Next-generation Magnetic Recording

CSCI 333
April 8, 2019
Last Class: SSDs

- Interface:
  - **Read** from pages
    - As many times as we want
  - **Program** (write to) pages
    - Once -> then need to erase before rewriting
    - Limited endurance -> need to wear level
  - **Erase** whole blocks
    - Erasing is slow
    - Need to perform GC -> migrate live data
- FTLs wear many hats
  - L2P page translation, wear leveling, GC, ECC, …
This Class: “Spinning Rust”

• (Abbreviated recap) Hard Disk Drives
  • Basic Design/Geometry
  • Performance characteristics
• Shingled Magnetic Recording
  • Concepts and interface
  • Position in the storage stack
• Other SMR Interfaces/Opportunities
• Skylight
• IMR
Next Class

- **Filters.** Why shift schedule?
  - Hopefully inspire final project ideas
  - Original DAM model paper is rough… looking for more interesting/clear presentation of material

- For next class: read the Bloom filter paper
  - Optionally read the quotient filter paper
  - Optionally read the cuckoo filter paper

- Goals:
  - Understand/articulate problem(s) that filters solve
  - Describe the high-level design and parameters
Hard Disk Drives (HDDs)

- High capacity, low cost
- Predictable performance

- “Unwritten contract”: LBAs near each other are more efficient to access than LBAs that are far away
HDDs

- Sector (unit of transfer)
- Tracks (concentric circles)
- Platters (rotate)
- Disk Head (seeks in/out)
Performance Observations

- **Setup** (placing the disk head) is expensive $O(10\text{ ms})$
  - seeking to target track
  - Up to a full **rotational delay** to locate target sector
  - Once the disk head is in place, data **transfer** is quite fast $O(100\text{ s MiB/s})$
**Performance Goal**: build a system where data is written sequentially (i.e., no random writes)
Keeping HDDs Relevant

- HDDs compete on $/GiB, not performance
- As capacity goes up, $/GiB down
- Problem:
  - Capacity gains traditionally result of reduced track width to increase density
  - Physical limits restrict our ability to shrink tracks further
- We’re stuck… unless?
ASTC Technology Roadmap

- **ASTC**
  - Advanced Storage Technology Consortium

- **HDMR**
  - Heated-Dot Magnetic Recording
  - (BPMR+HAMR+TDMR)

- **HAMR**
  - Heat Assisted Magnetic Recording

- **PMR**
  - Perpendicular Magnetic Recording

- **PMR+**
  - PMR with Two-Dimensional Magnetic Recording (TDMR) and/or Shingled Magnetic Recording (SMR)

- Graph showing areal density (Tb/in²) over years from 2013 to 2025

[https://blog.seagate.com/craftsman-ship/hamr-next-leap-forward-now/]
Shingled Magnetic Recording (SMR)

- Increases HDD density by overlapping tracks
Shingled Magnetic Recording (SMR)

- Increases HDD density by overlapping tracks

**Insight:** Read head is more precise than write head

**Technique:** Overlap next track, but leave enough of “lower” track visible for safe reading
SMR Introduces Challenges

- **Writing** data becomes harder
  - No random writes
  - No overwrites
  - Must garbage collect to reclaim space
No Random Writes

If we don’t write to zones append-only, we could lose data
No Overwrites

Must perform **out-of-place updates**, or suffer a read-modify-write of entire zone
Garbage Collection

1. Copy **live** data from source to destination
2. Reclaim old zone
Garbage Collection

1. Copy *live* data from source to destination
2. Reclaim old zone
Recall HDD Observations

- **Problem**: Seeking is slow
- **Solution**: perform large sequential I/Os

**Takeaway**: HDD *performance optimizations* translate into SMR *correctness*
Question: who enforces the SMR write constraints?
Drive Managed vs. Host Managed

- **SMR**
  - Zoned Access
  - SMR Translation Layer (STL)

**File System**
- Read and write LBAs
- Read LBAs, write to zones

**SMR Translation Logic**

**SMR Zoned Access**

Software
- + Easy to Deploy
- - Limited HW resources

Firmware
- + Flexible
  - + Shares host resources

---

Diagram:
- File System
  - SMR Translation Layer (STL)
    - Read LBAs, write to zones
  - Read and write LBAs

Software
- + Easy to Deploy
- - Limited HW resources

Firmware
- + Flexible
  - + Shares host resources
Zoned Block Commands

- Conventional Zones
  - Random write capabilities of “normal” disks

- Sequential-write-required zones
  - Query zone status
  - Append blocks to zone’s write pointer
  - Reset zone write pointer (reclaim space)
SMR Opportunities

• Other SMR interfaces have been proposed
  • Caveat Scriptor
  • Configurable zone layouts (Flex) [Feldman ’18]

• Interlaced Magnetic Recording (IMR)
  • Combines HAMR and overlapping tracks
Caveat Scriptor

[Kadekodi ’15 HotStorage]

Basic Idea:

• Drive characteristics are exposed to the user
• User can write anywhere, but data may be lost
Interlaced Magnetic Recording

[Hwang ‘16 Transactions on Magnetics]

Figure 3: Depiction of interlaced track recording

[Feldman ‘18 ;login:]
Magnetic Recording

Figure 1: Track layout for CMR, SMR, and IMR.

[Wu '18 HotStorage]
Open Questions

- Translation layer design
- Garbage collection schemes
- Creating and using new interfaces
- SMR-aware key-value stores
  - Integrating SMR maintenance with DS work
Let’s Think About Designs

• What are our options?

• Static or dynamic?

• What do you think is done in practice ("Archive" DM-SMR drives available at big box stores)?

  • Skylight designed & performed benchmarks to tease out drive parameters