SCHOOL OF INFORMATICS AND COMPUTING

Computer Science Colloquium Series

Torsten Hoefler

ETH Zürich

Friday, April 4, 2014

3:00 pm

Lindley Hall, Rm. 102

Modeling Communication in Cache-Coherent SMP Systems - A Case-Study with Xeon Phi

Abstract: Most multi-core and some many-core processors implement cache coherency protocols that heavily complicate the design of optimal parallel algorithms. Communication is performed implicitly by cache line transfers between cores, complicating the understanding of performance properties. We developed an intuitive performance model for cache-coherent architectures and demonstrate its use with the currently most scalable cache-coherent many-core architecture, Intel Xeon Phi.

Using our model, we develop several optimal and optimized algorithms for complex parallel data exchanges. All algorithms that were developed with the model beat the performance of the highly-tuned vendor-specific Intel OpenMP and MPI libraries by up to a factor of 4.3. The model can be simplified to satisfy the tradeoff between complexity of algorithm design and accuracy. We expect that our model can serve as a vehicle for advanced algorithm design, similar to established network models such as LogP.

Biography: I am Assistant Professor for Computer Science at ETH Zürich where I lead the Scalable Parallel Computing Laboratory (SPCL). Before I left the USA, I was working in the Blue Waters Directorate at the University of Illinois at Urbana-Champaign. I was responsible for performance modeling and simulation of the Blue Waters Petascale computer and applications running on it. In this effort, I worked with William Kramer, Marc Snir, and Bill Gropp and external partners. I am co-chair of the collective operations working group in the MPI Forum. I am interested in Collective Communications, Process Topologies, One Sided Operations, and Hybrid Programming in MPI. I am also a member ofACM SIGHPC, ACM, and IEEE.

My research interests revolve around the central topic of "Performance-centric Software Development". In the context of High-Performance Computing (HPC), one can identify three sub-branches that I am actively working on: (1) performance modeling, simulation, and optimization of large-scale parallel applications, (2) topologies, routing, and host interfaces of large-scale networks, and (3) advanced parallel programming techniques and runtime environments. An overview graph of my past and current research activities (relative to my publication list) shows their connections: Research Overview. My Erdős number is three (via Marc Snir and Shlomo Moran) and I am an academic descendant of Hermann von Helmholtz.

SCHOOL OF INFORMATICS AND COMPUTING
INDIANA UNIVERSITY
Bloomington