

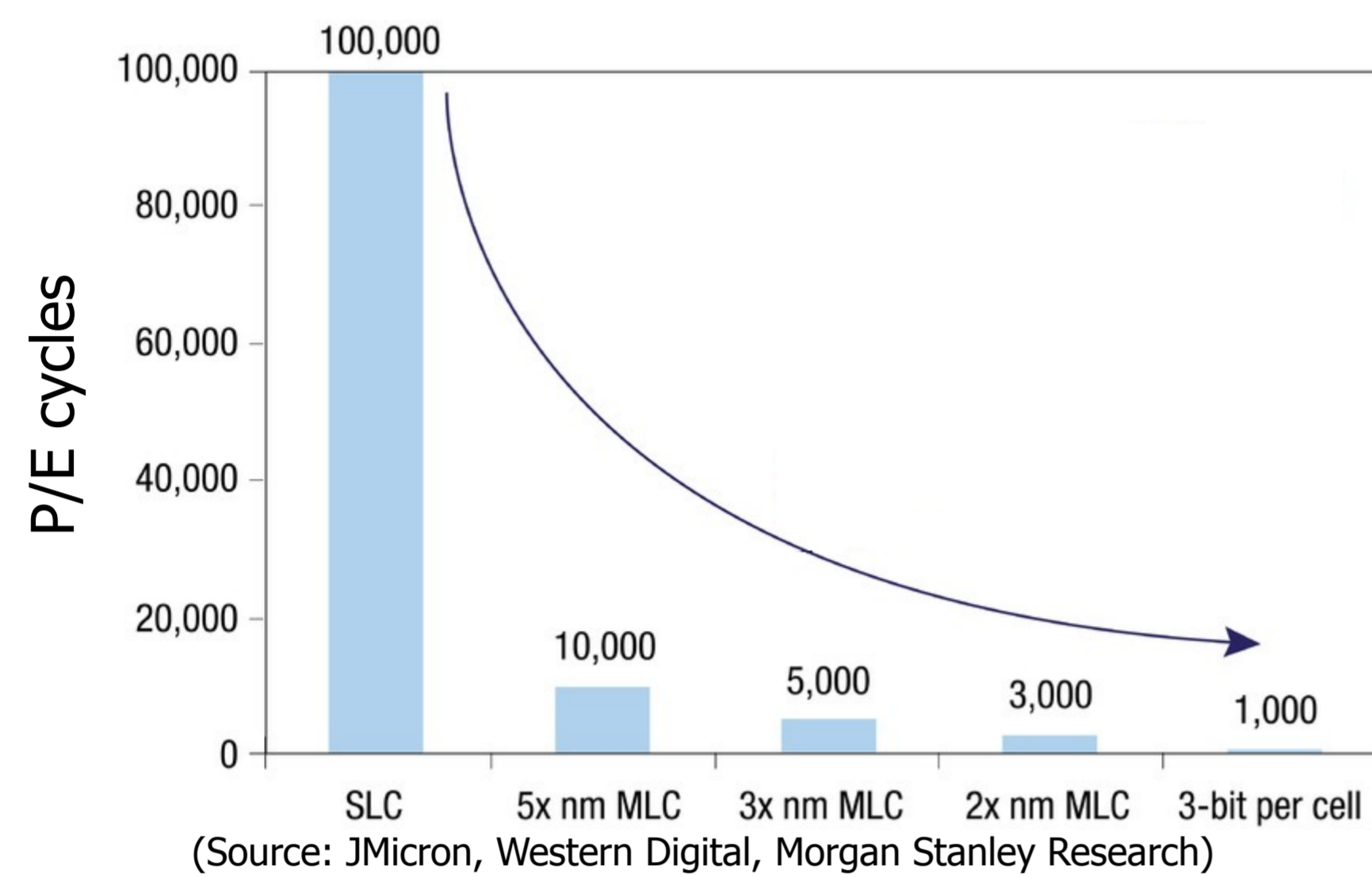


FineDedup: A Fine-Grained Deduplication Technique for Flash-Based SSDs

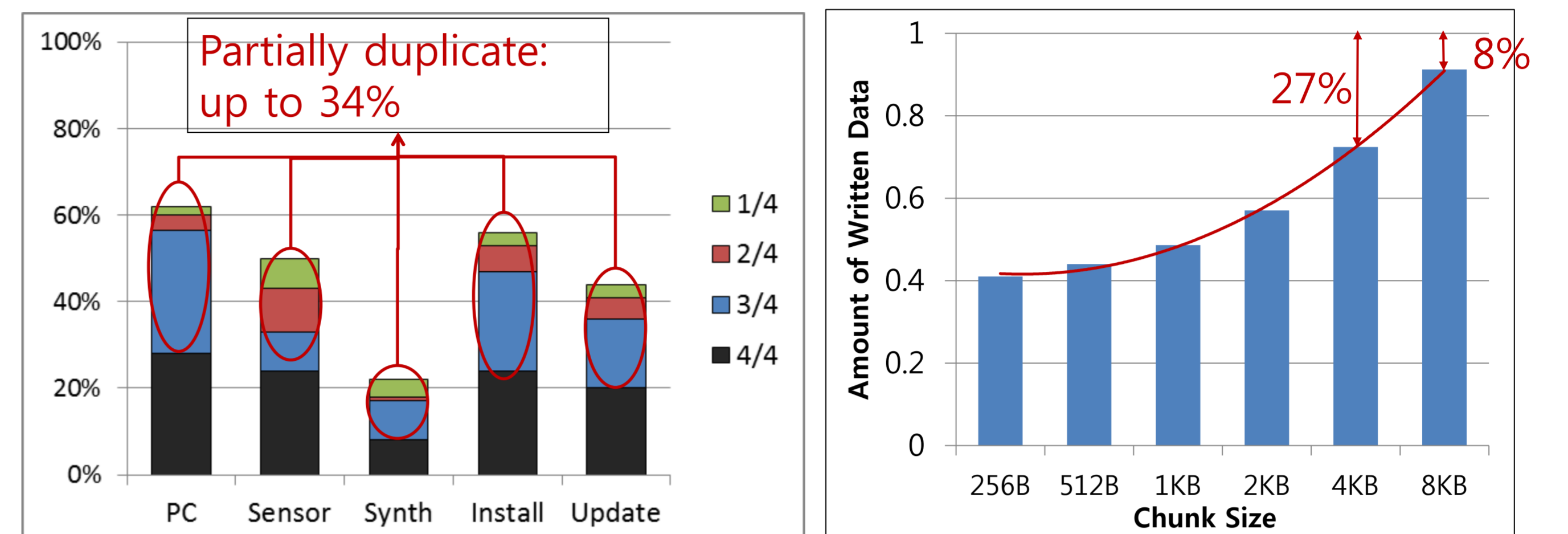
Taejin Kim, Sungjin Lee, and Jihong Kim
Seoul National University

Motivation

- The poor write endurance of SSDs is a main barrier for wider adoption of SSDs in the various environment.
- The endurance of SSDs is rapidly decreasing.
 - 100K P/E cycles (SLC) → 3K P/E cycles (2x nm MLC)



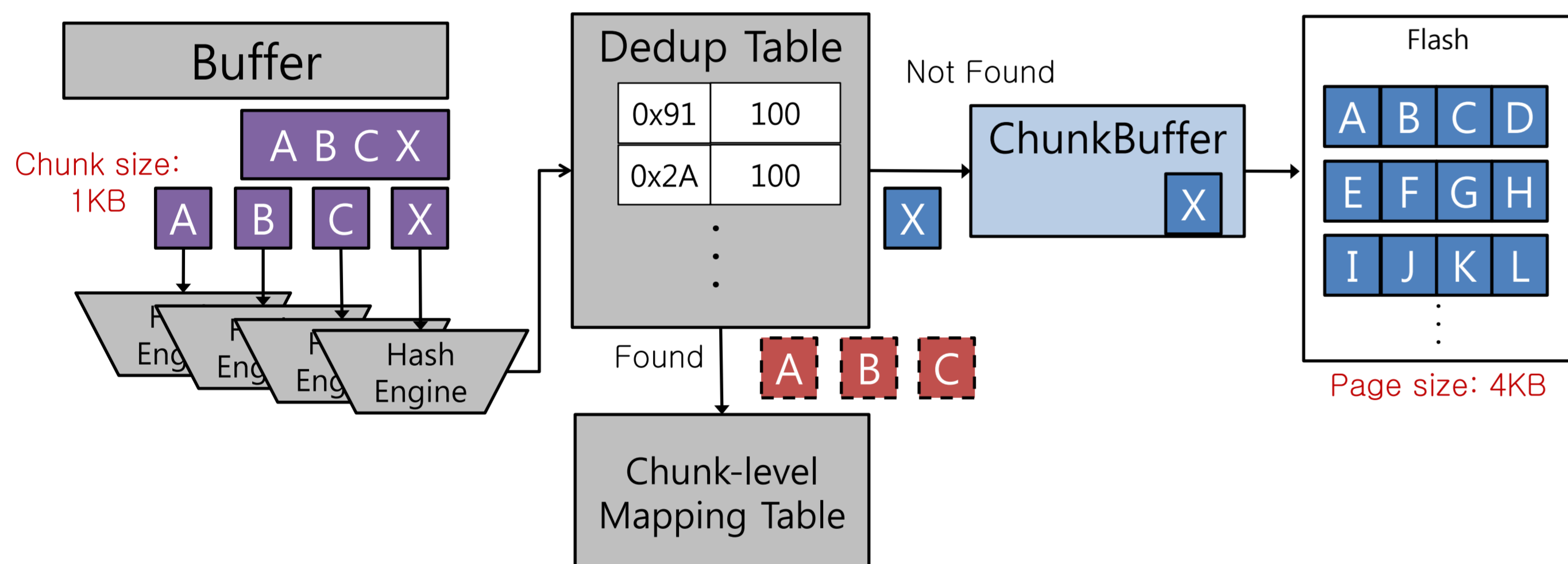
- Existing deduplication techniques
 - Use a fixed-size chunk whose size is equal to a flash page size as a unit of deduplication.



- Potential deduplication opportunities are lost
- Poor dedup ratio for bigger flash page sizes (e.g., 8 KB)

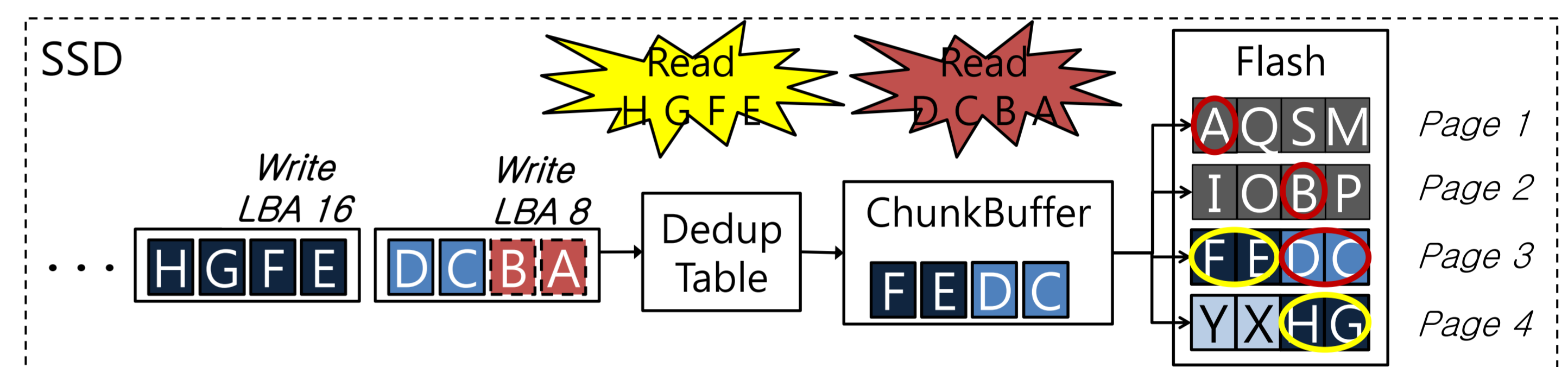
Our Approach

- Uses smaller chunk size than a flash page size
 - Able to eliminate duplicate data within a page chunk
- Exploits multiple hardware hash engines
- Add ChunkBuffer to hold chunks until 4 chunks arrive
- Causes performance and memory overheads
 - Multiple page reads
 - Additional memory space for mapping table



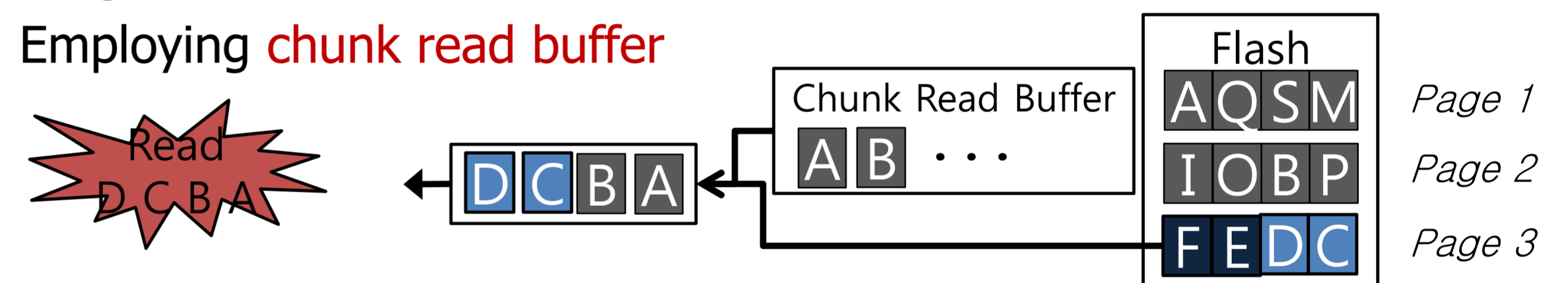
Read Overhead Management

- Read Overhead: due to page fragmentation by smaller chunk

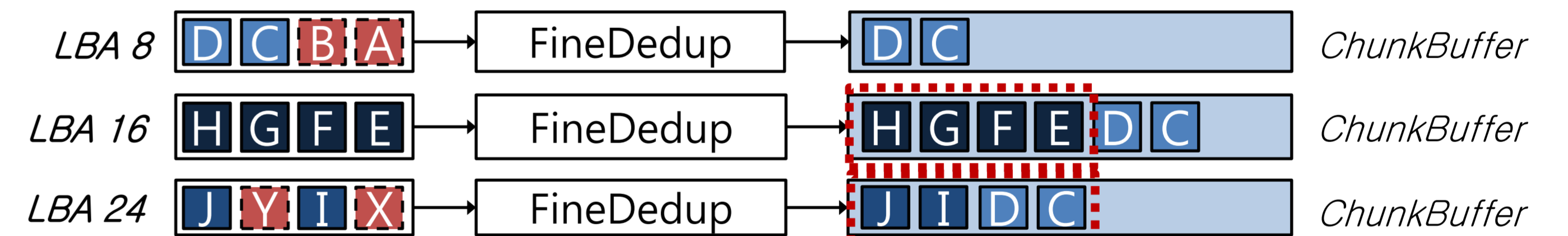


- Management Scheme

- Employing chunk read buffer

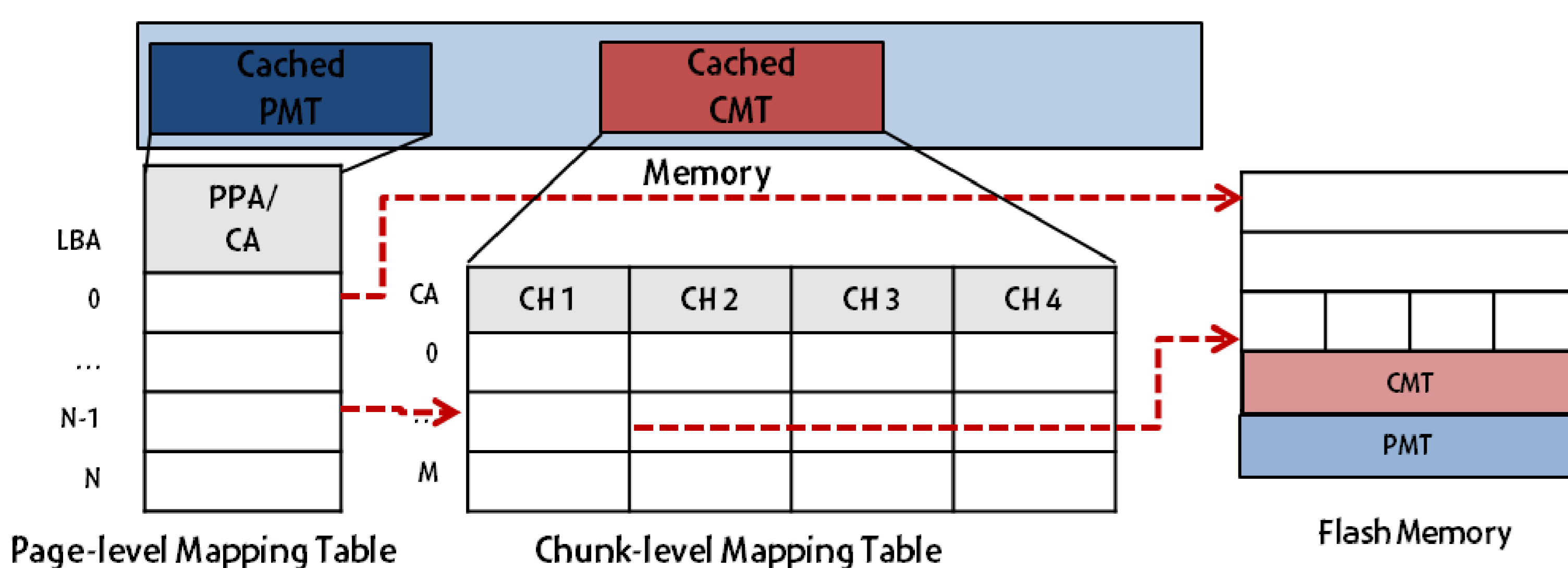


- Chunk packing scheme in CHunkBuffer



Memory Overhead Management

- Memory Overhead: due to finer-granularity of mapping
 - Chunk-level mapping requires 4x more memory than page-level
- Management Scheme
 - Employing hybrid mapping table: separate duplicate pages from pages with no duplicate chunks
 - Adopting demand-based strategy: in case of workload requiring large CMT



Results

- Evaluate the effectiveness of FineDedup in reducing duplicate data and overheads using real-world workloads

