J. Indian Inst. Sci., July–Aug. 2005, **85**, 195–200

Scaling characteristics of f_{NQS} and f_t in NMOSFETs with and without supply voltage scaling

R. SRINIVASAN* AND NAVAKANTA BHAT
Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore 560 012, India.
e-mail: sreenivaasan@ece.iisc.ernet.in

Received on February 10, 2005; Revised on April 19, 2005.

Abstract

Extensive process and device simulations are performed to investigate the nonquasi-static transition frequency (f_{NQS}) behaviour of the NMOSFETs at different technology nodes, 0.5 μm to 90 nm, having same off-state leakage current (I_{OFF}). These studies are done with and without supply voltage scaling. f_{NQS} exhibits a turnaround in the 100 nm regime when we do the supply voltage scaling along with the transistor scaling. We attribute this effect to the reduced gate overdrive (V_{GS}–V_t) and thereby to the degraded transconductance (g_m). The unity gain frequency (f_t) also shows a similar trend. The turnaround effect of f_{NQS} and f_t disappears when I_{OFF} is allowed to go up or the gate overdrive is increased or both.

Keywords: f_{NQS}, f_t, nonquasi-static, scaling.

1. Introduction

Scaling of CMOS technology not only promises gigabit integration, gigahertz clock rate, and systems on a chip, but also arouses great expectations for CMOS RF circuits in gigahertz range [1]. Low-frequency MOSFET models are usually derived from dc analysis by using the quasi-static (QS) approach, i.e. time-dependent behaviour is treated as a succession of steady-state solutions. But, at higher frequencies, the behaviour of MOSFETs cannot be treated as a succession of steady-state solutions [2, 3]. This behaviour is called the nonquasi-static (NQS) behaviour. NQS effects impose a frequency limit on the analog RF circuits. This is not only due to decrease in amplification, but also due to phase shift effects which may become significant even at lower frequencies [5, 6]. This is true even for the high-speed switching activities in digital circuits [4]. The small-signal NQS behaviour has already been investigated in several papers [7–12].

It is believed that f_{NQS} and f_t are inversely proportional to L_{eff}^2 [13–15]. Recently, NQS behaviour of long-channel transistors (10 to 1 μm) was analysed using device simulation [13], and it has been shown that the NQS transition frequency scales as $1/L_{\text{eff}}^2$, under fixed V_{DS} and V_{GS} bias condition. Burghartz *et al.* [14] and Dambrine *et al.* [15], who show an f_t of 100 GHz around 0.1 μm channel length, anticipate that f_t would scale as $1/L_{\text{eff}}^2$, under fixed bias condition. Some work has been done at the circuit level to investigate the submicron

*Author for correspondence.

MOSFET performance (f_t, etc.) for analog applications [16, 17]. We have recently shown that the NQS behaviour is quite different in the deep sub-micron (DSM) regime [18].

In this paper, we have used process, device and mixed-mode simulation approach to investigate the f_{NQS} behaviour of the scaled submicron MOSFETs. NMOSFETs with gate lengths 0.5 mm to 90 nm have been studied with and without scaling the supply voltage. The outline of the paper is as follows. In the next section, we discuss the scaling of fNQS with the channel length. Section 3 describes the simulation setup used in this work and the devices used in the simulation. The simulation results are given in Section 4, and in Section 5, we provide the conclusions.

2. Nonquasi-static frequency and scaling

When a small time varying ac signal (v_g) is applied to a gate of a transistor which is kept in strong inversion and saturation, the inversion charge takes some time to respond to the gate signal, i.e. there is a phase difference (f) between the signal and the inversion charge response. When the signal frequency increases, f also increases. According to the accuracy requirement, f can be fixed, and the frequency corresponding to this f (5° in this paper) is known as nonquasi-static frequency (f_{NQS}). f_{NQS} is given by [13],

$$f_{\mathrm{NQS}} = \frac{a m_{\mathrm{eff}} (V_{GS} - V_t)}{2p L_{\mathrm{eff}}^2} \tag{1}$$

where a, L_{eff}, m_{eff}, V_{GS}, and V_t are respectively the fitting parameter, effective channel length, mobility, gate bias, and threshold voltage of the transistor. The fitting parameter a is technology dependent. According to (1), with scaling, f_{NQS} is directly proportional to $1/L_{\mathrm{eff}}^2$, provided the other parameters are kept constant.

But, when the supply voltage is reduced to keep the field strength within the maximum limit, for reliability consideration, then the scenario changes. This is because V_{GS} is also reduced when the supply voltage is reduced. Generally, V_{GS} is chosen to be 0.5 times of V_{DS} to get the maximum output swing. As V_t is not scaled down at the same rate as that of supply voltage due to subthreshold conduction, the gate overdrive decreases with scaling which plays a major role in determining f_{NQS}. Therefore, we can hypothesize that the scaling will not always improve the device performance, especially in the DSM regime. This hypothesis is verified in Section 4 through simulation studies.

For analog/RF applications, unity gain frequency (f_t) is an important metric, and is defined as the frequency at which the short-circuit small-signal current gain of a transistor drops to unity, indicative of the speed of the intrinsic device. f_t is given by

$$f_t = \frac{g_m}{2p C_{Gin}} \tag{2}$$

where g_m is the transconductance and C_{Gin}, the total capacitance seen at the gate terminal. Even though the C_{Gin} scales as $1/L_{\mathrm{eff}}^2$, if the value of g_m degrades significantly due to the reduced gate overdrive, then we can expect f_t to behave similar to f_{NQS}, and show a turn-around effect.

Table I
Typical gate oxide thickness and gate voltages used in the devices for realistic scaling condition

Gate length (*m*m)	Gate oxide thickness (Å)	Halo dose (10^{11}/cm^2)	V_{ds}/V_{gs} (V)
0.5	100	6	5/2.5
0.35	70	7	3.3/1.65
0.25	50	11.28	2.5/1.25
0.18	36	17.45	1.8/0.9
0.12	24	55.5	1.2/0.6
0.09	18	61.25	0.9/0.45

3. Simulation methodology

All the simulations are done using ISE-TCAD simulator. The transistors are generated using a 2D process simulator, DIOS. We have used the disposable spacer technique to control the short channel effect. The background doping of the substrate is 10^{16}/cm^3. After the initial threshold voltage (V_t) adjustment implantation, SSRC implantation was done. NMOSFETs with poly gate lengths of 0.5 *m*m, 0.35 *m*m, 0.25 *m*m, 0.18 *m*m, 0.12 *m*m and 90 nm have been studied. Appropriate gate oxide thickness is used at each gate length (Table I). The boron halo with 30° tilt has been used through a screening oxide of 10 nm. The transistors are designed such that I_{OFF} is the constraint. I_{OFF} is the drain current with $V_{DS} = V_{DD}$ and $V_{GS} = 0$. To keep the I_{OFF} constant halo dose and SSRC dose were changed.

From the output structure of the process simulator, suitable mesh for device simulation was generated. DESSIS device simulator was used to extract the various device-related parameters. The physics section of the DESSIS includes the appropriate models for band to band tunneling, quantization of inversion layer charge, doping dependency of mobility, effect of high and normal electric fields on mobility, and velocity saturation. All the simulations are done at the 2D level, i.e. the results discussed in the paper are for transistor width of 1 *m*m.

Figure 1 shows the simulation setup used in this work. A time-varying ac signal (v_g) of amplitude 0.01 V is superimposed on top of the gate dc voltage (V_{GS}). The inversion charge

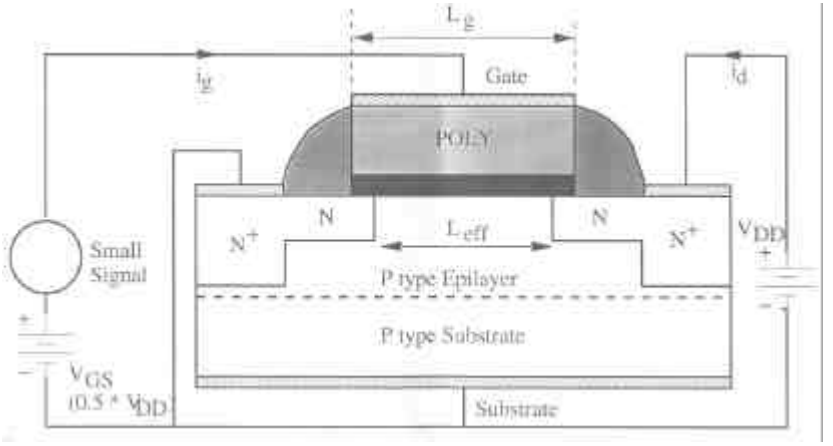

FIG. 1. Simulation setup.

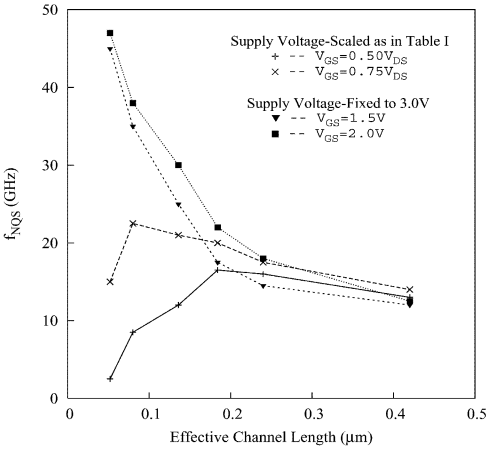

FIG. 2. f_{NQS} vs channel length for two cases, with and without supply voltage scaling, for two different V_{GS}.

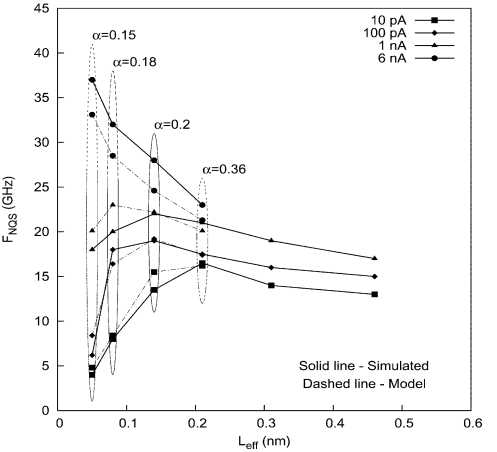

FIG. 3. f_{NQS} vs channel length for various off-state-leakage currents (I_{OFF}).

in the channel (in NMOS it is electrons) consists of two parts, one is because of the dc bias (N) and the other is due to the ac signal (n). N is constant once the bias is fixed. The time-varying quantity, n, which is the difference between the total number of electrons in the channel and the electrons due to dc bias, or equivalently the number of electrons responding to the varying gate voltage, is used to find out f_{NQS}. In the first set of experiments, NMOS devices with the I_{OFF} of 10 pA are chosen. The V_{GS} and V_{DS} are fixed to 1.5 and 3 V, respectively, for all the devices. In the second set of experiments (using the same devices), the V_{GS} and V_{DS} for different devices are set as shown in Table I based on constant electric field scaling. The same simulations are then repeated for devices with higher I_{OFF}.

4. Results and discussion

Figure 2 shows the variation of f_{NQS} vs channel length for two cases, with and without supply voltage scaling, for two different V_{GS}. When the supply voltage is not scaled down, with device scaling, f_{NQS} increases monotonically, i.e. L_{eff} is dominating in eqn (1) throughout all the gate lengths. But, when the supply voltage is scaled down, with device scaling, f_{NQS} shows a turnaround behaviour at 0.25 mm, i.e. the reduction in channel length below this does not give any improvement in terms of f_{NQS}. As we go towards the DSM regime, the gate overdrive is going down because the V_t is not scaled at the same rate as that of supply voltage to satisfy the leakage constraint, i.e. even though the reduction in L_{eff} tries to increase the f_{NQS}, the reduction in $(V_{GS}–V_t)$ starts dominating (refer eqn (1)). This is responsible for the f_{NQS} turnaround-effect.

From Fig. 2, it can be observed that increasing V_{GS} shifts the turnaround point towards the left or, in other words, it slowly disappears when V_{GS} is increased. V_{GS} increases the gate overdrive, hence the increases in f_{NQS}. The dependence of f_{NQS} on V_{GS} is stronger when the supply voltage is scaled with device scaling.

Next we shall see the effect of I_{OFF} on the turnaround behaviour. Figure 3 shows the plot of f_{NQS} vs channel length, for different I_{OFF}. As V_t trades off with the I_{OFF}, gate overdrive in-

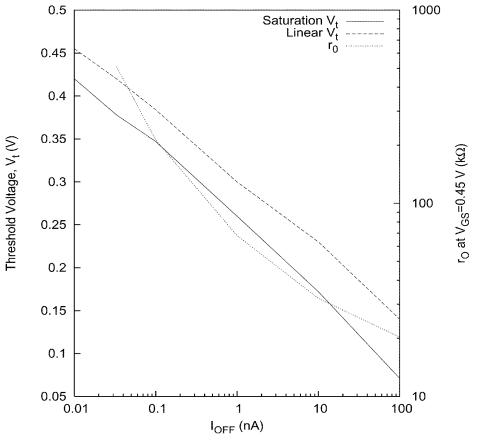

FIG. 4. Linear and saturation V_t vs I_{OFF} for 90 nm device.

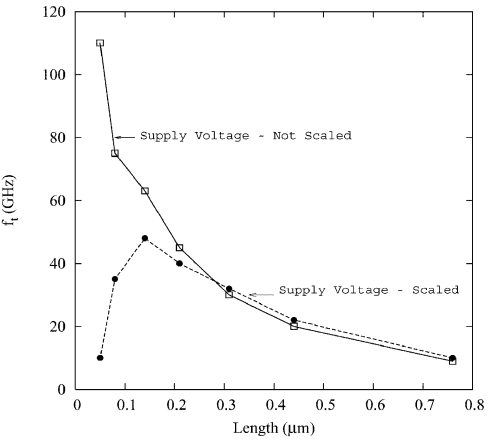

FIG. 5. Plot of f_t vs channel length for two cases, with and without supply voltage scaling.

creases. When the I_{OFF} constraint is relaxed by reducing the doping, then the mobility may also improve, which is an added advantage. It can be noticed from Fig. 3 that the turn-around effect slowly disappears if the I_{OFF} constraint is relaxed. However, it is interesting to note that the improvement in f_{NQS} comes at the expense of degradation in DIBL and output resistance. Figure 4 shows the linear and saturation V_t vs I_{OFF} plots of the 90 nm device. It can be noticed from Fig. 4 that the difference between linear and saturation V_t, i.e. the DIBL effect is slightly higher at higher leakages. The figure also shows the output resistance (r_O) vs I_{OFF}, at $V_{DS} = V_{GS} = 0.45$ V. We observe that when I_{OFF} is varied from 100 pA to 100 nA, r_O reduces from 212 to 20 kΩ.

We have also attempted to predict the f_{NQS} behaviour analytically. Equation (1) is modified to account for velocity saturation for $L_g < 0.25$ μm as below,

$$f_{NQS} = \frac{a m_{\text{eff}} (V_{GS} - V_t)}{2p L_{\text{eff}}^m}. \tag{3}$$

In eqn (3), m takes care of the velocity saturation effect. We have used $m = 1.867$ for all the devices. The other parameters, V_t and m_{eff} are extracted from the device behaviour. a is technology-dependent fitting parameter. Figure 3 shows the f_{NQS} values predicted analytically (dashed line), at four different gate lengths, for all the leakage currents. The value of a used is also indicated. We notice that there is a reasonably good agreement between the analytical values and the mixed mode simulation values.

Figure 5 shows the behaviour of f_t with L_{eff}, for the devices having $I_{OFF} = 10$ pA. f_t also shows a similar trend like f_{NQS}, i.e. increases monotonically with decreasing L_{eff}, for the case where the gate and drain bias voltages are 1.5 and 3 V, respectively, for all the devices. But, when the supply voltages are scaled down as in Table I, along with the device scaling, f_t increases till 0.18 μm and then decreases. As per eqn (2), f_t is directly proportional to g_m which is once again related to gate overdrive and the reasoning can be done as in the case of f_{NQS} behaviour.

5. Conclusion

On one hand, scaling increases the f_{NQS} monotonically as $1/L_{\text{eff}}^2$, provided the gate and drain voltages are fixed. On the other hand, because of the supply voltage scaling, f_{NQS} turns around at 100 nm regime. The f_t also shows similar behaviour. The performance degradation after scaling is due to reduced gate overdrive. The turnaround effect of f_{NQS} and f_t was not seen when I_{OFF} is allowed to go up or the gate overdrive is increased.

References

1. Qiuting Huang, Francesco Piazza and Tatsuya Ohgura, The impact of scaling down to deep submicron on CMOS RF circuits, *IEEE J. Solid St. Circuits*, **33**, 1023–1036 (1998).
2. Y. P. Tsividis and G. Masetti, Problems in the precision modeling of MOS transistor for analog applications, *IEEE Trans. Computer Aided Des.*, **3**, 72–79 (1984).
3. J. J. Paulous and D. A. Antoniads, Limitations of quasistatic capacitance models for the MOS transistors, *IEEE Electron Device Lett.*, **4**, 221–224 (1983).
4. Yuhua Cheng *et al.*, *BSIM3v3 manual*, Department of Electrical Engineering and Computer Sciences, University of California, Berkeley (1996).
5. T. Smeds and F. M. Klaasen, An analytical model for the nonquasi-static small signal behaviour of submicron MOSFETS, *Solid St. Electronics*, **38**, 121–130 (1995).
6. Yannis Tsividis, *Operation and modelling of the MOS transistor*, McGraw-Hill (1999).
7. M. Bagheri and Y. Tsividis, A small signal dc-to-high-frequency non-quasistaticmodel for the four-terminal MOSFET valid in all regions of operation, *IEEE Trans. Electron Devices*, **32**, 2383–2391 (1985).
8. P. J. V. Vandeloo and W. M. C. Sansen, Modeling of the MOS transistor for high frequency analog design, *IEEE Trans. Computer Aided Des.*, **8**, 713–723 (1989).
9. H-J. Park, P. K. Ko and C. Hu, A charge conserving non-quasi-static (NQS) MOSFET model for SPICE transient analysis, *IEEE Trans. Computer Aided Design*, **10**, 629–642 (1991).
10. H.-J. Park, P. K. Ko and C. Hu, A non-quasi-static MOSFET model for SPICE-AC analysis, *IEEE Trans. Computer Aided Des.*, **11**, 1247–1257 (1992).
11. J. M. Sallese and A.-S. Porret, A novel approach to charge-based non-quasi-static model of the MOS transistor valid in all modes of operation, *Solid St. Electronics*, **44**, 887–894 (2000).
12. Alain-Serge Porret, Jean-Michel Sallese and Christian C. Enz, A compact non-quasi-static extension of a charge-based MOS model, *IEEE Trans. Electron Devices*, **48**, 1647–1654 (2001).
13. Allen F. L. Ng, Ping K. Ko and Mansun Chan, Determining the onset frequency of non-quasi-static effects of the MOSFET in AC simulation, *IEEE Electron Device Lett.*, **23**, 37–39 (2002).
14. J. N. Burghartz *et al.*, RF potential of a 0.18 *m*m CMOS logic technology, *IEDM Tech. Dig.*, pp. 35.3.1–35.3.4 (1999).
15. G. Dambrine *et al.*, What are the limiting parameters of deep submicron MOSFETs for high frequency applications?, *IEEE Electron Device Lett.*, **24**, 189–191 (2003).
16. M. J. M. Pelgrom and M. Vertegt, CMOS technology for mixed signal ICs, *Solid St. Electronics*, **41**, 967–974 (1997).
17. Claudio Fiegna, The effects of scaling on the performance of small signal MOS amplifiers, *Solid St. Electronics*, **46**, 675–683 (2002).
18. R. Srinivasan and Navakanta Bhat, Effect of scaling on the non-quasi-static behaviour of the MOSFET for RF ICs, *Int. Conf. on VLSI Design*, New Delhi (2003).