

White Paper
**Using SMR Drives with Smart Storage Stack-Based HBA
and RAID Solutions**

October 2017



Contents

1 What Are SMR Drives and Why Are They Used?	1
2 SMR Drives in HBA or RAID Configurations	3
3 How Does This Work in the Real World?	4
4 Conclusion	6

1 What Are SMR Drives and Why Are They Used?

Shingled magnetic recording (SMR) is a technology used to increase the capacity of hard disk drives (HDDs) by writing overlapping tracks. With standard hard disk drives, also called conventional magnetic recording (CMR) HDDs to differentiate them from SMR HDDs, the minimal width of write tracks for hard disk drives is determined by the size of the write head, which is constructed to write data reliably. Conventional HDDs do not overlap tracks.

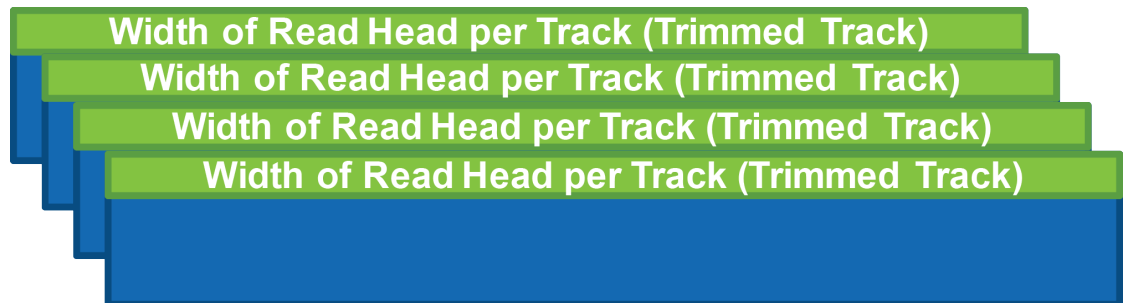
Figure 1 • Conventional Writes (CMR Drives and SMR Conventional Zone)



SMR drives take advantage of the fact that the track width needed to read the data, and with it the width of the HDD's read head, is much narrower, and therefore does not need the full width of the written track to read data reliably.

The following illustration shows how SMR drives overlap tracks during writes to increase data density, resulting in increased usable capacity. This works well as long as data is always appended at the end. Once data is inserted in the middle, a write would overwrite multiple tracks at the same time.

Figure 2 • Overlapped SMR Writes



To allow some randomness during writes or to store data from multiple sequential streams, such as when recording multiple video streams or performing simultaneous backups, each SMR drive has multiple bands (sometimes also called “zones”) with a write pointer that allows data to be appended to each band.

The pitfall of SMR drives is overwriting existing data, as shown in the following illustration. This is because the write head's width touches multiple overlapping tracks when writing any data. To avoid data loss, the entire band's data must first be read into a buffer, then the new data must be placed at the appropriate position in the buffer, and finally the whole band must be rewritten (read-modify-write cycle). This means that a random write operation of a potentially small block can balloon into read and write operations for a whole band.

Figure 3 • Overwriting Existing Data on an SMR Drive



On top of that, deleting data from a band leaves holes; therefore, occasional garbage collection must take place to clean up bands to allow new data to be appended at the end of the band. This process is quite similar to garbage collection in a flash-based SSD, although it occurs for a slightly different reason.

Overwriting data and garbage collection are very time-consuming processes on hard disk drives due to their mechanical nature, which involves head movements taking milliseconds. This makes SMR HDDs even less suitable for random write workloads than CMR drives, which already lag behind memory based storage devices such as flash-based SSDs.

SMR drives usually have a conventional zone where they do not overlap tracks, and this is used either as a buffer for random write data to improve response times or as metadata storage keeping track of things like deleted data or write pointer position in the various bands.

Why are SMR drives deployed despite the additional complexity that they introduce? The answer is that they increase the capacity of HDDs without increasing cost. Leveraging the same mechanics used for a CMR drive, usable capacity is between 15% to 20% higher using SMR technology.

With the introduction of new storage device technologies, such as flash-based SSDs or other non-volatile memory (NVRAM)-based storage, hard disk drives are less frequently used in transactional workloads. However, with the data growth predicted, there is still an increasing demand for cost-efficient, high-capacity storage devices optimized for dollars per GB rather than I/O performance.

Data is categorized as either hot or cold. Hot data is frequently accessed and might also change more often and is held on flash based storage or in memory. Warm data is still accessed from time to time but less likely to change, while cold data is archived and rarely changes. Cold data is usually written once, and then is occasionally read as time goes on. One example of cold data is photo albums archived in cloud storage. Newer photos are eventually edited before being printed, but older albums are left untouched, occasionally being looked through to revisit fond memories.

External USB HDDs, as well as cloud storage for archives and backups, are ideal use cases for SMR drives. At the time this article was written, millions of SMR drives have already been deployed as USB backup drives, as they offer more capacity at an affordable cost. In the future, SMR drives will likely be used for archival and cold storage. Different SMR drive variants have been developed to accommodate the different requirements for cloud drives and external USB backup hard disk drives, and these variants and the differences between them will be described in the following section.

2 SMR Drives in HBA or RAID Configurations

Device-managed (DM) SMR drives can be deployed with any standard HBA that is SMR-aware. These drives are commonly deployed as an external USB backup device for desktop or notebook systems.

The ideal HBA for a host-aware (HA) SMR drive is one that supports the ZBC/ZAC command sets. Supporting these command sets is mandatory for any host-managed (HM) SMR drive. HA drives are often used by applications that take direct control of the drive. Currently, OS and file system vendors are developing HM drive solutions. Requirements for HM drive support are different from other drives due to the fact that HM drives do not support any randomness whatsoever. For example, most modern operating systems use multiple threads running on different server CPU cores and multiple command queues to leverage the system's hardware resources, optimizing I/O operations for high-end memory-based storage devices. In such architecture, not all I/Os are submitted in sequence, a process not supported by HM SMR drives.

The previous example illustrates why SMR solutions are most likely proprietary and specific to certain applications in hyperscale data centers. These data centers are among the first to adopt HM SMR technology as they deploy massive amounts of storage and require a contained environment where they can control the software and all running applications.

A redundant array of independent disks (RAID) aggregates multiple storage devices into one logical unit that is treated like a single storage device by the host/operating system. The goal is to increase data availability by adding redundancy and scaling performance by leveraging multiple storage devices. To this end, the data can be mirrored (RAID1), striped (RAID0), or redundancy information can be added for increased availability (RAID5, RAID6).

Dedicated storage adapters, called RAID adapters, manage RAID arrays with nonvolatile memory, and can be used to further accelerate RAID by caching and offloading redundancy calculation from the host.

RAID configurations use a fixed pattern to determine how to distribute data among storage devices in a RAID array, and are therefore not suitable for HM SMR drives that require all write data to be sequential (no support for write-in-place). RAID can be used with DM SMR drives and HA drives that support drive-managed mode.

Despite RAID being compatible with both SMR and CMR drives, mixing the two drive types within the same RAID array is not a good idea, as they have very different performance characteristics. As the saying goes, "The chain is only as strong as its weakest link." Likewise, the performance of a RAID array that mixes SMR and CMR drives would be similar to an SMR-only RAID array. Due to their additional complexity, SMR drives have limits in the number of IOPS they can deliver and suffer from inconsistent latency when responding to I/O requests in random write workloads. Incorporating SMR drives into RAID arrays does not change this fact.

In summary, SMR drives in RAID arrays have the same limitations as individual SMR drives. However, the RAID configuration can help aggregate the performance of multiple SMR drives as it would for CMR drives. As a result, an overall higher level of performance can be achieved in workloads, while the RAID provides higher data availability.

In order to support SMR drives in RAID arrays, the RAID's software and nonvolatile cache can help buffer data. However, it cannot change the SMR drive's characteristics. That means the RAID adapter can only do so much in terms of buffering commands and providing more relaxed timeouts. SMR drives in such RAID arrays should only be used with applications and in environments that require very limited random write performance. This rule also applies to individual SMR drives connected to an HBA. Otherwise, the disk and adapter's buffers will quickly run over, starving the application due to I/O requests not getting processed in a timely manner.

If used the wrong way, I/O processing times of DM SMR HDD-based RAID arrays can extend far beyond OS and application timeouts, causing hangs and even data loss, much like an individual SMR device. The use of DM/HA SMR drives in RAID is therefore only recommended in such environments where the level of random write I/Os is low and controllable.










3 How Does This Work in the Real World?

Both RAID adapters and HBAs need optimization in order to work with suitable SMR drives. SMR support for the recommended operating modes described earlier is available in the Microsemi Adaptec® HBA 1100 Series, SmartHBA 2100 Series, and SmartRAID 3100 Series.

In order to help users to identify DM and HA SMR drives that look and behave like CMR drives, the Microsemi Smart Storage adapter checks the necessary identifiers of the device that report drive type. If it is identified as an SMR drive, the adapter will display it with a separate icon that alerts the user that it is not a standard CMR HDD. The following illustration shows the icons and describes their meanings.

Figure 4 • Enterprise View Icons

What do the Enterprise View Icons Mean?

Icon	Description
	System with controller and directly attached disk drives or enclosures
	Controller
	Enclosure
	Logical device
	maxCache Device (healthy)
	Hard disk drive
	Solid State Drive (SSD)
	SMR drive
	Connector or other physical device

These notifications are designed to highlight the device's type and to prevent the user from mixing CMR and SMR drives within the same RAID array or logical volume. The adapter will display a warning message if a user tries to create such RAID array of mixed devices and prevents from setting up the RAID or logical drive (LD). HBAs also use icons to indicate whether an SMR drive is HA or HM.

Equipped with this information, the user can create SMR-only logical volumes and RAID arrays to aggregate capacity and increase performance, capacity, and data availability (within the established limits of SMR technology). The same RAID levels that are supported with CMR drives are open to both SMR devices with similar device counts per logical volumes and RAID arrays.

Even OS boot is supported by modern DM SMR drives, as it mostly involves reads and limited random writes. However, there are still scenarios in which the SMR drives can impact performance, and therefore this use case is not recommended. If the SMR drive is exposed to heavy random write access, dirty data will build up on the drive's random write zone, and the device will eventually become unresponsive. This scenario can only be avoided if the user ensures that the drive is only used in an SMR-friendly environment.

The use of HM SMR drives and HA SMR drives in HA mode is limited to raw devices. These are supported with all Microsemi Smart Storage adapter variants, including the HBA 1100, SmartHBA 2100, and SmartRAID 3100 products. Microsemi has already successfully validated various HA SMR HDDs with its adapters, and is working with operating system and software vendors to implement full support for HM devices that are not currently widely supported by existing devices due to being unable to accept any random writes.

For more information about SMR drive compatibility and support, please either contact a Microsemi sales representative or visit Microsemi's website.

4 Conclusion

SMR increases HDD capacity by writing on overlapping tracks, something that conventional HDDs cannot do. RAID configurations use a fixed pattern to distribute data among storage devices in a RAID array, making them unsuitable for HM SMR drives that require all write data to be sequential. However, RAID can be used with DM SMR drives and HA SMR drives that support drive-managed mode. SMR drives in RAID arrays have the same limits as SMR drives used as individual drives. Despite this, RAID can help aggregate the performance of multiple SMR drives in the same way that it does with CMR drives, achieving higher overall performance levels and data availability in workloads.

Starting with the availability of the Microsemi Adaptec HBA 1100, SmartHBA 2100, and SmartRAID 3100 Series, Microsemi's storage adapter solutions support the needed command extensions for HM SMR or HA SMR drives: Zone Block Commands set or Zoned-Device ATA Command set (ZBC/ZAC). HM SMR and HA SMR drives operating in HA mode should only be used as raw devices and require support in the operating system, file system, or application in order to ensure successful operation.

For more information about Microsemi products, visit www.microsemi.com.



Microsemi Corporate Headquarters

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

© Microsemi Corporation. All rights reserved. Microsemi and the Microsemi logo are trademarks of Microsemi Corporation. All other trademarks and service marks are the property of their respective owners.

Microsemi makes no warranty, representation, or guarantee regarding the information contained herein or the suitability of its products and services for any particular purpose, nor does Microsemi assume any liability whatsoever arising out of the application or use of any product or circuit. The products sold hereunder and any other products sold by Microsemi have been subject to limited testing and should not be used in conjunction with mission-critical equipment or applications. Any performance specifications are believed to be reliable but are not verified, and Buyer must conduct and complete all performance and other testing of the products, alone and together with, or installed in, any end-products. Buyer shall not rely on any data and performance specifications or parameters provided by Microsemi. It is the Buyer's responsibility to independently determine suitability of any products and to test and verify the same. The information provided by Microsemi hereunder is provided "as is, where is" and with all faults, and the entire risk associated with such information is entirely with the Buyer. Microsemi does not grant, explicitly or implicitly, to any party any patent rights, licenses, or any other IP rights, whether with regard to such information itself or anything described by such information. Information provided in this document is proprietary to Microsemi, and Microsemi reserves the right to make any changes to the information in this document or to any products and services at any time without notice.

Microsemi Corporation (Nasdaq: MSCC) 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-2171613