Secure HfO_2 Based 10-kb EEPROM for Data Security Applications

Cheng Hao
School of Electrical and Computer Engineering
Oklahoma State University
Stillwater, OK 74078-5032
Email: cheng.hao@okstate.edu

***Abstract*—Secure systems rely on encryption keys to be stored safely. However, it has been shown that popular storage means such as DRAM or PROM possesses vulnerability to physical cryptographic attacks. This work proposed an on-chip secure 10-kb EEPROM with adjustable programming voltages using charge trapping mechanism. The on-chip feature allows encryption keys to be stored with the processor minimizing observability. The use of die dependent programming voltages enables the optimization of data retention time verses memory lifetime. A quick future view on modeling EEPROM lifetime and data retention time tradeoff is given as well.**

Keywords—charge trapping; HfO_2; high-k dielectric; EEPROM; programming; encryption; cyber security.

I. Introduction

High demands on data security lead to the encryption techniques. For most systems encryption keys are stored in local memory such as DRAM and PROM. After power removal gradual discharge of the storage cell capacitors results a finite amount of time before stored contents vanishing completely. This time is sufficient for adversaries to retrieve the content using techniques combined with physical access to the memory. Halderman et al. have shown seconds of data retention time of DRAM after the power is cut or removed from the motherboard at normal operating temperatures [1]. Data retention time can be prolonged to minutes when the memory die was kept at lower temperatures [1]. Peterson has provided a solution to this problem, called cryptkeeper, by dividing RAM into two different sized segments [2], one small and insecure segment for content access [2], the other larger segment for encrypted data storage [2]. However, this method still requires encryption key storage means in order to decode the stored contents. Therefore, a more promising solution is in order. The work presented in this paper proposed an on-chip 10-kb secure EEPROM employing charge trapping mechanism with 32nm HfO_2 based SOI CMOS process. The architecture presented here includes a full data communication interface between the CPU and EEPROM core, on-chip LDOs for programming voltages, write/erase, generation and off-chip DACs allowing optimal programming protocol execution. In addition to the EEPROM architecture, a quick view on modeling EEPROM lifetime and data retention time tradeoff is provided as future work. Section II of this paper briefly reviews the trapping mechanism of HfO_2 based transistor. Section III introduces the proposed EEPROM structure and its constituent parts. Section IV presents the implementation of the EEPROM. Section V gives a short future view on modeling charge trap based EEPROM lifetime and data retention time tradeoff and is followed by a conclusion in section VI.

The content of this work is appropriate to distributed within the U.S.

II. Charge Trapping Mechanism

Kothandaraman et al. [3] and Cartier et al. [4] have reported 32nm CMOS process with HfO_2 gate dielectric possesses programmability that suits for developing medium to high density EEPROMs. This approach enables on-chip hardware implementation and eliminates chemical etch back as a means of decoding and physical eavesdropping of the circuits.

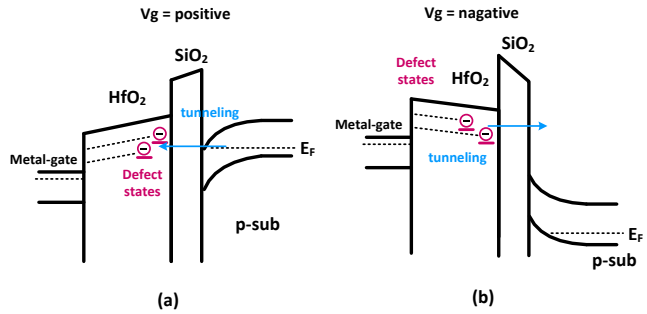


Fig. 1. Energy band diagram describing charge trapping mechanism for NMOS. (a) Positive gate voltage results carrier tunneling through the SiO_2 interfacial layer and trapped into the HK HfO_2 layer. (b) Negative gate voltage results carriers detrap from the HfO_2 layer.

The storage cells use HfO_2 based high-k metal gate (HKMG) transistor which can support multiple time programming (write and erase) [5]. The device threshold voltage, V_{th}, can be shifted by trapping and detrapping carriers in and from the high-k (HK) dielectric layer shown by fig. 1. These traps are due to oxygen vacancies in the HfO_2 HK layer [3]. Compared to floating gate devices, carriers are trapped in the dielectric layer rather than conductive gate electrode [6]. The retention of charge depends on dielectric material, thickness, quality and temperature [46][47]. The programming of the device is achieved by applying positive gate bias and elevated drain voltage causing energetic carriers tunnel through

DISTRIBUTION STATEMENT A.
Approved for public release: distribution is unlimited.

the SiO_2 energy barrier and trapped into the HK dielectric layer and results an increase in V_{th} [7]. When negative gate voltage is applied, trapped carriers are freed from traps resulting V_{th} reduction to the value close to its native [5].

III. THE 10 KB EEPROM

This section discusses the proposed EEPROM overall architecture, storage cell, operations, and programming voltage generation circuits.

A. Architecture

The EEPROM is comprised of four 256 words by 88 bits banks as the storage core, three input registers for data and command buffering, CUI (command user interface) for EEPROM-CPU communication, operation controller, and LDOs for programming voltage generation. Fig. 2 presents the top level block diagram of the complete design. The 88 bit words are expanded from 80 bits to support ECC [8]. The ECC function is reserved for implementation in future versions. Write and erase logic are developed with 2.5V 160nm legacy SOI CMOS while the storage core and I/O registers are 32nm HfO_2 CMOS. Three input registers are used for control, an initialization register, a command register, and a 16 words data register. The initialization register contains clock scaling byte and programming voltage amplitudes. The command register contains memory address, programming voltage stabilization time, application time and timeout for write, read and erase operations, the BUSY bit and memory operation command in binary values. The 16 words data register serves the role of data buffer which can be accessed independently by the EEPROM and the CPU under the control of CUI. The CUI logic provides access scheduling between the EEPROM and the CPU to the data register. The BUSY bit in the command is dedicated to indicate data register availability and serves as a semaphore flag. When the BUSY bit is reset, CPU has the access to the data register. EEPROM and CPU cannot access the data register simultaneously to avoid data conflict.

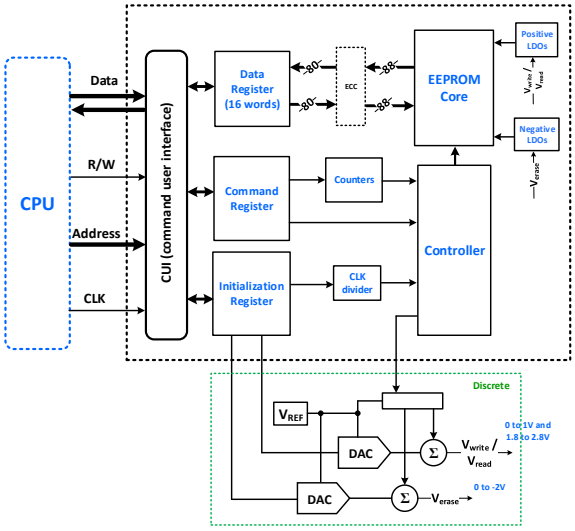


Fig.2. EEPROM architecture block diagram with discrete programming voltage control block.

This design utilizes off-chip DACs (MAX531) for controlling programming voltages via on die LDOs as device programming is better understood in order to balance the tradeoff between data retention and EEPROM lifetime. DACs together with on-chip LDOs provide accurate precision programming voltages for write/erase to achieve fine tuning of the programming voltages so that memory lifetime and data retention can be balanced. The proposed write/erase voltage ranges are from 1.5V to 2.7V and minus 2 to minus 3V. Read voltage is from 0.5V to 0.9V. In addition of setting precision programming voltages, the DACs allow die process threshold measurement for making smarter programming voltage decisions.

B. Storage Cell

The proposed storage cell uses differential structure implementation which is similar to [3] and [7]. The differential cell consists of high-threshold logic NMOSs, MC_1 and MC_2 whose gates are tied together forming word line. Fig. 3 shows the cell diagram. Two thick-oxide high-voltage NMMOS, M_1 and M_2 in each column are programming assist devices. The tail voltage, V_{TP}, serves as drain of the storage pair during wire and erase operations and source during read operation. During write V_{TP} is 1.5V providing a source of hot carriers while the gate of the storage pair is 2.5V as the nominal gate programming voltage. Either M_1 or M_2 is turned on depending on data to be written. The side with M_1 (or M_2) turned on supports trapping carriers into the gate dielectric. During erase gate voltage of the cell is -1.5V while V_{TP} is 1.5V. In this manner the cell is reversely biased with 3V detrapping the carriers. Both M_1 and M_2 are off. During read V_{TP} is pulled to V_{SS} with M_1 and M_2 are turned off. Word line is pulled to 0.9V. Current difference in the bit line (BL) and bit line bar (BL_bar) is applied to the sense amplifier to determine the stored logic.

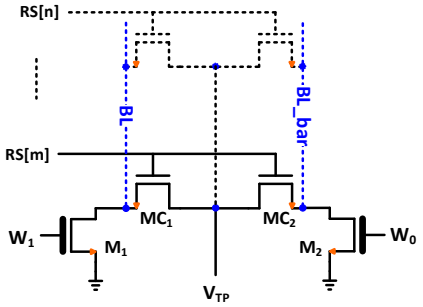


Fig.3. Differential storage cell (MC_1 and MC_2) circuit diagram.

C. Operation: Safe Write Erase and Read

Memory operation begins with the CPU loading the three input control registers. The loading order is the initialization register, the command register followed by the data register. During the fist load, the BUSY bit in the command register should be logic 0 so that CPU has access to the data register. After data register loading is complete, CPU reload command register and sets the BUSY bit while other register contents remain unchanged. With the BUSY bit set, EEPROM operation starts at the next raising clock edge. Write, read, and erase (W/R/E) are carried out in a similar manner. Based on the content of initialization register, EEPROM clock and programming voltages from the DACs are set. The EEPROM

address in the command register is decoded. Voltage stabilization counter starts counting down for programming voltages at the LDO outputs to be stable. When the programming voltages are stable, voltage is applied and application counter starts. Counter contents specify the duration or application time of the programing voltage. After programming is complete, timeout counter starts. Timeout duration controls the quench time of the programming voltages. This ensures no potential over voltages or circuit shorting occurs. This implementation supports single word write, a block (16 words) write and block erase. After write or erase operation is complete, read operation starts automatically to support confirmation of a valid operation. The contents at memory locations just written or erased are read back and stored in the data register for CPU to recover and verify. When EERPOM operation is complete, CPU can recover data from data register or over write with new contents.

The storage cells are implemented with thin oxide logic devices and programmed using thick oxide high voltage devices. Storage cells have four operational states: low or no voltage idle state, low voltage read state, high voltage write state and high voltage erase state. These four states are summarized in fig. 4. Fig. 4(a) depicts the idle state programm-

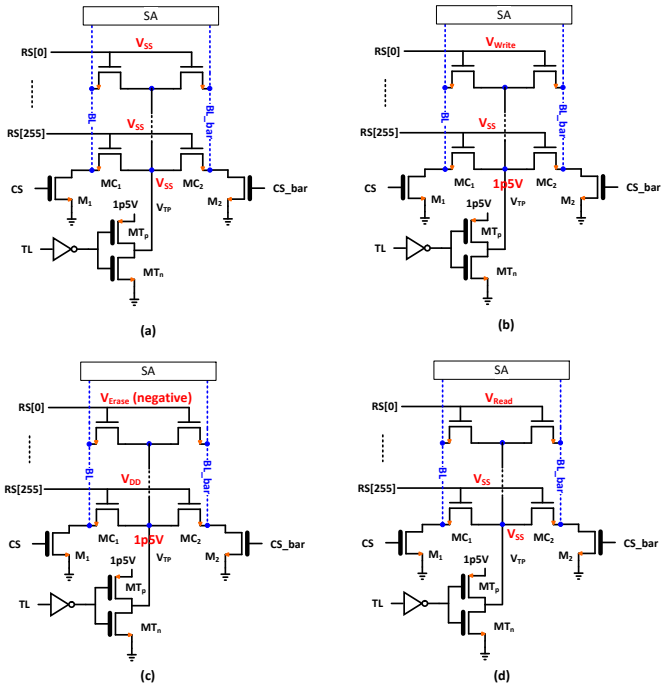


Fig. 4. Circuit of a column of storage cell structure with sense amplifier block. (a) Idle state programming voltage conditions. (b) Write operation voltage conditions. (c) Erase operation voltage conditions. (d) Read operation voltage condition.

-ing node voltage conditions. When the cell is in idle state, all word lines (RS[n]) are at zero volts. The drain of all the storage pair transistors is pulled to V_{SS} by the tail inverter composed by MT_p and MT_n. M_1 and M_2 are turned off. Write operation voltages are shown in fig. 4(b). During write operation the drain of the storage pair is pulled to 1.5V by the tail inverter providing a source of hot carriers for trapping. The word line (i.e. the gate of storage transistor) of the cell being written is supplied by write voltage, V_{write}. Either M_1 or M_2 will be turned on during write operation depending on the data applied to trap charge on one side of the differential cell. The side written with trapped charge will have positive threshold shift relative to the unwritten side. Fig. 4(c) summarizes the erase voltage of the storage cell. The gate of the cell being erased can be supplied by voltage in the range of 0V to -1.5V. The drain of the storage pair is 1.5V. Both M_1 and M_2 are turned on. This results in a reverse bias in the range of -1.5V to -3V on the storage transistors to de-trap the carrier from the high-k gate dielectric. Fig. 4(d) shows the cell voltage condition for read operation. The gate voltage of the word being read is at V_{DD} and at V_{SS} for non-reading words. The drain of the storage cell is pulled to V_{SS}. Both M_1 and M_2 are off. When the sense amplifier is enabled the current difference in the bit lines due to programing threshold difference will be sensed to reconstruct the written data.

All write, read and erase operations associated with high voltages are mutually exclusive to avoid excess voltage exposure of the storage array. Appropriate programming voltages are activated through handshake mechanism by the controller and applied for a duration defined by predetermined counter values. Storage cells bias is provided by thick oxide high voltage devices forming voltage steering circuits and high voltage level shifters. The high voltage level shifters are used for developing gate stress voltages for write and erase operations.

D. Programming Voltage Generation

Programming voltages are generated by on-chip low-dropout voltage regulators (LDO) and off chip DACs. The DACs provide adjustable programming voltages that allowing for optimal programming that enables the balance of data retention time and EEPROM lifetime. Higher write/erase voltages imposes greater stress on the gate dielectric material that leads to quick wear out or shorter life span. Therefore, we employed voltage ramp programming fashion using DACs to achieve precise control to the programming voltages. Fig. 5(a)

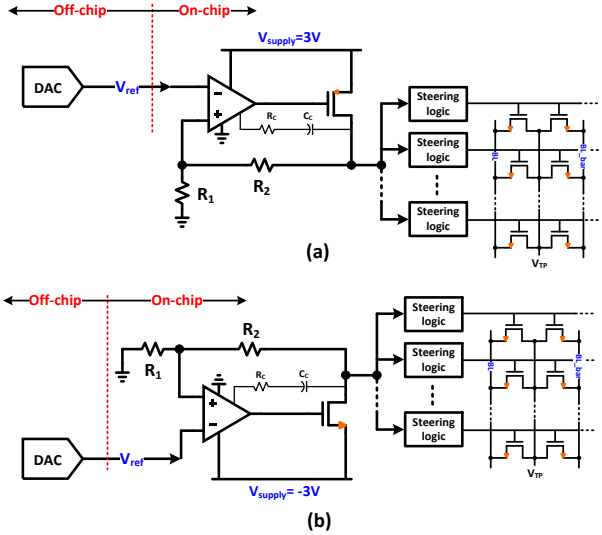


Fig. 5. On-chip LDO and off-chip DAC generate EEPROM programming voltages. (a) LDO for write and read voltages. (b) LDO for erase voltages.

and (b) show the LDO and DAC circuit diagram for generating the programming voltages. LDO in fig. 5 provides positive output voltages for write and read operation and negative output voltage for erase. The pioneer work [7] and [3] proposed a nominal programming voltage of 2V lasting 10ms for approximately 200mV threshold voltage shift. The erase voltage is -3V across storage transistors gate and source. In order to extract more accurate programming voltages for balancing data retention and wear out, we proposed a write gate voltage range from 1.5V to 2.7V and erase gate voltage range from -2V to -3V.

The presence of threshold variation across different EEPROM dies on a wafer due to CMOS process variation has non-negligible impact on lifetime. A given programming voltage has different stress strengths on different dies because of the process threshold differences. The statistical nature of the threshold variation allows us to use the precision DAC voltages to measure the threshold of each die. The programming voltage can then be adjusted so that the lower threshold dies experience alleviated stress. Therefore, the lifetime can be improved. The accuracy of the threshold measurement determines the amount of lifetime improvement. The idea can be illustrated using the time-to-failure equation provided in 1/E model [14] [18–21]

$$T_{BD} \cong \tau_0 e^{\frac{K\, t_{ox}}{V_g - V_{th}}} \tag{1}$$

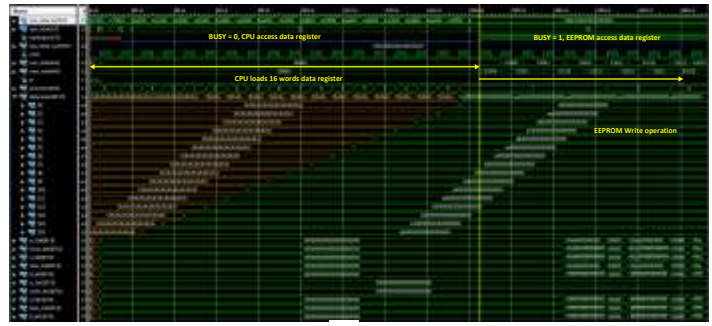
For a given CMOS process, the threshold variation, δV_{th}, is known from the PDK. The percentage lifetime improvement can be obtained using the following equation

$$\ln\left(1 + T_{BD,improve}\%\right) = \frac{1}{V_g - \delta V_{th} \mp \delta V_{th} \cdot V_{th,measure\ accuracy}\%} - \frac{1}{V_g - \delta V_{th}} \tag{2}$$

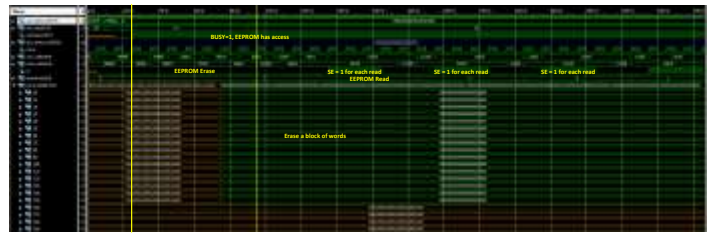
One thing needs to mention is the 1/E model was derived for SiO_2 dielectric material. The coefficients are yet to be determined for HfO_2 dielectric in our next stage of work. But we anticipate the exponential relation will be adopted for HfO_2 dielectric.

IV. SIMULATION RESULTS

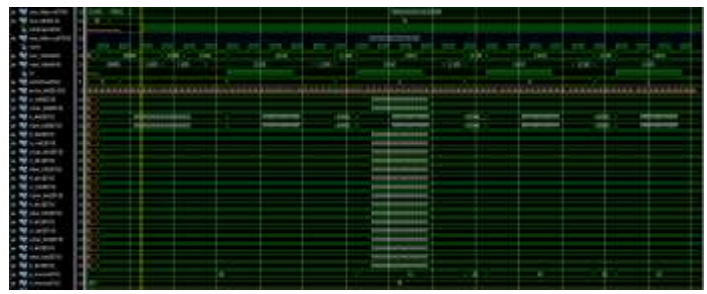
In order to validate the logic correctness of the interface between EEPROM core and CPU, the EEPROM core is modeled by four 88x256 arrays as the four banks whose write, erase and read operation are controlled by signals generated from the interface logic. The design is implemented using VHDL. Write operation is mimicked by assigning input values to the specified array cells based on asserted write control signals and address. Erase operation is simulated by assigning zeros to the corresponding array cells. Read operation is modeled by displaying the specified array cells value when read signals and address are present. Fig. 6(a) shows the waveform of writing 16 words from the data register to the EEPROM array. Fig. 6(b) shows the erase operation of a block (16 words) of words. Fig. 6(c) shows the interface control signals for read operation.



(a)



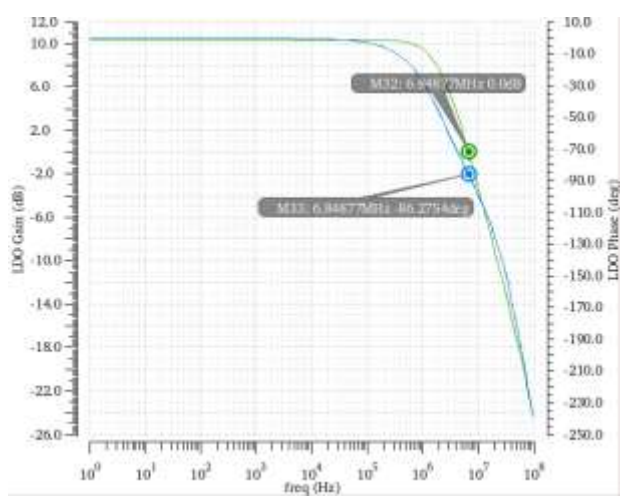
(b)



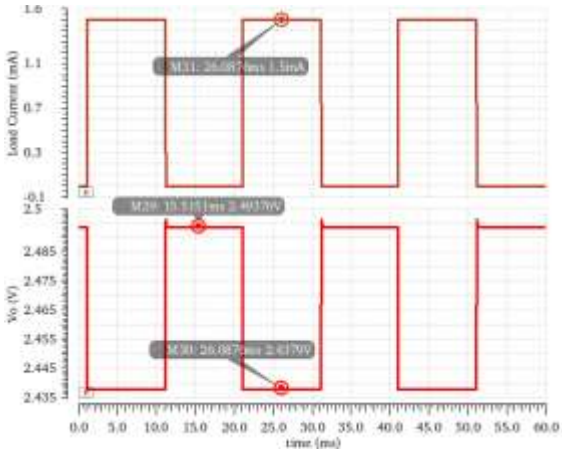
(c)

Fig. 6. (a) Simulation of writing 16 words from data register to EERPOM bank. (b) Simulation of erasing a block (16 words) of words. (c) Simulation of control signals for read operation.

Fig. 7 shows the LDO closed-loop AC and output transient simulations. The LDO has gain of 3.3 and 93.7 degree phase margin. It supports output voltage range from 0.5V to 2.5V.



(a)



(b)

Fig. 7(a) LDO closed-loop Bode magnitude and phase. (b) Output transient of the LDO with 2.5V output and 1.5mA output current.

V. FUTURE WORK

Research groups [3,5,7] and [9] have demonstrated the feasibility of non-volatile memory applications using the charge trapping transistors and extrapolated the loss of threshold shift only 16% over 10 years [3]. High dielectric transistor TDDB (time dependent dielectric breakdown) has been studied [10–40]. However, at this point, EEPROM cell lifetime (or wear out) retention tradeoff with programming voltages have yet to be addressed. The goal of our next stage of work is to develop a programming model that predicts the trap based EEPROM lifetime verses data retention time for different programming voltages while expecting lifetime improvement. Fig. 8 shows the conceptual block diagram of the proposed model. The comprehensive model will integrate two independent building blocks, TDDB model and trap charge retention model generating retention time and time to failure of the cell transistors.

Long term charge retention loss, required read voltage, and input offset voltages of storage cells and sense amplifier define the programming threshold shift requirement for successful read operations after a storage period. Knowing the gate stack capacitor model and the programming threshold shift, the amount of trapped charge can be estimated. Combining obtained the trapped charge with desired trap energy [41], gate stack thickness and Fowler-Nordheim tunneling model [42–45], programming voltages can be established. Storage cell lifetime can be extrapolated using TDDB model for various programming voltages. McPherson [14] and Padovani [15] have summarized frequently used TDDB models. Determining which model or combination of models with different weighting will be part of the next stage of work. With our current knowledge, the retention time model is a modified version of existing model that suit for HfO_2 dielectric.

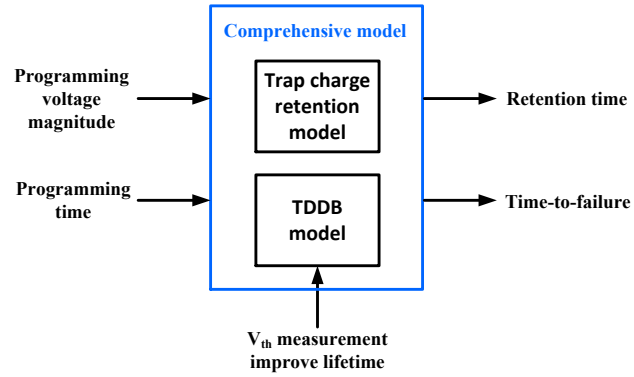


Fig. 8 Conceptual block diagram of EEPROM lifetime and data retention time tradeoff model.

VI. CONCLUSION

This work proposed a highly secured on-chip EEPROM architecture in order to solve the memory vulnerability issues caused by physical cryptographic attacks. This work fully implemented on-chip CPU-EEPROM interface and adjustable programming voltage generation block where DAC programmable LDOs are used to develop and control write/read/erase voltages ensuring more optimal programming voltages offering greater device yields and longer EEPROM lifetime. A model that enables to balance the tradeoff between EEPROM lifetime and data retention time for various programming voltages will be developed through the following work effort.

ACKNOWLEDGMENT

Author thanks AFRL for the funding of this project. Author thanks Dr. Chris Hutchens for the technical guidance throughout the work. Also thanks Mr. Vishal Reddy for his assistance on VHDL coding.

REFERENCES

[1] J. Halderman et al., “Lest we remember: cold-boot attacks on encryption keys,” Communications of the ACM, vol. 52(5), pp. 91–98, May 2009.

[2] P. A. H. Peterson, “Cryptkeeper: Improving security with encrypted RAM,” 2010 IEEE International Conference on Technologies for Homeland Security (HST), Waltham, MA, pp.120–126, 2010.

[3] C. Kothandaraman et al, “Oxygen vacancy traps in Hi-K/Metal gate technologies and their potential for embedded memory applications,” 2015 IEEE Internatioinal Reliability Physics Symposium, Monterey, CA, pp. MY.2.1-MY.2.4., 2015.

[4] E. Cartier, B. P. Linder, V. Narayanan and V. K. Paruchuri, “Fundamental understanding and optimization of PBTI in nFETs with SiO2/HfO2 gate stack,” 2006 International Electron Devices Meeting, San Francisco, CA, pp. 1-4, 2006.

[5] F. Khan, E. Cartier, J. C. S. Woo, and S. S. Iyer, “Charge trap transistor (CTT): an embedded fully logic-compatible multiple-time programmable non-volatile memory element for high-k-metal-gate CMOS technologies,” IEEE Electron Device Letters, vol. 38, no. 1, pp. 44-47, Jan. 2017.

[6] J. M. Rabaey, A. Chandrakasan, B. Nikolic, Digital Integrated Circuits A Design Perspective, 2nd ed., Pearson Education, Inc., 2003, pp. 647-653.

[7] B. Jayaraman et al., “80-kb logic embedded high-k charge trap transistor-based multi-time-programmable memory with no added

process compliexity," IEEE Journal of Solid-State Circuits, vol. 53, no. 3, pp. 949-960, March 2018.

[8] T. Haifley, "Endurance of EEPROM with On-Chip Error Correction," in IEEE Transactions on Reliability, vol. R-36, no. 2, pp. 222-223, June 1987.

[9] J. Portal et al., "Design and Simulation of a 128 kb Embedded Nonvolatile Memory Based on a Hybrid RRAM (HfO2)/28 nm FDSOI CMOS Technology," in IEEE Transactions on Nanotechnology, vol. 16, no. 4, pp. 677-686, July 2017.

[10] E. Cartier et al., "Fundamental aspects of HfO_2-based high-k metal gate stack reliability and implications on t_{inv}-scaling," 2011 International Electron Devices Meeting, Washington, DC, 2011, pp. 18.4.1-18.4.4

[11] A. Kerber, A. Vayshenker, D. Lipp, T. Nigam and E. Cartier, "Impact of charge trapping on the voltage acceleration of TDDB in metal gate/high-k n-channel MOSFETs," 2010 IEEE International Reliability Physics Symposium, Anaheim, CA, 2010, pp. 369-372.

[12] I. Hirano et al., "Time-dependent dielectric breakdown (TDDB) distribution in n-MOSFET with HfSiON gate dielectrics under DC and AC stressing", Microelectronics Reliability, 53, pp. 1868-1874, 2013.

[13] J. Le, "A finite weakest-link model of lifetime distribution of high-k gate dielectrics under unipolar AC voltage stress," Microelectronics Reliability, 52, pp. 100-106, 2012.

[14] J. W. McPherson, "Time dependent dielectric breakdown physics – models revisited," Microelectronics Reliability, vol. 52, pp. 1753-1760, 2012.

[15] A. Padovani, D. Z. Gao, A. L. Shluger and L. Larcher, "A microscopic mechanism of dielectric breakdown in SiO2 films: An insight from multi-scale modeling," Journal of Applied physics, 121, 155101, 2017.

[16] A. Berman, "Time-Zero Dielectric Reliability Test by a Ramp Method," 19th International Reliability Physics Symposium, Las Vegas, NV, USA, 1981, pp. 204-209.

[17] J. W. McPherson and H. C. Mogul, "Underlying physics of the thermochemical E model in describing low-field time-dependent dielectric breakdown in SiO2 thin films," Journal of Applied Physics, vol. 84, no. 3, pp. 1513-1523, 1998.

[18] I. C. Chen, S. Holland and C. Hut, "A quantitative physical model for time-dependent breakdown in SiO_2," 23rd International Reliability Physics Symposium, Orlando, FL, USA, 1985, pp. 24-31.

[19] J. Lee, I. Chen and C. Hu, "Statistical modeling of silicon dioxide reliability," 26th Annual Proceedings Reliability Physics Symposium 1988, Monterey, CA, USA, 1988, pp. 131-138.

[20] R. Moazzami, J. C. Lee and C. Hu, "Temperature acceleration of time-dependent dielectric breakdown," in IEEE Transactions on Electron Devices, vol. 36, no. 11, pp. 2462-2465, Nov. 1989.

[21] K. F. Schuegraf and C. Hu, "Hole injection oxide breakdown model for very low voltage lifetime extrapolation," 31st Annual Proceedings Reliability Physics 1993, Atlanta, GA, USA, 1993, pp. 7-12.

[22] A. W. Strong et al, Reliability wearout mechanisms in advanced CMOS technologies, IEEE Press Series on Microelectronic Systems, 2009.

[23] D. J. DiMaria and J. W. Stasiak, "Trap creation in silicon dioxide produced by hot electrons," Journal of Applied Physics, Vol. 65(6), 15 March 1989, pp. 2342-2356.

[24] P. E. Nicollian, W. R. Hunter and J. C. Hu, "Experimental evidence for voltage driven breakdown models in ultrathin gate oxides," 2000 IEEE International Reliability Physics Symposium Proceedings. 38th Annual (Cat. No.00CH37059), San Jose, CA, USA, 2000, pp. 7-15.

[25] F. Chen et al., "A Comprehensive Study of Low-k SiCOH TDDB Phenomena and Its Reliability Lifetime Model Development," 2006 IEEE International Reliability Physics Symposium Proceedings, San Jose, CA, 2006, pp. 46-53.

[26] K. H. Allers, "Prediction of dielectric reliability from I-V characteristics: Poole-Frenkel conduction mechanism leading to √E model for silicon nitride MIM capacitor," Microelectronics Reliability, vol. 44, p. 411, 2004.

[27] J. R. Lloyd, "On the physical interpretation of the impact damage model in TDDB of low-k dielectrics," 2010 IEEE International Reliability Physics Symposium, Anaheim, CA, 2010, pp. 943-946.

[28] S. C. Chen et al., "A reliable TDDB lifetime Projection model for advanced gate stack," 2013 IEEE International Integrated Reliability Workshop Final Report, South Lake Tahoe, CA, 2013, pp. 102-105.

[29] A. Aal, "TDDB Data Generation for Fast Lifetime Projections Based on V-Ramp Stress Data," in IEEE Transactions on Device and Materials Reliability, vol. 7, no. 2, pp. 278-284, June 2007.

[30] T. Nigam, A. Kerber and P. Peumans, "Accurate model for time-dependent dielectric breakdown of high-k metal gate stacks," 2009 IEEE International Reliability Physics Symposium, Montreal, QC, 2009, pp. 523-530.

[31] B. P. Linder, E. Cartier, S. Krishnan, J. H. Stathis and A. Kerber, "The effect of interface thickness of high-k/metal gate stacks on NFET dielectric reliability," 2009 IEEE International Reliability Physics Symposium, Montreal, QC, 2009, pp. 510-513.

[32] T. Nigam and A. Kerber, "Reliability modeling of HK MG technologies," Proceedings of the IEEE 2014 Custom Integrated Circuits Conference, San Jose, CA, 2014, pp. 1-8.

[33] G. Bersuker, N. Chowdhury, C. Young, D. Heh, D. Misra and R. Choi, "Progressive Breakdown Characteristics of High-K/Metal Gate Stacks," 2007 IEEE International Reliability Physics Symposium Proceedings. 45th Annual, Phoenix, AZ, 2007, pp. 49-54.

[34] S. Pae et al., "Reliability characterization of 32nm high-K and Metal-Gate logic transistor technology," 2010 IEEE International Reliability Physics Symposium, Anaheim, CA, 2010, pp. 287-292.

[35] A. Kerber and E. A. Cartier, "Reliability Challenges for CMOS Technology Qualifications With Hafnium Oxide/Titanium Nitride Gate Stacks," in IEEE Transactions on Device and Materials Reliability, vol. 9, no. 2, pp. 147-162, June 2009.

[36] F. F. Aghdam and H. Liao, "Reliability study on high-k bi-layer dielectrics," 2017 Annual Reliability and Maintainability Symposium (RAMS), Orlando, FL, 2017, pp. 1-7.

[37] K. Okada, H. Ota, T. Nabatame and A. Toriumi, "Dielectric Breakdown in High-K Gate Dielectrics - Mechanism and Lifetime Assessment," 2007 IEEE International Reliability Physics Symposium Proceedings. 45th Annual, Phoenix, AZ, 2007, pp. 36-43.

[38] X. Wu et al., "Role of oxygen vacancies in HfO2-based gate stack breakdown," Applied Physics Letters, Vol. 96(17), 26 Apr. 2010.

[39] J. H. Stathis, "Percolation models for gate oxide breakdown," Journal of Applied Physics, Vol. 86(10), 15 Nov. 1999, pp. 5757-5766.

[40] A. Kerber, L. Pantisano, A. Veloso, G. Groeseneken, and M. Kerber, "Reliability screening of high k dielectrics based on voltage ramp stress," Microelectronics Reliability, Vol. 47(4), 2007, pp. 513-517.

[41] N. A. Chowdhury and D. Misra, "Charge trapping at deep states in Hf-silicate based high-k gate dielectrics," Journal of The Electrochemical Society, 154 (2), pp. G30-G37, 2007.

[42] G. He and Z. Sun, High-k Gate Dielectrics for CMOS Technology, Wiley-Vch, 2012.

[43] J. Dabrowski and E. R. Weber, Predictive simulation of semiconductor processing status and challenges, Springer, 2004.

[44] S. Oh and Y. T. Yeow, "A modification to the Fowler-Nordheim tunneling current calculation for thin MOS structures", Solid-State Electronics, Vol. 31, Issue. 6, June 1988, pp. 1113-1118.

[45] S. N. Mohammad, G. Fiorenza, A. Acovic, J. B. Johnson and R. L. Carter, "Fowler-nordheim tunneling of carriers in mos transistors: two-dimensional simulation of gate current employing FIELDAY," Solid-state Electronics, Vol. 38, No. 4. pp. 807-814, 1995.

[46] T. Wang, H. C. Ma, C. H. Li, Y. H. Lin, C. H. Chien and T. F. Lei, "Charge Retention Loss in a HfO_2 Dot Flash Memory via Thermally Assisted Tunneling," in *IEEE Electron Device Letters*, vol. 29, no. 1, pp. 109-110, Jan. 2008.

[47] Hirotaka Hamamura *et al*., "Electron trapping characteristics and scalability of HfO_2 as a trapping layer in SONOS-type flash memories," *2008 IEEE International Reliability Physics Symposium*, Phoenix, AZ, 2008, pp. 412-416.