

 Open access • Journal Article • DOI:10.1016/0038-1101(93)90065-X

Rigorous theory and simplified model of the band-to-band tunneling in silicon

— [Source link](#) 

Andreas Schenk

Institutions: École Polytechnique Fédérale de Lausanne

Published on: 01 Jan 1993 - Solid-state Electronics (Pergamon)

Topics: Scanning tunneling spectroscopy, Stochastic tunneling and Quantum tunnelling

Related papers:

- [A new recombination model for device simulation including tunneling](#)
- [Theory of Tunneling](#)
- [Zener tunneling in semiconductors](#)
- [Tunnel field-effect transistors as energy-efficient electronic switches](#)
- [New Phenomenon in Narrow Germanium p – n Junctions](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/rigorous-theory-and-simplified-model-of-the-band-to-band-1qpossrrp>

RIGOROUS THEORY AND SIMPLIFIED MODEL OF THE BAND-TO-BAND TUNNELING IN SILICON

A. SCHENK

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

(Received 9 May 1992)

Abstract—The phonon-assisted band-to-band tunneling rate in crystalline silicon is calculated using the equilibrium Green's function formalism. Electron-phonon collisions, that balance the momentum, are included in the perturbation operator. Houston-type solutions are used for the time dependence of the Bloch states. RPA decoupling yields a tractable expression for the differential tunneling conductivity. Its evaluation is presented explicitly, taking exactly into account the anisotropy of the six conduction band valleys. A simplified rate formula for the purpose of device simulation is then derived from the general expression, restricting the field strength and using reasonable models for the matrix elements. It is shown that indirect, phonon-assisted tunneling largely exceeds direct tunneling at all events. Finally, band-to-band tunneling is compared with trap-assisted tunneling. We conclude that the pre-breakdown range in silicon is dominated by tunneling via traps.

1. INTRODUCTION

In 1934 Zener[1] proposed band-to-band tunneling as an explanation for the electrical breakdown. A modified Zener theory was used by McAfee *et al.*[2] in 1951 to describe the breakdown of reversed biased p - n -junctions, called *Zener diodes* since then. However, experimental work[3,4] in the following years showed that in such diodes with wide junctions the breakdown is not caused by tunneling, but by impact ionization. Only in narrow junctions, where the width of the transition region is less than 50 nm, the necessary field strength for tunneling is reached before the avalanche effect sets in. This was first clearly demonstrated by Chynoweth and McKay[5] in 1957 by the absence of microplasma noise and by the temperature coefficients of reverse and forward characteristics of junctions with different breakdown voltages. In the same year Esaki[6] discovered that narrow p - n -junctions between degenerate regions can have forward characteristics with a portion of negative differential conductivity, and that the tunnel "hump" is only weakly temperature dependent. Esaki's work initiated intensive experimental and theoretical investigations. Holonyak *et al.*[7] and Hall[8] observed structures in the I - U -characteristics of heavily doped Si-junctions at 4.2 K, which they attributed to the momentum conserving phonons in indirect band-to-band tunneling. Various phonon energies could be resolved in these characteristics. Chynoweth *et al.*[9,10] then found evidence that the excess current in silicon Esaki junctions, i.e. the current between the tunnel "hump" and the normal forward injection current, is essentially caused by the process of field ionization of impurity levels. This mechanism had been suggested by Yajima and Esaki[11]. It was confirmed by Sah[12],

who measured the excess current in gold-doped silicon tunnel diodes and found different onset-voltages for the various possible tunneling paths via the two gold levels. Brody[13] suggested another explanation for the excess current. He considered transitions between the tail states of the heavily doped material and tried to find a correlation between tunneling characteristic and shape of density of states. Chynoweth *et al.*[10] investigated the dependence of the excess current on radiation damage and observed a linear increase of that current with the bombardment dose. Furthermore, phonon-assisted tunneling was correlated to a S-bend shape of the peak current-versus-temperature curve, which did not occur for the excess current. By measuring the derivative of the conductance in silicon Esaki diodes at 4.2 K, Chynoweth *et al.*[14] could reveal 12 phonon and phonon-combination energies, which agreed well with results of neutron scattering studies. In a subsequent paper Logan and Chynoweth [15] succeeded to decompose the tunneling current of silicon Esaki junctions into a phonon-unassisted current, a TA phonon-assisted current and a TO phonon-assisted component. For biases below the peak voltage the amount of excess current was found to be negligible compared with the total current. The TA phonon- and TO phonon-assisted currents contributed equally to the band-to-band tunneling current. Tyagi[16] 1967 made extensive studies at silicon p - n -junctions to clarify the relative influence of avalanche generation and band-to-band tunneling. He found that narrow junctions with breakdown voltages up to 5 V show a soft breakdown, which is entirely caused by tunneling up to applied voltages of 3 V. At slightly higher voltages onset of multiplication of field-generated carriers was observed, but without a hard breakdown up to 6 V. Agreement

with the avalanche theory of Chynoweth[17] proved only for diodes with breakdown voltages larger than 15 V. Tyagi concluded that between 3 and 14 V both mechanisms operate simultaneously. Fair and Wivell [18] produced SIMS-characterized linearly graded junctions and investigated the influence of the junction grade constant on the current mechanism. They found that the sharp transition from a tunneling-dominated reverse current to a multiplied-generation current occurred at voltages, which increased with decreasing grade constant a until $a \approx 1 \times 10^{24} \text{ cm}^{-4}$. For smaller gradients the tunneling component vanished.

In recent years band-to-band tunneling in silicon has again received attention from scaled bipolar devices and trench transistor DRAM cells. As a result of scaling very shallow junctions with high doping levels and steep gradients came into use. Stork and Isaac[19] studied the influence of different energies for the boron implantation on the reverse tunneling current in emitter-base n^+p junctions. The drastically increased reverse current after high energy implantation was explained in terms of a flat base doping profile yielding higher maximum field strengths than a sharply decreasing B-profile after low energy implantation. The authors could rule out avalanche generation, surface recombination and damage-related bulk SRH recombination, but they did not include trap-assisted tunneling in their discussion. As an interesting fact, the boron doping level did not exceed 10^{18} cm^{-3} . Cuthbertson and Ashburn[20] observed a highly nonideal base current due to a peripheral p^+-n^+ junction, which arose as consequence of the lateral spread of the extrinsic base regions beneath the side-wall oxide of the polysilicon emitter in their self-aligned transistors. They attributed the nonideal characteristic to the excess current known from p^+-n^+ Esaki diodes. This was supported by the linearity of both the forward and reverse characteristics at small applied voltages, which is typical for the excess current. With an adequate spacer technology the perimeter tunneling could be suppressed. Del Alamo and Swanson[21] pointed out, that as scaling advances, tunneling must be expected at the whole emitter-base area posing a principal limit to the gain. Various papers[22–27] dealt with drain-substrate tunneling leakage in MOSFETs and its reduction, with hot carrier injection across the gate oxide initiated by tunneling generation[28,29], and with tunneling leakage in trench transistor DRAM cells[30,31]. In all these papers either band-to-band tunneling or tunneling via bulk or surface traps was considered the leakage mechanism.

The first calculation of the probability of phonon-assisted band-to-band tunneling transitions was presented by Keldysh in 1958[32]. He used second order perturbation theory, Houston approximations for the wave functions[33] and the saddle-point method. Independently, Price and Radcliffe[34] obtained an expression for the phonon-assisted tunneling current

with the Wentzel–Kramers–Brillouin (WKB) method. Keldysh's result was also adapted and improved by Kane[35]. In all these papers the tunnel diode problem was solved by determining the transmission coefficient of an electron striking the junction barrier and then calculating the current by the number of generated carriers. A new access to the diode problem was given by Fredkin and Wannier[36] who used a model Hamiltonian (the Fredkin–Wannier operator) of the Esaki diode. In this model the usual Bloch Hamiltonian was superimposed by a finite, step-like electric field of the transition region. The current was obtained by means of the scattering amplitude for scattering of the Bloch electrons at this inhomogeneous field.

The connection between transmission probability and current density is not necessary, if a macroscopic quantity is calculated which directly determines the band-to-band tunneling current. This was done for the first time by Enderlein and Peuker[37]. They used a Kubo formula[38] for the differential conductivity of a crystal in strong electric fields. It was shown that the conductivity can be split into an intraband part and a band-to-band part, if collisions are neglected. The latter is determined by the off-diagonal elements of the one-particle density matrix and arises, because the electrons change their place when they penetrate the barrier.

In this paper we will follow the lines of Enderlein and Peuker[37] and apply the Green's function formalism to the phonon-assisted band-to-band tunneling in silicon. In Section 2 the current density and generation rate are calculated. The procedure is related to the Effective Mass Approximation (EMA), thus no semi-classical approach is necessary. The conduction band anisotropy is exactly taken into account. In Section 3 we simplify the microscopic model to make it suitable for device simulation. Direct tunneling calculated with the same method is compared with the results for indirect tunneling at the end of this section. The discussion in Section 4 focusses on the comparison between band-to-band and trap-assisted tunneling in silicon. The major weakpoints of the tunneling theory are mentioned.

2. MICROSCOPIC MODEL

2.1. Kubo formalism for the tunneling conductivity

The starting point is Kubo's formula[38] applied to the tensor of conductivity $\hat{\sigma}(\omega, \mathbf{F})$ [37]:

$$\hat{\sigma}(\omega, \mathbf{F}) = i \frac{e^2 N}{m(\omega + i\epsilon)} \hat{\mathbf{i}} + \frac{1}{\hbar(\omega + i\epsilon)} \times \int_0^\infty dt e^{i\omega t} \langle [\mathbf{j}(0), \mathbf{j}(-t)]_- \rangle_\infty, \quad (1)$$

where the symbols have the following meaning: ω —frequency, \mathbf{F} —field strength, N —number of

electrons, \mathbf{j} —current density operator.† The time dependence of the latter is given by:

$$\mathbf{j}(t) = e^{iHt/\hbar} \mathbf{j}(0) e^{-iHt/\hbar}. \quad (2)$$

In eqn (1) for the correlation function the abbreviation:

$$\langle [\mathbf{j}(0), \mathbf{j}(-t)]_- \rangle_\infty = \lim_{t' \rightarrow \infty} \text{Tr}\{\rho(t') [\mathbf{j}(0), \mathbf{j}(-t)]_- \} \quad (3)$$

with

$$\rho(t) = e^{-iHt/\hbar} \rho_0 e^{iHt/\hbar} \quad (4)$$

was used. The quantity ρ_0 denotes the density operator of the grand canonical ensemble of the field-free system:

$$\rho_0 = \mathcal{Z}^{-1} e^{-\beta(H_0 - \mu\hat{N})}, \quad \mathcal{Z} = \text{Tr}\{e^{-\beta(H_0 - \mu\hat{N})}\} \quad (5)$$

($\beta = 1/kT$). We are interested in the first non-vanishing expression that describes indirect, i.e. phonon-assisted, transitions from $E_{v\Gamma_{25'}}$ to $E_{c\Delta_1}$ and vice versa (see Fig. 1). Therefore, the electron-phonon coupling has to be included in the unperturbed Hamiltonian H_0 in (5), which, however, states an unsolvable problem. For our purpose electron-phonon collisions have to be taken into account only in so far as they serve as a momentum source for the tunneling electrons. Thus, H_0 in (5) can be replaced by:

$$H_{00} = \sum_{\mu=c,v} \int \frac{d^3\mathbf{k}}{4\pi^3} E_\mu(\mathbf{k}) a_\mu^+(\mathbf{k}) a_\mu(\mathbf{k}) + \int d^3\mathbf{q} \hbar\omega_{\mathbf{q}} b^+(\mathbf{q}) b(\mathbf{q}). \quad (6)$$

The Hamiltonian is completed by the two terms describing the coupling with \mathbf{F} and the phonon ensemble:

$$H_F = iFe \sum_{\mu=c,v} \int \frac{d^3\mathbf{k}}{4\pi^3} a_\mu^+(\mathbf{k}) \frac{\partial}{\partial k_\parallel} a_\mu(\mathbf{k}), \quad (7)$$

$$H_{\text{int}} = \sum_{\mu=c,v} \int \frac{d^3\mathbf{k}}{4\pi^3} \int d^3\mathbf{q} V_\mu(\mathbf{q}) a_\mu^+(\mathbf{k} + \mathbf{q}) a_\mu(\mathbf{k}) \times [b_{\mathbf{q}} - b_{-\mathbf{q}}^\dagger]. \quad (8)$$

Equation (7) indicates that the constant field approximation is used in this calculation. A possible generalization to inhomogeneous fields in form of an analytical interpolation between WKB and EMA scheme was given in Ref. [39] for the case of direct transitions. At the end of Section 2 we will briefly discuss the application of this treatment to the case of phonon-assisted tunneling.

The a^+ , a , b^+ , b denote creation and annihilation operators of Bloch electrons and phonons, respectively. In silicon the direct gap at Δ_1 is comparable with that at Γ . After [40] we have $\Gamma_{15} - \Gamma_{25'} = 3.42$ eV and $X_1 - X_4 = 4.03$ eV with a monotonously increasing direct gap from Γ to X (see Fig. 1). Therefore,

†The volume of the crystal was set equal to unity throughout the paper.

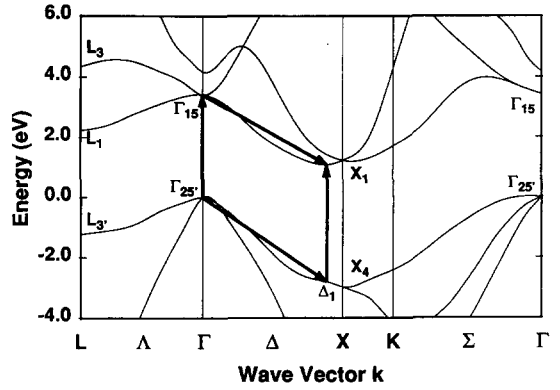


Fig. 1. Indirect (phonon-assisted) band-to-band tunneling in silicon. The two transitions via intermediate states opposite to the band extrema of the indirect gap are indicated by bold arrows.

both electron-phonon and hole-phonon collisions must be taken into account in (8).

The conductivity $\hat{\sigma}(\omega, \mathbf{F})$, given by eqn (1), can be split into an intraband and an interband part [37]. In the limit $\omega \rightarrow 0$ one may interpret the latter as tensor of differential tunneling conductivity[37]:

$$\hat{\sigma}(\mathbf{F}) = \frac{d}{d\omega} \hat{\mathbf{L}}(\omega, \mathbf{F}) \Big|_{\omega=0}, \quad (9)$$

$$\hat{\mathbf{L}}(\omega, \mathbf{F}) = \frac{1}{\hbar} \int_0^\infty dt e^{i\omega t} \langle [\mathbf{j}(0), \mathbf{j}(-t)]_- \rangle_\infty. \quad (10)$$

If we write the current density operator $\mathbf{j}(t) = e/m\mathbf{p}(t)$ in the form of second quantization using Bloch states, the tensor $\hat{\mathbf{L}}(\omega, \mathbf{F})$ becomes:

$$\hat{\mathbf{L}}(\omega, \mathbf{F}) = \frac{e^2}{m^2\hbar} \int_0^\infty dt e^{i\omega t} \int \frac{d^3\mathbf{k}}{4\pi^3} \frac{d^3\mathbf{k}'}{4\pi^3} + \sum_{\substack{\mu\nu \\ \mu'\nu'}} \mathbf{p}^{\mu'\nu'}(\mathbf{k}') \mathbf{p}^{\mu\nu}(\mathbf{k}) \mathbf{K}_{\mu\nu}^{\mu'\nu'}(\mathbf{k}', \mathbf{k}; t) \quad (11)$$

with

$$\mathbf{K}_{\mu\nu}^{\mu'\nu'}(\mathbf{k}', \mathbf{k}; t) = \langle [a_{\mu'\mathbf{k}'}^+ a_{\nu\mathbf{k}}, a_{\mu\mathbf{k}}^+(-t) a_{\nu\mathbf{k}}(-t)]_- \rangle_\infty. \quad (12)$$

For the momentum matrix elements the relation

$$\int d^3\mathbf{x} \varphi_{\mu\mathbf{k}}^*(\mathbf{x}) \mathbf{p}(0) \varphi_{\nu\mathbf{k}}(\mathbf{x}) = \delta_{\mathbf{k}\mathbf{k}'} \mathbf{p}^{\mu\nu}(\mathbf{k}), \quad (13)$$

has been assumed, which will be commented below. We change to the interaction representation in eqn (11) according to:

$$a_\sigma(\tau) = \mathcal{S}(0, \tau) \tilde{a}_\sigma(\tau) \mathcal{S}(\tau, 0). \quad (14)$$

The unitary \mathcal{S} -operator is given by:

$$\mathcal{S}(t, t') = \hat{\mathbf{1}} - \frac{i}{\hbar} \int_{t'}^t dt_1 H_{\text{int}}(t_1) \mathcal{S}(t_1, t'), \quad (15)$$

and obeys the relations:

$$\begin{aligned} \mathcal{S}(t', t) &= \mathcal{S}^\dagger(t, t'), \\ S(t, t_1) \mathcal{S}(t_1, t') &= \mathcal{S}(t, t'). \end{aligned} \quad (16)$$

The tilde indicates that the time development of the operator goes without H_{int} . Now, the correlation function (12) has the form:

$$\mathbf{K}_{\mu\nu}^{\prime}(\mathbf{k}', \mathbf{k}; t) = \langle [a_{\mu\mathbf{k}'}^{\dagger} a_{\nu\mathbf{k}'} \mathcal{S}(0, -t) \times \tilde{a}_{\mu\mathbf{k}}^{\dagger}(-t) \tilde{a}_{\nu\mathbf{k}}(-t) \mathcal{S}(-t, 0)]_{-} \rangle_{\infty}. \quad (17)$$

If the \mathcal{S} -operator is developed up to second order in H_{int} :

$$\mathcal{S}(t, t') = \mathcal{S}^{(0)}(t, t') + \mathcal{S}^{(1)}(t, t') + \mathcal{S}^{(2)}(t, t') \quad (18)$$

with:

$$\begin{aligned} \mathcal{S}^{(0)}(t, t') &= \hat{1} \\ \mathcal{S}^{(1)}(t, t') &= -\frac{i}{\hbar} \int_{t'}^t dt_1 H_{\text{int}}(t_1) \\ \mathcal{S}^{(2)}(t, t') &= \left(\frac{i}{\hbar}\right)^2 \int_{t'}^t dt_1 \int_{t'}^{t_1} dt_2 H_{\text{int}}(t_1) H_{\text{int}}(t_2), \end{aligned} \quad (19)$$

the zero and second order correlation functions turn out to be:

$$\mathbf{K}_{\mu\nu}^{(0)}(\mathbf{k}', \mathbf{k}; t) = \langle [a_{\mu\mathbf{k}'}^{\dagger} a_{\nu\mathbf{k}'} \tilde{a}_{\mu\mathbf{k}}^{\dagger}(-t) \tilde{a}_{\nu\mathbf{k}}(-t)]_{-} \rangle_{\infty}, \quad (20)$$

$$\begin{aligned} \mathbf{K}_{\mu\nu}^{(2),d}(\mathbf{k}', \mathbf{k}; t) &= \langle [a_{\mu\mathbf{k}'}^{\dagger} a_{\nu\mathbf{k}'} \mathcal{S}^{(2)}(0, -t) \tilde{a}_{\mu\mathbf{k}}^{\dagger}(-t) \\ &\times \tilde{a}_{\nu\mathbf{k}}(-t)]_{-} \rangle_{\infty} + \langle [a_{\mu\mathbf{k}'}^{\dagger} a_{\nu\mathbf{k}'} \\ &\times \tilde{a}_{\mu\mathbf{k}}^{\dagger}(-t) \tilde{a}_{\nu\mathbf{k}}(-t) \mathcal{S}^{(2)}(-t, 0)]_{-} \rangle_{\infty}, \end{aligned} \quad (21)$$

$$\begin{aligned} \mathbf{K}_{\mu\nu}^{(2),\text{ind}}(\mathbf{k}', \mathbf{k}; t) &= \langle [a_{\mu\mathbf{k}'}^{\dagger} a_{\nu\mathbf{k}'} \mathcal{S}^{(1)}(0, -t) \\ &\times \tilde{a}_{\mu\mathbf{k}}^{\dagger}(-t) \tilde{a}_{\nu\mathbf{k}}(-t) \mathcal{S}^{(1)}(-t, 0)]_{-} \rangle_{\infty}. \end{aligned} \quad (22)$$

The first order as well as odd orders vanish due to the trace over the phonon states. $\mathbf{K}^{(0)}$ is the term which describes direct transitions at a certain \mathbf{k} -value of the Brillouin zone, where the gap is $E_g^d = E_c(\mathbf{k}) - E_v(\mathbf{k})$. It is not sure from the start that these transitions occur at $\mathbf{k} = 0$, since the direct gap of silicon is almost constant over the whole Brillouin zone in [100]-direction.

The two terms $\mathbf{K}^{(2),d}$ represent second order corrections to the direct transition term caused by electron-phonon interaction. Here, direct transitions are accompanied by the emission or absorption of one phonon. The last term $\mathbf{K}^{(2),\text{ind}}$ contains the indirect transitions, which can be assumed to play the dominant role in silicon.

The evaluation of the correlation functions (20)–(22) requires the determination of the time development of the operators $\tilde{a}_{\mu\mathbf{k}}^{\dagger}(t) \tilde{a}_{\nu\mathbf{k}}(t)$ with the Hamiltonian $H_{\infty} + H_F$. An often used approximate solution (Houston-type [33]) is:

$$\begin{aligned} \tilde{a}_{\mu\mathbf{k}}^{\dagger}(t) \tilde{a}_{\nu\mathbf{k}}(t) &= \exp\left\{\frac{i}{\hbar} \int_0^t d\vartheta [E_{\mu}(\mathbf{k}_g) \right. \\ &\left. - E_{\nu}(H\mathbf{k}_g)]\right\} a_{\mu\mathbf{k}_t}^{\dagger} a_{\nu\mathbf{k}_t} \end{aligned} \quad (23)$$

with an ‘‘accelerated momentum vector’’ $\mathbf{k}_t = \mathbf{k} - e/\hbar \mathbf{F}t$ and the Bloch energy $E_{\nu}(\mathbf{k})$ (band structure).

2.2. Direct (zero phonon) transitions

We first consider the direct transitions. Using eqn (23), the correlation function $\mathbf{K}^{(0)}$ can easily be calculated. The result is:

$$\begin{aligned} \mathbf{K}_{\mu\nu}^{(0)}(\mathbf{k}', \mathbf{k}; t) &= \exp\left\{i \int_0^{-t} d\vartheta \omega_{\mu\nu}(\mathbf{k}_g)\right\} \\ &\times [\langle a_{\mu\mathbf{k}'}^{\dagger} a_{\nu\mathbf{k}'} \rangle_{\infty} \delta_{\mu\nu} \delta_{\mathbf{k}-\mathbf{k}'} \\ &- \langle a_{\mu\mathbf{k}'}^{\dagger} a_{\nu\mathbf{k}'} \rangle_{\infty} \delta_{\nu\mu'} \delta_{\mathbf{k}-\mathbf{k}'}], \end{aligned} \quad (24)$$

with the abbreviation

$$\hbar\omega_{\mu\nu}(\mathbf{k}) = E_{\mu}(\mathbf{k}) - E_{\nu}(\mathbf{k}). \quad (25)$$

Two simplifications are necessary for the further calculation. First, we neglect the off-diagonal elements of the one-particle density operator:

$$\langle a_{\mu\mathbf{k}}^{\dagger} a_{\mu'\mathbf{k}} \rangle_{\infty} \approx \delta_{\mu\mu'} f_{\mu}(\mathbf{k}). \quad (26)$$

Here $f_{\mu}(\mathbf{k}) = \langle a_{\mu\mathbf{k}}^{\dagger} a_{\mu\mathbf{k}} \rangle_{\infty}$ represents the Fermi function. Second, the time-dependent interband matrix element $\mathbf{p}^{\mu\nu}(\mathbf{k})$ are taken by their values for $t = 0$. Both approximations hold in the pre-breakdown range, but they are questionable beyond the breakdown voltage.

If we reduce the tensor $\mathbf{L}^{(0)}(\omega, \mathbf{F})$ to a scalar quantity, it can be written with (11) and (24) as

$$\begin{aligned} L^{(0)}(\omega, \mathbf{F}) &= \frac{e^2}{m^2 \hbar} \int \frac{d^3 \mathbf{k}}{4\pi^3} \sum_{\mu\nu} \\ &\times |\mathbf{p}^{\mu\nu}(\mathbf{k})|^2 [f_{\nu}(\mathbf{k}) - f_{\mu}(\mathbf{k})] \tau_{\mu\nu}^{\mathbf{k}}(\omega) \end{aligned} \quad (27)$$

with

$$\tau_{\mu\nu}^{\mathbf{k}}(\omega) = \int_0^{\infty} dt \exp\left\{i \int_0^t d\vartheta (\omega - \omega_{\mu\nu}(\mathbf{k}_g))\right\}. \quad (28)$$

The time $\tau_{\mu\nu}^{\mathbf{k}}(\omega)$ can be interpreted as the decay time of an electron-hole Bloch state under the action of an external electric field. According to eqn (13) the \mathbf{k} -dependence of the momentum matrix elements is due to the \mathbf{k} -dependence of the Bloch factors $u_{\mathbf{k}_t}(\mathbf{x})$. As is usual in effective mass approaches, we replace $u_{\mathbf{k}_t}(\mathbf{x})$ by $u_{\text{ov}}(\mathbf{x})$, and consequently $\mathbf{p}^{\mu\nu}(\mathbf{k})$ by $\mathbf{p}^{\mu\nu}(\mathbf{o})$.

Taking into account the lowest conduction and the highest valence band in (27), we have:

$$\begin{aligned} L^{(0)}(\omega, \mathbf{F}) &= \frac{e^2}{m^2 \hbar} |\mathbf{p}^{\text{cv}}(\mathbf{o})|^2 \int \frac{d^3 \mathbf{k}}{4\pi^3} \\ &\times [f_{\nu}(\mathbf{k}) - f_{\text{c}}(\mathbf{k})] \{\tau_{\text{cv}}^{\mathbf{k}}(\omega) - \tau_{\text{vc}}^{\mathbf{k}}(\omega)\} \end{aligned} \quad (29)$$

or, for the tunnel conductivity:

$$\begin{aligned} \sigma^{(0)}(\mathbf{F}) &= \lim_{\omega \rightarrow 0} \frac{d}{d\omega} \frac{e^2}{m^2 \hbar} |\mathbf{p}^{\text{cv}}(\mathbf{o})|^2 \int \frac{d^3 \mathbf{k}}{4\pi^3} \\ &\times [f_{\nu}(\mathbf{k}) - f_{\text{c}}(\mathbf{k})] \int_{-\infty}^{\infty} dt \\ &\times \exp\left\{i \int_0^t d\vartheta [\omega - \omega_{\text{cv}}(\mathbf{k}_g)]\right\}. \end{aligned} \quad (30)$$

In the case of direct semiconductors like GaAs the direct gap $E_g^d(\mathbf{k}) = \hbar\omega_{\text{cv}}(\mathbf{k})$ has a distinct minimum at

$\mathbf{k} = 0$. With increasing \mathbf{k} the exponential in (30) starts to oscillate rapidly and effectively cuts the integration. Therefore, it is reasonable to use the parabolic band approximation for those materials. We should dwell on this case for a moment in order to point out the essential differences with respect to silicon. Let us therefore assume parabolic and isotropic bands in the vicinity of $\mathbf{k} = 0$. In that case we can write for the last equation taking the Fourier transform:

$$\begin{aligned} \sigma^{(0)}(\mathbf{F}) = & \lim_{\omega \rightarrow 0} \frac{d}{d\omega} \frac{e^2}{m^2 \hbar} |\mathbf{p}^{cv}(\mathbf{o})|^2 \int \frac{d^2 \mathbf{k}_\perp}{4\pi^3} \int_{-\infty}^{\infty} dk_\parallel \\ & \times [f_v(\mathbf{k}) - f_c(\mathbf{k})] \times \int_{-\infty}^{\infty} dt \\ & \times \exp \left\{ -i \left[t \frac{E_g^d + E_\perp^d - \hbar\omega}{\hbar\Theta_r} \right. \right. \\ & \left. \left. + \frac{1}{3} \left(t - \frac{k_\parallel \hbar\Theta_r}{eF} \right)^3 + \frac{1}{3} \left(\frac{k_\parallel \hbar\Theta_r}{eF} \right)^3 \right] \right\}, \quad (31) \end{aligned}$$

where $\Theta_r = (e^2 F^2 / 2\mu_r \hbar)^{1/3}$ is the electrooptical frequency for the direct transition ($\mu_r =$ reduced effective mass), $E_\perp^d = \hbar^2 k_\perp^2 / 2\mu_r$ the transverse part of the kinetic energy, and k_\parallel the \mathbf{k} -component parallel to the field direction. The integration variable k_\parallel is scaled by a factor $eF/\hbar\Theta_r$ in the third-power terms of the exponential in (30), which cuts the k_\parallel -integration due to rapid oscillations of the integrand just for very small k_\parallel -values. The last assertion is a consequence of the general constraint to the field strength:

$$\frac{eFa}{\hbar\Theta_r} \ll 1, \quad (32)$$

which means that the characteristic energy of the interband coupling must remain small compared to the characteristic energy $\hbar\Theta_r$ of the intraband coupling. The used Houston-type functions are based on this condition.

As a consequence of (32) we may take the distribution functions $f_\mu(\mathbf{k})$ at $k_\parallel = 0$ and evaluate the k_\parallel - and t -integral:

$$\begin{aligned} \sigma^{(0)}(\mathbf{F}) = & \lim_{\omega \rightarrow 0} \frac{d}{d\omega} \frac{e^3 F}{\pi m^2 \hbar^2 \Theta_r^2} |\mathbf{p}^{cv}(\mathbf{o})|^2 \int d^2 \mathbf{k}_\perp \\ & \times [f_v(\mathbf{k}_\perp) - f_c(\mathbf{k}_\perp)] Ai^2 \left(\frac{E_g^d + E_\perp^d - \hbar\omega}{\hbar\Theta_r} \right). \quad (33) \end{aligned}$$

We find by partial integration:

$$\begin{aligned} \sigma^{(0)}(\mathbf{F}) = & \sigma_0 \sqrt{\frac{E_g^d}{\hbar\Theta_r}} \left\{ [f_v(0) - f_c(0)] Ai^2 \left(\frac{E_g^d}{\hbar\Theta_r} \right) \right. \\ & \left. + \int_0^\infty d\epsilon Ai^2 \left(\frac{E_g^d + \epsilon}{\hbar\Theta_r} \right) \frac{d}{d\epsilon} [f_v(\epsilon) - f_c(\epsilon)] \right\}, \quad (34) \end{aligned}$$

with

$$\sigma_0 = \gamma_{cv} \frac{e^2 \sqrt{2\mu_r^3 E_g^d}}{m\hbar^2}. \quad (35)$$

In (35) γ_{cv} stands for the oscillator strength:

$$\gamma_{cv} = 2 \frac{|\mathbf{p}^{cv}(\mathbf{o})|^2}{mE_g^d}. \quad (36)$$

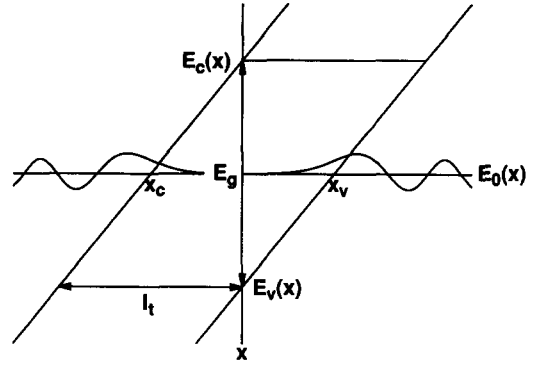


Fig. 2. Tunneling length $l_t = |x_v - x_c|$ and band diagram, assuming a constant electric field over l_t . The maximum overlap of the envelope wave functions determines the transition energy at a given position x in the device.

The difference of the Fermi functions $f_v - f_c$ accounts for the tunneling current direction. In thermodynamic equilibrium ($E_{F,n} = E_{F,p} = E_F$) we have:

$$\begin{aligned} E_c - \mu & \rightarrow E_c - e\varphi(x_c) - E_F \\ E_v - \mu & \rightarrow E_v - e\varphi(x_v) - E_F, \quad (37) \end{aligned}$$

and, since the electric field is assumed constant throughout the calculation:

$$|x_c - x_v| = l_t = \frac{E_g}{e|F|}, \quad (38)$$

is the tunneling length, which has to be overcome by the electrons (see Fig. 2).

The second term in eqn (34) is only a small correction and will be neglected. To obtain the tunneling current density, one has to integrate:

$$j_t^{(0)} = \int_0^F dF' \sigma^{(0)}(F'). \quad (39)$$

We find for the direct (zero-phonon) transitions:

$$\begin{aligned} j_t^{(0)} = & 3\gamma_{cv} \frac{e\mu_r^2 \Theta_r^2}{\hbar m} \left[Ai'^2 \left(\frac{E_g^d}{\hbar\Theta_r} \right) \right. \\ & \left. - \frac{E_g^d}{\hbar\Theta_r} Ai^2 \left(\frac{E_g^d}{\hbar\Theta_r} \right) \right] (f_v - f_c). \quad (40) \end{aligned}$$

This result is in agreement with that of a ‘‘Golden Rule’’ calculation[41], apart from a prefactor of the order 1.

In the case of silicon the situation is far more complicated. If we look at the band structure in Fig. 1, we notice that in [111]-direction the lowest conduction band and the uppermost valence band are almost parallel ($E_g^d(L) = 3.46$ eV, $E_g^d(\Gamma) = 3.42$ eV [40]). Therefore, the reduced effective mass of these bands is almost zero at Γ and direct tunneling in [111]-direction at the center of the Brillouin zone will involve the ‘‘ L_3 ’’-conduction band. On the other hand, a local minimum of the lowest conduction band at Γ exists in [110]-direction (right boundary of Fig. 1). Other direct transitions are possible at Δ_1 in [100]-direction and at L . Obviously, it is neither

possible to describe direct tunneling in silicon with a two-band model nor the anisotropy can be explained satisfactory. For the purpose of comparison with indirect, phonon-assisted tunneling we should consider the transition at Γ with effective masses such that the tunneling probability becomes maximum, i.e. with the light hole mass and an electron mass of the order 0.1 m. Nevertheless, we suppose indirect tunneling to be more probable, which then justifies the crude assumptions.

2.3. Indirect (phonon-assisted) transitions

The indirect gap of silicon has an absolute minimum of $E_g^i = E_c(\Delta_1) - E_v(\Gamma_{25'}) = 1.12$ eV at 300 K. This determines the intermediate states to have the same k -values. The two transition paths with intermediate states of $E = E_c(\Gamma_{15})$ and $E = E_v(\Delta_1)$ are illustrated in Fig. 1.

The correlation function eqn (22) now reads:

$$\mathbf{K}_{\mu\nu}^{(2),\text{ind}}(\mathbf{k}', \mathbf{k}; t) = \frac{1}{\hbar^2} \int_0^{-t} dt_1 \int_0^{-t_1} dt_2 \langle [a_{\mu\mathbf{k}'}^+ a_{\nu\mathbf{k}'}] \times H_{\text{int}}(t_1) \tilde{a}_{\mu\mathbf{k}}^+(-t) \tilde{a}_{\nu\mathbf{k}}(-t) H_{\text{int}}(t_2)]_- \rangle_{\infty}. \quad (41)$$

For simplicity we only insert the electron-phonon collisions here, i.e. the part with $\mu = c$ of (8):

$$\mathbf{K}_{\mu\nu}^{(2),\text{ind}}(\mathbf{k}', \mathbf{k}; t) = \frac{1}{\hbar^2} \int \frac{d^3\mathbf{k}}{4\pi^3} \int \frac{d^3\mathbf{k}'}{4\pi^3} \times \int d^3\mathbf{q} \int d^3\mathbf{q}' V(\mathbf{q}) V(\mathbf{q}') \left\langle \int_0^{-t} dt_1 \int_0^{-t_1} dt_2 \times [a_{\mu\mathbf{k}'}^+ a_{\nu\mathbf{k}'}] \tilde{a}_{c,\mathbf{k}+\mathbf{q}}^+(t_1) \tilde{a}_{c,\mathbf{k}}(t_1) \tilde{a}_{\mu\mathbf{k}}^+(-t) \times \tilde{a}_{\nu\mathbf{k}}(-t) \tilde{a}_{c,\mathbf{k}'+\mathbf{q}'}^+(t_2) \tilde{a}_{c,\mathbf{k}'}(t_2)]_- \times [b_{\mathbf{q}}(t_1) - b_{-\mathbf{q}}^+(t_1)] [b_{\mathbf{q}}(t_2) - b_{-\mathbf{q}}^+(t_2)] \right\rangle_{\infty}. \quad (42)$$

Finally we will generalize the result including also the hole-phonon collisions. If we again use Houston-type solutions [eqn (23)], it follows that:

$$\mathbf{K}_{\mu\nu}^{(2),\text{ind}}[\mathbf{k}', \mathbf{k}; t) = -\frac{1}{\hbar^2} \int \frac{d^3\mathbf{k}}{4\pi^3} \int \frac{d^3\mathbf{k}'}{4\pi^3} \times \int d^3\mathbf{q} V(-\mathbf{q}) V(\mathbf{q}) \int_0^{-t} dt_1 \int_0^{-t_1} dt_2 \times \exp\left\{ \frac{i}{\hbar} \int_0^{-t_1} d\vartheta [E_{\mu}(\mathbf{k}_g) - E_{\nu}(\mathbf{k}_g)] \right\} \times \exp\left\{ \frac{i}{\hbar} \int_0^{t_1} d\vartheta [E_c(\mathbf{k}_g - \mathbf{q}) - E_c(\mathbf{k}_g)] \right\} \times \exp\left\{ \frac{i}{\hbar} \int_0^{t_2} d\vartheta [E_c(\mathbf{k}'_g + \mathbf{q}) - E_c(\mathbf{k}'_g)] \right\} \times \langle [a_{\mu\mathbf{k}'}^+ a_{\nu\mathbf{k}'}] \tilde{a}_{c,\mathbf{k}_1-\mathbf{q}}^+ a_{c,\mathbf{k}_1} \tilde{a}_{\mu\mathbf{k}-\mathbf{q}}^+ a_{\nu\mathbf{k}-\mathbf{q}} \tilde{a}_{c,\mathbf{k}_2+\mathbf{q}}^+ a_{c,\mathbf{k}_2}]_- \rangle \times \{ N_{\mathbf{q}} \exp[i\omega_{\mathbf{q}}(t_1 - t_2)] + (N_{\mathbf{q}} + 1) \times \exp[-i\omega_{\mathbf{q}}(t_1 - t_2)] \}. \quad (43)$$

The last expression was derived by carrying out the average over the phonon ensemble after the time dependence of the boson operators had been inserted:

$$\langle b_{\mathbf{q}}^+(t_1) b_{\mathbf{q}}(t_2) \rangle_{\infty} = \langle b_{\mathbf{q}}^+ b_{\mathbf{q}} \rangle \exp(i\omega_{\mathbf{q}} t_1 - i\omega_{\mathbf{q}} t_2) = N_{\mathbf{q}} \delta_{\mathbf{q}\mathbf{q}'} \exp[i\omega_{\mathbf{q}}(t_1 - t_2)]. \quad (44)$$

$N_{\mathbf{q}}$ is the Bose function. The electron-phonon coupling constants are assumed to satisfy the relation $V(-\mathbf{q}) = V^*(\mathbf{q})$. For further simplification we neglect the phonon dispersion: $\omega_{\mathbf{q}} = \omega_0 = \text{const}$. This is not a serious restriction, since the fundamental indirect gap increases rapidly as the momentum difference, which must be provided by the phonon, deviates from $k_0 \approx 0.85 \pi/a$. Hence, the dispersion of the indirect gap effectively extracts only those phonons with $\mathbf{q} \approx \mathbf{k}_0$ (in [100]-direction!). Consequently we replace \mathbf{q} by \mathbf{k}_0 in the matrix element $V(\mathbf{q})$.

Calculating the commutator in eqn (43) yields a combination of three-particle Green's functions, which have to be decoupled by Random Phase Approximation (RPA). Instead of presenting the lengthy and combersome calculation, we only note that RPA gives the following product of Kroneckers:

$$\langle [\dots]_- \rangle \sim \delta_{\mathbf{k}_1, \mathbf{k}_2 + \mathbf{q}} \delta_{\mathbf{k}-\mathbf{q}, \mathbf{k}'} \delta_{\mathbf{k}_1, \mathbf{k}-\mathbf{q}} \delta_{\mu\nu} \delta_{\mu\nu}. \quad (45)$$

We adapt the result which was already derived in Ref. [42] (for $T = 0$ and the isotropic case there) to $L(\omega, \mathbf{F})$ defined by eqn (11):

$$L(\omega, \mathbf{F}) = \frac{e^2 |p^{\text{cv}}(0)|^2 |V(\mathbf{k}_0)|^2}{m^2 \hbar^3} \times \int \frac{d^3\mathbf{k}}{4\pi^3} \int d^3\mathbf{q} \int_{-\infty}^{\infty} dt \tau^*(\mathbf{k}, \mathbf{F}) \tau(\mathbf{k}, \mathbf{F}) \times \left[N_0 \exp\left\{ -i \int_0^t d\vartheta [\omega_c(\mathbf{k}_g, \mathbf{q}) - \omega \mp \omega_0] \right\} + (N_0 + 1) \exp\left\{ -i \int_0^t d\vartheta [\omega_c(\mathbf{k}_g, \mathbf{q}) - \omega \pm \omega_0] \right\} \right] \times [f_{\nu}(\mathbf{k}) - f_c(\mathbf{k} + \mathbf{q})] \quad (46)$$

with

$$\hbar\omega_c(\mathbf{k}_g, \mathbf{q}) = E_c(\mathbf{k}_g + \mathbf{q}) - E_v(\mathbf{k}). \quad (47)$$

The time

$$\tau(\mathbf{k}, \mathbf{F}) = i \int_0^{\infty} dt \exp\left\{ -i \int_0^t d\vartheta [\omega_{\text{cv}}(\mathbf{k}_g) - \omega] \right\}, \quad (48)$$

may be interpreted as the mean lifetime of the intermediate state characterized as a direct electron-hole pair. We suppose for the direct gap that $\hbar\Theta_{\mu} \ll E_g^d$. Then:

$$\tau(\mathbf{k}, \mathbf{F}) \approx \tau(\mathbf{k}, \mathbf{0}) = \frac{1}{\omega_{\text{cv}}(\mathbf{k}) - \omega - i\delta}, \quad (49)$$

can be approximated in the limit $\omega \rightarrow 0$ by:

$$\tau(\mathbf{k}, \mathbf{F}) \approx \frac{\hbar}{E_g^d}. \quad (50)$$

The expression for $L(\omega, \mathbf{F})$ contains different signs in front of the phonon frequency ω_0 . Kane[35]

discovered that the upper sign has to be applied in the case of reverse biased junctions (generation), whereas the lower sign holds for forward biased junctions (recombination) in order to account for energy conservation in the electron–phonon system.

The evaluation of the integrals in (46) requires the coordinate system to be specified properly. We choose for the z -axis the direction of the electric field vector, the corresponding components are labeled with the subscript “ \parallel ”. The major contributions to the tunneling conductivity arise from the local band extrema. Therefore, we may use the parabolic band approximation both for $E_c(\mathbf{k}_g + \mathbf{q})$ and $E_v(\mathbf{k})$ in (47). The iso-energy surface in the vicinity of the valence band maximum will be assumed isotropic, although both the light hole and the heavy hole bands are warped. This assumption is justified by the general uncertainty concerning the right hole tunneling mass to be used in the calculation. A strong electric field responsible for band-to-band tunneling leakage surely will remove the band degeneracy at $\mathbf{k} = 0$ and change the effective masses. Since it is nearly impossible to treat this effect quantitatively (an attempt was made in Ref. [43], where explicit expressions were found for the case of vanishing warping parameter C), we will concentrate on the anisotropy of $E_c(\mathbf{k}_g + \mathbf{q})$. We allow for an arbitrary angle between electric field vector and valley vector \mathbf{k}_0 . Because the complete conduction band energy surface consists of six valleys, one has to sum the tunneling conductivity over the six different \mathbf{k}_0 -vectors finally.

In Appendix A we derive the representation of $\hbar\omega_i(\mathbf{k}_g, \mathbf{q})$ in the properly chosen coordinate system. The result is:

$$\begin{aligned} \hbar\omega_i(\mathbf{k}_g, \mathbf{q}) = & E_g^i + \frac{\hbar^2}{2m_{\parallel}} [k_{g,\parallel} + q_{\parallel} - k_{0\parallel} + \kappa(k_{\perp 2} \\ & - k_{0\perp})]^2 + \frac{\hbar^2}{2m_v} k_{g,\parallel}^2 + \frac{\hbar^2}{2m_t} (k_{\perp 1} + q_{\perp 1})^2 \\ & + \frac{\hbar^2}{2\tilde{m}_{\perp 2}} (k_{\perp 2} + q_{\perp 2} - k_{0\perp})^2 \\ & + \frac{\hbar^2}{2m_v} (k_{\perp 1}^2 + k_{\perp 2}^2). \end{aligned} \quad (51)$$

Now we generalize the result including also hole–phonon collisions. Taking into account hole–phonon scattering instead of electron–phonon collisions leads basically to the same expression as (46), but with other matrix elements and another direct gap. We get for the quantity $L(\omega, \mathbf{F})$, which determines the tunneling conductivity:

$$\begin{aligned} L(\omega, \mathbf{F}) = & \frac{e^2}{m^2\hbar} \left(|V_c(\mathbf{k}_0)|^2 \left| \frac{p^{cv}(0)}{E_g^d(0)} \right|^2 \right. \\ & + |V_p(\mathbf{k}_0)|^2 \left| \frac{p^{cv}(\mathbf{k}_0)}{E_g^d(\mathbf{k}_0)} \right|^2 \left. \right) \int \frac{d^2\mathbf{k}_{\perp}}{4\pi^3} \int d^3\mathbf{q} \\ & \times [\bar{N}_0 Q_{\mp}(\mathbf{k}_{\perp}, \mathbf{q}) + (\bar{N}_0 + 1) Q_{\pm}(\mathbf{k}_{\perp}, \mathbf{q})] \\ & \times [f_v(\mathbf{k}_{\perp}) - f_c(k_{0\parallel}, (\mathbf{k} + \mathbf{k}_0)_{\perp})] \end{aligned} \quad (52)$$

with

$$\begin{aligned} Q_{\mp}(\mathbf{k}_{\perp}, \mathbf{q}) = & \int_{-\infty}^{\infty} dk_{\parallel} \int_{-\infty}^{\infty} dt \\ & \exp \left\{ -\frac{i}{\hbar} \int_0^t d\vartheta [\hbar\omega_i(\mathbf{k}_g, \mathbf{q}) - \hbar\omega \mp \hbar\omega_0] \right\}. \end{aligned} \quad (53)$$

As in the case of direct transitions we have fixed the distribution functions at $k_{\parallel} = 0$ in (52). The precondition for this (relation (32)) now reads $eFa \ll \hbar\Theta_{r,\parallel}$, which means that a field applied in [111]-direction must satisfy $F \ll 2.4 \times 10^7 \text{ V cm}^{-1}$. This is more than the breakdown field strength ($F_c = 1.9 \times 10^7 \text{ V cm}^{-1}$ in [111]-direction), but has the same order of magnitude. Furthermore, as in the electron–phonon matrix element we replaced \mathbf{q} by \mathbf{k}_0 also in the distribution function.

With eqn (51) we are able to evaluate $Q_{\mp}(\mathbf{k}_{\perp}, \mathbf{q})$ exactly. This is demonstrated in Appendix B. We find:

$$\begin{aligned} Q_{\mp}(\mathbf{k}_{\perp}, \mathbf{q}) = & (2\pi)^2 \frac{eF}{\hbar\Theta_{r,\parallel}^2} Ai^2 \left(\frac{\hbar^2 \tilde{q}_{\parallel}^2}{2m_{\Sigma}} + C_{\mp}^{\mp} \right) \\ & \times \exp \left(-i \frac{\hbar^2 \tilde{q}_{\parallel}^3}{2\mu_3 eF} \right), \end{aligned} \quad (54)$$

where Ai denotes the Airy function, and the various quantities have the following meaning:

$$\hbar\Theta_{r,\parallel} = \left[\frac{e^2 \hbar^2 F^2}{2\mu_{\parallel}} \right]^{1/3}, \quad (55)$$

$$\tilde{q}_{\parallel} = q_{\parallel} - k_{0\parallel} + \kappa(k_{\perp 2} - k_{0\perp}), \quad (56)$$

$$\begin{aligned} C_{\mp}^{\mp} = & \frac{\hbar^2}{2m_t} (k_{\perp 1} + q_{\perp 1})^2 + \frac{\hbar^2}{2\tilde{m}_{\perp 2}} (k_{\perp 2} + q_{\perp 2} - k_{0\perp})^2 \\ & + \frac{\hbar^2}{2m_v} (k_{\perp 1}^2 + k_{\perp 2}^2) - \hbar\omega \mp \hbar\omega_0 + E_g^i. \end{aligned} \quad (57)$$

The masses and related parameters are given by:

$$\kappa = \frac{(m_t - m_l) \cos \vartheta \sin \vartheta}{m_l \sin^2 \vartheta + m_t \cos^2 \vartheta}, \quad (58)$$

$$\frac{1}{m_{\parallel}} = \frac{1}{m_t} \sin^2 \vartheta + \frac{1}{m_l} \cos^2 \vartheta, \quad (59)$$

$$\frac{1}{\tilde{m}_{\perp 2}} = \frac{1}{m_t} \cos^2 \vartheta + \frac{1}{m_l} \sin^2 \vartheta, \quad (60)$$

$$\begin{aligned} \frac{1}{\mu_{\parallel}} = & \frac{1}{m_{\parallel}} + \frac{1}{m_v}, \quad \frac{1}{\mu_3} = \frac{1}{3\mu_{\parallel}} + \frac{1}{m_{\Sigma}}, \\ & m_{\Sigma} = m_{\parallel} + m_v. \end{aligned} \quad (61)$$

The angle ϑ is the angle between field strength vector \mathbf{F} and the vector \mathbf{k}_0 pointing to the valley under consideration. The longitudinal and transverse effective masses have the values [44] $m_l = 0.9163(4) m$ and $m_t = 0.1905(1) m$.

Inserting $Q_{\mp}(\mathbf{k}_{\perp}, \mathbf{q})$ into the tunneling conductivity $L(\omega, \mathbf{F})$, we are left with five integrations. First, we solve the q_{\parallel} -integral. With the new integration variable $z = \hbar \tilde{q}_{\parallel}^2 / 2m_{\Sigma} \Theta_{r,\parallel}$ we have

$$\int_{-\infty}^{\infty} dq_{\parallel} Q_{\mp}(\mathbf{k}_{\perp}, \mathbf{q}) = (2\pi)^2 \frac{eF\sqrt{2m_{\Sigma}}}{(\hbar\Theta_{r,\parallel})^{3/2}} \int_0^{\infty} \frac{dz}{\sqrt{z}} \times \text{Ai}^2\left(z + \frac{C_{\perp}^{\mp}}{\hbar\Theta_{r,\parallel}}\right) \left[\exp\left(-i \frac{m_{\Sigma}^{3/2}}{\mu_3 \sqrt{\mu_4}} z^{3/2}\right) + \text{c.c.} \right]. \quad (62)$$

The quantity C_{\perp}^{\mp} is positive in the limit $\omega \rightarrow 0$ and of the order E_g^i . As long as the electric field is not too strong, $E_g^i/\hbar\Theta_{r,\parallel}$ remains large enough in order to ensure a rapid exponential decay of Ai^2 with z . Therefore, together with $z^{-1/2}$ the integrand shows a sharp maximum at $z = 0$, which justifies a Taylor expansion of the exponentials and the restriction to the leading term. After [45]:

$$\int_0^{\infty} \frac{d\xi}{\sqrt{\xi}} \text{Ai}^2(\xi + x) = \frac{1}{2} \text{Ai}_1(\kappa x), \quad \kappa = 2^{2/3}, \quad (63)$$

and thus:

$$\int_{-\infty}^{\infty} dq_{\parallel} Q_{\mp}(\mathbf{k}_{\perp}, \mathbf{q}) = (2\pi)^2 \frac{eF\sqrt{2m_{\Sigma}}}{(\hbar\Theta_{r,\parallel})^{3/2}} \text{Ai}_1\left(\kappa \frac{C_{\perp}^{\mp}}{\hbar\Theta_{r,\parallel}}\right). \quad (64)$$

Ai_1 denotes the integral of the Airy function. The remaining integrations are straightforward, if the distribution functions are taken in front of the integrals at $\mathbf{k}_{\perp} = 0$. This can be justified by essentially the same argument as before. Then, we have to consider:

$$I = \int d\mathbf{q}_{\perp} \int d\mathbf{k}_{\perp} \text{Ai}_1\left(\kappa \frac{E_g^i - \hbar\omega \mp \hbar\omega_0 + \frac{\hbar^2}{2m_1}(k_{\perp 1} + q_{\perp 1})^2 + \frac{\hbar^2}{2\tilde{m}_{12}}(k_{\perp 2} + q_{\perp 2} - k_{0\perp})^2 + \frac{\hbar^2}{2m_v}(k_{\perp 1}^2 + k_{\perp 2}^2)}{\hbar\Theta_{r,\parallel}}\right). \quad (65)$$

The $q_{\perp 1}, q_{\perp 2}$ -integrations can be shifted by $k_{\perp 1}$ and $k_{\perp 2} - k_{0\perp}$, respectively. Introducing four new variables[46]:

$$\rho_1 = \frac{\hbar}{\sqrt{2m_1}} q_{\perp 1}, \quad \rho_2 = \frac{\hbar}{\sqrt{2\tilde{m}_{12}}} q_{\perp 2}, \\ \rho_3 = \frac{\hbar}{\sqrt{2m_v}} k_{\perp 1}, \quad \rho_4 = \frac{\hbar}{\sqrt{2m_v}} k_{\perp 2}, \quad (66)$$

we get:

$$I = \frac{4m_v\sqrt{m_1\tilde{m}_{12}}}{\hbar^4} \int d\rho_1 d\rho_2 d\rho_3 d\rho_4 \times \text{Ai}_1\left(\kappa \frac{E_g^i - \hbar\omega \mp \hbar\omega_0 + \rho^2}{\hbar\Theta_{r,\parallel}}\right), \quad (67)$$

where ρ is the radius in the four-dimensional space. Using:

$$d\rho_1 d\rho_2 d\rho_3 d\rho_4 = 2\pi^2 \rho^3 d\rho, \quad (68)$$

we find:

$$I = 8\pi^2 \frac{m_v\sqrt{m_1\tilde{m}_{12}}}{\hbar^4} \int_0^{\infty} d\rho \rho^3 \times \text{Ai}_1\left(\kappa \frac{E_g^i - \hbar\omega \mp \hbar\omega_0 + \rho^2}{\hbar\Theta_{r,\parallel}}\right) \\ = 4\pi^2 \frac{m_v\sqrt{m_1\tilde{m}_{12}}(\hbar\Theta_{r,\parallel})^2}{\hbar^4 \kappa^2} \int_0^{\infty} dt t \times \text{Ai}_1\left(t + \kappa \frac{E_g^i - \hbar\omega \mp \hbar\omega_0}{\hbar\Theta_{r,\parallel}}\right). \quad (69)$$

With Aspnæs[45] one can show that:

$$I = \int_0^{\infty} dt t \text{Ai}_1(t + x) \\ = \frac{1}{2} [\text{Ai}(x) + x \text{Ai}'(x) + x^2 \text{Ai}_1(x)]. \quad (70)$$

The first derivative of this expression with respect to x , which we need for the tunneling conductivity, turns out to be:

$$\frac{dI}{dx} = \text{Ai}'(x) + x \text{Ai}_1(x). \quad (71)$$

Inserting all results into the expression (52) for $L(\omega, \mathbf{F})$, taking the derivative with respect to ω and finally the limit $\omega \rightarrow 0$, we obtain for the differential tunneling conductivity:

$$\sigma^{\text{ind}}(\mathbf{F}) = \frac{8\pi e^2}{\hbar^5} \left(|V_c(\mathbf{k}_0)|^2 \left| \frac{p^{\text{cv}}(0)}{E_g^d(0)} \right|^2 + |V_p(\mathbf{k}_0)|^2 \left| \frac{p^{\text{cv}}(\mathbf{k}_0)}{E_g^d(\mathbf{k}_0)} \right|^2 \right) \frac{\sqrt{\tilde{m}_{12} m_{\parallel} m_1 m_v^{3/2}}}{m^2}$$

$$\times \left\{ (E_g^i \mp \hbar\omega_0) N_0 \left[\frac{\text{Ai}'(x^{\mp})}{x^{\mp}} + \text{Ai}_1(x^{\mp}) \right] + (E_g^i \pm \hbar\omega_0) (N_0 + 1) \left[\frac{\text{Ai}'(x^{\pm})}{x^{\pm}} + \text{Ai}_1(x^{\pm}) \right] \right\} [f_c(0) - f_c(\mathbf{k}_0)] \quad (72)$$

with:

$$x^{\pm} \equiv \kappa \frac{E_g^i \pm \hbar\omega_0}{\hbar\Theta_{r,\parallel}}, \quad \kappa = 2^{2/3}. \quad (73)$$

The tunneling current density follows from integrating the differential conductivity over the field strength:

$$j_i^{\text{ind}} = \int_0^F dF' \sigma[x(F')]. \quad (74)$$

We are left with the problem:

$$J = \int_0^F dF' \left[\frac{\text{Ai}'[x(F')]}{x(F')} + \text{Ai}_1[x(F')] \right], \quad (75)$$

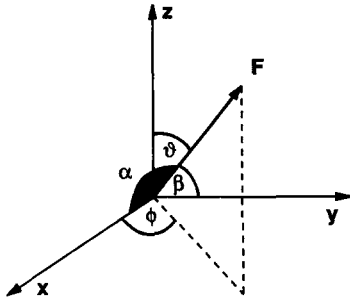


Fig. 3. Spherical coordinates of the field vector in the [100]-system. The angles α , β and ϑ are the angles of intersection with the valleys on the x -, y -, and z -axis, respectively.

which cannot be solved exactly. With a new integration variable we find:

$$J = \frac{3}{2}x^{3/2}F(x) \int_{F(x)}^{\infty} d\xi \xi^{-7/2} [Ai'(\xi) + \xi Ai_1(\xi)], \quad (76)$$

and with (70) and (71):

$$J \approx \frac{3}{4}F \left[\frac{Ai(x)}{x^2} + \frac{Ai'(x)}{x} + Ai_1(x) \right]. \quad (77)$$

So far we have considered transitions into one particular valley, where the field vector \mathbf{F} and the valley vector \mathbf{k}_0 encircle the angle ϑ , which determines the effective mass m_{\parallel} [eqn (59)] in field direction. The same angle occurs in the opposite valley ($-\mathbf{k}_0$) giving a factor 2. Both valleys are equivalent, since the electron-phonon interaction was assumed isotropic. The remaining two pairs of valleys are cut by the field vector with angles α (x -axis) and β (y -axis). An illustration is given in Fig. 3. In spherical coordinates of the $\{x, y, z\}$ -system defined by the crystal orientation these angles are given by:

$$\begin{aligned} \cos \alpha &= \cos \varphi \sin \vartheta, \\ \sin \alpha &= \sin \varphi \sin \vartheta. \end{aligned} \quad (78)$$

Consequently, we get the following cartesian representations of the effective mass m_{\parallel} :

$$\begin{aligned} m_{\parallel}^x &= \frac{m_t m_1}{m_1 - (m_1 - m_t) \cos^2 \varphi \sin^2 \vartheta} \\ &= \frac{m_t m_1}{m_1 - (m_1 - m_t) F_x^2 / F^2}, \\ m_{\parallel}^y &= \frac{m_t m_1}{m_1 - (m_1 - m_t) \sin^2 \varphi \sin^2 \vartheta} \\ &= \frac{m_t m_1}{m_1 - (m_1 - m_t) F_y^2 / F^2}, \\ m_{\parallel}^z &= \frac{m_t m_1}{m_1 - (m_1 - m_t) \cos^2 \vartheta} \\ &= \frac{m_t m_1}{m_1 - (m_1 - m_t) F_z^2 / F^2}. \end{aligned} \quad (79)$$

The corresponding values for the transverse mass $\tilde{m}_{\perp 2}^{\alpha}$ follow immediately from m_{\parallel}^{α} , if m_t and m_l are exchanged. The phonon-assisted tunneling current density now takes the form:

$$\begin{aligned} j_t^{\text{ind}} &= \frac{12\pi e^2}{\hbar^5} \left(|V_c(\mathbf{k}_0)|^2 \left| \frac{p^{cv}(0)}{E_g^d(0)} \right|^2 \right. \\ &\quad \left. + |V_p(\mathbf{k}_0)|^2 \left| \frac{p^{cv}(\mathbf{k}_0)}{E_g^d(\mathbf{k}_0)} \right|^2 \right) \frac{\sqrt{m_t m_1^3}}{m^2} F \\ &\quad \times \sum_{\alpha=x,y,z} \sqrt{\tilde{m}_{\perp 2}^{\alpha} m_{\parallel}^{\alpha}} [(E_g^i \mp \hbar\omega_0) N_0 H(x_{\alpha}^{\mp}) \\ &\quad + (E_g^i \pm \hbar\omega_0) (N_0 + 1) H(x_{\alpha}^{\pm})] \\ &\quad \times [f_v(0) - f_c(\mathbf{k}_0)], \end{aligned} \quad (80)$$

with:

$$H(x) = \frac{Ai(x)}{x^2} + \frac{Ai'(x)}{x} + Ai_1(x), \quad (81)$$

$$x^{\pm} = 2^{2/3} \frac{E_g^i \pm \hbar\omega_0}{\hbar\Theta_{\tau,\parallel}^{\alpha}}, \quad (82)$$

$$\hbar\Theta_{\tau,\parallel}^{\alpha} = \left(\frac{e^2 \hbar^2 F^2}{2\mu_{\parallel}^{\alpha}} \right)^{1/3}, \quad \frac{1}{\mu_{\parallel}^{\alpha}} = \frac{1}{m_{\parallel}^{\alpha}} + \frac{1}{m_v}. \quad (83)$$

The electron tunneling masses for the three cartesian directions of the crystal are given by:

$$\begin{aligned} m_{\parallel}^{\alpha} &= \frac{m_t m_1}{m_1 - (m_1 - m_t) F_x^2 / F^2}, \\ \tilde{m}_{\perp 2}^{\alpha} &= \frac{m_t m_1}{m_1 - (m_1 - m_t) F_x^2 / F^2}. \end{aligned} \quad (84)$$

Equations (80)–(84) represent the final result of a rigorous calculation of the current density resulting from indirect band-to-band tunneling in silicon. This general result is not based on the WKB method and includes finite temperatures as well as the anisotropy of the conduction band valleys.

Now we discuss briefly, how inhomogeneous electric fields can be included. The effect of a field variation over the tunneling length l_t becomes increasingly important for low reverse voltages and forward bias tunneling. In order to preserve the advantages of the EMA treatment, an analytical interpolation of the wave functions between band edge and semi-classical regions was developed in Ref. [39]. As a result, the argument of the Airy functions is not simply the ratio of band gap and electro-optical energy, but contains the action integral between the classical turning points.

For simplicity we assume the direction of the electric field to be parallel to the x -axis and consider only tunnel *generation*. Then, the essential change in eqn (80), which allows for inhomogeneous fields, is the replacement:

$$x^{\pm} = 2^{2/3} \frac{E_g^i \pm \hbar\omega_0}{\hbar\Theta_{\tau,\parallel}^{\alpha}} \rightarrow [3|S_c^{\pm}(x) - S_v(x)|]^{2/3}, \quad (85)$$

where $S(x)$ are action integrals defined by:

$$S_c^\pm(x) = \frac{\sqrt{2m_\parallel^*}}{\hbar} \int_{x_c^\pm}^x dx' \sqrt{e|\varphi(x') - \varphi(x_c^\pm)|}, \quad (86)$$

$$S_v(x) = \frac{\sqrt{2m_v}}{\hbar} \int_{x_v}^x dx' \sqrt{e|\varphi(x') - \varphi(x_v)|}. \quad (87)$$

Here φ denotes the electrostatic potential. The point x determines the energy level E_0 , at which the transition occurs, and therefore, the classical turning points by:

$$E_0(x) = E_c \pm \hbar\omega_{TA} - e\varphi(x_c^\pm), \quad (88)$$

$$E_0(x) = E_v - e\varphi(x_v). \quad (89)$$

The connection between x and E_0 can be seen by a qualitative argument, as illustrated in Fig. 2. The tunneling rate is given by the overlap of the envelope wave functions within the forbidden zone. The position of the maximum overlap can be estimated taking into account only the exponentials of the WKB approximation[39]. This yields an implicit relation:

$$(m_\parallel^* + m_v)\varphi(x) = m_\parallel^*\varphi(x_c^\pm) + m_v\varphi(x_v). \quad (90)$$

If we insert here eqns (88) and (89), the relation between x and E_0 becomes:

$$E_0(x) = \frac{m_v E_v(x) + m_\parallel^*[E_c(x) \pm \hbar\omega_{TA}]}{m_\parallel^* + m_v}. \quad (91)$$

With the help of equations (88)–(91) one can easily verify that a constant electric field exactly reproduces the old result according to (85).

3. MODEL FOR DEVICE SIMULATION

3.1. Simplifications

The microscopic model (80) contains four unknown matrix elements and is too cumbersome for an implementation into device simulation programs in that form. Therefore, we will specify all quantities as best as possible and replace the Airy function by asymptotic expressions.

For the momentum matrix element p^{cv} we apply Kane's two-band model[41] giving:

$$|p^{cv}(\kappa)|^2 \approx \frac{m^2 E_g^d(\kappa)}{4 \mu_\kappa}. \quad (92)$$

As mentioned before, the direct reduced effective masses at Γ and Δ_1 are not known. Hence, it also seems to make no sense to include the momentum matrix element (13) in the \mathbf{k} -integration, e.g. in form of Kane's model. The latter was done in recent papers (e.g. [47]) resulting in a very difficult \mathbf{k} -integral even for the easier case of direct tunneling, and it was claimed to having improved the theory essentially in this way. It should be mentioned that the \mathbf{k} -dependence of Kane's two-band model is not appropriate for the situation in silicon.

We took the Bloch factors at $\mathbf{k} = 0$ instead, in accordance with the EMA, which is well fulfilled in homo-junctions. The uncertainty concerning the momentum matrix element is in any case absorbed in the reduced effective mass μ_κ [eqn (92)]. In our simplification procedure we will assume the reduced mass at Δ_1 to be much larger than at Γ , where we use the estimate $\mu_\Gamma = 0.1 m$.

The next quantity to be considered is the matrix element of the linear electron–phonon coupling $V_c(\mathbf{k}_0)$. If we assume intravalley acoustic scattering, it follows that:

$$|V_c(\mathbf{k}_0)|^2 = \frac{\hbar\omega_{\mathbf{k}_0} \mathcal{E}_{ac,e}^2}{2\rho c_s^2}, \quad (93)$$

where $\mathcal{E}_{ac,e}$ is the deformation potential constant for electron scattering, ρ the density of silicon, and c_s the sound velocity[48]. The deformation potential constant $\mathcal{E}_{ac,e}$ was determined by Canali *et al.* [49] by fits in Monte Carlo simulation. They found $\mathcal{E}_{ac,e} = 9$ eV. For the phonon energy $\hbar\omega_0$ we choose the energy of the TA-phonon with $\mathbf{k} = \mathbf{k}_0$ in accordance with the experimental observations by Logan and Chynoweth [15]. Instead of summing over the different phonon branches we multiply by 2, because the TO-phonon assisted transitions were found to contribute equally, whereas the other branches had only little influence [15]. From [50] one can deduce $\hbar\omega_{\mathbf{k}_0}^{TA} = 18.6$ meV. Together with $\rho = 2.32831$ g cm⁻³ and $c_s = 9200$ ms⁻¹ [48] we find:

$$|V_c(\mathbf{k}_0)|^2 \approx 6.1 \times 10^{-25} (\text{eV})^2 \text{cm}^3. \quad (94)$$

It is worth noting that nonpolar optical scattering gives almost the same value [$\approx 7 \times 10^{-25} (\text{eV})^2 \text{cm}^3$].

The proper quantity for device simulation purposes is not the tunneling current density j_t , but the generation rate R_t , which is to implement in the right-hand side of the semiconductor transport equations. The relation between the two quantities reads:

$$j_t = eR_t l_t, \quad (95)$$

where l_t is the tunneling length $l_t = (E_g^i \pm \hbar\omega_0)/eF$ (see Fig. 2). Inserting (92)–(95) into eqn (80), the phonon-assisted band-to-band tunneling rate becomes:

$$R_t^{\text{ind}} = \frac{2.47 \times 10^{21}}{\text{cm}^3 \text{s}} F^2 \sum_{\alpha=x,y,z} \frac{\sqrt{\tilde{m}_{12}^2 m_\parallel^*}}{m} [N_0 H(x_\alpha^\mp) + (N_0 + 1) H(x_\alpha^\pm)] [f_v(0) - f_c(\mathbf{k}_0)], \quad (96)$$

where F is in V cm⁻¹. It is useful to define critical field strengths $F_c^{\alpha\pm}$ by:

$$E_g^i \pm \hbar\omega_{TA} = \left(\frac{3}{4}\right)^{2/3} \hbar\Theta_{r,\parallel}^\alpha, \\ F_c^{\alpha\pm} = \begin{cases} 8.30 \times 10^7 \sqrt{\mu_\parallel^*/m} \text{ V cm}^{-1} & \text{upper sign} \\ 7.90 \times 10^7 \sqrt{\mu_\parallel^*/m} \text{ V cm}^{-1} & \text{lower sign,} \end{cases} \quad (97)$$

with the smallest μ_\parallel^* . In [111]-direction $F_c^{\alpha+} = 2.61 \times 10^7$ V cm⁻¹ and $F_c^{\alpha-} = 2.48 \times 10^7$ V cm⁻¹. The

latter field strength can be considered not only as a breakdown threshold but also as an upper limit of the theory. In practical cases, even in advanced ULSI devices, the field strength remains below 10^7 V cm^{-1} . Hence, it is a good approximation to replace the various Airy functions by their asymptotic forms for large positive arguments[51]:

$$\begin{aligned} \text{Ai}(x) &\xrightarrow{x \rightarrow \infty} \left(1 - \frac{5}{48}x^{-3/2}\right) \frac{x^{-1/4}}{2\sqrt{\pi}} \exp\left(-\frac{2}{3}x^{3/2}\right) \\ \text{Ai}'(x) &\xrightarrow{x \rightarrow \infty} -\left(1 + \frac{7}{48}x^{-3/2}\right) \frac{x^{1/4}}{2\sqrt{\pi}} \exp\left(-\frac{2}{3}x^{3/2}\right) \\ \text{Ai}_1(x) &\xrightarrow{x \rightarrow \infty} \left(1 + \frac{7}{48}x^{-3/2}\right) \frac{x^{-3/4}}{2\sqrt{\pi}} \exp\left(-\frac{2}{3}x^{3/2}\right). \end{aligned} \quad (98)$$

In this way we find for the function $H(x)$ in eqn (96):

$$H(x) \rightarrow \frac{x^{-9/4}}{2\sqrt{\pi}} \exp\left(-\frac{2}{3}x^{3/2}\right). \quad (99)$$

Using (99) is closely related to the WKB approximation. The simplified model can be written now as:

$$\begin{aligned} R_t^{\text{ind}} &= \frac{3.8 \times 10^{20}}{\text{cm}^3 \text{ s}} F^{7/2} \sum_{\alpha=x,y,z} \frac{\sqrt{\tilde{m}_{12}^\alpha m_\parallel^\alpha}}{m} \\ &\times \left[\frac{(F_c^{\alpha\mp})^{-3/2} \exp\left(-\frac{F_c^{\alpha\mp}}{F}\right)}{\exp\left(\frac{\hbar\omega_{TA}}{kT}\right) - 1} \right. \\ &\left. + \frac{(F_c^{\alpha\pm})^{-3/2} \exp\left(-\frac{F_c^{\alpha\pm}}{F}\right)}{1 - \exp\left(-\frac{\hbar\omega_{TA}}{kT}\right)} \right] \\ &\times [f_v(0) - f_c(\mathbf{k}_0)], \end{aligned} \quad (100)$$

with

$$F_c^{\alpha\pm} = \frac{4}{3} \frac{\sqrt{2\mu_\parallel^\alpha}}{e\hbar} (E_g^\pm \pm \hbar\omega_{TA})^{3/2}, \quad (101)$$

and \tilde{m}_{12}^α , m_\parallel^α given in (84). As pointed out by Kane [35], the upper sign refers to tunneling generation (reverse bias, $f_v > f_c$), but the lower to recombination (forward bias, $f_v < f_c$).

There is still one point to be cleared. The band-to-band tunneling rate (100) is completely local, i.e. it only depends on the local field strength $F(\mathbf{x})$. This is a consequence of the constant field approximation (the field was assumed constant over the tunneling length l_t). But it is not obvious, at which position exactly the distribution functions must be evaluated. In other words, if we fix the position \mathbf{x} (in device simulation this is done by the mesh point), we have to prescribe the energy level at which the tunneling transition occurs [eqn (91)]. This level determines the local occupation probability by its distance to the local quasi Fermi level. For constant field eqn (90) takes the form $x/\mu_\parallel^\alpha = x_v/m_\parallel^\alpha + x_c/m_v$. The classical turning points x_c and x_v follow immediately from this equation and the obvious formulae $x_v - x_c = E_g^\pm/eF$ (phonon contribution neglected here). The energy

level $E_0(x)$ of the tunneling transition at x then is given by $E_0(x) = E_c(x) - eF(x - x_c)$. Inserting x_c in the last expression, we get:

$$E_0(x) = E_c(x) - \frac{\mu_\parallel^\alpha}{m_\parallel^\alpha} E_g^\pm = E_v(x) + \frac{\mu_\parallel^\alpha}{m_v} E_g^\pm. \quad (102)$$

Only if the effective masses are equal, the transition occurs in the middle of the gap. The Fermi functions in eqn (100) now take the form:

$$f_{c,v} = \left[\exp\left(\frac{E_0(\mathbf{x}) - E_{F_{n,p}}(\mathbf{x})}{kT}\right) + 1 \right]^{-1}. \quad (103)$$

3.2. Comparison of direct and indirect band-to-band tunneling

In order to compare phonon-assisted band-to-band tunneling with direct tunneling we use also for the direct transitions the estimate (92) for the momentum matrix element and the asymptotic expressions (98) of the Airy functions. Then, the rate of direct tunneling takes the form [see eqn (40)]:

$$R_t^{\text{d}} = \frac{3.88 \times 10^{12}}{\text{cm}^3 \text{ s}} F^3 \exp\left(-\frac{F_c^{\text{d}}}{F}\right) [f_v(0) - f_c(\mathbf{k}_0)], \quad (104)$$

where F_c^{d} denotes the critical field strength for the direct transition at $\mathbf{k} = 0$: $F_c^{\text{d}} = 1.37 \times 10^8 \text{ V cm}^{-1}$. For the reduced mass μ_r we inserted 0.1 m . Now we define the direction of the field vector to be along [111] and consider only the phonon-absorption term in (100). The difference of the Fermi functions is replaced by unity. Then we get for the ratio $R_t^{\text{d}}/R_t^{\text{ind}}$:

$$\begin{aligned} \frac{R_t^{\text{d}}}{R_t^{\text{ind}}} &\approx \frac{3.9 \times 10^{12} F^3 \exp(-1.37 \times 10^8/F)}{7.5 \times 10^8 F^{7/2} \exp(-2.48 \times 10^7/F)} \\ &\approx 5.2 \times 10^3 F^{-1/2} \exp(-1.12 \times 10^8/F). \end{aligned} \quad (105)$$

This result clearly shows that although the prefactors may be of the same order of magnitude, the probability of direct tunneling transitions is negligible because of the large gap energy. This conclusion holds for a much smaller value of the effective mass μ_r than 0.1 m , which, however, is not very likely.

4. DISCUSSION

We discuss the results using for the effective hole mass a value of 0.16 m (average light hole mass), room temperature for T and a step-like distribution function for f . The band-to-band tunneling rate after eqn (100) is shown in Fig. 4 in the field strength range from 1×10^5 to $2 \times 10^6 \text{ V cm}^{-1}$. Here, basically the $1/F$ -behaviour of the exponents in the half-logarithmic plot is reflected.

In Fig. 5 we illustrated the anisotropy effect of the silicon conduction band valleys in a more restricted field strength range from 5×10^5 to $1 \times 10^6 \text{ V cm}^{-1}$. Comparing three major directions of the field vector, the $\langle 100 \rangle$ -direction gives the maximum tunneling rate for all values of the field strength. Up to $8 \times 10^5 \text{ V cm}^{-1}$ the $\langle 111 \rangle$ -direction leads to the lowest tunneling rate as expected, then it is surpassed by the curve

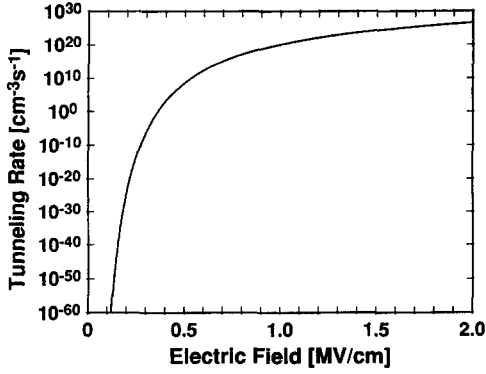


Fig. 4. Calculated band-to-band tunneling rate in silicon.

for the $\langle 110 \rangle$ -direction. However, the major result concerning the anisotropy effect is its general weakness, which is in contrast to the form of a single valley. At 0.65 MV cm^{-1} , where the band-to-band tunneling rate crosses a typical SRH rate of about $10^{14} \text{ cm}^{-3} \text{ s}^{-3}$, the maximum effect is not larger than half an order. Obviously, the presence of six valleys results in an effective averaging, and hence, weakens the anisotropy. Furthermore, we see that with rising field strength the effect becomes less pronounced, since the absolute value of the exponents decreases as the field strength approaches its critical value.

Figure 6 demonstrates the above mentioned uncertainty of the results with respect to the effective hole mass. The direction of the applied field is $\langle 111 \rangle$. Using the heavy hole mass instead of the light hole mass would decrease the rate by seven orders of magnitude at $4 \times 10^5 \text{ V cm}^{-1}$ and still by two orders of magnitude at $1.5 \times 10^6 \text{ V cm}^{-1}$. Although the heavy hole mass $0.49 m$ represents an extreme value, a change of the hole tunneling mass due to the strong field (which itself depends on the magnitude of the field) has to be considered the major cause of uncertainty in the present theory. It is interesting to note that two-level perturbation theory yields just twice the value of a reduced hole mass built from the heavy and light hole masses, i.e. $0.24 m$. Using this value gives the dotted curve in Fig. 6. The solid and dotted

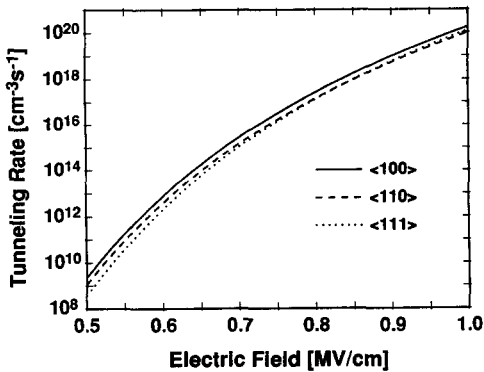


Fig. 5. Band-to-band tunneling rate in silicon for different directions of the electric field.

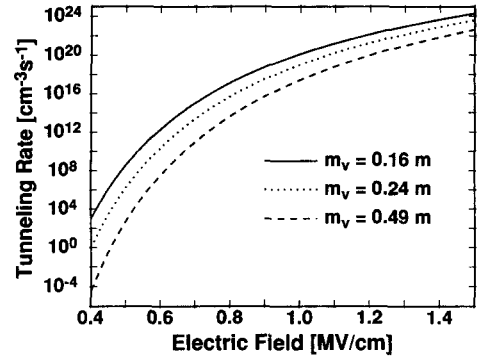


Fig. 6. Band-to-band tunneling rate in silicon for three values of the effective hole mass: $m_v = 0.16 m$ (light hole mass) and $m_v = 0.49 m$ (heavy hole mass) and $m_v = 0.24 m$.

curves should be realistically considered as describing the region of uncertainty due to the effective hole mass.

For application of advanced MOSFETs and DRAMs the nature of tunneling leakage currents is of considerable interest. We compared phonon-assisted band-to-band tunneling with trap-assisted tunneling in Fig. 7. Tunneling transitions via traps are direct because of the strong localization of the deep level wave function. The model used for comparison is based on a quantum-mechanical calculation outlined in Refs [52,53]. In this model trap tunneling is described as field-enhanced SRH recombination. It contains three physical parameters which define the individual recombination center: thermal binding energy E_t , effective phonon energy $\hbar\omega_0$ of the multi-phonon process, and lattice relaxation energy ϵ_R . The latter quantity is a measure of the coupling strength between localized state and lattice. The field effect is primarily influenced by ϵ_R . Therefore, we chose for comparison the cases of weak and strong coupling with values given in Fig. 7, respectively. The remaining parameters are: $E_t = 0.55 \text{ eV}$, $\hbar\omega_0 = 0.068 \text{ eV}$, and $T = 300 \text{ K}$. The field direction is $\langle 111 \rangle$. Since in the band-to-band tunneling rate the difference of the

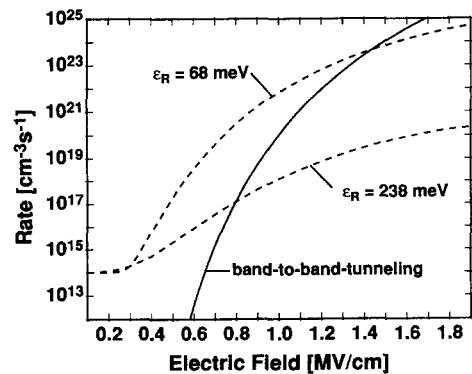


Fig. 7. Comparison of band-to-band tunneling rate and field-dependent SRH rate (trap-assisted tunneling). Trap-assisted tunneling depends sensitively on the lattice energy ϵ_R of the recombination center.

Fermi functions was replaced by unity, we apply the same simplification for the field-enhanced SRH rate, i.e. we write for the zero-field limit:

$$R_{\text{SRH}}^{F=0} = \frac{n_i}{\tau_{\text{SRH}}}. \quad (106)$$

A value of $1 \times 10^{10} \text{ cm}^{-3}$ for the intrinsic density n_i [54] and an upper limit of about 10^{-4} s for the SRH lifetime τ_{SRH} [55] then yield a lower limit for the SRH rate of about $10^{14} \text{ cm}^{-3} \text{ s}^{-1}$.

Figure 7 shows that the onset of trap tunneling is at a field strength of about $3 \times 10^5 \text{ V cm}^{-1}$, independent of the value of ϵ_r , whereas a band-to-band tunneling rate of $10^{14} \text{ cm}^{-3} \text{ s}^{-1}$ is first reached at about $6.6 \times 10^5 \text{ V cm}^{-1}$. There, the trap tunneling rate is already more than two orders of magnitude larger in the case of the strongly coupled recombination center, and about five orders for the weakly coupled. Only at higher field strengths the band-to-band tunneling mechanism starts to dominate, because its field dependence is much steeper. On the other hand, since the light hole mass was used, band-to-band tunneling should be at most weaker but not stronger. This holds also true if the constant field approximation is relaxed, since inhomogeneous fields always tend to lower the tunneling probability. Therefore, it can be concluded that at least in the pre-breakdown range trap-assisted tunneling outnumbers band-to-band tunneling. This is in agreement with the early measurements of the tunneling current in silicon Esaki diodes, e.g. with the findings of Chynoweth and McKay[5], who observed internal field emission in a p - n -junction at forward biases of up to 0.4 V, which cannot be explained by band-to-band tunneling. In addition, Logan and Chynoweth[15] showed by extrapolating the linear excess current region of the I - U -characteristics to lower biases, that at voltages below the peak position the amount of excess current (trap tunneling) is already negligible. Thus, going from large forward bias (low field strength) to zero voltage (higher field strength), first the excess current due to trap-assisted tunneling is observed, which is then superimposed by band-to-band tunneling, and finally, the current is determined by band-to-band tunneling alone.

Here we only considered the case of an upper limit for the SRH lifetime, resulting from unavoidable intrinsic defects (e.g. divacancies). In processed devices τ_{SRH} can be much shorter, which would result in an upward shift of the dashed curves in Fig. 7 and a still stronger dominance of the trap tunneling mechanism.

Acknowledgements—The author wishes to thank Dr H.-J. Wünsche (Humboldt University Berlin) for many valuable discussions.

REFERENCES

1. C. Zener, *Proc. R. Soc. A* **145**, 523 (1934).
2. K. B. McAfee, E. J. Ryder, W. Shockley and M. Sparks, *Phys. Rev.* **83**, 650 (1951).
3. K. G. McKay and K. B. McAfee, *Phys. Rev.* **91**, 1079 (1953); K. G. McKay, *Phys. Rev.* **94**, 877 (1954).
4. S. L. Miller, *Phys. Rev.* **99**, 1234 (1955).
5. A. G. Chynoweth and K. G. McKay, *Phys. Rev.* **106**, 418 (1957).
6. L. Esaki, *Phys. Rev.* **109**, 603 (1958).
7. N. Holonyak Jr, I. A. Lesk, R. N. Hall, J. J. Tiemann and H. Ehrenreich, *Phys. Rev. Lett.* **3**, 167 (1959).
8. R. N. Hall, *Proc. Int. Conf. Phys. Sem.*, Prague 1960, *Czech J. Phys.*, special publication, p. 193 (1960).
9. A. G. Chynoweth and R. A. Logan, *Proc. Int. Conf. Phys. Sem.*, Prague 1960, *Czech. J. Phys.*, special publication, p. 201 (1960).
10. A. G. Chynoweth, W. L. Feldmann and R. A. Logan, *Phys. Rev.* **121**, 684 (1961).
11. T. Yajima and L. Esaki, *J. Phys. Soc. Japan* **13**, 1281 (1958).
12. C. T. Sah, *Bull. Am. Phys. Soc.* **6**, 105 (1961).
13. T. P. Brody, *Bull. Am. Phys. Soc.* **6**, 105 (1961).
14. A. G. Chynoweth, R. A. Logan and D. E. Thomas, *Phys. Rev.* **125**, 877 (1962).
15. R. A. Logan and A. G. Chynoweth, *Phys. Rev.* **131**, 89 (1963).
16. M. S. Tyagi, *Solid-St. Electron.* **11**, 99 (1968).
17. A. G. Chynoweth, *Phys. Rev.* **109**, 1538 (1958).
18. R. B. Fair and H. W. Wivell, *IEEE Trans. Electron Devices* **ED-23**, 512 (1976).
19. J. M. C. Stork and R. D. Isaac, *IEEE Trans. Electron Devices* **ED-30**, 1527 (1983).
20. A. Cuthbertson and P. Ashburn, *IEEE Trans. Electron Devices* **ED-32**, 242 (1985).
21. J. A. Del Alamo and R. M. Swanson, *IEEE Electron Device Lett.* **EDL-7**, 629 (1986).
22. C. Duvvury, D. J. Redwine and H. J. Stiegler, *IEEE Electron Device Lett.* **EDL-9**, 579 (1988).
23. W. Bergner and R. Kircher, *Extended Abstracts of the 22nd Conference on Solid State Devices and Materials*, Sendai, p. 135 (1990).
24. Y. Otake, K. Kurimoto and S. Odanaka, *Extended Abstracts of the 22nd Conference on Solid State Devices and Materials*, Sendai, p. 131 (1990).
25. H. Hazama, *Extended Abstracts of the 22nd Conference on Solid State Devices and Materials*, Sendai, p. 303 (1990).
26. T. Endoh, R. Shiota, M. Momodomi and F. Masuoka, *IEEE Trans. Electron Devices* **ED-37**, 290 (1990).
27. I. Nedev, A. Asenov and E. Stefanov, *Solid-St. Electron.* **34**, 1401 (1991).
28. I.-C. Chen, D. J. Coleman and C. W. Teng, *IEEE Electron Device Lett.* **EDL-10**, 297 (1989).
29. Y. Igura, H. Matsuoka and E. Takeda, *IEEE Electron Device Lett.* **EDL-10**, 227 (1989).
30. S. Banerjee, D. Coleman Jr, W. Richardson and A. Shah, *IEEE Trans. Electron Devices* **ED-35**, 108 (1988).
31. S. H. Voldman, J. B. Johnson, T. D. Linton and S. L. Titcomb, *IEDM Tech. Digest*, p. 349 (1990).
32. L. V. Keldysh, *J. Eksp. Theor. Phys. (U.S.S.R.)* **33**, 994 (1957) transl. in *Soviet Phys. JETP* **6**(4) (1958); *J. Eksp. Theor. Phys. (U.S.S.R.)* **34**, 962 (1958); transl. in *Soviet Phys. JETP* **7**(4) (1958).
33. W. V. Houston, *Phys. Rev.* **57**, 184 (1940).
34. P. J. Price and J. M. Radcliffe, *IBM JI*, Oct., p. 364 (1959).
35. E. O. Kane, *J. appl. Phys.* **32**, 83 (1961).
36. D. R. Fredkin and G. H. Wannier, *Phys. Rev.* **128**, 2054 (1962).
37. R. Enderlein and K. Peuker, *Physica status solidi* (b) **48**, 231 (1971).
38. R. Kubo, *J. Phys. Soc. Japan* **12**, 570 (1957).
39. A. Schenk, M. Stahl and H.-J. Wünsche, *Physica status solidi* (b) **154**, 815 (1989).
40. M. L. Cohen and J. R. Chelikowsky, *Electronic Structure and Optical Properties of Semiconductors*. Springer, Berlin (1988).

41. E. O. Kane, *J. Phys. Chem. Solids* **12**, 181 (1959).
42. K. Henneberger, R. Strehlow and H.-J. Wünsche, *Phys. status solidi* (b) **61**, 455 (1974).
43. R. Enderlein, P. Renner and M. Scheele, *Physica status solidi* (b) **71**, 503 (1975).
44. J. C. Hensel, H. Hasegawa and M. Nakayama, *Phys. Rev. A* **138**, 225 (1965).
45. D. E. Aspnes, *Phys. Rev.* **147**, 554 (1966); *Phys. Rev.* **153**, 972 (1967).
46. L. Fritsche, *Physica status solidi* **11**, 381 (1965).
47. M. Takayanagi and S. Iwabuchi, *IEEE Trans. Electron Devices* **38**, 1425 (1991).
48. W. Brauer and H.-W. Streitwolf, *Theoretische Grundlagen der Halbleiterphysik*. Akademie-Verlag, Berlin (1977).
49. C. Canali, C. Jacoboni, F. Nava, G. Ottaviani and A. Alberigi-Quaranta, *Phys. Rev. B* **12**, 2265 (1975).
50. R. M. Martin, *Phys. Rev.* **186**, 871 (1969).
51. M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, National Bureau of Standards, Appl. Math. Ser. Nr. 55, (1964).
52. A. Schenk, *J. appl. Phys.* **71**, 3339 (1992).
53. A. Schenk, *Solid-St. Electron.* **35**, 1585 (1992).
54. Th. Wasserrab, *Z. Naturforsch.* **32a**(7), 746 (1977).
55. J. G. Fossum and D. S. Lee, *Solid-St. Electron.* **25**, 741 (1982).

APPENDIX A

Transformation of $\omega_i(\mathbf{k}_\theta, \mathbf{q})$

To be general, we assume an arbitrary angle between electric field vector and rotary axis of the energy ellipsoid. The coordinate system is chosen such that the z -axis (\parallel -axis) coincides with the field direction (see Fig. 8). The energy surface is given then by the expression:

$$E_c(\mathbf{k}) = E_g^i + \frac{\hbar^2}{2} [(\mathbf{k} - \mathbf{k}_0), \hat{m}^{-1}(\mathbf{k} - \mathbf{k}_0)], \quad (\text{A1})$$

with the effective mass tensor:

$$\hat{m}^{-1} = \begin{pmatrix} \frac{1}{m_1} & \frac{1}{m_{12}} & \frac{1}{m_{1\parallel}} \\ \frac{1}{m_{21}} & \frac{1}{m_2} & \frac{1}{m_{2\parallel}} \\ \frac{1}{m_{\parallel 1}} & \frac{1}{m_{\parallel 2}} & \frac{1}{m_{\parallel}} \end{pmatrix}. \quad (\text{A2})$$

Using

$$\mathbf{k} - \mathbf{k}_0 = (k_{\parallel} - k_{0\parallel})\mathbf{e}_{\parallel} + (k_{\perp 1} - k_{0\perp 1})\mathbf{e}_{\perp 1} + (k_{\perp 2} - k_{0\perp 2})\mathbf{e}_{\perp 2},$$

we find explicitly:

$$E_c(\mathbf{k}) = E_g^i + \frac{\hbar^2}{2} \left[\frac{(k_{\parallel} - k_{0\parallel})^2}{m_{\parallel}} + \frac{(k_{\perp 1} - k_{0\perp 1})^2}{m_1} \right.$$

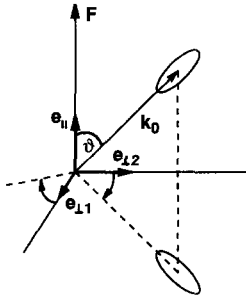


Fig. 8. Field vector \mathbf{F} , valley vector \mathbf{k}_0 , and coordinate system $\{\mathbf{e}_1, \mathbf{e}_{\perp 1}, \mathbf{e}_{\perp 2}\}$.

$$+ \frac{(k_{\perp 2} - k_{0\perp 2})^2}{m_2} + 2 \frac{(k_{\perp 1} - k_{0\perp 1})(k_{\perp 2} - k_{0\perp 2})}{m_{12}} + 2 \frac{(k_{\parallel} - k_{0\parallel})(k_{\perp 1} - k_{0\perp 1})}{m_{1\parallel}} + 2 \frac{(k_{\parallel} - k_{0\parallel})(k_{\perp 2} - k_{0\perp 2})}{m_{2\parallel}}]. \quad (\text{A3})$$

In deriving (A3) $m_{ij}^{-1} = m_i^{-1}$ has been used. The different masses follow from their definitions:

$$\frac{1}{m_1} = (\mathbf{e}_1, \hat{m}^{-1}\mathbf{e}_1), \quad \frac{1}{m_{1,2}} = (\mathbf{e}_{\perp 1,2}, \hat{m}^{-1}\mathbf{e}_{\perp 1,2}),$$

$$\frac{1}{m_{\parallel 1,2}} = (\mathbf{e}_{\parallel}, \hat{m}^{-1}\mathbf{e}_{\perp 1,2}), \quad \frac{1}{m_{12}} = (\mathbf{e}_{\perp 1}, \hat{m}^{-1}\mathbf{e}_{\perp 2}). \quad (\text{A4})$$

Now, $E_c(\mathbf{k})$ is rewritten identically into the form:

$$E_c(\mathbf{k}) = E_g^i + \frac{\hbar^2}{2m_{\parallel}} [(k_{\parallel} - k_{0\parallel}) + \kappa_{\perp}]^2 + \frac{\hbar^2}{2} [(\mathbf{k} - \mathbf{k}_0)_{\perp}, \hat{m}_{\perp}^{-1}(\mathbf{k} - \mathbf{k}_0)_{\perp}], \quad (\text{A5})$$

with the two-dimensional transverse tensor:

$$\hat{m}_{\perp}^{-1} = \begin{pmatrix} \frac{1}{m_1} - \frac{m_{\parallel}}{m_{12}^2} & \frac{1}{m_{12}} - \frac{m_{\parallel}}{m_{11}m_{12}} \\ \frac{1}{m_{12}} - \frac{m_{\parallel}}{m_{11}m_{12}} & \frac{1}{m_2} - \frac{m_{\parallel}}{m_{12}^2} \end{pmatrix}, \quad (\text{A6})$$

and the abbreviation:

$$\kappa_{\perp} = \frac{m_{\parallel}}{m_{11}}(k_{\perp 1} - k_{0\perp 1}) + \frac{m_{\parallel}}{m_{12}}(k_{\perp 2} - k_{0\perp 2}). \quad (\text{A7})$$

So far we have not specified the direction of the vectors $\mathbf{e}_{\perp 1}$ and $\mathbf{e}_{\perp 2}$. We turn the coordinate system until the tensor \hat{m}_{\perp}^{-1} becomes diagonal:

$$E_c(\mathbf{k}) = E_g^i + \frac{\hbar^2}{2m_{\parallel}} [(k_{\parallel} - k_{0\parallel}) + \kappa_{\perp}]^2 + E_{c,\perp}, \quad (\text{A8})$$

where $E_{c,\perp}$ denotes:

$$E_{c,\perp} = \frac{\hbar^2}{2\tilde{m}_{\perp 1}}(k_{\perp 1} - k_{0\perp 1})^2 + \frac{\hbar^2}{2\tilde{m}_{\perp 2}}(k_{\perp 2} - k_{0\perp 2})^2. \quad (\text{A9})$$

New components $k_{0\perp 1}$, $k_{0\perp 2}$, and new masses $\tilde{m}_{\perp 1}$, $\tilde{m}_{\perp 2}$ have been defined by (A8) and (A9). The fixed vectors $\mathbf{e}_{\perp 1}$ and $\mathbf{e}_{\perp 2}$ are the eigenvectors of the tensor \hat{m}_{\perp}^{-1} :

$$\hat{m}_{\perp}^{-1}\mathbf{e}_{\perp i} = \frac{1}{\tilde{m}_{\perp i}}\mathbf{e}_{\perp i}. \quad (\text{A10})$$

From Fig. 8 it follows that $k_{0\perp 1} = 0$ and $k_{0\perp 2} \equiv k_{0\perp}$, and consequently

$$E_{c,\perp} = \frac{\hbar^2}{2\tilde{m}_{\perp 1}}k_{\perp 1}^2 + \frac{\hbar^2}{2\tilde{m}_{\perp 2}}(k_{\perp 2} - k_{0\perp})^2. \quad (\text{A11})$$

Let the eigenvalues of the effective mass tensor in the main diagonal system of the rotary ellipsoid be $m_1^{-1} = m_2^{-1} \equiv m_i^{-1}$ and $m_3^{-1} \equiv m_{\parallel}^{-1}$. We have to express the vectors \mathbf{e}_{\parallel} , $\mathbf{e}_{\perp 1}$ and $\mathbf{e}_{\perp 2}$ within the main diagonal system $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ of the energy surface. We define the angle ϑ between the rotary axis (\mathbf{e}_3) and the field vector (\mathbf{e}_{\parallel}): $\vartheta = \angle(\mathbf{e}_{\parallel}, \mathbf{e}_3)$. Then $k_{0\parallel} = k_0 \cos \vartheta$ and $k_{0\perp} = k_0 \sin \vartheta$ (see Fig. 9). The vector \mathbf{e}_{\parallel} is given by spherical coordinates in the main diagonal system $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$:

$$\mathbf{e}_{\parallel} = (\cos \varphi \sin \vartheta, \sin \varphi \sin \vartheta, \cos \vartheta). \quad (\text{A12})$$

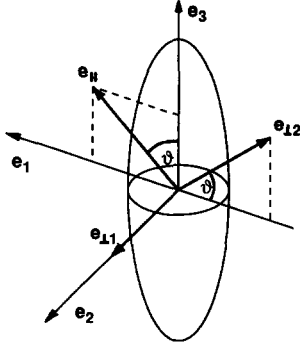


Fig. 9. The two coordinate systems $\{e_1, e_{11}, e_{12}\}$ and $\{e_1, e_2, e_3\}$.

If we turn the system $\{e_1, e_2, e_3\}$ so that $\varphi = 0$, we get (see Fig. 9):

$$\begin{aligned} e_1 &= (\sin \vartheta, 0, \cos \vartheta); \\ e_{11} &= (0, 1, 0); \\ e_{12} &= (-\cos \vartheta, 0, \sin \vartheta). \end{aligned} \quad (\text{A13})$$

All the masses now can be expressed by the longitudinal and transverse effective masses of the ellipsoid:

$$\begin{aligned} \frac{1}{m_{\parallel}} &= (e_{\parallel}, \hat{m}^{-1} e_{\parallel}) = \frac{1}{m_l} \sin^2 \vartheta + \frac{1}{m_t} \cos^2 \vartheta, \\ \frac{1}{\tilde{m}_{11}} &= (e_{11}, \hat{m}^{-1} e_{11}) = \frac{1}{m_t}, \\ \frac{1}{\tilde{m}_{12}} &= (e_{12}, \hat{m}^{-1} e_{12}) = \frac{1}{m_t} \cos^2 \vartheta + \frac{1}{m_l} \sin^2 \vartheta, \\ \frac{1}{m_{\parallel 1}} &= (e_{\parallel}, \hat{m}^{-1} e_{11}) = 0, \\ \frac{1}{m_{\parallel 2}} &= (e_{\parallel}, \hat{m}^{-1} e_{12}) = \left(\frac{1}{m_t} - \frac{1}{m_l} \right) \cos \vartheta \sin \vartheta. \end{aligned} \quad (\text{A14})$$

Furthermore we find:

$$\kappa_{\perp} = \kappa(k_{\perp 2} - k_{0\perp}), \quad \kappa \equiv \frac{(m_l - m_t) \cos \vartheta \sin \vartheta}{m_l \sin^2 \vartheta + m_t \cos^2 \vartheta} \quad (\text{A15})$$

and

$$E_{c\perp} = \frac{\hbar^2}{2m_l} k_{\perp 1}^2 + \frac{\hbar^2}{2\tilde{m}_{12}} (k_{\perp 2} - k_{0\perp})^2. \quad (\text{A16})$$

With these results the energy $\hbar\omega_i(\mathbf{k}_g, \mathbf{q})$ of eqn (47) takes the form:

$$\begin{aligned} \hbar\omega_i(\mathbf{k}_g, \mathbf{q}) &= E_g^i + \frac{\hbar^2}{2m_{\parallel}} [k_{g\parallel} + q_{\parallel} - k_{0\parallel} + \kappa(k_{\perp 2} - k_{0\perp})]^2 \\ &+ \frac{\hbar^2}{2m_v} k_{g\parallel}^2 + \frac{\hbar^2}{2m_l} (k_{\perp 1} + q_{\perp 1})^2 + \frac{\hbar^2}{2\tilde{m}_{12}} (k_{\perp 2} \\ &+ q_{\perp 2} - k_{0\perp})^2 + \frac{\hbar^2}{2m_v} (k_{\perp 1}^2 + k_{\perp 2}^2). \end{aligned} \quad (\text{A17})$$

APPENDIX B

Evaluation of a Double Integral

In this Appendix we derive the exact solution of:

$$\begin{aligned} Q_{\mp}(\mathbf{k}_{\perp}, \mathbf{q}) &= \int_{-\infty}^{\infty} dk_{\parallel} \int_{-\infty}^{\infty} dt \exp\left\{-\frac{i}{\hbar} \int_0^t \right. \\ &\left. \times d\vartheta [\hbar\omega_i(\mathbf{k}_g, \mathbf{q}) - \hbar\omega_{\mp} \mp \hbar\omega_0]\right\}. \end{aligned} \quad (\text{B1})$$

Using (A17) and the abbreviations:

$$\begin{aligned} \tilde{q}_{\parallel} &= q_{\parallel} - k_{0\parallel} + \kappa(k_{\perp 2} - k_{0\perp}), \\ C_{\perp}^{\mp} &= \frac{\hbar^2}{2m_l} (k_{\perp 1} + q_{\perp 1})^2 + \frac{\hbar^2}{2\tilde{m}_{12}} (k_{\perp 2} + q_{\perp 2} \\ &- k_{0\perp})^2 + \frac{\hbar^2}{2m_v} (k_{\perp 1}^2 + k_{\perp 2}^2) - \hbar\omega_{\mp} \mp \hbar\omega_0 + E_g^i \end{aligned} \quad (\text{B2})$$

this can be written as

$$\begin{aligned} Q_{\mp}(\mathbf{k}_{\perp}, \mathbf{q}) &= \int_{-\infty}^{\infty} dk_{\parallel} \int_{-\infty}^{\infty} dt \\ &\times \exp\left\{-\frac{i}{\hbar} \int_0^t d\vartheta \left[\frac{\hbar^2}{2m_l} \left(k_{\parallel} - \frac{e}{\hbar} F\vartheta + \tilde{q}_{\parallel} \right)^2 \right. \right. \\ &\left. \left. + \frac{\hbar^2}{2m_v} \left(k_{\parallel} - \frac{e}{\hbar} F\vartheta \right)^2 + C_{\perp}^{\mp} \right] \right\}. \end{aligned} \quad (\text{B4})$$

We carry out the ϑ -integration in the exponent, rearrange the terms and introduce a new integration variable to get:

$$\begin{aligned} Q_{\mp}(\mathbf{k}_{\perp}, \mathbf{q}) &= \frac{\tilde{q}_{\parallel}}{\beta} \exp\left(-\frac{i}{\hbar} C_{\perp}^{\mp} \beta\right) \int_{-\infty}^{\infty} dt \\ &\times \exp\left\{-i \left[\frac{1}{3} \Theta_{\parallel}^3 (\tau + \beta)^3 + \frac{1}{3} \Theta_v^3 \tau^3 + \frac{C_{\perp}^{\mp}}{\hbar} \tau \right] \right\} \\ &\times \int_{-\infty}^{\infty} dt \exp\left\{-i \left[\frac{1}{3} \Theta_v^3 (t + \beta)^3 \right. \right. \\ &\left. \left. + \frac{1}{3} \Theta_{\parallel}^3 t^3 + \frac{C_{\perp}^{\mp}}{\hbar} t \right] \right\}, \end{aligned} \quad (\text{B5})$$

with $\Theta_{\parallel}^3 = e^2 F^2 / (2m_{\parallel} \hbar)$ and $\beta = \hbar \tilde{q}_{\parallel} / (eF)$. In order to end up with the integral representation of the Airy function we again have to transform the exponents:

$$\begin{aligned} \frac{1}{3} \Theta_{\parallel}^3 (\tau + \beta)^3 + \frac{1}{3} \Theta_v^3 \tau^3 + \frac{C_{\perp}^{\mp}}{\hbar} \tau &= \frac{1}{3} \Theta_{r,\parallel}^3 \left(\tau + \frac{\Theta_{\parallel}^3}{\Theta_{r,\parallel}^3} \beta \right)^3 \\ &+ \left(\Theta_{\parallel}^3 \beta^2 + \frac{C_{\perp}^{\mp}}{\hbar} - \frac{\Theta_{\parallel}^3}{\Theta_{r,\parallel}^3} \Theta_{\parallel}^3 \beta^2 \right) \tau \\ &+ \frac{1}{3} \Theta_{\parallel}^3 \beta^2 - \frac{1}{3} \Theta_{r,\parallel}^3 \left(\frac{\Theta_{\parallel}^3 \beta^3}{\Theta_{r,\parallel}^3} \right). \end{aligned} \quad (\text{B6})$$

Since $\Theta_{\parallel}^3 / \Theta_{r,\parallel}^3 = \mu_{\parallel} / m_{\parallel}$, where μ_{\parallel} is the reduced effective mass in field direction: $\mu_{\parallel}^{-1} = m_{\parallel}^{-1} + m_v^{-1}$, the double integral becomes:

$$Q_{\mp}(\mathbf{k}_{\perp}, \mathbf{q}) = \frac{\tilde{q}_{\parallel}}{\beta} \exp\left(-\frac{i}{\hbar} C_{\perp}^{\mp} \beta\right) J(\Theta_{\parallel}, \Theta_v) J(\Theta_v, \Theta_{\parallel}) \quad (\text{B7})$$

with:

$$\begin{aligned} J(\Theta_{\parallel}, \Theta_v) &= \int_{-\infty}^{\infty} dt \exp\left\{-i \left[\frac{1}{3} \Theta_{r,\parallel}^3 t^3 \right. \right. \\ &+ \left(\beta^2 \Theta_{\parallel}^3 \left(1 - \frac{\mu_{\parallel}}{m_{\parallel}} \right) + \frac{C_{\perp}^{\mp}}{\hbar} \right) \left(t - \frac{\mu_{\parallel}}{m_{\parallel}} \beta \right) \\ &\left. \left. + \frac{1}{3} \Theta_{\parallel}^3 \beta^3 \left(1 - \left(\frac{\mu_{\parallel}}{m_{\parallel}} \right)^2 \right) \right] \right\}. \end{aligned} \quad (\text{B8})$$

Now, we can extract from J an Airy function according to [45]:

$$\text{Ai}(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dt \exp\{-i[\frac{1}{3}t^3 + yt]\}, \quad (\text{B9})$$

which gives for $J(\Theta_{\parallel}, \Theta_{\nu})$

$$J(\Theta_{\parallel}, \Theta_{\nu}) = \frac{2\pi}{\Theta_{r,\parallel}} \text{Ai} \left(\frac{\hbar\beta^2\Theta_{\parallel}^3(1 - \mu_{\parallel}/m_{\parallel}) + C_{\perp}^{\mp}}{\hbar\Theta_{r,\parallel}} \right) \times \exp \left\{ -i \left[-\beta \frac{C_{\perp}^{\mp}}{\hbar} \frac{\mu_{\parallel}}{m_{\parallel}} - \beta^3\Theta_{\parallel}^3 \left(1 - \frac{\mu_{\parallel}}{m_{\parallel}} \right) \frac{\mu_{\parallel}}{m_{\parallel}} + \frac{1}{3}\beta^3\Theta_{\parallel}^3(1 - (\mu_{\parallel}/m_{\parallel})^2) \right] \right\}. \quad (\text{B10})$$

Taking the product of $J(\Theta_{\parallel}, \Theta_{\nu})$ and $J(\Theta_{\nu}, \Theta_{\parallel})$ the exponential can be further simplified by introducing a new reduced mass μ_3

$$\frac{1}{\mu_3} = \frac{1}{3\mu_{\parallel}} + \frac{1}{m_{\parallel} + m_{\nu}}. \quad (\text{B11})$$

Inserting the expressions for β and \tilde{q}_{\parallel} , the double integral takes the form:

$$Q_{\mp}(\mathbf{k}_{\perp}, \mathbf{q}) = (2\pi)^2 \frac{eF}{\hbar\Theta_{r,\parallel}^2} \times \text{Ai} \left(\frac{\left(1 - \frac{\mu_{\parallel}}{m_{\parallel}} \right) \frac{\hbar^2}{2m_{\parallel}} [q_{\parallel} - k_{0\parallel} + \kappa(k_{\perp 2} - k_{0\perp})]^2 + C_{\perp}^{\mp}}{\hbar\Theta_{r,\parallel}} \right)$$

$$\times \text{Ai} \left(\frac{\left(1 - \frac{\mu_{\parallel}}{m_{\nu}} \right) \frac{\hbar^2}{2m_{\nu}} [q_{\parallel} - k_{0\parallel} + \kappa(k_{\perp 2} - k_{0\perp})]^2 + C_{\perp}^{\mp}}{\hbar\Theta_{r,\parallel}} \right) \times \exp \left\{ -i \frac{\hbar^2}{2\mu_3 eF} [q_{\parallel} - k_{0\parallel} + \kappa(k_{\perp 2} - k_{0\perp})]^3 \right\}. \quad (\text{B12})$$

We observe that

$$\left(1 - \frac{\mu_{\parallel}}{m_{\parallel}} \right) \frac{1}{m_{\parallel}} = \left(1 - \frac{\mu_{\parallel}}{m_{\nu}} \right) \frac{1}{m_{\nu}} = \frac{1}{m_{\parallel} + m_{\nu}}, \quad (\text{B13})$$

and therefore, the arguments of the Airy functions are identical. We define:

$$m_{\parallel} + m_{\nu} \equiv m_{\Sigma}, \quad (\text{B14})$$

and obtain the final result:

$$Q_{\mp}(\mathbf{k}_{\perp}, \mathbf{q}) = (2\pi)^2 \frac{eF}{\hbar\Theta_{r,\parallel}^2} \text{Ai}^2 \left(\frac{\frac{\hbar^2 \tilde{q}_{\parallel}^2}{2m_{\Sigma}} + C_{\perp}^{\mp}}{\hbar\Theta_{r,\parallel}} \right) \exp \left(-i \frac{\hbar^2 \tilde{q}_{\parallel}^3}{2\mu_3 eF} \right). \quad (\text{B15})$$