Guest Editorial: Special Section on Parallel and Distributed Computing Techniques for Non-Von Neumann Technologies

Scott Pakin, Christof Teuscher, and Catherine Schuman

The title of this special section, *Parallel and Distributed Computing Techniques for Non-Von Neumann Technologies*, is a bit misleading. Technically, parallel and distributed computers already diverge from the basic architecture that John von Neumann proposed in 1945, although they are still based on processors that execute a sequence of instructions, each of which performs a simple action such as computing an arithmetic result, reading or writing memory, or branching to a new location in the instruction sequence. But what would a computer that is not based on this model of execution look like?

Suppose it is not the CPUs that perform computations but the memory system. Suppose there were no instruction set at all, and computations are performed as they are in a mammalian brain, by multicasting electrical signals if the amount of energy stored locally increases beyond some threshold. Or suppose that computations are performed by establishing a set of constraints and letting the system settle into a solution that honors as many constraints as possible.

And who says that processors need to be made of silicon? Can verotoxin molecules extracted from *E. coli* or various types of biological tissue also be used for computation?

The technologies discussed in the following articles are more exotic, more innovative, and more intriguing than what you are likely to encounter in a typical collection of peer-reviewed computer-science articles. We hope that these articles will help you view computing in a new light and give you a sense of what the future of computing may look like.

Silent-PIM: Realizing the Processing-in-Memory Computing with Standard Memory Requests, by Kim *et al*. [A1], presents a new approach to performing computation within the computer's memory system. The proposed processing-in-memory (PIM) units are "silent" in the sense that they are controlled by standard memory operations issued by the CPU, which implies that no hardware modifications are required except for the PIMs themselves; high-speed DMA can be used for data transfers; and computation offloaded to the memory system can execute concurrently with the CPU.

Look-up-Table based Processing-in-Memory Architecture with Programmable Precision-Scaling for Deep Learning Applications, by Sutradhar *et al*. [A2], bases a PIM architecture on look-up-tables (LUTs) instead of biwise-parallel operations. By replacing complex computation with pre-calculated results stored in LUTs, the authors were able to improve both performance and energy efficiency on neural-network inference tasks relative to CPUs, GPUs, and other PIMs.

Memristors, non-linear nanoscale devices whose resistance can be changed by applying a current or voltage, have emerged as promising candidate building blocks for non-Von Neumann architectures over the last decade. *GIRAF: General Purpose In-Storage Resistive Associative Framework*, by Yavits, Kaplan, and Ginosar [A3], presents a new and parallel in-storage architecture and algorithm framework based on Resistive Content Addressable Memory (RCAM). As opposed to other architectures, in GIRAF, each memory bit is directly connected to processing transistors, thus providing an ultra-high peak memory bandwidth, which scales linearly with GIRAF size.

Memristors have also been shown to be a compelling candidate to implement neuromorphic hardware. However, the lack of endurance of these devices is a concern, especially with respect to the minimum lifetime of the neuromorphic system. *Endurance-Aware Mapping of Spiking Neural Networks to Neuromorphic Hardware*, by Titirsha *et al*. [A4], introduces an approach for mapping neural networks to memristive crossbar-based neu romorphic hardware that is endurance-aware. By incorporating the knowledge of the characteristics of the individual devices in the crossbar into the mapping procedure, the authors show a significant improvement in effective lifetime of the memristive crossbar over other mapping procedures.

With the explosion of non-von Neumann architectures used in post-Moore's law computing, it can be difficult to know which systems are most appropriate for which tasks and how much of an advantage these novel architectures provide over classical systems. *Fast Post-Hoc Normalization for Brain Inspired Sparse Coding on a Neuromorphic Device*, by Henke, Kenyon, and Migliori [A5], presents a normalization technique for sparse coding on the Loihi neuromorphic computing system, and they perform a comparison with classical systems. The authors demonstrate that the Loihi is

- *Scott Pakin is with the Computer, Computational, and Statistical Sciences Division, Los Alamos National Laboratory, Los Alamos, NM 87545 USA. E-mail: pakin@lanl.gov.*
- *Christof Teuscher is with the Department of Electrical and Computer Engineering, Portland State University, Portland, OR 97201 USA. E-mail: teuscher@pdx.edu.*
- *Catherine Schuman is with the Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831 USA. E-mail: schumancd@ornl.gov.*

Digital Object Identifier no. 10.1109/TPDS.2021.3093148

approaching the performance of classical systems on this task, and the new approach allows for faster, more accurate solutions than full simulation.

A *quantum annealer* searches for ground-state configurations of an Ising-model Hamiltonian function—or equivalently, the Boolean values that minimize a quadratic unconstrained binary optimization (QUBO) problem—through an annealing process based on quantum fluctuations. Due to the nature of quantum mechanics, the execution of quantum annealing is fundamentally opaque. However, *Inferring the Dynamics of Ground-State Evolution of Quantum Annealers*, by Pelofske, Hahn, and Djidjev [A6], demonstrates a technique for "peeking inside" the evolution of the quantum state by performing repeated runs that are quenched after increasing amounts of time. This technique leads to insights on ground-state evolution and the system's freeze-out characteristics.

As Moore's law approaches physical limits, devices are now being measured in units of atoms. *Protein Structured Reservoir Computing for Spike-based Pattern Recognition*, by Tsakalos, Sirakoulis, Adamatzky, and Smith [A7], demonstrates an approach for reservoir computing using a single protein molecule, where the atoms of the molecule are the neurons of the reservoir, and the molecular structure governs the reservoir's architecture. With this approach, the authors are able to demonstrate that the protein molecule, in combination with a readout layer, is able to achieve acceptable accuracy on a handwritten-digit classification task, laying the groundwork for exploring future protein molecule-based computing tasks.

A *membrane* or *P system* is an abstract and parallel computing model inspired by how biological cells process chemical compounds in their compartmental structure. In tissue P systems, the cells form a network. In *Monodirectional Evolutional Symport Tissue P Systems with Promoters and Cell Division* by Song, Li, and Zeng [A8], monodirectional evolutional symport tissue P systems with promoters (MESTP P systems) are introduced. The authors show that MESTP P systems are universal when two cells, at most one promoter, and all evolutional symport rules having a maximal length of two are employed. By introducing a cell division mechanism, MESTP P systems are also able to solve the SAT problem.

The editors wish to thank the paper reviewers for their efforts in reading each paper, judging its merits, providing actionable guidance to the authors, and helping us whittle down the initial collection of submissions to the eight papers described above. All of the papers in this Special Section underwent multiple rounds of reviews, which implies a large time commitment on the part of the review committee. The editors also wish to thank the authors for their hard work in performing the research and for preparing high-quality submissions for this Special Section on Parallel and Distributed Computing Techniques for Non-Von Neumann Technologies. Enjoy!

Appendix: Related Articles

[A1] C. H. Kim *et al.*, "Silent-PIM: Realizing the processing-in-memory computing with standard memory requests," *IEEE Trans. Parallel Distrib. Syst.*, vol. 33, no. 2, pp. 251–262, Feb. 2022.

[A2] P. R. Sutradhar *et al.*, "Look-up-table based processing-in-memory architecture with programmable precision-scaling for deep learning applications," *IEEE Trans. Parallel Distrib. Syst.*, vol. 33, no. 2, pp. 263–275, Feb. 2022.

[A3] L. Yavits, R. Kaplan, and R. Ginosar, "GIRAF: General purpose in-storage resistive associative framework," *IEEE Trans. Parallel Distrib. Syst.*, vol. 33, no. 2, pp. 276–287, Feb. 2022.

[A4] T. Titirsha *et al.*, "Endurance-aware mapping of spiking neural networks to neuromorphic hardware," *IEEE Trans. Parallel Distrib. Syst.*, vol. 33, no. 2, pp. 288–301, Feb. 2022.

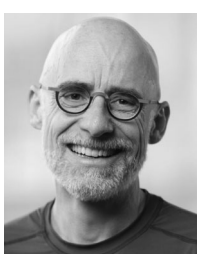
[A5] K. Henke, G. T. Kenyon, and B. Migliori, "Fast post-hoc normalization for brain inspired sparse coding on a neuromorphic device," *IEEE Trans. Parallel Distrib. Syst.*, vol. 33, no. 2, pp. 302–309, Feb. 2022.

[A6] E. Pelofske, G. Hahn, and H. Djidjev, "Inferring the dynamics of the state evolution during quantum annealing," *IEEE Trans. Parallel Distrib. Syst.*, vol. 33, no. 2, pp. 310–321, Feb. 2022.

[A7] K.-A. Tsakalos, G. Ch. Sirakoulis, A. Adamatzky, and J. Smith, "Protein structured reservoir computing for spike-based pattern recognition," *IEEE Trans. Parallel Distrib. Syst.*, vol. 33, no. 2, pp. 322–331, Feb. 2022.

[A8] B. Song, K. Li, and X. Zeng, "Monodirectional evolutional symport tissue P systems with promoters and cell division," *IEEE Trans. Parallel Distrib. Syst.*, vol. 33, no. 2, pp. 332–342, Feb. 2022.

Scott Pakin received the BS degree in mathematics/computer science from Carnegie Mellon University, and the MS and PhD degrees in computer science from the University of Illinois at Urbana-Champaign. He has been a scientist with Los Alamos National Laboratory since 2002. He has researched over time a variety of computer science topics related to high-performance computing, including programming models, application performance analysis, energy efficiency, and high-speed communication. Scott has recently begun investigating quantum computing and is currently the technical/scientific point of contact for LANL's D-Wave quantum annealer.



Christof Teuscher received the MSc and PhD degrees in computer science from the Swiss Federal Institute of Technology in Lausanne, in 2000 and 2004, respectively. He is currently a professor with the Department of Electrical and Computer Engineering with joint appointments with the Department of Computer Science and Systems Science Graduate Program. His research interests include next generation computing architectures and paradigms.

Catherine Schuman received the PhD degree in computer science from the University of Tennessee in 2015, where she completed her dissertation on the use of evolutionary algorithms to train spiking neural networks for neuromorphic systems. She is continuing her study of models and algorithms for neuromorphic computing with Oak Ridge National Laboratory (ORNL). She is currently a research scientist with ORNL. She has an adjunct faculty appointment with the Department of Electrical Engineering and Computer Science, University of Tennessee, where she co-leads the TENNLab neuromorphic computing research group. She has authored or coauthored more than 70 publications as well as seven patents in the field of neuromorphic computing. She is the recipient of U.S. Department of Energy Early Career Award in 2019. She was the program chair for International Conference on Neuromorphic Systems (ICONS) and is the co-chair of program committee for 2020 IEEE International Conference on Rebooting Computing.