VLSI Circuits for Optoelectronic Neural Network Weight Setting[1]

Chwei-Po Chew, Robert W. Newcomb, and Jen-Dong Yuh
Microsystems Laboratory, Electrical Engineering Department
University of Maryland, College Park, MD 20742
Fax:301-314-9281 Phone:301-405-3662 E-mail: mabc@eng.umd.edu newcomb@eng.umd.edu yuh@eng.umd.edu

Keywords Optoelectronics, Programmable, VLSI Circuit, Neural Networks, Weight

Abstract A VLSI optoelectronic implementation of matrix multiplication, which is useful for image-processing applications using neural networks, is presented. BiCMOS NPN and PNP phototransistors generate excitatory and inhibitory synapse currents when light shines on them. These synapse currents are controlled by complementary MOS transistors which allow us to program real valued weights. Differential amplifiers are used to sum the synapse currents while transmission gates are used as resistors. We present an example for digital weight interconnections with 5 bits resolution plus a sign. Circuit diagrams, SPICE3e1 circuit simulations, a NPN phototransistor magic layout, and a phototransistor measurement are also given.

I. Introduction

Matrix multiplication between input vector and weight matrix is the key operation for neural networks [1–4]. For hardware implementation, this has been done in electronic circuits which use floating gates, CCDs, or capacitors as weights [5–10]. Optoelectronic neural circuits, however, have advantages of providing massive interconnectivity and high parallelism which are useful for image-processing applications.

In Caudell's research [11–12], he proposed an electro-optical neural system, mainly built by photodetectors, miniature electro-optical cells, and laser diodes. Low resolution holographic mirrors are used to direct the global structure of network architecture [11]. In addition, Caudell uses a low cost filter, known as the binary phase-only filter (BPOF), and accompanied by a spatial light modulator (SLM) to implement the adaptive resonance neural network (ART1) [12]. Currently, holograms have been of interest in the implementation of optoelectronic neural networks because optoelectronic neurons fabricated from semiconductor materials can be connected by holograms [13–16]. However, these circuits are not easily integrated into a VLSI chip.

Some photoreceptors, based on CMOS technology, have been described by Mead [17]. A few VLSI oriented weight circuits are presented in references [18–22]. However, previous optoelectronic neural networks [19–22] have the limitations such as (1) only ±1 synapses are available and (2) synapses are not programmable.

Therefore, in this paper, we present a VLSI optoelectronic real valued matrix multiplication circuit which can be fabricated by ORBIT Analog 2.0 μ BiCMOS process.

II. Network Architectures and Circuit Diagrams

For simplicity, we build a digital weight interconnection with a resolution of b (b=5) bits plus a sign. Assume we have n input neurons (index j=1,...,n), and m output neurons (index k=1,...,m) (see Fig. 1). Let

$$\begin{aligned} X &= (x_1, \ldots, x_j, \ldots, x_n) \\ W_k &= (w_{k1}, \ldots, w_{kj}, \ldots w_{kn}) \\ V &= WX^T \end{aligned} \tag{1}$$

where X stands for a n-vector input, X^T is the transpose of X, W is a mxn weight matrix for output neurons, and V is a m-vector output. The notation, w_{kj}, is the weight value from input neuron j to output neuron k.

[1] Research supported in part by NSF Grant MIP 891122

We are going to do the matrix multiplications

$$v_k = \sum_{j=1}^{n} w_{kj} x_j \tag{2}$$

where v_k is the weighted sum of output neuron k.

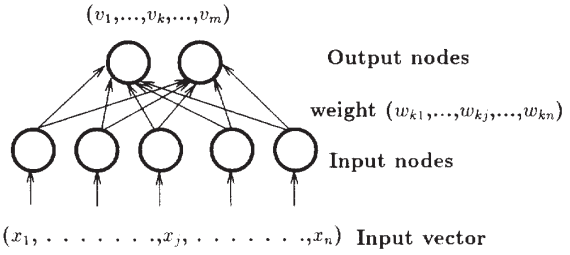


Figure 1: A Grossberg layer neural network [2,3].

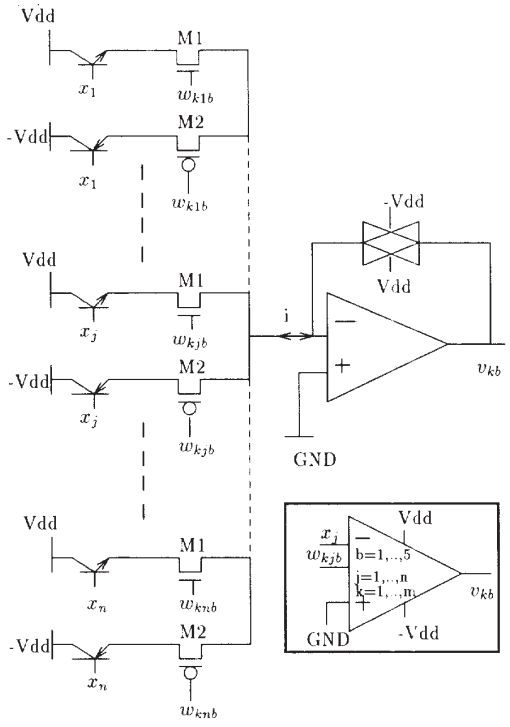


Figure 2: A circuit does a bit sum of products of an optoelectronic input vector and one bit of a digital weight vector W_k.

Figure 2 shows a bit multiplication between an optoelectronic input vector and one bit of a digital weight vector. A NPN phototransistor is responsible for the excitatory synapse current which flows into the summing node while a PNP phototransistor is for inhibitory synapse current which flows out of the summing node. Complementary MOS transistors (M1 & M2) are used to switch among excitatory synapse (when w_{kjb}= 5v, k=1,2,..,m, j =1,2,..,n, b=1,2,..,5), inhibitory synapse (when w_{kjb}= −5v), and 0 weight value (w_{kjb}= 0v), where w_{kjb} decides the bth bit (including sign) of weight w_{kj}. The optical input x_j has the binary value (0 or 1). Transmission gates are used as resistors. Therefore, we get the sum of products of weight (w_{kjb}, j=1,..,n) and input vector (x_j, j=1,..,n) for each bit line. Let us denote this bit line summation voltage as v_{kb}. The symbol for this circuit is shown in the inset.

Figure 3 shows the summation of all the 5 bit line voltages v_{kb} of Fig. 2. The input transmission gates are weighted by different L/W ratios from 1:1 to 1:16 while the output transmission gate has a 1:1 L/W ratio. Thus, we get an output voltage v_k for each output neuron k.

For operation, we use optical illumination (x_j from input neurons) (light on for logic 1, off for logic 0) applied to phototransistors on each bit (w_{kjb}) of weight value. If a bit weight is positive (negative), we apply 5v (-5v) to the bases of M1 & M2. If a bit weight is 0, the bases of M1 & M2 are grounded.

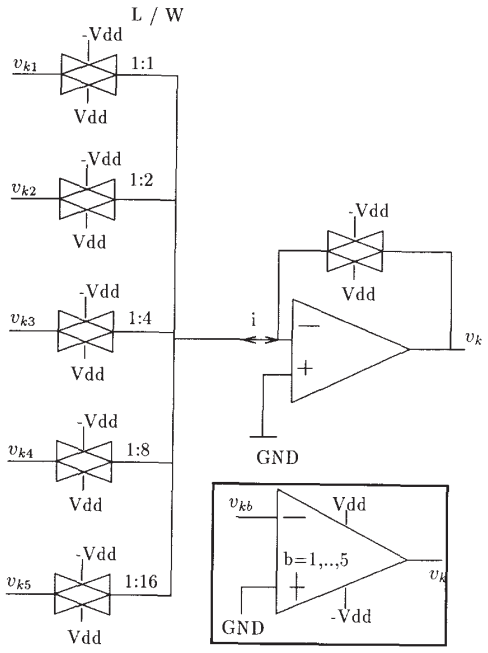


Figure 3: A circuit does the summation of 5 bit outputs from Fig. 2.

Figure 4 shows the differential amplifier we use in Figs. 2–3. Figure 5 is the transmission gate we use in Figs. 2–3.

Magic layouts of a NPN phototransistor and a summing amplifier are shown in Figs. 6 & 7 which have been fabricated by ORBIT Analog 2.0 μ BiCMOS process.

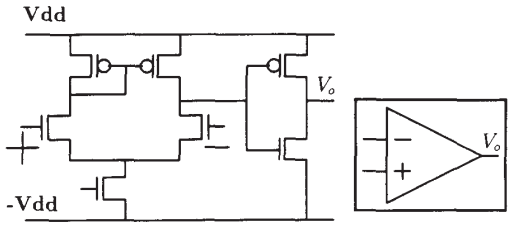


Figure 4: A differential amplifier circuit.

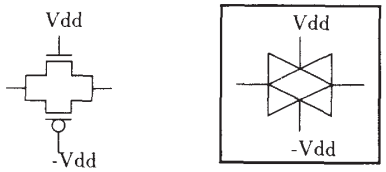


Figure 5: A transmission gate circuit.

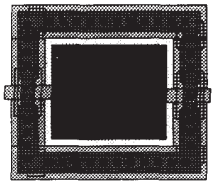


Figure 6: A magic layout of NPN phototransistor.

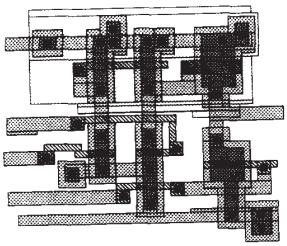
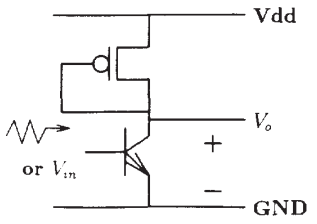


Figure 7: A magic layout of Fig. 4.

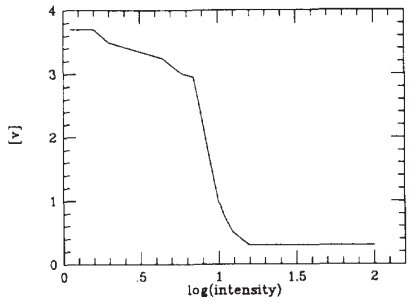
III. Simulations and Experiments

In this section, circuit simulation results obtained by SPICE3e1 are presented. In addition to simulation results, we present the characteristic curve of our photoreceptor.

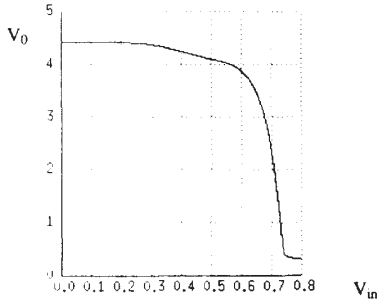
A test photoreceptor shown in Fig. 8a is considered. Figure 8b is the measured characteristic curve when light shines on the base of NPN phototransistor. Figure 8c is our SPICE3e1 simulation where we apply electric voltages on the base of NPN phototransistor. The consistency between our measurement and simulation is apparent. According to our measurement, when light shines on this NPN phototransistor, it generates a maximum current in the order of hundreds of microampere.



(a). A photoreceptor circuit.



(b). the measurement curve of Fig. 8a.

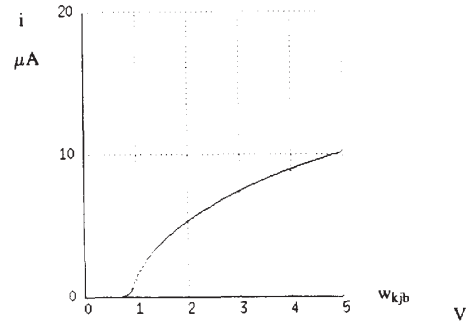


(c). the SPICE3e1 simulation result of Fig. 8a.

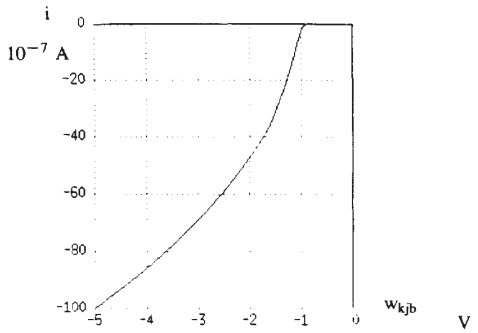
Figure 8: A photoreceptor circuit, a measurement curve, and a simulation curve.

Figure 9 shows the circuit simulations of Fig. 2. In our simulation, we apply electric voltages (0.667v for NPN and -0.732 for PNP) to the bases of phototransistors instead of optical illuminations. In Fig. 9a, we get an I-V curve by varying the input voltage of transistor M1 from 0 to 5v. In Fig. 9a, we get 10 μA current when the input to M1 is 5v. Similarly, we vary the input voltage of M2 transistor from -5 to 0v. In Fig. 9b, we obtain -10 μA when the input to M2 is -5v. In addition, we find that both M1 and M2 transistors are turned off when inputs are 0v (see Figs. 9a & 9b). Therefore, we get 10 μA excitatory current when the gate voltages of M1 & M2 are 5v, -10 μA for -5v gate voltages, and 0 μA for 0v gate voltages. The feedback transmission gate resistance is about 41 kΩ. Thus, we get each bit sum voltage V_{kb} (see Fig. 2) via a differential amplifier.

Figure 10 shows the characteristic I-V curves for 5 different L/W ratios of transmission gates. From Fig. 10, we ensure we can use those transmission gates as resistors. Figure 11 is the 5 bit summation of Fig. 3. In Fig. 11, each bit input voltage ranges from -1.5v to 1.5v. The result shows the summation is linear when inputs are all in the range of -1.5v to $+1.5$v. This range is easily obtained by controlling the value of feedback resistance in Fig. 2.



9a: An I-V curve for an excitatory synapse. When we vary the input voltage (w_{kjb}, k=1...,m, j=1,...,n, b=1,...,5) of transistor M1 from 0 to 5v, we get the current (i) which flows through M1.



9b: An I-V curve for an inhibitory. When the input voltage (w_{kjb}, k=1,...,m, j=1,...,n, b=1,...,5) of transistor M2 is varied from -5v to 0v, we get the current (i) which flows through M2.

Figure 9: The circuit simulation of Fig. 2. We apply 0.667v and -0.732v to the bases of NPN and PNP phototransistors respectively.

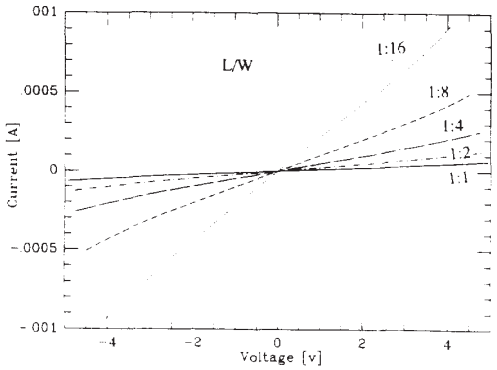


Figure 10: An I-V curves of 5 different ratios of transmission gates.

IV. Discussion & Conclusions

In our paper, a fabricated NPN phototransistor is presented along with measurements on the chip and SPICE3e1 simulations. A VLSI circuit for optoelectronic weight setting is designed and simulated. Magic layouts of the NPN phototransistor and a differential amplifier are shown. However, some points still need to be mentioned.

1. Our PNP phototransistor is a lateral bipolar transistor (see Fig. 12) because PNP transistor is not directly available in the MOSIS process. Due to a wide n-base and a low doping concentration in the n-base region, a lower gain than a NPN transistor is expected. However, it is still good enough for our circuits.

2. In our paper, we assume PNP and NPN phototransistors will generate equal amounts of photocurrents under moderate illuminations. In practice, we adjust the illumination areas of the PNP & NPN phototransistors to compensate for differences in the excitatory and inhibitory currents at a specific optical input level.

3. In order to guarantee the phototransistors operating properly, we maintain the correct voltage priority across the phototransistors.

4. In Fig. 11, the summation characteristic curve is not linear if all input voltages are not in the range of -1.5v to $+1.5$v. This is because we use a low gain differential amplifier instead of an ideal operational amplifier.

5. The optical input can be guided by waveguides or pass through a mask [23] to avoid interference.

In this paper, a VLSI optoelectronic circuit for weight setting is presented. Circuit diagrams are shown. Circuit simulations and experiments demonstrate this optoelectronic circuit theoretically works well.

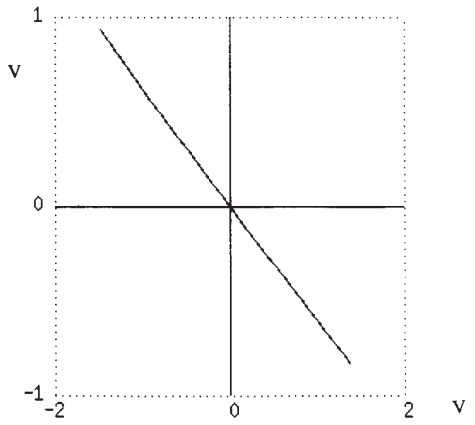


Figure 11: The simulation results of Fig. 3. Here, we assume v_{kb} (b=1,..,5) are equal and varied from -1.5v to 1.5v.

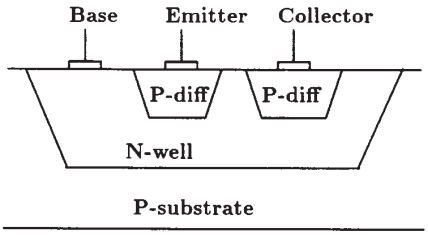


Figure 12: A lateral PNP phototransistor structure.

V. References

1. J. J. Hopfield, "Neurons with Graded Response Have Collective Computational Properties Like Those of a Digital Optical Processor," Proc. National Academy Science USA, Vol. 81, May 1984, pp. 3088–3092

2. Robert Hecht-Nielsen, "Applications of Counterpropagation Networks," Neural Networks, Vol. 1, 1988, pp. 131–139

3. Robert Hecht-Nielsen, "Counterpropagation Networks," Applied Optics, Vol. 26, No. 23, 1 Dec. 1987, pp. 4979–4984

4. R. P. Lippmann, "An Introduction to Computing with Neural Nets", IEEE ASSP Magazine, Vol. 4, No. 2, April, 1987, pp. 4-22

5. H. P. Graf, L. D. Jackel, "Analog Electronic Neural Network Circuits," IEEE Circuits and Devices Magazine, Vol. 5, No. 4, July, 1989, pp. 44–49

6. S. Y. Foo, L. R. Anderson, Y. Takefuji, "Analog Components for the VLSI of Neural Networks," IEEE Circuits and Devices Magazine, Vol. 6, No. 4, July, 1990, pp. 18–26

7. L. D. Jackel, H. P. Graf, R. E. Howard, "Electronic Neural Network Chips," Applied Optics, Vol. 26, No. 23, Dec. 1987, pp. 5077–5080

8. M. Holler, S. Tam, H. Castro, and R. Benson, "An Electrically Trainable Artificial Neural Network (ETANN) with 10240 Floating Gate Synapses" in Proc. IJCNN, vol. II (Washington, DC), June 1989, pp. 191–196

9. D. B. Schwartz, R. E. Howard, "A Programmable Analog Neural Chip," Proc. IEEE Custom Integrated Circuits Conf., IEEE Press, 1988, pp. 10.2.1–10.2.4

10. A. Agranat, A. Yariv, "Semiparallel Microelectronic Implementation of Neural Network Models Using CCD Technology,"Electronics Letters, Vol. 23, No. 11, 1987, pp. 580–581

11. T. P. Caudell, "Parametric Connectivity: Feasibility of Learning in Constrained Weight Space," in Proc. IJCNN, Vol. I (Washington, DC), June 1989, pp. 667–675

12. D. C. Wunsch, T. P. Caudell, C. D. Capps, R. J. Marks II, R. A. Falk, "An Optoelectronic Implementation of the Adaptive Resonance Neural Network," To be published in IEEE Trans. on Neural Network, special issue on hardware implementation of neural network.

13. H. J. Caulfield, D. Armitage, "Adaptive Resonance Theory of Optical Pattern recognition," Applied Optics, Vol. 28, No. 19, October 1989, pp. 4060–4061

14. D. P. Psaltis, D. Brady, X. Gu, S. Lin, "Holography in Artificial Neural Networks," Nature, Vol. 343, January 1990, pp. 325–330

15. K. Hsu, H. Li, D. Psaltis, "Holographic Implementation of a Fully Connected Neural Network," Proc. of IEEE, Vol. 78, No. 10, Oct. 1990, pp. 1637–1645

16. S. M. Arnold, "Electron Beam Fabrication of Computer-Generated Holograms," Optical Engineering, Sept./Oct. 1985, Vol. 24, No. 5, pp. 803–807

17. C. A. Mead, Analog VLSI and Neural Systems, Addison-Wesley, 1989, pp. 257–278

18. N. H. Farhat, "Optoelectronic Neural Networks and Learning Machines, "IEEE Circuits and Devices Magazine, Vol. 5, No. 5, Sep. 1989, pp. 32–41

19. R. V. Jones, "Photoneural Systems: an Introduction," Applied Optics, Vol. 26, No. 10, 15 May 1987, pp. 1948–1958

20. H. Yonezu, A. Miho, T. Himeno, K. Pak, Y. Takano, "Optoelectronic Neural Circuit with Variable Synaptic Weights," Electronics Letters, Vol. 25, No. 10, 1989, pp. 671–672

21. M. Otia, M. Takahashi, S. Tai, K. Kyuma, "Character Recognition Using a Dynamic Optoelectronic Neural Network with Unipolar Binary Weights," Optics Letters, Vol. 15, No. 21, 1 Nov. 1990, pp. 1227–1229

22. C. F. Neugebauer, A. Agranat, A. Yariv, "Optically Configured Phototransistor Neural Networks," Proceedings of the International Joint Conference on Neural Networks Vol. II, 1990, pp. 64–67

23. C. M. Gomes, H. Sekine, T. Yamazaki, S. Kobayashi, "Bipolar Optic Neural Networks Using Ferroelectric Liquid Crystal Devices," Neural Networks, Vol. 5, No. 1, 1992, pp. 169–177