



SANTA CLARA UNIVERSITY
THE JESUIT UNIVERSITY IN SILICON VALLEY



Data Layout and Management for Shingled Magnetic Recording

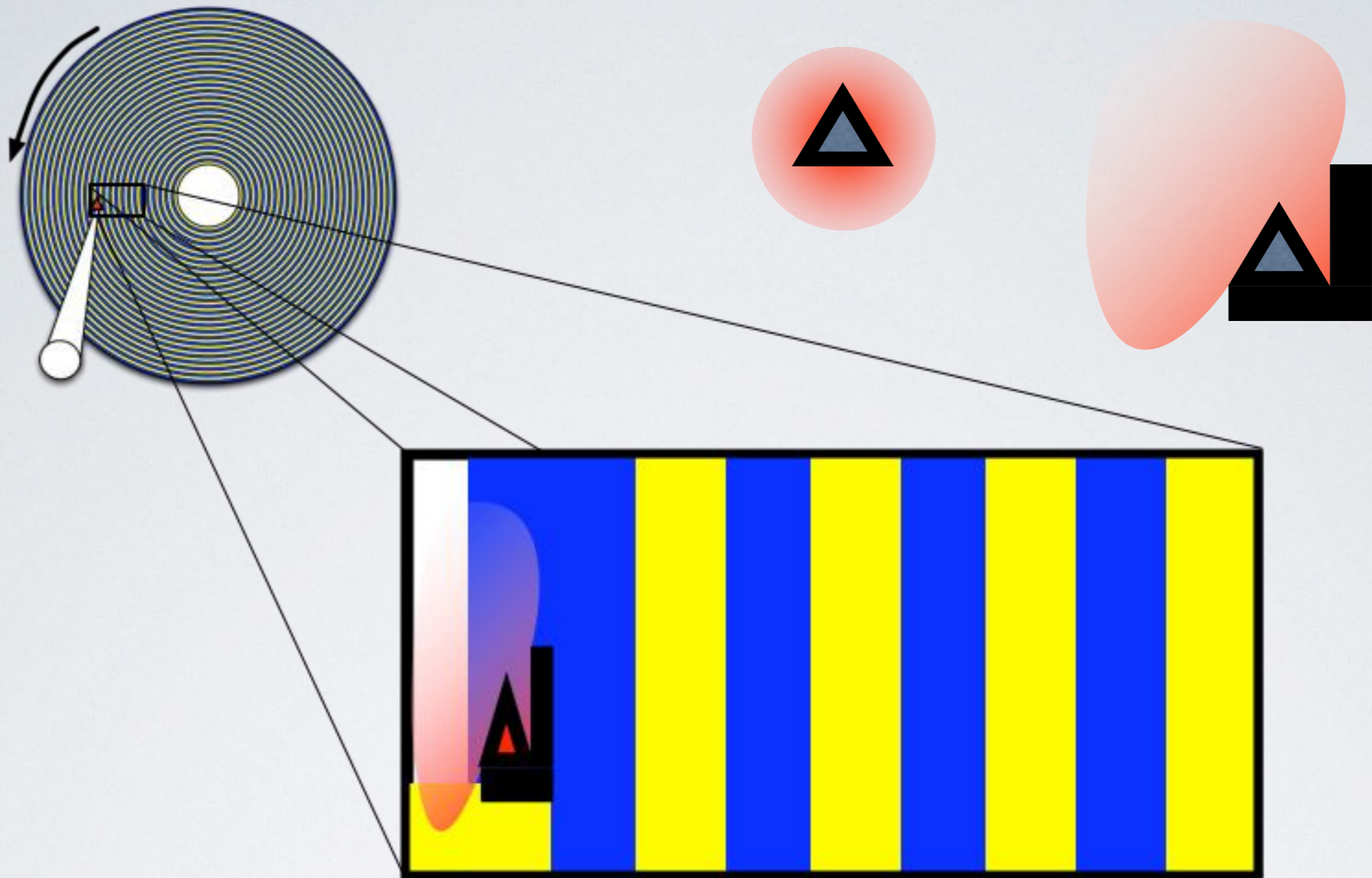
Ahmed Amer, JoAnne Holliday, Andy Hospodor,
Darrell Long, Ethan Miller,
Jehan-Francois Paris, Thomas Schwarz



Baskin
Engineering
UC SANTA CRUZ



INTRODUCTION



SHINGLED WRITING

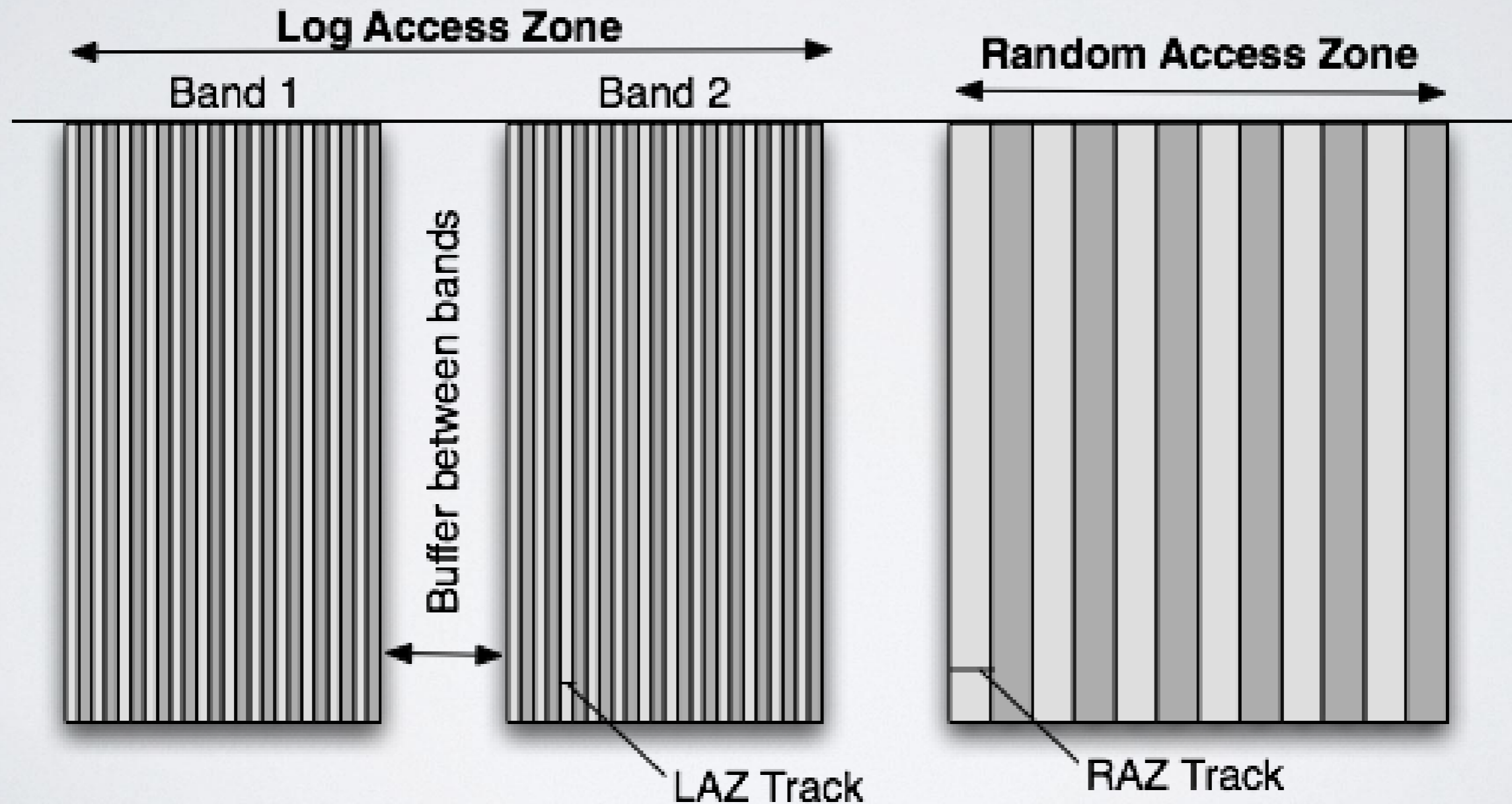
Thermal Stability vs. Writability

REDIRECTION SCHEMES

BASIC REMAPPING SCHEMES

- BANDS vs. Random Update Areas
- Log-structuring
 - Circular Logs
 - Segmented Logs
 - Hybrids and Hierarchies

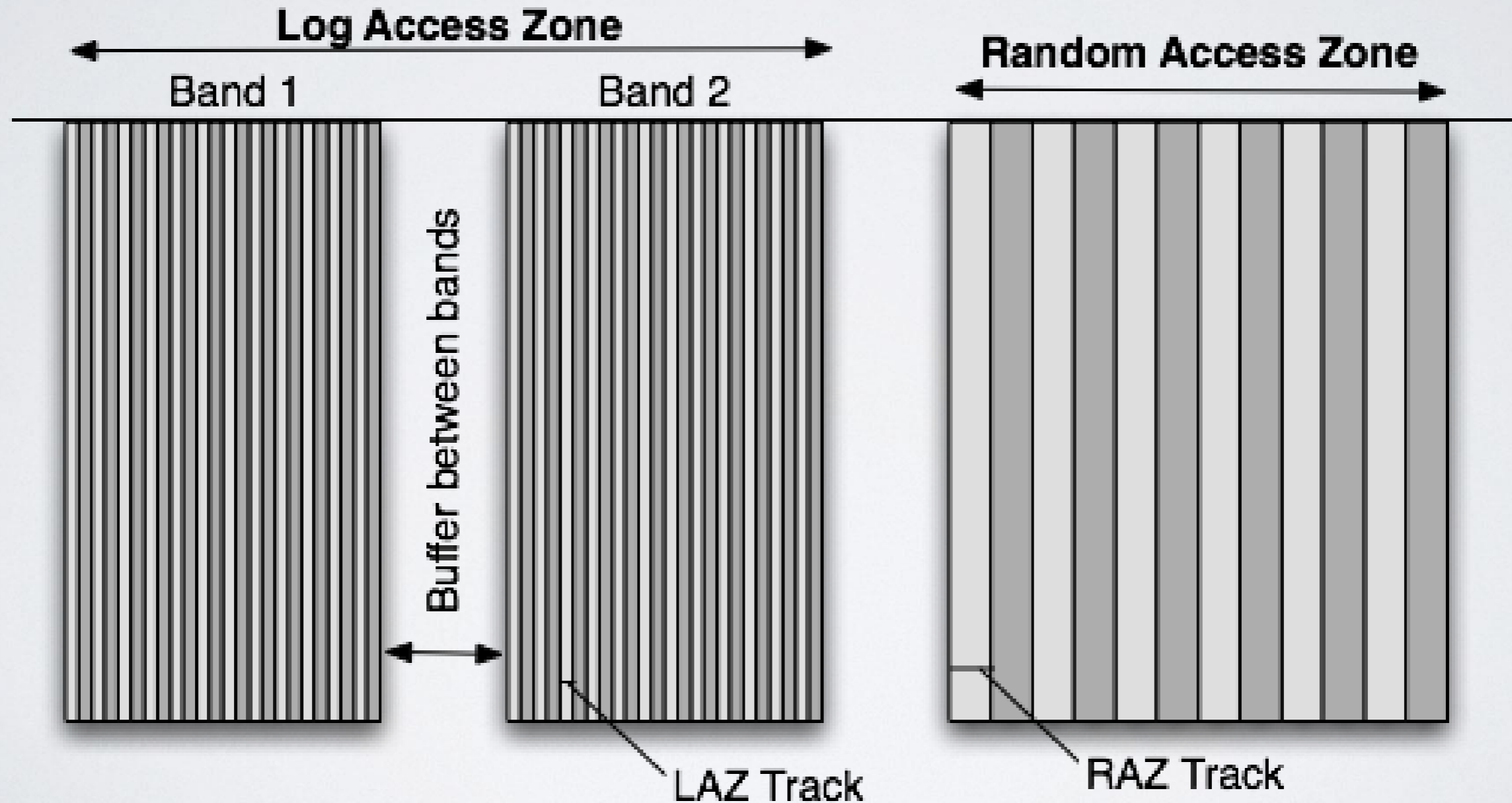
LOG-ACCESS & RANDOM ACCESS ZONES



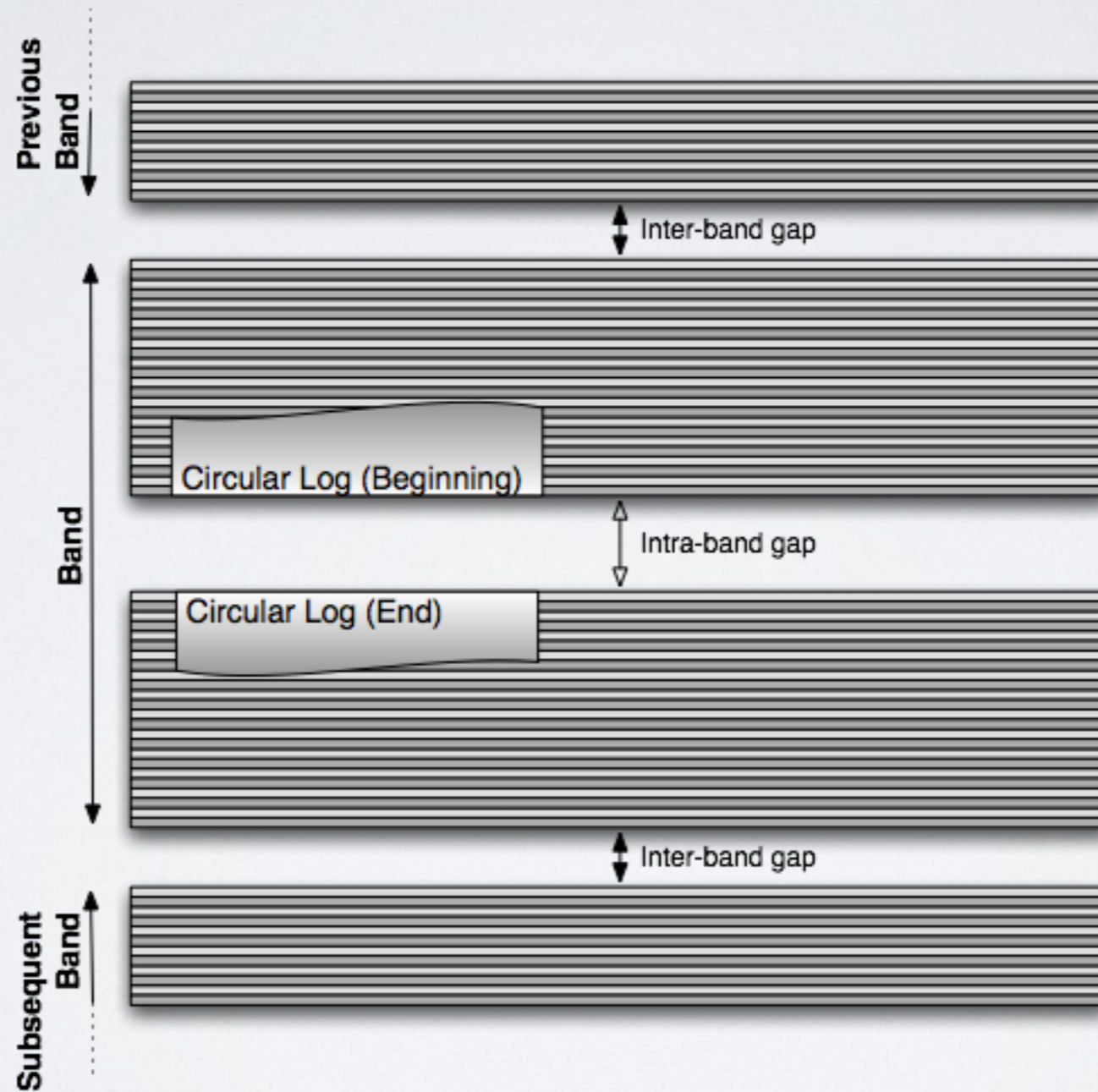
LOGICAL VIEW

- Append-only or destructive overwrite
- Divide into **bands**
- Allowing for an intra-band gap of **k** tracks
 - allows update-in-place **of the band**
- With adjustable bands, an SWD is logically similar to
 - tape with very flexible partition sizes
 - flash with variable bank sizes

LOG-ACCESS & RANDOM ACCESS ZONES

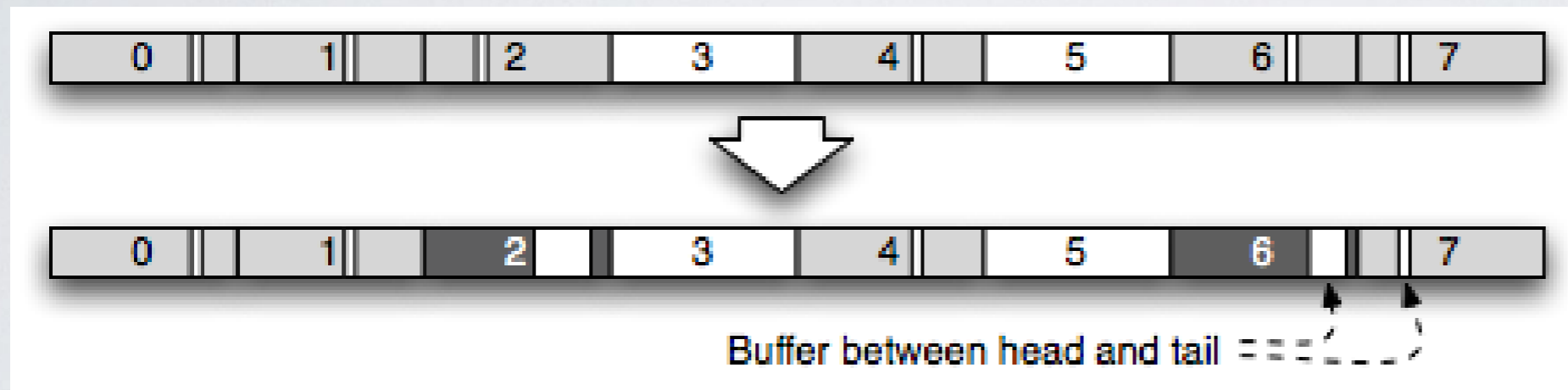


CIRCULAR LOG

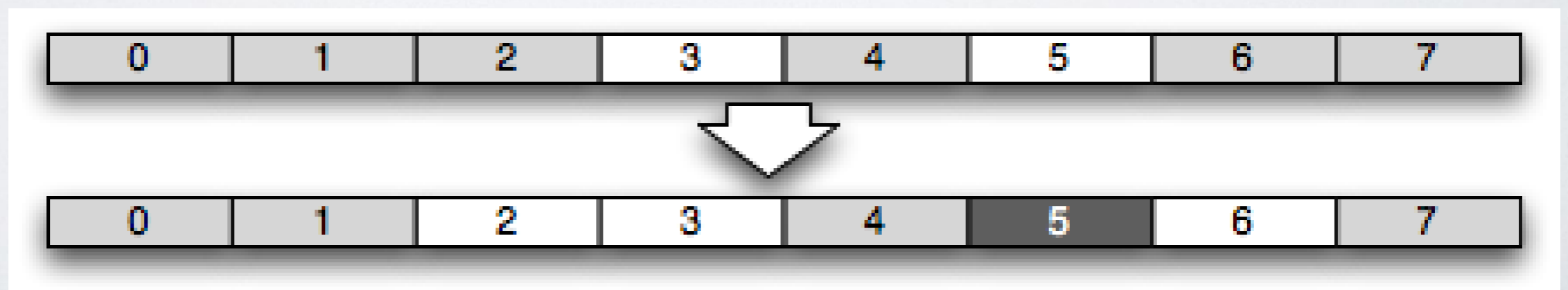


LOGS VS SEGMENTS

Example: Compacting Bands 2 & 6



In-band compaction with circular logs



Compaction of bands to new segments

WORKLOAD EFFECTS

3 MAJOR WORKLOAD ELEMENTS

- READ Operations
- WRITE Operations
- UPDATE Operations
 - WRITE Operations to previously written blocks.
 - It is no longer adequate to gauge the update rate as equivalent to a write rate.

BASIC RELOCATION SCHEMES

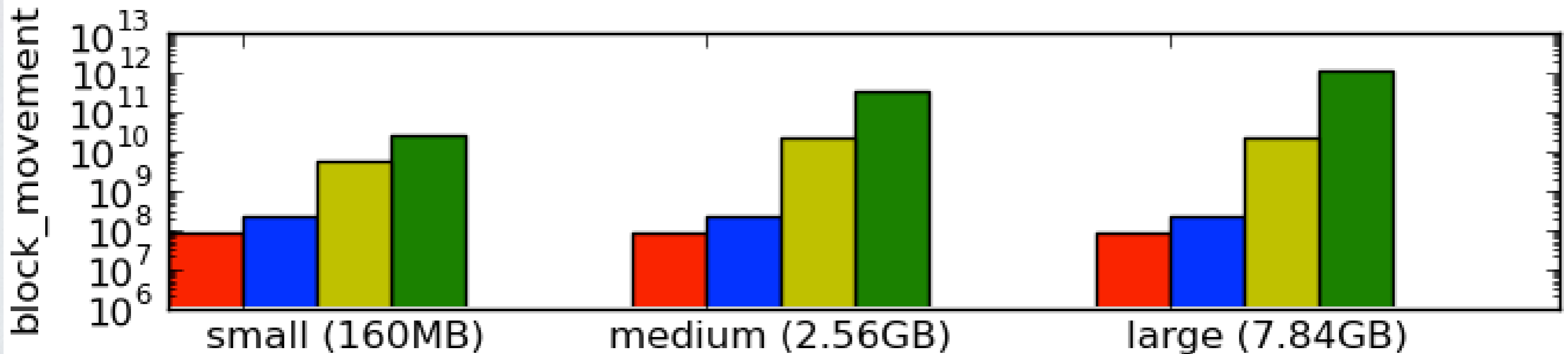
- SCHEME 0: NONE - no relocation necessary
- SCHEME 1: PURE LOGGING - logging with no overheads
- SCHEME 2: RE-ORGANIZATION - with **no** memory limit
- SCHEME 3: RE-ORGANIZATION - with memory limit

WUR - WORKLOAD

- Heavy update workload
- Most writes are unique writes to different blocks
- Very few reads or non-update writes

WUR - WORKLOAD

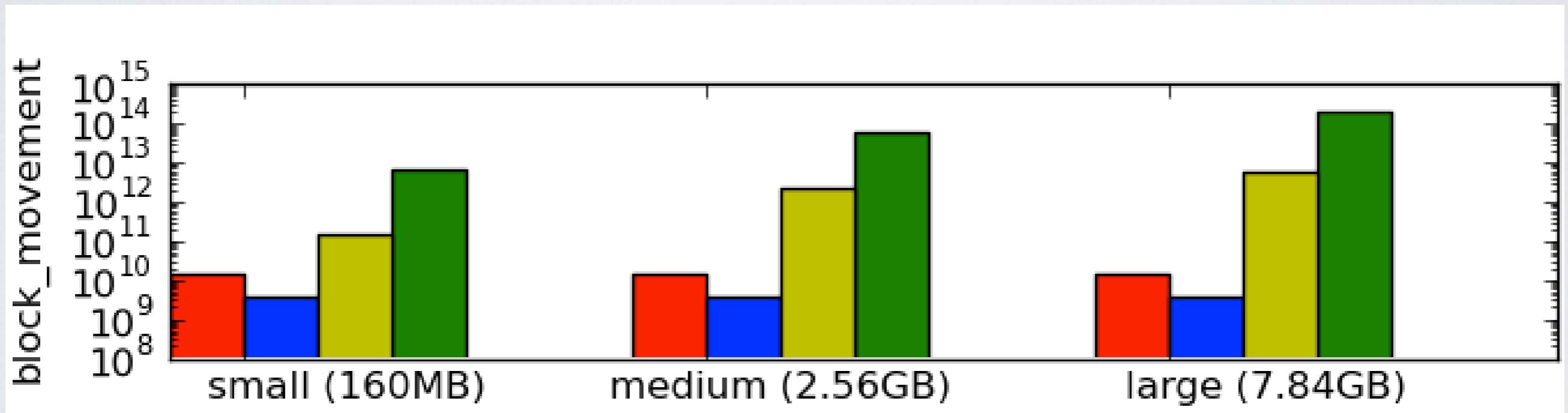
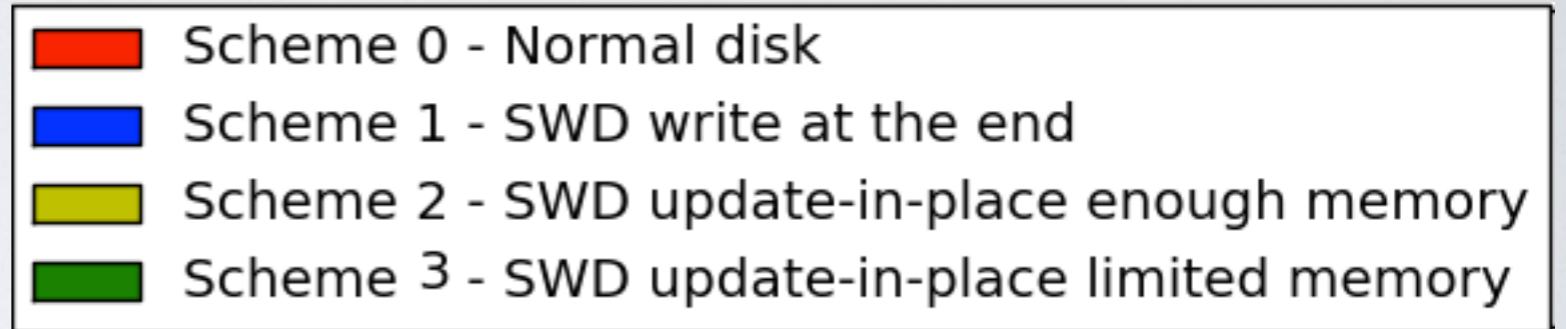
- Scheme 0 - Normal disk
- Scheme 1 - SWD write at the end
- Scheme 2 - SWD update-in-place enough memory
- Scheme 3 - SWD update-in-place limited memory



UWR - WORKLOAD

- Heavy update workload
- Most writes are repeated writes to the same blocks
- Very few reads or non-update writes

UWR - WORKLOAD

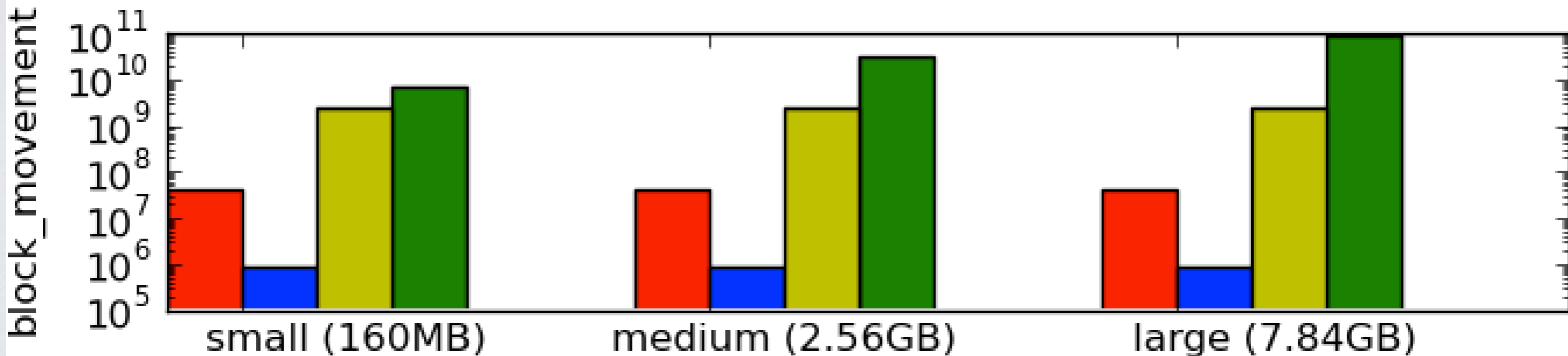


U+WR - WORKLOAD

- Heavy update workload
- Most writes are repeated writes to the same blocks
- Very few reads or non-update writes

U+WR - WORKLOAD

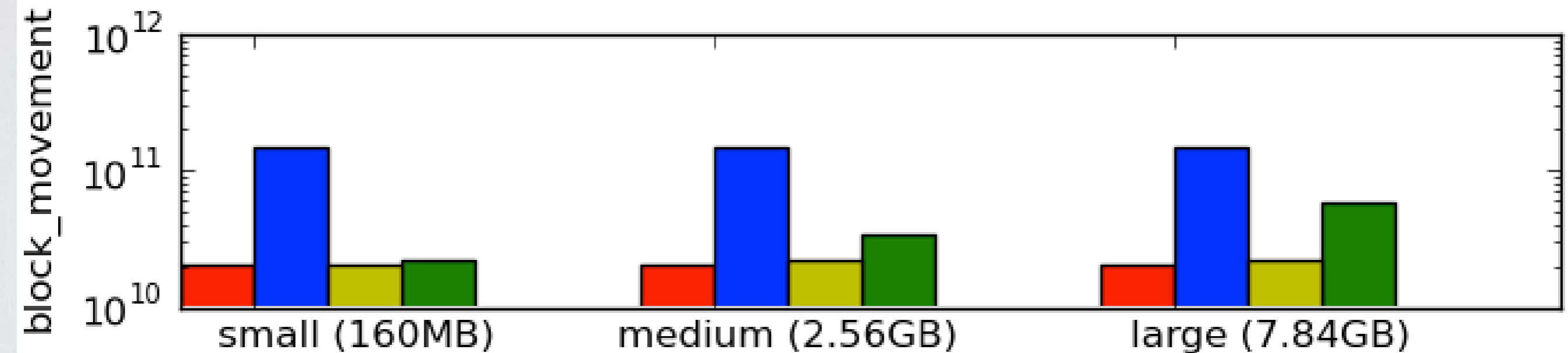
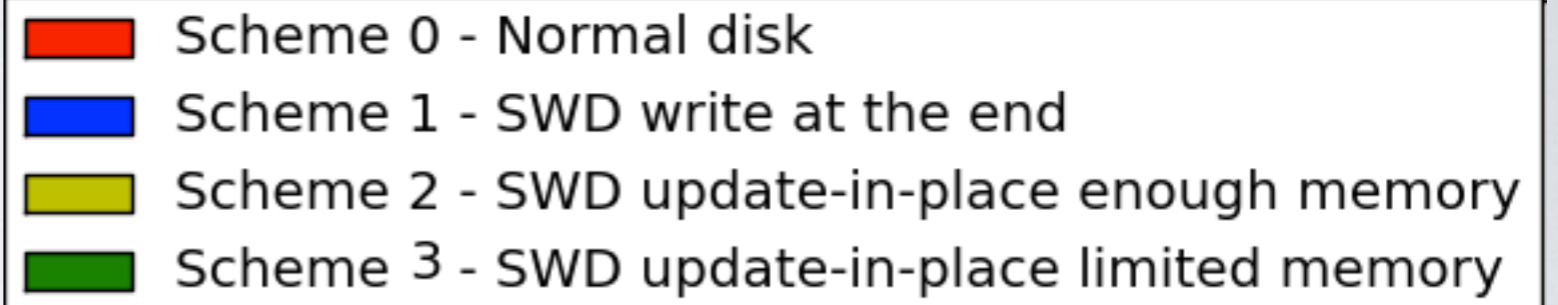
- Scheme 0 - Normal disk
- Scheme 1 - SWD write at the end
- Scheme 2 - SWD update-in-place enough memory
- Scheme 3 - SWD update-in-place limited memory



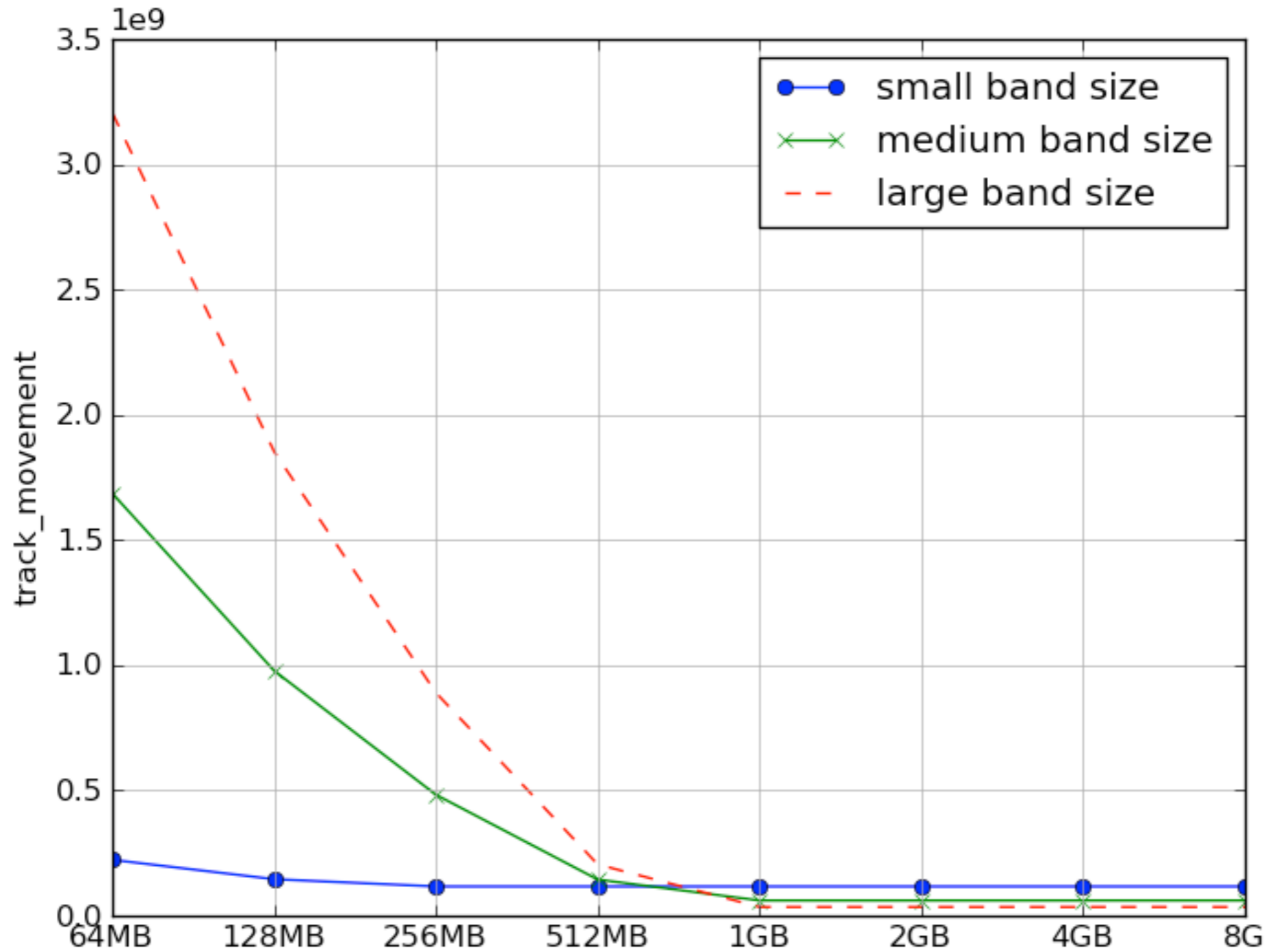
RRR - WORKLOAD

- Heavy read workload
- Very few writes

RRR - WORKLOAD



NVRAM BUFFERING



TDMR IMPACT

- TDMR = increased density (+)
- TDMR = increased latency (-)
 - Multiple rotational delays will be incurred
 - Functionally = Increased perceived seek times

CONCLUSIONS

- Shingled Writing is a promising new recording technology
- Redirection and log-structuring of writes
 - Circular logs or Segment-based band division
 - In-band or intra-band cleaning
- Self-sufficient and hybrid usage options
 - Metadata separation results support object-based interfaces and hybrid usage
 - Self-sufficient usage feasible with appropriate workloads
- While application is dominant factor, general purpose usage appears promising

INTERLEAVED WORKLOADS & INTERFACES

INTERFACE IMPACT

- **Block Level**

- Transparent implementation
- Drop-in device replacement

- **Object Level**

- Meta-data vs. data
- Reduced activity (free space awareness)

- **File Level (& Application-specific)**



Evaluating the behavior of shingled disks when used in an array configuration or when faced with heavily interleaved workloads from multiple sources.

Initial Findings

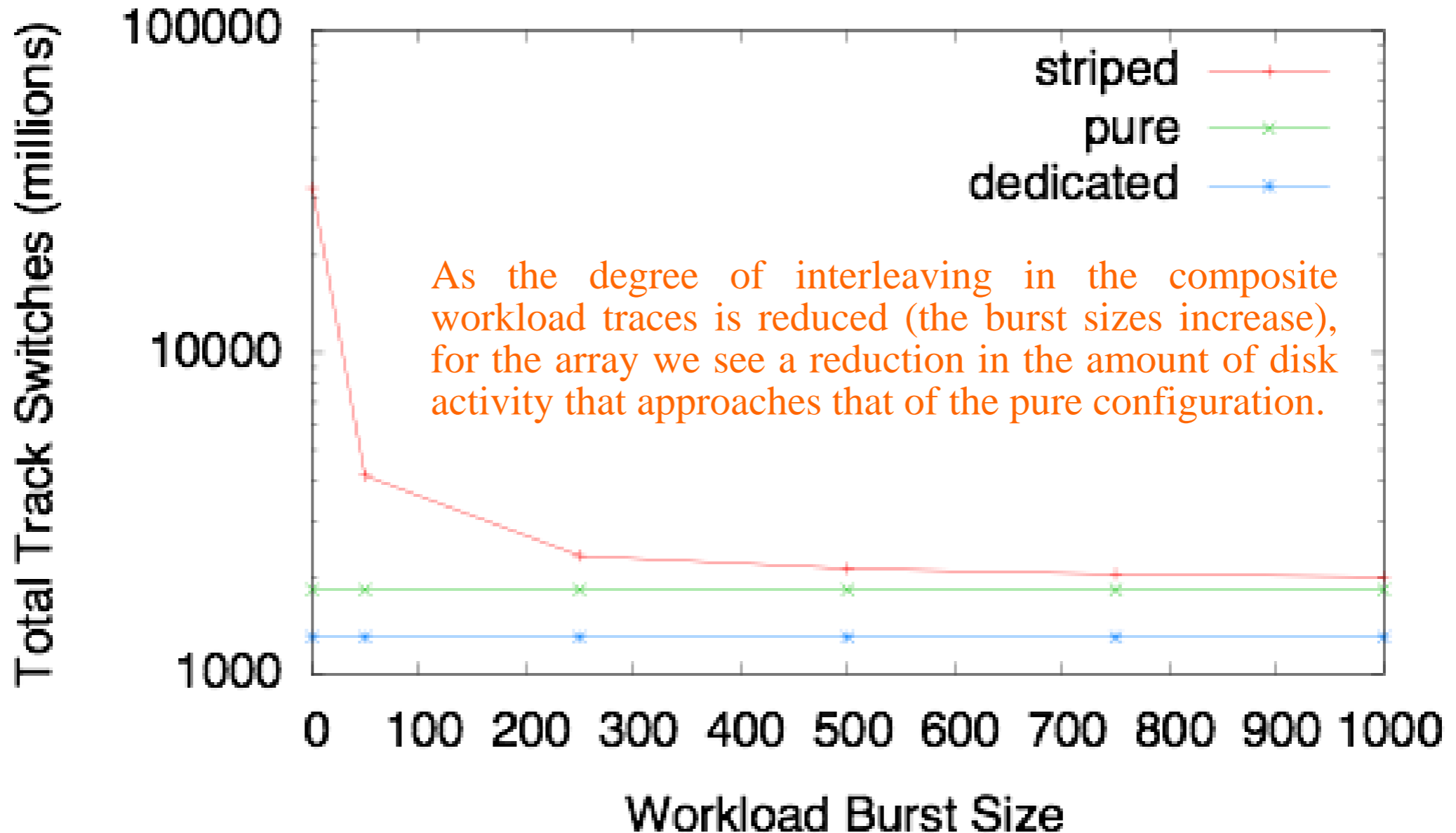
- Heavily interleaved workloads can have a dramatic negative impact on disk activity.
- Reducing interleaving has a significant positive effect.



Workload-Based Evaluation

- **Striped workload**
 - using composite of four workloads
 - workload mix varied by adjusting a random interleave
- **Pure workload**
 - disks arranged in sequence
 - time-varying workload, but not interleaving
- **Dedicated workload**
 - disks dedicated to individual workload sources
 - unlike "pure" and "striped" workloads: no interleaving per-disk

The measurement is total track switches across four disks



- Disk Layout Options (in lieu of basic striped arrays):
 - Dedicated disks and bands
 - Workload differentiation

INTERFACE IMPACT

- Block Level
- Object Level
- **File Level (& Application-specific)**
 - Object semantic knowledge + name-space awareness
 - Application-specific optimizations

QUESTIONS?