NSF Grant Number: ECS-0304026
PI: Sanjay Banerjee, University of Texas at Austin
Co-PIs: J. Ekerdt, G.Hwang and L.F. Register

Spatially Ordered Self-assembled Quantum Dot Gate Low voltage/power, High Speed Nanoscale Flash Memories

The demand for high-density, low-cost, low-power/voltage, and high speed (programming, erase and read operations) semiconductor memory will lead to current non-volatile flash memories to proliferate into new approaches such as nanoparticle floating gate and single/few electron memories. This allows the tunnel oxide to be scaled down more aggressively because weakspots now erase only the nanoparticles above it. If conventional channel hot electron (CHE) programming is used for nanoparticle arrays, it leads to non-uniform programming only near the drain end. Subsequently, Fowler-Nordheim (FN) erasure into the source will not erase these nanoparticles. Generally, programming and erasure has been demonstrated in individual cells using uniform FN tunneling between the channel and nanoparticle array floating gate, but that may be impractical in flash memory cell arrays from an architecture and speed point of view.

Our research on nanoparticle floating gate flash memories has therefore focused on three areas to address the above issues [Fig.1]. We replace the SiO_2 gate dielectric with high-K dielectrics to allow for thinner equivalent oxide thickness (EOT) in order to maintain the high coupling capacitance between the floating gate to the external world, without sacrificing non-volatility, and to allow for lower-voltage and/or higher-speed operation and increased device lifetime. The high-k dielectrics have a lower bandgap, but are physically thicker than SiO_2 for the same EOT. At high fields corresponding to programming or erase, the bands for the high-k dielectric bend sufficiently to cause FN tunneling through a narrow triangular barrier, while for the higher bandgap SiO_2, it stays in the direct tunneling regime. This allows programming and erase for the high-k based cells at lower voltages, as confirmed theoretically and experimentally [Fig.2]. At the same time, under retention conditions, since the fields are low, the bandbending is less and the barriers are much wider for the high-k case, leading to reduced charge loss [Fig.3].

Secondly, we have investigated Si and SiGe(C) nanoparticle formation on the high-k dielectrics. Since SiGe has a lower bandgap than Si, it has deeper potential wells when embedded in the gate dielectric and hence longer retention times. Preliminary data seems to confirm this though there are anomalies at longer storage times. This may be because the charge storage is not in the nanoparticle itself but at traps at the interface between the nanoparticles and the dielectric, which may be deeper for SiGe than for Si. There are concerns that Si nanoparticles embedded in SiO_2 will have too high of a charging energy as the dots are reduced in size in scaled memory cells because of reduced dot capacitance. Embedding them in a high-k dielectric and using lower bandgap SiGe dots may mitigate this concern.

The third approach we have adopted to address the scalability issue of the flash cells is to use Channel Initiated Secondary Electrons (CHISEL) instead of simply CHE, with low bandgap SiGe “cold cathodes” embedded in the channel of the MOSFET to enhance hot programming currents at lower voltages and shorter channel lengths [4]. Monte Carlo simulations show that CHISEL programming is achieved more uniformly into the gate than CHE, which is important for the discontinuous nanoparticle floating gates. The use of SiGe low bandgap SiGe cold cathodes can enhance the gate currents by factors of ~5X compared to Si channels.

The use of SiGe nanoparticle floating gates on high-k gate tunneling dielectrics, with SiGe cold cathodes in the channel are ways to enhance the low voltage/power operation of flash

cells, improve the speed and charge retention. Control of dot sizes and spatial distributions may be improved by templated growth using block-copolymers or protein-assisted assembly [Fig. 4]. It may possible to exploit Coulomb blockade and multi-level storage in single electron/ few electron charge memories.

We have modeled the growth of oxide-embedded SiGeC nanocrystals, with a focus on (1) investigation of the underlying mechanisms of SiGeC nanocrystal formation in(on) an oxide matrix(surface) using a Continuous Random Network (CRN) model to construct disordered structures of semiconductors, oxides, and their interfaces. Using first principles Density Functional Theory (DFT) calculations we have identified the fundamental behavior and properties of the silicon-rich oxide systems including: i) the relative energy increase by the presence of Si^{+1}, Si^{2+}, and Si^{3+} atoms in silicon-rich oxide parts (so called sub-oxide penalty energy) plays a major role in the silicon-oxide phase separation, while strain effect turns out to be insignificant, ii) oxygen diffusion mainly contributes to the phase separation, rather than excess silicon, iii) the structures and diffusion pathways of oxygen and excess silicon atoms are identified, iv) the oxygen diffusion is a strong function of local strain and the charge state of three neighboring Si lattice atoms directly involved. Using Metropolis/Kinetic Monte Carlo (KMC) MC simulations based on fundamental data from first principles calculations, we have found that silicon-oxide phase separation by oxygen atom diffusion is mainly responsible for formation of silicon particles in sub-oxides [Fig. 5]. This is in direct contradiction to the widely-accepted Ostwald ripening model based on excess silicon diffusion and agglomeration. Our simulations demonstrate that i) silicon particles grow very rapidly at the early stage of thermal annealing in a coalescence-like fashion, and the growth slows down significantly when the particle density becomes low, ii) the size of silicon particles is a strong function of initial silicon supersaturation, and iii) the shape of silicon particles is significantly affected by their spatial distribution. These results are in excellent agreement with experimental observations, while the Ostwald ripening theory cannot capture the growth behavior.

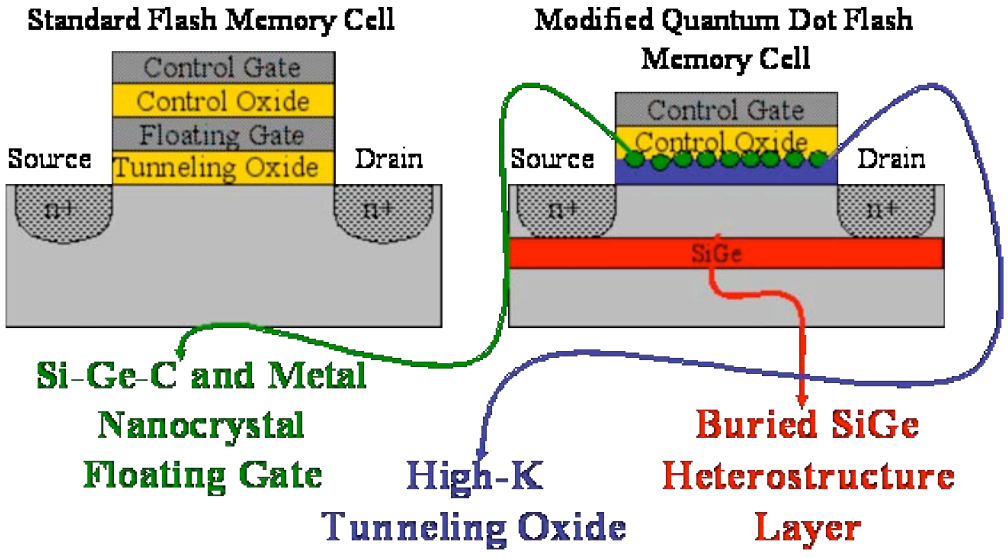

Fig. 1 Use of SiGe quantum dot floating gates with high-k tunnel dielectrics in non-volatile flash memories. SiGe "cold cathodes" in susbstrate can enhance the hot carrier programming currents.

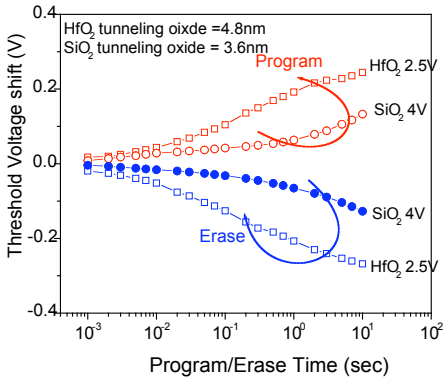

Fig.2. Lower voltage program and erase of flash memory using SiGe self-assembled nanocrystals and high-k versus SiO_2 gate dielectric because of lower bandgaps and thicker barriers.

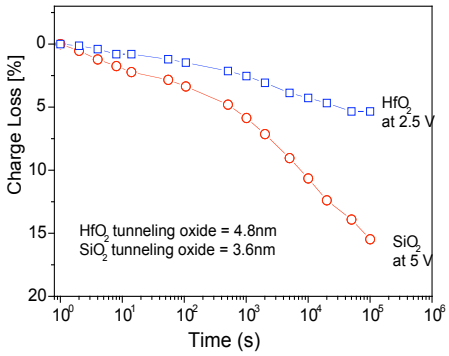

Fig. 3. Retention characteristics of HfO_2 and SiO_2 tunneling dielectric memories with SiGe nanocrystals after an initial +2.5 V and 5 V programming.

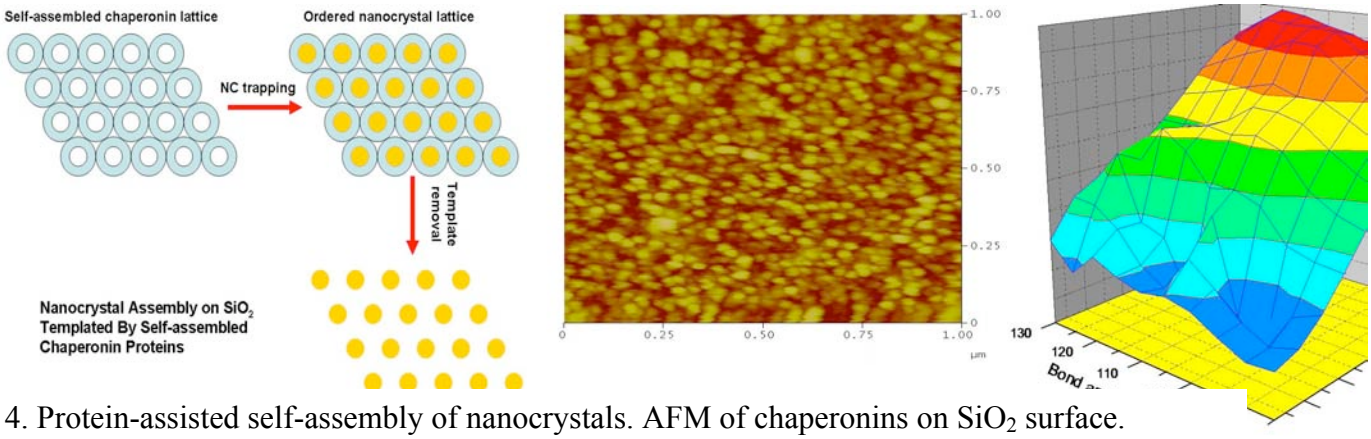

Fig. 4. Protein-assisted self-assembly of nanocrystals. AFM of chaperonins on SiO_2 surface.

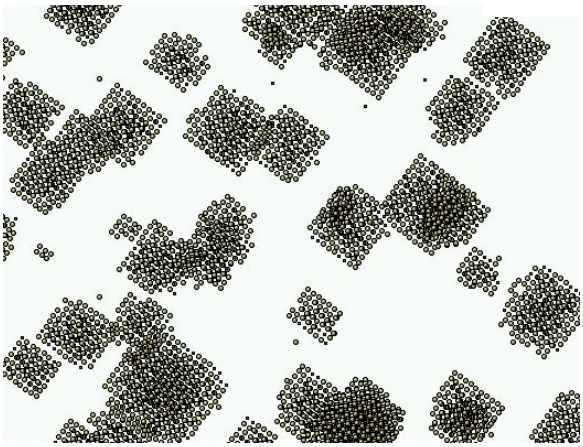

Fig. 5 Modeling of Si nanoparticle growth in SiO_2 using DFT and Kinetic Monte Carlo to study oxygen diffusion in oxide.