Analog Integrated Circuits and Signal Processing 2, 223-229 (1992)

Correlation-Based Associative Memory and Its MOS Implementation

S. OH, S. CHUNG, R.J. MARKS II, H.J. YOUN, AND D.C. PARK*
Department of Electrical Engineering, FT-10, University of Washington, Seattle, WA 98195
**Department of Electrical and Computer Engineering, Florida International University, Miami, FL 33199*

Received February 8, 1991.

Abstract. We propose an associative memory based on minimization of a free-energy function determined by the library vectors to be stored. When the library vectors are bipolar, the energy function contains minima at the library vector location. The minima can be sought by search techniques such as gradient descent. Significantly, if the correlation nonlinearity is chosen to be sufficiently strong, then convergence occurs in a single step. We demonstrate how the algorithm can be implemented using MOS circuitry.

1. Introduction

Associative memory architectures have been presented that are based on artificial neural network structures [1, 2] and iterative matched filters. Indeed, in certain configurations, the resulting procedures produce numerically identical results [3].

Similar associative memories based on nonlinear operations in the correlation domain have also been proposed [4-9]. Such memories have shift-invariant characteristics [6] and high capacity [10, 11]. If the correlation nonlinearity is stronger than $z(x) = N^{x/2}$ and the library vectors are bipolar (i.e., contain only ±1's), then the iterations normally required in such memories need not be used. Convergence, rather, occurs in one step [7-9].

In this paper, we examine the correlation based associative memory and show its relation to minimization of a free-energy function placed on the training data. Using gradient descent, the search is shown to converge in a single step for the case of a bipolar library.

A circuit implementation of the resulting associative memory is proposed. The correlation operation is performed using metal oxide semiconductor (MOS) capacitors and MOS switches. The subthreshold region of MOS is used to implement the exponential nonlinearity.

2. Probability of Occurrence and Free Energy

Let $\{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_N\}$ be library vectors of dimension L that we wish to store in the associative memory. Let $\mathbf{f}$ be a vector of like dimension. Then, the probability $P(\mathbf{f})$ is

$$P(\mathbf{f}) = \sum_{i=1}^{N} P(\mathbf{f}_i)P(\mathbf{f}\backslash\mathbf{f}_i) = \sum_{i=1}^{N} Q_i P(\mathbf{f}\backslash\mathbf{f}_i)$$

where $\mathbf{f}$ is an arbitrary vector and Q_i is the priori probability of the library vector $\mathbf{f}_i$. We define an energy function $F(\mathbf{f})$ as

$$F(\mathbf{f}) = -\gamma \ln[P(\mathbf{f})] = -\gamma \ln \left[\sum_{i=1}^{N} Q_i P(\mathbf{f}\backslash\mathbf{f}_i)\right]$$

where γ is a constant. Note that $F(\cdot)$ is analogous in concept to free energy [12]. Our associative memory paradigm is as follows. Let $\mathbf{g}$ denote a distorted version of one of the library vectors. With an initialization of $\mathbf{g}$, we search the landscape of F for a minimum. We will show that the minima occur at library vectors location in many important cases.

Example 2.1. If $P(\mathbf{g}\backslash\mathbf{f}_i)$ has an identically independent Gaussian distribution, then

$$P(\mathbf{g}\backslash\mathbf{f}_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{\| \mathbf{g} - \mathbf{f}_i \|^2}{2\sigma^2}\right]$$

Let $\gamma = 2\sigma^2$ and $a = 1/\gamma$. Then

$$F(\mathbf{g}) = -\frac{1}{a} \ln \left[\sum_{i=1}^{N} Q_i \exp\{-a \| \mathbf{g} - \mathbf{f}_i \|^2\}\right] + C \quad (1)$$

where C is a constant. An example of the energy landscape is illustrated in figure 1 for $a = 8$ and $a = 32$ with $C = 0$. The library vectors are $L = 1$ dimensional and are $\{-1, 0, 0.5, 1\}$ for both values of a. The energy function reaches local minima at the neighborhood of

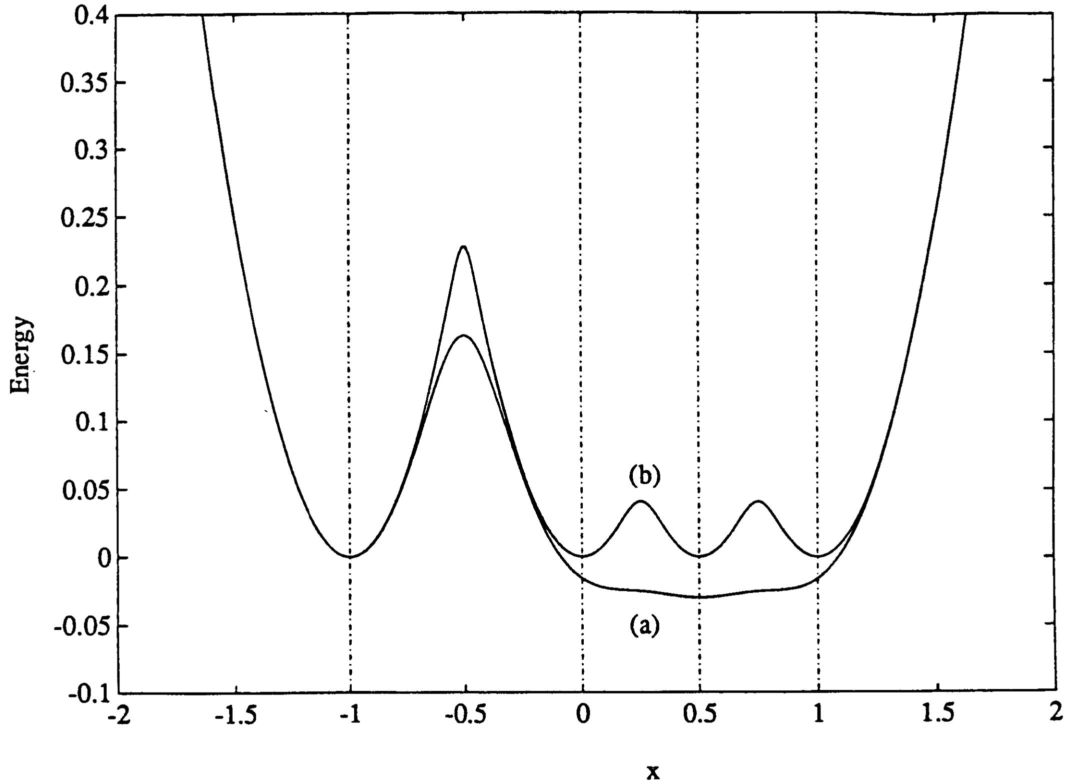


Fig. 1. The energy in the Gaussian noise for (a) $a = 8$ and (b) $a = 32$. The library vectors (scalars in this example) are −1, 0, 0.5, and 1.

the library vector location when $a = 32$, but this is not the case when $a = 8$. Thus we see that the minima become more distinct for large values of a (which, in the free-energy analogy, is inversely proportional to temperature).

Example 2.2. If all vectors are bipolar (i.e., contain only ± 1 components) and the noise is independent flip (Bernoulli) noise with probability $p < 1/2$, then, under the assumption that $\mathbf{f}$ is bipolar,

$$P(\mathbf{g} \backslash \mathbf{f}_i) = p^{k_i}(1 - p)^{L-k_i} \tag{2}$$

where k_i is the Hamming distance between $\mathbf{g}$ and $\mathbf{f}_i$. Equation (2) can be written as

$$P(\mathbf{g} \backslash \mathbf{f}_i) = (1 - p)^L \left(\frac{p}{1-p}\right)^{k_i} = (1 - p)^L \exp[-k_i b] \tag{3}$$

where

$$b = \ln\left(\frac{1-p}{p}\right)$$

and $\| \mathbf{g} - \mathbf{f} \|^2 = 4k_i$. Equation (3) can then be written as

$$P(\mathbf{g} \backslash \mathbf{f}_i) = (1 - p)^L \exp[-a \| \mathbf{g} - \mathbf{f}_i \|^2]$$

and let $a = b/4$. It follows that the energy function has the same form as (1)

$$F(\mathbf{g}) = -\gamma \ln \left[\sum_{i=1}^{N} Q_i \exp\{-a \| \mathbf{g} - \mathbf{f}_i \|^2\} \right] + C \tag{4}$$

3. Derivation of Operation

Searching for the maximum of $P(\mathbf{g})$ is equivalent to searching for the minimum of $F(\mathbf{g})$. The gradient descent method is a popularly used search technique and can be written as

$$\mathbf{g}(n + 1) = \mathbf{g}(n) - \beta \nabla_g F(\mathbf{g}) \tag{5}$$

where $\nabla_g F(\mathbf{g})$is gradient of $F(\mathbf{g})$. If initialization of the search is close to minimum, say $\mathbf{f}_1$, then we expect that $\mathbf{g}(\infty) = \mathbf{f}_1$.

3.1. Continuous Level Library Vector

In Example 2.1, the gradient is

$$\nabla_g F(\mathbf{g}) = -\frac{1}{a} 2a \frac{\sum_{i=1}^{N} Q_i(\mathbf{g} - \mathbf{f}_i) \exp(-a \parallel \mathbf{g} - \mathbf{f}_i \parallel^2)}{\sum_{i=1}^{N} Q_i \exp(-a \parallel \mathbf{g} - \mathbf{f}_i \parallel^2)}$$

$$= 2\mathbf{g} - 2\frac{\sum_{i=1}^{N} Q_i\mathbf{f}_i \exp(-a \parallel \mathbf{g} - \mathbf{f}_i \parallel^2)}{\sum_{i=1}^{N} Q_i \exp(-a \parallel \mathbf{g} - \mathbf{f}_i \parallel^2)}$$

Equation (5) is

$$\mathbf{g}(n+1) = \mathbf{g}(n) - \beta \nabla_g F(\mathbf{g})$$
$$= (1 - 2\beta)\mathbf{g}(n) + 2\beta \frac{\sum_{i=1}^{N} Q_i\mathbf{f}_i \exp(-a \parallel \mathbf{g} - \mathbf{f}_i \parallel^2)}{\sum_{i=1}^{N} Q_i \exp(-a \parallel \mathbf{g} - \mathbf{f}_i \parallel^2)}$$

Thus, if $\parallel \mathbf{f}_i \parallel$ is the same for all library vectors and $\beta = 1/2$, then

$$\mathbf{g}(n+1) = \frac{\sum_{i=1}^{N} Q_i\mathbf{f}_i \exp[2a\alpha_i(n)]}{\sum_{i=1}^{N} Q_i \exp(2a\alpha_i(n)]} \quad (6)$$

where $\alpha_i(n) = \mathbf{g}^T \mathbf{f}_i$ is the correlation of $\mathbf{g}$ with $\mathbf{f}_i$.

3.2. Bipolar Library Vector

For the bipolar library with equal priori probability, a sufficiently large value of a and corresponding value of β, assures that the gradient descent method converges in one step, if, after the iteration, we project onto the nearest bipolar vector. For the more general case where the nonlinearity is not sufficiently strong, we calculate from (4).

$$\nabla_g F(\mathbf{g}) = \gamma a \left\{ \mathbf{g} - \frac{\sum_{i=1}^{N} Q_i\mathbf{f}_i \exp(2a\alpha_i)}{\sum_{i=1}^{N} Q_i \exp(2a\alpha_i)} \right\}$$

where $\alpha_i = \mathbf{f}^T\mathbf{g}$. If we choose the gain

$$\beta = \frac{1}{\gamma a}$$

then the recurrent relation will be (6). Generally, $\mathbf{g}(n + 1)$ here is not bipolar, so we project it onto the nearest bipolar vector by the sign(•) operation

$$\mathbf{g}(n+1) = \text{sign}\left[\frac{\sum_{i=1}^{N} Q_i\mathbf{f}_i \exp[2a\alpha_i(n)]}{\sum_{i=1}^{N} Q_i \exp[2a\alpha_i(n)]} \right] \quad (7)$$

The block diagram of this operation is shown in figure 2, and this operation is the same as the exponential function at the correlation domain in iterative matched filter [7, 8].

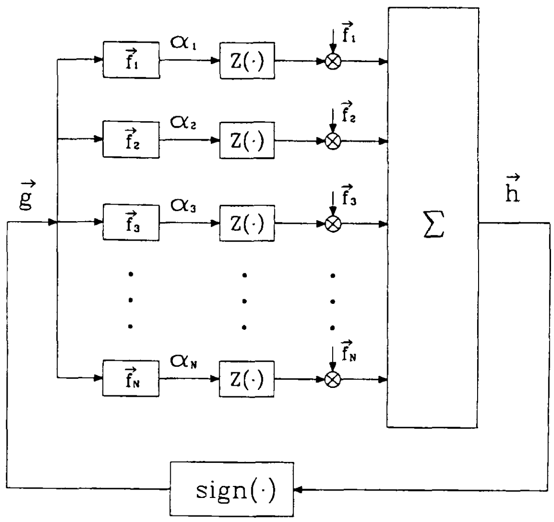


Fig. 2. Block diagram of a matched filter with a nonlinearity in the correlation domain. $A(x) = e^{2ax}$.

3.3. One Step Convergence for Bipolar Library

Consider the case where the priori probability, Q_i, is the same for all library vectors. One step convergence can then be assured for the iteration in (7). Specifically, if $a > \ln(N - 1)/2$, the operation in (7) converges to one of the library vector in one step [7, 8]. This condition is the same as requiring that $p < 1/N$. Under this condition, we establish the following

Lemma 3.1. Let $\mathbf{g}$ and $\mathbf{f}_i$ be bipolar vectors. If $a > \ln(N - 1)/2$, then $F(\mathbf{g}) > F(\mathbf{f}_i)$ for $\mathbf{g} \neq \mathbf{f}_i$ for all $i = 1, 2, \ldots, N$.

Proof. Let $c_{ik} = h(\mathbf{f}_i, \mathbf{f}_k)$ and $d_k = h(\mathbf{g}, \mathbf{f}_k)$ where $h(\mathbf{x}, \mathbf{y})$ is the Hamming distance between $\mathbf{x}$ and $\mathbf{y}$. Let

$$A(\mathbf{f}_i, \mathbf{g}) = \sum_{k=1}^{N} e^{-4ac_{ik}} - \sum_{k=1}^{N} e^{-4ad_k} \tag{8}$$

Because $h(\mathbf{f}_i, \mathbf{f}_k) \leq h(\mathbf{f}_i, \mathbf{g}) + h(\mathbf{f}_k, \mathbf{g})$,

$$e^{-4ac_{ik}} \geq e^{-4ad_i} e^{-4ad_k}$$

Equation (8) then becomes

$$A(\mathbf{f}_i, \mathbf{g}) \geq 1 + \sum_{k \neq i}^{N} [e^{-4ad_i} e^{-4ad_k}] - e^{-4ad_i} - \sum_{k \neq i}^{N} e^{-4ad_k}$$

$$= 1 - \sum_{k \neq i}^{N} e^{-4ad_k} - e^{-4ad_i} \left[1 - \sum_{k \neq i}^{N} e^{-4ad_k} \right]$$

$$= [1 - e^{-4ad_i}] \left[1 - \sum_{k \neq i}^{N} e^{-4ad_k} \right]$$

If $a > \ln(N - 1)/2$, then, for $n \geq 1$

$$e^{4an} \leq \frac{1}{N - 1}$$

This leads to conclude that $A(\mathbf{f}_i, \mathbf{g}) \geq 0$ and, as a consequence, $F(\mathbf{g}) > F(\mathbf{f}_i)$. □

The above lemma establishes that the free energy has local minima at all library vector locations. One step convergence using (7) attains this local minimum without iteration.

4. Circuit Implementation of the Operation

The operation in equation (7) consists of two parts. The first is calculation of the correlation between the input and each library element. The second is generation of the nonlinearity operation and sign(·) function operations. The implementation of correlation can be achieved a matrix vector multiplier and that of the nonlinear function by a cascade connection of the multiple differential amplifier and a high-gain amplifier.

We give an example of the implementation using CMOS technology. The bipolar matrix vector multiplier can be implemented by MOS switches and MOS capacitors. Figure 3 is an example for $\mathbf{f} = [1, -1, -1, 1]^T$. The connection between capacitors and the voltage sources depend on the values of elements of **f**. Also, the switches are operated by the values of input **g**. The operation is as follows

1. Charging mode:
   a. The switch S_5 is connected to V_{DD}.

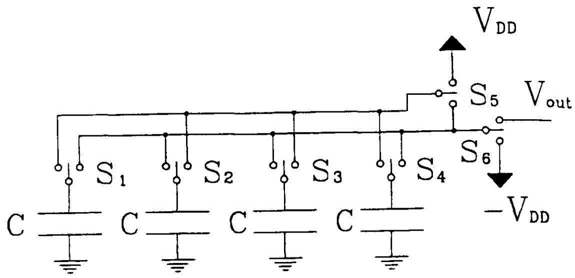


Fig. 3. An example of MOS implementation of correlation with a bipolar library vector.

   b. The switch S_6 is connected to $-V_{DD}$.
   c. The switches S_1, S_2, S_3, and S_4 are connected to the left side if the input is 1 and to the right side otherwise.
2. Distributing mode:
   a. The switches S_1, S_2, S_3, and S_4 are connected to the right side.
   b. The switch S_5 is closed to the lower terminal.
   c. The switch S_6 is closed to V_{out}.

We will explain the operation of the circuit for $\mathbf{g} = [1, 1, 1, -1]^T$. After the charging mode, the voltages of the capacitors are V_{DD}, $-V_{DD}$, $-V_{DD}$, $-V_{DD}$. In the distributing mode, because the switch 5 is connected to lower terminal, all capacitors are connected parallel. The voltage across each capacitor is the same and the value is $-2V_{DD}/4$. The value -2 is the same as $\mathbf{f}^T\mathbf{g}$, so this circuit gives the correlation operation.

Figure 4 shows the basic schematic circuit for nonlinearity implementation. If all of the MOS transistors are operated in the subthreshold region, then the current voltage relationship of the NMOS follows as [13–15].

$$I_D = I_{do} \exp\left(\frac{\kappa q V_g}{kT}\right) \left[\exp\left(-\frac{qV_s}{kT}\right) - \exp\left(-\frac{qV_d}{kT}\right)\right]$$

If $\kappa V_g - V_d$ is negative, then we can ignore the last term of the above equation. In figure 4,

$$I_1 = \sum_{i=1}^{p} I_{do_i}^{+} \exp\left[\frac{q(\kappa V_i^+ - V_s)}{kT}\right]$$

$$I_2 = \sum_{j=1}^{r} I_{do_j}^{-} \exp\left[\frac{q(\kappa V_j^- - V_s)}{kT}\right]$$

Using the Kirchhoff's law, $I_o = I_1 + I_2$. The current mirror action by T_1 and T_2 leads to $I_{out} = I_1 - I_2$. Then

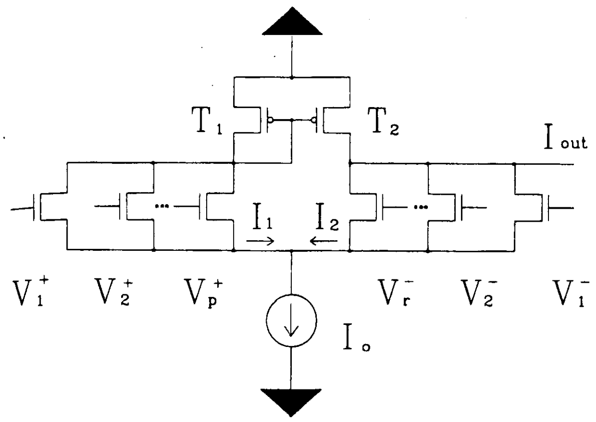


Fig. 4. A CMOS implementation of correlation with a bipolar library vector.

$$I_{\text{out}} = I_o \frac{\sum_{i=1}^{p} I_{do_i}^{+} \exp(uV_i^{+}) - \sum_{j=1}^{r} I_{do_j}^{-} \exp(uV_j^{-})}{\sum_{i=1}^{p} I_{do_i}^{+} \exp(uV_i^{+}) + \sum_{j=1}^{r} I_{do_j}^{-} \exp(uV_j^{-})}$$

where

$$u = \frac{\kappa q}{kT}$$

Here, the I_{do}'s are the function of the *W/L* value of gate.

Figure 5 is an example of the circuit implementation whose library vectors are

$$\mathbf{f}_1 = [1, -1, 1, -1, -1]^T$$
$$\mathbf{f}_2 = [1, -1, -1, -1, 1]^T$$
$$\mathbf{f}_3 = [1, -1, 1, 1, -1]^T$$

In this figure, the circuit has three stages. The first is the correlation operation using MOS switches and capacitors. The second is the nonlinear operation using the CMOS which is operated exponentially in the subthreshold region. The third is a saturated high gain amplifier, i.e., the comparator or buffer with switches.

5. Conclusion

In this paper, we found the most probable solution for correlation based associative memories using an energy function equal to the logarithm of the probability. For

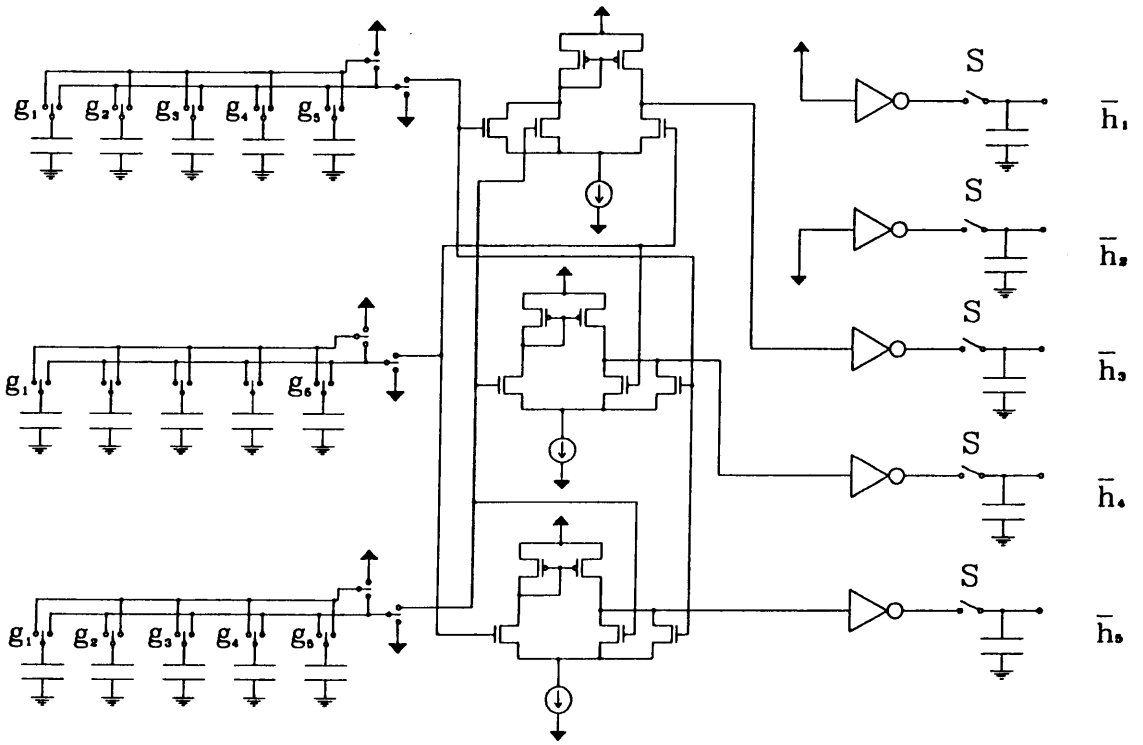


Fig. 5. The schematic of the CMOS implementation for the library $[1, -1, 1, -1, -1]^T$, $[1, -1, -1, -1, 1]^T$, and $[1, -1, 1, 1, -1]^T$. The switches in the correlation oprator operate by $g_1(n)$, $g_2(n)$, $g_3(n)$, $g_4(n)$, and $g_5(n)$. The outputs are $\bar{h}_1(n)$, $\bar{h}_2(n)$, $\bar{h}_3(n)$, $\bar{h}_4(n)$, and $\bar{h}_5(n)$ where $h_i(n) = g_i(n + 1)$. The switches S are closed in the distributing mode.

bipolar library vectors, the associative memory with an exponential nonlinearity in correlation domain minimizes the energy which is proportional to ln[P(f)]. Also, we showed that one-step convergence to the desired local minimum can occur for a bipolar library with equal priori probabilities for sufficiently strong nonlinearity. We demonstrated that this associative memory can be implemented using the CMOS technology.

References

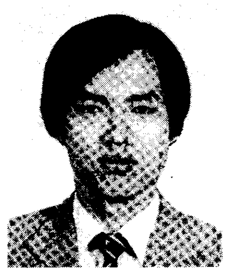
1. J.J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proc. Nat. Acad. Sci. USA*, pp. 2554–2558, 1982.
2. R.P. Lippman, "An introduction to computing with neural nets," *IEEE ASSP Mag.* p. 7, April 1987.
3. R.J. Marks II and L.E. Atlas, "Geometrical interpretation of Hopfield's content addressable memory neural network," in *Northcon/ 88 Conf. Record*, Seattle, Washington, 1988.
4. B.H. Soffer, G.J. Dunning, Y. Owechko, and E. Marom, "Associative holographic memory with feedback using phase-conjugate mirrors," *Opt. Lett.*, vol. 11, pp. 1118–120, 1986.
5. G.J. Dunning, E. Marom, Y. Owechko, and B.H. Soffer, "Optical holographic associative memory using a phase conjugate resonator," *SPIE*, pp. 205–213, 1986.
6. D. Psaltis and J. Hong, "Shift-invariant optical associative memory," *Opt. Eng.*, vol. 26, pp. 10–15, 1987.
7. R.J. Marks II, L.E. Atlas, J.J. Choi, S. Oh, K.F. Cheung, and D.C. Park, "Performance analysis of associative memories with nonlinearity in the correlation domain," *App. Opt.*, vol. 27, pp. 2900–2904, 1988.
8. R.J. Marks II, L.E. Atlas, J.J. Choi, S. Oh, K.F. Cheung, and D.C. Park, "Nonlinearity requirements for correlation based associative," *SPIE*, pp. 179–183, 1988.
9. J.J. Choi, *A Performance Analysis of Associative Memories with Nonlinearities in the Correlation Domain*. University of Washington, Master Thesis, 1987.
10. Y. Owechko, G.J. Dunning, E. Marom, and B.H. Soffer, "Holographic associative memory with nonlinearities in the correlation domain," *Appl. Opt.*, vol. 26, pp. 1900–1910, 1987.
11. T. Chiueh and R.M. Goodman, "High-capacity exponential associative memories," in *Proc. IEEE Int. Conf. Neural Networks*, San Diego, California, 1988, pp. I 153–I 159.
12. F. Reif, *Fundamentals of Statistical and Thermal Physics*, McGraw-Hill: New York, 1965.
13. C. Mead, *Analog VLSI and Neural Systems*, Addison-Wesley: Reading, MA, 1989.
14. Y. Tsividis and P. Antognetti, *Design of MOS VLSI Circuits for Telecommunications*, Prentice-Hall: Englewood Cliffs, 1985.
15. Carver Mead and Mohammed Ismail, *Analog VLSI Implementation of Neural Systems*, Kluwer Academic: Boston, 1989.

Seho Oh received the B.S. degree in electronics engineering from Seoul National University and the M.S. degree in electrical engineering from Korea Advanced Institute of Science and Technology, Seoul. From 1981 through 1986, he was in Central Research Laboratory of Goldstar Company. He received his Ph.D. in electrical engineering from the University of Washington, Seattle, in 1989. His research interests are in the area of signal analysis, artificial neural networks and pattern recognition. Dr. Oh is the coauthor of over 30 archival and proceedings papers and has been issued two United States patents.

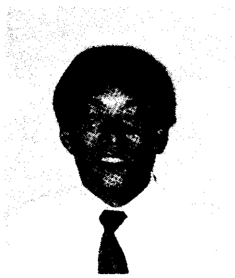
Suk-Yeon Chung received the B.S. and M.S. degrees in physics from Ehwa Womens University, Seoul and the University of Washington, Seattle, respectively. Currently, she is working toward a Ph.D. in electrical engineering at the University of Washington. Her research interests are in the areas of semiconductor physics and quantum wells.

Robert J. Marks II is a professor in the Department of Electrical Engineering at the University of Washington, Seattle. Professor Marks was awarded the Outstanding Branch Councilor award in 1982 by

IEEE and, in 1984, was presented with an IEEE Centennial Medal. He was chair of IEEE Neural Networks Committee and was the cofounder and first chair of the IEEE Circuits and Systems Society Technical Committee on Neural Systems and Applications. Professor Marks was also elected the first president of the IEEE Council on Neural Networks. He was named an IEEE Distinguished Lecturer in 1992. Professor Marks is a Fellow of the Optical Society of America and a senior member of IEEE. He was the cofounder and first president of the Puget Sound Section of the Optical Society of America and was elected that organization's first honorary member. He is cofounder and current president of Multidimensional Systems Corporation in Lynnwood, Washington. Professor Marks is the editor-in-chief of the *IEEE Transactions on Neural Networks*. He was the topical editor for Optical Signal Processing and Image Science for the *Journal of the Optical Society on America—A* (1989–91) and a member of the editorial board for *The International Journal of Neurocomputing* (1989–92). Dr. Marks served as North American liaison for the 1991 Singapore International Joint Conference on Neural Networks (IJCNN), program cochair of the 1993 Nagoya IJCNN, and international chair of the 1992 RNNS/IEEE Symposium on Neuroinformatics and Neurocomputing (Rostov-on-Don), USSR). Dr. Marks also served as the organizational chair of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis (Victoria, BC, 1992), the program and tutorials chair for the First International Forum on Applications of Neural Networks to Power Systems (Seattle, 1991), and the general chair of the International Symposium on Circuits and Systems (Seattle, 1995). He has published over 100 journal and proceeding papers in the areas of signal analysis, detection theory, signal recovery, optical computing, signal processing, fuzzy systems, and artificial neural processing. Dr. Marks has two U.S. patents in the field of artificial neural networks. He is the author of the book *Introduction to Shannon Sampling and Interpolation Theory* (Springer-Verlag, 1991). Dr. Marks is a cofounder of the Christian Faculty Fellowship at the University of Washington and serves as the faculty advisor to the University of Washington's chapter of Campus Crusade for Christ. He is a member of Eta Kappa Nu and Sigma Xi.



Hyoung Jo Youn received the B.S. degree in electronics engineering from the Seoul National University in 1979, the M.S. degree in electrical engineering from the Korea Advanced Institute of Science and Technology in 1981, and the Ph.D. in electrical engineering at the University of Washington in 1990. His Ph.D. work was on AlGaAs/GaAs photodetectors using both quantum well and bulk structures. From 1981 to 1985, he worked at the Daewoo Semiconductor Company on bipolar analog IC design. He is currently working at the Hewlett Packard Company, Santa Clara, CA, where he is responsible for the cell library design in CMOS and BiCMOS technology. His current interests are in optical detectors using III-V compound semiconductors and logic design in BiCMOS/CMOS.



Dong Chul Park received his B.S. degree from Sogang University, M.S. degree from Korea Advanced Institute of Science and Technology, Seoul, Korea, and Ph.D. in electrical engineering at the University of Washington, Seattle, in 1980, 1982, and 1990, respectively. During 1982–1985, he worked at the Central Research Lab. of Gold Star Co., Seoul, Korea. Currently, he is an assistant professor with the Department of Electrical and Computer Engineering at the Florida International University, the State University of Florida at Miami. He is an organizing committee member of INTERMAG Conference '93 and Southeastcon '94. Since 1988 he has published 14 archival journal and 20 proceedings papers in the areas of neural networks, optical computing, image processing, signal processing, optimal design of electromagnetic devices, and inverse problems.