香港浸會大學理學院
HKBU Faculty of Science

for favour of posting

DEPARTMENT OF COMPUTER SCIENCE

PhD Degree Oral Presentation

PhD Candidate:	Ms Xinxin MEI
Supervisor:	Dr. Xiaowen CHU
External Examiner:	Dr. Chuan WU Dr. Koon Kau CHOI (Proxy for Dr. Bingsheng HE)
Time:	9 August 2016 (Tuesday) 10:30 am - 12:30 pm (35 mins presentation and 15 mins Q & A)
Venue:	RRS732, Sir Run Run Shaw Bldg., HSH Campus

"Energy Conservation Techniques for GPU Computing"

Abstract

The emerging general purpose graphics processing units (GPGPU) computing has tremendously speeded up a great variety of commercial and scientific applications. The GPUs have become prevalent accelerators in current high performance clusters. Though the computational capacity per Watt of the GPUs is much higher than that of the CPUs, the hybrid GPU clusters still consume enormous power. To conserve energy on this kind of clusters is of critical significance.

In this thesis, we seek energy conservative computing on the GPU accelerated servers. We introduce our studies as follows.

First, we dissect the GPU memory hierarchy due to the fact that most of the GPU applications are suffering from the GPU memory bottleneck. We find that the conventional CPU cache models cannot be applied on the modern GPU caches, and the microbenchmarks to study the conventional CPU cache become invalid for the GPU. We propose the GPU-specified microbenchmarks to examine the GPU memory structures and properties. Our benchmark results verify that the design goal of the GPU has transformed from pure computation performance to better energy efficiency.

Second, we investigate the impact of dynamic voltage and frequency scaling (DVFS), a successful energy management technique for CPUs, on the GPU platforms. Our experimental results suggest that GPU DVFS is still promising in conserving energy, but the patterns to save energy strongly differ from those of the CPU. Besides, the effect of GPU DVFS depends on the individual application characteristics.

Third, we derive the GPU DVFS power and performance models from our experimental results, based on which we find the optimal GPU voltage and frequency setting to minimize the energy consumption of a single GPU task. We then study the problem of scheduling multiple tasks on a hybrid CPU-GPU cluster to minimize the total energy consumption by GPU DVFS. We design an effective offline scheduling algorithm which can reduce the energy consumption significantly.

At last, we combine the GPU DVFS and dynamic resource sleep (DRS), another energy management technique, to further conserve the energy, for the online task scheduling on hybrid clusters. Though the idle energy consumption increases significantly compared to the offline problem, our online scheduling algorithm still achieves more than 30\% of energy conservation with appropriate runtime GPU DVFS readjustments.

***** ALL INTERESTED ARE WELCOME *****