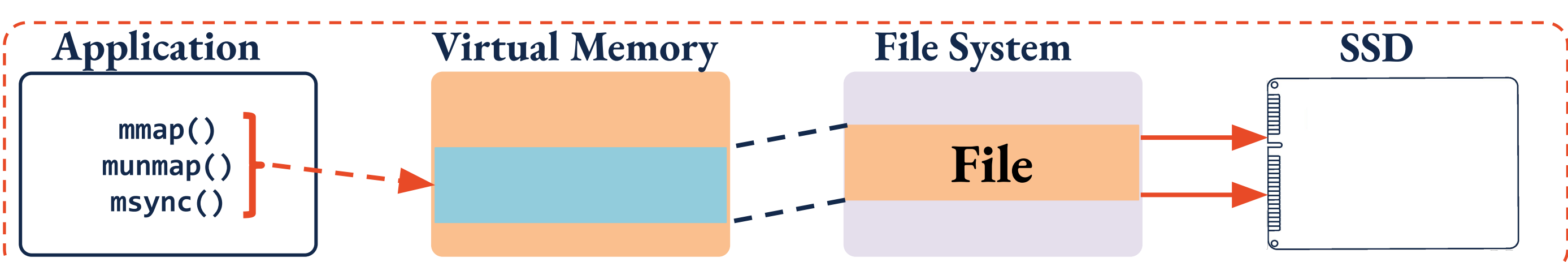# FlatFlash: Exploiting the Byte-Accessibility of SSDs Within A Unified Memory-Storage Hierarchy

Ahmed Abulila[1], Vikram Sharma Mailthody[1], Zaid Qureshi[2], Jian Huang[1], Nam Sung Kim[1], Jinjun Xiong[3], Wen-mei Hwu[1]
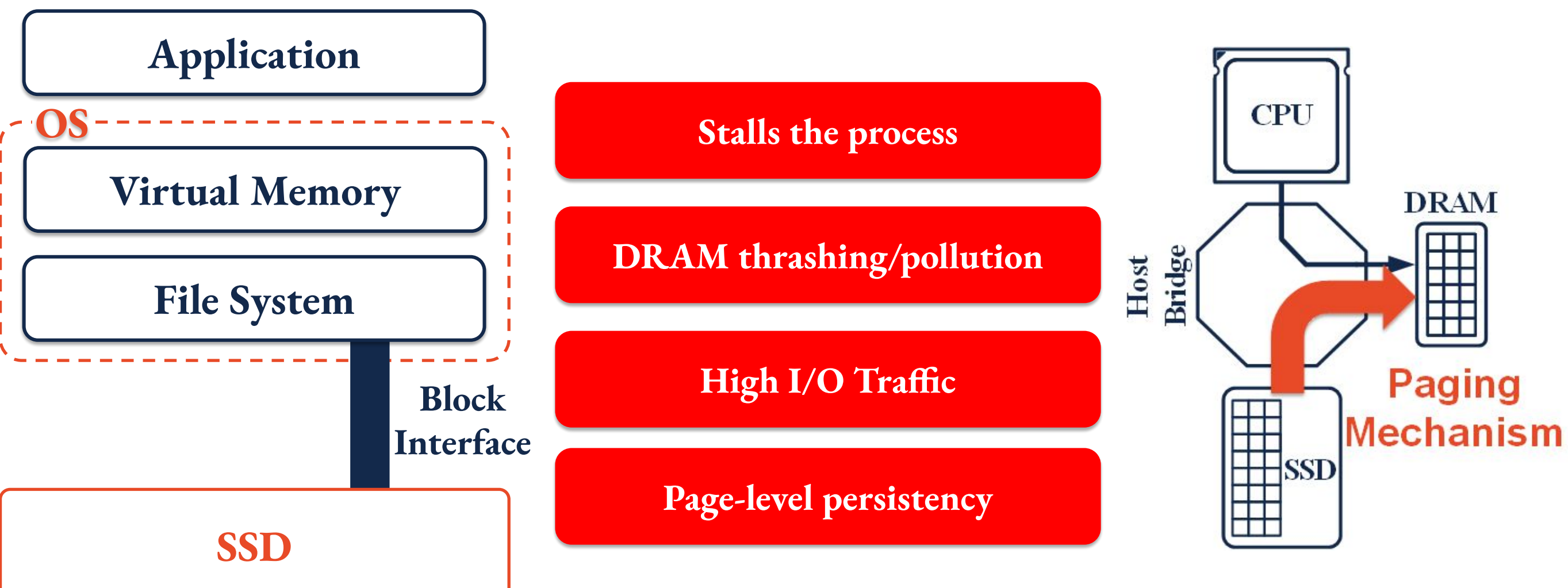
[1] ECE, [2] CS, University of Illinois at Urbana-Champaign, Urbana, IL 61801

[3] Cognitive Computing & University Partnership, IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598
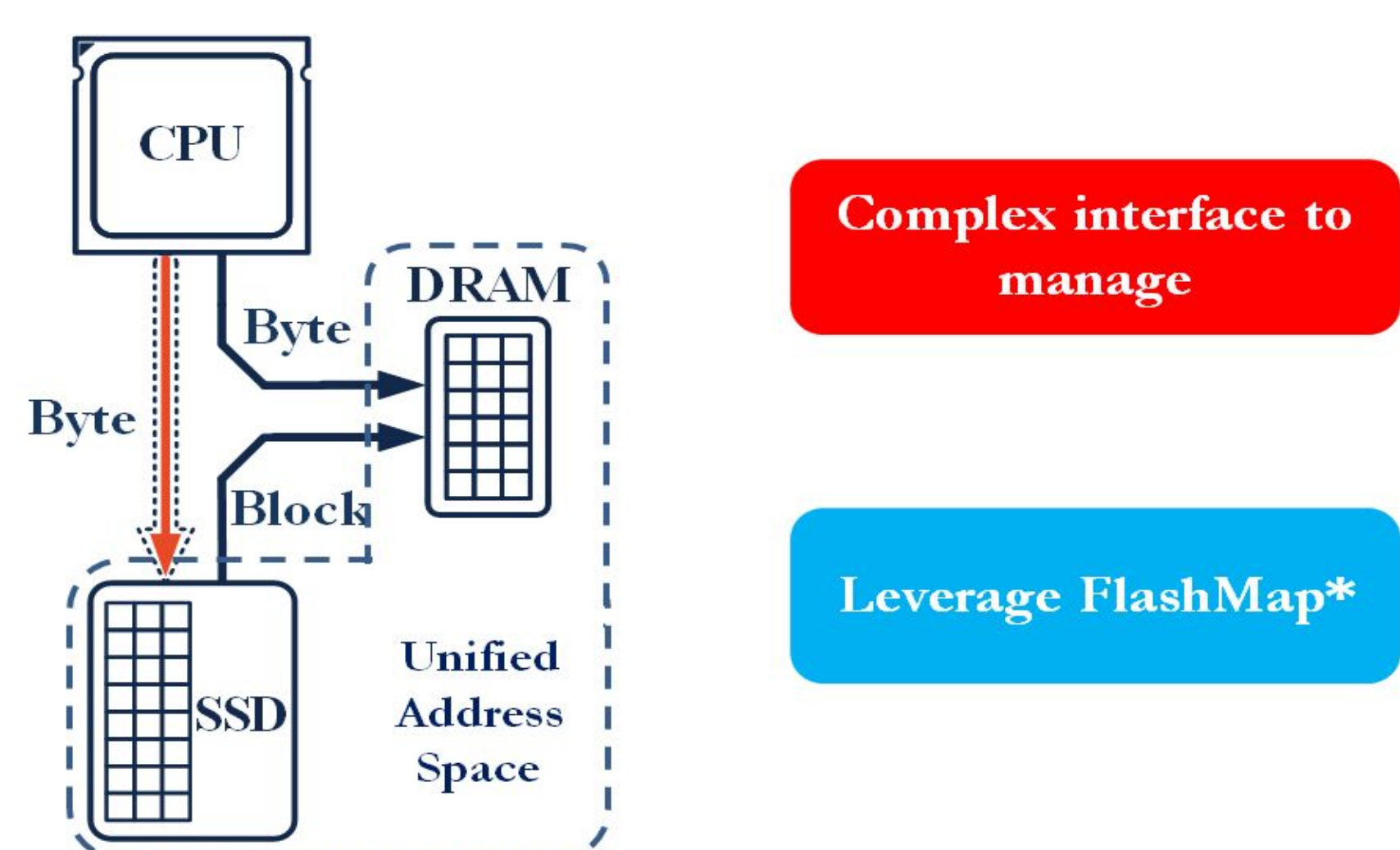
Using Flash as Non-Volatile Memory

Application: mmap(), munmap(), msync() — Virtual Memory — File System — SSD

Limitations of Block I/O Interface

Application / OS / Virtual Memory / File System / SSD / Block Interface

- Stalls the process
- DRAM thrashing/pollution
- High I/O Traffic
- Page-level persistency

Paging Mechanism

Enabling the Byte-Accessibility of SSDs

+ PCI EXPRESS + OpenCAPI

- DRAM in SSD
- PCIe MMIO
- Cache Coherence

Unifying the Memory and Storage

CPU — Byte — DRAM / Byte / Block — SSD / Unified Address Space

- Complex interface to manage
- Leverage FlashMap*

Performance Challenges

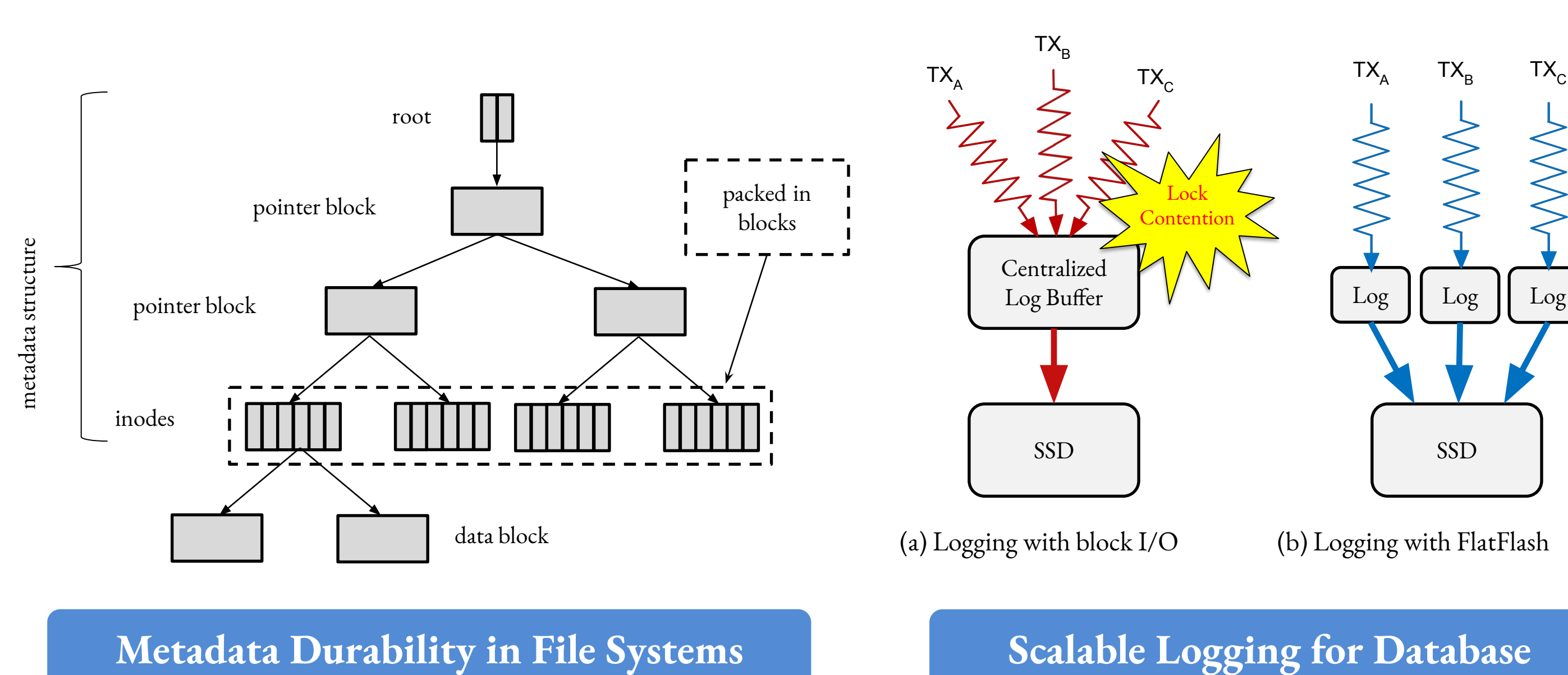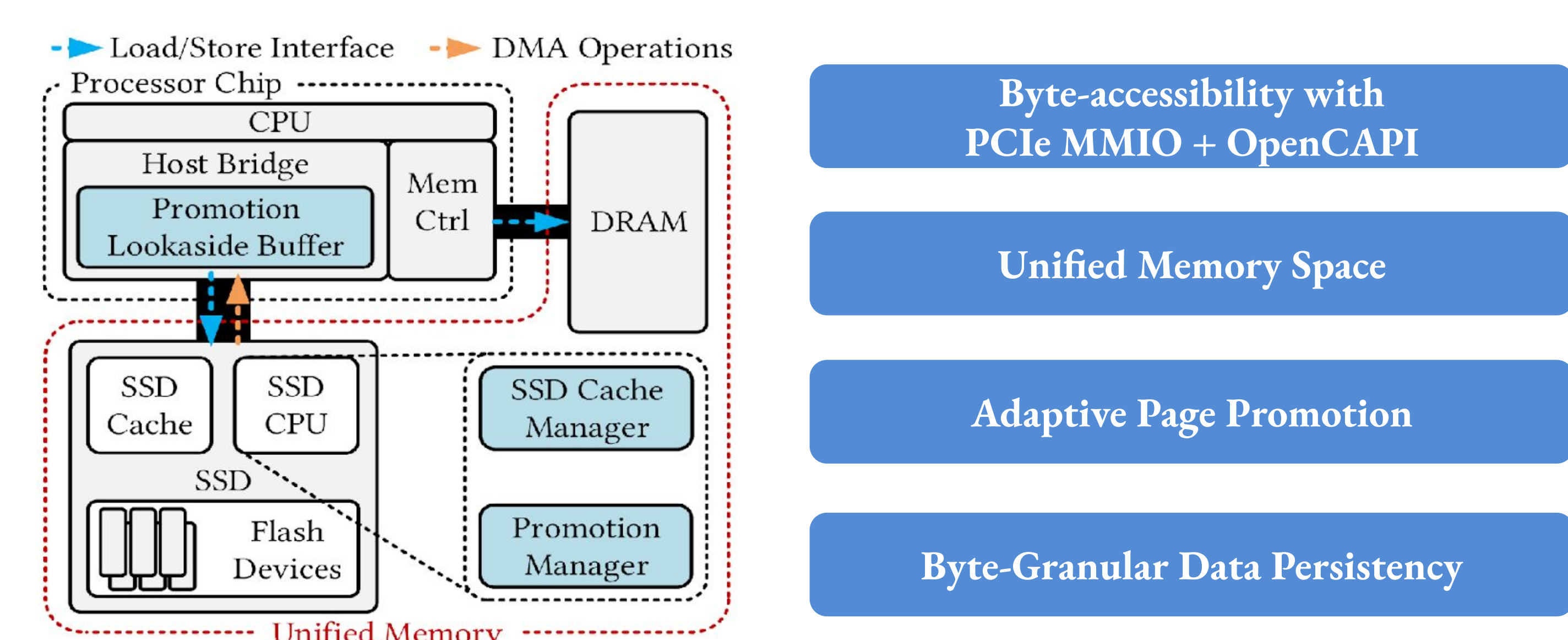| Overhead Source	Average (μsec)	
Read a cache line in SSD-Cache via PCIe MMIO	4.8	Page Promotion to Host DRAM
Write a cache line in SSD-Cache via PCIe MMIO	0.6	
Promote a page from SSD-Cache to host DRAM	12.1	Off-Critical Path Page Promotion
Update PTE and TLB entry in host machine	1.4	
Page table walking to get the page location	0.7	

Ensuring Data-Consistency During Page Promotion

CPU / Host Bridge / Promotion Lookaside Buffer / DRAM / SSD

In-flight Page Migration

In-Flight Promotion with Cache-Line Granularity

WriteBack (0x02, data)

Set Bits in Promotion Look-aside Buffer

Dropped

Handle Data Access Conflict
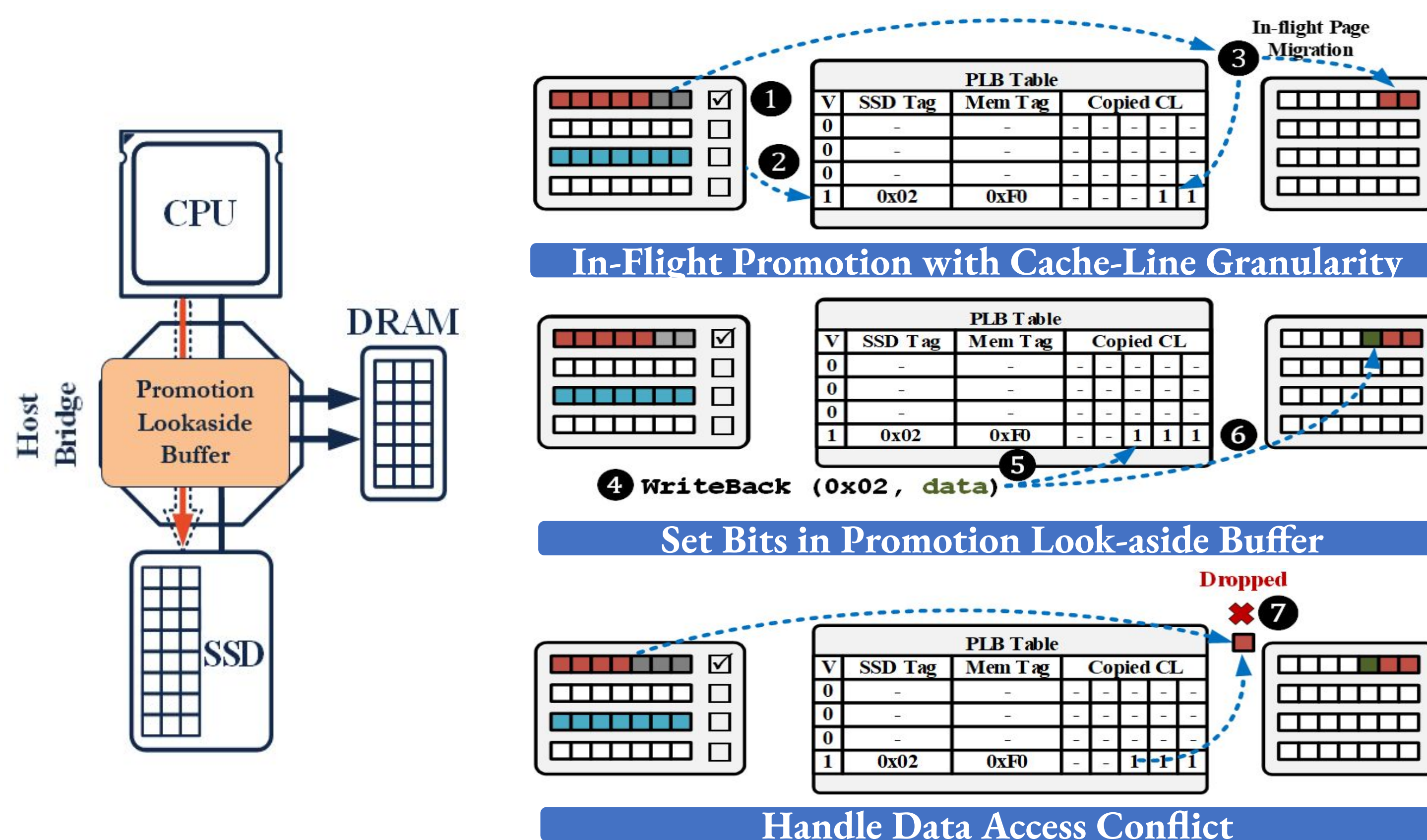
Enabling Byte-Granular Data Persistence

① flush cache lines: clflush/clwb
② verify addr & disable promotion
③ write-verify read

CPU — Cache — Promotion Lookaside Buffer / Host Bridge — PCIe — SSD-Cache / SSD

```
create_pmem_region (void* vaddr, size_t size)
```

FlatFlash disables page promotion for persistent memory regions
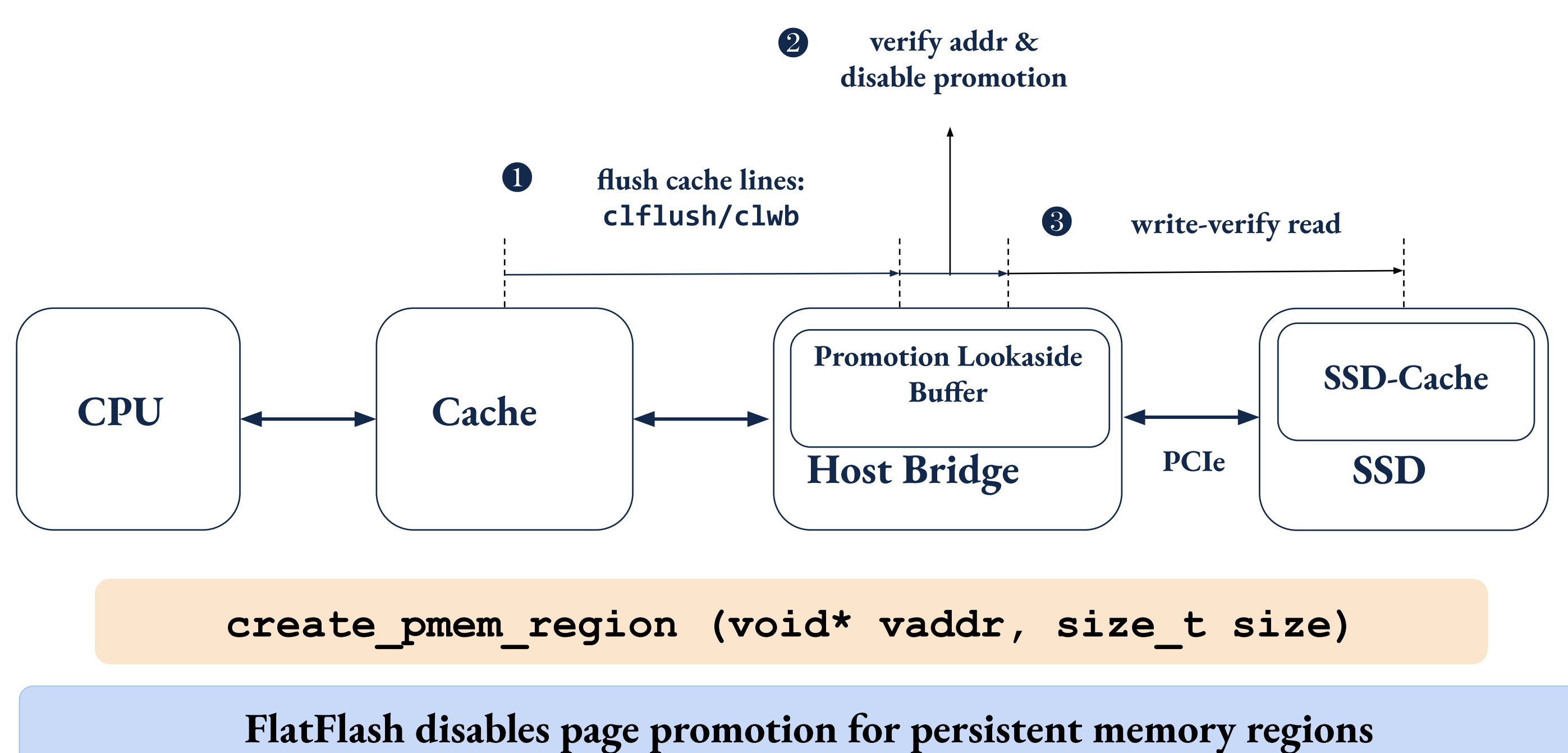
System Implications of FlatFlash

root / pointer block / pointer block / inodes / data block / packed in blocks

metadata structure

TX_A, TX_B, TX_C → Centralized Log Buffer → Lock Contention → SSD (a) Logging with block I/O

TX_A, TX_B, TX_C → Log, Log, Log → SSD (b) Logging with FlatFlash

Metadata Durability in File Systems

Scalable Logging for Database

Put It All Together

Load/Store Interface — DMA Operations

Processor Chip / CPU / Host Bridge / Promotion Lookaside Buffer / Mem Ctrl / DRAM / SSD Cache / SSD CPU / SSD Cache Manager / SSD / Flash Devices / Promotion Manager / Unified Memory

- Byte-accessibility with PCIe MMIO + OpenCAPI
- Unified Memory Space
- Adaptive Page Promotion
- Byte-Granular Data Persistency

Results

PageRank + Twitter

Connected Components + Twitter

Redis + YCSB-D (photo tagging app)

Redis + YCSB-D (social media status update)

Persistency Benefit in Databases

Persistency Benefit in File Systems

Paper Download

ILLINOIS