Neuromorphic Learning VLSI Systems: A Survey

Gert Cauwenberghs
Electrical and Computer Engineering
Johns Hopkins University, Baltimore, MD 21218
E-mail: gert@bach.ece.jhu.edu

Abstract

This [chapter] reviews advances in hardware learning and adaptation in synthetic neural systems. Over the last decade, research in the field has intensified, drawing inspiration across several science and engineering disciplines. This review briefly covers neural models, implementation technology, architectural constraints, and system applications of learning in hardware.

1 Introduction

Carver Mead introduced "neuromorphic engineering" [1] as an interdisciplinary approach to the design of biologically inspired neural information processing systems, whereby neurophysiological models of perception and information processing in biological systems are mapped onto analog VLSI systems that not only emulate their functions but also resemble their structure [2]. The motivation for emulating neural function and structure in analog VLSI is the realization that challenging tasks of perception, classification, association and control successfully performed by living organisms can only be accomplished in artificial systems by using an implementation medium that matches their structure and organization.

Essential to neuromorphic systems are mechanisms of adaptation and learning, modeled after the "plasticity" of synapses and neural structure in biological systems [3],[4]. Learning can be broadly defined as a special case of adaptation whereby past experience is used effectively in readjusting the system response to previously unseen, although similar, stimuli. Based on the nature and availability of a training feedback signal, learning algorithms for artificial neural networks fall under three broad categories: unsupervised, supervised and reward/punishment (reinforcement). Physiological experiments have revealed plasticity mechanisms in biology that correspond to Hebbian unsupervised learning [5], and classical (pavlovian) conditioning [6],[7] characteristic of reinforcement learning.

Mechanisms of adaptation and learning also provide a means to compensate for analog imperfections in the physical implementation of neuromorphic systems, and fluctuations and uncertainties in the environment in which it operates. To this end, it is crucial that the learning be continuously performed on the system in operation. This enables the system to be functionally self-contained, and to adapt continuously to the environment in which they operate.

For neuromorphic systems which involve a large number of parameters such as synapses in a densely connected neural network, it is imperative that the learning functions be an integral part of the hardware, implemented locally and interfacing directly with the synaptic functions.

Practical limits of integrated implementation of learning functions are imposed by the degree of locality implied by the learning rule, and the available memory bandwidth and fanout provided by the technology. This is an important point to consider in the design, and determines whether an electronic, optical, or hybrid implementation is most suited for the targeted application. A very important consideration as well is the need for locally storing the analog or digital parameter values, to retain the information being extracted during learning. Not surprisingly, technological issues of adaptation and memory are directly related, and both need to be addressed concurrently.

A vast research literature is dedicated to various styles of neural hardware implementations with provisions for learning, some of it with integrated learning functions. A selection of the literature (which is bound to be incomplete even at the time of printing!) is included in the list of references below. Some examples of early implementations of neural systems with integrated adaptation and learning functions can be found in edited volumes such as [8], [9] and [10], in conference proceedings such as NIPS, IJCNN (ICNN/WCNN) and ISCAS, and in special and regular issues of journals such as IEEE Transactions on Neural Networks (May 1992 and 1993 [12]), IEEE Micro (Micro-Neuro special issues) and Kluwer's International Journal of Analog Integrated Circuits and Signal Processing [13, 14]. The exposition below will serve as a brief description of a limited cross-section of research in the field over the last decade (mainly focusing on analog VLSI systems), as well as a general coverage of the important issues.

2 Adaptation and Learning

Definitions for the terms adaptation and learning come in several varieties, differing with the particular discipline in which it is formulated, such as cognitive science, neuroscience, neural computation, artificial intelligence, information theory, and control theory.

From a system level perspective, a general framework for adaptation and learning is depicted in Figure 1 [17]. A system with adjustable parameters p_i (vector $\mathbf{p}$) interacts with the environment through sensory inputs and activation outputs. An adaptive element, either internal or external to the system, adjusts the parameters of the system to "optimize" some performance index that is either defined or implied in relation to the system and its interaction with the environment. In most models of learning and adaptation, the measure of performance is quantified either as an error index $\mathcal{E}(\mathbf{p})$ which needs to be minimized:

$$\mathbf{p} = \text{argmin } \mathcal{E}(\mathbf{p}) \tag{1}$$

or, equivalently, a quality index which needs to be maximized. The optimization is subject to suitable constraints that have to do with physical limits on the system as well as other requirements on the system and the way it interacts with the environment.

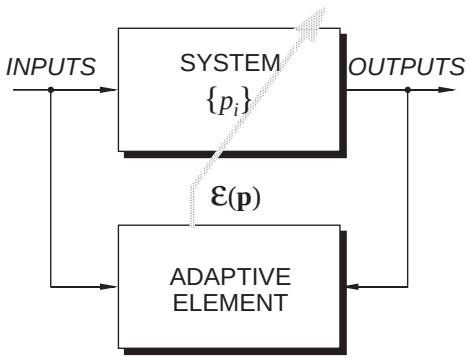

Figure 1: Adaptation and learning in an information processing system by adjusting the analog system parameters p_i to optimize a performance index $\mathcal{E}(\mathbf{p})$. The system interacts with the environment through its sensory inputs and activation outputs.

What distinguishes learning from more general forms of adaptation is the way in which the system uses past experience in trying to respond effectively to previously unseen, although similar, input stimuli. The distinct objective in learning is to *generalize* beyond the specifics of the presented input samples, and minimize the *expected* value of $\mathcal{E}(\mathbf{p})$ from the underlying statistics of the training samples:

$$\mathbf{p} = \text{argmin } E(\mathcal{E}(\mathbf{p})) \ . \tag{2}$$

Based on the nature and availability of a training feedback signal in the formulation of $E(\mathcal{E}(\mathbf{p}))$, learning algorithms for artificial neural networks (ANNs) fall under three broad categories: supervised [19], unsupervised [26], and reward/punishment (reinforcement) [33].

Supervised Learning [18]-[23] assumes that a "teacher" is continuously available to produce target values $y_k^{\text{target}}(t)$ for the outputs $y_k(t)$, whereby the (instantaneous) error index is quantified as the distance between actual and target outputs

$$\mathcal{E}(\mathbf{p}; t) = \sum_k |y_k^{\text{target}}(t) - y_k(t)|^\nu \quad , \tag{3}$$

using a distance metric with norm $\nu > 0$. Supervised learning is in a sense the easiest case of learning to implement, since the learning task is well defined and the performance index, directly quantified in terms of the target training signals, can be evaluated and optimized on-line. The most popular of all learning algorithms is backpropagation [20], which is effectively the chain rule of differentiation applied to gradient descent of (3) on a multilayer feedforward ANN, and which can be extended to more general feedforward structures [19], and to more complex structures with recurrent dynamics in the state variables [22, 23]. A system example of supervised learning in VLSI with recurrent dynamics is presented in [the next chapter].

Unsupervised learning [24]-[29] does not assume any feedback from an external teacher, and attempts to classify inputs based on the underlying statistics of the data. Classifiers

of this type are intended for tasks which inherently require some form of data compression or an otherwise more suitable data representation for subsequent information processing. The criterion for adjusting the boundaries between classes can be expressed in information-theoretic terms, either as to maximize the mutual information between the analog inputs and the discrete output classes [28], or (equivalently) to minimize the average description length of the output classes [30]. Typical unsupervised learning techniques include Hebbian learning [24] in a self-organizing neural network, auto-associative memories [26, 25], k-means clustering in a vector quantizer [27], and adaptive resonance theory [29]. A VLSI learning binary vector quantizer is described in [one of the following chapters].

Reinforcement learning [32]-[38] assumes the available external feedback on system performance is limited to discrete-time, delayed rewards and punishments, without a target specified for the system outputs. The difficulty in this type of learning is the assignment of proper credit to responsible actions in the past leading to the system failure or success indicated by the penalty or reward. Algorithms of the reinforcement learning type use internal mechanisms of credit assignment which make no prior assumptions on the causal relationships of the system and the enviroment in which it operates. Closely related to models of learning in artificial intelligence [31, 39], they include "time difference learning" or TD(λ) [35] as a special case of [33], Q-learning [38] using a value function on the state space for optimal policy iteration, and "advanced heuristic dynamic programming" [36] using vectorial gradient information for increased speed of convergence. Details on reinforcement learning system in analog VLSI and a system example are given in [the next chapter].

Hybrids: Unsupervised and supervised learning approaches can be combined in many ways with various networks architectures to generate internally self-organizing adaptive hetero-associative systems. This synthesis reaches beyond neural nets in the restricted sense of what is conventionally known as ANNs, and includes fuzzy neural systems [40, 41, 42] as well as "hierarchical mixture of experts" models trained with the expectation-maximization algorithm [43]. In both cases, internal structure is learned using unsupervised clustering techniques based on the input statistics, and the output structure is trained through (gradient-based and other) supervised learning.

3 Technology

Biological neural systems are built out of "wetware" components in an implementation medium which is necessarily different from technologies available to the implementation of artificial computing systems, such as semiconductors and optical propagation media. The neuromorphic engineering approach extends the functionality and structure of biological systems to artificial systems built with components and architectures that closely resemble their biological counterparts at all levels, transparent to differences in technology. Still, the physical limits

on size, density and connectivity depend strongly on the technology used.

Most neural hardware implementations use VLSI technology, which is functionally highly versatile but mostly restricted to two dimensions. The planar nature of VLSI technology is not necessarily a restriction for neural implementations since neural structures such as in the cerebral cortex are mostly two-dimensional as well— after all the brain is itself a folded 2-D structure. Optical free-space interconnects, on the other hand, allow synaptic densities presently unavailable in state-of-the-art VLSI technology. Hybrid opto-electronic systems combine the technological advantages of both worlds, with functionally rich local VLSI processing and global optical interconnects.

For learning and adaptation, a central issue in all implementation technologies is the local storage of synaptic parameters. This issue, together with the means of incrementally adapting the stored parameters, is addressed below in particular detail. For brevity, the exposition focuses mainly on electronic implementations in analog VLSI technology.

3.1 VLSI Subthreshold MOS Technology

MOS transistors operating in the subthreshold region [46] are attractive for use in medium-speed, medium-accuracy analog VLSI processing, because of the low current levels and the exponential current-voltage characteristics that span a wide dynamic range of currents [47] (roughly from 100 fA to 100 nA for a square device in 2 μm CMOS technology at room temperature). Subthreshold MOS transistors provide a clear "neuromorph" [1], since their exponential I-V characteristics closely resemble the carrier transport though cell membranes in biological neural systems, as governed by the same Boltzman statistics [45]. The exponential characteristics provide a variety of subthreshold MOS circuit topologies that serve as useful computational primitives (such as nonlinear conductances, sigmoid nonlinearities, etc.) for compact analog VLSI implementation of neural systems [2]. Of particular interest are translinear subthreshold MOS circuits, derived from similar bipolar circuits [47]. They are based on the exponential nature of current-voltage relationships, and offer attractive compact implementations of product and division operations in VLSI.

3.2 Adaptation and Memory

Learning in analog VLSI systems is inherently coupled with the problem of storage of analog information, since after learning it is most often desirable to retain the learned weights for an extended period of time. The same is true for biological neural systems, and mechanisms of plasticity for short-term and long-term synaptic storage are not yet clearly understood.

In VLSI, analog weights are conveniently stored as charge or voltage on a capacitor. A capacitive memory is generically depicted in Figure 2. The stored weight charge is preserved when brought in contact with the gate of an MOS transistor, which serves as a buffer between weight storage and the implementation of the synaptic function. An adaptive element in

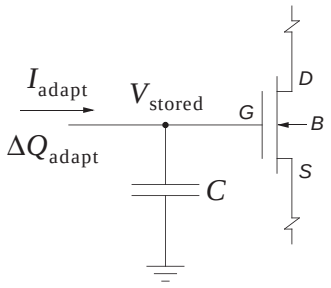

Figure 2: Adaptation and memory in analog VLSI: storage cell with charge buffer.

contact with the capacitor updates the stored weight in the form of discrete charge increments

$$V_{\text{stored}}(t + \Delta t) = V_{\text{stored}}(t) + \frac{1}{C}\Delta Q_{\text{adapt}}(t) \tag{4}$$

or, equivalently, a continuous current supplying a derivative

$$\frac{d}{dt}V_{\text{stored}}(t) = \frac{1}{C}I_{\text{adapt}}(t) \tag{5}$$

where $\Delta Q_{\text{adapt}}(t) = \int_t^{t+\Delta t} I_{\text{adapt}}(t')dt'$.

On itself, a floating gate capacitor is a near-perfect memory. However, leakage and spontaneous decay of the weights result when the capacitor is in volatile contact with the adaptive element, such as through drain or source terminals of MOS transistors. This distinguishes volatile from non-volatile storage VLSI technology. An excellent review of analog memories for neural computation is given in [48].

Non-volatile Memories [49]-[60] contain adaptive elements that interface with the floating gate capacitor by capacitive coupling across an insulating oxide. In standard CMOS VLSI technologies, charge transport through the oxide is typically controlled by tunneling [84, 134, 49, 51], hot electron injection [59] or UV-activated conduction [179, 52, 56, 74]. Flash memories offer fast adaptation rates (msecs) and long retention times (years) without the need for high programming voltages or UV light, but are not standardly available in CMOS processes.

Volatile Memories [49],[61]-[66] offer fast adaptation rates and instantaneous reprogramming of the parameter values, using a voltage-controlled ohmic connection to the capacitor in the form of MOS switches and switched current sources. A leakage current results from the reverse diode formed between source and drain diffusions and bulk connection of a switch transistor. The leakage typically resticts the retention time of the memory to the msec range, adequate for short-term storage. An active refresh mechanism is required for long-term storage [49],[62]-[64]. An adaptive element which combines active refresh storage and incremental adaptation, and which allows a random-access read and write digital interface, is described in [the next chapter].

Other implementations frequently use local or external digital storage of the parameters, combined with either local or multiplexed D/A conversion. This solution is less attractive for large-scale neural processors with local learning functions that require incremental adaptation of the parameters, since then the increment would need to be performed in digital as well. Both volatile and non-volatile analog memories allow incremental updates in direct analog format, according to (4) or (5).

The non-volatile solution is more attractive than volatile alternatives when long-term storage is a more pressing concern than speed of adaptation and flexibility of programming. The volatile scheme is particularly useful in multiplexed hardware implementations for multi-purpose applications or to realize virtual larger-scale systems, requiring frequent reloading of large blocks of partial weight matrices. This could be done with an external digital cache memory and an array of A/D/A converters for bi-directional digital read and write access to the synaptic array [65]. Random-access memory addressing in digital format is on itself a valuable feature for system-level interfacing.

3.3 Emerging Technologies

Innovation and continued progress in information technology benefits the design of learning neural systems of larger size and better performance, as it benefits other information processing systems. Some relatively new developments in VLSI include micro-electromechanical systems (MEMS) [67], wafer-scale integration [141, 143], chip-scale packaging [68], and silicon-on-insulator (SOI) integrated circuit fabrication [69, 70]. The latter is of special interest to analog storage, because significant reduction of leakage currents due to bulk reverse diodes in MOS switches allows longer retention times of capacitive memories.

Continued technology developments in optical and optoelectronic information processing in combination with mature VLSI technology hold the potential for significant performance improvements in artificial neural information processing systems [150]-[158], promising massive inter-chip connectivity as needed for larger size neural networks. High-density optical storage and adaptation for integrated learning could be achieved in 3-D optical media such as photo-refractive crystals.

4 Architecture

Learning algorithms that are efficiently implemented on general-purpose digital computers do not necessarily map efficiently onto analog VLSI hardware. The good news is that the converse is also true, as it is well known that special-purpose processors tuned to a given task easily outperform most general-purpose computing engines, on that particular task. From the perspective of computational efficiency, it is therefore important to closely coordinate the design of algorithms and corresponding VLSI architecture to ensure an optimal match.

Important guidelines in efficiency of computation dictate the usual principles commonly taught in modern VLSI design: locality, scalability, and parallelism. The principle of locality

confines intensive computations to the cell level, and restricts global operations to nearest-neighbor interactions. In addition, certain scalar global operations which can be easily performed with a single common wire in analog VLSI technology are allowed, such as global summing of currents or charges, and global communication of voltage-coded variables. Scalability implies that the implemented algorithms cannot scale stronger than second order in a linear parameter such as the number of neurons, since nothing more complex than a 2-D array can be implemented on an extended scale in planar VLSI technology. Parallelism in this context implies that the number of operations performed concurrently at any given time scales linearly with the number of cells.

Even if the learning algorithm supports a parallel and scalable architecture suitable for analog VLSI implementation, inaccuracies in the implementation of the learning functions may significantly affect the performance of the trained system. Neuromorphic principles call for a *distributed* architecture not only for the network of neurons but also to implement the learning functions, robust to localized errors in the implementation.

4.1 Incremental Outer-Product Learning in Distributed Systems

For networks with distributed neurons such as linear and multilayer perceptrons [20]

$$x_i = f(\sum_j p_{ij} x_j) \tag{6}$$

gradient descent of an LMS error functional $\mathcal{E}$ defined on the output neurons x_k^{out} gives rise to incremental outer-product learning rules of the form

$$\Delta p_{ij} = \eta \; x_j e_i \tag{7}$$

where the backpropagation of the error variables e_i is derived by application of the chain rule for differentiation as [19]

$$\begin{aligned} e_k^{\text{out}} &= -f_k' \, \frac{\partial \mathcal{E}}{\partial x_k^{\text{out}}} \\ e_j &= f_j' \, \sum_i p_{ij} e_i \end{aligned} \tag{8}$$

where f_j' denotes the derivative of the function $f(.)$ evaluated at its argument in (6). Outer-product rules of the form (7) are *local*: synaptic updates are constructed from intersecting variables at the location of the synapses. The general class of learning algorithms of the incremental outer-product type include

Supervised Learning: the delta rule [18] and backpropagation [20] for supervised learning in linear or multilayer feedforward perceptrons with a functional (3). Also included, with stochastic rather than deterministic neurons, are Boltzman learning in networks of stochastic neurons [21, 71], and pulse firing neural nets [89].

Unsupervised Learning: hebbian learning [24], where $e_i = f_i' x_i$ corresponding to a functional $\mathcal{E} \equiv -\sum_i {x_i}^2$. The k-means clustering algorithm for learning vector quantization (LVQ) [27] is a special case of the latter, where the nonlinearity in the output layer f_k selects a single winner across all outputs k. Kohonen topology-preserving maps [26] further include the neighbors of the winner $k \pm 1$ into the learning updates. Learning in ART networks [29] also fits in this category although it is slightly more involved. Learning in Hopfield networks [25] is hebbian in slightly modified form.

Hybrids and Variants: fuzzy maps, hetero-associative neural networks, radial basis networks, etc. which conform to the general structure of Eqns. (6)-(9) and their variants and combinations.

Reinforcement Learning: The reinforcement learning updates for both the action network and the adaptive critic in [33] are of the general incremental outer-product form (7), although modulated with a global (common) reinforcement signal, and low-pass filtered for credit assignment back in time. See [the next chapter] for more details on the equivalent gradient-descent outer-product formulation. An outer-product VLSI implementation is described in [116].

Since all of the above learning algorithm share essentially the same incremental outer-product learning rule, they can be cast into the same general VLSI architecture depicted in Figure 3. Clearly, this architecture exhibits the desirable properties of locality, parallelism and scalability. Forward and backward signal paths x_j and e_i traverse in horizontal and vertical directions through the array of synapse cells p_{ij}. The neuron nonlinearity $f(.)$ and its derivative are implemented at the output periphery the array. Several layers of this structure can be cascaded in alternating horizontal and vertical directions to form multi-layer perceptrons. The array architecture of Figure 3 (b) forms the basis for many of the implemented VLSI learning systems [71]-[103]. One example, described in [74], arguably contains the densest VLSI array for general outer-product learning developed to date, using only two transistors for synapse and learning operations per cell. An array of single-transistor learning synapses for certain classes of learning is presented in [59].

Digital VLSI implementations [139]-[149] differ from the analog architecture mainly in that contributions to the summations in (6) and (9) cannot be accumulated onto a single line. Global summations are most commonly implemented using a systolic array architecture.

4.2 Localized Outer-Product Learning in Cellular Neural Systems

Notice that the fully interconnected architecture of Figure 3 (b) becomes inefficient when the network that it implements has sparse connectivity. A limiting case of sparsely interconnected networks are cellular neural networks [105], in which neurons only interact with their immediate neighbors, conveniently arranged on a 2-D grid. Since the synaptic connections in networks of this type are only peripheral, the implementation architecture is determined directly by the topology of the neurons in relation with their neighbors. The synapse and learning functions

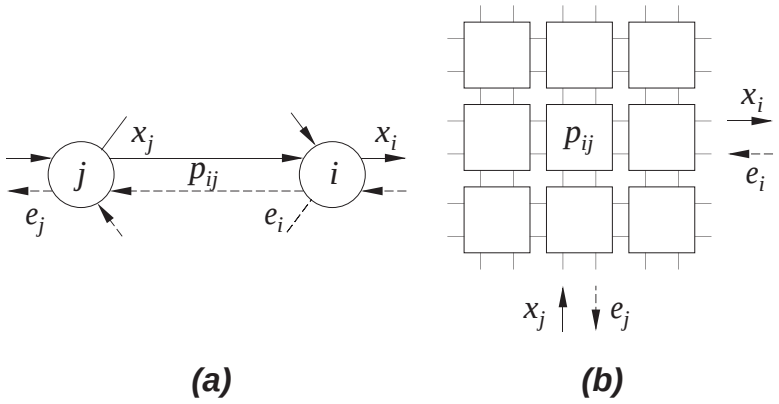

Figure 3: Incremental outer-product learning. (a) Feedforward and backpropagation model; (b) Simplified VLSI architecture.

are integrated at the neuron level, rather than distributed over an array as in Figure 3 (b). Other than that, the same principles hold, and rules of the outer-product type as illustrated in Figure 3 (a) are implemented locally at the neuron inter-cell level [104]-[107].

4.3 Model-Free Learning Approaches

Although model-based approaches for learning such as the outerproduct learning models described above are fairly robust to mismatches in the implementation of the learning functions owing to their distributed architecture [118, 119, 122, 123], the same can not be said *a priori* of more general classes of learning which do not fit the outerproduct type. This is particularly so for recurrent neural networks with hidden internal dynamics for which learning complexity rises sharply with the number of parameters [22, 23], or for more complex systems of which a model is difficult to derive or unknown to the learning element. Model-free approaches to learning [124] do not assume a particular model for the system nor the environment in which it operates, and derive parameter updates Δp_i by physically probing the dependency of the performance index $\mathcal{E}$ on the parameters p_i through perturbations π_i on the parameters.

The term "model-free" pertains to the learning, and not necessarily to the structure of the system itself being adapted, which can be anything and which clearly is parametric. The main advantage of model-free learning is that it leaves tremendous freedom in configuring the system, which is allowed to change structurally on-line as learning progresses, without the need to compile models. This is particularly useful for training reconfigurable architectures [135, 111]. The insensitivity of learning performance to inaccuracies in the implemented system, and the ability to learn systems with intractible models, are direct benefits of model-free learning. An additional benefit of stochastic perturbative learning approaches seems to be that the synaptic noise thus introduced improves generalization performance of the learned system [120].

Variants on perturbative model-free learning use some limited model information to train feedforward multilayer ANNs more effectively [128, 131, 133]. The question of how much

model information can be reliably used is important, although truly model-free approaches are most generally applicable and expandable, and their performance does not significantly suffer from the lack of complete gradient information on the error $\mathcal{E}$ as some asymptotic theory establishes [126].

The model-free nature of learning applies to general learning tasks beyond the traditionally supervised and unsupervised, and can be extended to reinforcement learning. An extensive study of model-free supervised and reinforcement learning architectures with examples of analog VLSI systems is the subject of [the next chapter].

5 Systems

Several examples of adaptive and/or learning VLSI systems with applications in vision, speech, signal processing, pattern recognition, communications, control and physics are included in the references [170]-[202]. This list is by no means complete, and the spectrum of applications will likely expand as the new application areas are discovered and research advances create new ways of using adaptation and learning in the design of intelligent neuromorphic information processing systems.

Covering such diverse range of disciplines across neurobiology, artificial intelligence, cognitive science, information theory, etc., research on learning systems is bound to develop further as different concepts and experimental evidence combine to bridge the gap between bottom-up and top-down modeling approaches, towards the engineering of truly intelligent autonomous learning systems, and towards a better understanding of learning mechanisms in biological neural systems at different levels of abstraction.

References

[1] C.A. Mead, "Neuromorphic Electronic Systems," *Proceedings of the IEEE,* vol. **78** (10), pp 1629-1639, 1990.

[2] C.A. Mead, *Analog VLSI and Neural Systems,* Reading, MA: Addison-Wesley, 1989.

6.1 Neurobiological Inspiration

[3] G.M. Shepherd, *The Synaptic Organization of the Brain,* 3rd ed., New York, NY: Oxford Univ. Press, 1992.

[4] P.S. Churchland and T.J. Sejnowski, *The Computational Brain,* Cambridge, MA: MIT Press, 1990.

[5] S.R. Kelso and T.H. Brown, "Differential Conditioning of Associative Synaptic Enhancement in Hippocampal Brain Slices," *Science*, vol. **232**, pp 85-87, 1986.

[6] R.D. Hawkins, T.W. Abrams, T.J. Carew, and E.R. Kandell, "A Cellular Mechanism of Classical Conditioning in *Aplysia*: Activity-Dependent Amplification of Presynaptic Facilitation," *Science,* iss. **219**, pp 400-405, 1983.

[7] P.R. Montague, P. Dayan, C. Person and T.J. Sejnowski, "Bee Foraging in Uncertain Environments Using Predictive Hebbian Learning," *Nature*, vol. **377** (6551), pp. 725-728, 1996.

6.2 Edited Book Volumes, Journal Issues and Reviews

[8] C.A. Mead and M. Ismail, Eds., *Analog VLSI Implementation of Neural Systems,* Norwell, MA: Kluwer, 1989.

[9] N. Morgan, Ed., *Artificial Neural Networks: Electronic Implementations,* CA, Los Alamitos: IEEE Computer Society Press, 1990.

[10] E. Sanchez-Sinencio and C. Lau, Eds., *Artificial Neural Networks: Paradigms, Applications, and Hardware Implementations*, IEEE Press, 1992.

[11] M.A. Jabri, R.J. Coggins and B.G. Flower, "Adaptive Analog VLSI Neural Systems," London, UK: Chapman Hall, 1996.

[12] E. Sanchez-Sinencio and R. Newcomb, Eds., Special Issues on Neural Network Hardware, *IEEE Transactions on Neural Networks,* vol. **3** (3), 1992; and vol. **4** (3), 1993.

[13] T.S. Lande, Ed., Special Issue on Neuromorphic Engineering, *Int. J. Analog Int. Circ. Signal Proc.,* March 1997.

[14] G. Cauwenberghs, M. Bayoumi and E. Sanchez-Sinencio, Eds., Special Issue on Learning in Silicon, to appear in *Int. J. Analog Int. Circ. Signal Proc.*

[15] "Learning on Silicon," special session, *Proc. Int. Symp. Circuits and Systems,* Hong Kong, June 1997.

[16] H.P. Graf and L.D. Jackel, "Analog Electronic Neural Network Circuits," *IEEE Circuits and Devices Mag.*,vol. **5**, pp 44-49, 1989.

[17] G. Cauwenberghs, "Adaptation, Learning and Storage in Analog VLSI," in *Proceedings of the Ninth Annual IEEE International ASIC Conference,* Rochester, NY, Sept. 23-27, 1996, pp 273-278.

6.3 Learning Models

6.3.1 Supervised Learning

[18] B. Widrow and M.E. Hoff, "Adaptive Switching Circuits," in *IRE WESCON Convention Record*, vol. **4**, pp 96-104, 1960.

[19] P. Werbos, *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences.* Ph.D. dissertation, 1974. Reprinted in P. Werbos, *The Roots of Backpropagation.* New York: Wiley, 1993.

[20] D.E. Rumelhart, G.E. Hinton and R.J. Williams, "Learning Internal Representations by Error Propagation," in D.E. Rumelhart and J.L. McClelland, Eds., *Parallel Distributed Processing, Explorations in the Microstructure of Cognition,* vol. **1**, chapter 8, Cambridge, MA: MIT Press, 1986.

[21] G.E. Hinton and T.J. Sejnowski, "Learning and Relearning in Boltzman Machines," in D.E. Rumelhart and J.L. McClelland, Eds., *Parallel Distributed Processing, Explorations in the Microstructure of Cognition,* vol. **1**, chapter 7, Cambridge, MA: MIT Press, 1986.

[22] R.J. Williams and D. Zipser, "A Learning Algorithm for Continually Running Fully Recurrent Neural Networks," *Neural Computation,* vol. **1** (2), pp 270-280, 1989.

[23] B.A. Pearlmutter, "Learning State Space Trajectories in Recurrent Neural Networks," *Neural Computation,* vol. **1** (2), pp 263-269, 1989.

6.3.2 Unsupervised Learning

[24] D.O. Hebb, *The Organization of Behavior,* New York, NY: Wiley, 1949.

[25] J. Hopfield, "Neural Networks and Physical Systems with Emergent Collective Computational Abilities," *Proc. Natl. Acad. Sci.,* vol. **97**, pp 2554-2558, 1982.

[26] T. Kohonen, "Self-Organisation and Associative Memory," Berlin: Springer-Verlag, 1984.

[27] A. Gersho and R.M. Gray, "Vector Quantization and Signal Compression," Norwell, MA: Kluwer, 1992.

[28] , R. Linsker, "Self-Organization in a Perceptual Network," *IEEE Computer,* vol. **21**, pp 105-117, 1988.

[29] G.A. Carpenter, "Neural Network Models for Pattern-Recognition and Associative Memory," *Neural Networks,* vol. **2** (4), pp 243-257, 1989.

[30] C.M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, 1995.

6.3.3 Reinforcement Learning and Related Models

[31] K.S. Narendra and M.A.L. Thatachar, "Learning Automata—A Survey," IEEE T. Syst. Man and Cybern., vol. **SMC-4**, pp. 323-334, 1974.

[32] S. Grossberg, "A Neural Model of Attention, Reinforcement, and Discrimination Learning," *International Review of Neurobiology,* vol. **18**, pp 263-327, 1975.

[33] A.G. Barto, R.S. Sutton, and C.W. Anderson, "Neuronlike Adaptive Elements That Can Solve Difficult Learning Control Problems," *IEEE Transactions on Systems, Man, and Cybernetics,* vol. **13** (5), pp 834-846, 1983.

[34] S. Grossberg and D.S. Levine, "Neural Dynamics of Attentionally Modulated Pavlovian Conditioning: Blocking, Inter-Stimulus Interval, and Secondary Reinforcement," *Applied Optics,* vol. **26**, pp 5015-5030, 1987.

[35] R.S. Sutton, "Learning to Predict by the Methods of Temporal Differences," *Machine Learning,* vol. **3**, pp 9-44, 1988.

[36] P.J. Werbos, "A Menu of Designs for Reinforcement Learning Over Time," in *Neural Networks for Control,* W.T. Miller, R.S. Sutton and P.J. Werbos, Eds., Cambridge, MA: MIT Press, 1990, pp 67-95.

[37] W.T. Miller, R. Sutton, and P. Werbos (eds.), *Neural Networks for Control.* Cambridge, MA: MIT Press, 1990.

[38] C. Watkins and P. Dayan, "Q-Learning," *Machine Learning,* vol. **8**, pp 279-292, 1992.

[39] W.-M. Shen, *Autonomous Learning from the Environment*, New York, NY: Freeman, Computer Science Press, 1994.

6.3.4 Hybrid Learning Approaches

[40] G.A. Carpenter, et al, "Fuzzy ARTMAP - A Neural Network Architecture for Incremental Supervised Learning of Analog Multidimentional Maps," *IEEE Transactions on Neural Networks,* vol. **3** (5), pp. 698-713, 1992.

[41] D. White and D. Sofge, Eds, "*Handbook of Intelligent Control: Neural, Adaptive and Fuzzy Approaches.*" New York: Van Nostrand, 1992.

[42] P.J. Werbos, "Neurocontrol and Elastic Fuzzy Logic: Capabilities, Concepts, and Applications," *IEEE Transactions on Industrial Electronics,*, vol. **40** (2), pp. 170-180, 1993.

[43] M. Jordan and R. Jacobs, "Hierarchical Mixtures of Experts and the EM Algorithm," *Neural Computation*, vol. **6**, pp 181-214, 1994.

[44] R.M. Sanner and J.J.E. Slotine, "Gaussian Networks for Direct Adaptive Control," *IEEE Transactions on Neural Networks,* vol. **3** (6), pp. 837-864, 1992.

6.4 Technology

6.4.1 Subthreshold MOS Operation

[45] A.L. Hodgkin and A.F. Huxley, "Current Carried by Sodium and Potassium Ions Through the Membrane of the Giant Axon of *Loligo*," *Journal of Physiology,* vol. **116**, pp. 449-472, 1952.

[46] E. Vittoz and J. Fellrath, "CMOS Analog Integrated Circuits Based on Weak Inversion Operation," *IEEE Journal on Solid-State Circuits,* vol. **12** (3), pp 224-231, 1977.

[47] A.G. Andreou, K.A. Boahen, P.O. Pouliquen, A. Pavasovic, R.E. Jenkins, and K. Strohbehn, "Current-Mode Subthreshold MOS Circuits for Analog VLSI Neural Systems," *IEEE Transactions on Neural Networks,* vol. **2** (2), pp 205-213, 1991.

6.4.2 Analog Storage

[48] Y. Horio, and S. Nakamura, "Analog Memories for VLSI Neurocomputing," in *Artificial Neural Networks: Paradigms, Applications, and Hardware Implementations*, E. Sanchez-Sinencio and C. Lau, Eds., IEEE Press, 1992, pp 344-363.

[49] E. Vittoz, H. Oguey, M.A. Maher, O. Nys, E. Dijkstra, and M. Chevroulet, "Analog Storage of Adjustable Synaptic Weights," in *VLSI Design of Neural Networks*, Norwell MA: Kluwer Academic, pp 47-63, 1991.

[50] M.A. Holler, "VLSI Implementations of Learning and Memory Systems," in *Advances in Neural Information Processing Systems,* San Mateo, CA: Morgan Kaufman, vol. **3**, pp 993-1000, 1991.

Non-Volatile Analog Storage

[51] A. Kramer, C.K. Sin, R. Chu and P.K. Ko, "Compact EEPROM-based Weight Functions," in *Advances in Neural Information Processing Systems,* San Mateo, CA: Morgan Kaufman, vol. **3**, pp 1001-1007, 1991.

[52] D.A. Kerns, J.E. Tanner, M.A. Sivilotti and J. Luo, "CMOS UV-Writable Non-Volatile Analog Storage," in *Proc. Advanced Research in VLSI Int. Conf.*, Santa Cruz CA, 1991.

[53] A. Soennecken, U. Hilleringmann and K. Goser, "Floating Gate Structures As Nonvolatile Analog Memory Cells in 1.0um-LOCOS-CMOS Technology with PZT Dielectrica," *Microel Eng*, vol. **15** (1-4), pp 633-636, 1991.

[54] B.W. Lee, B.J. Sheu and H. Yang, "Analog Floating-Gate Synapses for General-Purpose VLSI Neural Computation," *IEEE Trans. Circuits and Systems*, vol. **38**, pp 654-658, 1991.

[55] D.A. Durfee and F.S. Shoucair, "Low Programming Voltage Floating Gate Analog Memory Cells in Standard VLSI CMOS Technology," *Electronics Letters*, vol. **28** (10), pp 925-927, May 7, 1992.

[56] R.G. Benson, Ph.D. Dissertation, California Institute of Technology, 1993.

[57] O. Fujita and Y. Amemiya, "A Floating-Gate Analog Memory Device for Neural Networks," *IEEE Device*, vol. **40** (11), pp 2029-2055, Nov. 1993.

[58] A. Thomsen and M.A. Brooke, "Low Control Voltage Programming of Floating-Gate Mosfets and Applications," *IEEE Circ I*, vol. **41** (6), pp 443-452, June 1994.

[59] P. Hasler, C. Diorio, B. Minch and C.A. Mead, "Single Transistor Learning Synapses," in *Advances in Neural Information Processing Systems,* Cambridge, MA: MIT Press, vol. **7**, 1995.

[60] H. Won, Y. Hayakawa, K. Nakajima and Y. Sawada, "Switched Diffusion Analog Memory for Neural Networks with Hebbian Learning-Function and Its Linear-Operation," *IEICE T. Fund. El. Comm. Comp. Sci.d Elect Commun Comp Sci*, vol. **E79A** (6), pp 746-751, June 1996.

Volatile Analog Storage and Refresh

[61] D.B. Schwartz, R.E. Howard and W.E. Hubbard, "A Programmable Analog Neural Network Chip," *IEEE J. Solid-State Circuits*, vol. **24**, pp 313-319, 1989.

[62] B. Hochet, V. Peiris, S. Abdot, and M.J. Declercq, "Implementation of a Learning Kohonen Neuron Based on a New Multilevel Storage Technique," *IEEE J. Solid-State Circuits,* vol. **26**, pp 262-267, 1991.

[63] R. Castello, D.D. Caviglia, M. Franciotta and F. Montecchi, "Selfrefreshing Analog Memory Cell for Variable Synaptic Weights," *Electronics Letters*, vol. **27** (20), pp 1871-1873, 1991.

[64] G. Cauwenberghs, and A. Yariv, "Fault-Tolerant Dynamic Multi-Level Storage in Analog VLSI," *IEEE Transactions on Circuits and Systems II,* vol. **41** (12), pp 827-829, 1994.

[65] G. Cauwenberghs, "A Micropower CMOS Algorithmic A/D/A Converter," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. **42** (11), pp 913-919, 1995.

[66] J. Elias, D.P.M. Northmore and W. Westerman, "An Analog Memory Circuit for Spiking Silicon Circuits," *Neural Computation,* vol. **9** (2), pp 419-440, 1997.

6.4.3 Emerging VLSI Technologies

[67] B. Gupta, R. Goodman, F. Jiang, Y.C. Tai, S. Tung and C.M. Ho, "Analog VLSI System for Active Drag Reduction," *IEEE Micro Mag.*, vol. **16** (5), pp 53-59, Oct. 1996.

[68] T. Distefano and J. Fjelstad, "Chip-Scale Packaging Meets Future Design Needs," *Solid State Tech.*, vol. **39** (4), p 82, Apr. 1996.

[69] B. Elkareh, B. Chen and T. Stanley, "Silicon-On-Insulator - an Emerging High-Leverage Technology," *IEEE T. Comp. Pack. Man. Techn. Part A*, vol. **18** (1), pp 187-194, March 1995.

[70] C.M. Hu, "SOI (Silicon-On-Insulator) for High-Speed Ultra Large-Scale Integration," *Japan JAP 1*, vol. **33** (1B), pp 365-369, Jan. 1994.

6.5 Architecture

6.5.1 Outer-Product Supervised Learning

[71] J. Alspector, B. Gupta, and R.B. Allen, "Performance of a Stochastic Learning Microchip," in *Advances in Neural Information Processing Systems,* San Mateo, CA: Morgan Kaufman, vol. **1**, pp 748-760, 1989.

[72] F.M.A. Salam and Y.W. Wang, "A Real-Time Experiment Using a 50-Neuron CMOS Analog Silicon Chip with On-Chip Digital Learning," *IEEE T. Neural Networks*, vol. **2** (4), pp 461-464, 1991.

[73] C.R. Schneider and H.C. Card, "CMOS Mean Field Learning," *Electronics Letters*, vol. **27** (19), pp 1702-1704, 1991.

[74] G. Cauwenberghs, C.F. Neugebauer, and A. Yariv, "Analysis and Verification of an Analog VLSI Outer-Product Incremental Learning System," *IEEE Transactions on Neural Networks,* vol. **3** (3), pp 488-497, 1992.

[75] S.P. Eberhardt, R. Tawel, T.X. Brown, T. Daud and A.P. Thakoor, "Analog VLSI Neural Networks - Implementation Issues and Examples in Optimization and Supervised Learning," *IEEE T. Ind. El.*, vol. **39** (6), pp 552-564, Dec. 1992.

[76] Y. Arima, M. Murasaki, T. Yamada, A. Maeda and H. Shinohara, "A Refreshable Analog VLSI Neural Network Chip with 400 Neurons and 40k Synapses," *IEEE J. Solid-State Circuits*, vol. **27** (12), pp 1854-1861, Dec. 1992.

[77] R.G. Benson and D.A. Kerns, "UV-Activated Conductances Allow for Multiple Time Scale Learning," *IEEE Transactions on Neural Networks,* vol. **4** (3), pp 434-440, 1993.

[78] K. Soelberg, R.L. Sigvartsen, T.S. Lande and Y. Berg, "An Analog Continuous-Time Neural-Network," *Int. J. Analog Integ. Circ. Signal Proc.*, vol. **5** (3), pp 235-246, May 1994.

[79] T. Morie and Y. Amemiya, "An All-Analog Expandable Neural-Network LSI with On-Chip Backpropagation Learning," *IEEE J. Solid-State Circuits*, vol. **29** (9), pp 1086-1093, Sept. 1994.

[80] F.J. Kub and E.W. Justh, "Analog CMOS Implementation of High-Frequency Least-Mean Square Error Learning Circuit," *IEEE J. Solid-State Circuits*, vol. **30** (12), pp 1391-1398, Dec. 1995.

[81] Y. Berg, R.L. Sigvartsen, T.S. Lande and A. Abusland, "An Analog Feedforward Neural-Network with On-Chip Learning," *Int. J. Analog Integ. Circ. Signal Proc.*, vol. **9** (1), pp 65-75, Jan. 1996.

[82] J.W. Cho, Y.K. Choi, S.Y. Lee, "Modular Neuro-Chip with On-Chip Learning and Adjustable Learning Parameters," *Neural Proc. Letters*, vol. **4** (1), 1996.

[83] M. Valle, D.D. Caviglia and G.M. Bisio, "An Experimental Analog VLSI Neural-Network with On-Chip Backpropagation Learning," *Int. J. Analog Integ. Circ. Signal Proc.*, vol. **9** (3), pp 231-245, Apr. 1996.

6.5.2 Outer-Product Unsupervised Learning

[84] J.P. Sage and R.S. Withers, "Analog Nonvolatile Memory for Neural Network Implementations," 1988, reprinted in N. Morgan, Ed., *Artificial Neural Networks: Electronic Implementations,* CA, Los Alamitos: IEEE Computer Society Press, 1990, pp 22-32.

[85] K.A. Boahen, P.O. Pouliquen, A.G. Andreou and R.E. Jenkins, "A Heteroassociative Memory Using Current-Mode MOS Analog VLSI Circuits," *IEEE T. Circ. Syst.*, vol. **36** (5), pp 747-755, 1989.

[86] J.R. Mann and S. Gilbert, "An Analog Self-Organizing Neural Network Chip," in *Advances in Neural Information Processing Systems,* San Mateo, CA: Morgan Kaufman, vol. **1**, pp 739-747, 1989.

[87] A. Hartstein and R.H. Koch, "A Self-Learning Neural Network," in *Advances in Neural Information Processing Systems,* San Mateo, CA: Morgan Kaufman, vol. **1**, pp 769-776, 1989.

[88] M.R. Walker, S. Haghighi, A. Afghan and L.A. Akers, "Training a Limited-Interconnect, Synthetic Neural IC," in *Advances in Neural Information Processing Systems,* San Mateo, CA: Morgan Kaufman, vol. **1**, pp 777-784, 1989.

[89] A. Murray, "Pulse Arithmetic in VLSI Neural Networks," *IEEE Micro Mag.*, pp 64-74, Dec. 1989.

[90] Y. Arima, K. Mashiko, K. Okada, T. Yamada, A. Maeda et al., "A 336-Neuron, 28k-Synapse, Self-Learning Neural Network Chip with Branch-Neuron-Unit Architecture," *IEEE J. Solid-State Circuits*, vol. **26** (11), pp 1637-1644, 1991.

[91] B.J. Maundy and E.I. Elmasry, "A Self-Organizing Switched-Capacitor Neural Network," *IEEE T. Circ. Syst.*, vol. **38** (12), pp 1556-1563, Dec. 1991.

[92] D.A. Watola and J.L. Meador, "Competitive Learning in Asynchronous-Pulse-Density Integrated-Circuits," *Int. J. Analog Integ. Circ. Signal Proc.*, vol. **2** (4), pp 323-344, Nov. 1992.

[93] J. Donald and L. Akers, "An Adaptive Neural Processor Node," *IEEE Transactions on Neural Networks,* vol. **4** (3), pp 413-426, 1993.

[94] Y. He and U. Cilingiroglu, "A Charge-Based On-Chip Adaptation Kohonen Neural Network," *IEEE Transactions on Neural Networks*, vol. **4** (3), pp 462-469, 1993.

[95] D. Macq, M. Verleysen, P. Jespers and J.D. Legat, "Analog Implementation of a Kohonen Map with On-Chip Learning," *IEEE T. Neural Networks*, vol. **4** (3), pp 456-461, May 1993.

[96] B. Linares-Barranco, E. Sanchez-Sinencio, A. Rodriguez-Vazquez, and J.L. Huertas, "A CMOS Analog Adaptive BAM with On-Chip Learning and Weight Refreshing," *IEEE Transactions on Neural Networks,* vol. **4** (3), pp 445-455, 1993.

[97] P. Heim and E.A. Vittoz, "Precise Analog Synapse for Kohonen Feature Maps," *IEEE J. Solid-State Circuits*, vol. **29** (8), pp 982-985, Aug. 1994.

[98] G. Cauwenberghs and V. Pedroni,"A charge-based CMOS parallel analog vector quantizer," in *Advances in Neural Information Processing Systems,* Cambridge, MA: MIT Press, vol. **7**, pp 779-786, 1995.

[99] T. Shibata, H. Kosaka, H. Ishii and T. Ohmi, "A Neuron-MOS Neural-Network Using Self-Learning-Compatible Synapse Circuits," *IEEE J. Solid-State Circuits*, vol. **30** (8), pp 913-922, Aug. 1995.

[100] R.Y. Liu, C.Y. Wu and I.C. Jou, "A CMOS Current-Mode Design of Modified Learning-Vector-Quantization Neural Networks," *Int. J. Analog Integ. Circ. Signal Proc.*, vol. **8** (2), pp 157-181, Sept. 1995.

[101] C.Y. Wu and J.F. Lan, "CMOS Current-Mode Neural Associative Memory Design with On-Chip Learning," *IEEE T. Neural Networks*, vol. **7** (1), pp 167-181, Jan. 1996.

[102] K. Hosono, K. Tsuji, K. Shibao, E. Io, H. Yonezu et al., "Fundamental Device and Circuits for Synaptic Connections in Self-Organizing Neural Networks," *IEICE T. Electronics*, vol. **E79C** (4), pp 560-567, Apr. 1996.

[103] T. Serrano-Gotarredona and B. Linares-Barranco, "A Real-Time Clustering Microchip Neural Engine," *IEEE T. VLSI Systems*, vol. **4** (2), pp 195-209, June 1996.

6.5.3 Adaptive Cellular Neural Networks

[104] P. Tzionas, P. Tsalides and A. Thanailakis, "Design and VLSI Implementation of a Pattern Classifier Using Pseudo-2D Cellular Automata," *IEE Proc G*, vol. **139** (6), pp 661-668, Dec. 1992.

[105] T. Roska and L.O. Chua, "The CNN Universal Machine - an Analogic Array Computer," *IEEE T. Circ. Syst. II*, vol. **40** (3), pp 163-173, March 1993.

[106] Y. Miyanaga and K. Tochinai, "Parallel VLSI Architecture for Multilayer Self-Organizing Cellular Network," *IEICE T. Electronics*, vol. **E76C** (7), pp 1174-1181, July 1993.

[107] S. Espejo, R. Carmona, R. Dominguez-Castro and A. Rodriguez-Vazquez, "A CNN Universal Chip in CMOS Technology," *Int J. Circuit Theory Appl.*, vol. **24** (1), pp 93-109, Jan-Febr. 1996.

6.5.4 Adaptive Fuzzy Classifiers

[108] J.W. Fattaruso, S.S. Mahant-Shetti, and J.B. Barton, "A Fuzzy Logic Inference Processor," *IEEE Journal of Solid-State Circuits,* vol. **29** (4), pp. 397-401, 1994.

[109] Z. Tang, Y. Kobayashi, O. Ishizuka and K. Tanno, "A Learning Fuzzy Network and Its Applications to Inverted Pendulum System," *IEICE T. Fund. El. Comm. Comp. Sci.*, vol. **E78A** (6), pp 701-707, June 1995.

[110] F. Vidal-Verdu and A. Rodriguez-Vazquez, "Using Building Blocks to Design Analog Neuro-Fuzzy Controllers," *IEEE Micro*, vol. **15** (4), pp 49-57, Aug. 1995.

[111] W. Pedrycz, C.H. Poskar and P.J. Czezowski, "A Reconfigurable Fuzzy Neural-Network with In-Situ Learning," *IEEE Micro*, vol. **15** (4), pp 19-30, Aug. 1995.

[112] T. Yamakawa, "Silicon Implementation of a Fuzzy Neuron," *IEEE Fuz Sy*, vol. **4** (4), pp 488-501, Nov. 1996.

6.5.5 Reinforcement Learning

[113] C. Schneider and H. Card, "Analog CMOS Synaptic Learning Circuits Adapted from Invertebrate Biology," *IEEE T. Circ. Syst.*, vol. **38** (12), pp 1430-1438, Dec. 1991.

[114] T.G. Clarkson, C.K. Ng and Y. Guan, "The pRAM: An Adaptive VLSI Chip," *IEEE Trans. on Neural Networks,* vol. **4** (3), pp 408-412, 1993.

[115] A.F. Murray, S. Churcher, A. Hamilton, A.J. Holmes, G.B. Jackson et al., "Pulse Stream VLSI Neural Networks," *IEEE Micro*, vol. **14** (3), pp 29-39, June 1994.

[116] G. Cauwenberghs, "Reinforcement Learning in a Nonlinear Noise Shaping Oversampled A/D Converter," to appear in *Proc. Int. Symp. Circuits and Systems,* Hong Kong, June 1997.

6.5.6 Nonidealities and Error Models

[117] M.J.S. Smith, "An Analog Integrated Neural Network Capable of Learning the Feigenbaum Logistic Map," *IEEE Transactions on Circuits and Systems,* vol. **37** (6), pp 841-844, 1990.

[118] R.C. Frye, E.A. Rietman, and C.C. Wong, "Back-Propagation Learning and Nonidealities in Analog Neural Network Hardware," *IEEE Transactions on Neural Networks,* vol. **2** (1), pp 110-117, 1991.

[119] L.M. Reyneri and E. Filippi, "An Analysis on the Performance of Silicon Implementations of Backpropagation Algorithms for Artificial Neural Networks," *IEEE Comput*, vol. **40** (12), pp 1380-1389, 1991.

[120] A. Murray and P.J. Edwards, "Synaptic Noise During MLP Training Enhances Fault-Tolerance, Generalization and Learning Trajectory," in *Advances in Neural Information Processing Systems,* San Mateo, CA: Morgan Kaufman, vol. **5**, pp 491-498, 1993.

[121] P. Thiran and M. Hasler, "Self-Organization of a One-Dimensional Kohonen Network with Quantized Weights and Inputs," *Neural Networks*, vol. **7** (9), pp 1427-1439, 1994.

[122] G. Cairns and L. Tarassenko, "Precision Issues for Learning with Analog VLSI Multilayer Perceptrons," *IEEE Micro*, vol. **15** (3), pp 54-56, June 1995.

[123] B.K. Dolenko and H.C. Card, "Tolerance to Analog Hardware of On-Chip Learning in Backpropagation Networks," *IEEE T. Neural Networks*, vol. **6** (5), pp 1045-1052, Sept. 1995.

6.5.7 Model-Free Learning

[124] A. Dembo and T. Kailath, "Model-Free Distributed Learning," *IEEE Transactions on Neural Networks,* vol. **1** (1), pp 58-70, 1990.

[125] M. Jabri and B. Flower, "Weight Perturbation: An Optimal Architecture and Learning Technique for Analog VLSI Feedforward and Recurrent Multilayered Networks," *IEEE Transactions on Neural Networks,* vol. **3** (1), pp 154-157, 1992.

[126] G. Cauwenberghs, "A Fast Stochastic Error-Descent Algorithm for Supervised Learning and Optimization," in *Advances in Neural Information Processing Systems,* San Mateo, CA: Morgan Kaufman, vol. **5**, pp 244-251, 1993.

[127] J. Alspector, R. Meir, B. Yuhas, and A. Jayakumar, "A Parallel Gradient Descent Method for Learning in Analog VLSI Neural Networks," in *Advances in Neural Information Processing Systems,* San Mateo, CA: Morgan Kaufman, vol. **5**, pp 836-844, 1993.

[128] B. Flower and M. Jabri, "Summed Weight Neuron Perturbation: An $\mathcal{O}(n)$ Improvement over Weight Perturbation," in *Advances in Neural Information Processing Systems,* San Mateo, CA: Morgan Kaufman, vol. **5**, pp 212-219, 1993.

[129] D. Kirk, D. Kerns, K. Fleischer, and A. Barr, "Analog VLSI Implementation of Gradient Descent," in *Advances in Neural Information Processing Systems,* San Mateo, CA: Morgan Kaufman, vol. **5**, pp 789-796, 1993.

[130] G. Cauwenberghs, "A Learning Analog Neural Network Chip with Continuous-Recurrent Dynamics", in *Advances in Neural Information Processing Systems,* San Mateo, CA: Morgan Kaufman, vol. **6**, pp 858-865, 1994.

[131] P.W. Hollis and J.J. Paulos, "A Neural-Network Learning Algorithm Tailored for VLSI Implementation," *IEEE T. Neural Networks*, vol. **5** (5), pp 784-791, Sept. 1994.

[132] G. Cauwenberghs, "An Analog VLSI Recurrent Neural Network Learning a Continuous-Time Trajectory," *IEEE Transactions on Neural Networks,* vol. **7** (2), March 1996.

[133] A.J. Montalvo, R.S. Gyurcsik and J.J. Paulos, "Toward a General-Purpose Analog VLSI Neural-Network with On-Chip Learning," *IEEE T. Neural Networks*, vol. **8** (2), pp 413-423, March 1997.

6.5.8 Chip-in-the-Loop Training

[134] M. Holler, S. Tam, H. Castro and R. Benson, "An Electrically Trainable Artificial Neural Network (ETANN) with 10240 Floating Gate Synapses," in *Proc. Int. Joint Conf. Neural Networks*, Washington DC, pp 191-196, 1989.

[135] S. Satyanarayana, Y. Tsividis and H.P. Graf, "A Reconfigurable Analog VLSI Neural Network Chip," in *Advances in Neural Information Processing Systems,* San Mateo, CA: Morgan Kaufman, vol. **2**, pp 758-768, 1990.

[136] E. Sackinger, B.E. Boser and L.D. Jackel, "A Neurocomputer Board Based on the ANNA Neural Network Chip," in *Advances in Neural Information Processing Systems,* San Mateo, CA: Morgan Kaufman, vol. **4**, pp 773-780, 1992.

[137] J.A. Lansner, "An Experimental Hardware Neural-Network Using a Cascadable, Analog Chipset," *Int J Elect*, vol. **78** (4), pp 679-690, Apr. 1995.

[138] J.O. Klein, H. Pujol and P. Garda, "Chip-In-The-Loop Learning Algorithm for Boltzmann Machine," *Electronics Letters*, vol. **31** (12), pp 986-988, Jun 8, 1995.

6.5.9 Digital Implementations

[139] A. Johannet, L. Personnaz, G. Dreyfus, J.D. Gascuel and M. Weinfeld, "Specification and Implementation of a Digital Hopfield-Type Associative Memory with On-Chip Training," *IEEE T. Neural Networks*, vol. **3** (4), pp 529-539, July 1992.

[140] T. Shima, T. Kimura, Y. Kamatani, T. Itakura, Y. Fujita et al., "Neurochips with On-Chip Backpropagation and/or Hebbian Learning," *IEEE J. Solid-State Circuits*, vol. **27** (12), pp 1868-1876, Dec. 1992.

[141] M. Yasunaga, N. Masuda, M. Yagyu, M. Asai, K. Shibata et al., "A Self-Learning Digital Neural Network Using Wafer-Scale LSI," *IEEE J. Solid-State Circuits*, vol. **28** (2), pp 106-114, Febr. 1993.

[142] C. Lehmann, M. Viredaz and F. Blayo, "A Generic Systolic Array Building-Block for Neural Networks with On-Chip Learning," *IEEE T. Neural Networks*, vol. **4** (3), pp 400-407, May 1993.

[143] M. Fujita, Y. Kobayashi, K. Shiozawa, T. Takahashi, F. Mizuno et al., "Development and Fabrication of Digital Neural-Network WSIs," *IEICE T. Electronics*, vol. **E76C** (7), pp 1182-1190, July 1993.

[144] P. Murtagh, A.C. Tsoi and N. Bergmann, "Bit-Serial Systolic Array Implementation of a Multilayer Perceptron," *IEE Proc E*, vol. **140** (5), pp 277-288, Sept. 1993.

[145] T. Morishita and I. Teramoto, "Neural-Network Multiprocessors Applied with Dynamically Reconfigurable Pipeline Architecture," *IEICE T. Electronics*, vol. **E77C** (12), pp 1937-1943, Dec. 1994.

[146] Z. Tang and O. Ishizuka, "Design and Implementations of a Learning T-Model Neural-Network," *IEICE T. Fund. El. Comm. Comp. Sci.*, vol. **E78A** (2), pp 259-263, Febr. 1995.

[147] M.P. Perrone and L.N. Cooper, "The NI1000: High Speed Parallel VLSI for Implementing Multilayer Perceptrons," in *Advances in Neural Information Processing Systems,* Cambridge, MA: MIT Press, vol. 7, pp 747-754, 1995.

[148] J. Wawrzynek, et al., "SPERT-II: A Vector Microprocessor System and its Application to Large Problems in Backpropagation Training," in *Advances in Neural Information Processing Systems,* Cambridge, MA: MIT Press, vol. 8, pp 619-625, 1996.

[149] S. Rehfuss and D. Hammerstrom, "Model Matching and SFMD Computation," in *Advances in Neural Information Processing Systems,* Cambridge, MA: MIT Press, vol. 8, pp 713-719, 1996.

6.5.10 Optical and Optoelectronic Implementations

[150] J. Ohta, Y. Nitta and K. Kyuma, "Dynamic Optical Neurochip Using Variable-Sensitivity Photodiodes," *Optics Lett*, vol. **16** (10), pp 744-746, 1991.

[151] D.Z. Anderson, C. Benkert, V. Hebler, J.-S. Jang, D, Montgomery and M. Saffman, "Optical Implementation of a Self-Organizing Feature Extractor," in *Advances in Neural Information Processing Systems,* San Mateo, CA: Morgan Kaufman, vol. **4**, pp 821-828, 1992.

[152] Y. Nitta, J. Ohta, S. Tai and K. Kyuma, "Optical Learning Neurochip with Internal Analog Memory," *Appl Optics*, vol. **32** (8), pp 1264-1274, March 10, 1993.

[153] K. Wagner and T.M. Slagle, "Optical Competitive Learning with VLSI Liquid-Crystal Winner-Take-All Modulators," *Appl Optics*, vol. **32** (8), pp 1408-1435, March 10, 1993.

[154] M. Oita, Y. Nitta, S. Tai and K. Kyuma, "Optical Associative Memory Using Optoelectronic Neurochips for Image-Processing," *IEICE T. Electronics*, vol. **E77C** (1), pp 56-62, Jan. 1994.

[155] E. Lange, Y. Nitta and K. Kyuma, "Optical Neural Chips," *IEEE Micro*, vol. **14** (6), pp 29-41, Dec. 1994.

[156] A.J. Waddie and J.F. Snowdon, "A Smart-Pixel Optical Neural-Network Design Using Customized Error Propagation," *Inst. Phys. Conf. Series*, vol. **139**, pp 511-514, 1995.

[157] K. Tsuji, H. Yonezu, K. Hosono, K. Shibao, N. Ohshima et al., "An Optical Adaptive Device and Its Application to a Competitive Learning Circuit," *Japan JAP 1*, vol. **34** (2B), pp 1056-1060, Febr. 1995.

[158] W.E. Foor and M.A. Neifeld, "Adaptive, Optical, Radial Basis Function Neural-Network for Handwritten Digit Recognition," *Appl Optics*, vol. **34** (32), pp 7545-7555, Nov. 10, 1995.

6.5.11 Architectural Novelties

[159] J. Alspector, J.W. Gannett, S. Haber, M.B. Parker and R. Chu, "A VLSI-Efficient Technique for Generating Multiple Uncorrelated Noise Sources and Its Application to Stochastic Neural Networks," *IEEE T. Circ. Syst.*, vol. **38** (1), pp 109-123, 1991.

[160] P.A. Shoemaker, M.J. Carlin and R.L. Shimabukuro, "Back Propagation Learning with Trinary Quantization of Weight Updates," *Neural Networks*, vol. **4** (2), pp 231-241, 1991.

[161] Y.H. Pao and W. Hafez, "Analog Computational Models of Concept-Formation," *Int. J. Analog Integ. Circ. Signal Proc.*, vol. **2** (4), pp 265-272, Nov. 1992.

[162] T. Morie and Y. Amemiya, "Deterministic Boltzmann Machine Learning Improved for Analog LSI Implementation," *IEICE T. Electronics*, vol. **E76C** (7), pp 1167-1173, July 1993.

[163] S.P. Deweerth and D.M. Wilson, "Fixed-Ratio Adaptive Thresholding Using CMOS Circuits," *Electronics Letters*, vol. **31** (10), pp 788-789, May 11, 1995.

[164] M. Vandaalen, J. Zhao and J. Shawetaylor, "Real-Time Output Derivatives for on Chip Learning Using Digital Stochastic Bit Stream Neurons," *Electronics Letters*, vol. **30** (21), pp 1775-1777, Oct. 13, 1994.

[165] V. Petridis and K. Paraschidis, "On the Properties of the Feedforward Method - a Simple Training Law for On-Chip Learning," *IEEE T. Neural Networks*, vol. **6** (6), pp 1536-1541, Nov. 1995.

[166] H. Singh, H.S. Bawa and L. Anneberg, "Boolean Neural-Network Realization of an Adder Subtractor Cell," *Microel Rel*, vol. **36** (3), pp 367-369, March 1996.

[167] T. Lehmann, E. Bruun and C. Dietrich, "Mixed Analog-Digital Matrix-Vector Multiplier for Neural-Network Synapses," *Int. J. Analog Integ. Circ. Signal Proc.*, vol. **9** (1), pp 55-63, Jan. 1996.

[168] T. Serrano-Gotarredona and B. Linares-Barranco, "A Modified ART-1 Algorithm More Suitable for VLSI Implementations," *Neural Networks*, vol. **9** (6), pp 1025-1043, Aug. 1996.

[169] M.L. Marchesi, F. Piazza and A. Uncini, "Backpropagation Without Multiplier for Multilayer Neural Networks," *IEE P. Circ.*, vol. **143** (4), pp 229-232, Aug. 1996.

6.6 Systems Applications of Learning

6.6.1 General Purpose Neural Emulators

[170] P. Mueller, J. Van der Spiegel, D. Blackman, T. Chiu, T. Clare, C. Donham, T.P. Hsieh and M. Lionaz, "Design and Fabrication of VLSI Components for a General Purpose Analog Neural Computer," in *Analog VLSI Implementation of Neural Systems*, Norwell, MA: Kluwer, pp 135-169, 1989.

6.6.2 Blind Signal Processing

[171] E. Vittoz and X. Arreguit, "CMOS Integration of Herault-Jutten Cells for Separation of Sources," in *Analog VLSI Implementation of Neural Systems*, Norwell, MA: Kluwer, pp 57-83, 1989.

[172] M.H. Cohen and A.G. Andreou, "Current-Mode Subthreshold MOS Implementation of the Jutten-Herault Autoadaptive Network," *IEEE J. Solid-State Circuits*, vol. **27** (5), pp 714-727, 1992.

[173] R.P. Mackey, J.J. Rodriguez, J.D. Carothers and S.B.K. Vrudhula, "Asynchronous VLSI Architecture for Adaptive Echo Cancellation," *Electronics Letters*, vol. **32** (8), pp 710-711, Apr. 11, 1996.

6.6.3 Biomedical Adaptive Signal Processing

[174] R. Coggings, M. Jabri, M. Flower and S. Pickard, "ICEG Morphology Classification Using an Analogue VLSI Neural Network," in *Advances in Neural Information Processing Systems,* Cambridge, MA: MIT Press, vol. **7**, pp 731-738, 1995.

6.6.4 Speech Research

[175] J. Wawrzynek, et al., "SPERT-II: A Vector Microprocessor System and its Application to Large Problems in Backpropagation Training," in *Advances in Neural Information Processing Systems,* Cambridge, MA: MIT Press, vol. 8, pp 619-625, 1996.

[176] J. Lazzaro, "Temporal Adaptation in a Silicon Auditory Nerve," in *Advances in Neural Information Processing Systems,* San Mateo, CA: Morgan Kaufman, vol. **4**, pp 813-820, 1992.

6.6.5 Olfactory Sensory Processing

[177] P.A. Shoemaker, C.G. Hutchens and S.B. Patil, "A Hierarchical-Clustering Network Based on a Model of Olfactory Processing," *Int. J. Analog Integ. Circ. Signal Proc.*, vol. **2** (4), pp 297-311, Nov. 1992.

6.6.6 Focal-Plane Sensors and Adaptive Vision Systems

[178] J. Tanner and C.A. Mead, "An Integrated Analog Optical Motion Sensor," in *VLSI Signal Processing II,* S.Y. Kung, Ed., New York: IEEE Press, 1986, pp 59-76.

[179] C.A. Mead, "Adaptive Retina," in *Analog VLSI Implementation of Neural Systems*, C. Mead and M. Ismail, Eds., Norwell, MA: Kluwer Academic Pub., 1989, pp 239-246.

[180] M. Mahowald, *An Analog VLSI Stereoscopic Vision System,* Boston, MA: Kluwer Academic, 1994.

[181] T. Delbruck, "Silicon Retina with Correlation-Based, Velocity-Tuned Pixels," *IEEE Transactions on Neural Networks,* vol. **4**, pp 529-541, 1993.

[182] J.C. Lee, B.J. Sheu, and W.C. Fang, "VLSI Neuroprocessors for Video Motion Detection," *IEEE Transactions on Neural Networks,* vol. **4** (2), pp 78-191, 1993.

[183] R. Etienne-Cummings, J. Van der Spiegel, P. Mueller, "VLSI Model of Primate Visual Smooth Pursuit," in *Advances in Neural Information Processing Systems,* Cambridge, MA: MIT Press, vol. 8, pp 707-712, 1996.

[184] R. Sarpeshkar, J. Kramer, G. Indiveri and C. Koch, "Analog VLSI Architectures for Motion Processing - from Fundamental Limits to System Applications," *P IEEE*, vol. **84** (7), pp 969-987, July 1996.

[185] K. Boahen, "A Retinomorphic Vision System," *IEEE Micro,* vol. **16** (5), pp 30-39, October 1996.

[186] S.C. Liu and C. Mead, "Continuous-Time Adaptive Delay System," *IEEE T. Circ. Syst. II*, vol. **43** (11), pp 744-751, Nov. 1996.

6.6.7 Optical Character Recognition

[187] B.Y. Chen, M.W. Mao and J.B. Kuo, "Coded Block Neural Network VLSI System Using an Adaptive Learning-Rate Technique to Train Chinese Character Patterns," *Electronics Letters*, vol. **28** (21), pp 1941-1942, Oct 8, 1992.

[188] C.S. Miou, T.M. Shieh, G.H. Chang, B.S. Chien, M.W. Chang et al., "Optical Chinese Character-Recognition System Using a New Pipelined Matching and Sorting VLSI," *Opt Eng*, vol. **32** (7), pp 1623-1632, July 1993.

[189] S. Maruno, T. Kohda, H. Nakahira, S. Sakiyama and M. Maruyama, "Quantizer Neuron Model and Neuroprocessor-Named Quantizer Neuron Chip," *IEEE J. Sel. Areas Comm.*, vol. **12** (9), pp 1503-1509, Dec. 1994.

6.6.8 Image Compression

[190] W.C. Fang, B.J. Sheu, O.T.C. Chen and J. Choi, "A VLSI Neural Processor for Image Data-Compression Using Self-Organization Networks," *IEEE Transactions on Neural Networks,* vol. **3** (3), pp 506-518, 1992.

6.6.9 Communications and Decoding

[191] J.G. Choi, S.H. Bang and B.J. Sheu, "A Programmable Analog VLSI Neural-Network Processor for Communication Receivers," *IEEE T. Neural Networks*, vol. **4** (3), pp 484-495, May 1993.

[192] M.I. Chan, W.T. Lee, M.C. Lin and L.G. Chen, "IC Design of an Adaptive Viterbi Decoder," *IEEE T. Cons. El.*, vol. **42** (1), pp 52-62, Febr. 1996.

[193] R. Mittal, K.C. Bracken, L.R. Carley and D.J. Allstot, "A Low-Power Backward Equalizer for DFE Read-Channel Applications," *IEEE J. Solid-State Circuits*, vol. **32** (2), pp 270-273, Febr. 1997.

[194] B.C. Rothenberg, J.E.C. Brown, P.J. Hurst and S.H. Lewis, "A Mixed-Signal RAM Decision-Feedback Equalizer for Disk Drives," *IEEE J. Solid-State Circuits,* vol. **32** (5), pp 713-721, 1997.

6.6.10 Clock Skew Timing Control

[195] W.D. Grover, J. Brown, T. Friesen and S. Marsh, "All-Digital Multipoint Adaptive Delay Compensation Circuit for Low Skew Clock Distribution," *Electronics Letters*, vol. **31** (23), pp 1996-1998, Nov 9, 1995.

[196] M. Mizuno, M. Yamashina, K. Furuta, H. Igura, H. Abiko et al., "A GHz MOS Adaptive Pipeline Technique Using MOS Current-Mode Logic," *IEEE J. Solid-State Circuits*, vol. **31** (6), pp 784-791, June 1996.

[197] E.W. Justh and F.J. Kub, "Analog CMOS Continuous-Time Tapped Delay-Line Circuit," *Electronics Letters*, vol. **31** (21), pp 1793-1794, Oct. 12, 1995.

6.6.11 Control and Autonomous Systems

[198] Y. Harata, N. Ohta, K. Hayakawa, T. Shigematsu and Y. Kita, "A Fuzzy Inference LSI for an Automotive Control," *IEICE T. Electronics*, vol. **E76C** (12), pp 1780-1787, Dec. 1993.

[199] G. Jackson and A.F. Murray, "Competence Acquisition is an Autonomous Mobile Robot using Hardware Neural Techniques," in *Adv. Neural Information Processing Systems,* Cambridge, MA: MIT Press, vol. **8**, pp. 1031-1037, 1996.

6.6.12 High-Energy Physics

[200] T. Lindblad, C.S. Lindsey, F. Block and A. Jayakumar, "Using Software and Hardware Neural Networks in a Higgs Search," *Nucl Inst A*, vol. **356** (2-3), pp 498-506, March 15, 1995.

[201] C.S. Lindsey, T. Lindblad, G. Sekhniaidze, G. Szkely and M. Minerskjold, "Experience with the IBM ZISC036 Neural-Network Chip," *Int J. Modern Phys. C*, vol. **6** (4), pp 579-584, Aug. 1995.

[202] G. Anzellotti, R. Battiti, I. Lazzizzera, G. Soncini, A. Zorat et al., "Totem - a Highly Parallel Chip for Triggering Applications with Inductive Learning Based on the Reactive Tabu Search," *Int J. Modern Phys. C*, vol. **6** (4), pp 555-560, Aug. 1995.