



A Controller Redesign Slashes the Cost of Data-Center Storage

By Pearse Coyle, CEO, NVMdurance

The Project

The NAND market is approaching its “scaling limit,” a point beyond which storage cost can no longer be reduced by shrinking the production process alone. We sought a storage solution that would enable data centers to extend NAND device life, to slash storage costs, and to move between the most cost effective TLC NAND or advanced 3D NAND flash technologies while still successfully meeting the stringent reliability requirements that cloud and HPC data center storage customers demand.

The Design Challenge

NAND flash chips can sustain only a finite number of program/erase cycles before they wear out. There are a number of ways that this life can be extended:

1. Provide more flash cells to replace worn out cells (i.e. overprovisioning, costly)
2. Shorten the retention: the length of time the chips must retain data when the unit is switched off (usually set at a year in consumer-grade flash)
3. Reduce the amount of data to be stored, with technologies such as compression and deduplication
4. Use an error correcting scheme with more error correcting bits or a more powerful algorithm (e.g. LDPC).

But all of these options come with either a hardware monetary cost (e.g. substantially higher controller gate count required to implement LDPC) or they come with a performance cost. Many come with both.

This dilemma—either high cost for NAND devices that quickly wear out, or high cost for complex controllers and strategies to extend the chip life—is slowing the adoption of solid-state storage architectures, even though we know these systems can dramatically improve performance, predictability, and energy consumption of applications in a cloud environment.



The Design Team:

The design team at NVMdurance has evolved through 15 years of research in characterizing NAND flash devices. This research has led to automated techniques for characterization of NAND devices, generating optimized operating parameters, and developing programmable controllers to employ those parameters. The result is greatly extended operating life for production NAND devices in data center applications.

Challenge:

Flash memory devices wear-out, and data center/cloud use-models quickly drive these flash devices to exhaustion far too soon. Traditionally these flash devices will be on SSDs which will then be thrown out entirely - flash, controller, and all packaging - even though it is only the flash chips that are exhausted.

Solution:

By automatically characterizing the NAND chips, mostly before deployment but partly on the fly, and then using that data to set the operating parameters of a flexible controller, we enable the flash to last, typically, 4 to 7 times longer. We get an additional factor of 10 to 20 times longer by combining our approach with other efforts, such as LDPC.

NVMdurance, with a long background working in foundries to characterize NAND behavior and to develop optimal NAND management strategies, sought to break apart the dilemma. But that would require a controller that could continuously update parameters as NAND devices aged, and then be completely reconfigured when old NAND DIMMs were replaced with devices using a different flash technology requiring entirely different parameters. A software-based controller would offer the necessary flexibility, but software could not execute some of the parameterized operations quickly enough to preserve the inherent performance of the NAND devices.

The Design Solution

Our solution comes in the form of an analysis tool and a highly-configurable, SoC-FPGA-based solid-state storage controller. First, before a new NAND family goes into production, NVMdurance uses a proprietary analysis tool – NVMdurance Pathfinder - to characterize the optimal range of operating parameters for the family. This information is passed on to the reconfigurable controller, which has two primary functions. First, the controller continuously monitors the condition of the NAND flash during operation and automatically adjusts the controller operating parameters in real time. Second, the controller employs these parameters in managing the NAND array and in conducting individual transactions with the NAND DIMMs, greatly expanding the useful life of the NAND chips. Because the controller adjusts based on its continuous monitoring, it anticipates shifts in NAND behavior and avoids the read/write errors and retries that harm both performance and predictability in conventional solid-state storage.

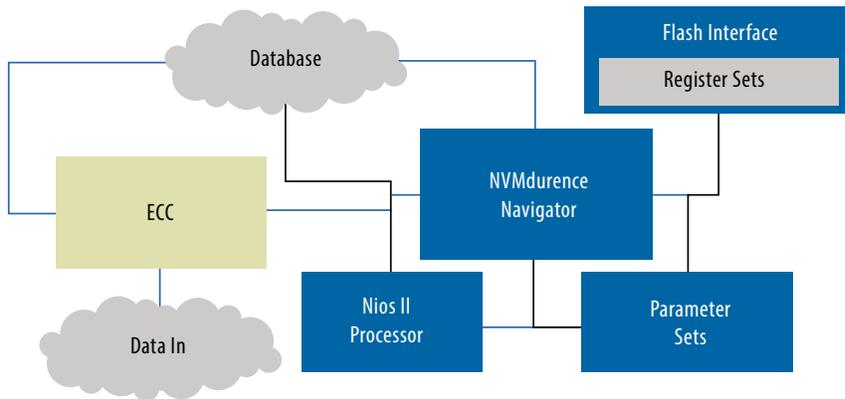
The controller also has more than adequate headroom for implementing compression and deduplication algorithms, strong error correction, and vendor-proprietary techniques for further improving storage performance and lifetime. This controller is implemented in an Altera SoC FPGA.

Theory of Operation

Before the memory product (e.g. an SSD) goes into production, NVMdurance Pathfinder, a custom built suite of machine learning techniques, determines multiple viable sets of flash register values. The power behind NVMdurance is the use of this off-line machine learning software that automatically learns the optimal parameter settings for the NAND device. It provides either a static set of parameters, which are set at the start of life for the device, or for optimal endurance a dynamic set, which allows the controller to periodically optimize the parameters over time as the device ages. The parameters provided are all relative to the manufacturers' specified parameters so they work even though individual chips/batches have slight variations in their absolute performance.

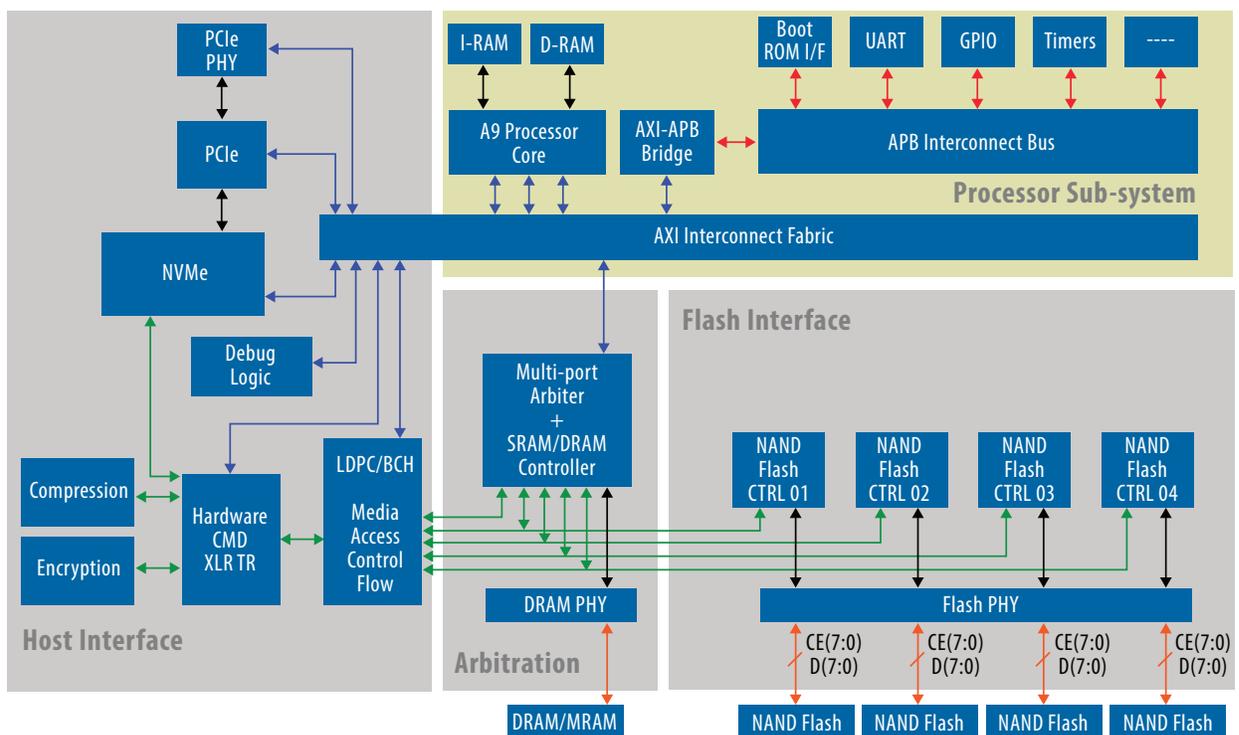
No factory determined operating parameters are ideally suited to each application, nor can they track the condition of the flash as it degrades through use. So our preferred method is to pass the conclusions from Pathfinder to the configurable controller. Then on the controller of the live memory product NVMdurance Navigator continuously monitors operation of each block of the solid-state memory, and on the fly chooses which of the predetermined sets to use for each stage of life to ensure that the flash lasts as long as possible (Figure 1).

Figure 1. Information Flow in the Controller.



The controller algorithms are implemented in an Altera SoC FPGA (Figure 2). In this device, the host interface and hardware Flash memory interface are implemented primarily in programmable logic, while software functions reside on the ARM Cortex-A9 subsystem.

Figure 2. Controller Implementation



Results

The actual extension of endurance depends on the particular use-case and the information available to the NVMdurance Navigator autonomic system. But as a typical example, using Altera's embedded CPU FPGA architecture in an optimized, cost-effective, single-chip solution integrating a solid-state disk (SSD) controller from Mobiveil and NAND optimization software from NVMdurance, we have proven to double the life of NAND flash and increase the number of program-erase cycles by up to 7 times. This longer life reduces the cost of flash storage per month to the data center operator by at least 75%. Implementation of other compression and error-correction methods—for which there is still more than adequate capacity in the SoC FPGA, would extend the life of the NAND DIMMs up to 10 to 20-fold.

And there is a second source of major savings. The cost per month of the overall solution goes down much more than 75% because the overall package (other than the flash) can be used indefinitely. When the NAND devices eventually reach the point where the controller can no longer achieve adequate performance from them, because they are mounted on DIMMs they can be replaced with newer devices in whatever technology—say, 3D NAND—is most cost-effective at the time. And the controller can be reconfigured to operate at full functionality with the new devices. This extends the life of the entire solid-state storage unit indefinitely and the savings go straight to the bottom line of a cloud provider that is selling GB by month because the hardware that is providing those GB can now last for many more months.

"This Design Solution describes an actual design that has been developed. However, it does not represent a supported product or reference design, and is not orderable from Altera. If you would like additional information, please contact Altera's authorized distributor".

Altera Corporation

101 Innovation Drive
San Jose, CA 95134
USA
Telephone: (408) 544 7000
www.altera.com

Altera European Headquarters

Holmers Farm Way
High Wycombe
Buckinghamshire
HP12 4XF
United Kingdom
Telephone: (44) 1 494 602 000

Altera European Trading Company Ltd.

Building 2100
Cork Airport Business Park,
Cork, Republic of Ireland
Telephone: +353 21 454 7500

Altera Japan Ltd.

Shinjuku i-Land Tower 32F
6-5-1, Nishi Shinjuku
Shinjuku-ku, Tokyo 163-1332
Japan
Telephone: (81) 3 3340 9480
www.altera.co.jp

Altera International Ltd.

Unit 11- 18, 9/F
Millennium City 1, Tower 1
388 Kwun Tong Road
Kwun Tong
Kowloon, Hong Kong
Telephone: (852) 2945 7000
www.altera.com.cn

Altera Corporation Technology Center

Plot 6, Bayan Lepas Technoplex
Medan Bayan Lepas
11900 Bayan Lepas
Penang, Malaysia
Telephone: 604 636 6100

