Thermal Expansion and Aging Effects in Neuromorphic Signal Processor

Amir Zjajo, Rene van Leuken
Circuits and Systems Group, Delft University of Technology, 2628 CD, Delft, The Netherlands

***Abstract*—In this paper, we propose an efficient methodology based on a real-time estimator and predictor-corrector scheme for accurate thermal expansion profile and aging evaluation of a neuromorphic signal processor circuit components. As the experimental results indicate, for comparable mesh size, the proposed method is 1~2 order of magnitude more accurate than corresponding, generalized finite element method.**

I. Introduction

Neuromorphic signal processing circuits are implemented on optimized, special purpose hardware, which can provide direct one-to-one mapping and low instruction execution redundancy [1]. Accurate computation of temperature at the length-scales of devices and interconnects in neuromorphic circuits requires the development of an analytical model for heat transport, and a detailed accounting of the heat flow from the power-sources through the nanometer-scale layout within the processor. Existing thermal-simulation methodologies, such as the ones based on finite-difference time domain [2], the image method [3], green function [4], or mesh based methods [5], when applied to a full-chip reduce the computational complexity by homogenizing the materials within a layer, limiting the extent of an eigenfunction expansion, or ignoring sources' proximity to boundaries. These simplifications render their results less accurate at fine length-scales, on wires, vias, or individual transistors.

In this paper, we propose efficient methodology for accurate thermal expansion and aging estimation of a neuromorphic circuits components on a nanometer-scale. Due to the different thermal expansions coefficients of materials which form the neuromorphic circuits, repetitive thermal cycling occurred cause time and stress dependent drift of device characteristics, and subsequently, their thermal resistance, which may be a key indicator for condition monitoring and control. In the proposed method, this indicator actively adapt the predictor-corrector scheme, which provide both, steady-state and transient temperature expansion profiles even for geometrically complicated physical structures.

II. Thermal Expansion and Aging Effects in a Neuromorphic Processor

The structure of the neuromorphic core (Figure 1) [6] implemented in 65 nm CMOS technology consists of input decoder that connects 1024×256 programmable synapses to 256 integrate-and-fire neurons, the I/O network communication layer, and activity-dependent dynamic voltage and frequency scaling (DVFS) circuits for active power reduction. The integrate-and-fire neuron circuits are current-mode, conductance-based, process input data on demand, in real time, and produce fast asynchronous digital output pulses. The circuit is compact, adaptive, low power, implements refractory period and spike-frequency adaptation, and has biologically realistic time constants.

We model the neuromorphic processor circuit as a regular mesh (according to the information provided by the layout geometry) containing N_i discretized elements, i.e. a thermal grid, with the ambient temperature corresponding to ground. All boundary conditions are determined by the operating environment. Each element i in the mesh is connected to n neighbors via resistors, and has a ground capacitance C_i and ground resistance R_{is}. Given n neighbors, T_{amb}, $T_i(t)$ and $T_n(t)$ are the ambient temperature, the temperature of element i at time t and the temperature of element i's neighbor n, respectively. $P_i(t)$ is the heat source coupled to element i, and R_{in} is the resistance between element i and its neighbor n. The modeled thermal grid is than expressed as

$$C_i \frac{dT_i(t)}{dt} + \left(\sum_{n \in N_i} \frac{T_i(t) - T_n(t)}{R_{in}} \right) + \frac{T_i(t) - T_{amb}}{R_{is}} - P_i(t) = 0 \quad (1)$$

which can be simplified to

$$\frac{dT_i'(t)}{dt} = \frac{1}{C_i} \left(\sum_{n \in N_i} \frac{T_n'(t)}{R_{in}} - \eta_i T_i'(t) + P_i(t) \right) \quad (2)$$

where $\eta_i = \sum_n (1/R_{in}) + 1/R_{is}$ and the normalized temperature of element i, $T_i' = T_i - T_{amb}$. The steady state temperature of each location (x,y) across the silicon die at time t, $T(x,y,t) = \sum_n R_{\Sigma,i} + 1/P_i(t)$, where $R_{\Sigma i}$ is the thermal correlation (thermal resistance) between heat source i and location (x,y). By the nature of device aging [7], a linear approximation of thermal resistance upon the number of its thermal cycles can be assumed

$$R_{i(aged)} = R_i \left(1 + \frac{T_{meas}(t) - T_{estim}(t)}{P_i(t) R_{tot}} \right) \quad (3)$$

where $R_{i(aged)}$ is the updated thermal resistance in the degraded thermal path, T_{meas} and T_{estim} are the measured and estimated temperatures, and R_{tot} is the sum of the thermal resistances without aging effects. Assuming that the j-th neighbor of element i is k_j, we can define G_i as

$$G_{i(aged)} = (\underbrace{0,\ldots,0}_{k_1-1}, \frac{1}{R_{i1(ageed)}}, \underbrace{0,\ldots,0}_{k_2-1}, \frac{1}{R_{i2(aged)}}, \ldots, -\eta, \ldots, \frac{1}{R_{in_i(aged)}}, \ldots, 0) \quad (4)$$

where $-\eta$ is the i-th entry of the vector, and all other entries are 0s, and $1/R_{ij(aged)}$ is the k_j-th entry of vector $G_{i(aged)}$, representing the degraded thermal conductance between i and its j-th neighbor k_j. To solve ODE in (3) we utilize a discontinuity detector as in [8]. Subsequently, the modified, third-order predictor-corrector Runge-Kutta scheme reads

This research was supported in part by the European Union and the Dutch government, as part of the CATRENE program under RESIST project.

$$
\begin{aligned}
&(1+\Delta tC^{-1}P(T^{l}))\hat{T}^{(1)} = \hat{T}^{l} + \Delta tC^{-1}G_{(aged)}(T^{l}) \\
&(1+1/4\,\Delta tC^{-1}P(T^{(1)}))\hat{T}^{(2)} = 1/4\,(3\hat{T}^{l} + \hat{T}^{(1)} + \Delta tC^{-1}G_{(aged)}(T^{(1)})) \\
&(1+2/3\,\Delta tC^{-1}P(T^{(2)}))\hat{T}^{l+1} = 1/3\,(\hat{T}^{l} + 2\hat{T}^{(2)} + 2\Delta tC^{-1}G_{(aged)}(T^{(2)}))
\end{aligned} \tag{5}
$$

for two time instants T^{l} and T^{l+1}. Convergence to steady state is further accelerated using a multigrid technique, e.g. the original fine mesh is coarsened a number of times, and the solution on the coarse meshes is used to accelerate convergence to steady state on the fine mesh.

III. Experimental Results

All the experimental results are carried out on a single processor Ubuntu Linux 9.10 system with Intel Core 2 Duo CPUs 2.66 GHz processor and 6 GB of memory. The circuit netlist is simulated in Cadence Spectre using 65nm CMOS model files. The simulation date points are processed with a PERL script and fed back into the MatLab code, where proposed method and all sparse techniques have been implemented. Modeled thermal conductivity of silicon is 148W/(mK), copper interconnect is 383W/(mK), and the silicon resistance is 0.02K/W. To illustrate in simulation plot only the temperature increment due to a device activity, the ambient temperature is assumed to be 0K. As a representative example, in Figure 2, we show a temperature profile estimation of the neuromorphic core interconnect. The profile is computed within 1.4 seconds, and it shows that our solver can calculate temperature of multiple wires very accurately: output in our simulator ranges from 0.114K to 0.471K. For comparison purposes, we implemented generalized finite element method, which can be found in several commercially available software packages (e.g. [9]). Figure 3 illustrates that the proposed method (several $G_{(aged)}$ are shown) is 1~2 order of magnitude more accurate for comparable mesh size than corresponding generalized finite element method. For more complex circuit structures, additional increase has been observed.

IV. Conclusion

This paper presents the real-time thermal expansion and aging effects estimator constrained with the surface boundary conditions. As the results indicate, for comparable mesh size, the estimation errors, e.g. arising from aging of physical layers, of the predictor-corrector scheme are 1~2 order of magnitude lower than corresponding, generalized finite element method.

References

[1] C. Zamarreno-Ramos, *et al.* "On spike-timing-dependent-plasticity, memristive devices, and building a self-learning visual cortex," *Frontiers in Neuroscience*, vol. 5, pp. 1-36, 2011.

[2] T.T. Wang, Y.M. Lee, C.C.P. Chen, "3D thermal ADI - an efficient chip-level transient thermal simulator," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems,* vol. 21, no. 12, pp. 1434-1445, 2002.

[3] K.J. Scott, "Electrostatic potential Green's functions for multi-layered dielectric media," *Philips Journaal of Research,* vol. 45, pp. 293-324, 1990.

[4] A.M. Niknejad, *et al.*, "Numerically stable Green function for modeling and analysis of substrate coupling in integrated circuits," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems,* vol. 17, no. 4, pp. 305-315, 1998.

[5] N. Allec, *et al.*, "ThermalScope: Multi-scale thermal analysis for nanometer-scale integrated circuits," *IEEE International Conference on Computer-Aided Design*, pp. 603-610, 2008.

[6] A. Zjajo, N. Mehta, R. van Leuken, "A 31 pJ/spike hybrid stochastic neuromorphic signal processor," *IEEE Signal Procesing in Medicine and Biology Symposium*, pp. 1-2, 2015.

[7] W. Wang, *et al.*, "Compact modeling and simulation of circuit reliability for 65-nm CMOS technology", *IEEE Transactions on Device and Materials Reliability,* vol. 7, no. 4, pp. 509-517, 2007.

[8] A. Zjajo, N. van der Meijs, R. van Leuken, "Thermal analysis of 3D integrated circuits based on discontinuous Galerkin finite element method", *IEEE International Symposium on Quality Electronic Design*, pp. 117-122, 2012.

[9] Ansys 10.0, http://www.ansys.com

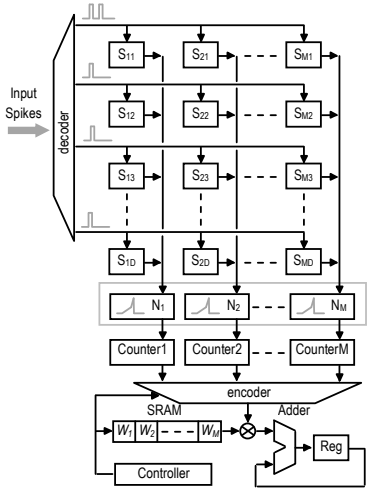

Figure 1: Overview of the neuromorphic core including adaptive, integrate-and-fire neuron-, and synapse circuits, respectively, SRAM, adder, multiplier, register and controller.

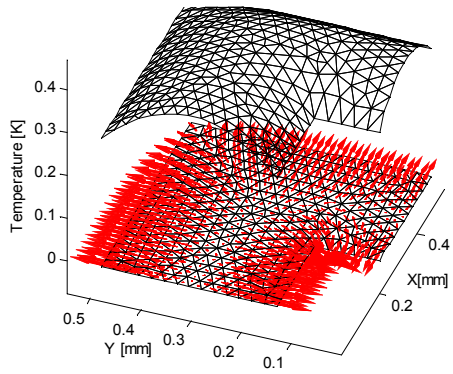

Figure 2: Temperature profile of neuromorphic processor interconnect at the (critical) crossings. Thermal expansion and aging effects are illustrated with arrows.

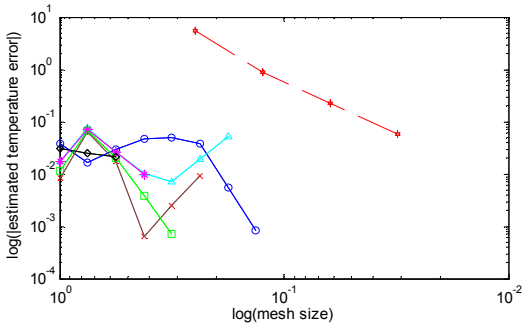

Figure 3: Estimated temperature error versus mesh size for the proposed (several $G_{(aged)}$ bold lines), and generalized finite element method (red dashed line); neuromorphic processor interconnect example.