Threshold Variations for Undoped Double-Gate MOSFET’s

Qiang Chen and James D. Meindl
Microelectronics Research Centre, Georgia Institute of Technology,
791 Atlantic Dr. N.W., Atlanta, Georgia 30332, the U.S.A.
E-mail: qchen@ee.gatech.edu

Abstract

A quantitative scaling theory of undoped double-gate MOSFET’s is developed based on unified dependencies of subthreshold swing (S) and threshold voltage roll-off (∆VTH) on the ratio of the channel length (L) to the scale length (λ1). Applying this theory, compact, analytical models of threshold variations are established. It is found that threshold variations are determined completely by three unique, well-characterised dependencies on the ratio of L/λ1, contributed by L, the silicon film thickness (tSi), and the oxide thickness (tOX). For the same process tolerance, tOX causes the least, relatively insignificant amount of threshold variations, while L causes about 30-50% more threshold variations than tSi does. For example, at L/λ1=6, which corresponds to S=66 mV/dec. and ∆VTH≈60 mV, 10% variations of L, tSi, and tOX lead to about 18, 14, and 5 mV threshold variations, respectively. Threshold variations would deteriorate very much if tSi is to be determined by lithography.

1. Introduction

The undoped (or lightly doped) double-gate (DG) MOSFET has been proposed as the most promising device structure that enables further CMOS scaling toward the end of the International Technology Roadmap for Semiconductors (ITRS) [1]. Although an undoped or lightly doped DG MOSFET does not suffer threshold variation due to random microscopic fluctuations of dopant atoms [2], severeness of threshold variations due to process tolerance is unclear. Numerical simulations carried out to address this concern [3] only dealt with specific sets of device parameters without giving a comprehensive picture of the general trend. This paper studies systematically and comprehensively the threshold variations in undoped symmetric DG MOSFET’s due to process tolerance of key device parameters, including the channel length (L), the silicon film thickness (tSi), and the gate oxide thickness (tOX). To this end, a quantitative scaling theory of undoped DG MOSFET’s is first proposed in Section 2 based on newly developed compact, analytical short-channel models of subthreshold swing (S) and threshold voltage (VTH). Applying this scaling theory, compact, analytical models of threshold variations are obtained and analysed in detail in Section 3. Section 4 concludes the paper with some key remarks.

2. Quantitative scaling theory

![Figure 1. S model (tOX=1.5nm) (Discrete points are Medici numerical simulations) [4].](image)

![Figure 2. A unified S vs. L/λ1 dependence, obtained from Figure 1. The inset shows the factor Γ1cos(tSi/4λ1) in (1) as a function of r.](image)

A compact, analytical short-channel S model of undoped symmetric DG MOSFET’s has been developed by solving the 2-D Poisson equation \( \nabla^2 \Phi = qN_d / \varepsilon_0 \) in the channel region using an evanescent-mode analysis [4]:
\[ S = \left( 1 - 2 \Gamma, \cos \frac{t_{o} B}{\lambda_{1}} e^{-\frac{t_{o} B}{\lambda_{1}}} \right)^{-1} kT \ln 10, \]  

(1)

The scale length \( \lambda_{1} \) is the lowest-order solution of \( \tan \gamma / 2 \lambda_{1} = r \lambda_{1} / S_{0} \), where \( r = e_{t} / e_{i} \). For estimation purposes, \( \lambda_{1} \) can be well approximated by

\[ \lambda_{1} = \frac{1 + 1 / r}{1 + \pi / 2}, \]  

(a)

\[ \lambda_{1} = \frac{1 + \sqrt{2} / r}{\sqrt{2} + \pi / 2}, \]  

(b)

for \( r \leq \pi / 2 \) and \( r > \pi / 2 \), respectively, over a practical range of \( r \)-values from 0.8 to 20. The parameter \( \Gamma_{1} \) is given as

\[ \Gamma_{1} = 2 \lambda_{1} \left( 1 + r_{s}^{2} \frac{1}{\lambda_{1}^{2}} \frac{1}{r_{s}^{2} / 2} \right). \]  

(3)

The \( S \) model (1) has been compared to Medici numerical simulations with good agreement [4], as shown in Figure 1. A more careful examination of (1) reveals that the factor, \( \Gamma_{1} \cos (t_{o} / 4 \lambda_{1}) \), varies insignificantly around unity, as shown by the inset of Figure 2. Thus, \( S \) is primarily determined by the ratio of the channel length to the scale length, \( L / \lambda_{1} \). Following this observation, the two \( S-L \) curves in Figure 1, if redrawn dividing the abscissa by their corresponding values of \( \lambda_{1} \), consolidate almost perfectly into a unique \( S-L / \lambda_{1} \) curve, as demonstrated in Figure 2.

In undoped DG MOSFET’s, a reasonable \( V_{TH} \) model must account for the mobile charges due to low doping concentration and volume conduction [5]. Such a \( V_{TH} \) model has been developed by solving the 2-D Poisson equation with the mobile charge term included \( \nabla^{2} \Psi = q n / e_{so} \) [6] as

\[ V_{th} = \Phi_{MS,i} + \eta \frac{kT}{q} \frac{\cosh(\theta)}{\cosh(\theta / 2)} \left( \frac{Q_{MS}}{e_{so}} \right) \left[ \frac{\cosh(\theta)}{\cosh(\theta / 2)} \eta^{-1} \right] \Phi_{MS}, \]  

(4)

where \( \theta = B_{t} / L_{\lambda}, \Phi_{MS,i} \) is the workfunction difference between the gate and intrinsic silicon, and \( Q_{MS} \) is a constant density of free electrons that is pre-selected to define \( V_{TH} \). Parameters \( \eta, \Phi_{MS}, \) and \( B \) are given in compact, explicit expressions [6]. The threshold voltage roll-off (\( \Delta V_{TH} \)) is readily derived from (4) as

\[ \Delta V_{TH} = \frac{kT}{q} \ln \left( \frac{Q_{MS}}{e_{so} n_{so}} \right) \Phi_{MS} \left[ \frac{\cosh(\theta)}{\cosh(\theta / 2)} \right]^{-1}, \]  

(5)

and compared to FIELDAY numerical simulations of [7] with good agreement [6] (Figure 3). It is worthwhile pointing out that the models (4)-(5) have no empirical, or “fitting” parameters.

All the eight \( \Delta V_{TH} \) curves in Figure 3, if redrawn dividing the abscissa by their corresponding values of \( \lambda_{1} \), converge nicely into a tight bunch, forming a unified \( \Delta V_{TH} / L / \lambda_{1} \) dependence, as shown in Figure 4. It is indeed striking to observe such convergence because the scale length \( \lambda_{1} \) and the \( V_{TH} \) model (4) are derived independently under conditions of fixed-charge-determined and mobile-charge-determined electric fields, respectively. And as such, \( \lambda_{1} \) does not figure explicitly in the \( V_{TH} \) and \( \Delta V_{TH} \) models at all. Nonetheless, the ratio of \( L / \lambda_{1} \) is proven to be an excellent quantitative measure of threshold roll-off in undoped DG MOSFET’s.

A quantitative scaling theory is thus formed by the unified dependencies of \( S-L / \lambda_{1} \) and \( \Delta V_{TH} / L / \lambda_{1} \), shown in Figure 2 and Figure 4. It characterises an undoped DG MOSFET quantitatively and with high accuracy by \( S \) and \( \Delta V_{TH} \) through a single ratio of \( L / \lambda_{1} \) for all possible design combinations of \( L, t_{so}, \) and \( t_{mo} \).
3. Threshold variations

\[ \frac{\partial V_{TH}}{\partial t_{Si}} = \frac{4\theta}{\beta r} \sinh \theta \left( \frac{\cosh \theta}{2} \right) + \frac{\beta q}{kT} \ln \left( \frac{Q_{in}}{n_{Si}} \right) \left( \frac{B}{2\beta \cosh (\theta/2)} \right) \frac{\sinh (\theta/2) + \cosh (\theta/2) + 8\cosh (\theta/2)}{1 + 3 \cosh (\theta/2) + \cosh (3\theta/2)} \right] \]

where \( \beta = q/kT \).

Figure 5. \( V_{TH} \) fluctuation per 10% tolerance of \( t_{Si} \) vs. (a) \( L \), and (b) the ratio of \( L/\lambda_1 \). Discrete points are calculated from numerical simulations of [3].

The analytical \( V_{TH} \) model (4) makes it straightforward to study threshold variations due to process tolerance by taking partial derivative of \( V_{TH} \) with respect to a device parameter, \( X \), as

\[ \frac{\partial V_{TH}}{\partial X} = \frac{\partial V_{TH}}{\partial X} \frac{\delta X}{X} \]

\( \delta V_{TH} \) is the threshold variation caused by the process tolerance of parameter \( X \). The relevant partial derivatives with respect to \( t_{Si} \), \( L \), and \( \lambda_1 \) are obtained as

\[ \frac{\partial V_{TH}}{\partial t_{Si}} = \frac{4\theta}{\beta r} \sinh \theta \left( \frac{1}{2} \right) \frac{\cosh (\theta/2)}{\beta \cosh (\theta/2)} \]

\[ \frac{\partial V_{TH}}{\partial L} = \frac{2\beta q}{kT} \sinh \theta \left( \frac{1}{2} \right) \frac{\cosh (\theta/2)}{\beta \cosh (\theta/2)} \ln \left( \frac{Q_{in}}{n_{Si}} \right) \left( 2\cosh (\theta/2) + 8 \cosh (\theta/2) \right) \]

\[ \frac{\partial V_{TH}}{\partial \lambda_1} = \frac{2\beta q}{kT} \ln \left( \frac{Q_{in}}{n_{Si}} \right) \left( 2\cosh (\theta/2) + 8 \cosh (\theta/2) \right) \]

Using (7), the threshold variation due to 10% process tolerance of \( t_{Si} \) (\( \delta V_{TH,t_{Si}} \)) is calculated and shown in Figure 5a for various designs. Increase of \( t_{Si} \) leads to degradation of short channel effects, and consequently, decreases threshold voltage, which explains the negativity of the partial derivative of \( \partial V_{TH}/\partial t_{Si} \). For exactly the same reason, \( \delta V_{TH,t_{Si}} \) worsens with thicker \( t_{Si} \) and shorter \( L \), both of which make short channel effects more severe. Following the quantitative scaling theory, all curves in Figure 5a are transformed dividing the abscissa by their corresponding values of \( \lambda_1 \). It results in a unified dependence of \( \delta V_{TH,t_{Si}} \) on the ratio of \( L/\lambda_1 \) as illustrated in Figure 5b. Three discrete data points calculated from numerical simulations of [3] are shown alongside for comparison. Figure 5b definitely manifests the power and effectiveness of the quantitative scaling theory. Similarly, threshold variations due to \( L \) and \( \lambda_1 \) tolerance (\( \delta V_{TH,L} \) and \( \delta V_{TH,\lambda_1} \)) are obtained using (8) and (9) and shown in Figure 6 and Figure 7, respectively. Their dependencies on device parameters are identically explained as in the case with \( t_{Si} \) variations. Unified dependencies of \( \delta V_{TH,L} \) and \( \delta V_{TH,\lambda_1} \) on the ratio of \( L/\lambda_1 \) are obtained in Figure 6b and Figure 7b, respectively.
Figure 7. $V_{TH}$ fluctuation per 10% tolerance of $t_{ox}$ vs. (a) $L$, and (b) the ratio of $L/\lambda_1$.

Figure 8. Stacked-bar plot of overall $\Delta V_{TH}$ vs. $L/\lambda_1$, consisting of contributions from $L$, $t_{Si}$, and $t_{ox}$.

The worst-case overall threshold variation ($\delta V_{TH}$) will result from the most unfavourable combination of $L$, $t_{Si}$, and $t_{ox}$ variations, and is given as:

$$\delta V_{TH} = \left| S_{L/\lambda_1} \right| + \left| S_{t_{Si}} \right| + \left| S_{t_{ox}} \right|.$$  (10)

Its dependence on the ratio of $L/\lambda_1$ is shown in Figure 8 by combining Figure 5b, Figure 6b, and Figure 7b in a stacked-bar fashion. Figure 8 clearly shows that threshold variations in undoped DG MOSFET's are sheerly determined by the ratio of $L/\lambda_1$, or equivalently, the degree of short channel effects. For the same process tolerance, $t_{ox}$ causes the least, relatively insignificant amount of threshold variations, while $L$ causes about 30-50% more threshold variations than $t_{Si}$ does. For example, at $L/\lambda_1=6$, 10% variations of $L$, $t_{Si}$, and $t_{ox}$ lead to about 18, 14, and 5 mV threshold variations, respectively. If $t_{Si}$ is determined by lithography that typically provides process control on the order of 10% of $L$, the resulting tolerance $\delta t_{Si}$ can be significantly greater than 10% due to much larger values of $L$ than $t_{Si}$. It may then render $t_{Si}$ the foremost cause of threshold variations.

4. Conclusions

A quantitative scaling theory of undoped double-gate MOSFET's is developed based on unified dependencies of $S$ and $\Delta V_{TH}$ on the ratio of $L/\lambda_1$. As an application of this theory, compact, analytical models of threshold variations are established. It is found that threshold variations are determined completely by three unique, well-characterised dependencies on the ratio of $L/\lambda_1$, contributed by $L$, $t_{Si}$, and $t_{ox}$. For the same process tolerance, $t_{ox}$ causes the least, relatively insignificant amount of threshold variations, while $L$ causes about 30-50% more threshold variations than $t_{Si}$ does. For example, at $L/\lambda_1=6$, which corresponds to $S=66$ mV/dec and $\Delta V_{TH}=60$ mV, 10% variations of $L$, $t_{Si}$, and $t_{ox}$ lead to about 18, 14, and 5 mV threshold variations, respectively. Threshold variations may deteriorate very much if $t_{Si}$ is to be determined by lithography.

This work is supported by DARPA (F33615-97-J-1132) and NSF (ECS-0108650).

References