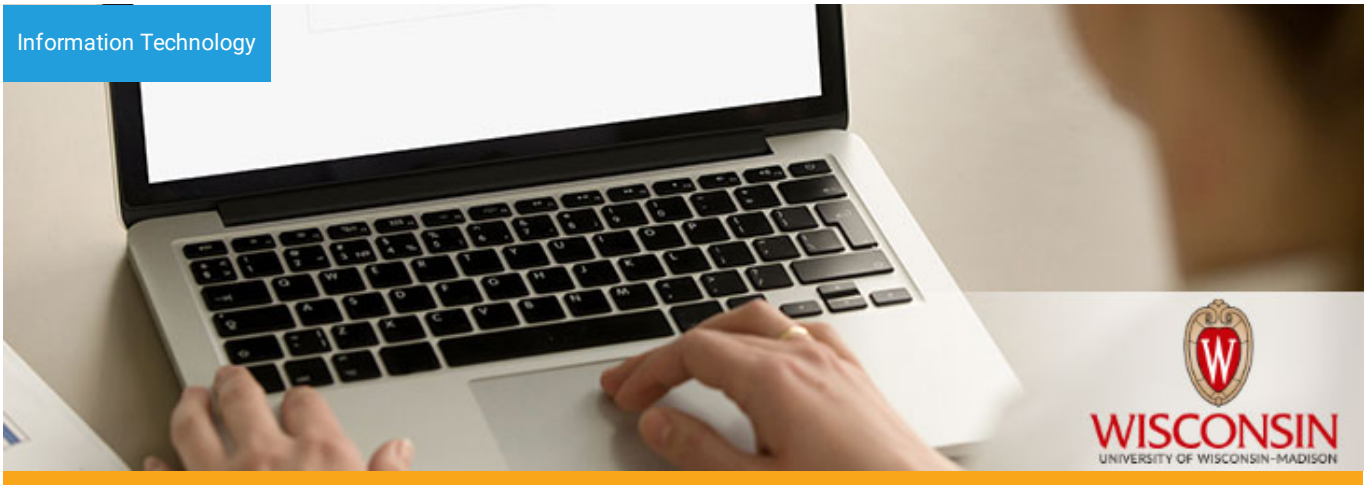
Information Technology



Memory Controller for Heterogeneous Processors

View U.S. Patent No. 9,501,227 in PDF format.

WARF: P140398US01

Inventors: Nam Sung Kim, Hao Wang

The Wisconsin Alumni Research Foundation (WARF) is seeking commercial partners interested in developing a method to improve memory access in common heterogeneous computers having a CPU and a GPU.

Overview

Many modern computers employ multiple processor types, for example, one or more conventional CPUs operating in tandem with specialized GPUs, the latter tailored to high-speed streaming processing.

The performance of these systems is limited by the speed of off-chip memory access. One method to improve speed is to prioritize access to CPU cores because they represent a smaller proportion of memory accesses and are less able to tolerate long access latency. In contrast, GPUs can offset long latency.

Deciding how to prioritize memory access between the two types of processors can be difficult because the relative importance of each task is unknown. The imperfect solution has been to rely on a 'fairness' system in which each processor is guaranteed a minimum amount of memory access.

The Invention

UW–Madison researchers have developed a memory controller providing improved access to common memory when a single parallel application is divided between different processor types, e.g., a CPU and a GPU. In these instances, fairness may not be a primary consideration and performance can be evaluated in terms of completing the entire application.

The controller works by dynamically adjusting access priorities between the different processors. It can predict sequential memory accesses by the processor having higher memory latency or fewer access requests to lockout the other processor during those sequences.

Applications

- Hardware and firmware to improve execution of single parallel programs running on heterogeneous processors

Key Benefits

- Improves memory access and performance
- Dynamic and flexible
- Can be implemented with existing hardware/minor modifications
- Accommodates changing program behavior to promote high priority memory access requests
- No undue delay on low priority access requests

Stage of Development

The new scheme achieves nearly 10 percent higher throughput than previous techniques. By employing several optimizations, an additional eight percent higher throughput can be achieved compared to FR-FCFS techniques (first-ready first-come-first-service).

Additional Information

Related Technologies

- WARF reference number P130132US01 describes a multiplier circuit that adjusts computing accuracy during run time to conserve energy.
- WARF reference number P120224US01 describes a method for improving GPU performance using memory-link compression.

Related Intellectual Property

- View Divisional Patent in PDF format.

Tech Fields

- Information Technology : Computing methods, software & machine learning
- Information Technology : Hardware

For current licensing status, please contact Jeanine Burmania at jeanine@warf.org or 608-960-9846