TRANSPORT IN SUBMICRON DEVICES
D. Ferry

To cite this version:
D. Ferry. TRANSPORT IN SUBMICRON DEVICES. Journal de Physique Colloques, 1981, 42 (C7), pp.C7-253-C7-261. <10.1051/jphyscol:1981730>. <jpa-00221667>

HAL Id: jpa-00221667
https://hal.archives-ouvertes.fr/jpa-00221667
Submitted on 1 Jan 1981

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
TRANSPORT IN SUBMICRON DEVICES

D.K. Ferry

Colorado State University, Fort Collins, Colorado 80523, U.S.A.

Résumé - Alors que la taille des composants semiconducteurs devient plus petite, on peut s'attendre à ce que les échelles de temps et de distance dans ces composants soient suffisamment faibles pour que l'approche semiclassique de la théorie du transport incorporée dans l'Equation de Transport de Boltzmann (BTE) ne soit plus valable. Les réponses transitoires spatiale et temporelle durent en fait pendant une fraction appréciable du temps de transit des porteurs dans le composant. Il faut donc développer des équations cinétiques de transport correctement retardées incluant les effets de mémoire et de durée non nulle des collisions. Dans cet article, celles-ci sont obtenues à partir d'une équation fondamentale du transport quantique. Nous en déduisons les équations cinétiques de variables dynamiques comme le moment et l'énergie.

Abstract - As semiconductor devices become smaller, it is expected that the relevant temporal and spatial time scales in these devices become sufficiently small that the semi-classical approach to transport theory, as embodied in the Boltzmann transport equation (BTE), is no longer valid. The temporal and spatial transient response in fact lasts for a sizable fraction of the transit time of the carriers through the device. Thus, kinetic transport equations must be developed that are properly retarded to include memory functional and non-zero collision duration effects. In this paper, properly retarded transport equations are derived from a basic quantum transport equation and kinetic equations for dynamic observables such as momentum and energy are developed.

1. Introduction. - Essentially all theoretical investigations of hot electron transport in semiconductors are based on a one-electron transport equation, usually the Boltzmann transport equation (BTE) [1]. Indeed, the overriding theoretical concern in such high-field transport is primarily the solution of the transport equation to ascertain the form which the distribution function takes in the presence of the electric field. However, for transport purposes, this is not an end product, since integrals must be carried out over the distribution function in order to evaluate the transport coefficients. In applications to semiconductor devices, however, the full solution of the BTE is usually too complicated to determine at each spatial point within the device, and transport equations for relevant observables such as energy and momentum are preferred [2,3]. Such transport equations are obtained by taking moments of the BTE and these directly relate to the normal hydrodynamic semiconductor equations. In general, the complicated nature of the BTE precludes solving it analytically, and the existence of the various moment equations is based upon

*This work supported in part by the U.S. Office of Naval Research

Article published online by EDP Sciences and available at http://dx.doi.org/10.1051/jphyscol:1981730
a number of assumptions, the most common of which is that the distribution function 
can be represented as a displaced Maxwellian [4].

In small semiconductor devices, however, the time scales are such that the use 
of the BTE must be questioned [5]. Traditional semi-classical approaches, such as 
that of the BTE, assume that the response of the carriers to any applied force 
occurs simultaneously with the applied force, even though the system may undergo 
subsequent relaxation to a nonequilibrium steady-state. On the short-time scale 
though, a truly causal theory introduces memory effects which lead to convolution 
integrals in the transport coefficients [6-10], so that the resultant kinetic equa-
tion is not of the Markovian type. For the steady-state, this results in a colli-
sion operator that depends upon the frequency of the driving field. The problem is 
complicated as well by the fact that the time scales of interest are usually suffi-
ciently short that the collision duration $\tau_c$ is no longer negligible in comparison, 
thus leading to further frequency dependences of the collision operators. In the 
BTE, the collision terms assume point collisions of zero duration, which is fine 
when the mean time between collisions is large, but at high fields in small semi-
conductor devices, this is no longer the case, and correction terms must be gener-
ated to account for these "memory" effects.

Previously, we have discussed the development of balance equations from a modi-
fied BTE [11], based upon a full quantum kinetic equation. Direct and general 
methods for the derivation of such quantum kinetic equations, for the one-electron 
density matrix for electrons in phonon and other arbitrary fields of force, have 
 existed for some time [12]. Such treatments avoid the random phase approximation 
and yield kinetic equations which indeed are non-Markovian. In addition, effects 
due to interference between the driving fields and the scatterers are obtained, 
leading to collisional shielding of the driving field as well as to an intra-collis-
sional field effect [12-14], which is a disturbance of the scattering by the driving 
field. Using such a quantum kinetic equation for the distribution function, the 
moment equations can be obtained in a straightforward manner and are modified from 
their more normal BTE-derived counterparts. These equations show memory effects as 
well as additional retardation due to the non-zero collision duration. However, the 
basis of these equations is still subject to rather stringent approximations. 
Indeed, a rather unique form must be assumed for the distribution function [11]; it 
is specified as a retarded, displaced Maxwellian of rather special form. It remains 
unclear to what extent the form assumed for the distribution function affects the 
resulting moment equations and whether the actual form assumed is a proper ansatz.

The above concerns over the detailed form of the moment equations can be 
removed by deriving these equations directly from the quantum transport equations. 
The exact solutions of the Liouville equations describe the time evolution of a 
statistical ensemble at any time interval after applying an external field. If the 
rate of inter-carrier scattering is high, then after a short time interval $\tau_1$, 

smaller than any appropriate time-scale of interest, the evolution of the nonequilibrium density matrix must be independent of the initial distribution, and there should be a reduction in the number of parameters necessary to describe the nonequilibrium response of the system [15]. It is therefore possible to assume a nonequilibrium statistical operator which is smoothed in its micro-fluctuations and from the very beginning describe the slow evolution of the system for time intervals which are larger than \( \tau_1 \) [16,17]. This approach has previously been applied to hot electron problems by Kalashnikov [18,19]. By utilizing such an approach, both the relevant moment equations and the form of the distribution function itself are obtained directly prior to the extension to the semi-classical transport properties. In this paper, we review this approach. The nonequilibrium density matrix and the moment equations are developed. It is shown that the equations obtained in [11] are correct under certain conditions, for which the assumed form of the distribution function is also correct. In this treatment, a homogeneous field is assumed, although the treatment is readily extendable to take into account the spatial inhomogeneities of the system and the derivation of the scatterers from thermal equilibrium [19], such as in intense picosecond laser excitation and light scattering [20].

2. The Nonequilibrium Distribution.—It is assumed, based upon the work of Bugoluibov [15], that for time \( t > \tau_1 \), correlations with lifetimes less than \( \tau_1 \) can be ignored and the state of the system may be described by a reduced set of macroscopic observables \( Q_m (t) \) which are the average values, over the nonequilibrium ensemble, of a set of dynamical variables \( P_m \). These dynamical variables, and their conjugate forces \( f_m (t) \), may then be used to define a nonequilibrium statistical ensemble. The approach is based upon Poincaré's theorem on integral invariants generalized to quantum systems [19,21]. Thus, the system density matrix should be constructable from the principle invariants of the system. Additionally, an auxiliary (local) equilibrium density matrix \( \rho_0 \) is defined as a quasi-equilibrium quantity. This latter function serves as an idealized initial condition for the system after the randomization, characterized by \( \tau_1 \), has occurred, and from which the system evolves under dynamical laws governed by its Hamiltonian. This means that at all times \( t > \tau_1 \), the nonequilibrium density matrix, which must satisfy the Liouville-Von Neumann equation, is a functional of \( \rho_0 \).

Thus, on a time scale greater than \( \tau_1 \), the system may be characterized by the average values of the set of operators \( P_m \), defined as

\[
P_m = \{ H_{c}, P_{c} N, \ldots \} ,
\]

where \( P_c \) is the operator for the total momentum of the carriers and \( N \) is the carrier number operator. The term \( H_c \) contains the carrier Hamiltonian, the coherent part of any interaction with the environment, and the carrier-carrier interaction. However,
it is assumed that carrier eigenfunctions can be chosen such that this term is diagonal. The set $P_m$ in (1) is a minimal basis set, in that it is assumed here that the lattice remains in equilibrium. The $P_m$ satisfy equations of motion

$$-i\hbar \dot{P}_m = \hbar P_m H_m,$$

where the dot over $P$ indicates the time derivative. Thus, for

$$H_S = H_c + H_{cp} - e\vec{F} \cdot \vec{r},$$

where $H_c$ contains the incoherent interaction with the lattice (and environment),

$$H_c = e\vec{F} \cdot P_c/m + i\hbar_{c(L)}/\hbar,$$

$$\dot{P}_c = e\vec{F} + i\hbar_{c(L)}/\hbar,$$

$$\dot{N} = 0,$$

where $\hbar_{c(L)} = \hbar_p P_m$ (the carat over $H$ in (2) and (4) indicates a commutator-generating super-operator). The next step is to introduce a set of time-dependent parameters $f_m(t)$ (which are c-numbers) constructed to be thermodynamic conjugates of the $Q_m$ as

$$f_m(t) = -\beta(t) \cdot \{1, v_d(t), \mu(t)-m \cdot v_d^2(t), \ldots\},$$

where $\beta(t)$ multiplies each term in the brackets and is the inverse carrier temperature ($= 1/k_B T$), $v_d(t)$ is the drift velocity, and $\mu(t)$ is the chemical potential.

The asymptotic, time-smoothed quantities $B_m$ may be defined from the $P_m f_m(t)$ as

$$B_m(t) = \int_0^\infty dt' e^{-\beta t'} f_m(t-t')P_m(t'),$$

where

$$P_m(t') = e^{+i\hbar t'/\hbar} P_m(0).$$

The quantities $B_m(t)$ satisfy the Liouville equation in the limit $s \to 0^+$ and are the quasi-invariant portions of the products $P_m f_m(t)$ with respect to evolution under the Hamiltonian $H_s$. Thus, they are integrals of motion in this limit and any statistical operator constructed from these quantities will also be an integral of Liouville's equation. The operation of taking the quasi-invariant part is also used in scattering theory to assure that retarded, causal solutions to Schrodinger's equation are used [22]. In order to assure that $f_m(t)$ and $Q_m = <P>_m^t$ (the t superscript denotes a time-varying average) are thermodynamically conjugate, we impose the condition

$$<P>_m^t = <P>_m^0.$$
where \( < \cdot >^t \) denotes an average over the nonequilibrium density matrix \( \rho(t) \) and \( < \cdot >^o \) indicates an average over the quasi-equilibrium \( \rho^o(t) \).

We can now write the nonequilibrium density operator in the form

\[
\rho(t) = \exp\{- \phi - \sum_m B_m(t)\}
\]

where

\[
\phi = \ln \text{Tr} \exp\{- \sum_m B_m(t)\}
\]

The choice of (9) for \( \rho(t) \) assures that this quantity reduces to the generalized Gibbs distribution in thermal equilibrium and results in a positive entropy production in the system. Indeed, by defining the entropy operator as

\[
S(t,o) = \phi^o + \sum_m P_m^f(t)
\]

with

\[
\phi^o = \ln \text{Tr} \exp\{- \sum_m P_m^f(t)\}
\]

we may combine (6) and (9) so that

\[
\rho(t) = \exp\{- S(t,o) - \int_0^\infty dt'e^{-st'} \dot{S}(t-t',t')\}
\]

after integration by parts, where

\[
\dot{S}(t-t',t') = \sum_m \{ \dot{P}_m(t')f_m(t-t') - [P_m(t') - \langle P_m^o \rangle_{^t} t' f_m(t-t')\}
\]

Clearly then \( \rho^o(t) = \exp\{- S(t,o)\} \) is the quasi-equilibrium operator. Here, \( \dot{P}_{-m} \) can be found from (4), and

\[
\frac{d f_m(t)}{dt} = \sum_{\lambda\delta} \frac{\delta f_m}{\delta P_{<\lambda\delta>}} \dot{P}_{<\lambda\delta>} = \frac{1}{A} \sum_{\lambda\delta} \frac{\delta f_m}{\delta P_{<\lambda\delta>}} \hat{H} P_{<\lambda\delta>}
\]

where the \( \delta \) indicates a functional derivative.

It is obvious that terms of first and second order in \( H_{cp} \) appear in (14). Therefore, to obtain the balance equations up to second-order in the electron-phonon interaction (the first non-vanishing order), we use the following iterative expansion of the density matrix:

\[
\rho(t) = \rho^o(t) + \tau S(t,o) \int_0^\infty dt'e^{-st'} \int_0^\tau dt e^{-\tau S(t,o)} \dot{S}(t-t',t') \rho^o(t)
\]

where \( \rho^o(t) \) is the quasi-equilibrium density matrix corresponding to the isolated carrier distribution at given values of the average energy and momentum.

3. The Balance Equations.- With the non-equilibrium density matrix of (16), we can now find explicit transport equations for the generalized operators. We suppose, as in (4), that the equations of motion of the operators \( P_m \) have the form
\[ \dot{\rho}_m = \sum_{n} \alpha_{mn} \rho_n + \dot{\rho}_m(L), \]  

where the \( \alpha_{mn} \) and \( \dot{\rho}_m(L) \) are defined in (4). Then, using (16) to construct the average, we find that

\[
\langle \rho_m \rangle = \sum_{n} \alpha_{mn} \rho_n + \int_0^t dt' \sum_{L} \langle \rho_m(L) \rangle \dot{\rho}_m(L)(t') - t^F_\xi(t-t') + 
\]

\[
+ \sum_{L} \alpha_{mn} \rho_n \rho_m(L)(t') - t^F_\xi(t') \text{ ,} \tag{18}\]

where the last term has been integrated by parts and the limit \( s \to o^+ \) has been taken throughout. The upper limit on the second term on the right-hand side follows from the assumption that \( f_L(t) = 0 \) for \( t < 0 \). The quantum correlation functions are defined by

\[
\langle \alpha; B(t') \rangle = \frac{1}{N} \int_0^t dt' \langle \alpha(t') B(t') \rangle - \langle \alpha \rangle \text{ .} \tag{19}\]

and \( B(t') \) is given by (7). The energy and momentum balance equations are then given by

\[
\frac{d\langle \epsilon \rangle}{dt} = \frac{1}{N} \frac{\hat{\epsilon}_c}{\rho_c} = eF - \int_0^t dt' \langle \epsilon(t') \rangle + \beta_e(t-t') \rho_c(L)(t') - \beta_e(t) \langle \rho_c(t) \rangle + v_d(t) \rho_c(t) \text{ ,} \tag{20}\]

where \( \beta' = \beta - \beta_e \), and

\[
\frac{m \frac{dv}{dt}}{dt} = \frac{1}{N} \frac{\hat{v}_c}{\rho_c} = eF - \int_0^t dt' \langle v(t') \rangle + \beta_e(t-t') \rho_c(L)(t') - \beta_e(t) \langle \rho_c(t) \rangle + v_d(t) \rho_c(t) \text{ .} \tag{21}\]

Over a broad interval of the electric field strength, the kinetic energy associated with the electron drift is much smaller than the mean energy of a carrier; i.e., \( mv^2_d \ll e \) [1]. Therefore, in the expression for the quasi-equilibrium density matrix \( \rho_c(t) \) we may neglect the drift velocity \( v_d \). Then the correlation functions which contain the products \( \rho_c(L) \dot{\rho}_c(L) \) and \( \dot{\rho}_c(L) \rho_c(L) \) vanish as in the case of thermal equilibrium [18]. Then, the correlation functions simplify and (20-21) become

\[
\frac{d\langle \epsilon \rangle}{dt} = eF v_d(t) \left[ 1 - \langle \rho_c(t) \rho_c(t) \rangle \beta_e(t)/m \right] - \int_0^t dt' \beta_e(t-t') \rho_c(L)(t') - \beta_e(t) \langle \rho_c(t) \rangle + v_d(t) \rho_c(t) \text{ ,} \tag{22}\]

and

\[
\frac{m \frac{dv}{dt}}{dt} = eF - \int_0^t dt' \beta_e(t-t') v_d(t-t') \rho_c(L)(t') - \beta_e(t) \langle \rho_c(t) \rangle + v_d(t) \rho_c(t) \text{ .} \tag{23}\]
where we have used $<\mathcal{P}_c> = mNvd$. The remaining correlation functions are the collision-relaxation memory functions [6], and the forms of (22), (23) are that obtained in the drifted Maxwellian approximation [11], although the details of the scattering function are somewhat different. Indeed, $\rho(t)$ is almost identical to that taken in [11] for this case. The quantity multiplying the force term in (22) has an additional term not found from the normal semi-classical equations, and represents an additional delay on the increase of energy. The second factor in the square brackets usually produces only a negligible change in the velocity response. However, it is significant during the transient phase for the energy response and appears even in classical Monte Carlo calculations [23]. That this memory effect should appear in the energy equation follows from the fact that $\mathcal{E} = \frac{3}{2} k_B T_e = \frac{3}{2} \beta_e^{-1}$. The temperature is in effect a measure of the fluctuations and cannot be changed until the correlation of these fluctuations is broken up by the field.

The relaxation functions have a gain-loss structure as is well known. For example, the energy-loss rate can be written as [18]
\[
\dot{\mathcal{E}}(t') = \frac{2\pi}{\hbar} \sum_{k,k',q} \langle \mathcal{H}_{c}(L) \rangle \langle \mathcal{H}_{c}(L) \rangle^{*} \frac{1}{\hbar^{2}}(\mathcal{H}_{q})^{2}(N_{q} + 1)f_{k},(1 - f_{k}) - N_{q} f_{k}(1 - f_{k}),g(\mathcal{E}_{k} - \mathcal{E}_{k}, - \mathcal{H}_{q},t), \tag{24}
\]
where [24]
\[
g(x,t) = \frac{2}{\pi t} \int_{0}^{t} \frac{\sin(xt')}{x} dt', \tag{25}
\]
from time-dependent perturbation theory, and
\[
f_{k} = \{\exp[(\mathcal{E}_{k} - \mu_{o}N)\beta_{e}] - 1\}^{-1}. \tag{26}
\]
The limit of $t >> \tau_{c}$ gives $g(x,t)$ as the collision-duration broadened joint-spectral-density function [14]. But in high electric fields, where the duration of $g(x,t)$ can be long, it is possible to have field acceleration during the collision due to the time variation of energy in (24) [12-14,24]. This intra-collisional field effect can be introduced into (22), (23) by an additional retardation due to the effective collision duration $\tau_{c}$ as [11]
\[
\dot{i}(t') = \frac{1}{\tau_{c}} \int_{0}^{t'} \frac{\sin(xt')}{x} \chi^{*} \langle \mathcal{H}_{c}(L) \rangle \langle \mathcal{H}_{c}(L) \rangle^{*} \frac{1}{\hbar^{2}}(\mathcal{H}_{q})^{2}(N_{q} + 1)f_{k},(1 - f_{k}) - N_{q} f_{k}(1 - f_{k}),g(\mathcal{E}_{k} - \mathcal{E}_{k}, - \mathcal{H}_{q},t), \tag{27}
\]
where $\dot{i}$ under the integral is evaluated with the long-time joint-spectral density function. Here $\tau_{c}$ can be evaluated directly from the principle duration of (25).

4. Discussion.- It is clear from the present approach that the normal semi-classical balance equations can be obtained from a more accurate quantum transport approach. However, several points must be made. First, as is evident in (20) and (21), the individual balance equations are a fully coupled pair unless $(mv_{d}^{2}/\langle\mathcal{E}\rangle) \ll 1$. It is clear that a difference arises here from the semi-classical approach.
as several treatments of semiconductor transport have found this term to be comparable to $I$ in certain operating regions of semiconductor devices [25,26]. The extra terms leading to the direct coupling between the equations appear to be residual terms from the field-scatterer interference effects [12,14]. In this role, they lead to a memory effect which modifies the effective mass equation at $t = 0^+$. Indeed, the extra correlation function in the driving term of (22) represents the effects on the motion in the single-particle space due to initial correlations with other particles [27]. While this term is common in quantum transport [27,28], it has generally been regarded as small and negligible in classical transport [29,30]. Such is not the case in the fast, transient response regime. The indirect coupling appearing in the relaxational correlation functions directly points out, that although the two correlation functions differ slightly, one cannot readily identify a solely momentum relaxation process and a solely energy relaxation process. The second critical point is that usage of balance equations in which the convolution integrals are omitted is restricted to cases of $t >> \tau_e, \tau_m$. Indeed, in this limit the relaxation terms can be simplified. However, the use of these un-retarded simplifications to treat transient response, where $t \leq \tau_e, \tau_m$, is clearly in error.

References
19. The nonequilibrium statistical operator is reviewed by Zubarev D. N., Non-equilibrium Statistical Thermodynamics (Consultants Bureau, New York, 1974). See also the discussion on this approach in Ferry D. K. and Barker J. R., Sol.-State Electron. 23 (1980) 531.
23. Lugli P., Zimmermann J. and Ferry D. K., these proceedings.