



Design optimization of multi-barrier tunneling devices using the transfer-matrix method

A. Gehring^{a,*}, T. Grasser^a, B.-H. Cheong^b, S. Selberherr^a

^a Institute for Microelectronics, TU Wien Gusshausstrasse 27–29, A-1040 Vienna, Austria

^b Computational Science Engineering Lab, Samsung Advanced Institute of Technology P.O. Box 111, Suwon 440-600, Republic of Korea

Received 9 December 2001; received in revised form 2 February 2002; accepted 7 February 2002

Abstract

We present simulations of a recently published multi-barrier phase-state low electron device memory cell. For the proper consideration of tunneling through the insulating barriers we implemented a one-dimensional Schrödinger solver based on the transfer-matrix formalism into the device simulator MINIMOS-NT. We investigate the effect of barrier size and position on the $I_{\text{on}}/I_{\text{off}}$ ratio of the memory cell. We find that the position and thickness of the central shutter barrier can be used for device tuning. For high $I_{\text{on}}/I_{\text{off}}$ ratios the central shutter barrier (CSB) should be placed near the upper contact. Furthermore, a reduction in the stack width leads to increasing $I_{\text{on}}/I_{\text{off}}$ ratios. Although the use of the transfer-matrix method in a device simulator requires a number of assumptions, it turns out to be a viable tool for deepening the understanding of tunneling effects in devices where other tunneling models fail.

© 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Device simulation; PLEDM; Tunneling; Quantum effects

1. Introduction

Multi-barrier tunneling devices propose a possibility to boost the density and performance of non-volatile memory cells. Phase-state low electron device memory (PLEDM) cells have been presented by Nakazato et al. [1] and promising results have been reported [2–5]. The principle of a PLEDM is to put a PLED transistor (PLEDTR) on top of the gate of a conventional MOSFET, see Fig. 1. The charge on the memory node is provided by tunneling of carriers through a stack of Si_3N_4 barriers sandwiched between layers of intrinsic silicon. Upper and lower barriers prevent diffusion from the poly-Si contacts, while the middle barrier blocks tunneling current in the off-state. In the on-state the height of the energy barriers is heavily reduced by the voltage on the word line, causing tunneling current to

flow at the interface to the side gate oxide. Since the charge on the memory node is used to control the MOSFET transistor, the cell has gain and only a small amount of charge has to be added to or removed from the memory node to change the state of the memory cell. The purpose of this paper is to investigate the effects of device design related issues on the performance of such PLEDTR-based memory cells.

2. Tunneling model

The key problem when simulating multi-barrier tunneling devices like the PLEDM cell is that the charging and discharging currents are pure tunneling currents. In common device simulators like DESSIS [6] or MEDICI [7] tunneling is usually taken into account by a Fowler–Nordheim analytical formula or a more sophisticated WKB or Gundlach approximation [8]. However, those models are based on approximations of the tunneling coefficient for triangular or trapezoidal barriers. They are not able to handle energy barriers

* Corresponding author. Tel.: +43-1-58801-36016; fax: +43-1-58801-36099.

E-mail address: gehring@iue.tuwien.ac.at (A. Gehring).

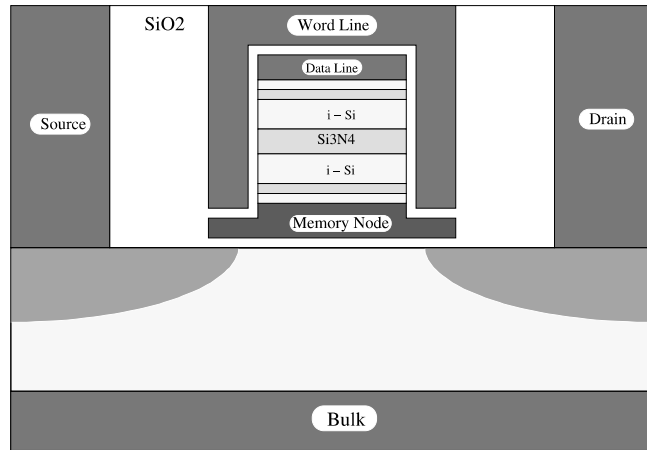


Fig. 1. Structure of the PLEDM cell.

which are of arbitrary shape. Additionally, the effect of resonances due to quasi-bound states can only be reproduced within the Gundlach model, but again only for trapezoidal barriers [9]. Thus, the PLEDM device needs a more rigorous approach, including the solution of the Schrödinger equation in the barrier region. Such a solution can be found using the transfer-matrix method. This formalism is based on the work of Tsu and Esaki on resonant tunneling diodes, see for example [10]. Descriptions can also be found in [11–13]. The main principle is to replace an energy barrier of arbitrary shape by a series of rectangular energy barriers as shown in Fig. 2. A modified transfer-matrix method where the barrier is

replaced by a piecewise linear (instead of constant) potential has also been presented [14]. Following the work of Tsu and Esaki, the tunneling current can be written as

$$J_t = \frac{4\pi m q k_B T}{h^3} \int_{E_{\min}}^{\infty} \text{TC}(E) \ln \left[\frac{1 + \exp\left(\frac{E_{f,1} - E}{k_B T}\right)}{1 + \exp\left(\frac{E_{f,2} - E}{k_B T}\right)} \right] dE \quad (1)$$

where $E_{f,1}$ and $E_{f,2}$ denote the Fermi levels in the data line and the memory node, respectively. The effective electron mass is denoted by m , T is the absolute temperature, and k_B the Boltzmann constant. The integra-

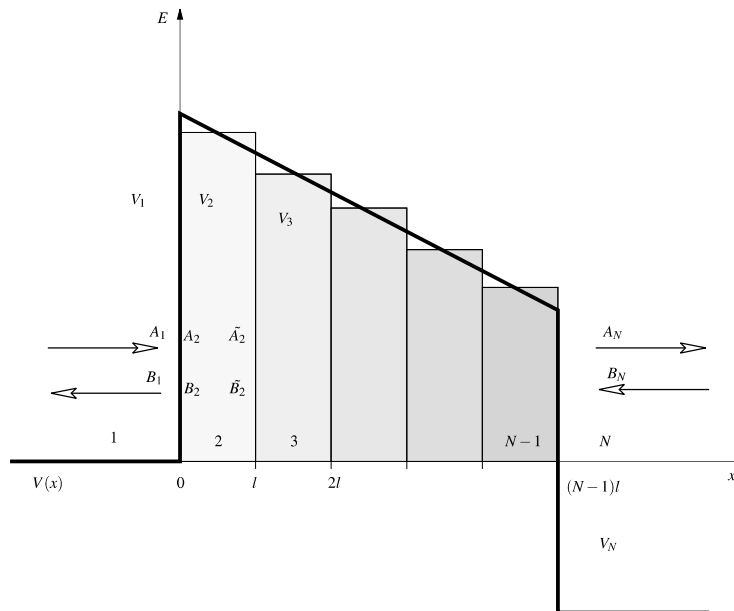


Fig. 2. The transfer-matrix method.

tion is performed starting from the higher of the two conduction band edges in the data line and the memory node, cf. Fig. 1. The energy barrier is divided into several regions $i = 1, 2, \dots, n$ with constant potential V_i . The wave function in each region $\Psi_i(x)$ is written as the sum of an incident and a reflected wave, with A_i and B_i being their amplitudes, and k_i the complex wave number

$$\Psi_i(x) = A_i \exp(jk_i x) + B_i \exp(-jk_i x) \quad (2)$$

$$k_i = \frac{\sqrt{2m_i(E - V_i)}}{\hbar}. \quad (3)$$

The amplitudes of the wave functions A_i and B_i , the carrier mass in the layers m_i and the potential V_i are constant for the region i . The boundary conditions for energy and momentum conservation

$$\Psi_i(x-) = \Psi_{i+1}(x+) \quad (4)$$

$$\frac{1}{m_i} \frac{d\Psi_i(x-)}{dx} = \frac{1}{m_{i+1}} \frac{d\Psi_{i+1}(x+)}{dx} \quad (5)$$

yield relations between the wave function amplitudes in region n and $n + 1$

$$A_i + B_i = A_{i+1} + B_{i+1} \quad (6)$$

$$\frac{k_i}{m_i} A_i - \frac{k_i}{m_i} B_i = \frac{k_{i+1}}{m_{i+1}} A_{i+1} - \frac{k_{i+1}}{m_{i+1}} B_{i+1} \quad (7)$$

which lead to

$$(A_{i+1} B_{i+1}) = \frac{1}{2} \begin{pmatrix} 1 + \alpha_i & 1 - \alpha_i \\ 1 - \alpha_i & 1 + \alpha_i \end{pmatrix} \begin{pmatrix} A_i \\ B_i \end{pmatrix} \quad (8)$$

with

$$\alpha_i = \frac{k_i}{k_{i+1}} \frac{m_{i+1}}{m_i}. \quad (9)$$

This equation holds only for $x = 0$. For the other interfaces we have to account for the distance l in the wave function amplitudes. The wave function $\tilde{\Psi}_{i+1}(x)$ at the right edge of layer $i + 1$ is

$$\tilde{\Psi}_{i+1}(x) = \Psi_{i+1}(l_{i+1}) = \exp(jk_{i+1} l_{i+1}) \Psi_{i+1}(0), \quad (10)$$

we thus write an expression for the wave function amplitudes at the right edge of layer $i + 1$:

$$\begin{pmatrix} \tilde{A}_{i+1} \\ \tilde{B}_{i+1} \end{pmatrix} = \frac{1}{2} \begin{pmatrix} \gamma_{i+1} & 0 \\ 0 & \gamma_{i+1}^{-1} \end{pmatrix} \begin{pmatrix} A_{i+1} \\ B_{i+1} \end{pmatrix} \quad (11)$$

with

$$\gamma_i = \exp(jk_i l_i). \quad (12)$$

If we define the matrices T_i and C_i as

$$T_i = \frac{1}{2} \begin{pmatrix} 1 + \alpha_i & 1 - \alpha_i \\ 1 - \alpha_i & 1 + \alpha_i \end{pmatrix} \quad (13)$$

$$C_i = \begin{pmatrix} \gamma_i & 0 \\ 0 & \gamma_i^{-1} \end{pmatrix} \quad (14)$$

and perform the above computation for each layer, we arrive at an expression for the wave function amplitudes in region n

$$\begin{pmatrix} A_n \\ B_n \end{pmatrix} = T_{n-1} C_{n-1} \cdot T_{n-2} C_{n-2} \cdot \dots \cdot T_2 C_2 \cdot T_1 \cdot \begin{pmatrix} A_1 \\ B_1 \end{pmatrix}. \quad (15)$$

If assumed that there is no reflected wave in region n and the amplitude of the incident wave is unity, we can write

$$\begin{pmatrix} A_n \\ 0 \end{pmatrix} = \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix} \begin{pmatrix} 1 \\ B_1 \end{pmatrix} \quad (16)$$

or, looking at the inverse matrix M :

$$\begin{pmatrix} 1 \\ B_1 \end{pmatrix} = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \begin{pmatrix} A_n \\ 0 \end{pmatrix} \quad (17)$$

with

$$T = M^{-1} = \frac{1}{\det M} \begin{pmatrix} M_{22} & -M_{12} \\ -M_{21} & M_{11} \end{pmatrix}. \quad (18)$$

It can be shown that $\det M = 1$, $M_{12} = M_{21}^*$ and $M_{11} = M_{22}^*$. In quantum mechanics the transmission coefficient (TC) is given as the ratio of outflowing to inflowing flux (see, for example, [15] or [13]). The flux is defined as

$$f = \frac{\hbar}{2jm} (\Psi^* \cdot \nabla \Psi - \nabla \Psi^* \cdot \Psi) \quad (19)$$

and has the unit (m/s). If we look at the ratio of the fluxes f_n and f_1 due to an incident wave $\Psi_1(x) = A_1 \times \exp(jk_1 x)$ and a transmitted wave $\Psi_n(x) = A_n \exp(jk_n x)$, we get the TC as

$$TC = \frac{f_n}{f_1} = \frac{k_n}{k_1} \frac{m_1}{m_n} \frac{A_n^2}{A_1^2}. \quad (20)$$

From (17) we know that $A_n = 1/M_{11}$, and a comparison of (18) and (16) yields $M_{11} = T_{22}$, hence we arrive at an expression which is usually found in literature for the tunneling coefficient:

$$TC = \frac{k_n}{k_1} \frac{m_1}{m_n} \frac{1}{T_{22}^2}. \quad (21)$$

This tunneling coefficient can now be inserted into expression (1) to yield the total tunneling current through the barrier stack.

For the simulation of the silicon regions between the nitride layers we used insulating layers with the band edge energy and permittivity of silicon. This assumption is justified by the fact that usually intrinsic silicon is used for the silicon layers between the barriers. In a recent

publication Mizuta et al. applied a Schrödinger–Poisson solver to simulate a single nitride barrier and used the resulting $I(V)$ data to calibrate a drift-diffusion device simulator [5]. However, this approach can only give a crude approximation since resonances between the barriers are not taken into account. Additionally, for a correct simulation of the carrier charge in the insulating layers, the carrier concentration due to the wave functions inside the oxide must be calculated by

$$n(x) = \frac{mk_B T}{\pi \hbar^2} \sum_i |\Psi_i(x)|^2 \ln \left[1 + \exp \left(\frac{E_f - E_i}{k_B T} \right) \right] \quad (22)$$

where E_i is the bound state energy level and E_f the Fermi energy. However, this expression does not account for quasi-bound states inside the oxide layers. Only [16] gives an expression for the electron density for quasi-bound states within oxides for the case of an applied bias using the Green function formalism which is far too complex for a multi-purpose device simulator. In resonant tunneling diodes this problem is usually solved by assigning either the left or the right Fermi level to each region. However, this choice is somewhat arbitrary and so will be the results [13]. It was also shown by several authors (for example [17], where Miranda et al. reproduced measured oscillations in the Fowler–Nordheim current, and [18], where Pan et al. successfully applied the transfer-matrix method to the simulation of a super lattice hetero-bipolar transistor), that the charge of the tunneling electrons has only minor effects on the total tunneling current. Thus we assumed that the transfer-matrix method is able to provide at least significant approximations for the tunneling process and, therefore, the contribution of the carrier charge in the classically forbidden regions can be omitted.

3. Simulation results

The model has been implemented in the device simulator MINIMOS-NT in a self-consistent manner. The gate oxide is divided into one-dimensional slices with constant potential and material properties, and the total current is found by a summation over all slices. We used the results of Mizuta et al. [5] for a single Si_3N_4 barrier diode to calibrate our simulator and found good agreement to their data, see Fig. 3. Electron and hole tunneling processes have been taken into account. For calibration the carrier mass in the oxide was used as a fit parameter. Electron and hole masses of $0.5m_0$ and $0.8m_0$ were found to reproduce Mizuta's results reasonably well. The Si_3N_4 barrier was modeled with a barrier height of 5 eV and a conduction band offset of 2 eV to the Si conduction band edge with the dielectric permittivity being 7.5. For SiO_2 , we used a barrier height of 3.2 eV. Fig. 4 shows a plot of the resulting TC for the three-barrier structure

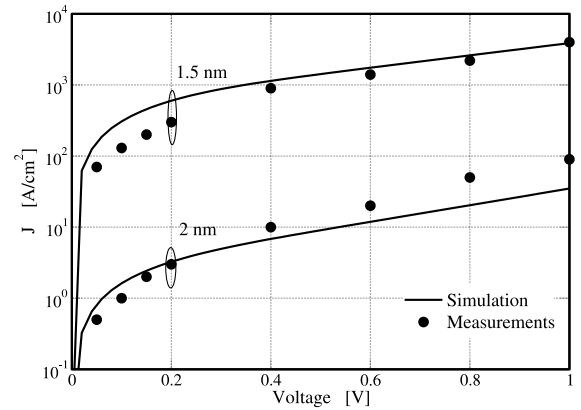


Fig. 3. Calibration to measurement data. The measurement values are taken from [5].

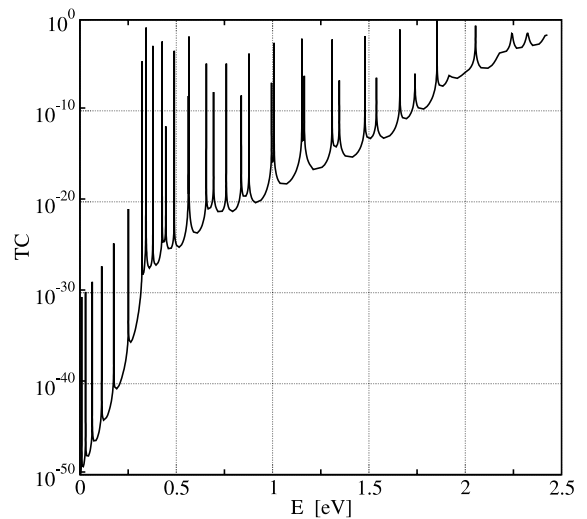


Fig. 4. The TC of the PLEDTR as a function of energy.

with a total stack height of 100 nm, 2 nm thick upper and lower barriers, and a 10 nm CSB. It can be seen that there are various resonances which correspond to quasi-bound states within the energy wells. These resonances are of fundamental importance for the accuracy of the total current and must be resolved appropriately. An energy grid in the range of peV has to be used in these regions to get precise results. Note that the number and steepness of the peaks depends on the number and width of the barriers, as well as on the applied voltage.

The absolute value of the normalized squared electron wave function for an energy level of 0.82 eV at an applied voltage at the word line of 2 V is shown in Fig. 5, together with the incident carrier energy (full line) and the energy barrier (dashed line). For this figure the stack height and the CSB thickness have been reduced to 35 and 5 nm, respectively.

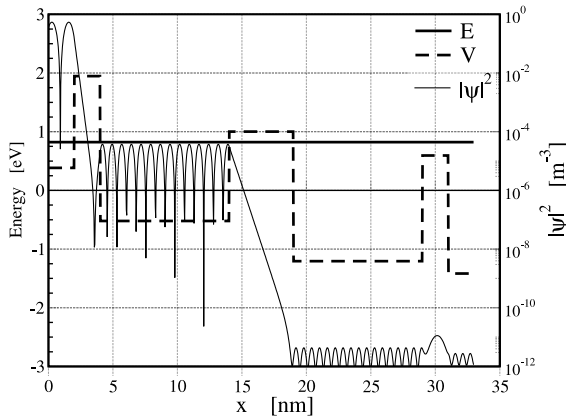


Fig. 5. The electron wave function for an energy level of 0.82 eV in the PLEDTR stack. E is the energy of the incident electron, V the energy barrier. The square of the electron wave function is denoted by $|\Psi|^2$.

We investigated the effect of the position and size of the CSB as well as the effect of shrinking the stack width. We assumed two cell states: an on-state with 3 V applied on the data line and the word line, and an off-state with 0.8 V applied on the memory node and 0 V on the word line. In both states we extracted the charging current (I_{on}) and the discharging current (I_{off}). The PLEDTR had a stack width of 180 nm and a stack height of 100 nm. The thickness of the upper and lower barriers was set to 2 nm. The thickness of the side gate oxide was 3 nm.

3.1. Position and thickness of the CSB

Fig. 6 shows the effect of different CSB thicknesses on the on- and off-current of the device. While the on-current is hardly influenced by the different thicknesses, the off-current is very sensitive to it. Also, the position of the CSB is critical, because for a CSB located near the

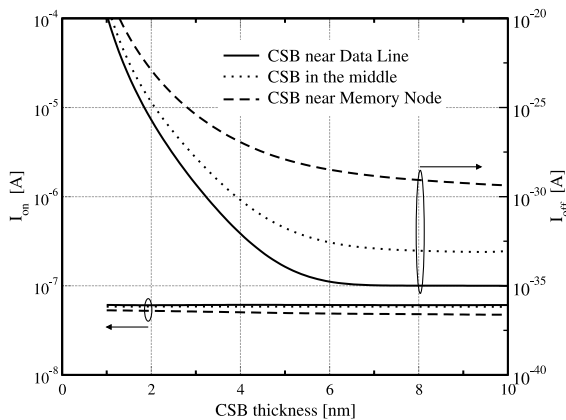


Fig. 6. I_{on} and I_{off} for different CSB thicknesses.

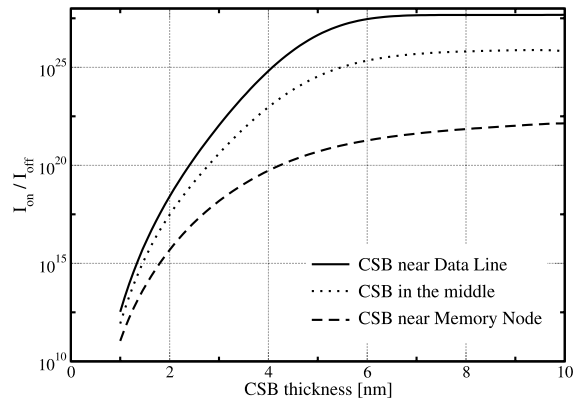


Fig. 7. I_{on}/I_{off} ratio for different CSB thicknesses.

memory node, the energy barrier will be reduced in the off-state by the charge on the memory node. If, on the other hand, the CSB is placed near the data line, the energy barrier is not suppressed and the off-current is much lower. This is caused by the lowering of the effective energy barrier which is given by

$$V(x) = E_{c,0} - q\phi(x) \tag{23}$$

where $\phi(x)$ is the electrostatic potential. The on-current is also reduced by this effect, but the amount of reduction is much lower as compared to the off-current, due to the fact that the on-current mainly depends on the voltage of the word line. In Fig. 7 we show the I_{on}/I_{off} ratio as a function of CSB thickness and position. It can be seen that the ratio increases with the thickness of the CSB and that it is best for a CSB located near the data line. Such an asymmetry in the $I-V$ characteristics depending on the position of the CSB has already been experimentally observed [5]. An I_{on}/I_{off} ratio of 10^{27} can be reached for a CSB thickness of 7 nm.

3.2. Width of the barrier stack

In [19] the feasibility of very narrow silicon-insulator stacks is shown. This encourages the assumption that a further reduction of the width of the gate stack is possible. We investigated the effect of shrinking the stack width on the device performance. Fig. 8 shows the on- and off-currents of the device with a CSB thickness of 10 nm for a stack width of 140 down to 20 nm. It can be seen that reductions of the stack width lead to increasing on-currents and decreasing off-currents. The reason is that the current in the on-state, which mainly flows as a surface current near the word line, is not reduced by the decreased width of the stack. It even increases for very low stack widths which may be due to the fact that the energy barriers at the side of the stack merge for very low stack widths. The off-current, on the other hand, is

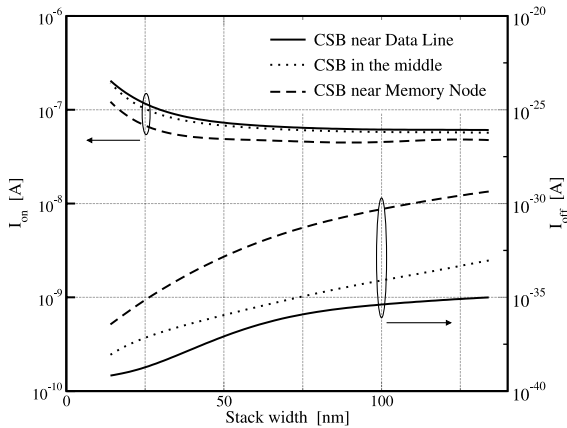


Fig. 8. I_{on} and I_{off} for different stack widths.

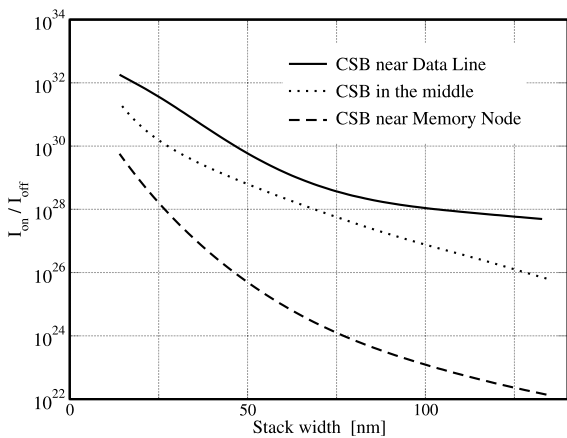


Fig. 9. $I_{\text{on}}/I_{\text{off}}$ ratio for different stack widths.

directly proportional to the stack area and can thus be directly downscaled by shrinking the stack width. The $I_{\text{on}}/I_{\text{off}}$ characteristics in Fig. 9 is greatly improved with decreasing width. For an extremely narrow gate stack of only 20 nm, $I_{\text{on}}/I_{\text{off}}$ ratios of more than 10^{32} could be reached, allowing an increase in the retention time by several orders of magnitude.

4. Conclusions

We showed quantum-mechanical simulations of a recently proposed PLEDM multi-barrier tunneling device by incorporating the transfer-matrix formalism into the device simulator MINIMOS-NT. Simulation results indicate that the device performance in terms of the $I_{\text{on}}/I_{\text{off}}$ ratio can be optimized by choosing a proper arrangement of the barriers. In particular, placing the CSB near the data line gives better performance inde-

pendent of the barrier thickness. The reason is that in the off-state, the voltage at the memory node reduces barriers which are located near the memory node, while it has hardly any influence on the barriers near the data line. Thus, for a low off-current, barriers should be placed near the data line where they are not lowered by the stored charge. The performance can also be increased by shrinking the stack width. Reductions of the stack width lead to lower off-current, but hardly influence the on-current, since the on-current flows only at the interface to the side gate oxide which is not influenced by the reduced stack width, while the off-current flows through the whole stack area. An on-current of 10^{-7} A and an off-current of as low as 10^{-38} A can be reached, values which correspond to results reported in [2].

Acknowledgements

This work has been partly supported by the Samsung Advanced Institute of Technology (SAIT), Seoul, Korea. We thank C. Pacher for stimulating discussions.

References

- [1] Nakazato K, Piotrowicz PJA, Hasko DG, Ahmed H, Itoh K. PLED—planar localised electron devices. In: International Electron Devices Meeting, 1997. p. 179–82.
- [2] Mizuta H, Nakazato K, Piotrowicz PJA, Itoh K, Teshima T, Yamaguchi K, et al. Normally-off PLED (planar localised electron device) for non-volatile memory. In: Symposium on VLSI Technology Digest of Technical Papers, 1998. p. 128–9.
- [3] Nakazato N, Itoh K, Mizuta H, Ahmed H. Silicon stacked tunnel transistor for high-speed and high-density random access memory gain cells. *Electron Lett* 1999;35:848–50.
- [4] Nakazato K, Itoh K, Ahmed H, Mizuta H, Kisu T, Kato M, et al. Phase-state low electron-number drive random access memory PLEDM. In: International Solid-State Circuits Conference, 2000.
- [5] Mizuta H, Wagner M, Nakazato K. The role of tunnel barriers in phase-state low electron-number drive transistors PLEDTR. *IEEE Trans Electron Dev* 2001;48(6): 1103–8.
- [6] Integrated Systems Engineering, Zürich, DESSIS User's Manual, 2000.
- [7] Avant!, Fremont, USA, MEDICI User's Manual, 2000.
- [8] Shanware A, Shiely JP, Massoud HZ. Extraction of the gate oxide thickness of n- and p-channel MOSFETs below 20 Å from the substrate current resulting from valence-band electron tunneling. In: International Electron Devices Meeting, 1999. p. 815–8.
- [9] Gundlach K. Zur Berechnung des tunnelstroms durch eine trapezförmige potentialstufe. *Solid-State Electron* 1966;9: 949–57.

- [10] Tsu R, Esaki L. Tunneling in a finite superlattice. *Appl Phys Lett* 1973;22(11):562–4.
- [11] Vassell MO, Lee J, Lockwood HF. Multibarrier tunneling in $\text{Ga}_{1-x}\text{Al}_x\text{As}/\text{GaAs}$ heterostructures. *J Appl Phys* 1983;54(9):5208–13.
- [12] Ando Y, Itoh T. Calculation of transmission tunneling current across arbitrary potential barriers. *J Appl Phys* 1987;61:1497–502.
- [13] Ferry DK, Goodnick SM. *Transport in nanostructures*. Cambridge: Cambridge University Press; 1997.
- [14] Lui W, Fukuma M. Exact solution of the Schrödinger equation across an arbitrary one-dimensional piecewise-linear potential barrier. *J Appl Phys* 1986;60(5):1555–9.
- [15] Gasirowicz S. *Quantenphysik*. München Wien: Oldenbourg Verlag; 1996.
- [16] Datta S. Nanoscale device modeling: the Green's function method. *Superlatt Microstruct* 2000;28(4):253–78.
- [17] Miranda E, Faigon A, Campabadal F. Analysis of experimental current oscillations in MOS structures using a semi-empirical tunneling model. *Solid-State Electron* 1997;41(1):67–73.
- [18] Pan H-J, Feng S-C, Wang W-C, Lin K-W, Yu K-H, Wu C-Z, et al. Investigation of an InGaP/GaAs resonant-tunneling heterojunction bipolar transistor. *Solid-State Electron* 2001;45:489–94.
- [19] Fukuda H, Hoyt JL, McCord MA, Pease RFW. Fabrication of silicon nanopillars containing polycrystalline silicon/insulator multilayer structures. *Appl Phys Lett* 1997;70(3):333–5.