How to Control the State Transformation from Short-Term Potentiation to Long-Term Potentiation of Charge Trapping Synapse?

Md. Hasan Raza Ansari and Nazek El-Atab
SAMA Labs, Division of Computer, Electrical and Mathematical Science and Engineering (CEMSE) Division,
Electrical and Computer Engineering, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia
(Email: nazek.elatab@kaust.edu.sa)

***Abstract*—In this work, the charge trapping-based Gate All Around Transistor (GAA) technology is utilized for the artificial synapse due to its superior performance in terms of low power consumption and high speed. The controlling of state transformation of electronic synapses from short-term potentiation (STP) to long-term potentiation (LTP) is crucial. We have investigated the effect of metal gate workfunction and channel doping on state transformation by using a feedback mechanism (impact ionization). This mechanism regulates channel potentiation and the level of charge trapping in the nitride layer. Our simulation results indicate that this approach can effectively control the STP to LTP transition, while also achieving long STP retention and better dynamic range. This study presents a promising direction for developing advanced electronic devices based on charge-trapping-based GAA transistors.**

Keywords*— *NSFET, Charge Trapping Memory, Synapse features, Short-Term Memory, Long-Term Memory, STP, and LTP

I. Introduction

Neuromorphic computing is a promising candidate for overcoming the limitation of Von-Neumann computing by simultaneously solving the logic applications and storage of the information [1]–[4]. In neuromorphic computing systems, artificial neurons and synapses are essential parts, which mimic the biological nervous system [1]–[4]. A synapse is a junction of two neurons that transfers the information from pre-neuron to post-neurons and vice versa. Therefore, it is important that the artificial device should mimic the synapse features, such as short-term potentiation, long-term potentiation, and long-term depression. Synaptic plasticity is an intrinsic property of the human brain, and plasticity can be controlled through external stimuli [5], [6]. Short-term plasticity is analogous to short-term memory (STM), which means the capacity of the brain to remember information only for a short period. In contrast, Long-term memory is intended for storing data over a long period. STM can be transformed into LTM through consolidation, involving rehearsal and meaningful association [7]. Therefore, it is important to control the transformation and enhance the retention time of short-term memory. In this work, a mature technology silicon-oxide-nitride-oxide-silicon (SONOS) based Gate All Around Transistor memory is utilized to show the synapse behaviour (STP and LTP). Also, the results highlight how we can control the transformation from STP to LTP by optimizing the device metal gate workfunction and channel doping and enhancing STM retention time.

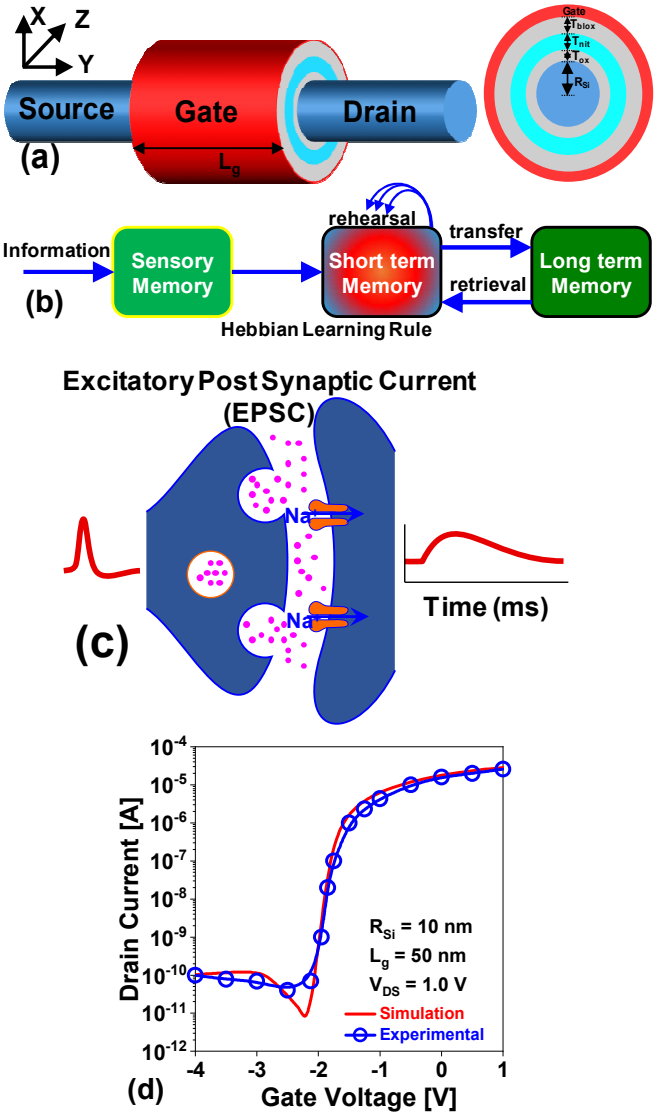

Fig. 1. Schematic representation of (a) Gate All Around transistor based SONOS memory and (b) Hebbien's learnig rules. (c) Schematic illustration of biological synapse mimics excitatory postsynaptic current (EPSC). Calibration of simulation results with experimental result.

II. Results and Discussion

The simulated device through the Silvaco ATLAS simulator is shown in Fig. 1 (a) [8]. The device is simulated with a gate length (L_g) of 200 nm and a channel thickness (R_{Si}) of 20 nm. The device consists of Oxide-nitride-Oxide (ONO) between silicon and gate electrode. The blocking oxide layer (SiO_2) of the ONO is T_{blox} = 6 nm thick, followed by a nitride layer (Si_3N_4) of T_{nit} = 4 nm thickness and a tunnelling oxide layer (SiO_2) of T_{ox} 2 nm

thickness. The metal gate workfunction varies from 4.7 eV to 5.0 eV and channel doping (N_A) from 10^{16} cm^{-3} to 10^{18} cm^{-3}. Fig. 1(b) shows the mechanism of short-term to long-term memory transition [7].

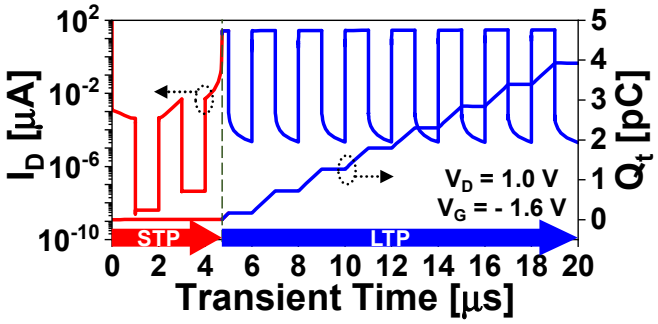

Fig. 2. Transient analysis and trapped charge in the nitride layer of the device during potentiation operation.

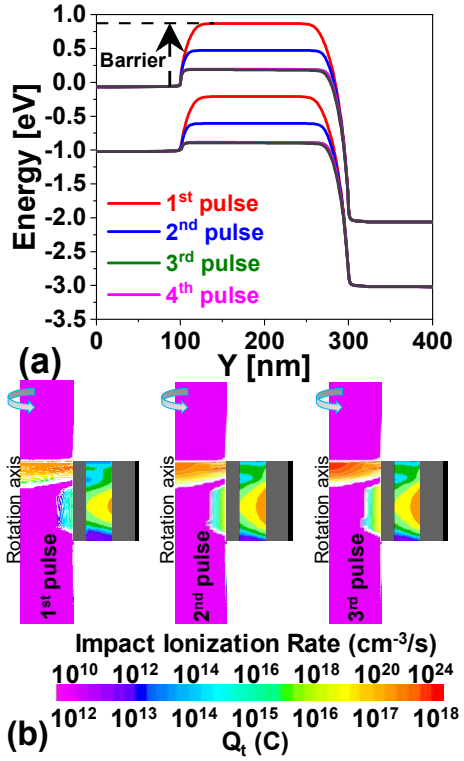

Fig. 3. Physical resaons of state transformation. (a) Variation of energy band diagram and (b) controur plots of the device during different pulses (1st, 2nd, and 3rd pulse) of the potentiation.

In this work, we are showing only excitatory synapse current (EPSC) as shown through the biological synapse (Fig. 1(c)). The artificial electronics synapse operations are based on charge trapping and de-trapping mechanism. Therefore, we have calibrated the simulated models as shown in Fig. 1(d). Fig. 1(d) shows the comparison of simulated transfer characteristics with experimental transfer characteristics, and it conveys that simulated models are well-calibrated with experimental data. The simulation models for this work are Fowler-Nordheim, hot carrier injection models. These models are used to simulate the behaviour of carriers in the silicon layer of the ONO structure. The non-local band-to-band tunnelling (BTBT) model describes the tunnelling of electrons across a potential barrier in the silicon layer, while the impact ionization (II) model describes the generation of electron-

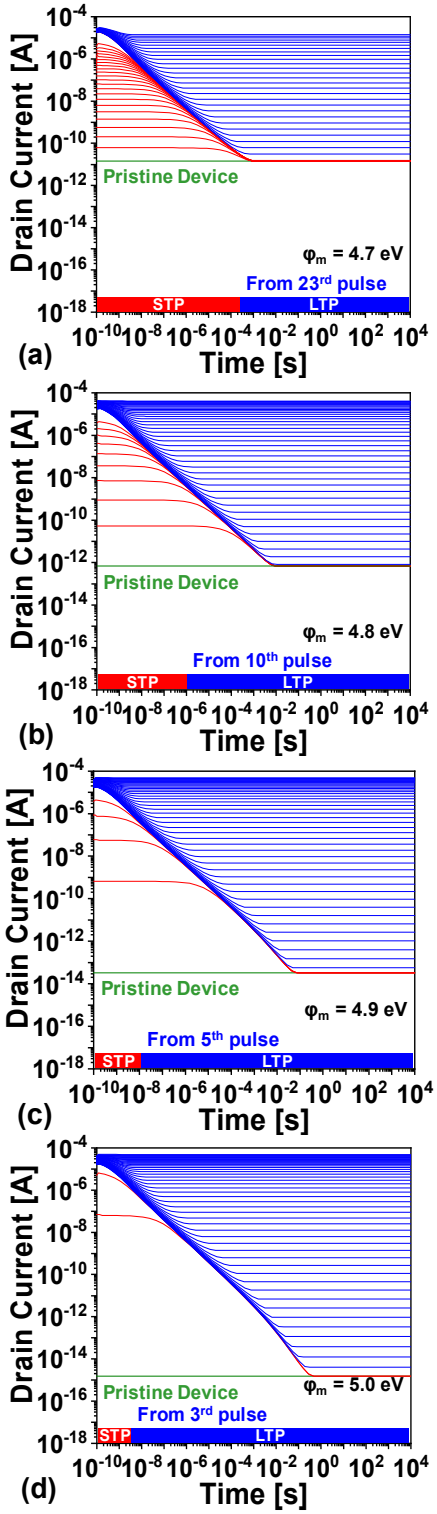

Fig. 4. Transient analysis of the different potentiaion pulses during inference operation for different gate workfunctions of (a) 4.7 eV, (b) 4.8 eV, (c) 4.9 eV, and (d) 5.0 eV. Inference operation is performed at V_D = 0.1 V.

hole pairs due to high electric fields. The SRH model describes the recombination of electrons and holes due to defects in the silicon lattice, while the Auger generation recombination model describes the recombination of carriers due to interactions with other carriers. By incorporating these models, the behaviour of carriers in the ONO structure can be accurately simulated and analyzed.

Fig. 2 shows the transient analysis and variation of trapped charges (Q_t) in the nitride layer of the device during the potentiation operation. We applied repetitive 50 pulses for the potentiation operation by applying a positive drain voltage (V_D = 1.0 V) and negative gate voltage (V_G = -1.6 V) with a pulse and interval time of 1 µs. As shown in Fig. 3(a) and 3(b), the first pulse of potentiation generates electron-hole (*e-h*) pairs due to band-to-band tunnelling. Subsequent pulses increase the generation of *e-h* pairs through the impact ionization mechanism, as shown in the energy band diagram and contour plot of the device. The holes generated through impact ionization get enough energy to tunnel from the tunnelling oxide and get trapped in the nitride layer, leading to a transformation from STP to LTP [9]. This is also reflected in Fig. 2 through the sharp transition in the drain current. Thus, it confirms that the device archives the LTP state at the 3rd pulse of the potentiation.

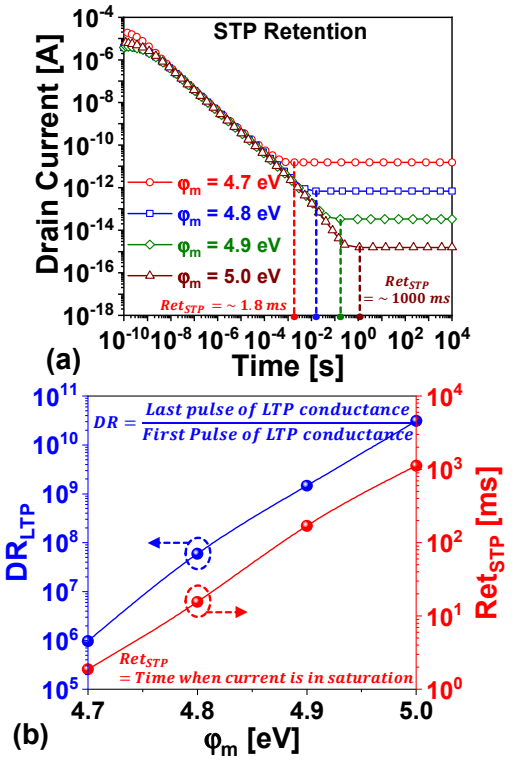

Fig. 5. Transient analysis (estimation of retention) of the 1st pulse of the potentiaion pulses during inference operation for different gate workfunctions. (b) Variation of *DR* and *Ret_STP* states with workfunciton.

Figs. 4(a)-(d) show the effect of the metal gate workfunction (φ_m) (4.7 to 5.0 eV) on the transition pulse, dynamic range (*DR*), and retention time (*Ret*) of each pulse of the potentiation during inference operation of the device. The dynamic range can be estimated through the equation (1) and theses parameters are essential for the neural network simulation [10].

$$DR = \frac{G_{LTP_{Last}}}{G_{LTP_{First}}} \qquad (1)$$

where $G_{LTPFirst}$ and $G_{LTPLast}$ are the first and last pulse of the conductance values.

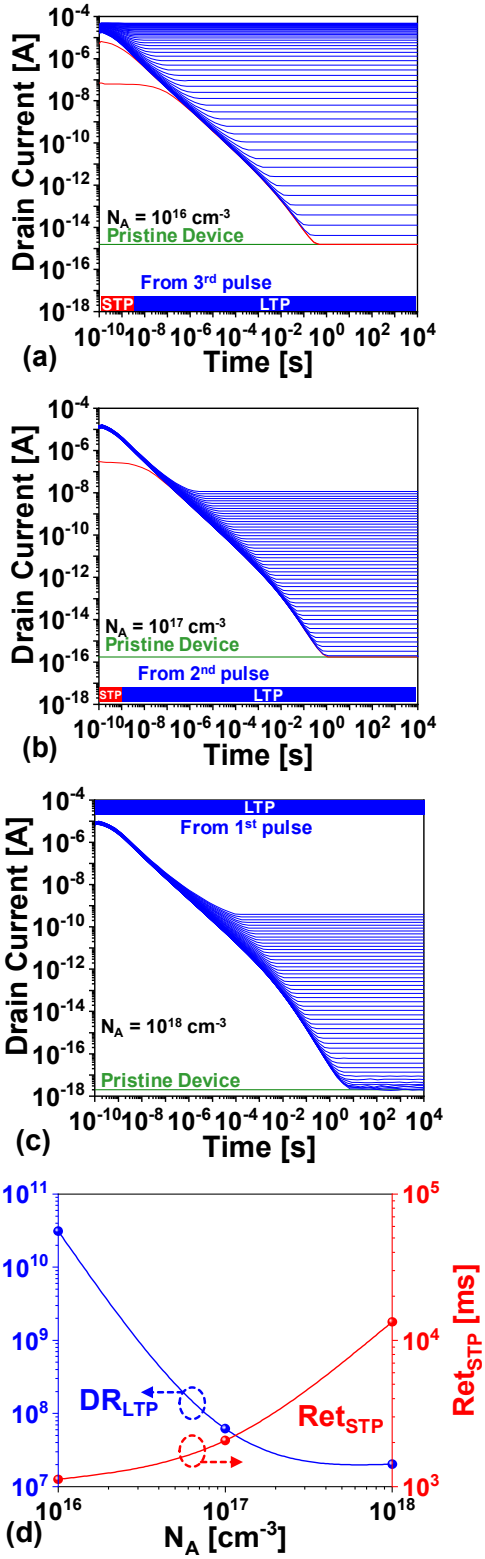

Fig. 6. Transient analysis of the different potentiaion pulses during inference operation for different channel doping (N_A) of (a) 10^{16} cm^{-3}, (b) 10^{17} cm^{-3}, and (c) 10^{18} cm^{-3}. (d) Variation of *DR* and *Ret_STP* states with N_A.

The inference operation is performed at a lower drain voltage (V_D) = 0.1 V. Pristine device defines the device at equilibrium condition (no charge in the nitride layer and floating body). The increase in φ_m accumulates holes in the silicon channel at the tunneling oxide and silicon interface, and thus increases the electric filed in the channel of the device at zero bias conditions. An increase in the electric

field increases the generations of electron-hole pairs due to band-to-band tunnelling and impact ionization in the device.

Therefore, the device archives LTP at a lower pulse. At the same, an increase in φ_m increases the barrier (potential depth) for charge storage in the floating body, which helps to enhance the retention time of the first pulse of short-term potentiation and dynamic range of LTP as shown in Fig. 5(a) and Fig. 5(b), respectively [11]. Thus, the analysis confirms that a higher metal gate φ_m is suitable for the maximum STP retention (*Ret*), requiring less pulse to achieve LTP and better dynamic *DR* of the synapse. Further, the effect of channel doping (N_A) with fixed φ_m =5.0 eV is also analysed to estimate the retention time, dynamic range, and transition pulse as shown in Figs. 6(a)-6(d). Increase in N_A increases the electric field, which shows that the device can achieve LTP at a lower pulse for higher doping as shown in Fig. 6(c). The device is in an LTP state from the first pulse of potentiation. The device also shows that the device achieves higher retention time for STP state due to deeper potential well for higher workfunction and channel doping.

III. Conclusion

This work demonstrated the short-term and long-term memory operations through Gate all-around-based SONOS memory. The generation of holes through band-to-band tunneling and impact ionization controls the transformation state from STP to LTP. The device shows that the workfunction and channel doping control the electric field of the device and thus, the transition state. The device LTP at lower pulse at high gate workfunction and channel doping.

Acknowledgment

This work was supported by the King Abdullah University of Science and Technology baseline fund.

References

[1] D. Ielmini and S. Ambrogio, "Emerging neuromorphic devices," *Nanotechnology*, vol. 31, no. 9, p. 092001, Feb. 2020, doi: 10.1088/1361-6528/ab554b.

[2] A. Sebastian, M. Le Gallo, R. Khaddam-Aljameh, and E. Eleftheriou, "Memory devices and applications for in-memory computing," *Nat. Nanotechnol.*, vol. 15, no. 7, pp. 529–544, Jul. 2020, doi: 10.1038/s41565-020-0655-z.

[3] J. Lee, B. G. Park, and Y. Kim, "Implementation of Boolean Logic Functions in Charge Trap Flash for In-Memory Computing," *IEEE Electron Device Lett.*, vol. 40, no. 9, pp. 1358–1361, 2019, doi: 10.1109/LED.2019.2928335.

[4] C. Mead, "Neuromorphic electronic systems," *Proc. IEEE*, vol. 78, no. 10, pp. 1629–1636, 1990, doi: 10.1109/5.58356.

[5] T. Chang, S. H. Jo, and W. Lu, "Short-term memory to long-term memory transition in a nanoscale memristor," *ACS Nano*, vol. 5, no. 9, pp. 7669–7676, 2011, doi: 10.1021/nn202983n.

[6] S. Kim *et al.*, "Artificial stimulus-response system capable of conscious response," *Sci. Adv.*, vol. 7, no. 15, pp. 1–9, 2021, doi: 10.1126/SCIADV.ABE3996.

[7] R. C. Atkinson and R. M. Shiffrin, "Human Memory: A Proposed System and its Control Processes," in *Journal of Physics A: Mathematical and Theoretical*, 1968, pp. 89–195. doi: 10.1016/S0079-7421(08)60422-3.

[8] Atlas User's Manual., *St. Clara, CA, USA Silvaco*.

[9] M. H. R. Ansari, U. M. Kannan, and S. Cho, "Core-Shell Dual-Gate Nanowire Charge-Trap Memory for Synaptic Operations for Neuromorphic Applications," *Nanomaterials*, vol. 11, no. 7, p. 1773, Jul. 2021, doi: 10.3390/nano11071773.

[10] 6 S. Yu, "Neuro-Inspired Computing With Emerging Nonvolatile Memorys," *Proc. IEEE*, vol. 106, no. 2, pp. 260–285, Feb. 2018, doi: 10.1109/JPROC.2018.2790840.

[11] M. H. R. Ansari, N. Navlakha, J.-T. Lin, and A. Kranti, "Doping Dependent Assessment of Accumulation Mode and Junctionless FET for 1T DRAM," *IEEE Trans. Electron Devices*, vol. 65, no. 3, pp. 1205–1210, Mar. 2018, doi: 10.1109/TED.2018.2789901.