White Paper

What's New in 24G SAS

July 2018

Contents

Introduction

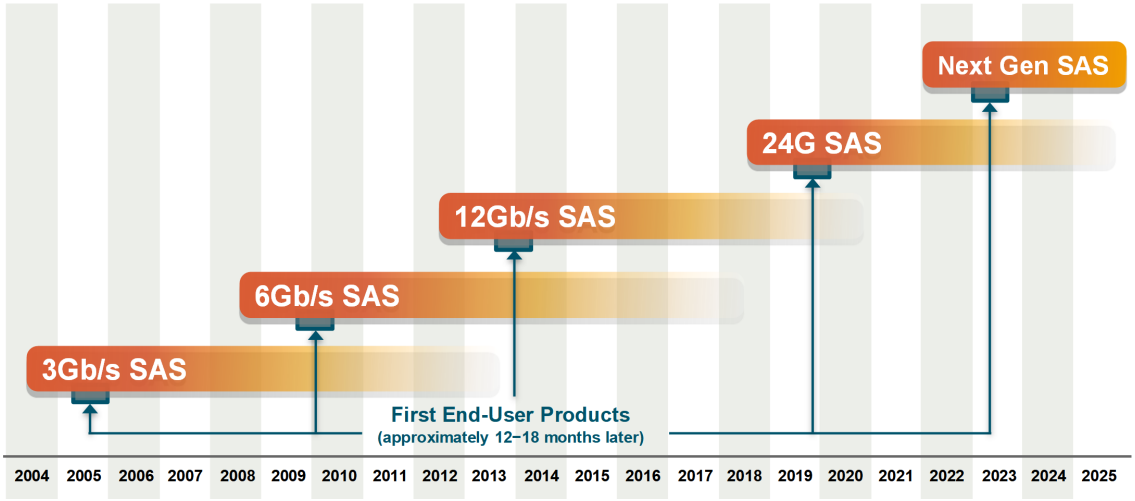
Serial attached SCSI, commonly known as SAS, has long been the interface of choice for moving data between the computing and storage sub-systems. SAS is a point-to-point serial protocol that uses the SCSI command set and is compatible with the Serial ATA (SATA) standard. The SCSI Trade Association, the body that governs the development of the SAS standard, has announced the next generation of SAS, 24G SAS (sometimes referred to as SAS-4, when in fact, 24G SAS is described by a collection of standards, including SAS-4). This revision of the standard includes protocol and block-level enhancements to take advantage of the changes in media technology and usage models. 24G SAS infrastructure is well positioned to enable end customers to take their key storage infrastructure use cases to the next level. This paper outlines the new features in 24G SAS and the performance benefits they provide from a storage system point of view.

24G SAS, while a major overhaul of the technology, continues to support backward compatibility for two generations, 12 Gbps SAS and 6 Gbps SAS, and leverages the existing ecosystem of tools and test equipment. 24G SAS extends the useful life of legacy equipment by providing significant performance improvements through aggregation while maintaining the low cost economics of the SAS infrastructure. As a result, new storage system designs can benefit from 24G SAS technology even with legacy SAS /SATA devices. It preserves the existing value proposition of SAS as an Enterprise-proven transport that can support all tiers of media storage with unparalleled reliability, availability, scalability, and manageability. While doubling the link bandwidth, it continues to provide low latency access to media that spans various tiers of performance and capacity, ranging from Enterprise class solid state drives and workhorse SAS HDDs to cost-optimized SATA HDDs.

The importance of higher bandwidth to the storage media has increased due to the pending availability of PCIe Gen 4 interfaces on mainstream CPUs. As the interface bandwidth between storage controllers and the CPU increases as a result of PCIe Gen 4, a bottleneck is created between the storage controller and the storage media when using 12G SAS. 24G SAS alleviates that bottleneck and allows the PCIe Gen 4 bus to the CPU to be saturated under full system load. Microsemi's innovative new Dynamic Channel Multiplexing (DCM) technology and enhanced SAS/SATA buffering technology enable full-use of the 24G SAS uplink between the SAS ROC/IOC and SAS expanders even when using lower speed 6G/12G drives.

The following figure shows the evolution of SAS technology.

SAS Technology Roadmap



24G SAS Characteristics

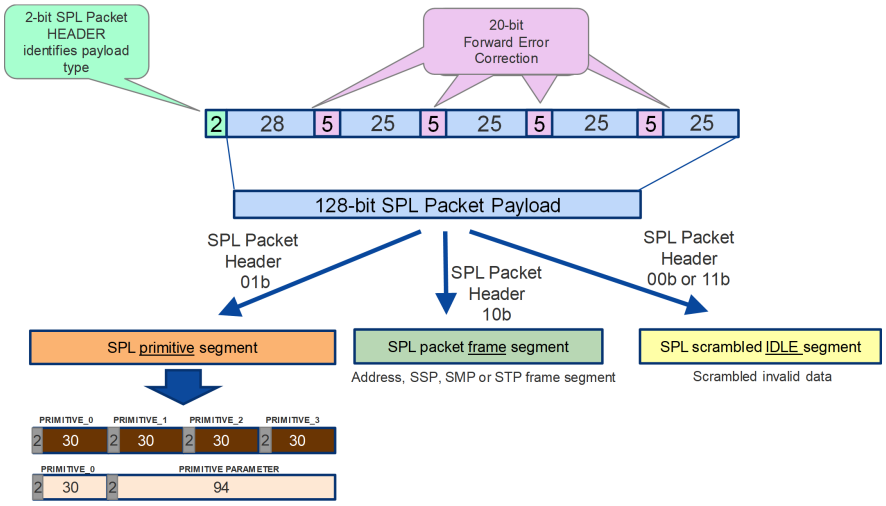
This section describes the changes to the Physical, PHY, and Link layers of the SAS protocol stack in order to double the link bandwidth, and are described in the SAS-4 and SPL-4 standards. The SAS-4 standard documents the electrical and functional requirements for the physical interconnect to operate at 24G SAS. The SPL-4 (SAS Protocol Layer 4) standard documents the protocol layers corresponding to 24G SAS and is intended to be used in conjunction with the SCSI and ATA command set standards.

Encoding Scheme

SAS-4 supports a line rate of 22.5 Gbps. This is a departure from previous generations in which the line rate was doubled with every generation. The SCSI Trade Association calls it 24G SAS to avoid confusion. SAS-4 achieves the doubling of the link bandwidth by changing the underlying data encoding scheme from the SAS dword mode, which uses 8b/10b encoding, to the SAS packet mode. All information transferred in the SAS packet mode is encoded into SPL packets that are 150-bit blocks that use 128b /150b encoding. SAS-3 operates at a baud rate of 9.6G baud (12 Gbps bit rate × 8b/10b encoding). SAS-4 operates at a baud rate of 19.2G baud (22.5 Gbps bitrate × 128b/150b encoding), which is effectively twice the baud rate of 12G SAS.

Each SPL packet consists of two header bits, four dwords (128 bits), and 20 bits of Reed-Solomon Forward Error Correction (FEC) code. The intent of the FEC bits is to detect all errors and correct up to two error bursts per packet. This provides an extra 5 dB of channel loss. DC balance and run length are maintained statistically by the scrambling algorithm and primitive encodings. CRC protection at the frame level is maintained. The following diagram shows the encoding changes in 24G SAS.

Encoding Changes



Packet Format

The SPL packet starts with a two-bit header that specifies the format of the subsequent packet payload as shown in the preceding figure. The packet payload can contain a scrambled idle segment, an idle dword segment, an SPL frame segment that contains data dwords or a primitive segment. 24G SAS defines a scrambled idle segment that is used for rate matching if the negotiated logical link rate in the pathway is faster than the requested connection rate. The idle dword segment is transmitted when the transmitter has no other data to transmit and serves the same purpose as in the SAS dword mode.

Only primitives can be encoded as payload in packets with header bits of 01. Unused primitive slots are populated with ALIGNs. 24G SAS also defines binary primitives, extended binary primitives, and primitive parameters. The binary and extended binary primitives serve various control and framing requirements in packet mode. Primitive parameters are a way to deliver additional information by associating parameter data with a primitive. As an example, the CLOSE primitive parameter may be used by an expander to send fairness priority to an attached expander device.

Scrambling

In generations previous to 24G SAS, the SAS dword mode performs scrambling at the link layer. The scrambler is initialized by start of frame primitives (SOF/SOAF/SATA_SOF), and the scrambler is free running for idle dwords outside of the frames. This raises the possibility of creating pathological reverse scrambler sequence frames that can result in signal degradation and possible loss of synchronization. 24G SAS addresses this issue by initializing the scrambler independent of frame boundaries while operating in the SAS packet mode. The PACKET_SYNC extended binary primitive is sent periodically and initializes the scrambler. This means that scrambling in SPL packet mode is performed in the PHY layer instead of in the link layer.

Packet header bits and the forward error correction bits are not scrambled and must not advance the scrambler. Packets with header bits of 00, 11, or 10 scramble the payload and advance the scrambler. Primitives are not scrambled and do not advance the scrambler.

Signal Integrity

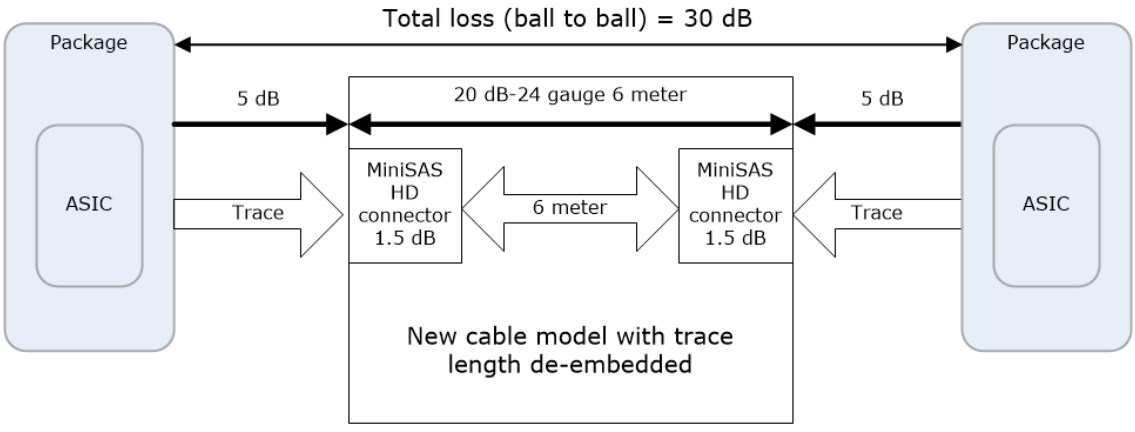
The SAS-4 specification has changed how SAS PHY transmitter characteristics are measured and the permitted amount of channel loss. Transmitter device jitter measurements for trained 22.5 Gbps signals are based on the methods defined in the Optical Internetworking Forum's CEI-25G-LR (OIF-CEI) specification. OIF-CEI is a proven standard that is already in the field and is used in I/O applications that exceed 25 Gbps. This approach significantly simplifies channel modelling by removing the need for special channel model tools like the SAS-3 EyeOpening or SAS-2 SASWDP tools. The transmitter device component edge and the receiver device component edge are defined as the compliance points where the S-parameters are measured for 24G operation. As a result, a designer can now analyze the channel characteristics by comparing them with the loss equations and limits that are published in the SAS-4 standard.

The process of establishing a link between two SAS-4 PHYs at 22.5 Gbps follows the same speed negotiation sequence as defined in SAS-3 with the exception that transmitter and receiver training are performed at the higher line rate. Also, the composition of the Train_Tx-pattern and the Train_Rx-pattern is different between the SAS dword and SAS packet modes.

The following figure shows the SAS-4 channel design and simulation improvements.

Note that the SAS-3 total loss measurement is from die to die. SAS-4 total loss measurement is from ball to ball.

Signal Integrity



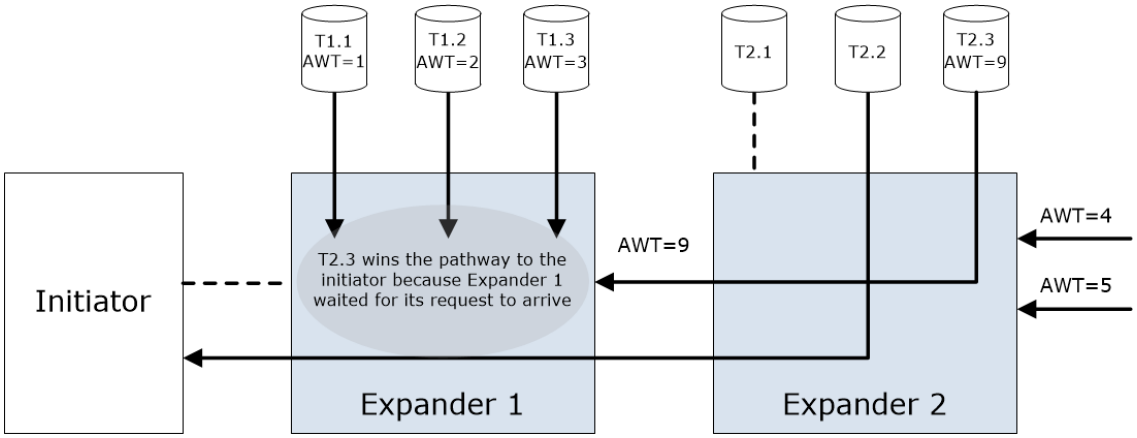
New Protocol Features

This section describes the new protocol features available in the SPL-4 standard.

CLOSE Primitive Parameter

The CLOSE primitive parameter may be used by expander devices to enhance inter-expander fairness. In cascaded SAS topologies, a bias exists that favors target devices that reside close to the initiator. This bias occurs because pending requests close to the initiator can occupy recently freed connection pathways before deeper requests have time to advance. 24G SAS defines a new CLOSE primitive parameter that sends reservation information along with the CLOSE primitive when tearing down a connection. Each expander in the chain compares the OPEN ARBITRATION WAIT TIME in the CLOSE primitive parameter with its local requests and updates the CLOSE primitive parameter with the result. Because each expander in the chain is informed of the highest priority request that is pending deeper in the chain, races are prevented that keep lower priority requests from winning because they are closer to the initiator. The following diagram shows the role of the new CLOSE primitive.

CLOSE Primitive Parameter



OPEN_REJECT Retry-Class Primitive Parameter

When a SAS device is temporarily unable to accept an OPEN Address Frame, it responds with an OPEN_REJECT retry class primitive. In generations prior to 24G SAS, there is no provision through which this device can inform the originator of the rejected OPEN address frame, and specify a time delay to observe before it sends another OPEN request. As a result, the requesting device has to try opening the connection without knowledge of when the recipient device might become available. In SAS-4, any retry-class OPEN_REJECT primitive can include a parameter that specifies a requested delay that the originator observes before retrying the OPEN address frame to the same SAS destination address. The delay can be in the range of 0 ms to 6.5 ms, 0 ms to 655 ms, or 0 sec to 65 sec. This delay provides a level of predictability to the device trying to open a connection and helps with better utilization of the link bandwidth. The OPEN_REJECT retry-class primitive parameter may be used by an end device or an expander device to request the delay.

SMP Frame Priority

Large SAS topologies can sometimes experience Serial Management Protocol (SMP) I/O timeouts and live-lock conditions during drive removals and insertions. Prioritizing SMP will allow discovery and management traffic to propagate through these heavily loaded systems at a faster rate. 24G SAS enables SMP frames to be delivered by raising the Arbitration Wait Time (AWT) priority of SMP connection requests above requests from all other protocol types. This feature does not disrupt normal traffic because SMP traffic is typically small relative to SSP.

Credit Advance

The credit advance feature reduces the time between a device sending an OPEN address frame (OAF) and receiving the first data frame from the destination. If an SSP PHY has receive resources available then it may advance credit by using the CREDIT ADVANCE bit in the OPEN address frame to automatically grant credit immediately when the connection is opened. If the recipient of the connection request implements the credit advance feature, it does not have to wait for a RRDY from the originator before it can send the first data frame. This feature was first introduced in SPL-3 and was revised in SPL-4 to provide additional clarifications.

Buffered PHY Burst Size

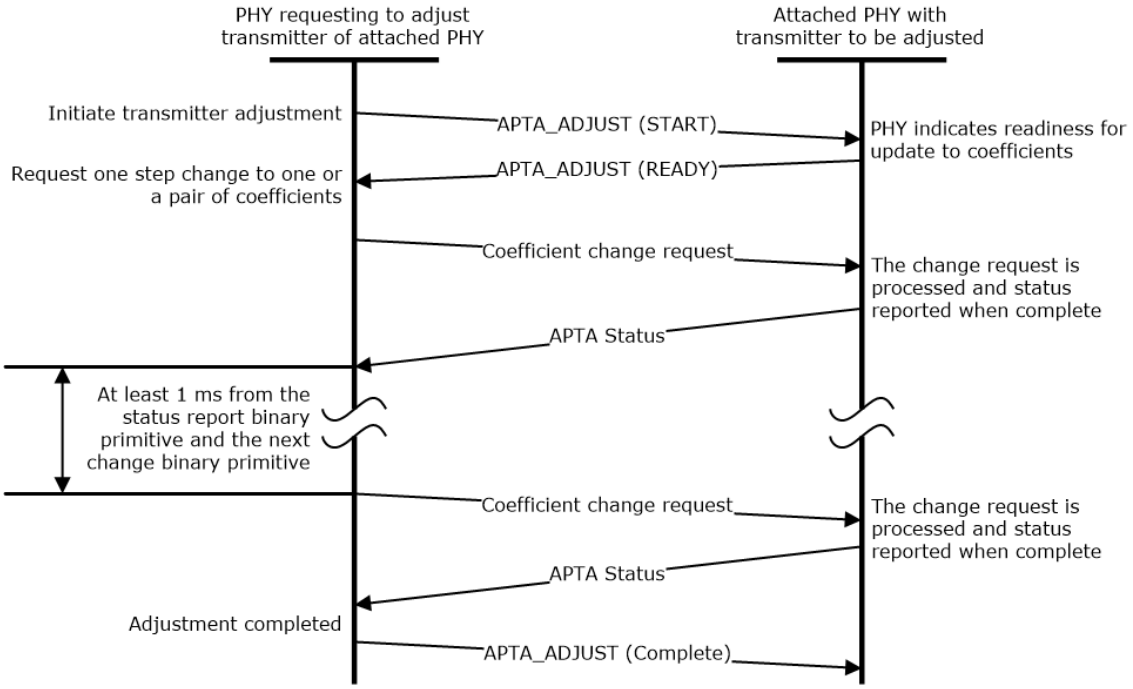
Transfers that are not optimized to match the buffer size in a PHY can result in additional flow control due to mismatches between the amount of data being transferred. In SAS-4, the DISCOVER response format adds a field to indicate the optimum transfer size for the expander PHY if that PHY contains buffers. An initiator device may use this information to optimize write data transfers by matching it to the size of the buffer implemented in the expander PHY. This feature improves the efficiency of expander buffering by exposing the expander buffer size to the SAS initiator such that it can match its requested transfer size for optimal buffer utilization.

Active Transmitter PHY Tuning (APTA)

Previous generations of SAS enabled transmitter training only during link negotiation. Once a link is established, if the operating conditions change, the protocol does not define a methodology to re-train the transmitter during normal operation without resetting the link. This can be troublesome for channels that are close to the limits of the loss budget. SAS-4 addresses this by defining new binary primitives that enable transmitter tuning by sending in-band change requests during normal link operation. This feature can improve link integrity and Bit Error Rate (BER) by making small adjustments to the transmitter to optimize the eye at the receiver. The result is active compensation for channel loss drift due to a change in operating conditions.

APTA primitives are deletable binary primitives and are sent outside of a connection. They are exchanged between attached PHYs to request changes to the transmitter coefficient settings and to report transmitter status without causing a reset sequence. Very small transmitter adjustments with long delays between each request maintain the link at its optimally tuned operating state. The delay allows for the receiver to adapt to the incoming signal over a long period of active data bits before it makes a calculation for the next change request. The following figure shows an example of APTA tuning.

SAS-4 APTA Primitives



Connectors and Cables

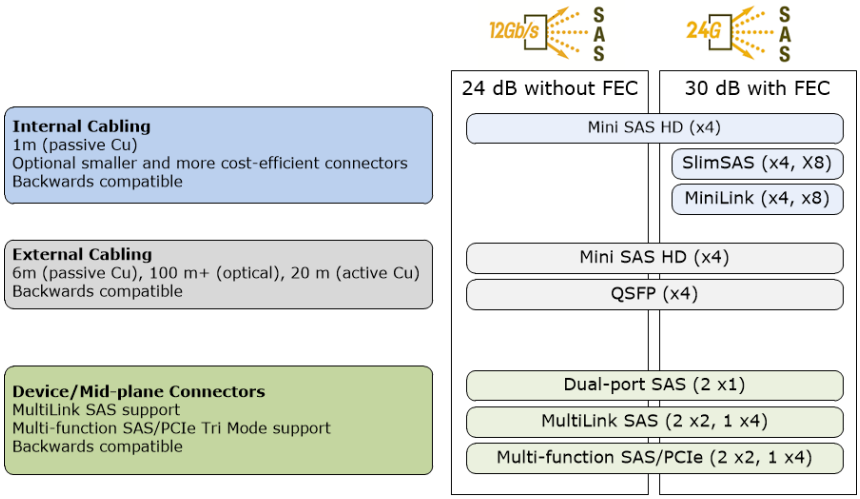
As with each generation of SAS technology, incremental improvement in cables and connectors is required to support the higher data rate of SAS-4 technology. The strong ecosystem of firms supporting the SAS standard has brought to market the full suite of required components to enable a smooth transition from 12G SAS-3 to the upcoming 24G SAS-4 infrastructure.

The current generation of internal (SFF-8643) and external (SFF-8644) 12G miniSAS HD connectors have been upgraded to support 24G signaling, and new SFF connector numbers have been assigned (SFF-8673 for internal and SFF-8674 for external). Mechanical dimensions remain the same, facilitating backward compatibility with earlier generation devices and systems. Additionally, two new SFF connector numbers are available for SAS-4 internal cabling; miniLink (SFF-8611) and SlimSAS (SFF-8612). External cable assemblies supporting 24G are available in passive, active copper and optical depending on the length of the reach.

Board materials used in SAS-3 deployments can be used in SAS-4 designs depending on the channel that is being implemented. Existing 25G+ technologies have significantly driven down costs for material that is used in very high data rate applications. As a result, the incremental cost difference between dielectrics used in 24G and 12G is expected to be minimal in the time frame that SAS-4 reaches production.

24G SAS analyzer platforms have been released by test equipment vendors and are available for testing and validation of 24G SAS implementations.

Connector and Cable Compatibility



24G SAS Ecosystem

Emerging technologies like the Internet of Things (IoT), machine learning, and artificial intelligence (AI) have accelerated the need for increased performance coupled with a requirement for increased storage capacity with faster access. CPU platforms from multiple suppliers are already transitioning to adopt PCIe Gen4 to support connectivity to storage infrastructure. An equivalent increase in the per PHY performance of the storage infrastructure is required to take advantage of the doubling of the PCIe bandwidth to 16 Gbps per lane. Without such an increase, twice the number of 12G SAS-3 lanes are required to saturate the new PCIe Gen4 CPU interfaces.

Hard drive technology continues to provide the most cost-effective access to storage and is forecasted to provide 85% of all capacity through 2021 (IDC). Revolutionary breakthroughs like microwave-assisted magnetic recording (MAMR) and heat-assisted magnetic recording (HAMR) are helping HDDs maintain a dramatic cost advantage over competing storage technologies. Innovations that increase areal density by placing tracks and bits closer together, and technologies like multi-actuator that provide finer positioning and control to harness these higher densities, have enabled the development of very high capacity HDDs without sacrificing reliability. HDDs are currently and will continue to be the lowest cost dollar/bit media compared to flash. For this reason, storage system designers will continue to invest in hybrid storage systems wherein media that spans various performance tiers will coexist. Multi-actuator drive architectures will multiply the I/O performance such that end users can make effective use of the growth in capacity while enjoying the highest level of data access. This innovation captures the performance of parallelism within a single drive. For example, in a dual actuator drive, the host can send two separate data requests simultaneously and the drive will service them in parallel. The new protocol level features introduced in SAS-4 that improve predictability and enhance fairness in connection management can help harness the significant performance gains that these rapidly evolving drive technologies are able to provide.

Capacity gains through technologies like shingled magnetic recording (SMR) have been advanced further by providing the ability to balance capacity and performance by supporting "realms" or regions in a single drive that can switch between SMR and CMR (conventional magnetic recording) as workloads require. This is referred to as "dynamic hybrid SMR" and is expected to deliver a new class of storage efficiencies and flexibility. Innovative ways of solving evolving storage needs through this new dynamic recording method is helping redefine the future of HDDs in hyperscale data centers. It provides the option to tune how data is stored on the media by changing the data density to serve the changing application requirements. The same drive can be used for warm and cold storage, which vastly simplifies the supply chain and deployment logistics. The SAS-4 buffered PHY feature optimizes system throughput and improves efficiency in storage configurations with media that spans various performance and capacity tiers.

Conclusion

24G SAS continues the evolution of SAS by increasing efficiency, doubling performance, and supporting the various advances in drive technology. SAS remains the protocol of choice, in particular for capacity storage, within the data center as a result of its enterprise-class features and price/performance point. It provides the lowest total cost of ownership through power and bandwidth efficiency and supports a full range of storage media that addresses end-user applications requiring different performance and bulk storage tiers.

24G SAS introduces new features that optimize the transport performance and enable a highly robust and scalable ecosystem for data transfer between initiators and end devices through a service delivery network of expanders and SAS switches. These new features enhance fairness by improving how connection requests are handled and reduce protocol latency by prioritizing SMP frame transfers, allowing for more devices to be serviced in a shorter time. Transfer sizes can now be matched to buffer sizes at the receive node and retry events, that unnecessarily extend connection times, can be minimized. This improves link utilization and increases system throughput by buffering data to and from lower rate end devices in expander attached topologies. Additionally, Microsemi has developed revolutionary link aggregation technology Dynamic Channel Multiplexing (DCM) that can provide near ideal link utilization with low latency and bandwidth fairness for legacy JBODs and drives when interoperating within the SAS-4 ecosystem. For more information about Microsemi's innovative link aggregation technology, "Dynamic Channel Multiplexing", please contact Microsemi support at sales.support@microsemi.com.

References

American National Standard, Information Technology, Working Draft Serial Attached SCSI-4 (SAS-4)

American National Standard, Information Technology, Working Draft, SAS Protocol Layer-4 (SPL-4)

Microsemi Headquarters
One Enterprise, Aliso Viejo,
CA 92656 USA
Within the USA: +1 (800) 713-4113
Outside the USA: +1 (949) 380-6100
Sales: +1 (949) 380-6136
Fax: +1 (949) 215-4996
Email: sales.support@microsemi.com
www.microsemi.com

Microsemi, a wholly owned subsidiary of Microchip Technology Inc. (Nasdaq: MCHP), offers a comprehensive portfolio of semiconductor and system solutions for aerospace & defense, communications, data center and industrial markets. Products include high-performance and radiation-hardened analog mixed-signal integrated circuits, FPGAs, SoCs and ASICs; power management products; timing and synchronization devices and precise time solutions, setting the world's standard for time; voice processing devices; RF solutions; discrete components; enterprise storage and communication solutions; security technologies and scalable anti-tamper products; Ethernet solutions; Power-over-Ethernet ICs and midspans; as well as custom design capabilities and services. Microsemi is headquartered in Aliso Viejo, California, and has approximately 4,800 employees globally. Learn more at www microsemi.com.

ESC-2180677