Failures in hard disks and SSDs

Tyler Bletsch
Duke University

Slides include material from Vince Freeh (NCSU), some material adapted from “Hard-Disk Drives: The Good, the Bad, and the Ugly” by Jon Elerath (Comm. ACM, Vol. 52 No. 6, Pages 38-45)
HDD/SSD failures

- Hard disks are the weak link
  - A mechanical system in a silicon world!
- SSDs better, but still fallible

- RAID: Redundant Array of Independent Disks
  - Helps compensate for the device-level problems
  - Increases reliability and performance
  - Will be discussed in depth later
Failure modes

- Failure: cannot access the data
- Operational: faults detected when they occur
  - Does not return data
  - Easy to detect
  - Low rates of occurrence
- Latent: undetected fault, only found when it’s too late
  - Returned data is corrupt
  - Hard to detect
  - Relatively high rates of occurrence
Fault tree for HDD

To learn more about individual failure modes for HDD, see "Hard-Disk Drives: The Good, the Bad, and the Ugly" by Jon Elerath (Comm. ACM, Vol. 52 No. 6, Pages 38-45)
Fault tree for SSD

- Out of sparing capacity
- Controller failure
- Whole flash chip failure
- Calculated limit on write cycles

Operational Failures

- cannot find data
- cannot read data

Latent Failures

- data missing
- written but destroyed

- error during writing

Degradation loss due to write cycles (probabilistic) – gate lost ability to ever hold data

Loss of gate state over time (“bit rot”) – gate lost its current data (due to time or adjacent writes)
What to do about failure

- Pull disk out
- Throw away
- Restore its data from parity (RAID) or backup
The danger of latent errors

- **Operational errors:**
  - Detected as soon as they happen
  - When you detect an operational error, the total number of errors is likely **one**

- **Latent errors:**
  - Accrue in secret over time!
  - In the darkness, little by little, your data is quietly corrupted
  - When you detect a latent error, the total number of errors is likely **many**

- In the intensive I/O of reconstructing data lost due to latent errors, more likely to encounter operational error
  - Now you’ve got multiple drive failure, data loss more likely
Minimizing latent errors

- Catch latent errors earlier (so fewer can accrue) with this highly advanced and complex algorithm known as **Disk Scrubbing**:

  Periodically, read everything
### Disk reliability

- **MTBF (Mean Time Between Failure):** a useless lie you can ignore

<table>
<thead>
<tr>
<th>Specifications</th>
<th>8TB</th>
<th>6TB</th>
<th>5TB</th>
<th>4TB</th>
<th>3TB</th>
<th>2TB</th>
<th>1TB</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model number</strong></td>
<td>WD80EFZX</td>
<td>WD60EFRX</td>
<td>WD50EFRX</td>
<td>WD40EFRX</td>
<td>WD30EFRX</td>
<td>WD20EFRX</td>
<td>WD10EFRX</td>
</tr>
<tr>
<td><strong>Formatted capacity</strong></td>
<td>8TB</td>
<td>6TB</td>
<td>5TB</td>
<td>4TB</td>
<td>3TB</td>
<td>2TB</td>
<td>1TB</td>
</tr>
<tr>
<td><strong>Form factor</strong></td>
<td>3.5-inch</td>
<td>3.5-inch</td>
<td>3.5-inch</td>
<td>3.5-inch</td>
<td>3.5-inch</td>
<td>3.5-inch</td>
<td>3.5-inch</td>
</tr>
<tr>
<td><strong>Advanced Format (AF)</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Native command queuing</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>RoHS compliant</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

**Performance**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Interface speed</td>
<td>178 MB/s</td>
<td>175 MB/s</td>
<td>170 MB/s</td>
<td>150 MB/s</td>
<td>147 MB/s</td>
<td>150 MB/s</td>
<td>147 MB/s</td>
</tr>
<tr>
<td>Internal transfer rate</td>
<td>128</td>
<td>64</td>
<td>64</td>
<td>64</td>
<td>64</td>
<td>64</td>
<td>64</td>
</tr>
<tr>
<td>Cache (MB)</td>
<td>5400 RPM Class</td>
<td>5400 RPM Class</td>
<td>5400 RPM Class</td>
<td>5400 RPM Class</td>
<td>5400 RPM Class</td>
<td>5400 RPM Class</td>
<td>5400 RPM Class</td>
</tr>
</tbody>
</table>

**Reliability/Data Integrity**

| Load/unload cycles*     | 600,000 | 600,000 | 600,000 | 600,000 | 600,000 | 600,000 | 600,000 |
| Non-recoverable read errors per bits read | <1 in 10^14 | <1 in 10^14 | <1 in 10^14 | <1 in 10^14 | <1 in 10^14 | <1 in 10^14 | <1 in 10^14 |
| **MTBF (hours)**         | 1,000,000 | 1,000,000 | 1,000,000 | 1,000,000 | 1,000,000 | 1,000,000 | 1,000,000 |
| **Limited warranty (years)** | 3       | 3       | 3       | 3       | 3       | 3       | 3       |

---

1,000,000 hours = 114 years

“Our drives fail after around a century of continuous use.”

-- A Huge Liar
Data from BackBlaze

- **BackBlaze**: a large scale backup provider
  - Consumes thousands of hard drives, publishes *health data on all of them publically*
  - **Data presented** is a little old – newer data exists (but didn’t come with pretty graphs)
- Other large-scale studies of drive reliability:
  - “*Failure Trends in a Large Disk Drive Population*” by Pinheiro et al (Google), FAST’07
  - “*Disk Failures in the Real World: What Does an MTTF of 1,000,000 Hours Mean to You?*” by Schroeder et al (CMU), FAST’07
General Predicted Failure Rates

- Decreasing Failure Rate
- Constant Failure Rate
- Increasing Failure Rate

Failure Rate

- Early "Infant Mortality" Failure
- Constant (Random) Failures
- Wear Out Failures

Observed Failure Rate

Time

BackBlaze
Annual Failure Rate Each Quarter

- X-axis: Quarter
- Y-axis: Annual Failure Rate

Bar chart showing the annual failure rate per quarter.
Interesting observation:
The industry standard warranty period is 3 years...
80% of Drives Last Four Years

Hard Drive Survival Rates - Chart 2
What about SSDs?

- From recent paper at FAST’16: “Flash Reliability in Production: The Expected and the Unexpected” by Schroeder et al (feat. data from Google)

**KEY CONCLUSIONS**

- Ignore Uncorrectable Bit Error Rate (UBER) specs. A meaningless number.
- **Good news:** Raw Bit Error Rate (RBER) increases slower than expected from wearout and is not correlated with UBER or other failures.
- High-end SLC drives are no more reliable that MLC drives.
- **Bad news:** SSDs fail at a lower rate than disks, but UBER rate is higher (see below for what this means).
- **SSD age, not usage, affects reliability.**
  - Bad blocks in new SSDs are common, and drives with a large number of bad blocks are much more likely to lose hundreds of other blocks, most likely due to die or chip failure.
  - 30-80 percent of SSDs develop at least one bad block and 2-7 percent develop at least one bad chip in the first four years of deployment.

Drive replacements

- Percentage of drives replaced annually due to suspected hardware problems over the first 4 years in the field:

  - Approximately 1-2% of drives replaced annually, much lower than hard disks!
  - 0.5-1.5% of drives developed bad chips per year
    - Would have been replaced without methods for tolerating chip failure

Average annual replacement rates for hard disks (2-20%)
Errors experienced during a drive’s lifecycle

- **Non-transparent errors** common:
  - 26-60% of drives with uncorrectable errors
  - 2-6 out of 1,000 drive days experience uncorrectable errors
  - Much worse than for hard disk drives (3.5% experiencing sector errors)

- **Transparent errors**
  - Correctable error
  - Read retry
  - Write retry
  - Erase error

- **Uncorrectable error**
  - Final write error
  - Meta error
  - Timeout error

- Symptoms of bad blocks or bad chips

---

Slide from "Flash Reliability in Production: The Expected and the Unexpected" by Schroeder et al. FAST’16.
Overall conclusions on drive health

• HDD:
  • Usually just die, sometimes have undetected bit errors.
  • Need to protect against drive data loss!

• SSD:
  • Usually have undetected bit errors, sometimes just die.
  • Need to protect against drive data loss!

• Overall conclusion?
  Need to protect against drive data loss!