A decorative graphic consisting of multiple parallel, wavy lines in various colors (purple, blue, orange, grey, green) that flow from the left side of the slide towards the right, creating a sense of movement and depth.

Shingled Magnetic Recording Models, Standardization, and Applications

Mary Dunn, Seagate
Timothy Feldman, Seagate

- ◆ The material contained in this tutorial is copyrighted by the SNIA unless otherwise noted.
- ◆ Member companies and individual members may use this material in presentations and literature under the following conditions:
 - ◆ Any slide or slides used must be reproduced in their entirety without modification
 - ◆ The SNIA must be acknowledged as the source of any material used in the body of any document containing material from these presentations.
- ◆ This presentation is a project of the SNIA Education Committee.
- ◆ Neither the author nor the presenter is an attorney and nothing in this presentation is intended to be, or should be construed as legal advice or an opinion of counsel. If you need legal advice or a legal opinion please contact your attorney.
- ◆ The information presented herein represents the author's personal opinion and current understanding of the relevant issues involved. The author, the presenter, and the SNIA do not assume any responsibility or liability for damages arising out of any reliance on or use of this information.

NO WARRANTIES, EXPRESS OR IMPLIED. USE AT YOUR OWN RISK.

- ◆ Shingled Magnetic Recording – Models, Standardization, and Applications
 - ◆ This session will appeal to storage developers that are seeking an understanding of shingled magnetic recording and the standards that support it. The session will delve into the various SMR device types: their models, rules, best practices, and extensions to the T10 SCSI and T13 ATA standards. With information for both developers and system designers, the session will also bring a clear understanding of the alternatives, and provide a framework for device selection. The audience will receive a grounding in SMR and how they can make best use of drives with this fundamental recording methodology.

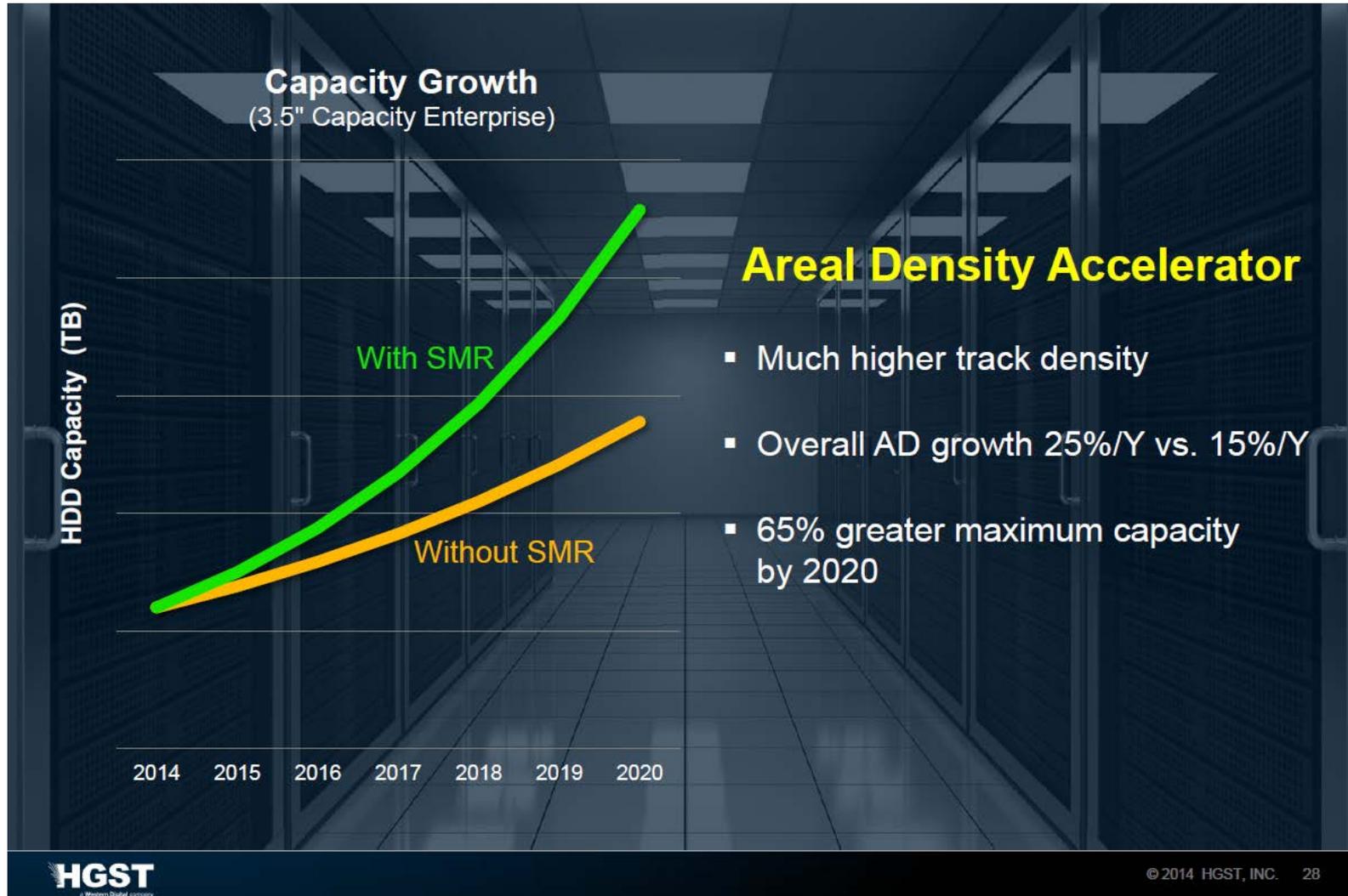
Agenda

- Introduction to Shingled Magnetic Recording
- SMR Implementation Options
- Drive Managed SMR Overview
- Host Supported SMR Standards
- Host Aware Zoned Block Device Overview
- Host Managed Zoned Block Device Overview
- Comparisons and Next Steps

Agenda

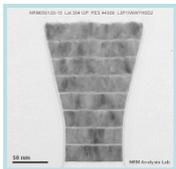
- Introduction to Shingled Magnetic Recording
- SMR Implementation Options
- Drive Managed SMR Overview
- Host Supported SMR Standards
- Host Aware Zoned Block Device Overview
- Host Managed Zoned Block Device Overview
- Comparisons and Next Steps

Why Shingled Magnetic Recording?

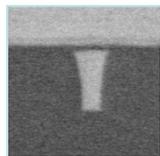


Conventional Perpendicular Writer

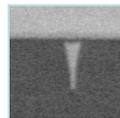
- The focus for new writers is continual improvement in writability, field contour and track width control at ever smaller geometries while reducing ‘side writing’ effects
 - ◆ Progressively higher TPI transforms write pole from well defined trapezoidal to triangular shape leading to increasingly greater losses in writability beyond what was previously associated with writer width reduction.
- Writer design is now close to geometry limitations that preclude further growth in track density without new innovation or new recording technology
 - ◆ Write pole SEM micrographs:



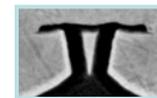
1st Generation
2006



2nd Generation
2008



3rd Generation
2009



4th Generation
2010

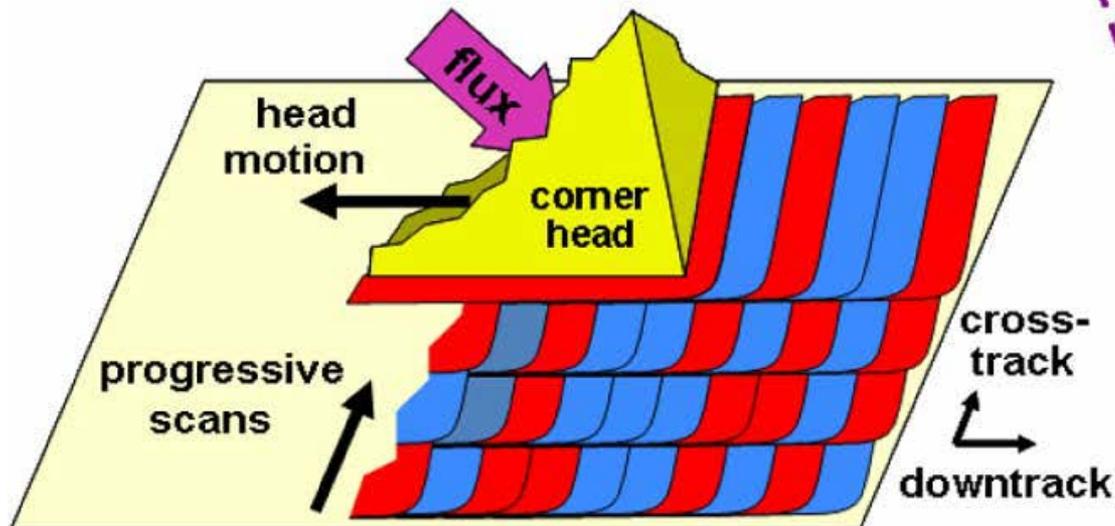


5th Generation
2011

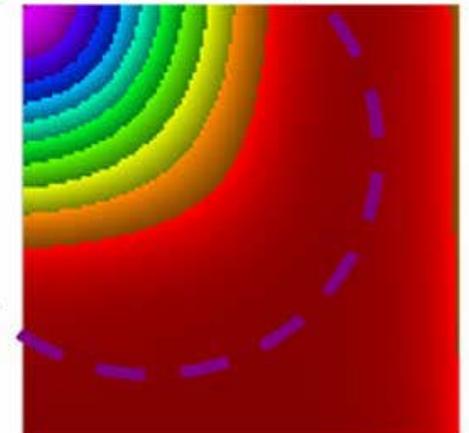
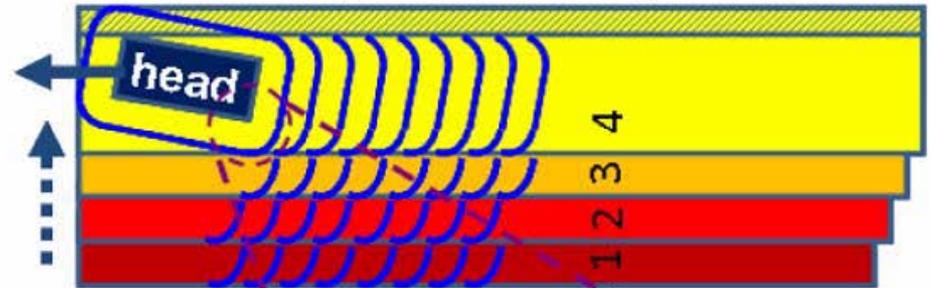
What is SMR?

Shingle-write Process

- Tracks are heavily overlapped.
- Insensitive to pole-width variation
- Only one corner of write-head is important for design.
- No flux constrictions into head
- No ATE (no repetitive writes)



track layout for shingle-write



head field contours

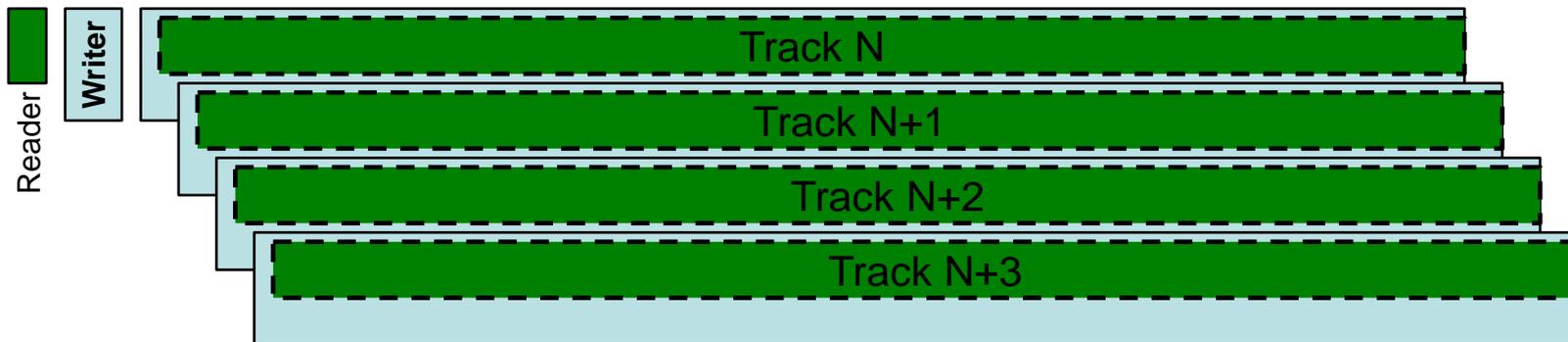
Wood, et al.: The Feasibility of Magnetic Recording at 10 Terabits Per Square Inch on Conventional Media, IEEE Transactions on Magnetics, Vol. 45, No. 2, February 2009

Conventional versus SMR Writing

Conventional Writes



SMR Writes



Agenda

- Introduction to Shingled Magnetic Recording
- **SMR Implementation Options**
- Drive Managed SMR Overview
- Host Supported SMR Standards
- Host Aware Zoned Block Device Overview
- Host Managed Zoned Block Device Overview
- Comparisons and Next Steps

Overview of SMR Drive Types

➤ Drive Managed

- ◆ Drive autonomously hides all SMR issues
- ◆ Workloads can affect performance

➤ Host Aware

- ◆ Superset of Drive Managed and Host Managed
- ◆ Backward compatible
- ◆ Extensions to ATA and SCSI command sets

➤ Host Managed

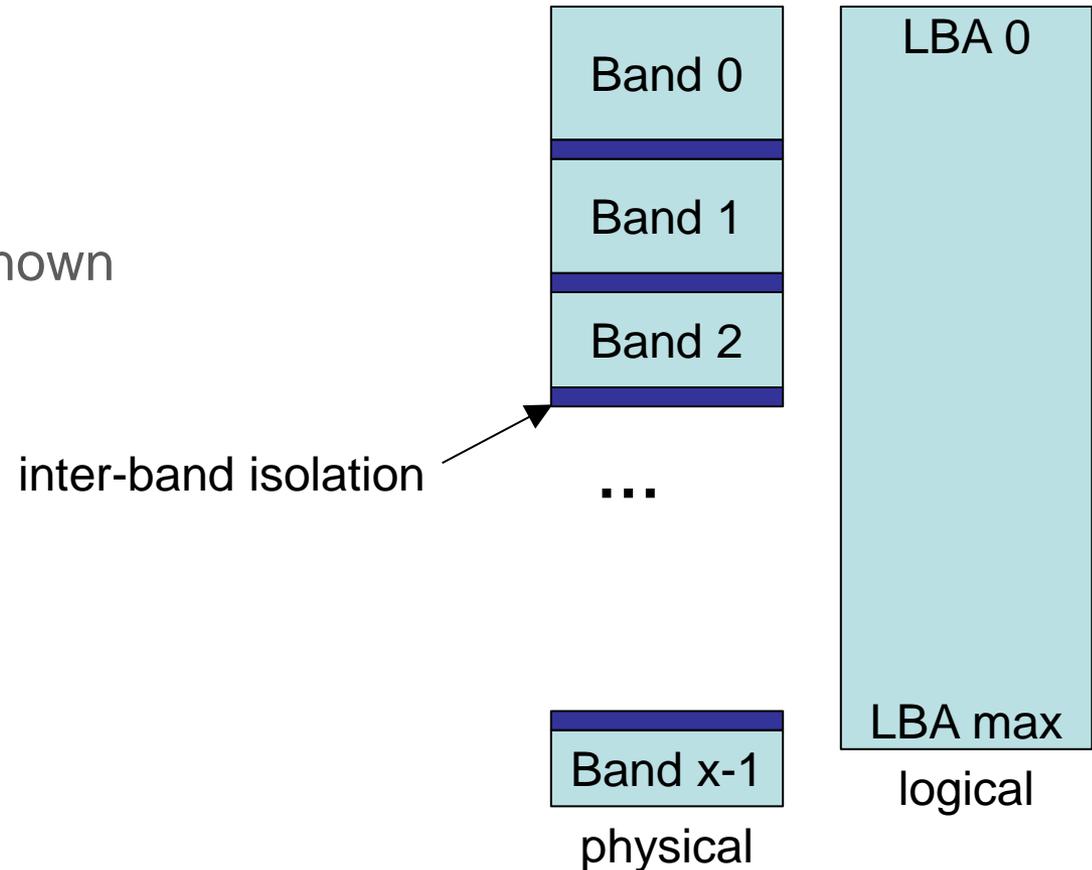
- ◆ New device type
- ◆ Extensions to ATA and SCSI command sets
- ◆ Error conditions for some reads and writes
- ◆ Not backward compatible



Overview of Bands and Zones

➤ SMR Bands

- ◆ Physical construct
- ◆ Boundaries are not known outside the drive



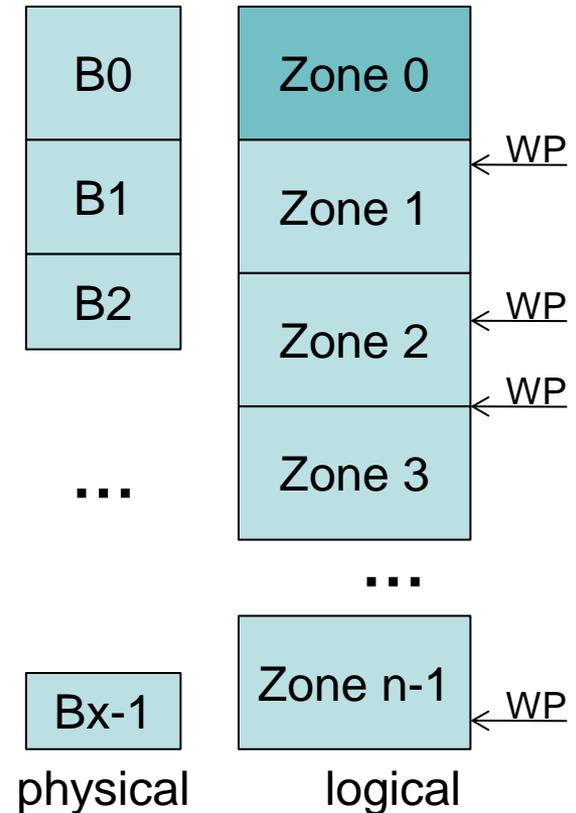
Overview of Bands and Zones

➤ SMR Bands

- ◆ Physical construct
- ◆ Boundaries are not known outside the drive

➤ Zones

- ◆ Logical space is divided into zones
 1. Conventional zones
 2. Write pointer zones
 - › Each has a write pointer
 - Automatically advances
 - Hosts issues reset before re-write
1. Sequential Write Preferred zones
 - Random writes are supported
 2. Sequential Write Required zones
 - Random writes are not allowed



Comparison of SMR Device Types

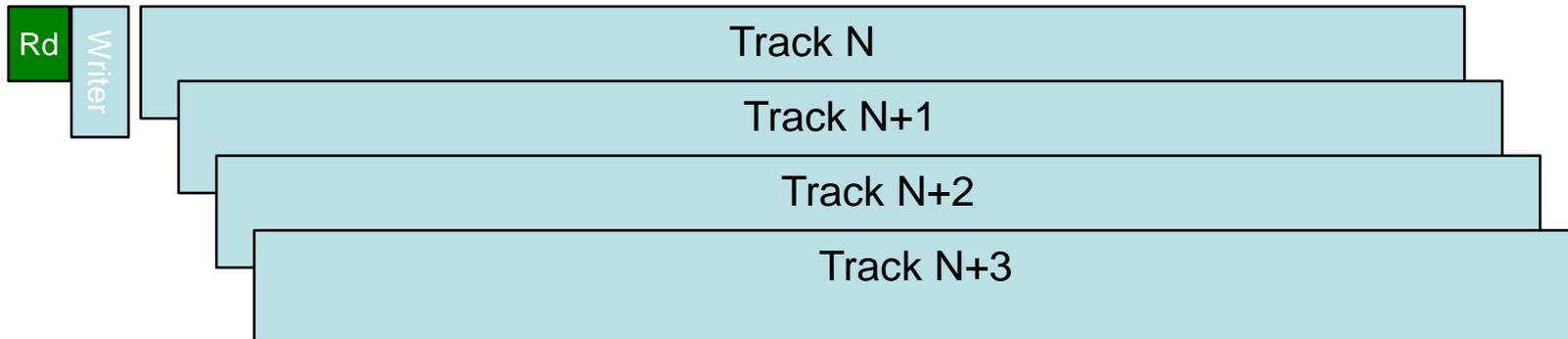
Style	SCSI Peripheral Device Type	ATA Device Signature	Zone Types	New Commands	New Rules
Drive Managed	00h: Direct Access Device	ATA	None	None	None
Host Aware	00h: Direct Access Device (with Host Aware flag)	ATA	Sequential Write Preferred *	<ul style="list-style-type: none"> • Report Zones • Reset Write Pointer 	None
Host Managed	14h: Host Managed Zoned Block Device	Host Managed Zoned	Sequential Write Required *	<ul style="list-style-type: none"> • Report Zones • Reset Write Pointer 	<ul style="list-style-type: none"> • No random writes to WP zones • No reads of unwritten data • Etc.

*conventional zones are optional

Agenda

- Introduction to Shingled Magnetic Recording
- SMR Implementation Options
- **Drive Managed SMR Overview**
- Host Supported SMR Standards
- Host Aware Zoned Block Device Overview
- Host Managed Zoned Block Device Overview
- Comparisons and Next Steps

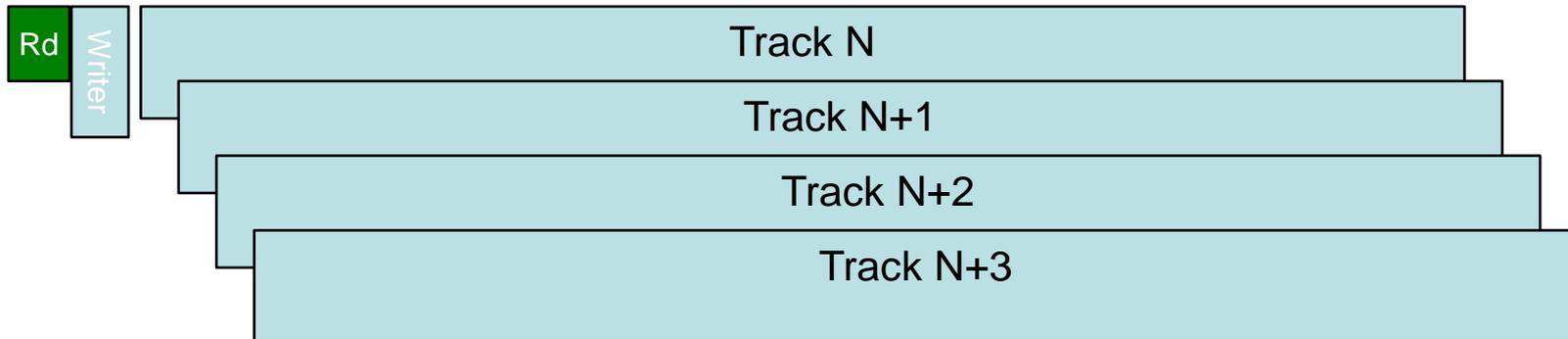
Updating a band with new data



Data Buffer	
Old Track N	Data
Old Track N+1	Data
Old Track N+2	Data
Old Track N+3	Data

1. Read old data

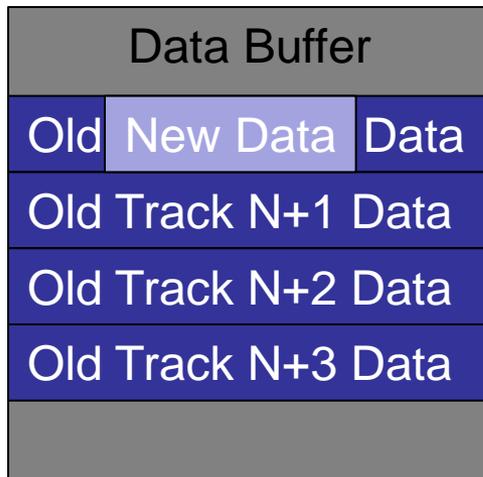
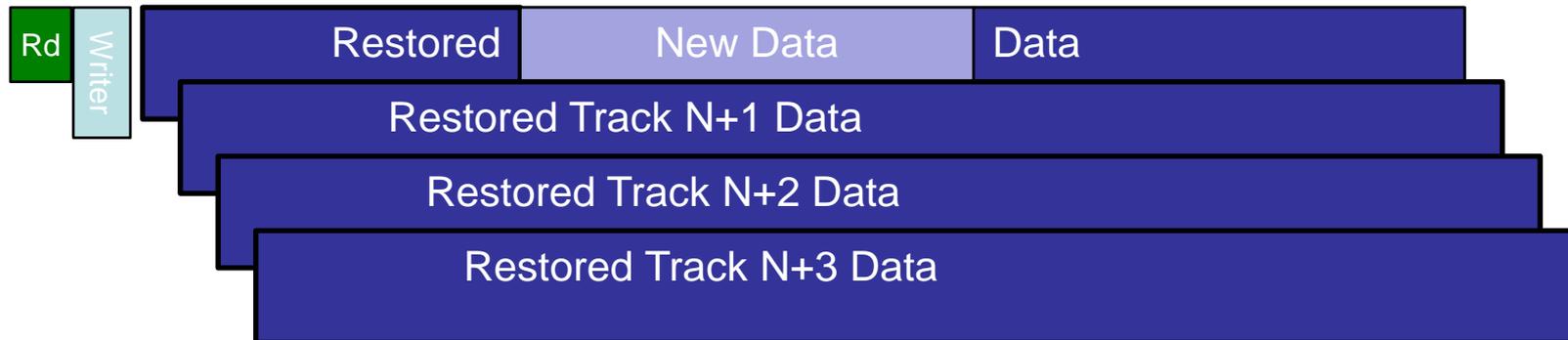
Updating a band with new data



Data Buffer		
Old	New Data	Data
Old Track N+1 Data		
Old Track N+2 Data		
Old Track N+3 Data		

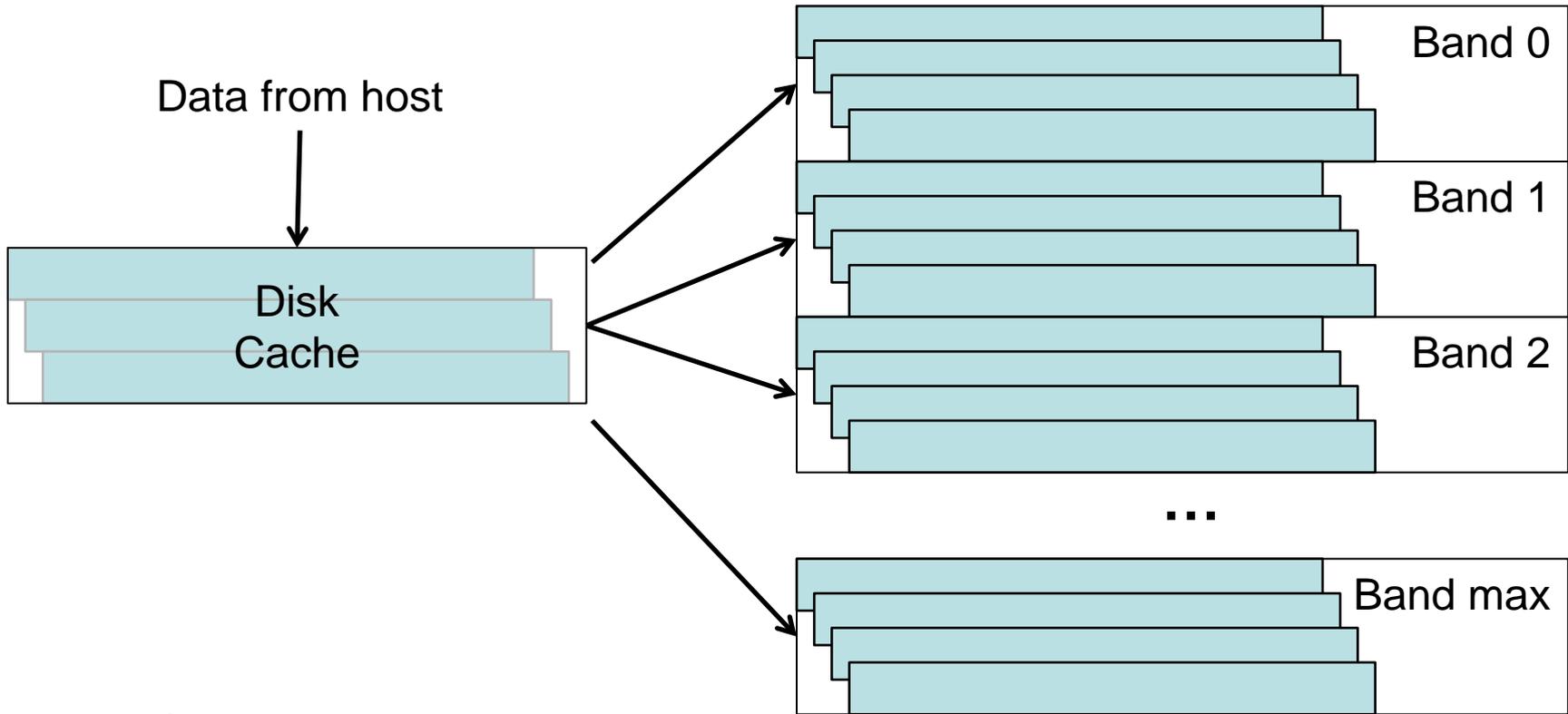
1. Read old data
2. Merge with new data

Updating a band with new data



1. Read old data
2. Merge with new data
3. Write new data, refreshing old data

Random Write Performance With Disk Cache



Summary of a Drive Managed Implementation

- **Write-back for random writes**
 - ◆ Large disk cache
 - ◆ Fast response for bursty workloads
 - ◆ Aggregation of multiple commands before band update
- **Write-around for sequential writes**
 - ◆ Conventional performance at media data rate
- **No host changes required**
 - ◆ Performance similar to conventional drives in client benchmarks
 - ◆ SMR Best Practices to extend into other applications

Drive Managed Best Practices

➤ Reads

- ◆ Same as non-SMR

➤ Writes

- ◆ **Maximize long sequential write runs**
 - › Align to 4-KiB physical sectors
 - › Limit number of interleaved sequential writes
- ◆ **Limit and concentrate random writes**
 - › Use a few small partitions

Agenda

- Introduction to Shingled Magnetic Recording
- SMR Implementation Options
- Drive Managed SMR Overview
- **Host Supported SMR Standards**
- Host Aware Zoned Block Device Overview
- Host Managed Zoned Block Device Overview
- Comparisons and Next Steps

◆ Host assisted SMR enables:

- ◆ Consistent, conventional performance
 - › When rules and best practices are followed, native drive fundamentals govern performance: spin speed, seek time, data rate
- ◆ Larger SMR bands
 - › Increase areal density entitlement
- ◆ SMR in more markets

- New models, feature sets, commands, logs, parameters
- SCSI first, then ATA
 - ◆ History has shown this is fastest
- T10 (SCSI) Zoned Block Commands – ZBC
 - ◆ Revision 1b is published
 - 1 new Peripheral Device Type
 - 2 new commands: Report Zones and Reset Write Pointer
 - 3 new VPD parameters in a new page
 - ◆ Additional proposals under development
 - ◆ January 2015 for Version 1
- T13 (ATA) Zoned block ATA Commands – ZAC
 - ◆ Fast follower to ZBC

◆ Report Zones

- ◆ Reports configuration and current state of zones
 - › Type, Condition, Size, Start LBA, Write Pointer
- ◆ Report can be restricted by type or condition
- ◆ SAME flag in returned header specifies that all zones are the same size and type
- ◆ No method to change the configuration in the field

◆ Reset Write Pointer

- ◆ Resets the write pointer of a zone to the start
 - › RESET ALL flag in command specifies that all zones are to be reset
- ◆ All LBAs become unwritten

Agenda

- Introduction to Shingled Magnetic Recording
- SMR Implementation Options
- Drive Managed SMR Overview
- Host Supported SMR Standards
- **Host Aware Zoned Block Device Overview**
- Host Managed Zoned Block Device Overview
- Comparisons and Next Steps

◆ Host Aware device type

- ◆ SCSI: Direct Access Device with HAW_ZBC=1
- ◆ ATA: ATA device signature with Host Aware Zones feature set

◆ Two zone types

1. Sequential Write Preferred zones

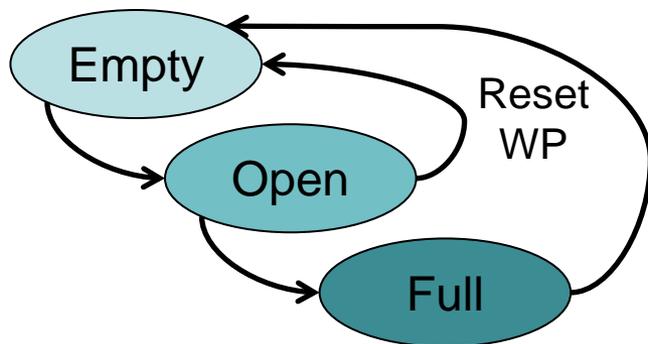
- Each has a write pointer to indicate preferred write location
- Can be any media including SMR and flash
- In one or multiple states
 - Empty, Open, Full, Offline, Read Only

2. Conventional zones

- No write pointer
- Can be any media including non-SMR, Drive Managed SMR and flash

Sequential Write Preferred Zones

- LBA space is made up of zones
- Each zone has a write pointer
- Writes at the write pointer have conventional performance
 - ◆ Write pointer automatically advances
- Writes not at the write pointer handled like Drive Managed
 - ◆ Write pointer may or may not advance
- Issue Reset Write Pointer before re-writing



- Empty
 - ◆ Write pointer is at start of zone
- Open
 - ◆ Write pointer is mid-zone
- Full
 - ◆ No write pointer value

Sequential Write Preferred Zones

➤ Write rules

- ◆ Writes preferably start at the write pointer
- ◆ Writes may start anywhere and span any extent
- ◆ Writes to zones in Offline and Read Only states are not allowed

➤ Read rules

- ◆ Reads may start anywhere and span any extent
- ◆ Unwritten logical blocks return initialization pattern
- ◆ Reads to zones in Offline state are not allowed

➤ Open zones

- ◆ Optimal Number Of Open Sequential Write Preferred Zones
 - › In new ATA log or SCSI VPD page
 - › Advises the largest number of zones that should be open for best performance

➤ Random write zones

- ◆ Optimal Number Of Non-Sequentially Written Sequential Write Preferred Zones
 - › In new ATA log or SCSI VPD page
 - › Advises the largest number of zones that should be randomly written for best performance

- Maximize long sequential write runs
 - ◆ Align commands to the drive's reported physical sectors
 - ◆ Limit number of open zones
 - › Stay within the advisory value
 - ◆ Issue Reset Write Pointer before re-use
 - › Eliminate unnecessary internal re-writes of stale data
- Limit and concentrate random writes
 - ◆ Allocate a few zones for random writes
 - › Stay within the advisory value

Agenda

- Introduction to Shingled Magnetic Recording
- SMR Implementation Options
- Drive Managed SMR Overview
- Host Supported SMR Standards
- Host Aware Zoned Block Device Overview
- **Host Managed Zoned Block Device Overview**
- Comparisons and Next Steps

➤ Host Managed device type

- ◆ SCSI: Host Managed Zoned Block Device
- ◆ ATA: Host Managed Zoned device signature

➤ Two zone types

1. Sequential Write Required zones

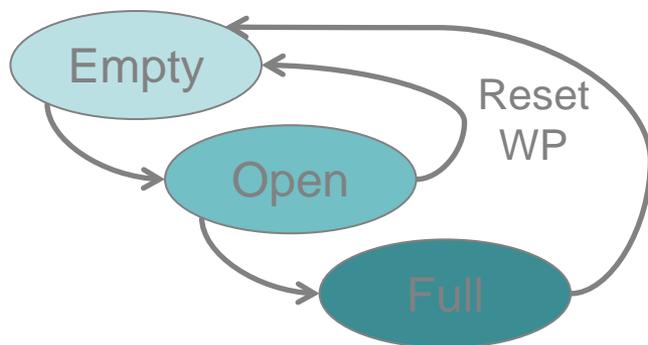
- Each has a write pointer to indicate ~~preferred~~ required write location
- Can be any media including SMR and flash
- In one or multiple states
 - Empty, Open, Full, Offline, Read Only

2. Conventional zones

- No write pointer
- Can be any media including non-SMR, Drive Managed SMR and flash

Sequential Write Required Zones

- LBA space is made up of zones
- Each zone has a write pointer
- Writes at the write pointer have conventional performance
 - ◆ Write pointer automatically advances
- ~~Writes not at the write pointer handled like Drive Managed~~
- ~~Writes not at the write pointer fail~~
 - ◆ ~~Write pointer may or may not advance~~
- Issue Reset Write Pointer before re-writing



- Empty
 - ◆ Write pointer is at start of zone
- Open
 - ◆ Write pointer is mid-zone
- Full
 - ◆ No write pointer value

Sequential Write Required Zones

Write rules

- ◆ Writes ~~preferably~~ must start at the write pointer
- ◆ ~~Writes may start anywhere and span any extent~~
- ◆ Writes must be aligned to physical sector boundaries
- ◆ Writes may not span a zone boundary
- ◆ Writes to zones in Offline and Read Only states are not allowed

Read rules

- ◆ ~~Reads may start anywhere and span any extent~~
- ◆ ~~Unwritten logical blocks return initialization pattern~~
- ◆ Reads must be below the write pointer
- ◆ Reads may not span a zone boundary
- ◆ Reads to Offline zones are not allowed

➤ Open zones

- ◆ **Maximum Number Of Open Sequential Write Required Zones**
 - > In new ATA log or SCSI VPD page
 - > Specifies the largest number of zones that may be open

Host Managed SCSI Supported Commands Reference

➤ Mandatory

- ◆ INQUIRY
- ◆ LOG SENSE
- ◆ MODE SELECT (10)
- ◆ MODE SENSE (10)
- ◆ READ (16)
- ◆ READ CAPACITY (16)
- ◆ REPORT LUNS
- ◆ REPORT SUPPORTED OPCODES
- ◆ REPORT SUPPORTED TASK MGMT FUNCS
- ◆ REPORT ZONES
- ◆ REQUEST SENSE
- ◆ RESET WRITE POINTER
- ◆ START STOP UNIT
- ◆ SYNCHRONIZE CACHE (16)
- ◆ TEST UNIT READY
- ◆ WRITE (16)
- ◆ WRITE SAME (16)

➤ Optional

- ◆ ATA PASS-THROUGH (12)
- ◆ ATA PASS-THROUGH (16)
- ◆ FORMAT UNIT
- ◆ LOG SELECT
- ◆ PERSISTENT RESERVE IN
- ◆ PERSISTENT RESERVE OUT
- ◆ READ BUFFER
- ◆ READ DEFECT DATA (12)
- ◆ REPORT TIMESTAMP
- ◆ SET TIMESTAMP
- ◆ SANITIZE
- ◆ SECURITY PROTOCOL IN
- ◆ SECURITY PROTOCOL OUT
- ◆ SEND DIAGNOSTIC
- ◆ VERIFY (16)

Host Managed SCSI Not Supported Commands Reference

Compared to SBC-4 commands for direct access block devices

- ♦ ACCESS CONTROL IN
- ♦ ACCESS CONTROL OUT
- ♦ CHANGE ALIASES
- ♦ COMPARE AND WRITE
- ♦ EXTENDED COPY
- ♦ GET LBA STATUS
- ♦ MAINTENANCE IN
- ♦ MAINTENANCE OUT
- ♦ MODE SELECT (6)
- ♦ MODE SENSE (6)
- ♦ ORWRITE (16), (32)
- ♦ POPULATE TOKEN
- ♦ PREFETCH (10), (16)
- ♦ PREVENT ALLOW MEDIUM REMOVAL
- ♦ READ (10)*, (12), (32)
- ♦ READ ATTRIBUTE
- ♦ READ CAPACITY (10)*
- ♦ READ DEFECT DATA (10)
- ♦ READ LONG (10), (16)
- ♦ REASSIGN BLOCKS
- ♦ RECEIVE COPY RESULTS
- ♦ RECEIVE DIAGNOSTIC RESULTS
- ♦ RECEIVE ROD TOKEN INFORMATION
- ♦ REDUNDANCY GROUP IN
- ♦ REDUNDANCY GROUP OUT
- ♦ REMOVE I_T NEXUS
- ♦ REPORT ALIASES
- ♦ REPORT IDENTIFYING INFORMATION
- ♦ REPORT PRIORITY
- ♦ REPORT REFERRALS
- ♦ REPORT TARGET PORT GROUPS
- ♦ SET IDENTIFYING INFORMATION
- ♦ SET PRIORITY
- ♦ SET TARGET PORT GROUPS
- ♦ SPARE IN
- ♦ SPARE OUT
- ♦ SYNCHRONIZE CACHE (16)
- ♦ UNMAP
- ♦ VERIFY (10), (12), (32)
- ♦ VOLUME SET IN
- ♦ VOLUME SET OUT
- ♦ WRITE (10), (12), (32)
- ♦ WRITE AND VERIFY (10), (12), (16), (32)
- ♦ WRITE ATTRIBUTE
- ♦ WRITE BUFFER
- ♦ WRITE LONG (10), (16)
- ♦ WRITE SAME (10), (32)
- ♦ WRITE USING TOKEN
- ♦ XDWRITE READ (10), (32)
- ♦ XPWRITE (10), (32)

* Mandatory for direct access block devices

Agenda

- Introduction to Shingled Magnetic Recording
- SMR Implementation Options
- Drive Managed SMR Overview
- Host Supported SMR Standards
- Host Aware Zoned Block Device Overview
- Host Managed Zoned Block Device Overview
- Comparisons and Next Steps

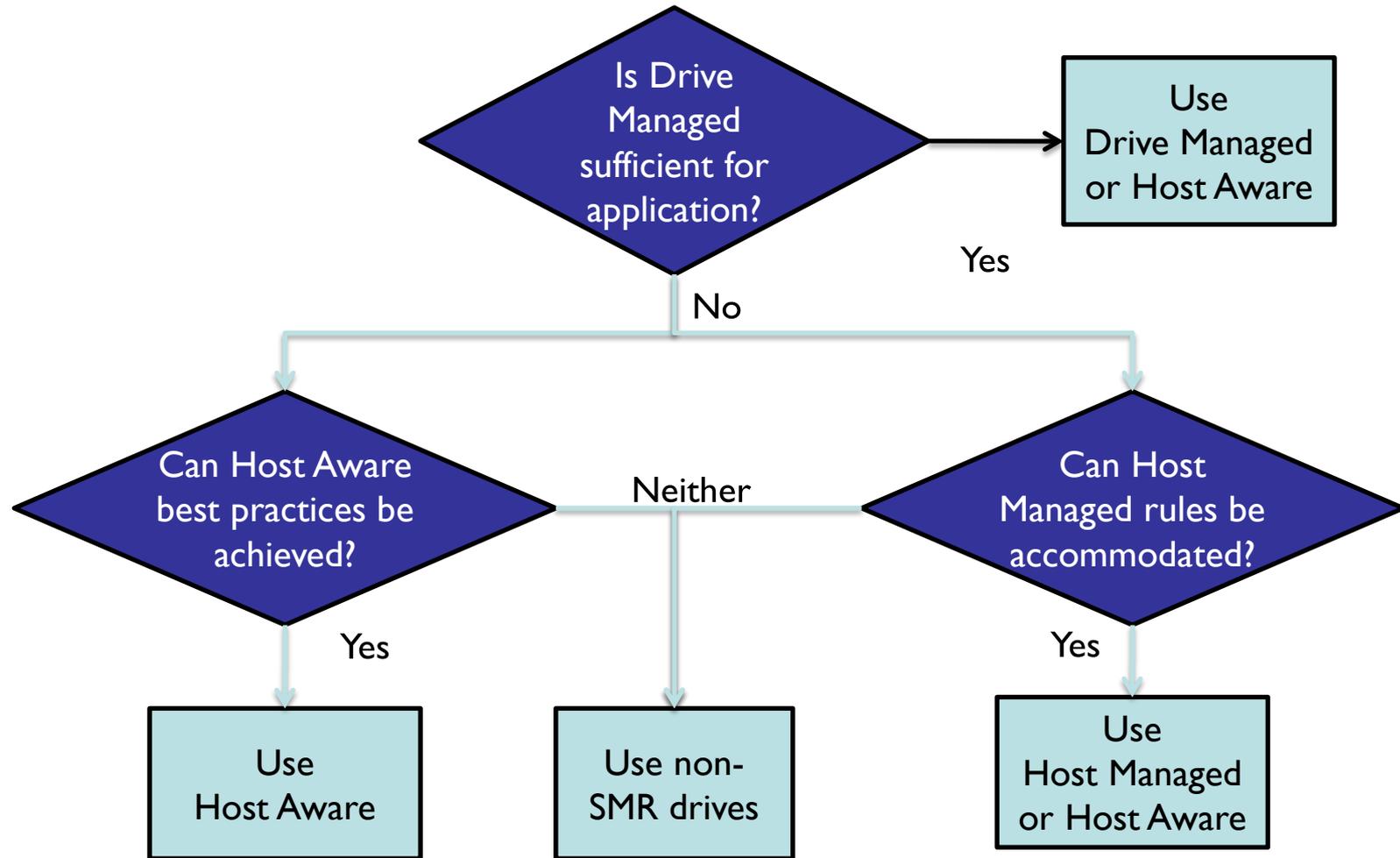
Comparison of Drive Managed, Host Aware and Host Managed

Capability	Drive Managed	Host Aware	Host Managed
SCSI Peripheral Device Type	00h: Direct Access Device	00h: Direct Access Device	14h: Host Managed Zoned Block Device
SCSI HAW_ZBC bit	0	1	0
ATA device signature	ATA	ATA	Host Managed Zoned
Command support REPORT ZONES RESET WRITE POINTER	SCSI: SBC-4 Disallowed Disallowed	SCSI: SBC-4 Mandatory Mandatory	SCSI: ZBC reduced set Mandatory Mandatory
Conventional zones	n/a	Optional	Optional
Sequential write preferred zones	n/a	Mandatory	Disallowed
Sequential write required zones	n/a	Disallowed	Mandatory

Comparison of Host Aware and Host Managed

- Performance on Host Managed workloads is the same
 - ◆ If a command sequence succeeds on HM then it succeeds on HA and with the same performance
- No fundamental cost difference
 - ◆ No mandated electronics for either type
- Host Aware is backwards compatible
 - ◆ Today's software runs successfully
 - ◆ A selected part of the software can be migrated
- Host Managed requires new software
 - ◆ Requires no non-sequential writing to sequential write required zones and other read and write rules

SMR Selection Guide



- Upgrade software stack components
 - ◆ Conform to rules and best practices
 - ◆ Applications
 - › Backup, archive, databases, DVR, surveillance, etc.
 - ◆ File systems and kernel components
 - › Space allocation
 - › Device mapper, block layer
 - › SMR Friendly File System – coming soon
 - ◆ Devices
 - › Host bus adapters, expanders, RAID controllers

The SNIA Education Committee thanks the following individuals for their contributions to this Tutorial.

Authorship History

Mary Dunn and Timothy Feldman, Seagate
September 16, 2014

Additional Contributors

none

Please send any questions or comments regarding this SNIA Tutorial to tracktutorials@snia.org