Abstract—It is well known that for the design and simulation of state-of-the-art circuits thermal effects like self-heating and coupling between individual devices must be taken into account. As compact models for modern or experimental devices are not readily available, a mixed-mode device simulator capable of thermal simulation is a valuable source of information. Considering self-heating and coupling effects results in a very complex equation system which can only be solved using sophisticated techniques. We present a fully coupled electrothermal mixed-mode simulation of an SiGe HBT circuit using the design of the μA709 operational amplifier. By investigating the influence of self-heating effects on the device behavior we demonstrate that the consideration of a simple power dissipation model instead of the lattice heat flow equation is a very good approximation of the more computation time consuming solution of the lattice heat flow equation.

Index Terms—Circuit simulation, electrothermal effects, heterojunction bipolar transistors, silicon, silicon alloys, simulation, thermal power generation.

I. INTRODUCTION

Due to the ever increasing packaging density of integrated circuits, self-heating, and thermal coupling effects become more and more important. As compact models for modern submicron devices and nonmainstream devices like HBTs or HEMTs are not readily available, a device simulator capable of mixed-mode simulation is a valuable source of information for the circuit and device designer. Two different aspects must be considered and modeled properly: first, self-heating of the individual devices caused by power dissipation inside the devices and secondly, heating of the devices due to thermal coupling with other power dissipating elements. In device simulation, self-heating effects are normally considered by solving the lattice heat flow equation whereas thermal coupling effects by inclusion of a thermal network [1]–[3]. This thermal network may either be hand-crafted or automatically extracted using available geometry and material information.

Many approaches for the solution of coupled electrothermal systems rely on the coupling of simulators [3], [4]. In this paper, we present an all-in-one solution implemented in the device simulator MINIMOS-NT [5]–[7]. This thermal network may either be hand-crafted or automatically extracted using available geometry and material information.

II. LATTICE HEAT FLOW EQUATION

To account for self-heating effects, the lattice heat flow equation

$$\nabla \cdot \left( - \rho_{L} \cdot C_{L} \cdot \nabla T_{L} \right)$$

is solved. $S_{L}$ is the lattice heat flow density and the coefficients $\rho_{L}$, $C_{L}$, and $\kappa_{L}$ denote the materials mass density, specific heat, and thermal conductivity, respectively. $H$ is the generated local heat density and is modeled by the expression given by Adler [8]

$$H = \nabla \cdot \left( \frac{E_{C}}{q} \cdot J_{n} + \frac{E_{V}}{q} \cdot J_{p} \right)$$

$E_{C}$, $E_{V}$, $J_{n}$, and $J_{p}$ are the conduction band edge energy, the valence band edge energy, the electron and hole current densities, respectively. Equation (3) accounts for both Joule heat generation and recombination heat. However, the influence of thermo-electric effects (Seebeck, etc.) is neglected in both (1) and (3) [9]. Of course, temperature dependent models are used for all physical parameters needed in the device equations, e.g., for the band edge energies, recombination rates, and mobilities [10], [11].

Two different thermal contact models are commonly used. The first model implements an isothermal contact by simply setting the lattice temperature at the interface points equal to the contact temperature (Dirichlet boundary condition)

$$T_{L} = T_{C}$$

The second model is of Cauchy type, it considers a thermal contact resistance at the contact boundary and determines the normal component of the flux. Thus, the expression for the thermal heat flow density $S_{L}$ at the contact reads

$$n \cdot S_{L} = \frac{T_{L} - T_{C}}{\rho_{th}}$$

with $\rho_{th}$ being the thermal contact resistivity and $n$ the normal vector to the surface. The thermal contact conductance $G_{th}$ is
related to the thermal resistivity $\rho_{th}$ by $G_{th} = A/\rho_{th}$ with $A$ as the contact area.

III. EVALUATION OF THE BOUNDARY CONDITIONS

As an example device we consider a SiGe HBT structure as investigated in [12]. However, an additional thermal contact has been added at the right side of the device assuming a mirror symmetry of the device structure. This contact models the thermal heat flow along the chip surface. The resulting device structure is shown in Fig. 1. Also shown is the region of maximum heat generation which is in the base-collector space-charge region were the maximum collector-emitter voltage drop occurs. It must be kept in mind that the ratio of the heat flows over these four contacts is determined by the design of the chip and the environment it is used in. For the following we assume properly designed heat sinks which drain the generated heat mainly toward the collector and along the chip surface.

For the simulation, unless otherwise noted, the following thermal contact conductances were used: $G_{th}^E = G_{th}^B = 10$ mW/K and $G_{th}^C = G_{th}^{Chip} = 50$ mW/K. Note that the thermal boundary conditions are determined by the thermal resistors at material transitions as much as by the bulk properties. As these unknown thermal resistivities of industrially relevant materials, such as, e.g., glue or thermal bumps, have to be verified by experiments, a self consistent fitting procedure and various assumptions will always be part of application oriented simulations. Thus the thermal boundary conditions have to be determined with respect to aspects normally not included in device simulation such as neighboring devices or chip mounting. Simulated temperature cross sections through the center of the device are shown in Figs. 2 and 3 for the isothermal and the resistance contact model, respectively, with $V_{CE} = 3.5$ V. Fig. 2 shows the temperature distribution for different base-emitter voltages $V_{BE}$ whereas for Fig. 3 $V_{BE} = 1.0$ V was used and $G_{th}^{Chip} = G_{th}^{Chip}$ were varied. For $V_{BE} = 1.0$ V both contact models generate temperature distributions of similar shape but in the case of the resistance contact model the temperature is shifted by an offset which exponentially depends on $G_{th}^{Chip}$. For $G_{th}^{Chip}$ as small as 10 mW/K no meaningful solution could be found as the lattice temperature exceeds 600 K which inhibits a successful simulation since the temperature dependent models leave their range of validity. Furthermore, other effects like impact ionization become important for higher collector-emitter voltages. Although $V_{BE} = 1.0$ V is quite high it must be pointed out that even for lower bias conditions the same situation occurs for improper choice of $G_{th}^{Chip}$.

These investigations show that the simulation result is very sensitive to the contact resistances. Furthermore, it follows that the isothermal model must be used with great care.
In Fig. 4, the heat generation inside the device is shown for different values of $G_{\text{Th}}^\text{Chip}$. As the current density remains approximately constant within this cross-section, the maximum of the heat generation is located at the base-collector space charge region where the electric field is maximal. As $V_{CE} = 3.5$ V was assumed which is quite moderate, even higher heat generation rates can be expected for power circuits. Although the final values may give reasonable temperature distributions, during iteration the bias voltages of a device in a circuit may vary considerably and can easily exceed $V_{BE} = 1.5$ V and $V_{CE} = 20$ V. This situation can occur during mixed-mode simulation of circuits with large supply voltages and cause excessive problems when simulating fully coupled electrothermal systems especially as measured values for $G_{\text{Th}}^\text{Chip}$ are in the range 1–10 mW/K.

In Fig. 5, the temperature distribution for different base-emitter voltages is shown generated with a quite large value of 100 mW/K for $G_{\text{Th}}^\text{Chip}$. All these figures indicate, that the heat generated inside the device accumulates because it cannot be drained off by the thermal contacts. Thus, the local temperature rise inside the device is much smaller than the temperature rise induced by the contact model. This is especially true for this example as silicon is a good thermal conductor and large temperature gradients are not likely to occur inside the device.

It might therefore be unnecessary to perform a fully consistent self-heating (SH) simulation by solving the lattice heat flow equation. Instead, we could use a global self-heating model (GSH) and calculate the dissipated power as

$$P = \sum_C I_C \cdot V_C$$

with $I_C$ and $V_C$ being the contact currents and voltages. The spatially constant lattice temperature is modeled as

$$T_L = T_C + P \cdot R_g$$

with $R_g$ being the global thermal resistance. This model is commonly used in compact modeling (e.g., [1], [3]). However, when applying this expression to mixed-mode device simulation, each device is modeled at a distinct device-specific temperature which has a significant impact on device performance. With this approach it is thus possible to make use of all temperature dependent physical parameter models, e.g., mobility, thus significantly increasing simulation accuracy. The GSH model gives only two additional unknowns ($T_L$ and $P$) compared with the pure electrical system. $R_g$ should be equal to the effective thermal contact resistance plus an equivalent resistance of the device which can be approximated as

$$R_g = R_{g}^{\text{eff}} + \frac{w}{A \cdot \kappa}$$

with $w$ being the average distance of the thermal contact to the region where the heat is generated and $A$ being the average area of the section connecting the junction with the thermal contact. $\kappa$ is the thermal conductivity of the underlying material which shows a strong temperature dependence [13] and must therefore be evaluated at an average temperature value. Of course, this formula is far too simple to give exact results and it is better to consider $R_g$ a mere fitting parameter approximated by (8). Due to the large uncertainties in the temperature dependence of the thermal contact resistances, $R_g$ was assumed to be temperature independent for the simulations presented here.

Two models are mainly used to model the electrical properties of semiconductor devices, both of which can be derived from Boltzmann’s transport equation: the drift-diffusion [13] and the more complex hydrodynamic transport model [14], [15]. Both models give an additional diffusion current caused by carrier temperature gradients. For the drift-diffusion model the carrier subsystems are assumed to be in thermal equilibrium with the lattice temperature, whereas in the hydrodynamic model carrier temperature relaxation to the lattice temperature is modeled by an energy relaxation time. Both transport models normally assume a constant lattice temperature but can be consistently extended to nonconstant lattice temperatures [16].
port model is irrelevant in the following context, the simpler drift-diffusion model has been used in all examples.

The GSH model has been implemented and verified against the SH model. Furthermore, simulation times and convergence properties have been compared to the purely electrical system which has been solved at $K_T (T=300 \text{ K})$. For the SH model the thermal contact resistances at the electrical contacts were neglected to yield a maximum nonuniform temperature distribution inside the device (worst-case). The transfer and output characteristics are shown in Figs. 6 and 7, respectively. The GSH model with $K/W$ nicely reflects the electrical terminal characteristics of the SH model. Furthermore, the GSH model shows convergence properties similar to the T300 model whereas the SH model causes convergence problems for large lattice temperatures and no solution could be obtained for $V_{BE} > 1.03 \text{ V}$. Computational details for the calculation of the transfer characteristic ($V_{BE} = 0.7$–$0.9 \text{ V}$, $V_{CE} = 3.5 \text{ V}$, 20 points) can be found in Table I. The GSH takes only 33% longer than the T300 model as opposed to the 144% of the SH model. This is a significant improvement considering that for the simulation a fully temperature dependent set of physical parameters has been used, including strong nonlinearities.

On the other hand, one should be aware of the simplifications introduced by the GSH model. As the device temperature is the same for the whole device, the additional component in the diffusion current caused by temperature gradients is neglected. Furthermore, it is obvious that no information about hot-spots inside the device can be extracted from these simulations.

For devices similar to the example HBT where self-heating is mainly determined by the contact model and thus the GSH forms a reasonable approximation, the results obtained by the GSH model give a good initial guess for a following SH simulation. Simulations based on this iteration scheme resulted in a 10% reduction of simulation time for lower bias. This is because the global temperature resulting from GSH model tends to overestimate the lattice temperature in many device regions, thus reducing the quality of the initial guess for higher biases where self-heating becomes more important.

### IV. EXAMPLE CIRCUIT

Thermal effects are of crucial importance for the chip design of integrated circuits. Typical operational amplifiers (OpAmps) can deliver powers of 50–100 mW to a load, and as the output stage internally dissipates similar power levels the temperature of the chip rises in proportion to the dissipated output power [17], [18]. As the transistors are very densely packed, self-heating of the output stage will affect all other transistors. This is especially true as silicon is a good thermal conductor, so the whole chip tends to rise to the same temperature as the output stage. However, small temperature gradients develop across the chip with the output stage being the heat source. These temperature gradients appear across the input components of the OpAmp and induce an additional input voltage difference which is proportional to the output dissipated power.

A SiGe HBT circuit based on the schematic of the $\mu A709$ [19], [20] as shown in Fig. 8 has been simulated considering thermal interaction between the input and the output stage. This circuit is of special interest as it is one of the SPICE benchmark circuits given in [21]. The dc transfer characteristic has been calculated with and without thermal interaction. Consideration of thermal interaction was first performed by solving the SH model for the transistors $T_1$, $T_2$, $T_5$ and $T_{15}$ and by assuming a thermal network as shown in Fig. 9. The thermal conductances were assumed to be $G_1 = G_2 = 2 \text{ mW/K}$ and

<table>
<thead>
<tr>
<th>Method</th>
<th>System-Size</th>
<th>CPU</th>
</tr>
</thead>
<tbody>
<tr>
<td>T300</td>
<td>2560</td>
<td>61 s</td>
</tr>
<tr>
<td>SH</td>
<td>3469</td>
<td>152 s</td>
</tr>
<tr>
<td>GSH</td>
<td>2563</td>
<td>81 s</td>
</tr>
</tbody>
</table>

**TABLE I**

**Computational Details for the Comparison of the SH with the GSH Model for the Example HBT. The T300 Model Is Shown as Reference**
No problems occurred during the solution of the purely electrical system. Even the consideration of thermal interaction using the GSH model caused no problems. The solution of the fully coupled electrothermal equation system considering the SH model was possible with a properly designed iteration scheme. This is because a small change in the output voltage during iteration causes a large change in the collector current of the conducting output transistor. Hence, the dissipated power changes and so does the temperature distribution inside the output transistor. This modified power alters the base-emitter voltages of the input transistors which produces a change in the base-emitter voltages of the output transistors. All these coupling mechanisms are highly nonlinear.

The iteration scheme works as follows: in the first block the thermal quantities were ignored until an electrical solution was found. In the second block, the lattice temperature was added to the solution vector without considering the coupling effects caused by the node temperatures. This was also found to be advantageous when stepping through the dc transfer curve hence this block was also used for the consecutive steps. After having established a proper temperature distribution inside the devices for the new voltage boundary conditions, the complete equation system can be used.

The dc transfer characteristic was calculated by stepping \( \varphi_{in} \) from \(-1\) mV to 1 mV with \( \Delta\varphi_{in} = 20 \) mV. From SPICE simulations the open-loop gain of the \( \mu A709 \) was known to be approximately 35 000 so for each step of \( \Delta\varphi_{in} \) a step of 0.7 V could be expected for \( \Delta\varphi_{out} \) which is quite large. However, no convergence problems occurred until \( \varphi_{out} \) approached 0 V. This was the most critical part of the simulation and several step reductions for the input voltage were necessary for the SH model. Details of the simulations are summarized in Table II.

The dc transfer characteristic is shown in Fig. 10 with the obvious humps in the SH models resulting from thermal feedback effects. The GSH model perfectly fits the results obtained by the more complex SH model. In Fig. 11 the open-loop voltage gain \( A_v \) is shown demonstrating the dramatic impact of thermal coupling. The thermal conductances assumed in this simulation were very optimistic and an even stronger impact of thermal coupling has been published [2], [4]. For stronger coupling, even the sign of the open-loop voltage gain may change and cause the OpAmp to become unstable.

The maximum temperature and the contact temperature of the output stage are shown in Fig. 12. Information about the maximum temperature is lost for the GSH model, though. As can be seen self-heating inside the transistor plays only a minor role at these current levels. However, the power dissipated inside the device heats up the NPN transistor due to the resistive
thermal boundary condition which obstructs the heat flow out of the transistor. This is the reason for the excellent results obtained by the simple GSH model. The PNP transistor has a $\beta$ of only approximately ten and comparable current levels have been obtained by increasing the emitter area of the transistor ($W_{\text{PNP}}/W_{\text{NPN}} = 5$). Hence the locally generated heat density $H$ is even smaller than for the NPN transistor and the temperature drop inside the device is negligible thus resulting in nearly no loss of information for the GSH model.

A similar situation occurs for the input transistors $T_1$ and $T_2$, as shown in Fig. 13. As they are biased with $I_C = 20 \mu A$ only self-heating is negligible and the contact temperature resembles the heat transferred from the output stage thus again resulting in negligible loss of information for the GSH model. As unsymmetric thermal conductivities have been assumed the temperature of $T_1$ is always slightly higher than the temperature of $T_2$. The maximum temperature difference of the input transistors $T_1 - T_2$ was found to be only 22 mK. Even this small temperature difference has such a strong impact on the output characteristic due to the high gain of the circuit.

V. CONCLUSIONS

We have investigated the impact of a computation time efficient approach to cover self-heating effects on device and circuit performance. It was shown in a realistic example that self-heating is dominated by the resistive thermal boundary conditions. Thus, the lattice heat flow equation can be substituted by a global self-heating model with nearly no loss of accuracy in the electrical terminal characteristic. This observation is of fundamental importance in the case of mixed-mode device simulations where thermal-coupling effects dramatically increase the complexity of the problem. Using this approximation the problem can be solved in considerably less time with reasonable
accurate inclusion of thermal effects. The benefits provided by this approach can be even better exploited in three-dimensional device simulations as there the reduction in the number of unknowns is obviously even more significant.

ACKNOWLEDGMENT

The authors would like to thank Dr. M. Knaipp for his basic work on the thermal models implemented in MINIMOS-NT and the valuable discussions resulting therefrom. They also appreciate the discussions with R. Quay during the evolution of this paper.

REFERENCES


Tibor Grasser was born in Vienna, Austria, in 1970. He received the Diplomingenieur and doctoral degrees in technical sciences from the Technische Universität Wien, Wien, Austria, in 1995 and 1999, respectively.

He joined the Institute for Microelectronics, Technical University of Vienna, Vienna, Austria, in April 1996. From October to December 1997, he was with Hitachi, Ltd., Tokyo, Japan, as a Visiting Research Engineer. In February 1998, he was appointed Assistant Professor at the Institute for Microelectronics. His scientific interests include circuit and device simulation, device modeling, physical and software aspects in general.

Siegfried Selberherr (M’79–SM’84–F’93) was born in Klosterneuburg, Austria, in 1955. He received the Diplomingenieur and doctoral degrees in electrical engineering and technical sciences from the Technische Universität Wien, Wien, Austria, in 1978 and 1981, respectively.

Dr. Selberherr has held the “venia docendi” in computer-aided design since 1984. Since 1988, he has been the Head of the Institute for Microelectronics, Technical University of Vienna, Vienna, Austria, and since 1999 he has been Dean of the Fakultät für Elektrotechnik, Technische Universität Wien. His current research topics are modeling and simulation of problems for microelectronics engineering.