HL NAND

ENABLING
Ultra-High Bandwidth
Scalable SSDs
with HLnand

HL NAND

INTRODUCTION

Solid State Drives (SSDs) are available in a wide range of capacities, employing a variety of system interfaces and form factors to satisfy different applications. For example, in the consumer PC segment, SSDs based on the 2.5" hard disk form factor and SATA 6Gbps interconnect, delivering up to 500 GB capacity, are dominant. For enterprise server applications a standard plug-in PCIe card with 8 or 16-lanes, and providing up to 4TB, is becoming the de-facto standard. High end rack-mounted memory appliances use multiple Fibre Channel, Gb Ethernet, or Infiniband channels with up to 32TB capacity. The interconnects employed in the SSD market range in bandwidth from 500MB/s for SATA to 40GB/s for 8 QDR Infiniband channels. Within the SSD large numbers of individual NAND flash memory devices must be connected in an efficient and scalable manner to achieve the wide range of performance and capacity for the different applications.

Figure 1. Evolution of System Interconnects

Conventional asynchronous NAND Flash interfaces are based on a parallel bus structure and are capable of operating at transfer rates of just 33 – 50MB/s, with only a few devices populated on each memory channel. This limitation necessitates SSD designs with a large number of channels in order to match the media throughput to the throughput of even older system interconnects. For example, eight to 10 flash channels of are commonly used for SATA consumer SSDs. In the enterprise server space, where PCIe is often used to connect storage hardware, SSDs require as many as 25 to 50 channels to provide the throughput demanded by the system interface. The need to increase capacity is another significant problem resulting in the proliferation of memory channels. With a limit of about eight loads on a parallel bus before channel bandwidth begins to roll off, it becomes necessary to increase the number of memory channels in order to achieve increased drive capacity. Advances in Flash interface design with respect to speed and capacity have, for the most part, been modest and lag both the growth in system interconnect throughput and the trend toward larger storage capacity.

EVOLUTIONARY FLASH INTERFACES SOLVE FEW PROBLEMS

To take advantage of the growing bandwidth of system interconnects, SSD developers are under pressure to upgrade the underlying Flash devices to newer varieties equipped with faster interfaces. However, second-generation Flash interfaces such as ONFi and toggle mode are not up to the job. As channel bandwidth is increased, fewer chips can be populated on a single channel before the channel speed must be reduced. Therefore, the higher the capacity of the drive, the more memory channels are required.

Developing high capacity SSDs without dramatically increasing the number of memory channels and, hence, the complexity, requires a memory interface that is more scalable. Likewise, as system interconnect throughput continues to grow, designers require Flash devices with an interface that can accommodate correspondingly higher speed operation without loading induced roll-off. Both characteristics are essential to producing high performance, high capacity SSDs.

HLNAND DELIVERS HIGH SCALABILITY AND THROUGHPUT MATCHED TO FASTER SYSTEM INTERCONNECT

HLNAND™ (HyperLink NAND) is the only new Flash memory interface equipped to deal with the issue of reducing the complexity of the SSD design as interconnect speeds increase, while simultaneously increasing the capacity. Thus, HLNAND is highly scalable offering developers a simpler approach to designing a wide variety of storage applications without the scaling problems inherent with parallel busses.

HLNAND with the HyperLink interface utilizes a serial, point-to-point, daisy-chain topology to connect up to 255 memory devices in a single memory channel, called a 'Ring'. Because each device is only connected to the next device in the Ring, it is driving just one load. Therefore, maximum operational speed is maintained regardless of the number of devices populated in the Ring.

HyperLink NAND, shown in Figure 2, has an 8-bit, synchronous, DDR data bus that operates with a 400MHz source-synchronous clock to deliver up to 800MB/s of throughput.
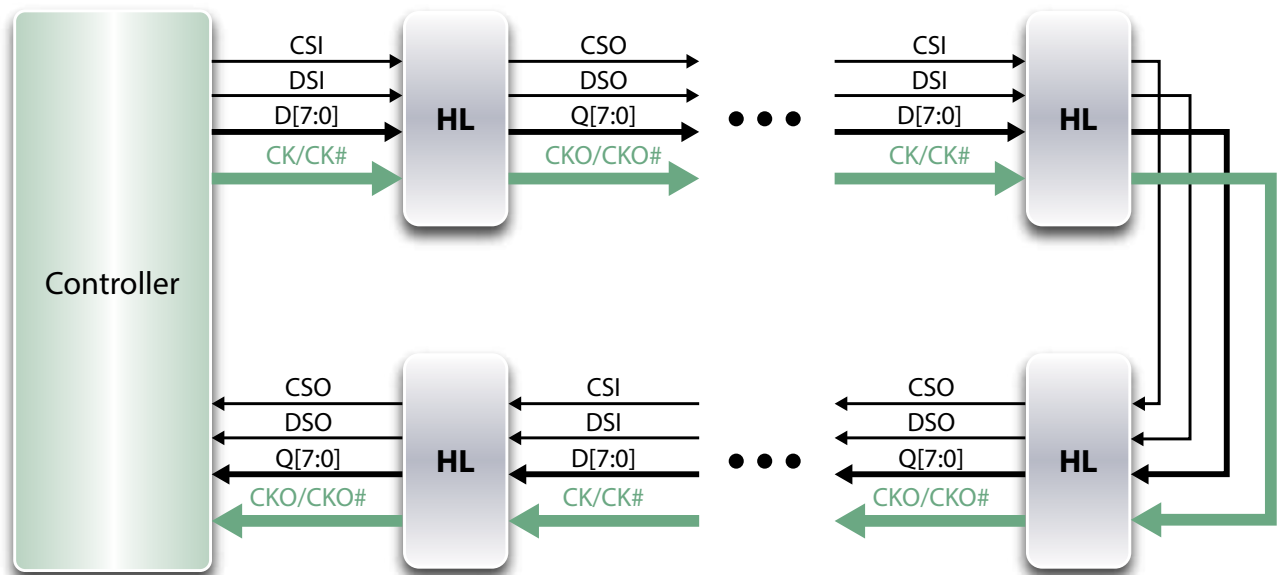
Figure 2. HLNAND with Source-Sychronous Clock

Figure 3 shows how the channel count increases as the system interconnect speed increases. The graph excludes command overhead and the need for redundant capacity to facilitate background operations and mitigate performance degradation as the Flash chips age—factors that increase the requirement for extra channels of Flash. Conventional Flash, operating at about 40MB/s, is rapidly becoming obsolete and is already virtually unworkable with interconnects like SATA 3, PCIe 1.x and PCIe 2.x, at mid-level lane counts. As the industry considers increasing PCIe lanes and moving to PCIe 3, we see that conventional 40MB/s Flash will require more than 100 channels to saturate the interconnect. ONFi 3.0 or Toggle Mode 2.0 Flash, operating at 400MB/s, requires several 10's of channels to saturate PCIe 2.x and PCIe 3.x storage systems.
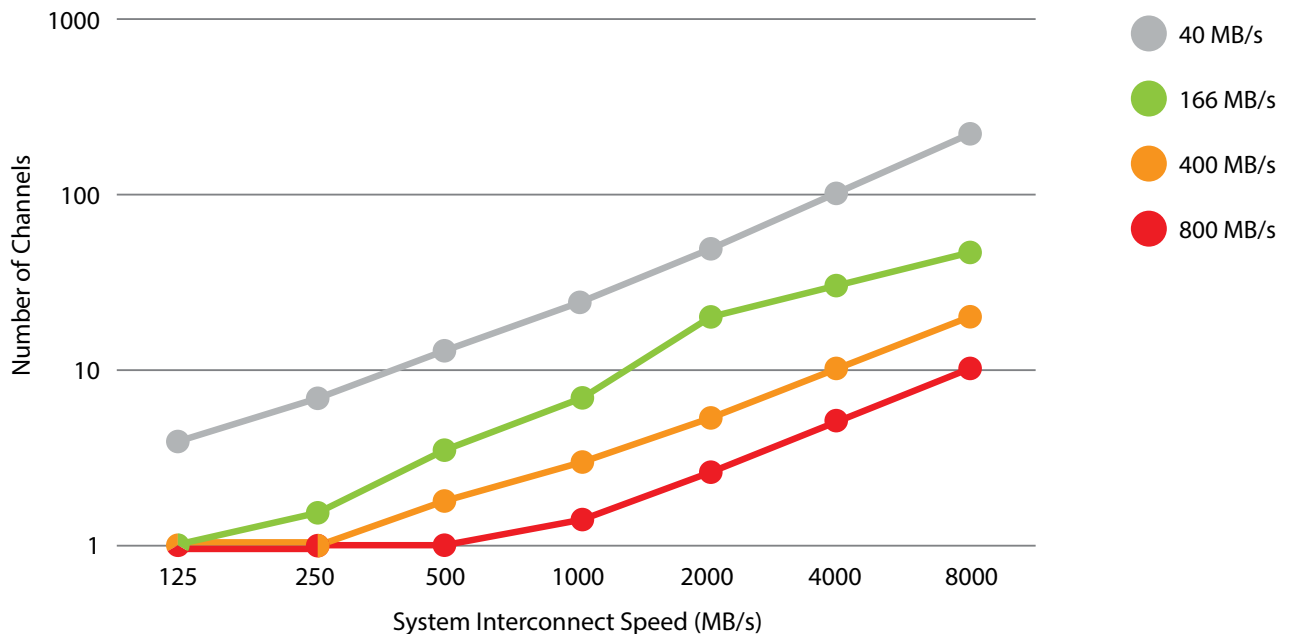

Figure 3. Growth in Memory Channels vs. Growth in System Interconnect Speed

The graph in Figure 3 shows that HLNAND, operating at 800MB/s, offers a solution that brings the memory channel count back down into the comfortable range of 8 to 12. HLNAND is well suited to deliver high performance to the enterprise segment, taking advantage of the high throughputs offered by PCIe 2.x and PCIe 3.x.

The scalability of HLNAND also makes it easier to hide the program latency on each memory channel as compared to parallel bus-based Flash interfaces. Over the years, NAND Flash program latency has increased as cell sizes have shrunk. Today, a typical program latency time is approximately 2ms, compared to about 0.8ms just a Flash generation ago. Since speed is not degraded on the HyperLink interface, it is possible to populate enough devices to hide the program latency and still maintain the full media bandwidth.

When we take program throughput into account, the situation deteriorates for Flash devices that utilize a parallel interface. Today's program latency times range from 800µs to 2ms, which translates to a program throughput range of 4MB/s to 10MB/s per device. In other words, ignoring the bus transfer time, each added device contributes up to 10MB/s of write throughput to a memory channel. Since it is impractical to add more than eight devices to a parallel bus, based on a 8KB page, the program throughput on a single channel of such Flash would range from 32MB/s to 80MB/s before topping out. That is well below the maximum transfer rate of the fastest parallel Flash bus today; 400MB/s. Therefore, parallel buses are ill-suited to high-speed Flash since it is not practical to populate the bus with enough devices to take full advantage of the higher speed.

This situation, depicted in the graph of Figure 4, is similar for all Flash interfaces that employ a parallel bus interface such as ONFi 2.0/3.0, Toggle Mode 1.0/2.0, and traditional asynchronous NAND Flash. Since a 40MB/s Flash interface is adequate for making drives with eight to 10 channels that saturate SATA 2 interconnects, the parallel bus was sufficient and well suited to earlier, slower system interconnects. The HyperLink interface is better suited to current and future higher-speed interconnects, since it utilizes a point-to-point daisy chain. Therefore, there is virtually no limit to the number of devices that can be populated on a single memory channel and the program throughput continues to increase as more devices are populated on the channel.
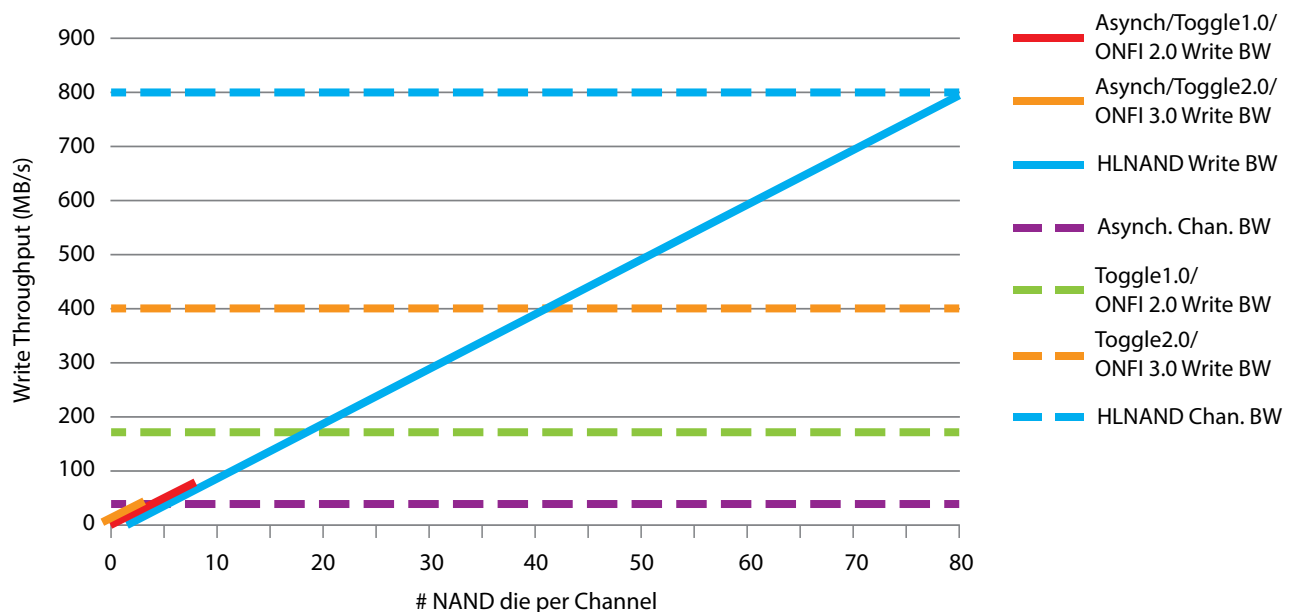

Figure 4. Write Throughput vs. Loads per Channel with a 8KB Page Size

Increasing the scalability of the Flash channel provides various opportunities. For example, drives that require greater capacity within a constant performance envelope need not be redesigned with a greater number of memory channels. Each memory channel may simply be populated with a greater number of devices to achieve the desired capacity. This property facilitates both scalability and upgradeability. HLNAND Memory Modules (HLDIMMs) may be used to upgrade the capacity of drives over a large range of capacities with no other hardware changes. This is not easy to do with a parallel bus-based interface since each channel is limited to a small number of loads and the use of memory modules exacerbates the loading. There is significantly more margin for employing memory modules with a point-to-point interface such as HyperLink, making their use far easier. With a logical limit of 255 Flash devices per memory channel, the range for upgradeability is clearly quite large.

INCREASED PAGE SIZE REQUIRES INCREASED INTERFACE DATA RATE

As NAND Flash device capacity has grown over the years, so too has the respective page size. In the early 2000s, devices with capacity in the 1Gb to 4Gb range had 2KB page sizes. These devices gave way to parts with 4Gb to 16Gb capacity and 4KB pages and more recently 32Gb to 64Gb devices with 8KB pages. In turn, 128Gb capacity devices with 16KB pages are now emerging. We expect the trend to continue, giving us 32KB pages in the foreseeable future.

However, as page size grows, a penalty is paid in per-page transfer times, which has a negative effect on IO operations per second (IOPS). The larger the page, the more time it takes to transfer a page of data at a given interface speed. Figure 5 shows the number of data transfers that can be completed on a given Flash interface type within the time it takes the Flash device to complete its internal read operation. Plots for four page sizes and their respective read times (tR) are given.

Figure 6 shows another way to express this relationship, demonstrating how the possible number of transfers decreases as the page size increases. As page size increases, the number of transfers that can be completed decreases, causing IOPS to decrease. As Flash vendors increase page sizes to compensate for slower program and read times without dramatically increasing the interface, data rate performance potential is left on the table—unused. Therefore it is important to maximize the Flash interface data rate to increase IOPS, especially as page size grows. HLNAND is capable of speeds of up to 800MB/s, corresponding to nearly eight full-page data transfers that may be completed during the time it takes to complete a read on a Flash device with an 8KB page size.
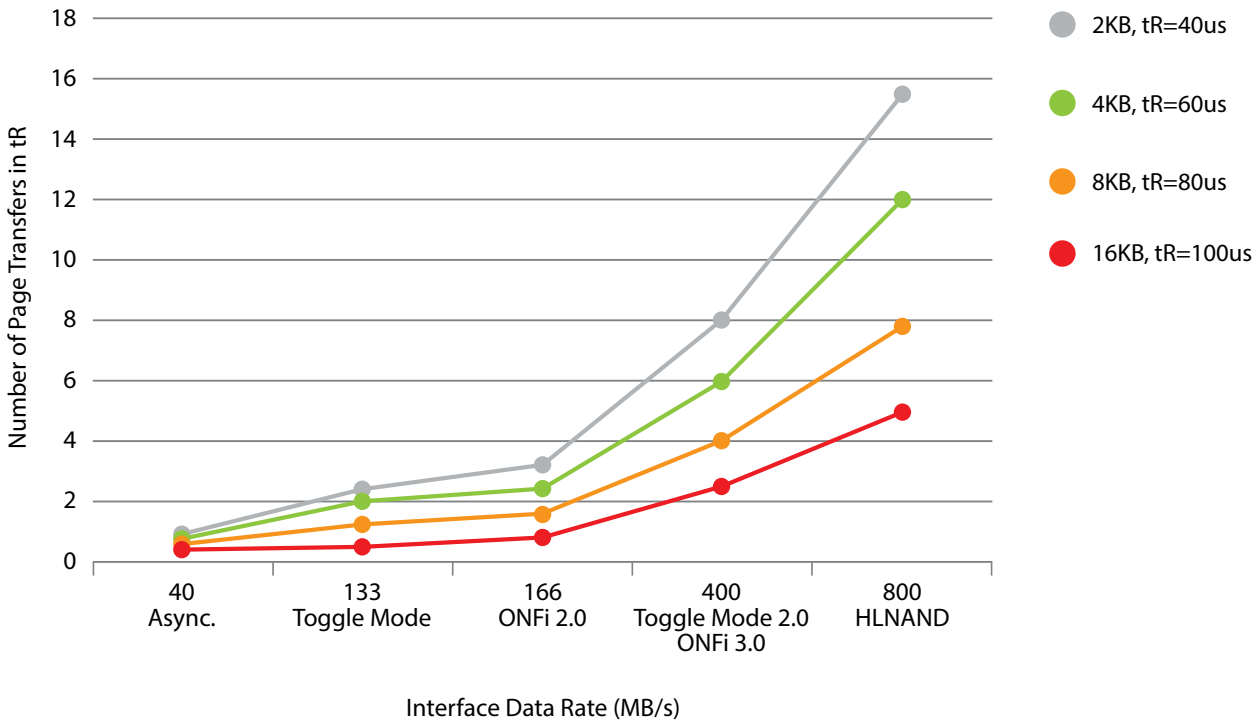
Figure 5. Number of Data Transfers vs. Interface Data Rate for Various Interfaces
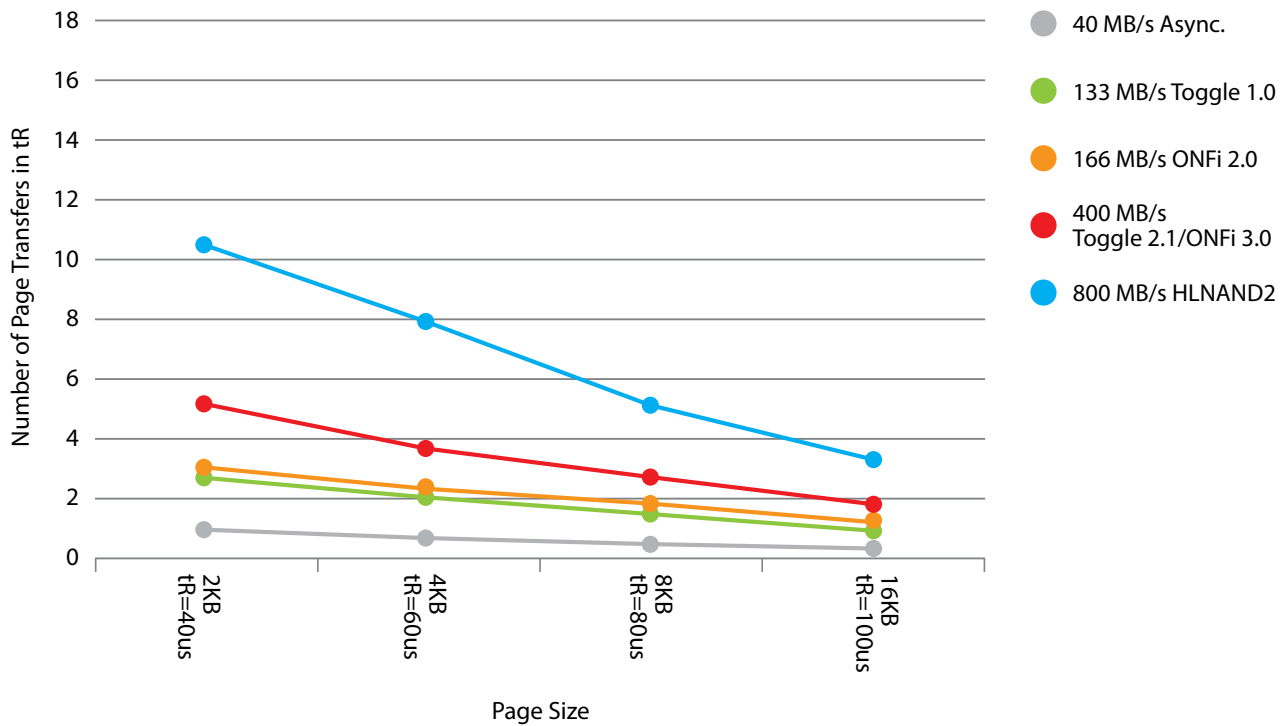

Figure 6. Possible Number of Data Transfers vs. Page Size for Various Interfaces

An important corollary to this performance metric is that, with a higher speed interface such as HLNAND that does not suffer from speed degradation with increased bus loading, it is possible to populate a large number of devices (up to 255) on single memory channel, thereby increasing parallelism without increasing the number of channels. With HLNAND, parallelism, one of the most effective ways to increase performance, can be increased with less cost in controller die area, as compared to parallel bus Flash interfaces, since the number of memory channels can be kept small.

2TB HLNAND SSD REQUIRES ONLY A SINGLE CONTROLLER

A 2TB HLNAND-based SSD (HLSSD) was developed as an HDD replacement in the industry standard 3.5" form factor. The 2TB HLSSD, shown in Figure 7 employs an FPGA controller and 4 HLNAND channels. Each of the channels connects to a removable 512GB HLDIMM, each containing sixteen 32GB HLNAND Flash MCPs. Each MCP contains eight 32Gb MLC NAND Flash die plus a single controller chip interfacing to the HLNAND ring. The controller interfaces to the host system via a SATA 3.0, 6Gbps port. Figure 8 shows that the HLSSD achieves 450MB/s sequential read speed and 390MB/s sequential write speed , virtually saturating the SATA3 interface. Using 64Gb or 128Gb NAND die, and 16 die stacking in each MCP, HLSSD capacities of 4TB, 8TB, and even 16TB can be achieved with the same architecture. Replacing the FPGA with a custom ASIC controller will reduce cost and power while fully saturating the SATA3 interface.

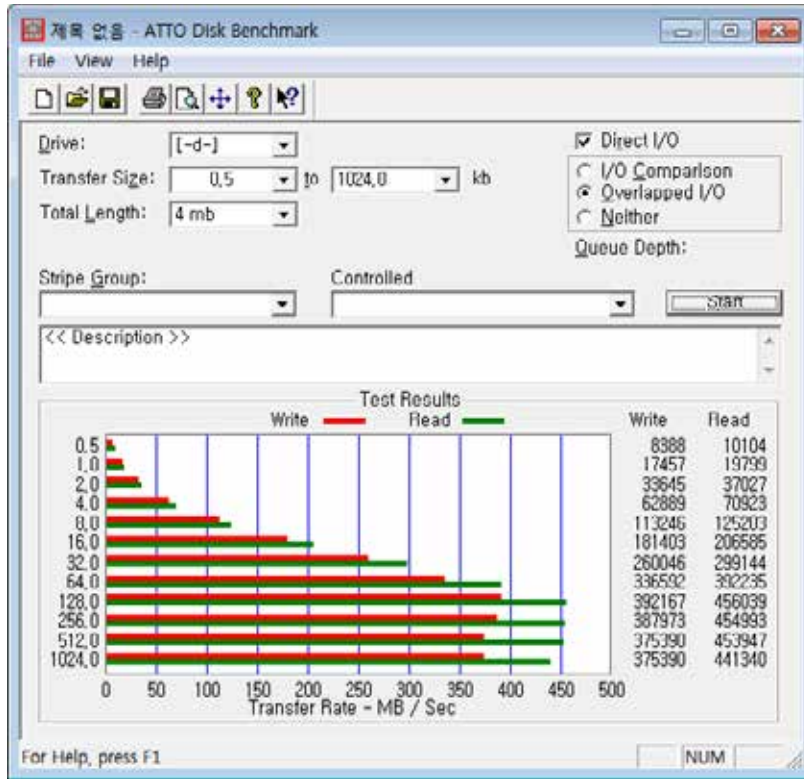*Figure 7. 2TB HLSSD in Standard 3.5" HDD Form Factor*

Figure 8. Atto Disk Benchmark Results for 2TB HLSSD

CONCLUSION

As the trend toward faster system interconnects continues, it is clearly desirable to have a NAND Flash interface that is capable of greater bandwidths and that is highly scalable, in order to accommodate greater capacity on fewer memory channels. HLNAND delivers on both counts to provide the fastest and most highly scalable Flash memory solution available today.

Both high bandwidth and scalability are facilitated by the point-to-point daisy chain Ring topology of the HyperLink interface. Higher bandwidths are possible because each device in the memory channel drives only one load, resulting in a channel that operates on a higher clock speed. The point-to-point, daisy chain topology means that, as more devices are added to the channel, loading is not increased and therefore speed is not decreased. When considering write throughput, more devices per channel provide greater throughput, thus highlighting the importance of scalability in a given Flash interface technology. As page sizes continue to grow, it is important to maximize the interface's data rate in order to achieve high IOPS. High interface data transfer rates in combination with a highly scalable interface results in SSD designs with greater parallelism and less controller complexity. A comparison of HLNAND, ONFi 3.0, and Toggle Mode 2.0 NAND are shown in Figure 9, further illustrating the greater scalability, speed and power savings of HLNAND.

	HLNAND	ONFI 3.0	Toggle 2.0
Synchronous IO	Yes	Yes	No
Transfer rate	800MT/s	400MT/s	400MT/s
Clock speed	400MHz	200MHz	200MHz
# chips before roll-off	Limitless	4	4
IO Voltage	1.2V	1.8V	1.8V
Termination (ODT)	Not required	Required	Required

Figure 9. Comparison of Key Flash Interface Characteristics

Finally, the capabilities of HLNAND are evidenced by the 2TB HLSSD, which demonstrates that a read throughput of 450MB/s and a write throughput of 390MB/s are achievable in a 4-channel HLNAND-based SSD. An 8-channel HLNAND-based SSd will have sufficient internal bandwidth for an 8-lane PCIe 2.x or a 4-lane PCIe 3.0 system interface. HLNAND outperforms any other Flash interface on the basis of per-channel throughput and per-channel capacity.

NOTES

c·nversant

INTELLECTUAL PROPERTY MANAGEMENT