

ARTICLES

Using neural networks to enhance the Higgs boson signal at hadron colliders

R. D. Field, Y. Kanev, and M. Tayebnejad

Institute for Fundamental Theory, Department of Physics, University of Florida, Gainesville, Florida 32611

P. A. Griffin

Rockefeller University, New York, New York 10021

(Received 1 September 1995)

Neural networks are used to help distinguish the $ZZ \rightarrow l^+ l^-$ -jet-jet signal produced by the decay of a 400 GeV Higgs boson at a proton-proton collider energy of 15 TeV from the “ordinary” QCD Z +jets background. The ideal case where only one event at a time enters the detector (no pileup) and the case of multiple interactions per beam crossing (pileup) are examined. In both cases, when used in conjunction with the standard cuts, neural networks provide an additional signal-to-background enhancement.

PACS number(s): 13.87.Fh, 13.38.Be, 13.85.Qk, 14.80.Bn

I. INTRODUCTION

A neural network is an information processing system that is nonlinear, nonalgorithmic, and intensely parallel [1,2]. Neural networks have been studied as a model for the human brain and as a model for certain condensed matter systems. In this paper, we are interested in neural networks only as a *tool* for high energy collider phenomenology. The great challenge at hadron colliders is to disentangle any new physics that may be present from the “ordinary” QCD background. Hadron collider events can be very complicated and quite often one has the situation where the signal is hiding beneath the background. In addition, there are many variables that describe a high energy collider event and it is not always obvious which variables best isolate the signal or precisely what data selection (or cuts) optimally enhance the signal over the background. Here neural networks are an excellent tool since they are ideal for separating patterns into categories (e.g., signal and background). We will “train” a network to distinguish between signal and background using a large number of variables to describe each event. The network computes a single variable that ranges from zero to one. If the training is successful the network will output a number near one for a signal event and near zero for a background event and a single cut can be made on the network output which will enhance the signal over the background.

An important final state at hadron colliders consists of a large transverse momentum charged lepton pair plus two accompanying jets (i.e., $l^+ l^- jj$). It is one of the relevant signals for the production of a Higgs particle and its subsequent decay into ZZ with one Z decaying leptonically and the other Z decaying hadronically into a $q\bar{q}$ pair which then manifests itself as a pair of jets. The predominant background for this process is a single large transverse momentum Z boson plus the associated jets that mimic the Higgs boson signal. Requiring the Z boson to have a large transverse momentum by demanding a large P_T lepton pair forces the background to have a large P_T “away-side” quark or gluon via subpro-

cesses such as $qg \rightarrow Zq$ or $q\bar{q} \rightarrow Zg$. This away-side parton often fragments via gluon bremsstrahlung, producing away-side jet pairs which resemble the signal. In this paper, we use neural networks to help distinguish the $ZZ \rightarrow l^+ l^- jj$ decay of a 400 GeV Higgs boson signal from the Z +jets background in proton-proton collisions at 15 TeV. The neural network will be used in conjunction with the standard data cuts to provide additional signal to background enhancements. The discovery mode for a Higgs boson of this mass at a hadron collider is the “gold-plated” four lepton decay, $ZZ \rightarrow l^+ l^- l^+ l^-$. Here we investigate whether neural networks can help with the “jet physics” of the $l^+ l^- jj$ mode, particularly in the environment of multiple interactions per beam crossing (i.e., pileup). Also, progress made here can be carried over to the $WW \rightarrow l\nu jj$ decay mode of the Higgs boson.

This paper is not intended to be a detailed simulation of an experiment at the CERN Large Hadron Collider (LHC) [3,4]. Higgs boson production at a 15 TeV proton-proton collider is used as an illustration of neural networks as a tool in high energy jet phenomenology. We have designed, constructed, and tested the networks presented here from the beginning with the emphasis on high energy data analysis. We begin in Sec. II by discussing the construction and training of our neural networks. In Sec. III we discuss event generation, data selection and cuts for the ideal case where only one event at a time enters the detector (no pileup). Our network analysis without pileup is presented in Sec. IV, while in Sec. V we examine the case of multiple interactions per beam crossing (pileup). Section VI is reserved for summary and conclusions.

II. CONSTRUCTING AND TRAINING NEURAL NETWORKS

Characteristics of the network

Our neural networks are information processing systems with a set of N_{in} inputs, $\{x\}$, which can have any value and

one output, z_{net} , which is restricted to the range, $0 < z_{\text{net}} < 1$. The net output is a function of the input set $\{x\}$ and the network “memory” parameters as follows:

$$z_{\text{net}} = F_{\text{net}}(\{x\}, \{w\}, \{T\}),$$

where the network memory consists of a set of weights, $\{w\}$, and a set of thresholds $\{T\}$. The goal is to construct a network that can distinguish between two patterns of input data, “signal” events and “background” events, where each event is characterized by the N_{in} variables. A “perfect” network responds with z_{net} near one for a signal input and with z_{net} near zero for a background input.

The networks we will be using are far from perfect and the net outputs will vary from zero to one for both the signal and the background events. One way to characterize the performance of a network on a sample of N_{sig} signal events and N_{bak} background events is to define a network “error function” as follows:

$$\chi_{\text{net}}^2 = \frac{1}{N_{\text{sig}}} \sum_{n=1}^{N_{\text{sig}}} [z_{\text{net}}(n) - 1]^2 + \frac{1}{N_{\text{bak}}} \sum_{n=1}^{N_{\text{bak}}} [z_{\text{net}}(n) - 0]^2,$$

where $z_{\text{net}}(n)$ in the first and second summation is the network response for the n th signal and background event, respectively. This quadratic error function ranges from zero to one. It is equal to zero for a “perfect” network and is equal to 0.5 for a network that responds with $z_{\text{net}} = 0.5$ for both signal and background (i.e., a “dumb” network).

Constructing the network

The basic building block of the network is a processing element called a “neurode.” A single neurode has a set of N input variables, $\{y\}$ and one output, z . There is a weight, w_i , associated with each input. The neurode forms the weighted sum of the inputs and adds a threshold T as

$$Z = \sum_{i=1}^N w_i y_i + T.$$

The output of the neurode is arrived at by evaluating the function f at the point Z ,

$$z = f(Z),$$

where the “sigmoid” function f is given by

$$f(Z) = \frac{1}{1 + e^{-Z}}.$$

The output z ranges from zero to one since the sigmoid function lies between zero and one for any real Z .

Every neurode has N inputs and N weights, but just one output and one threshold. Network nodes consisting of N_1 inputs and N_2 outputs are formed by clustering together N_2 neurodes each with N_1 inputs. For example, Fig. 1 shows a network node consisting of two inputs and four outputs. Finally, a complete network is constructed by combining several network nodes with the outputs of one node becoming the inputs of the next node.

Figure 2 shows a schematic of an $N_{\text{in}} - N_1 - N_2 - 1$ net-

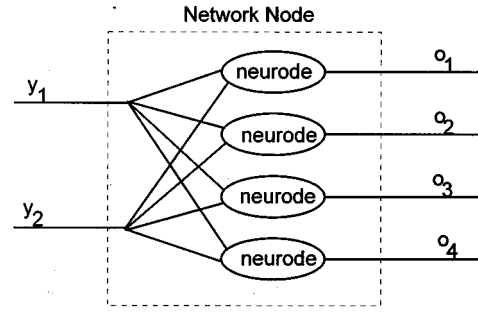


FIG. 1. Shows a neural network node with two inputs y_1 and y_2 and four outputs $o_1 \cdots o_4$. A node has the same number of processing elements called neurodes as outputs with the inputs being shared among all the neurodes.

work architecture. This network has N_{in} input variables, x_i , followed by a “hidden” layer with N_1 output variables, a_j , and another hidden layer with N_2 output variables, b_k . The output of the j th neurode of the first hidden layer has a value given by

$$a_j = f \left(\sum_{i=1}^{N_{\text{in}}} (w1)_{ij} x_i + (T1)_j \right),$$

where $(w1)_{ij}$ and $(T1)_j$ are the weights and threshold for the j th processing element in node 1. Similarly, the k th neurode of the second hidden layer has an output value given by

$$b_k = f \left(\sum_{j=1}^{N_1} (w2)_{jk} a_j + (T2)_k \right),$$

where $(w2)_{jk}$ and $(T2)_k$ are weights and threshold for the k th processing element in node 2. Finally, the output of the net is

$$z_{\text{net}} = f \left(\sum_{k=1}^{N_2} (w3)_k b_k + (T3) \right),$$

with node 3 consisting of just one processing element. In general, a network with two hidden layers and one output has

$$N_{\text{in}} N_1 + N_1 N_2 + 2 N_2 + N_1 + 1,$$

network memory parameters.

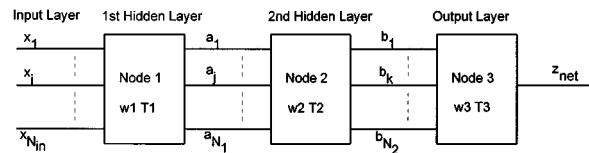


FIG. 2. Shows a schematic illustration of a network architecture with three nodes and two “hidden layers.” The network has N_{in} inputs followed by a node with N_1 output variables, a_j , and another node with N_2 output variables, b_k , and one output, z_{net} . There are $N_{\text{in}} \times N_1$ weights $W1$ and N_1 thresholds $T1$ associated with node 1; and $N_1 \times N_2$ weights $W2$ and N_2 thresholds $T2$ associated with node 2; and N_2 weights $W3$ and one thresholds $T3$ associated with node 3.

The most difficult part of using neural networks for data analysis is in finding networks that yield z_{net} near one for the signal and z_{net} near zero for the background or equivalently finding networks with a small χ_{net}^2 . To find networks that can distinguish between signal and background we minimize χ_{net}^2 with respect to the network memory parameters (i.e., the weights and thresholds). The process of minimizing χ_{net}^2 over a set of signal and background events is referred to as “training” the network. A well trained (i.e., smart) network has a small χ_{net}^2 .

Minimizing χ_{net}^2 in the large multidimensional space of the weights and thresholds is quite challenging and we will not present all the details here. One way is to simply generate network memories at random and keep the one with the best χ_{net}^2 . Another method which is referred to as “backward error propagation” [1,2] involves calculating the derivatives of χ_{net}^2 with respect to every weight and threshold and forming the gradient in $w-T$ space. One can lower χ_{net}^2 by moving in $w-T$ space in the direction of $-\nabla\chi_{\text{net}}^2$. Here one has to be careful not to get caught in a local minimum. We use a combination of the random method and the backward error propagation method to train our networks.

III. EVENT GENERATION, DATA SELECTION, AND CUTS—WITHOUT PILEUP

We consider first the ideal case where only one event at a time enters the detector. We want to determine whether neural networks can be trained to distinguish between the Higgs boson signal and the Z +jets background when there is no pileup. ISAJET version 7.06 is used to generate Higgs bosons with a mass of 400 GeV in 15 TeV proton-proton collisions. The generated width of the Higgs boson is about 30 GeV. The Higgs boson is forced to decay into two Z bosons with one Z decaying leptonically and the other Z decaying into a quark-antiquark pair. We refer to this as the “signal.” The “background” consists of single Z boson events generated with the hard-scattering transverse momentum of the Z , \hat{k}_T , greater than 100 GeV. Single Z bosons are produced at large transverse momentum via the “ordinary” QCD subprocesses $qg \rightarrow Zq$, $\bar{q}g \rightarrow Z\bar{q}$, and $q\bar{q} \rightarrow Zg$. These subprocesses, of course, generate additional gluons via bremsstrahlung off both incident and outgoing color nonsinglet partons, resulting in multiparton final states which subsequently fragment into hadrons, and is referred to as the Z +jets background.

We are not attempting to do a detailed simulation of an LHC detector [2,3]. Events are analyzed by dividing the solid angle into “calorimeter” cells having size $\Delta\eta\Delta\phi=0.2 \times 15^\circ$, where η and ϕ are the pseudorapidity and azimuthal angle, respectively. A single cell has an energy (the sum of the energies of all the particles that hit the cell *excluding* neutrinos) and a direction given by the coordinates of the center of the cell. From this the transverse energy of each cell is computed from the cell energy and direction. Large transverse momentum leptons are analyzed separately and are not included when computing the energy of a cell. Jets are defined using a simple algorithm. One first considers the “hot” cells (those with transverse energy greater than 5 GeV). Cells are combined to form a jet if they lie within a specified “distance” or “radius,” $R^2=\Delta\eta^2+\Delta\phi^2$, in $\eta-\phi$ space from each other. Jets have an energy given by the sum

of the energy of each cell in the cluster and a momentum \vec{p}_j given by the vector sum of the momentums of each cell. The invariant mass of a jet is simply $M_j^2=E_j^2-\vec{p}_j \cdot \vec{p}_j$.

We have taken the energy resolution to be perfect, which means that the only resolution effects are caused by the lack of spatial resolution due to the cell size. However, we are using a very crude calorimeter with large cells (960 cells with $|\eta|<4$). Experiments at, for example, the LHC [3,4] will have considerably smaller cell size and hence better spatial resolution. Even with the addition of energy resolution effects, the combined spatial and energy resolution at the LHC should be comparable to or better than in our analysis.

Lepton trigger

Our “zero-level” trigger is designed to select large transverse momentum Z bosons that have decayed into charged leptons. The first cut is made by demanding that the event contain *at least* two high transverse momentum leptons ($l^\pm=e^\pm$ or μ^\pm) in the central region as follows:

$$P_T(l^\pm)>25 \text{ GeV}, \quad |\eta(l^\pm)|<2.5.$$

Lepton pairs (e^+e^- and $\mu^+\mu^-$) are constructed for the events that survive this first cut. The pairs are ordered according to their invariant mass, with pair No. 1 having the mass closest to the Z boson and pair No. 2 being the second closest, etc. Finally, the event is rejected unless at least one lepton pair satisfies the following

$$P_T(l^+l^-)>100 \text{ GeV}.$$

Table I shows that for a 400 GeV Higgs boson at 15 TeV, roughly 10 000 events per year pass this “zero level” trigger. Here the integrated luminosity for one year is taken to be the expected LHC value of $10^5/\text{pb}$ [3,4]. About 2 million background events per year survive this “zero level” lepton cut.

This high transverse lepton pair cut is, of course, crucial. The transverse momentum spectrum of the single Z QCD background falls off rapidly, while for the heavy Higgs boson the signal is peaked at about half the mass of the Higgs boson. Here one wants to take as large of a cut on $P_T(l^+l^-)$ as possible without losing too much of the signal. However, even with this cut, the background is still more than 200 times the signal.

Jet-pair selection

The jet topology of events with at least one large transverse momentum lepton pair is analyzed by first examining only jet cores [i.e., narrow jets of size $R_j(\text{core})$]. Here one includes only those jet cores satisfying

$$E_T(\text{jet core})>25 \text{ GeV}, \quad |\eta(\text{jet core})|<3,$$

with

$$R_j(\text{core})=0.2.$$

With our large calorimeter cell size, a jet core consists of one, two, or three cells (distances are measured from the center of the cell).

In an attempt to find the two jets produced by the hadronic decay of the large transverse momentum Z boson, jet

TABLE I. 400 GeV Higgs bosons produced in 15 TeV proton-proton collisions. The table shows the number of events per year (with $\mathcal{L}=10^5/\text{pb}$) for the Higgs boson $\rightarrow ZZ$ signal and the Z +jets background for the ideal case where only one event at a time enters the detector (no pileup). The “zero-level” lepton trigger is used as a reference point and is normalized to 100%. The enhancement factor is defined to be the percentage of signal divided by the percentage of background surviving the given set of cuts. Both the overall and relative enhancement factors are shown.

Selection or cut	$H \rightarrow ZZ$ signal			Z +jets background			Enhancement		
	% Relative	% Overall	Events/ year	% Relative	% Overall	Events/ year	Bak/ sig	Relative	Overall
Lepton trigger:									
$PT(l) > 25$ GeV $PT(ll) > 100$ GeV	100%	100%	10 185	100%	100%	1 961 818	193	1.0	1.0
Jet pair selection ($R_{jj} < 1.6$):									
$ET(j) > 25$ GeV $PT(jj) > 100$ GeV	49.0%	49.0%	4 995	30.4%	30.4%	595 622	119	1.6	1.6
Z-mass cut:									
$81 < MZ < 101$ GeV	51.1%	25.0%	2 551	7.4%	2.3%	44 244	17	6.9	11.1
Higgs boson mass cut:									
$350 < MH < 450$ GeV	87.8%	22.0%	2 241	32.7%	0.7%	14 471	6.5	2.7	29.8
Z-mass and net cut:									
$81 < MZ < 101$ GeV $z_{\text{net}} > 0.75$	41.6%	10.4%	1 060	8.3%	0.2%	3 683	3.5	5.0	55.4
Higgs boson mass and net cut:									
$350 < MH < 450$ GeV $z_{\text{net}} > 0.75$	42.6%	9.4%	954	12.9%	0.1%	1 862	2.0	3.3	98.7

pairs are formed by demanding that the distance between the two jet cores in η - ϕ space, $d_{jj}^2 = (\eta_1 - \eta_2)^2 + (\phi_1 - \phi_2)^2$, be less than 1.6. Namely,

$$d_{jj}(\text{jet-jet cores}) < 1.6.$$

In addition, the jet-jet cores are required to satisfy

$$P_T^{jj} > 100 \text{ GeV}, \quad |\phi_{jj} - \phi_{ll}| > 90^\circ,$$

where P_T^{jj} is the total transverse momentum of the core jet pair and $\phi_{jj} - \phi_{ll}$ is the azimuthal angle between the leading lepton pair and the core jet pair. The jet pair is required to be in the opposite hemisphere (or “away-side”) from the lepton pair. If more than one jet pair meets all of these requirements than the pair with the largest total transverse energy is selected.

Table I shows that of the 10 000 signal events passing the “zero level” lepton trigger about 49% also pass the jet-pair selection criterion. Unfortunately, about 30% of the ordinary Z +jets background events that survive the “zero level” lepton trigger also have a jet-pair meeting the selection criteria.

Here it is useful to define two quantities that measure the effectiveness of a particular cut. The “enhancement factor” is defined as the percentage of signal divided by the percentage of background that survives the cut. Namely,

$$F_{\text{enh}} = \frac{\% \text{ of signal surviving cut}}{\% \text{ of background surviving cut}}.$$

The efficiency of a cut is defined as the percentage of signal that survives the cut:

$$F_{\text{eff}} = \% \text{ of signal surviving cut.}$$

The jet-pair selection criterion results in an enhancement of 1.6 with an efficiency of about 49%. The “zero level” lepton trigger is used as a reference point and is normalized to an efficiency of 100% and an enhancement of one. One might have expected to do better at this stage. However, once we require that the Z boson have a large transverse momentum, we force the background to have a large P_T away-side quark or gluon jet. This away-side parton often fragments via gluon bremsstrahlung into multiple away-side jets which then survive the selection criteria.

Invariant mass cuts

The invariant mass, $M_{jj}(\text{full})$, is constructed by using all cells that lie within a “distance” $R_{jj}(\text{full})$ in η - ϕ space of *either* of the two jets. Cells are *not* double counted. For example, a cell may lie within $R_{jj}(\text{full})$ of both jets, nevertheless it is counted just once. The aim here is, of course, to reconstruct the invariant mass of the Z boson as shown in Fig. 3. However, this full jet-jet invariant mass will *only* be used in the event selection. The Higgs boson mass will be reconstructed by setting $M_{jj} = M_Z$. At this stage, events are rejected unless the full jet-jet mass satisfies

$$81 < M_{jj}(\text{full}) < 101 \text{ GeV},$$

with

$$R_{jj}(\text{full}) = 0.6.$$

As can be seen in Fig. 3 and Table I, about 51% of the Higgs signal passing *both* the lepton cut and the jet-pair selection have M_{jj} within 10 GeV of the Z boson mass. On the other hand, only about 7% of the Z +jets background events surviving *both* the lepton cut and the jet-pair selection have a full jet-pair invariant mass within 10 GeV of the Z boson

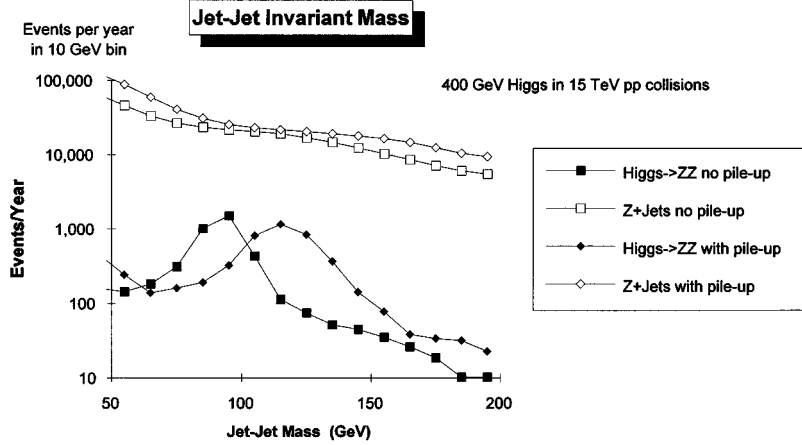


FIG. 3. Shows the away-side jet-jet mass for a 400 GeV Higgs boson produced in 15 TeV proton-proton collisions. The plot corresponds to the number of events per year (with $\mathcal{L}=10^5/\text{pb}$) in a 10 GeV bin for the of the Higgs boson $\rightarrow ZZ$ signal and the Z+jets background. The ideal case where only one event at a time enters the detector (no pileup) and the case of multiple interactions per beam crossing (with pileup) are shown. In all cases the events have survived the “zero-level” lepton trigger and the jet-pair selection criterion.

mass. This corresponds to an overall enhancement factor at this stage of about 11 with an overall efficiency of about 25%. The background lies well above the signal in Fig. 3 so that one cannot directly see the Z mass peak. Nevertheless, the jet-jet invariant mass cut is very important.

Reconstructing the Higgs boson mass

The Higgs boson invariant mass is constructed from the momentum vectors of the two charged leptons and the momentum vector of the jet pair as follows:

$$M^2 = (E_{l^+} + E_{l^-} + E_{jj})^2 - (\vec{p}_{l^+} + \vec{p}_{l^-} + \vec{p}_{jj})^2,$$

where

$$E_{jj}^2 = \vec{p}_{jj} \cdot \vec{p}_{jj} + M_z^2.$$

The mass of a jet is not a well defined quantity since it depends on the soft particles. The momentum vector of a jet is better defined and is determined primarily by the core cells. Thus, in constructing the Higgs boson mass we use the momentum vector of the jet-pair but *not* the jet-pair mass. The mass of the jet-pair is set equal to the mass of the Z boson.

Figure 4 shows the reconstructed Higgs boson mass for both the signal and background events that have passed the lepton cuts, the jet-pair selection, and have $81 < M_{jj}(\text{full}) < 101$ GeV. At this stage, there are about 2000

Higgs boson events and 14 000 QCD background events per year within 50 GeV of the true Higgs boson mass of 400 GeV. This corresponds to an overall enhancement factor of about 30 (see Table I) with an overall efficiency of about 22%. However, even with this enhancement the Z+jets background is still more than six times the signal. It is at this stage that neural networks will be used to provide an additional enhancement of signal over background.

IV. NETWORK ANALYSIS WITHOUT PILEUP

We will train a neural network to distinguish between the signal and background events that have already passed the lepton cuts, the jet-pair selection, and have $81 < M_{jj}(\text{full}) < 101$ GeV. These important cuts are made before sending the events to the network. Even though both the signal and background events have survived these cuts, there is still additional information in the events that is not the same for the signal and the background. The network can use these differences to further help distinguish signal from background.

Network inputs and training

Of course, the key to a good network lies in the selection of the input variables. These variables must characterize the differences between the signal and the background. In this analysis we choose the following nine input variables:

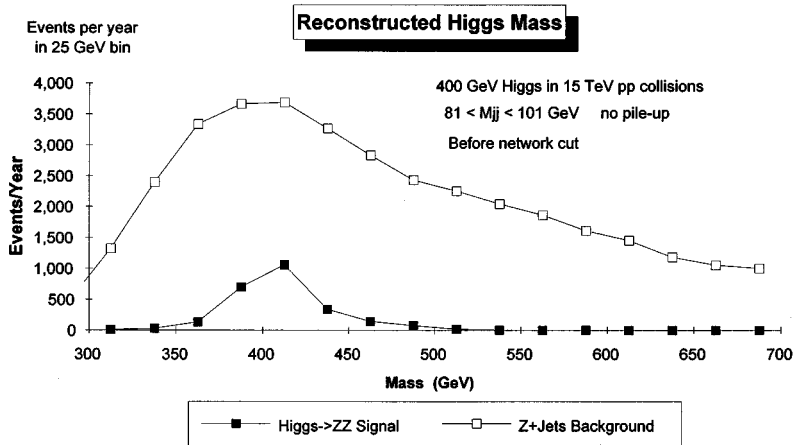


FIG. 4. Shows the reconstructed mass of a 400 GeV Higgs boson produced in 15 TeV proton-proton collisions. The plot corresponds to the number of events per year (with $\mathcal{L}=10^5/\text{pb}$) in a 25 GeV bin for the Higgs boson $\rightarrow ZZ$ signal and the Z+jets background for the ideal case where only one event at a time enters the detector (no pileup). The events have survived the “zero-level” lepton trigger and the jet-pair selection criterion with $81 < M_{jj}(\text{full}) < 101$ GeV. No network cut has been made.

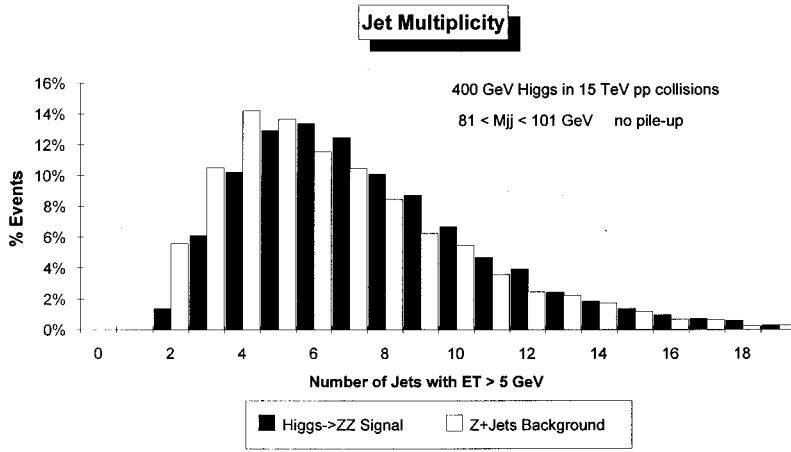


FIG. 5. Shows the multiplicity of jets for 400 GeV Higgs bosons produced in 15 TeV proton-proton collisions. The plot corresponds to the percentage of events with N jets with E_T greater than 5 GeV for the Higgs boson $\rightarrow ZZ$ signal and the Z +jets background for the ideal case where only one event at a time enters the detector (no pileup). The events have survived the “zero-level” lepton trigger and the jet-pair selection criterion with $81 < M_{jj}(\text{full}) < 101$ GeV.

$$x_1 = d_{jj},$$

$$x_2 = |E_T^j(1) - E_T^j(2)| / [E_T^j(1) + E_T^j(2)],$$

$$x_3 = N_{\text{jet}}(E_T > 5 \text{ GeV}),$$

$$x_4 = E_T(R_{jj} < 0.2) / E_T(R_{jj} < 1.0),$$

$$x_5 = E_T(0.2 < R_{jj} < 0.6) / E_T(R_{jj} < 1.0),$$

$$x_6 = E_T(0.6 < R_{jj} < 1.0) / E_T(R_{jj} < 1.0),$$

$$x_7 = M(R_{jj} < 0.2) / M(R_{jj} < 1.0),$$

$$x_8 = M(0.2 < R_{jj} < 0.6) / M(R_{jj} < 1.0),$$

$$x_9 = M(0.6 < R_{jj} < 1.0) / M(R_{jj} < 1.0).$$

The first variable is simply the distance in η - ϕ space between the two “away-side” jets selected in the jet-pair selection. For the signal this is related to the opening angle of the quark-antiquark pair resulting from the $Z \rightarrow q\bar{q}$ decay, while for the background this is the distance between, for example, an outgoing quark and the radiated gluon jet. The second variable is the “skewness” of the transverse energies of the two jets cores, while the third variable is simply the overall number of jets (with $E_T > 5$ GeV) in the event and is shown in Fig. 5.

The remaining variables depict the precise manner in which transverse energy and mass are distributed around the away-side jet pair. For example, x_6 is the ratio of the amount of transverse energy coming from calorimeter cells within the “halo” region $0.6 < R_{jj} < 1.0$ surrounding both jets to the total transverse energy of the *extended* jet pair [$R_{jj}(\text{extended}) = 1.0$]. As can be seen in Fig. 6, the fraction of transverse energy in this region is, on the average, slightly larger for the background than for the signal. Similarly, x_9 is fraction of the full jet-jet invariant mass that comes from calorimeter cells in the “halo” region $0.6 < R_{jj} < 1.0$. Figure 7 shows that more of the extended jet-jet mass lies in this region for the background than for the signal. The other halo regions also show slight variations between signal and background which the network can use to help distinguish between the two.

The idea here is similar to the jet-jet profile analyses we presented in Ref. [5]. For the signal, the away-side jet pair arises from the $q\bar{q}$ decay of a large transverse momentum Z boson. The Z boson is a color singlet and does not radiate gluons during flight. On the other hand, the large P_T away-side recoil quarks or gluons in the single Z background are not color singlets and produce additional gluons via bremsstrahlung. These radiated gluons deposit transverse energy around the jet-jet cores. This results in more transverse energy and invariant mass surrounding the jet-jet cores for the Z +jets background than for the Higgs boson signal. The

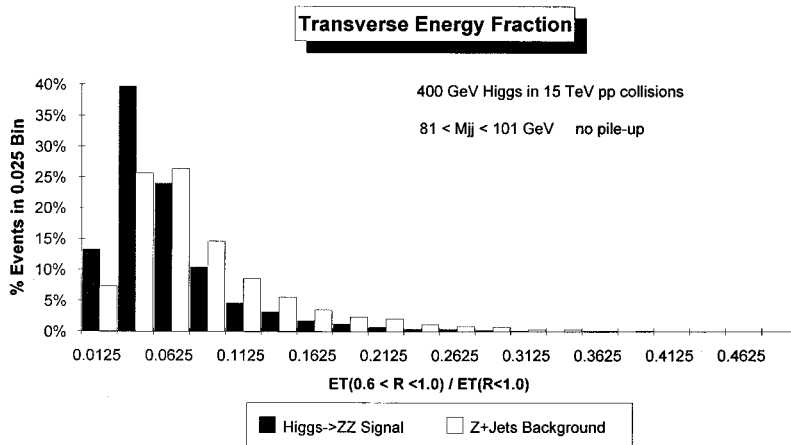


FIG. 6. Shows the fraction of transverse energy coming from calorimeter cells within the “halo” region $0.6 < R_{jj} < 1.0$ surrounding either of the away-side jets. The plot corresponds to the percentage of events with the jet-jet transverse energy fraction within the 0.025 bin for the Higgs boson $\rightarrow ZZ$ signal and the Z +jets background for the ideal case where only one event at a time enters the detector (no pileup). The events have survived the “zero-level” lepton trigger and the jet-pair selection criterion and have $81 < M_{jj}(\text{full}) < 101$ GeV.

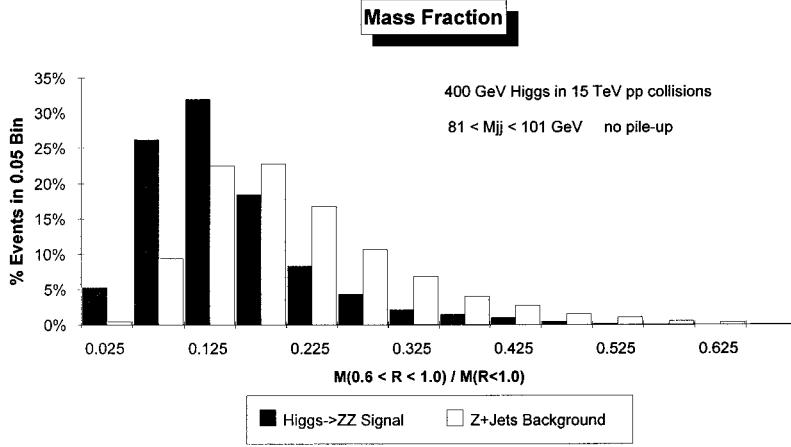


FIG. 7. Shows the fraction of invariant mass coming from calorimeter cells within the “halo” region $0.6 < R_{jj} < 1.0$ surrounding either of the away-side jets. The plot corresponds to the percentage of events with the jet-jet invariant mass fraction within the 0.05 bin for the Higgs boson $\rightarrow ZZ$ signal and the Z +jets background for the ideal case where only one event at a time enters the detector (no pileup). The events have survived the “zero-level” lepton trigger and the jet-pair selection criterion and have $81 < M_{jj}(\text{full}) < 101$ GeV.

distribution of transverse energy and invariant mass around the “away-side” jet pair is slightly different in the two cases.

The network is trained on a sample of 8 348 signal and 7 254 background events using the nine inputs shown above and where both signal and background events have already satisfied the lepton cuts, the jet-pair selection, and have $81 < M_{jj}(\text{full}) < 101$ GeV. To get this training sample, it was necessary to generate 80 000 Higgs boson events and 800 000 Z +jet events. There is no systematic procedure that provides the best network topology for a given problem. One looks for the simplest network that can discriminate signal from background. We experimented with a variety of network sizes and types and present here the results from a 9-16-8-1 net which has 305 memory parameters. After a lengthy training process we achieved $\chi_{\text{net}}^2 = 0.1678$ on the training sample.

Network performance

Figure 8 shows the network response (i.e., z_{net}) for the sample of signal and background events used in the training. The situation is far from the ideal. There are some events around $z_{\text{net}} = 0.5$ for which the net cannot distinguish between signal and background. Nevertheless, the net does allow for some separation of signal and background. The net clearly recognizes some events as signal or background, while for other events there is an overlap and the net cannot distinguish between the two. Ideally one would like a clean separation between the signal and background in Fig. 8. One would then perform a network cutoff and assign any event with $z_{\text{net}} > z_{\text{cut}}$ to be signal and events with $z_{\text{net}} < z_{\text{cut}}$ to be background.

Figure 8 also shows the network response (i.e., z_{net}) for an independent sample of signal and background events *not* used in the training. If the network generalized perfectly there would be no difference between the response of the network for the independent and the training samples. The small differences seen in Fig. 8 reflect that fact that we have trained the net on a small relatively sample of events. We could improve the ability of the network to generalize by starting with a larger training sample, but this result is sufficient for what we want to illustrate in this paper.

The enhancement and efficiency of the network cutoff depends on the value chosen for z_{cut} , where the network enhancement and efficiency are defined as

$$F_{\text{enh}}^{\text{net}} = \frac{\% \text{ of signal with } z_{\text{net}} > z_{\text{cut}}}{\% \text{ of background with } z_{\text{net}} > z_{\text{cut}}},$$

$$F_{\text{eff}}^{\text{net}} = \% \text{ of signal with } z_{\text{net}} > z_{\text{cut}}.$$

The overall network performance can be characterized by the single curve of the network enhancement versus the network efficiency shown in Fig. 9. Each point in Fig. 9 corresponds to a different choice for the network cutoff with the lower efficiencies and higher enhancements corresponding to larger values of z_{cut} . In the analysis presented here, we choose $z_{\text{cut}} = 0.75$ which for the training sample corresponds to a relative efficiency of about 42% with a relative enhancement of about 6.

Fisher discriminates

Another method of separating signal and background is to use Fisher discriminates [6]. This method is analogous to a neural network with no hidden layers. Here as with the network, one inputs a set of N_{in} variables, x_i , and there is one output, F . However, in this case F is a linear function of the inputs,

$$F = \sum_{i=1}^{N_{\text{in}}} \alpha_i x_i,$$

where the Fisher coefficients, α_i are chosen to maximize the separation between signal and background in F space,

$$\frac{(\mu_F^{\text{sig}} - \mu_F^{\text{bak}})^2}{(\sigma_F^{\text{sig}})^2 + (\sigma_F^{\text{bak}})^2},$$

where μ_F and σ_F are the mean and the standard deviation, respectively, of the Fisher output for the signal (sig) and background (bak) sample. The Fisher coefficients are given by

$$\alpha_i = \sum_j (V^{\text{sig}} + V^{\text{bak}})^{-1}_{ij} (\mu_j^{\text{sig}} - \mu_j^{\text{bak}}),$$

where $(V^{\text{sig}} + V^{\text{bak}})^{-1}$ is the inverse matrix and μ_i is the mean of the distribution x_i ,

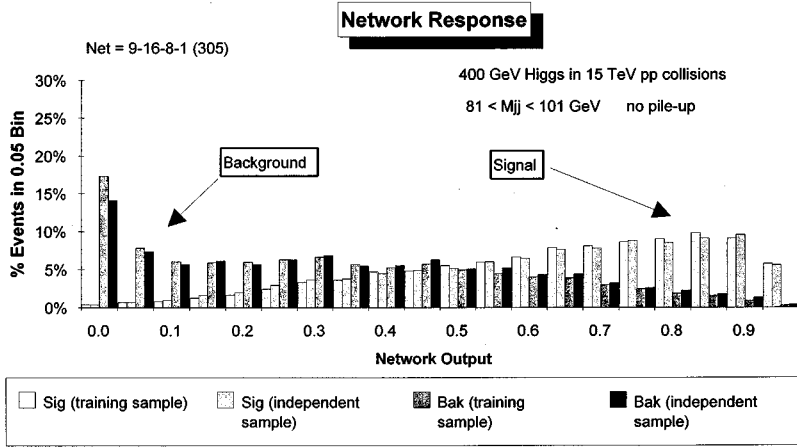


FIG. 8. Shows the network response, z_{net} , for the sample of signal and background events used in the training and for an independent sample of signal and background events. The plot corresponds to the percentage of events with z_{net} within a 0.05 bin for the Higgs boson \rightarrow ZZ signal and the Z+jets background for the ideal case where only one event at a time enters the detector (no pileup). The events have survived the “zero-level” lepton trigger and the jet-pair selection criterion and have $81 < M_{jj}(\text{full}) < 101$ GeV.

$$\mu_i = \frac{1}{N} \sum_{n=1}^N x_i(n),$$

and V is the covariance matrix,

$$V_{ij} = \frac{1}{N} \sum_{n=1}^N [x_i(n) - \mu_i][x_j(n) - \mu_j].$$

Here the sum over n corresponds to the sum over the training sample.

In this case training consists of calculating the Fisher coefficients which involves inverting an $N_{\text{in}} \times N_{\text{in}}$ matrix, but is easier than training a network. Once this is done the situation is similar to the network. For each input of N_{in} variables there is one output F . We have determined the Fisher coefficients for the sample of signal and background events used to train our network and the Fisher response for these events is shown in Fig. 10. The separation between signal and background is not as good as with the network. As with the network, the overall Fisher performance can be characterized by the single curve of the Fisher enhancement versus the Fisher efficiency which is shown in Fig. 9 together with the network performance. Each point corresponds to a different choice for the Fisher cutoff.

Using the network cutoff

We now analyze an *independent* sample of events using the trained network as a tool to help distinguish between

signal and background. Figure 11 shows the reconstructed Higgs boson mass for both the signal and background events that have passed the “zero level” lepton trigger, the jet-pair selection with $81 < M_{jj}(\text{full}) < 101$ GeV, and the network cutoff (with $z_{\text{cut}} = 0.75$). Now, there are about 1 000 Higgs boson events and 2 000 QCD background events per year within 50 GeV of the true Higgs boson mass of 400 GeV. This corresponds to an overall enhancement factor of about 100 (see Table I) with an overall efficiency of about 10%. Figure 11 shows that the signal and background are now comparable. Comparing the reconstructed Higgs boson mass in Fig. 4 with Fig. 11 shows the added enhancement the neural network provides.

Using network weighting

An alternative approach to using the network cutoff is to use network weighting. Here one weights the event with the network response, z_{net} , which lies between zero and one. If the network has been able to separate signal from background then signal events will be assigned a weight near one and background events will be assigned a weight near zero.

Figure 12 shows the *network weighted* reconstructed Higgs boson mass for both the signal and background events that have passed the lepton cuts, the jet-pair selection with $81 < M_{jj}(\text{full}) < 101$ GeV. The advantage here is that all the signal events are used (i.e., the relative efficiency is 100%), but in this case the network cutoff procedure provides a better enhancement of the signal.

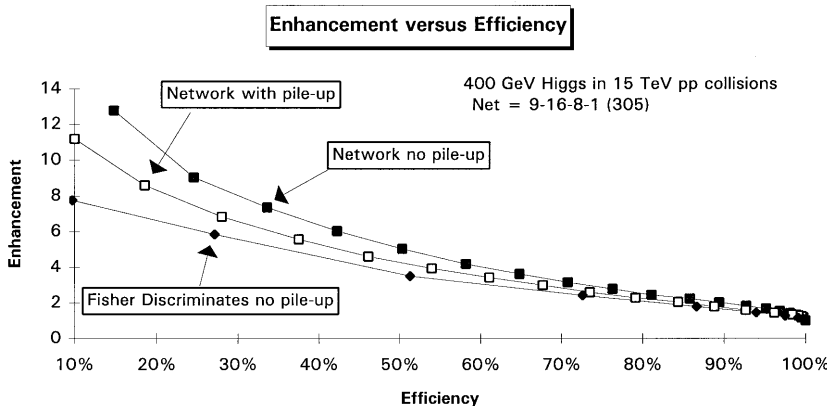


FIG. 9. Shows the enhancement versus the efficiency for the training sample of events for the 9-16-8-1 neural network with 305 memory parameters. Both the ideal case where only one event at a time enters the detector (no pileup) and for the case of multiple interactions per beam crossing (pileup) are shown. Each point in the plot corresponds to a different choice for the network cutoff with the lower efficiencies and higher enhancements corresponding to larger values of z_{cut} . The network enhancements are compared with the enhancements arrived at by the use of Fisher discriminates (no pileup).

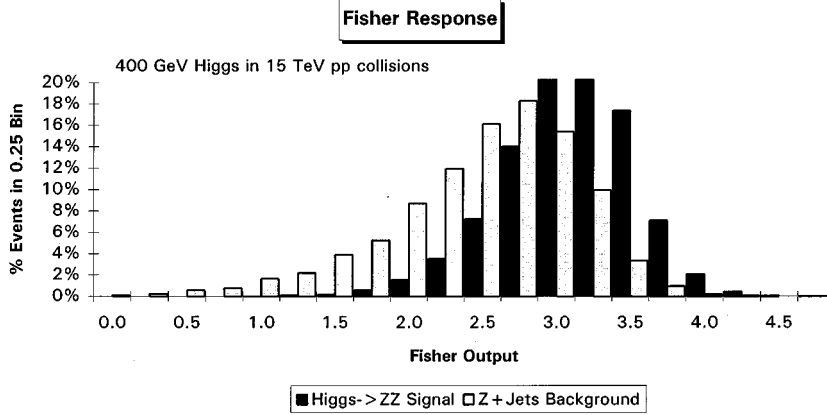


FIG. 10. Shows the Fisher response, F , for the sample of signal and background events used in the training of the neural network. The plot corresponds to the percentage of events with F within a 0.3 bin for the Higgs boson $\rightarrow ZZ$ signal and the Z +jets background for the ideal case where only one event at a time enters the detector (no pileup). The events have survived the “zero-level” lepton trigger and the jet-pair selection criterion and have $81 < M_{jj}(\text{full}) < 101$ GeV.

V. NETWORK ANALYSIS WITH PILEUP

Event generation and cuts

We now consider the case of multiple interactions per beam crossing. ISAJET is used to generate N_{pile} minimum bias events along with each Higgs boson $\rightarrow ZZ$ signal and each Z +jets background event. The number of pileup interactions per beam crossing, N_{pile} , that enter the calorimeter is generated according to a Poisson distribution with a mean of about 29 minimum bias collisions for each Higgs boson or Z +jets event as shown in Fig. 13. The mean of 29 collisions per beam crossing was arrived at by using a bunch crossing time of 25 ns, a peak luminosity of 10^{34} $\text{cm}^{-2} \text{sec}^{-1}$, and the ISAJET minimum bias cross section at 15 TeV of 116 mb. Our mean number is larger than the 20 collisions per beam crossing quoted for the LHC [3,4]. However, the ISAJET “minimum bias” events are softer than what is expected. To compensate for this we have increased that average number of pileup interactions per beam crossing to 29.

These pileup interactions greatly increase the particle multiplicity and the global transverse energy of each event. Nevertheless, they do not affect the lepton trigger. Table II shows that, as before, roughly 10 000 Higgs boson and about 2 million background events per year pass the “zero level” lepton trigger.

Events are again analyzed by dividing the solid angle into “calorimeter” cells having size $\Delta\eta\Delta\phi=0.2\times 15^\circ$, but in this case we ignore all cells with $E_T < 1$ GeV. This is done to reduce the number of nonzero cells which saves time and

improves the jet algorithm. Jets are defined as before, but the definition of a “hot” cells is changed to 10 GeV. This means that the minimum jet transverse energy is now 10 GeV (compared to 5 GeV in the analysis without pileup).

Except for these changes, the jet-pair selection is done as before with similar results. Table II shows that of the 10 000 signal events passing the “zero level” lepton cut about 50% also pass the jet-pair selection criterion. Also, about 30% of the ordinary Z +jets background events that survive the “zero level” lepton trigger have a jet pair that meets the selection criterion.

The jet-jet invariant mass for the signal and background events that have passed the “zero-level” lepton trigger and the jet-pair selection criterion is shown in Fig. 3. Comparison with the no pileup case shows that the Z mass peak has shifted up about 20 GeV and become somewhat broader. This is, of course, due to the pileup interactions which have contributed transverse energy and mass to the jet pair. Rather than trying to subtract out this effect, we simply shift our jet-jet mass cut to

$$100 < M_{jj}(\text{full}) < 120 \text{ GeV},$$

where $M_{jj}(\text{full})$ is defined as before with $R_{jj}(\text{full})=0.6$. As before, the invariant mass of the jet pair is used *only* in the selection of events, the Higgs boson mass is reconstructed from the momentum of the jet pair with M_{jj} set equal to M_z . As can be seen from Table II, in this case about 36% of the Higgs boson signal passing *both* the “zero-level” lepton cut

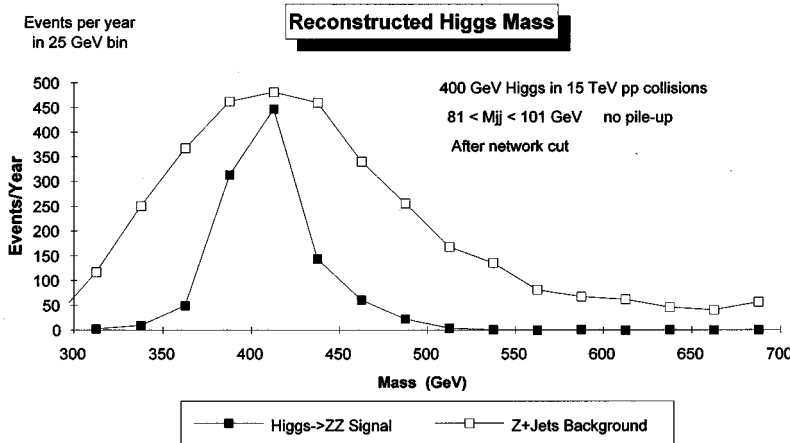


FIG. 11. Shows the reconstructed mass of a 400 GeV Higgs boson produced in 15 TeV proton-proton collisions. The plot corresponds to the number of events per year (with $\mathcal{L}=10^5/\text{pb}$) in a 25 GeV bin for the Higgs boson $\rightarrow ZZ$ signal and the Z +jets background for the ideal case where only one event at a time enters the detector (no pileup). The events have survived the “zero-level” lepton trigger and the jet-pair selection criterion with $81 < M_{jj}(\text{full}) < 101$ GeV and have passed the network cutoff (i.e., have $\tau_{\text{net}} > 0.75$).

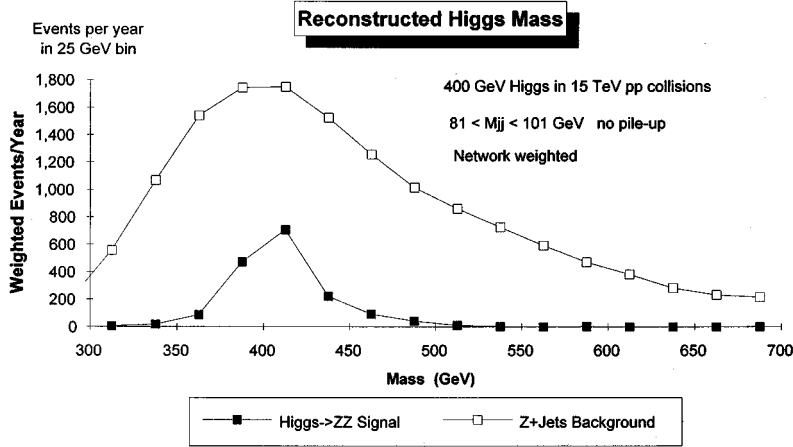


FIG. 12. Shows the reconstructed mass of a 400 GeV Higgs boson produced in 15 TeV proton-proton collisions weighted by the network output, z_{net} . The plot corresponds to the weighted number of events per year (with $\mathcal{L}=10^5/\text{pb}$) in a 25 GeV bin for the Higgs boson $\rightarrow ZZ$ signal and the Z+jets background for the ideal case where only one event at a time enters the detector (no pileup). The events have survived the “zero-level” lepton trigger and the jet-pair selection criterion with $81 < M_{jj}(\text{full}) < 101$ GeV.

and the jet-pair selection criterion have M_{jj} within this range, which is slightly less than the 51% for the no pileup case. About 7% of the Z+jets background events surviving both the “zero-level” lepton cut and the jet-pair selection criterion have a full jet-pair invariant mass in this range, which is about the same as the no pileup case. This corresponds to an overall enhancement factor at this stage of about 8 with an overall efficiency of about 19%, which is slightly worse than the no pileup case.

At this stage, Table II shows that there are about 1 500 Higgs events and 14 000 background events per year within 50 GeV of the true Higgs boson mass that pass the “zero-level” lepton trigger, the jet-pair selection criterion, and have $100 < M_{jj}(\text{full}) < 120$ GeV. This corresponds to an overall enhancement factor of about 21 with an overall efficiency of about 15%. With this enhancement, the Z+jets background is roughly nine times the signal. At this stage, we apply a neural network to improve the signal to background ratio beyond what can be achieved with these standard cuts.

Retraining the network

We use the same nine variables to characterize the events, but since these variables have changed dramatically, the network must be retrained. Figure 14 shows the new jet multiplicities. Figures 15 and 16 show that the fraction of transverse energy and mass, respectively, originating in the

extended region, $0.6 < R_{jj} < 1.0$, has greatly increased for both the signal and background events due to the pileup. Nevertheless, there are still slight differences between signal and background that the network can use to distinguish between the two.

The 9-16-8-1 (305) network is retrained on a sample of 2 741 signal and 3 566 background events that include the pileup interactions. Both signal and background events have already satisfied the “zero-level” lepton cuts, the jet-pair selection, and have $100 < M_{jj}(\text{full}) < 120$ GeV. To get this training sample it was necessary to generate 40 000 Higgs boson events with pileup and 400 000 Z+jet events with pileup. Running with pileup is a lot slower since a large number of events enter the calorimeter during each beam crossing. Because of this we are using a very small training sample. We could do better with a larger sample, but this is sufficient for what we want to illustrate in this paper. After training, we achieve a $\chi_{\text{net}}^2 = 0.1797$ with a network response for the training events shown in Fig. 17. Figure 17 also shows the network response (i.e., z_{net}) for an independent sample of signal and background events *not* used in the training. In spite of the small training sample, the network generalizes fairly well.

The network performance for the training sample is shown in Fig. 9 together with the no pileup case. Again we choose a network cutoff, z_{cut} of 0.75, which in this case for the training sample corresponds to a relative enhancement of about 6 with a relative efficiency of about 38%.

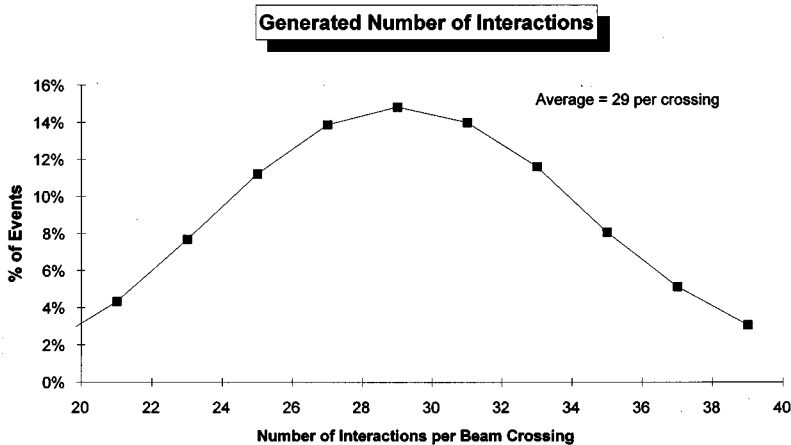


FIG. 13. Generated number of minimum bias interactions per beam crossing. These events enter the calorimeter together with one Higgs boson $\rightarrow ZZ$ signal event or one Z+jets background event to simulate the case of multiple interactions per beam crossing (pileup).

TABLE II. 400 GeV Higgs bosons produced in 15 TeV proton-proton collisions. The table shows the number of events per year (with $\mathcal{L}=10^5/\text{pb}$) for the Higgs boson $\rightarrow ZZ$ signal and the Z +jets background for the case of multiple interactions per beam crossing (i.e., with pileup). The “zero-level” lepton trigger is used as a reference point and is normalized to 100%. The enhancement factor is defined to be the percentage of signal divided by the percentage of background surviving the given set of cuts. Both the overall and relative enhancement factors are shown.

Selection or cut	$H \rightarrow ZZ$ signal			Z +jets background			Enhancement		
	% Relative	% Overall	Events/year	% Relative	% Overall	Events/year	Bak/sig	Relative Overall	
Lepton trigger:									
$PT(I) > 25$ GeV $PT(II) > 100$ GeV	100%	100%	10 212	100%	100%	1 973 919	193	1.0	1.0
Jet pair selection ($R_{jj} < 1.6$):									
$ET(j) > 25$ GeV $PT(jj) > 100$ GeV	53.3%	53.3%	5 440	33.6%	33.6%	662 850	122	1.6	1.6
Z-mass cut:									
$100 < M_Z < 120$ GeV	36.3%	19.3%	1 973	6.7%	2.3%	44 693	23	5.4	8.5
Higgs boson mass cut:									
$350 < M_H < 450$ GeV	75.5%	14.6%	1 489	30.5%	0.7%	13 615	9.1	2.5	21.1
Z-mass and net cut:									
$100 < M_Z < 120$ GeV $z_{\text{net}} > 0.75$	35.3%	6.8%	696	7.2%	0.2%	3 230	4.6	4.9	41.7
Higgs boson mass and net cut:									
$350 < M_H < 450$ GeV $z_{\text{net}} > 0.75$	38.1%	5.6%	568	11.2%	0.1%	1 525	2.7	3.4	72.0

Using the network cutoff

We now analyze an *independent* sample of signal and background events with pileup. Figure 18 shows the reconstructed Higgs boson mass for both the signal and background events that have passed the lepton cuts, the jet-pair selection with $100 < M_{jj}(\text{full}) < 120$ GeV, and the network cutoff (with $z_{\text{cut}} = 0.75$). Now, there are about 600 Higgs boson events and 1500 QCD background events per year within 50 GeV of the true Higgs boson mass of 400 GeV. This corresponds to an overall enhancement factor of about 72 (see Table II) with an overall efficiency of about 6%. Although the results are not quite as good as the no pileup case, signal and background are again roughly comparable and the network has improved the signal to background ratio by about a factor of 4.

VI. SUMMARY AND CONCLUSIONS

We have shown that neural networks are a useful tool in Higgs boson phenomenology. Using observables that measure how transverse energy and mass, respectively, are distributed around the away-side jet-jet system, a neural network can help to distinguish the two jet system originating from the $q\bar{q}$ decay of a color singlet Z boson from a random jet-pair coming from the “ordinary” QCD gluon bremsstrahlung of colored quarks and gluons. We have used the neural network in conjunction with the standard Higgs boson cuts to provide additional signal to background enhancements. Our procedure can be summarized by the following series of selections and cuts: lepton pair trigger, jet-pair selection, jet-jet profile cuts, jet-jet invariant mass cuts, and neural network cutoff. The invariant mass of the jet-pair is used *only* in the selection of events, the Higgs boson mass is recon-

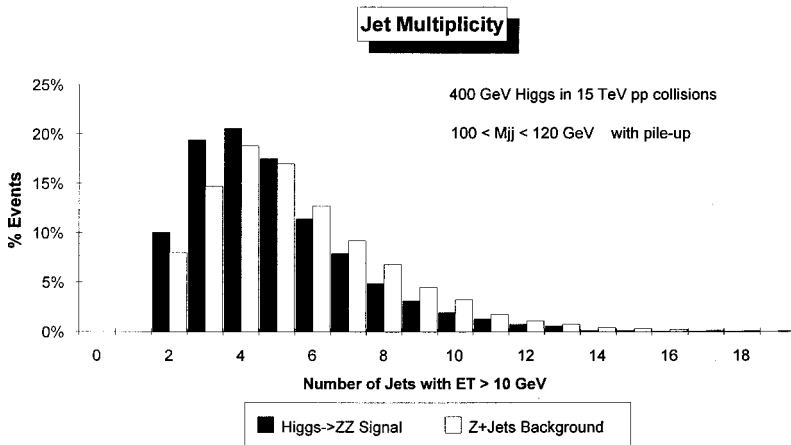


FIG. 14. Shows the multiplicity of jets for 400 GeV Higgs bosons produced in 15 TeV proton-proton collisions. The plot corresponds to the percentage of events with N jets with E_T greater than 10 GeV for the Higgs boson $\rightarrow ZZ$ signal and the Z +jets background for the case of multiple interactions per beam crossing (pileup). The events have survived the “zero-level” lepton trigger and the jet-pair selection criterion with $100 < M_{jj}(\text{full}) < 120$ GeV.

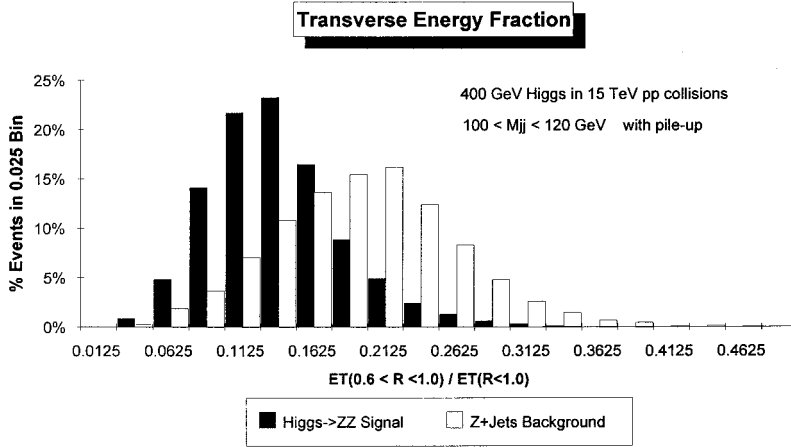


FIG. 15. Shows the fraction of transverse energy coming from calorimeter cells within the “halo” region $0.6 < R_{jj} < 1.0$ surrounding either of the away-side jets. The plot corresponds to the percentage of events with the jet-jet transverse energy fraction within the 0.025 bin for the Higgs boson $\rightarrow ZZ$ signal and the Z+jets background for the case of multiple interactions per beam crossing (pileup). The events have survived the “zero-level” lepton trigger and the jet-pair selection criterion and have $100 < M_{jj}(\text{full}) < 120$ GeV.

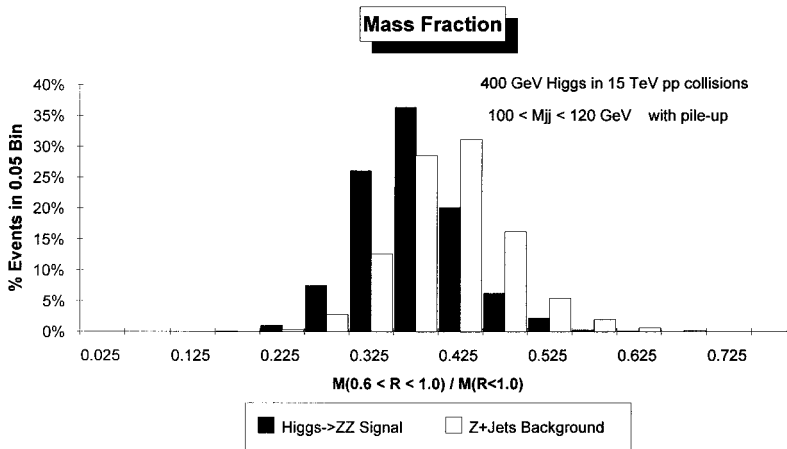


FIG. 16. Shows the fraction of invariant mass coming from calorimeter cells within the “halo” region $0.6 < R_{jj} < 1.0$ surrounding either of the away-side jets. The plot corresponds to the percentage of events with the jet-jet invariant mass fraction within the 0.05 bin for the Higgs boson $\rightarrow ZZ$ signal and the Z+jets background for the case of multiple interactions per beam crossing (pileup). The events have survived the “zero-level” lepton trigger and the jet-pair selection criterion and have $100 < M_{jj}(\text{full}) < 120$ GeV.

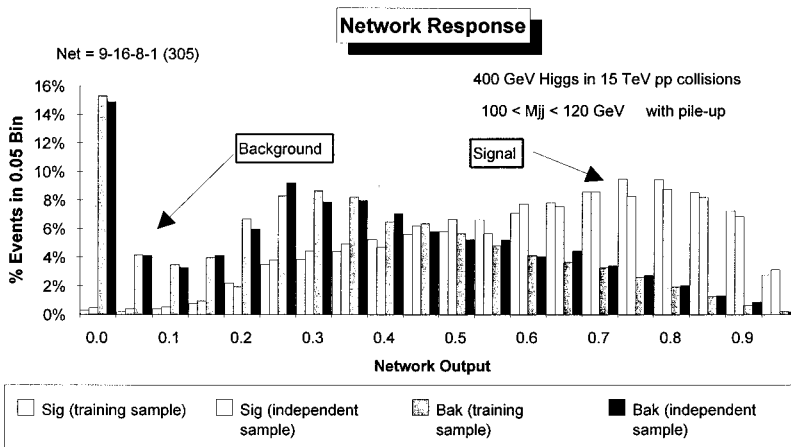


FIG. 17. Shows the network response, z_{net} , for the sample of signal and background events used in the training and for an independent sample of signal and background events. The plot corresponds to the percentage of events with z_{net} within a 0.05 bin for the Higgs boson $\rightarrow ZZ$ signal and the Z+jets background for the case of multiple interactions per beam crossing (pileup). The events have survived the “zero-level” lepton trigger and the jet-pair selection criterion and have $100 < M_{jj}(\text{full}) < 120$ GeV.

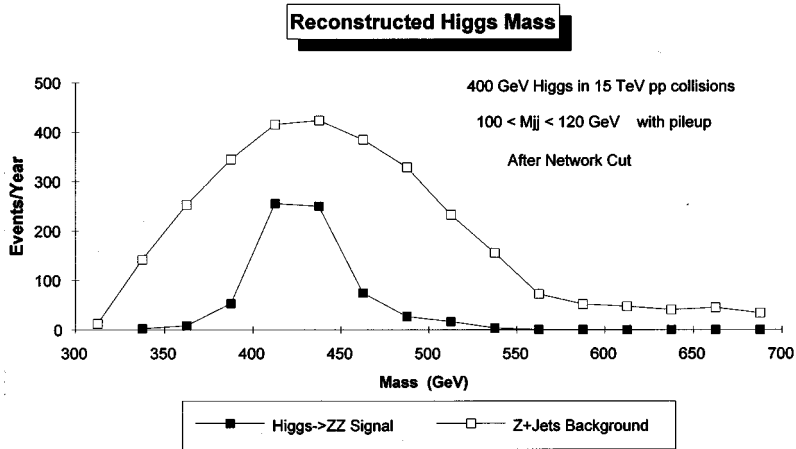


FIG. 18. Shows the reconstructed mass of a 400 GeV Higgs boson produced in 15 TeV proton-proton collisions. The plot corresponds to the number of events per year (with $\mathcal{L}=10^5/\text{pb}$) in a 25 GeV bin for the Higgs boson $\rightarrow ZZ$ signal and the Z+jets background for the case of multiple interactions per beam crossing (pileup). The events have survived the “zero-level” lepton trigger and the jet-pair selection criterion with $100 < M_{jj}(\text{full}) < 120$ GeV and have passed the network cutoff (i.e., have $z_{\text{net}} > 0.75$).

structed from the momentum of the jet pair with M_{jj} set equal to M_z . We are able to obtain an overall signal to background enhancement of around 10 with the standard Higgs boson cuts. The neural network provides an additional enhancement of 4–5 beyond what can be achieved with the standard data cuts resulting in an overall enhancement of about 50. We believe that we could further improve the network performance by using larger training samples and by increasing the number of input variables to include additional global information such as the number of forward jets in the event, etc.

Our method works even with a large number of interactions per beam crossing. This shows that some jet physics can be done even in the large pileup environment of the LHC. Although this paper is not a detailed simulation, experiments at the LHC should be able to do as well or better than our analysis. Furthermore, our procedure can be applied to W bosons and should help enhance the Higgs boson $\rightarrow WW \rightarrow l\nu jj$ signal at hadron colliders as well.

This work was supported in part by U.S. Department of Energy Grant No. DE-FG05-86ER-40272.

- [1] Patrick K. Simpson, *Artificial Neural Systems* (Pergamon, New York, 1990).
- [2] Maureen Caudill and Charles Butler, *Understanding Neural Networks* (MIT Press, Cambridge, MA, 1994).
- [3] CMS Collaboration, The Compact Muon Solenoid and Its Physics at the LHC, invited talk presented by Thomas Muller at the 10th Topical Workshop on Proton-Antiproton Collider

Physics, Fermilab, 1995 (unpublished).

- [4] ATLAS Collaboration, The Physics Potential of the ATLAS Experiment at LHC, invited talk presented by K. Jakobs at the 10th Topical Workshop on Proton-Antiproton Collider Physics, [3].
- [5] R. D. Field and P. A. Griffin, *Phys. Rev. D* **48**, 3167 (1993).
- [6] R. A. Fisher, *Ann. Eugenics* (1936).