Application Optimization with Flexible Zone Namespace Configurations in QLC-Based SSDs

David Wang
Director, SSD FW Development
Silicon Motion Technology Corp.

Legal Notice and Disclaimer

- The content of this document including, but not limited to, concepts, ideas, figures and architectures is furnished for informational use only, is subject to change without notice, and should not be construed as a commitment by Silicon Motion Inc. and its affiliates. Silicon Motion Inc. assumes no responsibility or liability for any errors or inaccuracies that may appear in the informational content contained in this document.
- Nothing in these materials is an offer to sell any of the components or devices referenced herein.
- Silicon Motion Inc. may have patents, patent applications, trademarks, copyrights, or other intellectual property rights covering subject matter in this document. Except as expressly provided in any written license agreement from Silicon Motion, Inc., the furnishing of this document does not give you any license to these patents, trademarks, copyrights, or other intellectual property.
- © 2023 Silicon Motion Inc. or its affiliates. All Rights Reserved.
- Silicon Motion, the Silicon Motion logo, MonTitan, the MonTitan logo are trademarks or registered trademarks of Silicon Motion Inc.

Agenda

Why use ZNS with QLC?

ZNS Usage Models

Flexible ZNS Config in SM8366

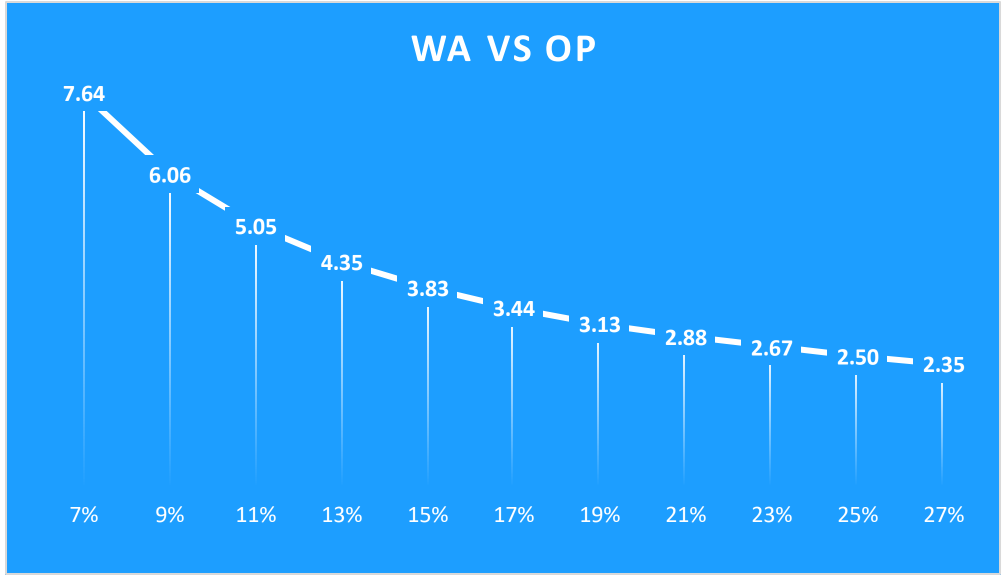
Summary

Why Use ZNS with QLC?

Why Use ZNS with QLC?

Why ZNS?

- GC/WA is expensive
- SSD doesn't know how to do it right
- → Let hosts worry about it
- → Only sequential writes allowed



Why Use ZNS with QLC?

Why QLC?

- Lower cost: 70~80% of TLC
- Good read performance
  - ~100us read latency
  - Less than 100 dies to saturate PCIe Gen 5 x4 under random read workload
- Fair write performance
- Poor endurance: 1.5K~3K PEC
  - TLC: WA of 5 → Good DWPD
  - QLC: WA of 1 → Good DWPD, as well

Source: ISSCC'2022

	Layer	Plane	Area density	Prog BW	tR	Interface
Kioxia/WD, 1Tb	162	4	15.0Gb/mm2	**60MB/s**	**65us**	2400MTs
Micron	176	4	14.7Gb/mm2	**40MB/s**	**90us**	1600MTs
SK-Hynix	176	4	14.8Gb/mm2	**40MB/s**	**90us**	1600MTs
Samsung	(NA) close to 176	4	11.55Gb/mm2	**164MB/s**	**45us**	2400MTs

SiliconMotion

Why Use ZNS with QLC?

Flash Memory Summit

ZNS + QLC can serve

hosts who

- Do sequential writes (low WA) only
  - Do no GC at all, or
  - Do GC on their own
- Desire good read performance
- Desire lower cost

Good Value!

- **OLAP**
- **Big data + AI**
- **Media streaming**
- **NoSQL**
- **And more …**

ZNS Usage Models

-defined by SNIA ZSTWG

Use Cases in Zone Storage Models

	Model A	Model B	Model C
Example Use Cases	Wide range, e.g., streaming application, database applications (WAL, log-structured writes), AFA	High host IO parallelism to achieve full media bandwidth and more compatible to ZBC/ZAC host software stack. Archival Storage as an example	Support parallel accesses by many users (e.g., VMs)
Open/Active Zone Resources	12 or more recommended.	Depends on device	Tens to thousands
Performance	Accessing 1 to 4 Zones concurrently should achieve the max throughput of the associated media to the NS and/or device.	Depends on device, recommend one zone span all Media Units (NAND dies) of one device.	Device specifies the minimal open zone number for achieving max IO throughput.
Mandatory IO Cmds	Read, Write	Read, Write	Read, Write
Adv Feature – Zone Append	Zone Append required	Not need	Not need
Adv Feature – ZRWA	Nice to have	Not need	Not needed
Media Reliability Mgmt.	Device	Device	Shared, RAID protection for die failure will be moved to host.

SM8366 ZNS Spec

	Model A	Model B	Model C
Example Use Cases	Supported	Supported	Supported
Open/Active Zone Resources	16 – full Max throughput per zone access 32 – half Max throughput per zone access 64 – ¼ Max throughput per zone access	16 – full max throughput per zone access	Up to 1024, minimal zone size is 1 or 2 die-block size
Performance	Accessing 1 to 4 Zones concurrently should achieve the max throughput of the associated media to the NS and/or device.	Single Zone access achieve the max IO throughput.	Host manages parallel access to utilize the max IO throughput.
Mandatory IO Cmds	Read, Write, Optional Simple Copy, Optional – in-order execution of N outstanding write commands)		
Adv Feature – Zone Append	Zone Append support		
Adv Feature – ZRWA	Support ZRWA, up to 1MB per Open/Active Zone or disable		
Media Reliability Mgmt.	Besides NAND ECC, advanced NAND Management and Wear leveling, Device provides RAID protection against die failure.		Shared, RAID protection against die failure will be moved to host.

Implementation of Flexible ZNS Config

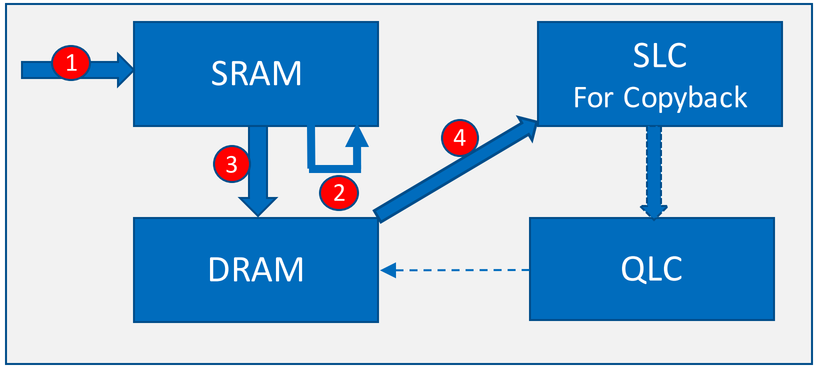
Challenges in Flexible Open Zone Support

- 1K+ Open Zones require SSD controller provide a fairly large size of high bandwidth write buffers.
  - Charge-trapping QLC NAND typically needs 2 pass prog with 2 times of data transfer with N WL gaps.
  - To deal with write-write collision and write-erase collision, also need the extra write buffer size to buffer host data during waiting time

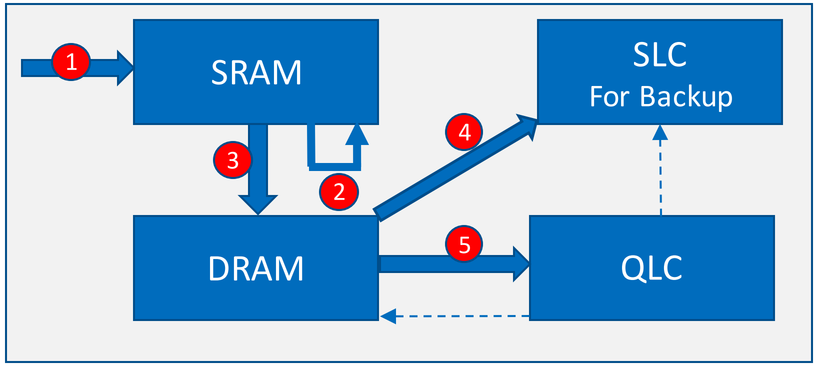
- SSD Controller Constraints
  - Internal SRAM (typically 8-16MB) is not big enough as data manipulation and program buffers for all open zones
  - Using DRAM as prog buffers is limited by the two factors DRAM bandwidth and Supercap capacitor.
  - RAID parity buffer size and capacity overhead.

Thoughts on ZNS Data Flow Design

- Use 2-stage buffering (SRAM+DRAM) for data manipulation and programming respectively
  - SRAM: RAID encoding, FW metadata insertion, RAID parity buffer (2)(3)
  - DRAM: data buffer for programming SLC (4) or directly programming QLC (5)



- Solution A/B for support 1K+ Open/Active Zone
  - A with SLC Copyback: less DRAM Bandwidth need, but additional flow to monitor SLC block RBER.
  - B with SLC Backup: more DRAM Bandwidth consumed



Configurable Zone Options

ZNS Options	Zone Capacity	Max Open Zone Num	RAID Protection Options
1	128 die-blocks (21-22GB)	8	Die/Plane/NA
2	64 die-blocks (10-11GB)	16	Die/Plane/NA
3	32 die-blocks (~5GB)	32	Die/Plane/NA
4	16 die-blocks (~2.5GB)	64	Plane/NA
5	8 die-blocks (~1.2GB)	128	Plane/NA
6	4 die-blocks (~0.6GB)	256	NA
7	2 die-blocks (0.34GB)	512	NA
8	1 die-blocks (0.17GB)	1024	NA

- *One Die-block is one group of plane blocks with the same Word-line offset, e.g. the die-block0 of Micron N48R QLC die comprises of 4 of block0 of the 4 planes.*
- *The ZNS config table is based on ZNS SSD with 128 N48R dies (or LUNs).*

Summary

- SM8366 with configurable data-flow supports flexible zone configurations in 3 ZNS usage models defined in SNIA Zoned Storage Task Group.
- SM8366 supports Turnkey ZNS solution or fully customized FTL with layered FW architecture.
- SM8366 demonstrates top ZNS IO performance and power efficiency.

Performance Metric	SM8366 ZNS QLC with Micron N48R (U.2 16TB)
PCIe Interface	Gen5
128K Read Bandwidth (GB/s)	14GB/s
128K Write Bandwidth (GB/s)	2.2GB/s
4K Random Read (IOPS)	2.3M
Sequential Write Latency (4K)	7.39us*
Sequential Read Latency (4K)	7.32us*
Random Read Latency (4K)	118.8us*

* Measurement based on SPDK

Meet us at booth #315

Scan to learn more!