Investigation of Simulated and Measured Program Characteristics in 4-bit/cell Charge-trap Flash (CTF) Memories

Jae Moo Kim, Yu Jeong Seo, Ho-Myoung An and Tae Geun Kim*

School of Electrical Engineering, Korea University, Seoul 136-713

Sung Wook Park and Dae Hwan Kim

School of Electrical Engineering, Kookmin University, Seoul 136-702

(Received 26 August 2008)

We investigate the simulated and measured program characteristics for 4-bit program operations in silicon-oxide-nitride-oxide-silicon (SONOS) devices by using two-dimensional (2-D) device simulations. For this calculation, the width of the region with charge trapped locally in the drain region is assumed to be as narrow as 44 nm to remove second-bit effects during the 2-bit and 4-bit operation. We determine the reverse read voltage for screening bit-1 to be 2.5 V and confirm that both 2-bit and 4-bit characteristics are successfully observed in our devices. From the threshold voltage shift, the densities of trapped charge are estimated to be \(0 \times 10^{19}\), \(3 \times 10^{19}\), \(6 \times 10^{19}\) and \(9 \times 10^{19}\) cm\(^{-3}\) at 4-level states, respectively. We also find that these simulation results are reasonably consistent with the experimental results achieved in this work.

PACS numbers: 85.30.De, 73.40.Qv

Keywords: 4-bit-per-cell, SONOS, NROM\(^{	ext{TM}}\), Localized trapped charge

I. INTRODUCTION

In recent years, the great demand for high-density flash memories has increased for solid-state mass storage applications such as mobile telephones and portable players. The charge-trap flash (CTF) based on a polysilicon-oxide-nitride-oxide-silicon (SONOS) structure has been considered a promising candidate for future flash memories beyond the floating-gate technology because of its simple cell structure and process, its low power operation, and its high density [1–4]. Moreover, as scale-down of the device size is faced with its physical limits, the 2-bit operation has received much attention because of its advantages of increasing the density [5,6]. Therefore, the 2-bit/cell capability of the SONOS structure is very attractive for high density memories. Generally, programming is performed by channel hot electron (CHE) injection above one of the junctions, and erasing is accomplished by injecting hot holes generated by band-to-band tunneling (BTBT) at the same junction. Reading is performed in the “reverse” direction by changing the source and the drain. The localization of the trapped charge in the nitride layer makes it possible to separate two physical bits because one bit does not affect the other bit due to the narrow storage region. Recently, this 2-bit/cell SONOS technology has evolved into 4-bit/cell SONOS devices, which combine physically localized storage and a multi-level cell [7]. The inherent advantage of the 4-bit SONOS is that only 4-threshold voltage (\(V_{th}\)) levels are required in each storage area. Compared to traditional multi-level cell (MLC) floating gate (FG) flash memories, a 4-bit/cell SONOS memory has several distinct advantages. First, the injected charges in a SONOS cell are trapped in a non-conducting nitride medium; therefore, the trapped charges are much less sensitive to the leakage than those in a conducting metal medium in a FG flash memory cell. Second, the geometrical symmetry for 2-bit storage in a SONOS cell allows only 4 different charge states in each location to achieve 4-bit-per-cell operation, compared to the required 16 different \(V_{th}\) levels within a single location in traditional MLC FG flash cell. Third, additional processes for 4-bit/cell operation are not required because the structure for 4-bit/cell operation is identical to that of a 2-bit SONOS cell. Lastly, 4-bit/cell SONOS technology has a potential for greater reduction to smaller process nodes than the conventional FG technology, which is faced with significant challenges for scaling down below 40 nm. However, in spite of the many advantage of 4-bit/cell SONOS technology, there are few reports on the spatial width and trap density for 4-bit program operation. In particular, the location and density of trapped charges in the nitride layer is decisive in achieving reliable 4-bit operation.

In this paper, we analyze the 4-bit operation of locally trapped charges in the 4-bit/cell SONOS device.

*E-mail: tgkim1@korea.ac.kr; Fax: +82-2-924-5119

-367-
Fig. 1. Cross-sectional view of the SONOS structure: (a) TEM image and (b) TCAD structure.

Particularly, the influences of the width of the region with locally trapped charges and of the density of locally trapped charges in the nitride region are investigated and explained through a two-dimensional (2-D) device simulator, Sentaurus\textsuperscript{TM}. In this simulation, the width of the region of locally trapped charge is investigated to verify 2-bit/4-bit operation without the second bit effect. Also, the dependences of the optimum reverse read voltage and trapped charge density on the 4-level $V_{th}$ are measured and simulated for reliable 4-bit operation.

### II. EXPERIMENTAL DETAILS

A TEM image of the fabricated device and its structure designed by using technology computer-aided design (TCAD) are shown in Fig 1. It is based on 0.35-um-gate-length n-channel metal-oxide-semiconductor (NMOS) device. The gate stack consists of 4 layers making up with a 34-Å oxide layer, a 74-Å silicon nitride ($Si_3N_4$) layer as a charge storage layer, and a 34-Å oxide layer between the channel and the poly (4.7 eV) gate. The $p$-type substrate doping concentration is $1 \times 10^{17} \text{ cm}^{-3}$.

Parameters used in the simulation were taken from the literature. In this work, a Shockley-Read-Hall (SRH) model is used for trapping and recombination of the traps, and a hydrodynamic transport model is used for carrier transport [8]. The SRH model has the following form:

$$R_{SRH}^{net} = \frac{n_p - n_{eff}^2}{\tau_p (n + n_1) + \tau_n (p + p_1)}$$

with

$$n_1 = n_{eff} \exp \left( \frac{E_{trap}}{kT} \right)$$

and

$$p_1 = n_{eff} \exp \left( -\frac{E_{trap}}{kT} \right)$$

where $E_{trap}$ is the difference between the defect level and the intrinsic level. The silicon default value is $E_{trap} = 0$.

The minority lifetimes $\tau_n$ and $\tau_p$ are modeled as

$$\tau_c = \tau_{dop} \frac{f(T)}{1 + g_c(F)} \quad (c = n, p) \quad (4)$$

The $\tau_{dop}$ is the doping dependence of the SRH lifetime, and $g(F)$ is the field dependence. They are modeled as

$$\tau_{dop} (N_{A,0} + N_{D,0}) = \tau_{min} + \frac{\tau_{max} - \tau_{min}}{1 + \left( \frac{N_{A,0} + N_{D,0}}{N_{ref}} \right)^\gamma} \quad (5)$$

and

$$g(F) = \int_0^F \exp \left[ u - \frac{2 \sqrt{u^3}}{3} E \right] \, du \quad (6)$$

We assumed the temperature to be 300 K; then, the temperature dependence was $f(T) = 1$. To define the trapped charge in the programming state and perform 4-bit operation, we laterally separate the nitride layer as charge packets, and the respective length of a packet is 44 nm because the hot carrier injection region was reported to be 40 – 50 nm in previous research [9,10]. The locally trapped charge above the drain junction has a negligible effect on the threshold voltage, but does reduce the slope. Therefore, it was neglected because the shift of threshold voltage was the main consideration of this paper. In addition, only the bulk electrons trapped in the nitride were considered, and the trap energy level was neglected because $\Delta V_{th}$ was due to the cumulative number of charges.

### III. RESULT AND DISCUSSION

In order to confirm the 2-bit storage operation, Fig. 2 shows the reverse/forward (RVS/FWD) read $V_{th}$ with respect to drain voltages after a drain-side channel hot
electron (CHE) injection program. The drain side bit is programmed to $\Delta V_{th} = 0.74$ V under $V_d = 2.5$ V RVS read conditions. In this figure, the threshold voltage of the programmed bit is shown in the RVS/FWD read operations as a function of the drain voltage ($V_d$). Under RVS read operation, $\Delta V_{th}$ decreases as a function of the $V_d$ level due to the localized drain-induced barrier lowering (DIBL) effect [9]. For $V_d > 2.5$ V, the barrier lowering effect saturates, and the threshold voltage becomes constant. Interestingly, as is clearly visible, an excellent agreement of $V_{th}$ difference between RVS and FWD read operations is achieved, between measurements and simulations, for a 44 nm width of injected region.

At the RVS read voltage determined in Fig. 2, the $I_d - V_g$ curve of RVS/FWD read operations was measured and simulated and was compared to the conventional read voltage of 0.1 V after programming the bit-1 (drain side), as shown in Fig. 3. In the measurement, the 2-bit device was programmed by using CHE injection into the silicon nitride ($Si_3N_4$) near the drain junction edge. In a TCAD simulation, the drain side charge packet is filled with a charge density $3 \times 10^{19}$ cm$^{-3}$, which is well matched with the measurements. Here, $V_{th}$ is defined as the applied gate voltage at which the drain current is 1 $\mu$A. When the drain voltage has a small value ($V_d = 0.1$ V), there is little change in the threshold voltage between the RVS and the FWD read operations. However, if the drain voltage has a sufficiently large value (2.5 V), there is a significant change in the threshold voltage of about 0.7 V. This shows that the difference between the RVS and the FWD read operations has a strong dependence on drain voltage.

To explain this observation in Fig. 3, we show the effect of the expansion of the depletion region in bit-1 programmed cell at drain voltages of 0.1 and 2.5 V in Fig. 4. Fig. 4(a) and (b) show the depletion region of the simulation structure at $V_d = 0.1$ and $V_d = 2.5$ V, respectively. When $V_d = 0.1$ V (sufficiently small value), as shown in Fig 4(a), there is little expansion of the depletion region. The charge packet is still in the channel.
region and affects both the RVS and the FWD read operations. Therefore, both operations of FWD/RVS read have almost the same $V_{th}$. As the $V_d$ increases, however, the depletion region expands because the reverse bias is increased and the effective channel length gets shorter. Eventually, the depletion region fully overlaps the whole charge packet region. Then, charge in the packet loses its influence on the $V_{th}$ shift because only electrons above the channel region increase the threshold voltage. Hence, the FWD read operation at value of the drain voltage under $V_d = 2.5$ V has little $V_{th}$ shift, but in the RVS read operation, bias is applied to the source side, and extension of depletion region hardly affects the drain side channel. Thus, the $V_{th}$ shift induced by trapped charge is similar to that of the $V_d = 0.1$ V state. Therefore, the effect of the depletion region mainly appears in FWD read operation. This obvious difference between the FWD and the RVS read operations at a certain point makes it possible to have a distinct programmed bit.

As the number of electrons in bit-1 increases, the programming effect on bit-1 cannot be fully screened out in the read operation. Therefore, the programmed charge in bit-1 affects on the $V_{th}$ shift of bit-2, which is called a second bit effect [7]. In order to achieve reliable 4-bit/cell operation in a SONOS structure, a prerequisite would be a 2-bit program operation without the second-bit effect. Fig. 5 shows the 2-bit program characteristics, which are the experimental and simulated result for the FWD and the RVS $V_{th}$ as functions of the time and the injected charge density, respectively. Bit-1 is programmed to high $V_{th}$ while bit-2 (source side) is in the erased state of 3 V and is sequentially programmed. Although the second-bit effect appears at a programming time of 50 ms, or longer, it seems to be no problem for 2-bit program operations because the $V_{th}$ difference ($\Delta V_{th}$) between bit-2 RVS and bit-2 FWD read is still as high as 0.74 V. Also, the simulated results for the RVS/FWD read $V_{th}$ as functions of the injected charge density from 0 to $3 \times 10^{-19}$ match the measured results well.

In order to perform the 4-bit/cell operation from the 2-bit result shown in Fig. 5, we investigated the program characteristics of MLC operations, as shown in Fig 6. The measured $I_d - V_g$ curves are programmed for various gate biases (4 – 6 V) at a fixed drain bias of 4 V for 10 ms. The threshold voltages are 3, 3.7, 4.4 and 5.1 V, and their margins are large enough to separate “00”, “01”, “10” and “11” states. At a gate voltage of 7 V, the $V_{th}$ shift is significantly reduced. This result shows that program characteristics can be readily controlled for 4-bit/cell operations in a SONOS structure. Also, from the experimental results, the trapped charge density at each condition can be obtained by using TCAD simulation. The simulation results are agree with the measured ones when the charge densities are $0 \times 10^{19}$, $3 \times 10^{19}$, $6 \times 10^{19}$ and $9 \times 10^{19}$ cm$^{-3}$. Thus, the charge density increases by about $3 \times 10^{19}$ for a $V_{th}$ shift of 0.7 V, and the trap density in the nitride layer is estimated to be about $9 \times 10^{19}$ cm$^{-3}$ because $V_{th}$ hardly increased after that point.

IV. CONCLUSION

The width of the area and the density of locally trapped charges for 4-bit program operations in SONOS devices are estimated by using 2-D device simulators. From the experimental results, the width of the region of locally trapped charge in the simulation is assumed to be 44 nm to verify 2-bit/4-bit operation without the second-bit effect, and the RVS read voltage for screening of bit-1 is determined to be 2.5 V. For the 4-bit program operation having a sufficient margin, we set the gap of the $V_{th}$ shift as 0.7 V between each state. In the simulation, the trapped charge densities of the 4-level states in
the drain region were estimated to be $0 \times 10^{19}$, $3 \times 10^{19}$, $6 \times 10^{19}$ and $9 \times 10^{19}$ cm$^{-3}$, respectively, by matching the experiment results. This information enables us to optimize the 4-bit program operation condition in the 4-bit/cell SONOS flash memory.

**ACKNOWLEDGMENTS**

This work was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korea government (MEST) (Quantum Photonic Science Research Center).

**REFERENCES**


