Modeling Gate-Pitch Scaling Impact on Stress-Induced Mobility and External Resistance for 20nm-node MOSFETs

Seong-Dong Kim°, Sameer Jain°, Hwasung Rhee*, Andreas Scholze+, Mickey Yu+, Seung Chul Lee*, Stephen Furkay+
°*+IBM Semiconductor Research and Development Center,
*Samsung Electronics, °*Hopewell Junction, NY 12533,
+Essex Junction VT 05452, USA.
E-mail: sdkim@us.ibm.com

Marco Zorzi†, Fabian M. Bufler†‡, Axel Erlebach†
†Synopsys Switzerland LLC, CH-8050 Zurich, Switzerland
‡IIS, ETH Zürich, CH-8092 Zurich, Switzerland

***Abstract*—The impact of gate-pitch scaling on device internal and external resistance is examined by advanced process and device modeling including distributed contact resistance model, mechanical stress and Monte Carlo (MC)-based stress-dependent mobility model. The contact resistance components and their major parameters in sub-50nm contact regime are analyzed by TCAD and transmission line modeling (TLM). The calibration method for the stress-induced channel mobility and the external resistance is proposed using R_{on}-L_{gate} measurements of 32nm-node devices with different gate-pitches. The significant performance degradation due to simple gate-pitch scaling is predicted for 20nm-node technology with sub-100nm gate-pitch.**

I. INTRDUCTION

Stress engineering using etch stop liner (ESL) or dual stress liner (DSL) has become one of the main performance element to enhance channel mobility in recent nano-meter scale CMOS technologies [1]. However, as the gate pitch aggressively scales down into sub-100nm regime with continuous technology scaling beyond 32nm-node [2], the two primary concerns and critical challenges for the device performance are being addressed as illustrated in Fig. 1: First, the stress effect associated with liner can be weakened by the narrower distance between neighboring gates which can be the effective space for transferring liner stress into channel. Second, this narrowed gate-space would lead to significant increase in the external resistance, R_{ext}, since the silicide contact length, L_{con} becomes shorter than the contact transfer length and hence will directly increase the contact resistance, R_{co} [3].

In this paper, the impact of gate-pitch scaling on device performance element is examined by advanced process and device modeling including the distributed contact resistance model, mechanical stress and Monte Carlo (MC)-based stress-dependent mobility model. The gate length dependence and the pitch dependence and tensile liner stress effect on the internal and external resistances are correlated to measured R_{on}-L_{gate} response of 32nm-node NFETs and applied for 20nm-node technology performance prediction.

This work is performed at IBM Microelectronics, Semiconductor Research and Development Center, Hopewell Junction, NY 12533 USA.

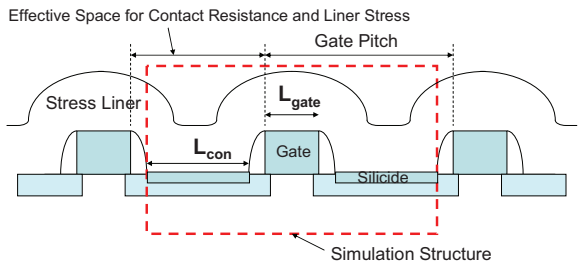


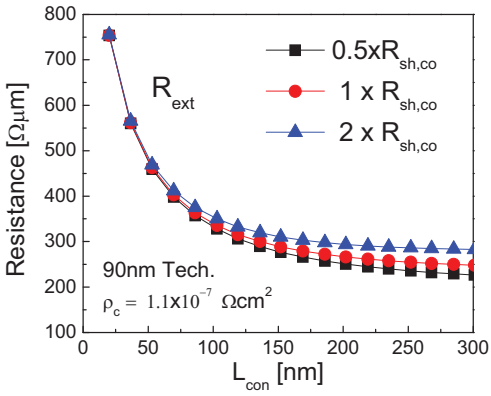
Figure 1. Schematic illustration of device structure with gate pitch. The effective space transferring liner stress into channel and the silicide contact resistance are scaled with the aggressive shrink of gate-pitch.

II. TCAD MODELING AND ANALYSIS

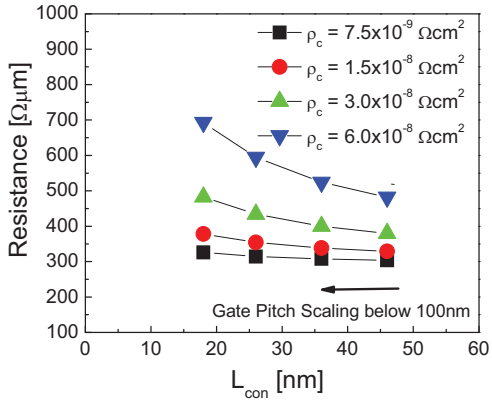
TCAD process and device modeling using Sentaurus process and device simulator are developed to reflect IBM 32nm high-k/metal gate bulk CMOS technology. Starting from the advanced calibration models, the basic 1-dimentional (1D) and 2-dimentional (2D) parameters are calibrated using experimental data such as device geometries, doping profiles and long and short channel electrical data.

A. Impact on Contact Resistance

In order to understand the gate-pitch dependence on contact resistance, the detailed silicide contact parameter study is performed by 1D TLM and 2D device simulation. Fig. 2(a) and (b) show the modeling results of the contact resistance behavior as a function of contact length with different sheet resistances of deep source/drain (S/D) junction in 90nm technology [4] and with different specific silicide contact resistivities in scaled contact length of 20nm technology, respectively. It can be seen from Fig 2(a) and (b) that sheet resistance of deep S/D has been an important parameter for determining the contact resistance level in previous technologies whose contact length is longer than contact transfer length. However, as the L_{con} diminishes below the transfer length, there is a sharp increase in R_{co}, and hence the silicide contact length and location are becoming dominant in recent technologies. In Fig. 2(a) and



(a)

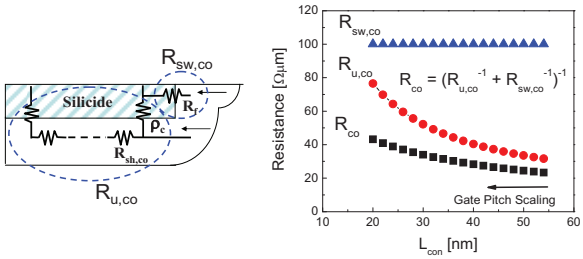


(b)

Figure 2. Modeled contact resistance behavior as a function of contact length for (a) different sheet resistance of deep S/D junction in 90nm technology [4] and for (b) specific contact resistivity of silicide contact in 20nm technology with short contact.

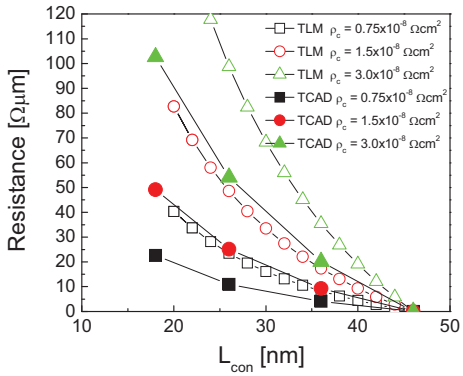
(b), it is found that major R_{co} contributors in sub-50nm contact length regime are the contact length and specific contact resistivity, ρ_c, while the sensitivity to sheet resistance of deep S/D junction is almost negligible. Therefore, continuous innovation focusing on silicide material property and activate doping level in silicide/silicon interface is required for the contact resistance improvement for future technologies.

The schematic descriptions of contact resistance components having sidewall contact resistance, $R_{sw,co}$ and planar contact, $R_{u,co}$ and their relative contributions in total contact resistance with contact length calculated by TLM are shown in Fig. 3(a) and (b), respectively. The uniform doping and uniform specific contact resistivity are assumed in TLM. R_{co} is determined mainly by $R_{u,co}$ in old technologies with relatively long contact length, but as gate-pitch scales further, the contribution of $R_{sw,co}$ becomes comparable to $R_{u,co}$ or can be even larger than $R_{u,co}$ for non-uniform S/D junction cases, so that the accurate modeling on $R_{sw,co}$ component would be desirable for accurate device design and performance prediction. The significance of $R_{sw,co}$ component is confirmed by comparing resistance variation with contact length scaling between TLM and TCAD simulation in Fig. 4. It is observed TLM without $R_{sw,co}$ component underestimates the contact resistance,

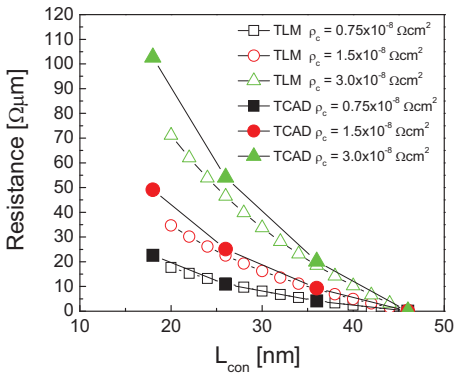


(a) (b)

Figure 3. (a) Contact resistance components having front sidewall contact resistance ($R_{sw,co}$) and planar contact ($R_{u,co}$) and (b) the contribution of each contact resistance component calculated by TLM with gate-pitch or contact length, indicating significant contribution of $R_{sw,co}$.



(a)



(b)

Figure 4. Comparison of contact resistance increase with contact length scaling between TLM and TCAD. (a) TLM neglecting sidewall contact resistance. (b) TLM including sidewall contact resistance. Accurate modeling on $R_{sw,co}$ component is emphasized by TCAD and TLM.

while TLM considering $R_{sw,co}$ component is in reasonable agreement with TCAD estimation with uniform contact resistivity. The discrepancy is getting larger for shorter L_{con} due to additional non-negligible device parameters like silicide geometry and the active doping levels.

B. External Resistance Calibraton

The external resistance can be separately calibrated from the channel mobility using 32nm R_{on}-L_{gate} measurements for the devices with 1x and 2x contacted pitch cases. Total on-

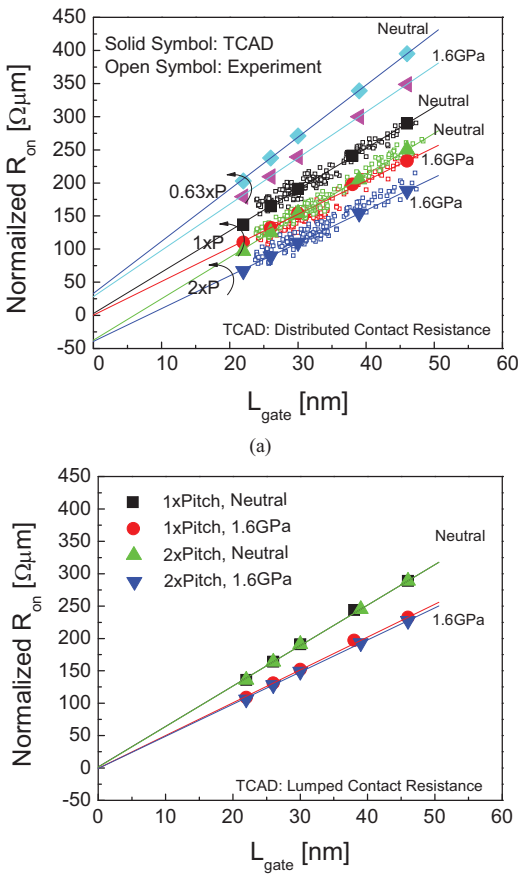


Figure 5. Stress-induced mobility and R_{ext} calibration to measured R_{on} vs. L_{gate} for 1x and 2x pitch of 32nm NFETs and prediction for 0.63x pitch of 20-nm technology. (a) TCAD employed distributed contact resistance model (b) TCAD employed lumped contact resistance model.

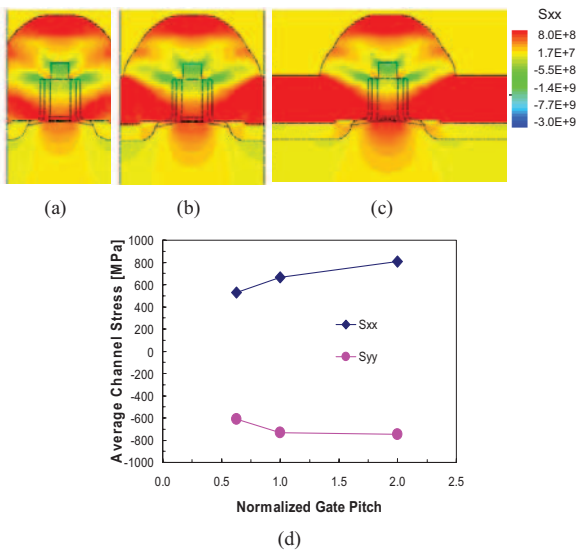


Figure 6. Calibrated lateral stress (Sxx) distribution for different gate pitches. (a) 0.63xPitch, (b) 1xPitch, (3) 2xPitch. (d) Average stress components along the channel for different gate-pitches. Reduction in both tensile Sxx and compressive Syy channel stress levels with pitch scaling leads to NFET electron mobility degradation.

resistance, R_{on} is defined at linear drain current measured at a fixed overdrive gate voltage (V_{gs}–V_{tlin}) [5][6]. By neglecting channel mobility dependence on gate length, the channel mobility can be assumed to be inversely proportional to slope of R_{on}-L_{gate} and the external resistance can be determined by the intercept of R_{on}-axis where the effective gate length diminishes.

First, the pitch impact on R_{ext} is calibrated by modifying specific contact resistivity employing distributed contact resistance model after 1D and 2D doping profile calibration as shown in Fig. 5. The external resistance calibration employing simple lumped resistive contact model may be acceptable for long contact or same gate- pitch devices, but cannot predict contact sizing effect driven by gate-pitch scaling, shown in Fig. 5(b). In this work, uniform distributed contact resistance model are fitted to match device on-resistance at the same overdrive gate voltage for characterizing contact sizing effect of 1x and 2x pitches. The calibrated specific contact resistivity of 3.6×10^{-8} Ωcm^2 is resulted from the distributed contact resistance model which is found to be close to the recent measurement range from IBM alliance research group [7]. The corresponding attached lumped resistance is extracted at 70 Ωμm from lumped resistive contact modeling for 1xpitch device case, which shows the limitation of the model for pitch dependence prediction in Fig 5(b). The more detailed contact model considering active doping-dependence and silicide Schottky barrier height would be required for accurately characterizing sophisticated silicide contact resistance behavior for next generation technologies in the short contact regime.

C. *Stress-Induced Channel Mobility*

Mechanical stress simulation is performed by Sprocess. The intrinsic stress is applied to dual stress liner depending on device type. The tensile intrinsic stress nitride film layers are deposited on the NFET devices after silicide formation. No stress relaxation processes are taken into account for DSL stress modeling except for the stress relaxation due to free surfaces that are generated in the structure during etching step and other structure formation. 30-multilayer tensile liner deposition technique [8] is employed in DSL process simulation for more accurate pitch dependence modeling. An MC-based model for stress-induced effective electron mobility enhancement is used similar to the model for holes [9]. The effective electron mobilities are obtained by self-consistent MC simulations of long-channel NFETs and the stress-dependence of the band structure is taken into account with an improved analytical two-band model [10]. The stress-induced channel mobility behavior is accurately emulated by fitting the slope change of R_{on}-L_{gate} responses between neutral stress liner and 1.6GPa tensile liner (dR_{on}/dL_{gate}) for two different gate pitch devices via fine-tuning of the employed mobility model.

TCAD calibrations for gate-pitch dependent R_{ext} and the relative channel mobility change with stress liner are in good agreement with the measurement data from 32nm NFET devices with different pitches as shown in Fig. 5(a). Then, simple 37% pitch scaling is applied to the device keeping the same gate length for predicting gate-pitch sizing effect on 20nm-node devices. From R_{on}-L_{gate} characteristics

of the measurement and TCAD prediction in Fig. 5(a), it can be seen the channel mobility enhancement due to tensile stress liner is significant for all gate-pitches but the slope change is much higher in 0.63xpitch devices in addition to R_{ext} increase, which is indicating that the simple scaling of gate-pitch for 20nm technology may result in significant performance degradation. Fig. 6(a), (b) and (C) show the calibrated 2D Sxx distribution for different gate-pitches and Fig 6 (d) shows Sxx and Syy stress components averaged out along the channel at 5nm depth under gate oxide for different gate-pitches where Sxx is lateral stress in the direction of the channel length and Syy is vertical stress perpendicular to channel plane. As shown in Fig. 6, the stress liner-induced channel electron mobility degradation with pitch scaling is well explained by the reduction in both lateral tensile stress and vertical compressive stress component, leading to decrease in the inverse slope of R_{on}-L_{gate} responses.

The gate-pitch sensitivity to stress-induced mobility and R_{ext} based on calibrated TCAD is shown in Fig. 7(a) and I_{dsat} performance with gate-pitch scaling is also predicted in Fig. 7(b). It should be noted that mobility degradation is more and more significant in 0.63xpitch of 20nm technology compared to 1x and 2x pitches of 32nm technology due to the pinching-off of stress-liner in the narrower gate space as is confirmed in 2D and 1D channel stress profiles in Fig 6. TCAD prediction indicates that 17% and 27% degradation in the external resistance and channel resistance with gate-pitch scaling, respectively, which may lead to about 10% degradation in I_{dsat} performance for 20nm-node technology.

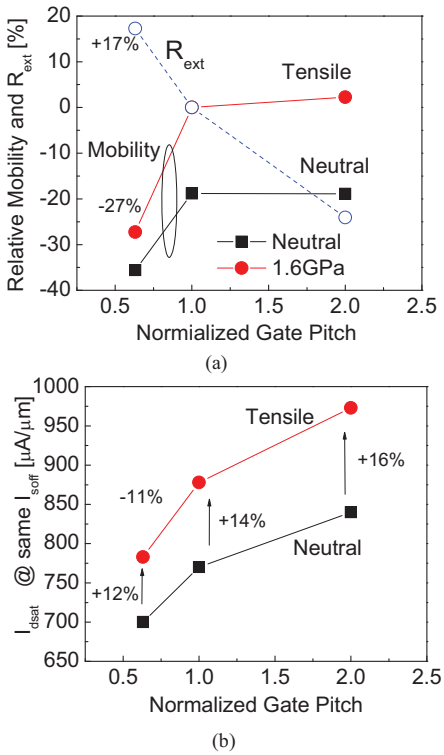


Figure 7. (a) Relative variation of stress-induced mobility and R_{ext} with gate-pitch based on 32nm tech showing 17% and 27% R_{ext} and R_{ch} degradation with smaller gate-pitch size below 100nm, respectively. (b) Relative I_{dsat} performance degradation with gate-pitch scaling.

III. Conclusion

Advanced TCAD modeling and analysis for gate-pitch scaling impact on external resistance and stress liner-induced mobility are presented. The importance of sidewall front silicide contact is emphasized by component analysis of contact resistance with 1D TLM and TCAD modeling. The gate length and pitch dependence and tensile liner stress effect on the channel mobility and the external resistances are well calibrated by using measured R_{on}-L_{gate} responses of 32nm-node NFET devices. The calibrated TCAD modeling indicates the simple gate-pitch sizing may have a severe impact on both internal and external resistance resulting in performance penalty of 10% for next generation technologies with sub-100nm gate-pitch, which highly demands novel channel mobility booster and external resistance engineering.

References

[1] K. Uejima, H nakamura, T. Fukase, S. Mochizuki, S. sugiyama, and M. Hane, "Highly efficient stress transfer techniques in dual stress liner CMOS integration," *VLSI technical Digest*, 2007. pp.220-221.

[2] S. Natarajan, M. Armstrong, M. Bost, M. Brazier, C.-H. Chang, V. Chikarmane, M. Childs, H. Deshpande, K. Dev, G. Ding, T. Ghani, O. Golonzka, W. Han, J. He, R. Heussner, R. James, I. Jin, C. Kenyon, S. Klopcic, S.-H. Lee, M. Liu, S. Lodha, B. McFadden, A. Murthy, L. Neiberg, J. Neirynck, P. Packan, S. Pae, C. Parker, C. Pelto, L. Pipes, J. Sebastian, J. Seiple, B. Sell, S. Sivakumar, B. Song, K. Tone, T. Troeger, C. Weber, M. Yang, A. Yeoh, K. Zhang, "A 32nm logic technology featuring 2nd-generation high-k+metak-gate transistors, enhanced channel strain and 0.171μm² SRAM cell size in a 291Mb array," *Tech. Dig. of International Electron Device Meeting (IEDM2008)*, 2008, pp.941-943.

[3] S.-D. Kim, "Optimum location of silicide/Si interface in ultra-thin body SOI MOSFETs with recessed and elevated silicide source/drain contact structure," *Solid-State Electron,* Volume 53, Issue 10, 2009, Pages 1112-1115.

[4] S.-D. Kim, S. Narasimha, and K. Rim, "An integrated methodology for accurate extraction of S/D series resistance components in nanoscale MOSFETs," *Tech. Dig. of International Electron Device Meeting (IEDM2005)*, 2005, pp.149 – 152.

[5] K. Terada and H. Muta, 'A new method to determine effective MOSFET channel length," *Jap.J. Appl. Phys*., vol 18, 1979, pp.953-959.

[6] J.G.J. Chern, P. Change, R.F. Motta, and N. Godinho, "A new method to determine MOSFET channel length," *IEEE EDL*, vol. EDL-1, 1980, pp. 170-173.

[7] K. Ohuchi, C. Lavoie, C. Murray, C. D'Emic, I. Lauer, J. O. Chu, B. Yang, P. Besser, L. Gignac, J. Bruley, G.U. Singco, F. Pagette, A.W. Topol, M.J. Rooks, J.J. Bucchignano, V. Narayanan, M. Khare, M. Takayanagi, K. Ishimaru, D.-G. Park, G. Shahidi, P. Solomon, "Extendibility of NiPt silicide conatact for CMOS technology demonstrated to the 22-nm Node," *Tech. Dig. of International Electron Device Meeting (IEDM2007)*, 2007, pp.1029 – 1031.

[8] K.V. Loiko, V. Adams, D. Tekleab, B. Winsted, X.-Z. Bo, P. Grunowski, S. Goktepeli, S. Filipiak, B. Goolsby, V. Kolagunta, and M.C. Foisy, "Multi-layer model for stressor film deposition," *SISPAD Proc*., 2006, pp123126.

[9] F.M. Bufler, A. Erlebach, and M. Oulmane, "Hole mobility model with silicon inversion layer symmetry and stress-dependent piezoconductance coefficients," *IEEE EDL*, vol. 30, No.9, 2009, pp.996-998.

[10] 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, 2008, pp.91-100.