Guest Editorial: Special Issue on Accelerated Computing

Aviral Shrivastava, Senior Member, IEEE and Fadi J. Kurdahi, Fellow, IEEE

Accelerated Computing refers to a computing model wherein some/all of the computation of an application is carried out on specialized hardware (known as an accelerator) in tandem with the traditional CPU. Accelerators are highly specialized hardware components that can execute a specific functionality at high performance and lower power, and often, even higher reliability than is possible on a traditional CPU. The demand of ever more computation and ever-higher power-efficiency of computation (Watts per MFlops) plus the brakes on Dennard scaling have brought the paradigm of Accelerated Computing to the front and center of computer architecture and computing system design. The question that architects and system developers are asking is no longer about whether they need an accelerator in their design or not, but rather, “what is the right accelerator for my application?” From the computing in data centers to the computing in the eye-glasses – all require acceleration.

Accelerators come in several forms, each with its own set of trade-offs. The main dimensions in which the accelerators operate are i) the potential for improvement, as in how much improvement in power and performance can be expected, ii) the scope of acceleration, as in how many and how important applications can be accelerated, and iii) ease of programmability, as in how easy/difficult is it to program for the accelerator. For example, implementing some functionality directly as an Application Specific Integrated Circuit (ASIC) will probably be the highest performance and lowest-power solution possible. However, such fixed functionality silicon accelerators suffer from limited reuse. Field Programmable Gate Arrays or FPGAs are on the other side of the spectrum. They can accelerate almost any functionality, and can even be reconfigured at runtime, but programming them remains hard, and for other than for streaming applications, they may not turn out to be power-efficient. Vector Processing Units are commonplace in contemporary processors, but it remains hard for compilers to exploit them. Consequently, they remain largely unused, except for when the application is important enough (e.g., games, high performance computing applications) that programmers can spend time vectorizing their applications to exploit the vector units. General-Purpose Graphics Processing Units or GPGPUs have become quite popular accelerators recently. Although GP-GPUs can spectacularly accelerate only data parallel applications, the main reason for their popularity is that, one: there are indeed a lot of applications that are embarrassingly parallel, including a large domain of image processing applications, and second, that easy to use programming models (e.g., CUDA and openCL) helped make it easy for programmers to start accelerating their applications.

While a lot of influential research is being done in the area of accelerated computing, there are still a lot of open issues that need further investigation to increase the effectiveness and efficiency of accelerated computing. Novel accelerator architectures are needed for emerging applications, e.g., deep neural networks, brain simulations, etc. New programming models are needed to effectively and dynamically balance the workload among the computational resources in the accelerator, and also between the accelerator and the CPU. Of course, the programming models should keep programming natural and easy; otherwise, it becomes hard to understand, debug and maintain code. Several models of communication between the CPU and accelerators have been developed. However, what is the right model for a given application must be determined. Given a computing fabric with several CPUs and several accelerators, how to specify the application, how to divide and execute the application to achieve efficient computing is still an open challenge. While covering all the aspects of accelerated computing is impossible in a single special issue, our goal is to provide a sampling of current research in this domain, and capture some of the influential research trends.

The articles in this special issue cover and highlight many important aspects of accelerator design. The first aspect that pops out is that accelerators are required or needed in so many application domains. The first article is about an accelerator for aerospace systems, the third one is to accelerate graphics on mobile platforms, the fourth is for real-time applications, fifth for security applications, and seventh for visualization. The second aspect is the diversity of the accelerator platforms. The first, fourth, fifth and seventh articles propose an FPGA-based solution, the second proposed a Resistive RAM memory-based solution, the third one is based off GPUs, and the sixth is an ASIC solution.

Advances in accelerated computing promise more computing and in a much more power-conscious manner. It will be a crucial driver in delivering on the promises at the nexus of artificial intelligence, machine learning, neural networks, cyber-physical systems, and big data solutions. All of these domains require large amounts of computing under tight power budget, which cannot be delivered without the underpinnings of accelerated computing.
Aviral Shrivastava received the bachelor's degree in computer science and engineering from the Indian Institute of Technology, Delhi, and the PhD and master's degrees in information and computer science from the University of California, Irvine. He is a 2011 NSF CAREER Award Recipient, and recipient of the 2012 Outstanding Junior Researcher in CSE at Arizona State University (ASU). His works have received the Best Student Paper Award at VLSI 2016, and a Best Paper Award nomination at DAC 2017 and ASPDAC 2008. His students have received outstanding the PhD Student Award in CSE at ASU in 2017, and Outstanding MS Student Award in CSE at ASU in 2012 and 2010, respectively. He is an associate professor in the School of Computing Informatics and Decision Systems Engineering at the Arizona State University, where he has established and heads the Compiler and Microarchitecture Labs (CML) (http://aviral.lab.asu.edu/). Prof. Shrivastava’s research lies in the broad area of “Software for Embedded and Cyber-Physical Systems.” More specifically, he is interested in topics around i) Compilers and microarchitectures for heterogeneous and many-core computing, ii) protecting computation from soft errors, and iii) Precise timing for Cyber-Physical Systems. His research is funded by NSF, DOE, NIST, and several industries including Microsoft, Raytheon Missile Systems, Intel, Nvidia, etc. He serves on the organizing and program committees of several premier embedded system conferences, including DAC, ICCAD, ISLPED, CODES+ISSS, CASES and LCTES. He is currently serving as an associate editor for the ACM Transactions Embedded Computing Systems (ACM TECS), IEEE Transactions on Computer-Aided Design (IEEE TCAD), IEEE Transactions on MultiScale Computing (IEEE TMSC), Springer International Journal on Parallel Processing (Springer IJPP), and the Springer Journal on Design Automation for Embedded Systems. He is the program chair of CODES+ISSS 2017 and 2018, one of the top conferences in embedded systems. He is a senior member of the IEEE.

Fadi J. Kurdahi (S’85-M’87-SM’03-F’05) received the BE degree in electrical engineering from the American University of Beirut in 1981 and the PhD degree from the University of Southern California in 1987. Since 1987, he has been a faculty member with the Department of Electrical Engineering and Computer Science, University of California at Irvine, where he conducts research in the areas of computer aided design, high level synthesis, and design methodology of large scale systems. He currently serves as the director of the Center for Embedded and Cyber-physical Systems, which is comprised of world-class researchers in the general area of embedded and cyber-physical systems. He is a fellow of the AAAS. He received the Best Paper Award of the IEEE Transactions on VLSI in 2002, the Best Paper Award in 2006 at ISQED, and in 2016 at ASP-DAC, and four other distinguished paper awards at DAC, EuroDAC, ASP-DAC, and ISQED. He also received the Distinguished Alumnus Award from his Alma Mater, and the American University of Beirut, in 2008. He was the program chair or the general chair of program committees for several workshops, symposia, and conferences in the areas of CAD, VLSI, and system design. He has served on numerous editorial boards. He is a fellow of the IEEE.