

C-8-4

A Double-Gate device architecture optimization for sub-45nm digital CMOS technologies using cell-based timing analysis

R.Surdeanu, G.Doornbos, R.Ng, P.Christie, V.H.Nguyen, B.J.Pawlak, J.J.G.P.Loo, M.J.H.van Dal, Y.V.Ponomarev

Philips Research Leuven, Kapeldreef 75, 3001 Leuven, Belgium

Phone: +32-1628-8319 E-mail: radu.surdeanu@philips.com

1. Introduction

One of the most important reasons for introduction of consecutive CMOS generations is the reduction of circuit propagation delay. So far, there has been no systematic evaluation of the impact different source/drain architectures on this quantity. Here we present results of such a study for sub-45nm CMOS technologies by performing a dynamic system-level analysis with realistic wire-loading. We show that a 50% gain in the drive current while lowering 10 x the off-state leakage current can be obtained by source/drain architecture optimisation and furthermore, the system delay can be improved by a factor 2.

2. Device structure and design of experiments

A combination of experimental results and numerical simulations was used to provide a complete set of parameters for circuit analysis. We consider thin-body Si double gate devices as the mainstream device architecture for 32 nm node and beyond [1-4], since single gate devices fail to deliver the needed performance for these dimensions [5]. A schematic double-gate device structure for is shown in Fig. 1. It features metal gates with a fixed mid-gap work-function, high-k dielectric with EOT = 1 nm and undoped channel. The ratio between channel thickness and gate length, L_g , was always kept at the optimal value of 1/3. The design of experiments included variations of activation level and abruptness of the so-called Lowly Doped Source/Drain regions (LDD), position and metal work function of metal contact (e.g. silicide) and spacer material. The digital transistor performance (drive current I_{on} , off-state current I_{off} , short channel behavior (DIBL), sub-threshold swing and overlap capacitance C_{ov}) was optimized to achieve required values for Low Operating Power (LP) and General Purpose (GP) targets for sub-45nm nodes.

3. Results

We have observed the same trends from all gate lengths investigated (10-100 nm), therefore only $L_g = 20$ nm results are shown. The abruptness and activation levels of the LDD were investigated in the range given by various implant and anneal techniques expected to be used in CMOS processing such as pre-amorphization, ion implants, followed by fast ramp-rate spike, laser melt or Solid Phase Epitaxially re-growth (SPER). Figure 2 illustrates the impact of various LDD abruptnesses (open symbols) and activation levels (closed symbols) on device performance I_{on} (squares) and I_{off} (stars). Clearly, junctions activated above $3e20$ at.cm⁻³ and with abruptness smaller than 2 nm/decade are needed, which can be realised *only* by SPER [6] or laser anneal [7]. At higher activations and small abruptness, the potential barrier at the metal – doped region interface becomes higher and narrower, resulting in the increase in I_{on} and the

decrease in I_{off} . These type of junctions can be fabricated with 1 nm precision in positioning the doped regions and the S/D metal by SPER: after gate patterning, pre-amorphization and doping implants are performed, followed by SPER process at low temperatures, in which the regrowth rate is well-controlled[6]. In Fig. 1b, the XTEM is presented for a thin-body Si device after the LDD SPER step. Subsequently, the amorphous/crystalline interface can be used to precisely position the silicide or the metal for source/drain. Furthermore, we have observed that both position and type of source/drain metal has the most significant impact on overall circuit performance. Figure 3 shows that, for a highly activated ($3e20$ at.cm⁻³) and abrupt (1nm/dec.) LDD, by correctly positioning the S/D metal, one can obtain a $8 \times$ higher I_{on} compared to the conventional “spike” junction, accompanied by a reduction of $\sim 250 \times$ of the off-state leakage. The low I_{on} and high I_{off} obtained for positioning the metal under the gate are due to the unwanted Schottky behavior resulted from the doped region consumption by the S/D metal. As shown in Fig. 4, for an optimised metal position, activation and abruptness, by appropriately selecting the S/D metal work-function, up to a factor 2 in drive current enhancement can be obtained (for a workfunction of 4.1 eV for NMOS) compared to the best DG + SPER device. As expected, by modifying the dielectric constant of the spacer material for the optimised device in the previous steps, C_{ov} was observed to change substantially (not shown here). The optimal devices were analysed further with a cell-based timing analysis.

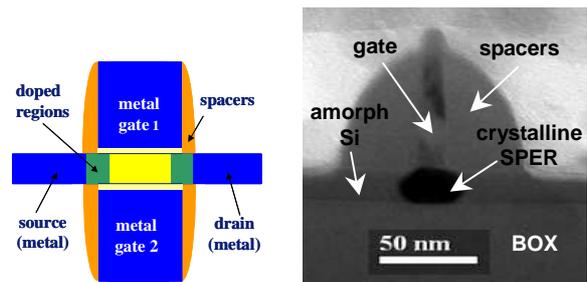


Fig. 1. (a) A schematic double-gate device structure for gate length $L_g = 20$ nm; (b) XTEM of a thin-body Si device after the S/D extensions SPER step.

4. Circuit analysis

To perform a dynamic system-level analysis of the various device options, we developed a inverter-based compact model, which allows the transient behavior of the devices to be evaluated based on just six parameters extracted from I - V data. These extracted parameters are (1) the nominal threshold voltage V_t , (2) the gain B , (3) the slope of the $I_d(V_d)$ curve in saturation λ , (4) the exponent n in $I_{d,sat} = (W/L) \cdot B \cdot (V_{gs} - V_t)^n$, (5) k and (6) the exponent m from: $V_{dsat} = k(V_{gs} - V_t)^m$.

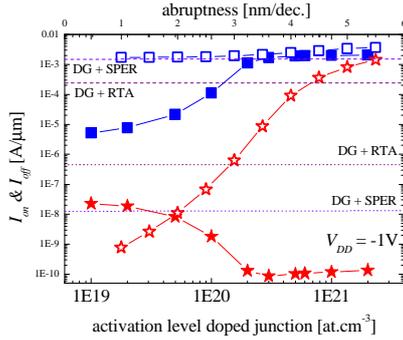


Fig. 2. Abruptness (open symbols) and activation level (closed symbols) impact on DG PMOS performance. The horizontal dashed (I_{on}) and dotted (I_{off}) lines represent the reference values obtained with a DG device similar to the one in Fig.1, where only “conventional” doped and annealed junctions are used, 40 nm spacers and NiSi for silicidation.

These parameters are used as inputs for the transient output of the inverter as a function of the input slew rate and output capacitive load. Calibration of our method was performed on 0.12 μ m technology circuits[8]. In this study, we use a signal path composed of seven inverters. The first inverter is driven with a step input voltage and the delay and output slew are calculated using closed form expressions derived from the output transient expression. The output slew of the first inverter is then used as the input slew of the second inverter, and so on, along an inverter chain. We have observed that after approximately four stages, the slew and delay reach their steady state values and no longer change with subsequent stages. Based on this observation, we have evaluated the stage delays for the different device options presented above assuming (a) the load is given by the cell inverter capacitance and therefore the delay corresponds to a ring oscillator stage delay and (b) the wire load additionally includes the interconnect load within a large standard cell array. Clearly, it is not possible to know individual wire lengths connecting cells within an array of cells prior to layout. However, for a long signal path, it is reasonable to replace each cell-to-cell wire by a wire of average length. Detailed analysis of average wire length can be found elsewhere [8]. Here we use a value of 31.3 μ m. This average wire length has been converted to a capacitive load by extracting the capacitance per unit length from the nominal 32 nm back-end shown in Figure 5, yielding the interconnect component of the capacitive load of 3.2 fF.

The results of the circuit analysis on the optimised devices are compared in Fig. 6, where the cumulative wire loaded delays with a realistic wire load are shown for various optimisation steps in source/drain architecture. By choosing the appropriate S/D metal, and correctly positioning it with respect to the gate and the highly-active highly-abrupt junction, the delay τ is improved by almost a factor 2 for a device with oxide spacers, giving $\tau=8.9$ ps.

6. Conclusions

We have presented a systematic evaluation of the impact of different source/drain architectures on digital MOS-FET and circuit propagation delay for sub-45nm CMOS

CMOS technologies by performing a dynamic system-level analysis with realistic wire-loading. We show that a 50% gain in the drive current while lowering 10 x the off-state leakage current can be obtained by source/drain architecture optimization and furthermore, the system delay can be improved by a factor 2.

References

- [1] ITRS 2004 update
- [2] H.S.P. Wong et al., *IEDM Tech. Dig.* **1997**
- [4] S. Monfray et al., *IEDM Tech. Dig.* **2001**
- [5] Y.V. Ponomarev et al., *VLSI Symp.* **2001**
- [6] R. Lindsay et. al, *Proc. MRS* **2003**
- [7] C.J.J. Dachs, *USJ Workshop* **2003**
- [8] P. Christie, *IEDM Tech. Dig.* **2004**

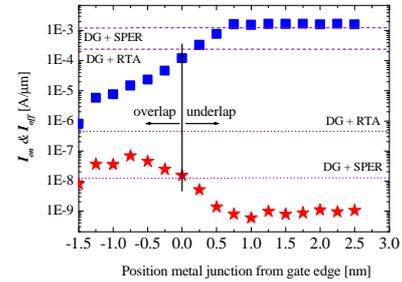


Fig. 3. Position of source/drain metal contact impact on device performance.

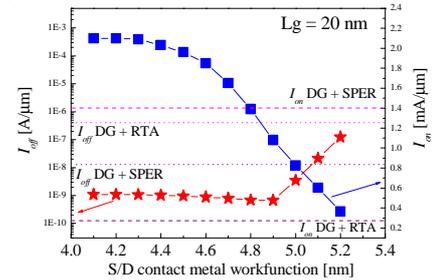


Fig. 4. Source/drain contact metal choice impact on device performance for NMOS transistors.

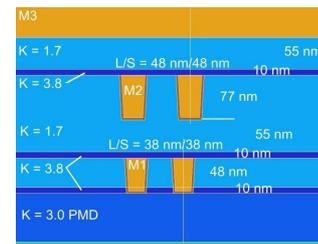


Fig. 5. 32 nm back-end in this study

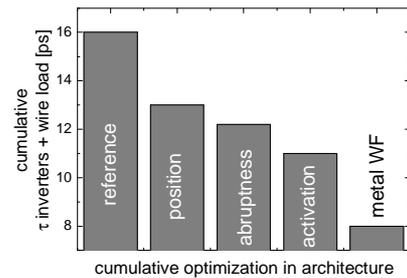


Fig. 6. Source/drain contact metal choice impact on device performance for NMOS transistors.