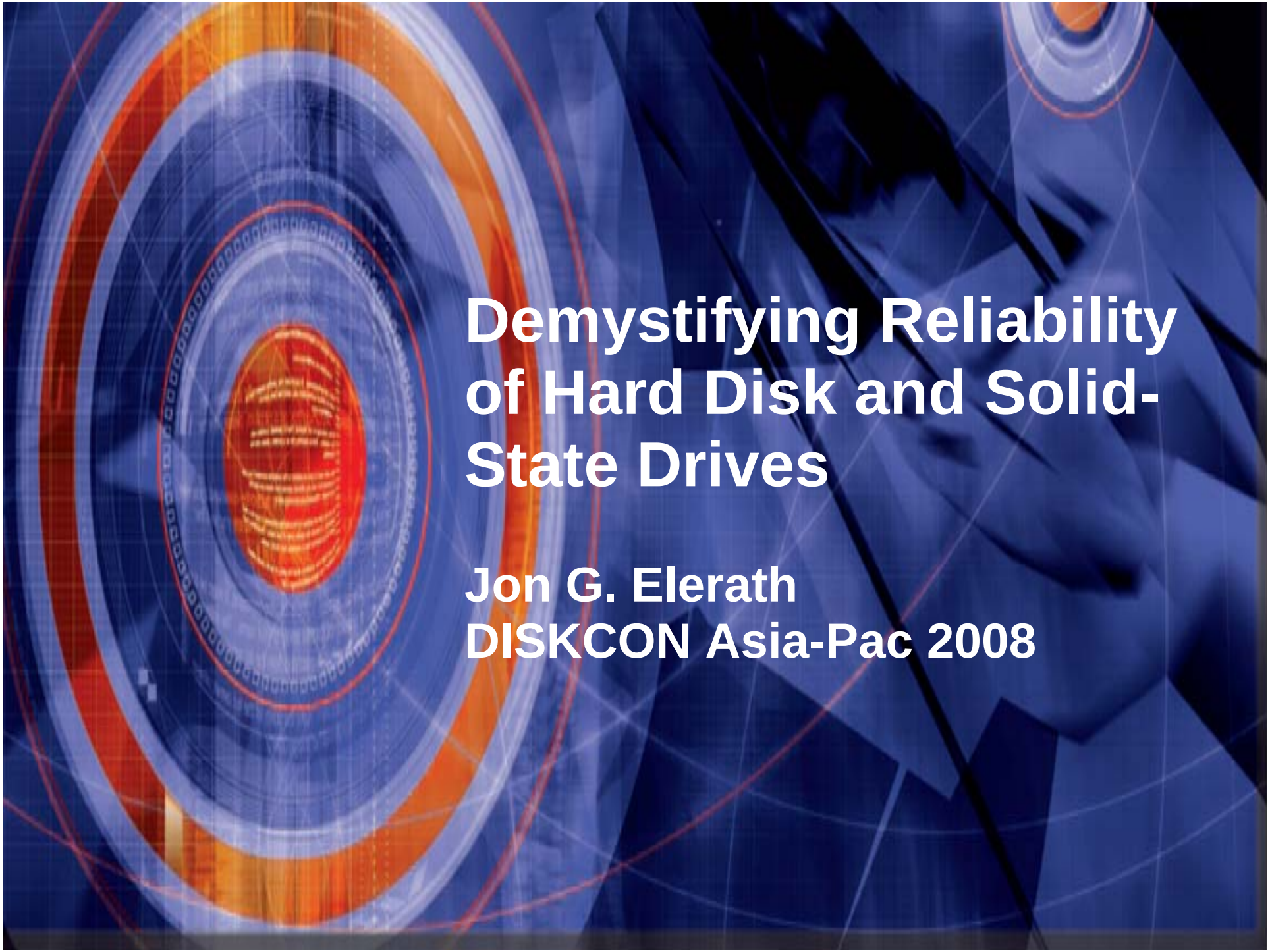


Demystifying Reliability of Hard Disk and Solid-State Drives

Jon G. Elerath

DISKCON Asia-Pac 2008

Foreword

This presentation IS:

- A comparative ***overview*** of reliability for hard disk drives (HDDs) and solid state disk drives (SSDs)
- Includes reliability specifications; major failure modes & mechanisms; and failure prevention, mitigation and remedies for HDDs, SLC-NAND memory and SSDs (made from the memory)
- A challenge to the HDD community

This presentation IS NOT:

- A comparison of costs or performance parameters
- A comprehensive discussion of *all* aspects of reliability
- An endorsement of any particular technology
- All one needs to know to recommend HDDs or SSDs

Agenda

- Reliability claims
  - Smoke and mirrors
  - Claims and statistics
- HDD vs. SSD
  - Operation
  - Failure modes and effects (some causes)
  - Reliability tests in development and manufacturing process
  - Problem prevention & remedies
- Open questions and the future
- Conclusions and closing thought

SSD Smoke and Mirrors

Source:

B. Crothers, “Samsung defends flash drive reliability” [1]

Quotes Michael Yang, Flash Marketing Manager, Samsung

“ [wear leveling] makes it virtually impossible to wear out a flash chip. Yang said a pattern could be perpetually repeated in which a 64GB SSD is completely filled with data, erased, filled again, then erased again every hour of every day for years and the user still wouldn't reach the theoretical write limit. He added that if a failure ever does occur, it will not occur in the flash chip itself but in the controller.”

Reliability Claims

Statistically Imprecise Reliability Specifications

Hard Disks Drives[1]

- MTBF[6] = 1.4 M hours [2]
- UCE = $1x10^{-16}$ bits read [3]
- 5 Year Warranty [2]

- None state useful life! Only warranty period.

[1]FCAL Enterprise HDDs
[6] Product Specification

Solid State Drives[2]

- MTBF
  - 2 M hours[3] [4]
  - 1.7 M hours[3] (64GB) [5]
  - 5 M hours[4] [6]
- UCE
  - $1x10^{-15}$ bits read [5]
  - $1x10^{-20}$ bits read [7]
- Warranty = 5 years [7]
- Program/Erase (P/E) Endurance Life of 5 year[5] [5]

[2] 1-bit SLC, NAND Flash
[3] Handbook prediction
[4] Extrapolation of test data and assumed usage profile
[5] 800GB per day,32GB drive; normal operating conditions. Static and dynamic wear-leveling.

Reliability Basics - MTBF & Bathtubs

MTBF Basics

Let $\lambda(t)$ = failure rate

- Only if $\lambda(t) = \lambda$, then MTBF = $1/\lambda$; so failure rate does not change over the product life, from time = 0 and going for ever!! There is no end at which the failure rate increases or decreases
- If the MTBF = 1.4 M hr (constant failure rate), >88% will survive over 20 years and 36.8% will survive >160 years
- The probability of failure in 1 week's time is the same for the first week as for week 10,000
- Stating the MTBF alone is vague. At some point, failure mechanisms change; I don't believe HDDs will survive for 160 years.

Reliability Basics - Bathtub Curve

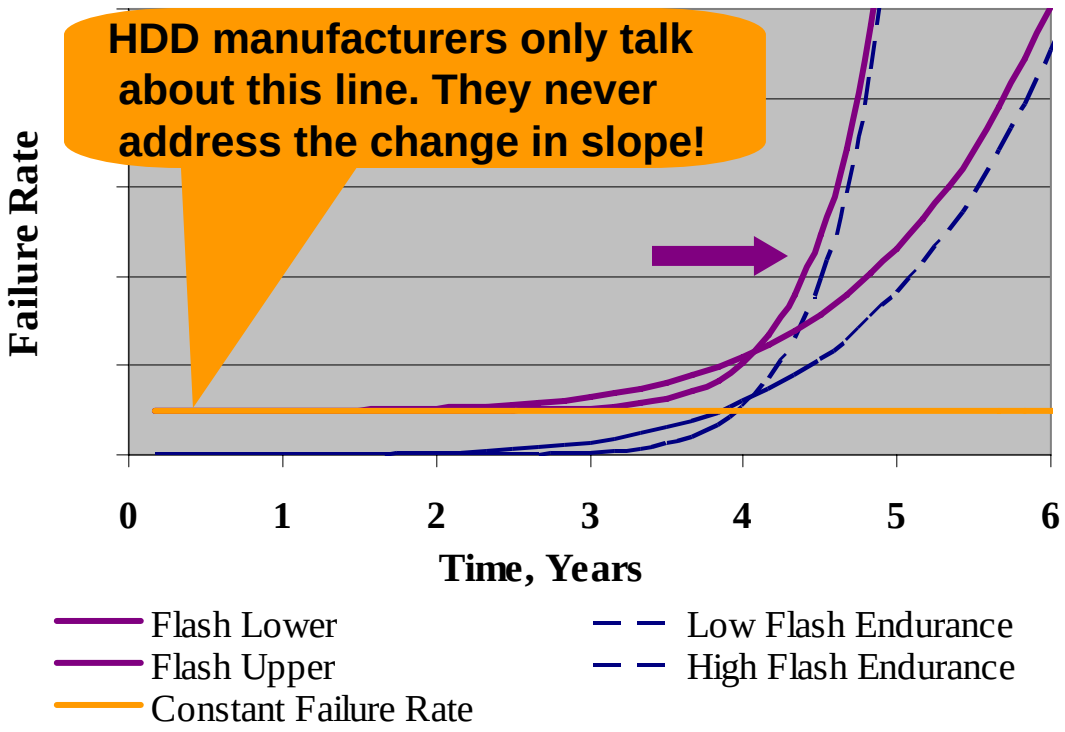
Bathtub Curves

Flash/SDD specifications are ***slightly better*** than HDD at describing assumptions for reliability statements.

Few HDD specifications discuss how usage affects reliability [8] and none have a relationship between usage and end of useful life. However, Flash does!

Endurance limit for NAND flash depends on usage, so this chart is WRONG!!

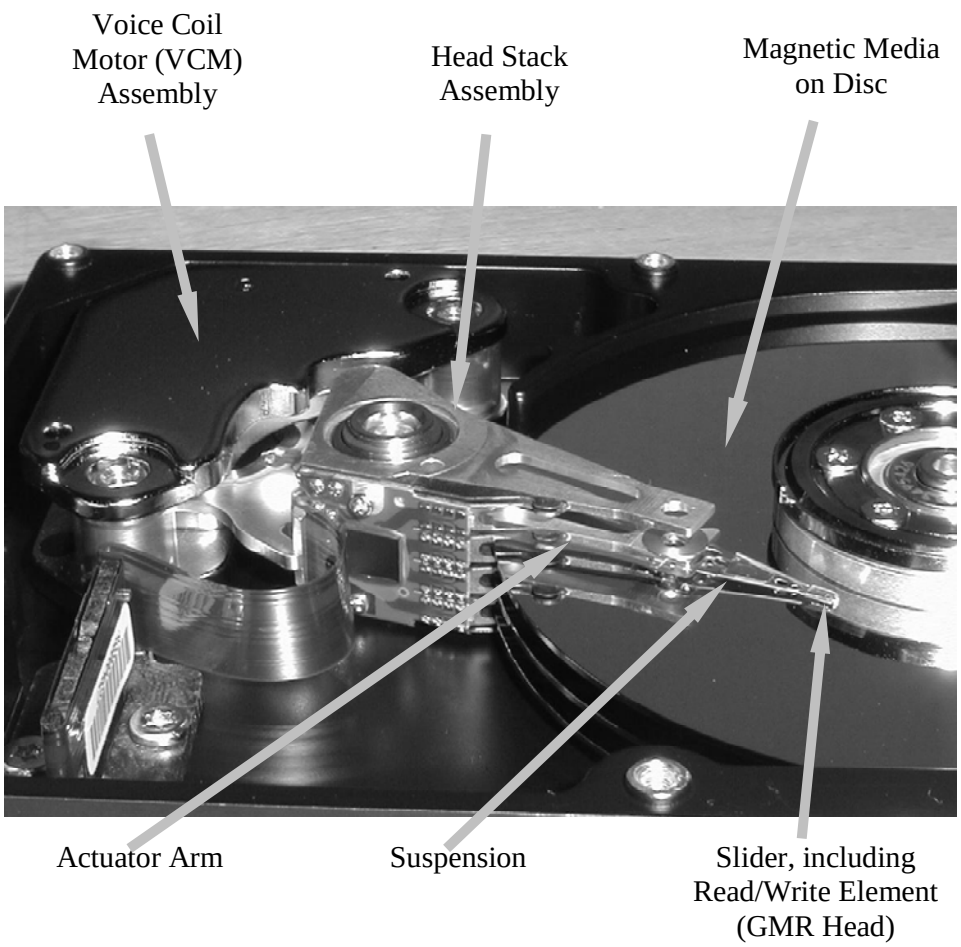
But where are the HDD specs.??? Where do I place the HDD “wear-out” curve?



HDD Construction & Operation

Discs spin and heads fly (Duh!)...closely!

- Actuator arm positions head
- Firmware keeps track of physical block location versus logic block
- Separate heads (on each slider) for read and write
- Write current changes direction to change induced magnetic field on disc
- Magnetic orientation on media is preserved as long as no “significant” magnetic fields come near



HDD Failure Modes

HDD Symptoms, Can't:

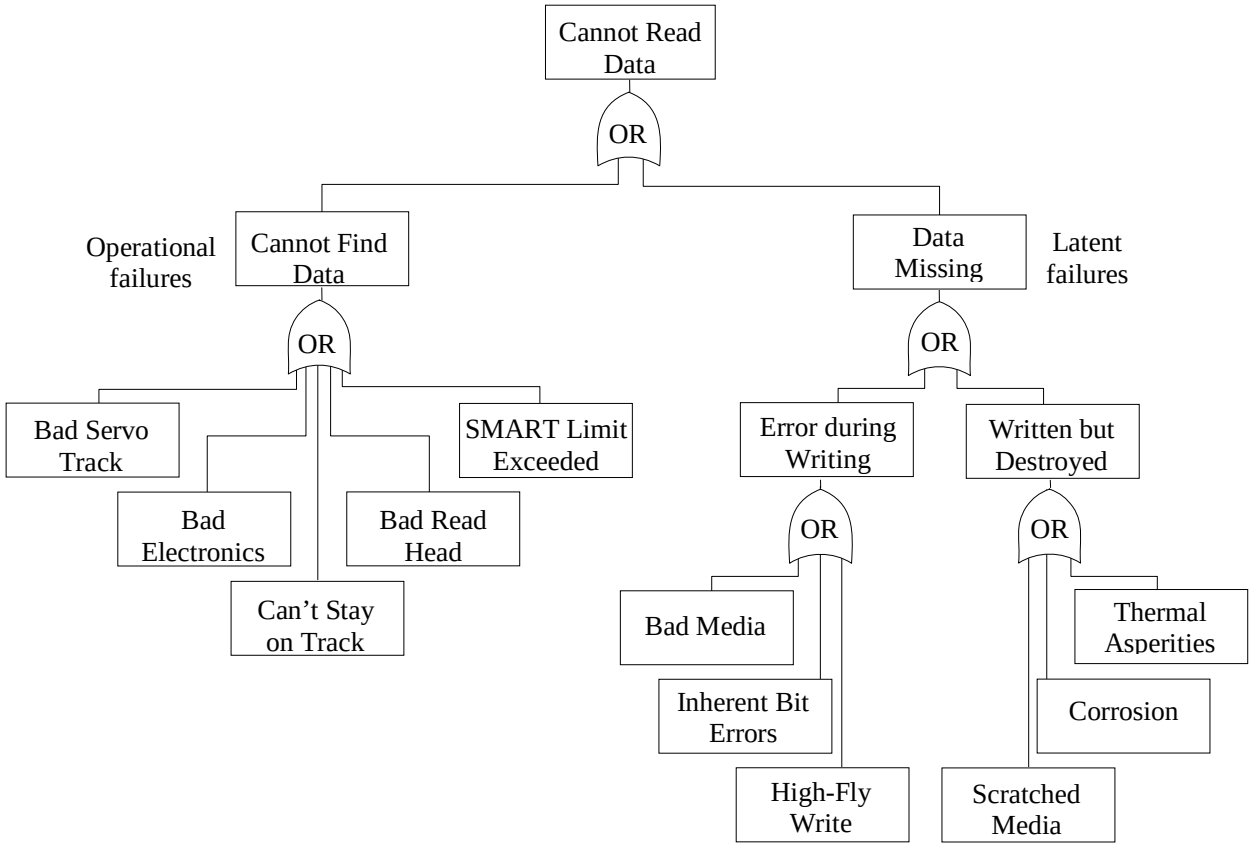
- get on track
- stay on track (NRRO, etc.)
- read data quickly enough (time-out)
- write data
- read data

Components involved:

- head related
  - unstable
  - failed
  - flying too high
- media
  - defects
  - scratches
- motors
- electronics (DRAM)
- firmware and servo

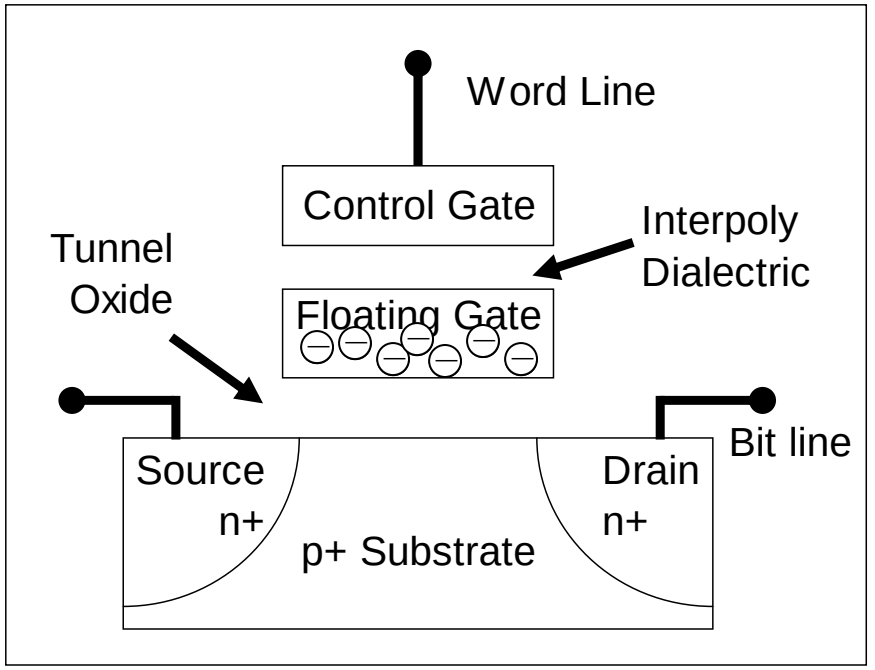
HDD Failure Modes & Effects

Symptoms, modes, mechanisms and causes



NAND-SLC Flash Construction & Operation

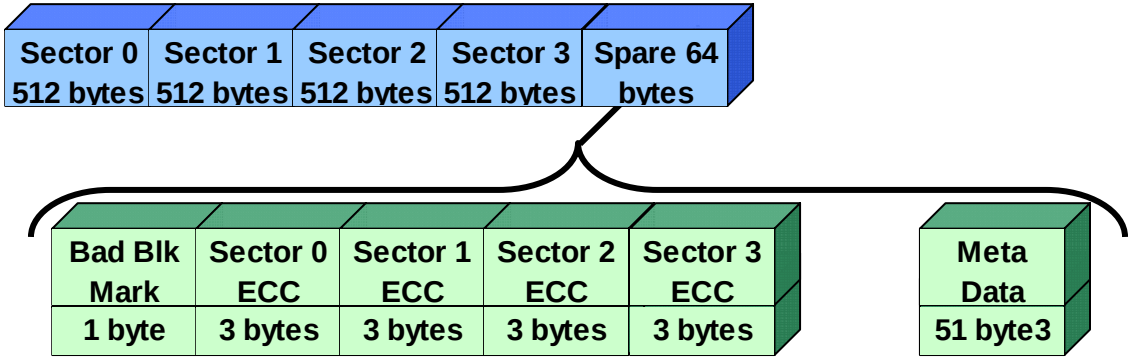
High energy is used to transport charge through the oxide by Fowler-Nordheim (FN) tunneling for erasure or channel hot electron (CHE) injection for programming [9]



At the Floating Gate:
+ is Logic 1 (erased)
- is Logic 0 (programmed)

NAND-SLC Flash Construction & Operation

Page Layout for 2048[7] NAND SLC Flash [10]



1 sector = 512 bytes
1 page = 4 sectors + ECC/spares = 2048 + 64 bytes
1 block = 64 pages = 128k + 4k bytes
1 device = 2048 blocks = 2,112MB

Blocks are the smallest erasable units (128k + 4k bytes)
Pages are the smallest programmable units (2048 bytes)

[7]Most manufacturers are moving towards 4096 bytes

Flash/SSD Failure Modes & Effects

- Permanent failure: bit is stuck and cannot change
- Endurance: Stuck cells due to charge-trapping in oxide layer or breakdown of the oxide. Not recoverable by an erase; mark cells as bad and do not use [11][8]
- Program disturb: Cells not being programmed receive elevated voltage stress. Always in block being programmed. Can be on unselected page or selected page that is not supposed to be programmed. Erase returns cells to undisturbed levels
- Read disturb: In the block being read, but always on pages NOT being read. Erase returns the cells to undisturbed levels
- Data Retention: Floating gate charge is reduced due to charge leakage through oxide defects. Block can be reprogrammed, but retention may not be as long as “pristine” oxide layer

[8][12] states trapped charge effects can be reversed with an erase

Flash/SSD Reliability Relationships

Time/Use dependent relationships

- Endurance: Specification is 100,000 P/E cycles (perfect wear-leveling)

$$\frac{100{,}000\ P/E\ cycles}{device} \times \frac{60{,}000\ MB}{P/E\ cycle} \times \frac{sec.}{50\ MB} \times \frac{hour}{3600\ sec.} \times \frac{year}{8760\ hours} = 3.8\ \frac{years}{device}$$

? ? ? ?

- Data Retention: Retention as function of time and temperature through Arrhenius relationship [12].

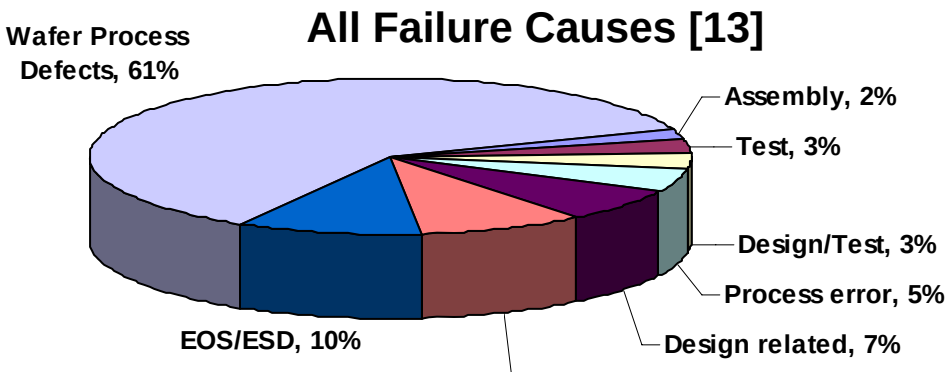
$$AF = e^{-\frac{E_a}{k}\left(\frac{1}{T_1}-\frac{1}{T_2}\right)}$$

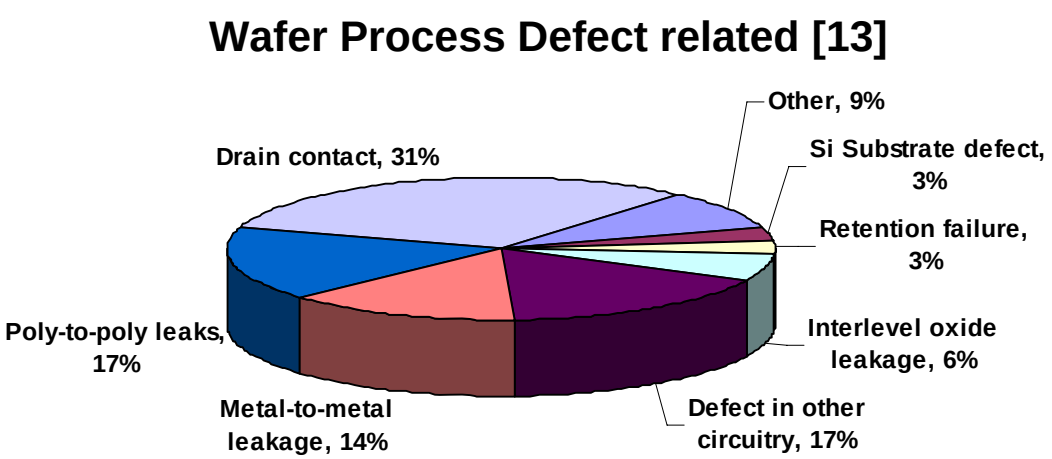
Where

AF = Acceleration factor
E_a = 0.6 eV = Activation energy
k = Speed Constant = 86.25×10^{-6}
T1 = Temperature 1 (K)
T2 = Temperature 2 (K)

Flash Memory Failures

- Field Data Compilation [13]
  - 2006, NOR Flash is *"largely applicable to NAND flash"*
  - Non-confirmed failures (at the FA lab) and repetitive cases are omitted
  - Wafer process defects are dominated by particle contamination. Are these the 0.4% that are accounted for in the MTBF specifications?
- Perspective: *"Quality levels for SSDs for PCs are very difficult to achieve in many aspects. Only companies that have direct fab control can succeed to get the appropriate quality levels."* [4]





HDD & SSD Tests

HDDs-Mature Industry

- Much consistency across suppliers
- Development Tests
  - DVT
  - ESS/EOS
  - Shock/vib.
  - Multiple RDTs (reliability demonstration tests)
- Test Conditions
  - Unproven acceleration factors
  - Vague relationship between usage failure rates
- Manufacturing Processes
  - Vintage to vintage life-time variability [last year's presentation]
- Complex f/w; years of experience

SSD-Immature Industry

- Little consistency across suppliers
- Development Tests
  - DVT
  - ESS/EOS
  - Shock/vib.
  - RDT (inconsistent from mfgr to mfgr.)
- Test Conditions
  - Inconsistent
  - No common set of conditions (sample size, environments)
  - Known acceleration factors
  - Known dependence on use
- Manufacturing Processes
  - Lot-to-lot variability REAL issue!
- Low extent of f/w maturity

HDD & SSD Problem/Mitigation Comparison

HDDs

- SMART (not very effective)
- ECC on fly (critical to reliability)
- Reallocation of bad sectors. Spare blocks distributed around the surfaces of the discs for reallocation
- Scrub and correct latent defects
- 4k sectors (greater ECC capability)
- Mechanisms that require replacing entire HDD are common (heads and particles)
- “the SSD still falls short when compared with HDDs, which have virtually unlimited write cycles per bit.” [14]
- Minimum number of bytes to map out is a sector (512 bytes or, soon, 4k bytes)

SSD

- SMART (Not the same as for HDDs; limited use; no standard)
- ECC - correct read and write disturbed bits (critical to reliability)
- Reallocate bad blocks (2-3%)
- Wear leveling (static & dynamic)
- Dynamic bad block detection and tracking before and after data is written. Map out questionable blocks. Not user selectable. Built into product
- Mechanisms causing entire SSD to fail are uncommon (very low probability)
- Minimum number of bytes to map out is a block (256 sectors, 128k bytes for 2k byte pages or 256k bytes for 4k byte pages)
- Reliability vs. time relationships better understood than for HDDs

SSD Open Questions

- Statistical Data
  - Distribution of cycles to failure
  - Field failure data
  - Server (NetApp) usage profile
- Qualification tests
  - How are they conducted
  - Definition of failure
  - Quantity tested
- Production Control, as it affects reliability
  - Lot-to-lot variability quantified/controlled?
  - ORT conducted?
  - Burn-in? Why (or why not?)
- Failure Consequences
  - Data retrieval from the remainder of the device after cell/sector/page/block failure?
  - What “whole device” failure modes/mechanisms/causes are there?
  - (controller itself..material limitation in the NAND devices

Conclusions

- SSD adoption depends on many performance issues for specific applications, not just reliability
- SSD reliability highly dependent on application. SSDs *not* more reliable than HDDs in *all* applications & less reliable in some
- SSD returns for performance problems (as is the case today in the PC industry) are perceived as reliability issues
- Flash has the entire semiconductor industry helping resolve some issues (density, contamination)
- Statistical relationships for reliability and time are better quantified for flash than HDDs (acceleration factors and endurance)
- SSD f/w in infancy; inconsistent across competitors
- SSDs not as reliable as all the “hype” in all cases, but potentially a strong competitor in many applications

Closing Thoughts

“You don’t sell the steak, you sell the sizzle.”

D. Rued, brother-in-law

References

1. B. Crothers, “Samsung defends flash drive reliability,” c/net News.com, February 22, 2008, 6:00 AM PST, http://www.news.com/8301-10784_3-9876557-7.html, last accessed 02/22/08.
2. “Product Overview, Cheetah 10K.7, 300-Gbyte enterprise disc drive,” Seagate, http://www.seagate.com/docs/pdf/marketing/Seagate_Cheetah_10K-7.pdf, last accessed 02/15/08.
3. “Data Sheet, Cheetah® 15K.5,” http://www.seagate.com/docs/pdf/datasheet/disc/ds_cheetah_15k_5.pdf, last accessed 02/15/08.
4. O. Balaban, “Bringing Solid State Drives to Mainstream Notebooks,” IDEMA, DISKCON-USA, 2007.
5. “RealSSD™ 2.5-Inch SATA, NAND Flash Solid State Drive (SSD),” Micron Specification Sheet, Dec. 2007, http://download.micron.com/pdf/datasheets/realssd/realssd_flash_drive_2.5.pdf, last accessed 02/15/08.
6. H. Pon, K. Rao, “A NAND Flash PC Platform Read Write Cache,” IEEE, 22nd Non-volatile Semiconductor Memory Workshop, 26-30 Aug, 2007, 21-22.
7. “Mach8 Solid State Drive Family Specification,” http://stec-inc.com/product/mach8.php, last accessed 02/25/2008.
8. “MTBF”, Hard disk Drive Knowledge Base, 2007, http://www.hitachigst.com/hddt/knowtree.nsf/cffe836ed7c12018862565b000530c74/a03ca96b145bc68e86256df6004d318e?OpenDocument&Highlight=0,mtbf, last accessed 09/10/07.
9. H. Handschuh and E. Trichina, “Securing Flash Technology,” IEEE, Workshop on Fault Diagnosis and Tolerance in Cryptography, 2007, 3-17.

References

10. W. Prouty, “NAND Flash Reliability and Performance, The Software Effect,” Flash Memory Summit, Santa Clara, CA, August 2007, http://download.micron.com/pdf/presentations/events/flash_mem_summit_waprouty_nand_reliability.pdf, last accessed 02/24/2008.
11. J. Cooke, “The Inconvenient Truths of NAND Flash Memory,” Flash Memory Summit, Santa Clara, CA, August 2007. http://download.micron.com/pdf/presentations/events/flash_mem_summit_jcooke_inconvenient_truths_nand.pdf, last accessed 02/24/2008.
12. P. Forstner, “MSP430 Flash Memory Characteristics,” Application Report SLAA334-September 2006, http://focus.ti.com/lit/an/slaa334/slaa334.pdf, last accessed 02/24/2008.
13. P. Muroke, “Flash Memory Field Failure Mechanisms,” IEEE 44th Annual International Reliability Physics Symposium, 2006, 313-316.
14. D. Reinsel et al., “Replacing HDDs with SSDs: The Business Case for Transition,” IDC, http://driveyourlaptop.com/images/download/reports.zip, last accessed 02/22/08.