of the depletion region or decrease of the capacitance slows. If the substrate has no p^+ region as well as having a lower doping as in some bipolar processes for RF applications [5], the n-well bias can reduce the parasitic capacitance even more.

A recent report has shown that in a SiGe bipolar technology, having floating n-doped regions under inductors can reduce the peak Q [6]. Measurements of inductors with and without an n-well fabricated in the CMOS technology used in this study, however showed that the peak Q is slightly higher when the well is included. It is speculated that this difference with the previous report [6] is due to the fact that resistivities in n-wells are similar or higher than in the havier doped field isolation regions, and the neutral n-well regions essentially act like the field isolation regions.

An ability to reduce the parasitic capacitance without affecting the inductance and series resistance is a useful feature for designing inductors for RF circuits. This brief demonstrated a method to decrease parasitic capacitance by a factor little less than 2. This, in theory, can be used to widen the metal line by a factor of 2 or enable use of the 1st level metal in a shunting scheme to improve the low-frequency Q by a factor of 2 while keeping the self-resonance frequency constant.

REFERENCES

[1] N. M. Nguyen and R. G. Meyer, “Si IC-compatible inductors and LC passive filters,” *IEEE J. Solid-State Circuits*, vol. 25, pp. 1028–1031, Aug. 1990.
[2] J. Y. Chang, A. A. Abidi, and M. Gaitan, “Large suspended inductors on Silicon and their use in a 2-μm CMOS RF amplifier,” *IEEE Electron Device Lett.*, vol. 14, pp. 246–248, May 1993.
[3] J. N. Burghartz, M. Soyuer, K. A. Jenkins, and M. D. Hulvey, “High-Q inductors in standard Silicon technology and its application to an integrated RF power amplifier,” in *IEDM'95*, 1995, pp. 1015–1017.
[4] H. M. Greenhouse, “Design of planar rectangular microelectronic inductors,” *IEEE Trans. Parts, Hybrids, Packag.*, vol. PHP-10, pp. 101–109, June 1974.
[5] K. O, P. Garone, C. Tsai, G. Dawn, B. Scharf, T. Tewksbury, C. Kermarrec, and J. Yasaitis, “A low-cost and low-power silicon npn bipolar process with NMOS transistors (ADRF) for RF and microwave applications,” *IEEE Trans. Electron Devices*, vol. 42, pp. 1831–1840, Oct. 1995.
[6] J. N. Burgartz, M. Soyer, K. A. Jenkins, and M. Kies, “RF components implemented in an analog SiGe bipolar technology,” in *IEEE BCTM Proc.*, 1996, pp. 138–141.

A New Algorithm for Faster Full-Thermodynamic Device Simulations

Ming-Yeh Chuang and Mark E. Law

***Abstract*—Full-thermodynamic device simulation solves drift-diffusion (DD), energy balance, and lattice self-heating equations simultaneously and the solution time is extensive. An algorithm to reduce the solution time is presented. The method uses a single lattice temperature to describe the self-heating in the whole semiconductor device. It is shown that the solution time is reduced by 20% and little error is induced compared with the full-thermodynamic simulations. This approach appears useful for submicron device structures.**

I. INTRODUCTION

The traditional drift-diffusion (DD) transport model in device simulations solves three semiconductor equations: Poisson's equation plus the electron and hole continuity equations. However, as device technologies improve, these three equations become insufficient to describe the characteristics of state-of-the-art semiconductor devices. The failure comes from the fact that a quasi-static condition is assumed while deriving the DD transport equations which is not true for carrier transport in a fast spatial and time variation of the local electric field. The large field and field gradient inside the scaled devices causes lattice self-heating and hot carrier nonlocal transport effects. Lattice self-heating and carrier energy balance equations [1]–[4] govern the energy transfer among electrons, holes, and the crystal lattice which take into account most of the possible mechanisms associated with nonlocal and self-heating effects. The full thermodynamic (TD) simulations solve these three energy equations together with DD equations producing better prediction capabilities compared with the DD simulations [5], [6].

The specific position of a simulated device geometry is represented by a node in the mesh. Equations are solved and solutions are stored at each node. For a simulated structure of N nodes, the total equations M solved in a TD simulation are $6 \times N$ whereas $3 \times N$ equations are solved in a DD simulation. The required memory and CPU time to solve the problem are approximately proportional to $M^{1.5}$. Thus, the memory and CPU time requirements of TD simulations are 2.8 times that of the DD simulations for the same number of nodes. A scaled device, in general, requires thousands of nodes to define the device structure. Although TD simulations are more accurate, the simultaneous solving of six equations is CPU and memory intensive compared with DD simulations. In this paper, an algorithm which reduces the solution time of the TD simulations with little sacrifice in the accuracy is presented.

II. ALGORITHM

The algorithm is implemented in FLOODS (FLorida Object Oriented Device Simulator) [5], [7] which has the capability to solve the complete TD equations. In FLOODS, The energy balance equations for electrons and holes are given by

$$\frac{\partial}{\partial t}(n \cdot U_n) = -\nabla \bullet \vec{S}_n - R_{\mathrm{wn}} \tag{1}$$

Manuscript received January 14, 1997; revised April 11, 1997. The review of this brief was arranged by Editor A. H. Marshak.

The authors are with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611-2044 USA.

Publisher Item Identifier S 0018-9383(97)06147-9.

0018–9383/97$10.00 © 1997 IEEE

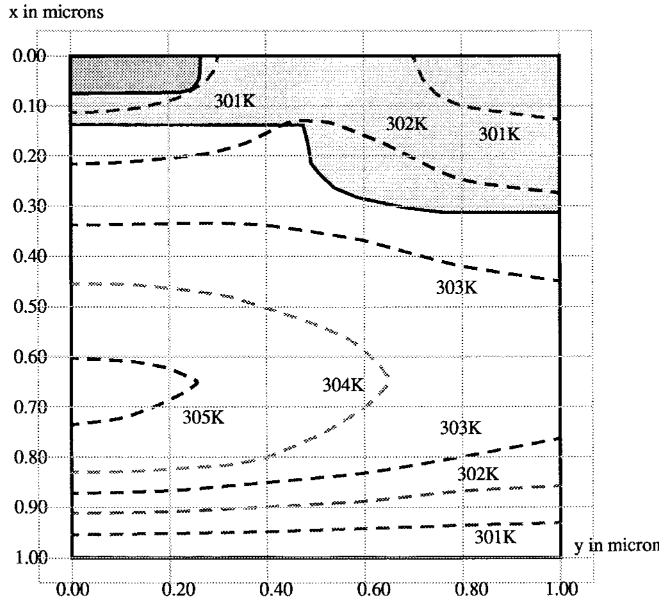

Fig. 1. The lattice temperature contour in an npn BJT at the biases of $V_{\rm be} = 1.0$ V and $V_{\rm cb} = 3$ V. The ambient temperature is 300 K and the thermal conductances at the contacts are infinite. The dark shaded area is the emitter and the light shaded area is the base.

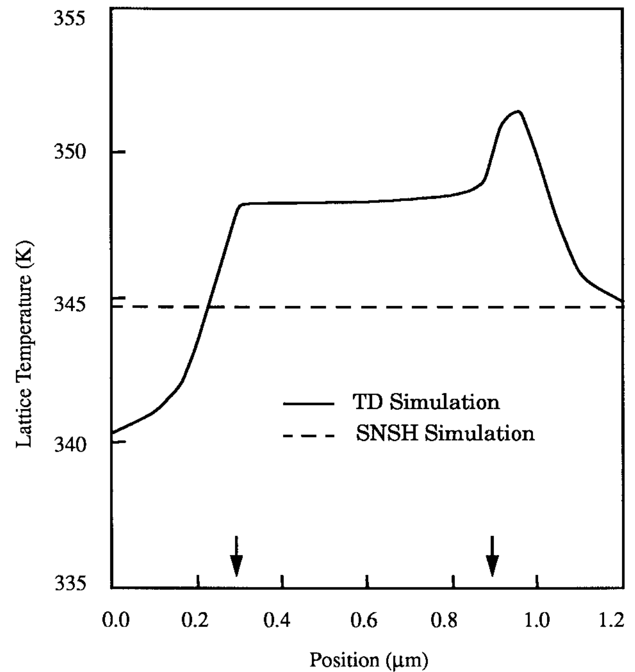

Fig. 2. Lattice temperature along the surface of an NMOS from TD and SNSH simulations. Both the biases at gate and drain are 7 V and the ambient temperature is 300 K. A thermal conductance of 9000 K/Wcm is attached to each contact. The arrows mark the gate width.

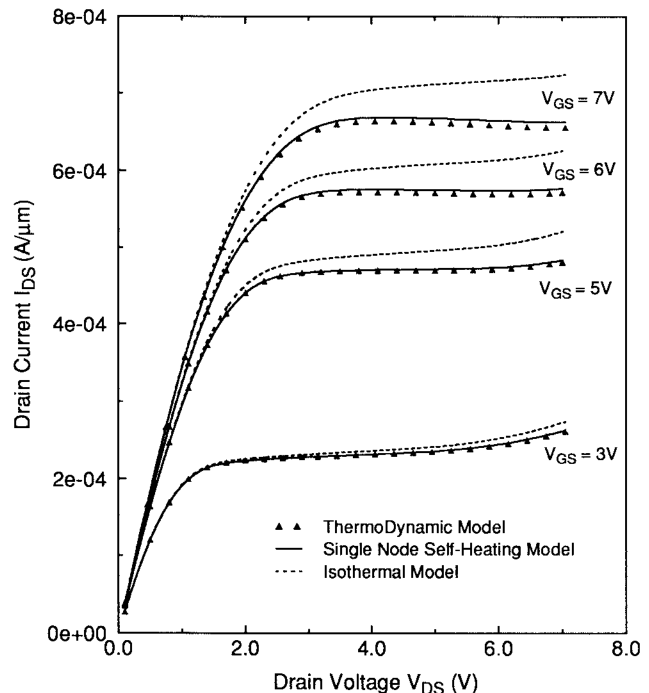

Fig. 3. The comparisons of the $I_{\rm DS}$–$V_{\rm DS}$ plot of an NMOS at different $V_{\rm GS}$'s between three models.

and

$$\frac{\partial}{\partial t}(p \cdot U_p) = -\nabla \bullet \vec{S}_p - R_{\rm wp} \tag{2}$$

where n and p are the electron and hole densities, respectively, $U_{\rm n,p}$ is the average local carrier energy, $\vec{S}_{\rm n,p}$ is the carrier energy flux, and $R_{\rm wn}$ and $R_{\rm wp}$ are the energy transfer rates from electron and hole, respectively, to other subsystems due to the scattering or carrier recombination/generation. The lattice self-heating equation governing the joule-heating and heat conduction in the semiconductor material is

$$\frac{\partial}{\partial t}(U_L) = -\nabla \bullet \vec{S}_L - R_{\rm wl} \tag{3}$$

where U_L is the local lattice energy, $\vec{S}_L$ is the lattice heat flux, and $R_{\rm wl}$ is the local energy transfer rate from crystal lattice to the carrier subsystems. If no external energy, (e.g., photon) is supplied, the complete thermodynamic system is energy conservative and we have $R_{\rm wl} + R_{\rm wn} + R_{\rm wp} = 0$.

$\vec{S}_L$ in the bulk semiconductor is primarily the heat flow due to thermal diffusion of the lattice temperature and can be expressed as $\vec{S}_L = -K_L \nabla T_L$ where K_L is the lattice's thermal conductivity and T_L is the lattice temperature. At the contact boundaries of the semiconductor, the heat flux is conducted to the ambience through interconnection contacts and can be represented by

$$\vec{S}_L = G_{\rm th} \cdot (T_L - T_A) \tag{4}$$

where $G_{\rm th}$ is the thermal conductance of the interconnection contact and T_A is the ambient temperature of the device. The active region of a scaled BJT or MOS transistor is usually small and the local self-heating of the silicon lattice is distributed over the active region. If we assume the lattice temperature is constant in the bulk region, we have $\nabla T_L = 0$ and $\vec{S}_L = 0$. Since the lattice temperature is constant in the bulk, the solution of T_L at one node can represent T_L's at the other $N-1$ nodes and only one instead of N lattice self-heating equations needs to be solved in the bulk mesh of N nodes. The size of the Jacobian matrix becomes $(5N+1) \times (5N+1)$ whereas the TD simulation requires solving a $6N \times 6N$ matrix. Although solving one equation, FLOODS loops over the N nodes and sums the energy transfer rate of each node in the same equation. Thus, (3) can be interpreted as follows: the total lattice energy generated in the device is represented by single temperature in one node and balanced by the heat flux at the contact boundaries. In the isothermal model, the self-heating equation is not taken into account and the size of Jacobian matrix is $5N \times 5N$ which is about the same size as the matrix in the single node self-heating (SNSH) model if $N \gg 1$.

TABLE I
THE COMPARISON OF THE CPU TIME AND ERROR BETWEEN THREE SIMULATION SCHEMES FOR NPN BJT SIMULATIONS

Model	Vbe = 0.8V, Vcb = 3.0V		Vbe = 0.9V, Vcb = 3.0V	
	CPU time	error (%)	CPU time	error (%)
Full-thermo-dynamic	1	0	1	0
Single Node Self-Heating	0.82	0.10	0.77	1.52
Isothermal	0.78	1.17	0.71	7.09

TABLE II
THE COMPARISON OF THE CPU TIME AND ERROR BETWEEN THREE SIMULATION SCHEMES FOR NMOS SIMULATIONS

Model	Vdrain = 7V, Vgate = 6V		Vdrain = 7V, Vgate = 7V	
	CPU time	error (%)	CPU time	error (%)
Full-thermo-dynamic	1	0	1	0
Single Node Self-Heating	0.80	1.02	0.72	1.16
Isothermal	0.76	8.69	0.78	10.5

III. RESULTS AND DISCUSSION

The lattice temperature distribution in an npn BJT with TD simulation of FLOODS is shown in Fig. 1. The base-emitter junction is forward biased at 1.0 V which drives the device into the high current injection regime and pushs the base-collector space-charge region to the edge of the collector buried layer where the hot-carrier effect takes place due to the high electric field. The electron temperature in the region is as high as 4500 K and the energy transfer from the electron to other subsystems is significant. However, the peak lattice temperature in the region is only 6 K higher than the boundaries. The base and collector currents are mainly determined by the drift-diffusion in the emitter and base regions, respectively, and the self-heating effect shows up in the physical models such as intrinsic carrier density and mobility. Since the temperature distribution is uniform in these regions, the constant temperature assumption is valid in the BJT devices. The comparisons of the running time and error between these three simulations are summarized in Table I. At the biases of $V_{\rm be} = 0.9$ V and $V_{\rm bc} = 3$ V, the error of the SNSH model is less than 2% but the solution time is 20% less than the TD simulation, whereas the error of the isothermal simulation is greater than 7% and running time is slightly faster than the SNSH simulation.

The lattice temperature along the surface of a 0.6-μm NMOS with TD simulation is plotted in Fig. 2 when the biases at drain and gate are 7 V. Due to the thermal conductances at all contacts, the device is heated up and the lattice temperature at the boundaries are at least 40 K higher than the ambient temperature. At the same operation condition, the lattice temperature of SNSH and isothermal simulations are 345 K and 300 K, respectively. Since the drain current is governed by the electron transport in the channel and the self-heating effect shows up in the mobility model, the large difference of channel lattice temperature between isothermal and TD simulations results in significant error in drain current. In contrast, the discrepancy of drain current between SNSH and TD simulations is little due to the small difference of the channel lattice temperature. Fig. 3 shows the comparisons between the TD, SNSH, and isothermal simulations of the drain current as a function of biases and the comparison between TD simulations and experimental data is shown in [5]. The drain current predicted by the TD model is always lower than that of the isothermal model because the channel lattice temperature in TD simulation is higher than that of the isothermal simulation which increases the lattice scattering and decreases the electron mobility. The degradation of the electron mobility predicted by TD model causes negative output conductance at high $V_{\rm gs}$ which is not observed in the isothermal simulation. The error of the isothermal simulation is larger than 10% when $V_{\rm gs} = V_{\rm ds} = 7$ V. For the SNSH simulation, the degradation of mobility and the negative conductance are also observed. The error of SNSH simulation is within 1.2% but the simulation time is 20% less than the TD simulation. The CPU time and error of the three simulations are summarized in Table II.

The accuracy of this approach is determined by the difference between the lattice temperature of SNSH simulation and actual lattice temperature in the critical active regions, i.e. channel in MOSFET's and intrinsic base, emitter and base-collector space charge regions in BJT's. For a device with large temperature variations in a big critical active area such as power transistor, the physical mechanisms can not correctly be described by a single lattice temperature. This could result in a larger error in the terminal current. However, this approach appears useful for submicron device structure where the lattice temperature variation in the critical active region is small and can be properly approximated by the SNSH simulations. The accuracy of SNSH model is insensitive to the size of noncritical active region since there is no heat generated in the area and has little effect to the lattice temperature of the SNSH simulations.

IV. CONCLUSION

A new algorithm which reduces the device simulation time of full-thermodynamic solutions is presented. This method utilizes one lattice temperature in single node to represent temperature in all nodes of the

device structure because of the approximately constant temperature in the device. The lattice self-heating equation is solved in the single node which reduces the size of the matrix and allows for rapid solving of the full-thermodynamic equations. The average solution time of the new method is about 20% less than the full-thermodynamic solution and good accuracy is achieved both in BJT and NMOS simulations.

REFERENCES

[1] R. Stratton, "Diffusion of hot and cold electrons in semiconductor barriers," *Phys. Rev.*, vol. 126, no. 9, pp. 2002–2014, 1962.

[2] R. K. Cook, "Numerical simulation of hot-carrier transport in silicon bipolar transistors," *IEEE Trans. Electron Devices*, vol. ED-30, pp. 1103–1110, Sept. 1983.

[3] D. Chen *et al.*, "Dual energy transport model with coupled lattice and carrier temperature," *Simulation Semiconduct. Devices Processes*, vol. 5, Sept. 1993.

[4] K. Fushinobu *et al.*, "Heat generation and transport in submicron semiconductor devices," *J. Heat Transfer*, vol. 117, pp. 25–31, 1995.

[5] M. Liang and M. E. Law, "Influence of lattice self-heating and hot-carrier transport on device performance," *IEEE Trans. Electron Devices*, vol. 41, pp. 2391–2398, Dec. 1994.

[6] V. M. Agostinelli *et al.*, "Two-dimensional energy-dependent models for the simulation of substrate current in submicron MOSFET's," *IEEE Trans. Electron Devices*, vol. 41, pp. 1784–1795, Oct. 1994.

[7] M. Liang and M. E. Law, "An object-oriented approach to device simulation—FLOODS," *IEEE Trans. Computer-Aided Design*, vol. 13, pp. 1235–1240, Oct. 1994.