Grain-Orientation Induced Work Function Variation in Nanoscale Metal-Gate Transistors—Part I: Modeling, Analysis, and Experimental Validation

Hamed F. Dadgour, Student Member, IEEE, Kazuhiko Endo, Member, IEEE, Vivek K. De, Senior Member, IEEE, and Kaustav Banerjee, Senior Member, IEEE

Abstract—This paper highlights and experimentally verifies a new source of random threshold-voltage ($V_{th}$) fluctuation in emerging metal-gate transistors and proposes a statistical framework to investigate its device and circuit-level implications. The new source of variability, christened work-function (WF) variation (WFV), is caused by the dependence of metal WF on the orientation of its grains. The experimentally measured data reported in this paper confirm the existence of such variations in both planar and nonplanar high-$k$ metal-gate transistors. As a result of WFV, the WFs of metal gates are statistical distributions instead of deterministic values. In this paper, the key parameters of such WF distributions are analytically modeled by identifying the physical dimensions of the devices and properties of materials used in the fabrication. It is shown that WFV can be modeled by a multinomial distribution where the key parameters of its probability distribution function can be calculated in terms of the aforementioned parameters. The analysis reveals that WFV will contribute a key source of $V_{th}$ variability in emerging generations of metal-gate devices. Using the proposed framework, one can investigate the implications of WFV for process, device, and circuit design, which are discussed in Part II.

Index Terms—Grain orientation, metal-gate devices, random variations, reliability, subthreshold leakage, threshold voltage, VLSI design, work-function (WF) variation (WFV).

I. INTRODUCTION

Traditional structure of silicon-based MOSFET devices (polysilicon–oxide–silicon) has proven to be extremely successful and has been the main technological driver of the electronics industry. However, to achieve higher performance, the channel length of these devices has been continuously scaled down during the past few decades, which, in turn, resulted in increased levels of short-channel effects. To counter these undesired effects, gate-oxide ($SiO_2$) thickness has also been scaled down in an attempt to sustain control of the gate terminal over the channel area. However, this technique results in elevated levels of gate leakage current (via direct tunneling) through the thinned gate-oxide layer. It has been predicted that for sub-65-nm technology nodes, the amount of gate leakage will surpass that of the subthreshold leakage if the oxide material remains unchanged [1], [2]. Therefore, insulating materials with higher dielectric constant (high-$k$) are being introduced to lower the gate leakage current (high-$k$ allows thicker dielectric layers while providing identical control over the channel).

Unfortunately, it turns out that the high-$k$ materials are not compatible with the polysilicon gate due to two major difficulties, namely, Fermi level pinning and phonon scattering. Fermi level pinning [3] occurs due to high density of the defects formed at the polysilicon/high-$k$ dielectric interface, which results in shifting of the threshold voltage ($V_{th}$) to higher values, thereby lowering the device drive current. The second phenomenon, phonon scattering [4], happens when optical phonon vibration (in the high-$k$ dielectric) interferes with movement of the electrons in the channel and, as a result, reduces the mobility of electrons. It has been shown that both these obstacles can be resolved by replacing the polysilicon-gate electrode by a metal electrode [5], [6]. Moreover, introduction of the metal gate results in lower gate resistance and eliminates the poly depletion layer, thereby increasing the ON current of the transistors.

Despite all these advantages, as has been recently highlighted for the first time, using metal-gate introduces a new source of random $V_{th}$ variations due to the dependence of the metal work function (WF) on the orientation of the metal grains [7], [8]. It is known that metal grains usually grow up to few nanometers in size under temperatures normally used in integrated circuit (IC) fabrication. Since the gate dimensions are in the range of few tens of nanometers, it is expected that the gate area will contain only a small number (~10–100) of grains. On the other hand, as will be discussed in Section II, the WF of each grain is a function of its orientation, which is not controllable during the growth period, and hence, the orientation of each grain is determined randomly [9]. The combined effect of low number of grains and randomness of their WF values will cause the overall WF of the fabricated metal gate to be a probabilistic distribution rather than a deterministic value. Since the $V_{th}$ of a MOS device is influenced by the WF of the gate, the WF...
variation (WFV) in metal-gate devices leads to $V_{th}$ fluctuations. Fluctuations in WF for metal-gate devices located at different positions on a wafer have been experimentally measured and attributed to varying proportions of different crystal orientations at different locations on the wafer caused by irregularities in the process conditions [10]. However, randomness of the grain orientation at the “device level” is not considered in [10].

In this work, a detailed modeling and analysis of WFV using a comprehensive statistical framework is presented, which can be applied to evaluate the impact of WFV on emerging high-$k\ell$ metal-gate devices, including nanowire [11] and nanotube-based FETs [12]. More specifically, the proposed model can be employed to achieve two important goals: (1) to identify appropriate materials and process conditions that can reduce the impact of random grain orientations and (2) to analyze the impact of such variations on the device characteristics as well as key circuit-level performance and reliability parameters.

The rest of this paper is organized as follows: Section II provides a physical understanding of the dependence of the WF on the microstructure of metals along with experimental verifications. Section III justifies and illustrates the formulation of the proposed statistical framework for analysis of WFV, and Section IV summarizes this paper.

II. DEPENDENCE OF METAL WF ON THE GRAIN ORIENTATION

The WF [measured in electronvolt (eV)] is defined as the minimum energy required for removing an electron from a solid material to a point immediately outside its surface (local vacuum level). The WF is the sum of the bulk chemical potential (due to electron–electron correlation and exchange effects) [13] and the surface dipole potential (discussed in the following subsections). Although the first component is constant for a given metal and process conditions, the second component depends on crystal orientation resulting in the dependence of the WF on crystal orientation.

A. Microstructure of Deposited Thin-Film Metals

Metals exist in the form of crystals (periodic lattice structure) in nature where each atom forms several bonds with adjacent atoms and are arranged in a regular array. However, due to defects and disorientations, a crystal cannot grow infinitely in size, and hence, metallic thin films are always composed of several “crystal grains” (regions of regularity) with different orientations, separated by “grain boundaries.” At the grain boundaries, atoms become misaligned. The grain size and structure are determined by “nucleation” and “growth” [14], which, in turn, are influenced by the deposition conditions and thermal treatment of the sample during IC processing. It has been shown that grain size increases with increasing substrate temperature during the deposition [15], [16]. In the temperature ranges compatible with IC fabrication, the grain size of the metal thin films can grow up to 5–20 nm, depending on the type of metal. Considering the fact that the gate length in current technology nodes is only a few tens of nanometers, it can be concluded that the gate area is composed of only a few grains.

Fig. 1. Unit cells corresponding to the most common metal-crystal structures (BCC, FCC, and HCP). Most metal crystals take only one of the aforementioned structures; however, few metals can possess two or more forms.

Fig. 2. Illustration of the surface-density concept for different orientations of the FCC crystal structure: (a) FCC unit cell. (b) Cross-sectional view of surface planes along (100), (c) (110), and (d) (111) directions.

B. Grain Orientation and Surface Density of Metal Thin Films

The origin of the dependence of the WF on grain orientation of metal has been explained in [17] and illustrated analytically by Lang and Kohn’s model for electron density at the metal surfaces [18]. To understand this model, it is important to first introduce two basic concepts: “grain orientation” and “surface density.” To achieve this goal, one can focus on analyzing the basic building blocks of the metal crystals called “unit cell.” It is named as unit cell because the crystal structure within a grain is absolutely uniform, and it is possible to construct the entire grain by repeating the unit cell in different directions.

There are about ten different feasible unit cells for metal crystals; however, three of them, namely, body-centered cubic (BCC), face-centered cubic (FCC), and hexagonal close packed (HCP) are more common. Most metals take only one of the aforementioned forms, but there are few metals with two or more crystal structures. Unit cells corresponding to BCC, FCC, and HCP crystals are shown in Fig. 1. In this figure, the dark circles represent metal atoms that are located further from the viewer, the gray ones are in the middle, and the white circles show the closer atoms.

The orientation of a metal crystal is identified by the way unit cells are terminated on the surface of the metal grain. This can be explained more effectively by visualizing the unit cells which are cut along different hypothetical planes that represent the surface of the metal grain. Fig. 2 shows this concept for the FCC crystal structure. Fig. 2(a) shows the arrangement of atoms in an FCC unit cell where eight atoms are located at the corners of the cube, and six more atoms are positioned at the centers of the sides. In this figure, the various surfaces of metal are represented by planes called “surface planes,” which are characterized by their normal vector (a vector that is perpendicular to the plane). For example, in Fig. 2(b), the surface of the grain is represented by a plane with normal vector $(x, y, z) = (1, 0, 0)$, which hereafter will be identified by its shorter form (100). Normal vectors of the surface planes are usually used to differentiate between the different grain orientations. Therefore, there are two other possibilities for the orientation of an FCC crystal structure: (110) and (111), which are shown in Fig. 2(c).
TABLE I
SURFACE DENSITY OF DIFFERENT FCC CRYSTALS (OF SIDE a)

<table>
<thead>
<tr>
<th>Orientation</th>
<th>Number of atoms</th>
<th>Area</th>
<th>Surface density</th>
</tr>
</thead>
<tbody>
<tr>
<td>⟨111⟩</td>
<td>2</td>
<td>(\frac{\sqrt{3}}{2}a^2)</td>
<td>(\frac{3}{2}\sqrt{3}a^2)</td>
</tr>
<tr>
<td>⟨100⟩</td>
<td>2</td>
<td>(a^2)</td>
<td>(\frac{3}{2}a^2)</td>
</tr>
<tr>
<td>⟨110⟩</td>
<td>2</td>
<td>(\sqrt{2}a^2)</td>
<td>(\frac{3}{2}\sqrt{2}a^2)</td>
</tr>
</tbody>
</table>

Fig. 3. Charge (both positive and negative) density at the surface of metal predicted by Lang and Kohn’s model [18]: (a) Less densely packed crystal (weaker dipole). (b) More densely packed (stronger dipole).

and (d), respectively. In this figure, dark circles are positioned on the intersection of the unit cell and the planes, whereas white circles represent those atoms that are not on the planes. Also, atoms which are located behind the surface plane and are supposed to be out of sight are shown by broken circles.

The other concept is surface density, which refers to the number of metal atoms per unit area of the metal surface. It can be observed from Fig. 2 that surface density is a function of the grain orientation. To calculate the surface density, the “effective” number of atoms located at the intersection of the surface plane and the unit cell (dark circles) must be divided by the area of the intersection. Such a calculation is performed for the three FCC grain orientations and is summarized in Table I. It can be observed that ⟨111⟩ orientation offers the highest surface density followed by ⟨100⟩ and ⟨110⟩. As will be explained in the next subsection, since WF is proportional to the surface density of atoms, the ⟨111⟩ orientation has the highest WF, followed by ⟨100⟩ and ⟨110⟩.

C. Dependence of Metal WF on the Grain Orientation

Metal WF is dependent on the surface potential created by dipoles formed near the surface [13]. The existence of these dipoles was introduced in [17] and can be explained by Lang and Kohn’s model for electron density [18]. According to this model, the electron-density distribution does not terminate at the surface of the metal, but it spills out. This means that a fraction of electrons can actually move outside the metal surface and create a negative charge outside the metal, close to its surface (shown in Fig. 3 as \(-Q_1\) and \(-Q_2\)). Since the total number of positive and negative charges in the neutral metal must be equal, the absence of those electrons that have moved out of the metal creates an imbalance between the number of electrons and metal ions, and hence, a positive charge will also appear inside the metal close to its surface (shown in Fig. 3 as \(+Q_1\) and \(+Q_2\)). These separated positive and negative charges create dipoles near the surface of the metal, which resist the removal of an electron from the metal and hence, increase the WF. Obviously, the stronger these dipoles are, the higher will be the WF. An interesting fact is that metal crystals with higher surface density create stronger dipoles as there are more number of atoms per unit area. This is shown in Fig. 3 where a densely packed metal surface [Fig. 3(b)] results in stronger dipoles than a less densely packed one (\(Q_1 < Q_2\)) [Fig. 3(a)]. Thus, it can be predicted that more densely packed surfaces also result in higher values of WFs, and hence, according to Table I, the WF of ⟨111⟩ FCC crystal is the highest followed by those of ⟨100⟩ and ⟨110⟩.

In fact, experimental measurements by Gaillard et al. [19] prove that this is indeed the case, and \(\Phi_{111} > \Phi_{100} > \Phi_{110}\), where \(\Phi_{111}, \Phi_{100}, \text{and } \Phi_{110}\) represent the WF value of crystals with orientations ⟨111⟩, ⟨100⟩, and ⟨110⟩, respectively. In their work, as shown in Fig. 4, a decananometer-scale copper thin-film wire with FCC crystal structure was used. Fig. 4(a) shows the measured WF results where the grains with high (H), medium (M), and low (L) WF values are colored as light, gray, and dark areas. The crystal orientations for grains on the same piece of metal are also identified as shown in Fig. 4(b). It can be observed that areas with high, medium, and low WF values correspond to ⟨111⟩, ⟨100⟩, and ⟨110⟩, respectively. Therefore, one can conclude that the WF values of metal grains (with different orientation) fluctuate, and the more densely packed crystal orientations exhibit higher WF values.

D. Experimental Validation of WFV

We have recently highlighted the WFV issue in metal-gate devices [7], [8], and such variations are confirmed by device-level measurements in [20] and [21]. WFV-induced \(V_{th}\) variations are reported for both bulk [Fig. 5(a)] and FinFET [Fig. 5(b)] devices in [20] and are compared with other sources of random \(V_{th}\) fluctuations. It can be observed that, for bulk transistors, the WFV component of \(\sigma V_{th}(\Delta \Phi_M)\) is even higher than the \(V_{th}\) variations caused by random fluctuations in the channel doping (\(\Delta N_A\)). More interestingly, for FinFET transistors, WFV is the most dominant source of random \(V_{th}\) variations according to these data.

E. Distribution of Grains With Different Orientations on Thin-Film Metals

It is a known fact from material science that the distribution of different grain orientations on metal films is not even. In other words, each metal has a “preferred orientation” for which
the crystal structure of the metal is more stable. For very high processing temperatures, all metal grains tend to grow in the preferred direction. However, for lower temperature ranges, such as those used in fabrication of ICs, the metal gate of transistors consists of both the preferred and other crystal orientations.

The distributions of grains with different orientations vary depending on the condition under which the metal films are grown; however, these distributions can be experimentally measured using a nondestructive-test technique called “X-ray diffraction (XRD).” In this method, the crystalline material is exposed to X-ray beams and then the reflected beams are examined to determine the different characteristics of the film. The basic principle behind the XRD method is called Bragg’s law [22], and can be explained using Fig. 6(a). In this figure, metal atoms are shown by dark circles, where the distance between two neighboring atoms is a. Two parallel X-ray beams with wavelength \( \lambda \) and angle \( \theta \) with respect to the surface of the solid are deflected by atoms in the lattice. Since the lower beam travels a longer distance, it lags behind by \( 2a \times \sin(\theta) \). The two deflected beam waves can interfere with each other constructively (overlapping waves add together to produce stronger peaks) only when the phase shift is equal to \( 2\pi \). According to the Bragg’s law, the distance \( 2a \times \sin(\theta) \) must be equal to an integer multiple of wavelength \( \lambda \) to guarantee a phase shift of \( 2\pi \).

It is interesting to know that, for different grain orientations of a solid material, deflected beams interfere constructively at different values of \( \theta \). Therefore, keeping all the other parameters constant, by varying \( \theta \) and observing the peaks, one can identify the different crystal orientations that are present in the sample. The value of \( a \), the spacing between the atoms, depends on the grain orientation. Hence, for each grain orientation, XRD peaks appear at different values of \( \theta \). For example, in Fig. 6(b), XRD results are shown for a TiN (titanium nitride) thin film grown on a silicon (Si) substrate [10]. It can be observed that this sample includes TiN grain orientations of \( \langle 111 \rangle \), \( \langle 222 \rangle \), and \( \langle 200 \rangle \) as well as Si substrate with the orientation of \( \langle 400 \rangle \).

It is also possible to identify the composition of the metal crystal using XRD experiment. The relative intensity of the different diffraction peaks can be used to estimate the number of grains with different orientations. For instance, by comparing the corresponding peaks from Fig. 6(b), it can be concluded that the concentration of the grains with orientation \( \langle 222 \rangle \) with respect to \( \langle 111 \rangle \) is approximately equal to 100/1058 \( \approx \) 10.5%. These values can also be stated in terms of probabilities; hence, in the growth process of this particular sample, the probability of growing TiN grain orientations \( \langle 111 \rangle \), \( \langle 222 \rangle \), and \( \langle 200 \rangle \) are roughly 90%, 8.5%, and 1.5%, respectively. This information on the occurrence probability of different grain orientations is essential for the modeling methodology that will be presented in the following section.

III. MODELING OF METAL-GATE WFV

Considering the approximate size to which the metal grains can grow in the IC fabrication process (5–20 nm) and the dimension of the minimum size metal gates (16–65 nm), one can conclude that the gates of nanoscale CMOS devices will be composed of only a few grains (~10–100) with a random distribution of orientations, as shown in Fig. 7. Since each grain orientation has a different WF value, the WF of the entire metal gate for one transistor cannot be predicted prior to its fabrication, and hence, the gate WF should be modeled as a probabilistic distribution rather than a deterministic value.

In this paper, a statistical approach is proposed to model the probability distribution of the WF of the metal gates. There are several parameters that have been used in this model and should be introduced. The symbols \( \Phi_1, \Phi_2, \ldots, \Phi_n \) and \( P_1, P_2, \ldots, P_n \) are used to identify the WF values of grains with different orientations and their corresponding probabilities (percentage share of a particular grain orientation in the total population of grains). The WF values \( \Phi_i \), where \( i = 1, 2, \ldots, n \), are typically determined from \( C-V \) measurements [24], and the probability values \( (P_i) \) can be extracted from analyses of XRD data [22]. It should be noted that the \( C-V \) measurement can provide the WF value of the grains with various orientations, if the grains are large enough to cover the entire area of the sample. This is difficult to achieve in the CMOS processes due to the
limited thermal budget of metal deposition steps. However, when fabricating metal thin films, it is possible to grow large grains simply by depositing the metal at elevated temperatures. These large grains can then be used to determine the values of \( \Phi_1, \Phi_2, \ldots, \Phi_n \). Moreover, note that assuming fixed probability values \( (P_i) \) for different devices (with identical gate metal) is reasonable because these values remain constant for each particular process conditions. Therefore, once the \( P_i \) values are measured, it is possible to use them repeatedly for analysis of the chips that would be fabricated under similar conditions.

It is also assumed that the grain size \( (G) \) of each type of metal film can be obtained by identifying grain boundaries on transmission electron microscope pictures of the surface of the metal gate. The average grain size for each orientation can also be extracted from the XRD data [Fig. 6(b)] by measuring the width of each diffraction peak. In the proposed model, it is assumed that all the grains with different orientations have a fixed grain size. This assumption is made based on an observation that the widths of the diffraction peaks in the XRD data were nearly identical for each particular process condition. This assumption enables us to reduce the complexity of the statistical analysis and to provide a closed-form analytical solution. If grains sizes are modeled by a statistical distribution, the mathematical treatment of the problem will be more complicated. Assuming a fixed grain size \( (G) \), for a transistor with gate length \( L \) and width \( W \), the total number of grains \( (N) \) within the metal-gate area can be calculated as \( (L/G) \times (W/G) \), assuming square-shaped grains, for simplicity.

Assuming \( X_1, X_2, \ldots, X_n \) to be the random variables that represent the number of grains with WF values of \( \Phi_1, \Phi_2, \ldots, \Phi_n \), respectively, one can calculate the WF of the metal gate \( (\Phi_M) \) as a weighted average of the WF of all the existing grains on the gate. Hence, the formula for \( \Phi_M \) can be written as follows:

\[
\Phi_M = \left( \frac{X_1}{N} \right) \Phi_1 + \left( \frac{X_2}{N} \right) \Phi_2 + \cdots + \left( \frac{X_n}{N} \right) \Phi_n. \tag{1}
\]

Here, \( (X_1/N) \) is the percentage of the gate area covered with grains whose WF is \( \Phi_1 \), and so forth. Note that \( X_1/N \) is not necessarily equal to \( P_1 \) for each metal gate. However, the average value of \( X_1/N \) for a large number of metal gates will equal \( P_1 \). Equation (1) intuitively makes sense, and it is also theoretically proven to be the case [25]. This equation indicates that the WF of a metal gate is higher if it consists of grains with higher WF values. Given the probabilities and WF values associated with each grain orientation, the goal is to calculate the mean and standard-deviation values for the random variable \( \Phi_M \).

Another approach that has been proposed for calculating \( \Phi_M \) is to consider a network of subgates (or subtransistors) with each grain modeled as an independent gate (transistor), as shown in Fig. 8 [26]. In this case, the overall current of the big (original) transistor [Fig. 8(c)] must be calculated from the combined currents provided by these subtransistors [Fig. 8(b)]. These subtransistors will have different \( V_{th} \) values, and their \( V_{gs} \) (gate–source bias) and \( V_{ds} \) (drain–source bias) values will be interdependent. This means that, although the gate voltage for all subtransistors is the same, the drain and source voltages for each subtransistor must be calculated based on Kirchhoff’s circuit laws, considering the distribution of the grains. Therefore, this complicated system must be modeled as a circuit, which is composed of a network of several subtransistors. Once the \( I_{d} - V_{gs} \) curve of the big transistor is calculated by simulating the network, one can obtain the overall WF \( (\Phi_M) \).

However, the aforementioned approach [26] does not yield accurate results for ultrashort-channel (sub-32 nm) devices since, for such ultrasmall gate areas, due to the small grain sizes (5–20 nm), the current must be modeled using a quasi-ballistic transport model, where the network model is not applicable. Additionally, to accurately estimate the \( \Phi_M \) distribution using this model, one should individually calculate the \( \Phi_M \) values for a large set of metal gates. The grain compositions of these metal gates must be generated in a random fashion (a Monte Carlo approach) in order to provide a meaningful statistical distribution, which is expensive in terms of computational time. Moreover, this is a purely simulation-based approach and hence provides limited physical insight.

A. Special Case With Only Two Grain Orientations

In this case, two random variables \( X_1 \) and \( X_2 \) represent the number of grains with WF values of \( \Phi_1 \) and \( \Phi_2 \) and probabilities of \( P_1 \) and \( P_2 \), respectively. Since the total number of grains \( (X_1 + X_2) \) is equal to \( N \), these two random variables are not independent; \( (X_2 = N - X_1) \). Therefore, once the distribution of the random variable \( X_1 \) is determined, the distribution of \( X_2 \) can be easily calculated. Since the probability of occurrence of a grain with WF \( \Phi_1 \) is \( P_1 \), the probability of getting exactly \( X_1 = k \) of such grains can be calculated by a “binomial distribution” where the probability density function (pdf) \((f_{X_1}(k))\) is as follows:

\[
f_{X_1}(k) = \binom{N}{k} P_1^k (1 - P_1)^{N-k}, \quad \text{where} \quad \binom{N}{k} = \frac{N!}{k!(N-k)!}. \tag{2}
\]

Knowing that the WF for the metal gate \( (\Phi_M) \) can be modeled with (1), one can write

\[
\Phi_M = \left( \frac{X_1}{N} \right) \Phi_1 + \left( \frac{X_2}{N} \right) \Phi_2 = \Phi_2 + \left( \frac{X_1}{N} \right) (\Phi_1 - \Phi_2). \tag{3}
\]
For example, if we assume a hypothetical gate with $P_1 = 0.6$ and $P_2 = 1 - P_1 = 0.4$, $\Phi_1 = 4.6$ eV, and $\Phi_2 = 4.4$ eV that consists of only one grain ($N = 1$), the WF distribution will be a simple discrete distribution, as shown in Fig. 9(a). This makes sense intuitively because there is a 60% chance of having a gate with $\Phi_M = 4.6$ eV and a 40% likelihood of $\Phi_M = 4.4$ eV. In the case of a metal gate with four grains ($N = 4$), it is possible to identify five distinct combinations for the composition of the metal gate as listed in Table II. The first row of this table represents a situation where all four grains on the metal gate have WF of $\Phi_2$ ($X_1 = k = 0$, $X_2 = N - k = 4$). The likelihood of such a scenario can be calculated by the formula of the binomial distribution (2), as shown in the third column from the left in this table. Moreover, it is possible to determine the WF of the metal gate by using (3), as calculated in the last column. The second row of Table II shows a situation in which three out of the four grains have WF of $\Phi_2$, and one grain has $\Phi_1$ ($X_1 = k = 1$, $X_2 = N - k = 3$), and so on.

The distribution of the metal WF associated with Table II is plotted graphically and shown in Fig. 9(b), where the probability values associated with each case are also shown. Repeating the same procedure for metal gates with higher number of grains, one can obtain the corresponding probability distributions. For instance, Fig. 9(c) and (d) show the distribution functions for the WF of metal gates consisting of 9 and 16 grains, respectively. It should be noted that as the number of grains increases, the shape of the distribution of the WF resembles that of a normal distribution as predicted by “the central limit theorem” [27].

### B. Special Case With Only Three Grain Orientations

In cases where three or more grain orientations are present, one can apply the generalized form of the binomial distribution known as “multinomial distribution” in order to model the distribution of $\Phi_M$. Therefore, the probability of having $X_1 = k_1$, $X_2 = k_2$, and $X_3 = k_3$ is

$$f_X(k_1,k_2,k_3) = \frac{N!}{k_1!k_2!k_3!} P_1^{k_1} P_2^{k_2} P_3^{k_3}$$  \hspace{1cm} (4)

where $k_1 + k_2 + k_3 = N$ and $P_1 + P_2 + P_3 = 1$.

Although (4) provides a formula to calculate the probabilities of $(X_1, X_2, X_3) = (k_1, k_2, k_3)$, it does not provide an analytical closed-form equation for the pdf of the individual random variables $X_1$, $X_2$, and $X_3$ [such as in (2)]. Therefore, the distribution of the WF $\Phi_M$ cannot be obtained in a closed form. However, according to the central limit theorem [27], it is expected that if the number of grains is large enough ($\approx 10$–15), the distribution of $\Phi_M$ (but not those of $X_1$, $X_2$, and $X_3$) is approximately a Gaussian distribution (Fig. 9). This is an important result because any Gaussian distribution (in this case, $\Phi_M$) can be fully identified by its first two moments: the expected value and variance. In order to calculate $E(\Phi_M)$ and $\text{var}(\Phi_M)$ from (1), one needs to initially determine $E(X_i)$, $E(X_i^2)$, and $E(X_i X_j)$ for the random variables $X_1$, $X_2$, and $X_3$.

Fortunately, the expected value $E(X_i)$ and variance $\text{var}(X_i)$ for the individual random variables $X_1$, $X_2$, and $X_3$ are known to be as in (5) and (6), respectively. Also, the covariance value between pairs of unidentical random variables $\text{cov}(X_i, X_j)$, where $i \neq j$, can be expressed as (7)

- $E(X_i) = NP_i$ \hspace{1cm} (5)
- $\text{var}(X_i) = NP_i(1 - P_i)$ \hspace{1cm} (6)
- $\text{cov}(X_i, X_j) = -NP_i P_j$. \hspace{1cm} (7)

By further manipulating (5)–(7), one can obtain the analytical expressions for $E(X_i^2)$ and $E(X_i X_j)$ as follows:

\begin{align}
\text{var}(X_i) &= E[(X_i - E(X_i))^2] = NP_i(1 - P_i) \\
&= E(X_i^2) - E^2(X_i) = NP_i(1 - P_i) \\
&= E(X_i^2)^2 - N^2 P_i^2 = NP_i(1 - P_i) \\
&= E(X_i^2) = NP_i(1 - P_i) + N^2 P_i^2 \hspace{1cm} (8)
\end{align}

\begin{align}
\text{cov}(X_i, X_j) &= E[(X_i - E(X_i))(X_j - E(X_j))] = -NP_i P_j \\
&= E[(X_i X_j - X_i E(X_i) - X_i E(X_j) + E(X_i) E(X_j))] \\
&= -NP_i P_j \\
&= E(X_i X_j) - E(X_i) E(X_j) \\
&= E(X_i X_j) - NP_i P_j = -NP_i P_j \\
&= E(X_i X_j) = N^2 P_i P_j - NP_i P_j. \hspace{1cm} (9)
\end{align}
Once the $E(X_1)$, $E(X_2)$, and $E(X_1X_2)$ values are calculated, (1) can be used to calculate $E(\Phi_M)$ and $\text{var}(\Phi_M)$

$$\Phi_M = \left( \frac{X_1}{N} \right) \phi_1 + \left( \frac{X_2}{N} \right) \phi_2 + \left( \frac{X_3}{N} \right) \phi_3$$

$$→ E(\Phi_M) = \left( \frac{X_1}{N} \right) E(X_1) + \left( \frac{X_2}{N} \right) E(X_2) + \left( \frac{X_3}{N} \right) E(X_3)$$

and using (5), one arrives at

$$E(\Phi_M) = P_1 \phi_1 + P_2 \phi_2 + P_3 \phi_3. \quad (10)$$

To calculate $\text{var}(\Phi_M)$, one can use the definition of $\Phi_M$ and its expected value $E(\Phi_M)$ as follows:

$$\text{var}(\Phi_M) = E(\Phi_M - E(\Phi_M))^2$$

$$→ \text{var}(\Phi_M) = E\left[ \left( \frac{X_1}{N} \right) \phi_1 + \left( \frac{X_2}{N} \right) \phi_2 + \left( \frac{X_3}{N} \right) \phi_3 - (P_1 \phi_1 + P_2 \phi_2 + P_3 \phi_3) \right]^2$$

$$= E\left[ \left( \frac{\phi_1}{N} \right) X_1^2 + \left( \frac{\phi_2}{N} \right) X_2^2 + \left( \frac{\phi_3}{N} \right) X_3^2 + (P_1 \phi_1 + P_2 \phi_2 + P_3 \phi_3)^2 - 2 \left( \frac{\phi_1}{N} \right) \left( \frac{\phi_2}{N} \right) X_1X_2 - 2 \left( \frac{\phi_2}{N} \right) \left( \frac{\phi_3}{N} \right) X_2X_3 - 2 \left( \frac{\phi_3}{N} \right) \left( \frac{\phi_1}{N} \right) X_1X_3 \right]$$

$$+ E\left[ \left( \frac{\phi_1 \phi_2}{N^2} \right) X_1^2 X_2^2 + \left( \frac{\phi_2 \phi_3}{N^2} \right) X_2^2 X_3^2 + \left( \frac{\phi_3 \phi_1}{N^2} \right) X_3^2 X_1^2 \right]$$

$$- E\left[ \left( \frac{\phi_1}{N} \right) (P_1 \phi_1 + P_2 \phi_2 + P_3 \phi_3) X_1 \right]$$

$$+ 2 \left( \frac{\phi_1 \phi_2}{N^2} \right) (P_1 \phi_1 + P_2 \phi_2 + P_3 \phi_3) X_1X_2$$

$$+ 2 \left( \frac{\phi_1 \phi_3}{N^2} \right) (P_1 \phi_1 + P_2 \phi_2 + P_3 \phi_3) X_1X_3$$

After a couple of algebraic simplifications, one can arrive at

$$\text{Var}(\Phi_M) = \frac{1}{N} \left( P_1 \phi_1^2 + P_2 \phi_2^2 + P_3 \phi_3^2 \right) - \frac{1}{N} \left( P_1 \phi_1 + P_2 \phi_2 + P_3 \phi_3 \right)^2. \quad (11)$$

**C. Generalization of the Model for More Than Three Grain Orientations**

Equations (10) and (11) present the expected value and variance of the gate WF if it is composed of metal grains with three different orientations. However, it is possible to generalize (10) and (11) for the occasions when the gate is composed of more than three grain orientations. Equations (12) and (13) express the expected value and variance of gate WF $\Phi_M$ for such cases

$$E(\Phi_M) = \sum_{i=1}^{r} P_i \phi_i \quad (12)$$

$$\text{var}(\Phi_M) = \frac{1}{N} \left( \sum_{i=1}^{r} P_i \phi_i^2 - \left( \sum_{i=1}^{r} P_i \phi_i \right)^2 \right) \quad (13)$$

$$N = (L/G) \times (W/G) \quad (14)$$

where $N$ is the total number of grains on the gate (assuming square-shaped grains with size $G$ on metal gate of length $L$ and width $W$); $r$ is the number of different grain orientations on the gate; $P_i$ and $\phi_i$ represent the probability and WF of the $i$th grain, respectively. Equation (13) is very important as it indicates that the standard deviation of WFV is inversely proportional to the number of grains [which, in turn, is proportional to the area of the metal gate or $W \times L$ (14)]. Therefore, assuming a constant grain size for successive technology nodes,

**TABLE II**

**POSSIBLE VALUES FOR PROBABILITY AND WF OF A METAL GATE CONSISTING OF FOUR GRAINS WITH TWO ORIENTATIONS**

<table>
<thead>
<tr>
<th>$X_1$</th>
<th>$X_2$</th>
<th>Probability (%)</th>
<th>$\Phi_M$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>4</td>
<td>$\binom{4}{0} \cdot (0.6)^0 \cdot (0.4)^4 = 2.56%$</td>
<td>$(\frac{0}{4}) \cdot (4.6)^1 \cdot (\frac{4}{4}) \cdot (4.4)^3 = 4.4$ eV</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>$\binom{4}{1} \cdot (0.6)^1 \cdot (0.4)^3 = 15.36%$</td>
<td>$(\frac{1}{4}) \cdot (4.6)^1 \cdot (\frac{3}{4}) \cdot (4.4)^2 = 4.45$ eV</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>$\binom{4}{2} \cdot (0.6)^2 \cdot (0.4)^2 = 34.56%$</td>
<td>$(\frac{2}{4}) \cdot (4.6)^1 \cdot (\frac{2}{4}) \cdot (4.4)^2 = 4.5$ eV</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>$\binom{4}{3} \cdot (0.6)^3 \cdot (0.4)^1 = 34.56%$</td>
<td>$(\frac{3}{4}) \cdot (4.6)^1 \cdot (\frac{1}{4}) \cdot (4.4)^2 = 4.55$ eV</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>$\binom{4}{4} \cdot (0.6)^4 \cdot (0.4)^0 = 12.96%$</td>
<td>$(\frac{4}{4}) \cdot (4.6)^1 \cdot (\frac{0}{4}) \cdot (4.4)^0 = 4.6$ eV</td>
</tr>
</tbody>
</table>
downsizing of the transistor gate dimensions by $0.7 \times$ will result in $(1/0.72) \approx 2$ times increase in the standard deviation of WFV.

Note that in (14), it is assumed that all grains have identical size, and the average grain size is used for $G$. Assuming a distribution for grain size makes it impossible to obtain a closed-form analytical formula for $\text{var}(\Phi_M)$. However, since the growth condition for all grains are identical, the distribution of grain size is expected to be narrow, and hence, using the average value ($G$) results in negligible error.

### D. Validation of the Model by Monte Carlo Analysis

The experimental measurement of the gate metal WF distribution is very complicated due to the large number of experiments that is required to obtain a meaningful distribution. It should be noted that WF measurement experiments (which are usually performed through $C-V$ measurements) are extremely time consuming, and, since obtaining an accurate distribution requires thousands of such measurements, it is not a practical way to obtain a reference model. Therefore, a computer-generated profile that closely resembles the actual WF distribution has been used in this paper. It should be noted that such an approach has also been used in other works in literature where experimental measurement of the physical properties was impossible. For instance, to evaluate the impact of random gate-oxide-thickness variations on the $V_{th}$ of MOSFETs, it is not practically possible to measure the oxide thickness for each device to obtain the distribution of oxide thickness, and hence, a computer-generated profile is employed to generate the oxide-thickness-variation distribution [28]. Similarly, in this paper, a Monte Carlo method is used to obtain the realistic gate metal WF distribution, which can be used as a reference for validation of the proposed model.

These reference distributions are generated by constructing 100,000 metal gates where the area of each metal gate is assumed to be $65 \text{ nm} \times 65 \text{ nm}$ and is composed of $7 \text{ nm} \times 7 \text{ nm}$ grids. Each cell of the grid represents one metal grain, and its orientation (and hence, WF) is assigned randomly based on the probability of that particular orientation. The overall WF of each metal gate is calculated by averaging the WF of its grains, similar to (1). Since grain orientation is assigned in a purely random fashion, the Monte Carlo method generates a WF distribution that would be similar to the profile obtained from measurements.

Table III provides several experimental data for $\Phi_i$ and $P_i$ values of a number of metal-gate materials, which are reported by various groups. These materials are grown under certain process conditions, and the measurements are performed under various settings. These details and the types of measurement methods used can be obtained from the cited reference papers. Therefore, these WF values can vary under other settings and process conditions. Note that the accuracy of the model proposed in this paper is entirely independent of the WF values used in the calculations.

Fig. 10(a) shows such a reference distribution for the WF of TaN metal gates consisting of three different grain orientations with WFs $\Phi_1 = 4 \text{ eV}$, $\Phi_2 = 4.15 \text{ eV}$, and $\Phi_3 = 4.8 \text{ eV}$ and probabilities $P_1 = 0.5$, $P_2 = 0.3$, and $P_3 = 0.2$, respectively. In this figure, the WF distribution predicted by the proposed model is also shown. As can be observed, the two distributions are almost indistinguishable, which indicates that the model is able to accurately predict the actual WF distribution. The close resemblance between the two distributions is also confirmed by a Q-Q plot [27] in Fig. 10(b). The Q-Q graph is
a plot of the quantiles (fraction or percent of points below a given value) of the Monte Carlo simulation results versus the quantiles obtained from the model. If the two distributions are identical, the Q-Q plot would be a linear line (the blue line). Fig. 10(b) indicates that the Q-Q plot is very close to the linear reference line, suggesting that the two distributions are almost identical.

IV. CONCLUSION

A new source of random $V_{th}$ fluctuation (WFV) in emerging high-$k$/metal-gate devices has been identified and analytically modeled. WFV is caused by the dependence of metal WF on the grains’ orientations. The existence of such random $V_{th}$ variations is validated through experimental measurements. The comprehensive analytical modeling framework developed in this paper indicates that the WF of metal gates exhibits a substantial distribution. The key parameters of such a distribution are calculated assuming that the physical characteristics of the gate (including gate size and metal type) are known. The proposed statistical framework has been verified against rigorous Monte Carlo simulations. According to the comprehensive analyses presented in this paper, WFV is going to be one of the dominant sources of $V_{th}$ variation in emerging nanoscaled CMOS transistors. The proposed framework can be used to investigate the important implications of WFV for process, device, and circuit design, as discussed in the companion paper.

ACKNOWLEDGMENT

The authors (H. F. D and K. B) would like to thank S. H. Rasouli and C. Xu for useful technical discussions and careful review of the manuscript.

REFERENCES


Hamed F. Dadgour (S’05) received the B.S. degree in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 1999 and the and M.S. degree in electrical engineering from the University of Tehran, Tehran, in 2001. He is currently working toward the Ph.D. degree in the Electrical and Computer Engineering Department, University of California, Santa Barbara (UCSB), Santa Barbara, where he is working in the Nanoelectronics Research Laboratory of Prof. Kaustav Banerjee.

His current research is focused on the design and implementation of energy-efficient circuits and systems using emerging nanoscale transistors. He has published several papers in leading international conferences and journals. Mr. Dadgour’s paper introducing a new source of random variability in high-k/metal-gate transistors was a finalist for the IEEE/ACM William J. McCalla ICCAD Best Paper Award in 2008. He was also the recipient of the Award of Distinction from UCSB in 2009 and the Peter J. Frenkel Foundation Fellowship from the Institute for Energy Efficiency at UCSB in 2010.

Kazuhiko Endo (M’99) received the Ph.D. degree in electrical engineering from Waseda University, Tokyo, Japan, in 1999.

He was with Silicon Systems Research Laboratories, NEC Corporation, from 1993 to 2003, where he worked on the research and development of multi-level interconnects and high-k gate-stack technologies for ULSI. From August 1999 to August 2000, he was a visiting scholar at the Center for Integrated Systems, Stanford University, Stanford, CA. Currently, he is a Senior Researcher with the Silicon Nanoscale Devices Group, National Institute of Advanced Industrial Science and Technology, Tsukuba, Japan. His research interests include nanometer-scale manufacturing for aggressively scaled multigate devices in advanced VLSI technologies.

Dr. Endo is a member of the IEEE Electron Devices Society and the Japan Society of Applied Physics. He was the recipient of the Best Paper Award at the 2003 Advanced Metallization Conference and at the 1998 Meeting of Japan Society of Applied Physics.

Vivek K. De (S’89–M’89–SM’07) received the B.S. degree in electrical engineering from the Indian Institute of Technology, Madras, India, in 1985, the M.S. degree in electrical engineering from Duke University, Durham, NC, in 1986, and the Ph.D. degree in electrical engineering from Rensselaer Polytechnic Institute, Troy, NY, in 1992.

He is an Intel Fellow and Director of Circuit Technology Research, Circuits Research Laboratory, Corporate Technology Group, Hillsboro, OR. In his current role, he provides strategic direction for future circuit technologies and is responsible for aligning Intel’s circuit research with technology-scaling challenges. He has published more than 150 technical papers in refereed conferences and journals and six books chapters on low-power circuits. He is the holder of 136 patents, with 57 more patents filed (pending).

Dr. De was the recipient of the Intel Achievement Award for his contributions to a novel integrated voltage-regulator technology.
Kaustav Banerjee (S’92–M’99–SM’03) received the Ph.D. degree in electrical engineering and computer sciences from the University of California, Berkeley, in 1999.

He was a Research Associate with the Center for Integrated Systems, Stanford University, Stanford, CA, from 1999 to 2001. From February to August 2002, he was a Visiting Faculty at the Circuits Research Laboratories, Intel, Hillsboro, OR. He has also held summer/visiting positions with Texas Instruments Incorporated, Dallas, from 1993 to 1997 and with the Swiss Federal Institute of Technology, Lausanne, Switzerland, in 2001. Since July 2002, he has been with the Faculty of the Department of Electrical and Computer Engineering, University of California, Santa Barbara (UCSB), where he has been a Full Professor since 2007. He is also an affiliated Faculty with the California NanoSystems Institute and the Institute for Energy Efficiency at UCSB. He is the author of over 200 journal and refereed international conference papers and several book chapters. He is also a Coeditor of the book *Emerging Nanoelectronics: Life With and After CMOS* (Springer, 2004). His current research interests include nanometer-scale issues in VLSI as well as circuits and system issues in emerging nanoelectronics. He is also involved in exploring the physics, technology, and applications of various carbon nanostructures for ultra energy-efficient electronics and energy harvesting/storage applications.

Dr. Banerjee was the recipient of numerous awards in recognition of his work, including the Best Paper Award at the Design Automation Conference in 2001, the Association of Computing Machinery Special Interest Group on Design Automation Outstanding New Faculty Award in 2004, the IEEE Micro Top Picks Award in 2006, and an IBM Faculty Award in 2008. He has served on the Technical and Organizational Committees of several leading IEEE and ACM conferences, including International Electron Devices Meeting, Design Automation Conference, International Conference on Computer Aided Design, International Reliability Physics Symposium, International Symposium on Quality Electronic Design, the EOS/ESD Symposium, and the International Conference on Simulation of Semiconductor Processes and Devices. From 2005 to 2008, he served as a member of the Nanotechnology Committee of the IEEE Electron Devices Society (EDS). At present, he serves on the IEEE/EDS GOLD Committee and the IEEE/EDS VLSI Circuits and Technology Committee. He has been a Distinguished Lecturer of the IEEE Electron Devices Society since 2008.