

# HEPiX Techwatch WG : Disk Storage

First Last (Affiliation), First Last (Affiliation), First Last (Affiliation), First Last (Affiliation), First Last (Affiliation), First Last (Affiliation), First Last (Affiliation), First Last (Affiliation), First Last (Affiliation), First Last (Affiliation), First Last (Affiliation), First Last (Affiliation)

“[Tape is Dead, Disk is Tape, Flash is Disk, RAM Locality is King](#)”, 12/2006, Jim Gray, Microsoft.

---

## Staging Area

## References

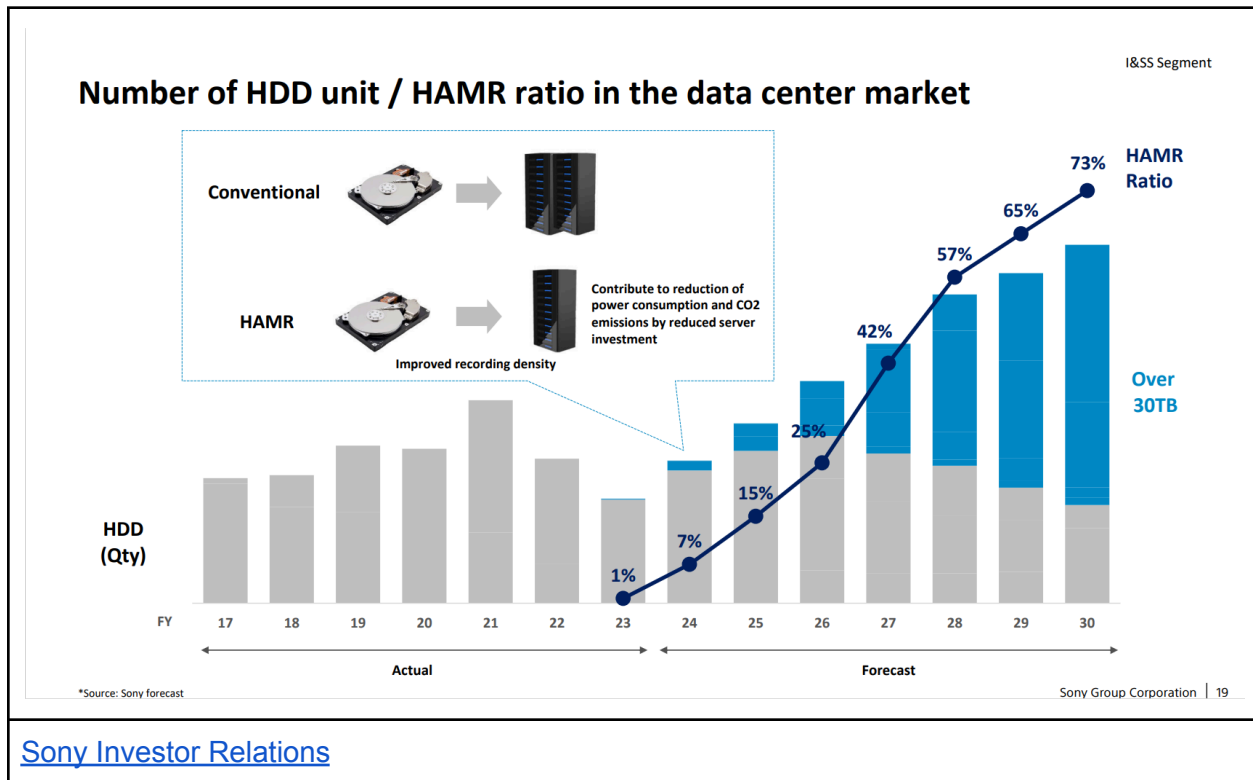
Following are links to articles and information that should be integrated into the document:

- Recent News (Add links to new articles for later integration in this document)
  - [Seagate 2024Q4](#) statements includes HAMR updates
  - [Toshiba HAMR drives](#)
  - [HDD Forecasts](#)
  - [Storage Software Stacks](#)
  - [Storage Roadmap](#)
  - [WD 2024Q4 presentation](#)
  - [WD Financials](#)
  - [Seagate AI Comments](#)
  - [Seagate Financials](#)
- “Full Picture” Sources
  - (Updated 9/24) Library of Congress [National Digital Information Infrastructure and Preservation Program](#) - The annual [Designing Storage Architectures Meetings](#) (DSA meeting) draws speakers from various industry and academic organizations.
  - (Updated 9/24) The presentations by Georg Lauhoff (IBM) at the DSA meetings are particularly relevant as they cover storage technology and market in detail.
  - (Updated 9/24) The [Information Storage Industry Consortium](#) (INSIC) is a consortium of academic, industry and government organizations in the field of information storage. They publish the [INSIC Tape Technology Roadmap](#) on a periodic basis.

- (Updated 9/24) [SpectraLogic](#) Data Storage Outlook Report - Annual storage market survey from SpectraLogic, a manufacturer of tape automation equipment. The most recent report is from [2023](#) and can be downloaded from SpectraLogic (requires registration). Older reports also appear to be available (The SpectraLogic website search tool can be used to locate reports from previous years.)
- News Sites and Market Analysts
  - (Updated 1/23) [Blocks and Files](#) is a website that provides news and analysis of the storage market
  - (Updated 1/23) [TrendFocus](#) - Market analysis company focusing on storage.
- Storage Trade Organization
  - (Updated 1/23) SNIA - The [Storage Networking Industry Association](#) is, to quote their website: “a not-for-profit global organization that leads the storage industry in developing and promoting vendor-neutral architectures, standards, and educational services that facilitate the efficient management, movement, and security of information.” SNIA sponsors multiple conferences and technical working groups and publishes numerous resources covering storage technology. Relevant SNIA sponsored conferences and meetings include the following:
    - SNIA [Storage Developer Conference](#)
    - SNIA [Persistent Memory and Computational Storage Summit](#)
    - SNIA [Preview](#)
- Standards and associated trade organizations
  - (Updated 1/23) [SATA-IO.org](#) - Custodians of the SATA standard
  - (Updated 1/23) [SCSI Trade Association](#) (SCSITA) - SAS/SATA trade organization
  - (Updated 1/23) [INCITS T10 Technical Committee](#) - Custodians of the SCSI and SAS standard
  - (Updated 1/23) [Fibre Channel Industry Association](#) - Custodians of the Fibre Channel storage area network standard
  - (Updated 1/23) [Infiniband Trade Organization](#) - Custodians of the Infiniband standard and the RoCE (RDMA over Converged Ethernet) standard
  - PCI-e standard is developed by the [PCI Special Interest Group \(PCI-SIG\)](#).
  - [NVMexpress.org](#) is the organization developing the NVMe including NVMeoF the standard.
    - Additional information on NVMe can be found at the [NVMe Developers Day conference](#).
  - [IEEE 802.3 Working Group](#) develops the Ethernet standard.
  - (Updated 1/23) [Zone Storage IO](#) - Website dedicated to the Zoned Storage (HDD) and Zoned Namespace (NVMe SSD) APIs.
- Vendor Pages
  - (Updated 1/23) [Seagate Technology](#) - HDD manufacturer website. Occasionally posts technical information in the Investor Relations [page](#). [2019](#) and [2021](#) Analyst Day presentations are particularly noteworthy. The Seagate corporate blog, <https://blog.seagate.com/>, has some articles on the status of Seagate HAMR products.

- (Updated 1/23) [Western Digital](#) - HDD manufacturer website. Their Investor Day presentations, available from their Investor Relations [page](#), provides technical information on the state of HDD and Flash technologies (and market) .Contents are volatile, although occasionally presentations have been available with market and technology information.
- (Updated 1/23) [Toshiba](#) - HDD manufacturer website. Occasionally posts technical information in the Investor Relations [page](#). Although from 2018, the [FY2018 Technology Strategy Briefing](#) documents Toshiba’s plan to introduce MAMR drives in the near future.
- (Updated 1/23) [The Magnetic Recording Conference](#) (TMRC) - An annual conference on magnetic recording technology sponsored by the [IEEE Magnetics Society](#). Websites for recent conferences are as follows:
  - TMRC 2024 - <https://sites.google.com/andrew.cmu.edu/tmrc2024/home/>
  - TMRC 2023 - <https://sites.google.com/umn.edu/tmrc2023/>
  - TMRC 2022 - <https://tmrc22.sites.stanford.edu/>
  - TMRC 2021 - <https://www.dssc.ece.cmu.edu/TMRC2021/index.html>
  - TMRC 2020 - <http://cml.me.berkeley.edu/TMRC2020/>
  - TMRC 2019 - <https://sites.google.com/umn.edu/mint-tmrc2019>
  - TMRC 2018 - <http://tmrc2018.ucsd.edu/>
  - TMRC 2017 - <https://www.nims.go.jp/mmu/tmrc2017.html>
  - TMRC 2016 - <https://tmrc16.stanford.edu/>.

1.



## Suggested Outline

1. Executive Summary
  2. Burning Questions
  3. Introduction
  4. Technology
    - a. PMR (Perpendicular)
      - i. ePMR
      - ii. OptiNAND
    - b. HAMR (Heat Assisted)
    - c. SMR (Shingled)
    - d. IMR (Interleaved)
    - e. Increased Platter Count
    - f. Multi Actuator
  5. Market
    - a. General competitive environment
    - b. Vendors
      - i. Western Digital
      - ii. Seagate
      - iii. Toshiba
  6. Supporting Technology
    - a. Interconnect
      - i. SATA
      - ii. SAS
      - iii. Fibre Channel
      - iv. NVMe/NVMeoF
    - b. System
-

## Executive Summary

Disk technology, specifically magnetic hard disks,  
Summary of key findings, including possible impact of cost and technology evolution on HEP/NP in the future. (Single Paragraph)

Big news is the announcement by Seagate of volume shipment of 30TB HAMR drives to Cloud Service Providers in Q1 of 2024. 2023 bad year financially, enterprise capacity HDD increasing its share of the HDD market. Three vendors are still in the market, Seagate, Western Digital, and Toshiba, but upheaval at WD as it plans to spin off its Flash business. PMR HDD drive capacity continues to increase through more platters, tweaks to PMR r/w heads, and other incremental improvements. Divergent write characteristics of SMR still inhibits adoption of SMR in the general market. However, SMR represents ~50% of exabytes shipped by Western Digital.

## Burning Questions with Answers

1. What is the long term viability, both technical and financial, of HDD
2. Will flash subsume the HDD market (Topic of a separate Techwatch WG report on data storage technologies.
3. Impact of Cloud Service Providers (aka Hyperscalers) on the HDD market

<Insert Answers Here>

## Introduction

Magnetic hard disk drives (HDDs) have been the “go to” technology for online persistent storage of data for High energy (HEP) and nuclear physics (NP) experiments. Historically, HDD based storage systems have provided fast and convenient access to large volumes of data at acceptable costs to enable researchers to analyze their data with minimal friction. In recent years advancements in storage technologies, particularly in the area of cost, have not kept pace with the escalating volumes of data being collected by these experiments. The HEP and NP communities have been working to reduce their storage requirements, but the success of these mitigation efforts will be partially determined by the rate of storage advancement over the next decade.

In the past few years, HDDs have been encountered the following obstacles:

1. Slow growth in HDD disk capacity
  - a. The maximum areal bit density (~1 Tbit/sq in) for Perpendicular Magnetic Recording (PMR) technology, the current recording technology, has been reached in current hard disk drives.
  - b. The availability of higher bit density recording technology (e.g. HAMR) has been delayed.
2. Limited improvements in HDD performance, resulting in lower BW/terabyte and IOPS/terabyte ratios over time as disk capacities have increased
  - a. Plateau in HDD platter RPM at 7200 RPM for capacity disk drives
  - b. Incremental improvements in I/O bandwidth (BW)
  - c. Virtually no change in random IOPS
3. Shrinking Market
  - a. Declining sales of desktop PCs resulting in lower desktop HDD revenues
  - b. Cannibalization of enterprise, desktop, and laptop HDD sales by flash storage
4. Established market pricing and market expectations - a disk costs less than ‘a few hundred’ Euros and no larger group will ever pay way more in the future (burned prices)

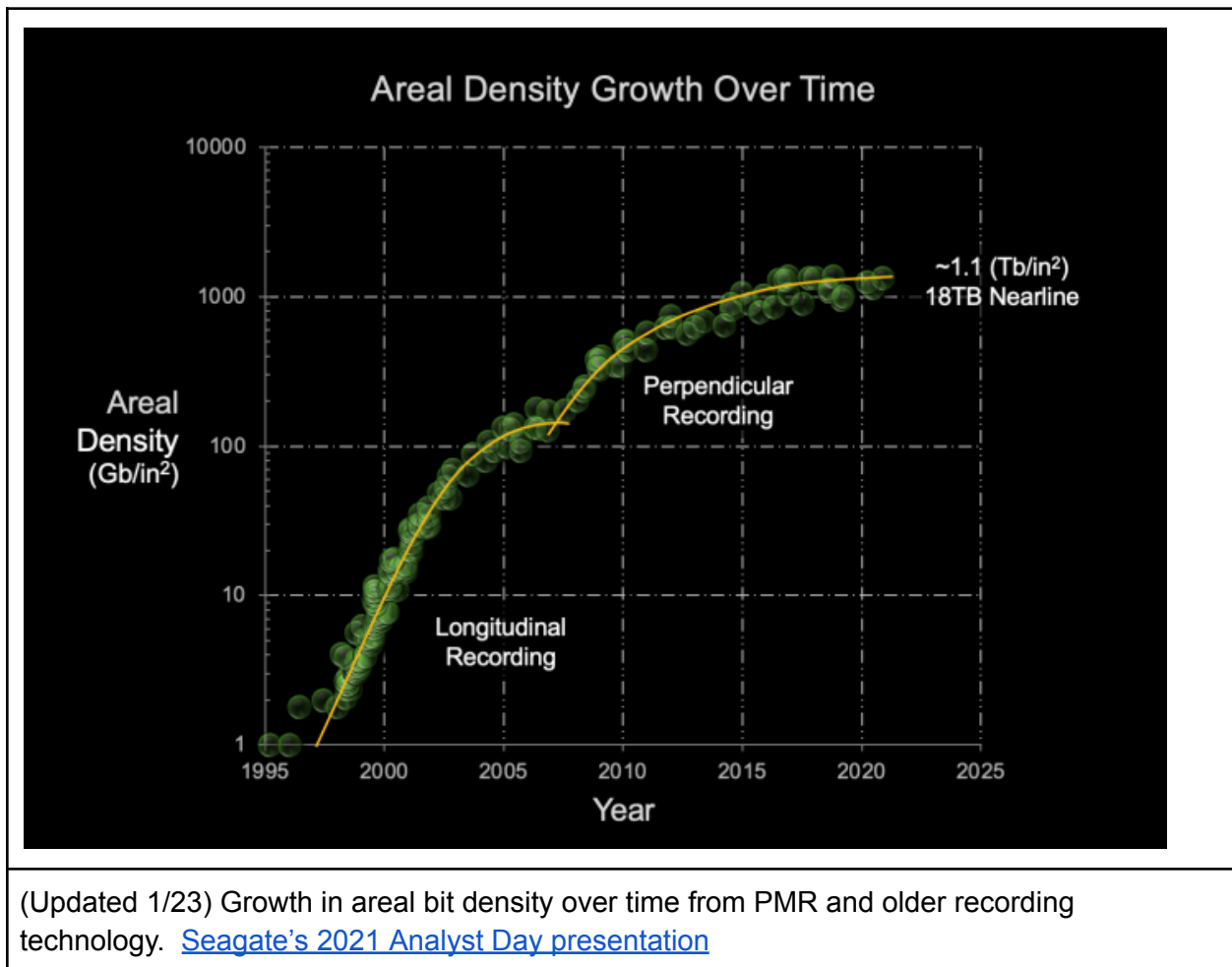
This document covers the current state and evolution of HDD technology as well as the state and financial health of the market and the major industrial “players” in disk storage.

## Technology

## Increasing HDD Capacity

### Perpendicular Magnetic Recording

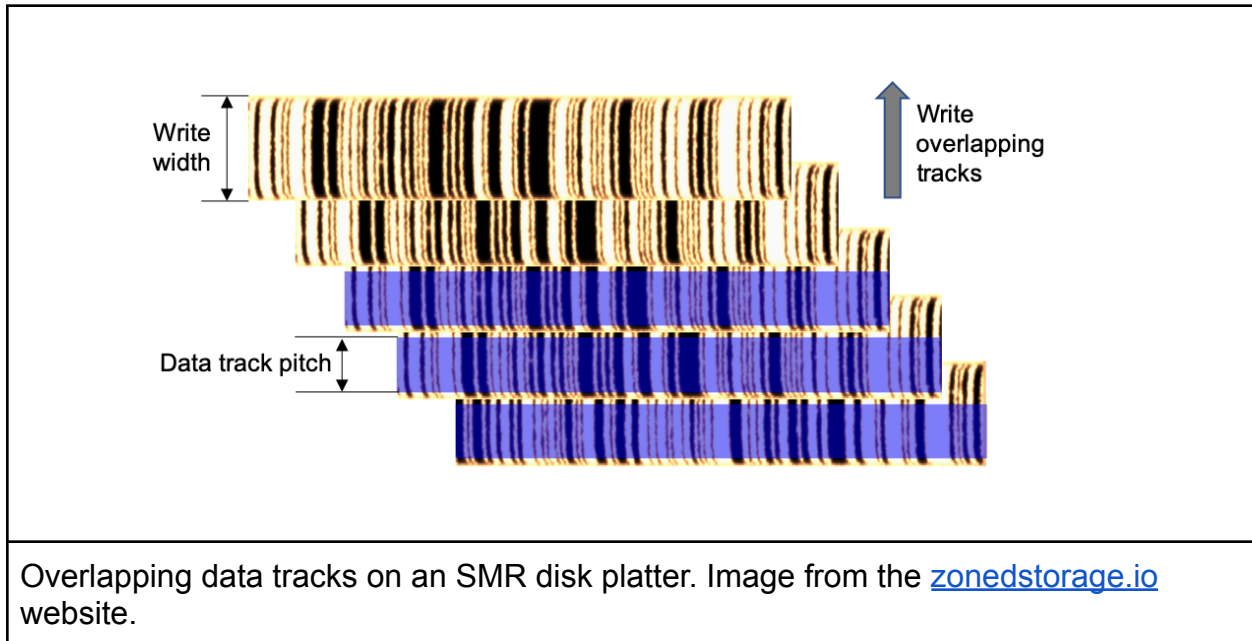
At this point in time, the major change in the outlook for HDDs is in the area of areal bit density, IOPS, and I/O bandwidth. The plateauing of areal bit density with perpendicular magnetic recording technology can be seen in the graph below from Seagate's 2021 Analyst Day presentation.



### Shingled Magnetic Recording (Zoned Storage)

[Shingled magnetic recording](#) (SMR) stretches PMR technology a bit further by overlapping adjacent data tracks resulting in a net gain of 20 - 25% capacity over CMR drives. SMR takes advantage of the fact that the width of the data track needed for reads is narrower than the

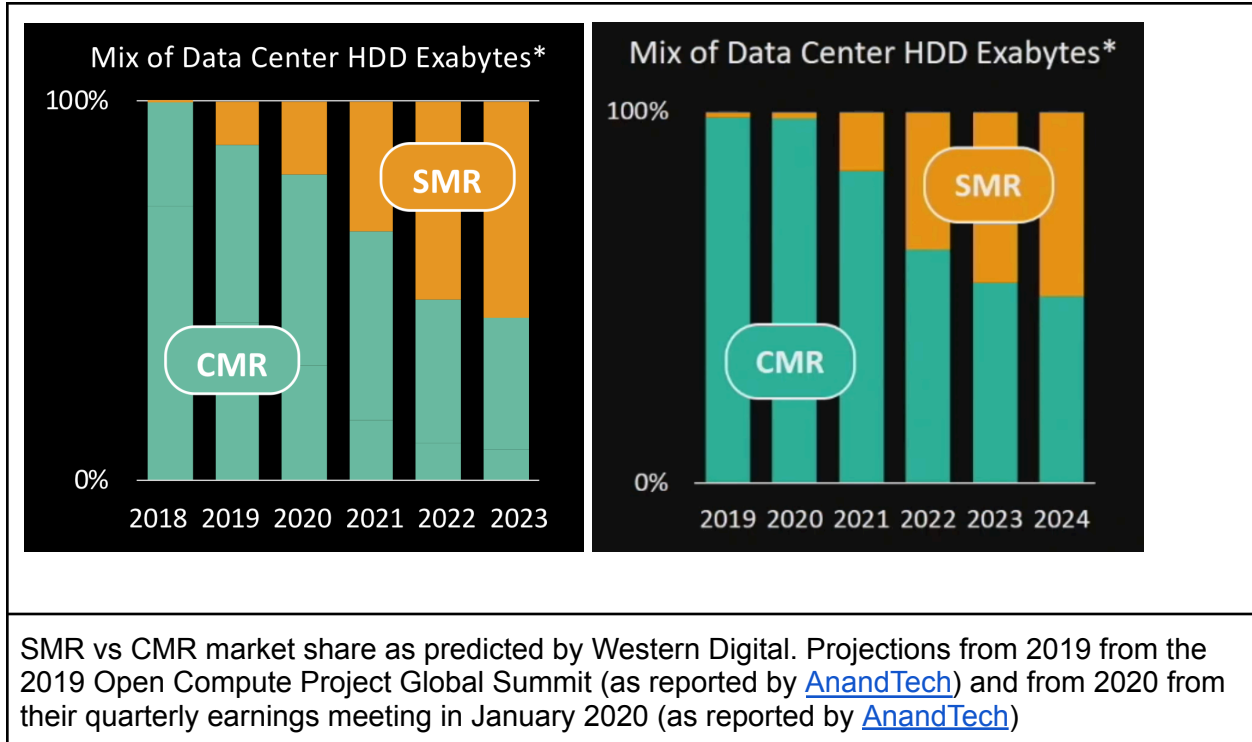
width of the data track laid down by the write head. The following photograph from the [zonedstorage.io](http://zonedstorage.io) web site shows overlapping data tracks in an SMR drive.



However, SMR is not a panacea as the native performance characteristics of the drive differs from normal PMR drives. To accommodate SMR drives, either the application (or OS) or drive electronics must compensate for the performance differences. This was made clear around 2020, when both Seagate and Western Digital surreptitiously changed the underlying recording technology in certain drive models from CMR to SMR. The change was noticed by consumers when the SMR drives failed during [rebuild](#) (resilver) of ZFS RAIDZ volumes. The zoned storage API, discussed later in this paper, was introduced by drive vendors to enable software developers to compensate for the quirks of SMR drives.

At the 2019 Open Compute Project Global Summit (as reported by [AnandTech](#)), Western Digital stated that they expect SMR based disk drives will represent the majority of hard drive storage capacity by 2023 and that host based SMR will be the model of choice. However, WD walked back their expectations during their quarterly earnings conference in January 2020. The figure on the left is from April 2019, the one on the right is from January 2020 (from [Anandtech](#)).





Western Digital currently sells only Host Managed SMR drives. On the other hand, Seagate currently sells “drive managed” SMR drives, requiring no host support, but appears to be looking at ZBD drives in the future. Note that both Toshiba and Western Digital are also expected to ship SMR based MAMR drives, in addition to “conventional” non-SMR based MAMR drives, in the near future.

## Interlaced Magnetic Recording

Interlaced Magnetic Recording (IMR) is a new technology that has been proposed as an alternative to SMR. IMR utilizes the capabilities of HAMR to write two layers of data tracks with the bottom and top tracks offset, resulting in an interleaving of bottom and top layers. This configuration has significantly lower performance impact on writes when compared to SMR.

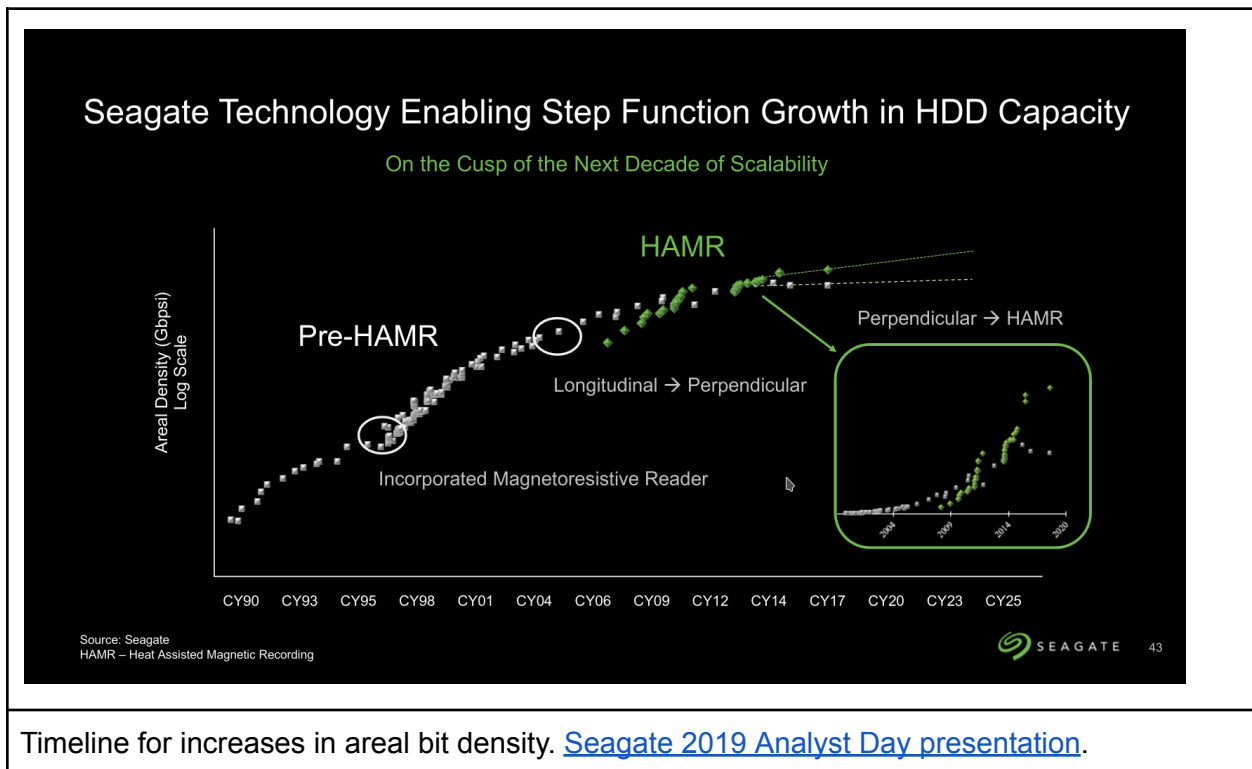
## Energy Assisted Magnetic Recording

### HAMR

For over a decade, heat assisted magnetic recording (HAMR) has been touted as the recording technology of the future. With PMR HDD technology reaching the end of the road, heat assisted magnetic recording (HAMR), in development for almost a decade, appears to be ready for production. Back in 2019, Seagate corporate [blog](#) stated customer integration testing (at NetApp) has occurred with HAMR drives, with production 20TB drives in CY2020. At this point

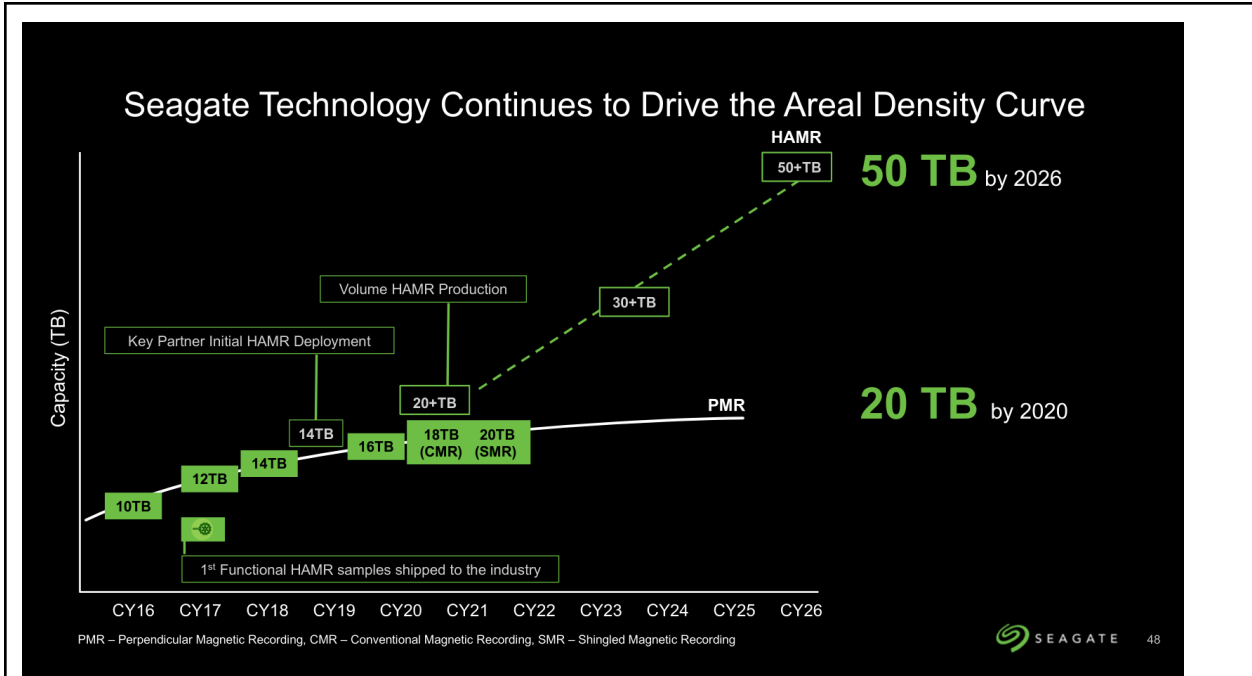
in time (Jan 2023) HAMR drives still seem to be relegated to selected customers. It should be noted that HAMR drives are a key component of the storage system for the Frontier supercomputer at Oak Ridge National Laboratory, scheduled for installation in 2021 ([Seagate Analyst Day Presentation](#) Sept 2019).

In 2019, Seagate demonstrated production media at [2 Terabits per square inch](#) (Tbps) with HAMR. Looking towards the future, Seagate has demonstrated [10Tbps](#) media with HAMR in the lab. Seagate expects 20% CAGR for areal bit density over the next decade with HAMR ([Seagate 2019 Analyst Day](#)).



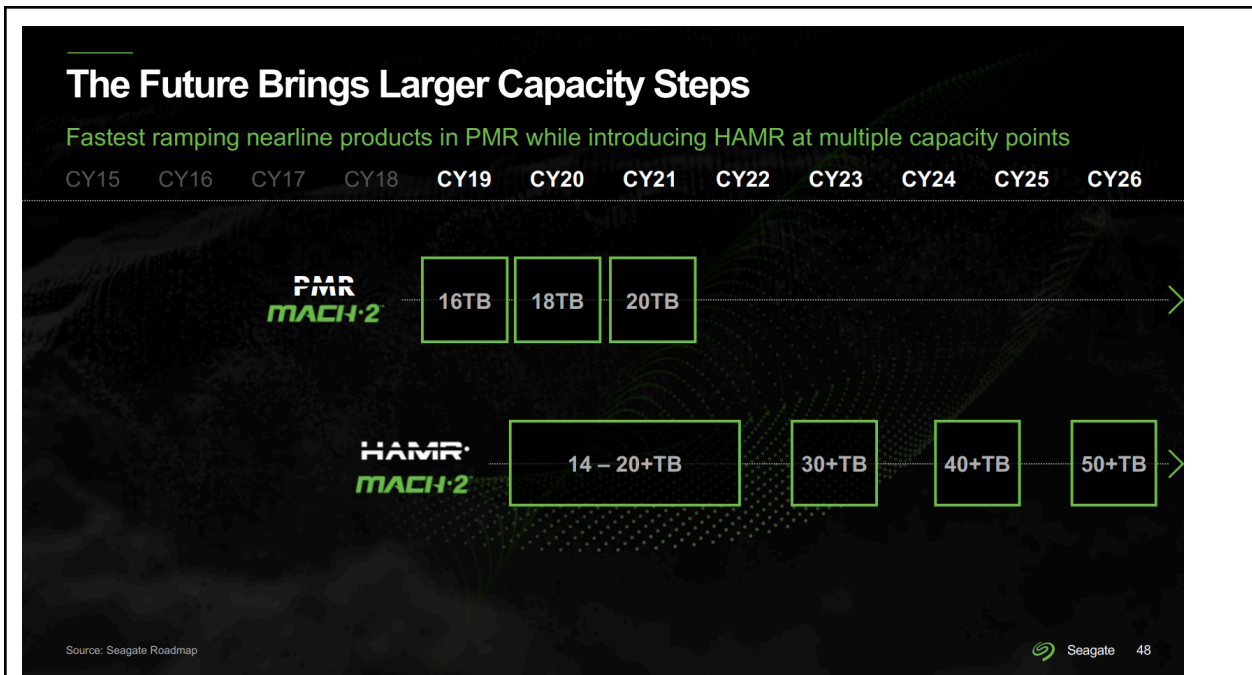
Timeline for increases in areal bit density. [Seagate 2019 Analyst Day presentation](#).

In the same presentation, Seagate predicts the following growth in HDD capacity over the next five years.



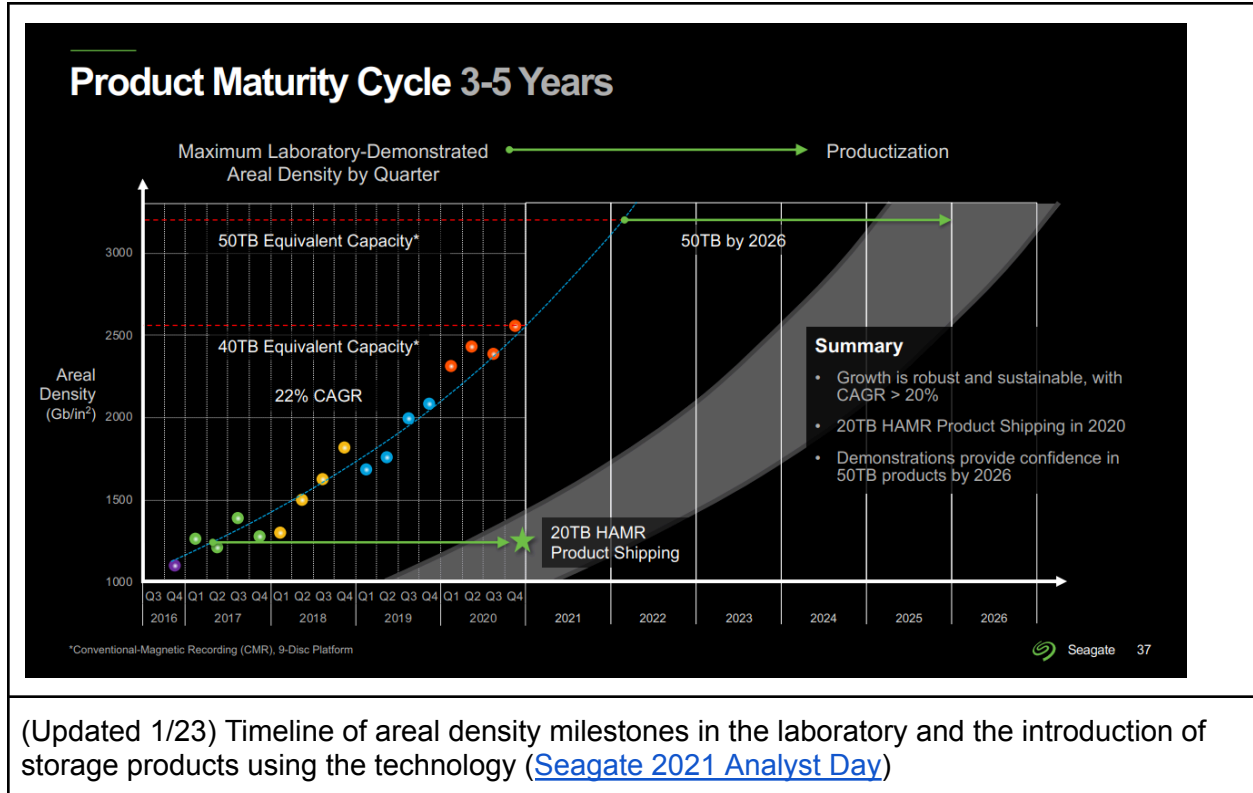
Timeline for increases in drive capacity. [Seagate 2019 Analyst Day presentation.](#)

An updated version of the drive capacity timeline from Seagate’s 2021 Analyst Day presentation is shown below:



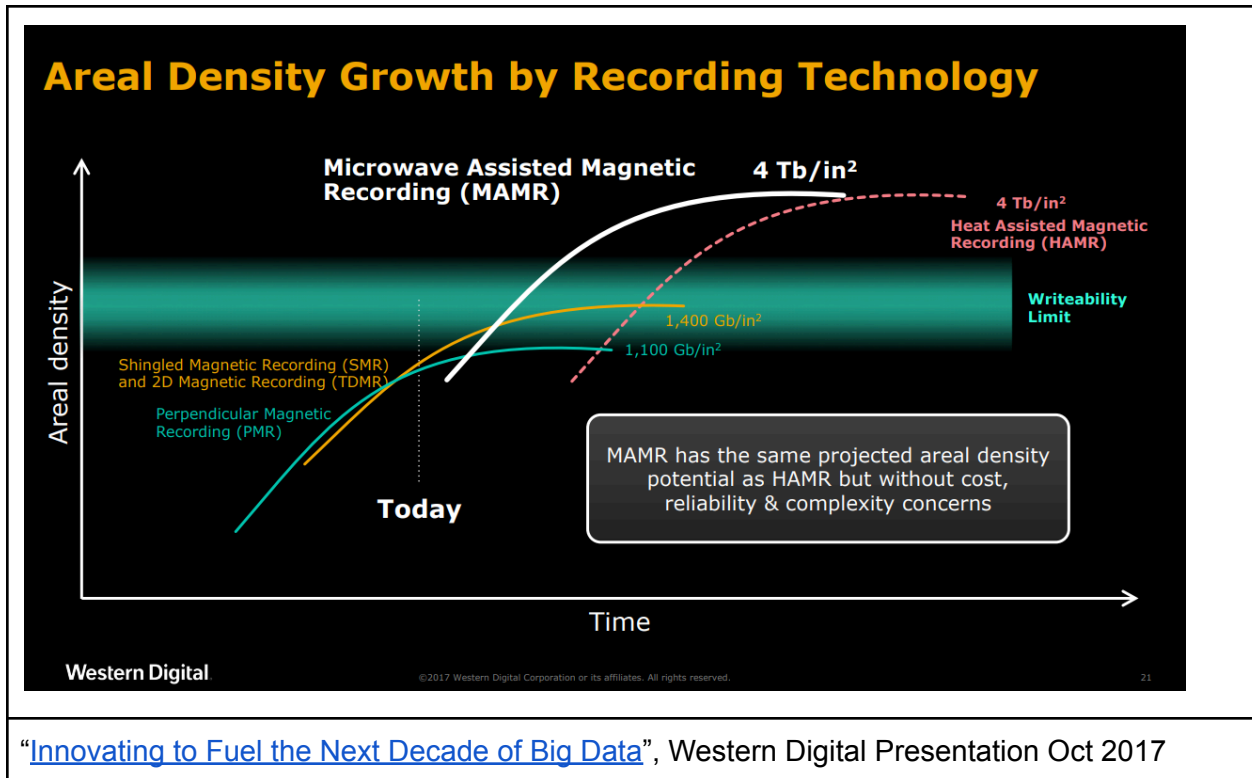
(Updated 1/23) Timeline for increases in drive capacity. [Seagate 2021 Analyst Day presentation.](#)

The following figure from [Seagate's 2021 Analyst Day](#) presentation shows the timeline of areal density milestones in the lab and a corresponding timeline showing the expected arrival of products utilizing the technology.



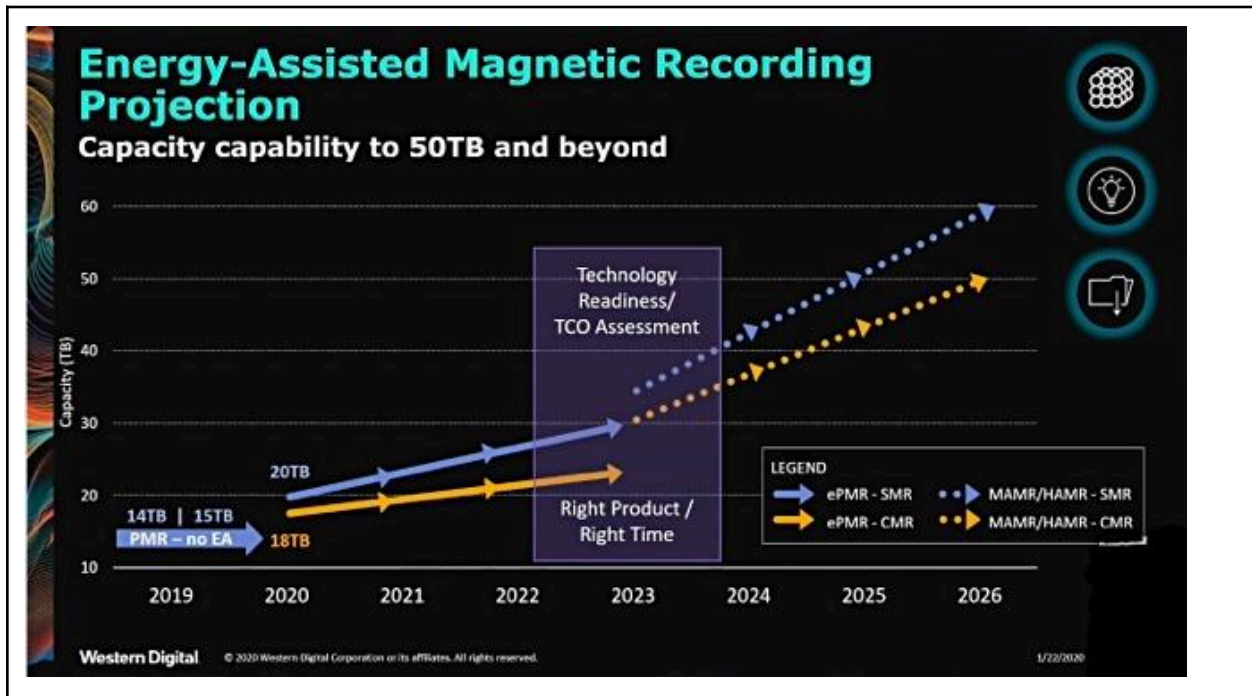
## MAMR

In contrast to Seagate, in 2019 Western Digital decided to push HAMR drives out into the future and concentrate on microwave assisted (MAMR) drives. As a stepping stone to HAMR, Western Digital saw MAMR having the potential to reach areal bit densities over [3Tbps](#), with WD projecting 40TB MAMR drives by 2025 in their [MAMR presentation](#) in 2017. In 2019, WD [announced](#) the sampling of 16TB MAMR-based disk drives.



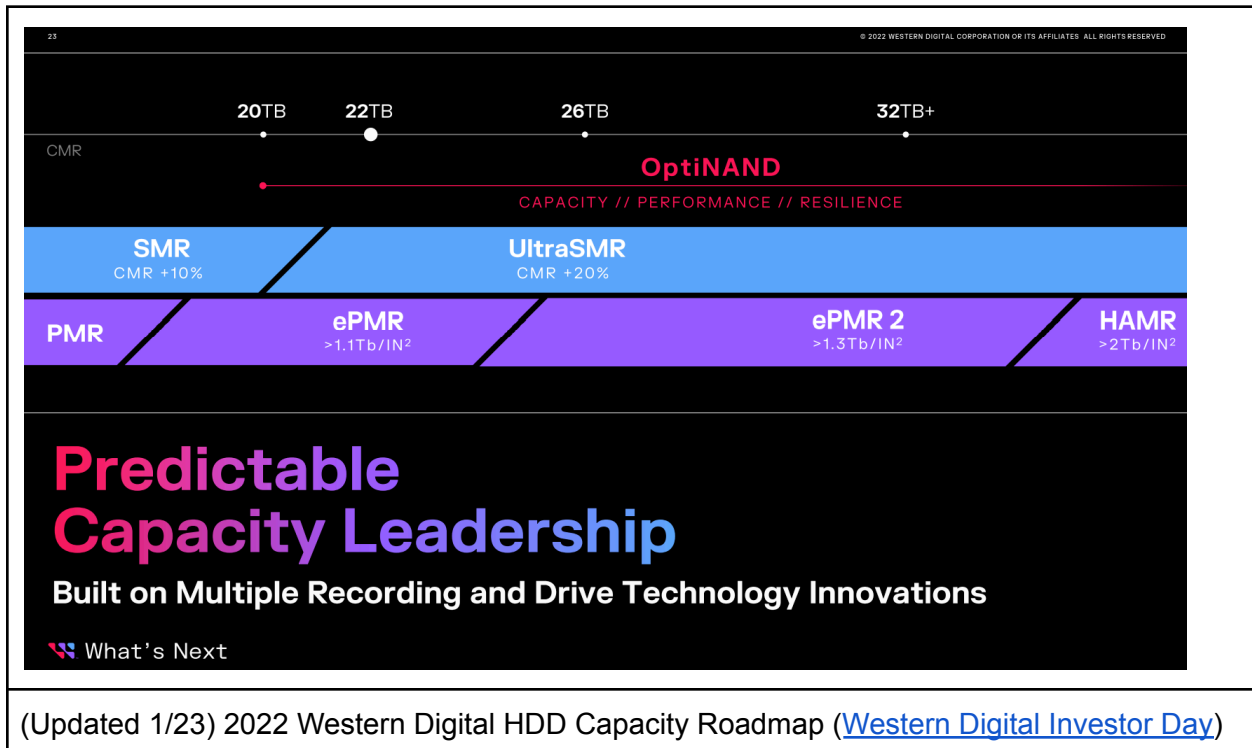
“[Innovating to Fuel the Next Decade of Big Data](#)”, Western Digital Presentation Oct 2017

In January 2020, Western Digital’s roadmap of drive capacity presented at the [Storage Field Day 19](#) (from [blocksandfiles.com](#)) had MAMR and HAMR drives pushed out into the future, with near term drives using “ePMR” technology.



Western Digital roadmap of drive capacity at the [Storage Field Day 19](#) (from [blocksandfiles.com](#))

Western Digital's high capacity drives 18TB PMR and 20TB SMR drives announced in 2020 were based on "ePMR" technology, an enhancement of PMR described in this Western Digital [technical brief](#). In May of 2022, MAMR drives completely disappeared from Western Digital's HDD roadmap and was replaced with "ePMR2" in the short term and HAMR in the long term.



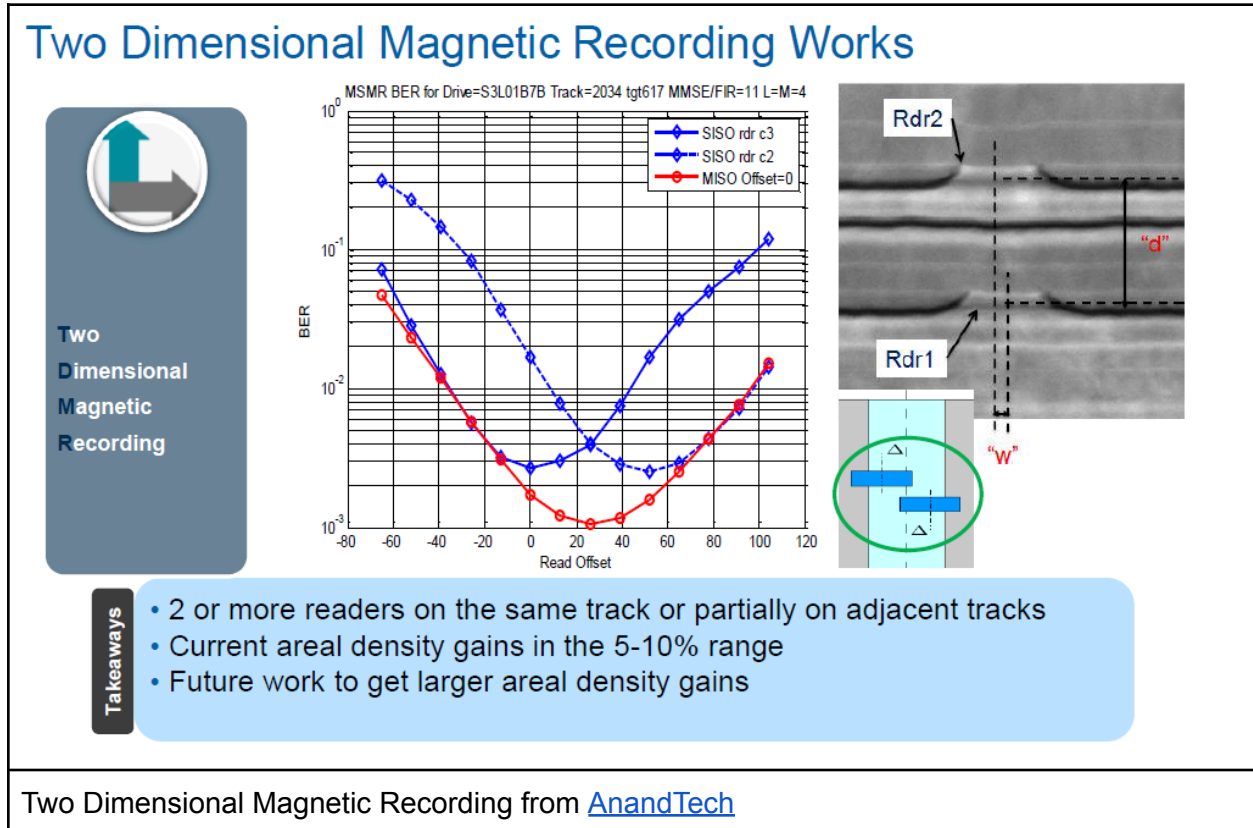
## OptiNand

Offload Adjacent Track Interference (ATI) and the need to rewrite data in sectors adjacent to sectors that are being frequently modified. Discuss OptiNand technology from Western Digital that uses embedded flash to hold metadata to track ATI.

## Beyond HAMR/MAMR

Beyond HAMR and MAMR, the technologies that have been discussed over the years by researchers are Bit Pattern Media (BPM) and Two Dimensional Magnetic Recording (TDMR). TDMR increases bit density by utilizing narrower data tracks, at the expense of lower signal to noise ratios caused by crosstalk between adjacent data tracks. To alleviate the crosstalk


problem, multiple read heads per track and more advanced signal processing are used to increase the noise margin. Note that TDMR is neither Shingle Magnetic Recording nor multi-actuator technology. The following diagram, from a 2015 [article](#) in AnandTech, shows how TDMR works.




Bit Pattern Magnetic recording (BPM) partitions magnetic particles or grains into magnetic “islands” that are more resistant to spontaneous bit flips than a uniform layer of grains. It is believed that both TDMR and BPM techniques can be applied to HAMR (Heat Dot Magnetic Recording) to increase areal bit density beyond pure HAMR. The following diagram from the same 2014 AnandTech article shows how BPM works.

## Heated Dot Magnetic Recording= BPM + HAMR

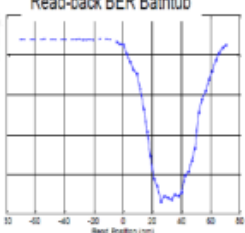
Continuous FePt film patterned @ 1Tdpsi to 5Tdpsi



**Heated  
Dot  
Magnetic  
Recording**



Read-back BER Bathtub



**BER = -2.43 @ 1Tb/in<sup>2</sup>**  
Spinstand testing and drive integration

	1 T	2 T	3 T
BCP			
C			
FePt			

**5 T**

**Takeaways**

- BPM: Multiple grains per bit to a single magnetic island per bit
- Demonstrated 1.5 Tdpsi Spinstand
- HDMR at 5Tdpsi and beyond looks feasible

Heated Dot Magnetic Recording from [AnandTech](#)

The Magnetic Recording Conference (TMRC), the annual IEEE sponsored conference for industry and academic researchers in the field of magnetic recording, surveys participants, pre and post conference, on the viability of different recording technologies. Prognostications by industry and academic researchers over the course of the past four conferences can be found at the following links:

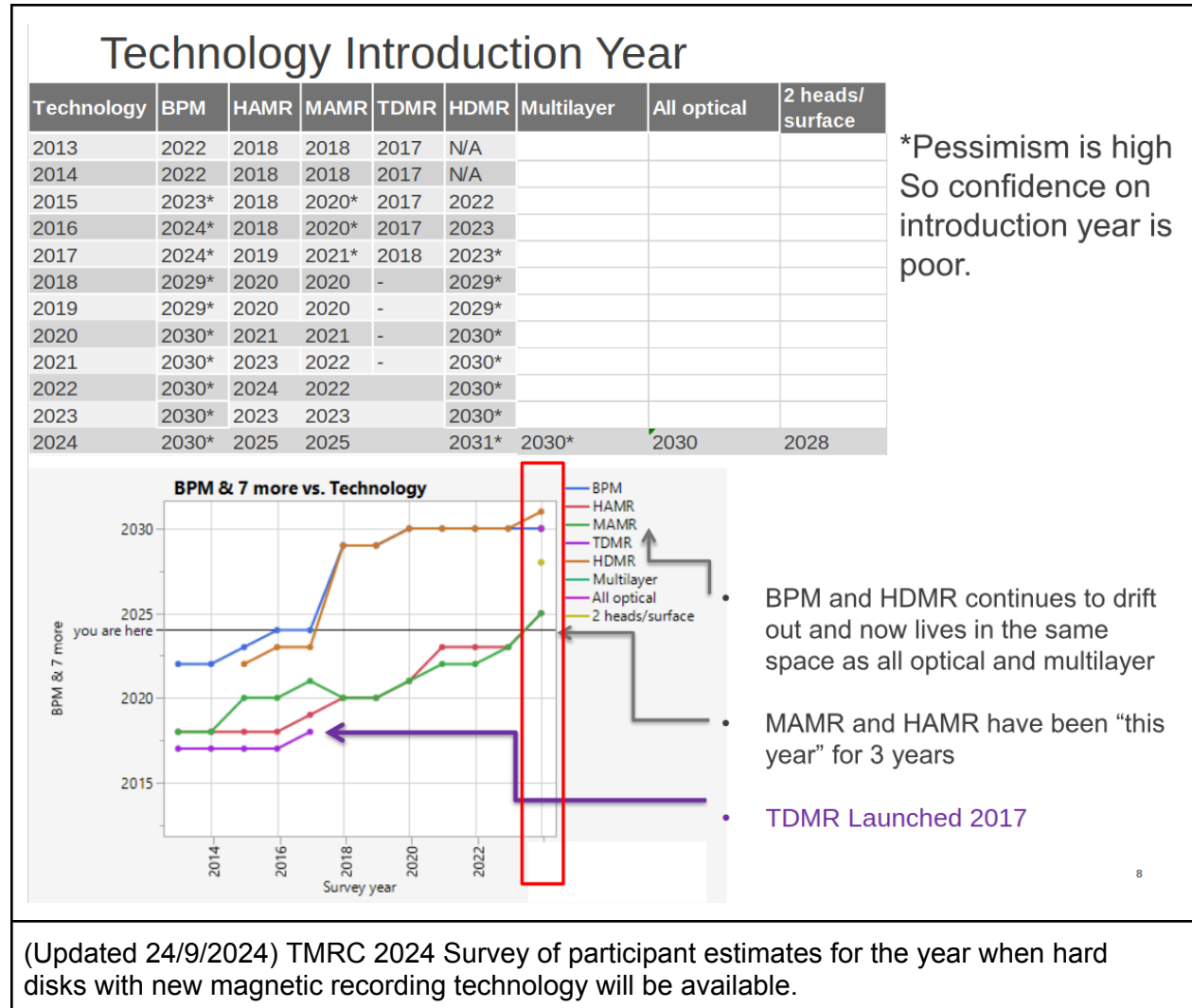
1. [TMRC 2022](#)
2. [TMRC 2020](#)
3. [TMRC 2019](#)
4. [TMRC 2018](#)

The results of the 2024 TMRC survey are shown in the figure below. The technologies being tracked are as follows :

1. BPM - Bit Patterned Media
2. HAMR - Heat Assisted Magnetic Recording
3. MAMR - Microwave Assisted Magnetic Recording
4. TDMR - Two Dimensional Magnetic Recording
5. HDMR (BPM+HAMR) - Heat Dot Magnetic Recording



Participants are particularly pessimistic with regards to BPM and HDMR. BPM is particularly problematic as it involves a [different manufacturing process and new production facilities](#) requiring significant investments.<sup>1</sup>



## Increasing Platter Count

For the past few years, HDD vendors have been stretching the capacity of PMR disk drives by using more platters, made possible through the use of helium gas.<sup>2</sup> Standard PMR disk drives, also referred to as “Conventional Magnetic Recording” (CMR) drives, from Seagate<sup>3</sup> and Western Digital<sup>4</sup> currently reach capacities of up to 20 TB and 22 TB respectively, using 10

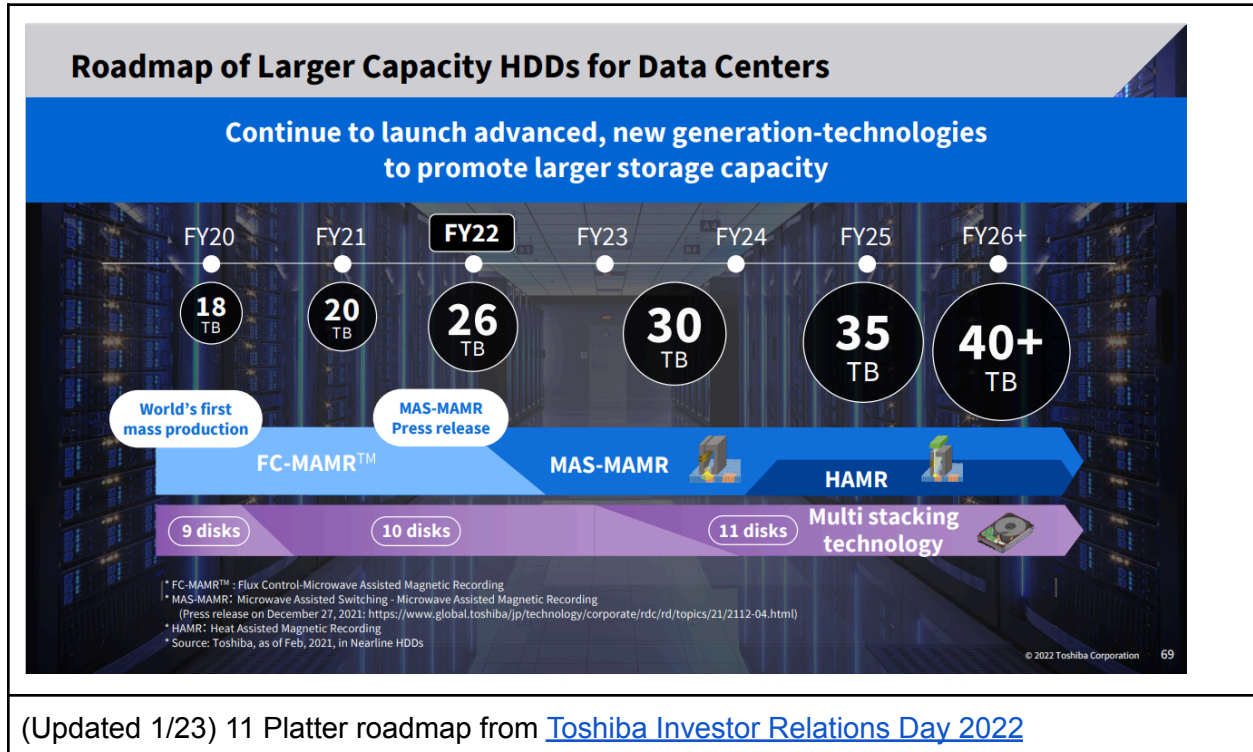
<sup>1</sup> [“HDD patterned media using jet-and-flash imprint lithography”](#), [Semiconductor-digest.com](#)

<sup>2</sup> “The Rise of Helium”, [Western Digital Blog July 2017](#).

<sup>3</sup> “Seagate launches 10-platter 20TB video surveillance disk drive”, [Blocks and Files](#).

<sup>4</sup> “Western Digital spins out 10-platter drives”, [Blocks and Files](#)

platters. [TrendFocus](#) believes that 11 or 12 platters are a path to larger capacity drives in the short term. Toshiba's HDD has demonstrated 11 platter MAMR drives,<sup>5</sup> from their [2022 Investor Relations Day](#). Beyond 20 TB vendors are looking at two separate approaches; Shingled Magnetic Recording and "Energy" Assisted Magnetic Recording.



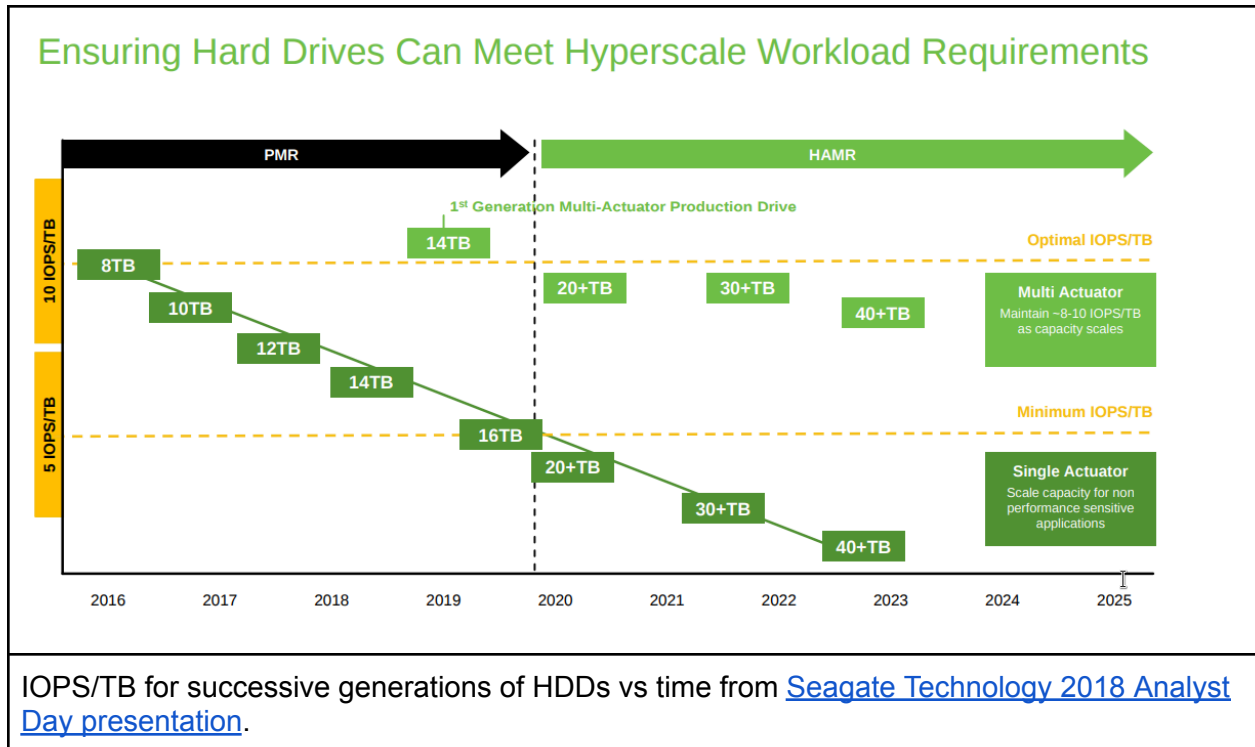
## Improving HDD IOPS

### Multi-Actuator HDD

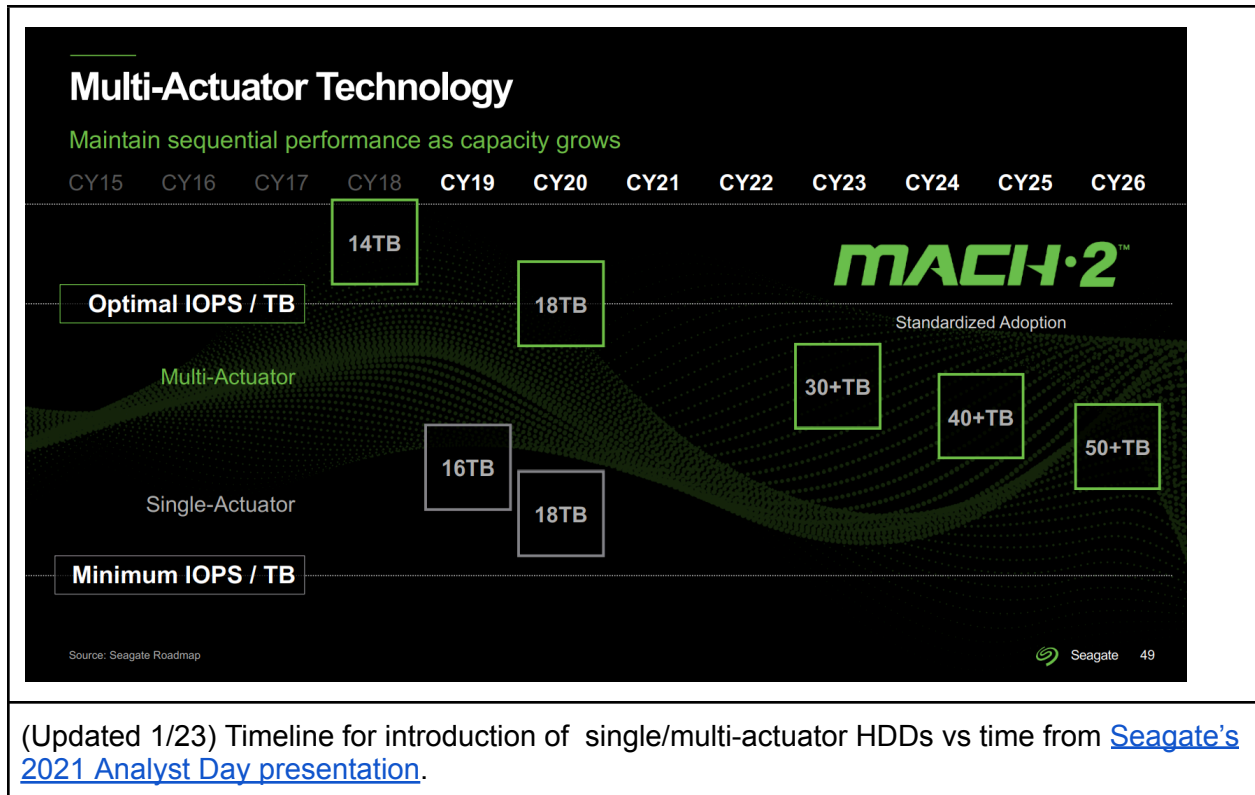
On the IOPS and I/O bandwidth front, [Seagate](#), [Western Digital](#) and [Toshiba](#) have announced dual actuator HDDs that are expected to provide a one time 2X increase in random IOPS and I/O bandwidth. This boost will stave off, at least temporarily, the declining IOPS/TB and BW/TB that is making HDD look more and more like magnetic tape from the perspective of access latency, as HDD capacities climb.

A schematic timeline for the availability of these technologies is in the [Seagate FY2018 presentation](#) to investors and is reproduced below. Whether intentional or not, during [Seagate's Analyst Day 2019](#) presentation, this diagram was reproduced, without the year explicitly mentioned on the time axis (x-axis).

<sup>5</sup> "Toshiba demos 32 TB HAMR and 31 TB MAMR disk drives", [Blocks and Files](#).



An updated timeline for multi-actuator drives from Seagate’s 2021 Analyst Day presentation is shown below:



The view from Western Digital is more ambiguous, as their President of Technology and Strategy Siva Sivaram stated “up to the 18TB capacity level we didn’t need dual actuators as customers are not saying they need them” ([blocksandfiles.com](http://blocksandfiles.com)).

All indicators suggest that dual actuator HDD are shipping products from both [Seagate](#) and [Western Digital](#), but are restricted to select [customers](#), most notably the public cloud [vendors](#). One interesting point of information with regards to dual actuator drives is that a single disk with total capacity 2X will be presented to the OS as two distinct LUNs of size X, for both [Seagate](#) and [Western Digital](#) drives.

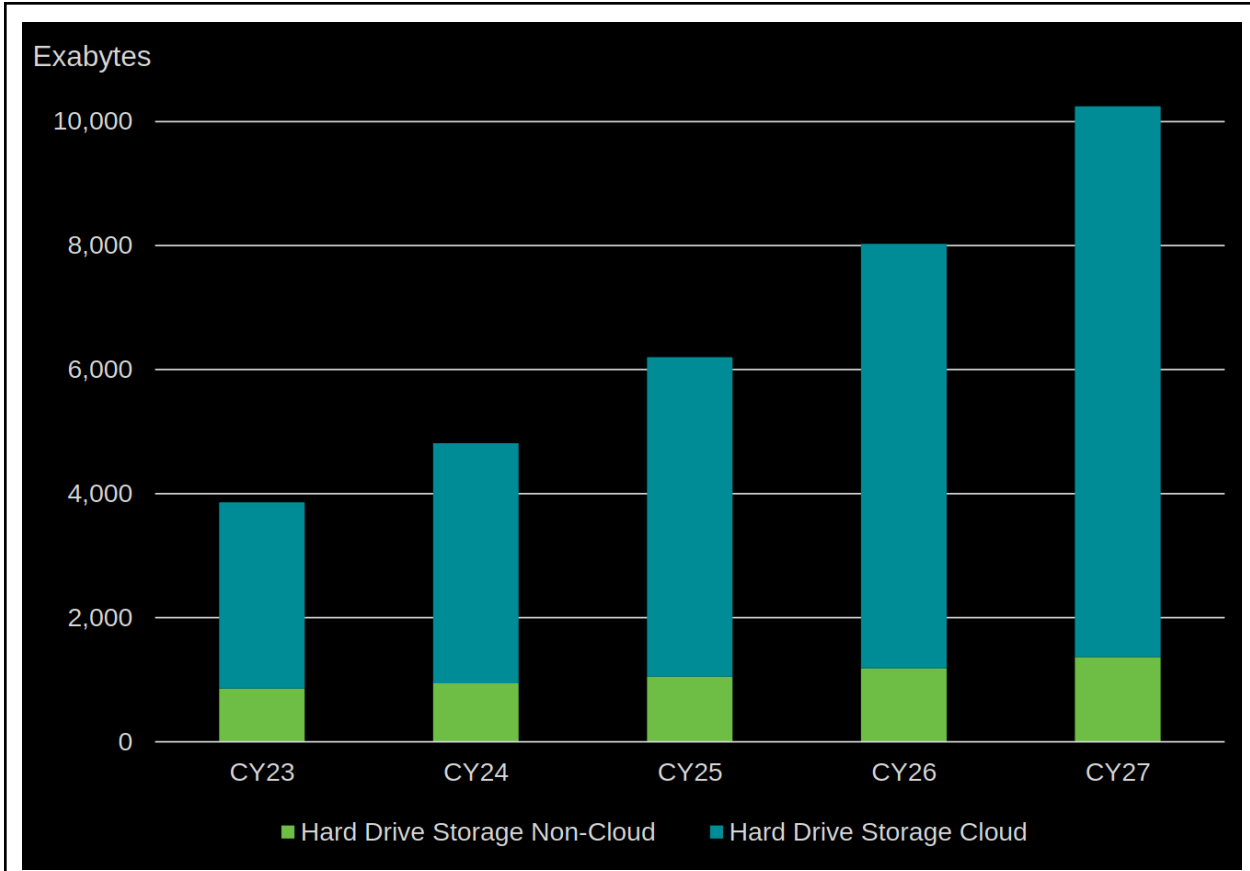
## Data Encryption and Sanitization

Although not a concern for bulk data storage systems holding HEP data, data protection is important for data storage systems holding proprietary data or potentially sensitive data, whether intentionally or by accident, e.g., mail servers and home directory servers. In the data center, HDDs factor into data protection at the end of their lifecycle. Proper disposal of HDDs typically involves sanitization of the drives to remove stored data. For those in the U.S., the relevant guidelines for sanitization is U.S. National Institute of Standards and Technology (NIST) [Special Publication 800-88 Revision 1](#). The relevant international standard is International Organization for Standardization (ISO) [ISO/IEC 27040:2015 standard](#). Selected Seagate and Western Digital HDD contain features that support proper sanitization of HDDs. Additional information can be found in the following documents from Seagate and Western Digital.

1. [Seagate Secure](#) technology paper
2. [Western Digital Instant Secure Erase](#) document

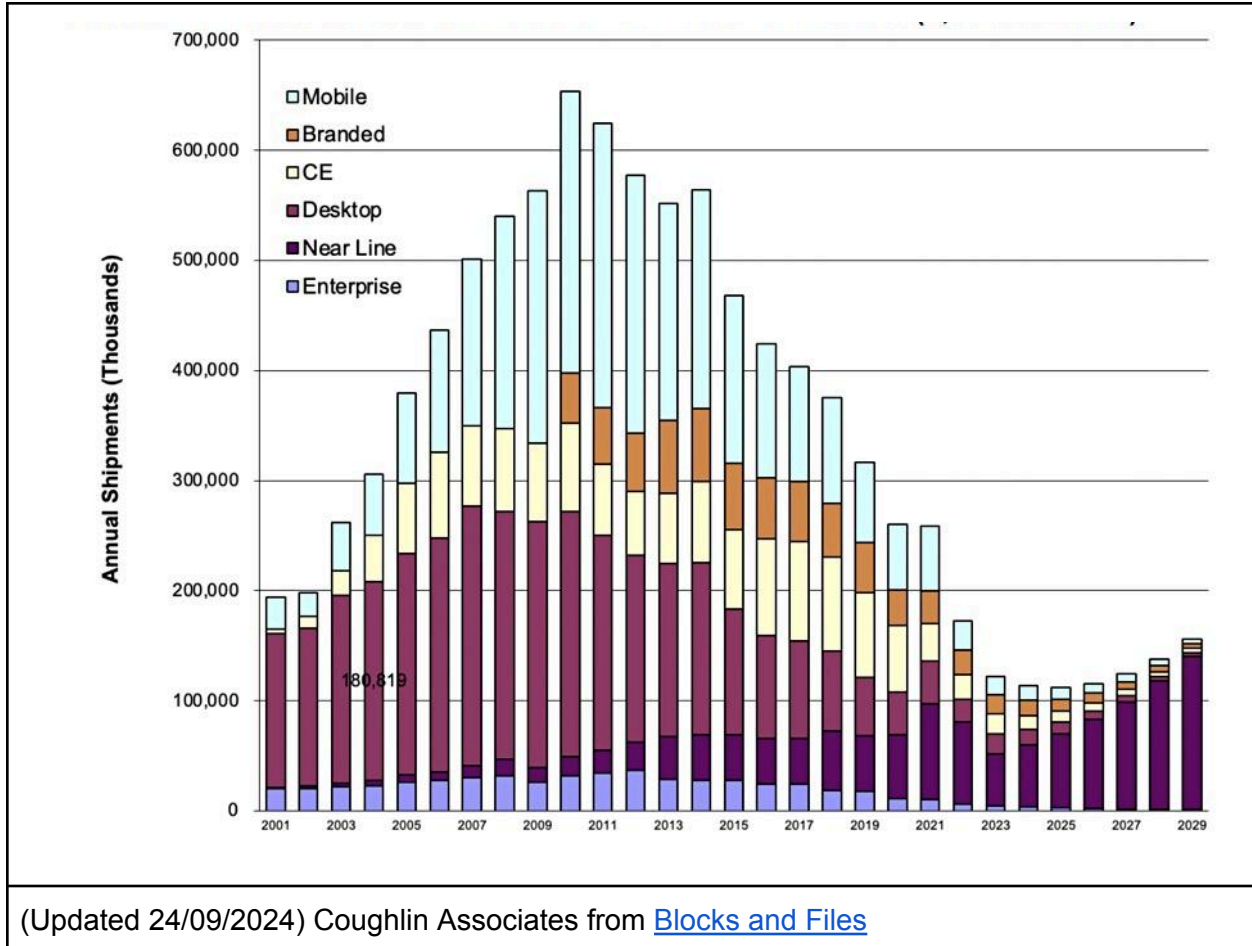
## Market Overview

Since the introduction of flash based solid state disks, HDDs have been reduced to near line disks, with an emphasis on capacity over performance and cost. As HDD has ceded market to SSD, the number of HDDs shipped has dropped dramatically from their peak in 2010. Within the nearline disk category, the vast majority of drives have been purchased by hyperscalers (cloud service providers).



HDD Market split between non-cloud and cloud storage. IDC Global StorageSphere, 2023 via ["Disk Trends"](#) presentation by Jon Trantham (Seagate) at the [Library of Congress Designing Storage Architectures for Digital Collections](#) meeting [April 2024](#)

In [Forbes](#) (May 29, 2020), industry analyst Tom Coughlin ([Coughlin Associates](#)) shows the contraction in unit sales of disk drives over the past 10 years.



Note the fact that drive shipments in all categories are either falling or stable, except “near line”, which are the type of disk used by the HEP/NP community. The implication is that HEP will not benefit from the economies of scale seen in previous years. Also, it is estimated that hyperscalers (Amazon, Google, Facebook, Microsoft) purchase almost [half of all disk drives](#), allowing them to potentially influence the direction of the market.

On this topic, it is interesting to note that Seagate, in their 2021 Analyst Day presentation, suggests that the hyperscaler and data center markets are diverging, with the former being more aggressive in moving to higher capacity disks (See the slide below). It should also be noted that current production HAMR and high capacity SMR drives are probably being directed to the public cloud. For example, DropBox stated in [2019](#) that they moved to SMR storage. In [2023](#), DropBox provided a status update on their migration to SMR. A [presentation](#) by Rick Kutcipal of Broadcom provides some additional insight into the utilization of HDDs by the hyperscalers and their influence on technical standards.

## Hyperscalers Pioneered a Blueprint for Mass Capacity

Mass Capacity Centric Architecture (90/10) with Rapid Adoption of highest capacity storage device

"Software-defined everything" with their own object storage software

Fleet of Edge Systems/Data movers/Shuttles to move Data into The Cloud

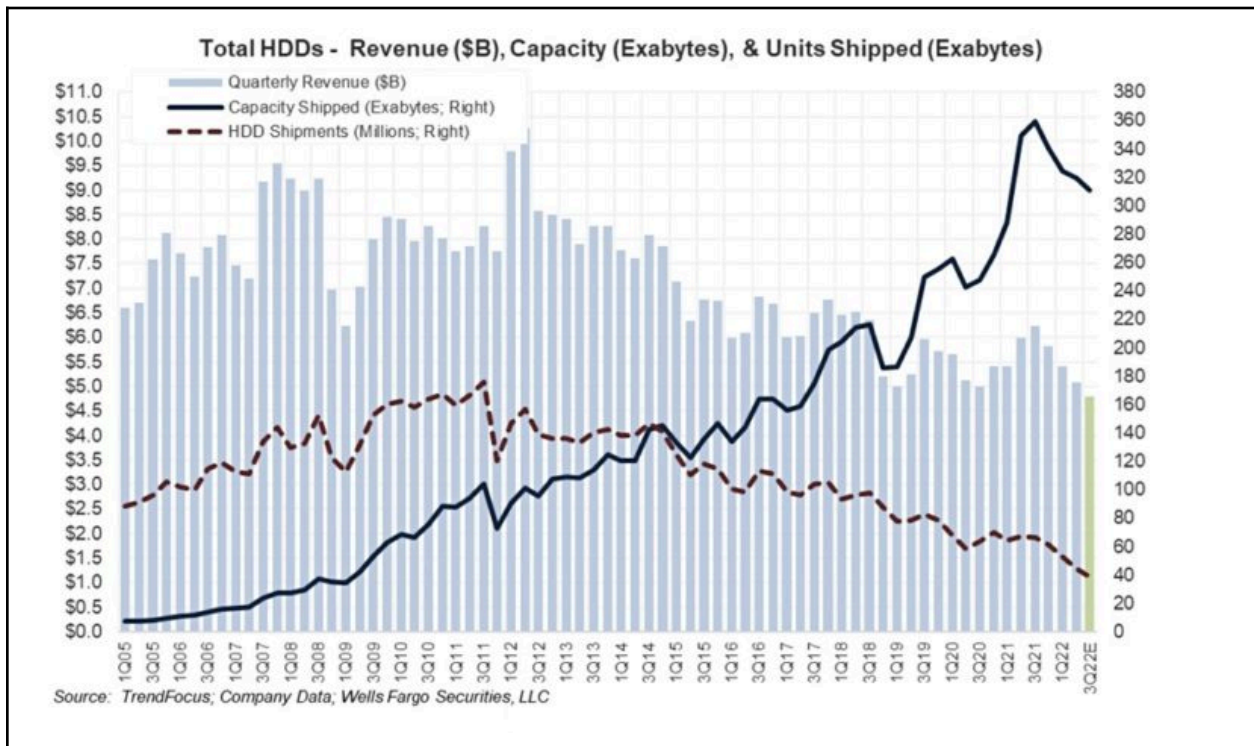
Seagate Enables an Open Mass Storage Architectural Platform

Chart for illustrative purposes only, deployment timeline based upon Seagate internal customer adoption data.

Seagate 65

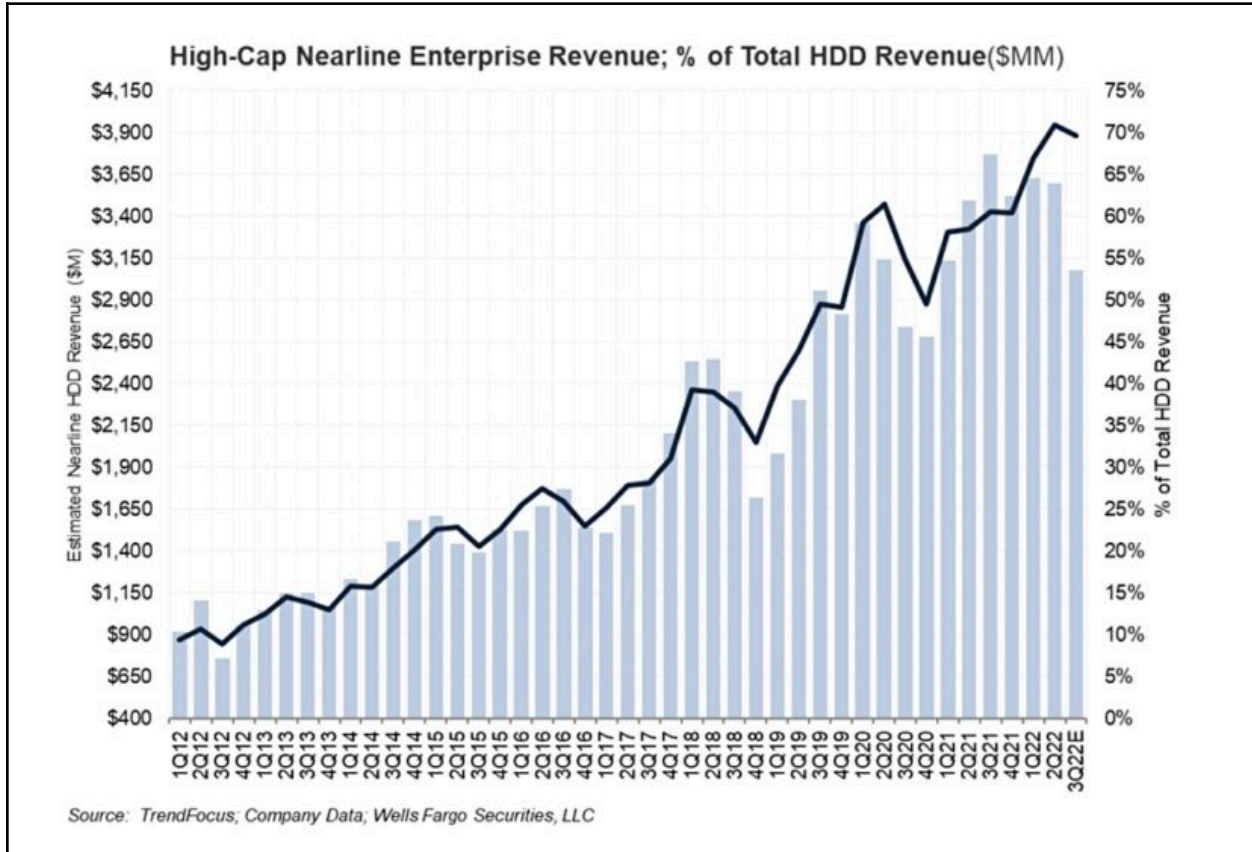
(Updated 1/23) Bifurcation in the nearline disk market (Hyperscaler vs Data Center) [Seagate's 2021 Analyst Day presentation](#).

The following graph from TrendFocus (via blocksandfiles.com) shows the revenue, capacity (exabytes) and units shipped for all HDD through the third quarter of 2022.



(Updated 1/23) Revenue, Capacity and Units shipped for all types of HDDs. Source: TrendFocus; Company Data;Wells Fargo Securities, LLC from [blocksandfiles.com](https://www.blocksandfiles.com)

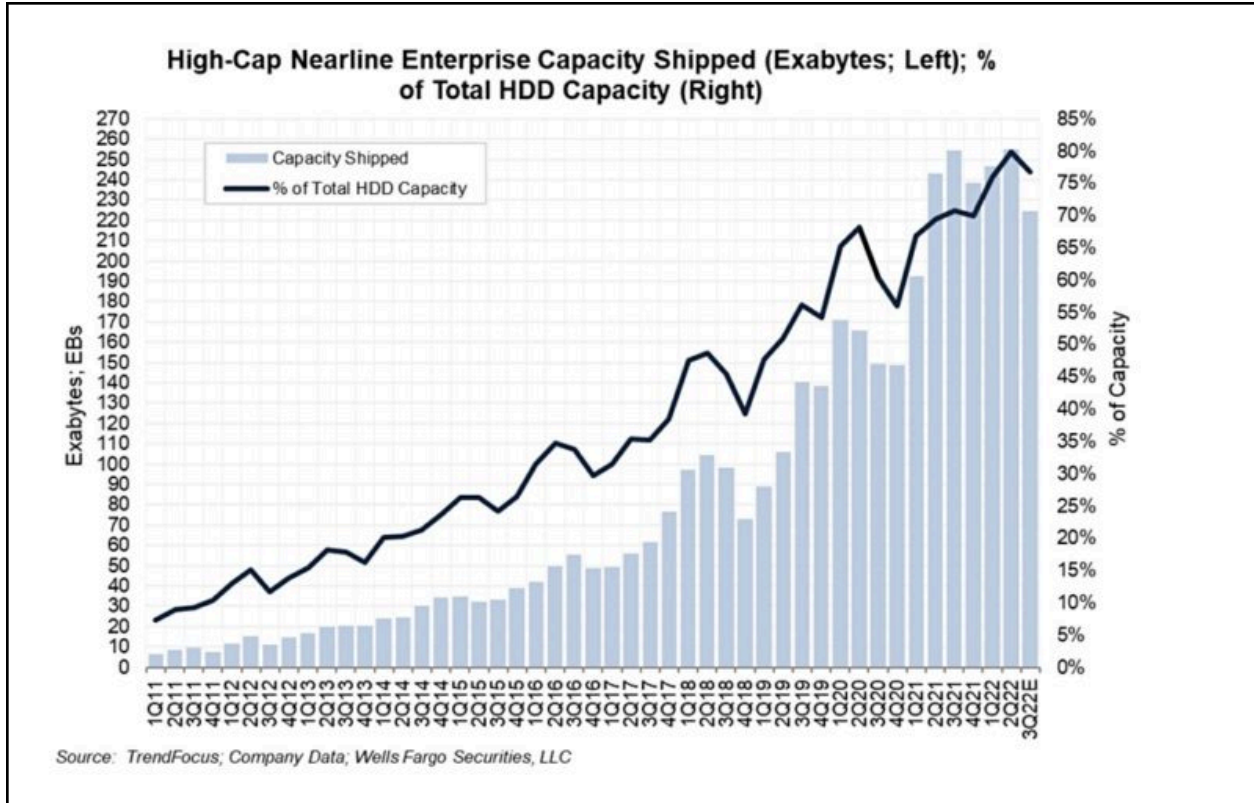
The following graph from [TrendFocus \(via blocksandfiles.com\)](https://www.blocksandfiles.com) shows the growth in revenue for nearline disks through the third quarter of 2022.



(Updated 1/23) High Capacity Enterprise HDD revenue as percentage of total HDD revenue. Source: TrendFocus; Company Data;Wells Fargo Securities, LLC from [blocksandfiles.com](https://www.blocksandfiles.com)

The following graph from [TendFocus](https://www.blocksandfiles.com) shows the growth in capacity shipped for high capacity nearline disks through the second quarter of 2022.





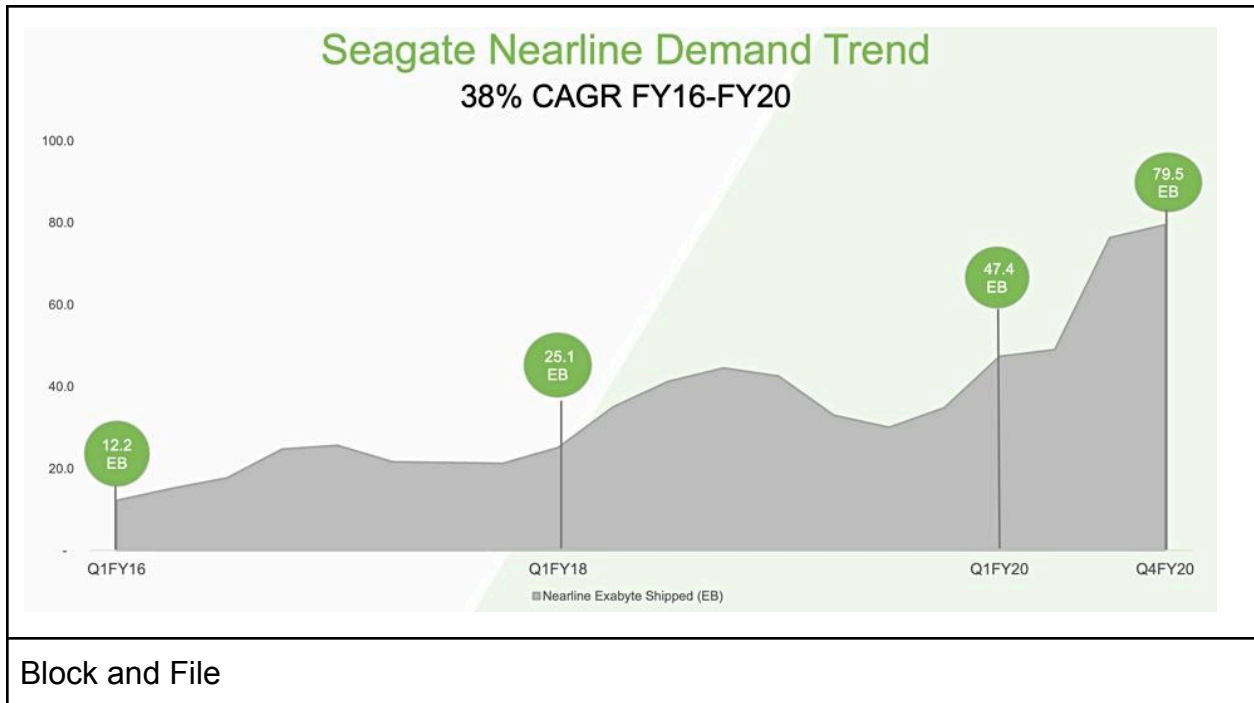
(Updated 1/23) High Capacity Enterprise HDD capacity shipped. Source: TrendFocus; Company Data;Wells Fargo Securities, LLC from [blocksandfiles.com](https://www.blocksandfiles.com)

Market share for high capacity enterprise HDD drives among the three HDD manufacturers is shown in the following table:

	Seagate	Toshiba	Western Digital
Capacity EB	~85.4	~27.9	~112.4
Capacity Share	~38%	~12%	~50%

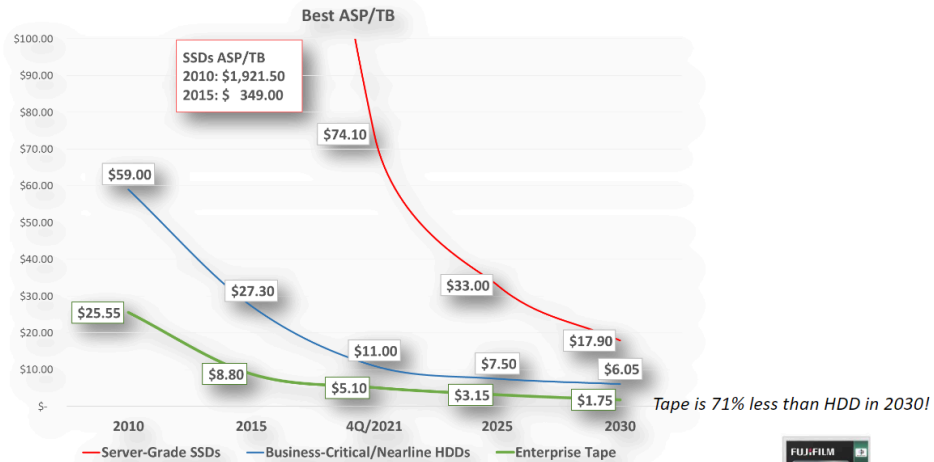
(Updated 1/23) High Capacity Enterprise HDD capacity share for the three HDD manufacturers. Source: TrendFocus; Company Data;Wells Fargo Securities, LLC from [blocksandfiles.com](https://www.blocksandfiles.com)

The growth in the nearline disk market is also reflected in the exabytes of nearline capacity shipped by Seagate through the fourth quarter of 2020 ([blocksandfiles.com](https://www.blocksandfiles.com)).



The following graph from FujiFilms presentation at the 2022 Flash Memory Summit, using data derived from “[The Escalating Challenge of Preserving Enterprise Data](#)” report from Further Market Research..

## Price Relationships: SSD, HDD, Tape

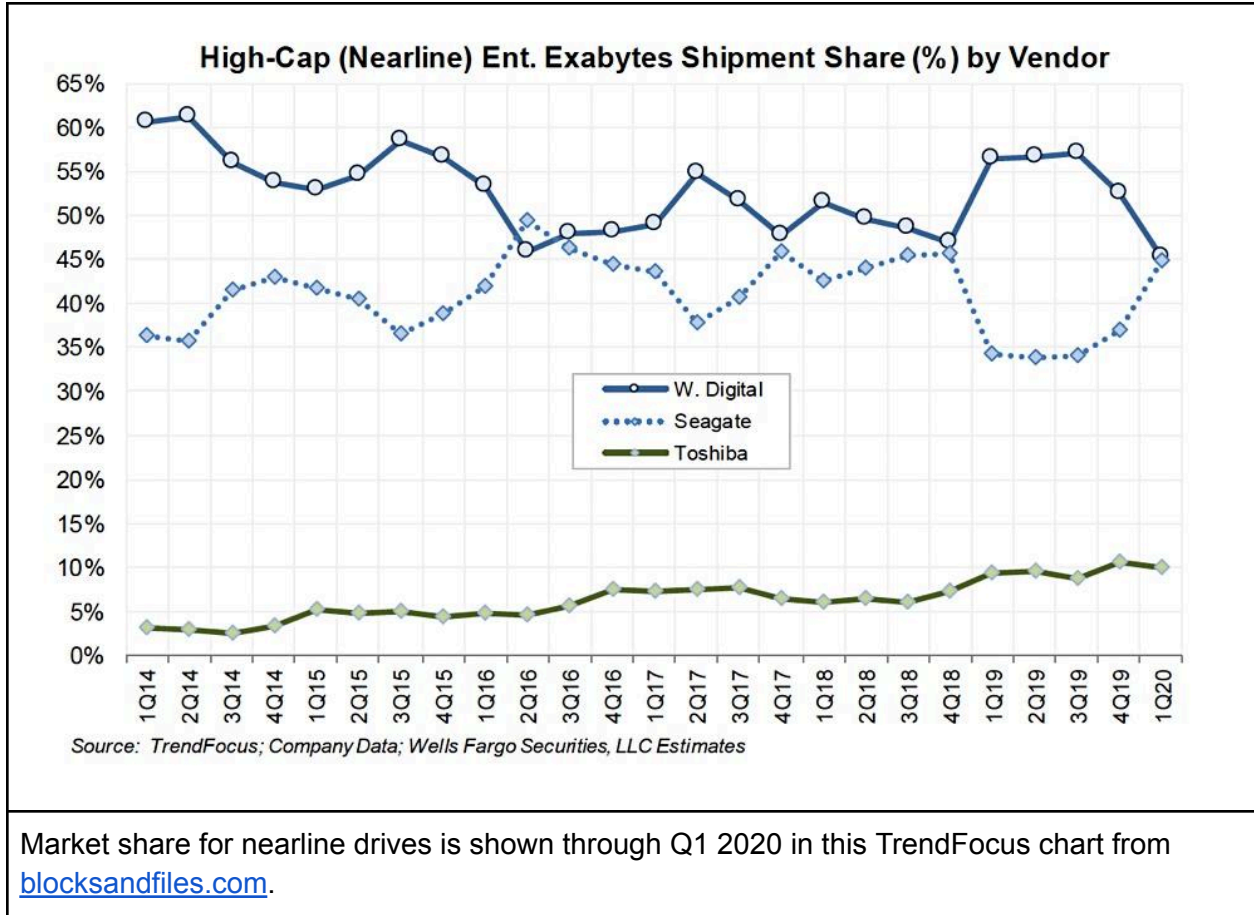


Source: The Escalating Challenge of Preserving Enterprise Data, Further Market Research, August 2022

4 | ©2022 Flash Memory Summit. All Rights Reserved.

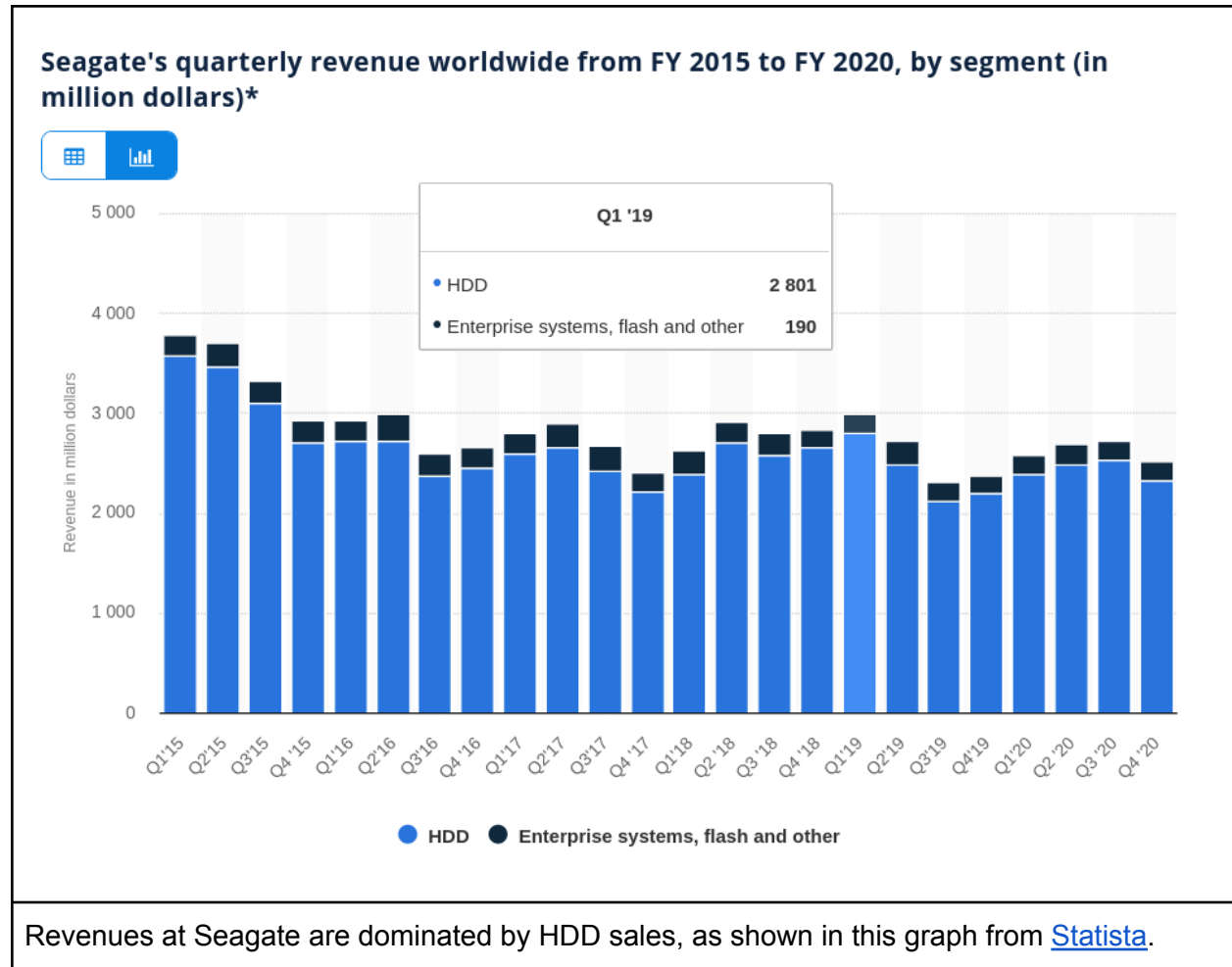


(Updated 1/23) Acquisition cost per TB for SSD, HDD, and Tape media. (Further Market Research via FujiFilms [presentation](#) at the 2022 Flash Memory Summit)



Market share for nearline drives is shown through Q1 2020 in this TrendFocus chart from [blocksandfiles.com](https://www.blocksandfiles.com).

While both Western Digital and Seagate are heavily invested in hard disks, only Seagate can be considered a “pure play”. Revenues at Seagate are dominated by HDD sales, as shown in this graph from [Statista](https://www.statista.com).



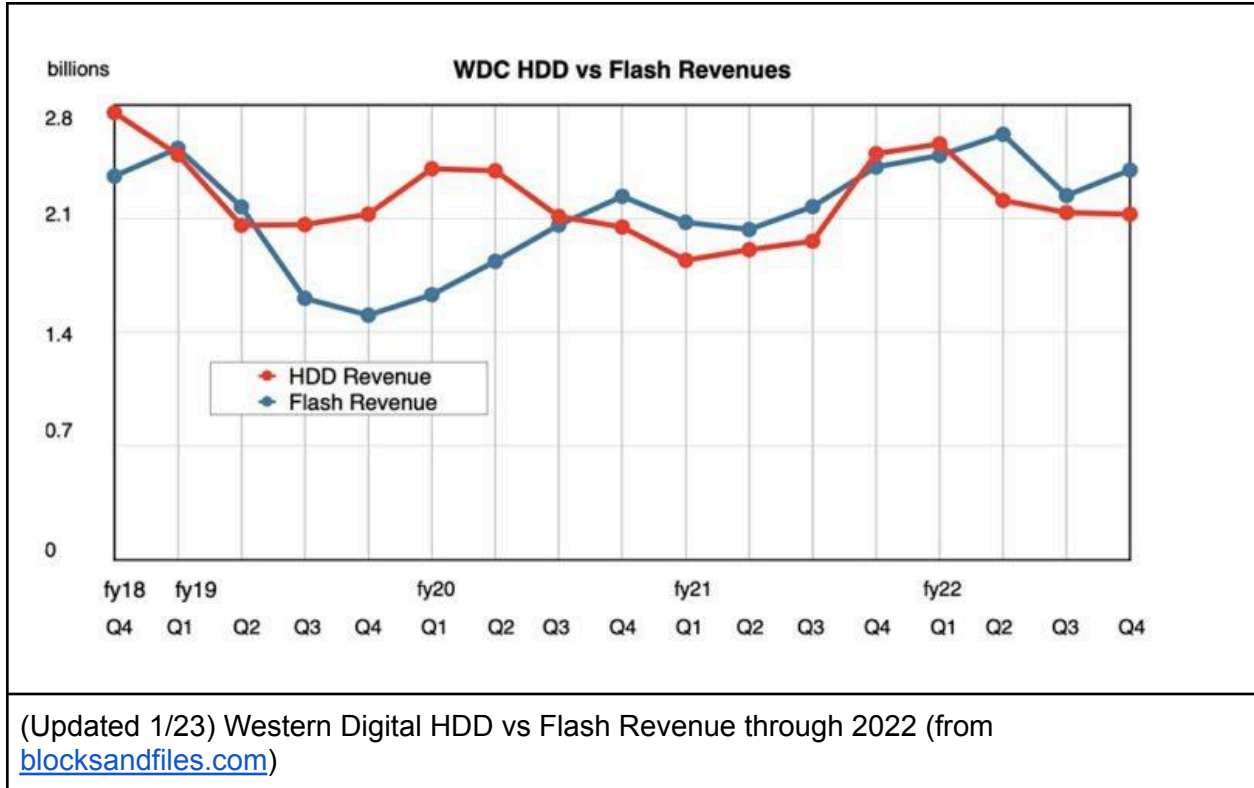
In contrast, WD flash revenues are roughly on par with its HDD revenues as shown in the following charts from [blocksandfiles.com](#). In October 2023, WD announced plans to split the company into two standalone, publicly traded companies, one focusing on HDD and the other focusing on flash.<sup>6</sup> WD reported progress on this split in March of 2024.<sup>7</sup>

<sup>6</sup>

<https://www.westerndigital.com/company/newsroom/press-releases/2023/2023-10-30-western-digital-to-form-two-independent-public-companies>

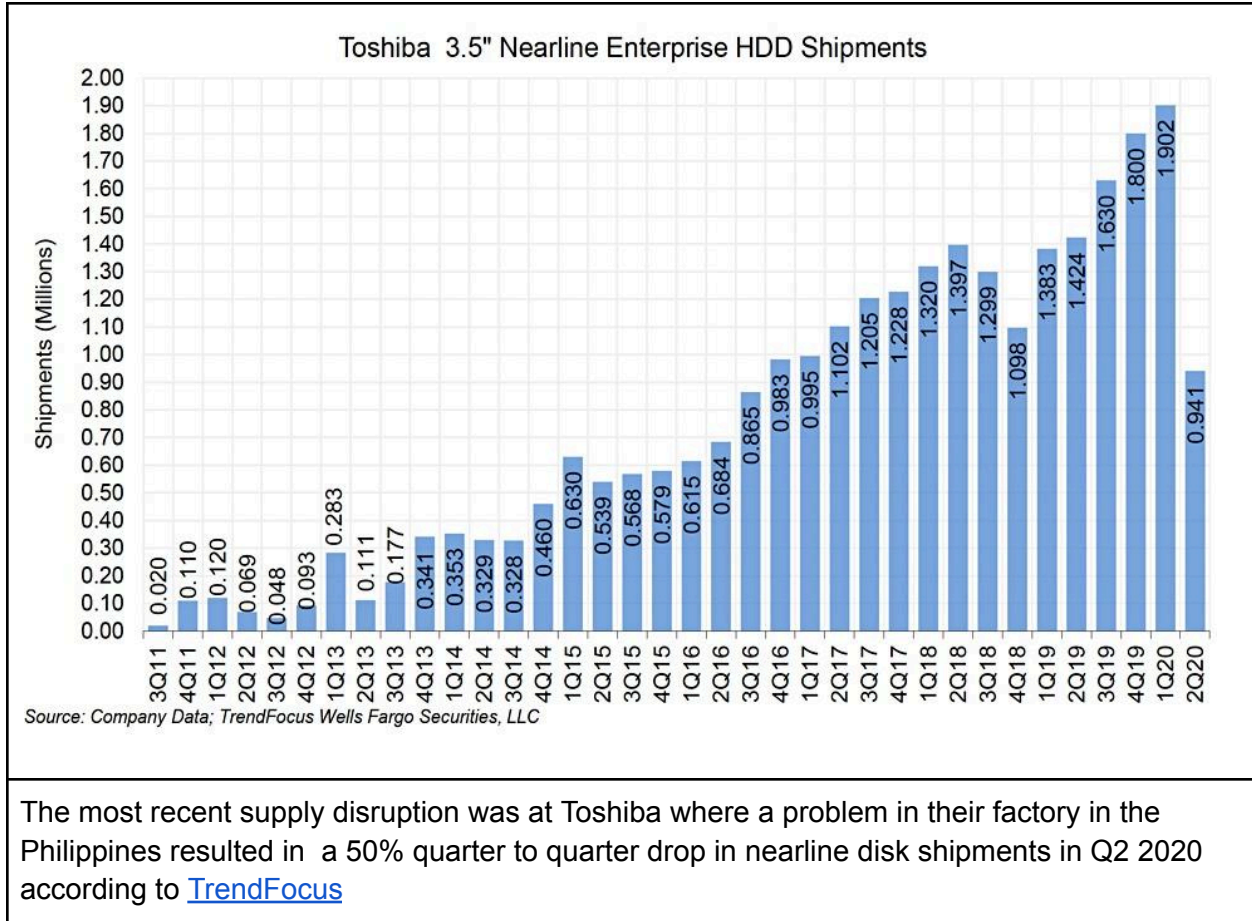
<sup>7</sup>

<https://www.westerndigital.com/company/newsroom/press-releases/2024/2024-03-05-western-digital-announces-update-on-company-separation>



## Supply Risks

With only three vendors, supply disruptions is an issue. This is further exacerbated by the reduction in manufacturing plants. Seagate and Western Digital have been closing HDD plants, the latest being a Western Digital plant in Malaysia according to the [WD 2018 annual report](#). Both companies are now down to [two plants each](#). The most recent supply disruption was at Toshiba where a problem in their factory in the Philippines resulted in a 50% quarter to quarter drop in nearline disk shipments in Q2 2020 according to [TrendFocus](#). To mitigate this risk, Toshiba has expanded production to a new factory in [China](#).



## Supporting Technologies

### Interconnect

1. Interconnect Technology
  - a. SATA
  - b. SAS
  - c. Fibre Channel
  - d. PCI-e/NVMe/NVMeoF
  - e. Infiniband ?
  - f. Ethernet ?

## SATA

Serial ATA (SATA) has been the interconnect of choice for consumer HDDs and looks to keep that position in the market. In the past it has been the interconnect of choice for SSDs, but consumer SSD are rapidly transitioning to NVMe in conjunction with the adoption of the M.2 form factor. PCI-e looks to be the interconnect of choice for SSD, excluding extremely cost sensitive and low performance applications. For these latter cases, SATA is likely to remain the interconnect of choice. The SATA standard continues to be updated, with the most recent version being [SATA 3.5a](#); however, maximum link speed has remained at 6Gbps. The bulk of the updates to the SATA 3.x standard have been mostly feature additions, not performance enhancements.

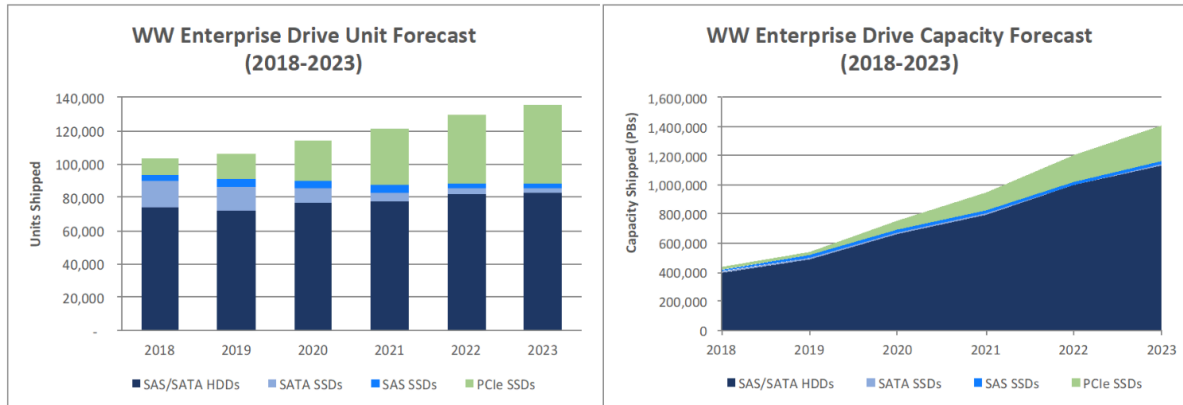
At this point in time, SATA seems to be relegated to low cost/low performance HDD device connectivity, but even this position may be tenuous given the addition of support for HDDs in the latest NVMe revision, [NVMe base specification 2.0c](#).

## SAS

Serial Attached SCSI (SAS), has been the interconnect of choice for “enterprise” HDD’s. In recent years, SAS has branched out to HDD expansion chassis connectivity and direct attached storage connectivity (DAS), particularly for JBOD deployments. For enterprise HDD connectivity and HDD DAS connectivity SAS continues to be the interconnect of choice. For DAS, this is the case as SAS supports SAS, SATA, and PCI-e (through SAS Express). Market penetration for SAS connected SSD is unclear. The slide shown below, from 2019 Storage Developer Conference presentation by Cameron Brett of the SCSI Trade Association shows the historical and projected future market share for connectivity types to “enterprise” storage drives.\



## SAS Remains Primary Enterprise Storage Interface



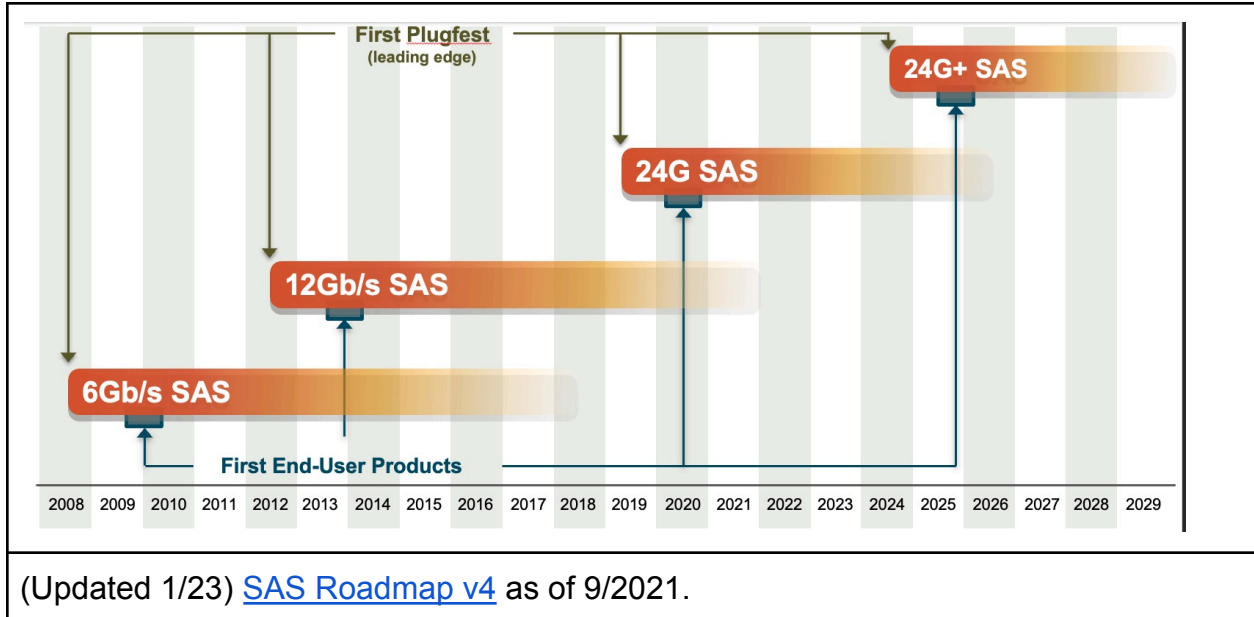
Source: IDC, May 2019

**SAS Infrastructure Enables >64% of Enterprise Storage Drives and >80% of Enterprise Storage Capacity thru 2023**

2019 Storage Developer Conference. © SCSI Trade Association. All Rights Reserved.



[SAS Express](#), SAS with PCI-e support appears to be a “transition” technology (with potentially limited life) for SSDs. The official [SAS/SCSI roadmap](#) was more recently updated in 2021. Underlying 24 Gbps SAS chip sets started appearing in 2019 and end user products started appearing in 2020. Moving forward, with the advent of NVMe, the open question is : What is the future of SAS ? The best guess at this point is that SAS will be relegated to “legacy” enterprise HDD connectivity and DAS connectivity, but like SATA, [NVMe is looking to replace SAS](#) as the storage interconnect standard of choice.



Fibre Channel

The official [Fibre Channel roadmap](#), now at version 23, was updated in 2020. Development and market availability of higher performance Fibre Channel versions (generations) have been pushed out by one to two years. The bandwidth chart is shown below

Product Naming	Throughput (Mbytes/s)*	Line Rate (Gbaud)	T11 Specification Technically Complete (Year) †	Market Availability (Year) †
8GFC	1,600	8.5 NRZ	2006	2008
16GFC	3,200	14.025 NRZ	2009	2011
32GFC	6,400	28.05 NRZ	2013	2016
64GFC	12,800	28.9 PAM-4	2017	2020
128GFC	24,850	56.1 PAM-4	2022	2024
256GFC	49,700	112.2 PAM-4	2025	Market Demand
512GFC	TBD	TBD	2029	Market Demand
1TFC	TBD	TBD	2033	Market Demand

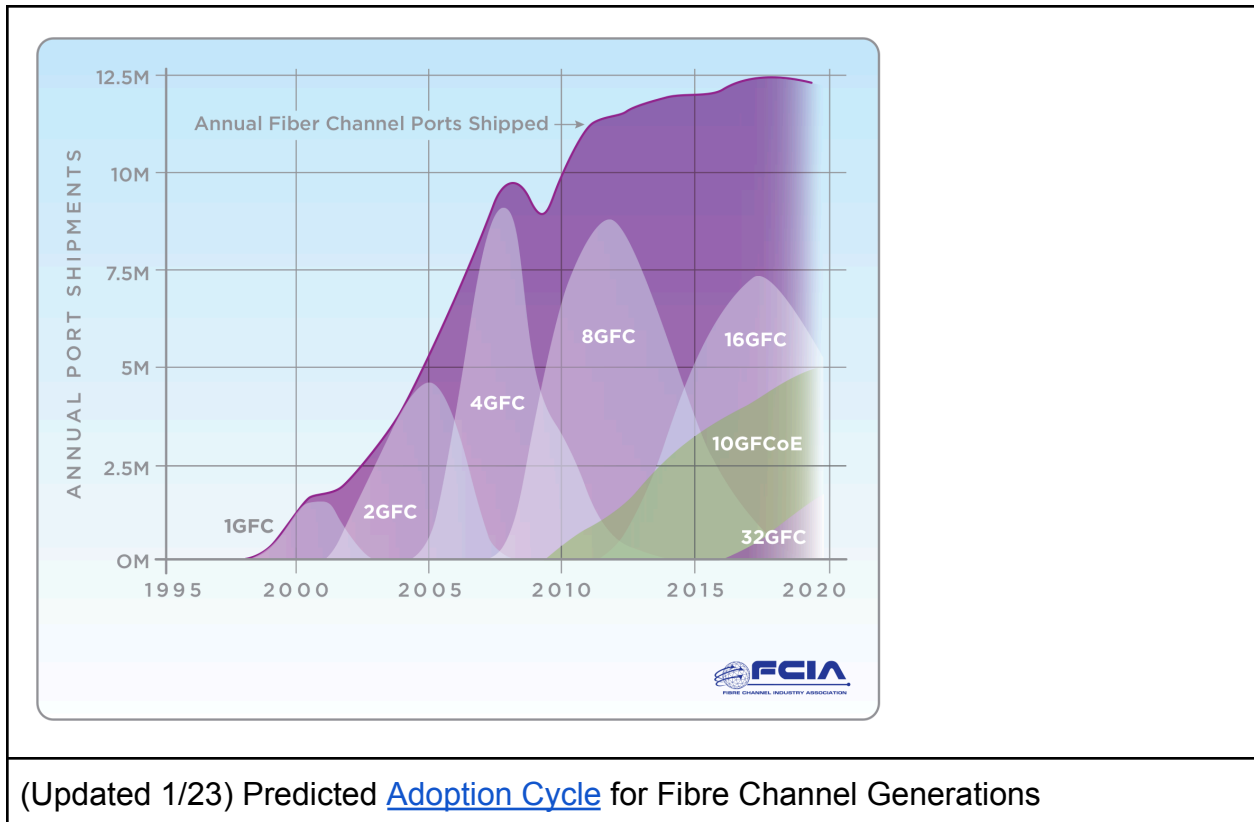
Fibre Channel Industry Association [2024 Fibre Channel Product Roadmap](#)

The end of 2018 saw the introduction of 64G Fibre Channel (“Gen 7”) [64Gb chip and HBAs](#) and 2019 saw the introduction of “Gen 7” [switches](#). Bandwidth and connectivity options (single mode, multi mode, NRZ/PAM4) mirror those for Ethernet.



(Updated 1/23) [Fibre Channel Roadmap](#) as of 2020

The predicted adoption cycles for the various generations of Fibre Channel from the Fibre Channel Industry Association are shown below:



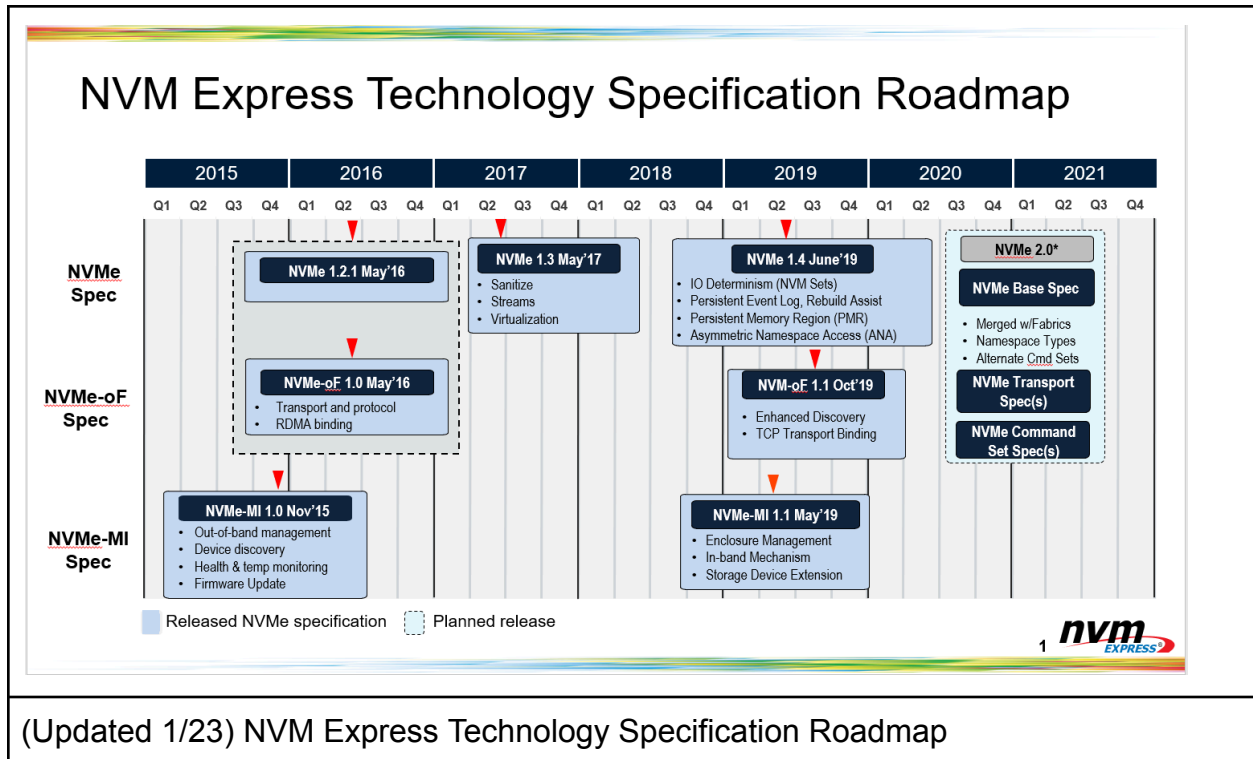
With only three vendors, Broadcom (switches/chips/HBA), Cisco (switches), and Marvel (HBAs), the Fibre Channel business shows all the signs of a mature market and technology. Fibre Channel continues to be the interconnect of choice for [Storage Area Networks and may yet have a role with NVMe-oF](#).

For HEP, Fibre Channel continues to be a critical interconnect technology, with FC adaptors being necessary within tape environments and represent a significant fraction of the tape mover cost.

NVMe/PCI-e/NVMe-oF

NVMe-oF storage connectivity - Future connectivity to “exotic”, high performance storage ?

The feature development roadmap for NVMe from the NVMeExpress website is shown in the image below. More recent updates may be available from the upcoming [2020 Storage Developers Conference](#).



(Updated 1/23) NVM Express Technology Specification Roadmap

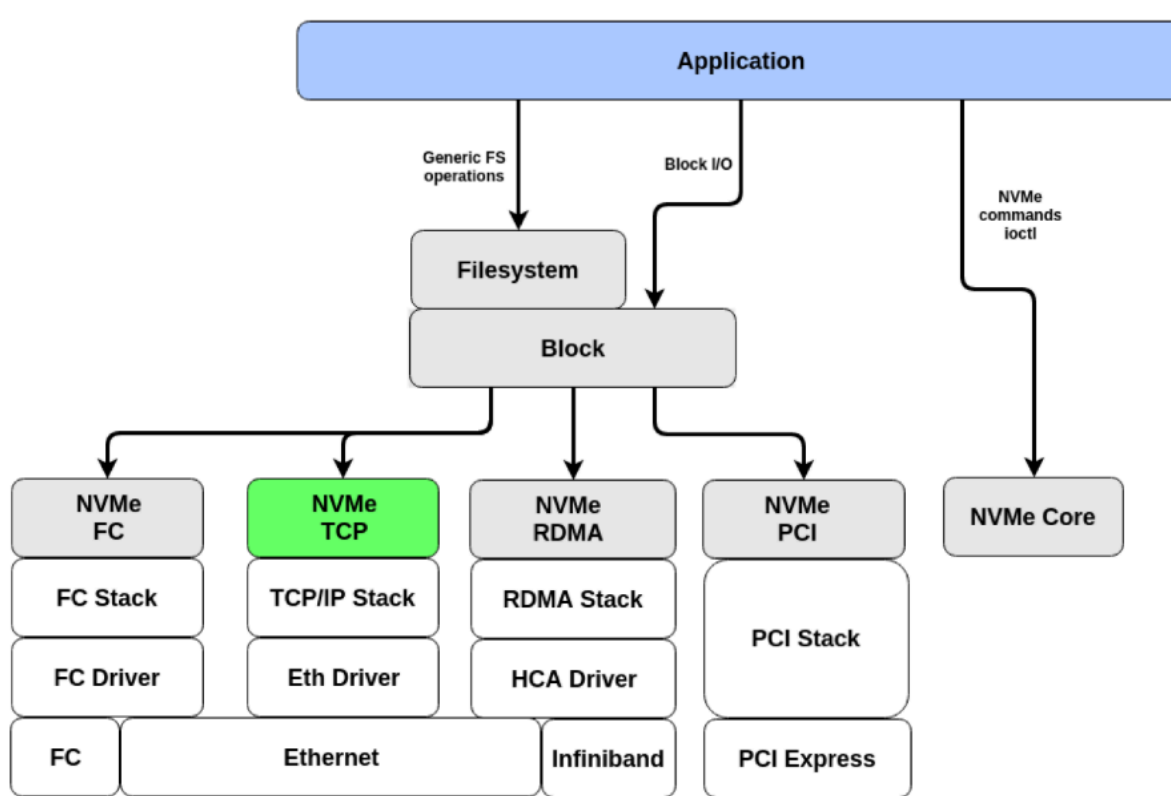
One of the goals of the NVMe developers is to have the NVMe software interface become the “lingua franca” of storage connectivity, regardless of the underlying physical layer, as shown in the diagram below from the SDC NVMe presentation.

<Prognostications about future of NVMe and Flash

<https://www.forbes.com/sites/tomcoughlin/2020/05/15/nvme-for-all-data-center-storage/#56ed87077e05>

<https://blocksandfiles.com/2020/05/27/nvme-universal-block-storage-access-protocol/>

Summary necessary>



Source: Kam Eshghi (Lightbits Labs)

The following graph from a [presentation](#) made by [G2M Research](#) at the 2018 NVMe Developer Days conference shows the expected growth and segmentation of the NVMe storage market.

2.

## Organization and Access

Advances in storage hardware is just one driver of improvements in capacity, cost, and performance of storage services. Two other drivers are advances in system architecture and improvements in software that take advantage of the new hardware.

<Information on Zoned Storage SCSI ZBC and SATA ZAC and application to SMR disks>

## Data Protection

### Traditional RAID

(Updated 1/23) RAID (Redundant Array of Independent Disks) has been the technology of choice for data protection for over three decades. RAID 1 (mirroring) emphasized I/O performance at the expense of higher cost (storage capacity). RAID 5 (data+parity bit) emphasized lower cost but sacrificed I/O performance, particularly with writes. As HDD capacities have increased, while I/O performance has effectively stalled, rebuild times failed disks in a RAID 5 have increased dramatically. When coupled with the increased likelihood of an additional drive failing during a rebuild, additional parity bits have been added over time, with RAID 6 (or ZFS RAIDz2) using 2 bits and ZFS RAIDz3 using 3 bits.

### Distributed RAID

(Updated 1/23) Increasing the number of parity bits in a traditional RAID LUN significantly reduces the probability of data loss due to multiple disk failures; however, they do not fix the fundamental problem of longer RAID rebuild times. Distributed RAID systems like ZFS [dRAID](#) and NetApp [Dynamic Disk Pools](#) were developed to address the rebuild time problem. With traditional RAID LUNs N data disks and P parity disks are explicitly assigned to a N+P disk LUN. Should a drive fail in the LUN, data from the failed disk is reconstructed from the remaining drives and written to a replacement disk (cold or hot spare). Reconstruction time is determined by the write performance of a single disk, ignoring contention on all disks in the LUN from normal data access I/O. With distributed (or declustered) RAID, data and parity bits can be distributed over any of the available disks. “Logical” hot spares are constructed from reserved capacity in all of the available drives. In the event of a disk failure, reconstructed data from all of the remaining disks holding the data and written to the reserved capacity in all of the remaining disks (but taking into account the need to maintain data protection). With distributed RAID, rebuild times are dramatically [reduced](#).

### Erasure Coding

(Updated 1/23) [Erasure coding](#), as colloquially defined, is the application of RAID-like data protection techniques across larger systems like RAID systems, storage server nodes, equipment racks, and “availability zones”. Erasure coding is primarily associated with non file system storage, typically but not restricted to object storage systems, e.g., [Ceph](#), [Cortx](#), [MinIO](#) and [Hadoop](#); however, it is also available in more traditional file system storage systems like [Lustre](#). Through the appropriate configuration of storage hardware and software, these erasure coding systems can

## HEPiX Techwatch : Disk Storage

tolerate (and recover) from the failure of storage systems (complete HW RAID arrays), storage servers (Lustre OSSes), storage/server racks (Hadoop) and entire data centers.