

System Reliability



Gary Kotzur – OCTO Storage Architecture

July, 2008

Enterprise Reliability

Goal: No data loss
under any condition!



Aspects of SSD Reliability

Metric	Rating
MTBF	Very good
BER	Very good
Write Endurance	Varies
Data Retention	Needs better understanding
Flash Infant Mortality	Needs to be addressed
Environmental	Excellent
Power loss data protection	Varies
In-flight data protection	Needs to be addressed
Monitoring	Needs to be addressed



MTBF

- Today SSDs advertise a $MTBF \geq 2.0M$, greater than a SAS drive
- Methods to measure MTBF vary between a SSD and a HDD, need to use consistent approach
- Concern over some SSDs using striping to improve performance. Using MTTDL calculation from RAID:
 - RAID0 : $MTTDL = MTBF_{Drive} / N$, N = number of drives
 - RAID5 : $MTTDL = MTBF_{Drive}^2 / N(N-1)$



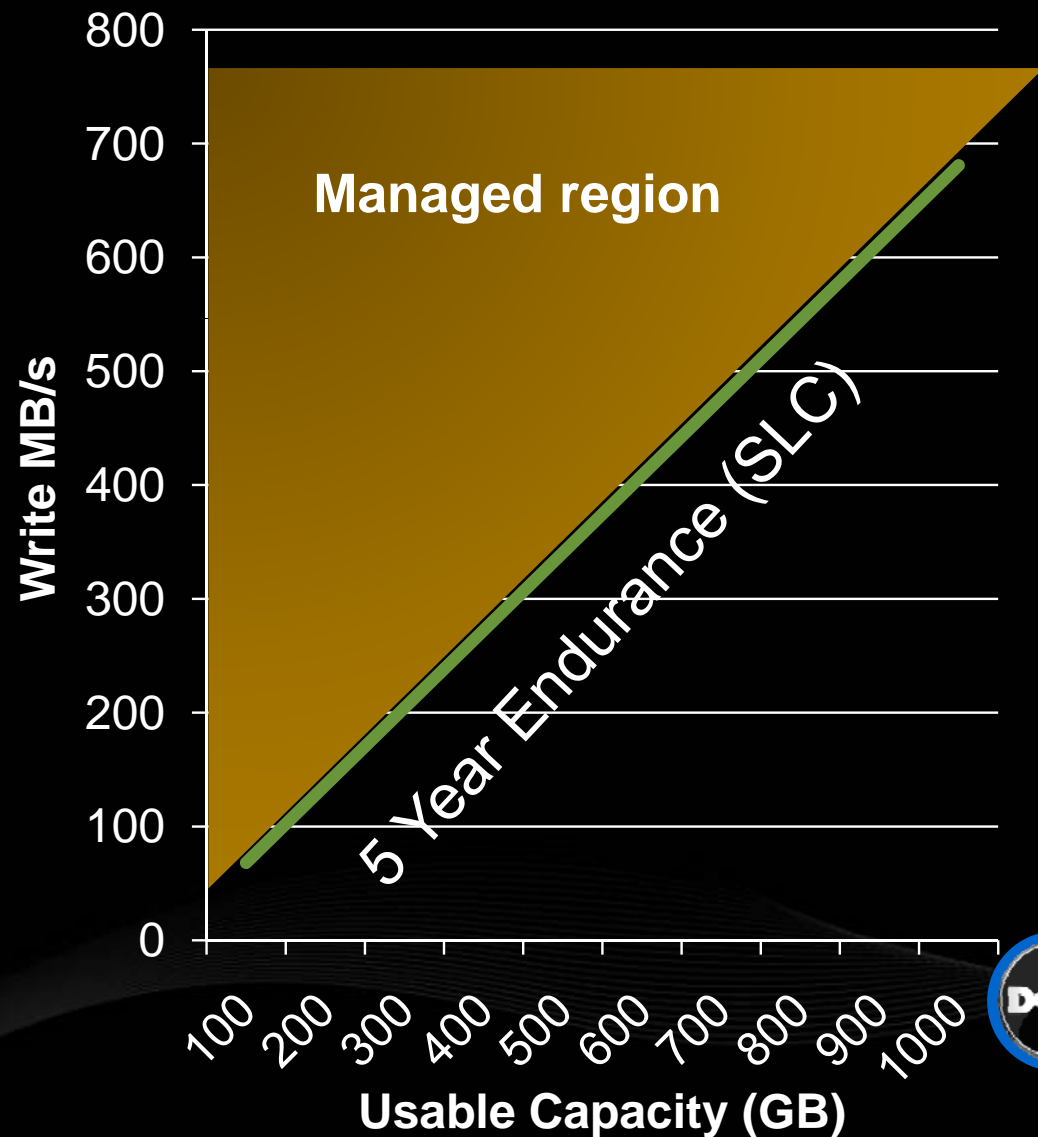
BER

- Today SSDs advertise a BER $\geq 10^{-16}$, which is equal to a SAS drive
- This is fine for “drive replacements”. Not sure what the requirements for future usage models will be.



Write Endurance

- Flash memory is “write limited”
- SSD vendors use wear-leveling to eliminate hotspots
 - Larger capacities improve write endurance
 - Higher bandwidths decrease “write life”
- High performance SSDs must be managed
 - Spare blocks
 - Throttle BW
 - User notification



Write Endurance (cont)

- Future geometry shrinks
 - Will cause write endurance to be reduced
 - SLC from 100K → 50K
 - MLC from 10K → 5K and lower
 - Flash density increases will move the lower capacity bound to meet performance requirements
- Need lifetime models for SSD products that take flash write endurance behavior as an input.
- Use models to determine minimum write endurance needed from flash memory suppliers.
- Drive that write endurance as an industry requirement vs. on an individual SSD supplier basis.



Data Retention

- Drive powered “on”
 - Read Disturbs – bits flipped on surrounding pages
 - Program Disturbs – bits flipped on surrounding pages
 - Partial page program – active and surrounding pages can have bits flipped
 - Currently protected by ECC and scrubbing
- Drive powered “off”
 - Cell voltage drifts, can cause bit flips
 - Worsens over drive use
 - Protected by ECC + ECC “margin”



Flash Infant Mortality

- Today if a flash device within a SSD fails, it is treated as a head crash
 - Drive is declared “failed”
 - Replacement is costly
- Need to find better approaches



Environmental

- Temperature ranges of SSDs are better than HDD
 - HDDs are usually limited to 60°C
 - SSDs can run up to 70°C, can be extended to 85°C
- SSDs can tolerate humidity better than HDD
 - HDD: 20%-85% RH humidity, non-condensing
 - SSD: 90%-98% RH humidity, non-condensing
- SSDs are more “rugged” than HDDs
 - HDD: 100G @ 1/2ms sine wave
 - SSD: 1500G @ 1/2ms sine wave



Power Loss Data Protection

- Enterprise system
 - RAID controller uses a cache to temporarily store data until it can write it to the drive(s). This cache is battery-protected
 - All drives use a cache to enhance performance, this cache be turned on or off
 - By default, SATA drive's cache is on
 - By default, SAS drive's cache is off
- SSD architecture
 - If cache setting is off, small data loss of “write in progress” is tolerated
 - If cache setting is on, no data loss is tolerated



In-Flight Data Protection

- Today's Enterprise system
 - RAID controller various means to protect data
 - ECC on memory
 - Internal data path protection
 - Parity across drives
 - SAS and PCIe use CRC
 - Drives use CRC and ECC
- Future Enterprise systems are moving to DIF for end-to-end data protection



Monitoring

- Currently SMART is being extended through the T13 and T10 forums to address new “behaviors” unique to SSDs
- Need to ensure that minimum requirements are met, plus need room for advanced features
- Need to drive consistency of parameters, today different SSDs respond differently on the same command





Dell

Enterprise Systems