Special Issue on Hot Interconnects 30

Scott Levy and Whit Schonbein, *Sandia National Laboratories, Albuquerque, NM, 87185, USA*

The IEEE Hot Interconnects Symposium celebrated its 30th year in 2023 with an exceptional series of presentations from industry and academia on the design, implementation, and effective use of high-performance interconnects. A core role of the Symposium is to promote the dissemination of cutting-edge research in the field. To this end, the 2023 Symposium included eight peer-reviewed presentations. This special issue of *IEEE Micro* presents revised and expanded versions of five of the best of these contributions.

A CORE ROLE OF THE SYMPOSIUM IS TO PROMOTE THE DISSEMINATION OF CUTTING-EDGE RESEARCH IN THE FIELD.

Recent trends in networking technology emphasize opportunities for offloading work traditionally done by the host CPU into the network. For example, contemporary smart network interfaces (SmartNICs) offer general-purpose (CPUs and field-programmable gate arrays) or task-specific (encryption and decryption and compression) hardware, allowing the NIC to process data as it passes between the host and the network. Two articles in this volume explore the opportunities afforded by SmartNICs to offload compression.

In "Compression Analysis for BlueField-2/-3 DPUs: Lossy and Lossless Perspectives,"[A1] Li et al. investigate SmartNIC compression performance on lossy and lossless algorithms using real-world datasets, identifying potential opportunities for offloading these workloads from the host. Similarly, in the article "Comprex: In-Network Compression for Accelerating IoT Analytics at Scale,"[A2] Oliveira and Gavrilovska consider how SmartNICs can exploit unique properties of messages generated within the Internet of Things (IoT) to compress network traffic, reducing the load on analytics servers and improving their scalability.

0272-1732 © 2024 IEEE
Digital Object Identifier 10.1109/MM.2024.3373338
Date of current version 9 April 2024.

APPENDIX: RELATED ARTICLES

A1. Y. Li, A. Kashyap, Y. Guo, and X. Lu, "Compression analysis for BlueField-2/-3 data processing units: Lossy and lossless perspectives," *IEEE Micro*, vol. 44, no. 2, pp. 8–19, Mar./Apr. 2024, doi: 10.1109/MM.2023.3343636.

A2. R. Oliveira and A. Gavrilovska, "Comprex: In-network compression for accelerating IoT analytics at scale," *IEEE Micro*, vol. 44, no. 2, pp. 20–30, Mar./Apr. 2024, doi: 10.1109/MM.2023.3343498.

A3. L. Dai, H. Qi, W. Chen, and X. Lu, "High-speed data communication with advanced networks in large language model training," *IEEE Micro*, vol. 44, no. 2, pp. 31–40, Mar./Apr. 2024, doi: 10.1109/MM.2024.3360081.

A4. D. Abts and J. Kim, "Enabling artificial intelligence supercomputers with domain-specific networks," *IEEE Micro*, vol. 44, no. 2, pp. 41–49, Mar./Apr. 2024, doi: 10.1109/MM.2023.3330079.

A5. D. Das Sharma and S. Choudhary, "Pipelined and partitionable forward error correction and cyclic redundancy check circuitry implementation for PCI Express 6.0 and Compute Express Link 3.0," *IEEE Micro*, vol. 44, no. 2, pp. 50–59, Mar./Apr. 2024, doi: 10.1109/MM.2023.3328832.

Artificial intelligence (AI) and machine learning continue to exert significant influence on cloud and high-performance computing, with generative AI at the vanguard. To better understand the challenges to traditional networks introduced by large-language models, Dai et al. profile the communication characteristics of various models utilizing different protocols and forms of parallelism in their article, "High-Speed Data Communication With Advanced Networks in Large Language Model Training."[A3] In addition, in "Enabling Artificial Intelligence Supercomputers With Domain-Specific

Networks,"[A4] Abts and Kim argue that supercomputers dedicated to AI require domain-specific processing involving low-latency, high-throughput interconnects capable of supporting tens of thousands of endpoints.

Finally, the Compute Express Link (CXL) standard unifies the memory of disparate components in heterogeneous systems by providing a cache-coherent protocol leveraging the PCI Express (PCIe) 6.0 physical layer. In "Pipelined and Partitionable Forward Error Correction and Cyclic Redundancy Check Circuitry Implementation for PCI Express 6.0 and Compute Express Link 3.0,"[A5] Das Sharma and Choudhary propose and evaluate a partitionable and pipelined implementation of forward error correction and cyclic redundancy checks supporting the CXL 3.0 specification.

In addition to the five articles published in this special issue, the entire proceedings of the Symposium are available in IEEE *Xplore* at https://ieeexplore.ieee.org/xpl/conhome/10287280/proceeding. Video recordings of many of the sessions are also available on the Hot Interconnects YouTube channel at https://www.youtube.com/@hoti-hotinterconnectssympo5358. Additional details on this and future symposia are available at http://www.hoti.org.

SCOTT LEVY is a principal member of technical staff in the Center for Computational Research, Sandia National Laboratories, Albuquerque, NM, 87185, USA. Contact him at sllevy@sandia.gov.

WHIT SCHONBEIN is a senior member of technical staff in the Center for Computational Research, Sandia National Laboratories, Albuquerque, NM, 87185, USA. Contact him at wwschon@sandia.gov.