Carbon Nanotube Memory by the Self-Assembly of Silicon Nanocrystals as Charge Storage Nodes

Mario Olmedo,† Chuan Wang,‡ Koungmin Ryu,‡ Huimei Zhou,† Jingjian Ren,† Ning Zhan,† Chongwu Zhou,‡ and Jianlin Liu†,*

†Department of Electrical Engineering, University of California, Riverside, California 92521, United States and ‡Department of Electrical Engineering, University of Southern California, California 90033, United States

ABSTRACT A memory structure based on self-aligned silicon nanocrystals (Si NCs) grown over Al₂O₃-covered parallel-aligned carbon nanotubes (CNTs) by gas source molecular beam epitaxy is reported. Electrostatic force microscopy characterizations directly prove the charging and discharging of discrete NCs through the Al₂O₃ layer covering the CNTs. A CNT field effect transistor based on the NC/CNT structure is fabricated and characterized, demonstrating evident memory characteristics. Direct tunneling and Fowler–Nordheim tunneling phenomena are observed at different programming/erasing voltages. Retention is demonstrated to be on the order of 10⁴ s. Although there is still plenty of room to enhance the performance, the results suggest that CNT-based NC memory with diminutive CNTs and NCs could be an alternative structure to replace traditional floating gate memory.

KEYWORDS: carbon nanotubes · carbon nanotube memory · electrostatic force microscopy · nanocrystal memory · nanocrystal self-assembly

Advances in carbon nanotube field effect transistor (CNT–FET) memory have been achieved in the past decade. Most prominently, the charge trapping memory with the CNT exposed to air has shown long memory window stability,¹ robust endurance,² sensitivity to various gases,³,⁴ and even resistive behavior.⁵ Other approaches such as oxide-nitride-oxide (ONO),⁶ nanocrystal (NC),⁷–¹¹ floating gate,¹²,¹³ ferroelectric,¹⁴,¹⁵ and oxide¹⁶ have also been implemented in CNT-based memories. These devices have shown interesting properties such as high-speed programming/erasing⁶ and single electron detection.⁷,⁸ Although the use of Au NCs on an oxide-covered CNT for memory was reported,⁷ the devices were fabricated based on randomly distributed CNTs. In addition, it remained unclear whether evaporation of the Au thin film by electron beam evaporation could give rise to good self-alignment of Au dots on oxide-covered CNTs and whether the NCs in the vicinity of the CNTs would add to the memory effect. In this article, a memory structure using self-aligned Si NCs as the charge trapping nodes on parallel-aligned CNTs is demonstrated. The use of parallel-aligned CNTs would enable scalable fabrication, minimize the tube-to-tube junctions, and improve the device performance. Electrostatic force microscopy (EFM) measurements show that the charges on the NCs can be programmed and erased from the CNTs through an Al₂O₃ tunneling layer. Furthermore, the device shows both direct tunneling (DT) and Fowler–Nordheim (FN) tunneling characteristics at corresponding gate voltages, as well as good retention performance.

A major advantage that CNT-based devices have over traditional MOSFET is its superior scalability.¹⁷ The channel width of a CNT is less than 2 nm, and with the addition of discrete NCs, the channel length for memory can also be decreased. A characteristic of CNT-based NC memory is that the NCs keep the charge discrete, which can withstand non-uniformities in the thickness of the tunneling oxide caused by CNT roughness. Areas where the CNT bends out of plane, causing the tunneling oxide layer to be thinner, would lead to easier charge leakage. If the memory would have been made using traditional poly-Si as a floating gate, it would be prone to very short retention because even a single leakage path caused by the CNT roughness would drain all charges on the continuous poly-Si gate. Moreover, NC memory in one-dimensional channels will have what is known as the “bottleneck” effect,¹⁸ for which as few as a single charged NC can keep the device in an ON or OFF state. This means that the ultimate scalability of this memory is limited only to the size of a single NC, which in this case is approximately 5 nm.
RESULTS AND DISCUSSION

The NC samples started with the growth of parallel-aligned CNTs on a quartz substrate by the method described in ref 19 and then transferred to a highly boron-doped p⁺-Si[100] substrate with 300 nm thermally grown SiO₂ by thermal tape (see the Supporting Information for experimental details). The CNT/SiO₂ structure was then introduced to an atomic layer deposition (ALD) chamber for the deposition of 11 nm Al₂O₃ with 0.09 nm/cycle rate at 300 °C. The average roughness for the ALD-grown Al₂O₃ surface was measured to be 0.3 nm by atomic force microscopy (AFM), which was slightly larger than the 0.2 nm average roughness of the SiO₂ only samples. This small increase in the film roughness was not found to affect NC growth. The Al₂O₃-covered CNTs were then introduced to a gas source molecular beam epitaxy (GSMBE) chamber for Si NC growth. Figure 1a shows a schematic of the structure, and Figure 1b shows an AFM image of a typical sample, which was used for device fabrication. The density of Si NCs is approximately 2 × 10⁹ cm⁻². For conventional NC memory applications, where a two-dimensional cell is implemented, this areal density is too low; however, it is acceptable for a one-dimensional memory cell as most of the NCs are on the apexes of the CNTs. More evidence of Si NC alignment on parallel-aligned CNTs covered by Al₂O₃ is reported in the Supporting Information. Similar preferential growth was noticed on the growth of Si NCs on randomly aligned CNTs covered with HfO₂. There it was studied how different growth conditions such as growth time, oxide thickness, and source gas flow rate influenced the NC density. There are two reasons that may contribute to the preferential growth of the Si dots on the apex of the oxide-covered CNTs. The first is related to strain in the Al₂O₃ layer as a result of the CNT underneath it. This strain can be manifested as a change in surface energy, with less energy at the surface of the oxide ridge on the CNT, or with an increase in point defects at the areas with strain. The second one is an effect from the trapping areas that are formed due to the slower ALD oxide growth on the top of the CNTs. The first reason is inspired by both theoretical and experimental work on single crystal quantum dot structures on a patterned dissimilar substrate, showing that places with less surface energy will be preferential areas of nucleation of dots. The second reason is proposed because the growth of Al₂O₃ by ALD uses H₂O as an oxygen precursor. This means the precursor cannot attach to the inert CNT so the layer has to overgrow the CNT from the sides. It has been claimed that at least 8 nm is needed to overgrow the CNTs to completely cover them, which is thinner than the 11 nm used here. It can be safely assumed yet that coverage will not be uniform on a CNT that has uneven contact with the substrate, and more point defects will be present at the places along the CNT where it is farther from the substrate.

To detect charging and discharging properties of the NCs compared to the surrounding areas, EFM was used to map the charge distribution across the surface. The scale bar represents 500 nm.
EFM measurement consists of a metalized Si tip first measuring the sample topography in tapping mode. Then the tip is raised at a constant height with respect to topography and does a second pass over the same area. The electrostatic force originated from the surface will interact with the tip by changing its resonant frequency. An attractive force lowers the frequency and appears dark in the image, while a repulsive force appears bright. A reading voltage ($V_{\text{str}}$) of $-3 \text{V}$ was applied at the tip (Figure 2a) during lift to be able to repel any capacitance effect from the CNTs and distinguish different types of charges present at the surface. Figure 2b shows the initial state EFM image of a CNT with NCs on top and adjacent to it, and the inset shows its corresponding topography. Although aligned NCs are clearly observed in the topographic image, the EFM image shows relatively even contrast due to the lack of charges on the surface. To program and erase the NC/CNT structure, the tip was put in contact with the surface by using the ramp command, which lowers the tip to the point at which the frequency of the tip is nullified. This ramp command was tuned to a level of $5 \text{ to } 10 \text{ mV}$ before the tip reached full contact with the surface in order to minimize charge leakage from the dots to the tip and vice versa. These experiments were also carried out in a nitrogen atmosphere to avoid anodic oxidation. While the tip was in ramping mode, a voltage ($V_p$) of $12 \text{V}$ was applied for programming and $-12 \text{V}$ for erasing ($V_e$). In both cases, the total ramping time was $5 \text{s}$. The ramp time defines the amount of charge that was in the ramp routine continuously. Since a $1 \text{Hz}$ ramp rate was used, it meant that the tip would only ramp $5 \text{ times}$ in the ramping time of $5 \text{s}$. This would account for a total contact time of about $500 \text{ ms}$ for which the tip was at the lowest position. The results indicate the change in charge from a dark CNT and light QDs after programming to bright CNT and dark QDs after erasing. The reason for this change in charge state was the large electric field that is being applied by the tip. At $V_p = 12 \text{V}$ (Figure 2c,d), the electrons would be attracted from the CNT through the oxide layer and into the NCs. This also happens with other NCs in the nearby area because of the tip size and the electric field distribution on the surface. An opposite effect occurs when $V_e = -12 \text{V}$ (Figure 2e,f) was introduced on the tip; here electrons were repelled and tunnel back from the NCs to the CNT. Since the $\text{Al}_2\text{O}_3$ layer outside the vicinity of the CNTs shows little change, that is, there is no evidence of charge injection outside of the NCs, it can be concluded that $\text{Al}_2\text{O}_3$ does not play a major role in the storage of charges. This transfer of charge from a channel, in this case the CNT, to the floating gate, here the Si NCs, is similar to the charge transfer process in a traditional MOSFET-based memory.

To further study the memory effects of the NC/CNT structure, back-gated field effect transistor (FET) devices were fabricated. The fabrication began with the passivation of the NCs with $20 \text{nm} \text{Al}_2\text{O}_3$, followed by etching down to the CNTs by photolithography and wet etching. Source and drain contacts consisting of a $5 \text{ nm} \text{Ti}$ adhesion layer followed by $50 \text{ nm} \text{Pd}$ layer were then deposited by electron beam evaporation. A second photolithographic step was performed to align the measurement pads to the source/drain contacts, and an additional $100 \text{ nm}$ layer of Pd was deposited on the resulting pattern. The channel width and length of this device are $5$ and $3 \mu\text{m}$, respectively. Figure 3a shows the schematic of the device, and Figure 3b shows a scanning electron microscopy (SEM) image of the fabricated device, which accommodated four parallel CNTs as its channel. The linear characteristics of the $I_{\text{d}}-V_{\text{g}}$ curves in Figure 3c indicate that the Ti/Pd contacts used are Ohmic with a current of $862 \text{nA}$ at $V_{\text{ds}} = 1 \text{V}$ and a grounded gate. The inset $I_{\text{d}}-V_{\text{g}}$ curve shows that the CNTs are p-type semiconducting, and the current on and off ratio $I_{\text{on}}/I_{\text{off}}$ is less than $1 \times 10^3$, which is similar to other FETs made with these CNTs. Semiconductor characteristics in the CNTs are important because the large on and off ratio will facilitate the characterization of the memory, including an accurate evaluation of the threshold voltage ($V_{\text{th}}$), shift and monitoring of the $I_{\text{on}}$ and $I_{\text{off}}$ currents as a function of time.

The change in $V_{\text{th}}$ ($\Delta V_{\text{th}}$) is tracked to measure the change in memory window. The method of linear extrapolation of $I_{\text{d}}-V_{\text{g}}$ is used to obtain $V_{\text{th}}$, that is, $V_{\text{th}} = V_{\text{gi}} - V_{\text{g}}/2$ with $V_{\text{gi}}$ being the gate voltage at which
$I_{ds} = 0\ \text{A}^{27}$ Due to the effect of the fringing electric fields originated from the two-dimensional back gate and one-dimensional CNT channel, as well as capacitive coupling between the NCs and the back gate,\textsuperscript{28} a positive voltage is used for programming while a negative voltage is used for erasing. In programming, $V_{th}$ shows an increase due to the presence of electrons in the NCs, while for erasing, $V_{th}$ decreases due to the lack of electrons or the presence of holes in the NCs. Figure 4 shows the change in $V_{th}$ with an increase of absolute value of $V_g$. The pulse width used was 1 s with the programming voltage applied first, followed by the erasing voltage. As the pulse amplitude increases, $V_{th}$ increases. At 20 V, there is an evident change in slope. This was attributed to two different types of tunneling mechanisms, that is, direct tunneling (DT) for $V_g < 20$ V and Fowler–Nordheim (FN) tunneling for $V_g > 20$ V. At low $V_g$ amplitudes, charges go through the insulator by DT, leading to the slow increase of $\Delta V_{th}$. As $V_g$ increases, the potential difference between the NCs and CNT channel increases to make the effective tunneling barrier width thinner, allowing more charges to tunnel in FN tunneling mode, and as a result, $\Delta V_{th}$ increases dramatically. It was calculated that FN tunneling happens when the electric field between the CNT and the NCs is 6.4 MV/cm, which is similar to the reported results that FN tunneling starts on other NC structures with an electric field of 7 MV/cm.\textsuperscript{29}

Retention characteristics were measured by monitoring $I_d$ for $2 \times 10^4$ s, as shown in Figure 5a after the device had been programmed and erased. Figure 5b shows the $I_d-V_g$ curves for the fresh device, programmed device after +30 V for 1 s, and erased device after −30 V for 1 s, indicating evident memory effect. In addition, the reading gate voltage can be tuned to show the largest window for the ON and OFF state currents. In this case, the reading voltage is −2.5 V. The trade-off is that the device will experience shorter electron retention due to the presence of the constant negative potential on the NCs. This can be avoided if a reading voltage of 0 V is used, although in this case, the memory window is smaller than that of the −2.5 V reading. In Figure 5a, the top curve is the retention performance in the programmed state while the bottom curve is the retention performance in the erased state. The retention curve for the programmed state has two slopes. The first slope ($\Delta I_d/t$, where $t$ is from the initial to $7 \times 10^3$ s), has a rapid loss of $\Delta I_d/t \approx -46$ pA/s, which is due to the fast leakage of the stored electrons at the higher energy levels in the NC quantum well or shallow defect levels such as interface trap levels between the CNTs and the surrounding oxide. The smaller slope of $\Delta I_d/t \approx -1.4$ pA/s for the waiting time longer than $7 \times 10^3$ s is due to the slower leakage of the stored charges occupying deeper energy levels in the NC quantum well. Hole retention differs from
electret retention due to a smaller slope of $\Delta \mu / t \approx 1$ pA/s. This phenomenon is attributed to the higher electron barrier from the CNT to the NCs. Figure 5c shows the band alignment of the device structure. The band gap ($E_g$) of Al$_2$O$_3$ is 8.8 eV.\(^3\) This defines an electron barrier of 2.8 eV for electrons at the NC side, which is indeed smaller than the electron barrier of 3.35 eV at the CNT side. It should be noted that for amorphous Al$_2$O$_3$, which is closer to the film grown by ALD, $E_g$ was observed to be 6.95 eV.\(^1\) This modification shall result in an electron barrier of 2.08 eV at the Si NC side and of 2.63 eV at the CNT side. These values though smaller would still yield the same relationship between the hole and electron retention results. By extrapolating the curve, the retention of this device can be approximately $1 \times 10^4$ s when $I_{on}/I_{off}$ ratio reaches 10 and $4 \times 10^4$ s at the point of complete loss of charge. In contrast, the longest reported retention time for a CNT-based memory by using the time lasted until the $I_{on}/I_{off}$ ratio reaches 10 is $1.5 \times 10^4$ s.\(^16\) Further extrapolation leads to a retention time of around $3 \times 10^4$ s assuming all charges are lost, which suggests the present device has a slightly better retention.

Nevertheless, the retention is still short compared with state-of-the-art Si-MOSFET-based floating gate memory where the requirement is $\sim 3 \times 10^7$ s. This suggests that there is plenty of room to optimize the CNT-based memory device, such as improving the quality of tunneling oxide.

**CONCLUSIONS**

A memory using Si NC floating gate on Al$_2$O$_3$-covered aligned semiconducting CNTs has been demonstrated. EFM measurement results show that the charges are transferred between the CNTs and the NCs in the process of programming and erasing and are kept discretely in the NCs during the retention. NC charging characteristics exhibit a change from DT at lower $V_g$ to FN tunneling at higher $V_g$. Finally, charge retention was measured by operating the device continuously and measuring the change in $I_d$, for which a moderate retention time was achieved. We believe that, with further optimization, this technology can enhance the scalability of NC memory. Therefore, it may be a potential candidate to replace traditional floating gate memory in the future.

**METHODS**

**CNT Transfer.** The parallel-aligned CNTs were grown on a quartz substrate by the method described in ref 19. Then a 100 nm Au layer was deposited by e-beam evaporation with a 0.01 nm/s rate for the first 5 nm then a 0.1 nm/s rate for the rest. A heavily boron-doped Si-p$^+$ wafer that had been covered with 300 nm thermally grown SiO$_2$ was heated to 150 °C on a hot plate. The thermal release tape (Nitto Denko Revalpha 3198 M tape) was subsequently used to remove the Au/CNT film from the quartz and transfer to the SiO$_2$ surface. Oxygen plasma at 40 W for 10 min was then used to get rid of the tape residue, and finally, the Au layer was etched with an Au etchant (Transene Gold etchant TFA).

**CNT-Based Si NC Memory Device Structure Growth.** The CNTs on SiO$_2$ samples were then prepared for ALD growth by introducing them to UV ozone treatment for 10 min to make the sample surface hydrophilic. The ALD sources used were aluminum tetrachloride (AlCl$_4$) and water (H$_2$O). The growth was carried out at 300 °C under low pressure (2 × 10$^{-7}$ Torr) with a 15 s interval between each source. A nominal 11 nm Al$_2$O$_3$ was grown on the sample to act as the tunneling layer. The samples were then introduced into a home-built GSMBE system, which was subsequently pumped to high vacuum ($1 \times 10^{-7}$ Torr). The sample was heated to 650 °C, and disilane was introduced at 4 sccm for 5 min. For fabrication of the FET memory devices, a passivation layer of 20 nm Al$_2$O$_3$ was grown using the same ALD conditions as used for the tunneling layer.

**Acknowledgment.** The authors thank the National Science Foundation (DMR-0807232), and DMEA Center on Nanomaterials and Nanodevices (3D Electronics program), and Focused Research Center Program (FCRP) on FENA for financial support of this research. The San Bernardino Community College District Applied Technology Training Center is also acknowledged for providing access to their AFM measurement system.

**Supporting Information Available:** AFM images of NC alignment change with temperature and flow rate. Transient memory characteristics of the device fabricated. Retention characteristics of reference sample without NCs. This material is available free of charge via the Internet at http://pubs.acs.org.