The pursuit of Moore’s Law is slowing and the exploration of alternative devices is underway to replace the CMOS transistor and traditional architectures at the heart of data processing. Moreover, the emergence of stringent application constraints, particularly those linked to energy consumption, require new system architectural strategies (e.g., manycore) and real-time operational adaptability approaches. Such complex systems require new and powerful design and programming methods to ensure optimal and reliable operation. Thus, this special issue aims at collating new research along all the dimensions of emerging technologies and architectures for computing in manycores.

The interest in this special issue was quite high, representing the broad interest that this special topic has within the community. Further, the range of subjects was also quite broad. To accommodate this high and broad interest we have elected to split our special issue into two issues. The first issue’s focus will be on hardware-oriented techniques. The next issue will focus on software-related techniques.

The first article in this special issue, “The Quest for Energy-Efficient IS Design in Ultra-Low-Power Clustered Many-Cores” by Igor Loi, Alessandro Capotondi, Davide Rossi, Andrea Marongiu, and Luca Benini, examines a critical constraint to energy efficiency in data-parallel, clustered many-core, IoT devices, the instruction caching hierarchy. This article examines mechanisms to share the instruction cache across multiple cores in clustered many-cores. Ultimately, the article shows that, for signal processing applications, the multi-port cache architecture can improve the performance significantly with respect to traditional private cache approaches.

With the scaling of many-core systems, a key component of their design will be the design of the interconnect between the cores. The correct design of this network-on-chip (NoC) interconnect requires successful and detailed pre-RTL design space exploration, a key component of which is detailed, and accurate, yet low-overhead simulation of realistic workloads. In the article “A Systematic and Realistic Network-on-Chip Traffic Modeling and Generation Technique for Emerging Many-Core Systems,” authors Weichen Liu, Zhe Wang, Peng Yang, Jiang Xu, Bin Li, Ravi Iyer, and Ramesh Illikkal present a systematic traffic modeling and generation methodology for efficient evaluation of NoC-based many-core systems. They then show that this traffic suite provides accurate and detailed traffic data with much less overhead than prior techniques.

The next article, “High-Precision Performance Estimation for the Design Space Exploration of Dynamic Dataflow Programs,” authored by Malgorzata Michalska, Simone Casale-Brunet, Endri Bezati, and Marco Mattavelli, explores the problem of mapping dynamic dataflow programs onto many/multi-core architectures. In particular, the problem of partitioning and scheduling processing elements of dynamic dataflow programs, together with their buffers and storage across a set of many-core processors is a challenge which requires accurate performance estimation. To this end, the article develops a set of performance estimation tools and heuristics to aid in the partitioning and scheduling of these programs across many-core machines.

Another important challenge in scaling workloads across many nodes in future many-core architectures lies in the synchronization of data movement between the nodes. In particular, synchronization via locking mechanisms often incurs significant latency, greatly lowering the scaling potential of multi-threaded applications. Prior work has shown that speculatively eliding locks can greatly improve application scalability, however at the cost of lost performance for roll back in the event the speculation is incorrect. In “On Approximate Speculative Lock Elision,” authors S. Karen Khatamifard, Ismail Aksay, and ulya R. Karpuzcu show that locks can be elided speculatively without the need for rollback if the errors induced by the lock elision failure are bounded and the result is approximated.

The integration of FPGA hardware with compute in dense many-core chips is a promising approach to achieve application specific accelerator-like performance in a general-purpose design. In the article “Scalable Light-Weight Integration of FPGA Based Accelerators with Chip Multi-Processors,” authors Zhe Lin, Sharad Sinha, Hao Liang,
Liang Feng, and Wei Zhang examine the problem of interconnecting multiple such FPGAs in with processor cores into the on-chip network. In particular, they propose distributed packet receivers and hierarchical packet senders to maintain scalability and reduce the critical path delay under a heavy task load. They further propose a technique to chain accelerators together via the on die interconnect.

With increased scaling of many-core processors, heterogeneity has become a key tool to achieve energy efficiency in application execution, as evidenced by recent big.LITTLE designs from ARM and other manufacturers. “CHOAMP: Cost Based Hardware Optimization for Asymmetric Multi-core Processors” by authors Jyothi Krishna Viswakaran Sreelatha, Shankar Balachandran, and Rupesh Nasre, examines the problem of optimizing the core-thread mapping in asymmetric many-core designs for best performance and energy-efficiency. In particular, they develop a probabilistic method which models program construct behavior under different run constraints, and uses this model to choose the cost function of choice to determine this mapping at compile time.

The final article in this issue, “Bi-Objective Cost Function for Adaptive Routing in Network-on-Chip” by authors Asma Bennessaoud Gabis, Pierre Bemol, and Marc Sevaux, further illustrates the importance of the interconnect with the continuing scaling of many-core processors. In this work, the authors develop a novel, new adaptive routing algorithm inspired by the A-star search algorithm. This approach allows the simultaneous co-optimization of both latency and throughput in their adaptive routing technique even in the presence of faults.

Research along all of the dimensions of emerging technologies and architectures for computing in manycores is of great and growing importance as semiconductors scale through the final generations of CMOS process technology. This topic area is exceptionally active and vast. Here, we present a sampling of techniques in this area with a focus on hardware oriented techniques, which we hope will foster new and innovated approaches to research in the community. We further hope the reader will return for the second half of our special issue focusing on software techniques, to be released in the next issue of the IEEE Transactions on Multi-Scale Computing Systems.

ACKNOWLEDGMENTS

This special issue resulted from the inputs and efforts of many individuals to whom we are thankful. The guest editors received several high quality articles, which were further strengthened through detailed comments and suggestions from a host of reviewers. They are grateful for their time and commitment. They thank the IEEE Transactions on Multi-Scale Computing Systems, editor-in-chief Partha Pande for his continued support throughout this process. Finally, the editorial staff of the IEEE Computer Society deserves special thanks for their wonderful job in editing and organizing this issue.

For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.

Sébastien Le Beux received the PhD degree in computer science from the University of Sciences and Technology of Lille, in 2007. He has been an associate professor with the Ecole Centrale de Lyon, University of Lyon, since 2010. Since 2017, he has also been an adjunct professor under CNRS delegation with the Ecole Polytechnique de Montreal in the Computer and Software Engineering Department. In 2013, he was invited to the University of Science and Technology Hong Kong as a visiting scholar. His research interests include design methods for emerging (nano)technologies and embedded systems, including nanophotonic interconnects and reconfigurable architectures. He has more than 70 scientific publications including journal articles, book chapters, patent, and conference papers and held pivotal positions in the organization of various international conferences. He is a senior member of the IEEE.

Paul V. Gratz (S’06–M’09–SM’17) received the BS and MS degrees in electrical engineering from the University of Florida, in 1994 and 1997, respectively, and the PhD degree in electrical and computer engineering from the University of Texas at Austin, in 2008. He is an associate professor with the Department of Electrical and Computer Engineering, Texas A&M University, currently visiting the University of Edinburgh on sabbatical. His research interests include efficient and reliable design in the context of high performance computer architecture, processor memory systems, and on-chip interconnection networks. From 1997 to 2002, he was a design engineer with Intel Corporation. His papers “Path Confidence based Lookahead Prefetching” and “B-Fetch: Branch Prediction Directed Prefetching for Chip-Multiprocessors” were nominated for best papers at MICRO ’16 and MICRO ’14, respectively. He has won three best paper awards at ASPLOS (2009) and the 2016 TRIPS Computer System Computer. In 2016, he received the “Distinguished Achievement Award in Teaching—College Level” from the Texas A&M Association of Former Students and in 2017 he received the “Excellence Award in Teaching, 2017” from the Texas A&M College of Engineering. He is a senior member of the IEEE.

Ian O’Connor (S’95–M’98–SM’07) received the PhD degree in electronics from the University of Lille, France, in 1997. He is currently a professor of heterogeneous and nanoelectronic systems design with the Ecole Centrale de Lyon, Ecully, France, head of the Heterogeneous Systems Design Group at the Lyon Institute of Nanotechnology, Lyon, France, and director of the SOC2 CNRS research network. Since 2008, he has also held a position of adjunct professor at the Ecole Polytechnique de Montréal, Montréal, QC, Canada. His research interests include novel computing architectures based on emerging technologies, associated with methods for design exploration. He has authored or co-authored more than 200 book chapters, journal publications, conference papers, and patents, has held various positions of responsibility in the organization of several international conferences, and has been package leader or scientific coordinator for several French and European projects. He also serves as an expert with the French Observatory for Micro and Nano Technologies (OMNT), IFIP (International Federation for Information Processing) WG10.5 (Design and Engineering of Electronic Systems), and ALLISTENE (Alliance for digital science and technology). He is a senior member of the IEEE.

Guest Editors

Sébastien Le Beux
Paul V. Gratz
Ian O’Connor

IEEE TRANSACTIONS ON MULTI-SCALE COMPUTING SYSTEMS, VOL. 4, NO. 2, APRIL-JUNE 2018