



北京大学高能效计算与应用中心学术报告

Invited Talk, Center for Energy-Efficient Computing and Applications

WHITHER ADVANCED GPU RESEARCH IN HPC? WHERE WE ARE, WHERE WE'ARE GOING

Dr. Shuaiwen Leon Song

High Performance Computing group at Pacific
Northwest National Lab (PNNL)

2017年4月13日 星期四 02:00pm

理科五号楼410会议室



ABSTRACT: Future large-scale high performance supercomputer systems require exceptional performance and energy efficiency to achieve exaflops computational power and beyond. To facilitate this effort, various kinds of accelerators have been added into top HPC systems due to their superior performance and power efficiency. Among them, the most typical and widely-adopted ones are GPUs. Major vendors like NVIDIA have recently proposed architectures for exascale acceleration requirement and even super machine learning nodes for big-data analytics (e.g., DGX-1). In this talk, I carefully sample several mini research topics related to advanced GPU research that I have been working on in the past year and half. They represent a range of high-interest GPU research topics including big graph analytics, approximate computing, emerging memory techniques on GPU, and software-architecture co-design. I hope these topics will invoke interesting conversations on future GPU research and eventually help us tackle various bottlenecks of the upcoming exascale era.

BIOGRAPHY: Dr. Shuaiwen Leon Song is currently a staff scientist of High Performance Computing group at Pacific Northwest National Lab (PNNL). He graduated with a Ph.D. from Computer Science department at Virginia Tech in May 2013. Before joining PNNL HPC group, he has worked with several government and industrial labs including Center for Advanced Computing (CASC) at Lawrence Livermore National Lab (LLNL), Performance Analysis Lab (PAL) at Pacific Northwest National Lab (PNNL), and the Architecture Research Division at NEC Research American at Princeton. He was a 2011 Livermore ISCR scholar, recipient of 2011 Paul E. Torgersen Excellent research award, 2016 PNNL PCSD outstanding performance award and two DOE Lab Directed Research and Development (LDRD) Awards. He has published in the major HPC-related conferences including ASPLOS, HPCA, PACT, HPDC, ICS, SC and IPDPS, etc. His SC'15 paper has been nominated for best student paper runner-up. He serves as organizing committee or PC member for several major HPC venues including ASPLOS, SC, ICS, IPDPS and HPDC. His past and current research are funded by several major U.S. agencies. Additionally, he has on-going collaboration with major U.S. Universities and industry labs such as Intel Lab and NVIDIA research.