not\!\!E_T)$.

The centrality, $\mathcal{C}$, has larger mean value (as is the case with S) for signal events than for backgrounds. The variable $\not\!\!E_T/\sqrt{E_T^{b1}}$ is a measure of the significance of the missing transverse energy. The smallest of azimuthal angles between $\not\!\!E_T$ and the jets in the event is expected to be smaller for $Wb\bar{b}$, $Zb\bar{b}$ as well as high multiplicity $t\bar{t}$ events than in signal events. We show the distributions of the variables and neural network outputs in Figs. 6(a–d). Again the signal distributions are for $M_H=100$ GeV/c^2. The results for this channel, after optimized cuts on network outputs, are listed in Table V.

VI. DISCUSSION AND SUMMARY

In Table VI we compare the results of our multivariate analysis with those based on the standard cuts, while Table VII and Figs. 7 and 8 show our final results, where we have combined all channels. The striking feature of these results is the substantial reduction in integrated luminosity required to make a 5σ discovery of the Higgs boson if one adopts a multivariate approach instead of the traditional method based on univariate cuts. In each of the three channels, the signal significance, which we define as $S/\sqrt{B}$, is seen to be 20–60% higher from our multivariate analysis as compared to an optimal conventional analysis. For example, at $M_H=110$ GeV/c^2 we find that the required integrated luminosity for a 5σ observation decreases from 18.3 fb^{-1} to 8.5 fb^{-1}. The results in Table VII include statistical errors only. The dominant systematic error will likely be due to background modeling. However, given the large data sets expected by the end of Run II we can anticipate that a thorough experimental study of the relevant backgrounds will have been undertaken. Therefore, it is possible that systematic errors could, eventually, be reduced to well under 10%. We can estimate the effect of systematic error by adding it in quadrature to the statistical error. If we assume a 10% systematic error on the total background the required integrated luminosity for a 5σ observation increases from 8.5 fb^{-1} to 12.8 fb^{-1}.

Run II at the Tevatron with the CDF and DØ detectors will begin in early 2001. Recently the scope of Run II has been expanded. The goal (hope) is to collect about 15-20 fb^{-1} per experiment in the period up to and including the start of the LHC. After 5 years of running, each experiment could see a 3σ-5σ signal of a neutral Higgs boson with $M_H\leqslant 130$ GeV/c^2. This exciting possibility for the Tevatron is the principal motivation for the recent important decision to expand the scope of Run II in order to accumulate as much data as possible. However, even with the expanded scope a discovery may be possible only if these data are analyzed with the most efficient methods available, such as the one we have described in this paper. It is important to note that the results we have presented are for a *single* experiment. That is, our conclusion is that each experiment has the potential of making an independent discovery. If the experiments combine their results the discovery of a low-mass Higgs boson at the Tevatron might be at hand a lot sooner.

ACKNOWLEDGMENTS

We thank the members of the Run II Higgs Working Group and, in particular, Ela Barberis, Alexander Belyaev, John Conway, John Hobbs, Rick Jesik, Maria Roco and Weiming Yao for useful discussions and for help with the event simulation. The research was supported in part by the U.S. Department of Energy under contract numbers DE-AC02-76CHO3000 and DE-FG02-97ER41022. This work was carried out by the authors as part of the Higgs Working Group[1] study at Fermilab. Fermilab is operated by Universities Research Association under contract to the U.S. Department of Energy.

[1]Run II Higgs Working Group (Run II SUSY/Higgs workshop). http://fnth37.fnal.gov/higgs.html.

[1] J. Erler and P. Langacker, in Proceedings of the 5th International WEIN Symposium: A Conference on Physics Beyond the Standard Model, Santa Fe, NM, 1998; e-print hep-ph/9809352.
[2] G. Altarelli, CERN report, CERN-TH/97-278; e-print hep-ph/9710434.
[3] P. C. Bhat, H. B. Prosper, and S. S. Snyder, Int. J. Mod. Phys. A **13**, 5113 (1998).
[4] LEP Electroweak Group, http://www.cern.ch/LEPEWWG/plots/summer99.
[5] See, for example, S. Martin, in *Perspectives on Supersymmetry*, edited by G. L. Kane (World Scientific, Singapore, 1998), e-print hep-ph/9709356 v3, 1999.
[6] M. Carena, M. Quiros, and C. E. M. Wagner, Nucl. Phys. **B461**, 407 (1996).
[7] B. C. Alanach *et al.*, Report of the Beyond the Standard Model Working Group of the 1999 UK Phenomenology Workshop on Collider Physics (Durham); e-print hep-ph/9912302.
[8] Run II Higgs Working Group of the Run II SUSY/Higgs workshop. http://fnth37.fnal.gov/higgs.html; T. Han, A. S. Turcot, and R. Zhang, Phys. Rev. D **59**, 093001 (1999).
[9] A. Sopczak, e-print hep-ph/0004015, IEKP-KA/2000-06. See also, http://l3www.cern.ch/conferences/talks99.html.
[10] M. Spira, e-print hep-ph/9810289; A. Djouadi, J. Kalinowski, and M. Spira, e-print hep-ph/9808312.
[11] DØ Collaboration, P. C. Bhat, in *Proceedings of the 10th Topical Workshop on Proton-Antiproton Collider Physics*, Batavia, IL (AIP, Woodbury, NY 1995), p. 308; C. M. Bishop, *Neural Networks for Pattern Recognition* (Clarendon, Oxford, 1998); R. Beale and T. Jackson, *Neural Computing: An Introduction* (Adam Hilger, New York, 1991).
[12] D. W. Ruck *et al.*, IEEE Trans. Neural Netw. **1**, 296 (1990); E.

A. Wan, *ibid.* **1**, 303 (1990); E. K. Blum and L. K. Li, Neural Networks **4**, 511 (1991).

[13] JETNET, C. Peterson, J. Rögnvaldsson, and L. Lönnblad, Comput. Phys. Commun. **81**, 185 (1994). We used JETNET version 3.0.

[14] PYTHIA, T. Sjöstrand, Comput. Phys. Commun. **82**, 74 (1994).

[15] COMPHEP, A. S. Belyaev, A. V. Gladyshev, and A. V. Semenov, e-print hep-ph/9712303; E. E. Boos *et al.*, e-print hep-ph/9503280.

[16] SHW 2.0, J. Conway, available at http://www.physics.rutgers.edu/~jconway/soft/shw/shw.html (unpublished).

[17] DØ Collaboration, H. B. Prosper *et al.*, in *Proceedings of the International Conference on Computing in High Energy Physics '95* (Rio de Janeiro, Brazil) (World Scientific, River Edge, NJ, 1996).