Partially Depleted SONOS FinFET for Unified RAM (URAM)—Unified Function for High-Speed 1T DRAM and Nonvolatile Memory

Jin-Woo Han, Seong-Wan Ryu, Chung-Jin Kim, Sungho Kim, Maesoon Im, Sung Jin Choi, Jin Soo Kim, Kwang Hee Kim, Gi Sung Lee, Jae Sub Oh, Myeong Ho Song, Yun Chang Park, Jeoung Woo Kim, and Yang-Kyu Choi

***Abstract*—Unified random access memory (URAM) is demonstrated for the first time. The novel partially depleted (PD) SONOS FinFET provides unified function of a high-speed capacitorless 1T DRAM and a nonvolatile memory (NVM). The combination of an oxide/nitride/oxide (O/N/O) layer and a floating-body facilitates URAM operation in PD SONOS FinFETs. An NVM function is achieved by FN tunneling into the O/N/O stack and, a 1T-DRAM function is achieved by excessive-hole accumulation in the PD body. The fabricated PD SONOS FinFET shows retention time exceeding 10 years for NVM operation and program/erase time below 6 ns for 1T-DRAM in a single-cell transistor. These two memory functions are guaranteed without disturbance between them.**

***Index Terms*—Capacitorless DRAM, FinFET, nonvolatile memory (NVM), SONOS, unified RAM (URAM), 1T DRAM.**

I. Introduction

INTEGRATION of various functional circuits in a single chip is desirable for embedded systems. System-on-chip (SoC) and system-in-package are examples of all-in-one integration. For a memory block, nonvolatile memory (NVM) and dynamic random access memory (DRAM) can be integrated on a single chip, but their poor process compatibility obstructs hybrid integration. Their dissimilar structure requires additional process steps, resulting in high cost and low yield. Therefore, if a single cell could perform the functions of both NVM and DRAM, lowered chip cost, increased product yield, and a user-friendly circuit design could be reasonably expected. The SONOS FinFET has been identified as a promising future NVM [1], [2], and capacitorless FinFET DRAM is currently among the most attractive candidates for embedded memory [3]–[5]. In this letter, we propose a novel partially depleted (PD) SONOS FinFET that unifies the functions of high-speed capacitorless 1T DRAM and NVM in a single cell. This memory is fabricated on a silicon-on-insulator (SOI) wafer by using a conventional SONOS memory process with minor modifications and performs multifunctions without sacrificing chip area. By implementation of oxide/nitride/oxide (O/N/O) as an electron-trapped zone for NVM and a PD floating-body as a hole storage zone for 1T DRAM, unified RAM (URAM) operation is realized. In a previous fully depleted (FD) 1T DRAM, a negative biased second gate was used in order to accumulate holes at the back interface [5]. However, the newly proposed structure allows excessive-hole accumulation in an extremely narrow fin without a second gate. Therefore, a gate separation step for the second gate is not needed, and additional bias to hold excessive holes is not necessary. The body of the proposed FinFET is composed of an FD zone for device scalability and a PD zone for hole accumulation.

In this letter, process details of the PD SONOS FinFET and multifunctional characteristics of NVM and 1T DRAM are presented. We also fabricate a conventional FD SONOS FinFET for comparison with the PD SONOS FinFET in terms of 1T-DRAM operation. Supportive simulation data are provided to explain the difference between the two structures.

II. Device Design and Fabrication

A p-type (100) SOI wafer is used as a starting material. After photoresist trimming to provide a narrow fin width, the fin is patterned with a nitride hard mask. High-density-plasma (HDP) SiO_2 is deposited and partially recessed until the 60 nm of the upper part of the Si fin is exposed. The remaining 50 nm of the lower part of the Si fin is covered by HDP SiO_2. Next, O/N/O with a thickness of 3/8/3 nm is deposited, and *in situ* n+ poly-Si is deposited sequentially. After gate patterning, S/D implantation and activation are carried out. The device has 30-nm fin width, 110-nm fin height, and 180-nm gate length. The process sequence for fabrication of the FD SONOS FinFET is fabricated by the same process sequence described in [2]. Fig. 1(a)–(c) show the PD SONOS FinFET, and Fig. 1(d) shows the FD SONOS FinFET prepared as a control group.

III. Results and Discussion

Fig. 2 shows the operation principle of the URAM. Data states of NVM are distinguished by the existence of trapped charges in the nitride layer. The program/erase (P/E) operation

Manuscript received April 15, 2008. This work was supported in part by the National Research Program for the 0.1-Terabit Nonvolatile Memory Development, sponsored by the Ministry of Knowledge Economy. The review of this letter was arranged by Editor Y. Taur.

J.-W. Han, S.-W. Ryu, C.-J. Kim, S. Kim, M. Im, S. J. Choi, and Y.-K. Choi are with the Division of Electrical Engineering, School of Electrical Engineering and Computer Science, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Korea (e-mail: ykchoi@ee.kaist.ac.kr).

J. S. Kim, K. H. Kim, G. S. Lee, J. S. Oh, M. H. Song, Y. C. Park, and J. W. Kim are with the National Nanofab Center, Daejeon 205-806, Korea.

Digital Object Identifier 10.1109/LED.2008.2000616

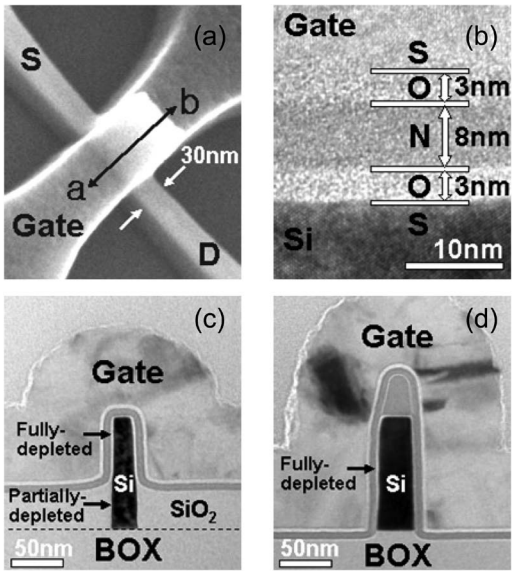

Fig. 1. Images of a PD SONOS FinFET. (a) Bird's eye view of PD SONOS FinFET. (b) TEM image of O/N/O layer. (c) TEM image of cross-sectional view along a–b direction in (a). (d) TEM image of FD SONOS FinFET.

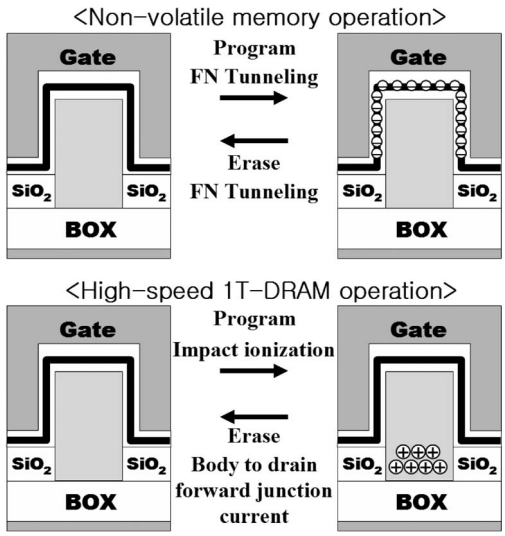

Fig. 2. Conceptual schematic of URAM. Implementation of an O/N/O layer and a floating-body into a single cell allows unified operation. A nitride trap in O/N/O is used as an electron storage node for NVM, and a floating body is utilized as a hole accumulation node for high-speed 1T-DRAM.

is carried out by Fowler–Nordheim tunneling or channel hot-electron injection, and the P/E voltage range is relatively high. Data states of 1T DRAM are stored in the floating-body, which is PD. In programming, holes created by impact ionization are accumulated in the floating-body, and the channel potential is lowered. In erasing, accumulated holes are eliminated by the negative drain bias, and the channel potential is raised. Therefore, the data state is identified by the presence of excessive holes. The key concept of the proposed URAM is that the hole accumulation region is spatially separated from the inverted channel. The upper part of the fin remains FD, and the lower part, which is uncovered by the gate, remains PD in order to hold excessive holes. Therefore, the floating-body effect is observed even at 30-nm fin width. Simulation data in Fig. 3 represent the body potential before/after impact ionization in

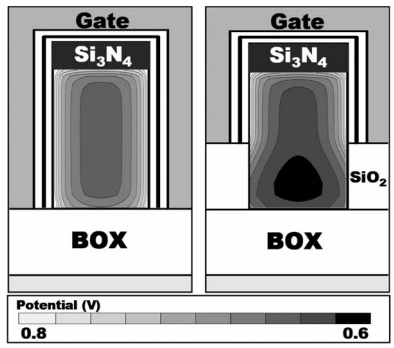

Fig. 3. Contour of body potential from the device simulation of (a) an FD FinFET and (b) a PD FinFET. In the PD FinFET, the lower body potential at the PD zone allows hole accumulation after impact ionization. Therefore, 1T DRAM can be operated due to the floating-body effect. However, in the FD FinFET, the floating-body effect is not observed because the body is FD.

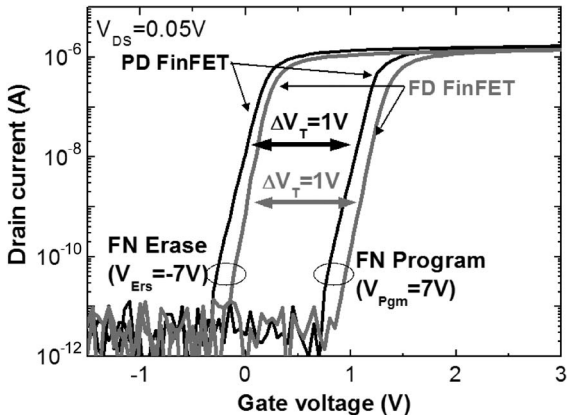

Fig. 4. Measured drain current (I_D)–gate voltage (V_G) characteristics before and after NVM programming. P/E is carried out by a Fowler–Nordheim mechanism. A 1-V threshold-voltage window is achieved at $\pm$7-V P/E voltage with an 80-μs pulse.

the FD and PD FinFETs. The existence of PD zone ensures that the PD FinFET can retain more holes and increase its retention time relative to the FD FinFET.

Fig. 4 shows measured I_D–V_G characteristics before and after programming for the NVM mode. P/E is carried out with a Fowler–Nordheim tunneling mechanism. A 1-V threshold-voltage window is achieved at $\pm$7-V P/E voltage with an 80-μs pulse for both the FD and PD FinFETs. Ten years of retention and 10^7 P/E cycles of endurance are measured with a 0.7-V of threshold-voltage window (data not shown). In NVM characteristics, FD and PD devices exhibit very similar V_T window and reliability. Fig. 5 shows the P/E characteristics for the 1T-DRAM mode. It should be noted that a high P/E voltage can increase the sensing margin but will lead to charge trapping into the nitride layer. Therefore, this parameter should be carefully selected so as to avoid undesired charge trapping. In this letter, program voltages of $V_{G,\mathrm{PGM}} = 1$ V and $V_{D,\mathrm{PGM}} = 1.5$ V are used for hole generation, and erase voltages of $V_{G,\mathrm{ERS}} = 1$ V and $V_{D,\mathrm{ERS}} = -1$ V are set for hole elimination. The read voltages are $V_{G,\mathrm{READ}} = 1$ V and $V_{D,\mathrm{READ}} = 0.6$ V. Before utilizing the 1T-DRAM mode, the initial threshold voltage should be set to 0.2 V. A high initial threshold voltage lowers the gate overdrive voltage (V_G–V_T), which in turn reduces the impact ionization efficiency, whereas a low initial threshold

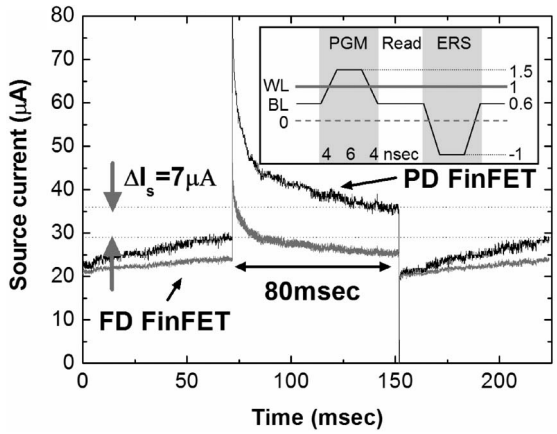

Fig. 5. Measured source current (I_S) for 1T-DRAM operation. Whereas two distinctive data states are clearly distinguished according to hole accumulation in the PD FinFET, the difference between these two states is relatively small in the FD FinFET.

voltage increases the OFF-state current. The optimized initial threshold voltage is found to be 0.2 V. In Fig. 5, the P/E states are clearly distinguished with a 7-μA sensing window in the PD FinFET, whereas the FD FinFET exhibits a smaller sensing window. This is attributed to the presence of an excessive-hole accumulation region, as shown in Fig. 3. It should be noted that the P/E speed is 6 ns. With a faster measurement system, a higher speed response could be expected.

The device simulation to show charge (hole) concentrations modulated by $V_{\rm sub}$ is performed to estimate a difference of the body capacitance for two structures. Because the gate, the source, and the drain structures are identical for both FinFETs, the difference originates from the $C_{\rm sub}$. Because $C_{\rm sub}$ can be expressed as $C_{\rm sub} = dQ/dV_{\rm sub}$, the relative body capacitance can be obtained from the slope of charge versus voltage graph. The body capacitance of PD FinFET is three times steeper than that of FD FinFET. As a result, the body capacitance of PD FinFET must be higher than that of FD FinFET. This estimated body capacitance can provide the enough insight of the difference between two structures.

IV. Conclusion

URAM, i.e., NVM and high-speed 1T-DRAM operation in a single cell, is demonstrated for the first time. The combination of an O/N/O layer as an electron storage node for NVM and a PD body as a hole storage node for 1T DRAM provided URAM operation in a single SONOS FinFET memory cell. The newly developed structure is fabricated using a standard SONOS FinFET with minor modifications, and it performs multifunctions without sacrificing chip area. The URAM is expected to be an attractive feature for fusion memory and future SoC applications.

References

[1] M. K. Cho and D. M. Kim, "High performance SONOS memory cells free of drain turn-on and over-erase: Compatibility issue with current flash technology," *IEEE Electron Device Lett.*, vol. 21, no. 8, pp. 399–401, Aug. 2000.

[2] P. Xuan, M. She, B. Harteneck, A. Liddle, J. Bokor, and T.-J. King, "FinFET SONOS flash memory for embedded applications," in *IEDM Tech. Dig.*, 2003, pp. 609–612.

[3] S. Okhonin, M. Nagoga, J. M. Sallese, and P. Fazan, "A capacitor-less 1T-DRAM cell," *IEEE Electron Device Lett.*, vol. 23, no. 2, pp. 85–87, Feb. 2002.

[4] C. Kuo, T.-S. King, and C. Hu, "A capacitorless double-gate DRAM cell," *IEEE Electron Device Lett.*, vol. 23, no. 6, pp. 345–347, Jun. 2002.

[5] E. Yoshida, T. Miyashita, and T. Tanaka, "A study of highly scalable DG-FinDRAM," *IEEE Electron Device Lett.*, vol. 26, no. 9, pp. 655–657, Sep. 2005.