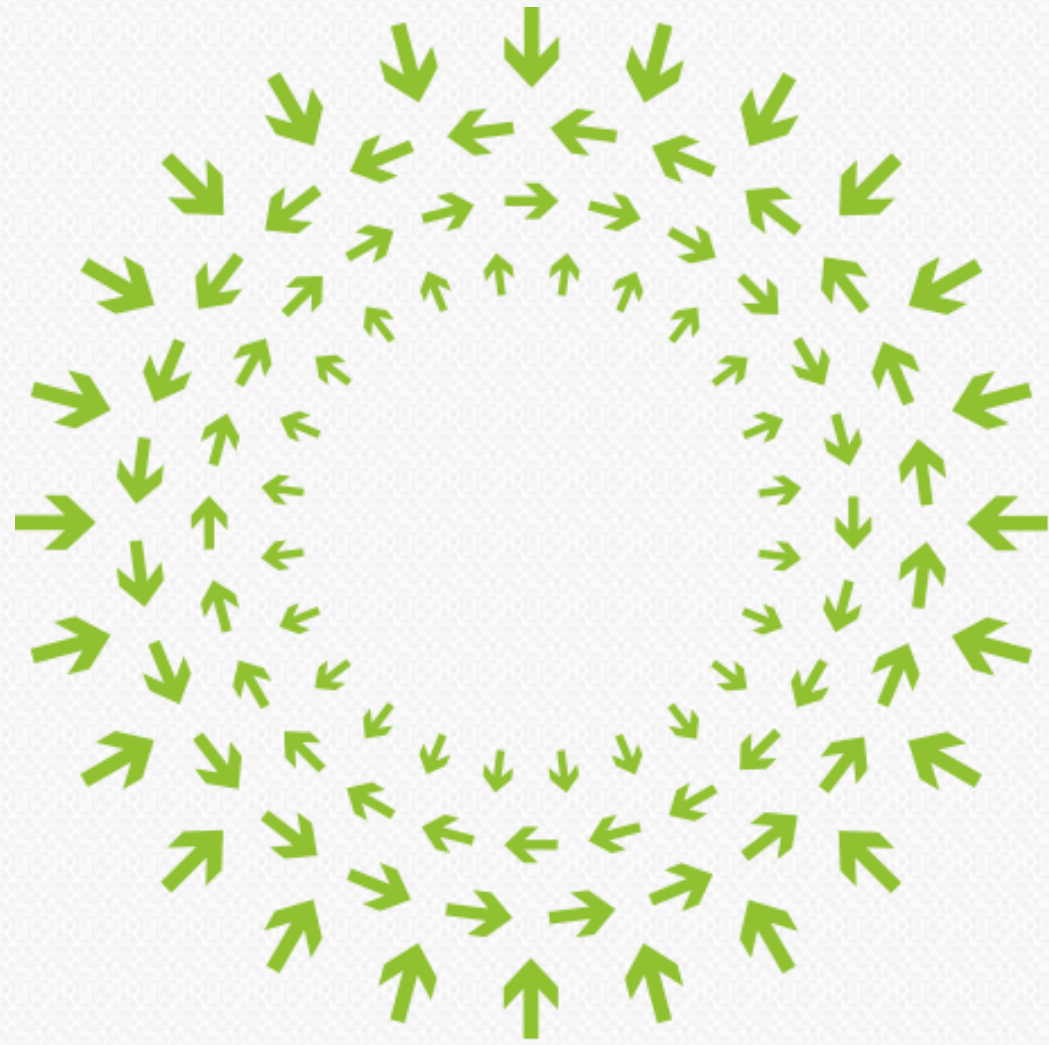




OPEN

Compute Summit
Engineering Workshop
October 30-31, 2014
Paris





SMR Directions:

Data Center Applications

Jim Malina

Western Digital

Technologist - System Architecture



Outline

WD Approach for SMR for DC

- What is Host Managed SMR
- Differences with Host Aware
- Costs of Host Aware / Benefits of Host Managed
- DC Markets / Host Managed Enablement

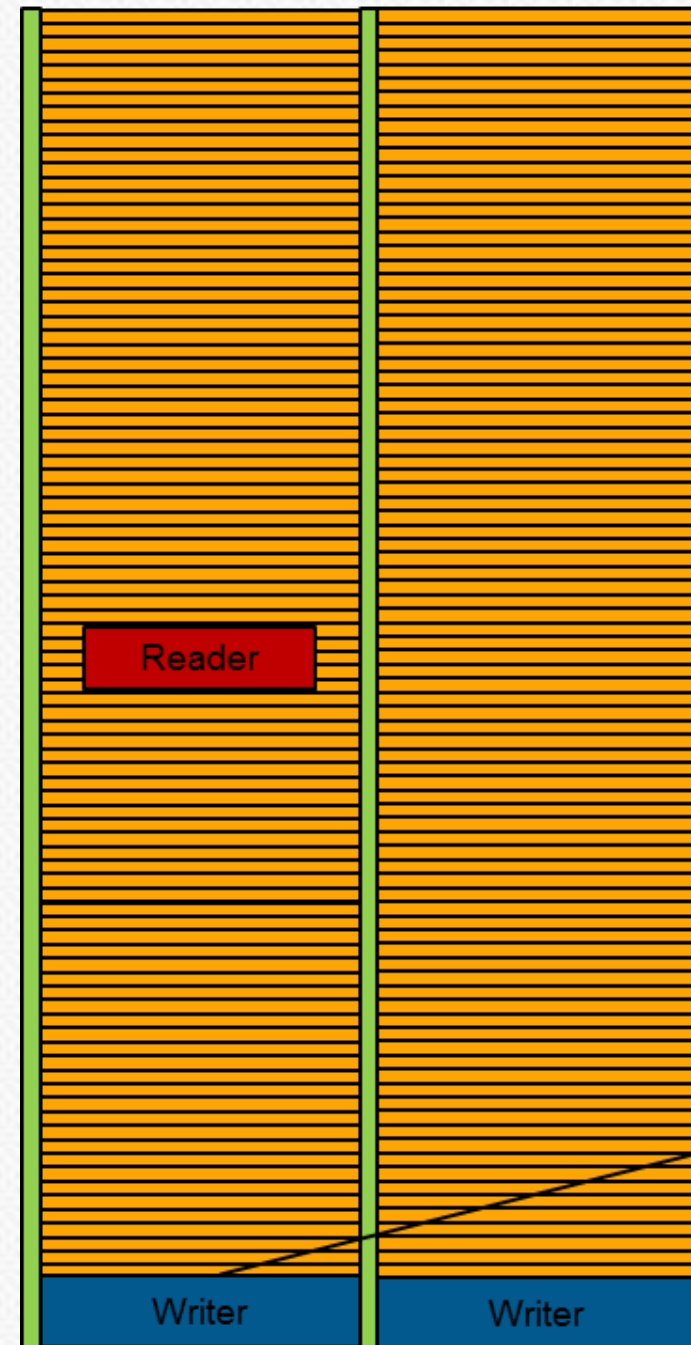
Zone Block Devices

- Benefits
- Leadership
- Status



What Is SMR?

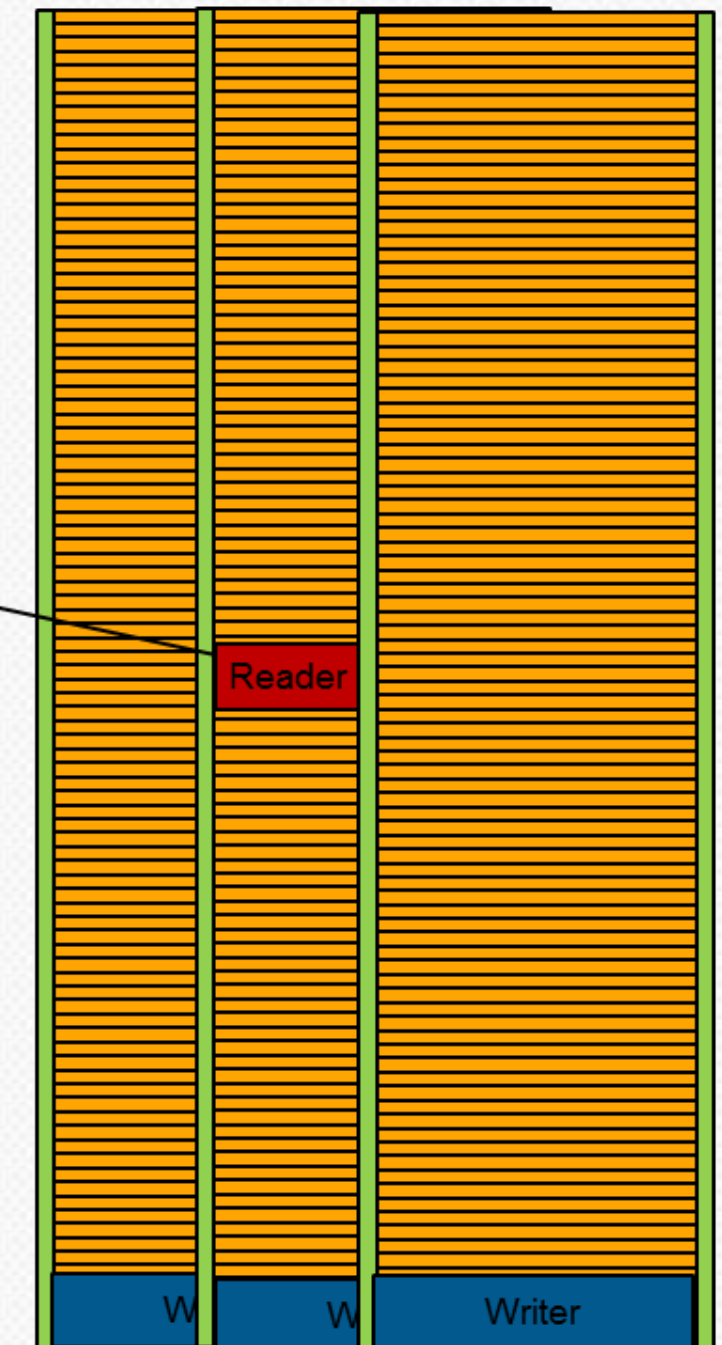
Conventional Recording



1

2

Shingled Magnetic Recording



1

2

3

Track width
determined
by reader width

**Can't rewrite track 1
without destroying
data on track 2**

Track width
determined
by writer width.
**Can't narrow the
write width anymore**



SMR Background

Tracks Physically overlap

- Tracks are written sequentially
- Groups of overlapping tracks are partitioned into zones
- Zone targeting 256MB each (100s of tracks)
- Zones take seconds to write or read completely



SMR Question

Who owns responsibility to write physically sequential?

- HDD?
- HDD + Host?
- Host?

Which is Most Efficient for OCP?

- Host!

Why?

- Host is content and system aware. HDD only knows about LBAs



Host Managed SMR

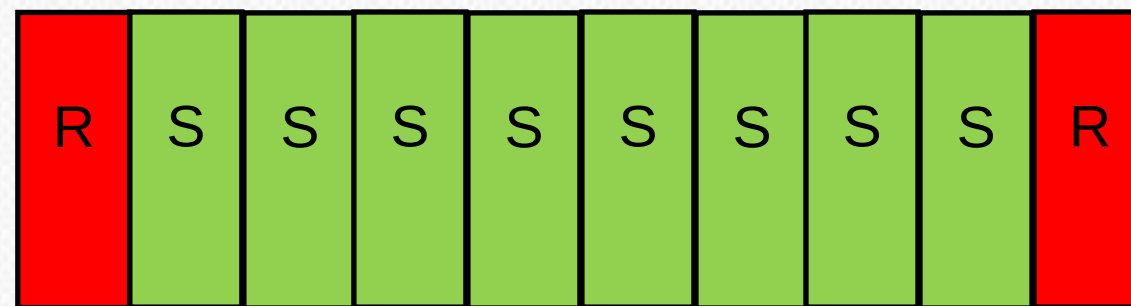
Host is aware of Zones & Zone Types

- Random Writeable
 - Sequential Write Only
- } Statically Mapped

New Commands to report Zone Boundaries and Write Pointers

Host writes at Write Pointer for Sequential Zones Only

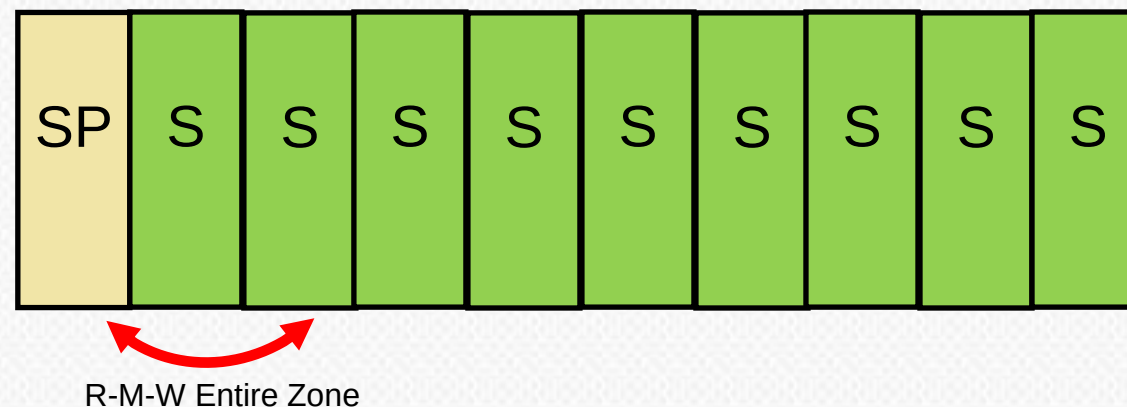
Drive does not do Read-Modify-Writes of Sequential Zones



Host Managed Compared to Host Aware

Host Aware has another zone type

- Sequential Write Preferred Zone
- Dynamic Mapping
- Indirection zone
 - Limited Resource (Media / SSD / DRAM / Metadata)
 - Indirection blends LBAs into the sequential preferred zone
 - Resource Recovered by Read – Modify – Write
- Out of order writes may change the state of a zone to use indirection resources



Costs of Host Aware

Costs of “Sequential Preferred” Zone

- Unpredictable Performance:
 - At some point, LBAs need to be un-blended
 - Latency penalties for indirection resource recovery,
 - R-M-W takes seconds
 - Performance will vary between vendors
- More Power
 - Un-blending granularity is done by R-M-W of entire zone
 - To retrieve all indirection resources, could require R-M-W many zones
- Increases Write Amplification
- Writes may be received by the drive out-of-order, with no host notification
 - A failed write command may lead to this condition & non-predictable performance



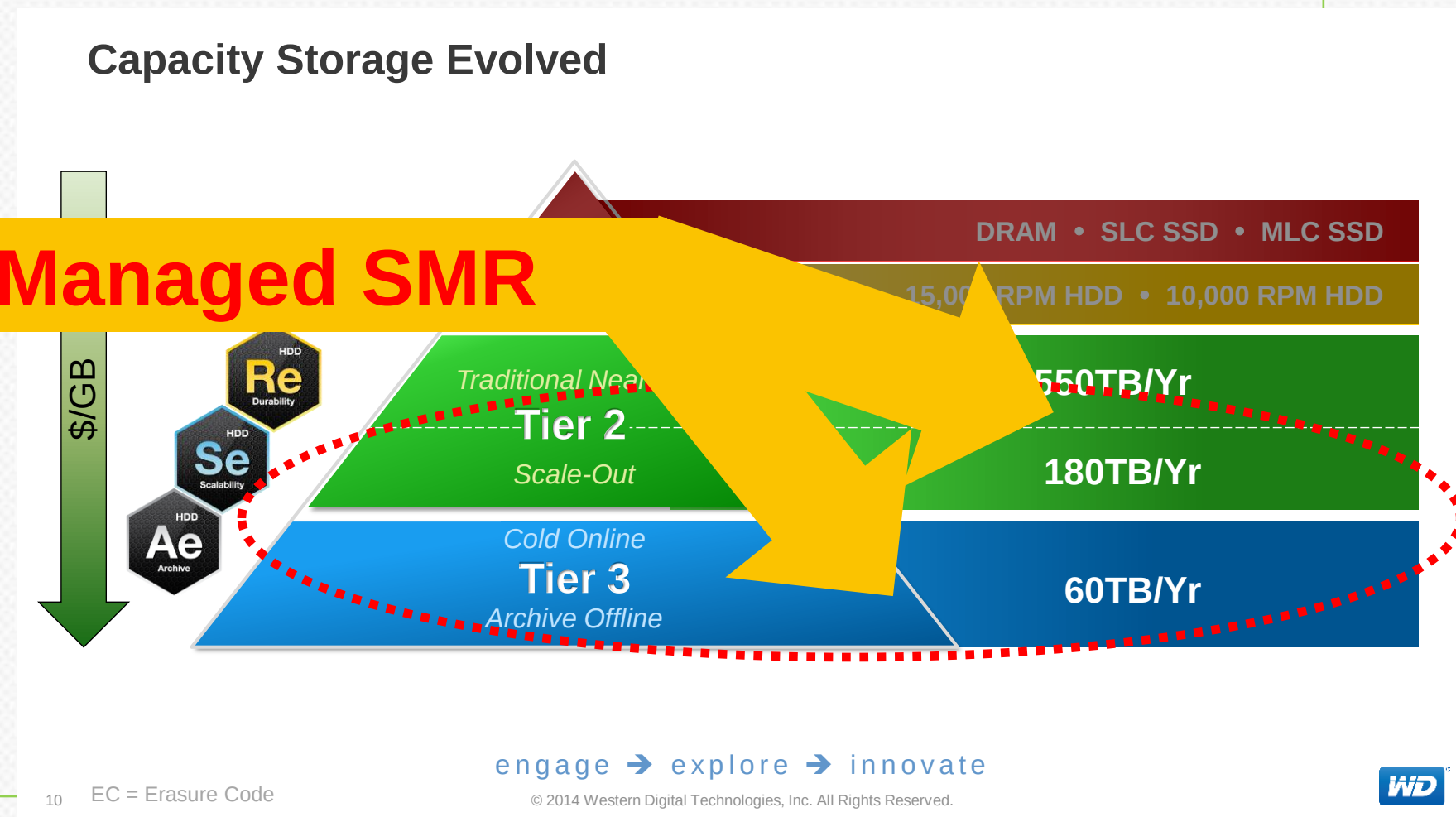
Advantages of Host Managed

Benefits of Host Managed

- More host control over device behavior
- Command response times similar to conventional HDD
- Lower \$/GB
- Simpler Device
- Lower Power
- No Write Amplification

If any of above is important
then HM is the right choice!

Strong Fit with Host Managed SMR



Host Managed Enablement - Evolution

First Generation (2011) – Single Zone (like tape)

- Enabled with Customer Application doing direct I/O
- LTFS + WD Driver

Next Generation (2012)– Multiple zones

- Enabled with NilFS

Upcoming (Today)

- Device Mapper – Linux Community buy-in 3/2014
 - WD is contributing SMR Simulator and Device Mapper assistance <https://github.com/westerndigitalcorporation>
- Bcache for remapping
- Btrfs / ZFS – “SMR Host Managed Friendly”
- Suse (Dr. Hannes Reinecke) demonstrating significant progress

<http://git.kernel.org/cgiit/linux/kernel/git/hare/scsi-devel.git/commit/?h=zac.v2&id=1792a4c3e928cce98d516c7bc63e935f09a33313>



Zone Bock Devices (ZBC / ZAC)

▪ **WD Leadership**

- WD championed exposing zone structures on Media in 2012 @T10 – WD Wrote Specification
- WD is standards editor (ATA/SCSI) – Curtis Stevens
- Multiple development partners embracing and moving forward

▪ **Benefits**

- Greater density @ CMR performance (\$/GB) for ZAC compliant hosts
- Enables Optimum Performance/Power/Workload highly aligned with Host Managed
- Better consistency across vendors for SMR

▪ **Status**

- Proposals nearly complete
- ZBC Letter Ballot – January 2015– Standard is stable
- ZAC (SATA) expected to follow by 2-4 months





Backup