
Copyright © 2013 Institute of Electrical and Electronics Engineers

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

Content must not be changed in any way or reproduced in any format or medium without the formal permission of the copyright holder(s)

When referring to this work, full bibliographic details must be given

http://eprints.gla.ac.uk/83180/

Deposited on: 26 August 2014
Geometry, Temperature and Body Bias
Dependence of Statistical Variability in 20nm Bulk CMOS Technology: A Comprehensive Simulation Analysis

Xingsheng Wang, Member, IEEE, Fikru Adamu-Lema, Member, IEEE, Binjie Cheng, Member, IEEE, and Asen Asenov, Fellow, IEEE

Abstract—Conventional bulk CMOS, which is arguably most vulnerable to statistical variability, has been the workhorse of the electronic industry for more than three decades. In this paper, the dependence of the statistical variability of key figures of merit on gate geometry, temperature and body bias in 25nm gate-length MOSFETs, representative for the 20nm CMOS technology generation, are systematically investigated using 3D statistical simulations. The impact of all relevant sources of statistical variability is taken into account. The geometry dependence of the threshold voltage is systematically investigated, and the dependence of other key transistor figures of merit) are well identified and investigated by means of 3D device simulations and experimental data analysis. The ever increasing doping, deployed to combat short channel effects, dominates the statistical variability in bulk MOSFETs which is reaching already critical levels at 28nm CMOS technology. As a result in its 22nm technology generation, Intel introduced the novel ‘tri-gate’ FinFET architecture, which has superior electrostatic integrity, tolerates low channel doping and has the potential to reduce significantly the statistical variability. Fully-depleted planar SOI transistors are also introduced by ST at 28nm CMOS technology in order to reduce the statistical variability. Many technology providers, however, continue to rely on conventional bulk transistors at the 20nm CMOS technology generation planned for early introduction in 2013.

Bearing in mind that the statistical variability can be reliably measured only in mature technologies, in this paper we deploy comprehensive 3D device simulations in order to provide early but accurate information for the level of statistical variability in the forthcoming bulk 20nm CMOS technology. This is based on meticulous validation of our simulation technology in respect of 45/40 nm [19] and 32/28-nm technology [14]. This paper also goes beyond most of the previously published simulation and measurement results, reporting a systematic study on the impact of geometry, substrate bias and temperature on the statistical variability of carefully designed ‘template’ 25nm gate-length MOSFETs meeting the specifications of 20nm CMOS.

I. INTRODUCTION

Conventional bulk complementary metal oxide semiconductor (CMOS) field-effect transistors (MOSFETs) have undergone a relentless down-scaling, driven by the demand for ever increasing functionality of the corresponding circuits and systems, and the constant pressure for reducing the cost per function. Controlling the short-channel effect, increasing the performance and reducing the power dissipation, have become key challenges on the trajectory of scaling bulk transistors to nano-scale regime. Break-through innovations including the introduction of strain at 90nm [1] and the integration of high-κ/metal gate to 45nm [2] have kept the bulk MOSFET scaling on track but at increasing technology development cost.

However, statistical parameter variations are becoming insurmountable challenge of the bulk CMOS technology [3][4], adversely affecting device integration and SRAM yield [5], and causing excess leakage and timing margin loss [5]-[10]. The main statistical variability (SV) sources including random discrete dopants (RDD) [11], line edge roughness (LER) [12], polysilicon (PSG) [13] and metal gate granularity (MGG) [4] are well identified and investigated by means of 3D device simulations and experimental data analysis. The ever increasing doping, deployed to combat short channel effects, dominates the statistical variability in bulk MOSFETs which is reaching already critical levels at 28nm CMOS technology [14]. As a result in its 22nm technology generation, Intel introduced the novel ‘tri-gate’ FinFET architecture [15] that has superior electrostatic integrity, tolerates low channel doping and has the potential to reduce significantly the statistical variability [16]. Fully-depleted planar SOI transistors are also introduced by ST at 28nm CMOS [17] to reduce the statistical variability. Many technology providers, however, continue to rely on conventional bulk transistors at the 20nm CMOS technology generation planned for early introduction in 2013.

Manuscript received December 11, 2012, revised February 19, 2013, accepted March 15, 2013. This work was supported in part by the EU ENIAC joint undertaking project “MODEling and DESign of Reliable, process variation-aware Nanoelectronic devices, circuits and systems” (MODERN) under Grant ENIAC-12003, and UK EPSRC platform grant “Atomic Scale Simulation of Nanoelectronic Devices” (EP/E033441/1).

X. Wang, and F. Adamu-Lema are with Device Modelling Group in the School of Engineering at University of Glasgow, G12 8LT, Glasgow, U.K. (e-mail: Xingsheng.Wang@glasgow.ac.uk, xingsheng.wang@ieee.org).

B. Cheng, and A. Asenov are with Device Modelling Group in the School of Engineering at University of Glasgow, G12 8LT, Glasgow, U.K., also with Gold Standard Simulations Ltd., Rankine Building, G12 8LT, Glasgow, U.K.
Section II of the paper describes the design of the 25nm gate length template transistor. The simulation methodology is outlined in Section III. The generic variability results highlighting the relative importance of the different variability sources are presented in Section IV. The geometry, temperature and back bias dependence of the statistical variability are reports in Sections V, VI and VII respectively before drawing the conclusions in Section VIII.

II. THE 20NM CMOS TEMPLATE TRANSISTOR DESIGN

The performance and leakage requirements and the related control of short-channel effects, have been in the past the premier targets when moving to a new CMOS technology generation. However, managing the device variability becomes a critical issue in sub 65 nm technology generations [20]. Statistical variability, introduced by the discreteness of charge and granularity of matter, has become a major concern. This has focused the attention on managing (and perhaps reducing) to certain degree the bulk MOSFET statistical variability through variability-aware device design. The template transistor, which has been used for this investigation has been designed to meet the performance and leakage prescriptions for the 20nm CMOS technology generation [21], keeping simultaneously the statistical variability low. Relatively simple but efficient process simulation engine [22] embedded in the GSS 3D ‘atomistic’ simulator GARAND [23] is used in the device design. Selected electrical characteristics and device parameters are given in Table 1.

The variability-aware design reflects two major considerations. 1) For bulk MOSFETs there is a strong correlation between the doping concentration close to the interface and the RDD induced statistical variability. Implementing retrograde vertical doping profile reduces the impact of random dopant. Careful design of the halo (pocket) doping profile can enable more aggressive retrograde doping profile without significantly compromising the RDD variability. However with the reduction of the channel length the overlapping halo region can dramatically increase channel doping and the corresponding statistical variability. This is illustrated in Fig. 1 with channel doping profiles for different gate-lengths. 2) Achieving a gradual threshold-voltage roll off around the nominal channel length of 25 nm mitigates the LER induced statistical variability [14]. The threshold voltage roll-off achieved in the template transistor is reported in Figure 2.

| TABLE 1: 25-NM GATE-LENGTH CMOS TEMPLATE DEVICE PARAMETERS AND CHARACTERISTICS. |
|---------------------|-----|
| L0 (nm)             | 25  |
| EOT (nm)            | 0.85|
| NCH (x10^18 cm^-3) | ~5.0|
| VDD (V)             | 1.0 |   |
| IDSAT (mA/µm) [N/P] | 1.35/0.99 |
| Ieff (mA/µm) [N/P]  | 0.765/0.543 |
| Ioff (nA/µm) [N/P]  | 89/106 |
| DIBL (mV/V) [N/P]   | 100/125 |

High-k/TiN is adopted for gate stack. Although TiN is a mid band-gap metal, appropriate work-functions (WF) are assumed when adjusting the threshold voltage and the corresponding leakage [24]. The tensile/compressive engineering is assumed for the n- and p- channel MOSFETs to achieve the desirable performance. At leakage current of Ioff ~100mA/µm, the corresponding saturation current is IDSAT ~1.35mA/µm for n-channel and IDSAT ~0.99mA/µm for the p-channel transistors. The effective drive currents (Ieff) [25] are 0.765 mA/µm and 0.543 mA/µm respectively.

Figure 1. The net doping profiles for different gate-lengths in NMOS. The channel doping concentration increases due to the tilted high halo (pocket) doping overlapping.

Figure 2. The device transfer characteristics (a) and threshold-voltage roll-off characteristics (b).

III. SIMULATION METHODOLOGY

GARAND is also used in this study for the simulation of the statistical variability associated with the individual and combined variability sources. The simulator is meticulously calibrated and validated in respect of statistical variability simulations and measurements at 45/40 nm [19], and at 32/28 nm technology generations [14]. It has been demonstrated that that RDD, LER and MGG are major statistical variability sources in 32/28nm high-k/metal gate bulk CMOS technology. The resolution of the individual discrete dopants in the RDF simulations employs fine meshing in conjunction with density gradient quantum corrections in resolving accurately the impact of each individual dopant. This prevents artificial charge trapping in the sharply resolved Coulomb wells of the ionised dopants and avoids acute mesh-spacing sensitivity [26]. Gate LER is introduced statistically using the power spectrum of a Gaussian autocorrelation function [12] and is parameterized by its correlation length of 30-nm and RMS of 1.33-nm. The metal gate granularity MGG has become important source of statistical variability. High temperature annealing and dopant activation in metal gate first technology result in poly-crystallization of the metal gate. Metal grains with different crystallographic orientation have different
atomic densities at the interface with the gate oxide and therefore different workfunctions. The MGG induced work-function variation has been introduced realistically using the simulation technology described in details in [4]. In this work, TiN was used as metal gate material with two dominant random grain orientations occurring with a probability of 40% and 60% and with a WF difference of 0.2V. The adopted average grain diameter is 6-nm according to [14]. Statistical 3D simulations have been carried out using individual and combined variability sources at different geometries, temperature, and substrate bias conditions. Ensembles of 1000 microscopically different transistors were simulated to allow accurate evaluation of the corresponding statistical distributions. Fig. 3 illustrates an example of a three-dimensional simulation domain, showing the combined effects of RDD, LER and MGG on the potential and carrier concentration distributions in one ‘random’ transistor.

![3D View of a Bulk Planar Device](image)

**Figure 3.** The 3D view of an ‘atomistic’ bulk planar device, showing electron density. It demonstrates the major statistical variability sources including RDD, LER, and MGG. The gate length and channel width both are 25-nm. The colour bar is in the logarithm scale.

### IV. STATISTICAL VARIABILITY

Fig. 4 illustrates the impact of the individual and combined sources of statistical variability on the threshold voltage standard deviation $\sigma V_T$ of the simulated ‘template’ transistors. It is clear that RDD remains the dominant statistical variability source in 20nm CMOS technology. The corresponding $\sigma V_T$ of the n-channel and p-channel MOSFETs are very similar, close to 60-mV and 54-mV at high and low drain-biases respectively. The gate LER induced $\sigma V_T$ is 23 mV for nMOSFET and 27 mV for pMOSFET at high drain-bias. The larger LER induced variability in the pMOSFET is related to the steeper $V_T$ roll off illustrated in Figure 2 due to the deeper p-n junctions associated with faster boron diffusion [27]. Both LER and RDD show drain-bias dependence of $\sigma V_T$. The MGG, which is expected to be present in metal gate first technology, in general has stronger impact on $\sigma V_T$ compared to LER. The corresponding combined statistical variability in the pMOSFET is larger than in nMOSFET due to slightly increased doping and worsened short-channel effects. It might be possible to achieve amorphous metal gate by limiting or avoiding all together the high-temperature thermal processing that results in metal poly-crystallization and therefore to eliminate the MGG as a source of statistical variability. This will result in overall reduction of the total variability in both the nMOSFET and pMOSFET by approximately 7-8mV. In contrast, FinFETs and FD SOI transistors tolerate low channel doping, practically eliminating the RDD effects, and dramatically reducing the statistical variability [16][17].

![Statistical Variability Plot](image)

**Figure 4.** The comparison of individual variability source induced threshold-voltage fluctuations in CMOS devices with square gate-area (W= L=25nm).

The correlations between the key transistor figures of merit are illustrated in Figure 5. $V_T$, $I_{OFF}$ (in logarithm scale), and $I_{ON}$ are closely correlated. The correlation coefficient between $V_T$ and Log($I_{OFF}$) is approximately 0.99, and the coefficient between $V_T$ and $I_{ON}$ is approximately 0.92. Despite the fact that the drift-diffusion simulation does not include ionized impurity scattering variations and underestimate $I_{ON}$ variations [28], the $\sigma V_T$ scattering cannot completely describe the $I_{ON}$ variation behavior. This is mainly because of transport variation due to dopants induced current percolation paths [29], and source/drain resistance variation due to the dopant number variation in the extensions [30] [31]. The threshold-voltage fluctuation also cannot fully represent the sub-threshold variation, as demonstrated in Figure 5. The correlation coefficient between drain-induced barrier lowering (DIBL) and threshold voltage is less than 0.5.

An insight in the DIBL variability and its de-correlation with the threshold voltage is provided in Fig. 6(a) by selecting as an example the ‘atomistic’ transistor with the largest DIBL illustrated. In this device the drain side of the channel has larger amount of acceptors and is under a metal grain with high work function. At high drain bias the channel potential near the drain is lowered by the drain potential, removing the influence of dopants and the grain close to the drain as

![Graph Illustrating DIBL Variability](image)

**Figure 5.** The correlations between figures of merit of 25nm gate-length NMOS (down-left) and PMOS (up-right) due to combined variability sources.
illustrated in Fig. 6(b) and resulting in low threshold voltage controlled by the dopants and the grains near the source end of the channel. At low drain bias however the device has large threshold voltage determined by the dopants crowding and the unfavorable work function near the drain. On the contrary, in the small DIBL cases, the dopants crowding and/or the grains with unfavorable work function are at the source side.

Figure 6. (a) Detailed observations on the NMOS device with the largest DIBL. The top slice shows the channel potential. The bottom shows the electron density. Blue dots indicate acceptors. H/L indicates the metal grain with high/low work-function (WF). (b) The schematic view of large DIBL mechanism caused by dopants and/or work-function variation in atomistic devices.

V. GEOMETRY DEPENDENCE

A. Gate Length

The threshold-voltage fluctuations at both high and low drain bias substantially increase with the reduction of the gate-length, as shown in Fig. 7. However this dependence does not follow the Pelgrorn’s law [32] according to which \( \sigma V_T \) at identical channel length should be inversely proportional to square root of channel-length. Indeed the inset of Fig. 7 shows a marked increase in the mismatch coefficient \( A_{VT} \) with the reduction of the channel length. Here we use the following definition of \( A_{VT} \): \( \sigma V_T = A_{VT} / \sqrt{LW} \). This behavior is consistent with previous measurements and characterization [33][34][35][36] and is related to the presence of halo implants. The tilted halo implantations from source and drain sides overlap in the channel below the gate with the reduction of the gate-length beyond 50nm, which leads to increase in channel doping and \( A_{VT} \). For example at high drain bias \( A_{VT} \) is 1.80 for NMOS and 1.87-mV/um for PMOS at \( L_G \geq 25 \text{nm} \) instead of the 1.5-mV/um flat value at \( L_G > 50 \text{nm} \).

Fig. 8 illustrated the DIBL distribution in the nMOSFET on a logarithmic scale indicating close to log-normal distribution, which is consistent with the observation in the experimental measurement [37]. The average values of the distributions and the spread are reduced with the increase in the gate length. As discussed in section IV the variation in the \( V_T \) difference at high and low drain-bias is associated with the asymmetry of the random dopant distribution and the metal grain workfunction distribution along the channel.

Figure 7. The gate-length dependence of threshold-voltage fluctuation. The channel width is fixed at 25-nm.

Figure 8. The Q-Q plots of DIBL for different gate-length n-channel MOSFETs. DIBL has log-normal distribution. The threshold-voltage is determined by fixed subthreshold current.

B. Channel Width

The channel width dependence of \( \sigma V_T \) for transistors with 25 nm gate length is illustrated in Figure 9. Again the width dependence shows a marked departure from the Pelgrom’s law. At a constant channel length the mismatch coefficient \( A_{VT} \) gradually increases with the increase of the channel width. The increase in \( A_{VT} \) is stronger pronounced at high drain voltage.

To understand the reasons for the departure from the Pelgrorn’s law when considering the width dependence of \( \sigma V_T \) we have studied the contributions of the individual SV sources, RDD, LER and MGG to the channel width dependence. Figure 10 illustrated the width dependence of \( \sigma V_T \) for each of the above SV sources. The expected from the Pelgrorn’s law \( A_{VT} / \sqrt{W} \) dependence is fitted to the corresponding 25nm \( \sigma V_T \) and also plotted in the same figure. For both RDD or MGG the width dependence accurately follows the Pelgrorn’s low. However in the case of LER \( \sigma V_T \) decreases slower than expected [38], and at large channel width starts to overtake MGG induced \( \sigma V_T \). Therefore the LER width dependence is responsible for the observed in Figure 9 upward trend of \( A_{VT} \) with increasing channel width. The width dependence of LER induced \( \sigma V_T \) is virtually independent on LER rms but depends on the LER correlation length. Moreover, fringing effects from shallow trench isolation at width ends will additionally modify the width dependence [39].
certain characteristics simulated statistical 246K, 273K, 300K transistors were carried temperature. can lead to variability is investigated in this section. Environmental design kits extraction and generation techniques are needed in order to Secondly, more sophisticated statistical compact model width needed individual variability sources.

It is important to note that in extremely scaled bulk MOSFETs both the channel length and the channel width dependence of $\sigma_{VT}$ strongly deviates from the Pelgrom’s law. The implications are twofold. Firstly special attention is needed in order to properly capture the channel length and width dependences when measuring and characterizing the mismatch in the corresponding technology generations. Secondly, more sophisticated statistical compact model extraction and generation techniques are needed in order to accurately capture the geometry dependence of the statistical variability in compact models and the corresponding process design kits (PDKs).

VI. TEMPERATURE DEPENDENCE

The lattice temperature dependence of the statistical variability is investigated in this section. Environmental temperature and self heating related to operation conditions can lead to global or local changes in the silicon lattice temperature. Statistical variability simulations for the template transistors were carried out for lattice temperatures of 219K, 246K, 273K, 300K and 327K. Fig. 11(a) illustrates the simulated statistical $I_D-V_G$ characteristics at two different temperatures. As expected, for each atomistic device the $I_D-V_G$ characteristics for the two different temperatures crossovers at certain gate-voltage above the threshold-voltage, however the crossover point is different for each individual atomistic device. The increase in the temperature degrades the device performance reducing simultaneously $V_T$ and $I_{ON}$, and increasing $I_{OFF}$. The scatter plots of $I_{ON}$ and $I_{OFF}$ in Fig. 11(b) show the impact of the temperature on the statistical performance of the simulated ensemble nMOSFETs.

![Figure 9: Channel-width dependence of threshold-voltage fluctuation.](image)

![Figure 10: Width dependence of threshold voltage variations induced by individual variability sources.](image)

The dependence of the averages and the standard deviations of key transistor figures of merit are plotted in Fig. 12 (a,b) as a function of the temperature. The average threshold-voltage $<V_T>$ linearly decreases with the temperature as expected [40][41], with slopes 0.55/0.72-mV/K at high drain-bias, 0.50/0.58-mV/K at low drain-bias for NMOS/PMOS respectively. The average subthreshold slope ($<SS>$) almost linearly increases with temperature. The $<\Delta I_{OFF}>$ increases due to decrease in $<V_T>$ and increase in $<SS>$. However, $\sigma_{VT}$ remains nearly constant with temperature. Simultaneously $\sigma_{SS}$ increases with the increase in temperature, and both $\alpha_{OFF}$ and $\sigma_{I_{OFF}}$ decrease with the increase of the temperature.

The gate-voltage crossover point in $I_D-V_G$ characteristics at different temperatures exhibits fluctuation from device to device as well. As the scatter plot for crossover gate-voltages and the corresponding threshold voltage at high/low drain biases is shown in Figure 13. There is a strong correlation between the two with correlation coefficient as high as 0.97.
VII. BODY BIAS DEPENDENCE

Applying body bias (V_{bb}) is a common practice in circuits for a performance boost or for leakage control. In this section, the interplay between body bias effect and statistical variability has been investigated. Reverse body bias can increase threshold-voltage simultaneously reducing leakage and drive current while forward bias reduces the threshold voltage simultaneously increasing leakage and drive current. This is illustrated in Fig. 14(a) for the statistical ensemble of nMOSFETs. The V_T at different body biases shows strong correlation with zero body bias V_T but with different slopes as shown in scatter plots on Fig. 14(b).

The threshold-voltage average values and standard deviations in the nMOSFET are monotonically reduced when sweeping the body bias from negative to positive values as shown in Fig. 15. From the inset of Fig. 15(b) is clear that the forward bias reduces the depletion width in the channel region, leading to early inversion and reduced threshold-voltage while the reverse body bias increases the gate depletion width and the threshold voltage.

The physical cause for larger threshold-voltage fluctuation at reverse bias compared to forward bias conditions is probably mainly due to the enhanced RDD impact from the increased bulk depletion region. Moreover, with the reverse body bias the increase of the source/drain junction depletion width reduces the effective channel-length and weakens the gate-control over the channel as shown in Fig. 16(a). LER variability is also increased with the application of negative body bias due to increased channel length sensitivity of the threshold voltage as illustrated in Fig. 16(b).

Figure 13. The correlation between gate-voltages of crossover points with threshold-voltages.

VIII. CONCLUSION

This paper presents the most comprehensive simulation study of statistical variability and its dependence on geometry, temperature and body bias in 20nm bulk planar CMOS technology. Reduction of the statistical variability can be achieved through variability-aware device design. The geometrical dependence of the statistical variability deviates from traditional Perglon’s rule due to complex doping profiles, and due to different properties of statistical variability sources. The DIBL follows log-normal distribution and is physically analyzed. The temperature shows strong impact on the transistor figures of merit and their variability. Depending on its polarity the body bias can greatly affect performance and statistical variability. All of the revealed dependences have to be carefully considered when characterizing the statistical variability and when incorporating statistical variability in compact models and corresponding PDKs.

REFERENCES


Xingsheng Wang (M’11) received Master degree in mathematics from Tsinghua University, Beijing China in Jan. 2007, and Ph.D. degree in electronics and electrical engineering from University of Glasgow, U.K. in 2010. His research interest includes nanoscale devices, variability and reliability, and modelling and simulations.
Fikru Adamu-Lema (M’12) received the Ph.D. degree in 2006. He is currently a research associate at the Device Modelling Group, University of Glasgow, UK.

Binjie Cheng (M’12) received the Ph.D. degree in Electronic Science and Technology from Xi’an Jiaotong University, Xi’an, China. Since 2002, he is a member of Device Modelling Group, University of Glasgow. Currently he is working on statistic compact modelling and statistical circuit simulation methodologies.

Asen Asenov (M’96–SM’05–F’11) is a founder and CEO of Gold Standard Simulations (GSS) Ltd. (www.goldstandardsimulations.com) and a Director of SureCore, Ltd, As James Watt Professor in Electrical Engineering at Glasgow University he also leads the Glasgow Device Modelling Group.