Ph.D. Thesis

“A Study on Cache Memory Systems for On-chip Multiprocessors”

by Takuya Terasawa
Keio University, Japan March 1996. (in Japanese)
Advisor : Prof. Hideharu Amano

Abstract

A bus connected multiprocessor is one of the most promising types of small scale parallel machines because of its simple and economical structure. Usually, all processors share a common address space of the shared memory. In order to reduce the access latency and the bus congestion, each processor provides a private cache with a snoop mechanism.

The existing snoop cache protocols are optimized for the current level of technologies: the access frequency of the processor, the transfer speed of the backplane bus, access latency of the SRAM used in the cache, and the bandwidth of the DRAM used in the shared memory.

However, the future advances of the devices and implementation technologies will change the structure of bus connected multiprocessors. Some number of processors and caches will be implemented on a single chip. It is called the multiprocessor chip. Bus connected multiprocessor is suitable for such implementation because of the limitation of the number of pins in a single chip. In this implementation, the speed of the bus inside the chip is far faster than that of the backplane bus, and the large gap of the transfer bandwidth between inside and outside of the chip will become a substantial problem. On the other hand, synchronous DRAM or Rambus DRAM with quick block transfer functions will be used in the shared memory.

According to these considerations, in this paper, a new snoop cache protocol for multiprocessor chip which minimizes the accesses to the shared memory is proposed. In this protocol, both write-invalidate and write-update type protocols can be used according to the nature of the shared data. It also supports the simple synchronization mechanism with Fetch&Dec operation and inter-processor interrupt.

Detailed simulation using the multiprocessor simulator MILL with practical parallel applications shows the efficiency of the protocol.