ADVANCED
AF
FORMAT

Microsoft Meeting
August 24th, 2011

Sponsored by IDEMA

HITACHI

Seagate

WD Western Digital

TOSHIBA
Leading Innovation >>>

IDEMA

SAMSUNG

Agenda

- **Introductions**
- **Advanced Format Roadmaps and Futures**
- **HDD LBA Indirection and Windows**
- **Hard Disk Drives, Hybrids, and SSDs**
- **Proposed WHQL Requirements for Storage**
- **Windows Features and timelines**

Meeting Expectations

- Shipped AF support in Windows 7 and Beyond
    - Thanks and Review
    - Impact of published AF support documents
    - How we partner on better educational tools
- Understanding SMR and LBA Indirection
    - SMR WG and hinting initiatives underway
    - Where are disk resources best managed?
- How upcoming requirements may impact the shipping market
    - HDD (legacy, AF, and SMR) Power Requirements
    - Booting in UEFI and POST time requirements
- How upcoming Windows features support HDDs and…

AF History & Direction

- Advanced Format HDDs shipping for 18 months in AF 512e Formats – No more education required, industry adopting

- Presentation and educational efforts focused on Enterprise customers in last year

- AF 4Kn conversations are on track with most OEMs

  - Most customers are looking for well-defined transition paths between legacy 512b and 4kn to include 512e as the real solution

  - IDEMA member companies having to counter publications that say AF HDDs are less reliable than legacy with unneeded qualification validation

- AF has lead to meaningful conversations on SMR and more LBA indirection to the host

  - Enterprise customers excited about the possibilities

Roadmaps

Get Ready!

- *Embrace Advanced Format 512e – It's here, it's good, it's a done deal…*

- *Use an Advanced Format Optimized OS*
  - *Windows 7 with SP1*

- *If you can't – Consult your HDD supplier for recommendations on how to best manage*

- *Began preparing for Advanced Format 4Kn – 4K native launch on enterprise-class drives coming in 2012*

- *Make sure utilities and other core software packages are AF Optimized – Check with you SW vendors!*

- *Use the free resources available through IDEMA*
  - *www.idema.org/af*

ARE YOU READY?

Advanced Format

ADVANCED AF FORMAT

ADVANCED AF FORMAT

Are you ready for ADVANCED FORMAT technology?

THE HARD DISK DRIVE INDUSTRY IS TRANSITIONING TO LONGER DATA SECTORS.

IN JANUARY 2011, THE PHASE OVER FROM 512 BYTES/SECTOR TO 4,096 BYTES WILL BEGIN.

We want to make sure you are ready

Why do hard disk drives need ADVANCED FORMAT technology?

- Continued storage capacity growth
- Increased format efficiencies
- Stronger error correction capabilities
- Better data integrity at higher densities

Why prepare for ADVANCED FORMAT hard drives

- Optimize product performance
- Ensure compatibility
- Deliver positive user experience
- Avoid technical support calls

What is affected by ADVANCED FORMAT?
Direct disk access operations used in:

- Software applications
- Software drivers
- Back-up programs
- Disk replication solutions
- Imaging tools
- Test tools
- Integration utilities
- Manufacturing processes
- Database applications
- SoC and BIOSes
- Operating Systems
- And more

HITACHI

Seagate

WD Western Digital®

TOSHIBA
Leading Innovation »»

IDEMA®

SAMSUNG

Get ready! Test your products today. Register for an AF Developer Kit at: **www.idema.org/af**

Hitachi GST AF Roadmap

- All new Hitachi hard drives will be Advanced Format

- First Advanced Format Mobile Product

  - Travelstar® 5K750 & 7K750
  - 750/640/500 GB
  - Shipping since 4Q 2010

- All upcoming Desktop Products <u>will</u> incorporate AF

- More information about Advanced Format

  - *http://www.hitachigst.com/internal-drives/advanced-format-drives*

  - *Advanced Format White Paper*

  - *"Hitachi Align Tool" is available for download on Hitachi GST website to address possibility of misaligned sector boundaries*
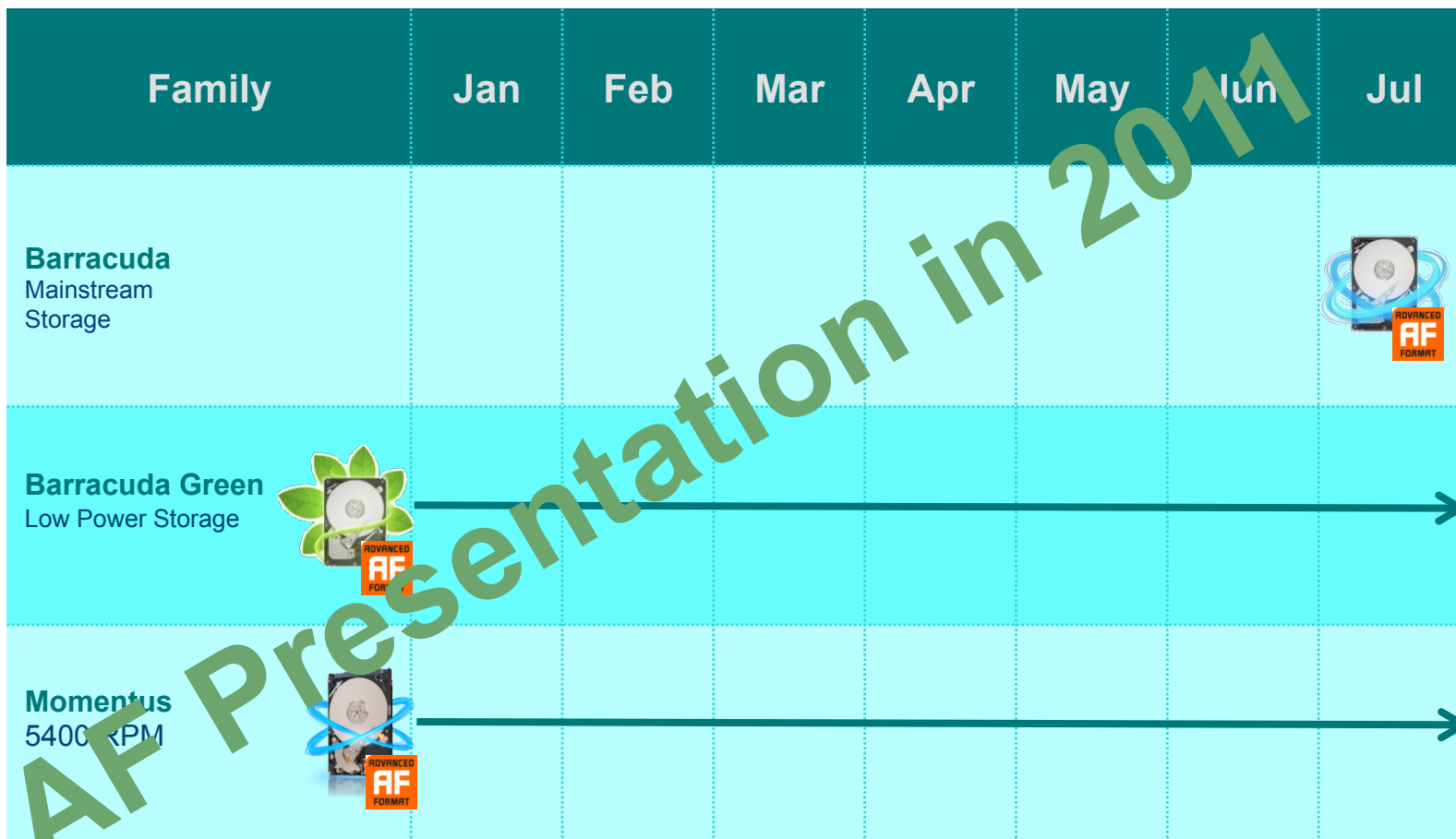
Western Digital AF Roadmap

- *Western Digital began shipping AF 512e HDDs 1H CY 2010 in the Caviar Green 3.5" HDD Lines*
- *AF 512e HDDs are phasing in across all WD desktop & mobile lines*
  - *Older platforms (legacy) may not transition*
- *Western Digital is supporting development of both 512e and 4K Native HDDs as demands increase or change*

- *First Green HDDs with AF in Market*
- *First AF Shipments to both OEM and Channel*
- *Championed Service Packs for OS Support*
- *Largest AF Manufacturing Capacity*

Seagate AF Roadmap

Family	Jan	Feb	Mar	Apr	May	Jun	Jul
Barracuda Mainstream Storage							
Barracuda Green Low Power Storage							
Momentus 5400 RPM							

Enterprise AF

Why 4K media?

ADVANCED AF FORMAT

SNR Trends for 4K vs 0.5K sector

- Gen 2 ECC Less
- Gen 1 ECC Less
- 1st 4K drive
- 2lvl Iterative channel

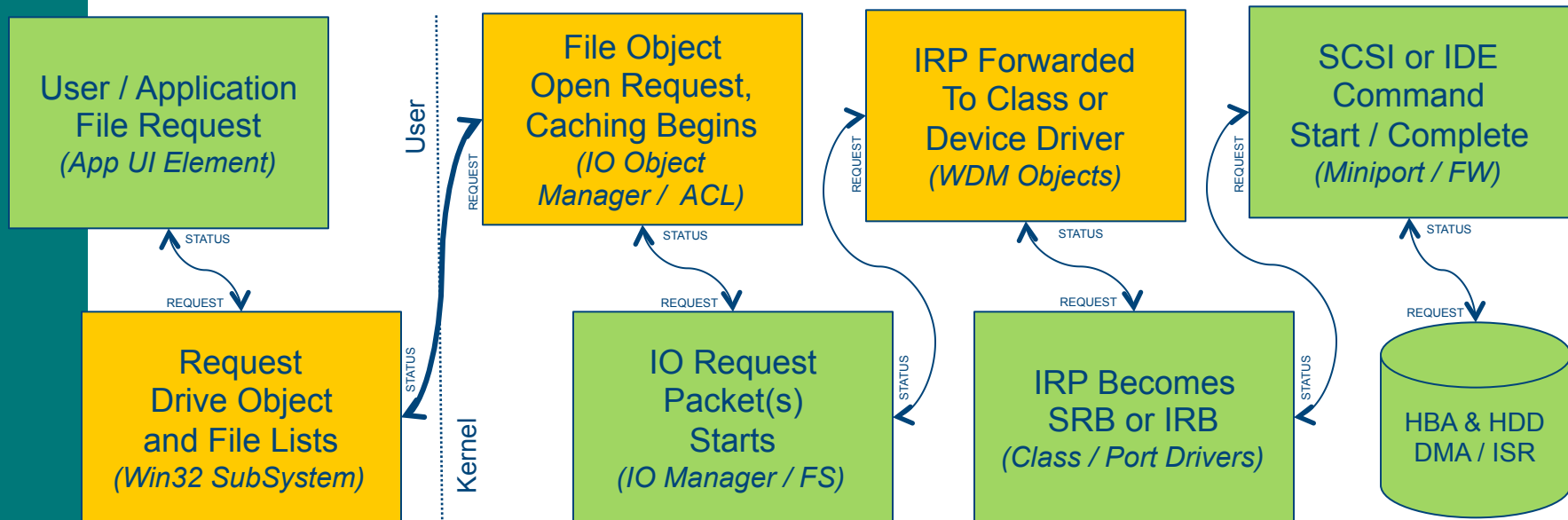Y-axis: Read Channel SNR (db)
X-axis: 2.5" GB/platter

Legend: 4K, 0.5K
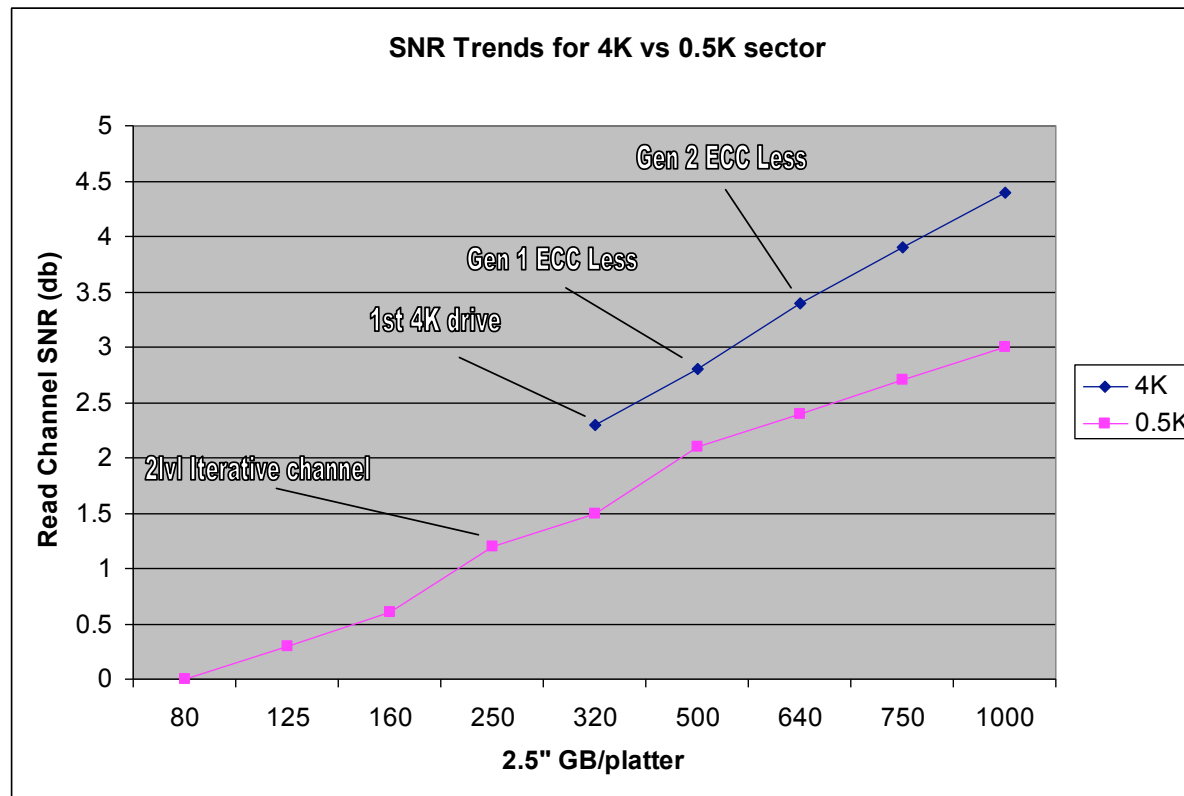
■ **4K sector SNR advantages**
   ■ Format efficiency gains due to removal of redundancy with a larger sector
      ■ PLO / gap reduction / overhead reduction
   ■ 4K SNR slope trend is much steeper than 0.5K slope
      ■ Limited coding tricks and signal processing gains available for 0.5K
      ■ Larger sector size enables longer codes and signal processing techniques

Windows & LBA Attributes

- IOCTL_STORAGE_QUERY_PROPERTY -> STORAGE_ACCESS_ALIGNMENT_DESCRIPTOR
  - This API allows Applications or Filters to determine LBA size
  - Exposure is very low in aligned systems

User / Application File Request *(App UI Element)*	File Object Open Request, Caching Begins *(IO Object Manager / ACL)*	IRP Forwarded To Class or Device Driver *(WDM Objects)*	SCSI or IDE Command Start / Complete *(Miniport / FW)*

User — Kernel

Request Drive Object and File Lists *(Win32 SubSystem)*

IO Request Packet(s) Starts *(IO Manager / FS)*

IRP Becomes SRB or IRB *(Class / Port Drivers)*

HBA & HDD DMA / ISR

STATUS / REQUEST

Awareness of Alignment

- SQL: http://msdn.microsoft.com/en-us/library/dd758814(v=sql.100).aspx
- VMWare: http://www.vmware.com/pdf/esx3_partition_align.pdf
- NTFS: http://ubuntuforums.org/archive/index.php/t-1565108.html
  - http://theether.net/kb/100104
- Microsoft: http://support.microsoft.com/kb/2385637
  - http://support.microsoft.com/kb/929491
- Other:
  - Wiki: http://en.wikipedia.org/wiki/XFS#Variable_block_sizes
  - Fat32: http://forum.easeus.com/viewtopic.php?p=24217&sid=18149153ac5fd04a17f576d0e30cc1ac
  - Oracle:
    - Unified Storage: http://blogs.oracle.com/dlutz/entry/partition_alignment_guidelines_for_unified
    - Solaris: http://osdude.wordpress.com/2010/09/10/aligning-solaris-x86-partitions-slices/
    - Solaris: http://www.oracle.com/technetwork/articles/systems-hardware-architecture/lun-alignment-163801.pdf
  - EMC Clariion: http://www.penguinpunk.net/blog/?p=499
  - Dell SQL: http://en.community.dell.com/support-forums/storage/f/1216/p/18697184/18820170.aspx
  - IBM Linux: http://www.ibm.com/developerworks/linux/library/l-4kb-sector-disks/

Futures – SMR/LBA

SMR Architecture

- **Read Modified Write**

  - Read-modify-write operation first reads a portion of data from the disk, then modifies part of that portion with the host provided write data, and finally writes the whole portion back to disk

- **Shingled Regions**

  - A group of tracks that is separated from neighboring shingled regions by a guard band. The purpose of the guard band is to prevent a write in a given region to interfere with data written on other regions. That isolation of interference guarantees that no read-modify-write will need to go beyond a region boundary

- **Indirection**

  - An indirection system is a collection of data structures and algorithms that assigns physical locations to logical block addresses and retrieves physical locations of logical block addresses. In the case of SMR, the indirection system is to be designed to provide good read/write performance for a wide variety of natural workloads

- **Hints**

  - Provide the device with information about the usage of data in the user area of the media. The purpose of hints is to allow the device to manage its data in a more intelligent way. Hints allow the device to increase reliability, performance and data integrity. Hints originate from the file-system

Shingled Regions

> **Implementing shingled regions is a good option to manage the random write performance impact of SMR**

Cylinder Number

ID   OD

Simplistic View
(e.g. if all tracks are shingled)

- Large sequential writes should have reasonable performance
- Small random writes may have very poor performance

Rotational Position
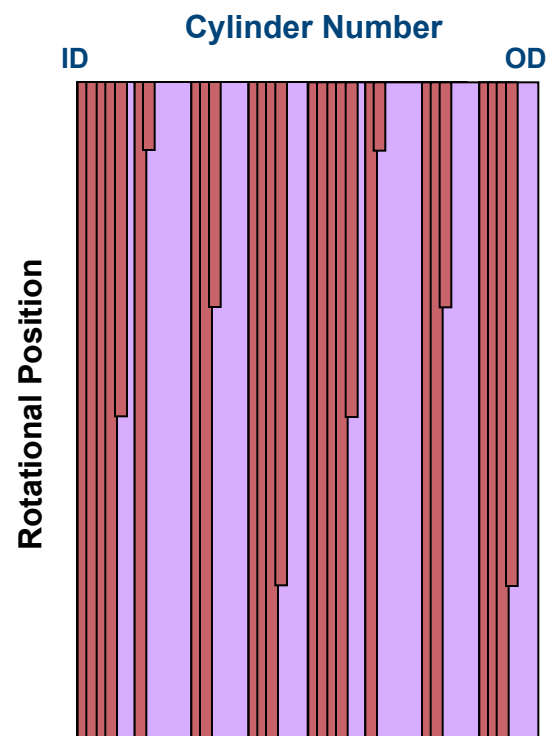
- HDD surface divided into regions of tracks
- Shingled writing within each region

'Write Staging' and Indirection

- Implementing a persistent 'Write Staging' solution will further improve the random write performance of SMR by delivering a persistent write caching effect

- Managing the 'Write Staging' likely involves background processes – data movement within the drive not in direct response to a host write

- A logical-to-physical indirection system will need to keep track of the data locations

One Possible Architecture

'Write Staging'
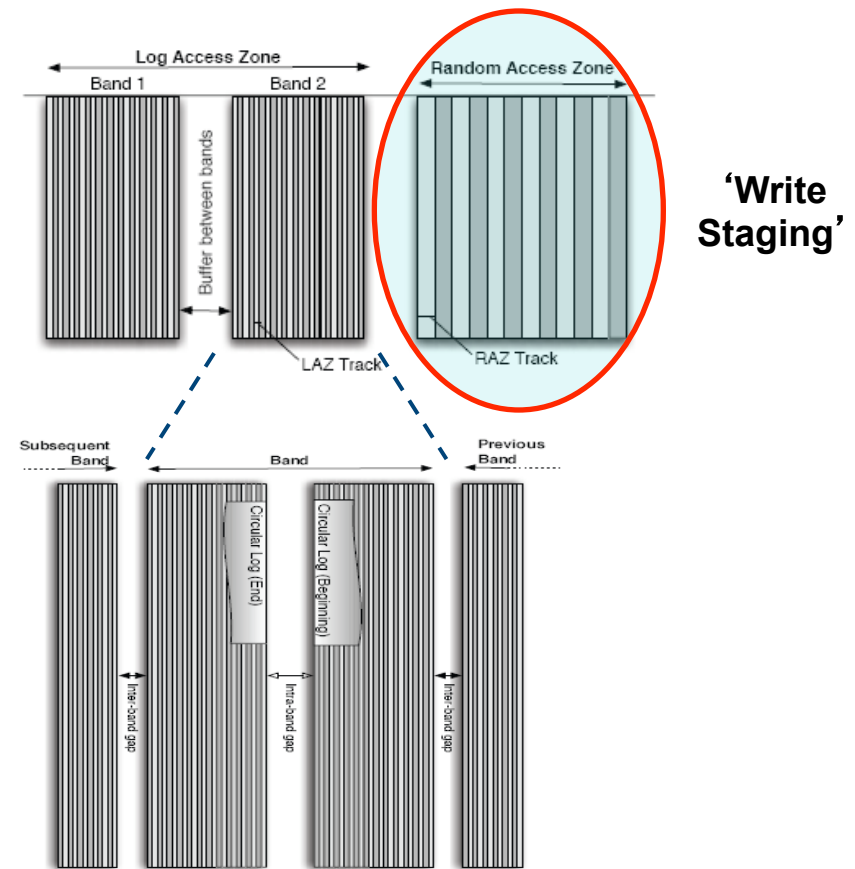
A. Amer et al. "Design Issues for a Shingled Write Disk System" MSST 2010

LBA Indirection

ADVANCED AF FORMAT

1 Host wants to rewrite **LBA-13** and **LBA-14**

2 Host seeks for next available unassigned block

3 Host assigns **LBA-13** and **LBA-14** to **PBA-32** and **PBA-33**

PBA-13 and **PBA-14** Now unassigned

PBA-11 LBA-11	PBA-12 LBA-12	PBA-13 LBA-N/A	PBA-14 LBA-N/A
PBA-21 LBA-21	PBA-22 LBA-22	PBA-23 LBA-23	PBA-24 LBA-24
PBA-31 LBA-31	PBA-32 LBA-13	PBA-33 LBA-14	PBA-34 LBA-N/A
PBA-41 LBA-N/A	PBA-42 LBA-N/A	PBA-43 LBA-N/A	PBA-44 LBA-N/A

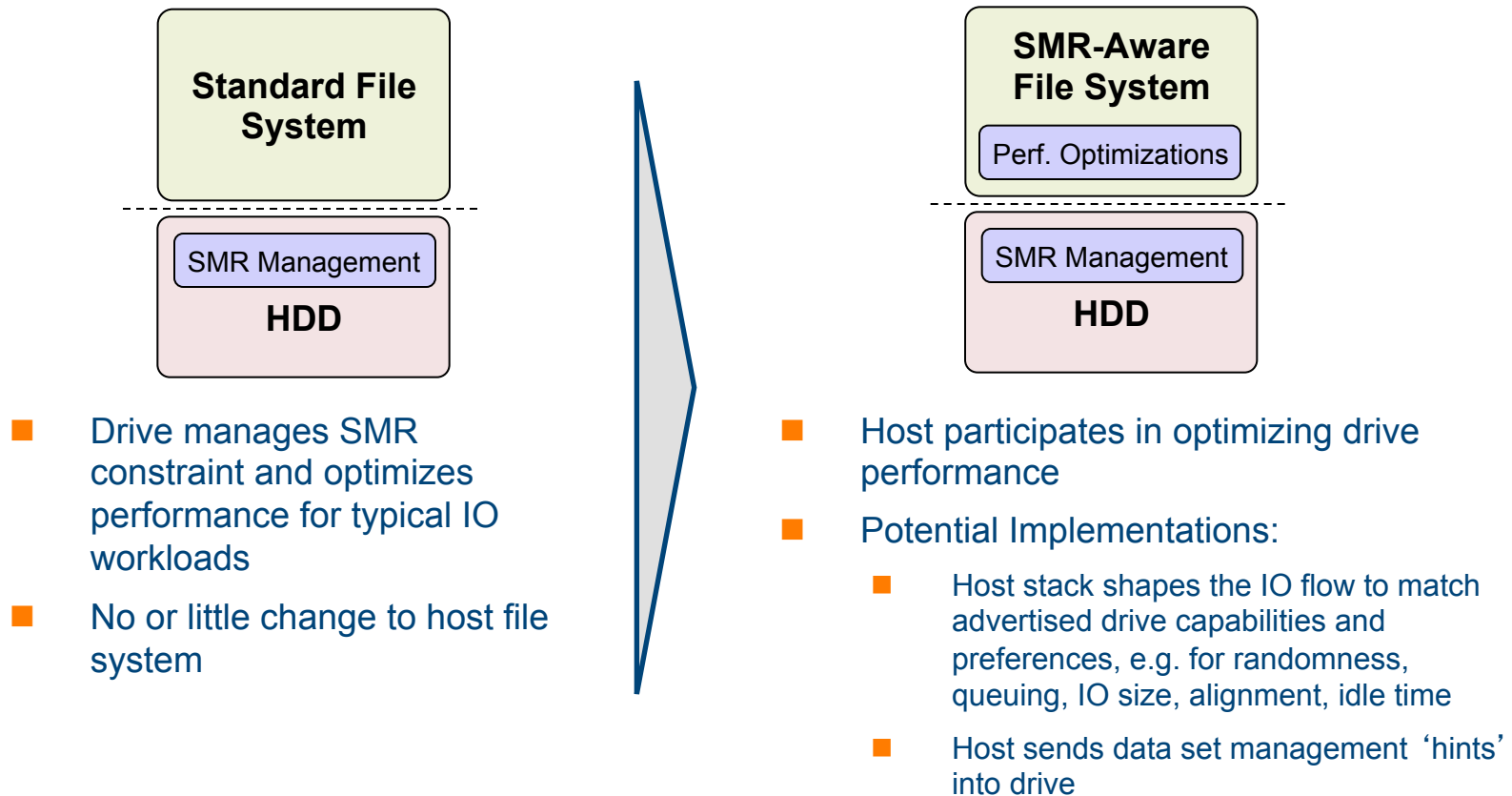Unassigned

TRIM/UNMAP for SMR Drives

> **Similar to a SSD, a SMR drive's performance benefits from actively managing the instantiation of its logical blocks**

Current Assumptions:

- SMR drives support the TRIM/UNMAP commands for basic dataset management
- A never written / freshly trimmed LB does not logically exist on the media
- Reads to a never written or freshly trimmed LBA return may return the original data, random data, or a sector filled with zeroes
- The drive reports its capabilities for how trimmed sectors are addressed
- The drive manages media health autonomously

Evolution of SMR

An SMR-aware host stack may be able to further optimize the system IO performance

Standard File System

SMR Management

HDD

SMR-Aware File System

Perf. Optimizations

SMR Management

HDD

- Drive manages SMR constraint and optimizes performance for typical IO workloads

- No or little change to host file system

- Host participates in optimizing drive performance

- Potential Implementations:
  - Host stack shapes the IO flow to match advertised drive capabilities and preferences, e.g. for randomness, queuing, IO size, alignment, idle time
  - Host sends data set management 'hints' into drive

IDEMA will continue to work on industry alignment for SMR

IDEMA SMR WG Charter

- Define industry standard Nomenclature & terminology
- Development strategies for interface commands to enable shingled writing technologies.
  - Bring requirements into standards and other organizations
  - Create material for host side implementers including OS and applications
- Develop marketing materials to describe the above
  - White papers, presentations

SMR Drive Options

- Standard HDD SMR: Storage device is built using SMR technology, but the architecture is designed to accommodate all IO requests that a current HDD can handle. The device will hide any internal operations (such as read-modify-write) needed to preserve data integrity.

  - Needed at first introduction – may cause performance issues

- New or Enhanced Interface: The storage device will use an interface protocol that will differ from the current generation HDD and SSD interfaces. It could differ by simply restricting the current protocols, adding extra functionality, adopting other protocols, etc. Here are some examples that would fall in this category:

  - Large Block SMR: The storage device restricts IO operations to be of a given, potentially fixed, large block size (e.g., 32MB).

  - HDD Emulation with Hints: Here the storage device maintains the current flexibility in IO request sizes, but adds extra functionality to allow a host system to provide the storage device with extra information pertaining to the data.

  - OSD: The storage device uses SMR technology and implements an OSD (Object Storage Device) interface.

OS Hinting for LBA Indirection
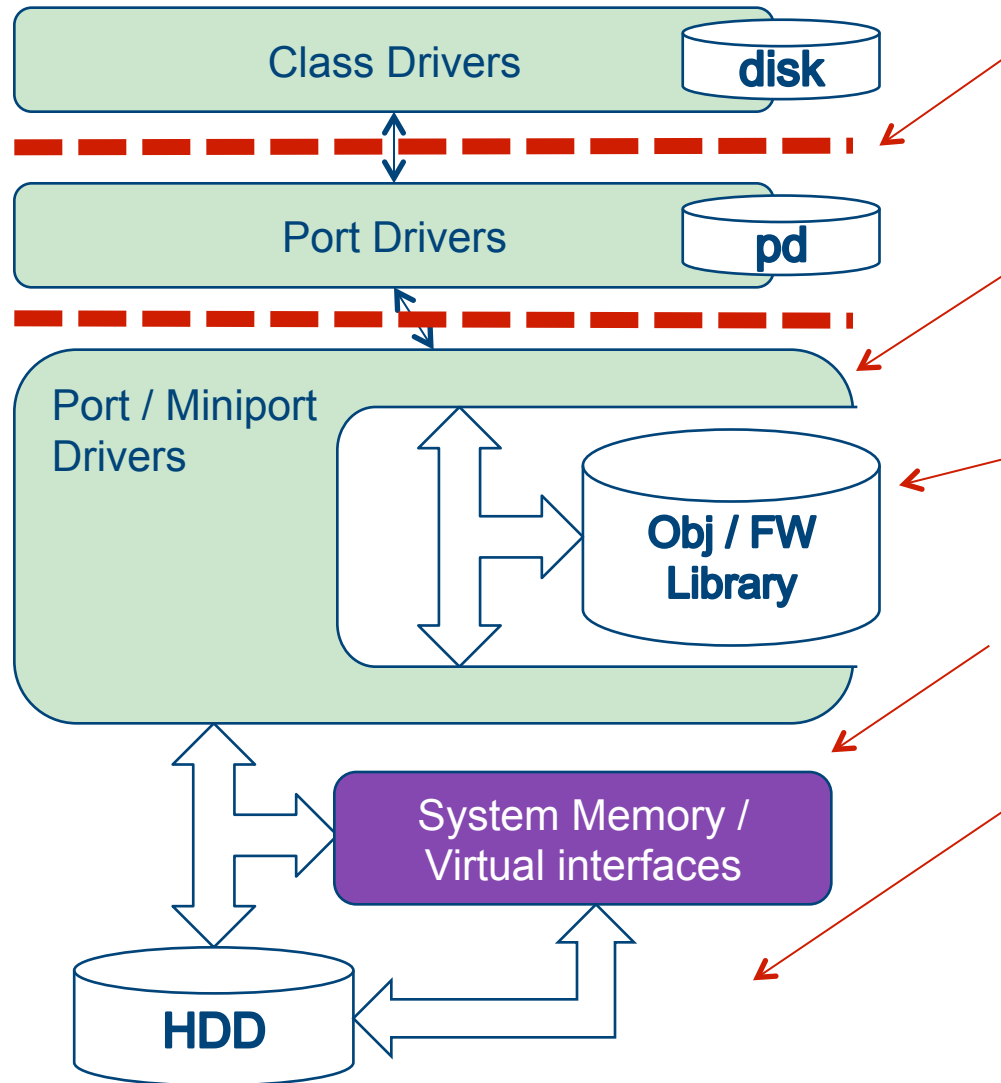
- **Filter Interfaces**
  - Defined by 3rd Party or OS Vendor
- **Port / Miniport Driver**
  - Are definitions needed?
- **Object Interface(s)**
  - Defined by IDEMA or Standards
- **Virtual Interfaces**
  - Defined by 3rd Party
- **HDD Interfaces**
  - Defined by IDEMA or Standards

Hybrids / SSDs

Ecosystem and Storage

- **Everything new and exciting trending towards**
  - Super thin, ultra-portable
  - Instant on, phone or TV like performance
  - Little to no power use at idle
- **Most recent Hard Disk Drive trends still pushing capacity**
  - Driven by existing market demands, explains AF
  - Desktop and Laptop markets still growing
  - Little to no tolerance for legacy to stop trend
  - Power management control expanding with TPM
- **AF HDD and SSD features converging in compute space**
  - New markets are important to both
  - Hybrid Hard Drives bridging the gap for capacity
    - Hybrid and SSD both use same NAND

AOAC, ARM, UltraBook

- **AOAC promoting both lower power budgets & instant on**
  - SSD and Hybrid well positioned to satisfy here
  - HDD also capable in this platform, with changes
- **Windows has flourished with HDDs, driven content**
  - ARM Port doesn't negate this success, advances it
  - Power requirements are changing, SATA can accommodate
- **Ultrabook will create demand back towards HDDs**
  - OEM and Customer demand cannot be satisfied by NAND
  - Cost and capacity in productions require Hybrids
- **Well known 7mm Z-axis designs raising interests**
  - Including reductions in power consumption
  - Including time to transfer data to near SSD
- **Hybrid will be AF, SMR or LBA Indirection**

Proposed WHQL Requirements

Questions

- System.Fundamentals.StorageAndBoot.IOLatency
  - Describe the need for the 300ms maximum latency in all storage
- System.Fundamentals.StorageAndBoot.SATABootStorage
  - Specifies SATA and power consumption as blocker in AOAC market
  - SATA power model set by previous architecture, discounts improvements
- Device.Storage.Hd.Flush
  - Business model justification lacking for testing burden at IHV and OEM
- Device.Storage.Hd.PhysicalSectorSizeReportsAccurately
  - No specified test criteria or metrics for success, why is this needed?
- System.Fundamentals.Firmware.UEFIPostTime
  - Specifically creates highest barrier for entry to SATA
  - SATA HDDs will have longest market tail given adoption curve
  - UEFI was unrequired in previous OS release

Feature Discussion

Feature Conversations

- Enabling telemetry for HDDs to reduce NDF for OEMs
  - Most important OEM topic for all IDEMA members
- Collapsing port driver count to establish higher quality
- Applying the requirements of booting to USB3
  - Most IDEMA members have USB HDD solutions
  - Is USB3 viable as internal boot technology?
- Will Hibernate as we know it go away?
- Adopting a model of flexible thin provisioning for all storage
- Will Windows backup, in some form, be revived again?
  - Does this map to existing HDDs, external storage?
- Is 4Kn throughout x64 architecture and UEFI a reality?
- Promoting self-encryption from novelty to reality