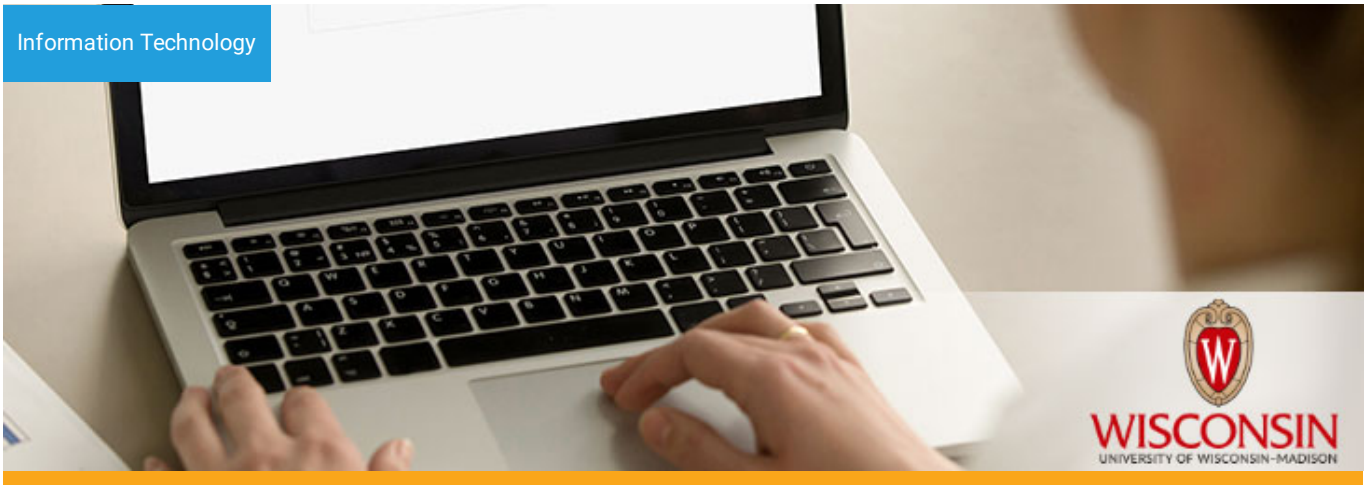
Information Technology



Cache Memory System to Reduce Invalidation Message Traffic

View U.S. Patent No. 9,223,717 in PDF format.

WARF: P130019US01

Inventors: Gurindar Sohi, Hongil Yoon

The Wisconsin Alumni Research Foundation (WARF) is seeking commercial partners interested in developing a method for cutting network traffic and energy consumption by delaying and combining invalidation messages.

Overview

Computer processors have cache memories that reduce delays spent waiting for data from the main memory. Data expected to be used by the processor in executing a program are loaded into these caches.

Cache coherence protocols are used when multiple processors and caches share access to a common main memory. The protocols coordinate changes to make data consistent across the different caches. One common cache coherence protocol is MSI, in which the data of each cache are given a status of modified, shared or invalid.

Generally, under such protocols, a change in data (for example, when one of the processors writes data to its cache) causes an invalidation message to be sent to the other caches to inform them that their corresponding data are no longer valid. However, transmitting such messages through complex multiprocessor systems consumes bandwidth and energy in the interconnection network between caches.

The Invention

UW–Madison researchers have developed a cache coherence system that streamlines multiple invalidation messages into a single transmission. The system works by delaying and collecting invalidation messages associated with common regions of memory during program executions where data-race conditions do not occur.

Specifically, the method features a cache memory holding lines of data that may be individually invalidated, as in a normal multiprocessor system with cache coherence, and a special cache controller. The cache controller delays communication of select coherence messages and then collects them into a single, combined invalidation message that can be transmitted with fewer demands placed on the interconnection network.

Applications

- Making microprocessors more efficient

Key Benefits

- Reduces number of messages that must be transmitted for cache coherence
- Reduces interconnection network traffic, energy use and heat
- Works with widely-deployed coherence protocols

Additional Information

For More Information About the Inventors

- Gurindar Sohi

Related Technologies

- WARF reference number P100343US01 describes a method of executing a program that computes data in parallel using multiple processors.

Tech Fields

- Information Technology : Computing methods, software & machine learning

For current licensing status, please contact Emily Bauer at emily@warf.org or 608-960-9842