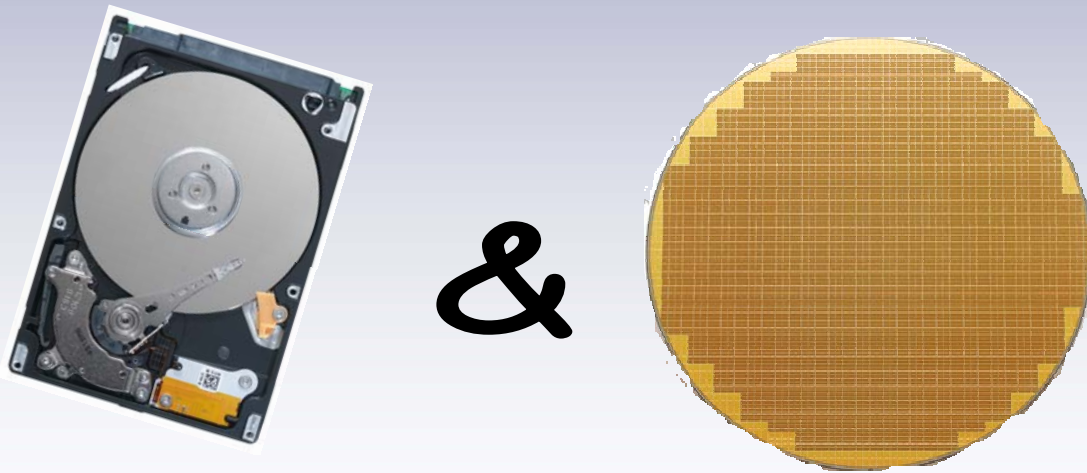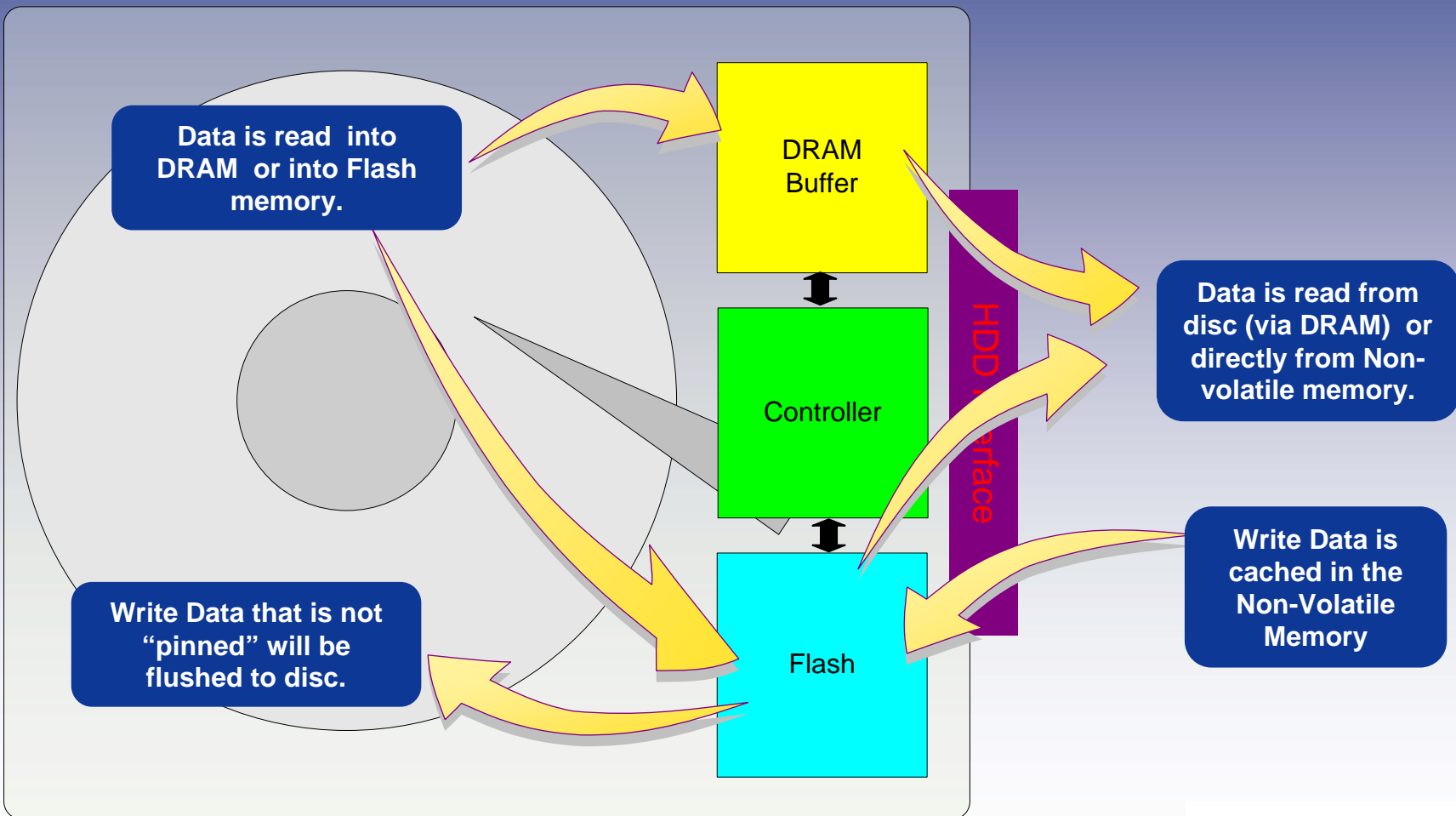# Hybrid HDD Nuts and Bolts

John E. Moon

Sr. Director, Systems Integration

Seagate Technology
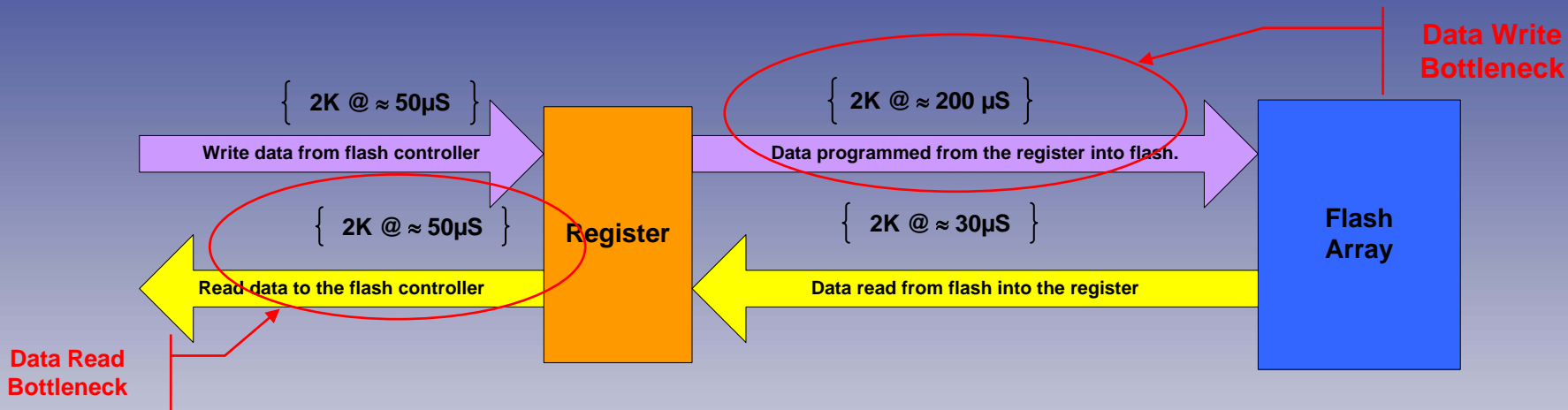
- Hybrid Drive Architecture
- Hybrid and Microsoft Vista
- The 'How's' on Hybrid Benefits

&

Hybrid Architecture - Hardware

Data is read into DRAM or into Flash memory.

DRAM Buffer

Controller

HDD Interface

Flash

Data is read from disc (via DRAM) or directly from Non-volatile memory.

Write Data is cached in the Non-Volatile Memory

Write Data that is not "pinned" will be flushed to disc.

Hybrid Architecture – Flash Interface

Data Write Bottleneck

{ 2K @ ≈ 50μS }
Write data from flash controller

{ 2K @ ≈ 200 μS }
Data programmed from the register into flash.

Register

Flash Array

{ 2K @ ≈ 50μS }
Read data to the flash controller

{ 2K @ ≈ 30μS }
Data read from flash into the register

Data Read Bottleneck

X 16 interface
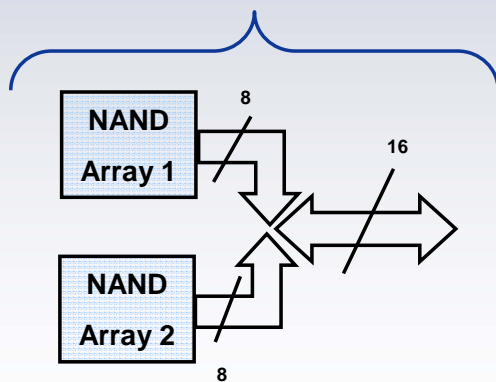
Parallel load and program.

NAND Array 1

8

16

NAND Array 2

8

Load Data Register | Program Flash Array 1

Load Data Register | Program Flash Array 2

Seagate

Hybrid Architecture – Firmware
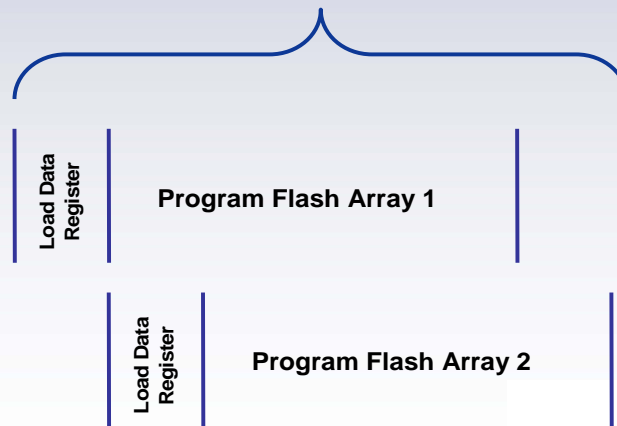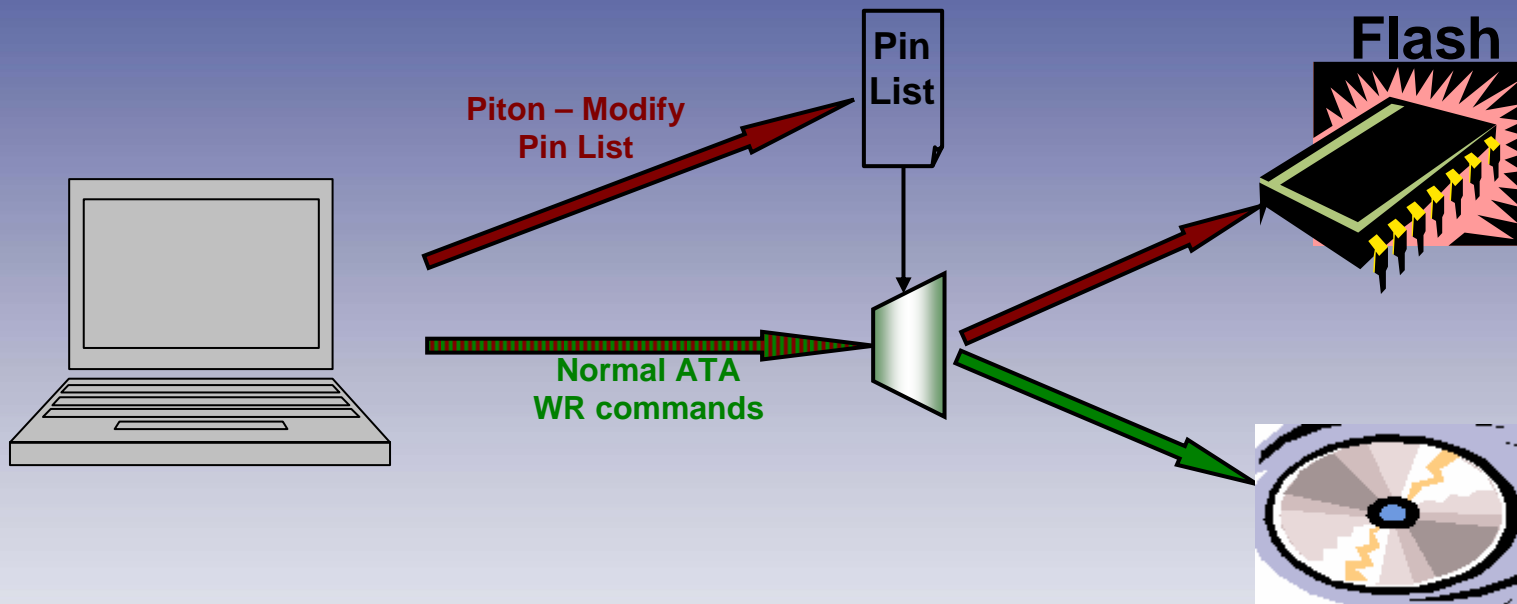
- Support new ATA – 8 commands.

- Bad Block Management.

- Wear – leveling.

- Garbage Collection.

Hybrid Commands - Modify Pin List
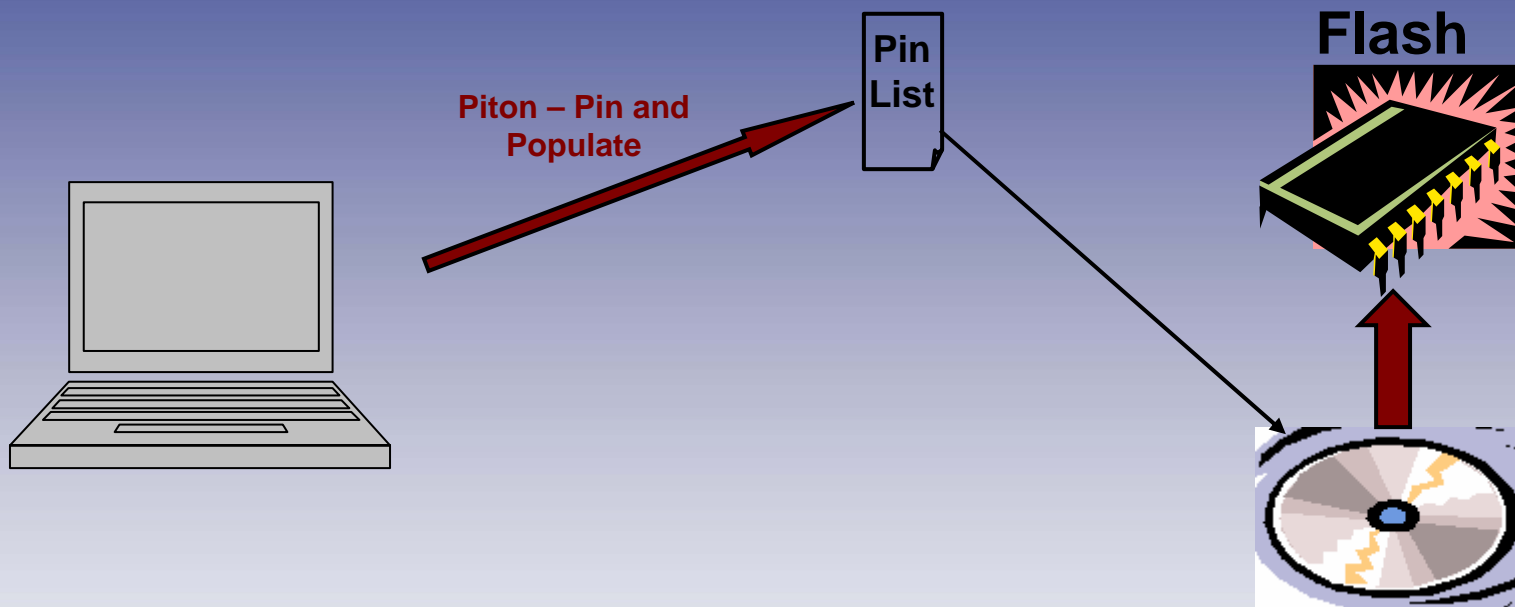

Flash

Pin List

Piton – Modify Pin List

Normal ATA WR commands

What: Used by the host to identify logical blocks for which host sent data should be stored in the Flash.

How Used: Host would Modify Pin List to cover addresses of data which when transferred from the host should be stored in the Flash. All subsequent writes to these addresses will be stored in Flash.

Hybrid Commands – Pin and Populate

Piton – Pin and Populate

Pin List

Flash

What: Used By Host to identify logical blocks to be added to the pin list AND populate pinned blocks from media.

How Used: Host would pin and populate addresses that are already written to the disc, but desired to be in Flash. Subsequent writes to these addresses will be stored in the Flash.
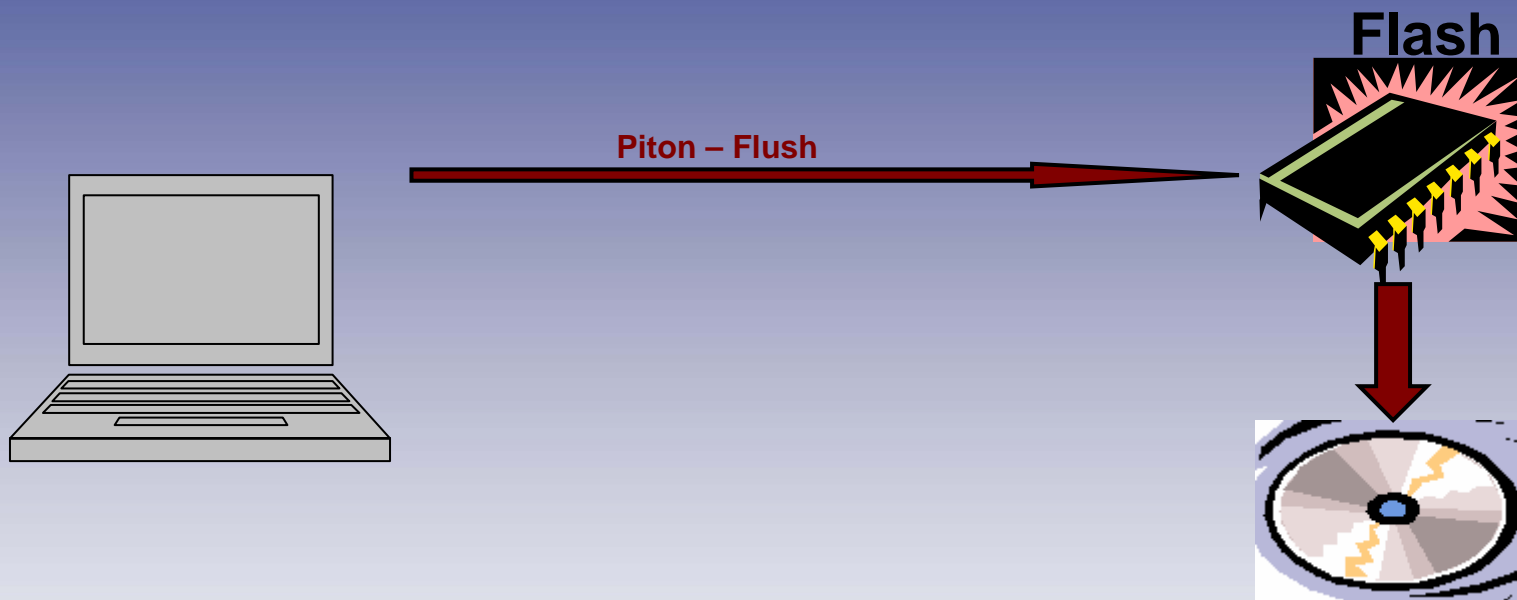
Hybrid Commands – NVC Power Mode

Flash

Piton – Flash Power Mode

Normal ATA WR commands
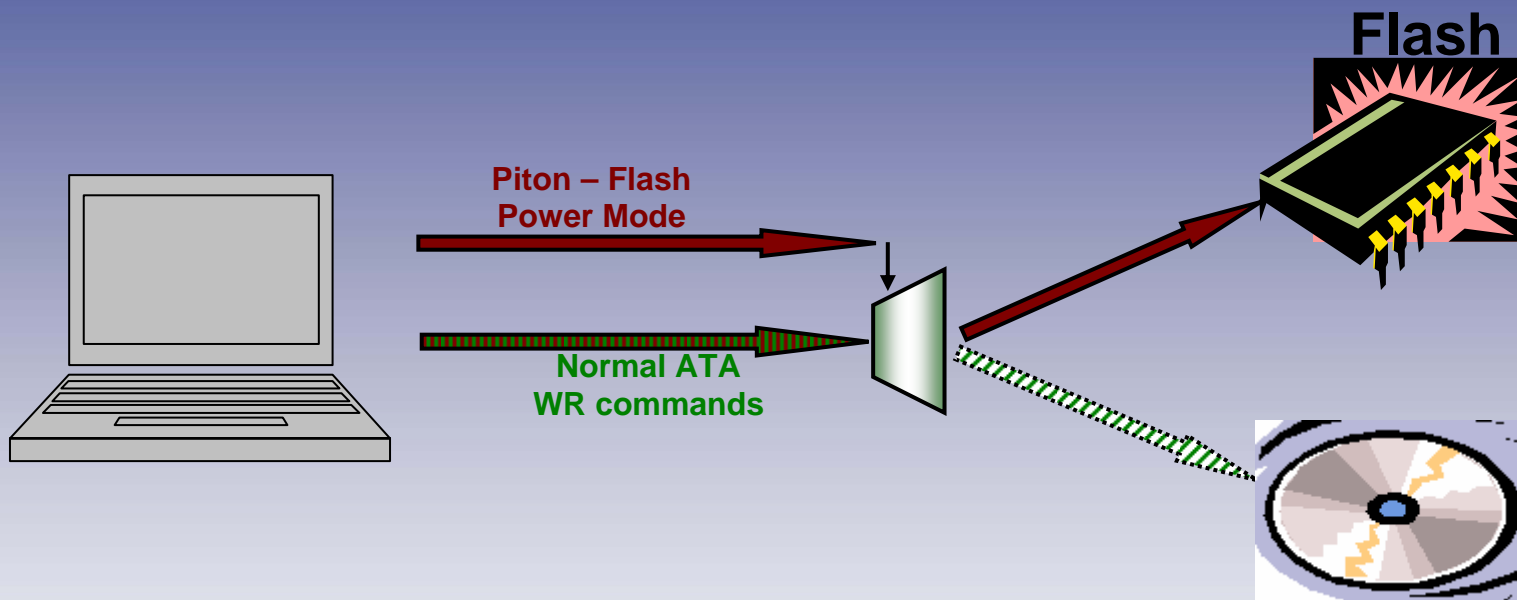
What: Used By Host to notify drive of low power environment.

How Used: Drive will spin down discs and attempt to put any host received write data to Flash. Drive will flush unpinned data as needed to manage FLASH space for low power operation.
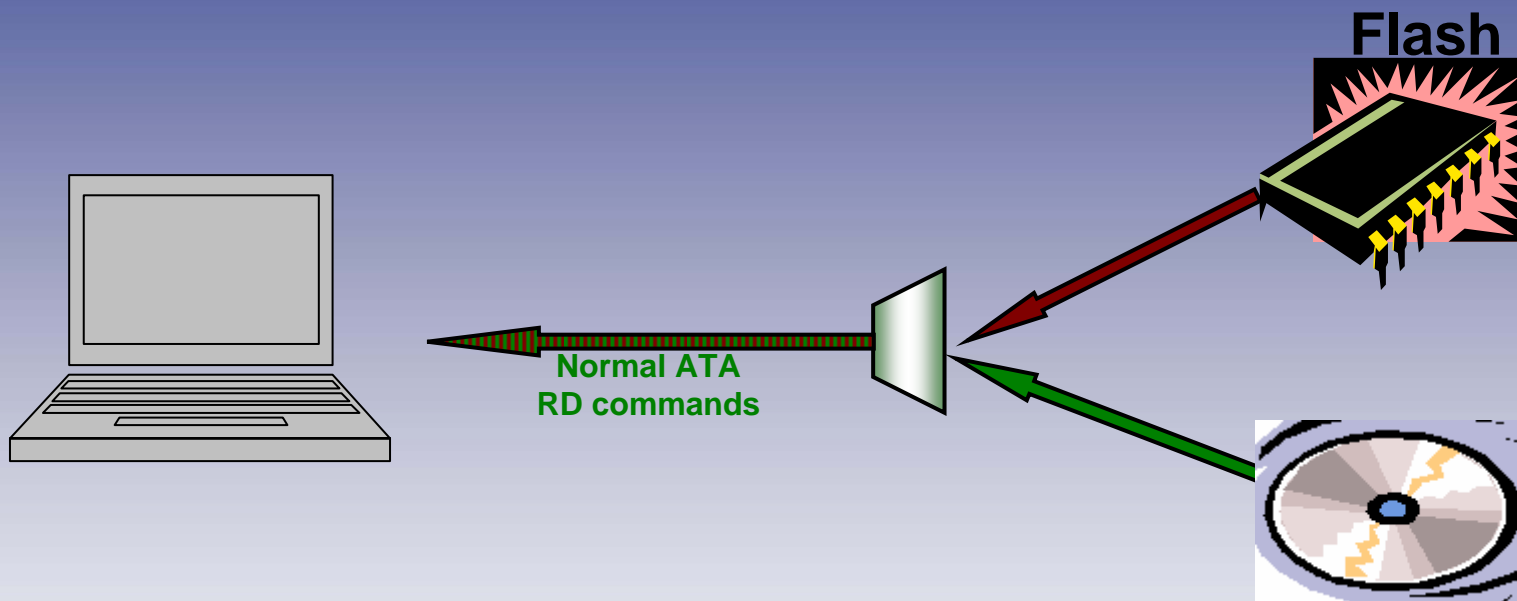
Flash

Piton – Flush

What: Used By Host to free space in the Flash. Host communicates requested number of blocks to be freed.

How Used: Host will request some number of blocks of space in the Flash. The drive will move to disc media unpinned blocks to meet space request from host. The host must unpin a sufficient number of blocks to ensure that space can be freed.   Once space is freed, the host can pin or pin and populate utilizing the freed space.

Flash

**Normal ATA
RD commands**

**What: Following normal cache check, the drive will check Flash for Read data. If
data is present, data is returned from Flash to host. If Read data is not in Flash
data is retrieved from media.**

After a hybrid drive is built the firmware must first determine the "as received" bad blocks before writing any data.

⬇

A bad block table must be generated and maintained throughout the life of the drive.

⬇

As additional bad blocks are created through wear of the NAND flash the drive must decommission the block and add it to the bad block table.

Flash Blocks

Defective as received

Defective as received

Defective as received

Added defect
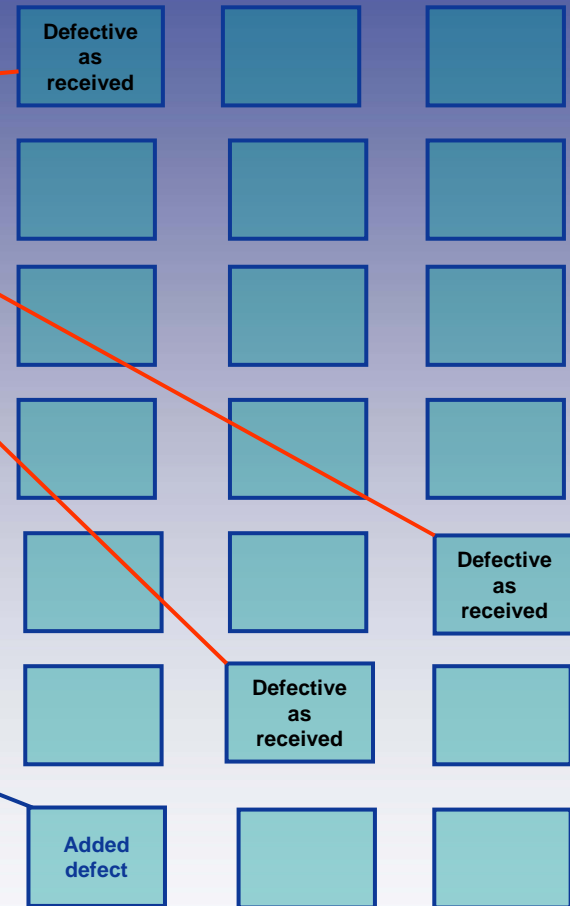
Hybrid Firmware – Wear-leveling

- **Wear-leveling algorithms must be employed to evenly distribute the wear of the NAND blocks by dynamically mapping logical LBAs to physical block locations.**
  - **Reduces the impact of frequently overwritten LBAs.**
  - **Sophistication of the algorithm is dependant on the usage. More critical on fluid data as in computer data storage.**

Impact

- **Data locations change with every re-write.**
- **Complex structures are needed to keep track of data locations.**
- **Dynamic mapping response time is critical.**
- **Structure integrity must be maintained through unexpected power loss.**

Flash Memory Blocks

LBA 3 → LBA 2 → LBA 1 → Dynamic Block Mapping → LBA 2 / LBA 1 / LBA 3
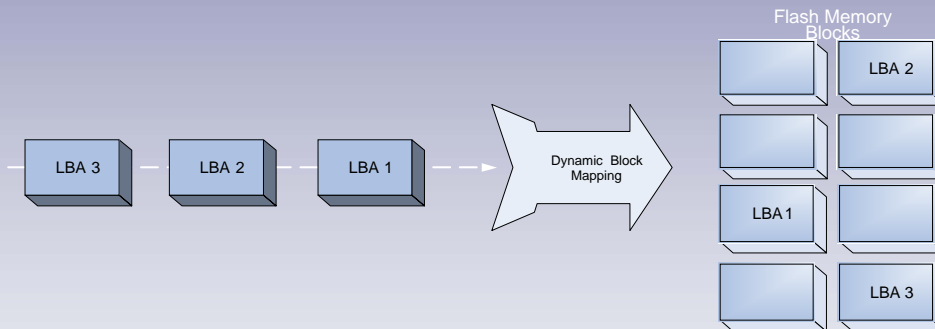
- **Program/Read disturb issues create errors in adjacent pages.**
- **Error is undetectable until the read of the corrupted page.**
- **The recovery mechanism is through application of ECC.**

Impact

- **New generations of Flash technology will require more powerful ECC.**
  - **Density increases.**
  - **Performance enhancements.**
- **As ECC requirements increase software correction will be insufficient without impacting performance. Hardware ECC engines are needed.**
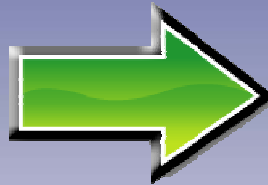
- ## Garbage Collection

  - Rewriting pinned addresses leaves 'stale' data pages.

  - Flash erasures are slow, 1.5mS, and are on block boundaries (64 pages).

  - 'Stale' data pages must be gathered for background erasure.

- ## Integration with standard APMs (Advanced Power Modes)
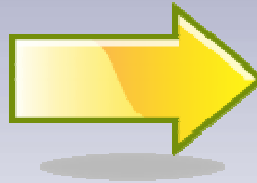
Windows Vista™ and Hybrid Drives

Windows Vista™

Features that work with Hybrid Drives.

SuperFetch™

ReadyDrive™
(Hybrid Drive)

ReadyBoost™
(System Level Flash Memory)

New Vista features targeted at improving system responsiveness.
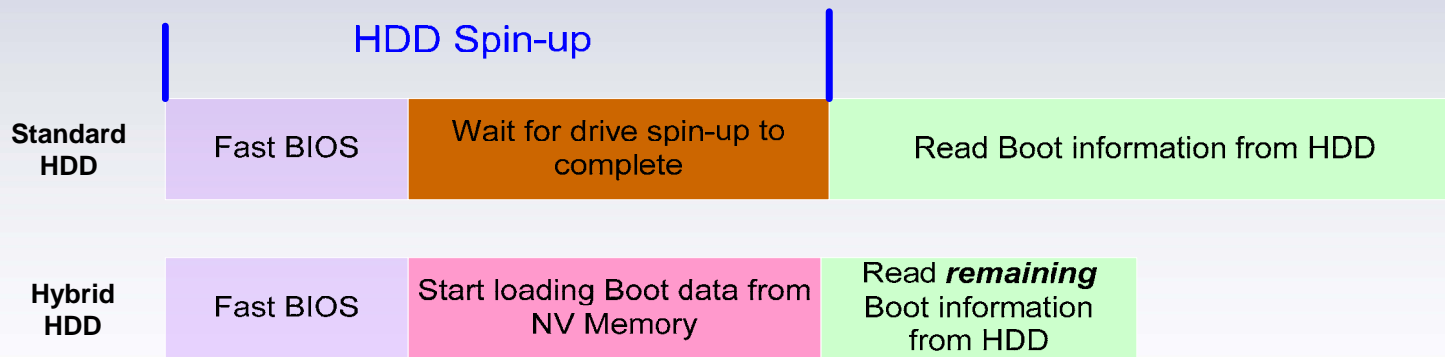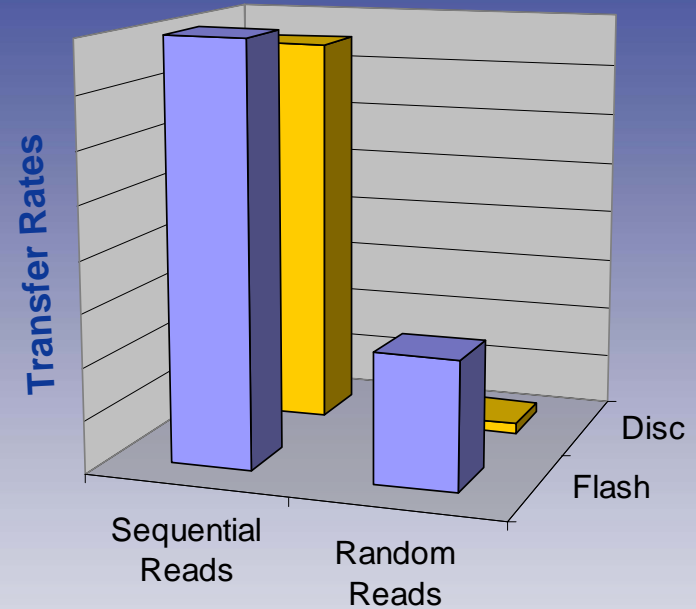
- **Decides what data to cache in Flash.**
- **Incorporates a learning algorithm based on usage patterns.**
- **Differentiates user, time of the day and day of the week.**
- **Works in conjunction with both ReadyDrive and ReadyBoost.**

- **Vista™ utilizes the new ATA-8 hybrid commands to manage the Hybrid Flash contents.**
- **During shutdown boot data is pinned in the Flash for faster response on next boot.**
- **OEMs can specify LBA's to pin in Flash for faster and more consistent application launch.**
- **When Hybrid Low Power Mode is selected then write data is cached in the Flash. Drive only spins up to empty a full Flash or for a read cache miss.**

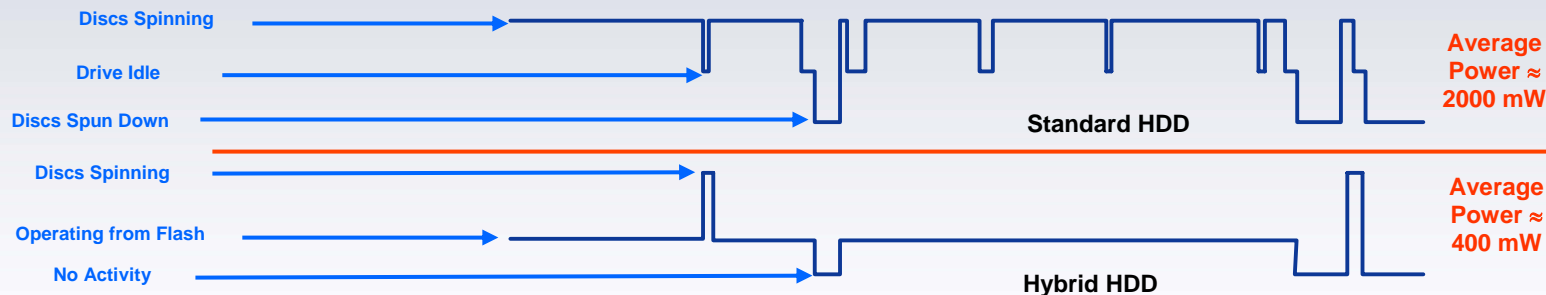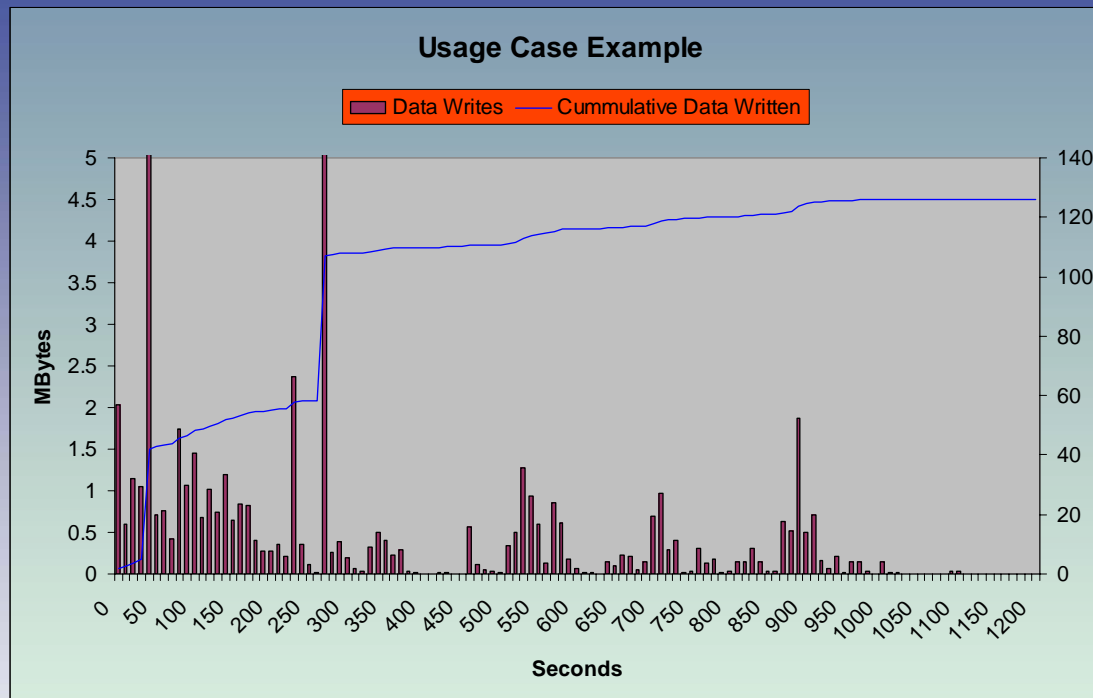# Hybrid Benefits - Faster Boot & Application Launch

- Sequential reads are comparable for Flash & Disc, but random reads are faster from Flash.

- Random files needed for Boot or application launch are pinned to Flash leading to improved response time.

Transfer Rates (chart: Sequential Reads, Random Reads — Disc, Flash)

- In systems with a fast BIOS, the disc spin-up times can be mitigated by pinning Boot data into Flash.

HDD Spin-up

Standard HDD	Fast BIOS	Wait for drive spin-up to complete	Read Boot information from HDD

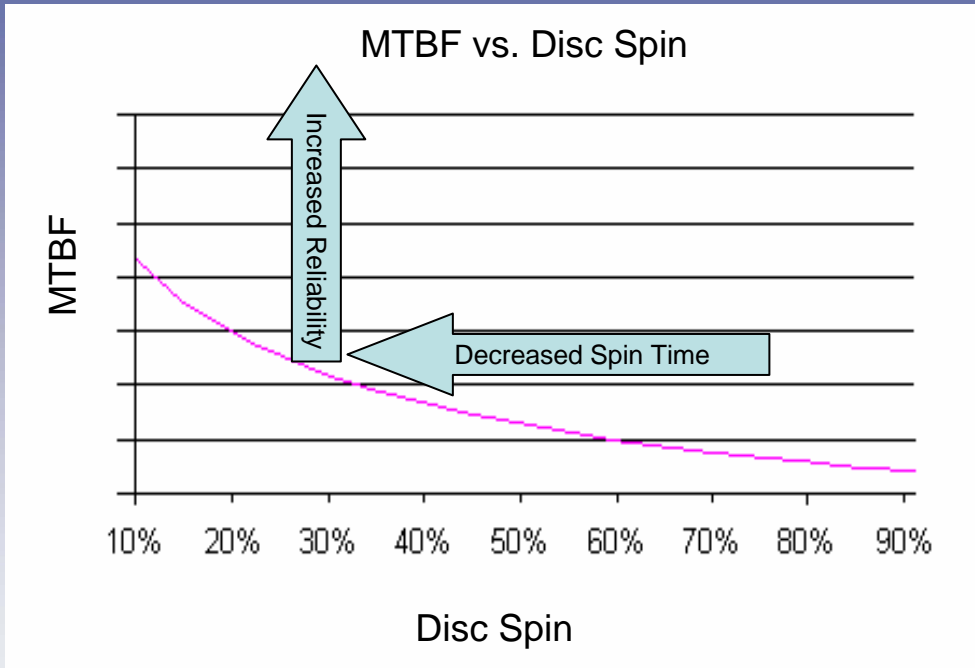|		
Hybrid HDD	Fast BIOS	Start loading Boot data from NV Memory	Read *remaining* Boot information from HDD

Hybrid Benefits - Power Reduction

- **Small block writes to the drive are cached in the non-volatile memory while the discs are kept spun down.**

- **When NV cache is full the drive will be spun up and the data will be transferred to the disc.**

- **This results in significant power reduction.**

Usage Case Example — Data Writes, Cummulative Data Written

Standard HDD — Discs Spinning, Drive Idle, Discs Spun Down — Average Power ≈ 2000 mW

Hybrid HDD — Discs Spinning, Operating from Flash, No Activity — Average Power ≈ 400 mW

MTBF vs. Disc Spin

Increased Reliability

Decreased Spin Time

MTBF

10% 20% 30% 40% 50% 60% 70% 80% 90%

Disc Spin

- Using the Flash memory as a write and read cache enables the hybrid drive to increase the time that the heads are parked and the discs are spun down.

- Increasing the time spun down extends the life of the components and increases reliability.

- This increase in spun down time also increases the probability that the drive is in a more robust state when shock events occur.

- There are many synergies in the handling of Flash memory and magnetic media:

  - Data flow management
  - ECC
  - Re-allocation of bad blocks

- The design of a hybrid drive merges two great storage technologies and utilizes the best benefits of each.

Thank You!

For questions or comments contact:

John Moon at

John.Moon@Seagate.com