

Student Research Seminar

Towards Loop Speculative Parallelization on GPGPUs

Mr ZHANG Chenggang

Date:

December 3, 2010
Friday
2:30 pm

Venue:

Room 313
Chow Yei Ching Building
The University of Hong Kong

Abstract:

The multi-core era has opened the Pandora box of parallel programming environments. The debut of general-purpose graphics processing units (GPGPUs) with hundreds of cores as auxiliary compute devices makes heterogeneous architectures such as multi-core with GPGPUs accelerators become mainstream as they have a better performance/power ratio.

Unfortunately, programming such heterogeneous architectures is a big headache for application developers. Even though GPGPUs are gradually adding more general-purpose programming features, programming them is still very hard. Programmers are required to analyze parallelism in applications and explicitly define kernels to express the parallelism. Besides, GPU APIs expose the device memory hierarchy and programmers have to explicitly manage data movements along a complex multi-level hierarchy and be accustomed to a mixed MIMD-of-SIMD execution model. This results in mind-bending codes and complicates the development of future HPC programs.

In this project, we propose a new runtime parallelization technique, which enables a multithreaded program to scale transparently on a GPGPU-based heterogeneous system. We focus on the auto-parallelization of loops with statically non-deterministic data dependencies. Two challenging issues while performing speculative loop parallelization on GPGPU will be addressed: (1) Space-efficient data dependency violation detection among loop iterations under the constraints of limited shared memory and lack of cross-block thread communication in GPGPU. (2) Effective rollback and retry mechanisms like kernel reconstruction and thread migration from GPU to CPU to make efficient use of the heterogeneous cores at runtime. In this talk, we will review related work in the above areas, discuss the technical challenges, and set out the direction for the future research.

About the Speaker:

Mr. ZHANG Chenggang is an M.Phil candidate under the supervision of Dr. Cho-Li Wang. His research interests include Parallel and Distributed Computing, Computer Architecture, Distributed Shared Memory (DSM) systems, GPGPU and Heterogeneous Many-Core Programming Model.

All are welcome!
For enquiries, please call 2859-2180 or email enquiry@cs.hku.hk
Department of Computer Science
The University of Hong Kong

