Enterprise SSD Competitive Analysis

Gary Tressler and Dean Sciacca
IBM Corporation

Agenda

- Introduction
- Methodology
- Performance
- Response Time
- Power Consumption
- Write Amplification
- Life Span
- Summary

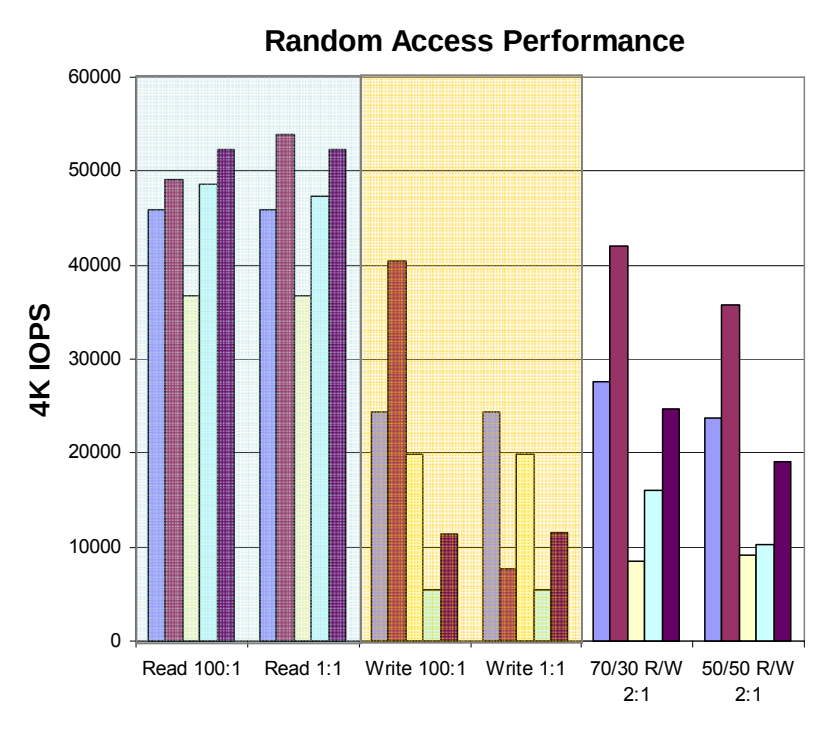
Introduction

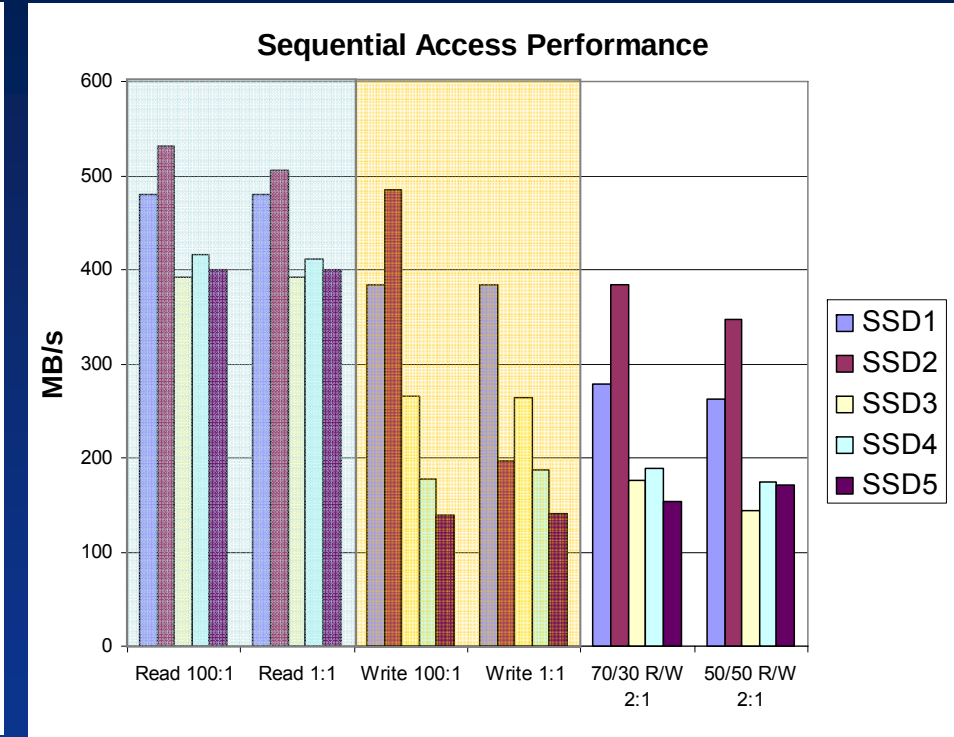
- IBM is driving MLC SSD adoption for Enterprise storage
- Five industry Enterprise MLC SSDs were characterized
  - SATA3 or SAS2 interface (6Gbps)
  - 200 or 400GB usable MLC Flash capacity
- SSDs evaluated were prototype or engineering samples
  - Firmware may not have been fully optimized

Enterprise SSD Characterization Methodology

- Characterization Platform
  - PC with high performance CPU, DDR3 DRAM
  - Windows 7 (64-bit) operating system
  - LSI 9212 host bus adapter
- Apply VDBench 5.03 Exerciser
  - SSDs in raw mode (no file system)
  - Use 4KB aligned writes/reads (512B mode)
  - 4KB transfer size for random workloads
  - 128K transfer size for sequential workloads
- Preconditioning
  - Initial 24-hour write random/sequential workload
  - Each individual measurement has its own preconditioning cycle
    - Preconditioning durations are based on the time necessary for SSD performance to stabilize against a given workload
  - Each measurement is the average performance for an extended interval

Enterprise MLC SSD Performance

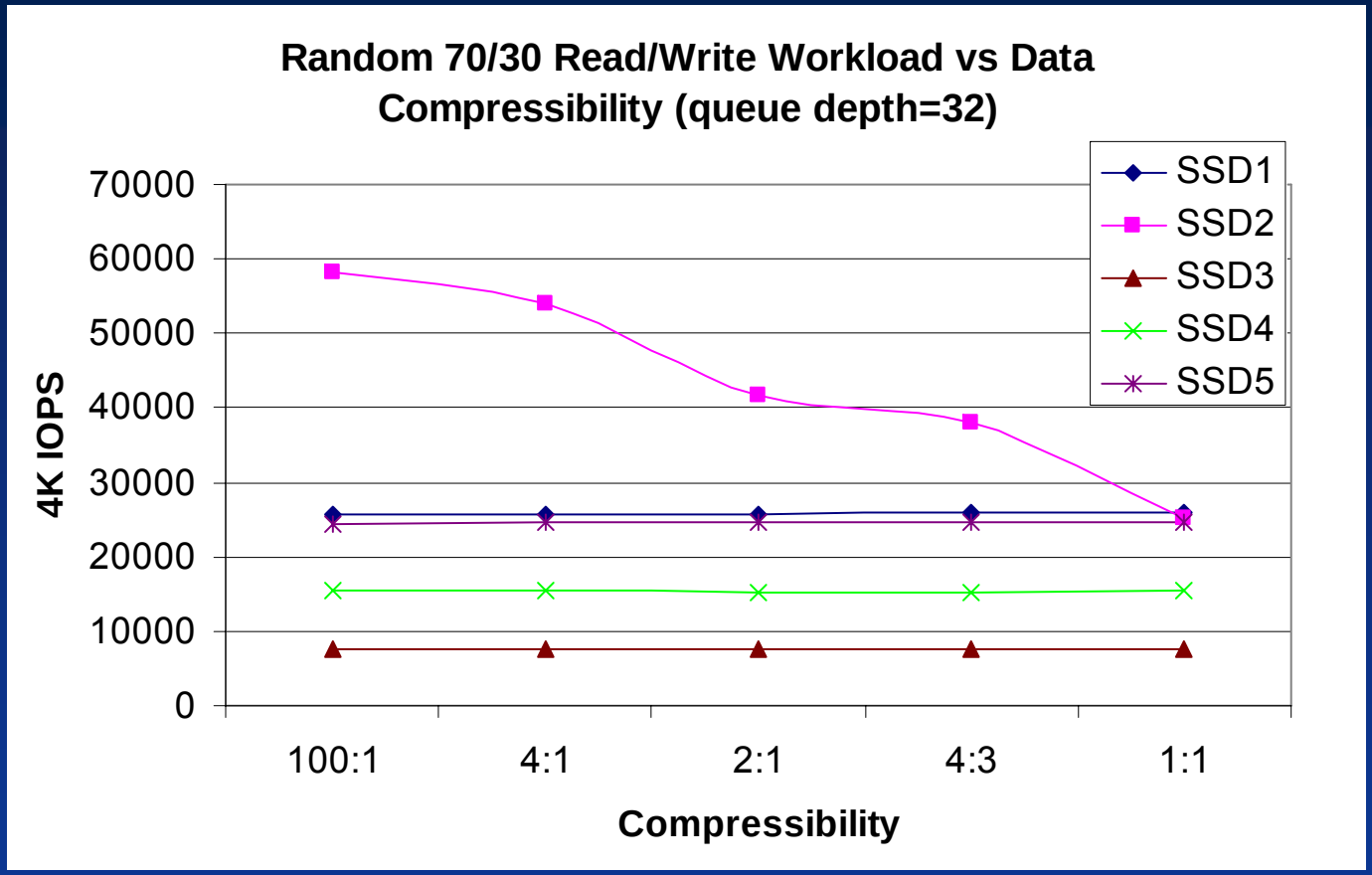




- *Random reads observed in the 35-55KIOPs range, and random writes in the 5-40KIOPs range*
- *Sequential reads observed in the 400-525MBps range, with sequential writes in the 150-480MBps range*
  - *Notable performance improvements observed over prior generation*
- *Mixed read-write mode performance not optimized in certain cases*

Note: Data compression ratios defined for each test (e.g.: 100:1, 1:1 and 2:1)

SSD Random Performance vs. Data Compressibility

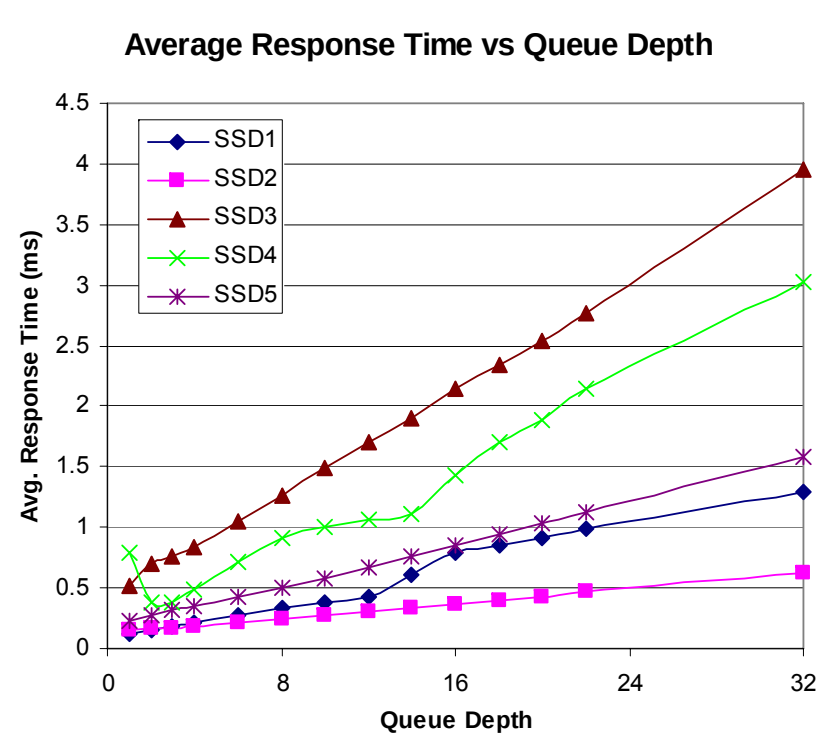


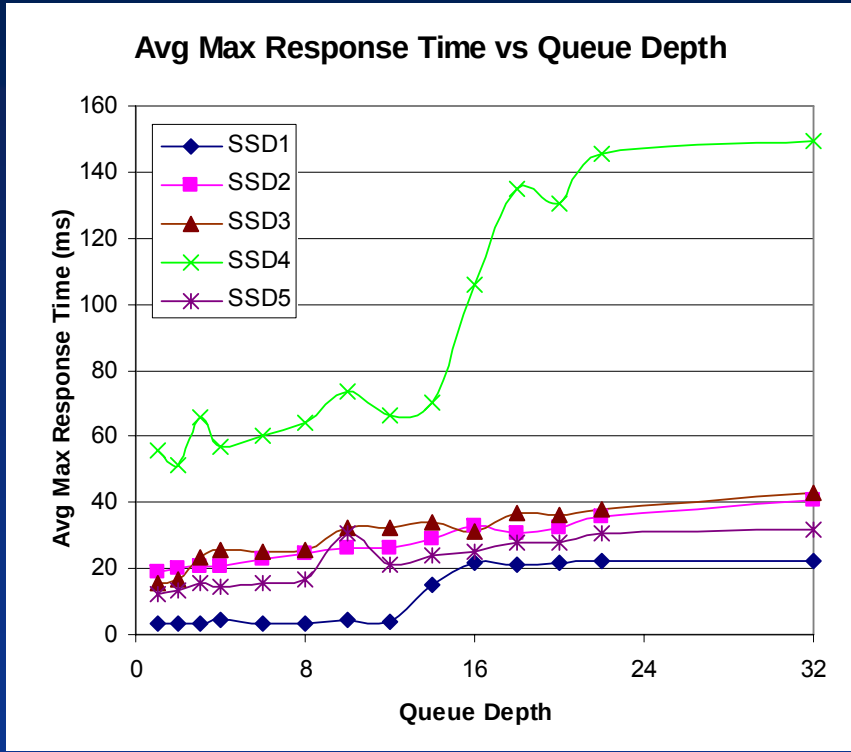
- *Most SSDs characterized showed consistent performance vs. data compressibility*
  - *Data compression techniques now under widespread development across industry*

SSD Response Time Characterization

- SSD response time is critical for overall Enterprise system performance and customer satisfaction
- A transaction type workload is applied
  - Mixed random read and write, 4KB and 8KB
- Response time is measured as a function of queue depth
  - Average maximum response time is the average of the maximum response times in thirty 10-second intervals
- Competitive response times require optimization of latency and frequency of various SSD background operations
- Response times are not adequately specified
  - Average read/write and average maximum read/write response time parameter specifications would provide users with valuable information

Average and Average Maximum Response Time vs. Queue Depth





- *Large variation in average response times observed – further optimization required*

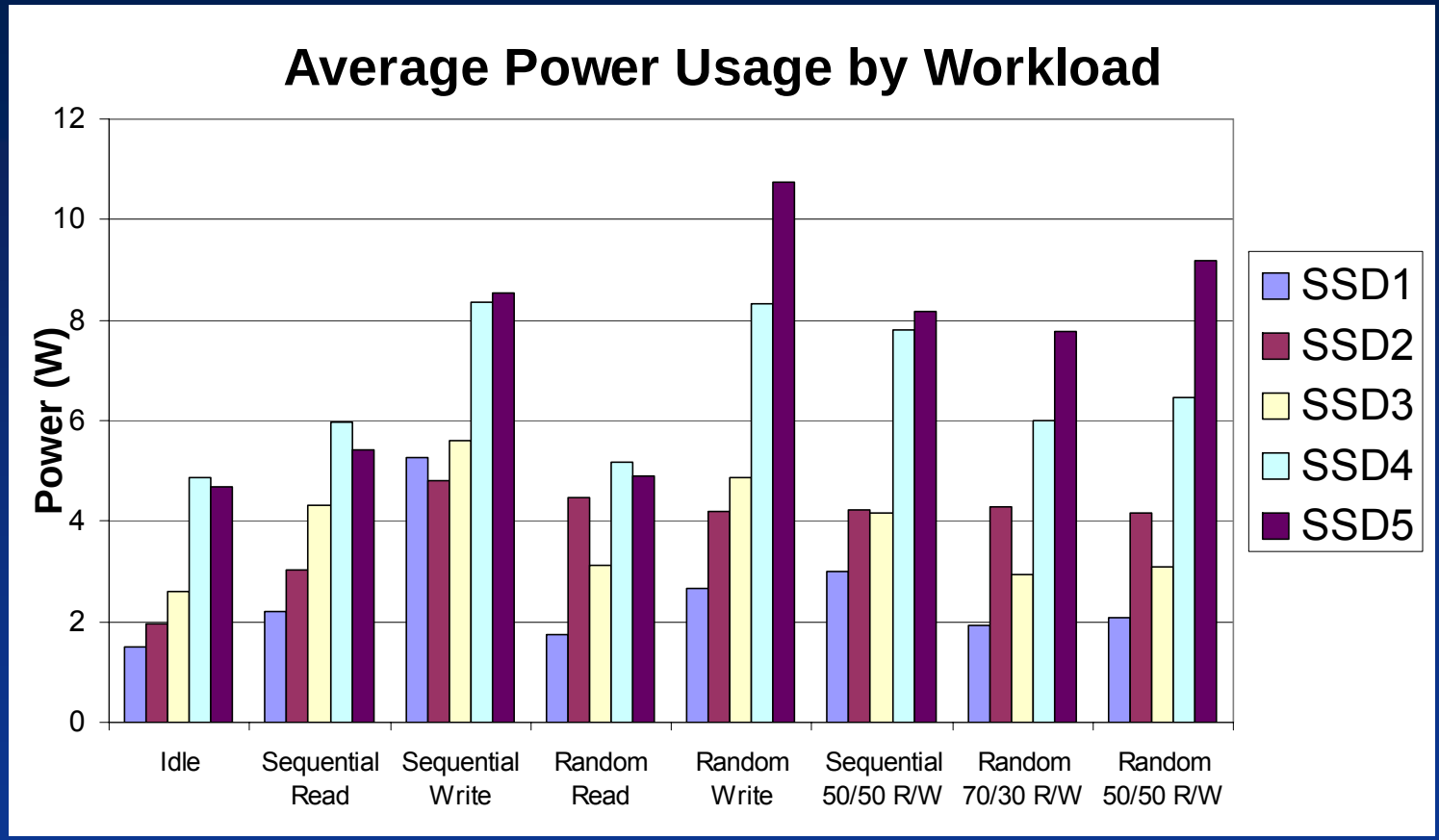
- *Frequency and latency of SSD non-data operations are key to reducing average maximum response times*

Note: Data compression of 5:1 applied

Enterprise SSD Power Consumption

- Power consumption is an increasingly important SSD attribute
  - Existing system requirements limit available power
  - Heat generation reduces intrinsic component reliability and can cause SSDs to fault
- SSD power is rising with improving SSD performance
  - Primarily due to higher Flash bandwidth and increasing number of active Flash die during write operations
  - Ongoing challenge require innovation
    - Thermal interface materials
    - Advanced throttling techniques

Enterprise SSD Power Consumption by Workload



- *SSD write power generally 1.5X higher than read power*
- *High idle power observed in certain cases*

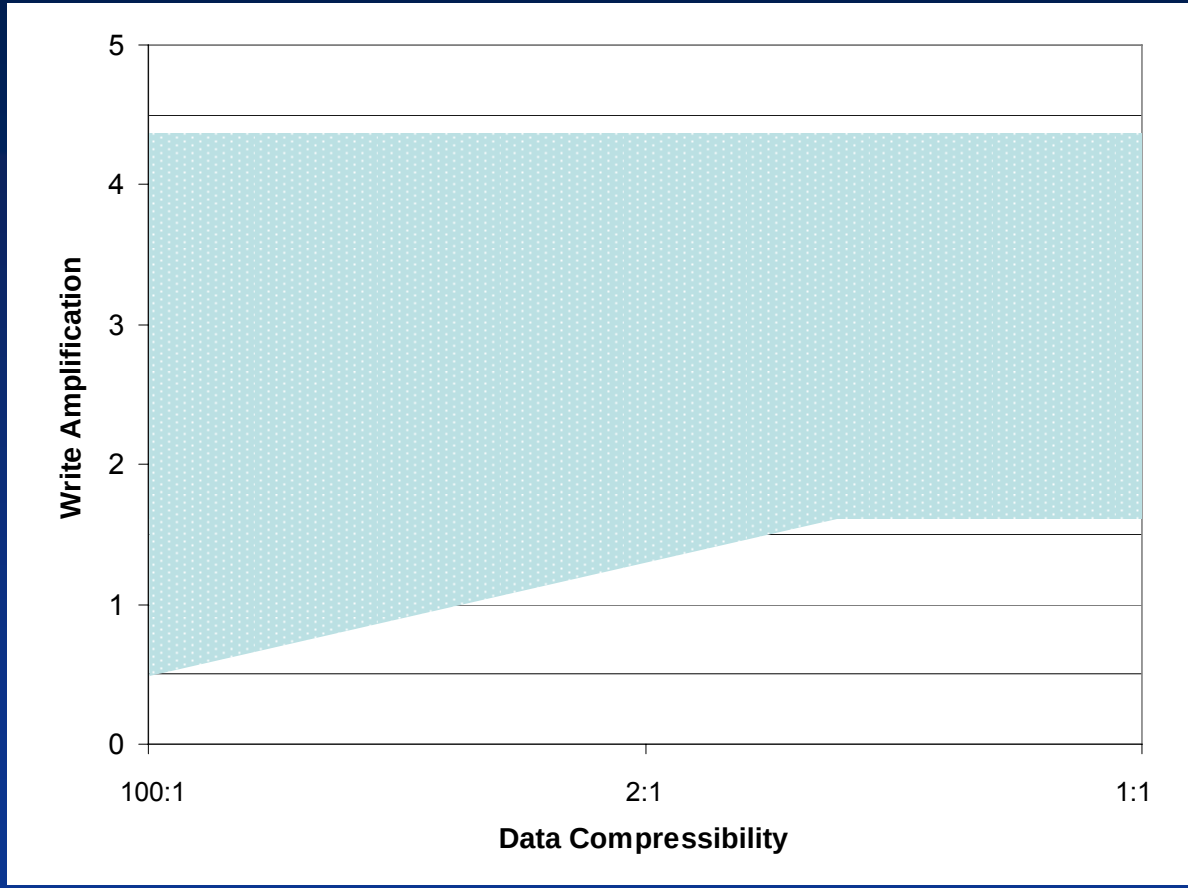
Note: Data compression of 2:1 applied

SSD Write Amplification

- SSD write amplification (WA) is the ratio of amount of data written to Flash to the amount of write data requested by the host
  - Enterprise MLC can sustain an order of magnitude more program-erase cycles than standard consumer grade MLC
    - Example - 30K or greater Enterprise MLC vs. 3K consumer MLC
- WA is a key SSD controller architectural parameter to assess Enterprise SSD lifespan for particular usage cases

$$\text{Disk Life Span} = \frac{\text{P/E cycles} * \text{Drive Capacity(MB)}}{\text{WA} * \text{Write Performance (MB/s)} * \text{Duty Cycle}}$$

Enterprise SSD Write Amplification (WA)



Note: Results reflect random access

- *View of industry Enterprise SSD write amplification status (8 suppliers)*
- *WA is a critical SSD architecture parameter with inverse relationship to life span*

Enterprise SSD Life Span Characterization

- Numerous industry Enterprise SSDs under long term continuous write testing at IBM
  - Effort in early stage – initiated 1Q11
- Approx. 10% of specified SSD usable life evaluated to date
  - Results reflective of specified life span expected by 3Q12
  - Monitoring Flash block retirement vs. time
  - No clear method identified for accelerated testing
- No observed SSD performance changes thus far

Summary

- Suppliers are now developing next generation Enterprise MLC SSDs targeted at 6Gbps (SATA3/SAS2)
  - Notable performance improvements observed over prior generation
- SSD response time is critical metric for overall Enterprise system performance and customer satisfaction
  - Opportunities for improved specifications exist
- SSD power characteristics have increased significantly and will present an ongoing systems challenge
  - Thermal mitigation innovation required
- Write amplification is a key controller architectural parameter to assess SSD life span for particular usage cases
  - With increased performance, write amplification and capacity must scale to maintain life span targets