Brain Inspired Computing

R. Stanley Williams
Hewlett Packard Labs
stan.williams@hpe.com

Abstract

The computer industry is facing a potential existential crisis: how to continue exponential increases in computing performance and efficiency after the end of transistor size scaling at roughly 5 nm feature sizes, which in principle is only two generations or as little as four years away. This challenge has demanded a great deal of attention from a variety of industrial, government and academic experts in various aspects of computing. In particular, the IEEE Rebooting Computing Initiative has held four Summits at which evolutionary and revolutionary approaches for computing ranging from new types of transistors to new architectures have been proposed and examined. In addition, the International Technology Roadmap for Semiconductors (ITRS), which for over two decades coordinated the efforts of the semiconductor industry and its supply chain to keep integrated circuits for processors and memory on the path of exponential increases in transistor count, has changed the emphasis of its efforts toward systems and applications, and is collaborating closely with Rebooting Computing.

The good news to come out of these efforts is that there are many proposals that have the potential to dramatically improve computer energy efficiency. Now the challenge is to begin a comprehensive research effort that can advance these ideas to the point where they can be implemented or rejected. This effort will require contributions from multiple institutions and disciplines, and will almost certainly involve international cooperation and competition as well. These issues have been recognized recently by the US White House Office of Science and Technology Policy, which announced a new Nanotechnology-Inspired Grand Challenge for Future Computing on October 20, 2015.

Of the many proposals for Future Computing, perhaps the most intriguing is to look to the human brain for inspiration on how to improve energy efficiency. A major problem is that there is very little agreement about how brains actually compute, so that the term 'neuromorphic computing' has taken on a wide variety of concepts and interpretations. I acknowledge that I know very little about what brains actually do, but in looking over various ideas that have been proposed I can already see significant opportunities for improving the computers we will have in the near future. Further, I anticipate that as neuroscientists learn more about cortical architecture and neuron function, electrical engineers will be inspired to design specialized devices and circuits that are hyper-optimized to perform particular valuable computations, and that these structures will be integrated onto heterogeneous system-on-chip or system-in-package engines as accelerators and co-processors. The reason why this strategy should succeed now is two-fold: the end of transistor scaling means that general purpose processors will no longer improve faster than special purpose hardware can be designed, and the power dissipation of scaled transistors will likely be so large that it will not be possible to operate all of the circuitry on a chip simultaneously – the 'dark-silicon' problem. The materials and technologies that will provide the new electronic functionalities for these neuromorphic accelerators have already been under development for over a decade and are the basis for the non-volatile memories that promise to change computing from a processor centric to a memory driven architecture. Can this be a path toward exponential improvements by increasing the amount of computation per unit of energy expended? The answer to this question likely involves the types of computation to be performed. I will discuss the speed versus limited precision of an example of a brain-inspired accelerator – an analog dot product engine that multiplies a fixed matrix by a variable vector in a single clock cycle, which is a type of computation in memory useful for a variety of algorithms. I will outline an idealized research program to harvest and utilize discoveries in neuroscience for brain inspired computing.

I will conclude by speculating on the possible relationship between chaotic behaviour of neurons and intuition, i.e. the ability to solve a problem for which there is no pre-existing pathway connecting two apparently unrelated states in a huge mass of data.

Categories and Subject Descriptors B.0 [GENERAL]

General Terms Performance, Design

Keywords neuromorphic computing; cortical architecture; system-on-chip; dark silicon; non-volatile memory; neuromorphic co-processors and accelerators

Bio

R. Stanley Williams is a Senior Fellow at Hewlett Packard Enterprise. He has received recognition for business, scientific and academic achievement, including being named one of the top 10 visionaries in the field of electronics by EETimes, the 2014 IEEE Outstanding Engineering Manager Award, the 2009 EETimes Innovator of the Year ACE Award, the 2007 Glenn T. Seaborg Medal for contributions to Chemistry, the 50th Anniversary Laureate Lecturer on Electrical and Optical Materials for the TMS, the 2004 Herman Bloch Medal for Industrial Research, the inaugural Scientific American 50 Top Technology leaders in 2002, and the 2000 Julius Springer Award for Applied Physics.

ASPLOS '16, April 02-06, 2016, Atlanta, GA, USA
ACM 978-1-4503-4091-5/16/04.
http://dx.doi.org/10.1145/2872362.2872417