Statistical Modeling with the Virtual Source MOSFET Model

Li Yu, Lan Wei, Dimitri Antoniadis, *Ibrahim Elfadel, Duane Boning
Massachusetts Institute of Technology,*Masdar Institute of Science and Technology
Email: {yul09, lanwei}@mit.edu, daa@mtl.mit.edu, ielfadel@masdar.ac.ae, boning@mtl.mit.edu

***Abstract*—A statistical extension of the ultra-compact Virtual Source (VS) MOSFET model is developed here for the first time. The characterization uses a statistical extraction technique based on the backward propagation of variance (BPV) with variability parameters derived directly from the nominal VS model. The resulting statistical VS model is extensively validated using Monte Carlo simulations, and the statistical distributions of several figures of merit for logic and memory cells are compared with those of a BSIM model from a 40-nm CMOS industrial design kit. The comparisons show almost identical distributions with distinct run time advantages for the statistical VS model. Additional simulations show that the statistical VS model accurately captures non-Gaussian features that are important for low-power designs.**

I. Introduction

Continued scaling of CMOS technology has introduced increased variations of process and design parameters, which profoundly affect all aspects of circuit performance [1]. While statistical modeling addresses the need for high product yield and performance, it inevitably increases the cost of computation. This problem is further exacerbated as future digital design becomes larger and more complex. Therefore, the simplicity of device models is a key factor in effective statistical design flows. Current compact transistor models consist of a large number of parameters and complex equations which do capture many (if not all) of the physical short-channel effects, but significantly slow down the simulation speed [2]. A distinct benefit of the ultra compact, charge-based statistical virtual source (VS) MOSFET model is that it directly addresses both the complexity and simulation problems of statistical circuit analysis for nanoscale CMOS devices [3] [4]. Indeed, it provides a simple, physics-based description of carrier transport in modern short-channel MOSFETs along with the capability of mapping the variability characterization in device behavior onto a limited number of underlying model parameters, which in turn enables the efficient prediction of variations in circuit performance.

The core of the ultra compact VS model is a simple physical description of channel minority carrier charges at the virtual source. It essentially substitutes the quasi-ballistic carrier transport concept for the concept of drift-diffusion with velocity-saturation. In doing so, it achieves excellent accuracy for the I-V and C-V characteristics of the device throughout the various domains of circuit operation. The number of parameters needed is considerably fewer (11 for DC and 24 in total) than in conventional models.

In this paper, we present the first derivation and validation of the statistical VS model. The development of the model is centered on a statistical extraction technique called the Backward Propagation of Variance (BPV) [5]. Although this is performed for the nominal V_{dd}, the resulting statistical model is valid over a whole range of V_{dd}'s, thus enabling the efficient analysis of power-delay tradeoffs in the presence of parameter variations.

978-3-9815370-0-0/DATE13/©2013 EDAA

The method we describe in this paper is applied to characterize the within-die (e.g., geometry-dependent) variability component due to manufacturing variations. It is well known that for the deeply-scaled technologies (65-nm CMOS and beyond), where the VS model is most appropriate, within-die variations can dominate inter-die (i.e., global) variations. However, the general idea of BPV could be applied to inter-die variation as well.

II. Virtual Source Charge-based Compact Model

A. Review of the VS Model Equations

The core concept of the Virtual Source (VS) compact model is that as the MOSFET operation in saturation approaches the ballistic limit, the virtual source velocity v_{x_o} becomes independent of V_{ds} except for the drain-induced barrier lowering (DIBL) effects. This behavior is to be contrasted with the drift-diffusion transport model where the velocity is directly proportional to the electrical field E and becomes saturated as the electrical field E passes beyond a critical value.

In saturation, the drain current I_D is calculated as the product of the charge areal density Q_{ixo} and the channel-injected carrier velocity v_{x_o} at the virtual source

$$I_D = F_s \cdot Q_{ixo} \cdot v_{x_o} \tag{1}$$

The function F_s is to account for non-saturation and provides continuity across all regions of operation

$$F_s = \frac{V_{ds}/V_{dsat}}{(1 + (V_{ds}/V_{dsat})^{\beta})^{1/\beta}} \tag{2}$$

β is a fitting parameter with a typical value of 1.8 [3].

B. Parameter Variations in VS Model

To support statistical circuit simulation, the measured IV and CV statistics need to be converted into variations of a complete set of independent VS model parameters. For modern MOSFETs, the primary sources of within-die variations include random dopant fluctuation (RDF), line-edge roughness (LER) and oxide thickness fluctuation (OTF) as well as local fluctuations of mechanical stress. To maintain the simplicity of the statistical VS model, we relate most of its parameters directly to standard device measurements rather than to manufacturing process parameters. The VS model parameters used for statistical modeling are listed in Table I. In the VS model, the threshold voltage is modeled as

$$V_T = V_{T0} - \delta(L_{eff})V_{DS} \tag{3}$$

where $\delta(L_{eff})$ is the L_{eff}-dependent DIBL coefficient [3]. The threshold voltage variation in Table I is determined by the variations in implantation energy and dose as well as fluctuations in substrate doping. These effects are modeled through variation in V_{T0} while length-dependent threshold

TABLE I
VS MODEL PARAMETERS LIST

Source	Model Parameter	Description
LER	$L_{eff}(nm)$	Effective channel length
LER	$W_{eff}(nm)$	Effective channel width
RDF	V_{T0}	Zero-bias threshold voltage
OTF	$C_{inv}(\mu F/cm^2)$	Effective gate-to-channel capacitance per unit area
Stress	$\mu(cm^2/V \cdot s)$	Carrier mobility
Stress	$v_{xo}(cm/s)$	Virtual source velocity

voltage variation is captured through variation in $\delta(L_{eff})$. Note that V_{T0} has a weak dependency on L_{eff} over the range considered here, and therefore its effect is negligible. A special feature of the VS model is that v_{xo} is independent of the bias voltages. Previous work has shown that the relative change in virtual source velocity is related to the change in mobility [6]. According to [7], v_{xo} also has a dependency on $\delta(L_{eff})$. Therefore variation on L_{eff} also has an impact on v_{xo}. In the VS model, both effects are described using an approximation for the sensitivity of v_{xo} with respect to μ and $\delta(L_{eff})$ as shown in the following expression:

$$\frac{\Delta v_{xo}}{v_{xo}} = [\alpha+(1-B)(1-\alpha+\gamma)]\frac{\Delta\mu}{\mu}+\frac{\partial v_{xo}}{v_{xo}\partial\delta(L_{eff})}\Delta\delta(L_{eff}) \quad (4)$$

Here $\alpha \approx 0.5$ and $\gamma \approx 0.45$ are both fitting indices to a power law and B is the ballistic efficiency given by the expression

$$B = \lambda/(\lambda + 2l) \quad (5)$$

where λ is the mean free path and l is the critical length for backscattering to the source at nominal L_{eff}. An approximate value for $\frac{\partial v_{xo}}{v_{xo}\partial\delta(L_{eff})}$ in the targeted technology is 2.

III. Statistical Extraction Method

A well-characterized nominal VS model is the foundation of variability analysis. The nominal values of important effects, such as DIBL, mobility and virtual source velocity are critical for determining the model sensitivity to parameter variations. The basis for mismatch modeling was proposed in [8]. For local variation, the fluctuations in the observed variation of parameters have a uniform area dependency

$$\frac{\sigma_p^2}{p^2} \propto \frac{1}{LW} \quad (6)$$

where the subscript p represents a process parameter such as effective channel length and width. For local mismatch, we have $\sigma_L = \sigma_{L_{eff}}$ and $\sigma_W = \sigma_{W_{eff}}$ and a complete equation considering the geometric dependence of each parameter is

$$\begin{bmatrix} \sigma_{V_{T0}} \\ \sigma_L \\ \sigma_W \\ \sigma_\mu \\ \sigma_{C_{inv}} \end{bmatrix} = [\alpha_1\ \alpha_2\ \alpha_3\ \alpha_4\ \alpha_5] \begin{bmatrix} \frac{1}{\sqrt{WL}} \\ \sqrt{\frac{L}{W}} \\ \sqrt{\frac{W}{L}} \\ \frac{1}{\sqrt{WL}} \\ \frac{1}{\sqrt{WL}} \end{bmatrix} \quad (7)$$

The ultimate goal of this statistical modeling is to extract a group of α_{1-5} that is appropriate for all transistor geometries and that match the statistical circuit performance. The mismatch variances of p_j cannot be characterized directly from measurement or device simulations. Instead, variations σ_{e_i} $(i = 1, 2, ..., m)$ of electrical performance parameters (e.g., I_{dsat}, I_{off}, etc.) are measured under different geometry and bias conditions and the σ_{p_j} are calculated from BPV [5] according to the formula

$$\sigma_{e_i}^2 = \sum_{j=1}^{n}(\frac{\partial e_i}{\partial p_j})^2\sigma_{p_j}^2 \quad (8)$$

Here p_j and p_k for any $j \neq k$ is assumed to be independent since we successfully decoupled variation sources into independent variation parameters in the VS model in Section II(B). Equation (8) assumes Gaussian distributions for both groups of $\{e_i\}$ and $\{p_j\}$. This assumption requires a careful selection of both $\{e_i\}$ and $\{p_j\}$. In our work and unlike the statistical modeling approach of [5], e_i is not applicable for all bias conditions. Other bias conditions such as I_d at the transition region between linear and saturation, or I_{off} are not appropriate $\{e_i\}$ because they do not strictly follow a Gaussian distribution.

The accuracy of Equation (8) hinges on the validity of approximating the electrical performance parameters as linear functions of the process parameters. We have found that such linear approximation is sufficiently accurate to extract σ_{p_j}.

A system of linear equations is set up after stacking a group of equations with different transistor sizes, as is shown in (9). The sensitivity matrix in (9) is calculated from SPICE simulation using the VS model. Virtual source velocity is not considered as a separate variation parameter in Equation (9) since its effect has been captured in the variation of L_{eff} and μ. Also, silicon dioxide films are created with a thermal oxidation process which historically has been extremely tightly controlled [9] with the σ variation of C_{inv} being less than 0.5% in our case. Because the BPV process tends to overestimate variation in tightly controlled process parameters, we directly measure C_{inv} through the oxide thickness, as suggested in [10].

Since the primary intrinsic mismatch corresponding to gate length and width variation is due to line edge roughness (LER), which is caused by etching and sub-wavelength photolithographic process, it is reasonable to assume the same roughness for both length and width. Therefore an empirical relationship $\alpha_2 = \alpha_3$ ($\sigma_L/\sigma_W = L/W$) is assumed to further reduce the unknown parameters in (9). A good match to data is achieved ($\alpha_2/\alpha_3 = 0.95 - 0.99$ under different geometries) in a 40-nm CMOS technology.

IV. Verification

To validate the accuracy of the VS statistical model as well as the statistical extraction method, we implement it using Verilog-A under the Cadence Virtuoso Design Environment. The method described in Section III was applied to characterize the SPICE-level benchmark circuit statistics of a 40-nm bulk CMOS technology. Although the BPV method is applicable to measurement data, here we have employed a BSIM based industrial design kit to validate the proposed VS statistical model. Various Monte Carlo simulations were performed, including several geometries of MOSFETs and different electrical tests (IV and CV). The sample sizes are more than 1000 to characterize the statistical variation and correlation for e_i. The extracted parameter statistics α_{1-5} are listed in Table II.

A. Validation of Device Variability

The percentage differences of σ/μ for I_{dsat} mismatch and the underlying process parameter contributions are shown in Fig. 1(a). Compared with previous results in a similar technology [11], we observe a similar extracted $\sigma_{V_{T0}}/\mu_{V_{T0}}$ and $\sigma_{L_{eff}}/\mu_{L_{eff}}$ but smaller σ_μ/μ_μ in the VS model. The latter result is due to the fact that in the context of the VS model,

$$\begin{bmatrix} \begin{pmatrix} \sigma^2_{I_{dsat}} - (\frac{\partial I_{dsat}}{\partial C_{inv}})^2\sigma^2_{C_{inv}} \\ \sigma^2_{log_{10}I_{off}} - (\frac{\partial log_{10}I_{off}}{\partial C_{inv}})^2\sigma^2_{C_{inv}} \\ \sigma^2_{C_{gg@V_g}} - (\frac{\partial C_{gg@V_g}}{\partial C_{inv}})^2\sigma^2_{C_{inv}} \end{pmatrix}_1 \\ \cdots \\ \begin{pmatrix} \sigma^2_{I_{dsat}} - (\frac{\partial I_{dsat}}{\partial C_{inv}})^2\sigma^2_{C_{inv}} \\ \sigma^2_{log_{10}I_{off}} - (\frac{\partial log_{10}I_{off}}{\partial C_{inv}})^2\sigma^2_{C_{inv}} \\ \sigma^2_{C_{gg@V_{dd}}} - (\frac{\partial C_{gg@V_{dd}}}{\partial C_{inv}})^2\sigma^2_{C_{inv}} \end{pmatrix}_m \end{bmatrix} = \begin{bmatrix} \begin{pmatrix} (\frac{\partial I_{dsat}}{\partial V_{T0}})^2\frac{1}{WL} & (\frac{\partial I_{dsat}}{\partial L})^2\frac{L}{W} & (\frac{\partial I_{dsat}}{\partial W})^2\frac{W}{L} & (\frac{\partial I_{dsat}}{\partial \mu})^2\frac{1}{WL} \\ (\frac{\partial log_{10}I_{off}}{\partial V_{T0}})^2\frac{1}{WL} & (\frac{\partial log_{10}I_{off}}{\partial L})^2\frac{L}{W} & (\frac{\partial log_{10}I_{off}}{\partial W})^2\frac{W}{L} & (\frac{\partial log_{10}I_{off}}{\partial \mu})^2\frac{1}{WL} \\ 0 & (\frac{\partial C_{gg@V_{dd}}}{\partial})^2\frac{L}{W} & (\frac{\partial C_{gg@V_{dd}}}{\partial W})^2\frac{W}{L} & (\frac{\partial C_{gg@V_{dd}}}{\partial \mu L})^2\frac{1}{WL} \end{pmatrix}_1 \\ \cdots \\ \begin{pmatrix} (\frac{\partial I_{dsat}}{\partial V_{T0}})^2\frac{1}{WL} & (\frac{\partial I_{dsat}}{\partial L})^2\frac{L}{W} & (\frac{\partial I_{dsat}}{\partial W})^2\frac{W}{L} & (\frac{\partial I_{dsat}}{\partial \mu})^2\frac{1}{WL} \\ (\frac{\partial log_{10}I_{off}}{\partial V_{T0}})^2\frac{1}{WL} & (\frac{\partial log_{10}I_{off}}{\partial L})^2\frac{L}{W} & (\frac{\partial log_{10}I_{off}}{\partial W})^2\frac{W}{L} & (\frac{\partial log_{10}I_{off}}{\partial \mu})^2\frac{1}{WL} \\ 0 & (\frac{\partial C_{gg@V_g}}{\partial L})^2\frac{L}{W} & (\frac{\partial C_{gg@V_g}}{\partial W})^2\frac{W}{L} & (\frac{\partial C_{gg@V_g}}{\partial \mu})^2\frac{1}{WL} \end{pmatrix}_m \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix} \quad (9)$$

TABLE II
EXTRACTED STANDARD DEVIATION COEFFICIENT USING THE BPV METHOD

	NMOS	PMOS
α_1 $(V \cdot nm)$	2.3	2.86
α_2 (nm)	3.71	3.66
α_3 (nm)	3.71	3.66
α_4 $(nm \cdot cm^2/V \cdot s)$	944	781
α_5 $(nm \cdot \mu F^2/cm^2)$	0.29	0.81

mobility and virtual source velocity have meanings that differ with those of [11]. I_{dsat} and $log_{10}I_{off}$ bivariate scatter plots for BSIM model and 1σ, 2σ and 3σ confidence ellipses for both the VS and BSIM models are shown in Fig. 1(b). Note that in the statistical VS model, the generated variation parameters L_{eff}, V_{T0}, and μ are non-correlated. This behavior confirms that the I_{dsat} and $log_{10}I_{off}$ variations are fully decoupled during the statistical extraction procedure.

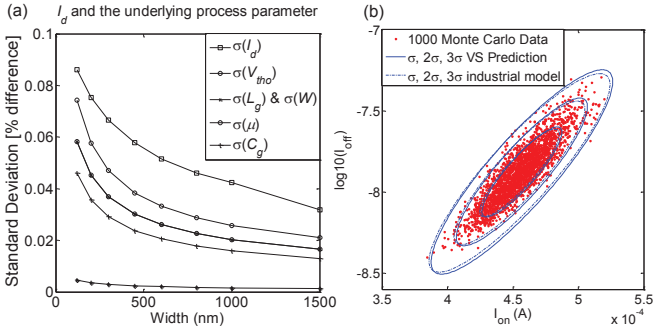

Fig. 1. (a) I_{dsat} mismatch and the underlying process parameter contributions for $L = 40nm$; (b) comparison of 1000 Monte Carlo simulation results for medium device ($W/L = 600nm/40nm$) between VS and BSIM statistical model. 1σ, 2σ and 3σ confidence ellipses for both model are also shown. The solid box represents $\pm 3\sigma$ limits for each variable from the BSIM model.

B. Statistical Validation Using Benchmark Circuits

We have performed statistical experiments on both the BSIM model and the VS model using a set of benchmark circuits, including standard library logic cells (INV, NAND2, DFF, etc.) and an SRAM cell.

Our first standard cell is a fanout-of-3 static INV gate having different geometries, as shown in Fig. 2. The V_{dd} in all cases is $0.9V$ which is the standard supply voltage for this particular technology. Excellent matching is achieved across a wide range of transistor sizes, which confirms that the geometric dependencies of the VS variation are well characterized. It is important to note that our statistical extraction procedure remains valid regardless of the specific functional dependence of the variations on device geometry.

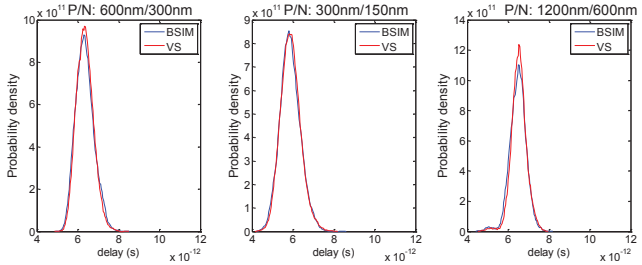

Fig. 2. Delay probability density comparison of 2500 Monte Carlo simulations for an INV gate (fanout of 3) with different sizes ($1\times$, $2\times$ and $4\times$).

Our second standard cell is a fanout-of-3 static NAND2 gate operating under a V_{dd} of $0.9V$, $0.7V$ and $0.55V$. Although power consumption decreases with supply voltage, local variations increase significantly, and as a result parametric yield is decreased. Even worse, the probability density of the delay becomes highly non-Gaussian at low supply voltage, and as a result, the application of statistical static timing analysis (SSTA) becomes more difficult [12]. Although all variation parameters in the VS model are assumed to be independent Gaussian variables, the non-Gaussian property of the delay distribution is correctly captured, as is shown in Fig. 3. The quantile-quantile plot for delay variation starts to deviate from a linear relationship when $V_{dd} = 0.7V$, and the non-linearity becomes pronounced at $V_{dd} = 0.55V$. In both cases, the VS prediction shows a good match with the BSIM model at the 3σ scale. Unlike the PSP model [13] where variances of extra electrical performance parameters have to be added to match the variance at different V_{gs}, no extra statistical fitting is needed in the VS model to adjust timing distributions in cases dynamic voltage scaling is used.

After verifying the approach on combinational logic cells, we now extend it to perform setup and hold time analysis on a D flip-flop. The schematic of the benchmark master-slave register is shown in Fig. 4(a). Fig. 4(b) shows a typical timing path for setup/hold analysis. The PDF's for setup/hold time for the registers simulated from VS model and BSIM models are shown in Fig. 4(c). One important note is that the characterization of the setup/hold time requires about 20 times more SPICE simulations than those of a combinational cell having the same number of transistors. This is because the setup/hold time can only be measured indirectly by varying clock to input signal delay. The ultra compact VS model plays a more important role in this case where tens of thousands of SPICE simulations are required.

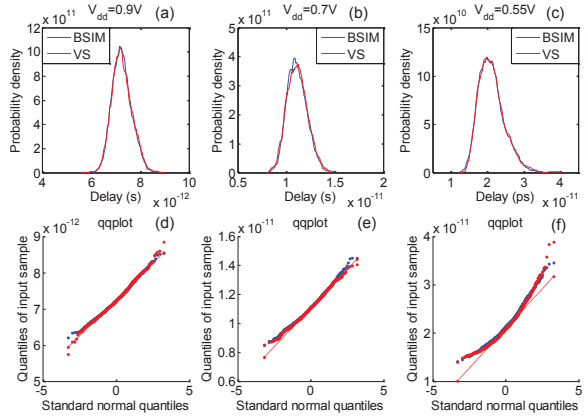

Fig. 3. Delay probability density comparison between BSIM and VS model for an NAND2 gate (fanout of 3) with a supply voltage of (a) $0.9V$, (b) $0.7V$ and (c) $0.55V$. The quantile-quantile plot for delay variation under each supply voltage in (d) $0.9V$, (e) $0.7V$ and (f) $0.55V$ shows a strongly nonlinear pattern in low power application.

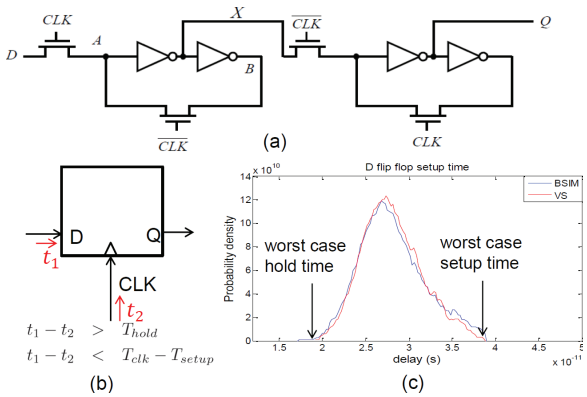

Fig. 4. (a) Master-slave register based on NMOS-only pass transistors, P/N sizes are $600nm/40nm$ and $300nm/40nm$, separately; (b) typical timing path for setup/hold analysis; and (c) probability density of the setup time in circuit (a) with 250 Monte Carlo runs.

The last circuit in our validation is a 6T SRAM cell, which is known to be highly sensitive to within-die variations, as shown in Fig. 5. Both the VS and BSIM models are employed to simulate the variability in SRAM READ and HOLD Static Noise Margin (SNM). The characteristic butterfly patterns generated with the statistical VS model are shown in Fig. 5 (a) and (d), for READ and HOLD, respectively. The SNM comparisons between the two models for READ and HOLD are shown in Fig. 5 (b) and (e). Even with this highly sensitive analog circuit, the ultra compact statistical VS model provides an excellent match to the "golden" BSIM model. In Fig. 5 (f), the quantile-quantile plot for SRAM HOLD SNR using both models shows a slightly non-Gaussian distribution.

Finally, the runtime speedup of the VS model (Verilog-A) with respect to BSIM4 (C code) is shown in Table III. We notice a $4.2\times$ speedup and $8.7\times$ reduction in memory usage. These favorable results can be further improved using an optimized C code implementation of the VS model in line with the optimized C code used for BSIM4.

V. Conclusion

In this paper, we have described the first statistical extension of the ultra-compact Virtual Source (VS) MOSFET model. The derivation of the statistical model is based on the backward

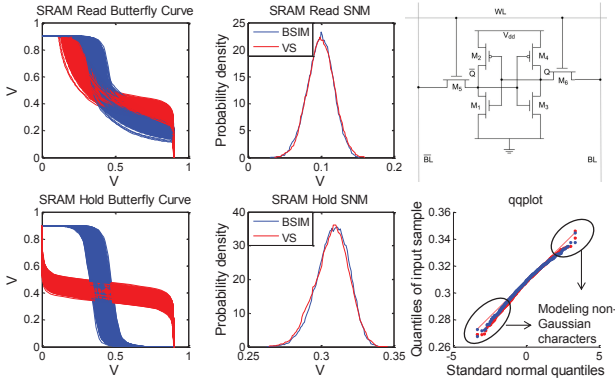

Fig. 5. 2500 Monte Carlo simulation for a miminum sized 6T SRAM cell; (a) butterfly pattern from VS model in static READ mode; (b) probability density for SRAM READ static noise margin (SNR); (c) schematic of the 6-T SRAM; (d) butterfly pattern from VS model in static HOLD mode;(e) probability density for SRAM HOLD SNR; and (f) quantile-quantile plot for SRAM HOLD SNR.

TABLE III
Speed and Memory Comparison for Monte Carlo Simulation between VS (in Verilig-A code) and BSIM4 model (in C code)

			VS		BSIM 4	
Cell	Sim.	Sample	Runtime	Memory	Runtime	Memory
NAND2	Tran	2000	$225s$	$14.9M$	$855s$	$126M$
DFF	Tran	250	$3.86ks$	$23.2M$	$13.5ks$	$157M$
SRAM	AC	2000	$405s$	$17M$	$2.15ks$	$187M$

propagation of variance (BPV), and nanometer-regime variation sources are mapped onto independent VS model parameters. The statistical VS model is validated in reference to a "golden" 40nm BSIM model using extensive Monte Carlo runs.

Acknowledgment

This work was funded by the Cooperative Agreement Between the Masdar Institute of Science and Technology and the Massachusetts Institute of Technology.

References

[1] D. Boning et al. Variation. In *ISQED*, pages 15–20, Mar. 2007.
[2] B.H. Calhoun et al. Digital circuit design challenges and opportunities in the era of nanoscale CMOS. *Proc. IEEE*, 96(2):343–365, Feb. 2008.
[3] A. Khakifirooz et al. A simple semiempirical short-channel MOSFET current-voltage model continuous across all regions of operation and employing only physical parameters. *T-ED*, 56(8):1674–1680, Aug. 2009.
[4] L. Wei et al. Virtual-source-based self-consistent current and charge FET models: From ballistic to drift-diffusion velocity-saturation operation. *T-ED*, (99):1–9, 2012.
[5] P.G. Drennan et al. A comprehensive MOSFET mismatch model. In *IEDM*, pages 167–170, 1999.
[6] A. Khakifirooz et al. Transistor performance scaling: The role of virtual source velocity and its mobility dependence. In *IEDM*, pages 1–4, Dec. 2006.
[7] A. Lochtefeld et al. On experimental determination of carrier velocity in deeply scaled NMOS: how close to the thermal limit? *EDL*, 22(2):95–97, Feb. 2001.
[8] M.J.M. Pelgrom et al. Matching properties of MOS transistors. *JSSC*, 24(5):1433–1439, Oct. 1989.
[9] M. Orshansky et al. *Design for Manufacturability and Statistical Design: A Constructive Approach*. Springer, 2010.
[10] C.C. McAndrew. Statistical modeling for circuit simulation. In *ISQED*, pages 357–362, Mar. 2003.
[11] W. Zhao et al. Rigorous extraction of process variations for 65nm CMOS design. In *ESSDERC*, pages 89–92, Sept. 2007.
[12] A. Agarwal et al. Statistical timing analysis for intra-die process variations with spatial correlations. In *ICCAD*, pages 900–907, Nov. 2003.
[13] X. Li et al. Statistical modeling with the PSP MOSFET model. *T-CAD*, 29(4):599–606, April 2010.