Unified Address Translation for Memory Mapped SSDs with FlashMap

Jian Huang

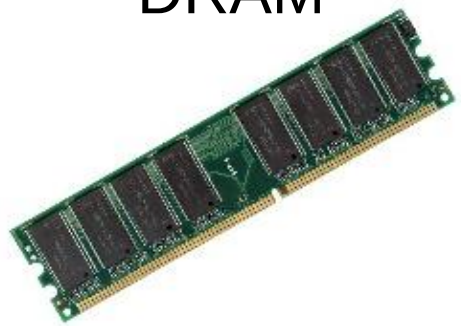
Anirudh Badam† Moinuddin K. Qureshi Karsten Schwan

Georgia Tech

† Microsoft

Bridging the DRAM-Disk Gap

DRAM



High Performance
Small Capacity

Application Memory Component

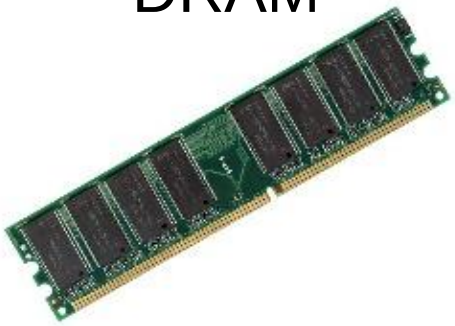
Disk

Low Performance
Large Capacity

Application Storage Component

Bridging the DRAM-Disk Gap

DRAM



High Performance
Small Capacity

SSD

Good Performance
Good Capacity

Disk

Low Performance
Large Capacity

Application Memory Component

Application Storage Component

Flash: Slow Memory or Fast Disk?

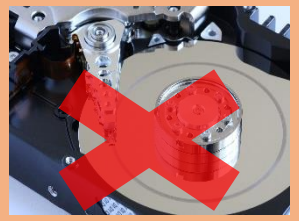
Flash behaves more like memory than disk

Flash: Slow Memory or Fast Disk?

Flash behaves more like memory than disk

Flash: Slow Memory or Fast Disk?

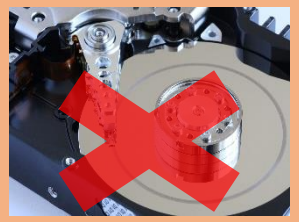
Flash behaves more like memory than disk



No Seek Latency

Flash: Slow Memory or Fast Disk?

Flash behaves more like memory than disk



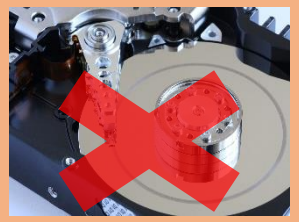
No Seek Latency

+

Internal Parallelism

Flash: Slow Memory or Fast Disk?

Flash behaves more like memory than disk



No Seek Latency

+

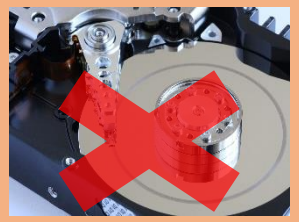
Internal Parallelism

+

High IOPS

Flash: Slow Memory or Fast Disk?

Flash behaves more like memory than disk



No Seek Latency

+

Internal Parallelism

+

High IOPS

Use Flash as Memory [Badam et al., NSDI'11]

DRAM

High Performance
Small Capacity

SSD

Good Performance
Good Capacity

Disk

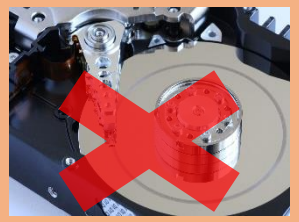
Low Performance
Large Capacity

Application Memory Component

Application Storage Component

Flash: Slow Memory or Fast Disk?

Flash behaves more like memory than disk



No Seek Latency

+

Internal Parallelism

+

High IOPS

Use Flash as Memory [Badam et al., NSDI'11]

DRAM

High Performance
Small Capacity

SSD

Good Performance
Good Capacity

Disk

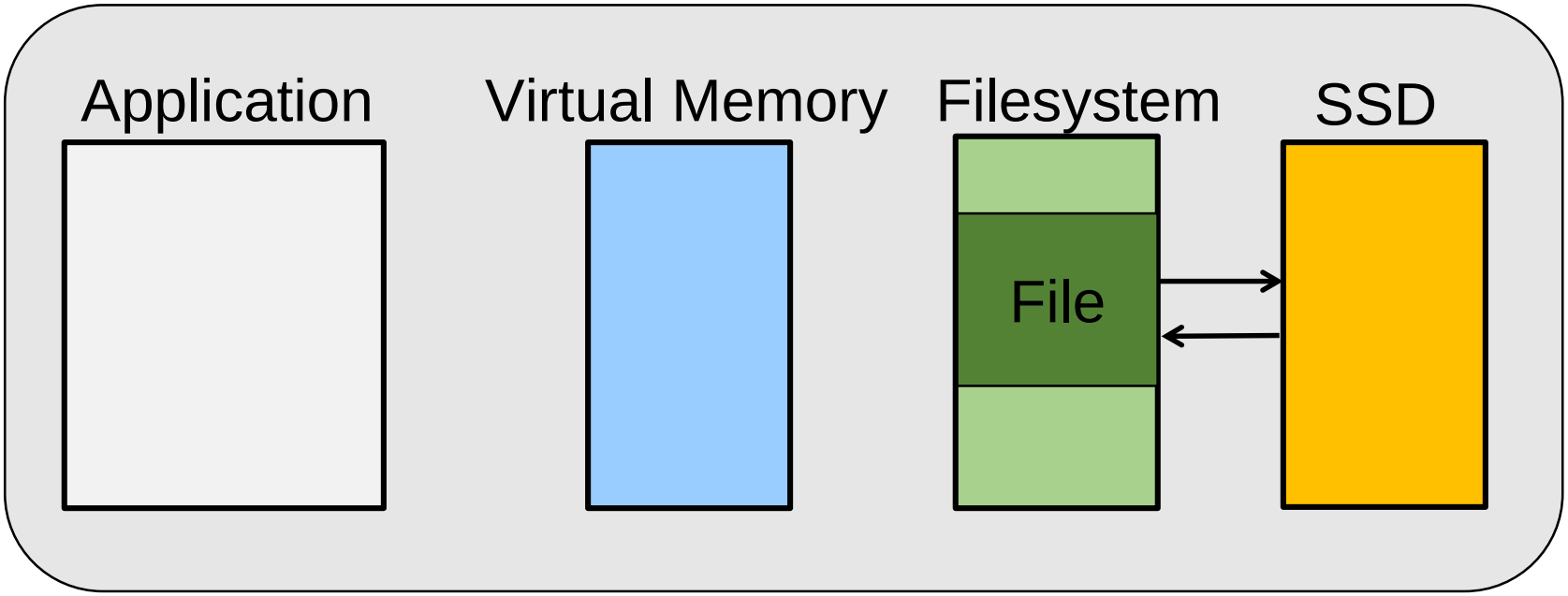
Low Performance
Large Capacity

Application Memory Component

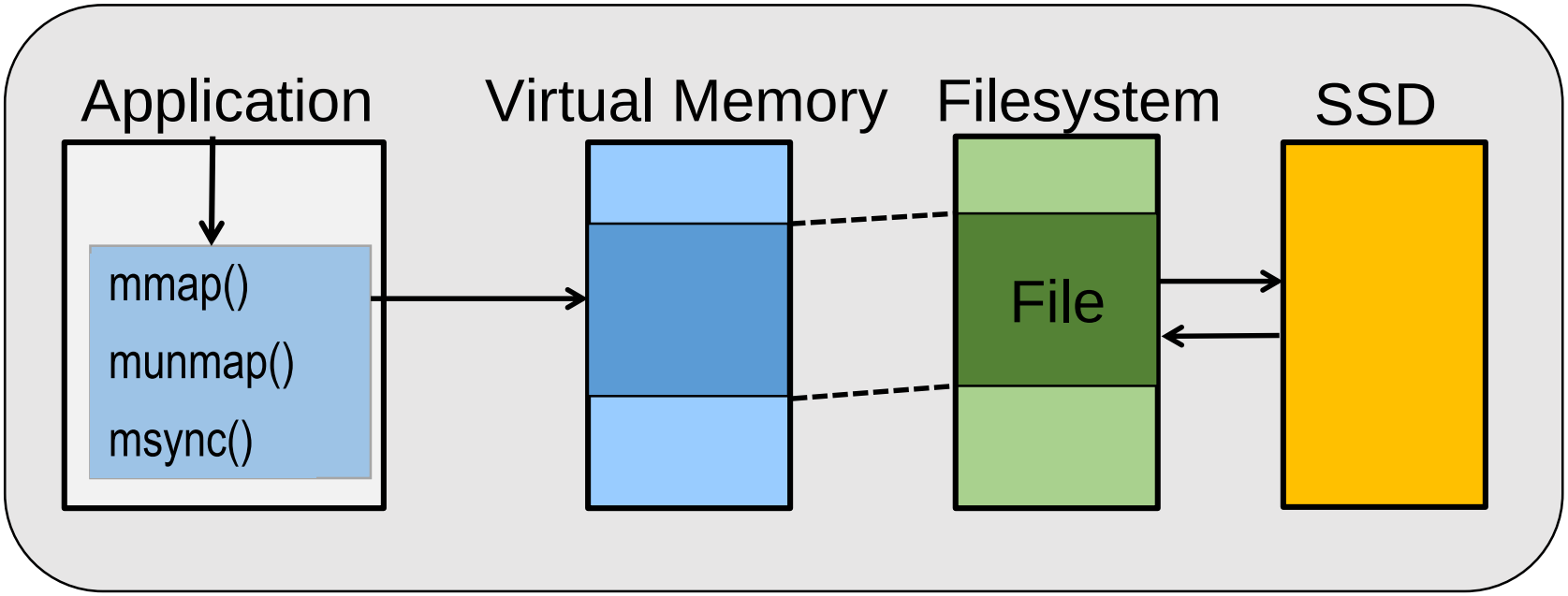
Application Storage Component

Memory Mapped SSDs

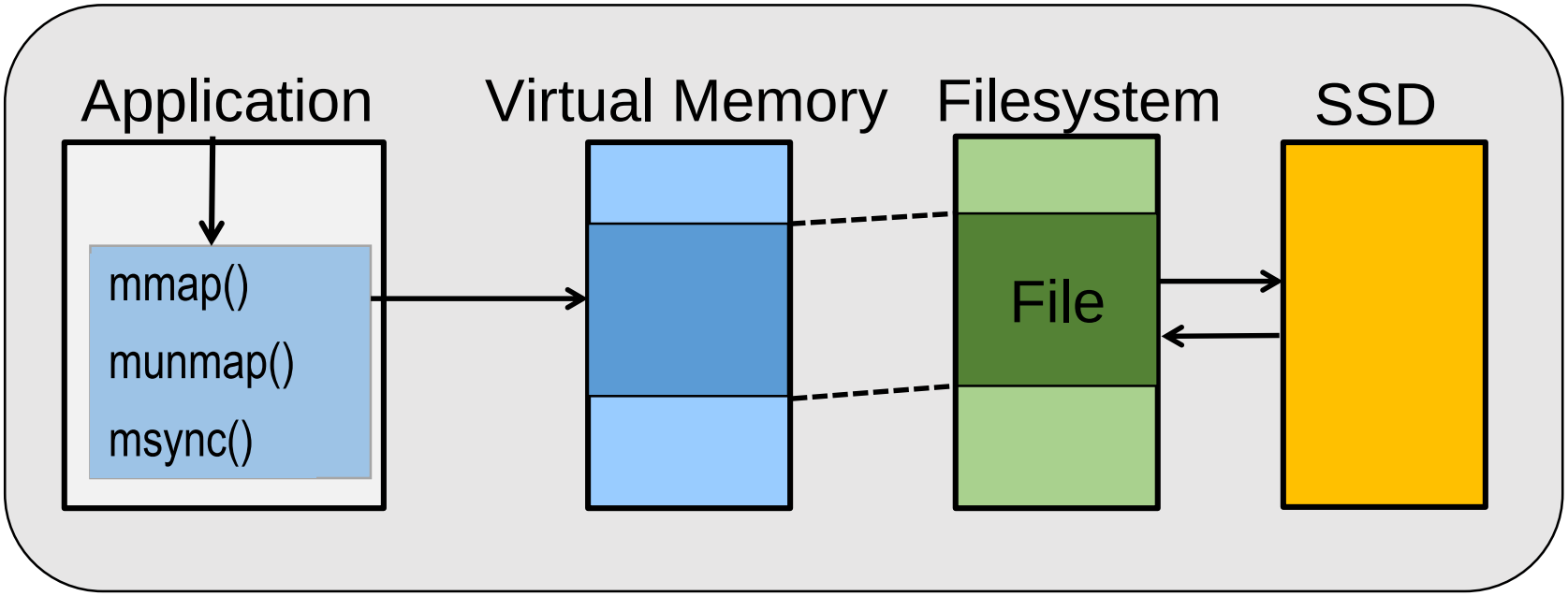
Memory Mapped SSDs



Memory Mapped SSDs

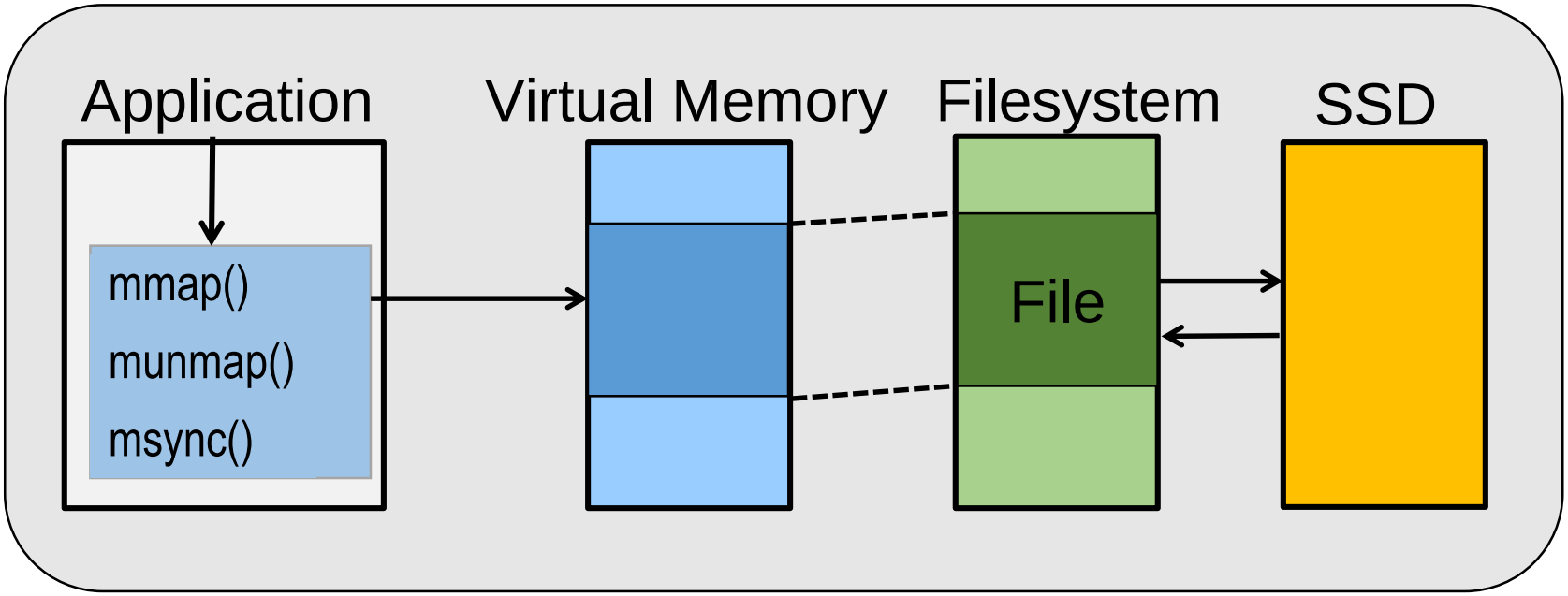


Memory Mapped SSDs



Extended Memory

Memory Mapped SSDs

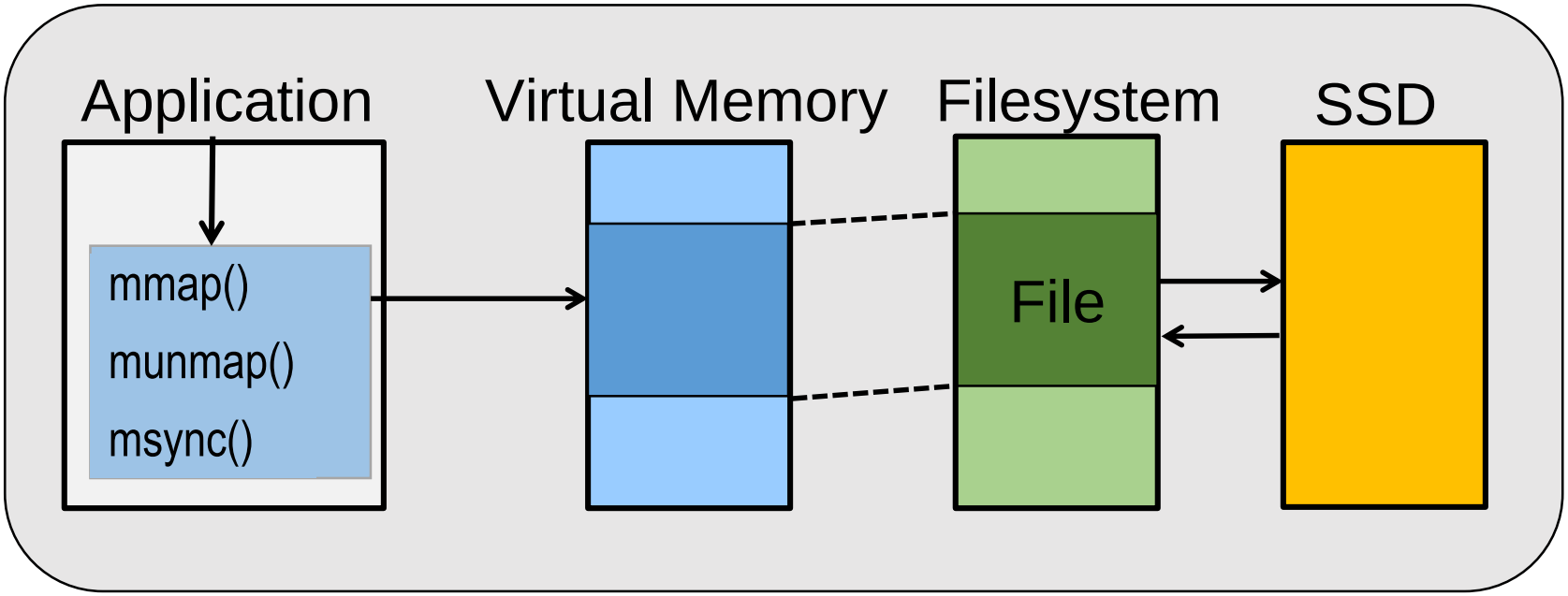


+

Extended Memory

Minimal Code Modifications

Memory Mapped SSDs

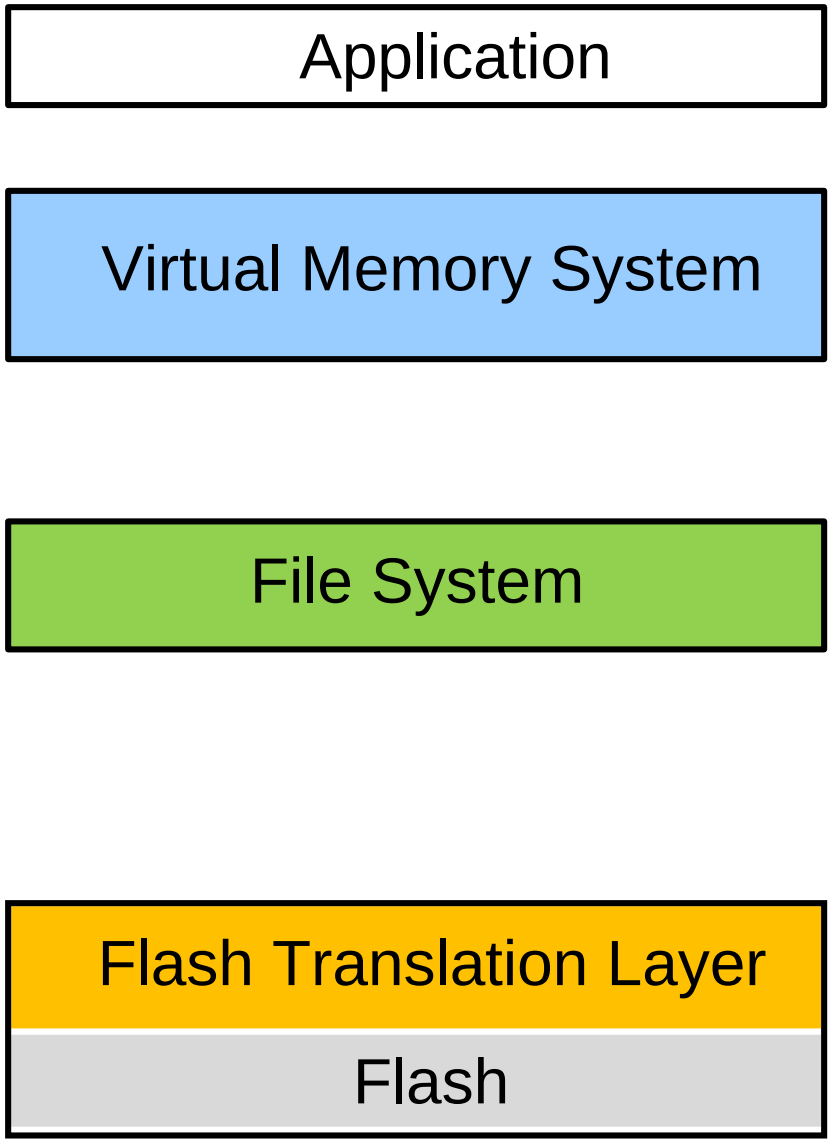


Extended Memory

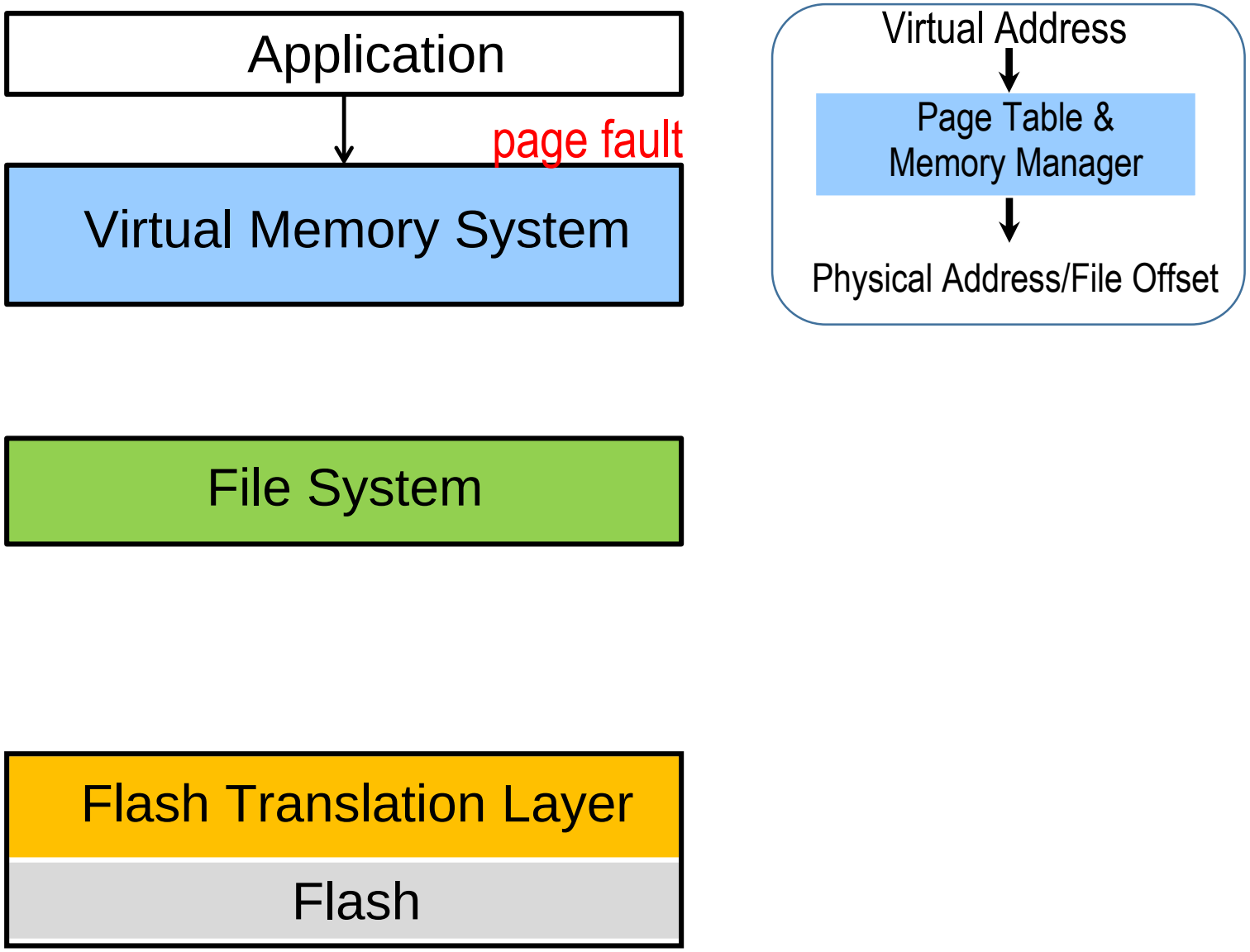
Minimal Code Modifications

Data Durability

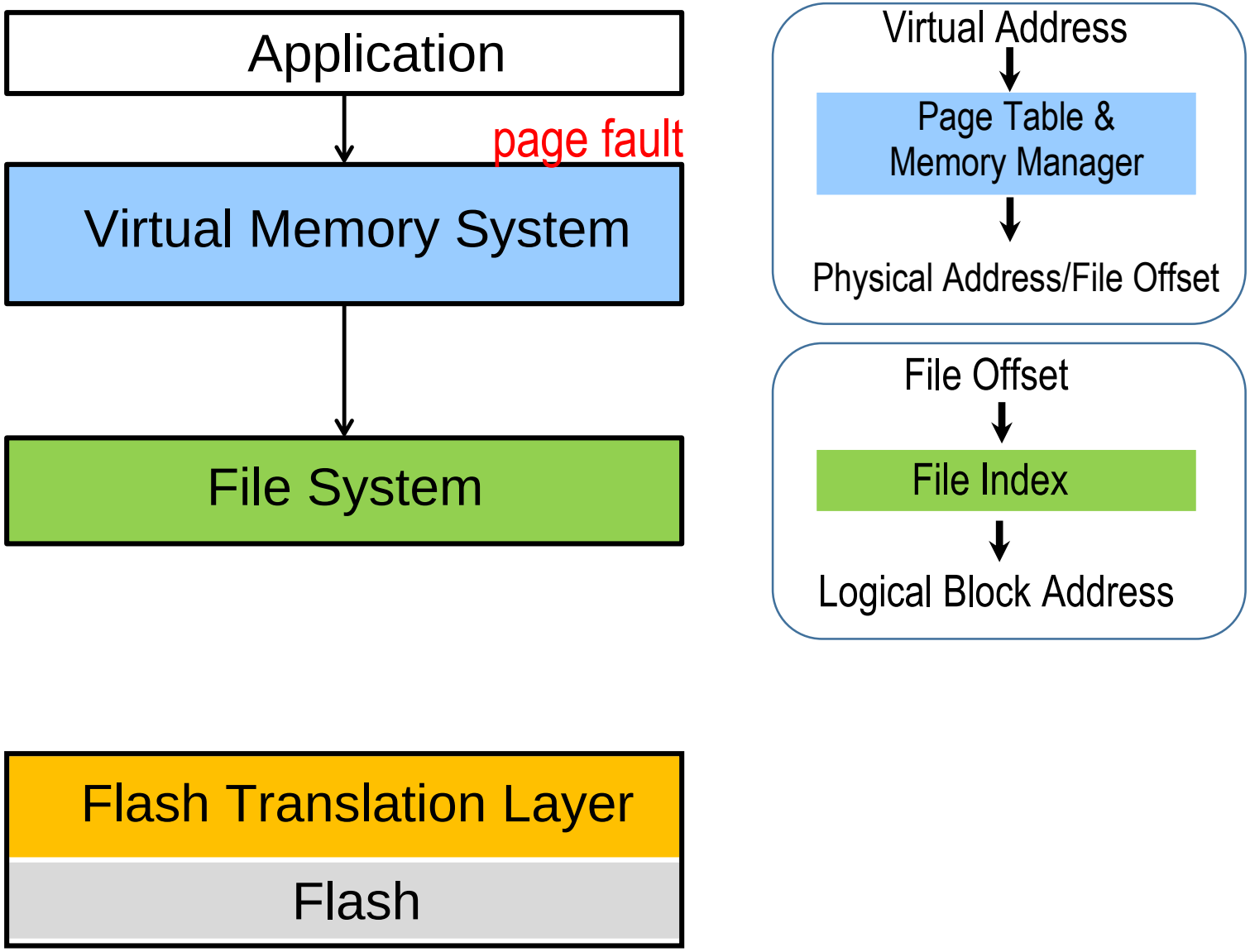
No Free Lunch: Software Overhead



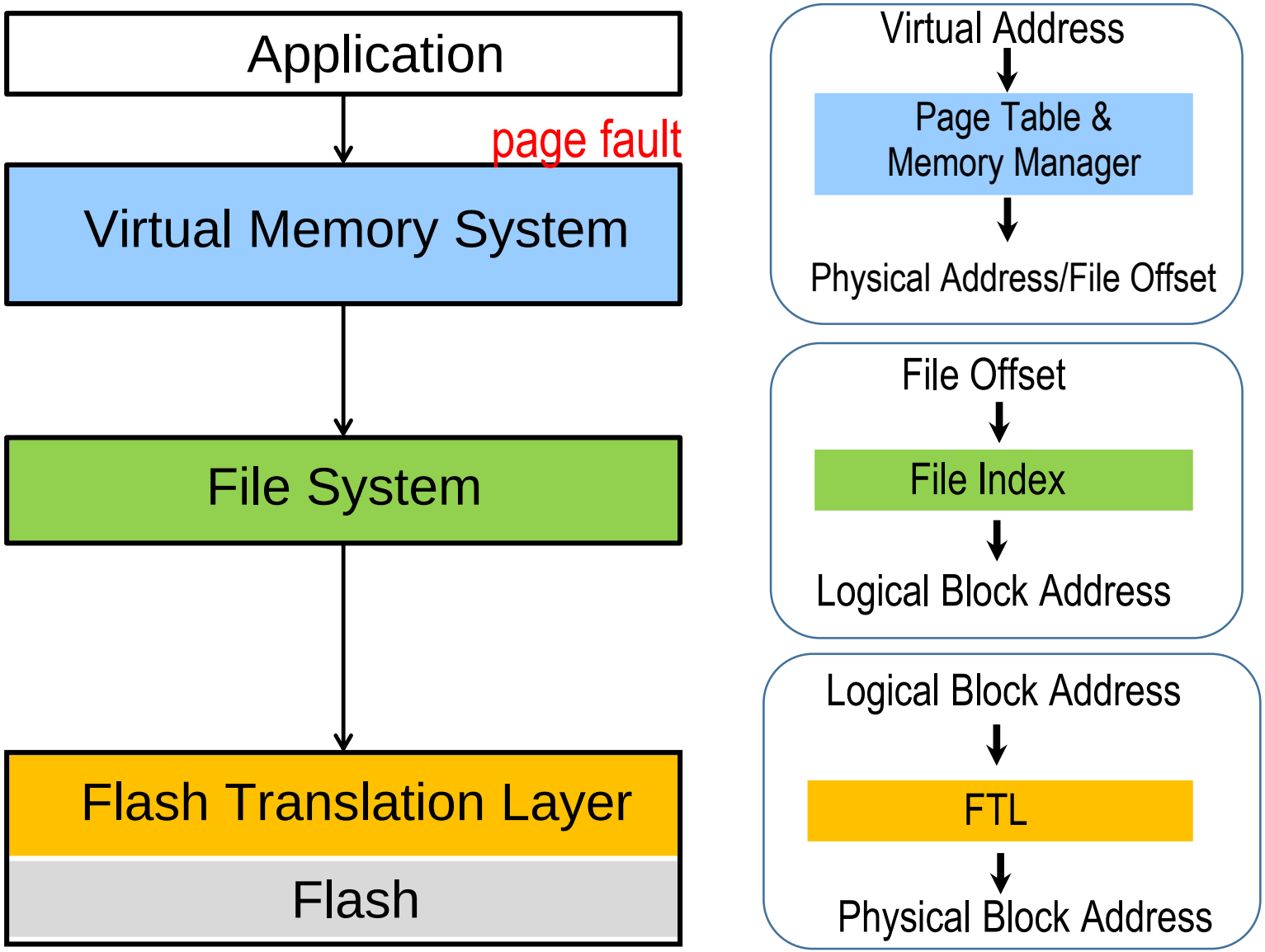
No Free Lunch: Software Overhead



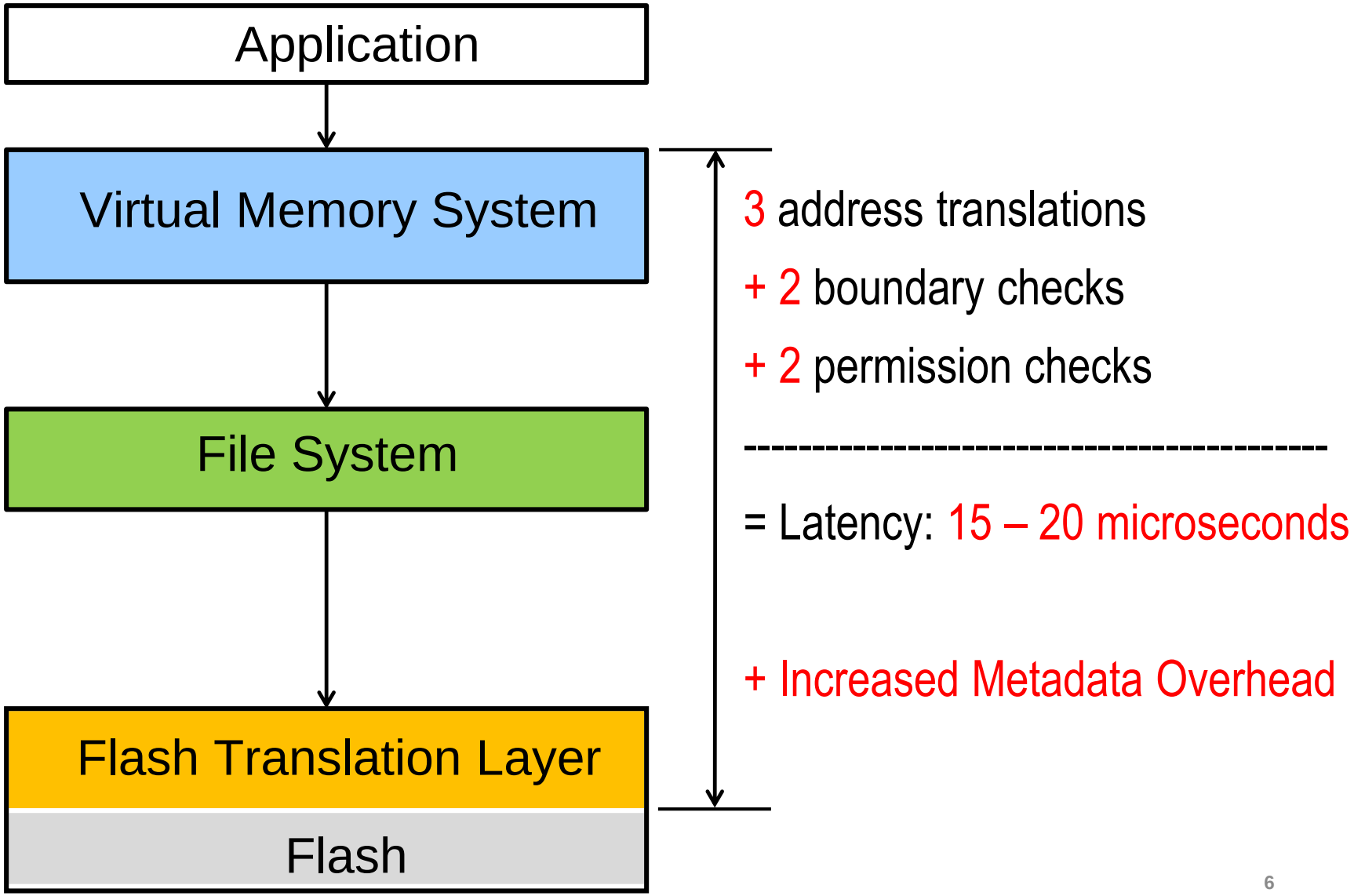
No Free Lunch: Software Overhead



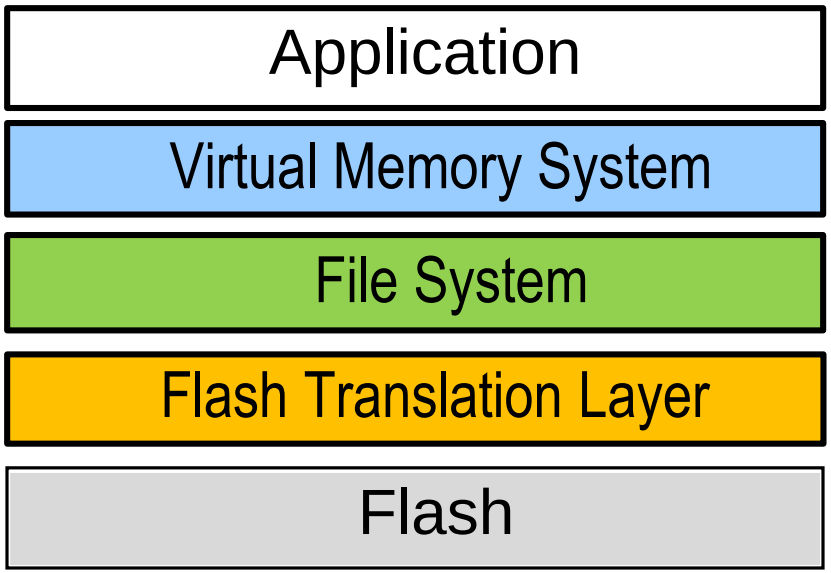
No Free Lunch: Software Overhead



Software Overhead Quantified



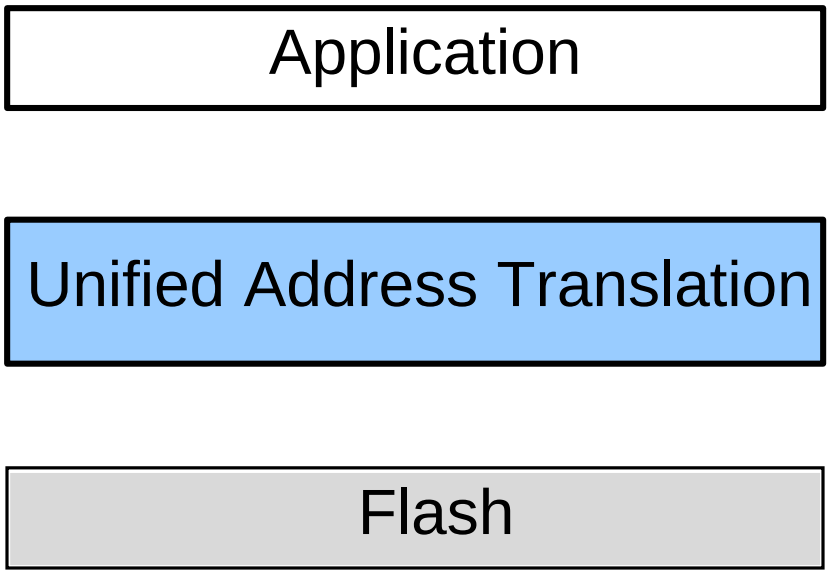
FlashMap: Unified Address Translation



1 Reduced Storage, only **1 mapping table**

2 Reduced Latency, only **1 address translation + 1 permission check +1 boundary check**

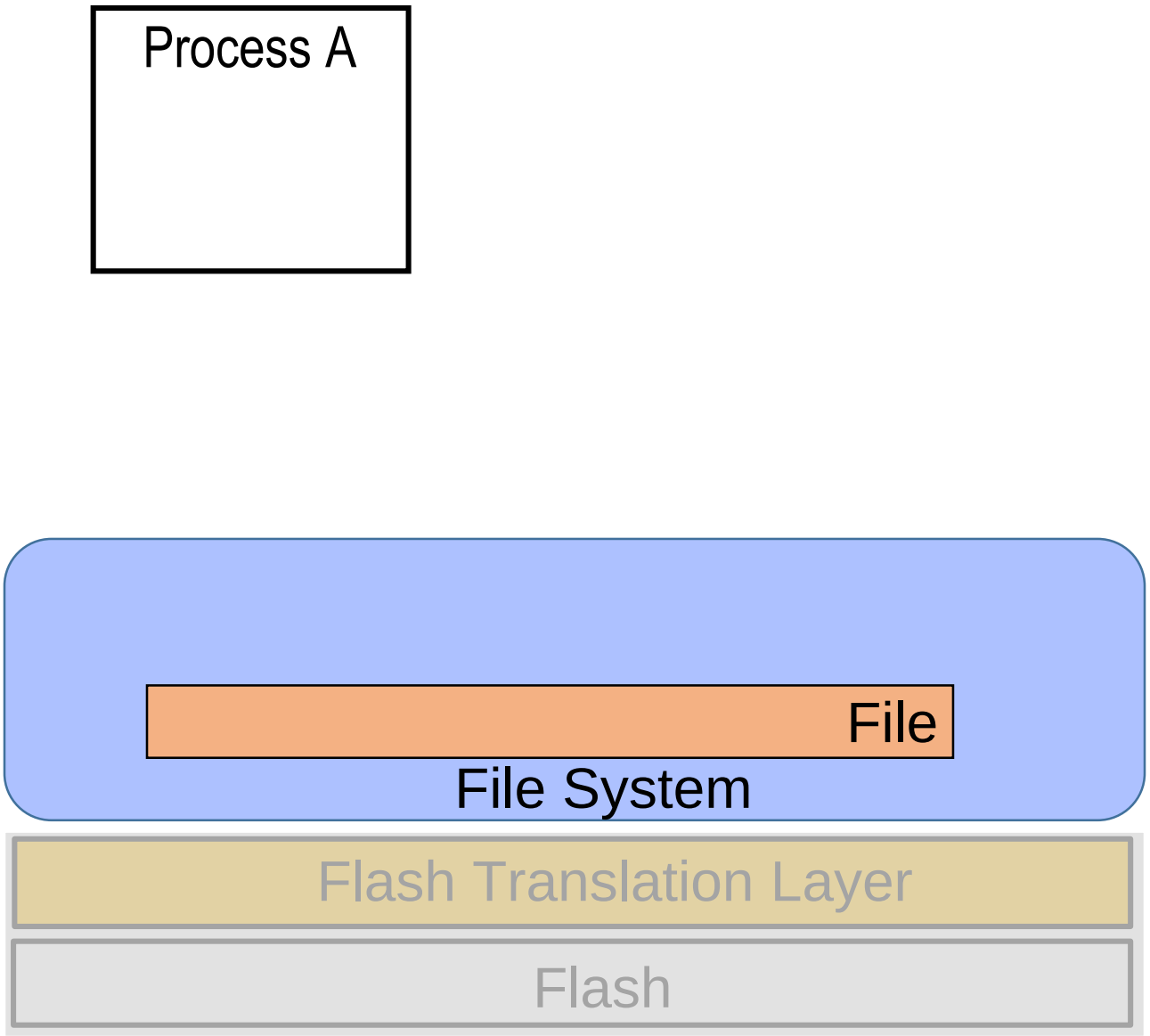
FlashMap: Unified Address Translation



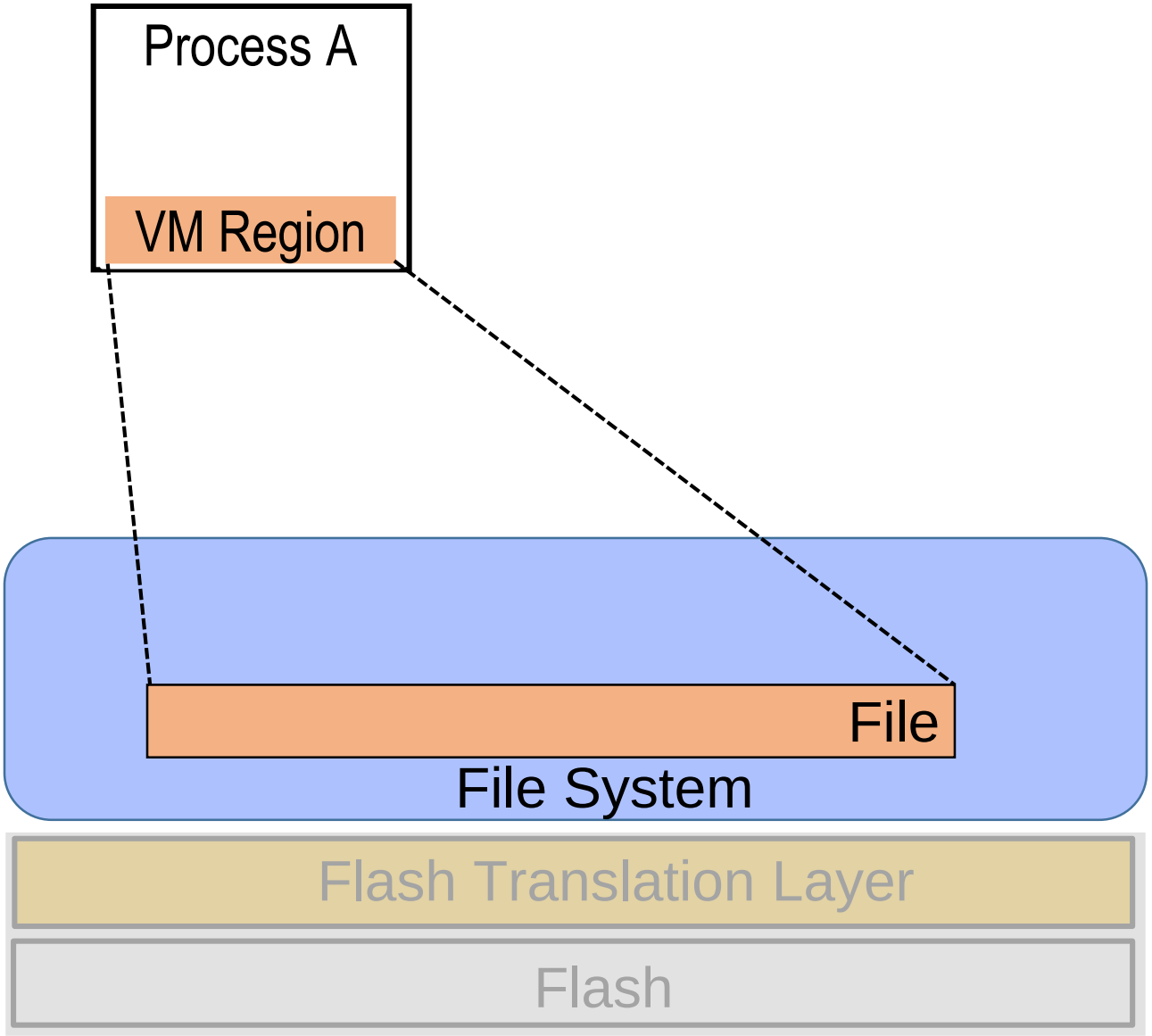
1 Reduced Storage, only **1 mapping table**

2 Reduced Latency, only **1 address translation + 1 permission check +1 boundary check**

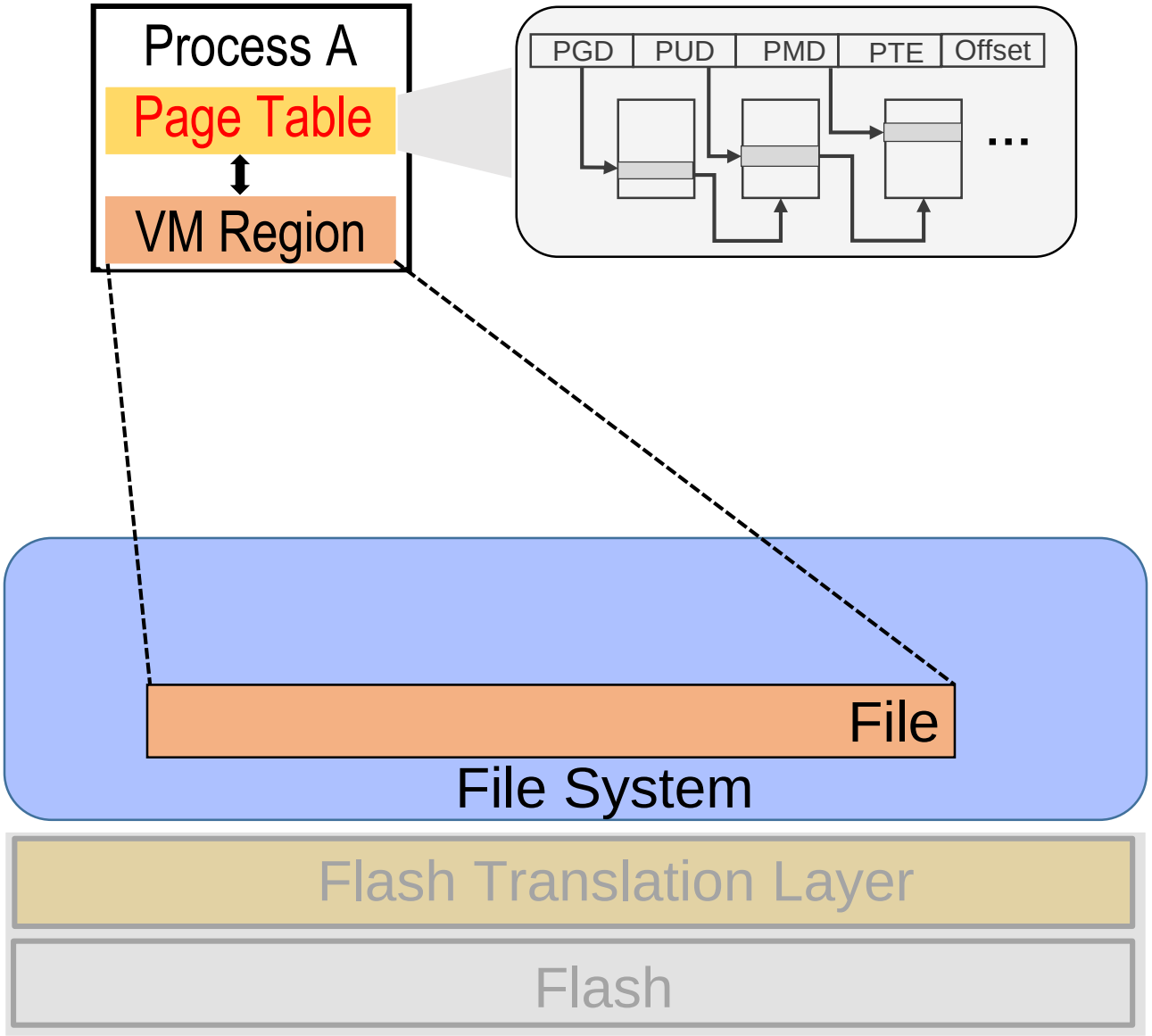
Combining Page Table and File System



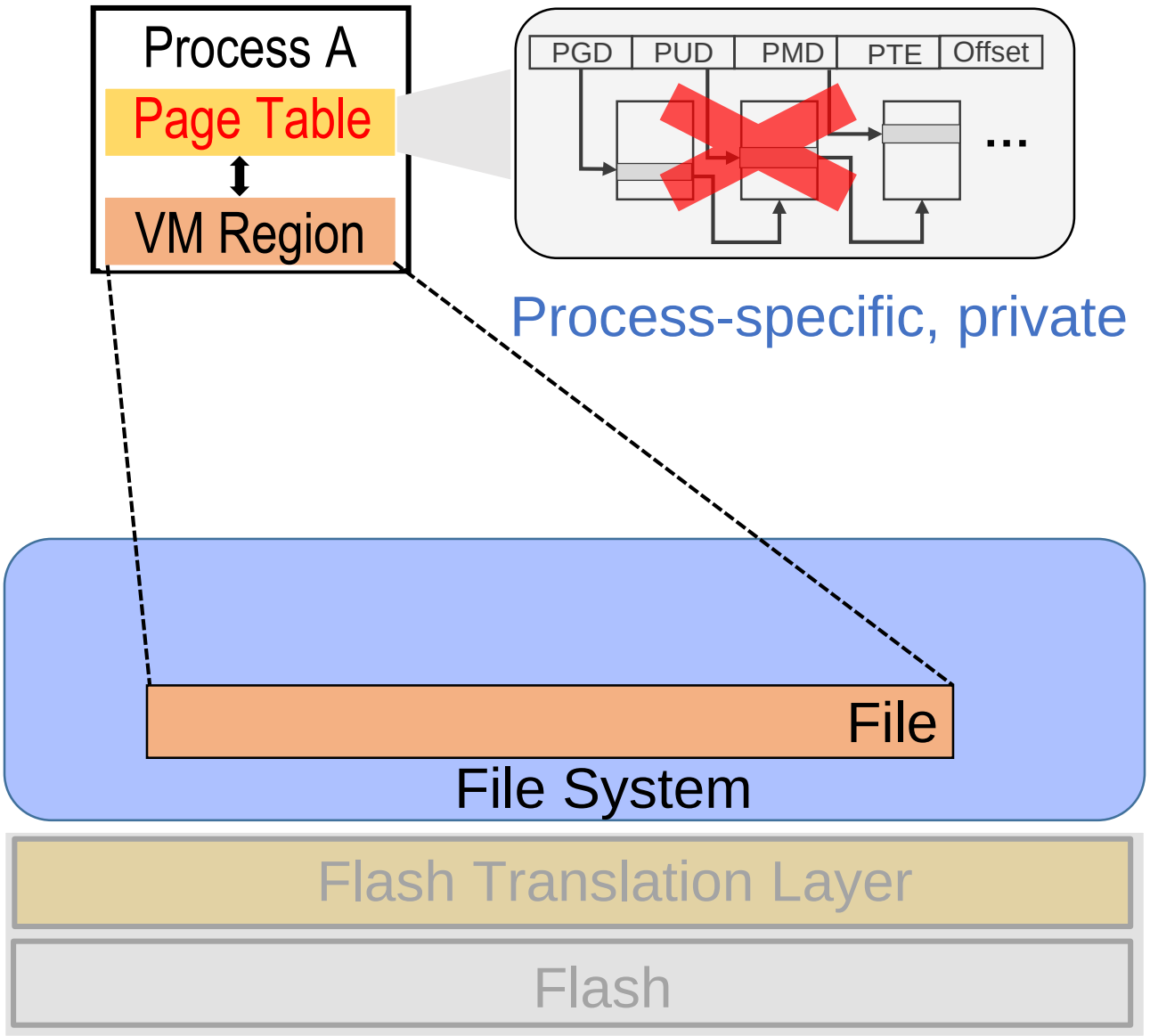
Combining Page Table and File System



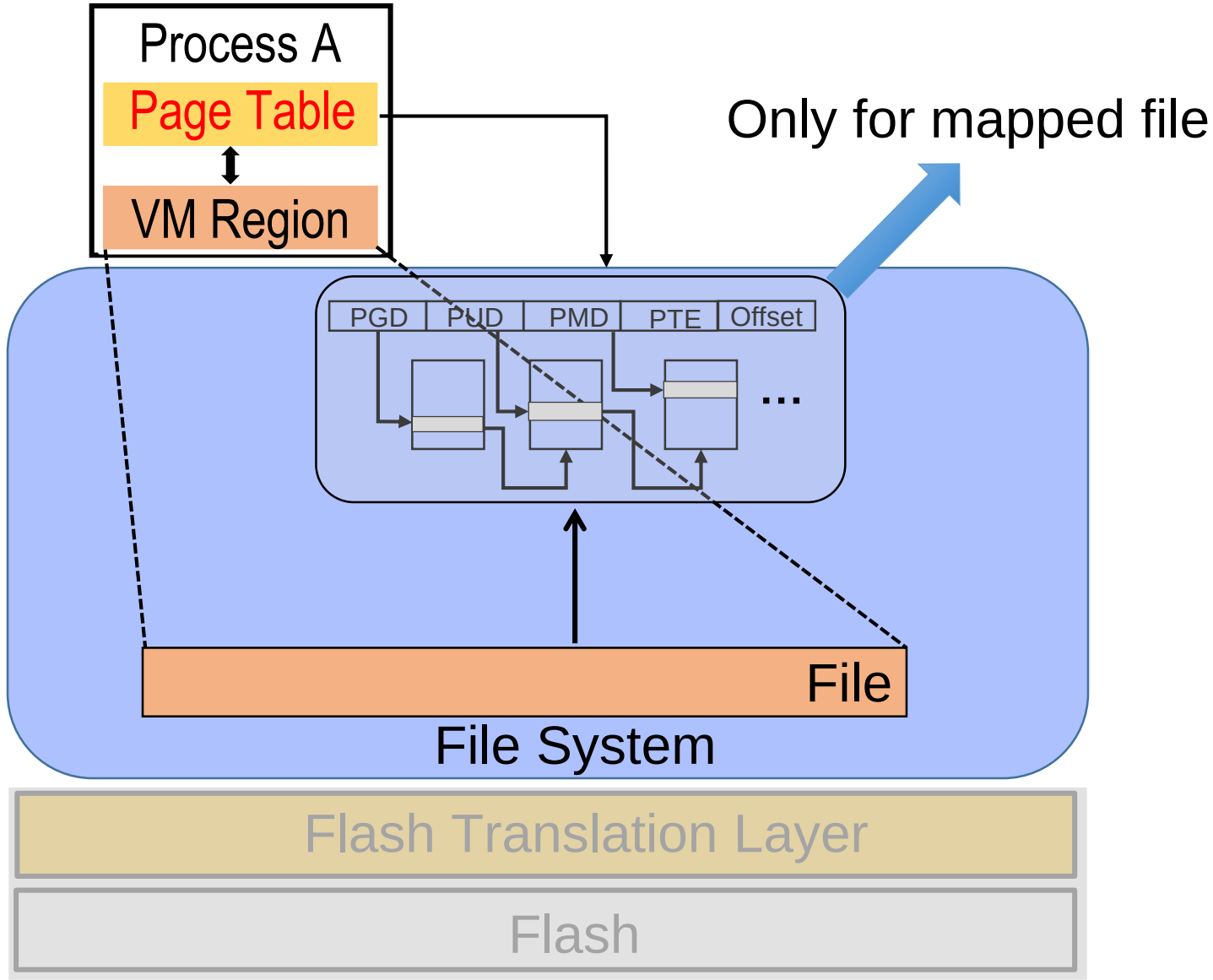
Combining Page Table and File System



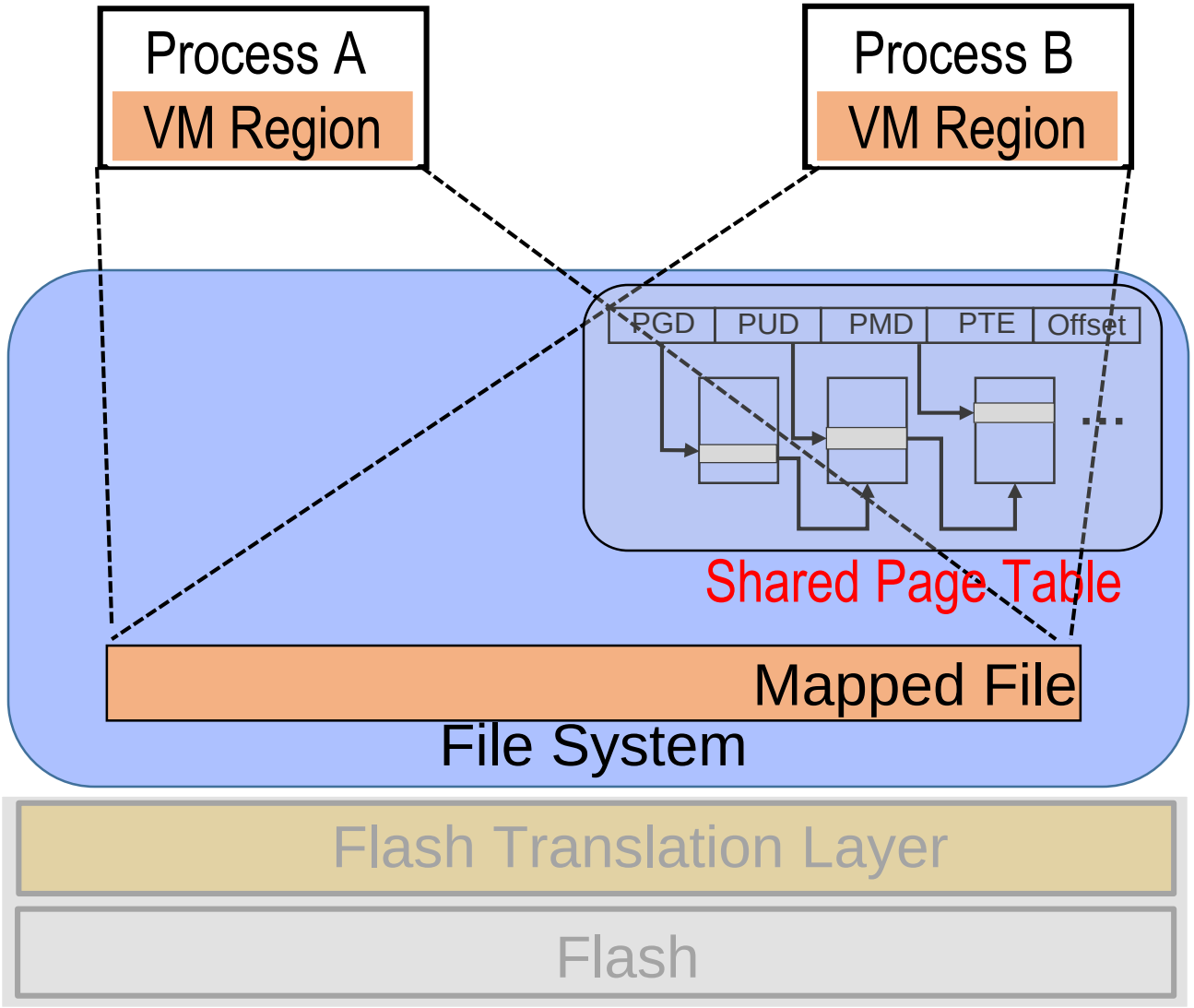
Combining Page Table and File System



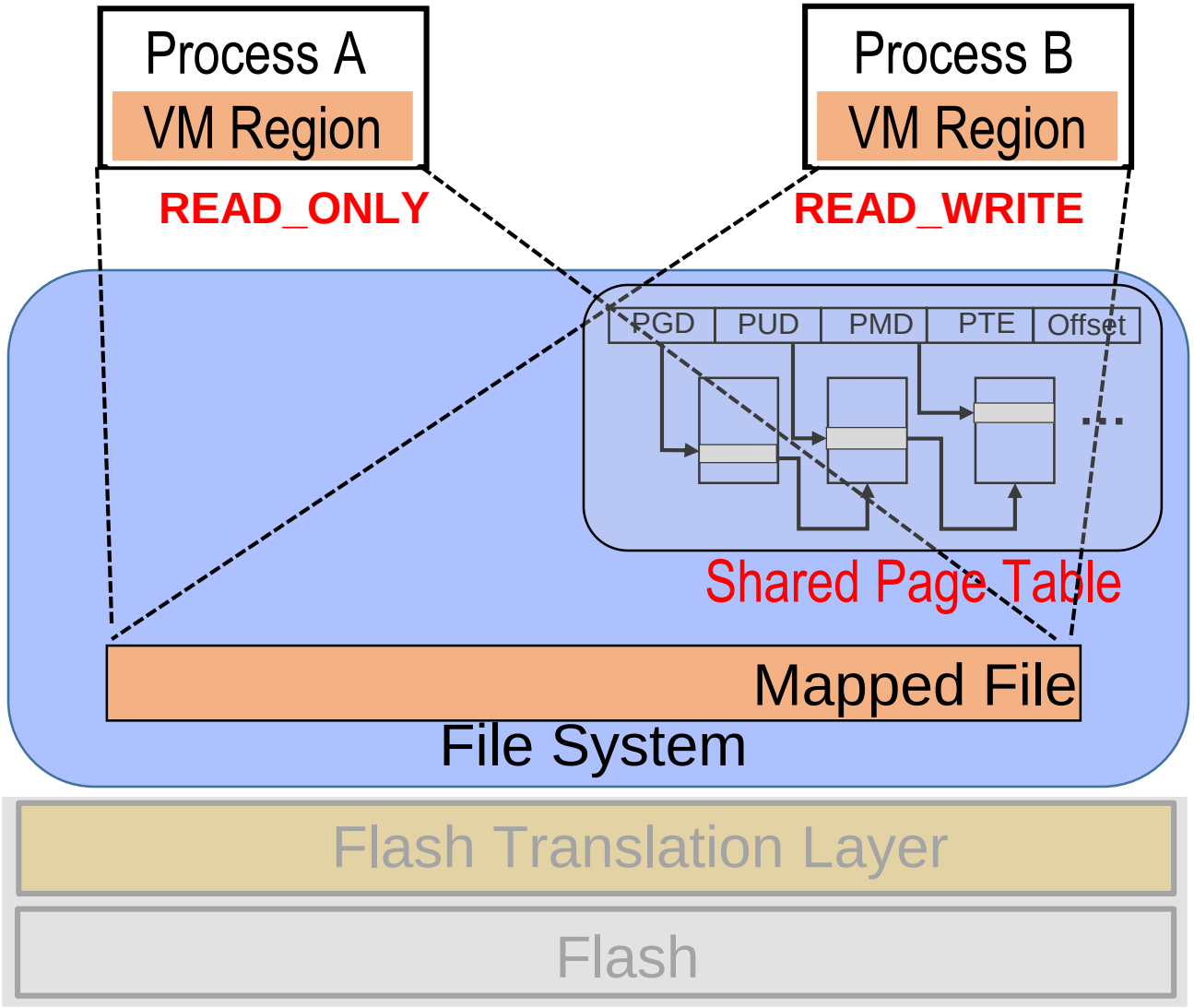
Combining Page Table and File System



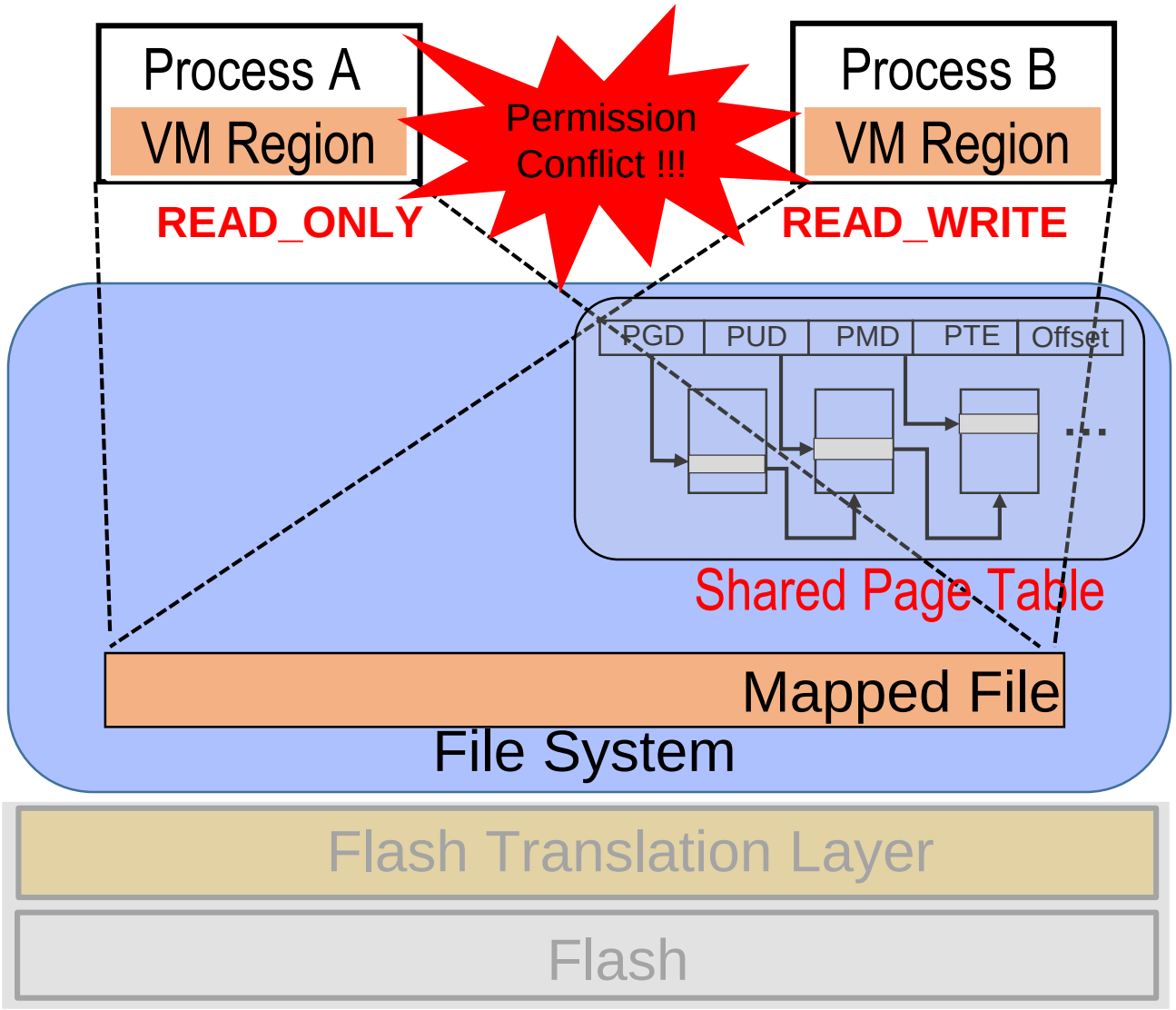
Preserving File System Permissions



Preserving File System Permissions

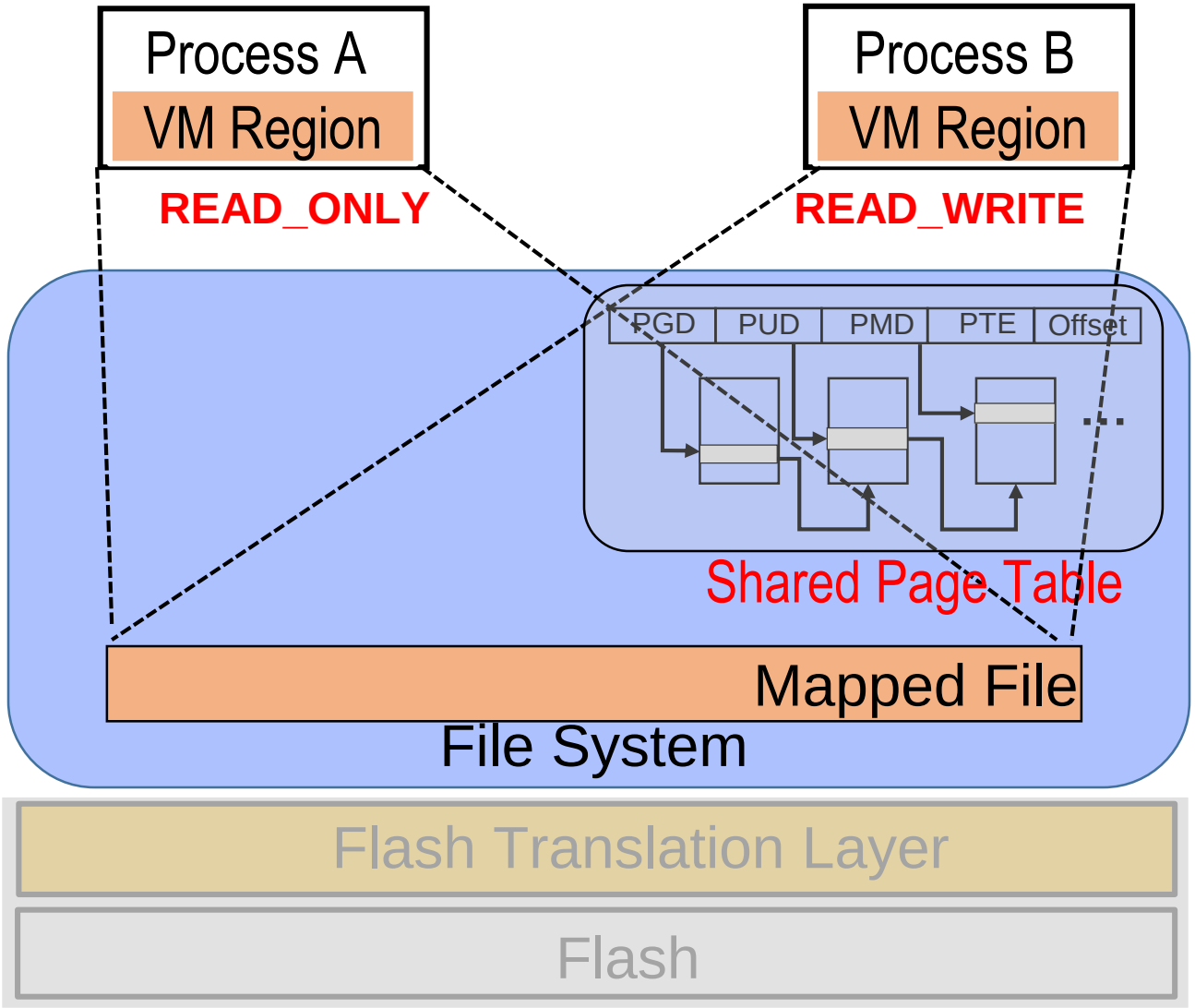


Preserving File System Permissions



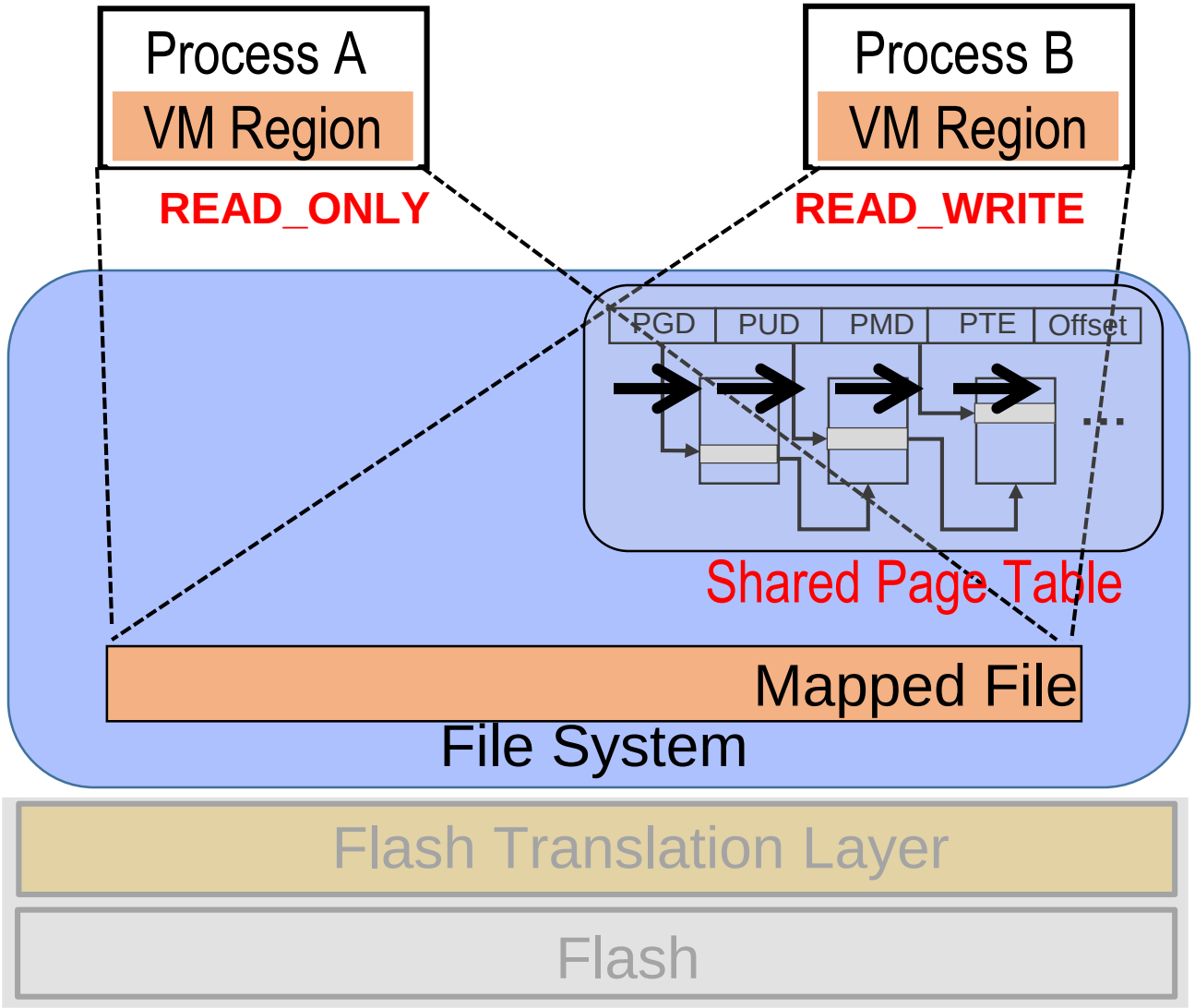
Preserving File System Permissions

Only share the leaf-level page table pages !

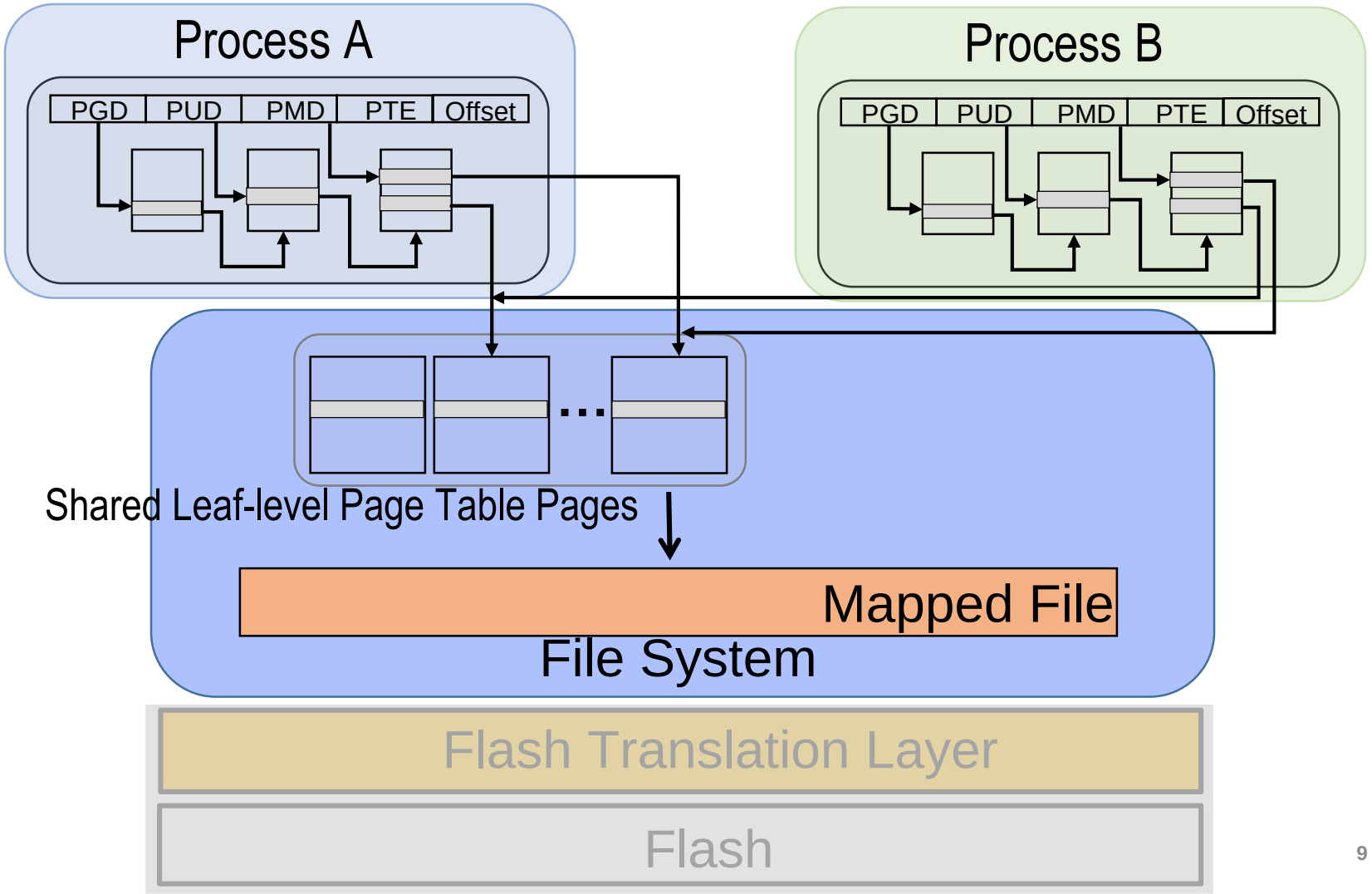


Preserving File System Permissions

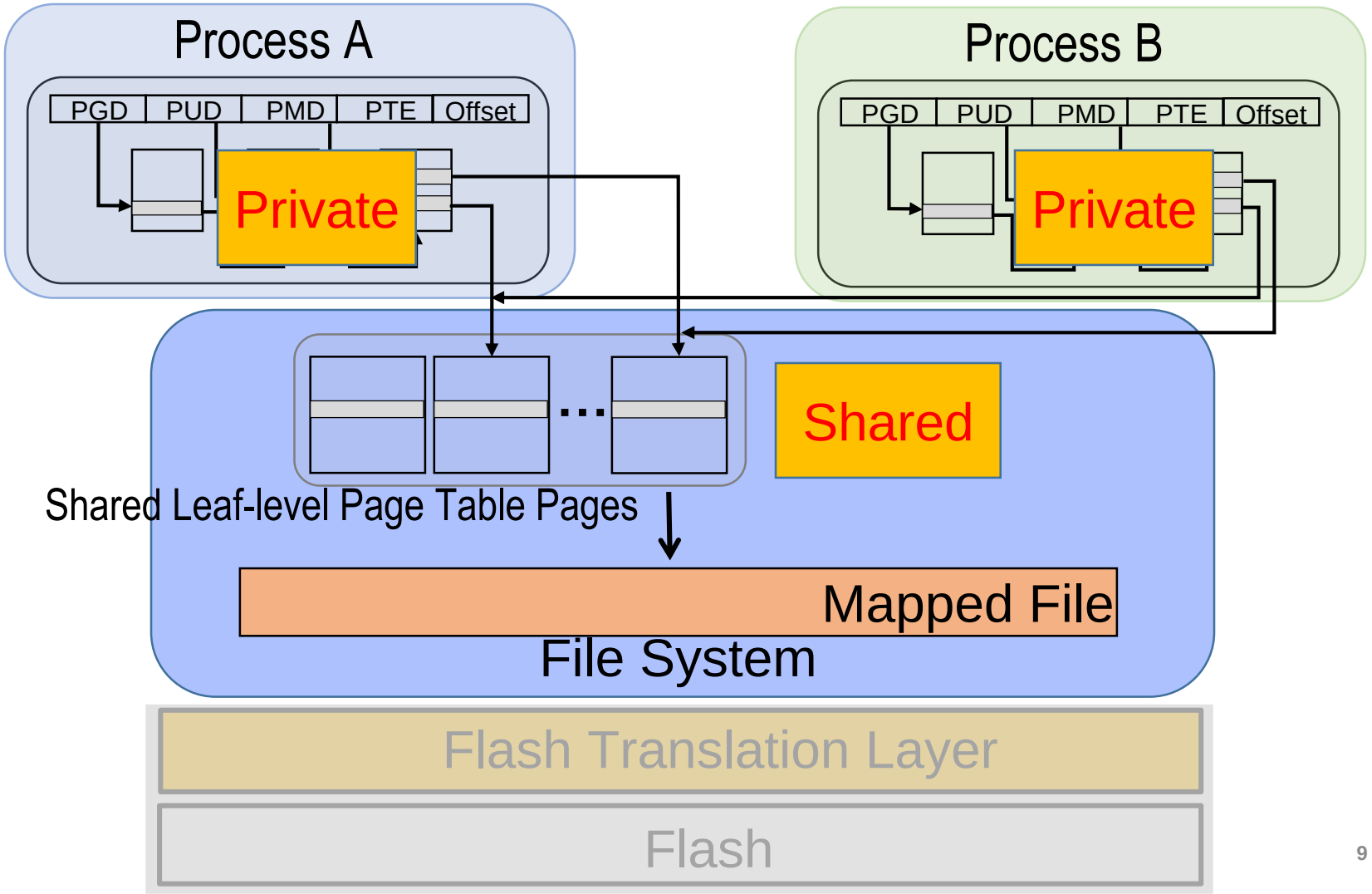
Only share the leaf-level page table pages !



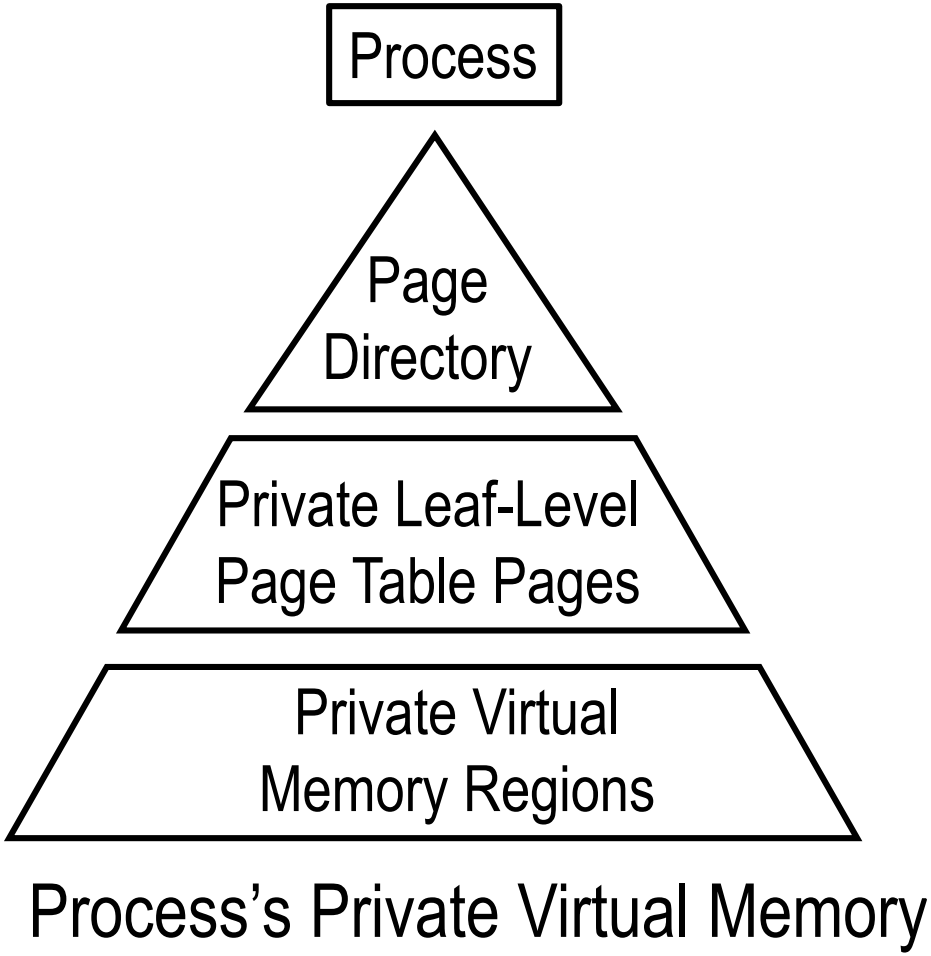
Preserving File System Permissions



Preserving File System Permissions

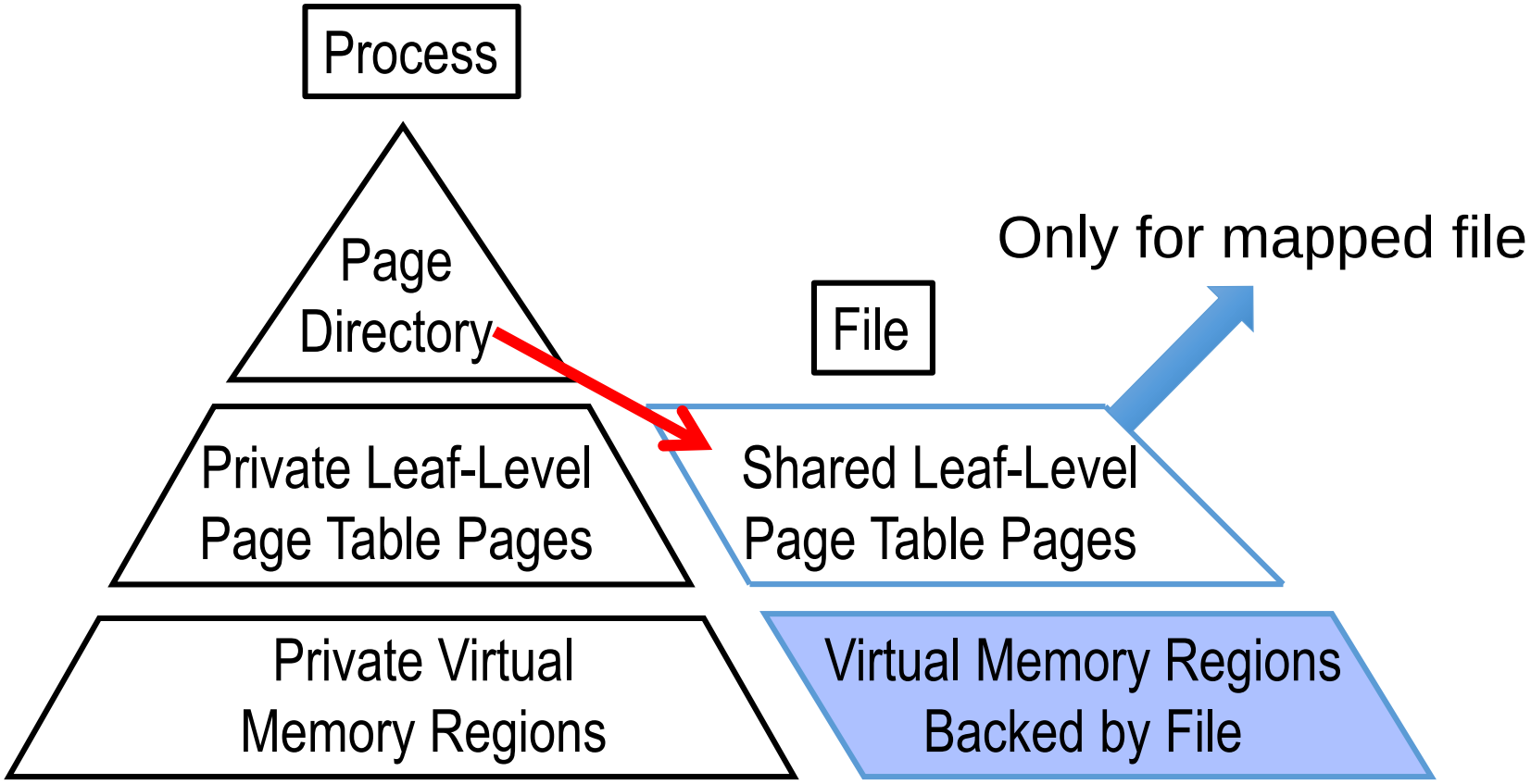


Page Table in FlashMap



Before Mapping a File

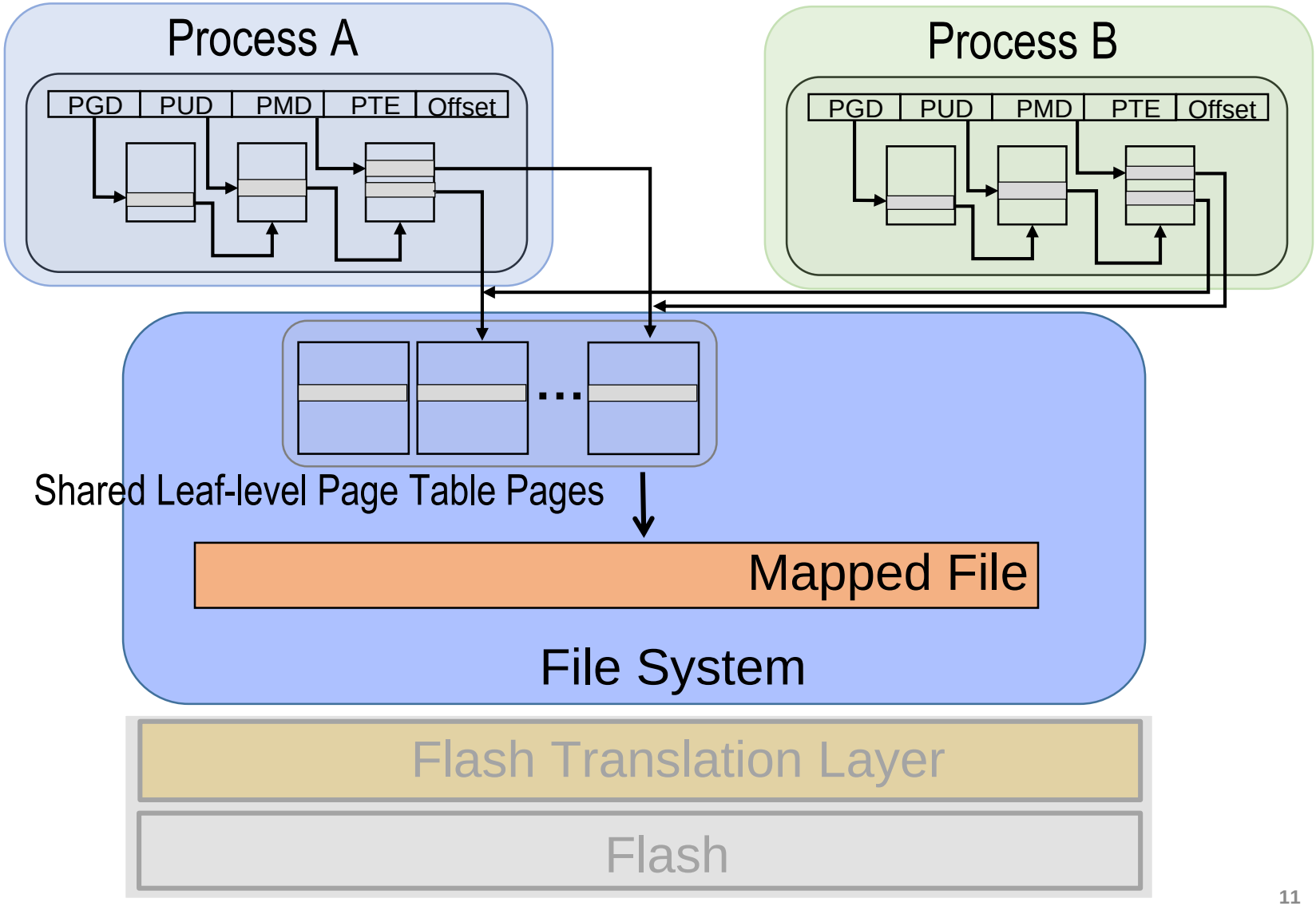
Page Table in FlashMap



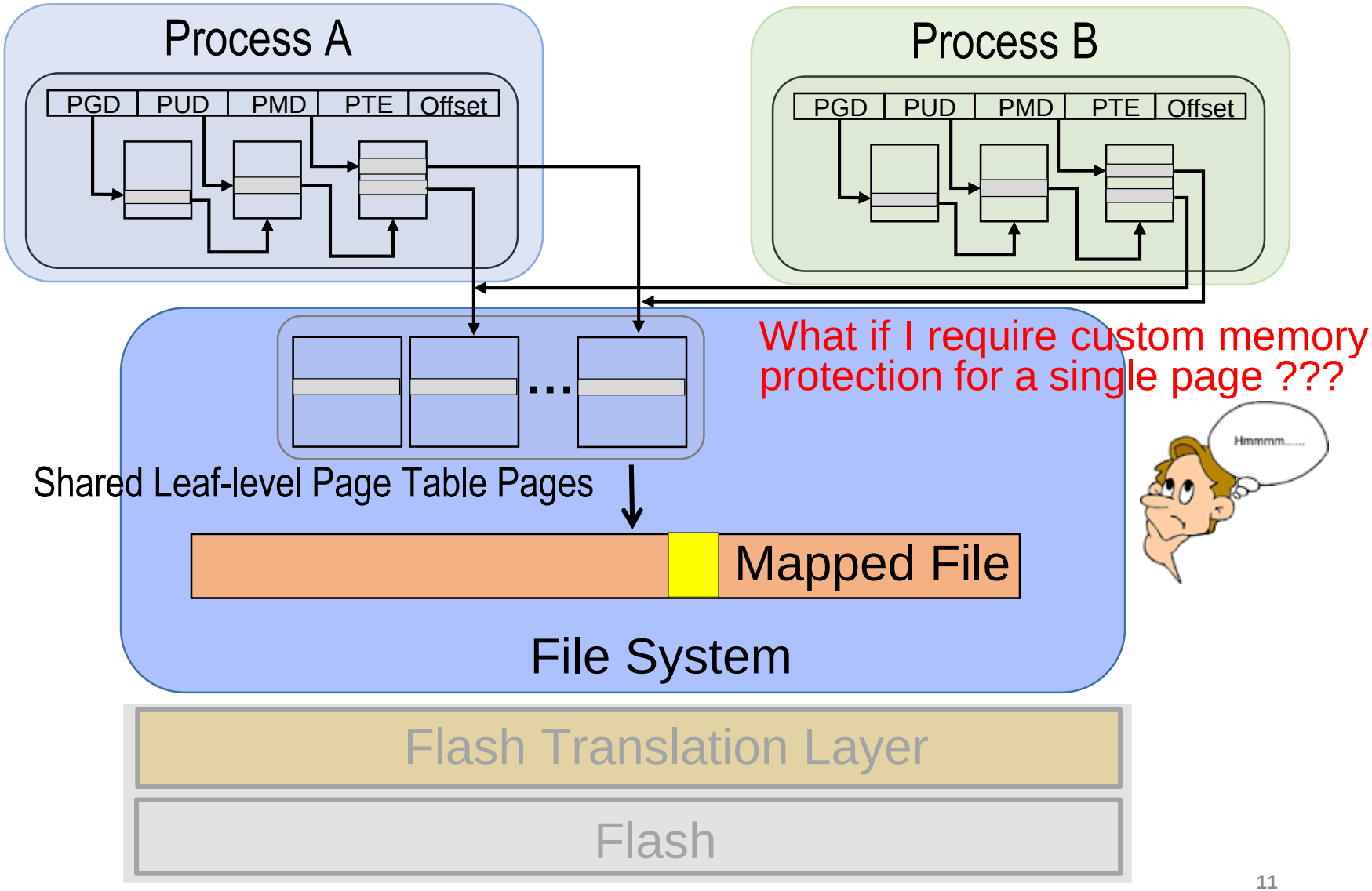
Process's Private Virtual Memory + File Backed Memory

After Mapping a File

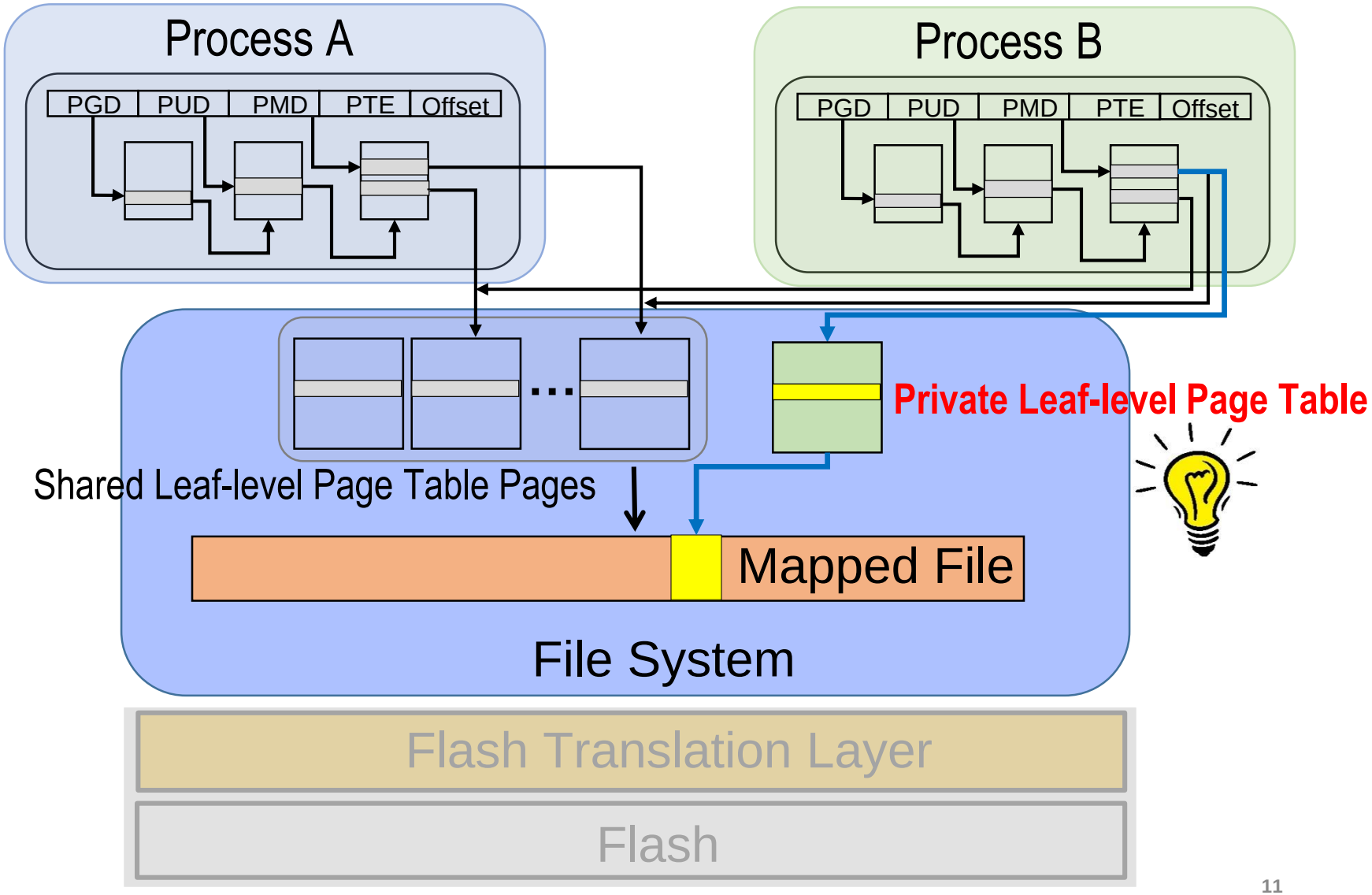
Preserving Memory Protection



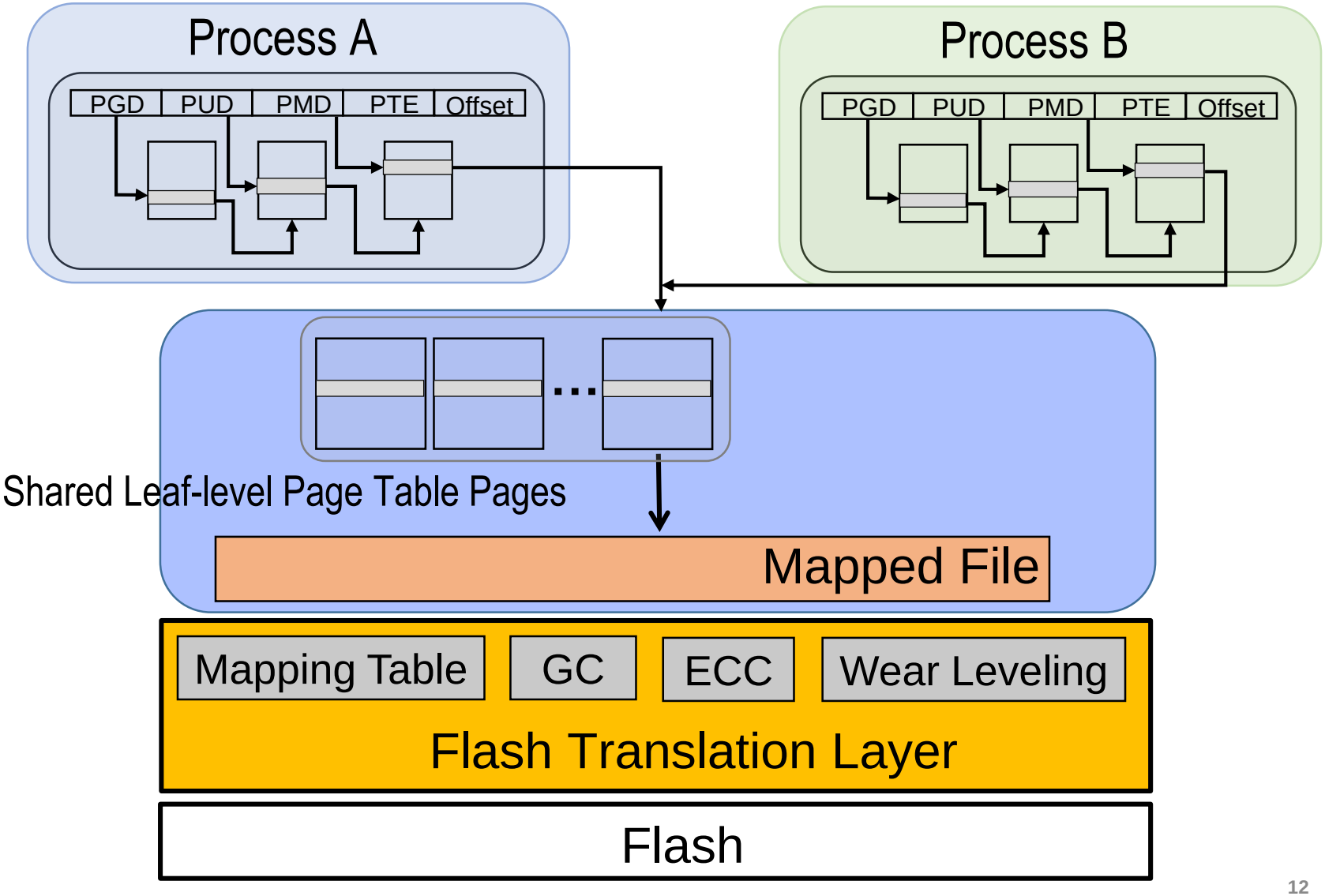
Preserving Memory Protection



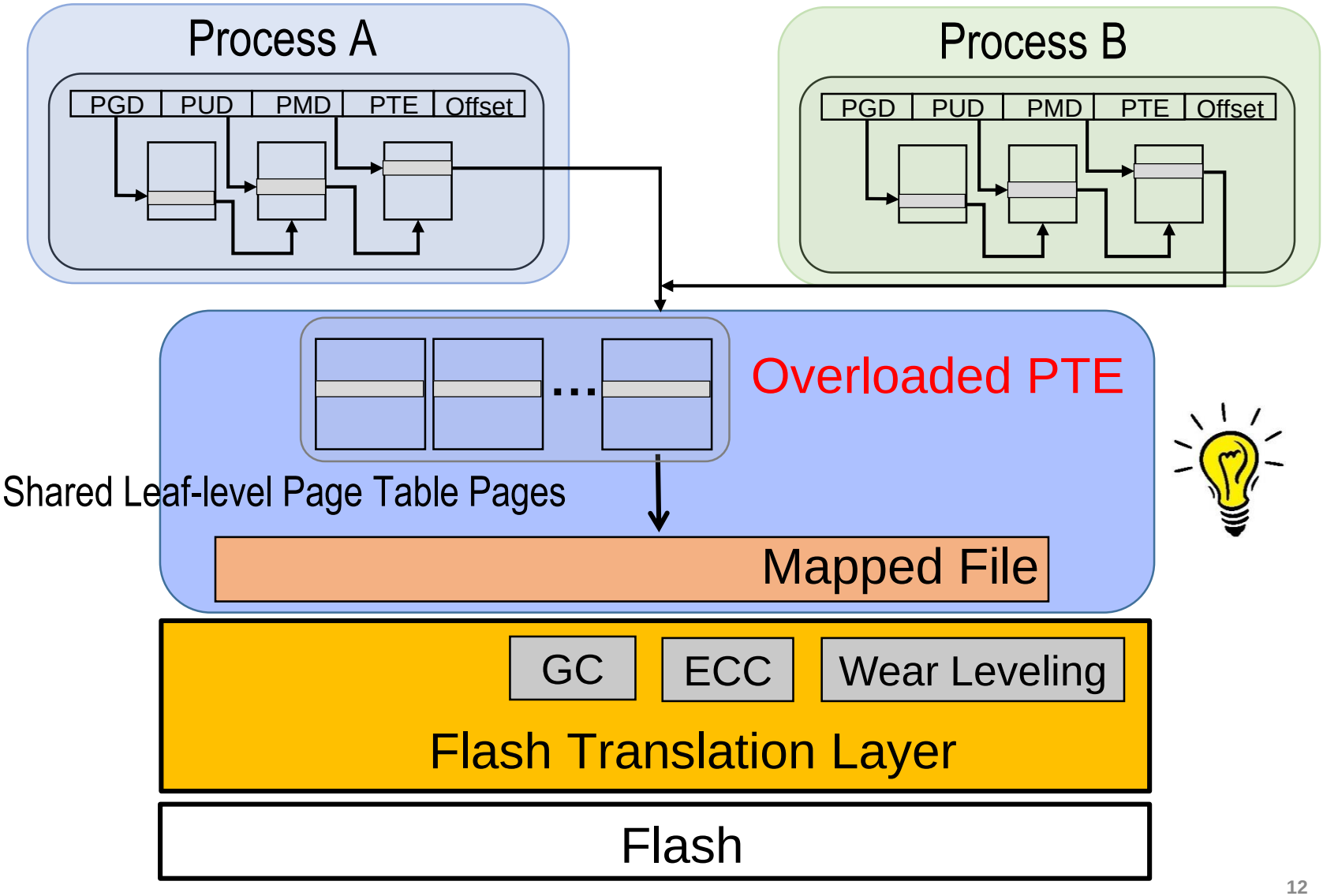
Preserving Memory Protection



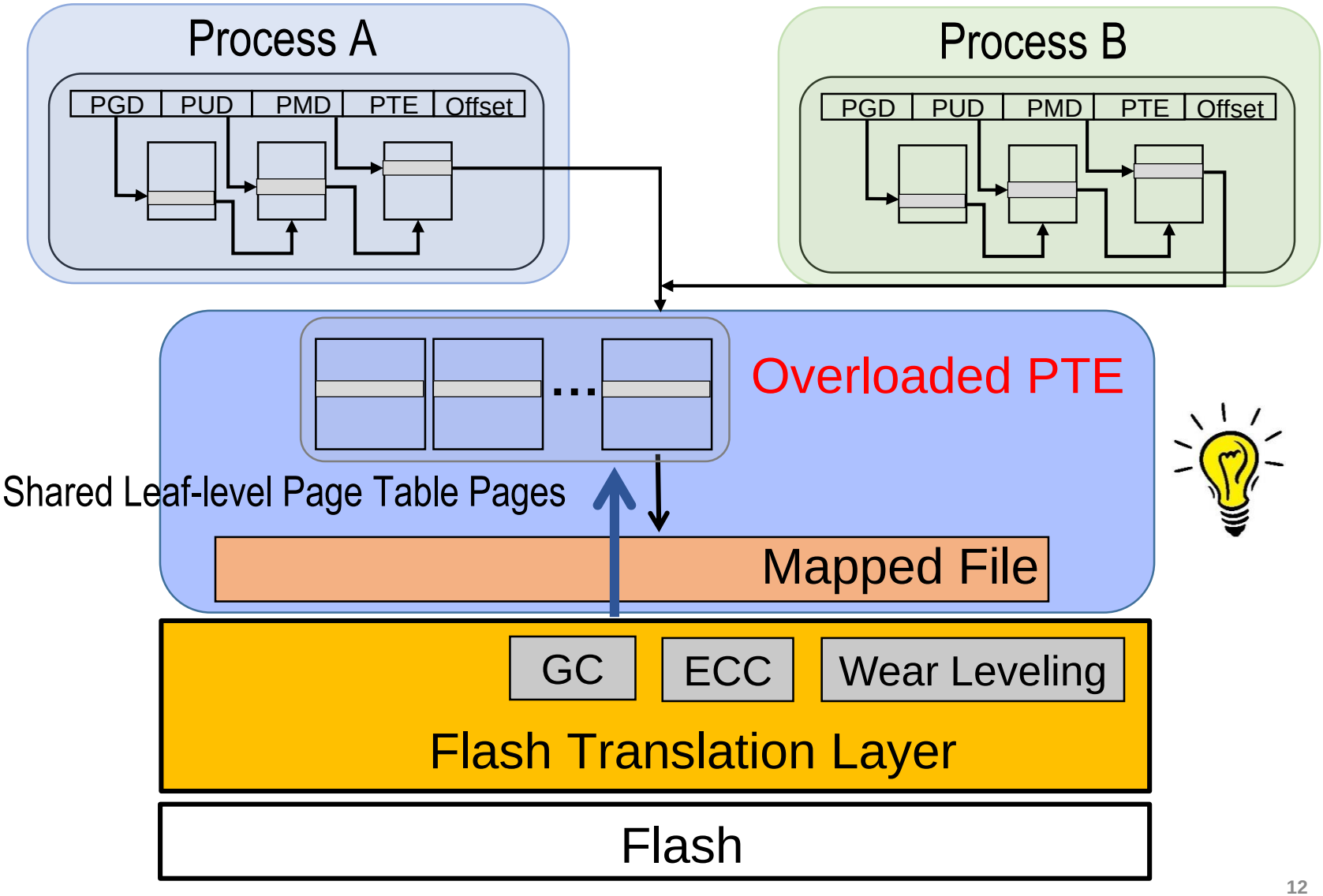
Combining FTL and Shared Page Table



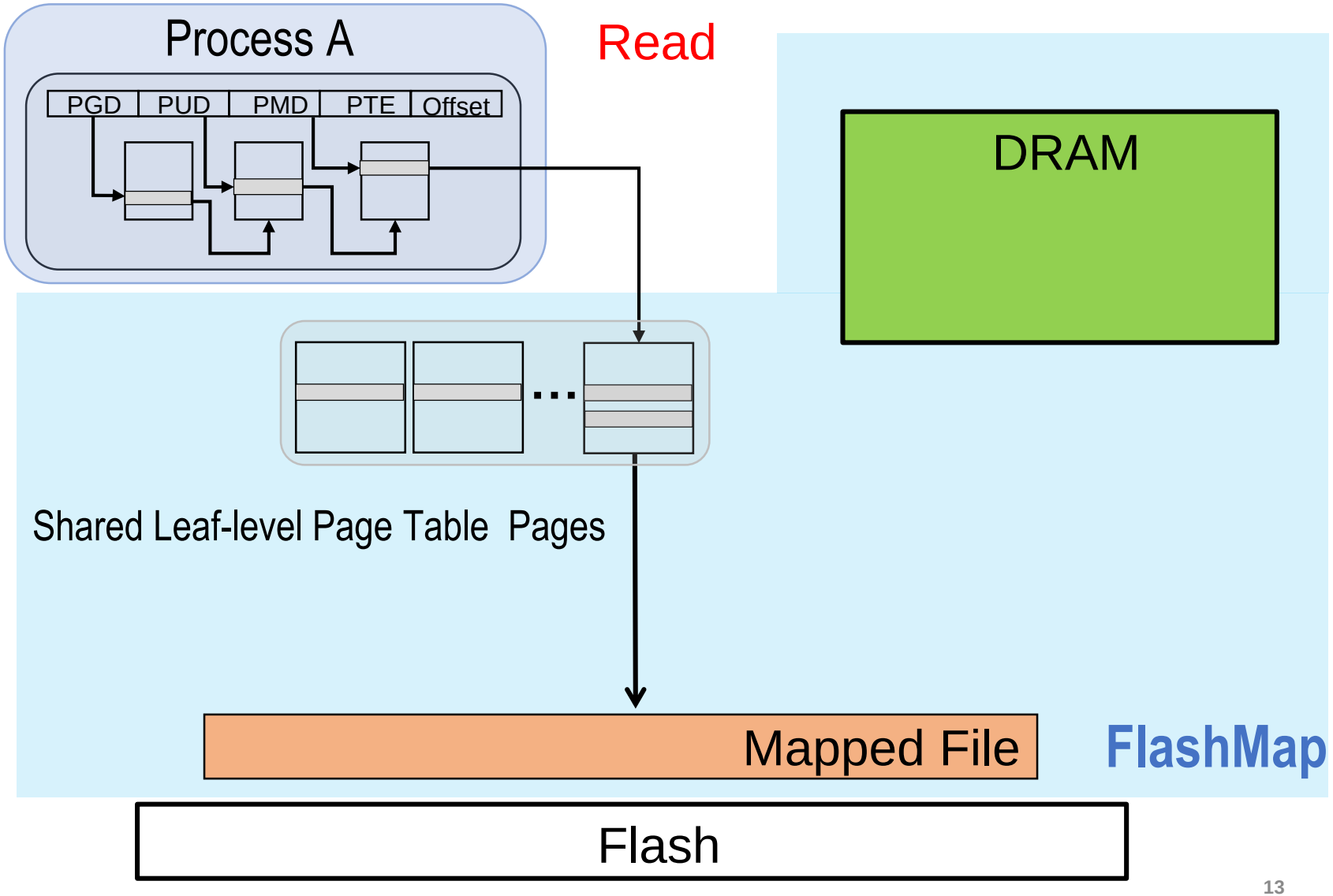
Combining FTL and Shared Page Table



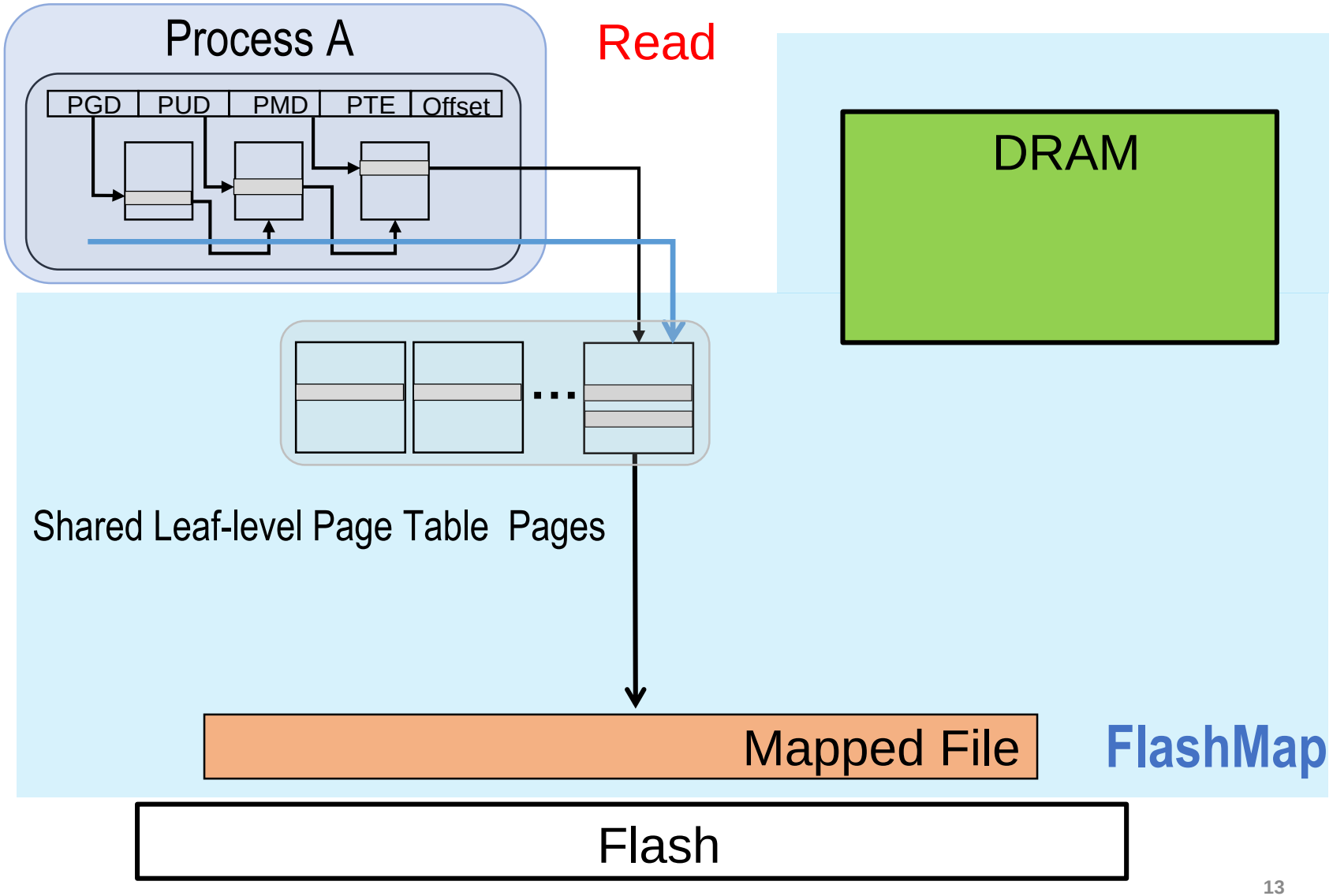
Combining FTL and Shared Page Table



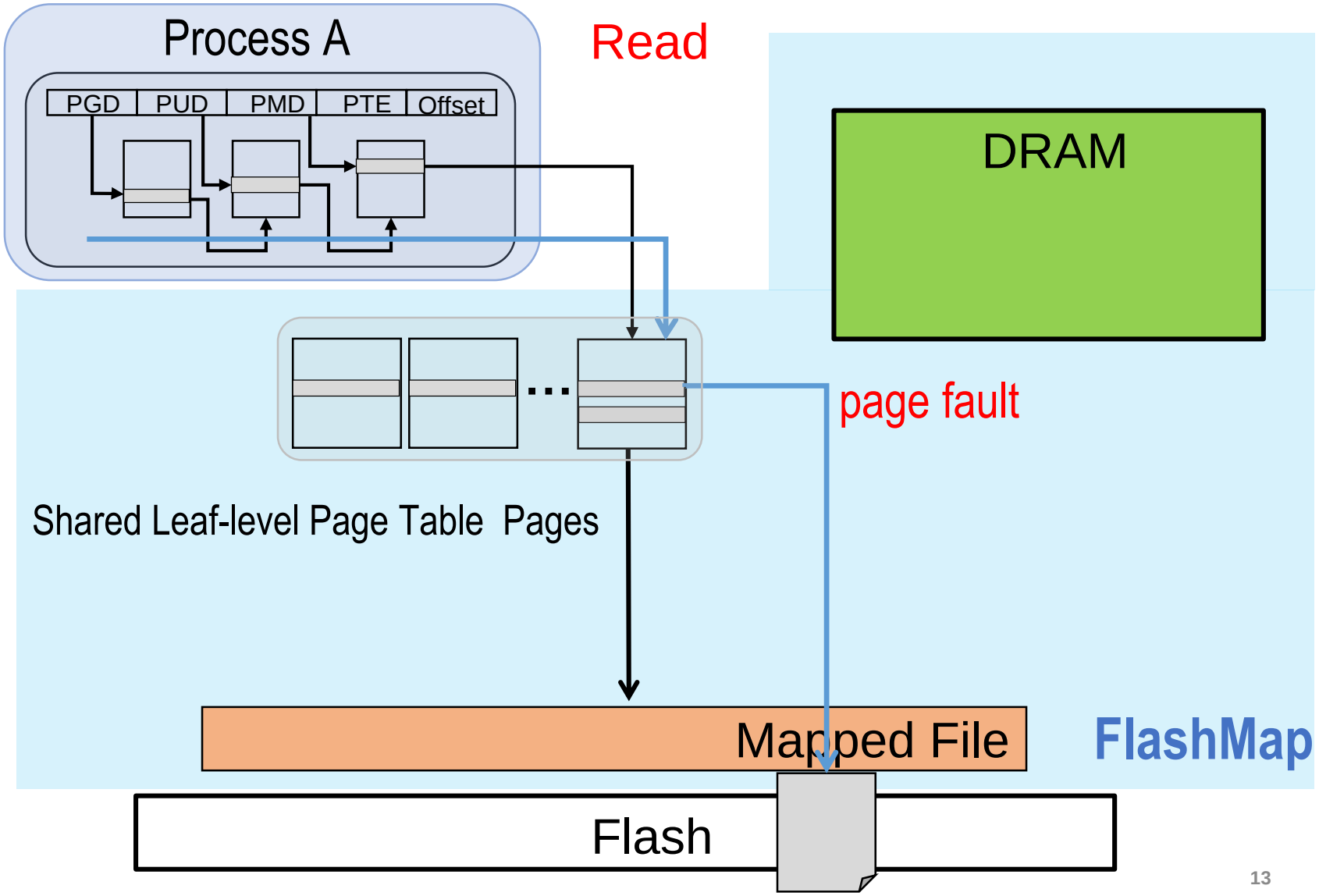
Putting It All Together



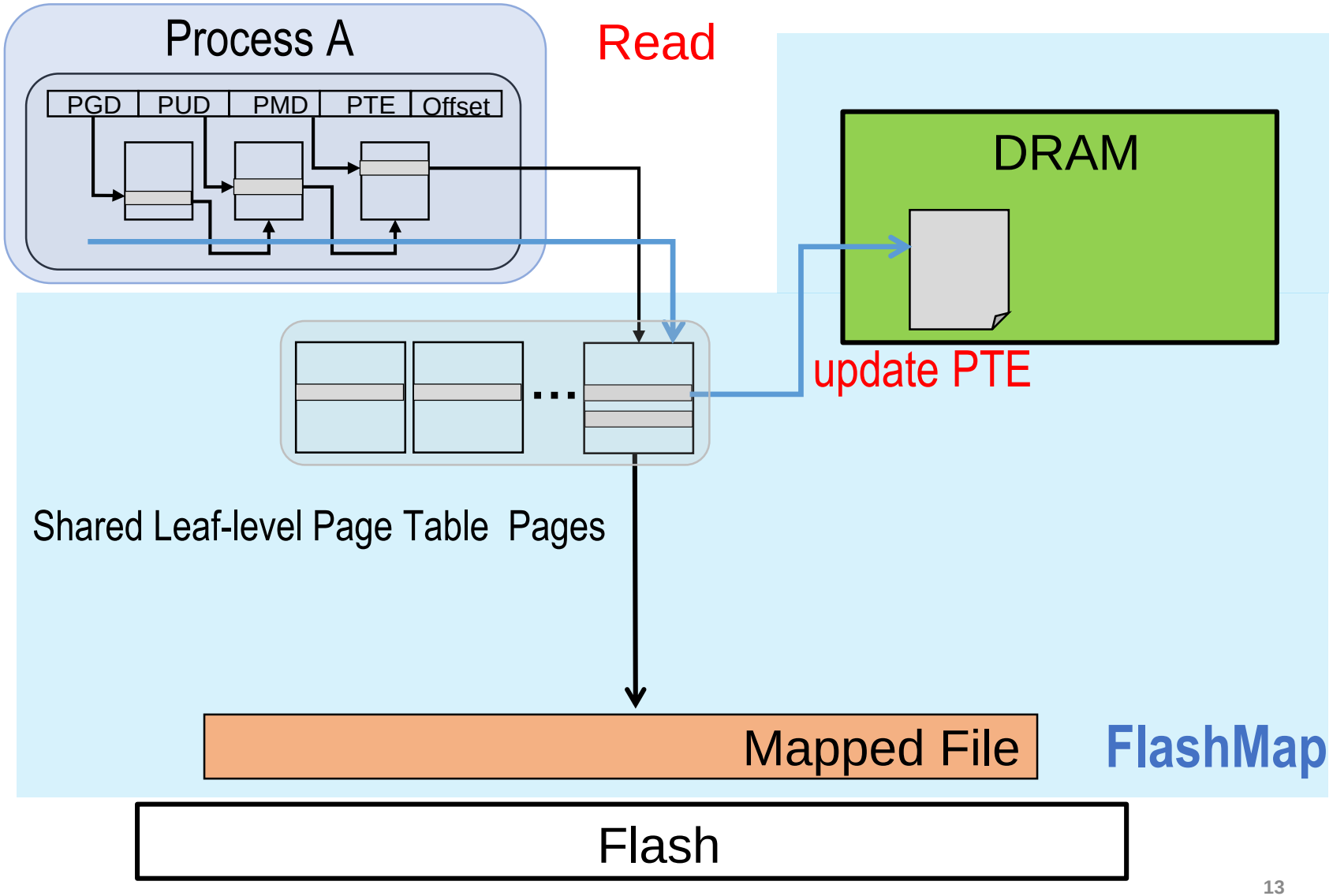
Putting It All Together



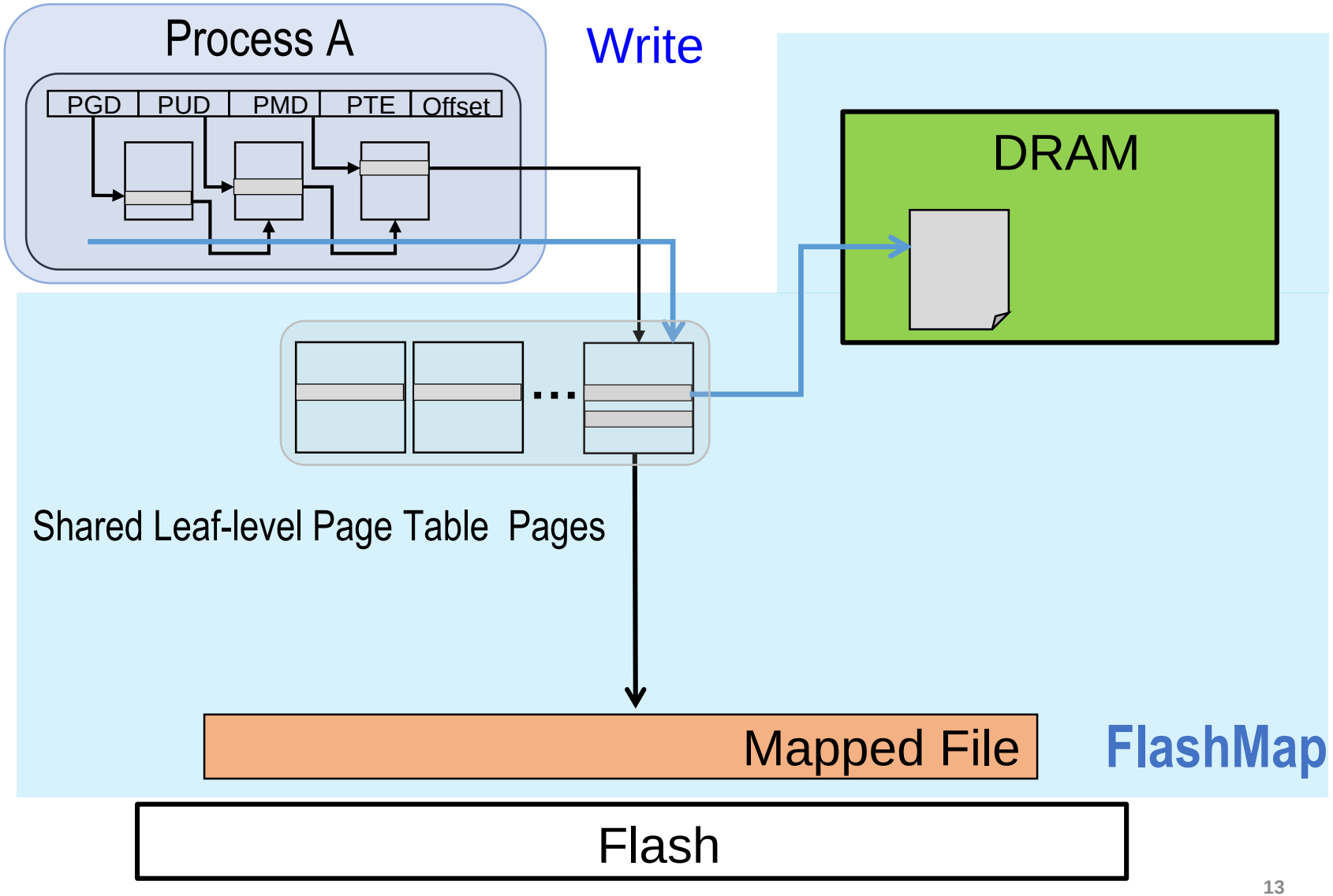
Putting It All Together



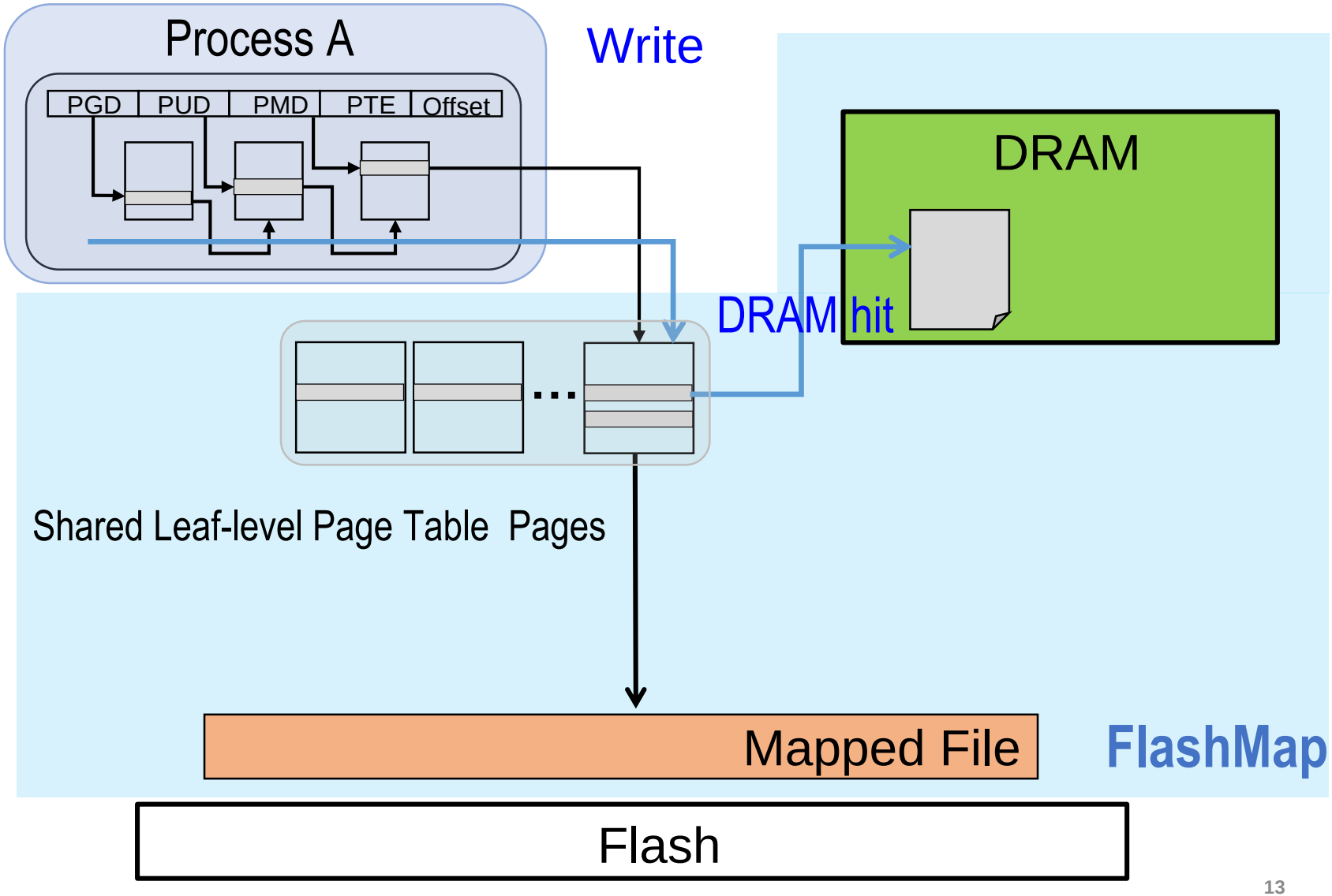
Putting It All Together



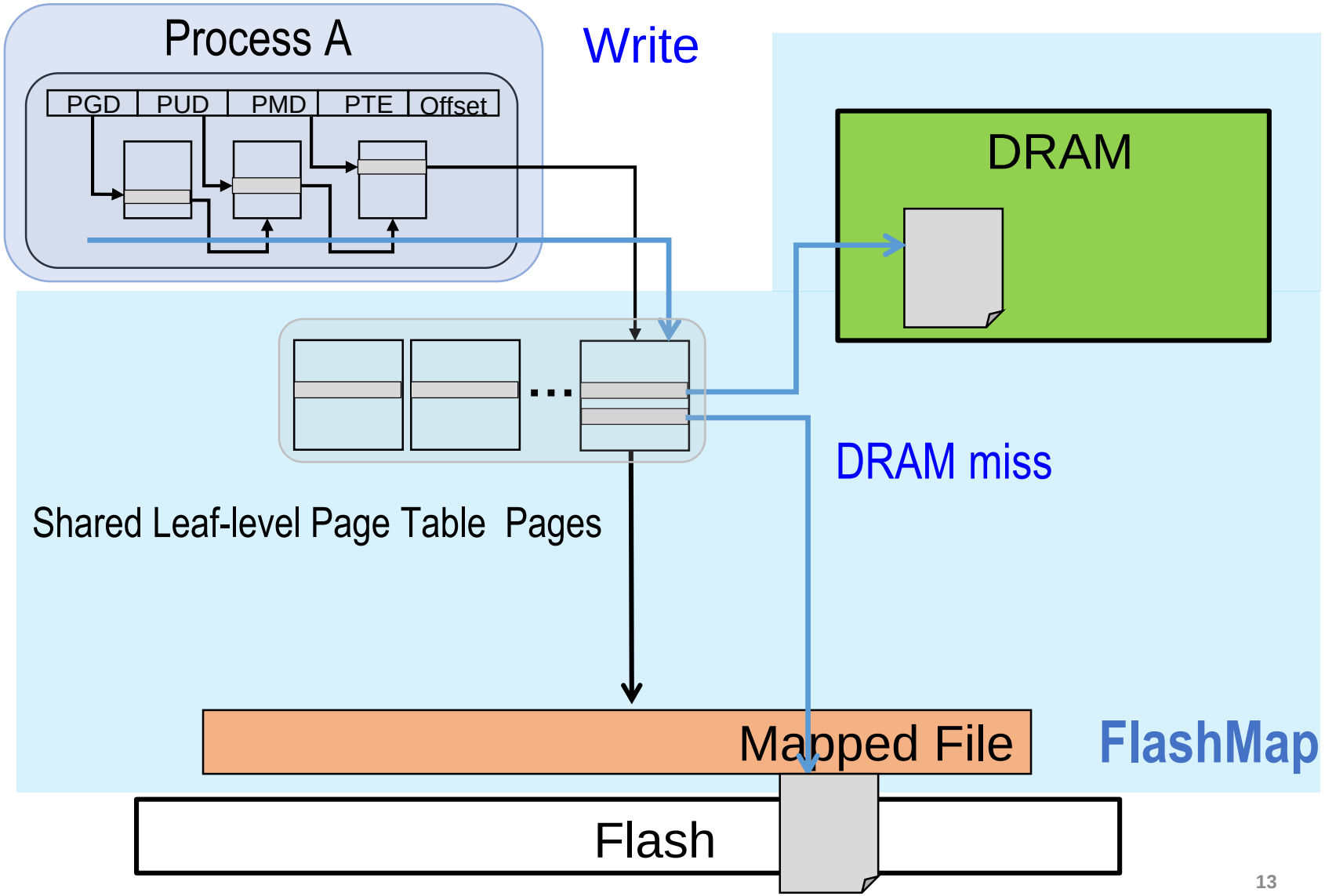
Putting It All Together



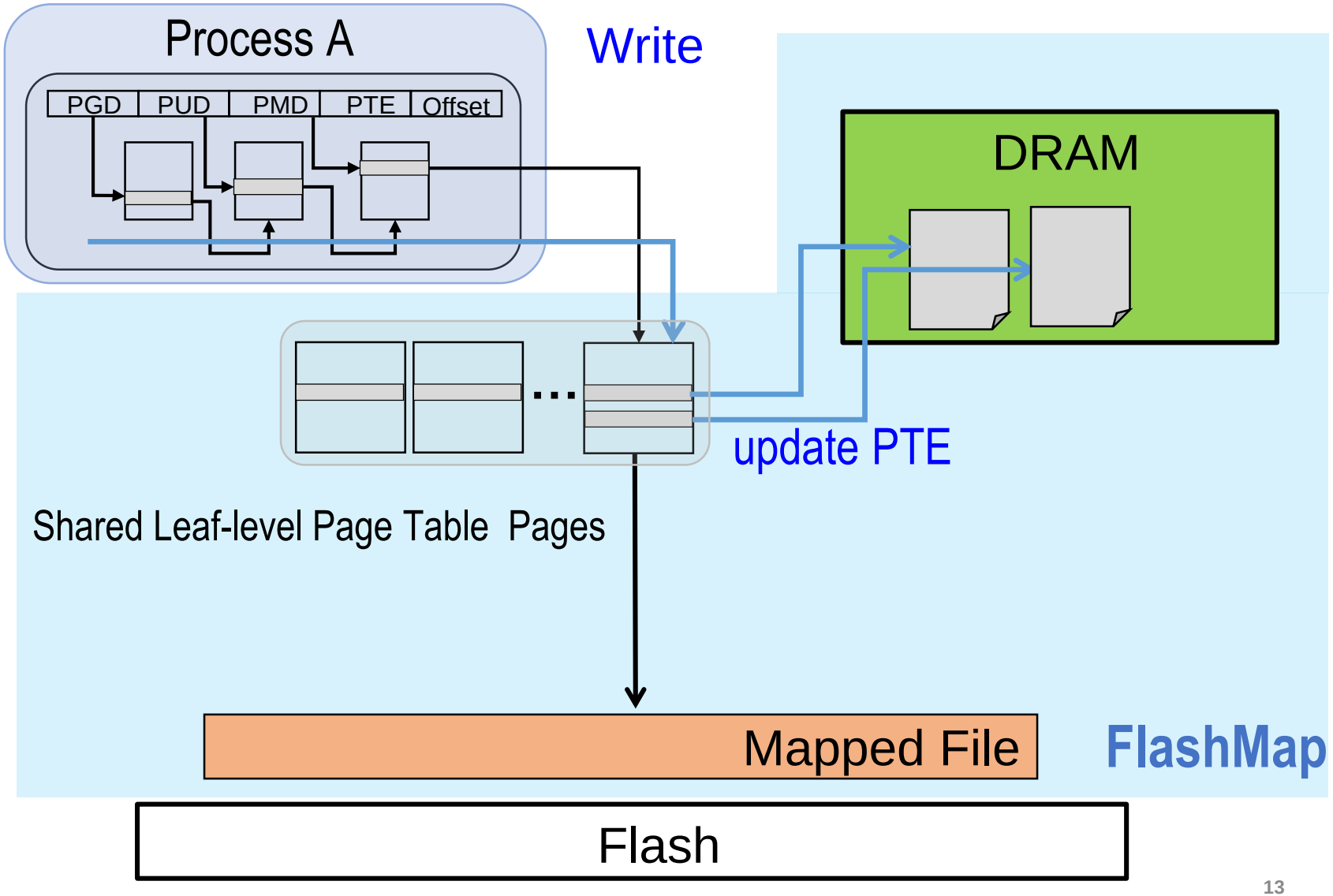
Putting It All Together



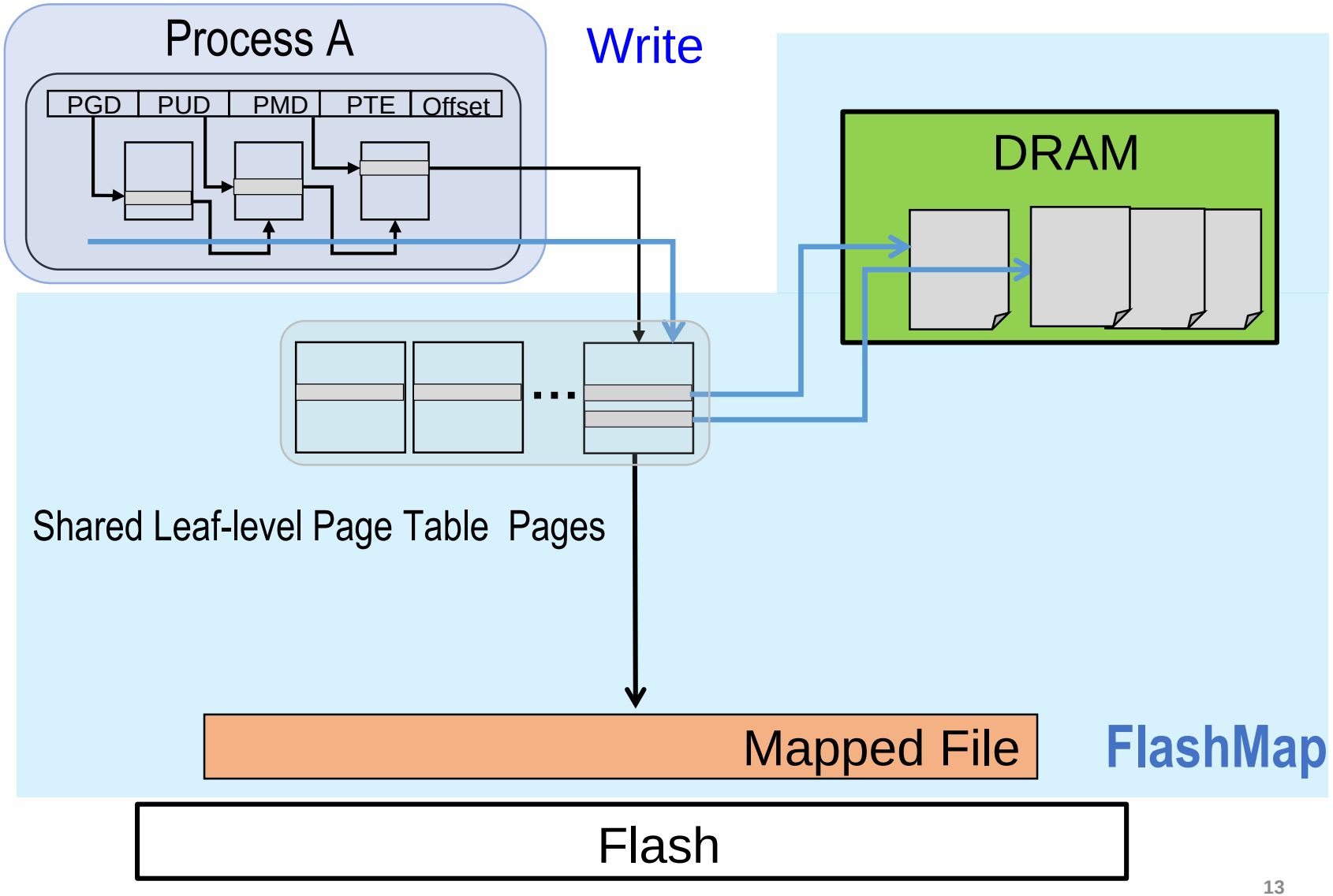
Putting It All Together



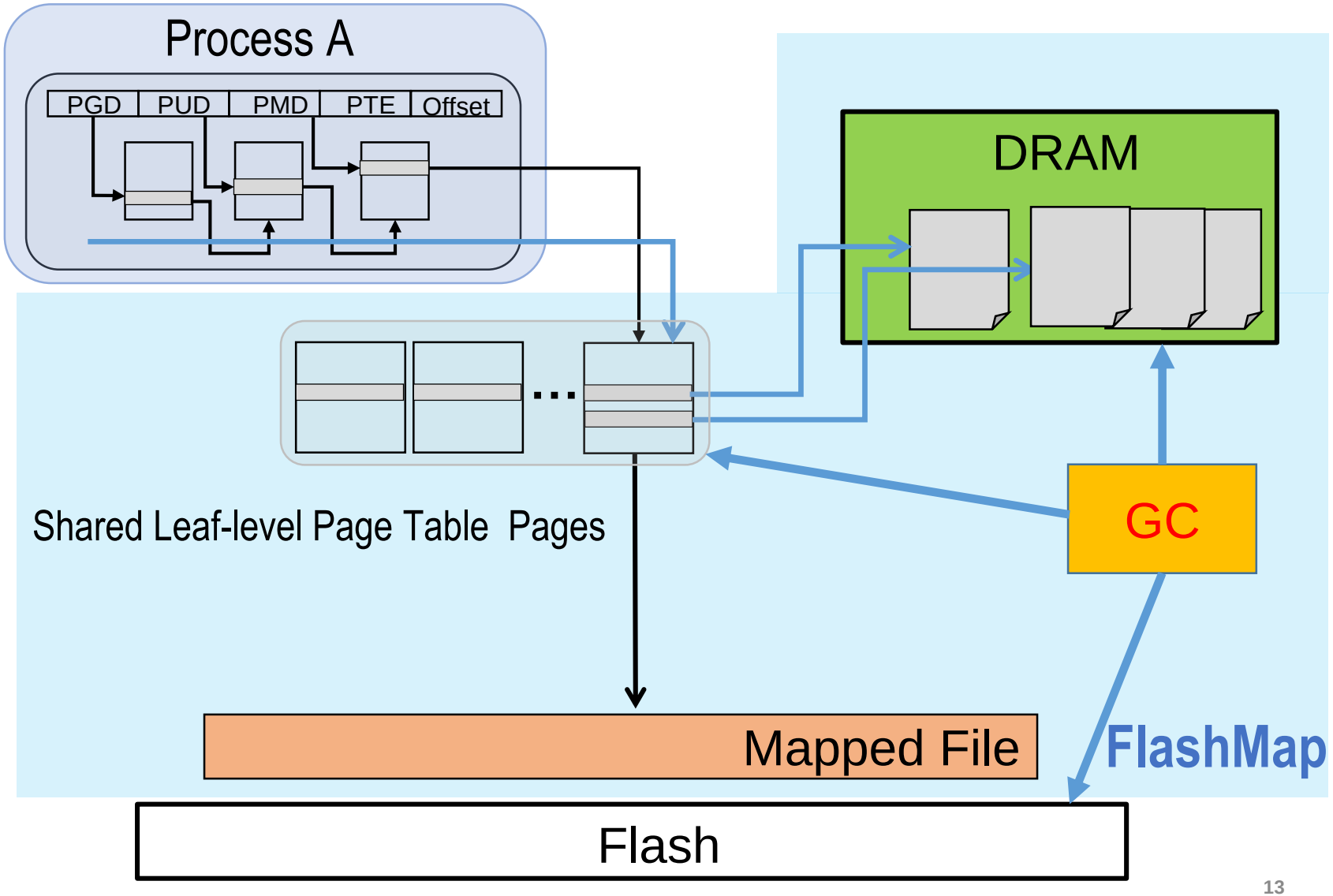
Putting It All Together



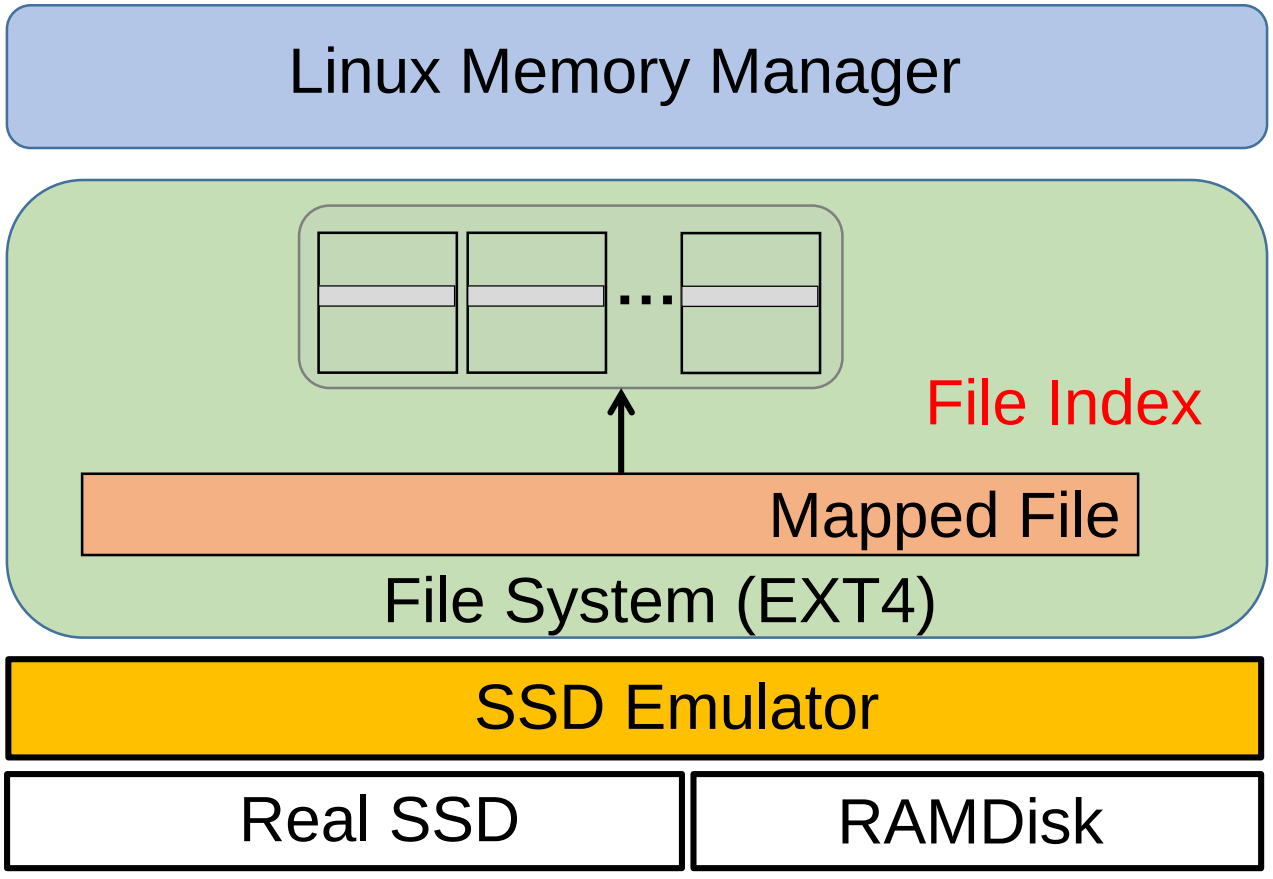
Putting It All Together



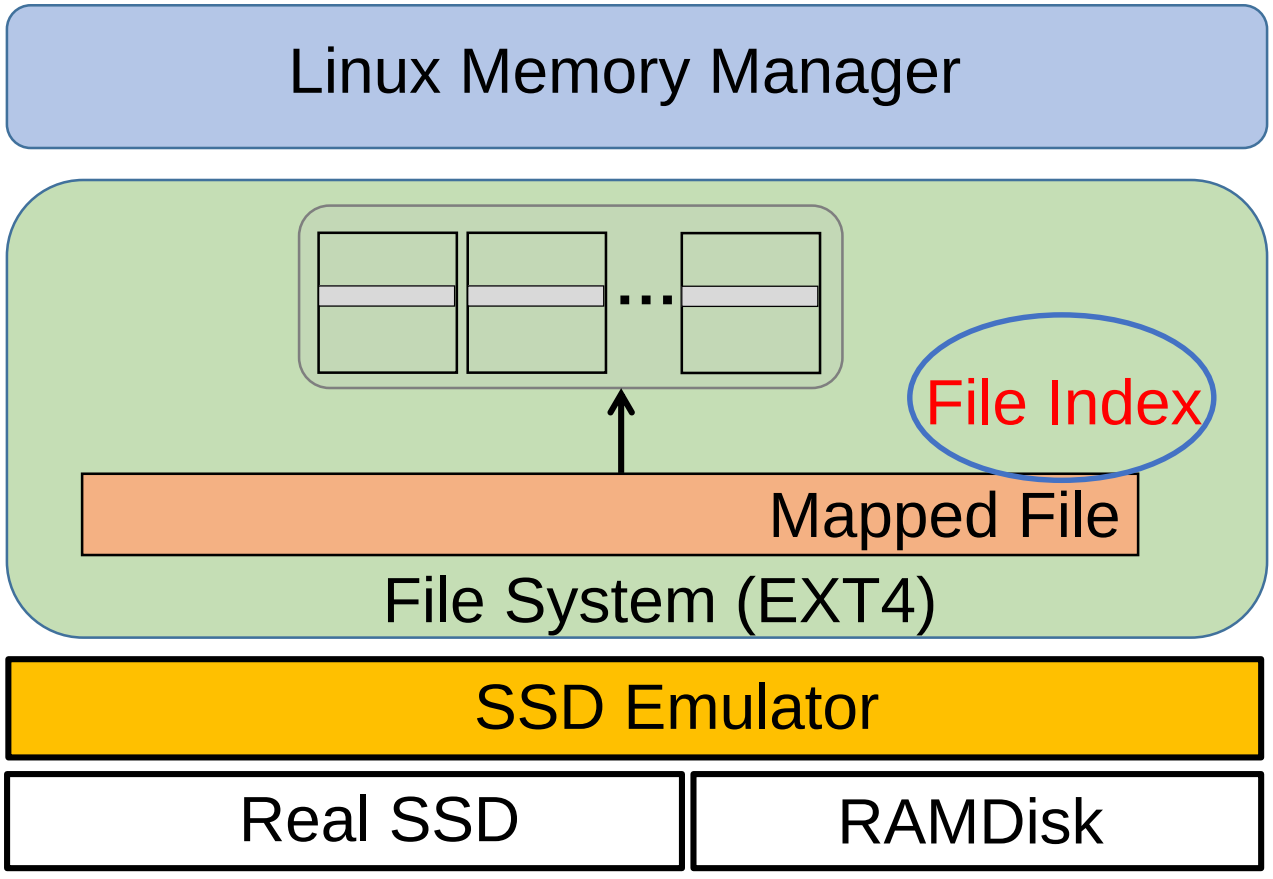
Putting It All Together



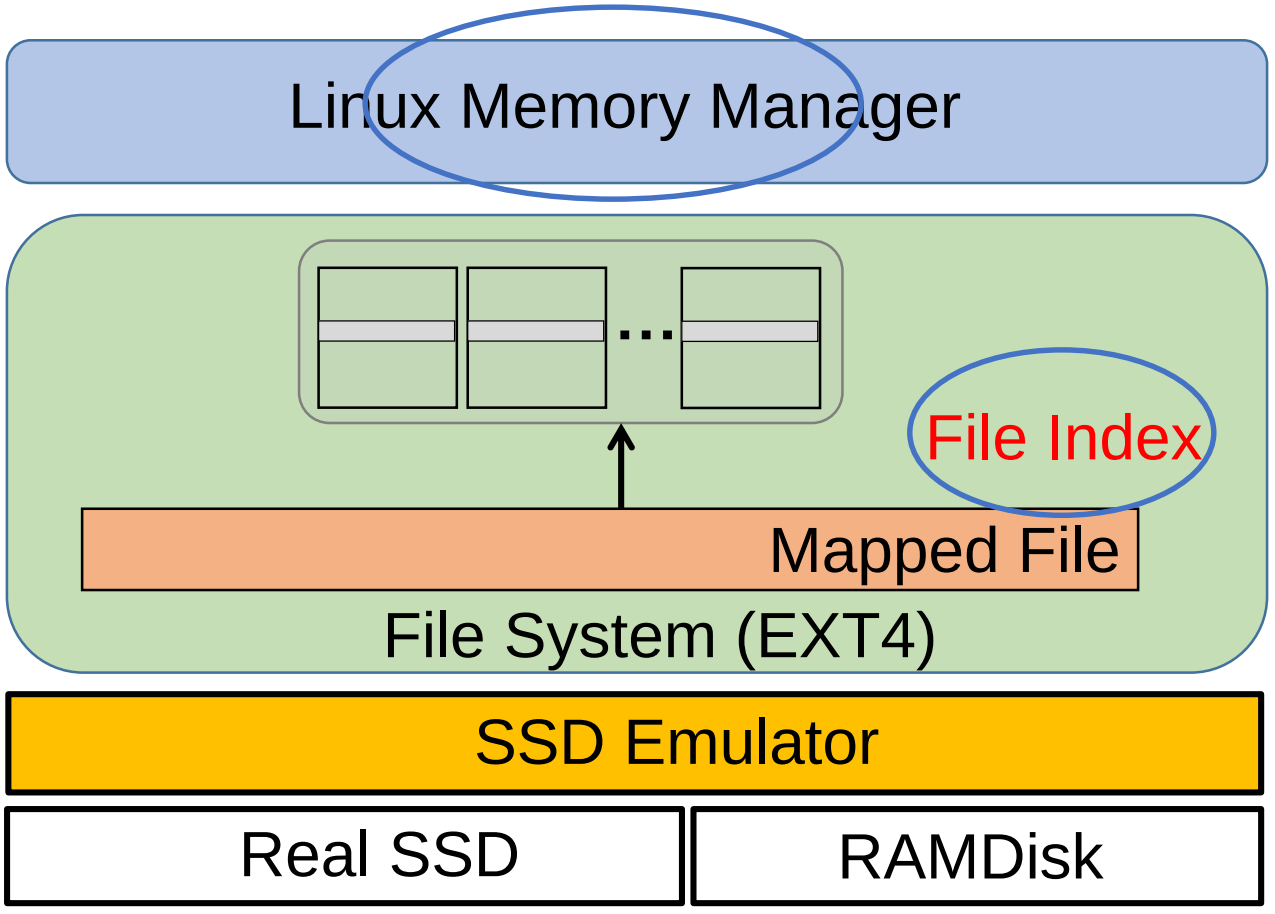
FlashMap: Implementation in Real System



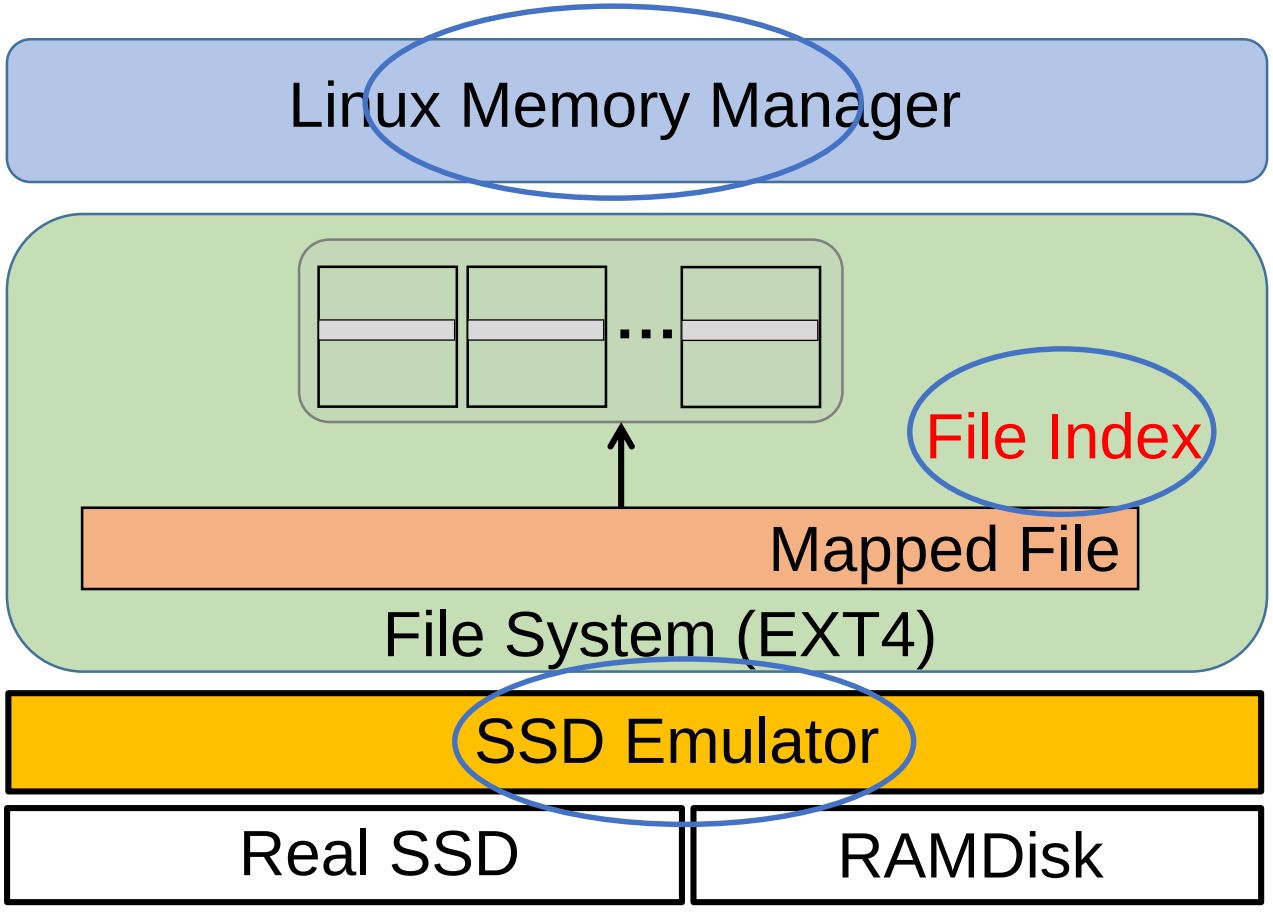
FlashMap: Implementation in Real System



FlashMap: Implementation in Real System



FlashMap: Implementation in Real System



Experimental Setup

Intel Xeon processors + 64 GB DRAM + 2 TB SSD

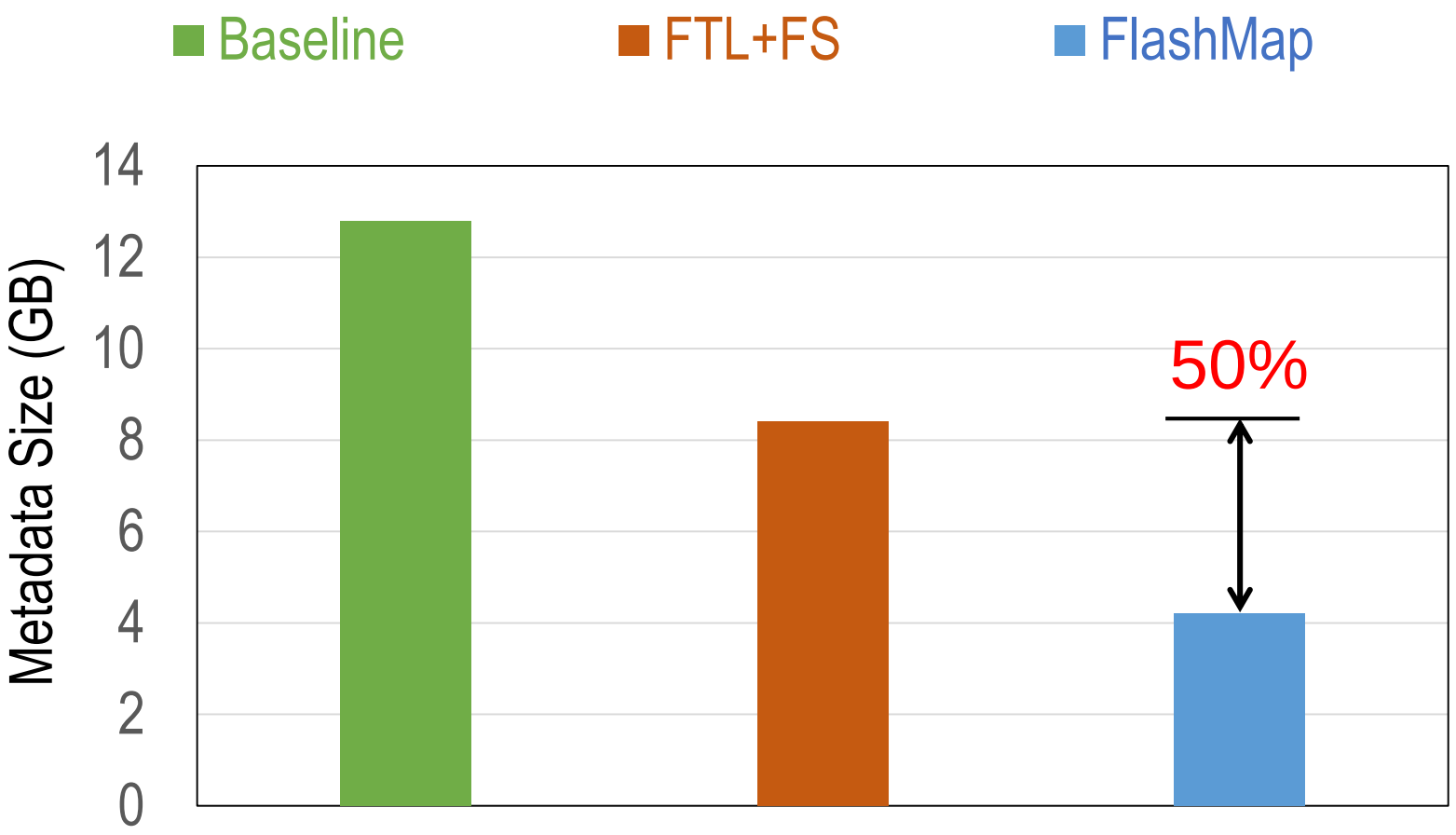
Baseline	unmodified Linux: mmap + EXT4 + FTL with page-level mapping
FTL+FS★	mmap + combined FTL & file system
FlashMap	unified address translation

★similar to Nameless Writes [Zhang et al., FAST'12] and DFS [Josephson et al., FAST'10]

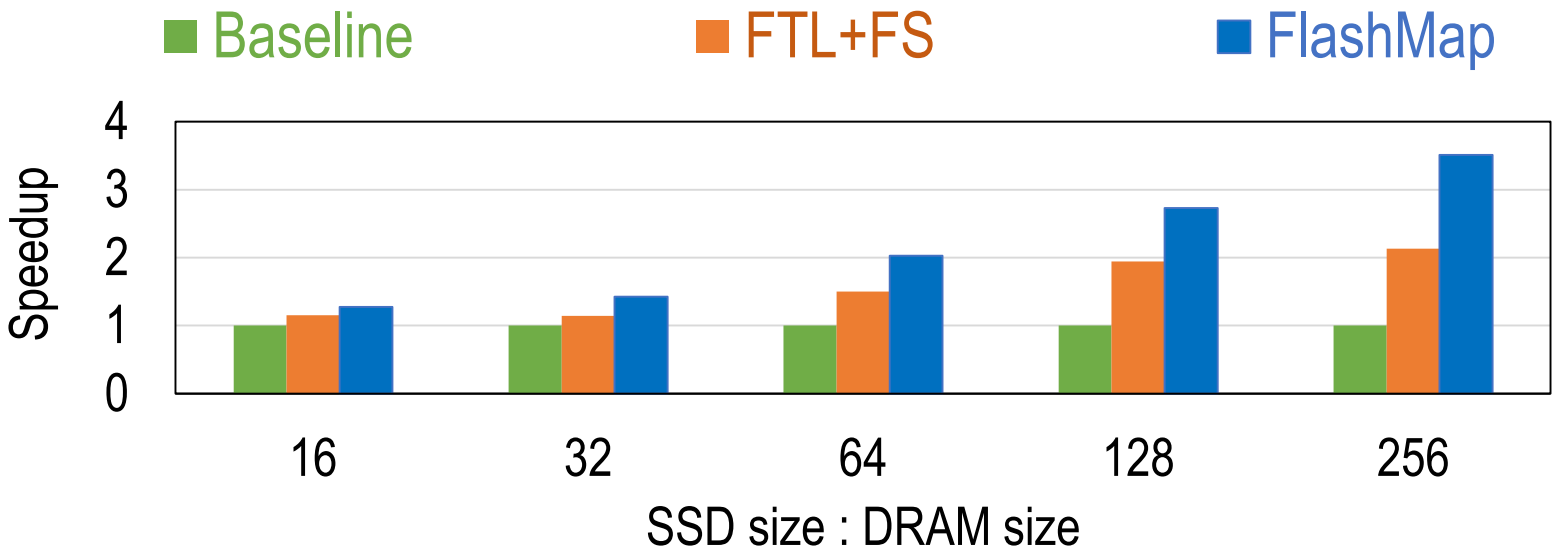
Real Application Workloads

NoSQL Store	 +  YCSB
SQL Database	Shore-MT + TPCC, TPCB, TATP
Graph Analytics	 + PageRank

Metadata Size for 2 TB SSD

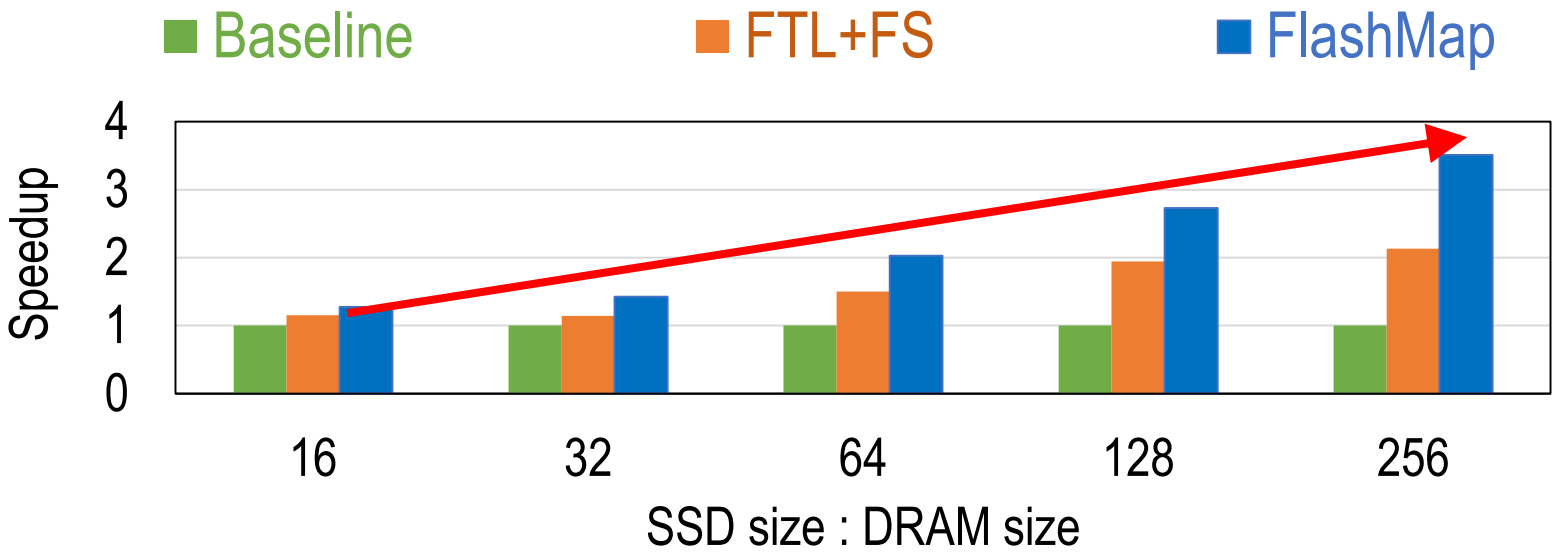


Benefits from Reduced Mapping Overhead



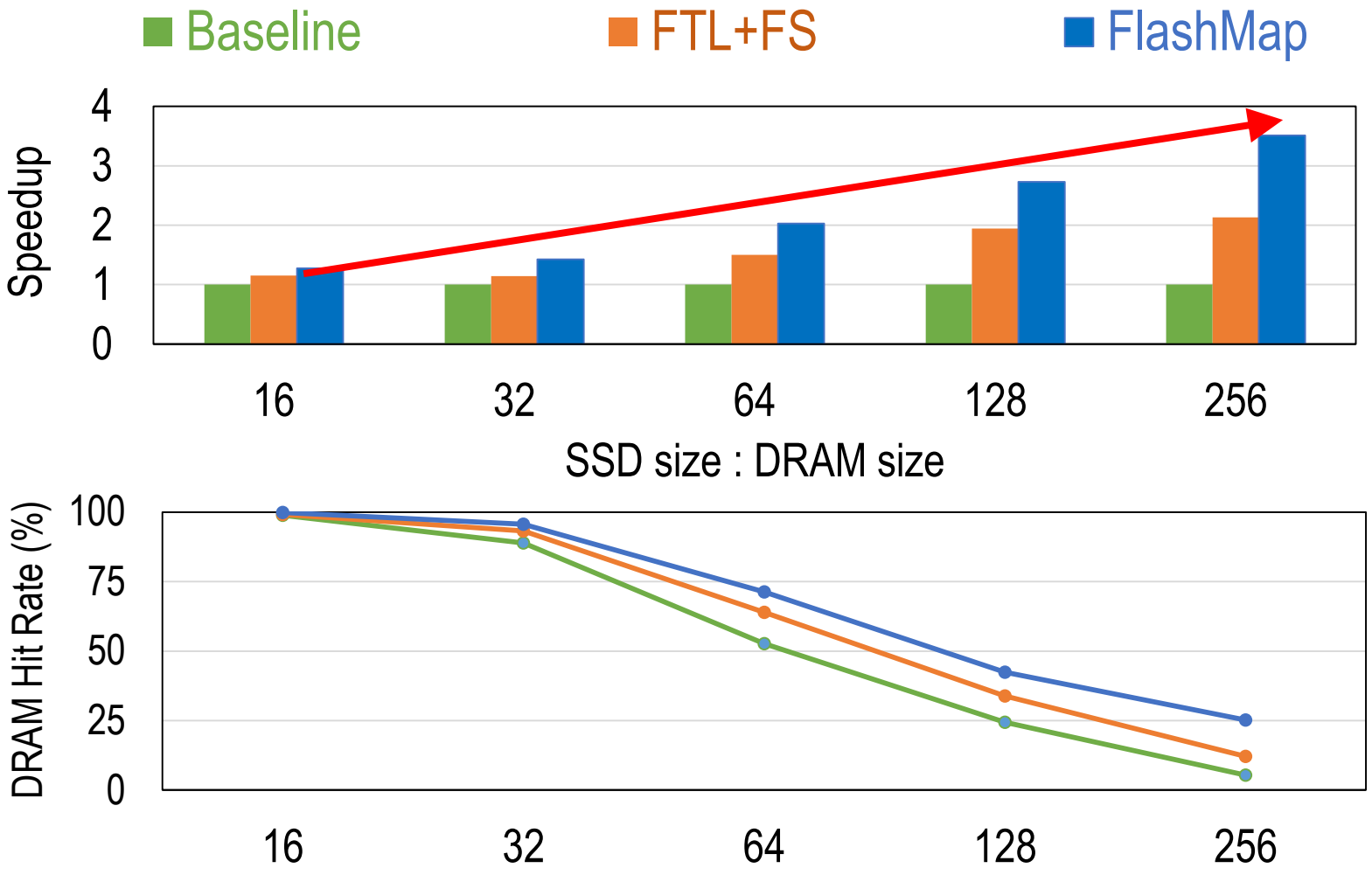
FlashMap: 1.7x performance improvement over FTL+FS

Benefits from Reduced Mapping Overhead



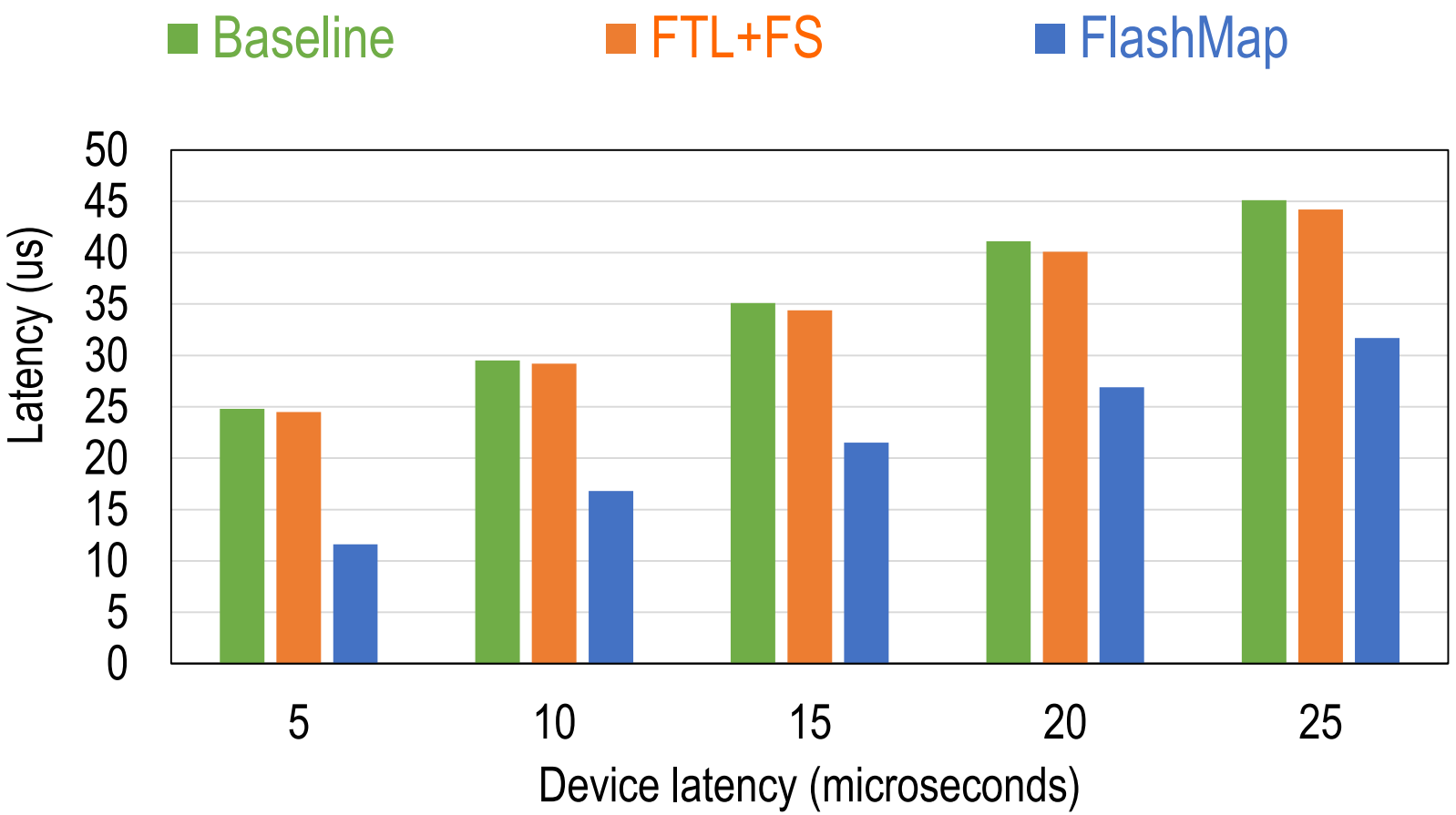
FlashMap: 1.7x performance improvement over FTL+FS

Benefits from Reduced Mapping Overhead



Reducing the mapping overhead improves the DRAM caching efficiency

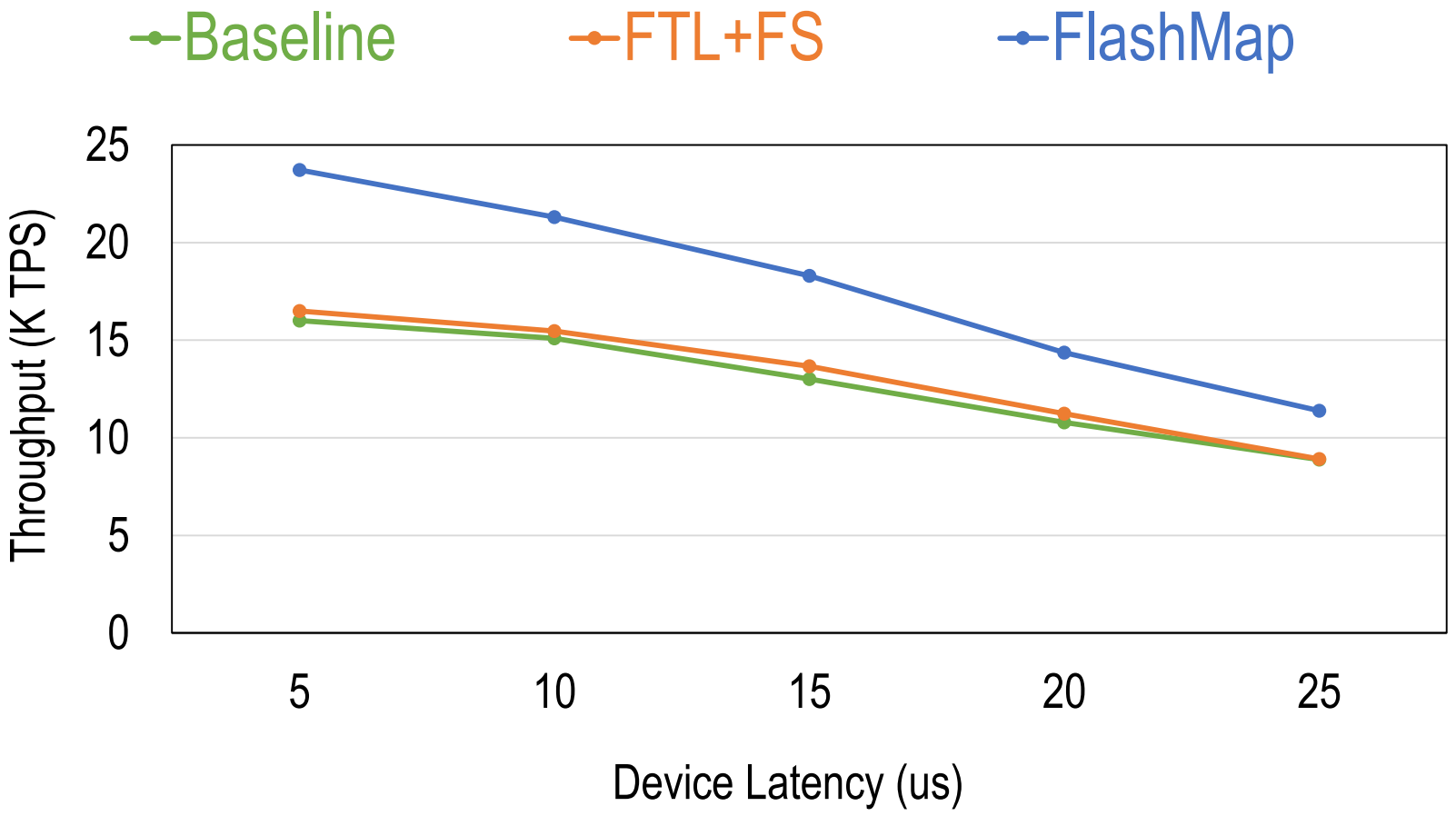
Latency Reduction



Benefit (up to 53% latency reduction)

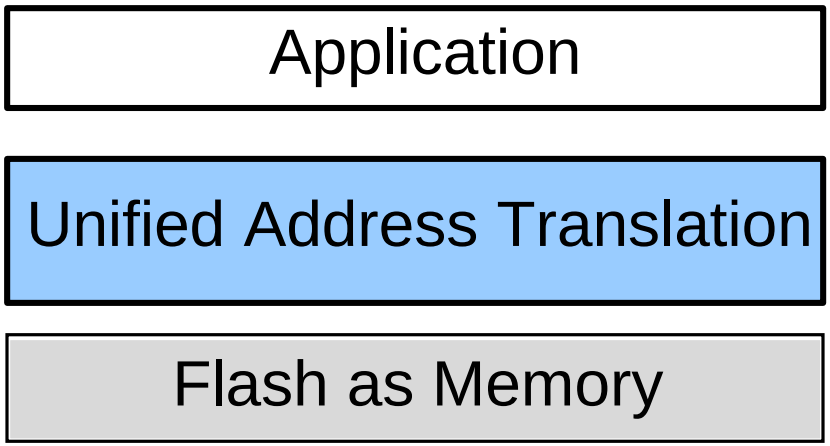
mainly comes from the combination of page table and file system

Benefits from Reduced Latency



FlashMap: 1.8x more TPS than baseline and FTL+FS

Conclusion



1 **Reduced Storage**

3.3x performance improvement for data-intensive applications

2 **Reduced Latency**

53% latency reduction for high-end SSDs, 1.8x more TPS for latency-sensitive applications, e.g., database systems

Thanks!

Jian Huang
jian.huang@gatech.edu

Anirudh Badam† Moinuddin K. Qureshi Karsten Schwan

Georgia Tech

† Microsoft

Q&A