Examining Latency of NVMe™ SSDs for Time-critical Applications Using Exceedance Analysis

John Gatch
Technologist, Data Propulsion Labs

Mark Lehrer
Technologist, Data Propulsion Labs

August 2017
SAFE HARBOR | DISCLAIMERS

Forward-Looking Statements

This presentation contains forward-looking statements that involve risks and uncertainties, including, but not limited to, statements regarding our addressable market, our product and technology positioning and compute platforms, the anticipated benefits of our new technologies, executing on our integrated strategic plans, realizing our strategic imperatives, including our solid-state drives and storage technologies. Forward-looking statements should not be read as a guarantee of future performance or results, and will not necessarily be accurate indications of the times at, or by, which such performance or results will be achieved, if at all. Forward-looking statements are subject to risks and uncertainties that could cause actual performance or results to differ materially from those expressed in or suggested by the forward-looking statements.

Additional key risks and uncertainties include the impact of continued uncertainty and volatility in global economic conditions; actions by competitors; difficulties associated with go-to-market capabilities; business conditions; growth in our markets; and pricing trends and fluctuations in average selling prices. More information about the other risks and uncertainties that could affect our business are listed in our filings with the Securities and Exchange Commission (the “SEC”) and available on the SEC’s website at www.sec.gov, including our most recently filed periodic report, to which your attention is directed. We do not undertake any obligation to publicly update or revise any forward-looking statement, whether as a result of new information, future developments or otherwise, except as otherwise required by law.
The Data Propulsion Labs (DPL)

The mission of the DPL is to design, develop, deploy, and document compelling solutions using Western Digital®, SanDisk®, and HGST-brand products

• SQL Server™ and Oracle® Testing
  – Many world records for database benchmarks have been set using Western Digital products and these tests are often done with key partners

• Application Testing
  – Multiple application benchmark suites are used, including HammerDB, Aerospike®, and various other performance test suites

• Software-Defined Storage Solutions
  – Much effort goes into understanding SDS and demonstrating its benefits, including all key players

• Ad Hoc Testing
  – New solutions are routinely tested, including Storage Class Memory, Memory Extension Technology, simulation of future devices, new data base designs, etc.

• Block Device Testing
  – The Block Performance Suite tests peak & sustained performance as well as Quality of Service
### Goals

1. Overview of Exceedance Charts and Exceedance Analysis
2. The problem statement and an answer (Exceedance Charts)
3. Several example Exceedance Charts
4. The easiest path to Exceedance Charts
5. One final issue...
This is an Exceedance Chart...

70:30 R:W QD1

- QD1 Read Exceedance
- QD1 Write Exceedance
This is an Exceedance Chart...

Key Points:
1. log-log scale.
2. Vertical axis = Number of Nines.
3. Horizontal axis = Latency.
4. Example: 90% of IOs complete in ~100us or less.
5. Closer to vertical is good.
6. Closer to horizontal is not good.
Problem Statement and an Answer
Problem Statement and an Answer

• The Problem
  – How to specify Quality of Service Requirements for high performance storage devices?
  – How to compare high performance storage devices?
  – How to diagnose performance problems in these devices?

• The Outliers
  – It is all about outliers, which can destroy QoS and make a device unusable for a given purpose.
  – Many ignore outliers, but the National Bureau of Standards says: Realistic performance parameters require the acceptance of all data that cannot be rejected for cause.

• An Answer
  – Exceedance Charts present QoS data in a very compact and understandable format that can be used for:
    • Product selection purposes
    • Comparative purposes
    • Diagnostic and performance monitoring purposes
Exceedance Chart 1

Two Curves for the same device:

BS=4K, QD=1, R/W=100/0 and 0/100.

Consumer Observations:
- Reads have better QOS.
- A few cache hits near 0 percent.
- Reads and writes shift ~20us.
- FIVE NINES < 100µs for R&W.
Horizontal movement is usually not good – you gain latency without increasing nines.

Gradual slope to the lower right is typical of many systems.

Plateaus or horizontal movement = large increase in latency with no gain in QoS.

Near vertical => increase in NINES with little increase in latency.

Controller resource contentions.
Exceedance Chart 3

- **Cache Hits provide Link estimate**
  ~448ns.

- **ECC engine is making this shape**
  and suggests BCH encoding but need more conversations with experts.

- **Near vertical behavior suggests no outer code word and no outer RAID failures at 6 NINES frequency.**

- **Assuming regular background activity, e.g., media management, Reads are impeded by the background activity.**

- **Suspect internal bottlenecks.**

- **Assuming background activity, e.g., media management.**

- **Suspect controller specific outlier such as a bug, controller background tasks (ADC check), logging, etc.**

©2017 Western Digital Corporation or its affiliates. All rights reserved. Confidential.
Exceedance Analysis Can Provide...

Rich detail and insights for storage devices

- Normal reads
- Reads blocked by other reads on the same die
- Reads blocked by programs
- Reads blocked by erases
- Reads blocked by an unknown effect

Uncertain causes. Possibly:
- 2 erases needed before write?
- HW bug?
- FW handling path where there was some GC that got in the way?

Escapes do stop though.
Methods and Procedures: The Easy Button = ezfio
Easy Exceedance Charts with ezfio

Select Drive for fio#

#fio: https://github.com/axboe/fio
ezfio produces the usual charts – IOPs, Lat & BW
ezfio produces well-formed Exceedance Charts
Contributors to the Current State of the Art

- **Earle Philhower III**, Western Digital
  - [https://github.com/earlephilhower/ezfio](https://github.com/earlephilhower/ezfio) A complete performance assessment tool including graphics of time series data, the standard performance description charts, and exceedance charts.

- **Vincent Fu**, Western Digital, [https://github.com/vincentkfu](https://github.com/vincentkfu)
  - [https://github.com/vincentkfu/fio-nanosecond-128-bit](https://github.com/vincentkfu/fio-nanosecond-128-bit) Modification of fio to support nanosecond resolution needed by today’s faster devices. This enhance reveals performance below the 1-microsecond floor.
  - Contributed to the implementation of exceedance charts in TAU, a test harness, this was foundational to the general availability of exceedance chart building blocks in fio.

- **Rebecca Cran**, Western Digital
  - Creator and overseer of the Windows port of fio: [http://bluestop.org/fio/](http://bluestop.org/fio/)

- **Dan Helmick**, Western Digital
  - Provided significant advice and technical assistance as we implemented Exceedance Charts.
One final issue ...
One Final Issue:
Exceedance Analysis is not a standalone figure of merit

• Here is a sample Exceedance Chart for TPC-C-like Transactions
• What’s missing?
One Final Issue:

*Exceedance Analysis is not a standalone figure of merit*

- Here is an Exceedance Chart for TPC-C-like transactions
- The sub chart in the lower left shows the performance of the three configurations
- The QOS and the performance, here in TPM units, provide a pretty complete picture of the alternative configurations

![Exceedance Chart](image)

~6.8X improvement
This is the end.
And, Thank You!