

Efficient quantum three-dimensional modeling of fully depleted ballistic silicon-on-insulator metal-oxide-semiconductor field-effect-transistors

M. J. Gilbert^{a)} and D. K. Ferry

Department of Electrical Engineering and Center for Solid State Electronics Research, Arizona State University, Tempe, Arizona 85287-5706

(Received 17 November 2003; accepted 16 February 2004)

We present an efficient, fully quantum mechanical approach to calculating ballistic transport in fully-depleted silicon-on-insulator metal-oxide semiconductor field effect transistor devices in three dimensions and apply the technique to the calculation of threshold voltages for realistic devices with narrow channels. We illustrate the fact that each dopant configuration becomes exceedingly important for quantum simulations. © 2004 American Institute of Physics.

[DOI: 10.1063/1.1699496]

INTRODUCTION

For many years the semiconductor industry has been pushing toward ever smaller devices. Industry has already demonstrated that it can produce devices that have gate lengths smaller than 15 nm that operate in a quite normal fashion.¹⁻³ While the creation of these devices is quite impressive, silicon-on-insulator (SOI) technology has been known to offer better performance over traditional bulk metal-oxide semiconductor field-effect-transistor (MOSFET) devices with better scaling capability. The relatively thin silicon film thicknesses associated with these devices (~5 nm) gives rise to quantization in the direction of the film growth, while recent experiments have shown it possible to fabricate MOS transistors in a SOI environment with channel widths as small as 2 nm.⁴⁻⁶ This gives rise to quantization in the lateral direction, which cannot be separated from the other two directions in the device. Therefore, as this push toward smaller devices continues, the need for an effective yet efficient simulation tool capable of describing the new generation of devices becomes more important.

When devices shrink, the method of transport begins to change and quantum phenomena manifest themselves in the active region of these devices. The traditional semiclassical tools of device simulation are fast becoming limited. There have been efforts to expand such methods as Monte Carlo and drift-diffusion to incorporate quantum effects^{7,8} via an effective quantum potential⁹⁻¹¹ mainly due to quick and easy results when compared to the computational resources devoted to run some current software. This effort has been most notable in Monte Carlo where the effective potential has found some success in predicting some of the quantum phenomena arising in next generation devices,⁷ such as charge setback from the gate. However, the effective potential and other quantum corrective tools cannot account for some physical phenomena, such as tunneling,¹² which will become much more important as the channel length decreases. With this in mind, the focus should turn to fully quantum mechanical methods to correctly model short-channel devices. There

have been many suggestions for different quantum methods to model these devices including simple analytical models,^{13,14} Green's function approaches,¹⁵⁻¹⁷ coupled Schrodinger approaches,^{18,19} and Pauli master equation approaches.²⁰ However, in each of these methods, the length and the depth are modeled rigorously, while the third dimension is usually included through the assumption that there is no interesting physics to capture in this dimension. Therefore, the third dimension is usually treated using a basis expansion which is then included in the Hamiltonian. It is assumed also that the mode does not change shape as it propagates from the source of the device to the drain of the device. Other simulation proposals have simply assumed that only one subband in the orthogonal direction is occupied, therefore making higher-dimensional transport considerations unnecessary. This is certainly not a valid assumption. While it is true that the mode will not change shape as it propagates from the source to the drain, it is important to consider the fact that this mode will couple to other modes, which is required to treat some of the interesting physics. In the source of the device, the modes that are excited are three-dimensional (3D) in nature. These modes are then propagated from the 3D section of the source to the channel. The excitation of different modes changes as one approaches the drain, due to the large source-drain bias. Moreover, as the doping and the Fermi level in short channel MOSFETs increases, we can no longer assume that there is only one occupied subband even at the source.

In an effort to address these issues, we have developed a more complete simulation method. This is a full 3D, quantum simulation tool based on the use of recursive scattering matrices, which have been used to simulate the threshold voltage characteristics of realistic short-channel, fully depleted SOI MOSFET devices, which consist of both thin silicon film thicknesses and narrow channel widths. The combination of the thin silicon film thickness and the narrow channel widths will give rise to the aforementioned quantization in both transverse directions leading to an inseparable solution space. In the following section, we outline the mathematics of the model in some detail. We then turn to a dis-

^{a)}Electronic mail: matthew.gilbert@asu.edu

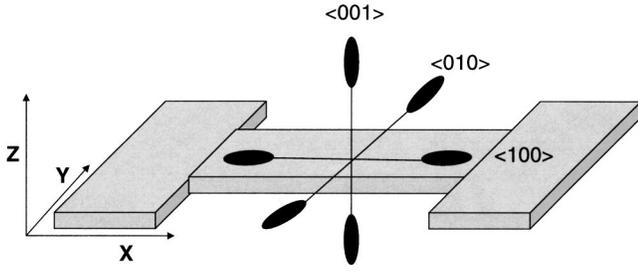


FIG. 1. Schematic showing the crystal orientation of the SOI MOSFET.

discussion of the SOI MOSFET structure and our results. Finally, we draw some important conclusions.

QUANTUM MODEL

We begin formulating our method by examining the crystal orientation of a silicon device, as shown in Fig. 1. We select the crystal orientation of the channel to be such that the current will flow along the $\langle 100 \rangle$ direction. This direction is chosen due to the fact that all of the other silicon surfaces will now line up with the coordinate axes. By this, we mean that the $\langle 010 \rangle$ direction lines up along the y direction and the $\langle 001 \rangle$ direction lines up with the z direction. It is important that the Cartesian axes are aligned with the principle axes of the six equivalent ellipsoids of the conduction band. The resulting quantization will split these ellipsoids into three pairs, and the choice of axes is most useful as the resulting Hamiltonian matrix will be diagonal. This allows for optimization in terms of the amount of memory needed to store the Hamiltonian, as well as the amount of computational time that is required to fill it. Had we chosen a different crystal direction the Hamiltonian of the system would no longer be diagonal. This, however, creates no new conceptual difficulties.

Now consider the discretization of the Schrödinger equation and the solution technique. The method that we use here is a numerically stabilized version of the recursive scattering matrix technique^{21,22} which proceeds as follows: We begin with the discrete Schrodinger equation in 3D

$$-\frac{\hbar^2}{2} \left(\frac{1}{m_x} \frac{d^2}{dx^2} + \frac{1}{m_y} \frac{d^2}{dy^2} + \frac{1}{m_z} \frac{d^2}{dz^2} \right) \psi(x,y,z) + V(x,y,z) \psi(x,y,z) = E \psi(x,y,z). \tag{1}$$

Here, we have taken the mass as a constant. While a more exact approach²³ would include the mass discontinuity across the silicon-oxide interface, the barrier height is such that the error in neglecting this is quite small. We note that one equation like Eq. (1) occurs for each pair of the six ellipsoids of the conduction band, and each pair has the appropriate mass values. We have labeled the mass corresponding to the principle coordinate axes, and these take on the values of m_L and m_T as appropriate to each valley. We then implement this on a finite difference grid with uniform spacing a . The derivatives appearing in the discrete Schrödinger equation are replaced with finite difference representations. The Schrödinger equation then reads

$$-t_x(\psi_{i+1,j,k} + \psi_{i-1,j,k}) - t_y(\psi_{i,j+1,k} + \psi_{i,j-1,k}) - t_z(\psi_{i,j,k+1} + \psi_{i,j,k-1}) + (V_{i,j,k} + 2t_x + 2t_y + 2t_z) \psi_{i,j,k} = E \psi_{i,j,k}, \tag{2}$$

where t_x , t_y , and t_z are the hopping energies

$$t_x = \frac{\hbar^2}{2m_x^* a^2},$$

$$t_y = \frac{\hbar^2}{2m_y^* a^2},$$

$$t_z = \frac{\hbar^2}{2m_z^* a^2}.$$

Each hopping energy corresponds to a specific direction in the silicon crystal. The fact that we are now dealing with three sets of hopping energies is quite important. This allows the true nature of the silicon crystal to enter into the transport by allowing coupling between the different energy surfaces. With this discrete form of the Schrödinger equation, we now seek to obtain the transfer matrices relating adjacent slices in our solution space. We begin by implementing Dirichlet boundary conditions on the upper and lower boundaries and expressing the wave function as a vector

$$\vec{\psi}_i = [\psi_{i,1} \dots \psi_{i,Ny-1} \psi_{i,Ny}]^T. \tag{4}$$

Thus far the implementation of the recursive scattering matrix method is identical for three dimensions as it is for the two-dimensional case. The first modification comes in the form of the wave function. In the 2D case, the wave function index N runs from the bottom of the slice to the top of the slice thereby creating the vector shown in Eq. (4). However, in the 3D case, the wavefunction at each point is no longer represented by a vector, but rather an extended vector defined as

$$\vec{\psi}_{i,EXT} = \begin{bmatrix} \Psi_{1,Ny} \\ \Psi_{2,Ny} \\ \dots \\ \Psi_{Nz,Ny} \end{bmatrix}. \tag{5}$$

In Eq. (5), each entry in the extended wave-function vector is itself a vector of length $1 \times Ny$. Therefore the extended wave-function forms a vector of size $1 \times Ny \times Nz$. Therefore, the discrete Schrödinger equation now may be written as

$$\vec{\psi}_{i+1} = \left(\frac{H_i - E}{t} \right) \vec{\psi}_i - \vec{\psi}_{i-1}. \tag{6}$$

With the discrete form of the Schrödinger equation realized, we must now define the Hamiltonian. To accomplish this we begin by defining it in parts. The first part of the Hamiltonian defines the energies and couplings between points along the rows in the y direction

After the modes have been propagated through the device, the resultant transmission may be found by

$$t = -(U_+ \lambda_+)^{-1} [C_1^{N+1} - U_+ (U_+ \lambda_+)^{-1}]^{-1}. \quad (24)$$

We now utilize the Landauer formula with finite temperature included, or

$$I(V_{sd}) = \frac{2e}{h} \int dE \times T(E) [f_s(E) - f_d(E)] \quad (25)$$

to calculate the current at different applied biases. In the above equation, f_s and f_d are the values of the Fermi functions at the source contact and drain contacts, respectively, and must be evaluated with the appropriate temperatures. In the computation of the integral in Eq. (25), we utilize ten uniformly distributed energy points with the effect of level broadening accounted for in the limits of the integration.

SOI MOSFET SIMULATION

Many other simulation techniques rely upon the assumption that the device under consideration is composed of a uniformly doped region in each of the source, channel, and drain. This is quickly becoming a very incorrect assumption, as the dimensions of the devices shrink. In the channel region of the devices that we consider there are approximately 3 million silicon atoms. For doping levels of 2×10^{18} and $5 \times 10^{18} \text{ cm}^{-3}$ we only expect to see approximately one and two acceptors present in the channel, respectively, for the narrow, thin, and short channels currently under consideration. Assuming that the channel region of the device is uniformly doped is erroneous. Therefore, we include the effects of random doping in the devices under consideration by utilizing the approach outlined by Wong and Taur.²⁴ In this approach, the silicon lattice is traversed and dopants are placed in the various regions if a generated random number is less than the ratio between the nominal doping of the region and the volume of silicon atoms in the region. The charge associated with the dopant atoms is then mapped back to the mesh and included in the self-consistent poisson solution for the potential. This dramatically changes the landscape of the potential as the normally smooth potential variation in the MOSFET is now replaced by a potential profile such as that shown in Fig. 2. Figure 2 is a slice of the potential profile in the xy -plane taken at a distance of 3 nm from the bottom of the silicon layer. The impinging electron waves must now navigate through large potential spikes present in the channel as they traverse from the source to the drain of the device. Moreover, we also see that there are large negative spikes in the source and the drain corresponding to the distribution of dopants. This changes the distribution of the electron density in the source and the drain as there are now energetically preferential sites in the source and the drain for the density.

We now utilize this technique to analyze the threshold voltage characteristics of the SOI MOSFET shown in Fig. 3. In this device, we have an oversized source and drain region which is doped $3 \times 10^{19} \text{ cm}^{-3}$ N -type. The dimensions of the source and the drain are 18 nm wide, 10 nm long, and 6 nm high corresponding to the thickness of the silicon layer. The

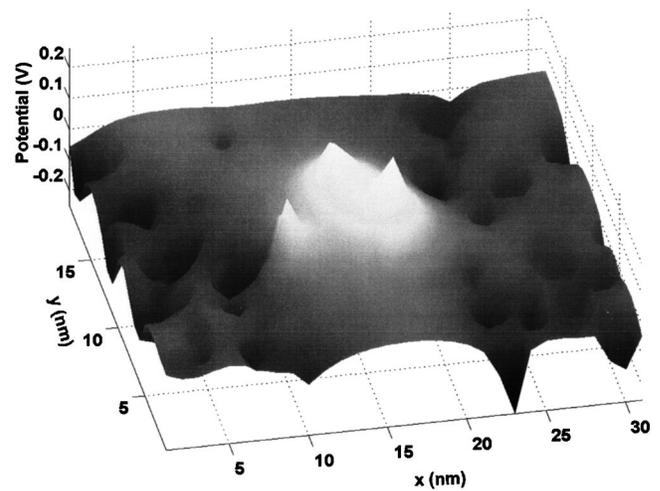


FIG. 2. Potential profile of a discretely doped SOI MOSFET taken at a distance of 3 nm from the bottom of the silicon layer. The oxide walls on either side of the channel have been removed so that the variation of the potential in the source, channel, and drain can be easily seen.

source and drain of the device have been given an exaggerated size to exacerbate the interaction of the modes excited in the source with the constriction present at the source-channel interface. The channel of our device is a P -type region. In this article, we consider dopings of 2×10^{18} and $5 \times 10^{18} \text{ cm}^{-3}$ for the channel. The channel is 10 nm in length, 6 nm in height, and 8 nm in width as shown in Fig. 3. In this type of device, we should see the benefit of including the third dimension in the calculation as there is undoubtedly quantization in the y -direction of the channel region which could not be properly accounted for through a simple basis expansion. For the following simulations, we have used a uniform grid spacing of 1 nm in all three spatial directions. The temperature of the system is set to 300 K and source-drain voltage of 10 mV.

The $I_d - V_g$ characteristics for a doping of $2 \times 10^{18} \text{ cm}^{-3}$ are shown in Fig. 4 for three different devices.

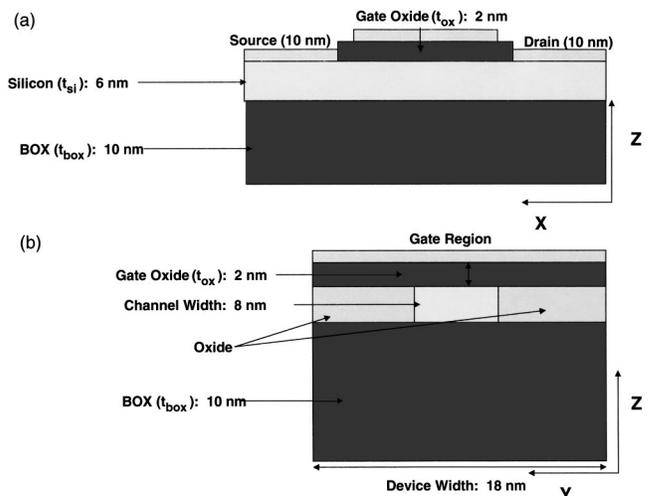


FIG. 3. (a) xz -plane schematic of the SOI MOSFET under consideration. (b) yz -plane schematic of the SOI MOSFET under consideration.

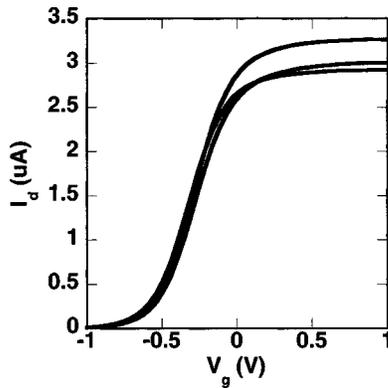


FIG. 4. I_d-V_g characteristics for a doping of $2 \times 10^{18} \text{ cm}^{-3}$. The curves represent different discretely doped cases.

These three discretely doped cases show an average variation of 0.013 V around an average value of -0.674 V with a maximum deviation of 0.05 V.

In Fig. 5, we plot the density distribution in the xy -plane for various gate voltages. In Fig. 5(a), we plot the density in the xy -plane for a gate voltage of -1 V which is below threshold. The density is shown at a distance of 3 nm above the bottom of the silicon layer in the device. As we can see, the vast majority of the excited density is reflected back from both the constriction and the source-channel interface. The reflection from the source-channel interface is expected as for this set of devices the dopants are mainly located at the source side of the channel, thereby providing a better barrier to the density. The nearer the dopants are to the source of the device, the more interaction they have with the source density, without interference from the drain contact. Therefore, the dopants give rise to larger reflections and larger effects on the threshold voltage. Moreover, we find that this reflection forms a standing wave pattern in the x and y directions in the source of the device. We can see that at this gate voltage there is virtually no density that is capable of propagating in the channel and, therefore, all of the resulting cur-

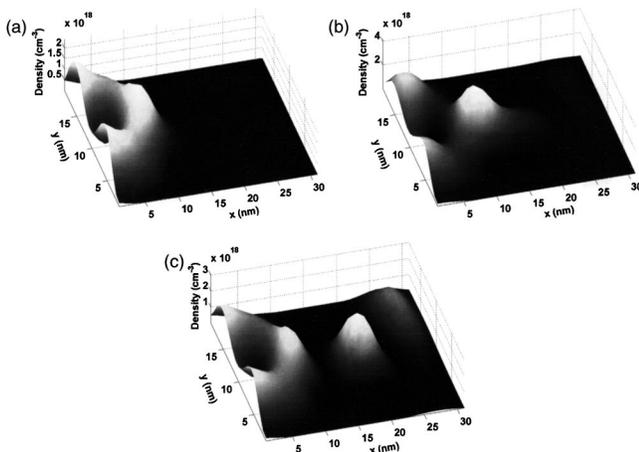


FIG. 5. Density in the xy -plane for a doping of $2 \times 10^{18} \text{ cm}^{-3}$ at a distance of 3 nm from the bottom of the silicon layer for a gate voltage of (a) -1 V , (b) -0.3 V , and (c) 1 V .

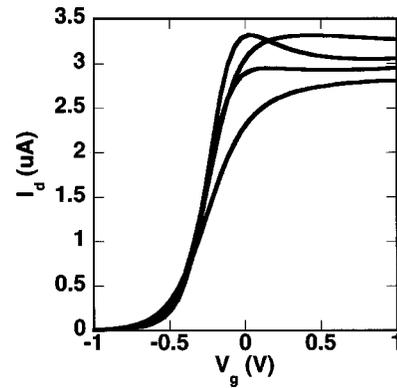


FIG. 6. I_d-V_g characteristics for a doping of $5 \times 10^{18} \text{ cm}^{-3}$. The curves represent different discretely doped cases.

rent is due to the tunneling of the density from the source to the drain.

In Fig. 5(b), we plot the density in the xy -plane at a distance of 3 nm from the bottom of the silicon layer for a gate voltage corresponding to -0.3 V , which is above threshold. We can see that the gate voltage has sufficiently decreased the barrier, and we now have propagating channel density which is beginning to then exit the channel. The density position in the drain corresponds to the dopant placement.

Finally, in Fig. 5(c), we plot the density in the xy -plane at a distance of 3 nm from the bottom of the silicon layer for a gate voltage of 1 V . We now have a fully propagating mode in the channel of the device which has formed a standing wave as the density reflects from the channel-drain opening and interferes with the incident density in the channel. We also find that the density that has exited the channel, which is noticeably less than is present in the source, still forms pools corresponding to the dopants locations in the drain. The resulting current is now saturated at this gate voltage as we have populated the channel with one full mode and the density in the channel is now maximized.

When the doping of the channel is raised from 2×10^{18} to $5 \times 10^{18} \text{ cm}^{-3}$, an additional dopant appears in the channel of the MOSFET. This will add an additional spike to the channel potential, which will in turn interact with the density causing more interference and additional reflections. Therefore, we should expect to see more of a discrepancy when we examine the threshold voltages of these devices. In Fig. 6, we plot the I_d-V_g characteristics corresponding to a channel doping of $5 \times 10^{18} \text{ cm}^{-3}$ for four discretely doped devices. The inclusion of the real doping profile results in increased interference effects due to the inclusion of the additional dopant. While the average threshold voltage is -0.49 V these interference effects cause an average variation of -0.1 V in the threshold voltage with a maximum variation of -0.14 V . We immediately see that the addition of one more dopant in the channel causes a greater degree of variation in the threshold voltage due to the increased reflections and interference than in the case of lighter doped devices. In Fig. 7, we once again plot the resultant densities in the xy -plane for various gate voltages to add additional insight into the resultant curves.

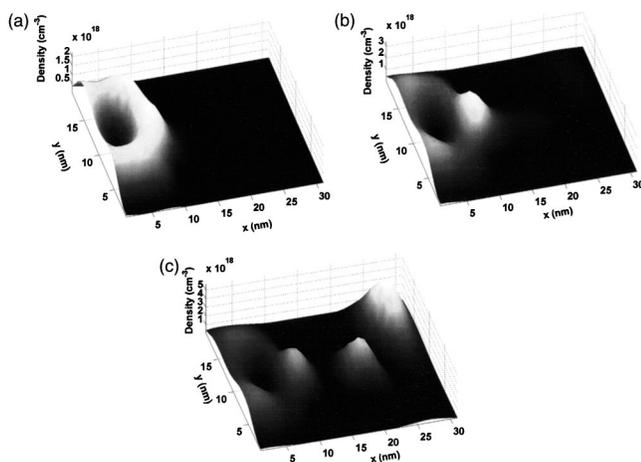


FIG. 7. (a) Density in the xy -plane for a doping of $5 \times 10^{18} \text{ cm}^{-3}$ at a distance of 3 nm from the bottom of the silicon layer for a gate voltage of (a) -1 V , (b) -0.3 V , and (c) 1 V .

In Fig. 7(a), we plot the density in the xy -plane at a distance of 3 nm from the bottom of the silicon layer with a gate voltage of -1 V . As in the case of the lighter doping, we find the same type of standing wave pattern in source. As expected, we see a great deal of reflection from the source-channel barrier and from the constriction. Nevertheless, when we compare this to Fig. 5(a), we find that the density is not quite as peaked at the source-channel barrier as in the previous case. This indicates that, for this sample, the location of the dopants is further in the channel than for the previous case. This allows the density to penetrate deeper into the channel before it is reflected. Nevertheless, as in the previous case, all of the current resulting at this gate voltage is due to tunneling from the source to the drain of the device.

In Fig. 7(b) we look at the density in the xy -plane in the device at a distance of 3 nm from the bottom of the silicon layer at a gate voltage of -0.3 V . Once again the device is above threshold and we see much more of the density entering the channel and propagating than in Fig. 5(b). However, we notice that the density is diminished from the case of lighter doping as is expected with the additional dopant in the channel. However, we can see some of the density exiting the channel and spreading out into the drain region according to the locations of the dopants in the drain.

In Fig. 7(c), we plot the density in the xy plane at a distance of 3 nm from the bottom of the silicon layer for a gate voltage of 1 V . Here we see a great departure from Fig. 5(c) in terms of the form of the channel density and the drain density. In Fig. 7(c), we see that there is much more channel density than in the lighter-doped case and that the resultant standing wave pattern is altered from the additional interactions with the channel dopants. The additional interactions then trap more of the density in the channel than is seen in lighter doped devices. With additional density trapped in the channel of the device, we see that the density that exits the channel is reduced. However, it does spread out more smoothly, but there is still a variation caused by the positions of the dopants in the drain. In the case of Fig. 5(c) more of the dopants were concentrated in the corner of the drain causing density to settle there, while in Fig. 7(c) the dopants

were more distributed leading to the greater spread of density in the drain.

CONCLUSION

We have proposed an efficient, fully quantum mechanical and straightforward method to calculate the transport in nanoscale SOI MOSFET devices in three dimensions. We have then applied the method to the calculation of threshold voltages in ballistic SOI MOSFET devices with various discrete doping densities. This has revealed the necessity for such a model as other methods cannot correctly capture and account for the interference effects we see in these realistic devices. We have seen that the position of the dopant atoms has a significant effect on the resultant device characteristics. This is a result of the interference that the potential spikes produce. Dopants that are positioned closer to the source of the device have a greater effect on the threshold voltage that do dopants positioned deeper in the channel of the device due to increased interaction with the waves incident at the source-channel interface, causing additional reflections. This can produce source-drain resonances as seen in Fig. 6, where two dopants have been positioned at the very boundary between the source and the channel at a depth corresponding to that of maximal density causing a resonant state to form. The fact that the transport properties can be dramatically affected not only by the position in the xy plane, but by their position in the z direction as well, adds additional importance to this work. Further, we see that the positions of the dopants in the source and the drain cause pools of electron density to form. This is again due to the negative potential spikes caused by the donors forming states for the density. This leads to noticeable variations in the density distribution in the source and, particularly, the drain. All of the preceding features caused by the interaction of the waves with the potential spikes present in the system illustrate the growing importance of the mechanisms of decoherence in ultra-short devices, which will be treated more completely in future work.

ACKNOWLEDGMENTS

The authors wish to thank R. Akis for useful discussions. This work is supported by the Office of Naval Research.

- ¹R. Chau *et al.*, Device Research Conference, 2003.
- ²B. Yu, H. Wong, A. Joshi, Q. Xiang, E. Ibok, and M.-R. Lin, Tech. Dig.-Int. Electron Devices Meet. **2001**, 937.
- ³B. Doris, M. Jeong, T. Kanarsky, Y. Zhang, R. A. Roy, O. Dokumaci, Z. Ren, F.-F. Jamin, L. Shi, W. Natzle, H.-J. Huang, J. Mezzapelle, A. Mocuta, S. Womack, M. Gribelyuk, E. C. Jones, R. J. Miller, H.-S. P. Wong, and W. Haensch, Tech. Dig.-Int. Electron Devices Meet. **2002**, 267.
- ⁴H. Namatsu, K. Kurihara, M. Nagase, and T. Makino, Appl. Phys. Lett. **70**, 619 (1997).
- ⁵J. Kedzierski, J. Bokor, and E. Anderson, J. Vac. Sci. Technol. B **17**, 3244 (1999).
- ⁶Y. Cui and C. M. Lieber, Science **291**, 851 (2001).
- ⁷S. M. Ramey and D. K. Ferry, Physica B **314**, 350 (2002).
- ⁸C. L. Gardner and C. Ringhofer, Phys. Rev. E **53**, 157 (1996).
- ⁹D. K. Ferry, Superlattices Microstruct. **28**, 419 (2000).
- ¹⁰J. R. Zhou and D. K. Ferry, IEEE Trans. Electron Devices **40**, 421 (1993).
- ¹¹A. Asenov, G. Slavcheva, A. R. Brown, J. H. Davies, and S. Saini, IEEE

- Trans. Electron Devices **48**, 722 (2001).
- ¹²D. K. Ferry, S. M. Ramey, L. Shifren, and R. Akis, J. Comp. Elect. **1**, 59 (2002).
- ¹³F. Assad, Z. Ren, D. Vasileska, S. Datta, and M. Lundstrom, IEEE Trans. Electron Devices **47**, 232 (2000).
- ¹⁴K. Natori, J. Appl. Phys. **76**, 3730 (1994).
- ¹⁵R. Lake, G. Klimeck, R. C. Bowen, and D. Jovanovic, J. Appl. Phys. **81**, 7845 (1997).
- ¹⁶J. Knoch, B. Lengeler, and J. Appenzeller, IEEE Trans. Electron Devices **49**, 1212 (2002).
- ¹⁷S. Datta, Superlattices Microstruct. **28**, 253 (2000).
- ¹⁸F. G. Pikus and K. K. Likharev, Appl. Phys. Lett. **71**, 3661 (1997).
- ¹⁹S. E. Laux, A. Kumar, and M. V. Fischetti, IEEE Trans. Nano **1**, 255 (2002).
- ²⁰M. V. Fischetti, J. Appl. Phys. **83**, 270 (1988).
- ²¹T. Usuki, M. Saito, M. Takatsu, R. A. Kiehl, and N. Yokoyama, Phys. Rev. B **52**, 8244 (1995).
- ²²R. Akis, D. K. Ferry, and J. P. Bird, Phys. Rev. B **54**, 17705 (1996).
- ²³G. Bastard, Phys. Rev. B **24**, 5693 (1981).
- ²⁴H. S. Wong, Y. Taur, Int. Electron Devices Meet. **1993**, 705.