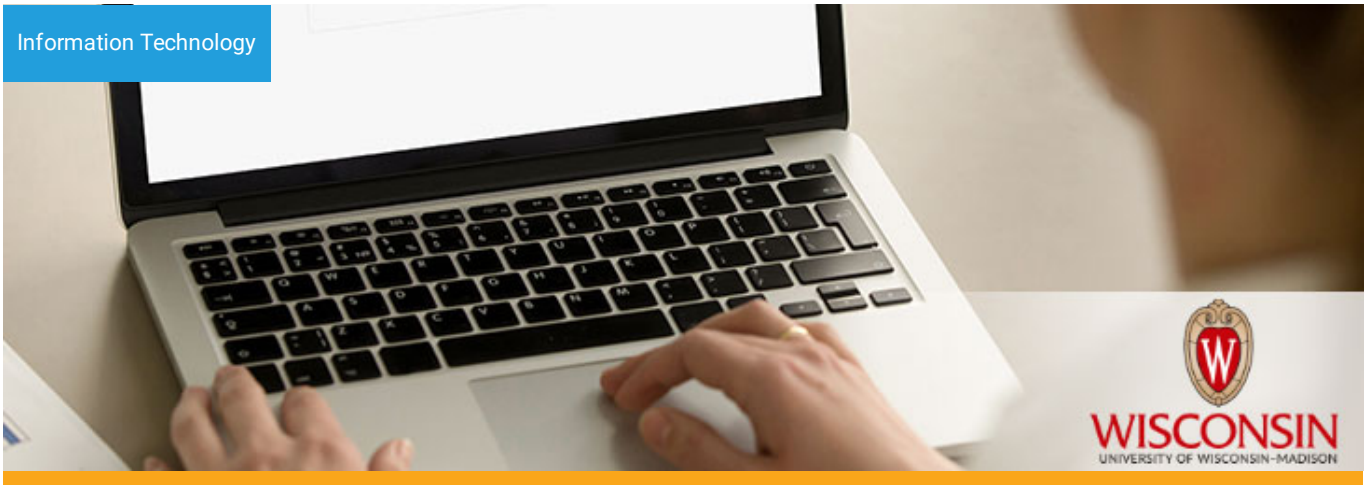
Information Technology



SuperTag Cache for Energy-Optimized Compression

View U.S. Patent No. 9,261,946 in PDF format.

WARF: P130076US01

Inventors: David Wood, Somayeh Sardashti

The Wisconsin Alumni Research Foundation (WARF) is seeking commercial partners interested in developing methods that improve compression effectiveness and reduce system energy by exploiting spatial locality of cache blocks.

Overview

Cache memories play a critical role in reducing a computer system's energy. Generally, they are fast access memories that store data reflecting locations in a corresponding main memory. Typically, the data stored in caches are organized into 'blocks.' Each block has a corresponding set of tags, which also are stored in cache. Such tags can include address tags identifying the area of main memory that maps to the corresponding block.

Although caches consume lots of power, they also can save system power by reducing costly off-chip accesses to main memory. Clearly, cutting energy usage requires making caches as effective as possible. Cache compression is one technique for increasing cache capacity and reducing misses. However, previous designs have focused on performance rather than energy efficiency.

A new compression approach could be more effective by exploiting the spatial locality of data blocks. This would recognize that several contiguous blocks often coexist in memory, having a similar compression ratio that typically is higher in larger block sizes.

The Invention

UW–Madison researchers have developed a compressed cache, called SuperTag, which exploits spatial locality to optimize compression effectiveness and energy use.

SuperTag cache manages cache at three granularities: 'super blocks,' single blocks and fractional data segments. Since contiguous blocks have the same tag address, SuperTag increases per-block tag space by tracking super blocks (for example, a group of four aligned contiguous blocks of 64 bytes each). It also breaks each cache block into smaller data segments for storage.

To improve compression ratio, the technique uses a variable-packing scheme allowing variable-size compression blocks without costly compaction. It also co-compresses contiguous blocks, including within the same super block, thereby producing data segments for storage.

Applications

- Memory system development

Key Benefits

- Improves energy and performance for memory-intensive applications
- Improves compression ratio

- Higher compressibility without high area overheads

Additional Information

Related Technologies

- WARF reference number P04273US describes a cache compression system that dynamically adapts to the costs and benefits of compression.

Tech Fields

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

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