Symbiotic Co-scheduling on GPU-based Heterogeneous Manycore Architectures

Ms. Hao Wu

Abstract:

GPU-based heterogeneous architectures have drawn widespread adoption because of their cost-effectiveness and power-efficiency. However, building an effective scheduler to explore the maximum power of such complex parallel architecture is still a big challenge. In this research, we propose a user-friendly software ecosystem which can automatically transform Java code to OpenCL code with a loop analyzer that performs automatic loop parallelization in a speculative manner, and a scheduling runtime that assigns workloads to CPUs and GPUs. We focus on the design of symbiotic scheduling algorithms which co-schedules workloads that have complementary resource requirements to reduce resource contention and minimize CPU-GPU communication. To determine the optimal launch order of concurrent kernels, we define a new metrics, called symbiotic factor, which is derived by identifying the shared data sets and data dependencies among GPU kernels, and their execution characteristics (e.g., access patterns of global memory) by leveraging compiler-guided runtime profiling techniques. The proposed software ecosystem will support the execution of Big Data applications with non-deterministic cross-iteration dependencies. The research challenges and the overall system design will be presented in this talk.

About the Speaker:

WU Hao is currently a full-time Ph.D. student at the Department of Computer Science, the University of Hong Kong. Her supervisor is Prof. C.L. Wang. Her research interest is in high-performance GPU computing.