OPEN Compute Summit

The Future of Cheap & Deep

Philippe Vaillant, Seagate

Abstract

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

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

Comparisons and Next Steps

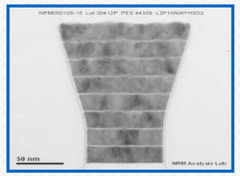
Conventional Perpendicular Writer

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

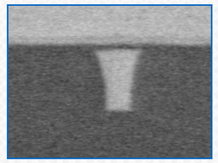
- Progressively higher TPI transforms write pole from well defined trapezoidal to triangular shape leading to ihcreasingly greater losses in writability beyond what was previously associated with writer width reduction.

Writer design is now close to geometry limitations that preclude further growth in track density without new innovation or new recording technology

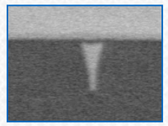
- Write pole SEM micrographs:



1st Generation
2006



2nd Generation
2008

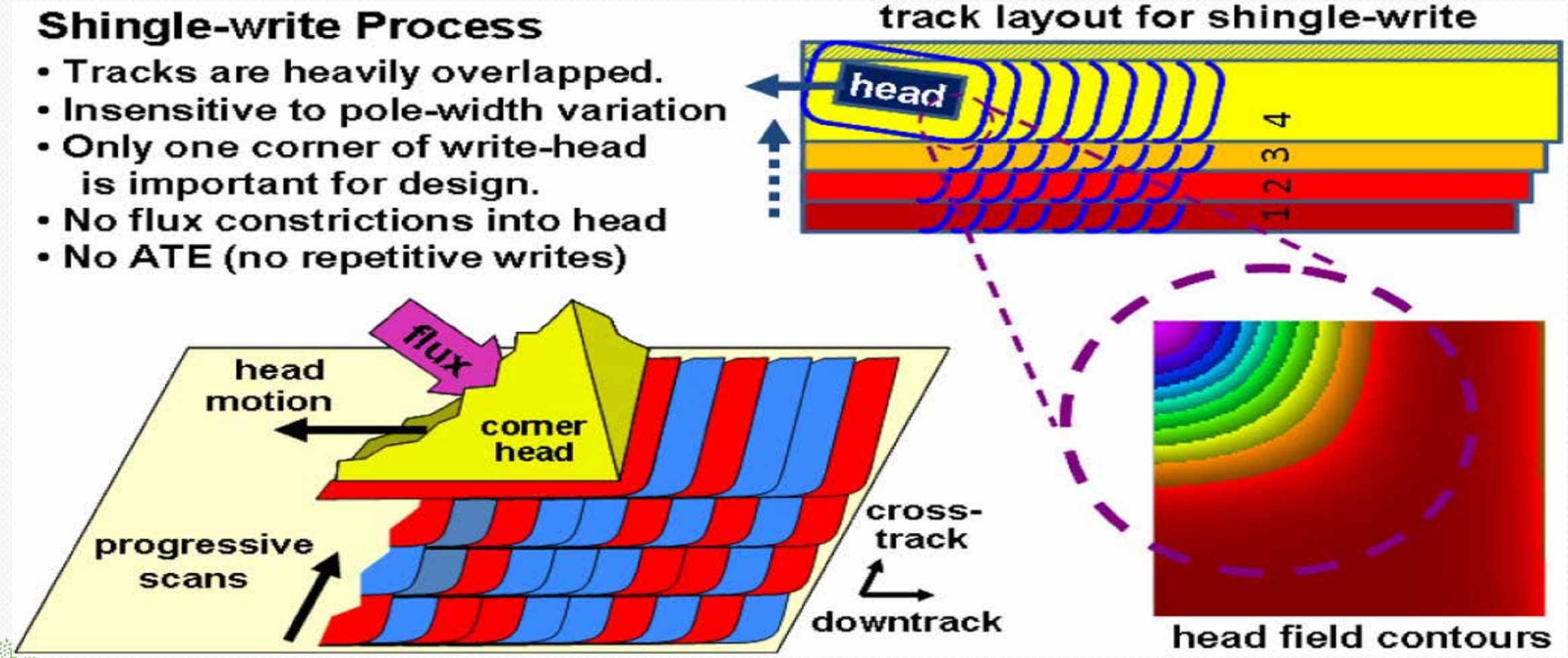


3rd Generation
2009

4th Generation
2010

5th Generation
2011

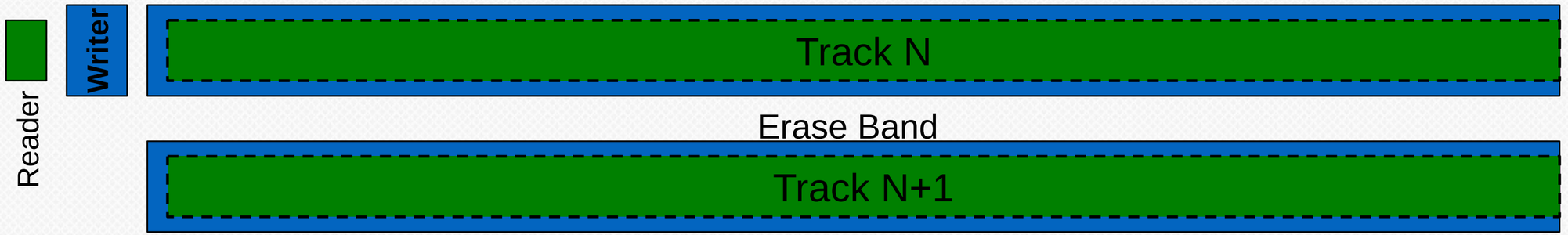
What is SMR?



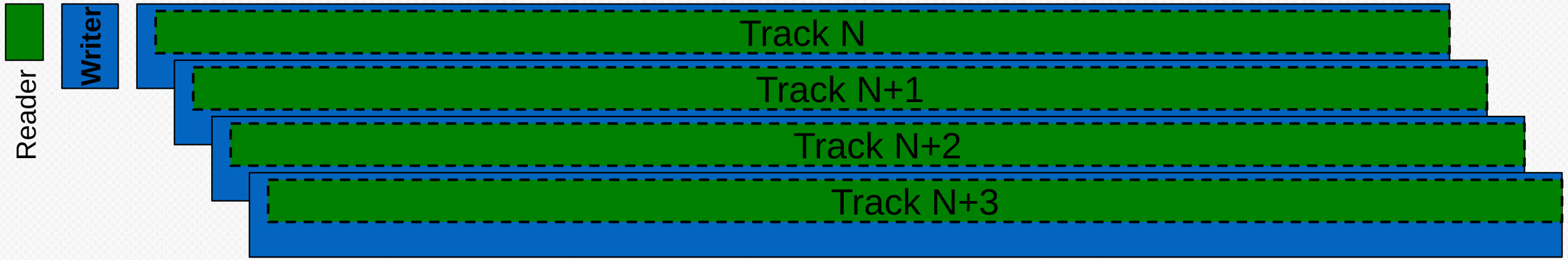
Wood, et al.: The Feasibility of Magnetic Recording at 10 Terabits Per Square Inch on Conventional Media, IEEE Transactions on Magnetics, Vol. 45, No. 2, February 2009

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

Comparisons and Next Steps

Overview of SMR Drive Types

Drive Managed

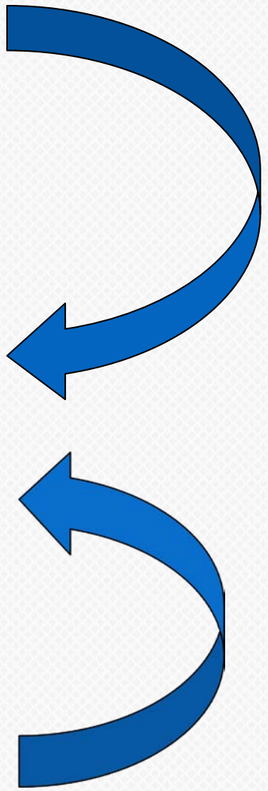
- Drive autonomously hides all SMR issues
- Workloads can affect performance

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



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* *	•Report Zones •Reset Write Pointer	None
Host Managed	14h: Host Managed Zoned Block Device	Host Managed Zoned	Sequential Write *Required* *	•Report Zones •Reset Write Pointer	•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

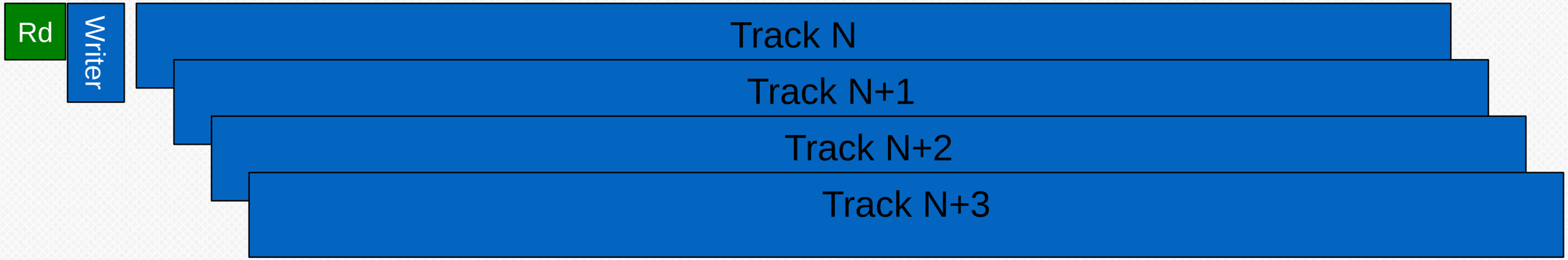
Drive Managed SMR Overview

Host Supported SMR Standards

Host Aware Zoned Block Device Overview

Comparisons and Next Steps

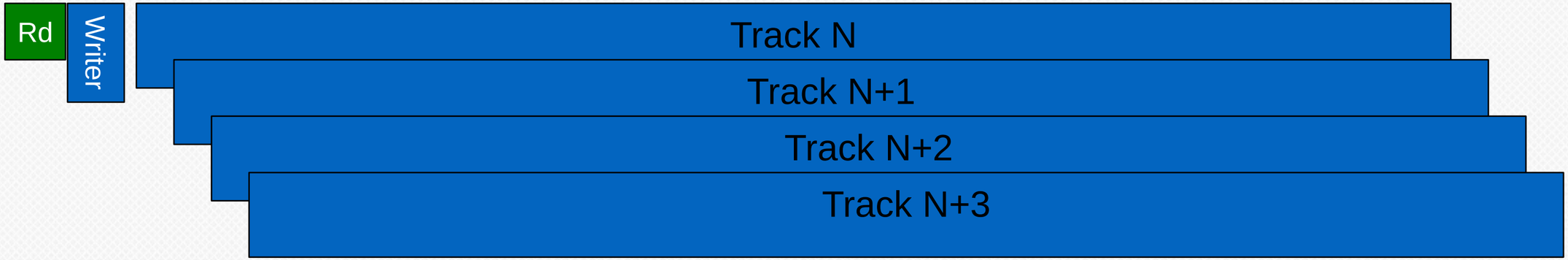
Updating a band with new data

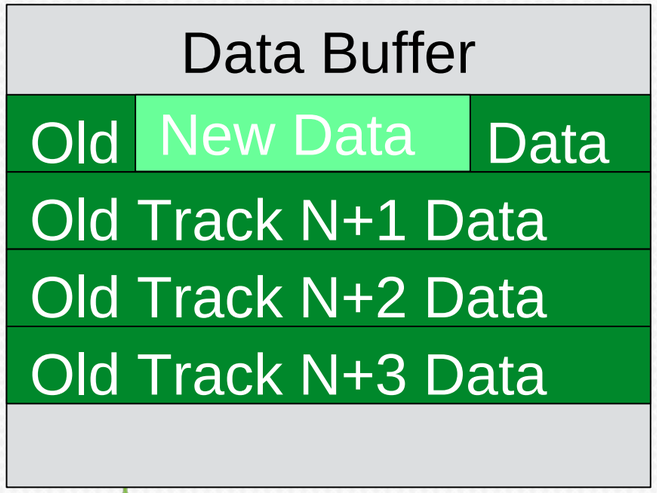


Data Buffer
Old Track N Data
Old Track N+1 Data
Old Track N+2 Data
Old Track N+3 Data

1. Read old data

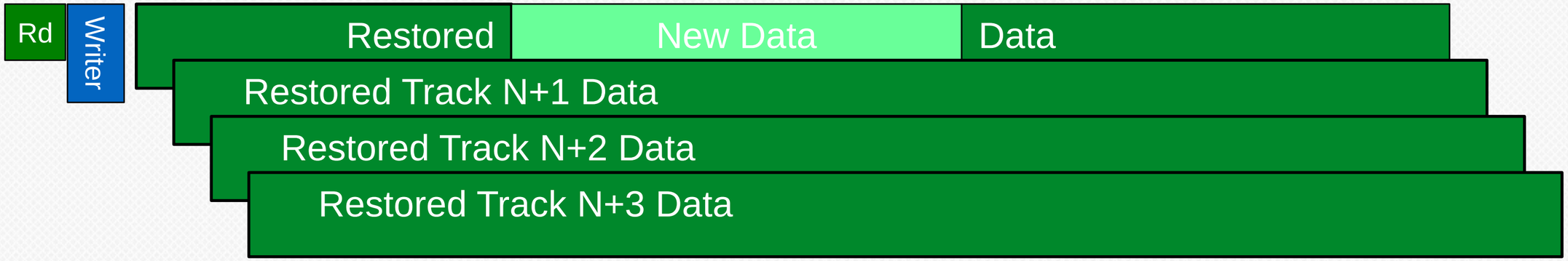
Updating a band with new data

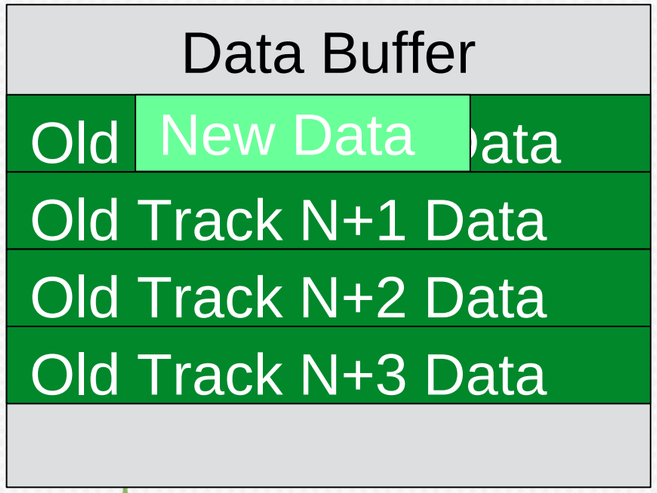




1. Read old data
2. Merge with new data

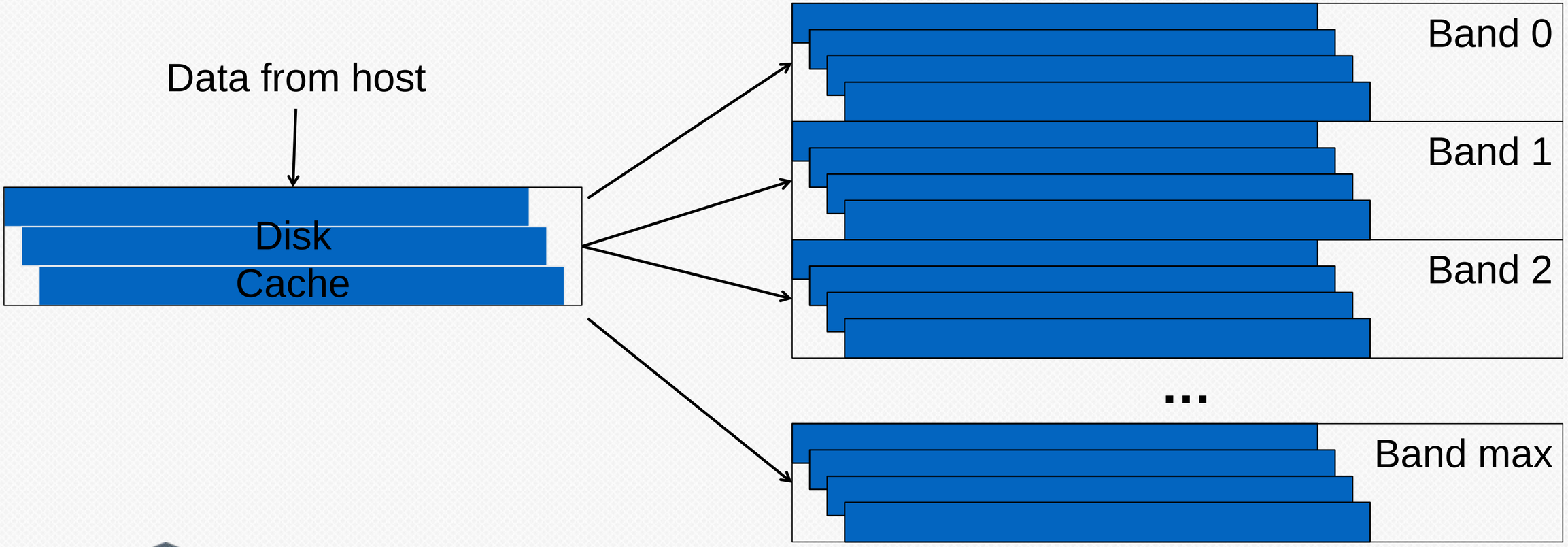
Updating a band with new data





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

Agenda

Introduction to Shingled Magnetic Recording

SMR Implementation Options

Drive Managed SMR Overview

Host Supported SMR Standards

Host Aware Zoned Block Device Overview

Comparisons and Next Steps

Host Support for SMR

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

Standard Activities

New models, feature sets, commands, logs, parameters

SCSI first, then ATA

- History has shown this is fastest

T10 (SCSI) Zoned Block Commands – ZBC

- Revision 1c 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
- March 2015 for Version 1

T13 (ATA) Zoned block ATA Commands – ZAC

- Fast follower to ZBC

Agenda

Introduction to Shingled Magnetic Recording

SMR Implementation Options

Drive Managed SMR Overview

Host Supported SMR Standards

Host Aware Zoned Block Device Overview

Comparisons and Next Steps

Host Aware Overview

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 of 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

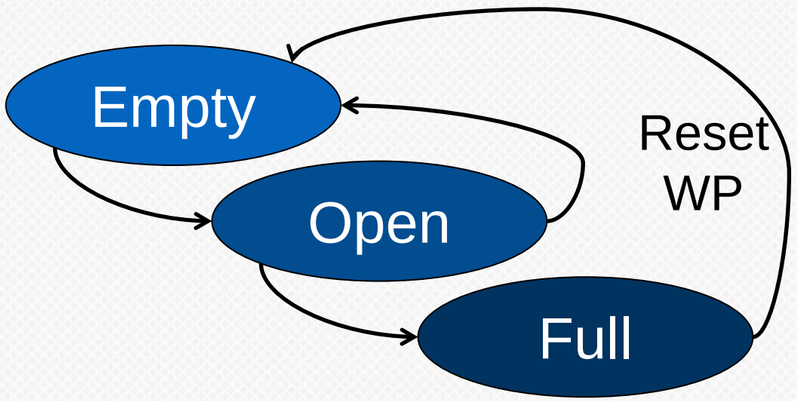
Writes at the write pointer have conventional performance

- 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

Agenda

Introduction to Shingled Magnetic Recording

SMR Implementation Options

Drive Managed SMR Overview

Host Supported SMR Standards

Host Aware Zoned Block Device Overview

Comparisons and Next Steps

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

Enabling Technologies

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 Future of Cheap & Deep

The capacity will come from SMR.

Some storage stacks will happily use Drive Managed with little or no change to the stack.

- Example 1: Archive with write-once filling of drives,low performance requirements and very low TCO targets.
- Example 2: Client markets with legacy software.

Other stacks will make use of upgraded file systems that

- take advantage of the flexibility of Host Aware for fast metadata updates and minimal garbage collection, and
- take advantage of the controls of Host Aware to deliver performance.

- Example 1: Active Archive with higher transaction rates than Archive
- Example 2: Large data analytics using copy-on-write file systems (btrfs)
- Example 3: Content Distribution using SMR-friendly file systems (next gen ext4)

For further information Contact:

Philippe Vaillant : philippe.vaillant@seagate.com

Josh Tinker : joshua.b.tinker@seagate.com

OPEN
Compute Summit

Thank you!