A silicon nanocrystals based memory

Sandip Tiwari,[a] Farhan Rana, Hussein Hanafi, Allan Hartstein, Emmanuel F. Crabbé, and Kevin Chan
IBM Research Division, Thomas J. Watson Research Center, Yorktown Heights, New York 10598

(Received 13 September 1995; accepted for publication 2 December 1995)

A new memory structure using threshold shifting from charge stored in nanocrystals of silicon ($\approx$5nm in size) is described. The devices utilize direct tunneling and storage of electrons in the nanocrystals. The limited size and capacitance of the nanocrystals limit the numbers of stored electrons. Coulomb blockade effects may be important in these structures but are not necessary for their operation. The threshold shifts of 0.2–0.4 V with read and write times less than 100's of a nanosecond at operating voltages below 2.5 V have been obtained experimentally. The retention times are measured in days and weeks, and the structures have been operated in an excess of 10^9 cycles without degradation in performance. This nanomemory exhibits characteristics necessary for high density and low power. © *1996 American Institute of Physics.* [S0003-6951(96)00310-0]

The use of multiple confinement of free carriers has led to several significant observations of physical phenomena and demonstrations of usefulness in electronic and optoelectronic devices. The observations include consequences of one dimension of freedom in field-effect transistors[1,2] and heterostructures[3] and zero dimensions of freedom in tunnel diodes.[4,5] Important device demonstrations that utilize multidimensional confinement include wire lasers with continuous room temperature threshold currents of sub-200 μA[6] and a prototypical memory in poly-silicon.[7] This letter reports a memory that utilizes direct tunneling into three-dimensionally confined nanocrystal for producing bistability in the conduction of a transistor channel. It is fast, has quasi-nonvolatile characteristics, can be fabricated with a minimum perturbation of conventional silicon technology, and operates at room temperature.

Figure 1 shows a schematic cross section and band diagrams during injection (write cycle), storage (store) and removal (erasure) of an electron in the device. A thin tunneling oxide (1.1–1.8 nm thick) separates the inversion surface of an n-channel silicon field-effect transistor (FET) from a distributed film of nanocrystals of silicon that covers the entire surface channel region. A thicker tunneling oxide (4.5 nm or higher) separates the nanocrystals from the control gate of the FET. An injection of an electron occurs from the inversion layer via direct tunneling when the control gate is forward biased with respect to the source and drain. The resulting stored charge screens the gate charge and reduces the conduction in the inversion layer, i.e., it effectively shifts the threshold voltage of the device to be more positive, whose magnitude for a single electron per nanocrystal is approximately given by:

$$\Delta V_T = \frac{q n_{\mathrm{well}}}{\epsilon_{\mathrm{ox}}} \left(t_{\mathrm{cntl}} + \frac{1}{2} \frac{\epsilon_{\mathrm{ox}}}{\epsilon_{\mathrm{Si}}} t_{\mathrm{well}} \right). \quad (1)$$

Here, ΔV_T is the threshold voltage shift, t_{cntl} is the thickness of the control oxide under the gate, t_{well} is the linear dimension of the nanocrystal well, ϵ's are the permittivities, q is the magnitude of electronic charge, and n_{well} is the density of nanocrystals. For nanocrystals that are 5 nm in dimension, 5 nm apart, i.e., a nanocrystal density of 1×10^{12} cm^{-2}, with a control oxide thickness of 7 nm, the threshold shift is nearly 0.36 V for one electron per nanocrystal, a large value that in a good transistor changes the subthreshold current by five orders of magnitude. This is easily current-sensed. The removal of the charge stored in nanocrystals requires the lowering of energy at the channel surface with respect to that of the control gate. Leakage of the charge occurs either laterally by conduction between the nanocrystals or to the silicon surface with an eventual path to the n^+-doped source and drain regions. The former is constrained by the tunneling probability between nanocrystal islands; SiO_2 gap exceeding 4.5 nm suffices in making the direct tunneling current component to be negligible. The latter is constrained by the quiescent bias state of the device which leaves the surface in depletion. The band alignment is unfavorable for tunneling, and the charge movement within the bulk silicon requires diffusion of carriers. Thus, the nanomemory has a long retention time.

Coulomb blockade effect can be very significant at these dimensions. The Coulomb charging energy ($q^2/2C_{\mathrm{tt}}$, where C_{tt} is the nanocrystal capacitance) for a 5 nm diameter

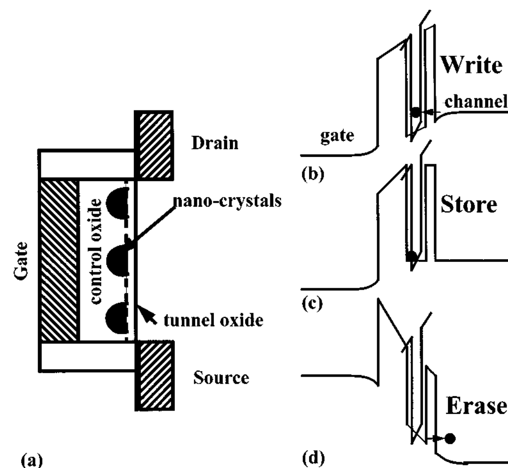

FIG. 1. A schematic cross section (a) and band diagram during injection (b), storage (c), and removal (d) of an electron from a nanocrystal.

[a]Electronic mail: tiwari@ibm.com

Downloaded¬04¬Mar¬2008¬to¬128.84.87.77.¬Redistribution¬subject¬to¬AIP¬license¬or¬copyright;¬see¬http://apl.aip.org/apl/copyright.jsp

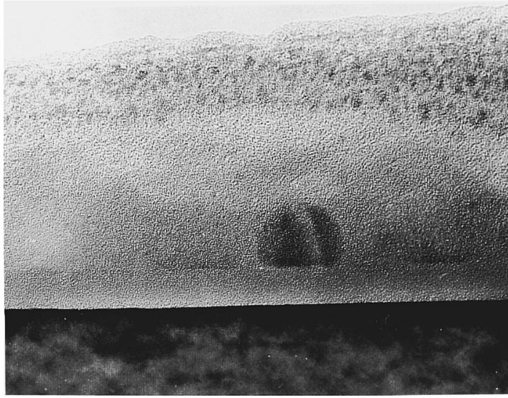

FIG. 2. A transmission electron micrograph showing silicon nanocrystals seeded on SiO_2 which is grown on silicon. The nanocrystal deposition is followed *in-situ* by a thicker control oxide. The micrograph also shows glue used for mounting samples. For reference, the tunneling oxide in this figure is 3 nm thick.

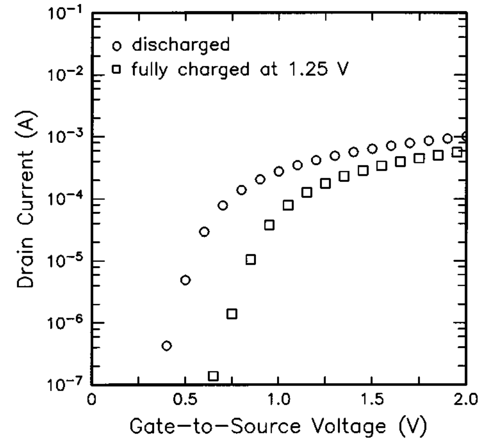

FIG. 3. A demonstration of the bistability in the structure as a result of injection of electrons into the nanocrystals. The nanomemory characteristics are measured with 20 ns gate voltage pulse.

nanocrystal of silicon in a matrix of SiO_2 is $\approx$74 meV. Not only is it larger than thermal energy, but it also limits additional injection of carriers into the nanocrystal. Thus, for a given write pulse, applied as a voltage signal between gate and source-drain, there is a saturation of a number of electrons that are accommodated in the nanocrystals, with each single electron causing a threshold shift of 0.36 V for this example. To write the charge in a short time, say 20 ns, requires a current density of 14 A/cm^2 through the tunneling oxide. With a gate voltage a volt, our theoretical and experimental results indicate a current density of 17 A/cm^2 at 1.5 nm oxide thickness dropping to about 1 A/cm^2 at 1.75 nm oxide thickness. The tunneling current is a very sensitive function of the thickness because of the large barrier height of the SiO_2/Si junction; a decade of change in the current results from a change in tunneling oxide thickness of 0.15 nm. However, it shows that if very thin oxide thicknesses can be achieved, a short time charging can be achieved while limiting the energy of carriers to less than the band gap of silicon. Thus, oxide hot carrier processes can be minimized during the tunneling process to achieve reliability. In a nanomemory of 0.18 μm effective channel length, with a 0.25 μm channel width, the total number of nanocrystals in our model example is $\approx$800, a fairly small number that allow achieving low power dissipation during write cycles.

The nanomemory structures have been fabricated at 0.4 μm and larger dimensions and within the constraints of silicon transistor technology. The critical new technique in the fabrication of the structures is the formation of the nanocrystals, which is accomplished through spontaneous decomposition during chemical vapor deposition. Figure 2 shows a cross-section view of the nanocrystals as deposited on the thermally grown tunneling oxide and followed by an *in-situ* coverage with deposited silicon dioxide. The unique attribute of this growth is the shape of the deposited nanocrystals. These are half spheres which expose the largest cross section to the tunneling source leading to the most efficient injection condition. Figure 3 shows characteristics in the presence and absence of electrons in the nanocrystals. In this structure, a device 0.4 μm in gate length and 20 μm in gate width, the drain current–gate voltage characteristics are shown with charging accomplished at 1.25 V. This example uses a 1.6 nm tunneling oxide, nanocrystals approximately 5 nm in size, a density of slightly less than 10^{12} cm^{-2}, and a control oxide of 7.3 nm. The characteristics are obtained using 20 ns pulses, applied between gate and grounded source, with the drain maintained at 100 mV. For the fully charged condition at 1.25 V, the gate is pulsed from the 1.25 V condition. The threshold shift of $\approx$0.25 V under 1.25 V static gate bias shows the effect of the storage of electrons which we estimate to be one electron per nanocrystal. Larger gate voltages leads to an increase in the number of stored electrons, e.g., in this same device structure, at 3 V static gate bias, the threshold shift is nearly 1.2 V, corresponding to larger number of electrons per nanocrystal.

The refresh times for the structures are very large, i.e., the stored charge does not leak out in accurately measurable times at room temperature. At elevated temperatures the time-constants are measurable and are summarized along with observed write times for charging to completion in Table I. The write-times are in the range of theoretical expectations, should scale exponentially with reduction in tunneling oxide thickness. The remarkable feature in the table is the hour or longer refresh times required by the nanomemory. The corollary to the long refresh time is an increase in erase times for the charge. Complete removal of the charge in the structures requires milliseconds at voltages of 3 V and above.

Direct tunneling is desired in the charging and discharging of the structure to prevent hot-carrier degradation. That this is successfully achieved is indicated by Figure 4 which shows the high and low threshold voltage in the presence and absence of charge as a function of number of write and erase cycles. The device of this structure, unlike that of Figure 3,

TABLE I. Extrapolated time-constants for measured structures. RTP stands for an oxide grown by rapid thermal processing.

Oxide	Write time constant	Refresh time constant
1.6 nm (Thermal)	100 ns	$>$ 1 wk (RT), 1 h (125 °C)
1.8 nm (Thermal)	100 ns	$>$ 1 wk (RT), $>$ 1 h (125 °C)
1.8 nm (RTP)	1 μs	$>$ 1 wk (RT), $>$ 1 h (125 °C)

Downloaded¬04¬Mar¬2008¬to¬128.84.87.77.¬Redistribution¬subject¬to¬AIP¬license¬or¬copyright;¬see¬http://apl.aip.org/apl/copyright.jsp

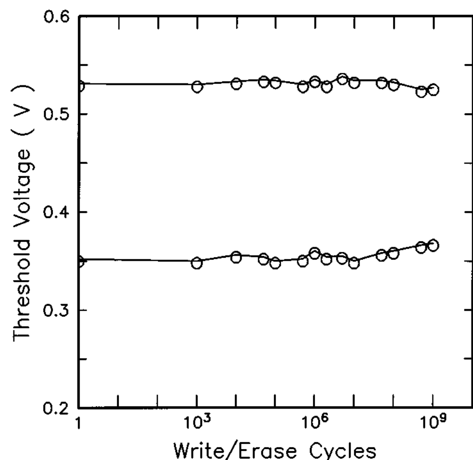

FIG. 4. Endurance characteristics of a positive threshold voltage nanomemory due to ± 2.5 V pulsing for writing ($1\,\mu$s) and erasing (1 ms) of the structure. The device in this test have a smaller nanocrystal density than that of Figure 3.

utilized a smaller nanocrystal density to limit the power dissipation. This is still a limited test, and cyclability exceeding 10^{13} is desired. The observation of a threshold for onset of degradation in thin oxides[8] suggests that low bias voltages in thin oxides ought to help in achieving high endurance. Localized field enhancement at an interface with two phases (Si and SiO_2) is known to cause enhanced electron injection. The dual electron injector structure of DiMaria *et al.*[9] utilizes this enhanced electron injection to achieve operation of an electrically-alterable memory at low voltages. Bisschop *et al.*[10] also show in poly-oxide structures, which have curved surfaces, an increase in injection current. The enhanced injection therefore is possibly a limiting mechanism in this nanocrystal memory structure through the total fluence. However, hemispherical grain structures have been employed in the formation of dynamic memory capacitors without catastrophic consequences and DiMaria *et al.* note an improvement due to localized field screening. Further experiments are required to truly show high endurance.

We have presented the theoretical basis and experimental operation of a nanomemory that uniquely takes advantage of Coulombic effects in presence of multidimensional confinement. A large bistability with a significant threshold voltage shift can be obtained at low bias voltages with in-built Coulomb blockade of additional carrier injection. The speeds in the structures are dependent on the tunneling oxide thickness and should scale to well below 100 ns for write time. The structures are quasi-non-volatile with refresh times in hours or more at 125 °C.

Over time, this work has been influenced and helped by a number of individuals. The early directions were influenced by the work of Dan DiMaria, and further refinements were helped by contributions of D. Buchanan and a number of people in the Silicon Laboratory.

[1] R. G. Wheeler, K. K. Choi, A. Goel, R. Wisnieff, and D. E. Prober, Phys. Rev. Lett. **49**, 1674 (1982).
[2] A. C. Warren, D. A. Antoniadis, and H. L. Smith, Phys. Rev. Lett. **56**, 1858 (1986).
[3] P. M. Petroff, A. C. Gossard, R. A. Logan, and W. Wiegmann, Appl. Phys. Lett. **41**, 635 (1982).
[4] T. A. Fulton and G. J. Dolan, Phys. Rev. Lett. **59**, 109 (1987).
[5] K. Hess, N. Holonyak, Jr., and T. A. Richard, Appl. Phys. Lett. **63**, 1408 (1993).
[6] S. Tiwari, G. D. Pettit, K. R. Milkove, R. Legoues, R. J. Davis, and J. M. Woodall, Appl. Phys. Lett. **64**, 3536 (1994).
[7] K. Yano, T. Ishii, T. Hashimoto, T. Kobayashi, F. Murai, and K. Seki, IEDM Tech. Dig. **1993**, 541.
[8] K. R. Farmer, M. O. Andersson, and O. Engstrom, Appl. Phys. Lett. **60**, 730 (1992).
[9] D. J. DiMaria, K. M. DeMeyer, and D. W. Dong, IEEE Electron Device Lett. **EDL-1**, 179 (1980).
[10] J. Bisschop, E. J. Korma, E. F. F. Botta, and J. F. Verwey, IEEE Trans. Electron Devices **ED-33**, 1809 (1986).

Downloaded¬04¬Mar¬2008¬to¬128.84.87.77.¬Redistribution¬subject¬to¬AIP¬license¬or¬copyright;¬see¬http://apl.aip.org/apl/copyright.jsp