SSD Series for Enterprise Use with Configurable Specifications to Flexibly Accommodate Applications

● MORO Hiroyuki ● KANEKO Atsushi

The use of solid-state drives (SSD) utilizing NAND flash memory is rapidly expanding in the field of cloud computing, which is highly dependent on enterprise servers and storage devices. Accompanying the dissemination of SSDs, their characteristics have become widely known and users' requirements have diversified, with priority being placed on specific features such as lower bit cost, higher read/write performance, larger capacity, higher endurance, or other characteristics.

Taking this trend into consideration, Toshiba has developed a new SSD series that not only provides improved performance, but also allows its specifications to be configured to accommodate the user's applications.

1. Introduction

Solid-state drives (SSDs) that support today's cloud and network computing are now moving from the introduction phase to the growth phase. Users are becoming well versed in SSDs and gaining significant practical expertise in their use. Accompanying the wide uptake of SSDs, users' requirements for SSDs have been changing and diversifying.

In response to this change, Toshiba has developed the PX04S family of SSDs for enterprise use that accommodates diverse user requirements while reducing the cost per bit, comparing with models of previous generation PX02S family. This report provides an overview of the PX04S family and describes the technologies used to meet diverse requirements.

2. Overview of the PX04S family

The PX04S family (**Figure 1**) comprises our third-generation SSDs designed for enterprise use. **Table 1** shows its main specifications.

The PX04S family provides a 12 Gbits/s Serial Attached Small Computer System Interface 3 (SAS-3) host interface and a maximum user storage capacity of 3.84 Tbytes (tera: 10^{12}). The host interface connector is compatible with both SAS and PCI Express[†], simplifying its applications for NVM Express[†] attached through the PCI Express bus (hereinafter referred to as "NVM Express[†]/PCI Express[†]) that is expected to come into widespread use, especially in server platforms. In order to meet the customer requirement for a reduction in the cost per bit of storage, the PX04S family uses multi-level-cell (MLC) NAND flash memory fabricated with the second-generation 19 nm process. Furthermore, to satisfy the increasing requirement for higher read/write

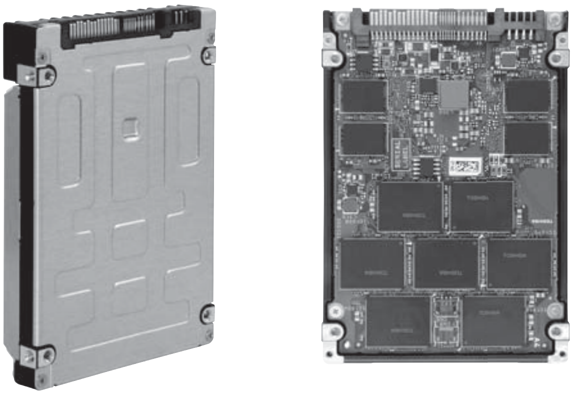


Figure 1. PX04S family SSD for enterprise use.
The new SSD family achieves high reliability and high performance by using a newly developed controller and NAND flash memory fabricated with a second-generation 19 nm process.

Table 1. Main specifications of PX04S family

Characteristic		Specification
Storage capacity (Gbytes)		200 to 3 840
NAND process		19 nm, second generation
Host interface		SAS-3 (12 Gbits/s)
Access performance	Sequential read (Mibytes/s)	1 500 to 1 900
	Sequential write (Mibytes/s)	750 to 1 900
	Random read (kIOPS)	270
	Random write (kIOPS)	22 to 145
Power efficiency (kIOPS/W)		27
Data reliability (bit error rate)		1×10^{-17}

Mibyte: mebi (2^{20}) bytes
IOPS: input/output per second

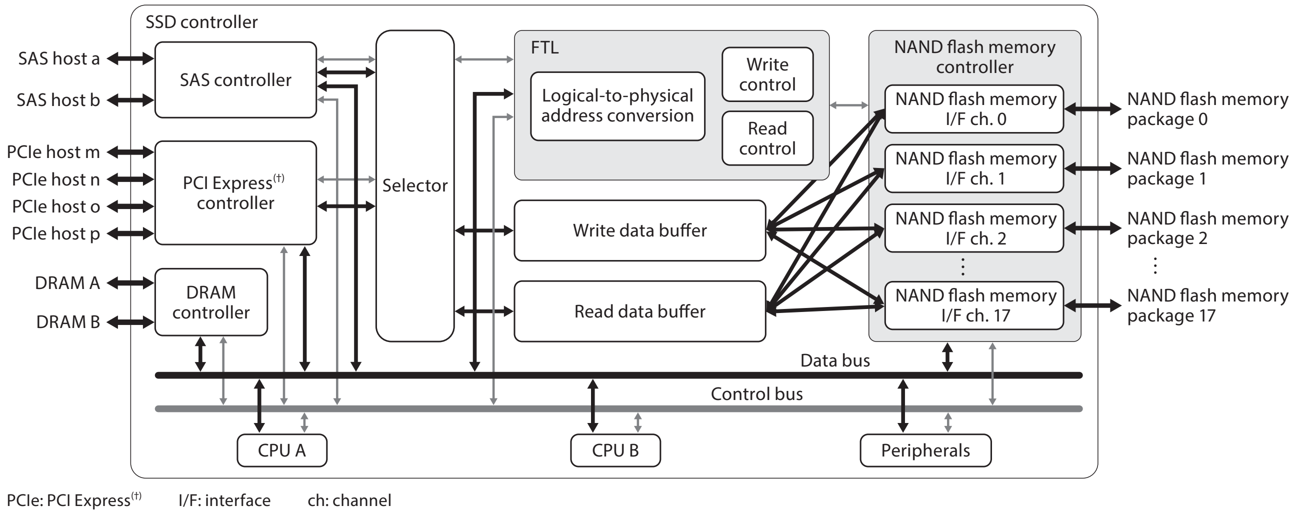

SSD controller
SAS host a
SAS host b
SAS controller
PCIe host m
PCIe host n
PCIe host o
PCIe host p
PCI Express(†) controller
DRAM A
DRAM B
DRAM controller
Selector
FTL
Logical-to-physical address conversion
Write control
Read control
Write data buffer
Read data buffer
NAND flash memory controller
NAND flash memory I/F ch. 0
NAND flash memory I/F ch. 1
NAND flash memory I/F ch. 2
NAND flash memory I/F ch. 17
NAND flash memory package 0
NAND flash memory package 1
NAND flash memory package 2
NAND flash memory package 17
Data bus
Control bus
CPU A
CPU B
Peripherals
PCIe: PCI Express(†)
I/F: interface
ch: channel

Figure 2. Block diagram of newly developed SSD controller.
The FTL is capable of removing power supplies from certain parts of the memory regions. It provides great flexibility in meeting customers' read/write performance requirements by enabling and disabling power supplies to individual regions.

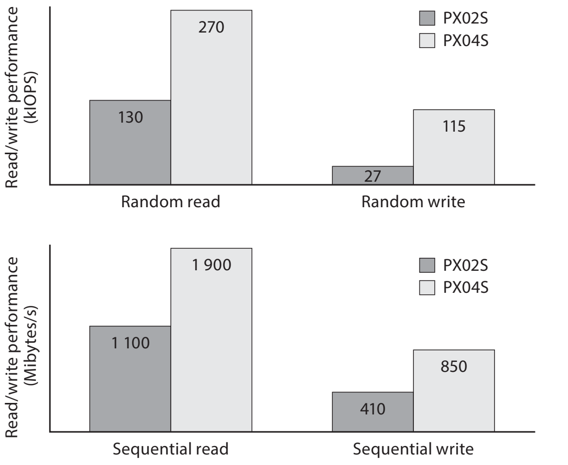
performance, we have developed a new SSD controller, which delivers a significantly improved data processing capability while inheriting the basic architecture of the controller for the second-generation enterprise SSDs(1). As a result of the foregoing, the PX04S family provides the industry's top-class read/write performance close to the maximum limit of the SAS-3 standard.

3. Controller

Figure 2 shows the organization of the SSD controller developed for use in the PX04S family. While the PX04S family uses an SAS host interface, its controller supports both SAS and PCI Express(†) host interfaces. This controller will also be used in an SSD family with NVM Express(†)/PCI Express(†) that is being developed. The new family will have a third-generation dual-port SAS interface and a third-generation four-lane PCI Express(†) interface.

The most notable feature of the controller is the flash translation layer (FTL) that converts read/write requests from the host interface into NAND flash memory commands. The FTL is capable of turning on and off the power supplies to specific FTL regions. Consequently, only the required FTL regions can be used in power-sensitive SAS-based products whereas all FTL regions can be used in PCI Express(†)-based products that prioritize high read/write performance. Thus, the controller can meet the requirements of both types of SSDs. Furthermore, while the second-generation enterprise SSDs have a 16-channel NAND flash memory interface, the PX04S family has an 18-channel interface. The operating frequency and data bus width have also been increased to boost the data processing performance of the FTL considerably. As a result, the PX04S family has a random-read performance more than twice that of the previous PX02S family and over four times the random-write performance (**Figure 3**). The PX04S family also exhibits a sequential-read performance approximately 1.7 times that of the PX02S family and over twice the sequential-write performance.

As NAND flash memory migrates to finer process geometries, the controller is uniquely architected to improve bit error correction capability and thereby increases data integrity and reliability. This is accomplished by a unique technology that switches the format in which data are written to the NAND flash memory.



Comparison conditions: SSD of the mid-endurance category*with a dual-port connection at the factory-default settings

* SSDs for main server and storage applications that require fast response time, high reliability and economic efficiency

Figure 3. Comparison of performance of PX04S family and previous PX02S family.
Compared with the PX02S family, the PX04S family provides over four times the random-write performance, over twice the random-read and sequential-write performances, and approx. 1.7 times the sequential-read performance.

This reliability technology is also applicable to the next-generation NAND flash memory, making it possible for a single controller to obtain the optimal performance from multiple generations of NAND flash memory.

4. User-programmable setup and configuration options

The principal characteristic of the PX04S family is that it allows users to modify the settings and configuration for three parameters that affect performance and service life. Two setup modes are available: online setup mode in which an SSD is set up while a system is active, and offline setup mode in which an SSD is configured using a PC before it is embedded in a system. It is unnecessary to reformat an SSD even when its parameters have been modified. New parameters are automatically reflected in the values that are reported via S.M.A.R.T. (Self-Monitoring Analysis and Reporting Technology), including changes in storage capacity and estimated endurance. Thus, parameters can be easily modified.

The user-programmable parameters are detailed in the following subsections.

4.1 Performance-versus-capacity control

There is a trade-off between improving performance and increasing the available user area capacity.

The PX04S family allows users to adjust the capacity of the available user area that has been fixed up to the preceding family. Reducing the user area increases random-write performance and write endurance. Conversely, increasing the user area reduces random-write performance and write endurance (**Figure 4**).

For example, for read-intensive applications with infrequent writes, users can opt to allocate a greater capacity to the user area at the expense of reduced random-write performance, thereby increasing the total storage space per server or storage system.

The ability to trade-off between performance and capacity makes it possible to meet diverse customer requirements.

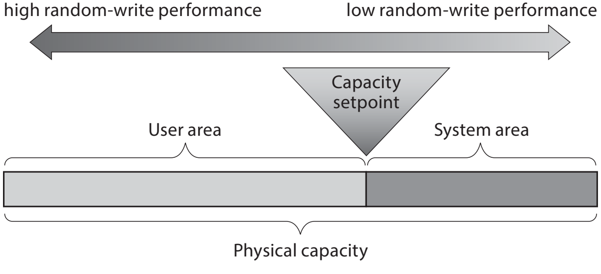


Figure 4. Outline of performance and capacity control.
Adjusting the capacity of the user area makes it possible to control write endurance and whether to prioritize performance or available capacity.

4.2 Performance mode

There is also a trade-off between reducing power consumption and increasing performance.

The factory default is set at an access performance level achieved at the typical power consumption. However, the PX04S family provides Performance mode in which users can adjust its access performance. In Performance mode, random-write performance is expected to increase by up to 30% at the expense of increased power consumption.

4.3 Write endurance control

The PX04S family has an internal useful-life indicator. The PX04S family can be configured to limit write performance as a means of maintaining write endurance when write operations per unit of time become more frequent than a baseline derived from the product life expectancy.

For example, prioritizing performance results in frequent writes. If it is expected that the remaining useful life will be shorter than a product's life expectancy, the PX04S family can automatically limit write performance to preserve life expectancy. See **Figure 5**. When the intersection of a useful-life indication and a product's life expectancy enters the region above the slope, write performance limiting is enabled until the intersection returns to the limit-free region.

This feature can be enabled and disabled, and the upper write performance limit is user-programmable.

5. Enclosure design

5.1 Mechanism design

Figure 6 shows a schematic diagram of the mechanical design of the PX04S family. The highest-capacity model has two printed circuit board assemblies (PCBAs) to accommodate up to 18 NAND flash memory chips. The use of an aluminum enclosure helps enhance heat dissipation.

The cover and base plates are formed using a metal-working technique called raising in order to increase

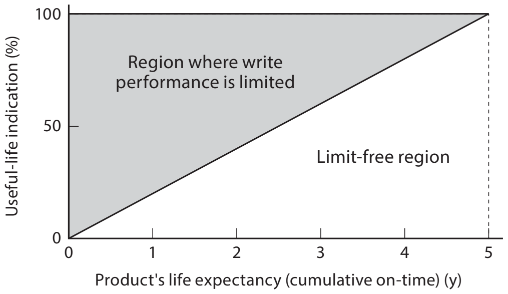


Figure 5. Outline of endurance control.
When the useful-life indicator shows an entry into the region above the slope, write performance is limited until an SSD returns to the limit-free region. The purpose of this feature is to control an SSD's life expectancy.

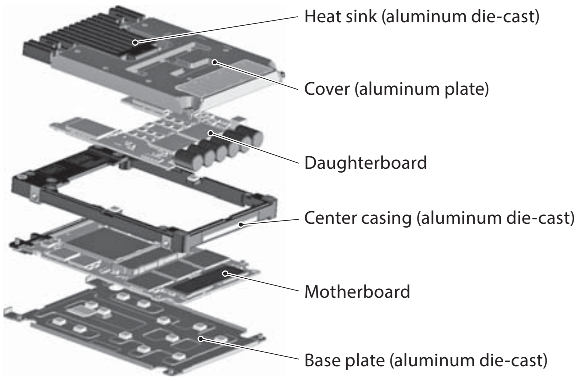

Heat sink (aluminum die-cast)
Cover (aluminum plate)
Daughterboard
Center casing (aluminum die-cast)
Motherboard
Base plate (aluminum die-cast)

Figure 6. Schematic diagram of PX04S family.
The aluminum enclosure efficiently removes heat generated by the NAND flash memory array, controller and other constituent components. The enclosure also provides a heat sink above the controller.

their rigidity and thereby secure mechanical shock resistance.

5.2 Thermal Design

As a result of improving SSD performance, the components that populate SSDs are becoming more power-hungry. Therefore, thermal design is crucial to ensure that the recommended operating conditions remain equivalent to those of the conventional enterprise SSDs. Most notably, the controller, the heart of an SSD, dissipates considerable heat. A variant of the PX04S family, which has an NVM Express(†)/PCI Express(†) interface and is housed in the same enclosure as the PX04S family, consumes approximately double the power of the PX04S family to obtain the best performance from the PCI Express(†) interface.

In order to efficiently remove the resulting heat, the SSD has a thermal conduction sheet immediately above the controller and part of the enclosure is formed as a heat sink. Additionally, heat conduction sheets are also placed between the enclosure and NAND flash memories as well as other constituent components, and aluminum is used as the enclosure to enhance heat removal.

Figure 7 shows an example of thermal simulation results. As shown in the figure, a considerable amount of heat is effectively removed from the controller.

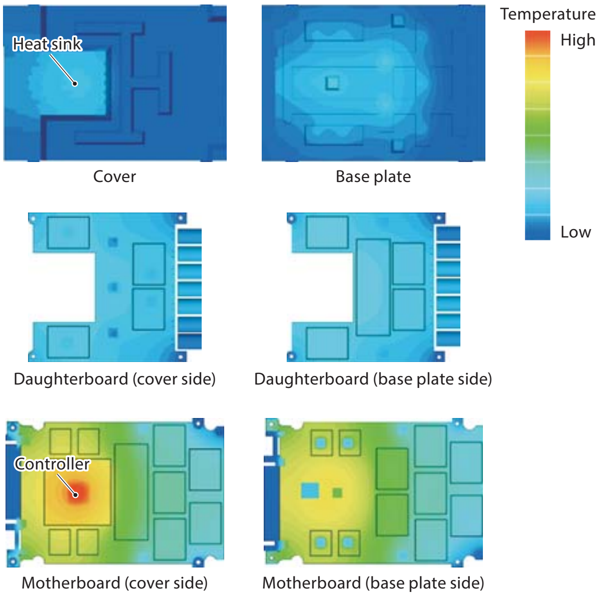

Heat sink
Cover
Base plate
Temperature
High
Low
Daughterboard (cover side)
Daughterboard (base plate side)
Controller
Motherboard (cover side)
Motherboard (base plate side)

Figure 7. Temperature distribution simulation on enclosure.
The heat sink, an integral part of the enclosure, efficiently removes a considerable amount of heat generated by the controller.

6. Conclusion

Drawing on our various innovative ideas and technologies, we have commercialized a family of enterprise SSDs that readily meet diversifying customer needs.

Based on the experience and expertise obtained from this project, we will accelerate the development of a variant of the PX04S family with NVM Express(†)/PCI Express(†) as well as the next-generation enterprise SSDs using BiCS FLASH™(*1), a three-dimensional (3D) flash memory with a stacked cell structure.

(*1) A structure stacking flash memory cells vertically on a silicon substrate to realize significant density improvements over planar NAND flash memory, where cells are formed on the silicon substrate.

References

(1) Kimura, A. et al. 2013. "1.6 Tbyte SSD for Enterprise Use Applying MLC NAND Flash Memory (in Japanese)." *Toshiba Review* 68 (9): 46–48.

(2) Kiuchi, H. 2011. "MK4001GRZB Solid-State Drive for Enterprise Use Achieving High Performance and High Reliability (in Japanese)." *Toshiba Review* 66 (8): 40–43.

MORO Hiroyuki

Chief Specialist. eSSD Engineering Department, Storage Products Division, Semiconductor & Storage Products Company. He is involved in the development of SSD controller chips.

KANEKO Atsushi

Specialist. Storage Products Application Engineering Department, Storage Products Division, Semiconductor & Storage Products Company. He is involved in technical customer support services for enterprise SSDs.

Notes:
This publication may include Toshiba's original estimation, and/or data based on Toshiba's evaluation.