Cache Data Compaction via Decoupled Tag Store

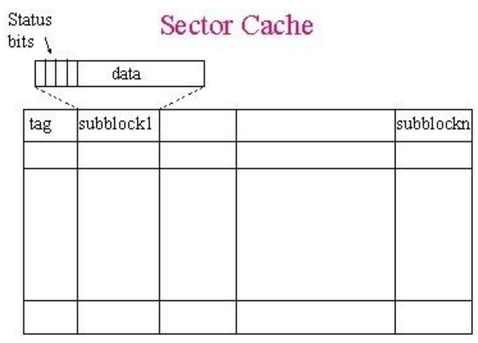
Edward Ma

15/18-740, Computer Architecture
Dec. 13, 2011

Concept and Motivation

- **Concerns**
  - Performance, latency and power are tied to memory bandwidth
  - Caches are expensive – can we reduce their size without significantly compromising performance?
  - Sectored caches provide both benefits of reduced cache block granularity and smaller tag store size by tying multiple blocks to one tag entry, thus reducing the number of entries. However, they suffer from poor performance due to unused space.

- **Data Compaction**
  - What if we increase the number of tag entries slightly and remove the restriction sector restrictions?

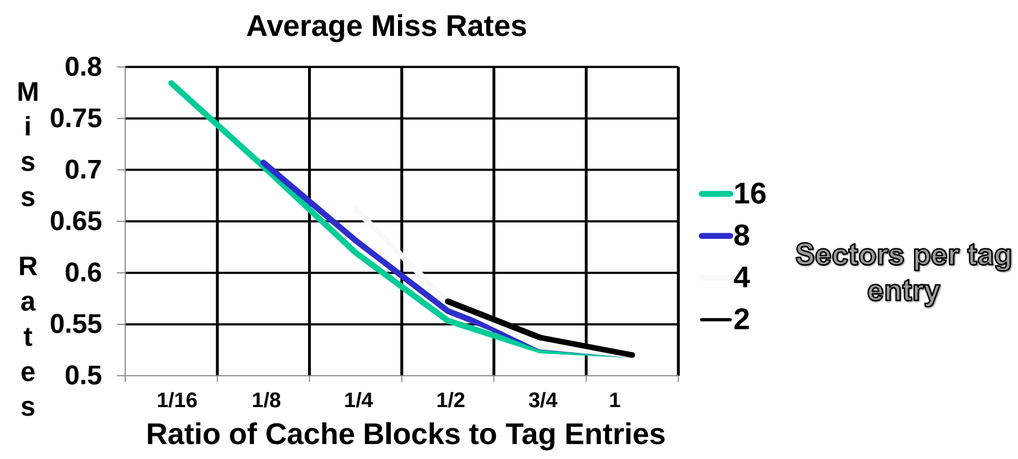


  - We can "compact" the data stored in the cache by allowing free "sectors" of the cache to be occupied by other blocks (tags).
  - Each block can have a variable size, can grow and shrink as needed.

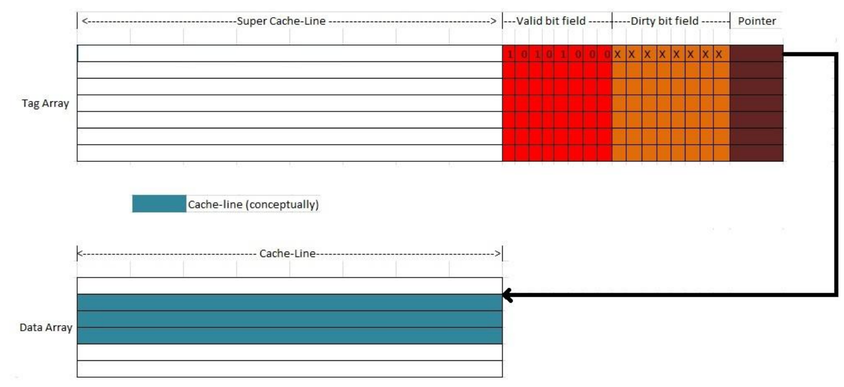
Big Questions

Just how many tag entries?

- For same number of tag entries, more sectors per tag entry is slightly better.



Where to insert a data block?

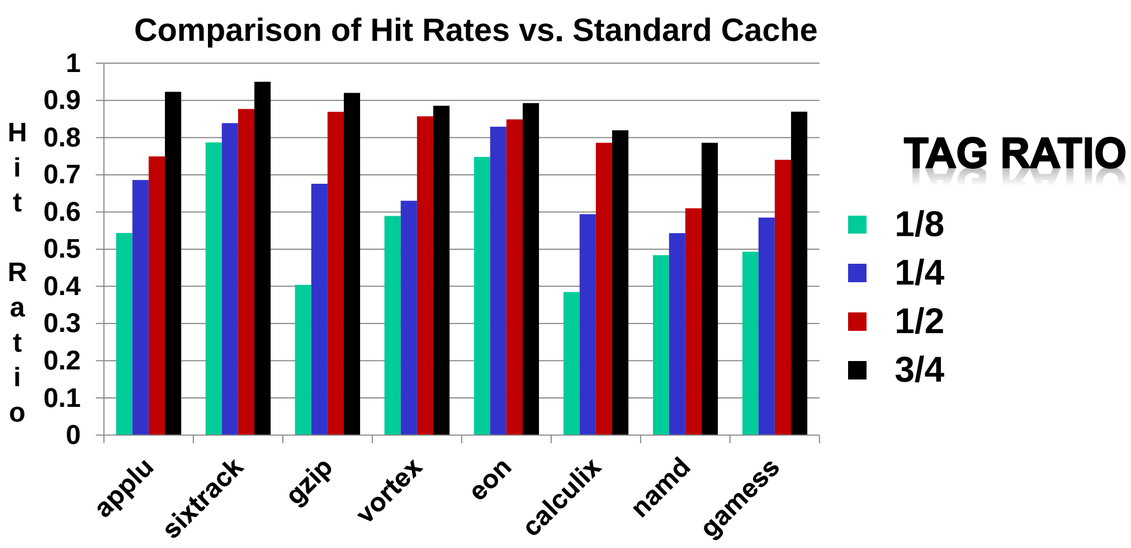


- Can't insert anywhere; reorder latency can be long
- Can evict block closest to sector "family"

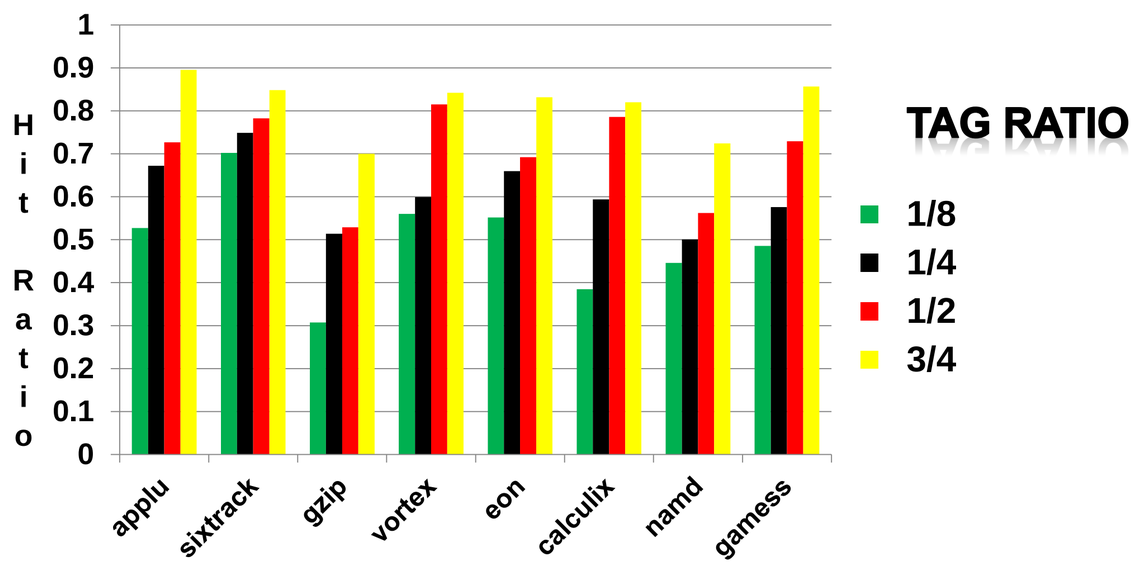
Benchmarks and Results

Savings versus Performance

- Number of Sectors per Block (nsb) = 8



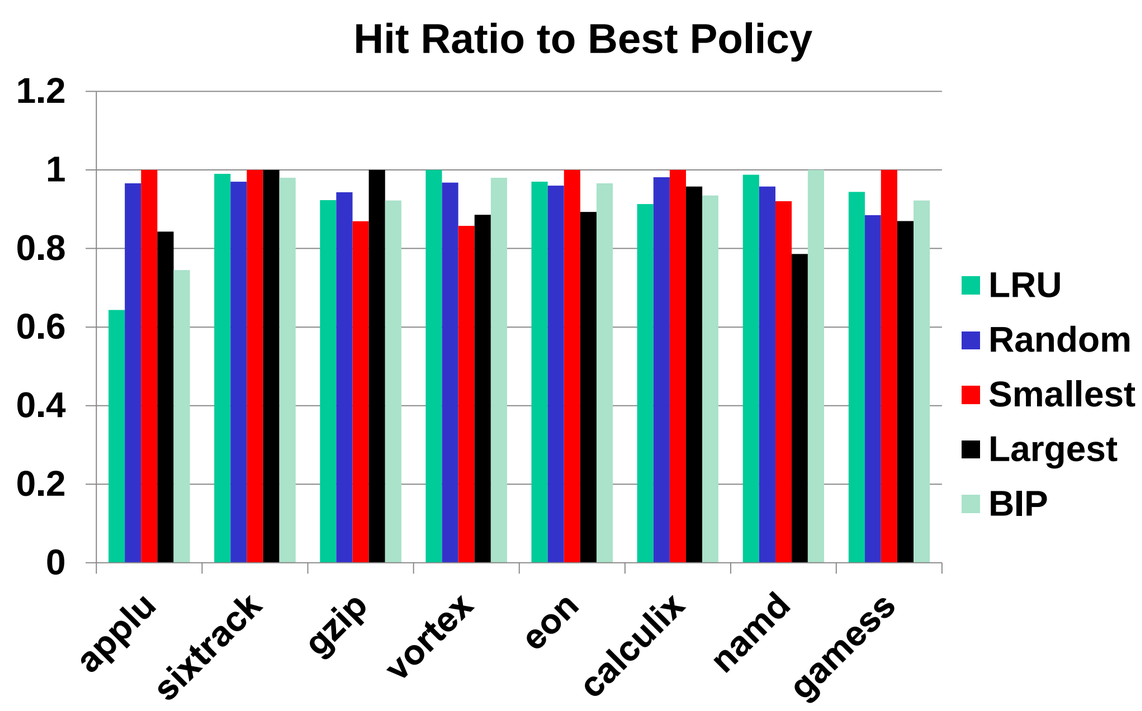
- Number of Sectors per Block (nsb) = 4
  - Marginal performance loss for more savings



For the 256kb, 8-way assoc. L2 cache above and a 4GB memory (4 sectors/block), we save 3% tag store space, pass 75% less data per cache transfer at roughly 70% of the hit rate (~75% better hit rate than a sectored cache with same nsb).

More Results & Future Work

Replacement Policy Comparison



Summary

- Cache Compacted Design Improves Hit Rate over Sectored Cache design, reduces data block granularity, reducing traffic.
- Some more improvements can be made to hit rate
  - Further exploration of replacement and insertion policies, especially in the data store.
  - Need better latency modeling for data store reorders
  - Techniques to reduce or "hide" reorder latency – offset bits in tag store or reorder amount restrictions….more tradeoffs!