A Highly Reliable Logic NVM “eCFlash (embedded CMOS Flash)”
Utilizing Differential Sense-Latch Cell with Charge-Trapping Storage

T. Ogura, M. Mihara, Y. Kawajiri, K. Kobayashi, S. Shimizu, S. Shukuri, N. Ajika and M. Nakashima
GENUSION, Inc., 7-1-3 Douicho, Amagasaki, Hyogo, Japan
Phone: +81-6-6416-6133, Fax: +81-6-6416-6134, e-mail: ogura.taku@genusion.co.jp

Abstract
A new Logic NVM “eCFlash (embedded CMOS Flash)” has been developed without any additional process steps in a 0.25um technology. In this architecture, a novel differential sense-latch cell with charge-trapping storage is adopted. This unique cell structure functions as a differential sense amplifier as well as a data latch, therefore mass data can be restored to each cell’s latch simultaneously for static data output with high sensitivity and low power consumption. Furthermore the handling scheme of twice the breakdown voltage with novel charge pump circuit and high voltage driver is demonstrated. By utilizing these techniques, a highly reliable Logic NVM is realized; 100K cycling endurance and data retention of 20years at 150C.

Introduction
In recent years, many ASICs require a low-cost, low-density (~1KB) non-volatile memory for analog trimming, redundancy data, ID code, parameter setting and so on. Conventional embedded non-volatile memory devices need additional layers or process steps. These approaches increase manufacturing costs and reduce wafer yield due to process complexity. To solve these problems, several Logic NVMs manufactured in a standard CMOS logic process have been developed [1]-[5]. Many of them utilize a floating poly as a storage node, so retention performance depends on the quality of gate oxide for I/O device as tunnel oxide. If the gate oxide is scaled down to 3-4nm for 1.8V I/O device, it could become more difficult to keep good data retention performance. In this paper, to investigate a technical possibility of another candidate, the charge-trapping device utilizing a side-spacer as a storage node instead of floating poly is demonstrated.

Architecture
Fig. 1 is a microphotograph of 2KB eCFlash test chip fabricated in TSMC 0.25um technology. The eCFlash architecture and operation conditions are illustrated in Fig. 2 and Fig. 3 respectively. The memory cell consists of NMOS transistors as NVM elements (MT, MB), PMOS cross-coupled transistors (P1, P2), PMOS pre-charge transistors (P3, P4), NMOS access transistors (N1, N2) and output inverters (INV1, INV2). Each memory cell has a flip-flop circuit composed of MT, MB, P1 and P2. This circuit functions as a differential sense amplifier as well as a data latch, which is defined as “differential sense-latch cell” structure.

Program operation is executed by Channel Hot Electron (CHE) injection to the side-spacer in SL-side of either MT or MB according to the programming data. Erase operation is executed by Avalanche Hot Hole (AHH) injection to the electron trapped area of each MT and MB at once. Electrons and holes trapped in the side-spacer shift the Vth of NVM element. Restore operation is executed at reverse read mode. The restored data is determined according to Vth difference between MT and MB. The output of INV1 indicates the data of memory cell (data"0": Vth_MT > Vth_MB, data"1": Vth_MT < Vth_MB).

Figure 1. Microphotograph of 2KB eCFlash test chip

Figure2. eCFlash architecture

Figure3. Operation conditions
Peripheral circuits

In the eCFlash architecture, WL, BLT and BLB of memory cell are controlled by logic signal level in each operation, so row and column decoder can utilize standard CMOS logic transistors. On the other hand, the high voltage (6V for programming, 9V for erasing in this 0.25um technology) is applied to SL of memory cell. The breakdown voltage of I/O device is not sufficient to handle 9V in erase operation, therefore novel charge pump circuit and high voltage driver which can handle twice the breakdown voltage are indispensable for eCFlash architecture.

The charge pump circuit is Dickson type [6] and consists of poly-diode [7] and cascade-capacitor as shown in Fig. 4. Poly-diode is composed of silicided P+ / N+ poly and non-doped poly. The length of non-doped poly between P+ poly and N+ poly is optimized to keep sufficient electrical diode characteristics (Ion / Ioff > 1E10, BVj > 16V). These characteristics are independent of gate oxide scaling. The cascade-capacitor has conventional MOS capacitors in series connection, so the gate electric field of each MOS capacitor can be reduced to a half of maximum voltage. For further scaling of gate oxide, the number of capacitors in series connection can be increased to satisfy the gate oxide reliability. The cascade-capacitor has an advantage that capacitance per unit area (2.2fF/um2) is larger than that of each MIM and PIP capacitor (1.0fF/um2) which needs additional process steps. Measured electrical characteristics of charge pump circuit have a good fitting to ideal Dickson equation and show that maximum output voltage 14V is enough to execute erase operation.

The high voltage driver to handle 9V in erase operation is shown in Fig. 5. The level shifter is composed of 5-stage circuits controlling the high voltage gradually. The power supplies for operating it consist of 0V, 2.25V (= 1/4VPP) for VP1, 4.5V (= 1/2VPP) for VP2, 6.75V (= 3/4VPP) for VP3 and 9V (= VPP). VP1, VP2 and VP3 are generated by push pull circuits. Thus, all terminals of each transistor in the high voltage driver are controlled so that the voltage difference between any two terminals is 4.5V (= 1/2VPP) or less, not over the breakdown voltage of I/O device.

Results and Discussion

Fig. 6 shows the program and erase characteristics. Program operation is executed by CHE injection and erase operation is done by AHH injection. Vth shifts > 1V are observed with programming at SL = 6V for 10ms and erasing at SL = 9V for 100ms, respectively.

Fig. 7 shows the Vth distributions after programming data “0”. In this case, MT is in program state and MB is in erase state. Each program and erase operation is executed without verify scheme. Fig. 8 shows the cycling endurance characteristics. By using the proposed high voltage driver, 100K cycling endurance is achieved without any failure of transistors. However Vth of each state is increasing gradually. It is considered that this phenomenon is due to accumulation of electrons in the side-spacer as shown in conventional charge-trapping SONOS device [8]. To compensate this phenomenon, the inverted data pre-programming scheme before erase operation is applied as a cycling stress leveling. Therefore ΔVth (Vth difference between MT and MB) can be kept nearly constant (> 0.97V@50%, > 0.70V@-3sigma) during 100K cycles.
Restore operation is similar to the latch-sense scheme of DRAM and executed simultaneously in one block (1KB in this test chip) within 1us.

Fig. 9 shows the simulated timing diagram of restore operation. Restore operation consists of three periods. In the first period, SL is set from 0V to VCC, each MT and MB is turned off. In the second period, each gate of MT and MB (VGT, VGB) is set to VCC through P3, P4. In the third period, SL is reset from VCC to 0V, then MT is turned on faster than MB on condition that Vth_MT is lower than Vth_MB in case of data”1”. As MT is turned on, the gate node of MB is pulled down to SL level, so MB is turned off. Consequently, the storage data in the side-spacer of each MT and MB is restored to the data latch.

Fig. 10 shows the Vgs-Ids characteristics of NVM element at reverse read mode after programming and erasing. In practical restore operation, Vgs bias condition shifts from 0V to VCC. As shown in Fig. 9, restoring to the data latch is accomplished within small Vgs (<1V), because even small Ids (<1uA) in erase state can pull down the gate node of NVM element in program state. This feature means that restore operation with differential sense-latch cell is suitable for low power operation. A static DC current does not flow during restore operation even though a transient current flows to SL in the first period (4.6uA/cell@peak) and VGT&VGB in the second period (3.3uA/cell@peak).

Fig. 11 shows the measured results of restore operation. ΔVth_min indicates the minimum Vth difference between MT and MB in all cells (1KB). These results show that the proposed restore operation can detect extremely small ΔVth (60mV) exactly.

By making use of these characteristics, mass data can be restored to each cell’s latch simultaneously for static data output with high sensitivity and low power consumption. This feature is useful for the applications which need the data restored during power-up sequence.
Retention characteristics under temperature range from 150C to 250C are shown in Fig. 12. $\Delta V_{th}@-5.89\sigma$ indicates the extrapolated worst bit when 512bit eCFlash macro is used at 1ppm chip failure rate. Experimental data is fitted by stretched exponential equation [9]. The activation energy is extracted as 1.32eV under this temperature range.

Fig.13 shows the cycling dependencies of retention lifetime at 150C. The retention lifetime is judged on condition $\Delta V_{th} = 60$mV. By utilizing the differential sense-latch cell, retention lifetime after 1K cycles is sufficient for 20 years at 150C; six orders of magnitude longer than that of the single cell structure.

These results prove the effectiveness of the differential sense-latch cell which can keep enough $\Delta V_{th}$ to restore the data exactly even after endurance cycles.

Figure 12. Retention characteristics (Variation of $\Delta V_{th}@50\%$, $\Delta V_{th}@-5.89\sigma$)

Figure 13. Cycling dependencies of retention lifetime

Conclusions
A 2KB “eCFlash” utilizing the differential sense-latch cell with charge-trapping storage has been developed without any additional process steps in a 0.25um technology. 10ms programming and 100ms erasing were achieved by CHE and AHH injection mode, respectively. 60mV sensitivity in restore operation and the handling scheme of twice the breakdown voltage have realized 100K cycling endurance and remarkable data retention of 20years at 150C.

References