Zoned Block Device Overview

July 2014

Timothy Feldman
Seagate

Non-Confidential

Zoned Block Devices

ZBC revision 0.1a

2 standards bodies
- T10 (SCSI): Zoned Block Commands (ZBC) www.t10.org
- T13 (ATA): Zoned-device ATA Commands (ZAC) www.t13.org

2 device type models
- Host Aware
- Host Managed

1 Peripheral Device Type and 1 type flag
- PDT = 14h: Host Managed Zoned Block Device
- HAW_ZBC: Host Aware Zoned Block Device

3 zone types
- Conventional zones
- Sequential write preferred zones
- Sequential write required zones

2 commands
- Report Zones
- Reset Write Pointer

How are all of these put together?

Zone Types

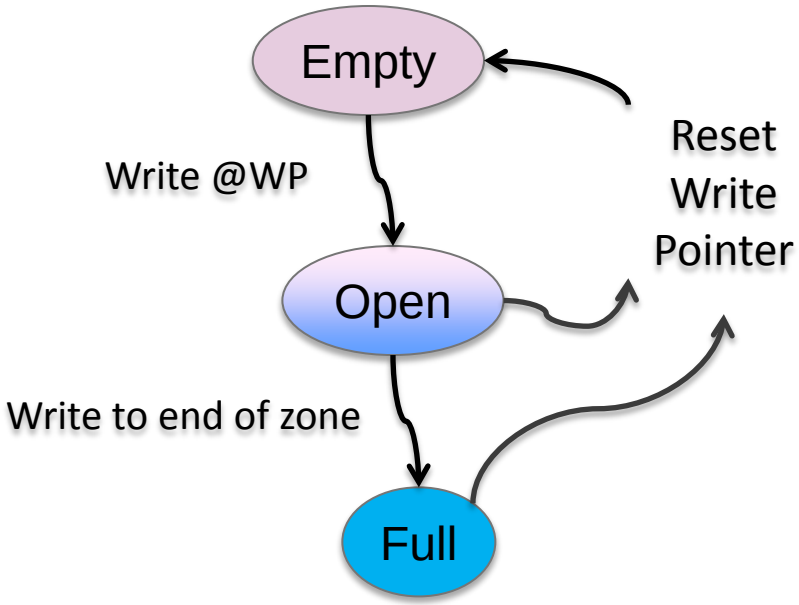
ZBC Overview

A drive's LBA space is separated into abutting zones

1. Conventional zone
   - » Drive autonomously manages all LBAs
   - » No write pointer
2. Sequential Write Preferred zone
   - Each zone has a write pointer to indicate the **preferred** write location
   - » Reset Write Pointer specifies
     - old data **changes to fill data (zeros)**
     - write pointer goes back to the start of the zone
   - Sequential writes after RWP have conventional performance
3. Sequential Write Required zone
   - Each zone has a write pointer to indicate the **required** write location
   - » Reset Write Pointer specifies
     - old data **is unreadable**
     - write pointer goes back to the start of the zone
   - Sequential writes after RWP have conventional performance

Write Pointer Zone State Machine (simplified)

- Each zone has its own Write Pointer
- Writes at the Write Pointer automatically advance the pointer
- Issue Reset Write Pointer to zone before re-writing



- Empty
  - Write pointer is at start of zone
- Open
  - Write pointer is mid-zone
- Full
  - Write pointer is not valid
- Reset Write Pointer
  - Go back to Empty

New Commands

ZBC Overview

- Report Zones command
  - Reports zone configuration for each zone
    - » Type, Condition, Size, Start LBA, Write Pointer
  - No method to change the configuration in the field
- Reset Write Pointer command
  - Resets the write pointer of a zone to the first LBA of that zone
    - » RESET ALL bit specifies that all zones are to have their write pointer reset

Device Types

ZBC and SMR Overview

1. Drive Managed
   - Direct Access Device without ZBC extensions
   - PDT = Direct Access Device 00h
     - » HAW_ZBC = 0
2. Host Aware
   - PDT = **Direct Access Device 00h**
     - » HAW_ZBC = 1
   - Collection of sequential write **preferred** zones
     - » Conventional zone(s) are optional
   - **Full** SCSI Block Commands (SBC) support
3. Host Managed
   - PDT = **Host Managed Zoned Block Device 14h**
   - Collection of sequential write **required** zones
     - » Conventional zone(s) are optional
   - **Partial** SCSI Block Commands (SBC) support
   - **Many other restrictions**

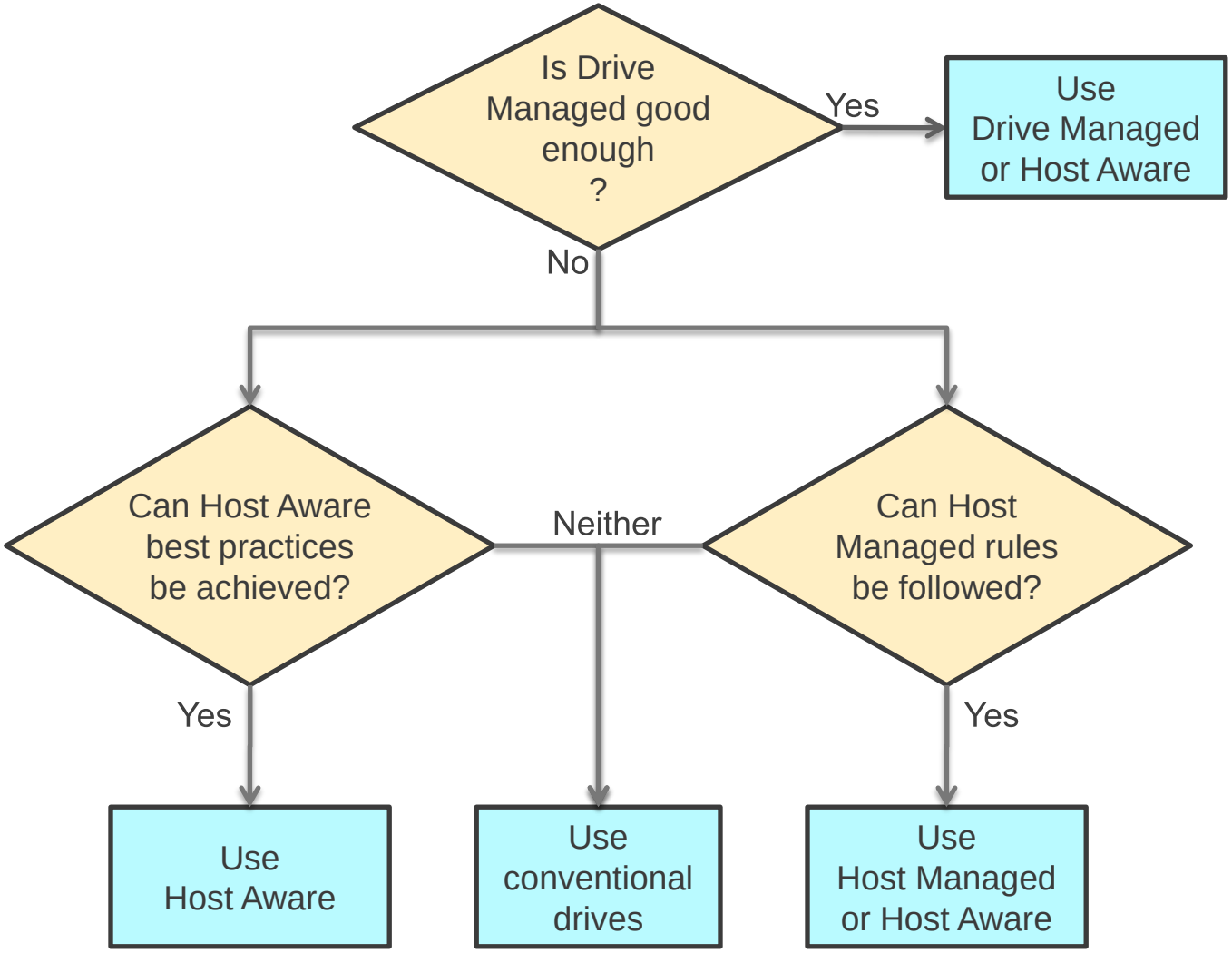
Device Types

ZBC and SMR Overview

Style	Zone Types	Write Rules	Read Rules	Best Practices
Drive Managed	none	Write any LBA in any order	Read any LBA in any order	• Maximize spatial density • Separate write bursts with idle
Host Aware	Sequential Write Preferred and Conventional*	Write any LBA in any order	Read any LBA in any order	• Write at Write Pointer • Minimize random write zone count • Minimize open zone count
Host Managed	Sequential Write Required and Conventional*	• Write only at Write Pointer • Do not write part of a physical sector • Do not span zone boundaries	• Read only below Write Pointer • Do not span zone boundaries	

*optional

SMR Drive Selection



ZBC Proposals In Progress

Current T10 Committee Activity www.t10.org

- Zone state machine
- New parameters
  - Max Open Zones
  - Allow fewer read and write restrictions for Sequential Write Required zones
  - Restrict writes to Conventional zones to be aligned to physical sector boundaries
- Command modifications
  - List of zones for Reset Write Pointer
- New commands
  - Open, Close and Finish

    or
  - Activate and Peg

Q & A

Additional slides

Comparison of Host Aware and Host Managed

Interface specification differences

Capability	Host Aware	Host Managed
Peripheral Device Type	00h (Direct Access Device)	14h (Host Managed Zoned Block Device)
HAW_ZBC bit	1	0
Command support Report Zones Reset Write Pointer SBC commands	SBC-4 Mandatory Mandatory Full support	ZBC reduced set Mandatory Mandatory Partial support
Conventional zones	Optional	Optional
Sequential write preferred zones	Mandatory	Not supported
Sequential write required zones	Not supported	Mandatory

Host Managed Zoned Block Device

Command Overview

- Host Aware commands and parameters
  - Reduced mandatory and optional list

Mandatory	Optional
Inquiry	ATA Pass-Through (12)
Log Sense	ATA Pass-Through (16)
Mode Select (10)	Format Unit
Mode Sense (10)	Log Select
Read (16)	Persistent Reserve In
Read Capacity (16)	Persistent Reserve Out
Report LUNs	Read Buffer
Report Supported Opcodes	Read Defect Data (12)
Report Supported Task Mgmt Funcs	Report Timestamp
Report Zones	Set Timestamp
Request Sense	Sanitize
Reset Write Pointer	Security Protocol In
Start Stop Unit	Security Protocol Out
Synchronize Cache (16)	Send Diagnostic
Test Unit Ready	Verify (16)
Write (16)	Write Buffer
Write Same (16)	

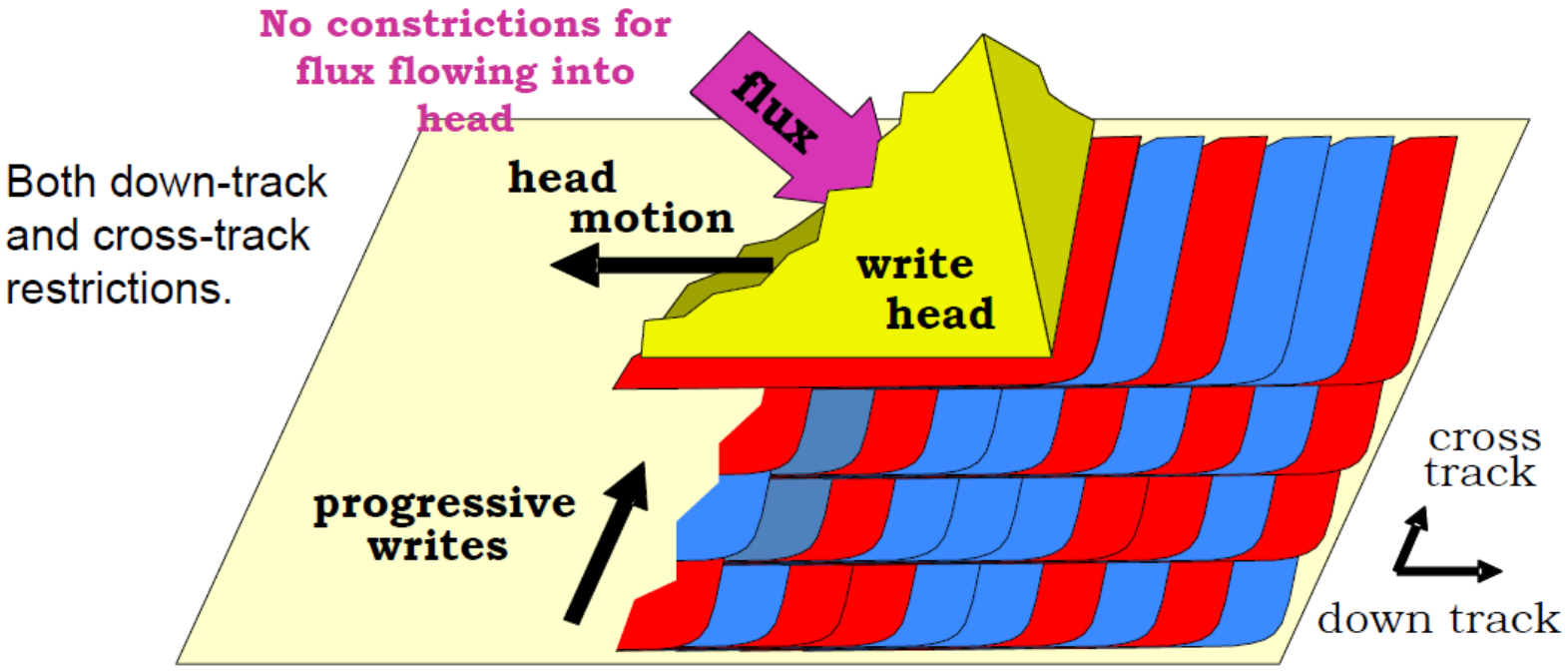
Comparison of Host Aware and Host Managed

ZBC Overview

- **Performance**
  - Performance of HM-compliant workloads is the same
    - If a command sequence succeeds on Host Managed then it succeeds on Host Aware and with the same performance
- **Cost**
  - No mandatory hardware difference
- **Compatibility**
  - Host Aware is backwards compatible
    - Today's software runs successfully
    - A selected part of the application and stack can be migrated
  - Host Managed requires new software
    - Guaranteeing no non-sequential writing and all the other read and write rules is challenging

What is SMR?

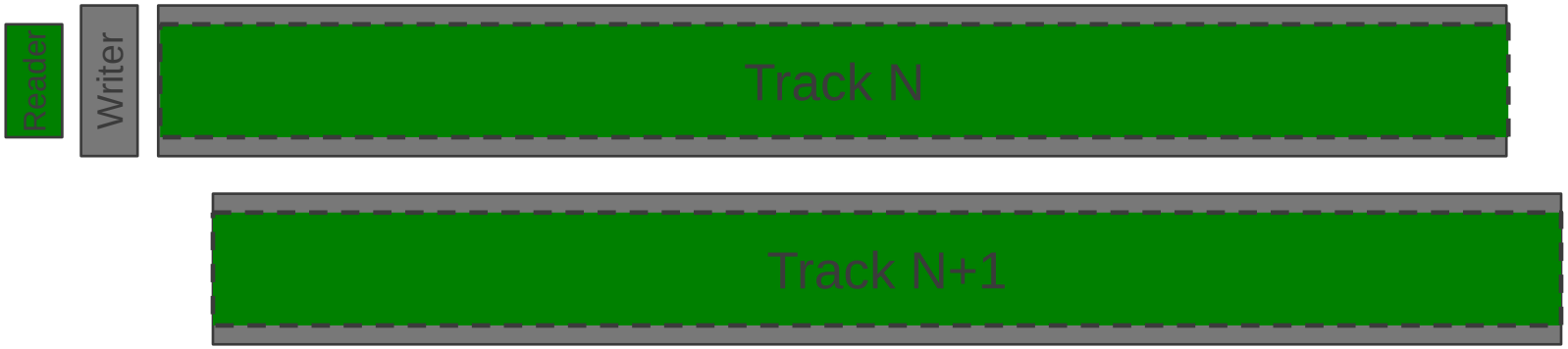
SMR write head geometry extends well beyond the track pitch



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

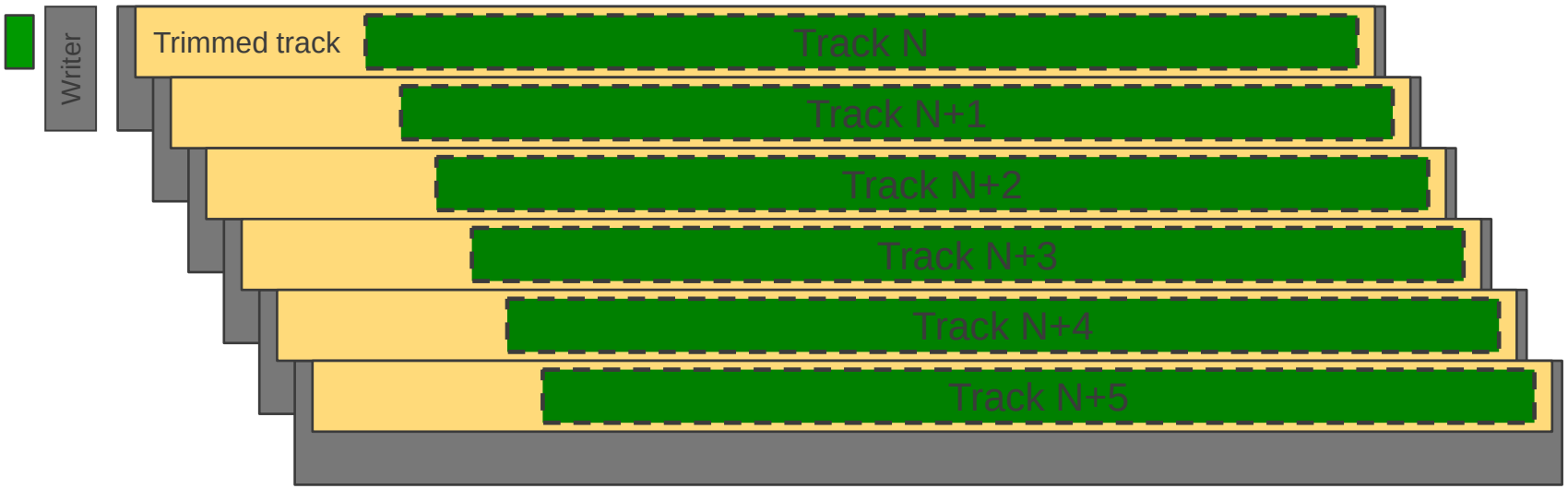
The larger SMR write head introduces the SMR Constraint

Conventional Writes



SMR Writes

- SMR reader is much narrower than writer



Host Aware Zoned Block Device

Best Practices

Writes:

- Sequential: **Maximize long sequential write runs**
  - Reset Write Pointer before zone re-use
    - » Eliminate unnecessary internal re-writes of unwanted data
  - Align to 4-KiB
    - » Align writes to the drive's reported physical sectors
  - Start and end at band boundaries (i.e.: 256 MiB)
  - Limit number of open zones
    - » Benefit: sequential writes will be performant if the number of concurrent threads does not exceed the advisory limit
- Random: **Limit and concentrate random writes**
  - Pick a few zones
    - » Benefit: random writes will be high performance if the number of zones conforms to the advisory limit

Other enabling technologies

Call to Action

Conform to rules and best practices

1. File systems and the storage stack
   - Host Bus Adapters, Expanders, RAID controllers
   - Device drivers, file systems and stack
     - » Open source: ext4, xfs, device mapper, etc.
     - » Closed source stacks
2. Applications
   - Backup and archive
   - Databases, log-structured applications
   - Content repositories, DVR, surveillance