Low-Cost and Scalable Embedded Non-Volatile Memory Using Quasi-Planar Bulk (Bulk+) Transistor with Standard CMOS Gate Stacks

Kensuke Ota, Masumi Saitoh, Chika Tanaka, Daisuke Matsushita, and Toshinori Numata
Advanced LSI Technology Laboratory, Corporate R&D Center, Toshiba Corporation, 1 Komukai-Toshiba-cho, Saiwai-ku, Kawasaki 212-8582, Japan, Phone: +81-44-549-2192 E-mail: kensuke.ota@toshiba.co.jp

Abstract
We propose low-cost and highly-scalable non-volatile memory using quasi-planar bulk transistor (Bulk+ Tr.) fabricated with standard CMOS process. Thanks to the electric field concentration at the channel corners in Bulk+ Tr., the efficiency of both programming and erase (charge trapping/detrapping in the gate oxide) is enhanced as compared to conventional bulk planar Tr. Scaling of gate length and channel width further improves the threshold voltage window up to more than 0.3V. Moreover, 300-times P/E cycles and 10-year data retention were experimentally demonstrated. Bulk+ Tr. memory is highly suitable for non-volatile memory embedded with low-power scaled CMOS.

1. Introduction
Recently, a low cost embedded non-volatile memory has attracted much attention for the code storage in SoC and microcontroller unit. Very recently, embedded memory operation with standard CMOS gate stacks was reported [1], whereas programming and erase characteristics of Bulk+ Tr. was fabricated on bulk-Si wafer with standard CMOS [2]. Bulk+ Tr. can be fabricated with the standard CMOS process so that the fabrication process of Bulk+ Tr. is more compatible to bulk planar Tr. and less expensive than that of FinFET or nanowire FET [3,4]. In this paper, we propose and demonstrate excellent non-volatile memory operation using Bulk+ Tr. Reliabilities for memory operations were also examined.

2. Program/erase operation using Bulk+ Tr.
Bulk+ Tr. was fabricated on bulk-Si wafer with standard CMOS gate stacks, that is, poly-Si gate and thermal SiO2 gate oxide. As shown in Fig. 1(a), Bulk+ Tr. consists of pseudo tri-gate structure. Therefore, strong gate controllability leads to high immunity to short-channel effects. Table 1 shows the comparisons among bulk planar Tr., nanowire Tr., and Bulk+ Tr. So far, better subthreshold slope, larger \( L_{th} \) and lower static noise margin in SRAM of Bulk+ Tr. at low supply voltage than bulk planar Tr. have been expected [2]. In the process of fabricating STI, additional wet etching of STI was introduced in order to construct the pseudo tri-gate structure. Figs. 1(b) and 1(c) show the cross-sectional TEM images of Bulk+ Tr. Figs. 2(a) and 2(b) show the typical voltage conditions for P/E operations in this study. Hot electron injection into the gate oxide at drain edge was utilized as program operation. By applying high \( V_{dd} \) and \( V_{gd} \) such as \( V_{dd} = 3.5 \text{V}, V_{gd} = 3.5 \text{V} \), positive \( V_{th} \) shift (\( \Delta V_{th} \)) was induced. In erase operation, negative \( V_{dd} \) and \( V_{gd} \) were applied. It has been already reported that drain-side forward-bias assist enabled erase operation [1]. Erase operation was not realized only by negative \( V_{dd} \). Forward-bias diode current at drain edge induced by negative \( V_{gd} \) might cause the electron ejection and/or hot hole injection, resulting in the negative \( V_{th} \) shift. Fig. 2(c) shows the \( L_{th} \) and \( V_{th} \) characteristics of program and erase states after 3, 10, and 30 P/E cycles. More than 0.3V \( V_{th} \) window was repeatedly obtained.

3. Size dependence of P/E characteristics
Fig. 3(a) shows the programming time dependence of \( \Delta V_{th} \) with different \( W \) under the same program voltage conditions. As \( W \) decreases, \( \Delta V_{th} \) increases. It was reported that the electric field at the channel corner in nanowire Tr. was enhanced and that the contribution of the channel corner increases as \( W \) decreases [5]. Since Bulk+ Tr. also have the channel corner in the pseudo tri-gate structure, Bulk+ Tr. with narrower \( W \) showed larger \( \Delta V_{th} \) leading to highly-efficient program operation. The size dependence of erase operation was also examined, as shown in Fig. 3(b). \( \Delta V_{th} \) in Bulk+ Tr. with narrow channel width was also larger than that with wide channel width indicating that channel corner effect also enhanced the erase operation caused by the electron ejection and/or hot hole injection. \( L_{th} \) dependences of \( \Delta V_{th} \) in program operation were investigated as shown in Fig. 4(a). As \( L_{th} \) decreases, \( \Delta V_{th} \) increases under the same program voltage conditions. In short \( L_{th} \), larger numbers of hot electrons were injected into the gate oxide due to the higher electric field in the lateral direction. Besides, short \( L_{th} \) achieved larger \( \Delta V_{th} \) in erase operation, as shown in Fig. 4(b). In erase operation, not only vertical electric field from gate voltage, but also drain-side forward bias was required [1]. As decreasing \( L_{th} \), lateral electric field was enhanced. Therefore, it is considered that forward bias current at the drain edge just below the gate oxide increased, resulting in the better erase operation. As summarized in Fig 5, narrower and shorter Bulk+ Tr. was found to be better for memory operation. Considering that the bulk planar Tr. does not have channel corner, Bulk+ Tr. shows better program and operation than bulk planar Tr.

4. Reliability for memory operation
Furthermore, we examined the reliability of Bulk+ Tr. memory. Figs. 6 and 7 show P/E cycling characteristics of \( V_{th} \) and \( I_{on}/I_{off} \), respectively. Here, we defined \( I_{on} \) and \( I_{off} \) as the drain current at \( V_{gd} = -0.2 \text{V} \) and \( 0.8 \text{V} \), respectively. \( V_{th} \) window larger than 0.3V and \( I_{on}/I_{off} \) ratio of 4 decades were achieved after 300 cycles. Subthreshold Slope (SS) was not degraded after 300 cycles, as shown in Fig. 8. These results suggest that the interface quality was maintained during P/E cycles, although charge trapping and detrapping in gate oxide were repeatedly induced by P/E. Fig. 9 shows the room temperature data retention in program and erase states. \( V_{th} \) margin as large as 0.3V was calculated after 10 years extracted by the linear fitting. Although the injected electrons were closer to the channel placed in the thin gate oxide, compared to the floating gate type memories, measured data retention was good enough for non-volatile memory operation, suggesting that the electrons were trapped in deep energy states of the gate oxide. These results suggest that proposed Bulk+ Tr. memory have enough reliability for non-volatile embedded memory operation.

5. Conclusion
We demonstrated the embedded memory operation in quasi-planar bulk Tr. with standard CMOS gate stacks. Thanks to the corner effect, quasi-planar bulk Tr. is more suitable for P/E operation than planar bulk Tr. Moreover, gate length and NW width scaling enhanced the P/E efficiency. P/E cycling and data retention showed enough reliability for memory operations. Taking into account the compatibility with CMOS process, proposed quasi-planar...
bulk Tr. is suitable for non-volatile memory embedded with low-power CMOS.

Acknowledgment

This work was partly supported by NEDO’s Development of Nanoelectronic Device Technology.

References


Table 1 Comparisons among bulk planar Tr., nanowire Tr., and Bulk+ Tr.

<table>
<thead>
<tr>
<th>Wafer</th>
<th>Bulk-Si</th>
<th>SOI</th>
<th>Bulk-Si</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gate structure</td>
<td>Planar</td>
<td>Tri-gate or Gate all around</td>
<td>Pseudo tri-gate</td>
</tr>
<tr>
<td>Typical Size</td>
<td>W: 200nm=1μm</td>
<td>No height (planar)</td>
<td>W: 10nm, H: 10nm, W:40nm=100nm, H: 10nm</td>
</tr>
</tbody>
</table>

Fig.1 (a) Schematic of Bulk+ Tr. Cross-sectional TEM image along the (b) NW width and (c) gate length direction.

Fig.2 I-V characteristics of P/E states after 3, 10, and 30 P/E cycles. $V_{th}$ window of more than 0.3V was repeatedly obtained.

Fig.3 Time dependence of $\Delta V_{th}$ with different NW width in (a) program and (b) erase operations. Program voltage was set to $V_g=V_d=3.5V$, while erase voltage was set to $V_g=V_d=-3.5V$.

Fig.4 Time dependence of $\Delta V_{th}$ with different gate length in (a) program and (b) erase operations. Voltage conditions were the same as Fig.3.

Fig.5 $L_g$ and $W$ scaling effects on P/E operations. Scaling of $L_g$ and $W$ improved both P/E efficiency, which leads to large $V_{th}$ window.

Fig.6 P/E cycle dependence of $V_{th}$. $V_{th}$ window more than 0.3V was obtained after 300 cycles.

Fig.7 P/E cycle dependence of $I_d/V_g$. $I_d/V_g$ ratio of 4 decades was maintained after 300 cycles.

Fig.8 Cycle characteristics of subthreshold slope after each operation.

Fig.9 Room temperature data retention after P/E cycles.