Modeling Statistical Variability with the Impedance Field Method

A systematic comparison between the Impedance Field and the "Atomistic" Method

Karim El Sayed and Eugeny Lyumkis
TCAD R&D, Synopsys Inc.
Mountain View, CA, USA
Karim.ElSayed@synopsys.com

Andreas Wettstein
TCAD R&D, Synopsys Switzerland LLC,
CH-8050 Zurich, Switzerland

Abstract— **As the transistor scaling continues, self-averaging of device properties for individual devices becomes less effective and thus the statistical variability of device properties become more prominent. Here we present the impedance field method (IFM) for statistical variability analysis, which provides a fast, convenient and accurate alternative to "atomistic" approaches and compare the two methods with respect to accuracy, calibration requirements and turn-around time.**

Keywords-TCAD; Variability Analysis; Impedance Field Method; Atomistic Simulations

I. THE IMPEDANCE FIELD METHOD

The basic idea behind the IFM is to treat the randomness as a perturbation of a reference device. Rather than solving the full, nonlinear Poisson and drift-diffusion equations for a large number of random device realizations, the 3D TCAD solution is obtained only once for the reference device. The current fluctuations at the device terminals caused by these random perturbations are then computed in linear response using a Green's function technique [1]. For modeling statistical variability we present three different variants of the IFM: The noise-like IFM, the statistical IFM and the deterministic IFM.

IFM can be applied to random doping fluctuations, bulk and interface trap fluctuations, geometric fluctuations (e.g. oxide thickness and line edge roughness), metal gate work function fluctuations as well as model parameter fluctuations [2]. As an illustration we discuss here these IFM variants for the example of random doping fluctuations. For all three IFM variants the 3D TCAD solution for the reference device with the doping profiles $N^s_{\mathrm{ref}}(\vec{r})$ is obtained first. The resulting terminal current deviation at a contact *c* follows from:

$$\delta I_{\nu,c} = \sum_s \int d^3r\, G^s_c(\vec{r})\,[N^s_\nu(\vec{r}) - N^s_{\mathrm{ref}}(\vec{r})] \quad , \quad (1)$$

where $N^s_\nu(\vec{r})$ is one particular random doping realization for the doping species *s*. (See for example Figure 1). The Green's function $G^s_c(\vec{r})$, which is also called the "impedance field", does not depend on the random doping realization. Within the noise-like IFM (nIFM) only second moment, i.e., the drain current variance $S_{c_1,c_2} = \langle \delta I_{c_1} \delta I_{c_2} \rangle$ is computed. (Also the threshold voltage variance can be computed from S_{c_1,c_2} .) Within the statistical IFM (sIFM) Eq. (1) is used directly, and a statistical sampling using a large number of random doping realizations $N^s_\nu(\vec{r})$ is performed to compute the small-signal current responses $\delta I_{\nu,c}$ at each contact c. From $\delta I_{\nu,c}$ the full IV curves of the randomized devices is constructed [3]. For deterministic IFM (dIFM), the variations are explicitly defined by the user. For example for deterministic doping variations in Eq. (1) the random doping realization $N^s_\nu(\vec{r})$ is replaced by a user defined actual modified doping profile.

The nIFM has been shown to yield accurate results for the standard deviations when the compared with "atomistic" methods [4]. The statistical IFM can also be applied when the knowledge of the standard deviations of terminal currents or voltages are insufficient. For example, for the statistical variation of the static noise margins of an SRAM cell. Fig. 2 shows the butterfly curves for 1000 randomizations of the 3D 6T SRAM shown in Fig. 1 for 4 different values of the supply voltage. Fig. 3 shows the butterfly curves for two selected randomizations for five different supply voltages. Consistent with experimental data given in Ref. [5] the sIFM results show that the shape of the butterfly curves at the full supply voltage is not a good predictor of the shape of the butterfly curves at lowered supply voltages. Even cells with very similar shaped butterfly curves at the full supply voltage can show quite different behavior at lowered supply voltages.

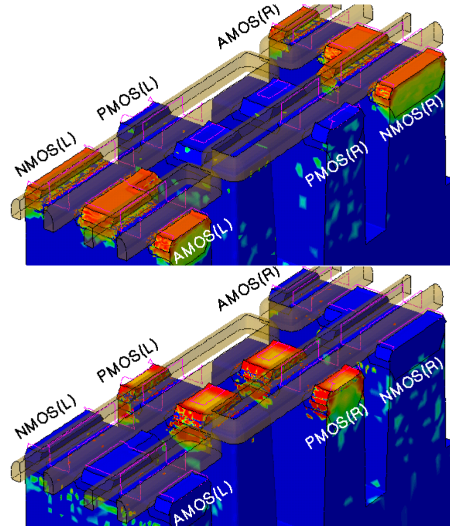

Figure 1: Example of a randomized realization of a 3D 6T SRAM Cell. Upper panel: donors. Lower Pannel: Acceptors. Oxide regions are suppressed for better viewing and nitride spacer are shown transparent.

ISBN 978-0-615-71756-2

II. COMPARISON IFM AND "ATOMISTIC" APPROACHES

Another approach, which is often used to analyze variability in semiconductor devices, is the so-called "atomistic" method [4]-[9] . In this section we investigate how the IFM method compares to "atomistic" approaches with respect to accuracy and turn-around time.

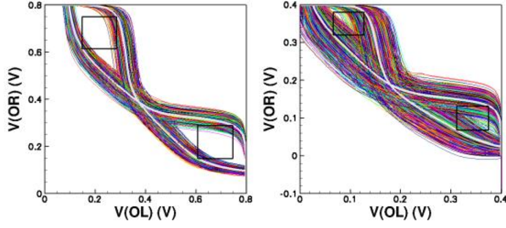

Figure 2: SRAM butterfly curves for 1000 randomizations of the 6T SRAM cell shown in Fig. 1 using sIFM. The white lines: reference device. "fitted" squares: average static noise margin. The supply voltages are (top- left to bottom-right): 0.8V, 0.4V

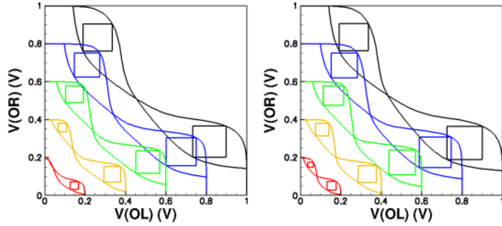

Figure 3: Butterfly curves at V_{dd}= 1.0, 0.8, 0.6, 0.4, and 0.2V for two devices with an effective SNM of 0.145V at V_{dd}= 1V. Left panel: a device for which one lob collapsed. Right: a device with the largest effective SNM at V_{dd}= 0.2V

A. *Large or Small Perturbation*

The accuracy of any linearization approach is inherently limited to "small" perturbations. However, it is not obvious how to judge the "smallness" of the perturbations associated with quantities relevant for variability analysis. For example in the channel of a typical MOS structure the doping concentration may be around 10^{18}/cm^3. After the "atomization" procedure in the channel area only 1 in about 100 mesh elements harbors a (single) dopant, resulting in an effective concentration of 10^{20}/cm^3 per mesh element, assuming a typical channel mesh element volume of 10 nm^3. All other channel mesh elements remain undoped. A deviation from the average doping concentration of 10^{18}/cm^3 to either zero or 10^{20}/cm^3 certainly does not seem "small". However, one has to remember that most relevant quantity is the resulting carrier concentration, not the doping concentration. It is the accepted standard to include the density gradient quantum correction model [10] for advanced MOSFET TCAD simulations in general and for random doping variability analysis in particular [8]. Even for such a large variation of the local doping concentration we find that when using the density gradient model the carrier concentration variability in the channel is quite well-behaved and limited to about one order of magnitude. This observation in itself does not prove the applicability of the linearization approach, but combined with the fact that nIFM has been shown to yield accurate results for the standard deviations when the compared with "atomistic" methods [4], lead us to conclude that the effective perturbations are in fact "small" enough such that the linearization approach remains valuable.

B. *Transport Model Parameter Selection*

The semi-empirical TCAD transport models have been developed and calibrated assuming continuous doping profiles, which represent an implicit "average" device. These TCAD transport models consequently contain semi-empirical non-linearities which are often expressed in terms of readily available numerical quantities. For example widely accepted mobility [11], [12] and band structure [13] models use the local doping concentration as an effective parameter instead of the carrier concentration in order to make the model more robust. The careful calibration of these semi-empirical TCAD transport models during the last two to three decades has ensured reliable TCAD results.

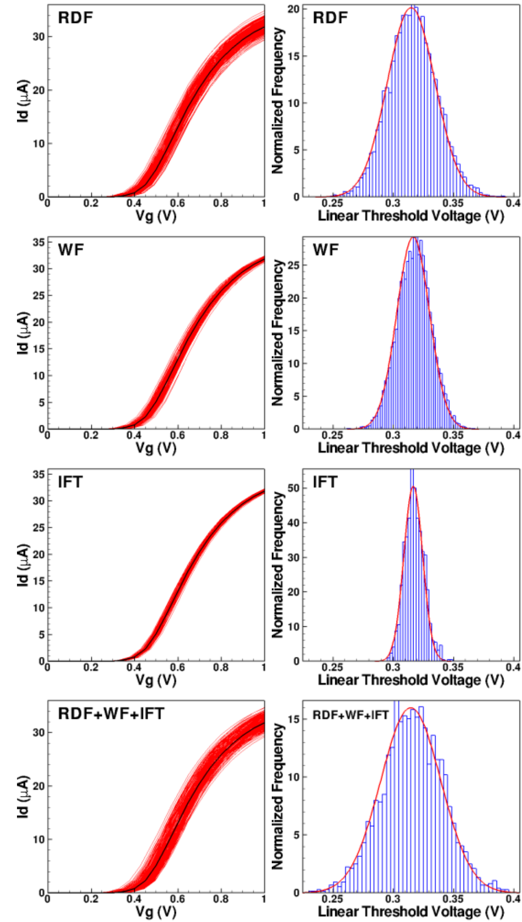

Figure 4: I_dV_g curves for the first 200 randomization and threshold voltage histogram for all 1000 randomizations of the NMOS transistor of the SRAM cell shown in Fig. 1. From top to bottom: random doping fluctuations, metal -gate work function variation, interface trap fluctuations, and all variability effects combined.

We find that certain TCAD transports models, such as the above mentioned ones, are ill-behaved when applied to randomized doping profiles. This indicates that many well-established TCAD transport models are not suitable for the "atomistic" approaches if used in a standard manner. To

illustrate this point we are comparing the TCAD simulations results for a reference device with the "atomistic" results for randomized devices when activating only a single transport parameter model at a time (otherwise, just the Poisson and carrier continuity equations are solved assuming a constant mobility and no band gap narrowing). We find that when the bandgap narrowing [13] model is activated the on-state drain current is systematically about 30% lower than that of the reference device. When activating the bulk Philips Unified Mobility [11] model current is systematically 500% higher. When activating the Lombardi surface degradation [12] models we find that the current is systematically 80% lower. This systematic deviation can be explained by considering that for all randomized devices the electrons in the channel experience about 99 out of 100 mesh elements as undoped – and consequently see little bandgap narrowing and bulk doping-dependent mobility degradation. For the Lombardi surface degradation models the screening contribution to the surface phonon scattering term uses the local doping concentration as an effective parameter, resulting in the artificial suppression of the mobility in undoped mesh elements. Such artifacts can be mitigated by carefully selecting a set of TCAD transport models, which are less susceptible to such effects and by carefully re-calibrating parameters [8],[14]. Also, it is also standard practice [14] to evaluate the mobility using the reference doping profiles instead of the "atomized" profile. While this trick certainly helps to mitigate the aforementioned artifact in "atomistic" TCAD simulations, this trick also excludes any doping variability effects on the mobility, and thus limits the predictability of the "atomistic" approach.

C. *Conductor network non-linearity*

Even after carefully selecting a set of TCAD transport models and parameters appropriate for "atomistic" simulations we find that the average of extracted quantity, such as for example V_{th} or I_{on}, over the "atomistic" simulations may not correspond to the value of the unperturbed reference device (a similar observation was first reported in [6]). Such a behavior can be understood in terms of the simple analytic model of a set of N parallel chains of M lumped conductors per chain. The IV relation of this model is given by:

$$I = V \sum_{i}^{N} \frac{1}{\sum_{j}^{M} \frac{1}{\sigma_{i,j}}} \quad , \quad (2)$$

Where $\sigma_{i,j}$ is the conductance of the jth conductor in the ith chain. The conductivities are randomized around an average value of σ_o, i.e., $\sigma_{i,j} = \sigma_o + \delta_{i,j}$. In linear response we find:

$$I = V\sigma_o \frac{N}{M} + \frac{V}{M^2} \sum_{i,j} \delta_{i,j} \quad . \quad (3)$$

For parallel lumped conductors ($M=1$) the linearization is exact, otherwise it is an approximation. This simple model qualitatively reproduces the residual average shift observed when comparing IFM and "atomistic" results.

D. *Re-Calibration requirements*

The average shift in "atomistic" simulations poses a conceptual problem. The model introduced in Sec. C shows that it is a direct consequence of the non-linear nature of the problem that the average response of the "atomistic" simulation does not correspond to the result of the reference simulations. This problem highlights the fact that the "atomized" structures (with 1 or 0 dopants per element) are qualitatively different from the reference device with about 1/100 dopants in all elements. It is therefore necessary to perform a separate re-calibration for the perturbed devices when using "atomistic" approaches [8],[14]. This re-calibration requirement makes it hard to leverage the decades of experience of continuum TCAD and also makes it difficult to directly compare the standard TCAD results with "atomistic" variability predictions.

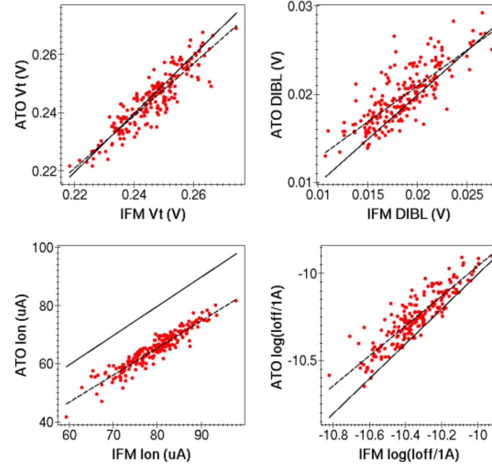

Figure 5: Correlation of threshold voltage, DIBL , on-state current, and logarithm of off-state current computed using the "atomistic" (ATO) and the statistical impedance field method (IFM).
Dashed line: linear regressions. Solid line: reference unity line.

Due to the linear nature of IFM it is automatically guaranteed that the average of any extracted quantity is the same as the corresponding value of the unperturbed reference device and consequently no re-calibration is necessary and the same set of well-established TCAD transport models and parameters can be used the reference device as well as for the variability analysis. We can therefore expect that at least for 1-2 standard deviations away from the average IFM will provide meaningful results and up to 3 standard deviations will still provide a good trend analysis.

Fig. 5 shows electrical parameters extract from IVs for 200 randomized MOSFET. The IVs are computed either using the "atomistic" or the sIFM method. Fig. 5 shows a clear correlation between the "atomistic" and the sIFM results. The systematic shift between the "atomistic" or the sIFM, which is particularly pronounces for the on-state current is mainly caused by the *conductor network non-linearity* effect, which also contributes to a slight deviation of the slope of the correlation from unity.

The distribution of the sIFM results are Gaussian by design. Observing that the correlation is of a "tight-cigar" type, this also means that, in spite of the non-linearities, the distribution of the "atomistic" results is also mainly Gaussian. We attribute that to the effect of the *Central Limit Theorem* (See Sec. E.).

E. *Distribution Tails*

While the linear nature of IFM makes the method unsuitable for the investigations of rare tail events,

experimental results [15] indicate that the variability of the threshold is indeed well-described by a Gaussian distribution. Even "atomistic" investigations have indicated that the dominant contributions to the threshold voltage variability results in mainly Gaussian-shaped distributions [4]. A linearization approach, such as the IFM method, is inherently incapable to make meaningful prediction about non-Gaussian tails due to non-linear effects. When performing "atomistic" simulations for the conductor network with a single chain ($N=1$) we indeed observe non-Gaussian distribution tails consistent with the non-linear nature of the problem. When adding more chains, however, the non-Gaussian distribution tails are diminished and for $N>10$ are no longer detectable. This observation is consistent with the *Central Limit Theorem* which states that self-averaging, here over the different conductor chains, results in Gaussian-like distributions, independent of the underlying statistics. For semiconductor devices the more-or-less statistical independent current paths from the source to the drain provide such a self-averaging mechanism, explaining why experimental observed distribution tails are often consistent with a Gaussian distribution.

F. Turn-Around Time

For the IFM the main computational overhead comes from the computations of the Green's function, which typically takes as long as the computation of the TCAD solution itself. The computation time for the linear response to a particular perturbation, however, is quite negligible. One can therefore compute the linear responses to for example 10,000 random perturbations in roughly the time that it would take to compute 2-3 "atomistic" responses. To obtain any statistically relevant results from "atomistic" a bare minimum of 200 3D TCAD simulations are needed – but often 1,000 [4] or even 100,000 [9] are used to obtain acceptable statistics. Because of the large computational effort "atomistic" simulations often resort to simplified geometries to speed up the individual TCAD simulation. For IFM on the other hand it is very feasible to include the actual potentially complex geometric details in the analysis. Further, "atomistic" simulations always have to be performed in 3D, even if the reference device structure of interest has a 2D symmetry. For IFM on the other hand, if the variability source is spatially delta-correlated (as is the case for RDF) 3D variability results can be obtained from 2D simulations without any loss of accuracy. In this case the integration over the third-dimension can performed analytically. This results in an additional tremendous speed up, when applicable.

III. Summary and Conclusion

IFM is based on linear response theory. This fact results in advantages and some disadvantages over "atomistic" approaches. The first advantage is turn-around time – a full analysis of variability effects takes only about 2-3 times as long as a standard continuum TCAD simulation run, while "atomistic" approaches would typically take 200-10,000 times as long. The second advantage is that IFM directly benefits from all calibration work done for the continuum device, while "atomistic" approaches require a possibly tedious re-calibration and even a distinctly different transport model parameter set.

A disadvantage of IFM is that as a linear response theory it is limited to "small" perturbation, when it is not a-priory clear how to verify the "smallness" of a given perturbation. While "atomistic" approaches can account for non-linarites of the transport model parameters, many of these non-linarities stem from convenient semi-empirical parameterizations and are not necessarily physical. It is therefore at least questionable if the "atomistic" results are more accurate compared to IFM in the limit of large perturbations.

In all the IFM method is a fast and attractive approach to investigate variability effects in semiconductor devices.

[1] A. Wettstein, O. Penzin, E. Lyumkis and W. Fichtner, "Random Dopant Fluctuation Modeling with the Impedance Field Method," in *Proc. SISPAD*, Sep. 2003, pp. 91-94.

[2] Sentaurus Device G-2012.06 User Guide. Synopsys Inc., Mountain View, CA, USA.

[3] K. El Sayed, A. Wettstein, S.D. Siemonov, E Lyumkis, and B. Polsky, "Investigation of the Statistical Variability of Static Noise Margins of SRAM Cells Using the Statistical Impedance Field Method", *IEEE Trans. Electron Devices*, vol. ED-59, no. 6, pp. 1738-1744, Jun. 2012

[4] G. Roy, A. Ghetti, A. Benvenuti, A. Erlebach, A. Asenov, "Comparative Simulation Study of the Different Sources of Statistical Variability in Contemporary Floating Gate Non-Volatile Memory," *IEEE Trans. Electron Devices*, vol. ED-58, no. 12, pp. 4155 - 4163, Dec. 2011.

[5] T. Hiramoto, M. Suzuki, X. Song, K. Shimizu, T. Saraya, A. Nishida, T. Tsunomura, S. Kamohara, K. Takeuchi, and T. Mogami, "Direct Measurement of Correlation Between SRAM Noise Margin and Individual Cell Transistor Variability by Using Device Matrix Array, " *IEEE Trans. Electron Devices,* vol. ED-58, no. 8, pp. 2249–2256, Aug. 2011.

[6] H.-S. Wong and Y. Taur, "Three-Dimensional "Atomistic" Simulation of Discrete Random Dopant Distribution Effects in Sub-0.1pm MOSFET's," in IEDM Tech. Dig., 1993, pp. 705-708.

[7] Y. Li, S.-M. Yu, J.-R. Hwang, and F.-L. Yang, "Discrete Dopant Fluctuations in 20-nm/15-nm-Gate Planar CMOS," IEEE Trans. Electron Devices, vol. 55, no. 6, pp. 1449-1455, Jun. 2008.

[8] G. Roy, A. R. Brown, F. Adamu-Lema, S. Roy, and A. Asenov, "Simulation study of individual and combined sources of intrinsic parameter fluctuations in conventional nano-MOSFETs," IEEE Trans. Electron Devices, vol. 53, no. 12, pp. 3063–3070, Dec. 2006.

[9] D. Reid, C. Millar, S. Roy, and A. Asenov, "Statistical Enhancement of the Evaluation of Combined RDD- and LER-Induced VT Variability: Lessons From 10^5 Sample Simulations," IEEE Trans. Electron Devices, vol. ED-58, no. 8, pp. 2257–2265, Aug. 2011.

[10] M. G. Ancona and G. J. Iafrate, "Quantum correction to the equation of state of an electron gas in a semiconductor," Physical Review B, vol. 39, no. 13, pp. 9536–9540, 1989.

[11] D. B. M. Klaassen, "A Unified Mobility Model for Device Simulation—I. Model Equations and Concentration Dependence," *Solid-State Electronics*, vol. 35, no. 7, pp. 953–959, 1992

[12] C. Lombardi, S Manzini, A. Saporito, and M. Vanzi, "A Physically Based Mobility Model for Numerical Simulation of Nonplanar Devices," IEEE Transactions on Computer-Aided Design, vol. 7, no. 11,pp. 1164–1171, 1988.

[13] J. W. Slotboom and H. C. de Graaff, "Bandgap Narrowing in Silicon Bipolar Transistors," IEEE Transactions on Electron Devices, vol. ED-24, no. 8, pp. 1123–1125, 1977.

[14] S.M. Amoroso, C.L. Alexander, S. Markov, G. Roy, and A. Asenov, "A Mobility Model Correction for 'Atomistic' Drift-Diffusion Simulation" Proc. SISPAD, 279 (2011)

[15] T. Tsunomura A. Nishida, F. Yano, A.T. Putra, K.Takeuchi, S. Inaba, S. Kamohara, K. Terada, T. Hiramoto, T. Mogami., "Analyses of 5σ V_{th} Fluctuation in 65nm-MOSFETs Using Takeuchi Plot," in *Symposium on VLSI Technology*, Honolulu, HI, USA, pp. 156–157, June 2008.