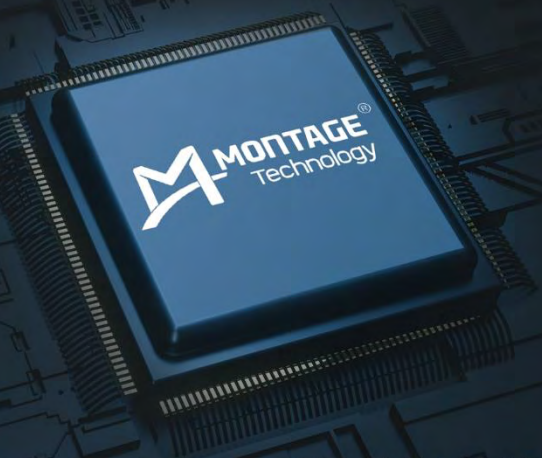
Samsung PCIe Gen5 NVMe SSD PM1743 with Montage PCIe 5.0/CXL 2.0 Retimer

High-Performance Signal Integrity Solution

White Paper



SAMSUNG

1. Background

With the growth of big data, cloud computing, 5G and other technologies, the demand for data computing and efficient storage is growing exponentially, and more terminal devices require high bandwidth and stable data transmission, so PCIe has become the mainstream solution for server buses with its high transmission rate. Currently, the transmission rate of PCIe bus has increased from 2.5GT/s in the first generation to 32GT/s in the fifth generation. PCIe Gen5-enabled devices can be more in line with the current needs of data centers and enterprise-level applications, and PCIe Gen5-related products have been launched gradually in various fields. Samsung's high-performance PCIe Gen5 NVMe SSD PM1743 was released in 2021, Montage started mass production of PCIe 5.0/CXL 2.0 Retimer in January 2023, and PCIe Gen5 enterprise servers from different vendors have all been released one after another. It is foreseeable that PCIe Gen5 will become the global market trend in the next few years, and the era belonging to PCIe Gen5 will soon arrive in full swing.

The extremely high signal transmission rate enables PCIe Gen5 to better support high-performance devices with ultra-high throughput requirements. However, facing the exponential growth of signal transmission rate in each generation of PCIe, the problem of signal attenuation and deformation has become more and more obvious and the stability and integrity of signal transmission has become the bottleneck in signal transmission quality, which limits further speedup of transmission.

In the field of data center storage applications, NVMe SSDs combined with PCIe interfaces can optimize the data transmission path, greatly increase the data transmission bandwidth, and reduce read and write latency, making them the preferred choice for enterprise storage devices. At present, the mainstream communication system link composition of PCIe Gen5 NVMe SSD includes CPU, motherboard PCB, PCIe expansion card PCB, cable and SSD, etc. Its total insertion loss significantly exceeds the end-to-end permissible channel loss budget of 36dB stipulated in the PCIe Gen5 standard, and it has become imperative for us to look for reliable link budget compression or high-performance link expansion solutions to minimize channel loss. Common high-performance link extension solutions on the market mainly involve the high-speed PCBs, Redriver or Retimer. As a protocol-aware device, Retimer can enhance the signal integrity between the server and SSD and improve transmission quality due to its good signal processing. Furthermore, it has lower costs compared with high-speed PCB boards, thus becomes the best solution for the system signal integrity design, providing a compatible solution for server OEMs to facilitate the design and development of servers.

This white paper focuses on the signal integrity solution based on Samsung PCIe Gen5 NVMe SSD PM1743 and Montage PCIe 5.0/CXL 2.0 Retimer M88RT51632 launched by Samsung and Montage Technology. The solution was tested for link stability and read/write performance on PCIe Gen5 servers to address the signal integrity challenges of PCIe Gen5 NVMe SSD system link design.

2. Hardware Configuration

2.1 Samsung PCIe Gen5 NVMe SSD PM1743

In 2021, Samsung released a high-performance SSD with its advanced V-NAND flash technology and the latest PCIe Gen5 interface: PM1743 (see Figure 1). It surpasses the predecessor in terms of performance and energy efficiency, and the core features are shown below. With these features, PM1743 is more than capable of handling large-load enterprise-class work environments, and will become the preferred cost-efficient choice in data centers.

- **Superior Performance:** PM1743 uses the new PCIe Gen5 interface with sequential read/write speeds up to 14GB/6GB per second and random read/write speeds up to 2500K/280K IOPS, and yields a 2x performance increase over PCIe Gen4 products.
- **Ultra-high Power Efficiency:** The power efficiency of PM1743 is improved by 40% compared with that of the previous generation, reaching up to 657MB/s per watt.
- **Cutting-edge Technology:** By virtue of Samsung's advanced V-NAND flash technology, PM1743 provides enterprise servers with a storage solution that offers greater capacity and performance with less latency.
- **Diverse Form Factors:** With the capacities ranging from 1.92TB to 15.36TB, PM1743 provides a traditional 2.5-inch form factor and an EDSFF (E3.S) package size (7.5mm thick)—an increasingly popular SSD form factor designed for next-generation enterprise-class servers and data centers, doubling the storage density of enterprise-class servers.
- **High Availability:** PM1743 supports dual-port operation, ensuring operational stability and business continuity of servers even if one port connection fails by shifting workloads to the second port, and thus facilitating consistent operations and high availability of servers and storage arrays.
- **High Reliability:** PM1743 utilizes enhanced telemetry for more effective remote monitoring and analysis, provides multi-tenancy capabilities without sacrificing performance, and ensures more secure storage with encryption and decryption authentication in order to provide excellent reliability for users.

Samsung PM1743	
Form factor	U.2/E3.S
Interface	PCIe 5.0 x4
NAND	V-NAND Technology
Port	Dual
Data Transfer Rate (128KB data size, QD=64)	
Seq.R/W (GB/s)	14/6
Data I/O Speed (4KB data size, QD=64)	
Ran.R/W (IOPS)	2500k/280k
Latency (sustained random workload, QD=1)	
Ran.R/W (us)	60/20
Capacity (TB)	2/4/8/16

Figure 1: Samsung PM1743

2. Hardware Configuration

2.2 Montage PCIe 5.0/CXL 2.0 Retimer M88RT51632

Montage PCIe 5.0/CXL 2.0 Retimer chip uses advanced signal conditioning technology to improve signal integrity and increase the effective transmission distance of high-speed signals. Compliant with PCI-SIG and CXL specifications, this chip adopts the industry mainstream package and becomes the first in the industry to support ultra-low transmission latency of less than 5ns with a transmission rate of up to 32 GT/s. The chip also supports complex system topologies such as SRIS and Retimer cascades, making it an ideal solution for addressing PCIe/CXL signal integrity challenges in next-generation servers, enterprise storage, and AI acceleration systems.

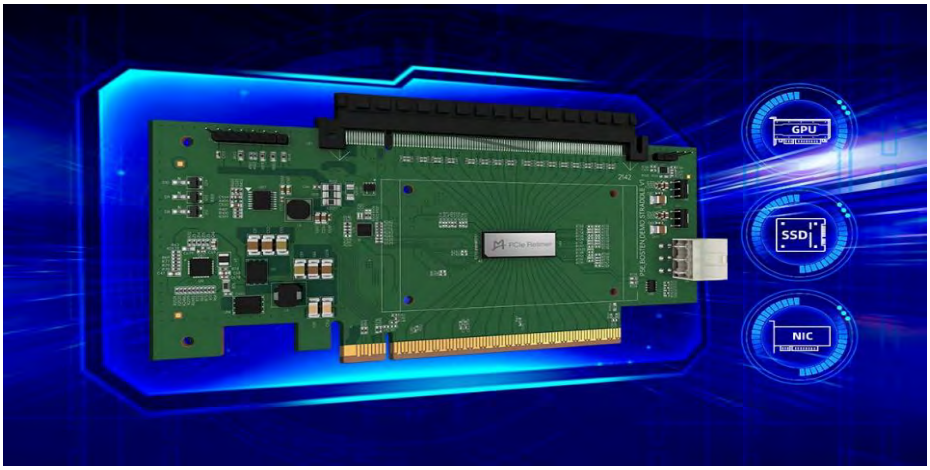


Figure 2: Montage M88RT51632

2.3 Signal Integrity Solution Based on PM1743 and M88RT51632

Samsung and Montage have jointly launched a signal integrity solution based on Samsung PCIe Gen5 NVMe SSD (PM1743) and Montage PCIe 5.0/CXL2.0 Retimer (M88RT51632), as shown in Figure 3. With a PCIe expansion card equipped with the M88RT51632, it ensures that the end-to-end link loss budget is less than 36dB, which is in line with the specification, and further eliminates the impact of crosstalk and reflections and other undesirable factors to ensure a stable and reliable system operation.

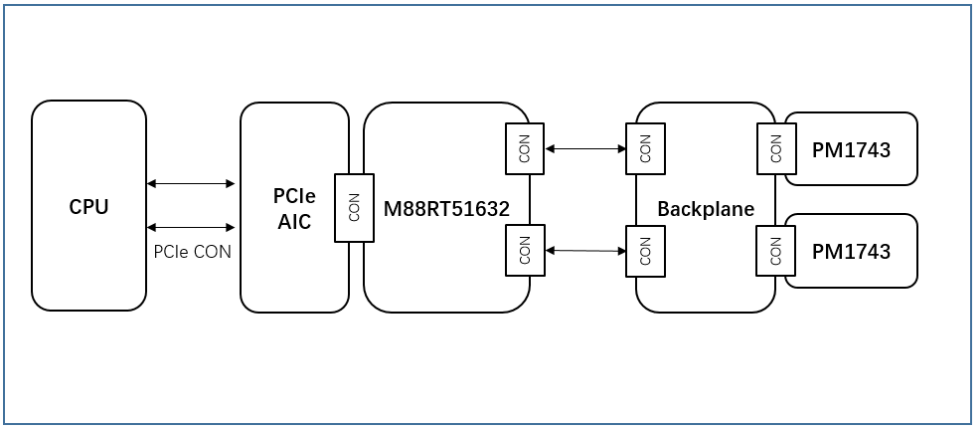


Figure 3: Samsung-Montage PCIe Gen5 NVMe SSD Solution

3. Test Configuration

We tested the following scenarios on signal integrity solution of Samsung PCIe Gen5 NVMe SSD PM1743 + Montage PCIe 5.0/CXL 2.0 Retimer M88RT51632 respectively on a PCIe Gen5 server.

- **Scenario 1: Link Stability Testing**
- **Scenario 2: Read/Write Performance Testing**

3.1 Link Stability Test Configuration

We mainly performed the link stability-related tests based on Intel CScripts and Power Cycle. Intel CScripts tests are performed during motherboard boot-up or debugging, and can handle memory errors, crash dumps, and other catastrophic errors encountered when debugging the firmware. Power Cycle tests can verify the stability and reliability of the system after multiple power-ups or reboots. The specific test parameter settings for Intel CScripts are shown below.

Table 1: Intel CScripts Test Parameters Setting

Test Items	Remarks	Value
Server	N/A	CPU: Intel Sapphire Rapids SP OS: CentOS 8.3.1-5
Intel CScripts	N/A	630113-EagleStream-CScripts-Rev2232-6000-Build2585
IOMT Version	N/A	632111/Windows_rev2p0p24
UEFI Shell Version	N/A	2.2
Reset Tests	Hot Reset, Secondary Bus Reset(SBR) Loop	Loop Counts:10k
Power Management	L1 PM(D3hot), L1 ASPM	Loop Counts:10k
New Feature	Dynamic Link Width	Loop Counts:10k
Link Training	Recovery via link retrain Loop Link Disable/Enable Loop TxEq Redo, Speed Change Loop	Loop Counts:10k

3. Test Configuration

3.2 Read/Write Performance Test Configuration

For the read/write performance test scenarios, we performed bare drive and database-based performance tests on Samsung PCIe Gen5 NVMe SSD PM1743 under this solution respectively, and compared the performance test results with the PM1743 without Montage M88RT51632 under the same test configurations to observe the performance differences.

3.2.1 FIO Test Configuration

In the bare drive performance test, we adopt FIO, a workload generator mainly used for SSD benchmark performance test and stress test. In order to achieve the expected performance of SSD, the test parameters of FIO should be properly configured. The FIO test methodology in this white paper refers to the ODCC enterprise SSD test standard, and the specific test parameters setting are shown below.

Table 2: FIO Parameters Setting

Parameters	Remarks	Value
Server	N/A	CPU: AMD EPYC 9554 OS: ubuntu22.04
FIO Version	N/A	3.35
Numjobs	Number of jobs carrying the same workload	Sequential read/write Numjobs = 1 Random read/write Numjobs = 16
QueueDepth	Queue depth required by asynchronous IO Engines	Sequential read/write QueueDepth = 64 Random read/write QueueDepth = 64
Block Size	The data size in one IO operation	Sequential read/write Block Size = 32/64/128KB Random read/write Block Size = 4/32/64KB

3.2.2 MySQL Test Configuration

In this white paper, the mainstream relational database MySQL is selected for testing. MySQL has the advantages of high performance, high reliability, and cross-platform flexibility, and has become a widely-used open-source database around the world thanks to its powerful online transaction processing capabilities.

The test tool is sysbench, an open-source, modular, cross-platform multi-threaded performance testing tool used for CPU, memory, disk I/O, threads, and database performance testing. It comes with a variety of workloads that

can be pressure tested in different scenarios. The specific test parameters setting in this white paper are shown in Table 3.

Table 3: MySQL Parameters Setting	
Server	CPU: AMD EPYC 9554/OS: ubuntu22.04
DB Version	MySQL-5.7.36
Test Tool	sysbench-1.0.20
Data Size	65% of SSD
Testname	oltp_read_write.lua
Threads	No more than CPU cores
Test Time	3600 Seconds
Test Result	TPS (transactions per second)/Average Latency

3.2.3 RocksDB Test Configuration

In the non-relational database tests, we selected RocksDB for testing. RocksDB relies on its fast and low-latency storage capability to fit different workloads and meets various data needs from a database storage engine, application data caching, embedded workloads, etc.

The testing tool is YCSB, an open-source NoSQL database performance testing tool that can test cloud databases. It comes with six workloads, and the parameters in the workloads can be modified to meet different testing requirements. The specific testing parameters setting for RocksDB in this white paper are shown below.

Table 4: RocksDB Parameters Setting	
Server	CPU: AMD EPYC 9554/OS: ubuntu22.04
DB Version	RocksDB-6.22
Test Tool	YCSB-0.18
Data Size	65% of SSD
Testname	Workloada(50%/50%,Reads/Writes)

3. Test Configuration

Table 4 (continued): RocksDB Parameters Setting	
Threads	No more than CPU cores
Test Time	3600 Seconds
Test Result	Throughput/Average Latency

4. Test Results

4.1 Link Stability Test Results

Referring to the parameter settings in Table 1, we tested the Samsung PM1743 with and without the Montage M88RT51632 respectively. Each item in Table 1 was tested 10,000 times, and Power Cycle tests were conducted 1,000 times.

The results of the tests are shown in Table 5. During the Power Cycle test of the Samsung PM1743 with Montage Retimer, there were no problems such as SSD being accidentally disconnected or unrecognized, meanwhile, the bandwidth and speed of the PM1743 were not reduced compared with using it alone. Intel CScripts test results also show that using Retimer has had no extra impact on the performance and functionality of the PM1743, and are consistent with the results of using the PM1743 without Retimer. This shows that the PM1743+M88RT51632 solution improves the quality of information interaction between the PCIe slots on the motherboard and the CPU while guaranteeing stable system operation and increasing the reliability of the system signals in a simulated real-world usage environment.

Table 5: Link Stability Test Results

Test Items	PM1743+M88RT51632	PM1743
SBR	10k pass/10k cycles	10k pass/10k cycles
Link Retrain	10k pass/10k cycles	10k pass/10k cycles
Link Disable/Enable	10k pass/10k cycles	10k pass/10k cycles
D3 Hot	10k pass/10k cycles	10k pass/10k cycles
TxEq Redo	10k pass/10k cycles	10k pass/10k cycles
Speed Change Retrain	10k pass/10k cycles	10k pass/10k cycles
Speed Change Max	10k pass/10k cycles	10k pass/10k cycles
Power Cycle	1k pass/1k cycles	1k pass/1k cycles

4.2 Read/Write Performance Test Results

4.2.1 FIO Test Results

In this white paper, we tested the difference in bare drive performance between the Samsung PM1743 used in conjunction with the Montage M88RT51632 and the Samsung PM1743 used alone. FIO-based read/write performance tests were primarily conducted with test items including 32KB/64KB/128KB sequential performance

tests and 4KB/32KB/64KB random performance tests.

The test results are shown in Figure 4. Compare to the bare drive performance of PM1743, it has no significant effect on the sequential bandwidth and random IOPS for different blocksize with Retimer. The sequential read and write performance fluctuation stays around 2% at the maximum, and the random read and write IOPS gap is 0.08% at the maximum. The average latency in the 4KB random read almost no difference, 4KB random write gap of 6%. This shows that the system read and write performance of Samsung PM1743 using Retimer maintains good consistency with the bare drive performance, and the solution meets the specification requirements.

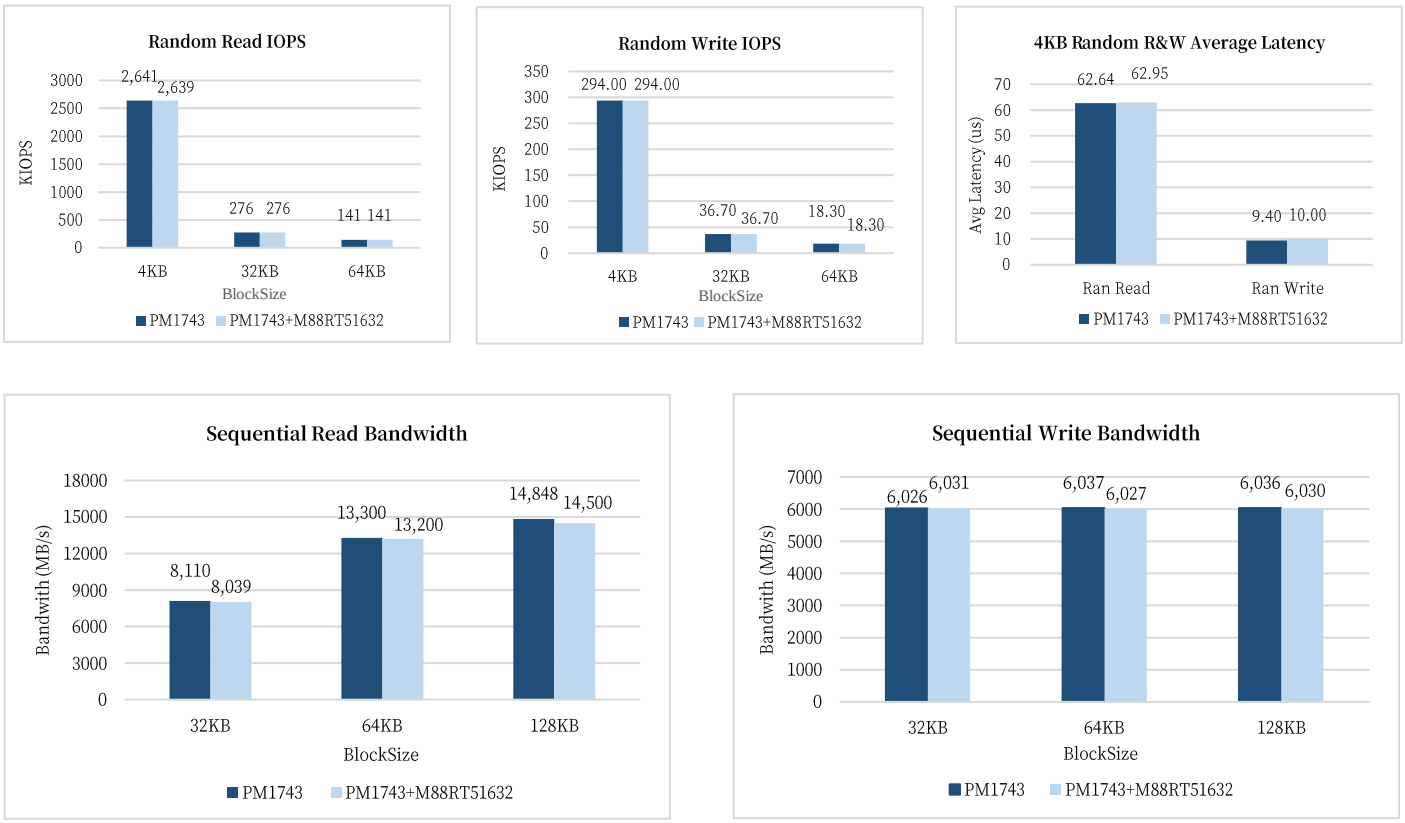


Figure 4: FIO Test Results

4.2.2 MySQL Test Results

In the test of mainstream relational database MySQL, we have tested PM1743 with and without Montage M88RT51632 through sysbench respectively. In order to be closer to the real workload, we use oltp_read_write.lua provided by sysbench to simulate the test environment in this white paper, and select the TPS and average latency in the test results as the metrics.

The test results are shown in Figure 5. Compared with the MySQL test results of PM1743 without Retimer, the changes in TPS and average latency are maintained at around 1.5%, which indicates that the system link of this solution is stable and the use of NVMe SSD together with Retimer has not caused any loss in the performance of the MySQL database application.

4. Test Results

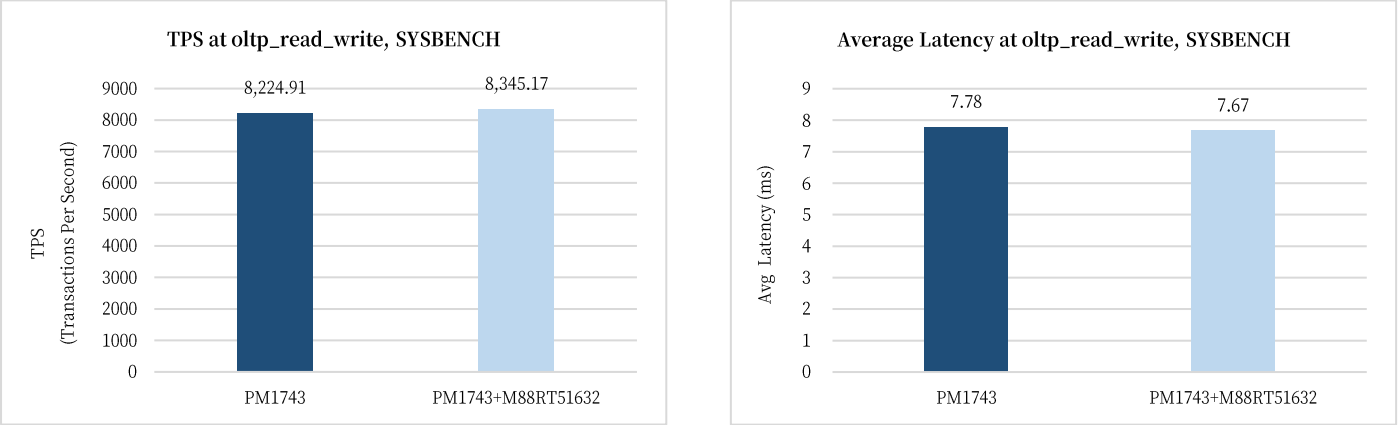


Figure 5: MySQL Test Results

4.2.3 RocksDB Test Results

In the test of the non-relational database RocksDB, we tested the solution using PM1743 alone and the solution equipped with Montage M88RT51632 under the same server configuration as the MySQL test, and selected the throughput and average latency in the test results as the metrics.

The test results in Figure 6 show a small change in both the throughput and average latency of RocksDB. For the throughput data, there is a fluctuation of about 9% for PM1743 directly connected to the PCIe slot compared with PM1743 using Retimer, while the average latency of both read and write operations has a change of about 7%. For the overall RocksDB test results, the solution's performance and latency variations are maintained within 10% as expected.

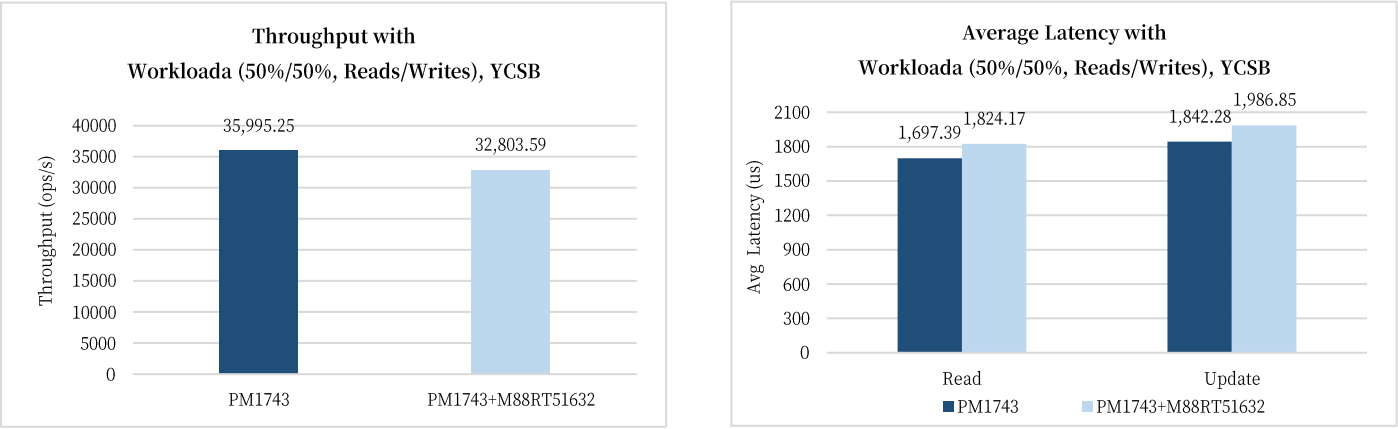


Figure 6: RocksDB Test Results

SAMSUNG

5. Conclusion

In this white paper, we have validated the signal integrity solution based on Samsung PCIe Gen5 NVMe SSD (PM1743) and Montage PCIe 5.0/CXL 2.0 Retimer (M88RT51632), and conducted link stability tests and read/write performance tests on the PCIe Gen5 server platform respectively. The link stability test results show that the solution can fully ensure the stability and integrity of the system signal. Read and write performance test results show that whether in the state of bare drive or in the state of carrying applications, the performance of the solution can achieve the expected results. Compared with the benchmark performance of NVMe SSD directly connected to the PCIe expansion card, the performance fluctuations of NVMe SSD are within the expected range when using the PCIe expansion card equipped with PCIe Gen5 Retimer. This solution meets current high-bandwidth and low-latency requirements, fully guarantees system link stability and system signal integrity, enhances the data transmission quality of NVMe SSD during read and write processes, and increases the effective transmission distance of high-speed signals.

Faced with cloud computing, cloud service vendors, and data centers' demands on high-speed reading and interaction, PCIe Gen5 NVMe SSD (with higher IO rate and lower energy consumption) has gradually become the first choice in building basic deployment of storage devices for major manufacturers in the PCIe Gen5 era. PCIe Gen5 not only provides strong support for devices with high throughput requirements, but further reduces the signal transmission attenuation distance. The solution of Samsung PCIe Gen5 NVMe SSD + Montage PCIe 5.0/CXL 2.0 Retimer guarantees the quality of information interaction between some of the farther PCIe slots on the motherboard and the CPU, while at the same time taking into account the needs of server OEMs, ODMs, and terminal users for system capacity, read/write bandwidth, and flexible topology. With further improvement of the PCIe Gen5 ecosystem, the demand for signal integrity solutions will greatly increase, and the solution will continue to evolve for broad use accordingly.