Formation of stacked ruthenium nanocrystals embedded in SiO_2 for nonvolatile memory applications*

Mao Ping(毛平)†, Zhang Zhigang(张志刚), Pan Liyang(潘立阳), Xu Jun(许军), and Chen Peiyi(陈培毅)

(Institute of Microelectronics, Tsinghua University, Beijing 100084, China; Tsinghua National Laboratory of Information Science and Technology, Beijing 100084, China)

Abstract: Two methods are proposed to fabricate stacked ruthenium (Ru) nanocrystals (NCs): rapid thermal annealing (RTA) for the whole gate stacks, and RTA before each SiO_2 layer deposition. The size and aerial density of Ru NCs are 2–4 nm and 3×10^{12} cm^{-2} for the former method, compared to 3–7 nm and 2×10^{12} cm^{-2} for the latter. Because of the higher surface trap density and more uniform electron tunneling path between upper and lower Ru NCs, a 5.2 V memory window and 1 V after a period of 10 years are observed in metal oxide semiconductor (MOS) capacitors fabricated by the former method, which are much better than 4.6 V and no window remaining after one year observed in the latter. The former method is compatible with conventional CMOS technology.

Key words: ruthenium nanocrystal; stacked; formation; nonvolatile memory
DOI: 10.1088/1674-4926/30/9/093003 **PACC:** 7125; 7300

1. Introduction

Recently, nanocrystals (NCs) have been proposed to solve the reliability issue of conventional continuous floating-gate non-volatile memory because of the nature of discrete charge storage[1–3]. Among the diverse NC materials, metal NCs can provide the advantages of a wide range of available work functions, smaller energy perturbation due to quantum confinement and stronger channel coupling[4]. Several metals have been proposed to improve the performance of NC memory[5–7]. However, thermal stability and contamination during conventional high temperature technology are critical concerns for metal materials. Multilayer NCs have also been introduced to obtain a larger memory window and longer retention time[8]. However, these benefits are compromised by the vertical variation of upper-layer metal NCs, because lower-layer metal NCs undulate the interlayer oxide surface[9].

Ruthenium (Ru) can be used to form NCs because of its high chemical and thermal stability and proper work function (~4.7 eV)[10]. Single-layer Ru NCs have been formed by atomic layer deposition (ALD)[11] and the memory potential has also been demonstrated by memory transistor devices[12]. In this paper, uniform stacked Ru NCs embedded in a SiO_2 matrix are fabricated by rapid thermal annealing (RTA) for the whole gate stacks, which is compatible with conventional complementary metal oxide semiconductor (CMOS) technology. This method is proposed to improve uniformity both in the shape and aerial distribution of metal NCs, especially for multilayer applications.

2. Experiment

Two different methods were used to fabricate Ru NCs: RTA for the whole gate stacks, and RTA before each SiO_2 layer deposition. Sample A was fabricated by RTA for the whole gate stacks. A 6 nm tunnel oxide was thermally grown on n-type (100) silicon wafers. Subsequently, a thin Ru layer (~1 nm) was deposited onto the tunnel oxide by magnetron RF sputtering (20 s). Then, a 4 nm SiO_2 interlayer was deposited by sputtering and followed by another Ru layer deposition. After that, a 30 nm thick blocking oxide was deposited by PECVD at 300 °C. When the whole gate stacks were formed, an RTA at 900 °C for 30 s was executed to form discrete Ru NCs in two Ru layers, which also improved the quality of the blocking oxide and the interface between the Ru NCs and the surrounding SiO_2 matrix. Finally, an aluminum electrode was patterned to form the MOS capacitors. For comparison, sample B was fabricated by the conventional sequent RTA method[9], which is to carry out RTA (600 °C, 60 s) before the deposition of each SiO_2 layer. In this way, Ru NCs with comparable size were obtained before the deposition of the blocking oxide. The process flow of the above methods is shown in Fig. 1.

The microstructure analysis was performed by high resolution transmission electron microscopy (HRTEM). To imitate real flash memory operating conditions, high voltage pulses were applied to execute the program/erase (P/E) operation for the MOS capacitors. The pulse amplitude and width refer to the P/E bias and time respectively. A relatively small range of

* Project supported by the State Key Development Program for Basic Research of China (No. 2006CB302702) and the National Hi-Tech Research and Development Program of China (No. 2008AA031403).

† Corresponding author. Email: achievemp@gmail.com

Received 13 February 2009, revised manuscript received 24 March 2009

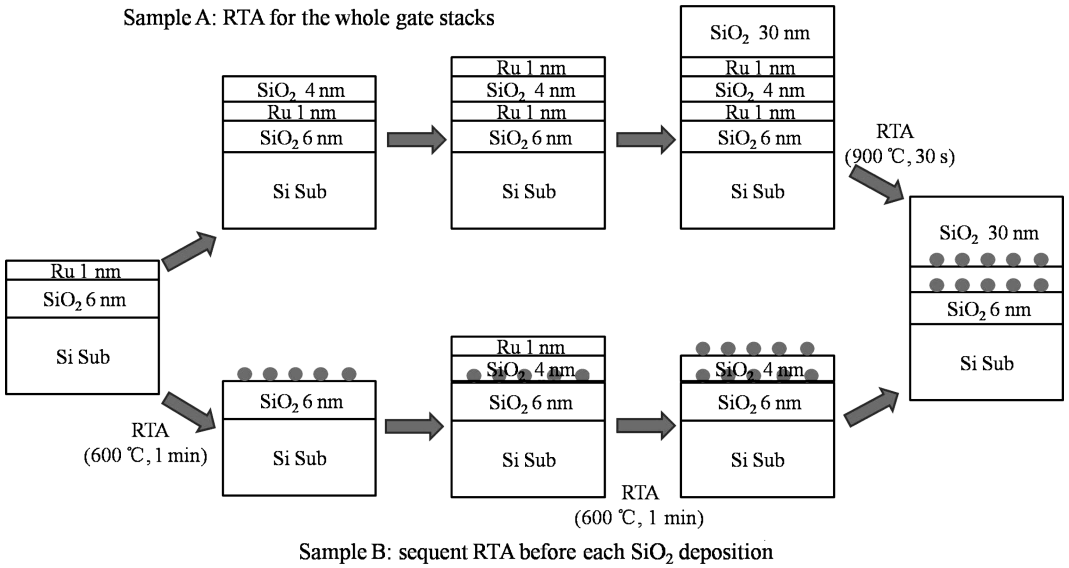

Fig. 1. Fabrication process flow of stacked Ru NCs.

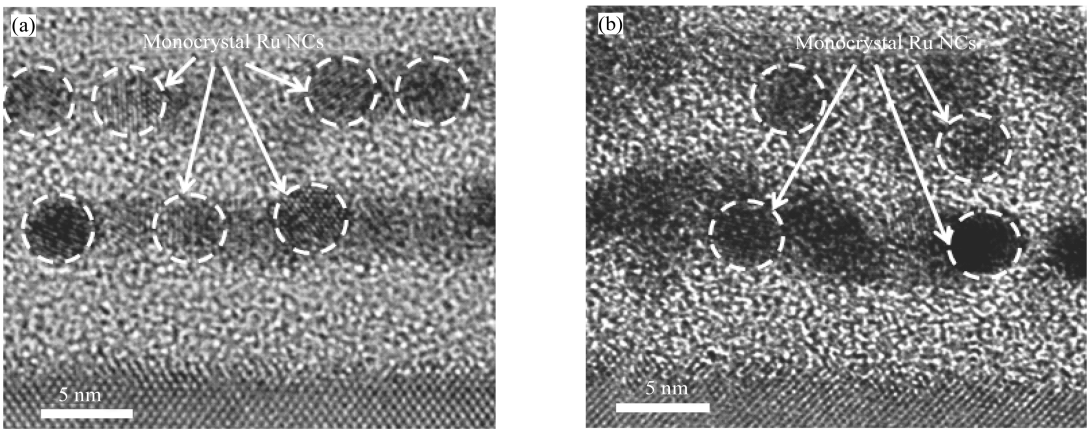

Fig. 2. Cross-sectional HRTEM images of stacked Ru NCs formed by two methods: (a) Sample A is by RTA (900 °C, 30 s) for the whole gate stacks; (b) Sample B is by RTA (600 °C, 1 min) before each SiO_2 deposition.

capacitance–voltage (C–V) sweep was used to read the flat-band voltage shift, which was measured at a high frequency of 1 MHz by a HP4284 Precision LCR meter.

3. Results and discussion

The mechanism of metallic NC fabrication is based on the thermally activated dewetting mechanism of ultra-thin metallic films. As described by Liu *et al.*[4], metallic NCs are self-assembled during the annealing treatment by lowering the total energy state of the initial metallic layer. Although the as-deposited film comes naturally with some thickness perturbation and even nano-islands before RTA, this deformation is not enough to obtain a clear and uniform separation. The high temperature offers sufficient energy to enhance atomic surface mobility and enables relaxation of the initial film stress. As a consequence, the film further breaks into islands along its initial roughness perturbations. Cross-sectional HRTEM images of stacked Ru NCs are shown in Fig. 2. The granular, separated and monocrystal Ru NCs are clearly observed embedded in the SiO_2 matrix in both ways. The size and aerial density per layer of the Ru NCs are about 2–4 nm and 3×10^{12} cm^{-2} for sample A [Fig. 2(a)], and 3–7 nm and 2×10^{12} cm^{-2} for sample B [Fig. 2(b)], respectively.

Obviously, stacked Ru NCs formed by RTA for the whole gate stacks exhibit less size dispersion and better aerial uniformity than those formed by RTA before each SiO_2 deposition, especially in the vertical distribution. This is attributed to the thermal stability of Ru and the vertical confinement introduced by the pre-existing SiO_2 matrix. Firstly, the reaction between Ru and SiO_2 is less favorable from theoretical thermal dynamic calculations, which expect a positive Gibbs free energy for the Ru and SiO_2 reaction (Metal+SiO_2 $\rightarrow$ M-oxide+M-silicide, $\Delta G > 0$)[13]. Secondly, the pre-existing vertical SiO_2 matrix exerts additional vertical confinement to Ru atoms in the RTA process. Therefore, the self-assembling process of Ru atoms mainly takes place along the lateral direction and the remaining vacancies are filled by SiO_2, which makes the formation of NCs more controllable. Furthermore, the size and density of the Ru NCs in each layer can be easily controlled by the initial thickness of the sputtering Ru layer. This fabrication method is compatible with conventional CMOS technology.

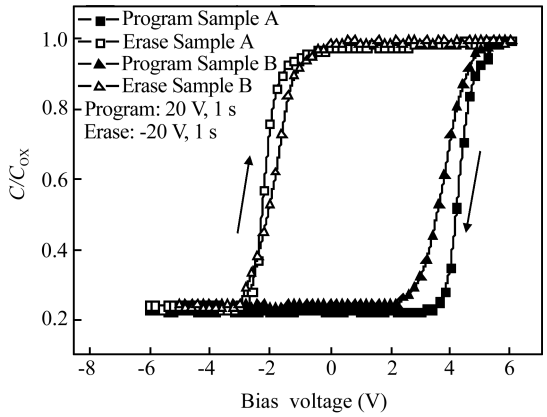

Fig. 3. Capacitance–voltage (C–V) characteristics for samples A and B. The amplitude and width of the voltage pulse are ±20 V and 1 s for the program and erase state respectively.

The flatband voltage shift ($\Delta V_{\rm FB}$) of the C–V sweep curve is shown in Fig. 3. The P/E voltage and time are ±20 V and 1 s, respectively. To avoid disturbing the charges stored in the Ru NCs, the amplitude of the low voltage C–V sweep read is set to ±6 V. A clockwise C–V hysteresis is obtained after the P/E operation. This proves the Ru NCs can store and release charge during the P/E operation. Electrons tunnel into the Ru NCs and tunnel back to the substrate through the tunnel oxide layer, under positive and negative bias respectively. A P/E window of 5.2 V (sum of positive and negative $\Delta V_{\rm FB}$) can be obtained in sample A, in comparison to 4.6 V in sample B under the same P/E operation. Moreover, a steeper transition region in C–V characteristics of sample A ensures fewer interface states, thus a better quality of interface between the Ru NCs and the surrounding SiO_2 matrix can be obtained after RTA for the whole gate stacks.

Using the classical formula of floating gate memories for estimation, $\Delta V_{\rm FB} = -D_{\rm s}q/C_{\rm eff}$, where q is the elementary charge, $D_{\rm s}$ is the equivalent surface charge density, and $C_{\rm eff}$ is the equivalent capacitance between the control gate and the NCs. Assuming that the charge is uniformly distributed among the well-isolated NCs, $C_{\rm eff}$ could be treated as equal because of the similar physical structure. The equivalent surface charge density can be calculated from the positive $\Delta V_{\rm FB}$, which are 3.2 V for sample A and 2.6 V for sample B. Therefore, the surface charge densities are 2.76×10^{12} cm^{-2} and 2.24×10^{12} cm^{-2}, respectively. The higher surface charge density of sample A is in accordance with the higher aerial density shown in Fig. 2.

The P/E characteristics of samples A and B are shown in Fig. 4. Because of the energy difference between the electron affinity of the silicon substrate and the work function of the Ru NCs, the tunnel barrier faced by the electrons is higher in the Ru NCs and lower in the Si substrate. So the electrons are more easily trapped in the Ru NCs than detrapped from them. This is proved by the fact that the positive $\Delta V_{\rm FB}$ is larger than the negative $\Delta V_{\rm FB}$. The P/E window of sample B is larger than

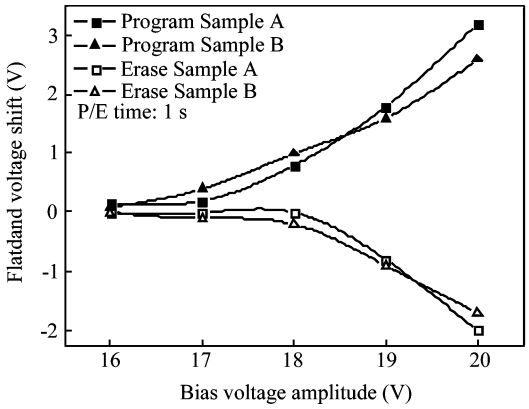

Fig. 4. Flatband voltage shifts of Ru NCs capacitors under various P/E operations.

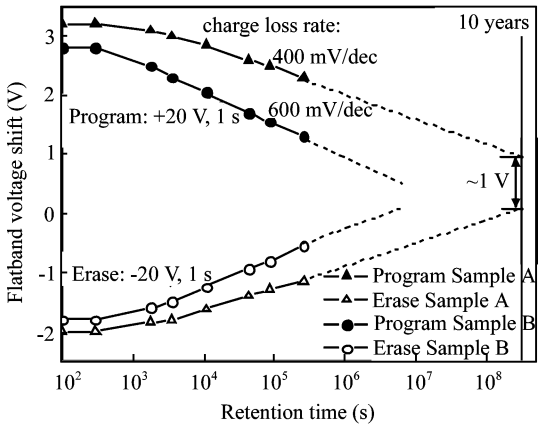

Fig. 5. Retention performance of stacked Ru NC MOS capacitors fabricated by the two methods.

that of sample A below 19 V bias, which can be explained by non-uniform electron tunneling between the lower and upper Ru NCs. Due to the larger distance dispersion between lower and upper NCs in sample B, it is easier for electrons to tunnel through a relatively thinner interlayer oxide under the above bias conditions.

The retention performance at room temperature is shown in Fig. 5. The initial P/E windows are 5.2 and 4.6 V for samples A and B, respectively. It can be predicted that about 1 V memory window remains after 10 years for sample A, while no window remains after one year for sample B. Thus, the corresponding charge loss rate of sample A (400 mV/dec) is smaller than that of sample B (600 mV/dec). The charge leakage from the upper Ru NCs is limited by Coulomb blockade and relatively thicker tunnel oxide for a double-layer NC memory device[9]. Furthermore, a more uniform and controllable vertical charge leakage path from the upper to lower NCs could be predicted because of a more uniform interlayer oxide obtained in sample A than in sample B, as shown in Fig. 2. Therefore, the charge loss rate of sample A is much smaller than that of sample B. This feature is more critical for retention performance when the tunnel oxide is reduced to below 3 nm. The introduction of a high temperature oxide or high-k material can further reduce the charge loss rate and improve the retention

performance.

4. Conclusion

In summary, stacked Ru NCs have been fabricated by two methods: RTA for the whole gate stacks and RTA before each SiO_2 deposition. Higher density (3×10^{12} cm^{-2}) and smaller sized (2–4 nm) stacked Ru NCs with better uniformity both in shape and aerial distribution are obtained by the former method, which achieves a higher surface trap density and more uniform electron tunneling path between the upper and lower NCs. Therefore, a MOS capacitor fabricated by the former method exhibits a larger memory window and much better retention performance. This fabrication method is compatible with conventional CMOS technology.

References

[1] Tiwari S, Rana F, Hanafi H, et al. A silicon nanocrystals based memory. Appl Phys Lett, 1996, 68: 1377

[2] Blauwe J D. Nanocrystal nonvolatile memory devices. IEEE Trans Nanotechnol, 2002, 1(1): 72

[3] Li Wei, Zhang Zhigang, Liang Renrong, et al. Storage characteristics of nano-crystal Si devices using different measurements for MOS capacitors. Chinese Journal of Semiconductors, 2007, 28(8): 1169

[4] Liu Z T, Lee C H, Narayanan V, et al. Metal nanocrystal memories-part I: device design and fabrication. IEEE Trans Electron Devices, 2002, 49: 1606

[5] Dufourcq J, Bodnar S, Gay G, et al. High density platinum nanocrystals for non-volatile memory applications. Appl Phys Lett, 2008, 92: 073102

[6] Samanta S K, Yoo W J, Samudra G, et al. Tungsten nanocrystals embedded in high-*k* materials for memory application. Appl Phys Lett, 2005, 87: 113110

[7] Chen W R, Chang T C, Liu P T, et al. Formation of stacked Ni silicide nanocrystals for nonvolatile memory application. Appl Phys Lett, 2007, 90: 112108

[8] Ohba R, Sugiyama N, Uchida K, et al. Non-volatile Si quantum memory with self-aligned doubly-stacked dots. IEDM, 2000: 13.5.1

[9] Lee C, Seetharam G A, Kan E C. Operational and reliability comparison of discrete-storage nonvolatile memories: advantages of single- and double-layer metal nanocrystals. IEDM, 2003: 22.6.1

[10] Zhong H, Heuss G, Misra V. Electrical properties of RuO_2 gate electrodes for dual metal gate Si-CMOS. IEEE Electron Device Lett, 2000, 21: 593

[11] Farmer D B, Gordon R G. High density Ru nanocrystal deposition for nonvolatile memory applications. J Appl Phys, 2007, 101: 124503

[12] Mao Ping, Zhang Zhigang, Pan Liyang, et al. Nonvolatile memory devices with high density ruthenium nanocrystals. Appl Phys Lett, 2008, 93: 242102

[13] Wang S Q, Hong S, White A, et al. Interfacial reactions in the SiO_2/Ru and SiO_2/Ru/Al-Si structures. J Appl Phys, 1995, 77: 5751