Open. Together.
SMR and ZNS – Two Sides of the Same Coin

Lee Prewitt
Principal Hardware Program Manager
Microsoft
Exploding Storage Growth
TLA Decoder Ring

SMR – Shingled Magnetic Recording
ZNS – Zoned Name Spaces (NVMe)
ZAC – Zoned ATA Commands (T13)
ZBC – Zoned Block Commands (T10)
Shingled Magnetic Recording

Needed to scale HDD recording density
HDDs can read much narrower than they can write
If you overlap the write tracks, you can increase areal density
But direct overwrite is no longer possible
ZAC and ZBC to the rescue

ZAC standardized in T13 for SATA drives
ZBC standardized in T10 for SAS drives
State machine is very similar to ZNS
Software must write sequentially into an open zone
Software can read randomly from any zone
Software deletes a zone as a whole
How an SMR Drive Works

Conventional Zone

Shingled Zone
QLC Flash

Quad Level Cell
4-bits of data per cell
16 programable voltage levels
Needed to keep scaling SSD capacity
But very difficult to program correctly
Very low endurance (<1,500 Program/Erase cycles)
ZNS to the Rescue

Zoned Name Spaces is being standardized in NVMe
State machine is very similar to ZAC/ZBC
Software must write sequentially into an open zone
Software can read randomly from any zone
Software deletes a zone as a whole
Allows for very efficient garbage collection in the drive
Drives Write Amplification (WAF) close to 1.0
The Software Stack

ReFS – Resilient File System
Copy on Write file system
Very good at sequential workloads
DAX
Direct Access driver for use on Persistent Memory
Allows for storage semantics on NVDIMMs
A Possible System Implementation

VM-1

VM-2

VM-n

NIC

Storage Node

Random Writes

NVDIMM

Sequential Writes

ZAC/ZBC HDD

ZNS SSD
Two Sides of the Same Coin

Zones concept enables continued scaling for both HDDs and SSDs.
Microsoft working together with industry and the standards committees to enable the protocols
Common state model between ZAC/ZBC and ZNS
One software stack works on both!
Open. Together.

OCP Global Summit | March 14–15, 2019