Breakdown of Universal Mobility Curves in Sub-100-nm MOSFETs

Savas Kaya, Asen Asenov, *Member, IEEE*, and Scott Roy

***Abstract*—We explore the breakdown of universal mobility behavior in sub-100-nm Si MOSFETs, using a novel three-dimensional (3-D) statistical simulation approach. In this approach, carrier trajectories in the bulk are treated via 3-D Brownian dynamics, while the carrier-interface roughness scattering is treated using a novel empirical model. Owing to the high efficiency of the transport kernel, effective mobility in 3-D MOSFETs with realistic Si–SiO_2 interfaces reconstructed from a Gaussian or exponential correlation function can be simulated in a statistical manner. We first demonstrate a practical calibration procedure for the interface mobility and affirm the universal behavior in the long channel limit. Next, effective mobility in ensembles of MOSFETs with a gate length down to 10 nm is investigated. It is found that the random-discrete nature of the Si–SiO_2 interface leads to a distribution of carrier mobility below the interface, which can deviate considerably from universal mobility curves when $L_{\mathrm{gate}} < 6\Lambda$, where Λ is the correlation length for the SiO_2 interface.**

***Index Terms*—Brownian dynamics, device simulation, interface roughness scattering, MOSFET, universal mobility.**

I. Introduction

THE miniaturization of Si MOSFETs has surpassed even the expectations of SIA Roadmap and should reach the sub-25-nm scale by 2007 [2]. The acceleration of the Roadmap has drawn highlighted attention to intrinsic parameter fluctuations in MOSFETs, because they constitute one of the main "showstoppers" for conventional CMOS scaling. Although intrinsic fluctuations often associated with discrete dopant charges [3], the atomicity of matter can also introduce substantial variations in the individual device characteristics [4]. The gate oxide thickness in 20-nm Intel devices [1], for example, is composed of only *three* silicon atomic layers. An interface roughness of even *one* layer then would have major implications for device electrostatistics [5] as well as transport adjacent to the Si–SiO_2 interface [6]. In particular, there is no clarification as to whether the random nature of the oxide interface can pose a significant threat to the universal mobility behavior [7]–[9], which has proved very useful in commercial device simulators used in conventional Si CMOS designs.

We recently demonstrated an effective approach to address the transport aspect of this problem and established a three-dimensional (3-D) statistical simulation framework [6]. The approach we have adapted is based on Brownian particle simulations in 3-D, uses realistic representation of the oxide interface with appropriate spectral models, and has been extremely efficient. In this work, we extend this approach to device simulations in order to investigate the impact of a random Si–SiO_2 pattern on the long-established universal mobility behavior in Si MOSFETs. We simulate effective mobility in MOSFETs down to 10 nm in gate length and show that significant departure from universal curves may occur when gate length is reduced below $\sim 6\Lambda$ where Λ is the correlation length assumed for the Si–SiO_2 interface. In the following, we first introduce basic features of the simulator, then describe device simulations, and finally discuss the simulation results. The major conclusions drawn are summarized at the end.

II. The Simulation Approach

In order to treat the universal mobility statistically, three main requirements may be identified: an accurate 3-D model for the interface based on its measured properties, a very efficient 3-D particle simulator, and an empirical treatment of surface scattering. In this section, we deal with these requirements and demonstrate an efficient simulator, which is then deployed to address the breakdown of universal mobility in decanano MOSFETs.

A. Brownian Transport Kernel

The requirements for an *exact* treatment of the transport problem associated with random pattern of the Si–SiO_2 interface in sub-100-nm MOSFETs are many-fold and impractical to address using current simulation techniques. We attempt to minimize the 3-D transport problem in bulk Si by using a classical Brownian motion. In Brownian formalism [10], Langevin's equation of motion including a damping term associated with drift mobility ($\mu = q\tau/m^*$, where τ is momentum relaxation time), is integrated along with a random force $\Delta\mathbf{F}_R$, which represents the diffusion term. We use a trapezoidal integration scheme [10] to obtain the discretized form of the Langevin equation

$$\beta\mathbf{v}_{n+1} = \gamma\mathbf{v}_n + \frac{\mu}{\tau}\mathbf{E}dt + \Delta\mathbf{V}_R \quad (1)$$

where dt is discrete time interval, $\mathbf{v}$ is velocity, $\mathbf{E}$ is the electric field, $\beta = 1 + dt/\tau$, and $\gamma = 1 - dt/\tau$. The term $\Delta\mathbf{V}_R$ represents the integrated random force and is selected at each time step from a normal distribution with a variance $\langle\Delta\mathbf{V}_R^2\rangle = 2k_BTdt/\tau m^*$. Equation (1) may be considered equivalent to using DD models in a particle-simulation frame-

Manuscript received June 17, 2002; revised August 9, 2002. The work of S. Kaya was supported by the Ohio University Baker Fund.

S. Kaya is with the School of Electrical Engineering and Computer Science, Russ College of Engineering and Technology, Ohio University, Athens, OH 45701 USA (e-mail: kaya@ohio.edu).

A. Asenov and S. Roy are with the Device Modeling Group, Department of Electronics and Electrical Engineering, The University of Glasgow G12 8LT, U.K.

Digital Object Identifier 10.1109/TNANO.2002.807385

1536-125X/02$17.00 © 2002 IEEE

work[1] and equivalent to the solution of Boltzmann transport equation in relaxation time approximation. Using this simple approach, not only is the phonon and Coulomb-limited bulk mobility reproduced in a particle simulation framework, but also the 3-D simulations speed up considerably [11]. As alternatives, though exact, Monte Carlo techniques would be substantially slower for a mere reproduction of the bulk mobility. As a further simplification, we avoid the solution of self-consistent field distribution in an appropriate doping landscape and in the presence of carriers, which would have severe overheads in a 3-D simulation. Thus, we assume a uniform transverse and longitudinal electric field along the interface. The limitations of such a simple, equilibrium-only approach may be justified by the fact that we are interested only at *low-field* mobility as a function of "effective" transverse field at the Si–SiO_2 interface. The main limiting factor in the Brownian iterations is the choice of dt. To reduce errors below 0.1%, we find $dt < \tau/20$. Also, we simulate up to 20 000 particles to get a reasonable average on the collected histories. Provided significantly long time is allowed for transients, this 3-D framework can be reliably used in a device configuration, which should include an accurate description of additional interface scattering events.

B. Random Interface Generation

Realistic 3-D interfaces considered in this work, such as the one shown in Fig. 1, have been reconstructed starting from an autocorrelation function with a given correlation length (Λ) and rms height (Δ) [5]. In the Fourier domain, the magnitude of elements of a two-dimensional (2-D) complex array representing the height function is determined from the spectral information, while the phase is selected at random, imposing necessary symmetries. The 2-D inverse Fourier transform is used to obtain a height function in real space, which is then quantized across one atomic layer of Si, $d_{\rm Si}$, which is approximately 3 Å. A detailed description of the algorithm can be found elsewhere [5]. The use of random phases ensures that a statistical ensemble can be built. The two options for the spectral model provide additional flexibility and are in line with earlier experimental and modeling works [13], [14].

C. Particle-Interface Interaction Scheme

As a viable and efficient interaction scheme, we showed in an earlier work [6] that "thermalizing" particle energies crossing *an* interface and replacing them randomly back in the device within a distance $dz = \delta d_{\rm Si}$ provides a suitable physical framework. By choosing an appropriate value for δ, it is possible to reduce artificial carrier heating[2] to negligible values and obtain stationery distribution of particles along the transverse direction. The replacement of carriers crossing boundaries back into device is equivalent to an effective repulsive force at the interface that drives particles away. We find that $\delta \approx 1.0$ is optimum for a wide range of transverse fields and δ does not have significant influence on the drift mobility in the x direction. The history of particle ensemble velocities in all directions is plotted in Fig. 2 for a device with a flat interface and $\delta \approx 1.0$. The inset in this figure shows velocity in the z direction for a range of $E_{\rm eff}$ values. This figure essentially demonstrates the success of our approach for the particle–interface interaction and shows that bulk mobility remains intact for a fictitious device with an ideal interface.

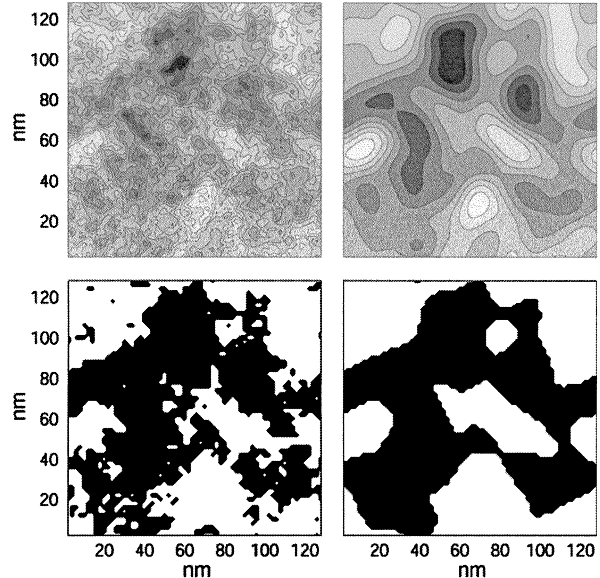

Fig. 1. Contour plots of realistic 3-D Si–SiO_2 interface model (top) reconstructed using Gaussian (right) and exponential (left) autocorrelations with $\Delta = 0.3$ nm and $\Lambda = 12$ nm. The equivalent surfaces at the lower half are digitized across a single Si(001) atomic layer.

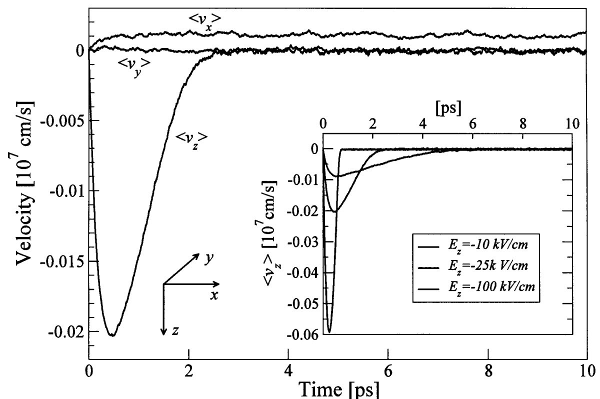

Fig. 2. Three-dimensional velocity history collected from 10 000 electrons undergoing Brownian motion inside a semi-infinite device with an ideal (flat) Si–SiO_2 interface and $\mu_{\rm Bulk} = 1100$ cm^2/Vs. Similar responses in the z direction for different effective field values are given in the inset.

III. Verification: Universal Mobility

Next, we investigate the universal mobility behavior of the random interfaces. For sufficient realism, we use a single-step interface having Gaussian or exponential autocorrelation functions and $dt = \tau/100$ and $\delta = 1.0$ throughout. The only ad-

[1]Note that (1) accounts for the carrier drift via the second term, while the diffusion process is modeled via the third term associated with a random walk in velocity space, which averages to zero in all time scales and has a variance proportional to the diffusion constant. As a further reading, Ferry explores the relationship between random Brownian motion and diffusion constant in the context of solution of BTE [12].

[2]Occurrence of artificial heating is not innate to (1), but would rather be a consequence of inaccurate handling of carrier dynamics when Brownian velocity iterations are halted at the device boundaries. The latter may or may not occur after particle–boundary interactions are included in 3-D simulations, depending on the nature of interaction scheme used. See [6] for more discussion.

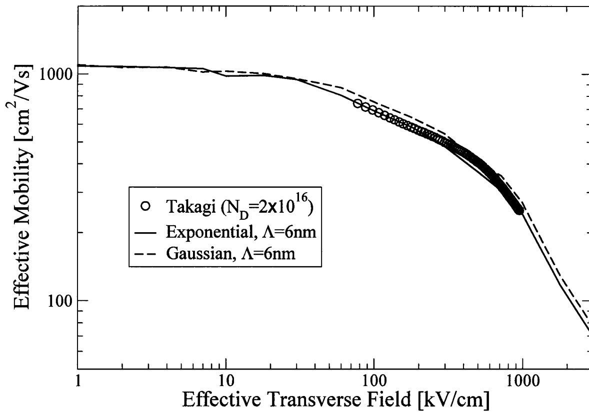

Fig. 3. Simulated effective mobility in a semi-infinitely long device, which has a random interface having a correlation length $\Lambda = 6$ nm. Gaussian and exponential correlations compared with experimental data [7] from a MOSFET with a bulk doping of 2×10^{16} cm^{-3}.

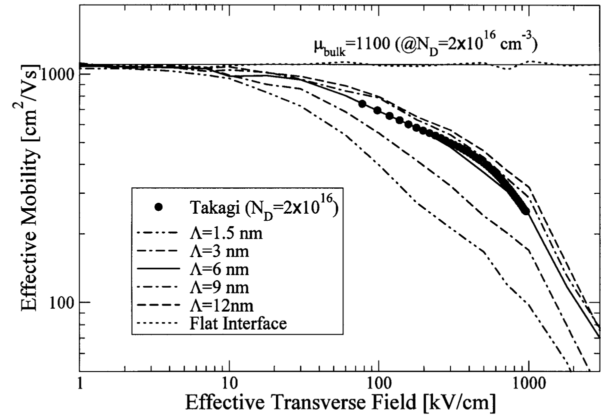

Fig. 4. Correlation length dependence (exponential model) of the simulated mobility in a semi-infinite device for a bulk mobility of 1100 cm^2/Vs. The mobility in the case of a flat interface ($\Lambda \rightarrow \infty$) and experimentally recorded mobilities are also shown.

justable parameter is the correlation length (Λ), which is not widely agreed upon, with conflicting values reported by various experimental methods and AFM measurements. However, calibrated MC simulations [15], [16] indicate that Λ falls in the range of 1–3 nm, and so does the HRTM data [13]. To mimic an infinitely long device, we use cyclic boundary conditions in the x–y direction of the motion. The simulator tracks particle trajectories in the vicinity of the interface steps, detects collisions, and implements the interaction scheme described above. Collision detection relies on the knowledge of the location of interface steps and takes into account multiple collisions at the corners. After the transients, ensemble average of the carrier velocities are collected while E_{eff} is altered in successive runs on the same device to build the universal mobility picture.

Fig. 3 shows the effective field dependence of the simulated mobility for a device with a bulk mobility 1100 cm^2/Vs, which corresponds to a doping level of 2×10^{16} cm^{-3} in Si. Both exponential and Gaussian correlation models are simulated for $\Lambda = 6$ nm. A surprisingly good agreement is obtained between our simulations and experimental data [7] for this value of correlation length, with the exponential model performing marginally better. Both the qualitative and quantitative features of the data are satisfactory for the chosen δ and Λ values. Fig. 4 shows the Λ dependence of the simulated mobility for the same conditions as in Fig. 3. Correlation lengths above 9 nm appear to have little influence in the surface limited mobility captured in this simple model.

The empirical interface scattering scheme we adapted above can be tuned successfully for a given set of parameters Λ and μ_{Bulk} in a MOSFET. Even though there is no universal δ value applicable to all cases, an appropriate set of dt and δ values may always be found for the chosen device parameters (μ_{Bulk} and Λ) [6]. Therefore, the simulation framework presented here can still be adequately used to analyze the impact of atomicity of the Si–SiO_2 interface on the interface roughness limited mobility in a decanano MOSFETs.

IV. Device Simulations

Two major modifications required for the adaptation of the above simulation method in device simulations are the imposition of proper boundary conditions at the source and drain end of a MOSFET and an appropriate collection of statistics in the "active" region of the channel. In device simulations, carriers are injected from the source with a shifted Maxwellian, while they are removed at the drain end and reallocated to the source at the next time step. To isolate the impact of source injection velocity from the effective mobility purely due to the SiO_2 interface, we first inject carriers from the source to a 200-nm "buffer zone" under the gate, which is immediately to the left of the actual "active" gate area. The interface above the buffer and active zones have similar statistical features and are reconstructed at the same time. Carrier statistics are collected only from the "active" area, which varies in size according to the gate length in question. Accordingly, "active" carrier number fluctuates during the simulation and becomes relatively low at small devices, for which simulation times must be sufficiently long to build proper averages. Finally, simulations are repeated for each MOSFET technology generation for up to 300 times so that average quantities may be built from large ensembles. For the device examples considered in the following, $\Lambda = 6$ nm and all other parameters are calibrated as before.

V. Results and Discussions

Fig. 5 shows the distribution of carrier effective mobility in 300 (50×100 nm) MOSFETs, each possessing a unique interface with a Gaussian autocorrelation. Note that mobility histogram fits reasonably well to a normal distribution, which is used to extract the average value and standard deviation of mobility in the ensemble. Such distributions are reminiscent of MOSFET threshold variations, which have been investigated in an earlier work [5] dealing with the electrostatic consequences of random interfaces in ultrasmall MOSFETs.

The dependence of average value and standard deviation of effective mobility on the gate length is studied in Figs. 6 and 7 for exponential and Gaussian autocorrelations. Both figures have common features: the average electron mobility and standard deviation rises as the channel length is reduced down to 10 nm. However, the exponential model reveals a sharper increase for the standard deviations below 30 nm as compared to the monotonous increase in the Gaussian case. In all cases,

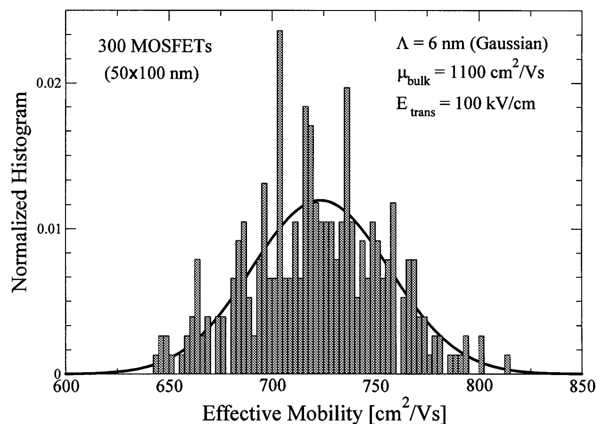

Fig. 5. Mobility histogram corresponding to 300 (50 × 100 nm) MOSFETs, each having a unique interface with a Gaussian autocorrelation ($\Lambda = 6$ nm). Effective mobility below the Si–SiO$_2$ interface distributes around a mean value expected from the universal mobility curves. Also shown in the background is a Gaussian fit intended as a guide to the eye.

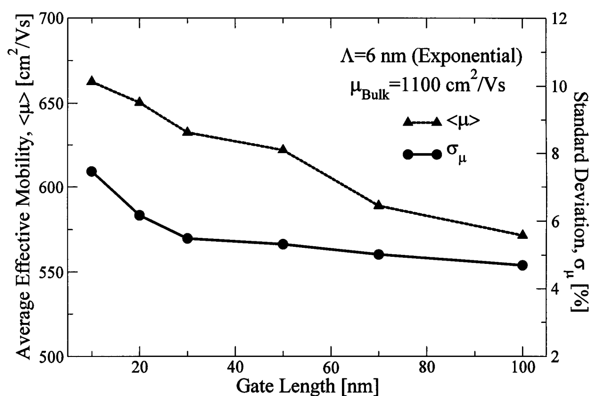

Fig. 6. Dependence of average mobility and standard deviation in sub-100-nm MOSFETs with an exponential interface autocorrelation ($\Lambda = 6$ nm) function. Standard deviation increases significantly for MOSFET gate length shorter than 30 nm.

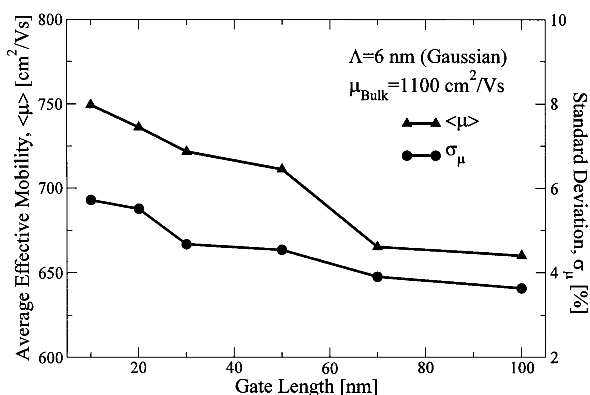

Fig. 7. Dependence of average mobility and standard deviation in sub-100-nm MOSFETs with a Gaussian interface autocorrelation ($\Lambda = 6$ nm) function. Standard deviation in this case increases in a monotonous fashion.

standard deviation appears to be limited to 10%. As the gate dimensions decrease in MOSFETs, carriers experience less number of scattering events with the interface during their journey from source to drain. This explains the increase in mobility. The standard deviation grows in size because in short-gate devices actual details of the interface are discernable and electron trajectories do not average out to a common behavior over many collisions. While this intuition is helpful, it is only through the statistical 3-D simulations we can describe the mobility distributions quantitatively.

It is also important to reconcile the difference between the meaning of Λ in this work and that of experimental and MC-based mobility calculations. Due to relative simplicity of the transport model and of the empirical carrier–interface interaction scheme used in this work, Λ should really be thought as a fitting parameter. Then all conclusions on the actual device scales may be normalized with respect to Λ, hence eliminating concerns about the interpretation of this parameter. Accordingly, we observe that universal mobility curves are more readily violated in MOSFETs when the gate length reaches below $\sim 6\Lambda$ scale (30–35 nm in this work). If Λ value is indeed limited to 1–2 nm in modern MOS structures, then our analysis would suggest that universal mobility curves will be reliable in MOSFETs as small as 10 nm in channel length. Thus, the approach presented here provides an answer to the original question regarding the extent of universal mobility curves in modern MOSFETs.

Despite their significance, results reported so far can be refined considerably if we introduce two significant changes to the simulator: 1) a self-consistent field solution in the channel which will take into account particle-mesh corrections, variation of transverse field across the channel and inclusion of random atomistic doping effects and 2) a quantum mechanical correction term which will alter particle distribution around the interface. Such refinements, which must be carefully optimized to retain the computational efficiency of the statistical simulator, are already underway and will allow a comparison between the classical and rigorous treatment of the problem.

VI. Conclusion

We have presented a 3-D statistical particle simulation framework for the investigation of universal mobility in the presence of realistic Si–SiO$_2$ interfaces. The framework is based on 3-D Brownian dynamics, is very fast, and can incorporate Gaussian or exponential correlations for the interface reconstruction. It employs a simple, efficient empirical algorithm for the particle–interface interactions, which results in proper distributions in velocity and real space and can be calibrated to reproduce MOSFET universal mobility behavior in the long-channel limit. We used this framework to investigate the statistical limit to the universal mobility curves in sub-100-nm MOSFETs. Statistical simulation of effective mobility below the Si–SiO$_2$ interfaces indicate that mobility fluctuations grow steadily in size and becomes more effective for devices with $L_{\rm gate} < 6\Lambda$, where Λ refers to the correlation length of the Si–SiO$_2$ interface.

Acknowledgment

The would like to thank to J. R. Watling, K. Kalna, C. Millar, and C. Arokianathan for useful discussions and suggestions during the course of this work.

REFERENCES

[1] R. Chau *et al.*, "30 nm physical gate length CMOS transistors with 1.0 ps n-MOS and 1.7 ps p-MOS gate delays," in *IEDM Tech. Dig.*, 2000, pp. 45–48.

[2] *The International Technology Roadmap for Semiconductors*, 2001.

[3] P. A. Stolk, F. P. Widdershoven, and D. B. Klaassen, "The effect of statistical dopant fluctuations on MOS device performance," in *IEDM Tech. Dig.*, 1996, pp. 627–630.

[4] S. Kaya *et al.*, "Implications of imperfect interfaces and edges in ultra-small MOSFET characteristics," Phys. Stat. Sol. B.

[5] A. Asenov, S. Kaya, and J. H. Davies, "Oxide thickness variation induced threshold voltage fluctuations in decanano MOSFET's: A 3D density gradient simulation study," *Superlat. Microstr.*, vol. 28, pp. 507–515, 2000.

[6] S. Kaya, A. Asenov, and S. Roy, "On the breakdown of universal mobility curves: A Brownian 3D simulation framework," J. Comput. Electron..

[7] S. Takagi *et al.*, "On the universality of inversion layer mobility in Si MOSFET's: Part I—Effects of substrate doping concentration," *IEEE Trans. Electron Devices*, vol. 41, pp. 2357–2362, 1994.

[8] D. Vasileska and D. K. Ferry, "Scaled silicon MOSFET's: Universal mobility behavior," *IEEE Trans. Electron Devices*, vol. 44, pp. 577–583, 1997.

[9] A. Pirovano *et al.*, "On the correlation between surface roughness and inversion layer mobility in Si-MOSFETs," *IEEE Electron Device Lett.*, vol. 21, pp. 34–36, 2000.

[10] C. R. Arokianathan, A. Asenov, and J. H. Davies, "An approach based on Brownian motion for the simulation of ultrasmall semiconductor devices," *J. Appl. Phys.*, vol. 80, pp. 226–232, 1996.

[11] ——, "Mesh-based particle simulation of sub-0.1 micron FET's," *Semicond. Sci. Technol.*, vol. 13, pp. A173–A176, 1998.

[12] D. K. Ferry, *Semiconductor Transport*. London, U.K.: Taylor & Francis, 2000.

[13] S. Goodnick *et al.*, "Surface roughness at the Si(100)-SiO_2 interface," *Phys. Rev. B*, vol. 32, pp. 8171–8186, 1985.

[14] T. Yoshinobu, A. Iwamoto, and H. Iwasaki, "Scaling analysis of SiO_2/Si interface roughness by atomic force microscopy," *Jpn. J. Appl. Phys.*, vol. 33, pp. 383–387, 1994.

[15] S. Yamakawa *et al.*, "Study of interface roughness dependence of electron mobility in Si inversion layers using Monte Carlo method," *J. Appl. Phys.*, vol. 79, pp. 911–916, 1996.

[16] G. H. Kruithof, T. M. Klapwijk, and S. Bakker, "Temperature and interface-roughness dependence of the electron mobility in high-mobility Si(100) inversion layers below 4.2 K," *Phys. Rev. B*, vol. 43, pp. 6642–6649, 1991.

Savas Kaya received the M.Phil. degree from the University of Cambridge, Cambridge, U.K., in 1994 and the Ph.D. degree from Imperial College of Science, Technology & Medicine, London, U.K., in 1998 for his work on strained Si quantum wells on vicinal substrates.

From 1998 to 2001, he was a Postdoctoral Researcher at the University of Glasgow, Glasgow, U.K., carrying out research in transport and scaling in Si/SiGe MOSFETs and fluctuation phenomena in decanano MOSFETs. He is currently with the Russ College of Engineering, Ohio University, Athens. His other interests include computer-aided design, transport theory, nanostructures, and process integration.

Asen Asenov (M'96) received the M.Sc. degree in solid state physics from Sofia University, Sofia, Bulgaria, in 1979 and the Ph.D. degree in physics from The Bulgarian Academy of Science, Sofia, Bulgaria, in 1989.

For ten years, he was Head of the Process and Device Modeling Group in IME, Sofia, developing one of the first integrated process and device CMOS simulators IMPEDANCE. From 1989 to 1991, he was a Visiting Professor at the Physics Department of the Technical University of Munich, Munich, Germany. He is currently Head of the Department of Electronics Engineering at the University of Glasgow, Glasgow, U.K. As the Leader of the Device Modeling Group and Academic Director of the Atomistic Device Simulation Center, he also coordinates the development of 2-D and 3-D device simulators and their application in the design of FETs, SiGe MOSFETs, and IGBTs. He has over 180 publications in process and device modeling and simulation, semiconductor device physics, "atomistic" effects in ultrasmall devices, and parallel computing.

Scott Roy received the Ph.D. degree from the University of Glasgow, Glasgow, U.K., in 1994 for investigations into the engineering and architectural aspects of extended single electronic systems.

He is a Lecturer and Researcher in the Department of Electronics and Electrical Engineering, University of Glasgow. His interests include high-performance computing (he represented Glasgow University in ESPRIT Project ZEUS; Zentren Europäischen Supercomputings) and its application to device modeling. He is presently developing codes to simulate and optimize n-channel SiGe FETs for VLSI and RF applications and designing, in collaboration with Motorola UK, multiprocessor systems on which they will run.