H-ARC: A Non-Volatile Memory Based Cache Policy for Solid State Drives

Ziqi Fan[1], David H.C. Du[1], and Doug Voigt[2]

[1]University of Minnesota-Twin Cities
[2]HP Storage

Agenda

- Background
- System Architecture
- Motivation
- Existing Work
- Our Approach
- Evaluation

Background: NVM

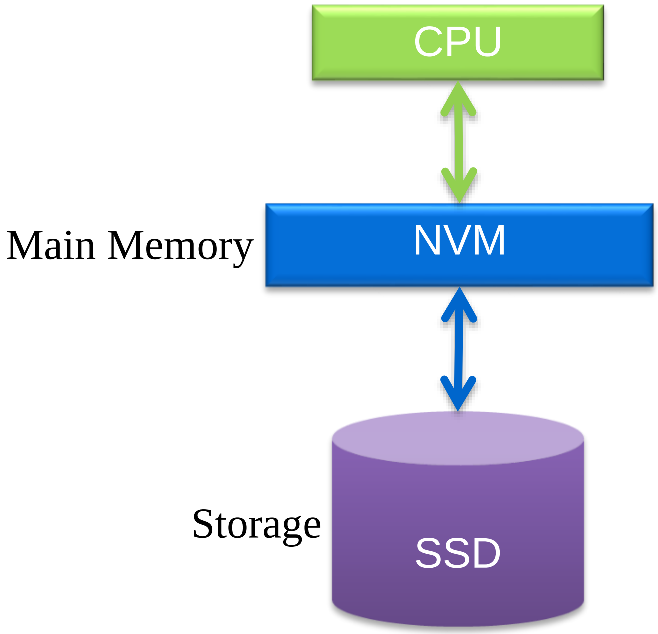
- Non-volatile memory (e.g. PCM, STT-RAM, Memristor, NV DIMMs)
  - Non-volatile (reliable)
  - Fast access speed (DRAM like)
  - High density
  - Low power consumption

Background: SSDs

- Read/write in page units
- Typical block = 128K;
- Page = 2K or 4K
- Must erase block before write
- Read = 25 microseconds
- Write = 200 microseconds
- Erase = 1500 microseconds
- Limited number of erases per block
  - 100K for SLC
  - 10K for MLC

SSD writes are expensive

System Architecture



Replace DRAM with NVM as a big buffer cache

Motivation

- Big difference
  - DRAM-based system, dirty pages must be flushed back frequently
  - NVM-based system, dirty pages can be cached much longer
- SSD writes are expensive
  - Wear out
  - Slow
- Hierarchical-ARC (H-ARC) tries to
  - Minimize write traffic from NVM to SSDs
  - Maintain hit ratio for good system performance (cache some hot clean pages)

Existing Work

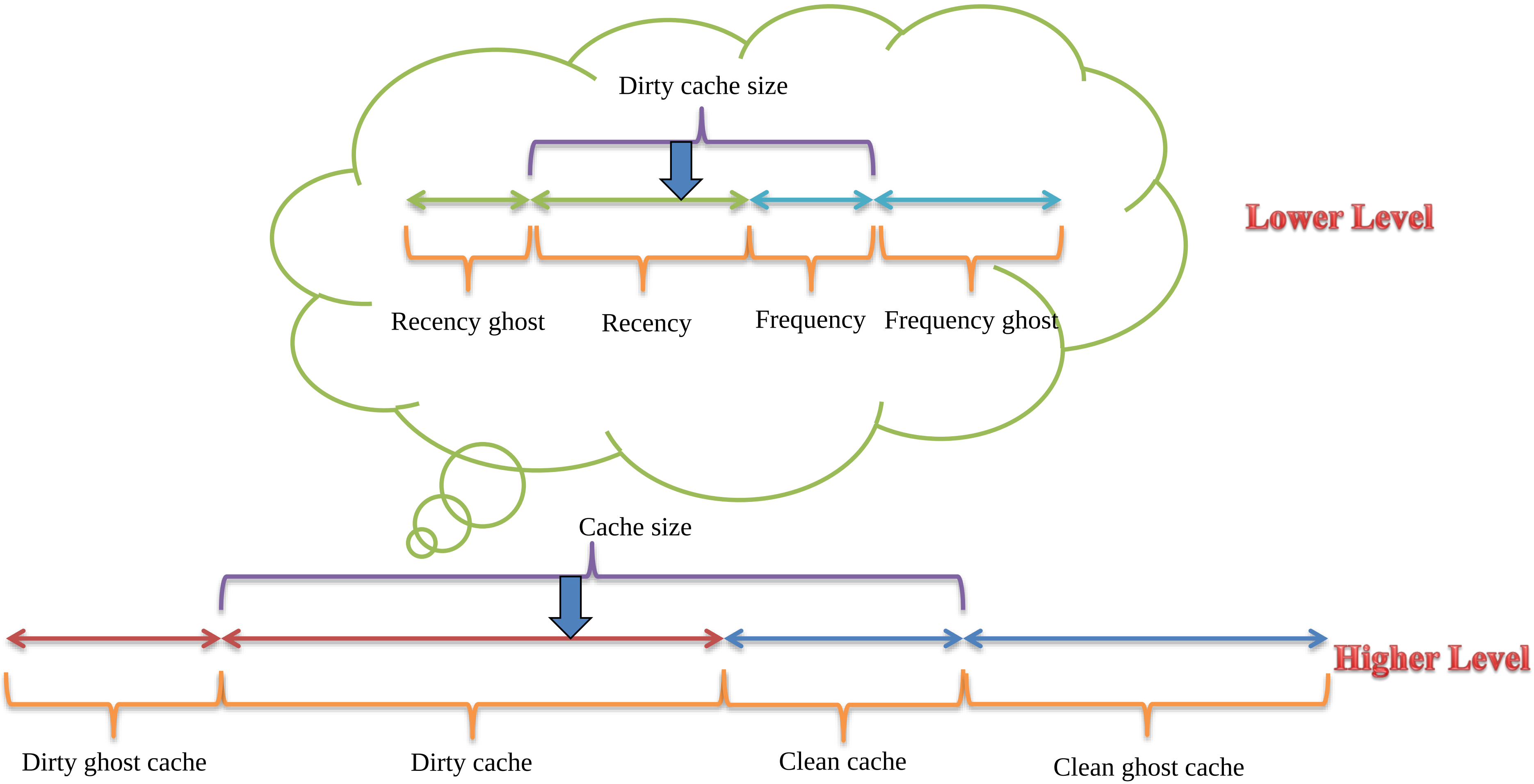
- Most DRAM-based systems mainly concentrate on improving read hit ratio
  - Recency
  - Frequency
  - ARC [1]: dynamically combine recency and frequency
- Most NVM-based systems mainly concentrate on decreasing storage write traffic
  - Dirty page
  - Clean page

[1] N. Megiddo, and D. S. Modha. ARC: a Self-tuning, Low Overhead Replacement Cache. In FAST ’03

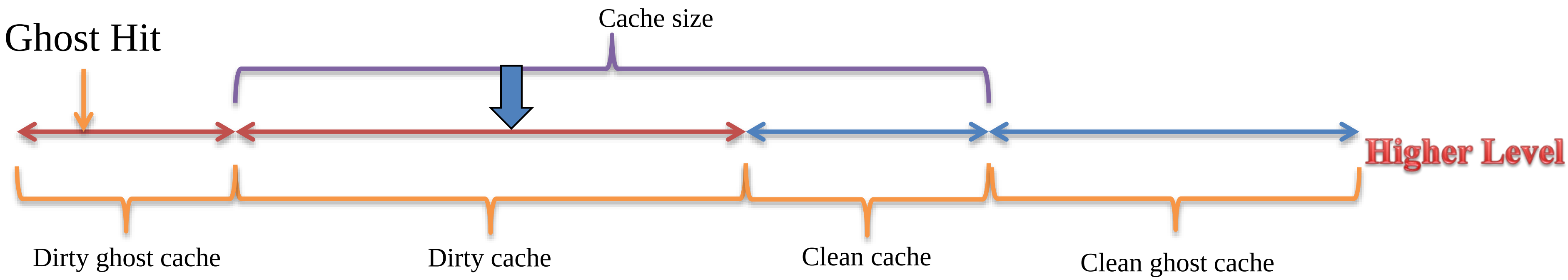
H-ARC: Overview

- Using the learning rule behind ARC
- Intelligently decide what kind of pages should be evicted based on their states, i.e. dirty, clean, accessed frequently and accessed recently
- Delay dirty page eviction by caching them longer on purpose

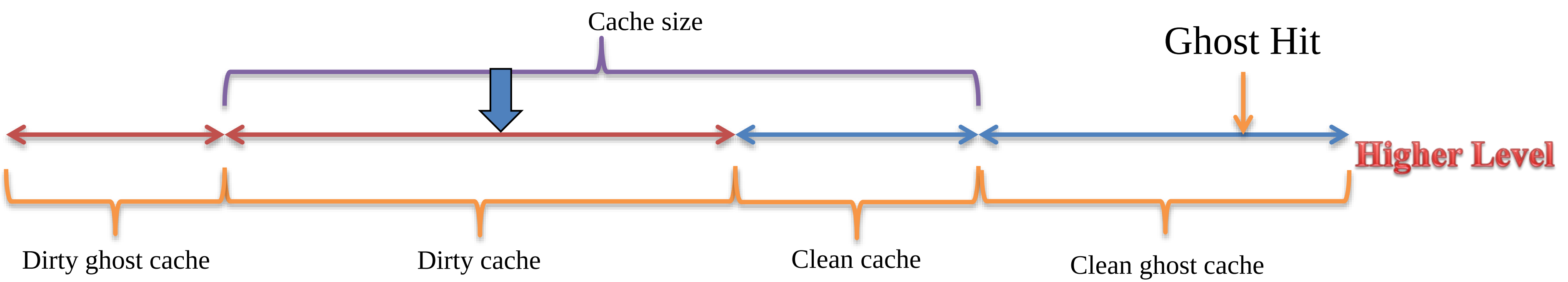
H-ARC: Overview



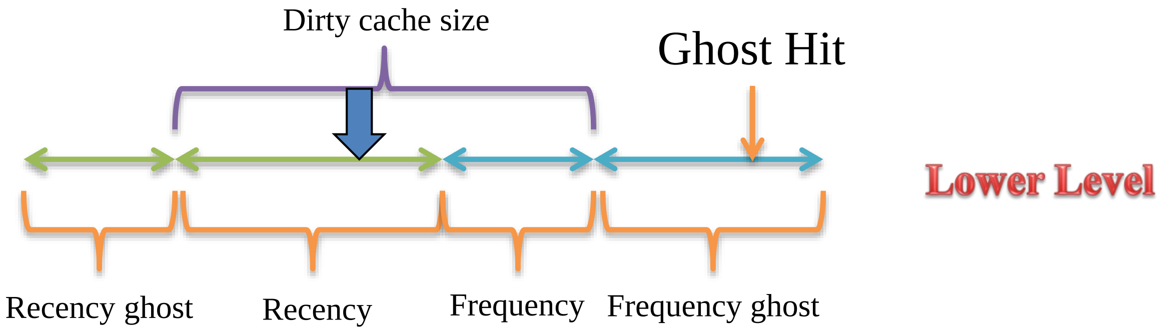
H-ARC: Learning (1/3)



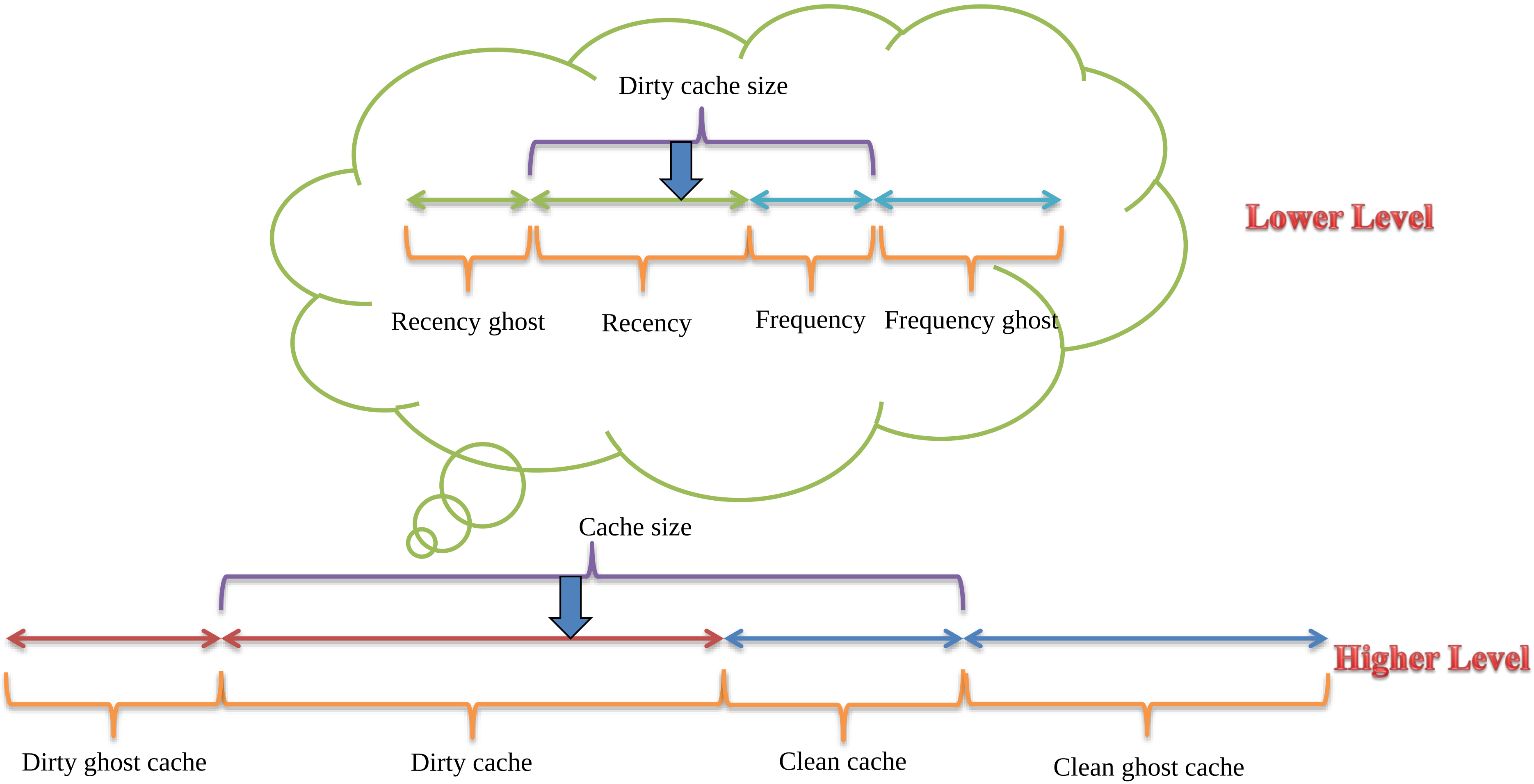
H-ARC: Learning (2/3)



H-ARC: Learning (3/3)



H-ARC: Eviction and Balance



Crash Recovery

- For system crashes or unexpected power failures
  - Goal: flush all cached dirty pages back
  - Solution:
    - Modify boot code
    - Page table and dirty pages well retained
    - Retrieve dirty pages from page table and flush them back
    - Reinitialize page table and continue booting process
- For hardware failures
  - Goal: mitigate the risk of losing dirty pages
  - Solution:
    - Add a timer to each dirty page
    - When time up, force flush back

Evaluation: Setup

- Simulator: Sim-ideal
- Comparison algorithms:
  - *MIN*: Belady MIN, optimal offline cache policy.
  - *LRU*: Least Recently Used cache policy. (hit ratio ↑)
  - *ARC*: Adaptive Replacement Cache policy. (hit ratio ↑)
  - *CFLRU [2]*: Clean First LRU cache policy. (write count ↓)
  - *LRU-WSR [3]*: Least Recently Used-Writes Sequence Reordering cache policy. (write count ↓)

[2] S. Park, D. Jung, J. Kang, J. Kim, and J. Lee. CFLRU: a Replacement Algorithm for Flash Memory. In CASES '06.
[3] H. Jung, H. Shim, S. Park, S. Kang, and J. Cha. LRU-WSR: Integration of LRU and Writes Sequence Reordering for Flash Memory. In IEEE Transactions on Consumer Electronics, Vol. 54, No. 3, August 2008

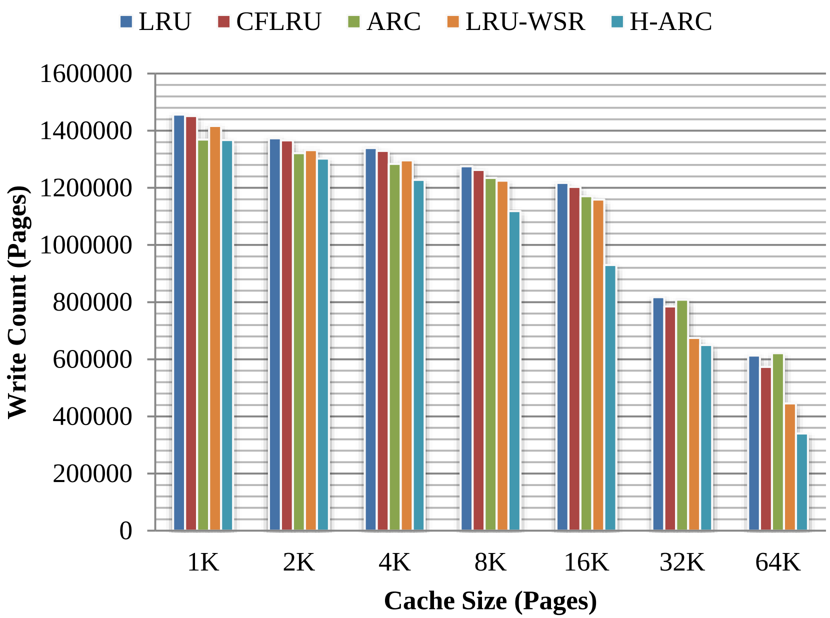
Evaluation: Workload

- Two types of workload
- MSR Cambridge traces [4]
  - Write intensive
  - Filtered by the buffer cache when collecting
- Synthetic workload
  - Fio and filebench
  - Read intensive
  - Use directI/O option to mimic workload issuing to buffer cache directly

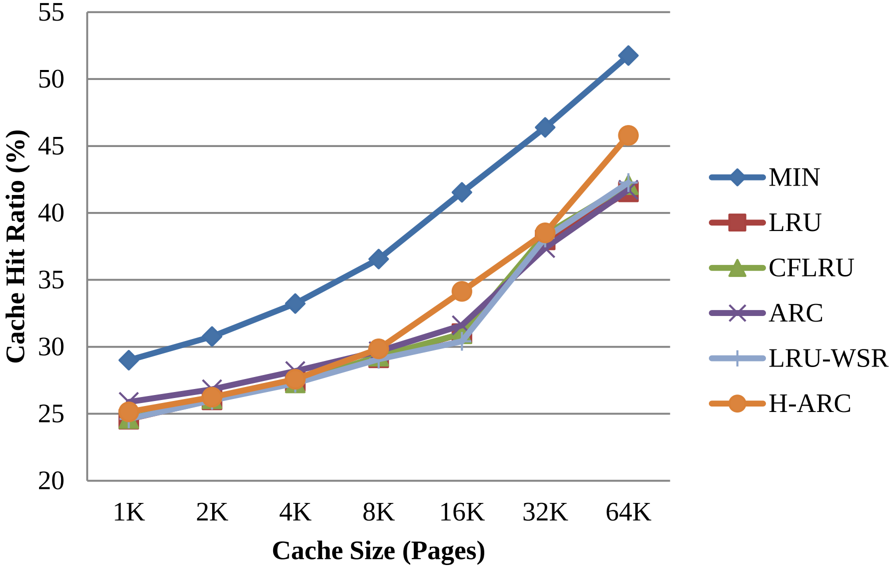
[4] SNIA. http://www.snia.org/

Evaluation: Results (1/3)

- ❖ Trace name: web_0
- ❖ Total requests: 7,129,953
- ❖ Unique pages: 1,724,201
- ❖ Read/write ratio: 1:0.76



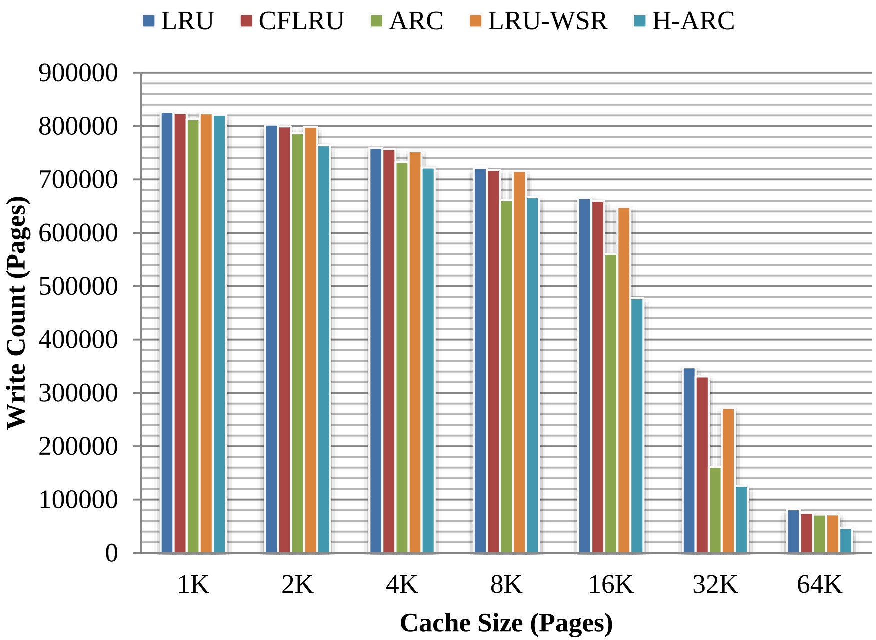
Write Count



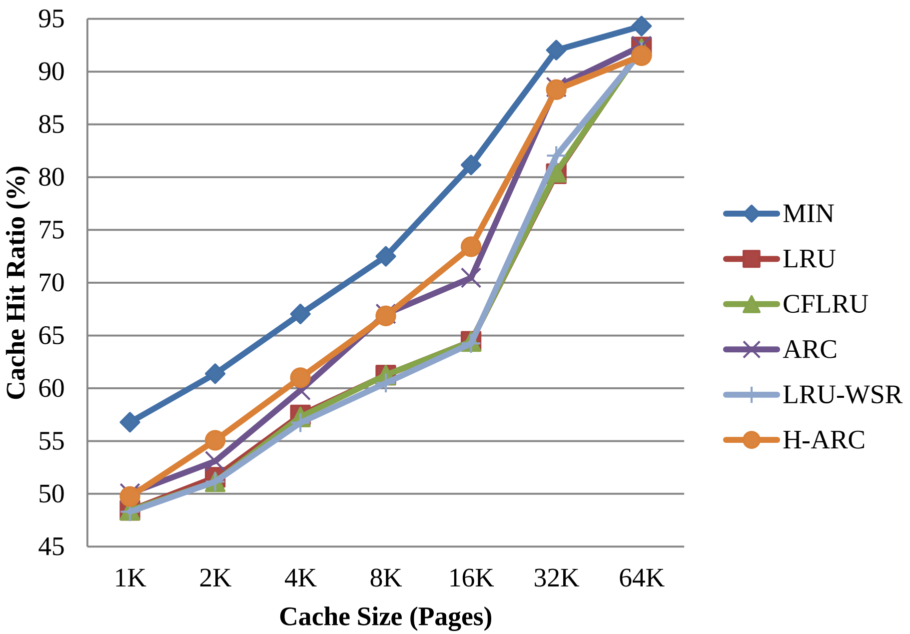
Hit Ratio

Evaluation: Results (2/3)

- Trace name: wdev_0
- Total requests: 2,368,194
- Unique pages: 128,870
- Read/write ratio: 1:3.73



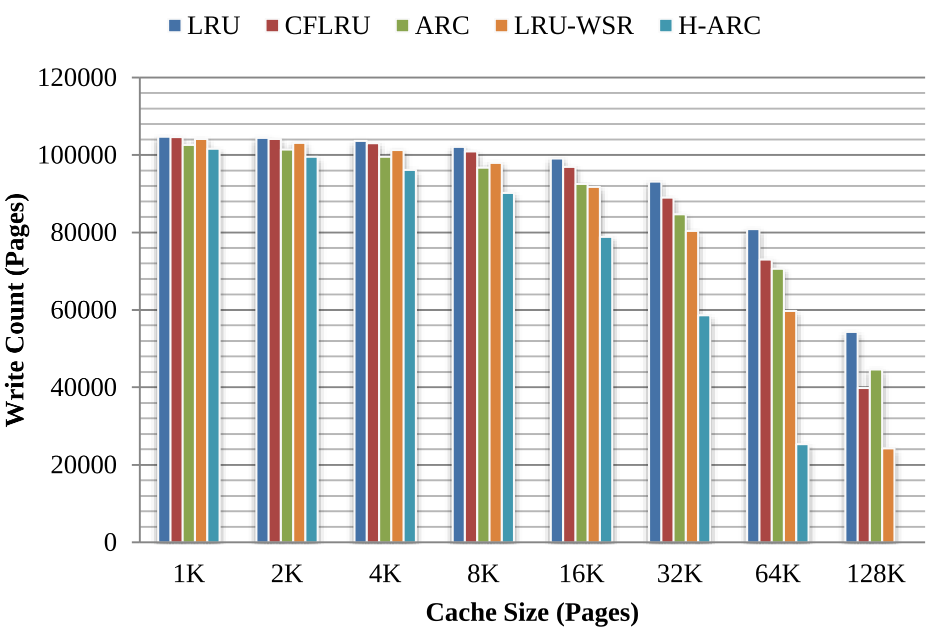
Write Count



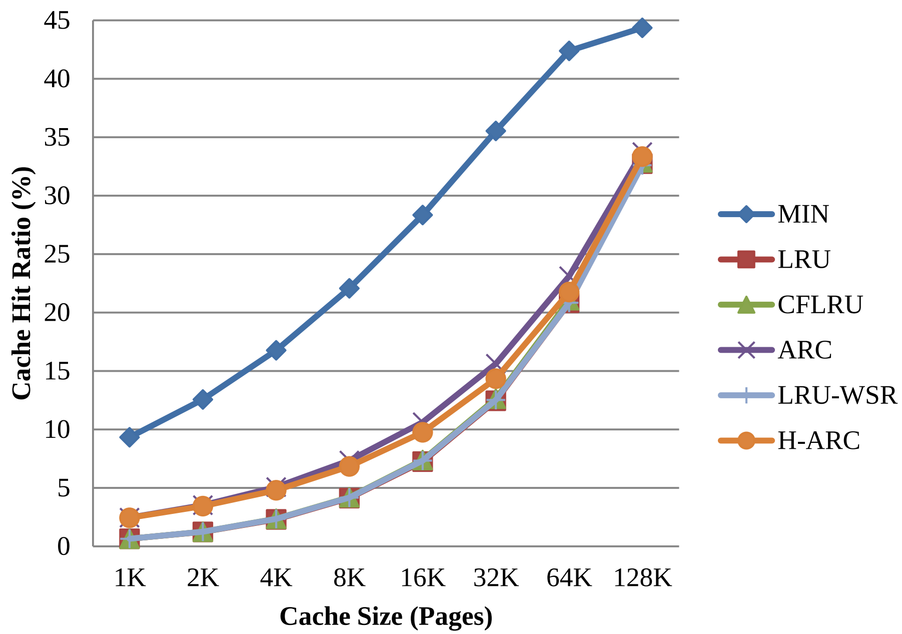
Hit Ratio

Evaluation: Results (3/3)

- ❖ Trace name: fio zipf
- ❖ Total requests: 524,411
- ❖ Unique pages: 291,812
- ❖ Read/write ratio: 1:0.25



Write Count



Hit Ratio

Questions?

THANK YOU

Ziqi Fan
fan@cs.umn.edu