OPEN Compute Summit

March 10–11, 2015
San Jose

OPEN Compute Summit

SMR

The Future of Cheap and Deep

Paul Renden
Seagate Technology
Staff Technical Customer Manager, Enterprise

Why Shingled Magnetic Recording?

Large capacity gains are increasingly requiring advanced technologies

- SMR
- HAMR
- BPM

SMR is currently the most cost-effective technology to meet these required capacity gains for the next several years

- Currently, SMR provides ~25% capacity gain with no other required changes

Lowest $/GB in Enterprise

Copyright © Seagate Technology, March 2015

Agenda

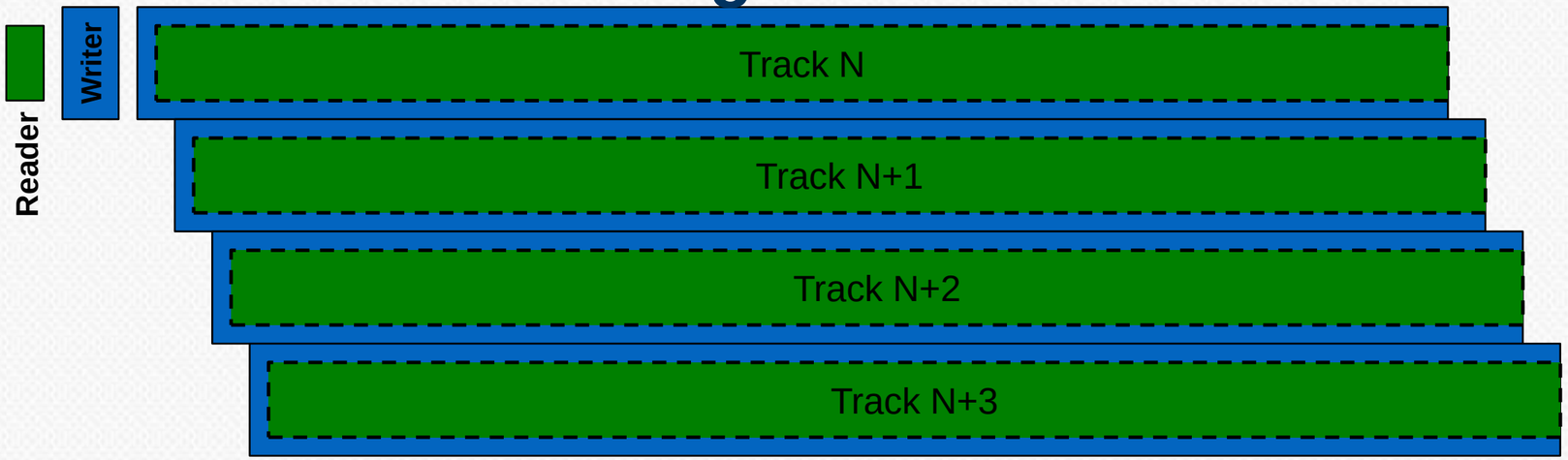
- Very quick basic refresh of SMR
- Drive Managed SMR
- Host Aware SMR
- Moving Forward with SMR

Copyright © Seagate Technology, March 2015

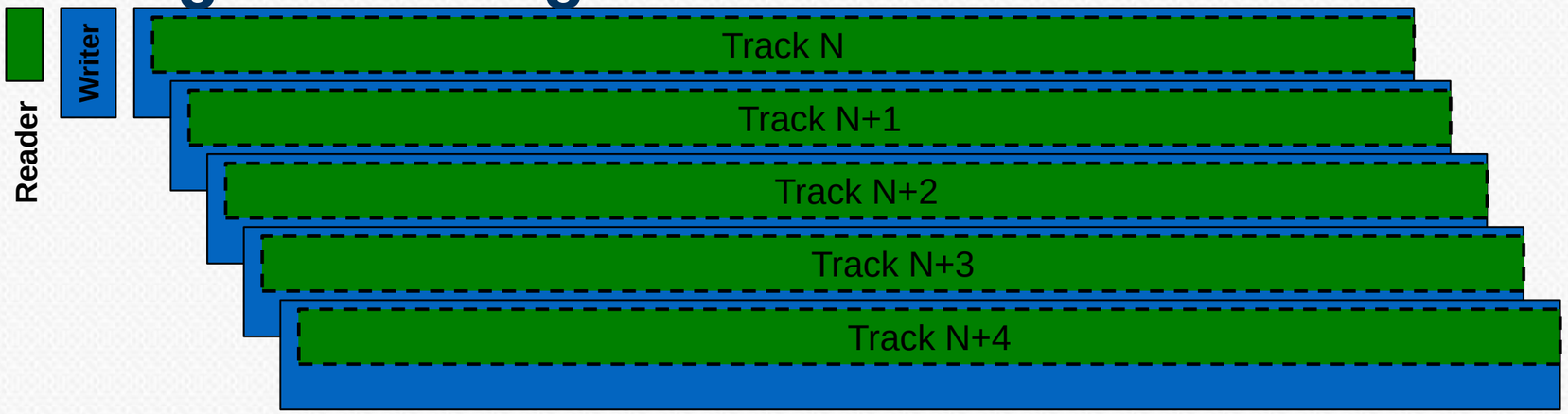
SMR Basics

Conventional versus SMR Writing

Conventional Writing



Shingled Writing

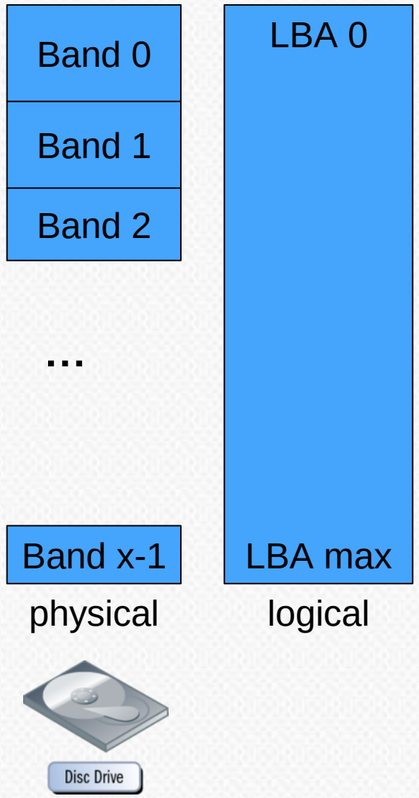


Copyright © Seagate Technology, March 2015

Drive Managed SMR

A disk as a set of bands

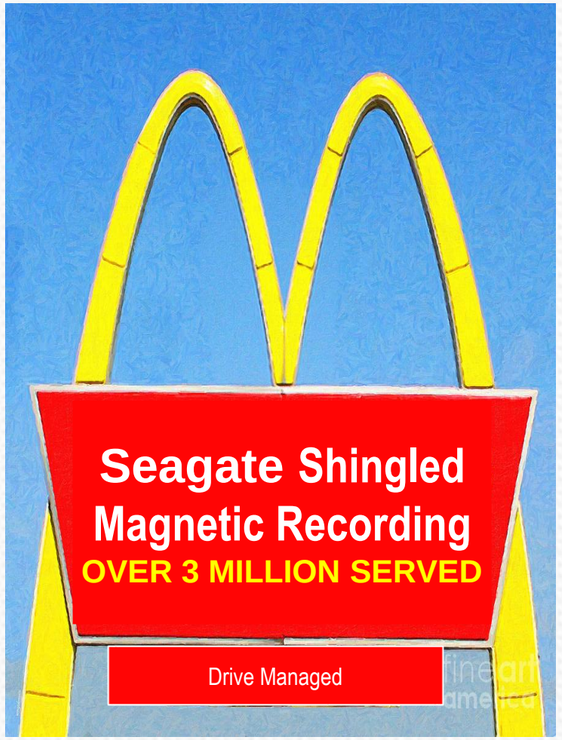
- SMR Bands
  - Physical Construct
  - Boundaries not known outside of drive



Copyright © Seagate Technology, March 2015

Drive Managed SMR

- The first SMR drive type
- Makes architectural details of SMR transparent to the host.
- Backward compatible



Copyright © Seagate Technology, March 2015

Drive Managed SMR

- Benefits
  - No host changes required
  - Extremely effective in Personal Compute and in some Enterprise applications (Ex: Archive)
  - Reads (seq, rndm) typically achieve conventional performance
  - Large write-back disk cache
  - Write-around for sequential writes = conventional performance

Copyright © Seagate Technology, March 2015

Drive Managed SMR

- Challenges
  - Disk cache is a limited resource
    - Full cache can degrade random write performance
    - Larger cache has areal density cost
    - Read performance impact if reads are fragmented
  - Cache Cleaning is complex
    - Potentially large command latency tails
  - Write-around is limited
    - Sequential detection is non-trivial
      - Multiple streams versus random
      - Long inter-command time versus end of stream
      - Multiple tracks needed: write one track after the next track is buffered or queued

Copyright © Seagate Technology, March 2015

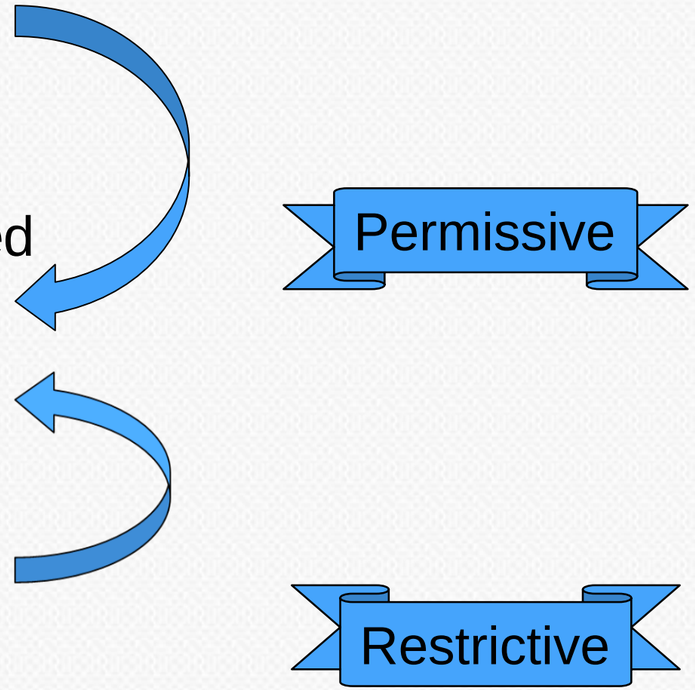
Drive Managed SMR

- Best practices to achieve optimal performance
  - Idle time allows cache to clear and minimize fragmentation conditions
  - Truly Sequential writes
  - Writes with high transfer length and queue depths
    - Maintains sequential streaming mode
  - Limit and concentrate random writes

Copyright © Seagate Technology, March 2015

SMR Drive Types

- Drive Managed
  - Drive autonomously manages all SMR operations
  - Backward compatible
- Host Aware
  - Superset of Drive Managed and Host Managed
  - Backward compatible
  - Extensions to ATA and SCSI command sets
- Host Managed
  - Extensions to ATA and SCSI command sets
  - Error conditions for some reads and writes
  - Not backward compatible
  - New device type



Copyright © Seagate Technology, March 2015

Host Aware SMR

Host Aware SMR Solution

Goals

- Performance parity vs conventional disks
  - Broadened use cases over Drive Managed
  - Trivial sequential detection
- Minimal interface changes
  - A few new commands and parameters
  - No changes to Read and Write commands
- Enable more markets
- Grow beyond Archive

Copyright © Seagate Technology, March 2015

Host Aware SMR Solution

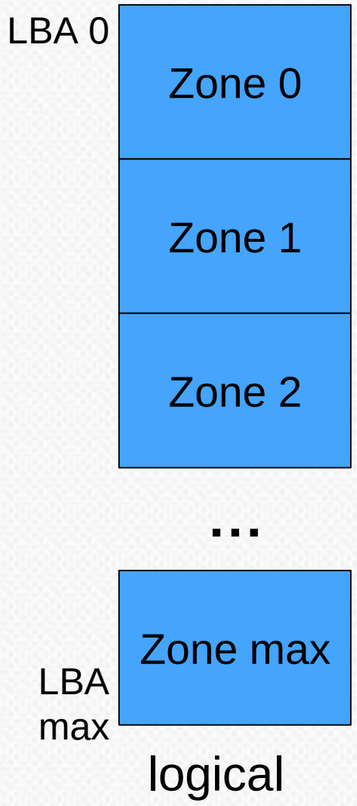
Achieving the goals

Goals

- Parity with conventional disks
- Minimal interface changes
- Enable more markets
- Grow beyond Archive

Methodology

- Zones
  - Logical address ranges exposed to host
- Write Pointers
  - Location of sequential writing
- Host controls zone usage
  - Tell drive what sectors are not in use, “unwritten”
- Leverage device capabilities
  - Number of active sequential streams at full performance
  - Amount of random write space at full performance

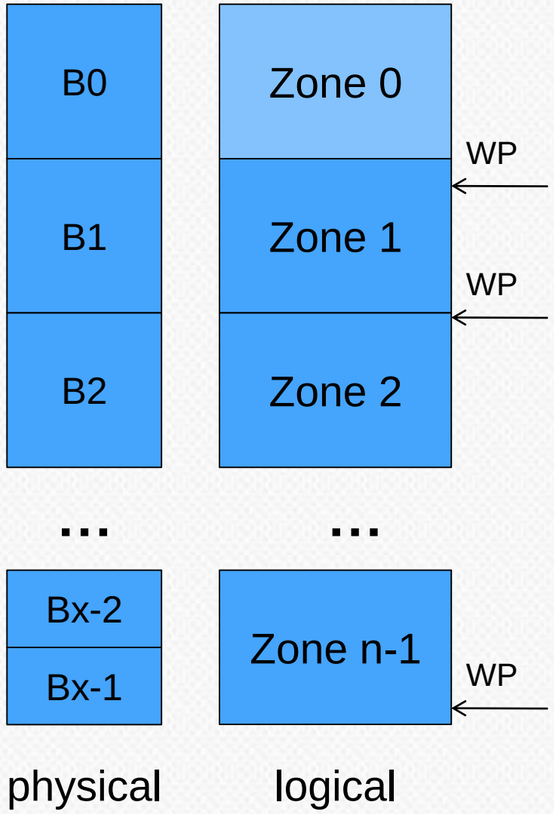


Copyright © Seagate Technology, March 2015

Host Aware SMR Solution

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
       - *"Sequential Write Preferred"*
       - Each zone has its own Write Pointer
       - Each zone has its own state



Copyright © Seagate Technology, March 2015

Host Aware SMR Solution

Write Pointer Zones

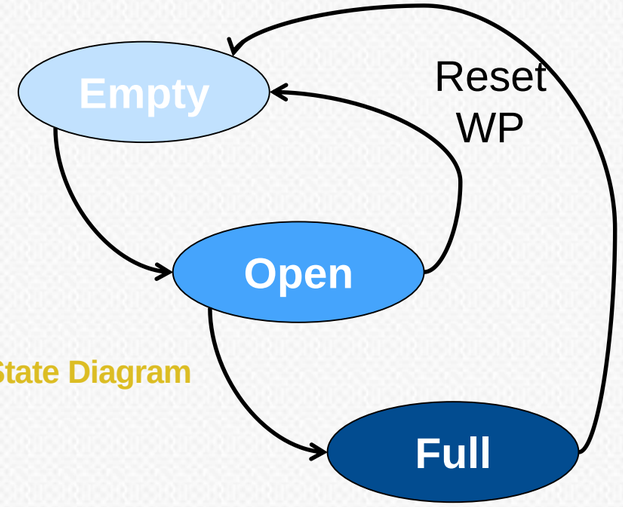
Writes at the write pointer have conventional performance

- 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



Simplified State Diagram

Empty

- Write pointer is at start of zone

Open

- Write pointer is mid-zone

Full

- No write pointer value

Copyright © Seagate Technology, March 2015

Host Aware SMR Solution

Intended Usage Model

1. Use **REPORT ZONES** and parameters to determine configuration

2. (*Judiciously*) Assign random write zones as needed
   - Limit to the device's Random Zones capabilities
   - Don't care about Write Pointer values
   - Don't issue Reset Write Pointer
   - Not preconfigured from factory
   - Can change during operation

3. Use the rest of the zones for sequential writing
   - Write Pointer is implicitly known

4. Control the number of Open zones
   - Limit to the device's Open zones capability

5. Garbage collect to evacuate zones for re-use
   i. Copy non-stale data to an open zone
   ii. Issue Reset Write Pointer
   iii. Move zone to free pool

Copyright © Seagate Technology, March 2015

Moving Forward with SMR

Host Aware SMR Solution

Considerations for adopting Host Aware

- Do you understand your Workload (from a drive's perspective)?
  - How close to "best practices"?
- What SW development resources are available to you?
  - App modifications
  - File System optimizations
  - Implementing ZAC/ZBC features, commands, HA intended usage model
- Familiarize yourself with ZAC/ZBC via T10, T13 website content / specs
  - ZBC rev. 2 is in letter ballot review
  - ZAC is in spec development
- Obtain drive samples, initiate a POC

Copyright © Seagate Technology, March 2015

Resources

t10.org	ZBC: SCSI Zoned Block Commands letter ballot review, now
t13.org	ZAC: Zoned-device ATA Commands spec development, now
github.com/hgst/libzbc	Linux user space libraries for ZBC emulation
github.com/Seagate/SMR_FS-EXT4	SMR Friendly Filesystem upgrades to ext4 See announcement at Linux Vault Conference March 11-12, 2015, Boston

Copyright © Seagate Technology, March 2015

Summary

- The most cost-effective near-term capacity gains require SMR
- Some applications run well on Drive Managed
  - “Don’t know until you try it”
- Host Aware broadens SMR’s appeal to a wider range of applications
  - Leveraging the intended usage model
- Understand your environment and look for opportunities to adopt Host Aware SMR

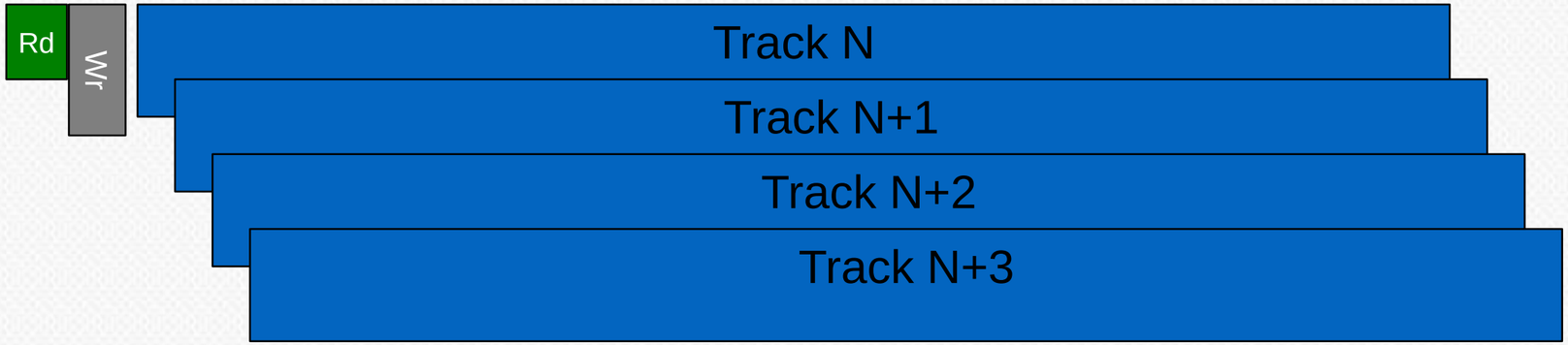
Copyright © Seagate Technology, March 2015

THANK YOU!

BACKUP

SMR Basics

Updating a band with new data



Data Buffer

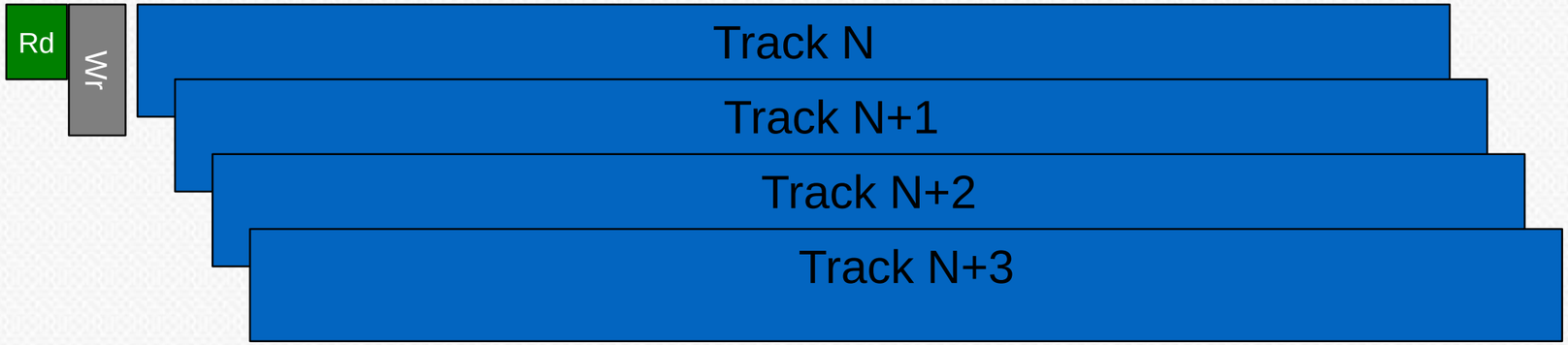
Old Track N Data
Old Track N+1 Data
Old Track N+2 Data
Old Track N+3 Data

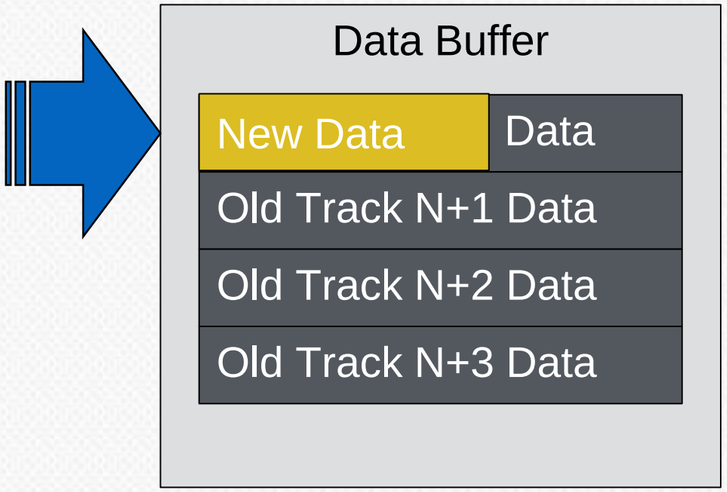
1. Read old data

Copyright © Seagate Technology, March 2015

SMR Basics

Updating a band with new data



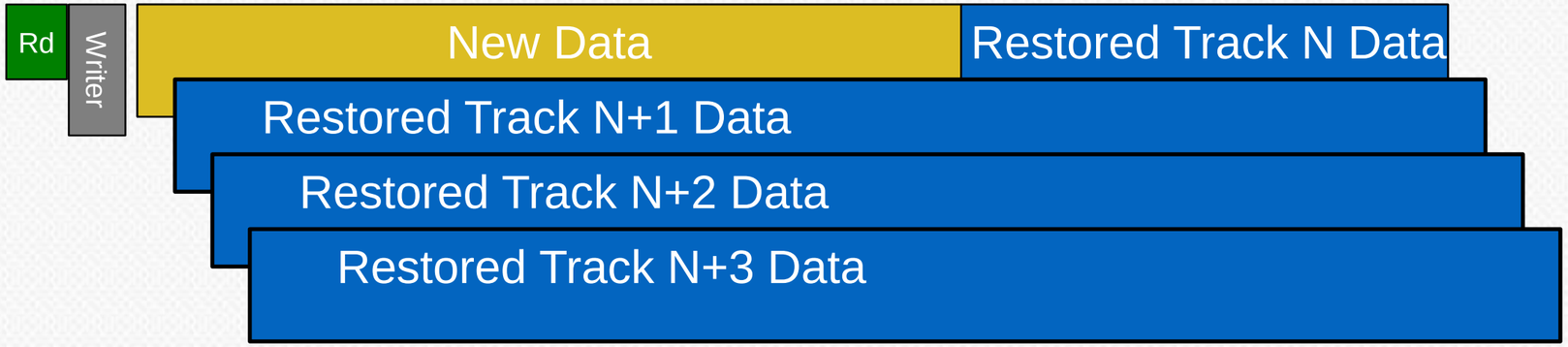


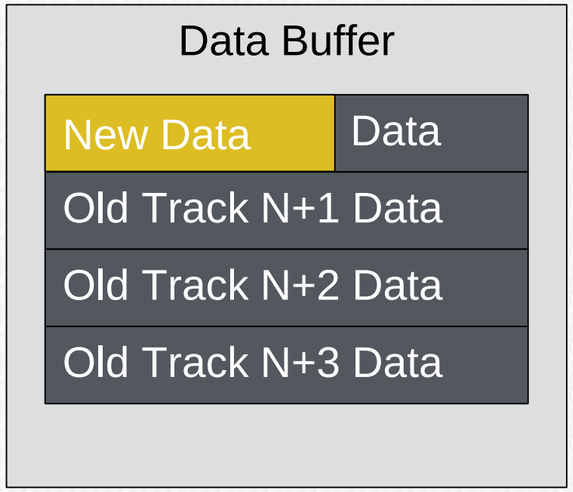
1. Read old data
2. Merge with new data

Copyright © Seagate Technology, March 2015

SMR Basics

Updating a band with new data



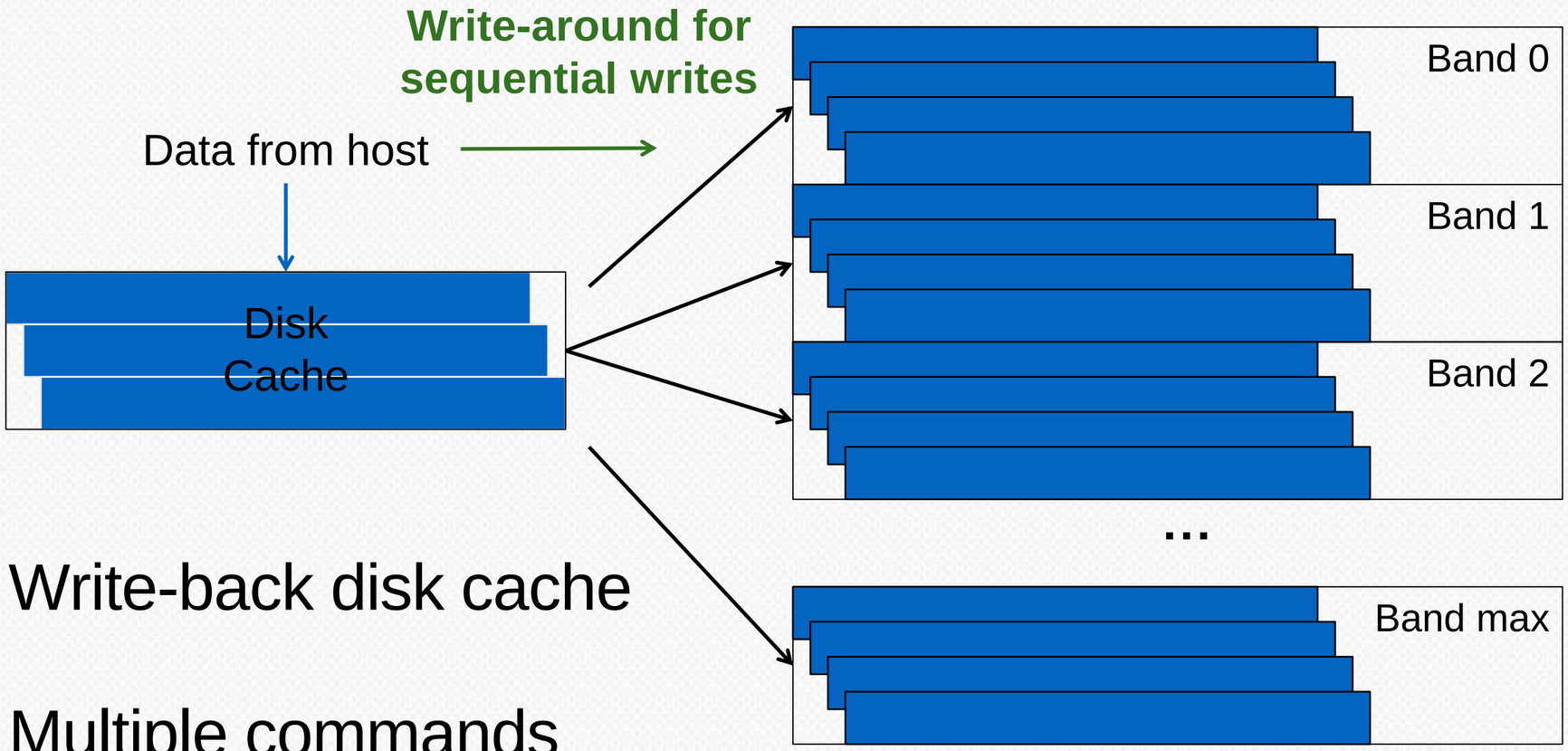


1. Read old data
2. Merge with new data
3. Write new data, refreshing old data

Copyright © Seagate Technology, March 2015

SMR Basics

Improving Write Performance



Write-back disk cache

Multiple commands cleaned during a band update

Copyright © Seagate Technology, March 2015