Analysis and reduction of mismatch in silicon neurons

Shuo, Sun; Basu, Arindam

2011

Shuo, S., & Basu, A. (2011). Analysis and reduction of mismatch in silicon neurons. IEEE Biomedical Circuits and Systems Conference (BioCAS) 2011.

<https://hdl.handle.net/10356/100795>

<https://doi.org/10.1109/BioCAS.2011.6107776>

© 2011 IEEE This is the author created version of a work that has been peer reviewed and accepted for publication by IEEE Biomedical Circuits and Systems Conference (BioCAS) 2011, IEEE. It incorporates referee's comments but changes resulting from the publishing process, such as copyediting, structural formatting, may not be reflected in this document. The published version is available at: <http://dx.doi.org/10.1109/BioCAS.2011.6107776>.

Downloaded on 26 Apr 2023 19:59:35 SGT

Analysis and Reduction of Mismatch in Silicon Neurons

Sun Shuo* and Arindam Basu†

School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore

Email: *ssun2@e.ntu.edu.sg; †arindam.basu@ntu.edu.sg

Abstract—In this paper, we describe a methodical approach for reducing errors due to mismatch in neuron circuits. We chose the neuron’s current-frequency (f-i) curve as the desired output and use a sensitivity analysis to determine which transistors contribute most significantly to its variation. This allows us to identify the most critical transistors that need to be matched. For the special case in which floating-gate (FG) transistors are used to reduce this mismatch, we propose a method to further reduce the number of FG devices to be used in the circuit resulting in a corresponding reduction in ‘calibration’ time. In addition to reducing mismatch between neurons, the usage of FG devices allows the user to independently set the parameters of each neuron. Since the calibration is based on f-i curve, it can be obtained through address-event representation (AER) circuits that are included in the neuron array for normal functionality. We use one example of commonly used integrate and fire neuron to illustrate this mismatch correction procedure. The method presented allows the corrected neurons to compute both rate codes and spike time codes in a mismatch resilient fashion.

I. INTRODUCTION AND MOTIVATION

Silicon neurons, initially proposed by Mead in 1989 [1], capture the characteristics of real nerve cells at different levels of abstraction. Since then, these circuits have been used in neuromorphic sensory devices emulating the retina and cochlea, for developing ‘intelligent’ systems that can interact with its environment in real-time and for high-speed modeling of large scale cortical circuits. As an example, the European FACETS project has created systems able to mimic the brain, with 200,000 neurons linked up by 50 million synaptic connections.

However, mapping a digital neural network simulation to an analog chip is prone to errors because ‘identical’ devices suffer from random mismatch which stem from microscopic fluctuations in dimensions, doping, oxide thickness and a host of other causes [2]. In sub-threshold circuit design, which is employed in neuromorphic systems to save power, the effect of statistical variations is exacerbated because of the exponential dependence of drain current on threshold voltage differences (ΔV_T) [3]. Typical analog design approaches consider only random mismatch assuming that systematic mismatch due to process gradients can be mitigated by techniques like common-centroid layout. However, in arrays of neurons, systematic mismatch also dominates increasing the effect of variations even more.

Several research groups have previously attempted to reduce the variation in silicon neuron arrays. The method proposed in [4] will work well for rate encoded processing but not when

exact timing of the neuron’s spike is important. Also, it uses multiple spikes for every connection which is costly in terms of AER bandwidth and hence power. [5] proposes another method that also uses off-chip memory to modify probability of connection between neural populations. This method also works with rate encoded processing and not for exact spike timings. In this paper, we provide a method of analyzing and reducing mismatch in silicon neurons for both rate and spike-timing codes.

II. STATISTICAL VARIATIONS IN SILICON NEURONS

A. General Analysis

A neuron circuit can be most generally described by a differential equation as follows:

$$\frac{d\vec{x}}{dt} = Q_{\vec{\alpha}}(\vec{x}, I_{in}, t) \quad (1)$$

where $\vec{x}$ is the vector of state variables including the membrane voltage (V_{mem}), I_{in} is the input current and $Q_{\vec{\alpha}}$ is the vector field that is dependent on a set of parameters $\vec{\alpha}$. Mismatch between parameters of different neurons will result in different vector fields $Q_{\vec{\alpha}}$. Analyzing the effect of variations rigorously will be very complicated since it needs us to compare the evolution operators of the corresponding mismatched differential equations.

Instead, we propose to use a particular signature of the evolution operator, the f-i curve to analyze neuron circuit mismatch. We make this choice because the f-i curve is a property shared by all neuron circuits and it would directly reflect overall neuron circuit’s spiking behavior. This is a commonly used method for encoding information in spiking neurons - hence, circuits are already needed on the chip to extract this information leading to no extra overhead for calibration. Moreover, eliminating mismatch in the f-i curve will allow exact spike timings after stimulus presentation to also be matched reducing variability in time-to-first-spike type encoding strategies.

Consider the same current injected into an array of neurons and assuming threshold voltage mismatch is the dominant source underlying f-i curve mismatch, the variance in time period of spiking can be expressed as

$$\sigma_T^2 = f(\sigma_{V_{T1}}^2, \sigma_{V_{T2}}^2, \dots, \sigma_{V_{Tn}}^2) \quad (2)$$

where the neuron circuit has ‘n’ transistors, σ_T^2 is the variance in the neuron spiking period and $\sigma_{V_{Ti}}^2$ is the threshold voltage variance of the i-th transistor. Equation (2) can be further

Fig. 2: **Monte Carlo analysis:** A Monte Carlo analysis is performed on the neuron circuit assuming σ_{V_T} of 2mV, 5mV, 10mV and 20mV. For each value of input current, 100 runs were performed to obtain the mean and standard deviation of the spike frequency. EKV models were used for the simulation.

To understand why these transistors stand out, we need to analyze the components of neuron spike interval. The spike interval of this neuron consists of refractory period, charging time and discharging time. The discharging time is much smaller compared to the other components; hence its variations can be ignored. The refractory period is directly affected by the voltage V_{rfr} on the gate of M9 and gate voltage of M12. Hence V_T of M9 and M12 affect f-i curve. During charging time, the total current injecting onto the membrane capacitor is

$$I_{total} = I_{in} - I_{leak} + I_{fd} \quad (6)$$

where I_{fd} is the feedback current from M4. M13 will determine the leakage current. M1, M3, M4 & M5 are on the positive feedback path and would both alter the neuron switching threshold and affect the feedback current. Hence, the number of critical transistors are found to be seven and V_{T1} , V_{T3} , V_{T4} , V_{T5} , V_{T9} , V_{T12} and V_{T13} are the α parameters mentioned in (1).

B. FGMOS based control

As discussed in section II-B, using a special device like FGMOS can allow us to reduce the number of critical transistors to be replaced by carefully designed devices. We shall illustrate this method using the example of the current neuron circuit. We begin by deriving the expression for the spiking time of the circuit for a fixed I_{in} . For this case, the differential equation can be approximated as follows assuming sub-threshold operation of transistors:

$$C_o \frac{dV_{mem}}{dt} \approx I_{in} - I_{13} e^{-\frac{\kappa(V_{leak} - V_{T13})}{U_T}}$$

Fig. 3: **Individual transistor V_T mismatch effect on f-i curve:** Same threshold voltage change on every transistor (a) Critical Transistors (b) NonCritical Transistors

$$+ I_1 e^{-\frac{\kappa(V_{mem} - V_{T1})}{U_T}} \frac{e^{\frac{\kappa(|V_{T3}| - |V_{T5}|)}{U_T}}}{1 + e^{\frac{\kappa(|V_{T4}| - |V_{T5}|)}{U_T}}} = I_{in} - \beta_1 I_{13} + \beta_2 I_1 e^{-\frac{\kappa V_{mem}}{U_T}} \quad (7)$$

If $V = V_{peak}$, $V \leftarrow V_{reset} = 0V$ for $\beta_3 t_{rfr}$ s. where

$$\beta_3 t_{rfr} \approx \frac{C_{g12}(V_{dd} - V_{T12} - \frac{U_T}{\kappa} \ln(\frac{I_{in} - I_{13} e^{-\frac{\kappa(V_{leak} - V_{T13})}{U_T}}}{I_{12}}))}{I_9 e^{-\frac{\kappa(V_{rfr} - V_{T9})}{U_T}}} \quad (8)$$

where I_x is the current that would flow in the transistor M_x if it had both gate voltage and threshold voltage equal to 0, U_T is the thermal voltage and κ is the inverse of the sub-threshold slope of a MOSFET.

Note that the total number of parameters to be controlled, denoted by β following (5), has reduced to three from seven in the earlier section. β_1 depends on V_{T13} , β_2 depends on V_{T1} , V_{T3} , V_{T4} and V_{T5} and β_3 depends on V_{T9} , V_{T12} and V_{T13} . From β_1 , we see M13 has to be a FG device. From β_3 , we see that either M9 or M12 can be a FG device; we chose M9 to be the FG element in our design. The case for β_2 is more complicated. Not all transistors in this equation have similar effects. For example, if M4 is chosen as the FG, it cannot make β_2 arbitrarily large. We decided to match M3, M4 and M5 using common centroid layout since matching the

Fig. 4: **Proposed neuron circuit:** Three FG transistors are used to calibrate out f-i mismatch. This neuron has a negative injection current but still retains positive feedback and controllable refractory period.

threshold voltages of M3 and M4 with M5 would eliminate the mismatch contribution of these three. M1 was chosen as the FG device. Also, this choice allows all three special devices to be a NMOS; as mentioned earlier this is useful while using special devices like CLBT. Even for the FG devices, we chose FG PMOS only since its programming characteristics are well-studied [8]. Hence, we flip the polarity of the neuron circuit by replacing PMOS by NMOS and vice versa to come up with the neuron circuit in Fig. 4. Now, we describe a methodical approach to correct for f-i curve mismatch by separating the effects of each control ‘knob’. We assume that one neuron can be tuned to achieve desired characteristic; other neurons need to be tuned to match with this target neuron.

- The first step is to correct the slope of the f-i curve for currents in the range of interest. Leakage current is first removed from both the target circuit and the mismatched circuit by turning M13 off. Also refractory period is minimized by keeping the current through M9 to be high. V_T of M1 is then modified using the FG to get a close approximation to the desired slope of the target f-i curve.
- The second step is to turn ON leakage current in both circuits; this creates an offset in the f-i curve. Tune M13 in the neuron to increase leak current in small steps to get correct offset.
- The third step is to tune M9 for refractory period correction.

IV. RESULTS

By changing some critical transistors into FG transistors and through calibration, we can correct for neuron circuit mismatch. The expected f-i curve correction procedure is shown in Fig. 5. Circuit in Fig. 1 is used in this simulation as an illustration since it has similar behavior as the one in Fig. 4. The target f-i curve is obtained from the circuit with all transistors having nominal V_T . The random f-i curve is created by simulating the neuron after randomly assigning V_T generated from a Gaussian distribution with $\sigma_{V_T} = 50$ mV to every neuron transistor. It can be seen that using this method, we can correct for very large mismatch in f-i curve through just

Fig. 5: **Mismatch correction:** Correction of f-i curve mismatch based on programming FG transistors

three FG devices. We have taped out a chip containing three arrays of neurons which are made from normal transistors, FG transistors and CLBT which will enable us to measure the matching properties of CLBT and our proposed FG based mismatch correction scheme.

V. CONCLUSION AND DISCUSSION

We presented a methodical approach for analyzing and reducing errors due to mismatch in neuron circuits. We analyzed the causes underlying f-i curve mismatch in a particular neuron example, and proposed to use FG transistor to correct for circuit mismatch. We showed that the number of FG devices needed for correcting mismatch can be smaller than number of critical transistors obtained in sensitivity analysis. Our simulation results show that we can correct for f-i curve mismatch by using only three FG transistors in the particular circuit considered.

REFERENCES

- [1] Carver Mead, *Analog VLSI and Neural Systems*, Addison-Wesley, Reading, MA, 1989.
- [2] A. Hastings, *The Art of Analog Layout*, Prentice Hall, 2006.
- [3] M.J.M. Pelgrom, A.C.J. Duinmaijer, and A.P.G. Welbers, “Matching Properties of MOS Transistors,” *IEEE Journal of Solid-State Circuits*, pp. 1433–1439, 1989.
- [4] M. Oster and Shih-Chii Liu, “A winner-take-all spiking network with spiking inputs,” in *Proceedings of the International Symposium on Circuits and Systems*, May 2004, pp. 203–6.
- [5] E. Neftci and G. Indiveri, “A device mismatch compensation method for VLSI neural networks,” *IEEE Biomedical Circuits and Systems*, pp. 262–265, 2010.
- [6] A. Van Schaik, E. Fragniere, and E. Vittoz, “Improved Silicon Cochlea using Compatible Lateral Bipolar Transistors,” *Advances in neural information processing system*, pp. 671–677, 1996.
- [7] B. Linares-Barranco, T. Serrano-Gotarredona, and R. Serrano-Gotarredona, “Compact low-power calibration mini-DACs for neural massive arrays with programmable weights,” *IEEE Transactions on Neural Networks*, vol. 14, no. 5, pp. 1207–16, Sept 2003.
- [8] A. Bandyopadhyay, G. Serrano, and P. Hasler, “Adaptive Algorithm Using Hot-Electron Injection for Programming Analog Computational Memory Elements Within 0.2% of Accuracy Over 3.5 Decades,” *IEEE Journal of Solid-State Circuits*, vol. 41, no. 9, pp. 2107–14, Sept 2006.
- [9] G. Indiveri, E. Chicca, and R. Douglas, “A VLSI Array of Low-Power Spiking Neurons and Bistable Synapses With Spike-Timing Dependent Plasticity,” *IEEE Transactions on Neural Networks*, vol. 17, no. 1, pp. 211–221, Jan 2006.