NOLTA, IEICE

Paper

High-fidelity pulse density modulation in neuromorphic electric circuits utilizing natural heterogeneity

Akira Utagawa [1 a)] *Tetsuya Asai* [1] *Yoshihito Amemiya* [1]

[1] *Graduate School of Information Science and Technology, Hokkaido University Kita 14, Nishi 9, Kita-ku, Sapporo, 060-0814, Japan*

[a)] *utagawa@lalsie.ist.hokudai.ac.jp*

Abstract: Hospedales *et al.* have recently proposed a neural network model of the "vestibulo-ocular reflex" (VOR) in which a common input was given to multiple nonidentical spiking neurons that were exposed to uncorrelated temporal noise, and the output was represented by the sum of these neurons. Although the function of the VOR network is equivalent to pulse density modulation, the neurons' non-uniformity and temporal noises given to the neurons were shown to improve the output spike's fidelity to the analog input. In this paper, we propose a CMOS analog circuit for implementing the VOR network that exploits the non-uniformity of real MOS devices. Through extensive laboratory experiments using discrete MOS devices, we show that the output's fidelity to the input pulses is clearly improved by using multiple neuron circuits, in which the non-uniformity is naturally embedded into the devices.

Key Words: vestibulo-ocular reflex, integrate-and-fire neuron, neural network, pulse density modulation

1. Introduction

Charge-based ultra-low-power analog circuits may suffer from physical limitations in their temporal responses because the present designing strategy is to decrease the power supply voltage or bias currents [1]. Is there a possible way to construct an analog circuit that can perform high-speed information processing with slow-but-low-power semiconductor devices? Neural networks are ideal for such an implementation because high-speed parallel information processing can be performed with slower neural elements as compared to present semiconductor devices. Recently, Hospedales *et al.* reported that in a neural network model of "vestibulo-ocular reflex" (VOR), the non-uniformity of neurons and the temporal noises applied to the neurons clearly improved the resulting output's fidelity to the analog input [2]. Figure 1 shows the proposed model. A common input is applied to N spiking neurons that are exposed to uncorrelated temporal noise ξ_i $(i = 1, 2, \cdots, N)$, and the output is represented by the sum of these neurons. When the neurons are identical and no noise is applied to the network, the generated spikes are synchronous to each other. On the other hand, when temporal noises is applied or when the neurons are non-identical, their spikes are no longer correlated. Note that that the average inter-spike interval of the noisy network's output is roughly N-times smaller than that of the synchronous one. Therefore, the response (or firing) frequency of the noisy network is always larger than that of the undisturbed network, resulting in a fidelity improvement at the input.

[Received: August 20, 2010. Revised: November 21, 2010. Published: April 1, 2011.]
[DOI: 10.1588/nolta.2.218]
Nonlinear Theory and Its Applications, IEICE, vol. 2, no. 2, pp. 218−225

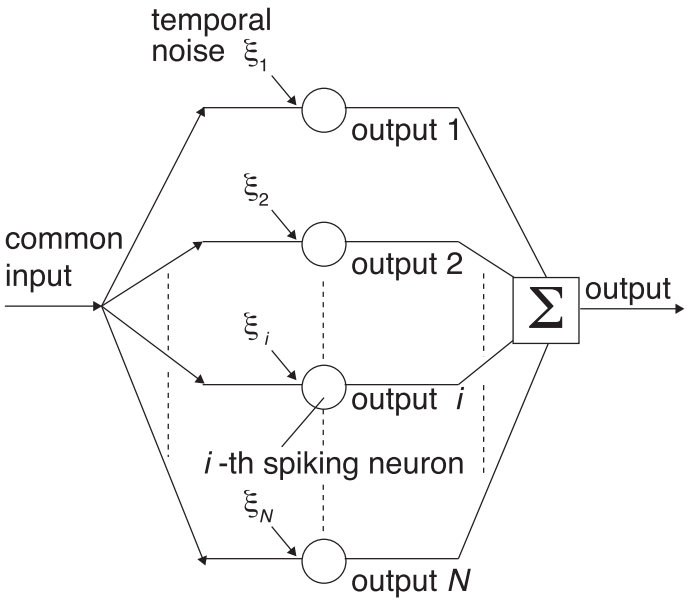


Fig. 1. Neural network model of VOR.

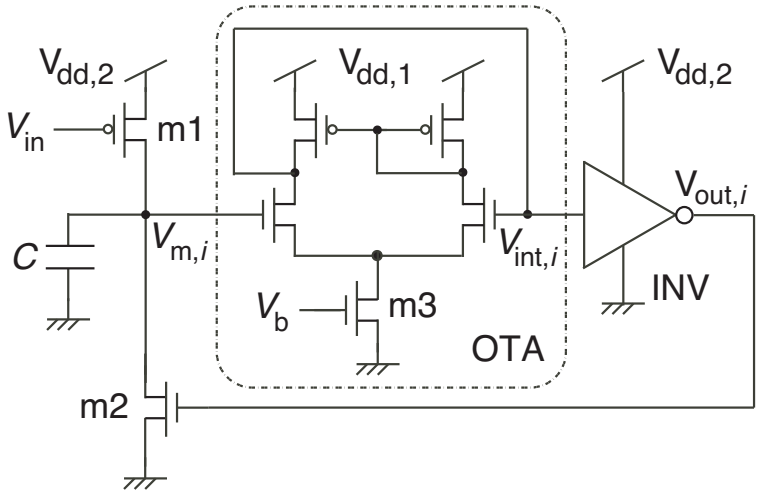


Fig. 2. i-th integrate-and-fire neuron circuit.

Inspired by these results, we propose a CMOS circuit for implementing a VOR network that exploits the non-uniformity of real MOS devices. Ideally, the responses of the VOR network circuit for sinusoidal inputs would be evaluated, i.e., the D/A-converted values of the pulse modulated outputs of the network would be calculated, and the outputs' signal-to-noise ratio (SNR) at the input frequency would be measured. However, in this paper, we instead show that an improvement in the output's fidelity can be achieved using multiple neuron circuits, whose non-uniformity is naturally embedded into the devices. This is demonstrated through extensive laboratory experiments.

2. A CMOS circuit implementing a neural network model of VOR

Based on the VOR model proposed by Hospedales *et al.* [2], we developed a network circuit that consisted of multiple MOS neuron circuits. In the network, we used integrate-and-fire neuron circuits. Figure 2 shows a schematic of the neuron circuit. In Fig. 2, $V_{\mathrm{m},i}$ and $V_{\mathrm{int},i}$ represent the membrane voltage and internal voltage of the i-th neuron, V_{b} the bias voltage, C the capacitance, $V_{\mathrm{dd},1}$ and $V_{\mathrm{dd},2}$ the supply voltages. The value of C sets the maximal follow-up frequency of the single neuron circuit. The circuit consists of a standard inverter (INV), an nMOS transconductance amplifier (OTA) acting as a comparator, an input transistor m1, and a reset transistor m2. When $V_{\mathrm{in}} < V_{\mathrm{dd},2}$, the membrane voltage $V_{\mathrm{m},i}$ is increased by charging C through transistor m1. Since the OTA has a

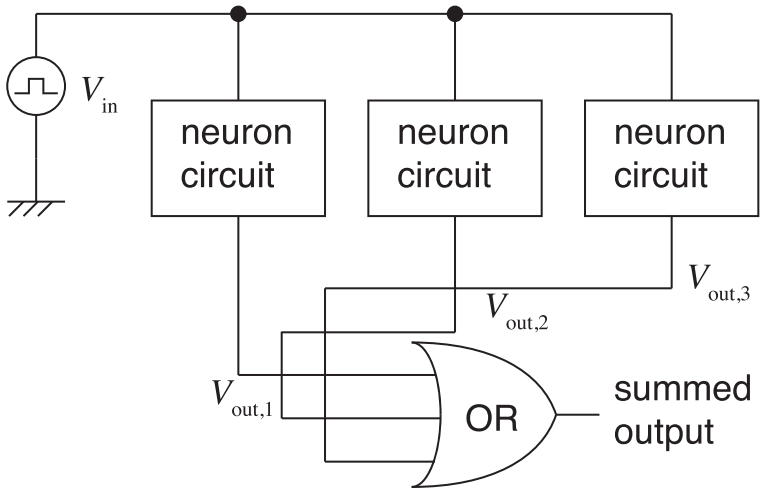


Fig. 3. VOR network with three neuron circuits.

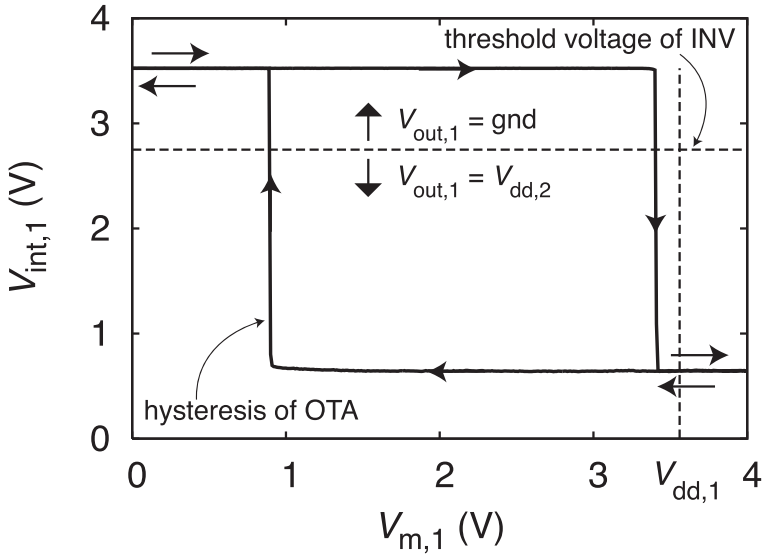
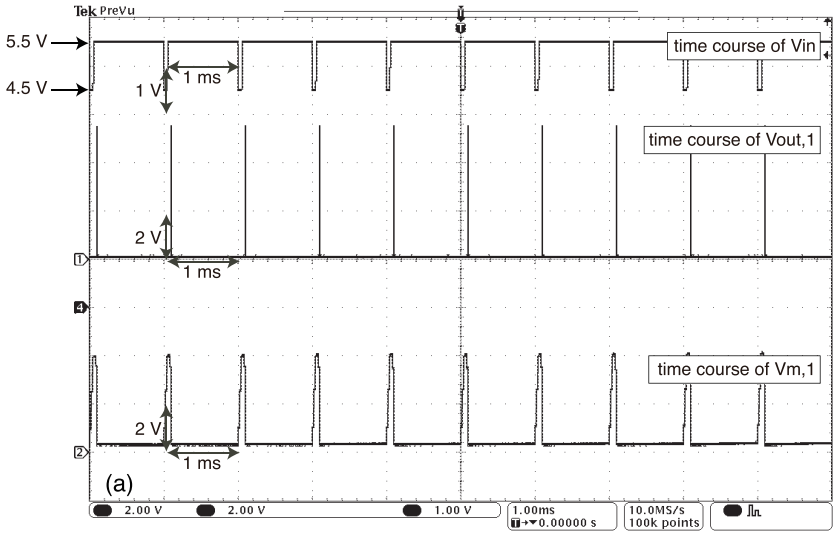


Fig. 4. Hysteresis of OTA (experimental results).

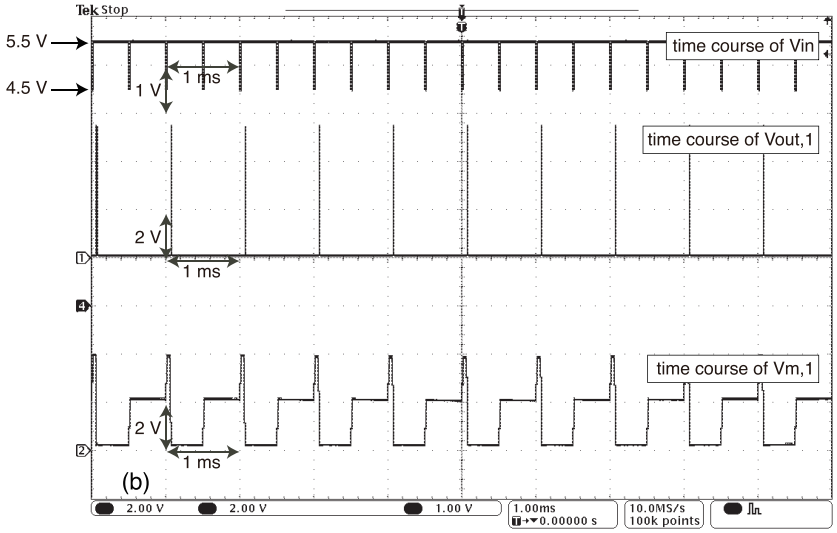
positive feedback loop (via node $V_{\mathrm{int},i}$), hysteresis exists between $V_{\mathrm{m},i}$ and $V_{\mathrm{int},i}$ (see [3] for details). When $V_{\mathrm{m},i}$ approaches $V_{\mathrm{dd},1}$, $V_{\mathrm{int},i}$ approaches gnd. The output of the INV ($V_{\mathrm{out},i}$) goes to $V_{\mathrm{dd},2}$, which results in shunting of $V_{\mathrm{m},i}$. Bacause of the hysteresis characteristics, shunting continues until $V_{\mathrm{m},i}$ reaches gnd. $V_{\mathrm{m},i}$ is therefore reset to gnd with a time delay. During this delay period the output voltage $V_{\mathrm{out},i}$ is at $V_{\mathrm{dd},2}$, and outside of this period it remains at gnd. In other words, $V_{\mathrm{out},i}$ is a narrow pulse signal that represents neuron's output spike exhibiting "integrate and fire" behaviors. Figure 3 shows a schematic of a VOR network circuit consisting of three neuron circuits. A common input (V_{in}) is applied to each neuron circuit. Since the neuron's output is represented by a spike (a logical "0" for the resting and refractory states, a logical "1" for firing state), the summed output can be represented simply by the logical OR of each individual neuron's output. It should be noted that in real world experiments every neuron circuit possesses slightly different physical parameters. Therefore, the non-uniformity in the original VOR model is naturally embedded into these circuits.

3. Experimental results

In the following experiments, we used discrete devices with C = 100 pF, $V_{\mathrm{dd},1}$ = 3.6 V, and $V_{\mathrm{dd},2}$ = 5.5 V. We used nMOS-FET (ALD1106) and pMOS-FET (ALD1107) for the OTA, m1, and m2 transistors, and a standard inverter (TC4069BP) for the INV. We used an HD4075BP for the three-input OR gate. The delay through the OR gate is small as compared to the delay of our proposed



(a) $f_{\mathrm{in}} = 1$ kHz.



(b) $f_{\mathrm{in}} = 2$ kHz.

Fig. 5. Responses of single neuron (experimental results).

circuit. The bias voltage V_{b} was set to 0.6 V to mimic operations in subthreshold CMOS circuits (the threshold voltage of the nMOS-FET used in these experiments was actually 0.7 V). We will show that even if the operating frequency of a single neuron is limited owing to its small bias voltage V_{b}, the overall network can still operate at higher frequencies.

Figure 4 shows the hysteresis curves of the OTA of a single neuron circuit. When V_{in} (which is less than $V_{\mathrm{dd},2}$) was applied, $V_{\mathrm{m},1}$increased and eventually approached $V_{\mathrm{dd},1}$, at which point$V_{\mathrm{int},1}$ sharply decreased, as shown in Fig. 4. $V_{\mathrm{int},1}$ then fell below the threshold voltage of the INV as shown by the dashed line in Fig. 4 and $V_{\mathrm{out},1}$ became $V_{\mathrm{dd},2}$. $V_{\mathrm{m},1}$ then decreased until $V_{\mathrm{int},1}$ rose above the threshold voltage of the INV.

Figure 5 shows temporal responses of a single neuron circuit for two different input frequencies ($f_{\mathrm{in}} = 1$ kHz and 2 kHz). V_{in} is a pulsed input and the offset, amplitude, and duty cycle ratio of the input are 5.0 V, 0.5 V, and 95%, respectively. When f_{in} was set to 1 kHz, as shown in Fig. 5(a), the neuron circuit generated a spike after every input pulse, which indicated that the circuit was sufficiently responsive. On the other hand, when f_{in} was set to 2 kHz, as shown in Fig. 5(b), the circuit only generated spikes on every second input pulse, thereby lowering the fidelity. To visualize the relationship between the frequency of the input pulse (f_{in}) and the generated spike (f_{out}), we

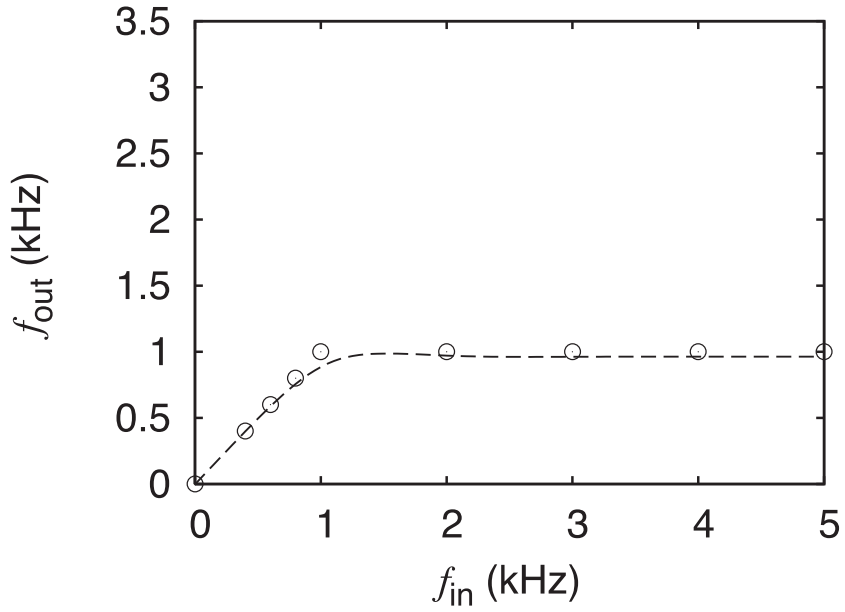


Fig. 6. Output vs. input frequency characteristics of single neuron circuit.

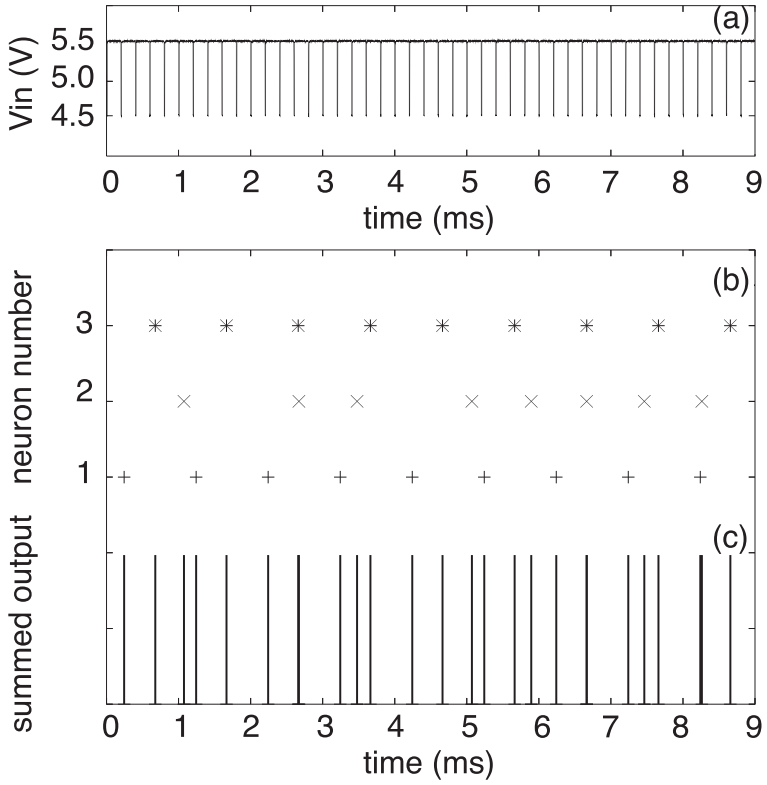


Fig. 7. Responses in noisy VOR network circuit.

plotted the f_{in} vs. f_{out} characteristics of a single neuron circuit in Fig. 6 (the "$\circ$" symbols represent the experimental raw data, and the dashed line was drawn based on bezier curve.) The estimated output frequency f_{out} indicates the total number of output spikes per second. The peak frequency of $V_{\text{out},1}$ was also calculated from FFT analysis and confirmed the 1 kHz upper limit.

The effect of noise (due to non-uniformity between neuron circuits) was then investigated using our network. Figure 7 shows the temporal responses. In this experiment, we applied an input pulse of $f_{\text{in}} = 5$ kHz, as shown in Fig. 7(a), which was five times larger than the operational limit of any single neuron circuit (1 kHz). Figure 7(b) shows each neuron's firing event, where the symbols "+", "$\times$", and "$*$" represent firings of neurons 1, 2, and 3, respectively. As shown, each neuron could not

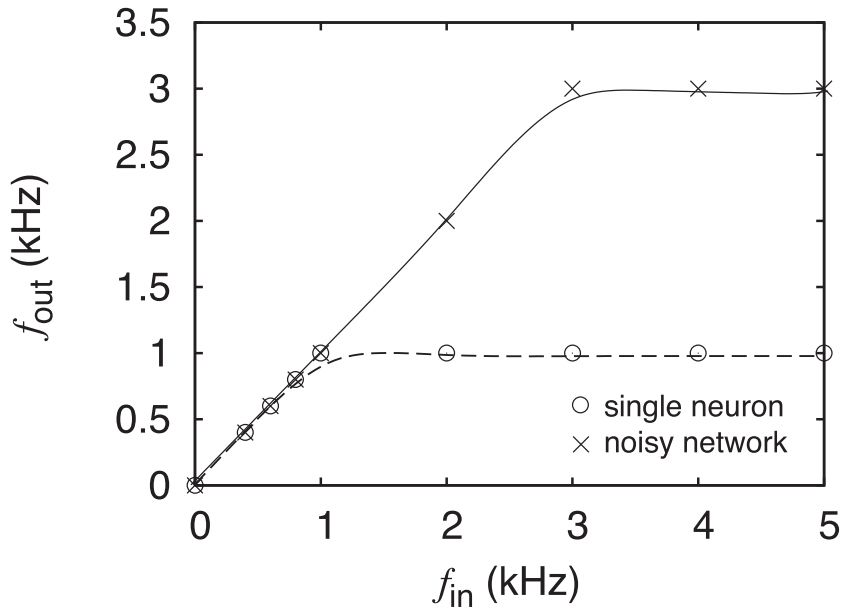


Fig. 8. Input-output frequency characteristics of single neuron circuit and noisy VOR network circuit.

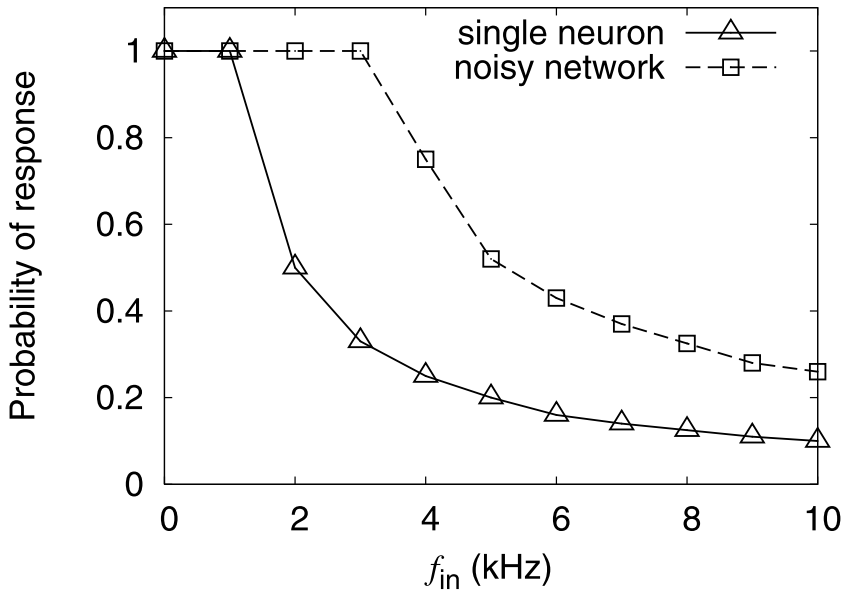


Fig. 9. Experimental probability of response in single neuron circuit and in noisy VOR network circuit.

reliably respond to each input pulse. Figure 7(c) shows the summed output of these three neuron circuits, as represented by the vertical lines. The number of spikes was increased by about three times as compared to the firing of a single neuron, and the response rate to the input pulse was similarly increased by about three times. This clearly shows the fidelity improvements obtained through the innate non-uniformity of the network. Figure 8 plots the $f_{\rm in}$ vs. $f_{\rm out}$ characteristics of the network circuit ("×") as well as that of the single neuron circuit ("○"). The peak frequency of $V_{\rm out}$ (the network output) was calculated from FFT analysis and was found to closely corresponds to $f_{\rm in}$, although the estimated output frequency $f_{\rm out}$ saturated at 3 kHz. Figure 8 clearly shows that the network circuit could respond to the input pulses at maximum fidelity when $f_{\rm in} \leq 3$ kHz. From these results, we have confirmed that the fidelity has been improved by three times and that the extent of the improvement is proportional to the number of neurons in the network. The probability of a response at higher frequencies was then calculated in order to evaluate its reliability. A probability of 1 means that the circuit responds to each input pulse with 100% probability. Figure 9 shows the probability of response

Table I. Summary of simulation condition.

technology	0.18-μm, CMOS	V_b	0.6 V
$V_{dd,1}$ (=$V_{dd,2}$)	1.8 V	μ_{vth0} of m1	-398 mV
V_{in}			
offset	1.4 V	σ_{vth0} of m1	20 mV
amplitude	0.4 V	μ_{vth0} of m3	365 mV
duty cycle ratio	98%		
frequency	1 MHz	σ_{vth0} of m3	15 mV
C	1 pF		

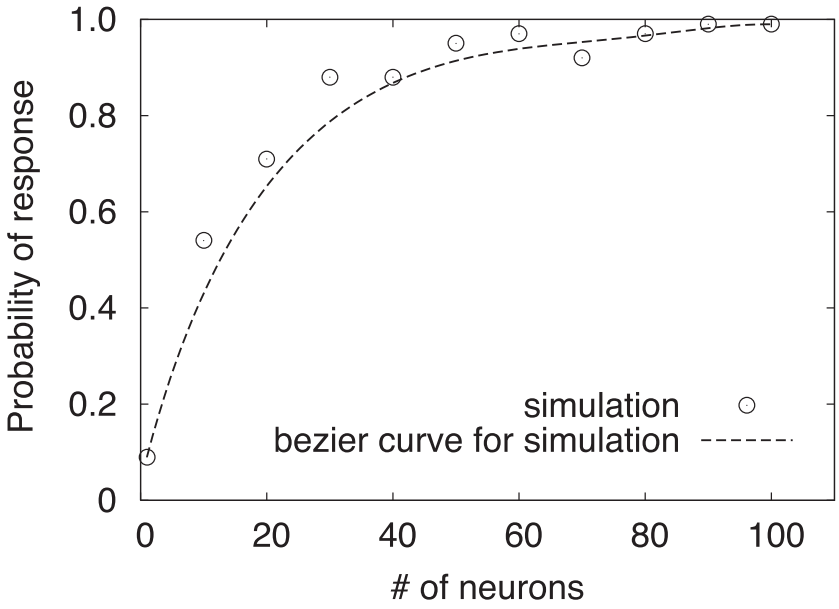


Fig. 10. Simulated probability of response in noisy VOR network circuit.

for a single neuron and a noisy network. Note that the initial state of the circuit was randomly given and the network was shown to be able to respond to the signal up to 3 kHz with 100% probability. At higher frequencies, owing to natural heterogeneity in the system, neurons fired incoherently and their probability of a response decreased.

By implementing N neurons in a network, one would expect a circuit whose maximum $f_{\rm out}$ is N-times faster than that of a single neuron circuit. However, $f_{\rm out}$ is in fact limited by the output spike width of a single neuron circuit, as explained hereafter. The output of the network is represented by a simple summation. Assume that output spike width is T_w s and the maximum follow-up frequency of a single neuron is 1 / T_p Hz. When the input frequency is larger than 1 / T_p Hz, a single neuron cannot follow the input; however, a network consisting of many neurons can in fact follow it. When input frequency is larger than 1 / T_w Hz, output spikes of any one neuron become overlapped with those of other neurons. When large numbers of neurons are employed, all of their output spikes overlap each other and therefore no output response is observed. In our experiments, we observed a 35 μs maximum spike width, giving a maximum $f_{\rm out}$ of 28 MHz. This could be achieved by using 28 $\times 10^3$ neurons in the circuit. Since constructing a network with thousands of neuron with discrete devices is not practical, we instead performed circuit simulations with 100 neurons, which confirmed that the maximum $f_{\rm out}$ is 10 times higher than the maximum follow-up frequency of the single neuron. Table I shows a summary of the simulation condition. Note that the device variation of the neuron circuit accounted for the standard deviation of two separate transistors and assumed a Gaussian distribution. The threshold voltages of m1 and m3 determine the value of T_p and T_w, respectively.

The mean maximum follow-up frequency of a single neuron was 117 kHz, and the initial condition of each neuron was chosen at random. Figure 10 shows the simulations results. The network circuit can be seen to follow a 1MHz input signal with almost 100% probability when 100 neurons are employed.

We have demonstrated the fidelity of the VOR network using integrate-and-fire neuron circuits. In the future, we will evaluate the response of the network to sinusoidal inputs, i.e., we will calculate the D/A-converted values of the pulse modulated outputs of the network and obtain the SNR for each input frequency. This will enable us to construct a useful pulse-density modulator based on our noise-driven circuits.

4. Conclusion

We developed a noise-driven pulse density modulator having high fidelity to input signals based on a neural network model of the vestibulo-ocular reflex (VOR). We constructed a network circuit consisting of multiple CMOS neuron circuits and a standard OR logic circuit. Through extensive laboratory experiments, we have shown that the fidelity to the input pulses was clearly improved by the multiple neuron circuits, in which non-uniformity was embedded in the MOS devices. We confirmed that a single neuron circuit could respond to input pulses of 1 kHz and that a network consisting of three neuron circuits could respond to input pulses of up to 3 kHz which is three times larger than the upper limit of the single neuron circuit.

Acknowledgments

This study was supported by a Grant-in-Aid for Scientific Research on Innovative Areas [20111004] from the Ministry of Education, Culture Sports, Science and Technology (MEXT) of Japan.

References

[1] M. Sanduleanu and E. van Tuijl, *Power Trade-Offs and Low Power in Analog CMOS ICs*, Kluwer Academic Publishers, Dordrecht, Netherlands, 2002.

[2] T.M. Hospedales, M.C.W. van Rossum, B.P. Graham, and M.B. Dutia, "Implications of noise and neural heterogeneity for vestibulo-ocular reflex fidelity," *Neural Computation*, vol. 20, no. 3, pp. 756–778, 2008.

[3] T. Asai, Y. Kanazawa, T. Hirose, and Y. Amemiya, "Analog reaction-diffusion chip imitating the Belousov-Zhabotinsky reaction with Hardware Oregonator Model," *International Journal of Unconventional Computing*, vol. 1, no. 2, pp. 123–147, 2005.

[4] A. Utagawa, T. Asai, T. Hirose, and Y. Amemiya, "An inhibitory neural-network circuit exhibiting noise shaping with subthreshold MOS neuron circuits," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer*, vol. E90-A, no. 10, pp. 2108–2115, 2007.