Revised Shockley–Read–Hall lifetimes for quantum transport modeling

Timm Höhr,[a] Andreas Schenk, and Wolfgang Fichtner

Swiss Federal Institute of Technology, Integrated Systems Laboratory, Gloriastrasse 35, CH-8092 Zürich, Switzerland

The inclusion of quantization effects on the carrier densities is now the state of the art in modern semiconductor device simulators and yields, for example, quantum-corrected threshold voltages and quantum-mechanical models of the channel mobility. However, the effect of charge quantization on nonradiative thermal generation–recombination has not received much attention. In this article, Shockley–Read–Hall recombination is examined for situations in which electrons and/or holes are confined in semiconductor devices. For the transitions between band states and a single deep level, a previously developed multiphonon description is adopted. It is found that the lifetimes have to be altered due to the same quantized local density of states that also accounts for the carrier distribution. Numerical evaluation of this model for one-dimensional potentials and small phonon energies results in spatially varying lifetime profiles that exhibit two opposite regimes. The additional nonclassical offset of the subband eigenenergies causes an increased lifetime in the limit of strong quantum confinement. For nondegenerate statistics, an analytical high-temperature approximation is presented for this limit, where the activation energy of the lifetime is increased by the lowest-subband offset. In the absence of confinement, however, high electric fields reduce the lifetime due to carrier tunneling into the bandgap. © 2004 American Institute of Physics. [DOI: 10.1063/1.1687992]

I. INTRODUCTION

The inclusion of quantum effects is crucial for the technology computer-aided design of modern deep-submicron devices. Corresponding enhancements of usual drift-diffusion device simulators are widely used by applying Schrödinger solvers or density gradient models.[1–3] These methods provide good results for the quantum-mechanical (QM) density profiles. However, in applying those methods, the question arises of how other quantities—namely, transport parameters such as mobility and generation–recombination lifetimes—can be modeled in a consistent way.

Two major types of nonradiative recombination must be considered: recombination by Auger emission and via deep trap levels in the bandgap. New mechanisms of Auger recombination in quantum wells were theoretically identified by Zegrya and co-workers.[4]

With regard to device simulation, this work presents an example for including the QM eigenstates in a model for the Shockley–Read–Hall (SRH) recombination.[5] For one-dimensional confinement, the energetic separation between the electronic states in conduction and valence bands and the deep trap levels increases due to subband formation. Therefore, in addition to an altered density, it is reasonable to also expect a change of the SRH lifetime in the presence of quantization.

The model adopted here describes capture and emission of electrons between the trap and band states as multiphonon processes, which transfer energy between the electronic and the vibrational system of the crystal. The corresponding lifetimes receive spatial dependence through a local density of states (DOS) that is composed of the eigenstates of the confining potential. This ansatz has also been used in order to describe enhanced recombination due to tunneling as a function of the local electric field.[6,7] This was done by accounting for tunnel-assisted transitions, but using densities as in classical device simulation (without quantum correction).

The article is organized as follows: in Sec. II, the model is introduced, Sec. III presents numerically obtained lifetime profiles for a simple triangular potential, and in Sec. IV, an analytical approximation is derived for the limit of strong quantization. In Sec. V, the model and the approximation are applied to quantum states resulting from simulated one-dimensional devices.

II. MODEL FOR THE SRH LIFETIME

A. Rate formula

SRH recombination occurs via deep trap levels in the energy gap. In the following, we assume that all traps are identical in nature and that they have a single level with thermal binding energy E_t measured from the local conduction band edge $E_c(z)$. Furthermore, the system is assumed to be homogeneous in the $x-y$ plane. Under stationary conditions, the net recombination rate is given by

$$R = \frac{np - n_1 p_1}{\tau_n(p + p_1) + \tau_p(n + n_1)}, \tag{1}$$

where n and p are, respectively, the electron and hole densities.[5] The quantities $n_1 = n(1 - f_t^n)/f_t^n$ and $p_1 = p f_t^p/(1$

[a]Electronic mail: hoehr@iis.ee.ethz.ch

       4875

$-f_t^p$) have their usual meaning. The lifetimes $\tau_n = n/(N_t \hat{c}_n)$ and $\tau_p = n/(N_t \hat{c}_p)$ contain the average capture coefficients

$$\hat{c}_n = \int dE\, N_c(E) c_n(E) f_c(E) \qquad (2)$$

and

$$\hat{c}_p = \int dE\, N_v(E) c_p(E) [1 - f_v(E)], \qquad (3)$$

where $N_{c,v}$ denote the DOS of conduction and valence band, respectively. The spectral capture rates (in units of cm³/s) $c_{n,p}(E)$ characterize the transitions from a certain energy level E in the conduction or valence band to the deep trap level in the energy gap. Different quasi-Fermi levels, E_F^n and E_F^p, are used for both carrier species and their respective distribution functions:

$$f_{c,v}(E) = \left[g \exp\left(\frac{E - E_F^{n,p}}{kT} \right) + 1 \right]^{-1}, \qquad (4)$$

with $g = 1$. The trap occupation probabilities $f_t^{n,p}$ are obtained by replacing E with the trap energy $E_c(z) - E_t$, and g with the ratio of the degeneracy factors of the empty and the occupied trap level. Expressions for the emission coefficients $\hat{e}_{n,p}$ are similar to Eqs. (2) and (3). They are related to the capture coefficients by $\hat{e}_n = \hat{c}_n(1 - f_t^n)/f_t^n$ and $\hat{e}_p = \hat{c}_p f_t^p/(1 - f_t^p)$.

B. Multiphonon capture rate

The capture and emission processes are modeled according to the theory of multiphonon emission and absorption.[8] See Refs. 6, 7 and references therein for the development of this theory. Here, the trap states are assumed to be strongly localized and, therefore, to relate only to the charge densities at the same point in space. Furthermore, only a single phonon mode of frequency ω_0 is assumed to interact with the electron. The spectral capture rates thus have the following form:

$$c_{n,p}(E) = \overline{c_{n,p}^0} \sum_{l \geqslant 0} \frac{(l \mp S)^2}{S} L(l)\, \delta(l \hbar \omega_0 \pm E_c \mp E_t \mp E), \qquad (5)$$

where the lower signs apply to c_p.[6,7] The function

$$L(l) = e^{-S(2f_B + 1)} \left(\frac{f_B + 1}{f_B} \right)^{l/2} I_l[2S\sqrt{f_B(f_B + 1)}] \qquad (6)$$

contains the modified Bessel function I_l of order l and the Huang–Rhys factor S defining the lattice relaxation energy $\epsilon_R = S\hbar\omega_0$. Further ingredients are the Bose–Einstein occupation probability for the phonon mode with energy $\hbar\omega_0$, $f_B = [\exp(\hbar\omega_0/kT) - 1]^{-1}$ and the energetic separation E_t of the trap levels in the bandgap from the local conduction band edge $E_c(z)$. The factor $(l \mp S)^2/S$ is replaced by unity to avoid the artificial disappearance of the probability of thermally induced transitions for $l^* = S$. As discussed in detail in Ref. 7, this artifact is related to the violation of first-order perturbation theory when, in a configuration-coordinate diagram, the lower potential parabola (bound state) crosses the upper parabola (band state) at its minimum, leading to a

completely anharmonic lattice potential around this crossing point.[9] It should be noted that the factor does not appear in a two-phonon model with accepting and promoting modes.[10]

We assume that the parameters of the recombination center (E_t, S, and ϵ_R) are not changed by the confining potential and hence will not become position dependent. Due to the assumption of a δ-like trap potential, the influence of the quantum confinement on binding energy and wave function of the center will be small as long as its distance to the interface remains larger than its localization radius. However, stronger deviations with respect to the bulk values must be expected in the case of charged centers with a long-range part of the potential. In addition, alterations of the phonon system due to confinement are ignored as well.

From the spectral capture rate one obtains the lifetimes

$$\tau_n(z)^{-1} = \frac{N_t \overline{c_n^0}}{n(z)} \sum_{l \geqslant 0} L(l) N_c(E_l, z) f_c(E_l) \qquad (7)$$

and

$$\tau_p(z)^{-1} = \frac{N_t \overline{c_p^0}}{p(z)} \sum_{l \geqslant 0} L(l) N_v(E_l, z)[1 - f_v(E_l)], \qquad (8)$$

with

$$E_l = \begin{cases} E_c - E_t + l\hbar\omega_0, & \text{for electrons} \\ E_c - E_t - l\hbar\omega_0, & \text{for holes.} \end{cases} \qquad (9)$$

From now on, all considerations are restricted to electrons. Holes can be treated analogously.

C. Density of states for quantization in one dimension

The DOS entering $\hat{c}_n$ is the same as in the densities n and n_1. If n is the QM density, then n_1, $\hat{c}_n$ and the lifetime τ_n are also affected by quantization.

In the following, quantization is considered for motion along the z axis. This leads to a description in terms of a local DOS which is further detailed in the appendix. Here, only the resulting DOS expressions are given. The band structure is assumed to be parabolic and, for the moment, only one valley is considered. Without confinement and for vanishing field, the envelope wave function consists of plane waves in all three dimensions, and the usual bulk DOS applies:

$$N_c(E, z) = \frac{1}{2\pi^2} \left(\frac{2m^*}{\hbar^2} \right)^{3/2} \sqrt{E - E_c(z)}\, \Theta[E - E_c(z)], \qquad (10)$$

with $m^* = (m_x m_y m_z)^{1/3}$ and m_i for $i = x$, y, z being the effective mass tensor components along the principal axes. These are chosen to coincide with the Cartesian coordinate axes for the sake of simplicity.

For a homogeneous electric field F in the z direction (only one classical turning point), Eq. (10) is replaced by

$$N_c(E, z) = \left(\frac{2m^*}{\hbar^2} \right)^{3/2} \frac{\sqrt{\hbar\theta_z}}{2\pi} \mathcal{F}\left(\frac{E_c(z) - E}{\hbar\theta_z} \right), \qquad (11)$$

with $\theta_z = [e^2 F^2/(2 m_z \hbar)]^{1/3}$, $E_c(z) = E_c(0) + eFz$, and the electro-optical function $\mathcal{F}(x) := [\text{Ai}'(x)]^2 - x \, \text{Ai}(x)^2$.[11] This expression contains a nonvanishing DOS contribution decaying into the energy gap, as described for the first time by Franz and Keldysh for optical absorption in semiconductors in a field.[12,13]

Secondly, consider that in the z direction, the particle is confined from both sides (two classical turning points). This leads to the following DOS [see Eq. (A6)]:

$$N_c(E,z) = \frac{m_{xy}}{\pi \hbar^2} \sum_i |\varphi_i(z)|^2 \Theta(E - E_i), \qquad (12)$$

where $\varphi_i(z)$ denotes the envelope wave function of the ith eigenstate, E_i its energy, and $m_{xy} = \sqrt{m_x m_y}$. Using this DOS, the QM density profile is obtained as follows:

$$n(z) = \int dE \; N_c(E,z) f_c(E)$$
$$= \frac{kT m_{xy}}{\pi \hbar^2} \sum_i |\varphi_i(z)|^2 \mathsf{F}_0 \left(\frac{E_{\mathrm{F}}^n - E_i}{kT} \right), \qquad (13)$$

where $\mathsf{F}_0(x) = \ln(1 + \exp(x))$ denotes the Fermi integral of order 0.

In the following, the different DOS expressions, Eq. (11) for continuous eigenspectrum in a constant field and Eq. (12) for a discrete spectrum, will be referred to as Franz–Keldysh and quantum-confined DOS (FKDOS and QCDOS), respectively.

D. Electron lifetime profiles

For a multivalley band with anisotropic, parabolic valleys (labeled by the additional index ν) the combination of Eqs. (7), (12), and (13) yields the electron lifetime

$$\tau_n(z)^{-1} = \frac{N_t \overline{c_n^0} \sum_\nu m_{xy}^\nu \sum_i |\varphi_{\nu i}(z)|^2 \sum_{l \geq l_0^{\nu i}(z)} L(l) f_c(E_l)}{kT \sum_\nu m_{xy}^\nu \sum_i |\varphi_{\nu i}(z)|^2 \mathsf{F}_0 \left(\dfrac{E_{\mathrm{F}}^n - E_{\nu i}}{kT} \right)}, \qquad (14)$$

where

$$l_0^{\nu i}(z) = \max \left\{ \frac{E_{\nu i} - [E_c(z) - E_t]}{\hbar \omega_0}, \; 0 \right\}$$

is the minimal non-negative number of phonons necessary to reach the subband at $E_{\nu i}$ from the local trap energy given by $E_c(z) - E_t$.

This expression will be further explored in two ways. In Sec. III, lifetime profiles are numerically calculated for a triangular potential in order to illustrate the effects of confinement as well as of an electric field. Secondly, the case of strong confinement (i.e., increased separation between the electronic states in the band and the trap level) is treated analytically in Sec. IV. The results will often be presented with respect to the classical lifetime [obtained from Eqs. (7) and (10)]:
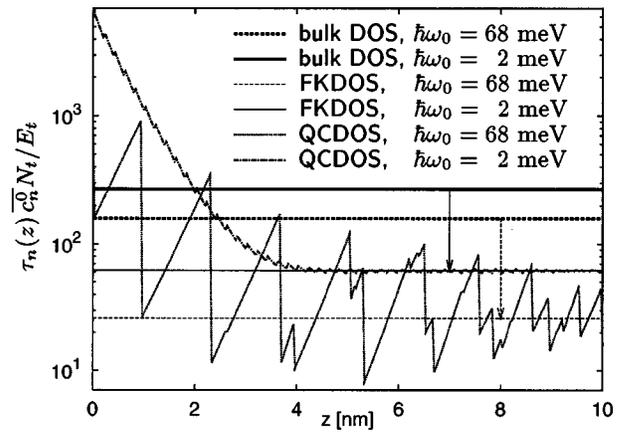
FIG. 1. Comparison of the lifetimes for a triangular well with a field strength $F = 5 \times 10^5$ V/cm for different DOS models and two different phonon energies $\hbar \omega_0$. Thick horizontal lines show the classical lifetime without field enhancement or quantum effects. Thin horizontal lines show the results of using the FKDOS. The arrows highlight the lifetime reduction due to the field. The remaining lines indicate the results for the QCDOS including all relevant eigenstates.

$$\frac{1}{\tau_{n,cl}} = \frac{2 \overline{c_n^0} N_t \sum_{l \geq E_t/\hbar \omega_0} L(l) \sqrt{l \hbar \omega_0 - E_t} f_c(E_l)}{\sqrt{\pi} \, (kT)^{3/2} \mathsf{F}_{1/2} \left(\dfrac{E_{\mathrm{F}}^n - E_c}{kT} \right)}, \qquad (15)$$

where $\mathsf{F}_{1/2}(\epsilon) = (2/\sqrt{\pi}) \int_0^\infty dx \sqrt{x} [1 + \exp(x - \epsilon)]^{-1}$ is the Fermi integral of order 1/2.

III. LIFETIME PROFILES FOR A TRIANGULAR WELL

The quasi-Fermi energy E_{F}^n is assumed to be constant throughout the system, although this will lead to unrealistically large separations from the conduction band edge for large distances z and high fields. The barrier at $z = 0$ is set to infinite height. The corresponding eigenstates are given in the appendix [Eq. (A7)]. Expression (7) was evaluated numerically for the following DOS types: bulk (10), FKDOS (11), and QCDOS (12), the latter producing (14). Here, a single valley semiconductor with isotropic effective mass equal to the longitudinal effective mass component m_l of silicon was considered. The parameters describing the trap state are taken from the gold acceptor level in silicon as used in Ref. 7, $E_t = 0.55$ eV, $\epsilon_R = 0.238$ eV and $\hbar \omega_0 = 68$ meV. The former are kept fixed as the phonon energy $\hbar \omega_0$ is varied. The temperature is always $T = 300$ K. The resulting lifetime profiles are shown in Fig. 1.

For each subband, the minimum phonon number necessary to reach the trap level changes by one at certain distances from the wall, which causes a sudden decrease of the lifetime. The amplitude of these spikes is reduced by choosing a smaller phonon energy. At about 5 nm distance, the lifetime reaches the FK result, which itself is reduced with respect to the bulk value (arrows in Fig. 1). This reduction is caused by field-enhanced tunneling into the bandgap. In contrast, the steep rise near the wall is caused by the additional separation of the lowest subband from the trap level due to quantization.
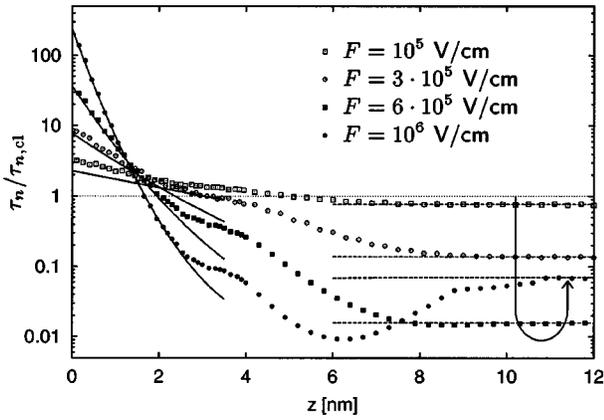
FIG. 2. Lifetime profiles for different field strengths and $\hbar\omega_0 = 2$ meV (symbols). The FK lifetime for nondegenerate statistics is given by dashed lines. The arrow indicates the direction of increasing field strength. Solid lines show an analytical approximation for the contribution of the lowest subband, $1/g_{QM}(z)$, as defined by formula (33).

In the following, the small phonon energy is retained, but the sixfold valley band structure of silicon is used. The coordinate axes are chosen to coincide with the [100]-directions. The following values were used for the effective mass components: longitudinal mass $m_l = 0.9163\,m_0$ and transverse mass $m_t = 0.1982\,m_0$, where m_0 is the electron rest mass.

Slope and magnitude of the lifetime profile in the part near the wall clearly depend on the field (Fig. 2). The curves are more complicated because two sets of nonequivalent valleys exist, which have different quantization masses, m_l and m_t (the latter are labeled with an additional prime on the corresponding subband index). The individual subband contributions segregate spatially for very high fields (Fig. 3, for $F = 10^6$ V/cm) as the energetic separation of the subband levels increases. Near the wall ($z < 3$ nm), the inverse lifetime is governed by the lowest subband of the unprimed set. The absolute maximum is determined by the lowest primed sub-
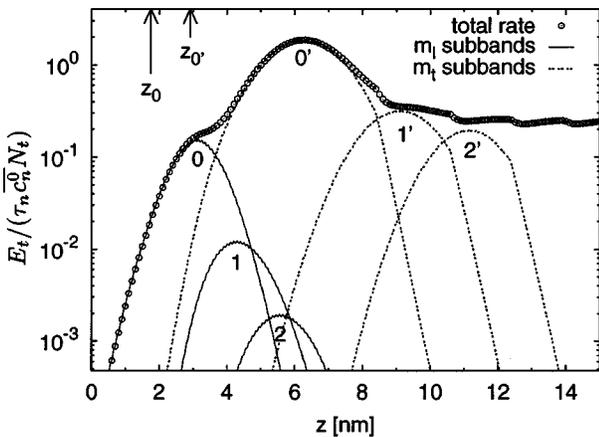
FIG. 3. Contributions of individual subbands to the inverse lifetime profile (circles) as they appear as summands in the numerator of Eq. (14). Only the respective first three of each set are shown. Primed subband indices correspond to the smaller transverse mass m_t (dashed lines), unprimed ones to the larger longitudinal mass m_l (solid lines). Parameters are $F = 10^6$ V/cm and $\hbar\omega_0 = 2$ meV. The arrows labeled z_0 and $z_{0'}$ indicate the classical turning points of the lowest subband in each set.

band because of a larger penetration depth of its wave function into the bandgap. This maximum, which corresponds to the subband energy $E_{0'}$ (see Fig. 3), is found more than 3 nm outside of the classically allowed region $z \in [0, z_{0'}]$. Accordingly, the separation between $E_{0'}$ and the trap level at this position is less than half of the separation in the classical treatment (which is E_t everywhere).

Farther from the wall, the primed subbands continue to dominate. They determine the constant lifetime in the distant part of the profile, which coincides with the corresponding FK result for the given field (dashed lines in Fig. 2). This field enhancement reaches a maximum around 7×10^5 V/cm for the chosen phonon energy and is reduced for higher fields (arrow in Fig. 2). This can be explained as follows: With increasing field strength, the DOS extends into the bandgap and consequently the density increases. The sum in the numerator of Eq. (14), however, is cut off at $l = 0$ (which corresponds to trap-assisted tunneling without phonon assistance). This means that tunneling cannot contribute for energies below the local trap level. Thus, in exceeding a certain field strength, the increase of the density prevails and reverses the trend of a decreasing lifetime.[14]

IV. ANALYTICAL APPROXIMATION FOR STRONG QUANTUM CONFINEMENT

The contribution of a certain subband i at a specified position z is the stronger the closer it is to the trap energy level and the larger $|\varphi_i(z)|^2$ is. All but the lowest subband can be neglected if the confinement is sufficiently strong to provide a high subband separation ($\gg kT$) and if one stays inside or close to the classically allowed region (in Figs. 2 and 3, this corresponds to $z \lesssim 2$ nm). In this case, the probability densities in Eq. (14) cancel. Assuming Boltzmann statistics and defining $\zeta = 2S\sqrt{f_B(f_B+1)}$, one obtains

$$\tau_n^{-1} = \frac{\overline{c_n^0}N_t e^{-S(2f_B+1)}}{kT \exp\left(\dfrac{E_F^n - E_0}{kT}\right)} \sum_{l \geqslant (E_0 - E_c(z) + E_t)/\hbar\omega_0} I_l(\zeta)$$

$$\times \exp\left(\frac{E_F^n - E_c(z) + E_t - l\hbar\omega_0}{kT} + \frac{l\hbar\omega_0}{2kT}\right). \quad (16)$$

Following Ref. 7, the summation over l is approximated by an integral assuming $E_t \gg \hbar\omega_0$:

$$\tau_n^{-1} = \frac{A e^{\Delta(z)/kT}}{\hbar\omega_0} \int_{\Delta(z)} dE\, e^{-E/2kT} I_{E/\hbar\omega_0}(\zeta), \quad (17)$$

where $A = \overline{c_n^0} N_t e^{-S(2f_B+1)}/kT$ and $\Delta(z) = E_0 - E_c(z) + E_t$ denotes the energetic distance between the local trap level and the lowest subband energy. In this approximation, $l = E/\hbar\omega_0$ can be regarded as large; hence, the modified Bessel function is replaced by its asymptotic form for large-order l:

$$I_l(z) \rightarrow \frac{\exp[\sqrt{l^2 + z^2} - l\ln(l/z + \sqrt{1 + l^2/z^2})]}{\sqrt{2\pi\sqrt{l^2 + z^2}}}, \quad (18)$$

yielding

$$\tau_n^{-1} = \frac{A e^{\Delta(z)/kT}}{\hbar \omega_0 \sqrt{2\pi}} \int_{\Delta(z)} dE \, W(E), \tag{19}$$

with the thermal weight function[7]

$$W(E) = \left(\frac{E^2}{(\hbar \omega_0)^2} + \zeta^2 \right)^{-1/4} \exp \left[\sqrt{\frac{E^2}{(\hbar \omega_0)^2} + \zeta^2} \right.$$
$$\left. - \frac{E}{\hbar \omega_0} \ln \left(\frac{E}{\hbar \omega_0 \zeta} + \sqrt{1 + \frac{E^2}{(\hbar \omega_0 \zeta)^2}} \right) - \frac{E}{2kT} \right]. \tag{20}$$

Assuming high temperature $kT \gg \hbar \omega_0$, one may use $\zeta \gg E/\hbar \omega_0$ to obtain

$$W(E) \approx \frac{\hbar \omega_0}{\sqrt{2akT}} \exp \left(\zeta + \frac{a}{4kT} - \frac{(E+a)^2}{4akT} \right), \tag{21}$$

where $a = (\hbar \omega_0)^2 \zeta/(2kT) = (\epsilon_R \hbar \omega_0/kT) \sqrt{f_B(f_B+1)}$. The integral over the Gaussian expression in $W(E)$ produces the complementary error function in the result

$$\tau_n^{-1}(z) = \frac{\overline{c_n^0} N_t}{2kT} C \exp \left(\frac{\Delta(z)}{kT} \right) \mathrm{erfc} \left(\frac{\Delta(z)+a}{\sqrt{4akT}} \right), \tag{22}$$

where

$$C = \exp \left(- \frac{\epsilon_R(2f_B+1)}{\hbar \omega_0} + \zeta + \frac{a}{4kT} \right). \tag{23}$$

Now, a is replaced by its high-temperature value $a \approx \epsilon_R$, except in the factor C, because compensating terms can produce large errors. Applying the asymptotic behavior of the complementary error function for large argument [$\mathrm{erfc}(x) \to \exp(-x^2)/(\sqrt{\pi}x)$] leads to

$$\tau_n^{-1}(z) = \frac{\overline{c_n^0} N_t C \sqrt{\epsilon_R}}{\sqrt{\pi kT}(\Delta + \epsilon_R)} \exp \left\{ - \frac{E_{\mathrm{act}}^{\mathrm{QM}}[\Gamma(z)]}{kT} \right\}. \tag{24}$$

Hence, τ_n is thermally activated with the activation energy

$$E_{\mathrm{act}}^{\mathrm{QM}}(\Gamma) = \frac{(E_t + \Gamma - \epsilon_R)^2}{4\epsilon_R}, \tag{25}$$

where $\Gamma(z) = \Delta(z) - E_t = E_0 - E_c(z)$ is the additional offset between the conduction band edge and the lowest eigenenergy. The barrier for electron capture is enlarged by the subband offset $\Gamma(z)$ with respect to the corresponding classical expression, which is $E_{\mathrm{act}}^0 = E_{\mathrm{act}}^{\mathrm{QM}}(0)$.

The same form of the thermal activation energy $E_{\mathrm{act}}^{\mathrm{QM}}$ was used by Michler *et al.* in order to deduce the trap energy E_t from time-resolved photoluminescence experiments on oxygen-doped GaAs-based quantum wells of varying widths.[15] They found good agreement with independent bulk measurements of a presumably damage-induced deep level.

The classical lifetime (15) can be treated in a similar manner. For Boltzmann statistics and small phonon energy, one obtains

$$\tau_{n,\mathrm{cl}}^{-1}(z) = \frac{B e^{E_t/kT}}{\hbar \omega_0} \int_{E_t} dE \, e^{-E/2kT} I_{E/\hbar \omega_0}(\zeta) \sqrt{\frac{E-E_t}{kT}}$$
$$\approx \frac{B e^{E_t/kT}}{\hbar \omega_0 \sqrt{2\pi}} \int_{E_t} dE \, W(E) \sqrt{\frac{E-E_t}{kT}}, \tag{26}$$

with $B = 2\overline{c_n^0} N_t e^{-S(2f_B+1)}/(\sqrt{\pi} kT)$. Applying the high-temperature approximation as before leads to

$$\tau_{n,\mathrm{cl}}^{-1}(z) \approx \frac{\overline{c_n^0} N_t C}{\pi \sqrt{a}(kT)^{3/2}} \int_{E_t} dE \, \exp[f(E)], \tag{27}$$

where

$$f(E) = - \frac{(E+a)^2}{4akT} + \frac{1}{2} \ln \left(\frac{E-E_t}{kT} \right).$$

As in Ref. 7, the integral is approximated by expanding f around its maximum at the dominant transition energy E^*, yielding the formula:

$$\int_{E_t}^{\infty} dE \, \exp[f(E)] \approx \exp[f(E^*)] \sqrt{\frac{\pi}{2|f''(E^*)|}}$$
$$\times \mathrm{erfc} \left(\sqrt{\frac{|f''(E^*)|}{2}} (E_t - E^*) \right). \tag{28}$$

From $f'(E^*) = 0$, one finds

$$E^* \approx E_t + \frac{akT}{E_t + a}, \tag{29}$$

$$f(E^*) \approx - \frac{(E_t - a)^2}{4akT} - \frac{E_t}{kT} + \frac{1}{2} \ln \left(\frac{a}{E_t + a} \right), \tag{30}$$

$$f''(E^*) \approx - \frac{(E_t + a)^2}{2(akT)^2}, \tag{31}$$

which give the following result for the lifetime (replacing a by ϵ_R):

$$\tau_{n,\mathrm{cl}}^{-1}(z) \approx \frac{\mathrm{erfc}(-1/2) \overline{c_n^0} N_t C \epsilon_R}{\sqrt{\pi kT}(E_t + \epsilon_R)^{3/2}} \exp \left(- \frac{E_{\mathrm{act}}^0}{kT} \right). \tag{32}$$

With the results (23) and (32), a "quantum correction" factor with respect to the zero-field "classical" SRH lifetime can be defined:

$$g_{\mathrm{QM}}(z) := \frac{\tau_{n,\mathrm{cl}}}{\tau_n} \approx \sqrt{\frac{\epsilon_R + E_t}{\epsilon_R}} \frac{\exp \left(\frac{E_{\mathrm{act}}^0 - E_{\mathrm{act}}^{\mathrm{QM}}(z)}{kT} \right)}{\mathrm{erfc}(-1/2)}, \tag{33}$$

where the z-dependence outside the exponents has been neglected ($\Delta \approx E_t$). The remaining exponential contains the difference of the activation energies (compare Ref. 7).

As long as the lowest subband alone dominates the lifetime, this approximation works very well, in case of the triangular potential for sufficiently high fields and $z \lesssim 2$ nm (Fig. 2). The case $\Delta = 0$ marks the classical turning point where g_{QM} is of the order of unity, with the parameters used here.

The calculation of g_{QM} still requires the knowledge of the eigenenergy E_0 of the lowest subband, in order to calculate the local value of Δ. However, E_0 may not be explicitly available, for example, if the density gradient model is used to account for quantization. In these cases E_0 can at least be obtained in the lowest subband approximation from formula (13) by expressing the dominating probability density as $|\varphi_0(z)|^2 = n(z)/\int n(z)dz$:

$$E_0 = E_F^n - kT\mathsf{F}_0^{-1}\left(\frac{kTm_{xy}}{\pi\hbar^2}\int n(z)dz\right). \tag{34}$$

For a device simulator, the problem would consist in choosing an appropriate interval for integrating the density. In addition, the device regions must be determined where expression (33) is a sufficient approximation.[16]

V. LIFETIME PROFILES FOR SIMULATED DEVICES

Up to this point, a very simple potential shape has been assumed. In order to study more realistic situations, the input for calculating the lifetime profile (band edges, quasi-Fermi energies, eigenenergies, and wave functions) was taken from device simulations,[17] that solved the coupled system of the Poisson equation and the continuity equations. In all examples, the relaxation energy $\epsilon_R = 0.238$ eV was retained regardless of the material, and E_t was chosen such that the trap level is located in the middle of the bandgap.

A. Metal-oxide-semiconductor diode

A metal-oxide-semiconductor (MOS) diode with a p-doped silicon substrate (10^{18} cm^{-3}), a strongly n-doped silicon gate (10^{20} cm^{-3}), and an oxide width of 2 nm was simulated using a self-consistent Schrödinger–Poisson solver.[1] Quantization was incorporated for electrons only. Direct tunneling through the oxide and SRH recombination were enabled.[18] For this simulation, constant SRH lifetimes were used ($\tau_n^0 \approx 9.9 \times 10^{-8}$ s and $\tau_p^0 \approx 2.97 \times 10^{-8}$ s). The simulated charge profiles are shown in Fig. 4. From the results of this simulation, a new electron lifetime profile was calculated using a phonon energy of $\hbar\omega_0 = 2$ meV (Fig. 4, right axis). The lifetime decreases by almost four orders of magnitude from the wall to the minimum located at 8 nm away. This corresponds to the behavior already analyzed for the triangular potential. It is well reproduced by the lowest subband approximation within the first 3 nm. Beyond the minimum, the lifetime increases arriving at a constant value around $z \approx 50$ nm. This can be explained by the ceasing field enhancement effect as the field strength decreases towards the end of the space charge region.

In the MOS diode, the SRH recombination rate is only influenced by the field enhancement in the space-charge region. The strong increase near the oxide has no effect because the hole density is very small there.

B. Quantum-well diode

The simulated structure corresponds to an intrinsic GaAs quantum well between two layers of $Al_{0.4}Ga_{0.6}As$ that form a pin diode. The p- and n-doping of the AlGaAs regions is 10^{17} cm^{-3}. Spherical parabolic bands are used
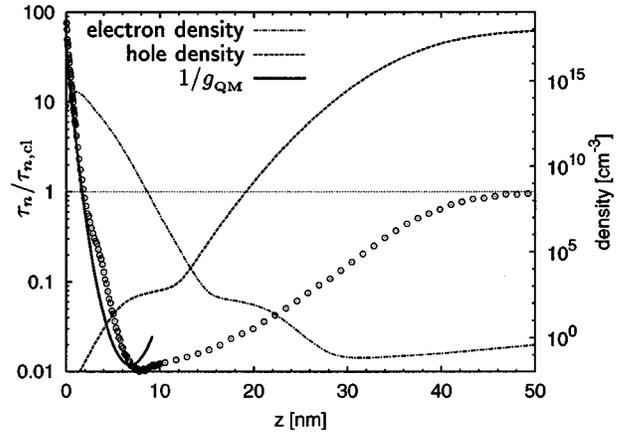
FIG. 4. Profiles of the electron and hole density (dashed lines, right axis), the electron lifetime τ_n (symbols), and lowest subband approximation for τ_n using E_0 from expression (34) (solid line). The horizontal dotted line indicates the zero-field lifetime.

for electrons and holes, with an electron effective mass of $m_e^* = 0.0672\,m_0$ and two hole effective masses: $m_{lh} = 0.0485\,m_0$ and $m_{hh} = 0.407\,m_0$.

Two well widths were considered: 30 and 5 nm. The Schrödinger equation was solved for both carrier types in a region containing the GaAs well and a few nanometers of the adjacent material. In the simulations, constant SRH lifetimes were used: $\tau_n^0 = 10^{-8}$ s and $\tau_p^0 = 10^{-9}$ s, for electrons and holes, respectively. Results are presented for 1 V forward bias. Band edges and carrier densities are shown in Fig. 5(a), the resulting lifetime profiles in Fig. 5(b) for $\hbar\omega_0 = 2$ meV. One can observe differences with respect to the confinement length and carrier type. For the 30 nm well, the electron lifetime shows a distinct field reduction in the left part of the well, corresponding to higher subbands, and an increasing slope towards the right wall. Both features are far less pronounced in the hole lifetime profile. This can be explained by the dominance of the heavy-hole band. The light-hole band does not contribute much because its in-plane DOS is smaller.

In the narrow well (5 nm), the electron lifetime is considerably higher than in the wide well. For the hole lifetime, this is not the case. The crucial parameter is the shift of the lowest subband with shrinking well width. The heavy hole subband is not lifted much with respect to the minimum potential (regarding hole energies), but the electron subband is.

The numerical lifetime profiles are compared with the lowest subband approximation (33). They agree very well for the electrons in the narrow structure and least for the holes in the wide structure [Fig. 5(b)]. This illustrates the limitations due to the influence of quantization mass and confinement on the subband separation.

The change of the SRH rate upon replacing the constant SRH lifetimes with the calculated profiles is shown in Fig. 5(c). For all rates, the QM densities are used. The ratio of the trap level degeneracy factors in Eq. (4) is set to $g = 1$. The 30 nm structure shows an overall increased rate. The enhancement reaches a factor of 2 and more towards the left end of the well. The maximum increases by approximately 70%.
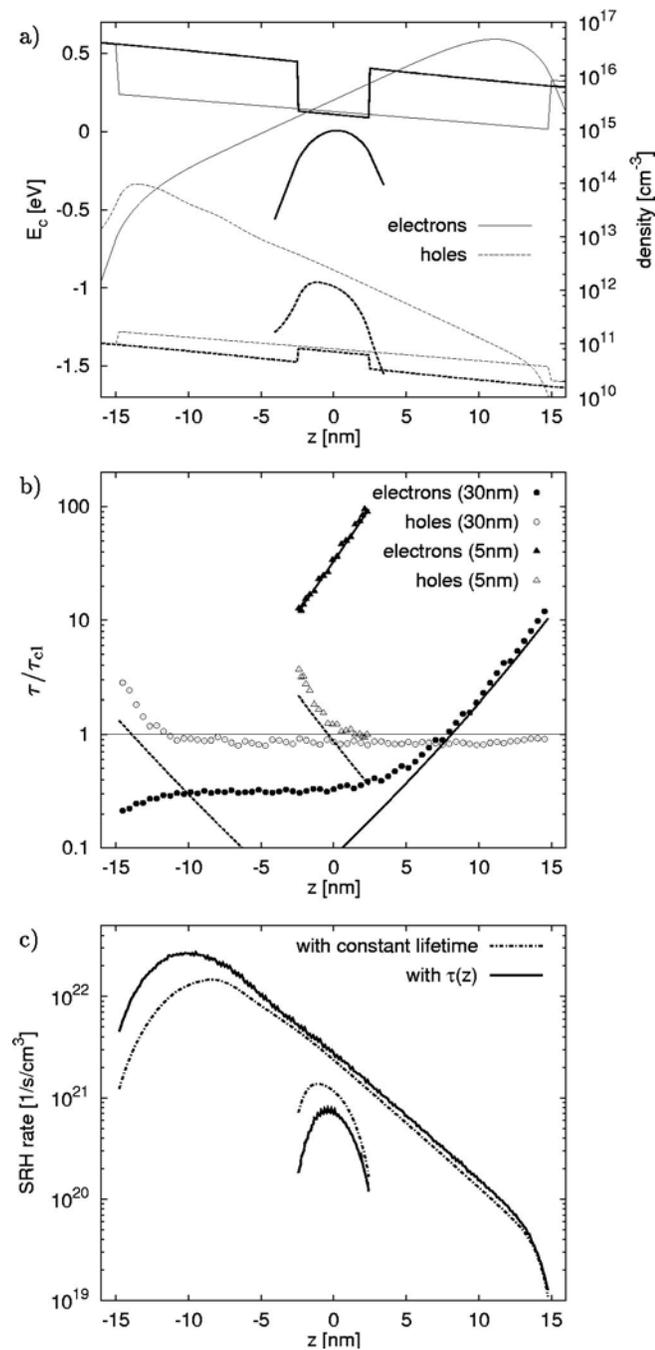
FIG. 5. Carrier density (a), lifetime (b), and SRH rate profiles (c) at 1 V forward bias inside two quantum wells of different widths (30 and 5 nm, to be discerned by the different plot ranges along z). (a) Simulation results for the conduction and valence band edges (left axis) and the carrier densities as calculated from the eigenstates (right axis). (b) Ratio of the QC and bulk lifetime, τ and τ_{cl}, respectively, for both electrons (filled symbols) and holes (open symbols). Lines indicate the lowest subband approximation $1/g_{QM}(z)$ with E_0 obtained from formula (34). (c) SRH rate obtained with constant lifetimes ($\tau_n^0 = 10^{-8}$ s and $\tau_p^0 = 10^{-9}$ s, dash-dotted lines) and with the QC lifetime profile from (b) after multiplying by $\tau_{n,p}^0$ (solid lines).

The rate in the 5 nm structure is reduced, mainly in the left half of the well. The maximum decreases by about 70%. Of course, the impact of the spatially varying lifetimes on the rate is controlled by the densities and $\tau_{n,p}^0$.

Reduced nonradiative capture rates due to confinement were also found by Delerue *et al.* in a theoretical study of recombination in small silicon crystallites.[19] Juršėnas *et al.* examined the recombination of photoexcited hot carriers in CdS nanocrystals.[20] The multiphonon contribution was divided into an interface and a volume channel with different activation energies. The authors state that the particle size was too large to see confinement effects; the observed size dependence was attributed to the changing surface-to-volume ratio.

VI. CONCLUSIONS

The SRH lifetime has been modified to consistently account for quantization effects. The model combines a local DOS composed of the eigenstates of the system with multiphonon recombination processes. Strongly localized traps and a single phonon energy are assumed. The model has been investigated for small phonon energies ($E_t \gg \hbar\omega_0$). The results exhibit two effects:

(1) On one hand, the electric field may enable tunneling of the wave function, which enhances the capture and emission processes and thus reduces the lifetime.
(2) On the other hand, the additional separation of the lowest subband from the band edge can cause a considerable increase of the lifetime. For this effect, an analytical approximation has been derived. The form of the activation energy is supported by an earlier experiment.[15]

Note that we did not consider recombination in the barriers of a quantum well or any other nonradiative channels for carrier loss such as thermionic escape. The latter causes an increasing nonradiative contribution with decreasing well width,[21] in contrast to recombination *within* the confined structure, as was considered here.

Modified lifetimes have been computed for quantized carriers in one-dimensional devices:

(1) In a MOS diode, the quantization leads to enhanced recombination in the depletion layer due to the electric field, but the subband offset near the oxide barrier is of no importance.
(2) The 30 nm quantum well *pin* diode shows both features for the lifetime profiles, but the SRH rate is governed by field-enhanced recombination. In the 5 nm device, the SRH rate is reduced due to the dominating subband separation.

Including these effects may be of importance for the modeling of small confining structures, e.g., quantum-well devices. To this end, a compact model for both quantum-confinement and field-enhancement effects would be desirable. A drawback in this respect may be the nonlocal form of the analytical "quantum correction" factor g_{QM}, requiring the lowest eigenenergy E_0.

APPENDIX: LOCAL DENSITY OF STATES

In this section, the concept of a local density of states is illustrated by a few examples. Firstly, the local DOS is defined as
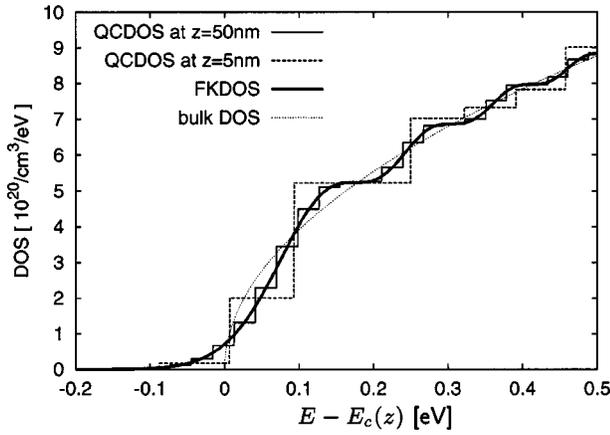
FIG. 6. Local DOS of bound states (QCDOS) in a triangular well at different distances from the wall for a field of $F=10^6$ V/cm. For comparison, the bulk DOS and FKDOS are also shown. Only one silicon valley with $m_z = m_l$ and $m_{xy} = m_t$ was considered.

$$N(E, \mathbf{r}) = 2 \sum_n |\psi_n(\mathbf{r})|^2 \delta(E - E_n), \qquad (A1)$$

with the wave function ψ_n corresponding to the eigenenergy E_n. A factor of 2 for the spin is included. For plane waves, $\psi_{\mathbf{k}}(\mathbf{r}) = 1/\sqrt{V} \exp(i\mathbf{k} \cdot \mathbf{r})$, with $V = L_x L_y L_z$ being the normalization volume, the summation transforms into $[V/(2\pi)^3] \Sigma_{\mathbf{k}}$ and the usual DOS expression is obtained: $N(E) = 2/(2\pi)^3 \Sigma_{\mathbf{k}} \delta(E - E_{\mathbf{k}})$.

Now, the potential is considered to be nonconstant in the z direction with one classical turning point. The motion in the z direction is then described by the eigenenergy $E_\parallel$ and the wave function $\varphi(E_\parallel, z)$, leading to the following ansatz for the total wave function:

$$\psi_{\boldsymbol{\kappa}, E_\parallel}(\mathbf{r}) = (L_x L_y)^{-1/2} \exp(i\boldsymbol{\kappa} \cdot \boldsymbol{\rho}) \varphi(E_\parallel, z), \qquad (A2)$$

and the total energy

$$E(\boldsymbol{\kappa}, E_\parallel) = E_\parallel + \frac{\hbar^2 \boldsymbol{\kappa}^2}{2m_{xy}}, \qquad (A3)$$

where parabolic bands have been assumed. The wave and position vectors perpendicular to the z direction are $\boldsymbol{\kappa}$ and $\boldsymbol{\rho}$, respectively. After integrating over $E_\parallel$, the local DOS retains a z dependence:

$$N_c(E, z) = \frac{1}{2\pi^2} \int d^2\boldsymbol{\kappa} \left| \varphi\left(E - \frac{\hbar^2 \boldsymbol{\kappa}^2}{2m_{xy}}, z\right) \right|^2. \qquad (A4)$$

In the case of a homogeneous field F, the one-dimensional Schrödinger equation has the solution

$$\varphi(E, z) = \sqrt{\frac{2m_z}{\hbar^2 q_z}} \mathrm{Ai}[q_z(z - z_E)],^{22} \qquad (A5)$$

where $z_E = E/(eF)$ denotes the classical turning point, $q_z = (2m_z eF/\hbar^2)^{1/3}$, and $\mathrm{Ai}(x) := (1/\pi) \int_0^\infty \cos(t^3/3 + tx) dt$ is the Airy function.[23] With this expression, the integration in Eq. (A4) produces the FKDOS in Eq. (11).

In the case of bound states, the integral over $E_\parallel$ is replaced by a sum Σ_i over discrete eigenstates φ_i and eigenenergies E_i:

$$N_c(E, z) = \sum_i \frac{2}{(2\pi)^2} \int d^2\boldsymbol{\kappa} |\varphi_i(z)|^2 \delta\left(E - E_i - \frac{\hbar^2 \boldsymbol{\kappa}^2}{m_{xy}}\right)$$

$$= \frac{m_{xy}}{\pi \hbar^2} \sum_i |\varphi_i(z)|^2 \Theta(E - E_i). \qquad (A6)$$

A reasonable approximation for an NMOS inversion layer is the triangular potential: The electron is only allowed to reside in the region $z > 0$ where the potential is Fz. Solutions are those Airy functions that vanish at $z = 0$. With proper normalization, these are

$$\varphi_i(z) = \sqrt{\frac{q_z}{\mathcal{F}(-q_z z_i)}} \mathrm{Ai}[q_z(z - z_i)], \qquad (A7)$$

where q_z is the same as above and $z_i = E_i/(eF)$ is the right classical turning point for the ith eigenstate (see Refs. 24 and 25 for E_i). With these solutions the local DOS (12) approaches expression (11) for large distances z (see Fig. 6).

[1] A. Wettstein, Quantum Effects in MOS Devices (Hartung-Gorre, Konstanz, 2000).
[2] M. G. Ancona, IEEE Trans. Electron Devices 47, 1449 (2000).
[3] A. S. Spinelli, A. Pacelli, and A. L. Lacaita, IEEE Trans. Electron Devices 47, 2366 (2000).
[4] G. G. Zegrya and A. S. Polkovnikov, JETP 86, 815 (1998).
[5] W. Shockley and W. T. Read, Phys. Rev. 87, 835 (1952).
[6] A. Schenk, J. Appl. Phys. 71, 3339 (1992).
[7] A. Schenk, Solid-State Electron. 35, 1585 (1992).
[8] K. Huang and A. Rhys, Proc. R. Soc. London, Ser. A 204, 406 (1950).
[9] E. Gutsche, Phys. Status Solidi B 109, 583 (1982).
[10] A. Schenk, D. Suisky, and R. Enderlein, Acta. Phys. Pol. A 71, 315 (1987).
[11] D. E. Aspnes, Phys. Rev. 147, 554 (1966).
[12] W. Franz, Z. Naturforsch. A 13, 484 (1958).
[13] L. V. Keldysh, Sov. Phys. JETP 7, 788 (1958).
[14] This effect strongly depends on carrier statistics and is less pronounced for the degenerate case.
[15] P. Michler, T. Forner, V. Hofsäß, F. Prins, K. Zieger, F. Scholz, and A. Hangleiter, Phys. Rev. B 49, 16632 (1994).
[16] In a MOS inversion layer, this region may be much smaller than reasonable integration intervals.
[17] ISE AG, ISE TCAD Manuals, Release 8.0, DESSIS (Zurich, 2000).
[18] A. Schenk and G. Heiser, J. Appl. Phys. 81, 7900 (1997).
[19] C. Delerue, G. Allan, and M. Lannoo, Phys. Rev. B 48, 11024 (1993).
[20] S. Juršėnas, G. Tamulaitis, G. Kurilčik, and A. Žukauskas, Phys. Scr., T 79, 187 (1999).
[21] J. P. Bergman, P. O. Holtz, B. Monemar, M. Sundaram, J. L. Merz, and A. C. Gossard, Mater. Sci. Forum 143–147, 629 (1994).
[22] L. Landau and E. Lifschitz, Lehrbuch der Theoretischen Physik, Quantenmechanik (Akademie, Berlin, 1985).
[23] Handbook of Mathematical Functions, edited by M. Abramowitz and I. Stegun (Dover, New York, 1970).
[24] F. Stern, Phys. Rev. B 5, 4891 (1972).
[25] T. Ando, A. B. Fowler, and F. Stern, Rev. Mod. Phys. 54, 437 (1982).