

Wigner transport in the presence of phonons: Particle models of electron kinetics

M. NEDJALKOV

*Department of Electrical Engineering, Arizona State University - USA
and Institute for Microelectronics, TU-Vienna - Austria*

1. – Introduction

Carrier transport in conventional microelectronic devices is governed by the Boltzmann equation (BE). The equation provides a physical picture where carriers are point-like particles (with well-defined position and momentum which compose the phase space) subject to drift and scattering events. During the drift process carriers are accelerated by the device electric field over classical Newton's trajectories. The interaction with the lattice imperfections, such as phonons, gives rise to scattering events which are local in time and space. This means that the interaction is instantaneous and causes a change of the carrier momentum but not of the carrier position. Averaged physical quantities such as mean free path or time between the collisions characterize the transport conditions. The channel of a conventional MOSFET is many mean free paths long. As the field is moderate and changes smoothly, the scattering with phonons maintains the carriers in a local balance with the lattice. The channel of a modern submicron device is only few mean free paths long. The field is strong and exhibits strong changes with the distance. Carriers are no more in a local balance with the lattice. Single carriers may cross the channel without any interaction with phonons. This transport regime is nearly ballistic.

The classical picture fails when the scales of position and momentum or energy and time become such that the Heisenberg uncertainty principle must be taken into account. These scales can be evaluated by the following considerations:

$$(1) \quad \Delta E \cdot \delta t \gtrsim \hbar, \quad \hbar \Delta k \Delta x \gtrsim \hbar,$$

where k is the wave vector.

The position-momentum uncertainty is usually associated with the picture of a moving wave packet. The packet must be well localized in the Brillouin zone, where k is changed by the electric field: Δk must be smaller than the dimension of the Brillouin zone of the order of the reciprocal lattice constant. Then the real space dimension of the wave packet extends above the lattice constant. As the external field must remain constant within the extension of the wave packet (see the discussion after eq. (24)), the Boltzmann picture assumes potentials which are slowly varying over the lattice. This condition is not satisfied around the energy band offsets in nanostructures.

The energy variation can be estimated using as a reference the mean thermal energy, $\Delta E = kT$ (so that $\Delta k = (2mkT)^{1/2}/\hbar$).

- The evolution time of a Boltzmann system must be much longer than the time between the collisions. The latter, in its turn, must be longer than $\delta t \geq kT/\hbar \simeq 0.1$ picoseconds in order to allow the build-up of the classical, energy-conserving delta-function. Otherwise collisional broadening occurs, which will be discussed in the section devoted to the Levinson equation.

- The minimum distance $\Delta x \geq \hbar/(2mkT)^{1/2} = \lambda_B$, where λ_B is called the thermal de Broglie wavelength, is estimated to be $\lambda_B > 10$ nm for most common semiconductors. The dimensions of the Boltzmann system must be much longer than the mean free path, which must be longer than λ_B .

These conditions are not satisfied in the active regions of nanoelectronic devices or during the evolution of systems considered by the modern femtosecond spectroscopy. In such cases the classical transport picture must be abandoned and a quantum description is necessary.

It is reasonable to assume that the carrier transport in the short active regions of nanoelectronic devices is quantum-ballistic. Nevertheless phonon interaction can strongly affect the neighboring regions and thus the device operation. Three types of such nanostructures are considered below.

Figure 1 shows the typical geometry of a FinFET. According to the industry trends, Intel and AMD will begin to utilize this structure in three years, as it offers superior control of short-channel effects. The source (S) carriers are guided through the thin fin to the drain (D) region. The gate potential controls the current flow by opening or closing the inversion channel in the fin. It is the active region of the device which requires quantum treatment. The S/D regions cannot be interpreted as two reservoirs of equilibrium particles: It has been observed that the output characteristics of the device significantly depend on the S/D contact schemes [1] which demonstrate resistance. These regions, large enough to allow classical treatment of the carrier transport, must be included in the considerations. Thus a coupling between the transport pictures in classical and quantum regions is necessary [2].

The resonant-tunneling diode (RTD) is a manifestly quantum device, whose operation is based on tunneling. The energy band diagram, fig. 2, shows the active region comprising the quantum well and the two barriers. The electric field in the region is approximated by a constant. The device operates at far-from-equilibrium conditions provided by a bias applied between the emitter (left) and collector (right) electrodes. The bias controls the

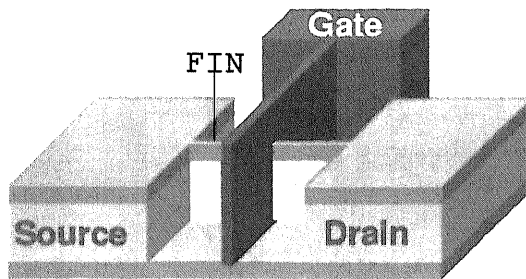


Fig. 1. – The FinFET structure. Typical dimensions for the fin are 10 nm width, 50 nm height and 40 nm length. The width and length of the S/D regions can be several tens of nanometers.

exchange of the particles which flow between the electrodes through the allowed (resonant) energy state formed in the well. The peak of the current occurs when the state lines up with the band minimum in the emitter (chosen in fig. 2 to be the reference energy). After this point the current drops with the increase of the bias, which forms a negative differential resistance region in the current voltage (I - V) characteristics of the device.

Surprisingly, ballistic quantum theories fail to model the device. Non-self-consistent models, where the electric field is frozen, show a significant difference in the I - V characteristics with and without dissipation by phonons [3]. Phonon scattering leads to an increase in the valley current and a resonance voltage shift. The effect is due to a repopulation of the electron states in the emitter. Inelastic scattering events dissipate the energy of the electrons entering from the left-hand electrode. Propagating electrons fall in the lower energy states in the notch just before the left barrier, fig. 2, and contribute to the current.

A self-consistent model has been investigated by Frensley [4] for a single-barrier device. The Schrödinger and Poisson equations have been iteratively solved at consecutive steps. The obtained self-consistent solution gives rise to a rather unphysical potential. The energy barrier lies near the bottom of a parabolic potential well spread across the simulation domain. This is due to a lack of accumulation of electrons in the emitter side of the barrier, corresponding to the notch in fig. 2. The main reason is that the notch states are formed by the states incidenting from the collector side, as indicated by the arrow in fig. 2. The correct accumulation is obtained only by taking into account the interaction with phonons, which links the emitter states with the states in the notch.

As a third device we consider a nanosensor for single-photon counting at microwave frequencies around 500 GHz. This far-infrared range (FIR) of $10\text{ }\mu\text{m}$ -1 mm photon wavelengths is very interesting for spectroscopic research, since it contains the quantum energy levels of semiconductor nanostructures, rotational spectra of molecules and vibrational spectra of solids. Individual counting of photons in this range is very difficult since the typical photon energies are as small as millielectronvolts. To compare, the visible and near-infrared phonon energies are three orders of magnitude higher and can be sensed by photomultipliers.

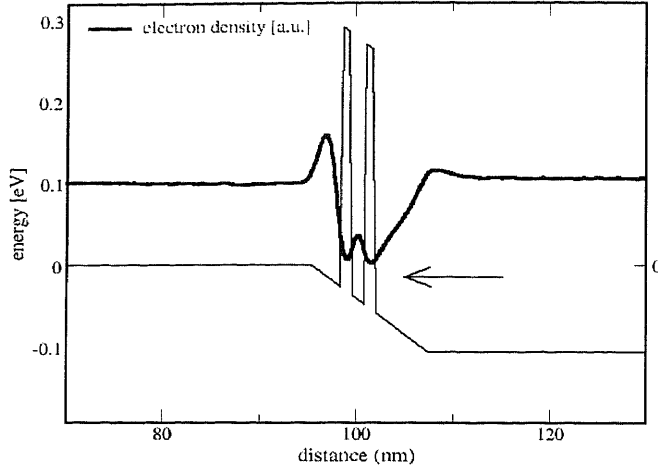


Fig. 2. – Central part of the energy band diagram of an RTD. The height of the barriers is 0.3 eV, their width is 3 nm, and the well width is 5 nm. A linear voltage drop is assumed over a distance of 40 nm.

The extreme sensitivity of single-electron transistors (SET) to charge has recently been utilized for developing sensors for single FIR photons. The latter are converted to a charge through an appropriate excitation mechanism. If cyclotron resonance is adopted as such mechanism, strong magnetic fields are needed. An alternative mechanism without magnetic field is utilized by the SET based on a double quantum dot system as represented in fig. 3. The plunger gate G and the pair gate 2 control the energy levels μ_A and μ_B in the two quantum dots A and B . The latter are separated by a potential barrier controlled by the pair gate 1. The purpose is to allow tunneling of the electrons excited in dot B to the open dot A . The pair gate 3 is strongly biased so that the tunneling between B and the left-hand side reservoir is suppressed. The wings of the pair gate 2 serve as a dipole antenna which converts the photons into charge energy in B . Finite current is transmitted between S and D regions when the Fermi energy E_F in the leads lines up with μ_A . Hence, if μ_A is changed with the bias of G , the conductance shows oscillations each time when $E_F = \mu_A$. The oscillation period is proportional to the distance between the carrier energy eigenvalues in A . The number of electrons in B affects the oscillation period due to the electrostatic coupling between the two quantum dots. An absorbed photon gives rise to an excited electron in B , which will tunnel to A and furthermore to the leads. There the dissipative electron-phonon interaction cools down the excited electron. The resulting ionization of B affects the oscillation period of the conductance. The detection of individual events is possible because the lifetime of the photoexcited positive charge in B is made sufficiently long. At low temperatures the probability (per unit time) for an electron to absorb energy from the lattice is negligible. Then the lifetime depends inversely on the probability of a cold electron to tunnel back in B .

These examples show that carrier-phonon interaction may play a crucial role for the

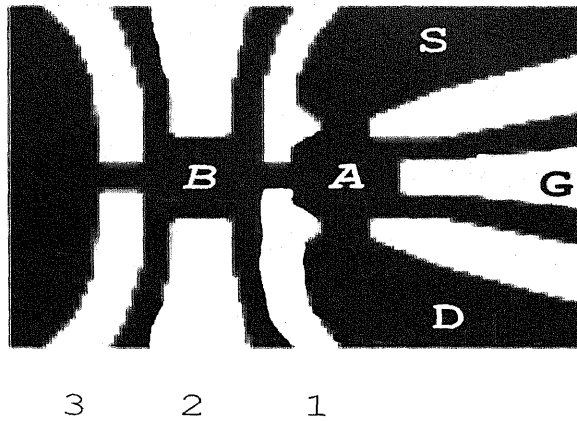


Fig. 3. – Single-photon sensor based on high-mobility GaAs/Al_xGa_{1-x}As heterostructure. The two quantum dots *A* and *B* are formed at $T = 4.2\text{ K}$ by the negatively biased gate *G* and pair gates 1, 2 and 3.

operation of some nanoelectronic devices. Along with the active regions of these devices, the description of the carrier transport must also include regions of classical transport. Different quantum approaches can be chosen for the active region. At the interface with the classical regions the quantities inherent for the chosen approach must be coupled with the Boltzmann distribution function or its moments. We will see that a very convenient approach is provided by the Wigner formulation of quantum mechanics. It retains many of the concepts and notions of the classical transport such as phase space and distribution function. Furthermore phonon interaction can be accounted for within a hierarchy of models, which range between first-principle quantum description and Boltzmann scattering. We will first introduce the single-electron Wigner function.

2. – Classical and quantum distribution functions

2.1. Classical statistical mechanics. – We recall some concepts and notions of the statistical mechanics. We consider a single particle in a potential field. The single-particle phase space is defined by the Cartesian product of the position \mathbf{r} and momentum \mathbf{p} . Physical quantities are functions $A(\mathbf{r}, \mathbf{p})$ defined in the phase space. The state of a single particle is presented by a point in the phase space. A statistical description is introduced if the coordinates of the point cannot be specified exactly, but with some probability. Alternatively we may consider a system of equivalent particles, which are sufficiently many to allow statistical description. The particles are considered non-interacting, but may interact with the environment.

According to the basic postulate of the classical statistical mechanics, the state of any particle system is completely specified by a function $f(\mathbf{r}, \mathbf{p}, t)$, called distribution

function with the following properties:

$$(2) \quad f(\mathbf{r}, \mathbf{p}, t) \geq 0, \quad \int d\mathbf{r} d\mathbf{p} f(\mathbf{r}, \mathbf{p}, t) = 1,$$

so that the mean value $\langle A \rangle$ of any physical quantity is given by

$$(3) \quad \langle A \rangle(t) = \int d\mathbf{r} d\mathbf{p} A(\mathbf{r}, \mathbf{p}) f(\mathbf{r}, \mathbf{p}, t).$$

The evolution equation for f is given by

$$(4) \quad \left(\frac{\partial}{\partial t} + \frac{\mathbf{p}}{m} \cdot \frac{\partial}{\partial \mathbf{r}} + \mathbf{F}(\mathbf{r}) \frac{\partial}{\partial \mathbf{p}} \right) f(\mathbf{r}, \mathbf{p}, t) = \left(\frac{\partial f}{\partial t} \right)_c.$$

Here m is the particle mass and the force $\mathbf{F} = -\nabla V$ is given by the gradient of the potential energy V . The characteristics of the differential operator in the brackets, called Liouville operator, are classical Newton's trajectories. Over such trajectories the left-hand side of (4) becomes a total time derivative. In the case of no interaction with the environment, $(\partial f / \partial t)_c = 0$ trajectories carry a constant value of f . Otherwise the particles are redistributed between the trajectories and the right-hand side of (4) equals the net change of the particle density due to collisions. In the rest of this section we derive a quantum analog of eqs. (2), (3) and the Boltzmann equation (4).

2'2. The Weyl transform. – Physical quantities in quantum mechanics are presented by Hermitian operators \hat{A} :

$$(5) \quad \hat{A}|\phi_n\rangle = a_n|\phi_n\rangle, \quad \langle\phi_n|\phi_m\rangle = \delta_{mn}, \quad \sum_n |\phi_n\rangle\langle\phi_n| = \hat{1}.$$

Such operators have real eigenvalues and a complete system of orthonormal eigenvectors which form an abstract Hilbert space. If the spectrum is continuous the sum transforms into integral and the Kronecker delta becomes a delta-function. We assume that the reader is familiar with the Dirac notations. A state $|\Psi_t\rangle$ of the system now is an element of the Hilbert space (quantum counterpart of the classical phase space point). In wave mechanics it is postulated that the state $|\Psi_t\rangle$ is a solution of the Schrödinger equation

$$(6) \quad \hat{H}|\Psi_t\rangle = i\hbar \frac{\partial |\Psi_t\rangle}{\partial t}, \quad \langle\Psi_t|\Psi_t\rangle = 1,$$

where \hat{H} is the operator of the energy. It can be shown that during the evolution the state remains normalized, which is often called conservation of probability.

How to define the operator \hat{A} associated to given physical quantity? According to the correspondence principle, to the classical position and momentum correspond the Hermitian operators $\hat{\mathbf{r}}$ and $\hat{\mathbf{p}}$, which, moreover, obey the following commutation relation:

$$(7) \quad \mathbf{r} \longrightarrow \hat{\mathbf{r}}, \quad \mathbf{p} \longrightarrow \hat{\mathbf{p}}, \quad \hat{\mathbf{r}}\hat{\mathbf{p}} - \hat{\mathbf{p}}\hat{\mathbf{r}} = [\hat{\mathbf{r}}, \hat{\mathbf{p}}]_- = i\hbar\hat{1}.$$

With the help of (7) and the knowledge of $A(\mathbf{r}, \mathbf{p})$, we can try to obtain \hat{A} explicitly. For example, we can use the Taylor expansion to establish the rule

$$A(\mathbf{r}, \mathbf{p}) = \sum_{i,j} b_{i,j} \mathbf{r}^i \mathbf{p}^j \longrightarrow A(\hat{\mathbf{r}}, \hat{\mathbf{p}}) = \sum_{i,j} b_{i,j} \hat{\mathbf{r}}^i \hat{\mathbf{p}}^j.$$

For the Hamiltonian of a particle in a potential field $H(\mathbf{r}, \mathbf{p}) = (\mathbf{p}^2/2m) + V(\mathbf{r})$ this gives the correct result, see (11). However, for general functions A the procedure is not straightforward, since the operators $\hat{\mathbf{p}}$ and $\hat{\mathbf{r}}$ do not commute. First, non-Hermitian operators can appear. Second, even for Hermitian operators there is ambiguity in the correspondence. Consider two equivalent expressions for the function $A(x, p_x)$:

$$A_1 = p_x x^2 p_x = A_2 = \frac{1}{2} (p_x^2 x^2 + x^2 p_x^2).$$

A replacement with \hat{x} and \hat{p}_x gives rise to the following operators:

$$A_1 \longrightarrow \hat{A}_1 = \hat{p}_x \hat{x}^2 \hat{p}_x, \quad A_2 \longrightarrow \hat{A}_2 = \frac{1}{2} (\hat{p}_x^2 \hat{x}^2 + \hat{x}^2 \hat{p}_x^2).$$

Now, while $A_1 = A_2$, the obtained operators differ by \hbar^2 : $\hat{A}_1 = \hat{A}_2 + \hbar^2$. Thus (7) is not sufficient to establish a unique correspondence between A and \hat{A} . The ambiguity is avoided by establishing a special rule of correspondence, which is postulated:

$$(8) \quad \hat{A} = \hat{A}(\hat{\mathbf{r}}, \hat{\mathbf{p}}) = \int d\mathbf{s} d\mathbf{q} \beta(\mathbf{s}, \mathbf{q}) e^{i(\mathbf{s}\hat{\mathbf{r}} + \mathbf{q}\hat{\mathbf{p}})}.$$

Equation (8) is called Weyl transform. Here β is adjoint to A via the Fourier transform:

$$(9) \quad A(\mathbf{r}, \mathbf{p}) = \int d\mathbf{s} d\mathbf{q} \beta(\mathbf{s}, \mathbf{q}) e^{i(\mathbf{r}\mathbf{s} + \mathbf{p}\mathbf{q})}, \quad \beta(\mathbf{r}, \mathbf{p}) = \frac{1}{(2\pi)^6} \int d\mathbf{r} d\mathbf{p} A(\mathbf{r}, \mathbf{p}) e^{-i(\mathbf{r}\mathbf{s} + \mathbf{p}\mathbf{q})}.$$

2'3. The density matrix. — The uncertainty principle imposes a low limit in precision for the joint knowledge of \mathbf{r} , \mathbf{p} . Wave mechanics uses only half of the phase space—coordinate or momentum representation—for the description of the physical system. We assume a coordinate representation, where

$$(10) \quad \hat{\mathbf{r}}|\mathbf{r}\rangle = \mathbf{r}|\mathbf{r}\rangle, \quad \int d\mathbf{r} |\mathbf{r}\rangle \langle \mathbf{r}| = \hat{1}, \quad \hat{\mathbf{p}} = -i\hbar \frac{\partial}{\partial \mathbf{r}}.$$

By denoting $\Psi_t(\mathbf{r}) = \langle \mathbf{r} | \Psi_t \rangle$, the Schrödinger equation for a particle in a potential V reads

$$(11) \quad \langle \mathbf{r} | \hat{H} | \Psi_t \rangle = \left(-\frac{\hbar^2 \partial^2}{2m \partial \mathbf{r}^2} + V(\mathbf{r}) \right) \Psi_t(\mathbf{r}) = i\hbar \frac{\partial \Psi_t(\mathbf{r})}{\partial t}.$$

The average value of a physical quantity A in state $|\Psi_t\rangle$ is expressed by

$$(12) \quad \langle A \rangle(t) = \langle \Psi_t | \hat{A} | \Psi_t \rangle = \int d\mathbf{r} \langle \Psi_t | \mathbf{r} \rangle \langle \mathbf{r} | \hat{A} | \Psi_t \rangle.$$

This formalism is too far away from the familiar classical concepts. Nevertheless, it is possible to reformulate the ideas of quantum mechanics in phase space. The first step is to double the variables in the representation. For this we recall how \hat{A} operates on $|\Psi_t\rangle$. With the help of (5) and (10) we obtain the following integral:

$$\langle \mathbf{r} | \hat{A} | \Psi_t \rangle = \int d\mathbf{r}' \sum_n a_n \langle \mathbf{r} | \phi_n \rangle \langle \phi_n | \mathbf{r}' \rangle \langle \mathbf{r}' | \Psi_t \rangle = \int d\mathbf{r}' \alpha(\mathbf{r}, \mathbf{r}') \Psi_t(\mathbf{r}').$$

A replacement in (12) shows that the physical average $\langle A \rangle(t)$ is actually evaluated in a “double half” of the phase space:

$$\langle A \rangle(t) = \int d\mathbf{r}' \int d\mathbf{r} \Psi_t^*(\mathbf{r}) \alpha(\mathbf{r}, \mathbf{r}') \Psi_t(\mathbf{r}').$$

This implies to investigate the quantity

$$\rho_t(\mathbf{r}, \mathbf{r}') = \Psi_t^*(\mathbf{r}') \Psi_t(\mathbf{r}) = \langle \mathbf{r} | \Psi_t \rangle \langle \Psi_t | \mathbf{r}' \rangle = \langle \mathbf{r} | \hat{\rho}_t | \mathbf{r}' \rangle,$$

called density matrix. It is straightforward to see that the corresponding density operator $\hat{\rho}_t$ is Hermitian. $\langle A \rangle(t)$ can be expressed as

$$(13) \quad \begin{aligned} \langle A \rangle(t) &= \int d\mathbf{r} d\mathbf{r}' \rho_t(\mathbf{r}, \mathbf{r}') \alpha(\mathbf{r}', \mathbf{r}) = \int d\mathbf{r} \int d\mathbf{r}' \langle \mathbf{r} | \hat{\rho}_t | \mathbf{r}' \rangle \langle \mathbf{r}' | \hat{A} | \mathbf{r} \rangle = \\ &= \int d\mathbf{r} \langle \mathbf{r} | \hat{\rho}_t \hat{A} | \mathbf{r} \rangle = \text{Tr}(\hat{\rho}_t \hat{A}). \end{aligned}$$

Using (6) we can prove that $\hat{\rho}_t$ conserves the probability in time:

$$(14) \quad \text{Tr}(\hat{\rho}_t) = \int d\mathbf{r} \langle \mathbf{r} | \hat{\rho}_t | \mathbf{r} \rangle = \int d\mathbf{r} \Psi_t^*(\mathbf{r}) \Psi_t(\mathbf{r}) = 1.$$

2.4. The Wigner function for pure state. – From eq. (11) and its conjugate we obtain the equation of motion of ρ_t (the time dependence is written now explicitly):

$$(15) \quad i\hbar \frac{\partial \rho(\mathbf{r}, \mathbf{r}', t)}{\partial t} = \langle \mathbf{r} | [\hat{H}, \hat{\rho}_t]_- | \mathbf{r}' \rangle = \left\{ -\frac{\hbar^2}{2m} \left(\frac{\partial^2}{\partial \mathbf{r}^2} - \frac{\partial^2}{\partial \mathbf{r}'^2} \right) + (V(\mathbf{r}) - V(\mathbf{r}')) \right\} \rho(\mathbf{r}, \mathbf{r}', t).$$

We change the variables in the above equation by using a center-of-mass transform:

$$(16) \quad \begin{aligned} \mathbf{r}_1 &= \frac{(\mathbf{r} + \mathbf{r}')}{2}, & \mathbf{r}_2 &= \mathbf{r} - \mathbf{r}', \\ \frac{\partial \rho(\mathbf{r}_1 + \mathbf{r}_2/2, \mathbf{r}_1 - \mathbf{r}_2/2, t)}{\partial t} &= \\ \frac{1}{i\hbar} \left\{ -\frac{\hbar^2}{m} \frac{\partial^2}{\partial \mathbf{r}_1 \partial \mathbf{r}_2} + \left(V\left(\mathbf{r}_1 + \frac{\mathbf{r}_2}{2}\right) - V\left(\mathbf{r}_1 - \frac{\mathbf{r}_2}{2}\right) \right) \right\} \rho\left(\mathbf{r}_1 + \frac{\mathbf{r}_2}{2}, \mathbf{r}_1 - \frac{\mathbf{r}_2}{2}, t\right). \end{aligned}$$

The Wigner function is defined by the Fourier transform with respect to \mathbf{r}_2

$$f_W(\mathbf{r}_1, \mathbf{p}, t) = \frac{1}{(2\pi\hbar)^3} \int d\mathbf{r}_2 \rho\left(\mathbf{r}_1 + \frac{\mathbf{r}_2}{2}, \mathbf{r}_1 - \frac{\mathbf{r}_2}{2}, t\right) e^{-i\mathbf{r}_2 \cdot \mathbf{p}/\hbar}.$$

We note that, due to the Wigner transform, \mathbf{r}_1 and \mathbf{p} are independent variables. It is easy to show that the corresponding operators commute. Thus \mathbf{r}_1 and \mathbf{p} define a phase space —the Wigner phase space.

After applying the Fourier transform to (16) we obtain on the right-hand side two terms which are evaluated by using the abbreviation $\rho(+, -, t)$ for $\rho(\mathbf{r}_1 + \mathbf{r}_2/2, \mathbf{r}_1 - \mathbf{r}_2/2, t)$:

$$\begin{aligned} I &= -\frac{1}{i\hbar} \frac{\hbar^2}{m(2\pi\hbar)^3} \int d\mathbf{r}_2 e^{-i\mathbf{r}_2 \cdot \mathbf{p}/\hbar} \frac{\partial^2 \rho(+, -, t)}{\partial \mathbf{r}_1 \partial \mathbf{r}_2} = \\ &= -\frac{1}{m(2\pi\hbar)^3} \mathbf{p} \cdot \frac{\partial}{\partial \mathbf{r}_1} \int d\mathbf{r}_2 e^{-i\mathbf{r}_2 \cdot \mathbf{p}/\hbar} \rho(+, -, t) = -\frac{1}{m} \mathbf{p} \cdot \frac{\partial f_W(\mathbf{r}_1, \mathbf{p}, t)}{\partial \mathbf{r}_1}, \end{aligned}$$

where we have integrated by parts using $\rho \rightarrow 0$ if $\mathbf{r}_2 \rightarrow \pm\infty$;

$$\begin{aligned} II &= \frac{1}{i\hbar(2\pi\hbar)^3} \int d\mathbf{r}_2 e^{-i\mathbf{r}_2 \cdot \mathbf{p}/\hbar} \left(V\left(\mathbf{r}_1 + \frac{\mathbf{r}_2}{2}\right) - V\left(\mathbf{r}_1 - \frac{\mathbf{r}_2}{2}\right) \right) \rho(+, -, t) = \\ &= \int d\mathbf{r}_2 \int d\mathbf{r}' \delta(\mathbf{r}_2 - \mathbf{r}') e^{-i\mathbf{r}_2 \cdot \mathbf{p}/\hbar} \left(V\left(\mathbf{r}_1 + \frac{\mathbf{r}_2}{2}\right) - V\left(\mathbf{r}_1 - \frac{\mathbf{r}_2}{2}\right) \right) \rho\left(\mathbf{r}_1 + \frac{\mathbf{r}'}{2}, \mathbf{r}_1 - \frac{\mathbf{r}'}{2}, t\right). \end{aligned}$$

After replacing the delta-function with the integral

$$(17) \quad \delta(\mathbf{r}_2 - \mathbf{r}') = \frac{1}{(2\pi\hbar)^3} \int d\mathbf{p}' e^{i(\mathbf{r}_2 - \mathbf{r}') \cdot \mathbf{p}'/\hbar},$$

we obtain

$$\begin{aligned} II &= \frac{1}{i\hbar(2\pi\hbar)^3} \int d\mathbf{p}' \int d\mathbf{r}_2 e^{i\mathbf{r}_2 \cdot (\mathbf{p}' - \mathbf{p})/\hbar} \left(V\left(\mathbf{r}_1 + \frac{\mathbf{r}_2}{2}\right) - V\left(\mathbf{r}_1 - \frac{\mathbf{r}_2}{2}\right) \right) \times \\ &\times \frac{1}{(2\pi\hbar)^3} \int d\mathbf{r}' e^{-i\mathbf{r}' \cdot \mathbf{p}'/\hbar} \rho\left(\mathbf{r}_1 + \frac{\mathbf{r}'}{2}, \mathbf{r}_1 - \frac{\mathbf{r}'}{2}, t\right) = \\ &= \int d\mathbf{p}' V_W(\mathbf{r}_1, \mathbf{p}' - \mathbf{p}) f_W(\mathbf{r}_1, \mathbf{p}', t). \end{aligned}$$

We summarize the results of these transformations. Equation (16) gives rise to the Wigner equation

$$(18) \quad \frac{\partial f_W(\mathbf{r}, \mathbf{p}, t)}{\partial t} + \frac{\mathbf{p}}{m} \cdot \frac{\partial f_W(\mathbf{r}, \mathbf{p}, t)}{\partial \mathbf{r}} = \int d\mathbf{p}' V_W(\mathbf{r}_1, \mathbf{p}' - \mathbf{p}) f_W(\mathbf{r}, \mathbf{p}', t),$$

where V_W is the Wigner potential,

$$(19) \quad V_W(\mathbf{r}, \mathbf{p}) = \frac{1}{i\hbar(2\pi\hbar)^3} \int d\mathbf{r}' e^{-i\mathbf{r}'\mathbf{p}/\hbar} \left(V\left(\mathbf{r} - \frac{\mathbf{r}'}{2}\right) - V\left(\mathbf{r} + \frac{\mathbf{r}'}{2}\right) \right).$$

Note that the sign of \mathbf{r}_2 has been changed to obtain the most frequently used definition of the Wigner potential V_W .

2.5. Properties of the Wigner function. – We first outline the equivalence between the Schrödinger equation and the Wigner equation in the case of a pure state. From Ψ_t we can obtain ρ and thus f_W . The opposite is also true: it can be shown that, if we know f_W , we can obtain Ψ_t up to a phase factor.

Comparing with (4), we can recognize, on the left-hand side of the Wigner equation, the fieldless Liouville operator. It is easy to see that the Wigner potential is a real quantity, $V_W = V_W^*$. It follows that f_W , being a solution of an equation with real coefficients, is real. The Wigner function conserves the probability in time:

$$(20) \quad \int d\mathbf{r} \int d\mathbf{p} f_W(\mathbf{r}, \mathbf{p}, t) = \int d\mathbf{r} \int d\mathbf{r}_2 \rho\left(\mathbf{r} + \frac{\mathbf{r}_2}{2}, \mathbf{r}_1 - \frac{\mathbf{r}_2}{2}, t\right) \delta(\mathbf{r}_2) = \int d\mathbf{r} \langle \mathbf{r} | \hat{\rho}_t | \mathbf{r} \rangle = 1.$$

In a similar way it can be shown that the position or momentum probability distributions are obtained after integration over momentum or \mathbf{p} or \mathbf{r} , respectively

$$(21) \quad \int d\mathbf{p} f_W(\mathbf{r}, \mathbf{p}, t) = |\Psi_t(\mathbf{r})|^2, \quad \int d\mathbf{r} f_W(\mathbf{r}, \mathbf{p}, t) = |\Psi_t(\mathbf{p})|^2.$$

The most important property of the Wigner picture is that the mean value $\langle A \rangle(t)$ of any physical quantity is given by

$$(22) \quad \langle A \rangle(t) = \int d\mathbf{r} \int d\mathbf{p} f_W(\mathbf{r}, \mathbf{p}, t) A(\mathbf{r}, \mathbf{p}),$$

where $A(\mathbf{r}, \mathbf{p})$ is the classical function (9). This will be proved in Appendix A.

Our goal formulated at the beginning of this section has been satisfied to a large extent. Equation (20) corresponds to the second equation in (2) and the Wigner function is real. Equation (22) is equivalent to (3). The left-hand sides of the Wigner equation (18) and the Boltzmann equation are given by the Liouville operator. Classical and quantum pictures become very close. Nevertheless there are basic differences. The Wigner function allows negative values and thus is not a probability function. It cannot be interpreted as a joint distribution of particle position and momentum. Actually the Wigner function can

have non-zero values in domains where the particle density is zero. As follows from (21), a physical interpretation is possible only after integration.

We discuss the classical limit of (18).

2'6. Classical limit of the Wigner equation. – Assume that the potential V is a linear or a quadratic function of the position:

$$V\left(\mathbf{r} \pm \frac{\mathbf{r}'}{2}\right) = V(\mathbf{r}) \pm \frac{\partial V(\mathbf{r})}{\partial \mathbf{r}} \frac{\mathbf{r}'}{2} + \dots = V(\mathbf{r}) \mp \mathbf{F}(\mathbf{r}) \frac{\mathbf{r}'}{2} + \dots,$$

where the dots stay for the quadratic term. The force \mathbf{F} can be at most a linear function of the position. As the even terms of the Taylor expansion of V cancel in (19), the Wigner potential becomes

$$V_W(\mathbf{r}, \mathbf{p}) = \frac{1}{i\hbar(2\pi\hbar)^3} \int d\mathbf{r}' e^{-i\mathbf{r}'\mathbf{p}/\hbar} \mathbf{F}(\mathbf{r}) \mathbf{r}'.$$

The right-hand side of (18) becomes

$$\begin{aligned} (23) \quad & \int d\mathbf{p}' V_W(\mathbf{r}_1, \mathbf{p}' - \mathbf{p}) f_W(\mathbf{r}, \mathbf{p}', t) = \\ & \frac{-i}{\hbar(2\pi\hbar)^3} \int d\mathbf{p}' \int d\mathbf{r}' e^{-i\mathbf{r}'(\mathbf{p}' - \mathbf{p})/\hbar} \mathbf{F}(\mathbf{r}) \mathbf{r}' f_W(\mathbf{r}, \mathbf{p}', t) = \\ & \frac{-\mathbf{F}(\mathbf{r})}{(2\pi\hbar)^3} \frac{\partial}{\partial \mathbf{p}} \int d\mathbf{p}' \int d\mathbf{r}' e^{-i\mathbf{r}'(\mathbf{p}' - \mathbf{p})/\hbar} f_W(\mathbf{r}, \mathbf{p}', t) = -\mathbf{F}(\mathbf{r}) \frac{f_W(\mathbf{r}, \mathbf{p}, t)}{\partial \mathbf{p}}, \end{aligned}$$

where we have used the equality $\mathbf{r}' e^{-i\mathbf{r}'(\mathbf{p}' - \mathbf{p})/\hbar} = -i\hbar(\partial/\partial \mathbf{p}) e^{-i\mathbf{r}'(\mathbf{p}' - \mathbf{p})/\hbar}$. Then the Wigner equation reduces to the collisionless Boltzmann equation

$$(24) \quad \frac{\partial f_W(\mathbf{r}, \mathbf{p}, t)}{\partial t} + \frac{\mathbf{p}}{m} \cdot \frac{\partial f_W(\mathbf{r}, \mathbf{p}, t)}{\partial \mathbf{r}} + \mathbf{F}(\mathbf{r}) \frac{\partial f_W(\mathbf{r}, \mathbf{p}, t)}{\partial \mathbf{p}} = 0.$$

Now consider a minimum-uncertainty wave packet as an initial condition. The Wigner function of such a packet is a Gaussian of both position and momentum [6]. The latter can equally well be interpreted as an initial distribution of classical electrons. Provided that the force is a constant or a linear function of position, the packet evolves according to eq. (24). The evolution resembles that of the classical distribution. Despite the spread in the phase space, the Gaussian components determine the general shape of the packet. f_W remains positive during the evolution.

However, stronger variations of the field with position introduce interference effects. Near band offsets the packet rapidly loses its shape and negative values appear. Figure 4 shows the negative values which appear at the initial stage of the interaction of a Gaussian wave packet with a single potential barrier [5].

We will show that, in the presence of phonons, (24) turns into the Boltzmann equation.

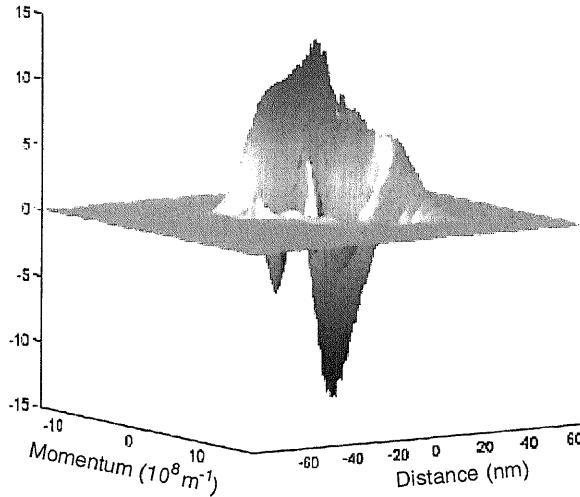


Fig. 4. – Gaussian wave packet Wigner function interacting with a 3 nm, 0.3 eV potential barrier. After Shifren and Ferry [5].

2.7. Quantum-statistical mechanics. – The density operator $\hat{\rho}_t = |\Psi_t\rangle\langle\Psi_t|$, used to obtain the Wigner function, corresponds to a system in pure state. Often the state of the system is not known exactly. Assume that a set of possible states $\hat{\rho}_t^i$ can be occupied with probabilities γ_i . Then the definition of density operator can be generalized (compare with subsect. 2.3). The density operator for a mixed state is defined as

$$(25) \quad \hat{\rho}_t = \sum_i \gamma_i \hat{\rho}_t^i, \quad \sum_i \gamma_i = 1, \quad \gamma_i > 0.$$

Then the mean value of a given physical quantity becomes a statistical average of “averages in states i ”. It is easy to see that (13) and (14) continue to hold in this case.

According to the basic postulate of quantum statistical mechanics, the state of the system is completely specified by the density operator $\hat{\rho}_t$. The mean value of \hat{A} and the equation of motion of $\hat{\rho}_t$ are given by

$$(26) \quad \langle \hat{A} \rangle(t) = \text{Tr}(\hat{\rho}_t \hat{A}), \quad \frac{\partial \hat{\rho}_t}{\partial t} = \frac{1}{i\hbar} [\hat{H}, \hat{\rho}_t]_-.$$

The mixed-state Wigner function and equation are derived from $\hat{\rho}$ and its equation of motion as in the case of pure state. Since the derivation is reversible, we can equivalently postulate $f_W(\mathbf{r}, \mathbf{p}, t)$ as a definition of the state of the system.

Note that if we know the set γ_i we can build up the density matrix from (25). This is for example possible in models where γ_i can be obtained from the boundary conditions [6]. However, for more complex physical systems, containing electrons which interact with other types of quasi-particles, γ_i are not known *a priori*. In this case $\hat{\rho}_i$ and γ_i are

obtained with the help of (26). Indeed the corresponding representation of the system is given by the basis vectors $|X_i\rangle|\mathbf{r}\rangle$, where the additional degrees of freedom X describing the quasi-particles are assumed enumerable. We are usually interested in the electron averages, so that \hat{A} does not affect X_i . By using (26) we obtain

$$\langle \hat{A} \rangle(t) = \text{Tr} (\hat{\rho}_t \hat{A}) = \sum_i \int d\mathbf{r} \langle \mathbf{r} | \langle X_i | \hat{\rho}_t \hat{A} | X_i \rangle | \mathbf{r} \rangle = \int d\mathbf{r} \langle \mathbf{r} | \hat{\rho}_t^e \hat{A} | \mathbf{r} \rangle = \text{Tr}_e (\hat{\rho}_t^e \hat{A}),$$

where $\hat{\rho}_t^e = \sum_i \langle X_i | \hat{\rho}_t | X_i \rangle$ is the electron, or the reduced density operator. We can introduce the set of probabilities γ_i and the set of electron density operators $\hat{\rho}_t^{e,i}$ according to

$$\gamma_i = \text{Tr}_e (\langle X_i | \hat{\rho}_t | X_i \rangle) \geq 0, \quad \sum_i \gamma_i = 1; \quad \hat{\rho}_t^{e,i} = \frac{\langle X_i | \hat{\rho}_t | X_i \rangle}{\text{Tr}_e (\langle X_i | \hat{\rho}_t | X_i \rangle)}, \quad \text{Tr}_e (\hat{\rho}_t^{e,i}) = 1.$$

The estimates follow from the fact that $\hat{\rho}_t$ is a positively defined operator and from the conservation of the probability. Hence

$$\hat{\rho}_t^e = \sum_i \gamma_i \hat{\rho}_t^{e,i}$$

and we formally arrive at (25). However, in order to obtain γ_i and $\hat{\rho}_t^{e,i}$ we need $\hat{\rho}_t$, *i.e.* we need to solve the evolution equation for the whole system. Usually this is not possible, moreover we are not interested in the detailed information about the state of the quasi-particles. This implies to approximate the evolution equation (26) to a closed equation for the electron density operator. Alternatively this can be done in terms of the Wigner functions obtained after a Wigner transform of the corresponding density operators.

2'8. Appendix. – Here we prove the important relation (22). From the Weyl transform we have

$$\hat{A} = A(\hat{\mathbf{r}}, \hat{\mathbf{p}}) = \frac{1}{(2\pi)^6} \int d\mathbf{s} d\mathbf{q} \int d\mathbf{r} d\mathbf{p} A(\mathbf{r}, \mathbf{p}) e^{-i(\mathbf{r}\mathbf{s} + \mathbf{p}\mathbf{q})} e^{i(\mathbf{s}\hat{\mathbf{r}} + \mathbf{q}\hat{\mathbf{p}})}.$$

Inserting in $\int d\mathbf{r}' \langle \mathbf{r}' | \hat{\rho}_t \hat{A} | \mathbf{r}' \rangle$ one obtains

$$\langle \hat{A} \rangle(t) = \int d\mathbf{r} d\mathbf{p} A(\mathbf{r}, \mathbf{p}) \int \frac{d\mathbf{s} d\mathbf{q}}{(2\pi)^6} e^{-i(\mathbf{r}\mathbf{s} + \mathbf{p}\mathbf{q})} \int d\mathbf{r}' \langle \mathbf{r}' | e^{i(\mathbf{s}\hat{\mathbf{r}} + \mathbf{q}\hat{\mathbf{p}})} \hat{\rho}_t(t) | \mathbf{r}' \rangle.$$

The average value $\langle \hat{A} \rangle(t)$ appears as the desired phase space integral with the classical function $A(\mathbf{r}, \mathbf{p})$. It remains to prove that the term I after A is the Wigner function f_W . We make use of the following relations [7]:

$$e^{i(\mathbf{s}\hat{\mathbf{r}} + \mathbf{q}\hat{\mathbf{p}})} = e^{-i\mathbf{s}\mathbf{q}\hbar/2} e^{i\mathbf{q}\hat{\mathbf{p}}} e^{i\mathbf{s}\hat{\mathbf{r}}}, \quad e^{i\mathbf{q}\hat{\mathbf{p}}} |\mathbf{r}\rangle = |\mathbf{r} - \mathbf{q}\hbar\rangle.$$

The calculation is long but straightforward:

$$\begin{aligned}
 I &= \int \frac{d\mathbf{s} d\mathbf{q}}{(2\pi)^6} e^{-i(\mathbf{r}\mathbf{s} + \mathbf{p}\mathbf{q})} \int d\mathbf{r}' \int d\mathbf{r}'' \langle \mathbf{r}' | e^{-i\mathbf{s}\mathbf{q}\hbar/2} e^{i\mathbf{q}\mathbf{p}} e^{i\mathbf{s}\mathbf{r}'} | \mathbf{r}'' \rangle \langle \mathbf{r}'' | \hat{\rho}_t(t) | \mathbf{r}' \rangle = \\
 &\int \frac{d\mathbf{s} d\mathbf{q}}{(2\pi)^6} e^{-i(\mathbf{r}\mathbf{s} + \mathbf{p}\mathbf{q})} \int d\mathbf{r}' \int d\mathbf{r}'' e^{-i\mathbf{s}\mathbf{q}\hbar/2} \langle \mathbf{r}' | e^{i\mathbf{q}\mathbf{p}} | \mathbf{r}'' \rangle e^{i\mathbf{s}\mathbf{r}''} \langle \mathbf{r}'' | \hat{\rho}_t(t) | \mathbf{r}' \rangle = \\
 &\int \frac{d\mathbf{s} d\mathbf{q}}{(2\pi)^6} e^{-i(\mathbf{r}\mathbf{s} + \mathbf{p}\mathbf{q})} \int d\mathbf{r}' \int d\mathbf{r}'' e^{-i\mathbf{s}\mathbf{q}\hbar/2} \delta(\mathbf{r}' - \mathbf{r}'' + \mathbf{q}\hbar) e^{i\mathbf{s}\mathbf{r}''} \rho(\mathbf{r}'', \mathbf{r}', t) = \\
 &\int \frac{d\mathbf{s} d\mathbf{q}}{(2\pi)^6} e^{-i(\mathbf{r}\mathbf{s} + \mathbf{p}\mathbf{q})} \int d\mathbf{r}' e^{-i\mathbf{s}\mathbf{q}\hbar/2} e^{i\mathbf{s}(\mathbf{r}' + \mathbf{q}\hbar)} \rho(\mathbf{r}' + \mathbf{q}\hbar, \mathbf{r}', t).
 \end{aligned}$$

Here we set $\mathbf{r}' = \mathbf{r}_1 - \mathbf{q}\hbar/2$:

$$\begin{aligned}
 I &= \int \frac{d\mathbf{s} d\mathbf{q}}{(2\pi)^6} e^{-i(\mathbf{r}\mathbf{s} + \mathbf{p}\mathbf{q})} \int d\mathbf{r}_1 e^{-i\mathbf{s}\mathbf{q}\hbar/2} e^{i\mathbf{s}(\mathbf{r}_1 + \mathbf{q}\hbar/2)} \rho\left(\mathbf{r}_1 + \frac{\mathbf{q}\hbar}{2}, \mathbf{r}_1 - \frac{\mathbf{q}\hbar}{2}, t\right) = \\
 &\int \frac{d\mathbf{s} d\mathbf{q}}{(2\pi)^6} e^{-i\mathbf{p}\mathbf{q}} \int d\mathbf{r}_1 e^{i\mathbf{s}(\mathbf{r}_1 - \mathbf{r})} \rho\left(\mathbf{r}_1 + \frac{\mathbf{q}\hbar}{2}, \mathbf{r}_1 - \frac{\mathbf{q}\hbar}{2}, t\right) = \\
 &\int \frac{d\mathbf{q}}{(2\pi)^3} e^{-i\mathbf{p}\mathbf{q}} \int d\mathbf{r}_1 \delta(\mathbf{r}_1 - \mathbf{r}) \rho\left(\mathbf{r}_1 + \frac{\mathbf{q}\hbar}{2}, \mathbf{r}_1 - \frac{\mathbf{q}\hbar}{2}, t\right).
 \end{aligned}$$

By setting $\mathbf{q} = \mathbf{r}'/\hbar$ we arrive at the definition of the Wigner function:

$$I = \int \frac{d\mathbf{r}'}{(2\pi\hbar)^3} e^{-i\mathbf{p}\mathbf{r}'/\hbar} \rho\left(\mathbf{r} + \frac{\mathbf{r}'}{2}, \mathbf{r} - \frac{\mathbf{r}'}{2}, t\right) = f_W(\mathbf{r}, \mathbf{p}, t).$$

3. – Generalized Wigner function of the coupled electron-phonon system

We consider a system of a single electron (or equivalently many non-interacting electrons) subject to a structure potential and interaction with lattice vibrations. The description of the system is provided by both the electron and the phonon degrees of freedom. We first generalize the Wigner function and the Wigner equation for the coupled electron-phonon system. The Hamiltonian of the considered system is given by

$$\begin{aligned}
 (27) \quad H &= H_0 + V + H_p + H_{e-p} = \\
 &-\frac{\hbar^2}{2m} \nabla_{\mathbf{r}} + V(\mathbf{r}) + \sum_{\mathbf{q}} b_{\mathbf{q}}^\dagger b_{\mathbf{q}} \hbar \omega_{\mathbf{q}} + i\hbar \sum_{\mathbf{q}} F(\mathbf{q}) (b_{\mathbf{q}} e^{i\mathbf{q}\mathbf{r}} - b_{\mathbf{q}}^\dagger e^{-i\mathbf{q}\mathbf{r}}),
 \end{aligned}$$

where H_0 is the free-electron part, $V(\mathbf{r})$ the structure potential, H_p the free-phonon Hamiltonian and H_{e-p} the electron-phonon interaction. In the above expressions $b_{\mathbf{q}}^\dagger$ and $b_{\mathbf{q}}$ are the creation and annihilation operators for the phonon mode \mathbf{q} , $\omega_{\mathbf{q}}$ is the energy of that mode and $F(\mathbf{q})$ is the electron-phonon coupling element, which depends on the type of phonon scattering analyzed. The state of the phonon subsystem is presented by the set $\{n_{\mathbf{q}}\}$, where $n_{\mathbf{q}}$ is the occupation number of the phonons in mode \mathbf{q} . Then

the representation is given by the vectors $|\{n_{\mathbf{q}}\}, \mathbf{r}\rangle = |\{n_{\mathbf{q}}\}\rangle|\mathbf{r}\rangle$. The generalized Wigner function [8, 9] is defined by (the prefactor $(2\pi\hbar)^{-3}$ is omitted for convenience)

$$f_W(\mathbf{r}, \mathbf{p}, \{n_{\mathbf{q}}\}, \{n'_{\mathbf{q}}\}, t) = \int d\mathbf{r}' e^{-i\mathbf{p}\mathbf{r}'/\hbar} \left\langle \mathbf{r} + \frac{\mathbf{r}'}{2}, \{n_{\mathbf{q}}\} | \hat{\rho}_t | \{n'_{\mathbf{q}}\}, \mathbf{r} - \frac{\mathbf{r}'}{2} \right\rangle.$$

The equation of motion of f_W is obtained from (26):

$$\frac{\partial f_W(\mathbf{r}, \mathbf{p}, \{n_{\mathbf{q}}\}, \{n'_{\mathbf{q}}\}, t)}{\partial t} = \frac{1}{i\hbar} \int d\mathbf{r}' e^{-i\mathbf{p}\mathbf{r}'/\hbar} \left\langle \mathbf{r} + \frac{\mathbf{r}'}{2}, \{n_{\mathbf{q}}\} | [H, \hat{\rho}_t]_- | \{n'_{\mathbf{q}}\}, \mathbf{r} - \frac{\mathbf{r}'}{2} \right\rangle.$$

For convenience we denote the right-hand side of the above equation by $WT(H)$. In the following we evaluate $WT(H)$ for each term of the Hamiltonian (27). $WT(H_0 + V(\mathbf{r}))$ has already been evaluated within the steps following after (15). The only difference is that f_W in terms *I* and *II* now contains the phonon coordinates. The free-phonon term is readily evaluated to

$$WT(H_p) = \frac{1}{i\hbar} (\epsilon(\{n_{\mathbf{q}}\}) - \epsilon(\{n'_{\mathbf{q}}\})) f_W(\mathbf{r}, \mathbf{p}, \{n_{\mathbf{q}}\}, \{n'_{\mathbf{q}}\}, t),$$

where $\epsilon(\{n_{\mathbf{q}}\}) = \sum_{\mathbf{q}} n_{\mathbf{q}} \hbar \omega_{\mathbf{q}}$. $WT(H_{e-p})$ gives rise to four terms which are evaluated separately. Inserting $\int d\mathbf{r}'' |\mathbf{r}''\rangle \langle \mathbf{r}''|$ in the first one, we get

$$\begin{aligned} & \int d\mathbf{r}' \int d\mathbf{r}'' e^{-i\mathbf{p}\mathbf{r}'/\hbar} \left\langle \mathbf{r} + \frac{\mathbf{r}'}{2}, \{n_{\mathbf{q}}\} | b_{\mathbf{q}'} e^{i\mathbf{q}'\mathbf{r}''} | \mathbf{r}'' \rangle \langle \mathbf{r}'' | \hat{\rho}_t | \{n'_{\mathbf{q}}\}, \mathbf{r} - \frac{\mathbf{r}'}{2} \right\rangle = \\ & \sqrt{n_{\mathbf{q}'} + 1} \int d\mathbf{r}' e^{-i\mathbf{p}\mathbf{r}'/\hbar} e^{i\mathbf{q}'(\mathbf{r} + \mathbf{r}'/2)} \left\langle \mathbf{r} + \frac{\mathbf{r}'}{2}, \{n_1, \dots, n_{\mathbf{q}'} + 1, \dots\} | \hat{\rho}_t | \{n'_{\mathbf{q}}\}, \mathbf{r} - \frac{\mathbf{r}'}{2} \right\rangle = \\ & \sqrt{n_{\mathbf{q}'} + 1} e^{i\mathbf{q}'\mathbf{r}} f_W \left(\mathbf{r}, \mathbf{p} - \frac{\hbar \mathbf{q}'}{2}, \{n_1, \dots, n_{\mathbf{q}'} + 1, \dots\}, \{n'_{\mathbf{q}}\}, t \right), \end{aligned}$$

where we have used the orthonormality relation $\langle \mathbf{r} | \mathbf{r}' \rangle = \delta(\mathbf{r} - \mathbf{r}')$ and $b_{\mathbf{q}}$ becomes a creation operator when operating to the left. The remaining terms are evaluated in a similar way. We are ready to formulate the generalized Wigner equation. The following short notations are used: $\{n_{\mathbf{q}}\}_{\mathbf{q}'}^+$, $(\{n_{\mathbf{q}}\}_{\mathbf{q}'}^-)$ are states of the phonon subsystem, obtained

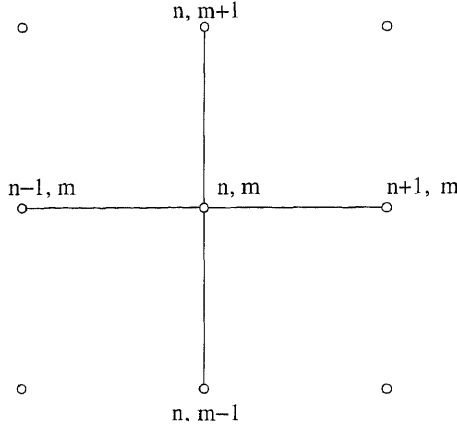


Fig. 5. – Coupled elements.

from $\{n_{\mathbf{q}}\}$ by increasing (decreasing) the number of phonons in the mode \mathbf{q}' by unity:

$$\begin{aligned}
 (28) \quad & \left(\frac{\partial}{\partial t} + \frac{\mathbf{p}}{m} \cdot \nabla_{\mathbf{r}} \right) f_{\mathbf{W}}(\mathbf{r}, \mathbf{p}, \{n_{\mathbf{q}}\}, \{n'_{\mathbf{q}}\}, t) = \\
 & \frac{1}{i\hbar} (\epsilon(\{n_{\mathbf{q}}\}) - \epsilon(\{n'_{\mathbf{q}}\})) f_{\mathbf{W}}(\mathbf{r}, \mathbf{p}, \{n_{\mathbf{q}}\}, \{n'_{\mathbf{q}}\}, t) + \\
 & + \int d\mathbf{p}' V_{\mathbf{W}}(\mathbf{r}, \mathbf{p}' - \mathbf{p}) f_{\mathbf{W}}(\mathbf{r}, \mathbf{p}', \{n_{\mathbf{q}}\}, \{n'_{\mathbf{q}}\}, t) + \\
 & + \sum_{\mathbf{q}'} F(\mathbf{q}') \left\{ e^{i\mathbf{q}' \cdot \mathbf{r}} \sqrt{n_{\mathbf{q}'} + 1} f_{\mathbf{W}}\left(\mathbf{r}, \mathbf{p} - \frac{\hbar \mathbf{q}'}{2}, \{n_{\mathbf{q}}\}_{\mathbf{q}'}^+, \{n'_{\mathbf{q}}\}, t\right) - \right. \\
 & - e^{-i\mathbf{q}' \cdot \mathbf{r}} \sqrt{n_{\mathbf{q}'}} f_{\mathbf{W}}\left(\mathbf{r}, \mathbf{p} + \frac{\hbar \mathbf{q}'}{2}, \{n_{\mathbf{q}}\}_{\mathbf{q}'}^-, \{n'_{\mathbf{q}}\}, t\right) - \\
 & - e^{i\mathbf{q}' \cdot \mathbf{r}} \sqrt{n'_{\mathbf{q}'}} f_{\mathbf{W}}\left(\mathbf{r}, \mathbf{p} + \frac{\hbar \mathbf{q}'}{2}, \{n_{\mathbf{q}}\}, \{n'_{\mathbf{q}}\}_{\mathbf{q}'}^-, t\right) + \\
 & \left. + e^{-i\mathbf{q}' \cdot \mathbf{r}} \sqrt{n'_{\mathbf{q}'} + 1} f_{\mathbf{W}}\left(\mathbf{r}, \mathbf{p} - \frac{\hbar \mathbf{q}'}{2}, \{n_{\mathbf{q}}\}, \{n'_{\mathbf{q}}\}_{\mathbf{q}'}^+, t\right) \right\}.
 \end{aligned}$$

The generalized Wigner equation couples an element $f_{\mathbf{W}}(\dots, \{n\}, \{m\}, t)$ to four neighborhood elements, as shown in fig. 5 for any phonon mode \mathbf{q} . For any such mode, $n_{\mathbf{q}}$ can be any integer between 0 and infinity and the sum over \mathbf{q} couples all modes.

The last two terms in the curly brackets can be obtained by the following operation applied to the first two terms:

- i changes the sign;
- instead of in the left state the phonon number in the mode determined by the summation index (\mathbf{q}') is changed in the right state;

– in the square roots $n_{\mathbf{q}'}$ is replaced with $n'_{\mathbf{q}'}$.

In what follows we denote the last two terms by i.c.

Consider the evolution of an initial state of the system defined at time $t = 0$ by the function $f_W(\mathbf{r}, \mathbf{p}, \{n_{\mathbf{q}}\}, \{n'_{\mathbf{q}}\}, 0)$. It is convenient to use the integral form of (28), which includes explicitly the initial condition:

$$\begin{aligned}
 (29) \quad f_W(\mathbf{r}, \mathbf{p}, \{n_{\mathbf{q}}\}, \{n'_{\mathbf{q}}\}, t) &= f_W(\mathbf{r}_{(\mathbf{p}, 0)}, \mathbf{p}, \{n_{\mathbf{q}}\}, \{n'_{\mathbf{q}}\}, 0) e^{-i/\hbar(\epsilon(\{n_{\mathbf{q}}\}) - \epsilon(\{n'_{\mathbf{q}}\}))t} + \\
 &+ \int_0^t dt' e^{-i/\hbar(\epsilon(\{n_{\mathbf{q}}\}) - \epsilon(\{n'_{\mathbf{q}}\}))(t-t')} \times \\
 &\times \left[\int d\mathbf{p}' V_W(\mathbf{r}_{(\mathbf{p}, t')}, \mathbf{p}' - \mathbf{p}) f_W(\mathbf{r}_{(\mathbf{p}, t')}, \mathbf{p}', \{n_{\mathbf{q}}\}, \{n'_{\mathbf{q}}\}, t') + \right. \\
 &+ \sum_{\mathbf{q}'} F(\mathbf{q}') \left\{ e^{i\mathbf{q}' \cdot \mathbf{r}_{(\mathbf{p}, t')}} \sqrt{n_{\mathbf{q}'} + 1} f_W\left(\mathbf{r}_{(\mathbf{p}, t')}, \mathbf{p} - \frac{\hbar \mathbf{q}'}{2}, \{n_{\mathbf{q}}\}_{\mathbf{q}'}^+, \{n'_{\mathbf{q}}\}, t'\right) - \right. \\
 &\left. \left. - e^{-i\mathbf{q}' \cdot \mathbf{r}_{(\mathbf{p}, t')}} \sqrt{n_{\mathbf{q}'} } f_W\left(\mathbf{r}_{(\mathbf{p}, t')}, \mathbf{p} + \frac{\hbar \mathbf{q}'}{2}, \{n_{\mathbf{q}}\}_{\mathbf{q}'}^-, \{n'_{\mathbf{q}}\}, t'\right) + \text{i.c.} \right\} \right].
 \end{aligned}$$

Here $\mathbf{r}_{(\mathbf{p}, t')} = \mathbf{r} - (\mathbf{p}/m)(t - t')$ is a classical trajectory initialized by \mathbf{r}, \mathbf{p} at time t . The integral form can be proved by taking the time derivative of (29) which should lead us to (28). To see this, we rewrite (29) by keeping only the relevant variables and writing explicitly the time dependence of the trajectory: $\mathbf{r}(t - t') = \mathbf{r}_{(\mathbf{p}, t')} = \mathbf{r} - (\mathbf{p}/m)(t - t')$,

$$f(\mathbf{r}(t - t''), t'')|_{t''=t} = f(\mathbf{r}(0), 0) e^{\lambda t''}|_{t''=t} + \int_0^{t''} dt' K(\mathbf{r}(t - t')) e^{\lambda(t'' - t')} f(\mathbf{r}(t - t'), t')|_{t''=t}.$$

The equation is written in this way to remind that the time derivative is taken over the trajectory, so that we have to differentiate with respect to t'' and set $t'' = t$ in the final result. The left-hand side readily gives the fieldless Liouville operator acting on f , while the right-hand side gives $\lambda f(\mathbf{r}, t) + K(\mathbf{r})f(\mathbf{r}, t)$, compare (28).

4. – Weak coupling and equilibrium phonons approximations

Of interest is the reduced Wigner function, which is obtained from the generalized Wigner function by taking the trace over the phonon states. An exact equation for the reduced Wigner function cannot be obtained since the trace operation does not commute with the electron-phonon interaction Hamiltonian. Derived is a model which approximates the generalized Wigner equation, but is closed with respect to the reduced Wigner function. The model is general enough to account for the quantum character of the interaction with the phonons. The electron-device potential part of the transport is treated on a rigorous quantum level. The model further gives rise to simplified model equations which end up with the classical Boltzmann equation. The derivation introduces a consistent hierarchy of assumptions and simplifications.

4.1. *Weak coupling.* – We begin with the assumptions which simplify (29) towards a model equation set for the electron Wigner function. Of interest are the diagonal elements of the generalized WF. Thus the initial condition is assumed diagonal with respect to the phonon coordinates, which corresponds to the evolution process of an initially decoupled electron-phonon system:

$$\begin{aligned}
 (30) \quad f_W(\mathbf{r}, \mathbf{p}, \{n_{\mathbf{q}}\}, \{n_{\mathbf{q}}\}, t) &= f_W(\mathbf{r}_{(\mathbf{p},0)}, \mathbf{p}, \{n_{\mathbf{q}}\}, \{n_{\mathbf{q}}\}, 0) + \\
 &+ \int_0^t dt' \left[\int d\mathbf{p}' V_W(\mathbf{r}_{(\mathbf{p},t')}, \mathbf{p}' - \mathbf{p}) f_W(\mathbf{r}_{(\mathbf{p},t')}, \mathbf{p}', \{n_{\mathbf{q}}\}, \{n_{\mathbf{q}}\}, t') + \right. \\
 &+ \sum_{\mathbf{q}'} F(\mathbf{q}') \left\{ e^{i\mathbf{q}' \cdot \mathbf{r}_{(\mathbf{p},t')}} \sqrt{n_{\mathbf{q}'} + 1} f_W\left(\mathbf{r}_{(\mathbf{p},t')}, \mathbf{p} - \frac{\hbar \mathbf{q}'}{2}, \{n_{\mathbf{q}}\}_{\mathbf{q}'}^+, \{n_{\mathbf{q}}\}, t'\right) - \right. \\
 &\left. \left. - e^{-i\mathbf{q}' \cdot \mathbf{r}_{(\mathbf{p},t')}} \sqrt{n_{\mathbf{q}'}} f_W\left(\mathbf{r}_{(\mathbf{p},t')}, \mathbf{p} + \frac{\hbar \mathbf{q}'}{2}, \{n_{\mathbf{q}}\}_{\mathbf{q}'}^-, \{n_{\mathbf{q}}\}, t'\right) + \text{i.c.} \right\} \right].
 \end{aligned}$$

A diagonal element is linked to elements, called first off-diagonal elements, which are diagonal in all modes but the current mode \mathbf{q}' of the summation. In this mode the four neighbors of $n_{\mathbf{q}'}$, $n_{\mathbf{q}'}$, namely $n_{\mathbf{q}'} \pm 1$, $n_{\mathbf{q}'}$ and $n_{\mathbf{q}'}$, $n_{\mathbf{q}'} \pm 1$, are concerned. This is schematically presented in fig. 6. According to (29), the first off-diagonal elements are linked to elements which in general are placed further away from the diagonal ones by increasing or decreasing the phonon number in a second mode, \mathbf{q}'' , by unity. These are the second off-diagonal elements. The only exception is provided by two contributions which recover diagonal elements. They are obtained from $(\{\{n_{\mathbf{q}}\}_{\mathbf{q}'}^+\}_{\mathbf{q}''}^-, \{n_{\mathbf{q}}\})$ and $(\{n_{\mathbf{q}}\}_{\mathbf{q}'}^+, \{n_{\mathbf{q}}\}_{\mathbf{q}''}^+)$ in the case when the two phonon modes coincide: $\mathbf{q}' = \mathbf{q}''$. Note that each link of two elements corresponds to a multiplication by the factor F .

The next assumption is that F is a small quantity. While the first off-diagonal elements give contributions to (30) by order of F^2 , the second off-diagonal elements give rise to higher-order contributions and can be neglected. The physical meaning of the assumption is that the interaction with a phonon in mode \mathbf{q}' which begins from a diagonal element completes at a diagonal element by another interaction with the phonon in the same mode, without any interference with phonons of other modes. The assumption allows to truncate the considered elements to those between the two lines parallel to the main diagonal in fig. 6. The corresponding equations for the first two off-diagonal terms in (30) are obtained by using (29). The remaining two elements, which compose the i.c. term in (30) give rise to two integral equations which are complex conjugate to the first two. In this way the relevant information is provided by (30) and the integral equations for the first two elements. The latter will not be stated explicitly, but only their peculiarities will be discussed. As the initial condition is assumed diagonal, it does not appear in these equations. The role of an initial condition is played by the terms containing the diagonal elements. It is important to note that the pre-factor of the first off-diagonal element $f_W(\dots, \{n_{\mathbf{q}}\}_{\mathbf{q}'}^+, \dots)$ in (30) is $\sqrt{n_{\mathbf{q}'} + 1}$. The same pre-factor appears in front of the diagonal terms in the equation for the first off-diagonal element. The same holds for $f_W(\dots, \{n_{\mathbf{q}}\}_{\mathbf{q}'}^-, \dots)$ which has a pre-factor $\sqrt{n_{\mathbf{q}'}}$. The two diagonal

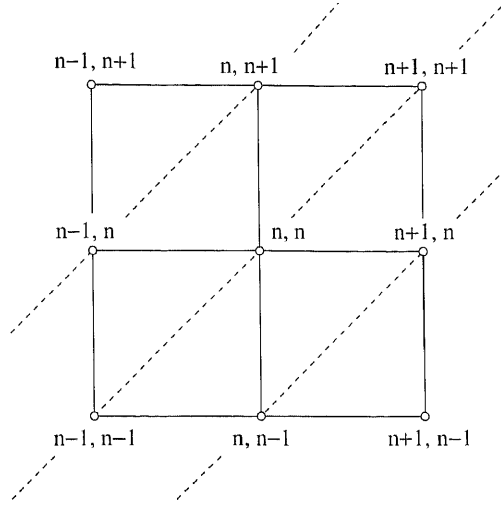


Fig. 6. – Diagonal and first off-diagonal elements.

terms in the equation for $f_W(\dots, \{n_{\mathbf{q}}\}_{\mathbf{q}'}^-, \dots)$ contain the same pre-factor $\sqrt{n_{\mathbf{q}'}}$. This feature will be used below, when we consider the trace over the phonon coordinates.

4.2. Equilibrium phonon averages. – The set of eq. (30) and the two equations for the first off-diagonal elements constitutes the basis for the derivation of the principal transport model for the reduced WF. The set is still infinite with respect to the phonon coordinates, which are to be eliminated by the trace operation. The next assumption is that the phonon system is a thermostat for the electrons, *i.e.* the phonon distribution remains in equilibrium during the evolution:

$$P(n_{\mathbf{q}}, t') = \int d\mathbf{r} \int d\mathbf{p} \sum'_{\{n_{\mathbf{q}'}\}} f_W(\mathbf{r}(\mathbf{p}, t'), \mathbf{p}, \{n_{\mathbf{q}'}\}, \{n_{\mathbf{q}'}\}, t') = P_{\text{eq}}(n_{\mathbf{q}}) = \frac{e^{-\hbar\omega_{\mathbf{q}}n_{\mathbf{q}}/kT}}{n(\mathbf{q}) + 1}.$$

Here $P(n_{\mathbf{q}}, t')$ is the probability of finding $n_{\mathbf{q}}$ phonons in mode \mathbf{q} at time t' , \sum' denotes summation over all phonon coordinates but the one in mode \mathbf{q} , and $n(\mathbf{q})$ is the mean equilibrium phonon number (Bose distribution),

$$(31) \quad n(\mathbf{q}) = \sum_{n_{\mathbf{q}}=0}^{\infty} n_{\mathbf{q}} P_{\text{eq}}(n_{\mathbf{q}}) = \frac{1}{e^{\hbar\omega_{\mathbf{q}}/kT} - 1}, \quad \sum_{n_{\mathbf{q}}=0}^{\infty} P_{\text{eq}}(n_{\mathbf{q}}) = 1.$$

If there exists a closed model for the reduced or electron Wigner function $f(\mathbf{r}, \mathbf{p}, t')$ which is equivalent to the obtained set of three equations for the generalized WF, then for any

time $0 \leq t' \leq t$ the following must hold:

$$(32) \quad f_W(\mathbf{r}, \mathbf{p}, \{n_{\mathbf{q}}\}, \{n_{\mathbf{q}}\}, t') = f(\mathbf{r}, \mathbf{p}, t') \prod_{\mathbf{q}} P_{\text{eq}}(n_{\mathbf{q}}).$$

This property holds for the initial condition which is the free term in (30). We assume that (32) holds until time t and investigate if it remains true after a time step Δt . From (30) we obtain

$$(33) \quad f_W(\dots, \{n_{\mathbf{q}}\}, \{n_{\mathbf{q}}\}, t + \Delta t) = (\dots) f_W(\dots, \{n_{\mathbf{q}}\}, \{n_{\mathbf{q}}\}, t) + \Delta t \sum_{\mathbf{q}'} F(\mathbf{q}') \{\dots\}.$$

The first term on the right satisfies (32) and we pursue the second term, which refers to the off-diagonal elements $f_W(\dots, \{n_{\mathbf{q}}\}_{\mathbf{q}'}^{\pm}, \dots)$. It can be shown that the second bracket includes terms which depend on the phonon coordinates by the following factors:

$$(34) \quad \begin{aligned} & (n_{\mathbf{q}'} + 1) P_{\text{eq}}(n_{\mathbf{q}'}) \prod_{\mathbf{q}}' P_{\text{eq}}(n_{\mathbf{q}}), & (n_{\mathbf{q}'} + 1) P_{\text{eq}}(n_{\mathbf{q}'} + 1) \prod_{\mathbf{q}}' P_{\text{eq}}(n_{\mathbf{q}}), \\ & n_{\mathbf{q}'} P_{\text{eq}}(n_{\mathbf{q}'}) \prod_{\mathbf{q}}' P_{\text{eq}}(n_{\mathbf{q}}), & n_{\mathbf{q}'} P_{\text{eq}}(n_{\mathbf{q}'} - 1) \prod_{\mathbf{q}}' P_{\text{eq}}(n_{\mathbf{q}}). \end{aligned}$$

We first conclude that (32) is violated after the time step, since, according to (34), the phonons in mode \mathbf{q}' are driven out of equilibrium. Here in help comes the assumption that the phonon system stays in equilibrium. It is convenient to think that the phonon system immediately recovers the equilibrium after the interaction.

Now we can take the trace over the phonon coordinates. The trace means to sum over $n_{\mathbf{q}}$ for all modes \mathbf{q} , which is readily done with the help of (31) and the following equalities:

$$n(\mathbf{q}) = \sum_{n_{\mathbf{q}}} (n_{\mathbf{q}} + 1) P_{\text{eq}}(n_{\mathbf{q}} + 1), \quad n(\mathbf{q}) + 1 = \sum_{n_{\mathbf{q}}} n_{\mathbf{q}} P_{\text{eq}}(n_{\mathbf{q}} - 1).$$

As a result, the phonon coordinates are replaced by the following numbers in the second bracket, compare with (34):

$$\begin{aligned} & (n(\mathbf{q}') + 1), & n(\mathbf{q}'), \\ & n(\mathbf{q}'), & (n(\mathbf{q}') + 1). \end{aligned}$$

This is an important step which allows to close the equation set for the electron Wigner function.

5. – Models for the electron Wigner function

5.1. *Main model for the electron Wigner function.* – In the obtained set of equations the reduced Wigner function f_W is coupled to two auxiliary functions f_1 and f_2 :

$$(35) \quad f_W(\mathbf{r}, \mathbf{p}, t) = f_W(\mathbf{r}_{(\mathbf{p},0)}, \mathbf{p}, 0) + \\ + \int_0^t dt' \left[\int d\mathbf{p}' V_W(\mathbf{r}_{(\mathbf{p},t')}, \mathbf{p}' - \mathbf{p}) f_W(\mathbf{r}_{(\mathbf{p},t')}, \mathbf{p}', t') + \right. \\ + \sum_{\mathbf{q}'} F^2(\mathbf{q}') \left\{ e^{i\mathbf{q}'\mathbf{r}_{(\mathbf{p},t')}} f_1\left(\mathbf{r}_{(\mathbf{p},t')}, \mathbf{p} - \frac{\hbar\mathbf{q}'}{2}, t'\right) - \right. \\ \left. \left. - e^{-i\mathbf{q}'\mathbf{r}_{(\mathbf{p},t')}} f_2\left(\mathbf{r}_{(\mathbf{p},t')}, \mathbf{p} + \frac{\hbar\mathbf{q}'}{2}, t'\right) + \text{c.c.} \right\} \right],$$

$$(36) \quad f_1\left(\mathbf{r}', \mathbf{p} - \frac{\hbar\mathbf{q}'}{2}, t'\right) = \\ \int_0^{t'} dt'' e^{-i\omega_{\mathbf{q}'}(t'-t'')} \left[\int d\mathbf{p}' V_W\left(\mathbf{r}'_{(\mathbf{p},\mathbf{q}',t'')}, \mathbf{p}' - \mathbf{p} + \frac{\hbar\mathbf{q}'}{2}\right) \times \right. \\ \times f_1(\mathbf{r}'_{(\mathbf{p},\mathbf{q}',t'')}, \mathbf{p}', t'') - e^{-i\mathbf{q}'\mathbf{r}'_{(\mathbf{p},\mathbf{q}',t'')}} \left\{ (n(\mathbf{q}') + 1) f_W(\mathbf{r}'_{(\mathbf{p},\mathbf{q}',t'')}, \mathbf{p}, t'') - \right. \\ \left. \left. - n(\mathbf{q}') f_W(\mathbf{r}'_{(\mathbf{p},\mathbf{q}',t'')}, \mathbf{p} - \hbar\mathbf{q}', t'') \right\} \right],$$

$$(37) \quad f_2\left(\mathbf{r}', \mathbf{p} + \frac{\hbar\mathbf{q}'}{2}, t'\right) = \\ \int_0^{t'} dt'' e^{i\omega_{\mathbf{q}'}(t'-t'')} \left[\int d\mathbf{p}' V_W\left(\mathbf{r}'_{(\mathbf{p},-\mathbf{q}',t'')}, \mathbf{p}' - \mathbf{p} - \frac{\hbar\mathbf{q}'}{2}\right) \times \right. \\ \times f_2(\mathbf{r}'_{(\mathbf{p},-\mathbf{q}',t'')}, \mathbf{p}', t'') + e^{i\mathbf{q}'\mathbf{r}'_{(\mathbf{p},-\mathbf{q}',t'')}} \left\{ n(\mathbf{q}') f_W(\mathbf{r}'_{(\mathbf{p},\mathbf{q}',t'')}, \mathbf{p}, t'') - \right. \\ \left. \left. - (n(\mathbf{q}') + 1) f_W(\mathbf{r}'_{(\mathbf{p},-\mathbf{q}',t'')}, \mathbf{p} + \hbar\mathbf{q}', t'') \right\} \right],$$

where $\mathbf{r}'_{(\mathbf{p},\mathbf{q}',t'')} = \mathbf{r}' - ((\mathbf{p} - \hbar\mathbf{q}'/2)/m)(t' - t'')$. Equations (36) and (37) are two integral equations with free terms given by the curly brackets (with the corresponding pre-factors and integrated over t''). The solutions f_1 and f_2 can be formally presented as a series of consecutive iterations of the kernel (the term with V_W) on the free terms. A replacement in (35) leads to a closed equation for f_W , which is the main model for the electron Wigner function. This equation is still too complicated and needs further approximations.

5.2. *The Levinson equation.* – An explicit solution of (36) and (37) is possible if the system is considered in a homogeneous semiconductor with an applied constant field \mathbf{F} . According to (23), the Wigner potential term becomes proportional to $\mathbf{F} \cdot \nabla_{\mathbf{p}}$. A single equation is obtained after few steps of transformations. First, eqs. (35) to (37) are differentiated with respect to time to obtain the integro-differential form. The term

containing \mathbf{F} is transferred to the left which completes the Liouville operator as in (24). It can be shown that, by using the following Ansatz:

$$\begin{aligned} f_1\left(\mathbf{r}', \mathbf{p} - \frac{\hbar \mathbf{q}'}{2}, t'\right) &= e^{-i\mathbf{q}' \cdot \mathbf{r}'} f_1\left(\mathbf{p} - \frac{\hbar \mathbf{q}'}{2}, t'\right), \\ f_2\left(\mathbf{r}', \mathbf{p} + \frac{\hbar \mathbf{q}'}{2}, t'\right) &= e^{i\mathbf{q}' \cdot \mathbf{r}'} f_2\left(\mathbf{p} + \frac{\hbar \mathbf{q}'}{2}, t'\right), \end{aligned}$$

the relation $(2\mathbf{q}' \cdot \mathbf{p} - \hbar \mathbf{q}'^2)/2m = (1/\hbar)(\epsilon(\mathbf{p}) - \epsilon(\mathbf{p} - \hbar \mathbf{q}'))$ and assuming a homogeneous initial condition, the space coordinate \mathbf{r} becomes irrelevant and can be omitted. The two auxiliary equations are transformed back into integral equations where, now, the field determines the trajectories: $\mathbf{p}_{(\tau)} = \mathbf{p} - \mathbf{F}(t - \tau)$. In this way f_1 and f_2 are expressed in terms of f_W . A replacement in the differential form of (35) along with the transformation $\sum_{\mathbf{q}'} \rightarrow \sum_{\mathbf{p}'} = (V/h^3) \int d\mathbf{p}'$ gives rise to the equation

$$\begin{aligned} (38) \quad \left(\frac{\partial}{\partial t} + e\mathbf{E} \cdot \nabla_{\mathbf{p}} \right) f_W(\mathbf{p}, t) &= \int_0^t dt' \int d\mathbf{p}' \times \\ &\quad \times \left(S(\mathbf{p}', \mathbf{p}, t, t') f_W(\mathbf{p}'_{(t')}, t') - S(\mathbf{p}, \mathbf{p}', t, t') f_W(\mathbf{p}_{(t')}, t') \right), \\ S(\mathbf{p}', \mathbf{p}, t, t') &= \frac{2V}{h^3} F^2(\mathbf{q}') \left[(n(\mathbf{q}') + 1) \cos \left(\int_{t'}^t d\tau \Omega(\mathbf{p}_{(\tau)}, \mathbf{p}'_{(\tau)}) \right) + \right. \\ &\quad \left. + n(\mathbf{q}') \cos \left(\int_{t'}^t d\tau \Omega(\mathbf{p}'_{(\tau)}, \mathbf{p}_{(\tau)}) \right) \right], \\ \Omega(\mathbf{p}_{(\tau)}, \mathbf{p}'_{(\tau)}) &= \frac{\epsilon(\mathbf{p}_{(\tau)}) - \epsilon(\mathbf{p}'_{(\tau)}) + \hbar \omega_{\mathbf{q}'}}{\hbar}, \quad \mathbf{q}' = \frac{\mathbf{p} - \mathbf{p}'}{\hbar}. \end{aligned}$$

This equation has been derived by Levinson from the homogeneous electron-phonon density matrix [10]. Barker and Ferry derived an equation which is equivalent to (38) with the only difference of an exponential damping factor in S which accounts for the finite lifetime of the carrier [11]. The Barker-Ferry equation can be derived [12] also from the single band model of the electron-phonon quantum kinetics in pulse excited semiconductors proposed in [13]. Equation (38) reveals interesting quantum effects of collisional broadening, retardation and the intra-collisional field effect. These effects have been theoretically predicted in quantum regimes of the electron-phonon interaction [14, 15].

The used physical model is for GaAs material with a PO phonon with constant energy $\hbar\omega$. The initial condition is a Gaussian function of the energy. A choice of a very low temperature, when the physical system has a transparent semiclassical behavior, allows conveniently to study quantum effects.

5.2.1. Collisional broadening and retardation. The effects of collisional broadening and retardation exist already at zero electric field. In this case it is convenient to use the spherical symmetry of the task. The equation is transformed into spherical coordinates (k, θ, ϕ) and integrated over θ and ϕ . The solution depends now only on $k = |\mathbf{p}|/\hbar$

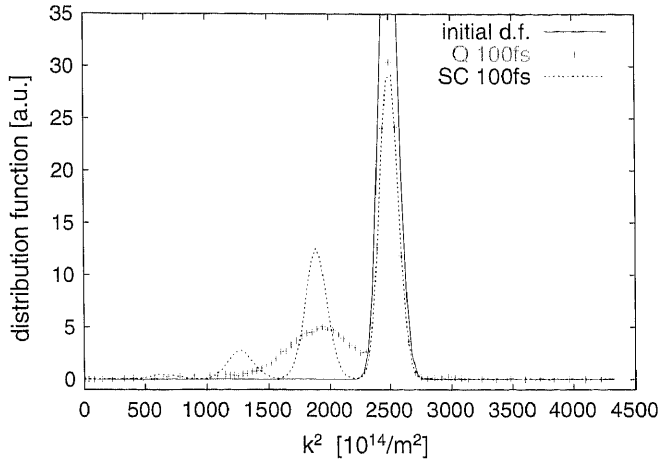


Fig. 7. – Initial distribution function (initial d.f.), semiclassical (SC) and quantum (Q) solutions $kf(0, k, t)$ for 100 fs evolution time at zero electric field.

and the time. Figures 7, 8 and 9 present snapshots of the evolution of the semiclassical and quantum solutions $kf(0, k, t)$ for times 100 fs, 200 fs and 400 fs as a function of k^2 . The quantity k^2 is proportional to the electron energy in units 10^{14} m^{-2} . Semiclassical electrons can only emit phonons and lose energy equal to a multiple of the phonon energy $\hbar\omega$. They evolve according to a distribution, patterned by replicas of the initial condition shifted towards low energies. The electrons cannot appear in the region above the initial distribution. The quantum solutions demonstrate two effects of deviation

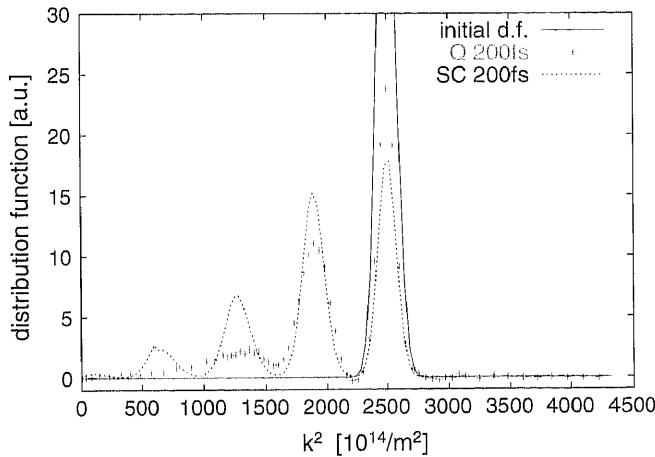


Fig. 8. – Initial distribution function (initial d.f.), semiclassical (SC) and quantum (Q) solutions $kf(0, k, t)$ for 200 fs evolution time at zero electric field.

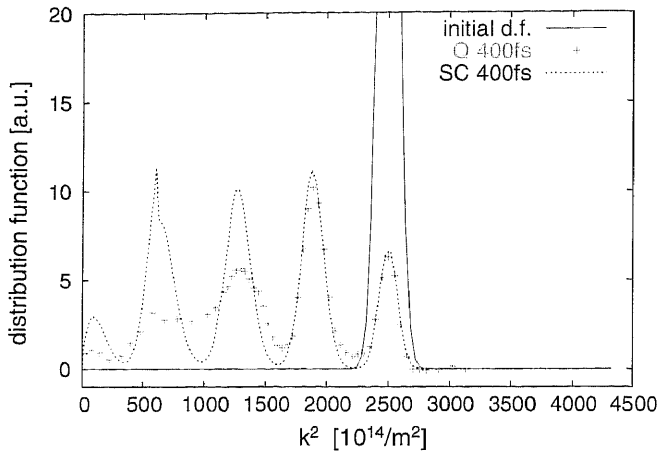


Fig. 9. – Initial distribution function (initial d.f.), semiclassical (SC) and quantum (Q) solutions $kf(0, k, t)$ for 400 fs evolution time at zero electric field.

from the semiclassical behavior. There is a retardation in the build-up of the remote peaks with respect to the initial-condition peaks. The replicas are broadened and the broadening increases with the distance to the initial peak. The broadening is due to the lack of energy conservation in the interaction. At low evolution times the cosine function in eq. (38) weakly depends on the phase space variables. With the increase of time, the cosine term becomes a sharper function of these variables and in the long time limit tends to the semiclassical delta-function. Accordingly, the first replica of the 100 fs is broadened. It resembles the corresponding replica of the semiclassical solution after 400 fs evolution time. The retardation of the quantum solutions is associated with the memory character of the equation. The time integral on the right-hand side of eq. (38) causes the delay of the build-up of the replicas.

5.2.2. The intra-collisional field effect (ICFE). We first discuss some numerical aspects of the task posed by the applied electric field. The field destroys the spherical symmetry of the equation. The dimensionality of the task increases and furthermore problems with the simulation domain arise due to the correlation of the phase space and time coordinates. The solution for a phase space point \mathbf{p} at instant t is related to the solutions at shifted points $\mathbf{p} - \mathbf{F}(t - t'')$. The shift depends on the electric field and the time interval $0 \leq t'' \leq t$ and hence no general integration domain can be specified in the phase space. This problem is solved by a transformation to a coordinate system moving with the field. A new variable p^t and function f^t are introduced such that

$$\mathbf{p}_1^t = \mathbf{p}_1 - \mathbf{F}t, \quad \mathbf{p}_1^t(\tau) = \mathbf{p}_1^t + \mathbf{F}\tau, \quad f(\mathbf{p}, t) = f(\mathbf{p}^t + \mathbf{F}t, t) \stackrel{\text{def}}{=} f^t(\mathbf{p}^t, t),$$

where \mathbf{p}_1 stands for \mathbf{p} and \mathbf{p}' , respectively. f_W on the right-hand side of (38) becomes

$$f(\mathbf{p}_{1,(\nu)}, t') = f(\mathbf{p}_1^t + \mathbf{F}t', t') = f^t(\mathbf{p}_1^t, t').$$

The transformation decouples the phase space and time arguments of the cosine functions in S according to

$$\epsilon(\mathbf{p}'(\tau)) - \epsilon(\mathbf{p}(\tau)) = \epsilon(\mathbf{p}^t) - \epsilon(\mathbf{p}^t) + 2\hbar\mathcal{F}(\mathbf{q}')\tau, \quad \mathcal{F}(\mathbf{q}) = \frac{\hbar}{2m}\mathbf{q} \cdot \mathbf{F}.$$

Due to this transform, the integral form of (38) simplifies to (the superscript t is omitted)

$$(39) \quad f(\mathbf{p}, t) = \phi(\mathbf{p}) + \int_0^t dt' \int_0^{t'} dt'' \int d\mathbf{p}' \{ S(\mathbf{p}', \mathbf{p}, t', t'') f(\mathbf{p}', t'') - \\ - S(\mathbf{p}, \mathbf{p}', t', t'') f(\mathbf{p}, t'') \},$$

where ϕ denotes the initial condition. The symmetry around the direction of the electric field can be used to reduce the variables in the equation. In cylindrical coordinates (r, k, θ) with r chosen normal to the field direction, the relevant variables become $x = (r, k)$, where x is a two-dimensional point. For zero lattice temperature ($n_q = 0$) the equation obtained reads

$$(40) \quad f(x, t) = \phi(x) + \int_0^t dt'' \int_G dx' \left[K(x, x') \times \right. \\ \left. \times \left\{ \int_{t''}^t dt' S_1(x, x', t', t'') \right\} f(x', t'') + \left\{ \int_{t''}^t dt' S_2(x, x', t', t'') \right\} f(x, t'') \right],$$

where $x \in G = (0, Q) \times (-Q, Q)$, Q determines the simulation domain and

$$(41) \quad K(x, x') = K(r, r', k, k') = \frac{\mathcal{G}r'}{\sqrt{((r - r')^2 + (k' - k)^2)((r + r')^2 + (k' - k)^2)}}, \\ S_1(x, x', t', t'') = -S_2(x', x, t', t'') = \cos \left(\left(\Omega(x, x') - \frac{\hbar}{2m} F(k' - k)(t' + t'') \right) (t' - t'') \right),$$

where \mathcal{G} is a constant. Note that K has a pole at $x = x'$. Equation (40) poses severe memory requirements for the deterministic numerical approaches. It has been solved by a randomized iterative Monte Carlo algorithm described in [16]. The solutions are obtained on cut lines parallel to the field ($k > 0, r = 0$), opposite to the field ($k < 0, r = 0$) and normal to the field ($k = 0, r > 0$). Figure 10 compares the 200 fs solutions as a function of $k < 0$ for different positive values of the electric force \mathbf{E} . The first replica peaks of the 6 and 12 kV/cm solutions are shifted in the field direction. For negative states the distance to the initial peak increases. Moreover, the solution in the classically forbidden region, to the right of the initial condition, demonstrates enhancement of the electron population. This effects can be associated with the structure of the first kernel of (39),

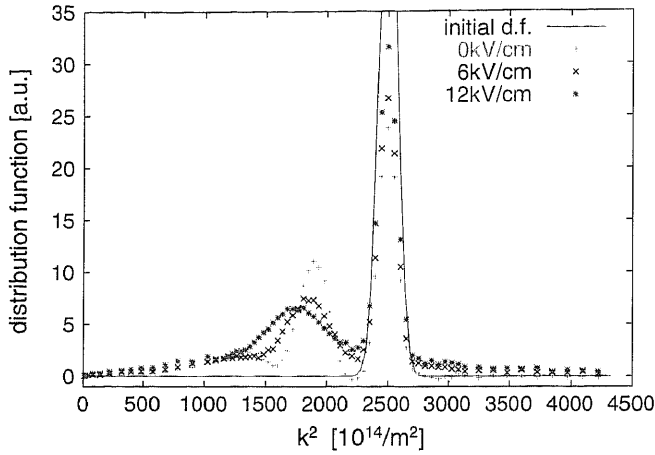


Fig. 10. – Solutions $|k|f(0, k, t)$, at negative k values, and evolution time 200 fs. The electric field of 0, 6 kV/cm, and 12 kV/cm points to the left.

which controls the electron transfer between the states. Responsible for the build-up of the peak is the first iteration term, obtained by replacing f with the initial condition ϕ . The cosine in eq. (41) has a permanent contribution to the solution if the pre-factor of $(t' - t'')$ is around zero. States with k' to the right of the k region of the first peak become important. For such states $k' - k < 0$ (the negative k region is examined) and since F is positive the energy of the field is added to the phonon energy. Accordingly, the solution behaves as if in the presence of a phonon with energy higher than $\hbar\omega$; the distance between the first replica and the initial condition increases. In the classically forbidden region, $k' - k > 0$, so that the energy of the field is subtracted from the phonon energy. The pre-factor is small for states k close to the k' region of the initial condition. Accordingly, the electron population in the vicinity to the right of the initial condition increases.

A comparison of the first replicas and the main peaks under the initial condition shows that the field has a pronounced influence on the collisional broadening and retardation. As demonstrated by additional numerical experiments, this effects depends on the field strength and direction.

From the above considerations it follows that just opposite effects should appear in the region of positive k values. Indeed, the first peaks in fig. 11 are shifted to the right since now $k' - k > 0$ and the energy of the field is subtracted from the phonon energy. In the semi-classically forbidden region, to the right of the initial condition, the pre-factor is large and there is no enhancement of the electron population.

No shift in the replicas of the solutions in the direction normal to the field should exist, due to the symmetry of the task. This can be used as a particular test for the consistency of the approach. As can be seen from fig. 12, the distance between the first replica peaks and the main peak does not depend on the field. Nevertheless, the field has

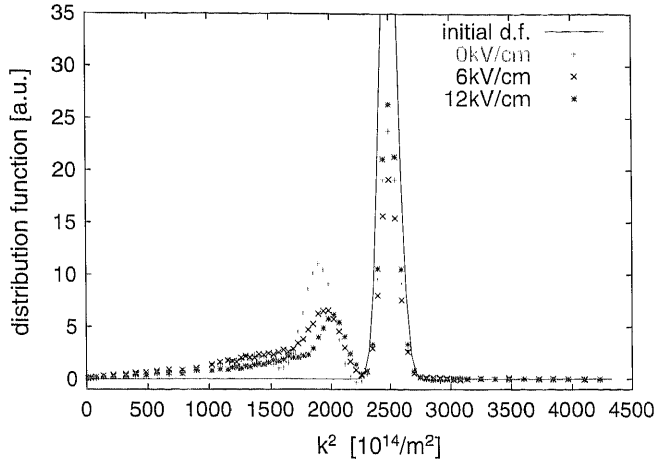


Fig. 11. – Solutions $kf(0, k, t)$ for positive k values and evolution time 200 fs. The electric field is 0, 6 kV/cm, and 12 kV/cm.

a pronounced effect on the broadening and retardation of the solutions: A comparison of the first replicas and the main peaks under the initial condition in figs. 10, 11, and 12 shows that the field influences the effects of collisional broadening and retardation.

5.3. Approximate model for the reduced WF. – Under general conditions for the potential we look for approximate solutions for the functions f_1 and f_2 . One option is to consider only the zero-order terms in the Neumann expansions. That is, only the free

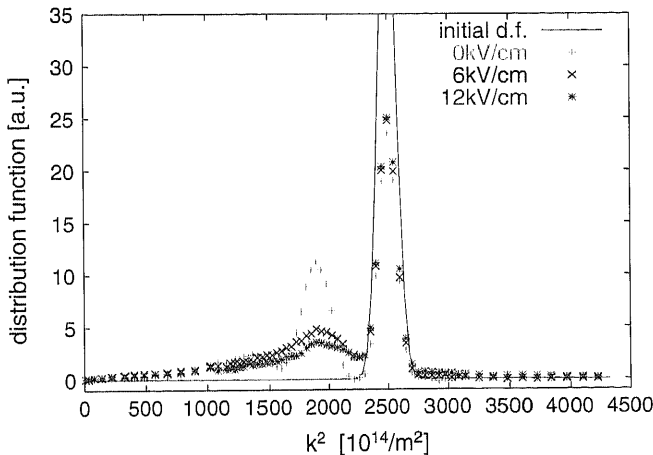


Fig. 12. – Solutions $rf(r, 0, t)$ for an evolution time of 200 fs. The electric field \mathbf{E} is 0, 6 kV/cm, and 12 kV/cm.

terms of (36) and (37) can be used in (35). Unfortunately in this case the ICFE is entirely neglected. Alternatively, in order to avoid the whole complexity of the general model, we can utilize some of the results derived for a homogeneous field. The potential V in the two auxiliary equations is approximated to a linear function $-\mathbf{F} \cdot \mathbf{r}$. The two equations can be solved by using Newton's trajectories defined by

$$(42) \quad \mathbf{p}_{(\tau)} = \mathbf{p} - e\mathbf{E}(t - \tau), \quad \mathbf{r}_{(\mathbf{p}, \mathbf{q}', \tau)} = \mathbf{r} - \int_{\tau}^t dy \frac{\mathbf{p}_{(y)} - \hbar \mathbf{q}'/2}{m} = \mathbf{r}_{(\mathbf{p}, \tau)} + \frac{\hbar \mathbf{q}'}{2m} (t' - \tau).$$

After replacement in the integro-differential form of (35) one obtains

$$(43) \quad \left(\frac{\partial}{\partial t} + \frac{\mathbf{p}}{m} \cdot \nabla_{\mathbf{r}} \right) f_W(\mathbf{r}, \mathbf{p}, t) = \int d\mathbf{p}' V'_W(\mathbf{r}, \mathbf{p}' - \mathbf{p}) f_W(\mathbf{r}, \mathbf{p}', t') + \int_0^t dt' \times \\ \times \left(S(\mathbf{p}', \mathbf{p}, t, t') f_W \left(\mathbf{r} + \frac{\hbar \mathbf{q}}{2m} (t - t'), \mathbf{p}'_{(t')}, t' \right) - \right. \\ \left. - S(\mathbf{p}, \mathbf{p}', t, t') f_W \left(\mathbf{r} + \frac{\hbar \mathbf{q}}{2m} (t - t'), \mathbf{p}_{(t')}, t' \right) \right), \\ S(\mathbf{p}', \mathbf{p}, t, t') = \frac{2VF_{\mathbf{q}}^2}{(2\pi\hbar)^3} (n(\mathbf{q}) \cos(\Omega(\mathbf{p}', \mathbf{p}, t, t')) + (n(\mathbf{q}) + 1) \cos(\Omega(\mathbf{p}, \mathbf{p}', t, t'))), \\ \Omega(\mathbf{p}, \mathbf{p}', t, t') = \int_{t'}^t d\tau \frac{\epsilon(\mathbf{p}_{(\tau)}) - \epsilon(\mathbf{p}'_{(\tau)}) + \hbar\omega_{\mathbf{q}}}{\hbar}, \quad \mathbf{p}_{(\tau)} = \mathbf{p} - \mathbf{F}(t - \tau), \quad \mathbf{q} = \frac{\mathbf{p}' - \mathbf{p}}{\hbar}.$$

This equation is a generalization of the Levinson equation in the full six-dimensional phase space. The latter is recovered in the case when V is linear and a space-independent initial condition is assumed. Thus the ICFE has been properly taken into account. Moreover no approximations are introduced for the coherent part of the transport process: if the phonon interaction is neglected, the common Wigner equation for an electron in a potential field is recovered. The equation becomes more transparent for a physical analysis when it is transformed into an integral equation. The main peculiarity of (43) is the non-locality in real space. The Boltzmann distribution function in point \mathbf{r} , \mathbf{p} at time t collects contributions only from the past of the real space part of the trajectory β passing through this point. Because of the quantum character of the phonon interaction, the solution of (43) can collect contributions from all points in the phase space. Formally a test particle that carries the value of f_W can be assigned to a given phase space point at time t'' . The interaction begins at t'' when the particle absorbs a half of the phonon momentum $\hbar \mathbf{q}'/2$. The real space part of the trajectory of the particle changes accordingly. The interaction has a finite duration so that the particle can appear on β at some time $t' \leq t$. At time t' the second half of the phonon momentum is absorbed by the particle. The particle now has the right coordinates to continue to \mathbf{r} , \mathbf{p} at time t . The process corresponds to a real absorption of a phonon with mode \mathbf{q}' . Alternatively, virtual absorption occurs when the particle releases half of the phonon mode. Note that a next interaction can begin only after completion of the current one. This is in accordance with the assumption of weak scattering we did for the generalized Wigner function. It is

concluded that besides ICFE the finite duration of the phonon interaction gives rise to a space non-locality of the quantum transport process. The momentum conservation law holds after the interaction completes as in the Boltzmann case. But there is no energy conservation even in the most simple homogeneous case, where the electric field is zero. The energy-conserving delta-function in the Boltzmann type of interaction is obtained after a limit which neglects the duration of the collision process.

6. – Classical limit in the electron-phonon interaction

We consider the classical limit of the electron-phonon interaction in (43). The same result will be obtained if we take the limit in (36) and (37) of the general model and replace the results in (35). The time integral in (43) is of the form

$$(44) \quad \int_0^t d\tau e^{(i/\hbar)\epsilon\tau} \phi(\tau).$$

The following formal limit holds in terms of generalized functions:

$$(45) \quad \lim_{\hbar \rightarrow 0} \frac{1}{\hbar} \int_0^t d\tau e^{(i/\hbar)\epsilon\tau} \phi(\tau) = \phi(0) \left\{ \pi \delta(\epsilon) + i\mathcal{P} \frac{1}{\epsilon} \right\}.$$

The actual meaning of the limit is that the product of the energy and time scales become much larger than \hbar . The mathematical aspects of the derivation are considered in [17]. As applied to the right-hand side of (43), limit (45) leads to cancellation of all principal values \mathcal{P} . This follows from the fact that (43) contains only real quantities. The energy and momentum conservation laws are incorporated in the obtained equation. The differential form of the latter resembles the Boltzmann scattering operator as added to the Wigner potential operator:

$$(46) \quad \left(\frac{\partial}{\partial t} + \frac{\mathbf{p}}{m} \cdot \nabla_{\mathbf{r}} \right) f_W(\mathbf{r}, \mathbf{p}, t) = \int d\mathbf{p}' V_W(\mathbf{r}, \mathbf{p}' - \mathbf{p}) f_W(\mathbf{r}, \mathbf{p}', t') + \\ + \int d\mathbf{p}' (f_W(\mathbf{r}, \mathbf{p}', t) S(\mathbf{p}', \mathbf{p}) - f_W(\mathbf{r}, \mathbf{p}, t) S(\mathbf{p}, \mathbf{p}')), \\ S(\mathbf{p}', \mathbf{p}) = \frac{V}{h^3} \frac{2\pi}{\hbar} \left\{ |\mathcal{F}(\mathbf{q})|^2 \delta(\epsilon(\mathbf{p}) - \epsilon(\mathbf{p}') - \hbar\omega_{\mathbf{q}}) n(\mathbf{q}) + \right. \\ \left. + |\mathcal{F}(-\mathbf{q})|^2 \delta(\epsilon(\mathbf{p}) - \epsilon(\mathbf{p}') + \hbar\omega_{-\mathbf{q}}) (n(-\mathbf{q}) + 1) \right\},$$

where $\mathbf{q} = (\mathbf{p} - \mathbf{p}')/\hbar$ and $|\mathcal{F}|^2 = \hbar^2 F^2$ is the electron-phonon matrix element.

Equation (46), called Wigner-Boltzmann equation, will be in the main focus in what follows. The interaction with phonons is treated classically, while the interaction with the Wigner potential is considered on a rigorous quantum level. A classical limit in the potential term of the obtained Wigner-Boltzmann equation recovers the Boltzmann equation, see (24). Thus eq. (46) is one step higher in the derived hierarchy of transport models shown in fig. 13.

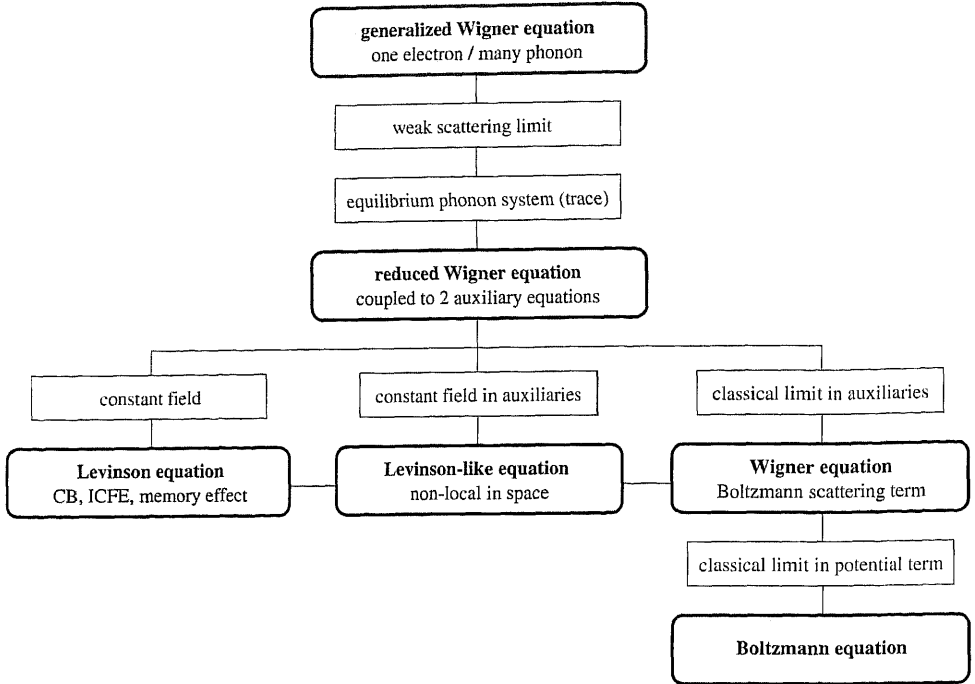


Fig. 13. – Hierarchy of transport models.

7. – Particle models for the Wigner-Boltzmann equation

While the flow of approximations in fig. 13 goes downwards, the numerical approaches climb from the bottom of the figure. In the last 30 years Monte Carlo method has been widely recognized as the most efficient numerical approach to the Boltzmann equation. The Levinson equation has been solved by a backward Monte Carlo method which can be efficient up to 0.5 picoseconds evolution time. During the last decade the method has been generalized for solving the semiconductor Bloch equations [18] and other topics of the ultrafast phenomena in photo-excited semiconductors, see [19] and references therein. The base of any Monte Carlo approach is the underlying particle model. We first introduce some concepts of Monte Carlo integration. They are used to develop a particle model for the Wigner-Boltzmann equation (46). In the obtained model particles are associated with a sign and thus become positive and negative. The sign is the only property of the particles related to the quantum information. All other aspects of their behavior resemble Boltzmann-like particles. This shows that the analogy between classical and Wigner transport pictures can be even closer. The sign is taken into account in the evaluation of the physical averages. The sign has a physical meaning, since positive and negative particles which meet in the phase space annihilate one another. The Wigner and Boltzmann transport pictures are explained in a unified way by the processes drift,

scattering, generation and recombination of positive and negative particles. The model ensures a seamless transition between the classical and quantum regions.

7.1. Concepts of Monte Carlo method.

7.1.1. Random variable. Consider a random variable ψ which takes values $\psi(Q)$ with probability density $p_\psi(Q)$. Here Q is a multi-dimensional point. The expectation value E_ψ of ψ is given by

$$(47) \quad E_\psi = \int dQ p_\psi(Q) \psi(Q).$$

The simplest Monte Carlo method evaluates E_ψ by performing N independent realizations of the probability density p_ψ . Generated are N points Q_1, \dots, Q_N , called sampling points for the random variable ψ . The sample mean η estimates the expectation value E_ψ ,

$$(48) \quad E_\psi \simeq \eta = \frac{1}{N} \sum_{i=1}^N \psi(Q_i),$$

with a precision which depends on the number of independent realizations N and the variance of σ_ψ of the random variable. According to the “rule of the three sigma” [20]

$$(49) \quad \mathcal{P} \left(|E_\psi - \eta| \leq \frac{3\sigma_\psi}{\sqrt{N}} \right) \simeq 0.997,$$

the probability \mathcal{P} for η to be inside the interval $3\sigma_\psi/\sqrt{N}$ around E_ψ is very high (0.997).

7.1.2. Evaluation of integrals. The concept of Monte Carlo approach for evaluation integrals is to present the given integral as an expectation value:

$$(50) \quad I = \int f(Q) dy = \int p(Q) \frac{f(Q)}{p(Q)}, \quad p(Q) \geq 0, \quad \int p(Q) dy = 1$$

of the random variable $\psi = f/p$. The probability density function p can be arbitrary, but admissible for f : $p \neq 0$ if $f \neq 0$. Different random variables can be introduced, depending on the choice of p . All of them have the same expectation value I but different variance and higher moments. It can be shown that the lowest variance is obtained if p is chosen proportional to $|f|$.

7.1.3. Evaluation of integral equations. Consider a Fredholm integral equation of the second kind with a kernel K and a free term f_0 :

$$(51) \quad f(Q) = \int dQ' f(Q') K(Q', Q) + f_0(Q).$$

Iteratively replacing the equation into itself, we obtain the solution expanded into a Neumann series. The terms of the series are consecutive iterations of the kernel on the initial condition. Each term in the series is a multiple integral of the type of (50) and can be evaluated by a Monte Carlo method. Now assume that we are not interested in the solution itself, but in the inner product of f with a given function A :

$$\langle A \rangle = (A, f) = \int dQ A(Q) f(Q).$$

Consider the following equation, called adjoint to (51):

$$(52) \quad g(Q') = \int dQ K(Q', Q) g(Q) + A(Q').$$

If (51) is multiplied by g , (52) is multiplied by f and the two equations are integrated and compared, it is obtained that

$$(53) \quad \langle A \rangle = (A, f) = (f_0, g) = \int dQ f_0(Q') g(Q').$$

To evaluate $\langle A \rangle$ we can use the Neumann expansion of (52)

$$(54) \quad g(Q') = \int dQ \left(\delta(Q - Q') + \sum_{n=1}^{\infty} K^n(Q', Q) \right) A(Q),$$

$$K^n(Q', Q) = \int dQ_1 K(Q', Q_1) K^{n-1}(Q_1, Q).$$

This gives rise to a series expansion for $\langle A \rangle = \sum_i \langle A \rangle_i$. For example, the second term becomes

$$\langle A \rangle_2 = \int dQ' dQ_1 dQ_2 P_0(Q') P(Q', Q_1) P(Q_1, Q_2) \frac{f_0(Q') K(Q', Q_1) K(Q_1, Q_2) A(Q_2)}{P_0(Q') P(Q', Q_1) P(Q_1, Q_2)}.$$

Actually the term has been augmented with the help of the probabilities P_0 and P in order to resemble (50). These probabilities are used to build the so-called numerical trajectories:

- $P_0(Q')$ selects the initial point Q' of the trajectory. P_0 must take non-zero values in the points where f_0 is different from zero.
- From a given point Q' the probability $P(Q', Q)$ selects the next trajectory point Q . In order for P to be a probability, it is required that $\int dQ P(Q', Q) = 1 \forall Q'$. Furthermore P must be different from zero where K is non-zero.

The random variable in $\langle A \rangle$ is a product of weights $(f_0/P_0)(Q')$, $(K/P)(Q_{1,2})$ evaluated at each selected point in the row $Q_0 \rightarrow Q_1 \rightarrow Q_2$, obtained by application of $P_0 \rightarrow P \rightarrow P$. The mean value of N realizations of the random variable calculated over the trajectories $(Q' \rightarrow Q_1 \rightarrow Q_2)_i$, $i = 1, \dots, N$, evaluates $\langle A \rangle_2$. Here we used the iterative character of the multiple integral for $\langle A \rangle_2$ to introduce a consecutive procedure for the construction of the trajectory. In this way the trajectory can be continued by consecutive applications of P . Such trajectory can be used for the evaluation of the consecutive terms in $\langle A \rangle$. Apparently a trajectory can be associated with a “numerical” particle which jumps between the consecutive points. Our goal is to find a particle model where the concrete choice of P_0 and PP incorporates physical features in the numerical particles as much as possible.

8. – Physical averages in the stationary Wigner-Boltzmann transport: probabilistic analysis

8.1. *Formulation and reformulation of the transport problem.* – From now on we adopt the common notations in the field and use the wave vector $\mathbf{k} = \mathbf{p}/\hbar$ instead of the momentum \mathbf{p} . The Wigner phase space is thus composed of the real space coordinate \mathbf{r} and the wave vector \mathbf{k} . We consider the case of stationary transport, where the physical conditions imposed on the boundaries determine the device behavior. The device exchanges carriers with two or more reservoirs through the contacts denoted by b . Open-system boundary conditions are provided by the Fermi-Dirac distribution functions $f_b(\mathbf{r}_b, \mathbf{k})$ in the contacts [4]. The solution $f_W(\mathbf{r}, \mathbf{k})$ of the Wigner-Boltzmann equation is used to obtain the average values of all physical quantities of interest, recall eq. (22):

$$(55) \quad \langle A \rangle = \int_D d\mathbf{r} \int d\mathbf{k} f_W(\mathbf{r}, \mathbf{k}) A(\mathbf{r}, \mathbf{k}) = (f_W, A),$$

where D is the device domain. Equation (55) asserts that, in order to evaluate the averaged value of interest $\langle A \rangle$, one needs to know the solution f_W inside the device. An alternative expression for the mean value $\langle A \rangle$ can be found from the adjoint integral form of the WB equation. The derivation of this expression begins with the revision of the integro-differential form of the equation.

8.1.1. *Integro-differential form revised.* The stationary form of eq. (46) is

$$(56) \quad \mathbf{v}(\mathbf{k}) \cdot \nabla_{\mathbf{r}} f_W(\mathbf{r}, \mathbf{k}) = \int d\mathbf{k}' V_W(\mathbf{r}, \mathbf{k}' - \mathbf{k}) f_W(\mathbf{r}, \mathbf{k}) + \\ + \int d\mathbf{k}' f_W(\mathbf{r}, \mathbf{k}') S(\mathbf{k}', \mathbf{k}) - f_W(\mathbf{r}, \mathbf{k}) \lambda(\mathbf{k}).$$

Here $\mathbf{v}(\mathbf{k}) = \hbar \mathbf{k} / m$ is the velocity. $S(\mathbf{k}', \mathbf{k})$ has the meaning of the rate for scattering from state $(\mathbf{r}, \mathbf{k}')$ to state (\mathbf{r}, \mathbf{k}) due to phonon interaction. $\lambda(\mathbf{k}) = \int S(\mathbf{k}, \mathbf{k}') d\mathbf{k}'$ is the

phonon out-scattering rate. The Wigner potential (19) loses \hbar^3 from the denominator because of the transition to the wave vector:

$$(57) \quad V_W(\mathbf{r}, \mathbf{k}) = \frac{1}{i\hbar(2\pi)^3} \int d\mathbf{s} \exp[-i\mathbf{k} \cdot \mathbf{s}] \left(V\left(\mathbf{r} - \frac{\mathbf{s}}{2}\right) - V\left(\mathbf{r} + \frac{\mathbf{s}}{2}\right) \right).$$

The characteristics of the Liouville operator in (56) are particularly simple Newton trajectories (no force):

$$(58) \quad \mathbf{r}(t) = \mathbf{r} + \mathbf{v}(\mathbf{k})t, \quad \mathbf{k}(t) = \mathbf{k}.$$

We note that the results derived in what follows do not rely on the particular form of (58) (which is due to the stationary fieldless form of the Liouville operator in (56)), but can be obtained for the general form of the Liouville operator. As discussed in Appendix A, a stationary trajectory $(\mathbf{r}(t), \mathbf{k}(t))$ can always be initialized by the phase space point (\mathbf{r}, \mathbf{k}) at time 0. Equation (56) can be reformulated by a decomposition of the antisymmetric Wigner potential into two complementary parts:

$$V_W(\mathbf{r}, \mathbf{k}) = V_W^+(\mathbf{r}, \mathbf{k}) - V_W^+(\mathbf{r}, -\mathbf{k}), \quad \gamma(\mathbf{r}) = \int d\mathbf{k} V_W^+(\mathbf{r}, \mathbf{k}).$$

The function V_W^+ equals V_W if the Wigner potential is positive and is zero otherwise: $V_W^+ = V_W \xi(V_W)$ with ξ the Heaviside step function. The meaning of the function γ is discussed in the next section. By adding $\gamma(\mathbf{r})f_W(\mathbf{r}, \mathbf{k})$ to both sides of (56), the equation becomes

$$(59) \quad (\mathbf{v}(\mathbf{k}) \cdot \nabla_{\mathbf{r}} + \mu(\mathbf{r}, \mathbf{k})) f_W(\mathbf{r}, \mathbf{k}) = \int d\mathbf{k}' \Gamma(\mathbf{r}, \mathbf{k}, \mathbf{k}') f_W(\mathbf{r}, \mathbf{k}'),$$

$$(60) \quad \Gamma(\mathbf{r}, \mathbf{k}', \mathbf{k}) = V_W^+(\mathbf{r}, \mathbf{k}' - \mathbf{k}) - V_W^+(\mathbf{r}, \mathbf{k} - \mathbf{k}') + S(\mathbf{k}', \mathbf{k}) + \gamma(\mathbf{k})\delta(\mathbf{k} - \mathbf{k}'),$$

$$(61) \quad \mu(\mathbf{r}, \mathbf{k}) = \lambda(\mathbf{k}) + \gamma(\mathbf{r}).$$

With the help of (58), eq. (59) is transformed into an integral equation.

8.1.2. Integral form. To obtain the integral form of the stationary Wigner-Boltzmann equation, we consider a given phase space point (\mathbf{r}, \mathbf{k}) . This point determines uniquely a phase space trajectory, $(\mathbf{r}(t), \mathbf{k}(t))$ in backward parameterization, $t \leq 0$. Consider the parameterized equation (59)

$$(62) \quad (\mathbf{v}(\mathbf{k}(t)) \cdot \nabla_{\mathbf{r}(t)} + \mu(\mathbf{r}(t), \mathbf{k}(t))) f_W(\mathbf{r}(t), \mathbf{k}(t)) = \int d\mathbf{k}' \Gamma(\mathbf{r}(t), \mathbf{k}(t), \mathbf{k}') f_W(\mathbf{r}(t), \mathbf{k}').$$

If both sides of (62) are multiplied by the integrating factor $\exp[-\int_t^0 \mu(\mathbf{k}(y), \mathbf{r}(y)) dy]$, the left-hand side represents a total time derivative:

$$(63) \quad \frac{d}{dt} \exp \left[- \int_t^0 \hat{\mu}(y) dy \right] \hat{f}(t) = \exp \left[- \int_t^0 \hat{\mu}(y) dy \right] \hat{\Gamma}[f](t).$$

Here $\hat{\Gamma}[f](t)$ denotes the right-hand side of (62) and

$$\hat{\mu}(y) = \mu(\mathbf{k}(y), \mathbf{r}(y)) dy, \quad \hat{f}(t) = f_W(\mathbf{r}(t), \mathbf{k}(t)).$$

This equation can be integrated straightforwardly. The upper bound of integration should be $t = 0$ to obtain $\hat{f}(0) = f(\mathbf{k}, \mathbf{r})$, the value of f at the given phase space point. The lower time bound has to be chosen such that $\mathbf{r}(t)$ and $\mathbf{k}(t)$ take on values at which the distribution function is known. In the steady state the distribution function is known only at the device boundary. An appropriate lower time bound is therefore the time, say t_b^- , at which the trajectory crosses the simulation domain boundary. Apparently, this time depends on the point \mathbf{k}, \mathbf{r} under consideration. For trajectories closed in D the time t_b is $-\infty$. Integration of (63) in the time bounds discussed above results in the integral form of the stationary Wigner-Boltzmann equation

$$(64) \quad f(\mathbf{r}, \mathbf{k}) = \int d\mathbf{k}' \int_{t_b^-}^0 f(\mathbf{r}(t'), \mathbf{k}') \Gamma(\mathbf{r}(t'), \mathbf{k}', \mathbf{k}(t')) \times \\ \times \exp \left[- \int_{t'}^0 \mu(\mathbf{r}(y), \mathbf{k}(y)) dy \right] + f_0(\mathbf{r}, \mathbf{k}), \\ f_0(\mathbf{r}, \mathbf{k}) = f_b(\mathbf{r}(t_b^-), \mathbf{k}(t_b^-)) \exp \left[- \int_{t_b^-}^0 \mu(\mathbf{r}(y), \mathbf{k}(y)) dy \right], \quad t_b^- = t_b^-(\mathbf{k}, \mathbf{r}).$$

The equation can be understood in analogy with the integral form of the Boltzmann equation. The latter is obtained by formally setting the Wigner potential to zero. Then, the exponent in (64) becomes the probability for a particle to drift without scattering by phonons during the time interval $(t', 0)$ on the proper trajectory $\beta(\mathbf{r}, \mathbf{k})$, which arrives at (\mathbf{r}, \mathbf{k}) . There are two contributions to the value of f in the point (\mathbf{r}, \mathbf{k}) . f_0 is the value of the boundary function f_b , which survives on β despite the action of the phonons. The other term gives cumulative contributions from previous times t' : the values of f located at $\mathbf{r}(t')$ scatter according to S from everywhere in the wave vector space to the proper $\mathbf{k}(t')$. The particular values fS are further multiplied by the exponent to filter out the part which is scattered out of β by the phonons. This picture can be maintained when the Wigner potential is switched on. Then V_W^+ in (60) has a clear meaning of scattering due to the Wigner potential. The function γ can be interpreted as an out-scattering rate due to the Wigner potential, in strict analogy with the phonon out-scattering rate λ . Then $\gamma\delta$ becomes a self-scattering function. The major difference between S and Γ comes from the fact that, while the former is strictly non-negative, there is a minus sign in (60). This sign precludes a direct probabilistic treatment of the equation in terms of classical particles.

After this step, the boundary conditions f_0 appear explicitly in (64). f_0 , along with the solution g of an equation adjoint to (64), give rise to the desired expression for the physical averages.

8.1.3. Adjoint equation. Equation (64) can be formally written in the standard form (51) of a Fredholm integral equation of the second kind:

$$(65) \quad f(\mathbf{r}, \mathbf{k}) = \int d\mathbf{r}' \int d\mathbf{k}' f(\mathbf{r}', \mathbf{k}') K(\mathbf{r}', \mathbf{k}', \mathbf{r}, \mathbf{k}) + f_0(\mathbf{r}, \mathbf{k}).$$

The kernel K has been augmented to account for the \mathbf{r}' integration by a spatial delta-function:

$$(66) \quad K(\mathbf{r}', \mathbf{k}', \mathbf{r}, \mathbf{k}) = \int_{-\infty}^0 dt' \Gamma(\mathbf{r}', \mathbf{k}', \mathbf{k}(t')) \times \\ \times \exp \left[- \int_{t'}^0 \mu(\mathbf{r}(y), \mathbf{k}(y)) dy \right] \delta(\mathbf{r}' - \mathbf{r}(t')) \theta_D(\mathbf{r}').$$

The indicator function of the simulation domain θ_D ensures the proper lower bound $t_b^-(\mathbf{r}, \mathbf{k})$ of the time integral. The adjoint equation has the same kernel as (65) but the integration is carried out over the unprimed variables, see (52). The free term is chosen to be the physical quantity of interest A :

$$(67) \quad g(\mathbf{r}', \mathbf{k}') = \int d\mathbf{r} \int d\mathbf{k} K(\mathbf{r}', \mathbf{k}', \mathbf{r}, \mathbf{k}) g(\mathbf{r}, \mathbf{k}) + A(\mathbf{r}', \mathbf{k}').$$

As obtained from (64), eq. (67) assumes a backward parameterization of the trajectories. Forward trajectories are introduced by first changing the integration variables from \mathbf{r} , \mathbf{k} to $\mathbf{r}'' = \mathbf{r}(t')$, $\mathbf{k}'' = \mathbf{k}(t')$ back in time over the trajectory initialized by \mathbf{r} , \mathbf{k} . Applying (A.3) and (A.4), the adjoint equation (67) is obtained in forward parameterization. The integration on \mathbf{r}'' can be achieved using the delta-function in K . A replacement of \mathbf{k}'' by \mathbf{k} and $-t'$ by t gives rise to the following compact form:

$$(68) \quad g(\mathbf{r}', \mathbf{k}') = \int d\mathbf{k} \int_0^\infty dt \theta_D(\mathbf{r}') \Gamma(\mathbf{r}', \mathbf{k}', \mathbf{k}) \times \\ \times \exp \left[- \int_0^t \mu(\mathbf{r}'(y), \mathbf{k}(y)) dy \right] g(\mathbf{r}'(t), \mathbf{k}(t)) + A(\mathbf{r}', \mathbf{k}').$$

Here, $(\mathbf{r}'(t), \mathbf{k}(t))$ is a forward trajectory initialized by $(\mathbf{r}', \mathbf{k})$. The equation has the desired property that the integration is carried out over final states and that the time variable is positive. According to (53) the averaged value of the physical quantity A is expressed through the boundary conditions f_0 and the solution of (68),

$$(69) \quad \langle A \rangle = \int_D d\mathbf{r} \int d\mathbf{k} f_b(\mathbf{r}(t_b^-), \mathbf{k}(t_b^-)) \exp \left[- \int_{t_b^-}^0 \mu(\mathbf{r}(y), \mathbf{k}(y)) dy \right] g(\mathbf{k}, \mathbf{r}),$$

where t_b^- and the backward trajectory $(\mathbf{r}(t), \mathbf{k}(t))$ are determined by (\mathbf{r}, \mathbf{k}) .

Since f_b is defined only at the boundary ∂D , a transformation is needed that leads from a volume to a boundary integral. A phase space point (\mathbf{r}, \mathbf{k}) is bijectively mapped

onto $(\mathbf{k}(t_b), \mathbf{r}_b = \mathbf{r}(t_b))$ by the boundary time t_b . (We consider the subspace of points $(\mathbf{k}', \mathbf{r}')$ having finite t_b . Fortunately this is the relevant sub-domain for integral (69), since in the complementary subspace the integrand vanishes.) This implies that the transformation must replace one of the space integrals by a time integral. One can formally augment (69) by a time integral $\int dt' \delta(t' - t_b)$ in the limits $(0, -\infty)$. After a rearrangement of the integrals (see [21]), (69) is transformed into

$$(70) \quad \langle A \rangle = \oint_{\partial D} d\sigma(\mathbf{r}_b) \int_{P_+} d\mathbf{k}_b \int_0^\infty dt_0 |v_\perp(\mathbf{k}_b)| f_b(\mathbf{r}_b, \mathbf{k}_b) \times \\ \times \exp \left[- \int_0^{t_0} \mu(\mathbf{r}_b(y), \mathbf{k}_b(y)) dy \right] g(\mathbf{r}_b(t_0), \mathbf{k}_b(t_0)).$$

Here P_+ is the part of the wave vector space with \mathbf{k}_b vectors pointing towards the device and $|v_\perp|$ is the modulus of the velocity component normal to ∂D .

By replacing g with the iteration series (54), $\langle A \rangle$ may be expanded into the series

$$(71) \quad \langle A \rangle = \langle b, (I - K)^{-1} A \rangle = \langle \mathbf{v}_\perp f_b, (I - \tilde{K})^{-1} \tilde{A} \rangle = \sum_{i=0}^{\infty} \langle A \rangle_i.$$

The second term is reformulated to facilitate further analysis. The multipliers in each term of the sum $\langle b, (I - K)^{-1} A \rangle$ are formally regrouped. Now \tilde{K} is the repeating term in the pattern, which is obtained from K after absorbing the exponent on the left and releasing the exponent to the right for the next \tilde{K} . In this way A is assigned with the last exponent to become \tilde{A} . The zero-order term $\langle A \rangle_0 = \langle \mathbf{v}_\perp f_b, \tilde{A} \rangle$ is given by the right-hand side of (70), with $A(\mathbf{r}_b(t_0), \mathbf{k}_b(t_0))$ in place of g . The first term is

$$(72) \quad \langle A \rangle_1 = \oint_{\partial D} d\sigma(\mathbf{r}_b) \int_{P_+} d\mathbf{k}_b \int_0^\infty dt_0 \int d\mathbf{k}_1 \int_0^\infty dt_1 |v_\perp(\mathbf{k}_b)| f_b(\mathbf{r}_b, \mathbf{k}_b) \times \\ \times \exp \left[- \int_0^{t_0} \mu(\mathbf{r}_b(y), \mathbf{k}_b(y)) dy \right] \theta_D(\mathbf{r}_b(t_0)) \Gamma(\mathbf{r}_b(t_0), \mathbf{k}_b(t_0), \mathbf{k}_1) \times \\ \times \exp \left[- \int_0^{t_1} \mu(\mathbf{r}_1(y), \mathbf{k}_1(y)) dy \right] A(\mathbf{r}_1(t_1), \mathbf{k}_1(t_1)).$$

Here the trajectory $(\mathbf{r}_1(t), \mathbf{k}_1(t))$ is initialized by $(\mathbf{r}_b(t_0), \mathbf{k}_1)$. The next term, $\langle A \rangle_2$, is derived by augmenting (72) with integrals on \mathbf{k}_2 and t_2 , and replacing A with $\theta\Gamma$. The product

$$(73) \quad \tilde{A}_2 = \exp \left[- \int_0^{t_2} \mu(\mathbf{r}_2(y), \mathbf{k}_2(y)) dy \right] A(\mathbf{r}_2(t_2), \mathbf{k}_2(t_2))$$

appears at the end of the expression for $\langle A \rangle_2$ as an integrand on time t_2 . Higher-order terms in (71) are derived by induction.

The series expansion (71) is the key quantity in the treatment of the boundary value problem. It proves that knowledge of the boundary distribution is sufficient to determine arbitrary volume integrals defined by (55) and therefore to determine $\langle A \rangle$ uniquely. Only the subspace P_+ of boundary states \mathbf{k}_b having an inward directed velocity component \mathbf{v}_\perp appear in (70), and thus they determine the boundary condition. The complementary part is *a priori* unknown and comes out as a result of the transport process. The series expansion (71) will be analyzed in terms of probability densities.

8.2. Analysis of $\langle A \rangle$. – The basic particle methods, used to date for simulation of semiconductor devices, were originally devised by considerations where the simulation was an emulation of the physical process: the transport picture has been used to establish the corresponding stochastic method. The link between such physically based methods and the numerical methods of Monte Carlo integration has been established later [22-27]. Here, we follow the opposite approach: the numerical Monte Carlo theory is used to propose a common particle picture of the Wigner and Boltzmann transport processes. As already discussed, the approach requires all integrals in (71) to be decomposed into probability densities and random variables. An advantage is provided by the common structure of these integrals, which are built by the boundary term b , the consecutive applications of \tilde{K} and end up with the quantity \tilde{A} . It is then sufficient to extract from each of these three quantities the proper probability densities. As these quantities appear in (72), we focus on that equation.

8.2.1. Injection from the boundaries. The boundary term allows a simple probabilistic interpretation. For the purpose of normalization, we introduce the integrals

$$(74) \quad j_\perp(\mathbf{r}_b) = \int_{P_+} d\mathbf{k} |v_\perp(\mathbf{k})| f_b(\mathbf{k}, \mathbf{r}_b), \quad \Phi = \oint_{\partial D} j_\perp(\mathbf{r}) d\sigma(\mathbf{r}),$$

which represent the normal component of the incident particle current density and the total incident particle current. Then the quantity

$$p_b(\mathbf{r}_b, \mathbf{k}_b) = \frac{j_\perp(\mathbf{r}_b)}{\Phi} \frac{|v_\perp(\mathbf{k}_b)| f_b(\mathbf{r}_b, \mathbf{k}_b)}{j_\perp(\mathbf{r}_b)} = p_{b1}(\mathbf{r}_b) p_{b2}(\mathbf{r}_b, \mathbf{k}_b)$$

has the proper normalization of a conditional probability density. p_b generates a phase space point on the boundary by first selecting the position \mathbf{r}_b proportional to the incident particle current density. \mathbf{k}_b is then selected according to the velocity-weighted equilibrium distribution p_{b2} . In this way, the boundary term is factorized into a product of p_b with the normalization constant Φ . The selection of the boundary point follows the classical rules used in the device MC method and is thus associated with a particle that is injected into the device.

8'2.2. Probability factors in \tilde{K} . \tilde{K} is augmented by a multiplication and a division by μ , which gives rise to the product

$$\tilde{K}(\mathbf{r}', \mathbf{k}', \mathbf{k}, t) = p_t(t, \mathbf{r}', \mathbf{k}') \theta_D(\mathbf{r}'(t)) \frac{\Gamma(\mathbf{r}'(t), \mathbf{k}'(t), \mathbf{k})}{\mu(\mathbf{r}'(t), \mathbf{k}'(t))},$$

$$p_t = \mu(\mathbf{r}'(t), \mathbf{k}'(t)) \exp \left[- \int_0^t \mu(\mathbf{r}'(y), \mathbf{k}'(y)) dy \right].$$

The structure of the first term p_t is well known from the classical MC method: it is the probability for a drift without scattering provided that the scattering frequency is μ . The normalization to unity is readily proven by integration over time with the limits $(0, \infty)$. p_t generates a value of t associated with a free flight time of a particle which drifts over a piece of a Newton trajectory between the initial state $(\mathbf{r}', \mathbf{k}')$ and $(\mathbf{r}'(t), \mathbf{k}'(t))$, and which, as we will show, has a meaning of a before-scattering state. It is used as an input in the conditional probabilities composing the remaining term $\Gamma(\mathbf{r}'(t), \mathbf{k}'(t), \mathbf{k})/\mu(\mathbf{r}'(t), \mathbf{k}'(t))$ to generate the output value of \mathbf{k} :

$$(75) \quad \frac{\Gamma(\mathbf{r}', \mathbf{k}', \mathbf{k})}{\mu(\mathbf{r}', \mathbf{k}')} = p_\lambda(\mathbf{r}', \mathbf{k}') p_{\text{ph}}(\mathbf{k}', \mathbf{k}) + p_\gamma(\mathbf{r}', \mathbf{k}') \times$$

$$\times \left(\frac{1}{3} p_W^+(\mathbf{r}', \mathbf{k}' - \mathbf{k}) - \frac{1}{3} p_W^-(\mathbf{r}', \mathbf{k} - \mathbf{k}') + \frac{1}{3} p_\delta(\mathbf{k} - \mathbf{k}') \right) 3,$$

$$p_\lambda(\mathbf{r}', \mathbf{k}') = \frac{\lambda(\mathbf{k}')}{\mu(\mathbf{r}', \mathbf{k}')}, \quad p_\gamma(\mathbf{r}', \mathbf{k}') = \frac{\gamma(\mathbf{r}')}{\mu(\mathbf{r}', \mathbf{k}')},$$

$$p_{\text{ph}}(\mathbf{k}', \mathbf{k}) = \frac{S(\mathbf{k}', \mathbf{k})}{\lambda(\mathbf{k}')}, \quad p_W^\pm(\mathbf{r}', \mathbf{k}) = \frac{V_W^\pm(\mathbf{r}', \mathbf{k})}{\gamma(\mathbf{r}')}.$$

Here the time argument has been omitted, and $p_W^- = p_W^+$ is introduced for convenience. According to (61), p_λ and p_γ are two complementary probabilities, which can be used to select either p_{ph} or the term in the brackets in (75). The first branch occurs with the probability p_λ , which selects the type of interaction to be scattering with phonons. The application of the probability density p_{ph} is readily understood as a generation of the phonon after-scattering state $(\mathbf{r}', \mathbf{k})$. The second branch can be interpreted as a generation of an after-scattering state due to interaction with the Wigner potential. It is comprised of the three terms enclosed in the brackets. p_2 has been introduced with the purpose to select which of the three densities p_W^+ , p_W^- and p_δ generates an after-scattering state $(\mathbf{r}', \mathbf{k})$. In this way, the action of the Wigner potential is realized by a scattering generated by either of these three probability densities. They will be discussed in detail in the next section. Here, we conclude that the consecutive application of the conditional probabilities, comprising \tilde{K} , generates a transition between $(\mathbf{r}', \mathbf{k}')$ and $(\mathbf{r}'(t), \mathbf{k})$ which is associated with a particle which undergoes a free flight followed by a scattering event.

What remains for the random variable associated with \tilde{K} is the term $w\theta_D = (\pm 3)^i \theta_D$. The power i depends on the type of the interaction: $i = 0$ and $w = 1$ if the scattering is due to phonons. If the Wigner potential is selected as a scattering source, $i = 1$ and

$w = (\pm 3)$, where the minus sign applies if $p_{\bar{w}}$ in (75) is selected. The quantity w is called a weight factor. The domain indicator θ_D is unity if the particle is inside the device at the end of the free flight and is zero otherwise. \tilde{K} factorizes into a product of the random variable and an evolution operator composed of conditional probability densities.

8'2.3. Recording averages. The integrand (73) can be written as follows:

$$\tilde{A} = p_t(t, \mathbf{r}, \mathbf{k}) \frac{A(\mathbf{r}(t), \mathbf{k}(t))}{\mu(\mathbf{r}(t), \mathbf{k}(t))}.$$

The random variable ψ_A associated with the physical quantity A is the term A/μ evaluated at the end of the free flight. It must be noted that, as we are interested in physical averages in a given region Ω inside the device domain, A contains implicitly the indicator θ_Ω of that region. If the end point of the free flight is outside Ω , the random variable is zero. Another way to express the random variable ψ_A can be obtained by integration by parts of the t integral

$$(76) \quad \tilde{A} = p_t(t, \mathbf{r}, \mathbf{k}) \int_0^t dy A(\mathbf{r}(y), \mathbf{k}(y)).$$

The value of ψ_A is identified as the path integral over y . Actually, due to the indicator θ_Ω , only the part of the path belonging to Ω contributes to the integral. The two functionals of A are known in the classical single-particle MC method as synchronous ensemble and time integration techniques [28].

So far we are ready to state the stochastic approach for evaluation of (72). Numerical trajectories are built up with the help of p_b , the conditional probabilities identified from \tilde{K} and the probability p_t . The random variable $\psi_1 = \Phi\theta(\pm 3)^i \psi_A$ is calculated for each trajectory. The sample mean (48) over N trajectories estimates $\langle A \rangle_1$. A generalization for the n -th term in (71) is straightforward. Numerical trajectories are built up with the help of p_b , n consecutive iterations of the conditional probabilities identified from \tilde{K} , and the probability p_t . The corresponding random variable ψ_n is given by the product

$$(77) \quad \psi_n = \Phi \prod_{k=1}^n \theta_{D_k} (\pm 3)^{i_k} \psi_A = \Phi W_n \psi_A, \quad \psi_{\langle A \rangle} = \sum_n \psi_n.$$

We first note that a given trajectory can be used to evaluate all terms with order lower than n . Trajectories which leave the device domain after k iterations give zero contribution to the sample mean (48) for any term with $n > k$. Nevertheless, such trajectories are counted as independent realizations in the denominator N of the sample mean. It follows that a given trajectory can be used for evaluation of all terms in (71): a trajectory which begins at a domain boundary and ends at a domain boundary becomes an independent realization of the random variable $\psi_{\langle A \rangle}$ in (77). The sample mean over N such trajectories estimates $\langle A \rangle$.

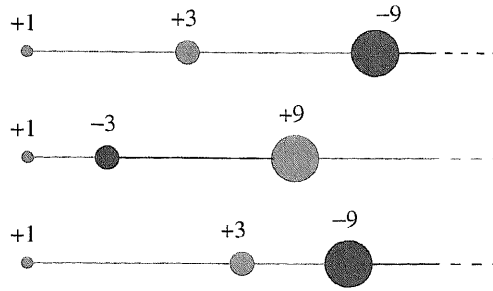


Fig. 14. – Schematic evolution of the weight of Wigner particles in the scattering model.

9. – Particle models

9.1. Classical transport. – In the case of Boltzmann transport, the Wigner potential completes the Liouville operator with a force term. The trajectories have the general form (A.1). On the right-hand side of (56), only the two terms related to the phonons remain. This allows to formally set $\gamma = 0$, $\mu = \lambda$ in (61). The right-hand side of (75) reduces to p_{ph} . It is easily seen that the numerical trajectories coincide with the real trajectories of the Boltzmann carriers evolving in the device. Indeed, the generation of the initial point of the trajectory corresponds to an injection of a classical particle. p_t becomes the usual classical probability for the free-flight duration due to phonons. The scattering is determined by the phonon scattering rate S through p_{ph} in (75). The weight W_n in (77) remains unity for all n . The domain indicator takes into account only the trajectories which are inside the device. Thus, numerical particles contribute to the averages in the same way as Boltzmann carriers in the single-particle MC method. We conclude that the resulting particle picture coincides with the picture of classical particles that is emulated by the device MC method.

9.2. Coherent transport. – The coherent transport, which considers only events of quantum interaction, is obtained by setting $\lambda = 0$, $\mu = \gamma$ in (61). According to the term in brackets in (75), these are scattering events which change the statistical weight in (77). The scattering model (corresponding to the scattering interpretation of the Wigner potential) leads to a weight evolution shown in fig. 14.

We examine the model as applied to a tunneling process. In the simulated experiment, particles are injected between the two 1 nm thin, 50 meV high barriers of an unbiased resonant-tunneling device (RTD). The injected particles are evenly distributed in the middle 2 nm part inside the 4 nm wide potential well and have a Maxwell-Boltzmann distribution in energy, the tail of which is truncated at 50 meV such that the injected particles can overcome the barriers only by tunneling. Material parameters for GaAs at 300 K temperature are assumed. Particles crossing the barriers can leave the device through the left or right absorbing contacts.

The experiment is especially designed for comparison with a classical system. Under

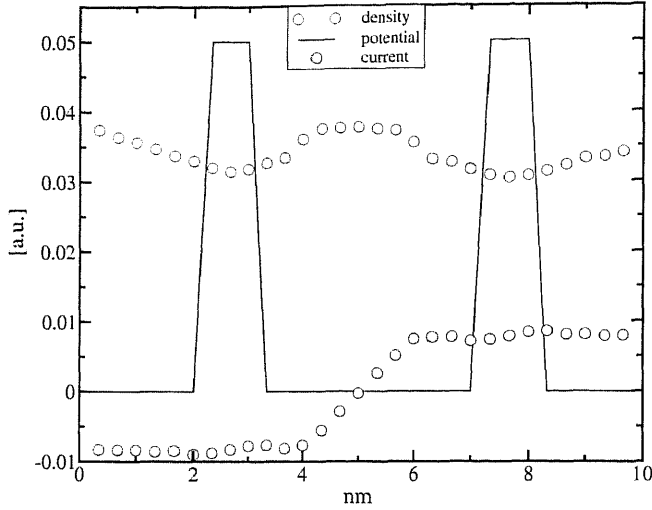


Fig. 15. – Device potential, current and particle density distribution in the device.

the assumed conditions, classical particles have a simple behavior which will be used as a reference for evaluation of the QMC results. Classical particles cannot overcome the barriers. The particle density would grow with time inside the well and no stationary solution exists for this case. The current density outside the well would be zero because no particles leave the well region. It would be zero also inside the well, due to the fact that the injection region is centered in the well and that the reflection from the barrier walls does not destroy the equilibrium distribution.

The QMC method provides the stationary solution which consistently characterizes the quantum nature of the transport process. The electron density plotted in fig. 15 remains independent of the simulation time. The density remains constant in the region of injection (middle part of the quantum well). It decreases outside the injection region well before the physical location of the barriers, because of the non-local character of the quantum potential. Indeed the potential out-scattering rate γ is remarkably high around the barriers on a distance determined by the coherence length $L_c = N\Delta x$. As shown in fig. 16, γ , which is on the order of 10^{-14} s^{-1} , assumes even higher values outside the barriers than inside.

The current density in fig. 15 clearly demonstrates the tunneling process. The density is non-zero in contrast with the classical case. Particles tunneling through the left barrier give rise to negative current, as opposite to those leaving through the right barrier. Outside the injection region the current densities to the left and right contacts are space-independent because of the current continuity.

A quantitative characterization of the tunneling process is given by the mean kinetic energy, fig. 16. As proportional to the square of the wave vector, the kinetic energy becomes negative in the barriers. This is the place where the wave vector of the tunneling

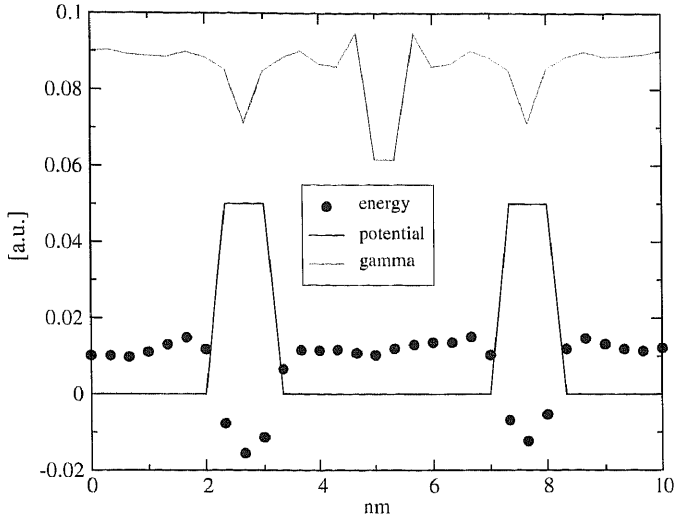


Fig. 16. – Device potential, out-scattering rate γ and kinetic energy distribution in the device.

particles transforms from a real to an imaginary quantity. The particles are injected in the well with mean energy corresponding to the equilibrium value $\simeq 13$ meV for one direction. Each individual scattering by the quantum potential changes the energy of the particle. In this experiment particles can gain energy well above 1 eV. Despite the fact that scattering occurs in the whole device, the equilibrium value is maintained in the regions with zero potential energy.

These results show that the scattering model works and we can model tunneling by using particles. The problem is that it works for the chosen energy and dimension scales which are both one order of magnitude less than their counterparts in real nanoelectronic devices. The problem comes from the rapid increase of the weight of the particles. Consider (48) with $E_\psi \simeq 1$ and $\psi_i \simeq \pm 10^{20}$ (the accumulated weight). One needs $N \simeq 10^{20}$ trajectories for a reliable estimate of E_ψ . One can estimate the mean accumulated weight W from the mean time T a trajectory spends in the device. T is given by the sum of all free-flight times. The number n of the scattering events is then $n = T\gamma$, and W is estimated as

$$W = \pm(3)^n = \left(1 + \frac{2\gamma}{\gamma}\right)^n = \left(1 + \frac{2\gamma T}{n}\right)^n \simeq \exp[2\gamma T].$$

It follows that the mean weight, and thus the variance, grows exponentially with the magnitude of the Wigner potential and the dwell time T . This result is in accordance with the exponential growth with time of the variance of the MC approach to Feynman path integrals [29]. If the device dimensions are larger than ten nanometers, T is commonly larger than a picosecond, while $\gamma \simeq 10^{15} \text{ s}^{-1}$ for a 0.3 eV potential barrier. This precludes the application of the approach to mesoscopic devices.

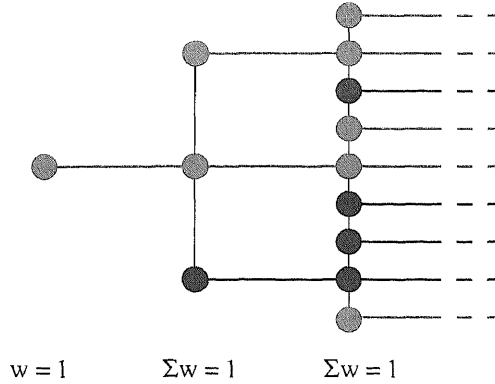


Fig. 17. – Schematic evolution of a Wigner particle in the weight-split model.

An alternative interpretation is needed in order to solve the problem with the accumulated weight. We explore the idea of particle splitting, which is an established approach [30] for statistical enhancement in classical Monte Carlo simulations. A particle entering a rarely visited region of the phase space can be split into sub-particles, each of which carries a fraction of the particle weight. This can be achieved by setting $p_2 = 1$ in (75). This modification changes entirely the interpretation of the quantum term. Now, the state $Q' = (\mathbf{r}', \mathbf{k}')$ which enters the interaction gives rise to three states, so that the Wigner potential is understood as a generation term. After the interaction, the initial particle survives in the same state, due to the delta-function in the brackets of (75). Two additional particles are generated by p_W^\pm in states $Q^\pm = (\mathbf{r}', \mathbf{k}^\pm)^{(1)}$. The trajectory now branches so that the weight carried by any branch keeps a constant magnitude and can change only the sign. It can be seen from (72) that the branching corresponds to a splitting of the integral into three integrals. Hence each branch continues with a free flight to contribute to the sample mean of ψ_1 . One of the contributions carries the minus sign of p_W^- . It is beneficial to assign a sign to the particle associated to each trajectory. Then the following transport process can be imagined. A positive particle is injected from the device boundaries. It drifts over a trajectory (58) until the interaction time generated according to p_t is reached. The particle does not “feel” the Wigner potential since after the interaction it remains in the same state. The next drift process continues on the same trajectory. The action of the potential is realized through the creation of two new particles in two phase space states. The particle related to p_W^+ (p_W^-) has the same (the opposite) sign as the primary particle. The created particles follow the same evolution process over their own trajectories. The schematic evolution of the particles in the weight decomposition model is shown in fig. 17. After each individual interaction any

⁽¹⁾ We note that there are always two states generated: $p^\pm(\mathbf{r}', \mathbf{0}) = 0$, since V_W is antisymmetric.

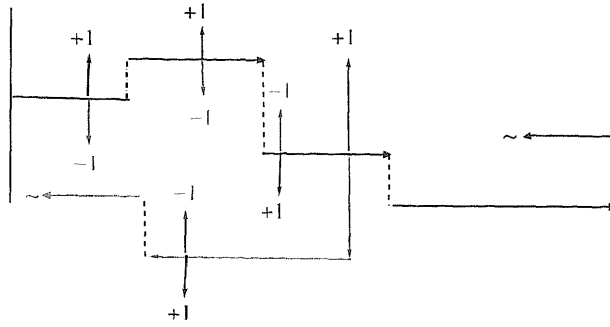


Fig. 18. – Schematic evolution of a Wigner particle interacting with phonons. The Wigner potential in the right part of the figure is negligible, which marks a classical region. No generation is presented. On the contrary, generation is very active in the middle part. The particle does not feel the generation process so that the overall evolution is classical. The trajectory at the bottom belongs to a generated particle. It evolves in the same way and carries the initial sign.

positive (negative) particle contributes to the estimator of $\psi_{\langle A \rangle}$ with $+(-)\Phi\psi_A$. Two particles which are in the same phase space point follow a common trajectory. If they have opposite sign, they give opposite contributions to $\psi_{\langle A \rangle}$. Moreover such particles create with the same probability for any point of the phase space particles with opposite sign. The net contribution of such particles to the physical averages as well as to the generation process is zero. It follows that particles with opposite sign which meet in the phase space can be annihilated. The coherent transport is characterized by processes of generation and annihilation of positive and negative particles.

9.3. Quantum transport with dissipation. – The above two limiting cases of the Wigner-Boltzmann transport can be combined without interference into a general picture of quantum transport with dissipation. The phonon interaction is inserted on top of the coherent picture, and affects the dynamics of the particles. They no longer remain on a single trajectory throughout the device, but are scattered to different trajectory segments after each process of drift, as shown in fig. 18. According to (75), the events of phonon and quantum interactions are complementary. The action of the Wigner potential on the interacting particle is equivalent to a self-scattering event, since it does not change the trajectory. From this analogy, it follows that the duration of the free flight on a given trajectory depends only on λ . The after-scattering state is selected by the phonon rate S through p_{ph} . We conclude that particles have the same Boltzmann-like behavior in both classical and quantum regions. The quantum character of the transport is marked by the generation process and the sign of the particles. The possibility to annihilate particles with opposite sign also remains true. The reason is that the evolution does not change its Markovian character in the case of phonons. Particles at a given phase space point still have a common probabilistic future and the considerations from the previous section apply. A condition for this is that the phonons are treated in a classical way. If the interaction is quantum, this property is not generally true because of the memory

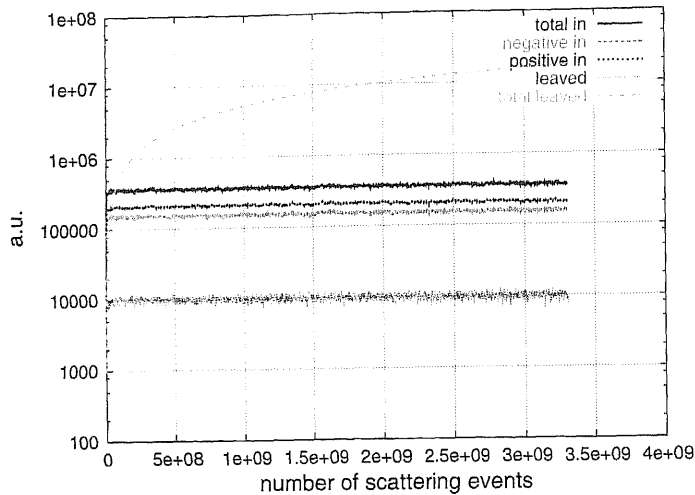


Fig. 19. – Total weight of the particles as a function of quantity proportional to the simulation time.

character of the evolution. We note that other interpretations conserving the absolute weight on a trajectory are possible. The interaction with the Wigner potential can be chosen twice as rare on the expense that four particles are created per such event. A reformulation of (75) can lead to events where quantum and phonon interactions occur

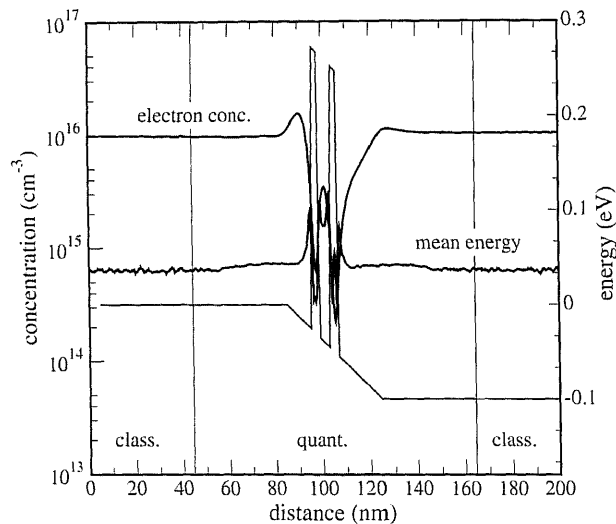


Fig. 20. – Energy and density distribution in a resonant tunneling diode. The simulation domain is large enough to include the classical regions of the device.

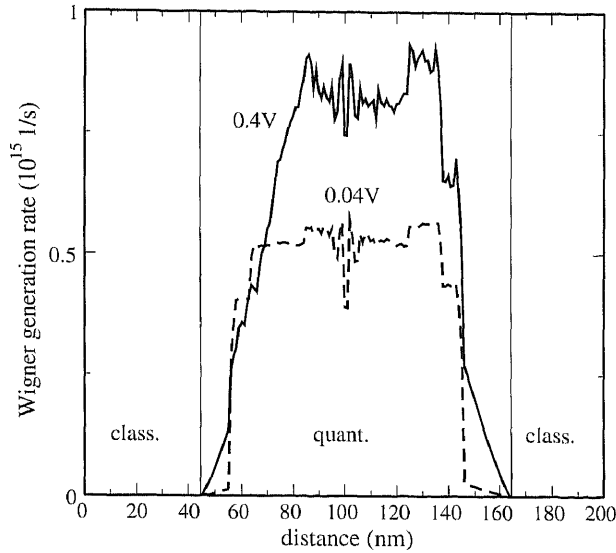


Fig. 21. – Wigner potential out-scattering rate for two bias values.

in the same instances [31]. The proposed particle picture is the most straightforward one which follows from this approach.

The decomposition of the weight greatly improves the performance of the Monte Carlo approach. The particle annihilation precludes the variance from growing exponentially by keeping the total particle weight in the device (now equal to the particle number inside) under a reasonable limit. Figure 19 shows that the particle number inside the device remains constant. The constant is much smaller than the total number of particles injected from the boundaries reached at the end of the simulation.

Actual nanostructures can be now simulated. Figure 20 shows the energy and carrier density distribution in a resonant tunneling diode. A seamless transition between the classical and quantum domains exists despite the strong onset of the Wigner potential shown in fig. 21. For a detailed considerations about the numerical aspects of the method, see [3].

We conclude that the presented particle approach provides both a model which facilitates the understanding of the Wigner transport picture of and a useful tool for quantum simulations.

* * *

Many of the results in these lecture notes are obtained in collaboration with H. KOSINA, Technical University of Vienna, T. GUROV and I. DIMOV, Bulgarian Academy of Sciences, C. JACOBONI, University of Modena, and D. VASILESKA and D. K. FERRY, Arizona State University.

APPENDIX A.

The freedom of choice of the initialization time directly follows from the fact that the wave vector \mathbf{k} remains constant in time for an unaccelerated trajectory. Actually this property is valid for a general Liouville operator with a force term which does not depend explicitly on time. A useful relation which follows from the stationarity assumed will be derived.

A trajectory is initialized by a phase space point (\mathbf{k}, \mathbf{r}) and a time t_0 :

$$(A.1) \quad \begin{aligned} \mathbf{R}(t; t_0, \mathbf{r}, \mathbf{k}) &= \mathbf{r} + \int_{t_0}^t \mathbf{v}(\mathbf{K}(y; \cdot)) dy, \\ \mathbf{K}(t; t_0, \mathbf{r}, \mathbf{k}) &= \mathbf{k} + \int_{t_0}^t \mathbf{F}(\mathbf{R}(y; \cdot)) dy. \end{aligned}$$

The order of t_0 and t is irrelevant. A trajectory is called forward if the evolution time is greater than the initialization time: $t > t_0$. Otherwise the trajectory is called backward. The limits of the time integration in (A.1) can be exchanged by changing the sign of the integrals. To describe a time-invariant system an absolute time scale is not needed. Only the time difference between two consecutive events is important. Invariance under time translation can be proven:

$$(A.2) \quad \begin{aligned} \mathbf{R}(t + \tau; t_0 + \tau, \mathbf{r}, \mathbf{k}) &= \mathbf{R}(t; t_0, \mathbf{r}, \mathbf{k}), \\ \mathbf{K}(t + \tau; t_0 + \tau, \mathbf{r}, \mathbf{k}) &= \mathbf{K}(t; t_0, \mathbf{r}, \mathbf{k}). \end{aligned}$$

This property will be used to adjust conveniently the time reference $t_0 = 0$ for each trajectory. A shortcut notation, $\mathbf{r}(t) = \mathbf{R}(t; 0, \mathbf{r}, \mathbf{k})$ and $\mathbf{k}(t) = \mathbf{K}(t; 0, \mathbf{r}, \mathbf{k})$, is introduced. A particularly useful relation is obtained from (A.2):

$$(A.3) \quad \int d\mathbf{r} d\mathbf{k} \int_{-T}^0 dt \phi(\mathbf{r}, \mathbf{k}, \mathbf{r}(t), \mathbf{k}(t)) = \int d\mathbf{r}' d\mathbf{k}' \int_0^T dt \phi(\mathbf{r}'(t), \mathbf{k}'(t), \mathbf{r}', \mathbf{k}').$$

Here $\mathbf{r}(t)$, $\mathbf{k}(t)$ is a backward trajectory initialized by \mathbf{r} , \mathbf{k} , while $\mathbf{r}'(t)$, $\mathbf{k}'(t)$ is a forward trajectory initialized by \mathbf{r}' , \mathbf{k}' . The relation is proven by introducing new integration variables $\mathbf{r}' = \mathbf{R}(t; 0, \mathbf{r}, \mathbf{k})$, $\mathbf{k}' = \mathbf{K}(t; 0, \mathbf{r}, \mathbf{k})$. Then

$$\begin{aligned} \mathbf{r} &= \mathbf{R}(0; t, \mathbf{r}', \mathbf{k}') = \mathbf{R}(-t; 0, \mathbf{r}', \mathbf{k}'), \\ \mathbf{k} &= \mathbf{K}(0; t, \mathbf{r}', \mathbf{k}') = \mathbf{K}(-t; 0, \mathbf{r}', \mathbf{k}'). \end{aligned}$$

According to the Liouville theorem, the phase volume is invariant under this transformation: $d\mathbf{r}' d\mathbf{k}' = d\mathbf{r} d\mathbf{k}$. The last step is to reverse the time by switching the sign of t and to recall that $\mathbf{R}(t; 0, \mathbf{r}', \mathbf{k}')$ and $\mathbf{K}(t; 0, \mathbf{r}', \mathbf{k}')$ are forward trajectories initialized by $(\mathbf{r}', \mathbf{k}')$ and denoted by $\mathbf{r}'(t)$, $\mathbf{k}'(t)$. In particular, for a given function μ it follows that

$$(A.4) \quad \int_{-T}^0 \mu(\mathbf{r}(y), \mathbf{k}(y)) dy = \int_0^T \mu(\mathbf{r}'(y), \mathbf{k}'(y)) dy.$$

REFERENCES

- [1] KAM H., CHANG L. and KING T., in *IEEE 2004 Silicon Nanoelectronics Workshop, Honolulu, Hawaii, June 2004*.
- [2] KATHAWALA G., MOHAMEND M. and RAVAIOLLI U., *J. Comput. Electron.*, **2** (2003) 85.
- [3] KOSINA H., KLIMECK G., NEDJALKOV M. and SELBERHERR S., in *Proceedings of Simulation of Semiconductor Processes and Devices, Boston, USA, September 2003*, pp. 171-174.
- [4] FRENSLEY W., *Rev. Mod. Phys.*, **62** (1990) 745.
- [5] SHIFREN L. and FERRY D. K., *J. Comput. Electron.*, **1** (2002) 55.
- [6] KLUKSDAHL N. C., KRIMAN A., FERRY D. K. and RINGHOFER C., *Phys. Rev. B*, **39** (1989) 7720.
- [7] TATARSKII V., *Sov. Phys. Usp.*, **26** (1983) 311.
- [8] ROSSI F., JACOBONI C. and NEDJALKOV M., *Semicond. Sci. Technol.*, **9** (1994) 934.
- [9] BORDONE P., BERTONI A., BRUNETTI R. and JACOBONI C., *Math. Comput. Simulat.*, **62** (2003) 307.
- [10] LEVINSON I., *Sov. Phys. JETP*, **30** (1970) 362.
- [11] BARKER J. and FERRY D. K., *Phys. Rev. Lett.*, **42** (1979) 1779.
- [12] NEDJALKOV M. and DIMOV I., *Math. Comput. Simulat.*, **47** (1998) 383.
- [13] SCHILP J., KUHN T. and MAHLER G., *Phys. Rev. B*, **50** (1994) 5435.
- [14] FERRY D. K. and BARKER J., *Solid-State Electron.*, **23** (1980) 545.
- [15] RAMMER J., *Rev. Mod. Phys.*, **63** (1991) 781.
- [16] GUROV T. and WHITLOCK P., *Math. Comput. Simulat.*, **60** (2002) 85.
- [17] RINGHOFER C., NEDJALKOV M., KOSINA H. and SELBERHERR S., *SIAM J. Appl. Math.*, **64** (2004) 1933.
- [18] HAAS S., ROSSI F. and KUHN T., *Phys. Rev. B*, **53** (1996) 12855.
- [19] ROSSI F. and KUHN T., *Rev. Mod. Phys.*, **74** (2002) 895.
- [20] SOBOL I., in *Numerical Methods Monte Carlo* (Nauka, Moscow), 1973.
- [21] NEDJALKOV M., KOSINA H. and SELBERHERR S., *Math. Comput. Simulat.*, **62** (2003) 453.
- [22] JACOBONI C., POLI P. and ROTA L., *Solid-State Electron.*, **31** (1988) 523.
- [23] JACOBONI C., in *International Electron Devices Meeting, Washington, D.C. (IEEE Electron Devices Society) 1989*, pp. 469-472.
- [24] ROTA L., JACOBONI C. and POLI P., *Solid-State Electron.*, **32** (1989) 1417.
- [25] NEDJALKOV M. and VITANOV P., *Solid-State Electron.*, **32** (1989) 893.
- [26] NEDJALKOV M. and VITANOV P., *Solid-State Electron.*, **33** (1990) 407.
- [27] KOSINA H., NEDJALKOV M. and SELBERHERR S., *IEEE Trans. Electron Devices*, **47** (2000) 1898.
- [28] JACOBONI C. and REGGIANI L., *Rev. Mod. Phys.*, **55** (1983) 645.
- [29] SCHMIDT T. and MOEHRING H., *Phys. Rev. A*, **48** (1993) R3418.
- [30] WORDELMAN C., KWAN T. and SNELL C., *IEEE Trans. Computer-Aided Design*, **17** (1998) 1230.
- [31] KOSINA H., NEDJALKOV M. and SELBERHERR S., *Monte Carlo Methods Appl.*, **10** (2004) 359.