Understanding NVMe Zoned Namespace (ZNS) Devices

Nick Tehrany and Animesh Trivedi Contact: <u>N.Tehrany@student.tudelft.nl</u> Homepage: <u>https://nicktehrany.github.io/</u> June 8-10 Compsys'22



SSD

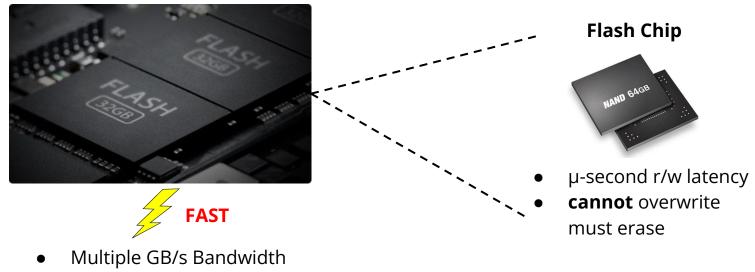




- Multiple GB/s Bandwidth
- Millions IOPS

Nitin Agrawal, Vijayan Prabhakaran, Ted Wobber, John D. Davis, Mark Manasse, and Rina Panigrahy. Design tradeoffs for SSD performance. 1. 2008. USENIX ATC'08.

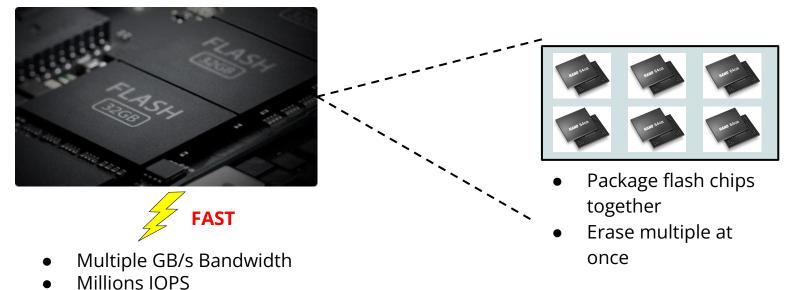
SSD



• Millions IOPS

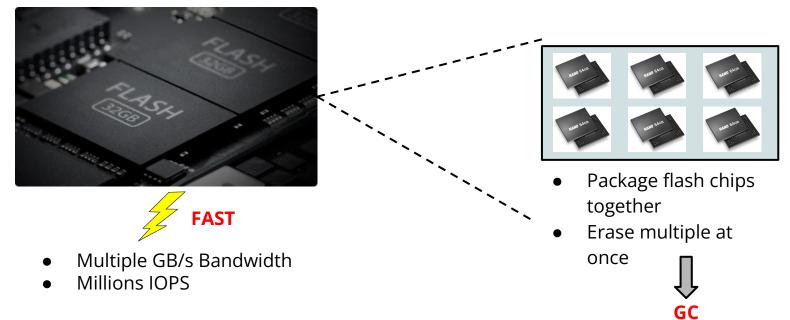
1. Nitin Agrawal, Vijayan Prabhakaran, Ted Wobber, John D. Davis, Mark Manasse, and Rina Panigrahy. *Design tradeoffs for SSD performance*. 2008. USENIX ATC'08.

SSD

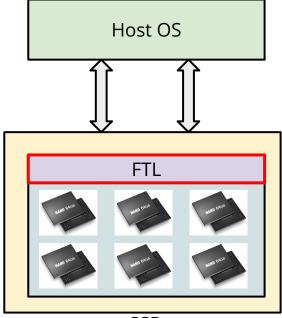


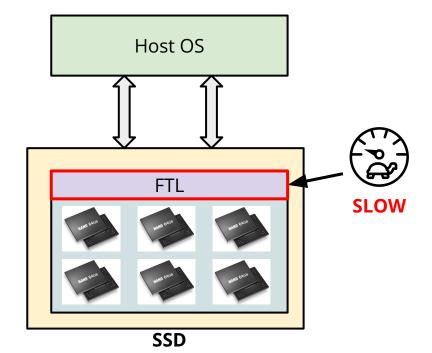
^{1.} Nitin Agrawal, Vijayan Prabhakaran, Ted Wobber, John D. Davis, Mark Manasse, and Rina Panigrahy. *Design tradeoffs for SSD performance*. 2008. USENIX ATC'08.

SSD

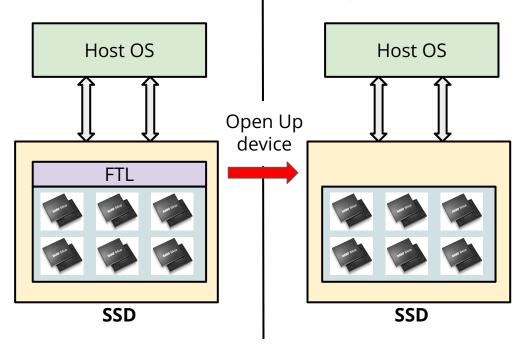


1. Nitin Agrawal, Vijayan Prabhakaran, Ted Wobber, John D. Davis, Mark Manasse, and Rina Panigrahy. *Design tradeoffs for SSD performance*. 2008. USENIX ATC'08.

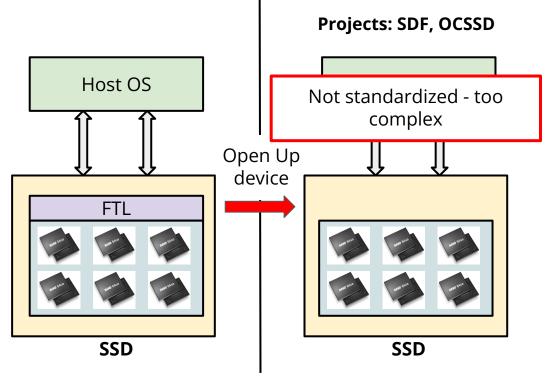




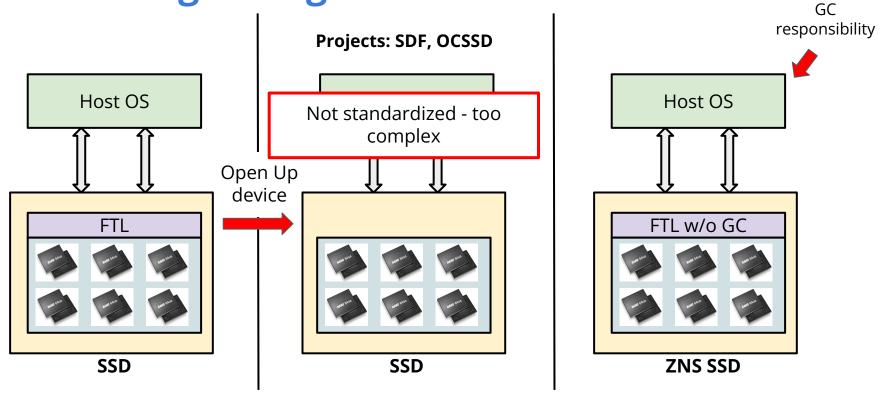
Projects: SDF, OCSSD



- 1. Ouyang, Jian, et al. "SDF: Software-defined flash for web-scale internet storage systems." Proceedings of the 19th international conference on Architectural support for programming languages and operating systems. 2014.
- 2. Picoli, Ivan Luiz, et al. "Open-Channel SSD (What is it Good For)." CIDR. 2020.

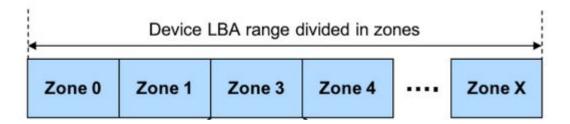


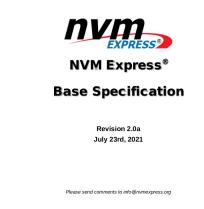
- 1. Ouyang, Jian, et al. "SDF: Software-defined flash for web-scale internet storage systems." Proceedings of the 19th international conference on Architectural support for programming languages and operating systems. 2014.
- 2. Picoli, Ivan Luiz, et al. "Open-Channel SSD (What is it Good For)." CIDR. 2020.



- 1. Ouyang, Jian, et al. "SDF: Software-defined flash for web-scale internet storage systems." Proceedings of the 19th international conference on Architectural support for programming languages and operating systems. 2014.
- 2. Picoli, Ivan Luiz, et al. "Open-Channel SSD (What is it Good For)." CIDR. 2020.

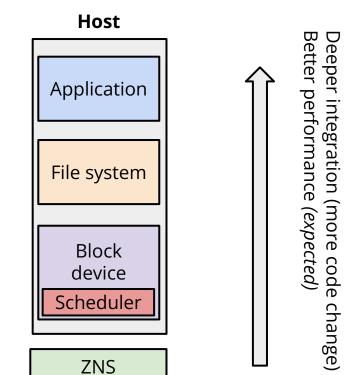
Zoned Namespace (ZNS) Devices



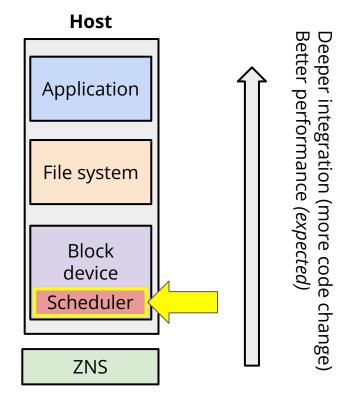


- Append-only sequential writes
- Zone erase before overwrite
- 1 outstanding write per zone

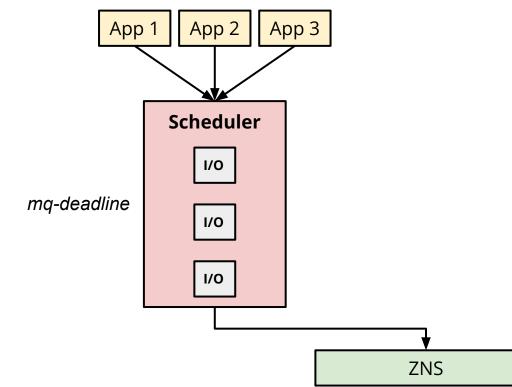
ZNS Integration - Tradeoff



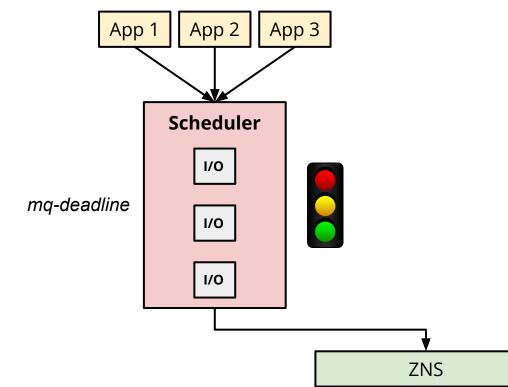
ZNS Integration - Tradeoff

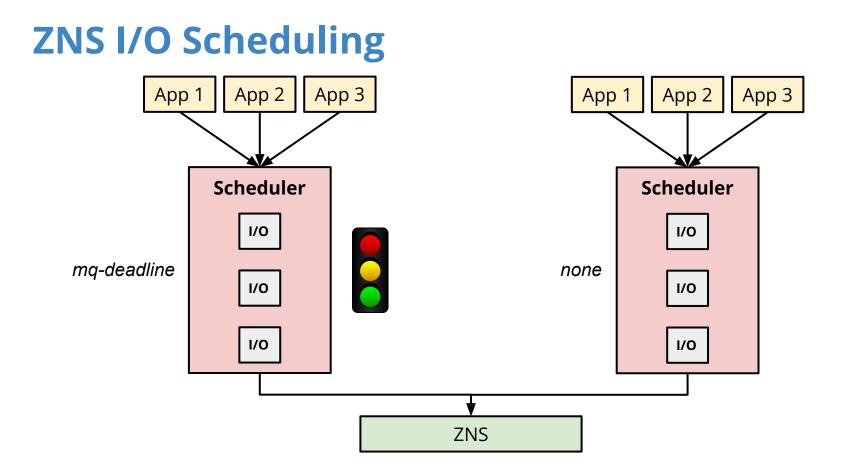


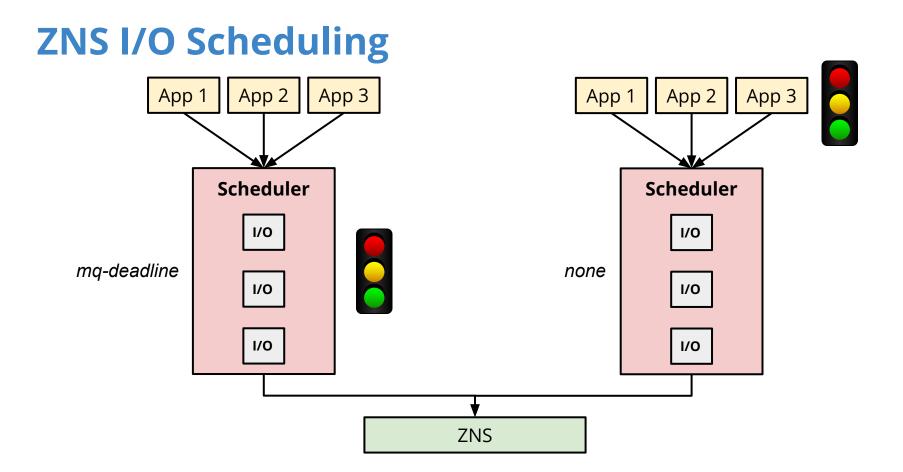
ZNS I/O Scheduling



ZNS I/O Scheduling

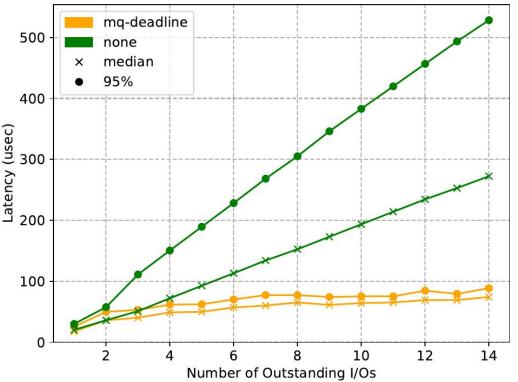




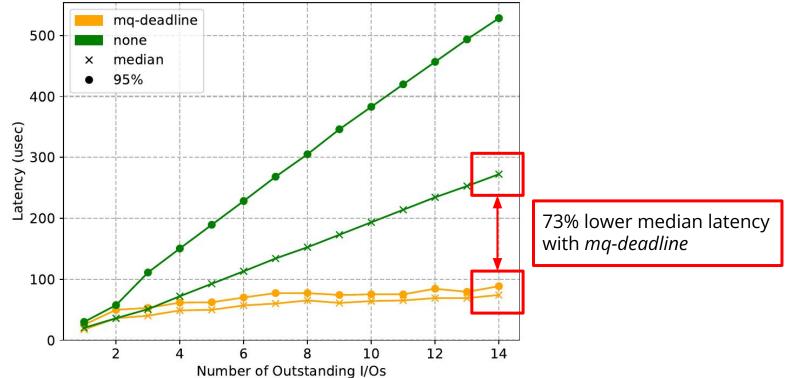


Initial Results

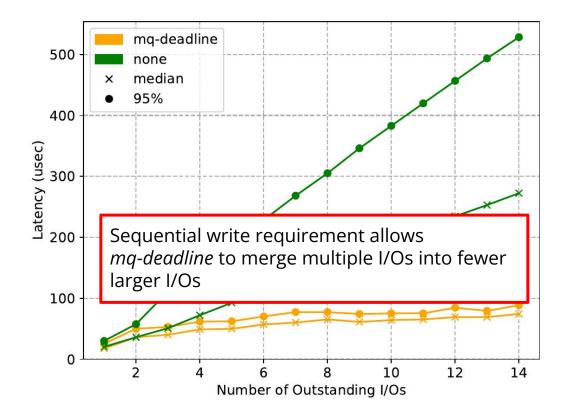
- <u>Write</u> performance of both scheduling configurations:
 - *mq-deadline* increasing number of outstanding I/Os in a single zone
 - *none* single I/O per zone and increasing concurrent zones



Initial Results



Initial Results



Ongoing Work

- Performance at the varying levels of integration
- GC implications at the different levels
- Concurrent instances of different levels

Summary

- ZNS present unique addition to host storage stack
- Gaining significant attention in research community

	NVMe Zoned Namespace (ZNS) Flas	h SSD Storage Devices
Jnderstanding lick Tehrany, Animesh	NVMe Zoned Namespace (ZNS) Flas	h SSD Storage Devices
lick Tehrany, Animesh		Sectore Content C Sectores
	nvedi	
collection overheads, ald However, with the curren Linux Kernel, and provide performance gains, requi- solutions: Operating Syste CM classes: D.4.2 arXiv:2206.0154 (or arXiv:2206.0154	rg with lower device write amplitication. Trixis, additionally increasi software stack there are numerous ways of Integrating 2NS devic a Initial set of performance measurements. Our main findings at ing careful consideration of 2NS Integration and configuration dep ms (ca.05); Performance (cs.PF)	a ploement and garhage collection out on the device to the hout. This allows the hout has allows ener option grant mode allows as a result. 216 dovices arguing spatial adjustment in the research community, as into a hout garhan and an arguing spatial adjustment of the spatial and and an arguing spatial adjustment in the spatial or t
ubmission history		
rom: Animesh Trivedi (view) v1] Fri, 3 Jun 2022 12:54:55		
Bibliographic Tools	Code & Data Demos Related Papers About arXiv	abs

https://arxiv.org/abs/2206.01547

P master - P 1 branch 🛇	Go tags	to file Code -	About
2 nicktehrany Update	376762b 18 days e	igo 🕲 75 commits	ZNS St scripts,
IO_Performance	Added missing data back	22 days ago	🖽 Rea
figures/IO_Perf	Update	18 days ago	화 MIT 쇼 0 st
plot	Update	18 days ago	⊕ 1w
gitignore	Updated plotting script and data loc	3 months ago	¥ 0 to
P LICENSE	Initial commit	3 months ago	
README.md	Merge branch 'master' of https://github.com/nicktehrany/ZNS-Study	last month	Langua
aet zoned device info	Updated scripts	3 months ago	
			 Shell
plot.py	Figure update	2 months ago	
Di un braite	And the second second	0	
run_benchs README.md ZNS-Study	Added write seq bench	2 months ago	
README.md README.md This repository contains all rui additionally contains pointing a present in this repository here Requirements:	wing sorigts for the benchmarks run in the evaluation of NVMe ZNS dev organ, as well as collected data and figures. The final paper of this evalu	ices. It	
README.md ZNS-Study This repository contains plotting s present in this repository here	wing sorigts for the benchmarks run in the evaluation of NVMe ZNS dev organ, as well as collected data and figures. The final paper of this evalu	ices. It	
README.md README.md This repository contains all ru additionally contains potiting s present in this repository here Requirements: Linux Kernel 5.9+ (for ZN	wing sorigts for the benchmarks run in the evaluation of NVMe ZNS dev organ, as well as collected data and figures. The final paper of this evalu	ices. It	
README.md Image: This repository contains all run additionally contains all run additionally contains and run additionally contains proting is present in this repository here Requirements: • Linza Kernel 5.9 v(or ZN) • linzar	wing sorigts for the benchmarks run in the evaluation of NVMe ZNS dev organ, as well as collected data and figures. The final paper of this evalu	ices. It	
README.md README.md This repository contains all ru additionally contains poting as present in this repository have Requirements: Linux Kennel 5.9+ (dor ZM immme mmic-d	wing sorigts for the benchmarks run in the evaluation of NVMe ZNS dev organ, as well as collected data and figures. The final paper of this evalu	ices. It	
■ README.md EXSS-Study Drust-study prosent in this repository term requirements:	nning scripts for the benchmarks run in the evaluation of NVMe ZNS dev cripts, as well as collected data and figures. The final paper of this evalu S support)	ices. It	
README.md README.md This repository contains gifting additionally contains pointing present in the repository here Requirements: Linux Kennel S.9+ (for ZN Linux Kennel S.9+ (for ZN	nning scripts for the benchmarks run in the evaluation of NVMe ZNS dev cripts, as well as collected data and figures. The final paper of this evalu S support)	ices. It	

Acknowledgements

- This work is generously supported by Western Digital Donations
- Matias Bjørling, Hans Holmberg, and ZNS team at Western Digital provided helpful comments and feedback