Verification and Management of Endurance in NAND SSDs

Venkatesh Vasudevan, Hanmant Belgal, Neal Mielke
NVM Solutions Group, Intel® Corporation

Flash Memory Summit, Aug 2012

Legal Notices and Disclaimers

INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL PRODUCTS. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN INTEL'S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF INTEL PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT.

A "Mission Critical Application" is any application in which failure of the Intel Product could result, directly or indirectly, in personal injury or death. SHOULD YOU PURCHASE OR USE INTEL'S PRODUCTS FOR ANY SUCH MISSION CRITICAL APPLICATION, YOU SHALL INDEMNIFY AND HOLD INTEL AND ITS SUBSIDIARIES, SUBCONTRACTORS AND AFFILIATES, AND THE DIRECTORS, OFFICERS, AND EMPLOYEES OF EACH, HARMLESS AGAINST ALL CLAIMS COSTS, DAMAGES, AND EXPENSES AND REASONABLE ATTORNEYS' FEES ARISING OUT OF, DIRECTLY OR INDIRECTLY, ANY CLAIM OF PRODUCT LIABILITY, PERSONAL INJURY, OR DEATH ARISING IN ANY WAY OUT OF SUCH MISSION CRITICAL APPLICATION, WHETHER OR NOT INTEL OR ITS SUBCONTRACTOR WAS NEGLIGENT IN THE DESIGN, MANUFACTURE, OR WARNING OF THE INTEL PRODUCT OR ANY OF ITS PARTS.

Intel may make changes to specifications and product descriptions at any time, without notice. Designers must not rely on the absence or characteristics of any features or instructions marked "reserved" or "undefined". Intel reserves these for future definition and shall have no responsibility whatsoever for conflicts or incompatibilities arising from future changes to them. The information here is subject to change without notice. Do not finalize a design with this information.

The products described in this document may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

Contact your local Intel sales office or your distributor to obtain the latest specifications and before placing your product order.

Copies of documents which have an order number and are referenced in this document, or other Intel literature, may be obtained by calling 1-800-548-4725, or go to: http://www.intel.com/design/literature.htm%20

This document contains information on products in the design phase of development.

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products.

Results have been estimated based on internal Intel analysis and are provided for informational purposes only. Any difference in system hardware or software design or configuration may affect actual performance.

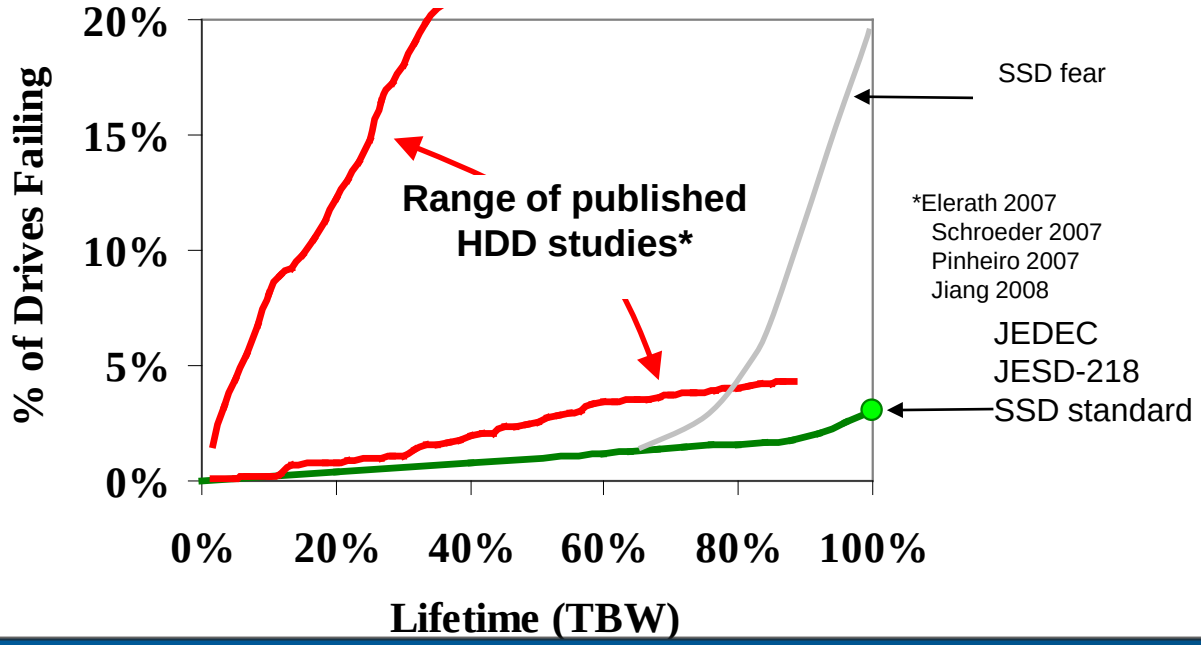
Results have been simulated and are provided for informational purposes only. Results were derived using simulations run on an architecture simulator or model. Any difference in system hardware or software design or configuration may affect actual performance.

Intel does not control or audit the design or implementation of third party benchmark data or Web sites referenced in this document. Intel encourages all of its customers to visit the referenced Web sites or others where similar performance benchmark data are reported and confirm whether the referenced benchmark data are accurate and reflect performance of systems available for purchase.

*Other names and brands may be claimed as the property of others.

SSD Endurance Challenge

- Endurance is the ability to withstand repeated writing of data to SSD
- SSD endurance is finite because NAND Flash memory endurance is finite
- Though some field reliability studies show that HDDs have high failure rates…
- …there has been some *fear* that SSDs will *wear out* from limited endurance



Endurance was a Concern for Widespread Adoption of SSDs in Early Days….

SSD Endurance Challenge Response

- **Challenge to the industry:**
  - Develop Standards and Methods to Address SSD Endurance

- **Response:**
  - JEDEC Technical Committee (JC64.8) formed in May 2008 to tackle the challenge of creating SSD endurance and retention standards
  - Broad membership of SSD customers and suppliers [Samsung, Toshiba, Micron, Intel, IBM, HP, Dell, Seagate, Micron, LSI (Sandforce), others]

- Committee's work resulted in standards: JESD-218 and JESD-219, issued in Sept 2010
  - Client workload supplement approved in 2011 and awaiting formal publication
- Standards required specified low failure rate through the end of the rated endurance life

JEDEC Successfully Published Standards to Clearly Specify and Verify SSD Endurance..

JEDEC (JESD218/219) Solution

- **Two Components**:

1. Define an endurance rating as the TeraBytes Written (TBW) that can be written to the SSD

2. Define a rigorous endurance verification test (EVT) to ensure that an SSD meets the endurance rating

Specify the Endurance…

Verify the Specification

Component #1: Specify the Endurance

Application Class	Workload (see JESD-219)	Active Use (power on)	Retention Use (power off)	Functional Failure Requirement (FFR)	UBER Requirement
Client	Client	40ºC 8 hrs/day	30ºC 1 year	≤3%	$\leq 10^{-15}$
Enterprise	Enterprise	55ºC 24hrs/day	40ºC 3 months	≤3%	$\leq 10^{-16}$

- Separate requirements for client and enterprise classes
- Use conditions defined: detailed I/O workload, temperatures (See JESD218/219 for details)
- Endurance limit encompasses all the criteria mentioned above
  - Data retention time
  - Functional failure requirement (defects in NAND)
  - Uncorrectable bit error rate (write bit errors)
- A drive verified to meet these requirements can be claimed to have the stated JEDEC endurance rating
  - If a customer or supplier has different target use conditions, the verification part of the standard can be customized to match those particular requirements (coming up later)

Endurance spec is total TeraBytes Written (TBW) Over Which the Drive Meets All of These Requirements

Why Should We Care About Details?

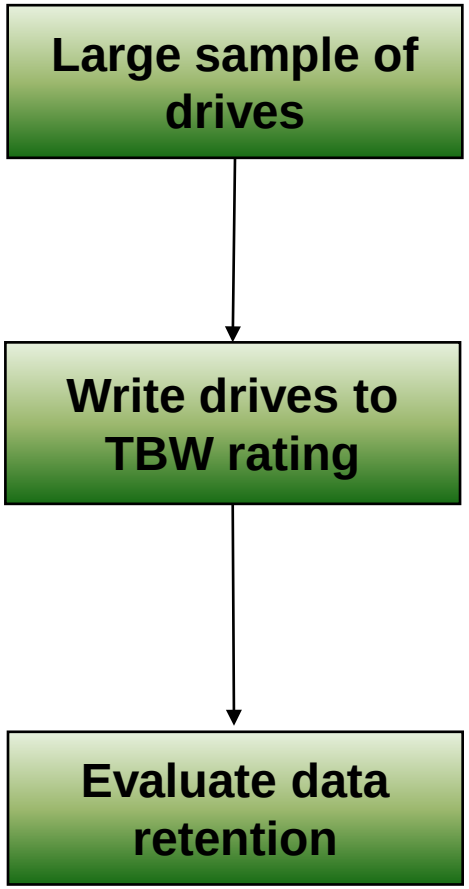
- **NAND has finite endurance because a number of things degrade thru the program/erase cycling of memory cells**
  1. Physical blocks can go bad and will be retired.
  2. Increasing raw bit error rate and chance of an uncorrectable data error
  3. Time for data retention can go down
  4. High voltages inside the ICs can trigger latent defects, shorting out a circuit

- **In addition, characteristics of NAND mechanisms need to be understood and accounted for**
  - Mechanisms can depend on temperature
  - Bit error rate events in NAND can be erratic and requires data to be checked continually

- **Rigorous approach is to consider all factors that drive endurance of SSD:**
  - For example: Endurance cannot be determined solely by counting # of bad blocks or spare capacity
  - Merely measuring NAND endurance can be misleading, since SSD's endurance depends on the drive's error management features
  - Stress on NAND is dependant on Write Amplification factor, which is highly dependant on workload.

JEDEC Standard Covers All Important Factors that Impact Endurance of SSD

Component #2: Endurance Verification Test



- Enough drives to resolve 3% functional failures
- Enough total bits to resolve the UBER

- Similar to HDD “RDT” (Reliability Demonstration Test)
- JESD-219 specifies workload
- Requires high and low temperature stressing
- Data continuously read and verified
- <3% fail, UBER meeting requirement

- High-temperature bake of the drive, aligned to known temperature acceleration for data retention
- Unaccelerated room-temp retention test required too, to address one data retention mechanism that is not temperature accelerated
- Must meet UBER requirement as advertised in datasheet

TBW Claim is Verified for ALL Mechanisms, Taking Into Account ALL the Considerations Discussed

What About High-Endurance Drives?

- With high-write applications requiring ultra-high-endurance SSDs the EVT cannot perform a lifetime's worth of writes on the entire SSD (not enough time)

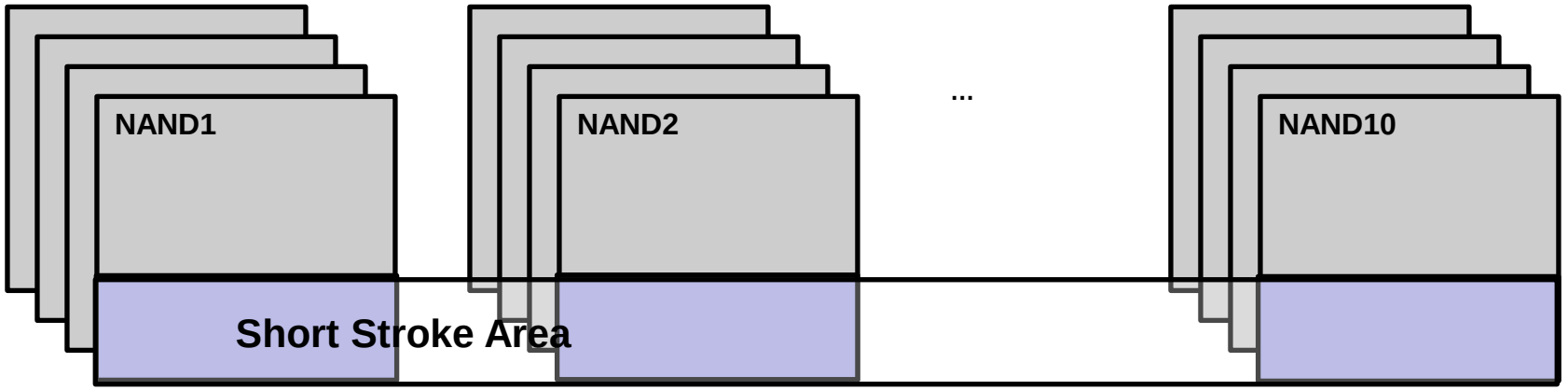
- For those cases, the standard provides extrapolation methods for evaluating full-lifetime reliability

- Intel believes that the most rigorous approach is to reduce the drive capacity ("short stroke") so that some % of the NAND gets stressed to the full endurance limit
  - Exercises the NAND to the correct lifetime P/E cycles in the correct SSD/controller environment

- But the standard defines NAND-based methods as well, when needed
  - Sometimes the endurance target may be so high that short-stroke is not sufficient to complete in reasonable time.
  - For the lower-temperature retention mechanism, it is impractical to take a year to verify the retention capability of a client SSD

JEDEC Standard is Comprehensive and Rigorous..
…and Yet Flexible and Customizable

Short Stroking Method: SSD Endurance Verification



- Writing whole drive to high endurance, such as 30K P/E cycle, would take **months** of just writes, plus system overhead and time to do reads
- Instead, write partial LBA span ('short stroke') to get corresponding % of NAND to life time cycles (see JESD-218)
  - Requires FW modified to prevent wear leveling that would spread cycles across all NAND
- SSD vendors have to artificially reduce the capacity of drives, thru **special firmware** while not distorting the internal working of the drives
- **Gold standard for endurance verification**, because actual SSD controller/FW used to get to final endurance

SSD Customers Should Demand Such Data and Make that Qualification Requirement..

Challenge: Inconsistent Application of JEDEC Standards – Enterprise SSD

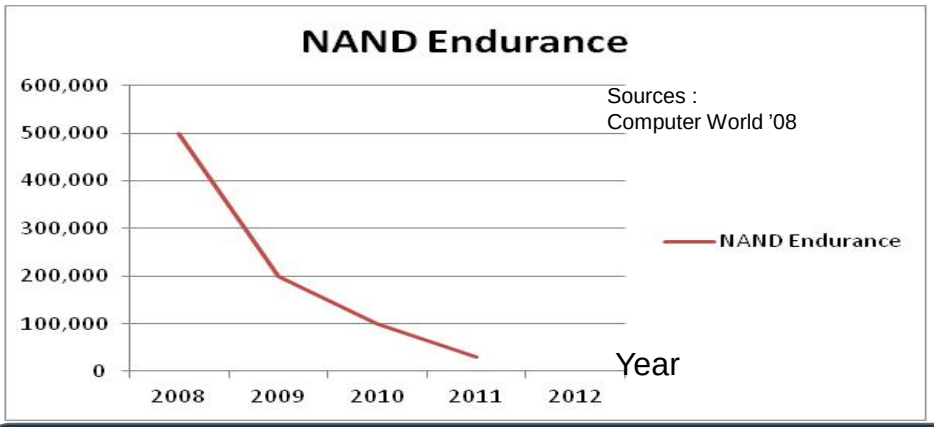
Supplier	Endurance Claims/Requirements	Workloads Information	Failure rate criteria (UBER, AFR, failure mechanisms, retention)	Verification method
JEDEC	TBW- Supplier Specified	JESD 219	JESD 218: UBER 1E-16; 3% lifetime FFR; Data retention specified	JESD218
Intel	TBW	JESD 219 and 4K/8K Random	1E-17 UBER; AFR; JESD218 retention	JESD218
Example of SSD Suppliers	TBW	JESD 219	1E-16 UBER; AFR; retention	JESD218
	Endurance in NAND writes!	70/30 R/W ; possibly random pattern	1E-20 UBER;no FFR; no retention	Not mentioned
	30 drive writes/day;	No workload specified	1E-20 UBER; no FFR ; no retention	Not mentioned
	TBW	Unclear workloads mentioned	1E-18 UBER;no FFR; no retention	Not mentioned
	TBW	50%seq/50% random	Not Mentioned	Not mentioned
	TBW	No reference to wokload	1E-17 UBER;no FFR; no retention	Not mentioned
Example of SSD Customers	Drive write/day	No specific workload	1E-17 UBER; AFR; retention specified	Not mentioned
	Not Specified	No specific workload	1 E-17 UBER; AFR;retention specified	Not mentioned

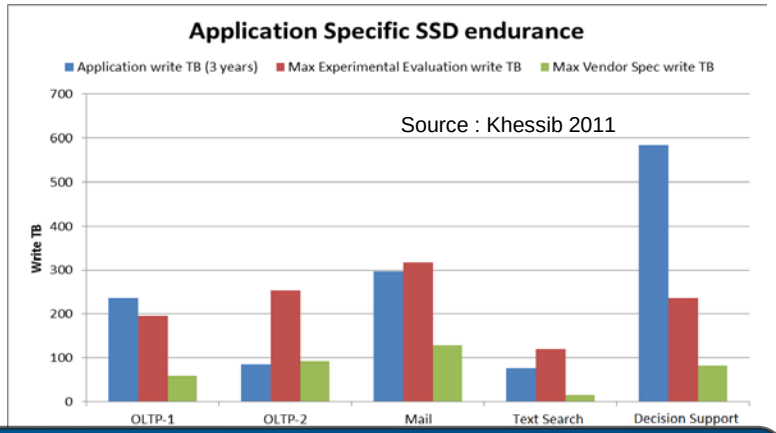
- Write endurance claims includes host writes and even NAND WRITES!
- Workloads are often not specified or are inconsistent with what others are doing
- Failure criteria often do not explicitly comprehend functional failure rates and data retention
- Unclear whether claims of UBER of 1E-18 or better are verified rigorously thru life time usage (very difficult), or merely ‘estimated’

Inconsistency Regarding Workloads and Endurance Claims continues…

Trend of Endurance and Usages – Enterprise Class

- NAND P/E Cycle Requirements have been Trending Lower for Enterprise Applications (see the left graph)
  - 500K P/E cycle was heavily guard banded requirement when SSDs were new
  - Industry has reacted to the cost imperatives (MLC vs. SLC) and removed excessive guardband
- This Trend of Reducing Guardbands is Likely to Continue…
  - Example: Customer guardbanding 2x because "endurance may be 2x worse than claimed" would be better served with rigorous endurance verification to remove the uncertainty
  - Example: Different requirements for different applications (see right graph)
  - Example: Enhanced use of SMART endurance-used indicators provides safety net, so that endurance requirement does not need to be guard-banded to extremes





Usages and Workload Conditions Should be Clear and Form the Basis of Endurance Assessment

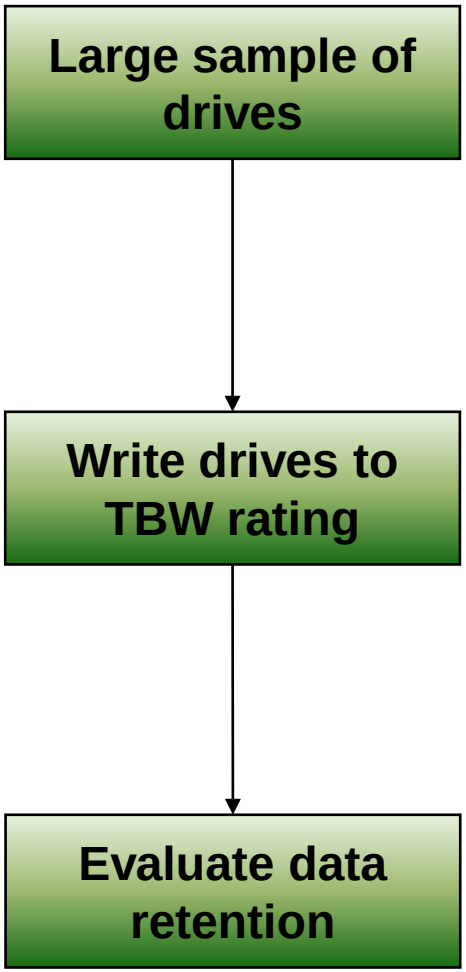
Customization within JEDEC Framework

Application Class	Workload (see JESD-219)	Active Use (power on)	Retention Use (power off)	Functional Failure Requirement (FFR)	UBER Requirement
Client	Client	40°C 8 hrs/day	30°C 1 year	≤3%	$\leq 10^{-15}$
Enterprise	Enterprise	55°C 24hrs/day	40°C 3 months	≤3%	$\leq 10^{-16}$

- The standard allows the EVT method to be customized
- Examples of customizing the requirements
  - Good
    - Microsoft's paper from 2011 Flash Memory Summit (Khessib et al):
      - Wants days (not weeks) of data retention
      - Endurance estimated for a variety of workloads (not just one)
  - Bad
    - Customer requirements for UBER such as 1E-20 are not realistic and cannot be verified thru rigorous testing.
    - Customer requires a particular endurance but without specifying workload
    - Customer wants more stringent targets based on "we assume that the endurance will only be half what is claimed"

Win-Win for Industry is the Right Targets, Rigorously Verified

Following JEDEC…
but with Flexibility and Rigor



- Enough drives to resolve 3% functional failures
- Enough total bits to resolve specified UBER **(ex: 1E-17)**

- JEDEC supplies the workload **-> can be changed to any specified workload and measure WAF (see JESD219)**
- Data written and verified **-> Short-stroke method**
- <3% fail, UBER meeting requirement

- High-temperature bake of the drive, aligned to known temperature acceleration for data retention **-> can be adjusted to meet data retention for specified time (see JESD218- Table 3.0)**

Following JEDEC for Consistency and Rigor ..but Taking Advantage of its Flexibility

Complete Endurance Specification for SSD: Example*

Workload	Drive Writes/day	Retention	UBER (thru life time usage)	Functional Failure Rate (thru life time usage)
JESD 219	10	3 months at 40C	1E-16	3%
JESD 219	15	1 month at 40C	1E-16	3%
JESD 219	5	1 year at 40C	1E-16	3%
4K random-100% span	10	3 months at 40C	1E-16	3%
100 % Sequential	20	3 months at 40C	1E-16	3%

* Hypothetical values

Publishing Endurance Specifications Based on JEDEC will Benefit the SSD Industry

Call to Action

- SSD customers should clearly understand and specify workloads and usages for endurance estimation
  - Segmented for application types and reasonably guard banded for uncertainties
  - Customers should utilize SMART indicators and other tools to actively manage endurance consumption of drives in field

- The JEDEC standards are excellent ones that the industry should rally around
  - Standards are complete and stringent in their coverage of endurance/retention limits
  - Standards allow for customization including different application conditions
  - But simply ignoring the standards, and the rigor behind them, is a recipe for getting field reliability that does not match needs
  - Customers should demand and suppliers should provide rigorous testing to meet the spirit of JESD

SSD Customers and Suppliers Should Follow JEDEC Standards for SSD Endurance

References

Elerath 2007: Elerath (NetApp) & Pecht (U. Maryland), “Enhanced Reliability Modeling of RAID Storage Systems,” 37th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN'07)

Jiang 2008: Weihang Jiang, Chongveng Hu, Yuanyan Zhou (all U. Illinois), and Arkady Kanevsky (NetApp), “Are Disks the Dominant Contributor for Storage Failures? A Comprehensive Study of Storage Subsystem Failure Characteristics,” ACM Transactions on Storage, Vol. 4, No. 3, Article 7, Nov. 2008

Kavalanekar 2008: S. Kavalanekar et al. (Microsoft), Characterization of Storage Workload Traces from Production Windows Servers,” IIWSWC 2008

Khessib 2011: B. Khessib et al. (Microsoft), “Opportunities and Challenges of Using Solid State Drives in Large Scale Datacenters,” Flash Memory Summit 2011

Pinheiro 2007: Eduardo Pinheiro, Wolf-Dietrich Weber and Luiz Andr´e Barroso (Google), “Failure Trends in a Large Disk Drive Population,” 5th USENIX Conference on File and Storage Technologies (FAST’07), February 2007

Riska 2009: A. Riska and E. Riedel (Seagate/EMC), Evaluation of Disk-Level Workloads at Different Time-Scales,” IISWC 2009

Schroeder 2007: Bianca Schroeder and Garth Gibson (Carnegie Mellon), “Understanding Disk Failure Rates: What Does an MTTF of 1,000,000 Hours Mean to You?” ACM Transactions on Storage, Vol. 3, No. 3, Article 8, October 2007

Computer World ‘08: Agam Shah, “Sun calls for SSD companies to unite on standards” Computer World, Aug 5 2008.