Extended Abstracts of the 1994 International Conference on Solid State Devices and Materials, Yokohama, 1994, pp. 343-345 S-IV-2

Invited

Comparison between Optical and Electrical Implementation of Neural Networks

Kazuo Kyuma, Jun Ohta, and Yutaka Arima
Mitsubishi Electric Corporation, Semiconductor Research Laboratory
8-1-1 Tsukaguchi-Honmachi, Amagasaki-city, 661 Japan
TEL: Japan 6-497-7078, FAX: Japan 6-497-7288
E-MAIL: kyuma@qua.crl.melco.co.jp

This paper makes clear the features and the potential applications of three types of neurochips, which include VLSI analog-, digital-, and optical-neurochips. The recent progress on our digital neurochip and artificial retina chip is also reported in detail. Our digital neurochip contains twelve 24-bit floating point processing units in it and the execution speed of 1.2 GFLOPS is achieved. Our artificial retina integrates the image sensing and processing with high speed as fast as 300 μs. Their applications are also introduced.

1. Introduction

There has been much interest in neural networks for solving ill-defined problems because of their abilities of "learning" or "generalization" from examples. However, neural network computations are very time consuming if implemented on the conventional sequential computers. Therefore, the development of the neural network hardware is important for real time applications.

Until today, more than 50 types of neurochips have been reported. These chips are classified as electronic (VLSI) circuits and optical circuits. The electronic neurochips are classified into analog circuits and digital circuits. Among the neurchips reported so far, several chips with a very high speed performance, exceeding that of the conventional super-computers, have been commercially available at a fraction of their cost. The purposes of this paper are to make clear the features and the potential applications of these three types of neurochips, and to report our digital neurochips and artificial retina chips.

Table1 Hardware Classification and Comparison

Type	Components (Upper:Neuron / Lower:Synapse)	Features & Typical Performances	Applications
Digital Neurochip	Multiplier + Adder + Nonlinear Function Table / RAM, ROM, EPROM, etc.	General Purpose, High Accuracy / Several~64 Parallel PA Integrated, 1.2GFLOPS	General Purpose Neural Processor (Neurocomputer, Controller)
Analog Neurochip	Inverter, OP Amp., etc. / MOS Resister, Stored Carrier, etc.	Large Scale Integration, High Speed / 64~400 Neurons Integrated, 2TCPS	Sensor Information Processing, Space & Defense Equipments
Optical Neurochip	Light Source + Electronic Circuit + Photodetector / Variable Sensitivity Photo-detector, Hologram, Liquid Crystal SLM, etc.	Parallel Processing, Direct Image Processing / 128 Neurons Fully Connected, Artificial Retina Chip	Intelligent Sensor, Optical Interconnection

2. Comparison of Neurochips

The typical configurations, features, performance, and potential applications of the three types of neurochips are summarized in Table 1.

The digital neurochips offer many desirable features over the analog chips, such as high accuracy (16-32 bits), expandable capability, and fast processing (~GCPS). The disadvantages are necessity of large chip area. Since the digital circuits have the expandable capability to large-scale systems, the computations of almost all neural networks are performed by mapping the algolithm on the parallel processors. Thus the digital chips are expected to play a role as a general-purpose neural hardware that provides adaptive and flexible platform.

On the other hand, the analog neurochips are advantageous in terms of ultra-fast processing (~TCPS), robustness to hardware errors and compatibility with sensor outputs. The drawbacks are low accuracy (5-6 bits), lack of on-chip flexibility. Therefore the analog chips are not suitable for the general-purpose applications (BP model requires 12-16 bits accuracy). The analog chips may find the applications in special-purpose preprocessing fields such as early vision and speech recognition, using appropriate models without differential calculus. Based on this idea, we have developed fast (2 TCPS) and large-scale (400 neurons) analog chips for the Boltzmann machine model [1].

The VLSI neurochips suffer from the tradeoff between the circuit complexity and the integration density, and the I/O bottleneck between chips. The optics is expected to solve these problems. The optical devices can be oriented normal to the surface of the substrate in the form of 2-D arrays. This means that the light beams can connect optical devices on the arrays through 3-D free space in a massively-parallel way. Another advantage of optics is a direct image processing capability which leads to the novel intelligent image processing.

3. VLSI Digital Neurochip

The requirements for the general-purpose digital neurochips are high speed, large network expandability, and high accuracy operation. We have developed a neurochip[2] (NEURO4) that satisfies these requirements. The chip block diagram and the characteristics are shown in Fig.1 and Table 2, respectively.

The NEURO4 operates based on a single-instruction multi-data stream (SIMD) architecture and have twelve 24-bit floating-point processing-units (FP-PUs), a nonlinear function unit (NFU), and a control unit (CU). Each PU includes a 24b×1.28kw local memory and communicates with its neighbor through a shift register ring. The execution speed of 1.2 GFLOPS is achieved at 50 MHz clock frequency. Two external memory ports and a ring expansion port permit to construct the large-scale networks. The neural tasks over one million synapses can be executed using multi-chips together with memory chips. The FP representation is essential to the general-purpose hardware because the hardware should be used for various network models with different dynamic range and accuracies. To realize the FP-based hardware, we have implemented a 24 bit FP-adder and a 24 bit FP-multiplier for each PU. A neuro-computer with a few tenth of GFLOPS performance is under development for various kinds of applications such as quality control and inspection equipments, millitary and aerospace applications, robotics, and so on.

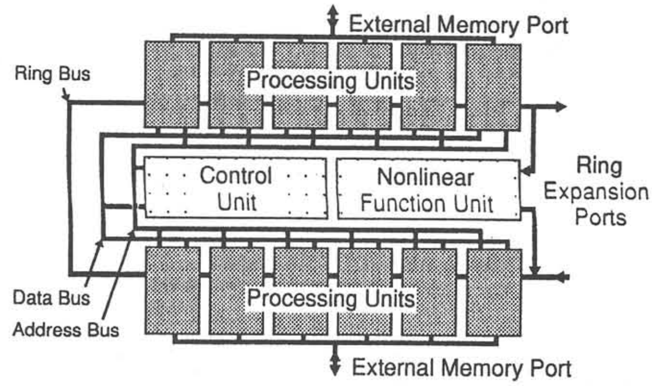

Fig1. Digital Neurochip Block Diagram

Table2 Digital Neurochip Characteristics

Process Technology	0.5 μm CMOS (1Poly-3Al)
Number of Transistors	3.4Million
Local Memories	24b×1.28kw×12
External Memories	24b×768kw max.
Accuracy	24b Floating-Point
Peak Performance	1.2GFLOPS (@50MHz)
Power Dissipation	7.1W (@50MHz, Vcc=3.3V)

4. Artificial Retina Chip

Figure 2 shows our GaAs-based artificial retina (RETINA) consisting of a 2-D variable sensitivity photodetector (VSPD) array[3]. The VSPD is a fundamental element of RETINA with three functions; as a conventional photodetector, a spatial light modulator, and an analog memory. The RETINA has the simultaneous functions of parallel sensing and parallel processing of the input image. The input image is directly projected on the chip as the matrix. The chip performs the matrix-vector multiplication, where the vector is an electric control signal applied to the VSPD electrodes. The chip performs a variety of processing, such as TV-camera like image sensing, edge extraction, Fourier transform, image storage, pattern matching, image compression, character recognition by changing the electric control signal. The experimental results of image compression for finger prints, which are made by the chip with 128×128 VSPD elements, are shown in Fig. 3. The processing time is shorter than 300 μs. The Retina has been recognized quite useful as an artificial "eye" in the moving vehicles because of the extremely fast response. For example, the Retina can predict the time to collision with an approaching object, avoid obstacles. More sophisticated chips with adaptation and learning capabilities, consisting of a 2-D VSPD array vertically integrated on a 1-D LED array[4], are also under development.

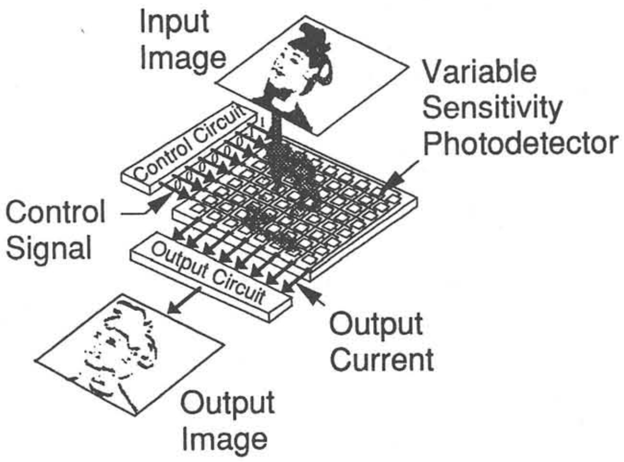

Fig.2 Artificial Retina Chip Scheme

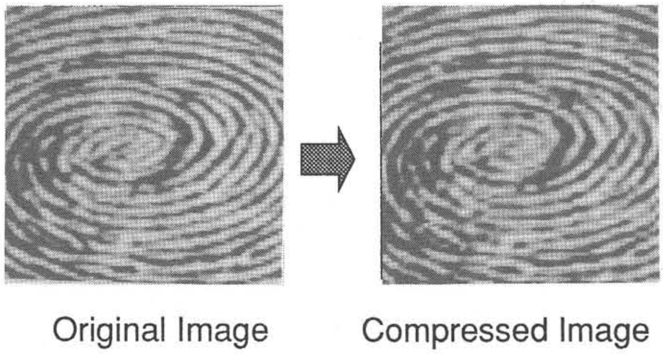

Fig.3 Image Compression of Fingerprint

5. Conclusion

The neurochips will be gradually utilized in the practical applications such as visual pattern recognition systems. In future, the neurochips are expected to play an important role as real time and intelligent processing elements in the super-information-highways and/or multi-media systems.

References

(1) Y. Arima, et al., IEEE J. Solid-State Circuits, 27, 12, 1854, 1992.
(2) Y. Kondo, et al., ISSCC DIGEST OF TECHNICAL PAPERS, 218, 1994.
(3) E. Lange, et al., Proceedings of IJCNN'93, 801, 1993.
(4) Y. Nitta, et al., Proceedings of IJCNN'93, 805, 1993.