The Techniques for exploiting the plane-level parallelism in NAND Flash based Storage Device

[1), 2)] Wontaeck Jung and [1)] Eui-Young Chung

[1)] *School of Electrical and Electronic Engineering, Yonsei University, Seoul 03722, Korea*

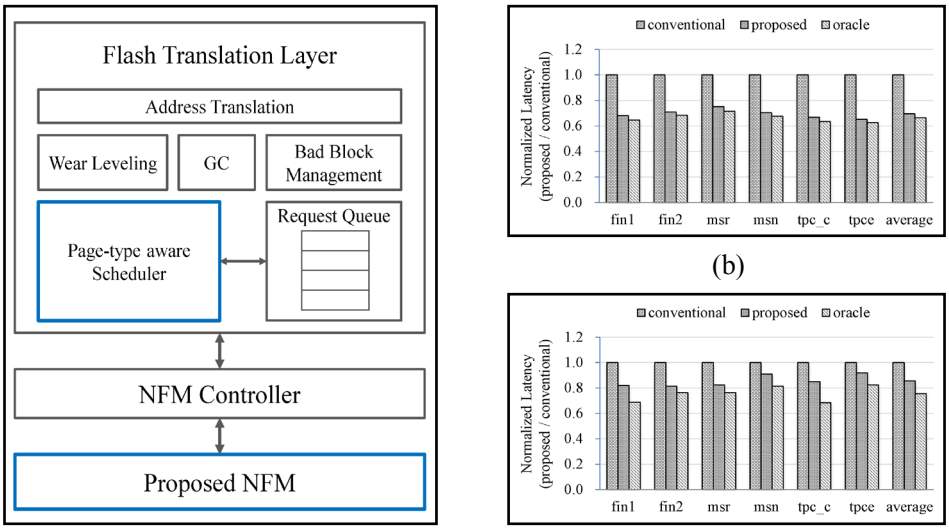
[2)] *Samsung Electronics Co. Ltd., Suwon 16677, Korea*

E-mail : rootian@dtl.yonsei.ac.kr

Recently, NAND flash memory-based storage systems (NFMS) are widely used in various applications. These systems are composed with many NAND Flash Memories (NFMs) and use multi-level parallelism in a concurrent or interleaving manner between NFMs to improve a performance. The multi-level parallelism is divided into three parts such as channel-level, way-level and plane-level. However, most of the previous works had focused on parallelism except plane-level because there are restrictions and complexity in using the multi-plane operation [1], [2]. The restriction is that the page address must be identical for multi-plane operations [3]. That implies that the physical location and the page-type, which is LSB and MSB, of the selected pages should be same. To overcome the restrictions, large area is required. However, the most of overhead is related to the page-type which determines voltage and timing.

In this paper, we propose a novel NFM architecture and a plane-level parallelism-aware-FTL. Fig 1(a) is shown the overall architecture of proposed NFMS. The proposed NFM is able to select different word-line, which is the physical location of page, between planes within a NFM by adding small circuits related to a row address. The plane-level parallelism-aware-FTL avoids the restriction on the page-type and increases the frequency of exploiting the plane-level parallelism by scheduling requests in the queue based on the page-type.

The evaluation results show that our proposed methods can achieve an average reduction of 31% in read latency and 14% in write latency, as shown in Fig.1(b) and (c). Within minimal area overhead, the proposed method follows closely the oracle which exploits plain-level parallelism for all requests. Consequently, the proposed methods improve a system performance by overcoming the restriction and exploiting the plane-level parallelism maximally.



(a) (b) (c)

Fig 1. (a) The overview of NFMS for proposed NFM, (b) Normalized read latency, (c) Normalized write latency

Acknowledgement: This work supported in part by the IC Design Education Center (IDEC), by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2013R1A1A2011208), and by Samsung Electronics.

[1] Guo, Xufeng, Jianfeng Tan, and Yuping Wang. "PAB: parallelism-aware buffer management scheme for NAND-based SSDs." Modeling, Analysis & Simulation of Computer and Telecommunication Systems (MASCOTS), 2013 IEEE 21st International Symposium on. IEEE, 2013..

[2] Kang, Jeong-Uk, et al. "A multi-channel architecture for high-performance NAND flash-based storage sys-tem." Journal of Systems Architecture 53.9 (2007): 644-658.

[3] NAND Flash Memory Datasheet, Micron Technology, Boise, ID, USA, 2009.