frontiers in
NEUROSCIENCE

OPINION ARTICLE
published: 22 September 2011
doi: 10.3389/fnins.2011.00108

Neuromorphic silicon neurons and large-scale neural networks: challenges and opportunities

Chi-Sang Poon*[1] *and Kuan Zhou*[2]

[1] *Harvard–MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA, USA*
[2] *Intel Corporation, Hillsboro, OR, USA*
**Correspondence: cpoon@mit.edu*

NEUROMORPHIC SILICON NEURONS: STATE OF THE ART

Complementary metal-oxide-semiconductor (CMOS) transistors are commonly used in very-large-scale-integration (VLSI) digital circuits as a basic binary switch that turns on or off as the transistor gate voltage crosses some threshold. Carver Mead first noted that CMOS transistor circuits operating below this threshold in current mode have strikingly similar sigmoidal current–voltage relationships as do neuronal ion channels and consume little power; hence they are ideal analogs of neuronal function (Mead, 1989). This unique device physics led to the advent of "neuromorphic" silicon neurons (SiNs) which allow neuronal spiking dynamics to be directly emulated on analog VLSI chips without the need for digital software simulation (Mahowald and Douglas, 1991). In the inaugural issue of this Journal, Indiveri et al. (2011) review the current state of the art in CMOS-based neuromorphic neuron circuit designs that have evolved over the past two decades. The comprehensive appraisal delineates and compares the latest SiN design techniques as applied to varying types of spiking neuron models ranging from realistic conductance-based Hodgkin–Huxley models to simple yet versatile integrate-and-fire models. The timely and much needed compendium is a tour de force that will certainly provide a valuable guidepost for future SiN designs and applications.

NEUROMORPHIC SILICON NEURONS VS. DIGITAL NEURAL SIMULATIONS

For all the impressive technical feats in neuromorphic engineering, a basic question often from those uninitiated to the field is: why SiN? To be sure, neural modeling "*in silico*" as practiced today is still largely digital software-based rather than analog silicon chip-based, and for obvious reasons. Digital simulation enjoys double-precision and virtually noise-free numerical outputs that are highly reproducible and readily re-programmable; certainly no way with analog. However, this archetypal reasoning misses the fact that neural computing itself is inherently analog and noisy/variable even though neuron-to-neuron communication is predominantly digital via all-or-none spiking. Computational precision and repeatability are really the least of a neuron's concerns. Instead, biological neural networks excel in performing massive high-speed computations in parallel under noisy and variable environments, all at minute power consumption within a tiny anatomic space. By contrast, even today's most high-power supercomputing clusters cannot simulate neocortical or thalamocortical connectivity in real time (Markram, 2006; Izhikevich and Edelman, 2008; Ananthanarayanan et al., 2009). In comparison, SiNs offer a practical computational medium that is intermediate between biological neurons and digital computers in terms of power and space efficiencies, and is orders of magnitude faster than real neurons in terms of computing speed (**Figure 1**). Unlike numerical simulation on general-purpose serial computers, computation delay of analog SiN circuits is independent of network size except for signal propagation. SiNs will therefore be most useful when large-scale dedicated neural computing is desired in real time and under stringent power and space/weight

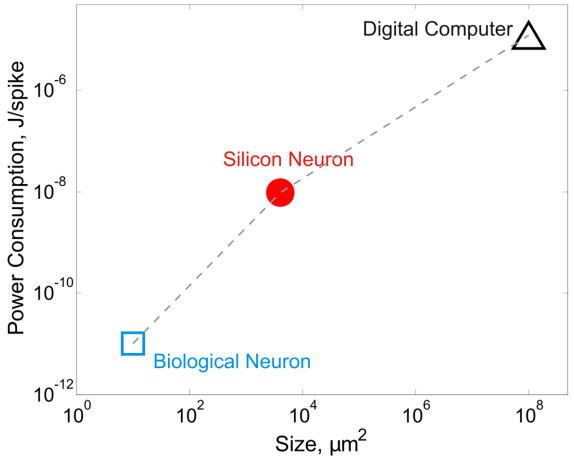

FIGURE 1 | Biological and silicon neurons have much better power and space efficiencies than digital computers. A biological neuron consumes approximately 3.84×10^8 ATP molecules in generating a spike (Attwell and Laughlin, 2001; Lennie, 2003). Assuming 30–45 kJ released per mole of ATP (Berg et al., 2007; or 5–7.5 × 10^{-20} J per ATP molecule), the energy cost of a neuronal spike is in the order of 10^{-11} J. The density of neurons under cortical surface in various mammalian species is ~100,000/mm^2 (Braitenberg and Schüz, 1998), which translates to a span of ~10 μm^2 per neuron. Silicon neurons have power consumption in the order of 10^{-8} J/spike on a biological timescale. For example, an Integrate-and-Fire neuron circuit consumes 3–15 nJ at 100 Hz (Indiveri, 2003) and a compact neuron model consumes 8.5–9.0 pJ at ~1 MHz (Wijekoon and Dudek, 2008), which translates to 85–90 nJ at 100 Hz. For silicon neurons, the on-chip neuron area is estimated to be ~4,000 μm^2 (70 μm × 40 μm in Wijekoon and Dudek, 2008, ~3750 μm^2 in Vogelstein et al., 2007, and 70 μm × 70 μm in Wijekoon and Dudek, 2009). According to Liu and Delbrück (2007), digital computers are 10^4–10^8 less efficient than biological neurons. The power efficiency of digital computers is therefore estimated to be 10^{-3}–10^{-7} J/spike. Most current multi-core digital microprocessor chips have dimensions from 263 to 692 mm^2. A single core has an average size from ~50 to ~90 mm^2.

constraints, such as in neuroprosthetic, brain–computer interface, or embedded machine intelligence applications.

BUILDING ROBUST LARGE-SCALE IONO-NEUROMORPHIC SILICON NEURAL NETWORKS

Emulating neuronal spiking on SiNs is only the first step in neuromorphic modeling. Building large-scale SiN networks on VLSI chips to mimic complex brain functions remains a great challenge, not least because subthreshold CMOS circuits are notoriously highly susceptible to mismatch in transistor threshold voltage and current factor caused by fabrication imperfections and temperature variations (Pavasovic et al., 1994; de Gyvez and Tuinhout, 2004; Kinget, 2005; Andricciola and Tuinhout, 2009). The intrinsic VLSI process variability severely limits the scalability of SiN networks since it is not practicable to fine-tune a large number of analog transistor circuits to correct for mismatch on chip after fabrication. Although SiN designs that operate in the above-threshold regime are less sensitive to transistor mismatch (Indiveri et al., 2011), such circuits require much higher currents and power consumption (several orders of magnitude higher), and are non-ideal for VLSI fabrication at a high transistor density.

Another layer of complexity for large-scale neuromorphic modeling is the need for biological realism as constrained by experimental data (Markram, 2006; Djurfeldt et al., 2008). Biological neural networks are not just a large collection of spiking neurons interconnected via plastic chemical synapses or gap junctions (Sporns et al., 2005; DeFelipe, 2010). Instead, neurons and glia cells are endowed with a panoply of membrane and intracellular properties which confer a myriad of complex glial/neuronal, dendritic, axonal, and synaptic dynamics at the cellular level and resultant emergent dynamics at the network level. Such multiscale spatiotemporal dynamics in large-scale SiN networks are difficult to emulate on chip because of inevitable transistor mismatch which may vary across the entire VLSI network. To put this in perspective, traditional subthreshold current-mode differential pair circuits commonly used for emulating the sigmoidal current–voltage relationship of ion channels have a limited input voltage dynamic range of $< \pm 100$ mV. Since typical CMOS threshold voltage may vary by ±20 mV (3 standard deviations) or more (ITRS, 2007), the worst-case mismatch errors for single devices could be in excess of 100 mV or 100% across the chip and may be further compounded as network size increases and the temperature varies especially for deep submicron processes. Similar dynamic range limitations also apply to other basis circuits such as transconductance amplifier, translinear multiplier, current mirror, etc. Although integrate-and-fire models of neuronal spiking and bursting behaviors are more amenable to implementation on chip because of their simplicity (Indiveri et al., 2011), these phenomenological SiN models are non-mechanistic and their requisite parameter settings are not readily adjustable *en masse* post-fabrication or during computation in response to changes in neuronal inputs. Indeed, all subthreshold circuits including phenomenological SiN circuits are subject to similar threshold voltage mismatch limitations regardless of whether they are designed in voltage or current mode, although the sensitivity of phenomenological circuits to mismatch has not been systematically evaluated.

To circumvent these difficulties, a novel subthreshold current-mode circuit design approach has been recently proposed that emphasizes the importance of a wide dynamic range for input voltages in neuromorphic modeling of ion channel and intracellular ionic dynamics (iono-neuromorphic dynamics) in large-scale SiN networks (Rachmuth and Poon, 2008). With judicious circuit designs using source degeneration and other negative feedback techniques, the dynamic range for input voltages of neuromorphic ion channels and other circuits has been extended to >1 V, making them much more robust to mismatch errors. The increased robustness of subthreshold SiNs significantly improves the scalability of iono-neuromorphic networks and allows more faithful reproduction of complex neuronal dynamics, such as chaotic bursting in pacemaker neurons (Rachmuth and Poon, 2008).

NEXT-GENERATION LARGE-SCALE IONO-NEUROMORPHIC SILICON NEURAL NETWORKS

Current subthreshold SiN circuits are susceptible to mismatch because they rely on commercial CMOS processes that are intended for mainstream performance-driven VLSI digital circuits rather than analog or mixed-signal electronics. The good news is that the trend is rapidly changing in recent years due to phenomenal demands for low-power system-on-chip (SoC) applications such as smartphones, wearable electronics, portable medical devices, etc. Processes that are dedicated to low-power subthreshold SoC circuits are already on the horizon. For example, a newly available fully depleted silicon-on-insulator technology optimized for ultra-low-power subthreshold circuit applications allows significant reduction in threshold voltage variation and device capacitance when compared with conventional CMOS transistors (Vitale et al., 2011). This novel subthreshold CMOS technology is already making its way to iono-neuromorphic analog VLSI circuits (Meng et al., 2011).

As the performance of subthreshold SiNs continues to improve, the scale of SiN networks will be ultimately limited at the chip level only by the ever-growing capacity of VLSI technology. The recent introduction of the 22-nm bulk CMOS three-dimensional (non-planar) Tri-Gate process (Intel, 2011) brings a new dimension to VLSI device scaling that is likely to continue to fuel Moore's Law for the next decade. Concurrently, recent advances in three-dimensional integrated circuit technology have made it possible to stack multiple interconnected bulk CMOS dies on top of one another, allowing increased effective VLSI footprint with dramatically shortened interconnect wire lengths (Davis et al., 2005; Mak et al., 2011). These enabling technologies open new and exciting avenues for the next generation of large-scale iono-neuromorphic SiN networks.

Another emerging approach to large-scale neuromorphic modeling employs nanotechnologies such as nanowires, carbon nanotubes, memristors, etc. as the computational analog, either independently or in combination with CMOS technology. For example, memristor-based neuromorphic devices have been recently reported that are capable of emulating spike-timing-dependent synaptic plasticity in a crossbar network with massive connectivity between CMOS-based SiNs (Jo et al., 2010; Zamarreno-Ramos et al., 2011). Despite their promise, these nanocircuits are still fraught with similar (if not greater) problems of

non-robustness as CMOS devices, and are less efficient in terms of yields and other performance benchmarks (Chau et al., 2005). From a neuroscience perspective, a fundamental drawback of such nano-sized neuronal analogs is that they are largely phenomenological (artificial) rather than iono-neuromorphic (realistic) models of neuronal function. Future neuromorphic modeling efforts should therefore target not only the integration density and computation and/or power efficiencies of the neuronal analogs (**Figure 1**) but also their degree of biological realism and robustness, if the full glory of the brain is ever to be captured on chip.

REFERENCES

Ananthanarayanan, R., Esser, S. K., Simon, H. D., and Modha, D. S. (2009). "The cat is out of the bag: cortical simulations with 10^9 neurons, 10^{13} synapses," in *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis* (Portland, OR: ACM), 1–12.

Andricciola, P., and Tuinhout, H. P. (2009). The temperature dependence of mismatch in deep-submicrometer bulk MOSFETs. *IEEE Electron Device Lett.* 30, 690–692.

Attwell, D., and Laughlin, S. B. (2001). An energy budget for signaling in the grey matter of the brain. *J. Cereb. Blood Flow Metab.* 21, 1133–1145.

Berg, J. M., Tymoczko, J. L., and Stryer, L. (2007). *Biochemistry*, 6th Edn. New York: W. H. Freeman.

Braitenberg, V., and Schüz, A. (1998). *Cortex: Statistics and Geometry of Neuronal Connectivity*, 2nd Edn. Berlin: Springer.

Chau, R., Datta, S., Doczy, M., Doyle, B., Jin, J., Kavalieros, J., Majumdar, A., Metz, M., and Radosavljevic, M. (2005). Benchmarking nanotechnology for high-performance and low-power logic transistor applications. *IEEE Trans. Nanotechnol.* 4, 153–158.

Davis, W. R., Wilson, J., Mick, S., Xu, M., Hua, H., Mineo, C., Sule, A. M., Steer, M., and Franzon, P. D. (2005). Demystifying 3D ICs: the pros and cons of going vertical. *IEEE Des. Test Comput.* 22, 498–510.

de Gyvez, J. P., and Tuinhout, H. P. (2004). Threshold voltage mismatch and intra-die leakage current in digital CMOS circuits. *IEEE J. Solid-State Circuits* 39, 157–168.

DeFelipe, J. (2010). From the connectome to the synaptome: an epic love story. *Science* 330, 1198–1201.

Djurfeldt, M., Ekeberg, O., and Lansner, A. (2008). Large-scale modeling – a tool for conquering the complexity of the brain. *Front. Neuroinform.* 2:1. doi: 10.3389/neuro.11.001.2008

Indiveri, G. (2003). "A low-power adaptive integrate-and-fire neuron circuit," in *Proceedings of the 2003 International Symposium on Circuits and Systems*, Bangkok, IV-820–IV-823.

Indiveri, G., Linares-Barranco, B., Hamilton, T. J., van Schaik, A., Etienne-Cummings, R., Delbruck, T., Liu, S.-C., Dudek, P., Häfliger, P., Renaud, S., Schemmel, J., Cauwenberghs, G., Arthur, J., Hynna, K., Folowosele, F., Saïghi, S., Serrano-Gotarredona, T., Wijekoon, J., Wang, Y., and Boahen, K. (2011). Neuromorphic silicon neuron circuits. *Front. Neurosci.* 5:73. doi: 10.3389/fnins.2011.00073

Intel. (2011). *New Transistors for 22 Nanometer Chips Have an Unprecedented Combination of Power Savings and Performance Gains*. Hillsboro, OR: Intel Corp.

ITRS. (2007). *International Technology Roadmap for Semiconductors*. Available at: http://www.itrs.net/

Izhikevich, E. M., and Edelman, G. M. (2008). Large-scale model of mammalian thalamocortical systems. *Proc. Natl. Acad. Sci. U.S.A.* 105, 3593–3598.

Jo, S. H., Chang, T., Ebong, I., Bhadviya, B. B., Mazumder, P., and Lu, W. (2010). Nanoscale memristor device as synapse in neuromorphic systems. *Nano Lett.* 10, 1297–1301.

Kinget, P. R. (2005). Device mismatch and tradeoffs in the design of analog circuits. *IEEE J. Solid-State Circuits* 40, 1212–1224.

Lennie, P. (2003). The cost of cortical computation. *Curr. Biol.* 13, 493–497.

Liu, S.-C., and Delbrück, T. (2007). *Introduction to Neuromorphic and Bio-Inspired Mixed-Signal VLSI*. Pasadena, CA: California Institute of Technology.

Mahowald, M., and Douglas, R. (1991). A silicon neuron. *Nature* 354, 515–518.

Mak, T., Al-Dujaily, R., Zhou, K., Lam, K.-P., Meng, Y., Yakovlev, A., and Poon, C.-S. (2011). Dynamic programming networks for large-scale 3D chip integration. *IEEE Circuits Syst. Mag.* 11, 51–62.

Markram, H. (2006). The blue brain project. *Nat. Rev. Neurosci.* 7, 153–160.

Mead, C. (1989). *Analog VLSI and Neural Systems*. Reading, MA: Addison-Wesley.

Meng, Y., Zhou, K., Monzon, J. J. C., and Poon, C.-S. (2011). "Iono-neuromorphic implementation of spike-timing-dependent synaptic plasticity," in *33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC '11)*, Boston, MA.

Pavasovic, A., Andreou, A. G., and Westgate, C. R. (1994). Characterization of subthreshold MOS mismatch in transistors for VLSI systems. *J. Signal Process. Syst.* 6, 75–85.

Rachmuth, G., and Poon, C. S. (2008). Transistor analogs of emergent iono-neuronal dynamics. *HFSP J.* 2, 156–166.

Sporns, O., Tononi, G., and Kotter, R. (2005). The human connectome: a structural description of the human brain. *PLoS Comput. Biol.* 1:e42. doi: 10.1371/journal.pcbi.0010042

Vitale, S. A., Kedzierski, J., Healey, P., Wyatt, P. W., and Keast, C. L. (2011). Work-function-tuned TiN metal gate FDSOI transistors for subthreshold operation. *IEEE Trans. Electron Devices* 58, 419–426.

Vogelstein, R. J., Mallik, U., Vogelstein, J. T., and Cauwenberghs, G. (2007). Dynamically reconfigurable silicon array of spiking neurons with conductance-based synapses. *IEEE Trans. Neural Netw.* 18, 253–265.

Wijekoon, J. H. B., and Dudek, P. (2008). Compact silicon neuron circuit with spiking and bursting behaviour. *Neural Netw.* 21, 524–534.

Wijekoon, J. H. B., and Dudek, P. (2009). "A CMOS circuit implementation of a spiking neuron with bursting and adaptation on a biological timescale," in *Biomedical Circuits and Systems Conference, 2009, BioCAS 2009, IEEE*, Beijing, 193–196.

Zamarreno-Ramos, C., Camunas-Mesa, L. A., Perez-Carrasco, J. A., Masquelier, T., Serrano-Gotarredona, T., and Linares-Barranco, B. (2011). On spike-timing-dependent-plasticity, memristive devices, and building a self-learning visual cortex. *Front. Neurosci.* 5:26. doi: 10.3389/fnins.2011.00026

Received: 05 July 2011; accepted: 29 August 2011; published online: 22 September 2011.

*Citation: Poon C-S and Zhou K (2011) Neuromorphic silicon neurons and large-scale neural networks: challenges and opportunities. Front. Neurosci. **5**:108. doi: 10.3389/fnins.2011.00108*

This article was submitted to Frontiers in Neuromorphic Engineering, a specialty of Frontiers in Neuroscience.