NOVEL NEGATIVE-FEEDBACK METHOD FOR WRITING VARIATION SUPPRESSION IN FEFET-BASED COMPUTING-IN-MEMORY MACRO

Weikai Xu[1], Jin Luo[1], Yide Du[1], Qianqian Huang[1,2], and Ru Huang[1,2*]*
[1] Key Laboratory of Microelectronic Devices and Circuits (MOE), School of Integrated Circuits,
Peking University, Beijing 100871, China
[2] Chinese Institute for Brain Research (CIBR), Beijing 102206, China
*Corresponding Author's Email: hqq@pku.edu.cn; ruhuang@pku.edu.cn

ABSTRACT

In this work, a novel one-shot negative-feedback writing method is proposed for suppression of writing variation in ferroelectric FET (FeFET) based analogy computing-in-memory (CIM) macro for high-accuracy artificial neural network (ANN). By utilizing the source-voltage negative-feedback mechanism and the voltage and time dependent multi-domain switching dynamics, a source-follower writing structure with an FeFET and a NMOS is proposed and simulated to reduce variation of target programming V_{TH} for FeFET with even multi-level state, revealing significant decrease of FeFET synaptic conductance variation. Furthermore, based on the proposed FeFET writing variation suppression method, the CIM macro of FeFET array is demonstrated to improved accuracy of image recognition ANN by 40% compared with direct writing, showing great potential for high-accuracy neural network system.

INTRODUCTION

Emerging non-volatile memory devices (NVM) [1][2] based computing-in-memory (CIM) architectures which perform matrix vector multiplication (MVM) operations of artificial neural network (ANN) have triggered a lot of interests. Ferroelectric FET (FeFET) is considered as a NVM candidate for synaptic weight cell of CIM macro due to low write power consumption and high on/off ratio [3][4]. However, the writing variation of FeFET caused by non-uniformity of domain distribution or stochasticity of domain switching [5], resulting in memory window collapse [6] and network accuracy decline [7], which is one of the main bottlenecks in the realization of high-accuracy large-scale ANN. Recently, a current-limiting circuit with an FeFET and a resistor was proposed to restrain I_{on} variation for 1-bit storage [8], while sacrificing the dynamic range. The write-and-verify scheme was proposed to restrain FeFET V_{TH} variation [9], while suffering from complex sequence control circuit and high program energy consumption.

In this work, a one-shot negative-feedback writing method based on source-follower structure and voltage and time dependent multi-domain switching dynamics, is proposed to suppress writing variation of FeFET based multi-level weight cell. Without influencing the FeFET dynamic range, only one-shot write operation is required to form the target state. Furthermore, based on the proposed writing variation suppression method, FeFET based CIM macro for image recognition with ANN is demonstrated to achieve accuracy improvement by 40%.

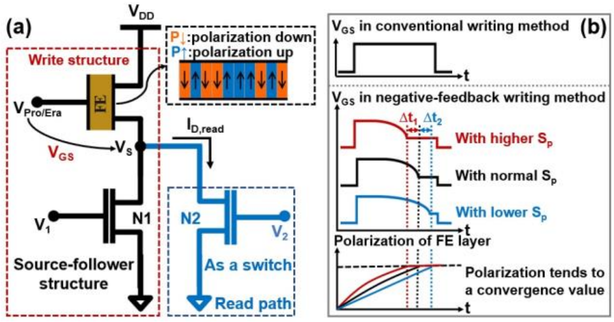

Figure 1: (a) The novel one-shot negative-feedback writing circuit; (b) The principle of suppressing variation

THE NEGATIVE-FEEDBACK WRITING METHOD FOR WRITING VARIATION SUPPRESSION OF FEFET-BASED CELL

Principle of the negative-feedback writing method

As shown in Fig.1(a), the one-shot negative-feedback writing circuit based on FeFET is proposed, which is composed of one FeFET and two NMOSs. During the writing process, N2 is turned off so that the source of FeFET is connected to the drain of N1 with a fixed gate voltage, which forms a source-follower structure. During the reading process, N2 is turned on with sufficiently high conductivity to short the N1 and the FeFET channel conductance is obtained. Taking the program process as an example to analyze the principle of suppressing writing variation, as shown in Fig.1(b). The amount of polarization switching per unit time during program pulse applied is decided by the polarization program speed (S_p). Sp is different due to the variation of coercive field (E_c) distribution and remanent polarization (P_r) of multi-domain ferroelectric (FE) layer, leading to different polarization switching amount per unit time. For conventional open loop direct writing operation where the source of FeFET is connected to GND and V_{GS} remains constant during pulse programming process, the FeFET programming V_{TH} state shows a severe deviation due to the different S_p of FE layer. For the proposed negative-feedback method of source-follower structure, the source voltage of FeFET will change along the dynamic switching process of FE polarization, which will play a negative-feedback role to adaptively adjust V_{GS} of FeFET. Higher S_p of FeFET results faster V_{TH} decrease and channel conductance increase during positive pulse, and V_{GS} decreases faster to inhibit excess polarization switching as shown in the red line in Fig.1(b). On the contrary, for FeFET with lower S_p, V_{GS} decreases slower to allow longer programming duration as shown in the blue line in Fig.1(b). Therefore, the final polarization of FE layer tends to a relatively convergence value due to the dynamic negative-feedback process of V_{GS}, which suppresses the variation of programmed FeFET V_{TH}. For the erase process, the principle is the same as above.

Modeling of FeFET variation

As shown in Fig.2(a), the FeFET model composed of the multi-domain dynamic Preisach model for FE layer and BSIM model for MOSFET [10][11] is established to analyze the function of the proposed negative-feedback writing method by adding additional Gaussian distribution to reflect the variation of dispersion degree of E_c and P_r for FeFETs. As the FE layer is integrated within the MOSFET gate stack in FeFET, the Eq.1 governing the law of charge conservation and voltage division are solved on iterative method in HSPICE.

$$\begin{cases} Q_{Fe}(V_{Fe}) = Q_{MOS}(V_{MOS}) \\ V_G = V_{Fe} + V_{MOS} \end{cases} \quad (1)$$

The E_c and Pr of the FE model is calibrated with experimental results in [12]. Fig.2(b) shows the *P-V* saturation loop and corresponding *I-V* curve of metal-ferroelectric-metal (MFM) capacitor. There are various variation sources which mainly change the E_c distribution and P_r in the *P-V* loop of MFM capacitor. The variation of dispersion degree of multi-domain Ec distribution and FE polarization density is analyzed respectively, and the simulation results are shown in Fig.2(c)(d).

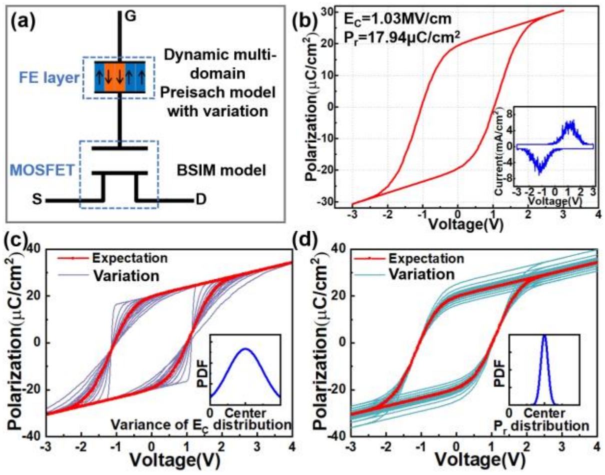

Figure 2: (a) The FeFET model; (b) P-V saturation loop of MFM capacitor; (c)(d) P-V loops with variation

Variation suppression with negative-feedback method

When applying the program/erase voltage pulses with the conventional open-loop direct writing operation and the proposed negative-feedback writing operation respectively, I_D-V_{GS} read-out curves of FeFET with high/low V_{TH} states are obtained with backward/forward gate-voltage scanning, as shown in Fig.3. For the variation of Ec distribution and Pr, the negative-feedback writing method reduces the standard deviations (σ) of FeFET V_{TH} distribution by more than 50% from Fig.3(a)(b) to Fig.3(c)(d), without significantly reducing of the memory window width compared with the conventional method, indicating that the negative-feedback writing method can effectively suppress the writing variation of FeFET.

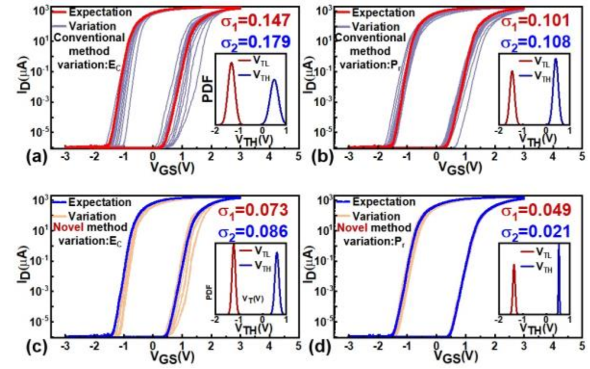

Figure 3: I_D-V_{GS} curves with the two types of variation mentioned above (a)(b) In conventional writing method; (c)(d) In one-shot negative-feedback writing method

Furthermore, because the multilevel characteristics of FeFET can improve the performance of FeFET-based CIM macro [3], we further study the influence of negative-feedback writing method on the multilevel states of FeFET with variation. We use the characteristic that FE polarization follows the minor-loop under unsaturated applied voltage to obtain four different V_{TH} states by

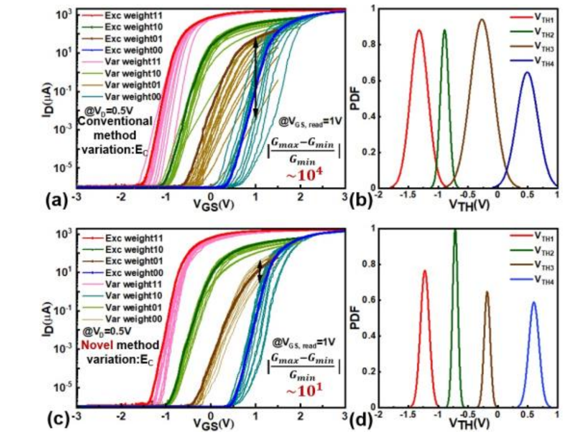

Figure 4: Simulation results of multilevel I_D-V_{GS} curves and threshold voltages distribution of FeFET with variation (a)(b) In conventional writing method; (c)(d) In one-shot negative-feedback writing method

applying four different program voltages, representing and storing four different weight values respectively. The I_D-V_{GS} curves with the variation of dispersion degree of E_c distribution of FE layer are shown in Fig.4. The four threshold voltage states in conventional method have obvious overlapped and are difficult to distinguish, while there is still a significant sense margin in negative-feedback method. For different FeFETs with device-to-device variation or different write operations in one FeFET with cycle-to-cycle variation, the maximum shift of conductance state which is defined as $\left|\frac{G_{max}-G_{min}}{G_{min}}\right|$ at V_{GS}=1V in this work, is reduced by about 1000 times from nearly 10^4 in conventional method to 10 in negative-feedback method, The variation of FeFET based multi-weight on MVM operation is significantly reduced, which is of great significance to improve the accuracy of analog synaptic ANN based on FeFET CIM macro.

FEFET-BASED CIM MACRO FOR IMAGE RECOGNITION

Based on the proposed negative-feedback writing method and multiplexing principle, a negative-feedback FeFET pseudo-crossbar array architecture is proposed as shown in Fig.5(a).The memory cell consists of a NMOS for selection and an FeFET for storage, which can be accurately programmed/erased separately. Therefore, the whole array only needs to share a feedback transistor through a multiplexer for the negative-feedback write operation, which has negligible area and energy efficiency loss compared with the conventional method.

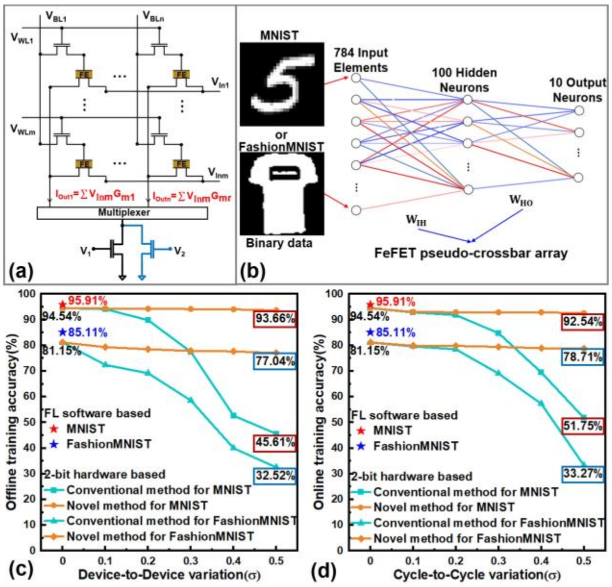

Figure 5: (a) FeFET pseud-crossbar array based on negative-feedback writing method; (b) Two-layer MLP neural network; (c) Offline training accuracy with device-to-device variation; (d) Online training accuracy with cycle-to-cycle variation

A two-layer multilayer perceptron (MLP) ANN with quantifiable weight accuracy is established, which can further add device-to-device during inference and cycle-to-cycle variation during training respectively, as shown in Fig.5(b). The classification accuracy of 2-bit FeFET weight-based hardware without variation reaches 94.54% and 95.91% respectively, close to 95.91% and 85.11% of full precision weight based on software for MINST and FashionMNIST databases. However, the classification accuracy in conventional method will decline severely when adding the variation of FeFET. Based on the proposed novel negative-feedback writing method, the image classification accuracy of ANN with variation is demonstrated with significantly improved compared to the conventional method in a large variation range, as shown in Fig.5(c)(d). Furthermore, the improvement of accuracy is more significant with variation increasing. The offline training accuracy improvement can reach 48.05% and 44.52% based on MNIST and FashionMNIST databases respectively with the biggest device-to-device variation considered in this model as shown in Fig.5(c), and the online training accuracy improvement can reach 40.79% and 45.44% with the biggest cycle-to-cycle variation as shown in Fig.5(d).

CONCLUSION

In this work, a novel one-shot negative-feedback writing method of FeFET is proposed, which can suppress the writing variation effectively utilizing the negative-feedback mechanism and the voltage and time dependent multi-domain switching dynamics. Furthermore, based on the proposed FeFET-based CIM macro design, high-accuracy image recognition ANN with variation is demonstrated, which is of great significance to the implementation of high-accuracy large-scale neural network system.

ACKNOWLEDGEMENTS

This work was supported by National Key R.D Program of China (2018YFB2202800), NSFC (61927901, 61822401, 61851401, 61421005) and 111 Project (B18001).

REFERENCES

[1] C. Xue et al., *ISSCC*, pp. 388-390, 2019.
[2] L. Gallo et al., *IEEE TED*, pp. 99.1-99.9, 2018.
[3] S. Yu, *Proceedings of the IEEE*, pp. 260-285, 2018.
[4] J. Luo et al., IEDM, pp.19.5.1-19.5.4, 2021.
[5] K. Ni et al., *IRPS*, pp. 1-5, 2020.
[6] K. Ni et al., *VLSI*, pp. T40-T41, 2019.
[7] P. Chen et al., *IEDM*, pp. 6.1.1-6.1.4, 2017.
[8] T. Soliman et al., *IEDM*, pp. 29.2.1-29.2.4, 2020.
[9] H. Zhou et al., *IEDM*, pp. 18.6.1-18.6.4, 2020.
[10] J. Luo et al., *IEDM*, pp. 6.4.1-6.4.4, 2019.

[11] Z. Fu et al., *CSTIC*, pp. 1-4, 2020.
[12] V. Gaddam et al., *IEEE TED*, pp. 745-750, 2020.