IEEE CAS Distinguished Lecture and SSCS/CAS Atlanta Joint Chapter Invited Seminar

BRAINWAY: Bio-inspired Architectures for Cognitive Computing and Applications in Health Care and Personalized Medicine

Speaker: Professor Andreas G. Andreou

Affiliation: Electrical and Computer Engineering, Center for Language and Speech Processing and Whitaker Biomedical Engineering Institute Johns Hopkins University

Abstract: Since the invention of the integrated circuit -the chip in short- in the 1950's, the microelectronics industry has seen a remarkable evolution from the centimeter scale devices created by Jack Kilby millimeter scale integrated circuits fabricated by Robert Noyce to today's 8nm feature size MOS transistors. During this time, not only have exponential improvements been made in the scaling of size and the density of devices, but CAD and workstation technologies have advanced at a similar pace enabling the design of complete truly complex Systems On a Chip (SOC). The advances in the microelectronics industry have also enabled the proliferation of computational fields for bio-informatics, systems biology imaging and multi-scale multi-domain modeling. Semiconductor technology is contributing to the advancement of biotechnology, medicine and health care delivery in ways that it was never envisioned; from scientific grade CMOS imagers to silicon photomultiplier and ion sensing arrays. The stunning convergence of semiconductor technology and life science research is transforming the landscape of the pharmaceutical, biotechnology, and healthcare industries, signaling the arrival of personalized and molecular-level imaging diagnosis and treatment therefore speeding up the pace of scientific discovery, and changing the practice and delivery of patient care. Whether through tissue and organ imaging, Labs-on-Chip or genome sequences, biotechnology and modern medical diagnostics are generating a staggering amount of data stored in data centers! However, computing in data centers, the engines behind our insatiable desire for global communication, instant connectedness and interaction comes at an economic and environmental cost. Future projected needs in data centers are data intensive applications in Cognitive Computing Technology (CCT). CCT aims at advancing intelligent software and hardware that can process, analyze, and distill knowledge from vast quantities of text, speech, images and biological data ultimately with and as much nuance and depth of understanding as a human would. To meet the scientific demand for future data-intensive CCT for every day mundane tasks such as searching via images to the uttermost serious health care disease diagnosis in personalized medicine [1], we urgently need a new cloud computing paradigm and energy efficient i.e. green technologies. The TrueNorth platform from IBM is an excellent example of such as system.

In this talk I will discuss our work in approximate computing, that is processing of data with fixed point arithmetic on parallel computer architectures for high velocity BigData applications. I will first discuss our work with the SpiNNaker system and will present some results from parallelizing and mapping sampling algorithms (neural sampling and Gibbs sampling on the SpiNNaker. I will the introduce the BRAINWAY project in my lab is aimed at the design of an energy efficient Cognitive Processor Unit (CogPU) that combines Ultra-Low-Voltage (ULV) circuit techniques with brain-inspired chip-multiprocessor network-on-chip (NoC) architecture, in 3D CMOS technology. The design of the CopPU

architecture is based on the recently developed mathematical framework for architecture exploration and optimization [2], where neurons are abstracted as arithmetic units, processing information using stochastic or deterministic unary representations [3,4]. Data in the system represent probabilities a choice that is well suited for probabilistic inference and machine learning. Such highly energy efficient CogPU inference engine will provide an energy efficiency gain of about ×65 by using ULV techniques and massive parallelism, a gain of about ×10 by relying on its SOC 3D DRAM, and a gain of about ×15 by relying on new memory based Bayesian inference computational structures. This yields an estimate aggregate improvement factor in energy efficiency of about ×10000, roughly four to five orders of magnitude with respect to present day state-of-the-art. Preliminary results from fabricated chips in the Global Foundries 55nm technology confirm our estimates and to the best of our knowledge these are the first CMOS computer architecture that computes natively with probabilities.

This work is supported by an ONR-MURI N000141010278, National Science Foundation grants NSF-SCH INT-20132017 and NSF-INSPIRE: SMA-1248056 and by DARPA-UPSIDE.

[1] A. G. Andreou, "Johns Hopkins on the chip: microsystems and cognitive machines for sustainable, affordable, personalized medicine and health care (invited paper)," IEE Electronics Letters (special supplement on semiconductors for personalized medicine), pp. s34–s37, Dec. 2011. http://digital-library.theiet.org/dbt/dbt.jsp?KEY=ELLEAK&Volume=47&Issue=26

[2] A. S. Cassidy and A. G. Andreou, "Beyond Amdahl's Law: an objective function that links multiprocessor performance gains to delay and energy," IEEE Transactions on Computers, vol. 61, no. 8, pp. 1110–1126, Aug. 2012.

[3] A. S. Cassidy, J. Georgiou, and A. G. Andreou, "Design of silicon brains in the nano-CMOS era: spiking neurons, learning synapses and neural architecture optimization," Neural Networks, pp. 1–28, Jun. 2013.

[4] D. H. Goldberg, G. Cauwenberghs, and A. G. Andreou, "Probabilistic synaptic weighting in a reconfigurable network of VLSI integrate-and-fire neurons," Neural Networks, vol. 14, pp. 781–793, 2001.

Speaker Biography (http://www.ece.jhu.edu/~andreou/andreas/**): Andreas G. Andreou** is a professor of electrical and computer engineering, computer science and the Whitaker Biomedical Engineering Institute, at Johns Hopkins University. Andreou is the co-founder of the Johns Hopkins University Center for Language and Speech Processing. Research in the Andreou lab is aimed at brain inspired microsystems for sensory information and human language processing. Notable microsystems achievements over the last 25 years, include a contrast sensitive silicon retina, the first CMOS polarization sensitive imager, silicon rods in standard foundry CMOS for single photon detection, hybrid silicon/silicone chip-scale incubator, and a large scale mixed analog/digital associative processor for character recognition. Significant algorithmic research contributions for speech recognition include the vocal tract normalization technique and heteroscedastic linear discriminant analysis, a derivation and generalization of Fisher discriminants in the maximum likelihood framework. In 1996 Andreou was elected as an IEEE Fellow, "for his contribution in energy efficient sensory Microsystems."

Seminar Time: 1:30PM-3:00PM on Sept 11th 2015 **Seminar Location:** TSRB 530, Georgia Tech.

Organizer: Dr. Hua Wang, IEEE SSCS/CASS Atlanta Joint Chapter Chair, Assistant Professor, School of ECE, Georgia Technology. Email: hua.wang@ece.gatech.edu. Phone: (404) 385-6003