An ultra-compact virtual source FET model for deeply-scaled devices: Parameter extraction and validation for standard cell libraries and digital circuits

An Ultra-Compact Virtual Source FET Model for Deeply-Scaled Devices: Parameter Extraction and Validation for Standard Cell Libraries and Digital Circuits

Li Yu, Omar Mysore, Lan Wei, Luca Daniel, Dimitri Antoniadis, *Ibrahim Elfadel, Duane Boning
Massachusetts Institute of Technology,*Masdar Institute of Science and Technology

Abstract— In this paper, we present the first validation of the virtual source (VS) charge-based compact model for standard cell libraries and large-scale digital circuits. With only a modest number of physically meaningful parameters, the VS model accounts for the main short-channel effects in nanometer technologies. Using a novel DC and transient parameter extraction methodology, the model is verified with simulated data from a well-characterized, industrial 40-nm bulk silicon model. The VS model is used to fully characterize a standard cell library with timing comparisons showing less than 2.7% error with respect to the industrial design kit. Furthermore, a 1001-stage inverter chain and a 32-bit ripple-carry adder are employed as test cases in a vendor CAD environment to validate the use of the VS model for large-scale digital circuit applications. Parametric V_{dd} sweeps show that the VS model is also ready for usage in low-power design methodologies. Finally, runtime comparisons have shown that the use of the VS model results in a speedup of about $7.6\times$.

I. Introduction

As technology scales down to the nanometer range, accuracy requirements on digital cell timing rules are becoming more stringent while the complexity of the transistor models used to generate these rules is steadily increasing. Existing BSIM [1], PSP [2] and PTM [3] models are being constantly augmented to account for the emerging physical phenomena at the nanometer regime. As a result, the management of transistor parameters in CAD environments or during library cell characterization for timing and power becomes a more complex undertaking. The issue becomes even more acute in the context of Monte Carlo simulations if we have to account for parameter variations of a large number of parameters. One way to address these issues for digital design in the nanometer regime is to make use of ultra-compact transistor models specifically developed for short- and ultra-short-channel devices. The Virtual Source (VS) model is one such ultra-compact model [4][5][6].

The widely adopted threshold-voltage-based compact models (such as BSIM [1] and PTM [3]) and the surface-potential based compact models (such as PSP [2]), include as many as several hundred parameters related to the manufacturing process, the geometry of the device, and to achieve smoothing or transitions between different equation regimes. On the other hand, the VS model restricts itself to a simple physical description for channel minority carrier charges at the virtual source by substituting 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 domain of operation required for digital timing and power analysis. The number of parameters needed is considerably fewer (11 for DC and 24 in total) than in conventional models. It is worth noting that the ultra-compact model developed in [7] is based on the alpha-power model of [8], which is purely empirical and aims at maximizing the timing accuracy of an inverter. On the other hand, the VS model is physics-based and achieves higher timing accuracy than [7] with a similar number of parameters.

The major contributions of our work are as follows:

- A consistent parameter extraction methodology for the static and dynamic parameters of the VS model.
- A robust Verilog-A implementation of a VS model calibrated to work in a 40nm digital design environment. Aside from the use of the VS model no other changes are needed in the CAD circuit simulation environment.
- An accurate timing validation of the VS model on standard library cells characterization with an average error less than 1% with respect to the industrial models.
- A systematic timing, power, and waveform validation of the VS model on two large-scale digital circuits in the presence of a parametric V_{dd} sweep. Finally, it is worth noting that as a result of the use of the VS ultra-compact model, we have observed a $7.5\times$ speedup with respect to the industrial MOSFET models.

II. Review of the virtual source charge-based compact model

A. Static VS Model

In the VS model, the drain current of a MOSFET normalized by width (I_D/W) can be described using the fol-

lowing general equation:

$$I_D/W = Q_{ix_o} v_{x_o} F_s \tag{1}$$

valid for both the saturation and non-saturation regions.

The virtual source velocity, denoted as v_{x_o}, refers to the velocity of carriers located in the MOSFET channel at the top of the energy barrier near the source (virtual source). The core concept in VS modeling is that in short-channel devices v_{x_o} does not depend on V_{ds} except for drain-induced barrier lowering (DIBL) effects. This is to be contrasted with a drift-diffusion transport model where the velocity is directly proportional to low electrical field E and is saturated when E is larger than the critical electrical field. The VS model also uses the fact that although ballistic velocity increases with V_{gs}, the virtual source velocity v_{x_o} is almost constant at high V_{gs} [9].

The VS inversion charge density Q_{ix_o} can be approximated by the empirical function [4][10]:

$$Q_{ix_o} = C_{inv} n\phi_t \ln(1 + \exp \frac{V'_{GS} - (V_T - \alpha\phi_t F_f)}{n\phi_t}) \tag{2}$$

where C_{inv} is the effective gate-to-channel capacitance per unit area in strong inversion, ϕ_t is the thermal voltage ($k_B T/q$), V'_{GS} represents the internal gate-source voltage and n is the subthreshold coefficient. The function F_f in (2) denotes a Fermi function that allows a smooth 0 to 1 transition, and α is introduced to adjust the V_T shift for which $3.5\phi_t$ is a good approximation.

The function F_s in (1) serves to account for the continuous transition from non-saturation to saturation and is given by

$$F_s = \frac{V'_{ds}/V_{dsat}}{(1 + (V'_{ds}/V_{dsat})^{\beta})^{1/\beta}} \tag{3}$$

where V'_{ds} accounts for the intrinsic drain-to source voltage after deducting the IR drop for both the source R_s and drain R_d resistances using $V'_{ds} = V_{ds} - I_D(R_s + R_d)$. β is an empirical parameter for the transition from low-field non-saturation region to high-field saturation region with a typical value of about 1.8 [4].

B. Dynamic VS Model

The transient behavior of the VS model is described in [6], where the intrinsic channel charge is partitioned into that of the source and drain terminals:

$$\begin{cases} Q_S = \int_0^{L_g} (1 - x/L_g) Q'_i(x) dx \\ Q_D = \int_0^{L_g} \frac{x}{L_g} Q'_i(x) dx \end{cases} \tag{4}$$

The channel charge areal density, $Q'_i(x)$, is calculated according to [6], using a non-saturation (NVsat), saturation (Vsat), or quasi-ballistic (QB) model. Since the nominal transistor gate length in this work is 40 nm, we have used the quasi-ballistic version of the channel charge model.

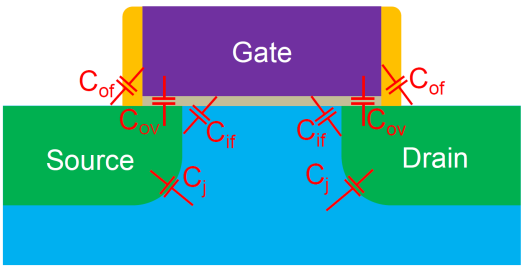

Fig. 1. Schematic of a short-channel n-MOSFET showing the VS model parasitic capacitances.

Under the assumption that the source and drain are symmetric, we introduce four capacitances to model parasitic effects, as shown in Fig. 1. C_{ov} is the overlap capacitance, C_{of} is the outer-fringing capacitance, C_{if} is the inner-fringing capacitance and C_j is the junction capacitance. In this work, C_{ov} and C_{of} are considered voltage independent while C_{if} and C_j are considered voltage dependent.

III. Parameter Extraction and Model Calibration

The VS model described in (1) through (4) is calibrated using data for transistors with different sizes. To extract all of the parameters of the VS model, a full set of I-V and C-V measurement data is needed. The parameter extraction flow is described in Section A, followed by validation through analysis of I-V curves in Section B.

A. Parameter Extraction Flow

While previous work has identified the key VS model parameters that need to be extracted from measurements [4] [6], a systematic, consistent and optimized flow to extract the full I-V and C-V model parameters has been lacking. One of the major contributions of this paper is to describe, implement and test such a model extraction flow. As shown in Fig. 2, our novel parameter extraction method consists of the following steps.

1. First, the effective gate-to-channel C_{inv} is extracted by subtracting the C_{gs} curves of two long channel devices where the short-channel parasitic capacitances are negligible. This step needs to be done before DC parameter extraction since C_{inv} affects the distribution of charges Q_{ix_o}.

2. Once C_{inv} is properly extracted, the I-V curve calibration is achieved under a separate flow in both subthreshold and full regions. First, V_{th0} is adjusted to achieve consistency with respect to Q_{ix_o}. Then the

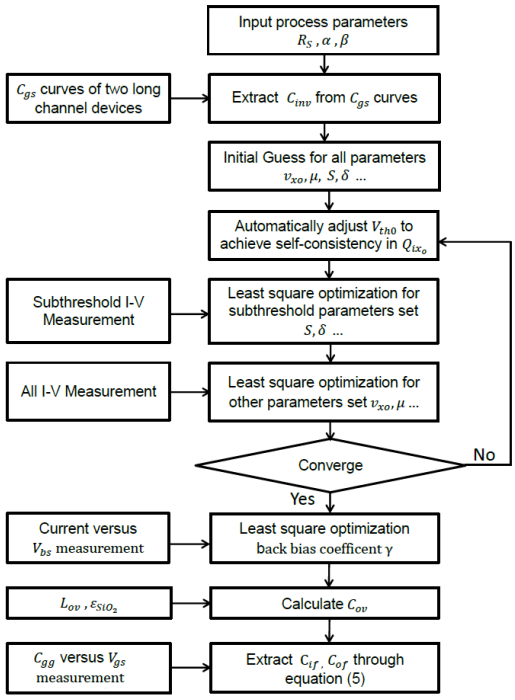

Fig. 2. A novel optimization flow for I-V parameter extraction in the VS model.

sub-threshold parameter set (S, δ) and full region parameter set (v_{xo}, μ, etc.) are optimized separately using non-linear least-squares error minimization. The solution is iterated until convergence. In most of our tests, good convergence results have been achieved within 5 iterations for transistors with various sizes.

3. The back bias coefficient is extracted from the IV_{bs} measurement as the last step in DC parameters extraction procedure.

4. We then extract the parasitic capacitances (C_{if} and C_{of}) by fitting the $C_{gg} - V_g$ curve, as shown in Fig. 3(a). The equations employed to extract the parasitics of the VS model are

$$\begin{cases} C_{gg}(V_g = 0) = 2W(C_{if} + C_{of} + C_{ov}) \\ C_{gg}(V_g = V_{dd}) = W[(C_{inv}(L - L_{ov}) + 2C_{of} + 2C_{ov})] \end{cases} \quad (5)$$

Table I lists the key parameters of the VS model and the parasitic capacitances obtained from the parameter extraction methodology as illustrated in Fig. 2. Note that the VS model calculates the drain current normalized by width, so that the extracted parameters are applicable for all rectangular devices having the same channel length. Good consistency and accuracy are achieved in devices with various widths using a single parameter set extracted by the aforementioned parameter extraction flow. This scalability is illustrated in Fig. 3(b). Since the $40nm$

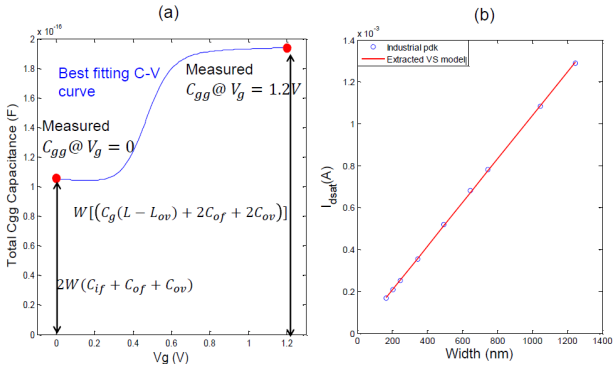

Fig. 3. (a) Gate capacitance versus V_{gs} at $V_{ds} = 0$V with the equation used to extract parasitic capacitances. (b) I_{dsat} versus effective channel width

channel is close to the ballistic transport regime, a quasi-ballistic (QB) model is employed to calculate the channel charge areal density [6]. We note, however, that for this technology, the non-saturation (NVsat) model achieves a similar accuracy.

TABLE I
KEY PARAMETERS FOR THE VS MODEL FITTED TO A $40nm$ INDUSTRIAL DESIGN KIT. CHANNEL WIDTHS ARE $300nm$ AND $600nm$ FOR NFET AND PFET, RESPECTIVELY.

Parameters	NMOS	PMOS	Description
$L_g(nm)$	40	40	Channel length drawn
$L_{ov}(nm)$	8	8	Total overlap channel length on both sides
$C_{inv}(\mu F/cm^2)$	1.40	1.35	Effective gate-to-channel capacitance per unit area
$S(mV/dec)$	89	98	Subthreshold swing
$DIBL(mV/V)$	93	159	Drain-induced barrier lowering
$v_{xo}(cm/s)$	1.39e7	0.855e6	Virtual source velocity
$\mu(cm^2/V \cdot s)$	248	146	Low-field mobility
$R_s(ohm - \mu m)$	60	80	Series resistance per side
γ	0.34	0.39	Body effect coefficient
m^*	$0.2m_e$	$0.2m_e$	Carrier effective mass

B. I-V Curve Analysis

Once all the parameters are extracted, the VS model is validated by comparing its I-V curve with that of a BSIM4 model from a $40nm$ bulk industrial design kit, as shown in Fig. 4. We see good agreement in both sub-threshold and above threshold regions for both NFETs and PFETs. The accuracy of the VS model fitting is comparable to other popular industrial models [2][3] and much better than other ultra-compact models [7] with similar complexity as the VS model. Previous work has demonstrated that the VS model has indeed good DC agreement with real measurement data fabricated in various nodes (32-nm, 45-nm, 65-nm) and processes (poly-SiON gate, high-k metal-gate) from various foundries (IBM, Intel) [4][6].

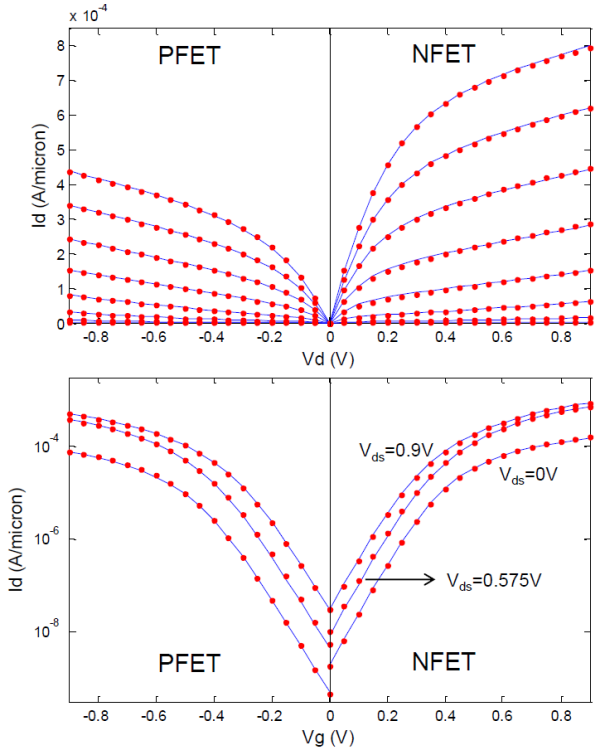

Fig. 4. VS model fitting with data from a 40nm BSIM4 industrial design kit. The channel width is $300nm$ for NFET, $600nm$ for PFET.

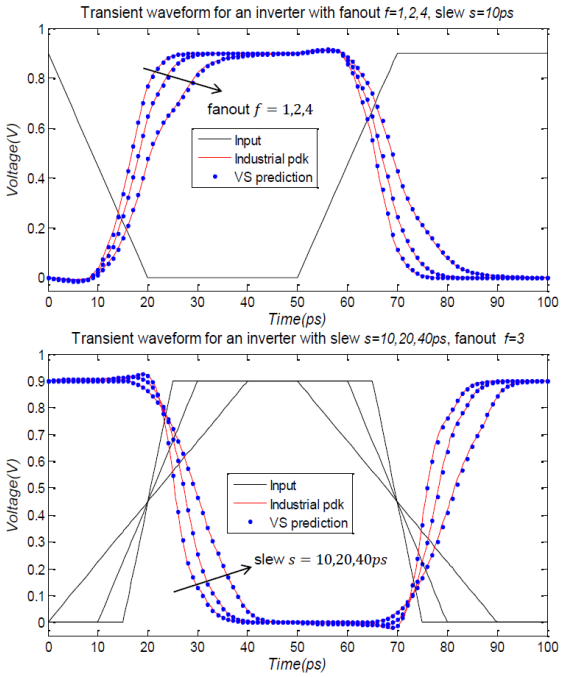

Fig. 5. Transient response waveform for an inverter chain with various input slews and fanouts.

However, systematic validation of the VS model for timing and power verification of digital circuits has been missing so far in the literature; the results of this paper are the first such validation. Once the device I-V and C-V curves are calibrated, we can proceed to timing or power comparison for standard library cells and other large-scale digital circuits. This is illustrated in the next two sections.

IV. Standard Cell Characterization Using Calibrated VS Model

To validate the accuracy of the VS model as calibrated in the previous section, we implement it using Verilog-A under the Cadence Virtuoso Design Environment. We then use it to characterize the SPICE-level circuits of a set of generic library standard cells from an industrial design kit in 40-nm bulk CMOS technology. The first circuit we consider is an inverter undergoing trapezoidal input transitions. This basic example is used to illustrate several important features of our calibrated VS model. It is well known that the charging and discharging activities during input gate transitions require precise balancing of both static and dynamic behavior of the NFET and PFET transistors. The output voltage waveforms using the VS model in comparison with the industry-standard BSIM4 model are depicted in Fig. 5. The conducted tests are similar to those used in static timing cell characterization as they have sweeps with respect to both output loads and input slews. In the first sweep test, the input slew is fixed at $10ps$ and the load (fanout) is 1, 2 and 4, while in the second test, the load (fanout) is fixed at 3 and the input slew rate is $10ps$, $20ps$, and $40ps$. The average delay error between the VS model and the "golden" or baseline BSIM4 model is 0.88% and the $10\% - 90\%$ rising/falling time error is 0.92%/1.11%. This is a good indication of the accuracy of the transient calibration of the proposed VS model.

TABLE II
Delay (in ps)comparison between VS and BSIM4 for various gates with input slew of $10ps$ and a fanout of 3.

	VS		BSIM 4		Error	
	D_{l-h}	D_{h-l}	D_{l-h}	D_{h-l}	E_{l-h}	E_{h-l}
INV	5.75	5.59	5.71	5.62	0.7%	0.5%
$NAND2_1$	6.3	9.86	6.3	10.2	0.1%	2.6%
$NAND2_2$	6.97	11.5	6.97	11.3	0.1%	2.3%
$NAND3_1$	7.4	16.2	7.4	16.5	0.1%	1.9%
$NAND3_2$	7.25	16.2	7.23	15.9	2.7%	2%
$NAND3_3$	7	15.1	6.97	14.8	0.4%	2.5%
$NOR2_1$	9.92	6.4	10.1	6.4	1.6%	0.1%
$NOR2_2$	8.84	6.12	8.93	6.08	0.9%	0.6%
$NOR3_1$	15.3	6.96	15.58	6.96	1.9%	0.1%
$NOR3_2$	14.8	6.81	15.01	6.8	1.4%	0.1%
$NOR3_3$	13.0	6.49	13.07	6.44	0.4%	0.7%

In Tables II and III we summarize the computed delays and rise/fall times for 2- and 3-input symmetrical NAND/NOR gates. The input slew is $10ps$ and the out-

put load is a fanout of 3. The average and the maximum relative error, VS vs. BSIM4, for all gates under test are 1.5% and 2.6%, respectively. This compares favorably with [7] where that ultra-compact model achieves only 6% (delay) and 11% (slew) accuracy on a similar set of tests.

TABLE III
SLEW COMPARISON BETWEEN VS AND BSIM4 FOR VARIOUS GATES WITH INPUT SLEW OF $10ps$ AND A FANOUT OF 3.

(ps)	VS		BSIM 4		Error	
	rise	fall	rise	fall	E_r	E_f
INV	9.13	8.84	9.03	8.67	1.1%	2.0%
$NAND2_1$	10.3	17.7	10.5	18.1	1.7%	2.2%
$NAND2_2$	11.4	18	11.2	17.6	1.6%	2.2%
$NAND3_1$	12.3	27.8	12.2	27.1	0.9%	2.5%
$NAND3_2$	12.1	27.8	11.9	27.1	1.6%	2.5%
$NAND3_3$	11.8	27.8	11.6	27.1	1.9%	2.5%
$NOR2_1$	16.0	9.94	15.7	9.82	1.0%	1.2%
$NOR2_2$	16	9.37	15.9	9.25	0.7%	1.3%
$NOR3_1$	25.0	11.22	24.9	11.2	0.7%	0.2%
$NOR3_2$	25.0	10.51	24.8	10.38	0.8%	1.2%
$NOR3_3$	24.9	10.07	24.6	9.9	1.3%	1.7%

V. VS MODEL VALIDATION FOR DIGITAL CIRCUITS

To further verify the calibrated VS model, a 32-bit ripple-carry adder is designed in the targeted technology (40-nm CMOS) and the transient waveform of the critical path compared using VS and BSIM4 models. The simulation environment and the SPICE convergence setting using both models are exactly the same; this is important to demonstrate that no major work is required to adapt the circuit simulation environment to the presence of the new transistor model. The test circuit includes 0.9k transistors in total belonging to various library cell types (INV, NAND, NOR and XOR). We select the worst-case delay for a 32-bit add operation. This requires setting input A at 100...00 and input B at 111...11. The input carry on signal of the very first bit C_{in0} has a $0 \rightarrow 1$ transition and then a $1 \rightarrow 0$ transition, and the output carry on signal of the very last bit will reflect the critical path delay. To show the robustness of the VS model for low-power design, the supply voltage V_{dd} is swept from $0.6V$ to $0.9V$. The transient signal C_{in0} and C_{out32} at different V_{dd} from both VS and BSIM4 model are shown in Fig. 6, which demonstrates that the output signals of the two models have excellent matching. The average delay mismatch under all V_{dd} conditions is about 0.3%.

The second digital circuit we consider is a 1001-stage inverter chain designed in the same technology. This test circuit includes 2k transistors in total and the delays under different V_{dd} from $0.6V$ to $0.9V$ are compared between the VS and BSIM4 models. The average delay mismatch under all V_{dd} is about 0.25%. In both digital circuit cases, the delay mismatch for VS vs. BSIM4 model is smaller

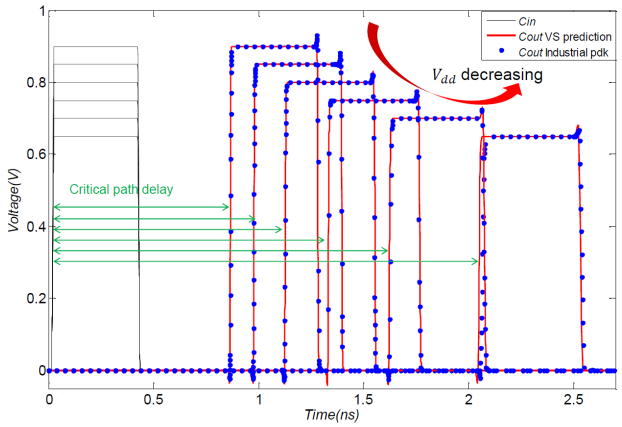

Fig. 6. Critical path transient waveform for a 32-bit ripple-carry adder with V_{dd} from $0.65V$ to $0.9V$.

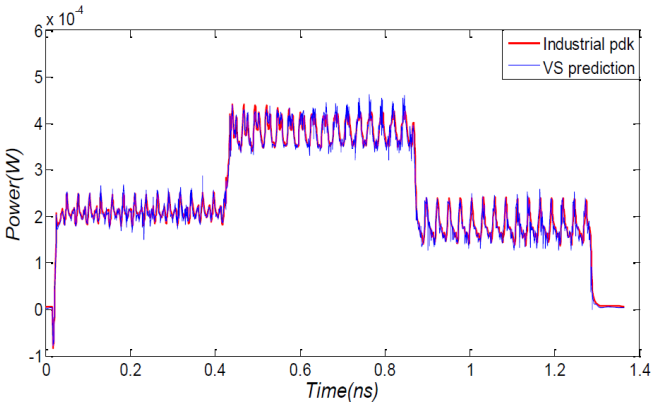

Fig. 7. Transient power consumption for a 32-bit ripple-adder with V_{dd}=0.9V.

than the mismatch observed in library cell delays. This is because the rise/fall mismatches tend to cancel each other in the inverter chain case.

TABLE IV
TRANSIENT SIMULATION SPEED COMPARISON BETWEEN VS MODEL AND BSIMSOI MODEL[11]

Model	VS run time	BSIMSOI run time	VS speed up
1001-stage inverter chain	$621s$	$3470s$	5.6×
32-bit ripple-carry adder	$111s$	$1060s$	9.6×

Power consumptions of the critical path transitions in the aforementioned test cases are also compared. Transient power consumed by the 32-bit ripple-adder under $V_{dd} = 0.9V$ is shown in Fig. 7, which demonstrates good agreement between the VS and BSIM4 models. The average power consumption mismatch under all V_{dd}'s is 1.3% for the 32-bit adder and 1.8% for the 1001-stage inverter chain. Finally, the power-delay curves for both cases under different V_{dd} are shown in Fig. 8

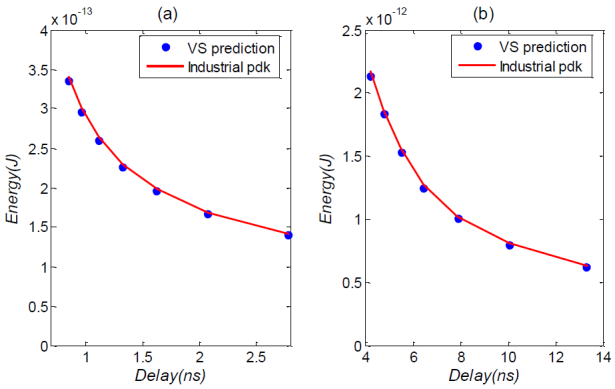

Fig. 8. Energy delay curve for (a) a 32-bit ripple-adder and (b) 1001 stages inverter chain under V_{dd} from 0.65 to 0.9V.

The runtime speedup of the VS model is further compared with an open source BSIM SOI compact model implemented in Verilog-A [11], which has a model complexity similar to BSIM4. The main reason for using the BSIM SOI for runtime comparison is because the industrial BSIM4 has been implemented in C, which is more computationally efficient than Verilog-A. The transient simulation runtime comparison in the two aforementioned test circuits are shown in Table IV. An average speed up of 7.6× is achieved which is in line with the order of magnitude reduction in the number of VS parameters. The simulation environment, the SPICE convergence setting and maximum iteration setting for both models are exactly the same for fair comparison. Also, in both models, the transition time and the delay time of the library cells are tuned to be similar to ensure a comparable computing effort in both cases. Since the simplicity of device models is key to a statistical design flow [12], the 7.6× speed up achieved by VS model points to its great potential for highly-efficient variation-aware statistical analysis.

VI. Conclusion

In this paper, an ultra-compact virtual source (VS) transport model which includes the main short-channel effects in nanometer MOS transistors has been developed and validated. The model has been coded in Verilog-A, implemented in a vendor CAD environment, and compared with that of a 40-nm industrial design kit. According to our SPICE-level simulation results, the model tracks, to within 2.7%, a much more complex reference model having an order of magnitude more parameters than the VS model. The VS model has also been used to characterize the timing of a standard cell library with excellent matching to that of the industrial design kit. Simulations of two large-scale digital circuits demonstrate that the VS model has been enabled to support SPICE-level timing and power analysis at an industrial degree of accuracy while having an order of magnitude fewer parameters than the BSIM4 industry standard. Further simulations have shown that the use of the VS model results in a runtime speedup of about 7.5×. In summary, this paper provides a solid validation for using the VS model in transistor-level digital circuit design and verification in a deeply scaled technology node. Its usage has been validated for both high-performance and low-power design methodologies.

Acknowledgment

This work was funded by the Cooperative Agreement Between the Masdar Institute of Science and Technology (Masdar Institute), Abu Dhabi, UAE and the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, Reference No. 196F/002/707/102f/70/9374.

References

[1] M. Chan, K.Y. Hui, C. Hu, and P.K. Ko. A robust and physical BSIM3 non-quasi-static transient and ac small-signal model for circuit simulation. *IEEE Transactions on Electron Devices*, 45(4):834 –841, Apr. 1998.

[2] G. Gildenblat, X. Li, W. Wu, H. Wang, A. Jha, R. van Langevelde, G.D.J. Smit, A.J. Scholten, and D.B.M. Klaassen. PSP: An advanced surface-potential-based mosfet model for circuit simulation. *IEEE Transactions on Electron Devices*, 53(9):1979 –1993, Sept. 2006.

[3] W. Zhao and Y. Cao. New generation of predictive technology model for sub-45 nm early design exploration. *IEEE Transactions on Electron Devices*, 53(11):2816 –2823, Nov. 2006.

[4] A. Khakifirooz, O.M. Nayfeh, and D. Antoniadis. A simple semiempirical short-channel MOSFET current-voltage model continuous across all regions of operation and employing only physical parameters. *IEEE Transactions on Electron Devices*, 56(8):1674 –1680, Aug. 2009.

[5] C. Jeong, D.A. Antoniadis, and M.S. Lundstrom. On backscattering and mobility in nanoscale silicon mosfets. *IEEE Transactions on Electron Devices*, 56(11):2762 –2769, Nov. 2009.

[6] L. Wei, O. Mysore, and D. Antoniadis. Virtual-source-based self-consistent current and charge FET models: From ballistic to drift-diffusion velocity-saturation operation. *IEEE Transactions on Electron Devices*, PP(99):1 –9, 2012.

[7] E. Consoli, G. Giustolisi, and G. Palumbo. An accurate ultra-compact i-v model for nanometer MOS transistors with applications on digital circuits. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 59(1):159 –169, Jan. 2012.

[8] T. Sakurai and A.R. Newton. Alpha-power law mosfet model and its applications to cmos inverter delay and other formulas. *Solid-State Circuits, IEEE Journal of*, 25(2):584 –594, apr 1990.

[9] M. Lundstrom, Z. Ren, and S. Datta. Essential physics of carrier transport in nanoscale MOSFETs. In *International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, pages 1 –5, 2000.

[10] G.T. Wright. Threshold modelling of MOSFETs for CAD of CMOS-VLSI. *Electronics Letters*, 21(6):223 –224, 14 1985.

[11] Bsimsoi compact mosfet model. [online]. available: http://www-device.eecs.berkeley.edu/bsim/?page=bsimsoi.

[12] B.H. Calhoun, Y. Cao, X. Li, K. Mai, L.T. Pileggi, R.A. Rutenbar, and K.L. Shepard. Digital circuit design challenges and opportunities in the era of nanoscale CMOS. *Proceedings of the IEEE*, 96(2):343 –365, Feb. 2008.