Top 5 Reasons

Why Deploy PCIe® 5.0 NVMe™ SSDs In Your Data Center?

With the advent of the PCI Express® (PCIe) 5.0 specification revision, data can move through the interface up to twice as fast when compared with the previous PCIe 4.0 generation. This enables devices, such as SSDs, graphics processing units (GPUs) and network interface controllers (NICs), to deliver input/output (I/O) even faster than before. The speed upgrade is beneficial to data-intensive and computational applications such as cloud computing, databases, data analytics, artificial intelligence (AI), machine learning (ML), container orchestration and media streaming, to name a few. Key server and SSD vendors are developing solutions to comply with the PCIe 5.0 standard (and associated NVMe 2.0 protocol) for immediate availability. There are a number of reasons to deploy PCIe 5.0 SSDs in your data center, but here are the top five:

1. *Twice the Performance*
2. *Device Density and Power Benefits*
3. *Designed for Multiple Markets*
4. *Support for Multiple SSD Form Factors*
5. *PCIe 5.0 and EDSFF E3.S SSD Availability*

#1 Twice the Performance

The PCIe 5.0 interface standard developed by PCI-SIG® is an upgrade that enables twice the data transfer speed and bandwidth versus the PCIe 4.0 specification (Table 1). It increases data transfer speeds from 16 gigatransfers per second (GT/s) to 32 GT/s. In other words, the PCIe 5.0 interface can move data at approximately 4 gigabytes per second (GB/s) per lane versus the almost 2 GB/s per lane supported by PCIe 4.0, doubling performance while delivering a theoretical 4-lane (x4) bandwidth of approximately 16 GB/s. As a result, PCIe 5.0 NVMe SSDs are able to deliver bandwidths from 13,000 to 14,000 megabytes per second (MB/s). When compared to top-end enterprise SATA SSD bandwidths from 500 to 600 MB/s, PCIe 5.0 bandwidth performance is eye popping.

PCIe Revision	Transfer Rate (in GT/s)	x16 Link Bandwidth (in GB/s)	Year Ratified
PCIe 1.0	2.5	8	2003
PCIe 2.0	5.0	16	2007
PCIe 3.0	8.0	32	2010
PCIe 4.0	16.0	64	2017
PCIe 5.0	32.0	128	2019
PCIe 6.0	64.0	256	2021

Table 1: PCIe specification generations (source: PCI-SIG)

#2 Device Density and Power Benefits

The PCIe 5.0 specification is an update to the PCIe 4.0 standard and is backwards compatible. The doubling of performance enables devices to communicate faster with system CPUs and transfer large chunks of data with lower latency. It also means that devices may be able to achieve the same throughput with fewer PCIe lanes, making more lanes available. For example, a PCIe 4.0 SSD that used to require four lanes to run at full speed may now be able to run at the same speed with two PCIe 5.0 lanes, with two additional lanes available for other workloads. This is an important benefit because CPUs provide a limited number of lanes for distribution among devices so they can communicate with each of them.

By utilizing half of the lanes, the PCIe 5.0 interface enables many more devices in a system versus the PCIe 4.0 interface. The PCIe 5.0 interface is expected to become a critical part of data centers as it possesses high-speed networking capabilities that can easily handle a 400 Gigabit Ethernet® (GbE) network connection.

The increased performance gains of the PCIe 5.0 interface also has a positive effect on energy efficiency within a system. Performance per watt is one example where a system measures the rate of computation for each watt of power it consumes. To demonstrate system performance per watt efficiency, KIOXIA CM7-V Series PCIe 5.0 SSDs utilizing 25-watt (25W) power envelopes and KIOXIA CM6-V Series PCIe 4.0 SSDs utilizing 19W power envelopes are compared[1] in 3.2 terabyte[2] (TB) capacities (Table 2).

		KIOXIA CM7-V Series *(25 W power)*		KIOXIA CM6-V Series *(19 W power)*		
Read / Write Operation		**PCIe 5.0 Performance**	**PCIe 5.0 Performance/Watt**	**PCIe 4.0 Performance**	**PCIe 4.0 Performance/Watt**	**PCIe 5.0 Advantage**
Sequential Read	*(128 KB; QD=32)*	14,000 MB/s	560 MB/s per watt	6,900 MB/s	363 MB/s per watt	+54%
Sequential Write	*(128 KB; QD=32)*	7,000 MB/s	280 MB/s per watt	4,200 MB/s	221 MB/s per watt	+26%
Random Read	*(4 KB; QD=256)*	2,700,000 IOPS	108K IOPS per watt	1,400,000 IOPS	73K IOPS per watt	+47%
Random Write	*(4 KB; QD=32)*	600,000 IOPS	24K IOPS per watt	350,000 IOPS	18K IOPS per watt	+33%

Table 2: Performance comparisons between PCIe 5.0 and PCIe 4.0 using KIOXIA CM Series SSDs

In this example, the PCIe 5.0 interface delivers significant performance/power efficiency advantages over the previous PCIe 4.0 interface, and reflects the performance per watt increases, improving energy efficiency by 54% for sequential reads, 26% for sequential writes, 47% for random reads and 33% for random writes.

#3 Designed for Multiple Markets

Server and storage vendors have recently brought new Enterprise and Datacenter Standard Form Factor (EDSFF) E3.S and PCIe 5.0 platforms to market. The EDSFF E3.S form factor targets NVMe SSDs with x4 PCIe link widths (though they can mechanically fit an x16 card edge). With power profiles up to 25W, E3.S form factors are well suited for general-purpose NVMe server storage subsystems where 2.5-inch[3] form factor SSDs are typically used, including those with modular and short depth chassis. The E3.S form factor additionally delivers:

- *Support for multiple device types such as compute express link (CXL™) devices, storage class memory (SCM) devices, computational storage devices, low-end accelerators, and front-facing I/O devices such as NICs*
- *Better flexibility relating to SSD and device placement*
- *Greater storage density than the 2.5-inch format per rack unit*
- *Support for the PCIe 5.0 interface as well as next-generation, high-frequency interfaces - E3 connector provides robust signal integrity characteristics and is designed to support the PCIe 6.0 interface*
- *Support for up to 25W power profiles required to saturate a PCIe 5.0 4x 32 GT/s link*

#4 Support for Multiple SSD Form Factors

PCIe is an interface technology not tied to any SSD form factor - therefore, PCIe 5.0 SSDs could be available in any form factor as long as their electrical/connector specifications support the PCIe 5.0 characteristics. SSD form factors supported by the PCIe 5.0 interface include 2.5-inch, M.2 and new EDSFF E3.S/E3.L and E1.S/E1.L variants. Today, all industry standard NVMe SSDs use the PCIe interface as a physical transport.

The 2.5-inch form factor originated with hard disk drives and has served the industry for three decades. It is not optimal for flash memory chip packages or optimized for flash memory channels. With faster communication protocols and robust interfaces, storage media based on 2.5-inch drive formats are limited and cannot keep technological pace with new server demands, especially those based on PCIe 5.0 technologies and beyond.

The M.2 form factor is more recent (introduced in 2012). Its small and slim size makes it ideal for lightweight and portable systems such as laptops, notebooks, ultrabooks and tablets. M.2 SSDs take up less space than 2.5-inch SSDs or hard drives and typically feature up to 3.84 TB storage capacity. However, M.2 SSDs have limitations. For example, an M.2 SSD includes a 22 mm wide printed circuit board (PCB) that constrains flash memory placement and limits higher drive capacities. These SSDs are also not hot pluggable, requiring a server power-down to remove them.

Both form factors lack support for signal integrity issues, multiple device types, multiple link widths and different server platform sizes. They also do not support reasonable power envelopes in the future that scale to higher performing devices, and they have thermal challenges. The industry realized years ago that a new set of form factors could address these limitations and support future enterprise architectural requirements. The EDSFF working group formed to address these objectives, and from this effort, the E3 family and E1.S family of form factors were created.

The EDSFF E3 family includes the following variants:

E3 Short Thin (E3.S) 76 mm x 112.75 mm x 7.5 mm Up to 25W power envelope	**E3 Short Thick (E3.S 2T)** 76 mm x 112.75 mm x 16.8 mm Up to 40W power envelope	**E3 Long Thin (E3.L)** 76 mm x 142.2 mm x 7.5 mm Up to 40W power envelope	**E3 Long Thick (E3.L 2T)** 76 mm x 142.2 mm x 16.8 mm Up to 70W power envelope
Targeted to NVMe SSDs with x4 PCIe link widths. Positioned to be a primary form factor for mainstream NVMe server storage subsystems and used across a wide variety of platforms including modular and short depth chassis.	Targeted to higher performance NVMe SSDs (PCIe 5.0 x4 saturation), CXL, SCM, computational storage or front I/O implementations. It supports x4, x8 or x16 PCIe link widths.	Targeted to be a primary form factor for storage subsystems and server platforms requiring maximum capacity for each 'U' configuration that utilize deeper chassis, and for high-capacity NVMe SSDs or SCM devices with support for x4, x8 or x16 PCIe link widths.	Targeted to FPGAs and accelerators, with support for x4, x8 or x16 PCIe link widths.

The E3 form factor platform enables system designers to increase overall airflow significantly in compatible systems with the ability to deploy a large number of SSDs to increase storage subsystem performance. E3.S systems and SSDs driven by the PCIe 5.0 interface and NVMe 2.0 protocol are available today (see next section).

The EDSFF E1.S family includes the following variants:

E1.S 9.5 mm 33.75 mm x 118.75 mm x 9.5 mm Up to 20W power envelope	**E1.S 15 mm** 33.75 mm x 118.75 mm x 15 mm 25W+ power envelope	**E1.S 25 mm** 33.75 mm x 118.75 mm x 25 mm 25W+ power envelope
Targeted to small footprint systems such as 1U edge compute servers, storage systems, blade servers and densely scaled servers.	Targeted to 1U, 2U, 4U and special-purpose systems including compute blades and systems, performance and capacity-optimized storage systems, AI/ML systems, high performance computing (HPC) systems and edge systems.	Targeted to 2U and larger rack systems such as storage appliances, storage rich servers, storage rich database servers and performance-oriented storage systems.

#5 PCIe 5.0 and EDSFF E3.S Availability

If end users require PCIe 5.0 performance today from a leading server OEM, their only option as of this publication is EDSFF E3.S-compatible systems and SSDs. PCIe 5.0 SSDs with E3.S system compatibility are only available from leading server and storage OEMs. KIOXIA Corporation offers the KIOXIA CM7 Series enterprise NVMe SSDs supporting the PCIe 5.0 interface and E3.S form factors in one device. The series includes:

KIOXIA CM7 Series SSD[4]

CM7-R Series:
The KIOXIA CM7-R Series (Table 3) are read intensive E3.S NVMe SSDs optimized to support a broad range of enterprise applications and workloads such as data warehousing, business intelligence, AI/ML, online transaction processing (OLTP), software defined storage (SDS) and virtualization.

CM7-V Series:
The KIOXIA CM7-V Series E3.S NVMe SSDs (Table 3) for higher endurance mixed use workloads support a broad range of enterprise applications that could include HPC, OLTP, IoT/IIoT, edge computing and media streaming.

Metric	KIOXIA CM7-R E3.S Series	KIOXIA CM7-V E3.S Series
Interface/Protocol	PCIe 5.0 / NVMe 2.0	PCIe 5.0 / NVMe 2.0
Interface Speed	128 GT/s	128 GT/s
Capacities	1,920 / 3,840 / 7,680 / 15,360 GB	1,600 / 3,200 / 6,400 / 12,800 GB
Drive Write(s) per Day[5] (DWPD)	1	3
Sequential Read Performance[1]	up to 14,000 MB/s	up to 14,000 MB/s
Sequential Write Performance[1]	up to 7,000 MB/s	up to 7,000 MB/s
Random Read Performance[1]	up to 2,700,000 IOPS	up to 2,700,000 IOPS
Random Write Performance[1]	up to 310,000 IOPS	up to 600,000 IOPS

Table 3: Overview of KIOXIA CM7 Series E3.S SSDs

More information on KIOXIA CM7 Series E3.S SSDs is available here.

Notes:

[1] Performance metrics provided by KIOXIA Corporation. Actual results may vary due to system configuration usage and other factors.

[2] Definition of capacity - KIOXIA Corporation defines a megabyte (MB) as 1,000,000 bytes, a gigabyte (GB) as 1,000,000,000 bytes and a terabyte (TB) as 1,000,000,000,000 bytes. A computer operating system, however, reports storage capacity using powers of 2 for the definition of 1Gbit = 2^{30} bits = 1,073,741,824 bits, 1GB = 2^{30} bytes = 1,073,741,824 bytes and 1TB = 2^{40} bytes = 1,099,511,627,776 bytes and therefore shows less storage capacity. Available storage capacity (including examples of various media files) will vary based on file size, formatting, settings, software and operating system, and/or pre-installed software applications, or media content. Actual formatted capacity may vary.

[3] 2.5-inch indicates the form factor of the SSD and not its physical size.

[4] The product image shown is a representation of the design model and not an accurate product depiction.

[5] Drive Write(s) per Day: One full drive write per day means the drive can be written and re-written to full capacity once a day, every day, under the specified workload for the specified lifetime. Actual results may vary due to system configuration, usage, and other factors.

Trademarks:

CXL is a trademark of Compute Express Link Consortium, Inc. Ethernet is a registered trademark of Fuji Xerox Co., Ltd. NVMe is a registered or unregistered trademark of NVM Express, Inc. in the United States and other countries. PCIe, PCI Express and PCI-SIG are registered trademarks of PCI-SIG. All other company names, product names and service names may be the trademarks or registered trademarks of their respective companies.

Disclaimers: