Research Statement

Ajay Joshi
Boston University
Department of ECE, 8 Saint Mary's Street, Boston MA 02215
April 2, 2015

I lead the Integrated Circuits and Systems Group (http://www.bu.edu/icsg) in the ECE department at Boston University (BU). The overarching focus of my research group is to develop novel architecture and circuit techniques for designing low-power, high-performance, robust and secure computing systems using both traditional as well as emerging technologies. Currently the three main ongoing projects in my research group are (1) Design of silicon-photonic networks for intra-chip and inter-chip communication in manycore and GPU systems; (2) Development of neural network-based accelerator architectures for accelerating general-purpose as well as specialized applications; and (3) Design of optical antennas in digital ICs for detecting hardware Trojans and counterfeit ICs. A detailed description of these projects and other projects is provided later in this document.

Since starting at Boston University in September 2009, my group has published 27 refereed conference/workshop papers, 7 journal papers, 1 book chapter and 1 technical report, and filed 1 US patent. Overall, I have published 40 refereed conference/workshop papers, 10 journal papers, 1 book chapter and 2 technical reports, and filed 1 US patent. My work has been cited 791 times, and I have an h-index of 12 (as of March 22, 2015). My research group's current/past research sponsors include NSF, DARPA, NASA and IARPA, with equipment donation from Intel and Xilinx. I have received the **NSF CAREER award in 2012** and have so far secured **more than US $1.3M in external funding**. Currently, I have 5 PhD students, 3 MS/MEng students and 3 BS students in my research group. I graduated my first PhD student in Spring 2014, who now works at Freescale Semiconductor in Austin, Texas. I supervised a postdoctoral researcher from 2012 to 2014. He is now an Assistant Professor in Catholic University of Murcia (Spain). I have also had 6 MS Research advisees (now working at MIT Lincoln Labs, HP Labs, Intel, Juniper Networks, Atlas Wearables and Netronome Systems) and 2 BS Research advisees (now working at MIT Lincoln Labs and Intrepid Pursuits). I have hosted 7 high-school research interns (now undergraduate students at MIT, U C Berkeley, USC and UCSD) through the BU Research Internship in Science and Engineering (RISE) Program.

On the professional service front, I have served on the TPC of several conferences including Symposium on High-Performance Interconnects (HOTI) 2010, 2011, 2013, 2014, International Symposium on Networks-on-Chip (NOCS) 2014, 2015, International Symposium on Quality Electronic Design (ISQED) 2013, 2014, 2015, International Conference on VLSI Design 2011, 2012, 2013, 2014, 2015, and International Green Computing Conference 2014. I have also served as an external reviewer for Design Automation Conference (DAC) 2012, 2013, 2014, International Symposium on Computer Architecture (ISCA) 2015 and High Performance Computer Architecture (HPCA) 2014. I have served as a reviewer for IEEE Transactions on VLSI Systems, IEEE Transactions on Circuits and Systems, IEEE Journal of Solid-State Circuits, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, IEEE Journal for Emerging and Selected Topics in Circuits and Systems, ACM Transactions on Architecture and Code Optimization and IEEE Computer Architecture Letters. I have served as an Associate Editor for the Journal of Circuits, Systems and Computers in 2012 and 2013. I was also invited to serve on review panels by NSF, ARO and ETH Zurich Research Commission. I restarted the Boston area ARChitecture (BARC) workshop in 2015. BARC is attended by computer architecture experts (from both academia and industry) in the Greater Boston area and is composed of numerous informal oral and poster presentations and invited talks. The informal setting of BARC enables students to present their research ideas and receive valuable feedback on preliminary studies. BARC 2015 was attended by over 90 researchers from the Greater Boston area.

Since starting at BU, I have worked on single-PI projects and have also been involved in multi-PI projects. All my projects are highly inter-disciplinary in nature and adopt an integrated approach to develop novel solutions to challenging problems. A detailed description of the various ongoing and completed projects in my research group is given below.

Design of silicon-photonic networks for intra-chip and inter-chip communication in manycore and GPU systems *(Funded by NSF CAREER and DARPA)*

Silicon-photonic link technology has been proposed as a potential replacement to the electrical link technology for on-chip/off-chip communication due to its large bandwidth density and link-length independent data dependent energy. However, the large fixed power in the form of laser power and thermal tuning power negates these advantages making it difficult to justify the adoption of silicon-photonic link technology for communication in manycore systems. Moreover, a lack of architectures and applications that can leverage the large bandwidth offered by silicon-photonic networks further exacerbates the case. Under this project, there are two sub-projects: (a) Development of run-time techniques for power and thermal management in silicon-photonic manycore systems (b) Exploration of electro-photonic NoC architectures for 1000+ core systems and GPU systems.

The objective of the first sub-project is to investigate architecture-level runtime techniques to manage laser power and thermal tuning power of on-chip/off-chip silicon-photonic networks in manycore systems. We have designed a silicon-photonic multibus NoC architecture (between private L1 caches and distributed L2 cache banks), which uses weighted time-division multiplexing to minimize laser power at run time [1]. In addition, we have proposed to reduce the laser power dissipation at run time by dynamically activating/deactivating L2 cache banks and switching ON/OFF the corresponding silicon-photonic links in the NoC [2]. Moreover, we have also developed a methodology to explore the design space for laser source sharing (among waveguides) and placement to minimize laser power [3]. This methodology would in turn be used to develop techniques for run-time selection of laser sources to power the silicon-photonic network. For thermal management, we have developed a novel job allocation technique that minimizes the temperature gradients among the ring modulators/filters to maximize NoC bandwidth and in turn improve the application performance [4].

The second sub-project is on the exploration of novel architectures and systems that can leverage the large bandwidth that is offered by silicon-photonic link technology. Under this sub-project we have developed a manycore architecture that uses silicon-photonic links for efficient global sharing of computing resources located in a globally-shared Execution Unit Cloud (EUCloud) [5]. In addition, we have also explored the use of silicon-photonic link technology in the design of NoC in next generation GPUs [6]. In particular, we have developed a GPU-specific design of a hybrid photonic crossbar NoC that is more energy-efficient than the conventional electrical crossbar that used in GPUs today.

Under the silicon-photonic networks projects, we have also proposed a redesign of the DRAM main-memory system using a monolithically integrated silicon-photonic technology and show that our photonically interconnected DRAM (PIDRAM) can address the memory bandwidth limitation imposed by the processor and DRAM pin-bandwidth density and by total DRAM chip power, including off-chip signaling, cross-chip interconnect, and bank access energy [7]. We have also developed a cross-layer iterative approach that moves between the architectural level, the microarchitectural level, and the physical levels to design a silicon-photonic NoC that meets application requirements given the technology constraints [8].

Development of neural network-based accelerator architectures for accelerating general-purpose as well as specialized applications *(Funded by NASA, NSF CELEST and Xilinx)*

With the availability of limited power budgets and slow down in the CMOS technology scaling,

it has become imperative to explore 'More-than-Moore' approaches to sustain the historic scaling trends of performance, power and cost of computing systems. As a possible solution, computer architects have been investigating heterogeneous manycore architectures that incorporate dedicated accelerators with improved energy efficiency. The main goal of this project is to explore novel neural network-based accelerator architectures that can be used to acceleratore both general-purpose and special-purpose applications. This effort is part of a larger project whose overarching goal is to develop Programmable Smart Machines (PSMs) that are hybrid computing systems that behave as programmed but transparently learn and automatically improve their operation [9].

As part of this project we have developed a generic ISA interface between applications, machine code, and the underlying implementation of neural network accelerating hardware [10]. The interface is independent of the NN type, configuration, and implementation, and enables generic support for neural-network-based computation in both single- and multithreaded scenarios. In addition, we have also explored the use of NN-based accelerators to approximate mathematical functions in the GNU C Library (`glibc`) that commonly occur in application benchmarks [11]. For an energy-efficient design of these neural network accelerators, we have also investigated the use of memristors for designing multi-bit 1T1R Resistive memories, which could form one of the building blocks of the accelerator architecture [12]. Moving forward, we propose to design and evaluate a complete single-chip heterogeneous manycore processor architecture (containing general-purpose cores as well as accelerator cores) that would serve as underlying hardware for the envisioned Programmable Smart Machines.

Design of optical antennas in digital ICs for detecting hardware Trojans and counterfeit ICs *(Funding Proposals under review)*

With the increasingly fragmented and global nature of the supply chain and development cycle of IC chips, it is becoming easier for intruders to insert malicious circuitry in the IC chips or for vendors to sell counterfeit IC chips. The goal of this project is to explore the use of optical watermarking techniques at the hardware level that would enable us to rapidly detect the insertion of hardware Trojans in digital ICs as well as identify counterfeit ICs. In our approach, the optical watermarks take the form of advanced metal nanoantennas designed into the metal layers within the standard cells. These nanoantennas are engineered to elicit unique optical scattering signatures that are highly sensitive to the smallest details of the standard cell geometry. The combined optical response of a set of nanoantennas (that span across tens of standard cells) can then be tested post-fabrication to detect any insertion of hardware Trojans through replacement or rearrangement of the standard cells.

As part of this effort, as a first step, we have engineered the fill cells in a standard cell library to be highly reflective at near-IR wavelengths so that they can be readily observed in an optical image taken through the backside of the chip. The pattern produced by their locations produces an easily measured watermark of the circuit layout. Replacement, modification or re-arrangement of these fill cells to add a Trojan can therefore be detected through rapid post-fabrication backside imaging [13]. In addition to fill cells, we are also exploring the insertion of optical antennas (that will part of the optical watermark) in functional logic cells to make the IC chip more secure against hardware Trojans. An integral component of our approach is the backside imaging methodology. So we are developing a high-throughput near-IR multi-spectral imaging approach to image at sub-100 nm dimensions for rapid detection of hardware Trojans [14]. Moving forward, we will also be exploring the design of an optical PUF for detecting counterfeit IC chips.

Other projects *(Funded by Dean's Catalyst Award, IARPA, NSF)*

In addition to above described ongoing projects, my group has also worked on several other projects. Taking inspiration from communication theory, for sub-threshold digital logic we have developed a tunable adaptive feedback equalizer circuit that adjusts the switching thresholds of

the gate before the flip flop based on the gate's output in the previous cycle. This enables a faster switching of the gate outputs, which mitigates timing errors. This mitigation of timing errors provides an opportunity to reduce the dominant leakage energy component in the sub-threshold regime and in turn improve the energy efficiency of the digital logic circuit and the system as a whole [15, 16]. We have also explored the use of these equalization techniques in the design of logical circuits operating above threshold. In particular, we have explored the use of equalization techniques in the design of static complementary CMOS logic [17] and pass transistor logic [18].

On the electrical NoC design front we have proposed a novel, low-cost method to reconfigure the network channel width at run time to maximize energy efficiency of applications [19]. Here, we analyzed the effect of channel width choices for two commonly used cache hierarchies, private and distributed caches, on manycore systems with an underlying bus or crossbar architecture running parallel workloads. In addition, to address the performance limitations imposed by the use of snoop-based cache coherence protocol on switched network-on-chip (NoC), we have proposed a new network flow control technique, Express Virtual Channel with Taps (EVC-T), for transmitting both broadcast packets and data packets [20]. In addition, we have also proposed a low-latency broadcast packet notification tree network that maintains the order of broadcast packets on an unordered NoC.

To detect and correct errors in the multi-level cell (MLC) NAND Flash memories, we have proposed two general constructions of nonlinear multi-error correcting codes based on concatenations or generalized from Vasil'ev codes [21]. The proposed constructions can generate nonlinear bit-error correcting or digit-error correcting codes with very few or even no errors undetected or miscorrected for all codewords. We have also proposed a novel error detection technique based on the random selection of linear arithmetic codes and explore dthe use of this technique for the protection of the multiplier, which is a basic block in many public-key cryptographic devices [22]. The error detection technique does not imply any limitations on the types of errors at the output of the device, e.g., the multiplicity of the error does not have to be small.

Our group has also been involved developing optical techniques for detecting faults in advanced CMOS technology nodes. The decrease in the dimensions of integrated circuits and the increase in component density with technology scaling has introduced resolution challenges for optical failure analysis techniques. Using laser voltage imaging as our technique for the optical failure analysis technique, we have shows how an overcomplete dictionary-based sparse representation can improve resolution and localization accuracy [23].

Future plans

Moving forward, in addition to these existing projects, my research group will be involved in the following three projects: 1) Exploration of cross-layer optimization techniques for designing energy-efficient silicon-photonic networks for on-chip/off-chip communication – The goal of the cross-layer optimization project is to develop a novel cross-layer design automation flow that connects photonic device design, physical design, network architecture and overlying applications for design-time and run-time optimization silicon-photonic networks; 2) Designing energy-efficient digital signal processing units for small satellites – The goal of this project is to develop integrated energy-efficient single-chip solutions for the various sensing and processing operations performed by satellites; and 3) Designing energy-efficient hardware platforms for biomedical imaging – The goal of this project is to develop hardware platforms that can enable the use of novel imaging techniques like shear wave imaging and molecular imaging for biomedical applications. Though these projects span across various application areas, the common theme across all projects is the development of novel architecture and circuit techniques for designing low-power, high-performance, robust and secure computing systems.

References

[1] C. Chen and A. Joshi, "Runtime management of laser power in silicon-photonic multibus noc architecture," *Selected Topics in Quantum Electronics, IEEE Journal of*, vol. 19, pp. 3700713–3700713, March 2013.

[2] C. Chen, J. Abellan, and A. Joshi, "Managing laser power in silicon-photonic noc through cache and noc reconfiguration," *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, to appear.

[3] C. Chen, T. Zhang, P. Contu, J. Klamkin, A. Coskun, and A. Joshi, "Sharing and placement of on-chip laser sources in silicon-photonic nocs," in *Networks-on-Chip (NoCS), 2014 Eighth IEEE/ACM International Symposium on*, pp. 88–95, Sept 2014.

[4] T. Zhang, J. Abellan, A. Joshi, and A. Coskun, "Thermal management of manycore systems with silicon-photonic networks," in *Design, Automation and Test in Europe Conference and Exhibition (DATE), 2014*, pp. 1–6, March 2014.

[5] J. Abellan, C. Chen, and A. Joshi, "Leveraging silicon-photonic links for efficient global sharing of computing resources in manycore systems," *in preparation*, 2015.

[6] A. Ziabari, J. Abellan, R. Ubal, C. Chen, A. Joshi, and D. Kaeli, "Leveraging silicon-photonic noc for designing scalable gpus," in *Proc. International Conference on Supercomputing*, 2015.

[7] S. Beamer, C. Sun, Y.-J. Kwon, A. Joshi, C. Batten, V. Stojanović, and K. Asanović, "Re-architecting dram memory systems with monolithically integrated silicon photonics," in *Proceedings of the 37th Annual International Symposium on Computer Architecture*, ISCA '10, (New York, NY, USA), pp. 129–140, ACM, 2010.

[8] C. Batten, A. Joshi, V. Stojanovic, and K. Asanovic, "Designing chip-level nanophotonic interconnection networks," *Emerging and Selected Topics in Circuits and Systems, IEEE Journal on*, vol. 2, pp. 137–153, June 2012.

[9] J. Appavoo, A. Waterland, K. Zhao, S. Eldridge, A. Joshi, M. Seltzer, and S. Homer, "Programmable smart machines: A hybrid neuromorphic approach to general purpose computation," in *Proc. Neuromorphic Architectures (NeuroArch) Workshop at 41th International Symposium on Computer Architecture (ISCA-41)*, 2014.

[10] S. Eldridge, A. Waterland, Z. Kaufman, M. Seltzer, J. Appavoo, and A. Joshi, "Dana: An architecture for the integration of neural network acceleration into general purpose isas," in *under review*, 2015.

[11] S. Eldridge, F. Raudies, D. Zou, and A. Joshi, "Neural network-based accelerators for transcendental function approximation," in *Proceedings of the 24th Edition of the Great Lakes Symposium on VLSI*, GLSVLSI '14, (New York, NY, USA), pp. 169–174, ACM, 2014.

[12] M. Zangeneh and A. Joshi, "Design and optimization of nonvolatile multibit 1T1R resistive RAM," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 22, pp. 1815–1828, Aug 2014.

[13] B. Zhou, R. Adato, M. Zangeneh, T. Yang, A. Uyar, B. Goldberg, M. S. Unlu, and A. Joshi, "Detecting hardware trojans using backside optical imaging of embedded watermarks," in *Design Automation Conference (DAC) 2015*, June 2015.

[14] R. Adato, A. Uyar, M. Zangeneh, B. Zhou, A. Joshi, B. Goldberg, and M. S. Unlu, "High-throughput gate-level mapping of integrated circuits via multi-spectral near-infrared microscopy," in *in preparation*, 2015.

[15] M. Zangeneh and A. Joshi, "Sub-threshold logic circuit design using feedback equalization," in *Design, Automation and Test in Europe Conference and Exhibition (DATE), 2014*, pp. 1–6, March 2014.

[16] M. Zangeneh and A. Joshi, "Designing tunable sub-threshold logic circuits using adaptive feedback equalization," *to appear in Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*.

[17] Z. Takhirov, B. Nazer, and A. Joshi, "Error mitigation in digital logic using a feedback equalization with schmitt trigger (fest) circuit," in *Quality Electronic Design (ISQED), 2012 13th International Symposium on*, pp. 312–319, March 2012.

[18] Z. Takhirov, B. Nazer, and A. Joshi, "Energy-efficient pass-transistor-logic using decision feedback equalization," in *Low Power Electronics and Design (ISLPED), 2013 IEEE International Symposium on*, pp. 335–340, Sept 2013.

[19] J. Meng, C. Chen, A. K. Coskun, and A. Joshi, "Run-time energy management of manycore systems through reconfigurable interconnects," in *Proceedings of the 21st Edition of the Great Lakes Symposium on Great Lakes Symposium on VLSI*, GLSVLSI '11, (New York, NY, USA), pp. 43–48, ACM, 2011.

[20] C. Chen, J. Meng, A. Coskun, and A. Joshi, "Express virtual channels with taps (evc-t): A flow control technique for network-on-chip (noc) in manycore systems," in *High Performance Interconnects (HOTI), 2011 IEEE 19th Annual Symposium on*, pp. 1–10, Aug 2011.

[21] Z. Wang, M. Karpovsky, and A. Joshi, "Nonlinear multi-error correction codes for reliable mlc nand flash memories," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 20, pp. 1221–1234, July 2012.

[22] Z. Wang, M. Karpovsky, and A. Joshi, "Secure multipliers resilient to strong fault-injection attacks using multilinear arithmetic codes," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 20, pp. 1036–1048, June 2012.

[23] T. B. Cilingiroglu, M. Zangeneh, A. Uyar, W. C. Karl, J. Konrad, A. Joshi, B. Goldberg, and M. S. Unlu, "Dictionary-based sparse representation for resolution improvement in laser voltage imaging of cmos integrated circuits," in *Proc. Design, Automation and Test in Europe (DATE)*, 2015.