This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Compact Models for MOS Transistors: Successes and Challenges

Colin C. McAndrew, *Fellow, IEEE*

***Abstract*— This paper provides an industry perspective on the present state of compact models for MOS transistors. It highlights the complexity of layout-dependent effects in modern transistors, reviews some common misunderstandings of MOS transistor capacitances, and introduces a new, structurally symmetric, "gate-as-input" equivalent network for these capacitances. Areas of focus for the future model developments are also proposed.**

***Index Terms*— Compact models, MOS transistors, SPICE.**

I. Introduction

MOS transistors are the basic building block of the Information Age. Although the structure of transistors has evolved, from planar bulk transistors to ultrathin body and box (UTTB) transistors to FinFETs to gate-all-around transistors, the basic principal of operation has remained unchanged: an electric field is applied transverse to the drain–source axis of the transistor to induce a conducting charge density, then another electric field is applied longitudinally to the drain–source axis to induce current flow. The details of the electrostatics of how the conducting charge density depends on the applied fields vary between the different transistor structures, but the fundamental principle of operation does not.

In the 45 years since SPICE [1], [2] was unleashed on the world, integrated circuit (IC) design has been based on circuit simulation. The accuracy of those simulations depends both on the numerical algorithms used in a simulator and the compact models available.[1] There have been significant improvements in compact models for MOS transistors over the past 30 years; however, issues still remain. From my perspective in industry, here I will briefly review what I think have been the biggest steps forward, where the most pressing needs for continued model improvement are, and where I see misunderstandings about how MOS transistors, and models for them, work. My comments are mostly generic, and independent of specific transistor type (planar, FinFET, etc.).

Manuscript received May 10, 2018; revised June 13, 2018; accepted June 14, 2018. The review of this paper was arranged by Editor B. Iñiguez.

The author is with NXP Semiconductors N.V., Chandler, AZ 85224 USA (e-mail: mcandrew@ieee.org).

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

Digital Object Identifier 10.1109/TED.2018.2849943

[1] One of the core strengths of SPICE was that it placed as much importance on models as on algorithms; to quote from [1]: "particular emphasis is placed upon the circuit models for the BJT and the FET" (today "circuit models" are commonly referred to as "compact models" or "SPICE models").

II. Path to the Present

Models for MOS transistors were developed more than 50 years ago [3], [4]. What is considered to be the reference approach, the Pao–Sah model, followed shortly thereafter [5], although, because of its double integral formulation, it was impractical for use in circuit simulators for IC design. The first MOS transistor model implemented in SPICE was the Shichman–Hodges model [6], which apart from some "if" conditions and absolute value operations was close to being symmetric. The first practical MOS transistor model based on a surface potential ψ_s appeared a decade later [7]. However, there were difficulties with this approach that required years of effort to solve before useful implementations became available. Consequently, MOS transistor model development went down the path of source-referenced, threshold voltage-based models. At the time, the fundamental shortcomings of this approach, which eventually doomed it, were not recognized. An alternative approach based on inversion charge was also proposed [8].

Although complaints about problems with MOS transistor models were raised over 35 years ago [9], no progress on improvement was made until specific benchmarks were defined, starting in [10] and culminating in [11] and [12, Appendix K]. Once these became expected minimum requirements for models, the source-referenced, threshold voltage-based approach was discarded. Today's models are generally inversion charge based, like BSIM-BULK [13], or ψ_s-based, for FinFET [14], UTTB [15], and bulk [16] structures. Specifically, the symmetric linearization approach used in PSP [16] was a major step forward. An innovative formulation for double-gate MOS transistors was pioneered in [17], but has not found its way into any standard models. The "virtual source" approach to field-effect transistor modeling [18] has not (yet) found its way into mainstream CMOS models, but has proven to be effective for a wide range of devices, especially for GaN HEMT modeling [19].

Note the time horizons associated with model development, implementation, and adoption. It took more than 30 years for problems initially noted in [9] to be addressed in bulk MOS transistor models, and over 5 years from the announcement of [13] until parameter sets started to be provided by foundries.

III. Current Challenges

One of the major issues facing MOS transistor modeling is complexity, not of the devices themselves, as their principle of operation is relatively simple (as noted earlier, induce a conducting charge with a transverse field, drive current with a longitudinal field), but of the influence of their surroundings.

0018-9383 © 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE I
PROXIMITY EFFECTS [20], [21]

Name(s)	Acronym(s)	Category
lithographic proximity effect layout rounding effect	LPE LRE	lithography/shape
active diffusion poly corners		rounding$^{(1)}$
transient enhanced diffusion	TED	damage/strain$^{(2)}$
well proximity effect	WPE	dopant scattering
shallow trench isolation length of diffusion length of thin-ox definition	STI LOD	strain
(embedded) SiGe proximity	eSiGe siGe	strain
contact induced strain contact proximity		strain
contact etch-stop layer	CESL	strain
polysilicon spacing effect	PSE	strain
dual stress-liner effect	DSL	strain
near diffusion effect neighboring diffusion effect diffusion spacing effect oxide spacing effect OD spacing effect active spacing effect	NDE DSE OSE	strain
metal boundary effect	MBE	NMOS/PMOS metal gate work-function interaction
gate line end effect	GLE	variation an ends of FinFET gates
gate cut effect		strain$^{(3)}$

(1) not really a proximity effect
(2) indirect, affects the doping profile, which affects V_{T0}
(3) from cuts in FinFET gates used for adjacent channel interconnect

Table I summarizes some of the proximity effects, also known as layout-dependent effects (LDEs), that are found in modern technologies. Each of these can have a significant effect on transistor behavior, and the range of influence of some effects can be of the order of a few micrometers [22]. Today, many surrounding devices or influencing edges can lie within that distance of a specific transistor.

The transistor layouts used to characterize LDEs commonly use "regular" layouts (i.e., with constant spacing to strain sources). Fortuitously, parameterizing LDEs using integrated "effective" spacings has proved to give reasonably accurate results for irregular layouts as well [22], [23]. Characterization of LDEs is usually done by varying one effect at a time, but recently it has been discovered that different LDEs interact and cannot be modeled as separate, additive effects [24]. If LDE characterization requires seven different transistor geometry selections, and three "aggressor" edge spacings for

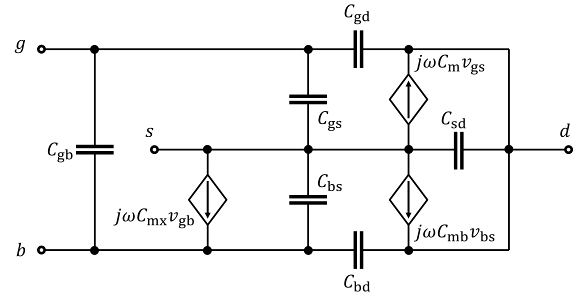

Fig. 1. Gate-as-output-centric capacitance equivalent circuit (asymmetric).

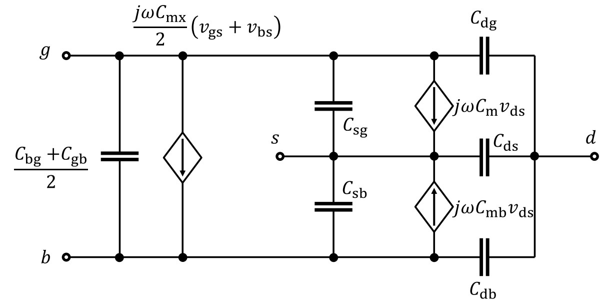

Fig. 2. Gate-as-input-centric capacitance equivalent circuit (symmetric).

each LDE, then if five proximity effects are to be modeled without interactions, this requires 105 separate test structures per device type. However, if effect interactions are taken into account, this would require 1701 test structures, which is impractical.

In addition, shrinking device dimensions and the move to FinFET structures introduce many significant parasitic resistances and capacitances from and between interconnects, fins, and access regions [25].

Modeling the LDEs and the parasitics is as important as and can be more challenging than modeling the intrinsic transistor; this is what I mean by "complexity" being a significant issue in modeling MOS transistors in modern technologies.

IV. CAPACITANCES

Capacitances are, in my opinion, the least well-understood aspect of MOS transistor behavior. I often find that people do not know that or do not understand why $C_{sd} < 0$ and $C_{dg} \neq C_{gd}$, although clear explanations are provided in [12]. Sometimes they cannot articulate the difference between C_{dg} and C_{gd}, and are not sure which capacitances are the most important for common circuit applications. In part, this is because of the simplified, incomplete capacitance models presented in IC design textbooks. This section presents a simple way to conceptualize MOS transistor capacitances, details which capacitances are the most important for amplifiers operating in saturation, proposes an alternative capacitance representation that is more "design centric" than the commonly used equivalent network, and discusses the root cause on nonreciprocity at $V_{DS} = 0$ in ψ_s-based models.

MOS transistor capacitances are often denoted generically as C_{ij}, but this gives no mental hint to help interpret what

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

the capacitance means.[2] Using the notation C_{mf} simplifies interpretation: here m stands for *measure* and f stands for *force*; C_{mf} then represents the current in port m when port f is excited by a small-signal sinusoidal voltage.[3]

Fig. 1 shows the conventional quasi-static small-signal capacitance representation for a MOS transistor [12], [26], where $C_{\mathrm{m}} = C_{\mathrm{dg}} - C_{\mathrm{gd}}$, $C_{\mathrm{mb}} = C_{\mathrm{db}} - C_{\mathrm{bd}}$, and $C_{\mathrm{mx}} = C_{\mathrm{bg}} - C_{\mathrm{gb}}$. In compact models, a charge is associated with each port, and each charge is a function of the port voltages. However, the sum of the port charges must be zero, and one port needs to be used as a voltage reference, so there are only three independent charges and three independent voltages, therefore, nine independent capacitances (which are the imaginary components of the elements of the 3×3 y-parameter matrix of the device [12]). A capacitor symbol represents the imaginary component of current flowing between two ports due to a small-signal excitation between the same two ports; topologically only six such connections can be made for a four-port device, so there must be at least three "nonself-controlled" capacitance elements.

The topology of Fig. 1 was selected [26] because if reciprocity is assumed, i.e., if $C_{\mathrm{m}} = C_{\mathrm{mb}} = C_{\mathrm{mx}} = 0$, and C_{sd} is removed, then the equivalent circuit and capacitances are identical to the common (but incomplete) five-capacitance textbook representation. There are several problems with this selection.

First, it does not embody the symmetry of the front and back gates.[4]

Second, when excitation is applied to the gate, the capacitive currents out of the drain, source, and body are determined by C_{dg}, C_{sg}, and C_{bg}, respectively (g is the "force" port), none of which is directly included in the representation of Fig. 1.

Third, consider operation of a long-channel transistor in saturation, with $V_{\mathrm{bs}} = 0$, is a typical usage in analog circuits (this is a common-source amplifier). Ignoring parasitics, the 2×2 capacitance matrix (port 1 is g and port 2 is d) is

$$\begin{bmatrix} C_{\mathrm{gg}} & -C_{\mathrm{gd}} \\ -C_{\mathrm{gd}} - C_{\mathrm{m}} & C_{\mathrm{dd}} \end{bmatrix} \approx \begin{bmatrix} C_{\mathrm{gg}} & 0 \\ -C_{\mathrm{dg}} & 0 \end{bmatrix} \quad (1)$$

where the approximation recognizes that for a long-channel transistor in saturation, derivatives with respect to the drain voltage are small. The C_{dg} element is not immediately apparent in Fig. 1, it is "hidden" in the C_{m} element; in fact, thinking in terms of the common five-capacitance textbook representation, the C_{dg} component of current is completely missing. For a typical substrate doping in a planar device, approximately 35% of the imaginary component of gate current comes out of the drain, the other 65% out of the source and body; the "capacitive feedthrough" current to the drain is omitted, or obscured, in common MOS transistor small-signal model representations.

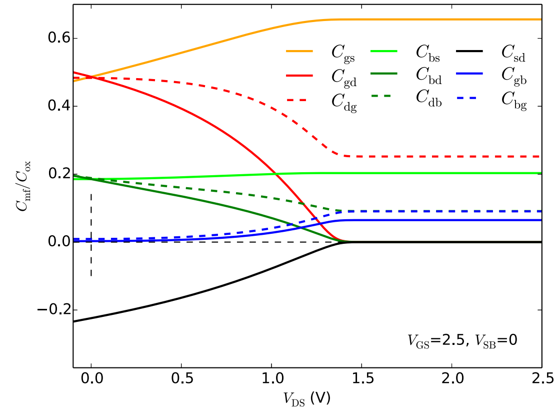

Fig. 3. Capacitances of the small-signal capacitance model of Fig. 1.

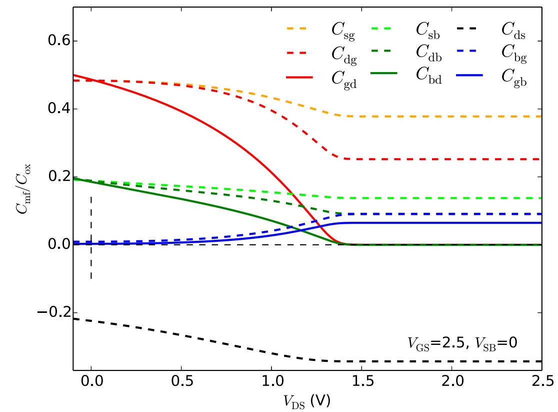

Fig. 4. Capacitances of the small-signal capacitance model of Fig. 2.

Fourth, Fig. 1 gives the appearance that the gate–drain capacitance C_{gd} is Miller multiplied. This is true for the overlap and fringing components of capacitance but is not true for the intrinsic component; the "feedthrough" capacitance C_{dg} is *not* Miller multiplied.

Finally, and more importantly, in normal circuit usage, the gate is the input port, not the output port, so it is more natural to represent the small-signal capacitances in terms of C_{mg} rather than C_{gf}. Fig. 2 is an equivalent small-signal capacitance representation for a MOS transistor that emphasizes the gate as an input. For a common-source amplifier in saturation, it is immediately apparent in Fig. 2 that C_{dg} models the gate-to-drain capacitative current.

Figs. 3 and 4 show intrinsic capacitances, normalized to the total gate capacitance C_{ox}, from an ideal long-channel model in (relatively) strong inversion, for the small-signal capacitance models of Figs. 1 and 2, respectively. There are only three differences C_{sg}, C_{sb}, and C_{ds} replacing C_{gs}, C_{bs}, and C_{sd}, respectively. Compatibility with the five-capacitance textbook model is lost, but the importance of C_{dg} becomes obvious. In addition, Fig. 2 reflects the structural gate symmetry of UTTB devices, whereas Fig. 1 does not.

At $V_{\mathrm{DS}} = 0$, both detailed physical simulation and thermodynamic analysis show that MOS transistor capacitances should be reciprocal, i.e., $C_{\mathrm{m}} = C_{\mathrm{mb}} = C_{\mathrm{mx}} = 0$ should hold.

[2] Sometimes the term "transcapacitance" is used if $i \neq j$, but they all appear as $\pm j\omega C_{\mathrm{ij}}$ entries in the y-parameter small-signal representation, so following [12], the name capacitance is used here for all C_{ij}.

[3] More precisely, the imaginary component of the current entering port m, if $m = f$, or leaving port m, if $m \neq f$, divided by ω, for a voltage excitation at port f of $1+j0$.

[4] Fig. 1 can be made symmetric by replacing C_{gb} with $(C_{\mathrm{gb}}+C_{\mathrm{bg}})/2$, scaling the C_{mx} source by 1/2, and adding an additional source $j\omega C_{\mathrm{mx}} v_{\mathrm{gb}}/2$ from the source to the (top) gate. However, this leads to an equivalent circuit with 10 elements, more than the minimum requirement of nine, and four "nonself-controlled" capacitance elements, which obfuscates the intent of the equivalent circuit as these are the least well intuitively understood components

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

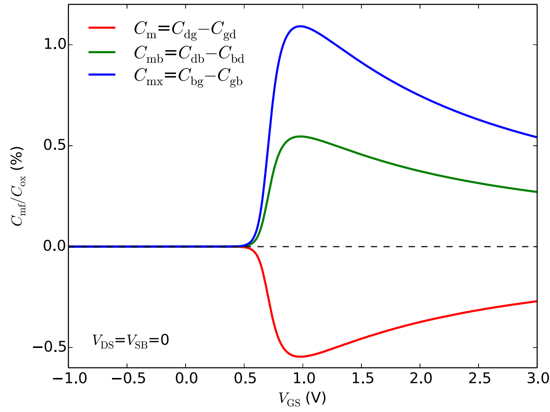

Fig. 5. Capacitance nonreciprocity.

The inversion charge and ψ_s approaches to modeling MOS transistors violate this. What is not widely known is that the degree of violation of reciprocity can be quantified. Detailed analysis in [27] gives, at $V_{\mathrm{DS}} = 0$

$$C_{\mathrm{mx}} = 2C_{\mathrm{mb}} = -2C_{\mathrm{m}} = C_{\mathrm{ox}}\left(\alpha_m \frac{\partial \psi_s}{\partial V_{\mathrm{GB}}} + \frac{\partial \psi_s}{\partial V_{\mathrm{SB}}} - 1\right) \quad (2)$$

where α_m is the bulk charge linearization factor. Fig. 5 shows capacitance nonreciprocity for an ideal long-channel ψ_s-based model [16]. The relationship between the capacitances clearly matches the prediction of (2), but overall the magnitude of the capacitance nonreciprocity is small, less than about 1% of C_{ox}. Sometimes the cause of the nonreciprocity is attributed to the charge-sheet approximation. However, (2) shows that the underlying cause is that ψ_s is not pinned to $2\phi_F + V_{\mathrm{SB}} + \phi_Z$ [12] in strong inversion, where ϕ_Z is several times the thermal voltage ϕ_t; this is underscored because the strong inversion charge model of [12] is based on the charge-sheet approximation, but assumes pinning of ψ_s and has reciprocal capacitances at $V_{\mathrm{DS}} = 0$.

V. Variability

Over the past decade, analog design has increasingly relied on Monte Carlo like simulations (on extracted netlists, because of parasitics and LDEs, see Section III) to verify yield and manufacturability. However, there is still a strong belief that corner (also known as worst case) simulations are sufficient, but this is false [28].

Consider some performance measure e_m for a circuit. If it depends on variability in a set of process parameters p_i, e.g., oxide thickness, channel length variation, doping or sheet resistance variation, and so on, then to first order

$$\delta e_m = \sum_i \frac{\partial e_m}{\partial p_i} \delta p_i \quad (3)$$

is the variation in e_m. The sensitivities $\partial e_m/\partial p_i$ are specific to each e_m, to circuit topology, and to the geometry and bias for each device in the circuit. However, the δp_i are fixed for each corner. This means that a corner simulation guarantees *nothing* about the prediction of the σ level for a specific e_m (see examples provided in [28]). In addition, obviously corner simulations vary each p_i by the same amount for each device, so they do not account for device mismatch (i.e., local variability), which is often the main source of variability for analog circuits.

Furthermore, propagation of variance gives

$$\sigma^2_{\delta e_m} = \sum_i \left[\left(\frac{\partial e_m}{\partial p_i}\right)^2 \sigma^2_{\delta p_i} + 2\sum_{j>i} \frac{\partial e_m}{\partial p_i}\frac{\partial e_m}{\partial p_j}\sigma_{p_i p_j}\right] \quad (4)$$

where $\sigma_{p_i p_j}$ is the covariance of p_i and p_j. If correlations between parameters are ignored, or are improperly modeled, then any prediction of $\sigma_{\delta e_m}$ is inaccurate. Without knowing *a priori* the signs of $\partial e_m/\partial p_i$ and $\partial e_m/\partial p_j$, it is not even possible to predict whether (4) will overpredict or underpredict $\sigma_{\delta e_m}$. Both are bad, i.e., the former leads to overdesign (increased area or power) and the latter to yield loss. In addition

$$\sigma^2_{\delta e_m(\mathrm{total})} = \sigma^2_{\delta e_m(\mathrm{global})} + \sigma^2_{\delta e_m(\mathrm{local})}(L, W, \ldots) \quad (5)$$

where the local component of variation depends (reciprocally) on geometry (gate length L, gate width W, number of gates, and so on). This means that the correlations between different parameters depend on geometry because, by definition, local variations for each parameter for each device are uncorrelated.

The simplest approach to statistical modeling, I believe, is to use uncorrelated parameters; a general technique to model parameter correlations using uncorrelated parameters, and to use correlations between measured e_ms to characterize the variances of the uncorrelated parameters, is available in [29].

VI. Path to the Future

Although many of the historic problems with MOS transistor models have been resolved, some remain unsolved or have not been addressed. The complexity of LDEs, parasitics, and variability in modern processes were noted in Section III. Additional areas that I believe warrant attention include:

1) modeling of sensitivities to layout;
2) efficient modeling of both fast digital transients and analog behavior;
3) efficient and accurate modeling of transistors with nonuniform lateral-channel doping (NULD);
4) development of "design centric" models and proactive use of models for the design;
5) resolution of the discrepancies in electrical behavior between transistors in test arrays, which are used for characterization, and transistors found in typical circuits;
6) techniques and tools to automatically detect unphysical behavior, or numerical problems, in models.

Note that some of these relate to "core" model developments and some to "infrastructure." Each item will now be discussed.

Since device, and therefore circuit, performances depend strongly on LDEs, it has been recognized for some time that the design of precision analog circuits can no longer be done at the schematic level, with a final verification on an extracted netlist, but must integrate schematic and layout in a single design flow. Not all devices affect circuit performance equally, and not all devices are equally affected by LDEs. Being able to use models and simulations to evaluate which transistors to take particular care of in layout could significantly simplify

the design process. Impediments to this are that the pace of technology and device innovation has shrunk below the time horizon of model updates, implementation, and characterization, and the understandable desire to protect proprietary device optimizations. LDE models are now commonly delivered through "black box," nonhuman-readable, mechanisms that make a proactive determination of layout sensitivities difficult.

Occasionally, IC designers will consult me about spikes they see in simulation, and they want to know whether it is a real phenomenon that needs to be factored into reliability analysis, or whether they can ignore it. Usually, the cause turns out to be, in essence, a long transistor with a resistive load being driven by a fast switching waveform. C_{dg} couples the rapidly changing gate voltage to the drain, which causes the spike. However, in practice, the drain charge cannot respond until the carriers have time to transit the length of the channel, and that time varies to first order as L^2 (besides, the distance to travel increasing as L, the electric field, and therefore, the drift speed, decreases as $1/L$). There can be some small spike, from coupling through overlap and fringe capacitances, but mostly it is a simulation artifact, from a quasi-static model being used where the transistor is operating in a nonquasi-static (NQS) manner. NQS models exist but are rarely used because of their computational expense. The spike can be suppressed by selecting a different partitioning of the inversion charge, allocating it all to the source in saturation rather than splitting it between the source and the drain. This artificially forces C_{dg} to be zero in saturation, and unphysically gives the wrong sign for C_{sd} and, in some cases, C_{bs} in nonsaturation, thereby messing up modeling of analog behavior. At present, there is no simple way to simulate both analog and (fast) digital circuits with the same model (a transistor does not "know" if it is analog or digital). It would be better if an efficient technique for modeling charges/capacitances for both analog applications and fast digital transients, without the need to invoke a complex NQS model, could be developed.

Although FinFETs and UTTB transistors have negligible channel doping, high voltage and power transistors are still predominantly planar, and they can have NULD from halo implants and/or from out-diffused channel regions. A good "effective" threshold voltage model has been proposed for halo-implanted devices [30], but this still does not enable accurate modeling of the electrical characteristics of NULD devices. In such transistors, the threshold voltage is determined by the most highly doped portion of the channel; when a transistor is turning ON the effective gate length that reduces to the length of this region, then asymptotically increases to approach the total channel length as V_{GS} increases. This gives a "peaked" $g_{\mathrm{m}}(V_{\mathrm{GS}})$ characteristic (Fig. 6) that can only be approximated with a single-section model by using nonphysical values for some model parameters.

In addition, because the lowly doped portion of the channel inverts before the highly doped portion(s), the capacitance behavior cannot be accurately represented using any standard MOS transistor model. Fig. 7 shows capacitances, at $V_{\mathrm{DS}} = 0$, for a uniformly doped transistor and a transistor with the same doping as the uniformly doped transistor at its source

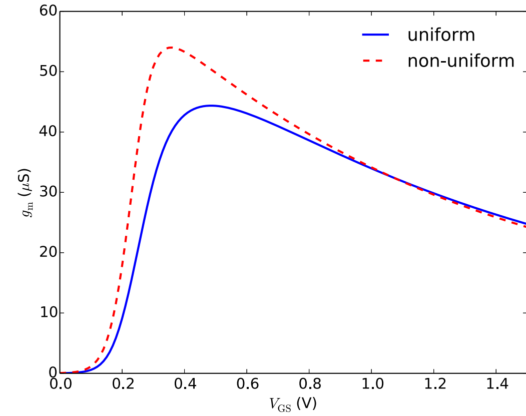

Fig. 6. g_{m} for uniform and nonuniform lateral doping, $V_{\mathrm{DS}} = 0.1$ V.

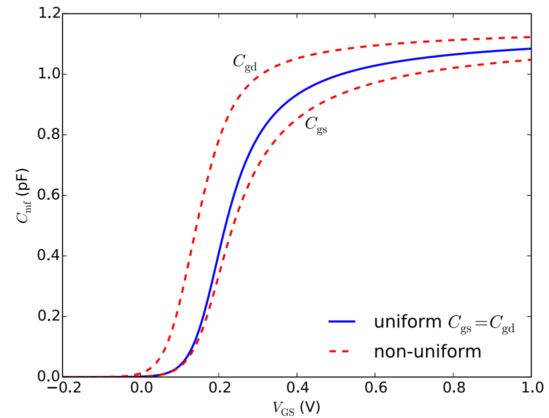

Fig. 7. C_{mf} for uniform and nonuniform lateral doping, $V_{\mathrm{DS}} = 0$ V.

and a lower doping toward the drain. Clearly, C_{gs} and C_{gd} are asymmetric. At present, the only known way to simulate such behavior is with a sectional transistor model, but this significantly increases simulation time. It would be better if a single-section model that can fit such behavior were developed.[5]

Rather than defining transistor geometries and biases and then simulating their performance, IC design would be more efficient if requirements for f_{T}, g_{m}, and matching were specified and an automated design tool could, based on the actual design models and not simplifications, determine appropriate geometries and biases. Steps in this direction have been taken [33], [34] but the approach has not become mainstream.

As discussed in Section III, LDEs from parasitics to surrounding interconnect, devices, and material and layer boundaries can significantly affect the electrical characteristics of devices. Transistor behavior, therefore, depends strongly on the surroundings. However, the test structures used for device characterization and for process monitoring and control are typically isolated devices, either individual or a mismatch pair, connected to probe pads. Their surroundings are, therefore, significantly different from those of devices in circuits, which

[5]Capacitances of NULD MOS transistors are not just asymmetric, they are nonreciprocal at $V_{\mathrm{DS}} = 0$ [31], [32]; this behavior cannot be modeled even quantitatively today using port charges of a single-section model.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

are densely packed with many interconnect lines. There is substantial opportunity to improve how well test structures reflect the behavior of "real" transistors by using devices in representative circuits for characterization and modeling.

Finally, all novice, and even sometimes experienced, model developers will occasionally formulate equations that violate the laws of physics or implement code that is not numerically robust. Following best practices [35] and using code standard checkers can significantly help to improve model quality, but these do not guarantee that a model obeys the laws of physics or has no numerical problems. This is a hard problem but it would be useful if an automated checker for passivity, order of continuity, symmetry, numerical singularities, and so on, over all possible parameter and bias conditions, could be developed.

VII. Conclusion

MOS transistor models continue to evolve, although in a more evolutionary than revolutionary manner than in the past. Apart from BSIM-BULK [13], most mainstream models are ψ_s based, with differences for different device structures in how the electrostatics is formulated and solved. A primary emphasis of modeling today is not the core transistor behavior, but rather on proximity effects and parasitics, which have a significant influence on transistor behavior.

This paper presented a perspective, from industry, on various aspects of MOS transistor modeling. An alternative, "gate-as-input-centric" capacitance model was proposed, some limitations of existing models were highlighted, and what I think are the fruitful areas for further development were presented.

References

[1] L. W. Nagel and D. O. Pederson, "Simulation program with integrated circuit emphasis (SPICE)," in *Proc. 16th Midwest Symp. Circuit Theory*, 1973, Paper VI.1.

[2] L. W. Nagel, "SPICE2: A computer program to simulate semiconductor circuits," Ph.D. dissertation, Dept. EECS, Univ. California, Berkeley, Berkeley, CA, USA, May 1975.

[3] H. K. J. Ihantola and J. L. Moll, "Design theory of a surface field-effect transistor," *Solid-State Electron.*, vol. 7, pp. 423–430, Jun. 1964, doi: 10.1016/0038-1101(64)90039-5.

[4] C.-T. Sah, "Characteristics of the metal-oxide-semiconductor transistors," *IEEE Trans. Electron Devices*, vol. ED-11, no. 7, pp. 324–345, Jul. 1964, doi: 10.1109/T-ED.1964.15336.

[5] H. C. Pao and C. T. Sah, "Effects of diffusion current on characteristics of metal-oxide (insulator)-semiconductor transistors," *Solid-State Electron.*, vol. 9, no. 10, pp. 927–937, Oct. 1966, doi: 10.1016/0038-1101(66)90068-2.

[6] H. Shichman and D. A. Hodges, "Modeling and simulation of insulated-gate field-effect transistor switching circuits," *IEEE J. Solid-State Circuits*, vol. SSC-3, no. 3, pp. 285–289, Sep. 1968, doi: 10.1109/JSSC.1968.1049902.

[7] J. R. Brews, "A charge-sheet model of the MOSFET," *Solid-State Electron.*, vol. 21, no. 2, pp. 345–355, Feb. 1978, doi: 10.1016/0038-1101(78)90264-2.

[8] M. A. Maher and C. A. Mead, "A physical charge-controlled model for MOS transistors," in *Advanced Research in VLSI*, P. Losleben, Ed. Cambridge, MA, USA: MIT Press, 1987.

[9] Y. Tsividis, "Problems with precision modeling of analog MOS LSI," in *IEDM Tech. Dig.*, Dec. 1982, pp. 274–277, doi: 10.1109/IEDM.1982.190272.

[10] Y. P. Tsividis and K. Suyama, "MOSFET modeling for analog circuit CAD: Problems and prospects," in *Proc. IEEE Custom Integr. Circuits Conf.*, May 1993, pp. 14.1.1–14.1.6, doi: 10.1109/CICC.1993.590710.

[11] X. Li, W. Wu, G. Gildenblat, C. C. McAndrew, and A. J. Scholten, "Benchmark tests for MOSFET compact models," in *Compact Modeling: Principles, Techniques and Applications*, G. Gildenblat, Ed. New York, NY, USA: Springer, 2010, ch. 3, pp. 75–104, doi: 10.1007/978-90-481-8614-3.

[12] Y. Tsividis and C. McAndrew, *Operation and Modeling of the MOS Transistor*, 3rd ed. New York, NY, USA: Oxford Univ. Press, 2011.

[13] Y. S. Chauhan *et al.*, "BSIM6: Analog and RF compact model for bulk MOSFET," *IEEE Trans. Electron Devices*, vol. 61, no. 2, pp. 234–244, Feb. 2014, doi: 10.1109/TED.2013.2283084.

[14] M. V. Dunga, "Nanoscale CMOS modeling," Ph.D. dissertation, Dept. EECS, Univ. California, Berkeley, Berkeley, CA, USA, Mar. 2008.

[15] S. Khandelwal *et al.*, "BSIM-IMG: A compact model for ultrathin-body SOI MOSFETs with back-gate control," *IEEE Trans. Electron Devices*, vol. 59, no. 8, pp. 2019–2026, Aug. 2012, doi: 1109/TED.2012.2198065.

[16] G. Gildenblat *et al.*, "Surface-potential-based compact model of bulk MOSFET," in *Compact Modeling: Principles, Techniques and Applications*, G. Gildenblat, Ed. New York, NY, USA: Springer, 2010, ch. 1, pp. 3–40, doi: 10.1007/978-90-481-8614-3.

[17] Y. Taur, X. Liang, W. Wang, and H. Lu, "A continuous, analytic drain-current model for DG MOSFETs," *IEEE Electron Device Lett.*, vol. 25, no. 2, pp. 107–109, Feb. 2004, doi: 10.1109/LED.2003.822661.

[18] A. Khakifirooz, O. M. Nayfeh, and D. Antoniadis, "A simple semi-empirical short-channel MOSFET current–voltage model continuous across all regions of operation and employing only physical parameters," *IEEE Trans. Electron Devices*, vol. 56, pp. 1674–1680, Aug. 2009, doi: 10.1109/TED.2009.2024022.

[19] U. Radhakrishna, S. Lim, P. Choi, T. Palacios, and D. Antoniadis, "GaNFET compact model for linking device physics, high voltage circuit design and technology optimization," in *IEDM Tech. Dig.*, 2015, pp. 9.6.1–9.6.4, doi: 10.1109/IEDM.2015.7409664.

[20] J. V. Faricelli, "Layout-dependent proximity effects in deep nanoscale CMOS," in *Proc. IEEE Custom Integr. Circuits Conf.*, Sep. 2010, pp. 1–8, doi: 10.1109/CICC.2010.5617407.

[21] A. L. S. Loke *et al.*, "Analog/mixed-signal design challenges in 7-nm CMOS and beyond," in *Proc. IEEE Custom Integr. Circuits Conf.*, Apr. 2018, pp. 1–8, doi: 10.1109/CICC.2018.8357060.

[22] K.-W. Su *et al.*, "A scaleable model for STI mechanical stress effect on layout dependence of MOS electrical characteristics," in *Proc. IEEE Custom Integr. Circuits Conf.*, Sep. 2003, pp. 245–248, doi: 10.1109/CICC.2003.1249396.

[23] B. Peddenpohl, M. Otrokov, and J. Wells, "Validation of the BSIM4 irregular LOD SPICE model by characterization of various irregular LOD test structures," in *Proc. Int. Conf. Microelectron. Test Struct.*, 2018, pp. 31–34, doi: 10.1109/ICMTS.2018.8383758.

[24] Q. Lin, H. Pan, and J. Chang, "Test structures for debugging variation of critical devices caused by layout-dependent effects in FinFETs," in *Proc. Int. Conf. Microelectron. Test Struct.*, 2018, pp. 3–6, doi: 10.1109/ICMTS.2018.8383751.

[25] N. Lu, P. M. Kotecha, and R. A. Wachnik, "Modeling of resistance in FinFET local interconnect," in *Proc. IEEE Custom Integr. Circuits Conf.*, Sep. 2014, pp. 1–4, doi: 10.1109/CICC.2014.6946027.

[26] C. Turchetti, G. Masetti, and Y. Tsividis, "On the small-signal behaviour of the MOS transistor in quasistatic operation," *Solid-State Electron.*, vol. 26, pp. 941–949, Oct. 1983, doi: 10.1016/0038-1101(83)90070-9.

[27] C. McAndrew. *Complete and Consistent All-Region Body-Referenced Capacitance Model Based on Symmetric Linearization*. Accessed: May 2018. [Online]. Available: http://global.oup.com/us/companion.websites/9780195170153/pdf/CapCompAl%lRegionBodyRefModel.pdf

[28] C. C. McAndrew, I.-S. Lim, B. Braswell, and D. Garrity, "Corner models: Inaccurate at best, and it only gets worst," in *Proc. IEEE Custom Integr. Circuits Conf.*, Sep. 2013, pp. 219–224, doi: 10.1109/CICC.2013.6658428.

[29] C. C. McAndrew, "Statistical modeling using backward propagation of variance (BPV)," in *Compact Modeling: Principles, Techniques and Applications*, G. Gildenblat, Ed. New York, NY, USA: Springer, 2010, ch. 16, pp. 491–520, doi: 10.1007/978-90-481-8614-3.

[30] R. Rios *et al.*, "A three-transistor threshold voltage model for halo processes," in *IEDM Tech. Dig.*, Dec. 2002, pp. 113–116, doi: 10.1109/IEDM.2002.1175791.

[31] H. K. Dirks, "Kapazitätskoeffizienten nichtlinear dissipativer systeme," Habilitation thesis, Dept. Inst. Theoretische Elektrotechnik, RWTH Aachen Univ., Aachen, Germany, 1988.

[32] A. C. T. Aarts, R. van der Hout, J. C. J. Paasschens, A. J. Scholten, M. B. Willemsen, and D. B. M. Klaassen, "New fundamental insights into capacitance modeling of laterally nonuniform MOS devices," *IEEE Trans. Electron Devices*, vol. 53, no. 2, pp. 270–278, Feb. 2006, doi: 10.1109/TED.2005.862235.

[33] D. Stefanovic, M. Kayal, M. Pastre, and V. B. Litovski, "Procedural analog design (PAD) tool," in *Proc. 4th Int. Symp. Quality Electron. Design*, 2003, pp. 313–318, doi: 10.1109/ISQED.2003.1194751.

[34] C. L. Recker, B. Braswell, P. G. Drennan, and C. C. McAndrew, "A Web tool for interactive exploration of analog design tradeoffs," in *Proc. IEEE Custom Integr. Circuits Conf.*, Sep. 2006, pp. 631–634, doi: 10.1109/CICC.2006.320838.

[35] C. C. McAndrew *et al.*, "Best practices for compact modeling in Verilog-A," *IEEE J. Electron Devices Soc.*, vol. 3, no. 5, pp. 383–396, Sep. 2015, doi: 10.1109/JEDS.2015.2455342.

Colin C. McAndrew (S'82–M'84–SM'90–F'04) received the B.E. degree (Hons.) in electrical engineering from Monash University, Melbourne, VIC, Australia, in 1978, and the Ph.D. degree in systems design engineering from the University of Waterloo, Waterloo, ON, Canada, in 1984.

From 1987 to 1995, he was with AT&T Bell Laboratories, Allentown, PA, USA. He is currently a fellow with NXP Semiconductors, Chandler, AZ, USA.