GUEST EDITORS' INTRODUCTION

Special Issue on Hot Interconnects

Sayan Ghosh, *Pacific Northwest National Laboratory, Richland, WA, 99354, USA*
Ryan E. Grant, *Queen's University, Kingston, ON, K7L 3N6, Canada*
Min Si, *Meta AI, Menlo Park, CA, 94025, USA*

Welcome to the *IEEE Micro* Special Issue on High Performance Interconnects (Hot Interconnects). In this issue, you will find the best articles on the hottest, most cutting-edge interconnects design occurring in industry and academia from this year's IEEE Symposium on High Performance Interconnects (Hot Interconnects). Like our sister conference, Hot Chips, Hot Interconnects is the venue of choice for revealing the latest advances in hardware/software support for modern interconnects targeting data center, cloud, and supercomputing communities. This is the second year that Hot Interconnects was held virtually due to the COVID-19 pandemic. Hot Interconnects attendance was excellent, welcoming approximately 1,000 attendees.

THIS YEAR AT HOT INTERCONNECTS, WE WERE PLEASED TO WELCOME TALKS FROM THE NETWORK LEADERS AT INTEL, NVIDIA, IBM, ARISTA, LIQID, ALGO-LOGIC, TIDALSCALE, META (FORMERLY FACEBOOK), GIGAIO, AYAR LABS, AND LENOVO.

This year at Hot Interconnects, we were pleased to welcome talks from the network leaders at Intel, NVIDIA, IBM, Arista, LIQID, Algo-Logic, Tidalscale, Meta (formerly Facebook), GigaIO, Ayar Labs, and Lenovo. Dr. Debendra Das Sharma (Intel Fellow) talked about the recent advent of compute express link (CXL), a new open standard for cache-coherent interconnect, to meet the demands of the emerging heterogeneous computing landscape. Douglas Gourley (Arista's VP) discussed mitigating the challenges of adapting cloud connectivity frameworks to on-prem environments. Andy Bechtolsheim (co-founder of Arista Networks and industry veteran) talked about the advances in on-package optics technologies for providing higher performance at a lower cost and power. Alan Benjamin (CEO of GigaIO) laid out the importance of converging compute, memory, storage, and communication I/O into a single-system cluster fabric. John W. Lockwood (CEO, Algo-Logic) and Gary Smerdon (CEO, Tidalscale) introduced new network technologies.

A major theme of the submissions at this year's Hot Interconnects was network customizations for machine learning/deep learning workloads. To that end, SmartNICs and DPUs provide interesting tradeoffs for offloading computations on network hardware. Following the theme, a panel moderated by Scott Schweitzer of Achronix had a very lively discussion on Von-Neumann versus programmable logic-based architectures of SmartNIC platforms. The experts in the discussion included leaders from three of the most powerful Von Neumann-architecture SmartNIC platforms: Marvell, NVIDIA, and Pensando, as well as all three from the programmable-logic architecture side: Achronix, Intel, and Xilinx.

Our 2021 conference featured several rigorously peer-reviewed papers, with at least four reviews per submission. These articles served as the basis for the high-quality talks at this year's Hot Interconnects. Five of the top-rated papers from Hot Interconnects were invited to prepare updated versions of their articles for inclusion in this special issue. These articles went through further rounds of peer review followed by rebuttal and revision phases.

The first article in this issue, "A Low-Latency and Low-Power Approach for Coherency and Memory Protocols on PCI Express 6.0 PHY at 64.0 GT/s With PAM-4 Signaling" by Das Sharma, describes mechanisms to use PCIe/circled R 6.0 PHY at 64.0 GT/s with PAM-4 signaling for coherency and memory protocols in heterogeneous platforms. The article also discusses new mechanisms for power savings with low latency.

The article "Accelerating Allreduce With In-Network Reduction on Intel PIUMA" by Lakhotia *et al.*

0272-1732 © 2022 IEEE
Digital Object Identifier 10.1109/MM.2022.3151214
Date of current version 28 March 2022.

introduces novel architectural features of the upcoming Intel/circledR PIUMA/circledR system that allow logical collective topologies to be directly embedded into the network and support pipelined embeddings for high-throughput computation. The authors use the popular Allreduce collective operation as a case study. The authors use the PIUMA hardware and network topology to develop a methodology that generates low-latency embeddings for offloading the Allreduce operation. The proposed in-network Allreduce exhibits less than 1-μs latency on 16,000 nodes using a simulator.

Jain *et al.* explore the performance tradeoffs for offloading deep learning (DL) training on recent Nvidia DPUs in "Optimizing Distributed DNN Training Using CPUs and BlueField-2 DPUs." Traditionally, DL phases are executed either on the central processing units (CPUs) or graphics processing units (GPUs). The authors provide multiple design strategies for offloading DL training phases on latest Nvidia Bluefield-2 data processing units (DPUs), which combine the capabilities of traditional ASIC-based network adapters with an array of ARM processors. Their experimental results show that the proposed designs can deliver up to 17.5% improvement in the overall DL training time.

The article "Polarized Routing for Large Interconnection Networks" by Camarero *et al.* addresses the topology-aware or agnostic routing problem on modern scalable network topologies. The authors introduce the polarized routing algorithm, an adaptive nonminimal hop-by-hop mechanism that can be used in most network topologies. Polarized routing follows two design criteria: source–destination route symmetry and backtracking avoidance. Their experimental evaluation demonstrates that the polarized algorithm outperforms other routings in random graph topologies and low-diameter interconnection networks.

The overall performance of machine learning/deep learning (ML/DL) workloads on contemporary dense-GPU nodes is highly dependent on intranode interconnects, such as NVLink/circledR and PCIe/circledR. Temuçin *et al.* address this problem in "Accelerating Deep Learning Using Interconnect-Aware UCX Communication for MPI Collectives" by designing an interconnect-aware multipath GPU-to-GPU communication mechanism using UCX, targeting both NVLink and PCIe-based systems. The authors propose a multipath data transfer mechanism that pipelines and stripes the message across multiple intrasocket communication channels and memory regions to maximize the bandwidth of data transfers. They also design topology-aware Allreduce algorithms and observe major performance improvements compared to Horovod/TensorFlow frameworks for a variety of DL models.

WE HOPE THAT YOU WILL ENJOY THE BEST OF THIS YEAR'S HOT INTERCONNECTS AND JOIN US FOR HOT INTERCONNECTS 29 IN AUGUST 2022 AT QCT, SAN JOSE, CA, USA.

In addition to the best of Hot Interconnects 28 articles presented in this issue, several other papers and additional content from our invited talks and keynotes are available in the IEEE Xplore Digital Library at https://ieeexplore.ieee.org. We hope that you will enjoy the best of this year's Hot Interconnects and join us for Hot Interconnects 29 in August 2022 at QCT, San Jose, CA, USA. Participating in person is the only way to access some of the best interactive portions of the conference, including our popular technical panel sessions and talks by industry leaders. In the day preceding the conference, we offer an array of technical tutorials to keep attendees up to date on the latest advances in networking and give them an opportunity to have hands-on learning experiences facilitated by leaders from industry and academia. Further details can be found at http://www.hoti.org.

SAYAN GHOSH is a Computer Scientist with the Data Science and Machine Intelligence Group (a part of the Advanced Computing, Mathematics, and Data Division), Pacific Northwest National Laboratory (PNNL), Richland, WA, USA. Contact him at sg0@pnnl.gov.

RYAN E. GRANT is an Assistant Professor at Queen's University, Kingston, ON, Canada. He is a Senior Member of IEEE. Contact him at ryan.grant@queensu.ca.

MIN SI is a Research Scientist at Meta AI, Menlo Park, CA, USA. She is a Member of IEEE. Contact her at minsi.atwork@gmail.com.