Direct Tunneling Memory

● Tatsuya Usuki ● Kouji Tsunoda ● Akira Sato ● Toshiro Nakanishi
● Hitoshi Tanaka

(Manuscript received February 24, 2003)

To realize a random accessible storage device, we propose the Direct Tunneling Memory (DTM). This is a sub-0.1 μm floating gate memory that uses direct tunneling at a lower operating voltage than flash EEPROM: the applied word line voltage of DTM is 6 V, and the bit line voltage is 2 V. The oxide barrier of DTM memory is 3 nm or less, which is much thinner than the 10 nm oxide barrier of flash EEPROM (10 nm). However, the new structure has a long retention time. The DTM structure can be made using a logic compatible fabrication-process, since, unlike FeRAM and MRAM, it does not use any new materials. We will use DTM to fabricate a large, low power consumption memory that can be embedded into a System-on-a-Chip (SOC).

1. Introduction

The demand for highly efficient digital equipment such as cell phones and mobile computers continues to increase, and there is a strong need to embed large-scale, low-power-consumption memory into SOCs. However, SRAM, DRAM, flash, and the other types of conventional memory are unsuitable choices for these applications (**Table 1**) and much work is being done to develop a memory that is based on a new principle of operation.

Table 1
Features and problems of typical memories.

	Feature	Problems
SRAM	High-speed operation	• Large cell size • Cell reduction increases standby current
DRAM	High capacity	• Special material and process • Hard to embed DRAM in SOCs • Hard to decrease standby current owing to refresh
Flash	Nonvolatile	• High-voltage and low-speed operation • Writing restriction (< 10^6 times)

Fundamental memory performance is being improved with FeRAM and MRAM by introducing a new material. However, the new types of memory that have been developed for embedding in SOCs overcome the problems listed in Table 1 without using new materials. They therefore have the big advantage that they do not require investments in new equipment. This paper describes one of these new types of memory: the Direct Tunneling Memory (DTM).[1]

Like flash memory, DTM uses a floating gate (FG) and can be fabricated from silicon and silicon-oxide films. Also, DTM has a simple device structure and is fabricated using the existing process technology of logic transistors. To achieve non-volatility and low-leakage in a flash memory, the gate oxide between the FG and channel must be about 10 nm thick. On the other hand, the oxides in DTM are 3 nm or less and the oxides in the MOSFETs used in the latest logic circuits are even thinner. However, even though these MOSFETs use extremely thin oxides, the physical properties of these oxides are gradually

becoming clear.

2. Principle of operation and structure of DTM

When the oxide thickness is reduced to the nano scale, its physical properties change considerably. One of these properties is the tunnel current through the oxide film, which increases exponentially as the thickness is reduced. Although this increase poses a problem in logic circuits, flash memory and DTM use this tunnel current to operate. DTM can obtain sufficient tunnel current at a much lower operating voltage than flash memory. As a result, DTM can be operated at a higher speed and lower voltage. Moreover, because DTM oxide films are subjected to a lower voltage, their reliability is greatly superior to that of flash-memory oxide fims.[2] In flash memory, the Fowler-Nordheim (FN) tunneling and channel hot electron (CHE) injection that are performed during writing and erasing degrades the oxide and restricts the lifetime to about 10^6 writes. On the other hand, with the right oxide thickness and operating voltage, DTM might be capable of practically an unlimited number of writes.

However, when the oxide thickness is reduced, the electric charge at the FG of conventional flash memory quickly leaks to the channel region {**Figure 1 (a)**}. The structure of DTM suppresses this leakage {**Figure 1 (b)**} because of two features:

- The source and drain regions do not overlap the FG. The control gate (CG) is formed on both sides of the FG using a self-aligned process. Since this structure also strengthens the electrical connection between the CG and FG, it also enables the operation voltage to be reduced.
- While the channel impurities are highly concentrated as in logic, a 20 to 30 nm depletion region is established in the channel side of the FG. In this structure, a large potential ΔE of about 0.7 eV arises between the channel and FG. This large potential prevents the leakage that occurs when electrons stored in the FG become thermally excited and tunnel to the channel. Moreover, since the electric charge accumulates away from the oxide interface, the leakage current through the interface state can also be reduced.[3)]

Because of this structure, DTM achieves high-speed writing and a long retention time at a low voltage. In the next section, we describe the main parameters of DTM.

3. Main parameters of DTM

The main parameters of DTM are the retention time τ_R and the write-in time τ_w. These two parameters are strongly related to the tunnel oxide thickness T_{OX} as follows:

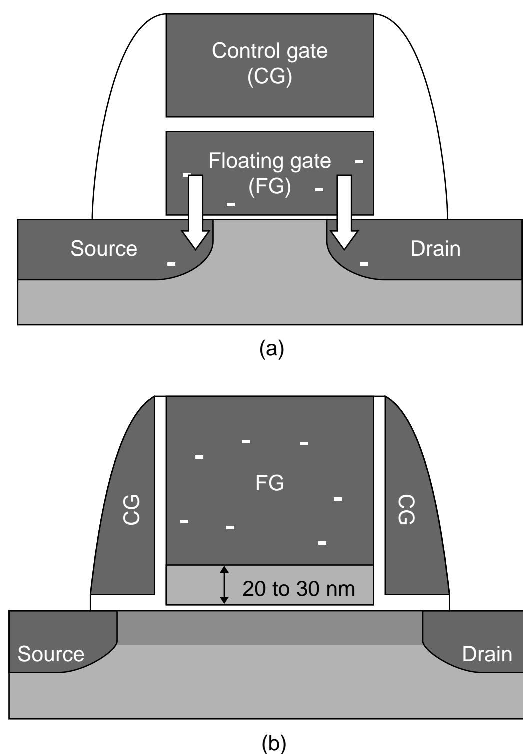

Figure 1
(a) Conventional floating gate (FG) memory structure. Stored charge in FG escapes to source and drain regions when the tunnel oxide is thin.
(b) DTM structure. Control gate (CG) of DTM is formed by a self-aligned process. In this arrangement, there is no overlap between the floating gate (FG) and the source or drain.

$$\tau_R, \tau_W \propto \exp\left(\frac{T_{OX}}{T_0}\right), \quad (1)$$

where T_0 is a constant that depends on the oxide quality and applied voltage. In the case of an oxidization film, T_0 is about 0.09 nm. The other important parameter is the ratio of τ_R and τ_w, which depends on the built-in potential ΔE. If the applied voltage dependence of tunnel probability is neglected as the 0th approximation, then:[note)]

$$\frac{\tau_R}{\tau_W} \approx \exp\left(\frac{\Delta E}{k_B T_e}\right), \quad (2)$$

where k_B is the Boltzmann constant and T_e is the electronic temperature. When T_{OX} is small, both τ_R and τ_w become short and the above equations indicate high-speed operation equivalent to that of DRAM. On the other hand, it is clear that retention time can be prolonged when T_{OX} is large. The built-in potential ΔE in Eq. (2) depends on the thickness and channel concentration of the FG depletion layer. ΔE does not exceed the band gap of Si, since it is generated by the p-n junction formed between the channel and FG. We can also apply Eq. (2) to DRAM that use an Si-MOSFET to control an electric charge stored in a capacitor. When only the ratio of τ_R and τ_w is considered, DTM and DRAM are both restricted by the band gap of Si. In summary, by adjusting T_{OX}, the memory retention time of DTM can be greatly increased while maintaining the same τ_R/τ_w as DRAM.

Next, we will describe the electrical properties of an experimental DTM cell with a T_{OX} of 2.8 nm that we fabricated. The memory retention characteristics of this device are shown in **Figure 2**. The figure shows how the threshold voltage V_{th}, which equals V_{CG} when I_D = 10 pA/μm and V_D = 0.1 V, changes after writing a 0 and 1 (equivalent to an erase in flash memory). As can be seen, even after 70 days, there is a clear difference in V_{th}. This demonstrates that even though the oxide of DTM is 7 nm or more thinner than that of flash memory, it can still effectively suppress leakage current.

Figure 3 shows the write-in characteristics for a positive write-in pulse. The horizontal axis is the pulse width of the write-in signal applied to the CG. Even when the applied voltage V_{CG} is 6 V, which is quite low compared to the V_{CG} in flash memory, a 0 can be written within 50 μs. Although it is not shown in this figure, when a negative pulse is applied to the CG, V_{th} shifts in the opposite direction and a 1 is written. The write-in characteristics for a negative pulse mirror those for a positive pulse, and 0 writing and 1 writing can be done in the same time scale.

Flash memory has several problems, for example, over-erase and a complicated write/erase algorithm. However, in DTM, since memory main-

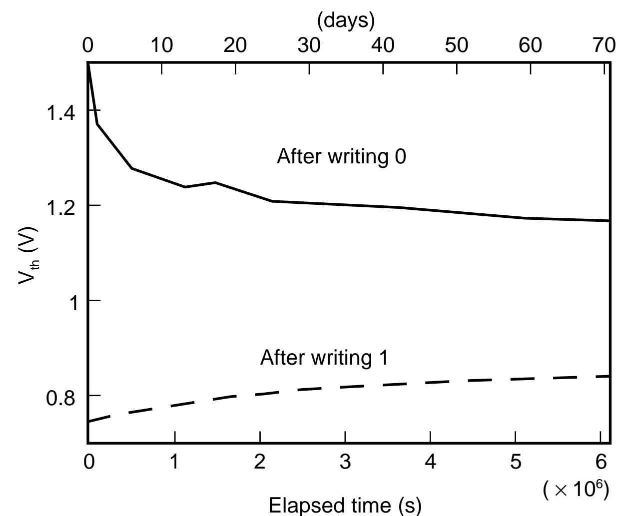

Figure 2
Retention characteristics after writing 0 and 1 with a CG voltage at I_D = 10 pA/μm for V_D = 0.1 V.)

note) The conditions of flash memory operation are mentioned here as an exception. In FN tunneling, a sufficiently high voltage applied to the insulator greatly affects the tunneling probability. In Si oxide, this occurs when the applied voltage exceeds 3 V. On the other hand, if T_e increases by a large amount at writing, τ_R/τ_w becomes large and the effect is equivalent to the effect of CHE injection. Since the electron energy in both cases increases, the oxide can be seriously damaged. Therefore, if τ_R/τ_w is so large that Eq. (2) no longer applies, the device cannot be used as a RAM.

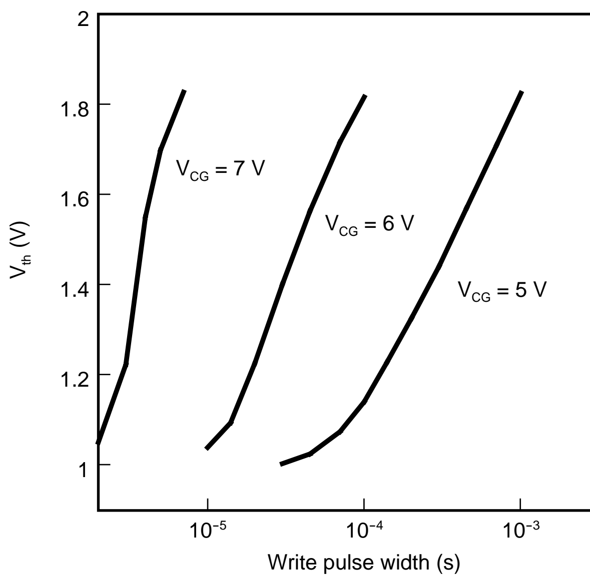

Figure 3
Write-in characteristics for writing 0. Write-in speed increases exponentially as applied voltage is increased.

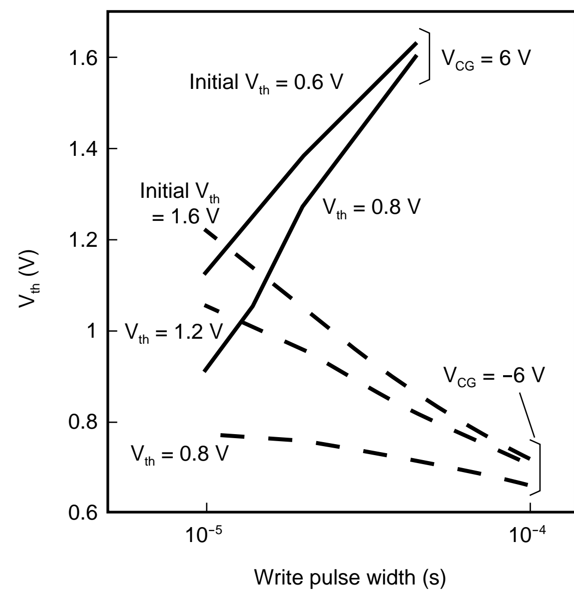

Figure 4
Convergences of write-in characteristics under various initial conditions.

tenance is performed by ΔE, over-write and over-erase do not happen. **Figure 4** shows how the write-in characteristics at various initial conditions converge. As can be seen, for 0 and 1 writing, the characteristics converge at the same value of V_{th}. Note that V_{th} does not become negative when a 1 is written.

Although DTM is not a nonvolatile memory, a retention time of several months has already been attained for quasi-nonvolatile DTM. Moreover, compared with conventional nonvolatile memory, 0 and 1 can be written simply and quickly at a low voltage. If DTM is used for non-ROM applications, it must have outstanding write-in endurance.[4] We therefore put our DTM through 10^9 0-1 write cycles at ±6 V and found that there was no change in V_{th}. We did not test beyond 10^9 cycles because of time restrictions. However, in high-speed DTM (T_{OX} = 1.9 nm), we did not observe a change in V_{th} even after 10^{11} cycles. During writing, the tunnel oxide of DTM is subjected to much less stress than the same oxide in a logic transistor. However, because the oxide between the CG and channel is subjected to the full V_{CG}, stress-induced leakage current (SILC) will occur if the oxide is too thin. Although SILC in the CG-channel oxide is not related to electric charge retention in the FG, it affects the reading of 0 and 1. Therefore, the CG-channel oxide should be 8 nm or thicker.

We also investigated the influence of write-in on the non-selected cells of an AND array composed of our DTM.[5] The investigation showed that the influence, which is called write disturbance, was sufficiently controlled.

4. Conclusion

In this paper, we introduced a new type of floating gate memory called direct tunneling memory (DTM). Measurements of the write-in and retention characteristics of DTM indicate that the new memory attains a large τ_R/τ_W of about 10^{10}. This value is almost the same as that of DRAM, and even larger values of τ_R/τ_W can be realized with DTM. The memory retention time of DTM can be greatly increased by adjusting the tunnel oxide. DTM has a simple structure, uses the same materials as a logic transistor, and has passed the trial-production stage.

We will now develop the array peripheral cir-

cuits and architecture for a memory chip that will use our DTM technology. By using an optimized architecture, we can lengthen the DTM refresh interval to a different order from that of DRAM. Although the write-in speed of this optimized DTM will be less than that of DRAM, its read-out speed will only be slightly influenced by the oxide thickness. By designing the architecture so that the write-in processing delay does not degrade the performance, we can achieve a memory with an extremely low power consumption.

The use of alternative structures to build memory that is controlled using a logic process has been proposed[6] and is being studied aggressively. However, as long as we continue to use the materials used to fabricate logic, we should bear in mind the relationship between τ_R and τ_w shown in Eq. (2). Therefore, to lengthen the refresh period, we have to lengthen the write-in time. It is now important to design an architecture that takes this relationship into full consideration.

References

1) N. Horiguchi, T. Usuki, K. Goto, T. Futatsugi, T. Sugii, and N. Yokoyama: A Direct Tunneling Memory (DTM) utilizing novel floating gate structure. IEDM Tech. Dig., p.922 (1999).
2) E. Y. Wu, J. Aitken, E. Nowak, A. Vayshenker, P. Varekamp, G. Hueckel, J. McKenna, D. Harmon, L.-K. Han, C. Montrose, and R. Dufresne: Voltage-Dependent Voltage Acceleration of Oxide Breakdown for Ultra-Thin Oxides. IEDM Tech. Dig., p.541 (2000).
3) A. Ghetti, E. Sangiorgi, J. Bude, T. W. Sorsch, and G. Weber: Low Voltage Tunneling in Ultra-Thin Oxides: A Monitor for Interface States and Degradation. IEDM Tech. Dig., p.723 (1999).
4) T. Usuki, N. Horiguchi, and T. Futatsugi: Direct Tunneling Memory: Trade-off between Nonvolatility and High Endurance with Low Voltage Operations. NVSMW2001, p.80 (2001).
5) T. Usuki, N. Horiguchi, and T. Futatsugi: Advantage of a quasi-nonvolatile memory with ultra thin oxide. SSDM2001, p.532 (2001).
6) K. Nakazato, K. Itoh, H. Ahmed, H. Mizuta, T. Kisu, M. Kato, and T. Sakata: Phase-state Low Electron-number Drive Random Access Memory (PLEDM). ISSCC2000, p.132 (2000).

Tatsuya Usuki received the B.S., M.S., and Ph.D. degrees in Department of Applied Physics from Osaka University, Osaka, Japan, in 1986, 1988, and 1991, respectively. He joined Fujitsu Laboratories Ltd., Atsugi, Japan in 1991, where he has been engaged in research of semiconductor physics and development of novel electron devices. He is a member of the Japan Society of Applied Physics, IEEE Electron Device Society, and American Physical Society.

E-mail: usuki.tatsuya@jp.fujitsu.com

Kouji Tsunoda received the B.S. and M.S. degrees in Electronic Engineering from the University of Tokyo, Tokyo, Japan in 1995 and 1997, respectively. He joined Fujitsu Laboratories Ltd., Atsugi, Japan in 1997, where he has been engaged in research and development of Si memory devices.

E-mail: tsunoda.kouji@jp.fujitsu.com

Akira Sato received the B.S. degree in Electronic Engineering from Niigata University, Niigata, Japan in 1986. He joined Fujitsu Laboratories Ltd., Atsugi, Japan in 1986, where he was engaged in research of Si surfaces and interfaces. He was responsible for the development of MIM capacitors at the Fujitsu/Toshiba DRAM joint development project from 1999 to 2001. He is currently engaged in memory device development.

E-mail: aksatoh@jp.fujitsu.com

Toshiro Nakanishi received the B.E. and M.E. degrees in Electronics from Kobe University, Kobe, Japan, in 1981 and 1983, respectively, and the D.Eng. degree from Tohoku University, Sendai, Japan in 1998. He joined Fujitsu Laboratories Ltd., Kawasaki, Japan in 1983, where he was engaged in research of Si and SiO_2. He was responsible for the development of hot processes at the Fujitsu/Toshiba DRAM joint development project from 1999 to 2001. He is currently engaged in memory device development.

E-mail: nakani@jp.fujitsu.com

Hitoshi Tanaka received the B.S., M.S., and Ph.D. degrees from the Department of Reaction Chemistry of the University of Tokyo, Tokyo, Japan, in 1979, 1981, and 1988, respectively. He joined Fujitsu Laboratories Ltd., Atsugi, Japan in 1984, where he has been engaged in research and development of semiconductor technologies for memory and other high-speed devices. He is currently responsible for research and development of nonvolatile semiconductor memory devices.

E-mail: hit.tanaka@jp.fujitsu.com