Memristive behavior in a junctionless flash memory cell
Ikram Orak, Mustafa Ürel, Gokhan Bakan, and Aykutlu Dana

View online: http://dx.doi.org/10.1063/1.4922624
View Table of Contents: http://scitation.aip.org/content/aip/journal/apl/106/23?ver=pdfcov
Published by the AIP Publishing

Articles you may be interested in
Performance improvement of metal-Al2O3-HfO2-oxide-silicon memory devices with band-engineered Hf-
aluminate/SiO2 tunnel barriers
J. Vac. Sci. Technol. B 31, 041201 (2013); 10.1116/1.4807842

Memristive operation mode of floating gate transistors: A two-terminal MemFlash-cell

Suppression of dielectric crystallization on metal by introduction of SiO2 layer for metal floating gate memory
blocking oxide

Band bending effect induced by gate voltage on the charge loss behavior of charge trap flash memory devices
Appl. Phys. Lett. 96, 052106 (2010); 10.1063/1.3295697

Equivalent oxide thickness reduction of interpoly dielectric using ALD- Al 2 O 3 for flash device application
J. Vac. Sci. Technol. B 22, 2295 (2004); 10.1116/1.1781659
Memristive behavior in a junctionless flash memory cell

Ikram Orak,1,2 Mustafa Ürel,3 Gokhan Bakan,4 and Aykutlu Dana3,a)

1Vocational School of Health Services, Bingöl University, 12000 Bingöl, Turkey
2Department of Physics, Faculty of Science and Art, Bingöl University, 12000 Bingöl, Turkey
3UNAM Institute of Materials Science and Nanotechnology, Bilkent University, 06800 Ankara, Turkey
4Faculty of Engineering, Antalya International University, 07190 Antalya, Turkey

(Received 27 April 2015; accepted 3 June 2015; published online 12 June 2015)

We report charge storage based memristive operation of a junctionless thin-film flash memory cell when it is operated as a two-terminal device by grounding the gate. Unlike memristors based on nanoionics, the presented device mode, which we refer to as the flashristor mode, potentially allows greater control over the memristive properties, allowing rational design. The mode is demonstrated using a depletion-type n-channel ZnO transistor grown by atomic layer deposition (ALD), with HfO2 as the tunnel dielectric, Al2O3 as the control dielectric, and non-stoichiometric silicon nitride as the charge storage layer. The device exhibits the pinched hysteresis of a memristor and in the unoptimized device, Roff/Ron ratios of about 3 are presented with low operating voltages below 5 V. A simplified model predicts Roff/Ron ratios can be improved significantly by adjusting the native threshold voltage of the devices. The repeatability of the resistive switching is excellent and devices exhibit 106 s retention time, which can, in principle, be improved by engineering the gate stack and storage layer properties. The flashristor mode can find use in analog information processing applications, such as neuromorphic computing, where well-behaving and highly repeatable memristive properties are desirable. © 2015 AIP Publishing LLC. [http://dx.doi.org/10.1063/1.4922624]

Memory and logic operations form the basis of modern information processing systems, where increasingly high density and lower power operation enable higher performance computing and a wide variety of contemporary mobile platforms. Among the non-volatile data storage technologies, floating gate charge storage memory is becoming the dominant technology. Originally, conceived in 1967 by Kahng and Sze,1 the floating gate memory has evolved over the decades to higher performance and density, and is now a ubiquitous element of computing platforms. Another type of memory and logic element, the memristor was conceptualized, few years later in 1971, by Chua.2 The memristor is a very general concept which can be implemented with a large variety of physical mechanisms and potentially offers high density, low power, and nonvolatile operation, placing such devices among the candidates of advanced components of future computing systems. Reports of resistive switching in macroscopic and microscopic systems are older than their theoretical description; however, the concept of nanoscale memristive devices for use in memory and logic operations has attracted significant attention in the last decade.3–12

Previously, we have observed resistive switching based on electrochemistry in graphene and even in organic systems.13,14 Non-traditional and conventional computing architectures can be potentially implemented using memristive devices.15–17 One particularly important field is neuromorphic computing, where the nonvolatile operation of solid-state memristors is expected to find application as synaptic elements.18–21 In order to realize high-density, high-performance neuromorphic computing schemes based on memristive devices, it would be highly desirable to integrate such devices with well-established silicon technology. Although some of the systems that exhibit memristive behavior are potentially not compatible with silicon complementary metal-oxide-semiconductor (CMOS) technology, some readily are, and resistive switching has been observed even in silicon based materials.22–26

According to Waser and Aono, among the various mechanisms that result in resistive switching, memristive operation based on charge storage can be cited as a separate category.4 Charging based memristor operation has been observed more than 40 years ago by Simmons and Verderber,27 and the mechanism is loosely explained by the presence of carriers that are injected by Fowler–Nordheim tunneling at high electric fields and subsequently trapped at sites such as defects or metal nanoparticles in the insulator. The modification of the electrostatic barrier of the Metal-insulator-metal (MIM) structure results in changes of its resistance, similar to the changes in channel resistance in a Flash field-effect transistor (FET) upon charging of the floating gate. Resistive switching in metal nanoparticle doped organic conductor films has been recently studied for their memristive properties and applications, and can be considered in this category.28–30

In this article, we demonstrate that a thin film junctionless field-effect transistor with a floating charge trap layer also exhibits memristive behavior, when operated as a two-terminal device. The significance of this demonstration is that the presented mechanism is well understood, highly repeatable and scalable, CMOS compatible, and allows rational design, while exhibiting nearly ideal memristor characteristics. The demonstration connects the two families of devices at the forefront of contemporary and future computing, namely, the flash memory and the memristor.

― E-mail: aykutlu@unam.bilkent.edu.tr
The schematic description of the device is shown in Figure 1(a). A 15 nm thick atomic layer deposited (ALD) ZnO layer is used as the channel between the source and drain contacts made of aluminum. The channel is separated from a 5 nm thick charge storage layer (non-stoichiometric silicon nitride layer) by a 15 nm thick ALD HfO₂ tunnel barrier, followed by a 15 nm thick ALD Al₂O₃ control oxide on p-type silicon acting as the gate. The band alignment of the layers under flatband condition. The bandgaps and affinities of the layers are 

\[
E_g^{SiO_2} = 2.65 \text{ eV for HfO}_2, \quad E_g^{SiO_3} = 2.05 \text{ eV for Al}_2O_3, \quad E_g^{Si_N_2} = 5.3 \text{ eV for SiO}_2, \quad E_g^{Si_N_4} = 5.7 \text{ eV for Si}_{3}N_{4}, \quad E_g^{ZnO} = 3.3 \text{ eV for SiO}_2, \quad E_g^{SiO_3} = 3.3 \text{ eV for SiO}_2.
\]

Silicon rich (refractive index \(n_{SiN_4} = 1.98\)) Si_N₄ provides trap states for charge storage. (c) When the source-drain voltage \(V_{DS}\) is swept, a pinched hysteresis is observed in the channel current \(I_{DS}\). A rapid scan exhibits smaller hysteresis (d), whereas a slower scan exhibits larger hysteresis. The pinched hysteresis is attributed to channel conductivity modulation upon charging and discharging of the silicon nitride layer.

In order to understand and relate the memristor operation to device parameters, we resort to a simplified model of the device. In general, a memristor can be described by the constitutive equations as

\[
i = g(Q,v),
\]

(1)

\[
\frac{dQ}{dt} = h(Q,v),
\]

(2)

where \(g(Q,v)\) is the memductance and the evolution of the state variable \(Q\) is now written in terms of the applied voltage, with the dynamic relation defined by a function \(h(Q,v)\). In order to determine \(g(Q,v)\) and \(h(Q,v)\), the current-voltage relation of the junctionless n-channel depletion type device is required. The relation can be given by an extension of the square-law piecewise description of a field effect transistor

\[
I_{DS} = \begin{cases} 
\mu_{eff}C_{ox} \frac{W}{L} [(V_{GS} - V_{TH})V_{DS} - \frac{1}{2}V_{DS}^2] (1 + \lambda V_{DS}), & \text{if } V_{GS} - V_{TH} > V_{DS} > 0 \\
\frac{1}{2}\mu_{eff}C_{ox} \frac{W}{L} (V_{GS} - V_{TH})^2 (1 + \lambda V_{DS}), & \text{if } V_{DS} > V_{GS} - V_{TH} > 0 \\
0, & \text{if } V_{DS} > 0 > V_{GS} - V_{TH} \\
-\mu_{eff}C_{ox} \frac{W}{L} [(V_{GS} + |V_{DS}| - V_{TH})|V_{DS}| - \frac{1}{2}V_{DS}^2] (1 + \lambda |V_{DS}|), & \text{if } V_{GS} - V_{TH} > 0 > V_{DS} \\
-\frac{1}{2}\mu_{eff}C_{ox} \frac{W}{L} (V_{GS} + |V_{DS}| - V_{TH})^2 (1 + \lambda |V_{DS}|), & \text{if } 0 > V_{GS} - V_{TH} > V_{DS} \\
0, & \text{if } 0 > V_{DS} > V_{GS} - V_{TH}.
\end{cases}
\]

(3)

where \(V_{GS}\) is the gate-source voltage, \(V_{DS}\) is the source-drain voltage, \(V_{TH}\) is the threshold voltage, \(\mu_{eff}\) is the effective channel mobility, \(C_{ox}\) is the areal capacitance of the gate stack, \(w\) and \(L\) are the width and length of the channel, and \(\lambda\) is the Early parameter. The extension of \(I_{DS}-V_{DS}\) relation to cover the negative \(V_{DS}\) bias regime is achieved by noticing that, due to symmetry of the junctionless transistor in this regime, the negative drain voltage can be taken as the ground reference and source and gate voltages can be re-referenced with respect to the drain voltage. The model can be used to fit experimental data, as shown in Figure 2(a).

In this simplified model, \(V_{TH}\) is assumed to be constant through the length of the channel. This corresponds to the floating gate being at a single voltage level, which is a first order approximation only, due to the fact that the Si₃N₄ storage layer allows spatially varying charge density to be stored. In order to derive full memristive dynamics, it is needed to establish the relation between stored charge and threshold voltage. which is given by the following equation:

\[
V_{TH} = V_{TH0} - \frac{Q}{C_{con}}.
\]

(4)

Here, \(V_{TH0}\) is the native threshold voltage of the device having no stored charge. The charging model can be approximated by a lumped model, where Fowler-Nordheim tunneling between the channel and the storage layer is the only current...
leading to charging. Fowler-Nordheim current depends on the field through the tunnel barrier, \( F_{\text{tun}} \), which is described by the following equation:

\[
F_{\text{tun}} = \frac{Q}{(C_{\text{con}} + C_{\text{tun}})t_{\text{tun}}} + \frac{C_{\text{con}}}{(C_{\text{con}} + C_{\text{tun}})t_{\text{tun}}} \left( V_{\text{GS}} - \frac{1}{2} \xi V_{\text{DS}} \right).
\]

(5)

Here, \( C_{\text{con}} \) is the control oxide capacitance, \( C_{\text{tun}} \) is the tunnel oxide capacitance, \( t_{\text{tun}} \) is the tunnel oxide thickness, and \( \xi \sim 1 \) is a geometric correction factor, which accounts for the spatially varying field distribution along the channel length.

The stored charge dynamics is then approximately given by

\[
dQ/dt = -\text{sgn}[F_{\text{tun}}] \frac{n_e q^2 m F_{\text{tun}}^2}{16 \pi^2 m_{\text{tun}} \hbar V_B} \exp \left( -\frac{2V_B^{3/2} \sqrt{8m_{\text{tun}} q}}{3\hbar F_{\text{tun}}} \right),
\]

(6)

where \( q \) is the electronic charge, \( m \) and \( m_{\text{tun}} \) are the electron masses for the channel and the tunnel barrier, respectively, \( V_B \) is the barrier height, and \( n_e \) is the capture efficiency of tunneling electrons by the nitride storage layer.

The change in the sign of threshold voltage shift is further emphasized in Figures 3(a) and 3(b), where +5 V and −5 V voltage pulses of 10 s duration are applied to the gate (flash mode operation) and to the drain (flashristor mode operation) and small signal resistance at \( V_{\text{DS}} = 0.1 \) V is measured and plotted. The measurements also show that the written and erased states are highly repeatable, showing better than 5% variation (note the linear scale of the graphs).

The flashristor mode can also be used by assigning a common gate to multiple devices. In order to investigate the variation of the \( R_{\text{off}}/R_{\text{on}} \) ratio on gate voltage, we measure it by setting a fixed gate voltage (Figure 3(c)). It is observed that the greatest \( R_{\text{off}}/R_{\text{on}} \) ratio is achieved for \( V_{\text{GS}} \). This is attributed to the reduction of tunnel field during write or erase operation due to the asymmetry generated by a non-zero \( V_{\text{GS}} \). The repeatability of the \( R_{\text{off}}/R_{\text{on}} \) ratio prompts use of multilevel pulses to write different charges onto the device. A four level pulsing scheme is shown in Figures 3(d) and 3(e), demonstrating feasibility of multilevel operation.

The model presented in the above equations can be used to estimate the dynamic behavior of the device, as shown in Figures 4(a) and 4(b). During positive and negative voltage pulses applied to the drain or to the gate, the source-drain current is recorded. Using a calibration data set, which was acquired rapidly in order to avoid charging of the storage layer, the threshold voltage shifts are extracted, as shown in Figure 4(a). It is observed that, when the pulse is applied to \( V_{\text{GS}} \), a greater threshold voltage shift can be induced as compared to the same voltage applied to \( V_{\text{DS}} \) (keeping \( V_{\text{GS}} = 0 \)).
This is well explained by the theoretical model, as shown in Figure 4(b). Here, $\eta$, $V_B$, and $m_{\text{run}}$ are used as fitting parameters. One important point that should be noted is, through fitting, the capture efficiency $\eta$ is found to be a very small value ($10^{-4}$). Assuming the model properly describes the operation of the device, this small capture efficiency can be attributed to the small number of trap density in the silicon nitride layer and to the fact that HfO$_2$ barrier only limits half of the energy range below the silicon nitride conduction band, not prohibiting all of the deposited carriers from rapidly escaping back into the channel. Based on the fitted values, which quantitatively explains the observed threshold voltage shifts, the model is used to investigate the dependence of $R_{\text{off}}/R_{\text{on}}$ ratio on device parameters. Assum- ming an infinite off-on ratio for the transistor, it is observed that the threshold voltage of the uncharged device, $V_{\text{TH}0}$, has a strong effect on the achievable $R_{\text{off}}/R_{\text{on}}$ ratio for a given voltage range (Figure 4(c)). Theory predicts that as the charged threshold value becomes positive, the transistor is turned off and the off resistance becomes large. In this case, the $R_{\text{off}}/R_{\text{on}}$ ratio can also be made arbitrarily large. In our devices, the uncharged threshold voltage is measured to be around $-2$ V, and according to Figure 4(a) typical threshold voltage shifts are on the order of 0.5 V (for 2.5 V write voltage) to 1 V (for 5 V write voltage). Therefore, in our measurements, we do not achieve a completely turned off device, hence, the $R_{\text{off}}/R_{\text{on}}$ ratio remains low. In order to increase the threshold voltage shift, one can wait for a longer period with the voltage pulse turned on (Figure 4(c)). In our devices this is partly confirmed, a prolonged write/erase pulse duration is indeed observed to improve the $R_{\text{off}}/R_{\text{on}}$ ratio (Figure 4(d)).

In order to better understand the device operation and obtain a clue to why the carrier capture probability is found to be unexpectedly small in the fitting process (Figure 4(b)), we use electrostatic force microscopy (EFM) (Figures 5(a) and 5(b)). A multi-frequency version of electrostatic force microscopy is used to simultaneously measure topography and surface potential. A metal coated cantilever with a resonance frequency of 71 kHz and nominal spring constant of 2.7 N/m was used in tapping mode, while a sinusoidal voltage (0.5 V amplitude) was applied to the cantilever at a frequency of 11 kHz to avoid harmonics of electrostatic forces from interfering with the topography measurement. The electrostatic forces at 11 kHz were measured using a Stanford Research Systems Lock-in amplifier, SR830. Contact potential difference along the device is measured prior to biased measurements by grounding $V_{DS}$ and $V_{GS}$, and used as a baseline. Multiple electrostatic force maps acquired at different bias conditions were used to calibrate the system. It is seen that, for $V_{DS} = 5$ V and $V_{GS} = 0$ V, there are significant voltage drops at the source-channel and channel-drain contacts as shown in the voltage profile across the channel in Figure 5(c). As a result of the voltage drops, a significantly smaller voltage develops across the channel during operation of the device, reducing the tunnel oxide field. The tunnel rate (Eq. (6)) exponentially depends on the tunnel field, and therefore, the capture probability $\eta_c$ is estimated to be very small, due to an incorrect assumption about the tunnel field.

The performance of the device is low, considering that the write/erase times are on the order of 1 s, and retention is on the order of $10^6$ s. Improvements of the device can, in principle, be made by optimizing the storage layer and gate 

---

**Figure 4** (a) Experimentally measured threshold voltage shift for the ZnO channel (70 x 60 nm$^2$) device operated in the flashristor mode ($V_{DS}$ pulsed with $-0.5$ to $-2.5$ V, 0.5 to 2.5 V, and $V_{GS} = 0$ V) and flash mode ($V_{DS} = 0.1$ V and $V_{GS}$ pulsed with $-0.5$ to $-2.5$ V and 0.5 to 2.5 V). The threshold voltages are extracted by measuring $I_{DS}$ as a function of time and fitting to a data set describing transistor operation under various bias conditions. (c) Based on the charging model (Eqs. (4)–(6)), the charging behavior is simulated by using $\eta_c$, $V_B$, and $m_{\text{run}}$ as fitting parameters. The geometric correction parameter in Eq. (5) is found to be $\zeta = 1.4$. (d) Dependence of the $R_{\text{off}}/R_{\text{on}}$ ratio on the native threshold voltage $V_{\text{TH}}$ for write/erase voltages of $+5/-5$, $+4/-4$, and $+3/-3$ V. The simplified model predicts that, if enough charge can be stored in the floating gate region, the $R_{\text{off}}/R_{\text{on}}$ ratio can be made arbitrarily high. (e) According to the model, increasing write/erase durations also improve the $R_{\text{off}}/R_{\text{on}}$ ratio. (f) Experimental results show an increase in the $R_{\text{off}}/R_{\text{on}}$ ratio as a function of pulse duration.

**Figure 5** (a) Topography and (b) EFM data on the channel region shows that (c) there are significant voltage drops between the source-channel ($50 \times 45$ nm$^2$) and channel-drain ($50 \times 45$ nm$^2$) contacts. This non-ideal behavior is due to the large contact resistance in the unannealed contacts, and inhibits development of larger fields between the channel and the floating gate during write/erase cycles.
stack properties. Previously we have predicted that, in a flash memory, a large retention time accompanied with fast write/erase times are possible by using high-$\kappa$ dielectrics in the gate stack and nanocrystals in the storage layer. The sizes of the devices are also large for high density neuromorphic computing, however, there is no significant variation of threshold voltages and charging characteristics by changing the dimensions within the micrometer range. We therefore estimate that the devices can, in principle, be scaled into the nanoscale regime, unless short channel effects become significant. There are a large number of variations on flash memory devices and programming modes. Among the mechanisms used for writing into storage layers, Hot Electron Injection (HEI) is widely used to speed up write times. In the presented devices, the HEI are insignificant as the size of the device is much greater than the thermalization length of carriers. HEI can become more important as the devices are scaled down to the nanometer range. In conclusion, the flashristor mode may allow high density neuromorphic computing applications using the readily available flash memory technology, extending the use of charge storage memories to future computational architectures.

This work was partially supported by TUBITAK under Grant BIDEB 2218 to Ikram Orak.