RC23546 (W0502-152) February 28, 2005
Electrical Engineering

IBM Research Report

Monte-Carlo Simulations of Performance Scaling in Strained-Si nMOSFETs

Arvind Kumar, Massimo V. Fischetti*, Steven E. Laux
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

*Department of Electrical and Computer Engineering
University of Massachusetts
Amherst, MA 01003

Research Division
Almaden - Austin - Beijing - Haifa - India - T. J. Watson - Tokyo - Zurich

LIMITED DISTRIBUTION NOTICE: This report has been submitted for publication outside of IBM and will probably be copyrighted if accepted for publication. It has been issued as a Research Report for early dissemination of its contents. In view of the transfer of copyright to the outside publisher, its distribution outside of IBM prior to publication should be limited to peer communications and specific requests. After outside publication, requests should be filled only by reprints or legally obtained copies of the article (e.g, payment of royalties). Copies may be requested from IBM T. J. Watson Research Center , P. O. Box 218, Yorktown Heights, NY 10598 USA (email: reports@us.ibm.com). Some reports are available on the internet at http://domino.watson.ibm.com/library/CyberDig.nsf/home.

Monte-Carlo Simulations of Performance Scaling in Strained-Si nMOSFETs

Arvind Kumar[1], Massimo V. Fischetti[2], and Steven E. Laux[1]
[1]IBM Semiconductor Research and Development Center (SRDC), T.J. Watson Research Center, P.O. Box 218, Yorktown Heights, NY 10598; E-mail: arvkumar@us.ibm.com, phone: (914)945-3786, fax: (914)945-2911
[2] Dept. of Electrical & Computer Engineering, Univ. of Massachusetts, Amherst, MA 01003

Introduction: Biaxial strain of Si, realized through epitaxial growth of a pseudomorphic Si layer on relaxed SiGe, is widely considered a promising candidate for achieving performance enhancement. Although enhanced mobility has been clearly observed in long-channel field-effect transistors (FETs) [1], with some enhancement reduction measured in 40-nm gate length FETs [2], uncertainty remains about the sustenance of performance enhancements for future technology nodes. Cai et al [3] have experimentally shown only moderate channel length dependence in strained-Si nFETs on SiGe-on-insulator scaled to sub-50-nm channel lengths, provided self-heating is correctly accounted for. The aim of this work is to provide a theoretical basis for understanding the scaling of strained-Si nFETs to the 20-nm regime by studying the dependence of performance on both channel length and Ge mole fraction through full-band Monte-Carlo simulations.

Device Structure: We study the intrinsic performance enhancement scaling of strained Si by utilizing three bulk n-channel field effect transistors (nFETs) scaled by applying the rules of generalized scaling [4] in which a disproportionately weak decrease of the supply voltage is compensated through additional channel doping. As summarized in Table 1, device simulations using the drift-diffusion simulator FIELDAY [5] confirm the proper scaling of device parameters as channel length, junction depth, and gate insulator thickness are decreased; channel doping is increased; and supply voltage is reduced. These scaled devices were then simulated using the full-band Monte-Carlo simulator DAMOCLES [6], replacing the unstrained-Si channel with a 13-nm thick strained-Si layer atop a relaxed $Si_{1-x}Ge_x$ buffer having Ge mole fractions x of 0 (relaxed), 0.l, 0.17, or 0.25, as shown in Fig. 1.

Monte-Carlo Simulations: The DAMOCLES simulations utilize full band-structure from nonlocal empirical pseudopotentials with spin-orbit coupling; details of the band structure of strained Si and $Si_{1-x}Ge_x$ are given in Ref. [7]. Surface roughness scattering is taken to be 75% specular and 25% diffusive for both relaxed and strained Si, giving long-channel enhancements in reasonable agreement with Ref. [1].

Simulation Results: Figure 2 shows the saturation threshold voltage V_{tsat} as a function of gate length L for various x. Here V_{tsat} is computed by linear extrapolation in the high gate bias regime at drain voltage V_{dd} where the drain current I_d is linear in V_{gs}. Note that at L=66 nm, V_{tsat} decreases with increasing x, as expected from the bandgap reduction effect, but V_{tsat} is a weaker function of L for increasing x, leading to a different ordering of V_{tsat} in x for small L. We use as a performance metric the effective drive current $I_{eff}=(I_{low}+I_{hgh})/2$, where I_{low} is I_d for $V_{gs}^{(L)}=V_{dd}/2+\Delta V_t(x,L)$, $V_{ds}^{(H)}=V_{dd}$, and I_{hgh} is I_d for $V_{gs}^{(H)}=V_{dd}+\Delta V_t(x,L)$, $V_{ds}^{(H)}=V_{dd}/2$ [8]. Here $\Delta V_t(x,L)$ refers to the shift in V_{tsat} from the relaxed Si (x=0) case at the same gate length L, allowing us to compare performance at the same gate overdrive. Figure 3(a) shows I_{eff} as a function of x for different L. Note the acceleration of I_{eff} enhancement for x>0.10. Figure 3(b) shows the I_{eff} enhancement for each x relative to the relaxed (x=0) case. Consistent with experiment [3] and previous Monte-Carlo studies [9], only a weak degradation in performance is observed upon channel length scaling. In Fig. 3(c) we estimate the overall speed scaling based on the dependence of delay as $C_L V_{dd}/I_{eff}$ (load capacitance C_L is assumed to be independent of L). Note that a substantial enhancement is achieved for x=0.17. Although intrinsic enhancement continues to increase with x, other considerations -- such as the higher doping that would be required to raise the threshold voltage, higher junction capacitance and leakage, and defect control -- suggest that continuing to increase x is unlikely to be beneficial [10]. Figure 4 compares the enhancement at low V_{ds} (0.05 V) to that at high V_{ds} (0.9 V) for L=22 nm, for both low ($V_{gs}=V_{gs}^{(L)}$) and high ($V_{gs}=V_{gs}^{(H)}$) gate overdrive. Even for this short channel length, a significant fraction of the enhancement at low V_{ds} -- generally correlated with the long-channel mobility -- persists at high V_{ds}. This enhancement fraction is seen to increase with increasing x. Finally Figs. 5(a-b) show the longitudinal velocity (density-weighted average in the transverse direction) along the channel for the L=66 nm and L=22 nm cases, under the I_{hgh} condition described above. Even for $V_{ds}^{(H)}=V_{dd}/2$, significant velocity overshoot above the saturation velocity is seen. Peak carrier velocity roughly doubles under channel length scaling from L=66 nm to L=22 nm.

Conclusions: In agreement with recent experimental results, we find using full-band Monte Carlo simulations only modest dependence of strained-Si performance enhancement on channel length. Sustenance of performance enhancement under scaling is confirmed both through persistence of low drain bias gain at high drain bias and through velocity gain arising from overshoot. Intrinsic performance enhancement improves as Ge mole fraction is increased, and these enhancements are robustly sustained under gate length scaling. Ge mole fractions near 0.17 should provide a good design point, balancing performance enhancement with practical considerations.

Acknowledgments: The authors wish to thank W. Haensch, J. Cai, R. Dennard, and K. Rim for useful discussions.

References:

[1] K. Rim *et al.*, VLSI Tech. Dig., p. 98, 2002.
[2] K. Rim *et al.*, IEDM Tech. Dig., p. 43, 2002.
[3] J. Cai *et al.*, IEDM Tech. Dig., p. 165, 2004.
[4] Y. Taur and T. Ning, Fundamentals of Modern VLSI Devices, Cambridge University Press, New York, 1998.
[5] E. Buturla *et al.*, NASECODE VI Proc., p. 291, 1989.
[6] M.V. Fischetti and S.E. Laux, Phys. Rev. B, vol. 38, p. 9721, 1988.
[7] M.V. Fischetti and S.E. Laux, J. Appl. Phys., vol. 80, p. 2234, 1996.
[8] M.H. Na *et al.*, IEDM Tech. Dig., p. 121, 2002.
[9] F.M. Bufler and W. Fichtner, IEEE Trans. Elec. Dev., vol. 50, p. 278, 2003.
[10] J.G. Fossum and W. Zhang, IEEE Trans. Elec. Dev., vol. 50, p. 1042, 2003.

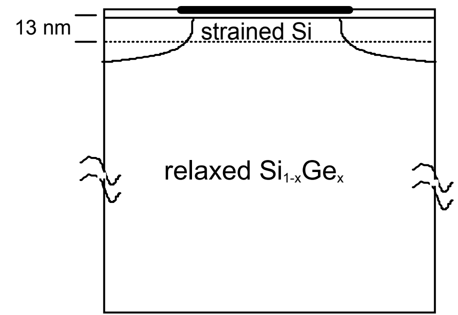

Fig. 1: Device structure used in DAMOCLES simulations consisting of a 13-nm strained Si layer atop a relaxed $Si_{1-x}Ge_x$ buffer.

L (nm)	t_{ox} (nm)	X_j (nm)	N_{ch} ($\times10^{18}cm^{-3}$)	V_{dd} (V)	V_{tsat} (V)	SS_{sat} (mV/dec)	DIBL (mV/V)
66	3	5.8	1.2	1.5	0.37	88	114
44	2	3.8	2.2	1.2	0.32	88	118
22	1	1.9	6.5	0.9	0.28	85	112

Table 1: Basic device input parameters under generalized scaling (gate length L, oxide thickness $t_{ox,}$ junction depth X_j, channel doping N_{ch}, supply voltage V_{dd}) and resulting device characteristics (threshold voltage V_{tsat}, subthreshold swing SS_{sat}, and DIBL).

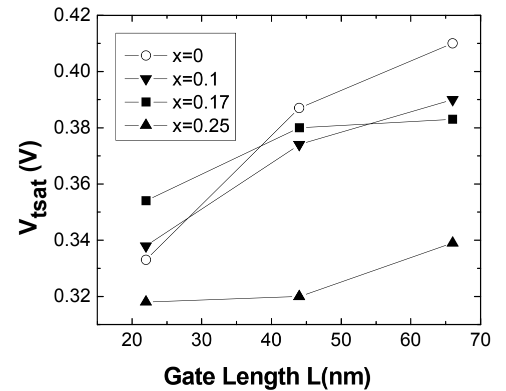

Fig. 2: Saturation threshold voltage extrapolated from high gate bias points for different Ge mole fractions, as a function of gate length.

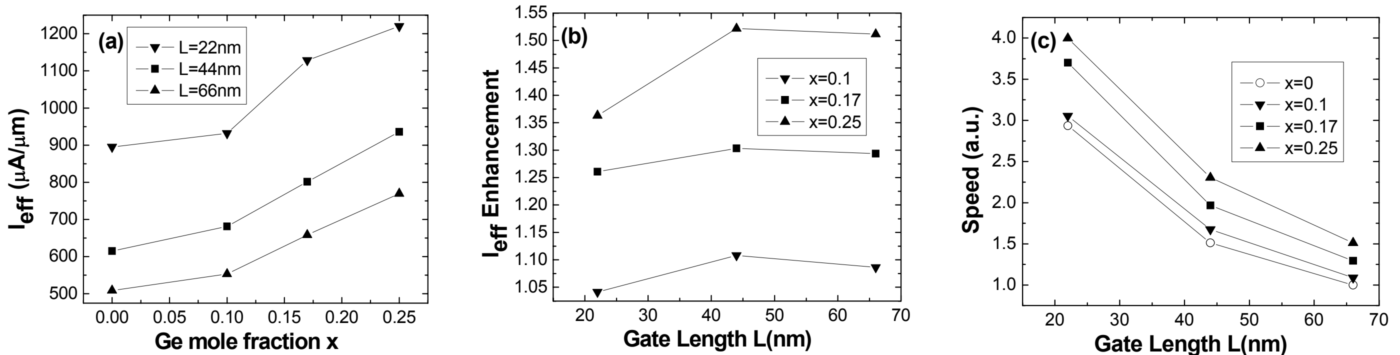

Fig. 3: (a) Effective drive current as a function of Ge mole fraction for different gate lengths. (b) Effective drive current enhancement (normalized to relaxed Si) as a function of gate length for various Ge mole fractions. (c) Speed scaling with channel length for various Ge mole fractions.

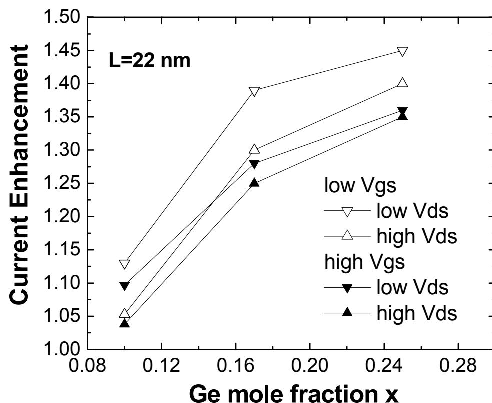

Fig. 4: Current enhancement, relative to relaxed Si, as a function of Ge mole fraction for L=22 nm. For both low and high gate overdrive, a significant fraction of the enhancement seen at low drain bias persists at high drain bias.

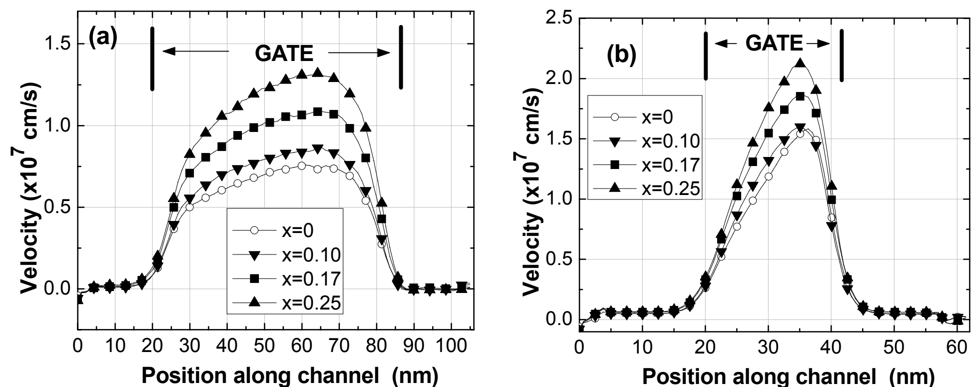

Fig. 5: Density-weighted average of component of electron velocity along channel as a function of position along channel for (a) L=66 nm and (b) L=22 nm, under biasing conditions corrresponding to I_{hgh}.