Generative Modeling of NAND Flash Memory Voltage Level

Ziwei Liu, Yi Liu and Paul Siegel

Center for Memory and Recording Research, University of California, San Diego

February 25, 2020

Table of Contents

Introduction

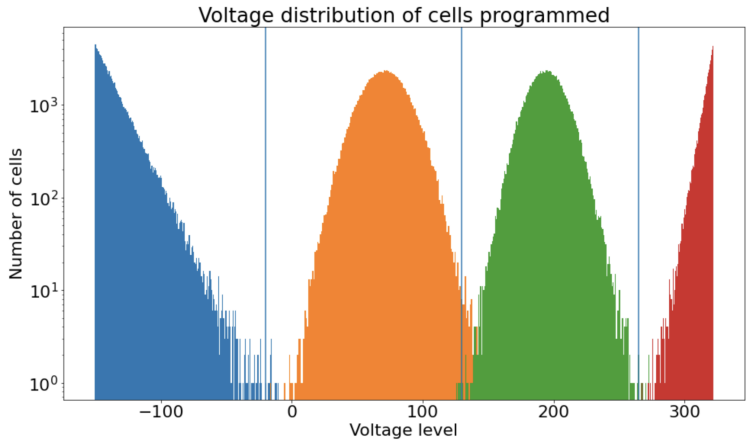
Motivation

- Both flash memory hardware characterization and error-correcting code design rely heavily on the flash memory cells' voltage level data obtained from the experiments conducted on flash memory devices.
- Yet the experiments to acquire such data is very time-consuming.

- Thus, we proposed a *voltage level generator* to generate a large amount of NAND flash memory cell's voltage level using a relatively small amount of measured such data.
- This generator can generate authentic voltage level distributions over a range of possible Program/Erase cycles and for each specified program level after learning from measured target distributions.

The Experimental Dataset

- The experimental dataset used in this demonstration is obtained from a Normal-Laplace based model of MLC flash memory voltage levels distribution.
- This model is proposed T. Parnell et al. in 2014 for sub-20nm NAND flash memory. Parnell et al. (2014)
- The Normal-Laplace based model could generate realistic MLC NAND flash memory voltage level distributions for Program/Erase (P/E) cycles from 10 to 1000 and for each program level.



- When the training is enforced, we draw the same amount of samples from the Normal-Laplace generator and our deep learning-based generator, and then compute the loss function.

Generative Moments Matching Network (GMMN)

- We model such voltage level distributions using the generative moments matching network. Li et al. (2015)
- Input to the network: samples drawn from normal distribution of feature size $1 \Rightarrow$ the generated samples are i.i.d.
- Target: output collectively assembles the true voltage level distribution.
- Measurement of the mean squared difference between generated and target distributions, i.e. the loss function: maximum mean discrepancy.

$$\mathcal{L}_{MMD^2} = \left\| \frac{1}{N} \sum_{i=1}^{N} \phi(x_i) - \frac{1}{M} \sum_{j=1}^{M} \phi(y_i) \right\|^2 \tag{1}$$

$$\mathcal{L}_{MMD^2} = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{i'=1}^{N} \phi(x_i)^T \phi(x_{i'}) - \frac{2}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} \phi(x_i)^T \phi(y_j) + \frac{1}{M^2} \sum_{j=1}^{M} \sum_{j'=1}^{M} \phi(y_j)^T \phi(y_{j'}) \tag{2}$$

Generative Moments Matching Network (GMMN)

$$\mathcal{L}_{MMD^2} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{i'=1}^{N} k(x_i, x_{i'}) - \frac{2}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} k(x_i, y_j) + \frac{1}{M^2}\sum_{j=1}^{M}\sum_{j'=1}^{M} k(y_j, y_{j'}) \tag{3}$$

- Then we utilize the Gaussian kernel trick to implicitly lifts the sample vectors into an infinite dimensional feature space. Gretton et al. (2006)

$$k(x, x') = \exp(-\frac{1}{2\sigma}\left|x - x'\right|^2) \tag{4}$$

- For Gaussian kernel, we can use a Taylor expansion to get an explicit feature map that contains an infinite number of terms and covers all orders of statistics.
- Thus, minimizing MMD in this setting is equivalent to minimizing a distance between all moments of the two distributions.

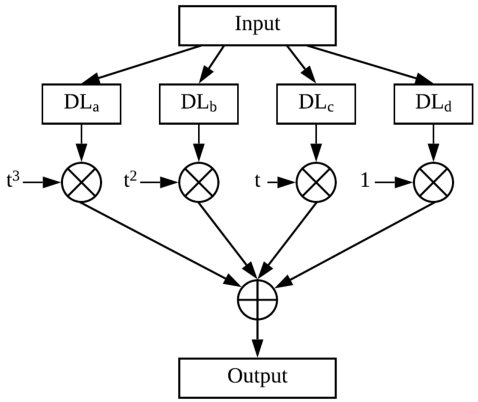
Time-dependent neural network

- Observation: voltage level distributions at different P/E cycles are different because of the program disturb effects and the wear-out of flash memory cells.
- A time-dependent generative moments matching network to capture the time-varying property. Holden et al. (2017) Liu et al. (2020)
- Each time-dependent dense layer: $F = \sum_{i=1}^{n} A_i(t)N_i(t) + B(t)$
- The weights and biases are estimated using degree-3 polynomial functions

$$A_i(t) = a_1 t^3 + a_2 t^2 + a_3 t + a_4 \tag{5}$$

$$A_i(t) \Rightarrow a_{i1}, a_{i2}, a_{i3}, a_{i4} \; B_i(t) \Rightarrow b_1, b_2, b_3, b_4 \tag{6}$$

- Training: Determine each continuous function for weight or bias using 4 parameters.



$$t = \frac{\text{present P/E cycle} - \text{minimum P/E cycle}}{\text{maximum P/E cycle} - \text{minimum P/E cycle}}$$

Time-dependent GMMN

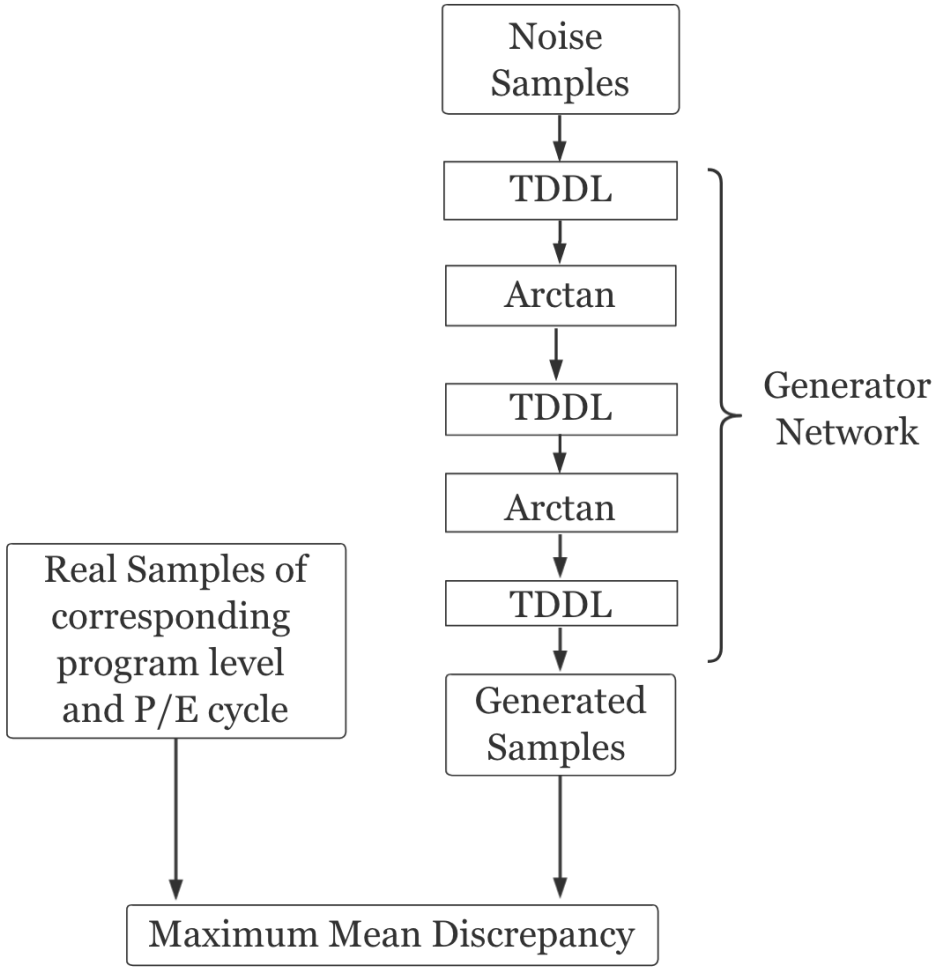
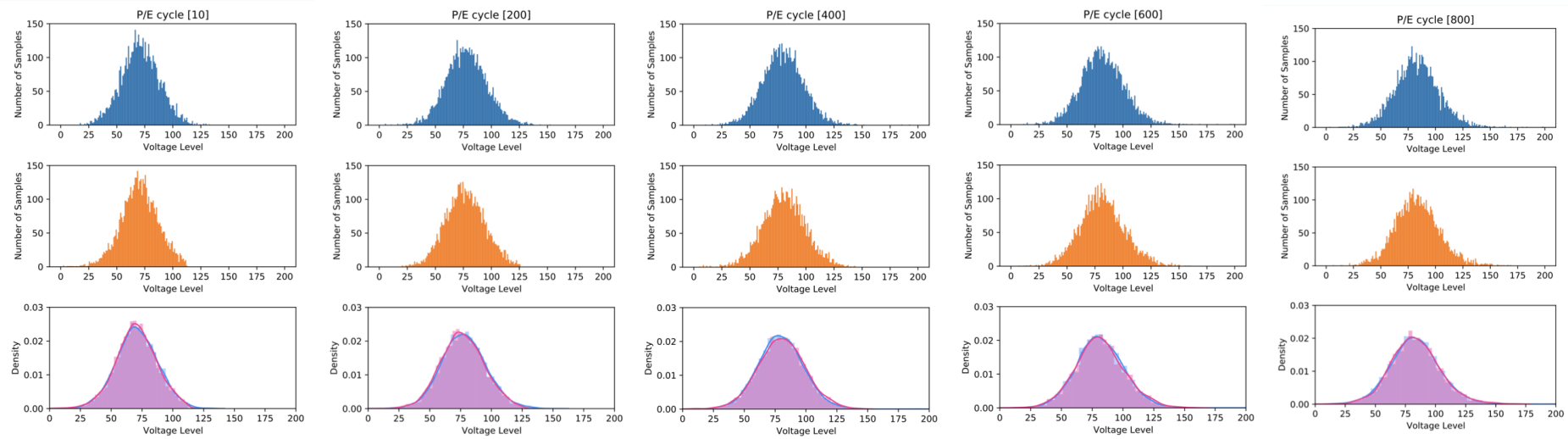


Figure: The generation workflow

Experimental Results



P/E Cycle 10 P/E Cycle 200 P/E Cycle 400 P/E Cycle 600 P/E Cycle 800

Figure: The histogram of samples drawn from the Normal-Laplace model (top), generated voltage levels (middle), and the estimated probability density function for both (bottom) for MLC program level 1 at P/E cycles 10, 200, 400, 600 and 800, separately. Number of samples: 5,000.

References

Gretton, A., Borgwardt, K. M., Rasch, M., Schölkopf, B., and Smola, A. J. (2006). A kernel method for the two-sample-problem. NIPS'06, page 513–520, Cambridge, MA, USA. MIT Press.

Holden, D., Komura, T., and Saito, J. (2017). Phase-functioned neural networks for character control. *ACM Trans. Graph.*, 36(4).

Li, Y., Swersky, K., and Zemel, R. (2015). Generative moment matching networks. In *International Conference on Machine Learning*, pages 1718–1727.

Liu, Y., Wu, S., and Siegel, P. (2020). Bad page detector for NAND flash memory. In *11th Annual Non-Volatile Memories Workshop (NVMW)*.

Parnell, T., Papandreou, N., Mittelholzer, T., and Pozidis, H. (2014). Modelling of the threshold voltage distributions of sub-20nm NAND flash memory. In *2014 IEEE Global Communications Conference*, pages 2351–2356.