SSD Performance, Latency, And A Better Experience

SSD Performance, Latency, And A Better Experience

There are many meaningful ways to characterize the performance of a storage device. Bandwidth is one of the most common metrics, defining the amount of data—sustained or up to some peak value—a drive can transfer in a given time interval. Typically, that's conveyed in megabytes per second. A second specification, I/O operations per second (IOPS), is usually reported right next to bandwidth, giving you some indication of the rate at which a subsystem can service, read, or write requests.

Both figures are important in different contexts. If you're a video editor working with large files, big throughput numbers mean you'll spend less time waiting for high-def content to make its way through your storage hierarchy. Databases, on the other hand, are sensitive to a drive's ability to field many small transfers, shifting the emphasis to IOPS. No two workloads are the same though, so even when competing products are compared on an even playing field, the parameters used for testing may not map over to the application that's important to you. The unfortunate reality is that storage benchmarks are easy to manipulate and even easier to misinterpret.

A third measurement—latency—can help give those other two performance specifications more relevance. Latency is the time it takes after issuing a request to receive a response. And as you might imagine, drives optimized for very low latency tend to offer a better user experience. They "feel" faster, since less time is spent waiting. Conversely, high or inconsistent latency manifests as stalls and hitches. In I/O-heavy workloads, even brief delays can compound as storage operations pile up behind each other.

Latency In The Enterprise

In most countries, enterprise-class businesses are defined as those with more than 1000 employees. Such large organizations commonly have very diverse storage requirements. Databases, online transaction processing, cloud-based services, and archiving might all exist under one roof. To varying degrees, each application interacts with storage differently. Some operations will be inherently random, while others are accessed sequentially. Further, transfer size affects the speed at which data can be read or written. Bandwidth and IOPS figures attempt to capture those behaviors.

What they fail to tell you—particularly as the workload becomes more random in nature—is how long operations take to service. As a result, you miss out on perhaps the most important measure of a storage device's suitability to enterprise applications (other than reliability): performance consistency, or the latency of I/Os over time.

Naturally, you want a storage device able to deliver low *average* latency, translating to a snappy, responsive experience. Even more important to many customers, however, is low *maximum* latency. This affects the perceived quality of service, since worst-case delays tend to wipe out the benefits of lower averages when the going gets tough. Achieving high marks in both disciplines, along with big bandwidth numbers and high IOPS throughput, requires a powerful storage processor and meticulous firmware optimization to eradicate overhead wherever possible.

AHCI Vs. NVMe: Balancing Performance and Compatibility

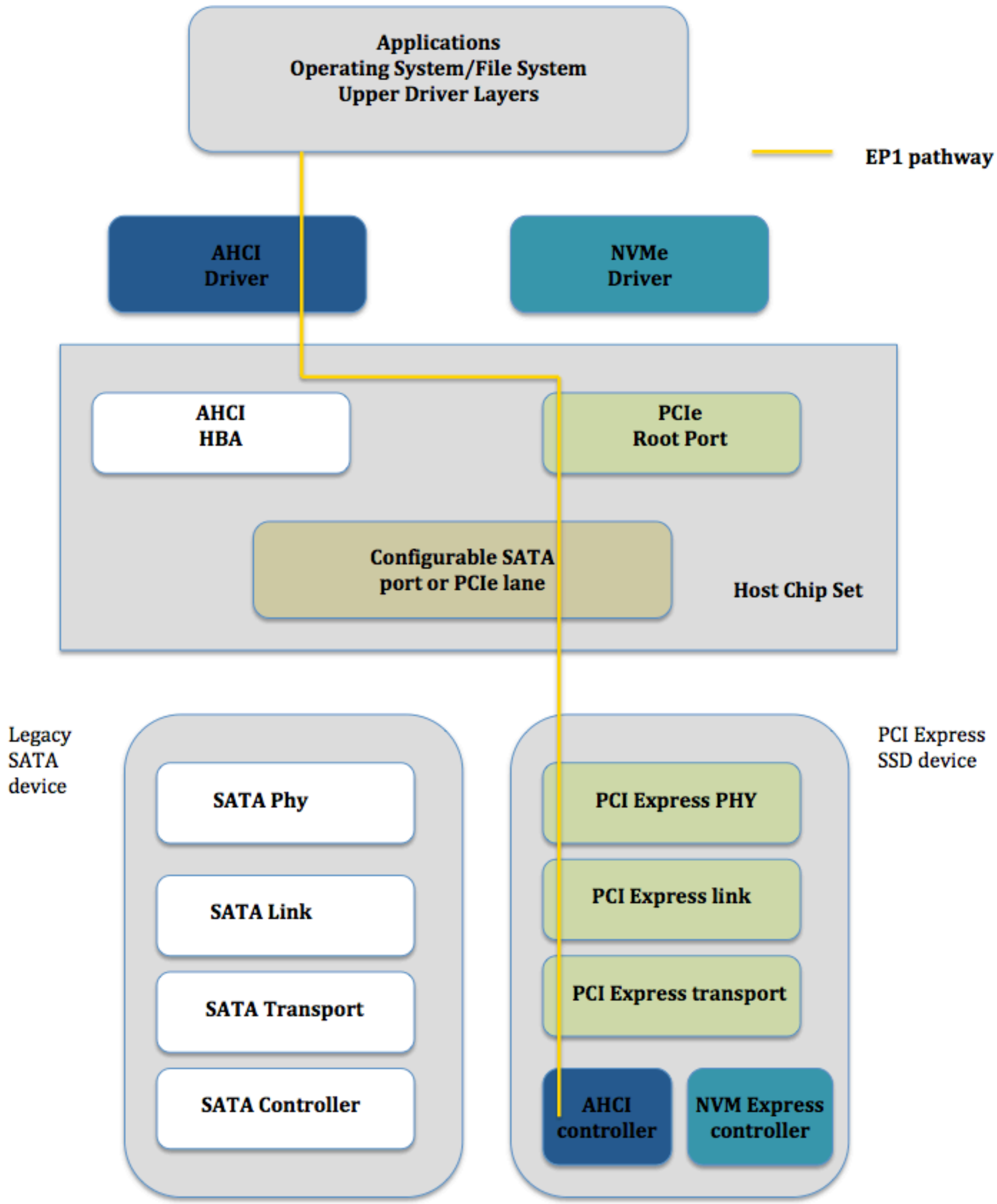
Of course, there's always going to be some amount of latency in the storage pipeline. Beyond a device's internal architecture, where a controller talks to flash memory over a number of channels with a DRAM cache buffering information on the fly, communication out across an interface takes time too.

As the Serial ATA bus evolved, the Advanced Host Controller Interface (AHCI) surfaced as a bridge between SATA-based drives and system memory. The disks themselves communicated over SATA to a host-bus adapter, which attached to the platform via PCI Express. AHCI-enabled HBAs also added support for hot plug functionality, power management, and, most significant, native command queuing. Translation services took place on both sides of the adapter, adding small but acceptable delays. In a world of mechanical storage, these were almost negligible compared to latency across the rest of the access path.

Solid-state drives challenged the need for an aggregation point to manage that disparity between fast RAM and slow mechanical disks, which is the role AHCI served. Given significantly higher performance and their low-latency characteristics, SSDs didn't require the same considerations as hard drives. In fact, the peak sequential throughput of most modern SSDs is limited by SATA's 6Gb/s physical layer. Rather than attaching to a SATA interface and communicating through AHCI, they could operate more efficiently on the PCI Express bus.

The mechanism for facilitating such a transition has been in the works for years. Dubbed the Non-Volatile Memory Host Controller Interface specification (more conventionally referred to as NVM Express), it standardizes the means by which storage devices communicate over PCIe. Like SSDs themselves, NVMe was designed with parallelism and scalability in mind. Nowhere will NVMe's impact be felt more profoundly than the enterprise, where demanding workloads create deep command queues and tax a storage subsystem's ability to churn through I/O operations in a consistent manner.

Most modern operating systems already incorporate some degree of NVMe support. The technology is still far from ubiquitous, though. As an example, in order to boot from an NVMe-based SSD, you need a motherboard with compatible UEFI software. Board vendors are slowly updating current-generation platforms. However, products built on older chipsets may never receive the requisite support. NVMe will undoubtedly mature over time. But there remains a lot of qualification work to demonstrate its reliability in the enterprise.



SATA Express NVMe Device Interface Architecture

For now, AHCI enjoys the benefit of broad software compatibility. Its protocols had to work with all of the applications written with IDE in mind. As a result, AHCI-based HBAs can either emulate that legacy interface or expose the more enhanced capabilities of its specifications. In contrast, the NVMe architecture began with a clean slate. It was written to exploit modern flash-based devices and is unquestionably the future of solid-state storage.

Lite-On's EP1-Series SSDs: The Best Of Both Worlds

However, NVMe's newness leaves plenty of room for transitional products that utilize an AHCI driver, yet attach directly to PCI Express. This naturally confers the compatibility advantages of AHCI while circumventing the performance limitations of legacy SATA, since a multi-lane PCIe link has more available throughput. We already know that a wider pipe isn't the sole enabler of a high-performance storage device, though. Minimizing latency is critical as well.

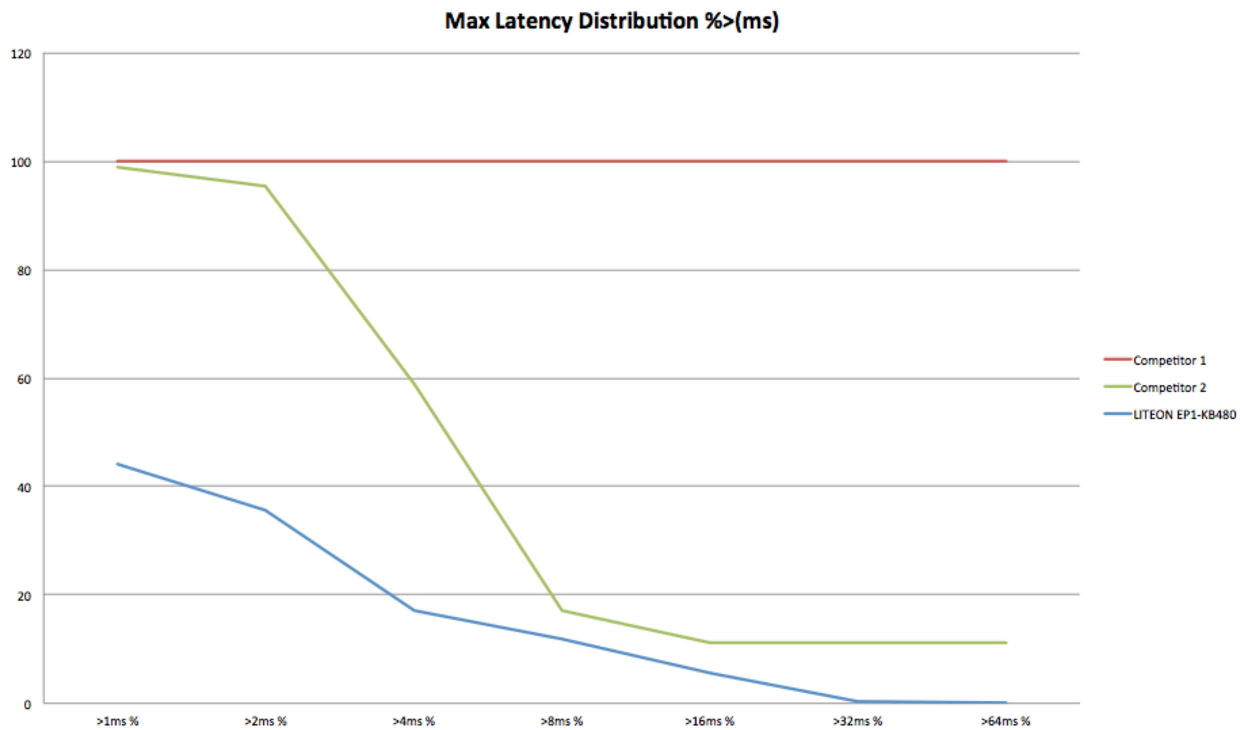
EP1

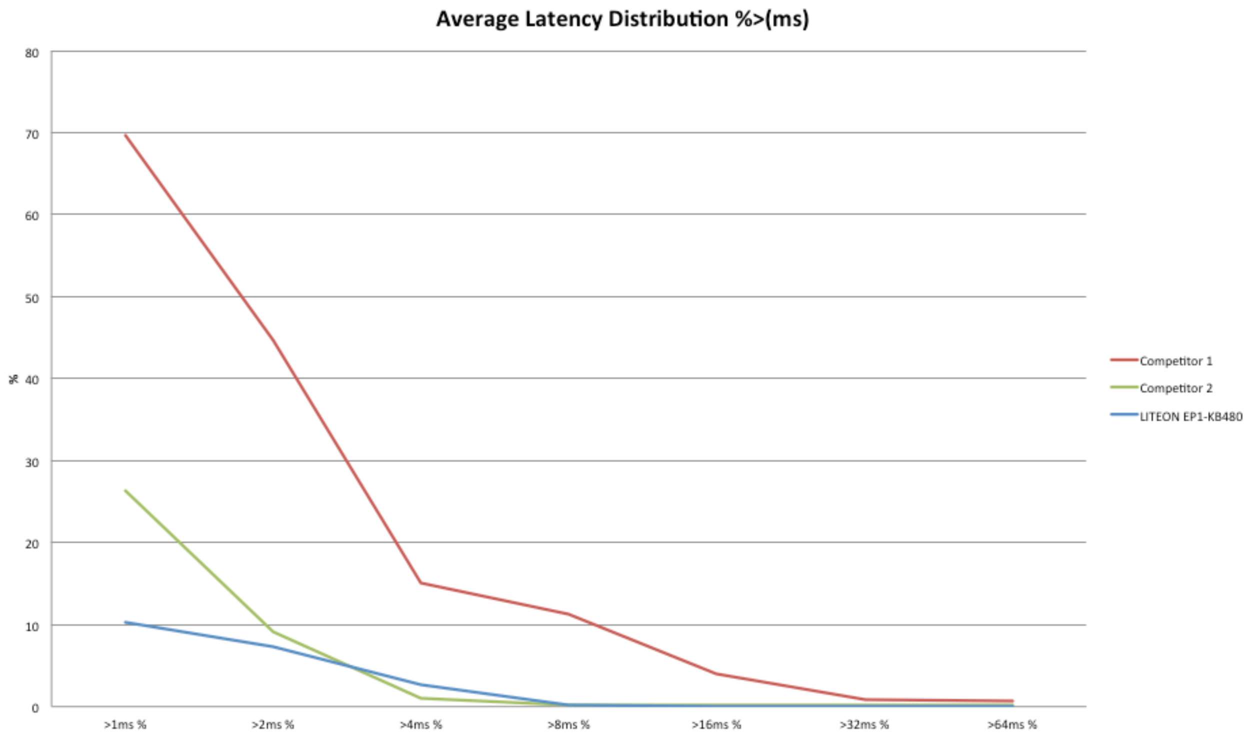
Lite-On's EP1-series SSDs are based on the M.2 form factor, which creates a four-lane PCIe 2.0 link that far exceeds the peak bandwidth imposed by SATA 6Gb/s. This allows the drive's Marvell 88SS9293 controller to operate unconstrained—a good thing since its two Cortex-R4 processing cores have plenty of horsepower in reserve. Consequently, Lite-On is able to specify the EP1 in its highest-power trim at up to 1100 MB/s sequential reads and up to 900 MB/s sequential writes. The company also publishes compelling random read and write figures, citing 4KB reads at up to 110,000 IOPS and writes at up to 25,000 IOPS using a queue depth of 32.

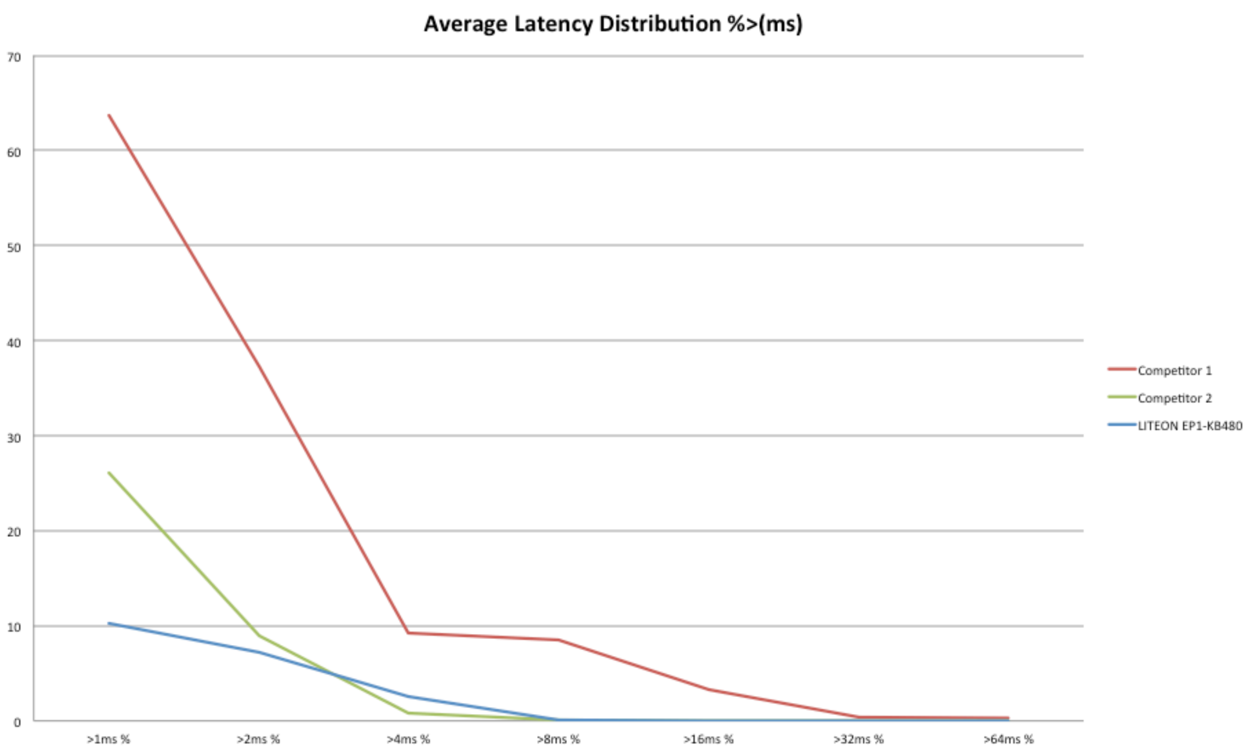
Lite-On even offers a performance consistency specification—that measure of I/O latency over time—of 90% for 4KB reads and 85% for 4KB writes. This is a potentially confusing figure since it's only just gaining popularity and is still somewhat obscure. Simply, the percentage is measured as IOPS in the 99.9th percentile slowest one-second interval divided by average IOPS during the test. As you get closer to 100%, a drive's least-favorable behavior starts looking like the average across the run, serving as an indicator of "evenness." Quality of service—defined as the time it takes for 99.9% of commands to travel round-trip from the host to the drive and back—more accurately reflects maximum latency. Again, measuring 4KB transfers at a queue depth of 32, Lite-On specifies its 480GB EP1 at 4ms for reads and 20ms for writes.

Aerospike Certification latency measure test results

	data is act version 3														
		trans %>(ms)							Device %>(ms)						
		1	2	4	8	16	32	64	1	2	4	8	16	32	64
		>1ms %	>2ms %	>4ms %	>8ms %	>16ms %	>32ms %	>64ms %	>1ms %	>2ms %	>4ms %	>8ms %	>16ms %	>32ms %	>64ms %
LITEON EP1-KB480	avg	10.31	7.31	2.62	0.12	0.01	0	0	10.25	7.23	2.58	0.11	0	0	0
	Max	44.29	35.82	20.42	19.5	17.2	12.57	3.87	44.06	35.54	17.13	11.85	5.65	0.42	0.03
Competitor 1	avg	69.67	44.75	15.07	11.3	3.94	0.82	0.66	63.63	37.3	9.28	8.47	3.35	0.35	0.3
	max	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Competitor 2	avg	26.28	9.16	1.04	0.26	0.21	0.19	0.17	26.12	8.96	0.88	0.06	0.02	0.01	0.01
	max	100	100	100	100	100	100	100	98.94	95.33	58.86	17.11	11.18	11.18	11.18







Optimizing for what can only be described as enterprise-class performance consistency and quality of service necessitates specialized expertise. Sometimes this needlework sacrifices peak throughput in the interest of a better user experience. But for IT professionals who understand the importance of minimizing latency, this is not only acceptable, but in fact, welcome.

Notably, not all controller vendors give their drive partners firmware-level access. And not all manufacturers have the resources to go under a controller's hood and tune its firmware for lower latency. Lite-On's EP1 does receive this treatment though, and the result is a big advantage over other AHCI-based SSDs attached directly to the PCI Express bus.

Quantifying a drive's responsiveness isn't as straightforward as pushing big files through it to measure sequential throughput. To get a truer sense of the EP1's competitive position, Lite-On ran a series of benchmarks using the Aerospike Certification Tool. The open source test measures latency under a load of large-block writes, returning a percentage of read operations that take longer than a set time period to complete. Ideally, you want those figures to be as low as possible, indicating a short interval between the host's request and data returned by the SSD.

According to Lite-On's results, just over 10% of requests take longer than one millisecond for the EP1 to service. The nearest competitor sees more than 25% of its requests exceed one millisecond. That's a massive lead in average latency. But it grows if the focus shifts to maximum latency—those worst-case delays you want to avoid at all costs. There, the EP1 is in a class unto itself. Take the total transaction round-trip and carve out just the time spent in the SSD itself. Lite-On can get the maximum percentage of read requests that take longer than 64 ms to complete down to just .03%. Its nearest competitor sits above 11% in the same metric, while another popular drive has to endure 100% of its requests exceeding 64 ms when maximum latency is isolated.

Just The Right Mix

At any given performance level, an SSD tuned for low latency is going to feel faster and more responsive. This is still a relatively new discipline in flash-based storage evaluations though, so it often goes underreported. Fortunately, the benefits of minimizing average latency and, even more important in many applications, maximum latency are becoming apparent as real-time workloads evolve. SSD vendors able to really tune a processor's firmware for consistency are earning recognition for delivering a smoother user experience, while those still focused on sequential throughput and small-block I/O struggle with latency measurements.

In the years to come, a transition to PCI Express-attached storage communicating through NVMe will allow us to better-utilize the parallel nature of SSD architectures. Until that interface matures, though, AHCI offers the broad software compatibility required by large businesses. Lite-On's EP1 simultaneously delivers the ubiquity of AHCI, performance unconstrained by SATA (thanks to its four-lane PCIe link), and class-leading response times enabled by the company's experience tuning firmware. Low power consumption and data loss protection only serve to strengthen the drive's enterprise-class pedigree.

LITEON

http://www.liteonssd.com/

America

Address: 42000 Christy Street, Fremont, CA 94538
Tel: 510-824-0066
Fax: 510-687-1900
E-mail: SSD.USSupport@liteon.com

Taiwan & ROW

Address:16F, 392, Ruey Kuang Road, Neihu, Taipei 114, Taiwan, R.O.C
Tel:+886-2-8798-2886
Fax:+886-2-8798-2825
E-mail:SSD@liteon.com

Industrial Application

Address:16F, 392, Ruey Kuang Road, Neihu, Taipei 114, Taiwan, R.O.C
Tel:+886-2-8798-2886
Fax:+886-2-8798-2825
E-mail:Sales_IPC@liteon.com