High trap density and long retention time from self-assembled amorphous Si nanocluster floating gate nonvolatile memory

Daigil Cha and Jung H. Shin
Department of Physics, KAIST, Guseong-dong, Daejeon 305-701, Korea

Sangjin Park, Eunha Lee, Yoondong Park, Youngsoo Park, In-Kyeong Yoo, and Kwang Soo Seo
Samsung Advanced Institute of Technology, P.O. Box 111, Suwon 440-600, Korea

Suk-Ho Choi
Department of Physics and Applied Physics, College of Electronics and Information, Kyung Hee University, Yongin 449-701, Korea

(Received 23 August 2006; accepted 6 November 2006; published online 14 December 2006)

The memory performance of floating gate nonvolatile memory based on amorphous Si (a-Si) nanoclusters self-assembled during low-temperature oxidation is investigated. A 2 nm thick a-Si layer was grown on a 5.6 nm thick thermal oxide tunneling layer by ultrahigh vacuum ion beam sputter deposition and subsequently oxidized by annealing in flowing N2O (9:1) environment for 0–540 s at 900 °C. After oxidation, a 14 nm thick Al2O3 control oxide layer was grown by atomic layer deposition. The authors find that the a-Si layer breaks up upon oxidation, self-assembling into a dense array of <3 nm sized a-Si nanoclusters separated by thermal oxide after 180 s. This combination of discrete Si nanoclusters separated by thermal oxide and modest thermal budget enabled by the use of amorphous cluster enables achieving a trap density in the excess of 10^13 cm^{-2} and a retention time of >1000 s at 150 °C. © 2006 American Institute of Physics. [DOI: 10.1063/1.2404586]

The ever-increasing need for ubiquitous computing has increased the demand for nonvolatile memories that can offer high density and scalability. In particular, there is a great interest in developing a technology that can replace the polysilicon floating gate of conventional flash memories with a nanometer-scale charge trapping layer that can keep pace with continued scaling of nonvolatile memories. Of the many possible approaches, nanocluster Si (nc-Si) floating gate memory (NFGM) that seeks to replace polysilicon gate with a layer of nc-Si dispersed in the gate oxide matrix has attracted particular attentions because of its all-Si construction that is compatible with existing silicon processing industry. Furthermore, it had been reported that the presence of nc-Si passivated by oxygen leads to the formation of deep states with increased capture cross sections. This, coupled with the fact that charges in NFGM are trapped by discretely dispersed nc-Si and therefore prevented from easy lateral transport, has led to expectations that NFGM can provide a high reliability with thin gate oxides and low-power operation without sacrificing speed.

However, while many methods such as direct deposition, aerosol formation, and precipitation out of silicon-rich silicon oxides have been used to demonstrate the principle of NFGM, the charge trap densities that have been reported are in the range of only 10^{11}–10^{12} cm^{-2}. Such a density not only is much lower than ~10^{13} cm^{-2} that has been achieved by other nanoscale nonvolatile memory technologies such as silicon-oxide-nitride-oxide-silicon but can also limit the future scalability of NFGM. Trap density can be increased by increasing the nc-Si density. However, due to the difficulty of maintaining sufficient isolation between nc-Si as the nc-Si density is increased, obtaining both long retention time and high trap density (>10^{13} cm^{-2}) from NFGM has remained challenging.

In this letter, we report on using thermal oxidation of an ultrathin amorphous silicon layer to induce self-assembly of the continuous layer into an array of amorphous nc-Si separated by thermal oxide. While comparable in overall structure to previous NFGMs, this combination of thermal oxidation and amorphous nc-Si has several important differences that sets it apart. First, because neither precipitation nor crystallization is required, only a modest thermal budget of 180 s anneal at 900 °C is necessary for full transformation. Second, the low thermal budget preserves the integrity of ultrathin nc-Si and tunnel oxide layers. Finally, due to the use of oxidation, the amorphous nc-Si are separated by high quality thermal oxide and not defective oxides. We find that these factors combine to result in all-Si NFGM with a trap density in the excess of 10^{13} cm^{-2} and a long retention time, with a total charge loss of less than 25% even after 1000 s at 150 °C.

5.6 nm thick tunnel oxide layer was grown thermally on (100)-oriented, p-type silicon wafers by annealing them in flowing O2 gas at 850 °C for 5 min. Afterwards, a 2 nm thick amorphous silicon layer was deposited at room temperature using ultrahigh vacuum ion beam sputter deposition method using a Si target and 600 eV Ar ions. After deposition, the samples were oxidized by rapid thermal annealing at 900 °C for 0–540 s in a flowing N2O (9:1) environment. Finally, a 14 nm thick Al2O3 control oxide layer was deposited at 350 °C by atomic layer deposition using trimethylaluminum and O3 as precursor and oxidant, respectively. The deposited films were analyzed with transmission electron microscopy (TEM). For electrical characterizations, metal-oxide-semiconductor (MOS)-type diodes were fabricated by form-
FIG. 1. Cross-sectional high-resolution TEM images and schematic descriptions of the samples right after deposition (a), after a 60 s oxidation (b), and after a 180 s oxidation (c), respectively. The arrows in (b) indicate parts of the a-Si layer that are thinner than the others. The inset of (c) is the enlarged image of nc-Si.

ing 100 m diameter circularelectrodes of Al with 2% Nd by dc sputtering. Capacitance-voltage (C-V) curves were measured with an Agilent 4288A capacitance meter at 1 MHz with Keithley 230 voltage source. Program/erase (P/E) characteristics were measured by monitoring the flatband voltages of C-V curves after an initial P/E pulse of ±15 V/1 s bias and subsequent ±15 V pulses of various time durations between 10−5 and 1 s to the gate electrode. The retention characteristics were measured by monitoring the flatband voltages of C-V curves after applying a ±15/1 s P/E pulse, both at room temperature and 150 °C.

Figures 1(a)–1(c) show the cross-sectional high-resolution TEM images and the schematic descriptions of the samples right after deposition, after a 60 s oxidation, and after a 180 s oxidation, respectively. The dark band that runs across the top of the images is the a-Si layer. We find that the as-deposited a-Si layer is smooth, with well-defined, planar interfaces both at the top and at the bottom. Such is not the case after the 60 s oxidation. We find that oxidation proceeds in a nonplanar manner, resulting in parts that are thinner than others (marked by arrows). Such nonuniformity is accentuated by further oxidation such that after the 180 s, the a-Si layer breaks up into isolated nc-Si separated by thermal oxide regions that reach down to the underlying tunnel oxide layer. The nc-Si, however, are still amorphous, as can be seen in the inset of Fig. 1(c). Furthermore, their diameters are still close to the original a-Si layer thickness of 2 nm, consistent with previous reports that very small Si clusters are resistant to oxidation.14

Figure 2(a) shows C-V traces of the MOS diodes that were prepared after thermal oxidation with voltage sweep from 15 to −15 and back to 15 V. Also shown is the C-V trace of a control sample that had the 5.6 nm SiO2 tunnel oxide and 14 nm Al2O3 control oxide layers without the a-Si charge trapping layer. We observe very little hysteresis from the control sample. On the other hand, clear and large hysterese are observed from all the MOS diodes with the a-Si charge trapping layer, demonstrating their capability as nonvolatile memory elements. The flatband voltage shifts, and the corresponding total charge trap densities (electron +hole), are summarized in Fig. 2(b). We find that the initial oxidation leads to a rapid reduction of the flatband voltage shift and trap density, which we attribute to annealing of defects in the as-deposited a-Si layer.15 However, after 60 s, the flatband voltage shift and the trap density remain about the same up to an oxidation time of 540 s, even though the charge trapping layer changes from continuous a-Si layer to nc-Si.

Figure 3 shows the (P/E) characteristics of the MOS diodes with charge trapping layers that were oxidized for 60 or 180 s. We observe little difference between them, even though one is in the form of a continuous layer (60 s) and the other is in the form of an array of nc-Si (180 s). Other samples that were oxidized for different times showed similar P/E characteristics as well, except for the sample that underwent a 20 s thermal oxidation. In that case, we could not discriminate between programed and erased state of MOS diode due to the extremely rapid loss of charges after bias pulses (data not shown).

The fact that the trap density and P/E characteristics of the 60 and 180 s oxidized diodes are nearly the same indicates that the charge trapping/detrapping mechanisms are the same between them as well, consistent with Figs. 1 and 2 which show that the charge trapping layer consists of a-Si for both diodes. The exact nature of the charge trapping centers is not clear at this point. But the fact that the transformation of the charge trapping layer from a continuous, two-
dimensional layer into isolated, zero-dimensional nc-Si does not affect the charge trapping/detrapping mechanism suggests that the trap centers are likely to be located to be at the $a$-Si/oxide interface.\textsuperscript{3,8} We note, however, that the exact location of the trap centers, whether at the interface or inside $a$-Si, does not affect the final analysis or implications.

Transformation of the charge trapping layer into isolated nc-Si, however, is critical for obtaining stable charge retention characteristics, as is demonstrated in Fig. 4. We find that while charge losses at room temperature after 1000 s are negligible for all samples, at 150 °C, the MOS diodes with continuous $a$-Si film charge trapping layer (40 and 60 s) lose most of their trapped charges before 1000 s. On the other hand, the MOS diodes with nc-Si charge trapping layer (180, 300, and 540 s) lose at most 25% of their total trapped charges even after 1000 s at 150 °C. We note that the erasing state starts with no or very small flatband voltage shift compared to the programmed state. This is mainly due to the back tunneling of electrons from the gate electrode through the $\text{Al}_2\text{O}_3$ layer. We are currently investigating the effects of using different combinations of tunnel/gate oxide layers and thicknesses on the retention characteristics for the erasing state.

While such improvement in charge retention characteristics by forming isolated nc-Si is expected, it is still noteworthy that such a near complete isolation of nc-Si can be achieved while still maintaining a trap density of $>10^{13}$ cm$^{-2}$ that is much higher than had been reported for nc-Si based NFGM. We attribute this to the use of $a$-Si layer and low-temperature oxidation. Since nc-Si are formed by the thermal oxidation procedure, they are expected to be separated by high-quality, stoichiometric $\text{SiO}_2$. Furthermore, since the nc-Si are resistant to oxidation once formed,\textsuperscript{10} such high-quality oxide isolation can be attained without sacrificing either the nc-Si size or the nc-Si density. Finally, since all such transformation occurs within 180 s at low temperatures, the layered structure and gate oxide integrity are maintained. This is in contrast to, for example, precipitating crystalline nc-Si out of silicon-rich silicon oxides that requires annealing in the excess of 1100 °C for several hours\textsuperscript{19} and can lead to degradation of the gate oxide integrity. Lowering the annealing temperature results in layer structure preservation and formation of smaller nc-Si,\textsuperscript{17} but the nc-Si are separated by Si suboxide,\textsuperscript{12,13} and are thus not fully isolated.

Clearly, nonuniform oxidation of $a$-Si layer that leads to local oxidation while leaving other parts unoxidized is critical. At this moment, it is not clear what initiates such nonuniform oxidation. However, since the formation of $\text{SiO}_2$ at the top and bottom of silicon surface requires a volume expansion, the newly formed $\text{SiO}_2$ pulls Si together at the convex surface of $a$-Si layer, but pushes Si apart at the concave surface of a Si layer. Since the Si thickness is very thin at the concave surface of $a$-Si layer, the Si layer breaks up easily at the concave surface. It is also possible that this stress causes substantial portion of atoms to relocate from the thinner part of the layer to the thicker part, completing the self-assembly process. The oxidation of the remaining Si becomes self-limited by compressive normal stress at the Si/$\text{SiO}_2$ interface. Thus, an array of nc-Si isolated by thermal oxide would naturally self-assemble upon further oxidation. We expect that by understanding and controlling the initial formation of undulating oxidation front, nc-Si size and density, and thus charge trapping and retention properties, can be improved further.

In conclusion, we have investigated the memory performance of floating gate nonvolatile memory based on amorphous Si nanoclusters self-assembled during low-temperature oxidation. We find that the thermal oxidation of $a$-Si layer for more than 180 s induces self-assembly of the continuous layer into a dense array of $<3$ nm sized $a$-Si nanoclusters separated by thermal oxide. The use of that array enables achieving the $a$-Si nanocluster floating gate memory with a trap density $>10^{13}$ cm$^{-2}$ and a very high retention rate at 150 °C.

FIG. 4. (Color online) Retention characteristics of the MOS diodes that had $a$-Si charge trapping layer underwent thermal oxidation of 40 s (star), 60 s (diamond), 180 s (triangle), 300 s (circle), and 540 s (square). The pulses of (+15 V, 1 s) and (−15 V, 1 s) were applied to the gate electrode for program (closed)/erase (empty) operations, respectively.