The Direct-to-Data (D2D) Cache: Navigating the Cache Hierarchy with a Single Lookup

Andreas Sembrant, Erick Hagersten, David Black-Schaffer
Challenges

Traditional cache access routine is power inefficient, and time consuming because of decouple of TLB and tag.

Data request from CPU

Virtual Addr

Physical Addr

L1 TLB

L2 TLB

Mem TLB

Flag

Physical Addr

L1 Tag

L1 Data

L2 Tag

L2 Data

Main memory
Main Idea

To optimize cache hierarchy from energy and time, our goals are:

1. Skip levels in the hierarchy (by accessing the right cache level directly)
2. Eliminate extra data array read (by reading the right way directly)
3. Avoid tag comparisons (by eliminating the tag array)
4. Go directly to DRAM on cache misses (by checking the TLB)

To realize these targets, the solution proposed in this paper is Direct-to-Data (D2D) cache design. This design uses a single lookup to an extended TLB to identify the cache level and way for the desired data.
Main Idea

Direct-to-data cache design couples the function of TLB and tag by using an extended TLB.

Data request from CPU

L1 eTLB

L2 TLB

Hub

Mem TLB

Virtual Addr

Physical Addr

L1 Data

L2 Data

Main memory
In addition to virtual to physical address translation, eTLB also holds cache-line location table (CLT) for each page. The CLT provides cache line location information for all the cache lines belonging to the same page, including 2-b cache location and 4-b way index.
The hub functions as a second-level storage for cache line locations (CLT) for each page. Different from L2 TLB, a hub is indexed by physical address, which can solve synonym handling issues. Besides, Hub guarantees invariant: every cache line has only one information in Hub.
Cache line

Each cache line is extended with a Hub pointer (HP) containing the location of the page in the Hub to which the cache line belongs. The pointer is only used during replacement to identify the CLT to update.
Examples:

(a) eTLB Hit + L1D Hit
(b) eTLB Hit + L2 Hit
(c) eTLB Hit + Cache Miss

Figure 2: D2D Cache Examples: eTLB Hit
Examples:

(d) eTLB Miss + Hub Hit
(e) Hub Miss + Replacement
(f) L1 eviction + Hub/eTLB Update

Figure 3: D2D Cache Examples: eTLB Miss