Solid-State Drive System Optimizations In Data Center Applications

Tahmid Rahman

Senior Technical Marketing Engineer

Non Volatile Memory Solutions Group

Intel Corporation

Monday, August 15, 2011

Legal Disclaimer

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.

UNLESS OTHERWISE AGREED IN WRITING BY INTEL, THE INTEL PRODUCTS ARE NOT DESIGNED NOR INTENDED FOR ANY APPLICATION IN WHICH THE FAILURE OF THE INTEL PRODUCT COULD CREATE A SITUATION WHERE PERSONAL INJURY OR DEATH MAY OCCUR.

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.

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. For more information go to http://www.intel.com/performance

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.

Material in this presentation is intended as product positioning and not approved end user messaging.

Intel and the Intel logo are trademarks of Intel Corporation in the U.S. and other countries.

*Other names and brands may be claimed as the property of others.
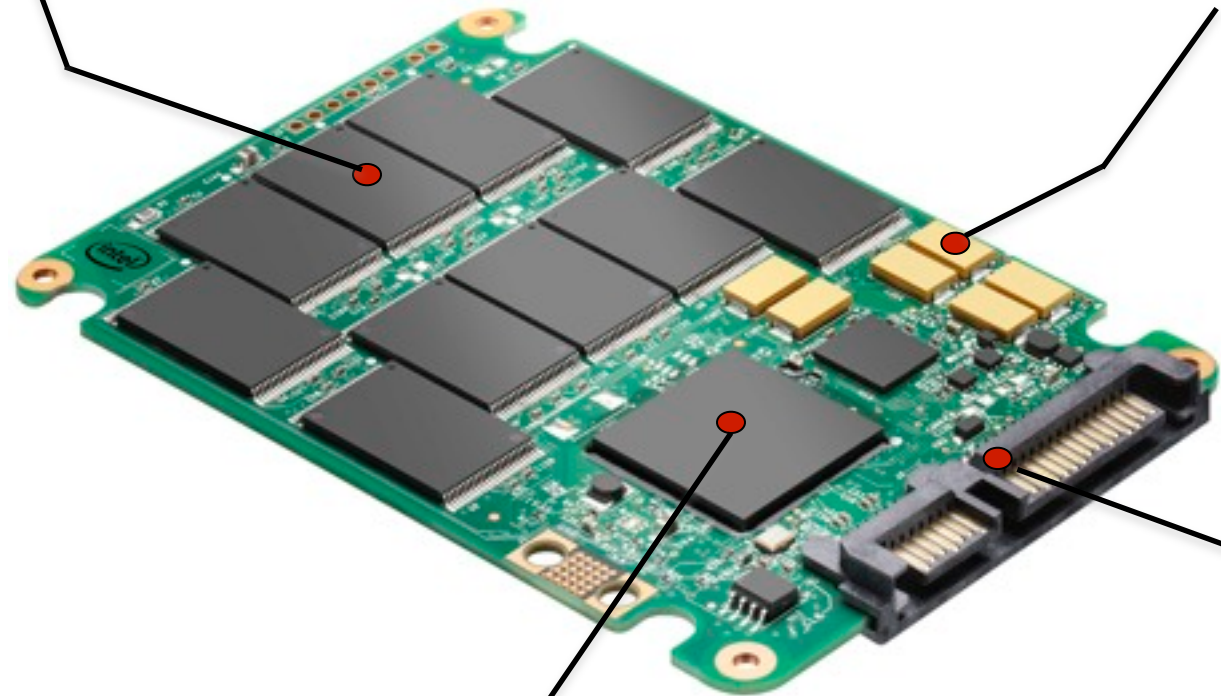
Monday, August 15, 2011

Agenda

- What is important in data center applications
- Retention and endurance limiters for SSDs
- Overcoming the limiters through system optimization
- NAND management and Quality of Service
- Data path protection
- Summary

What Is Important In Data Centers?

**NAND Endurance /
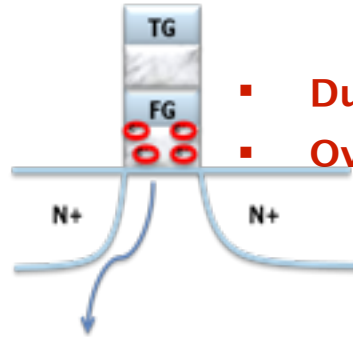Retention Management**

Robust Power Management

Quality of Service
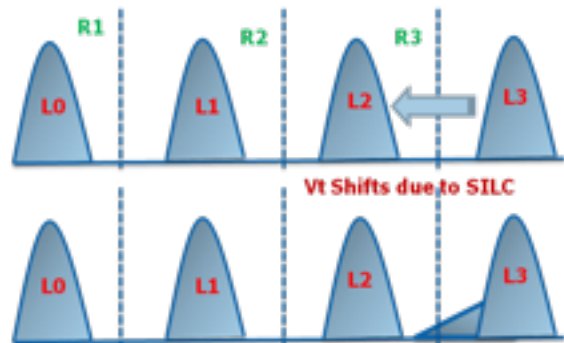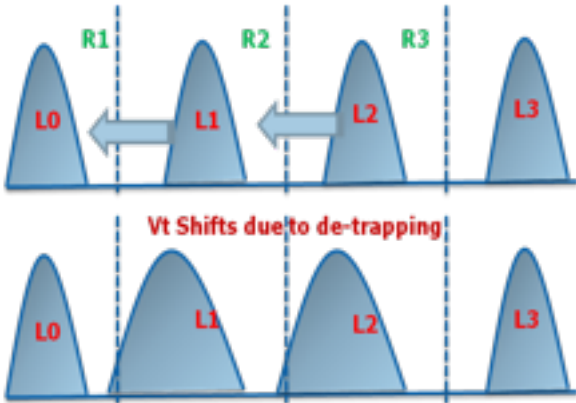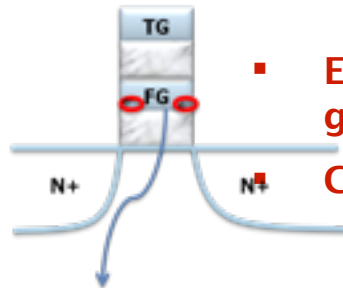
**Data
Integrity**

Retention Limiters For SSDs

Intrinsic Charge Loss (de-trapping) Effect

- **During P/E cycles, charge gets trapped in oxide**
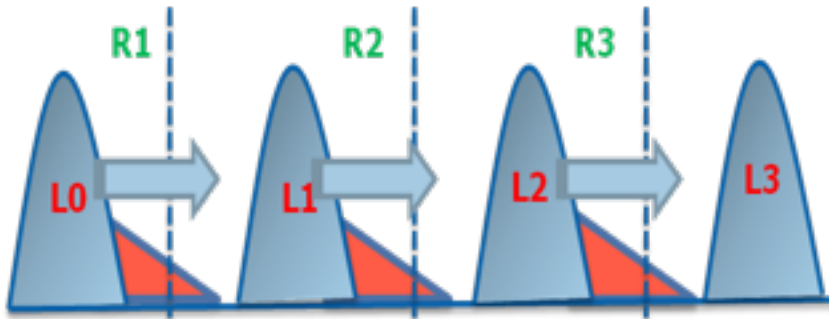- **Over time, de-trapping creates retention issues**

Stress Induced Charge Loss Effect

- **Electrical stress introduces leakage via floating gate**
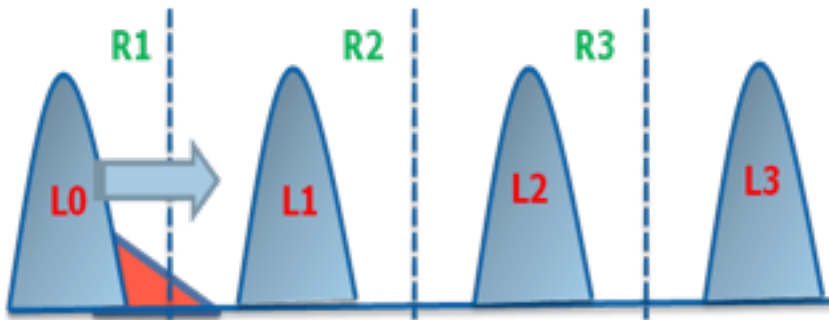- **Causes Vt shifts of L3 states**

Endurance Limiters For SSDs

Program Disturb

FG-FG coupling tends to shift the cell Vts upwa
Over-programming may also caused Vt shifts

Read Disturb

During Read, inhibited cells can get programm
Creates Vt shifts upward on erased cells

Overcoming Retention/Endurance Limiters

- Well-characterized NAND based on
  Read Window Margin and Intrinsic charge loss (ICL)

- Finer granularity of programming steps or slow programming to widen Read Window Margin

- Early detection and monitoring of ECC fatal events

- Re-allocation of active area to mitigate read/program disturbs – wear-leveling and data refresh

- Additional spare area reduces the burden on NAND and enables parity protection during catastrophic failure – such as bad die

SSD Endurance/Retention Specification For Data Center

- SSD Power Off Retention and UBER target standardized through JESD-218/219 specification

| Application Class	Workload (JESD-219)	Active Use (power on)	Retention Use (power off)	Functional Failure Requirement (FFR)	UBER Requirement
Client	Client (Draft)	40C 8 hrs/day	30C 1 year	≤3%	≤10-15
Enterprise	Enterprise	55C 24hrs/day	40C 3 months	≤3%	≤10-16

3 months of retention enables higher P/E cycl

- SSD Total Endurance Rating is also defined based on a given mixed workload
  - ➢ Much higher write amplification than a typical client workload

(0.5k) 4%,(1k) 1%,(1.5k) 1%,(2k) 1%, (2.5k) 1%, (3k) 1%, (3.5k) 1%, (4k) 67%, (8k) 10%, (16k) 7%(32k) 3%, (64k) 3% – Example of Enterprise work, workload uses different span sizes too (see JESD-219)
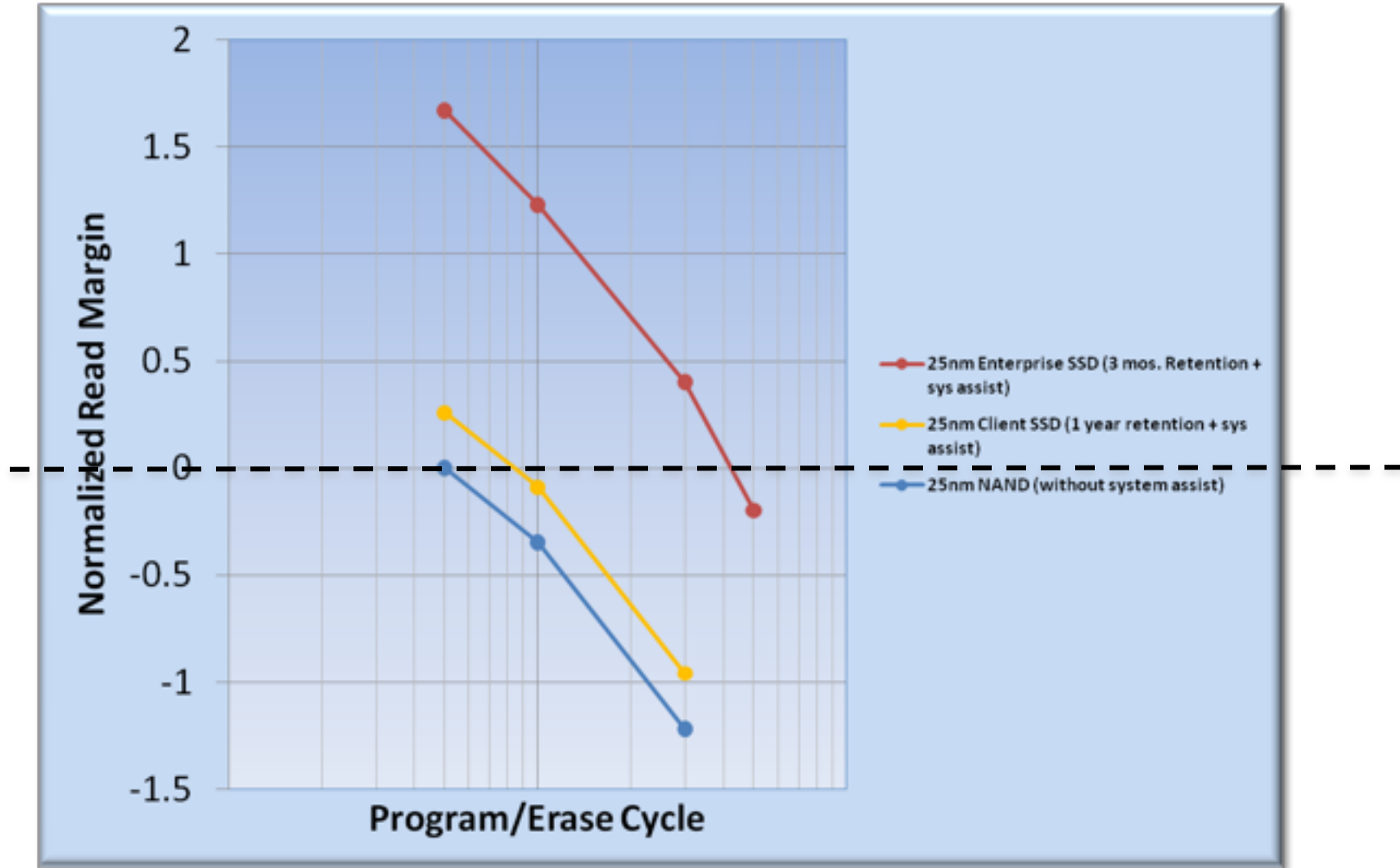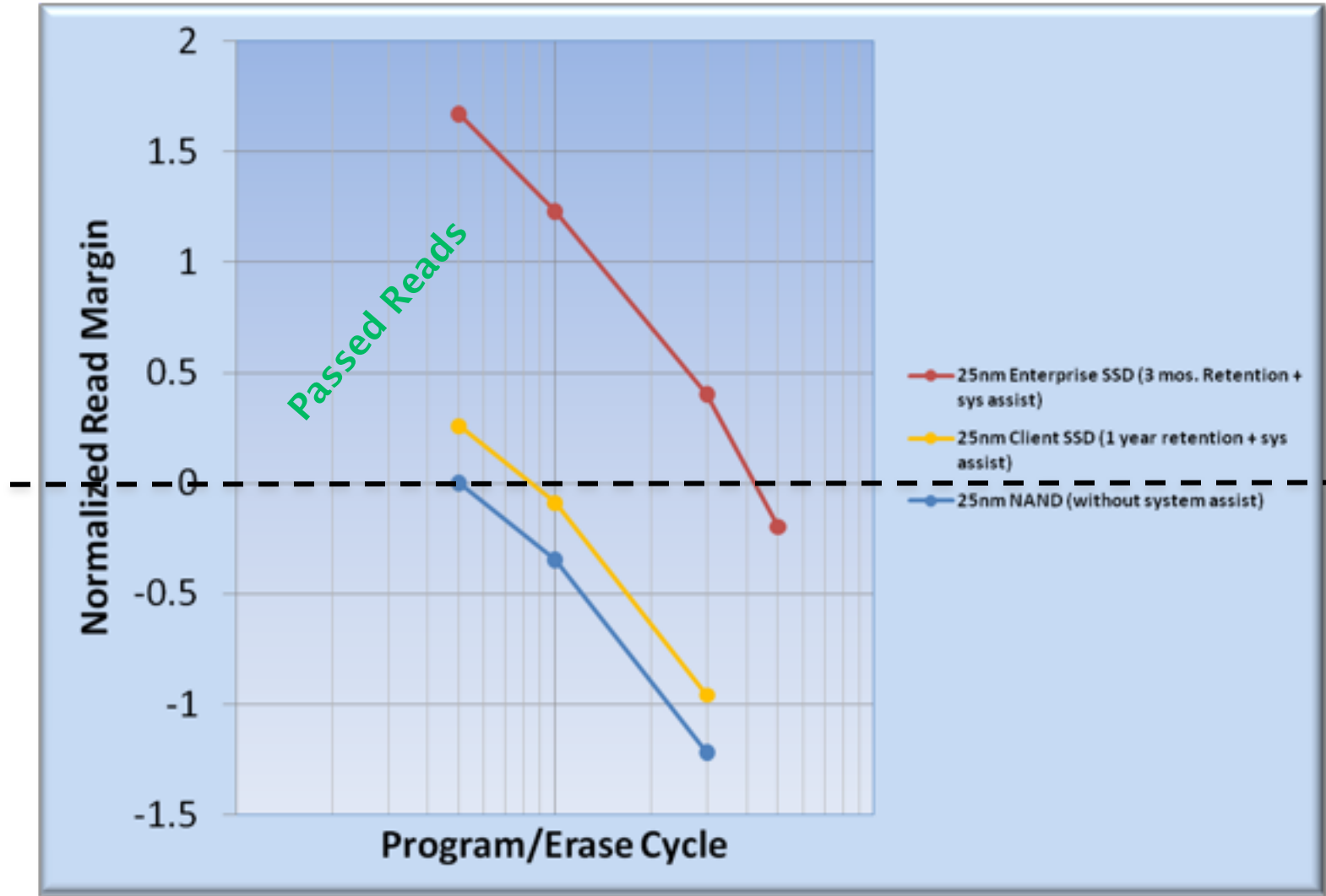
Higher Read Margin Enables Data Center Ready SSDs
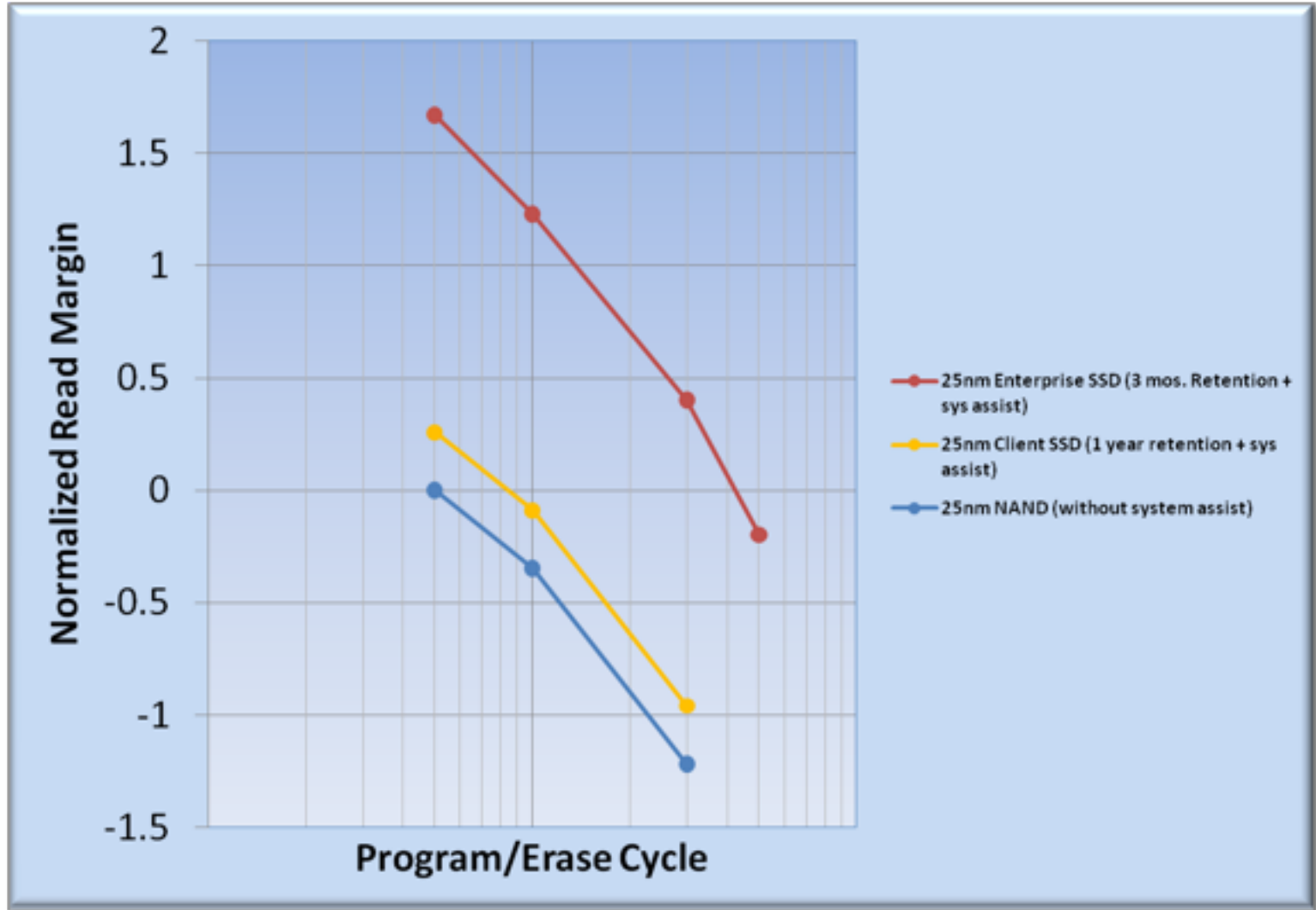
Source : Intel 25nm SSD Analysis

Higher Read Margin Enables Data Center Ready SSDs
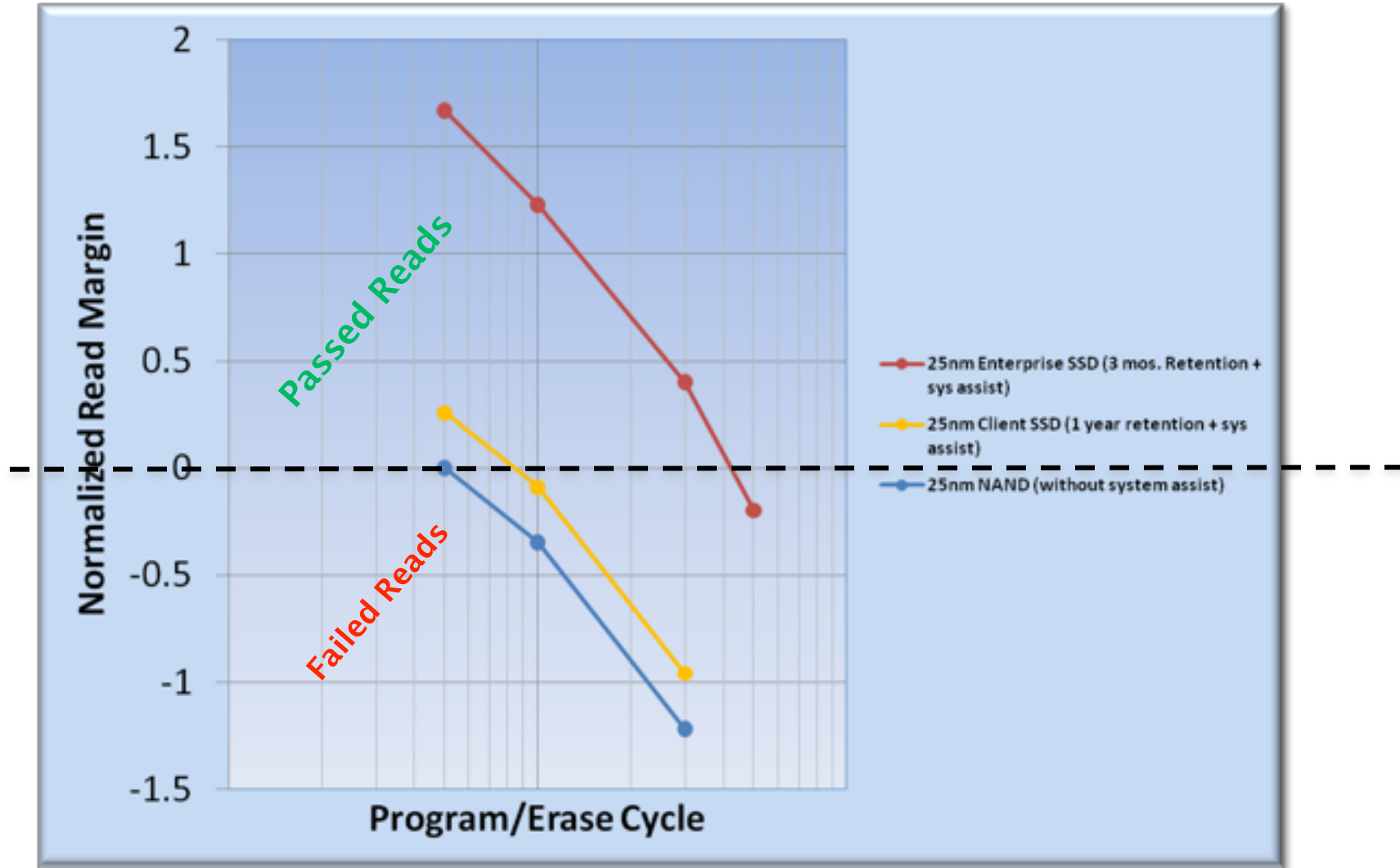

Source : Intel 25nm SSD Analysis

Higher Read Margin Enables Data Center Ready SSDs

Source : Intel 25nm SSD Analysis

Higher Read Margin Enables Data Center Ready SSDs

Source : Intel 25nm SSD Analysis

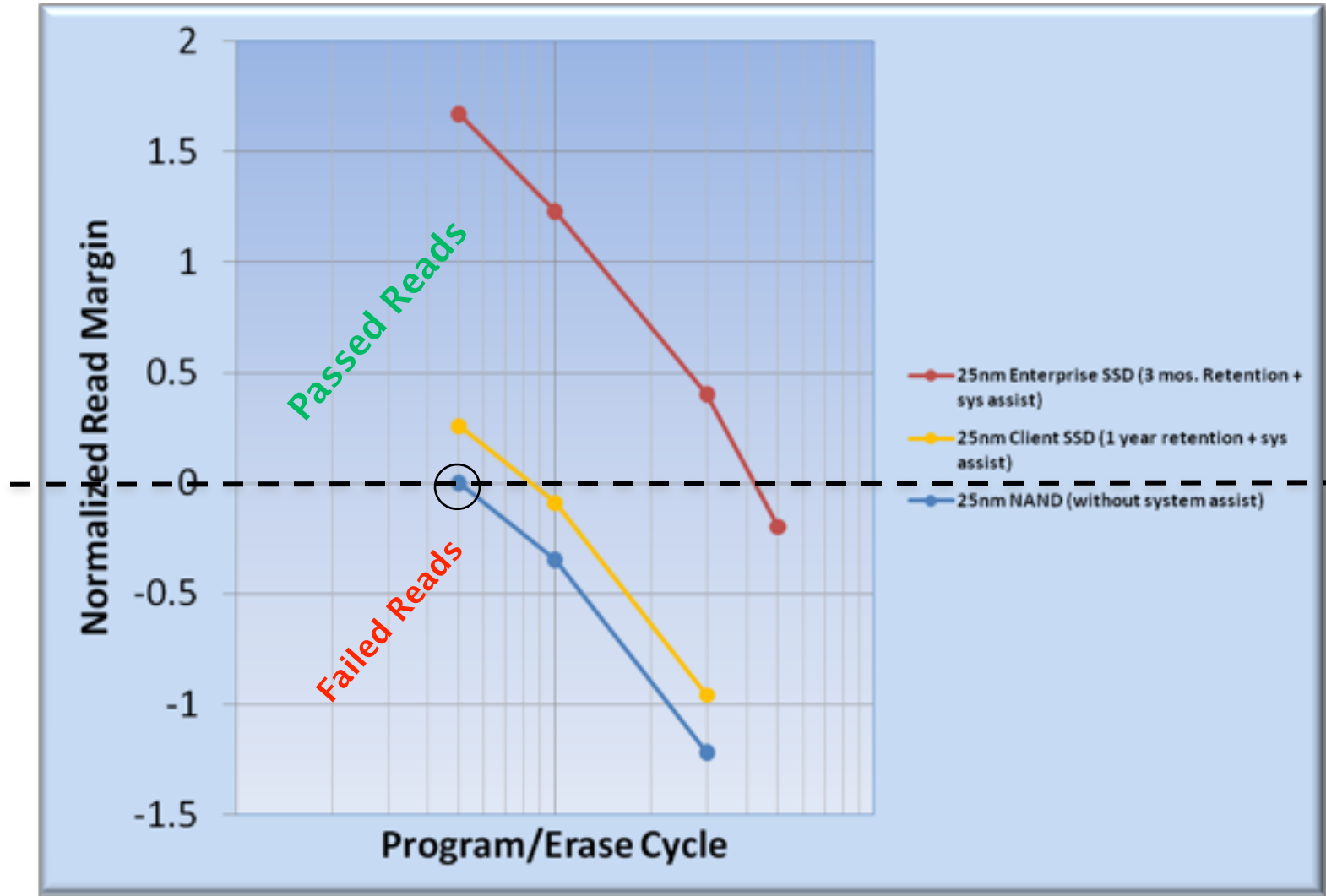# Higher Read Margin Enables Data Center Ready SSDs

Source : Intel 25nm SSD Analysis
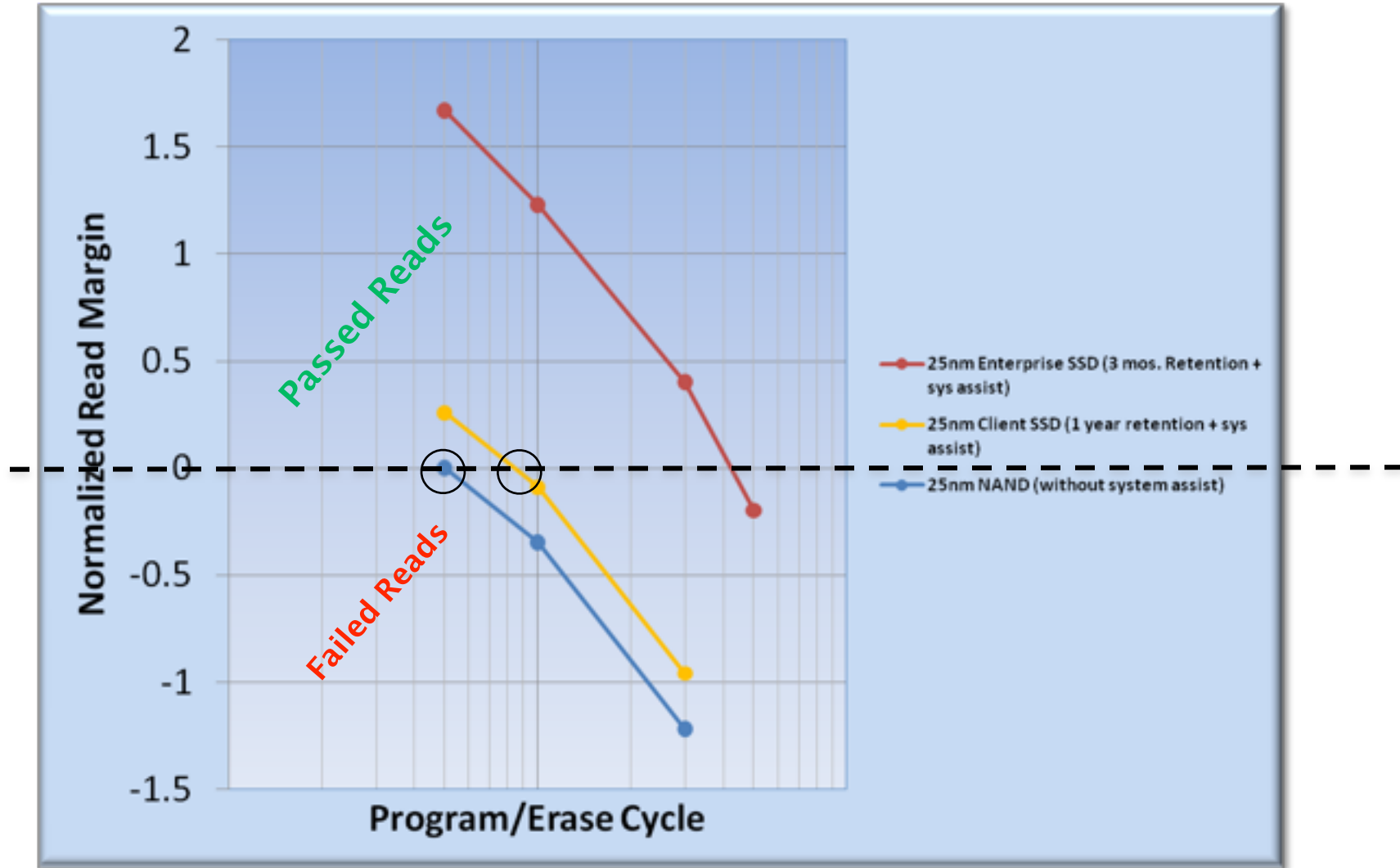
Higher Read Margin Enables Data Center Ready SSDs

Source : Intel 25nm SSD Analysis

Higher Read Margin Enables Data Center Ready SSDs

Source : Intel 25nm SSD Analysis

Monday, August 15, 2011
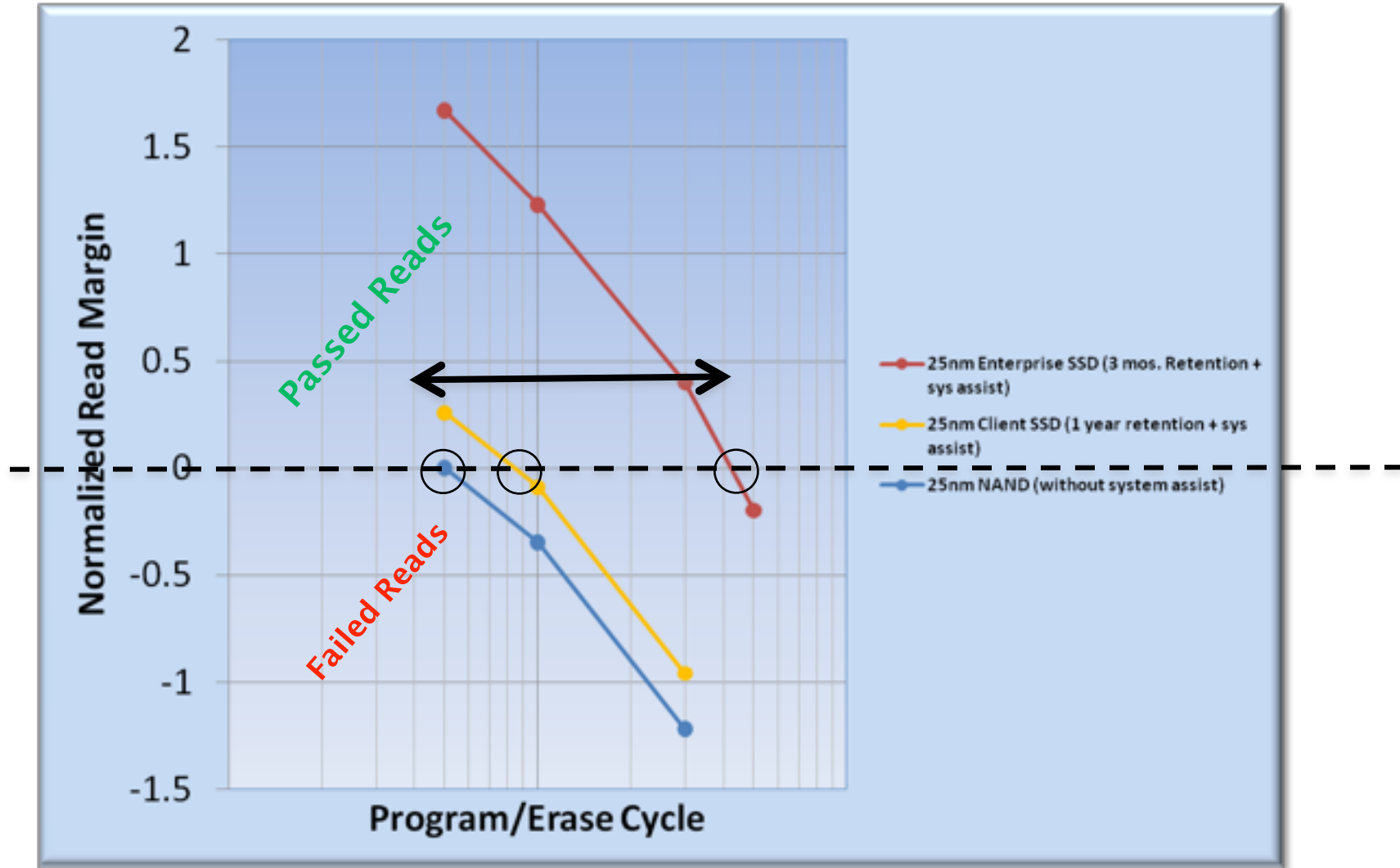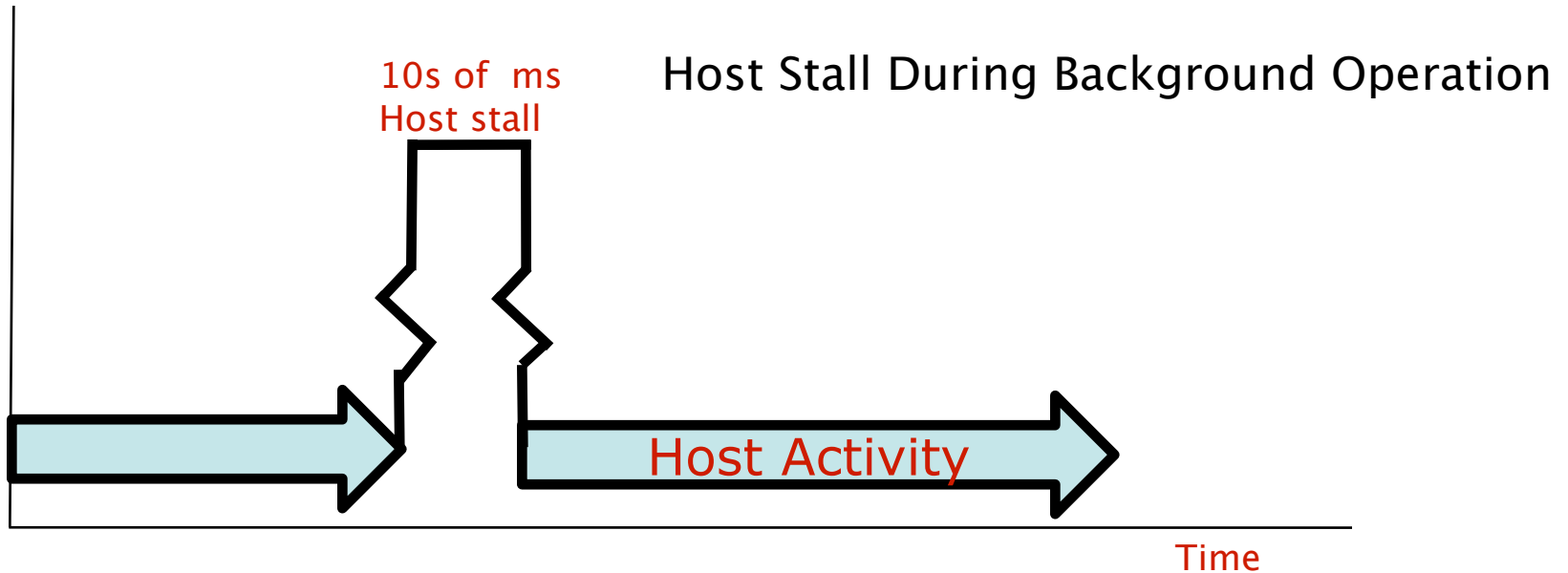
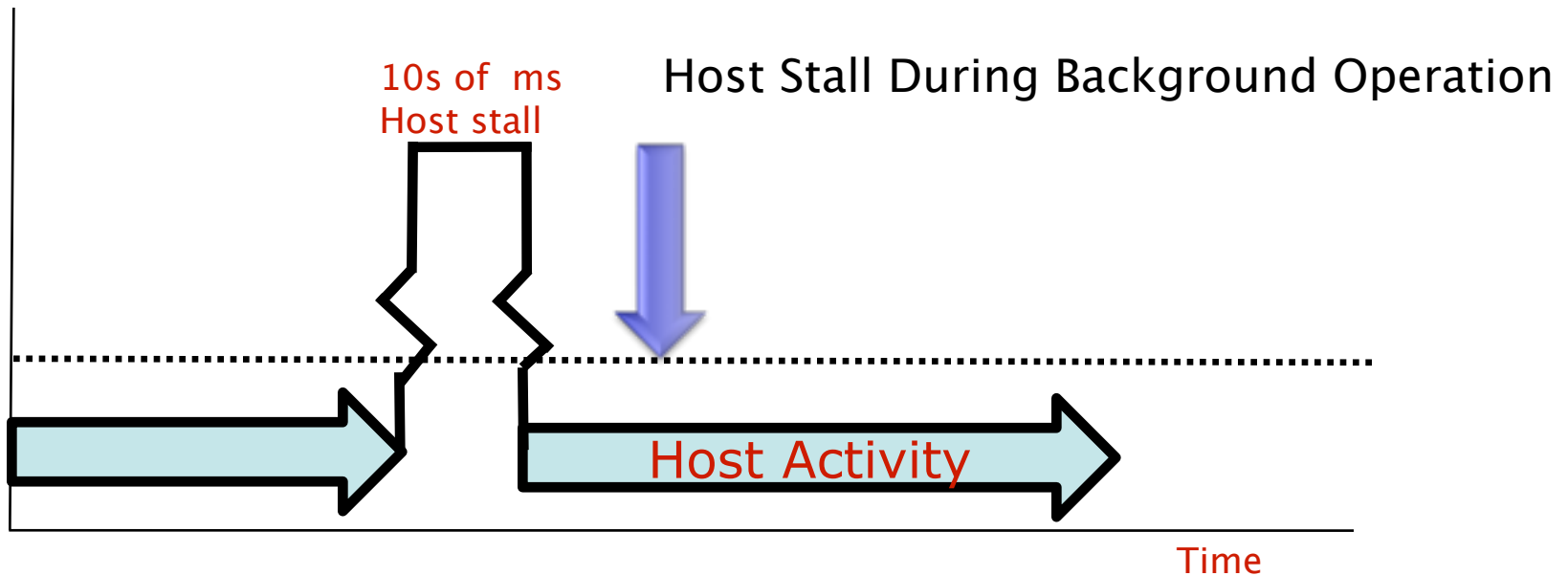# NAND Management Without Quality o Service (QoS) Impact

- NAND Management algorithm should execute
  - Without creating considerable host halts
  - Without creating high write amplifications
  - Giving priority to other critical non-host managements such as wear-leveling or defragging

10s of ms Host stall

Host Stall During Background Operation

Host Activity

Time

- **NAND Management algorithm should execute**
  - Without creating considerable host halts
  - Without creating high write amplifications
  - Giving priority to other critical non-host managements such as wear-leveling or defragging

10s of ms
Host stall

Host Stall During Background Operation
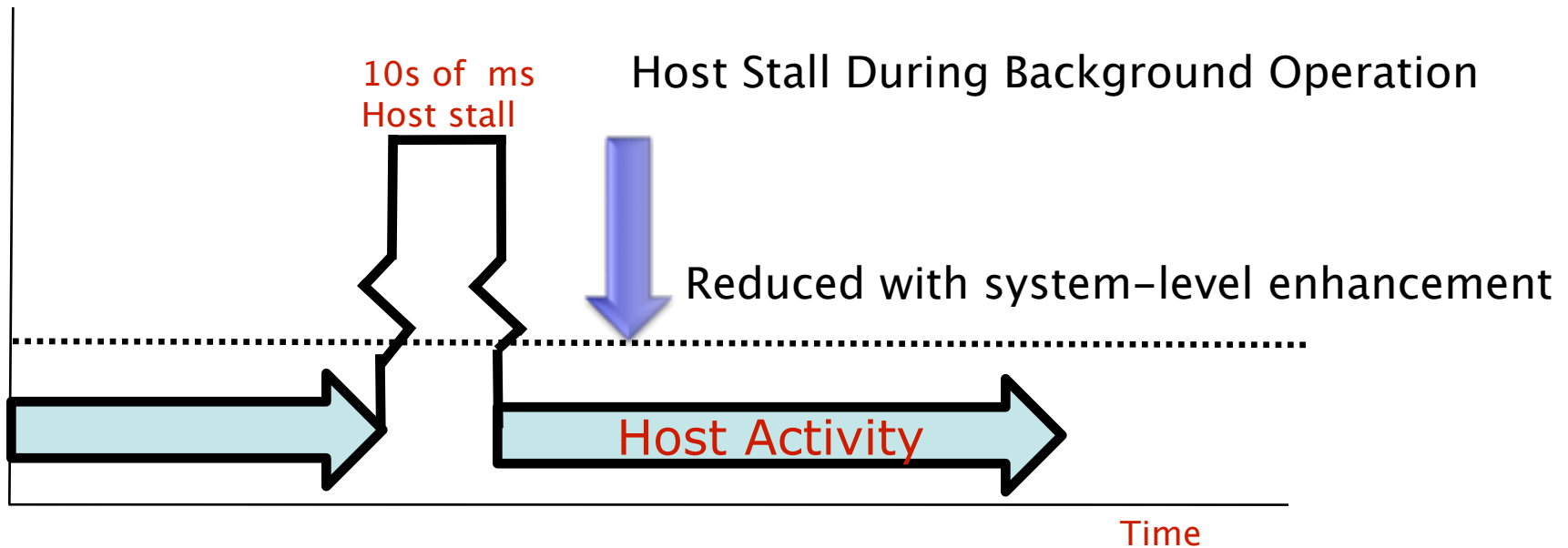
Host Activity

Time

- **NAND Management algorithm should execute**
  - Without creating considerable host halts
  - Without creating high write amplifications
  - Giving priority to other critical non-host managements such as wear-leveling or defragging

10s of ms
Host stall

Host Stall During Background Operation

Reduced with system-level enhancement

Host Activity

Time

Protecting Data Path Against Error Conditions

- SSDs containing temporary buffers as part of internal cache

- Temporary buffers can store host or non-host data

- SSDs need to
  - A) Protect buffers with data in flight during power loss
  - B) Detect errors on the data path

← **Mechanism to flush stored data during power l...**

Data +CRC   Data + CRC + LBA Tag   Data + CRC + LBA Tag +ECC

Checkpoints For integrity →

| Host | → | Buffer Management | → | Write Cache | → | NAND |

Summary

- Data center requires better NAND management to meet high endurance and retention targets
- QoS is key and should not be compromised
- Internal/external memory on the data path must be protected and monitored