Process Technology Variation

Kelin J. Kuhn, *Fellow, IEEE*, Martin D. Giles, *Fellow, IEEE*, David Becher, Pramod Kolar, Avner Kornfeld, Roza Kotlyar, Sean T. Ma, Atul Maheshwari, and Sivakumar Mudanai

***Abstract*—Moore's law technology scaling has improved performance by five orders of magnitude in the last four decades. As advanced technologies continue the pursuit of Moore's law, a variety of challenges will need to be overcome. One of these challenges is the management of process variation. This paper discusses the importance of process variation in modern transistor technology, reviews front-end variation sources, presents device and circuit variation measurement techniques, including circuit and memory data from the 32-nm node, and compares recent intrinsic transistor variation performance from the literature.**

***Index Terms*—Complementary metal–oxide–semiconductor (CMOS), static random access memory (SRAM), variation, V_{ccmin}.**

I. Introduction

ALTHOUGH there has been a trend in the complementary metal–oxide–semiconductor (CMOS) literature in recent years to describe process variation as a new challenge associated with advanced CMOS technologies [1], process variation has always been a critical aspect of semiconductor fabrication [2].

Similar to other critical areas in semiconductor processing, meeting Moore's law scaling with variation challenges requires combining new innovations with established continual improvement strategies. Fig. 1 illustrates this condition by showing static random access memory (SRAM) cell design from 90 nm to 22 nm. The 90-nm SRAM cell was a "tall" architecture with significant diffusion corner rounding. The 65-nm cell added the "wide" cell to eliminate diffusion corners, the 45-nm cell added double patterning to eliminate poly rounding, and the 32- and 22-nm cells refined the basic "wide" design with patterning and polish improvements. This cycle of innovation and continual improvement is a recurring theme in addressing the challenges of variation across technology generations.

Manuscript received September 29, 2010; revised February 7, 2011; accepted February 9, 2011. Date of publication April 25, 2011; date of current version July 22, 2011. The review of this paper was arranged by Editor H. S. Bennett.

K. Kuhn is with the Components Research Group, Intel Corporation, RA3-254, Hillsboro, OR 97124 USA (e-mail: kelin.ptd.kuhn@ intel.com).

M. Giles, R. Kotlyar, S. T. Ma, and S. Mudanai are with the Process Technology Modeling Department, Intel Corporation, RA3-254, Hillsboro, OR 97124 USA.

D. Becher, P. Kolar, A. Kornfeld, and A. Maheshwari are with Logic Technology Development, Intel Corporation, Hillsboro, OR 97124 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TED.2011.2121913

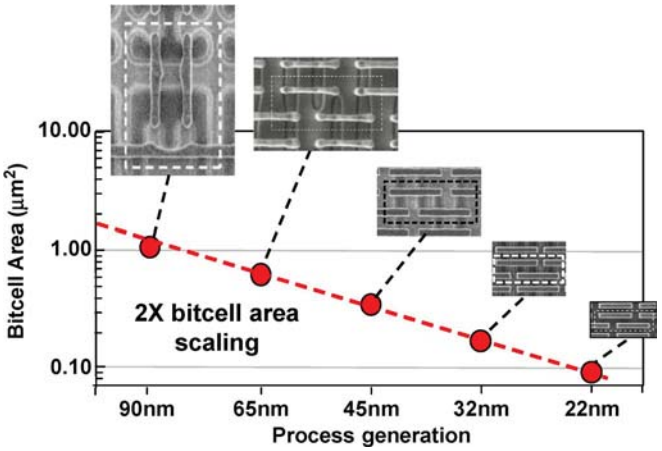

Fig. 1. SRAM cell design from 90 nm to 22 nm illustrates the combination of new innovations and continual improvement strategies for addressing variation challenges while maintaining a 2× area Moore's Law scaling.

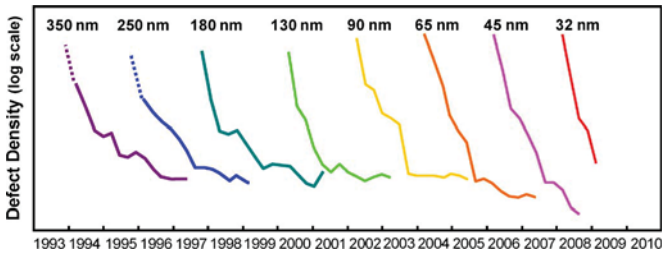

Fig. 2. Improvements in yield on each process generation arising from new innovations and established continual improvement strategies [3], [4].

A. Systematic Variation

Historically, systematic process variation has been of interest to semiconductor manufacturers, because systematic process variation is a strong driver for yield. The management of systematic variation is critical to maintaining competitive yield over multiple technology nodes. A variety of new innovations and established continual improvement strategies are needed to deliver improvements in yield on each process generation; see Fig. 2, [3] and [4]. Process technology developers have met this challenge and continue to deliver high yield at the Moore's law cadence.

B. Random Variation

In recent years, the importance of maintaining low random variation has increased due to the importance of achieving lower minimum operating voltages (V_{ccmin}) for memory elements such as SRAMs and register file arrays. The increased emphasis on lower V_{ccmin} arises from a dramatic increase in

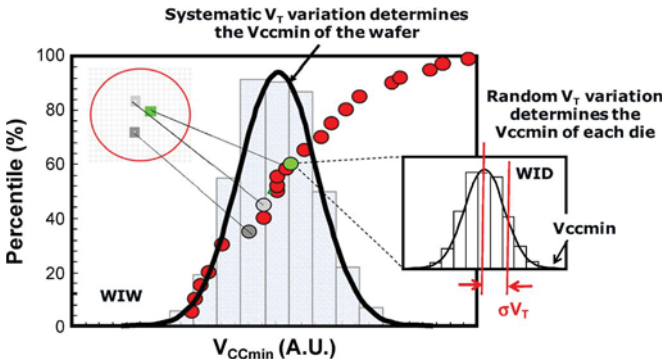

Fig. 3. SRAM minimum operating voltage (V_{ccmin}) is affected by both systematic and random transistor variations [5].

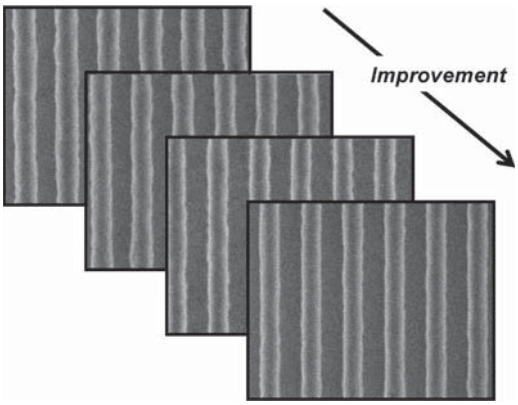

Fig. 4. Sequence of process enhancements for 32 nm LWR/LER illustrating long term variation management through continual improvements.

demand for mobile and ultramobile products, which must meet aggressive average power targets. In an SRAM array (see Fig. 3 and [5]), threshold voltage (V_T) random variation is the most significant contributor for the die V_{ccmin}, whereas systematic V_T variation is the most significant contributor for the wafer V_{ccmin}.

II. Front-End Variation Sources

A. Historical Sources

Variation sources in the CMOS front end (see [6] and references therein) can be categorized into two groups. The first group consists of historical variation sources that will continue to offer challenges moving forward. This group includes patterning proximity effects [both classical and optical proximity correction (OPC) [7]], line-edge roughness (LER) and linewidth roughness (LWR) [8], polish variations [shallow trench isolation (STI) [9] and gate [10]], and variations in the gate dielectric (oxide thickness variations [11], fixed charge [12], and defects and traps [13]). Long-term variation management requires continuing to drive aggressive improvements for these historical sources; see Fig. 4, [14], and [15].

B. Emerging Sources

The second group includes variation sources that were historically of minor impact but have emerged as significant challenges in recent years. This group includes random dopant

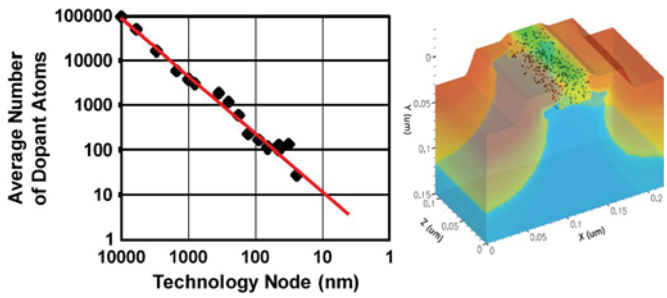

Fig. 5. Reduction in the number of dopant atoms per generation illustrates the increasing importance of RDF as a significant variation source [6], [26].

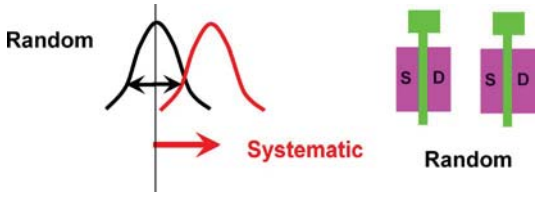

Fig. 6. (Left) Random variation is variation around a mean. (Right) Random variation in devices can be determined by measuring the standard deviation (σ) of the difference between parameters (e.g., V_T and I_D) for two closely spaced identical devices and dividing by $\sqrt{2}$ [14].

fluctuation (RDF; see Fig. 5 and [16]–[18]), variation associated with implants and anneals [pocket implants [19] and rapid-thermal anneal (RTA) [20]], variation associated with strain [wafer-level biaxial [21], high-stress capping layers [22], and embedded silicon–germanium (SiGe) [23]], and variation associated with gate material granularity (poly gates [24] and metal gates [25]). Long-term variation management requires focused effort on understanding these sources so that new innovations and continual improvement strategies can be developed to address them.

III. Measurement of Variation

A. Paired-Device Measurements of Variation

Random variation can be defined as variation around a mean, whereas systematic variation can be defined as the movement of the mean (see Fig. 6). In practice, random variation for a single device has historically been determined by measuring the standard deviation (σ) of the difference between critical parameters (e.g., V_T and I_D) for two closely spaced identical devices and dividing the difference by $\sqrt{2}$. The $\sqrt{2}$ assumes that these distributions are independent (an assumption that can be validated from array data).

One key issue in computing σ is assuring a large-enough sample set for the computation. The concern here is that, for small sample sets, repeated measurements of σ (as might be done in trending), will yield a distribution of σ values. This distribution can be interpreted as process variation but is actually sampling variation. Fig. 7, left, shows the tradeoff between the number of points and σ variation due to sampling. Fig. 7, right, shows how typical trending might vary, depending on the sample size. Although this issue is not common in manufacturing, it is quite common in research, where a

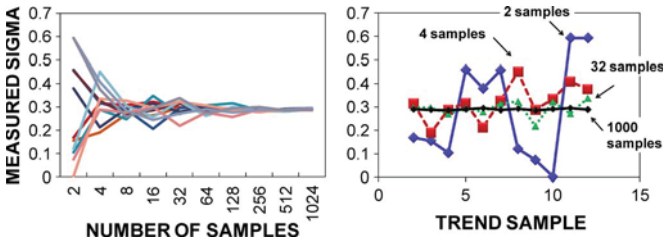

Fig. 7. (Left) Sensitivity of the measured sigma to the number of samples. (Right) Sampling variation can be confused with process variation if the sampling size is very small.

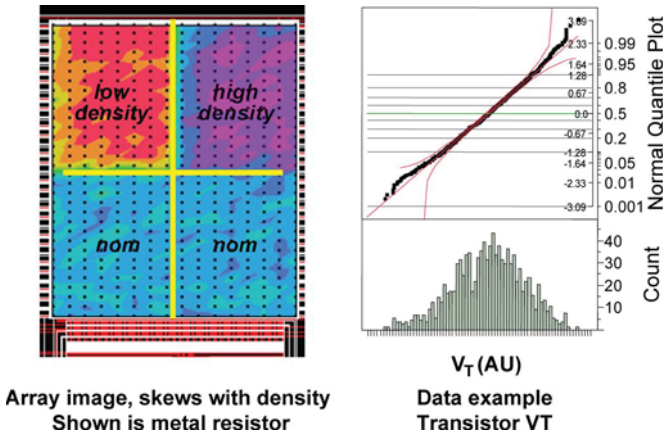

Fig. 8. The value of arrays is that they provide σX values for a single physical test structure. (Left) Use of arrays for density measurements. (Right) V_T measurements from a single array.

single-wafer experiment may not provide sufficient data for accurate σ computation.

Another significant difficulty with the paired-device method is that there may be a systematic variation of the random variation. (As one common example, the random variation measured from matched pairs may systematically be higher at the edge of the wafer than at the center.) Unfortunately, the obvious mitigation (filtering the data) carries the associated risk of dropping the sample size below the level necessary for good statistical sampling.

B. Device Array Measurement of Variation

One way of addressing issues with paired devices is to use measurement arrays (see Fig. 8). A measurement array consists of a large number of identical devices, each individually addressable, such that all parametric measurements (e.g., V_T, I_D, and subthreshold slope) can be done on a large set of closely spaced devices. The size of the array is chosen to be large enough to avoid sampling issues but smaller than all known process systematic effects (typically $100 <$ array size < 1000).

In an array of this type, the process systematic variation is typically assumed to be negligible, and thus, the σX of all the values in the array is assumed to describe the process random variation. However, because the parametric data are available for all the devices, the following two critical assumptions can quantitatively be validated: 1) the systematic variation is negligible and 2) the devices are independent. Array measurements bypass both the differential measurement and the $\sqrt{2}$ assumption of independence of the historical method but involve significantly more test structure complexity and test time.

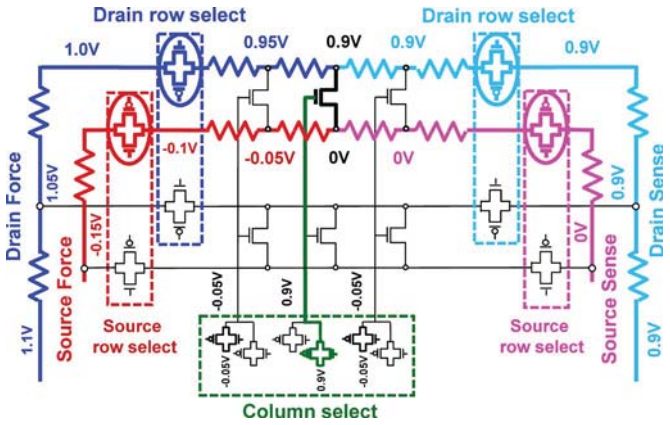

Fig. 9. Measurement configuration for determining debias through the use of force and sense connections on the drain and source.

One concern with arrays is the need to compensate for the systematic debiasing of devices due to the differences in bus length across the array (see Fig. 9). This compensation can be accomplished through feedback techniques such as sensing the voltage applied to the device through redundant connections, but at the expense of added complexity and test time. If logic circuitry is used to activate a specified row and column in an array, it adds the constraint that only wafers with functional logic circuitry can provide useful data from the array. Although it is possible to design an array of devices that does not require logic circuitry or various compensation techniques, such designs come at the expense of significant reduction in array efficiency. In either case, once the price has been paid in terms of die area and test complexity, arrays enable a level of visibility to random variation not possible with standard matched-pair analyses.

C. Ring Oscillator Circuit Measurement of Variation

Another approach for variation measurement is to use ring oscillators (see Figs. 10 and 11, [6], [14], and [26]). By analogy with individual devices, closely spaced ring oscillators (or ring oscillators with interlaced transistors) are used to obtain random variation data, and large populations of oscillators, with random variation removed through root-mean-square (rms) analysis, are used to obtain systematic data.

Ring oscillators need not be limited to simple inverters. As an example, Fig. 12 shows a circuit that can measure V_T variation and can be instantiated multiple times across a die. The circuit consists of an oscillator formed by an odd number of inverters connected in a chain and connected to different kinds of power gates. *MPref* and *MNref* transistors connect the oscillator power supplies to $V_{\rm CC}$ and $V_{\rm SS}$, respectively, and act as a reference measurement. $MN1$ and $MN2$ are the pair of n-type metal–oxide–semiconductor (NMOS) transistors whose V_T variation will be measured by the oscillator. For

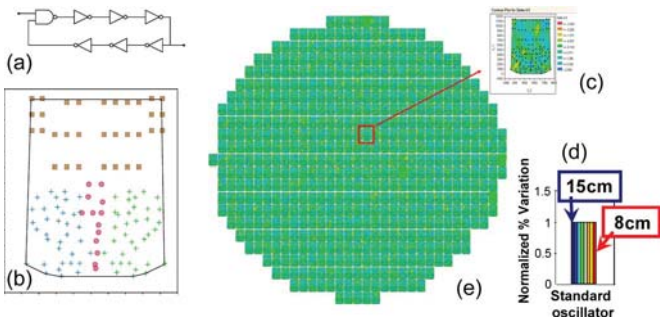

Fig. 10. (a) Ring oscillators can be densely distributed in the (b) product die to provide die-level evaluation of both (c) random and systematic variation (not illustrated), as well as rich statistics for (d) cross-wafer effects and (e) overall maps [6], [14], [26].

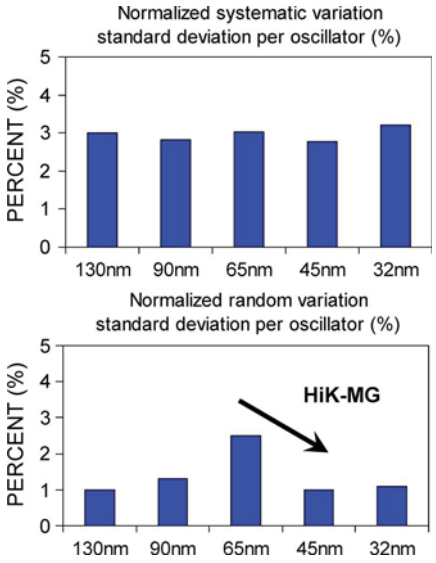

Fig. 11. (Top) Systematic and (bottom) random variation as measured by benchmark ring oscillator circuits for the last five generations. Note that systematic variation is constant (expected if all modules follow Moore's law scaling). In addition, note that random variation is constant, except for 65 nm (the last generation of SiON gate technology, which did not have $T_{\rm ox}$ scaling). HiK-MG in 45 nm restored $T_{\rm ox}$ scaling and a constant trend [14], [15].

NMOS V_T-related measurements, *MNref* is kept ON by setting *EnNref* to $V_{\rm CC}$. This condition makes $Vsvtp$ close to $V_{\rm SS}$. Now, by setting $EnN1$ or $EnN2$ to $V_{\rm CC}$, transistors $MN1$ and $MN2$ are used to power up the CMOS oscillator. Because these transistors are of the NMOS type, node $Vcvtn$ is set at $V_{\rm CC} - V_{\rm TN1}$ and $V_{\rm CC} - V_{\rm TN2}$, respectively, during these measurements, where $V_{\rm TN1}$ and $V_{\rm TN2}$ are, respectively, the threshold voltages of transistors $MN1$ and $MN2$. The resulting frequency is a function of $V_{\rm TN1}$ and $V_{\rm TN2}$ and can be used to estimate device V_T variation. Transistor *MPref* is used to calibrate the oscillator. [The converse structure to measure the p-type metal–oxide–semiconductor (PMOS) V_T is also shown in Fig. 12.]

The structure has some nonidealities. First, the voltage at node $Vcvtn$ is affected by the resistive drop across the transistor. This resistance is a function of transistor characteristics other than threshold voltage. This condition can be mitigated by making the oscillator draw less current or by increasing the size of the devices. (The choice of increasing the device size averages out variation and is less preferable.) Second, note that the threshold voltage drop across the NMOS device $(V_{\rm CC} - V_{\rm TN})$ is not exactly the same threshold voltage as defined by discrete device measurements (as will be discussed in Section IV of this paper). Finally, the oscillator switching injects some noise into the circuit.

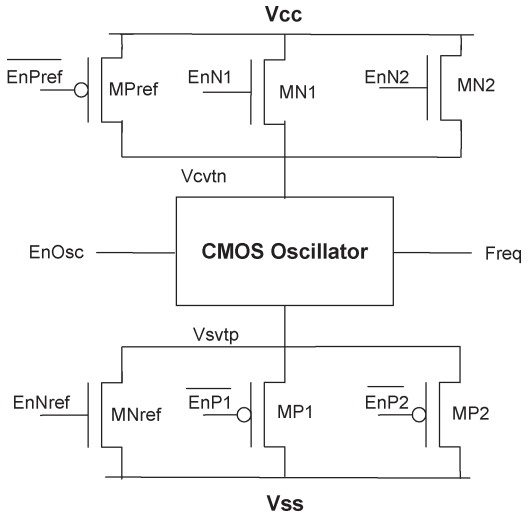

Fig. 12. Circuit for determining NMOS or PMOS V_T variation.

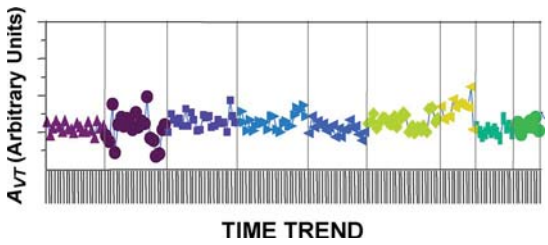

Fig. 13. Data from V_T oscillators similar to Fig. 10 have been used to construct an $A_{\rm VT}$ trend from a full-loop material.

One example of the benefit of this type of circuit is given in Fig. 13, where data from V_T oscillators similar to Fig. 12 have been used to construct an $A_{\rm VT}$ (for the definition of $A_{\rm VT}$, see Fig. 17) trend from a full-loop material. These data can then be compared with full-loop $V_{\rm ccmin}$ or yield data to permit rapid debug of process issues.

One positive feature of ring oscillators is that they can be implemented on the product die, because frequency data can easily be multiplexed out with other end-of-line tests. The value of product implementation is a large increase in the number of samples.

The major negative feature of ring oscillators is the difficulty in correlating the measured frequency variation back to simple parametric terms (e.g., V_T, I_D, and subthreshold slope) or individual process mechanisms (e.g., gate length, mobility, and external resistance). Despite these difficulties, product ring oscillators remain an excellent way of benchmarking overall product variation between technologies, because they represent a simple circuit design that can be implemented generation after generation (see Figs. 11 and 13).

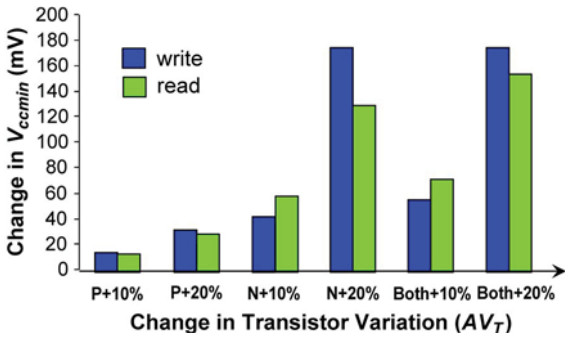

Fig. 14. Simulation of the sensitivity of the die-level $V_{\rm ccmin}$ to changes in $A_{\rm VT}$. Note the increasing sensitivity as the amount of variation is increased.

D. SRAM Modeling and Measurement

The $V_{\rm ccmin}$ of an SRAM cell is critical to product development, particularly in the mobile and handheld markets, where low power is a key requirement. Random variation analyses of static noise margin (SNM), dynamic read stability, and writability can be performed to model the $V_{\rm ccmin}$ of a 6-T SRAM cell in a large cache design [27]. Of particular interest is the sensitivity of the $V_{\rm ccmin}$ to the random variation in V_T (see Fig. 14).

However, although the random σV_T component is still the dominant source of variation when determining the die level $V_{\rm ccmin}$, systematic process variations are also contributors to the wafer-level die $V_{\rm ccmin}$ distribution (see Fig. 3, earlier). Therefore, it is necessary to include both random and systematic terms in modeling to accurately predict the statistical $V_{\rm ccmin}$ distribution at high percentiles [5].

In addition to device variation, cell size and array size determine $V_{\rm ccmin}$ for memory cells. Although it is relatively easy to produce a few small SRAM arrays that operate at low $V_{\rm ccmin}$, the important goal is to produce large arrays that operate at low $V_{\rm ccmin}$. Array density, which includes memory cells, sense amps, and control circuitry, is another important SRAM metric to report. Thus, when measuring and comparing low-voltage memory operation, cell size, $V_{\rm ccmin}$ distribution, array size, and array density must all be considered.

Fig. 15 shows the measured $V_{\rm ccmin}$ distribution for 32-nm-generation SRAM cells for a 3.25-Mb array with cell sizes of 0.171, 0.199, and 0.256 um^2. As expected, the larger cell sizes support smaller $V_{\rm ccmin}$ due to reduced random variation for larger devices. Fig. 16, left, shows the effect of array size on $V_{\rm ccmin}$ for array sizes of 3.25 and 91 Mb for a 0.199-um^2 cell. Fig. 16, right, compares array density, which includes memory cells, sense amps, and control circuitry, and technology node [15].

IV. V_T Benchmarking

A. Analytical Description of Variation

As discussed earlier, random threshold voltage variation (σV_T) is a key factor in determining the $V_{\rm ccmin}$ of memory elements such as SRAMs and register file cells. To benchmark random V_T variation, it is necessary to have an analytical expression for variation as a function of fundamental process parameters.

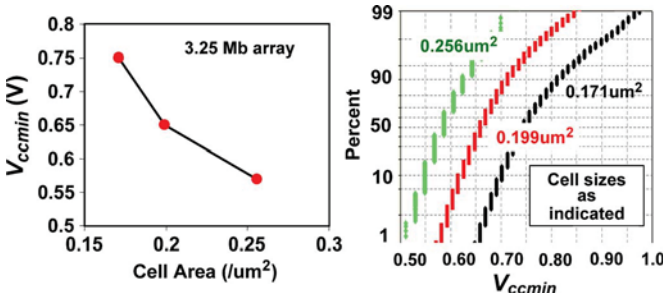

Fig. 15. Measured $V_{\rm ccmin}$ distribution for 32-nm-generation SRAM cells for a 3.25-Mb array with cell sizes of 0.171, 0.199, and 0.256 um^2. Larger cell sizes support smaller $V_{\rm ccmin}$ due to reduced random variation for larger devices [15].

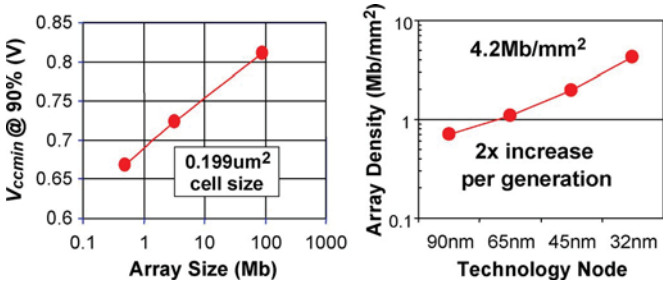

Fig. 16. (Left) Effect of array size on $V_{\rm ccmin}$ for array sizes of 3.25 and 91 Mb for a 0.199-um^2 cell. (Right) Comparison of array density (including memory cells, sense amps, and control circuitry) versus technology node [15].

In the pioneering work of Mizuno *et al.* [28], the analytical expression for σV_T in planar devices due to random dopant fluctuations was shown to be

$$\sigma V_T = \frac{\sqrt[4]{4q^3\varepsilon_{\rm Si}\phi_B}}{2} \cdot \frac{T_{\rm ox}}{\varepsilon_{\rm ox}} \cdot \frac{\sqrt[4]{N_{\rm tot}}}{\sqrt{L_{\rm eff}W_{\rm eff}}} \tag{1}$$

where the key features are a linear dependence on the oxide thickness $T_{\rm ox}$, an inverse square-root dependence on the effective length and width ($L_{\rm eff}$ and $W_{\rm eff}$), and an inverse fourth-root dependency on $N_{\rm tot}$ (where $N_{\rm tot}$ is the total doping concentration per unit volume of the same type of species). An expression of a similar form was shown by Stolk *et al.* [16] with slightly different coefficients as

$$\sigma V_T = \frac{\sqrt[4]{4q^3\varepsilon_{\rm Si}\phi_B}}{\sqrt{3}} \cdot \frac{T_{\rm ox}}{\varepsilon_{\rm ox}} \cdot \frac{\sqrt[4]{N_{\rm tot}}}{\sqrt{L_{\rm eff}W_{\rm eff}}}. \tag{2}$$

For a long-channel double-gate metal–oxide–semiconductor field-effect transistor (MOSFET) with a fully depleted channel, the total space charge in the channel region is determined by the device geometry and channel dopant concentration. This condition leads to an expression for threshold voltage [29], [30] of the form

$$V_T = V_{T0} + qN_{\rm tot}t_{\rm Si}/2C_{\rm ox} \tag{3}$$

where the factor 2 in the denominator counts the two sides of the double-gate device. Generalizing this condition to the case of a multigate device on a silicon-on-insulator (SOI) substrate gives

$$V_T = V_{T0} + q \cdot \left(\frac{T_{\rm ox}}{\varepsilon_{\rm ox}}\right) \cdot N_{\rm tot} \cdot \left(\frac{L_{\rm eff} \cdot W_{\rm si} \cdot H_{\rm si}}{L_{\rm eff} \cdot W_{\rm eff}}\right) \tag{4}$$

where $L_{\rm eff}$ is the effective gate length, $W_{\rm si}$ is the fin width, $H_{\rm si}$ is the fin height, and $W_{\rm eff}$ is the active gate width, given by $W_{\rm eff} = 2H_{\rm si}$ for a FinFET and $W_{\rm eff} = 2H_{\rm si} + W_{\rm si}$ for a Trigate device. If the number of dopant atoms in the channel is the result of random processes, the standard deviation is given by the square root of the number of atoms. The resulting standard deviation in threshold voltage is given by

$$\sigma V_T = q\left(\sqrt{\frac{W_{\rm si}\cdot H_{\rm si}}{W_{\rm eff}}}\right)\cdot\frac{T_{\rm ox}}{\varepsilon_{\rm ox}}\cdot\left(\frac{\sqrt{N_{\rm tot}}}{\sqrt{L_{\rm eff}\cdot W_{\rm eff}}}\right) \quad (5)$$

which further simplifies in the case of FinFETs to

$$\sigma V_T = q\left(\sqrt{\frac{W_{\rm si}}{2}}\right)\cdot\frac{T_{\rm ox}}{\varepsilon_{\rm ox}}\cdot\left(\frac{\sqrt{N_{\rm tot}}}{\sqrt{L_{\rm eff}\cdot W_{\rm eff}}}\right). \quad (6)$$

As with the expression derived earlier by Mizuno and by Stolk, this long-channel model captures variation due to the change in the number of dopant atoms in the channel but does not consider variation due to changes in their position within the channel.

B. Measurement Description and Nuances

The random variation of transistor pairs is determined by measuring the difference in V_T (ΔV_T) between a number of sets of closely spaced paired transistors (e.g., all the transistor pairs on a wafer) and computing the standard deviation of the difference ΔV_T $(\sigma\Delta V_T)$. This paired transistor result is divided by $\sqrt{2}$ to obtain the random σV_T for the individual device as

$$\sigma_{\rm random-pair} = \sigma(V_{\rm TA} - V_{\rm TB}) = \sigma(\Delta V_T) \quad (7)$$

$$\sigma_{\rm random-one-device} = \frac{\sigma(V_{\rm TA} - V_{\rm TB})}{\sqrt{2}} = \frac{\sigma(\Delta V_T)}{\sqrt{2}}. \quad (8)$$

Systematic variation is determined by taking the rms difference between the total and random variation as

$$\sigma_{\rm systematic-one-device} = \sqrt{(\sigma V_{T-\rm pop})^2 - \left(\frac{\sigma(\Delta V_T)}{\sqrt{2}}\right)^2} \quad (9)$$

where the total variation $(\sigma V_{T-\rm pop})$ is determined by taking the standard deviation of the entire population of transistors.

In benchmarking σV_T, an assumption is made of an inverse square-root dependence on $L_{\rm eff}$ and $W_{\rm eff}$, as shown in (1), and the threshold voltage is expressed in terms of a Pelgrom plot [31], where $\sigma\Delta V_T$ is plotted versus $1/\sqrt{L_{\rm eff}W_{\rm eff}}$, and the slope of the resulting line is termed $A_{\rm VT}$ (see Fig. 17).

Although the definition of a Pelgrom plot appears simple, there are a number of nuances in implementation.

1) Methods for the Measurement of V_T: There are a number of measurement techniques for V_T [32]. The simplest (and most commonly used) techniques are the "extrapolation in the linear regime" (ELR; sometimes called the *peak* g_m) and the "constant current" methods. In the ELR method, a linear extrapolation of the I_d–V_g curve at the maximum (positive) slope (i.e., the peak transconductance g_m) is determined, and the zero intercept of this line defined as V_T (see Fig. 18, left).

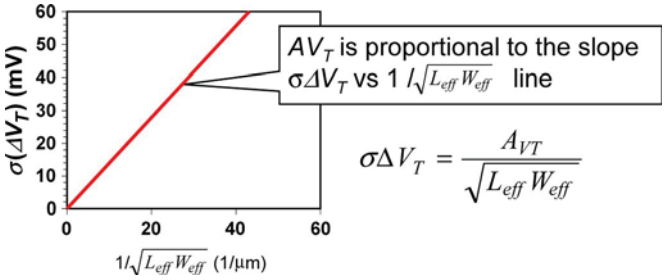

Fig. 17. Pelgrom plot of $\sigma\Delta V_T$ versus $1/\sqrt{L_{\rm eff}\,W_{\rm eff}}$, where the slope of the resulting line is termed AV_T [30].

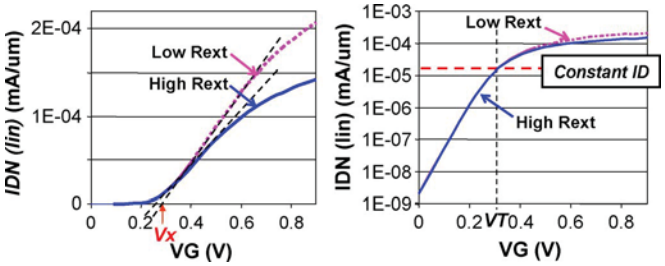

Fig. 18. (Left) ELR method [32] of V_T measurement, showing sensitivity to external parasitic resistance $R_{\rm ext}$. (Right) "Constant current" method [32] of V_T measurement, showing the arbitrary selection of current.

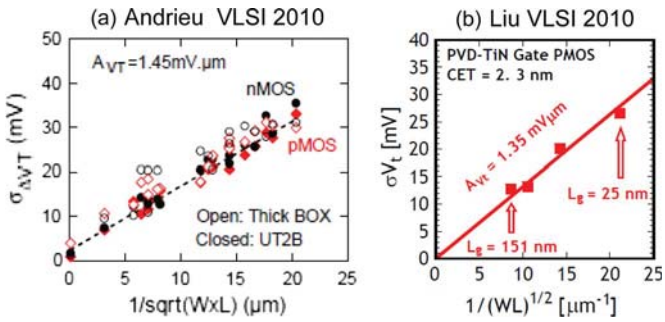

Fig. 19. Left plot (a, [33]) defines AV_T as per Pelgrom (from $\sigma\Delta V_T$). Right-hand plot (b, [34]) defines AV_T from σV_T (a $\sqrt{2}$ difference from plot a).

In the "constant current" method, a current value is selected, and V_T is the gate voltage at that current (see Fig. 18, right). Although the ELR method bypasses the requirement for an arbitrary current choice, the method is sensitive to both external parasitic resistance $R_{\rm ext}$ (a higher $R_{\rm ext}$ gives a lower V_T; see Fig. 18) and mobility (mobility changes the peak value of g_m). In contrast, the "constant current" method is insensitive to both $R_{\rm ext}$ and mobility (g_m) changes but is sensitive to changes in the subthreshold slope (as might occur with poor dielectrics with high $D_{\rm it}$).

2) σV_T (or $\sigma\Delta V_T$) Definition of $A_{\rm VT}$: In the translation between Pelgrom's historical notation (derived for analog devices) and modern CMOS analysis, an alternative plot of σV_T (rather than $\sigma\Delta V_T$) versus $1/\sqrt{L_{\rm eff}W_{\rm eff}}$ has become common, where the slope is also (unfortunately) termed $A_{\rm VT}$ (see Fig. 19). This case has created the situation that reported that $A_{\rm VT}$ benchmark numbers may differ by $\sqrt{2}$, depending on the background of the author.

3) Zero Intercept of the $A_{\rm VT}$ Line: If the measured σV_T (or $\sigma\Delta V_T$) values are extrapolated to the axis, the intercept may

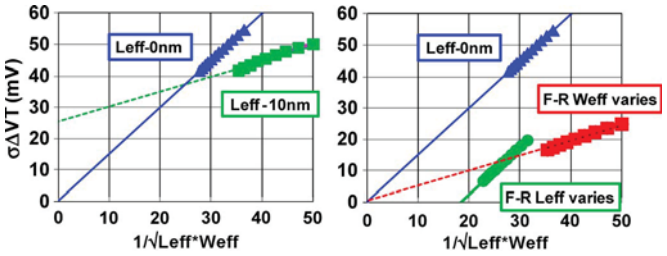

Fig. 20. (Left) It is common to see a nonzero intercept in a Pelgrom plot of small devices when incorrect dimensions have been used for $L_{\rm eff}$ or $W_{\rm eff}$. (Right) F–R measurements with $L_{\rm eff}$ varied will typically have a nonzero x-axis intercept, whereas measurements with $W_{\rm eff}$ will typically have a conventional appearance.

not be zero. A nonzero intercept carries the physical meaning that, for very large-gate-area devices, there is still random V_T variation between devices. However, in a healthy process, infinitely large devices are expected to identically match, and thus, a nonzero intercept is a clue to some potential anomaly in the data.

The most common origin of a nonzero intercept is incorrectly assigning values for $L_{\rm eff}$ and $W_{\rm eff}$. $L_{\rm eff}$ and $W_{\rm eff}$ should correspond to the electrical channel area of the device (as opposed to the drawn or physical dimensions). It is common to see a nonzero intercept in a Pelgrom plot when the data set is from relatively small devices (such as SRAM cells) and incorrect dimensions have been used for $L_{\rm eff}$ and $W_{\rm eff}$ (see Fig. 20, left).

4) Normalization With $T_{\rm ox}$ and Doping Differences: The following two other nomenclatures have also been proposed: 1) $B_{\rm VT}$ and 2) $C_{\rm VT}$. $B_{\rm VT}$ was suggested by Takeuchi [17] as a way of normalizing out the contribution of the gate oxide thickness as

$$\sigma V_T = B_{\rm VT}\sqrt{\frac{T_{\rm inv}(V_T + 0.1\ {\rm V})}{L_{\rm eff}W_{\rm eff}}} \quad \text{and} \tag{10}$$

$C_{\rm VT}$ was suggested by Putra [35] as a normalization that includes the effective doping profile and is given by

$$\sigma V_T = C_{\rm VT}\sqrt{\frac{q}{\varepsilon_{\rm ox}}}\cdot\frac{T_{\rm inv}\sqrt{N_{\rm eff}W_d}}{\sqrt{L_{\rm eff}W_{\rm eff}}}. \tag{11}$$

$B_{\rm VT}$ and $C_{\rm VT}$ represent interesting new metrics (particularly for the comparison of cross-generational or cross-site data) but are not yet widely used.

5) F–R Measurement Techniques: Additional information can be obtained from measuring the $\sigma\Delta V_T$ between forward and reverse (F–R) measurements on the same device. The rationale for this F–R measurement is that a single device has identical structural properties on each measurement (typically assumed to be $L_{\rm eff}$, $W_{\rm eff}$, and $T_{\rm ox}$), and thus, an F–R measurement should highlight variations in channel doping. The qualitative argument [36] is that the drain voltage V_d depletes channel impurities on the drain edge, whereas V_T is controlled by ionized channel impurities in the local region on the source edge. Therefore, comparing the symmetry of linear and saturated V_T [and drain-induced barrier lowering (DIBL)] should give insight into the doping profile in the channel [37], [38].

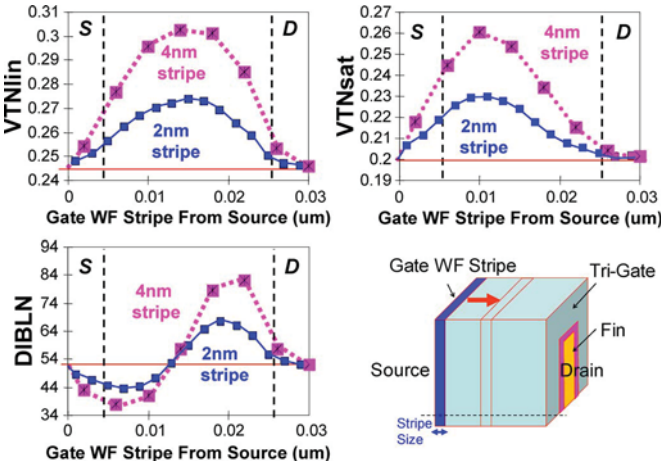

Fig. 21. Three-dimensional simulation of an undoped double-gate device with a midgap workfunction stripe scanned across the channel. When the device is biased in the linear region, the barrier is near the center of the channel, giving symmetric linear threshold voltage ($V_{\rm Tlin}$) performance. When the device is biased into the saturation region, the barrier moves toward the source side, creating an asymmetry shown in both saturated threshold voltage ($V_{\rm Tsat}$) and DIBL.

Note that $\sigma\Delta V_T$ for F–R measurements may also be plotted on a Pelgrom plot (see Fig. 20, right). In this case, (1)–(5) do not strictly apply, because there is an additional length dependency associated with the asymmetric source–drain depletion effect. Thus, the Pelgrom plot will look conventional (although typically of smaller value) if the length is held constant and the width is varied. However, the Pelgrom plot will show a nonzero x-axis intercept if the width is held constant and the length is varied.

Care must be taken in interpreting the F–R methodology for experimental data, because it implicitly assumes that RDF is the only contributor to local variation in a single device. (In fact, this measurement has been proposed as a metric for pure-RDF variation in a device [39].) The issue is that other sources of variation may also be sensitive to source–drain potential differences. As an example, workfunction may also vary across a device. Three-dimensional simulation of an undoped double-gate device with a midgap workfunction stripe scanned across the channel can be used to illustrate a similar effect for a situation with negligible RDF. When the device is biased in the linear region, the barrier is near the center of channel, giving symmetric $V_{\rm Tlin}$ performance. When the device is biased into the saturation region, the barrier moves toward the source side, creating an asymmetry shown in both $V_{\rm Tsat}$ and DIBL (see Fig. 21).

6) Nonplanar Devices: Some difficulties may occur when communicating $L_{\rm eff}$ and $W_{\rm eff}$ for nonplanar devices. The following two problems commonly occur: 1) the normalization per total electrical area versus the normalization per footprint area (see Fig. 22) and 2) the normalization for devices where current conduction is potentially at the center of the device rather than at the surface (i.e., nanowires).

For the first point, the total electrical area of a nonplanar device may significantly be larger than the total footprint area. One example is given in Figs. 22 and 23, where the total electrical area is given by $L_{\rm eff} * W_{\rm eff} = L_{\rm eff} * 4 * 8\ {\rm nm} * 3(wires) * 3(stacks)$, whereas the total footprint is given by

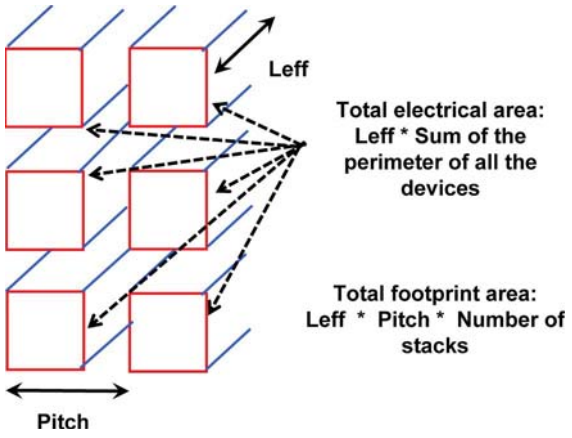

Fig. 22. Total electrical area versus the total footprint area.

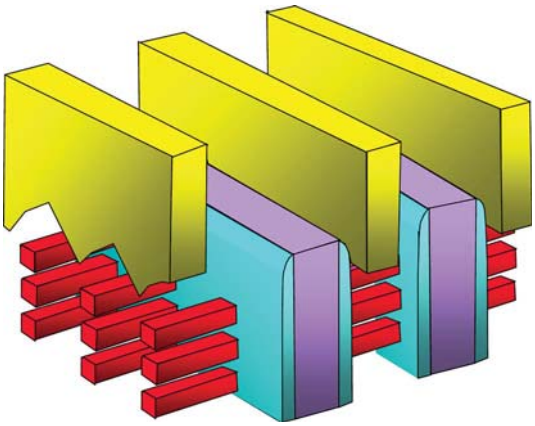

Fig. 23. Idealized nanowire device with square 8 nm × 8 nm wires, three stacks of three high wires, and a pitch of 50 nm between wire stacks.

$L_{\rm eff} * W_{\rm eff} = L_{\rm eff} * 3(stacks) * pitch$. If the pitch is 50 nm and $L_{\rm eff} = 50$ nm, then the total electrical area is $\sim 2\times$ the footprint area.

Based on the analyses in (1)–(5), we expect the variation to scale as the total electrical area. Thus, from the standpoint of a device engineer, the standard normalization should be to the total electrical area, because this condition permits comparison between device types. However, from the standpoint of a circuit designer, normalizing to the total electrical area is unhelpful, because the designer cares about footprint dimensions (because they affect product density and die cost) and not the total electrical area. In other words, from a designer's perspective, a nonplanar device folds a 3-D space (with potential variation improvements due to a larger total electrical area) into a 2-D footprint area. As a consequence, care must be taken to use the correct variation normalization when performing $V_{\rm ccmin}$ simulations on nonplanar circuits.

For the second point, there is the issue of the normalization of V_T as a function of carrier conduction. Wide nonplanar inversion-mode Si devices conduct current on the surface of the device (see Fig. 24, top), and thus, both drive current and variation can be scaled to the total electrical width ($W_{\rm eff} = 2 * Width + 2 * Height$). Smaller nonplanar devices conduct current at the center of the device (see Fig. 24, bottom), and thus, the drive current is frequently normalized to the diameter

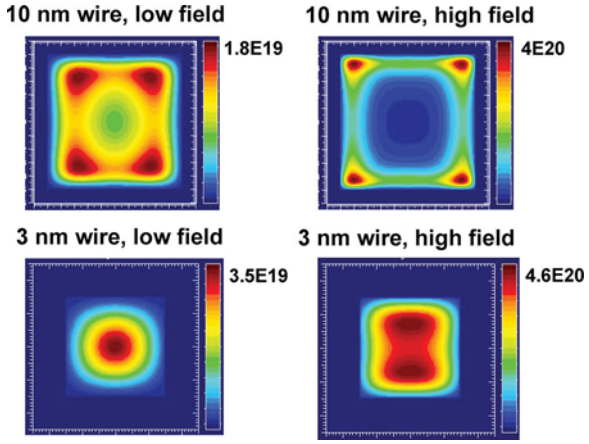

Fig. 24. Small-diameter nanowire devices conduct at the center of the wire rather than at the perimeter. W-normalization for drive current shifts to the diameter (rather than the perimeter) but the normalization for variation remains to the perimeter.

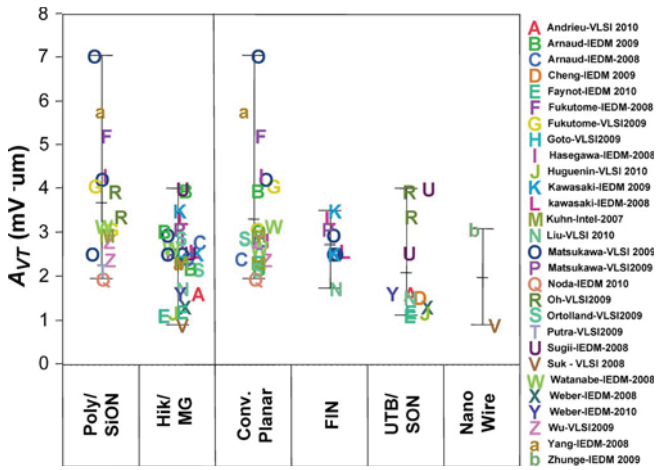

Fig. 25. Comparison of recent $A_{\rm VT}$ values, plotted per technology structure and per year [22], [33]–[35], [39]–[54].

rather than the perimeter. Given that the normalization for small wires is to the diameter, it would be tempting to also normalize random variation to the diameter. This approach is not correct, because random variation is a capacitive-like phenomenon [see (1)–(5)] and is thus controlled by the total electrical area, independent of the location of the charge center.

C. $A_{\rm VT}$ Literature Benchmarking

A comparative plot of measured threshold voltage variation (corrected for the definition of $A_{\rm VT}$ as appropriate; see earlier discussion) as a function of device type and key technology parameters for recent reported data is shown in Fig. 25. Fig. 26 illustrates the percentage variance due to RDF (estimated based on the device type and the probable channel doping) based on the following equation:

$$V_{\rm extra} = \frac{\left(A_{\rm VT(measured)}\right)^2 - \left(A_{\rm VT(RDF-only)}\right)^2}{\left(A_{\rm VT(measured)}\right)^2}. \quad (12)$$

This plot only includes data for which a Pelgrom plot was provided. No attempt was made to reconcile different

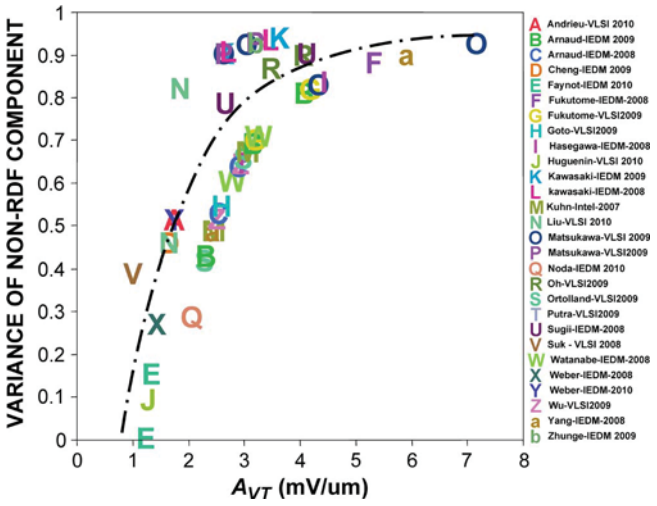

Fig. 26. Same data as shown in Fig. 25 but plotted as a function of non-RDF variance (providing a measure as to the amount of non-RDF variation).

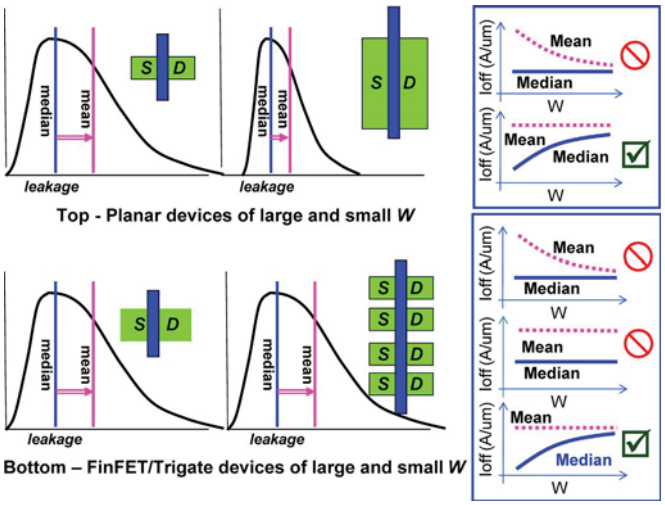

Fig. 27. FinFET devices stimulate new questions about the nature of $I_{\rm off}$ variation as a function of W.

normalization strategies for $L_{\rm eff}$ and $W_{\rm eff}$ (the plotted value for $1/\sqrt{LW}$ was directly used). Data from F–R measurements on a single device was not included. Data with an excessively nonzero intercept for $A_{\rm VT}$ (suggesting a normalization issue with $L_{\rm eff}$ $W_{\rm eff}$) were not included.

Key conclusions from this summary are 1. HiK delivers better variation than poly/SiON, i.e., 2. The record-breaking devices are high-k with ultrathin body or nanowire architectures.

V. Off-State Leakage Variation

A. Off-State Leakage Variation With Device Width W

The unique challenge of the OFF-state leakage $I_{\rm off}$ variation (versus all other types of transistor parametric variations) is that $I_{\rm off}$ follows a log-normal distribution rather than a normal distribution. In modeling $I_{\rm off}$ variation, it is common to assume that the base simulation model reflects the median leakage of the device, with additional random variation δL (as a function of W) to obtain the mean leakage. With this set of modeling assumptions, in the absence of any narrow-W effects, the median leakage variation would be flat, and the mean leakage variation would increase as W decreases (as shown in Fig. 27, top). This case poses a challenge for segmented diffusion devices (such as FinFET or Tri-gate devices), because it suggests that segmented devices of large effective W (which comprised several small W devices) would have an increase in the mean leakage at large W (see Fig. 27, bottom).

Looking at this question more deeply suggests a flaw in the reasoning due to the assumption of a flat median distribution. In the absence of any narrow-W effects, the central limit theorem (CLT) predicts that combining independent log-normal samples will result in a distribution with a flat *mean* and a *decreasing median* with decreasing W (in contrast with the aforementioned typical model, which predicts a flat *median* and an *increasing mean* with decreasing W; see Fig. 28).

The physical assumptions underlying the CLT result are that each of the segments is independent. In the case of a FinFET or a Tri-gate device, it is intuitively reasonable to assert that each fin is independent of the other fins (although an interesting subquestion appears about whether fins under the same gate are different from fins under different gates). However, it is less intuitively reasonable that planar devices act as if they comprised several small independent segments.

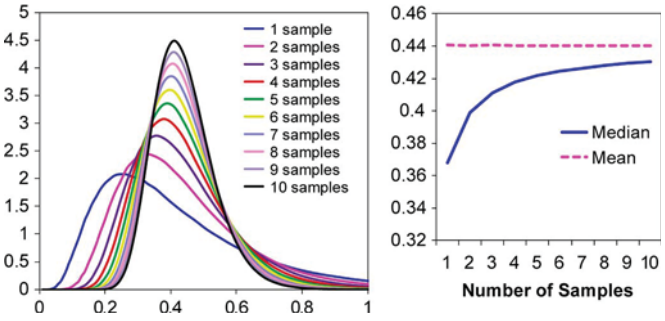

Fig. 28. In the absence of any narrow-W effects, the CLT predicts that combining independent log-normal samples will result in a distribution with a flat mean and a decreasing median with decreasing W (contrasting with the typical simulation model that predicts a flat median and an increasing mean with decreasing W).

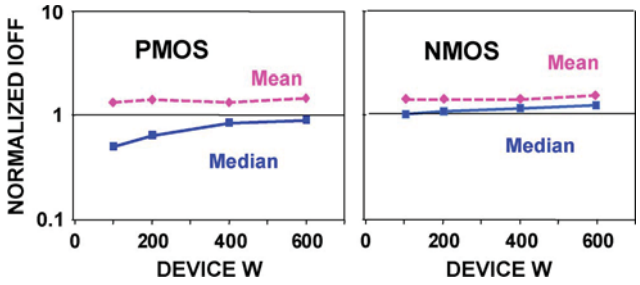

Fig. 29. A population of 32-nm planar devices has a flat mean and decreasing median with decreasing W as suggested by the CLT analysis of Fig. 26. In other words, the OFF-state leakage $I_{\rm off}$ variation of 32-nm planar devices behaves as if constructed from numerous independent small-W devices, suggesting no significant increase in $I_{\rm off}$ with the implementation of segmented diffusion devices such as FinFETs or Tri-gates.

We examined ~4000 die from the 32-nm process to determine if the planar 32-nm device followed the CLT prediction in Fig. 28. The results (shown in Fig. 29) suggest that planar

data behave similar to the CLT expectation. Physically speaking, this case suggests that planar devices behave as if they are constructed from numerous independent small-W devices. This result is important, because it suggests that no significant increase in $I_{\rm off}$ variation for large W devices will occur with the implementation of FinFET and Tri-gate devices.

VI. Conclusion

Similar to other critical areas in semiconductor processing, meeting Moore's law scaling in the presence of variation challenges requires combining new innovations with established continual improvement strategies. Yield data and random and systematic measurements of ring oscillators have shown that Moore's law scaling is maintained for recent generations through these efforts. Benchmarking variation data from recent publications have further shown that improvements are possible through innovations such as new transistor architectures and continual improvement in dielectrics and gate materials.

References

[1] S. Borkar, "Designing reliable systems from unreliable components: The challenges of transistor variability and degradation," *IEEE Micro.*, vol. 26, no. 6, pp. 10–16, Nov./Dec. 2005.

[2] W. Shockley, "Problems related to p–n junctions in silicon," *Solid State Electron.*, vol. 2, no. 1, pp. 35–67, Jan. 1961.

[3] M. Bohr, private communication, 2009.

[4] K. Kuhn, "Moore's law past 32 nm: Future challenges in device scaling," presented at the Structures, Structural Dynamics, and Materials Conference (SSDM), Sendai, Japan, Oct. 7, 2009.

[5] Y. Wang, U. Bhattacharya, F. Hamzaoglu, P. Kolar, Y. Ng, L. Wei, Y. Zhang, K. Zhang, and M. Bohr, "A 4.0-GHz 291-Mb voltage-scalable SRAM design in 32-nm high-κ metal-gate CMOS with integrated power management," in *Proc. IEEE Int. Solid-State Circuits Conf., Dig. Tech. Papers*, 2009, pp. 456–457.

[6] K. Kuhn, C. Kenyon, A. Kornfeld, M. Liu, A. Maheshwari, W.-K. Shih, S. Sivakumar, G. Taylor, P. VanDerVoorn, and K. Zawadzki, "Managing process variation in Intel's 45-nm CMOS technology," *Intel Tech. J.*, vol. 12, no. 2, pp. 93–110, Jun. 2008.

[7] L. Capodieci, "From optical proximity correction to lithography-driven physical design (1996–2006): 10 years of resolution enhancement technology and the roadmap enablers for the next decade," in *Proc. SPIE*, 2006, vol. 6154, p. 615 401.

[8] A. Asenov, S. Kaya, and A. R. Brown, "Intrinsic parameter fluctuations in decananometer MOSFETs introduced by gate line edge roughness," *IEEE Trans. Electron Devices*, vol. 50, no. 5, pp. 1254–1260, May 2003.

[9] S. Nag, A. Chatterjee, K. Taylor, I. Ali, S. O'Brien, S. Aur, J. D. Luttmer, and I. C. Chen, "Comparative evaluation of gap-fill dielectrics in shallow trench isolation for sub-0.25-μm technologies," in *IEDM Tech. Dig.*, Dec. 1996, pp. 841–845.

[10] J. Steigerwald, "Chemical mechanical polish: The enabling technology," in *IEDM Tech. Dig.*, Dec. 2008, pp. 37–40.

[11] M. Koh, W. Mizubayashi, K. Iwamoto, H. Murakami, T. Ono, M. Tsuno, T. Mihara, K. Shibahara, S. Miyazaki, and M. Hirose, "Limit of gate oxide thickness scaling in MOSFETs due to apparent threshold voltage fluctuation introduced by tunneling leakage current," *IEEE Trans. Electron Devices*, vol. 48, no. 2, pp. 259–264, Feb. 2001.

[12] V. S. Kaushik, B. J. O'Sullivan, G. Pourtois, N. Van Hoornick, A. Delabie, S. Van Elshocht, W. Deweerd, T. Schram, L. Pantisano, E. Rohr, L.-A. Ragnarsson, S. De Gendt, and M. Heyns, "Estimation of fixed charge densities in hafnium–silicate gate dielectrics," *IEEE Trans. Electron Devices*, vol. 53, no. 10, pp. 2627–2633, Oct. 2006.

[13] H. C. Wen, H. R. Harris, C. D. Young, H. Luan, H. N. Alshareef, K. Choi, D.-L. Kwong, P. Majhi, G. Bersuker, and H. Lee, "On oxygen deficiency and fast transient charge-trapping effects in high-k dielectrics," *IEEE Electron Device Lett.*, vol. 27, no. 12, pp. 984–987, Dec. 2006.

[14] K. Kuhn, "Variation in 45 nm and implications for 32 nm and beyond," presented at the Int. Conf. CMOS Variability, London, U.K., May, 2009.

[15] P. Packan, S. Akbar, M. Armstrong, M. Brazier, H. Deshpande, K. Dev, G. Ding, T. Ghani, O. Golonzka, W. Han, J. He, R. Heussner, R. James, J. Jopling, C. Kenyon, S.-H. Lee, M. Liu, S. Lodha, B. Mattis, A. Murthy, L. Neiberg, J. Neirynck, S. Pae, C. Parker, L. Pipes, J. Sebastian, J. Seiple, B. Sell, S. Sivakumar, B. Song, A. St. Amour, K. Tone, T. Troeger, C. Weber, and K. Zhang, "High performance 32-nm logic technology featuring 2nd-generation high-$k+$ metal gate transistors," in *IEDM Tech. Dig.*, Dec. 2009, pp. 659–662.

[16] P. Stolk, F. Widdershoven, and D. Klaassen, "Modeling statistical dopant fluctuations in MOS transistors," *IEEE Trans. Electron Devices*, vol. 45, no. 9, pp. 1960–1971, Sep. 1998.

[17] K. Takeuchi, T. Fukai, T. Tsunomura, A. T. Putra, A. Nishida, S. Kamohara, and T. Hiramoto, "Understanding random threshold voltage fluctuation by comparing multiple fabs and technologies," in *IEDM Tech. Dig.*, Dec. 2007, pp. 467–470.

[18] A. Asenov, "Random dopant induced threshold voltage lowering and fluctuations in sub-0.1-μm MOSFETs: A 3-D 'atomistic simulation study'," *IEEE Trans. Electron Devices*, vol. 45, no. 12, pp. 2505–2513, Dec. 1998.

[19] T. Tanaka, T. Usuki, T. Futatsugi, Y. Momiyama, and T. Sugii, "$V_{\rm th}$ fluctuation induced by statistical variation of pocket dopant profile," in *IEDM Tech. Dig.*, Dec. 2000, pp. 271–274.

[20] I. Ahsan, N. Zamdmer, O. Glushchenkov, R. Logan, E. J. Nowak, H. Kimura, J. Zimmerman, G. Berg, J. Herman, E. Maciejewski, A. Chan, A. Azuma, S. Deshpande, B. Dirahoui, G. Freeman, A. Gabor, M. Gribelyuk, S. Huang, M. Kumar, K. Miyamoto, D. Mocuta, A. Mahorowala, E. Leobandung, H. Utomo, and B. Walsh, "RTA-driven intradie variations in stage delay and parametric sensitivities for 65-nm technology," in *VLSI Symp. Tech. Dig.*, Jun. 2006, pp. 170–171.

[21] Y. L. Tsang, S. Chattopadhyay, S. Uppal, E. Escobedo-Cousin, H. K. Ramakrishnan, S. H. Olsen, and A. G. O'Neill, "Modeling of the threshold voltage in strained Si/$Si_{1-x}Ge_x$/$Si_{1-y}Ge_x$ CMOS architectures," *IEEE Trans. Electron Devices*, vol. 54, no. 11, pp. 3040–3048, Nov. 2007.

[22] O. Weber, O. Faynot, F. Andrieu, C. Buj-Dufournet, F. Allain, P. Scheiblin, J. Foucher, N. Daval, D. Lafond, L. Tosti, L. Brevard, O. Rozeau, C. Fenouillet-Beranger, M. Marin, F. Boeuf, D. Delprat, K. Bourdelle, B.-Y. Nguyen, and S. Deleonibus, "High immunity to threshold voltage variability in undoped ultrathin FDSOI MOSFETs and its physical understanding," in *IEDM Tech. Dig.*, Dec. 2008, pp. 245–248.

[23] L.-T. Pang, K. Qian, C. J. Spanos, and B. Nikolic, "Measurement and analysis of variability in 45-nm strained-Si CMOS technology," *IEEE J. Solid-State Circuits*, vol. 44, no. 8, pp. 2233–2243, Aug. 2009.

[24] A. R. Brown, G. Roy, and A. Asenov, "Poly-Si-gate-related variability in decananometer MOSFETs with conventional architecture," *IEEE Trans. Electron Devices*, vol. 54, no. 11, pp. 3056–3063, Nov. 2007.

[25] Y. Zhang, J. Li, M. Grubbs, M. Deal, B. Magyari-Kope, B. M. Clemens, and Y. Nishi, "Physical model of the impact of metal grain work function variability on emerging dual metal gate MOSFETs and its implication for SRAM reliability," in *IEDM Tech. Dig.*, Dec. 2009, pp. 57–60.

[26] K. J. Kuhn, "Reducing variation in advanced logic technologies: Approaches to process and design for manufacturability of nanoscale CMOS," in *IEDM Tech. Dig.*, Dec. 2007, pp. 471–474.

[27] E. Seevinck, F. List, and J. Lohstroh, "Static-noise margin analysis of MOS SRAM cells," *IEEE J. Solid-State Circuits*, vol. 22, no. 5, pp. 748–754, Oct. 1987.

[28] T. Mizuno, J.-I. Okamura, and A. Toriumi, "Experimental study of threshold voltage fluctuation due to statistical variation of channel dopant number in MOSFETs," *IEEE Trans. Electron Devices*, vol. 41, no. 11, pp. 2216–2221, Nov. 1994.

[29] M. H. Chiang, J.-N. Lin, K. Kim, and C.-T. Chuang, "Random dopant fluctuation in limited-width FinFET technologies," *IEEE Trans. Electron Devices*, vol. 54, no. 8, pp. 2055–2060, Aug. 2007.

[30] H. K. Lim and J. G. Fossum, "Threshold voltage of thin-film silicon-on-insulator (SOI) MOSFETs," *IEEE Trans. Electron Devices*, vol. 30, no. 10, pp. 1244–1251, Oct. 1983.

[31] M. J. M. Pelgrom, A. C. J. Duinmaijer, and A. P. G. Welbers, "Matching properties of MOS transistors," *IEEE J. Solid-State Circuits*, vol. 24, no. 5, pp. 1433–1439, Oct. 1989.

[32] A. Ortiz-Conde, F. J. García Sánchez, J. J. Liou, A. Cerdeira, M. Estrada, and Y. Yue, "A review of recent MOSFET threshold voltage extraction methods," *Microelectron. Reliab.*, vol. 42, no. 4/5, pp. 583–596, Apr./May 2002.

[33] F. Andrieu, O. Weber, J. Mazurier, O. Thomas, J.-P. Noel, C. Fenouillet-Beranger, J.-P. Mazellier, P. Perreau, T. Poiroux, Y. Morand, T. Morel, S. Allegret, V. Loup, S. Barnola, F. Martin, J.-F. Damlencourt, I. Servin, M. Casse, X. Garros, O. Rozeau, M.-A. Jaud, G. Cibrario, J. Cluzel, A. Toffoli, F. Allain, R. Kies, D. Lafond, V. Delaye, C. Tabone, L. Tosti, L. Brevard, P. Gaud, V. Paruchuri, K. K. Bourdelle, W. Schwarzenbach, O. Bonnin, B.-Y. Nguyen, B. Doris, F. Buf, T. Skotnicki, and O. Faynot,

"Low-leakage and low-variability ultrathin body and buried oxide (UT2B) SOI technology for 20-nm low-power CMOS and beyond," in *VLSI Symp. Tech. Dig.*, Jun. 2010, pp. 57–58.

[34] Y. X. Liu, K. Endo, S. O'uchi, T. Kamei, J. Tsukada, H. Yamauchi, Y. Ishikawa, T. Hayashida, K. Sakamoto, T. Matsukawa, A. Ogura, and M. Masahara, "On the gate-stack origin threshold voltage variability in scaled FinFETs and multi-FinFETs," in *VLSI Symp. Tech. Dig.*, Jun. 2010, pp. 101–102.

[35] A. T. Putra, T. Tsunomura, A. Nishida, S. Kamohara, K. Takeuchi, S. Inaba, K. Terada, and T. Hiramoto, "A new methodology for evaluating VT variability considering dopant depth profile," in *VLSI Symp. Tech. Dig.*, Jun. 2009, pp. 116–117.

[36] T. Tanaka, T. Usuki, Y. Momiyama, and T. Sugii, "Direct measurement of $V_{\rm th}$ fluctuation caused by impurity positioning," in *VLSI Symp. Tech. Dig.*, Jun. 2010, pp. 136–137.

[37] H. S. Wong and Y. Taur, "Three-dimensional 'atomistic' simulation of discrete random dopant distribution effects in sub-0.1-μm MOSFETs," in *IEDM Tech. Dig.*, Dec. 1993, pp. 705–708.

[38] M. Miyamura, T. Nagumo, K. Takeuchi, K. Takeda, and M. Hane, "Effects of drain bias on threshold voltage fluctuation and its impact on circuit characteristics," in *IEDM Tech. Dig.*, Dec. 2008, pp. 447–450.

[39] N. Sugii, R. Tsuchiya, T. Ishigaki, Y. Morita, H. Yoshimoto, K. Torii, and S. Kimura, "Comprehensive study on vth variability in silicon on thin BOX (SOTB) CMOS with small random dopant fluctuation: Finding a way to further reduce variation," in *IEDM Tech. Dig.*, Dec. 2008, pp. 249–252.

[40] F. Arnaud, A. Thean, M. Eller, M. Lipinski, Y. W. Teh, M. Ostermayr, K. Kang, N. S. Kim, K. Ohuchi, J.-P. Han, D. R. Nair, J. Lian, S. Uchimura, S. Kohler, S. Miyaki, P. Ferreira, J.-H. Park, M. Hamaguchi, K. Miyashita, R. Augur, Q. Zhang, K. Strahrenberg, S. ElGhouli, J. Bonnouvrier, F. Matsuoka, R. Lindsay, J. Sudijono, F. S. Johnson, J. H. Ku, M. Sekine, A. Steegen, and R. Sampson, "Competitive and cost effective high-k-based 28-nm CMOS technology for low-power applications," in *IEDM Tech. Dig.*, Dec. 2009, pp. 651–654.

[41] K. Cheng, A. Khakifirooz, P. Kulkarni, S. Ponoth, J. Kuss, D. Shahrjerdi, L. F. Edge, A. Kimball, S. Kanakasabapathy, K. Xiu, S. Schmitz, A. Reznicek, T. Adam, H. He, N. Loubet, S. Holmes, S. Mehta, D. Yang, A. Upham, S.-C. Seo, J. L. Herman, R. Johnson, Y. Zhu, P. Jamison, B. S. Haran, Z. Zhu, L. H. Vanamurth, S. Fan, D. Horak, H. Bu, P. J. Oldiges, D. K. Sadana, P. Kozlowski, D. McHerron, J. O'Neill, and B. Doris, "Extremely thin SOI (ETSOI) CMOS with record low variability for low-power system-on-chip applications," in *IEDM Tech. Dig.*, Dec. 2009, pp. 49–52.

[42] H. Fukutome, K. Okabe, K. Okubo, H. Minakata, Y. Morisaki, K. Ikeda, T. Yamamoto, K. Hosaka, Y. Momiyama, M. Kase, and S. Satoh, "(110) NMOSFETs competitive to (001) NMOSFETs: Si migration to create (331) facet and ultrashallow Al implantation after NiSi formation," in *IEDM Tech. Dig.*, Dec. 2008, pp. 59–62.

[43] H. Fukutome, Y. Hori, L. Sponton, K. Hosaka, Y. Momiyama, S. Satoh, R. Gull, W. Fichtner, and T. Sugii, "Comprehensive design methodology of dopant profile to suppress gate-LER-induced threshold voltage variability in 20-nm NMOSFETs," in *VLSI Symp. Tech. Dig.*, Jun. 2009, pp. 146–147.

[44] M. Goto, S. Kawanaka, S. Inumiya, N. Kusunoki, M. Saitoh, K. Tatsumura, A. Kinoshita, S. Inaba, and Y. Toyoshima, "The study of mobility-tin: Tradeoff in deeply scaled high-k/metal gate devices and scaling design guideline for 22-nm node generation," in *VLSI Symp. Tech. Dig.*, Jun. 2009, pp. 214–215.

[45] S. Hasegawa, Y. Kitamura, K. Takahata, H. Okamoto, T. Hirai, K. Miyashita, T. Ishida, H. Aizawa, S. Aota, A. Azuma, T. Fukushima, H. Harakawa, E. Hasegawa, M. Inohara, S. Inumiya, T. Ishizuka, T. Iwamoto, N. Kariya, K. Kojima, T. Komukai, N. Matsunaga, S. Mimotogi, S. Muramatsu, K. Nagatomo, S. Nagahara, Y. Nakahara, K. Nakajima, K. Nakatsuka, M. Nishigoori, A. Nomachi, R. Ogawa, N. Okada, S. Okamoto, K. Okano, T. Oki, H. Onoda, T. Sasaki, M. Satake, T. Suzuki, Y. Suzuki, M. Tagami, K. Takeda, M. Tanaka, K. Taniguchi, M. Tominaga, G. Tsutsui, K. Utsumi, S. Watanabe, T. Watanabe, Y. Yoshimizu, T. Kitano, H. Naruse, Y. Goto, T. Nakayama, N. Nakamura, and F. Matsuoka, "A cost-conscious 32-nm CMOS platform technology with advanced single exposure lithography and gate-first metal gate/high-k process," in *IEDM Tech. Dig.*, Dec. 2008, pp. 938–940.

[46] H. Kawasaki, V. S. Basker, T. Yamashita, C.-H. Lin, Y. Zhu, J. Faltermeier, S. Schmitz, J. Cummings, S. Kanakasabapathy, H. Adhikari, H. Jagannathan, A. Kumar, K. Maitra, J. Wang, C.-C. Yeh, C. Wang, M. Khater, M. Guillorn, N. Fuller, J. Chang, L. Chang, R. Muralidhar, A. Yagishita, R. Miller, Q. Ouyang, Y. Zhang, V. K. Paruchuri, H. Bu, B. Doris, M. Takayanagi, W. Haensch, D. McHerron, J. O'Neill, and K. Ishimaru, "Challenges and solutions of FinFET integration in an SRAM cell and a logic circuit for 22-nm node and beyond," in *IEDM Tech. Dig.*, Dec. 2009, pp. 289–292.

[47] H. Kawasaki, M. Khater, M. Guillorn, N. Fuller, J. Chang, S. Kanakasabapathy, L. Chang, R. Muralidhar, K. Babich, Q. Yang, J. Ott, D. Klaus, E. Kratschmer, E. Sikorski, R. Miller, R. Viswanathan, Y. Zhang, J. Silverman, Q. Ouyang, A. Yagishita, M. Takayanagi, W. Haensch, and K. Ishimaru, "Demonstration of highly scaled FinFET SRAM cells with high-k/metal gate and investigation of characteristic variability for the 32-nm node and beyond," in *IEDM Tech. Dig.*, Dec. 2008, pp. 237–240.

[48] C. W. Oh, H. J. Bae, J. K. Ha, S. J. Park, B. K. Park, D.-W. Kim, T.-Y. Chung, K. S. Oh, and W.-S. Lee, "A novel Thin BOX SOI technology using bulk Si wafer for system-on-chip (SoC) application," in *VLSI Symp. Tech. Dig.*, Jun. 2009, pp. 96–97.

[49] C. Ortolland, L.-A. Ragnarsson, P. Favia, O. Richard, C. Kerner, T. Chiarella, E. Rosseel, Y. Okuno, A. Akheyar, J. Tseng, J.-L. Everaert, T. Schram, S. Kubicek, M. Aoulaiche, M. J. Cho, P. P. Absil, S. Biesemans, and T. Hoffmann, "Optimized ultralow thermal budget process flow for advanced high-k/metal gate first CMOS using laser-annealing technology," in *VLSI Symp. Tech. Dig.*, Jun. 2009, pp. 38–39.

[50] S.-D. Suk, Y. Y. Yeoh, M. Li, K. H. Yeo, S. H. Kim, D. W. Kim, D. Park, and W. S. Lee, "TSNWFET for SRAM cell application: Performance variation and process dependency," in *VLSI Symp. Tech. Dig.*, Jun. 2009, pp. 38–39.

[51] R. Watanabe, A. Oishi, T. Sanuki, H. Kimijima, K. Okamoto, S. Fujita, H. Fukui, K. Yoshida, H. Otani, E. Morifuji, K. Kojima, M. Inohara, H. Igrashi, K. Honda, H. Yoshimura, T. Nakayama, S. Miyake, T. Hirai, T. Iwamoto, Y. Nakahara, K. Kinoshita, T. Morimoto, S. Kobayashi, S. Kyoh, M. Ikeda, K. Imai, M. Iwai, N. Nakamura, and F. Matsuoka, "A low-power 40-nm CMOS technology featuring extremely high density of logic (2100 kGate/μm2) and SRAM (0.195 μm2) for wide range of mobile applications with wireless system," in *IEDM Tech. Dig.*, Dec. 2008, pp. 641–644.

[52] S. Y. Wu, J. J. Liaw, C. Y. Lin, M. C. Chiang, C. K. Yang, J. Y. Cheng, M. H. Tsai, M. Y. Liu, P. H. Wu, C. H. Chang, L. C. Hu, C. I. Lin, H. F. Chen, S. Y. Chang, S. H. Wang, P. Y. Tong, Y. L. Hsieh, K. H. Pan, C. H. Hsieh, C. H. Chen, C. H. Yao, C. C. Chen, T. L. Lee, C. W. Chang, H. J. Lin, S. C. Chen, J. H. Shieh, S. M. Jang, K. S. Chen, Y. Ku, Y. C. See, and W. J. Lo, "A highly manufacturable 28-nm CMOS low-power platform technology with fully functional 64-Mb SRAM using dual/triple-gate oxide process," in *VLSI Symp. Tech. Dig.*, Jun. 2009, pp. 210–211.

[53] H. S. Yang, R. Wong, R. Hasumi, Y. Gao, N. S. Kim, D. H. Lee, S. Badrudduza, D. Nair, M. Ostermayr, H. Kang, H. Zhuang, J. Li, L. Kang, X. Chen, A. Thean, F. Arnaud, L. Zhuang, C. Schiller, D. P. Sun, Y. W. Teh, J. Wallner, Y. Takasu, K. Stein, S. Samavedam, D. Jaeger, C. V. Baiocco, M. Sherony, M. Khare, C. Lage, J. Pape, J. Sudijono, A. L. Steegen, and S. Stiffler, "Scaling of 32-nm low-power SRAM with high-k metal gate," in *IEDM Tech. Dig.*, Dec. 2007, pp. 467–470.

[54] J. Zhuge, R. Wang, R. Huang, J. Zou, X. Huang, D.-W. Kim, D. Park, X. Zhang, and Y. Wang, "Experimental investigation and design optimization guidelines of characteristic variability in silicon nanowire CMOS technology," in *IEDM Tech. Dig.*, Dec. 2009, pp. 61–64.

Kelin J. Kuhn (M'78–SM'04–F'11) received the B.S. degree in electrical engineering from the University of Washington, Seattle, in 1980 and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Palo Alto, CA, in 1985.

Prior to 1997 she was a tenured faculty Member with the Department of Electrical and Computer Engineering, University of Washington. In 1997, she joined Intel Corporation, Hillsboro, OR, working on the Intel 0.35-μm process technology and has been involved in the Intel manufacturing process technology development for the 0.35-μm, 130-nm, 90-nm, 45-nm, and 22-nm technology nodes. She is currently an Intel Fellow and the Director of Advanced Device Technology for the Components Research Group, Intel Corporation. She is the author of more than 70 technical papers in electronics and photonics, as well as the textbook *Laser Engineering* (Prentice-Hall, 1997).

Dr. Kuhn is the recipient of a National Science Foundation Presidential Young Investigator Award for her work on strained layer III–V materials and two Intel Achievement awards (one award for her work on HiK-metal gate and another award for her work on the 22-nm device architecture).

Martin D. Giles (S'82–M'84–SM'97–F'06) received the B.A. and M.A. degrees in natural sciences from Cambridge University, Cambridge, U.K., in 1981 and 1985, respectively, and the M.S.E.E. and Ph.D. degrees from Stanford University, Stanford, CA, in 1983 and 1984, respectively.

From 1984 to 1990, he was a Member of the VLSI Device Analysis and Simulation Group, AT&T Bell Laboratories. From 1990 to 1994, he was an Assistant Professor with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, working on silicon process modeling and related topics. In 1994, he was with the Technology CAD Department, Intel Corporation, Hillsboro, OR, where he became the Program Manager of the Process and Device Modeling Group. He is currently a Senior Principal Engineer with the Process Technology Modeling Department, Intel Corporation. He has published more than 70 journal papers and conference proceedings and is the holder of 11 issued patents. His research interests include process technology variation effects on devices and circuits and the modeling of advanced device technologies.

Dr. Giles is a member of the IEEE Electron Devices Society Technical Committee on TCAD.

David Becher received the B.S. degree in electrical engineering from the University of Nebraska, Lincoln, in 1998 and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Illinois at Urbana-Champaign, Urbana, in 2000 and 2002, respectively.

In 2002, he joined the Portland Technology Development Team, Intel Corporation, Hillsboro, OR, as a Senior Process Engineer and has worked on the logic manufacturing processes for the 90-, 45-, 32-, and 15-nm technology nodes.

Pramod Kolar received the B.E degree in electronics and communications engineering from the National Institute of Technology, Surathkal, India, in 1998 and the M.S. and Ph.D. degrees in electrical engineering from Duke University, Durham, NC, in 2002 and 2005, respectively.

Since 2005, he has been with the Advanced Design Group, Logic Technology Development, Intel Corporation, Hillsboro, OR, where he works on SRAM bitcell development, statistical circuit design, and yield analysis. He is the author or a coauthor of 16 papers published in international conference proceedings and technical journals and is the holder of two U.S. patents. He was a graduate Intern with Qualcomm in 2004.

Dr. Kolar received the Inventor Recognition Award from the Semiconductor Research Corporation in 2005.

Avner Kornfeld received the B.Sc., M.Sc., and D.Sc. degrees in electrical engineering from the Technion—Israel Institute of Technology, Technion City, Haifa, Israel, in 1980, 1982, and 1986, respectively.

From 1986 to 1991, he was a Lecturer with the Department of Electrical Engineering, Technion, working on cryogenic focal plane signal processing and device flicker noise. In 1992, he joined Intel Corporation, Hillsboro, OR, and worked for the Intel Communication and the Mobile Microprocessor Groups. He is currently a Principal Engineer with the Technology and Manufacturing Group, working on analog design, variation, and the interaction between design and process. Prior to his current role, he worked on power and frequency optimization and projection for Intel mobile microprocessors in 130-, 90-, 65-, and 45-nm nodes.

Dr. Kornfeld received the 1997 Intel Achievement Award for his work on the Intel 100/10 Mb/s fast Ethernet transceiver design.

Roza Kotlyar received the B.S. degree in physics from the Polytechnic University, Brooklyn, NY, in 1992 and the Ph.D. in theoretical condensed matter physics from the University of Maryland, College Park, in 1998.

From 1998 to 2000, she was a National Research Council Postdoctoral Fellow with the Electronics Science and Technology Division (ESTD), Naval Research Laboratory, Washington, DC. Since 2000, she has been with the Process Technology Modeling Department, Intel Corporation, Hillsboro, OR, working on the modeling of the quantum confinement effects, tunneling, high-k scattering, and the effect of stress in nanoscale devices. She is the author or a coauthor of 38 journal papers and conference proceedings. Her research activities and interests include NEGF quantum transport modeling and atomistic modeling approaches in scaled MOSFET devices.

Sean T. Ma received the B.S., M.S., and Ph.D. degrees in electrical engineering from the University of Arizona, Tucson, in 1996, 1999, and 2004, respectively.

Since 2004, he has been with the Process Technology Modeling Department, Intel Corporation, Hillsboro, OR, working on process and device modeling of Intel logic and flash transistors. His research interests include process/device optimization and variation.

Atul Maheshwari received the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Massachusetts, Amherst, in 2001 and 2004, respectively.

He is currently a Senior Design Engineer with the Technology and Manufacturing Group, Logic Technology Development Department, Intel Corporation, Hillsboro, OR. He is responsible for the delivery of performance and process variation probing circuits for 32- and 22-nm process technology. His research interests include performance and variation monitors, thermal sensors, and PLLs.

Sivakumar Mudanai received the B.Tech. degree from the Indian Institute of Technology, Madras, Chennai, India, and the M.S and Ph.D. degrees from the University of Texas at Austin, Austin.

In 2001, he joined Intel Corporation, Portland, OR. He is currently with the Process Technology Modeling Department. His research focus is on compact device modeling for digital, analog, and RF applications. He is also involved in MOSFET device variation analysis and modeling.