

On-board neural-processor design for an intelligent multi-sensor microspacecraft

Wai-Chi Fang¹, Bing J. Sheu², James Wall¹

¹Advanced Computing Resources Group and Center for Space Microelectronics Technology
Jet Propulsion Laboratory, California Institute of Technology
4800 Oak Grove Drive, Pasadena, CA 91109-8099, USA

²Department of Electrical Engineering and Center for Neural Engineering
University of Southern California, Los Angeles, CA 90089-0271, USA

ABSTRACT

A compact VLSI neural processor based on the Optimization Cellular Neural Network (OCNN) has been under development to provide a wide range of support for an intelligent remote sensing microspacecraft which requires both high bandwidth communication and high-performance computing for on-board data analysis, thematic data reduction, synergy of multiple types of sensors, and other advanced smart-sensor functions. The OCNN is developed with emphasis on its capability to find global optimal solutions by using a hardware annealing method. The hardware annealing function is embedded in the network. It is a paralleled version of fast mean-field annealing in analog networks, and is highly efficient in finding globally optimal solutions for cellular neural networks. The OCNN is designed to perform programmable functions for fine-grained processing with annealing control to enhance the output quality. The OCNN architecture is a programmable multi-dimensional array of neurons which are locally connected with their local neurons. Major design features of the OCNN neural processor includes massively parallel neural processing, hardware annealing capability, winner-take-all mechanism, digitally programmable synaptic weights, and multisensor parallel interface. A compact current-mode VLSI design feasibility of the OCNN neural processor is demonstrated by a prototype 5x5-neuroprocessor array chip in a 2- μ m CMOS technology. The OCNN operation theory, architecture, design and implementation, prototype chip, and system applications have been investigated in detail and presented in this paper.

Keywords

array processors, image sensor, multichip module, neural networks, optoelectronic integrated circuit, parallel processing, small satellite, microspacecraft, smart pixels, VLSI.

1. INTRODUCTION

1.1. Advanced Microspacecraft Overview [1, 2]

Advanced microspacecraft are intended to provide affordable, comprehensive access to space. Increasingly, space scientists are asked to deal with the economics of their data and the systems used to collect it. NASA is moving from a model of spacecraft design where science instruments are designed and built separately and then attached to the spacecraft late in the integration process to one where science collection needs are built into a spacecraft from the beginning and designed along with all other subsystems. This type of spacecraft has been dubbed a "sciencecraft" typified by small size and low cost. Low mass (10 kg to 100 kg) also provides for a 10x reduction in launch vehicle cost over traditional spacecraft (1000 kg+) when sent on the same trajectory. One major tradeoff of smaller spacecraft is reduced power which, in turn, results in reduced communication bandwidth. Thus, scientists must turn to several techniques including more powerful on-board processing of science data, to maximize the return of useful, interesting information. The synergy arising out of simultaneous processing of data from multiple sensors is another way of enriching the downlinked science data.

The economics of space science in the last decade of the twentieth century follows that of many other areas of endeavor. More results must be obtained from fewer resources. Intellectual capital is the one resource which can be applied without restraint. The "mantra" of NASA which captures this economic trend is "faster/better/cheaper" and "smarter" is the add-on which recognizes the central role of creativity and invention. This paper explores the application of neural processing to science data. The overall goal is to do more exciting science, both filling in the blanks left from earlier exploration and providing additional capabilities to capture unanticipated phenomena. The increasing autonomy of sciencecraft supported by a variety of processing paradigms (general purpose, interval logic, neural net, digital signal, etc.) have the potential to give this new generation of spacecraft abilities to recognize and capture important science without prior explicit instruction from the ground.

Two important computing trends are emerging in current spacecraft infrastructure design: packaging and processor technology. Aggressive packaging using multichip modules (MCMs), MCM stacking, chip stacking, chip-scale packaging (CSP), plastic packaging, etc. is driving down mass and increasing the functional density of flight avionics. The decreasing lag time between the commercial introduction of microprocessors and their use in space systems supports state-of-the-art processing performance along with the advantages of contemporary software and hardware development tools. These advances in packaging for microelectronics coupled with *low-power* design and the continued massive increases in memory capacity form a powerful foundation for new types of spacecraft processing approaches,

Packaging advances also support the physical integration of microsensors, associated control/conditioning electronics, buffering, processing, data compression and many other functions Traditionally separated, Data from one type of sensor may be used to dynamically calibrate other sensors. Signals from multiple sensors may be combined to provide improved signal-to-noise ratios. These and many other techniques are made possible by the physical proximity of sensors and the ability of advanced packages to provide controlled electrical environments (Faraday shielding, controlled impedance, extremely short signal paths, etc.) The stacking of MCMS and integrated circuits (3D microelectronics), allows this environmental control to extend into the third dimension. It also readily supports the evolution of sensor structures from linear arrays to area arrays [o solid arrays.

As mentioned earlier, packaging and processor technologies are key to the goals of twenty-first century spacecraft 3D microelectronics packaging has already provided a two order-of-magnitude increase in functional density over the previous generation of spacecraft microelectronics. Advanced 3D VLSI packaging techniques promise another two or three order-of-magnitude increase over the next 10 years. It should be noted that this is independent of any increase in functional density achieved by advances in IC process technology which may provide another 4x to 9x improvement in area efficiency. Processor technologies, through a variety of techniques, will extend the performance of spacecraft processors from the current 20 MIPS [o over 500 MIPS over the next 10 years. Some of these Techniques involve chip-level architectures, others are based on multi-computing arrays.

1.2. Intelligent Remote Sensing Imager

The Intelligent Multisensory Imager (IMI) is a highly integrated and versatile remote imaging instrument which performs various types of data sensing, acquisition, processing, synergy, compression, communication, and autonomous multisensory control. The IMI is proposed for NASA's advanced remote sensing microspacecrafts which automatically recognize, localize, and classify point, area, and volume objects and phenomena in real-time on the Earth and on planets [3].

The IMI includes two major subsystems: IMI Eye, and IMI Brain. The IMI Eye is a compact optoelectronic subsystem which integrates a wide range of different sensors with geometric, radiometric, and spectral parameters meeting the actual science and mission requirements. The IMI Brain is a high performance control and data subsystem which provides on-board computing resources for the IMI to perform various on-board tasks. Due to the adaptation of multiple sensor parameters for various dedicated tasks it is possible to optimize and to minimize the science payload complexity for microspacecraft applications. By means of on-board smart sensor functions and data analysis of the generated geometric and spectral data, a minimum task specific data volume can be produced and transmitted [o the Earth, meanwhile a maximum information may be provided for real-time users with PC based receiver stations [4]. The success of the IMI development will enable new science capabilities, enhance scientific return, improve timely information availability, reduce operation costs, and alleviate downlink limitations for the late- 1990's and future missions.

1.3. The Role of a High Performance OCN Neural Processor

A compact VLSI neural processor based on the optimization Cellular Neural Network (OCNN) has been under development to provide a wide range of supports for an intelligent remote sensing microspacecraft which requires both high bandwidth communication and high-performance computing for on-board data analysis, thematic data reduction, synergy of multiple types of sensors, and other advanced smart-sensor functions. Incorporating of the proposed VLSI neural neuroprocessor into the IMI Brain offers orders-of-magnitude computing performance enhancements for on-board real-time intelligent multisensor processing and control tasks. The OCN is targeted for the multisensor applications although it is universal as the Turing machine [7]. The neuroprocessor based on the OCN has great potential in solving many important scientific and engineering problems by the use of different cloning templates. For the maximum flexibility over a variety of applications, many CNN functions has been verified via system simulation. These functions include noise filtering, isolated pixel elimination, hole filling, morphological operations, image enhancement, edge detection, connected component detection, feature extraction, motion detection, motion estimation, motion compensation, object counting, size estimation, path tracking, collision avoidance, minimal and maximal detection, etc. The proper network operation is also confirmed for known and arbitrary cloning templates,

The OCN operation theory, architecture, design and implementation, prototype chip, and system applications have been investigated in detail and presented in the following sections.

2. INCORPORATING A NEURAL PROCESSOR INTO AN INTELLIGENT MULTISENSOR MICROSPACECRAFT

Cellular Neural Networks is a multi-dimensional array of mainly identical cells which are dynamic systems with continuous state variables and locally connected with their local cells within a finite radius. Since its original publication by Chua and Yang [5,6] in 1988, the CNN paradigm has evolved rapidly and provides an unified framework for many computation-intensive applications such as signal processing and optimization. Moreover, the CNN architecture is a locally connected, massively paralleled computing system with simple synaptic operators so that it is very suitable for VLSI implementation in real-time, high-speed applications [7,8]. The behavior of cellular neural networks depends on the computing model, network topology, and coefficient templates. The operation for different applications depends primarily on the coefficients of the templates and the procedure to apply them. A template includes the information for synapse weights, threshold value-s, and boundary conditions.

In this paper, the OCN is introduced with emphasis on its capability to find global optimal solutions by using a hardware annealing method [8,9]. The hardware annealing function is embedded in the network. It is a paralleled version of fast mean-field annealing in analog networks, and is highly efficient in finding globally optimal solutions for cellular neural networks. The OCN was designed to perform programmable functions for fine-grained processing with annealing control to enhance the output quality.

2.1. Major Features of the Optimization Cellular Neural Network

The OCN proposed for the Intelligent Multisensory System is an improved version of the original CNN. It has four more significant features than the basic CNN:

(A) *Optimal Solutions of Energy Function:*

Under the mild conditions [5], a CNN autonomously finds a stable solution for which the Lyapunov function of the network is locally minimized. To improve the local minimized energy function of the basic CNN, the annealing capability is included to accommodate the applications in which the optimal solutions of energy function are needed. Hardware annealing [9] is a highly efficient method of finding optimal solutions for cellular neural networks.

(B) *Multiple Layers with Embedded Maximum Evolution Functions:*

In the original CNN every pixel is represented by one neuron. In the OCN every pixel can be represented by multiple neurons which form a hyperneuron and execute the maximum evolution function for various profile selection or data synergy. For instance in the OCN designed for motion detection, every image pixel is represented by multiple mutually exclusive neurons which form a hyperneuron for velocity selection. Only the winning neuron is active high and the other neurons of the same hypercolumn are turned off. The network operation will be terminated whenever the energy function of the net work reaches a minimum.

(c) *Digitally Programmable Synapse Weigh(s):*

To improve the fixed synapse weights of the basic CNN, the digitally programmable synapse weights [10] are designed for the OCN to accommodate the applications in which programmable pre-determined operators are needed.

(D) *High-speed Parallel External Image I/O:*

To improve the global interconnections and external image I/O of the basic CNN, a 2,-D array of optical receivers and transmitters is integrated with the OCN to accommodate the applications in which high-speed parallel external image I/O and optical interconnections are needed [11].

2.2. System Architecture of the OCN Neural Processor

There are three major different system computing architectures of an OCN processing engine: (a) a front-end sensory information processor, (b) a processing accelerator, and (c) a universal machine. datapath. For IMI applications, the first two system architectures are selected,

2.2..1. Smart-Sensor OCN

The OCN can be used as a front-end sensory information processor to provide. high throughput real-time computing power at neighborhood of the sensory circuit. Figure 2.1 shows a data flow diagram of a smart-sensor OCN neuroprocessor design by taking advantage of the high-speed parallel optical I/O. The functional blocks for the optoelectronic OCN neuroprocessor include: (a) an optical receiver array, (b) an electronic neuroprocessors array, and (c) a programmable

synaptic-weight matrix memory. It is feasible to put the active-pixel sensor [15] on the top of a die stack of the OCNN. This optoelectronic smart-pixel OCNN design can take the combined advantages of the optics and electronics to achieve ultra-high-speed smart sensory information processing and analysis at the focal plane. In order to achieve a reliable implementation for an optoelectronic integrated computing system, the fabrication and packaging technology should be well supported for the optical functions as well as the electrical functions.

2.2.2, processing Accelerator OCNN

The OCNN can be used as a processing accelerator in a heterogeneous computing system. A heterogeneous computing processor with embedded neuroprocessors is shown in Fig. 2.2. This integrated computing module includes an 256x256-array OCNN neuroprocessor, dual RISCs, data and program memories, mass memory, high speed data links, spacecraft bus interface, multi-sensor interface, and optional host bus interface. It is implemented in a multichip module (MCM) of substrate size 1250 mm x 500 mm. Table 2.1 shows its projected performance data based on a 0.5- μm CMOS technology and a 1997 delivery schedule. The power dissipation of the proposed IMI Brain MCM is about 15 W at 50 MHz nominal operation. The network operation clock for the OCNN is about 0.2 MHz. Tire IMI Brain MCM provides 55x2 MIPS and 82x4 GCPS (gigs-connections per second). Its volume is less than 100 cm³ and its mass is about 200 gm.

A functional operation of the IMI processor is briefly described in the following. The master RISC gets an interrupt if the system is receiving commands. The master RISC has a task link to the slave RISC. In the interrupt procedure the slave RISC will be initialized by the master RISC and both RISC's are waiting for the next trigger. The master RISC has the highest priority and manages the tasks and procedures. In the command phase, the master RISC initializes the neuroprocessor array and controller. The OCNN is used as a high-performance processing accelerator for the master RISC to perform on-board computation and control tasks. The neural processing commands are pre-stored in the programmable synaptic-weight memory. Update of the synaptic weights is depended on the application algorithms and their associated learning process. Meanwhile the slave RISC has to control the multisensory electronics and direct the multi-sensor data [o the local mass memory or to the neuroprocessor data memory. The buffered raw data and the processed data are under the control of the master RISC. The master RISC builds up data frames with synchronize labels and house keeping data. The frames are then transmittal to the microspacecraft for further storage or downlink to the ground. A scaleable multiprocessor architecture can also be realized through high speed data links for advanced parallel processing tasks.

2.4. Advanced Packaging and Ultra-Low-Power Electronics [2]

An ultra-low-power 1024x1024-array OCNN neural processing engine has been under a development study by using a 3-D VLSI die stacking technology combined with a sub-0.25 μm low power ($V_{DD} = 0.9 \text{ V}$) SOI CMOS process technology which is under development in MIT Lincoln Laboratory. A 3-D VLSI stack of dimensions 3 cm x 3 cm x 0.5 cm is projected to accommodate a complete 1024x1024-array OCNN processor with one 1024x1024 active-pixel sensor on the top of the die stack. A miniaturized highly-integrated IMI Brain is therefore feasible to be implemented into a compact cube at a manageable power dissipation rate.

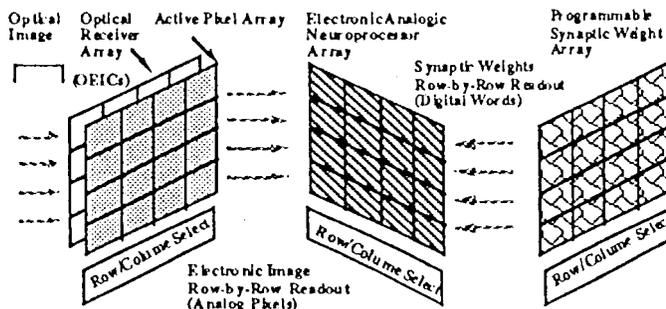


Fig. 2.1. Data flow diagram of the optoelectronic smart-sensor OCNN neuroprocessor array.

| Projection Year | 1997 | 2001 |
|---------------------|---------------------|--------------------------|
| CMOS Technology | 0.5 μm | 0.25 μm (SOI) |
| RISC Performance | 55MIPS@50MHz | 150MIPS@66MHz |
| RISC Chip Example | RAD6000-5L | PowerPC620 |
| OCNN neuron size | 470x746 λ^2 | |
| OCNN die size | 15 mmx 15 mm | 28.5mmx28.5 mm |
| OCNN array size | 128x128 | 512x512 |
| OCNN network speed | 5 Usec | 2.5 us |
| OCNN performance | 82 GCPS/die | 1048 GCPS/die |
| Program/Data Memory | 2.5 MB | 80 MB |
| Mass Memory | 1 GB | 32 GB |
| Power | 15 Watts | 10 Watts |
| Volume | 100 cm ³ | 50 cm ³ |
| Mass | 200 gm | 100 gm |
| Voltage | 5 V | 3.3 V |

Table 2.1. Performance data for an IMI Brain MCM

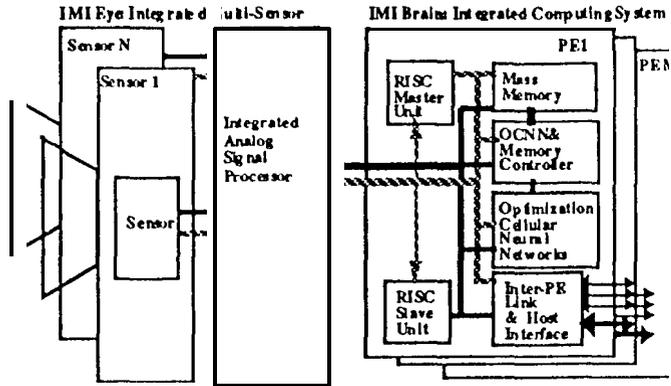


Fig. 2.2. System structure of the Intelligent Multisensory Imager with embedded OCNN neuroprocessors.

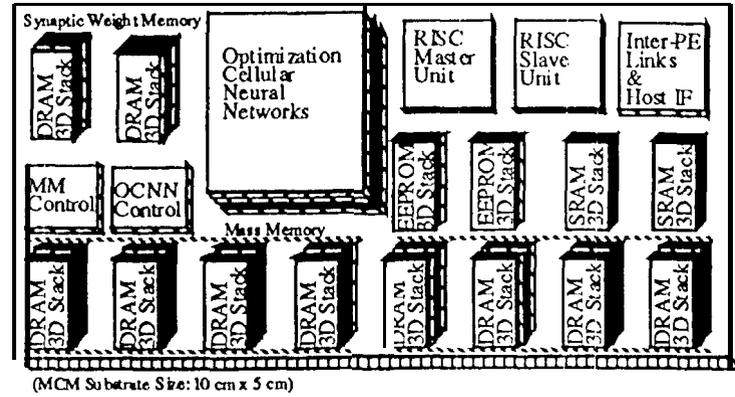


Fig. 2.3. The IMI Brain MCM silicon substrate layout.

3. THE OPTIMIZATION CELLULAR NEURAL NETWORK

Cellular neural network (CNN) is a class of recursive neural networks with locally-interconnected neural processors. The behavior of cellular neural networks depends on the computing model, network topology, and coefficient templates. The operation for different applications depends primarily on the coefficients of the templates and the procedure to apply them. A *template* includes the information for synapse weights, threshold values, and boundary conditions. In this paper, the optimization Cellular Neural Network is developed with emphasis on its capability to find global optimal solutions by using a hardware annealing method and a high-speed optoelectronic I/O. The *hardware* annealing function is embedded in the network. It is a paralleled version of fast mean-field annealing in analog networks, and is highly efficient in finding globally optimal solutions for cellular neural networks. The OCNN was designed to perform programmable functions for fine-grained processing with annealing control to enhance the output quality.

3.1 .Operation Theory

Consider the OCNN with $n \times m$ neurons on rectangular grid as shown in Fig. 3.1. Fig. 3.2 shows a model of the OCNN neuron. Each neuron has the piecewise-linear transfer function $f_{pw}(\cdot)$ and its gain is variable. The gain is controlled by a monotonically increasing function $g(t)$ such that

$$v_y = f_{pw}(g v_x) = \begin{cases} +1, & \text{if } v_x > \frac{1}{g} \\ g v_x, & \text{if } -\frac{1}{g} \leq v_x \leq \frac{1}{g} \\ -1, & \text{if } v_x < -\frac{1}{g} \end{cases} \quad (1)$$

The OCNN dynamic behavior satisfies a set of differential equations in the matrix notation as given:

$$C_x \frac{dx}{dt} = -\frac{1}{R_x} x + A y + B u + I_b w \quad (2)$$

where $x = [v_{x1} v_{x2} \dots v_{xN}]^T$, $y = [v_{y1} v_{y2} \dots v_{yN}]^T$, $u = [v_{u1} v_{u2} \dots v_{uN}]^T$, and $w = [1 \dots 1]^T$ is an $N \times 1$ unity vector, $N = n \times m$. The capacitance C_x , and resistance R_x at the state node of the processing element, annealing gain control $g(t)$, and bias current I_b are assumed to be the same for the whole network. In (2), A and B are two-real N -by- P matrices consisting of feedback and feed-forward synapse weights and determined by given cloning templates T_A and T_B , respectively. For the shift-invariant CNNs, they are *real* symmetric. Since a piecewise-linear function is used as the transfer function of the network, the generalized energy function is a scalar-valued quadratic function of the output vector y ,

$$\begin{aligned}
 E &= -\frac{1}{2} \sum_{i,j} \sum_{C(k,l) \in N(i,j)} A(i,j;k,l) v_{yij} v_{ykl} + \frac{1}{2gR_x} \sum_{i,j} (v_{yij})^2 - \sum_{i,j} \sum_{C(k,l) \in N(i,j)} B(i,j;k,l) v_{yij} v_{ykl} - \sum_{i,j} I_b v_{yij} \\
 &= -\frac{1}{2} y^T \left(A - \frac{1}{gR_x} I \right) y - y^T B u - I_b y^T w = -\frac{1}{2} y^T M_g y - y^T b.
 \end{aligned} \tag{3}$$

where $M_g = \left(A - \frac{1}{gR_x} I \right)$ and $b = Bu + I_b w$. Notice that $I' = \frac{1}{R_x}$.

Here the factor $1/g$ in the second term stems from the energy associated with the piecewise-linear function with a neuron gain other than unity. The process of finding the optimal solutions takes place during the change of M_g from negative definite to indefinite matrix, as the annealing gain increases. Since M_g is also a real symmetric matrix, it can be diagonalized as $M_g = Q \Lambda_g Q^T$, where Λ_g is the diagonal matrix of eigenvalues λ_k , $k = 1, 2, \dots, N$, and Q is an $N \times N$ matrix whose columns are made of orthonormal set of eigenvectors e_k 's. In an annealed neural network, the elements of Λ_g are time-varying. However, Q is independent of the neuron gain because M and M_g commute. By noting that $M_g = A - T_x^{-1} I - ((1-g)T_x^{-1} I) = M - ((1-g)T_x^{-1} I)$, the relationship between the eigenvalues of un-annealed and annealed network can be easily shown to be

$$\lambda_k = \lambda'_k - \frac{(1-g)}{gR_x} = \lambda'_k - \frac{(1-g)T_x}{g}, \quad k = 1, 2, \dots, N \tag{4}$$

where λ'_k are the eigenvalues of M . In the hardware annealing, the eigenvalues λ_k 's are changed from all negative initial values to the final values λ'_k 's by increasing the neuron gain g , such that the energy function (3) which is initially a convex function of y , is transformed gradually into a concave function. The initial neuron gain g_0 must be chosen such that

$\lambda_k(g_0) < 0, \forall k$. After the state is initialized to $x = x(0)$, the initial gain at time = 0 can be set [0 a very small, positive value such that $0 < g(0) \ll 1$. Therefore, the network can be effectively linearized. It then increases continuously for $0 < t \leq T_a$ to the nominal gain of 1 required by the cellular network. Regardless of initial state values, the network results in the optimal solution at which its energy is minimized globally. Despite of the time-varying nature of the hardware annealing method, the stability of the network is still maintained as long as the gain control function $g = g(t)$ are non-negative.

Figure 3.3 shows plots of the energy landscape for increasing neuron gain values. The energy function is convex, indefinite, and concave when the gain is in the range $0.1 < g \leq 0.4$, $0.4 < g \leq 2/3$, and $2/3 < g \leq 1$, respectively. The solution moves toward the global minimum point as the gain value increase. The global minimum is guaranteed when the gain value reaches 1.

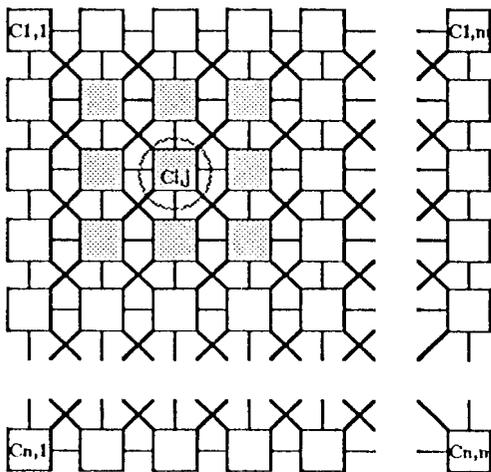


Fig. 3.1. An n-by-m OCNN on rectangular grid. The shaded boxes are the neighborhood cells of $C(i,j)$.

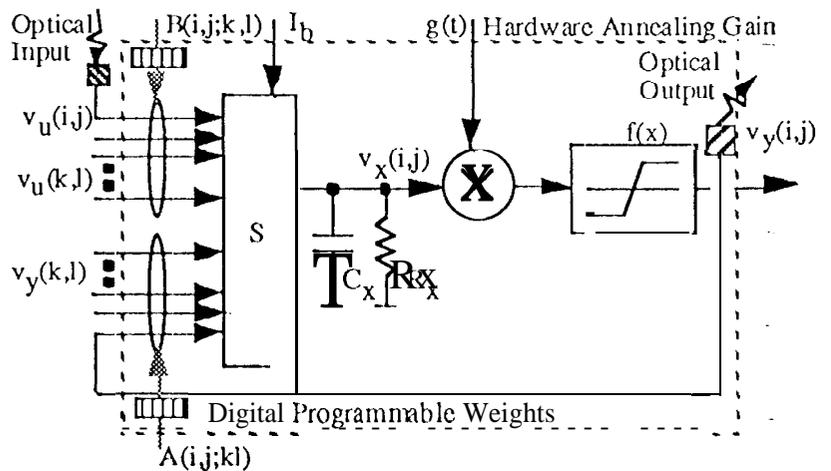


Fig. 3.2. Model of the OCNN neuron $C(i,j)$.

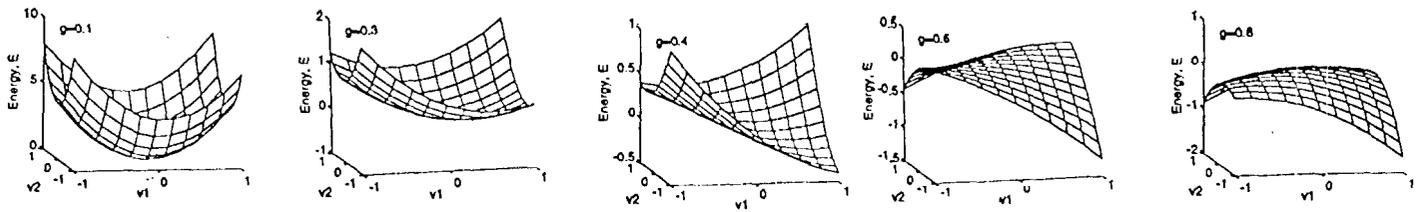


Figure 3.3. A sequence of snapshot the energy landscape for increasing neuron gain values from 1.0 to 0.8.

3.2 .Two-Neuron OCNN Experiment

To clearly illustrate the local-minimum issue in a network, a two-neuron network as shown in Figure 3.4 is considered. For a symmetric cellular neural network, a Lyapunov energy function can be found. The generalized energy function associated with the network possesses many local-minimum points which could trap the network into unwanted solutions when the network is used for optimization applications.

Assume that the piecewise-linear transfer function is used. The integration capacitance C_x is normalized to a unity, the feedback weights are symmetric such that $A_{1,1} = A_{2,2} = A_0 > T$, $A_{1,2} = A_{2,1} = A_1$, the feed-forward weights $B_{1,1} = B_{2,2} = 1$, and the network bias $I_b = 0$. Then, the generalized energy function of (2) can be simplified as

$$E_c = \frac{1}{2} \begin{bmatrix} v_{y1} \\ v_{y2} \end{bmatrix}^T \begin{bmatrix} A_0 - T_x & A_1 \\ A_1 & A_0 - T_x \end{bmatrix} \begin{bmatrix} v_{y1} \\ v_{y2} \end{bmatrix} - \begin{bmatrix} v_{y1} \\ v_{y2} \end{bmatrix}^T \begin{bmatrix} v_{u1} \\ v_{u2} \end{bmatrix}$$

where $-1 < v_{y1}, v_{y2} < +1$. After solving the equation $Mx = \lambda x$, the eigenvectors and eigenvalues of M can be obtained as $x_1 = [1 \ -1]^T$, $x_2 = [1 \ 1]^T$, and $\lambda_1 = A_0 - A_1 - T_x$, $\lambda_2 = A_0 + A_1 - T_x$, respectively. If the eigenvalues λ 's are positive and the bias is quite small, then the minimal occur at all corners.

The lowest energy function value corresponds to the global minimum while the others correspond to local minimal. A Matlab program was written to simulate the two-neuron network with $A_0 = 2$ and $A_1 = -0.25$. This simulation is performed with four different initial condition points, namely point A = (0.4, 0.2), point B = (0.5, 0.7), point C = (-0.9, -0.8), point D = (0.5, 0.6). The trajectories of output values for un-annealed and annealed cases with $VU = 0.2$ and $v_{u2} = -0.5$ are shown in Figure 3.5 (a), (b).

An HSPICE circuit simulation based on the electronic neurons described in Section 4 was performed. An initial condition set-up time of $1 \mu s$ was used. The time constant is $0.3 \mu s$. In the simulation, an annealing period of $15 \mu s$ (50 time constants) was employed. The results are shown in Figure 3.6. The outputs of both neurons, the state variables of both neurons and the annealing control signal are plotted. In the un-annealed case, the stable output is (-1, -1) for initial condition C, while in the annealed case, the stable output (1, -1). The whole optimization process completed with 50 time constants. The measure results are shown in Fig. 3.7.

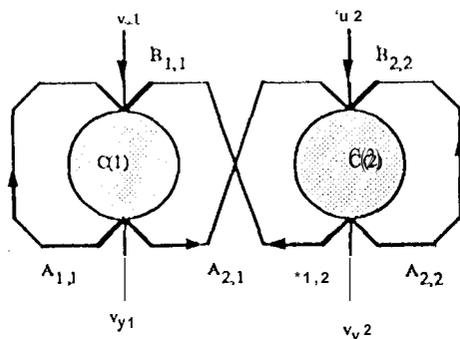


Fig. 3.4. Block diagram of a two-neuron OCNN with $B_{1,1} = B_{2,2} = 1$ and $B_{1,2} = B_{2,1} = 0$

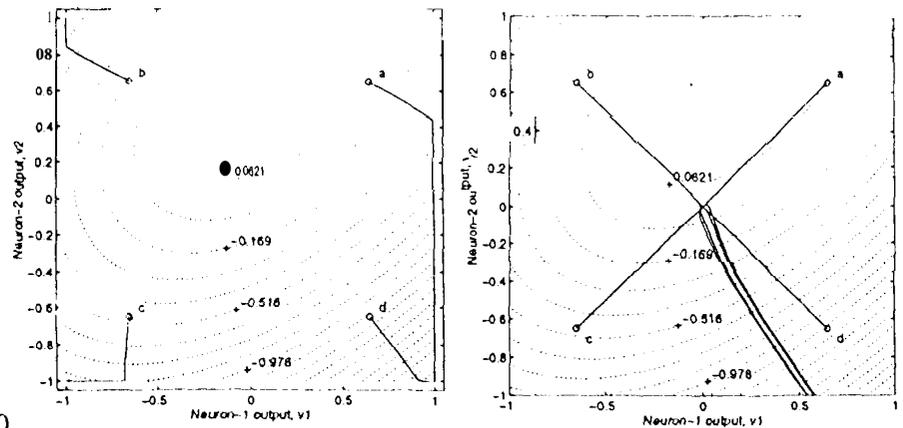


Fig. 3.5. Trajectories of optimization process. (a) Un-annealed case. (b) Annealed case.

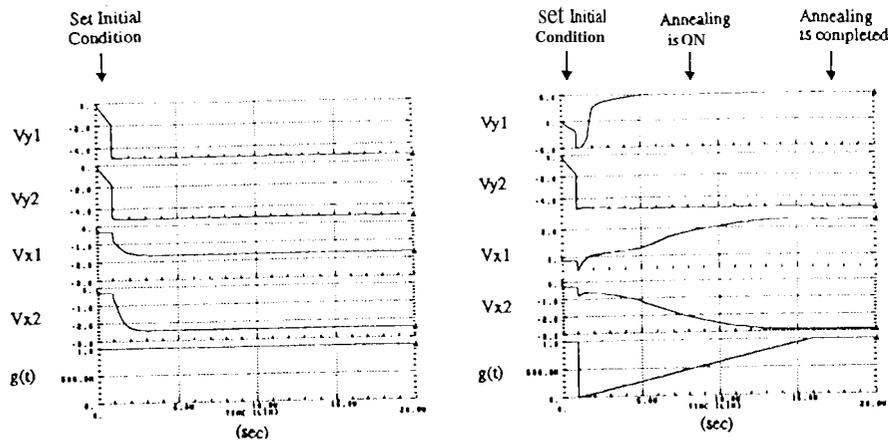


Fig. 3.6. HSPICE simulation result. (a) Un-annealed ease. (b) Annealed case at $g(t) = 1$ V

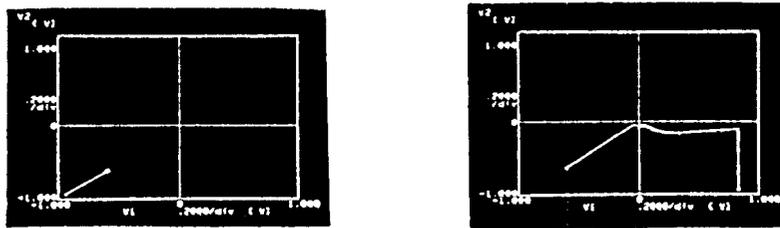
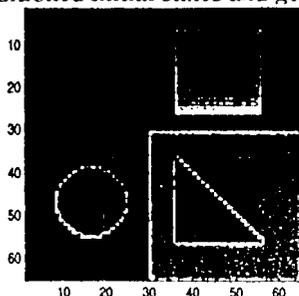


Fig. 3.7. Measurement results of network operation. (a) Un-annealed case. (b) Annealed case.

3.3. Applications

To illustrate the OCNN prime performance, the edge detection results of a 64-by-64 OCNN for un-annealed and annealed conditions is shown in Figure 3.8. The hardware annealing provides enough stimulation to those frozen neurons caused by such ill-conditioned initial states and gets better results.



(a) Original 64x64 gray-scale test image.

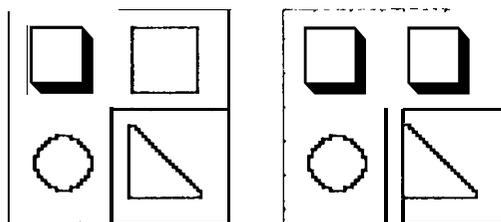
TA

| | | |
|---|---|---|
| 0 | 0 | 0 |
| 0 | 2 | 0 |
| 0 | 0 | 0 |

TB

| | | |
|-------|-------|-------|
| -0.25 | -0.25 | -0.25 |
| -0.25 | 2 | 0 |
| -0.25 | -0.25 | -0.25 |

(b) Feedback and Feedforward cloning templates.



(un-annealed operation) (annealed operation).
(c) Zero-initial conditions: $v_x(0)=0$, v_u = input image.



(un-annealed operation) (annealed operation).
(d) Ill-initial conditions: $v_x(0)>1$, v_u = input image.

Fig. 3.8. Edge detection application of a 64-by-64 CNN.

4. A COMPACT CURRENT-MODE VLSI NEURAL PROCESSOR DESIGN

4.1. Neuroprocessor Architecture and Operation

Figure 4.1 shows a complete OCNN multiprocessor architecture which is an $n \times m$ neuroprocessor array arranged in an n -by- m rectangular grid with appropriate interconnections. The block diagram of a neuroprocessing element is shown in Figure 4.2. Each OCNN neuroprocessing element consists of a core neuron cell, synaptic weights, input/output circuits, and digital interface. The network considered in this design is a continuous-time, rectangular-type OCNN with $r = 1$. A multilayer OCNN can be realized in a time-multiplexed fashion or in a multiple OCNNs configuration.

The digital interface provided are control data buses and read data lines. Four control buses for weights $a_0, a_1, a_2,$ and b_0 are 5-bit wide each. On the other hand, a read data line is common to all the cells in a row, and a column select line is activated at a time to read cell outputs in a specific column. The data for synaptic weights are written into operator registers and the network outputs are fetched from the cell output latches to the output buffer registers in column parallel. Depending on the applications, $x(0)$ can be initialized to zero, a scaled v_{μ} , or the weighted external inputs. One terminal of the capacitor C is switched to the voltage $x(0)$ during the initialization operation. At the same time, the outputs of the synapses go into the high-impedance state by control signal ϕ and ϕ , and the state node is connected to the ground to avoid possible spurious operation caused by the closed loop with the parasitic capacitance at the state node.

4.2. Building Blocks and Circuit Design

This section briefly describes the building blocks and associated circuits based on the current-mode approach [13] for a low-power compact OCNN neuroprocessor implementation.

Programmable synapses: The digital programmable synapse is realized using a binary-weighted current source array. Programmability of -3.75 to 3.75 for each synaptic weight is provided. This synaptic weight function can achieve 5-bit programmability and a resolution of higher than 8 bits.

Transimpedance multiplier: The hardware annealing is performed by the pre-multiplication of the state v_{xij} by the gain control function g before the nonlinear function $f(x)$ takes place. The basic element of the proposed circuit is the double-MOS differential resistor operating in triode region

Summing circuit: If the summation of the weighted currents from the neighboring cells is carried out directly in the transimpedance multiplier, the value of resistance R_x is inversely proportional to the gain-control voltage V_c . In order to accommodate a constant R_x , the constant input impedance current inverter is used at the input stage of the multiplier.

Nonlinear function: The circuit for the nonlinear function $y = f(x)$ is accomplished by a simple transconductor consisting of a differential amplifier. Its large signal transfer function is a smooth, sigmoid-like characteristics. A weak positive feedback is applied to increase the transconductance value without increasing the (W/L) ratio of the differential-pair transistors.

optical receiver: A vertical CMOS-BJT is utilized as a phototransistor and further integrated with two bipolar transistors in a Darlington configuration. A 60x60- μm Darlington-phototransistor provides a 20 mA with 100 dB dynamic range [16],

4.3. Current-Mode VLSI OCNN Neural Chip: Prototype and Demonstration

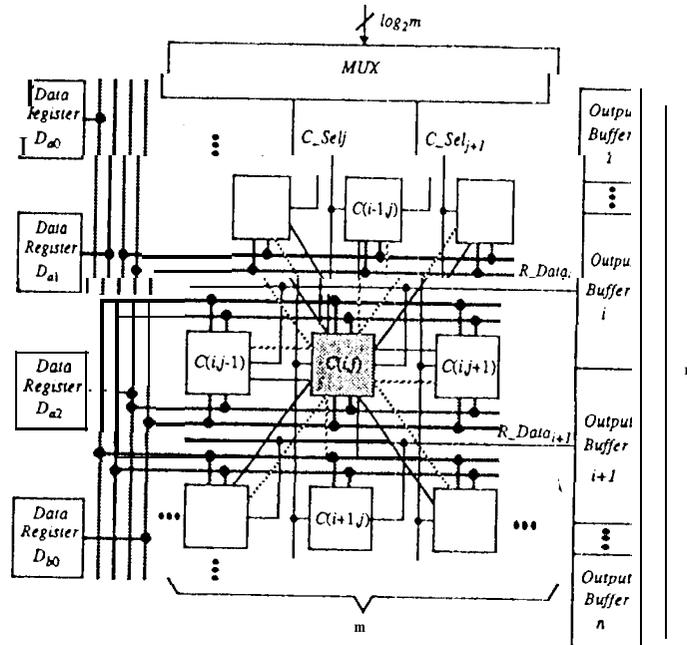
To illustrate the implementation feasibility, a programmable 5x5 cellular neural processing chip of active dimensions 1380 μm x 746 μm was designed and fabricated in a 2- μm CMOS technology through MOSIS Services. The die photo and the layout design of each neuron is shown in Fig. 4.3. There are a total of 11,250 transistors on the chip. The power consumption is 989 μW per neural cell, and that for the whole active array is less than 25 mW. A cell density of 505 cells/ cm^2 is achieved.

If an ultra-low-power sub-0.25 μm SOI-CMOS technology is used, a cell density of 16,160 cells/ cm^2 can be achieved. A network of 512 x 512 annealed neurons can be realized in a 2.9 cm x 2.9 cm chip. A network of 1024 x 1024 annealed neurons is feasible to be designed with 4 512x512 OCNN chips and packaged into a 3-D die stack.

A circuit board was built to demonstrate the operation of this prototype chip. The output arc connected to an array of LED displays. Experiments on edge detection were performed. The templates for edge detection operation are

$$T_A = \begin{bmatrix} 0 & -0.5 & 0 \\ -0.5 & 2 & -0.5 \\ 0 & -0.5 & 0 \end{bmatrix}, \text{ and } T_B = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The input pattern and the measured result of the edge detection experiment are shown in Fig. 4.4(a) and Fig. 4.3 (b), respectively. The measured result agrees well with the C-based simulation result. The CPU time for the C-based simulation is 2.53 seconds. The speedup is about 160,000.



4.1. An OCNN multiprocessor implementation based on an nxm neuroprocessor array architecture.

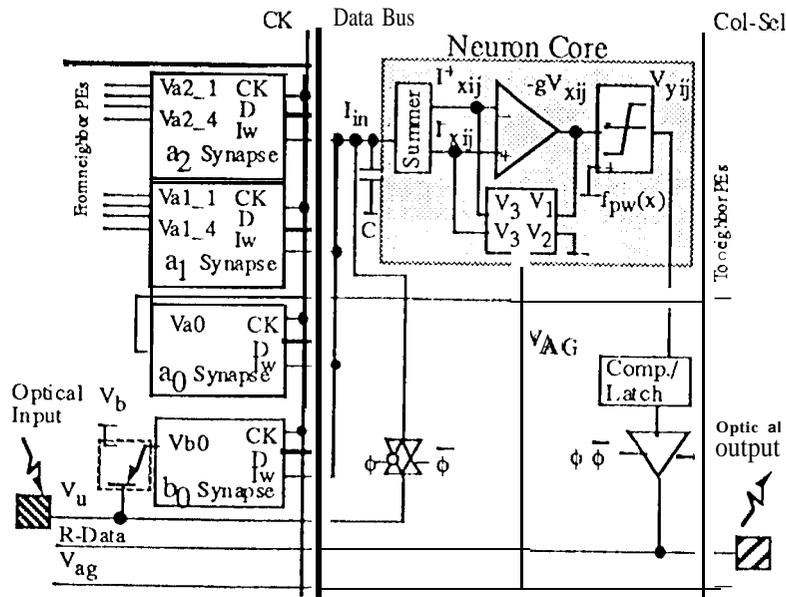


Fig.4.2. An OCNN unit consists of a core neuron cell, synaptic weights, I/O circuits, and digital interface.

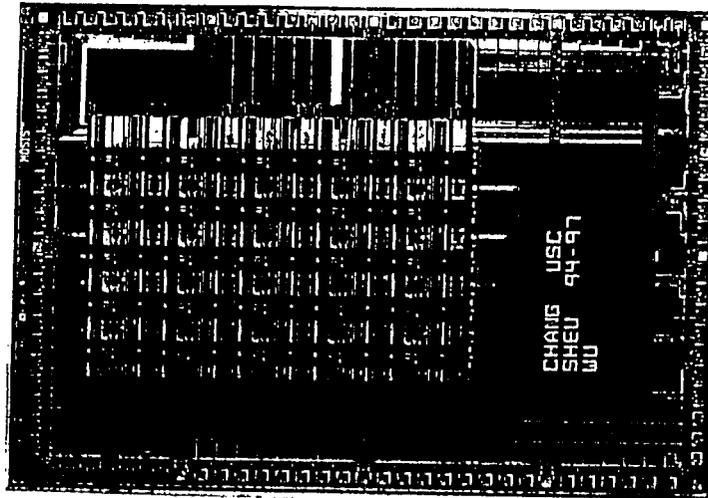


Fig. 4.3(a). A 2-µm CMOS die photo of the OCN chip with 5x5 neural cells,

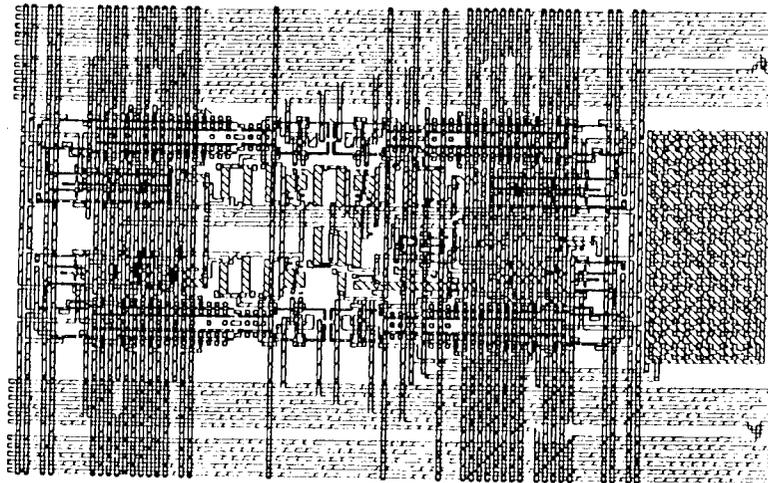


Fig. 4.3 (b). A 2-µm CMOS layout design of each OCN neural cell.

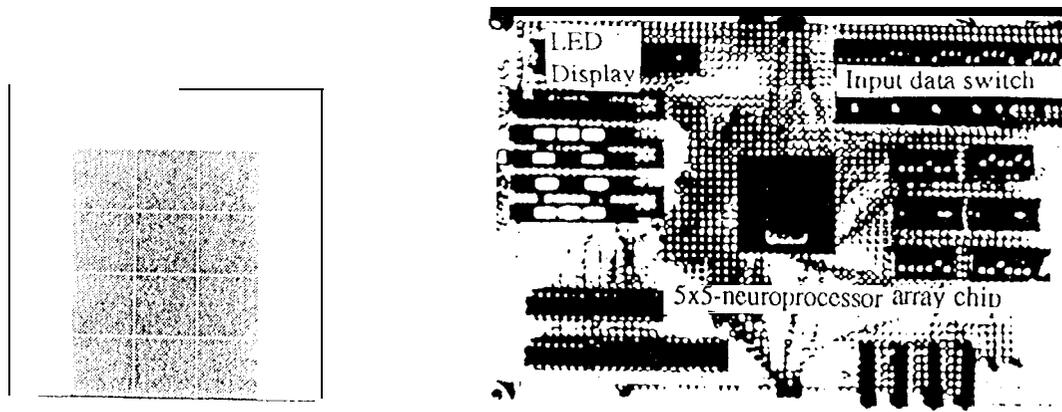


Fig. 4.4. Edge detection function, (a) Initial condition. (b) Board-level experimental result.

5 . CONCLUSIONS

Different system computing architectures of an OCNN processing engine were studied for microspacecraft on-board computing applications. Integration of an OCNN neuroprocessor array into the IMI Brain offers orders-of-magnitude computing performance enhancements for on-board intelligent multisensory processing and autonomous control tasks.

The OCNN algorithmic property for optimization solution is explored and illustrated. Important features of the OCNN implementation are well developed which include effective VLSI neural computing architecture, programmability for various algorithms, global optimization capability, compact neurons and synapses, fast data manipulation and routing, photo-sensors for front-end image acquisition, multisensory parallel interface, low power consumption, high cell integration density, and low manufacturing cost.

A compact current-mode VLSI 5x5-array OCNN chip was developed in a 2- μm CMOS to illustrate the implementation feasibility. A 3-D VLSI stack of dimensions 3 cm x 3 cm x 0.5 cm is projected to accommodate a complete 1024x1024-array OCNN embedded processor with one 1024x1024 active-pixel sensor on the top of the die stack. A miniaturized highly-integrated IMI Brain is therefore feasible to be implemented into a compact cube at a manageable power dissipation rate.

6 . ACKNOWLEDGMENTS

The research described in this paper was partially carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. Prof. Bing Sheu were supported by ONR under NOOO 14-94-1-0568. Valuable discussions with Dr. Robert Chang and Eric Chou are highly appreciated.

7. REFERENCES

- [1] D. Collins, et. al., "Second Generation Microspacecraft: Final Reports for 1994", JPL D- 12645, Rev A. 1994.
- [2] I. Alkalai, J. Klein, M. Underwood, "The New Millennium Program Microelectronics Integrated Product Development Team Technology Roadmap", JPL D-13276, November, 1995.
- [3] Workshop Proceedings on Future NASA Miniature Spacecraft Technology, Pasadena, CA, Feb. 1995.
- [4] D. Barbe, Ed., Smart Sensors, Proceeding of the SPIE, Vol. 178, April 1979; Vol. 252, July 1980.
- [5] L. O. Chua, L. Yang, "Cellular neural network Theory", IEEE Trans. on Circuits and Systems, vol. 35, pp. 1257-1272, Oct. 1988.
- [6] L. O. Chua, L. Yang, "Cellular neural networks: Applications," IEEE Trans. on Circuits and Systems, vol. 35, pp. 1273-1290, Oct. 1988.
- [7] T. Roska, J. Vandewalle, Eds., Cellular Neural Networks, West Sussex; John Wiley & Sons, 1993.
- [8] B. J. Sheu, J. Choi, Neural Information Processing and VLSI, Kluwer Academic Publishers: MA, Jan. 1995.
- [9] B. J. Sheu, S. Bang, W.-C. Fang, "optimal solutions of selected cellular neural network applications by the hardware annealing method," Proceedings of the third IEEE International Workshop on Cellular Neural Networks and their Applications (CNNA-94), pp. 279-284.
- [10] H. Halonen, V. Porra, et al., "programmable analog VLSI CNN chip with local digital logic," Proc. IEEE Int'l Symp. on Circuits and Systems, pp. 1291-1294, 1991.
- [11] B. J. Sheu, S. Bang, W.-C. Fang, "VLSI design of cellular neural networks with annealing and optical input capabilities," Proceedings of the IEEE International Symposium of Circuit and System (ISCAS-95), April 1995.
- [12] W.-C. Fang B. J. Sheu, S. Bang, "Compact Receivers and Smart-Pixel Chips for Optical Interconnections and Signal Processing," Proceedings of SPIE international Symposium on Optoelectronics, San Jose, CA, Feb. 4-10, 1995.
- [13] A. Rodriguez-Vazquez, et al., "Current-mode techniques for the implementation of continuous- and discrete-time cellular neural networks," IEEE Trans. on Circuits and Systems, vol. 40, pp. 132-146, Mar. 1993.
- [14] A. Cichocki, R. Unbehauen, Neural Networks for Optimization and Signal Processing, New York, NY; John Wiley & Sons, 1993.
- [15] Fossum and P. Wong, "Future prospects for CMOS active pixel sensors, " 1995 IEEE Workshop "on CCDs and Advanced Image Sensors, Dana Point, CA, April 20-22, 1995.
- [16] C. A. Mead, Analog VLSI and Neural Systems, Addison-Wesley Publishing Company, MA, 1989.