Robust Bioinformatics Recognition with VLSI Biochip Microsystem

Jaw-Chyng L. Lue[1], and Wai-Chi Fang[2]
[1]Department of Electrical Engineering– Electrophysics, University of Southern California, Los Angeles, CA
[2]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA

***Abstract*—A microsystem architecture for real-time, on-site, robust bioinformatic patterns recognition and analysis has been proposed. This system is compatible with on-chip DNA analysis means such as polymerase chain reaction (PCR) amplification. A corresponding novel artificial neural network (ANN) learning algorithm using new sigmoid-logarithmic transfer function based on error backpropagation (EBP) algorithm is invented. Our results show the trained new ANN can recognize low fluorescence patterns better than the conventional sigmoidal ANN does. A differential logarithmic imaging chip is designed for calculating logarithm of relative intensities of fluorescence signals. The single-rail logarithmic circuit and a prototype ANN chip are designed, fabricated and characterized.**

I. Introduction

Microdevice-based genetic expression assay analysis for extracting genome sequencing information using PCR amplification and capillary electrophoretic (CE) techniques make portable and real-time DNA analysis instrument feasible [1]. Recently, the integration of on-chip PCR with electro-chemical (EC) transduction functionality for DNA amplification and detection was introduced [2]. A dual-labeled (*i.e.* for sample and reference channels) assay design is commonly used for identifying differentially expressed genes. This method also reduces the sources of variability due to aspects of individual spot that affects both specimens similarly [3]. Logarithm of relative intensities of two fluorescent dye labeled specimens measured at a single spot is calculated from the assay's fluorescence image for analysis. The image analysis task is typically processed pixel by pixel sequentially on a digital computer. Finding and analyzing desired genetic patterns in real time and on a hazardous or dangerous field is difficult for the existing techniques.

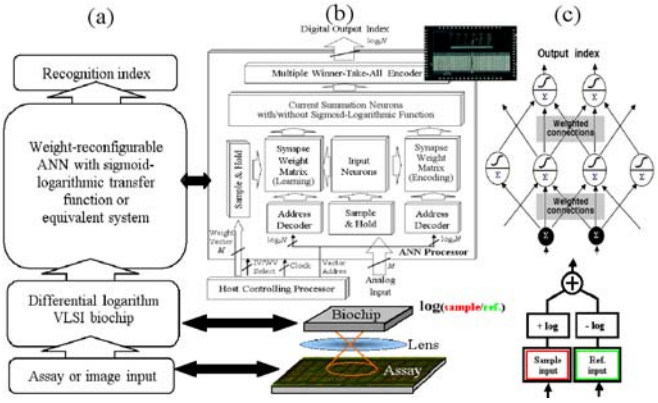

Fig. 1. (a) Hierarchy of the proposed biochip microsystem for genetic assay recognition. (b) A system-on-chip architecture design made of differential logarithm imaging chip and the weight reconfigurable ANN mix-signal chip. An optical micrograph of an ANN chip with 25 input neurons is shown at top right (4.6 mm by 6.8 mm). (c) Schematics of a three-layered ANN with new transfer function and the preceding differential logarithm

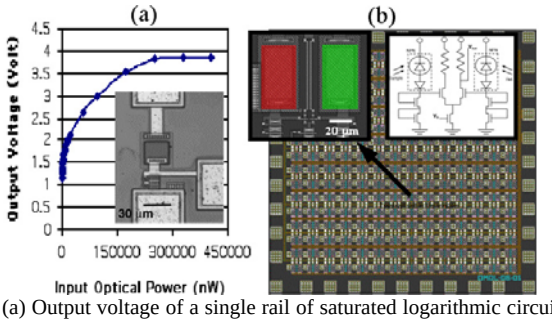

Fig. 2. (a) Output voltage of a single rail of saturated logarithmic circuit as a function of the input optical power (input wavelength: 830 nm). An optical micrograph of the single-rail logarithmic circuit is also shown. (b) Proposed layout of a 15 by 15 array of the differential logarithm circuitry (2.2 mm by 2.2 mm). An enlarged view of a single cell with pseudo color filter layers (red filter is for the sample fluorescence light; green filter is for the reference one.) and its schematics diagram are shown on top of the layout.

II. Biochip Module Architecture

We proposed a hardware microsystem suitable for real-time, on-site, robust genetic fluorescence data analysis (Fig.1). A bio-imaging chip made of an array of differential logarithm circuitry that can collect the fluorescence inputs is designed (Fig.2 (b)). Calculation of the logarithm of the ratio of sample to reference fluorescence lights is accomplished in this stage. A posterior weight-reconfigurable ANN stage is attached for sorting and recognizing desired bioinformatic patterns. By using massively paralleled neural computing interconnections and mixed-signal deep sub-micro technology, the ANN can be implemented on a single VLSI chip [5]. A parallel row/column data flow architecture is used to connect all on-chip systems and eliminate data bandwidth bottlenecks due to

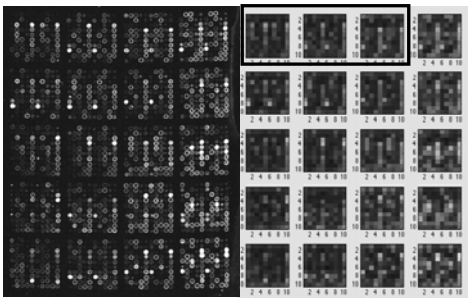

Fig. 3. (Left) Fluorescence image of a sampled microarray of cDNA, Cy3 dye and Cy5 dye mix (only Cy5 red fluorescence is shown) [4]. Each cluster consists of a 10 by 10 grid of sample dots. Each dot corresponds to the location of a cDNA probe to which mRNA from the cells of interest have been bound. (Right) Pixilated images of the clusters from the left microarray photo. Each cluster consists a 10 by 10 pixel array that mimics a bio-signature pattern. The first three randomly picked desired patterns (enclosed in the solid frame) have indices (0,0), (0,1) and (1,0) accordingly. The rest unwanted patterns share index (1,1).

conventional bus architectures. The neural processor chip performs recognition of a single input vector at a time complexity $O(1)$. This ANN consists of the input neurons, programmable weight synapses, summing and inner product cells, sigmoid-logarithmic output neurons, and an output multi-winner-take-all encoder. The programmable synapse matrix is composed of $M \times N$ cells for $N \times M$-dimensional codevectors. The output neuron array is composed of N summing neurons with sigmoid-logarithmic transfer function. The multi-winner-take-all block consists of N competitive circuit cells and uses binary codes to encode N classes. The weights set is generated and obtained from an off-chip supervised learning algorithm. This miniature system is expected to help increase the throughput of on-site assay analysis especially in a scenario for finding desired or suspicious bio-patterns in a massive amount of data.

A prototype ANN chip with 25 input neurons made of a scalable CMOS technology (2 μm) was designed and tested (top right of Fig.1(b)). It contains 25 x 64 weight cells, 64 summing neurons, 64 winner-take-all cells and a 64-to-8 encoder. Its power dissipation is approximately 100 mW at 10 MHz. Its equivalent computation power is about 16 giga-operations per second.

III. Simulations and Empirical Results

A. Saturated Nonlinear Circuitry

The saturated logarithmic circuit is fabricated by using MOSIS available AMI 1.5 μm ABN BiCMOS *n*-well process. (Fig.2 (a)) In conjunction with the N-well/P-base/N+emission bipolar detector, diode-connected NMOS transistors are connected in series. The detecting area collects fluorescence inputs. Mainly due to the diode configured NMOS transistors, this normally "OFF" and low power circuit's behavior is more logarithmic-like while operating in the sub-threshold region and square-root-like ($\sim\sqrt{I_{DS}}$) while operating in the saturation region. It consumes from 10 nW to 2 μW (V_{DD}: 5 Volts) depending on the incident intensity and wavelength.

B. Bio-signature Sorting & OCR Numerical Simulation

To emulate the capability of recognizing dim patterns, a bio-signature sorting task and a pattern recognition task were numerically simulated. MATLAB programs were created to train an ANN and examine its performance by using a novel sigmoid-logarithmic transfer function as the following.

$$A(h) = \begin{cases} \dfrac{1}{1+\exp(-h)} & h < -2 \\ -\alpha \ln(\beta(\delta - h)) & -2 \le h < 0 \\ \alpha \ln(\beta(h + \delta)) & 0 \le h < 2 \\ \dfrac{1}{1+\exp(-h)} & 2 \le h \end{cases}$$

Here, α = 0.050095635, β = 1000, δ = 0.01, h is the net weighted inputs to the transfer function $A(h)$.

A 100-100-2 (100 inputs, 100 hidden neurons and 2 output neurons) artificial feedforward neural network is chosen to perform both recognition jobs. For bio-signature sorting, 20 patterns/clusters of bio-informatics were prepared (Fig.3). Another seven datasets by rescaling the gray level of the original dataset with different factors were also produced

TABLE I
Bio-signature and OCR Recognition Results
(Unit: counts of patterns correctly recognized in one test dataset.
OCR: Each test dataset contains 100 characters.
BIO: bio-signature sorting task, each test dataset contains 20 patterns.)

Transfer function (learning rate) / Gray level to original data	Hybrid sigmoid-logarithmic (η = 0.03)		1/(1+exp(-h)) (η = 0.03)		1/(1+exp(-5h)) (η = 0.0018)	
	BIO	**OCR**	**BIO**	**OCR**	**BIO**	**OCR**
1 (original)	20	64	20	79	20	63
1 / 3.16	20	58	6	48	13	65
1 / 10	20	58	1	25	1	37
1 / 31.6	20	47	1	25	1	25
1 / 100	19	32	-	-	-	25
1 / 316	17	26	-	-	-	-
1 / 1000	17	25	-	-	-	-
1 / 10000	17	25	-	-	-	-

(factors: 1/3.16, 1/10, 1/31.6, 1/100, 1/316, 1/1000, 1/10000 of the original gray level). Noticing that, both brightness and contrast level of these new bio-patterns were reduced by the rescaling factor. The three desired (framed in Fig.3 that were randomly picked) bio-patterns are what we were searching for. EBP training using sigmoid-logarithmic transfer function and gradient descent method was conducted to find a convergent weight configuration (fixed learning rate η = 0.03). Result of recognizing all datasets by the trained network is shown in Table I (BIO). The network using new transfer function can still find one desired bio-signature in the dataset with factor 1/100 of the original gray level. However, the networks using conventional sigmoid transfer functions cannot distinguish bio-patterns well in the dataset with factor 1/3.16 of the original gray level. Consistent recognition result was obtained for the optical character recognition (OCR) case (Table I (OCR)).

IV. Conclusion

A new multichip microsystem for field applicable DNA analysis was proposed based on a mixed-signal ANN hardware architecture. A software engine/algorithm using a novel sigmoid-logarithmic function corresponding to the proposed microsystem was invented and proven to function effectively. Simulation results show that a trained ANN using this new transfer function can classify low fluorescence patterns better than using the conventional transfer functions. A differential logarithmic bio-image chip is designed. Its single-rail logarithmic circuit was fabricated and characterized. This proposed microsystem is expected to make on-site, real-time, high-throughput genetic expression analysis feasible.

References

[1] Lagally, E. T. *et. al.*, "Integrated Portable Genetic Analysis Microsystem for Pathogen/Infectious Disease Detection", *Analytical Chemistry*, v.76, no. 11, 2004.

[2] Lee, Thomas M. -H. *et. al.*, "Microfabricated PCR-electrochemical device for simultaneous DNA amplification and detection", *Lab Chip*, vol. 3, 2003.

[3] Dobbin, K. *et. al.*, "Questions and Answers on Design of Dual-Labeled Microarrays for Identifying Differentially Expressed Genes", *J. of the National Cancer Institute*, vol. 95, no. 18, pp. 1362-1369, 2003.

[4] Hu, L. *et. al.*, "Obtaining Reliable Information From Minute Amounts of RNA Using cDNA Microarrays", *BMC Genomics*, vol. 3, 16, 2002.

[5] Fang, W.-C. *et. al.*, "A Low Power Smart Vision System based on Active Pixel Sensor Integrated with Programmable Neural Network Processor," 1997 International Conference on Computer Design, Austin, Texas, Oct. 11-14, 1997.