(i.e., up to 25k P/E cycles), the distribution mean and standard deviation can be modeled to increase *linearly* with P/E cycles:

$$V_{th}^{mean,std}(PEcycle) = D + E \times PEcycle \qquad (5)$$

where D and E are constant coefficients. The average accuracy of the linear model for the mean value is 89.4%, 93% and 92%, and for the standard deviation it is 95.9%, 94.5% and 89.1%. The linear model is not as accurate as the exponential model, but it is much simpler and can still achieve > 89% accuracy.

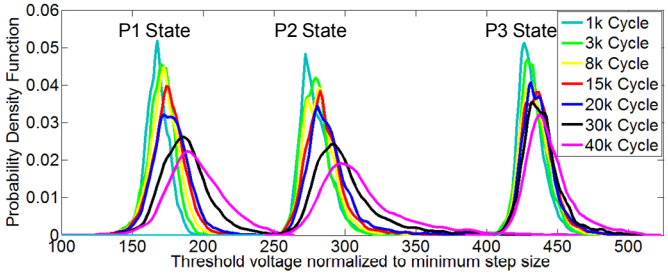

Fig. 11 Threshold voltage distribution under various P/E cycles using non-parametric kernel density estimation

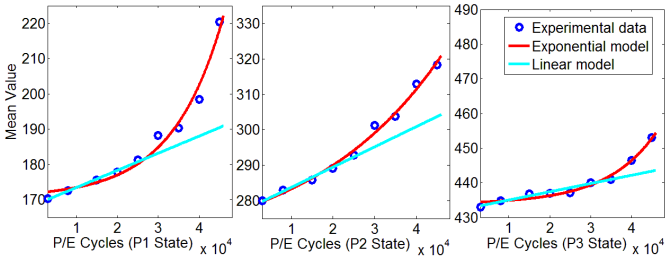

Fig. 12 Mean value (μ) of voltage distributions vs. P/E cycles

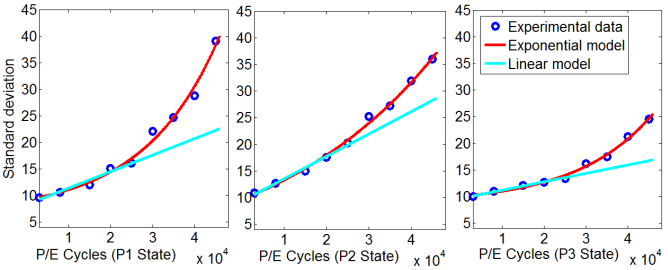

Fig. 13 Standard deviation value (σ) of voltage distributions vs. P/E cycles

The standard deviations for all three programmed states are close to each other at 3k P/E cycles, which is just at the end of flash nominal lifetime endurance. This means that ISPP is carefully designed to guarantee equal distribution width for all states within the nominal flash endurance. However, the rates of increase in both mean and variance beyond the nominal flash endurance vary among different programming states. The mean and variance change over P/E cycles for the distribution for the P1 or P2 states is much larger than the mean and variance change over the same number of P/E cycles for the P3 State. In other words, the voltage distribution of a state with a lower threshold voltage gets distorted much more with P/E cycles (this effect can be observed in Fig. 11 as well, which shows the distributions for states P1 and P2 shifting and widening much more than that for state P3 as P/E cycles increase). This is because a flash cell in the state with a higher threshold voltage has more electrons on the floating gate and the effective electric field across the tunnel oxide is smaller compared to a cell in a state with a low threshold voltage. Thus, given the same condition for the last iteration of ISPP, a cell with a high threshold voltage (i.e., in the P3 state) is less likely to get more electrons programmed than a cell with a low threshold voltage (i.e., in the P2 or P1 state).

Signal-to-Noise Ratio Analysis: We can model the 2-bit MLC NAND flash programming as 4-state pulse-amplitude modulation (PAM) in digital communication [17]. The distance between the mean values of neighboring states can be modeled as the signal. The variance of the distributions can be modeled as the noise. The noise increases with P/E cycles, and as we can see in Fig. 13, the distribution variances (or standard deviation) increase with P/E cycles. An important property of communication channels, the signal-to-noise-ratio (SNR) is shown in Fig. 14. We can see that the SNR drops linearly as P/E cycles increase. From 3k P/E cycles to 45k P/E cycles, the SNR drops from about 11dB to 5.5dB. On average, the SNR inside the flash memory deteriorates about 0.13dB every 1k P/E cycles. This information can be used to develop adaptive error correction mechanisms that provide higher error correction strength with the number of P/E cycles, an idea whose exploration we leave for future work.

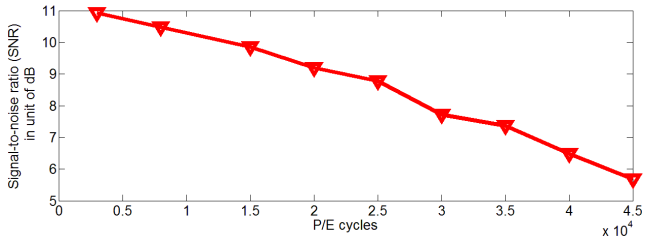

Fig. 14 Signal-to-noise ratio (SNR) vs. P/E cycles

VI. Conclusions

We have characterized, analyzed, and developed models for threshold voltage distributions in state-of-the-art 2Y-nm NAND flash memory, with the goal of enabling predictive models that can aid the design of more sophisticated error correction methods, such as LDPC codes, in future flash memories. Our experimental evaluations show that the developed model can accurately predict the changes in the threshold voltage distribution as the number of P/E cycles increases. As flash memory continues to scale to smaller feature sizes, we hope that the characterization, understanding, and models provided in this work would enable the design of new and more effective error tolerance mechanisms that can make use of the observed characteristics and the developed models.

Acknowledgments

We thank Justin Meza of CMU for help with writing and clarification of an earlier draft of this paper.

References

[1] A. Maislos et al., "A New Era in Embedded Flash Memory", FMS 2011.
[2] Shu Lin and D. J. Costello, Error control coding, Prentice Hall, 2004.
[3] H. Choi et al. "VLSI Implementation of BCH Error Correction for Multilevel Cell NAND Flash Memory", IEEE Transactions on VLSI, 2010.
[4] Y. Cai et al., "Error patterns in MLC NAND flash memory: measurement, characterization, and analysis", DATE 2012.
[5] Y. Cai et al., "Flash Correct-and-Refresh: Retention-Aware Error Management for Increased Flash Memory Lifetime", ICCD 2012.
[6] R. G. Gallager, Low-Density Parity Check Codes. Cambridge: MIT Press, 1963.
[7] T. Richardson et al, "Design of capacity-approaching irregular low-density parity-check codes", IEEE Transaction on Information Theory, pp: 616-637, 2001.
[8] Y. Lee, DSP: Shaping with Digital Storage World, 2012.
[9] G. Dong et al, "On the Use of Soft-Decision Error-Correction Codes in NAND Flash Memory", IEEE Transaction on Circuits and Systems-I, pp. 429-439, 2011.
[10] S. Chen et al., "Reliability Analysis and Improvement for Multi-Level Non-Volatile Memories with Soft Information", DAC 2011.
[11] N. Mielke et al, "Bit error rate in NAND Flash memories", IRPS 2008.
[12] H. Shim et al., "Highly Reliable 26nm 64Gb MLC E2NAND (Embedded-ECC & Enhanced-efficiency) Flash Memory with MSP (Memory Signal Processing Controller)", VLSI 2011.
[13] J. Yang, "High-Efficiency SSD for Reliable Data Storage Systems", FMS 2011.
[14] Y. Cai et al., "FPGA-based Solid-State Drive prototyping platform", FCCM 2011.
[15] K.-D. Suh et al., "A 3.3 V 32 Mb NAND flash memory with incremental step pulse programming scheme," JSSC, 1995.
[16] Kirk Prall, "Scaling non-volatile memory below 30nm", NVMW 2007.
[17] J. Proakis et al., Digital Communication, McGraw-Hill Companies, Inc. 2007