Lehrstuhl für Rechner Architektur und Parallele Systeme
Fakultät für Informatik
Technische Universität München

Master's Thesis, Guided Research:

Power-Performance Optimization for HPC Systems with Emerging Memory Technologies

Background

High Performance Computing (HPC) systems are facing severe limitations in both power and memory bandwidth/capacity. By now, both limitations have been addressed individually: To optimize performance under a strict power constraint, Power Shifting, which allocates more power budget on the bottleneck component, and Co-Scheduling, which launches multiple jobs on one single node, are promising approaches; for memory bandwidth/capacity increase, the industry has begun to support Hybrid Memory Architecture that utilizes multiple different technologies (e.g., 3D stacked DRAM, Non-Volatile RAM) in one main memory.

Approach

In this thesis, you will look at the combination of both technology trends and develop one (or both of) the following techniques: (1) Footprint-Aware Power Shifting and/or (2) Footprint-Aware Co-Scheduling. Both ideas are based on the same observation: in spite of the system software's efforts to optimize data allocations on a hybrid memory based system, the effective memory bandwidth decreases considerably when we scale the problem size of applications (e.g., using finer-grained or larger-scaled mesh models). As a result, the performance bottleneck changes among components depending on the footprint size, the memory consumption of the executed application, which then significantly affects the power-performance optimization strategies such as Power Shifting and Co-Scheduling. You will provide a basic solution for this emerging problem and develop a framework to realize it.

Requirements

- Basic knowledge in Computer Architecture and High Performance Computing
- English communication skills
- Motivation to understand fundamental principles and provide novel approaches for the emerging problems
- Experiences in working with HPC clusters is preferable

Contact

In case of interest, please contact Eishi Arima or Carsten Trinitis at the Chair for Computer Architecture and Parallel Systems (Prof. Schulz)
Eishi Arima, Dr. , Raum 01.04.060, +49 (89) 289 - 18461, arima@cc.u-tokyo.ac.jp
Carsten Trinitis, Dr.-Ing., Raum 01.04.033, +49 (89) 289 - 18458, carsten.trinitis@tum.de

Date: 31. Oktober 2018
The results obtained within this work will be published in a scientific paper on a first-class workshop/conference, where you will be a co-author.