On CMOS Circuit Reliability from the MOSFETs and the Input Vectors

Valeriu Beiu [1, 2] and Walid Ibrahim [1, 3]
[1] *College of Information Technology, UAE University, Al-Ain, P.O. Box 17551, UAE*
[2] *School of Intelligent Systems, University of Ulster, Magee Campus, UK*
[3] *Department of Systems & Computer Engineering, Carleton University, Ottawa, Canada*
{vbeiu, walidibr}@uaeu.ac.ae

Abstract

From the circuit reliability point of view transistors (and also wires) have mostly been ignored, with gates taking the lion's share. With scaling, this gate-level approach is becoming less-and-less accurate, as transistors (and wires) are starting to err more-and-more. That is why approximating the probability of failure of a gate by a fixed value (i.e., a constant) cannot hold in the long run. Trying to do justice to the elementary (nano-)transistors, this paper will present a failure analysis following on those papers which have made transistors' variations their concern. We will review the state-of-the-art, and start from the latest reported results on variations of the threshold voltage of MOSFETs (obtained using the Glasgow 3D atomistic drift/diffusion simulator) for estimating the probability of failure of classical CMOS gates. This approach will also expose the significance of the input vectors, and point to large differences between the upper and lower reliability bounds of elementary gates (which cannot be ignored when estimating the reliability of large circuits). The results reported here should have implications for forward-looking investigations on and design of emerging nano-architectures.

1. Introduction

Computers are currently heading towards several fundamental limitations [1]. The most daunting one comes from the fact that we are still relying on the classical physics for the design of the hurly-burly rush of trillions of electrons through billions of wires and transistors [2]. But, the chips at the heart of today's computers are running "out of steam," as electrons and atoms are starting to reveal their discreetness [3].

That is why the massive scaling of the CMOS devices deep into the nanometer regime is expected to introduce larger-and-larger (both static and dynamic) parameter fluctuations/variations at material, device and circuit levels 0–[6]. Extreme parameter variations are a major barrier to achieving reliable and predictable system implementations. At the material level, one of the potentially significant sources of fluctuation is the randomness in the exact location of doping atoms [5]. Although the average concentration of doping is well controlled by ion implantation and annealing processes, it was shown in [7] that the standard deviation of the number of doping atoms per device could increase without bound as the doping area decreases. This will lead to device-to-device fluctuations in key parameters, including the threshold voltage (V_{TH}), as device dimensions scale down.

In this paper we shall consider only the effects of V_{TH} variations on the probability of failure of standard CMOS gates, while the results presented here are being integrated into EDA reliability tools (allowing for very accurate system-level estimates). We shall start by revisiting prior work on V_{TH} variations of MOSFETs. Afterwards, we shall use the latest reported results for estimating the probability of failure of classical CMOS gates. As expected, the results obtained show strong dependence on the input vectors. Conclusion and further directions of research are ending the paper.

2. Variability of nano-MOSFETs

One of the fundamental limitation of MOSFETs is the accuracy of reproducing V_{TH} over the large number of transistors in a chip. It is long known [8], [9] that V_{TH} depends on the doping level, which varies due to the random distribution of impurities/dopants (the theoretical model was based upon the percolation theory for explaining this dependence). The problem was left mostly dormant for almost two decades as it did not significantly affect the integrated circuits of the 80's [10] and 90's [11], [12].

Towards the end of the last decade, with MOSFETs being scaled down into the nanometer regime (below 100nm), the problem has emerged with a vengeance [13]–[27] (this being only a partial list). The random fluctuations of both *the number of the dopants* and of *their physical locations* have started to attract very thorough investigations. This is due to the fact that the particular microscopic dopant distribution in the MOSFET channel has a non-negligible influence on the electrical performances. For analyzing such influences while scaling, three-dimensional (3D) device simulations have had to be refined. Among the first results were those reported in [11], [13], where 24 MOSFETs with different dopant distributions were analyzed using a 3D drift-diffusion simulator. They showed that the effects are aggravated by decreasing the channel volume, and pointed to a strong correlation between the deviation of V_{TH} and the deviation (from the mean) of the number of dopants. The discrete nature of the dopants results in an non-homogenous channel potential, which in turn allows for the early turn-on in parts of the channel, leading to shifts in V_{TH}. These results were improved upon by Asenov who significantly increased the size of the sample to 2500 microscopically different devices [14], more adequate for quantitative statistical predictions. It was hypothesized that V_{TH} fluctuations associated with random dopants follow a normal distribution, but while the atomistically simulated standard deviation of V_{TH} followed the dependence predicted by analytical models $(L_{eff})^{-1/2}$ [16], [18], its magnitude was higher.

With the continuous stride in computing power, 3D numerical simulations on larger scales have been used for understanding V_{TH} fluctuations due to *oxide thickness variations* [15], *line edge roughness* [20], and *polysilicon granularity*. It is only during the last few years that the *combined effects of all these factors* have been investigates. Some of the largest simulations have been carried out using the Glasgow 3D atomistic drift-diffusion device simulator, featuring density gradient quantum corrections, and calibrated with respect to a Toshiba 35nm MOSFET [20]. Similar results have been recently reported in [25]. These have shown that:

$$\sigma \approx t_{ox} N_A^{0.4} / (L_{eff} W_{eff})^{-1/2} \tag{1}$$

being in good agreement with [18], which shows an exponent of 0.45 instead of 0.4 (for the doping level N_A).

Finally, during the last year, quite a few results have been reported [21]–[25], and presented [26], [27]. These are based on analyzing a very large statistical sample of 100,000 microscopically different 35nm transistors, and suggest that V_{TH} distribution *deviates from the Gaussian distribution*. Still, these references show $\mu = 225.9 \pm 0.1$ [mV] as well as a standard deviation $\sigma = 30.30 \pm 0.05$ [mV] with a skew of 0.1597 ± 0.003 and a kurtosis of 0.0487 ± 0.006.

3. The CMOS inverter revisited

From eq. (1), $\sigma \approx (L_{eff} W_{eff})^{-1/2}$, and assuming $L_{eff} \approx a$, and $W_{eff} \approx ka$, it follows that $\sigma \approx a^{-1} k^{-1/2}$. For simplicity in this paper we shall assume a Gaussian distribution (while it would not be too difficult to modify the distribution, as suggested in the conclusions). Taking into account that for a transistor of 35nm the standard deviation is $\sigma = 30.28$mV, we can quickly estimate σ for a transistor of another size "a" [nm] as:

$$\sigma = \frac{30.28 \times 35}{a\sqrt{k}} \text{[mV]}. \tag{2}$$

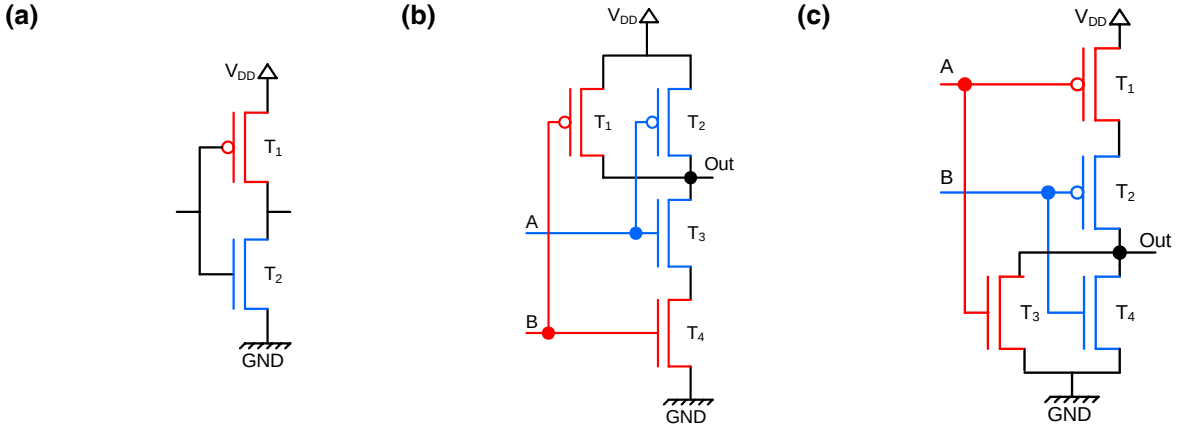

Fig. 1. CMOS standard implementation for: (a) Inverter; (b) NAND-2; (c) NOR-2.

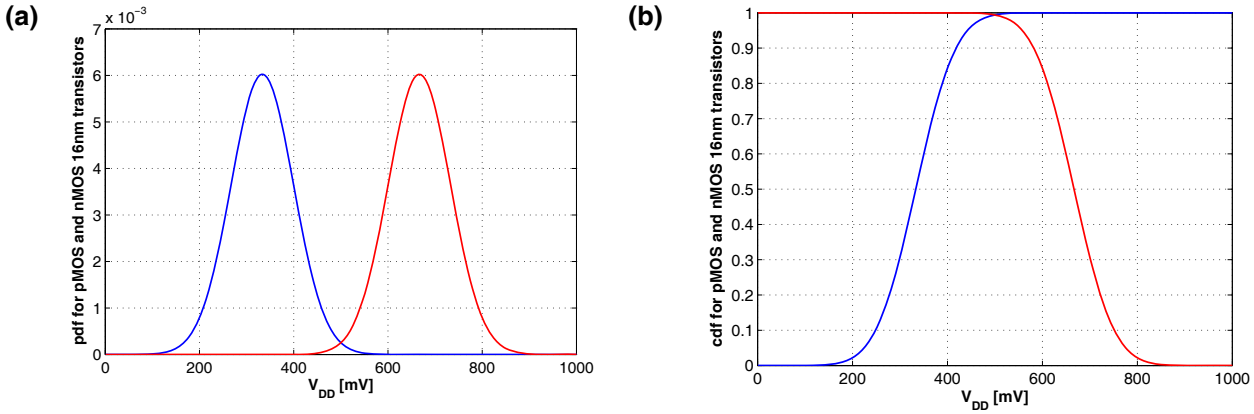

Fig. 2. nMOS and pMOS transistors (16nm): (a) pdf; (b) P_{ON} (probability that a transistor is ON).

It follows that the V_{TH} probability density function (*pdf*) and the cumulative distribution function (*cdf*) can be calculated as:

$$pdf(v) = \exp[-(v - V_{TH})^2 / 2\sigma^2] / (\sigma\sqrt{2\pi}) \quad (3)$$

$$cdf(v) = \{1 + erf[(v - V_{TH}) / (\sigma\sqrt{2})]\} / 2 . \quad (4)$$

One could immediately consider that the probability that an nMOS transistor is ON (P_{ON}) at a given voltage v is $P_{ON}(v) = cdf(v)$, while the P_{ON} for a pMOS transistor is $1 - cdf(v)$. Fig. 2 shows the *pdf* and P_{ON} for 16nm nMOS and pMOS transistors assuming that the nMOS and the pMOS transistor are being "switched" at $V_{DD}/3$ and $2V_{DD}/3$ respectively. Fig. 2(b) shows that $P_{ON}(250\text{mV}) = 0.1042$ for the nMOS transistor and respectively $P_{ON}(250\text{mV}) \approx 1$ for the pMOS transistor. It also shows that the probability of failure depends (to a great extent) on the allowed range of the input signal. If an input signal up to 250mV (*i.e.*, $V_{DD}/4$) is allowed as logic 0, there is a 0.1042 probability that the nMOS transistor will be ON while a logic 0 is applied on its gate. This means that due to the large variations of V_{TH} at 16nm (σ = 73.1 mV), 10.42% of the nMOS transistors are expected to be ON when a logic 0 is applied on their gates (*i.e.*, to FAIL). This probability of failure is reduced to only 2.2% if the maximum allowed value for a logic 0 signal is reduced to 200mV (or if the nMOS transistor is made larger).

The probability of failure of a CMOS inverter (PF_{INV}) was estimated by Mead and Conway [28] as:

$$PF_{\text{INV}} = \exp(-2V_{DD} / \sigma) . \quad (5)$$

However, the INV shown in Fig. 1(a) is expected to work correctly in case of a logic 0 input signal only when transistor T_1 is ON and transistor T_2 is OFF, and vice versa in case of a logic 1 input signal. Here we neglect the ON-ON (representing fighting, and increasing power) as well as OFF-OFF (representing sub-threshold operation, and being slow) cases. Assuming that the two transistors are identical (while in practice $\sigma_{\text{pMOS}} < \sigma_{\text{nMOS}}$), PF_{INV} can be calculated starting from the transistor-level as:

$$PF_{\text{INV}} = 1 - P_{ON}(T_1) \times [1 - P_{ON}(T_2)]. \quad (6)$$

Fig. 3 illustrates the effect of the V_{TH} settings on PF_{INV} calculated using eq. (6). It shows that adjusting the value of V_{TH} can affect the probability of failure of the inverter. Reducing the value of V_{TH} makes PF_{INV} for logic 0 input signal higher than for logic 1 input signal. Setting $V_{TH} = V_{DD}/2$, simultaneously minimizes the probabilities of failure for both logic 0 and logic 1.

Fig. 4 compares PF_{INV} based on the estimation of Mead and Conway (eq. (5)) against the values calculated using eqs. (4) and (6). It shows that eq. (5) is underestimating PF_{INV}. It also shows that PF_{INV} depends (to a great extend) on the allowed noise margins (which was varied when applying eqs. (4) and (6)). As an example, reducing the noise margin from $V_{DD}/3$ to $V_{DD}/4$ at the 25nm reduces PF_{INV} from 3.78E–5 to 6.39E–10.

4. Classical CMOS gates

Fig. 1 shows the CMOS standard implementation of the NAND-2 and NOR-2 logic gates. In [4], Forshaw *et al.* suggested that PF_{GATE} could be (roughly but quickly) estimated as:

$$PF_{\text{GATE}} = 1 - (1 - \varepsilon)^n, \quad (7)$$

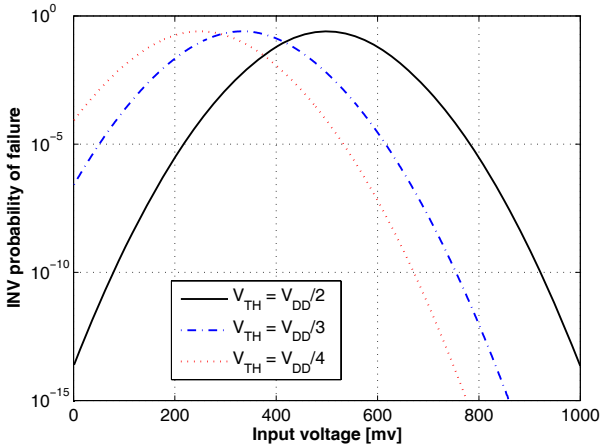

Fig. 3. PF_{INV} of an inverter using 16nm transistors for different values of V_{TH}.

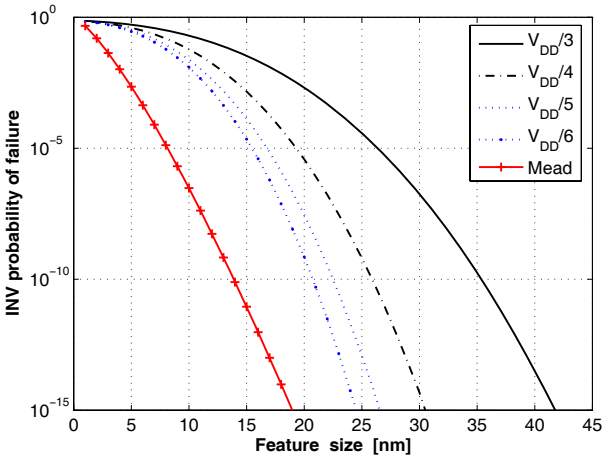

Fig. 4. PF_{INV} against the feature size (of the transistors) for different noise margins, as well as Mead and Conway estimate [28].

where PF_{GATE} denotes the probability of failure of a gate, ε denotes the probability of failure of a transistor, and n is the number of transistors the gate has. This estimation assumes that a gate fails if any of the n transistors fail, regardless of the internal construction of the gate and the particular input vector. To calculate more accurately the probability of failure at the gate level, both the internal construction of the gate as well as the applied input vector should be taken into account, in addition to the probability of failure of the individual transistors.

In case of input vectors "00", "01", "10", the expected output signal of a NAND-2 gate is logic 1. Therefore, a NAND-2 gate fails if both T_1 and T_2 are OFF, or when T_3 and T_4 are both ON (regardless of the state of T_1 and T_2). Thus, $PF_{\text{NAND-2}}$ in case of logic 1 (high) output signal can be calculated as:

$$PF_{\text{NAND-2}}\,(\text{Output High}) = P_{OFF}(T_1) \times P_{OFF}(T_2) + P_{ON}(T_3) \times P_{ON}(T_4) - P_{OFF}(T_1) \times P_{OFF}(T_2) \times P_{ON}(T_3) \times P_{ON}(T_4), \quad (8)$$

where $P_{OFF}(T_i) = 1 - P_{ON}(T_i)$.

On the other hand, in case of input vector "11", the NAND-2 gate fails if any of the four transistors fails. Consequently, $PF_{\text{NAND-2}}$ can be calculated as:

$$PF_{\text{NAND-2}}\,(\text{Output Low}) = 1 - P_{OFF}(T_1) \times P_{OFF}(T_2) \times P_{ON}(T_3) \times P_{ON}(T_4) \quad (9)$$

Fig. 5 shows the probability of failure of a NAND-2 gate using 16nm transistors and $V_{TH} = V_{DD}/2$ (best possible). It reveals that $PF_{\text{NAND-2}}$ for the input vector "00" is significantly lower than for the input vectors "01", "10", and "11." It also shows that the value of $PF_{\text{NAND-2}}$ depends on the allowed input range (of the input signals). As an example, in case of input vector "00", the $PF_{\text{NAND-2}}$ increases from 2.8596E–19 to 1.0117E–15 when the value of the allowed input signals is increased from 50mV to 100mV.

Fig. 6 compares the estimated value of $PF_{\text{NAND-2}}$ using eq. (7) against the more accurate values obtained using eqs. (8) and (9). The value of V_{TH} was set to $V_{DD}/2$ for both the nMOS and the pMOS transistors. The maximum allowed logic 0 input signal was set to $V_{DD}/3$ and the minimum allowed logic 1 input signal was set to $2V_{DD}/3$. The simulation results show that the value estimated using eq. (7) provides a very tight upper bound for $PF_{\text{NAND-2}}$. This can be understood because eq. (7) assumes that if any of the transistors fail the gate will fail, which is equivalent to the case when the input vector is "11." Fig. 6 also shows that (as expected) $PF_{\text{NAND-2}}$ increases exponentially with the reduction of the feature size. When going to 22nm, a NAND-2 gate using minimum feature size transistors is expected to have a probability of failure of about 10^{-3}. Increasing the size of all transistors by 4 (as customary done in VLSI) immediately reduces this value to 10^{-12}.

In case of a NOR-2, the results are identical to those for NAND-2 (Figs. 5 and 6) except that they are swapped between the "11" and the "00" input vectors.

5. Conclusions

This paper has presented an extension from a low-level transistor variability model to the gate-level reliability model (for classical CMOS gates), allowing

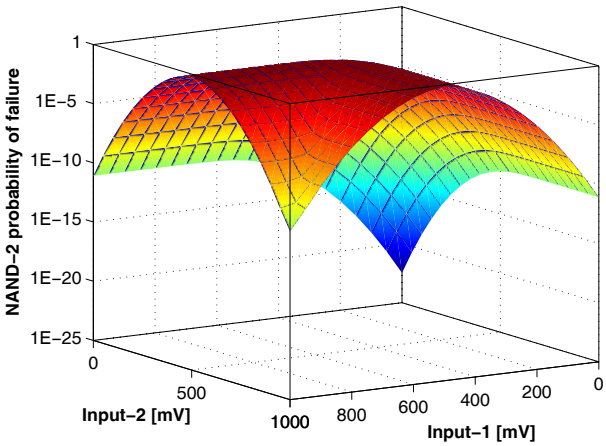

Fig. 5. PF_{NAND-2} for a NAND-2 gate using transistors of 16nm.

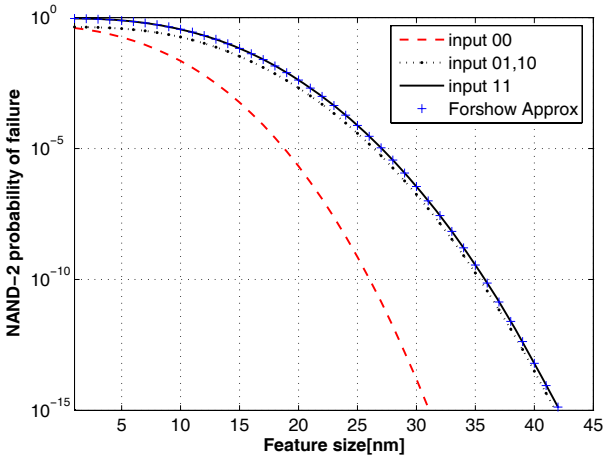

Fig. 6. PF_{NAND-2} against the feature size for different input vectors, as well as when using Forshaw *et al.* estimation [4].

for better (*i.e.*, more accurate and detailed) estimates of the probability of failure at the gate-level. It includes transistors' variations and also takes into account the different input vectors.

The results presented here are for CMOS gates only, while they should be considered in combination with more precise error models for the wires [30], and detailed analyses of larger circuits [31]. All of these different models (for devices and wires) are currently being integrated in our in-house nano-CR-EDA[2] (nano-circuit EDA tool for Evaluating Design Alternatives). When properly accounting for the input vectors, our expectation is that much more precise reliability estimates at the circuit-level will be achievable.

Future directions of research involve replacing the Gaussian with a stable distribution (heavy tailed). Preliminary results show PF_{NAND-2} in Fig. 7, while P_{INV} when slightly varying the distributions are presented in Fig. 8. Such results are still being refined, while the plan is to integrate them with Bayesian methods for allowing very accurate reliability estimates at the system level.

Acknowledgements. This research was partly supported by the UAE National Research Foundation through the recently established Emirates Center for Nanoscience and Nanoengineering.

6. References

[1] *International Technology Roadmap for Semiconductors (ITRS)*, SIA, Austin, TX, USA, 2007 & 2008. http://public.itrs.net/

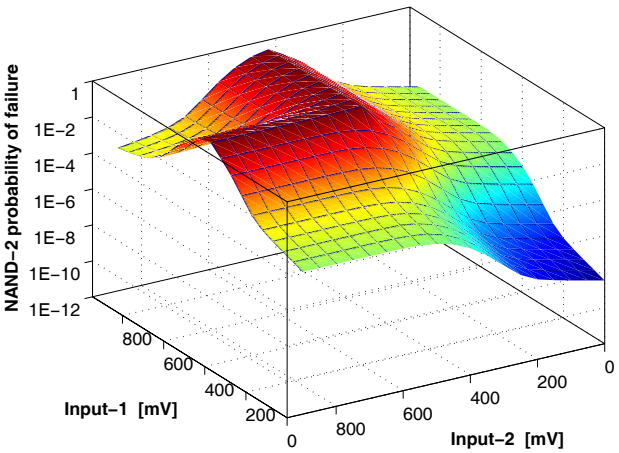

Fig. 7. PF_{NAND-2} for a NAND-2 gate using transistors of 16nm and stable distribution with $\alpha = 1.97$ (instead of the classical Gaussian).

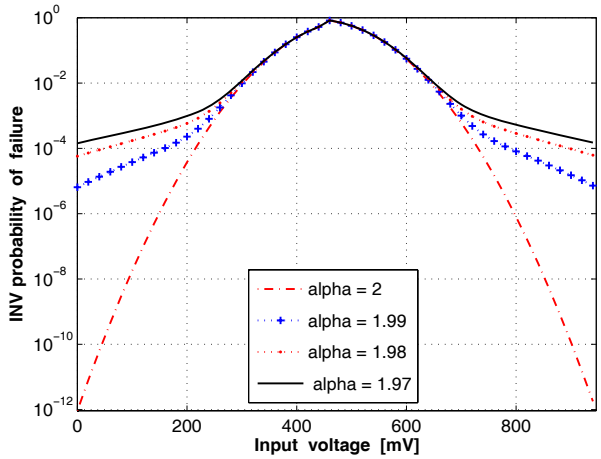

Fig. 8. PF_{INV} using 16nm transistors for Gaussian as well as a few stable distributions (very close to the Gaussian).

[2] V. Beiu, "Electrons behaving badly—Quo vadis nano-architectures?" (invited), *Proc. IES'08*, Sendai, Japan, Jul. 2008, pp. 24–27.

[3] R.K. Cavin *et al.*, "Research directions and challenges in nanoelectronics," *J. Nanoparticle Res.*, vol. 8, Dec. 2006, pp. 841–858.

[4] M. Forshaw, K. Nikolić, and A.S. Sadek, "ANSWERS: Autonomous Nanoelectronic Systems With Extended Replication and Signaling," *MEL-ARI #28667*, 3rd Year Report, 2001. http://ipga.phys.ucl.ac.uk/research/answers/reports/3rd_year_UCL.pdf

[5] D.J. Frank *et al.*, "Device scaling limits of Si MOSFETs and their application dependencies," *Proc. IEEE*, vol. 89, Mar. 2001, pp. 259–288.

[6] R.E. Bryant *et al.*, "Limitations and challenges for computer-aided design technology for CMOS VLSI," *Proc. IEEE*, vol. 89, Mar. 2001, pp. 341–365.

[7] J.D. Meindl, Q. Chen, and J.A. Davis, "Limits on silicon nanoelectronics for Terascale integration," *Science*, vol. 293, 14 Sep. 2001, pp. 2044–2049.

[8] R.W. Keyes, "Effect of randomness in the distribution of impurity ions on FET thresholds in integrated electronics," *IEEE J. Solid-State Circ.*, vol. 10, Aug. 1975, pp. 245–247.

[9] R.W. Keyes, "The effect of randomness in the distribution of impurity ions on FET thresholds," *Appl. Phys. A*, vol. 8, Nov. 1975, pp. 251–259.

[10] T. Hagivaga, K. Yamaguchi, and S. Asai, "Threshold voltage variation in very small MOS transistors due to local dopant fluctuations," *Proc. Symp. VLSI Tech.*, 1982, pp. 46–47.

[11] H.-S.P. Wong, and Y. Taur, "Three dimensional 'atomistic' simulation of discrete random dopant distribution effects in sub-0.1 µm MOSFETs," *Proc. Intl. Electr. Dev. Meeting IEDM'93*, Washington, DC, USA, Dec. 1993, pp. 705–708.

[12] T. Mizuno, J. Okamura, and A. Toriumi, "Experimental study of threshold voltage fluctuation due to statistical variation of channel dopant number in MOSFETs," *IEEE Trans. Electr. Dev.*, vol. 41, Nov. 1994, pp. 2216–2221.

[13] H.-S.P. Wong, Y. Taur, and D.J. Frank, "Discrete random dopant distribution effects in nanometer-scale MOSFETs," *Microelectr. Reliab.*, vol. 38, Sep. 1998, pp. 1447–1456.

[14] A. Asenov, "Random dopant induced threshold voltage lowering and fluctuations in sub-0.1 µm MOSFET's: A 3-D "atomistic" simulation study," *IEEE Trans. Electr. Dev.*, vol. 45, Dec. 1998, pp. 2505–2513.

[15] A. Asenov, S. Kaya, and J.H. Davies, "Intrinsic threshold voltage fluctuations in decanano MOSFETs due to local oxide thickness variations," *IEEE Trans. Electr. Dev.*, vol. 49, Jan. 2002, pp. 112–119.

[16] S. Toriyama, and N. Sano, "Probability distribution functions of threshold voltage fluctuations due to random impurities in deca-nano MOSFETs," *Physica E*, vol. 19, Jul. 2003, pp. 44–47.

[17] A. Asenov *et al.*, "Simulation of intrinsic parameter fluctuations in decananometer and nanometer-scale MOSFETs," *IEEE Trans. Electr. Dev.*, vol. 50, Sep. 2003, pp. 1837–1852.

[18] J. Yin, X. Shi, and R. Huang, "A new method to simulate random dopant induced threshold voltage fluctuations in sub-50 nm MOSFET's with non-uniform channel doping," *Solid-State Electr.*, vol. 50, Sep.-Oct. 2006, pp. 1551–1556.

[19] K. Bernstein *et al.*, "High-performance CMOS variability in the 65-nm regime and beyond," *IBM J. R&D*, vol. 50, Jul.-Sep. 2006, pp. 433–449.

[20] G. Roy *et al.*, "Simulation study of individual and combined sources of intrinsic parameter fluctuations in conventional nano-MOSFETs," *IEEE Trans. Electr. Dev.*, vol. 53, Dec. 2006, pp. 3063–3070.

[21] C. Millar *et al.*, "Accurate statistical description of random dopant-induced threshold voltage variability," *IEEE Electr. Dev. Lett.*, vol. 29, Aug. 2008, pp. 946–948.

[22] M.F. Bukhori, S. Roy, and A. Asenov, "Stastistical aspects of reliability in bulk MOSFETs with multiple defect states and random discrete dopants," *Microelectr. Reliab.*, vol. 48, Aug.-Sep. 2008, pp. 1549–1552.

[23] U. Kovac *et al.*, "Statistical simulation of random dopant induced threshold voltage fluctuations for 35 nm channel length MOSFET," *Microelectr. Reliab.*, vol. 48, Aug.-Sep. 2008, pp. 1572–1575.

[24] C. Alexander, G. Roy, and A. Asenov, "Random-dopant-induced drain current variation in nano-MOSFETs: A three-dimensional self-consistent Monte Carlo simulation study using *"ab initio"* ionized impurity scattering," *IEEE Trans. Electr. Dev.*, vol. 55, Nov. 2008, pp. 3251–3258.

[25] Y. Li *et al.*, "Large-scale atomistic approach to random-dopant-induced characteristic variability in nanoscale CMOS digital and high-frequency integrated circuits," *Proc. Intl. Conf. CAD ICCAD'08*, San Jose, CA, USA, Nov. 2008, pp. 278–285.

[26] A. Asenov, "Statistical device variability and its impact on design" (keynote), *Intl Symp. Asynch. Circ. & Syst. ASYNCH'08*, Newcastle upon Tyne, UK, Apr. 2008, pp. xv–xvi. http://async.org.uk/async2008/async-nocs-slides/Thursday/Keynote/Asenov_Async_08.pdf

[27] A. Asenov *et al.*, "Advanced simulation of statistical variability and reliability in nano-CMOS transistors" (tutorial), *Proc. Intl. Electr. Dev. Meeting IEDM'08*, San Francisco, CA, USA, Dec. 2008, p. 1. http://userweb.elec.gla.ac.uk/a/abrown/papers/IEDM_2008.pdf

[28] C. Mead, and L. Conway, *Introduction to VLSI Systems*, Addision-Wesley, 1980.

[29] M. Alam, "Reliability- and process-variation aware design of integrated circuits," *Microelectr. Reliab.*, vol. 48, Aug.-Sep. 2008, pp. 1114–1122.

[30] V. Beiu, W. Ibrahim, and R.Z. Makki, "On wires operated at low electron densities," *Intl. North Eastern Workshop Circ. & Syst. NEWCAS'09*, Toulouse, France, Jun. 2009, accepted/in press.

[31] W. Ibrahim, V. Beiu, and H. Amer, "Why should we care about the input vectors?" *Intl. North Eastern Workshop Circ. & Syst. NEWCAS'09*, Toulouse, France, Jun. 2009, accepted/in press.