Memories in the Future of Information Processing

By SANDIP TIWARI, *Fellow, IEEE*

In science and engineering, we usually start by defining terms and axioms—a minimal set—from which a predictive edifice is built. Memory is definable as a frozen representation of an instance of the past. Information, however, has no such clear and concise definition. Shannon's approach of a sequence of bits and information as the sequence's separation from randomness can be recast as the number of questions with "yes" and "no" answers to decode the message from the source. Shannon information entirely skirts the issue of bit patterns—the linkages between bits, which certainly has relevance to information. Pattern matching exists as a very successful information processing approach in the natural world and the physical world. The reason for this puzzle, no different from that for entropy, is that information is both objective with symbolic references and subjective with a meaning. It is this latter subjectiveness that makes it opaque for science. This information quandary is also not dissimilar to that of a biologist seeking to define "life." Biologists use a list of criteria that query the observable effects of "life," a quite unsatisfactory answer to purists.

This special issue aims to explore, with the future potentiation of information processing, the subject of memory in its multitudinous forms in circuits and systems.

A pattern of symbols is a representation consisting of rules, rules that distinguish patterns from randomness and the physical state of the observable in the Landauer sense of "information is physical." To us, in information processing an algorithm is a representation for evaluating a function (a Hamiltonian) and data is a representation of measurements (a state). In the information machine, the algorithm representation transforms the data representation. Place them both together, the executable is data to the compiler but an algorithm to the user. A machine state is the entirety together, just like the phase space, of the position and the flow.

This consequential complexity of bits and patterns has been endemic to information processing whether one tries to force-fit the Turing test for intelligence, or the evolving definitions of what computing is over the ages, or the dilemma that one finds in deterministic approaches near the end of scaling of transistors as minimum atom-sized dimensions appear on the horizon. That memory is central to processing of information in this objective–subjective view follows from the evolution of this processing in the phase space (the machine state space, where memory is essential to symbolically represent the position canonical coordinate). Information processing is a sequence of representations. Algorithms, where an image is processed and local neighbors attempt to be alike to belong with the same object, are "Ising"-like in attempting to align, a rapid interpretation of the image results when conflicting constraints are resolved. This vision-system-like approach works well in physical computation as well employing techniques of statistical mechanics and neural networks. Patterns matter. A perceptual recognition automaton can also learn and evolve using corrective feedback through hidden layer(s) that permits backpropagation for corrective adjustments. Deep learning, with bidirectional multilayers to augment perception and reasoning with sleeping and dreaming, makes it a more powerful information processing approach.

What happened in the past matters and it is in this sense that the future of information processing must place memory front and center. Their importance in traditional approach to information processing, whether in on-chip cache or higher level cache and memory, and as storage, is of course quite well recognized.

This is the rationale of this issue which aims to explore memories and

Digital Object Identifier: 10.1109/JPROC.2015.2448912

information processing multidimensionally by not restricting itself to the traditional scaling and deterministic style of the past six plus decades. It is organized to explore, with the future potentiation of information processing, the subject of memory in its multitudinous forms in circuits and systems.

The first paper, a scanning of the subject by Tiwari, explores the interrelationship between the physics of devices and the statistical consequences when one places them together in an integrated ensemble. Specifically, it draws out from the fluctuations the energy dissipation effects, the behavior of the two classes that memories used in semiconductors can be partitioned in, the dynamic and static effects, the effects of scale connections within the device and finally the interconnection between these and the attempts at subthreshold swing reduction in devices. It also sets the stage for probability-based approaches to computation.

The first section addresses two relatively recent approaches to memory that remain poorly understood, i.e., in the sense that due to the variety of underlying physico–chemical interactions and to an extent lack of control, a variety of behaviors are observed and reasonings posited. They are interesting because of the variety of attributes they show, many of which are also of interest to the alternatives under exploration for information processing. Wouters *et al.* discuss resistive switching memories where reduction–oxidation and phase change are consequential down to the nanometers scale. These memory structures allow 3-D in integration, potential multilevel operation, and some properties in tune with neuromorphic approaches. In the second paper, Zhou and Ramanathan explore metal–insulator phase transition as a path to memories. Phase transitions based on correlation effects, accompanied by changes in crystal symmetry, endow these structures with unique properties that are again potentially of benefit to memories at nanoscale. This paper, in particular, emphasizes VO_2, a rather semiconductor-compatible material in which a number of phase-transition-induced significant property shifts have been observed.

The second section follows with an exploration of memory architectures that are currently foreseen with not too-distant industrial-scale impact. Karam *et al.* explore computing that is centered around memory, newer memory architectures appropriate to such forays as content-addressable, associative, and transactional memory, and tie them to the more recent device explorations. It particularly stresses the mapping that exists between the properties of the memory to the role it can play in information processing. Nair, in the next paper, looks at the future, with the benefit of hindsight, to argue that the problems and applications of interest will force a change in the architecture of computers, that this change will be memory-centric, and that several of the new memory ideas remain with question marks attached. The architectural change related to memory is already visible at the highest level. Applications are moving toward near-data processing. Its descent to lower levels, including in-memory processing, is quite foreseeable because it reduces the flow of data across memory hierarchies.

From here, in the third section, we jump to exploring the current understanding of how information is tackled in the primate's brain. Balasubramanian argues that the enormously low power compared to that of our physical engines and the rapidity with which it performs the informational task is due to specialization of computational tasks within its heterogeneous parts with a hierarchy that spans from the neuron at the smallest scale to the brain at the largest scale. This architectural style firmly founded on memory-based evolutionary-selection-derived learned procedures is very different from the von Neumann form. The visual system is perhaps the most studied part of the brain. Lee, in his paper, relates these accumulated observations to develop specific computational principles that govern how a model of the world is constructed and represented. The Bayesian approach provides a conceptual way to look at the principles and the architectures and perception follows as a statistical inference from an internal model that employs the hierarchy. Taken together, these two papers provide a comprehensive physical science and engineering view that IEEE's community can understand.

The fourth section explores what we understand from our attempts at using some of these natural world principles in our information engines. This is a distinctly different direction of architecture where memory again plays a central role. Neural or neuromorphic is one example of this—pattern recognition being a well-known application of it. Indiveri and Liu review the understanding of memories and information processing via the neuromorphic route. Relative robustness despite slow, variable, and even faulty components is one of the defining attributes of these systems that is very appealing. The work argues that experience with different approaches of the form that the neural network takes and their physical implementation needs further effort so that learning, cognition, and substantially better performance over traditional computing attributes is achieved. In the second paper, Querlioz *et al.* show an application of the neural style through demonstration of an inference engine using spike-timing-dependent plasticity, and in particular characterize the attribute of robustness.

The reader will also find inexact computing (another alternative information processing style) sprinkled in a few of the different publications, particularly the first one that analyzes the landscape.

The fifth section returns back to the scientific foundation and a peek further out through DiVincenzo's discussion of quantum information processing's memory issue. Information, being represented in a quantum superposition, is susceptible to corruption due to noise by interference with the environment. The paper discusses the quantum error correction techniques and topological embodiments that potentially provide memory robustness.

So, this issue starts with a discussion that connects physics to engineering that information processing and its devices including memory bring together in an integrated ensemble, explores two mainstream emerging memory-oriented directions and then repeats this with a discussion of architectures and their evolution. This sets the stage for a discussion of memory in a deeper sense, by an exploration of the understanding of the brain and a model of its mechanisms. This is followed by a review of the practical implementations of the bioinspired information processing efforts and examples of their successes. We complete this loop of science and its practice by returning to a speculative and advanced theme exploring the questions related to memory in quantum computation.

It is hoped that the reader will find the issue an incisive and appealing peek at the future of information processing in its memory-centric foundations.

It is only by tackling these challenges of the underlying edifice that we will reach domestication of the computer through simplicity from complexity, rather than the state of Marshall McLuhan's "the medium is the message" that is currently seen in its daily use in our mobile life. ■

ABOUT THE GUEST EDITOR

Sandip Tiwari (Fellow, IEEE) is the Charles N. Mellowes Professor in Engineering at Cornell University, Ithaca, NY, USA and a Visiting Professor at Université de Paris-Sud, Orsay, France. His contributions to engineering have included the invention of nanocrystal memories, as a team partner the first demonstration of SiGe bipolar transistor and a variety of others of fundamental importance—theoretical and experimental—in electronic and optical devices, circuits, and architectures. Starting in 2015, Oxford University Press will publish an *Electroscience* series—a sequence of textbooks authored by him—whose first issue is a text *Nanoscale Device Physics* for graduate engineering and science curricula. He also served as the Director of U.S. networks for nanotechnology facilities during 1999–2012.

Prof. Tiwari was the founding Editor-in-Chief of the IEEE TRANSACTIONS ON NANOTECHOLOGY. Among different recognitions of his contribution is the Cledo Brunetti award of the IEEE.