Exploring MOL Design Options for a 20nm CMOS Technology using TCAD

Andreas Scholze, Stephen Furkay,
IBM Semiconductor Research and Development Center,
Essex Junction, VT 05452, USA.

Seong-Dong Kim, Sameer Jain,
IBM Semiconductor Research and Development Center,
Hopewell Junction, NY 12533, USA.

E-mail: ascholz@us.ibm.com

***Abstract*—A mixed-mode simulation framework is presented to study the AC performance of a 20nm bulk CMOS technology with respect to various options for contact design at the middle-of-line design level. These simulations combine the predictive capabilities of a calibrated two-dimensional TCAD model for a MOSFET with three-dimensional simulations for the layout dependent parasitic capacitances to extract the characteristic parameters of a multi-stage ring-oscillator circuit, such as the ring delay, and the effective switching capacitance. Significant performance degradation is predicted comparing the simulation results for a conventional contact design versus a typical 20nm design considering raised source-drain and a contact bar.**

I. Introduction

Aggressive pitch scaling, characteristic of state-of-the-art CMOS device designs, drives the need for quantitative evaluation of various middle-of-the-line (MOL) design options early in the technology development cycle. This becomes increasingly important for advanced technology nodes where the degradation in drive current due to an increase in the external resistance components is countered by choosing alternative contact design options which still allow reasonable short channel control [1]. Raised source drains (RSD), completely- or partially-strapped contact cones and contact bars are among said options. Their growing contributions to parasitic capacitance, resistance and circuit switching delays are typically quantified via ring-oscillator (RO) test structures (Fig. 1) [2].

We present herein an analogous TCAD-based approach to evaluation of these circuit implications, in the context of an industry standard 20nm bulk technology. These simulations couple 2D MOSFET structures with MOL-specific 3D capacitance extractions, in a mixed-mode (device/circuit) simulation scheme that allows modeling the complete RO layout and associated high-frequency response. We extract ring-delay, effective capacitance, C_{eff}, and effective resistance R_{eff} values for various MOL configurations, as the basis for identification of an optimal design strategy.

II. TCAD Modeling Approach

Process and device simulations based on the Sentaurus tool suite [3] are performed using a TCAD simulation environment implemented for a 20nm bulk CMOS technology that includes

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

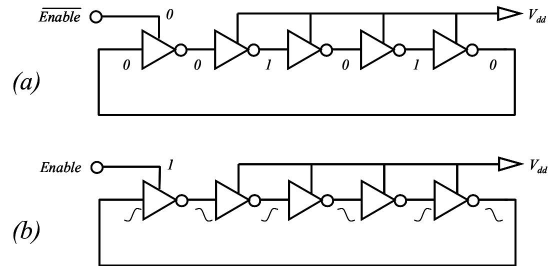

Figure 1. Simplified circuit diagram of a ring-oscillator with 5 stages: (a) quiescent and (b) active mode.

mechanical stress effects in a high-k/metal replacement-gate (RMG) flow. The process and device simulation models for the physical MOSFET were calibrated using experimental data such as measured SIMS profiles, TEM cross-sections and long and short channel electrical characteristics for various 32/28/20nm bulk n- and p-type MOSFETs. Circuit simulations for the ring-oscillator circuit are performed using the TCAD mixed-mode capability in Sentaurus Device [3]. Parasitic capacitances are extracted from small-signal simulations for a 3D MOSFET structure including the MOL contact elements.

A. RO Mixed-Mode Simulations

Typical RO test structures are comprised of many (>100) inverter stages each of which consists of active FET devices and parasitic loading capacitances. We consider a fan-out-of-three (FO3) design (Fig. 2) with 4 active FET devices and another 8 MOSFETs configured as MOS-capacitors. Five inverter stages were found to be sufficient for our mixed-mode simulations, each of which contained 2D numerical n- and p-MOSFETs and 5 layout-specific parasitic capacitors ($C_{1\text{-}5}$) coupling IN, OUT, Vdd, and GND nodes for our FO3 configuration.

Mixed-mode simulations are run for our 5-stage RO circuit in both active and quiescent modes. The ring delay, d, was extracted from the periodicity of the ring oscillations as

$$d = \frac{1}{2n \cdot f}, \tag{1}$$

where n is the number of inverter stages and f the oscillation frequency (Fig. 3). In addition to the switching frequency, the

978-1-61284-418-3/11/$26.00 © 2011 IEEE

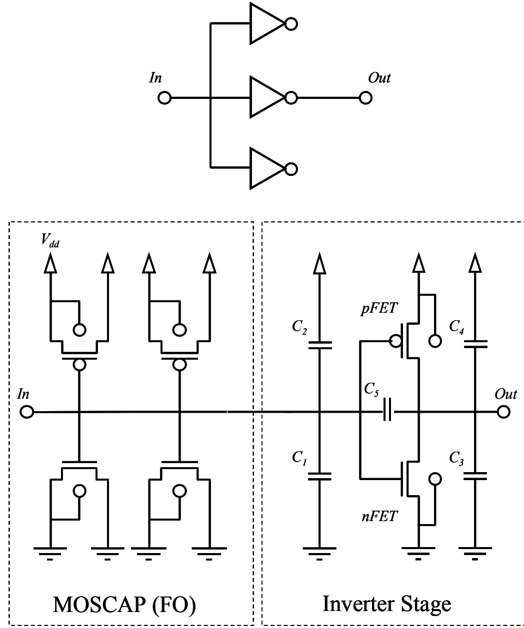

Figure 2. Circuit diagram for a ring-oscillator inverter stage with a FO3 configuration.

current drawn by the RO power supply during its active and quiescent states - *IDDA* and *IDDQ* respectively, is used to calculate the effective switching capacitance, C_{eff}, as

$$C_{eff} = 2d \frac{IDDA - IDDQ}{V_{dd}} . \tag{2}$$

The delay per stage is also expressed in terms of C_{eff}, and the switching resistance, R_{eff}, according to $d=C_{eff}R_{eff}$. The simulation time for the transient mixed-mode simulations has to be chosen such that at least one complete oscillation period is included. Usually, the simulation time was chosen large enough to prevent distortions due to a transient response at the first oscillations (Fig. 3).

B. *Parasitic Capacitance Extraction*

The loading capacitances are derived from 3D finite-element simulations of the complete layout (Fig. 4) using an industry standard field solver for a simplified, typically rectangular, geometry. We now assume that changes in the

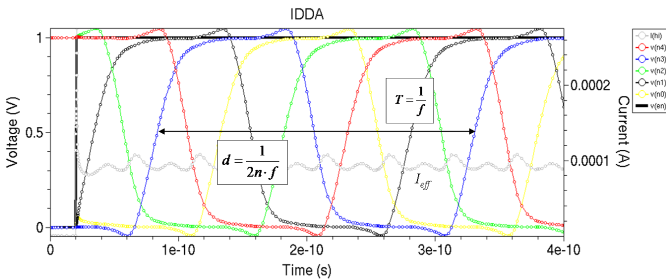

Figure 3. Transfer characteristics of a RO in active operation The nodal voltages for a five-stage oscillator are shown. I_{eff} is the total current drawn by the circuit. The delay is measured at $1/2V_{dd}$.

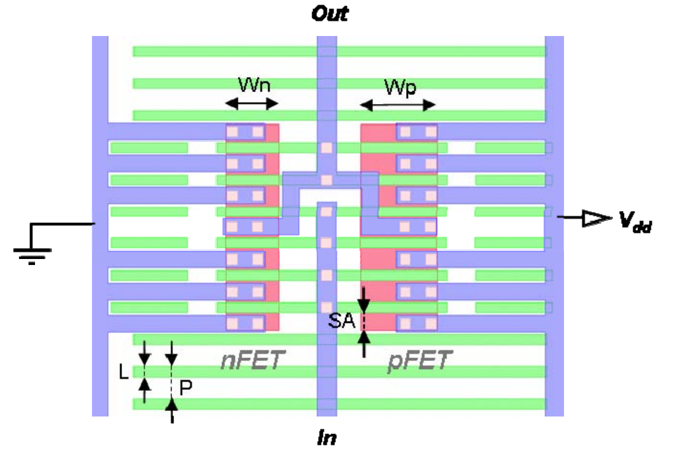

Figure 4. Layout of the inverter stage with FO3. The gate length is represented by L, and the gate-to-gate distance (pitch) by P. The width of the n- and the p-MOSFETs is represented by Wn and Wp respectively.

layout or the vertical structure for the inverter circuit can be captured by changes in the gate-to-contact capacitance which we call C_{ca2pc}. This capacitance is a component of each of the five parasitic loading capacitances, C_i, in the RO circuit. Therefore, we can write each parasitic capacitance as a sum according to

$$C_i = C_{i,0} + n \cdot C_{ca2pc}(n) + p \cdot C_{ca2pc}(p) . \tag{3}$$

The quantity $C_{i,0}$ denotes all capacitive components of C_i, independent of the structural differences between the different MOL options. Also, the gate-to-contact capacitance for an n- and a p-MOSFET are different. The multipliers – n, and p respectively, account for the number of incidences of the n- and p-type capacitor, C_{ca2pc}, in the layout. Consequently, the difference for each of the loading capacitances between two different device structures can be written as

$$\Delta C_i = n \cdot \Delta C_{ca2pc}(n) + p \cdot \Delta C_{ca2pc}(p) . \tag{4}$$

The parasitic loading capacitances and the structure multipliers for the inverter structure considered are given in Table I.

TABLE I. PARASITIC CAPACITANCES FOR THE INVERTER CIRCUIT AS IN FIG. 2 AND FIG. 4

		n	p	C_i (fF)
C_1	IN-GND	6	0	0.146
C_2	IN-Vdd	0	6	0.177
C_3	OUT-GND	4	0	0.152
C_4	OUT-Vdd	0	4	0.121
C_5	IN-Out	2	2	0.150

According to Eq. 4 the task of accounting for the effect of layout changes for the parasitic loading capacitances is reduced to calculating the gate-to-contact capacitance for all simulation scenarios. The MOL contact geometry has to be modeled in 3D since not all contacts fully strap the active device area (Fig. 5).

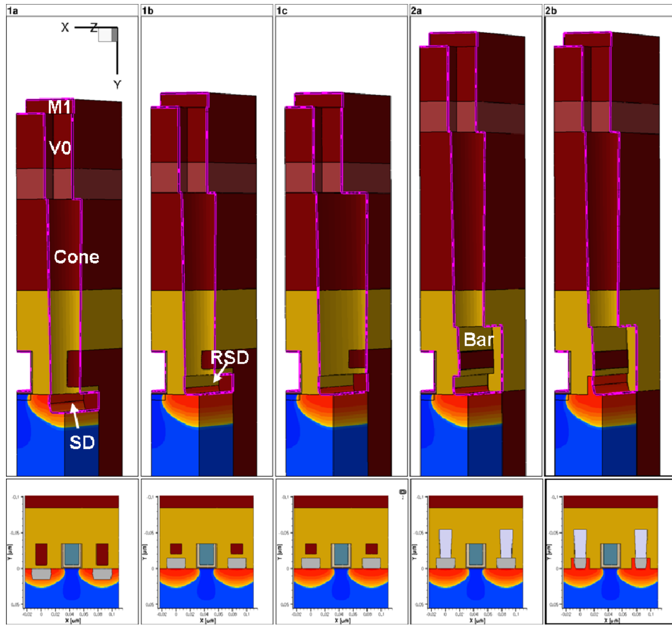

Figure 5. Three-dimensional and two-dimensional n-MOSFET structures representing the five design options. Shown is one quarter of the complete structure including the active area, the RMG, the STI, and the MOL contacts. The contact materials are removed in the 3D structure and corresponding boundary conditions are applied.

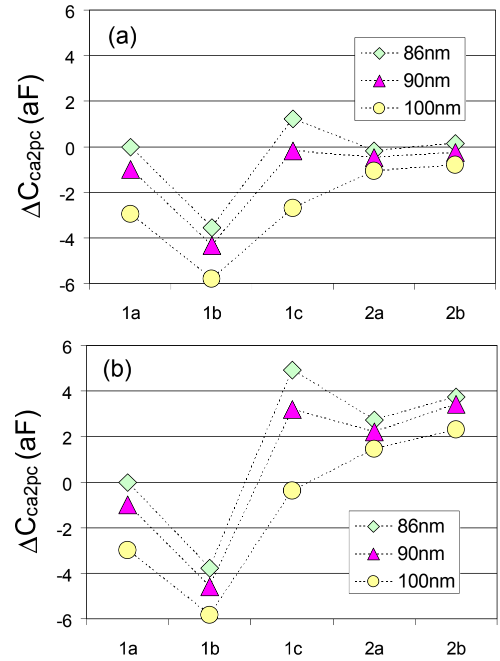

Figure 6. Relative changes in gate-to-contact capacitance for (a) n-MOSFET and (b) p-MOSFET. The capacitance is given for three different values of the gate-to-gate distance (gate-pitch).

The simulation geometry is generated by extruding the 2D MOSFET cross-section into the third dimension. All active regions of the device, the Silicide contact, and the RMG are included in the extruded structure. We add the MOL elements such as the contact bar and contact cone, and also include a shallow-trench isolation Oxide to terminate the device and provide accurate electrostatic boundary conditions. This is done after the extrusion of the 2D cross-section using a 3D device editor (Fig. 5). Small-signal capacitance simulations are then used to extract C_{ca2pc} for the n- and p-MOSFET.

III. Analysis and Results

A. Simulation Structures

We explored five design options with different contact configurations. The structures representing the device options are show in Fig. 5 wherein the upper-panel shows the three-dimensional structures used to extract the gate-to source capacitance. The lower panel shows the corresponding two-dimensional structures from the process and device simulations for the front-end-of-line (FEOL) which are used in the mixed-mode simulations for the RO.

We will refer to option *1a* as the base option, i.e. a flush-filled source-drain region and a contact cone that connects the contact Silicide with the metal levels which are referred to as V0 and M1 in Fig. 5. Options *1b* and *1c* introduce a RSD of 15nm height; option *1b* with a contact cone which fully straps the active device area. Note that we keep the doping profile the same for a flush-filled- and a raised source-drain based structure. This artificial constraint has been introduced since we intend to investigate the effect of changes to the MOL contact configuration only, ignoring for instance the difference that this may entail for the doping profiles, i.e. the FEOL. This constraint does not have to be introduced necessarily, and follow-up work beyond the exploratory stage of device design we are in will take this into account.

Design options *2a* and *2b* introduce a contact bar which connects the contact Silicide with the contact cone. The contact bar is always assumed to fully strap the active device area. Design *2a* differs from design *2b* as it allows the contact Silicide to consume the complete length of the RSD. In option *2b* the Silicide growth is only allowed to consume the Silicon immediately below the contact bar.

B. Gate-to-Contact Capacitances

We show the relative changes of the gate-to-contact capacitance for all design options depending on gate-pitch in Fig. 6. All results are normalized with respect to option *1a* and a value for the pitch of 86nm, i.e. ΔC_{ca2pc} is zero for option *1a* and a gate-to-gate distance of 86nm since the parasitic capacitances in Table I are extracted for a layout that assumes a 86nm pitch and vertical structure according to option *1a*. The values of ΔC_{ca2pc} for all other structures are then used to calculate the change in the parasitic capacitances using Eq. 4 and the polarity dependent multipliers in Table I.

The capacitive change due to the RSD is captured in a drop in C_{ca2pc} by ~4aF between *1a* and *1b*. Note that the gate-to-contact capacitance for the RSD case is lower then for the structure with a flush filled source-drain. The reason is the smaller capacitive coupling of gate and contact because of the reduced gate side-wall area. Of course there will be an increase

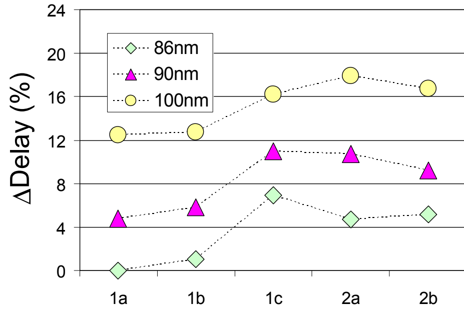

Figure 7. Percentage change in ring-oscillator delay for three different values of gate-pitch.

in the capacitive component which couples the RSD to the gate. In the mixed-mode simulations this component is included in the 2D MOSFET model that accounts for the overlap capacitance. A large increase of the gate-to-contact capacitance is observed for a fully strapped contact cone (*1c*). This increase is smaller for larger pitch values since the contact cone has the same width for all values of pitch. The introduction of a contact bar with a narrow cone on top lowers the capacitance again (*2a, 2b*) by ~1aF.

C. Ring-Oscillator Delay

The percentage change in RO delay based on the mixed-mode analysis of the full five-stage circuit is shown in Fig. 7. Strong gate-pitch dependent degradation is observed. The relative increase in the delay is small between *1a* and *1b,* and between *1c*, *2a*, and *2b*. The largest degradation occurs when introducing a fully strapped contact between design options *1b* and *1c*. In order to understand the origin of the degradation in RO delay, we need to extract C_{eff} according to Eq. 2, and R_{eff} according to $d=C_{eff}R_{eff}$ as discussed in Section II.

Fig. 8a shows an increase in the effective capacitance of around 4% for the RSD (*1a-1b*) and 8% for the design options with fully strapped contacts. There is only a relatively small gate-pitch dependence observed. The contact cone has a fixed thickness and does not scale with the gate-pitch. Therefore, we see the strongest gate-pitch dependence for case *1c*, where the cone fully straps the active device area. The contact bar in contrast does scale with the pitch. So, gate-pitch dependence is eliminated for cases *2a* and *2b* since the spacing between the bar and the gate remains unchanged. Fig. 8b shows a strong response of R_{eff} to the gate-pitch. With increasing pitch there is less Halo dose shadowing which increases the channel doping and therefore the threshold-voltage. This in turn results in a lower drive current and a higher effective resistance since R_{eff}~$1/IDDA$. Also observed is a drop in R_{eff} of ~3% for the RSD and another small drop of around 1% for the option *2b* which limits the Silicide to the area below the contact bar. These effects can be understood as being related to the location and shape of the Silicide. In the first case (*1b*) the elevated contact is surrounded by an area with higher doping concentration whereas in the second case (*2b*) the overall

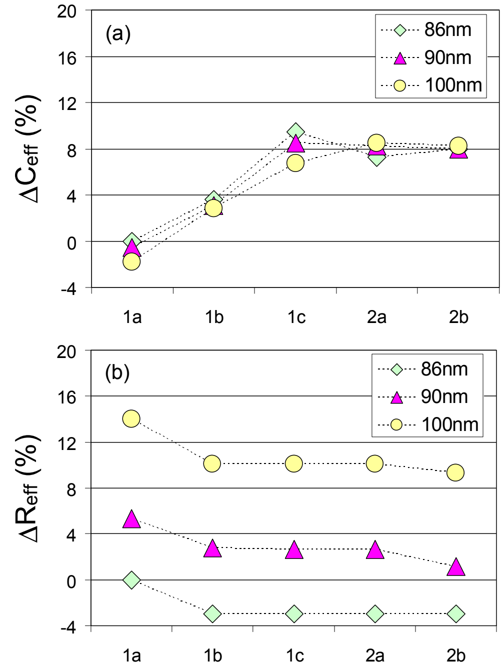

Figure 8. Percentage change in effective capacitance (a) and effective resistance (b) for three different values of gate-pitch.

contact area exposed to the current is larger. Both design options are in effect lowering the external resistance (see [1] or [5] for instance). The strong gate-pitch dependence of the effective resistance seen in our simulations is identified as the root cause for the pitch dependences in the ring-delay.

IV. CONCLUSIONS

Mixed-mode TCAD simulation was demonstrated to be a powerful tool to study the impact of various MOL designs on the high-frequency response of a CMOS logic component such as an inverter cell or a ring-oscillator. An industry standard 20nm technology provided the vehicle for this exercise. In particular the capacitive and resistive penalties associated with certain design options such as raised source-drain and various contact designs can be evaluated in detail. The causes for ring oscillator slowdown such as gate-pitch effect and contact geometry can be identified.

REFERENCES

[1] T. Uchino, T. Shiba, K. Ohnishi, A. Miyauchi, M. Nakata, Y. Inoue, and T. Suzuki, “A Raised Source/Drain Technology Using In-situ P-doped SiGe and B-doped Si for 0.1-um CMOS ULSIs,” *Tech. Dig. of International Electron Device Meeting (IEDM2007)*, 2007, pp.479-482.

[2] M. B. Ketchen, and M. Bhushan, “Product-representative ‘at speed’ test structures for CMOS characterization,” *IBM J. Res. & Dev*., vol. 50, 2006, pp.451-468.

[3] Sentaurus TCAD Tools, Synopsys, Inc.

[4] M. B. Ketchen, M. Bhushan, and D. Pearson, “High Speed Test Structures for In-Line Process Monitoring and Model Calibration,” *Proc. IEEE 2005 Int. Conf. om Microelectronic Test Structures*, vol. 18, 2005, pp.33-38.

[5] S.-D. Kim, S. Jain, H. Rhee, A. Scholze. M. Yu, S. Furkay, M. Zorzi, F. M. Bufler, and A. Erlebach, “Modeling Gate-Pitch Scaling Impact on Stress-Induced Mobility and External Resistance for 20nm node MOSFETs,” *SISPAD Proc*., 2010, pp.72-82.