VLSI DESIGN
1998, Vol. 8, Nos. (1-4), pp. 429-435
Reprints available directly from the publisher
Photocopying permitted by license only

Study of Electron Velocity Overshoot in NMOS Inversion Layers

WEI-KAI SHIH, SRINIVAS JALLEPALLI, MAHBUB RASHED, CHRISTINE M. MAZIAR* and AL. F. TASCH JR.

Microelectronics Research Center, The University of Texas at Austin, Austin, TX 78712, USA

Non-local electron transport in *n*MOSFET inversion layers has been studied by Monte Carlo (MC) simulations. Inversion layer quantization has been explicitly included in the calculation of density of states and scattering rate for low-energy electrons while bulk band structure is used to describe the transport of more energetic electrons. For uniform, high-lateral field conditions, the effects of quantization are less pronounced due to the depopulation of electrons in the lower-lying subbands. On the other hand, Monte Carlo results for carrier transport in spatially varying lateral fields (such as those in the inversion layer of MOSFETs) clearly indicate that depopulation of the low-lying subbands is less evident in the non-local transport regime. Quasi-2D simulations have shown that, at high transverse effective field, the inclusion of a quantization domain does have an impact on the calculated spatial velocity transient .

Keywords: *n*MOS, inversion layer, velocity overshoot, quantization, Monte Carlo simulation

INTRODUCTION

With the continued scaling of the feature size of MOS devices, carrier transport in the MOS inversion layers has entered the regime where non-local effects are no longer negligible. In a previous study, velocity overshoot was found to account for approximately 20% of the disagreement in drain current between measurement and drift-diffusion simulation in a 0.12 μm SOI MOSFET [1]. Despite a few pioneering studies on probing the electron spatial transient velocity in compound semiconductor devices [2], a mature experimental technique that allows the direct measurement of carrier velocity in silicon MOSFETs has not been available. In order to accurately express these non-local effects in physically based models, the Monte Carlo (MC) technique has become increasingly indispensable for device scientists and engineers. In conventional MC tools, effects due to channel quantization resulting from the steep bulk band bending are routinely ignored. With phenomenological surface-roughness scattering models, some conven-

* Corresponding author.

tional MC tools [3,4] using bulk silicon band structure have demonstrated the ability to reproduce the experimental inversion-layer velocity-field characteristics in the local transport regime within a limited range of temperatures. However, the validity and accuracy of this type of approach remains questionable in the non-local transport regime.

In favor of the semi-classical approach, one might argue that non-local transport is most generally observed under conditions of high electric field in the transport direction. A condition in which the inversion layer electrons populate higher energy states and the effects of quantization are less pronounced. However, the issue regarding how fast the 2DEG can respond to an abruptly changing field, hence become more classical-like, has not been investigated. The purpose of this paper is to qualitatively address these issues with simulations performed on quasi-2D test structures. As will be shown, inversion-layer quantization does have an impact on the carrier average velocity in the non-local regime, despite the presence of high lateral electric field.

THE MONTE CARLO SIMULATOR

The MC program used in this work is an integrated tool consisting of two simulation domains that partition the entire inversion-layer electron population into quantum (2D) and classical (3D) components. The 2D domain is used to simulate transport of low-energy electrons by directly solving the multi-subband 2D Boltzmann transport equation [5]. The 3D domain, employing three fitted conduction bands, is used for treating more energetic electrons that have motions that are more classical-like [6]. The design of the integrated tool allows the 3D domain to execute without invoking the 2D domain. In its standalone mode, the 3D domain takes into account phonon scattering and impact ionization and has been calibrated to reproduce uniform transport characteristics in bulk silicon. It also employs a phenomenological model, which was previously validated, to take into account surface roughness scattering [4]. In the 2D domain, the effective-mass approximation (EMA) including bulk nonparabolicity is used and effects of scattering due to bulk phonon and surface roughness have been accounted for. Coulomb scattering is greatly suppressed due to carrier screening in the strong-inversion regime and has been ignored in the present treatment. The envelope functions obtained within the EMA are used to calculate the electron-phonon form factors needed for computing the phonon scattering rate. The parameters used in phonon and surface roughness scattering models have been adjusted to reproduce the experimental mobility values. As shown in Figure 1, the inversion layer ohmic mobilities obtained with the 2D domain agree well with those obtained in previous experimental studies (at 413 K and 77 K) and the extensively validated UT-mobility model (at 300 K) [7,8]. It should be emphasized that all the MC mobility values are obtained with a single set of scattering model parameters.

Physical boundary conditions between the 2D and 3D domains are employed. The implementation is similar to that proposed by Fischetti *et al.* [9] except for the use of a 2D/3D boundary (in the energy space) more consistent with the EMA and our previous pseudo-potential calculation [10]. Within the EMA, the envelope function of a given

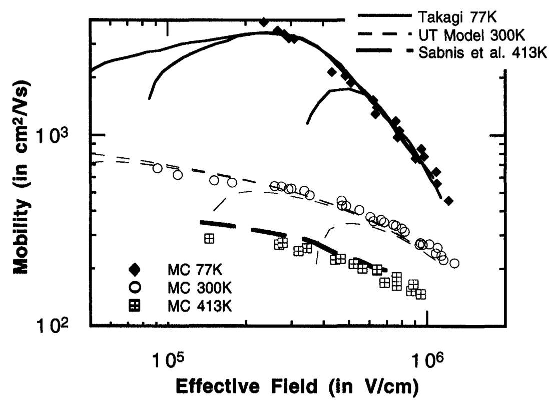


FIGURE 1 low-field mobility obtained with the single particle MC compared with experimental and UT-mobility model values.

subband is independent of the parallel momentum. Such an insensitivity, confirmed in [10], indicates that an electron with a large parallel momentum is as well quantized as that with a small parallel momentum in the same subband, despite the difference between their total energies. Instead of defining the 2D/3D boundary in terms of total energy, as suggested in [9], we use a criterion based on the electron total energy in the quantization (z-) direction. The concept of "buffer subbands" suggested in [9] has been adopted to allow 2D electrons to enter the 3D domain via scattering. The six lowest-lying subbands and 64 buffer subbands have been used. Applying the above criteria, as illustrated in Figure 2, a 3D electron is converted to the quantum domain when the following two conditions are met: (1) The electron is located in the lowest bulk valley. (2) The total energy of the 3D electron in the z-direction (E_z) is less than $E_z^{(th)}$, where $E_z^{(th)}$, is defined to be the edge of the lowest buffer subband and E_z is evaluated within the parabolic approximation. On the other hand, a 2D electron is converted to the classical domain when it enters one of the buffer subbands via intersubband scattering. Conservation of the electron in-plane momentum is imposed in the conversion process. Since the energy spectrum in the 3D domain is a continuum, energy conservation is enforced for the 2D to 3D transition. However, energy cannot be conserved during the 3D to 2D transition because only discrete subband levels are available in the 2D domain. In this case, the subband that minimizes the energy mismatch is chosen. If the energy conservation law were strictly enforced during the simulation, the 2D domain would be depleted of electrons at steady state under uniform, high-lateral-field condition, which is clearly unphysical. To circumvent this problem without violating the energy conservation, one needs to directly calculate the scattering rate between the 3D and the 2D states. This alternative approach is expected to be comparable to our current implementation since the 2D density of states obtained with the quantum treatment is expected to be comparable to that obtained with the classical treatment at sufficiently high energy (the correspondence principle).

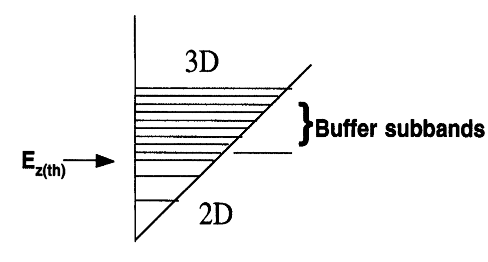


FIGURE 2 Illustration of the partition of simulation domain in the inversion-layer quantum well.

With the integrated MC tool, the velocity-versus-field relations obtained at various transverse E_{eff} are observed to agree well with the UT-mobility model (see Fig. 3). As expected, the calculated saturation velocity is the same as that in bulk silicon since at high field, the 3D population dominates and surface roughness scattering is less effective. This is further evidenced in Figure 4, where the populations of electrons in the quantum and classical domains are separately plotted with respect to the lateral field. At low lateral field, the 3D electrons have a larger population for the less inverted channel due to the weaker quantum confinement at lower E_{eff}. For the same reason, the 3D component takes over the 2D component at a lower lateral field in the less inverted channel.

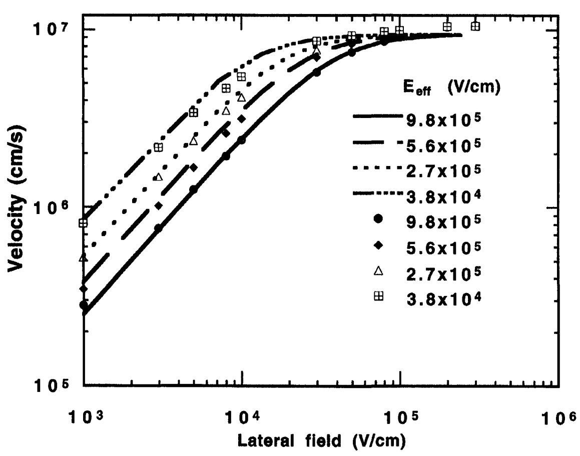


FIGURE 3 Velocity versus field obtained from the MC tool including both 2D and 3D domains (symbols) compared to that obtained with the UT mobility model (lines).

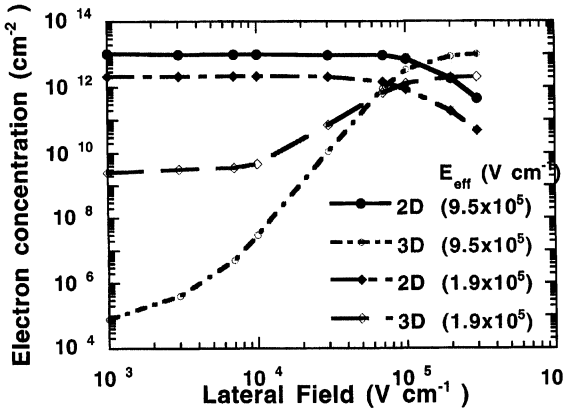
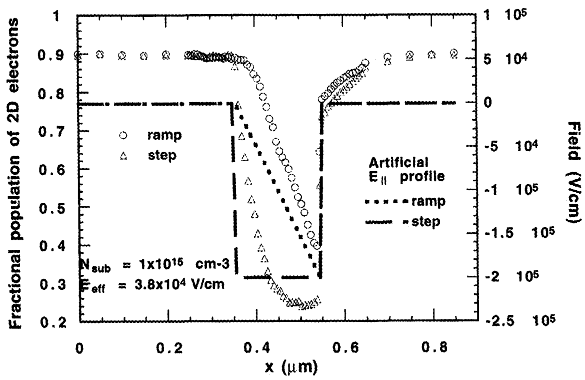


FIGURE 4 Populations of electrons in the quantum and classical domains.

SIMULATION PROCEDURE AND RESULT

Quasi-2D structures with uniform MOS inversion layers subjected to artificial lateral field profiles that vary abruptly along the channel (x-) direction are studied. A 1D Schrödinger-Poisson iteration following classical drift-diffusion simulation is used to obtain the self-consistent subband dispersion and scattering rate for given substrate doping levels (N_{sub}) and gate biases (V_G). Either a step or a ramp profile is chosen for the lateral field across a 0.9 μm channel. The low-field region is sufficiently long for transport to reach the steady state corresponding to the local field. A periodic boundary condition is applied to both ends of the channel with carrier injection conserving both momentum and subband index. To directly assess velocity overshoot due solely to the applied field, Poisson feedback resulting from the carrier redistribution along the channel has been ignored.

To appreciate the importance of the quantum domain in the non-local transport regime, the fractional electron population in the 2D domain under step-like and ramp-shaped lateral field profiles is examined. The ramp field profile corresponds to a 2V voltage drop across the 0.2 μm channel and roughly conforms with the power-supply scaling trend associated with the scaling of device dimensions. The step profile represents a worse case scenario where a power supply of 4V is applied across the channel. In Figure 5, it is observed that the 2D electron population in the high lateral-field region remains significant (>40%) in the overshoot regions (around $x = 0.5$ μm for the ramp profile and $x = 0.35$ μm for the step profile), suggesting that quantization might be of importance in the non-local transport regime in spite of the presence of the high lateral fields. It should be noted that, had there been no non-local effects, inspection of Figure 4 would have lead to the estimated fractional 2D populations at these locations of less than 20% even without taking into account



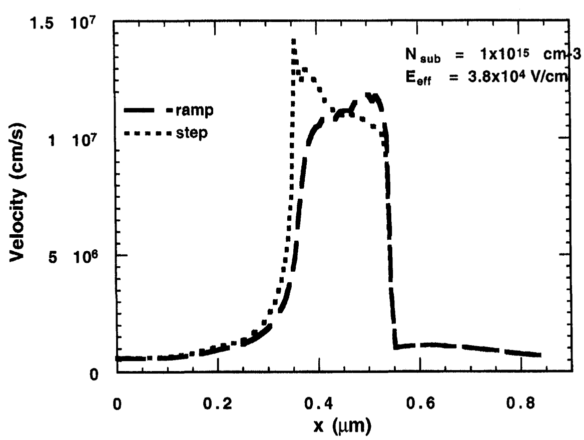


FIGURE 5 (a) Fractional population of electrons in the 2D simulation domain (symbols). Applied lateral field profiles are also shown (lines). (b) Velocity profiles along the lateral direction obtained with the lateral field profiles in (a). Overshoot occurs near $x = 0.5$ μm and $x = 0.35$ μm for the step profile and ramped profile, respectively.

the fact that the E_{eff} used in Figure 5 is lower than those used in Figure 4.

With the same ramped lateral field profile shown in Figure 5(a), Figure 6 shows that the transient velocity obtained at different E_{eff}'s and N_{sub}'s. In general, the transient velocity seems to have a stronger dependence on E_{eff} than on N_{sub}. In the low-field region, this is expected since the quantum domain explains universal mobility quite well. The universality of velocity (and mobility) dependence on E_{eff} seems to be retained along the rising edge of the transient velocity until the latter becomes sufficiently high, where N_{sub} starts to have some impact. At higher E_{eff}, the lower velocity observed on the rising edge can be qualitatively explained by the lower mobility of the ground subband due to stronger surface roughness scattering. Near the velocity peak, on the other hand, a simple explanation for the different velocities is not available since all subbands in the quantum domain and the classical domain are equally significant.

To see the quantization effects on the MC-simulated velocity profile, Figures 7(a) and 7(b) compare the velocity profiles obtained with conventional MC with those obtained with the integrated MC tool at $E_{\mathrm{eff}} = 5.6 \times 10^5$ V/cm and 9.8×10^5 V/cm, respectively. In both cases, lower

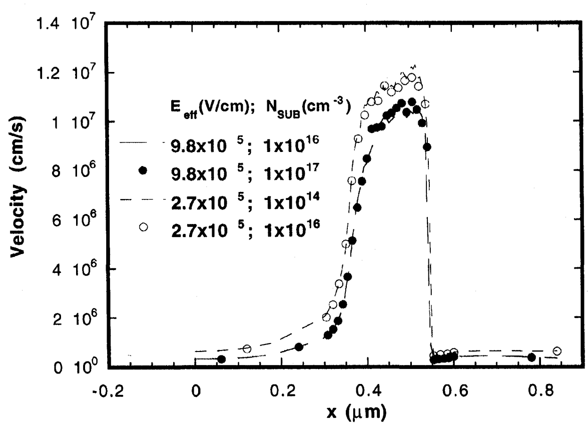


FIGURE 6 Velocity along the channel with ramped field profile at two different E_{eff}. More pronounced overshoot is seen at the lower E_{eff}. For each E_{eff}, two different substrate doping levels are used.

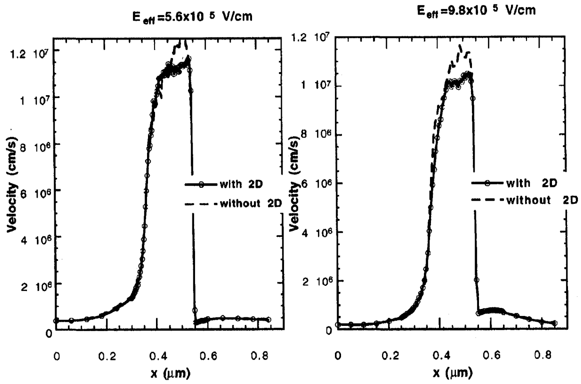


FIGURE 7 Velocity profiles obtained from conventional MC simulations compared to those obtained from the integrated tool. Different E_{eff}'s and same substrate doping level (1×10^{16} cm^{-3}) are used in (a) and (b).

peak velocities are predicted by the integrated MC. Although the difference between two MC results slightly increases as quantization becomes stronger, we are reluctant make the intuitively appealing statement that the difference is indeed enhanced as channel becomes more strongly inverted. If such a trend does exist, the impact of including channel quantization on the predicted velocity overshoot in conventional deep-submicron MOSFETs might be limited as the quantum confinement is weaker near the drain. However, as aggressive device design with asymmetric channel doping profile has been carried out [11] to take advantage of velocity overshoot near the source, where the silicon surface is strongly inverted, quantization effects on the spatial velocity transient is likely to be more significant in these devices.

CONCLUSION

Quasi-2D Monte Carlo simulations that take into account the effects due to size quantization have been used to study carrier velocity overshoot in nMOS inversion layers. Non-local effects are seen not only in carrier velocity but also in the carrier population. With a given ramp-field profile, the spatial transient velocity is observed to depend on the transverse effective field more strongly than on the substrate doping. In this work, more pronounced velocity overshoot is observed at lower E_{eff}.

The impact of quantization on simulated velocity profiles has been observed by comparing results obtained with the conventional MC to those obtained with the integrated MC tool. While quantization seems to have a slightly larger impact at higher E_{eff}, more thorough study is needed to support such a viewpoint, given the uncertainties in models used in both MC tools. Since the comparison is based on two different simulation models, each of which contains certain number of parameters calibrated such that uniform-condition transport characteristics can be reproduced, we cannot exclude the possibility that there exists a different region in the parameter space where one can draw a conclusion different from ours. In fact, even among the currently existing well-calibrated classical MC codes, disagreement in spatial transient velocity is very likely to exist. The same statement can also apply to the quantum MC codes. More studies are needed to clarify the issues related to the impact of quantization on velocity overshoot.

Acknowledgements

This work was supported, in part by the Semiconductor Research Corporation (SRC), the Joint services Electronics Program (F49620-95-C-0045), the Texas Advanced Technology Program (TATP) and Motorola. The United States Government is authorized to distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

References

[1] Assaderaghi, F., Sinitsky, D., Gaw, H., Bokor, J., Ko, P. K. and Hu, C. (1994). "Saturation velocity and velocity overshoot of inversion layer electron and holes", *IEDM Tech. Dig.*, pp. 479–482.

[2] Grann, E. D., Tsen, K. T., Sankey, O. F., Ferry, D. K., Salvador, A., Botcharev, A. and Morkoc, H. (1995). "Electron velocity overshoot in a GaAs-based p-i-n nanostructure semiconductor observed by transient subpicosecond Raman spectroscopy", *Appl. Phys. Lett.*, **67**, 1760–1762.

[3] Sangiorgi, E. and Pinto, M. R. (1992). "A semi-empirical model of surface scattering for Monte Carlo simulation of silicon *n*-MOSFETs", *IEEE Trans. Electron Devices*, **39**, 356–361.

[4] Jallepalli, S. (1992). "Surface scattering in a self-consistent Monte Carlo analysis of electron transport in *n*-MOSFETs", Master Thesis, The University of Texas at Austin.

[5] Shih, W.-K., Jallepalli, S., Yeap, C.-F., Rashed, M., Maziar, C. M. and Tasch, A. F. (1995). "A Monte Carlo study of electron transport in Silicon nMOSFET inversion layers", presented at International Workshop on Computational Electronics, Phoenix, Arizona, USA.

[6] Wang, X., Chandramouli, V., Maziar, C. M. and Tasch, A. F. (1993). "Simulation program suitable for hot carrier studies: an efficient multiband Monte Carlo model using both full and analytical band structure for silicon", *J. Appl. Phys.*, **71**, 3339–3347.

[7] Shin, H., Yeric, C. M., Maziar, C. M. and Tasch, A. F. (1991). "Physically-based models for effective mobility and local-field mobility of electrons in MOS inversion layers", *Solid-State Electronics*, **34**, 545.

[8] Khan, S. A., Hasnat, K., Tasch, A. F. and Maziar, C. M. (1995). "Detailed evaluation of different inversion layer electron and hole mobility models", presented at *Proceeding of the Eleventh Biennial University Government Industry Microelectronics Meeting*, Austin, Texas.

[9] Fischetti, M. V. and Laux, S. E. (1993). "Monte Carlo study of electron transport in silicon inversion layers", *Phys. Rev. B*, **48**, 2244.

[10] Jallepalli, S., Bude, J., Shih, W.-K., Pinto, M. R. and Maziar, C. M., "Quantization of electrons and holes in silicon inversion layers beyond the effective mass approximation", Submitted to *Phys. Rev. B*.

[11] Hiroki, A., Odanaka, S. and Hori, A. (1995). "A high performance 0.1 μm MOSFET with asymmetric channel profile", *IEDM Tech. Dig.*, pp. 439–442.

Authors' Biographies

Wei-Kai Shih received the B.S. degree in physics from National Taiwan University in 1989 and M.A. degree in physics from the University of Texas at Austin in 1994. He is currently working towards a Ph.D. degree in electrical engineering at the University of Texas at Austin. His research interests include theoretical and computational study of hot-carrier effects, non-local phenomena, effects due to inversion-layer quantization on carrier transport and development of physical models for hydrodynamic and drift-diffusion simulation of deep submicron silicon MOSFETs.

Srinivas Jallepalli was born in Srikakulam, India. He received the B. Tech degree from the Indian Institute of Technology, Madras, in 1991 and the M.S.E. and Ph.D. degrees from the

University of Texas at Austin, in 1993 and 1996, respectively. He is currently with the Advanced Device Development group at STL, Motorola, Austin, TX. His research interest include device physics and modeling with special emphasis on the use of the Monte Carlo tool for advanced transport analysis and for studying the physics and effects of carrier transport in deep submicron MOS structures.

Mahbub Rashed received his Bachelor of Science in Electrical Engineering from Bangladesh University of Engineering and Technology in May, 1991. He joined in the Electrical and Computer Engineering department of the University of Texas at Austin in January 1992. He obtained his Masters degree from this department in the summer of 1993. He continued his education at the same university, pursuing a Ph.D. degree in Electrical Engineering. He focused on the applications of Monte Carlo tools for studying carrier transport in silicon and strained-Si/$Si_{1-x}Ge_x$ based devices. In December, 1996 he was awarded the Ph.D. degree. He is currently with the Semiconductor Technologies Laboratory, Motorola, Austin.

Christine Maziar received the B.S.E.E. (with Highest Distinction), the M.S.E.E. and the Ph.D. degrees from Purdue University in 1981, 1984 and 1986. In January of 1987, she joined the faculty of the University of Texas at Austin, where she is a Professor of Electrical and Computer Engineering and Vice Provost. She holds the Archie W. Straition Endowed Teaching Fellowship in Electrical and Computer Engineering. Her research interests include modeling of charge transport in high performance device structures, semiconductor device physics and device simulator enhancement. She has authored or coauthored over 100 journal or conference publications. She is a member of IEEE, ASEE, APS, Eta Kappa Nu, Tau Beta Pi and Sigma Pi Sigma.

Al. F. Tasch received his Ph.D. degree in Physics in 1969 form the University of Illinois, Urbana-Champaign. He worked in industry for 17 years at Texas Instruments and Motorola and joined the faculty of the Department of Electrical and Computer Engineering at the University of Texas at Austin in 1986 as a Chair Professor. He has been awarded 38 U.S. patents and was elected to the National Academy of Engineering in 1989.