Abstract

Tunneling generation becomes increasingly important in modern devices both as a source of leakage and for special applications. Mostly, the observed phenomena are attributed to band-to-band tunneling, although from early investigations of Esaki diodes it is well known that at lower field strengths trap-assisted tunneling is responsible for non-ideal IV-characteristics. In this paper we apply microscopic models of trap-assisted and band-to-band tunneling, which were derived from first-principle quantum-mechanical calculations, in a general multi-device simulator. Special simplified versions of the models were developed for the purpose of fast numerical computations. We investigate pn-junctions with different doping profiles to reveal the relative contribution of the two tunneling mechanisms. Simulated currents as function of voltage and temperature are presented for each individual process varying the basic physical parameters. It turns out that the slope of reverse IV-characteristics dominated by trap-assisted tunneling is similar to those which are determined by band-to-band tunneling, if the localized state of the recombination center is only weakly coupled to the lattice. In the model such a slope is produced by field-enhancement factors of the Shockley-Read-Hall lifetimes expressing the probability of tunneling into (or out of) excited states of the electron-phonon system. The temperature dependence of these field-enhancement factors compensates to a certain extent the expected strong temperature effect of the Shockley-Read-Hall process. The latter remains larger than the temperature variation of phonon-assisted band-to-band tunneling, but not as much as often stated. Consequently, the slope of the IV-characteristics and their temperature dependence are not the strong criteria to distinguish between trap-assisted and band-to-band tunneling. The origin of tunnel currents in silicon rather depends on the sum of physical conditions: junction gradient, nature and concentration of defects, temperature and voltage range.

1 Introduction

In recent years tunneling generation has been found to be a source of leakage in advanced silicon devices. Forward bias tunneling in the emitter-base junction of bipolar transistors [1], gate-induced drain and trench-gated diode leakage in DRAM cells [2], gate-induced drain leakage in MOSFETs [3] and device degradation due to tunneling-created carriers [4] are well known examples. Besides, tunneling generation is intentionally used for (band-to-band?) tunneling induced substrate hot-electron injection (BBISHE) in non-volatile memories [5].

In most cases band-to-band tunneling (BBT) has been assumed for the generation process. By measuring the drain current of MOSFETs BBT was inferred from the $\ln[I_D/(V_{DG}-1.2)]$ versus $1/(V_{DG}-1.2)$ dependence [3,4] showing basically a constant slope. However, a similar behaviour is also to be expected from trap-assisted tunneling (TAT). It is not possible to extract a "critical" field strength from the cumbersome field dependence of TAT, but numerical calculations [6,7] gave comparable slopes at least for recombination centers with a more weakly coupled localized state. As another proof whether TAT or BBT is responsible for the above mentioned effects, the temperature dependence of $I-V$ characteristics has been used [3]. It is generally assumed that TAT has a much stronger temperature effect than BBT because of the Shockley-Read-Hall (SRH) statistics and a possible strong temperature dependence of the capture cross sections. However, the field-enhancement factors of SRH lifetimes exhibit a contrary temperature effect leading to a compensation of the T–dependence of the intrinsic density n_i in the tunneling regime [6].

In this paper we study the relative influence of TAT and BBT in tunneling generation by numerical simulation. To do this, microscopic models of TAT [6] (field-enhanced multiphonon recombination) and phonon-assisted BBT [7], outlined in the next section, were implemented into the device simulator **SIMUL** [8].

2 Theory

2.1 Trap-Assisted Tunneling

A simplified model of field-dependent SRH lifetimes was developed in Ref. [6]. It is based on multiphonon capture and emission processes that are enhanced by the presence of a strong electric field. The analytical simplifications required two basic assumptions: first, the number of involved phonons has to remain "large" and second, the field strength must not become too low or too high, respectively. These assumptions are fairly well fulfilled in the pre-breakdown range of silicon junctions at room temperature, since pure (zero phonon) trap tunneling has only a negligible probability there.

In order to keep the number of model parameters as small as possible, further assumptions are necessary. Only one kind of single-level recombination centers is considered with a concentration

$N_t(\boldsymbol{x})$ and (thermal) binding energy E_t measured from the conduction band minimum. The level is assumed to be neutral in its empty state and to couple to a single representative phonon mode with energy $\hbar\omega_0$ only. The constant field approximation (constant over an average tunneling length) allows for a local version of the field-enhanced SRH rate, i.e. the rate is a function of the trap position only and does not involve a numerical integration. A typical total tunneling length in the pre-breakdown range is 100 Å, which also gives the separation of generated or recombining electron-hole pairs in real space, since carriers appear or disappear in the classical picture at their classical turning points. The consequences of this separation will be discussed in the conclusions.

With these assumptions trap-assisted tunneling can be incorporated in the SRH lifetimes τ_n and τ_p in form of field-enhancement factors. The SRH net rate has its usual form, but with field-dependent lifetimes:

$$R_{net}^{SRH} = \frac{n \cdot p - n_i^2}{\hat{\tau}_p(F)[n + n_1] + \hat{\tau}_n(F)[p + p_1]} \tag{1}$$

where

$$\hat{\tau}_\nu(F) = \frac{\tau_\nu}{[1 + g_\nu(F)]} \quad , \; \nu = n,\, p \tag{2}$$

with

$$\begin{aligned} g_n\big(F(\boldsymbol{x})\big) \;=\; & \tfrac{1}{3} \sum_{i=x,\,y,\,z} \left(1 + \frac{\left(\hbar\Theta_{i,\|}\right)^{3/2} \sqrt{E_t - E_0}}{E_0 \hbar\omega_0} \right)^{-\frac{1}{2}} \frac{\left(\hbar\Theta_{i,\|}\right)^{3/4} (E_t - E_0)^{1/4}}{2\sqrt{E_t E_0}} \left(\frac{\hbar\Theta_{i,\|}}{kT} \right)^{\frac{3}{2}} \times \\ & \times \exp\!\left\{ -\frac{E_t - E_0}{\hbar\omega_0} + \frac{\hbar\omega_0 - kT}{2\hbar\omega_0} + \frac{E_t + kT/2}{\hbar\omega_0} \ln(E_t/\epsilon_R) - \frac{E_0}{\hbar\omega_0} \ln(E_0/\epsilon_R) \right\} \exp\!\left(\frac{E_t - E_0}{kT} \right) \exp\!\left[-\tfrac{4}{3} \left(\frac{E_t - E_0}{\hbar\Theta_{i,\|}} \right)^{3/2} \right] . \end{aligned} \tag{3}$$

In Eq. (3) besides the energy level E_t and the effective phonon energy $\hbar\omega_0$ the following characteristic energies were defined:

kT – thermal energy,

$\hbar\Theta_{i,\|} = \left(e^2\hbar^2 F^2 / 2m_{i,\|} \right)^{1/3}$ – electrooptical energy for electrons,

$\epsilon_R = S\,\hbar\omega_0$ – lattice relaxation energy,

$E_0 = E_0(F, kT, \epsilon_R, E_t)$ – optimum transition energy.

The mass $m_{i,\|}$ is the electron tunneling mass in field direction, which differs for the three pairs of equivalent conduction band valleys ($i = x, y, z$)

$$m_{i,\|} = \frac{m_t}{1 - (1 - m_t/m_l)\; F_i^2(\boldsymbol{x}) / F^2(\boldsymbol{x})} \; . \tag{4}$$

The energy E_0 is a field and temperature dependent optimum transition energy which can be given approximately by

$$E_0 = 2\sqrt{\epsilon_F}\left[\sqrt{\epsilon_F + E_t + \epsilon_R} - \sqrt{\epsilon_F}\right] - \epsilon_R \ , \qquad \epsilon_F = \frac{(2\epsilon_R kT)^2}{\left(\hbar\Theta_{i,\|}\right)^3} \ . \tag{5}$$

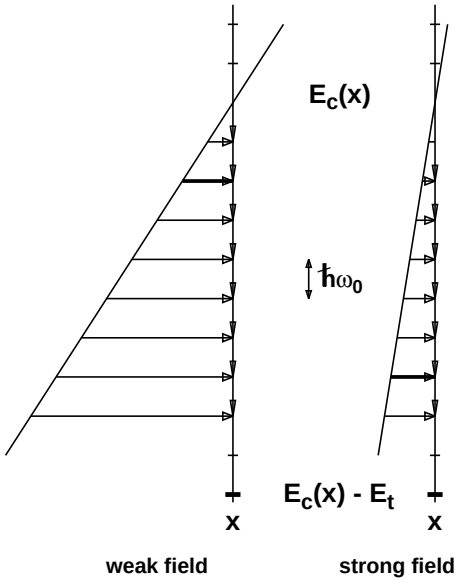

Figure 1: Change of the most probable transition path with electric field strength. $E_c(\boldsymbol{x})$ denotes the conduction band edge at $\boldsymbol{x}$, $E_c(\boldsymbol{x}) - E_t$ is the trap energy level and $\hbar\omega_0$ the effective phonon energy.

Fig. 1 illustrates the change of E_0 as the field strength increases. Formula (3) is the field enhancement factor for electrons, the corresponding expression for holes follows by skipping the sum over the three directions (isotropic valence band assumed) and replacing $m_{i,\|} \rightarrow m_v$ and $E_t \rightarrow E_g - E_t$. The pre-breakdown due to trap-assisted tunneling is dominated by the interplay of the three exponential factors. The first exponent results from the difference of multiphonon transition probabilities with and without electric field, respectively. Here, the low-temperature approximation of multiphonon theory was used. The second exponent describes an increase of occupation probability as the trap depth is effectively lowered by the electric field. The last exponential is the well known tunneling factor for the penetration of a triangular barrier, but with a field dependent tunneling depth $\Delta = E_t - E_0$ measured from the conduction band edge (compare Fig. 1).

According to the low-temperature approximation of multiphonon theory one can derive the following simplified expression for the temperature dependence of (zero field) SRH lifetimes τ_ν

$$\tau_\nu(T) = \tau_\nu(300) \left(\frac{300}{T}\right)^{\frac{3}{2}} \quad . \tag{6}$$

A decrease of lifetime as the temperature rises is consistent with the physical idea that the stronger the lattice vibrations the more efficient carriers can be captured or emitted by the recombination center.

In contrast, measurements of power diodes [9] could be fitted with a T^2–law.

2.2 Band-to-Band Tunneling

Whereas trap-assisted tunneling is a first-order process due to the strong localization of the bound state, indirect band-to-band tunneling requires an additional collision partner for the change in momentum. Since the direct gap of silicon depends only weakly on momentum, both electron-phonon and hole-phonon collisions must be taken into account. Since the early studies of Esaki diodes revealed that TO and TA phonons contribute equally to the phonon-assisted tunneling current in silicon, and other branches are negligible, we use $\hbar\omega_{TA} = 18.6\, meV$ for the phonon energy and multiply by 2 for the total BBT rate finally. Intravalley acoustic scattering was assumed for the electron-phonon matrix element.

For device simulation as in the case of TAT a simplified model was developed ([7]) assuming a constant electric field over the interband tunneling length and WKB approximation for the wave functions.

With these simplifications the rate of band-band tunneling in silicon can be expressed by the formula

$$\begin{aligned} R_{net}^{BBT} &= \frac{3.8\times 10^{20}}{cm^3 s} F^{7/2} \sum_{\alpha=x,y,z} \frac{\sqrt{\tilde{m}_{\perp 2}^{\alpha} m_{\parallel}^{\alpha}}}{m} \left[\frac{(F_c^{\alpha\mp})^{-\frac{3}{2}} \exp\left(-\frac{F_c^{\alpha\mp}}{F}\right)}{\exp\left(\frac{\hbar\omega_{TA}}{kT}\right)-1} + \frac{(F_c^{\alpha\pm})^{-\frac{3}{2}} \exp\left(-\frac{F_c^{\alpha\pm}}{F}\right)}{1-\exp\left(-\frac{\hbar\omega_{TA}}{kT}\right)} \right] \times \\ &\quad \times \left[f_v(0) - f_c(\boldsymbol{k}_0)\right] \ , \end{aligned} \tag{7}$$

with "critical" field strengths

$$F_c^{\alpha\pm} = \frac{4}{3} \frac{\sqrt{2\mu_{\parallel}^{\alpha}}}{e\hbar} \left(E_g^i \pm \hbar\omega_{TA}\right)^{3/2} \quad . \tag{8}$$

The effective masses are given by

$$m_{\parallel}^{\alpha} = \frac{m_t m_l}{m_l - (m_l - m_t) F_\alpha^2/F^2} \ , \qquad \tilde{m}_{\perp 2}^{\alpha} = \frac{m_t m_l}{m_t - (m_t - m_l) F_\alpha^2/F^2} \ , \tag{9}$$

where $1/\mu_{\|}^{\alpha} = 1/m_{\|}^{\alpha} + 1/m_v$ is the reduced effective mass of direction α (interband tunneling mass). For the hole mass m_v we took the value of light holes, although this quantity is uncertain because the strong electric field will remove the band degeneracy at $\boldsymbol{k} = 0$ and therefore deform the dispersion relation. In Eq. (7) the upper sign refers to tunneling generation (reverse bias, $f_v > f_c$), but the lower to recombination (forward bias, $f_v < f_c$), E_g^i denotes the indirect gap of silicon, and the field strength F is measured in V/cm.

Complete locality of the BBT rate results from the assumption that the overlap of the envelope wave functions near the middle of the gap (see Fig. 2) determines a correlation between position $\boldsymbol{x}$

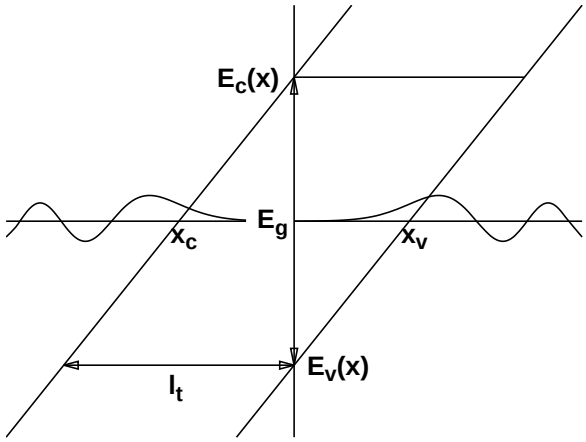

Figure 2: Tunneling length $l_t = |x_v - x_c|$ and band diagram, assuming a constant electric field over l_t. The maximum overlap of the envelope wave functions determines the transition energy at a given position $\boldsymbol{x}$ in the device.

and transition energy $E_0(\boldsymbol{x})$. This energy is approximately given by

$$E_0(x) = E_c(x) - \frac{\mu_{\|}^{\alpha}}{m_{\|}^{\alpha}} E_g^i = E_v(x) + \frac{\mu_{\|}^{\alpha}}{m_v} E_g^i \; . \tag{10}$$

The distribution functions in Eq. (7) then take the form

$$f_{c,v} = \left[\exp\left(\frac{E_0(\boldsymbol{x}) - E_{Fn,p}(\boldsymbol{x})}{kT} \right) + 1 \right]^{-1} \; , \tag{11}$$

i.e. the occupation is calculated at the point of maximum overlap of initial and final states, which coincides with the mesh point per definition of $E_0(\boldsymbol{x})$. The choice of $\boldsymbol{x}$ for the evaluation of the distribution functions is of course ambiguous to the same extent as the incorporation of a pure quantum effect into a classical drift-diffusion simulator.

3 Numerical Simulation

The relative influence of TAT and BBT in tunneling generation has been studied by numerical simulation. The microscopic models of TAT [6] (field-enhanced multiphonon recombination) and phonon-assisted BBT [7] were implemented into the device simulator **SIMUL** [8]. The investigated structure is a $n^+ - p$ diode with an error function profile and bulk concentrations of $10^{20}\,cm^{-3}$ for N_D and $10^{17} - 10^{19}\,cm^{-3}$ for N_A (junction depth $= 140\,nm$ at $N_A = 3 \times 10^{17}\,cm^{-3}$ and zero voltage). Such a structure is relevant for advanced bipolar devices and the MOS-gated diode. The conclusions are less applicable to BBISHE where confined states in the channel and localized interface states modify the tunneling probability (not modeled in **SIMUL**). In order to exclude artificial electron-hole pair generation directly at the oxide-silicon interface where no final conduction band states are available, a thin sheet with a thickness of half the BBT length does not contribute to the BBT rate in **SIMUL**. In this way a time-consuming searching procedure is avoided. Similarly, a corresponding thin sheet is excluded from the TAT domain. For the simulated structure in the present paper this problem is not relevant.

In addition, models of the following physical effects were used, all described in [8]: temperature-dependent bandgap, bandgap narrowing, doping and field dependent mobilities, avalanche generation (local field version), Auger recombination (band-to-band and trap-assisted), and doping dependence of SRH lifetimes.

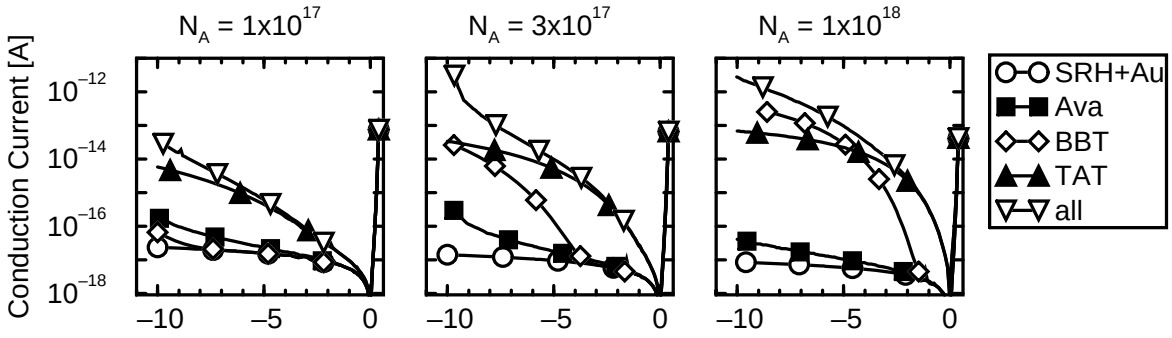

Figure 3: Reverse I–V characteristics for the individual generation–recombination processes and the resulting curve ('all') for $1\mu m \times 1\mu m$ diodes. The doping dependence of the SRH recombination was suppressed in order to obtain a lower limit for TAT. The SRH net recombination is included in every curve. Parameters: ϵ_R=240meV and T=300K.

Fig. 3 shows reverse $I - V$ characteristics of the individual generation-recombination processes ("SRH+Au" - Auger and SRH recombination without field enhancement, "Ava" - Avalanche generation) and the resulting curve (all) for a $1\,\mu m \times 1\,\mu m$ diode and N_A-concentrations ranging from $10^{17}\,cm^{-3}$ to $10^{18}\,cm^{-3}$. Here the doping dependence of the SRH lifetimes (Scharfetter relation)

was suppressed in order to obtain a lower limit for TAT. A midgap level and $0.240\,eV$ for the lattice relaxation energy ϵ_R of the recombination center were used. Such a value of the lattice relaxation energy is typical for defects in silicon, e.g. the gold impurity and the A-center [10]. The influence of BBT increases as the junctions become steeper, but TAT dominates the pre-breakdown up to $-4\,V$. Impact ionization remains unimportant because the junctions are too narrow.

Effects in the forward characteristic are only observed if the p- concentration is further increased (Fig. 4). In the case $N_A = 10^{19}\,cm^{-3}$ the forward bias tunneling is still merely determined by TAT.

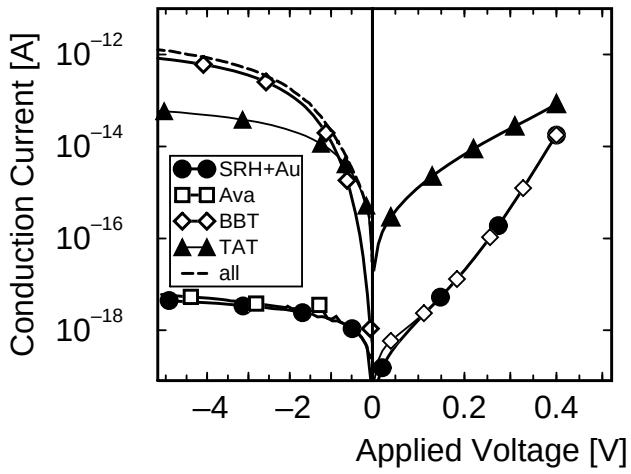

Figure 4: I-V characteristics of a p–n junction with $N_A = 10^{19}\ cm^{-3}$ and $N_D = 10^{20}\ cm^{-3}$ (T=300K, ϵ_R=240 meV). The forward characteristics show the onset of the characteristic bump for band–to–band tunneling and trap-assisted tunneling. The latter is the dominant part.

To obtain a possible lower limit of TAT, the lattice relaxation energy ϵ_R was varied in Fig. 5 from $0.4\,eV$ (strong coupling) to $0.05\,eV$ (weak coupling). Strongly coupled centers exhibit the lowest TAT, but the pre-breakdown remains in any case dominated by TAT up to the onset of BBT. For still stronger coupling than $\epsilon_R = 0.4\,eV$ the $I-V$ curve tends to the "reference" plateau, which is determined by SRH and Auger recombination. However, extremely strong coupling is not covered by the underlying multiphonon theory. Since real devices contain different kinds of defects, the latter case is not very likely to dominate the $I-V$ characteristics. The slope increases with decreasing coupling strength and becomes similar to that of BBT for $\epsilon_R \approx 0.1\,eV$. If one assumes the stronger coupled recombination centers to determine the zero field SRH lifetimes, and if one further assumes that other centers with a smaller coupling strength (and therefore with larger SRH lifetimes and a steeper slope) come into play at higher voltages, the TAT characteristic becomes a superposition of these curves. Therefore, from a $\ln\left[I_D/(V-1.2)\right]$ versus $1/(V-1.2)$ dependence

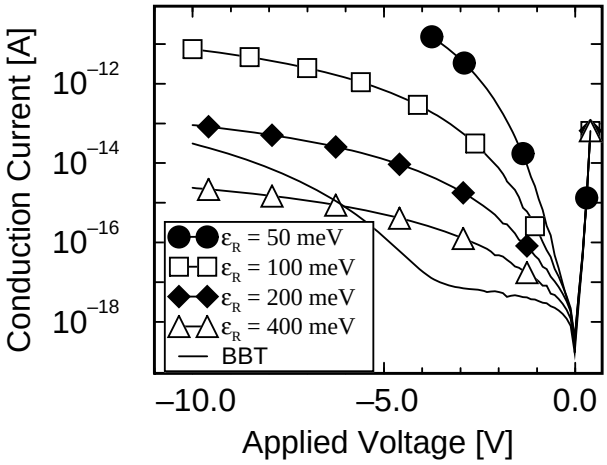

Figure 5: To obtain a possible lower limit of TAT the lattice relaxation energy ϵ_R was varied from 0.4 eV (strong coupling) to 0.05 eV (weak coupling). Parameters: $N_A = 3 \cdot 10^{17}\ cm^{-3}$, $N_D = 1 \cdot 10^{20}\ cm^{-3}$, and $T = 300K$.

one cannot unambiguously distinguish between BBT and TAT.

Fig. 6 shows the simulated temperature dependence of BBT and TAT together with SRH+Au. Before tunneling sets in, the difference is mainly due to n_i. In the BBT regime (Fig. 6a) higher temperatures increase the current because of the shrinked bandgap and the larger phonon occupation. In the case of TAT (Fig. 6b) the field enhancement factors increase the lifetimes with rising temperature and compensate the strong temperature effect from n_i – the more the lower the temperature. At higher field strengths the temperature difference of $I - V$ curves in the TAT regime is largly reduced and comparable with that of the BBT curves after the onset of BBT, i.e. for reverse biases larger than 3 V. For the zero field lifetimes a T^2-law was used. Thus, even the temperature dependence is not the strong criterion to decide wether BBT or TAT dominates the $I - V$ characteristics.

4 Discussion

Our results indicate that the field and temperature dependences of TAT may be similar to those of BBT. The low-voltage range of $I - V$ characteristics of $n^+ - p$ junctions is determined by TAT. If a model for the doping dependence of SRH lifetimes is used in the simulation, the dominance of TAT in the pre-breakdown range further increases. BBT strongly depends on the steepness of the junctions, i.e. requires high doping levels on both p and n side. The simulations show that it becomes dominant over TAT at a certain bias, which decreases as the N_A-concentration increases.

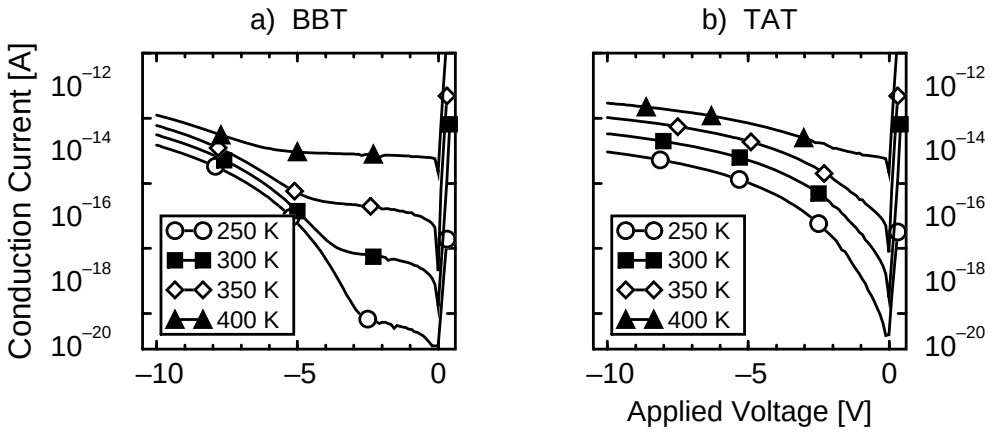

Figure 6:
a)
Band–to–band tunneling at different temperatures with $N_A = 3 \cdot 10^{17}\ cm^{-3}$, and ϵ_R=240 meV. In the BBT regime higher temperatures increase the current because of bandgap narrowing and the larger phonon occupation.
b)
Trap–assisted tunneling at different temperatures with $N_A = 3 \cdot 10^{17}\ cm^{-3}$, and ϵ_R=240 meV. The temperature dependence of the field enhancement factors compensates the strong temperature effect from n_i.

(In Esaki diodes the reverse bias branch is fully determined by BBT.)

The implemented local tunneling models are economical, since no search for turning points and no numerical integrations are necessary. If the doping gradients are still steeper, both the local generation models and the drift-diffusion approach become wrong. Simulating very narrow junctions, we observed the following effect: If the field is high enough, at the point of maximum field strength trap-assisted tunneling can generate such a large number of electron-hole pairs, that the np-product reaches the value of n_i^2, and consequently, the rate gets a pronounced minimum at this point. In the selfconsistent solution of the transport equations this leads to a saturation of the current untill impact ionization starts to dominate. Since the spatial separation of generated "carriers" (i.e. objects having a drift velocity) is not taken into account, an artificially high carrier density at the field peak falsifies the results. Furthermore, the transport of generated carriers out of the narrow junction is rather ballistic than a drift process.

References

[1] J. del Alamo and R. M. Swanson, "Forward-Bias Tunneling: A Limitation to Bipolar Device Scaling", IEEE Electron Device Letters **7**, no. 11, pp. 629–631, Nov. 1986

[2] S. H. Voldman, J. B. Johnson, T. D. Linton, and S. L. Titcomb, "Unified Generation Model with Donor- and Acceptor Type Trap States for Heavily Doped Silicon", IEDM Tech. Digest, pp. 349–352, Dec. 1990

[3] T. Y. Chan, J. Chen, P. K. Ko, and C. Hu, "The Impact of Gate-Induced Drain Leakage Current on MOSFET Scaling", IEDM Tech. Digest, pp. 718–721, Dec. 1987

[4] Y. Igura, H. Matsuoka, and E. Takeda, "New Device Degradation Due to "Cold" Carriers Created by Band-to-Band Tunneling", IEEE Electron Device Letters **10**, no. 7, pp. 227–229, May 1989

[5] I.-C. Chen, D. J. Coleman, and C. W. Teng, "Gate Current Injection Initiated by Electron Band-to-Band Tunneling in MOS Devices", IEEE Electron Device Letters **10**, no. 7, pp. 297–300, July 1989

[6] A. Schenk, "A Model for the Field and Temperature Dependence of Shockley-Read-Hall Lifetimes in Silicon", Solid-State Electronics **35**, no. 11, pp. 1585–1596, 1992

[7] A. Schenk, "Rigorous Theory and Simplified Model of the Band-to-Band Tunneling in Silicon", Solid-State Electronics **36**, no. 1, pp. 19–34, 1993

[8] **SIMUL** Manual, IIS, ETH Zürich, Switzerland, 1.0 (alpha) edition, 1992

[9] H. Schlangenotto and D. Silber, "High Current Behavior of pin-Rectifiers and Thyristors with Different p-Emitter Structures", ESSDERC München, Sep. 1973

[10] A. Schenk, K. Irmscher, D. Suisky, R. Enderlein, F. Bechstedt, and H. Klose, "Electric Field Effect on Multiphonon Transitions at Deep Centres", Proceedings of the 17th International Conference on the Physics of Semiconductors, p. 613, Springer-Verlag New York, 1985