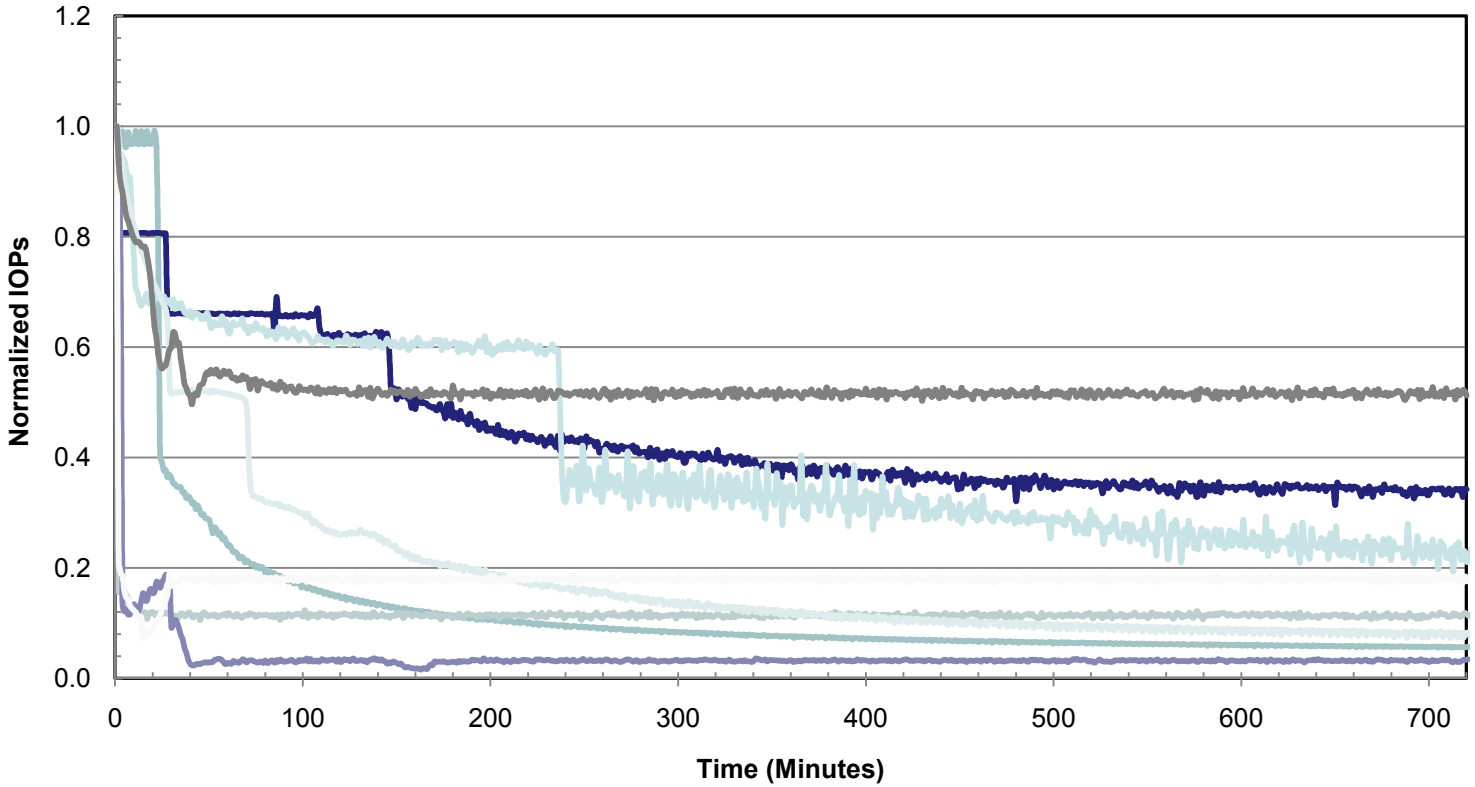
NAND-Based SSDs:
Performance States and Performance Measurement

Doug Rollins
Micron Technology
DougRollins@Micron.com

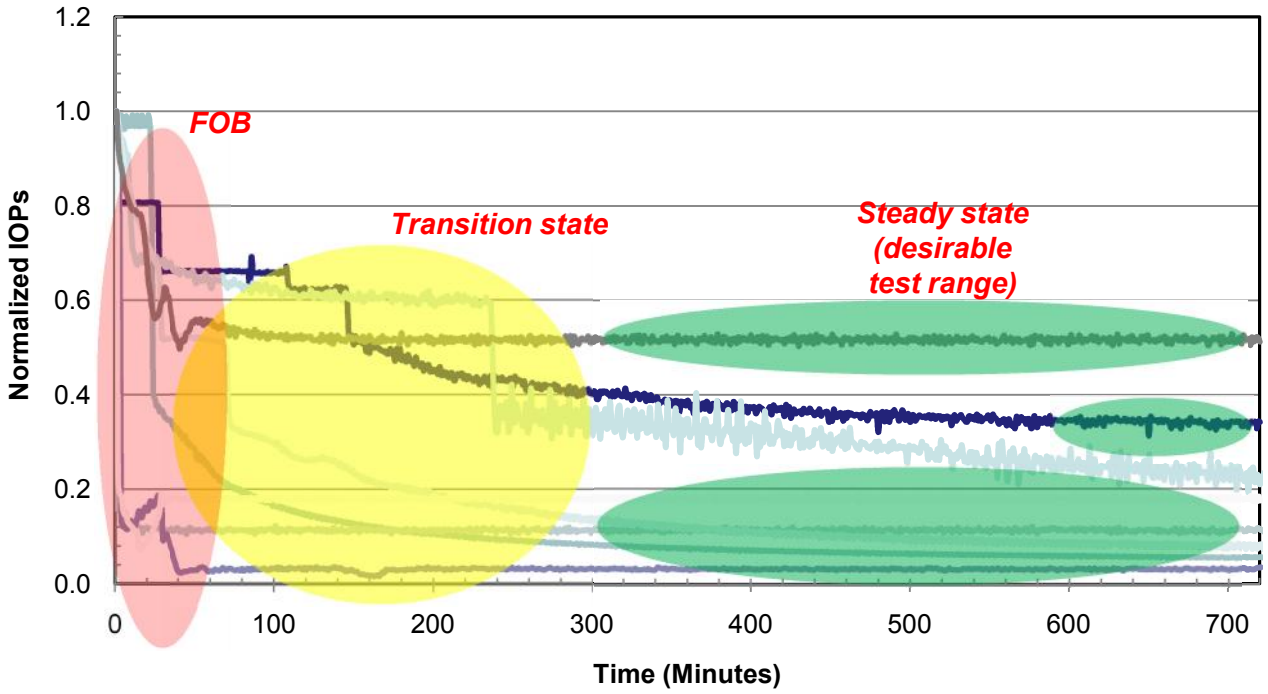
Time-Variant Performance



- Empirical data from several SSDs
- Drives were securely erased, then written with a fixed stimulus
- Amplitude and shape of curves differ from drive to drive

Common Terms

FOB state (red): The FOB state is visible at the extreme left of the plot. This state is reached when the drive has little to no user data; all the NAND cells are erased and available to receive new data. As the drive is written, the IOPs of this state decrease.



Transition state (yellow): Immediately following the FOB state, the drive enters a transition state, marked by steadily decreasing performance.

Steady state (green): Drive performance is as described in the SNIA SSSI TWG PTS*, so this area is a steady state (by inspection).

Note that:

- Different drives reach steady state at different times.
- One sample has yet to reach steady state.

***Storage Networking Industry Association Solid State Storage Initiative Technical Working Group Performance Test Specification**

Stimulus Sequence-Variant Performance

Example drive stimulated with:

128K sequential write (into steady state)

(followed by…)

4K random write (into steady state)

(followed by…)

128K sequential write (into steady state)

Result:

This drive *never recovers* original performance level

Example drive stimulated with:

4K random write (into steady state)

(followed by…)

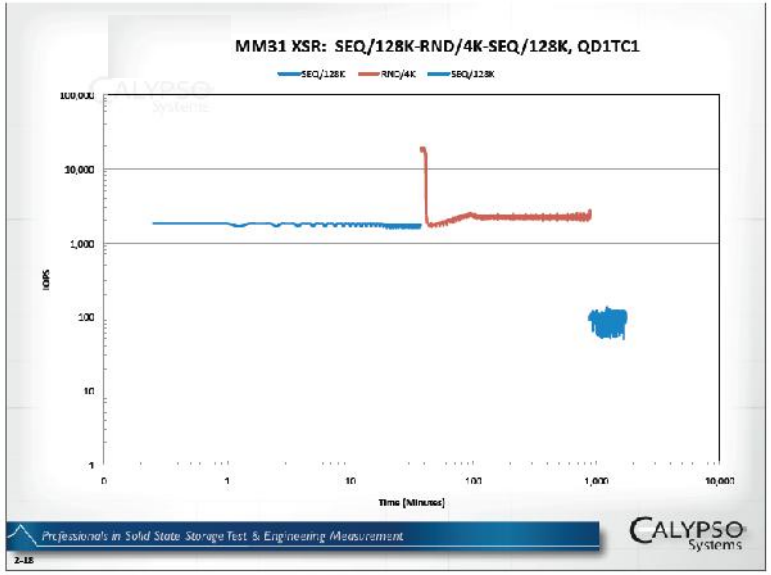
128K sequential write (into steady state)

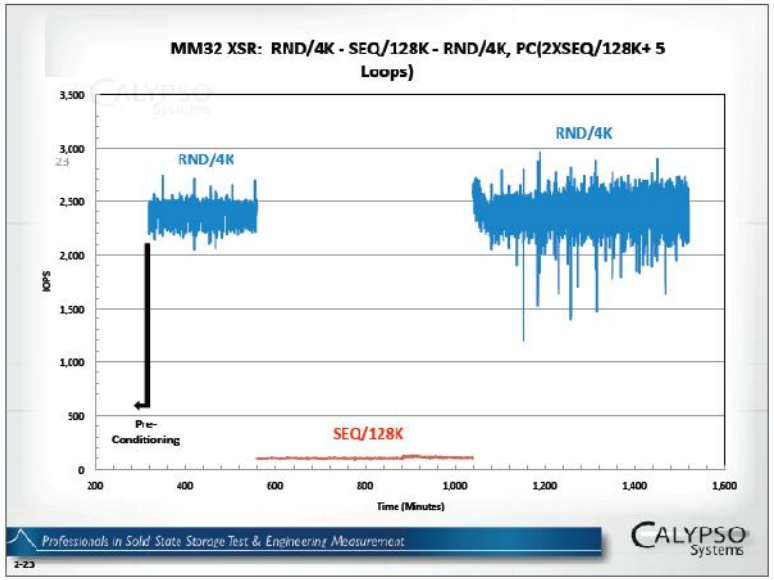
(followed by…)

4K random write (into steady state)

Result:

This drive *does recover* original performance level





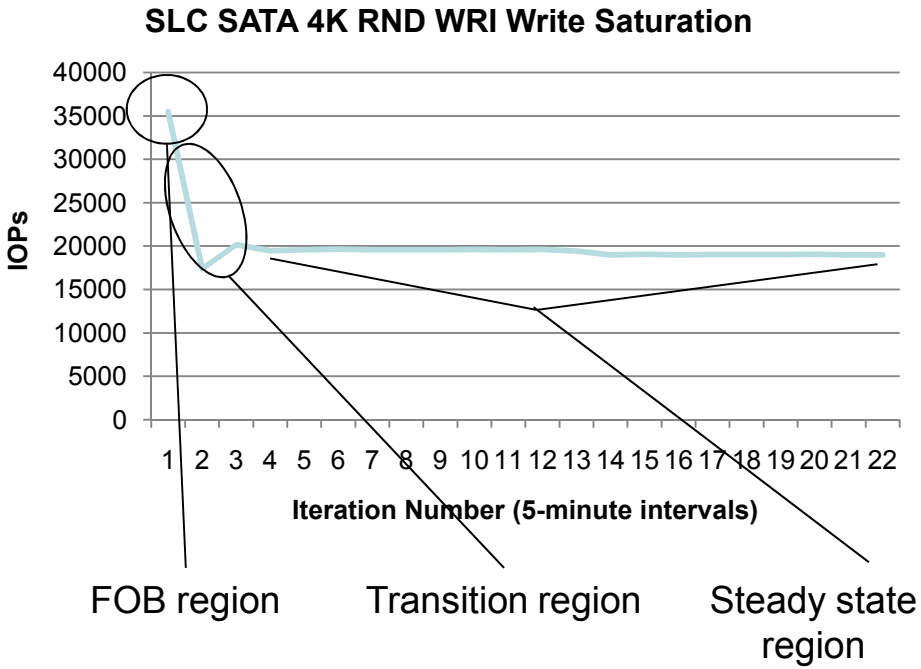
Key Care-Abouts and Assumptions: Enterprise

- **Enterprise assumptions:**
  - Drive is always full
  - Drive is always being accessed
  - Decisions are made on steady state performance
  - Steady state ≠ full drive ≠ worst case
- **Steady state defined (from SNIA PTS):**
  - Make sure the difference between the maximum and minimum performance is close to the average (no wide variance) **AND**
  - Maximum performance (linear curve fit of the data with the measurement window) – minimum performance (linear curve fit of the data with the measurement window) is small (within 10% of the average)
- **Full drive defined:** Drive has been written over some multiple (could be 1X) of the user-accessible LBA space by a fixed pattern that may be invariant from the test stimulus (i.e. 2X user LBA space written with 128K SEQ WRI)
- **Worst case defined:** Drive has been stimulated over some fixed time with a workload ***intentionally designed*** to demonstrate the drive's worst possible performance. Typically this includes:
  - Small transfers mixed with large transfers
  - Intentionally misaligned writes

Consistent Results

- **Always start from a repeatable, known, fixed point:**
  - Examples: purge, secure erase, low-level format
- **Always precondition the drive in the same way:**
  - Workload-independent – 128K SEQ WRI over 2X user capacity, page-aligned
  - Workload-dependent – apply the test stimulus iteratively until steady state is reached
- **Always stimulate the drive with a single, fixed stimulus until steady state is reached:**
  - Write history affects performance
  - Stimuli application order can affect performance
  - Plot results
  - Determine steady state via PTS definition
- **Example test sequence for every stimulus of interest:**
  - Purge, SE, or LLF
  - PC
  - Stimulate until steady state is reached

Example: SLC SATA SSD



Raw IOMeter Data

Iteration Number	Transfer Size	% Read	% Random	IOPs
1	4096	0	100	35460.98405
2	4096	0	100	17382.48189
3	4096	0	100	20179.53067
4	4096	0	100	19460.9479
5	4096	0	100	19568.64219
6	4096	0	100	19657.92814
7	4096	0	100	19573.7974
8	4096	0	100	19561.21846
9	4096	0	100	19572.46237
10	4096	0	100	19597.42278
11	4096	0	100	19577.95913
12	4096	0	100	19595.8534
13	4096	0	100	19442.13415
14	4096	0	100	18986.06345
15	4096	0	100	19065.45231
16	4096	0	100	18993.24672
17	4096	0	100	19016.87946
18	4096	0	100	19011.96523
19	4096	0	100	19024.61239
20	4096	0	100	19045.18973
21	4096	0	100	18990.85739
22	4096	0	100	18997.49995

Summary

- **Assumptions and working practices:**
  - Assume drive is always full and always under maximum load
  - Assume steady state is a region of interest
  - Always start from a repeatable, known, fixed point
  - Always precondition each drive in the same way
  - Always stimulate the drive with a single, fixed stimulus until steady state is reached
- **Examples of stimuli of interest (all page-aligned):**
  - 4K R/W 100% random
  - 8K R/W 100% random
  - 128K R/W 100% sequential
  - 8K 67R/33W 100% random OLTP
  - 64K R/W 100% sequential

  ➢ All are measured full-span

  ➢ Ensure that the host does not impede performance (when making relative comparisons)

Comments?

Revisit Micron's FMS 2011 presentations at:
www.micron.com/fms