Extended Abstracts of the 2004 International Conference on Solid State Devices and Materials, Tokyo, 2004, pp. 174-175

A-5-2

Optimization of Low Power and High Speed DTM for Embedded RAM Applications

H. Tashiro, K.Tsunoda, A.Sato, T.Nakanishi and H.Tanaka

Fujitsu Laboratories Ltd.
10-1 Morinosato- Wakamiya, Atsugi, Kanagawa 243-0197, Japan
Phone: +81-46-250-8215, Fax: +81-46-250-8804, E-mail: tashiro.hiroko@jp.fujitsu.com

1. Introduction

Direct tunneling memory (DTM) is an attractive candidate for embedded random access memory (RAM) because of its simple structure, fabrication process, principle of operation and needlessness of new materials [1]. We have realized an ultra-fast operation of less than 10ns at 5V so far [2]. However, lower power consumption is necessary for mobile communication applications, which requires both lower voltage operation and excellent retention time. In this paper, we report the result of device design for low voltage (3.3V) operation of DTM using device simulation.

2. Device Structure and Model

The structure of DTM is shown in Fig. 1. DTM is a floating gate (FG) memory featuring the sidewall control gate (CG) and offset source/drain extension to enhance the retention time with ultra-thin tunneling oxide. Fig. 2 shows the simulated device structure and its parameters. We used ISE TCAD DESSIS code [3]. Trap-assisted tunneling component was not evolved in this simulation.

3. Results and Discussion

Fig. 3 shows the effect of offset spacing on retention characteristics at erased state. Even if L_{offset} is small, a great improvement as compared with the source/drain overlap case was obtained at smaller ΔV_{th}, showing the excellent scalability of DTM cell structure. In Fig. 4, the influence of N_{fg} and N_{sub} on the retention time was also evaluated. Gate depletion and higher substrate concentration improve the retention time by several orders of magnitude. This is due to the leakage current suppression caused by the reduced band bending at substrate surface. Because of the large flatband voltage, the band bending of substrate surface and resultant increase of conduction band (CB) energy gap between FG and substrate are also effective to improve the retention time of weakly programmed state [4] as shown in Fig. 5 and Fig. 6. Moreover, since the electron density decreases by the band bending at FG/oxide interface, the leakage current component through the interface states can be suppressed. Experimental N_{fg} dependence of retention time shown in Fig. 7 supports the simulation results.

Fig. 8 shows the relationship between programming time and ΔV_{th}. Programming speed degrades only slightly (1/2 to 1/3) by gate depletion, which is relatively insignificant impact as compared with the drastic improvement of retention time. Furthermore, the value of L_{offset} and N_{sub} hardly influences the programming transient. Such an insensitivity of programming performance is a feature of DTM because the direct tunneling current is determined mainly by the voltage difference between FG and source/drain region once sidewall CG transistor turns on. Scaling the control oxide thickness (T_{con}) is effective to enhance the FG voltage by increasing the coupling constant defined as $\gamma=C_{CG}/(C_{FG}+C_{CG})$. Therefore, injected charge into FG with same programming condition increases with thinner T_{con} as shown in Fig. 9. However, ΔV_{th} decreases with increasing C_{CG}. As a result, programming performance is not significantly improved by the scaling of T_{con}. On the other hand, if the thickness of CG bottom oxide (T_{oxc}) thins, ΔV_{th} of the erase side increases because V_{th} is limited by CG transistor at erased state. As a consequence, the programming time improves without changing charge injection transient as shown in Fig. 10.

Effective parameters to improve the retention time are N_{sub} of high concentration and N_{fg} of low concentration with arbitrary L_{offset} as summarized in Table I. Thinning the T_{oxc} is effective for improving programming time. Fig. 11 shows the relationship between programming time and retention time with various parameters. For memory design, T_{oxf} should be fixed at first to keep the operation speed. In our optimized parameter set with T_{oxf}=1.2nm, programming time of about 30ns with V_{cg}=3.3V and retention time of more than 10s can be attained, which is enough competitive to the conventional DRAM cell.

4. Conclusions

Our analysis revealed that the poly-Si/substrate band engineering based on gate depletion effectively improves the retention time of DTM by several orders. By optimizing the device parameters, we obtained the prospect of achieving the excellent program/retention performance that equals to DRAM with lower operating voltage of 3.3V. Low power and high speed DTM of its simple structure and fabrication process is one of the promising embedded memories for near future.

Acknowledgements

This work is supported by National Institute of Information and Communications Technology (NiCT) Grant.

References

[1] N.Horiguchi et al., *IEDM Tech. Dig.* (1999) 922.
[2] K.Tsunoda et al., *VLSI Tech. Dig.* (2004)15.4.
[3] ISE TCAD Release 9.5 (Device Simulation)
[4] A.Ghetti et al., *IEDM Tech. Dig.* (1999) 723.

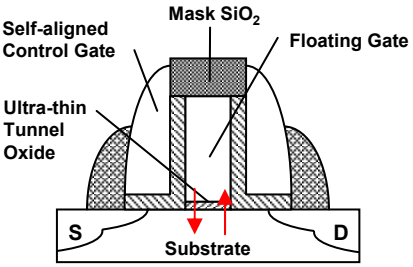

Fig.1 Schematic cross-sectional view of the DTM having the ultra-thin tunnel oxide, mask SiO_2 layer and self-aligned CG.

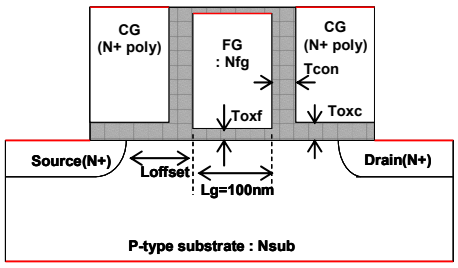

Fig.2 Simulated DTM structure and device parameters.

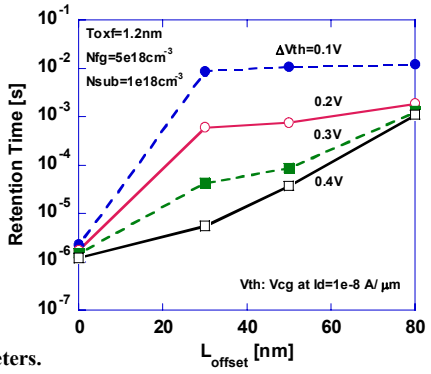

Fig.3 Retention characteristics at erased state with different source/drain offset spacing. ΔV_{th} is the threshold voltage difference to distinguish programmed and erased states.

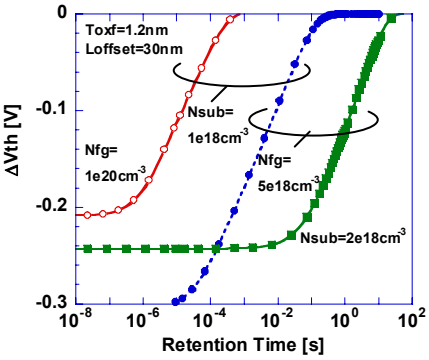

Fig.4 Retention characteristics at erased state with different N_{fg} and N_{sub}.

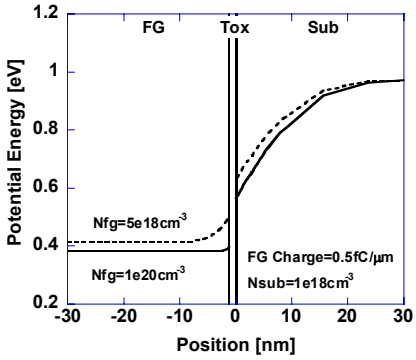

Fig.5 Energy band diagram of conduction band edge at programmed state with different N_{fg}.

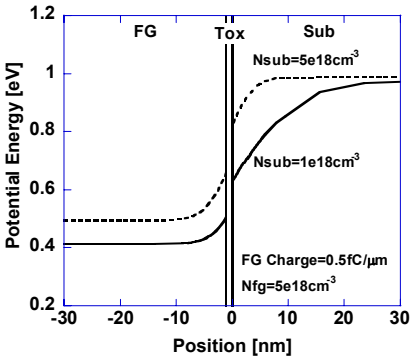

Fig.6 Energy band diagram of conduction band edge at programmed state with different N_{sub}.

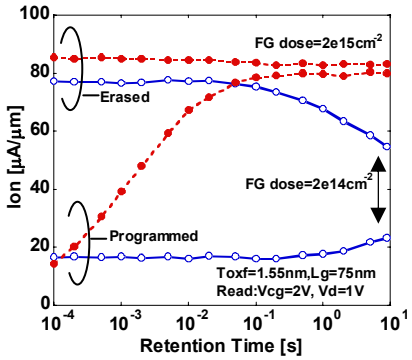

Fig.7 Measured relationship between retention time and drain current(I_{on}) with different FG dose(N_{fg}).

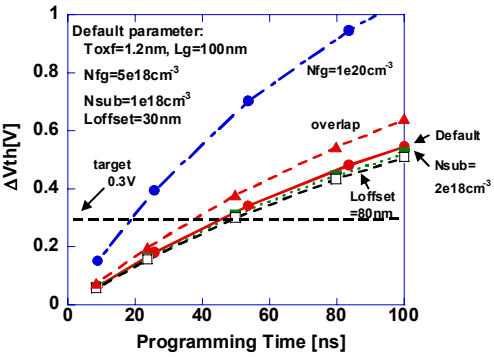

Fig.8 Programming transient at Vcg=3V with different device parameters.

TableI Effect of each parameters on retention time and programming time. Each value denotes the ratio of optimized time to initial one.

Parameter	Initial	Optimized	Ratio of Optimized/Initial	
			Retention time	Programming time
N_{fg}[cm-3]	1e20	5e18	1000	2
N_{sub}[cm-3]	1e18	3e18	1e6	1
L_{offset}[nm]	overlap	80	1000	1.2
T_{oxf}[nm]	1.5	1.2	0.01	0.08
T_{con}[nm]	10	5	10	1
T_{oxc}[nm]	8	4	1	0.5

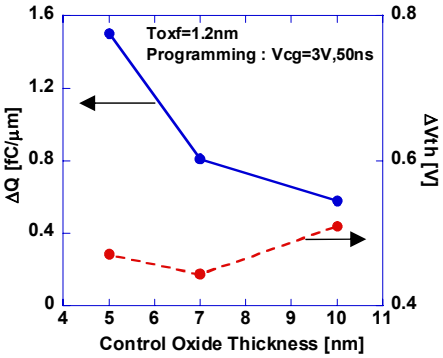

Fig.9 Dependence of injected charge and V_{th} shift on control oxide thickness (T_{con}) after programming.

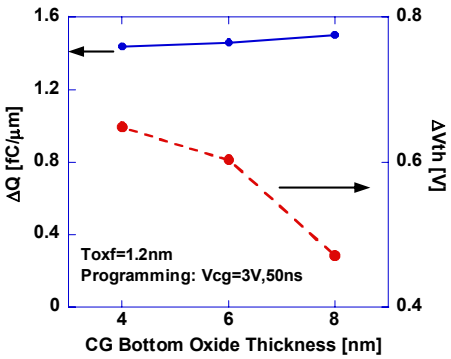

Fig.10 Dependence of injected charge and V_{th} shift on CG bottom oxide thickness (T_{oxc}) after programming.

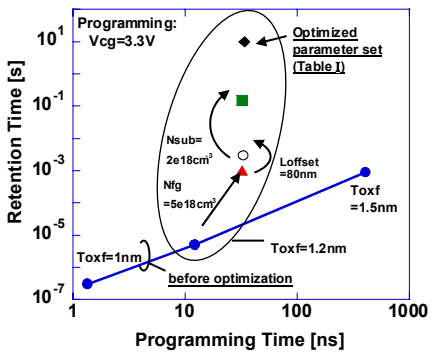

Fig.11 Improvement of retention/programming time ratio by optimizing device parameters.
Parameters before optimization:
N_{fg}=1e20cm-3, N_{sub}=1e18cm-3 and L_{offset}=30nm