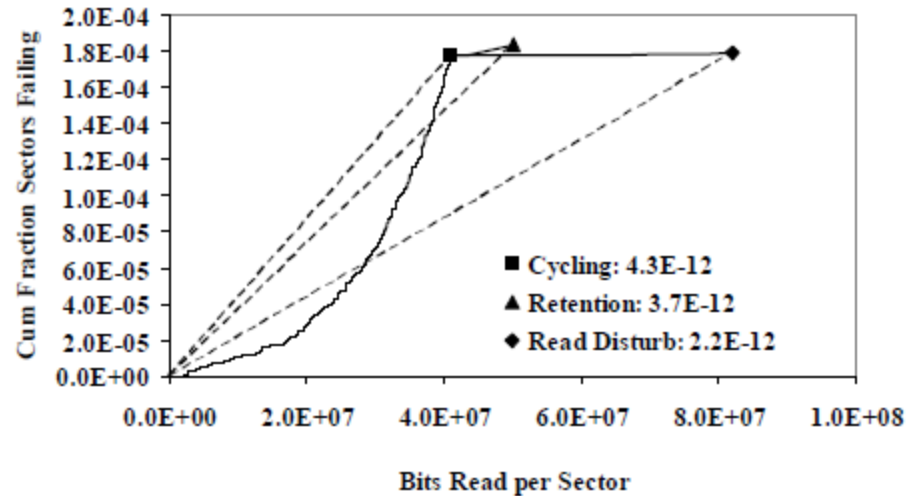# Three Surprises Concerning SSDs

Ted Wobber

MSR Silicon Valley

HPTS – October 2009

NAND-flash media errors

- Program/erase cycle errors

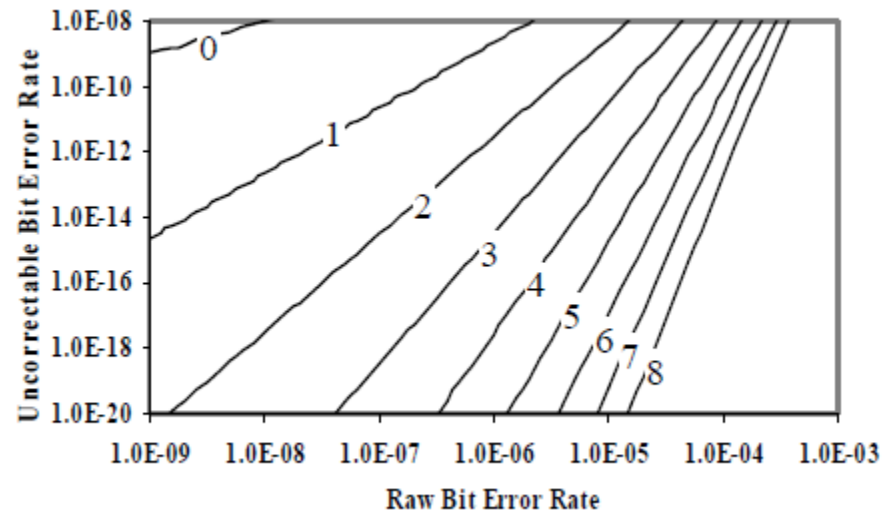- Retention errors

- Read-disturb errors

With 72nm MLC, rated at 10000 cycles, and 1-bit ECC:

Graphs from Mielke, et al, 46th Annual International Reliability Physics Symposium, 2008
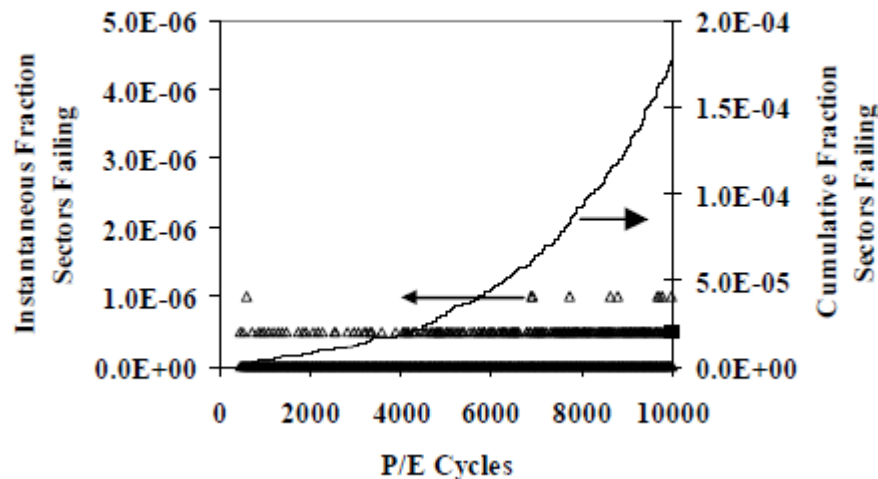
ECC is your friend

- As scale decreases … bit error rate increases
- More ECC bits imply:
  - more memory to store them
  - more logic to compute them
  - larger codeword

Program/erase cycle errors dominate

- Lifetime is defined by cycle count

- These numbers are beginning to get big!

- I don't like the shape of this curve!

(the graph doesn't go beyond 10000 cycles ☺)

Surprise #1

- Greater SSD capacity -> longer lifetime
  - with the same workload
  - assuming appropriate wear-leveling
- No such correlation for rotating disks

- The FTL distributes the write load
- More flash chips = more aggregate write cycles
- Given a workload, you can compute lifetime

Surprise #2

- <u>SSD lifetime varies with workload</u>
  - Reads vs. writes
  - Random I/O vs. sequential I/O
  - FTL efficiency: write-amplification varies
- Rotating disk lifetime is time-based

Log-structured disk buffer

2X Lifetime
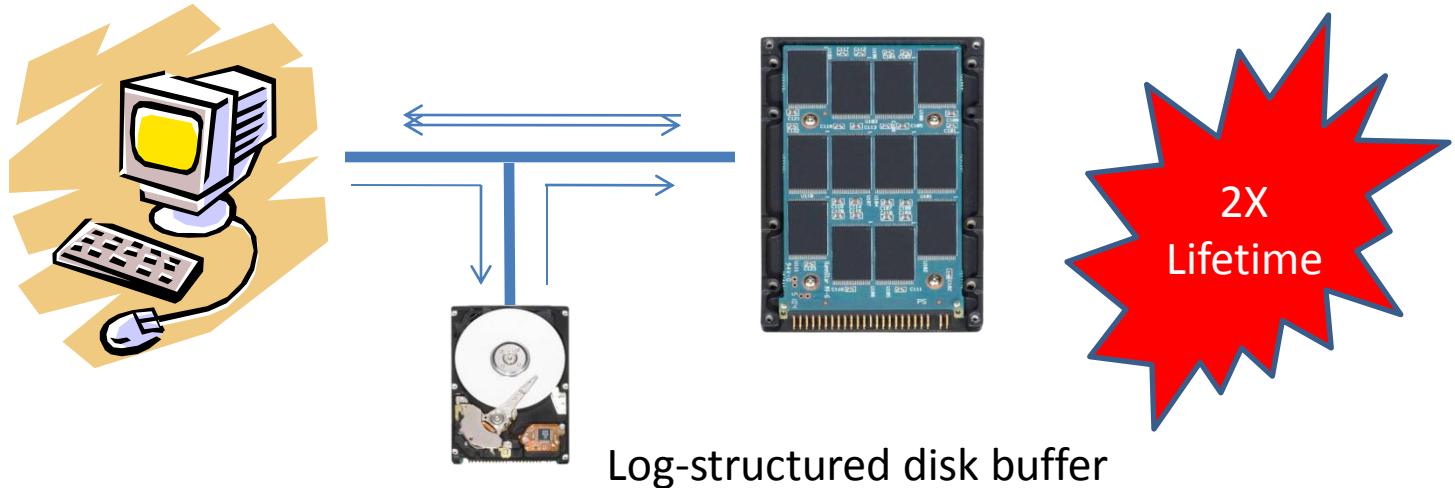
Surprise #3

- Forget about the "R" in SSD-RAID
  - a clever RAID5 of SSDs will load-balance writes
  - intent is to distribute parity-bits
  - so ... SSDs will all fail at same time
  - not optimal for long-term redundancy
- Greater variance in rotating disk failures

- Better to distribute write load unevenly?
- Better yet ... redundancy at flash chip level

The End

- I'm out of surprises.
- Questions?