



Flash Memory Summit

Reaching Petabyte-Scale Systems with ZNS SSDs

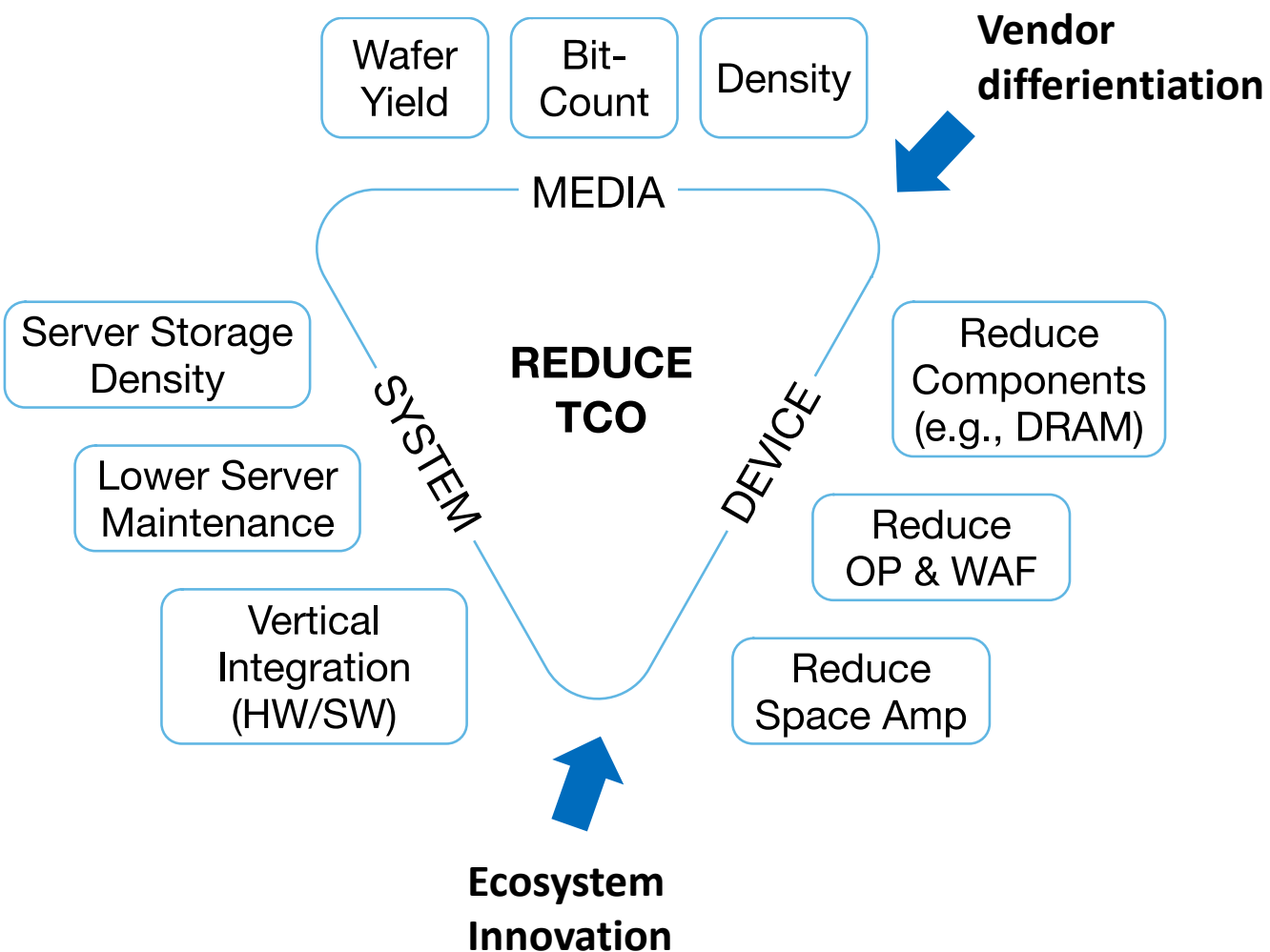
Javier González, Principal Engineer, Samsung Electronics

Wonchul Lee, Principal Engineer, Samsung Electronics

Reducing TCO



Flash Memory Summit



Reduce TCO through Density

Media

QLC NAND

- QLC reduces media TCO
- QLC reliability is 3x lower than TLC

Device

Enable QLC adoption through ZNS

- ZNS balances QLC reliability

System

Increase Server Density

- Move from TB to PB Scale
- Reduce per-server maintenance costs
 - Administration, Licenses, Power, Space, etc.

Provide an open ecosystem

- Focus on open-source and standards

Zoned Namespaces

Concept & Benefits

Core Concepts

- Organize LBA space in zones
- Write sequentially to zones
- Manage zone resets in host
- Use of Append Command
- Use of Simple Copy & ZRWA

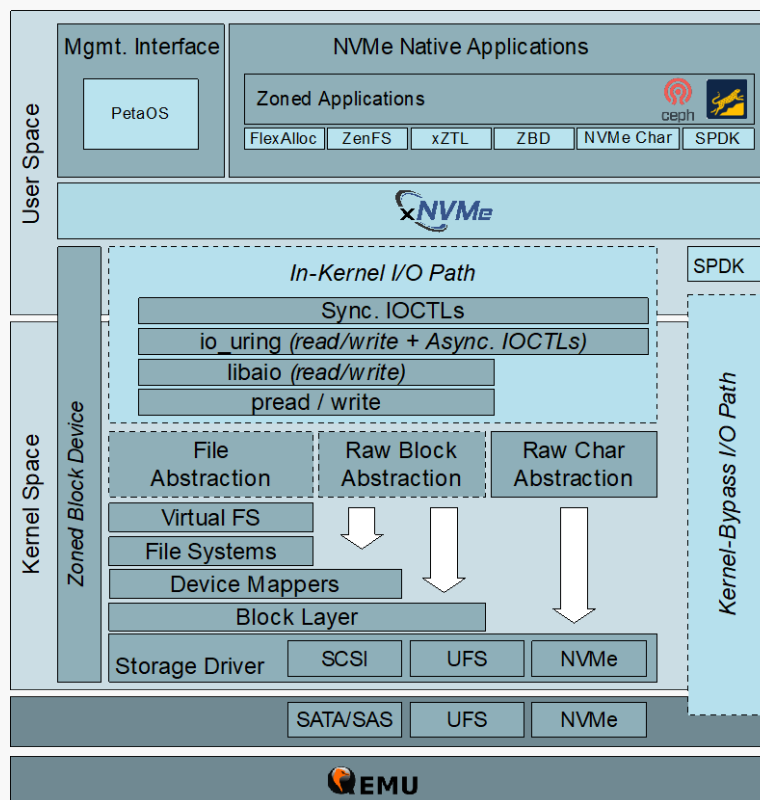
TCO Benefits

- Reduce WAF and OP
- Reduce device DRAM
- Remove device-side GC

➤ *ZNS is a great vehicle to enable adoption of QLC*

Ecosystem

Strong Linux Ecosystem



Standardization

NVMe

- Core spec available
- Small TPs in progress

SNIA Zoned TWG

- Working on spec to align industry around zoned models
- Target specification this year

D2PF

- Linux Foundation Initiative to aid adoption of new interfaces

[BMKT-101-1: Data Storage Strategies](#)

Samsung's Peta Scale Vision

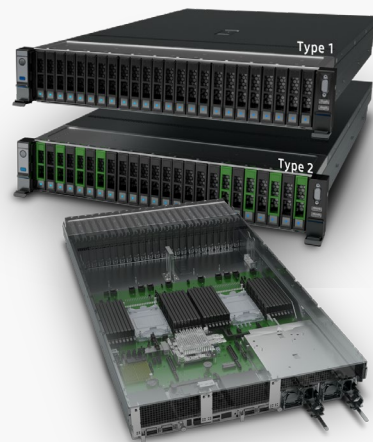
QLC + ZNS

- Reduced TCO
 - WAF, OP, & DRAM
- Increase Device Reliability
 - Use ZNS to reach TLC reliability levels in QLC
- Real workloads: Percona 8.0 on top of RocksDB



FMS'22 Demos

- Setup
 - Type 1 System
 - 128TB QLC ZNS
 - ZNS (QLC) : E3.S 128TB x 12ea (1.5PB)
 - 16 TB TLC CNS
 - CNS (TLC) : E3.S 16TB x 8ea (128TB)
- 2 Demos
 - ZNS QLC
 - PBSSD System



Peta Scale Storage Box

- Peta-byte scale storage
 - Use failure domain
 - Degraded performance due to physical isolation
 - Minimize impact of component failure
 - Trade-off between performance and blast radius
- Use ZNS: Increased management overhead
 - Host to reduce storage mgmt and data movement
- Use PB scale storage
 - Increased bring-up time.
 - Requires reduction of S/W initialization time
- Reach Exa-/Zeta-byte scale
 - Leverage EDSFF high capacity SSDs.
 - Maximization through zero OP of ZNS.

PBSSD: New storage solution for Peta-byte scale

- Fabric-attached disk array enclosure
- Multi-SSD management software (PetaOS)
 - Usability, performance, and reliability

Data/Storage Mgmt.

Zone abstraction
Copy offload
Volume (NS) per media

Space Efficiency

Triple SSD media
Media tiering

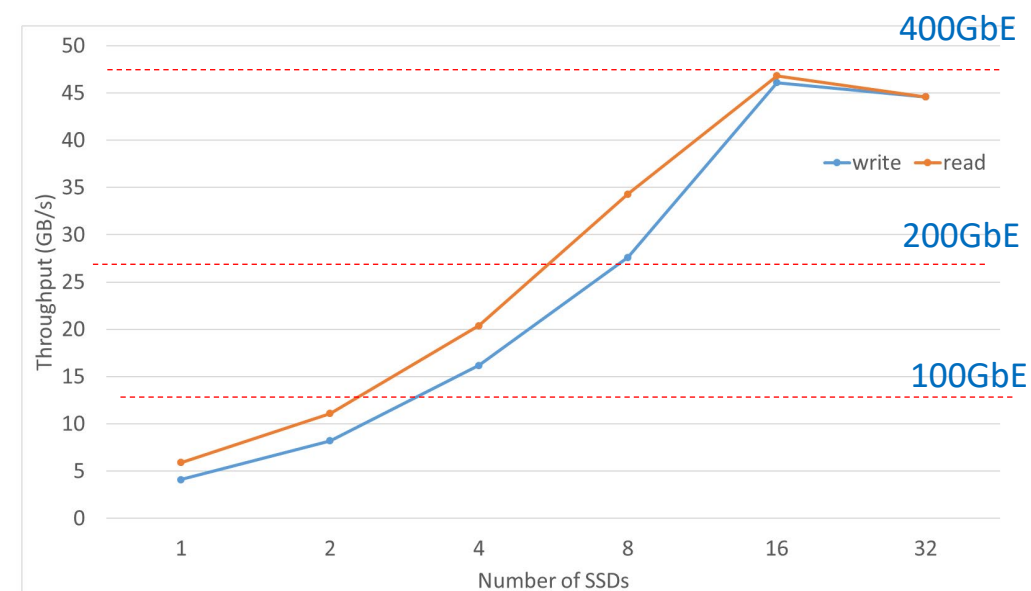
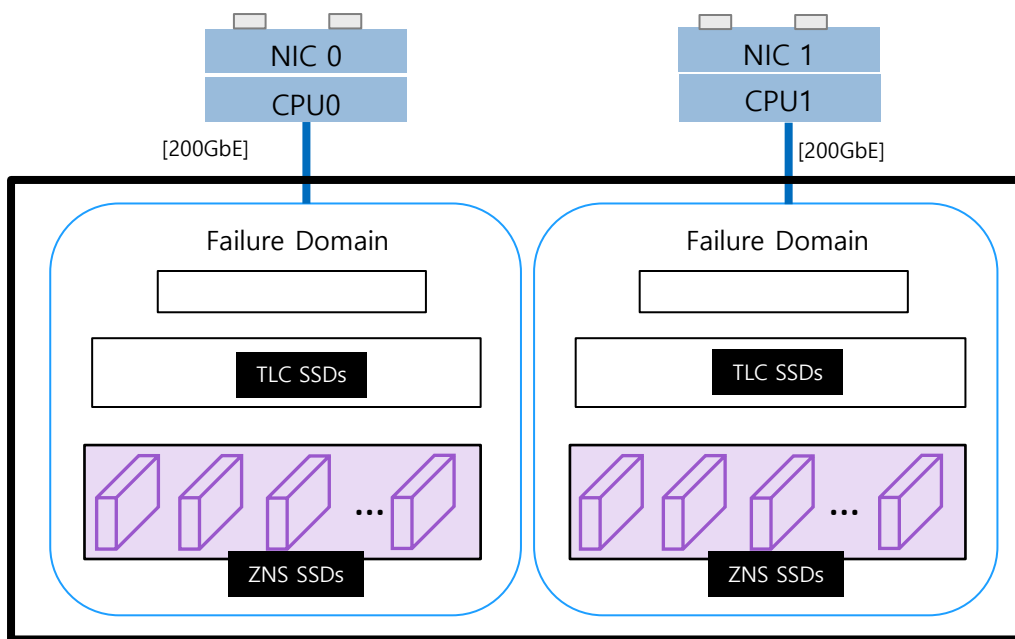
Interface

Conventional NS (CNS)
Zoned NS (ZNS)
Key-value (KV)

NAND Solutions

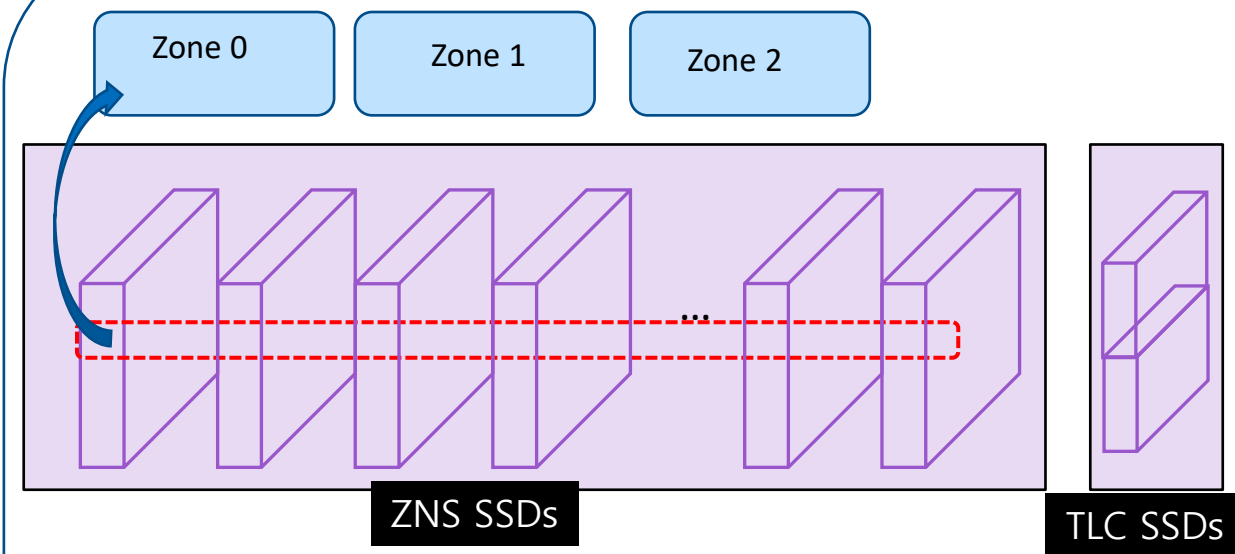
Turn-key solution
Next generation NANDs

- Failure domain (FD) : Isolate to minimize blast radius due to component failure.
 - Use ZNS to maximize user capacity → apply FD rather than RAID.
 - How many SSDs are in a FD?



- Storage & Data Managements : Offload host S/W stacks.

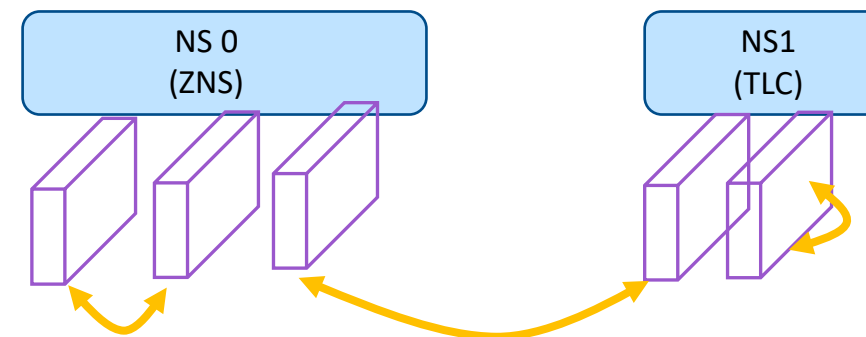
[Storage Management]



It manages next-gen NANDs, or new concept SSDs (ZNS) within the box

- **Service both TLC and ZNS (TLC/QLC) SSDs.**
- **Abstract to enlarge device zones to 20GB ~ 40GB**

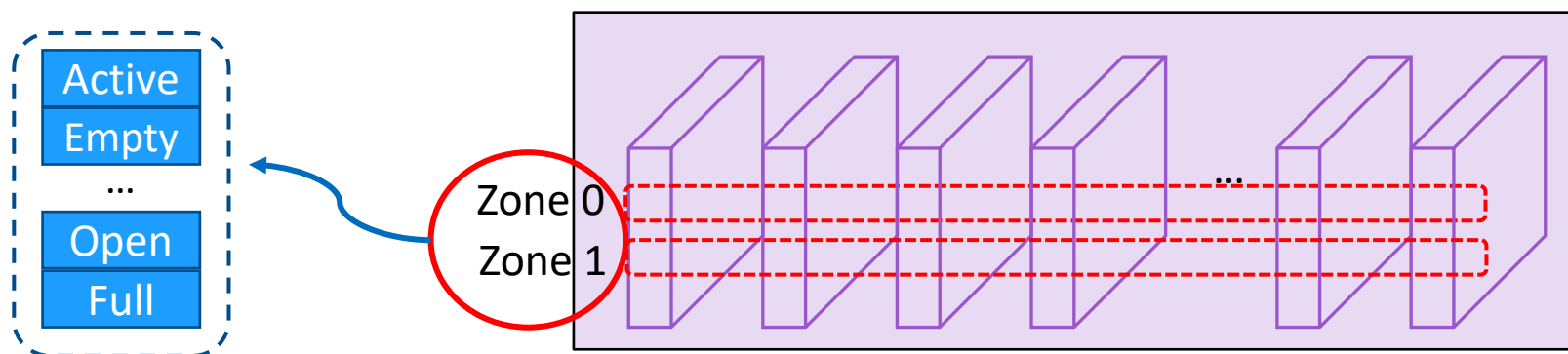
[Data Movement]



Copy offload

- **Control tiering with different type's storages.**
- **Performance can exceed network max.**
 - Intra copy : data move inside the volume (namespace)
 - Inter copy : data move between the volume (namespace)

- Shortening the init. time of the PB scale system will be competitive.
 - $128\text{TB} * 24\text{ea} = 3\text{PB}$
 - Bring-up time : Server H/W init. + Device map load + PetaOS initial.



Use zone characteristics as meta info.

- Save state info. in the CNS => 4sec
- Reconfigure last zone states using limited devices. => 40sec

Takeaways

- Question:
 - How we can reduce the TCO for NAND-based systems?
- Proposed Solution:
 - Reduce TCO by increasing storage density
- 3 Key Contributions:
 - QLC + ZNS: NAND OP, WAF, Space Amp, and server maintenance
 - Petabyte Scale: HW + SW
 - Open Ecosystem
- Visit Samsung's booth

Reaching Petabyte-Scale Systems with ZNS SSDs

Wonchul Lee <wonchul08.lee@samsung.com>

Javier González <javier.gonz@samsung.com>