Modeling and Simulation of Electrothermal Effects in Power Semiconductor Devices

H. Brand and S. Selberherr

Institute for Microelectronics,Technical University of Vienna
Gußhausstraße 27-29, A-1040 Vienna, Austria

An advanced model for self-heating effects in power semiconductor devices is presented. Based on irreversible thermodynamics it is valid in both the stationary and transient regimes. Numerical methods to solve the governing equations for the coupled transport of charge carriers and heat are described. Finally results obtained in simulating thermal runaway in a GTO-thyristor and latch-up in an IGT are discussed.

1. INTRODUCTION

Device simulators are widely used to optimize semiconductor devices and to develope new technologies. To allow accurate computation of device characteristics and analysis of parasitic effects in modern power semiconductor devices new simulation tools are needed. In order to account properly for the electrothermal nature of such phenomena as thermal runaway, current crowding, avalanche injection, secondary breakdown and latch-up, they must be based on transport models for the simultaneous flow of charge carriers and heat. From a numerical point of view the simulation of transient self-heating effects in semiconductor devices requires the self-consistent solution of Poisson's equation, the continuity equations for electrons and holes and the heat flow equation in space and time.

2. THE MATHEMATICAL MODEL

To derive a closed set of equations for the simultaneous flow of charge carriers and heat in a semiconductor device, principles of irreversible hydrothermodynamics are utilized [2], [5], [8]. The idea is to explicitly set up the entropy balance equation for the semiconductor, to identify thermody-

namic currents and conjugated forces and to formulate the phenomenological equations. To obtain governing equations describing mutually coupled transport phenomena, transport parameters have to be defined and specific state functions have to be evaluated [2], [5], [8].

Provided that the concept of local equilibrium is valid, the Gibbs fundamental equation for the semiconductor takes the form:

$$\frac{\partial u}{\partial t} = T \cdot \frac{\partial s}{\partial t} - q \cdot \varphi_n \cdot \frac{\partial n}{\partial t} + q \cdot \varphi_p \cdot \frac{\partial p}{\partial t} \tag{1}$$

u denotes the internal energy per unit volume, s the entropy density. n means the electron, p the hole concentration. T is the temperature, φ_n, φ_p are the quasi-Fermi levels for electrons and holes, respectively.

The Gibbs equation (1) interrelates balances of electrodynamic and thermodynamic quantities. Continuity equations for charge carriers together with Poisson's equations have to be derived from Maxwell's equations [9]. The conservation equation for the energy density u represents the first axiom of thermodynamics in terms of vector analysis for continuous systems. Momentum balance equations are not considered for reasons of simplicity. The actual form of the entropy balance equation is obtained by substitution of continuity equations for the carrier concentrations n, p and the energy density u into the Gibbs fundamental equation (1),

$$\begin{aligned} \frac{\partial s}{\partial t} + \operatorname{div}\left(\frac{1}{T} \cdot \vec{J}_q\right) = \quad & \vec{J}_q \cdot \left(-\frac{1}{T^2} \cdot \operatorname{grad} T\right) + \vec{J}_n \cdot \left(-\frac{1}{T} \cdot \operatorname{grad} \varphi_n\right) + \\ & \vec{J}_p \cdot \left(-\frac{1}{T} \cdot \operatorname{grad} \varphi_p\right) + R \cdot \left(\frac{1}{T} \cdot q \cdot (\varphi_p - \varphi_n)\right) \end{aligned} \tag{2}$$

where

$$\vec{J}_q \equiv \quad \vec{J}_u - \varphi_n \cdot \vec{J}_n - \varphi_p \cdot \vec{J}_p \tag{3}$$

is a definition of the heat flux $\vec{J}_q$. $\vec{J}_n$ denotes the electron, $\vec{J}_p$ the hole current density, respectively.

The right-hand side of (2) is the local production of entropy. Due to the first postulate of irreversible thermodynamics it can be interpreted as the sum of products of thermodynamic currents and conjugated forces. Furthermore, exploiting the second postulate of irreversible thermodynamics, that each flux linearly depends on all thermodynamic forces [2], the so-called phenomenological equations result:

$$\begin{pmatrix} \vec{J}_n \\ \vec{J}_p \\ \vec{J}_q \\ R \end{pmatrix} = \begin{pmatrix} L_{nn} & 0 & L_{nT} & 0 \\ 0 & L_{pp} & L_{pT} & 0 \\ L_{Tn} & L_{Tp} & L_{TT} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} -\frac{1}{T} \operatorname{grad} \varphi_n \\ -\frac{1}{T} \operatorname{grad} \varphi_p \\ -\frac{1}{T^2} \operatorname{grad} T \\ \frac{1}{T} \cdot q \cdot (\varphi_p - \varphi_n) \end{pmatrix} \tag{4}$$

Due to Onsager's symmetric reciprocal relations, known as the third postulate of irreversible thermodynamics, the number of independent kinetic coefficients in equation (4) can be reduced. Thus only conductivities σ and thermoelectric powers P for electrons n and holes p, respectively, have to be defined, together with the thermal conductivity κ. The resulting current relations (5) - (7) for the thermoelectric transport in a semiconductor include both limiting cases, (i) flow of electric charge due to the imposition of the quasi-Fermi potential, and (ii) flow of heat caused by a temperature gradient:

$$\vec{J}_n = \sigma_n \cdot (-\operatorname{grad} \varphi_n + P_n \cdot \operatorname{grad} T) \tag{5}$$

$$\vec{J}_p = \sigma_p \cdot (-\operatorname{grad} \varphi_p - P_p \cdot \operatorname{grad} T) \tag{6}$$

$$\vec{J}_q = -T \cdot P_n \cdot \vec{J}_n + T \cdot P_p \cdot \vec{J}_p - \kappa \cdot \operatorname{grad} T \tag{7}$$

With the heat flux (7) the entropy balance equation (2) can be transformed into the heat flow equation (8):

$$T \cdot \frac{\partial s}{\partial t} + \operatorname{div}(-T \cdot P_n \cdot \vec{J}_n + T \cdot P_p \cdot \vec{J}_p - \kappa \cdot \operatorname{grad} T) = -\vec{J}_n \cdot \operatorname{grad} \varphi_n - \vec{J}_p \cdot \operatorname{grad} \varphi_p + q \cdot R \cdot (\varphi_p - \varphi_n) \tag{8}$$

The expansion of the divergence operator and some algebraic operations yield:

$$T \cdot \frac{\partial s}{\partial t} + \operatorname{div}(-\kappa \cdot \operatorname{grad} T) = \frac{\vec{J}_n \cdot \vec{J}_n}{\sigma_n} + \frac{\vec{J}_p \cdot \vec{J}_p}{\sigma_p} + \vec{J}_n \cdot (\operatorname{grad}(T \cdot P_n) - P_n \cdot \operatorname{grad} T) - \vec{J}_p \cdot (\operatorname{grad}(T \cdot P_p) - P_p \cdot \operatorname{grad} T) + q \cdot R(\varphi_p - \varphi_n) + T \cdot P_n \cdot \operatorname{div} \vec{J}_n - T \cdot P_p \cdot \operatorname{div} \vec{J}_p \tag{9}$$

The right-hand side of equation (9) is the heat generation. It can be shown that (9) implies Thomson's laws.

To proceed the entropy has to be regarded as a state function of the temperature and the carrier concentrations $s = s(T, n, p)$. Furthermore, the quasi-Fermi levels are considered as state functions of the electrostatic potential ψ, the carrier concentrations n, p, the temperature T, and the effective intrinsic carrier concentration n_{ie} to fit heavy doping effects [9]. Using Maxwell's relations [2], [10] and Boltzmann statistics following system of nonlinear, coupled, partial differential equations results:

$$\operatorname{div} \operatorname{grad} \psi = \frac{q}{\varepsilon}(n - p - C) \tag{10}$$

$$\operatorname{div} q\cdot\mu_n\cdot n\cdot\left(\vec{E}-\frac{kT}{q}\operatorname{grad}(\ln n_{ie})+\frac{kT}{q}\frac{1}{n}\operatorname{grad} n+P_n^{eff}\operatorname{grad} T\right)= q\cdot\frac{\partial n}{\partial t}+q\cdot R \quad (11)$$

$$\operatorname{div} q\cdot\mu_p\cdot p\cdot\left(\vec{E}+\frac{kT}{q}\operatorname{grad}(\ln n_{ie})-\frac{kT}{q}\frac{1}{p}\operatorname{grad} p-P_p^{eff}\operatorname{grad} T\right)= -q\cdot\frac{\partial p}{\partial t}-q\cdot R \quad (12)$$

$$C\cdot\frac{\partial T}{\partial t}+\operatorname{div}(-\kappa\operatorname{grad} T)=\frac{J_n^2}{\sigma_n}+\frac{J_p^2}{\sigma_p}+q\cdot R\left(2\cdot T\cdot\frac{kT}{q}\frac{1}{n_{ie}}\frac{\partial n_{ie}}{\partial T}\right)+ \vec{J}_n T\operatorname{grad} P_n-\vec{J}_p T\operatorname{grad} P_p+TP_n^{eff}\operatorname{div}\vec{J}_n-TP_p^{eff}\operatorname{div}\vec{J}_p \quad (13)$$

$$\left(\frac{k}{q}\ln\frac{c}{n_{ie}}-\frac{kT}{q}\frac{1}{n_{ie}}\frac{\partial n_{ie}}{\partial T}\right)+P_c\equiv P_c^{eff} \quad (14)$$

Equation (10) is the well known Poisson equation. ψ denotes the electrostatic potential, C the total net concentration of all ionized impurities. q is the elementary charge, ε the permittivity. Equations (11), (12) are continuity equations for electrons n and holes p including the so-called extended-drift extended-diffusion approximation of the current relations. That is, an effective electric field is introduced because of the possible dependence of the intrinsic concentration on position, and the diffusion not only accounts for concentration gradients but also for temperature gradients. μ denotes the mobility, k the Boltzmann constant. $\vec{E}$ is the electric field. P denotes the thermoelectric power, c means 'carrier' and may take the value n or p for electrons and holes, respectively. Equation (13) is the heatflow equation. The right-hand side represents the heat generation. Four contributions to the heat generation can be distinguished: Joule heat, recombination heat, Thomson heat and carrier source heat. Equation (14) is a definition of the effective thermoelectric power P_c^{eff}.

The governing equations depend non-linearly on the lattice temperature. If lattice heating is significant, the thermal system becomes tightly coupled to the electrical system. Recombination, mobility, thermoelectric power and the effective intrinsic carrier density n_{ie} depend on temperature [9], while the temperature gradient acts as a driving force and the heat generation is a function of the electrical variables.

Auger recombination and carrier-carrier scattering - known as limiting physical effects for high injection conditions in power semiconductor devices [9] - have been taken into account. As heavy doping effects limit thyristor operation, the effective intrinsic carrier concentration is com-

puted using [9]. One obtains the relation:

$$\frac{\partial n_{ie}}{\partial T} = n_{ie} \cdot \left(\frac{E_g}{2 \cdot k \cdot T^2} + \frac{3}{2 \cdot T} \right)$$
$$- n_{ie} \cdot \left(\frac{q \cdot V_1 \cdot \left(\ln\left(\frac{N_D^+ + N_A^-}{N_0} \right) + \sqrt{\ln^2\left(\frac{N_D^+ + N_A^-}{N_0} \right) + 0.5} \right)}{2 \cdot k \cdot T^2} \right)$$
$$+ n_{ie} \cdot \left(\frac{9.025 \cdot 10^{-5} \ \frac{eV}{K} + 6.10 \cdot 10^{-7} \ \frac{eV}{K} \cdot \left(\frac{T}{K} \right)}{2 \cdot k \cdot T} \right)$$
$$+ n_{ie} \cdot \left(\frac{3}{4} \cdot m_0 \cdot \left(\frac{1.4 \cdot 10^{-4} \ \frac{1}{K} \cdot m_n^* + 4.5 \cdot 10^{-4} \ \frac{1}{K} \cdot m_p^*}{m_n^* \cdot m_p^*} \right) \right) \quad (15)$$

Note also that the dependence of the bandgap and the effective masses on temperature is accounted for in equation (15). Following [3], [6] the effective thermoelectric powers P_n^{eff}, P_p^{eff} can be expressed as functions of contributions to the overal mobility:

$$P_c^{eff} = \frac{k}{q} \left(\frac{2 - x_c^2 + x_c^3(\mathrm{Ci}(x_c)\sin(x_c) - (\mathrm{Si}(x_c) - \frac{\Pi}{2})\cos(x_c))}{1 + x_c^2(\mathrm{Ci}(x_c)\cos(x_c) + (\mathrm{Si}(x_c) - \frac{\Pi}{2})\sin(x_c))} - \frac{3}{2} \right) \quad (16)$$

$$x_c = \sqrt{6 \cdot \frac{\mu_c^L}{\mu_c^I}} \quad (17)$$

The index c denotes the carrier type. μ^L is the mobility due to lattice scattering, μ^I due to impurity scattering.

For the electrical subsystem, either Dirichlet boundaries at ohmic contacts or Neumann boundaries are assumed. Mixed boundary conditions for the heat flow equation are mandatory in order to be able to model realistic imperfect cooling conditions. This is of special importance for transient electrothermal simulations, as the time constant for self-heating increases with increasing external thermal resistance.

3. SOLUTION METHODS

The simulation of the coupled transport of heat and charge carriers requires the self-consistent solution of Poisson's equation (10), the continuity equations for electrons n (11) and holes p (12) and the heat flow equation (13) in space and time. Spatial discretization is done using finite boxes [4], a generalization of finite differences, while time is discretized with the backward Euler method.

The solution of the semiconductor equations typically exhibits a smooth behavior in neutral regions of the device domain whereas in space charge regions it varies rapidly. Thus an adaptive, strongly nonuniform mesh is mandatory. Using finite boxes the number of grid points is drastically reduced compared to classical finite differences since the grid can be refined in regions where the solution changes rapidly, e.g. in space charge regions. Furthermore non-rectangular device geometries can be resolved with finite boxes. During transient simulations the grid is adapted to the actual solution after every timestep. Thus the grid generator provides a 'moving grid'.

In a mesh consisting of finite boxes meshlines may end within the simulation area. At the termination point of a meshline a neighbour is missing. In order to obtain a difference equation of the same order as at a regular inner point the first derivatives have to be used for linear interpolation [4], [9], resulting in a 6-point instead of a 5-point scheme for the difference approximation of the Laplace operator.

Specific difficulties are involved in the discretization of the current relations. As the carrier concentrations exhibit in general an exponential behavior between neighbouring mesh points depending upon the electrostatic potential, the estimation of the current relations at half points is a tricky task. Straight forward discretization turns out to be numerically unstable. Usually the midinterval values of the quantities J_{nx}, J_{ny} are computed by solving an ordinary differential equation which has to be modified for nonuniform temperature simulations:

$$\frac{\partial n}{\partial x} - \frac{\frac{\partial}{\partial x}\left(\psi - w \cdot \frac{kT}{q}\right)}{\frac{k \cdot T}{q}} \cdot n = \frac{1}{\mu_n \cdot \frac{k \cdot T}{q}} \cdot Jnx|_{i+\frac{1}{2},j} \tag{18}$$

w is a weighting factor implying the effective thermoelectric power. The potential ψ in equation (18) contains the contribution due to bandgap narrowing. The solution to equation (18) can be calculated in a straightforward manner and reads:

$$J^n_{i+\frac{1}{2},j} = \frac{\mu_n}{h_i} \cdot \frac{\frac{k}{q} \cdot (T_{i+1,j} - T_{i,j})}{\ln \frac{T_{i+1,j}}{T_{i,j}}} \cdot \left(n_{i,j} \cdot B(-x_B) - n_{i+1,j} \cdot B(x_B) \right) \tag{19}$$

B denotes the Bernoulli function defined as follows:

$$B(x_B) = \frac{x_B}{e^{x_B} - 1} \tag{20}$$

x_B is the argument of the Bernoulli function:

$$x_B = \left(\frac{(\psi_{i+1,j} - \psi_{i,j}) - w \cdot \frac{k}{q} \cdot (T_{i+1,j} - T_{i,j})}{\frac{k}{q} \cdot (T_{i+1,j} - T_{i,j})} \cdot \ln \frac{T_{i+1,j}}{T_{i,j}} \right) \tag{21}$$

The electrothermal problem is solved selfconsistently following a decoupled approach. At each time step the electrical subsystem of equations is solved first, the lattice temperature being regarded as an independent variable. Then the temperature distribution is updated by solving the heat flow equation. Newton's method and LU-decomposition are used to solve the electrical and thermal subsystems alternately until convergence is attained.

Usually the characteristic time for the electrical and the thermal subsystem differ by several orders of magnitude. Thus a device is in steady state from the electrical point of view, whereas thermally it is still in transition. Therefore the electrical transient can be neglected with respect to the thermal transient, an assumption which is well satisfied in practice.

4. APPLICATIONS AND RESULTS

The first relevant application investigated is thermal runaway in a GTO-thyristor. Thermal runaway is a phenomenon of electrothermal interaction where the dissipation of electrical energy causes a temperature rise over the entire area of the device, resulting in increased current flow and further dissipation, until unrecoverable device failure or burn out of the device occur. The heat sink temperature is 300 K. Double sided cooling is assumed.

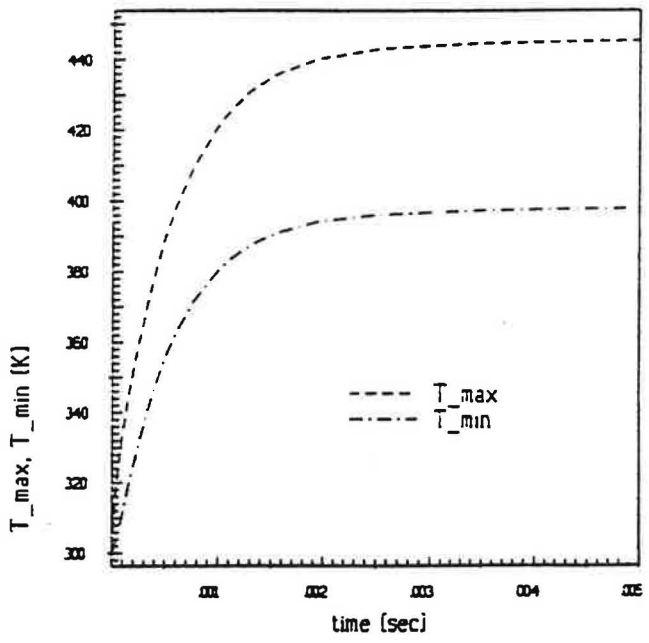

Fig. 1: time evolution of maximum and minimum temperature when the heat sink thermal conductance is 50 $Wcm^{-2}K^{-1}$ and the heat sink temperature is 300 K

Geometry and doping data are taken from [7]. The heat sink thermal conductance is 50 W/cm^2K in the first simulation. Fig. 1 shows the evolution of the internal maximum and minimum temperatures in time. Thermoelectric equilibrium is reached after 3 milliseconds. The data agree very well to the analytical solution of a thermal network, consisting of a thermal capacitor and a thermal resistor.

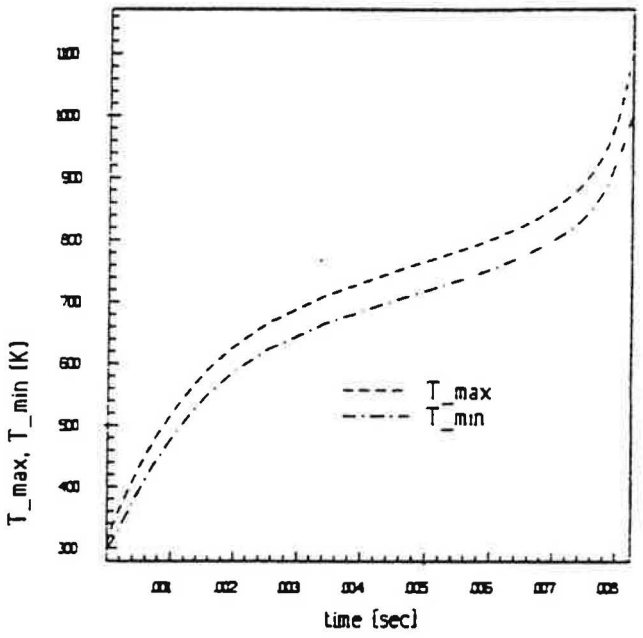

Fig. 2: time evolution of maximum and minimum temperature when the heat sink thermal conductance is 5 $Wcm^{-2}K^{-1}$ and the heat sink temperature is 300 K (thermal runaway)

In order to induce severe thermal runaway, bad cooling conditions are defined by choosing 5 W/cm^2K for the heat sink thermal conductance. Fig. 2 shows the exponential increase of the temperature.

The onset of thermal runaway occurs when thermal generation of carriers takes place not only in the small p-base region near the gate contact, but covers a considerable part of the whole device. Fig. 3 shows the thermal generation after 7.6 milliseconds. Carriers are already thermally generated across the whole base of the thyristor. Fig. 4 shows the corresponding exponential increase of the anode current. As current increases the Joule heating increases too, thus leading to a further increase of the device temperature.

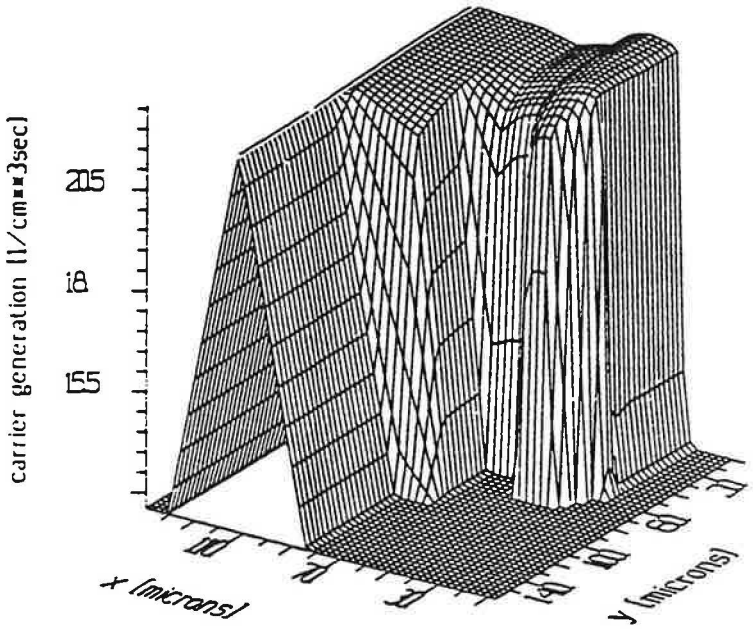

Fig. 3: thermal generation of carriers per cubic centimeter and second in case of thermal runaway after 7.6 milliseconds

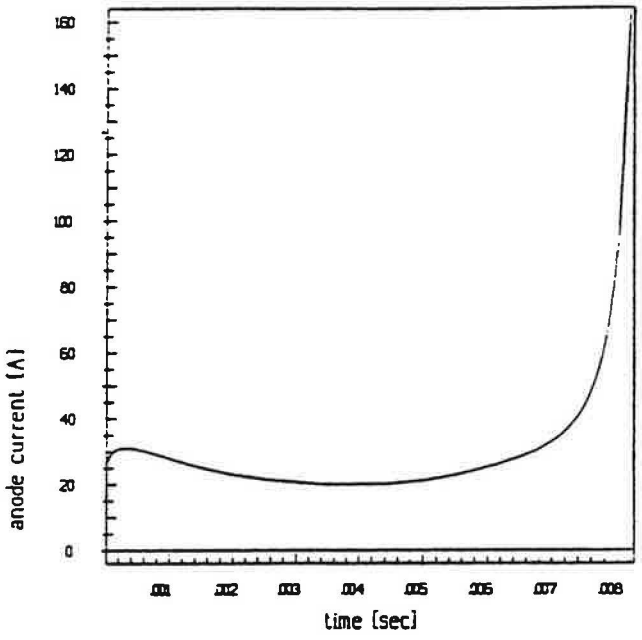

Fig. 4: time evolution of anode current in case of thermal runaway

Another application is the simulation of latch-up in an insulated-gate transistor (IGT). An IGT combines bipolar conduction with MOS gate control of the current. If, however, the parasitic thyristor in an IGT structure turns on, the control of the collector current by the applied gate voltage is lost. In DC circuits latch-up usually produces catastrophic failure of the device as a result of excessive heat dissipation.

Static latch-up has been investigated in an IGT. Geometry and doping data were taken from [1].

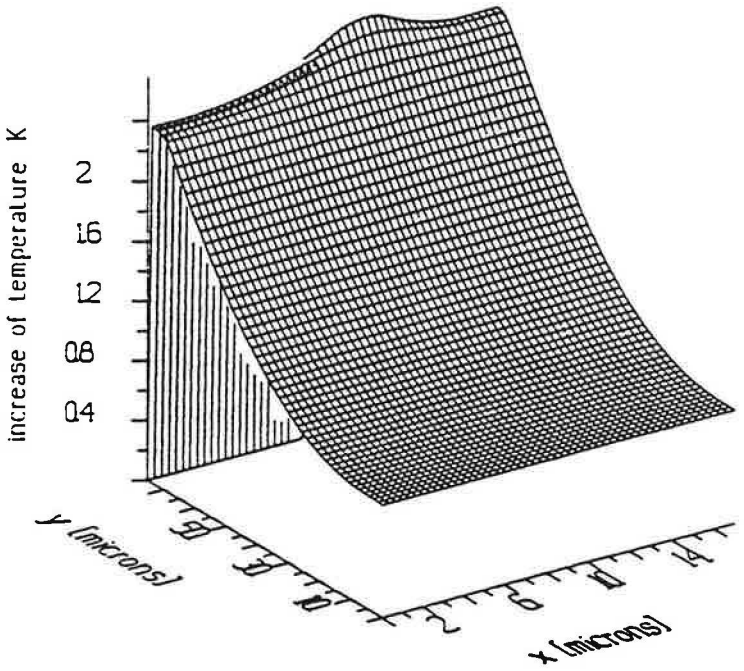

Fig. 5: increase of lattice temperature in Kelvin before latch-up ($50 W cm^{-2} K^{-1}, V_g = 15V, V_c = 2.5V$)

Fig. 5 shows the increase of the temperature in the channel region before latch-up occurs. Fig. 6 is a snapshot of the electron concentration after onset of latch-up. The heat sink thermal conductance is 50 W/cm^2K.

The IGT starts to latch up when the emitter injects electrons into the p-base during device operation. In silicon the critical value is exceeded when the n^+-emitter-p-base junction becomes forward-biased by more than 0.7 volt, usually because of lateral current flow within the p-base. Thus latch-up limits the current handling capability of the IGT. It is found that the latching current is reduced with increasing temperature.

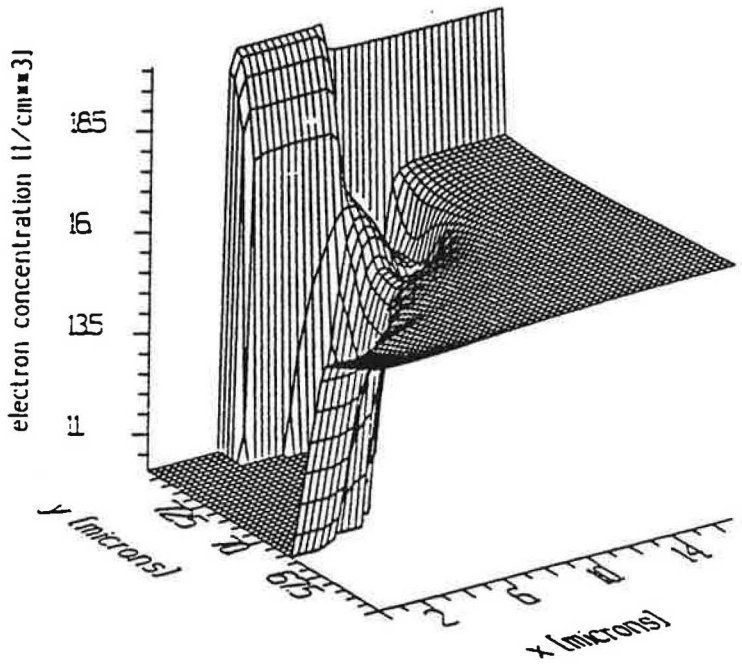

Fig. 6: snapshot of electron concentration during latch-up, $(50 W cm^{-2} K^{-1}, V_g = 15V, V_c = 2.903V)$

5. CONCLUSION

An advanced model for self-heating effects has been derived from principles of irreversible thermodynamics. It is valid in both the stationary and the transient regimes.

Thermal stability problems of a GTO in the on-state have been analysed. The simulation results allow extraction of the thermal relaxation time and the value of the total thermal resistance and capacitance for the equivalent thermal circuit model of the device under investigation.

The electrothermal nature of device behavior after onset of latch-up in an IGT has been simulated. It has been found that accurate analysis of latch-up phenomena requires electrothermal simulation to account for the reduction of latching current due to self-heating.

To summarize, a powerful tool to investigate electrothermal problems in power semiconductor devices based on a new model of coupled transport of heat and charge carriers is presented.

6. ACKNOWLEDGEMENT

The work has been supported by SIEMENS Austria, PSE 3. Special thanks to PSE 362 for the provision of computer resources.

7. REFERENCES

[1] BALIGA, B.J. – Modern Power Devices, John Wiley & Sons, New York-Chichester-Brisbane-Toronto-Singapore, 1987.

[2] CALLEN, H.B. – Thermodynamics, John Wiley & Sons, New York-London-Sidney, 1966.

[3] DORKEL, J. M. – On Electrical Transport in Non-Isothermal Semiconductors, Solid State Electronics, Vol.26, No.8, pp. 819-821, 1983.

[4] FRANZ, A.F., FRANZ, G.A., SELBERHERR, S., MARKOWICH, P. – Finite Boxes – A Generalisation of the Finite-Difference Method Suitable for Semiconductor Device Simulation, IEEE Trans. Electron Devices, Vol.30, No.9, pp. 1070-1083, 1983.

[5] GLANSDORF, P., PRIGOGINE, I. – Thermodynamic Theory of Structure, Stability and Fluctuation, John Wiley & Sons, London-New York-Sydney-Toronto, 1971.

[6] MADELUNG, O. – Grundlagen der Halbleiterphysik, Springer, Berlin-Heidelberg-New York, 1970.

[7] NAKAGAWA, A., NAVON, D.H. – A Time- and Temperature-Dependent 2-D Simulation of the GTO Thyristor Turn-Off Process, IEEE Trans. Electron Devices, Vol.31, No.9, pp. 1156-1163, 1984.

[8] PRIGOGINE, I. – Thermodynamics of Irreversible Processes, Interscience, New York-London, 1961.

[9] SELBERHERR, S. – Analysis and Simulation of Semiconductor Devices, Springer, Wien-New York, 1984.

[10] WACHUTKA, G. – Rigorous Thermodynamic Treatment of Heat Generation and Conduction in Semiconductor Device Modeling, IEEE Trans. on CAD, Vol.9, No.11, pp. 1141-1149, 1990.