CS267 Homework 0

Omer Spillinger
omers88@eecs.berkeley.edu
Electrical Engineering & Computer Science
University of California, Berkeley

Biography

I am a first-year graduate student in the EECS department's five-year BS/MS program. My research interests include distributed systems as well as communication-avoiding algorithms. As an undergraduate, I worked with Professor Jim Demmel's BeBOP (Berkeley Benchmarking and OPtimization) group on a communication-avoiding recursive matrix multiplication algorithm. Currently, I am working with Professor Armando Fox's SEJITS (Selective, Embedded Just-in-Time Specialization) group on a framework to enable productivity-layer programmers to easily take advantage of communication-avoidance techniques for dense linear algebra operations.

This semester I would like to refine my low-level programming skills, gain a deeper understanding of the theory behind communication-avoiding algorithms, and get more exposure to the real-world applications of high performance computing.

Cognitive Computing

The human brain is a highly efficient, event driven, massively parallel distributed processor which excels at tasks such as pattern recognition. In addition, processing and memory are combined. On the other hand, von Neumann computers are clock-driven with separate processing and memory [1]. The fact that the rate at which the CPU can work far exceeds the throughput of the shared bus between the program memory and the data memory has serious performance implications. This problem is also known as the von Neumann bottleneck [2].

A joint effort by Lawrence Berkeley National Lab and IBM Research is developing *TrueNorth*, a low-power, non-von Neumann architecture based on organic brains. In order to better understand cognitive computing, the research group built Compass, a scalable simulator for *TrueNorth* cores. Compass was implemented using the OpenMP and MPI libraries. It achieves "near-perfect" weak scaling on a 16 rack IBM® Rochester Blue Gene®/Q (262144 CPUs, 256 TB memory, 66th fastest supercomputer in the Top 500). The simulation parameters were within just two orders of magnitude from the human cortex "256 million neurosynaptic cores containing 65 billion neurons and 16 trillion synapses running only 388x slower than real time with an average spiking rate of 8.1 Hz" [3].

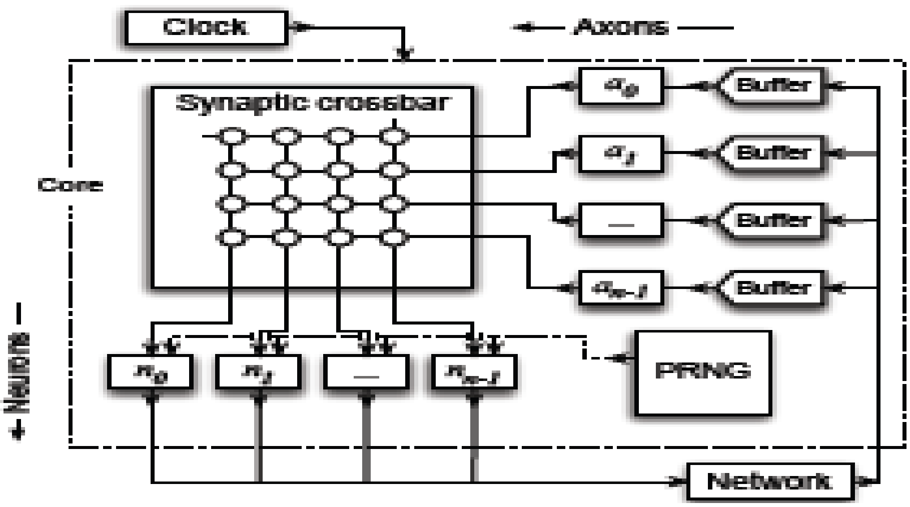

Fig. 1. Conceptual architecture of a neurosynaptic core incorporating a synaptic crossbar that contains axons (horizontal lines), dendrites (vertical lines), synapses (intersection of horizontal axons and vertical dendrites), and neurons attached to dendrites. A buffer for incoming spikes precedes each axon to account for axonal delays [3].

The *TrueNorth* architecture is based on many individual computation and memory units called neurosynaptic cores. These cores communicate over a network. In the brain, these tasks map to neurons, synapses, and axons respectively. Cores model gray matter and the communication infrastructure models white matter. The most significant challenge facing the *TrueNorth* project is that the CMOS technology developed for von Neumann computers poses limitations, such as the inability to support three-dimensional physical wiring, for architectures that behave like organic brains [3]. In order for this architecture to become practical, new fabrication technologies and a new programming paradigm are required. Cognitive computing is meant to complement, not replace, the current computing model. It will offer a low-power alternative which is not limited by clock rates and memory-processor latency.

References

[1] Kay, Roger. "Cognitive Computing: When Computers Become Brains." *Forbes*. Forbes Magazine, 09 Dec. 2011. Web. 07 Feb. 2013.

[2] "Von Neumann Architecture." *Wikipedia*. Wikimedia Foundation, 02 July 2013. Web. 07 Feb. 2013.

[3] R. Preissl, T. M. Wong, R. Appuswamy, P. Datta, M. Flickner, R. Singh, S. K. Esser, E. McQuinn, W. P. Risk, H. D. Simon, and D. S. Modha, "Compass: A scalable simulator for an architecture for Cognitive Computing," in *Proceedings of the International Conference for High Performance Computing, Networking, Storage, and Analysis* (SC 2012), Nov. 2012.