TaC: An Anti-Caching Key-Value Store on Heterogeneous Memory Architectures

Yunhong Ji*, Wentao Huang†, Xuan Zhou‡,, Bingsheng He†, Kian-Lee Tan†

*Renmin University of China
†National University of Singapore
‡East China Normal University

In-Memory Key Value Stores

- **Fundamental storage layer**
  - E.g. Redis, Memcached
- **Operation simple**
  - Get, Set
- **Performance critical**

Memcached

The Problem

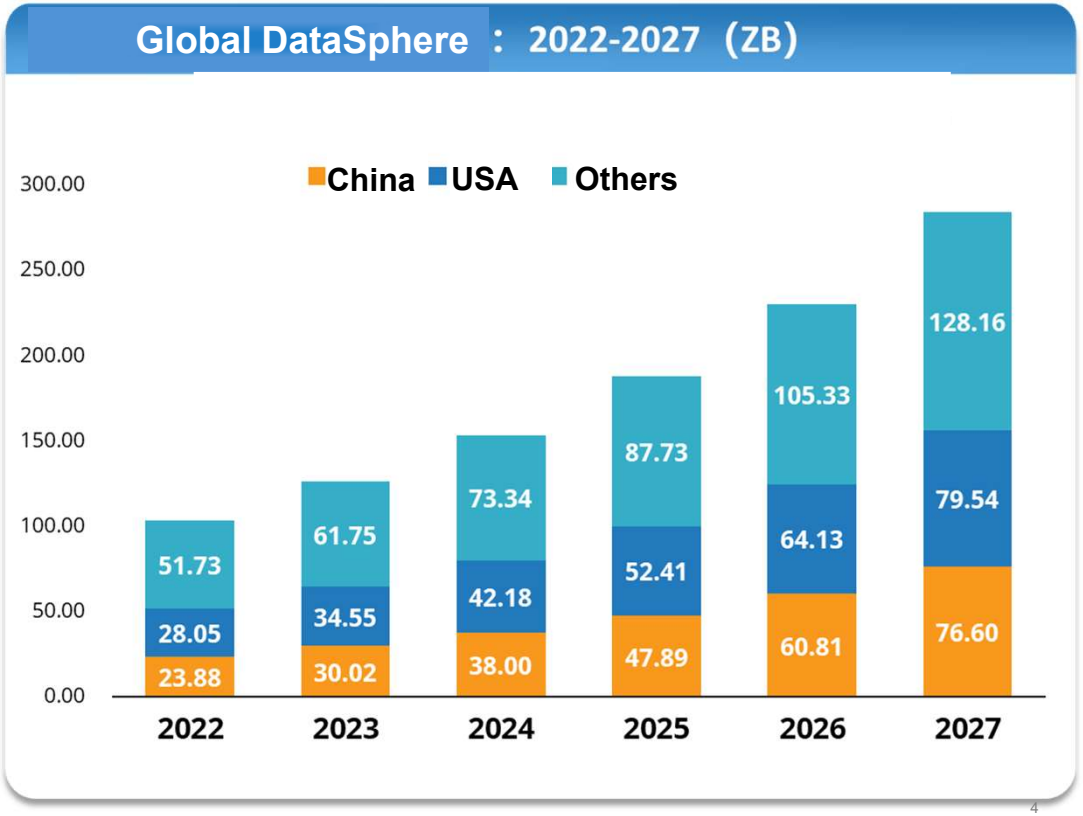
- The increasing data volume

  V.S.

- The memory scaling wall

The Increasing Data Volume

- **Data volume remains increasing**



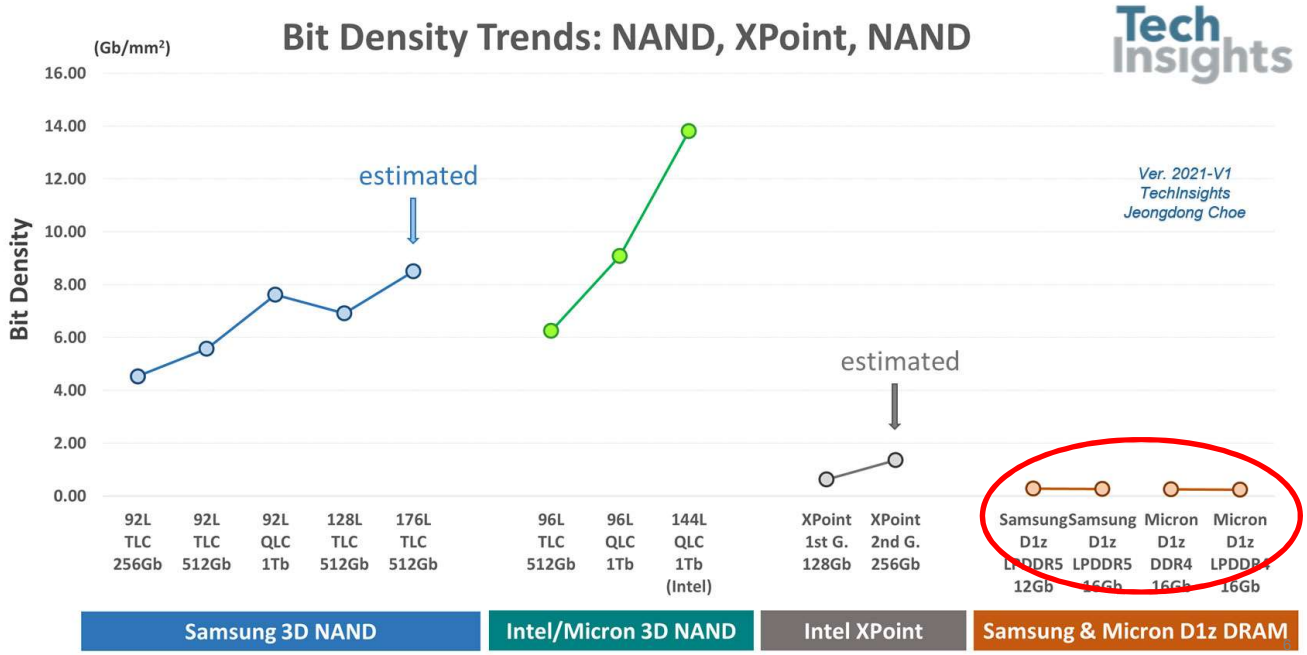
Source: IDC Global DataSphere 2023

The Memory Scaling Challenge

- **DRAM scaling wall**

The Memory Scaling Challenge

- **DRAM scaling wall**
  - Scaling DRAM technologies to sub-20nm is challenging



The Problem

- **Growing imbalance between Memory demand and supply**
  - Increasing demand of memory in applications
  - DRAM capacity pre core dropping by 30% biannually

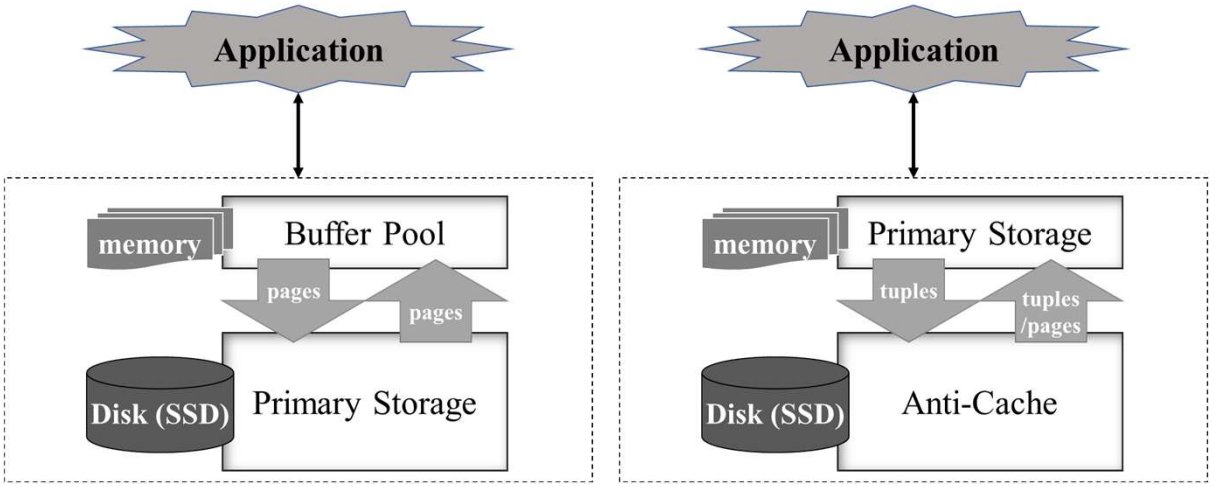
The Problem

- **Growing imbalance between Memory demand and supply**
  - Increasing demand of memory in applications
  - DRAM capacity pre core dropping by 30% biannually

- **Memory becomes more expensive compared to other devices**
  - 2TB M.2 SSD $< \$500$
  - 2TB DRAM $\approx \$20{,}000$

Anti-caching Architecture

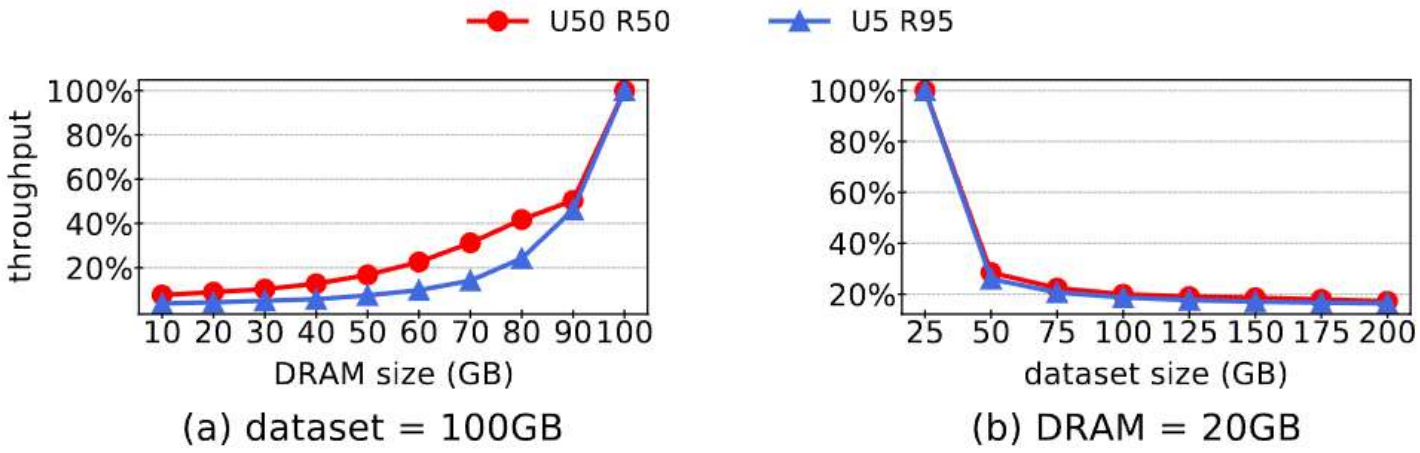
- **Expand the capacity while providing excellent performance**



Disk-based v.s. Anti-caching

Anti-caching Architecture

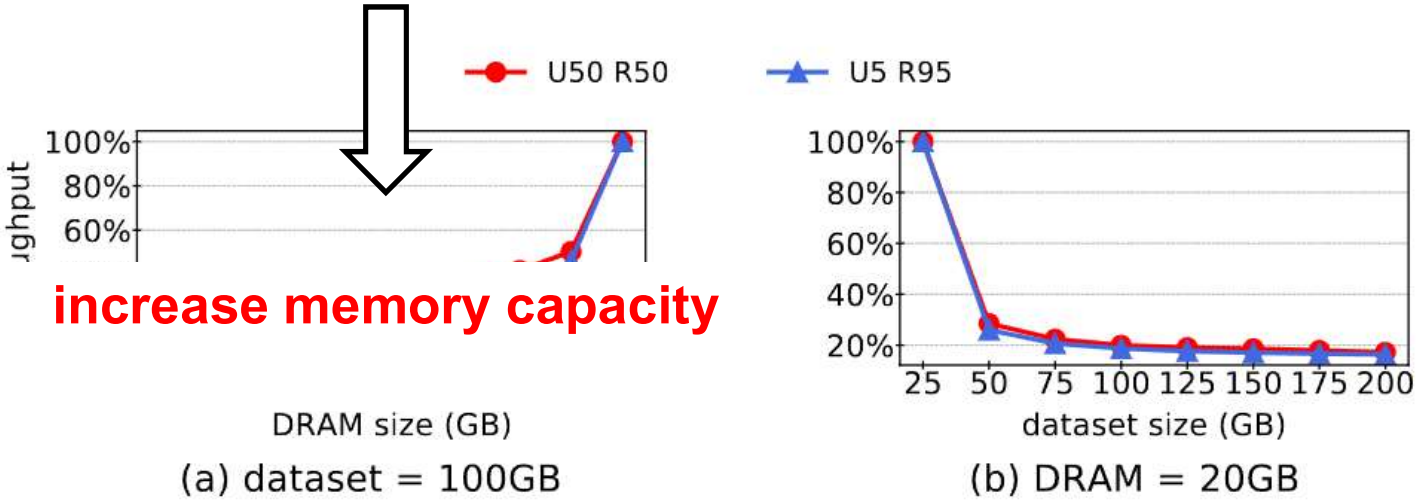
- **Performance diminishes when hot data surpasses the available memory size**



Normalized performance of FASTER with different data volume and DRAM configuration

Anti-caching Architecture

- **Performance diminishes when hot data surpasses the available memory size**



Normalized performance of FASTER with different data volume and DRAM configuration

Non-Volatile Memory

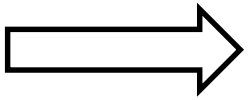
- **Byte-addressable**
- **Larger capacity**
- **Lower prices**

Table 1: Capacity and prices ($/GB) of different memory.

Capacity	128GB	256GB	512GB
DDR5 DRAM DIMM [12]	11.3	12.5	-
Intel Optane PMem [32]	8.6	8.4	8.2

Non-Volatile Memory

- **Byte-addressable**
- **Larger capacity**
- **Lower prices**



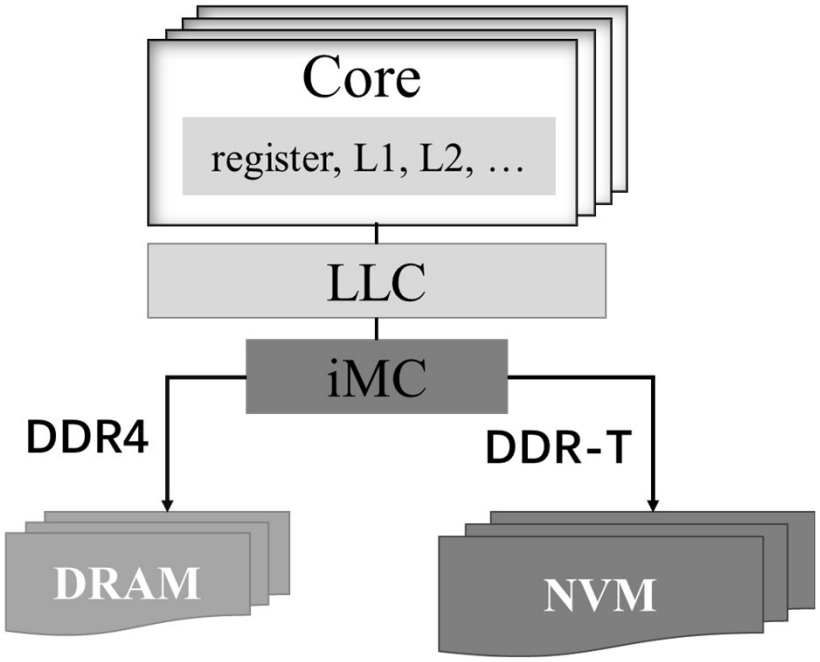
Leverage NVM to expand the memory in anti-caching Key-Value stores

Table 1: Capacity and prices ($/GB) of different memory.

Capacity	128GB	256GB	512GB
DDR5 DRAM DIMM [12]	11.3	12.5	-
Intel Optane PMem [32]	8.6	8.4	8.2

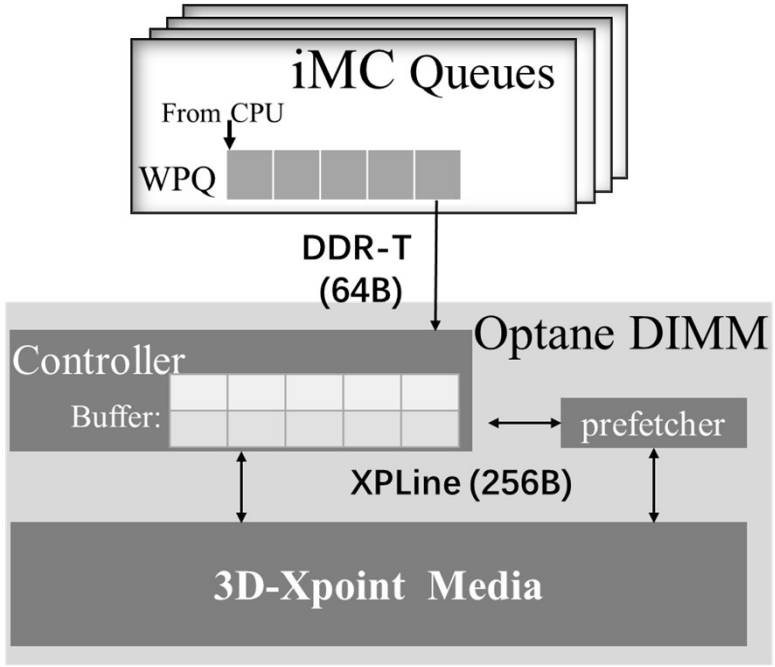
Non-Volatile Memory

- **NVDIMM-P**



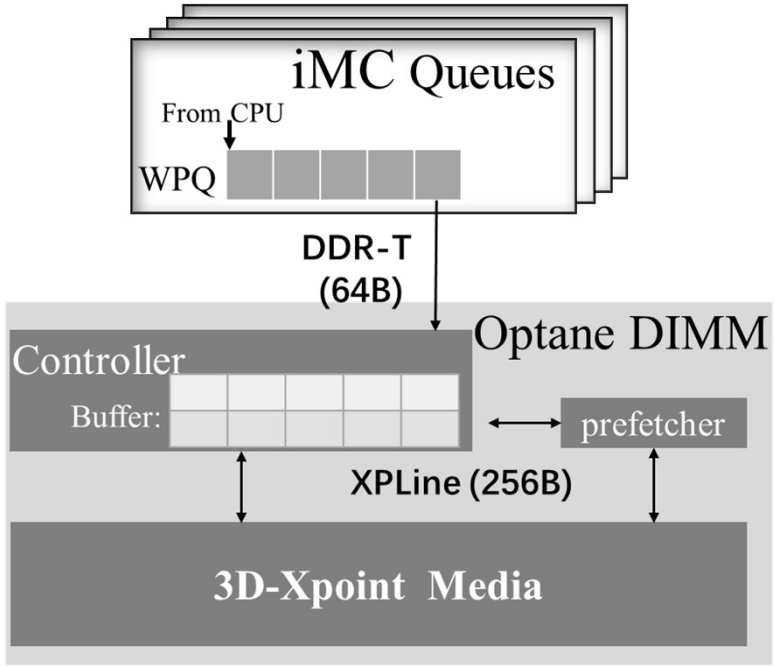
Non-Volatile Memory

- **Performance characteristics**



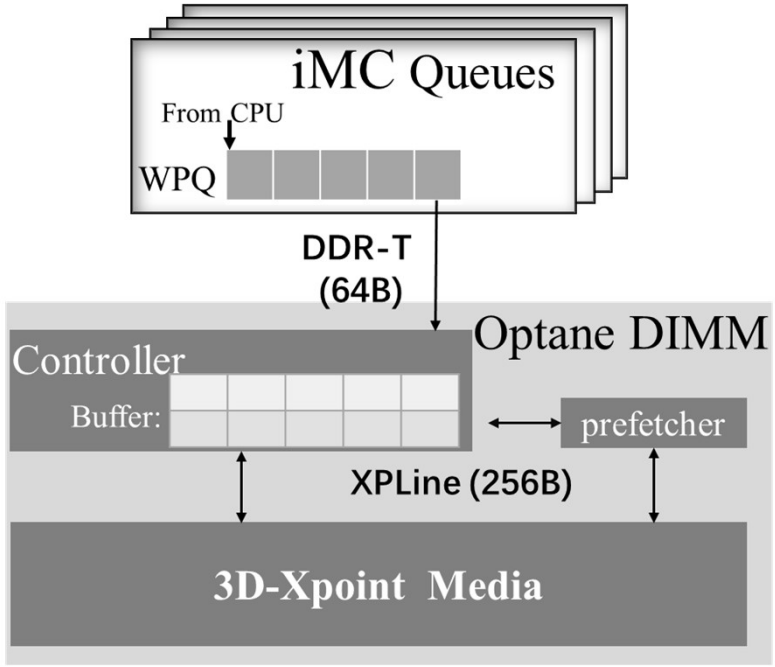
Non-Volatile Memory

- **Performance characteristics**
  - $\mathfrak{p}1$: Inferior performance
  - $\mathfrak{p}2$: Read-write asymmetry



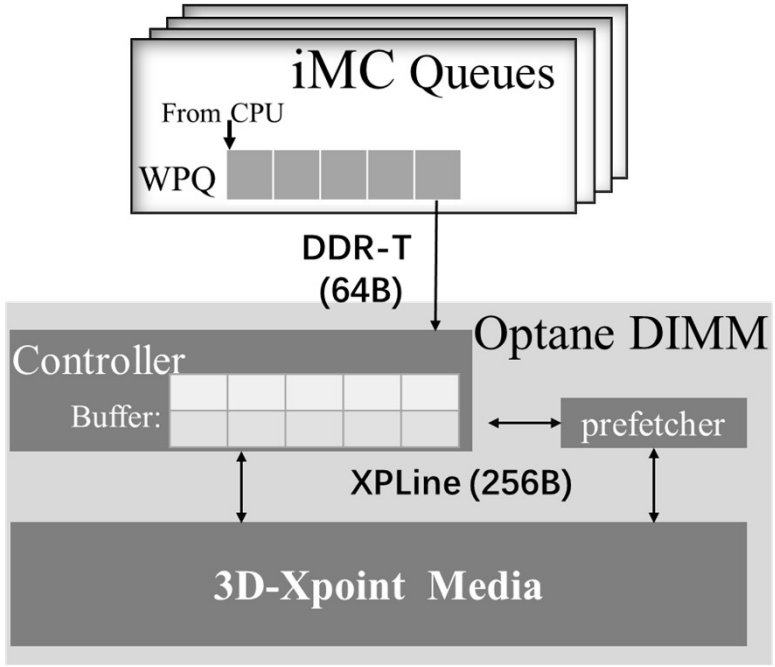
Non-Volatile Memory

- **Performance characteristics**
  - $\mathfrak{p}1$: Inferior performance
  - $\mathfrak{p}2$: Read-write asymmetry
  - $\mathfrak{p}3$: Inferior performance on small and random accesses



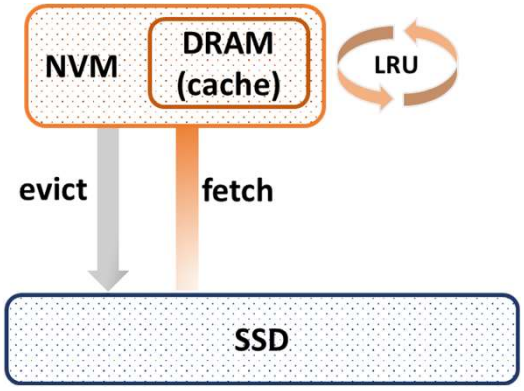
Non-Volatile Memory

- **Performance characteristics**
  - $\mathfrak{p}1$: Inferior performance
  - $\mathfrak{p}2$: Read-write asymmetry
  - $\mathfrak{p}3$: Inferior performance on small and random accesses
  - $\mathfrak{p}4$: Limited concurrency
  - $\mathfrak{p}5$: Interference with DRAM



Anti-Caching with NVM

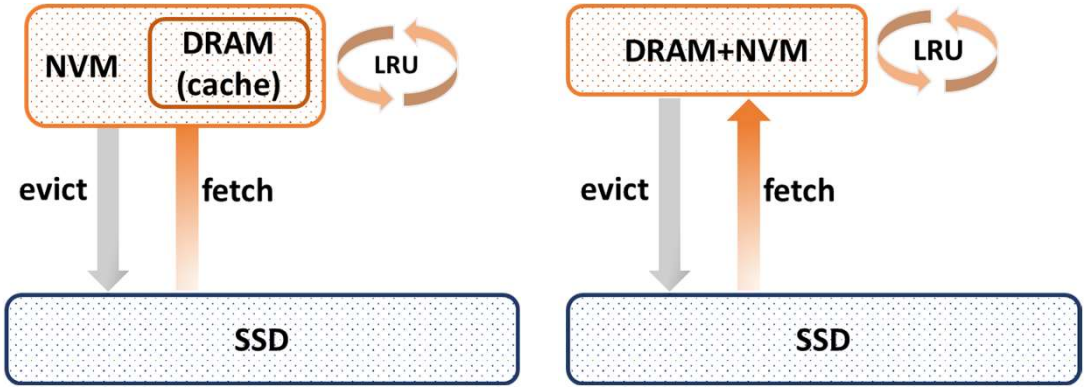
- **Potential choices**
  - Anti-NVM: Replace DRAM with NVM, similar to “Memory Mode”



Anti-NVM

Anti-Caching with NVM

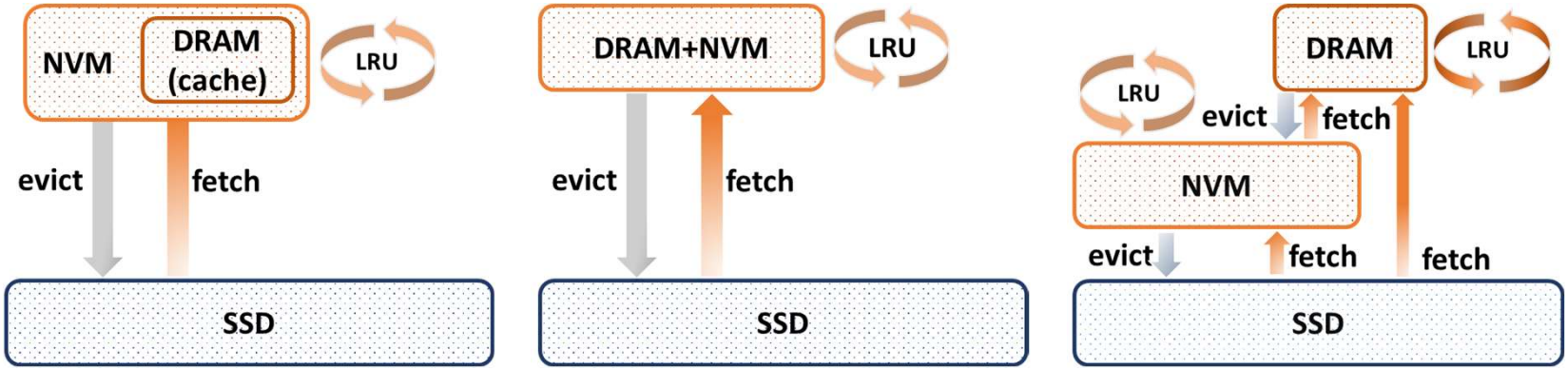
- **Potential choices**
  - Anti-NVM: Replace DRAM with NVM, similar to “Memory Mode”
  - Anti-2: Utilize NVM in the same way as DRAM



Anti-NVM Anti-2

Anti-Caching with NVM

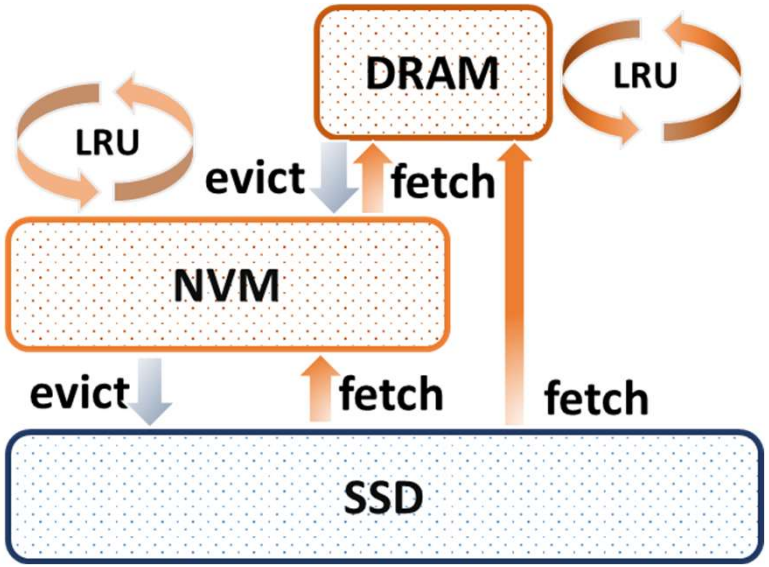
- **Potential choices**
  - Anti-NVM: Replace DRAM with NVM, similar to “Memory Mode”
  - Anti-2: Utilize NVM in the same way as DRAM
  - TaC: Three-tier anti-Caching



Anti-NVM Anti-2 TaC

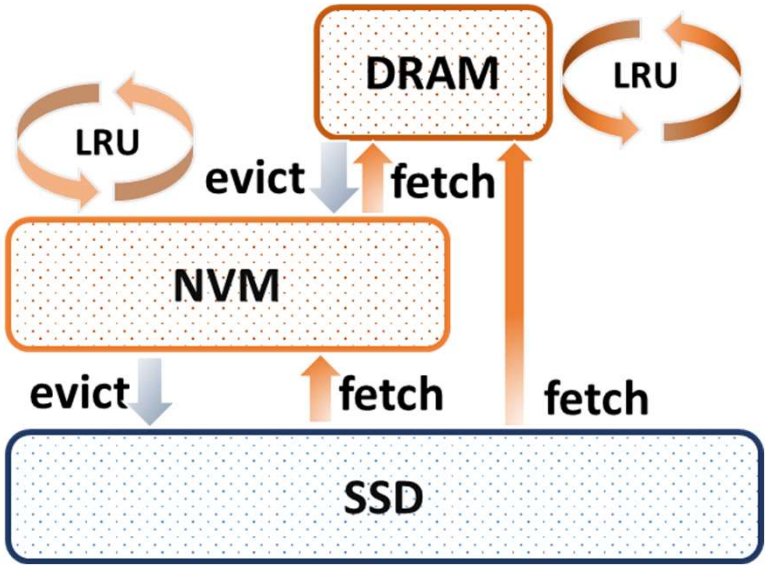
Three-tier anti-Caching

- **Strengths**
  - Leveraging DRAM Strengths ($\mathfrak{p}1$)



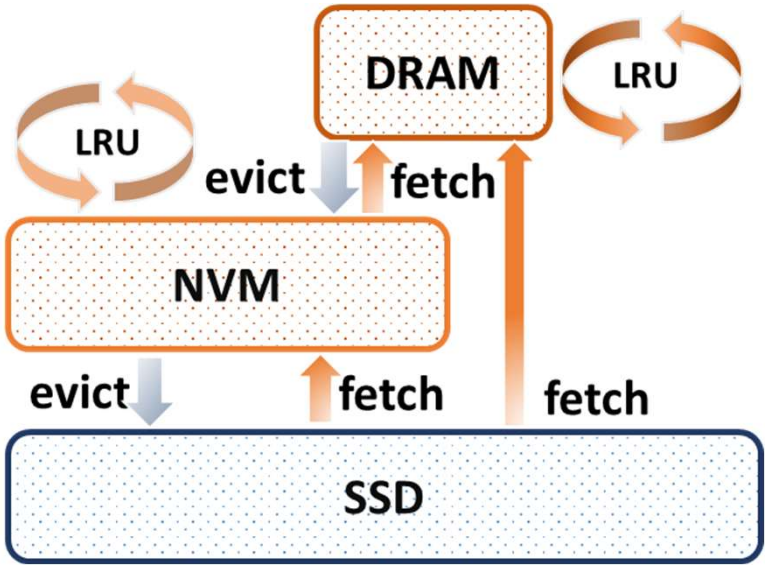
Three-tier anti-Caching

- **Strengths**
  - Leveraging DRAM Strengths ($\mathfrak{p}1$)
  - Hot Data Retention



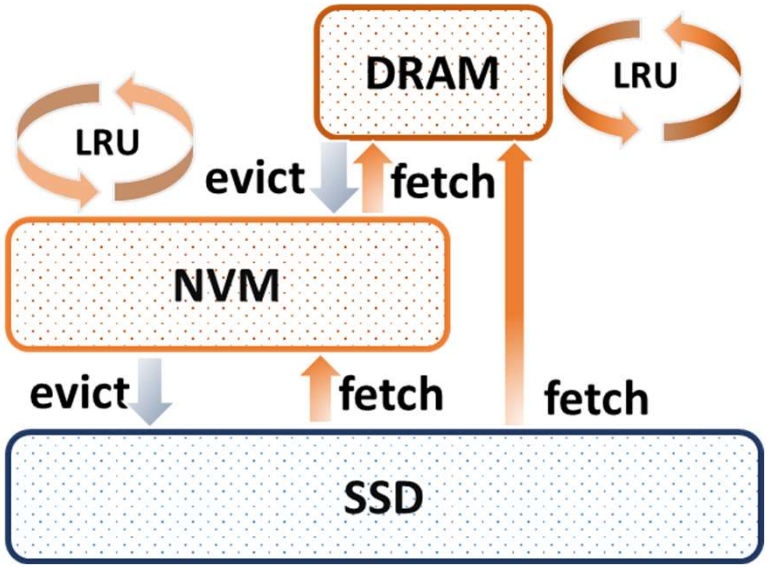
Three-tier anti-Caching

- **Strengths**
  - Leveraging DRAM Strengths ($\mathbf{p}1$)
  - Hot Data Retention
  - Controlled Writing to NVM ($\mathbf{p}4, \mathbf{p}5$)



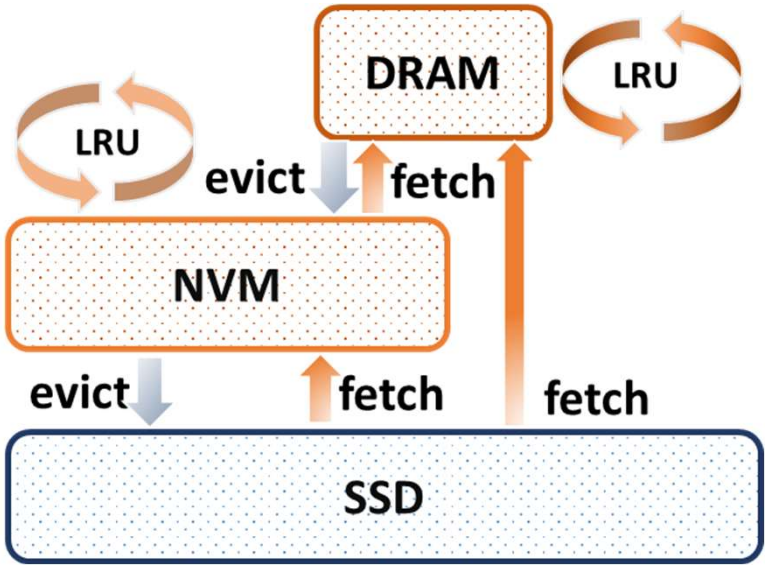
Three-tier anti-Caching

- **Challenges**
  - More complex data swapping paths



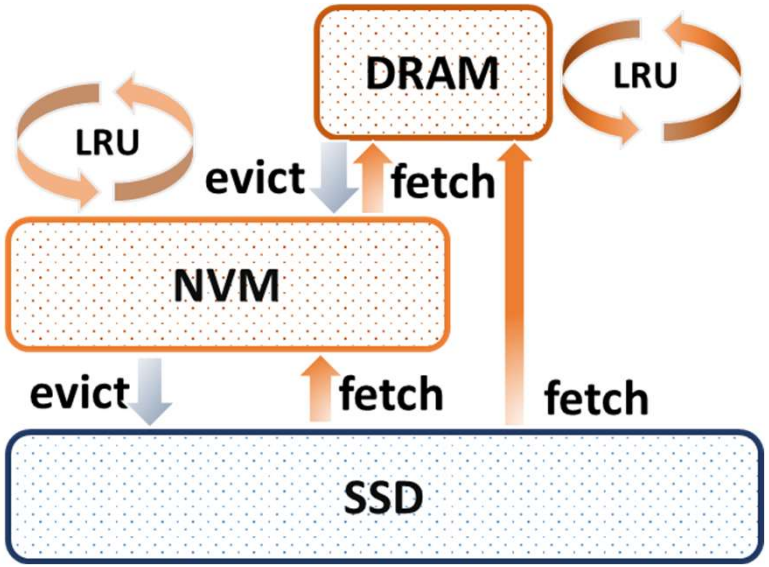
Three-tier anti-Caching

- **Challenges**
  - More complex data swapping paths
  - Multi-level data classification



Three-tier anti-Caching

- **Challenges**
  - More complex data swapping paths
  - Multi-level data classification
  - NVM-specific characteristics

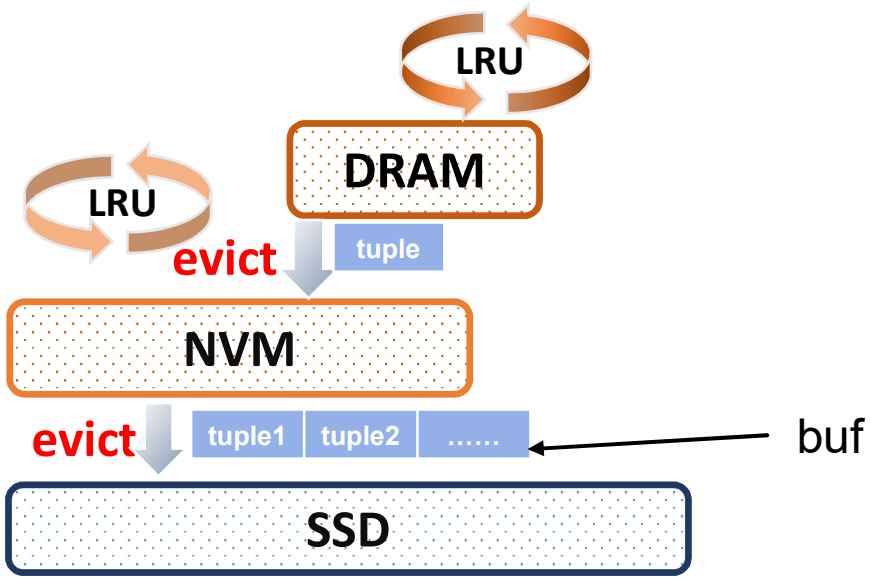


Three-tier anti-Caching

- **Challenges**
  - More complex data swapping paths ⇨ Hybrid data swapping
  - Multi-level data classification ⇨ Lazy LRU
  - NVM-specific characteristics ⇨ NVM-optimized data arrangement

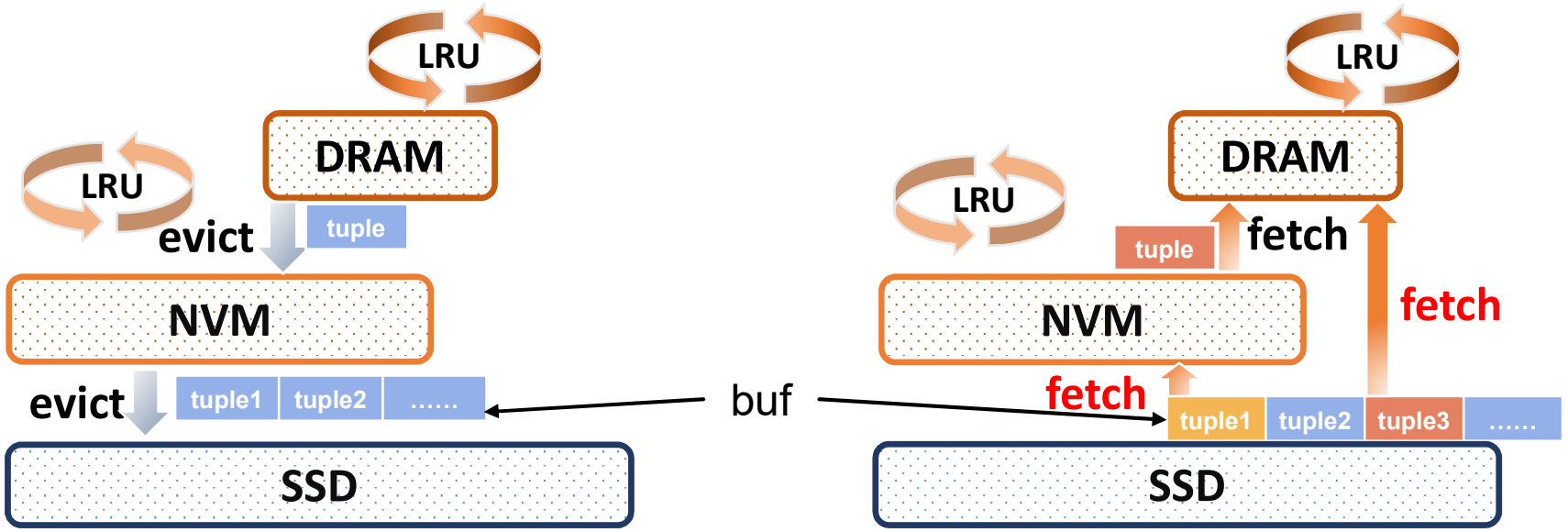
Three-tier anti-Caching

- Hybrid data swapping



Three-tier anti-Caching

- Hybrid data swapping



Three-tier anti-Caching

- **Lazy LRU**
  - $time$, the previous access time
  - $readCount$, the recent access frequency

Table 2: Temperature classes of tuples.

$t.readCount$	Temperature	Action
$> THR_{hot}$	Hot	Fetch from SSD or NVM to DRAM; Move atop the LRU list if in DRAM.
$> THR_{warm}$	Warm	Fetch from SSD to NVM; Move atop the LRU list if in NVM.
Others	Cold	Become ready for eviction.

Three-tier anti-Caching

- **Lazy LRU**
  - $time$, the previous access time
  - $readCount$, the recent access frequency

- Multi-level data classification
- asynchronous updates of LRU lists

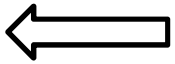


Table 2: Temperature classes of tuples.

$t.readCount$	Temperature	Action
$> THR_{hot}$	Hot	Fetch from SSD or NVM to DRAM; Move atop the LRU list if in DRAM.
$> THR_{warm}$	Warm	Fetch from SSD to NVM; Move atop the LRU list if in NVM.
Others	Cold	Become ready for eviction.

Three-tier anti-Caching

- **Lazy LRU**
  - $time$, the previous access time
  - $readCount$, the recent access frequency

Table 2: Temperature classes of tuples.

$t.readCount$	Temperature	Action
$> THR_{hot}$	Hot	Fetch from SSD or NVM to DRAM; Move atop the LRU list if in DRAM.
$> THR_{warm}$	Warm	Fetch from SSD to NVM; Move atop the LRU list if in NVM.
Others	Cold	Become ready for eviction.

Three-tier anti-Caching

- **Lazy LRU**
  - $time$, the previous access time
  - $readCount$, the recent access frequency

- Sample-based thresholds

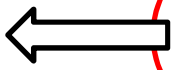
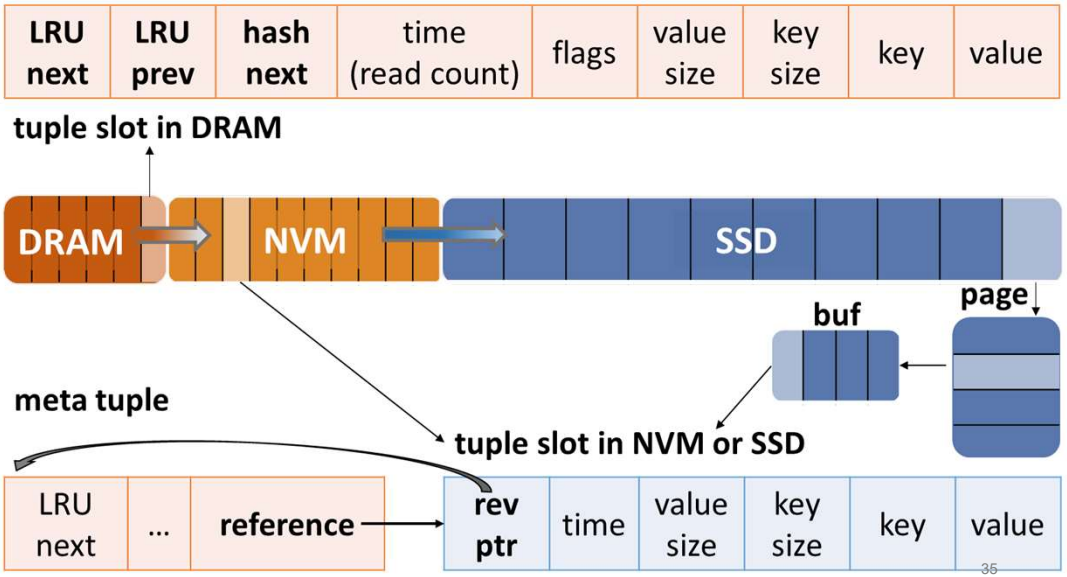


Table 2: Temperature classes of tuples.

$t.readCount$	Temperature	Action
$> THR_{hot}$	Hot	Fetch from SSD or NVM to DRAM; Move atop the LRU list if in DRAM.
$> THR_{warm}$	Warm	Fetch from SSD to NVM; Move atop the LRU list if in NVM.
Others	Cold	Become ready for eviction.

Three-tier anti-Caching

- **NVM-optimized data arrangement based on Memcached**
  - Manage NVM space at tuple-level like DRAM
  - Maintain the metadata in DRAM

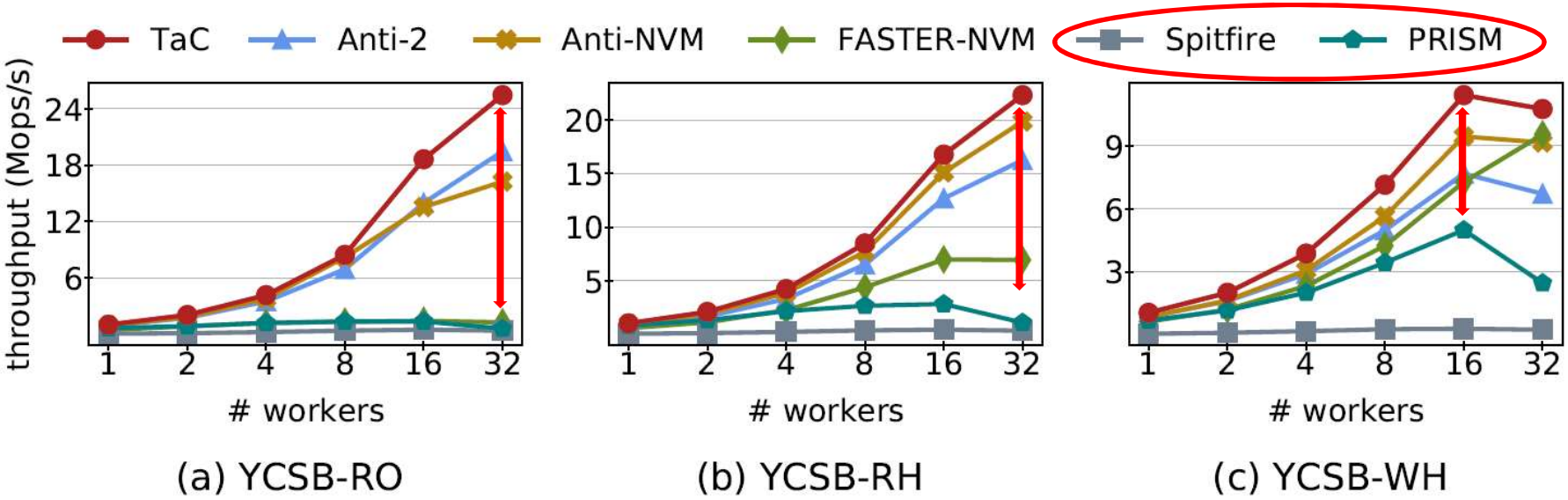


Evaluation

- **Compared systems**
  - Anti-NVM
  - Anti-2 & FASTER-NVM
  - Spitfire & PRISM
- **Set-Up**
  - 4GB DRAM
  - 32GB NVM
- **Workloads**
  - Read-Only (YSCB-RO): 100% reads
  - Read-Heavy (YCSB-RH): 95% reads, 5% updates
  - Write-Heavy (YCSB-WH): 50% reads, 50% updates

Evaluation

- **Observations**
  - The anti-caching architecture outperforms the caching architecture regarding the throughput

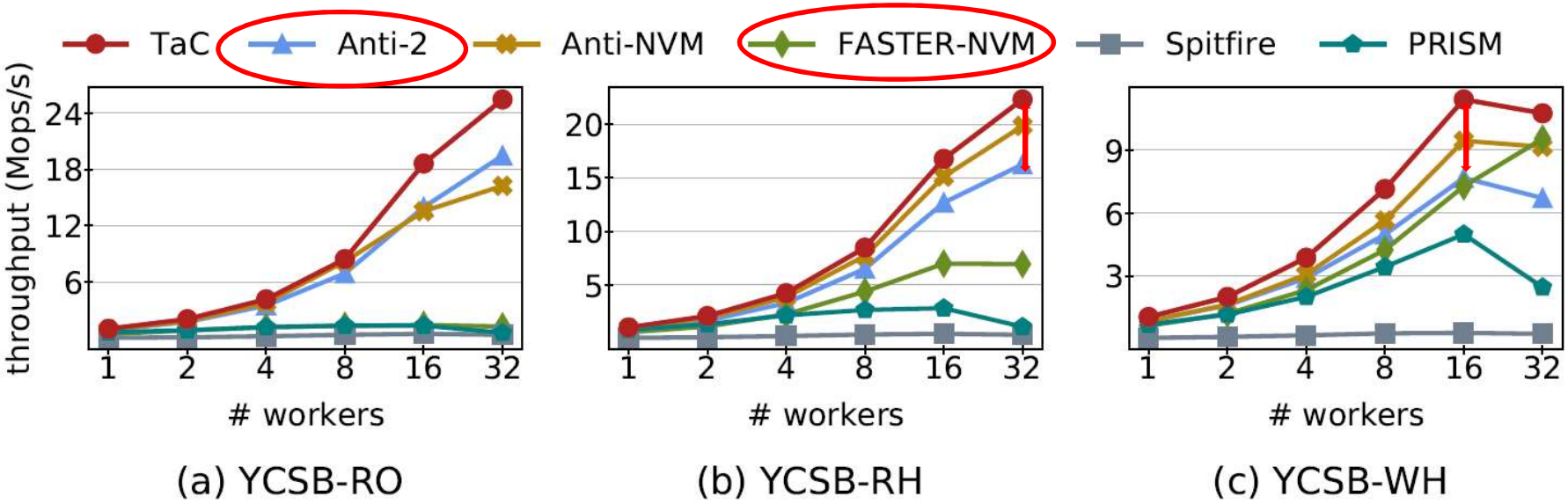


Evaluation

- **Observations**
  - Distinguishing between NVM and DRAM enhances system performance by improving DRAM hit rates and NVM accessing efficiency

Table 3: Memory hit rates on the YCSB-RH workload.

	TaC	Anti-2	Anti-NVM
DRAM	61.4%	9.1%	59.8%
NVM	21.0%	76.3%	18.8%

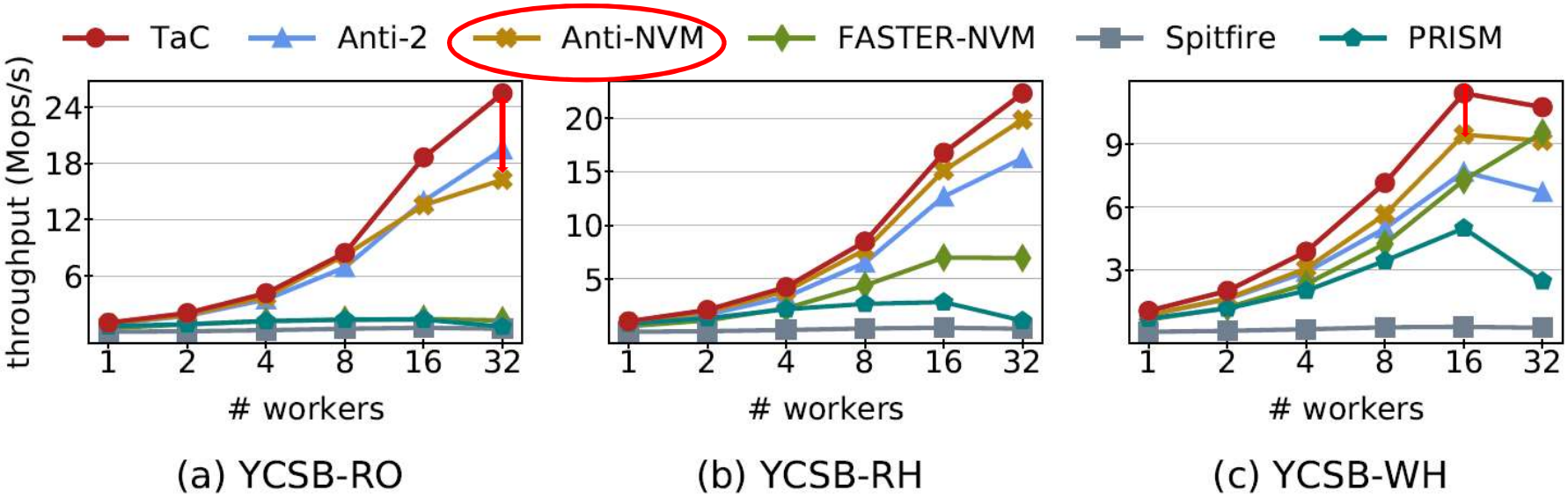


Evaluation

- **Observations**
  - The three-tier architecture outperforms the ``Memory Mode'' architecture due to better memory utilization and reduced data synchronization cost

Table 3: Memory hit rates on the YCSB-RH workload.

	TaC	Anti-2	Anti-NVM
DRAM	61.4%	9.1%	59.8%
NVM	21.0%	76.3%	18.8%



(a) YCSB-RO (b) YCSB-RH (c) YCSB-WH

Evaluation

- **More experiments**
  - Experiments on varying DRAM and NVM sizes
  - Experiments on varying data volume
  - Optimization Impact Analysis
    - Partitioned memory management
    - Hybrid data swapping
    - Lazy LRU
  - ……

Thanks!