TECHNOLOGY BRIEF

Data Center Memory and Storage
Intel® Optane™ Technology

Achieve Consistent Low Latency for Your Storage-Intensive Workloads

Intel Optane technology SSDs deliver more reliable performance through highly predictable access latency.

Frank T. Hady, Ph.D.
Intel Fellow
Chief Optane Systems Architect

Intel Non-Volatile Memory Solutions Group

Memory and Storage Technical Series

The Direct Connection to Intel Fellows and Principal Engineer

This paper is part of a series designed to help system architects, engineers, and IT administrators understand the technological limitations of traditional memory and storage, how those limitations have led to performance and capacity gaps in the data center, and how Intel Optane technology and Intel® QLC 3D NAND technology help fill those gaps with a new industry-disrupting architecture. The series examines several topics that affect storage performance and capacity, including bandwidth, latency, queue depth, quality of service (QoS), and reliability.

We've all seen product specifications that, while useful, don't quite tell us how a product will perform for the way we intend to use it. With this in mind, open the specification for your favorite solid state drive (SSD) and look for latency specifications. All of my favorites can be found at intel.com/content/www/us/en/products/memory-storage/solid-state-drives.html. Access latency is the time required by the SSD to fetch data requested by the system. In the SSD specification, you will see a typical latency quoted. Typical latencies are generally close to the fastest times that an SSD can accept write data or return data for a read, under the best of conditions. Of course, conditions are not always best-case. Sometimes, maybe for your workloads, the SSD is servicing a heavy workload. Sometimes there is background work going on within the SSD, causing it to take longer. This paper is about those times—when the latency is longer.

We call the extent to which access latency to an SSD varies the quality of service (or QoS) of that SSD. The QoS of an SSD refers to both the typical latencies and the less-frequent, longer latencies. Sometimes these less-frequent latencies are much longer. To understand QoS, an analogy is useful. Imagine visiting a coffee shop. The latency to your cup of coffee begins the moment you walk in the door. If the place is empty, you can order your coffee and get it quickly. This is the "typical" latency the coffee shop would quote. Sometimes you will find a line—maybe then it takes three times longer because you have to wait for three other customers. You might go in once and find a dozen customers in line ordering complex drinks for themselves and their friends, and you aren't even in line yet. This almost never happens, but you certainly still care when it does. If this happens more than once, and you are impatient like me, you will be looking for a new coffee shop. For getting coffee, you care about typical wait time, how long you have to wait, and the longest you'll expect to ever have to wait. You care about the QoS of the coffee shop, not just the typical amount of time it takes to get your coffee.

My SSD is always faster than my coffee shop, but I still care about the QoS of my SSD. If you are assessing the performance of a particular SSD, for example, just looking at average latency won't tell you enough to gauge how that device will perform when you are accessing it for time-critical operations or when you have service-level agreements (SLAs) to meet. If some fraction—even a small percentage—of latencies for a data request are long outliers, they can have a major impact on the performance of your app and hurt user experience. You'll find that you care about two numbers: the fraction of accesses that take longer than typical to complete, and the time those accesses take. These two numbers can be unambiguously described together by using a statistics tool called percentiles.

Using Percentiles to More Accurately Measure Drive Latencies

Percentiles allow us to specify the percentage of accesses that will be (and have been measured to be) faster than any specific latency. Assume we operate the SSD for a given workload and record the latencies. Then we sort those latencies from smallest (fastest) to largest (slowest) and list them left (smallest) to right (largest). The position of a particular latency value in that list tells you how likely it is to occur, which can be effectively specified as a percentile. Using percentiles, we can avoid listing the percentile measurement for every possible latency—that would be too many—by picking interesting percentiles and specifying the latency. The 50th percentile (in the middle of the list) is interesting; that's another name for average. The 90th percentile (one tenth of the way from the right in the list), the 99th percentile (99th out of 100), and percentiles up to the 99.999th percentile (99,999th out of 100,000—way on the right of the list) are all interesting points that often find their way into SSD specifications. Application performance for an application will be most dependent on a particular percentile measurement, as we will show.

Here's an example of how to use QoS percentiles. Suppose you are responsible for a system hosting web sites in which each page contains numerous elements fetched in real time from storage, including customer-specific data, social-media posts, real-time status on inventory, and advertising from multiple feeds. When a user accesses the site, dozens or even hundreds of SSD accesses are made in order to gather content for the page. The time it takes to load that page is dependent on the *slowest* item to appear—in other words, the longest latency. In this particular scenario, you care about the QoS of the SSD. Typical or even average latency doesn't tell you enough. If 100 items are required to make up the page, the 99th-percentile latency is the interesting measure, because it provides a relevant latency for the longest someone would have to wait for the slowest access out of the 100 accesses required, on average. It's that longest latency that determines the time the user has to wait to see the full view. Complete SSD specifications include percentile latency measures to help you understand the QoS you should expect.

This paper applies those percentile measurement principles to Intel Optane SSDs and 3D NAND SSDs to fully describe their relative performance. These measurements show that Intel Optane SSDs deliver a *much* better QoS than NAND-based SSDs for application-usage points that really matter.

The QoS Limitations of NAND Technology

For all SSDs, the QoS story starts with the behavior of the underlying media. For NAND SSDs, typical latency is determined by the T_{read} time of the media, as it tends to be significantly longer (10s of microseconds) than the typical latencies in the rest of the SSD and system. More important is the requirement of NAND to be written only after a large block has been erased. And erases are slow. Writes are slow too, though not as slow as erases. Because erases take place in large blocks on NAND, good data within a block must be moved to a new, freshly erased block before the source block can be erased. This is a process called garbage collection. Imagine that your coffee shop has only one employee to make coffee and clean up. During clean up (garbage collection), you wait. If cleanup includes taking the accumulated trash to the landfill and buying more cups at the store, you'd have to wait a long time.

There are lots of opportunities for longer-latency reads or writes. Accesses might need to wait for the NAND chip holding the data to finish a previous read. Accesses might get caught behind multiple accesses to a given chip. Accesses might even have to wait for a write to that chip to complete. These reads and writes that are in the way might be from internal garbage collection operations within the SSD. A read could be caught behind an erase operation that has already started. Even worse, a large number of writes to an SSD could back up garbage collection, forcing the writes to wait for

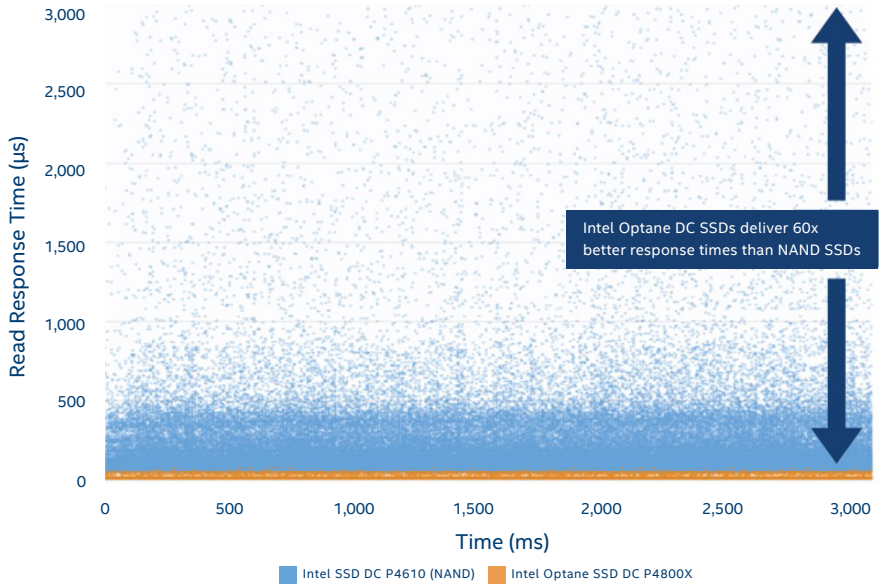

Figure 1. Intel Optane DC SSDs deliver up to 60 times better response times than NAND SSDs at 99-percent QoS[1]

many operations to complete. And, of course, the SSD is also responsible for returning highly accurate data, sometimes requiring significant error correction that might even require multiple reads.

NAND is a technology built for low-cost bulk storage, not always highest performance, and this choice is readily apparent when viewing the QoS of NAND SSDs. This is true of all NAND-based SSDs. Different vendors' NAND drives will exhibit different QoS characteristics, but all NAND drives show a relatively wide latency distribution. An example of such a distribution is shown by the blue dots in Figure 1. This figure plots individual latencies as the height of the dot, for 70-percent reads and 30-percent writes of 4 KB at the throughputs shown. Notice that, while infrequent, some latencies for this high-performance NAND SSD are as high as 3 milliseconds. We'll continue to work with this dataset, comparing it to Intel Optane SSDs and showing better ways to characterize QoS.

Intel Optane Technology Is a Game Changer for QoS

Latency levels and distributions are completely different for Intel Optane DC SSDs, compared to NAND SSDs, because they are built on a revolutionary new memory technology—Intel Optane memory media. This new type of memory features much lower latency reads and writes than NAND—orders of magnitude faster. It also allows for writes to occur in place, without an erase occurring first. Finally, it is byte-addressable, meaning that only the sector or sectors being modified must be written. No large block erases are required, so no garbage collection is required. It provides fast reads and writes, with no erases. All this results in much better QoS than NAND SSDs—orders of magnitude better.

With an Intel Optane SSD, you have walked into a coffee shop with the fastest espresso maker ever built, and several staff ready and waiting to serve you, with no need to take out the garbage. A NAND SSD is like a regular shop, sometimes it takes a long time to get your coffee. Referring to Figure 1, the orange dots represent the Intel Optane SSD latency measurements for the same workload described earlier for the NAND SSD. The orange area at the bottom isn't the x-axis; it's the tightly clustered read latencies recorded for the Intel Optane SSD DC P4800X. Remember, the blue dots in Figure 1 show individual read-response times for an Intel® 3D NAND SSD. Intel Optane DC SSDs exhibit consistently low latency and excellent QoS, with a 99th-percentile read-response time up to 60 times better than that of a high-endurance NAND SSD under the random-write workload measured.[1]

Figure 1 clearly shows the QoS advantage of the Intel Optane SSD, but you cannot read a particular percentile latency from this chart, so it cannot be used to correctly set expectations for a specific performance. For this, we need statistics to come to the rescue with a cumulative density function (CDF). Figure 2 shows the same data as Figure 1, but in two CDFs, one for each type of SSD. Here, for any given percentile (x-axis), it is possible to read the latency that should be expected (y-axis). Average latencies are about 14 microseconds (µsecs) for the Intel Optane SSD and about 200 µsecs for the NAND SSD. 90th-percentile values are still about 14 µsecs for the Intel Optane SSD, but they have increased to about 400 µsecs for the NAND SSD. And so on, up to the 99.999th percentile. With CDFs, the stark QoS advantage of the Intel Optane SSD is obvious and large.

The CDF in Figure 2 is useful, but it too is limited because it shows relative QoS for only a single input/output (I/O) operations per second (IOPS) load tested. Different applications will exercise not only in different percentile measurements but also in those different QoS measurements at different loads. To better understand the QoS advantage for Intel Optane memory media compared to NAND technology at different loads, Figure 3 graphs the latency distribution the way you typically would for memory rather than storage. Figure 3 shows a family of curves at varying percentiles (such as 99 and 99.9) for latency versus IOPS for an Intel NAND SSD and an Intel Optane DC SSD. Each load would generate a different snowflake chart and a different CDF. With a linear scale, the Intel Optane DC SSD curves would be clustered near the bottom of the y-axis, so Figure 3 instead uses a logarithmic scale for clarity.

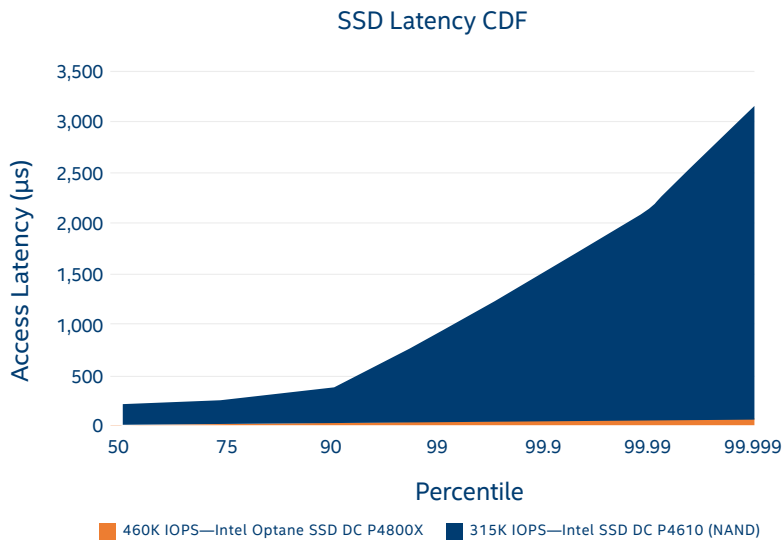

Figure 2. CDFs show that average latencies for the Intel Optane SSD DC P4800X are much lower than for the Intel SSD DC P4610 (NAND)

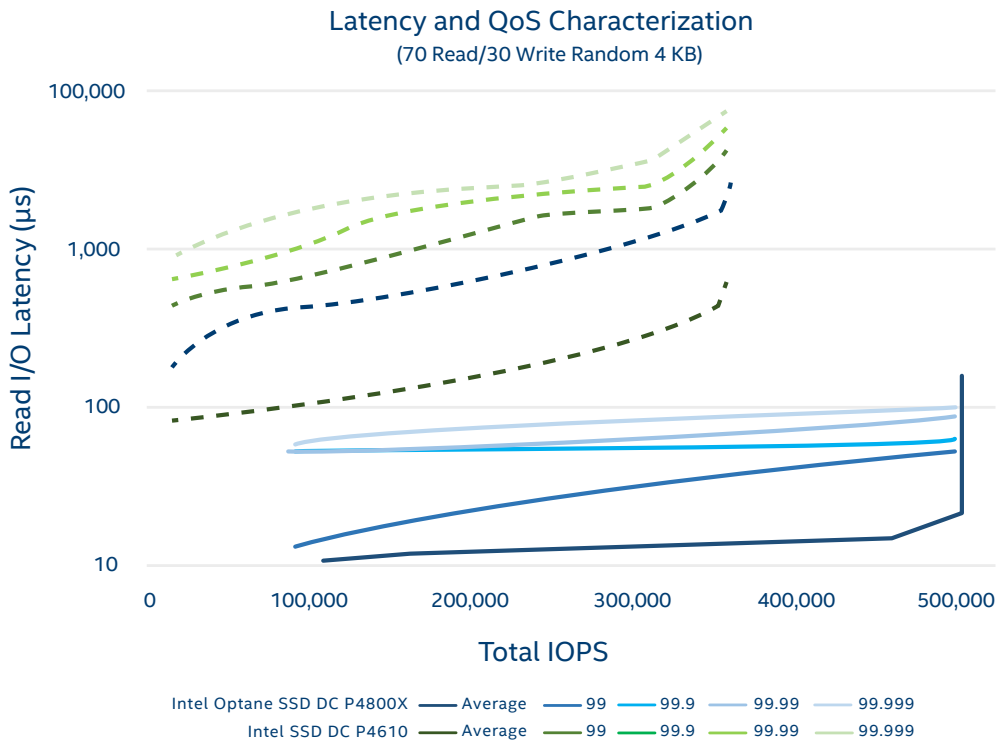

Figure 3. Intel Optane DC SSD latency distributions compared to Intel NAND SSD latency distributions[1]

Figure 3 is a useful chart for understanding and comparing SSD QoS. For each SSD (Intel NAND SSDs are in green, and Intel Optane DC SSDs are in blue), you can see a range of curves corresponding to different percentiles or levels of QoS. You can home in on the curve that represents the level of guaranteed latency needed for a particular application in your data center. For example, if you're writing an app that needs to support 300,000 IOPS, and you care about the 99.9 percentile—the case representing the slowest out of 1,000 accesses—you can see from the curves that you'll experience about 1,500 microseconds (1.5 milliseconds) of latency for the NAND SSD. That's quite different from the 77 microseconds of *average* read latency specified in the datasheet for the Intel SSD DC P4610.[2]

If you look at the 99.9 percentile for that same 300,000 IOPS load for the Intel Optane DC SSD in Figure 3, you'll see that it exhibits less than 80 microseconds of latency. So, for 300,000 read/write IOPS, we see almost a 20x QoS advantage for the Intel Optane SSD. Because of its unique memory technology and design, Intel Optane technology maintains low, consistent latency and higher QoS, even as IOPS increase. That benefit is critical for modern businesses because it enables consistent, high-level performance for large datasets and latency-sensitive workloads.

From the chart, it's obvious that the Intel Optane SSD delivers a large QoS advantage for any workload, with a larger advantage for more intense workloads and higher percentiles. It's also clear that the Intel Optane SSD delivers a more consistent performance, changing in latency less as load increases than the NAND SSD. This more ideal response makes the Intel Optane SSD easier for programmers to use, with less attention required for limiting load to maintain acceptable latency. This scenario is a reality with NAND SSDs for workloads with SLAs. Regardless of the QoS percentile you require for your applications, Intel Optane DC SSDs consistently deliver significantly lower latencies than NAND SSDs, even under load. Those lower latencies directly translate to faster, more predictable application performance for your business and can help you better meet your SLAs.

Intel Fellow Frank Hady

Frank Hady is an Intel Fellow and the Chief Optane Systems Architect in Intel's Non-Volatile Memory Solutions Group (NSG). Frank leads research and definition of Intel Optane technology products and their integration into the computing system. Frank has served as Intel's lead platform I/O architect, delivered research foundational to Intel® QuickAssist Technology, and driven significant platform performance advances. He has authored or co-authored more than 30 published papers on topics related to networking, storage, and I/O innovation and presents often on memory and storage. He holds more than 30 U.S. patents. Frank received his Bachelor's and Master's degrees in electrical engineering from the University of Virginia, and his Ph.D. in electrical engineering from the University of Maryland.

Intel Optane SSDs Benefit Real Uses

Working with customers, we've found that Intel Optane SSDs deliver differentiating solution-level benefits for a number of different applications. The consistently strong QoS of Intel Optane SSDs, coupled with their high endurance levels, enables these benefits.

Due to their excellent QoS, Intel Optane SSDs effectively provide a data caching or tiering level of storage. Data that is more frequently accessed is stored in an Intel Optane SSD, and the data that is less frequently accessed is stored in the much more capacious NAND SSDs. Often, 10 percent of the data will be stored in the fast Intel Optane SSD tier, which responds to 90 percent of the total data accesses. The remaining 10 percent of the accesses reference the colder 90 percent of the data stored in the NAND tier. Again, the Intel Optane SSDs deliver IOPS at a much higher rate and are able to do so with consistently low latency. That low latency translates directly to enabling a data store to maintain a strong 90-percent throughput rate and QoS quality from just a 10-percent Intel Optane SSD capacity.

File systems, databases, and high-reliability disaggregated stores all employ a double-write strategy to ensure that writes are not lost. The first write is to a log file that is much smaller than the actual database, and the second write is to the database itself. With this method, a loss of power at any point only disrupts one of the writes, ensuring that the database transaction can be correctly reconstructed when power is returned. This means a database write requires two writes to be completed before the operation is completed, and one (the log write) is to a much smaller store. That results in a higher IOPS rate for the SSD drive storing the log. The strong QoS of an Intel Optane SSD at a high IOPS load makes it an ideal log data store. With an Intel Optane SSD, the log file writes don't slow down database commits.

As these examples show, when used in the right place in a system, Intel Optane SSDs provide high QoS under heavy load, which can return strong benefits in important, real-world use cases.

Max Latency Values Aren't Noise, They're Data

If you're trying to measure real-world performance for your applications, average latencies are only a starting point. Occasional spikes in read-response times can make the difference between acceptable performance and an unhappy customer.

With their high capacity at lower cost and significant advantages over hard disk drives (HDDs), NAND SSDs will continue to play a crucial role in the data center as affordable capacity storage of less-frequently-accessed data. But for time-critical operations demanding excellent QoS, Intel Optane DC SSDs deliver a far superior solution.

Intel Optane technology offers a revolutionary approach to delivering consistently high QoS in the data center by providing tightly controlled, consistently low latencies even for intense SSD workloads.

Learn More

Learn more about how Intel Optane technology is disrupting the memory and storage hierarchy in the data center by exploring other papers in the **Memory and Storage Technical Series**.

To learn about Intel Optane DC persistent memory, visit: **intel.com/content/www/us/en/architecture-and-technology/optane-dc-persistent-memory.html**

To learn more about Intel Optane DC SSDs, visit: **intel.com/content/www/us/en/products/memory-storage/solid-state-drives/data-center-ssds/optane-dc-ssd-series.html**

[1] Based on Intel testing as of November 15, 2018: Response time refers to average read latency measured at queue depth 1 during 4K random write workload using FIO 3.1. Configuration: 4K 70/30 read/write performance at low queue depth (QD). Measured using FIO 3.1. Common configuration: Intel® 2U Server System, CentOS 7.5, kernel 4.17.6-1.el7.x86_64, 2 x Intel® Xeon® Gold 6154 processor at 3.0 GHz (18 cores), 256 GB DDR4 RAM at 2,666 MHz. Configuration: 375 GB Intel Optane SSD DC P4800X compared to 1.6 TB Intel SSD DC P4610. Intel microcode: 0x2000043; system BIOS: 00.01.0013; Intel® Management Engine (Intel® ME) firmware: 04.00.04.294; baseboard management controller (BMC) firmware: 1.43.91f76955; FRUSDR: 1.43. The benchmark results may need to be revised as additional testing is conducted.

[2] Intel. "Intel® SSD DC P4610 Series." https://ark.intel.com/content/www/us/en/ark/products/140103/intel-ssd-dc-p4610-series-1-6tb-2-5in-pcie-3-1-x4-3d2-tlc.html.

Performance results are based on testing as of the date set forth in the configurations and may not reflect all publicly available security updates. See configuration disclosure for details. No product or component can be absolutely secure.

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit intel.com/benchmarks.

Intel technologies may require enabled hardware, specific software, or services activation. Check with your system manufacturer or retailer.

Printed in USA 1219/JR/PRW/PDF Please Recycle 341929-001US