第26卷 第10期 半 导 体 学 报 Vol. 26 No. 10
2005年10月 CHINESE JOURNAL OF SEMICONDUCTORS Oct., 2005

An Improved Charge Pumping Method to Study Distribution of Trapped Charges in SONOS Memory*

Sun Lei, Pang Huiqing, Pan Liyang, and Zhu Jun

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

Abstract: In silicon-oxide-nitride-oxide-silicon (SONOS) memory and other charge trapping memories, the charge distribution after programming operation has great impact on the device's characteristics, such as reading, programming/erasing, and reliability. The lateral distribution of injected charges can be measured precisely using the charge pumping method. To improve the precision of the actual measurement, a combination of a constant low voltage method and a constant high voltage method is introduced during the charge pumping testing of the drain side and the source side, respectively. Finally, the electron distribution after channel hot electron programming in SONOS memory is obtained, which is close to the drain side with a width of about 50nm.

Key words: flash memory; SONOS; charge trapping memory; charge pumping method; charge distribution

EEACC: 2560A; 2530F

CLC number: TP333.5 **Document code:** A **Article ID:** 0253-4177(2005)10-1886-06

1 Introduction

Recently, studies of nonvolatile semiconductor memory have focused on charge trapping memory[1], rather than stack gate flash memory. Because charge trapping memory uses the gate dielectric layer to store the injected charges, it has the advantages of a simple structure, low programming voltage, and high technology compatibility. Silicon-oxide-nitride-oxide-silicon (SONOS), a type of charge trapping memory, has gained more attention due to high reliability and well-developed technology[2].

If SONOS and other charge trapping memories are programmed by channel hot electron (CHE) injection or channel initialed secondary electron (CHISEL) injection, the injected electrons will be trapped in a localized region[3,4], not in the whole floating gate like the stack gate flash memory. This localized charge distribution has been proved to influence the characteristics of the memory, such as reading, programming/erasing, and reliability[5~7]. This makes charge trapping memory quite different from the stack gate flash. Furthermore, the charge distribution will have further impact with the scaling down of the memory device. Therefore, study of the injected charges, especially the charge distribution, becomes more and more important. On the other hand, we can obtain the physical mechanisms of CHE or/and CHISEL injections, by studying the positions and ranges of injected charges. So charge trapping memory is suitable for the study of the programming mechanisms.

Methods to investigate the charge distribution usually include the capacitance-voltage (CV) method, the direct-current current voltage (DCIV) method[8], and the charge pumping method[9,10]. Compared to other methods, the charge pumping

* Project supported by the National Natural Science Foundation of China (No. 60306007)

Sun Lei male, was born in 1980, PhD candidate. His research interests are in devices, technologies, and reliability of nonvolatile semiconductor memory.

Received 19 April 2005, revised manuscript received 14 June 2005

method is more accurate and simple to implement. It can get information about trapped charges and interface states in the gate dielectric by the charge-discharge procedure of the interface states in the channel. The charge pumping method includes the constant low voltage (CLV) method, the constant high voltage (CHV) method, and the constant voltage amplitude method, which will be discussed in the following sections. However, in the actual measurement, the precision still cannot meet the requirements due to some factors, for example the leakage current. In addition, the charge distribution measured by the charge pumping method is usually not exactly the same as the actual one, and some narrow distribution may be neglected as a result. In this paper, we propose an improved charge pumping method to solve these problems and analyze the charge distribution of SONOS memory.

2 Method

In the early years, the charge pumping method was mainly used in the analyses of MOS device reliability. When the device is tested, a series of pulses with certain patterns of amplitude are added to the gate, while the drain, source, and substrate are all grounded. The channel surface then will change from accumulation to inversion or contrariwise. During this procedure, the interface states between gate oxide and silicon act as the recombination centers and contribute to the recombination current (called charge pumping current or I_{cp}), which can be detected through the substrate and presents the information about interface states and oxide charges as well.

SONOS and other charge trapping memories have a similar structure to the MOS device, and are also suitable for the charge pumping measurement. If we have the assumption that the virgin cell has a uniform interface state distribution along the channel, which will not vary a lot during a few cycles of programming/erasing operations, only the charges injected into the silicon nitride layer will influence the local threshold voltage in the channel, and then influence the I_{cp}.

2.1 CLV method: test from the drain and the source

To distinguish the injected charge distribution of the source side and the drain side, we test the charge pumping current from the source and the drain, respectively. As illustrated in Fig. 1(a), the source is floating when tested from the drain side. With a constant low voltage bias of gate pulses (lower than the flat band voltage in the channel), the high voltage bias increases, allowing more and

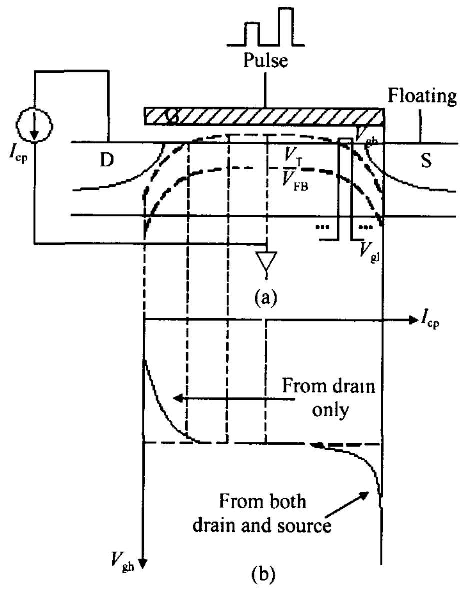

Fig. 1 (a) Schematics of charge pumping testing; (b) I_{cp}-V_{gh} curve of constant low voltage method

more channels to reach the inversion region, and letting more and more interface states contribute to the recombination current. Because of the floating state of the source, only the current near the drain can be detected, so the threshold voltage profile of the drain side and the calculated charge distribution can be obtained by I_{cp}. A typical relationship between charge pumping current and high voltage of gate pulses (I_{cp}-V_{gh}) is shown in Fig. 1(b). After the high voltage exceeds the maximum threshold voltage in the channel, the recombination current near the source side can also be measured due to the turning on of the channel, and I_{cp} will increase dramatically and reach saturation. The same testing and analyses methods are applicable to the source side (refer to Fig. 2), and we can get the charge distribution in the whole channel by combining the results of the drain and the source.

Our SONOS device used for charge pumping testing has the W/L of 20μm/0.8μm, with the effective channel length of about 0.6μm. The bottom oxide thickness of the gate dielectric is 4.5nm, the silicon nitride 8nm, and the top oxide 6nm. Before programming, the charge pumping current was measured in the virgin cell, and the I_{cp} curves tested from the drain and the source should be the same. The cell is then programmed using the CHE injection (V_d = 6.5V, V_g = 8V), and I_{cp} is measured again. The injected electrons will elevate the local threshold voltage, forming a V_T profile similar to the one in Fig. 2. Because of the asymmetry of the drain side and the source, the I_{cp} curves differ as well. Figure 3 shows the actual testing results. The gate pulses have a low voltage of -4V, with a high voltage sweeping from -1 to 4V, and a frequency of 5MHz.

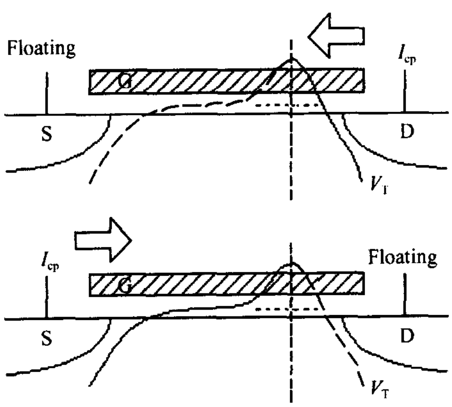

Fig. 2 Threshold voltage profile along the channel of a programmed SONOS The charge pumping currents are measured from the drain and the source independently.

During the CLV method of the charge pumping measurement, the higher V_{gh} is, the more current can be detected in the channel. However, the injected charges have very narrow distribution, usually below 100nm, so when testing from the drain side in the high gate voltage condition, the recombination current of the source side can flow through the channel into the drain, which makes I_{cp} increase rapidly and unite with the source current (refer to the overlap region of the drain current and the source current in Fig. 3). Consequently, the information about the narrow peak of injected charges will be reflected in the large I_{cp} region. The narrow peak may be influenced because of the saturating of I_{cp}, which results in a large error when profiling the narrow peak.

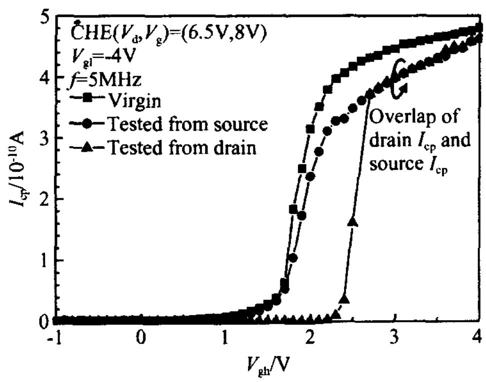

Fig. 3 I_{cp} tested by the constant low voltage method

2.2 CHV method

To solve the problem of narrow peak profiling, the CHV method is proposed, which has a fixed high gate pulses voltage (higher than the maximum threshold voltage in the channel) and an increasing low voltage. When the low voltage is higher than the flat band voltage of the entire channel, none of the interface states can contribute to the charge pumping current, so I_{cp} equals to zero. If some electrons are injected in a certain region of the channel, the local flat band voltage will increase, adding some additional I_{cp}. Different from the CLV method, the change of I_{cp} induced by injected charges is reflected in the small current region, so this small current region is also sensitive to the small number of injected charges and the measurement results are more accurate.

On the other hand, because the gate pulses high voltage is set higher than the channel threshold voltage throughout the CHV testing, the channel is always in the "on" state. This means currents tested from the drain and the source are the same, and we cannot get the position and shape of injected charges in the channel, apart from the width and height of the distribution. Figure 4 shows the test results of CHV method. The fixed high voltage is 4V, and the low voltage sweeps from -2 to 3V, and the frequency is 5MHz. From this figure we can see the curves of the drain side and the source are completely the same, and the electron injection can be observed obviously in the small current region under logarithmic coordinates.

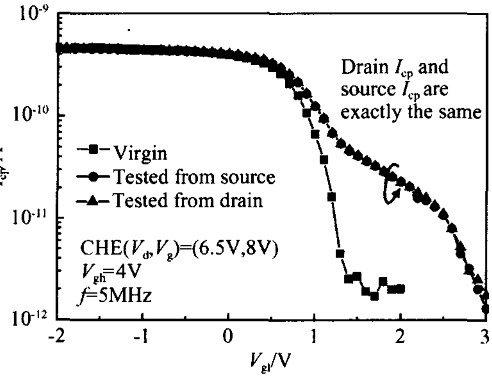

Fig. 4 I_{cp} tested by the constant high voltage method

These two kinds of testing methods have their own advantages and disadvantages. We can combine them in the actual testing: obtaining the distribution's width and height by the CHV method, and the location in the channel by the CLV method. By this means, we could get a precise distribution of injected charges.

3 Results and discussion

From I_{cp} measured by the constant low voltage method, the relation between injected charges density and channel length can be calculated using Eqs. (1) similar to[9]:

$$N_{ONO}(x) = \frac{C_{ONO}}{q}\Delta V_{gh},\ x = \frac{I_{cp}(V_{gh})}{I_{cpmax}}L \qquad (1)$$

where N_{ONO} is the electron density injected into ONO, and C_{ONO} is the capacitance of ONO, L is the channel length, and x refers to the position along the channel, where $x = 0$ refers to the source and $x = L$, the drain. ΔV_{gh} means, under a certain charge pumping current, the V_{gh} shift between programmed cell and virgin cell. $I_{cp}(V_{gh})$ is the charge pumping current under a certain V_{gh}. I_{cpmax} refers to the maximum of I_{cp}, which all the interface states in the channel contribute to.

The calculated results are illustrated in Fig. 5 including the data tested from the source side and the drain side. As mentioned above, the CLV method is not precise enough in the saturation current region, so the region where I_{cp} tested from the drain and the source are equal (labeled in Fig. 5) needs to be corrected by the CHV method. The distribution width W_{ie} and peak value $N_{ONOpeak}$ of injected electrons can be obtained by Eqs. (2) and (3):

$$W_{ie} = \frac{I_{cp}(V_{gl})}{I_{cpmax}}L \qquad (2)$$

$$N_{ONOpeak} = \frac{C_{ONO}}{q}\Delta V_{glmax} \qquad (3)$$

where $I_{cp}(V_{gl})$ is the charge pumping current under a certain V_{gl}, ΔV_{glmax} means the maximum shift of V_{gl} between programmed cell and virgin cell. After the correction, we finally get the injected electron distribution, as shown in Fig. 6.

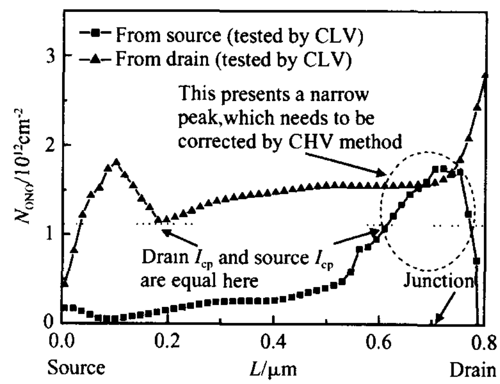

Fig. 5 Charge distribution calculated from the drain and the source measurement data by the CLV method

The results show that after CHE programming, the injected electrons mainly distribute near the drain side of SONOS memory. Furthermore, there are two peaks with a width of about 50nm. One peak is at the edge of the gate, and the other is

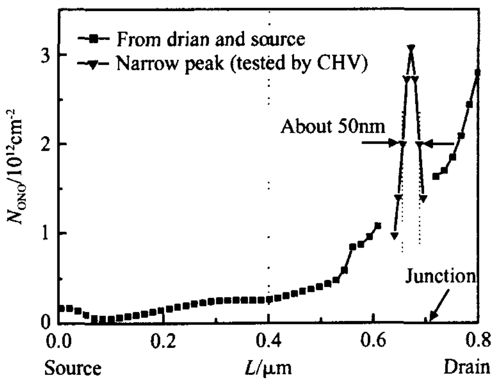

Fig. 6 Charge distribution after CHE programming

inside the channel near the drain junction. These two peaks are calculated by the CLV method and the CHV one, respectively. The information of the second peak can be measured precisely due to the sensitivity of the CHV method to the narrow charge distribution, as discussed before.

Under different drain and source structures, such as the LDD structure, injected charges will have different distributions. We can use this charge pumping method to evaluate the distribution character and contribute to junction engineering optimization. The charge distribution obtained using this charge pumping method will also help us study the programming mechanisms, optimize the programming conditions, and analyze and improve the reliability. Furthermore, when scaling down below 0.18μm, the injected charges near the drain junction will have influence on the source, especially in the two-bit application. So studies on the charge distribution will become more important in the small size devices.

4 Conclusion

Using the charge pumping method, the charge distribution of programmed SONOS memory and other charge trapping memories is measured and calculated. Because of its charge trapping character, SONOS memory is suitable for the study of CHE, CHISEL, and other charge injection mechanisms. To improve the test precision, the constant low voltage method and constant high voltage method are combined, and the current is measured from drain and source separately. Finally, by calculating we get the precise charge distribution.

References

[1] White M H, Adams D A, Bu J K. On the go with SONOS. IEEE Circuits and Devices, 2000: 22

[2] Cho M K, Kim D M. High performance SONOS memory cells free of drain turn-on and over-erase: compatibility issue with current flash technology. IEEE Electron Device Lett, 2000, 21(8): 399

[3] Lusky E, Yosi S D, Bloom I, et al. Characterization of channel hot electron injection by the subthreshold slope of NROM (tm) device. IEEE Electron Device Lett, 2001, 22(11): 556

[4] Mahapatra S, Shukuri S, Bude J. CHISEL flash EEPROM-part Ⅰ: performance and scaling. IEEE Trans Electron Devices, 2000, 49(7): 1296

[5] Chang Y W, Lu T C, Pan S, et al. Modeling for the 2nd-bit effect of a nitride-based trapping storage flash EEPROM cell under two-bit operation. IEEE Electron Device Lett, 2004, 25(2): 95

[6] Larcher L, Verzellesi G, Pavan P, et al. Impact of programming charge distribution on threshold voltage and subthreshold slope of NROM memory cells. IEEE Trans Electron Devices, 2002, 49(11): 1939

[7] Sun L, Pan L Y, Zeng Y, et al. Distribution and impact of local trapped charges in SONOS memory. International Conference on Solid State Devices and Materials, Tokyo, Japan, 2004: 650

[8] Neugroschel A, Sah C T, Han K M, et al. Direct-current measurements of oxide and interface traps on oxidized silicon. IEEE Trans Electron Devices, 1995, 42(9): 1657

[9] Chen C, Liu Z Z, Ma T P. Analysis of enhanced hot-carrier effects in scaled flash memory devices. IEEE Trans Electron Devices, 1998, 45(7): 1524

[10] Su Y, Zhu J, Chen Y C, et al. Charge pumping measurement for determining band-to-band-tunneling induced interface damage during erasing operation of flash. Chinese Journal of Semiconductors, 2001, 22(1): 69

利用改进的电荷泵法研究 SONOS 存储器陷阱电荷的分布特性*

孙　磊　庞惠卿　潘立阳　朱　钧

(清华大学微电子学研究所，北京　100084)

摘要：在 silicon-oxide-nitride-oxide-silicon(SONOS)等电荷俘获型不挥发存储器中，编程操作后注入电荷的分布会对器件的读取、擦写以及可靠性带来影响. 利用电荷泵方法可以有效而准确地测量出注入电荷沿沟道方向的分布. 为了提高测试精度，在进行电荷泵测试时，采用固定低电平与固定高电平相结合的方法，分别对 SONOS 器件源端和漏端进行注入电荷分布的测试. 通过测试，最终获得 SONOS 存储器在沟道热电子注入编程后的电子分布. 电子分布的峰值区域在漏端附近，分布宽度在 50nm 左右.

关键词：flash 存储器；SONOS；电荷俘获型存储器；电荷泵法；电荷分布

EEACC：2560A；2530F

中图分类号：TP333.5　　**文献标识码**：A　　**文章编号**：0253-4177(2005)10-1886-06

* 国家自然科学基金资助项目(批准号:60306007)

孙　磊　男，1980 年出生，博士研究生，主要从事不挥发存储器工艺、器件以及可靠性的研究.

2005-04-19 收到，2005-06-14 定稿