

1995-118837

40904
P.9

From Neural-Based Object Recognition toward Microelectronic Eyes

Bing J. Sheu, Ph.D. Senior Member, IEEE
Sa Hyun Bang, Ph.D. Student Member, IEEE

Department of Electrical Engineering, Powell Hall-604
University of Southern California, Los Angeles, CA 90089-0271, U.S.A.
Also with Center for Neural Engineering & the Signal and Image Processing Institute

Abstract

Engineering neural network systems are best known for their abilities to adapt to the changing characteristics of the surrounding environment by adjusting system parameter values during the learning process. Rapid advances in analog current-mode design techniques have made possible the implementation of major neural network functions in custom VLSI chips. An electrically programmable analog synapse cell with large dynamic range can be realized in a compact silicon area. New designs of the synapse cells, neurons, and analog processors are presented. A synapse cell based on Gilbert multiplier structure can perform the linear multiplication for back-propagation networks. A double differential-pair synapse cell can perform the Gaussian function for radial-basis network. The synapse cells can be biased in the strong inversion region for high-speed operation or biased in the subthreshold region for low-power operation. The voltage gain of the sigmoid-function neurons is externally adjustable which greatly facilitates the search of optimal solutions in certain networks. Various building blocks can be intelligently connected to form useful industrial applications. Efficient data communication is a key system-level design issue for large-scale networks. We also present analog neural processors based on Perceptron architecture and Hopfield network for communication applications. Biologically inspired neural networks have played an important role towards the creation of powerful and intelligent machines. Accuracy, limitations, and prospects of analog current-mode design of the biologically inspired vision processing chips and cellular neural network chips are key design issues.

I. Introduction

Rapid progresses in the research of intelligent information processing paradigms, architectures, and electronic hardware implementations based on artificial and biologically-inspired neural network models have helped to establish a rich knowledge base for practical applications. Studies of engineering neural network models were motivated by the investigation of human perceptron. The Von Neumann computing approach incorporates a single central processing unit and the main memory unit. It can execute instructions sequentially with a reasonable speed and accuracy for conventional data-processing applications. However, these digital machines, when packaged in a small physical size, can not perform computationally-intensive tasks with satisfactory performance in such areas as intelligent perceptron, including visionary and auditory signal processing, recognition, understanding, and logical reasoning where human being and even living animals can do a superb job.

Recent advances in artificial and biological neural networks research have provided excited evidence for high-performance information processing with a more efficient use of computing resources. The secret lies in the design optimization at various levels of computing and communication. Each neural network system consists of massively paralleled and distributed signal processors with every processor performing very simple operations. Large computational capabilities of these systems are derived from collectively parallel processing and efficient data routing through well-structured interconnection networks. Two different operation modes are associated with a typical neural information processing network: the data retrieving process and the learning process.

II. General Properties

Many important issues need to be carefully addressed in constructing electronic neural network systems:

1. A balanced exploration on the computing algorithms and architectures which are suitable for digital VLSI implementations and analog networks;
2. Emphasis of both artificial neural networks and biologically-inspired neural models; and
3. Solving real-world, large-scale problems.

In electronic implementation, the options are digital, analog, a combination of both, or pulsed-stream forms. Analog approaches can be divided into continuous-time [1, 2, 3], and discrete-time schemes [4, 5]. In continuous-time analog VLSI, some additional options arise relating to the operation mode of transistors: weak inversion [6] and strong inversion [7]. The pulsed-stream approach [8] is more biologically motivated than other approaches. Lyon and Mead [9] described the VLSI implementation of an analog electronic cochlea for speech recognition. Koch et al. [10] reported a real-time chip for computer vision and robotics. Satyanarayana et al. [11] presented a reconfigurable analog VLSI neural chip for general-purpose applications. Hollis and Paulos [12] proposed a current-summing neuron with binary data registers. Boser and Sackinger [13] presented an analog neural chip for hand-written character recognition. Fang, Sheu, et al. [14] presented a mixed-signal neural network processor chip for self-organizing networks.

There are three basic neural network architectures: the iterative networks, the multi-layer perceptron networks, and the self-organizing networks. The iterative neural networks, which are also called recurrent neural networks, are promising for temporal pattern recognition and generation. Recurrent neural networks can solve optimization problems because of their constraint-satisfaction capabilities. Data is retrieved from an iterative network through associative recalling. Representative iterative networks include the Hopfield network [15] and bidirectional associative memory [16]. In a multi-layer perceptron network, supervised learning [17] is used. The effective errors for the output layer and hidden layers are calculated from the actual outputs and expected outputs. Synapse weights are updated according to the delta rules or the derivatives. Layered neural networks are effective for spatial pattern recognition. The multi-layer perceptron networks are widely used in industrial applications.

A self-organizing network consists of two layers of neurons: the input layer and the competitive layer, which is also called the output layer [18]. A winner-take-all function is performed among the neurons in the competitive layer. The self-organizing network has the desirable property of effectively producing spatially organized presentation of various features of the input signals [19]. Competitive learning depends on the competition among the output neural units. Self organization is required in several image and vision processing applications such as pattern recognition, vector quantization for image compression, and motion estimation. In addition, it may be applied in the selection of optimal inference paths in symbolic computers. Such an application can systematically reduce the knowledge inference operation from an NP complete problem to a much simplified problem in a very efficient way.

III. Analog Building Blocks

Power consumption, required silicon area, and the number of packaged pins are also important figures of merit in practical hardware implementation. The required silicon area for a given function will be gradually decreased with the advances of microelectronic fabrication technologies. Therefore, the number of packaged pins for information communication could become a fundamental limitation for information exchange. Each package pin can be shared by several functional outputs through time-multiplexing scheme or frequency-multiplexing scheme.

A. Memory in Synapse Cells

An important component in hardware implementation of learning is memory. In analog neural network processor chips, synapse weight information can be stored in various formats. In the early design, fixed-resistance synapses were implemented with the well regions or an amorphous-silicon layer. Complementary-MOS transmission gates were also proposed to achieve programmable synapse resistance. Continuous-time synthesized resistance [20] is

made of four MOS transistors which are connected in a cross-coupled fashion. The threshold voltage mismatch effect is minimized by using symmetric control voltage.

A basic transconductance amplifier which is made of five MOS transistors requires a simple control signal for the programmable synapses [8]. Such a compact and programmable synapse provides the first- and third-quadrant multiplication capability. The synapse weight can be stored on the gate capacitance and refreshed periodically. A modified wide-range Gilbert multiplier is suitable for general-purpose programmable synaptic operation because it provides four-quadrant multiplication capability [21]. Long-term memory information can be stored in the floating-gate devices fabricated by a special EEPROM technology [22] or by a conventional double-polysilicon technology for analog circuits for over 20 years in room temperature [23].

B. Neurons

The summed synaptic current is converted to the voltage through a current-to-voltage converter. The feedback resistance of the converter can be implemented with six MOS transistors. The voltage gain of the neurons can be controlled continuously to perform the hardware annealing operation [24, 25] for the quick searching of optimal solutions in nonlinear optimization applications. Such a hardware implementation of mean-field annealing can be used in recurrent neural networks and multi-layered perceptron networks to avoid local minima problems.

C. Winner-Take-All Circuit

A high-precision VLSI winner-take-all circuit can achieve high-speed operation by biasing transistors in the strong-inversion region. It uses the cascade configuration to significantly increase the competition resolution and maintain a high speed operation for a large-scale network. The total bias current increases in proportion to the number of circuit cells so that a nearly constant response time is achieved. In addition, a unique dynamic current steering method is used to ensure only a single winner exists in the final output. Experimental results of the prototype chip fabricated by a 2- μm CMOS technology show that a cell can be a winner if its input is larger than those of the other cells by 15 mV. The measured response time is around 50 nsec at a 1-pF load capacitance. This analog winner-take-all circuit is a key module in the competitive layer of self-organization neural networks.

D. Radial-Basis Function Circuit

The circuit schematic diagram and transistor sizes for a Gaussian function synapse cell is shown [26]. This circuit consists of MOS differential pair and several arithmetic computational units in the current-mode configuration. Transistors with non-minimum channel lengths are used to avoid the channel-length modulation effect. The input voltage is applied to the gate terminal of one transistor in the differential pair and the synapse weight value is stored on the capacitance at the gate terminal of the other transistor. Measured results of the Gaussian synapse cell are shown.

IV. Design Methodology

Mixed-signal VLSI implementation is suitable for novel signal processing applications such as image restoration [45] and optical flow computing [46]. The mixed analog-digital circuit design techniques are used to take advantages of efficient numerical computation in analog domain with long-distance communication in digital data bus. The multiplexed scheme can also be used to transmit signals over a long distance in an electronic system. Additional system-level integration results can be found in [47].

Hybrid approach using combined analog dynamics and digital logic represents very powerful and appealing design. For example, the programmable CNNs provide a new quality of artificial neural networks through a kind of analog software, a simple way to solve CNN algorithms. In our design, we give the network instructions and templates information just like we had done with the general-purpose CPU. The whole system will work like a SIMD machine and each local cell will execute the given commands to accomplish the functions we want. There are two distinct portions

but they both use the analog and digital circuits. One part is consisted of global digital control circuits and global analog memory; the other one has one duplications in each local cell which contains small local control circuits and local analog and digital memory. A timing diagram of the global digital circuit is shown in figure 8.

One other novel way to implement the neural network is a hybrid neurocomputer that utilized electro-optic components for the input processing and analog electronics for implementation of the remainder of the transfer function. This type of neurocomputer was shown to be capable of successfully implementing simple Hopfield neural networks with weight values restricted to the set $\{-1, 0, +1\}$. B. Soffer et. al also developed a first all-optical neurocomputer [27].

V. Cellular Neural Network

1. General

A cellular neural network (CNN) is a continuous-time or discrete-time artificial neural network that features a multi-dimensional array of neuron cells and local interconnections among the cells. The basic CNN proposed by Chua and Yang [28, 29] in 1988 is a continuous-time network in the form of an n -by- m rectangular-grid array where n and m are the numbers of rows and columns, respectively. However, the geometry of the array needs not to be rectangular and can be such shapes as triangle or hexagon [30]. A multiple of arrays can be cascaded with an appropriate interconnect structure to construct a multi-layered CNN. Structural variations of the continuous-time, shift-invariant, rectangular-grided network include discrete-time CNN [31], CNN with nonlinear and delay-type templates [32], etc. CNN and its variations provide a natural and universal model of analog processor arrays on a geometrical grid. Their local connectivity and regular structure appear most efficient for electronic implementation for high-speed, real-time applications. Several hardware implementations of the CNN have been reported in the literatures [33]-[39].

2. Hardware Annealing

The hardware-based annealing technique [25], has an analogy to the metallurgical annealing in the metallurgy and simulated annealing in the Boltzmann machine, which are the optimal stochastic procedures. It is a paralleled, electronic version of the deterministic mean-field learning rule [42, 43] directly incorporated with the Hopfield neural network or CNN. It is a dynamic relaxation process for finding the optimum solutions in the recurrent associative neural networks such as Hopfield network and CNN. Even with a correct mapping of the cost function onto a neural network, the desired combinatorial solution is not guaranteed because a concave optimization problem always involves a large number of local minima. True combinatorial solutions can be achieved by applying the hardware-based annealing technique with which the global minimum of E is found in a real-time speed.

3. Applications

The CNN's can be used in many computation-intensive, adaptive signal processing applications. Due to its two-dimensional array architecture, CNN's are suitable for real-time image processing applications in the following areas [30].

- (a) Image processing: Feature extraction, motion detection & estimation, path tracking, collision avoidance, and mage halftoning,
- (b) 3-D surface analysis: Min/max detection and gradient estimation,
- (c) Solving partial differential equations,
- (d) Non-visual data imaging: Thermographic images, antenna array images, and medical maps and images.

A CNN has similar collective computational behaviors with Hopfield neural networks. Thus, the quadratic nature of the Lyapunov function allows us to map it into optimization problems [41, 43].

VI. Conclusion

There is a strong need to develop new neural network architectures and design techniques to extend the size of electronic implementation to a larger scale for solving real-world problems in science, engineering, and business. Extension of the hardware annealing to large-scale networks for complex problems is highly desirable. Chip-level and system-level packaging technologies will be crucial for future computing machines with one-million-unit neural networks on silicon wafers that interact with the external environment and change the structures adaptively. Reusable software modules and hardware modules are to be invented. For large scientific problems, neural networks with 10 tera connection updates per second will be needed. A flexible framework for representing various kinds of information efficiently and effectively will be the key for successful hardware/software co-designed systems.

Acknowledgement

The authors would like to thank Mr. Tony H.-Y. Wu for preparing some of the figures.

References

- [1] B. W. Lee, B. J. Sheu, "Design of a neural-based A/D converter using Hopfield network," *IEEE J. of Solid-State Circuits*, vol. 24, pp. 1129-1135, Aug. 1989.
- [2] B. E. Boser, E. Sačhinger, et al., "An analog neural network processor with programmable topology," *IEEE J. Solid-State Circuits*, vol. 26, pp. 2017-2025, Dec. 1992.
- [3] M. A. C. Maher, C. A. Mead, et al., "Implementing neural architectures using analog VLSI circuits," *IEEE Trans. on Circuits and Systems*, vol. 36, pp. 643-652, May 1989.
- [4] J. E. Hansen, D. J. Allstot, et al., "A time-multiplexed switched-capacitor circuit for neural network applications," *IEEE Int. Symp. on Circuits and Systems*, vol. 3, pp. 2177-2180, 1989.
- [5] R. Dominguez-Castro, E. Sanchez-Sinencio, et al., "Analog neural networks for real-time constrained optimization," *IEEE Int. Symp. on Circuits and Systems*, vol. 3, pp. 1867-1870, 1990.
- [6] C. A. Mead, et al., "Analog VLSI model of binaural hearing," *IEEE Trans. on Neural Networks*, vol. 2, pp. 230-236, Mar. 1991.
- [7] B. W. Lee, B. J. Sheu, "General-purpose neural chips with electrically programmable synapses and gain-adjustable neurons," *IEEE J. of Solid-State Circuits*, vol. 27, pp. 1299-1302, Sept. 1992.
- [8] A. Hamilton, et al., "Integrated pulse stream neural networks: results, issues, and pointers," *IEEE Trans. on Neural Networks*, vol. 3, pp. 385-393, May 1992.
- [9] R. F. Lyon, C. A. Mead, "An analog electronic cochlea," *IEEE Trans. on Signal Processing*, vol. 26, pp. 1119-1134, July 1988.
- [10] C. Koch, et al., "Real-time computer vision and robotics using analog VLSI circuits," *Advances in Neural Information Processing Systems 2*, pp. 750-757, Morgan Kaufmann, 1990.
- [11] S. Satyanarayana, Y. Tsividis, H. P. Graf, "A reconfigurable analog VLSI neural network chips," Editor: D. Touretzky, pp. 758-768, Morgan Kaufmann: San Matao, CA, 1990.
- [12] P. W. Hollis, J. J. Paulos, "Artificial neural networks using MOS analog multipliers," *IEEE J. Solid-State Circuits*, vol. 25, pp. 849-855, June 1990.

- [13] B. E. Boser, E. Sackinger, "An analog neural network processor with programmable network topology," *IEEE Tech. Digest of Inter. Solid-State Circuits Conference*, pp. 184-185, San Francisco, CA, Feb. 1991.
- [14] W.-C. Fang, B. J. Sheu, O. T.-C. Chen, J. Choi, "A VLSI neural processor for image data compression using self-organizing networks," *IEEE Trans. on Neural Networks*, vol. 3, pp. 506-518, May 1992.
- [15] D. W. Tank, J. J. Hopfield, "Simple 'neural' optimization networks: an A/D converter, signal decision circuit, and a linear programming circuit," *IEEE Trans. on Circuits and Systems*, vol. 33, pp. 533-541, May 1986.
- [16] B. Kosko, *Neural Networks and Fuzzy Systems*, Prentice Hall: Englewood Cliffs, NJ, 1992.
- [17] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representation by error propagation," in *Parallel Distributed Processing, vol. 1*, Eds. D. Rumelhart & J. McClelland, MIT Press: Cambridge, MA, 1986.
- [18] T. Kohonen, *Self-Organization and Associative Memory, 2nd ed.*, Springer-Verlag: New York, NY, 1988.
- [19] T. Kohonen, "The self-organizing map," *Proc. of IEEE*, vol. 78, pp. 1464-1480, Sept. 1990.
- [20] S. Bibyk, M. Ismail, "Issues in analog VLSI and MOS techniques for neural computing," in *Analog VLSI Implementation of Neural Systems*, Eds. C. Mead, M. Ismail, pp. 116-125, Kluwer Acad, 1989.
- [21] B. J. Sheu, J. Choi, C.-F. Chang, "An analog neural network processor for self-organizing mapping," *IEEE International Solid-State Circuits Conference*, pp. 136-137, 266, San Francisco, CA, Feb. 1992.
- [22] M. Holler, S. Tam, et al., "An electrically trainable artificial neural network (ETANN) with 10240 'Float gate' synapses," *Proc. IEEE/INNS Inter. Joint Conf. on Neural Networks*, vol. 2, pp. 191-196, Washington, DC, June 1989.
- [23] B. W. Lee, H. Yang, B. J. Sheu, "Analog floating-gate synapses for general-purpose VLSI neural computation," *IEEE Trans. on Circuits and Systems*, vol. 38, pp. 654-658, June 1991.
- [24] B. W. Lee, Bing. J. Sheu, *Hardware Annealing in Analog VLSI Neurocomputing*, Kluwer Academic Publisher: Boston, MA, 1991.
- [25] B. W. Lee, B. J. Sheu, "Paralleled hardware annealing for optimal solutions on electronic neural networks," *IEEE Trans. on Neural Networks*, vol. 4, no. 4, pp. 588-599, July 1993.
- [26] J. Choi, B. J. Sheu, C.-F. Chang, "A Gaussian synapse circuit for analog VLSI neural networks," *IEEE Trans. on VLSI Systems*, vol. 2, no. 1, Mar. 1994.
- [27] G.J. Dunning, E. Marom, Y. Owechko, B.H. Soffer, "Optical holographic associative memory using a phase conjugate resonator," *SPIE Proc.*, 625, Bellingham WA, Jan. 1986.
- [28] L.O. Chua, L. Yang, "Cellular neural network: Theory," *IEEE Trans. Circuits Syst.*, vol. 35, pp. 1257-1272, Oct. 1988.
- [29] L.O. Chua, L. Yang, "Cellular neural network: Applications," *IEEE Trans. Circuits Syst.*, vol. 35, pp. 1273-1290, Oct. 1988.
- [30] L.O. Chua, T. Roska, "The CNN paradigm," *IEEE Trans. Circuits Syst. Part I*, vol. 40, pp. 147-156, Mar. 1993.

- [31] H. Harrer, J. A. Nossek, "Discrete-time cellular neural networks," T. Roska, J. Vandewalle, Eds., *Cellular Neural Networks*, West Sussex; England, John Wiley & Sons, 1993.
- [32] T. Roska, L.O. Chua, "Cellular neural networks with non-linear and delay-type template elements and non-uniform grids," *Int J. Circuit Theory and Applications*, vol. 20, pp. 469-481, 1992.
- [33] J.M. Cruz, L.O. Chua, "A CNN chip for connected component detection," *IEEE Trans. Circuits Syst.*, vol. 38, pp. 812-817, July 1991.
- [34] A. Rodriguez-Vazquez, et al., "Current-mode techniques for the implementation of continuous- and discrete-time cellular neural networks," *IEEE Trans. Circuits Syst. Part II*, vol. 40, pp. 132-146, Mar. 1993.
- [35] J.E. Varrientos, E. Sanchez-Sinencio, J. Ramirez-Angulo, "A current-mode cellular neural network implementation," *IEEE Trans. Circuits Syst. Part II*, vol. 40, pp. 147-155, Mar. 1993.
- [36] H. Harrer, J.A. Nossek, R. Stelzl, "An analog implementation of discrete-time cellular neural networks," *IEEE Trans. Neural Networks*, vol. 3, pp. 466-476, May 1992.
- [37] I.A. Baktir, M.A. Tan, "Analog CMOS implementation of cellular neural networks," *IEEE Trans. Circuits Syst. Part II*, vol. 40, pp. 200-206, Mar. 1993.
- [38] G.F.D. Betta, S. Graffi, Zs.M. Kovacs, G. Masetti, "CMOS implementation of an analogically programmable cellular neural network," *IEEE Trans. Circuits Syst. Part II*, vol. 40, pp. 206-215, Mar. 1993.
- [39] M. Anguita, F.J. Pelayo, A. Prieto, J. Ortega, "Analog CMOS implementation of a discrete time CNN with programmable cloning templates," *IEEE Trans. Circuits Syst. Part II*, vol. 40, pp. 215-218, Mar. 1993.
- [40] B.W. Lee, B.J. Sheu, *Hardware Annealing in Analog VLSI Neurocomputing*, Norwell, MA: Kluwer Academic Publishers, 1991.
- [41] S. Bang, B.J. Sheu, "Optimal solutions for cellular neural networks by parallel hardware annealing," submitted for journal publication.
- [42] C. Peterson, J.R. Anderson, "A mean field theory learning algorithm for neural networks," *Complex Systems*, vol. 1, no. 5, pp. 995-1019, 1987.
- [43] C. Peterson, "Mean field theory neural networks for feature recognition, content addressable memory and optimization," *Connection Science*, vol. 3, pp. 3-33, 1991.
- [44] N. Fruehauf, E. Lueder, G. Bader, "Fourier optical realization of cellular neural networks," *IEEE Trans. Circuits Syst. Part II*, vol. 40, pp. 156-162, Mar. 1993.
- [45] J.-C. Lee, B. J. Sheu, J. Choi, R. Chellappa, "A mixed-signal VLSI neuroprocessor for image restoration," *em Trans. on Circuits and Systems for Video Technology*, vol. 2, no. 3, pp. 319-324, Sept. 1992.
- [46] J.-C. Lee, B. J. Sheu, W.-C. Fang, R. Chellappa, "VLSI neuroprocessors for video motion detection," *IEEE Trans. on Neural Networks*, vol. 4, no. 2, pp. 178-191, Mar. 1993.
- [47] E. Snachez-Sinencio, C. Lau, Eds., *Artificial Neural Networks*, IEEE Press: New York, 1992.

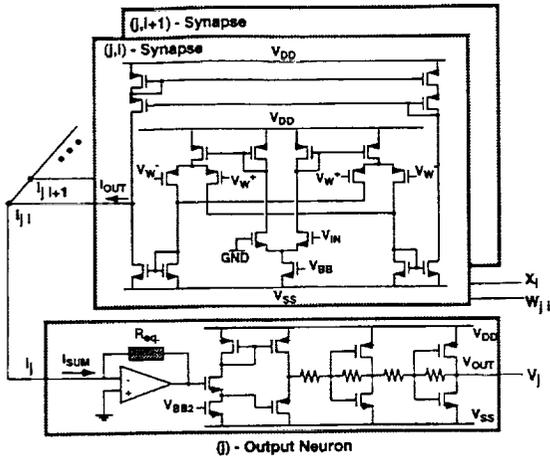


Fig. 1 Circuit schematic of the synapse cell and the output neuron.

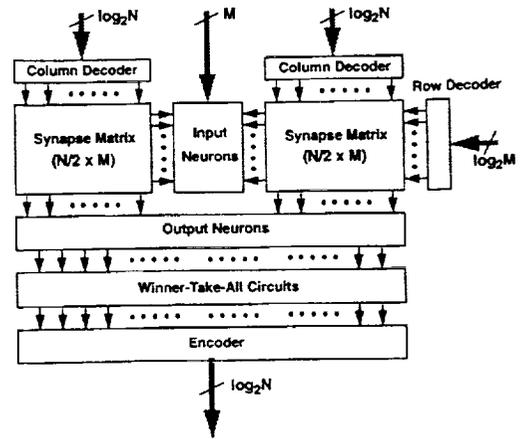


Fig. 2 Schematic diagram of a self-organizing analog neural processor.

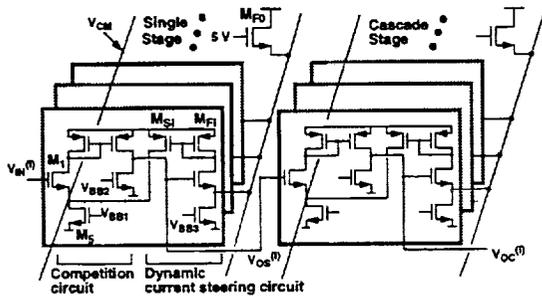
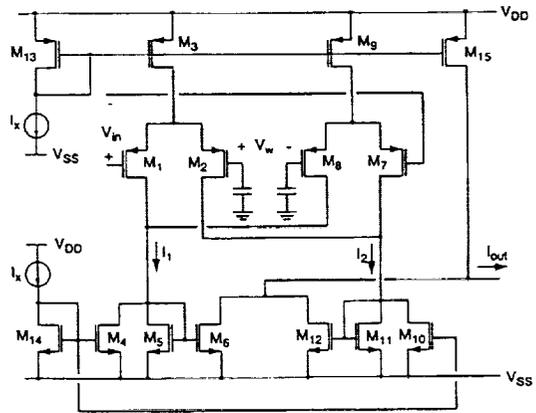


Fig. 3 Circuit schematic of the winner-take-all function.



(a) Circuit schematic diagram.

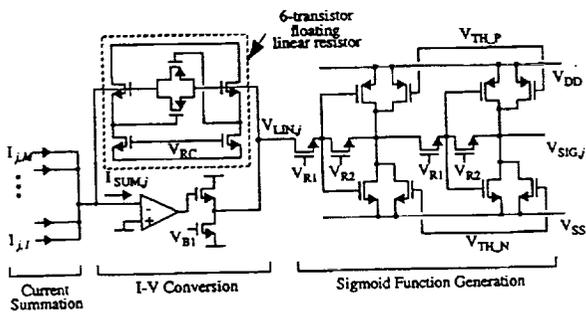
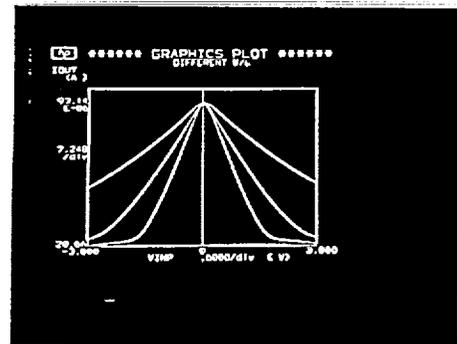


Fig. 5 Circuit schematic of neuron for multi-layered network.



(b) Measured results.

Fig. 4 The Gaussian function synapse cell.

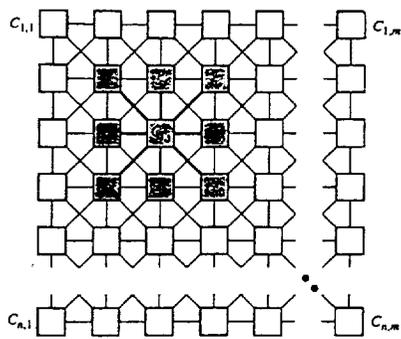


Fig. 6 Cellular neural network.

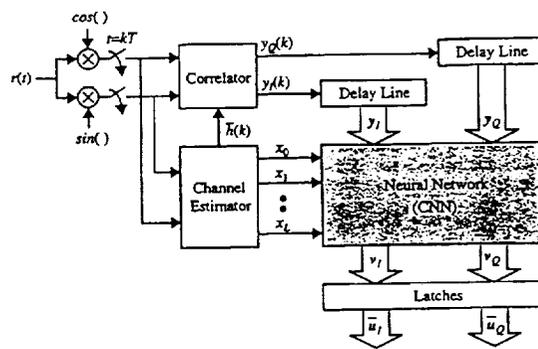


Fig. 7 MLSE application of CNN.

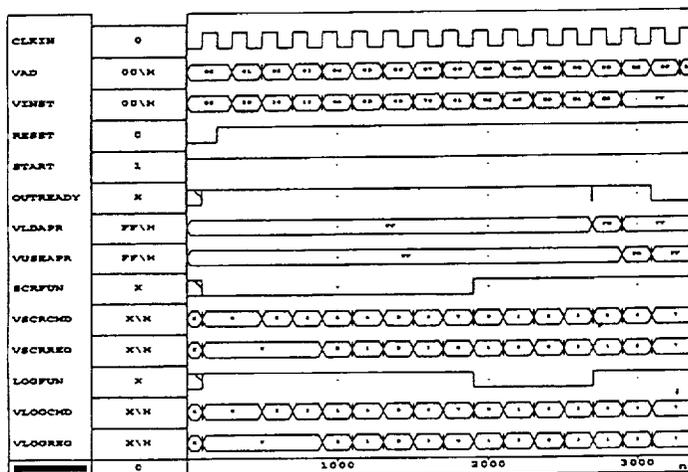


Fig.8 Timing diagram of global control circuit.

