Computer Architecture

Quiz 5 → Guide to Reading Chapter 5

April 21, 2021

Abstract

The underlying idea of Chapter 5 is multi-processors and their interaction with memory.

1 Vocabulary

We gain power over ideas which we can associate with a word. Instead of being fleeting collections of activations of neurons, alone, named ideas can be recalled, communicated, built upon.

Thus, we need vocabulary. Provide a definition for:

- cache coherence – since the view of memory held by two different processors is through their individual caches, the processors could very well end up seeing different values for the same memory location, which is what causes cache coherence.

- insufficient parallelism – a lack of parallelism which is a type of computing architecture that utilizes several processors to simultaneously execute smaller calculations and is one of the biggest performance challenges in multiprocessors. In order to combat it, new algorithms in software need to be implemented that offer better parallel performance as well as by

software systems that maximize the amount of time spent executing with the full complement of processors.

- consistency – determines when a written value will be returned by a read

- ownership of a block – denoted by an additional status bit which indicates that a block may be shared for reads, but only the owning processor can write the block, and that same processor is responsible for updating any other processors and memory when the block is changed or replaced.

- inclusion (in a cache hierarchy) – a property which requires that L3 must always have a copy of any data item in L1 or L2. In essence, in order for

inclusion to be present, all blocks in the higher-level cache must be also present in the lower-level cache, causing the lower level to be inclusive of the higher-level.

2 Relationships

Describe some relationship between the two terms (equality is possible), or explain why they are not related:

- centralized shared memory vs. symmetric shared memory vs. distributed shared memory – centralized and symmetric shared memory are related since both consist of multiprocessors sharing a single centralized memory that all processors have equal access to, but they are not related to distributed shared memory which consists of multiprocessors with physically distributed memory that must be distributed to processors rather than being centralized.

- directory-based coherence vs. snooping-based coherence – they are not related, directory-based coherence keeps the sharing status of a particular block of physical memory in one location, known as the directory. Snooping

based coherence, rather than keeping the state of sharing in a single directory, every cache that has a copy of the data from a block of physical memory could track the sharing status of the block.

- uniform memory access vs. non-uniform memory access – they are not related. Uniform memory access uses a single controller whereas non-uniform access uses a different memory controller. In addition to this, uniform memory access has a balanced memory access time whereas non-uniform memory access does not.

- per-core cache vs. shared cache – They are not related. A per-core cache is only accessible to an individual visitor, whereas a shared cache can be accessed by multiple cores, and since it is shared, each block is unique which results in a higher hit rate since there are no duplicate blocks.

- exclusive vs. owned (in a coherence protocol) – They are not related. Exclusive occurs when the status of the owner's cache block is changed from shared to unshared when an invitation is set as opposed to a block being owned which means that it is the core with the sole copy of a cache block.