第 23 卷第 7 期 半 导 体 学 报 Vol. 23, No. 7
2002 年 7 月 CHINESE JOURNAL OF SEMICONDUCTORS July, 2002

A Novel Flash Memory Using Band-to-Band Tunneling Induced Hot Electron Injection to Program*

Pan Liyang, Zhu Jun, Liu Zhihong, Zeng Ying and Lu Yong

(*Institute of Microelectronics, Tsinghua University, Beijing* 100084, *China*)

Abstract: A novel band-to-band hot electron programming flash memory device, which features programming with high speed, low voltage, low power consumption, large read current and short access time, is proposed. The new memory cell is programmed by band-to-band tunneling induced hot-electron (BBHE) injection method at the drain, and erased by Fowler-Nordheim tunneling through the source region. The work shows that the programming control gate voltage can be reduced to 8V, and the drain leakage current is only 3μA/μm. Under the proposed operating conditions, the program efficiency and the read current rise up to 4×10^{-4} and 60μA/μm, respectively, and the program time can be as short as 16μs.

Key words: flash memory; band-to-band; channel hot electron; Fowler-Nordheim
EEACC: 2560A; 1265D
CLC number: TP333 **Document code:** A **Article ID:** 0253-4177(2002)07-0690-05

1 Introduction

In recent years, a novel semiconductor non-volatile memory, which is referred to as the flash memory, has been developed. Typically, flash memory is programmed by channel hot electron (CHE)[1] injection at the drain edge or by Fowler-Nordheim tunneling injection[2]. The traditional programming of channel hot electron injection has a high program speed. However, its programming power consumption is as high as 24mW/byte and its efficiency of injection is very low, because the memory transistor is turned on when programming. The Fowler-Nordheim tunneling injection programming, though has high injection efficiency, requires relatively high voltage biasing on the terminals of the memory cell.

In 1992, Hsu *et al*. proposed using band-to-band hot electron(BBHE) injection to realize low voltage, low current and high speed programming[3]. After that, there were several flash memory cells[4~6] introduced by using this programming mechanism. Because the band-to-band tunneling induced hot electrons are impossibly emitted to the gate in NMOS transistor due to the negative gate voltage, all of those cells are based on p-channel, whose read current is smaller than one half of the one with n-channel. Chen *et al*. also proposed a band-to-band and tunneling induced substrate hot-electron (BBISHE) injection programming mechanism used in the n-channel flash memory[7]. However, the surface concentration at the channel should be higher than $1 \times 10^{18}\,cm^{-3}$, which makes the threshold voltage almost out of control.

In this paper, we described a novel band-to-band

* Project supported by the Ninth 5-Year Plan Project(No. 97-760-01-01)
Pan Liyang male, was born in 1975, PhD candidate. His research field is in the device, technology, and reliability of flash memory.
Received 3 November 2001, revised manuscript received 17 January 2002
©2002 The Chinese Institute of Electronics

tunneling induced hot electron injection programming flash memory cell with n-channel. The structure of the cell and the programming mechanism were firstly described. Furthermore, the programming, erasing and reading characteristics were discussed, and some testing results were also shown at the end.

2 Cell structure

A schematic illustration of the proposed flash memory cell is shown in Fig. 1. A stacked gate structure, which includes a control gate and a floating gate, is firstly formed on a p-type substrate, then an n-type implantation is used to form the source region and the n-halo region; a p^+ region is finally formed in the n-halo region and comes into being the drain of the cell. This cell structure is nearly the same as the traditional ETOX structure, except for an accessorial p^+ drain region. The concentration in the n-halo region and the p^+ drain region are nearly $1 \times 10^{19} cm^{-3}$ and $6 \times 10^{19} cm^{-3}$, respectively. The thickness of tunneling oxide is 10nm and the electric thickness of the oxide/nitride/oxide interpoly dielectric is 20nm. The gate length of the samples is between 0.8μm and 1.6μm, and the coupling ratio of the control gate is about 0.5.

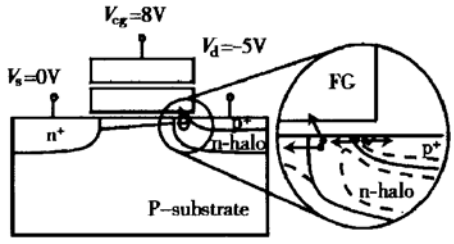

Fig. 1　Cell structure of the proposed flash memory cell and insert schematic illustration of the band-to-band tunneling induced hot electron injection programming method for the n-channel cell

When a negative drain voltage and a positive control gate voltage are applied to the cell with the source and p-substrate being set to 0V, the high positive control gate voltage can turn on the memory cell, and the floating n-halo region will be set to the same potential as the source (as shown in Fig. 1). As a result, the drain p^+/n junction is negatively biased, and electrons in the valence band can be emitted to the conduct band by band-to-band tunneling at the surface of the p^+ region. These band-to-band tunneling generated electrons may also flow into the n-halo region and induce hot electrons by gaining energy through the junction electric field and/or impact ionization. Most of the hot electrons will be collected by the cell source, and other hot electrons with high energy can overcome the barrier and are injected into the floating gate through the tunnel oxide with the aid of the positive bias on the control gate. The injection of such band-to-band hot electrons and the negative gate source side Fowler-Nordheim tunneling were used to perform the programming process for the proposed flash memory.

In order to perform the read operation, the drain was set to between 1 and 1.5V, the control gate was set to the access voltage, and other ends were set to ground. Under this bias condition, the drain pn junction was positively biased and the voltage of the n-halo region was clamped at 0.4~0.9V. With the access voltage, the p-type channel region may be inverted (in the "0" cell) so that the electrons can flow from the source to the n-halo and the read operation is performed. Table 1 illustrates the operating bias of the proposed flash memory cell.

Table 1　Operation condition of the flash cell

Operation	V_d/V	V_{cg}/V	V_s/V	V_b/V
Write	-5	8	0	0
Erase	floating	-8	5	0
Read	1.5	3	0	0

3 Results and discussion

3.1 Read

It can be seen that the p-drain, the p-substrate and the n-halo will make up of a parasitic PMOS transistor, as shown in Fig. 2. When V_{cg-d} ($V_{cg-d} = V_{cg} - V_d$) is smaller

than its threshold voltage, the parasitic PMOS will be on. The lower the control gate voltage is set, the smaller the drain voltage is needed to turn on the parasitic PMOS.

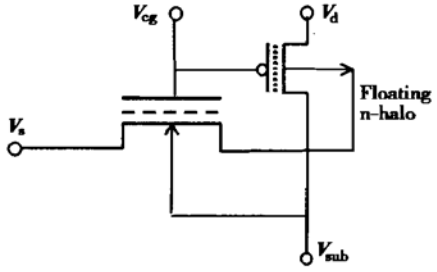

Fig. 2 Schematic illustration of the parasitic PMOS made up of the p-drain, the p-substrate and the n-halo

Figure 3 shows the I_d-V_d curves of the proposed flash memory cell. If drain voltage is higher than 1.8V, the PMOS will be turned on when control gate voltage equals to 0(which is typical in unselected cells), and the leakage current from the p-drain to the p-substrate will increase and cause the read error. In order to prevent the PMOS turning-on effect and offer a sufficient read current, the drain voltage is controlled between 1V and 1.5V.

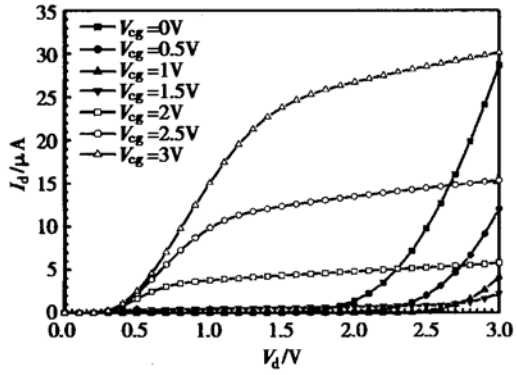

Fig. 3 I_d-V_d curves of the proposed flash cell When V_d> 1.8V, the increasing of the drain current is caused by the parasitic PMOS turn-on (gate length equals to 1.6μm).

Figure 4 shows the I_d-V_{cg} characteristic with a 1.5V drain voltage. When the control gate voltage is 3V, the read current can be larger than 60μA/μm in the 0.8μm cell, which is twice the access currents in other introduced p-channel flash memory cells.

3.2 Band-to-band programming

In order to study the mechanism of the band-to-band tunneling hot electron injection and the programming currents, the programming characteristics are firstly simulated with the simulator Silvaco. Figure 5 shows the electric field distribution with the control gate voltage equaling to

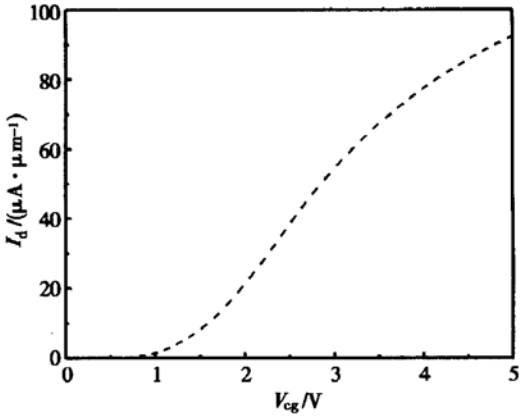

Fig. 4 Read current of the flash cell Read operation is biased at V_d= 1.5V, V_{cg}= 3V, V_s= 0; the read current is 60μA/μm with 0.8μm gate length.

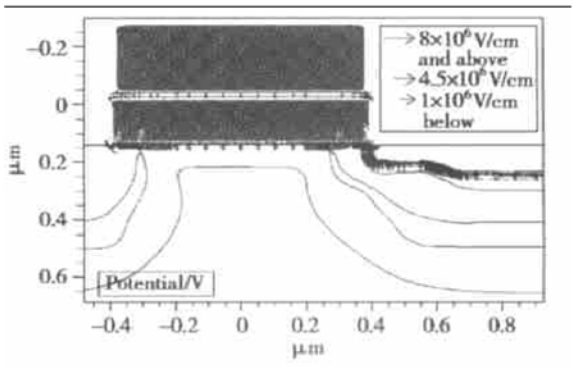

Fig. 5 Electric field distribution with band-to-band hot electron injection programming Program condition: V_{cg}= 8V, V_s= 0V, V_d= −5V

8V, and the drain voltage equaling to −5V. It can be seen that the electric field is highest at the front point of the p/n-halo junction. The maximum lateral and the vertical electric fields are as high as 2.5MV/cm and 4MV/cm, respectively. With these electric field stressing, the band-to-band tunneling is easily induced and the hot electrons can inject to the floating gate. The simulation results also show that the floating gate current is impossibly induced by Fowler-Nordheim tunneling mechanism due to

the graded gate oxide (GGO) phenomenon and the depletion of the surface of the p-drain.

Figure 6 shows the drain and floating gate currents with different floating gate voltages. The floating gate voltages were set to 0, 3, 4 and 5V, respectively, and the drain was stressed from 0V to − 5V. Because the drain junction was negatively biased during programming, the drain leakage current was only about 3μA/μm. However, the floating gate current can be up to 1.2nA when the floating gate and the drain were set to 4V and − 5V, respectively. The efficiency of the proposed program method is 4×10^{-4}, which is two to three orders of magnitude higher than traditional channel hot electron injection programming.

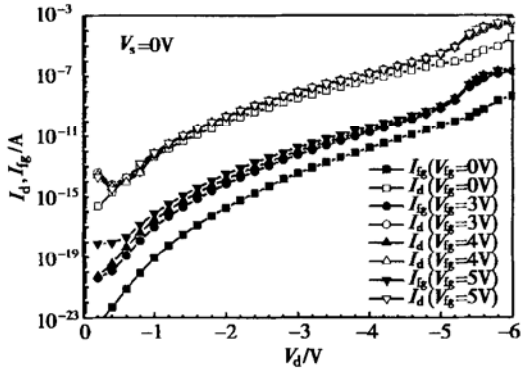

Fig. 6　Simulation results of the programming current. Drain current I_d is the band-to-band tunneling induced leakage current of the negative drain p-n junction. Floating gate current I_{fg} is caused by the band-to-band tunneling induced hot electron injection.

Figure 7 shows the time-dependent programming characteristics of the 1/0.8μm flash memory cell. During programming, the drain is set to − 5V, and the control-gate is set to 8V and 10V, respectively. It can be seen that the programming speed can get up to about 16μs and 12μs with the minimum threshold voltage shift of 4V.

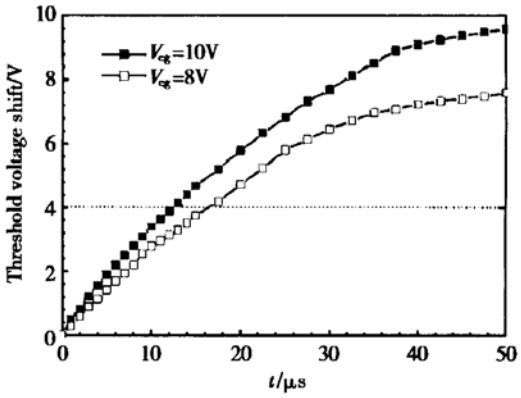

Fig. 7　Time-dependent programming characteristics at V_d equaling to − 5V

The program window is shown in Fig. 8. Here, the write condition is: V_{cg}= 8V, V_s= 0V, V_d= − 5V, t_{write}= 20μs. The erase operation used source side negative gate Fowler-Nordheim tunneling erasure, with which, the source, drain, control gate, and p-substrate voltages are set to 5V, floating, − 8V and 0V, respectively. As a result, the threshold voltages after writing and erasing are 5.7V and 1.0V, and the program window is up to 4.7V.

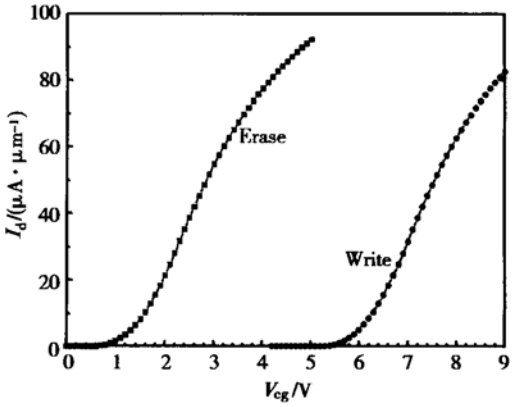

Fig. 8　Program window of the proposed flash cell

4　Conclusion

In this paper we proposed a novel n-channel flash memory cell, which used band-to-band hot electron injection to perform program, and used source side negative gate Fowler-Nordheim tunneling to erase. Because the drain junction is negatively biased when programming, the drain leakage current largely decreases, and the vertical electric field can be reduced to lower than 4MV/cm in the BBHE programming. From the testing result, the proposed cell can be programmed with a low voltage of 8V and the drain leakage current is only about 3μA/μm. Under the proposed operating conditions, the program efficiency and the read current rise up to 4×10^{-4} and 60μA/μm, respectively, and the program time can be as short as 16μs. Owning to the band-to-band program and the n-channel structure, the proposed flash memory cell realizes high programming speed, low programming voltage, low power consumption, large read current and short access time.

Acknowledgement The authors would like to express their thanks to Dr. Charles C. H. Hsu for his continuous encouragement, and thank Mr. Tao Zhong, Ms Yuxiang Fu, Ms Shu Xia, and all other members in the process line of IMETU for their great work on the sample manufacture.

References

[1] Lucero E M, Challa N, Feilds J. A 16K-bits smart 5V-only EEPROM with redundancy. IEEE J Solid-State Circuits, 1983, SC-18(10): 539

[2] Yoshihisa Iwata, Masaki Momodomi, Tomoharu Tanaka, et al. A high-density NAND EEPROM with block-page programming for microcomputer applications. IEEE J Solid-State Circuits, 1990, SC-25(4): 417

[3] Hsu C C H, Acovic A, et al. A high speed, low power p-channel flash EEPROM using silicon rich oxide as tunneling dielectric. In: Ext Abstract of 1992 SSDM, 1992: 140

[4] Ohnakado T, Mitsunaga K, Nunoshita M, et al. Novel electron injection method using band-to-band tunneling induced hot electron (BBHE) for flash memory with a p-channel cell. In: IEDM Tech Dig, 1995: 279

[5] Ohnakado T, Ajika N, Hayashi H, et al. 1.5V operation sector-erasable flash memory with bipolar transistor selected P-channel cells. In: Symp VLSI Technology Dig, 1998: 14

[6] Frank R L Lin, Cheng H P, et al. A novel hot carrier mechanism: band-to-band tunneling hole induced bipolar hot electron (DDHBHE). In: IEDM Tech Dig, 1999: 31

[7] Chen I C, Kaya C, et al. Band-to-band tunneling induced substrate hot-electron (BBISHE) injection: a new programming mechanism for nonvolatile memory devices. In: IEDM Tech Dig, 1989: 263

一种采用带-带隧穿热电子注入编程的新型快闪存贮器*

潘立阳 朱 钧 刘志宏 曾 莹 鲁 勇

(清华大学微电子学研究所, 北京 100084)

摘要: 提出一种采用带-带隧穿热电子注入编程的新型快闪存贮器结构,在便携式低功耗的 code 闪存中有着广泛的应用前景. 该结构采用带-带隧穿热电子注入(BBHE)进行"写"编程,采用源极 Fowler-Nordheim 隧穿机制进行擦除. 研究显示控制栅编程电压为 8V,漏极漏电流只有 3μA/μm 左右,注入系数为 4×10^{-4},编程速度可达 16μs,0.8μm 存贮管的读电流可达 60μA/μm. 该新型结构具有高编程速度、低编程电压、低功耗、大读电流和高访问速度等优点.

关键词: 快闪存贮器; 带-带隧穿; 沟道热电子; Fowler-Nordheim

EEACC: 2560A; 1265D

中图分类号: TP333 **文献标识码**: A **文章编号**: 0253-4177(2002)07-0690-05

* 国家九五计划资助项目(No. 97-760-01-01)

潘立阳 男,1975 年出生,主要从事快闪存贮器及其工艺、可靠性研究.

2001-11-03 收到,2002-01-17 定稿

©2002 中国电子学会