Backpropagating neurons from bichromatic interaction with a three-level system

David Kagan and Harry Friedmann

Optical implementation of a backpropagating neuron by means of a nonlinear Fabry-Perot etalon requires thresholding a forward signal beam while the transmittance of a backpropagating beam is multiplied by the differential of the forward signal. This is achievable by inputting a bichromatic field to a three-level system in an optical cavity. The response characteristics of this device have the added possibility of adaptability of the threshold by the backward probe input intensity.

1. Introduction

Considerable interest has been shown recently in optical implementation of neural network models for associative memory applications. One of the most interesting properties of a neural network is the ability to learn dynamically the interconnections that correspond to a desired behavior through an iterative adaptation of the weight matrices. One of the methods for learning or training the neural system involves error driven adaptation. Error driven behavioral modification has the property that it can sense the system performance and adapt the weights in a manner which will compensate for device imperfections that cause the unwanted behavior. In Ref. 2 a multilayer optical learning network is described that uses self-aligning volume holograms to interconnect nonlinear elements acting as individual neurons. The basic learning algorithm is now reviewed; more details can be found in the above-mentioned reference.

A set of training patterns is presented repetitively in a random sequence to the network. Initially the output patterns differ from the desired response. An error pattern is formed, optically, by taking the difference between the output and the desired response. This error pattern is sent backward through the output neurons into the system to correct the weights of the interconnection matrices. For most problems of interest it is necessary to adapt complex nonlinear decision surfaces. This is done by stacking single layer networks to form a multilayer system which can learn to classify more complex patterns. The layers are connected holographically and adaptive training is performed by the backpropagating error signal through the system’s various layers. The backpropagating algorithm requires that the transmittance function of the error signal be proportional to the derivative of the forward signal sigmoid transfer function. The derivative will be peaked where the nonlinear sigmoid forward transfer function has large differential gain. If the neuron is operating at this level the connecting weights will be strongly modified by the error signal and the neuron is forced to a decision. This last condition requires a specific nonlinear optical gate.

Nonlinear optical gates are usually Fabry-Perot etalons containing some optically sensitive material. The nonlinear transfer characteristics are easily obtained but a backward signal which is the derivative of the forward mode presents a problem. In Ref. 2 this was solved by using two closely spaced cavities both addressed by the same forward and backward beams. In this case one cavity is optimized to produce a sigmoid response to the forward beam while blocking the backward signal, the other being resonant to the error signal. The response of the cavity which is resonant to the backpropagating beam is dependent on the forward propagating beam, which tunes the cavity right across the resonance. If the forward signal is of low intensity, the cavity is detuned from resonance. As the forward signal intensity increases the cavity is tuned into resonance. When the forward signal saturates, the cavity is detuned once more so that its transmission decreases. This method has some obvious difficulties. The forward signal is not allowed to resonate in the cavity resonant to the backpropagating beam. This being the case, a very intense signal will be required to tune the cavity response right through the resonance (linearly scan the Airy pattern). A further disadvantage of this system is the requirement of two cavities per neuron, which add volume and expense to the system. In this paper we suggest a nonlinear cavity which will give a forward nonlinear signal and an error probe which approximates the desired derivative extremely well. The gate will also have an adaptive threshold which is a distinct advantage for learning procedures.

1 May 1989 / Vol. 28, No. 9 / APPLIED OPTICS 1697
Perot contains a three-level system irradiated by two beams, one acting as the forward signal, and the other, of a different frequency, as the backward error correcting probe.

The layout of the paper is as follows. Section II presents the model of the three-level system in an optical cavity, and coupled input–output equations are derived. In Sec. III the input–output equations are solved and the desired response is shown to result.

II. Model

Consider the system, shown in Fig. 1, which consists of three energy levels. The system is irradiated by two fields, \( E_1 \) and \( E_2 \), given by

\[
E(t,z) = E_1 \exp(i\omega_1 t) + E_2 \exp(i\omega_2 t),
\]

where \( \omega_1 \) is close to the resonant frequency for the 1–3 transition and \( \omega_2 \) is tuned to the resonant frequency for the 2–3 transition. Using the density matrix formalism, the time development of the system can be written as

\[
\frac{\partial \rho}{\partial t} = i \frac{\rho}{\hbar} [\rho, H],
\]

where \( \rho \) is the density operator, \( H \) is the Hamiltonian, and the square brackets represent the commutator. After making the rotating wave approximation and writing Eq. (2) explicitly, we then write the optical Bloch equations for this system as follows:

\[
\frac{\partial \rho_{12}}{\partial t} = \frac{i E(t)}{\hbar} \left[ \mu_{13} \rho_{23} - \mu_{32} \rho_{13} - \rho_{13}(\gamma - i \omega_{21}) \right],
\]

\[
\frac{\partial \rho_{13}}{\partial t} = -\frac{i E(t)}{\hbar} \left[ \mu_{13} \rho_{32} + \rho_{12}(\gamma - i \omega_{31}) \right],
\]

\[
\frac{\partial \rho_{23}}{\partial t} = -\frac{i E(t)}{\hbar} \left[ \mu_{23} \rho_{32} - \rho_{23}(\gamma - i \omega_{32}) \right],
\]

\[
\frac{\partial r_{13}}{\partial t} = \frac{i E(t)}{\hbar} \left[ 2(\mu_{13} \rho_{32} - \mu_{31} \rho_{13}) \right.
\]
\[
- (\mu_{23} \rho_{23} - \mu_{32} \rho_{23}) - (r_{13} - r_{31})(\gamma),
\]

\[
\frac{\partial r_{32}}{\partial t} = \frac{i E(t)}{\hbar} \left[ \mu_{31} \rho_{13} - \mu_{13} \rho_{31} \right]
\]
\[
+ 2(\mu_{23} \rho_{23} - \mu_{32} \rho_{23}) - (r_{32} - r_{32})(\gamma),
\]

where \( r_{13} \) and \( r_{32} \) are the population differences given by \( r_{13} = \rho_{13} - \rho_{31} \) and \( r_{32} = \rho_{32} - \rho_{23} \). The superscript eq represents equilibrium, and so refers to the equilibrium population differences. In these equations \( \mu_{ij} \) represents the dipole moment connecting level \( i \) to \( j \) and \( \omega_{ij} \) is given by

\[
\omega_{ij} = \frac{(\epsilon_i - \epsilon_j)}{\hbar},
\]

where \( \epsilon_i \) is the energy of level \( i \). The population decay rate and the phase decay rate are taken as equal and have been written as \( \gamma \). The above equations have been solved for steady state in Refs. 3 and 4 and, for exactly resonant fields, the optical susceptibilities for the two frequencies \( \omega_1 \) and \( \omega_2 \) are given by

\[
x(\omega_1) = -\frac{2\omega_1}{\hbar_1} \frac{4 + X_1^2 + 4X_1^2}{(1 + X_1^2 + X_2^2)(4 + X_1^2 + X_2^2)},
\]

\[
x(\omega_2) = \frac{2\omega_2}{\hbar_2} \frac{3X_1^2}{(1 + X_1^2 + X_2^2)(4 + X_1^2 + X_2^2)},
\]

In the above equations we have scaled the fields by the square root of the saturation intensities for the respective transitions so that we can write

\[
X_1 = \frac{I_{E_1}}{1 + X_1^2 + X_2^2}, \quad X_2 = \frac{I_{E_2}}{1 + X_1^2 + X_2^2},
\]

\[
\alpha_{\omega_1} = \frac{N \rho_{13}^2 k_1}{V \gamma 2 \rho_{12} h}, \quad \alpha_{\omega_2} = \frac{N \rho_{32}^2 k_2}{V \gamma 2 \rho_{23} h},
\]

where \( k_1 \) and \( k_2 \) are the magnitudes of the wave vectors at frequencies \( \omega_1 \) and \( \omega_2 \). Following Ref. 5 and making the mean field assumption, the cavity input–output steady state equations are written as

\[
Y_1 = X_1 \left[ 1 + 2\frac{C_1}{F} (4 + X_1^2 + X_2^2) \right],
\]

\[
Y_2 = X_2 \left[ 1 + 2\frac{C_2}{F} (3X_1^2) \right],
\]

where \( F \) is given by

\[
F = (1 + X_1^2 + X_2^2)(4 + X_1^2 + X_2^2),
\]

where \( Y_1 \) and \( Y_2 \) are the scaled input fields given by

\[
Y_1 = \frac{|E_{in}|}{h \gamma \sqrt{2}}, \quad Y_2 = \frac{|E_{in}|}{h \gamma \sqrt{2}},
\]

Here \( C_1 \) and \( C_2 \) are given by

\[
C_1 = \alpha_{\omega_1} LR_1, \quad C_2 = -\alpha_{\omega_2} LR_2
\]

where \( R_i \) is the reflection coefficient for the cavity dielectric mirrors at frequency \( \omega_i \), \( L \) is the length of the cavity, \( \alpha_{\omega_1} \) and \( \alpha_{\omega_2} \) are the absorption and emission coefficients as previously defined. The factor \( C \) is basically equal to the absorption or gain per pass through the cavity divided by the transmission coefficient. A large \( C_1 \) implies more absorption (until saturation), and a large \( C_2 \) leads to increased gain of the input with frequency \( \omega_2 \). To make the mean field...
assumption the gain or absorption per pass through the cavity should not be large. Since our model is only intended to give qualitative results this assumption has been made for simplicity.

It is instructive at this point to plot the on-resonance absorption and emission as given by Eqs. (8) and (9) as functions of the scaled fields $X_1$ and $X_2$. This is shown in Figs. 2 and 3. In Fig. 2 one notes that raising the pump intensity $X_1$ at first increases the susceptibility of frequency $\omega_1$ until it saturates. For frequency $\omega_2$, increasing $X_1$ increases its susceptibility and thus its gain, but as $X_1$ reaches saturating intensities the emission decreases as can be seen in Fig. 3. The reason for this decrease of the gain results from the splitting of level 3 by the pump such that the $X_2$ beam is now detuned from the 3–2 resonance. It is this property that immediately suggests the use of this system to obtain the desired response for the backpropagating neuron. As the field $X_1$ increases the emitted field first rises and then falls such that the maximum can be expected at the point of highest gain for $X_1$. The nonlinear response is obtained by putting the whole system in a cavity as usual. The cavity response is described in the following section.

III. Input–Output Cavity Response

Using Eqs. (10) and (11) the cavity response to the two beams can be evaluated. In Fig. 4 three curves have been plotted: $a$, the forward signal output $X_1$ vs input $Y_1$; $b$, the backward signal transmission $X_2/Y_2$; and $c$, the differential of the forward signal $dX_1/dY_1$. Let us now consider the application of this cavity as an optical neuron. If the forward propagating signal has the frequency $\omega_1$, output $X_1$ has a typical nonlinear response to the input, as in curve $a$ in Fig. 4. If an error signal is detected and sent back to the system at frequency $\omega_2$, as input $Y_2$, transmission $X_2/Y_2$, shown as $b$, is then proportional to the differential of the forward propagating signal $c$. It should be noted that the output of the forward signal $X_1$ is not a sigmoidal nonlinearity, the gate being transparent at saturating intensities. This is generally the case for the optical etalons suggested for implementation of nonlinear gates. When a true sigmoidal output in which the transmission levels off at high intensities is required, an optical limiter is used in conjunction with the gate. In the system we propose, the backward transmission is proportional to the differential of an ideal sigmoidal output, falling to zero as the intensity increases. In other words, the backward signal has a zero response after the forward signal has passed the decision threshold, and the weights need not be changed.
The three-level system also displays the possibility of tunability. If the control beam intensity $Y_2$ is charged, the output response will be shifted somewhat and the width and slope of the high gain will be changed. This is more clearly seen when parameters which give optical bistable output of $X_1$ are chosen. In Figs. 5 and 6 these outputs are plotted for different values of the control $Y_2$ input. We note that the tuning is accomplished by low power inputs of the control beam since both beams are allowed to resonate in the cavity. The idea of using a control beam to shift the width and slope of the output curves can be utilized to some advantage. Specifically, if one wishes to correct errors by simply moving the threshold, the control beam $Y_2$ could act as an input in the forward direction. Other applications of the tunability of optically bistable outputs have been presented elsewhere.\textsuperscript{6-8} The case of tunable optical bistability was also pointed out in Ref. 9, but there the control was of the same order of intensity as the main output, since it was not allowed to resonate in the cavity.

IV. Conclusion

The three-level system analyzed here shows the nonlinear optical properties required for backpropagating neurons as explained. The reason for the decrease of the gain of the frequency tuned to the 3–2 transition is the splitting of level 3 by the intense saturating pump. In a cavity, a resonant mode of the transition from the split level to level 2 might be excited if the gain is large. A difficulty in implementing this neuron is the requirement of generating an error signal with a different frequency. This would have to be done using a nonlinear interaction, such as a phase conjugate frequency shifter, at the output level. Although we recognize this limitation, it does not seem too much of a price to pay for an all-optical neuron which by its physical properties implements the backpropagating algorithm. It should also be pointed out that whatever system one is using, the backpropagating signal must be distinguishable from the forward beam.

The three-level system also shows the possibility of tunable optical bistability and adaptive thresholding by low power control beams, which could well be of use in optical computing.

References