Latent Sector Error Modeling and Detection for NAND Flash-based SSDs

Guanying Wu, Chentao Wu, and Xubin He
Department of Electrical and Computer Engineering
Virginia Commonwealth University, Richmond, VA
{wug, wuc4, xhe2}@vcu.edu

Extended Abstract

Latent Sector Error (LSE) is a well-known problem in HDD-based storage systems. LSEs, which occur silently, may result in data loss during RAID recovery from disk failure. LSEs in HDDs are caused by various reasons such as write errors or media imperfections [1], which result in bit/symbol errors that cannot be corrected with ECC. Disk scrubbing is used to detect LSEs by scrubbing the disk in the background. As pointed out in [2], the scrubbing strategy optimization requires a good model of LSE development, i.e., when and where LSE would likely happen. Inspired by previous work [1] [2], we are investigating the problem of modeling and detecting LSEs of NAND flash-based SSDs.

Due to increased density, NAND flash memory is becoming more and more prone to bit errors [3], which increases the probability of LSE hazards in SSDs. For example, reduced feature size potentially shrinks the volume of the floating gate, which is dedicated to store the electric charge. Therefore, the threshold voltage differences (determined by the amount of charge in the cell) among the cell levels are reduced. In addition, for the MLC technique, the more bits per cell, the more cell levels are there to share the threshold window. Resulted from the feature scaling and MLC, the reduced charge difference between cell levels are more vulnerable to errors caused by noise or disturbs. In addition, with a thinner oxide layer that isolates the floating gate, the feature size scaling amplifies the impact of P/E cycling, which introduces bit errors and reduces the lifetime of the flash. The high bit error rate may be addressed by stronger ECC. However, dealing with large-size sectors (due to MLC), most ECC schemes require more extra bits to store ECC. In addition, the ECC decoding latency grows with increased codeword length and ECC strength [4].

The model of LSE development in HDDs is built upon the real field data, which are not available for SSDs due to the limited population. However, we may model LSE according to the underlying mechanism of NAND flash bit errors. Specifically, the model we are currently working on considers the following factors:

1) The first and leading factor in bit error development of NAND flash is P/E cycling of individual flash block/page. P/E cycling degrades and eventually breaks down the oxide layer that insulates the floating gate in the flash cell. The manifestation of P/E cycling's impact on the threshold voltage of the flash cell is that the difference between the *erased* state and *programmed* state shrinks as the number of cycling increases. The bit error rate is linearly related to P/E cycling, which may be recorded using the wear information collected by the wear-leveling algorithm.
2) The second factor is retention, during which the charge status (thus the data) of the flash evolves. The threshold voltage is affected mostly by charge loss, which is more severe if the flash cell has a larger number of P/E cycling. Intuitively, we can keep track of retention using a time stamp that marks the creation of the data on a flash page.
3) The third factor is disturbs caused by read, programming, and erase operations, which are applied on the flash page directly or on the nearby units. Taking disturbs into account of the model is quite a challenge due to the fact that the impact of disturbs on the charge status is erratic, i.e., it could be either charge loss or charge gain.

To verify the model, intuitively, there are basically two ways: first, we may verify the LSE development features proposed by the model against the real field data, hoping they would be available soon; second, without the real field data, we may simulate the bit errors with known device characteristics.

In addition, the factors discussed above will be inspected with the processing technology of NAND flash. In particular, by taking into account the technology trend of NAND flash design, we aim at a guiding model of SSD's LSE development for the future. Specifically, modeling the impact of feature size scaling requires thorough understanding/knowledge about the electronic physics of NAND flash, while modeling the impact of MLC of increased bits/cell is more straightforward.

Qualitatively, the above factors reveal that LSE development of SSDs also exhibit temporal and spatial localities as in HDDs [2]. Specifically, aged pages (of high P/E cycles) have higher bit error rate, i.e., LSE would appear on these pages more frequently. The cold ones, which have long retention time, are also prone to LSEs. Due to the fact that the pages in the same block have the same P/E cycles, LSEs would likely happen in clusters on the flash blocks. Apparently, the disk scrubbing strategy for SSDs should leverage the localities by adopting the *localized* and *staggered* policies [1]. Based on the P/E cycling information, the scrubber should first scrub the aged blocks, and then decide upon the detection of an error on a flash page to immediately scrub the rest of the pages in the same block. In addition, for staggered scrubbing (to exploit the spatial locality), the scrubber may begin by reading the first page of each block, then the subsequent pages of each block. These two policies are demonstrated in Fig 1 and 2, respectively.

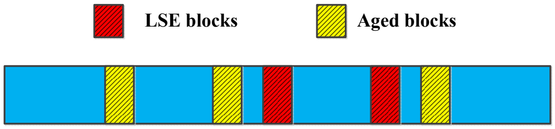

Fig. 1. Localized scrubbing is applied on two types of flash blocks: blocks that incurred LSEs and the aged blocks.

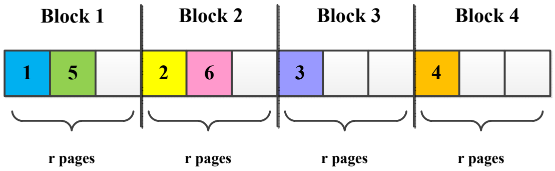

Fig. 2. Staggered scrubbing: the numbers marked on individual pages represent the order of scrubbing, i.e., the first page of each block is read at the beginning, followed by the subsequent pages.

The optimization of the scrubbing strategy requires to quantify these factors and to identify the complex correlation among them, which are the most challenging tasks of this work. For example, in addition to the duration, the impact of retention is also related to P/E cycling and the disturbs; the retention not only causes charge loss but also remedies the P/E cycling. Thus, the model must depict the full and quantified aspects of LSE development in SSDs.

To conclude, LSE development in SSDs is not merely a problem of the physical features of NAND flash but also a problem of data access patterns. The lack of real field data as well as the increased error rate due to higher density of NAND flash serve as the motivation of this work. Based on the preliminary investigations and our previous work on NAND flash reliability [5] [6], we are making progress towards the following objectives: 1) a comprehensive model of LSE development; 2) optimized disk scrubbing strategy guided by our model. We expect to report more results in the near future.

Acknowledgement

This research is sponsored in part by National Science Foundation grant CCF-0937799. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] B. Schroeder, S. Damouras, and P. Gill, "Understanding latent sector errors and how to protect against them," *ACM TOS*, vol. 6, no. 3, pp. 1–23, 2010.

[2] A. Oprea and A. Juels, "A clean-slate look at disk scrubbing," in *Proceedings of FAST'10*. USENIX Association, 2010, p. 5.

[3] J. Brewer and M. Gill, "Nonvolatile memory technologies with emphasis on flash," *IEEE Whiley-Interscience, Berlin*, 2007.

[4] T. Kgil, D. Roberts, and T. Mudge, "Improving nand flash based disk caches," in *Computer Architecture, 2008. ISCA'08. 35th International Symposium on*. IEEE, 2008, pp. 327–338.

[5] G. Wu, B. Eckart, and X. He, "BPAC: An adaptive write buffer management scheme for flash-based Solid State Drives," in *Proceedings of MSST'10*. IEEE, 2010, pp. 1–6.

[6] G. Wu, X. He, N. Xie, and T. Zhang, "DiffECC: Improving SSD Read Performance Using Differentiated Error Correction Coding Schemes," *Modeling, Analysis, and Simulation of Computer Systems, International Symposium on*, pp. 57–66, 2010.