## 17.2 Nonequilibrium electron ensembles

The term ‘‘hot electrons’’ purports a non-equilibrium ensemble of high-energy carriers. It is often possible to pump external energy (e.g., by shining light or applying an electric field) directly into the system of carriers. If the power input into the electronic system exceeds the rate of energy loss by that system to the lattice, then the carriers ‘‘heat up’’ and their velocity distribution $f(\mathbf{v})$ deviates significantly from the equilibrium Maxwellian form.[1] In general, the time-dependent distribution function $f(t, \mathbf{r}, \mathbf{v})$ can be determined by solving the Boltzmann transport equation,

$$\frac{\partial f}{\partial t} + \mathbf{v}\cdot\frac{\partial f}{\partial \mathbf{r}} + \mathbf{a}\cdot\frac{\partial f}{\partial \mathbf{v}} = \left[\frac{\partial f}{\partial t}\right]_{\text{coll}}, \tag{1}$$

where **a** is the acceleration and the collision integral in the right-hand side is a linear functional on $f$. In a steady state, $f(\mathbf{r}, \mathbf{v})$ does not explicitly depend on time but it may still be a function of the spatial position.

Solution of Eq. (1) is a complicated task even for the simplest scattering models involving scattering of electrons among themselves as well as with phonons and impurities. It is expected from a reasonable model of interaction between the electronic system and the thermal bath that collisions will *restore* the equilibrium distribution function $f_{\text{eq}}(\mathbf{r}, \mathbf{v})$ from any initial distribution − although this property is often difficult to prove mathematically. Since we know that thermodynamic equilibrium cannot be destroyed by scattering of electrons by a thermal bath, the collision integral must vanish if $f_{\text{eq}}(\mathbf{r}, \mathbf{v})$ is substituted for $f$. These properties are manifestly expressed in the model which approximates the collision integral by

1. Note that one can also *cool* the carrier system by making it do work against an external field at a fast rate compared to power replenished by the lattice.

$$\left[ \frac{\partial f}{\partial t} \right]_{\text{coll}} = \frac{f\,(t\,,\mathbf{v}) - f_{\text{eq}}(\mathbf{v})}{\tau(\mathbf{v})}\ , \tag{2}$$

where for brevity we have omitted the possible position dependence. The characteristic time constant $\tau(\mathbf{v})$ is called the relaxation time and the whole model (2) is called the relaxation time approximation. In this approximation the perturbed distribution will exponentially relax to equilibrium when the perturbing influence is removed.

The relaxation time approximation (2) is too crude in practice, because different characteristics of the distribution relax with different rates. Thus when collisions are predominantly elastic, it is natural that the first moment $\langle\mathbf{v}\rangle$ of $f\,(\mathbf{v})$ relaxes rapidly while the second moment $\langle\mathbf{v}^2\rangle$ takes a long time to relax. The first moment characterizes the electron drift velocity (or average crystal momentum) and the corresponding time constant, $\tau_m$, is called the momentum relaxation time, while the longer time $\tau_\varepsilon$ is called the energy relaxation time. It is often a good approximation to characterize the electron ensemble by a model distribution function, which embodies the relevant relaxation kinetics. The choice of an appropriate distribution function may also depend on time $t$ if the electric fields rapidly vary (on the scale of $\tau_\varepsilon$ or $\tau_m$). Similarly, the choice of a model $f\,(\mathbf{v})$ may depend on the position $\mathbf{r}$ in the device. Consider hot-electron models, commonly occurring in devices.

*17.2.1 Quasi-thermal.* An effective temperature $T_e$ is always established in an electronic ensemble of sufficiently high concentration, when the electron-electron (ee) interaction dominates over both phonon and impurity scattering. Starting from any initial distribution, ee collisions lead to an equilibrium within the electron gas. Since the center-of-mass velocity of colliding electrons does not change, the drift velocity $\langle\mathbf{v}\rangle$ of the electron gas remains a constant of the motion.[2] The distribution function is then of the form

2. This is true, so long as Umklapp processes are insignificant. When Umklapp is included, momentum relaxation occurs even under the pure ee interaction.

$$f(\mathbf{v}) = \left\{ 1 + \exp[(m\mathbf{v}^2/2 - m\langle\mathbf{v}\rangle\cdot\mathbf{v} - E_F)/kT_e] \right\}^{-1} , \qquad (3)$$

called the displaced Fermi distribution. It is characterized by the effective temperature $T_e$, the drift velocity $\langle\mathbf{v}\rangle$, and the Fermi level $E_F$, determined respectively by the conservation of energy, momentum, and the number of particles. In the reference frame that travels with the velocity $\langle\mathbf{v}\rangle$, the distribution (3) looks like an ordinary Fermi function. For a non-degenerate gas, Eq. (3) reduces to the form of a displaced Maxwellian distribution,

$$f(\mathbf{v}) = e^{-\frac{m\mathbf{v}^2/2 - m\langle\mathbf{v}\rangle\cdot\mathbf{v}}{kT_e}} . \qquad (3')$$

In the nondegenerate limit $T_e \gg E_F$, the effective temperature $T_e$ is related to the average electron energy $\langle E\rangle$ by the well-known formula,

$$\langle E\rangle = \frac{3}{2}kT_e + \frac{1}{2}m\langle\mathbf{v}\rangle^2 . \qquad (4)$$

In the opposite limit, $E_F \gg kT_e$, the average energy does not depend on $T_e$ and is determined only by the carrier density $n$:

$$\langle E\rangle = \frac{3}{5}E_F + \frac{1}{2}m\langle\mathbf{v}\rangle^2 , \quad \text{where } E_F = (3\pi^2)^{2/3}\frac{\hbar^2}{2m}n^{2/3} . \qquad (5)$$

In the intermediate range, the expression of $\langle E\rangle$ in terms of parameters $T_e$ and $E_F$ can be written in the form of a quadrature (see, e.g., Landau and Lifshitz, 1980).

Any function of the form (3) will make vanish the collision integral that contains only ee interaction. If interaction with impurities is included, the collision integral will vanish only provided $\langle\mathbf{v}\rangle = 0$. Including interaction with phonons, the collision integral will vanish only if $T_e$ coincides with the lattice temperature $T$.

In the situation when energy is continuously pumped into the electron gas by an electric field or electromagnetic radiation, one cannot ignore other forms of scattering, however slow they may be compared to ee interaction. Indeed, otherwise there would be nothing to check a runaway of the average energy and/or drift velocity. The correct approach is to use the so-

called *adiabatic* approximation in which the distribution function is assumed in the form (3) with the effective parameters assumed to follow the variations in the external input, as governed by the balance equations for energy and momentum,

$$\frac{d<E>}{dt} = <\mathcal{P}> - \frac{<E>}{\tau_\varepsilon} \quad , \tag{6a}$$

$$\frac{d<\mathbf{v}>}{dt} = <\mathbf{a}> - \frac{<\mathbf{v}>}{\tau_m} \quad , \tag{6b}$$

where $<\mathcal{P}>$ and $<\mathbf{a}>$ are, respectively, the average power input and acceleration. For the electronic ensemble (3) moving under a force $\mathbf{F}$, one has approximately $<\mathbf{a}> = \mathbf{F}/m$ and $<\mathcal{P}> = \mathbf{F}\cdot<\mathbf{v}>$. Parameters $\tau_\varepsilon$ and $\tau_m$ in these equations are determined self-consistently from the Boltzmann equation − by substituting into the collision integral for slow processes the distribution function in the form (3) − which is assumed maintained by the rapid ee interaction.

Thus, in the adiabatic approximation parameters $\tau_\varepsilon$ and $\tau_m$ of the balance equations (6) themselves become functions of the effective temperature $T_e$. The momentum relaxation time is usually much shorter than the energy relaxation time, $\tau_m \ll \tau_\varepsilon$. Indeed, elastic collisions are dominant for momentum relaxation, whereas an effective energy relaxation requires several inelastic interactions with phonons. Therefore, on the time scale of the momentum relaxation, the average electron energy can be considered quasi-static and while $\tau_m = \tau_m(T_e)$ in the balance equation (6b), the effective temperature $T_e$ itself is a function of time, ‘‘slowly’’ rising with a characteristic time $\tau_\varepsilon$. Thus, at short times ($t < \tau_E$) after the imposition of a strong electric field, the carrier drift occurs with a low-field mobility and the velocity can substantially overshoot its steady-state value (Ruch, 1972). The overshoot phenomenon has become quite important in determining the speed of modern transistors with ultra-short gate lengths (Sai-Halasz et al. 1988). For a 0.25 μm $n$-channel Si MOSFET this hot-electron effect contributes a 20% enhancement in the transistor speed (Pinto, 1991).

The qualitative explanation given above is quite adequate for describing the overshoot in Si and Ge, where there is a strong dependence of the phonon scattering rates on the carrier

energy. In GaAs and some other III-V compounds, the dominant scattering process in the Γ valley is due to polar optical phonons, and the rate of these processes at sufficiently high electron energies becomes nearly independent of energy. In such materials the mobility degradation at high energies is associated with the transfer of electrons to the lower-mobility upper valleys (Ridley and Watkins, 1961). Consequently the overshoot is not seen below the threshold field for the negative differential mobility effect ($F \gtrsim 3.2$ kV/cm in GaAs and $\gtrsim 11$ kV/cm in InP).

*17.2.2 Ballistic.* If the force acting on electrons suddenly changes, their subsequent motion for a short time may be considered without taking collisions into account.[3] The time interval $t \lesssim \tau_\beta$ when this is possible, is determined not only by the momentum relaxation time $\tau_m$ but also by the characteristic time $\tau_{ee}$ of interelectron scattering. By definition, the ee scattering rate is included neither in the rate $1/\tau_m$ which enters Eq. (6b) nor in the expression $\mu = (e/m)\tau_m$ for steady-state mobility, because ee collisions have no direct effect on $\langle \mathbf{v} \rangle$. Of course, they can have a very strong effect indirectly, by influencing other collision processes through the shape of the distribution function $f(\mathbf{v})$. At carrier concentrations $n \gtrsim 10^{17}\ \mathrm{cm}^{-3}$ the characteristic time $\tau_\beta$ for ballistic transport,

$$\frac{1}{\tau_\beta} = \frac{1}{\tau_m} + \frac{1}{\tau_{ee}} \ , \tag{7}$$

may be considerably shorter than $\tau_m$.

The concept of ballistic motion also applies to the *steady-state* transport. In this case one considers regions a short distance $d \lesssim \lambda_\beta$ away from an abrupt potential variation, Fig. 1. The characteristic length $\lambda_\beta$ is related to $\tau_\beta$ by $\lambda_\beta = \langle v \rangle \tau_\beta$ and both quantities depend on the shape

3. The common term ‘‘ballistic’’ applied to this time of motion, conjures up the image of a projectile moving in airless space. This image is not very apt, since interesting properties of the electronic motion in the ballistic regime often depend crucially on the band structure, which defies cannonball analogy.

of the distribution function $f$ (**v**). Instead of parameters pertaining to an electron ensemble, one often defines similar parameters for a given state of electron motion, viz. its lifetime $\tau(\mathbf{v})$ and the mean free path $\lambda(\mathbf{v})$, as limited by collisions with the lattice and other electrons. These definitions practically coincide for narrow distributions.

Steady-state ballistic transport was first demonstrated experimentally in unipolar heterostructures, illustrated in Fig. 2. The idea of ballistic-electron spectroscopy (Hesto et al., 1982) consists in the following: one measures the dependence of the collector current $I_C$ on the collector-base bias $V_{\rm CB}$ at a fixed emitter-base bias $V_{\rm BE}$ and plots $(\partial I_C/\partial V_{\rm CB})_{V_{\rm BE}}$ versus $V_{\rm CB}$. If certain important conditions are met, the resultant curve is proportional to the number of carriers arriving at the collector barrier with a normal component of the kinetic energy (i.e., portion of the energy corresponding to the motion normal to the barrier) equal to the barrier height $\Phi_C$. For this to be true, one must ensure that the collector barrier height linearly depends on the bias, $\delta\Phi_C \propto \delta V_{\rm CB}$, and that the collector bias does not affect the hot-electron energy distribution in the base, the emitter injection efficiency, and the above-barrier QM reflection. Studying the $I_C(V_{\rm CB})$ dependence of a GaAs/AlGaAs heterostructure similar to that illustrated in Fig. 2, Heiblum et al (1985) found a sharp peak in the derivative characteristic. According to authors interpretation, the majority of electrons contributing to the peak arrive at the analyzer without a single scattering event − otherwise, the peak would be displaced by at least 36 mV (corresponding to the optical-phonon energy $\hbar\omega_{\rm op}$) to lower voltages, which could be accounted for only by postulating an unrealistically low barrier height.

In an external field ballistic electrons can be accelerated to velocities much higher than the steady-state saturated velocity $v_{\rm sat}$ [up to the maximum band velocity $dE/d(\hbar\mathbf{k})$; for electrons in the conduction band of GaAs accelerated in a <100> direction this limit is $\simeq 10^8$ cm/sec] and such an enhancement is important and beneficial for the performance of semiconductor devices. It appears attractive (Shur, 1987) to realize ballistic transport in short-channel field effect

transistors, so that the carrier velocity at any point in the channel would be determined by the conservation of energy. Such a situation is realized in vacuum diodes, where the current is space charge limited and is described by the Child-Langmuir law. Despite considerable efforts, no semiconductor device structure has been demonstrated to-date in which the current-voltage characteristics would convincingly conform to a similar law. In practice, conditions for collisionless transport in a 2D channel (where sheet carrier concentrations typically much exceed $10^{11}$ cm$^{-2}$) are very difficult to realize because of ee collisions (which typically result in $\lambda_\beta < 1000\,\mathring{A}$ at these concentrations).

When the rate of ee collisions is very high, as is usually the case in the channel of a field-effect transistor in its *on* state, the electron ensemble behaves rather like a fluid than a gas. In the absence of *other* collisions (with phonons and impurities), the electronic fluid in the FET channel is described by hydrodynamic equations similar to those for shallow water (Dyakonov and Shur, 1993). Based on this analogy, Dyakonov and Shur (1995a,b) discussed several new effects related to plasma oscillations in the 2D electron fluid. In particular, a short-channel high-moblity transistor has a resonant response to an electromagnetic radiation at the plasma wave frequency of the 2D electrons (Dyakonov and Shur, 1996). This effect can be used to implement detectors, mixers and multipliers at terahertz frequencies. As pointed out by the authors, these devices should operate at much higher frequencies than conventional, transit-time limited devices, since the plasma waves propagate much faster than electrons. Moreover, their responsivities and conversion efficiencies can be expected to greatly exceed those of Schottky diodes currently used as detectors, mixers and multipliers in the terahertz range.

*17.2.3 Mesoscopic.* When the carrier concentration is sufficiently low and ee collisions are rare, the shape of their distribution function under external perturbation may depart considerably from the quasi-equilibrium form (3). The Maxwellian shape (Gaussian in velocities) can be viewed − in the spirit of the well-known central limit theorem of statistics −

as resulting from large number of independent scattering events, each contributing or withdrawing a random amount of energy.[4] Besides ee collisions, the electron distribution function can be maxwellized by other interactions as well, provided independent scattering events exchange random energies with electrons. Scattering by acoustic phonons has this property, while optical phonon scattering does not. If the latter were the only inelastic interaction, the equilibrium shape of the electron distribution function would be rather strange and thermodynamic properties of the electron ensemble rather different. The peculiarity of interaction with optic phonons stems from their largely monochromatic nature, which quantizes the energy exchange in units of $\hbar\omega_{op}$.

At sufficiently high temperatures, the electron energy relaxation rate due to optical phonons ($1/\tau^{(op)}$) is higher than that due to acoustic phonons ($1/\tau^{(ac)}$) by several orders of magnitude. Typically, in semiconductors $\tau^{(op)} \lesssim 10^{-12}$ s and $\tau^{(ac)} \gtrsim 10^{-9}$ s (Conwell, 1967). This disparity of the inelastic relaxation times can lead to the formation of an electronic ensemble that is in equilibrium with the optical-phonon system but has not yet appreciably interacted with acoustic phonons. Manifestation of these properties in electronic transport can be conveniently referred to as the ‘‘classical mesoscopic effects’’ (drawing a parallel to and a distinction from the quantum mesoscopic effects that occur when the coherence length or time of an electronic wave function exceeds characteristic system dimensions. Grinberg and Luryi (1990) considered the kinetics of an electron ensemble initially characterized by a Maxwellian distribution with $T_e = T_i$, subject to interaction with the lattice at equilibrium temperature $T$. Since the initial distribution is not the equilibrium Boltzmann function, it evolves in time because of the electron interactions with optical phonons, acoustic phonons, and due to the e-e scattering.

4. The form of distribution (3) does not contradict the central limit theorem. In the presense of the quantum correlation between electrons, maintained by the Pauli exclusion principle, multiple collisions lead to a Fermi rather than Maxwell distribution.

Under the assumption that $\tau^{(op)}/\tau^{(ac)} \ll 1$ and $\tau^{(op)}/\tau_{ee} \ll 1$, electrons rapidly establish a quasi equilibrium with the optical phonon field (the mesoscopic state) and then − on a longer scale − the true equilibrium is established by other inelastic scattering processes. Even though the e-e scattering does not change the average electron energy, it counts as an inelastic interaction, because it changes the shape of the non-stationary distribution.

The distribution function of the mesoscopic state can be determined from the statistical consideration alone, without actually solving the kinetic equation. The shape of this function is illustrated in Fig. 3 both for the narrow ($T_i < T$) and broad ($T_i > T$) initial distributions. Even though the electron system is in perfect equilibrium with optical phonons at temperature $T$, thermodynamic properties of electrons in the mesoscopic state are very different from those in true equilibrium, e.g., the average energy $<E> \neq (3/2)kT$ and the specific heat deviates from the classical value $3k/2$ (Grinberg and Luryi, 1990). Also the electron mobility in the mesoscopic state shows a strong overshoot. The time scale of this effect is very different from the conventional velocity overshoot; the mesoscopic state is established in less than 1 ps and persists for a long time (up to nanoseconds!) − controlled by acoustic phonon and interelectron scattering. Systematic numerical study of the classical mesoscopic kinetics has been carried out by Grinberg et al (1991).