Efficient 3D Monte Carlo Simulation of Orientation and Stress Effects in FinFETs

F. M. Bufler*†, F. O. Heinz†, and L. Smith‡
*Institut für Integrierte Systeme, ETH Zürich, Gloriastrasse 35, CH-8092 Zürich, Switzerland
Email: bufler@iis.ee.ethz.ch
†Synopsys Schweiz GmbH, Thurgauerstrasse 40, CH-8050 Zürich, Switzerland
‡Synopsys Inc., 700 E. Middlefield Road, Mountain View, California 94043, USA
Email: {bufler, fheinz, lees}@synopsys.com

***Abstract*—The stress and orientation dependence of FinFET performance is studied by parallelized 3D Monte Carlo (MC) device simulation. The long-channel mobility for holes in devices with (110)/⟨110⟩ sidewall/channel orientation was found to double relative to the (100)/⟨100⟩ configuration; electron mobility decreased by 20%. This agrees with recent measurements. In 15nm–FinFETs quasi–ballistic velocity overshoot is strongly enhanced by mechanical stress, leading to more than 10% increase in the on–current (ION). The wallclock time for computing ION with about one percent statistical error is less than ten minutes with 16 threads making MC viable for standard TCAD applications.**

I. Introduction

Compared to bulk MOSFETs, exploration of device performance by TCAD becomes more difficult for FinFETs. Four new aspects need to be considered and reconciled with the treatment of performance boosters such as stress engineering: (i) the device behavior is inherently 3D, (ii) realistic FinFETs involve geometric features such as fin corner rounding which require an unstructured mesh, (iii) surface scattering needs to consider arbitrary crystallographic orientations, and (iv) quasi–ballistic overshoot effects are decisive as FinFETs are intended for still smaller gate lengths at and below 15 nm. At the same time, computational speed is crucial but CPU times of parallelized ensemble 3D MC simulations are still quite long for typical TCAD applications [1], [2]. It is the aim of this paper to present an MC approach which considers the four aspects above on a physical basis and exploits the single–particle approach [3]–[5] to achieve fast 3D MC device simulations.

II. Monte Carlo Approach

As input, Monte Carlo simulation needs models for band structure and scattering. In this work, analytical band structure descriptions are used. Their advantage over tabulated pseudopotential band structures is that the computation time and disk space for generating and storing numerical band structure tables can be saved. For stress engineering this is a crucial aspect because each different stress tensor corresponds to a different band structure. Our two–band model with anisotropic nonparabolicity for electrons and the six–band $\mathbf{k}\cdot\mathbf{p}$ band structure for holes were found to be in good agreement with corresponding pseudopotential results [6], [7].

Phonon scattering, ionized impurity scattering and surface roughness scattering are considered. The standard model together with the corresponding values for the coupling constants

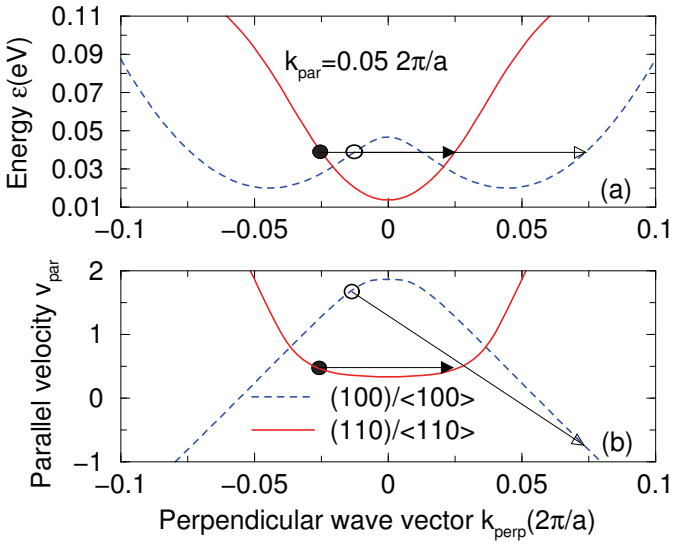

Fig. 1. Heavy–hole energy ϵ (a) and group velocity component parallel to the surface $v_{\rm par}$ (b) both as a function of the wave vector component perpendicular to the surface $k_{\rm perp}$ with the wave vector component parallel to the surface fixed at $k_{\rm par}$= 0.05 × 2π/a. The two configurations compared refer to (100) surface orientation with transport in ⟨100⟩ direction and to (110) surface orientation with transport in ⟨110⟩ direction, respectively.

for intervalley and elastic intravalley processes is applied for electron–phonon interaction [8]. Only the value of the acoustic intravalley deformation potential is changed to $\mathcal{E}$= 8.52 eV in order to reproduce the experimental velocity–field characteristics with our analytical band model, which differs from the usual nonparabolic ellipsoidal band model employed in Ref. [8]. Hole–phonon scattering, which comprises elastic acoustic and optical phonons, is described in Ref. [9]. Ionized impurity scattering is based on the Brooks–Herring model (see [8]), but the screening length is computed with degenerate statistics and the carrier temperature. Furthermore, its rate involves a doping–dependent prefactor which corrects the mobility overestimation at high doping levels.

Of particular importance is surface scattering because it is directly affected by the crystallographic surface orientation. It is modeled by a combination of 85 % specular and 15 % diffusive scattering. The specular part is governed by conservation of energy and parallel–momentum and its effect is illustrated in Fig. 1 for the six–band $\mathbf{k}\cdot\mathbf{p}$ heavy–hole band structure. Two

978-1-4673-5736-4/13/$31.00 ©2013 IEEE

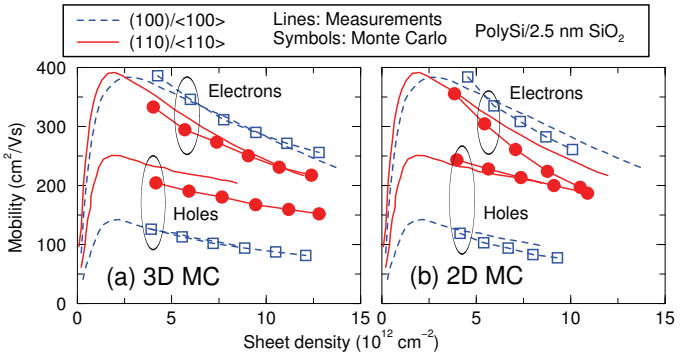

Fig. 2. Long–channel effective electron and hole mobilities in a FinFET with gate length L=0.5μm, fin height H=40nm, fin width W=30nm and a nitride hard mask on top of the fin. Measurements [13] are compared with (a) 3D Monte Carlo and (b) 2D Monte Carlo simulation referring to a double-gate device obtained from a horizontal cut through the FinFET.

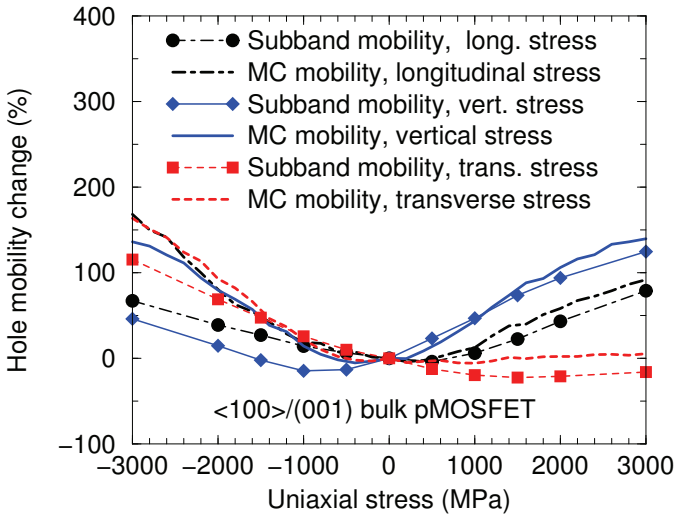

Fig. 3. Long-channel effective hole mobility change in a $\langle 100\rangle/(001)$ bulk pMOSFET at an effective field of $\mathrm{E_{eff}} = 1\mathrm{MV/cm}$ as a function of uniaxial stress in longitudinal, vertical and transverse direction. Subband mobility calculations based on 1D Schrödinger/Poisson solutions are compared with 2D Monte Carlo simulation.

situations are investigated: (100) surface with hole propagation in $\langle 100\rangle$ direction and (110) surface with hole propagation in $\langle 110\rangle$ direction. For both cases the hole energy and the parallel group velocity component are shown as a function of the perpendicular wave vector component; the parallel wave vector component is fixed since it is conserved upon specular scattering. The initial state for (100) and (110) surface orientation is indicated by the open and closed circles, respectively, and the final states by the end of the corresponding arrows. Whereas for the (110)/$\langle 110\rangle$ configuration energy conservation leads to an unchanged parallel group velocity, the parallel group velocity can change and even change its sign for the (100)/$\langle 100\rangle$ configuration thus degrading the surface mobility. This strong influence of the band structure on surface mobility was first observed for SiGe pMOSFETs [10]; in Fig. 1, the effect of the orientation of the band structure with respect to the surface is demonstrated for relaxed silicon.

The quantum–induced threshold voltage shift is considered in terms of modified effective oxide thickness and work function [11] based on a previous density-gradient simulation [12]. Thus our approach permits to consider simultaneously surface roughness via diffusive surface scattering, the orientation–dependence of surface mobility via specular surface scattering and the quantum–induced threshold voltage shift via effective oxide thickness and work function.

III. Orientation– and Stress–Dependent Mobility

Figs. 2 and 3 compare the effective mobilities of semi-classical MC simulation with measurements of FinFETs for (110)/$\langle 110\rangle$ and (100)/$\langle 100\rangle$ sidewall/channel orientations [13] and with subband mobility calculations for a (100)/$\langle 100\rangle$ oriented bulk pMOSFET under uniaxial stress, respectively. In Fig. 2, the mobility is as in measurements [14] computed via

$$\mu_{\mathrm{eff}} = \frac{L \cdot I_{\mathrm{D}}}{W_{\mathrm{eff}} \cdot q \cdot N_{\mathrm{inv}} \cdot V_{\mathrm{D}}} \qquad (1)$$

from the drain current I_{D} in a FinFET with a gate length of $L = 0.5\,\mu\mathrm{m}$ at a drain voltage of $V_{\mathrm{D}} = 50\,\mathrm{mV}$, q is the elementary charge, N_{inv} the inversion density per gate area, and the effective gate width is $W_{\mathrm{eff}} = 2 \cdot H$ because of the presence of a nitride hard mask on the top surface of the fin. In contrast, the MC mobility in Fig. 3 is calculated from internal quantities according to

$$E_{\mathrm{eff}} = \frac{q}{\epsilon_{\mathrm{Si}}} \int dz \left(N_{\mathrm{D}}^{+} - n(x,z) - N_{\mathrm{A}}^{-} + \eta\, p(x,z) \right) \qquad (2)$$

$$\mu_{\mathrm{eff}} = \frac{\int dz\, \frac{v_x(x,z)}{E_x(x,z)} p(x,z)}{\int dz\, p(x,z)} \qquad (3)$$

in analogy to subband mobility computations. N_{D}^{+}, n, N_{A}^{-} and p denote the donor, electron, acceptor and hole density, respectively, and η = 1/3 for the computation of the effective hole mobility. v_x and E_x are the x–components of the average hole velocity and of the electric field. The x–axis points in channel direction and integration over z is from the gate–oxide/silicon interface into the substrate. The channel doping is $N_{\mathrm{D}}^{+} = 2 \times 10^{17}\ \mathrm{cm}^{-3}$ and $N_{\mathrm{A}}^{-} = 0$.

As explained in the previous section, the orientation– and stress–dependence originates from the anisotropic band structures in conjunction with energy and parallel–momentum conservation of specular surface scattering, see also [15]. Figure 2 (a) shows that 3D MC is, with the *same* diffusive percentage for both orientation configurations, in good agreement with the measured orientation–dependence of mobility in FinFETs. But in the 2D approximation in Fig. 2 (b) — despite the hard mask — the difference between electron and hole mobility becomes smaller in the (110)/$\langle 110\rangle$ configuration due to the absence of the (001) top surface which is favorable for electrons and unfavorable for holes. It should be noted that previous measurements showed a much stronger degradation of the (110) electron mobility compared to the (100) surface orientation. A recent analysis of FinFET mobility measurements [16] found the (110) surface to be much rougher than the (100) surface in n–channel FinFETs, while roughnesses were equal in p–channel devices. This suggests that improved processing could enhance n–FinFET mobility; this might have been achieved in the works that report an electron mobility reduction of only 20% [13], [17]. The discrepancy with previous (110)

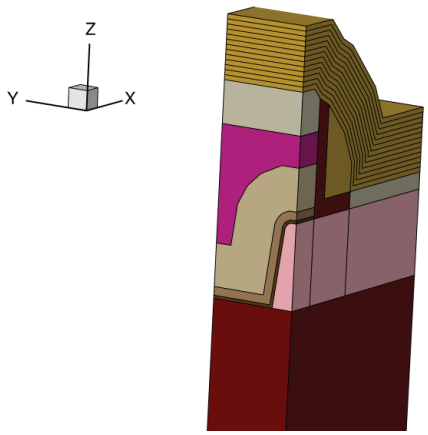

Fig. 4. Geometry of one quarter of the short–channel FinFET with gate length L=15nm, fin height H=20nm, fin width W=5.5nm at the top side (sidewall angle = 7°) and a 15nm thick cap liner. The complete structure used for Monte Carlo device simulation comprises 164540 tetrahedal elements and 39879 vertices leading to a memory consumption of 2.3 GByte.

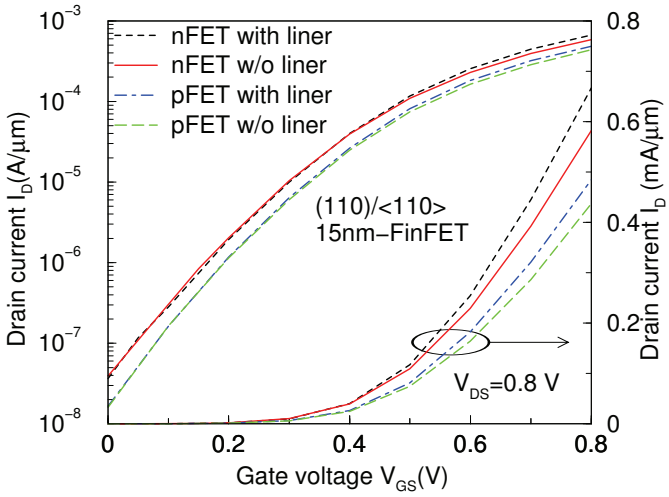

Fig. 5. Transfer characteristics in logarithmic and linear scales for the FinFET shown in Fig. 4 according to 3D MC simulation. The cap liner has 2 GPa intrinsic stress (tensile for nFinFET and compressive for pFinFET) and the workfunction in the strained FinFETs is adjusted to obtain the same gateover drive as in the unstrained cases.

electron mobility measurements, therefore, may be explained by improved smoothness of the sidewalls and the influence of the top surface on the measured mobility.

Figure 3 shows that like subband–based mobility models MC breaks the equivalence of uniaxial stress in (010) and (001) direction which would yield identical mobilities in bulk silicon [15]. This reduced symmetry is due to correctly incorporating the gate interface as a boundary condition, either via energy and parallel–momentum conservation in semiclassical MC or via quantization in subband mobility.

IV. Short–Channel Simulation

In order to study the influence of stress on quasi–ballistic transport, we simulate the 15nm–FinFET with a stressed cap liner in Fig. 4 in analogy to the experimental Ref. [18]. The 2 GPa cap liner leads to an averaged channel stress of 432 MPa in channel direction, −388 MPa in height direction and −43 MPa in width direction for the nFinFET (opposite signs in the pFinFET). More than 10 % increase in ION can be observed in the transfer characteristics in Fig. 5 which originates from higher velocities in the source-side of the channel as shown in Fig. 6; in particular, the velocity overshoot is very strong for electrons, but it is still enhanced by stress.

V. Parallelization

The bottleneck in 3D ensemble MC simulation is the frequent solution of the Poisson equation [1] involving typically 20000 updates [2]. This problem is absent in the single–particle approach [3], [4], where a stationary solution of Boltzmann and Poisson equations is obtained in an iterative scheme involving only about 20 steps in the on–state. A simple parallelization approach consists in independently propagating one particle per thread for faster sampling of the MC density to be used in the next Poisson update. Figure 7 shows the wallclock time and speed–up of our OpenMP–based implementation as function of the number of threads. A good speed–up is obtained for the off–current with less than 2 hours wallclock time on 16

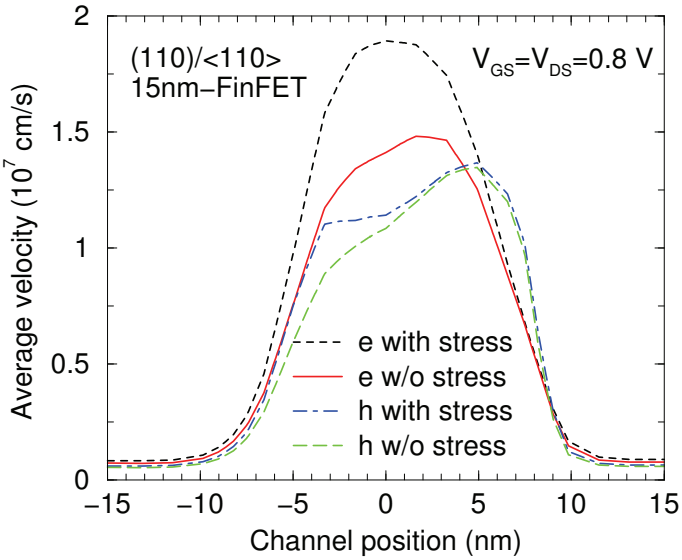

Fig. 6. Velocity profiles along the channel in the on–state of the transfer characteristics in Fig. 5. The velocities result from averaging with the carrier densities over the cross–section of the fin.

cores. Scalability is reduced for ION, but the wallclock time is already less than ten minutes on 16 cores compared to 20.7 hours for ensemble MC [2].

VI. Conclusion

We have presented a computationally efficient MC approach which captures the measured orientation–dependence of FinFET mobilities without orientation–dependent calibration. Quasi–ballistic overshoot and its enhancement by stress is demonstrated in short–channel devices. Thanks to the single–particle approach and parallelization, on–current computation in less than 10 minutes are possible on 16 core machines which makes this MC approach suitable for TCAD applications.

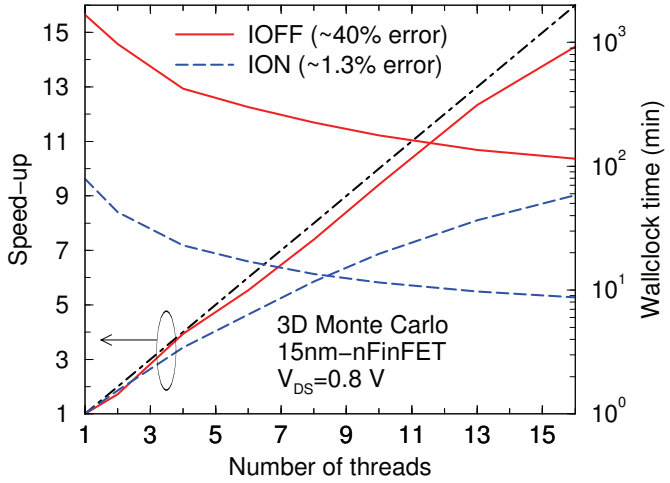

Fig. 7. Speed–up in linear scale and wallclock time in logarithmic scale for 3D Monte Carlo simulation of short–channel off–current IOFF and on–current ION on a 16 core 2.60 GHz Intel Xeon CPU E5-2670.

REFERENCES

[1] M. Aldegunde, A. J. García-Loureiro, A. Martinez, and K. Kalna, "Tetrahedral elements in self–consistent parallel 3D Monte Carlo simulations of MOSFETs," *J. Comput. Electron.*, vol. 7, pp. 201–204, 2008.

[2] W. Zhang, G. Du, Q. Li, A. Zhang, Z. Mo, X. Liu, and P. Zhang, "A 3D parallel Monte Carlo simulator for semiconductor devices," in *Proc. IWCE*, Bejing (China), May 2009, pp. 223–226.

[3] F. Venturi, R. K. Smith, E. C. Sangiorgi, M. R. Pinto, and B. Riccó, "A general purpose device simulator coupling Poisson and Monte Carlo transport with applications to deep submicron MOSFET's," *IEEE Trans. Computer–Aided Des.*, vol. 8, pp. 360–369, 1989.

[4] F. M. Bufler, C. Zechner, A. Schenk, and W. Fichtner, "Single–particle approach to self–consistent Monte Carlo device simulation," *IEICE Trans. Electron.*, vol. E86–C, pp. 308–313, 2003.

[5] Synopsys Inc., "Sentaurus Device Monte Carlo User Guide," Release H-2013.03, Mountain View (USA), 2013.

[6] F. M. Bufler, F. O. Heinz, A. Tsibizov, and M. Oulmane, "Simulation of ⟨110⟩ nMOSFETs with a tensile strained cap layer," *ECS Trans.*, vol. 16, no. 10, pp. 91–100, 2008.

[7] F. M. Bufler, A. Tsibizov, and A. Erlebach, "Scaling of bulk pMOSFETs: (110) surface orientation versus uniaxial compressive stress," *IEEE Electron Device Lett.*, vol. 27, pp. 992–994, 2006.

[8] C. Jacoboni and L. Reggiani, "The Monte Carlo method for the solution of charge transport in semiconductors with application to covalent materials," *Rev. Mod. Phys.*, vol. 55, pp. 645–705, 1983.

[9] F. M. Bufler and B. Meinerzhagen, "Hole transport in strained $Si_{1-x}Ge_x$ alloys on $Si_{1-y}Ge_y$ substrates," *J. Appl. Phys.*, vol. 84, pp. 5597–5602, 1998.

[10] C. Jungemann, S. Keith, and B. Meinerzhagen, "Full–band Monte Carlo simulation of a 0.12 μm–Si–PMOSFET with and without a strained SiGe–channel," in *IEDM Tech. Dig.*, 1998, pp. 897–900.

[11] R. Hudé, D. Villanueva, R. Clerc, G. Ghibaudo, and E. Robilliart, "A simple approach to account for the impact of quantum confinement on the charge in semiclassical Monte Carlo simulations of bulk nMOSFETs," in *Proc. ULIS*, Bologna (Italy), Apr. 2005, pp. 159–162.

[12] F. M. Bufler, R. Hudé, and A. Erlebach, "On a simple and accurate quantum correction for Monte Carlo simulation," *J. Comput. Electron.*, vol. 5, pp. 467–469, 2006.

[13] K. Akarvardar, C. D. Young, M. O. Baykan, I. Ok, T. Ngai, K.-W. Ang, M. P. Rodgers, S. Gausepohl, P. Majhi, C. Hobbs, P. D. Kirsch, and R. Jammy, "Impact of fin doping and gate stack on FinFET (110) and (100) electron and hole mobilities," *IEEE Electron Device Lett.*, vol. 33, pp. 351–353, 2012.

[14] T. Rudenko, V. Kilchytska, N. Collaert, M. Jurczak, A. Nazarov, and D. Flandre, "Carrier mobility in undoped triple–gate FinFET structures and limitations of its description in terms of top and sidewall mobilities," *IEEE Trans. Electron Devices*, vol. 55, pp. 3532–3541, 2008.

[15] F. M. Bufler, A. Erlebach, and M. Oulmane, "Symmetry reduction by surface scattering and mobility model for stressed ⟨100⟩/(001) MOSFETs," in *Proc. SISPAD*, Bologna (Italy), Sep. 2010, pp. 275–278.

[16] J. W. Lee, D. Jang, M. Mouis, G. T. Kim, T. Chiarella, T. Hoffmann, and G. Ghibaudo, "Mobility analysis of surface roughness scattering in FinFET devices," *Solid–State Electron.*, vol. 62, pp. 195–201, 2011.

[17] V. S. Basker, T. Standaert, H. Kawasaki, C.-C. Yeh, K. Maitra, T. Yamashita, J. Faltermeier, H. Adhikari, H.Jagannathan, J. Wang, H. Sunamura, S. Kanakasabapathy, S. Schmitz, J. Cummings, A. Inada, C.-H. Lin, P. Kulkarni, Y. Zhu, J. Kuss, T. Yamamoto, A. Kumar, J. Wahl, A. Yagishita, L. F. Edge, R. H. Kim, E. McIellan, S. J. Holmes, R. C. Johnson, T. Levin, J. Demarest, M. Hane, M. Takayanagi, M. Colburn, V. K. Paruchuri, R. J. Miller, H. Bu, B. Doris, D. McHerron, E. Leobandung, and J. O'Neill, "A 0.063 μm^2 FinFET SRAM cell demonstration with conventional lithography using a novel integration scheme with aggressively scaled fin and gate pitch," in *Symp. on VLSI Tech.*, Honolulu (Hawaii), Jun. 2010, pp. 19–20.

[18] C. Smith, S. Parthasarathy, B. E. Coss, J. Williams, H. Adhikari, G. Smith, B. Sassman, M. M. Hussain, P. Majhi, and R. Jammy, "Strain engineering in nanoscale CMOS FinFETs and methods to optimize $R_{S/D}$," in *VLSI-TSA*, Hsinchu (Taiwan), Apr. 2010, pp. 156–157.