Capacitive Gate Insulator Thickness and its Impact on Static and Dynamic Behavior of Scaled PD-SOI-MOSFET

T. Herrmann*, W. Klix*, R. Stenzel*, T. Feudel, J. Hoentschel**, M. Horstmann****
*University of Applied Sciences Dresden, Germany
E-mail: tom.herrmann@et.htw-dresden.de
**AMD Saxony LLC & Co. KG, Germany

Abstract
The influence of technological parameters on the capacitive gate insulator thickness and the electrical transistor behavior of different technology nodes were simulated. Starting from calibrated PD-SOI-MOSFETs, designed for the 130 nm technology, scaled 45 nm technology transistors were examined. The inverter propagation delay time and corresponding saturation drive currents of different technologies with varying gate insulator thicknesses were compared under the aspect of manufacturability. Further improvements due to the implementation of metal gates were estimated.

1. INTRODUCTION

Downscaling is a well known performance enhancer for high performance logic products. The International Technology Roadmap for Semiconductors (ITRS) [1] sets strict guidelines for several transistor parameters, which could not be obtained without new manufacturing processes. Performance enhancer in the last years was the reduced physical and capacitive equivalent thickness (CET) of the gate insulator. Further scaling of them is almost impossible without high-k insulator materials. Barring the gate leakage issue, thinner gate insulators increase the electric field in the channel region. This will increase the inversion charge density, but also reduce the carrier mobility and force quantum effects in the vicinity of the gate oxide. The ITRS determines constant voltage scaling towards the 45 nm technology, which yields these two opposed effects. Aim of the gate insulator scaling is a better control of the channel charge in the sub threshold region and higher saturation currents due to the increased inversion charge density. Performance improvement was examined through simulations with a single inverter structure, which include drive currents and capacities.

2. MODELS AND VALIDATION

For this study the tools FLOOPS and DESSIS were used for process and device simulation. The manufacturing process is based on the 130 nm technology with a gate length of 55 nm similar to [2]. The amount of active dopants in the poly silicon gate is constant with $1 \cdot 10^{20}$ cm^{-3} for n-MOS and $5 \cdot 10^{19}$ cm^{-3} for p-MOS devices. A validation based on measured doping profiles for different annealing and implantation conditions was done.

Device simulation combines the drift diffusion model with the density gradient model and several other models like band to band tunneling. The drift diffusion model is limited especially for the 45 nm technology transistors, but deliver fast and acceptable results with some parameter modifications [3]. Carrier mobility degradation at interfaces is described with the enhanced Lombardi model and adjusted for the density gradient model, based on the parameter set [4]. The density gradient model introduces a quantum potential which modifies the band energy in the carrier density distribution formula. Quantum effects shift the maximum of the carrier density away from the interface and therefore reduce the influence of the Lombardi model. This leads to a decreasing gate capacitance and increasing CET. Besides the quantum effects, poly depletion affects the gate insulator thickness and device performance mainly.

Electrical characteristics of the 130 nm technology transistor were validated with measured data from a comparable process. All simulations were done with partially depleted silicon on insulator (PD-SOI)-MOSFETs of constant film thickness and general process conditions, because of the well known and optimized manufacturing process and comparability of the results.

The inverter structure is simplified to one single inverter and a capacitive load. The load capacitance consists of a constant part for wiring capacitance and a varying part for the following gate capacitance according to the simulated device. The input pulse is chosen constant with 2 ps rise and fall time.

3. SIMULATION RESULTS

3.1 BASIC TRANSISTOR

The standard process of the 130 nm technology transistor yields drive currents of 681 µA/µm (n-MOS) and 363 µA/µm (p-MOS). A gate insulator with 2 nm physical thickness and material parameters of silicon dioxide achieves a CET of 2.63 nm (n-MOS) and 2.92 nm (p-MOS) arising from quantum effects and depletion effects in the poly silicon and the silicon film. The simplified inverter results in a propagation delay time of 5.1 ps, which is a little below the measured values because of its simplified structure ignoring further parasitic elements. A load capacitance of 3 fF is used for the unscaled MOSFET.

Table 1. Performance of the standard transistor for physical gate insulator scaling

	n-MOS		p-MOS	
T_{ins} (nm) reduction	0.1	0.2	0.1	0.2
$I_{D,sat}$ (%) improvement	5.6	11.9	3.8	8.5
$I_{D,sat}$ (%) improvement @const. $I_{D,off}$	4.1	8.7	3.5	7.1
CET (nm) reduction	0.1	0.2	0.1	0.2

Effects of a reduced physical gate insulator thickness are shown in Table 1. The inverter propagation delay time decreased 4.8% and 8.6% compared to the basic transistor for 0.1 nm and 0.2 nm thinner insulator. Electron density at the poly silicon- and silicon-insulator interface is shown in Figure 1. It is slightly increased for a thinner insulator inside the channel region. There is also a strong influence of the poly depletion visible, which reduces electron density at the interface to $2 \cdot 10^{18}$ cm^{-3} and affects the first 2.5 nm above gate insulator.

Simulation of the inverter without capacitance matching shows degraded propagation delay time improvement compared to the drive current. An alignment of load capacitance to simulated transistor capacitance lowers the propagation delay time, further. The values are improved 3.6% for 0.1 nm and 6.2% for 0.2 nm reduced gate insulator thickness. If the width of the p-MOS transistor is aligned, the propagation delay time for the falling and raising edge would be equal.

Therefore increasing gate capacitance is the main reason for slower enhancement of the inverter speed. The maximum of the inversion carrier density moves little closer to the interface but poly silicon depletion is more pronounced. So the CET reduction is equal to the thinning of physical gate insulator thickness.

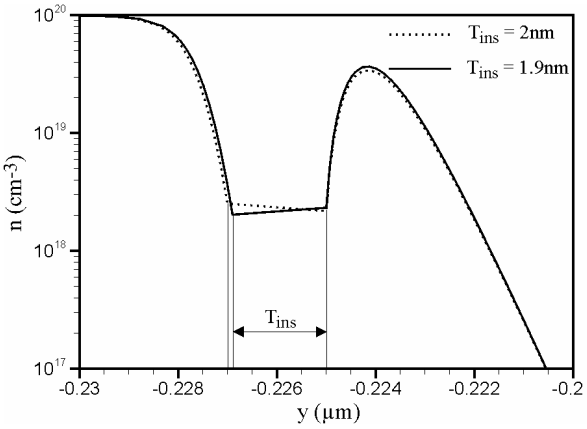

Fig. 1. Electron density (n) for 2 nm and 1.9 nm physical gate insulator thickness (T_{ins})

The universal curve in Figure 2 shows the behavior of the basic n-channel transistor with reduced physical gate insulator thickness. Remarkable is the flattening of this curve, which leads to a higher gain for super nominal transistors. The leakage current increases and therefore the advantage of pure insulator scaling will be further reduced for adjusted threshold voltages. The threshold voltage behavior for smaller gate length is improved with thinner gate insulators.

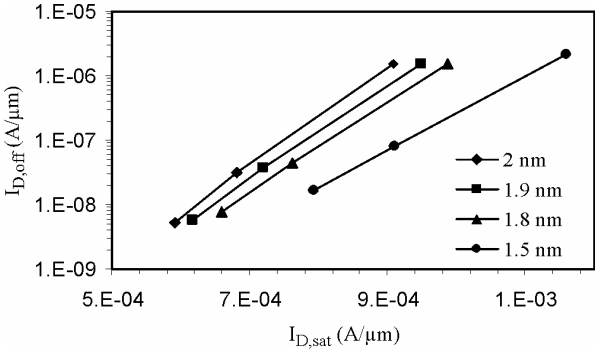

Fig. 2. Universal curve for n-MOSFET with different insulator thickness

3.2 SCALED TRANSISTOR

The transistor according to the 45 nm technology from the ITRS has 20 nm gate length. To obtain a useful channel length with

appropriate doping concentration at the gate edge, advanced annealing has to be introduced. The annealing time is set to 5 ms and temperature to 1300 °C. The doping concentration at gate edge is about 1.5 times higher than in basic transistors. Physical insulator thickness is set to 0.9 nm, which is 0.2 nm higher than the ITRS guidelines. This value is the state of the art limit for SiON insulators to achieve the leakage requirements. Advanced annealing can also improve the amount of active dopants in the poly silicon. Therefore the doping concentration is set to $1.3 \cdot 10^{20}$ cm^{-3} (n-MOS) and $1 \cdot 10^{20}$ cm^{-3}.Threshold voltage is aligned through the halo implantation dose, which is 2.4 times higher for n-MOS and 3.5 times higher for p-MOS. The implantation energy and tilt are not changed. Therefore the halo regions overlap in the channel and form an almost constant lateral doping profile.

The drive current of the scaled transistor is 1020 µA/µm (n-MOS) and 537 µA/µm (p-MOS). The n- and p-MOS drive current improvement is equal, because of the compensation of the reduced CET by less rising p-MOS drive current per Angstrom gate insulator thickness reduction. Thereby a CET of 1.55 nm (n-MOS) and 1.78 nm (p-MOS) is achieved. The inverter propagation delay is 2.9 ps with a 2 fF load capacitance. The reduced capacitance takes into account lower gate capacitances and 20% lower parasitic capacitances due to scaling. Furthermore, the p-MOS width is aligned to the n-MOS drive current.

The physical insulator thickness of transistors is reduced from 1 nm to 0.9 nm. Hence, the saturation current increases 6.4% (n-MOS) and 8.9% (p-MOS). A constant leakage current reduces these values to 4.8% (n-MOS) and 4.1% (p-MOS), which is in the range of a basic transistor structure for 0.1 nm reduced gate insulator thickness. A CET reduction of 0.11 nm (n-MOS) and 0.1 nm (p-MOS) is therewith linked. The maximum carrier density is approximately 0.5 nm and 0.8 nm away from the interface.

The scaled transistor has a noticeably flatter universal curve (Figure 3) than the basic transistor. A thinner gate insulator did not flatten the universal curve as much as in case of the basic transistor. The saturation threshold voltage reduction for shorter gate length is worse because of the short channel and the overlapping halo regions. So, the process induced gate length fluctuation needs better control.

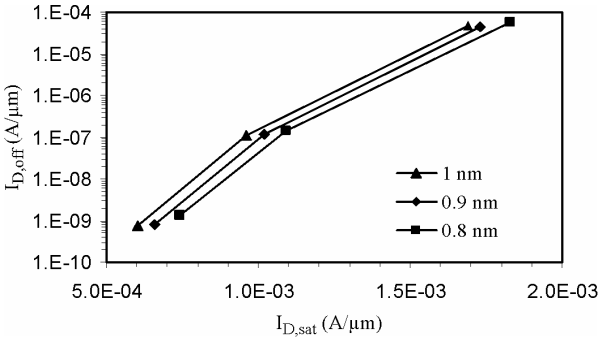

Fig. 3. Universal curve for scaled n-MOSFET with different insulator thickness

The reduction of gate insulator thickness results in a stronger increase of carrier density in the channel region. Due to a stronger electrical field, the poly depletion effect is more pronounced compared to the basic transistor, shown in Figure 4. In spite of the increased carrier density, drive current improvement is only slightly higher. A higher carrier surface concentration in the channel region could be the reason, which reduces the carrier mobility. Furthermore, the peak carrier concentration is greater and closer at the surface. The electron quantum potential differs more as a consequence of the larger carrier density difference.

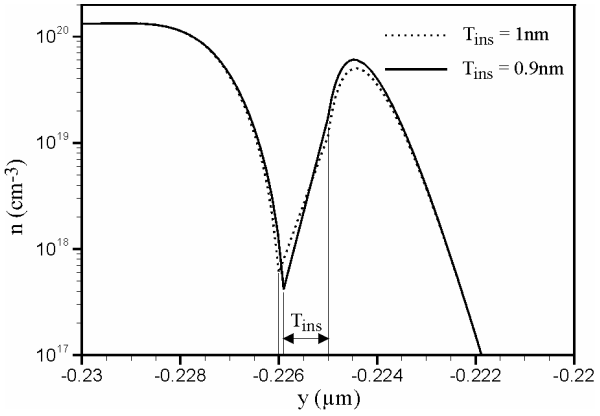

Fig. 4. Electron density (n) and for 1 nm and 0.9 nm physical gate insulator thickness (T_{ins})

The vertical hole density distribution is wider due to stronger hole confinement in silicon. Furthermore, density at the semiconductor insulator interface is almost equal and peak concentration does not increase that much like electron density. This partly explains the smaller

drive current improvement and higher CET. Another factor is the stronger poly depletion, which results in a carrier concentration at the poly silicon insulator interface of only $1 \cdot 10^{17}$ cm^{-3}.

Inverter propagation delay times of 3.1 ps are achieved with the 1 nm thick gate insulator. A gate insulator shrinking to 0.9 nm yields a 5.5% faster inverter, which does not fully reach the saturation current improvement of the single transistors. A matched load capacitance of 1.97 fF for the small gate insulator results in a 4.9% faster inverter.

The simulated drive currents for the 45 nm technology transistors do not meet the ITRS guidelines with current possible process conditions. A 0.7 nm thick gate insulator as promised by the ITRS will only have about 1200 µA/µm drive current, which is well below the 1900 µA/µm.

3.3 METAL GATE

Using a metal gate instead of poly silicon decreases the CET to about 1.12 nm for the n-MOS and 1.19 nm for the p-MOS side at 1 V gate to source voltage. The physical gate insulator thickness is thereby 0.9 nm. In this case, the influence of the quantization effects in the channel region can be seen without being overlaid by poly silicon effects, which are eliminated by the metal band structure. This lowering of the CET is combined with a 28% and 31% increased drive current for n- and p-MOS. A higher drive current at the p-MOS side is caused by stronger CET reduction, despite fewer drive current benefit per Angstrom. To achieve an identical threshold voltage, the halo implantation dose has to be reduced and therefore the doping dependent mobility increases.

The barrier height between metal gate and silicon is set to -0.5 and 0.5 for n- and p-MOSFET and is equal according to its amount. Consequently the gate parameters are fixed for varying electrical boundary conditions.

The propagation delay time decreases 24% with an identical load capacitance, which is less than the drive current improvement. Simulations with the matched load capacitance showed only a 15.6% faster inverter.

Drive currents of 1300 µA/µm (n-MOSFET) and 670 µA/µm (p-MOSFET) are also below the ITRS values for the 45 nm technology. The intrinsic switching delay time according to the ITRS will not exceed 0.5 ps, which is approximately 0.1 ps higher than the predicted value. The surface carrier density of metal and poly gate transistor is almost equal due to the stronger quantum effect, which could not be compensated by the vertical electrical field. Nevertheless, the peak carrier density is nearly two times higher, but also slightly more away from the interface. That's why there is still a difference between physical and capacitive gate insulator thickness.

4. CONCLUSIONS

Pure gate insulator scaling leads to approximately 4-5% enhanced drive current for 1 A lower physical gate insulator thickness at the same leakage current. The scaled transistor has slightly more improvement without leakage current matching. This enlargement could be not completely transferred to the inverter propagation delay time, due to higher intrinsic capacitances of transistors with reduced gate insulator thickness. The scaled transistor has only about 1% higher gain in terms of inverter propagation delay time for 0.1 nm gate insulator reduction. Usage of metal gates shows that problem too with even worse transition conditions. The drive current and intrinsic switching delay time requirements of the ITRS could not be reached with scaling and usage of metal gates. For this reason it is necessary to further improve other performance enhancers like strained silicon, which do not have negative impact on the capacitances.

The project described in this publication was funded by the German Federal Ministry of Education and Research, registered under funding number 01M3048A. The author named in the publication bears responsibility for all published contents.

REFERENCES

[1] ITRS 2004.

[2] M. Horstmann, D. Greenlaw, Th. Feudel, A. Wei, K. Frohberg, et al., "Sub-50nm gate length SOI transistor development for high performance microprocessor", *Mater. Sci. Eng. B*, vol. 114, 2004, pp. 3-8.

[3] J.D. Bude, "MOSFET Modeling Into the Ballistic Regime", *SISPAD 2000*, pp. 23-26.

[4] M.N. Darwish, J.L.Lentz, M.R. Pinto, P.M. Zeitzoff, T.J. Krutsick, and H.H. Vuong, "An Improved Electron and Hole Mobility Model for General Purpose Device Simulation", *IEEE Trans. Electron Devices*, vol. 44, pp. 1529-1538, September 1997.