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
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

Operational Failures

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

or

cannot find data

Latent Failures

- 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)

or

data missing

error during writing

written but destroyed

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>Model number</td>
<td>WD80EEXZ</td>
<td>WD60EFRX</td>
<td>WD50EFRX</td>
<td>WD40EFRX</td>
<td>WD30EFRX</td>
<td>WD20EFRX</td>
<td>WD10EFRX</td>
</tr>
<tr>
<td>Form factor</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>Advanced Format (AF)</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>Native command queuing</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>RoHS compliant</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>Performance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<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>150 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>
<tr>
<td>Performance Class</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Load/unload cycles</td>
<td>600,000</td>
<td>600,000</td>
<td>600,000</td>
<td>600,000</td>
<td>600,000</td>
<td>600,000</td>
<td>600,000</td>
</tr>
<tr>
<td>Non-recoverable read errors per bits read</td>
<td>&lt;1 in 10^14</td>
<td>&lt;1 in 10^14</td>
<td>&lt;1 in 10^14</td>
<td>&lt;1 in 10^14</td>
<td>&lt;1 in 10^14</td>
<td>&lt;1 in 10^14</td>
<td>&lt;1 in 10^14</td>
</tr>
<tr>
<td>MTBF (hours)</td>
<td>1,000,000</td>
<td>1,000,000</td>
<td>1,000,000</td>
<td>1,000,000</td>
<td>1,000,000</td>
<td>1,000,000</td>
<td>1,000,000</td>
</tr>
<tr>
<td>Limited warranty (years)</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

1,000,000 hours = 114 years

“Our drives fail after around 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

Observed Failure Rate

Constant (Random) Failures

Wear Out Failures

Time
Interesting observation: The industry standard warranty period is 3 years...
80% of Drives Last Four Years

Hard Drive Survival Rates - Chart 2

80% drives live
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:

<table>
<thead>
<tr>
<th>MLC-A</th>
<th>MLC-B</th>
<th>MLC-C</th>
<th>MLC-D</th>
<th>SLC-A</th>
<th>SLC-B</th>
<th>SLC-C</th>
<th>SLC-D</th>
</tr>
</thead>
<tbody>
<tr>
<td>~1-2%</td>
<td>~2%</td>
<td>~1.5%</td>
<td>~1%</td>
<td>~1%</td>
<td>~1%</td>
<td>~1%</td>
<td>~1%</td>
</tr>
</tbody>
</table>

  Average annual replacement rates for hard disks (2-20%)

- ~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
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)
Overall conclusions on drive health

- **HDD:**
  Need to protect against drive data loss!

- **SSD:**
  Need to protect against drive data loss!

- **Overall conclusion?**
  Need to protect against drive data loss!