



中国科学技术大学
University of Science and Technology of China

计算机学院青年学术沙龙—学术报告

Improving HPC and AI Systems and Applications via Error-Bounded Lossy Compression



报告人：陶鼎文 助理教授（美国华盛顿州立大学）

时 间：2020年10月16日（周五）10:00

地 点：腾讯会议：692 892 886

报告摘要：

The next generation of supercomputers will be exascale systems, which are capable of at least 10¹⁸ floating-point operations per second. These systems can help scientists and researchers tackle extremely complex high-performance computing (HPC) and artificial intelligence (AI) problems. Due to the gap between ever-increasing computation power and limited storage capacity and I/O bandwidth, HPC researchers must develop smart and effective methods to efficiently manage the large amounts of data generated by those HPC and AI applications. Error-bounded lossy compression has been considered as a promising solution, because it can significantly reduce the data size while maintaining high data fidelity. However, the current HPC community is facing several severe challenges to effectively and efficiently utilize error-bounded lossy compression in practice. This talk will cover two related topics: (1) optimizing error-bounded lossy compression for heterogeneous HPC systems (especially for GPUs and FPGAs), and (2) exploring error-bounded lossy compression for diverse HPC and AI applications such as cosmological simulations and deep neural network training.

报告人简介：

Dr. Dingwen Tao is an Assistant Professor in the School of Electrical Engineering and Computer Science at Washington State University. He received his B.S. degree in Mathematics from the University of Science and Technology of China in 2013 and his Ph.D. degree in Computer Science from the University of California, Riverside in 2018. Prior to joining WSU, he worked in the University of Alabama and interned at several U.S. Department of Energy national laboratories. His research interests include high-performance computing, parallel and distributed systems, and large-scale machine learning. He has published in major top-tier HPC and big data conferences and journals, including SC, ICS, HPDC, PPOPP, PACT, IPDPS, Cluster, DAC, BigData, ICPP, TPDS, etc. He is the receipt of 2020 IEEE Computer Society TCHPC Early Career Researchers Award for Excellence in High Performance Computing, 2020 NSF CRII Award, and 2017 UCR Dissertation Year Program Award. His current research has been supported by the U.S. NSF, DOE, NOAA, Xilinx, and AMD.

主办单位：计算机科学与技术学院

欢迎全校师生参加！