SMR Standardization

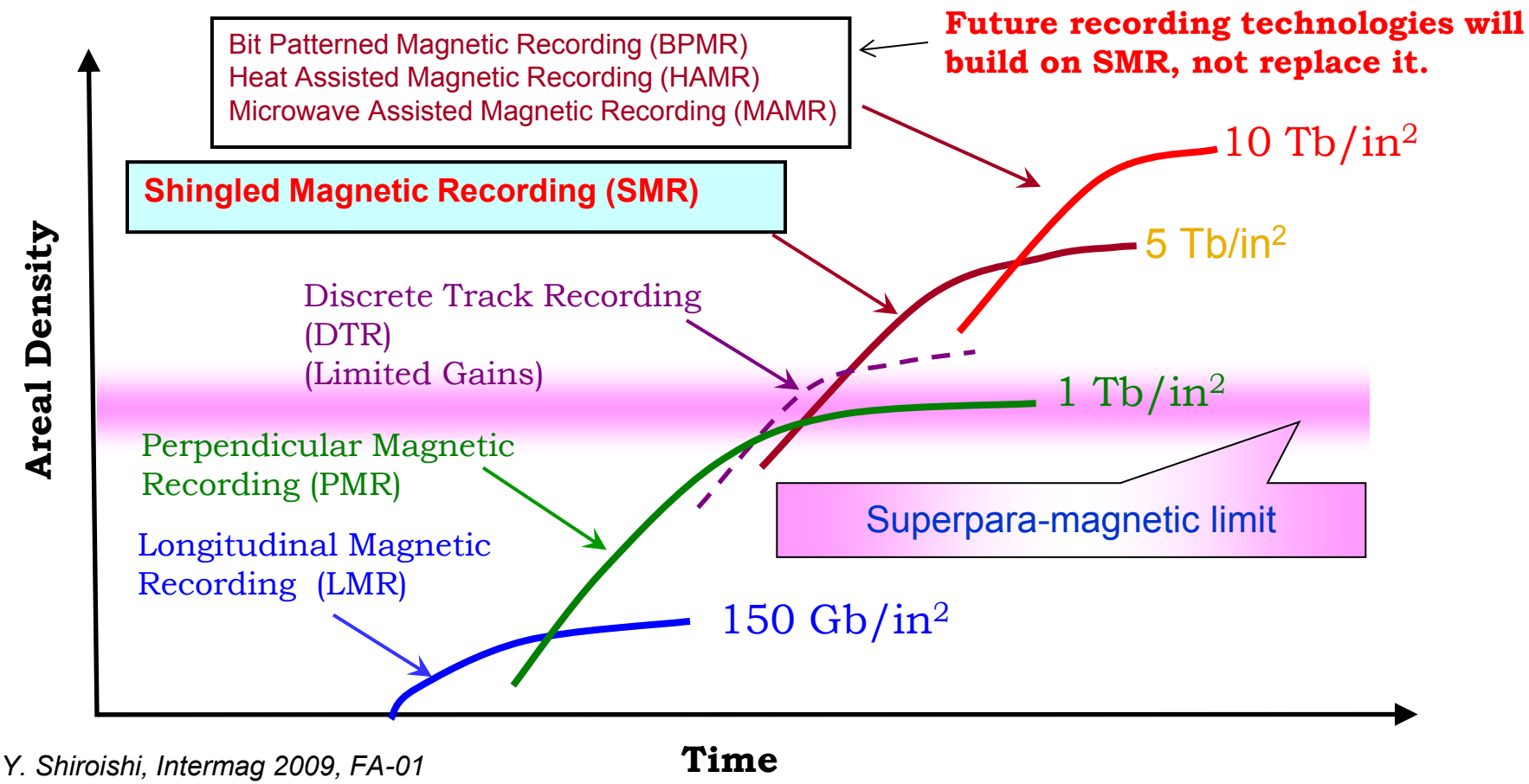
HGST, Seagate, Toshiba, Western Digital

Introduction

- The drive industry is introducing a new areal density enabling technology called Shingled Magnetic Recording.
- This technology alters the throughput and response time behavior of IO, especially random writes.
- To optimize the IO behavior of SMR drives, the drive industry is pursuing a couple of proposals to submit to the standardization bodies.
- Storage industry feedback is being solicited on these proposals.

Magnetic Recording System Technologies

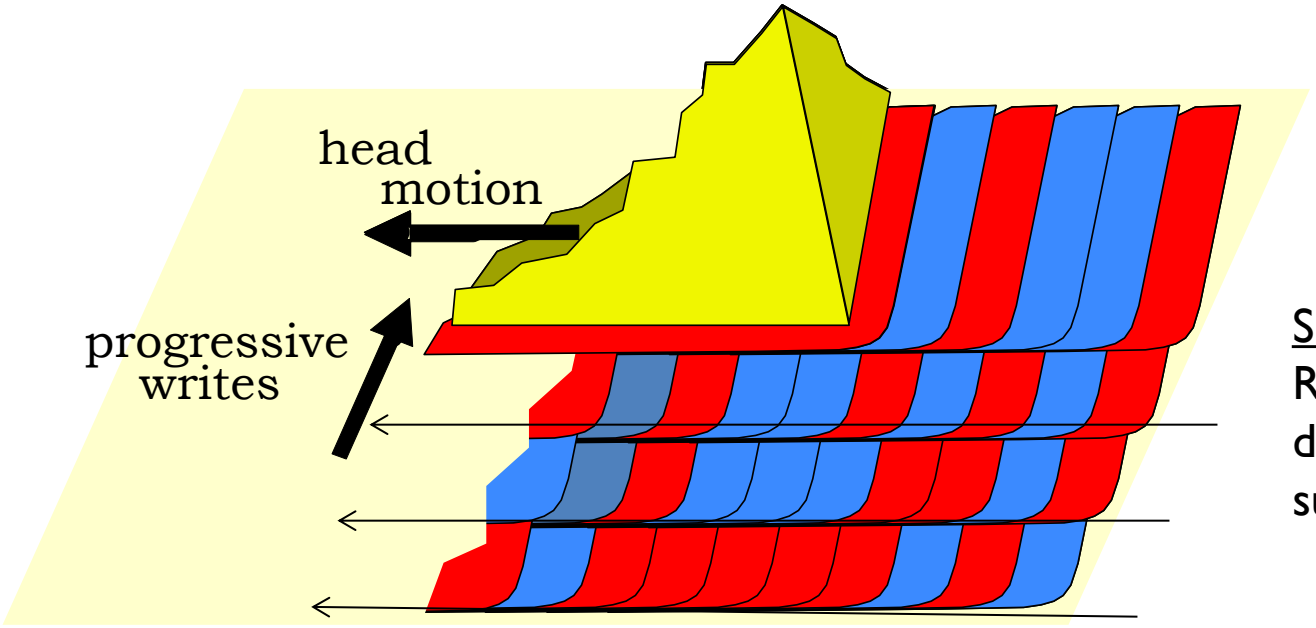
New recording system technologies are needed to keep the HDD industry on its historical track of delivering capacity improvements over time



Y. Shiroishi, Intermag 2009, FA-01

What is Shingled Magnetic Recording (SMR)?

SMR write head geometry extends well beyond the track pitch in order to generate the field necessary for recording. Tracks are written sequentially in an overlapping manner forming a pattern similar to shingles on a roof.



SMR Constraint:
Rewriting a given track will damage one or more subsequent tracks.

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

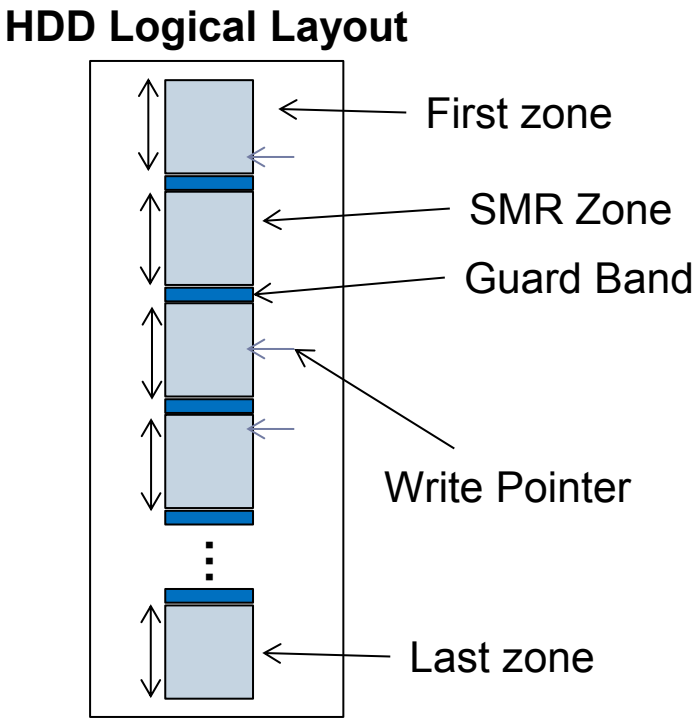
SMR Types

SMR category	Description
Drive managed	No host changes. SMR device manages all requests. Performance is unpredictable in some workloads. Backward compatible
Host aware	Host uses new commands & information to optimize write behavior. If host sends sub-optimal requests the SMR device accepts the request but performance may become unpredictable. Backward compatible
Restricted	Host uses new commands & information to optimize write behavior. ***Performance is predictable***. If host sends sub-optimal requests the SMR device rejects the request. Not backward compatible

Benefits of SMR
(Sequential Writing vs. Device-Managed random writing)

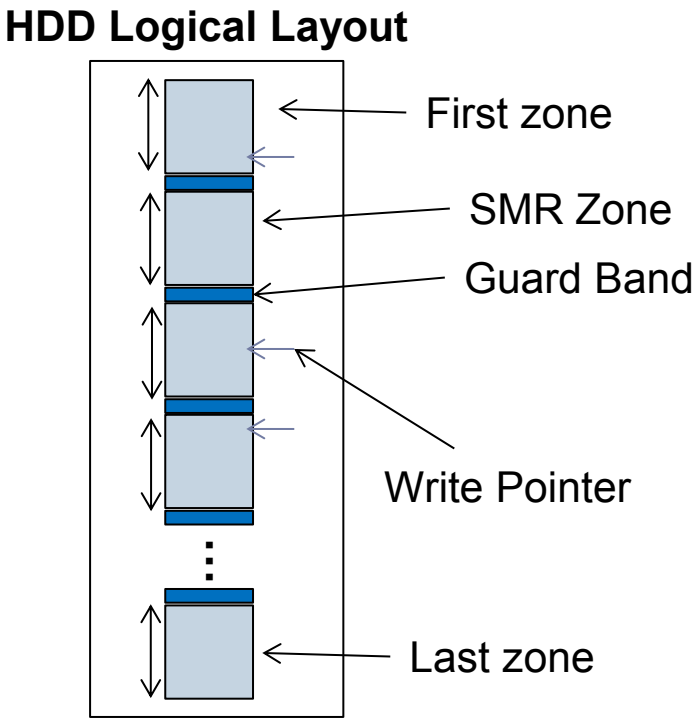
- Design
  - SMR drives are designed to write sequentially to the media
  - Some drives have multiple zones with guard bands between them
  - Other drives may have a single zone that encompasses the entire drive
- Effect if the rules are followed
  - SMR drives have better mechanisms to limit adjacent track interference and wide area track effects
    - These effects are minimized because guard bands separate the SMR zones
  - No Background or foreground tasks are required to mitigate issues associated with breaking the rules
- End result if the rules are followed
  - Cost /TB is reduced due to SMR technology
  - Less data movement relative to a device-managed solution
    - Power consumption and related metrics are reduced
  - Performance is optimal

SMR Interface Principles



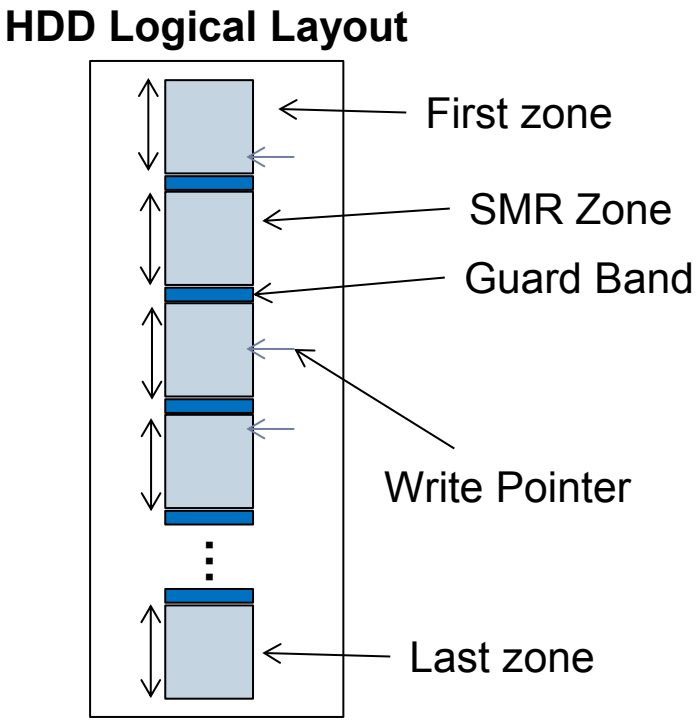
- SMR drives may be divided into zones: Random Write and SMR
- A Random Write zone is a range of LBAs that may be written randomly
  - The device provides media in this area that is optimized for random writing.
  - This type of zone may raise the cost of the device or lower the overall capacity of the device
- An SMR zone is a zone that should be written sequentially
  - The drive maintains a write pointer for each zone
    - If the initiator does not write starting at the write pointer, ***then the SMR zone rules have been violated***
  - The drive may provide a mechanism to retrieve the write pointer
    - Important for surprise power removal and data recovery operations
    - May be a part of the zone map described below
  - The device provides a mechanism to reset the write pointer to the beginning of the zone

SMR Interface Principles (continued)



- The drive provides a mechanism to report the zone map
  - Zones (Start and Length) and zone types (SMR or Random Write)
- If the SMR zone rules are violated
  - Devices may have unpredictable performance of non-sequential write commands
  - Devices may have unpredictable performance of future read and write commands
  - Devices may fail commands to maintain consistent performance
- New capabilities to be specified in INCITS T10 and T13 committees

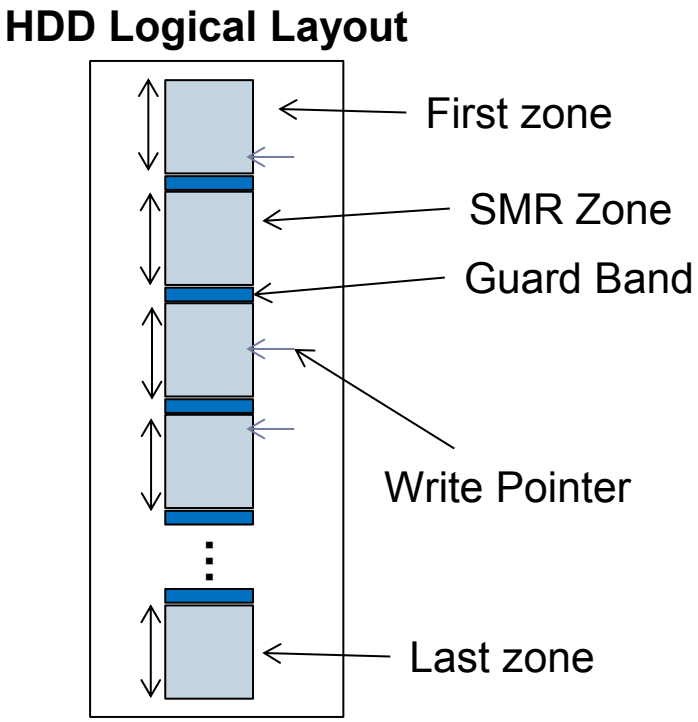
Restricted



- For SMR zones the following is true
  - The device expects writes to be at the write pointer
  - The device treats writes at location other than the write pointer as an exception causing a failed write command
  - No mitigations for non-sequential writes
  - Maintains consistent performance
    - Minimizes data movement
  - Not backward compatible
- Random access zones are optional
  - Random access zones do not have a write pointer
  - Backward compatible

Host Aware

(Drive accepts random write commands by providing assistance)



- For SMR zones the following is true:
  - The device expects writes to be at the write pointer
  - Assistance enable random access to be used throughout the drive volume
  - Assistance may cause inconsistent performance on current and future commands
  - Backward compatible
- Random access zones are optional
  - Random access zones do not have a write pointer
  - Backward compatible

Discussion Questions

- Do you prefer to fail commands that would have an adverse effect on performance or data movement, or is compatibility more important?
  - Does reducing background operations affect your position?
- Do you require compatibility?
  - Do you prefer sequential write pointer mitigations to be done during the offending write command? or
  - Is it better to maintain consistent write performance and slow down future reads or writes
- Is it better to follow the rules and have the system perform garbage collection activities?
  - Versus having the drive perform garbage collection

Drive formatting trade-offs

- How much drive capacity needs to be random write versus sequential write?
- How much of the workload is expected to be random writes?
- How many random write zones are needed?
- Are there other application client specific factors?
- Is 256 MiB SMR zone size a good choice?
- Are Random Write Zones needed?
- What is impact of restricting the number of "active" zones (i.e., zones with write pointers that are neither zero nor the max size)?

Continued Participation

- The SMR study group encourages input in this process
- This development is currently taking place as an INCITS T10 study group
  - See www.t10.org for more information
- Contact the following people for more information
  - Curtis Stevens
    - WD (a Western Digital company) – Curtis.Stevens@wdc.com
  - Gerry Houlder
    - Seagate – Gerry.Houlder@seagate.com
  - Jorge Campello
    - HGST (a Western Digital company) – Jorge.Campello@hgst.com
  - Patrick Hery
    - Toshiba – Patrick.Hery@taec.toshiba.com