



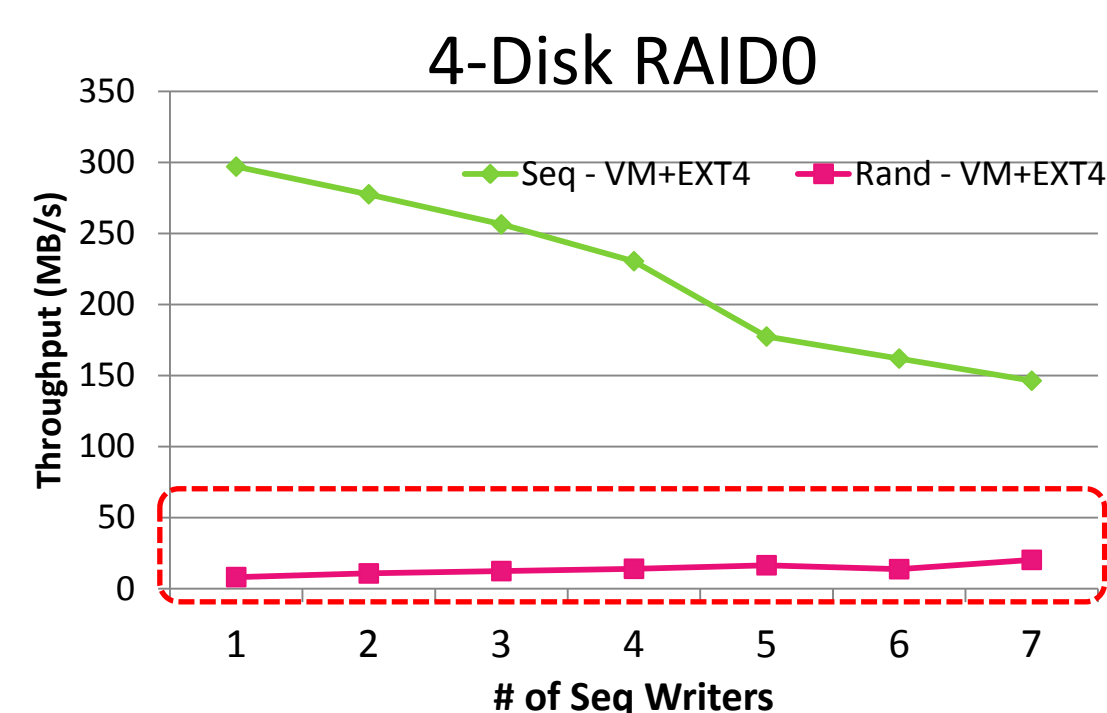
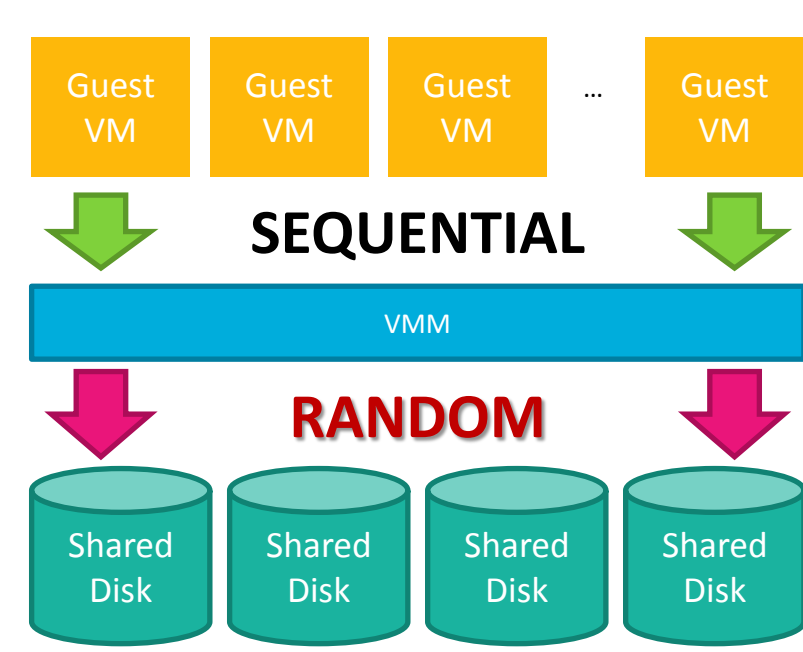
Gecko: Contention-Oblivious Disk Arrays for Cloud Storage

Ji-Yong Shin¹, Mahesh Balakrishnan², Tudor Marian³, and Hakim Weatherspoon¹

¹Cornell University, ²Microsoft Research, ³Google

Motivation

- Cloud/Virtualization accelerates consolidation of servers
 - Numbers of CPU cores and VMs increase per server
 - Storage is typically poorly virtualized



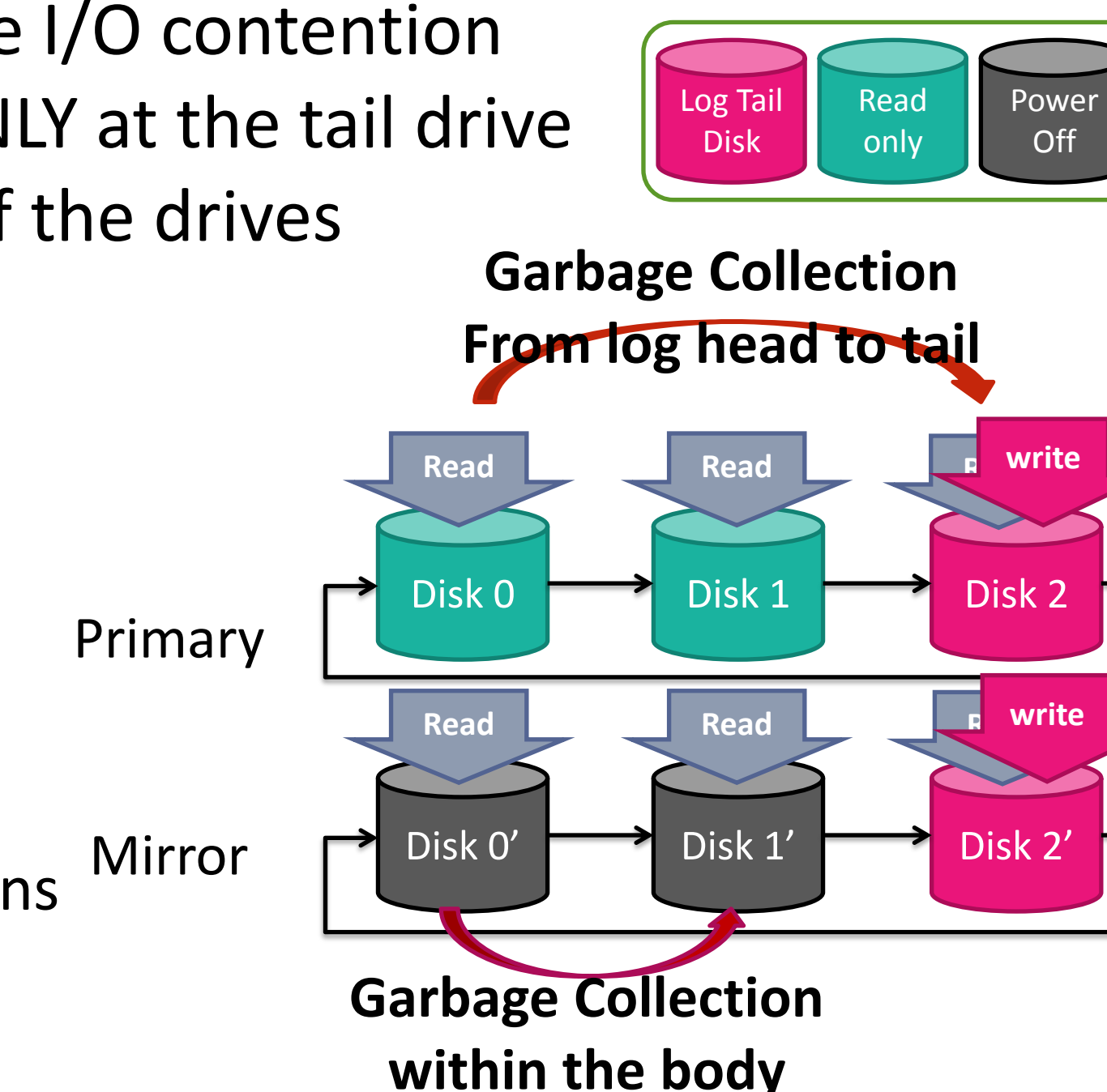
I/O contention destroy the performance

Challenges to I/O contention

- I/O contention to overcome
 - I/O contention moves the disk head and destroys performance
 - **Write-write, write-read, read-read, write-GC, and read-GC**
- RAID cannot preserve high throughput
 - IO performance varies depending on coexisting VMs
 - **Vulnerable to I/O Contention**
- LFS only solves write-write contention
 - **GC (Garbage Collection) interferes** with logging
 - First class **reads interfere** with logging

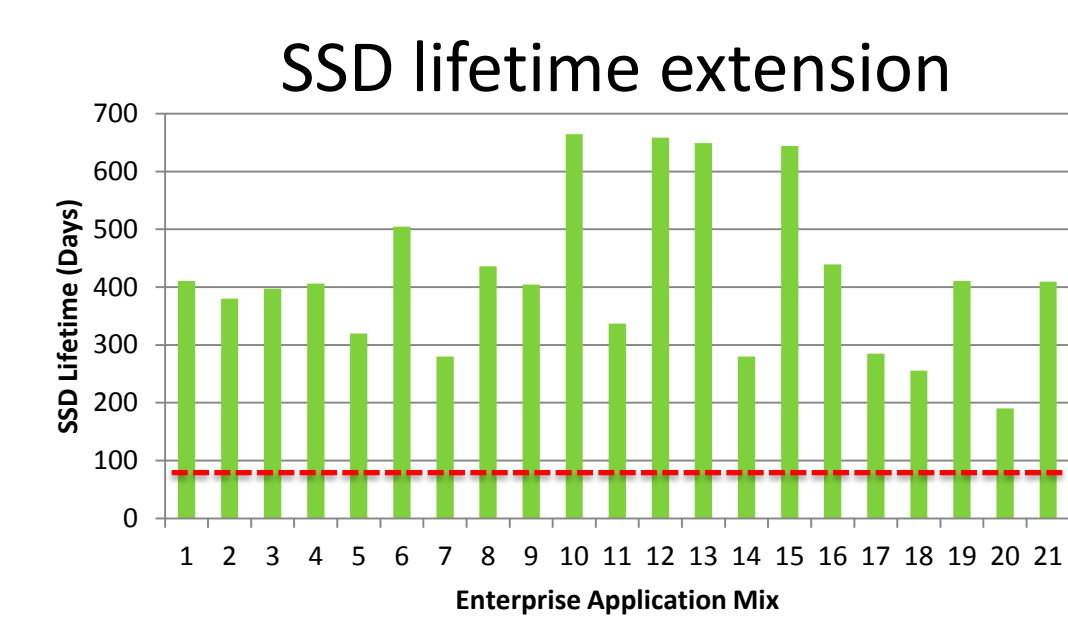
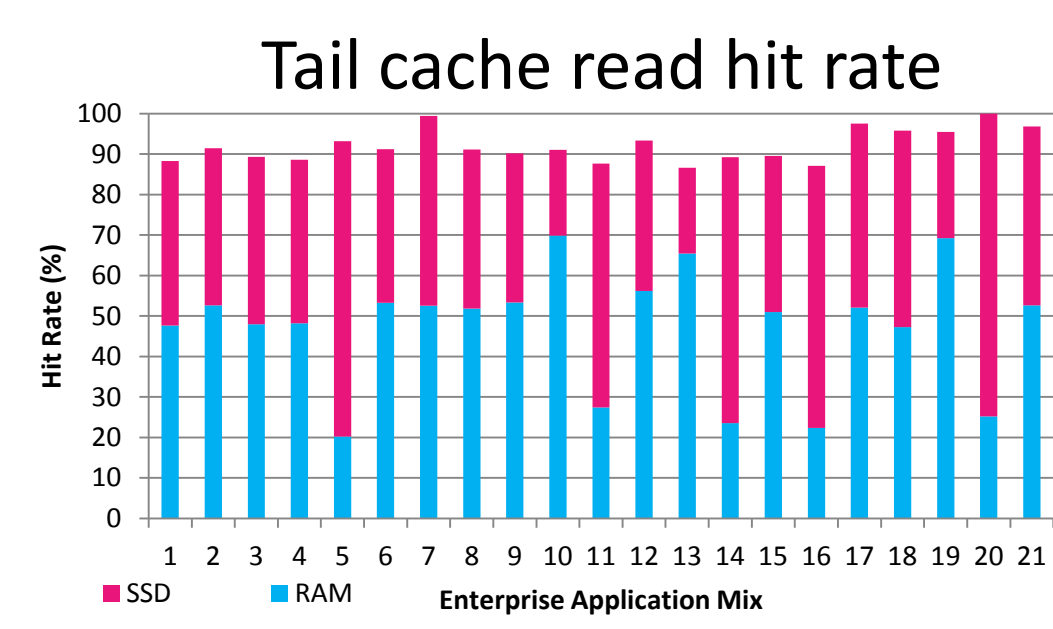
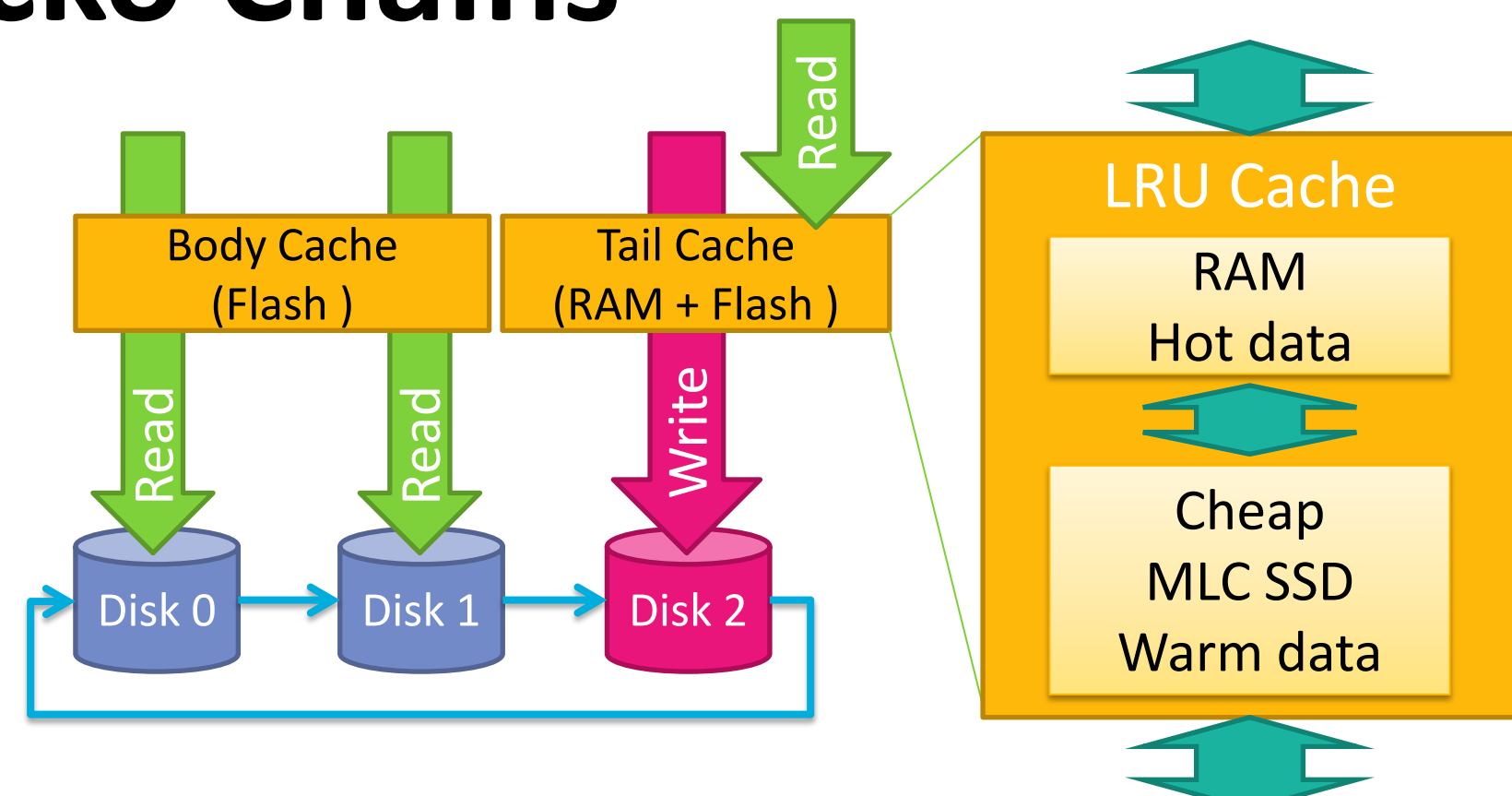
Gecko: A Chain Logging Design

- Logs to one disk at a time to reduce I/O contention
 - Logging (writing) takes place **ONLY** at the tail drive
 - GC and read occurs at the rest of the drives
- Logging solves
 - **Write-write contention**
 - **Write-GC-write contention**
- Chaining solves/reduces
 - **Write-GC-read contention**
 - **Write-read contention**
- Mirroring/Striping chains enables
 - **Power saving** w/o consistency concerns
 - High performance
 - High fault tolerance



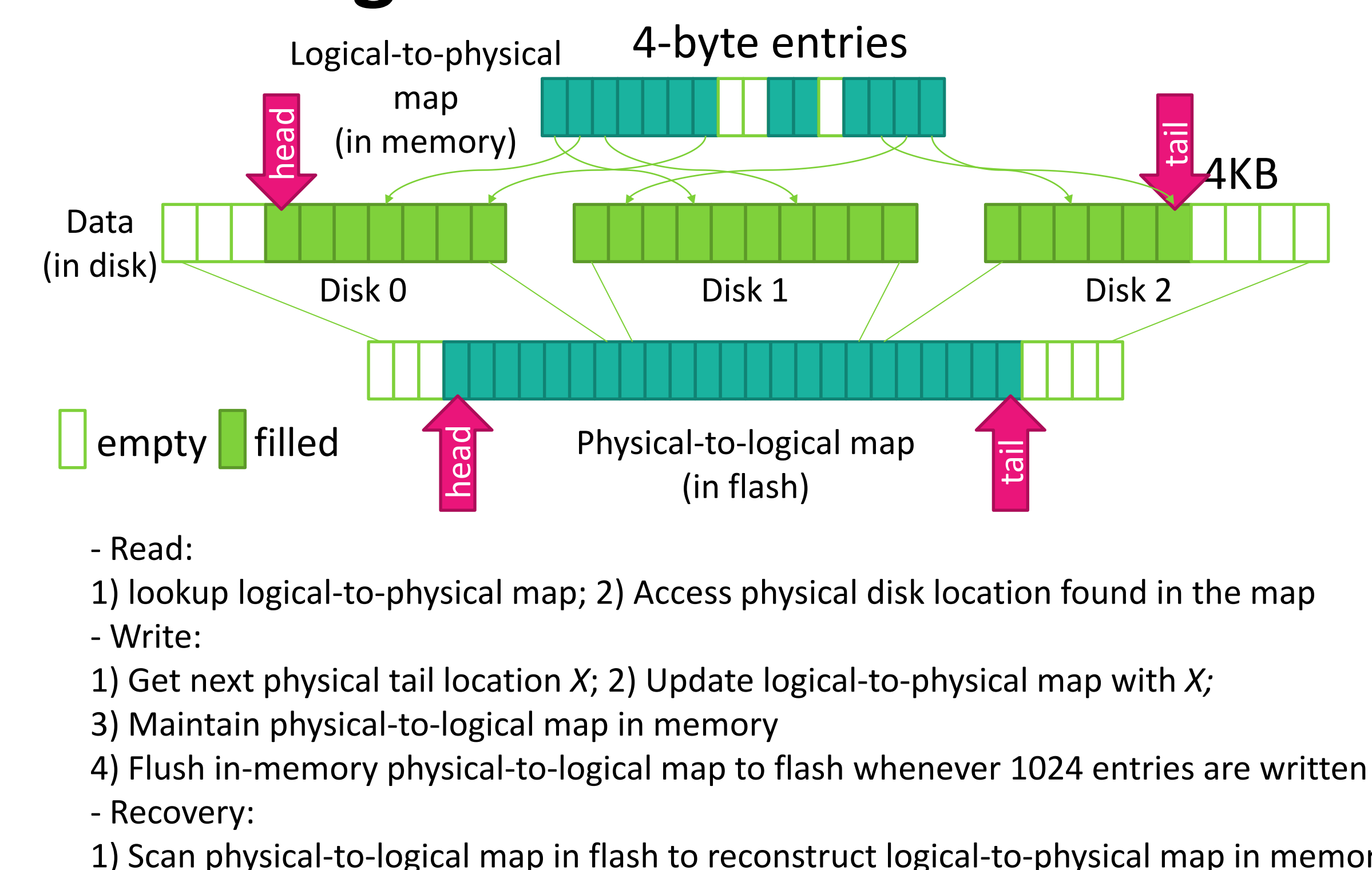
Caching Gecko Chains

- Caching tail drive
 - Reduces **write-read contention**
 - 86% of reads absorbed from mix of MS I/O traces
 - RAM + SSD cache
 - RAM: Small amount of hot data
 - SSD: Large amount of warm data
 - **RAM extends SSD lifetime by 8X**
- Caching body drives
 - SSD only cache
 - Reduces **read-read contention**
 - Relatively low hit rate



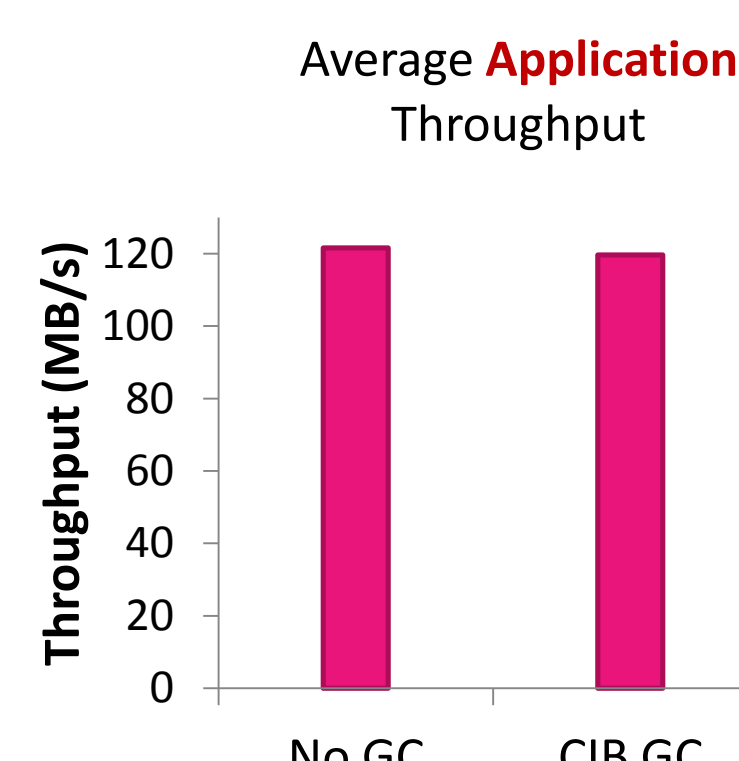
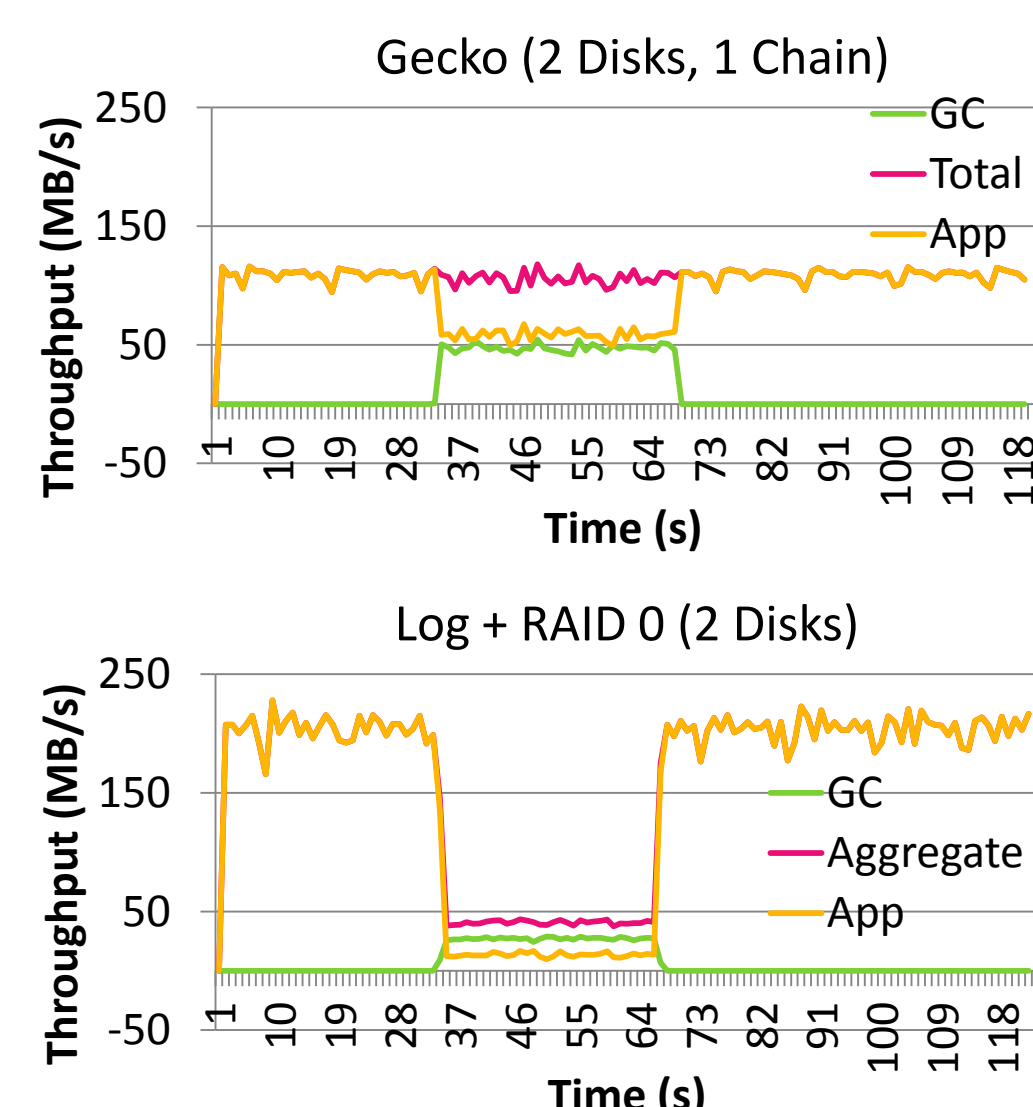
Metadata Management

- In memory logical-to-physical map
 - 4-byte entries per page
 - 8GB for 8TB storage
- In flash physical-to-logical map
 - Maintains persistence
 - Flushed to flash every 1024 page writes
 - Written in sequential order
 - High flash performance
 - Good for flash lifetime



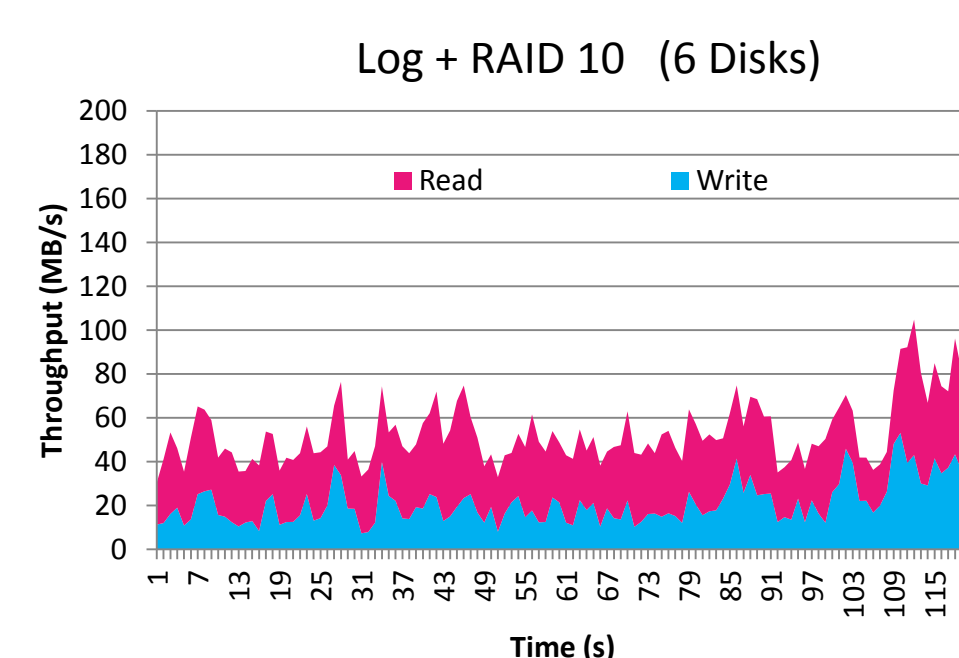
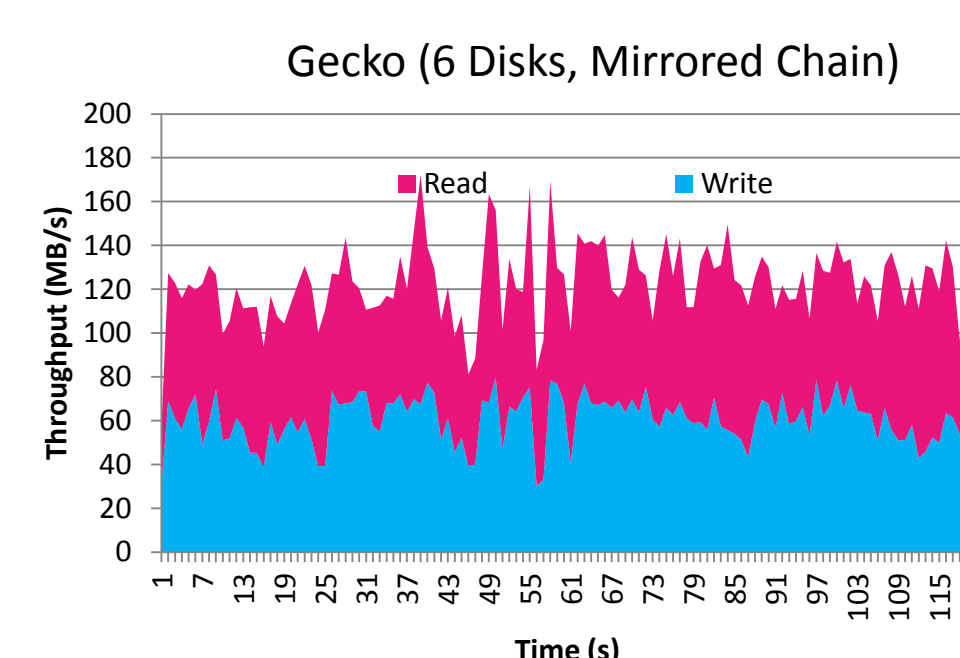
Throughput under GC

- Write only synthetic workload on 2 disks
- **Gecko's aggregate throughput always remains high**



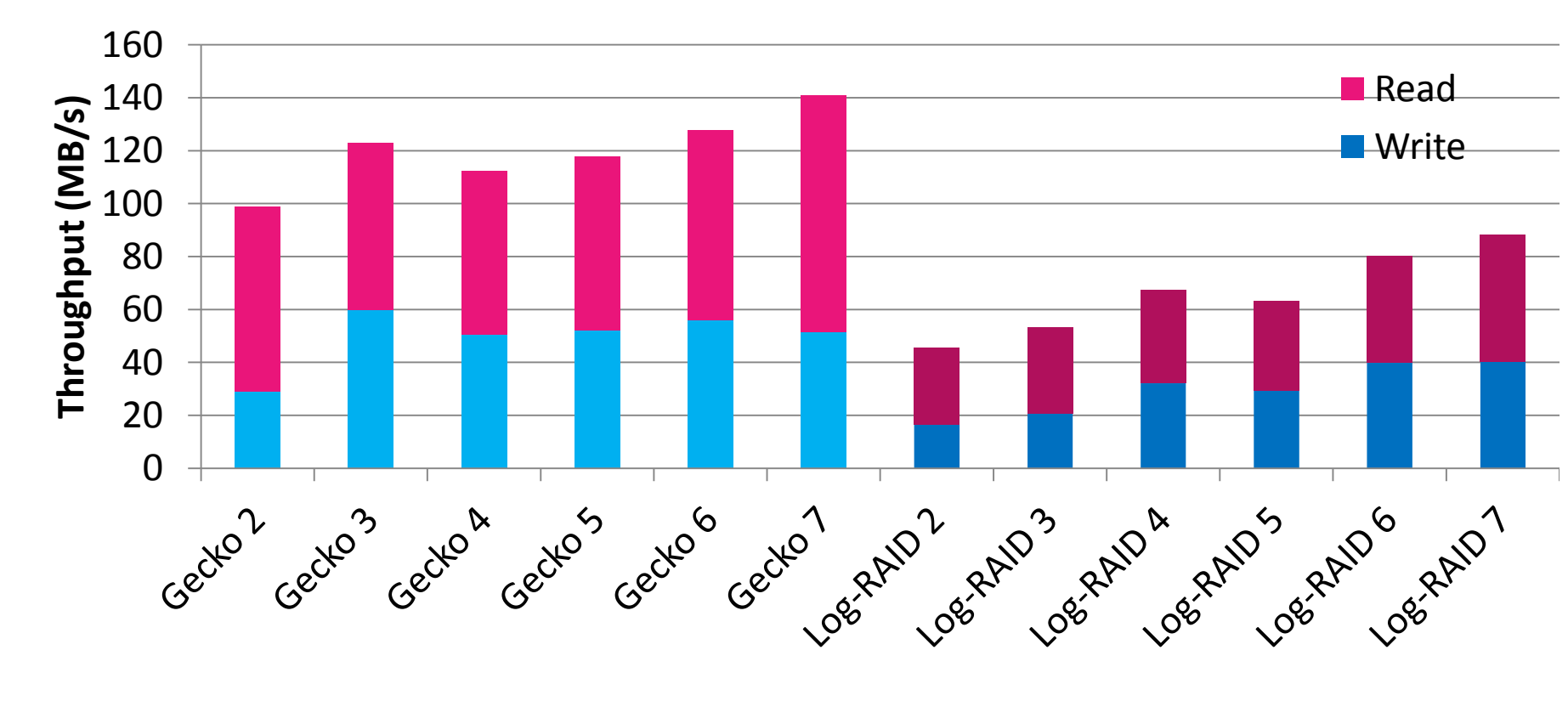
Running Enterprise Workloads

- Mix of 8 MSR Cambridge and MS enterprise workloads
- 6 Disk Configurations
- **Gecko performs**
 - **2-3X better than Log+RAID10**
 - **Order of magnitude better than RAID 0**



Varying Chain Length

- **SINGLE uncontended disk performs better than SEVEN contended disks**
- **Separating read and write reduces contention**
- Typically 3 to 4 disk chains perform reasonably



Conclusion

- Gecko maintains high I/O performance
 - Securing single uncontended disk
 - Separating reads and writes
- Log-structured designs
 - Oblivious to write-write contention
 - Sensitive to GC-write and read-write contention
- Gecko fixes GC-write and read-write contention
 - Log tail and head separation using chain logging
 - Use of RAM+SSD tail disk cache