The SSD Endurance Race: Who's Got the Write Stuff?

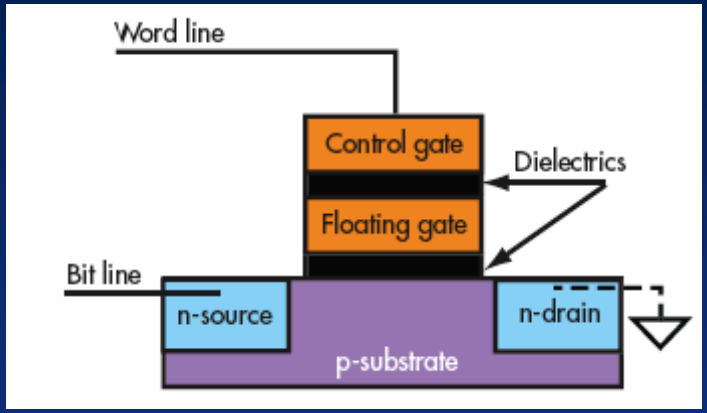
Ulrich Hansen

Director of Market Development

Santa Clara, CA
August 2012

Write Endurance – Why it Exists …

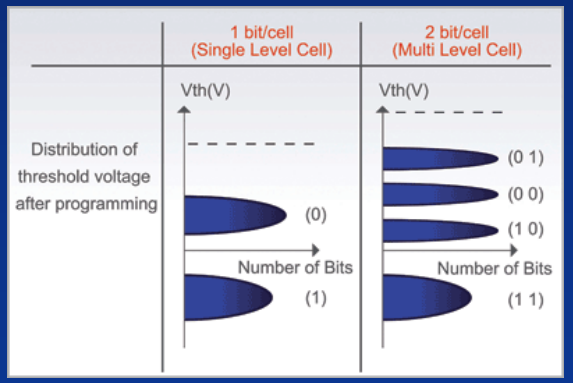
- Programming and erasing of a typical NAND cell involves forcing charge carriers through the dielectric isolators onto the floating gate via a tunneling effect
- This introduces ‘wear’ in the dielectric materials, e.g. by trapping / detrapping of charges in the materials and other defects



- With more and more program and erase activity, ‘wear’ decreases the isolating properties of the dielectric and changes its tunneling effectiveness, ultimately resulting in:
  - Incorrect reads from the NAND cell after relatively short retention times
  - Program / erase failures

NAND Component Endurance: Program & Erase Cycles

- NAND component endurance is specified an number of Program/Erase cycles under certain specifications and testing conditions
  - Meet Bit Error Rate under certain controller ECC capability assumptions and data retention specifications
  - Most test conditions are defined in JEDEC standards
- NAND Flash types and factory optimizations can create various endurance levels
  - NAND Factory optimizations include trimming, settings, and sorting & binning
- Typical P/E specifications for 3Xnm and 2Xnm NAND generations
  - SLC: 100K – 200K P/E cycles
  - eMLC: 10K – 30K P/E cycles
  - Compute-Grade MLC: 2,000 – 5,000 P/E Cycles
  - 3 Bit / Cell NAND: 200 – 500 P/E Cycles



- NAND Physics: Delivering the P/E cycling capability gets more challenging with smaller lithography -> fewer electrons on floating gate to distinguish a cell's levels

SSD Write Endurance – A Definition

- Random Drive Writes per Day for 5 Years (DW/D)
  - 'Drive Write': amount of host write data equivalent of the drive's capacity; in GB
- An equivalent metric is 'Lifetime Random PB Written' (PBW):
  - *Endurance in PBW =*

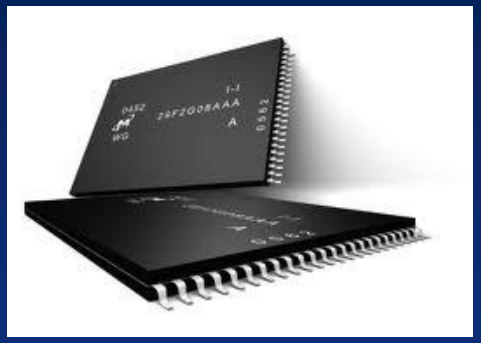
    *Endurance in DW/D * Drive Capacity in PB * 365 Days * 5 Years*

- Conditions are vendor-specific, example for HGST's SSDs:
  - Random workload definition: IO size of 4KB or 8KB, 4K-aligned, full-volume random (access is uniformly distributed across the full drive volume)
  - Drive is 100% full – all LBA used, no LBA Trimmed / Unmapped
  - Data is completely random, e.g. not compressible
  - End-of-life, power-off data retention of 3 months in 40DegC storage temperature
  - UBER of 1 block in 10E16 Bytes read
  - Allowed performance degradation at end of rated product life <10%

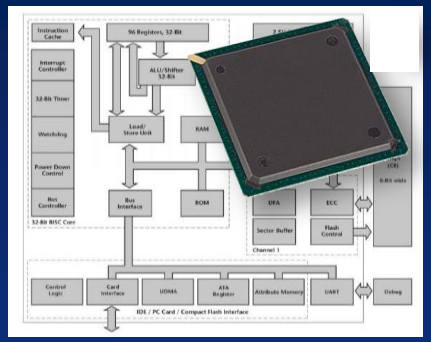
The definition of the endurance specification varies across vendors – you need to make sure you are comparing equivalent definitions

SSD-Level Endurance: Controller and NAND Over-Provisioning

NAND Components



Controller



SSD

Endurance Metric:
Number of P/E Cycles

Typical Conditions:
- Controller ECC Capability
- Data Retention

JEDEC Test Conditions

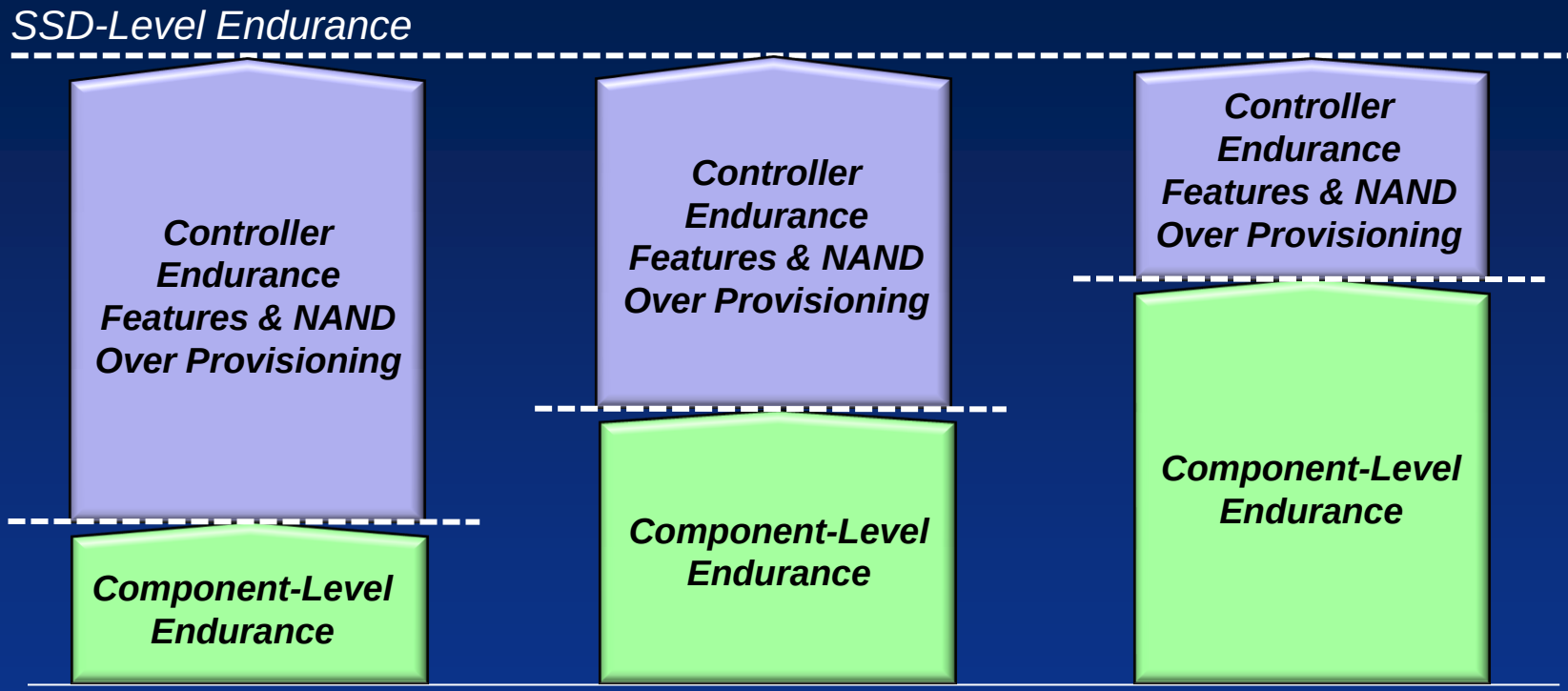
Controller & drive design implement endurance management:
- *ECC*
- *Wear Leveling*
- *Data Refresh*
- *Data Redundancy*
- *Over-Provisioning*
- *Dynamic Read Trim / Read-Retry and/or Dynamic Write Trimming*

Endurance Metric:
Amount of Data Written to the SSD

Endurance Conditions:
- Write workload definition
- Effective Over-Provisioning / Drive 'Fill-Level'
- Data Retention
- Drive UBER
- Date Entropy (Compressibility)

SSD-Level Endurance: Design Approaches

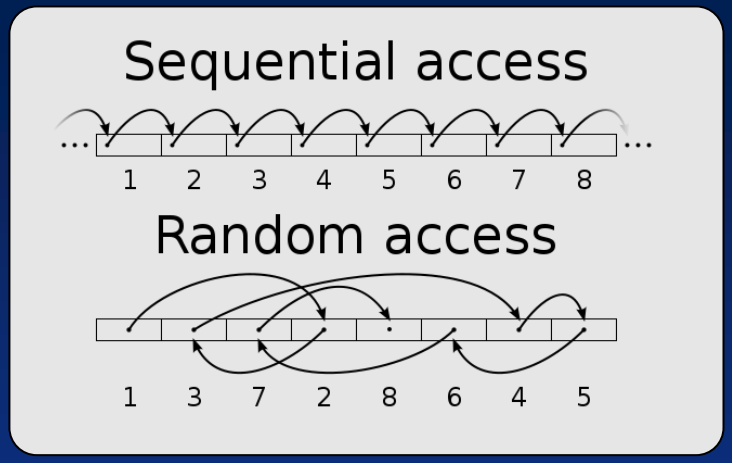
Illustrative



Different design approaches yield equivalent SSD endurance points – SSD vendors typically select a design based on their NAND component access and other design trade-offs (performance, cost)

Random vs. Sequential Write Endurance

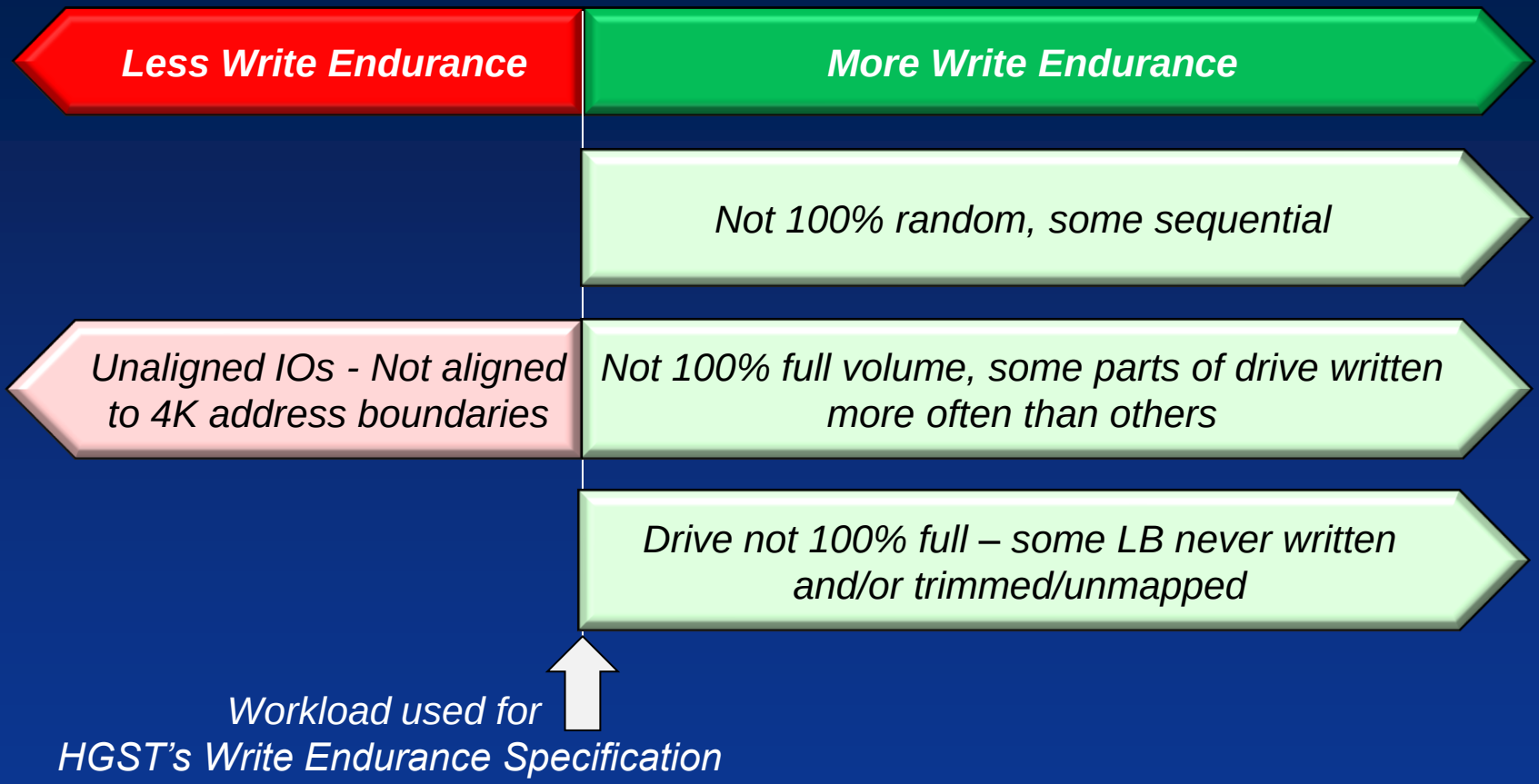
- Due to random write amplification, sequential write endurance is typically higher than random write endurance
- The larger the random WA of a particular SSD design, the larger the difference between the pure random and the pure sequential write endurance



- But keep in mind that sequential write throughput is typically also higher than random write throughput, so endurance can be consumed more quickly as well

Note: Unless noted otherwise, 'Write Endurance' refers to random write endurance in this presentation.

Real Life Write Workload Considerations



Depending on the actual write workloads, observed endurance can be significantly higher than HGST's specification

'Time to Wear-Out'

Time-to-Wear Out (Years) =
*(Endurance in DW/D * Drive Capacity in GB * 5 * 1000) /*
*(Average Write Throughput in MB/s * 60 * 60 * 24)*

HGST Ultrastar SSD400S.B

Drive Capacity (GB)	Random Write Endurance (DW/D)	Max Sustained Random Write TP (MB/s)	Time-to-Wear-Out @ 100% Max Rand Write TP (Years)
100	50	104	2.8
200	50	104	5.5
400	50	104	11

HGST Ultrastar SSD400M

Drive Capacity (GB)	Random Write Endurance (DW/D)	Max Sustained Random Write TP (MB/s)	Time-to-Wear-Out @ 100% Max Rand Write TP (Years)
200	10	98	1.2
400	10	98	2.4

'Worst-case' numbers for 'time-to-wear-out' are typically a poor approximation of real-life Enterprise SSD use

How Much Write Endurance is Enough?

'Worst Case'	'Typical Case'	'Time-to-Wear-Out' Multiplier
- 100% Write Mix & Max Write Throughput (i.e. 100% very high QD)	- Avg. of 30% of Max Write Throughput	3.3x
- 100% Duty Cycle	- 80% Duty Cycle	1.25x
- HGST Workload Spec - 100% Random - Full-Volume Random - Drive 100% Full	- Typical workload: - Some sequential writes - Some write data locality - Drive not 100% full – TRIM/UNMAP	1.25x
		5.2x

'Time-to-Wear Out' Examples: HGST Ultrastar SSD400M

200GB:	1.2 Years	~ 6.2 Years
400GB:	2.4 Years	~ 12.5 Years

End of Rated Product Life – What to Expect

- Depends on SSD product design: One or more of the stated specifications will not be met any longer if SSD is used beyond the rated product life:
  - Power-off data retention
  - Read UBER
  - Write performance
    - As more NAND is retired due to program / erase failures, effective over-provisioning is reduced
  - Read performance
    - E.g. as more read ‘re-try’s are needed
- SSD is not likely to fail catastrophically immediately

Use SSD SMART statistics to monitor rated life use

Qualifying Drive Endurance – Best Practices

- Qualifying the SSD endurance specification is a time-consuming and costly effort
- Need to test large populations in order to get statically relevant data, especially when validating 'rare event' specifications like UBER
- Need to validate the actual end-of life behavior of the SSD
  - Utilize specially configured, very small capacity SSDs to bring the SSD to wear-out point in ~2 months and validate all specifications
  - Also ensures end-of life drive firmware paths are executed and tested thoroughly

Leading SSD vendors have built significant experience and know-how related to endurance validation testing over time

Write Endurance – Industry Direction and Summary

- Endurance specification standardization efforts have not seen wide industry adoption yet
  - Similar challenges as performance specification standardization …
- Common points of data sheet specifications seem to emerge
  - 50, 25, 10, 3 DW/D for Enterprise, <1 DW/D for Client
- However, the underlying endurance specification assumptions and conditions vary by SSD vendor
- Also, the effort and expertise put into endurance validation testing varies significantly by SSD vendor

Continued customer education is needed, and SSD buyers are well advised to be diligent when comparing endurance specifications across different SSD vendors