Superlattices and Microstructures, Vol. 23, No. 3/4, 1998

Kinetic modelling of electron tunneling processes in quantum dots coupled to field-effect transistors

FARHAN RANA
Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.

SANDIP TIWARI†, J. J. WELSER
IBM Research Division, Thomas J. Watson Research Center, Yorktown Heights, NY 10598, U.S.A.

Transport of electrons in semiconductor nano-structures exhibits many features that are a consequence of quantum confinement and Coulomb blockade. A quantum dot coupled to a metal-oxide-semiconductor transistor's channel region is one example of such a structure with utility as a dense semiconductor memory. The memory state of this unit cell is a function of the number of electrons stored in the quantum dot and is sensed by the conduction in the channel. We describe a kinetic approach, based on a master equation, for modelling the injection and ejection of electrons into and from the quantum dot, and compare numerical results with experimental results for the silicon/silicon dioxide system where such memory structures have been achieved.

Key words: quantum dot, nano-crystal, tunneling, memory, field-effect transistor, charging, discharging, modeling, rate equation, master equation.

1. Introduction

When stories are recounted from IBM Research Division's past, there are often quite a few of the dual role of Rolf Landauer as a scientist and a manager. Of the former, his early elucidation of concepts of mesoscopic physics, and of the latter, his role in the move from vacuum technology to solid-state technology and then to integration are particularly relevant to the present times. Small critical dimensions are bringing together these two areas of interest to the scientist and the engineer.

In many laboratories, the silicon metal-oxide-semiconductor field effect transistor is rapidly approaching length scales of $\approx$ 10 nm—a dimension which is of the same order of magnitude as the thickness of inversion layer, Bohr radius, screening lengths, and mean free paths. Inevitably, quantum effects, and sometimes, single-electron effects, are important in these devices. Nanocrystal [1, 2] memories and quantum dot [3, 4] memories are two examples of such devices. In the quantum dot memory device a single quantum dot is placed between the gate and the channel of a field effect transistor. The quantum dot is quantum mechanically coupled to the transistor channel via a very thin layer of oxide (see Fig. 1). Electrons can be stored in the quantum dot and the memory state of the device is defined by the number of stored electrons. This stored charge electrostatically

† E-mail tiwari@watson.ibm.com

0749–6036/98/030757 + 14 $25.00/0 sm970539

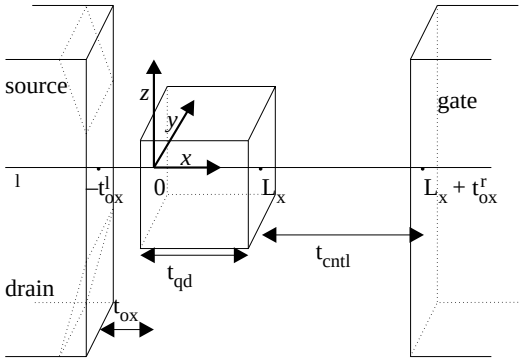

Fig. 1. A geometry showing the coupling of a quantum dot with a gate and a silicon channel through silicon dioxide.

influences the flow of electrons in the transistor channel. Changes in the transistor current are a measure of the number of electrons stored in the quantum dot and, therefore, a way of determining the state of the device. Electrons can be stored in the quantum dot by lowering the potential energy of the quantum dot below that of the channel by applying a positive voltage at the gate with respect to the source and drain. The ejection of the electrons occurs when the channel potential energy is lowered by a positive voltage applied at the source and drain or when the gate potential energy is raised by a negative voltage applied at the gate. Nano-crystal memory devices, which historically appeared first, are similar to quantum dot devices, except that, instead of a single patterned quantum dot, nano-crystal devices have a large number of silicon nano-crystals embedded in the oxide. These nano-crystals are randomly deposited instead of being patterned through processing.

Both of these memory devices offer the possibility of lower power dissipation and denser memory chips. The consequence is a decrease of write and erase times and an increase in sensitivity to random charge fluctuations. Here we present an approach, starting from the density matrix, to obtaining the stationary and time-dependent responses of these devices and for analysing the charge fluctuations.

2. Quantum kinetic equation

The Hamiltonian for the system can be written using the tunneling Hamiltonian,

$$H = H_{2deg} + H_{qd} + H_T, \tag{1}$$

where

$$\begin{aligned} H_{2deg} &= \sum_n (\epsilon_n + eV) a_n^\dagger a_n, \\ H_{qd} &= \sum_m \epsilon_m b_m^\dagger b_m + E_s(\nu), \end{aligned} \tag{2}$$

and

$$H_T = \sum_{n,m} T_{nm} a_n^\dagger b_m + c.c., \tag{3}$$

where H_{2deg} and H_{qd} are the Hamiltonians for the two-dimensional channel electron gas, and the quantum dot. H_T represents the coupling due to tunneling of electrons between the channel and the quantum dot. eV is the difference in potential energy between the quantum dot and the channel in the absence of any electrons

in the quantum dot. It is also equal to the work done by the external voltage source when one electron tunnels from the channel into the quantum dot. $E_S(\nu)$ is a function which describes the electrostatic energy of the system when there are ν electrons in the quantum dot. We will elaborate on this later in this paper.

We start with the equation of motion for the complete density matrix $\hat{P}_H$ in the Heisenberg representation as

$$i\hbar\frac{\partial\hat{P}_H(t)}{\partial t} = [H, \hat{P}_H(t)]. \tag{4}$$

Density matrix $\hat{P}_I(t)$ in the interaction representation is

$$\hat{P}_I(t) = \exp\left[\frac{i}{\hbar}\int_{t_0}^{t} H_0 dt'\right]\hat{P}_H(t)\exp\left[-\frac{i}{\hbar}\int_{t_0}^{t} H_0 dt'\right], \tag{5}$$

where $H_0 = H_{2deg} + H_{qd}$. The equation of motion in the interaction representation then becomes

$$i\hbar\frac{\partial\hat{P}_I(t)}{\partial t} = [H_T(t), \hat{P}_I(t)], \tag{6}$$

where

$$H_T(t) = \exp\left[\frac{i}{\hbar}\int_{t_0}^{t} H_0 dt'\right]H_T\exp\left[-\frac{i}{\hbar}\int_{t_0}^{t} H_0 dt'\right]. \tag{7}$$

The solution can be written as,

$$\hat{P}_I(t) = \hat{P}_I(t_0) - \frac{i}{\hbar}\int_{t_0}^{t}[H_T(t'), \hat{P}_I(t')]dt', \tag{8}$$

which yields the equation of motion

$$\frac{\partial\hat{P}_I(t)}{\partial t} = -\frac{i}{\hbar}[H_T(t), \hat{P}_I(t_0] + \left(\frac{i}{\hbar}\right)^2\int_{t_0}^{t}[H_T(t), H_T(t'), \hat{P}_I(t')]dt'. \tag{9}$$

Coherence times are usually much smaller than evolution times of $\hat{P}_I(t)$, allowing for substitution of $\hat{P}_I(t')$, in Markoff approximation, by $\hat{P}_I(t)$ inside the integral in the above equation.

2.1. Rate equation

Let the many body state of the entire system be described by the quantum state $|\{n_n\}, \{n_m\}\rangle$. This state is characterized by a particular configuration of occupation numbers $\{n_n\}$ and $\{n_m\}$ for the states belonging to the channel and the quantum dot. Let $p(\{n_m\})(t)$ be the probability that the quantum dot is described by occupation numbers $\{n_m\}$ at time t. $p(\{n_m\})(t)$ can be calculated from the density matrix $\hat{P}_I(t)$ by taking a trace over all the quantum states belonging to the electrons in the channel:

$$p(\{n_m\})(t) = \sum_{\{n_n\}}\langle\{n_n\}, \{n_m\}|\hat{P}_I(t)|\{n_n\}, \{n_m\}\rangle. \tag{10}$$

$p(\{n_m\})(t)$ contains a lot more information than needed to model the system. Usually, one would only be interested in the probability $p_\nu(t)$ that the quantum dot contains a total of ν electrons at time t. This can be obtained from $p(\{n_m\})(t)$ by summing over all possible configurations with occupation numbers $\{n_m\}$ such that $\sum_m n_m = \nu$. Thus,

$$\begin{aligned} p_\nu(t) &= \sum_{\{n_m\}} p(\{n_m\})(t)\ \delta(\sum_m n_m, \nu) \\ &= \sum_{\{n_m\}}\sum_{\{n_n\}}\langle\{n_n\}, \{n_m\}|\hat{P}_I(t)|\{n_n\}, \{n_m\}\rangle\ \delta(\sum_m n_m, \nu). \end{aligned} \tag{11}$$

Using eqn (9), and performing the indicated summations in the above equation, we get the master equation

$$\frac{\partial p_\nu(t)}{\partial t} = W_{\nu+1\to\nu}\, p_{\nu+1}(t) + W_{\nu-1\to\nu}\, p_{\nu-1}(t) - W_{\nu\to\nu+1}\, p_\nu(t) - W_{\nu\to\nu-1}\, p_\nu(t). \tag{12}$$

$W_{\nu\to\mu}$ are the transition rates for going from the state of the quantum dot with ν electrons inside it to the state where there are μ electrons inside it. These transition rates are:

$$W_{\nu+1\to\nu} = \frac{2\pi}{\hbar}\sum_{n,m}|T_{nm}|^2\delta(\epsilon_n + E_s(\nu+1) + eV - \epsilon_m - E_s(\nu))(1 - f_{FD}(\epsilon_n - E_f))\, f_{\nu+1}(\epsilon_m),$$

$$W_{\nu-1\to\nu} = \frac{2\pi}{\hbar}\sum_{n,m}|T_{nm}|^2\delta(\epsilon_n + E_s(\nu-1) + eV - \epsilon_m - E_s(\nu))\, f_{FD}(\epsilon_n - E_f)(1 - f_{\nu-1}(\epsilon_m)),$$

$$W_{\nu\to\nu+1} = \frac{2\pi}{\hbar}\sum_{n,m}|T_{nm}|^2\delta(\epsilon_n + E_s(\nu) + eV - \epsilon_m - E_s(\nu+1))\, f_{FD}(\epsilon_n - E_f)(1 - f_\nu(\epsilon_m)),$$

$$\text{and}\quad W_{\nu\to\nu-1} = \frac{2\pi}{\hbar}\sum_{n,m}|T_{nm}|^2\delta(\epsilon_n + E_s(\nu) + eV - \epsilon_m - E_s(\nu-1))(1 - f_{FD}(\epsilon_n - E_f))\, f_\nu(\epsilon_m). \tag{13}$$

f_{FD} and E_f are the Fermi–Dirac distribution function and the quasi-Fermi level for electrons in the channel. $f_\nu(\epsilon_m)$ is the probability that the state with energy ϵ_m in the quantum dot is occupied given that there are a total of ν electrons in the quantum dot. Note that the quantum dot is not attached to a particle reservoir in the sense of 'grand canonical ensemble.' Thus, Fermi-Dirac statistics may not be used to describe occupation statistics of electrons inside the quantum dot. Since the lifetimes of electrons in the quantum dots are expected to be much larger than the relaxation times, we can compute $f_\nu(\epsilon_m)$ for a canonical ensemble with ν electrons

$$f_\nu(\epsilon_m) = \frac{\sum_{\{n_{m'}\}}\exp\left(-\frac{1}{kT}\sum_{m'}\epsilon_{m'}n_{m'}\right)\delta(\sum_{m'}, \nu n_{m'})\delta(n_m, 1)}{\sum_{\{n_{m'}\}}\exp\left(-\frac{1}{kT}\sum_{m'}\epsilon_{m'}n_{m'}\right)\delta(\sum_{m'} n_{m'}, \nu)}. \tag{14}$$

For a quantum dot with few electrons, $f_\nu(\epsilon_m)$ can be calculated numerically.

2.2. Coupling constants

The coupling constant, T_{mn}, between a state $\psi_m(\boldsymbol{r})$ in the quantum dot and a state $\psi_n(\boldsymbol{r})$ in the two dimensional electron gas is given by,

$$T_{mn} = -\frac{\hbar^2}{2m_{ox}}\int[\psi_m^*(\boldsymbol{r})\boldsymbol{\nabla}\psi_n(\boldsymbol{r}) - \psi_n(\boldsymbol{r})\boldsymbol{\nabla}\psi_m^*(\boldsymbol{r})]\cdot d\boldsymbol{S}, \tag{15}$$

where m_{ox} is the electron effective mass in the oxide and the surface integral is performed over any surface separating the quantum dot and the channel and lying inside the oxide barrier. The wavefunction of an electron belonging to the two-dimensional channel takes the following form in the oxide adjacent to the channel,

$$\psi_{n,q_y,q_z}(\boldsymbol{r}) \approx \left(\frac{2}{AL_w}\right)^{1/2}\exp(iq_y y + iq_z z)\frac{(\beta(-t_{ox})/\beta(x))^{1/2}q_x m_{ox}\exp\left(-\int_{-t_{ox}}^{x}\beta(x')dx'\right)}{\{(m_{Si}^x\beta(-t_{ox}))^2 + (m_{ox}q_x)^2\}^{1/2}}, \tag{16}$$

where q_y and q_z are the wavevectors of the electron parallel to the silicon/silicon dioxide interface, L_w is the approximate depth of the inversion layer, and A is the area of the region over which the wavefunction is normalized. The wavevector q_x is defined as,

$$q_x = \left(\frac{2m_{Si}^x}{\hbar^2}\epsilon_n\right)^{1/2}. \tag{17}$$

The effective mass m^x_{Si} in silicon can either be the transverse (m^t_{Si}) or the longitudinal (m^l_{Si}) mass depending on the valley (of the six equivalent ones) to which the electron belongs. And,

$$\beta(x) = \left[\frac{2m^x_{Si}}{\hbar^2}(e\phi_o - e\mathcal{E}_{ox}x - \epsilon_n)\right]^{1/2}, \tag{18}$$

where $\mathcal{E}_{ox}$ is the field in the oxide and ϕ_o is the barrier height [5]. For direct tunnelling, such a WKB based approach provides a description with excellent accuracy.

Quantum dot states can be computed accurately in the Hartree approximation. However, coupling constants can be determined with reasonable accuracy using the following approximate form for wavefunction inside the oxide:

$$\psi_{m,k_y,k_z}(\boldsymbol{r}) \approx \left(\frac{8}{L_yL_zL_x}\right)^{1/2} \sin(k_y y)\sin(k_z z)\frac{(\alpha(0)/\alpha(x))^{1/2}k_x m_{ox}\exp\left(\int_0^x \alpha(x')dx'\right)}{[(m^x_{Si}\alpha(0))^2 + (m_{ox}k_x)^2]^{1/2}}, \tag{19}$$

where

$$k_x = \left(\frac{2m^x_{Si}}{\hbar^2}\epsilon_m\right)^{1/2}. \tag{20}$$

L_x, L_y, and L_z are dimensions of the quantum dot, and the extinction coefficient $\alpha(x)$ is

$$\alpha(x) = \left[\frac{2m^x_{Si}}{\hbar^2}(e\phi_o - e\mathcal{E}_{ox}x - \epsilon_m)\right]^{1/2}, \tag{21}$$

where ϵ_m is the energy associated with motion in the x-direction.

The wavefunctions of eqns (16) and (19) lead to the coupling constant $T_{\{n,q_y,q_z\},\{m,k_y,k_z\}}$:

$$\begin{aligned} T_{\{n,q_y,q_z\},\{m,k_y,k_z\}} = &-\frac{\hbar^2}{2m_{ox}}\left(\frac{16}{AL_wL_xL_yL_z}\right)^{1/2} \\ &\times \frac{(\alpha(0)\beta(0))^{1/2}k_xq_xm^2_{ox}}{((m^x_{si}\beta(-t_{ox}))^2 + (m_{ox}q_x)^2)^{1/2}((m^x_{Si}\alpha(0))^2 + (m_{ox}k_x)^2)^{1/2}} \\ &\times \left[\left(\frac{\alpha(-t_{ox}/2)}{\beta(-t_{ox}/2)}\right)^{1/2} + \left(\frac{\beta(-t_{ox}/2)}{\alpha(-t_{ox}/2)}\right)^{1/2}\right] \\ &\times \exp\left[\int_0^{-t_{ox}/2}\alpha(x')dx' + \int_{-t_{ox}/2}^{-t_{ox}}\beta(x')dx'\right] \\ &\times \left[\int_0^{L_y}\exp(iq_yy)\sin(k_yy)dy\right]\left[\int_0^{L_z}\exp(iq_zz)\sin(k_zz)dz\right]. \end{aligned} \tag{22}$$

The coupling constants and the transition rates determined from the equations described above are quite general, i.e. are valid when the state inside the quantum dot is coupled to a two-dimensional channel (inversion condition) or a three dimensional continuum (depletion condition). The summations are, however, different in the expressions for transition rates W given by eqn. (13).

2.3. Electrostatic interaction energy and coulomb charging effects

In the Hamiltonian for the quantum dot we introduced a term $E_s(\nu)$ that accounted for the Coulomb charging effects associated with the presence of electrons inside the quantum dot. Here, we briefly describe the method used to compute $E_s(\nu)$ numerically. The method is different from the capacitance model commonly employed to analyze single electron devices.

When an electron tunnels from the channel into the quantum dot (or vice versa) the stored electrostatic energy in the device changes. In addition, some current flows through the external circuit and thus the external voltage source performs some work. Both these effects must be taken into account in order to calculate the final state of the electron after tunneling. The electrostatic energy of the system $E_s(\nu)$ can be written as

$$E_s = \frac{1}{2}\int \epsilon(\boldsymbol{r})\boldsymbol{E}(\boldsymbol{r})\boldsymbol{E}(\boldsymbol{r})d^3\boldsymbol{r}. \tag{23}$$

Since $\boldsymbol{E}(\boldsymbol{r}) = -\boldsymbol{\nabla}\phi(\boldsymbol{r})$, we can write

$$\begin{aligned} E_s &= -\frac{1}{2}\int \epsilon(\boldsymbol{r})\boldsymbol{E}(\boldsymbol{r})\cdot\boldsymbol{\nabla}_r\phi(\boldsymbol{r})d^3\boldsymbol{r} \\ &= -\frac{1}{2}\boldsymbol{\nabla}.(\epsilon(\boldsymbol{r})\phi(\boldsymbol{r})\boldsymbol{E}(\boldsymbol{r})\int d^3\boldsymbol{r} + \frac{1}{2}\int \phi(\boldsymbol{r})\boldsymbol{\nabla}\cdot(\epsilon(\boldsymbol{r})\boldsymbol{E}(\boldsymbol{r}))d^3\boldsymbol{r} \\ &= \frac{1}{2}\sum_n U_n Q_n + \frac{1}{2}\int \phi(\boldsymbol{r})\rho(\boldsymbol{r})d^3\boldsymbol{r}. \end{aligned} \tag{24}$$

The summation is over all the conducting bodies in the system which are at potential U_n and have a total charge of magnitude Q_n. $\rho(\boldsymbol{r})$ is the volume charge density in the region between the conducting bodies.

Consider the situation shown in Fig. 1 where a silicon quantum dot is placed in an oxide matrix in between the channel and the gate electrode of a metal-oxide-semiconductor transistor. For simplicity, we may take the channel to be a metallic electrode at potential U. This approximation is valid as long as the channel consists of an inversion or an accumulation layer. The gate electrode is connected to ground. Suppose there are ν electrons in the quantum dot giving rise to a volume charge density $\rho_\nu(\boldsymbol{r})$. The potential $\phi_\nu(\boldsymbol{r})$ can be calculated by solving the Poisson equation with appropriate boundary conditions and the solution will take the general form

$$\phi_\nu(\boldsymbol{r}) = \phi_0(\boldsymbol{r}) + \int K(\boldsymbol{r},\boldsymbol{r}')\rho_\nu(\boldsymbol{r}')d^3\boldsymbol{r}'. \tag{25}$$

Equation (24) can then be used to calculate $E_S(\nu)$. However, the important quantity is the change in electrostatic energy when one electron is added to (or removed from) the quantum dot. Suppose an electron with energy ϵ_n tunnels from the channel into a state inside the quantum dot with energy ϵ_m and changes the number of stored electrons from ν to $\nu+1$. ϵ_n and ϵ_m represent energies associated with only kinetic degrees of freedom. We can write the following energy conserving equation:

$$\begin{aligned} \epsilon_n - \epsilon_m &= \{\text{increase in electrostatic energy }\} - \{\text{work done by external voltage source}\} \\ &= \left\{\frac{1}{2}\int \Delta\phi(\boldsymbol{r})\rho_\nu(\boldsymbol{r})d^3\boldsymbol{r} + \frac{1}{2}\int \phi_\nu(\boldsymbol{r})\Delta\rho(\boldsymbol{r})d^3\boldsymbol{r} + \frac{1}{2}\int \Delta\phi(\boldsymbol{r})\Delta\rho(\boldsymbol{r})d^3\boldsymbol{r}\right. \\ &\quad \left. + \frac{1}{2}U\delta Q\right\} - \{U(\delta Q + e)\}, \end{aligned} \tag{26}$$

where $\Delta\phi(\boldsymbol{r}) = \phi_{\nu+1}(\boldsymbol{r}) - \phi_\nu(\boldsymbol{r})$, and $\Delta\rho(\boldsymbol{r}) = \rho_{\nu+1}(\boldsymbol{r}) - \rho_\nu(\boldsymbol{r})$. ΔQ can be calculated from simple electrostatics as

$$\Delta Q = -\frac{1}{U}\int \phi_0(\boldsymbol{r})\delta\rho(\boldsymbol{r})d^3\boldsymbol{r}. \tag{27}$$

This equation can be used in eqn (26), and with eqn (25), results in

$$\epsilon_n + eU = \epsilon_m + \int \phi_\nu(\boldsymbol{r})\Delta\rho(\boldsymbol{r})d^3\boldsymbol{r} + \frac{1}{2}\int \Delta\phi(\boldsymbol{r})\Delta\rho(\boldsymbol{r})d^3\boldsymbol{r}. \tag{28}$$

The last term on the left hand side does not depend upon the number of electrons in the quantum dot and can, therefore, be absorbed in the definition of the energy ϵ_m. It is a renormalization of the energy of an electron

inside the quantum dot due to image charges on the conducting electrodes. Ignoring this term, above equation could have been written intuitively since all it says is that total initial energy of the electron is equal to the total final energy, provided one uses the electrostatic potential prior to the tunneling event. But this statement cannot be completely correct since the potential changes during the tunneling process. The last term provides the necessary corrections. This equation can be generalized to treat the channel more realistically and not simply as a conducting/metal electrode. The term eU on the right hand side is replaced by the average initial potential energy of the electron eigenstate in the channel plus a small correction term similar to the one just discussed.

The formalism we have described in this section is ideal for numerical simulations and can be used where simple capacitance models are not adequate. It also forms a basis for the numerical simulations whose results are presented in this paper.

2.4. Carrier statistics and charge fluctuations

The master equation (eqn (12)) can be written in a compact form as

$$\frac{\partial \boldsymbol{P}(t)}{\partial t} = \mathbf{W} \,.\, \boldsymbol{P}(t), \tag{29}$$

where the vector $\boldsymbol{P}(t) = [p_0(t),\, p_1(t), \cdots p_{\nu_{max}}(t)]$, and we restrict the number of electrons in the quantum dot to a maximum of ν_{max} for computational ease. The transition matrix $\mathbf{W}$ is of dimension $\nu_{max} \times \nu_{max}$. Stationary solution of eqn (29) is found by setting $\partial \boldsymbol{P}(t)/\partial t = 0$, and the stationary probabilities for $\nu \neq 0$ are

$$\overline{p}_\nu = \frac{\prod_{q=1}^{\nu} \frac{W_{q-1\to q}}{W_{q\to q-1}}}{1 + \sum_{r=1}^{\nu_{max}-1} \prod_{r=1}^{q} \frac{W_{r-1\to r}}{W_{r\to r-1}}}, \tag{30}$$

and for $\nu = 0$, it is

$$\overline{p}_0 = \frac{1}{1 + \sum_{r=1}^{\nu_{max}-1} \prod_{q=1}^{r} \frac{W_{q-1\to q}}{W_{q\to q-1}}}. \tag{31}$$

The mean and variance of the number of electrons in the quantum dot is

$$\langle \nu \rangle = \sum_{\nu=0}^{\nu_{max}} \nu \overline{p}_\nu, \tag{32}$$

$$\sigma_\nu^2 = \langle \nu^2 \rangle - \langle \nu \rangle^2 = \sum_{\nu=0}^{\nu_{max}} \nu^2 \overline{p}_\nu - \left(\sum_{\nu=0}^{\nu_{max}} \nu \overline{p}_\nu \right)^2. \tag{33}$$

The number fluctuation spectrum $S_\nu(\omega)$ can be found from

$$S_\nu(\omega; t_o) = \frac{1}{2} \int_0^\infty \exp(i\omega(t - t_o))(\langle \nu(t)\nu(t_o) + \nu(t_o)\nu(t) \rangle - 2\langle \nu(t_o) \rangle^2) d(t - t_o), \tag{34}$$

and

$$\langle \nu(t)\nu(t_o) \rangle = \sum_{\nu=0,\mu=0}^{\nu_{max},\nu_{max}} \nu\mu p(\nu, t|\mu, t_o)\; p(\mu, t_o), \tag{35}$$

where, $p(\mu, t_o)$ is the probability of having μ electrons in the quantum dot at time t_o and $p(\nu, t|\mu, t_o)$ is the conditional probability that the quantum dot will have ν electrons at time t given that it had μ electrons at time t_o. In the Markoff approximation all correlation functions have time invariance, e.g., $\langle \nu(t)\nu(t_o) \rangle =$

$\langle \nu(t+t')\nu(t_o+t')\rangle$. Therefore, the spectrum is also time invariant. Also $p(\nu, t|\mu, t_o) = p(\nu, t-t_o|\mu, 0)$, and for large t', $p(\mu, t_o+t') = \overline{p}_\mu$. Using these results, eqn (34) can be cast into the more useful form

$$S_\nu(\omega) = \frac{1}{2}\int_{-\infty}^{\infty} \exp(i\omega t)\left\{ \sum_{\nu=0,\mu=0}^{\nu_{max},\nu_{max}} \nu\mu p(\nu, t|\mu, 0)\overline{p}_\mu \theta(t) + \sum_{\nu=0,\mu=0}^{\nu_{max},\nu_{max}} \nu\mu p(\nu, 0|\mu, t)\,\overline{p}_\mu \theta(-t) - 2\left(\sum_{\nu=0}^{\nu_{max}} \nu\overline{p}_\nu\right)^2 \right\} dt. \tag{36}$$

$p(\nu, t|\mu, 0)$ and $p(\nu, 0|\mu, t)$ can be calculated from eqn (12).

3. Discussion

To obtain solutions for eqn (12), we first solve self-consistently for all the eigenstates in the channel for a given number of electrons stored in the quantum dot, for all gate voltages. We repeat this for different values of the number of stored electrons. The transition rates are then determined from eqn (13). We follow with a discussion of thenumerical results.

3.1. Threshold voltage shifts and fluctuations

The storage of electrons in the quantum dot results in a shift in the threshold voltage of the metal-oxide-semiconductor transistor, which, for a single quantum dot device with ν electrons in the quantum dot, is approximately given by,

$$\Delta V_T(\nu) \approx \frac{\nu e}{A\epsilon_{ox}}\left(\frac{t_{qd}\epsilon_{ox}}{2\epsilon_{Si}} + t_{cntl}\right), \tag{37}$$

where t_{cntl} is the control oxide thickness, and t_{qd} is the linear dimension of the quantum dot of cross-section A.

Figure 2 shows the calculated mean number of electrons in a silicon quantum dot of dimensions 10 nm × 10 nm × 6 nm as a function of the static gate bias. This calculation assumed a tunnel oxide of thickness 1.5 nm, control oxide of thickness 5.0 nm, substrate doping of 10^{17} cm^{-3} p-type with a $\langle 100\rangle$ orientation for the substrate and the quantum dot. The first electron does not appear in the quantum dot until the gate voltage exceeds the threshold voltage V_{TO} and an inversion layer is formed to supply that electron. When the quantum dot has one electron, the threshold voltage of the device is shifted up by $\Delta V_T(\nu = 1)$. For this device $\Delta V_T(\nu = 1)$ is about 0.3 V. To place a second electron in the quantum dot the gate voltage needs to be increased by at least $\Delta V_T(\nu = 1)$ beyond the voltage needed to place the first electron, which is approximately V_{TO}. The electron already inside the quantum dot tries to occupy the lowest available states, leaving the higher ones for the second electron. Thus the second electron can be placed in at a gate voltage which is slightly higher than $V_{TO} + \Delta V_T(\nu = 1)$. Placing the ν'th electron in the quantum dot requires a gate voltage of approximately $V_{TO} + \Delta V_T(\nu - 1)$.

Figure 2 shows the variance in electron number in the quantum dot as a function of gate bias. The variance is $1/2$ at gate voltages for which the mean number of electrons is $integer + 1/2$. The actual number of electrons in the quantum dot can take only integer values. A mean electron number of $integer + 1/2$ implies that the actual number of electrons is fluctuating rapidly between $integer$ and $integer + 1$. An analogy with a two state system is helpful here. Suppose the stationary probabilities $\overline{p}_\nu$ are all almost zero for all $\nu < \mu$ and for all $\nu > \mu + 1$. In this case the rate equations that follow from eqn (12) are:

$$\frac{\partial}{\partial t}\begin{bmatrix} p_\nu(t) \\ p_{\nu+1}(t) \end{bmatrix} = \begin{bmatrix} -z & w \\ z & -w \end{bmatrix}\begin{bmatrix} p_\nu(t) \\ p_{\nu+1}(t) \end{bmatrix}, \tag{38}$$

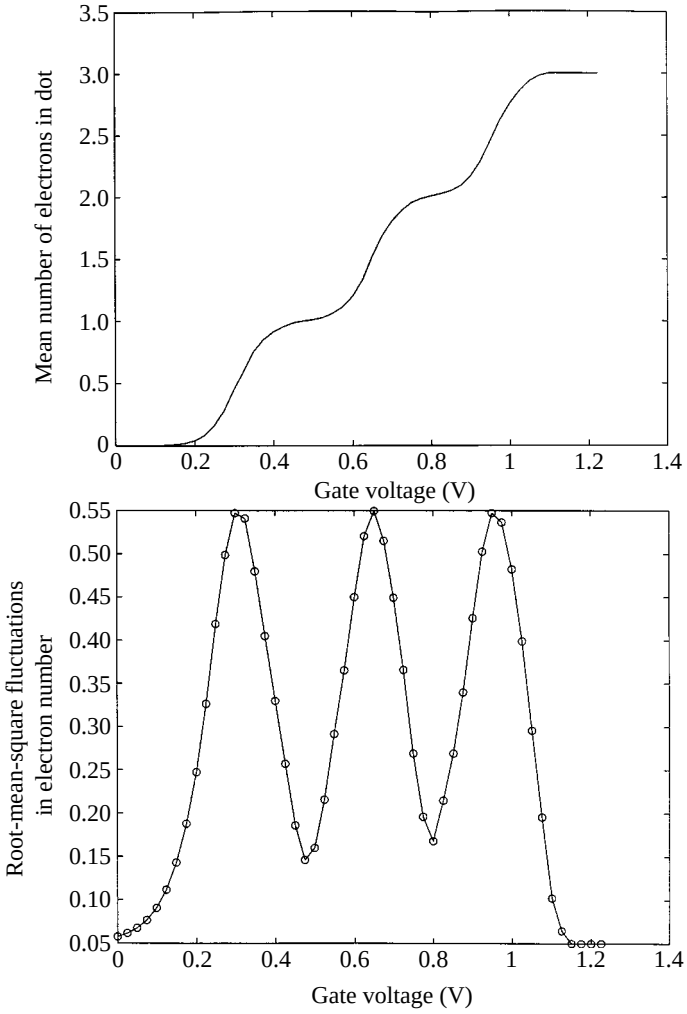

Fig. 2. Mean number of electrons in a silicon quantum dot (10 nm × 10 nm × 6 nm) as a function of applied gate voltage and the variance in electron number. Calculation is limited to a maximum of 3 electrons in the quantum dot.

where z and w are the rates $W_{\nu\to\nu+1}$ and $W_{\nu+1\to\nu}$. The solutions to the above equation are:

$$p(\nu, t|\mu, 0) = \frac{w}{z+w} + \exp[-(z+w)t]\left[\frac{z}{z+w}\delta(\mu, \nu) - \frac{w}{z+w}\delta(\mu, \nu+1)\right], \tag{39}$$

and

$$p(\nu+1, t|\mu, 0) = \frac{z}{z+w} - \exp[-(z+w)t]\left[\frac{z}{z+w}\delta(\mu, \nu) - \frac{w}{z+w}\delta(\mu, \nu+1)\right]. \tag{40}$$

For large times, $t \to \infty$,

$$p(\nu, t|\mu, 0) \to \overline{p}_\nu = \frac{w}{z+w}, \tag{41}$$

and

$$p(\nu+1,t|\mu,0) \to \overline{p}_{\nu+1} = \frac{z}{z+w}. \tag{42}$$

The mean number of electrons in the quantum dot is

$$\langle\nu\rangle = \frac{w}{z+w}\nu + \frac{z}{z+w}(\nu+1) = \nu + \frac{z}{z+w}. \tag{43}$$

When the transition rates are nearly identical (i.e., $z = w$) and the mean is $\nu + 1/2$, the variance in electron number becomes

$$\sigma_\nu^2 = \langle\nu^2\rangle - \langle\nu\rangle^2 = \frac{zw}{(z+w)^2} = \frac{1}{2}. \tag{44}$$

The spectrum of number fluctuations can be found using equation (36),

$$S_\nu(\omega) = \frac{z+w}{\omega^2 + (z+w)^2}\sigma_\nu^2. \tag{45}$$

The spectrum has a Lorentzian shape with a width directly related to the sum of the transition rates characterizing the system.

3.2. Channel current fluctuations

Fluctuations in number of electrons in the quantum dot can be related to the fluctuations in channel current which can be directly observed. At small values of drain to source bias, the channel current of the device can be approximated [6] by

$$I_{DS} = K(V_{GS} - V_T)V_{DS} \tag{46}$$

where K is a constant factor, and V_{GS} is the gate-to-source bias. The mean value of channel current is

$$\langle I_{DS}\rangle = K(V_{GS} - \langle V_T\rangle)V_{DS}, \tag{47}$$

where, for a single quantum dot device,

$$\langle V_T\rangle = V_{TO} + \frac{\langle\nu\rangle e}{A\epsilon_{ox}}\left(\frac{t_{qd}\epsilon_{ox}}{2\epsilon_{Si}} + t_{cntl}\right). \tag{48}$$

The variance of fluctuations in current, $\sigma_{I_{DS}}$, can be related to the variance of fluctuations in the threshold voltage, σ_{V_T}, by

$$\sigma_{I_{DS}}^2 = K^2\,\sigma_{V_T}^2\,V_{DS}^2, \tag{49}$$

and, for a single quantum dot device,

$$\sigma_{V_T}^2 = \sigma_\nu^2\left(\frac{e}{A\epsilon_{ox}}\right)^2\left(\frac{t_{qd}\epsilon_{ox}}{2\epsilon_{Si}} + t_{cntl}\right)^2. \tag{50}$$

The spectrum of current fluctuations for a single quantum dot device can also be found as

$$S_{I_{DS}}(\omega) = K^2\,S_{V_T}(\omega)\,V_{DS}^2, \tag{51}$$

and

$$S_{V_T}(\omega) = S_\nu(\omega)\left(\frac{e}{A\epsilon_{ox}}\right)^2\left(\frac{t_{qd}\epsilon_{ox}}{2\epsilon_{Si}} + t_{cntl}\right)^2. \tag{52}$$

The case of nanocrystal memory device is more complex and a complete analysis even for small drain to source voltages is beyond the scope of this paper. In a nano-crystal device there are regions of the channel,

located just underneath each charged nanocrystal, where the local threshold voltage is higher and, therefore, the carrier density is lower. During current flow positive and negative charges pile up around each such region to form electric dipoles. Charges on these dipoles are imaged on the gate electrode and also on the drain and source. The total current flowing through the device is a function of the device size, nanocrystal density and the spatial distribution of nanocrystals. Only in the case where,

1. the device is large enough so that effective medium theories become applicable and a very small fraction of dipoles are imaged on the source and drain, and
2. the nanocrystal density is small enough so that the device operates far away from the percolation threshold,

can one write versions of Equations 48, 50 and 52 which hold approximately for nanocrystal devices:

$$\langle V_T \rangle = V_{TO} + \frac{n_{qd}\langle \nu \rangle e}{\epsilon_{ox}} \left(\frac{t_{qd}\epsilon_{ox}}{2\epsilon_{Si}} + t_{cntl} \right), \tag{53}$$

$$\sigma^2_{V_T} = \frac{n_{qd}\sigma^2_\nu}{A} \left(\frac{e}{\epsilon_{ox}} \right)^2 \left(\frac{t_{qd}\epsilon_{ox}}{2\epsilon_{Si}} + t_{cntl} \right)^2, \tag{54}$$

$$\text{and } S_{V_T}(\omega) = \frac{n_{qd}S_\nu(\omega)}{A} \left(\frac{e}{\epsilon_{ox}} \right)^2 \left(\frac{t_{qd}\epsilon_{ox}}{2\epsilon_{Si}} + t_{cntl} \right)^2, \tag{55}$$

where n_{qd} is the average density of nanocrystals and A is the total area of the device. We see that in this case the fluctuations in threshold voltage and current will get scaled with the area of the device.

3.3. Time evolution of charging and discharging processes

So far we have limited our analysis to the case where static voltages had been applied to the quantum dot device. However, in practice a voltage pulse is usually applied to change the state of a memory cell. It is important to understand the time dependent behavior of quantum dot memory cells since this provides the write and erase times for this memory. The method based upon master equation can also be used to study the time dependent behavior. If a square voltage pulse of time duration T is applied to the gate at time $t = 0$, then eqn (12) can be solved with appropriate boundary conditions to yield the time dependent probabilities $p_\nu(t)$. If the quantum dot at time $t = 0$ was empty then boundary conditions are $p_0(t = 0) = 1$, and $p_\nu(t = 0) = 0$ for $\nu \neq 0$. The time dependent mean number of electrons $\langle \nu \rangle(t)$ in the quantum dot is

$$\langle \nu \rangle(t) = \sum_{\nu=0}^{\nu_{max}} \nu p_\nu(t). \tag{56}$$

The threshold voltage shift at the end of the pulse is $\Delta V_T(\nu = \langle \nu \rangle(t = T))$. In general, therefore, the threshold voltage shifts depends upon the magnitude and also the duration of the pulse.

Figure 3 shows the number of electrons in a quantum dot as a function of time for three different values of the magnitude of the applied voltage pulse. A general feature is that larger the number of electrons already inside the quantum dot the longer it takes to add one more electron to the quantum dot. The initial state has no electrons in the quantum dot. Consequently, the potential drop across the injecting oxide is large, and the electric field in the tunneling oxide is also large, resulting in a large coupling constant. Electrons are injected into high energy states of the quantum dot where density of states is also large. These states are also empty. All these factors lead to a large injection rate. As the quantum dot gets filled up with electrons, the potential drop across the injecting oxide becomes smaller with a corresponding reduction in the coupling constant. Injection now takes place in relatively lower energy states of the quantum dot, where density of states is also smaller, and these states now have a finite occupation probability for electrons. The presence of electrons in the quantum dot also changes the threshold voltage of the device which results in less electrons available in

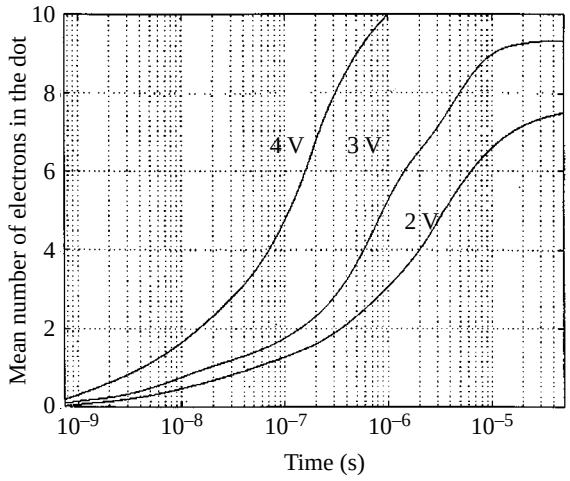

Fig. 3. The evolution of mean number of electrons in a silicon quantum dot (10 nm × 10 nm × 6 nm) for four different values of the gate bias. The initial number of electrons in each case is assumed to be zero.

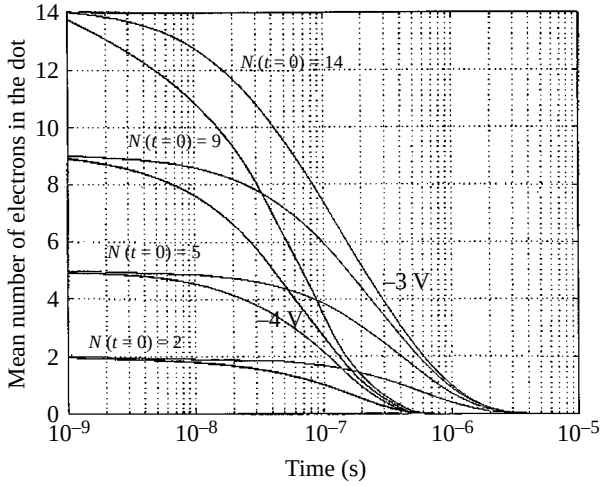

Fig. 4. The evolution of mean number of electrons in a silicon quantum dot (10 nm × 10 nm × 6 nm) when negative bias is applied to the gate. The figure shows the case when the initial number of electrons is 2,6 and 10 for three different values of the gate bias. The geometry is identical to the previous figure.

the channel that can tunnel into the quantum dot. All these factors contribute to a gradual reduction in the rate of electron injection into the quantum dot.

Figure 4 shows the numerical results when a negative voltage pulses are applied to the gate to eject the stored electrons into the substrate. The results are shown for three different initial number of stored electrons. The rate of discharge is higher for an initial condition with larger number of electrons in the quantum dot, a condition where there is a higher electric field across the tunneling oxide and hence a corollary to the low electron condition of the injection process. The behavior does not have as detailed features as the injection process; the injection process reveals more of the details of states being tunneled into.

Table 1: Extrapolated time-constants for measured structures with comparable nanocrystal density.

Oxide thickness	Write condition	ΔV_T
1.6 nm	200 ns, 3 V	≈ 0.65 V
2.1 nm	400 ns, 3 V	≈ 0.48 V
3.0 nm	1 μs, 3 V	≈ 0.55 V
3.6 nm	5 μs, 4 V	≈ 0.50 V

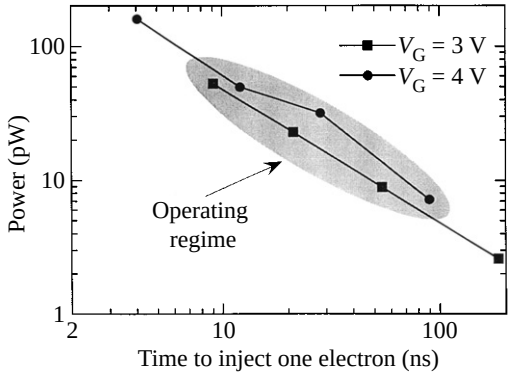

Fig. 5. Power required for injecting a single electron into the silicon dot as a function of the write speed.

The quantum dot can be charged with one electron with a 3.0 V pulse in 20 ns. But, discharging the quantum dot with a -3.0 V requires a pulse with time duration of 2 μs. Write times are, therefore, much smaller than erase times.

We can currently compare theoretical calculations with experimental results obtained on only nanocrystal memories. We have already mentioned the difficulties in analyzing the behavior of nanocrystal memory devices as compared to single quantum dot memory devices. But, rough comparisons can still be made for order of magnitude comparisons. Table 1 shows the injection time constants for devices, with different oxide thickness, and with approximately comparable nanocrystal density. The nanocrystal density and the observed threshold voltage shift in device corresponds to an average of 2–3 electrons per nanocrystal quantum dot. Although, in this case the smallest oxide thickness is greater than that simulated, time constants are roughly in the same range.

The interaction of the confinement and operating voltages in the time constants can be summarized in the power-delay figure shown in Fig. 5 for the write process. With the practical considerations of noise-margin taken into account through storage of a large number of electrons, the figure points to large advantages in power density that accrue through limiting the charge, but at the expense of speed.

4. Conclusions

We have outlined a density matrix approach for analyzing electron tunneling processes in quantum dots coupled to channel of a metal-oxide semiconductor transistor, i.e. to the two-dimensional inversion layer and

to unconfined substrate. We have also modeled the time dependent charging and discharging processes and calculated write and erase times. These calculations show that write times can be of the order of tens of nanoseconds whereas erase times can be as large as tens of micro-seconds, reasonably in good agreement with experimental results on nanocrystal memory devices.

References

[1] S. Tiwari, F. Rana, H. Hanafi, A. Hartstein, E. F. Crabbé, and K. Chan, Appl. Phys. Lett. **68**, 1377 (1996).
[2] S. Tiwari, F. Rana, K. Chan, L. Shi, and H. Hanafi, Appl. Phys. Lett. **69**, 1232 (1996).
[3] J. J. Welser, S. Tiwari, S. Rishton, K. Y. Lee, and Y. Lee, IEEE Electron Device Letters **18**, 278 (1997).
[4] L. Guo, E. Leobandung, and S. Chou, Science **175**, 649 (1997)
[5] Z. A. Weinberg, *Tunneling of Electrons from Si into Thermally Grown* SiO_2 , Solid State Electronics, **20**, 11 (1977).
[6] Yannis P. Tsividis, *Operation and Modeling of the MOS Transistor*, (New York, McGraw-Hill, 1987).