CS698Y: Modern Memory Systems
Lecture-9 (Cache Hierarchy)

Biswabandan Panda

biswap@cse.iitk.ac.in

https://www.cse.iitk.ac.in/users/biswap/CS698Y.html

Logistics

Quiz 1.0: Sept. 23, 13:00 to 15:00 hrs

P.A. 1: Sept. 10, 17:00 hrs, code + report @Canvas

Presentation: Sept. 14, 10:30 hrs, (9 + 1) mins

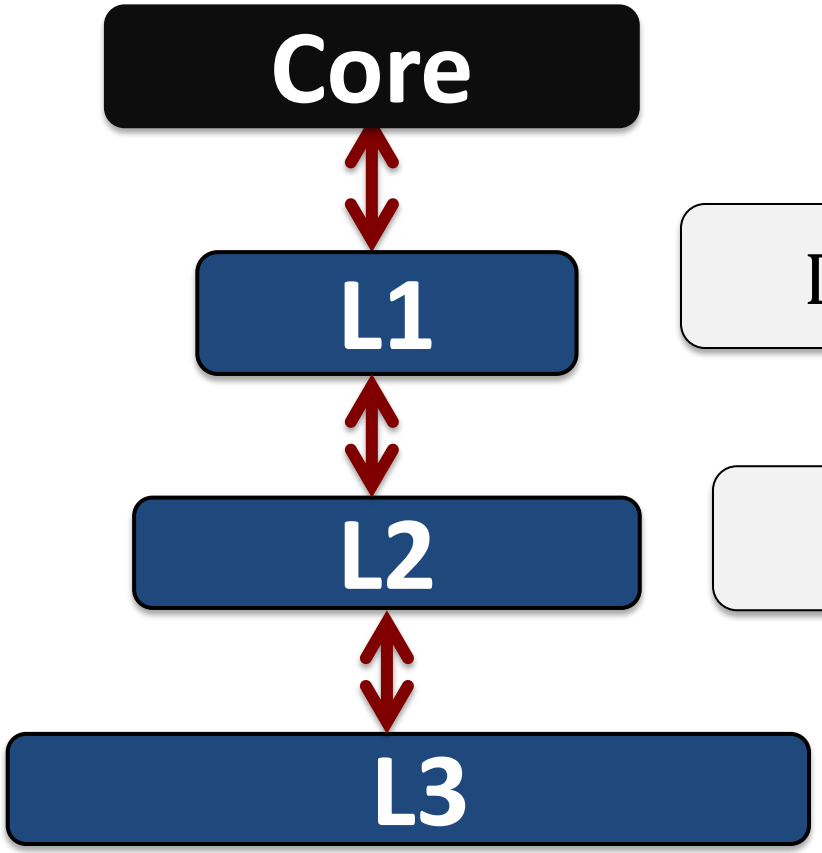
Insights: Why and how of what? For example:

What: astar, Hawkeye performs x% worse than LRU.

Why: x% of blocks have feature y, which is not captured by Hawkeye.

How: We can augment with a blackbox that does some black magic to improve Hawkeye.

Cache Hierarchy

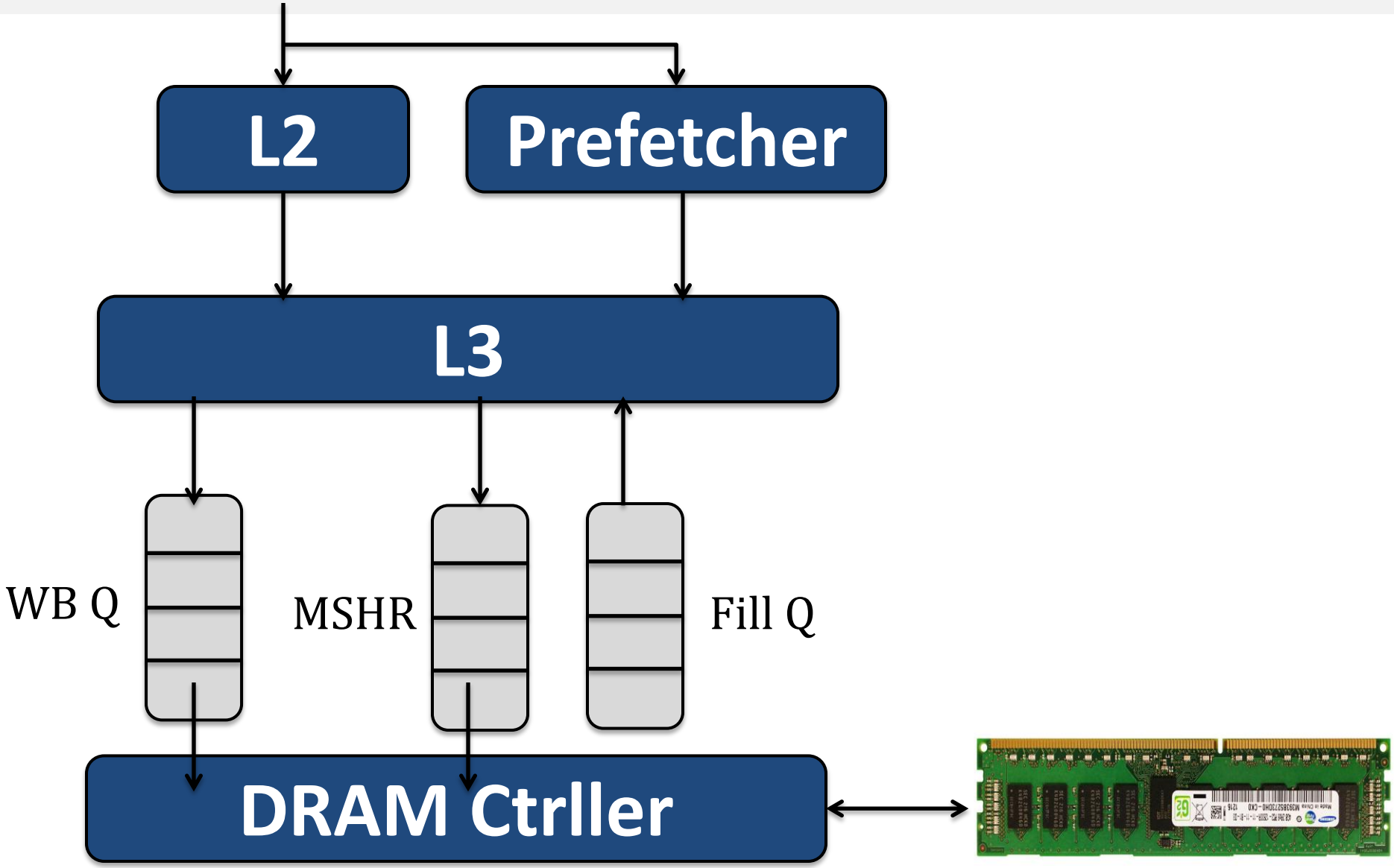


Designed for *bandwidth*

Designed for *latency*

Designed for *capacity*

Revisiting Cache Hierarchy (The Organization)

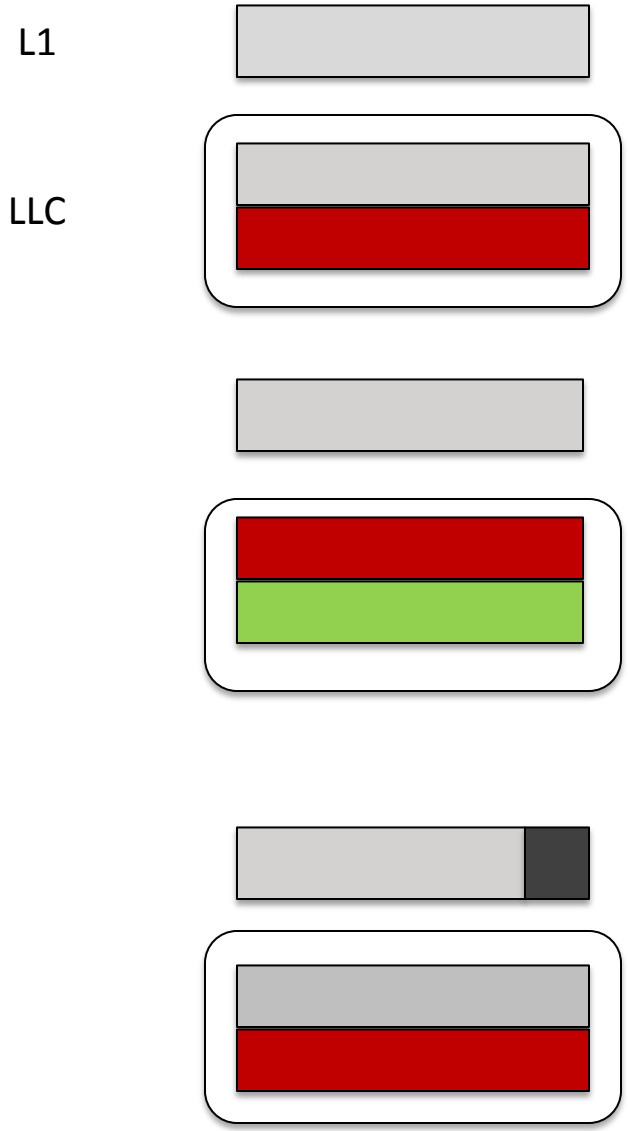


Cache Hierarchies

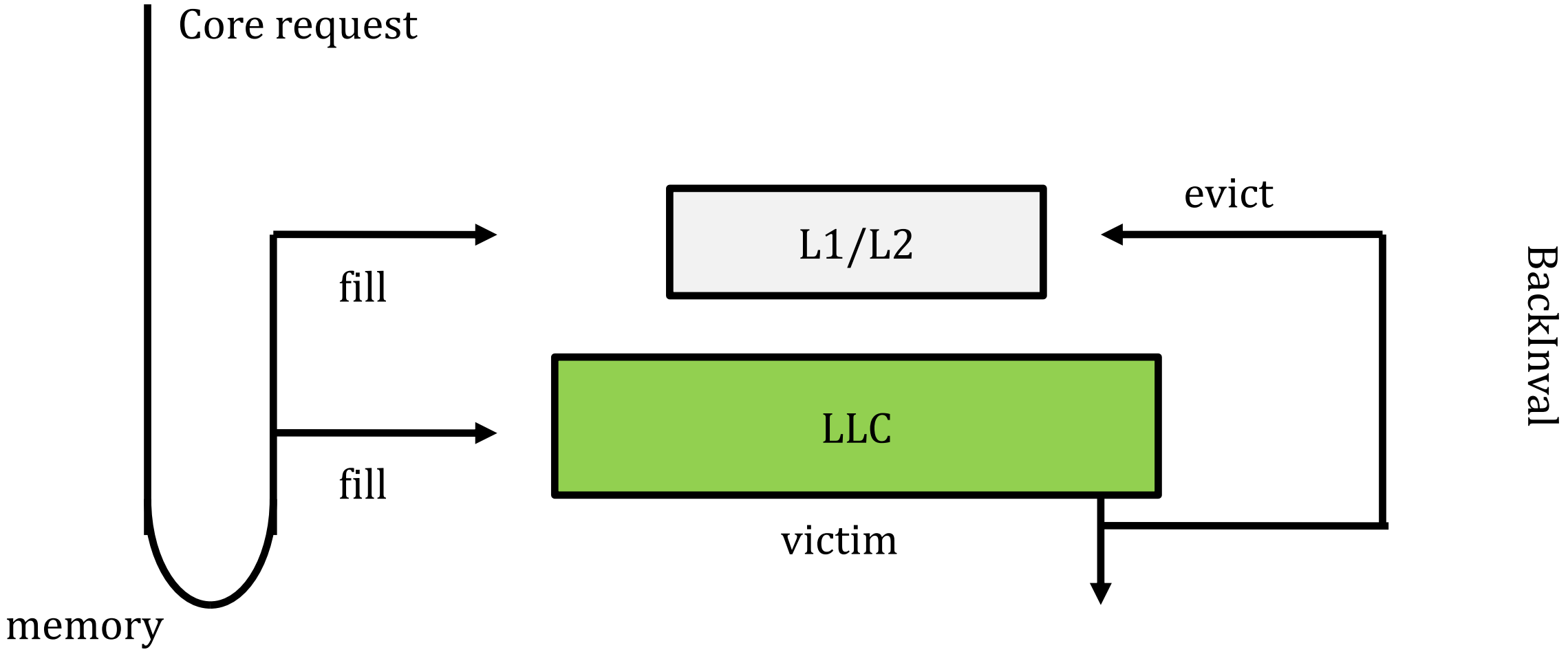
Inclusive: L1 cache blocks are present in L3

Exclusive: L1 cache blocks are absent in L3

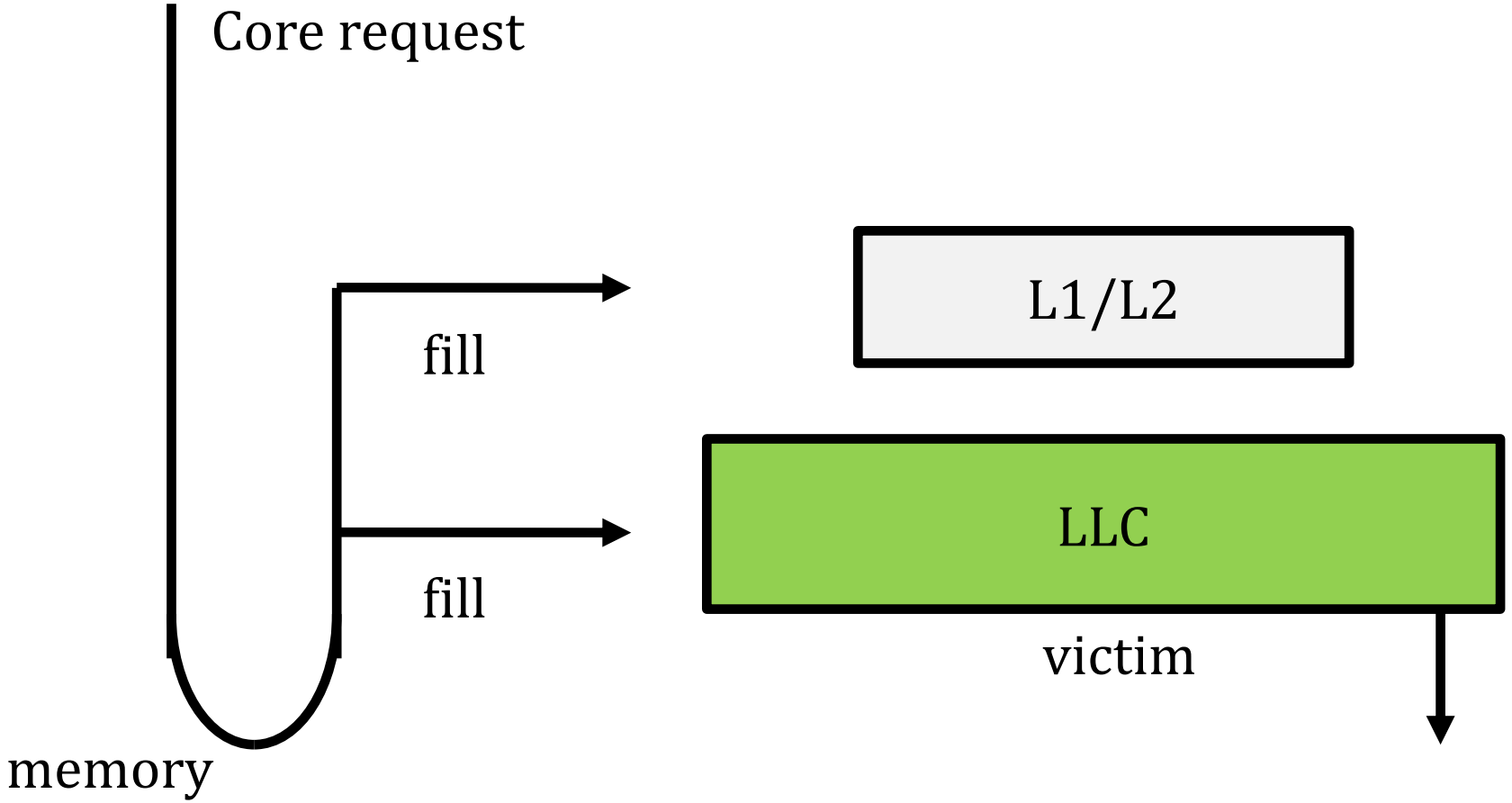
Non-inclusive: L1 cache blocks may/may not be in L3



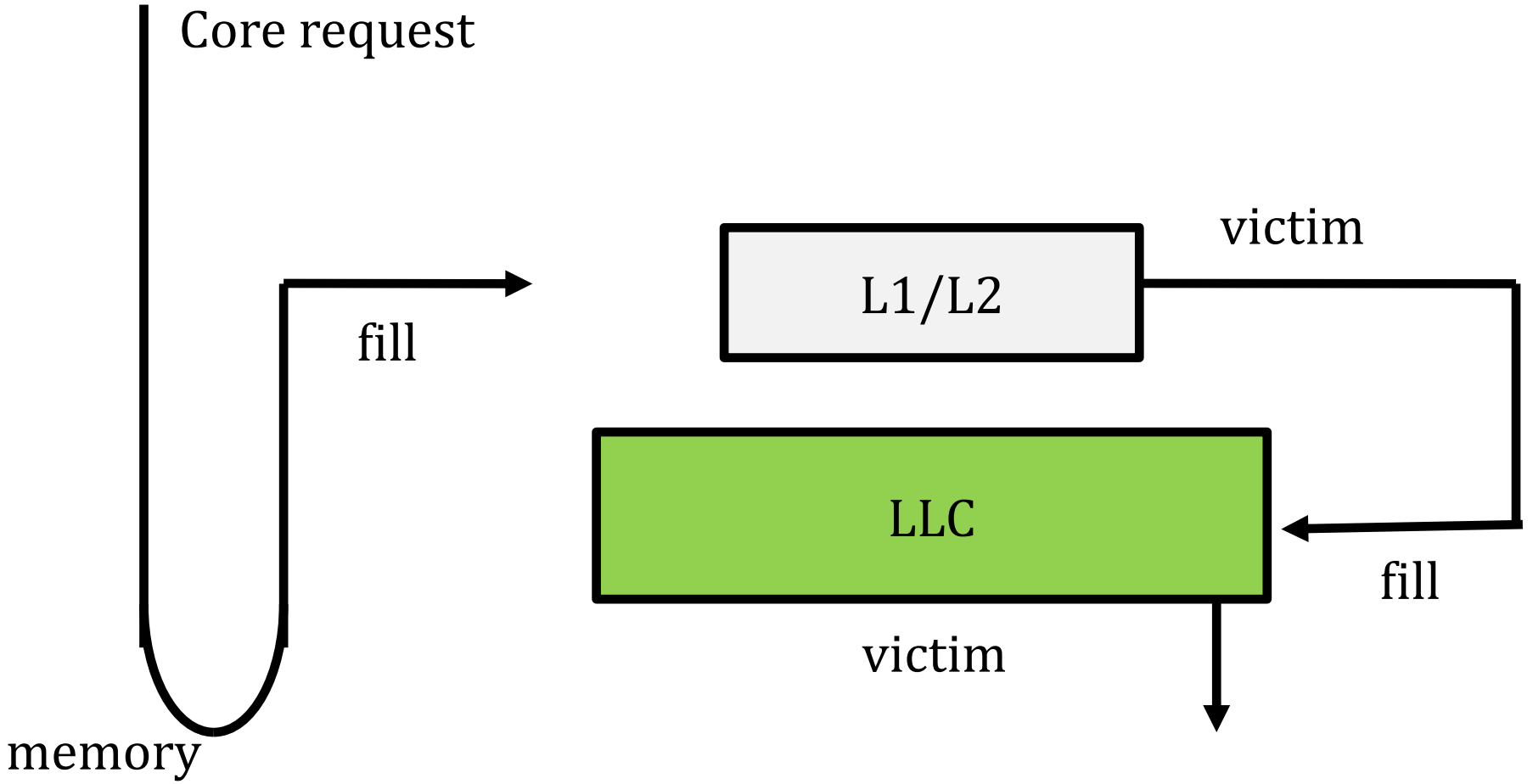
Inclusive Hierarchy



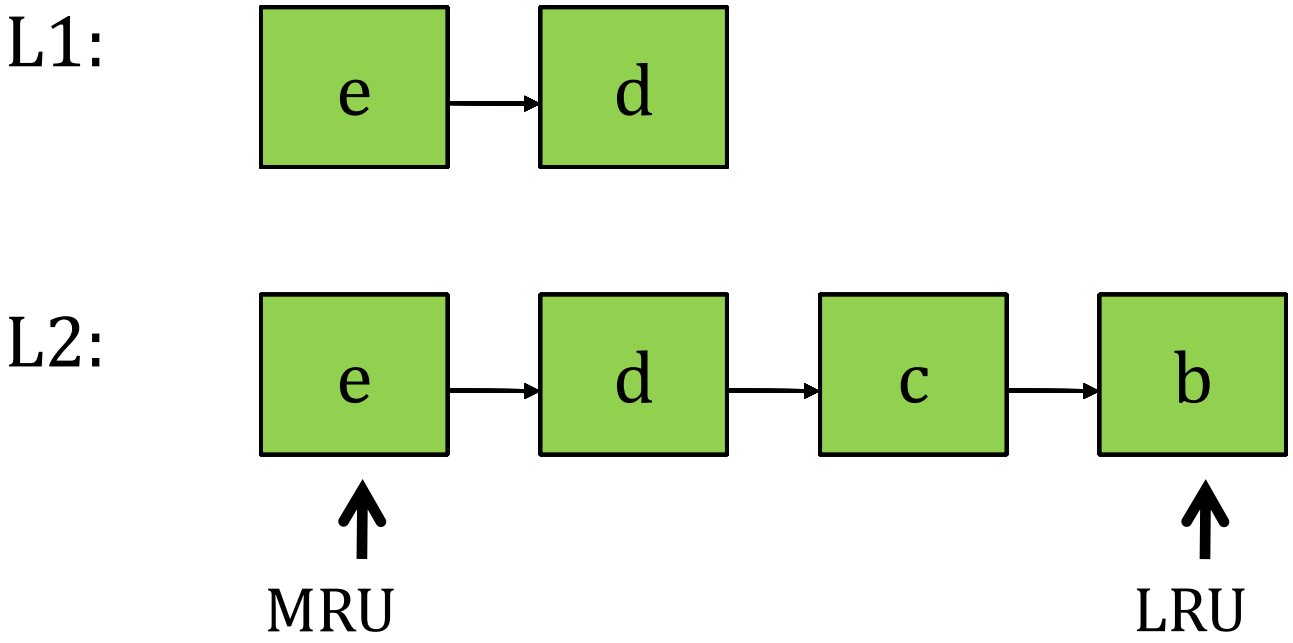
Non-inclusive



Exclusive Hierarchy



Back-invalidation [TLA, MICRO '10]



Reference 'e' misses and evicts 'a' from hierarchy

Next Reference to 'a' misses ☹