香港中文大學
The Chinese University of Hong Kong

CSCI5550 Advanced File and Storage Systems

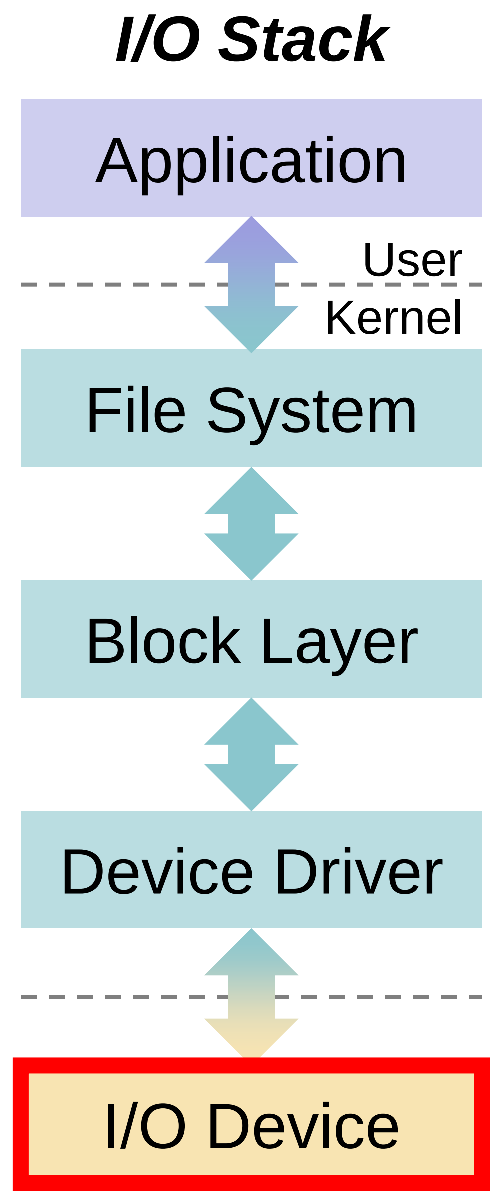
Lecture 07: Next-generation Hard Disk Drive

Ming-Chang YANG

mcyang@cse.cuhk.edu.hk

Outline

- Traditional Hard Disk Drive
  - Why and How
  - Development Bottleneck
- New Magnetic Recording Technologies
- Shingled Magnetic Recording (SMR)
  - Basics and Inherent Challenges
  - General Solution: Persistent Cache
  - Various SMR Drive Models and Designs
    - Drive-Managed SMR (DM-SMR)
    - Host-Aware SMR (HA-SMR)
    - Host-Managed SMR (HM-SMR)
    - Hybrid SMR



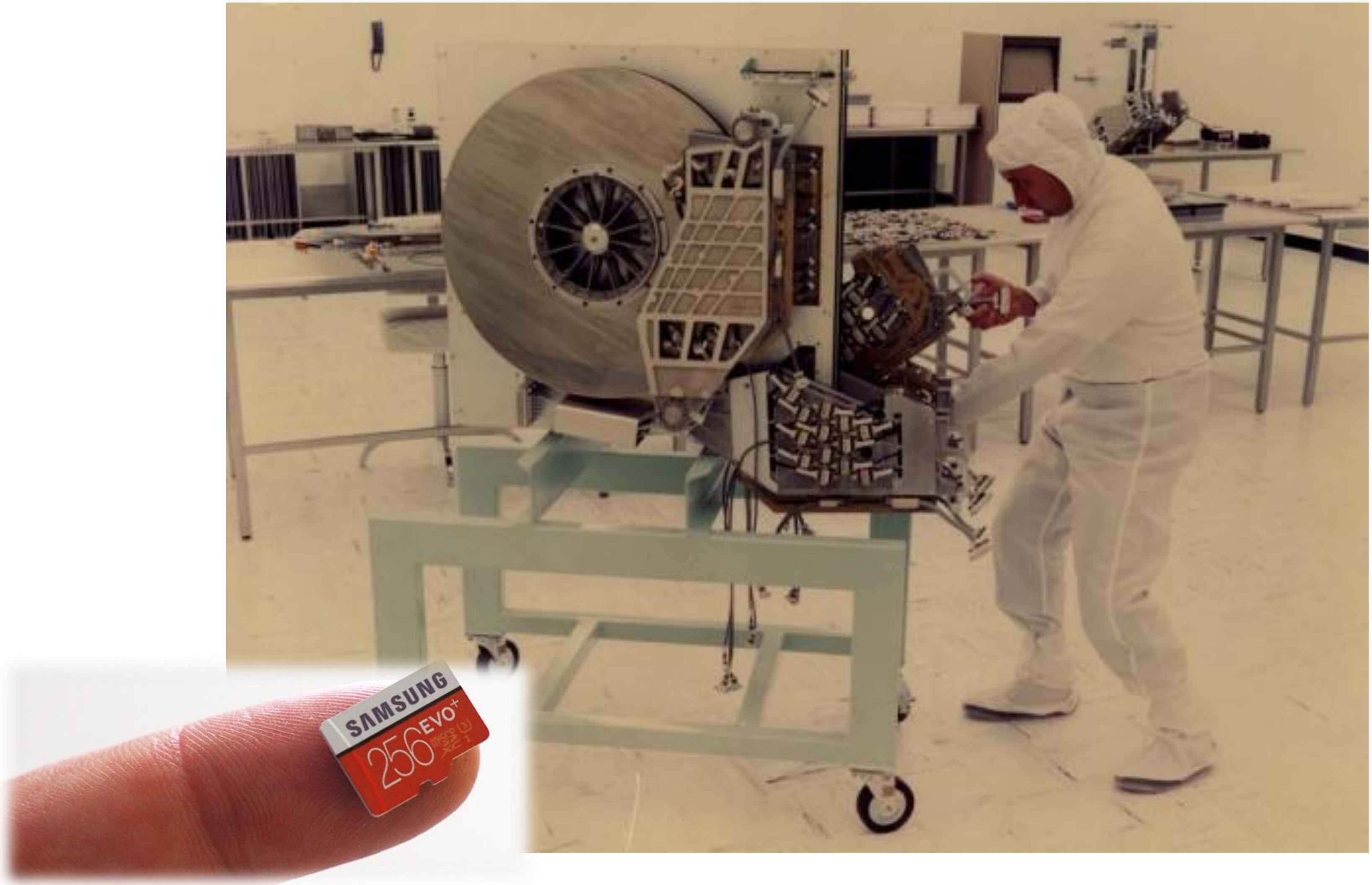
History of Hard Disk Drives

- HDDs have been the main form of **persistent data storage** in computer systems for decades.
  - In 1953, IBM recognized the urgent need.
  - The first commercial usage of HDD began in 1957.

- Many file and storage systems are designed and optimized based on HDD characteristics.

Amazing Photos about HDD

- Below is a 250 MB hard disk drive in 1979 …



Amazing Photos about HDD

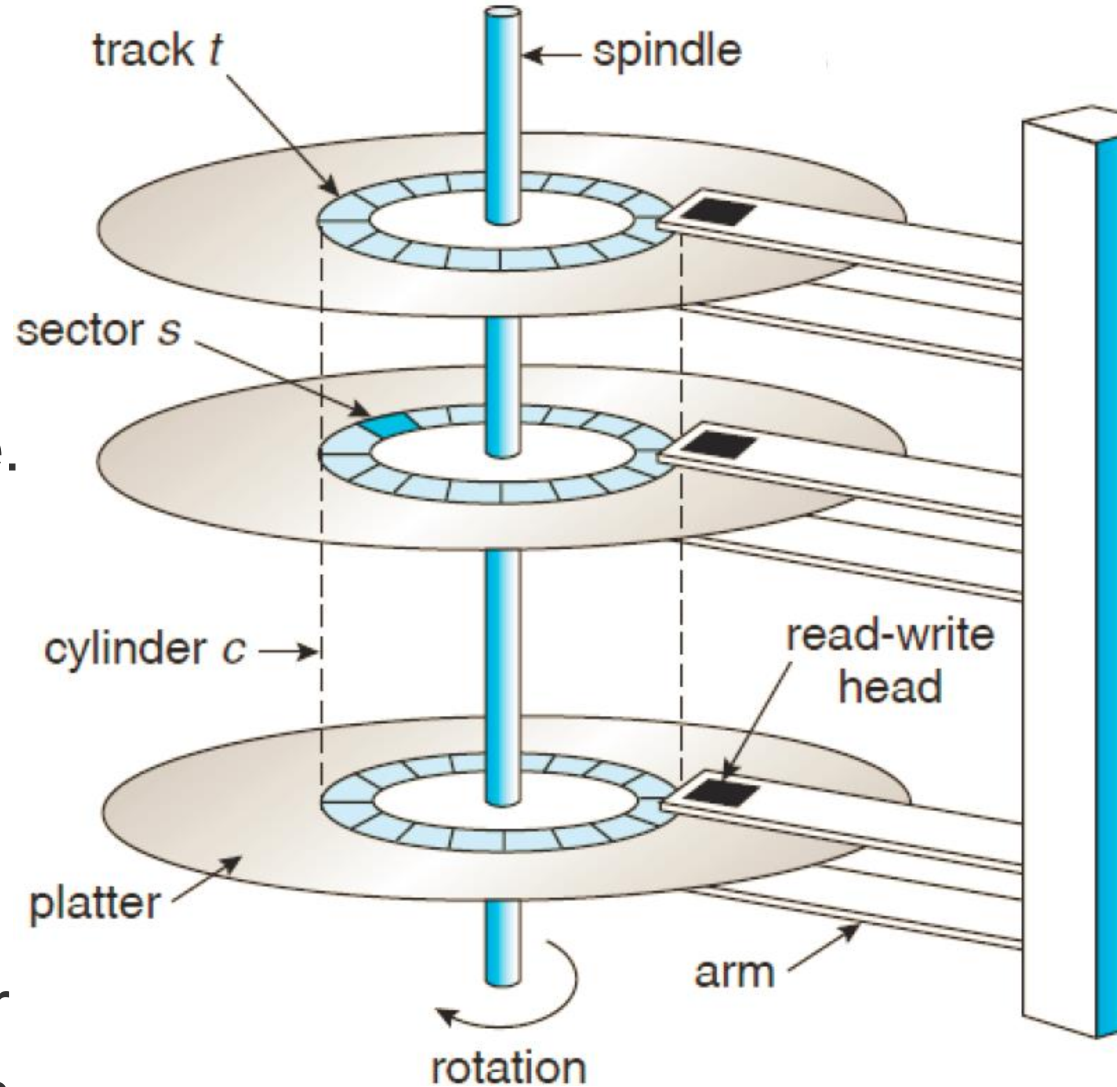
- From ‘80s to today: **8-inch → 3.5, 2.5, 1.8**-inch drives

Recall: Disk Organization: Logical View

- HDD: Accessed in **blocks** but organized in **sectors**
  - **Sector**
    - The most common sector size is 512 bytes.
      - The sector size is fixed on an HDD.
    - All sectors are numbered from 0 to $n - 1$ (i.e., the **address space**).
      - The disk can be logically viewed as **an array of n sectors**.
  - **Block**
    - Disk I/Os are in units of **blocks**.
    - A **block** may refer to one or multiple sectors.
- In an HDD, only a single 512-byte write is **atomic**.
  - It will either complete in entirety or fail at all.
    - Torn Write: Only a portion of a larger write may complete.

Recall: Disk Organization: Physical View

- A hard disk has one or multiple **platters**.
  - Each platter has 2 sides (**surfaces**).
  - Platters are bound together by a **spindle**.
- Each surface has multiple *concentric circles* called **tracks**.
  - A track is further divided into **sectors**.
- A **disk head** reads or writes data of sectors.

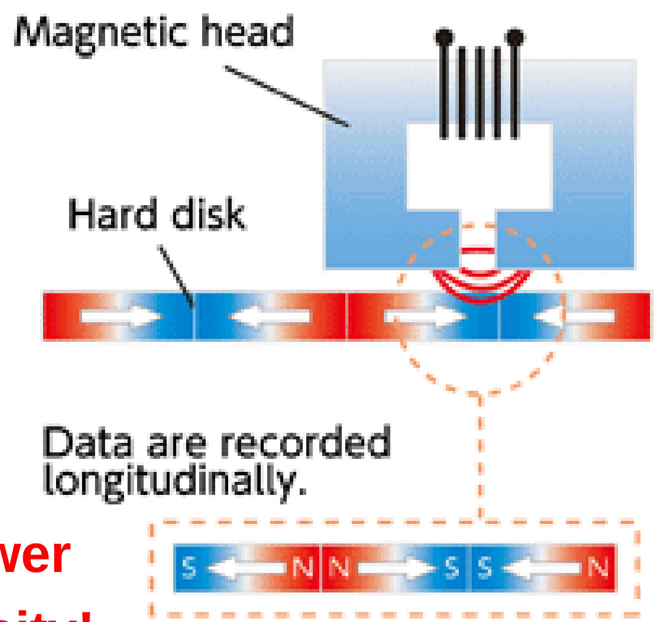


Silberschatz et al., "Operating System Concepts Essential".

Magnetic Recording Technology

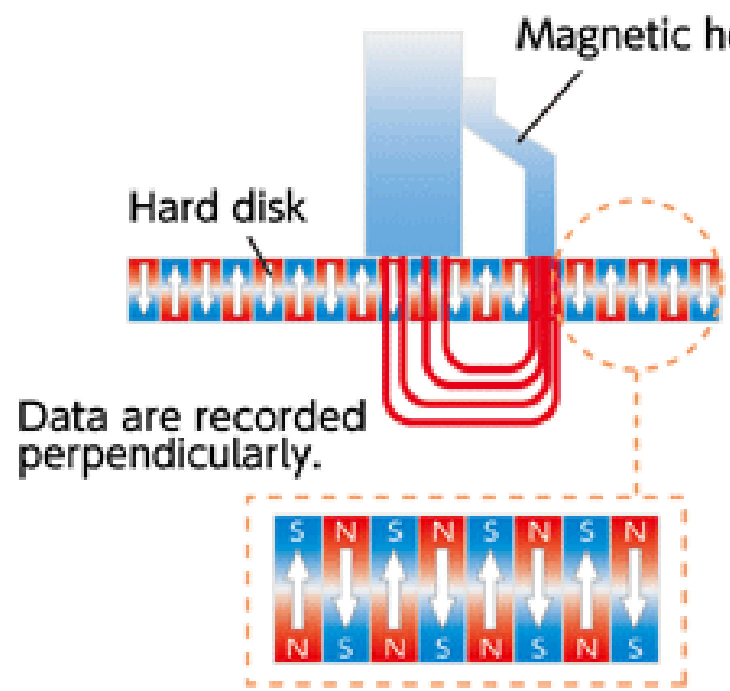
- HDDs store the data as **tiny areas** of either positive or negative magnetization on the disk surfaces.
  - Each tiny area represents a “bit” of information.
  - Based on the magnetization direction, there are two **Conventional Magnetic Recording (CMR)** technologies:

① **L**ongitudinal **M**agnetic **R**ecording



Lower density!

② **P**erpendicular **M**agnetic **R**ecording



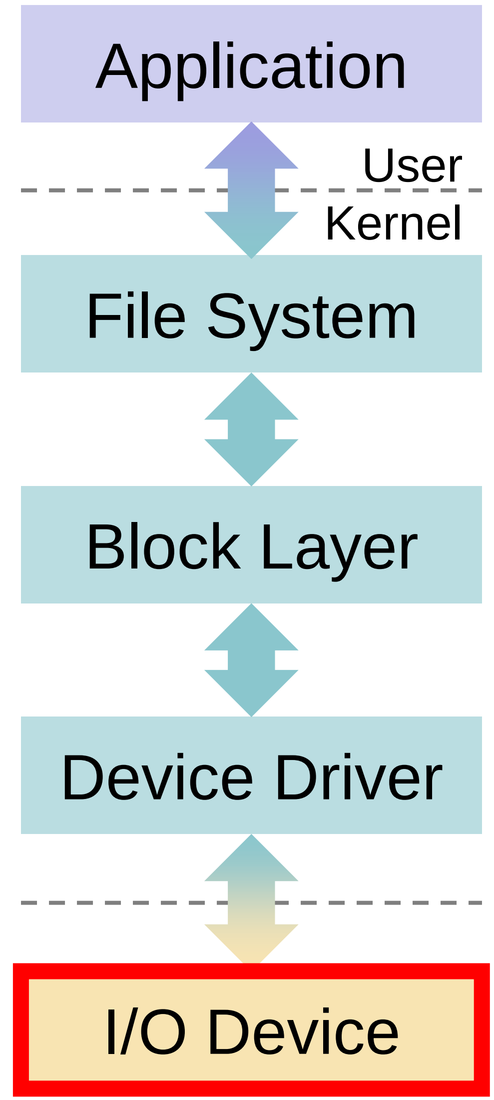
Higher density!

http://www.spring8.or.jp/en/news_publications/research_highlights/no_51/

Outline

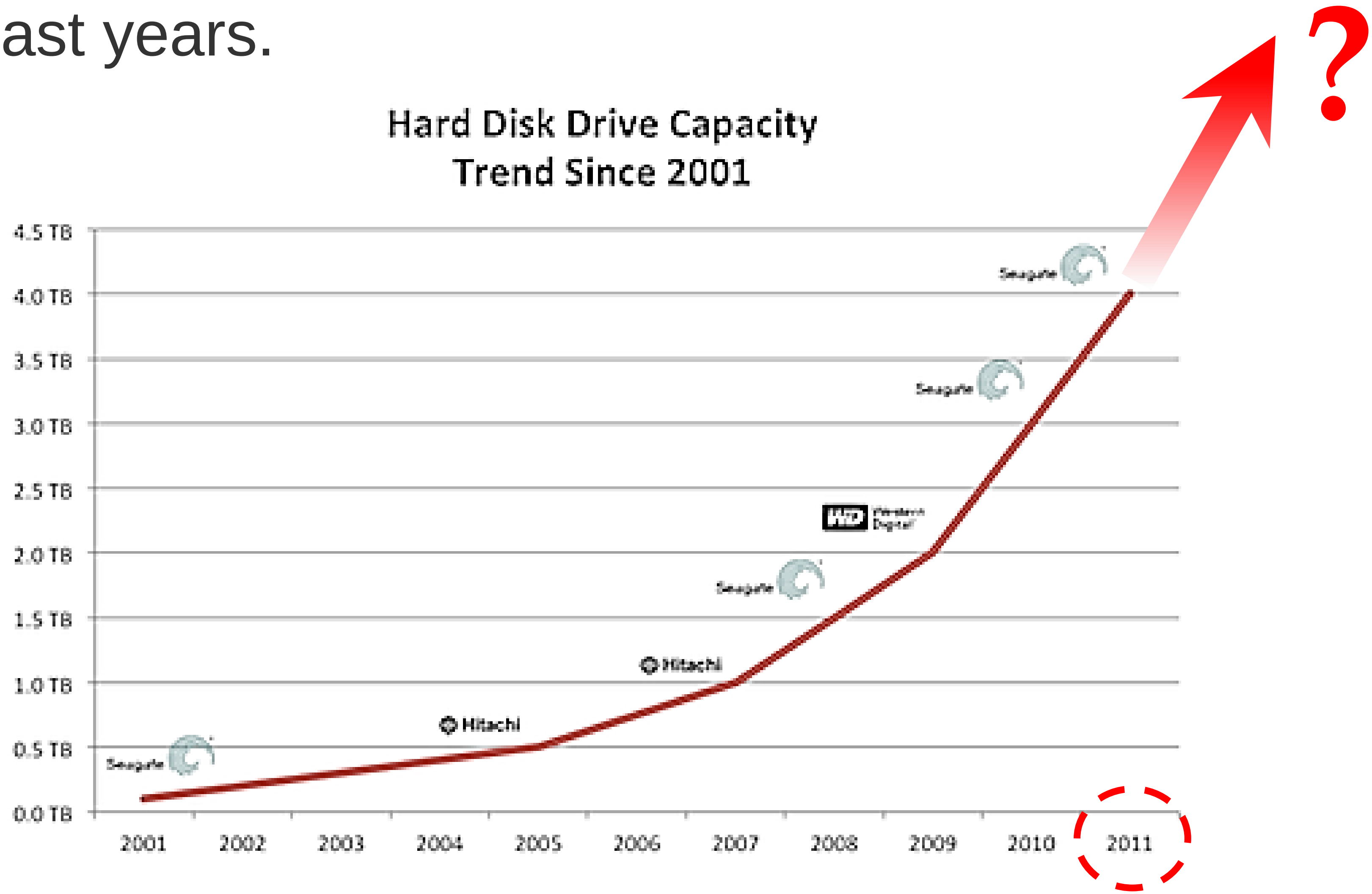
- Traditional Hard Disk Drive
  - Why and How
  - Development Bottleneck
- New Magnetic Recording Technologies
- Shingled Magnetic Recording (SMR)
  - Basics and Inherent Challenges
  - General Solution: Persistent Cache
  - Various SMR Drive Models and Designs
    - Drive-Managed SMR (DM-SMR)
    - Host-Aware SMR (HA-SMR)
    - Host-Managed SMR (HM-SMR)
    - Hybrid SMR

I/O Stack



HDD Capacity Trend

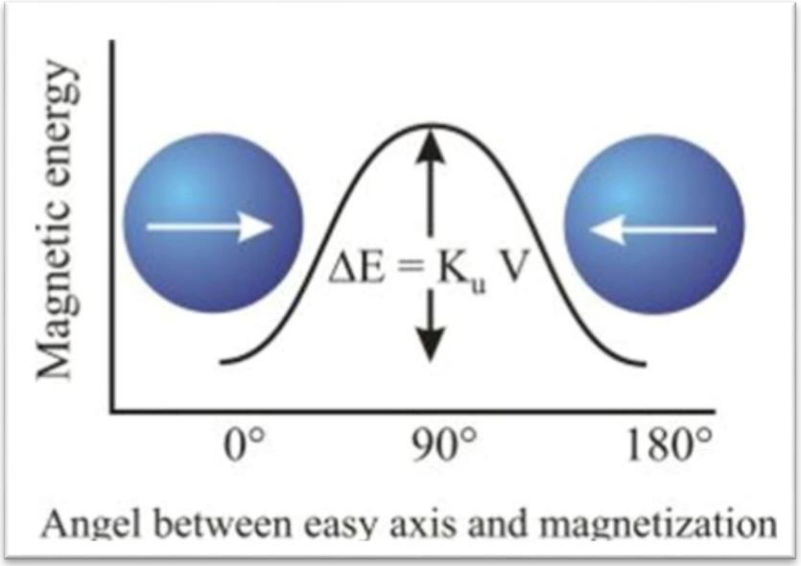
- The capacity of HDDs increased at breakneck speed in past years.

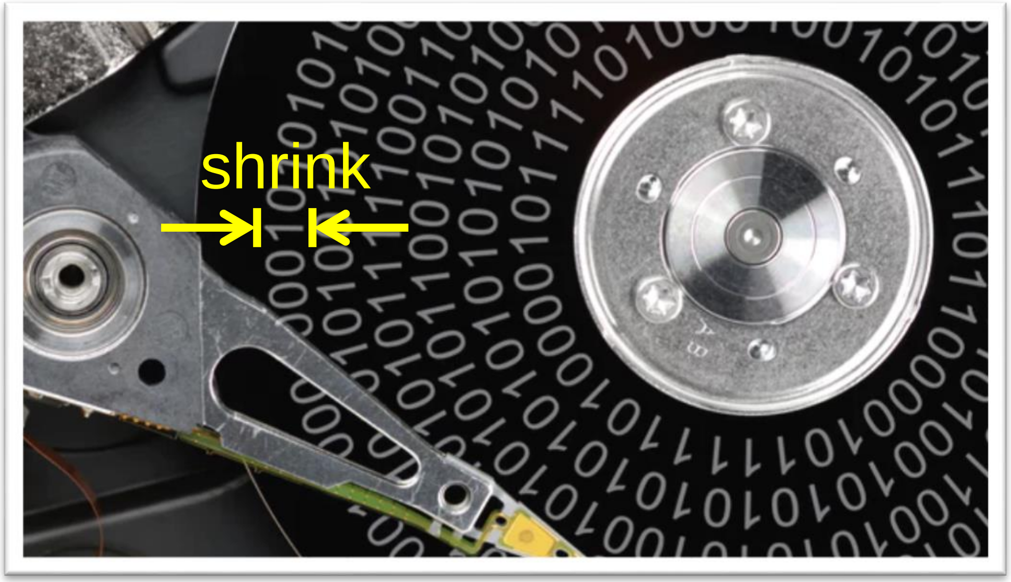


http://blog.fosketts.net/2011/09/08/seagate-goflex-desk-4tb-hitachi-deskstar/

Shrinkage Bottleneck

- PMR has reached the **bottleneck** in providing higher areal density.
  - The maximal areal density is **1 TB per square inch**.
  - Because of the **superparamagnetic effect (SPE)**, it is hard to continuously shrink the volume of magnetic grains.

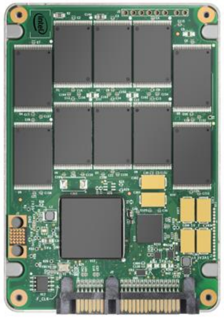




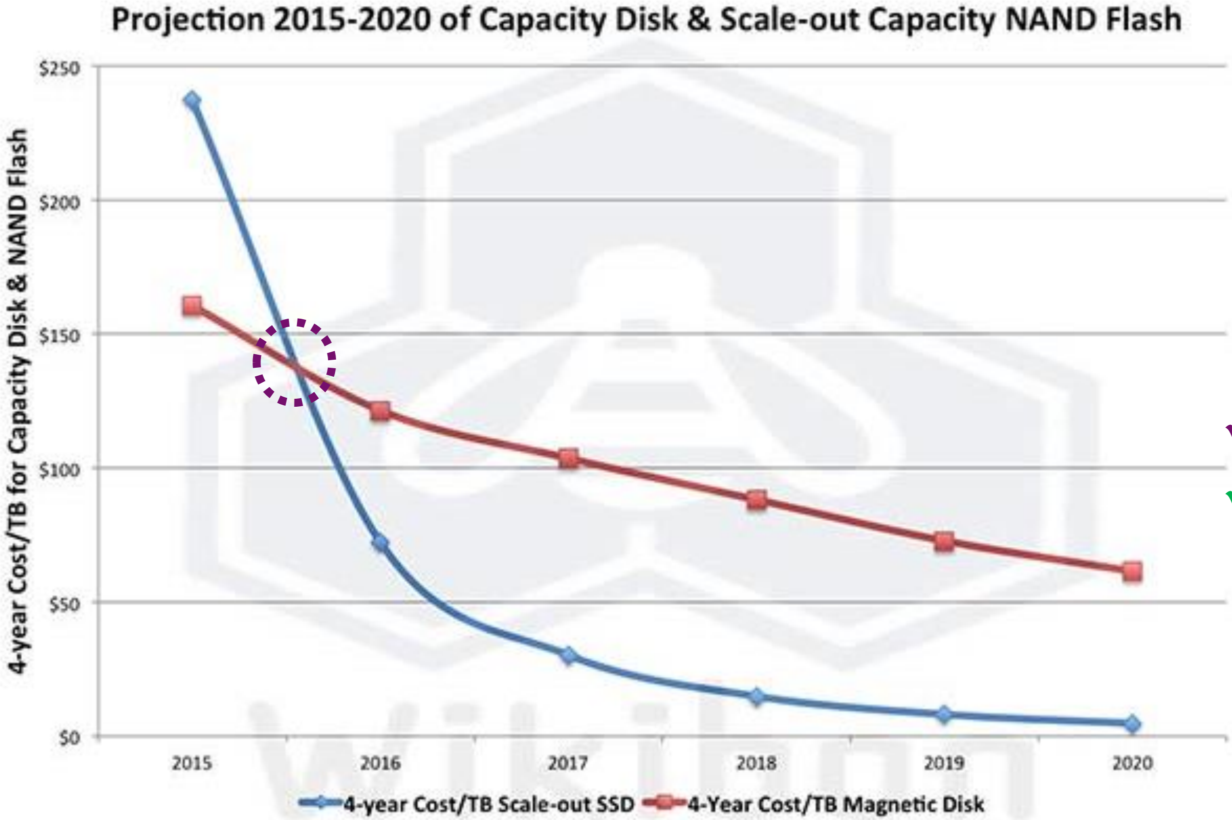
Breaking the Bottleneck or Dying …

- **Strong Rival**: Solid-State Drive (SSD)
  - **Flash memory** demonstrates several good advantages …
  - Flash memory is getting **cheaper** with **degraded reliability**.

SSD



- ✓ Faster
- ✓ Silent
- ✓ Shock-resistant
- ✓ Energy efficient
- ✓ Lighter
- ✓ Smaller



Source: Wikibon 2014. 4-Year Cost/TB Magnetic Disk includes Power, Maintenance, Space & Disk Data Reduction. 4-year Cost/TB SSD includes Power, Maintenance, Space, SSD Data Reduction & Data Sharing.

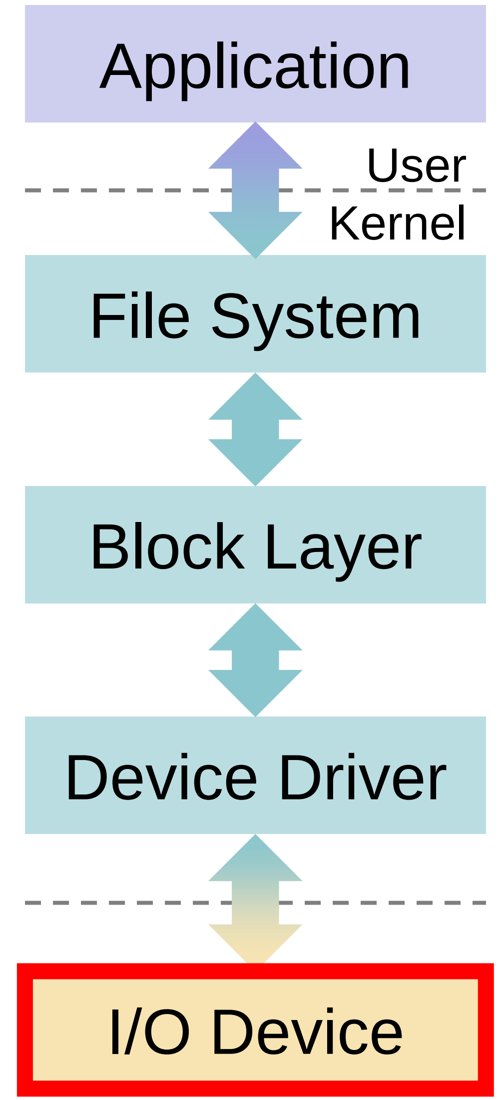
HDD

- ✓ Cheaper?
- ✓ Reliability

Outline

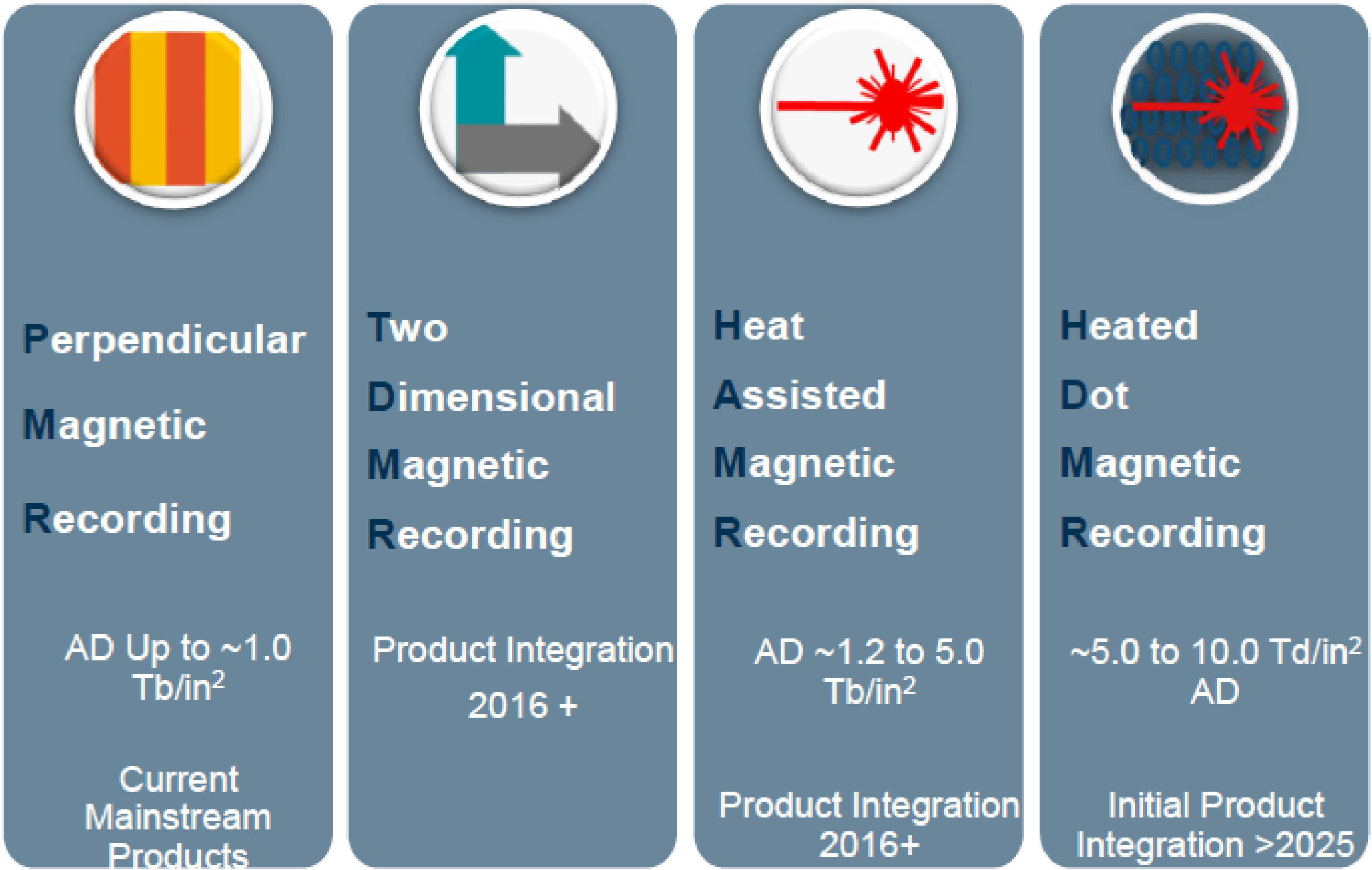
- Traditional Hard Disk Drive
  - Why and How
  - Development Bottleneck
- New Magnetic Recording Technologies
- Shingled Magnetic Recording (SMR)
  - Basics and Inherent Challenges
  - General Solution: Persistent Cache
  - Various SMR Drive Models and Designs
    - Drive-Managed SMR (DM-SMR)
    - Host-Aware SMR (HA-SMR)
    - Host-Managed SMR (HM-SMR)
    - Hybrid SMR

I/O Stack



New Magnetic Recording Technologies

- HDDs are poised to keep evolving for the **increased areal density (AD)** with various new technologies:
  - **Less** things to do with firmware/software ☹

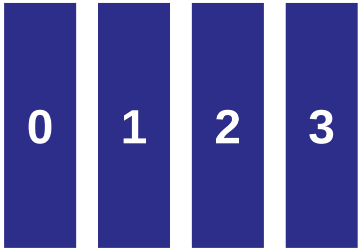


https://www.anandtech.com/show/10470/the-evolution-of-hdds-in-the-near-future-speaking-with-seagate-cto-mark-re

New Track Layouts

- HDDs are poised to keep evolving for the **increased areal density (AD)** with different track layouts:
  - **More** things to do with **firmware/software** ☺

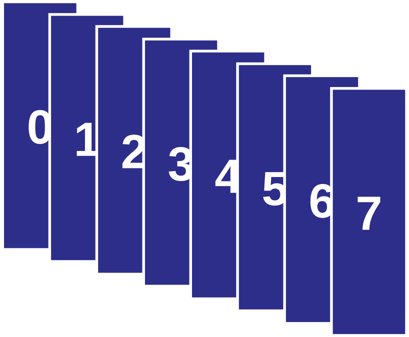
Conventional
Magnetic Recording



✓ Tracks **non-overlap**

➢ **Tracks can be rewritten freely!**

Shingled
Magnetic Recording



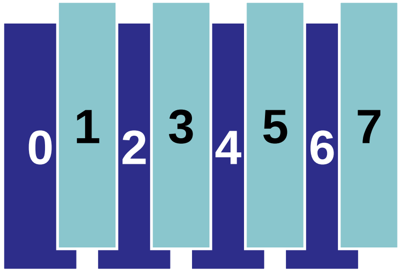
✓ Tracks **overlap**

✓ **25% higher capacity** than CMR

✓ **Commercially Available** for 5 years

✓ **Track rewrite issue**

Interlaced
Magnetic Recording



✓ Tracks **overlap**

✓ **40% higher capacity** than CMR

✓ **Not commercially available yet**

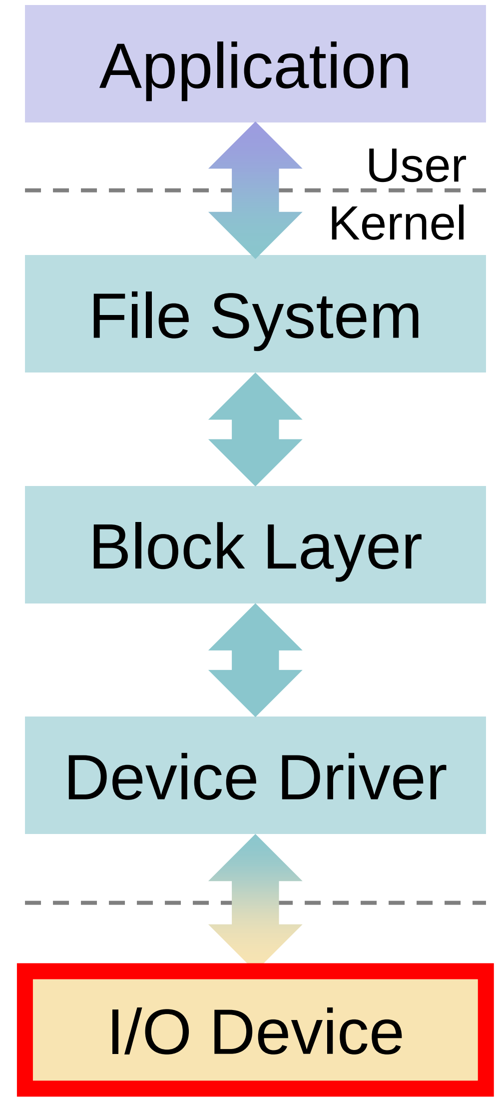
✓ **Track rewrite issue**

https://www.usenix.org/conference/atc19/presentation/hajkazemi

Outline

- Traditional Hard Disk Drive
  - Why and How
  - Development Bottleneck
- New Magnetic Recording Technologies
- Shingled Magnetic Recording (SMR)
  - Basics and Inherent Challenge
  - General Solution: Persistent Cache
  - Various SMR Drive Models and Designs
    - Drive-Managed SMR (DM-SMR)
    - Host-Aware SMR (HA-SMR)
    - Host-Managed SMR (HM-SMR)
    - Hybrid SMR

I/O Stack

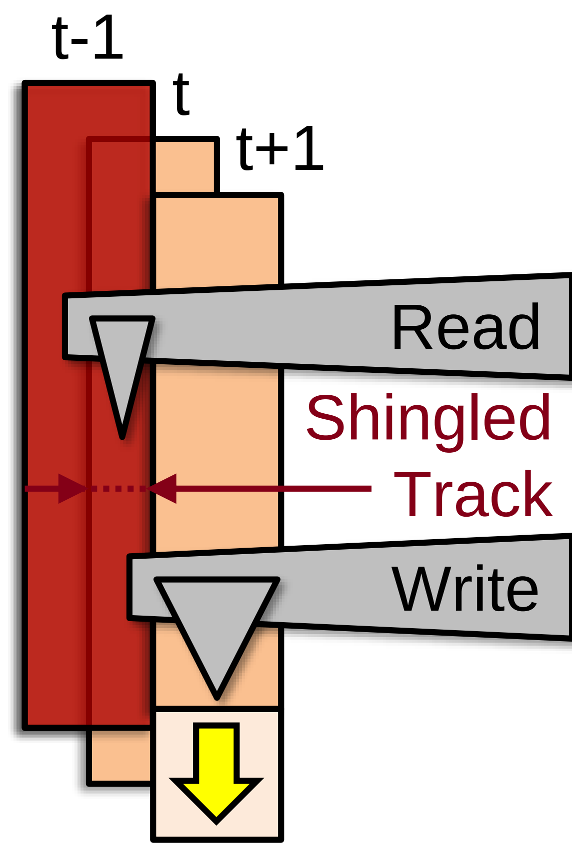


Shingle?

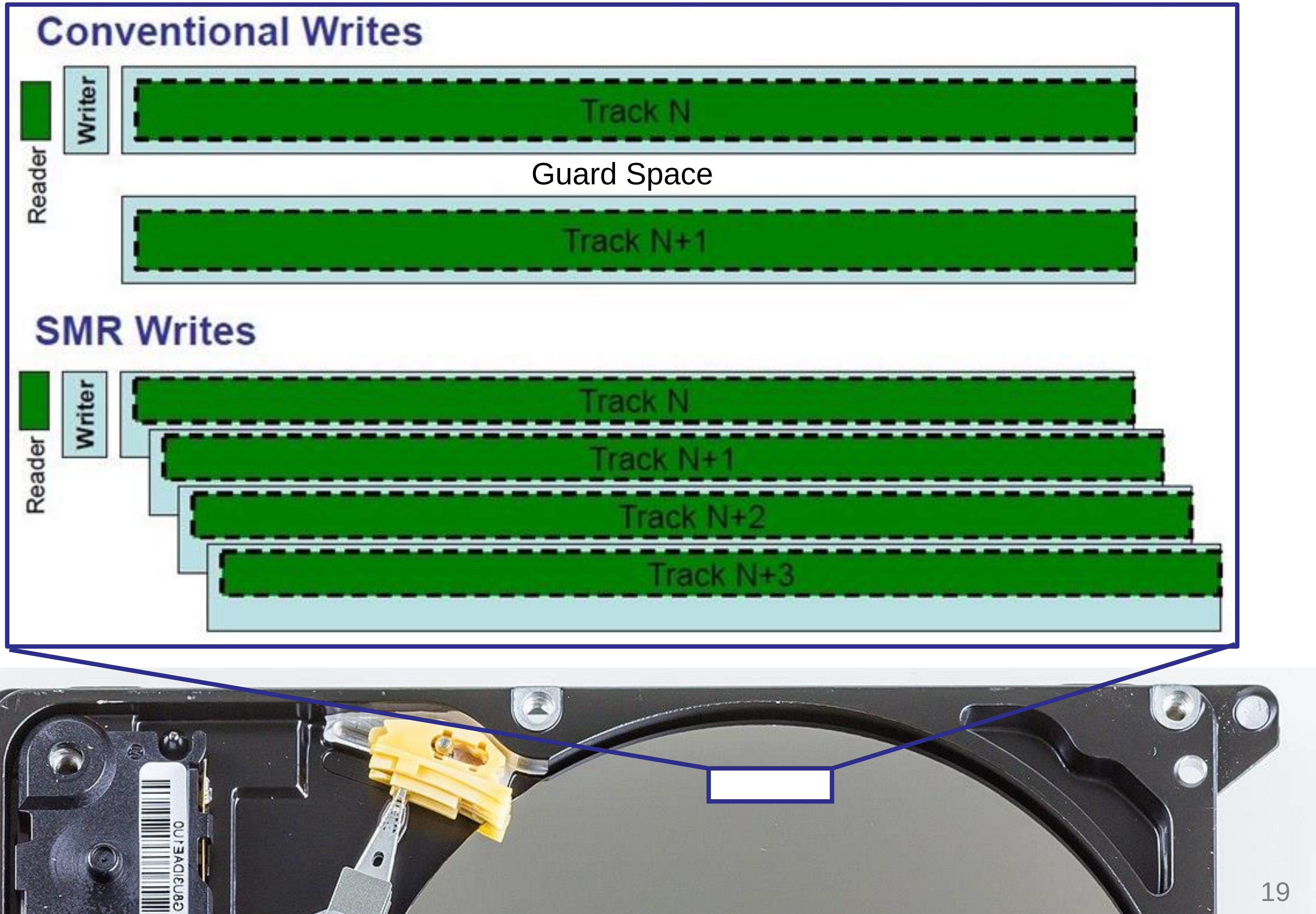
Shingled Magnetic Recording (SMR)

- **Key**: Read head is **more precise** than write head.
- SMR is based on
  - **Writing** in a sequential way with tracks overlapped with the previous ones.
  - **Reading** the “exposed” data from shingled tracks.
- **Advantages** of Shingled Tracks:
  - Areal density↑, Capacity↑, and Cost↓
  - No major changes of disk are needed.
- **Design Challenge**:
  - Updating data to an existing track may destroy data on the subsequent tracks.
  - SMR management is needed.

Shingled Tracks



Conventional vs. Shingled Writes

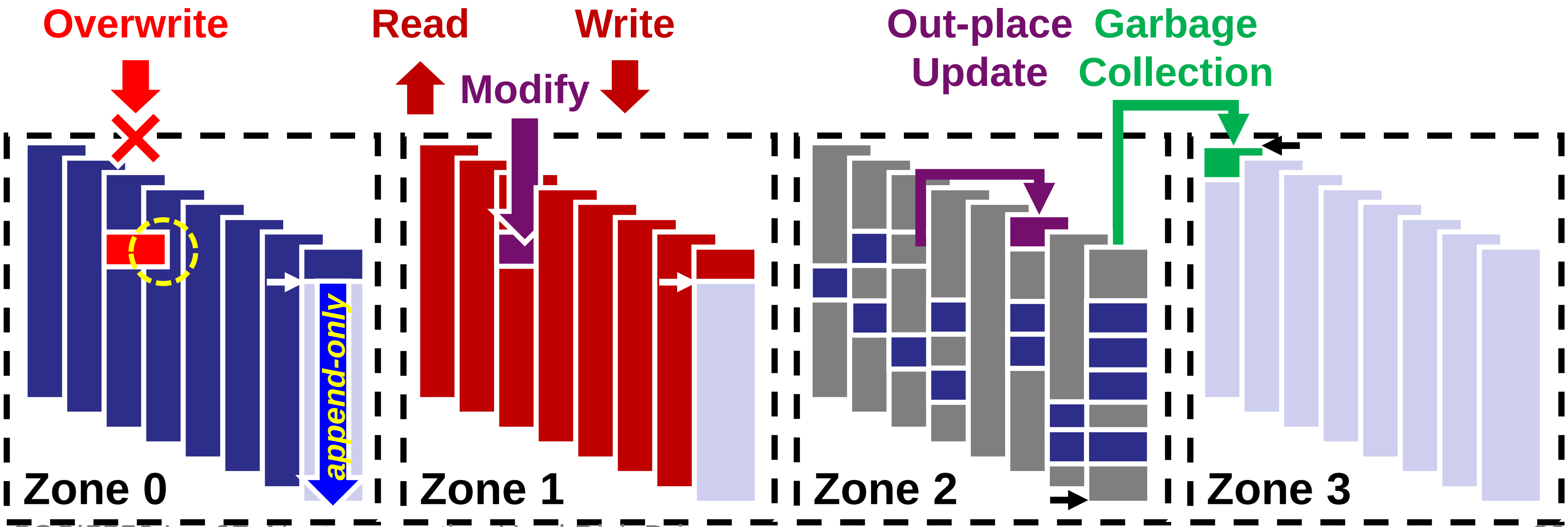


SMR Inherent Challenges

- **Constraint**: Writes to SMR must be strictly append-only on the current write position (i.e., write pointer).
  - **Zone-based management** can mitigate this challenge.
  - **Overwrite** is still prohibited in a zone.

Approach 1) **Read-modify-write (RMW)** is expensive over a zone.

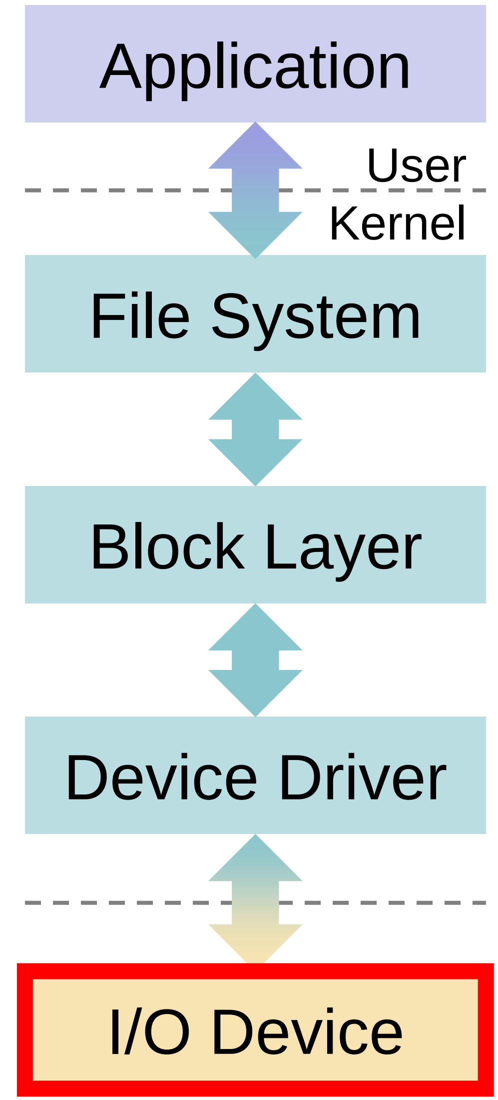
Approach 2) Although **out-place-update** may service updates more efficiently, **address mapping** and **garbage collection** are needed.



Outline

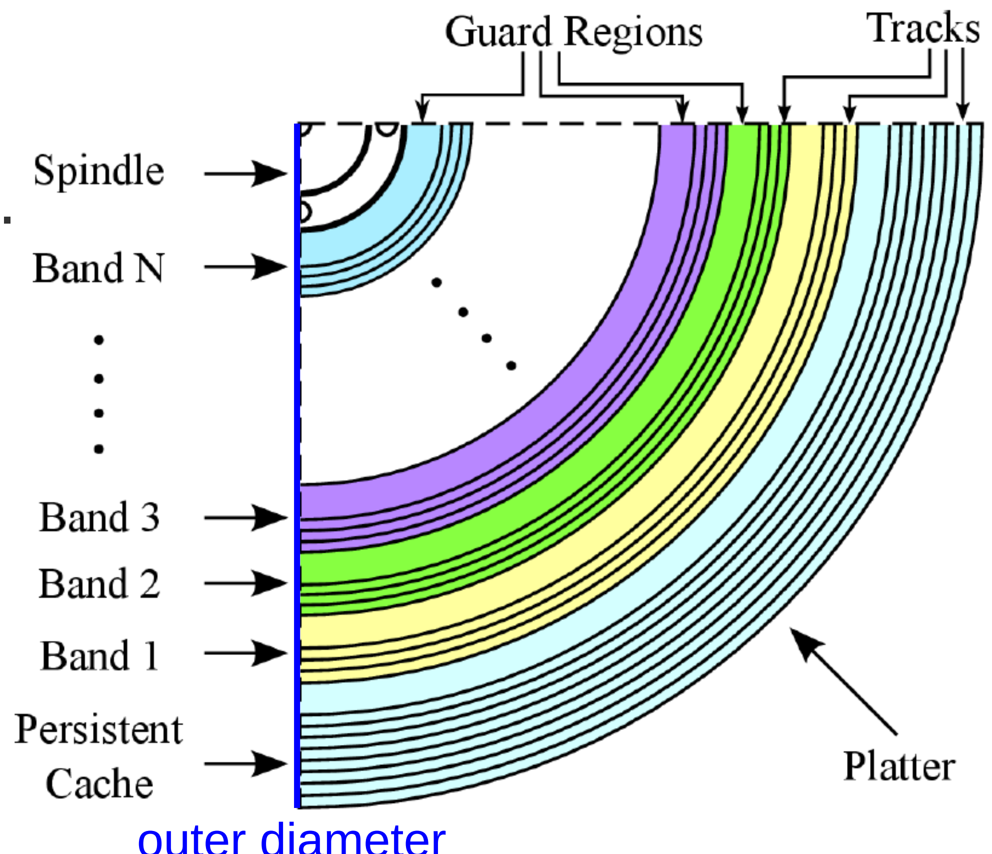
- Traditional Hard Disk Drive
  - Why and How
  - Development Bottleneck
- New Magnetic Recording Technologies
- Shingled Magnetic Recording (SMR)
  - Basics and Inherent Challenges
  - General Solution: Persistent Cache
  - Various SMR Drive Models and Designs
    - Drive-Managed SMR (DM-SMR)
    - Host-Aware SMR (HA-SMR)
    - Host-Managed SMR (HM-SMR)
    - Hybrid SMR

I/O Stack



General Solution to Non-Seq. Writes

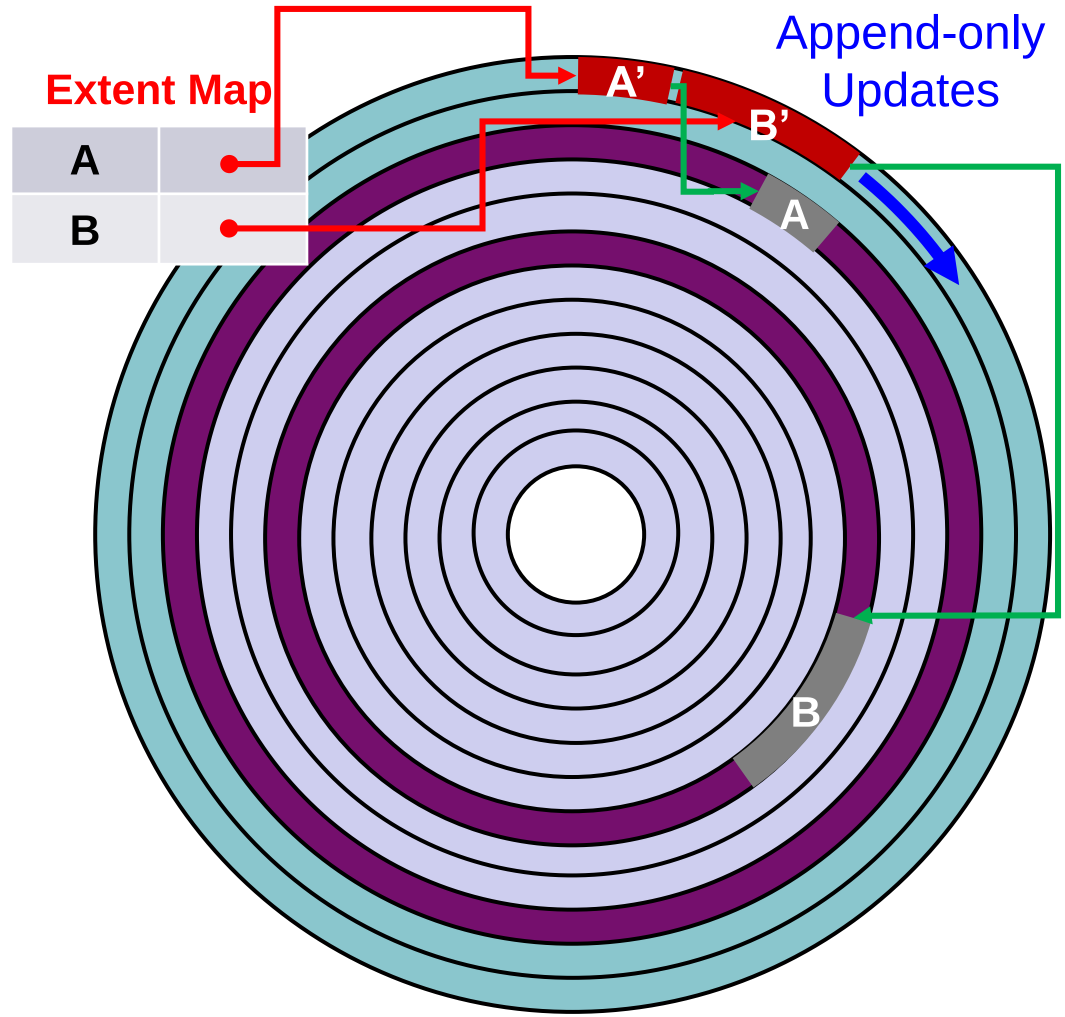
- **Persistent Cache**: A small region to **stage** non-sequential writes to all zones (via out-place updates).
  - **Non-sequential Write**: Data are not written to the current "write pointer" of a **zone** (also called a **band**).
  - It can postpone updates and reduce the number of costly RMWs to zones.
  - It can be made by any persistent storage technology such as:
    - **Disk itself**: SMR tracks at outer diameter (OD).
    - **Flash memory**; or
    - **Non-volatile memory**.



Skylight—A Window on Shingled Disk Operation (FAST'15)

Persistent Cache

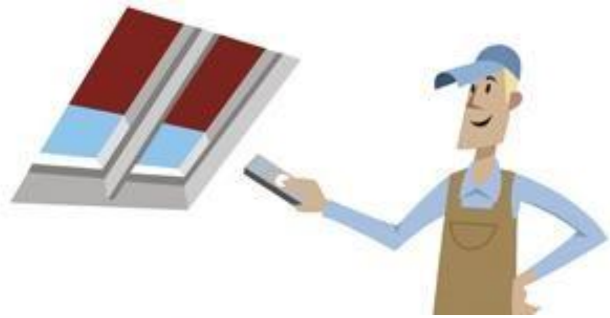
- **Hybrid Mapping**:
  - Persistent cache uses **extent mapping**;
  - Disk space is mapped at **zone granularity**.
- **Garbage Collection** is needed to read-modify-write a zone.
  - **Lazy Cleaning**
    - Start until the persistent cache is **almost full**
  - **Aggressive Cleaning**
    - Start as soon as disk **idleness** is detected



Zones are shown in purple;
Persistent cache is shown in green.

"Observational" Results (1/4)

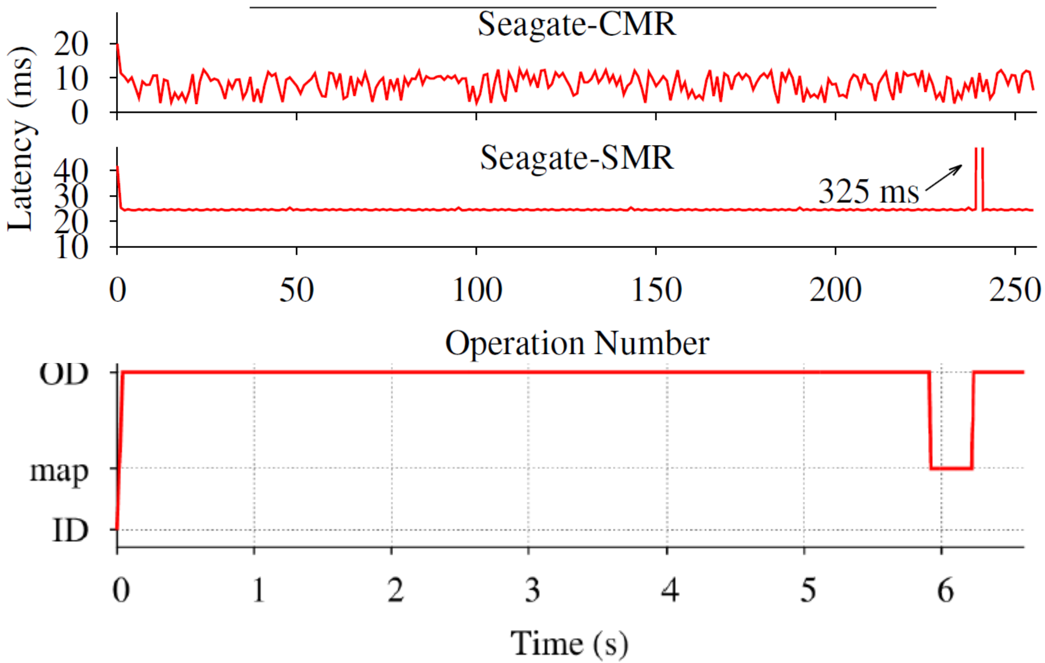
- **Idea**: Reverse engineer key properties of a DM-SMR drive.
  - **Software**: Launch crafted I/O operations using `fio`.
  - **Hardware**: Install a "**skylight**" on the drive to track the head movements using a high-speed camera.
- Real DM-SMR drive was tested.
  - Seagate ST5000AS0011
  - 5900RPM (rotation time 10 ms)
  - Four platters
  - Eight heads
  - 5TB capacity

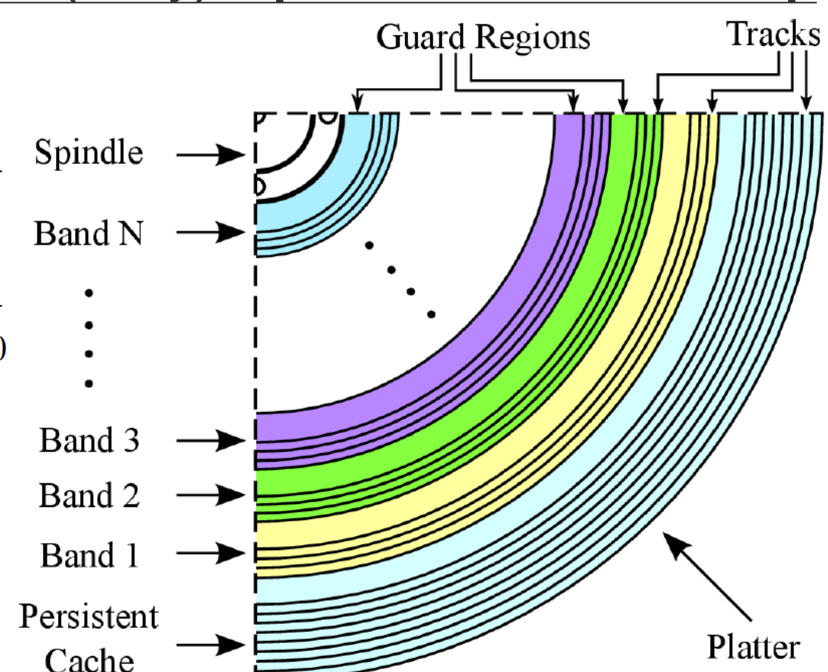


Skylight—A Window on Shingled Disk Operation (FAST'15)

“Observational” Results (2/4)

- Random Write Test: Write the first GB blocks randomly.
- **Observational Results**:
  - Seagate-CMR shows varying latencies (due to repositions).
  - Seagate-SMR shows a fixed $\approx 25ms$ latency with a **bump**.
    - **Fixed Latency**: There is a **persistent cache** at the outer diameter.
    - **Bump at the 240th Write**: There is (likely) a **persistent cache map** stored at the middle diameter.





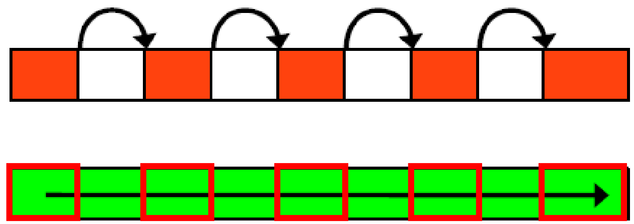
Skylight—A Window on Shingled Disk Operation (FAST'15)

"Observational" Results (3/4)

- Fragmented Read Test:
  - ① Choose a small region and writing every other block in it
  - ② Read the region sequentially, forcing fragmented reads

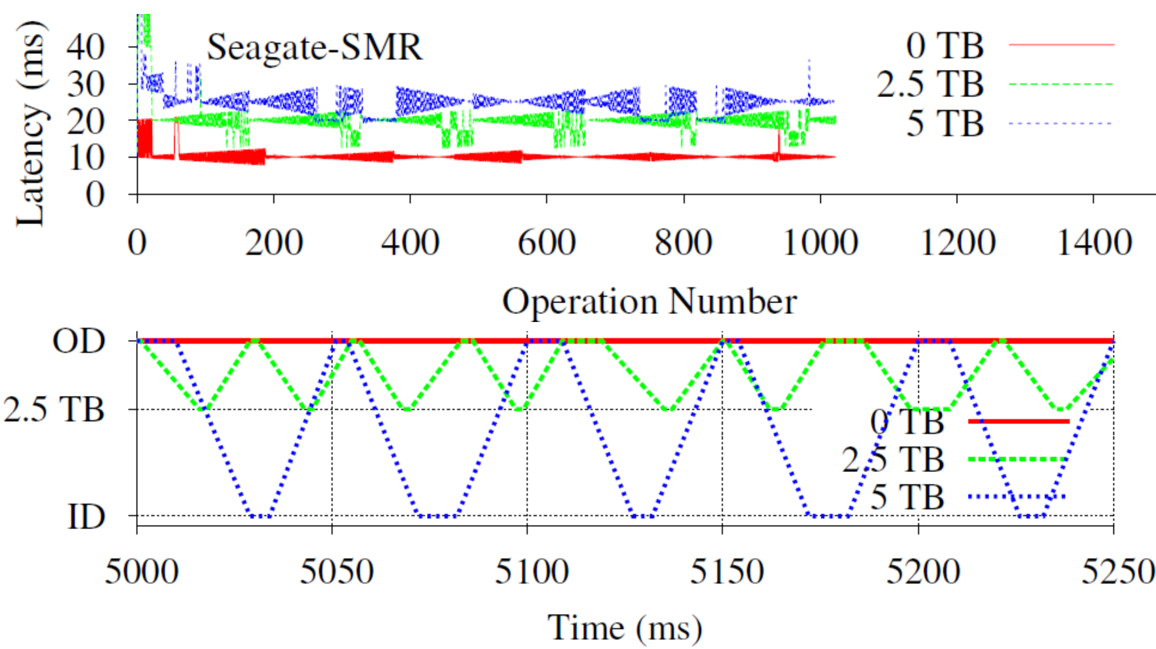
① Skip Write

② Sequential Read



Should be read from persistent cache!

- **Observational Results:**



- A fragmented read at **low LBAs** (0 TB) incurs negligible seek;
- A fragmented read at **high LBAs** (5 TB) incurs high seek.
- The persistent cache locates **at the OD**.

"Observational" Results (4/4)

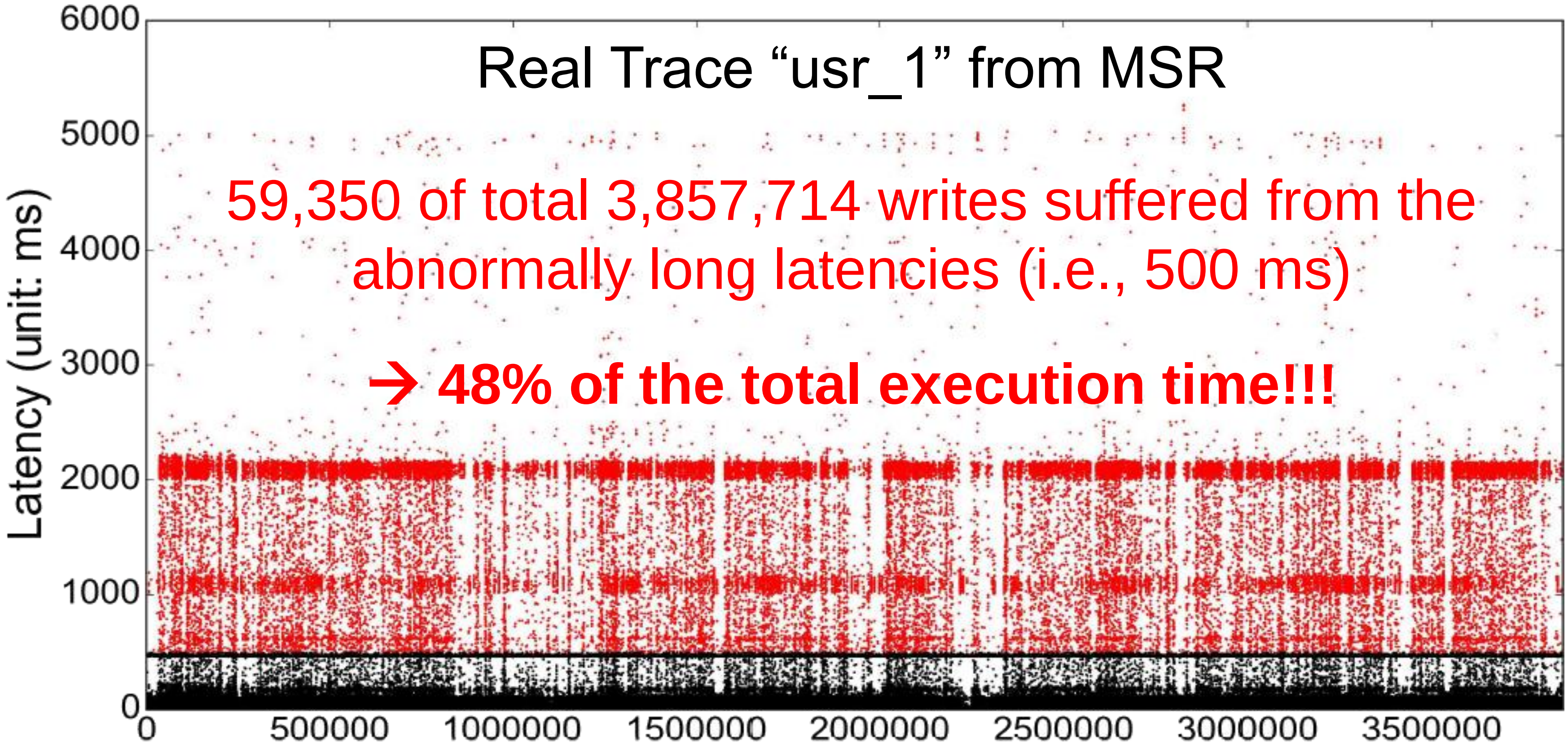
- All properties discovered by Skylight methodology:

Property	Drive Model ST5000AS0011	Drive Model ST8000AS0011
Drive Type	SMR	SMR
Persistent Cache Type	Disk	Disk
Cache Layout and Location	Single, at the OD	Single, at the OD
Cache Size	20 GiB	25 GiB
Cache Map Size	200,000	250,000
Band Size	17–36 MiB	15–40 MiB
Block Mapping	Static	Static
Cleaning Type	Aggressive	Aggressive
Cleaning Algorithm	FIFO	FIFO
Cleaning Time	0.6–1.6 s/band	0.6–1.6 s/band
Zone Structure	4–20 GiB	5–40 GiB
Shingling Direction	Towards MD	N/A

Skylight—A Window on Shingled Disk Operation (FAST'15)

“Evaluated” Results

- Long latency issue is observed under real workloads.
  – Real workloads were replayed on a Seagate 8TB HA-SMR.
  – Persistent cache cleaning may be the bottleneck.

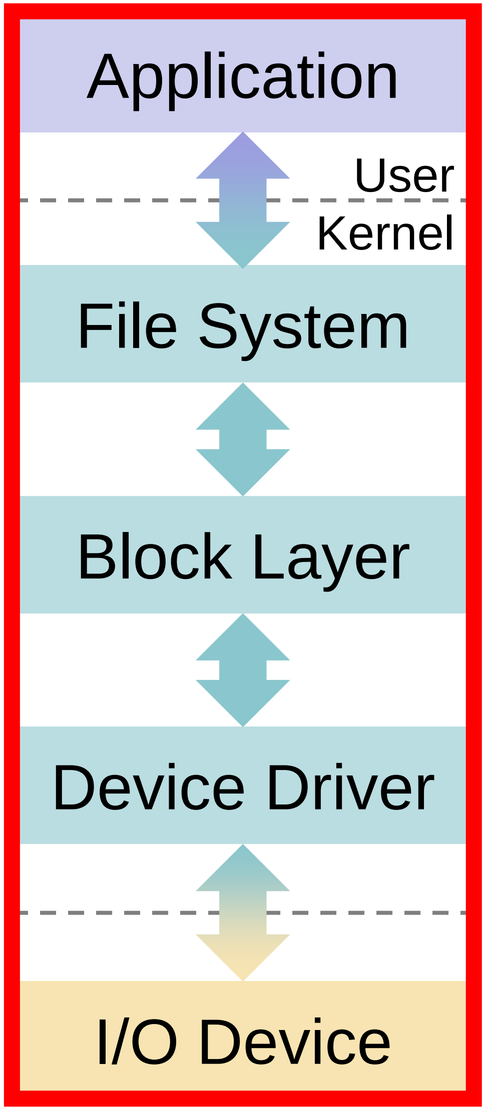


Persistent cache is not the panacea!

Outline

- Traditional Hard Disk Drive
  - Why and How
  - Development Bottleneck
- New Magnetic Recording Technologies
- Shingled Magnetic Recording (SMR)
  - Basics and Inherent Challenges
  - General Solution: Persistent Cache
  - Various SMR Drive Models and Designs
    - Drive-Managed SMR (DM-SMR)
    - Host-Aware SMR (HA-SMR)
    - Host-Managed SMR (HM-SMR)
    - Hybrid SMR

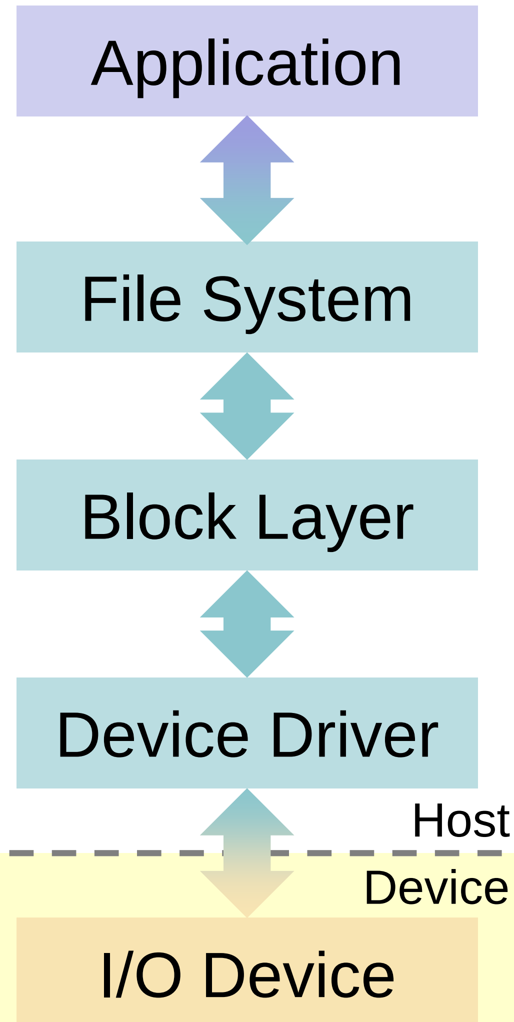
I/O Stack



Who Should Take Care of SMR

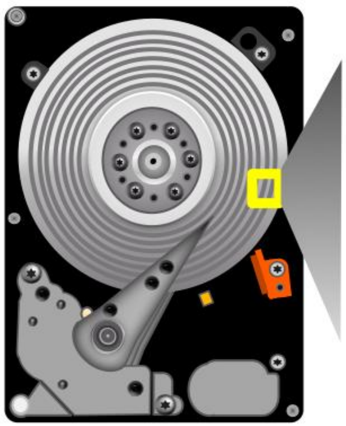
- **T10** defines three possible models:
  - **Host-Managed (HM)**
    - − Host must write a zone sequentially
      - ➢ **"Non-sequential write" is prohibited**
    - − Require lots of system software redesigns
    - + High predictability on raw I/O performance
  - **Host-Aware (HA)**
    - + Host is suggested to write a zone sequentially
      - ➢ **"Non-sequential write" is handled by drive**
    - − Require moderate system software redesigns
  - **Drive-Managed (DM)**
    - + **Transparent to the host side**
      - ➢ Drop-in replacement for traditional drives
      - ➢ **A firmware handles "non-sequential write"**
    - − Low predictability on I/O performance of drives

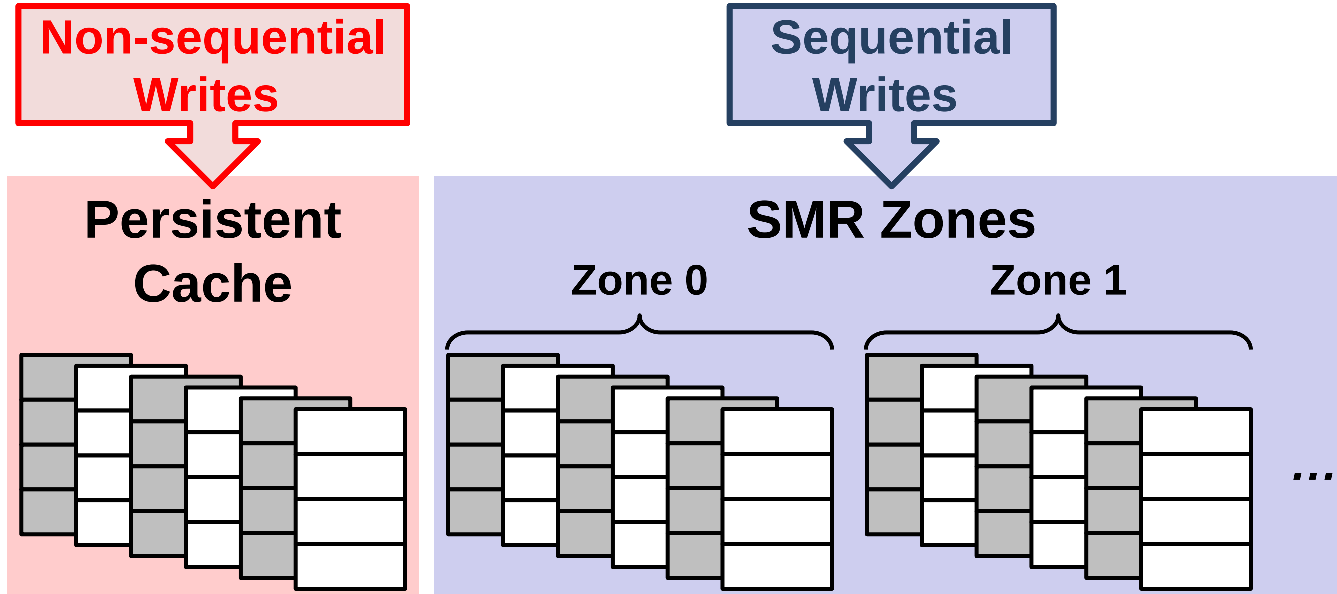
Technical Committee T10



Various SMR Drive Models

Technical Committee T10

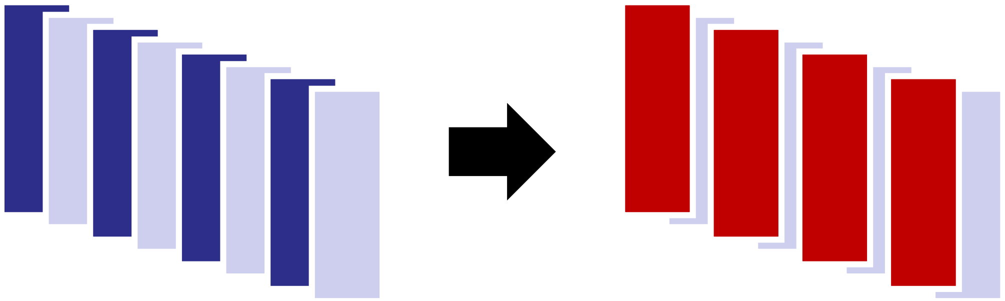




SMR Drive Model	Persistent Cache	SMR Zone
Host-Managed (HM)	X	O
Host-Aware (HA)	Δ *(not accessible by the host)*	O
Drive-Managed (DM)	Δ *(transparent to host)*	Δ *(transparent to host)*

Case Study of DM-SMR: SMaRT

- DM-SMR can manage all details of the drive to:
  - Improve overall I/O performance
  - **Remove or mitigate the use of persistent cache**
- **Observation**: A track supports in-place update if its following track is free or contains stale data.



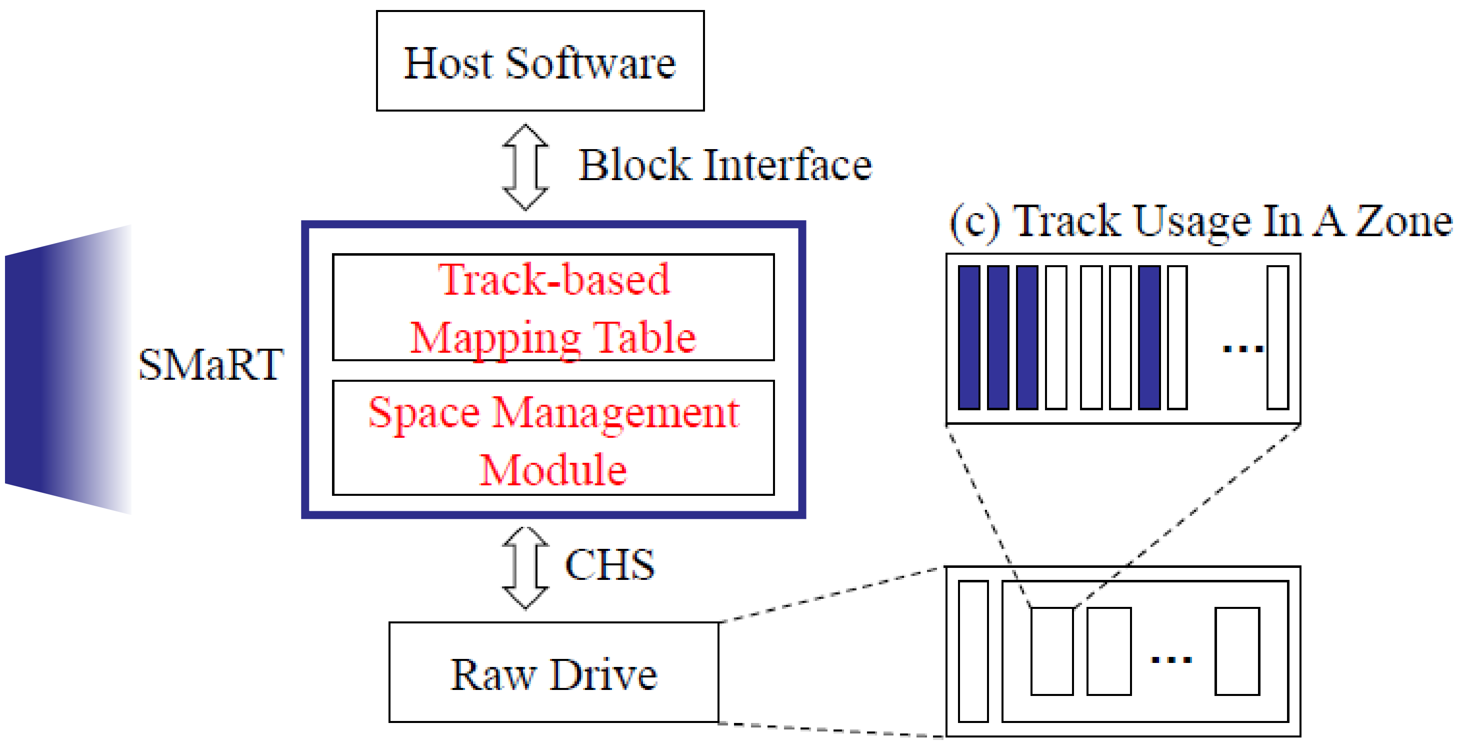
- **Track-based Management**:
  - **Sector** is too small: Sector-based mapping creates huge mapping table and introduces garbage blocks/sectors.
  - **Track** is moderate: Tracks can be managed more flexibly.
  - **Zone** is too big: Non-sequential writes are troublesome.

SMaRT - An Approach to Shingled Magnetic Recording Translation (FAST'17)

SMaRT: Overall Architecture

- Two major modules in the **firmware** (i.e., **STL**):
  - A **track-based mapping** to support track-level translation
  - A **space management module** to manage free track allocations and garbage collection

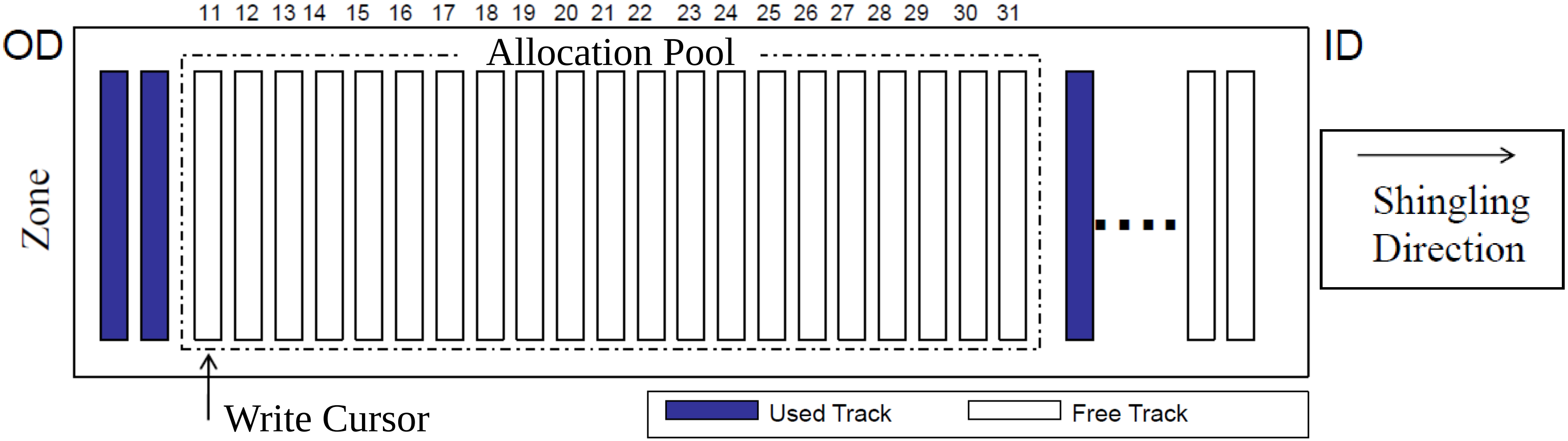
Shingled **T**ranslation **L**ayer (**STL**)



(a) General Architecture (b) Drive Physical Layout

SMaRT - An Approach to Shingled Magnetic Recording Translation (FAST'17)

SMaRT: Track Allocation in a Zone

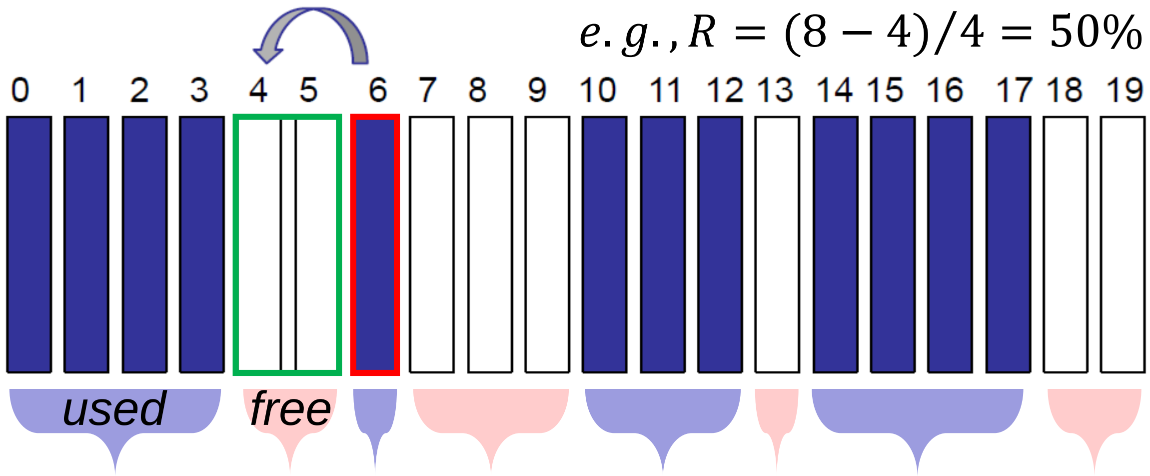


- All writes (new data and updated data) go to the **write cursor** of an **allocation pool** sequentially.
- Newly updated tracks are deemed as hot data.
  - SMaRT allocates an extra track as **safety gap** for each hot track if space utilization is less than 50%.
- When the current pool is full, choose the new one of the **largest** number of consecutive free tracks.

SMaRT: Garbage Collection

- SMaRT invokes GCs in each zone whenever the fragmentation ratio R is smaller than a *threshold*:

$e.g., R = (8-4)/4 = 50\%$



Fragmentation Ratio

$$R = \frac{F-N}{F}, where\ 1 \le N \le F$$

- F: Num. of free tracks
- N: Num. of free space elements

Space Elements

① **Pick Victim** $e.g., W = (12/20)\ /(1-(12/20)) = 1.5$

- Search for a victim space element in a zone (from the leftmost to rightmost), where its size is smaller than the used-to-free ratio W

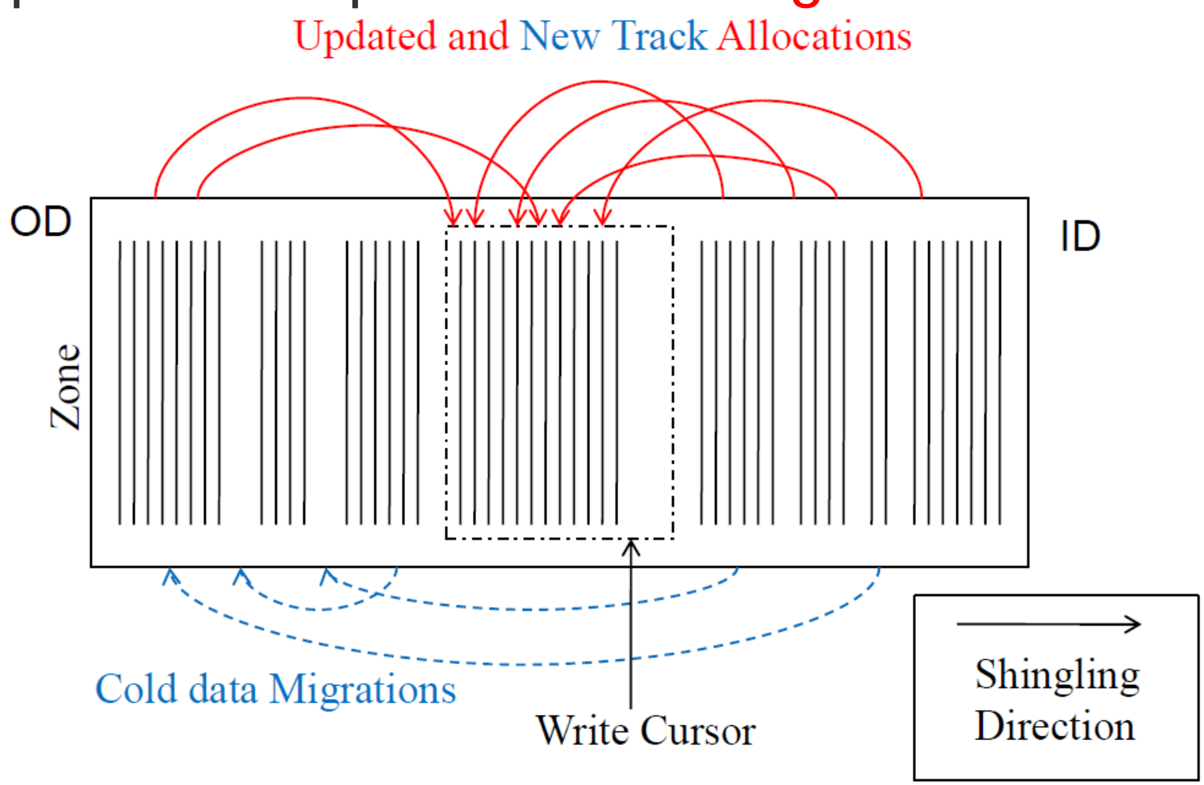
$$W = \frac{U}{1-U}, where\ U\ is\ the\ drive\ space\ usage$$

② **Pick Destination**

- Allocate to the first free space element to the left that fits it; or
- Shift left and append to its left neighbor

SMaRT: Auto Cold Data Progression

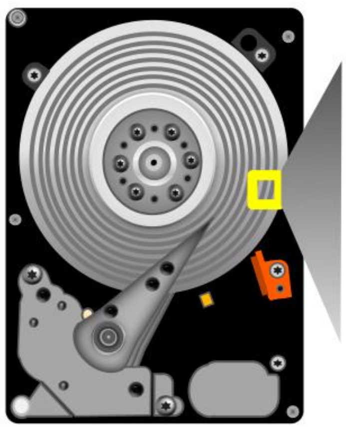
- The free track allocation of GC provides a good opportunity of **automatic cold data progression**.
  - The cold data will mostly stay at the left side, while hot data gets updated and pushed to the right side of the zones

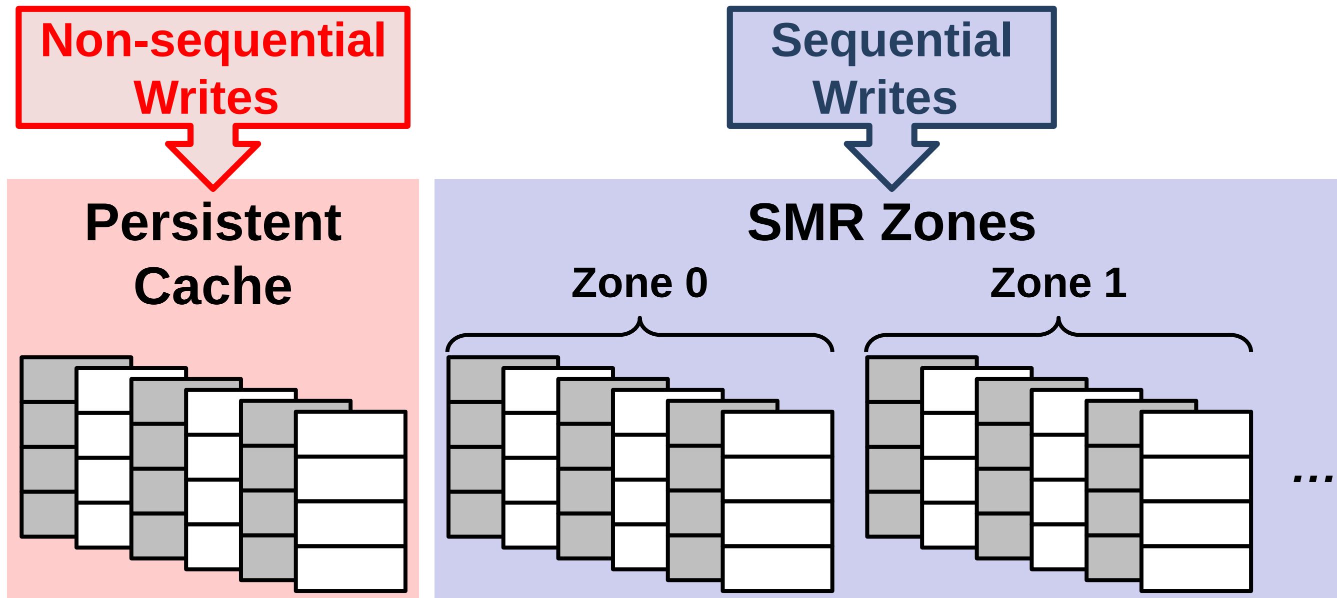


- Unnecessary cold data movements can be avoided.

Various SMR Drive Models

incits
Technical Committee T10

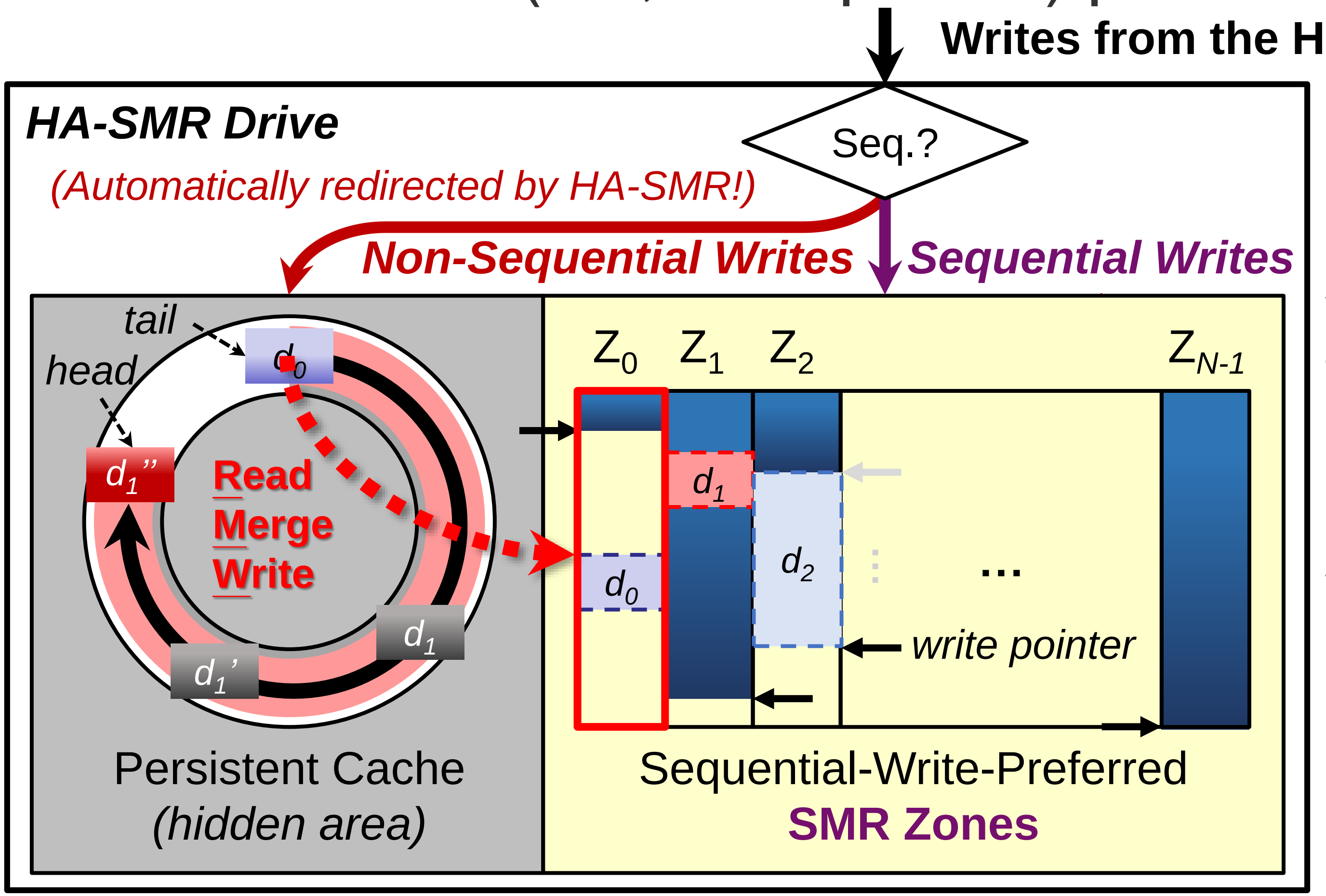




SMR Drive Model	Persistent Cache	SMR Zone
Host-Managed (HM)	X	O
Host-Aware (HA)	Δ *(not accessible by the host)*	O
Drive-Managed (DM)	Δ *(transparent to host)*	Δ *(transparent to host)*

Case Study of HA-SMR: VPC

- HA-SMR drive reports "zone information" to the host,
- HA-SMR automatically handles non-sequential writes in a "**hidden**" (i.e., transparent) persistent cache.



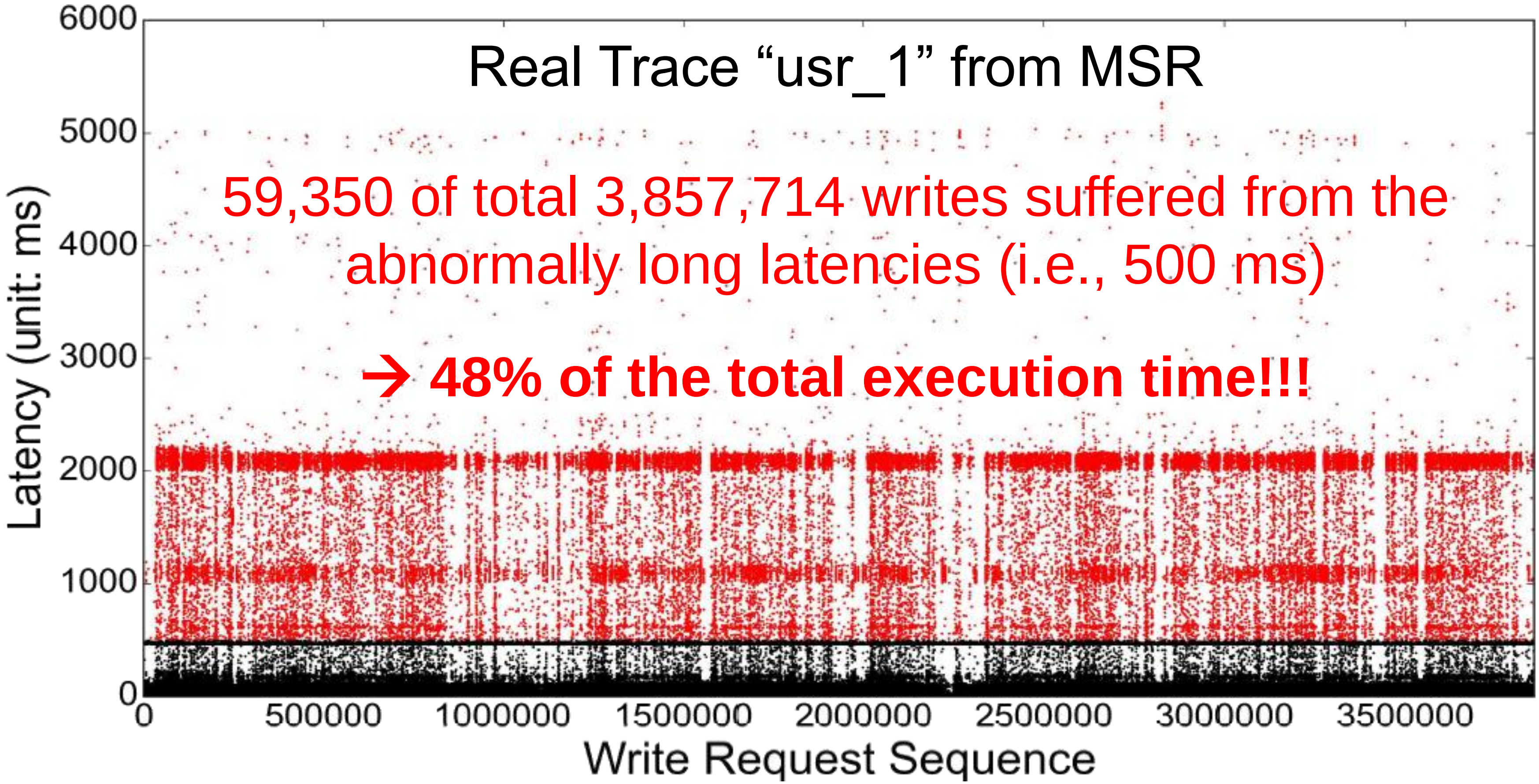
Sequential Write: Written data are on the current write pointer of a zone.

Non-sequential Write: Written data are not on the current write pointer of a zone.

On Improving the Write Responsiveness for Host-Aware SMR Drives (TC'19)

Long Latency Behavior of HA-SMR

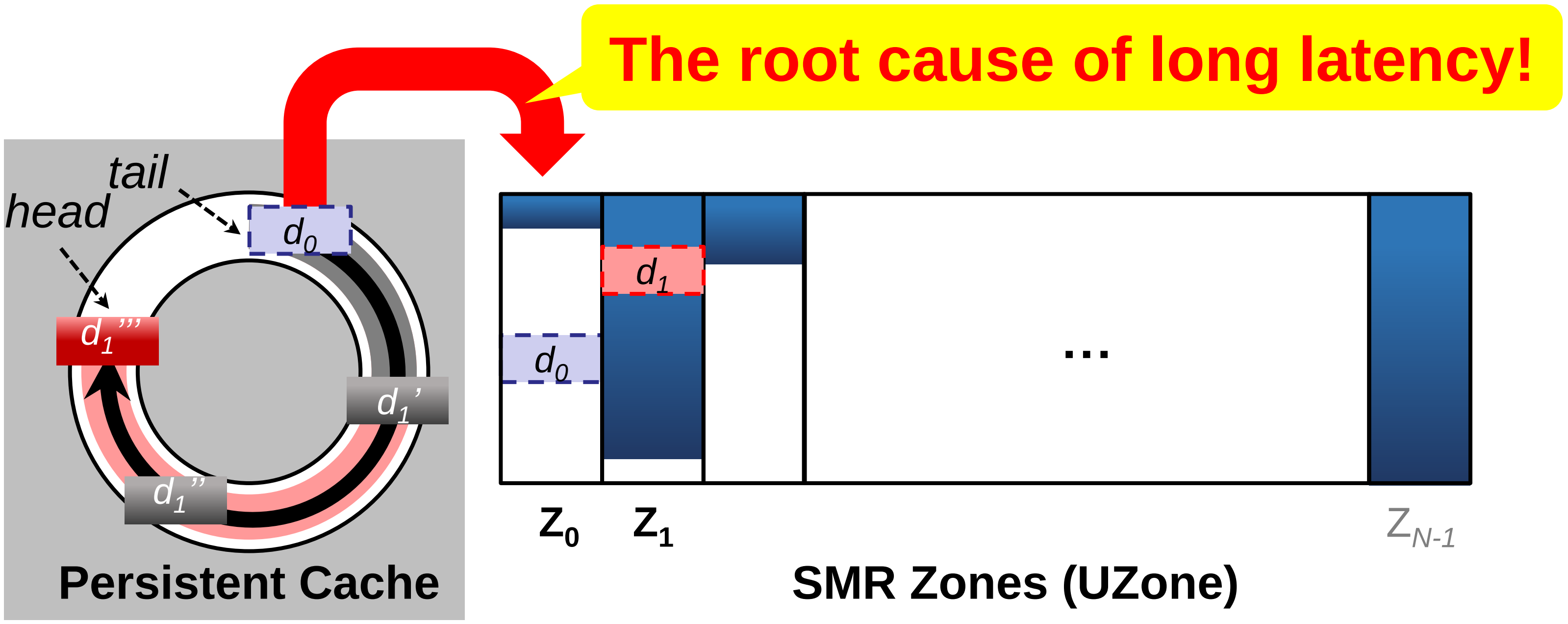
- Long latency issue is observed under real workloads.
  - Real workloads were replayed on a Seagate 8TB HA-SMR.
  - Persistent cache cleaning may be the bottleneck.



On Improving the Write Responsiveness for Host-Aware SMR Drives (TC'19)

Root of Long Latency Behavior

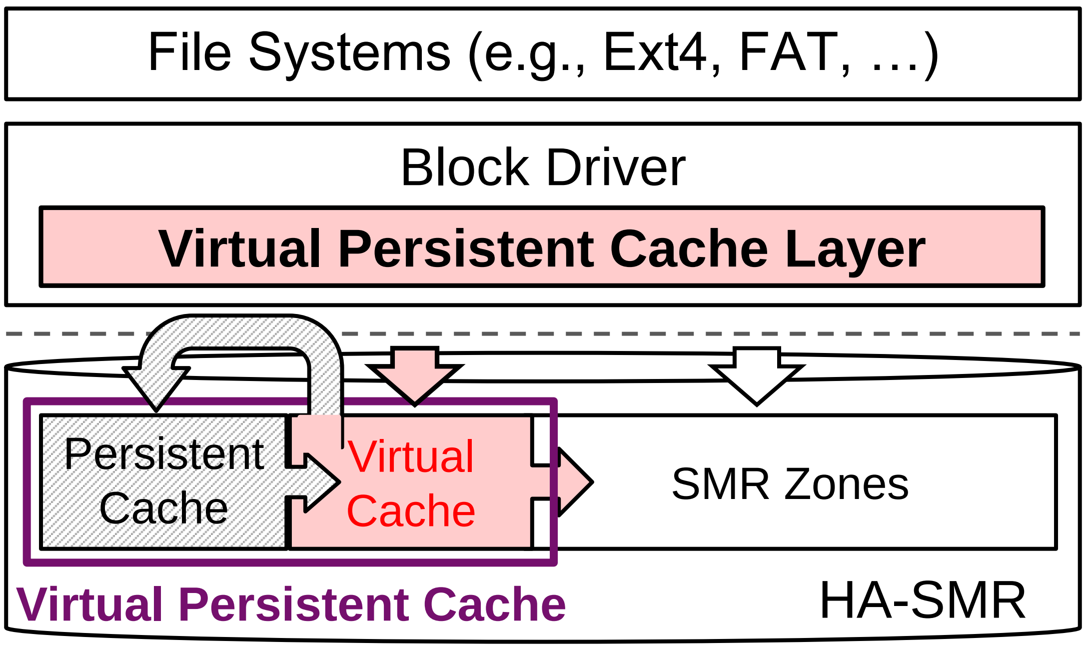
- **Fact**: Some are updated more frequently than others.
  - Non-sequential writes of **different update frequencies** are **mixed** in the persistent cache.
  - Read-merge-writes must **often** merge back non-frequently-updated data residing at the tail of the persistent cache.



On Improving the Write Responsiveness for Host-Aware SMR Drives (TC'19)

Virtual Persistent Cache for HA-SMR

- **VPC**: Virtually enlarge the persistent cache
  - Virtual Persistent Cache = Persistent Cache + Virtual Cache
  - **Goal:** Avoid overwhelming the persisten cache by non-sequential writes of a wide range of update frequenceies.
- Virtual Cache
  - Take a few zones based on the needs.
    - **Trade little space for great performance!**
  - Take advantage of the **computing and memory resources** of the host system.

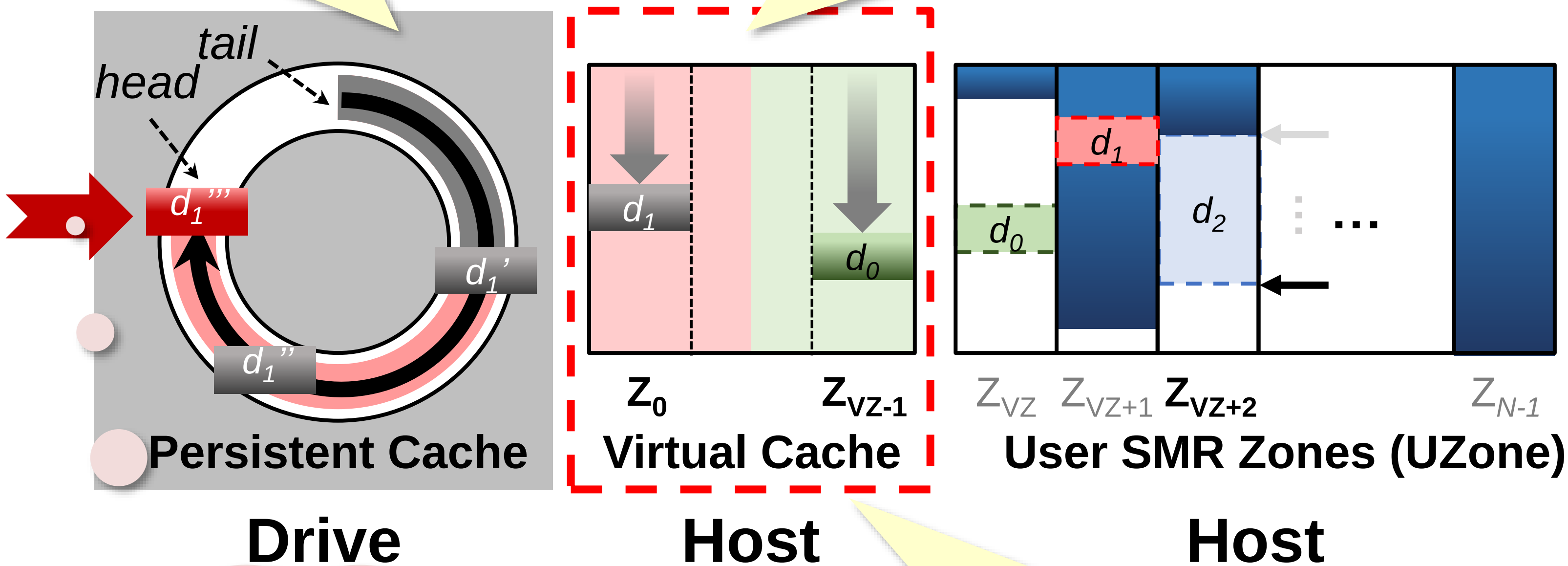


In practice, the proposed design can be realized at the block drive layer as a general solution for various applications.

VPC: Host-Drive Collaboration

(3) **Only allow freq. updated data** to be redirected into Persistent Cache

(2) Always sequentially accommodate the received non-seq. writes into Virtual Cache by **remapping** the LBA



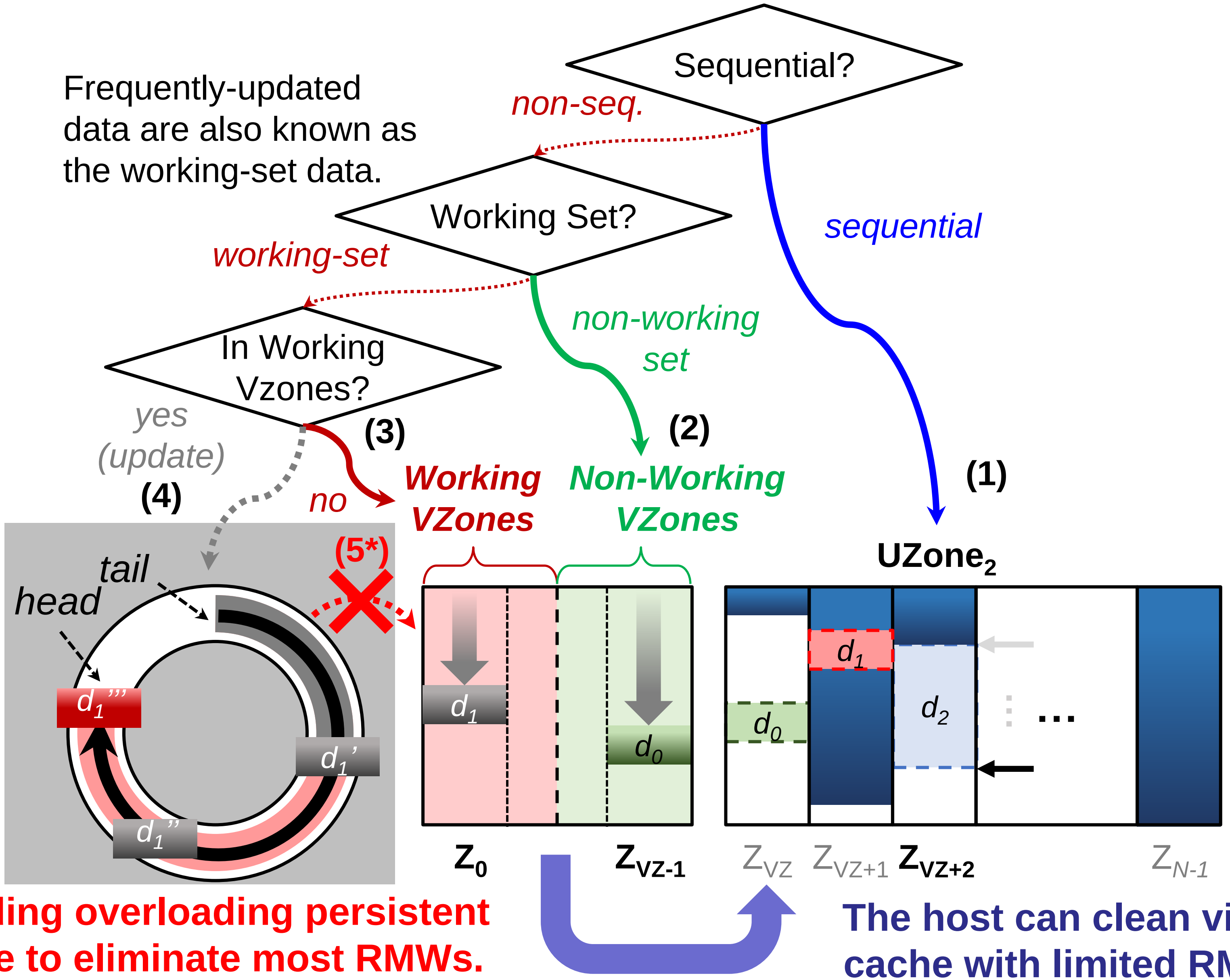
Making it a non-seq. update to $\mathbf{d_1}$ in Virtual Cache!

(1) Adaptively allocate few SMR zones to work as Virtual Cache

On Improving the Write Responsiveness for Host-Aware SMR Drives (TC'19)

VPC: Hot/Cold Separation in Cache(s)

Frequently-updated data are also known as the working-set data.

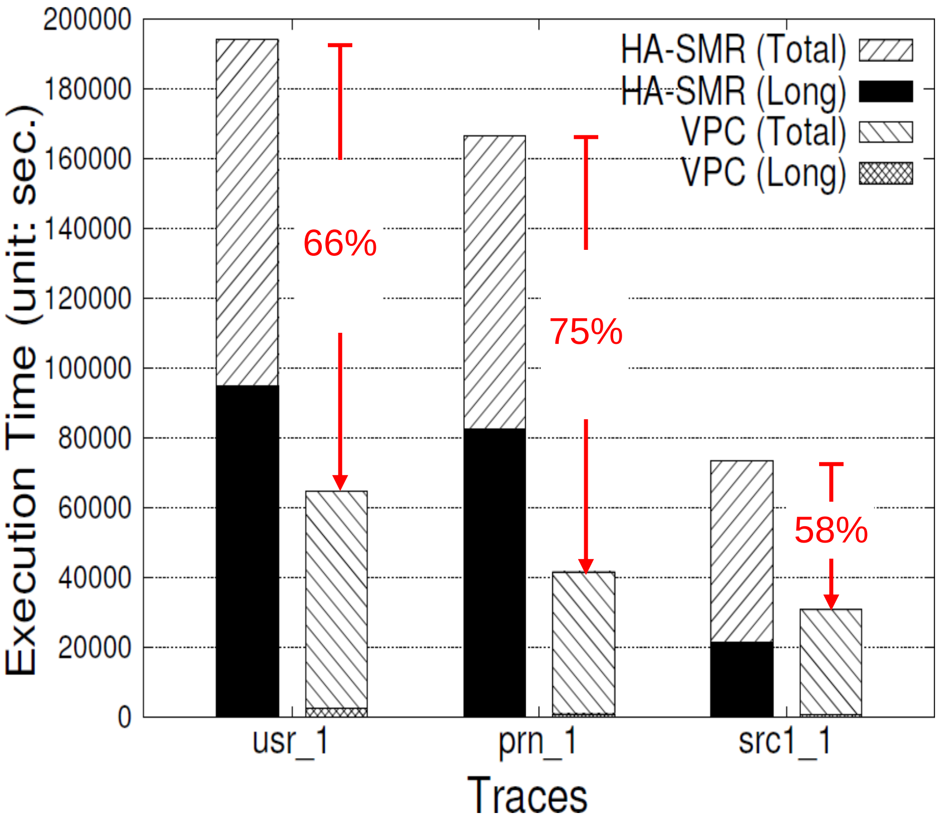


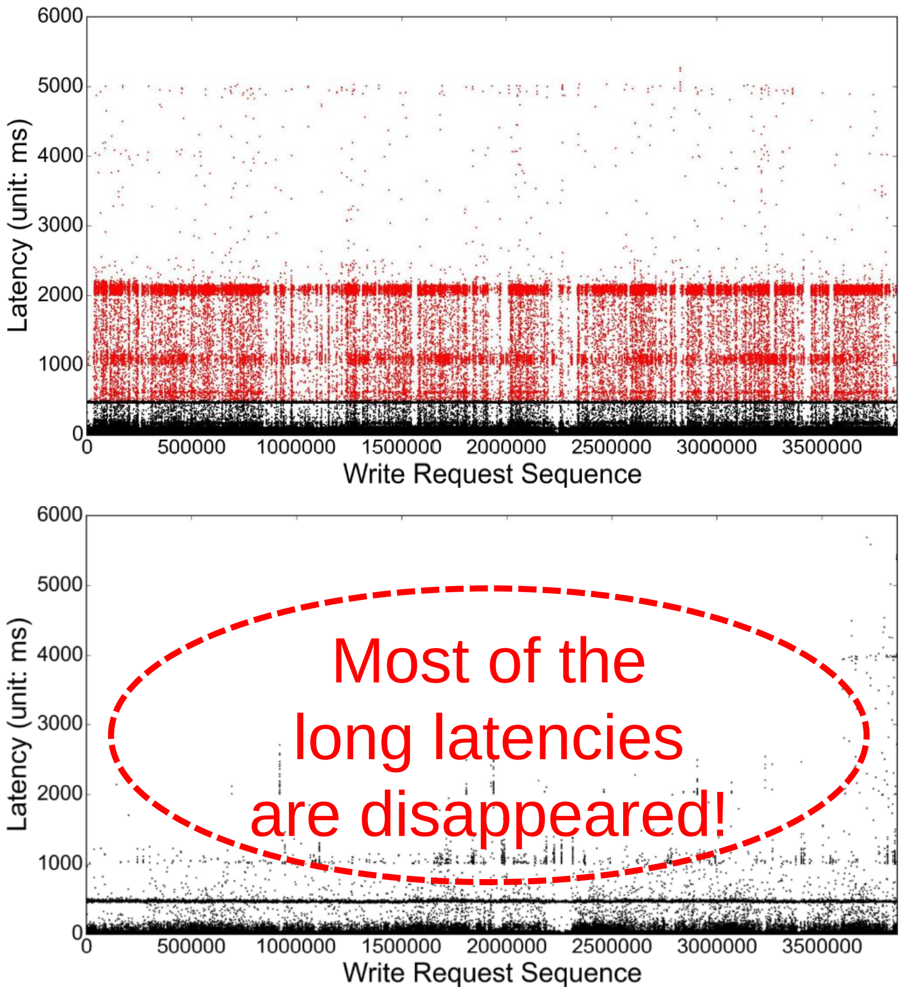
Avoiding overloading persistent cache to eliminate most RMWs.

The host can clean virtual cache with limited RMWs.

VPC: Evaluation Results

- Great Improvement of Total Execution Time

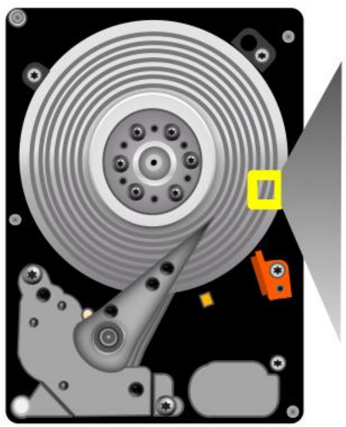


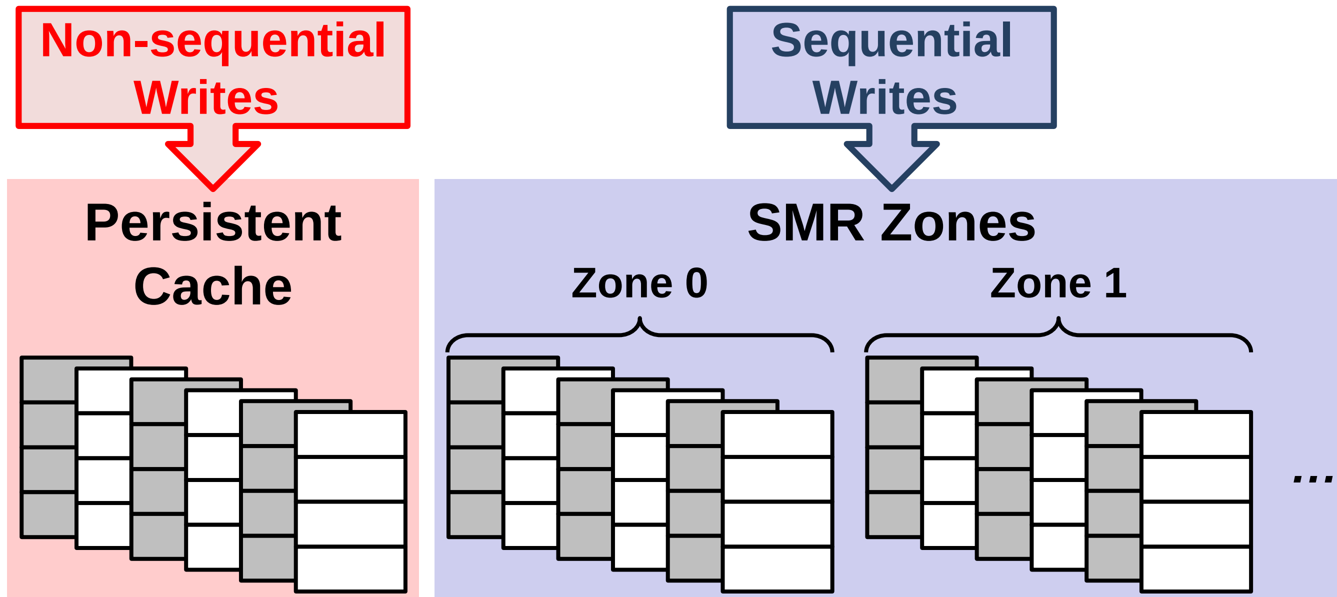


On Improving the Write Responsiveness for Host-Aware SMR Drives (TC'19)

Various SMR Drive Models

incits
Technical Committee T10

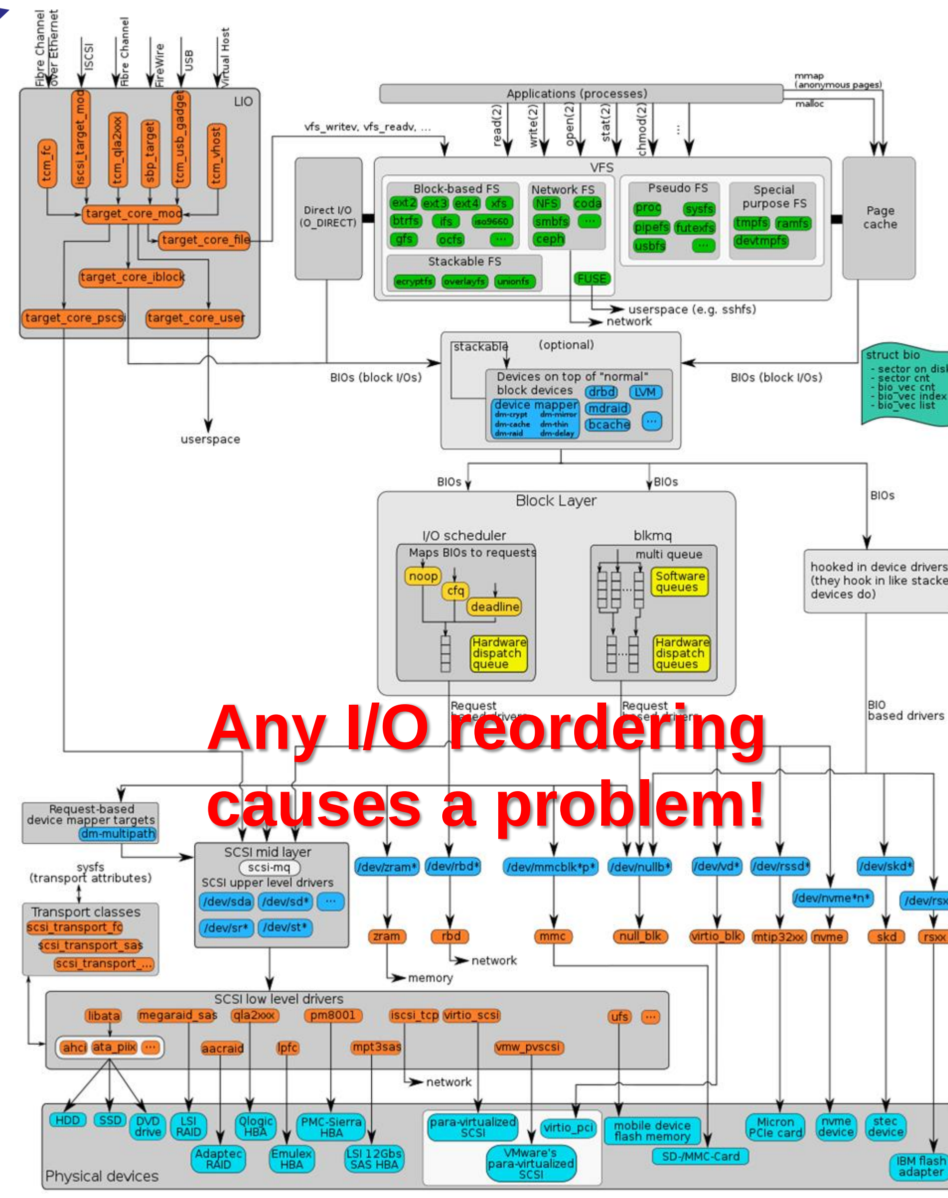




SMR Drive Model	Persistent Cache	SMR Zone
Host-Managed (HM)	**X**	**O**
Host-Aware (HA)	**Δ** *(not accessible by the host)*	**O**
Drive-Managed (DM)	**Δ** *(transparent to host)*	**Δ** *(transparent to host)*

Challenges of HM-SMR

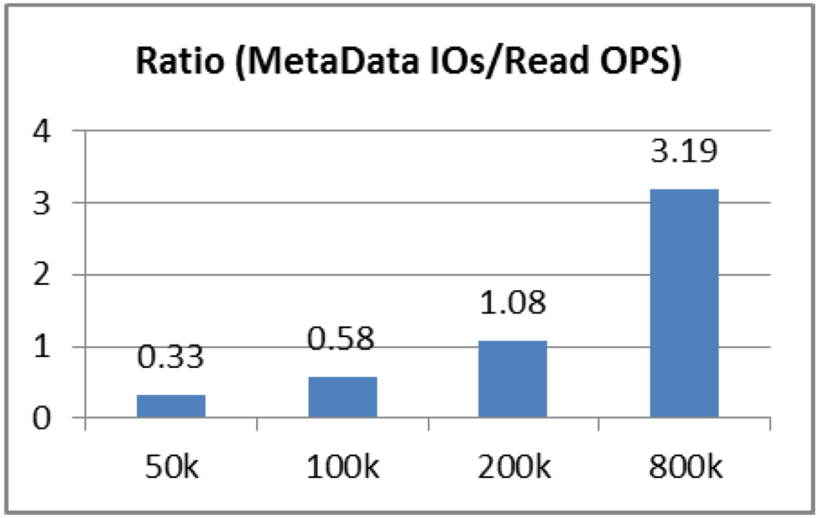
- The host **must** write a zone sequentially.
  - Non-sequential write is strictly prohibited.
- **All of layers** must be SMR compatible.
  - A dedicated FS must be designed for SMR.
- The host also knows more about the system.
  - Possible to substantially improve the performance and I/O predictability.



ZEA, A Data Management Approach for SMR (HotStorage'16)

Case Study of HM-SMR: HiSMRfs

- **HiSMRfs** is high performance FS for SMR drives.
  - It can manage **SMR zones** and support random writes without remapping layer implemented inside SMR drives.
    - Random Writes? Always writing at the end through appending.
  - It further separates data and metadata storage, and manages them differently to achieve high performance.
    - **Observation**: As the number of files increases, a larger number of metadata IOs is needed to access a single file data.

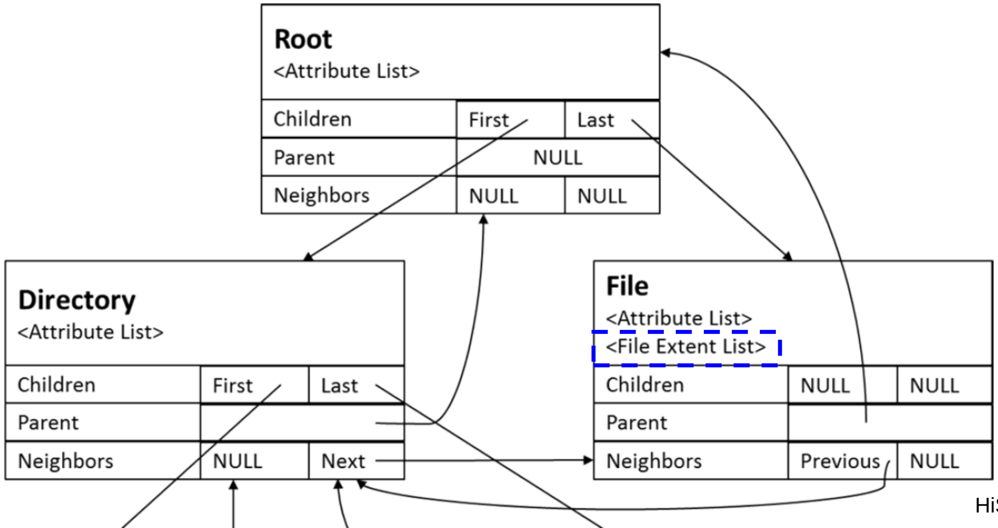


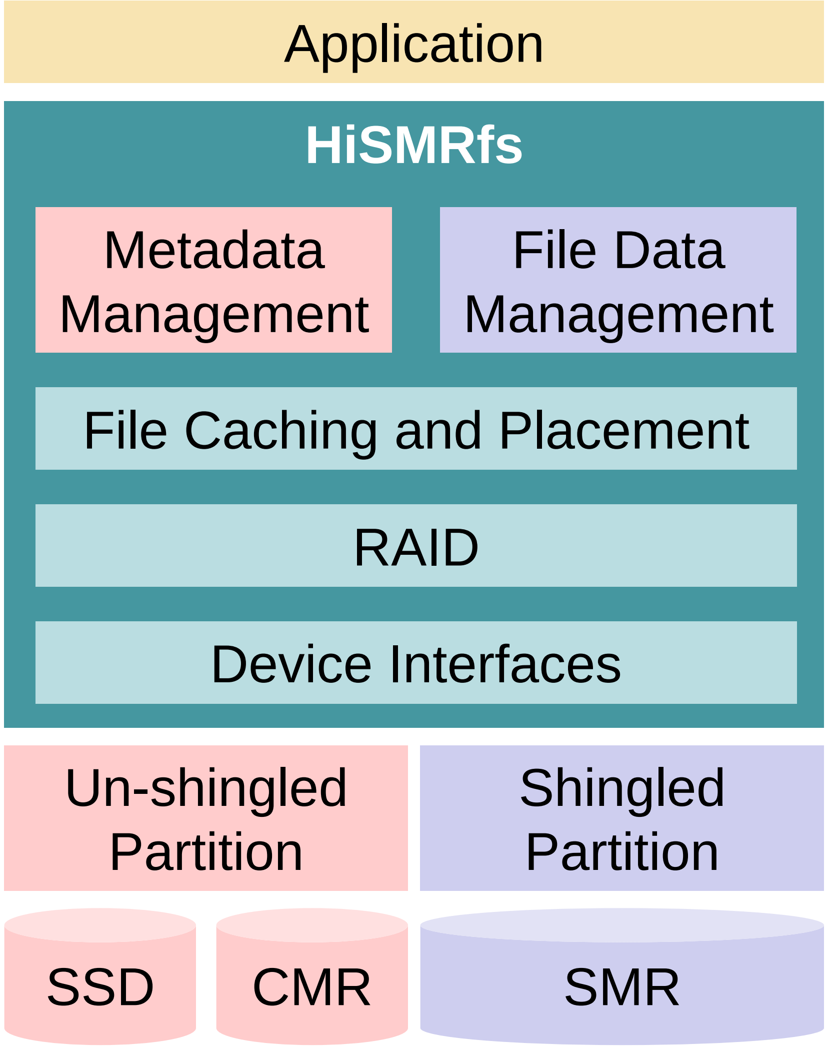
With 800K files stored, reading a data file requires, on average, **three** metadata accesses.

HiSMRfs - a High Performance File System for Shingled Storage Array (MSST'14)

HiSMRfs: Overall Architecture

- The **metadata** and **data** are managed **separately**.
  - **Meta**: Un-shingled Partitions
    - File extent list indicates where the file contents are kept in SMR.
  - **File Data**: Shingled Partitions
    - *File Caching and Placement module further **caches** hot file data in un-shingled partitions.*





HiSMRfs: File Data Management

- HiSMRfs is a SMR-compliant file system.
  - File data are **sequentially appended** at the end of each SMR zone (which is used as a **data log**).
  - File Data Management has **four** major modules:

① The **File Data Allocation** module determines where the data will be written.

② The **File Request Queuing and Scheduling** module arranges file read/write requests into queues.

③ The **Garbage Collection** module is responsible to reclaim released space from zones.

- File deletion and modification will cause invalid data blocks in the data log, and these invalid blocks need to be reclaimed.

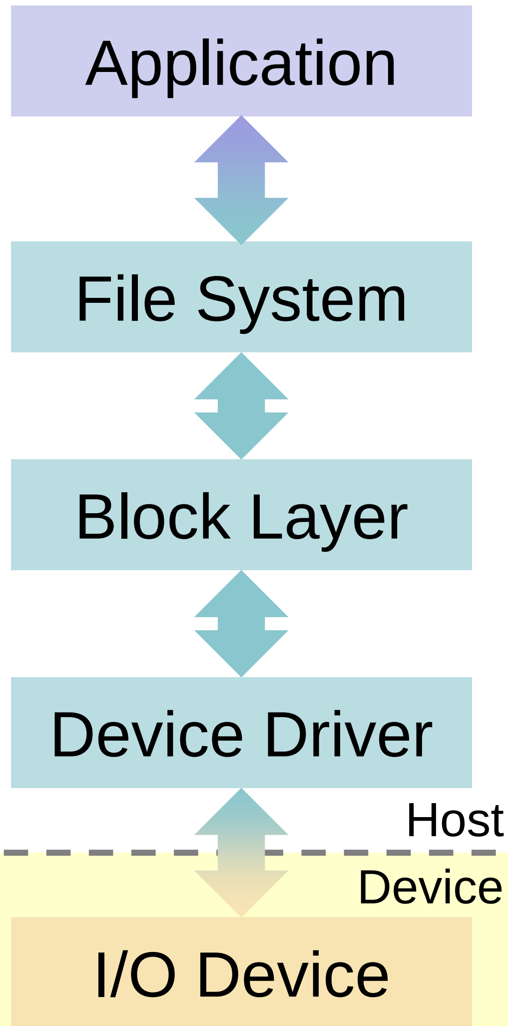
④ The **Zone Layout** module emulates a SMR data layout and its related information.

HiSMRfs - a High Performance File System for Shingled Storage Array (MSST'14)

Wrap-up: Who Should Take Care of SMR

- **T10** defines three possible models:
  - **Host-Managed (HM):** e.g., HiSMRfs
    − Host must write a zone sequentially
      - **"Non-sequential write" is prohibited**
    − Require lots of system software redesigns
    + High predictability on raw I/O performance
  - **Host-Aware (HA):** e.g., VPC
    + Host is suggested to write a zone sequentially
      - **"Non-sequential write" is handled by drive**
    − Require moderate system software redesigns
  - **Drive-Managed (DM):** e.g., SMaRT
    + **Transparent to the host side**
      - Drop-in replacement for traditional drives
      - **A firmware handles "non-sequential write"**
    − Low predictability on I/O performance of drives

incits Technical Committee T10



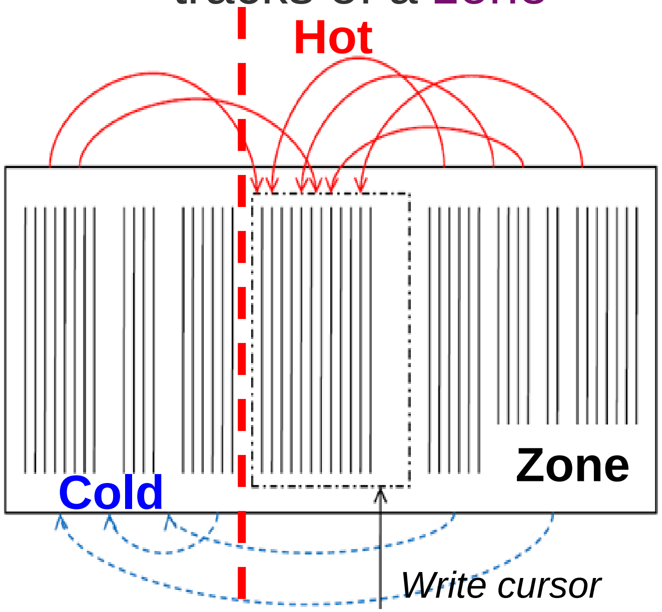
Wrap-up: Hot/Cold Separation

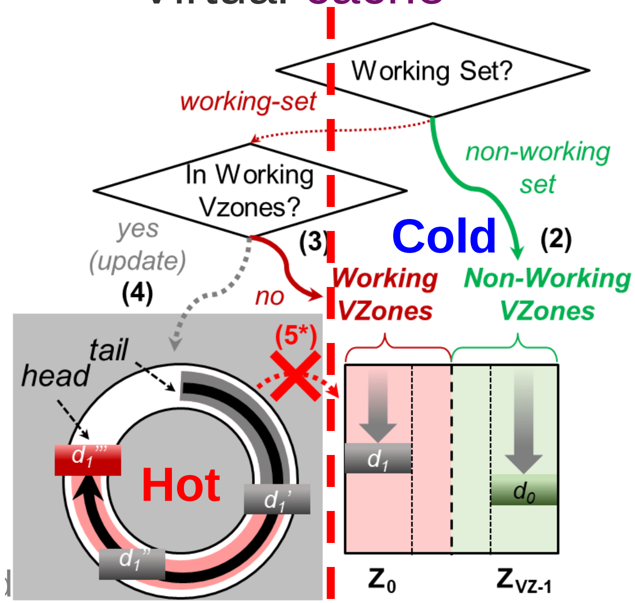
- These designs share one key technique, **hot/cold separation**, to mitigate the non-sequential write issue.
- They achieve it by different policies and granularities:

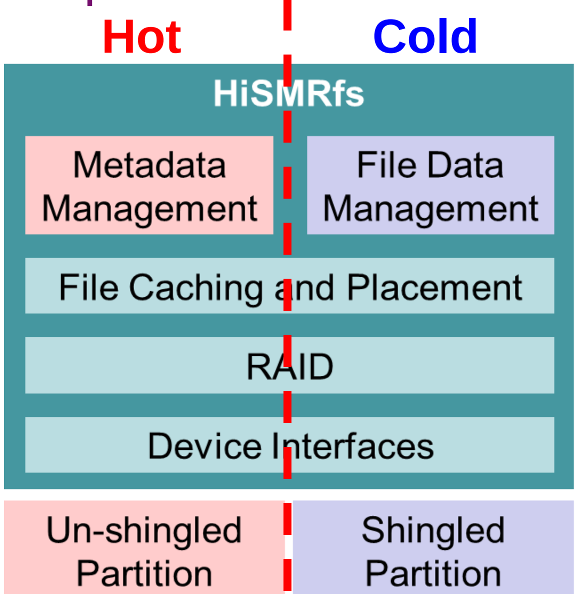
➢ **HM-SMR**: SMaRT
  - Hot Data: RHS tracks of a zone
  - Cold Data: LHS tracks of a zone

➢ **HA-SMR**: VPC
  - Hot Data: Persistent cache
  - Cold Data: Virtual cache

➢ **HM-SMR**: HiSMRfs
  - Metadata: Un-shingled partition
  - Data: Shingled partition

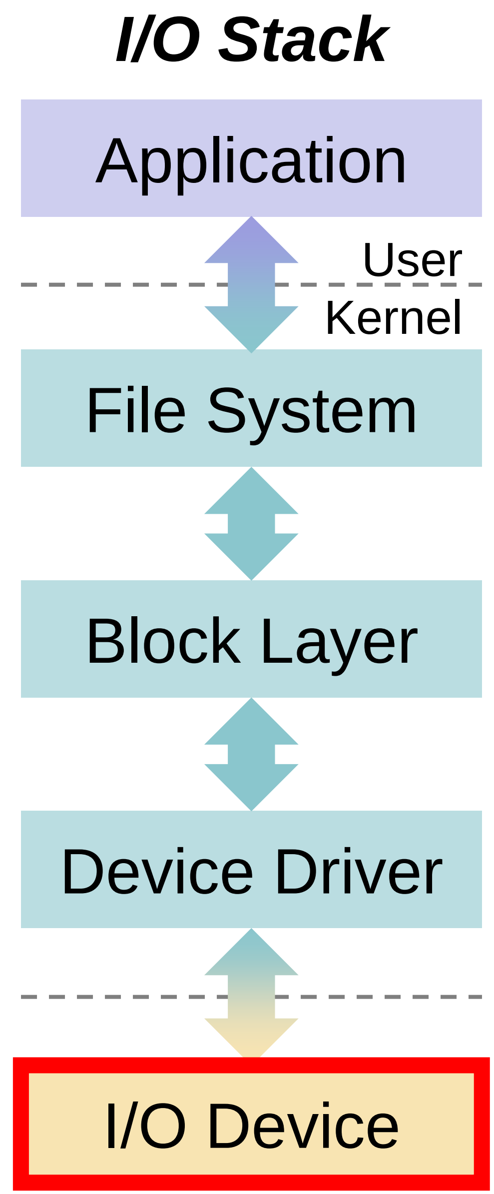






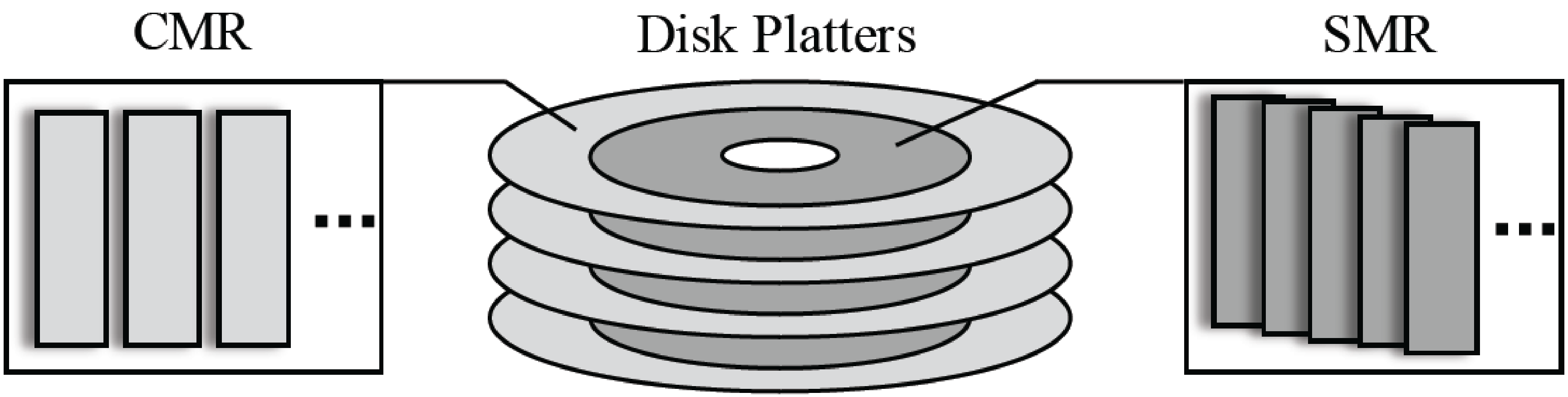
Outline

- Traditional Hard Disk Drive
  - Why and How
  - Development Bottleneck
- New Magnetic Recording Technologies
- Shingled Magnetic Recording (SMR)
  - Basics and Inherent Challenges
  - General Solution: Persistent Cache
  - Various SMR Drive Models and Designs
    - Drive-Managed SMR (DM-SMR)
    - Host-Aware SMR (HA-SMR)
    - Host-Managed SMR (HM-SMR)
    - Hybrid SMR



Hybrid SMR (H-SMR)

- Google introduces the idea of Hybrid SMR recently.
  - A single H-SMR drive can have both CMR and SMR areas.
  - The format can be changed from one type to the other.
  - The goal is to balance the IOPS and the capacity.

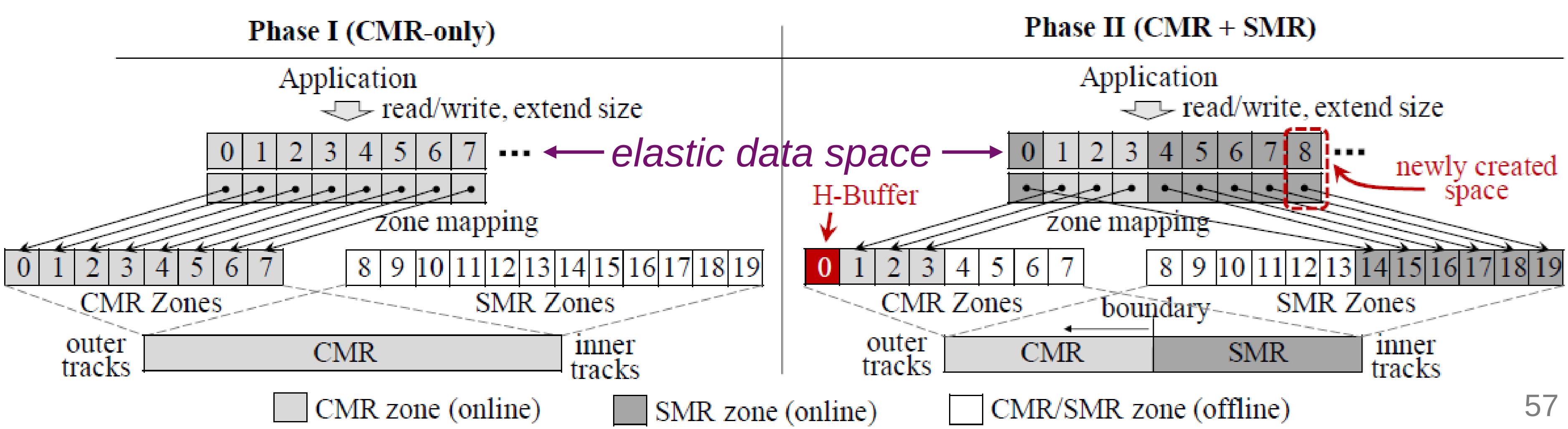


- There are still many challenges:
  - How to efficiently arrange the format layout and place data?
  - How to reduce SMR update overhead?
  - How to adapt to dynamic workloads?

ZoneAlloy - Elastic Data and Space Management for Hybrid SMR Drives (HotStorage'19)

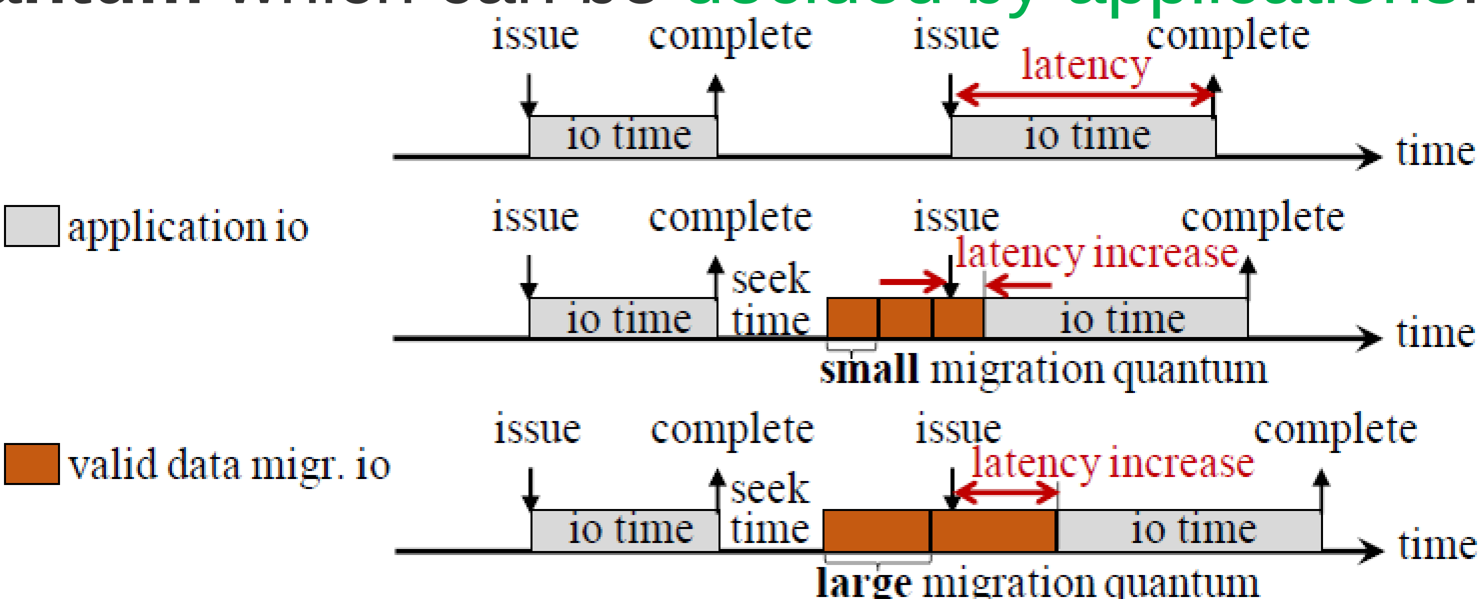
Case Study of H-SMR: ZoneAlloy

- **ZoneAlloy** hides the H-SMR details and presents upper layer applications with an elastic data space.
  - Elastic Data Space: an address space with extendable size.
  - It can be implemented at the block driver layer or firmware.
- **ZoneAlloy** adopts a **two-phase** elastic allocation:
  - Phase I: Initially allocating CMR space only
  - Phase II: Converting from CMR to SMR as necessary
    - Through a quantized migration to control the impact.



ZoneAlloy: Quantized Migration

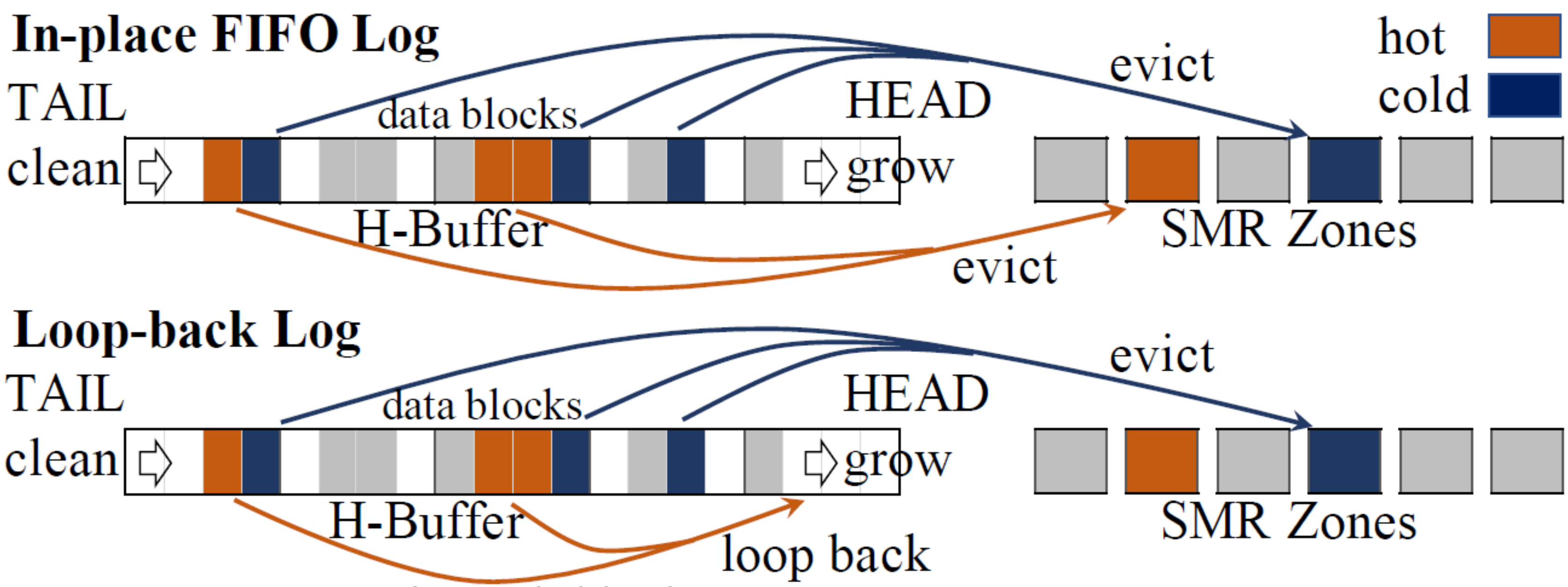
- Conversion is time consuming and intrusive.
  - The application I/O will be delayed by seconds, or even minutes, depending on the size of the requested space.
  - There is a trade-off between the performance of the application I/O and the conversion efficiency.
- Performing the migration in the unit of **migration quantum** which can be decided by applications.



ZoneAlloy - Elastic Data and Space Management for Hybrid SMR Drives (HotStorage'19)

ZoneAlloy: Host-controlled Buffer

- Updating an SMR zone directly using RMW still introduces significant performance overhead.
- A small CMR buffer (called **H-Buffer**) can accumulate multiple updates and migrate them to SMR in a batch.
  - **In-place FIFO Log** organizes the redirected data in a log.
  - **Loop-back Log** keeps hot data in log without evicting them.



ZoneAlloy - Elastic Data and Space Management for Hybrid SMR Drives (HotStorage'19)

Summary

- Traditional Hard Disk Drive
  - Why and How
  - Development Bottleneck
- New Magnetic Recording Technologies
- Shingled Magnetic Recording (SMR)
  - Basics and Inherent Challenges
  - General Solution: Persistent Cache
  - Various SMR Drive Models and Designs
    - Drive-Managed SMR (DM-SMR)
    - Host-Aware SMR (HA-SMR)
    - Host-Managed SMR (HM-SMR)
    - Hybrid SMR

I/O Stack