Shingled Magnetic Recording (SMR)

Presented by

Curtis E. Stevens

IDEMA SMR Committee Chair

IDEMA SMR

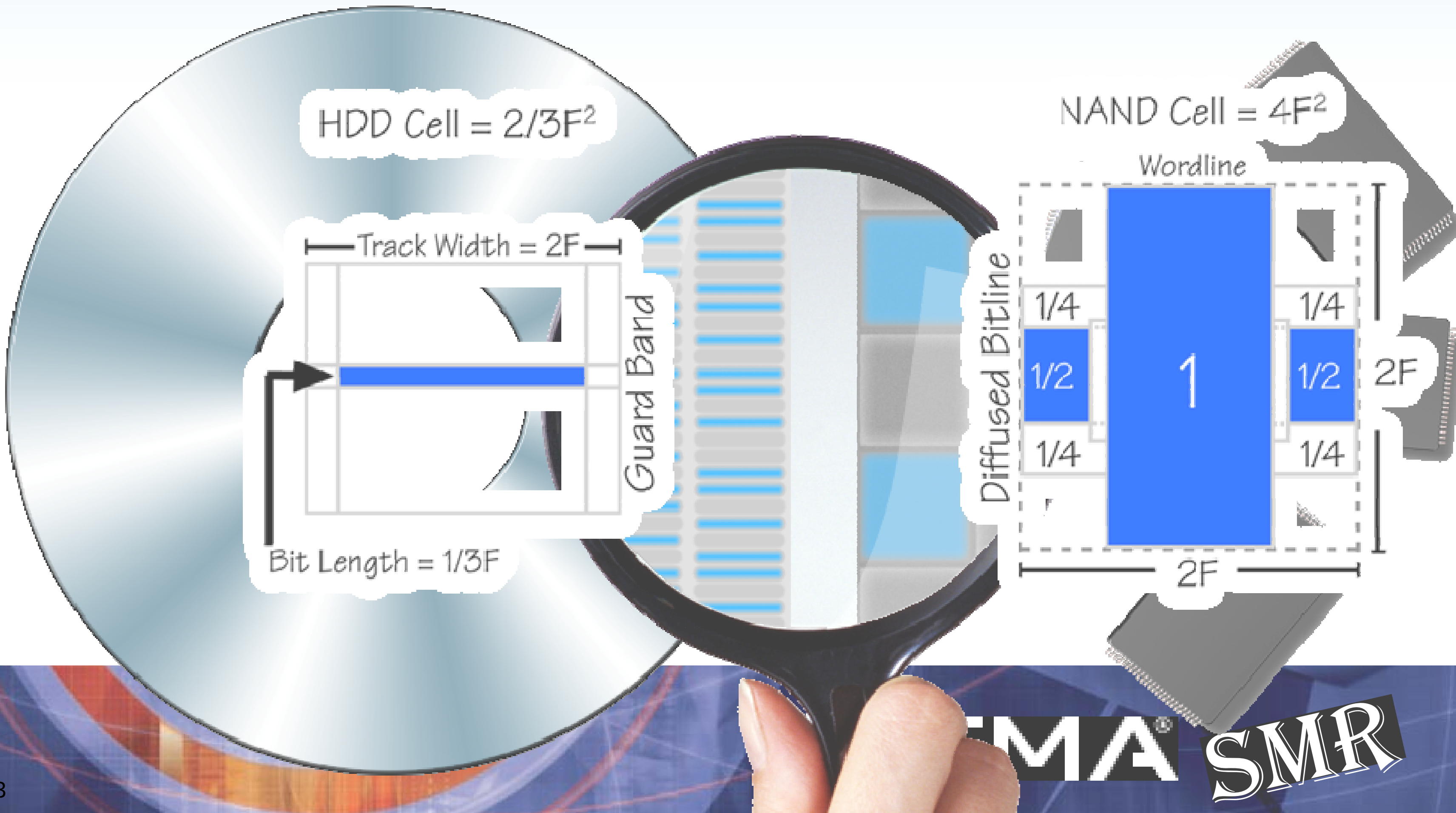
Contents

Bit Cell Size: HDD vs. NAND

HDD has a 6x smaller cell area for a given lithographic feature size (F)
HDD one generation behind on F: 40 nm for HDD vs 3x nm for NAND

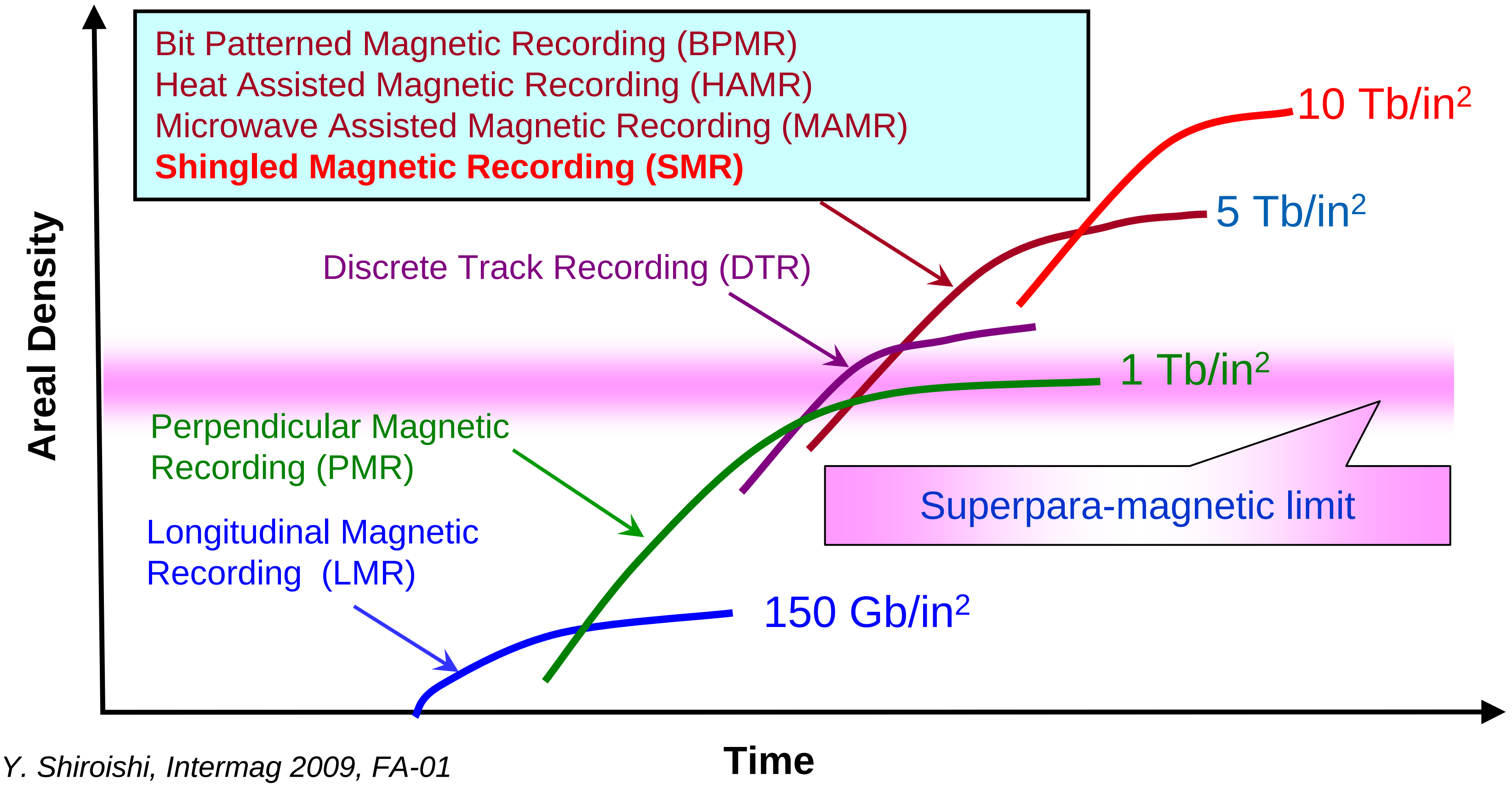
HDD scales better than SSD as long as BAR > N

HDD cell = $4F^2$/BAR | **SSD cell = $4F^2$/N**



Magnetic Recording System Technologies

New recording system technologies are needed to keep the HDD industry on its historical track of delivering capacity improvements over time



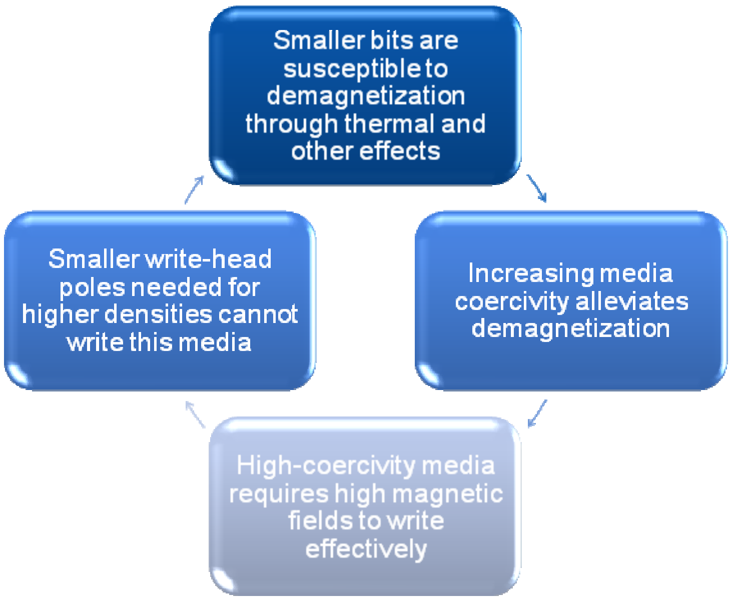
Y. Shiroishi, Intermag 2009, FA-01

Contents

Write Head Design Considerations

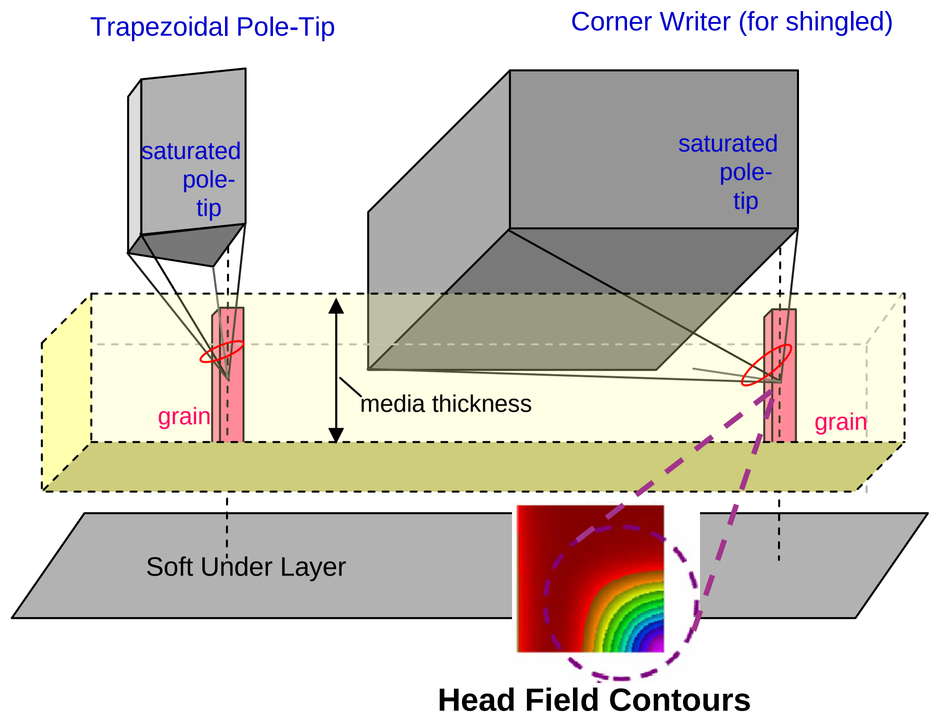
A larger write head addresses the scaling challenge of conventional magnetic recording

Scaling Challenge of Conventional Recording



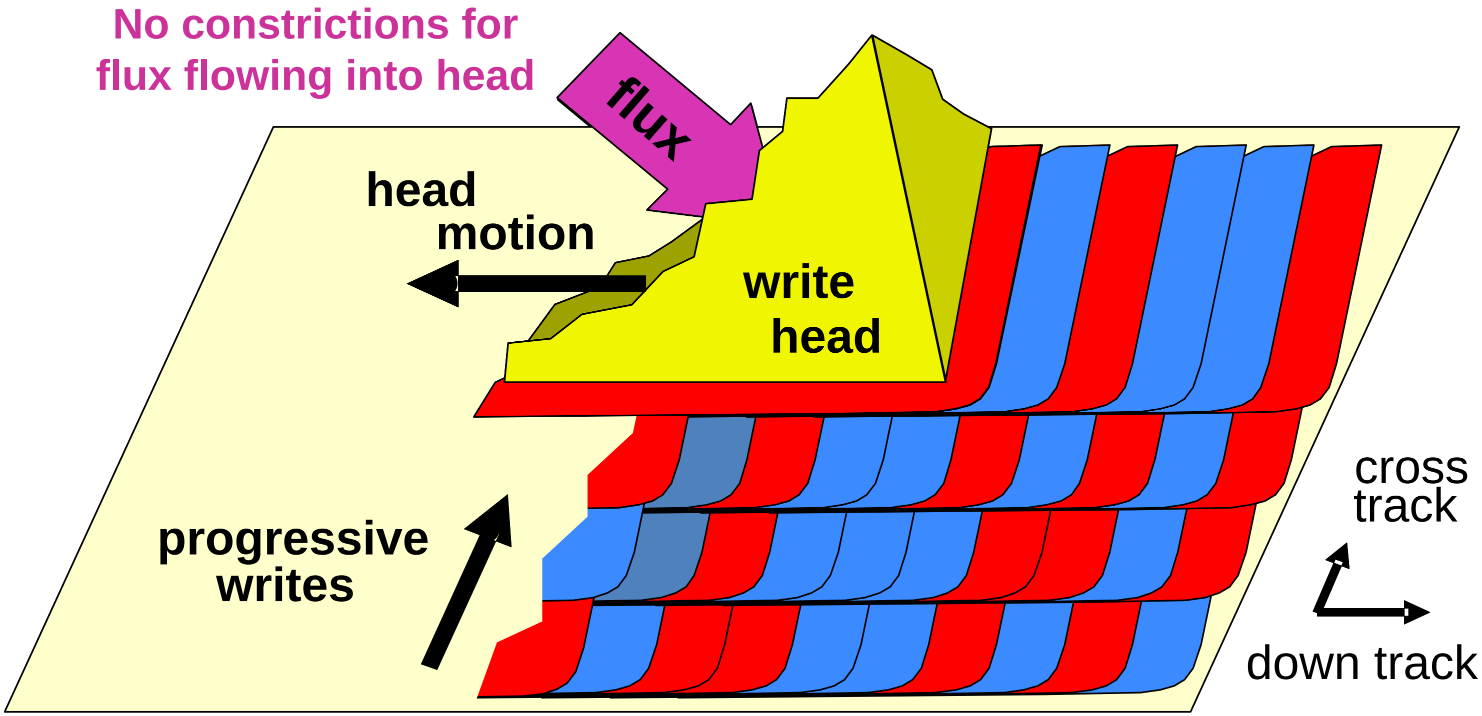
Write Head Comparison

Vertical head field is proportional to solid angle



What is SMR?

SMR write head geometry extends well beyond the track pitch

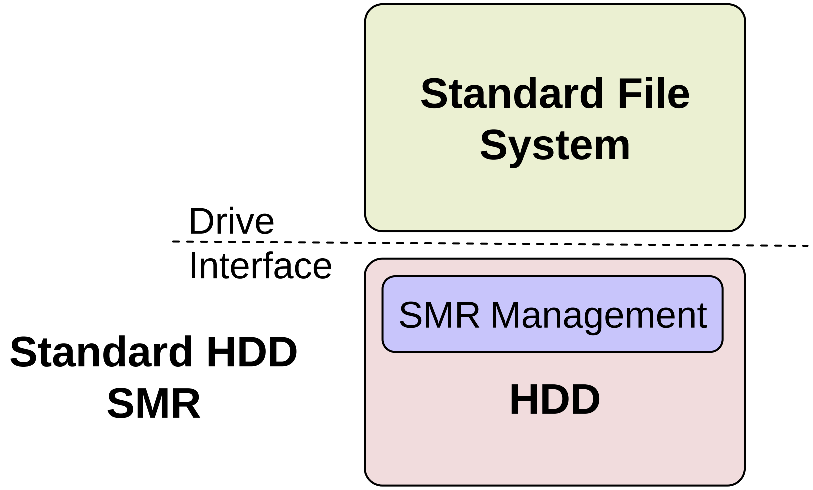


Wood, Williams, et al., IEEE TRANSACTIONS ON MAGNETICS, VOL. 45, NO. 2, FEBRUARY 2009

The larger SMR write head introduces the SMR Constraint

The SMR Constraint

In SMR, the large write head overwrites adjacent, previously written tracks, creating a need for SMR management



- No or little host modifications required
- Drive ensures data integrity and optimizes performance for a balanced set of typical IO workloads

"If [SMR] behavior is exposed to operating systems directly, there will be very low acceptance of these products. However, disk controller software can emulate full compliance with existing interfaces, and may be able to mask almost all performance implications as well."

Gibson and Polte, "*Directions for Shingled-Write and Two-Dimensional Magnetic Recording System Architectures*",
May 2009, Carnegie Mellon University

Contents

SMR Architecture Constructs

- **Read Modified Write**

  read-modify-write operation first reads a portion of data from the disk, then modifies part of that portion with the host provided write data, and finally writes the whole portion back to disk.

- **Shingled Regions**

  a group of tracks that is separated from neighboring shingled regions by a guard band. The purpose of the guard band is to prevent a write in a given region to interfere with data written on other regions. That isolation of interference guarantees that no read-modify-write will need to go beyond a region boundary.

- **Indirection**

  An indirection system is a collection of data structures and algorithms that assigns physical locations to logical block addresses and retrieves physical locations of logical block addresses. In the case of SMR, the indirection system is to be designed to provide good read/write performance for a wide variety of natural workloads.

- **Hints**

  Provide the device with information about the usage of data in the user area of the media. The purpose of hints is to allow the device to manage its data in a more intelligent way. Hints allow the device to increase reliability, performance and data integrity. Hints originate from the file-system.

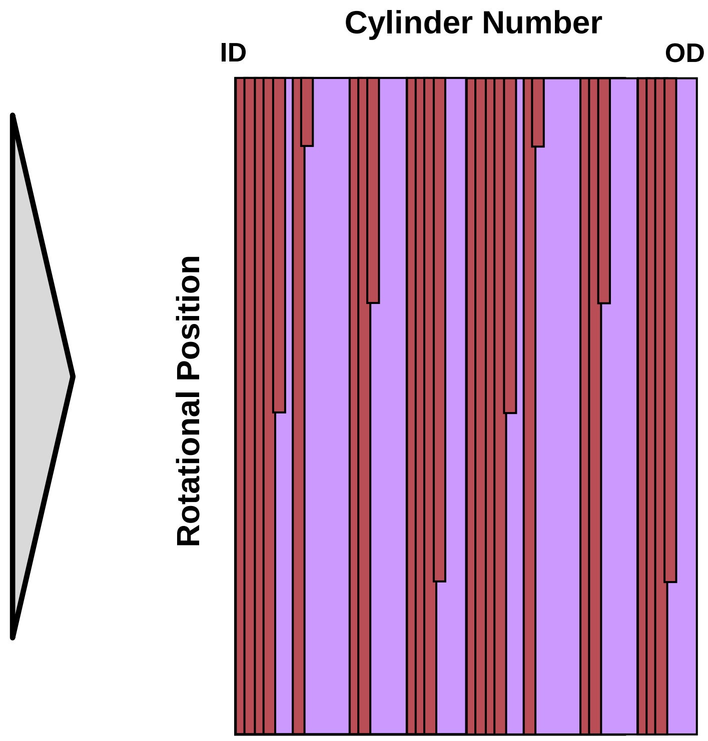
Shingled Regions

Implementing shingled regions is a good option to manage the random write performance impact of SMR

<u>**Simplistic View**</u>
(e.g. if all tracks are shingled)

•Large sequential writes should have reasonable performance

•Small random writes may have very poor performance

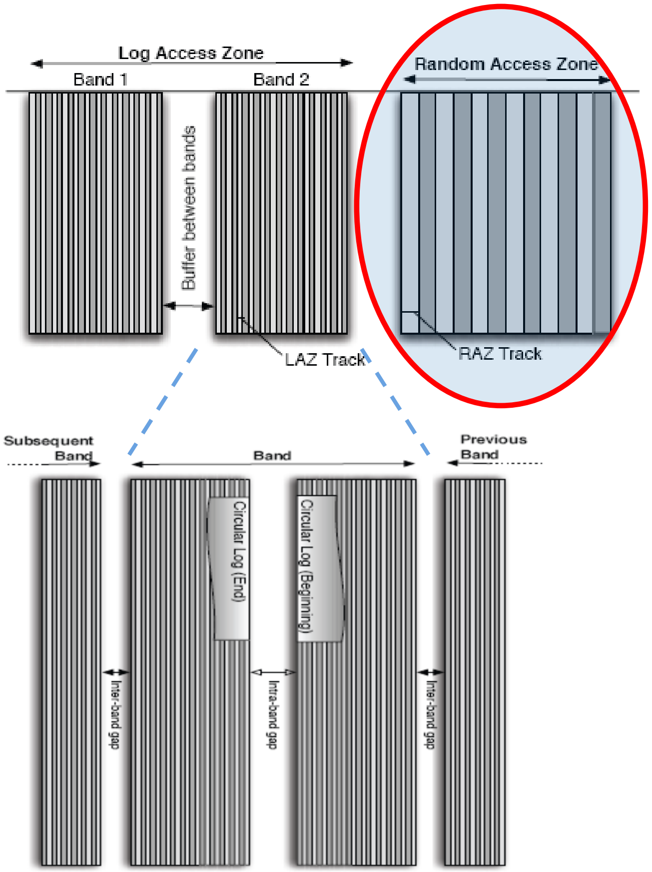


- HDD surface divided into regions of tracks
- Shingled writing within each region

'Write Staging' and Indirection

- **Implementing a persistent 'Write Staging' solution will further improve the random write performance of SMR by delivering a persistent write caching effect**
- **Managing the 'Write Staging' likely involves background processes – data movement within the drive not in direct response to a host write**
- **A logical-to-physical indirection system will need to keep track of the data locations**

One Possible Architecture



'Write Staging'

A. Amer et al. "Design Issues for a Shingled Write Disk System" MSST 2010

LBA Indirection

1 Host wants to rewrite **LBA-13** and **LBA-14**

2 Host seeks for next available unassigned block

3 Host assigns **LBA-13** and **LBA-14** to **PBA-32** and **PBA-33**

PBA-13 and **PBA-14**
Now unassigned

PBA-11 LBA-11	PBA-12 LBA-12	PBA-13 LBA-N/A	PBA-14 LBA-N/A
PBA-21 LBA-21	PBA-22 LBA-22	PBA-23 LBA-23	PBA-24 LBA-24
PBA-31 LBA-31	PBA-32 LBA-13	PBA-33 LBA-14	PBA-34 LBA-N/A
PBA-41 LBA-N/A	PBA-42 LBA-N/A	PBA-43 LBA-N/A	PBA-44 LBA-N/A

Unassigned

IDEMA SMR

TRIM/UNMAP for SMR Drives

Similar to a SSD, a SMR drive's performance benefits from actively managing the instantiation of its logical blocks

Current Assumptions:

•SMR drives support the TRIM/UNMAP commands for basic dataset management

•A never written / freshly trimmed LB does not logically exist on the media

•Reads to a never written or freshly trimmed LBA return may return the original data, random data, or a sector filled with zeroes

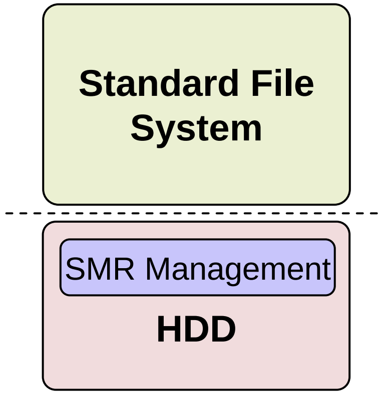
- The drive reports its capabilities for how trimmed sectors are addressed

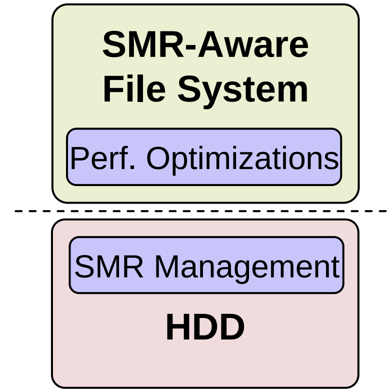
•The drive manages media health autonomously

Contents

Evolution of SMR

An SMR-aware host stack may be able to further optimize the system IO performance





- Drive manages SMR constraint and and optimizes performance for typical IO workloads
- No or little change to host file system

- Host participates in optimizing drive performance
- Potential Implementations:
  - Host stack shapes the IO flow to match advertised drive capabilities and preferences, e.g. for randomness, queuing, IO size, alignment, idle time
  - Host sends data set management 'hints' into drive

IDEMA will continue to work on industry alignment for SMR

IDEMA SMR Working Group Charter

- **Define industry standard Nomenclature & terminology**
- **Development strategies for interface commands to enable shingled writing technologies.**
  - Bring requirements into standards and other organizations
  - Create material for host side implementers including OS and applications
- **Develop marketing materials to describe the above**
  - White papers, presentations

SMR Drive Host Interface Options

- **Standard HDD SMR:** Storage device is built using SMR technology, but the architecture is designed to accommodate all IO requests that a current HDD can handle. The device will hide any internal operations (such as read-modify-write) needed to preserve data integrity.
  - Needed at first introduction – may cause performance issues
- **New or Enhanced Interface:** The storage device will use an interface protocol that will differ from the current generation HDD and SSD interfaces. It could differ by simply restricting the current protocols, adding extra functionality, adopting other protocols, etc. Here are some examples that would fall in this category:
  - **Large Block SMR:** The storage device restricts IO operations to be of a given, potentially fixed, large block size (e.g., 32MB).
  - **HDD Emulation with Hints:** Here the storage device maintains the current flexibility in IO request sizes, but adds extra functionality to allow a host system to provide the storage device with extra information pertaining to the data.
  - **OSD:** The storage device uses SMR technology and implements an OSD (Object Storage Device) interface.

Current Work

- **Shingled Magnetic Recording Technical Paper**
  - Introduces the technology and associated issues
  - Defines industry standard terms for SMR technology
- **This presentation**
- **Segmentation of requirements**
  - Enterprise, Commercial, Consumer, Consumer Electronics
  - Each segment has its own requirements
- **Document repository at wwww.IDEMA.org**
- **Planning a seminar at Microsoft, date TBD**