# ReFlex4ARM: Supporting 100GbE Flash Storage Disaggregation on ARM SoC
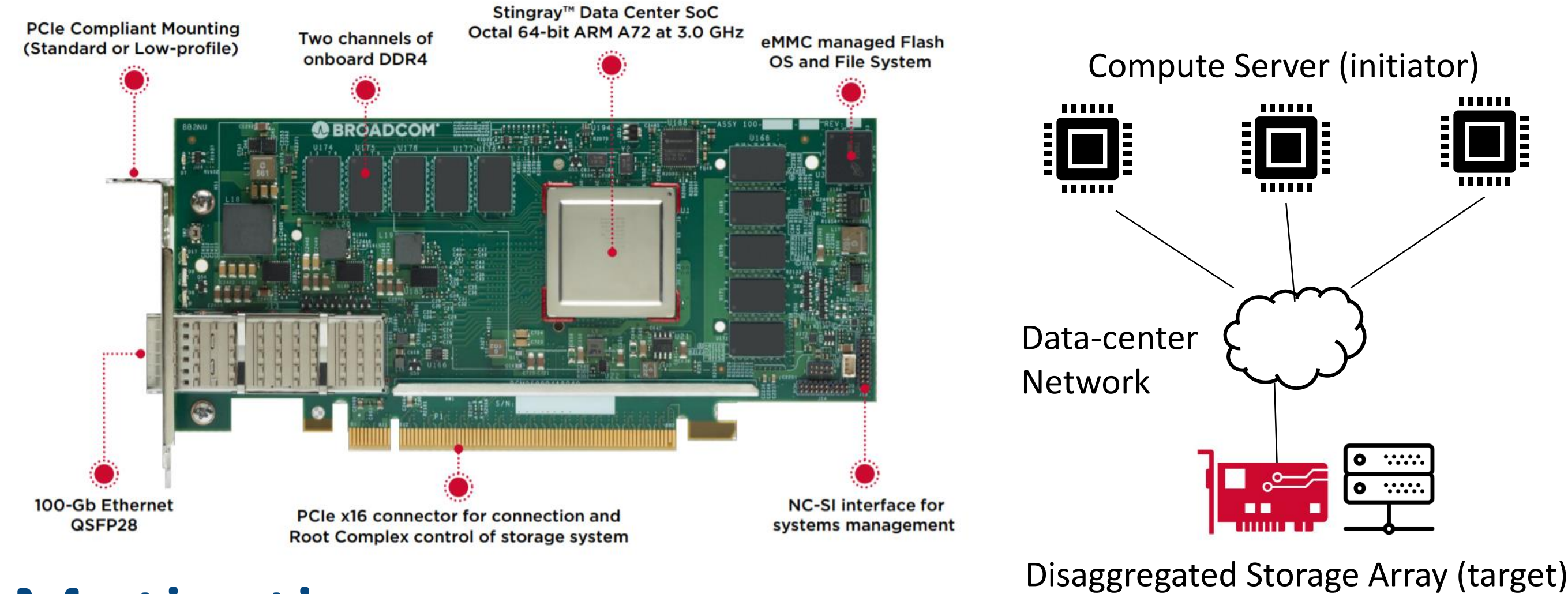
Minghao Xie, Chen Qian, Heiner Litz
Computer Science and Engineering, University of California, Santa Cruz

Contacts:
Minghao Xie, mhxie@ucsc.edu
Chen Qian, cqian12@ucsc.edu
Heiner Litz, hlitz@ucsc.edu

## Overview

Flash Disaggregation [1] enables to share flash storage across the data center, improving resource utilization and reducing the total cost of ownership (TCO). Previous work on flash disaggregation utilized costly server processors leaving significant headroom for optimizing TCO. In this work, we develop a new flash disaggregation system based on a cost-effective and power-efficient ARM-based Smart NIC. This work introduces our architecture and provides a comprehensive evaluation outperforming previous work in TCO by 2.57x.



Compute Server (initiator)

Data-center Network

Disaggregated Storage Array (target)

## Performance

| 2,000,000+ | 100μs | 1.32x10$^{-6}$ $ | 95th |
|---|---|---|---|
| Capable 4K IOPS | Remote Latency | Per million IOs | Tail Read Latency SLOs |

| 257% | 2.3x | | 5-10ms |
|---|---|---|---|
| TCO saving in a 10-year span | Power Saving by Offloading tasks to the smart NIC | | RTT SLA |

## Evaluation



Throughput vs. Latency (4K Rand Read, 4 threads)
- Average
- 10th percentile
- 95th percentile



Throughput vs. Cores (Rand Read, avg = 300us)
- 1K local
- 1K remote
- 4K remote
- 16K remote



ReFlex4ARM (left) v.s. LightOS (right)
- CAPEX
- Cooling
- Power
- Space

Table 1: Hardware Configuration Overview

| Solution | Front-end | Memory | SSDs | NIC |
|---|---|---|---|---|
| ReFlex4ARM | Stingray | 16GB | 16x | 200Gbps |
| LightOS | Xeon | 128GB | 24x | 200Gbps |

## Implementation

1. Added support for multiple SSDs by sharding storage capacity across devices, managed by the centralized control plane;

2. Performed comprehensive networking parameter tuning to minimize overheads for 100GbE;

3. Extended configuration interface to specify the service level objective (SLO) of a flow.

## Discussion

- Telemetry data from NVMe SSDs
- Dynamic IOPS++ by prediction
- Better performance isolation
- Scalable to changing workloads

## Motivation

**Can** we achieve a competitive performance on ARM64 to Intel x86?

**What** are the challenges of porting an application from x86 to aarch64?

**What** is the TCO optimal platform for Flash storage disaggregation?

### Portability

Our design is based on ReFlex [2] which is limited to run on x86 machines depends on specific inline assembly code for some critical sections. To enable the ARM ISAs we have refactored original ReFlex by replacing all arch-specific code with generic implementation.

### Performance

Achieving x86-like performance on ARM processors is challenging as the ARM A72's micro-architecture does not provide the same level of performance than modern x86 architectures. To increase the throughput, we also resort to domain specific hardware acceleration.
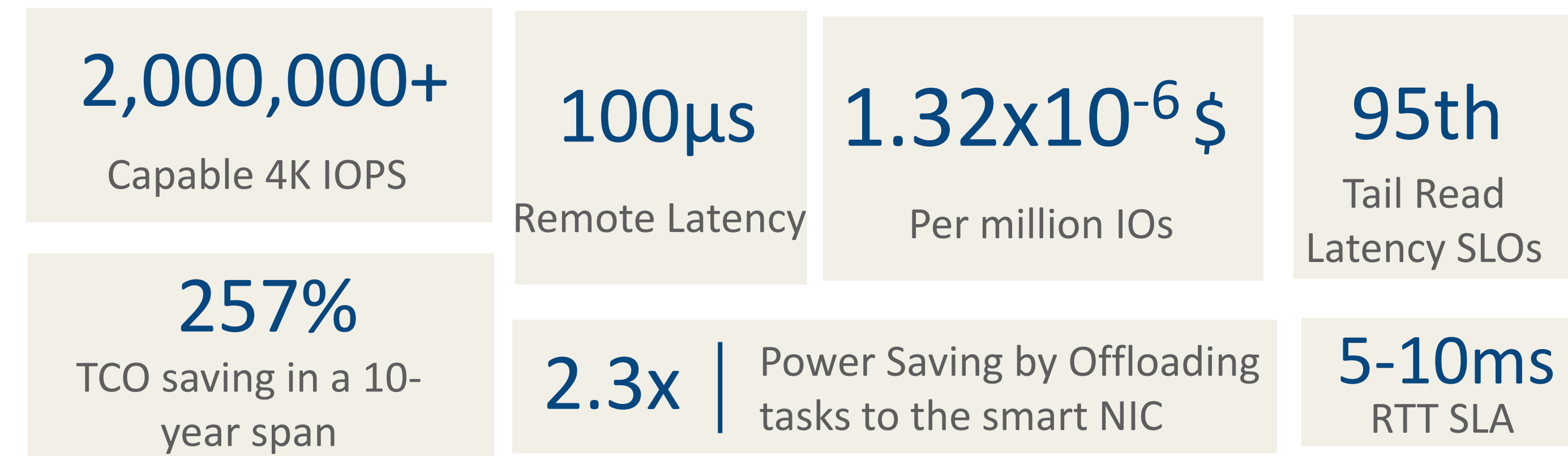
### TCO

TCO represents the primary optimization metric for data center operators [4]. We have developed a cost model for ReFlex4ARM. We utilize the same cost model to get the TCO for Lightbits [5]. ReFlex4ARM reduces TCO by 2.6× and increases 1.57× IO operations with the same cost.

## Contribution

1. Provide a portable flash disaggregation software
2. Explore the differences between ARM and x86
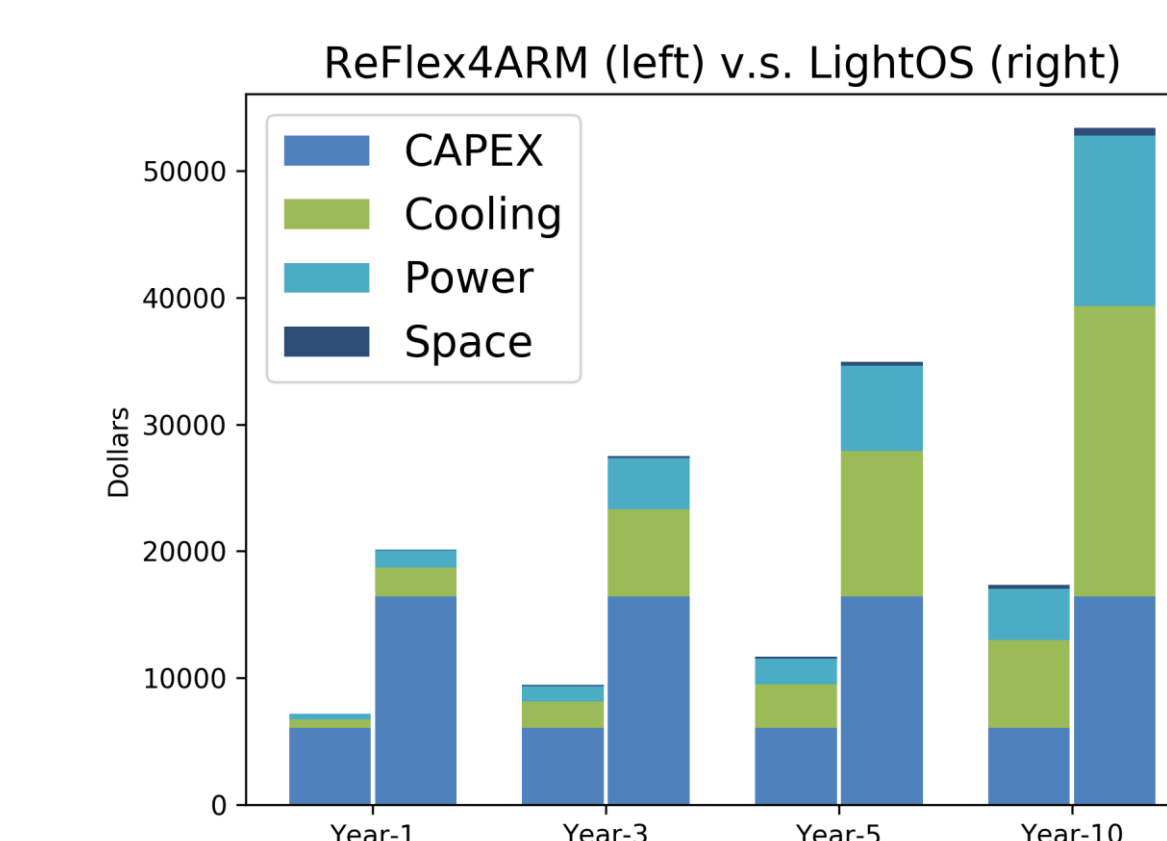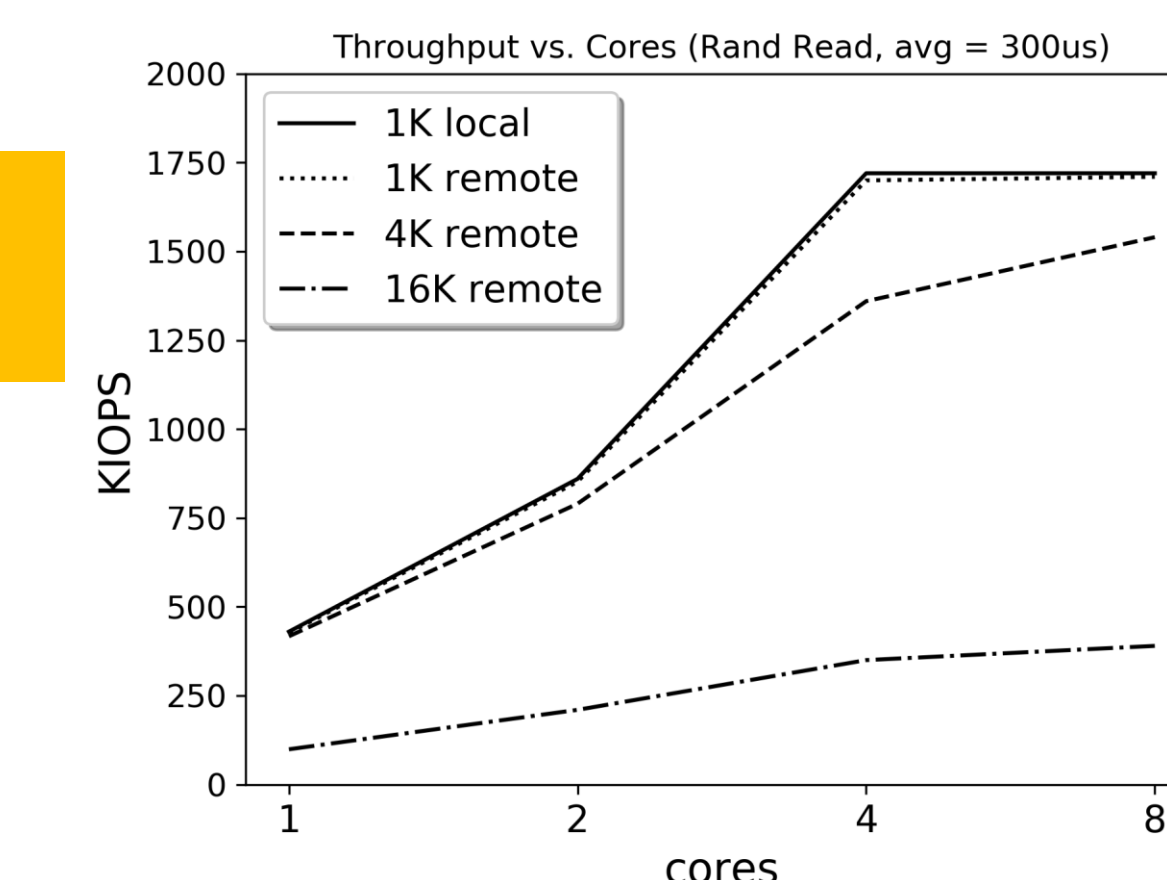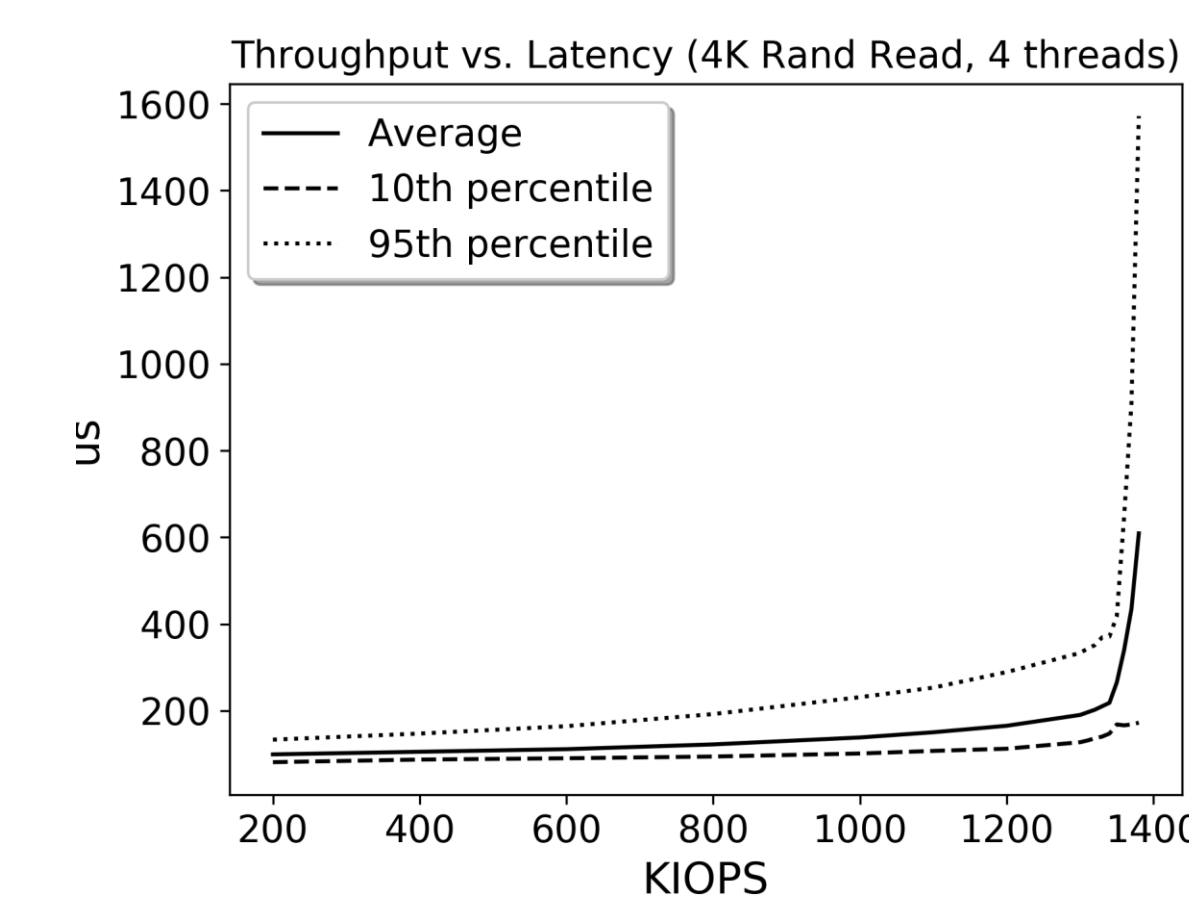3. Achieve a state-of-the-art throughput with better TCO

*Reference:*

[1] Klimovic, A., Kozyrakis, C., Thereska, E., John, B., & Kumar, S. (2016, April). Flash storage disaggregation. In Proceedings of the Eleventh European Conference on Computer Systems

[2] Klimovic, A., Litz, H., & Kozyrakis, C. (2017). Reflex: remote flash≈ local flash. ACM SIGOPS Operating Systems Review, 51(2), 345-359.

[3] https://www.broadcom.com/products/storage/ethernet-storage-adapters-ics/ps1100r

[4] Patterson, M. K., Costello, D., Grimm, P., & Loeffler, M. (2007). Data center TCO; a comparison of high-density and low-density spaces. Thermal Challenges in Next Generation Electronic Systems (THERMES 2007), 42-49.

[5] lightbits Inc., https://www.lightbitslabs.com/dell-poweredge/, 2020

CENTER FOR RESEARCH IN STORAGE SYSTEMS

BROADCOM®

Baskin Engineering

UC SANTA CRUZ

OCP FUTURE TECHNOLOGIES SYMPOSIUM

**2020 OCP Global Summit**