JEDEC SSD Endurance Workloads

Alvin Cox, Seagate

Chairman, JC-64.8

Workloads & SSD Applications

Endurance is dependent on the applications being run and the workloads presented to an SSD

Client SSD

- 8/7 operation, bursty workload (considerable relaxation times)
- Modest write activity
- More localized access footprint, smaller than entire LBA space
- Optimized cost

Enterprise SSD

- 24/7 operation, continuous workload
- Heavy write activity
- Mostly random accesses across the entire LBA space
- Highest data reliability, availability, and integrity

Endurance ratings must reflect the expected use cases. JEDEC workloads and testing are targeted at achieving this

JEDEC SSD Standards

- **JESD218A**: *Solid State Drive (SSD) Requirements and Endurance Test Method*
  - SSD Definitions
  - SSD Capacity
  - Application Classes
  - Endurance Rating
  - Endurance Verification

- **JESD219**: *Solid State Drive (SSD) Endurance Workloads*
  - Client
  - Enterprise

Endurance Rating (TBW Rating)

- Establishes a rating system for comparing SSDs
- Provides ratings based on application class
- Rating for user-measurable interface activity

This rating is referred to as TBW. Requirements for UBER, FFR, and retention are defined for each application class.

Endurance Rating (TBW Rating)

The SSD manufacturer shall establish an endurance rating for an SSD that represents the maximum number of terabytes that may be written by a host to the SSD, using the workload specified for the application class, such that the following conditions are satisfied:

1) The SSD maintains its capacity

2) The SSD maintains the required UBER for its application class

3) The SSD meets the required functional failure requirement (FFR) for its application class

4) The SSD retains data with power off for the required time for its application class

Application Class	Application Class Attributes
• Client • Enterprise	• Workload • Daily Active Use • Data Retention • BER

SSD Endurance Classes And Requirements

Application Class	Workload	Active Use (power on)	Retention Use (power off)	Functional Failure Rqmt (FFR)	UBER
Client	Client	40°C 8 hrs/day	30°C 1 year	$\leq 3\%$	$\leq 10^{-15}$
Enterprise	Enterprise	55°C 24hrs/day	40°C 3 months	$\leq 3\%$	$\leq 10^{-16}$

From JESD218A, Copyright JEDEC. Reproduced with permission by JEDEC

- Program/erase (P/E) cycles

  - The writing of data to one or more pages in an erase block and the erasure of that block, in either order.

  - NAND has a limit on how many P/E cycles it can withstand until data retention is not reliable. UBER and data retention are adversely affected by P/E cycles.

Program **Erase**
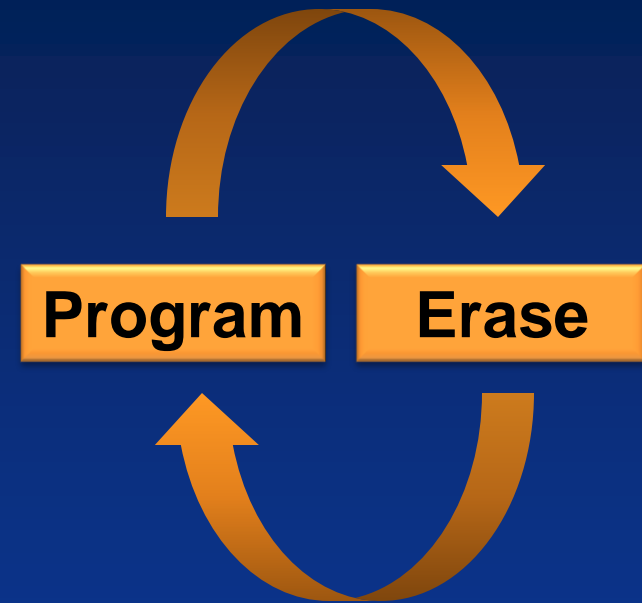
- **UBER (Unrecoverable Bit Error Ratio)**
  - UBER could be considered a short-term data retention measure although it is measured over the life of the SSD

- **Retention failure**
  - A data error occurring when the SSD is read after an extended period of time following the previous write. This can be considered a long-term data retention measure

WAF – A Key Endurance Factor

The higher the WAF, the faster the Flash wears out

- **Write Amplification Factor (WAF)**
  - The amount of data written to the NVM divided by the amount of data written by the host to the SSD

- WAF is directly associated with P/E cycles
  - Higher WAF = More P/E cycles happen

- An SSD usually writes more data to the memory than it is asked by the host to write

- The nature of the workload has a significant impact on WAF

- Workload factors that impact WAF:
  - Sequential versus random
  - Large transfers versus small transfers
  - Boundary alignment
    - Transfer size vs program page size/alignment
    - Transfers crossing erase blocks
  - Data content/patterns (especially for SSDs using data compression)

Workload Comparison

Client workload

- Based on actual client trace
- Real trace commands replayed
- Includes trim commands

- Does not touch every user LBA

- Full random data pattern to stress write amplification factor – required
- Non-random pattern - optional
  - Random selection of data from a confirmed entropy data file
  - Reported in addition to full random results

Enterprise workload

- Similar to SPC1 profile
- Synthetically generated by script
- Does not include TRIM or UNMAP commands
- Touches every user LBA resulting in 100% full utilization
- Full random data pattern to stress write amplification factor

- **The client workload consists of the following:**

  - **Precondition phase**
    - **Write all user LBAs**
    - **Testing at 100% full still under discussion**

  - **Run the test trace**
    - **Based on standard ATA I/O commands**
    - **Adjusted to stay within the SSD user capacity**
    - **Includes all commands from a real trace capture**
    - **The trace only has commands, including TRIM**
    - **Data payload is generated separately as random data**

  - **Replay the test trace to verify TBW rating**

SPC-1C Enterprise Workload

SPC-1C…

- **Is a set of I/O operations designed to demonstrate the performance of a small storage subsystem while performing the typical functions of a business critical application**

- **Represents a segment of applications characterized by predominately random I/O operations and requiring both queries as well as update operations**

- **Focuses on small storage solutions 1 to 24 drives**

- **Uses 100% 4K aligned transfers**
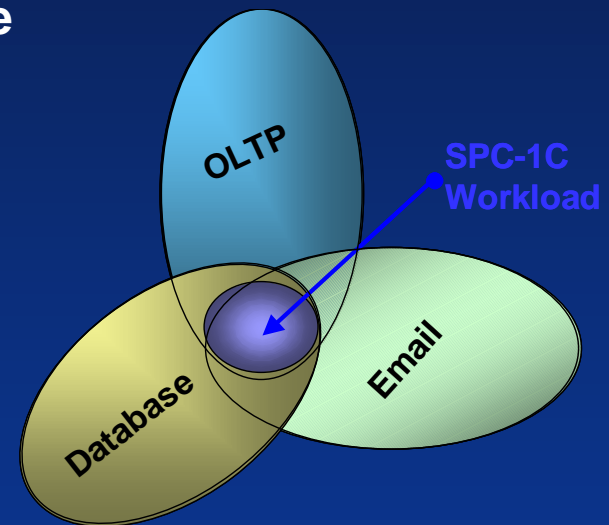
- Adjusted SPC-1C transfer size distribution to include 10% <4K transfers
  - Trace data indicated many applications have some non-alignment characteristics
- Random data is used for the data payload

512 bytes (0.5K) 4%
1024 bytes (1K) 1%
1536 bytes (1.5K) 1%
2048 bytes (2K) 1%
2560 bytes (2.5K) 1%
3072 bytes (3K) 1%
3584 bytes (3.5K) 1%
4096 bytes (4K) 67%
8192 bytes (8K) 10%
16,384 bytes (16K) 7%
32,768 bytes (32K) 3%
65,536 bytes (64K) 3%

- **Several workload studies show that less than 5% of the data get >50% of the accesses and 20% of the data get >80% of the accesses**

- **Distribution:**
  - **SSD under test:**
    - **50% of accesses to first 5% of user LBA space**
    - **30% of accesses to next 15% of user LBA space**
    - **20% of accesses to remainder of user LBA space**
  - **Distribution is offset through the different DUTs.**
  - **Address distribution to simulate application usage except making contiguous simplifies test apparatus requirements**
    - **Each segment access are randomized**
    - **Segment accesses are intermixed**

- **Enterprise and Client SSDs have different design points, application uses and workloads**

- **Different endurance rating methods are required for the for the different use cases**

- **JESD218A and JESD219 define the methods to derive a means of comparing the endurance ratings of devices within each class of device**

Thank You