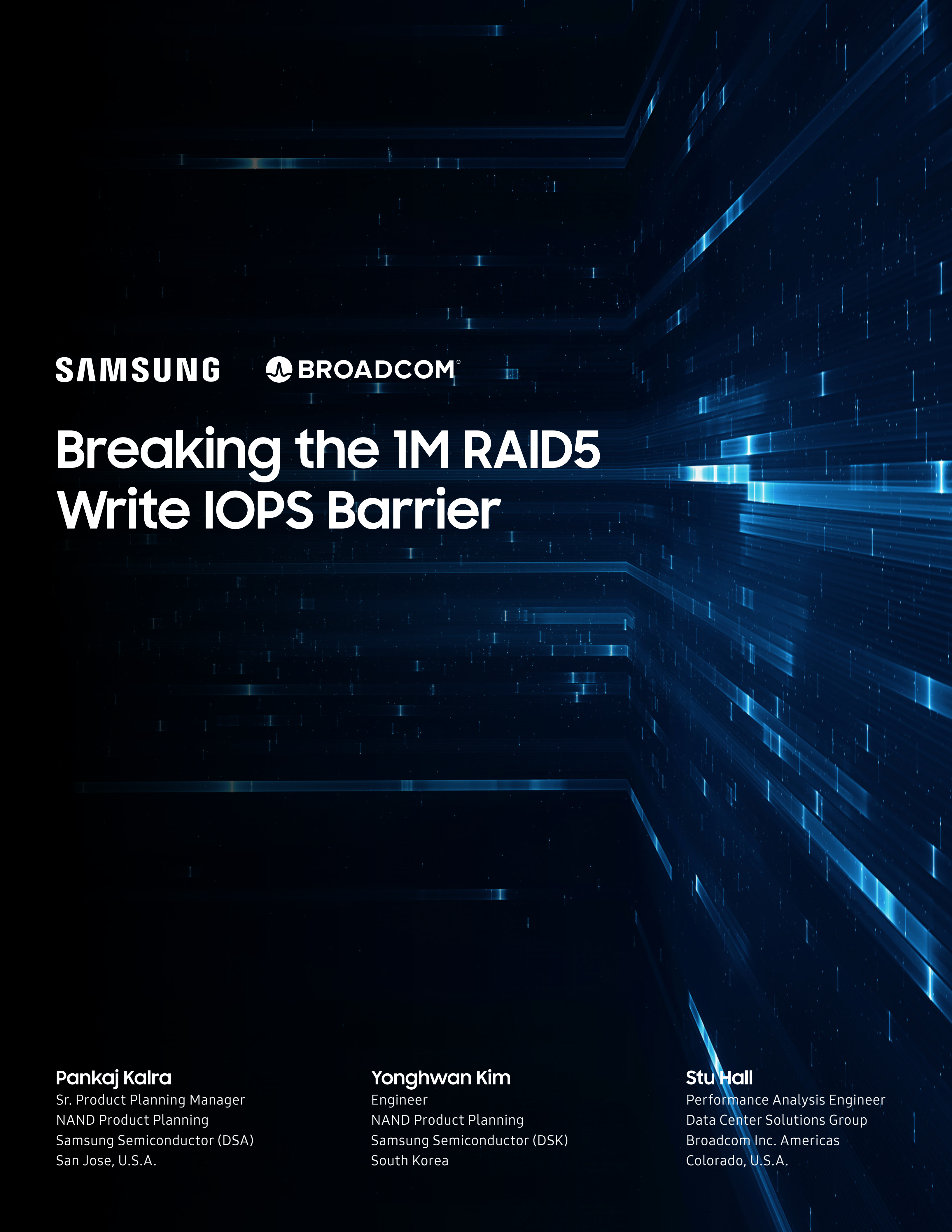


SAMSUNG BROADCOM®

Breaking the 1M RAID5 Write IOPS Barrier

Pankaj Kalra
Sr. Product Planning Manager
NAND Product Planning
Samsung Semiconductor (DSA)
San Jose, U.S.A.

Yonghwan Kim
Engineer
NAND Product Planning
Samsung Semiconductor (DSK)
South Korea

Stu Hall
Performance Analysis Engineer
Data Center Solutions Group
Broadcom Inc. Americas
Colorado, U.S.A.

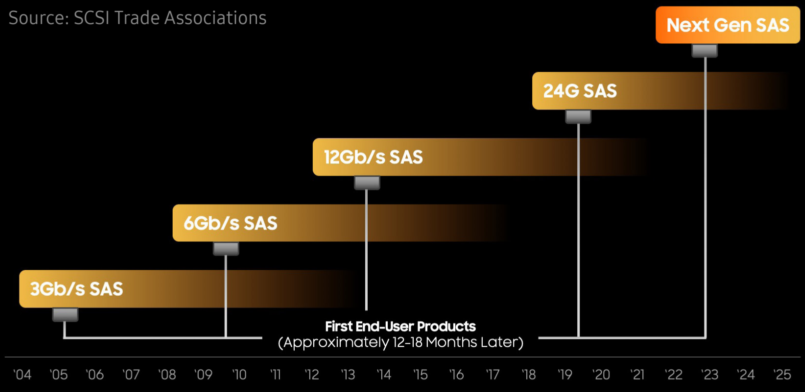
1. Introduction

In today's data-centric age, enormous amounts of data are generated, stored and processed at an unprecedented rate. Businesses are utilizing this data to make better decisions, drive greater efficiencies, develop more desirable products, improve profitability and ultimately increase user satisfaction. To continue deriving a high degree of value from a rapidly-expanding data flow, today's enterprise storage systems are constantly challenged to increase throughput while providing reliable data protection. This white paper highlights key challenges and offers a breakthrough solution that integrates highly innovative 24G SAS products from Samsung and Broadcom.

Serial-attached SCSI, commonly known as SAS, has been an industry standard for moving enterprise data between compute and storage systems, for more than 15 years. Since its introduction in 2004, SAS has proven to be a trusted and sustainable interface for demanding mission-critical workloads due to its high performance, exceptional reliability, remarkable scalability, outstanding flexibility and ease of management. In addition, the SAS ecosystem is well established, offering a wide range of inter-compatible HBAs, RAID adapters, expanders, cables, connectors and system backplanes.

"SAS continues to be one of the most trusted interfaces thanks to its ability to deliver the capacity, performance, reliability and scalability needed for the most demanding business-critical applications in enterprise data centers," said Jeff Janukowicz, Research Vice President at IDC. "With a clear roadmap to 24G, IT customers can expect continued enhancements to help data-intensive applications reach new levels of performance."

As demand for higher performance continues to rise, SAS has kept pace by transitioning from its original 3Gb/s to 12 Gb/s, and now to the latest performance threshold — 24G SAS. In fact, 24G SAS doubles the data throughput of its predecessor, offers improved reliability and is backward-compatible with the legacy infrastructure of 12Gb/s and 6Gb/s SAS, as well as with 6Gb/s SATA devices. It employs the more efficient 128b/130b encoding scheme with 20-bit forward error correction (FEC). This encoding scheme provides considerably better link efficiency at higher speeds. Moreover, FEC allows for detection and correction of transmission errors over high frequencies. With a very high degree of reliability, 20-bit FEC can detect and correct up to 2-bit errors without requiring re-transmission, thereby enabling maximum data throughput even in noisy channels.



Source: SCSI Trade Associations

Figure A - SAS Technology Roadmap

2. Data Protection

Data protection is important with any mission-critical enterprise storage application to enable full availability and smooth business continuity in the event of a drive failure. All drives eventually fail. This simple fact requires that data protection be designed into every storage solution — hard disk storage configurations and solid state storage alike. SSDs have a considerably lower failure rate than HDDs, yet regardless of the type of storage in an enterprise storage system, the potential for drive failure must be managed without triggering downtime, data loss or file inaccessibility. Any of these outcomes can translate into significant delay and/or greater costs for an enterprise.

The use of RAID5 (Type 5 of a Redundant Array of Independent Disks) represents what can often be the most cost-effective solution for protecting data. That is because RAID5 takes advantage of the logical relationship between an enterprise's data and the "exclusive-or" (XOR) condition of that data, to compute a parity value from which lost data may be reproduced.

A popular alternative RAID protection method involves RAID10, which utilizes data redundancy so that half of the drives keep a copy of all of the data at any given moment. For example, compare a four-drive RAID5 and a four-drive RAID10 array data protection system from the standpoint of capacity that's available to the user. In this case, a RAID5 solution can yield a user storage capacity that is 50 percent greater than the user capacity of a RAID10 array. For a 10-drive array, the capacity advantage of RAID5 relative to RAID10 grows to 80 percent.

Though RAID5 has a distinct cost advantage, RAID5 performance historically has been limited. It's particularly limited by the extra I/O operation overhead required to implement the RAID5 Read Modify Write algorithm under the control of loss-protection firmware. However, Broadcom's new 9670W-16i controller sharply reduces dependence on FW for RAID5 configurations. Moreover, data center managers now have a viable option other than RAID1 when coupling a high degree of protection with a high-performance priority.

In order to complete a RAID5 write operation, the following steps must be taken:

1. Existing data must be rapidly read;
2. Parity information for that data must also be read;
3. Data to be written must replace all old data that is superseded by new data;
4. An XOR protective operation must be run on any new data;
5. All new data as well as all XOR data must be written to the appropriate drives.

The diagram below shows a simplified example of the steps required to conduct a RAID5 ‘Read Modify Write’.

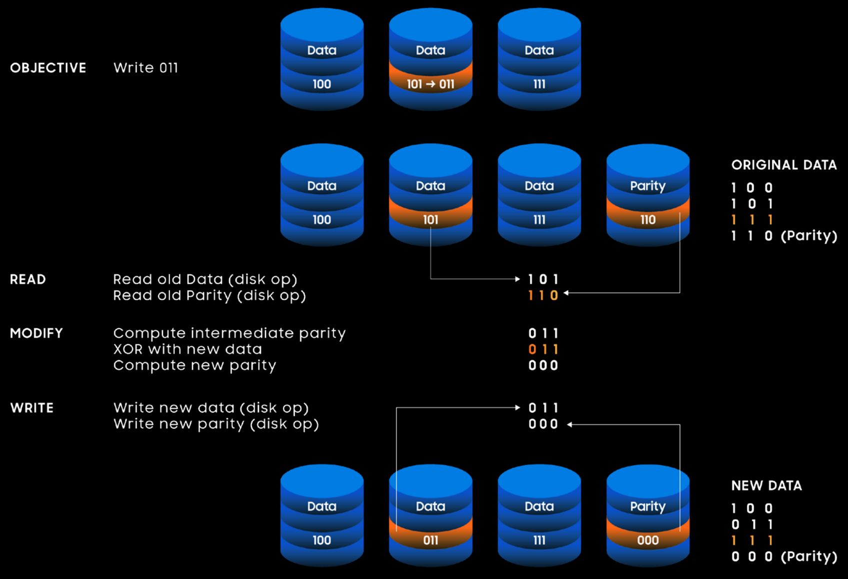


Figure B - RAID5 Read Modify Write Operation

As you can see, each RAID5 write needs to translate into an entire series of read/write and compute steps. To achieve maximum IO performance while using any RAID5 protection methodology, enterprise storage drives must deliver very high reads and writes. Coupled with a RAID controller, they must be able to execute the entire ‘Read Modify Write’ sequence at the same demanding level of performance as other critical operations.

A new storage solution combining Samsung’s PM1653 24G SSD and Broadcom’s 9670W 24G RAID controller accomplishes this requirement nicely. It increases the effectiveness of RAID5 by taking advantage of the very high read and write performance of a 24G SSD, and by automating all of the operations described above. In this way, the RAID5 approach can become highly cost effective for mission-critical data protection with stellar enterprise performance.

3. Innovations in PM1653

As the world leader in SAS storage, Samsung has been offering highly advanced and reliable enterprise storage solutions for a decade. Its latest offering, PM1653, is an exceptionally high-performing 24G SAS SSD. Based on the most recent SAS interface, the PM1653 is twice as fast as the previous 12G SAS storage (PM1643a). The PM1653 is also the first 24G SAS SSD made with sixth-generation (100+ layer) V-NAND technology, enabling storage capacities from 800GB to 30.72TB.

Samsung PM1653 employs an idle power-saving scheme and can be kept to less than 5W to meet the EU Eco regulation. Furthermore, the PM1653 uses a native 22.5Gbps SAS 4.0 Rhino controller with a dual port and a 2.5 inch form factor, which allows the drive to plug into almost any existing enterprise server system. The Rhino controller features command decoding and has a DMA description for its hardware automation to maximize performance while minimizing firmware interference.

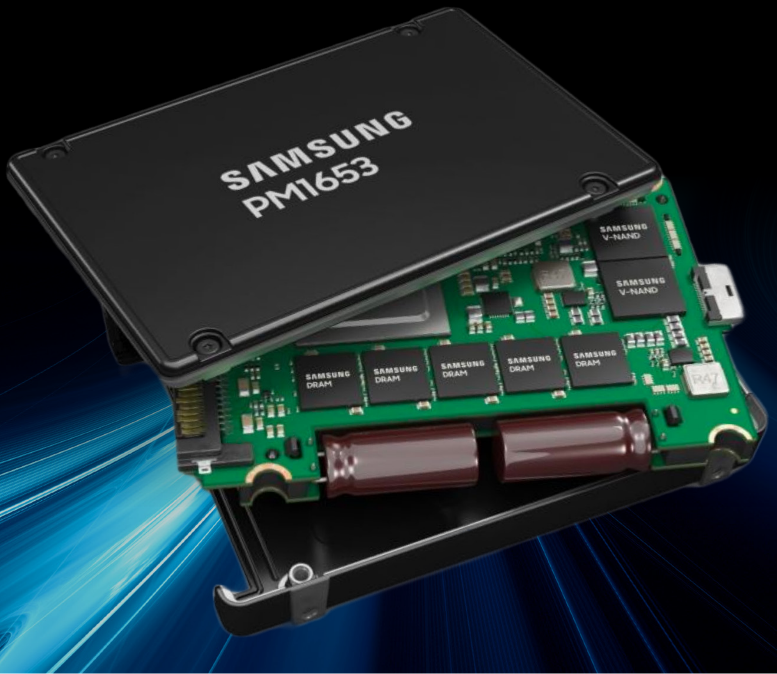
The PM1653 offers a high random read speed — a key metric for server storage performance — tested at up to 740K IOPS. It also delivers a sequential read speed of up to 4,300MB/s — the maximum available speed for the 24G SAS interface. Leveraging a dual port, the new SSD provides enterprise server OEMs with the flexibility of using one or both ports depending on the system environment. In the rare instance of a port experiencing a failure during operation, data can be transferred to and accessed through the other port with enterprise-grade reliability. The PM1653 can support 24G and the legacy 12G SAS platform. By using 3D V-NAND TLC, the PM1653 is capable of storing many terabytes of data every day on a single drive over a period of five years without experiencing a failure.

Overall, Samsung has optimized the PM1653's architecture with premium DDR4 DRAM for metadata and cache, a high-end controller, and highly advanced FW. In collaboration with Broadcom, Samsung is now able to considerably enhance PM1653 communication over the demanding SAS4 interface. These enhancements enable the PM1653 to run with a very high level of efficiency when connected to Broadcom's SAS4 Expanders. Additionally, SAS Expanders give users the opportunity to scale to higher drive counts.

Storage Capacity
up to **30.72TB**

Random Read Speed
up to **740K IOPS**

Sequential Read Speed
up to **4,300MB/s**



4. Innovations in 9670W-16i

The evolution of the data center is now being driven by increased network performance, higher CPU frequency and greater core counts. Solid state drives (notably the PM1653) as a storage medium are driving AID controllers to take storage performance to new levels of achievement in improving data center capabilities. To allow this, Broadcom's 9600 Series went through a revolutionary architecture change from supporting previous generations of ROC/IOC controllers to the goal of achieving breakthrough performance by focusing on six core metrics.

Key innovations include the following*:

- HW Automation and RAID: In previous generations, FW was required to help in the processing of RAID IOs, which limited maximum achievable performance. Now, the 9600 Series leverages HW acceleration almost exclusively, pushing performance well beyond levels enabled by FW.

- HW Automation and Cache Flush: Previously, FW was required to help flush the cache. Due to the slow nature of FW, the write cache would fill up quickly. This would seriously impact IO and latency. However, the 9600 Series with HW Automation has no such impediment.

Consider the six key metrics derived from these advancements by focusing on end-user applications and the requirements needed to best improve performance. Refer to Table 1 below.

	Performance	9400 Series	9500 Series	9600 Series
4x	Bandwidth (MiBs)	6,850	13,700	27,900
4x	Random Read (IOPS)	1.7M	3.0M	6.4M
5x	RAID5 Ramdom Write (IOPS)	185K	240K	1.1M
10x	Rebuild Time Under Load (Min/TB)	~300	~90	~33
60x	Write Latency (μs)	>500	>200	8
80x	Performance Under Rebuild (IOPS)	30K	43K	2.7M
	Application Performance	1x	2x to 3x	5x to 12x

Table 1 - Broadcom 9600 Series Performance Comparison

5. How Samsung and Broadcom are Optimizing the Use of 24G SAS for Today's Applications

The innovation collectively deployed in Samsung's PM1653 and Broadcom's 9670W-16i RAID controller provides performance exceeding one million RAID5 Write IOs per second using as few as 20 SAS4 Samsung SSDs (3.2TB PM1653). When configured with five drive volumes, RAID5 is able to retain approximately 80% of total capacity for the user, compared to RAID1 which only allows access to approximately 50% of installed capacity.

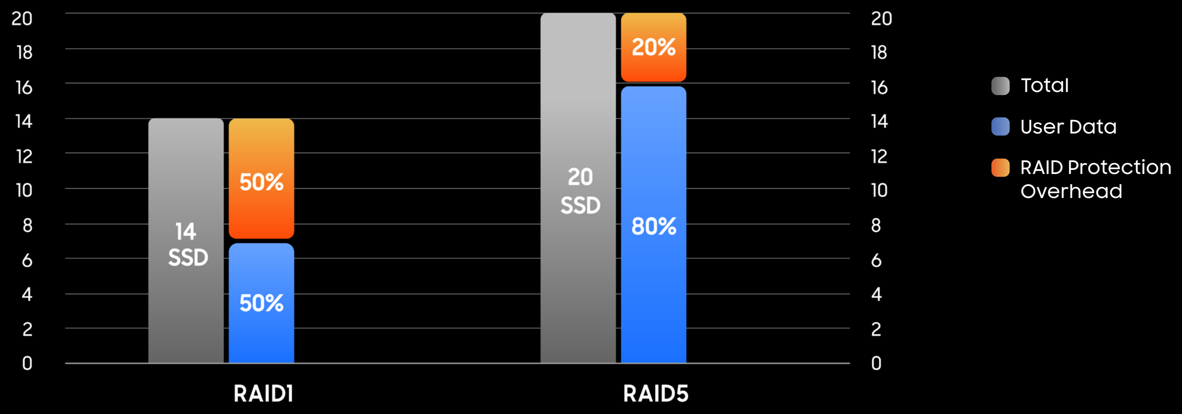


Figure C - SSD Count for 1 Million 4K RW IOPs

At 1 million IOPs, RAID5 provides more than double the user capacity of RAID1. A user can access 16 SSDs for RAID5 compared to only 7 with RAID1. At the same time, RAID5 requires 37% less storage per usable terabyte, while providing an equal amount of performance and protection against a single drive failure.

RAID	Total SSD Required	SSD Overhead for Protection	User Accessible SSD	User Capacity	Drives per User TB	Savings
RAID1	14	7 (50% of SSDs)	7 (50% of SSDs)	22.4TB (7*3.2TB)	0.625 (14SSD / 22.4TB)	0%
RAID5	20	4 (20% of SSDs)	16 (80% of SSDs)	51.2TB (16*3.2TB)	0.391 (20SSD / 51.2TB)	37% (625 - 391) / 625

Table 2 - Efficiency Comparison Between RAID1 and RAID5 @ 1M IOPs

Samsung's PM1653 delivers the 4K random read and write IO performance required to take full advantage of the performance of Broadcom's 9670W-16i RAID controller, while breaking the 1M RAID5 4K random write barrier.

6. Conclusion

24G SAS is in a unique position to provide a very reliable, highly scalable storage infrastructure upon which Broadcom and Samsung can address specific performance challenges. The pairing of Samsung’s PM1653 with Broadcom’s 9670W 24G RAID controller can deliver over one million RAID5 random write IOPS. This improvement in performance can greatly enhance the user experience for online transactional workloads such as when employing Microsoft SQL.

*(For more information about Broadcom 9600 Series RAID I/O controller performance contact your Broadcom Sales representative.)

Disclaimer

This White Paper is proprietary to Samsung and Broadcom – all rights reserved. This information shall not be duplicated or disclosed, in whole or in part, by any means, person or organization, for any purpose, without express written permission by Samsung and Broadcom. Neither receipt nor possession of this White Paper, from any source, constitutes such permission. Infringement of this proprietary right may result in criminal and/or civil liability.

Samsung and Broadcom have made reasonable efforts to ensure the accuracy of the information contained in this White Paper. However, it is presented AS IS and Samsung and Broadcom make no warranties, conditions, or representations of any kind, either express or implied, with respect to the content or the completeness or accuracy of this White Paper. IN NO EVENT SHALL SAMSUNG OR BROADCOM BE LIABLE TO ANY PERSON FOR ANY SPECIAL, INCIDENTAL, INDIRECT OR CONSEQUENTIAL DAMAGES, ARISING OUT OF OR IN CONNECTION WITH THE RELIANCE, USE OR PERFORMANCE OF THE INFORMATION PRESENTED HEREIN.