OPENFABRICS ALLIANCE

SSD Architectures and Experiences

Author: Tony Roug
Date: March 27, 1012

Agenda

- Flash Overview
- Performance Characterization Methodology
- Technology Overview

Virident Corporate Overview

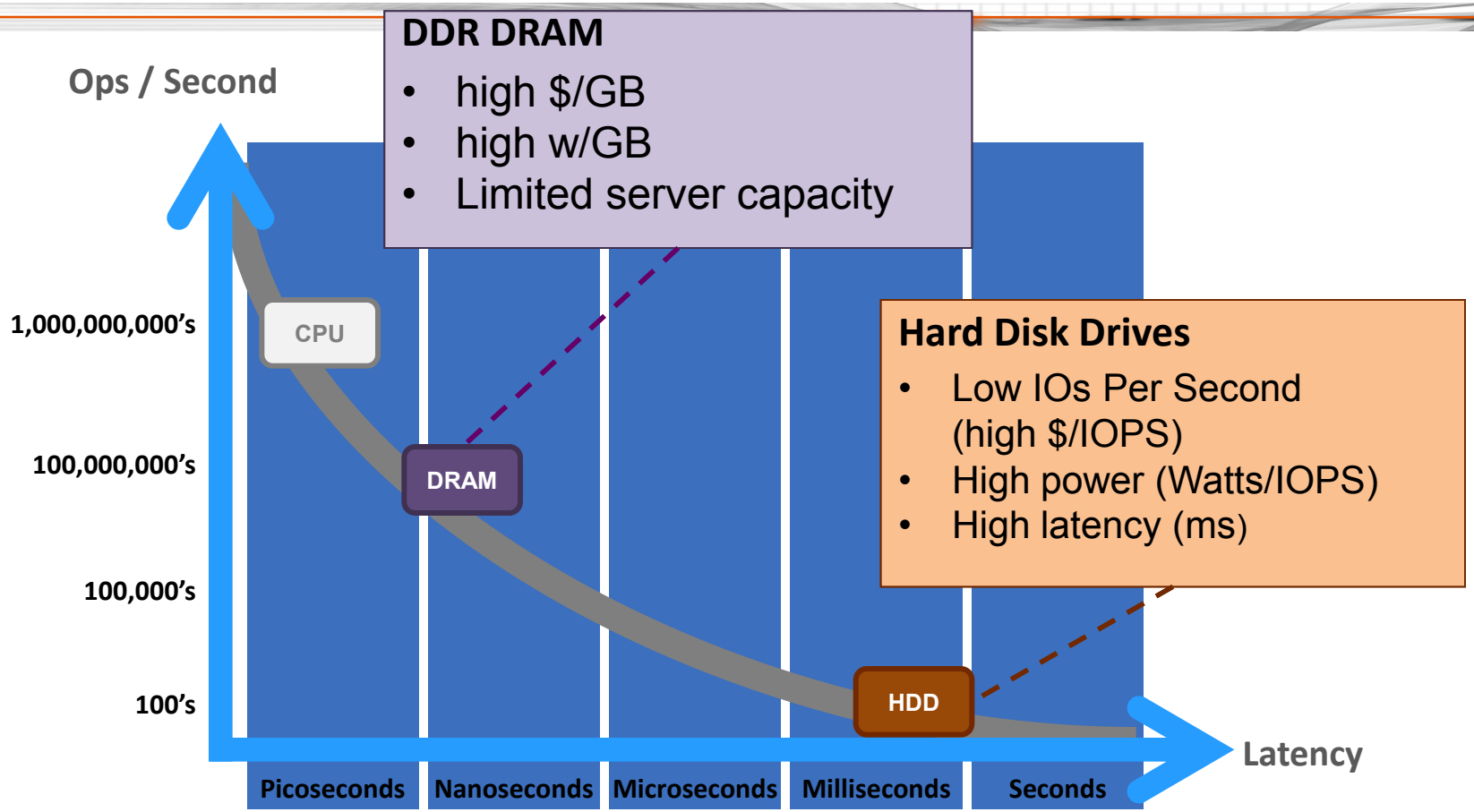
- Mission: Unconditional Performance Technology
  - Experts at system-level NAND management for performance and endurance
  - Blue-chip investors: Sequoia Capital, Globespan Capital, Artiman, Intel Capital, Cisco Ventures
  - Over 100 technical staff in Milpitas, CA and Bangalore, India

- Product: FlashMAX storage class memory
  - Unconditional performance: Highest IOs/G, Lowest latency (µs), Sustained over time
  - Enterprise class reliability: Guaranteed writes, ECC/RAID on board, 7+1 distribution
  - Highest Density: PCIe, half width, half height
  - Delivers 2-5X **application-level** sustained performance[1]

- Customers: High performance datacenter storage
  - Computer and Storage OEMs, Web Portals, Cloud Providers, Fortune 1000 Enterprise
  - U.S., Japan, APAC

[1]Based on 4K Random 70% Read, 30% Write – the typical TPC-C DB workload model,

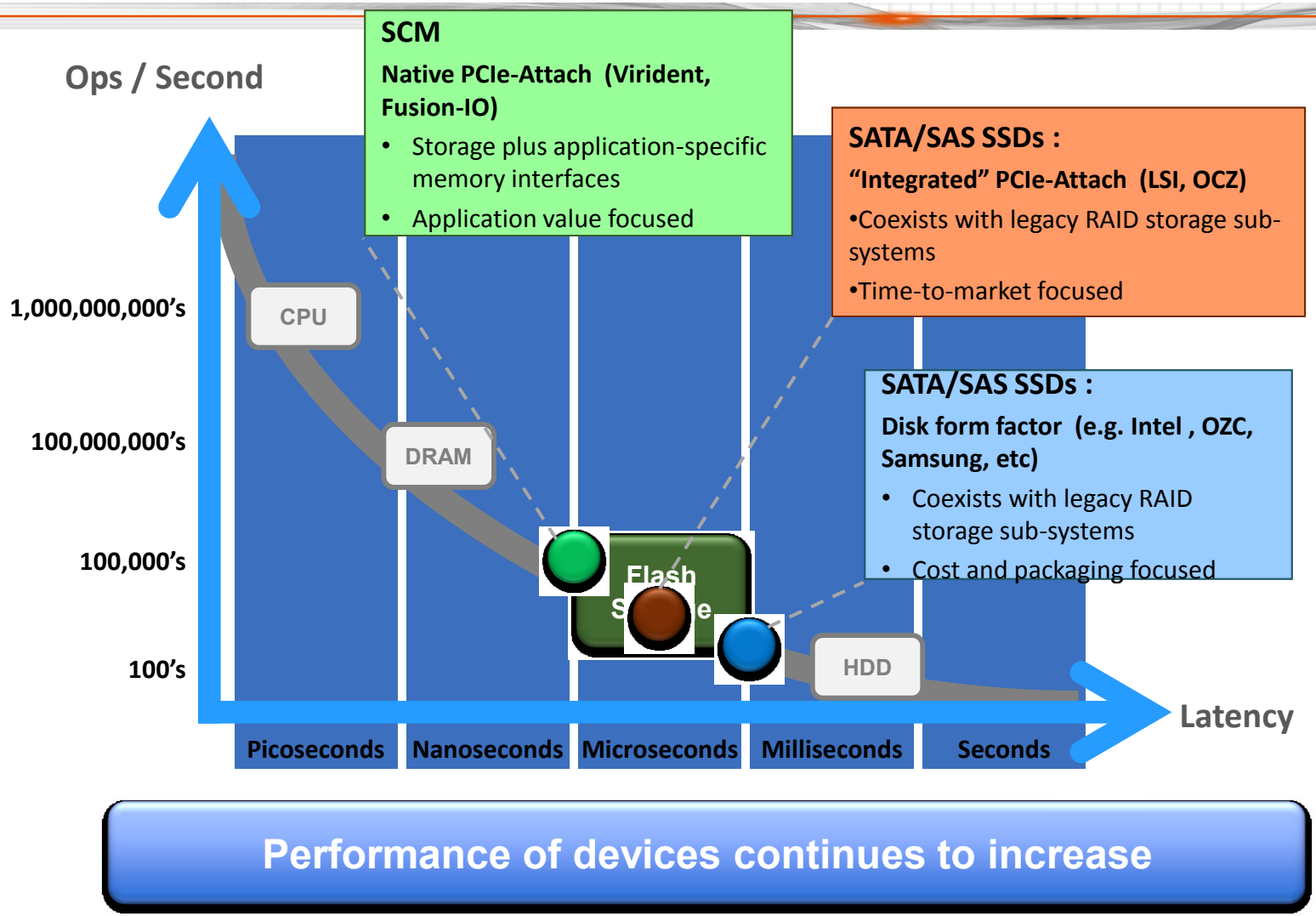
Traditional Storage Challenges



Performance gap between HDD and DDR

Flash Storage's Promise

Filling the Performance Gap

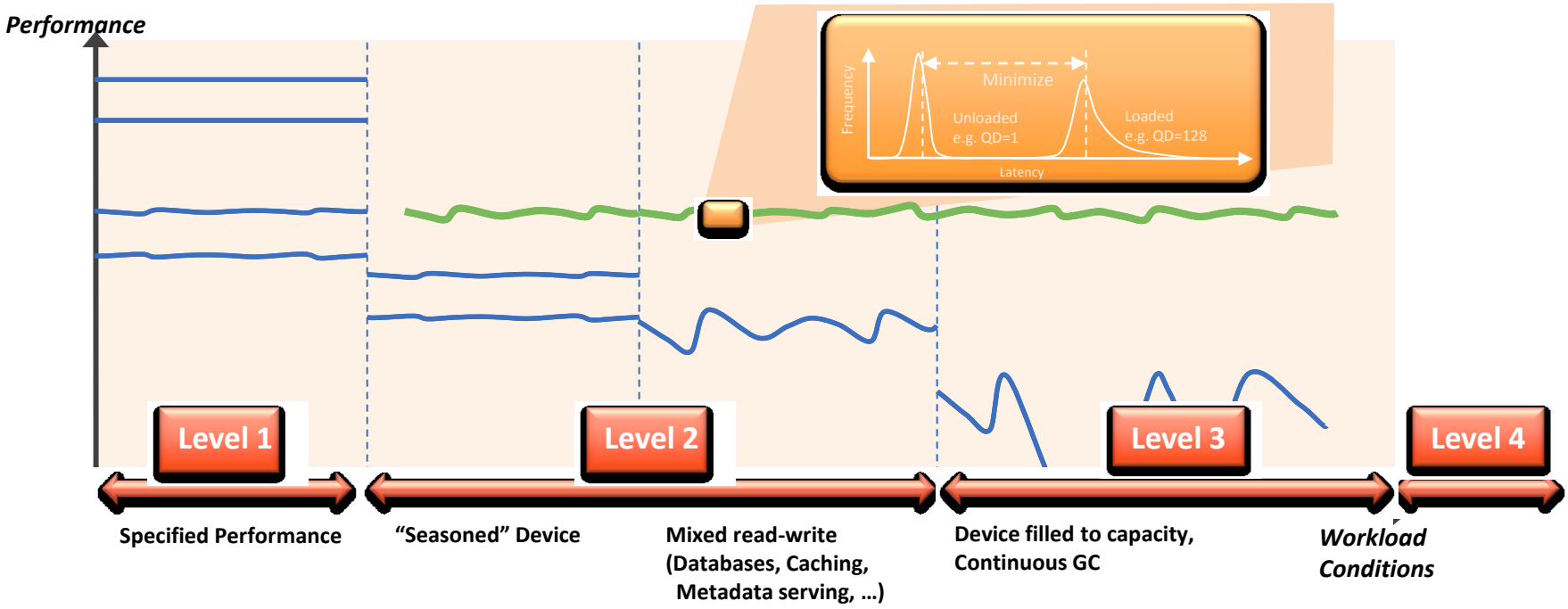


Agenda

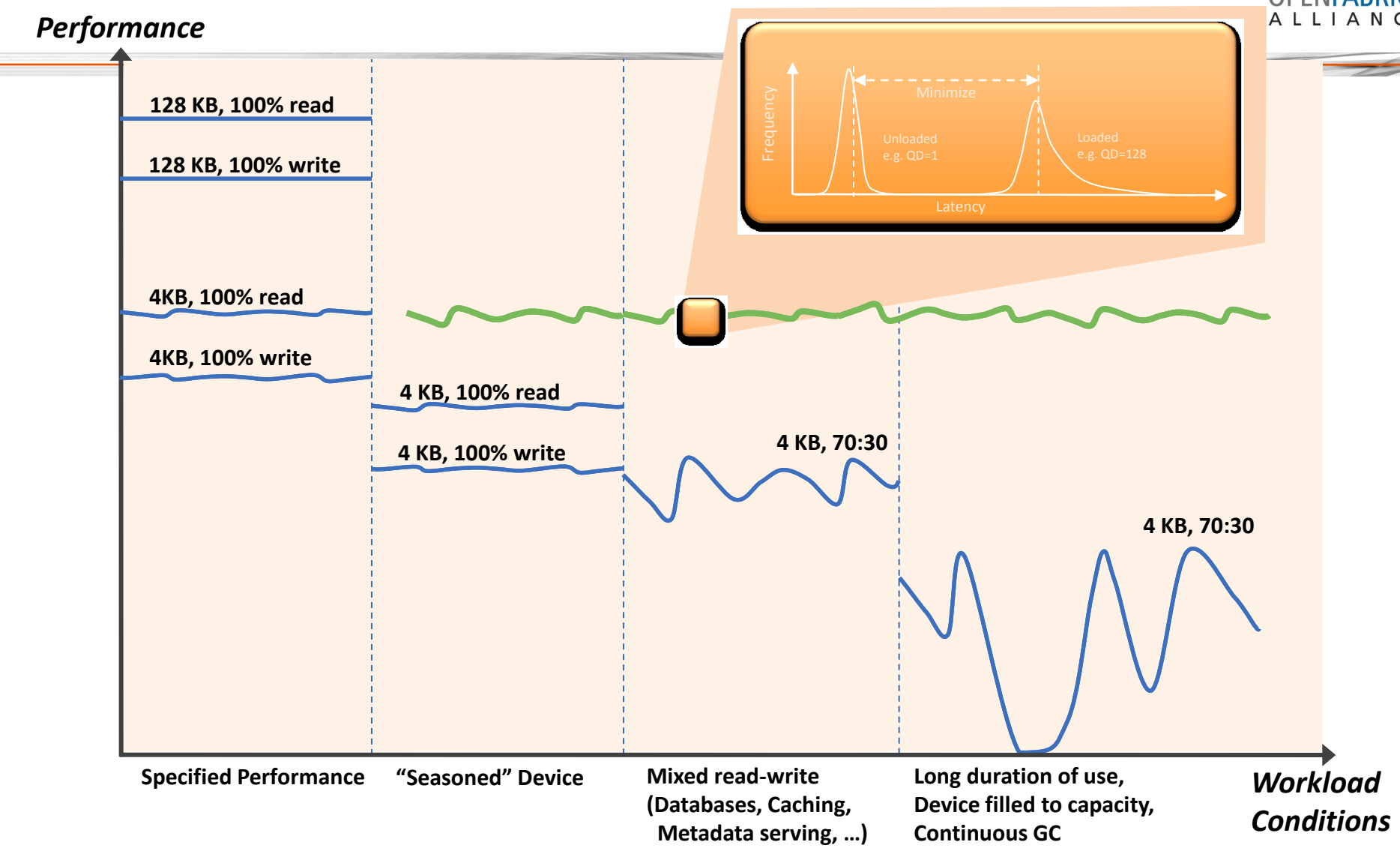
- Flash Overview
- Performance Characterization Methodology
- Technology Overview

Performance Characterization Methodology

- Level 1: Initial Performance (SNIA FOB)
- Level 2: Baseline performance (SNIA Transition)
- Level 3: Sustained performance, IO-QoS metrics (SNIA Steady State)
- Level 4: Application performance and real-world workloads

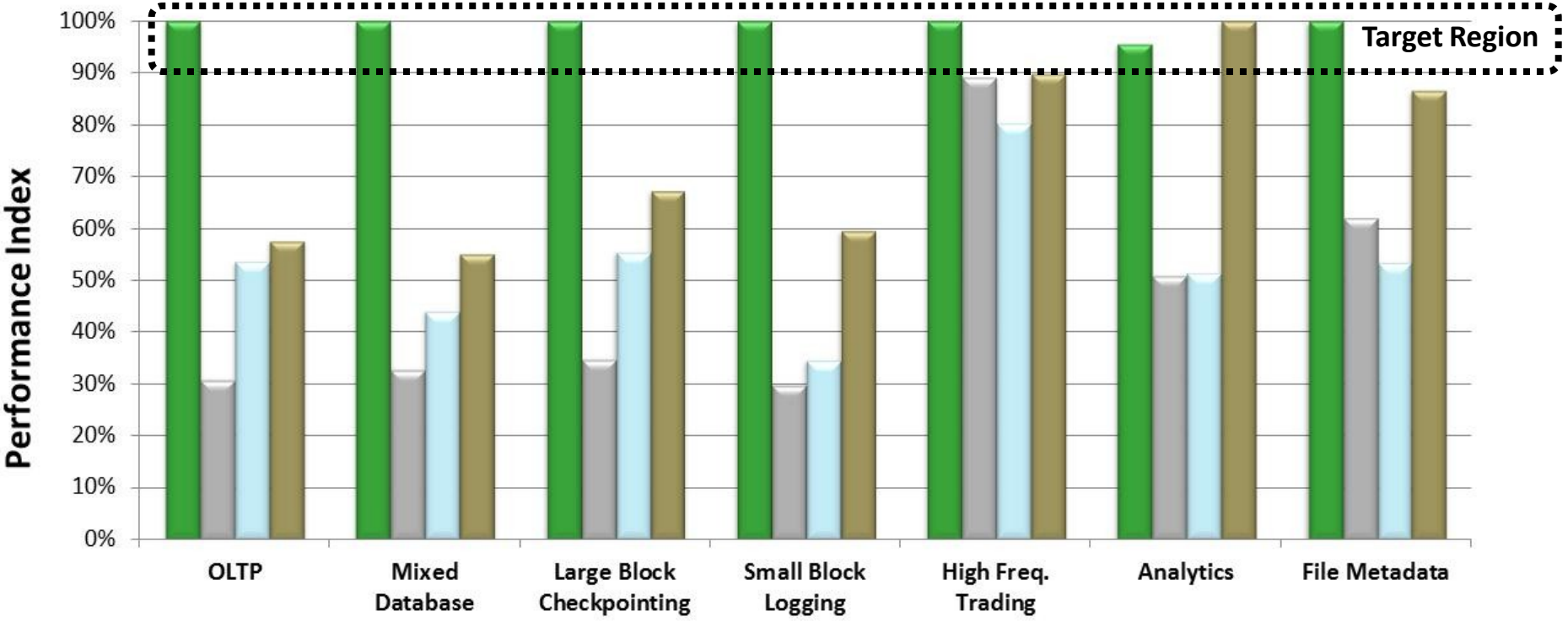


Enterprise Flash Storage: from IOPS to "IO QoS"



Goal: Highest, Sustained Performance

Across Wide Range of Applications



Example performance for sample of solid state devices

Agenda

- Flash Overview
- Performance Characterization Methodology
- Technology Overview

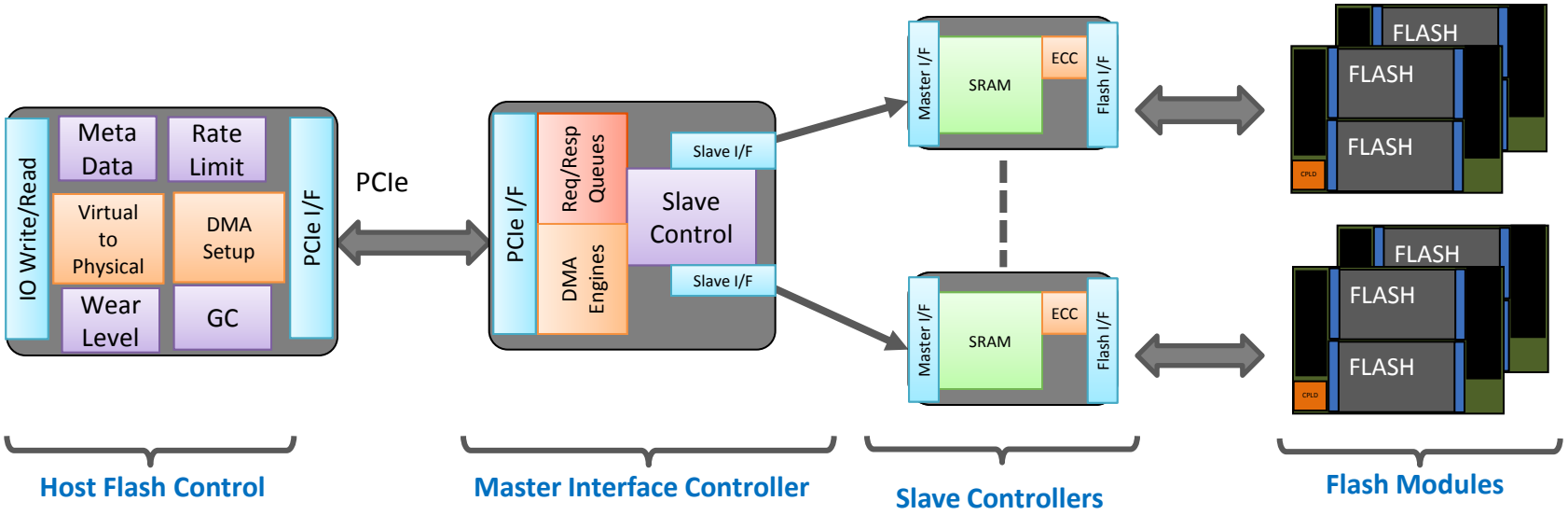
What Makes Delivering “IO QoS” Difficult?

NAND Flash characteristics:

- Asymmetry
  - Difference between read and write granularity
  - Operation latencies
- Lack of update-in-place capability
  - Needs constant balancing between user traffic and flash management
- Endurance, reliability
  - Worse with increasing capacity (or span)

Each can have an order-of-magnitude impact…

An IO-IOS flash architecture



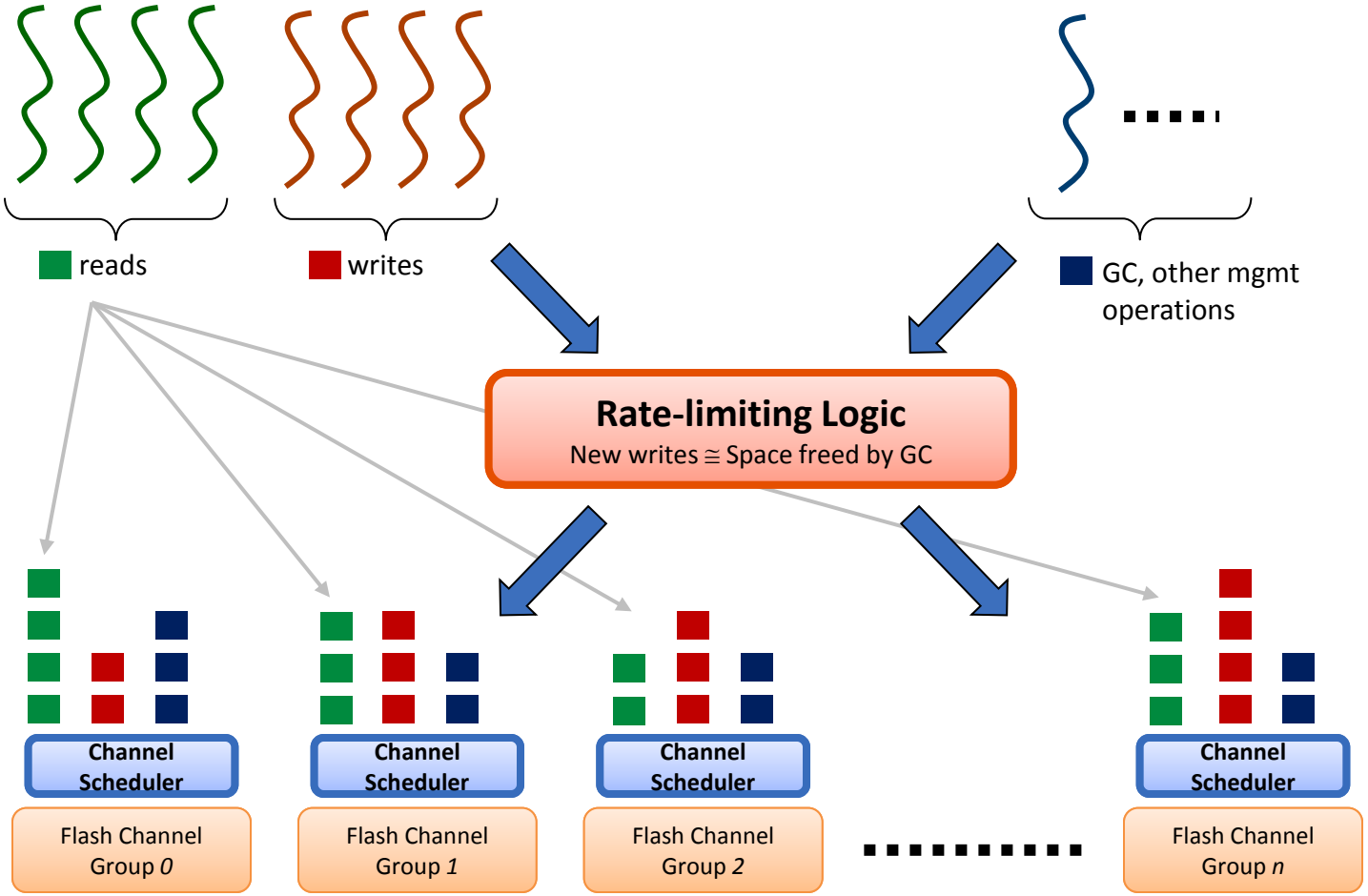
- Optimize write amp
- Advanced scheduling/rate limit
- Direct DMA to user space (no copy)
- By-Pass SCSI overhead
- Minimize host communication overhead
- Linear scalability
- Low resource overhead

- Guarantee sufficient performance during constant garbage collection
- Enable capacity scalability for more/higher density flash
- Provides two-level scheduling: across channels, within channel
- Supports Global Wear Leveling: Holistic view of full Flash capacity
- Advance ECC (e.g. BCH)

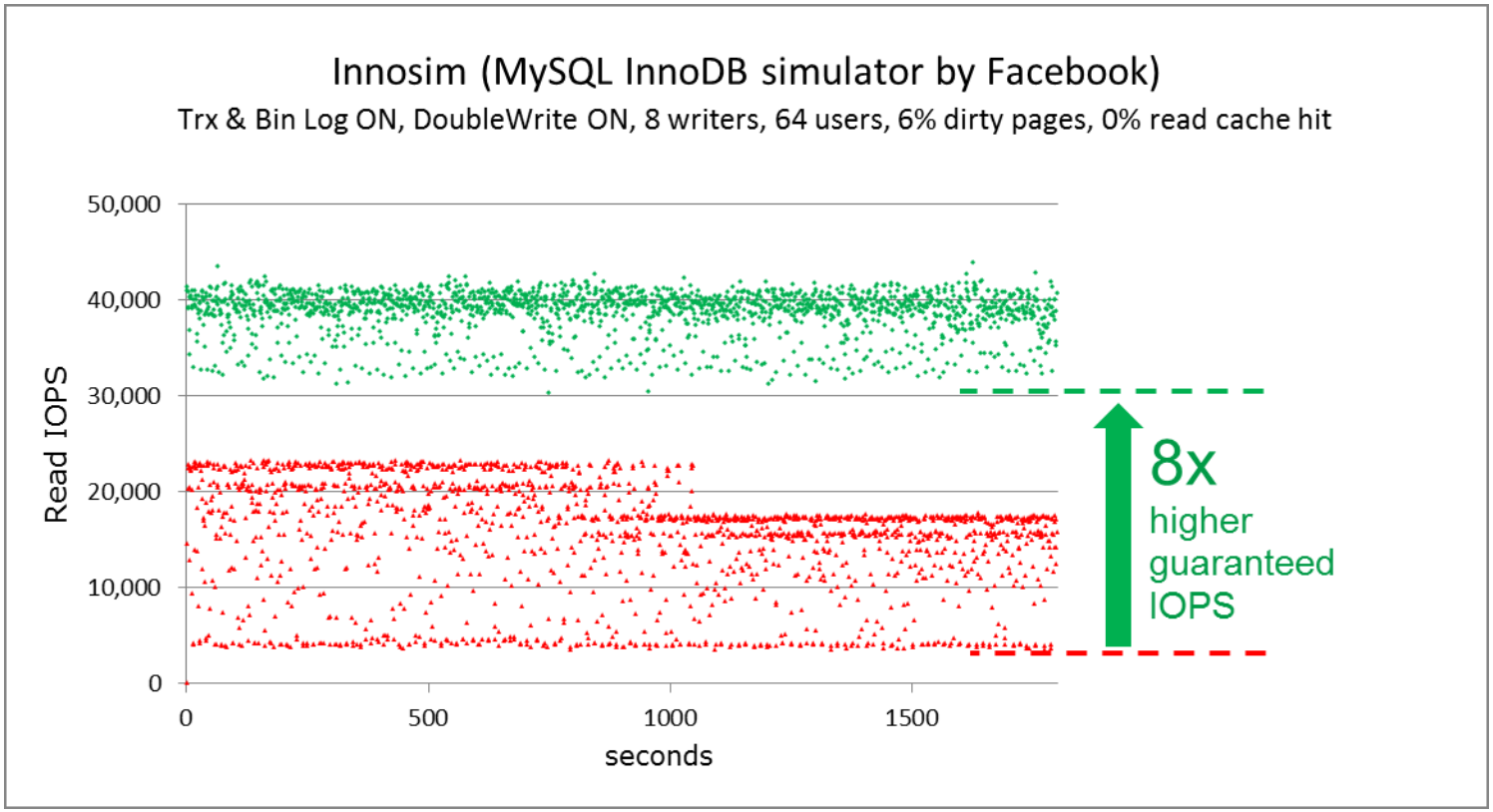
- Designed for NAND flexibility
- Resilience on module failure

Sustained Performance Mechanisms

Scheduling and Rate-Controlled Garbage Collection

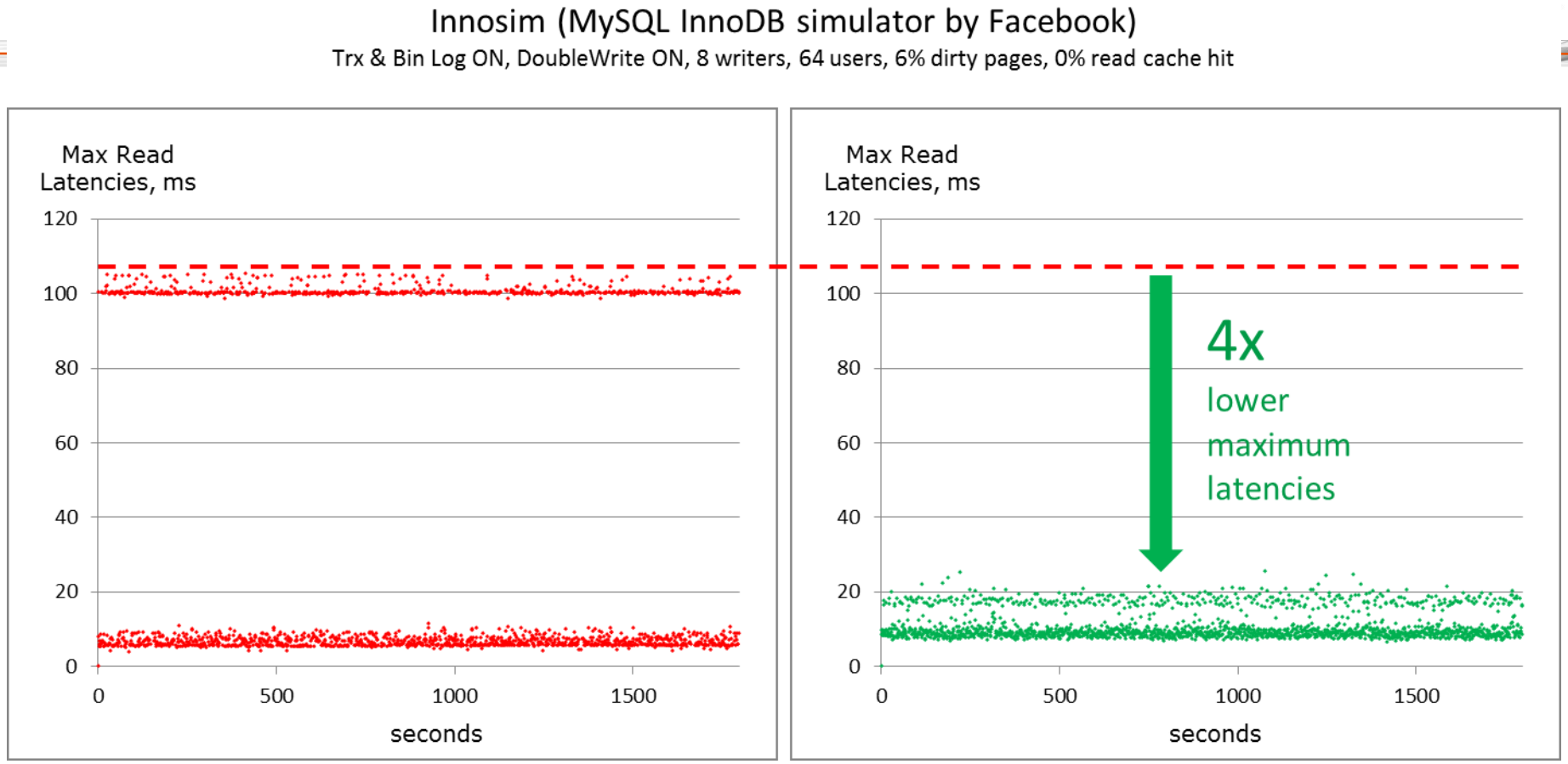


Low-level benchmark: IOPS



Test command: “run_all.sh 1 1800 0”
Innosim rev 78: http://bazaar.launchpad.net/~mdcallag/mysql-patch/mytools/files/head:/bench/innosim
CPU: 2 x Intel X5690 (6 cores @3.43GHz, HT on)
OS: CentOS 6, 2.6.32-220.el6
File system: XFS with 4KB sector size (mkfs.xfs -s size=4096). Full drive capacity pre-filled with data before creating file system.

Low level benchmark: latencies



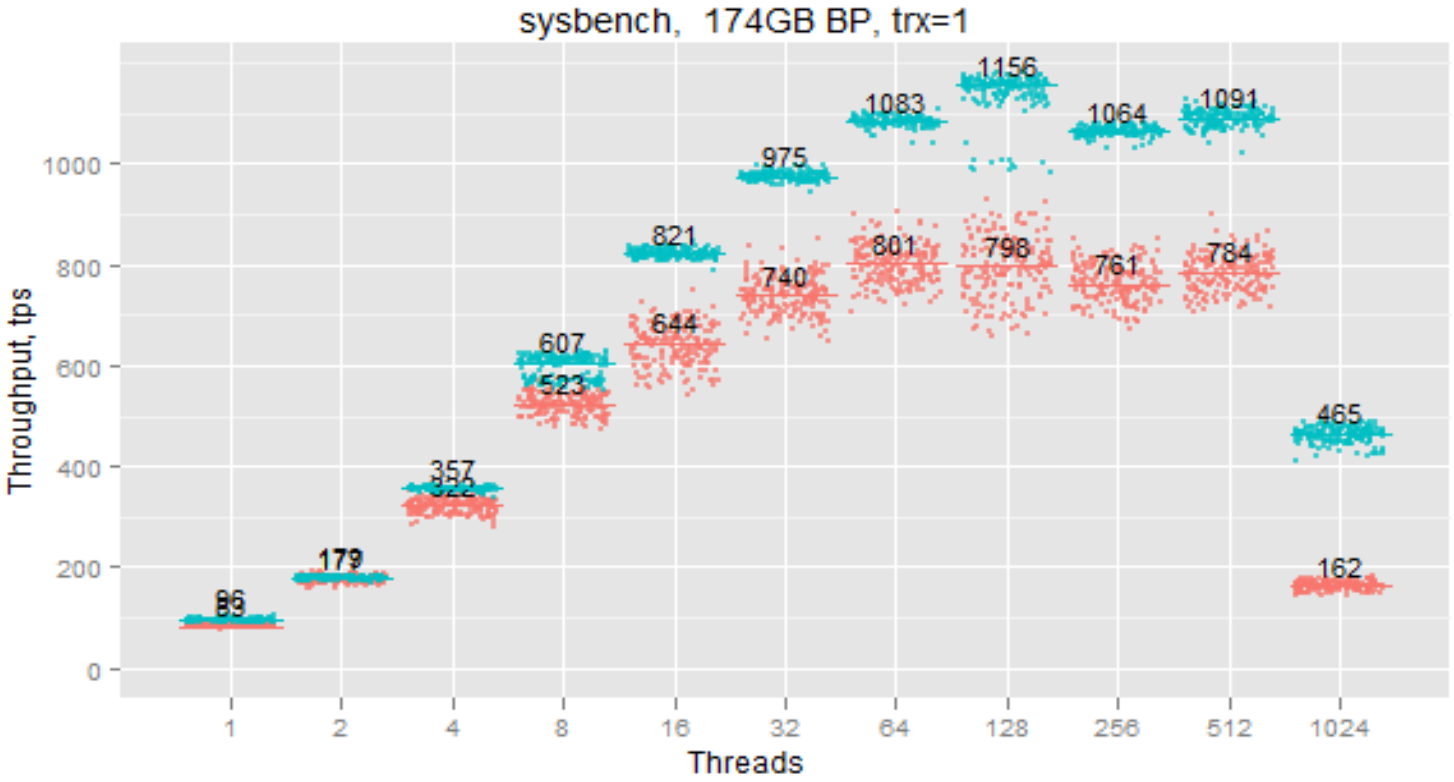
Test command: “run_all.sh 1 1800 0”
Innosim rev 78: http://bazaar.launchpad.net/~mdcallag/mysql-patch/mytools/files/head:/bench/innosim
CPU: 2 x Intel X5690 (6 cores @3.43GHz, HT on), OS: CentOS 6, 2.6.32-220.el6
File system: XFS with 4KB sector size (mkfs.xfs -s size=4096). Full drive capacity pre-filled with data before creating file system.

Application Benchmark: MySQL

Very Steady Performance, Scales with Threads

OPENFABRICS ALLIANCE

3rd party benchmarking by Percona – MySQL Sysbench



Steady High Performance Goal – Minimize Jitter

Summary

- Flash needs guaranteed high performance
- Requirements based on SNIA Steady-State and Application Benchmarks
- Architecture built to plug-n-play performance in worst case today
- Future product direction aligned with NVMe