Extended Abstracts of the 2009 International Conference on Solid State Devices and Materials, Sendai, 2009, pp808-809

Exploration of Device Design Space to Meet Circuit Speed Targeting 22nm and Beyond

Lan Wei*, Frédéric Boeuf+, Dimitri Antoniadis^, Thomas Skotnicki+, H.-S. Philip Wong*
* {lanw, hspwong}@stanford.edu, Stanford University, Stanford, California, USA
+{frederic.boeuf, thomas.skotnicki}@st.com, STMicroelectronics, Crolles, France
^daa@mtl.mit.edu, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

INTRODUCTION

Historically, the delay of CMOS devices is benchmarked by the intrinsic delay $C_{gc}V/I_{on}$, where C_{gc} is the intrinsic gate-to-channel capacitance, and I_{on} is the on-state drain current in saturation [1]. However, for a CMOS logic gate, the devices usually do not operate at the bias point which gives I_{dsat}. Instead, the whole trajectory of the device operating points matters. [2-5] studied the impact of different regions of device IV characteristics on the circuit delay performance and developed different effective current metrics for a better estimation of the circuit delay. On the other hand, at the device level, MOSFETs are benchmarked by characteristics such as on-state drain current (I_{on}), off-state drain current (I_{off}), DIBL, current booster factor (k_{vs}) and so on. These device-level characteristics are often used as targets or specifications for technology development. Establishing the connection of these device-level targets with circuit-level performance is the subject of this paper. In this paper, we examine the impact of several device-level characteristics on circuit performance for various applications and we study the scaling trends.

DEVICE MODEL AND CIRCUIT SIMULATION METHODOLOGY

As the devices are scaled down, short channel device suffers from poor gate electrostatics and other parasitic effects. High-k materials are introduced to improve the electrostatic control. Mobility and velocity boosters, e.g. strain, are introduced to keep I_{dsat} boosted. At the device level, these improvements are revealed as lower subthreshold swing (SS), smaller DIBL, higher k_{vs} and so on. A new semi-empirical, physical IV model [6] is used which directly captures these behaviors. Parasitic capacitances are modeled following [7]. Parameters are characterized to fit devices characteristics at 32nm technology [8] (Fig. 1 & 2, fitting parameters in Table 1). Devices for 22nm technology are projected based on parameters and features listed in Table 2 [1]. Three different applications, i.e. high performance (HP) with low threshold voltage (LVT), low operational power (LoP) with standard threshold voltage (SVT) and low stand-by power (LstP) with high threshold voltage (HVT), are included. Circuit operation of 4-stage FO4 inverter chain and 2-input NOR chain of these devices are modeled with MASTAR [9]. Transient simulation with step input is done based on the IV model [6] and the parasitic capacitance models [7]. The average delay from the 50% input level to the 50% output level is extracted.

IMPACT OF DEVICE CHARACTERISTICS ON CIRCUIT SPEED

DIBL and I_{on} : During the switching process, the MOSFET that is turned on operates at the above-threshold region, where the gate voltage (V_g) is larger than the threshold voltage (V_{th}). Larger DIBL and k_{vs} can both result in a larger I_{on}. However, the origins of the larger I_{on} are quite different and the impacts on *IV* characteristics are different. Compared with long channel devices, devices with higher DIBL have larger drain current under high V_d, with the same current at low V_d. However, devices with large k_{vs} current boosters (mobility booster, injection velocity booster) have larger current for both high and low V_d. Different combinations of DIBL and k_{vs} may result in the same I_{on}, however, the different I_d-V_d behaviors result in different trajectories during switching as shown in Fig 3. Though I_{on} is kept at the same level in both cases by adjusting k_{vs}, the inverter chain built by devices with higher DIBL switches at a lower speed. This is universally true for HP, LoP and LstP cases. The LstP case is the most sensitive to DIBL, while HP is the least sensitive (Fig 3, 4). This is because when DIBL is changed by a fixed amount from the nominal value for all three cases, the absolute values of I_{eff} [2], i.e., the current passed by the transistors averaged over the switching trajectory, are changed by a similar amount for all three cases (Fig 3). However, since LstP devices have the smallest nominal I_{eff}, the change in delay is the largest, as shown in Fig 3(c). To achieve a 10% speed-up of inverter-chain at 32 nm node, k_{vs} should be boosted to get a 15% higher I_{on} for all three applications; or DIBL can be reduced to 80mV/V, 95mV/V and 110mV/V for HP, LoP and LstP, respectively (Fig 4).

I_{off} and V_{dd} : I_{off} and V_{dd} are key factors for power consumption. I_{off} is specifically important for static power, while V_{dd} plays an important role for both static and dynamic power. Depending on the application, power and delay are traded off by engineering V_{th} (thus, I_{off}) and V_{dd}. The I_{off} drops exponentially with increasing V_{th} while I_{on} drops proportionally to (V_{dd}-V_{th}) to the first. Among HP, LoP and LstP, delay is the most sensitive to I_{off} and V_{dd} for the LstP case, and the least sensitive for the HP case. This is because a change in (V_{dd}-V_{th}) results in a proportional change of I_{eff}, but the percentage change is the most in LstP and the least in HP. At the 32nm technology, in order to achieve a 10% speed-up by trading-off power consumption, I_{off} should be increased by 2.5x, 2x and 1.7x for HP, LoP and LstP, respectively; or V_{dd} should be increased by 0.1V, 0.07V and 0.03V for HP, LoP and LstP, respectively (Fig 5).

More complex circuits : In NAND and NOR circuits, the stacked devices operate in the linear region most of the time [4]. For a balanced design with the same equivalent I_{on} as inverters, NAND and NOR gates have larger gate capacitances, thus the switching trajectory of the parallel devices in NAND and NOR chains are further from the saturation point than the inverter chains since the devices are turned on more slowly. (Fig 3(d)). Thus, the delay is less sensitive to I_{on}. For the same balanced NMOS and PMOS, I_{eff} of NAND and NOR circuits [4] are much smaller than that of the inverter. The delay is more sensitive to DIBL, I_{off} and V_{dd}, since the percentage change of I_{eff} due to a certain DIBL, I_{off} and V_{dd} is larger for NAND and NOR which have a smaller nominal I_{eff} than inverter. In general, the further the switching trajectory is from I_{on}, the more sensitive the delay is to DIBL, V_{th}(I_{off}) and V_{dd}.

SCALING TRENDS

As devices are scaled down, oxide thickness and channel length scaling are less and less effective. Increasing efforts are put into engineering current boosters and improving gate electrostatic control. With the nominal design parameters listed in Table 2, 22nm device delay performance are predicted as in Fig 7. For example, in HP application, the 22nm device with DIBL of 80mV/V and I_{on} of $1.1I_{on\_32nm}$ has a 20% smaller switching delay than the nominal 32nm device for an FO4 inverter chain. To get a 17% speed-up over 32nm technology, devices have to fall into the upper-left regime above the dashed lines in the DIBL-I_{on} contour map. The MASTAR program [9] is used to engineer physical device design, to meet the requirement as predicted by Fig 7. For HP (Fig 7(a)(d)), it is not practical to meet 17% speed-up requirement by solely engineering DIBL by aggressively designed junctions. Effort of boosting I_{on} by continuing the engineering of strain or introducing high mobility material such as III-V/Ge, is necessary. This demand is more strict on complex circuit (such as NOR), than simple inverter chains. (Fig 7(d)). For LoP, switching to UTBSOI at 22nm can practically bring DIBL down to the regime which satisfies the scaling requirement of 17% speed-up per generation. For LstP, the requirement on DIBL improvement is further relaxed. Engineering the DIBL is more effective than introducing new channel material for LstP and LoP.

CONCLUSION

Starting from the device behavior model, we carefully analyze the impact of device characteristics on the circuit performance. The requirements of device characteristics are predicted as a guide for continuing technology scaling. With the prediction, UTBSOI structure is practical for LoP and LstP applications at 22nm technology. However, a combination of improved DIBL and transport enhancement is still necessary for HP application.

Acknowledgement

The authors would like to thank Z. Yao and S. Li for programming help. D. Antoniadis, L. Wei, and H.-S. P. Wong are supported in part by the Focus Center Research Program MSD and C2S2. L. Wei and H.-S. P. Wong are also supported by the member companies of the Stanford Initiative for Nanoscale Materials and Processes (INMP). L. Wei is additionally supported by the Stanford Graduate Fellowship.

Table 1 Key parameters for nominal cases at 32nm

32nm	Physical meaning	NMOS HP/LoP/LstP	PMOS HP/LoP/LstP
R_s	Series resistance(Ω-μm)	350	350
$DIBL$	Drain induced barrier lower (mV/V)	130	160
SS	Subthreshold slope (mV/dec)	98	98
T_{inv}	Electric oxide thickness in inversion (nm)	1.2	1.4
CPP	Contact poly pitch (nm)	112	112
L_g	Channel length (nm)	30	30
V_{dd}	Supply voltage (V)	1	1
I_{on}	On-state current (μA/μm)	1584/1293/756	1257/1034/619
I_{off}	Off-state current (nA/μm)	100/10/0.1	100/10/0.1

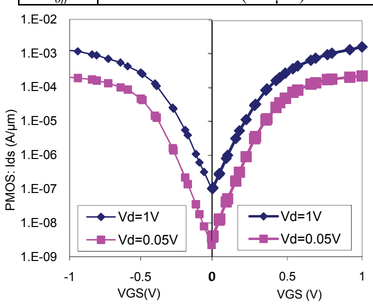

Figure 1 I_d-V_g characteristics in [8] is fitted. Lines = model. Symbols = experimental data [8].

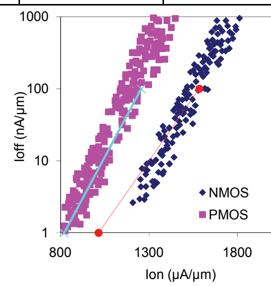

Figure 2 I_{on}-I_{off} The model is fitted across a wide range of I_{off}. Lines = model. Symbols = experimental data [8].

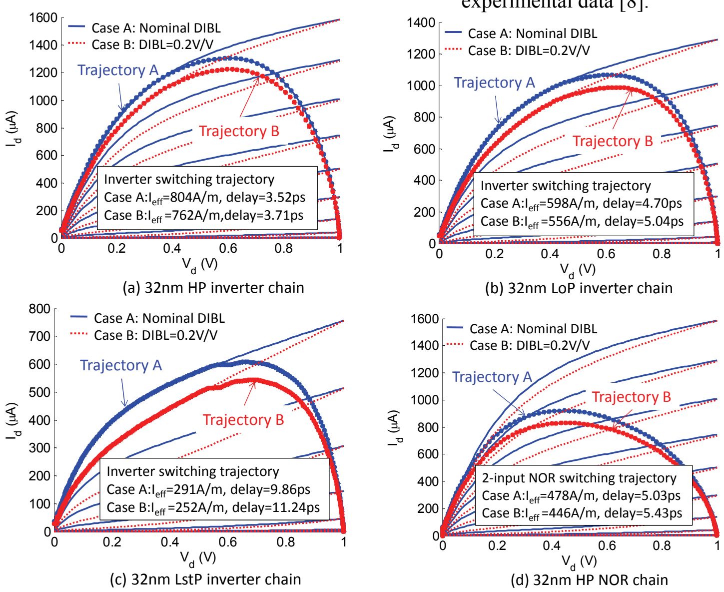

Figure 3 Switching trajectories I_{on} in Case A and Case B are the same, while Case B has a degraded DIBL of 200mV/V. The absolute difference in I_{eff} are similar (~40A/m) for inverter chains with (a) HP, (b) LoP and (c) LstP devices.

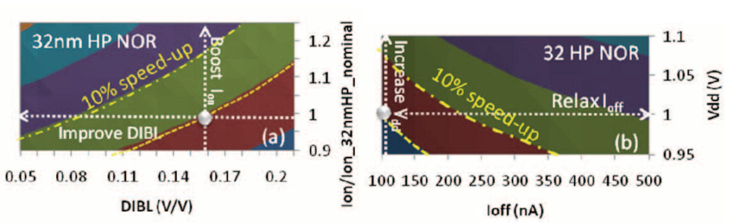

Figure 6 NOR chain delay improvement contours on (a) DIBL-I_{on} booster plane and (b) I_{off}-V_{dd} plane, with 32nm HP devices. Delay of a NOR chain are more sensitive to DIBL, I_{off} and V_{dd} than that of an inverter chain (Fig 5(a) and Fig 6(a)).

References

[1] ITRS http://www.itrs.net/reports.html
[2] M. Na et al., *IEDM*, p. 1391, 2002.
[3] E. Yoshida et al, *IEDM*, p. 195, 2006.
[4] K. von Arnim et al., *IEDM*, p.483, 2007
[5] J. Deng and H.-S. P. Wong, *T-ED*, p. 317-1322, 2006.
[6] A. Khakifirooz et al, *T-ED*, p. 1674-1680, 2009.
[7] L. Wei et al, *VLSI-TSA*, p. 78, 2009
[8] S. Natarajan et al, *IEDM*, p. 941, 2008.
[9] MASTAR http://public.itrs.net/models.html

Table 2 Key parameters for nominal cases at 22nm

22nm	HP	LoP	LstP
	nFET/pFET	nFET/pFET	nFET/pFET
CPP (nm)	90	90	90
V_{dd} (V)	0.9	0.9	0.9
EOT (nm)	0.7	0.8	1
T_{inv} (nm)	0.8/1.1	1.0/1.2	1.1/1.3
L_g (nm)	22	22	22

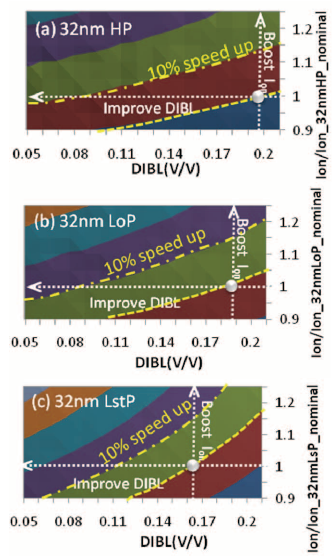

Figure 4 Inverter chain delay improvement contours on DIBL-I_{on} imrpovement plane with (a) HP, (b) LoP and (c) LstP at 32nm. I_{on} improvement includes effect of DIBL and k_{vs}. Each stripe is a 10% improvement of the delay with a combination DIBL and I_{on} booster over that of the 32nm nominal device. The slopes of the contours correspond to the sensitivity of the delay over the parameter indicated by the axis. The dot-dashed line refers to 10% speed-up over the nominal case, indicated by the big white dot.

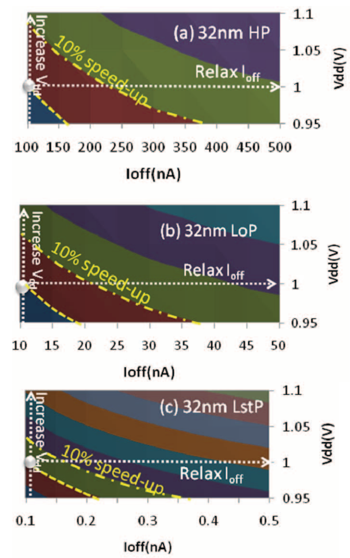

Figure 5 Inverter chain delay improvement contours on I_{off}-V_{dd} booster plane with (a) HP, (b) LoP and (c) LstP at 32nm. Each stripe is a 10% improvement of the delay with certain I_{off} and V_{dd} over that of the 32nm nominal device. The slopes of the contours correspond to the sensitivity of the delay over the parameter indicated by the axis. The dot-dashed line refers to 10% speed-up over the nominal case, indicated by the big white dot.

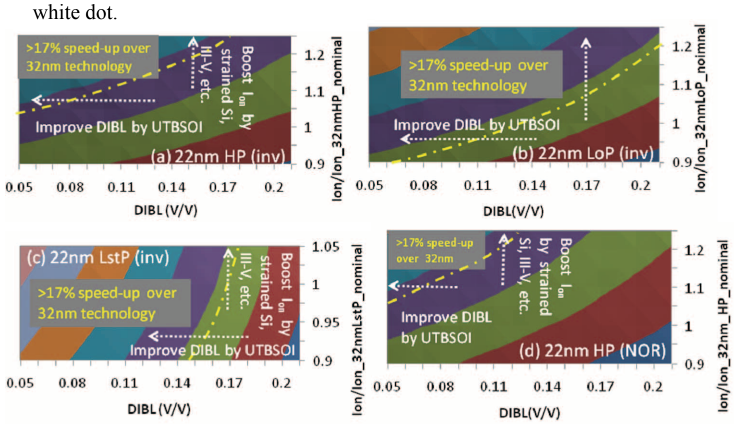

Figure 7 Inverter chain delay improvement contours on DIBL-I_{on} booster plane, with (a) HP inverter chain, (b) LoP inverter chain, (c) LstP inverter chain and (d) HP NOR chain, all at 22nm. The reference device is the nominal case at 32nm. To satisfy 17% speed-up per generation, UTBSOI structure can be introduced to improve DIBL for LoP and LstP; while current booster such as strained Si, Ge, or III-V channel materials are necessary for HP.