



Article A High-Density Memory Design Based on Self-Aligned Tunneling Window for Large-Capacity Memory Application

Chen Wang 🔍, Xiuli Zhao, Hao Liu, Xin Chao, Hao Zhu * and Qingqing Sun *

School of Microelectronics, Fudan University, Shanghai 200433, China; chenwang19@fudan.edu.cn (C.W.); xiulizhao19@fudan.edu.cn (X.Z.); liuhao16@fudan.edu.cn (H.L.); xchao19@fudan.edu.cn (X.C.) * Correspondence: hao_zhu@fudan.edu.cn (H.Z.); qqsun@fudan.edu.cn (Q.S.)

Abstract: Despite the continuous downscaling of complementary metal–oxide–semiconductor (CMOS) devices, various scenarios of technology have also been proposed toward the shrinking of semiconductor memory. In this paper, a high-density memory (HDM) has been proposed on the basis of band-to-band tunneling (BTBT) for low-power, high density, and high-speed memory applications. The geometric structure and electrical properties have been demonstrated by using TCAD tools. Typical memory operations including read, program, and erase have been designed and performed. High operation speed, lower power consumption, as well as good reliability characteristics have been achieved by simulation, which indicates that the HDM may have potential application value as a novel semiconductor memory device.

Keywords: high-density memory; band-to-band tunneling; lower power; reliability characteristic; high operation speed

1. Introduction

Static random-access memory (SRAM) has been used in speed-sensitive cache due to its fast speed with low power. However, the conventional 6T-SRAM has encountered severe challenges in the implementation of large-capacity memory due to the large cell size (~80 F^2 –120 F^2). A memory technology with comparable speed but higher density has been urged by the development of on-chip memory [1–3]. In the past fifty years, various new devices have become the driving force for the development of the semiconductor industry. Specifically, in the past decade, semiconductor memory has driven an explosive growth in the industry. However, although the complementary metal–oxide–semiconductor (CMOS) logic devices have reached the physical limits by landing at the sub-5 nm technology node, the scaling of CMOS memory is still far behind due to technical difficulties. The use of the storage capacitor and its leakage in conventional memory technology is the major bottleneck impeding further down-scaling, which seriously hinder the further development of semiconductor memory [4–8]. Therefore, a variety of new structures and new materials have been used to realize advanced semiconductor memories to overcome the shortcomings of traditional memory devices.

Band-band tunneling mechanisms have been extensively studied and applied to semiconductor devices with various structures, such as tunneling transistors. Studies have shown that devices based on this mechanism have the characteristics of fast speed and low power consumption, which makes it possible to study a semiconductor memory device based on the band-band tunneling mechanism. Compared with semiconductor memory based on the hot carrier injection mechanism and F-N tunneling mechanism, devices based on band-to-band tunneling have the advantages of low operation voltage and high speed [9–13]. In order to meet the requirements of fast-speed, high-density, and low-power memory, alternative architectures have been considered as promising candidates to replace the current product. In this paper, we propose a design of a high-density memory (HDM)



Citation: Wang, C.; Zhao, X.; Liu, H.; Chao, X.; Zhu, H.; Sun, Q. A High-Density Memory Design Based on Self-Aligned Tunneling Window for Large-Capacity Memory Application. *Electronics* **2021**, *10*, 1954. https://doi.org/10.3390/ electronics10161954

Academic Editor: Elias Stathatos

Received: 25 June 2021 Accepted: 11 August 2021 Published: 13 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). that utilizes tunneling current for program and erase operations based on the band-to-band tunneling effect.

The structure of the HDM was modified on the basis of traditional floating gate memory. Through simulation, we integrated the process of HDM and logic device, and we realized the device structure in the form of simulation. The simulation results of device characteristics and process steps were based on TCAD tools. Through the simulation results, we could clearly see that the band-to-band tunneling of electrons caused the tunneling current, and the tunneling occurred in the gate-drain overlap region when negative voltage was applied on the gate ($V_{CG} < 0$) and positive voltage was applied on the drain ($V_D > 0$). The simulation results indicated that the HDM has superior performance such as lower power consumption, high density (8 F²), higher speed (operation speed less than 5 ns), and enhanced anti-interference capability (~1 s), which indicate that the HDM may have potential application values in high-density SoC cache.

2. Device Structure and Main Process of HDM

This HDM was simulated under the advanced semiconductor device simulation platform, which can be used to accurately simulate the process and electrical characteristics of the device. This HDM device achieves high speed and low power consumption by changing the structure of the traditional floating gate memory. By opening a part of the gate oxide layer, the programming mode of the device changes from the hot carrier injection mode as in traditional floating gate memory to the band-to-band tunneling mode, which allows the device to work under a smaller operating voltage and faster storage speed.

Through the simulation by using TCAD tools, the device structure and the main process steps are shown in Figure 1. All the processes have good compatibility with standard CMOS fabrication, which greatly reduces the manufacturing cost of the device, enabling possible application in the large-scale integration. The main processes are as follows: P-type substrate is used and the floating gate (FG) is p+ doped. The source, control gate (CG), and drain are self-aligned during the ion implantation process, and the doping is n+ type. The oxide thickness between the CG and FG is 4 nm, while the gate oxide thickness between the channel and FG is 8 nm. The CG and FG are formed with polysilicon. The thicknesses of the FG and CG are 250 nm and 175 nm, respectively. The widths of the band-to-band tunneling window and MOSFET channel are 110 nm and 60 nm, respectively. Figure 2 shows the main process simulation results of the HDM.



Figure 1. (a) The device structure with concentration profile and (b) the main process flow of HDM.

3 of 8



Figure 2. The main process simulation results (**a**) Gate oxide oxidation (8 nm). (**b**) Gate oxide etch to fabricate tunneling window. (**c**) FG poly deposit. (**d**) FG and CG barrier oxide oxidation (4 nm). (**e**) Gate etch. (**f**) Fabricate spacer of the HDM.

3. Simulation Results and Discussion

Based on the device structure described above, the electrical characteristics of this device were simulated in addition to the process simulation using the TCAD tools. The electrical simulation process simulates the program, erase, read, retention, and other characteristics of the device. These electrical simulation results showed that the device has excellent storage characteristics, making it a new solution for semiconductor memory in the future. The operation voltage and time of each electrode of the device are shown in Table 1. The band-to-band generation rate was calculated during the simulation process by the Kane BTBT E2 model, which is expressed as follows [14–17]:

$$G_{BTBT} = A_{BTBT} \times \frac{E^2}{E_g^{\frac{1}{2}}} \times exp\left(-B_{BTBT} \times \frac{E_g^{\frac{3}{2}}}{E}\right)$$
(1)

where G_{BTBT} is tunneling efficiency, and A_{BTBT} and B_{BTBT} are tunneling constants. The programming operation of HDM consists of two processes, namely write-0 and write-1. Through these two different write operations, the device has two different storage states of 0 and 1. The difference between write-0 and write-1 processes lies in the difference in V_D voltage in the two processes. Larger V_D voltage suggests higher programming efficiency under the same operating time, so different numbers of carriers are written into the floating gate. The equivalent circuit is shown in Figure 3. During the program operation, the C_{fg} is charged by BTBT tunneling. The band-to-band tunneling rate profile of the write-1 operation and the bandgap during the write-1 operation are shown in Figure 4a,b, respectively.

	Erase	Write0	Write1	Read	Retention
V _{CG} (V)	1.2	-1.5	-1.5	0.6	0.0
V _D (V)	0.6	1.2	0.6	0.6	0.6
$V_{S}(V)$	0.0	0.0	0.0	0.0	0.0
V _{Bluk} (V)	0.0	0.0	0.0	0.0	0.0
Time (ns)	5.0	5.0	5.0	5.0	5.0

Table 1. The operation parameters of the HDM.



Figure 3. The equivalent circuit of this device.



Figure 4. (a) The BTBT rate profile and (b) bandgap during writing-1 operation. The tunneling generation rate is highest where the band bending is largest. (c) The normalized reading and writing operation currents of the HDM. (d) The FG potential with the change in the erase-write 0-read-erase-write 1-read operations of the HDM.

According to the simulation results, the tunneling rates of the write-1 and write-0 processes were 1.551×10^{27} cm⁻³s⁻¹ and 4.812×10^{28} cm⁻³s⁻¹, respectively. This means that during the write-1 process, more carriers were written into the floating gate. Such tunneling follows the mechanism of band-to-band tunneling, and it occurred in the region of the tunneling window (shown in Figure 1) with $V_{CG} = -1.5$ V and $V_D = 1.2$ V. In this process, the energy band is stretched by the vertical electric field induced by V_{CG} . Under the horizontal electric field formed by V_D , the tunneling electrons flow to the drain, which causes the rise in FG potential. The write-1 and write-0 processes induce two different FG potentials so that during the read operation, we can obtain two different currents. As no vertical electric field is stretching the energy band, the bandgap broadens during the hold operation and tunneling is not likely to occur for the electrons. Under the application of the horizontal electric field, the carriers are trapped in the floating gate, preserving the information in a long-term manner. The length of the retention time is directly related to the size of the holding voltage, and the storage time of the information can be changed by changing the size of the retention voltage.

Figure 4c,d show the normalized tunneling current and FG potential curves of the HDM extracted from the second cycle because the currents are more stable and reproducible from our previous simulation experience. One operation period consists of the erase, retention, write-0, write-1, and read steps, as shown in Figure 4c,d. The operation voltages of the source and substrate were 0 V during the operation period, and the V_D and V_{CG} changed along with different operations. The operation time of each step was 5 ns. The simulation results showed that the device can achieve high-speed and continuous voltage operation, and has excellent storage performance and reliability. Compared with the traditional floating-gate semiconductor memory device, the device has a lower operating voltage and a faster operating speed, and the information retention time can be adjusted by tuning the retention voltage, which enables both volatile and nonvolatile memory capabilities in the device. In addition, compared with traditional dynamic random-access memory, this device has a faster speed, longer retention time (>1 s), and no capacitance bottleneck, which makes this device promising for future stand-alone and on-chip memory applications.

The change in drain current (I_D) during the second operation period is shown in Figure 4c. The CG voltage of the write-0 operation was the same as that of write-1, but the drain voltage of the write-1 operation was twice as high as that of write-0, resulting in the different amounts of tunneling electrons during the write operation. As discussed above, a different threshold voltage shift was observed with the change in FG potential. A different read current was obtained when the same read voltage was applied for read-0 and read-1 operations. The difference between the currents of read-0 and read-1 was larger than 2 μ A (0.74 μ A and 3.12 μ A, respectively), which is sufficiently distinguishable. Besides, the leakage was lower than 10^{-12} A of I_D during the hold step, which indicates that the device is reliable in the operation period.

Figure 4d shows the change in FG potential during the second operation period. The erase voltages were $V_{CG} = 1.2$ V and $V_D = 0.6$ V. The potential of FG decreased as expected due to the effect of the vertical electric field. The erase operation made the device return to the initial state. The write-1 voltages were $V_{CG} = -1.5$ V and $V_D = 1.2$ V, and the write-1 voltages were $V_{CG} = -1.5$ V and $V_D = 1.2$ V, and the write-0 and write-1 operations due to the band-to-band tunneling, and the charge was stored into the FG, which will change the potential of the FG. The FG potentials were 0.02 V and 0.10 V after the write-0 and write-1 operations, respectively. The different FG potential resulted in the difference in V_{th} during the read operation. As a result, we can obtain the read currents mentioned above, and the state "0" and state "1" can be distinguished by checking and comparing the read current. The change in the FG potential during the second operation period was consistent with I_D , which shows that the electrical performance of the device is very reliable.

During the write-0 and write-1 operations, the variation in FG potential is shown in Figure 5a. With different write-0 and write-1 voltages ($V_D = 0.6$ V and 1.2 V, respectively,

and higher programming efficiency under the condition of large V_D), the different amounts of charges are stored in the FG, so that the information of "0" and "1" can be established. The retention characteristic curve is shown in Figure 5b. State-0 and state-1 profiles are separately saved after write-0 and write-1 operations for further simulation. The operation voltages of retention characterization were $V_{CG} = 0$ V and $V_D = 0.6$ V. According to the simulation results, the retention time of state-0 and state-1 can reach up to ~1 s, which means the device does not have to be refreshed frequently. Due to the existence of the tunneling window region, a leakage current from the FG to drain always exists, and the FG potential and the drain potential are slowly consistent during the standby step. The increment in standby time of the HDM can be performed by altering V_D , which can affect the volatile and nonvolatile properties of the device. Therefore, the long retention time can be obtained by increasing the hold voltage V_D if necessary.



Figure 5. (a) The changes in the FG potential of the HDM with different write-0 and write-1 times with fixed voltages. (b) FG potential of state-0 and state-1 extracted at different retention times. (c) V_D write-1 disturb performance of state-0 and state-1. (d) V_D read disturb performance of state-0 and state-1.

The disturbance persistence capability is an important property of a memory array. As the array operation of the HDM is page-operated, the disturbance of the word line voltage (V_{CG}) can be ignored, while all disturbance is attributed by the bit line voltage (V_D). Moreover, the write-0 and write-1 operations share the same programming mechanism of varying V_D . In this work, four types of crosstalk mechanisms were presented, and the crosstalk performance of the HDM is shown in Figure 5c,d. The second operation cycle was used for crosstalk calculation, which is also consistent with previous studies. The state-0 and state-1 profiles were loaded to carry out the simulation by TCAD. The disturb time of state-0 and state-1 V_D -write-1 disturb remained ~1 s, as shown in Figure 5c, when disturb

voltages were $V_{CG} = -1.5$ V and $V_D = 1.2$ V. The operation time of the write process was 5 ns, which means the memory array can be written $\sim 10^8$ times. The disturb voltages of V_D-read disturb were V_{CG} = 0.6 V and V_D = 1.2 V, and the state-0 and state-1 V_D-read disturb remained ~ 1 s, as shown in Figure 5d, which indicates that the memory array can be executed by the read operation for $\sim 10^8$ times continuously. The simulation results of the four types of V_D disturb showed that HDM structure preserved very reliable performance within a memory array operation. The main characteristics are compared with SRAM in Table 2.

Parameters	SRAM	HDM
Area of cell	80–120 F ²	8 F ²
P/E Speed	<1 ns	<5 ns
Retention Time	-	>1 s
Non-destructive Readout	yes	yes

Table 2. The compared benchmarks of SRAM and the HDM [1-3].

4. Conclusions

In this work, a new mechanism of high-density memory was explored. This HDM relies on the band-to-band tunneling effect, which makes it a good alternative in low-power and high-speed memory applications. The process flow, electrical performance, and stability of the HDM device were simulated by using TCAD tools. The HDM process flow had good CMOS compatibility, and it could enable further downscaling of the semiconductor memory, which makes it have high density. The charging and discharge of the floating gate was assisted by BTBT tunneling in this device, which provides a solid basis for high speed, low power consumption, and high density as compared with 6T-SRAM, which may have potential applications in future memory cache applications.

Author Contributions: Writing—Original Draft, C.W.; Validation, X.Z.; Methodology, H.L. and X.C.; Writing—Review and Editing, H.Z.; Supervision, Q.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Science and Technology Commission of Shanghai Municipality (21DZ1100700), 02 State Key Project (2017ZX02315005), Program of Shanghai Subject Chief Scientist (18XD1402800), and the Support Plans for the Youth Top-Notch Talents of China.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Nayak, D.; Acharya, D.P.; Mahapatra, K. A Read Disturbance Free Differential Read SRAM Cell for Low Power and Reliable Cache in Embedded Processor. *AEU Int. J. Electron. Commun.* **2017**, 74, 192–197. [CrossRef]
- Kim, T.H.; Jeong, H.; Park, J.; Kim, H.; Song, T.; Jung, S.-O. An Embedded Level-Shifting Dual-Rail SRAM for High-Speed and Low-Power Cache. *IEEE Access* 2020, *8*, 187126–187139. [CrossRef]
- Prasad, G.; Tandon, D.; Isha; Mandi, B.C.; Ali, M. Process Variation Analysis of 10T SRAM Cell for Low Power, High Speed Cache Memory for IoT Applications. In Proceedings of the 2020 7th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 27–28 February 2020. [CrossRef]
- 4. Kerber, P.; Zhang, Q.; Koswatta, S.; Bryant, A. GIDL in Doped and Undoped FinFET Devices for Low-Leakage Applications. *IEEE Trans. Electron. Devices* 2012, 34, 154–159. [CrossRef]
- Lee, J.-M.; Choi, P.-H.; Kim, S.-K.; Oh, J.-H.; Shin, S.-H.; Noh, J.-Y.; Kim, H.-S.; Choi, B.-D. New Method for Reduction of the Capacitor Leakage Failure Rate Without Changing the Capacitor Structure or Materials in DRAM Mass Production. *IEEE Trans. Electron. Devices* 2018, 65, 4839–4845. [CrossRef]
- Yoshida, E.; Tanaka, T. A capacitorless 1T-DRAM technology using gate-induced drain-leakage (GIDL) current for low-power and high-speed embedded memory. *IEEE Trans. Electron. Devices* 2006, 53, 692–697. [CrossRef]
- Kim, K.; Jeong, M.-Y. The COB stack DRAM cell at technology node below 100 nm-scaling issues and directions. *IEEE Trans.* Semicond. Manuf. 2002, 15, 137–143. [CrossRef]

- 8. Chen, J.-H.; Wong, S.-C.; Wang, Y.-H. An analytic three-terminal band-to-band tunneling model on GIDL in MOSFET. *IEEE Trans. Electron. Devices* **2001**, *48*, 1400–1405. [CrossRef]
- 9. Wang, T.; Chang, T.-E.; Chiang, L.-P.; Wang, C.-H.; Zous, N.-K.; Huang, C. Investigation of oxide charge trapping and detrapping in a MOSFET by using a GIDL current technique. *IEEE Trans. Electron. Devices* **1998**, *45*, 1511–1517. [CrossRef]
- Mandelman, J.A.; Dennard, R.; Bronner, G.; Divakaruni, J.D.R.; Radens, C. Challenges and future directions for the scaling of dynamic random-access memory (DRAM). *IBM J. Res. Dev.* 2002, 46, 187–212. [CrossRef]
- 11. Choi, S.-J.; Han, J.-W.; Moon, D.-I.; Choi, Y.-K. Analysis and evaluation of a BJT-based 1T-DRAM. *IEEE Electron Device Lett.* 2010, 31, 393–395. [CrossRef]
- 12. *Sentaurus User's Manual*; ver C-2013.03; Synopsys, Inc.: Mountain View, CA, USA, 2013. Available online: https://picture.iczhiku.com/resource/eetop/wYItAzFDYgQzfNvV.pdf (accessed on 20 May 2021).
- Hisamoto, D.; Kimura, S.; Kaga, T.; Nakagome, Y.; Isoda, M.; Nishida, T.; Takeda, E. A New Stacked Cell Structure for Giga-Bit DRAMs Using Vertical Ultra-Thin SO1 (DELTA) MOSFETs. In Proceedings of the International Electron Devices Meeting 1991 [Technical Digest], Washington, DC, USA, 8–11 December 1991; pp. 959–961.
- 14. Wang, P.; Lin, X.; Liu, L.; Sun, Q.; Zhou, P.; Liu, X.; Liu, W.; Gong, Y.; Zhang, D.W. A Semi-Floating Gate Transistor for Low-Voltage Ultrafast Memory and Sensing Operation. *Science* **2013**, *341*, 640–643. [CrossRef] [PubMed]
- 15. Jiang, S.Y.; Yuan, Y.; Wang, X.; Chen, L.; Zhu, H.; Sun, Q.Q.; Zhang, D.W. A Semi-Floating Gate Transistor with Enhanced Embedded Tunneling Field-Effect Transistor. *IEEE Electron Device Lett.* **2018**, *39*, 1497–1499.
- Lin, X.; Liu, X.Y.; Zhang, C.M.; Liu, L.; Shi, J.S.; Zhang, S.; Wang, W.B.; Bu, W.H.; Wu, J.; Gong, Y.; et al. Investigation of temperature dependence device scalability and modeling of semifloating-gate transistor memory cell. *IEEE Trans. Electron Devices* 2015, 62, 1177–1183. [CrossRef]
- 17. Wan, J.; Le Royer, C.; Zaslavsky, A.; Cristoloveanu, S. Progress in Z2-FET 1T-DRAM: Retention time, writing modes, selective array operation, and dual bit storage. *Solid State Electron.* **2013**, *84*, 147–154. [CrossRef]