

Optimization of V_{DD} and V_{TH} for Low-Power and High-Speed Applications

Koichi Nose and Takayasu Sakurai

Institute of Industrial Science, University of Tokyo
7-22-1 Roppongi, Minato-ku, Tokyo, 106-8558 Japan

Tel: +81-3-3403-1643

Fax: +81-3-3403-1649

e-mail: {nose, tsakurai}@iis.u-tokyo.ac.jp

Abstract - Closed-form formulas are presented for optimum supply voltage (V_{DD}) and threshold voltage (V_{TH}) that minimize power dissipation when technology parameters and required speed are given. The formulas take into account short-channel effects and the variation of V_{TH} and temperature. Using typical device parameters, it is shown that a simple guideline to optimize the power consumption is to set the ratio of maximum leakage power to total power about 30%. Extending the analysis, the future VLSI design trend is discussed. The optimum V_{DD} coincides with the SIA roadmap and the optimum V_{TH} for logic blocks at the highest temperature and at the lowest process variation corner is in the range of 0V~0.1V over generations.

I. Introduction

Decreasing power consumption of VLSI's is getting one of the key design issues. Lowering the supply voltage (V_{DD}) is the most effective to decrease the power consumption, since CMOS power quadratically depends on V_{DD} . Low V_{DD} , however, degrades the performance of circuits. It is possible to maintain the performance by decreasing the threshold voltage (V_{TH}) at the same time, but then the subthreshold leakage power increases exponentially. Therefore, there are optimum V_{DD} and V_{TH} that achieve the required performance and the lowest power.

In this context, V_{DD} - V_{TH} optimization has been investigated extensively but previous publications on V_{DD} - V_{TH} optimization have following three problems.

First, Energy-Delay product (ED product) has been often used as an object function in optimizing CMOS circuit power consumption [1]-[3]. In practice, however, the objective of the optimization is to minimize the power consumption while satisfying a speed constraint. When we take the ED product as an object function, we get only one pair of the optimized V_{DD} and optimized V_{TH} if the technology is fixed. This is not what we want, since the optimized V_{DD} and V_{TH} should be different if the target circuit speed is different. In this paper, the optimization is carried out taking the power as an object function and the speed as a constraint to make the optimization results more practical.

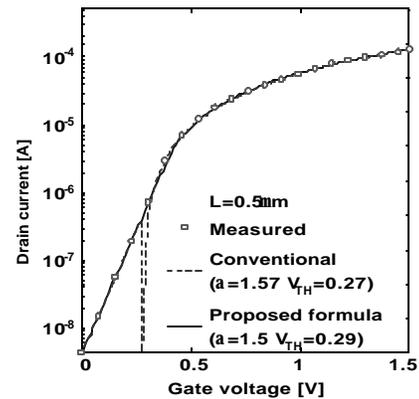


Fig.1 Drain current models used in power optimization.

The second issue is on the drain current modeling of MOSFET's. Figure 1 shows a comparison between the present model and the previous model that has been used in power optimization papers [1][2]. It is seen that the previous drain current model has discontinuity around the V_{TH} while the present model rectifies the issue, details of which is discussed in the text.

The last problem is that the previous calculation has not considered the effects of both V_{TH} fluctuation and temperature variation. Since these effects are getting more important in the deep submicron region, the analysis should take these effects into account.

In this paper, closed-form formulas are presented for optimum V_{DD} and V_{TH} that minimize power dissipation when the technology and required speed are given. Above-mentioned problems are eliminated in the analysis. $V_{TH,min}$ is considered in this paper to incorporate V_{TH} fluctuation effects. The resultant formulas have been applied to the technology roadmap to discuss the future VLSI design trend.

II. Closed-form formulas for optimum V_{DD} and V_{TH}

A new drain current model for short-channel MOSFET's is proposed that provides smooth transition across subthreshold region and above-threshold region. By using the model, accurate calculation of power and delay near the threshold is possible. The model is described as the following expressions.

$$I_D = \begin{cases} I_0 e^{\alpha \left(\frac{V_{GS} - V_{TH}}{\alpha N_S} \right)} & (V_{GS} \geq V_{TH} + \alpha N_S) \\ I_0 e^{-\frac{V_{GS} - V_{TH}}{N_S}} & (V_{GS} \leq V_{TH} + \alpha N_S) \end{cases} \quad (1)$$

The notations for these formula as well as the notations for other quantities used in this paper are tabulated in Table I.

Figure 1 shows a comparison between the proposed model and the conventional model [1]. The previous drain current model has discontinuity around V_{TH} and the present model does not have one. The difference between the proposed formula and the measured result is within 4% when $V_{GS}=0-1.5V$.

Here, as a basis of optimization, the delay and the power dissipation models are explained that take into consideration the V_{TH} variation through process and temperature. The two main sources of power dissipation in CMOS VLSI's are the dynamic power dissipation due to charging and discharging of load capacitance, and the power dissipation due to subthreshold leakage. There may be short-circuit power dissipation as the third source of power dissipation but it is less than 10% in total power dissipation [5] and has been neglected in this study.

The main device parameters that depend on the temperature are mobility, μ , V_{TH} , and subthreshold slope, N_S . The temperature dependence of these parameters are written as [7]

$$\mu = \mu' \cdot \left(\frac{T_{max}}{T_{min}} \right)^{-m} \quad (2)$$

$$V_{TH, min} = V_{TH, max} - \Delta V_{TH} - \kappa \Delta T \quad (3)$$

$$N_S = N'_S \cdot \frac{T_{max}}{T_{min}} = \frac{nkT_{max}}{q} \quad (4)$$

where μ' and N'_S are the mobility and the subthreshold slope at the lowest temperature in use, T_{min} , respectively. $V_{TH, max}$ and $V_{TH, min}$ are the maximum and minimum threshold voltage

Table I Notations used in this paper

notation	meaning
a	switching activity
L_d	logic depth of critical path
f	given clock frequency
C_L	load capacitance
α	velocity saturation index [6]
I_0	drain current when $V_{GS}=V_{TH}$ at lowest temperature
T_{min}	lowest operation temperature
T_{max}	highest operation temperature
ΔT	$T_{max}-T_{min}$
N_S	nkT_{max}/q (n: subthreshold slope factor)
K	coefficient of delay
ΔV_{TH}	peak-to-peak V_{TH} variation through process
κ	temperature coefficient of V_{TH}
$V_{TH, max}$	highest V_{TH} in operation temp. and process variation range
$V_{TH, min}$	lowest V_{TH} in operation temp. and process variation range
$V_{DD, opt}$	optimum V_{DD}
$V_{TH, opt}$	optimum $V_{TH, min}$
$I_{ON, min}$	drain current when $V_{GS}=V_{DD}$ at lowest temp. and highest V_{TH} corner in process variation
$I_{OFF, max}$	leakage current at highest temp. and lowest V_{TH} corner in process variation
$P_{LEAK, max}$	leakage power at highest temp. and lowest V_{TH} corner in process variation

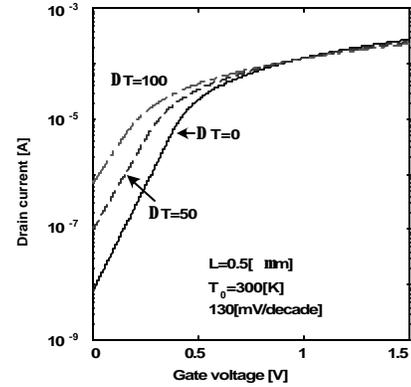


Fig.2 Temperature characteristics of MOSFET.

under the temperature and process fluctuation. κ is a temperature coefficient of V_{TH} , which is typically 2.4mV/K in 0.5 μ m process, and m is a temperature exponent of mobility whose typical value is 1.5. Figure 2 shows the temperature dependence of drain current. It is seen that, in sub-1V region, CMOS circuits show positive temperature dependence, because the effect caused by V_{TH} lowering is stronger than the effect caused mobility degradation [8][9]. If our interest is in sub-1V region, the worst-case delay occurs at the lowest operation temperature. The delay of interest is written as

$$t_d = K \frac{C_L V_{DD}}{\beta (V_{DD} - V_{TH, max})^\alpha} \quad (5)$$

where $\beta = I_0 \left(\frac{e}{\alpha N_S} \right)^\alpha$. (6)

On the other hand, the worst power consumption is observed at the highest operation temperature, because the dynamic power component, P_D , which is written as

$$P_D = afC_L V_{DD}^2, \quad (7)$$

does not have temperature dependence and the main temperature dependence comes from the leakage component. The leakage component also increases when V_{TH} is lowered by V_{TH} fluctuation. Therefore, the maximum leakage current appears when the threshold voltage is $V_{TH,min}$. Consequently the maximum leakage power, $P_{LEAK,max}$ is written as

$$P_{LEAK,max} = I_0 e^{\frac{-V_{TH,min}}{N_S}} V_{DD}. \quad (8)$$

The frequency is expressed using t_d (Eq.5) and the logic depth of a critical path, L_d .

$$f = \frac{1}{L_d \cdot t_d}. \quad (9)$$

Equations 7, 8 and 9 are the basic equations for the power optimization. Now we try to solve the equation system. First, by solving Eq.9 in terms of $V_{TH,min}$, we get

$$\begin{aligned} V_{TH,min} &= V_{DD} - \left(\frac{fL_d K C_L}{\beta} \right)^{1/\alpha} V_{DD}^{1/\alpha} - \Delta V_{TH} - \kappa \Delta T \\ &= V_{DD} - \chi V_{DD}^{1/\alpha} - \Delta V_{TH} - \kappa \Delta T, \end{aligned} \quad (10)$$

where $\chi = (fL_d K C_L / \beta)^{1/\alpha}$.

Substituting Eq.10 in Eq.7 and Eq.8 the formula of power dissipation can be derived, which is denoted as $P(V_{DD})$. In order to obtain $V_{DD,opt}$ and $V_{TH,opt}$ when the clock frequency is given, we differentiate $P(V_{DD})$ with respect to V_{DD} and set the resultant expression to zero. The resulting equation is transcendental and cannot be solved exactly. Here we can assume $V_{DD} \gg N_S$, since N_S is normally less than 0.05V. Then, the equation becomes as follows.

$$V_{DD} - \chi V_{DD}^{1/\alpha} = -N_S \ln \left(\frac{2afC_L N_S}{I_0} \frac{\alpha}{\alpha - \chi} \right) + \Delta V_{TH} + \kappa \Delta T \quad (11)$$

Still the above equation cannot be solved for V_{DD}

analytically, but optimum $V_{TH,min}$, which is denoted as $V_{TH,opt}$, can be calculated using Eq. 10 and Eq.11 easily.

$$V_{TH,opt} = -N_S \ln \left(\frac{2afC_L N_S}{I_0} \frac{\alpha}{\alpha - \chi} \right) \quad (\alpha > \chi), \quad (12)$$

where

$$\chi = \left(\frac{fL_d K C_L}{\beta} \right)^{1/\alpha}, \quad N_S = \frac{nkT_{max}}{q}. \quad (13)$$

As is described above, it is difficult to solve $V_{DD,opt}$. Some approximations are used. By using Taylor expansion of the equation around $V_{DD}=1$, $V_{DD,opt}$ can be solved as

$$\begin{aligned} V_{DD,opt} &= \frac{-N_S \ln \left(\frac{2afC_L N_S}{I_0} \frac{\alpha}{\alpha - \chi} \right) + \Delta V_{TH} + \kappa \Delta T + \frac{\alpha - 1}{\alpha} \chi}{1 - \frac{\chi}{\alpha}} \\ &(\alpha > \chi). \end{aligned} \quad (14)$$

Equations 12 and 14 are the optimum V_{DD} and V_{TH} .

Let us make the simpler guideline for the power optimization. This is possible by using either the ratio between $P_{LEAK,max}$ and P_D or the ratio between $I_{ON,min}$ and $I_{OFF,max}$. $P_{LEAK,max}$, $I_{ON,min}$ and $I_{OFF,max}$ are defined in Table I. Using Eq.7 and Eq.8, the ratio of $P_{LEAK,max}/P_D$ can be expressed as

$$\frac{P_{LEAK,max}}{P_D} = \frac{2}{\frac{V_{DD,opt}}{N_S} - 1 - \frac{V_{DD,opt} - V_{TH,opt}}{\alpha N_S}}, \quad (15)$$

where $P_{LEAK,max}$ is leakage power dissipation at the highest temperature and at the lowest V_{TH} corner in process variation. If we confine V_{DD} around 1V (0.5V~1.5V) and $V_{TH,opt} \ll 1$, the ratio can be simplified as

$$\frac{P_{LEAK,max}}{P_D} = \frac{2N_S \alpha}{\alpha - 1} \quad (\alpha > 1.1). \quad (16)$$

In terms of $I_{ON,min}$ and $I_{OFF,max}$, it is rewritten as

$$\frac{I_{ON,min}}{I_{OFF,max}} = K \frac{\alpha - 1}{2N_S \alpha} \frac{L_d}{a} \quad (\alpha > 1.1). \quad (17)$$

Assuming typical values for the parameters such that $N_S=0.048$ (S-factor=80mV/decade and $T_{max}=400K$) and $\alpha=1.3$, $P_{LEAK,max}$ is calculated to be about 30% of the total power dissipation. This value of about 30% is not changed over a wide range of design parameters such as a , L_d and f .

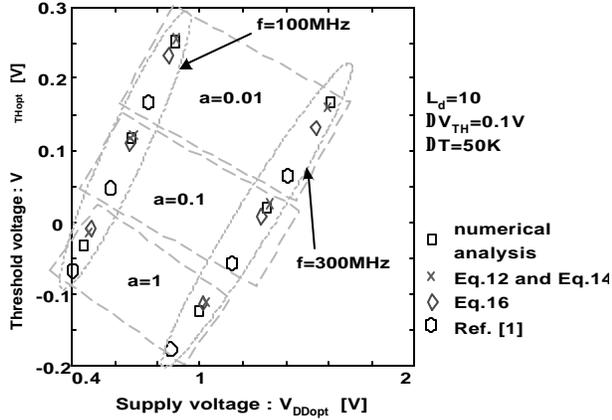


Fig. 3 V_{DDopt} and V_{THopt} comparison among the proposed analysis formula, the proposed simple expression and the expression in [1].

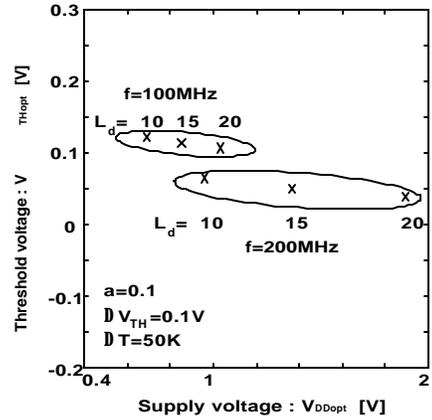


Fig. 5 V_{DDopt} and V_{THopt} dependence on logic depth, L_d .

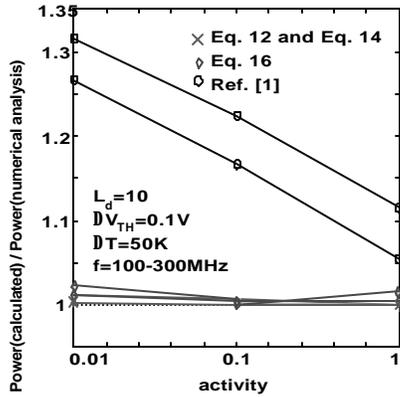


Fig. 4 Comparison of power consumption among calculations using Eq. 12 and 14, Eq. 16 and previously published expression in [1].

This is understood like below. When the target speed is changed, V_{THopt} changes slightly but V_{DDopt} changes much because V_{TH} changes the power exponentially while the dependence of power on V_{DD} is quadric. The amount of change in V_{TH} and V_{DD} cancels out the dependency of power on these parameters.

III. Comparison with numerical solutions

In order to confirm the validity of the V_{DDopt} and V_{THopt} formulas of Eq.12 and 14 and the simple expression of Eq.16, the proposed formulas are compared with the results of numerical solutions by Eqs. 7, 8 and 9, and the conventional formula in [1] where it is stated that the ED product is minimized when $P_{LEAK,max}/P_D=1$.

Figure 3 shows the result. In this analysis, the activity, a , is varied from 1, 0.1, to 0.01 and the logic depth, L_d is set to 10,

which is typical. ΔV_{TH} is set to 0.1V and ΔT is set to 50K. It is seen from the figure that the discrepancy in V_{THopt} between the numerical solution and the conventional calculation [1] is 0.11V, while the discrepancy is suppressed to 0.03V for the proposed formula calculation.

Figure 4 shows the accuracy of the proposed formulas together with the formula in the previous publication [1]. The calculated values are compared with the results of direct numerical analysis using Eqs.7, 8 and 9. It is seen that the proposed formulas are in good accordance with the numerical solutions and above-mentioned approximations are found to be reasonable.

IV. Discussions

It is clearly seen from Fig.3 that V_{THopt} decreases only 0.1V when the required frequency changes from 100MHz to 300MHz. On the other hand, V_{THopt} increases 0.3V when activity, a , changes from 1 to 0.01. Figure 5 shows the V_{DDopt} and V_{THopt} dependence on the logic depth, L_d . In this figure, the variation of V_{THopt} when L_d is changed from 10 to 20 is only 0.03V. From these results, it can be said that V_{THopt} is not a strong function of either the clock frequency or the logic depth but strongly depends on the activity. Therefore, it is effective to decide V_{TH} according to the activity of macro blocks (ex. high V_{TH} for memory blocks, low V_{TH} for logic blocks and further lower V_{TH} for clock circuits). The power increases exponentially when V_{TH} decreases. Hence, to improve the speed, V_{DD} tends to increase and V_{TH} tends to stay the same. This is the reason why V_{THopt} is not a strong function of speed related constraints.

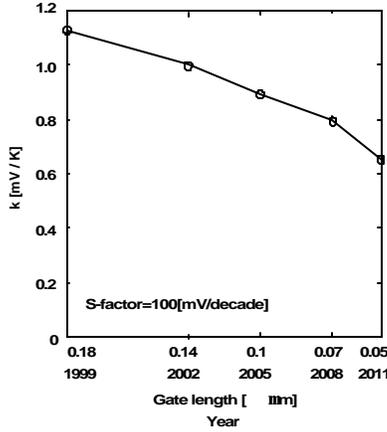


Fig.6 Change in κ on generations

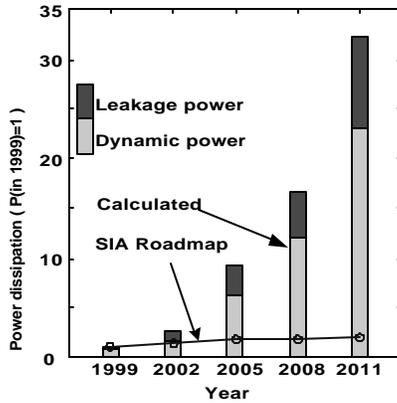


Fig.7 Power consumption trend by the estimation through proposed formulas and that by SIA roadmap

V. Future trend of optimum V_{TH} and logic depth

A future trend in V_{DD} and power dissipation has been shown in the SIA Roadmap [4]. V_{TH} and the logic depth, however, are not discussed in the roadmap. In this section, the trend of the optimum V_{TH} , the logic depth, and the number of transistors in logic blocks is discussed for the first time using the parameter values given in the SIA roadmap.

When a certain device parameter is given in the SIA roadmap, it is used in the analysis. For parameters that are not listed in the roadmap, reasonable assumptions are made as follows. α , K and N_S , are assumed to be constant in all generations, being equal to 1.3, 0.78, and 0.05, respectively. T_{min} and T_{max} are set equal to 300K and 400K, respectively. The activity, a , is set to 0.1 for logic blocks [11].

κ is a function of impurity density and can be estimated using the formula in [10]. Figure 6 shows the change of κ on generations. In 0.18 μ m technology, V_{TH} increases about 0.11V when the temperature goes up by 100K, but when the

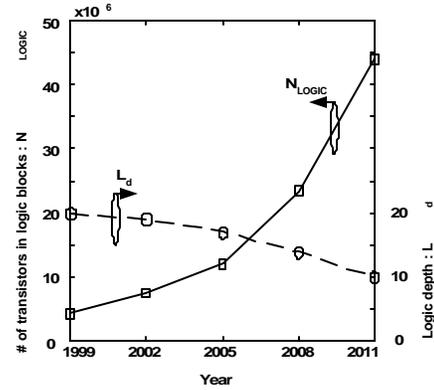


Fig.8 Trend of N_{LOGIC} and logic depth, L_d

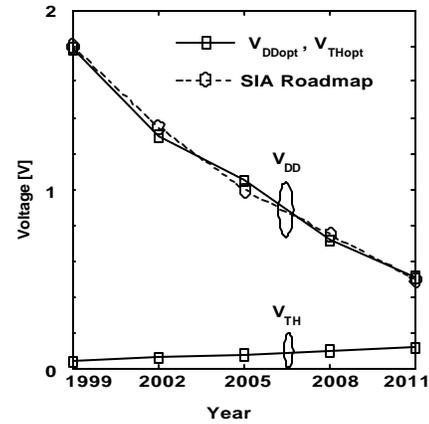


Fig. 9 Prediction of V_{DDopt} and V_{THopt}

feature size becomes 0.05 μ m in 2011, the V_{TH} change will be less than 0.07V.

The total number of transistors on a chip, N_{CHIP} , consists of the number of transistors in logic blocks and that in memory blocks. N_{CHIP} in 2011 is predicted to be about 70 times as large as that in 1999. The power dissipation in memory blocks can be neglected when leakage cutoff techniques are used (for example, see dynamic leakage cut-off scheme [12]). Therefore, the number of transistors in logic blocks, N_{LOGIC} , is of importance in calculating the power consumption. At present, the ratio of N_{LOGIC} to N_{CHIP} is about 20%. For a moment, let us suppose the ratio is invariant over time. L_d is also set constant at 20.

Figure 7 shows the power consumption trend by the estimation through proposed formulas and that by the SIA roadmap. In the calculation, the power will increase by a factor of 30. On the other hand, the SIA roadmap tells that the total power in 2011 should be within 2 times the power in 1999. It is clear that the target in the SIA roadmap cannot be

achieved without some modifications in the scaling scenario. The main parameters, which can be modified in the design level, are the logic depth and the ratio of $N_{\text{LOGIC}}/N_{\text{CHIP}}$.

Three scenarios are considered here. In the first scenario, $N_{\text{LOGIC}}/N_{\text{CHIP}}$ remains constant at 20%, while the logic depth can be changed freely. The logic depth is a function of architecture, a pipeline scheme and a design style. There are no official values for the L_d change in time. The estimated logic depth in 2008 becomes 1. Although there is a tendency that the logic depth is being decreased, this is totally unrealistic.

In the second scenario, L_d is kept constant at 20 and $N_{\text{LOGIC}}/N_{\text{CHIP}}$ are changed freely. Then, N_{LOGIC} in 2011 will be 1.1 times of N_{LOGIC} in 1999. This scenario again is unrealistic, since it basically says that the number of transistors for logic blocks should not be increased.

Now, in the third scenario, more realistic values for L_d and N_{LOGIC} are searched for. In this scenario, the minimum achievable L_d is set equal to 10, a half of the current typical value and then N_{LOGIC} in 2011 can be calculated and fixed. From 1999 through 2011, N_{LOGIC} are interpolated assuming an exponential change in time. The resultant figure is shown in Fig. 8. This can be one possible scenario. The point is that memories can be using more transistors while logic part cannot be. Figure 9 shows the trend in V_{DDopt} and V_{THopt} in this scenario. V_{THopt} for the logic part is 0.05V in 1999 and 0.12V in 2011. Note that this V_{THopt} is the lowest V_{TH} in the operation temperature and at the lowest process variation corner. The conclusion that the optimum $V_{\text{TH,min}}$ is in the range of 0V~0.1V over generations is basically unchanged even if activity increases up to 0.3 from 0.1.

V_{DDopt} coincides with the SIA roadmap. There are many ideas presented to reduce stand-by power but up to now there are eventually no successful proposals on reducing the active power except for changing the supply voltage. In this circumstance, this third scenario is a compromised approach.

V. Conclusion

Closed-form formulas for optimum V_{DD} and V_{TH} are presented for low power and high-speed LSI's. These formulas take into account the variation of threshold voltage and temperature. From the calculation using these formulas, it is shown that V_{THopt} is not a strong function of either the clock frequency or the logic depth but strongly depends on the activity.

It is shown that a simple guideline for power optimization is to set the ratio of the maximum leakage power to the total

power around 30%. Note that the maximum leakage power is observed at the highest temperature and at the lowest V_{TH} corner in process variation.

The trend in V_{THopt} and V_{DDopt} is calculated using the device parameters given in the SIA roadmap. The V_{DDopt} coincides with the SIA roadmap and V_{THopt} , that is, the optimum $V_{\text{TH,min}}$ is in the range of 0V~0.1V over generations.

Acknowledgement

Useful discussions with K.Sasaki, K.Ishibashi, M.Miyazaki and H.Mizuno are acknowledged. The work was supported by a grant from Hitachi, Ltd.

References

- [1] J. Burr and A. Peterson, "Ultra low power CMOS technology," *NASA VLSI Design Symposium*, pp. 4.2.14.2.13, 1991.
- [2] R. Gonzalez, B. M. Gordon and M. A. Horowitz, "Supply and threshold voltage scaling for low power CMOS," *IEEE Journal of Solid-State Circuit*, vol. 32, pp. 1210-1216, Aug., 1997.
- [3] Z. Chen, C. Diaz, J. D. Plummer, M.Cao and W. Greene, "0.18um dual Vt MOSFET processing and energy-delay measurement," *IEDM tech. digest*, pp. 851-854, 1996.
- [4] The National Technology Roadmap for Semiconductors, SIA Handbook, 1998.
- [5] K. Nose and T. Sakurai, "Closed-Form Expressions for Short-Circuit Power of Short-Channel CMOS Gates and Its Scaling Characteristics," *Proceedings of ITC-CSCC*, pp.1741-1744, July, 1998.
- [6] T. Sakurai and A. R. Newton, "Alpha-power law MOSFET model and its application to CMOS inverter delay and other formulas," *IEEE Journal of Solid-State Circuits*, vol.25, pp. 584-593, Apr., 1990.
- [7] A. Bellaouar, A. Fridi, M. I. Elmasry and K. Itoh, "Supply voltage scaling for temperature insensitive CMOS circuit operation," *IEEE Transaction on Circuit and Systems II*, vol. 45, pp. 415-417, Mar., 1998.
- [8] C. Park et al, "Reversal of temperature dependence of integrated circuits operation at very low voltages," *IEDM Tech., Digest*, pp. 71-74, 1995.
- [9] K. Kanda, K. Nose, H. Kawaguchi and T. Sakurai, "Design Impact of Positive Temperature Dependence of Drain Current in Sub 1V CMOS VLSI's," *Proceedings CICC'99*, pp.563-566, May, 1999.
- [10] Yuan Taur and Tak H. Ning, "Fundamental of Modern VLSI Devices," by Cambridge University Press, pp. 131, 1998.
- [11] J. Burr and J. Shott, "A 200mV encoder-decoder circuit using Stanford Ultra Low-Power CMOS," *ISSCC Digest of Tech. Papers*, pp. 84-85, Feb., 1994.
- [12] H.Kawaguchi, Y.Itaka and T.Sakurai, "Dynamic Leakage Cut-off Scheme for Low-Voltage SRAM's," *Symp. on VLSI Circuits*, pp.140-141, June, 1998.