Shingled Magnetic Recording

Presented by

Ernest A. Louis

Western Digital Corporation

Acknowledgement

Acknowledging my colleagues:

William Cain

Sining Mao

Curtis Stevens

Shaoping Li

Daniel Bai

Chavalratana Puckdinukul

For their for their help and guidance with this presentation.

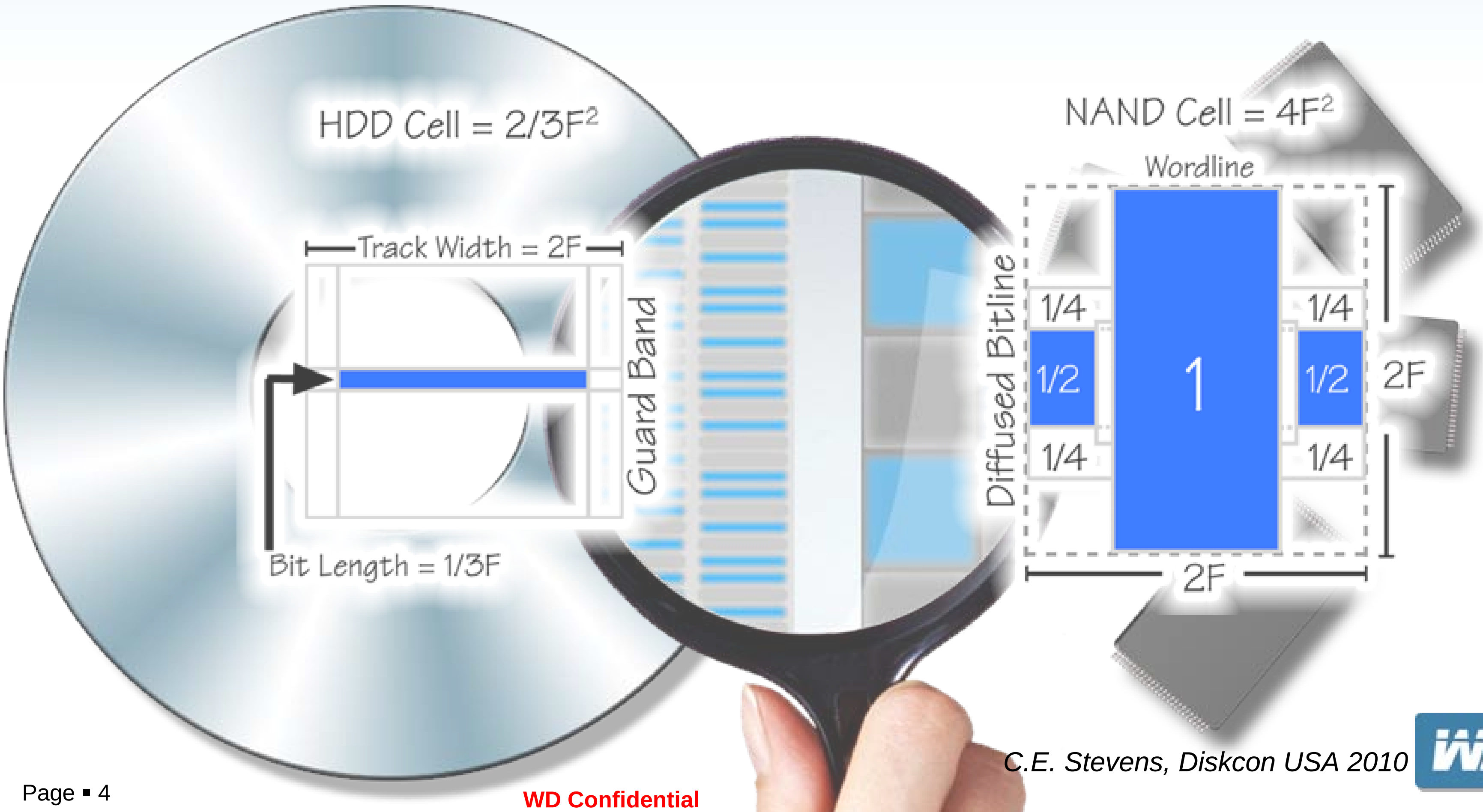
Contents

- **Introduction - Overview of SMR.**
- **Data Architectures and Performance.**
- **SMR Deployment and System Implications.**

Bit Cell Size: HDD vs. Flash (NAND)

HDD scales better than SSD as long as BAR > N

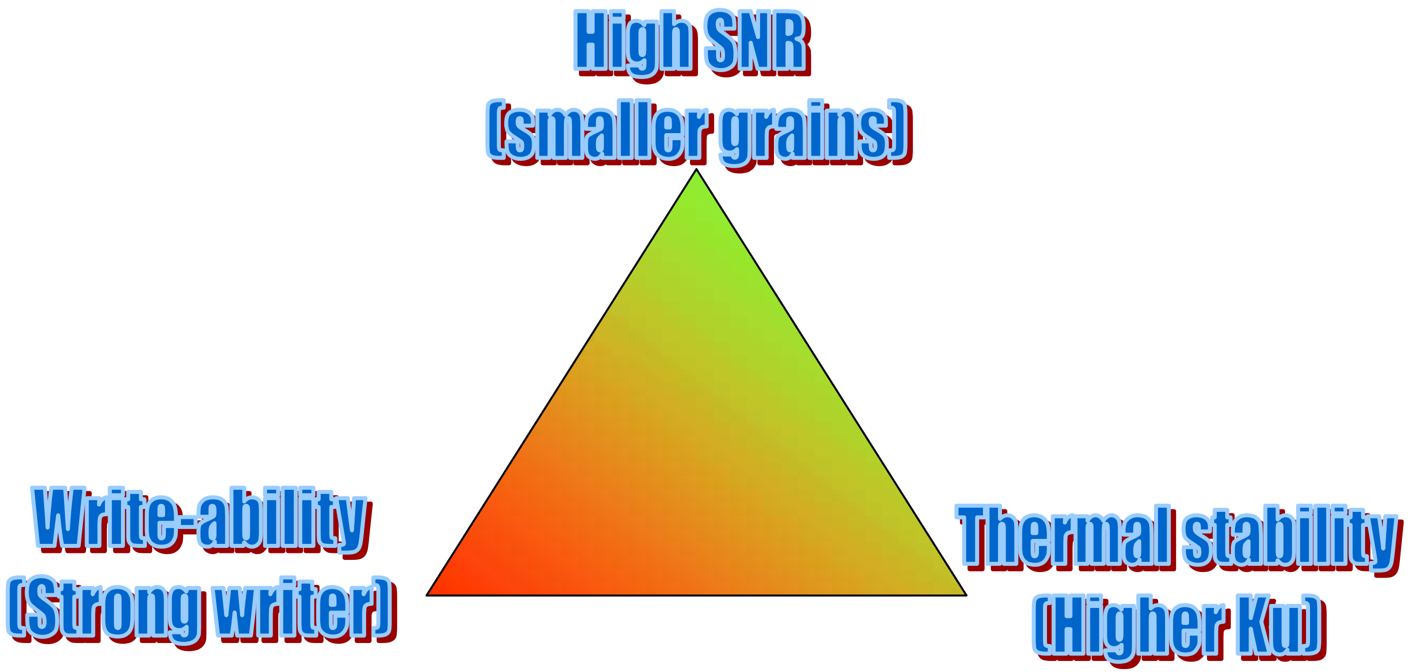
HDD cell = $4F^2/BAR$ SSD cell = $4F^2/N$



C.E. Stevens, Diskcon USA 2010

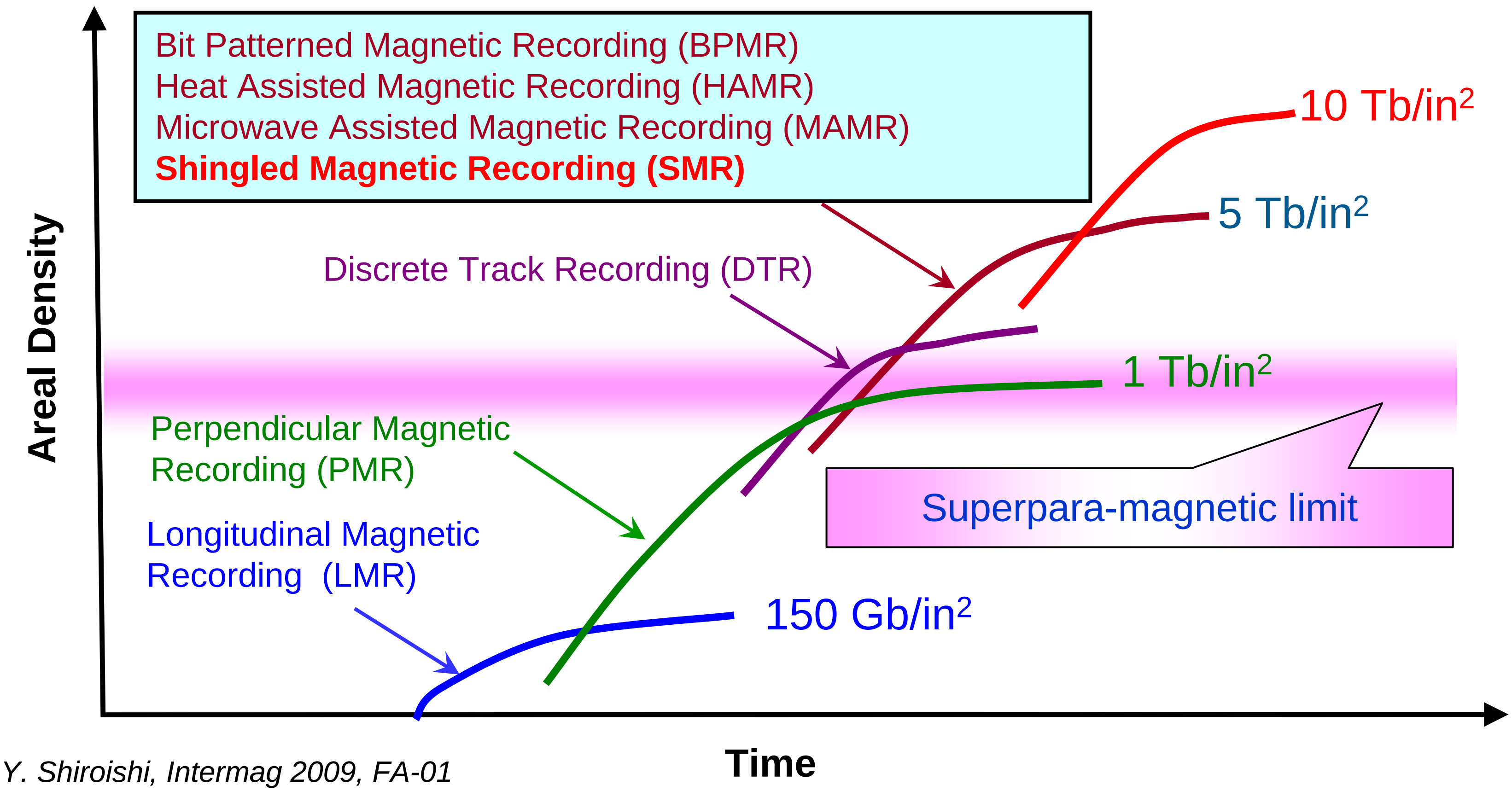
Media "Trilemma"

Superparamagnetic effect forces a trade-off for media design



Magnetic Recording System Technologies

New recording system technologies are needed to advance areal density

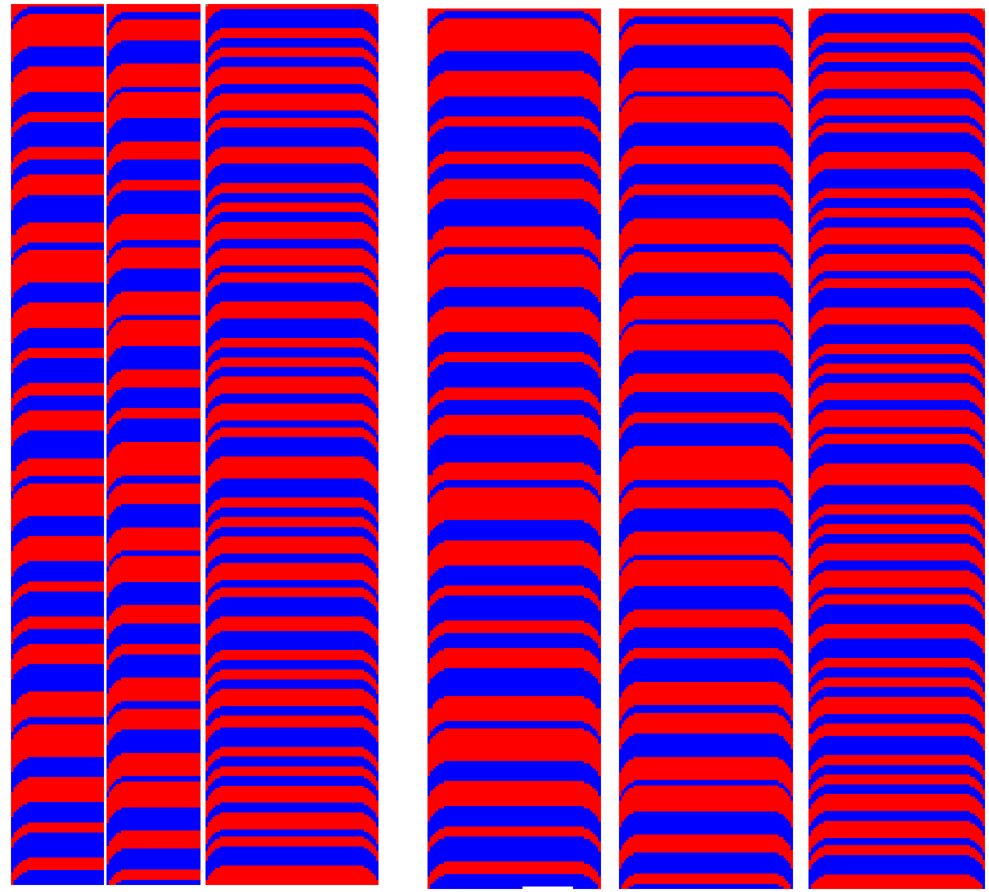


Y. Shiroishi, Intermag 2009, FA-01

What is Shingled Magnetic Recording

Shingled write

- Shingling overlaps tracks written sequentially by a wider writer.
- Residual tracks narrower than originally written widths.
- Shingled writing is not new - long used in WORM tape storage.



Shingled write

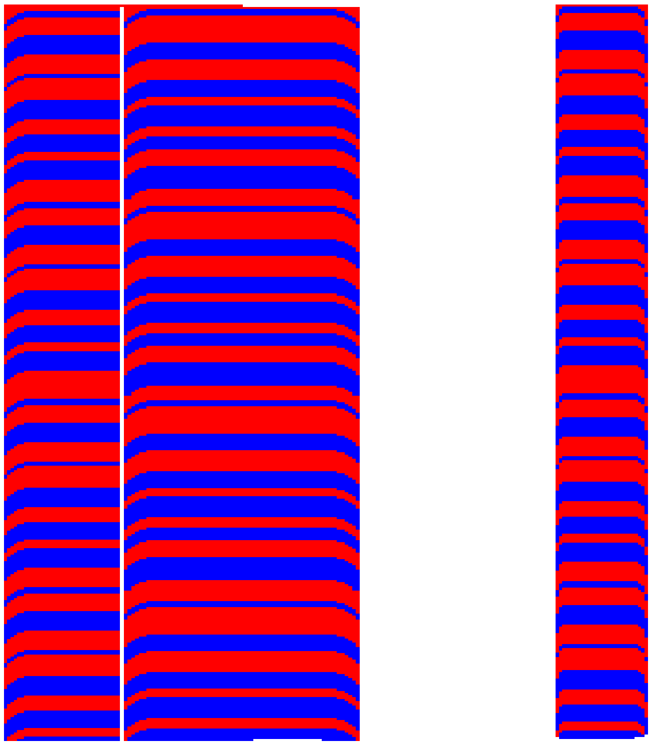
Conventional write

Advantages of Shingled Magnetic Recording

A wider than track-width writer pole addresses the scaling challenge

Advantages

- Stronger write field and gradient.
- ATE no longer a concern, can use stronger write fields.
- Can achieve 1.6X higher writer field with the same write head materials and processes.



Advantages of Shingled Magnetic Recording

Bit cell size

RW is now TPI driver and almost equal to smallest lithographic feature size
Bit cell size reduced by almost 2X - gap to NAND becomes more favorable.

Writer fabrication

Relaxes the exacting controls required for writer geometry – high yields easily attainable.

Extendability

Its a pre-cursor for TDMR - TDMR can attain even higher AD capability

Domain configurations and control

Wider writer pole provides more desirable domain configurations and reduces DLU induced erasure

No Adjacent track erasure

ATE is no longer a concern – we no longer trade off write-ability for ATE.

Implications for HDD technology

Media design

Higher media Ku - bit cell volume can be reduced.
Smaller grain size (20%) - results in higher SNR and narrower erase bands by 50%.

Head Design

Writer design for corner writing with higher write fields – asymmetric writer for high side-field gradients
Reader driven AD capability - reader response and side reading is more important now.

Areal density

Higher areal density with current component technology.
Simulations show 2 Tbits/sq.in. or more possible with continuous media

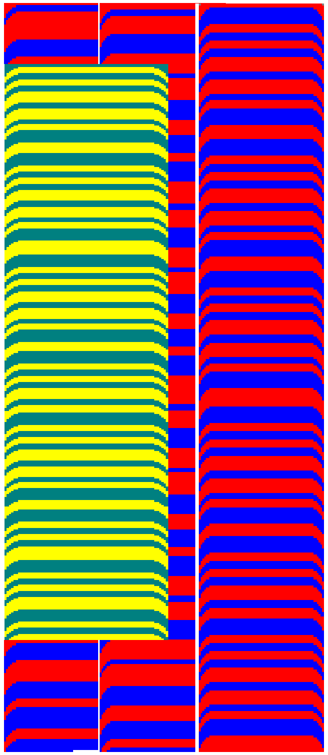
Mechanical

Tighter TMR requirement especially NRRO.
Micro-actuation systems may be necessary.

Challenges to Shingled Magnetic Recording

Random write access restrictions

- No update in place possible or direct write to random physical blocks.
- Challenge is to allow unrestricted write access (from the host's perspective).
- Common solution/approach is an embedded software layer - an indirection system - to mask and manage write access restrictions.

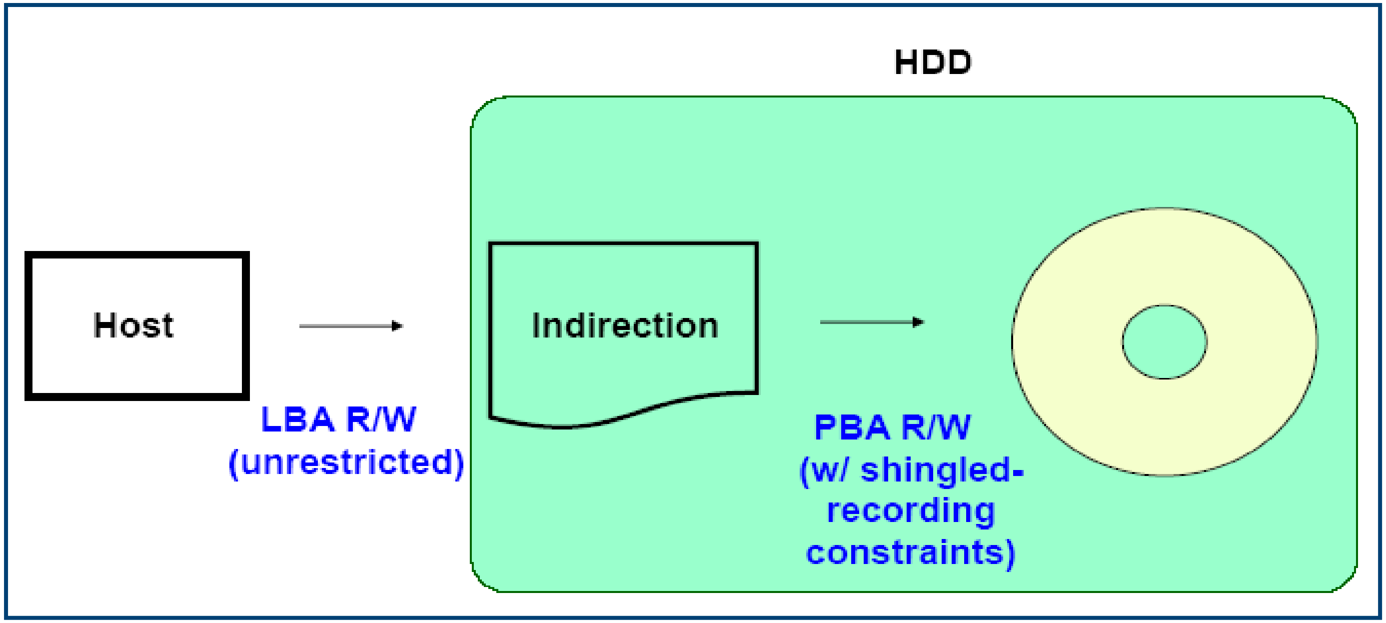


Contents

- **Introduction - Overview of SMR.**
- **Data Architectures and Performance.**
- **SMR Deployment and System Implications.**

Indirection Systems

Indirection System	A collection of data structures and algorithms that assigns physical locations to logical block addresses (LBA) during write and retrieves physical locations of LBAs during read.

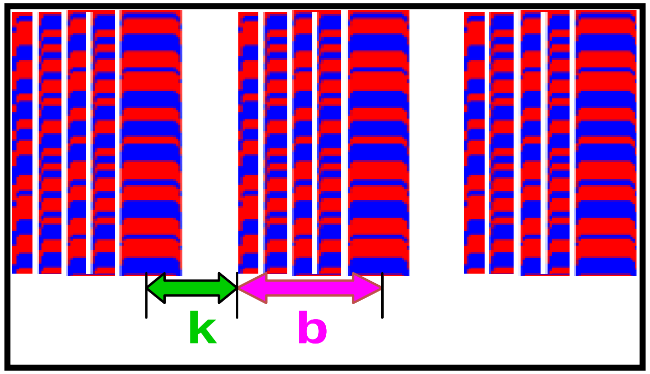


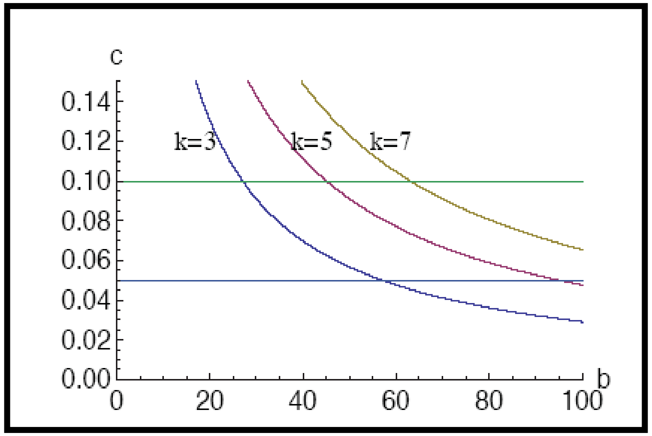
Y. Cassuto et al , MSST 2010, IEEE

Shingled layout and RMW operations

Managing random writes with RMW

- Partition disk surface into independent shingled regions with guard bands to isolate interference.
- Use read modify write (RMW) as a simple first level approach - RMW one or more regions for each write request or update.
- Trade off performance Vs capacity.
  Large shingled regions result in poor performance for small host writes.
  Small shingled regions have good performance but reduce capacity.



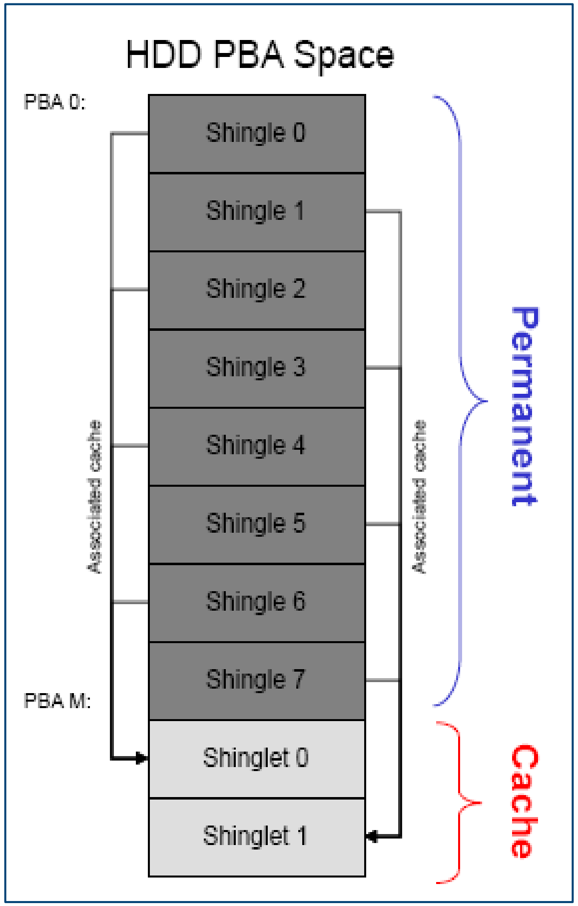


A. Amer et al , MSST 2010, IEEE

Set-associative disk-cache architecture

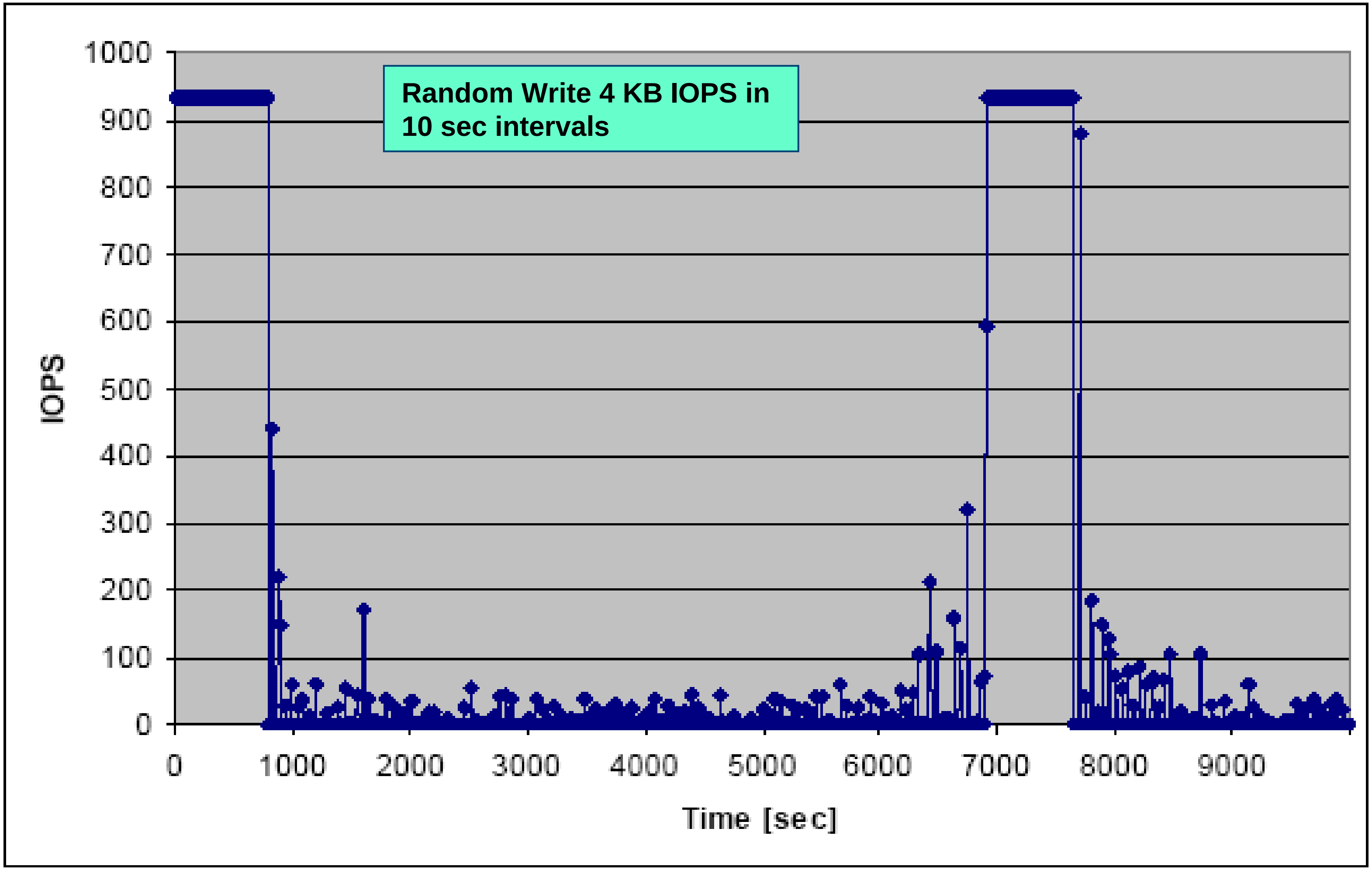
Disk-cache based architecture

- Disk partitioned to LBAs and cached regions (all shingled).
- Each LBA is associated with one cache region – multiple LBAs served by same cache regions.
- Incoming writes assigned to consecutive physical blocks in the cache region starting from the first.
- When the cache is full, blocks are re-written to native regions with RMW operations.



Y. Cassuto et al , MSST 2010, IEEE

Performance of disk-cache architectures

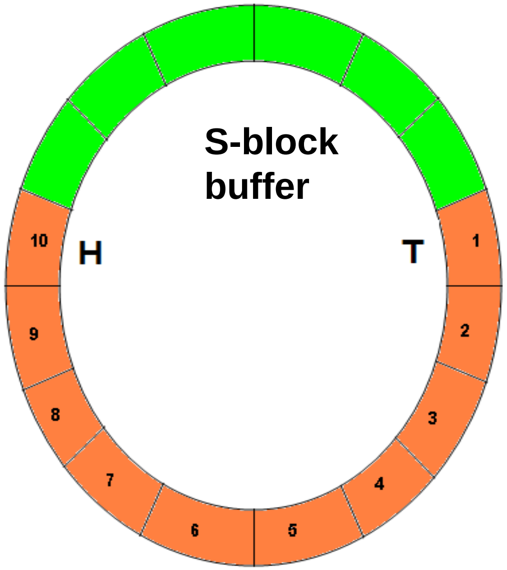


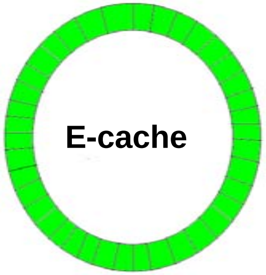
Y. Cassuto et al , MSST 2010, IEEE

Shingled Block architecture

S-block architecture

- Add an intermediate buffer in the storage layer of S–blocks.
- S-blocks store multiple blocks - consecutive LBAs on continuous PBAs.
- Two separate buffers now - E-cache and S-block
- Shingled regions managed as circular buffers with head/tail pointers and guard bands.



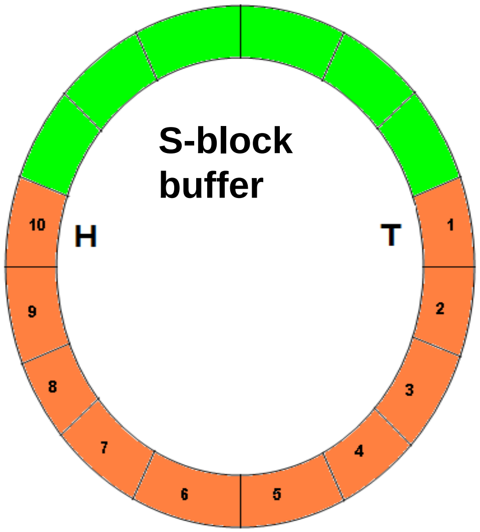


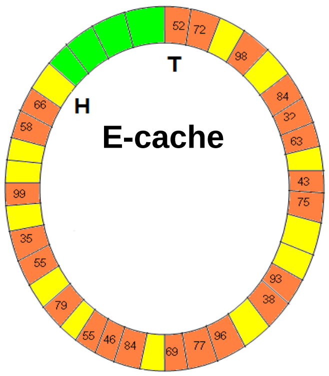
Y. Cassuto et al , MSST 2010, IEEE

Shingled Block architecture

S-block architecture

- Incoming writes always added to the head of the cache. Addition done at head and removal at tail.
- Data in cache transferred by RMW operation to S-block buffer with “de-stage” process
- “Garbage cleaning” is done to reclaim invalid blocks with “buffer defrag” process.

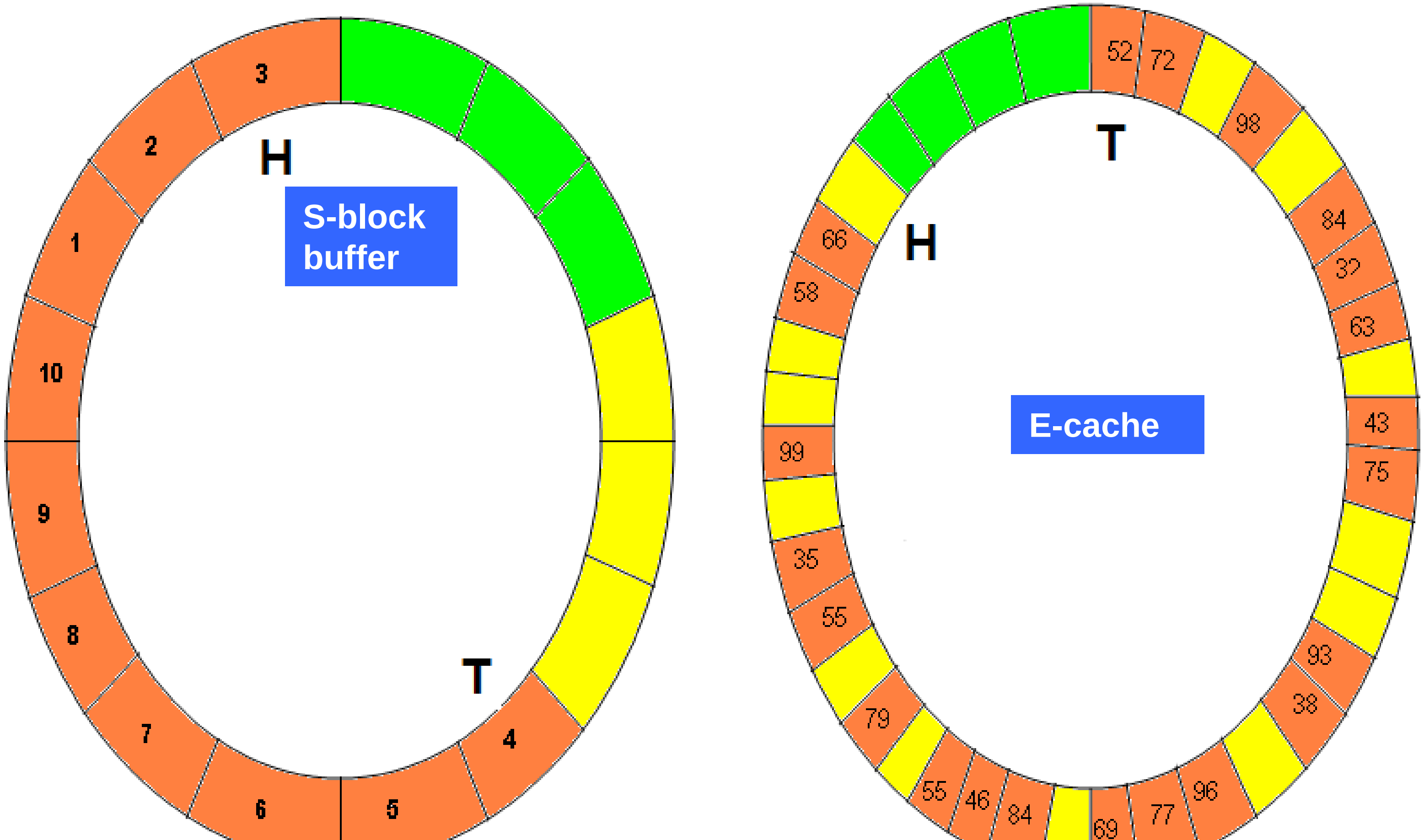




Orange – valid blocks in use
Yellow – used but invalid blocks
Green – empty available blocks

Y. Cassuto et al , MSST 2010, IEEE

Destage process with S-Block architecture

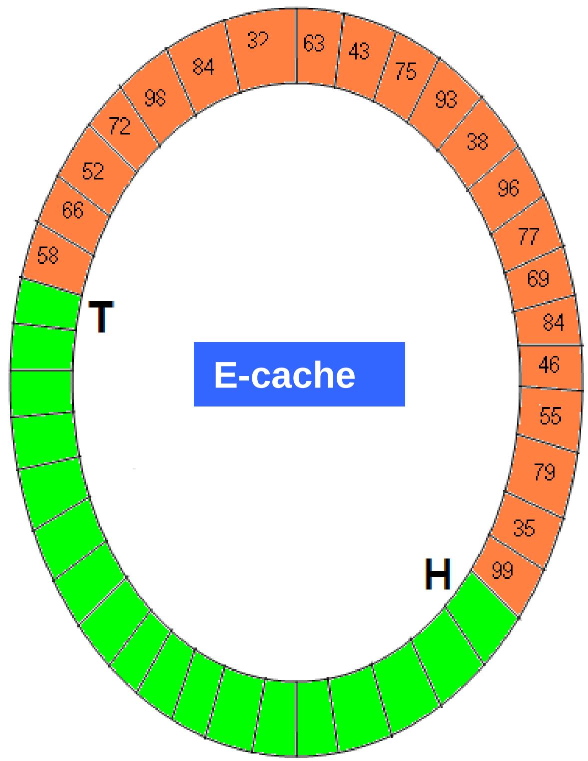


Y. Cassuto et al , MSST 2010, IEEE

De-stage Process	Data transferred from E-cache to S-block during destage - vacant blocks are shown in green, valid blocks in use in orange and yellow for used but invalid blocks.

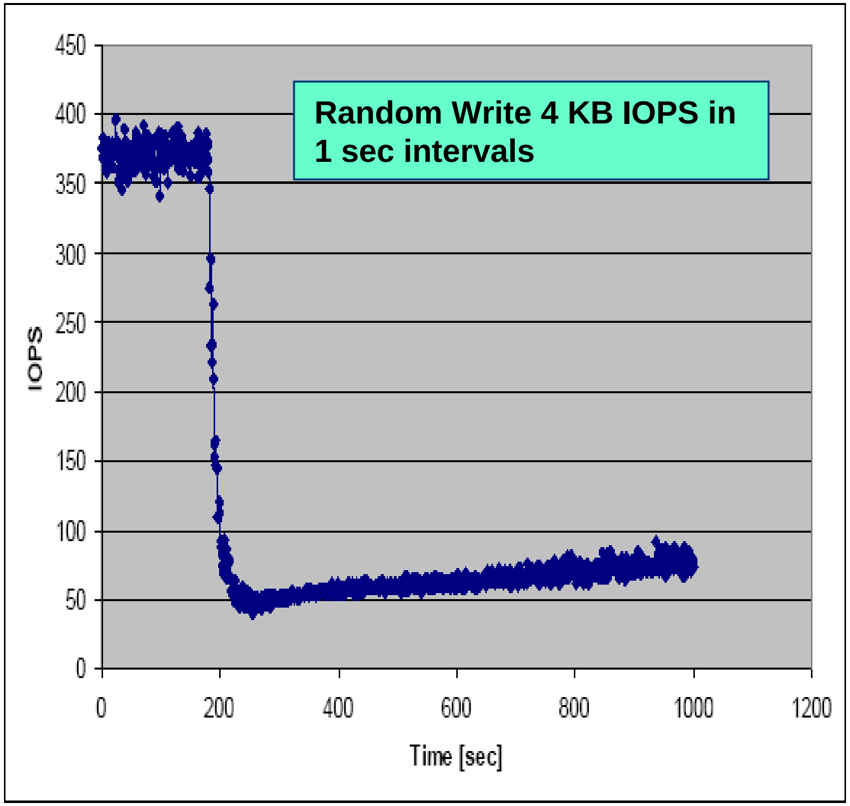
Defrag process with S-Block architecture

Defrag Process	Data moved from tail to head of buffer during defrag - green signifies vacant blocks, orange valid blocks in use and yellow for used but invalid blocks.

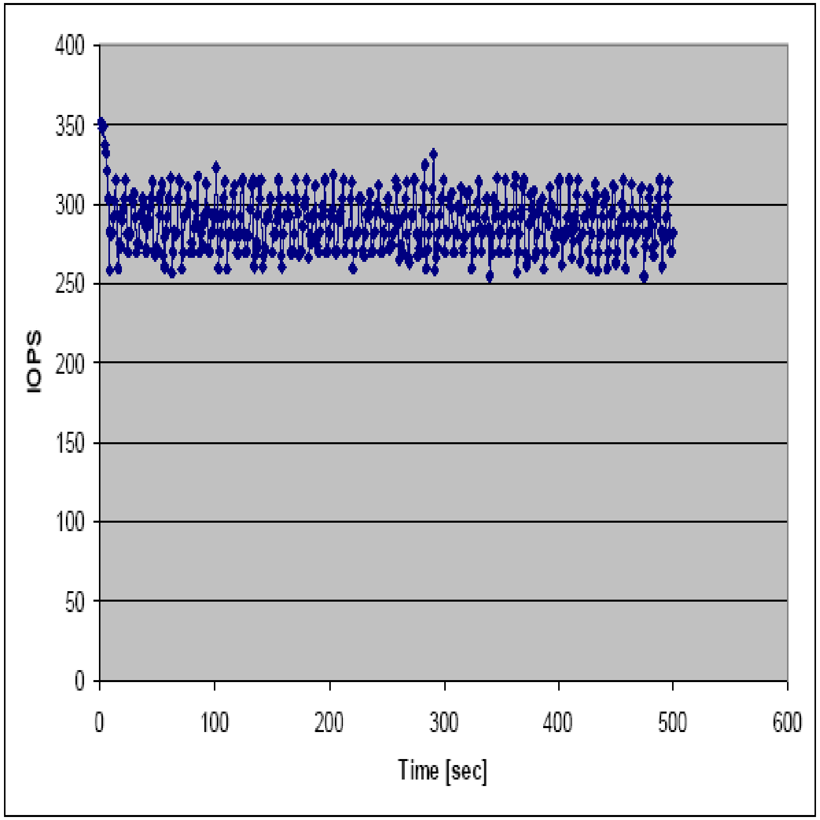


Y. Cassuto et al , MSST 2010, IEEE

Performance of S-block architecture



Lazy Garbage collection



Constant Garbage collection

Y. Cassuto et al , MSST 2010, IEEE

Viability of SMR

Impact of indirection system

- SMR challenges and opportunities are mostly at systems architecture level.
- An effective indirection scheme moves SMR from an impractical to a more manageable realm of performance.
- Indirection systems are not new but need to be adapted to the unique properties and constraints of SMR. Architectural design and heuristics will need further optimization according to host system type and workload.
- Goal is to present to the host a shingled drive that essentially has the same performance behavior as a non-shingled drive.

Contents

- Introduction - Overview of SMR.
- Data Architectures and performance.
- SMR deployment and System Implications.

SMR Implementation

IDEMA SMR Working Group Charter

▪ Define industry standard Nomenclature & terminology
▪ Development strategies for interface commands to enable shingled write.
    ▪ Bring requirements into standards and other organizations.
    ▪ Create material for host implementers including OS and applications
▪ Develop white papers, presentations to describe the above.

C.E. Stevens, Diskcon USA 2010

SMR Implementation

Host Interface Options

- **Standard HDD SMR:** architecture designed to accommodate all IO requests that a current HDD can handle. The device will mask internal operations (such as RMW). Needed at first introduction – may have performance gap.
- **New or Enhanced Interface:** different protocol Vs. current - examples:
    - **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:** maintains current flexibility in IO request sizes, but adds extra functionality enabling host system to provide extra information pertaining to the data.
    - **OSD:** The storage device implements an OSD (Object Storage Device) interface.

C.E. Stevens, Diskcon USA 2010