

Resonant Fowler-Nordheim Tunneling through Layered Tunnel Barriers and its Possible Applications

Alexander Korotkov and Konstantin Likharev
State University of New York, Stony Brook, NY 11794-3800, U.S.A.

Abstract

We have analyzed electron transfer through layered, symmetric, "crested" tunnel barriers, including the effect of resonant Fowler-Nordheim tunneling via electron subbands formed at the layer interfaces. The high sensitivity of the current to an applied electric field implies that the barriers may be used in terabit-scalable, nonvolatile, fast, bit-addressable memory (NOVORAM), as well as for ultradense, electrostatic data storage (ESTOR) and electronic cooling.

Introduction

Electric-field-induced ("Fowler-Nordheim") tunneling is the basic process used for writing and erasing data in electrically-alterable floating gate memory cells (1). Unfortunately, in standard, uniform tunnel barriers (Fig. 1a) this process is not sufficiently sensitive to applied voltage (see thin lines in Fig. 2). This is why such barriers do not allow the standard 10-year retention time to be combined with sub-microsecond write/erase, at least for the maximum voltages (< 12 V) which guarantee reasonable endurance under electric stress. The goal of this report is to show that the charge injection may be sped up significantly by using profiled ("crested") tunnel barriers with a potential maximum in the middle. As a result of such speed-up and inherent better scalability, floating memories using crested barriers may be able to compete with DRAM for bit-addressable applications.

Crested Tunnel Barriers

The relatively low dependence of the uniform barrier transparency on the electric field is due to the fact that the highest part of the barrier, closest to the electron source, is only weakly affected by the applied voltage: $U_{\max}(V) \approx U_{\max}(0)$ - see the dashed line in Fig. 1a. On the other hand, if the barrier potential peaks in the middle (see, e.g., Fig. 1b), its height is rapidly suppressed by the applied voltage: $U_{\max}(V) \approx U_{\max}(0) - eV/2$. As a result, the barrier transparency grows much faster with voltage.

In order to ensure acceptable retention time in nonvolatile memory applications at room temperature, U_m should be of the order of 3 eV, so that the most suitable crested barrier materials are wide-bandgap semiconductors. Unfortunately, no good shallow acceptors have been found for these

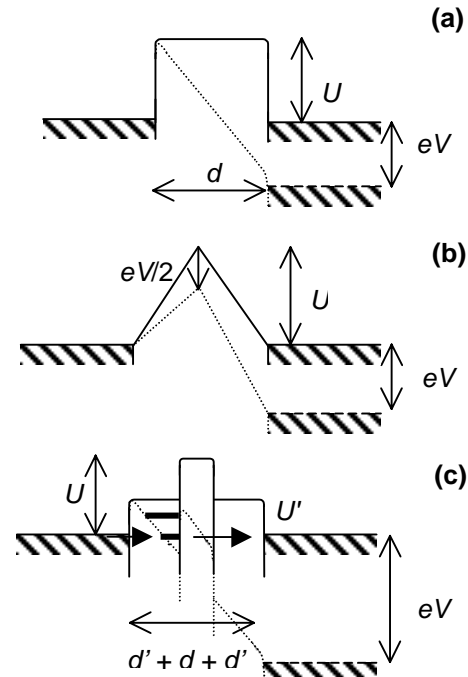


Fig. 1. Conduction band edge diagrams for various tunnel barriers without (solid lines) and with (dashed lines) applied electric field: (a) traditional, uniform barrier; (b) idealized crested barrier; and (c) trilayer crested barrier with resonant tunneling (bold arrows) through subbands which are naturally formed in the quantum well at the layer interface.

materials, which makes the implementation of the graded barriers (Fig. 1b) using modulation doping (2) a problem. A gradual change in the layer composition during its growth, which allows another natural implementation of graded barriers in III-V materials (3, 4), is also hardly feasible for wide-bandgap semiconductors.

Fortunately, there is another possible solution to this problem, which seems much more practical (5, 6): the triangular barrier (Fig. 1b) may be reasonably well approximated by the step-like potential formed in a trilayer crested barrier (Fig. 1c). The thick lines in Fig. 2 show a typical result of calculation of current through the trilayer barrier with parameters corresponding to the published data for the following materials: n^+ -Si for both electrodes, Si_3N_4 for two external layers, and AlN for the internal layer. One can see that the barriers may combine a 10-year retention time (at applied voltage V below $V_1 = 5.3$ V) with a sub-10-ns write/erase time (at $V > V_2 = 9.6$ V).

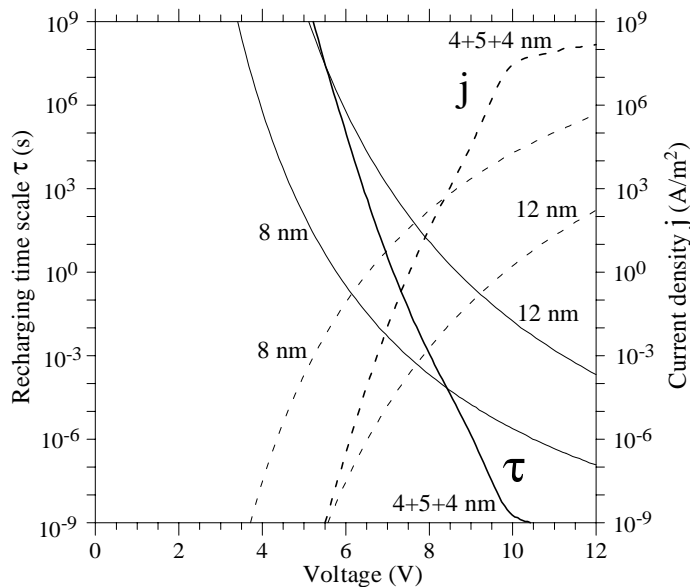


Fig. 2. Tunneling current density j (in A/m^2 , dashed lines) and the floating gate recharging time scale $\tau = CV/j$ (in seconds, solid lines) for the barriers shown in Figs. 1a and 1c, as functions of applied voltage V , calculated using the quasiclassical theory. Thin lines: uniform barriers with parameters corresponding to $n^+Si/SiO_2/n^+Si$ ($U = 3.2$ eV, $m = 0.3 m_0$, $d = 8$ nm and 12 nm). Thick lines: trilayer crested barrier with parameters corresponding to $n^+Si/Si_3N_4/AlN/Si_3N_4/n^+Si$ ($U = 2.0$ eV, $m' = 0.2 m_0$, $\epsilon' = 7.5$, $d' = 4$ nm; $U = 3.6$ eV, $m = 0.48 m_0$, $\epsilon = 8.5$, $d = 5$ nm).

Resonant Tunneling

An important ingredient of the physics responsible for the fast growth of the crested barrier transparency is the resonant tunneling of electrons through subbands which are naturally formed at the interface between the inner and one of the outer layers when an external field is applied – see Fig. 1c. In our calculations the tunneling has been calculated self-consistently, taking into account the electric field of the electrons in the subbands (8). The only important physical assumption in the theoretical treatment was that of conservation of the electron momentum component along the layers. This assumption is not valid if the electron scattering on interface roughness and/or impurities is too high. It is important to notice, however, that if the resonant tunneling is disregarded altogether (which may give a reasonable presentation of the effects at strong scattering), theory still shows almost the same degree of barrier sensitivity to the field, though at a different external layer thickness (5).

NOVORAM

If our predictions of the high sensitivity of crested barriers to applied voltage are confirmed experimentally, this effect may create unique opportunities for nonvolatile memories.

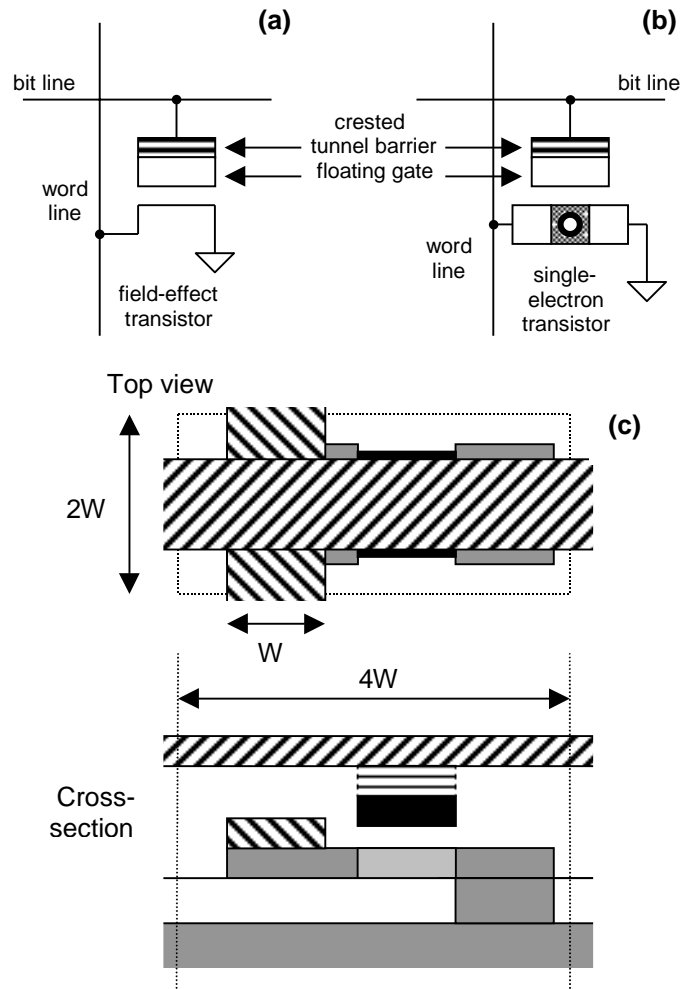


Fig. 3. Schematics of (a) NOVORAM and (b) SET/FET hybrid memory cells and (c) a possible layout of the NOVORAM cell using a SOI MOSFET.

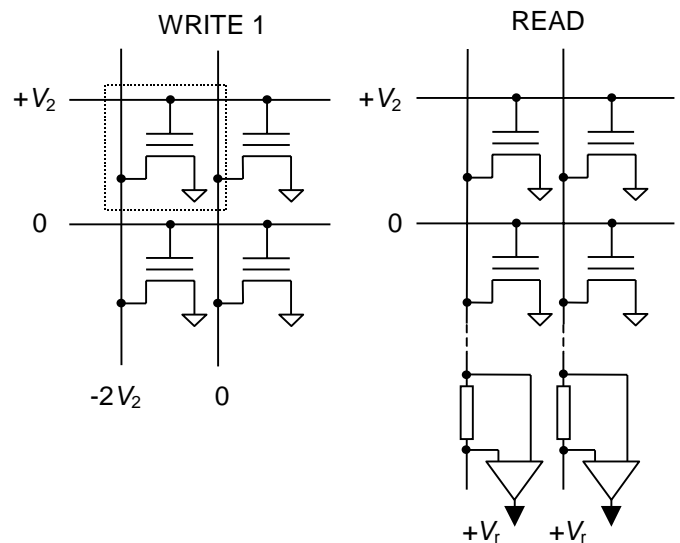


Fig. 4. NOVORAM: structure and basic operations. (For WRITE 0, the applied voltages are equal but opposite in sign to those shown for WRITE 1). Dashed line shows the selected cell. $V_r \ll V_2$.

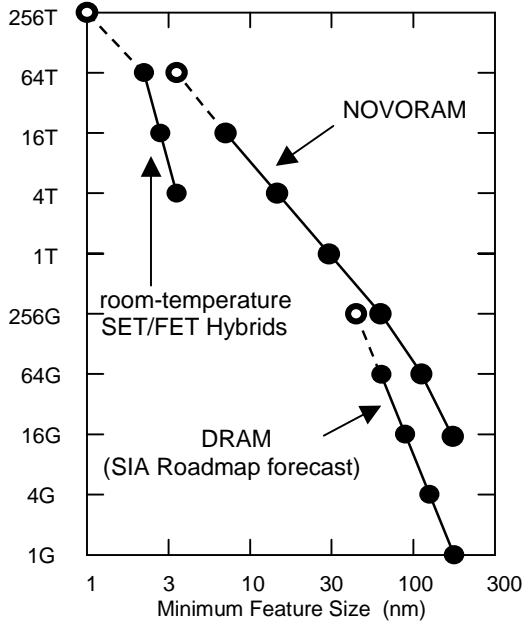


Fig. 5. Estimated integration scale of various fast random access memories as a function of minimum feature size. Dashed lines show anticipated scaling problems. The die size projections accepted by the SIA Roadmap (16) have been used to translate the calculated memory density to integration scale.

First, sub-10-ns write/erase time at low maximum field (below 8 MV/cm, see Fig. 2) may allow nonvolatile, random-access memories (NOVO RAM) with practically unlimited endurance. Second, since the ratio V_2/V_1 of the write/erase voltage to retention voltage is below 2 (Fig. 2), NOVO RAM can have a very simple, single-transistor cell structure (Fig. 3a) and architecture (Fig. 4), with NOR-type readout and without the necessity for disturb effect compensation (1). Third, the simple cell structure may result in a very small cell area (about $8W^2$, where W is the minimum feature size), physically scaleable all the way down to $W \sim 7$ nm, restricted only by the limits of SOI MOSFET scaling (9, 10) – see Fig. 5. We believe this may allow NOVO RAM to compete with DRAM for the whole bit-addressable memory market.

Hybrid SET/FET Memory

Even higher density may be achieved using the hybrid SET/FET memory (11). A low-temperature prototype of a single cell of such memory has already been successfully tested by a NEC-Tsukuba group (12). Physically this cell (Fig. 3b) is very similar to that of NOVO RAM (Fig. 3a), except that the field-effect transistor is replaced by a single-electron transistor – SET (13-15). However, the operation mode of this memory is very specific, in order to circumvent a common drawback of single-electron devices, the random background charge problem (14, 15). In this mode the readout is combined with “write 1” operation and is destructive, so that the memory needs a writeback operation

similar to refresh in DRAM (11). Otherwise, write/erase operations are similar to those in NOVO RAM.

Although the SET/FET memory was proposed some time ago (11), it may be made practical only if it uses crested barriers which enable high operation speed, high endurance, and low disturb effect. Unfortunately, due to the high sensitivity of single-electron charging effects to thermal fluctuations (14, 15), the maximum operating temperature of this memory reaches 300 K only when the minimum feature size reaches ~ 3 nm – see Fig. 5. Though fabrication of VLSI circuits with such design rules is a very distant prospect indeed, the bridge over the 3-to-30-nm gap provided by NOVO RAM (Fig. 5) may eventually lead to the implementation of this opportunity.

ESTOR

One more idea which may become a reality due to the use of crested barriers is ultradense electrostatic data storage – ESTOR. Fig. 6 shows a possible system for such storage (5, 17). A read/write head (tip) is flown over a substrate with the crested barrier separating a conducting (ground) layer and a layer of nanometer-size metallic grains, not necessarily all of the same size or shape. The binary unity is coded by the few-electron charging of a small group of grains. Write 1 operation is achieved by the application of a sufficiently high voltage to both inputs of the circuit on the tip. The voltage creates a local electric field which suppresses the tunnel barrier under a group of grains and pulls electrons from the ground electrode into the grains.

The recorded data may be read out by the application of the opposite voltages to the inputs. This voltage biases the

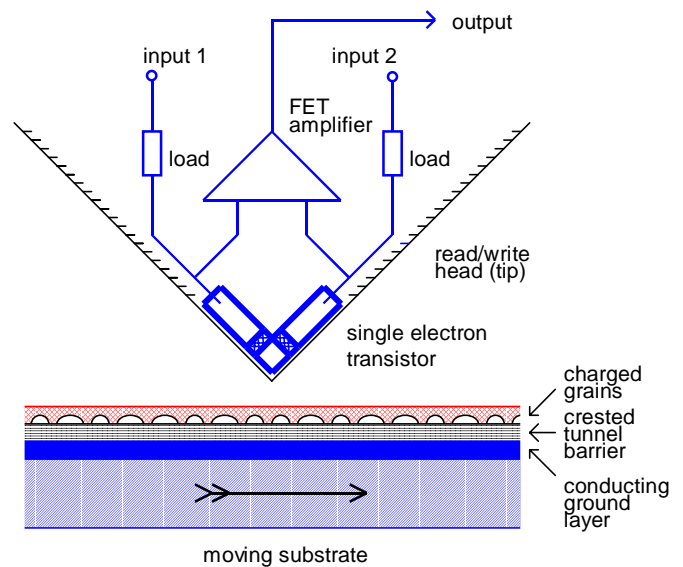


Fig. 6. The basic idea of electrostatic storage (ESTOR) – see the text.

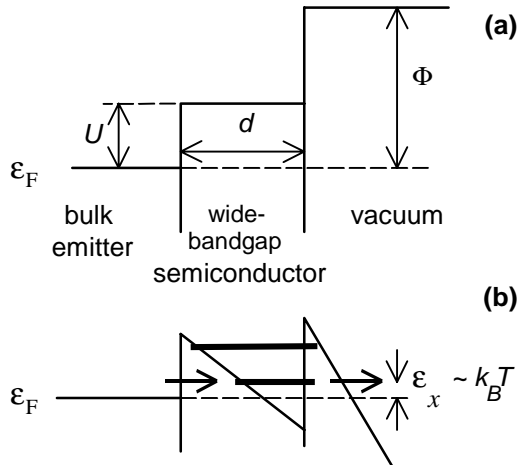


Fig. 7. The band edge diagram of the proposed structure for electronic cooling using resonant Fowler-Nordheim tunneling to vacuum: (a) in the absence of the external electric field and (b) at the field which aligns the lowest resonant subband with the hottest electrons in the cathode, leading to their effective removal.

single-electron transistor which is extremely sensitive to an electric field, in this case created by the grain charge. The SET output signal is further amplified by a closely located, FET-based sense amplifier and then sent out. Recent experiments (18) may be considered as the first step toward application of such readout.

Preliminary estimates (5) show that electrostatic recording may provide data storage density beyond 1 Tbits/in², i.e., about two orders of magnitude higher than the presently demonstrated magnetic recording density, provided that the read/write head can be flown at a comparable height (~20 nm) above the substrate surface. In contrast to earlier approaches to electrostatic data recording, the use of crested tunnel barriers may make possible a write/read speed up to 1 Gb/s per channel, which seems adequate even for this unparalleled bit density.

Electronic emission cooling

In addition to the nonvolatile memory and data storage systems, resonant Fowler-Nordheim tunneling may find one more application: for electronic cooling (8). Fig. 7 shows the necessary energy profile which is even simpler than those discussed above and may be provided by just one thin (a few nm) semiconductor film deposited on a conducting bulk substrate. If such a structure is placed in vacuum (necessary to avoid the phonon backflow) and biased by an appropriate electric field, the intensive resonant tunneling of the hottest electrons from the cathode may lead to its effective cooling. Estimates (8) show that heat removal density up to 30 W/cm² at 100K should be possible, even taking into account a moderate roughness of the deposited film.

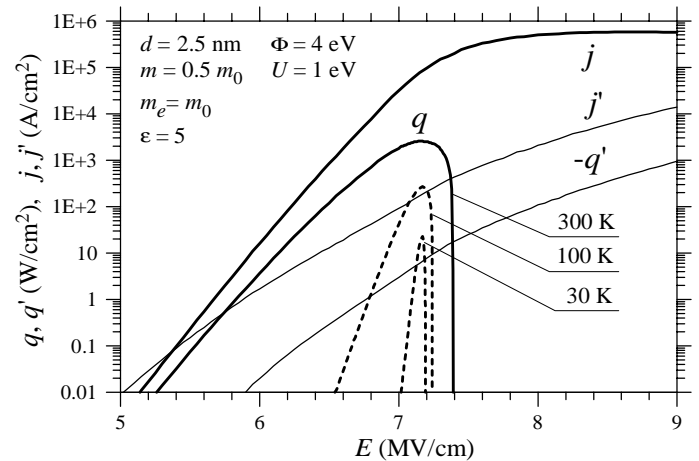


Fig. 8. Calculated density of resonant current j , the removed heat q (for several cathode temperatures), non-resonant current j' , and the related cathode heating q' , as functions of the applied electric field.

Acknowledgment

This work was supported in part by DARPA within the framework of the ULTRA and AME programs (via ONR), and the HERETIC program (via ARO).

References

- (1) *Nonvolatile Semiconductor Memory Technology*, W.D. Brown and J.E. Brewer, Eds., New York: IEEE Press, 1998.
- (2) R.J. Malik, T.R. AuCoin, R.L. Ross, K. Board, C.E.C. Wood, and L.F. Eastman, *Electron. Lett.*, vol. 16, pp. 836-837, 1980.
- (3) C.L. Allyn, A.C. Gossard, and W. Weigmann, *Appl. Phys. Lett.*, vol. 36, pp. 373-376, 1980.
- (4) F. Capasso, F. Beltram, R.J. Malik, and J.F. Walker, *IEEE Electron. Dev. Lett.*, vol. 9, pp. 377-379, Aug. 1988.
- (5) K. Likharev, *Appl. Phys. Lett.*, vol. 73, pp. 2137-2139, Oct. 1998.
- (6) The effect of speeding electron transfer in layered barriers with stepped profile has been noticed long ago (4, 7). However, asymmetric barriers considered in these works could not maintain high speed for both write and erase operations.
- (7) D. J. DeMaria, *J. Appl. Phys.*, vol. 50, pp. 5226-5228, Sep. 1979.
- (8) A. N. Korotkov and K.K. Likharev, *Appl. Phys. Lett.*, vol. 75, Oct. 1999.
- (9) F. Pikus and K. Likharev, *Appl. Phys. Lett.*, vol. 71, pp. 3661-3663, Dec. 1997.
- (10) Y. Naveh and K. Likharev, "Silicon MOSFET shrinking limits: A quantitative study", accepted for presentation at the 1999 ISDRS, Charlottesville, VA, Dec. 1999.
- (11) K. K. Likharev and A. N. Korotkov, in *Proc. of 1995 ISDRS*, Charlottesville, VA: U. Virginia, 1995, pp. 355-359.
- (12) C.D. Chen, Y. Nakamura, and J.S. Tsai, *Appl. Phys. Lett.* vol. 71, pp. 2038-2040, Oct. 1997.
- (13) K. K. Likharev, *IEEE Trans. Magn.*, vol. 23, pp. 1142-1145, March 1987.
- (14) K. K. Likharev, *Proc. IEEE*, vol. 87, pp. 606-632, Apr. 1999.
- (15) A. N. Korotkov, *Int. J. Electron.*, vol. 86, pp. 511-547, May 1999.
- (16) *International Technology Roadmap for Semiconductors*, 1998 Update, available on the Web at <http://www.itrs.net/itrs/itrs.nsf>.
- (17) A.N. Korotkov and K.K. Likharev, in *Int. Workshop on Computational Electronics Abstracts*, Tempe, AZ, 1995, p. 42; *VLSI Design (Amsterdam)*, vol. 3, pp. 341-344, 1998.
- (18) M.J. Yoo, T.A. Fulton, H.F. Hess, R.L. Willett, L.N. Dunkleberger, R.J. Chichester, L.N. Pfeiffer, and K.W. West, *Science*, vol. 276, pp. 579-582, Apr. 1997.