C. R. Physique 17 (2016) 1047–1059

Contents lists available at ScienceDirect

Comptes Rendus Physique

www.sciencedirect.com

Thermoelectric mesoscopic phenomena/Phénomènes thermoélectriques mésoscopiques

Linear and nonlinear mesoscopic thermoelectric transport with coupling with heat baths

Transport thermoélectrique mésoscopique linéaire et non linéaire avec couplage à des sources de chaleur

Jian-Hua Jiang [a], Yoseph Imry [b,*]

[a] *School of Physical Science and Technology, Soochow University, 1 Shizi Street, Suzhou, Jiangsu, 215006, China*
[b] *Department of Condensed Matter Physics, Weizmann Institute of Science, Rehovot 76100, Israel*

ARTICLE INFO

Article history:
Available online 25 August 2016

Keywords:
Mesoscopic thermoelectric transport
Inelastic transport
Three terminal transport
Nonlinear effects

Mots-clés :
Mésoscopique thermoélectrique transport
Transport d'inélastique
Trois transport terminal
Effets non linéaires

ABSTRACT

Decades of research on thermoelectrics stimulated by the fact that nano- and meso-scale thermoelectric transport could yield higher energy conversion efficiency and output power has recently uncovered a new direction on inelastic thermoelectric effects. We introduce the history, motivation, and perspectives on mesoscopic inelastic thermoelectric effects.

RÉSUMÉ

Des décades de recherche sur la thermoélectricité, stimulée par le fait que le transport thermoélectrique aux échelles nanométriques et mésoscopiques pourrait améliorer le rendement de la conversion énergétique et la puissance délivrée, ont permis de découvrir récemment une nouvelle utilisation des effets thermoélectriques mésoscopiques. Nous présentons l'histoire et les motivations des recherches sur les effets thermoélectriques mésoscopiques inélastiques, ainsi que les perspectives ouvertes par ceux-ci.

1. Introduction

Research on the science and technology of the thermoelectric phenomenon has a long history since its discovery almost two centuries ago [1]. The modern theory of thermoelectric transport in bulk semiconductors was established, with the help of energy band theory and semi-classical transport theory, in the middle of the last century [1]. The key concepts such as the figure of merit and the power factor were introduced then, which facilitated and stimulated many theoretical and experimental studies. It was found that the figure of merit (a measure of the optimal thermoelectric efficiency in a material),

* Corresponding author.
E-mail addresses: joejhjiang@sina.com (J.-H. Jiang), yoseph.imry@weizmann.ac.il (Y. Imry).

http://dx.doi.org/10.1016/j.crhy.2016.08.006
1631-0705/

$$ZT = \frac{\sigma S^2 T}{\kappa} \tag{1}$$

is limited by the following interrelated transport quantities: the electrical conductivity σ, the Seebeck coefficient S, and the thermal conductivity κ. The latter has contributions from the electronic transport κ_e and other mechanisms (mainly from phonons in semiconducting materials, κ_p), i.e. $\kappa = \kappa_e + \kappa_p$. Increasing S usually leads to reduce the electrical conductivity σ. In addition, σ and κ_e are interrelated. In metals, these two quantities mostly follow the Wiedemann–Franz law

$$\kappa = LT\sigma \tag{2}$$

where $L \equiv a_L(k_B/e)^2$ is the Lorenz number ($a_L \sim 1$, depending on the material and the temperature). In semiconductors, similar relations are usually approximately confirmed. When the Wiedeman–Franz law holds, the thermoelectric figure of merit is approximately S^2e^2/k_B^2, which is often smaller than 1.

Mahan and Sofo proposed to improve the figure of merit by using conductors with very narrow energy bands, so that the thermal conductivity κ_e is reduced from the Wiedemann–Franz law [2]. However, this was found lately to be less effective as, for example, the output power is suppressed in the zero band width limit [3] due to, e.g., suppressed electrical conductivity [4].

Alongside with those developments, the scientific community was pursuing a better understanding of (charge) transport in nano- and meso-scale systems, localization phenomena, and strongly interacting electron systems. These studies activated research on thermoelectric transport in non-standard (including bulk) semiconductors, which is still ongoing [5]. The most influential studies are based on two seminal works by Hicks and Dresselhaus [6], who found that in reduced-dimension semiconductors, such as quantum wells (2D), quantum wires (1D) and quantum dots (0D) heterostructures, the interrelation between the electrical conductivity, the Seebeck coefficient, and the thermal conductivity can be partially decoupled. This is mainly because the density of states can be effectively enhanced and modulated (at different energies) when the spatial confinement along any direction is small enough to induce significant quantum-confinement effects. Opportunities for enhancing the thermoelectric figure of merit ZT and the power factor $S^2\sigma$ by engineering nanostructures and nanomaterials thus emerge [7–11]. In addition, when low-dimensional structures are packed up (densely) into macroscopic structures, the abundant interfaces effectively reduce the phonon heat conductivity [7,8]. If the phonon heat conductivity is reduced more significantly (e.g., by introducing effective phonon scattering centers, or packing materials with mismatched phonon impedance), then the electrical conductivity, the thermoelectric figure of merit can be improved [9]. This was demonstrated in BiTe quantum well superlattices and PbTe quantum dot superlattices [8]. In the latter, a high density of quantum dots forming a regular array induces a large electronic density of states in the PbSeTe alloy layer. Other methods, such as energy filtering and semimetal–semiconductor transition, in engineering the electronic density of states and the energy dependent conductivity, are also introduced [10,11]. The underlying physics is manifested in the Mott–Cutler formula [12,13]

$$S = \frac{\langle E-\mu\rangle}{eT}, \quad \kappa = \sigma\frac{\mathrm{Var}(E-\mu)}{e^2T} \tag{3}$$

where $\mathrm{Var}(E-\mu) = \langle (E-\mu)^2\rangle - \langle (E-\mu)\rangle^2$ is the variance of the transport electronic energy; the averages of powers of the electronic energy are defined as

$$\langle (E-\mu)^n\rangle = \frac{1}{\sigma}\int \mathrm{d}E\sigma(E)(E-\mu)^n[-\partial_E f_F(E)], \quad n = 1, 2 \tag{4a}$$

$$\sigma = \int \mathrm{d}E\sigma(E)[-\partial_E f_F(E)] \tag{4b}$$

Here $\sigma(E)$ is the energy-dependent conductivity and $f_F(E) = 1/[\exp(\frac{E-\mu}{k_BT}) + 1]$ is the Fermi distribution function. The energy dependence of the conductivity $\sigma(E)$ is the key quantity for engineering the thermoelectric transport properties. It can be written as $\sigma(E) = ek_BT\rho(E)b(E)$, where $\rho(E)$ is the density of states (DOS) and $b(E)$ is the energy-dependent mobility.

Nanostructures and nanocomposites have been demonstrated as effective ways to tune the carrier density of states and the energy dependence of carrier scattering, as well as to reduce the phonon thermal conductivity. These methods lead to high thermoelectric efficiency and power density for applications. Nevertheless, several nontrivial aspects of mesoscopic electron transport have not been exploited, which might give a chance to a further enhancement of thermoelectric efficiency and power. These are: (1) nonlinear effects, (2) inelastic effects (due to strong carrier–carrier interaction and carrier–phonon interaction), (3) thermoelectric transport with spatially separated electrical and thermal currents. The main purpose of this review is to emphasize those aspects and their potential values for improving thermoelectric performance, as well as the realization of thermoelectric diodes and transistors. We mainly focus on the inelastic effects and show that they can offer an alternative (new) route toward high-performance thermoelectric structures that may reduce the need for novel functional materials. We further discuss nonlinear effects and the spatial separation of heat and electrical currents in thermoelectric transport through inelastic thermoelectric transport. We will illustrate how these aspects may lead to better thermoelectrics. Several example systems are used to demonstrate the principles.

In the next section, we consider bounds (which, when reached, are sometimes referred to as "quantized values") of thermoelectric response coefficients. These bounds determine sometimes [14] the scales over which the maximal efficiency is reduced from Carnot. In section 3, we consider linear and nonlinear transport above a barrier. The latter is crossed by thermal activation, the energy for which is taken from the (usually phonons, or the other electrons) associated heat baths. In section 4, we discuss a simple, quite generic, model [15,16] for inelastic transport using a heat bath. Synergistic heat and electrical rectification and transistor effects in systems with pronounced inelastic thermoelectric transport [17] is reviewed in section 5. We conclude and summarize in section 6.

2. Bounds and quantized values of thermal-electric transport coefficients (This section is dedicated to Jacob Bekenstein, a highly esteemed colleague, some of whose early work is reviewed here)

In this section, we discuss bounds (upper limits) on electric, thermal and thermoelectric transport coefficients. For the first two, these bounds can be reached in ideal conducting channels and give rise to "quantized" values of these coefficients.

It makes physical sense that maximal values of currents carried in a pipe are proportional to the cross sectional area of that pipe. The same is valid for the transport of charge, energy, and heat through a conducting wire. This is of interest by itself and because, as found by Whitney [14], these bounds also give the scales on which the maximal efficiency is reduced by Carnot's theorem.

We start with the simplest case of charge transport, which is not the first example to have been discussed [18], but the easiest to understand [19]. Consider a 1D wire carrying electrical current, at a temperature $T \ll E_F$ with E_F being the Fermi energy. We imagine, following Landauer [20], that the strictly 1D wire (so narrow that the lowest transverse excitation energy is larger than E_F plus several $k_B T$, so only one transverse state is relevant) connecting two reservoirs each at equilibrium, but with slightly different electrochemical potentials μ_L and μ_R and, at this stage, identical temperatures, T. The current flowing between L and R is:

$$I = e \int \mathrm{d}E (f_L - f_R) v(E) \rho(E) \tag{5}$$

$v(E)$ and $\rho(E)$ being the velocity and the density of states at energy E. f_L and f_R are the Fermi functions of L and R. Since the 1D single-particle DOS, spin included, is given by

$$\rho(E) = \frac{1}{\pi \hbar v(E)} \tag{6}$$

the velocity factors cancel. For the *linear response* $f_L - f_R = -eVf'$ with $f' \equiv \partial_E f$ (f is the equilibrium Fermi distribution), the integral trivially gives $I = \frac{e^2}{\pi\hbar} V$. Therefore, the conductance is given by

$$G = I/V = \frac{e^2}{\pi \hbar} \tag{7}$$

This inconspicuous result is amazingly deep! The conductance of an *ideal* 1D ("single-channel") wire is universal and totally independent on the material and the properties of the wire! No single-channel wire can have a larger conductance than an ideal one. So, this value is both an upper limit and a so-called "quantized value" of this conductance. The upper limit is realized in the ideal conductor (no scattering) limit. We shall soon see that a similar phenomenon occurs for thermal conductance, but before that we discuss the effect of a finite-thickness wire, comprising several channels.

If the wire has n transverse states below the Fermi energy, there will be a continuum of longitudinal (parallel to the wire) states for each, which will create a $\frac{e^2}{\pi\hbar}$ conductance. Thus, the total low-temperature conductance will be $n\frac{e^2}{\pi\hbar}$. Similarly, if the temperature is higher than the separation, Δ, to a higher transverse state, its channel will be populated and it too will contribute another conductance quantum. Since $\Delta \sim \hbar^2/ma^2$, where a is the relevant dimension of the wire, the condition $T \gg \Delta$ (still, with $T \ll E_F$) is equivalent to $a \gg \lambda_T$. (Similarly, $\Delta \gg E_F$ can also by described by $a \ll k_F^{-1}$.) Thus, the number of channels, n, in a wire is on the order of the number of elementary k_F^{-2} or λ_T^2 areas at zero or higher temperatures, respectively. At the higher temperatures, the measured conductance, related to the number of channels, is a quantum property, dependent on $\hbar$! This has been mentioned by Whitney [14] in the context of thermal conductivity, which is the subject we shall discuss shortly.

The quantized conductance has been observed experimentally, concurrently by two groups [21], in suitable ballistic quantum point contacts in GaAs, within a percent accuracy, about two years after the prediction. More recently, it became a standard item in mesoscopic Physics.

Now we discuss the electronic thermal conductance; to that end we assume that the two Landauer reservoirs have the same electrochemical potentials and different temperatures T_L and T_R (in linear response, the effects of $T_L - T_R$ and $\mu_L - \mu_R$ are simply additive). We want [22] the heat current between L and R. Thermodynamics tells that an electron of energy transferring from L to R carries heat given by [23,24] $E - \mu = TS$. Thus, the heat current, $\int \mathrm{d}E (f_L - f_R)(E-\mu) v(E)\rho(E)$ between L and R, is given by

$$J_Q = (T_L - T_R) \int \mathrm{d}E (-f') \frac{(E-\mu)^2}{T} v(E)\rho(E) \tag{8}$$

where we used $f_L - f_R = (-f')\frac{(E-\mu)}{T}(T_L - T_R)$. We now use the Sommerfeld expansion (see, e.g., [23]) $\int_0^\infty dE(E-\mu)^2(-f') = (\pi^2/3)(k_B T)^2$, to find

$$K = (\pi/3\hbar)k_B^2 T \tag{9}$$

This is the upper bound on the electronic thermal conductance per channel, or the "quantum of thermal conductance" (for two spin directions). Not surprisingly (both based on the Sommerfeld expansion), it satisfies the Wiedemann–Franz relationship, $K = (\pi^2/3)(k_B/e)^2 TG$, with the conductance quantum $\frac{e^2}{\pi\hbar}$.

Somewhat more surprising is the situation with the phonon thermal conductivity. The two differences are that here heat and energy are identical ($\mu = 0$) and that the Bose function should be used, instead of the Fermi one. As noticed already by Bekenstein [18], the result is nevertheless the same. The conductance quantum per phonon mode is $K = (\pi/6\hbar)k_B^2 T$. This was first observed experimentally by Schwab et al. [25].

Finally, we turn to the upper bound on the thermoelectric coefficient of 1D systems. A systematic way to get that would be to write the thermopower in terms of the energy-dependent conductance $G(E)$ and vary the latter to maximize the former. However, there are physical limitations on the behavior of $G(E)$, which are not obvious and need more work. A simpler approach is to appeal to the requirement of stability, i.e., positive definiteness of the thermoelectric transport (Onsager) matrix, $\hat{\mathcal{L}}$ (i.e., $\mathcal{L}_{12}^2 \le L_{11}L_{22}$), using the above bounds on the diagonal elements. This gives,

$$\mathcal{L}_{12} \le \sqrt{\mathcal{L}_{11}\mathcal{L}_{22}} = \frac{e}{\sqrt{3\hbar}}k_B T \tag{10}$$

as the upper bound for the coefficient $\mathcal{L}_{12}$ for spin-$\frac{1}{2}$ 1D electron systems. As for the thermopower, it is given by $\mathcal{L}_{12}/(T\mathcal{L}_{11})$ and it can in principle be as large as we please, for a small enough electrical conductance $\mathcal{L}_{11}$. As Eq. (3) shows, however, when a large enough energy is transferred by each electron, S is large. But then the output power is (usually, exponentially) small as well (due to a small electrical conductance).

3. Activated transport above a barrier

3.1. Linear transport

Here we consider a very simple system where the electronic transport is effectively funneled to a narrow band. We imagine an ordinary barrier in one dimension (1D, generalized later), depicted in Fig. 1. The barrier is chosen so that its height W (measured from the averaged chemical potential μ) and its thickness, d, satisfy

$$W \gg k_B T; \quad d \gg \hbar/\sqrt{2mW}, 1/k_F \tag{11}$$

The second inequality is more strongly obeyed than the first one, so that the dominant transport is via thermal activation above the barrier and not by quantum-mechanical tunneling. We assume that the barrier is tapered (see Fig. 1) so that the transmission through it changes rather quickly from 0 to 1 when the electron energy E increases through W. This is certainly the case in the 1D clean tunnel junction of the type discussed here, or in the quantum point contact [19,26]. Its validity in a high-dimensional system will be confirmed later on. When some disorder exists, rendering the electron motion diffusive, it makes sense that the transmission still changes from 0 to 1 when the electron energy increases through W. This increase may become slower than in clean systems, but that should not change the qualitative behavior.

A significant feature of our setup is that the electronic heat conductance can be made to be very small while the thermopower stays finite

$$K_e = K_e^0 - GS^2T^2 \ll K_e^0 \tag{12}$$

where G is the conductance, S is the thermopower, K_e is the electronic thermal conductance (defined at vanishing electrical current), while K_e^0 can be termed a "bare" thermal conductance (defined at vanishing electrochemical potential difference). As mentioned above, according to Ref. [2], the largest two-terminal figure of merit is achieved in systems with the smallest $K_e/(GS^2T^2)$. Here this ratio is very small, and then ZT is mainly limited by the phonon heat conductivity K_p between the two metallic contacts [see Eq. (17) below]. K_p can be small in nanosystems [7,8]. Our system is then expected to possess a high figure of merit. Another way to understand the situation (shown by the formulae below) is that S is the average energy transferred by an electron, divided by eT, while K_e^0/G is the average of the square of that energy, divided by e^2. Therefore K_e/G is proportional to the variance of that energy. The latter obviously vanishes for a very narrow transmission band. In this case, the transmission band is the range of a few $k_B T$ above W. Thus, it is not surprising that K_e is of the order of $(k_B T)^2$.

The thermoelectric linear transport problem is fully characterized, for this two-terminal situation, by:

$$\begin{pmatrix} I_e \\ I_Q^e \end{pmatrix} = \begin{pmatrix} G & L_1 \\ L_1' & K_e^0 \end{pmatrix} \begin{pmatrix} \delta\mu/e \\ \delta T/T \end{pmatrix} \tag{13}$$

where I_e is the charge current and I_Q^e is the heat current, $\delta T = T_L - T_R$, and $\delta\mu/e \equiv (\mu_L - \mu_R)/e \equiv V$ is the voltage between the left and right terminals. The 2×2 matrix contains the regular conductance G, the bare electronic thermal

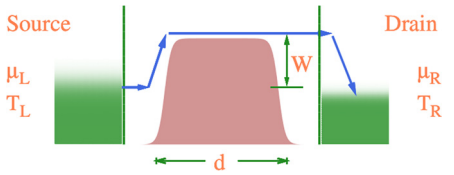

Fig. 1. The suggested device: a long ($d \gg$ characteristic tunneling length) and high ($W \gg k_B T$) barrier separating two electron gases. The transferred electron gets an energy $W \pm \mathcal{O}(k_B T)$ from the LHS thermal bath and deposits it in the RHS one.

conductance K_e^0, and the (off-diagonal) thermoelectric coefficients L_1 and L_1'. That the latter two are equal is the celebrated Onsager relation (valid for time-reversal symmetric systems). We remind the reader that $S = L_1/(TG)$. All currents and transport coefficients in Eq. (13) are given in 1D in terms of the energy-dependent transmission of the barrier, which we take as $\mathcal{T}(E) \simeq \Theta(E - W)$. We measure all energies from the common chemical potential μ (i.e. $\mu \equiv 0$). The currents are

$$I_e = \frac{2e}{h}\int_0^\infty dE\mathcal{T}(E)[f_L(E) - f_R(E)]$$
$$I_Q^e = \frac{2}{h}\int_0^\infty dE E\mathcal{T}(E)[f_L(E) - f_R(E)] \tag{14}$$

and hence

$$G \simeq \frac{2e^2}{h}\int_W^\infty dE(k_B T)^{-1} f(E)[1 - f(E)]$$
$$L_1 \simeq \frac{2e}{h}\int_W^\infty E dE(k_B T)^{-1} f(E)[1 - f(E)]$$
$$K_e^0 \simeq \frac{2}{h}\int_W^\infty E^2 dE(k_B T)^{-1} f(E)[1 - f(E)] \tag{15}$$

with $f(E) = 1/[\exp(\frac{E}{k_B T}) + 1]$ being the equilibrium Fermi distribution. It reduces for $W \gg k_B T$ to the Boltzmann distribution and then $f(E)[1 - f(E)] \propto \exp(-\frac{E}{k_B T})$. Therefore,

$$\langle E\rangle = k_B T(x_w + 1), \quad \langle E^2\rangle - \langle E\rangle^2 = (k_B T)^2 \tag{16}$$

with $x_w \equiv W/(k_B T)$.

From the transport coefficients, one readily obtains the electronic figure of merit:

$$ZT = (ZT)_e \frac{K_e}{K_e + K_p} \tag{17}$$
$$(ZT)_e = \frac{\langle E\rangle^2}{\langle E^2\rangle - \langle E\rangle^2} = (x_w + 1)^2 = \left(\frac{W}{k_B T} + 1\right)^2 \tag{18}$$

In two and three dimensions ($d = 2, 3$), the calculation proceeds similarly. The energy is the sum of the longitudinal part which goes over the barrier and the transverse part which should be integrated upon. The latter has the usual density of states in $d-1$ dimensions. This is $E^{-1/2}$ for $d = 2$ and constant for $d = 3$. The overall factors do not matter for ZT. The final result differs from Eq. (18) just by numerical factors. At $d = 2$, the result is:

$$(ZT)_e = \frac{2}{3}\left(\frac{W}{k_B T} + \frac{3}{2}\right)^2 \tag{19}$$

The result for $d = 3$ is

$$(ZT)_e = \frac{1}{2}\left(\frac{W}{k_B T} + 2\right)^2 \tag{20}$$

The large order of magnitude of ZT remains. The advantage of the last case, $d = 3$, is twofold: it is easier to make (two thick layers of the conducting material, separated by an appropriate barrier) and the total current for a given W, and hence the power, is proportional to the cross-section of the device.

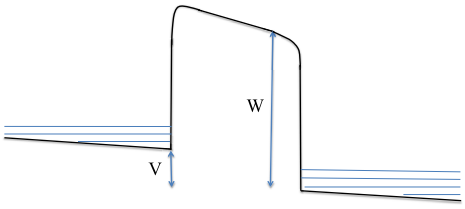

Fig. 2. Schematics of a barrier (W) with a large bias (V).

A vacuum junction has too high a barrier for most applications [27]. The ballistic quantum point contact is a very effective realization of the model, when biased in the pinch-off regime and in the region where activated conduction is dominant. It requires, however, rather high technology and can handle only small powers. We believe that the two more realistic straightforward ways to effectively achieve the requirements of the model are:

1. a metal–semiconductor–metal junction, with a properly chosen difference between work functions and electron affinities (an example might be Au–p-Si–Au). A large area will increase the power of the device;
2. a superlattice [28] separating the two metallic electrodes, where the Fermi level is inside the gap between the "valence" and a "conduction" mini-bands of the superlattice. Electron–hole symmetry should be strongly broken either by intrinsic lack of band symmetry or by the placement of E_F away from the middle of the gap. Obviously, highly-doped semi-conductors can be substituted for the metallic electrodes.

The last remark brings us to the issue of semiconducting systems. Imagine first an intrinsic small-gap semiconductor, such as BiTe, the current work-horse of applied thermoelectricity. The first thought would be that since conductivity is thermally activated, with effectively a Boltzmann distribution at temperatures much below the gap; it is a simple realization of our model. This would be the case only if electron–hole symmetry will be strongly broken. If that symmetry prevails, the electron and hole thermopowers will cancel. That symmetry breaking can be achieved by judicious selective doping and alloying, and seems to dominate the art of the present manipulations of BiTe and its derivatives. However, what is proposed here is a different approach: let a larger band-gap semiconductor, even Si, Ge, graphite, or a member of the GaAs family, bridge two layers of metal so chosen that their Fermi level is significantly closer to either the conduction or the valence band. In a large temperature regime, this will, as discussed in the above, realize our activated junction model. We add, finally, that a yet another way to realize the model is via such a semiconductor, e.g., n-doped so that the Fermi level is, say, $(3 \sim 10)k_BT$ below the conduction band and much further from the valence band. The conduction band should then play the role of our activation barrier. Some experimentation should show which of these schemes is superior.

3.2. Nonlinear activated transport above a barrier

Here we show that a relatively high barrier, W, in the nonlinear regime, $W \gtrsim V \gg k_BT$, is a rather efficient thermoelectric device. Neglecting return currents and phonon parasitic thermal conductance, the efficiency is found to reach W/V, where W is the barrier the electron has to cross (setting $e \equiv 1$ in this section). This is rather high, especially for such a simple device. The cooling power is IW, I being the current. Ways to reduce the effect of the phonon thermal conductance are suggested.

We consider a wide (to neglect tunneling) solid-state or vacuum, electronic barrier, of height W, bridging two conductors (see Fig. 2). It is idealized, and taken as biased by a voltage V, where $W \gtrsim V \gg k_BT$, *such that all electronic conduction is from the left to the right terminal*. We have in mind $W \sim eV$. Now, even if we keep the LHS (RHS) at temperatures T_c (T_h), where $T_c < T_h$, the device may transfer heat between the cool left terminal and the warmer right one; i.e. we have a (thermoelectric) refrigerator. That heat per transported electron is given in the ideal case roughly by $W - V \sim W \gtrsim V$ [27]. This is because an electron has to borrow an energy of about W to cross the barrier from the LHS heat bath, it then returns it to the RHS heat bath. Thus, the cooling power is IW and the invested power is IV, where I is the current. Therefore, the efficiency and power are

$$\eta = W/V, \quad P = IW \tag{21}$$

where we neglected the "parasitic" phonon thermal conduction back from right to left and took the energy price per transferred electron with the bias as V. We remark that this is not a linear response (small V and $\delta T \equiv T_h - T_c$) calculation. It is valid for finite driving forces, as long as V is the self-consistent voltage.

By essentially reversing the process (i.e. employing temperature differences to generate electrical energy), one may use this device for energy harvesting as well.

Here we give a proof, within the linear transport approximation, that the W/V efficiency does not exceed the Carnot one. Write the heat current as $I_Q^e = -K\delta T/T + LV$, where $L = SGT$ with S being the Seebeck coefficient, K and G are effective thermal and electrical conductances. The "working condition" is that to refrigerate, $V > (K/L)\delta T/T \Rightarrow V/W > K/(SGWT)\delta T/T = K/(S^2GT^2)\delta T/T$. (Remembering that ST is the average heat transferred by an electron, $S = W/T$.) Now

we use the stability of the Onsager matrix: $K/G > (ST)^2$, to get $W/V < T/\delta T = \eta_c$. η_c is the Carnot efficiency ("figure of merit") for a refrigerator. Thus η_c is indeed the upper limit on the efficiency.

The efficiency obtained is quite high. Even pushing the voltage up to, say, $W/2$, we get 2 in the ideal case. If the idealizing factor of negligible phonon heat conductance decreases the efficiency by a factor of two, a value of 1 is still respectable for such a simple device. It is comparable to the values for current thermoelectric devices having the parameter $ZT \cong 1$. Of course, we have the usual conundrum between power and efficiency.

As far as the negative effect of the phonon thermal conductance, besides decreasing its relative influence using a larger V (decreasing the phonons' relative contribution, see below), it is suggested to use a soft metal on one side of the barrier and a hard semiconductor or metal on the other side, to reduce the phonon thermal conductance via acoustic impedance mismatch. We also reemphasize that the phonon conductance is largely independent of the applied voltage, so that *larger values of the latter are favorable also in that respect*.

4. The inelastic thermoelectric transport assisted by a heat bath: linear transport

In true inelastic transport processes, energy has to be absorbed/emitted from heat baths to compensate for the energy difference between the initial and final states. This energy can come in the form of phonons, plasmons, charge (spin) fluctuations or other collective excitations. In bulk materials, those collective excitations usually thermalize quickly with electrons. Therefore, the local temperature of those collective excitations is nearly the same as that of the electronic ensemble, unless the system is out of the linear-response regime. Interaction with phonons and other collective excitations may modify thermoelectric transport coefficients via, e.g., phonon drag effect, which is here an unessential quantitative correction.

In mesoscopic regimes, the abundance of interfaces and disorder may significantly reduce the thermalization processes. Substantial temperature or electrochemical potential differences can be maintained between adjacent nanoscale regions. Thus inelastic and nonlinear transports become significant and prevail. Moreover, geometric configurations that support inelastic transport are richer. It turns out that the latter opens an important direction for improving thermoelectric performance that has not been fully explored so far. The fact that mesoscopic thermoelectric transport can have significant spatial separation of electrical and thermal currents, particularly in inelastic transport processes, enables the disentanglement of electrical and thermal conductivity and provides new opportunities for improving the thermoelectric figure of merit.

The simplest nontrivial geometry configuration for inelastic transport processes is the three-terminal set-up where energy absorbed/emitted by the electron comes from a third heat bath differing from the source and drain [see Fig. 3]. Inelastic thermoelectric transport in three-terminal geometry was first examined by Entin-Wohlman, Imry, and Aharony [29] for molecular junctions, and later by Sánchez and Büttiker [30] for capacitively coupled double quantum dots. For the former, the additional energy is provided by a phononic heat bath; for the latter, the additional energy is provided by a capacitively coupled electronic reservoir. Entin-Wohlman, Imry, and Aharony [29] pointed out, based on thermodynamic arguments, that the full description of thermoelectric transport in such three-terminal system must include two heat currents (one is the electronic heat current from source to drain and the other is the heat current from the third terminal), coupled with a single electrical current. Such a property of three-terminal mesoscopic system immediately leads to two *correlated* thermopowers (i.e. the electrical current can be induced by two temperature differences that contribute additively in the linear-response regime). The merit of such an effect has not been fully appreciated until recently in Refs. [16,35]. Another important property is that the direction of heat current from the third terminal is different from the direction of electrical current. This means spatial separation of electrical and thermal currents [15,29–32], which allows independent tuning of electrical and thermal conduction, and achieve high thermoelectric figure of merit and power factor [16,35].

As in many situations, the above findings share some similarities with some earlier works. In 1993, Edwards et al. [33] proposed to use inelastic transport to cool an island of electrons, which can be useful for sub-Kelvin cooling of the electron gas following conventional cooling methods [34]. Nevertheless, the physics uncovered in recent studies of the mesoscopic inelastic thermoelectric effect goes far beyond a novel method of cooling [15,17,35–40].

The remaining of this paper will present a pedagogical review of the essential aspects using a simple picture which is depicted in Fig. 3. In this case, electrons are transported, under certain bias, from source to drain. The dominant path for such transport is jumping elastically to a quantum dot (or a localized energy level) with energy E_1 and then hopping inelastically to another quantum dot with energy E_2, and finally tunneling into the drain. Because the energies E_1 and E_2 do not match, the electrons have to exchange energy with a heat bath. This heat bath can have a temperature T_p differing from the temperatures of the source (T_L) and drain (T_R).

In this picture, there are three reservoirs: the source, the drain, and the heat bath. Their states are characterized by their temperatures and chemical potentials. The heat bath can be electronic or phononic (or consisting of another type of bosons that carry energy). Without loss of the essential physics, we focus on the situation that there is no electrical charge flowing out of the heat bath. Under this assumption we can treat the heat bath as bosonic. In this regime, the relevant thermodynamic variable of the heat bath is just its temperature T_p. The elementary excitation that provides energy to electron transport can be phonons, charge fluctuations (i.e. charge density waves or plasmons) [36,41–43], spin fluctuations (i.e. spin density waves or magnons [44]), photons [45], etc. This energy is sometimes termed as energy gain in the literature [30].

The energy and charge flows into the reservoirs are defined as the time derivatives of their total energy and charge of the reservoirs, $\dot{E}_i$ $(i = L, R, p)$ and $e\dot{N}_i$ $(i = L, R)$ where $L/R/p$ stand for source/drain/heat bath, respectively. N_i is the

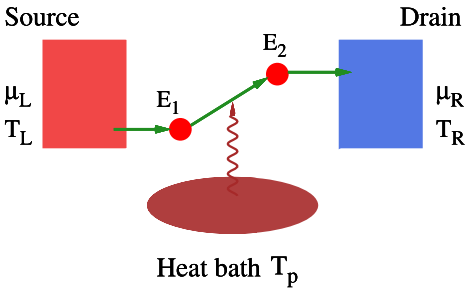

Fig. 3. Schematic of three-terminal inelastic thermoelectric mesoscopic systems. Electron transport from source to drain has to absorb energy from the heat bath. There are three reservoirs: source, drain, and the heat bath. They can have different temperatures.

total number of electrons in the i reservoir and $e<0$ is the charge of an electron. Total energy and charge conservation gives $\sum_i \dot{E}_i = 0$ and $\sum_i \dot{N}_i = 0$. Therefore, there are only three independent currents, which can be organized into two heat currents and one electrical current,

$$I_{\rm Q}^{e} \equiv \frac{1}{2}(\dot{Q}_{\rm R} - \dot{Q}_{\rm L}), \quad I_{\rm Q}^{p} \equiv -\dot{Q}_{\rm p}, \quad I_{\rm e} \equiv e\dot{N}_R \tag{22}$$

where $\dot{Q}_i = \dot{E}_i - \mu_i \dot{N}_i$ for $i = L, R$ and $\dot{Q}_{\rm p} = \dot{E}_{\rm p}$.

The thermodynamic affinities conjugated to those three currents satisfy the following relation [46],

$$\dot{S}_{tot} = I_{\rm e} A_1 + I_{\rm Q}^{e} A_2 + I_{\rm Q}^{p} A_3 \tag{23}$$

We found that [17] these conjugated affinities are

$$A_1 = \frac{\mu_{\rm L} - \mu_{\rm R}}{e}\left(\frac{1}{2T_{\rm L}} + \frac{1}{2T_{\rm R}}\right), \quad A_2 = \frac{1}{T_{\rm R}} - \frac{1}{T_{\rm L}}$$
$$A_3 = \frac{1}{2T_{\rm L}} + \frac{1}{2T_{\rm R}} - \frac{1}{T_{\rm p}} \tag{24}$$

The phenomenological Onsager transport equation is then

$$I_i = \sum_j M_{ij} A_j + \mathcal{O}(A^2) \tag{25}$$

where $\hat{M}$ is the 3×3 thermoelectric response matrix. In the linear response regime,

$$A_1 \simeq \frac{\mu_{\rm L} - \mu_{\rm R}}{eT}, \quad A_2 \simeq \frac{\delta T}{T^2}, \quad \delta T \equiv T_{\rm L} - T_{\rm R}$$
$$A_3 \simeq \frac{\Delta T}{T^2}, \quad \Delta T \equiv T_{\rm p} - \frac{1}{2}(T_{\rm L} + T_{\rm R}) \tag{26}$$

where T is the equilibrium temperature. And

$$\hat{M} = T\begin{pmatrix} G & L_1 & L_2 \\ L_1 & K_1 & K_{12} \\ L_2 & K_{12} & K_2 \end{pmatrix} \tag{27}$$

where G is the conductance, K_1, K_2, and K_{12} are the diagonal and off-diagonal heat conductances. The two thermopowers are

$$S = \frac{L_1}{TG}, \quad S_{\rm p} = \frac{L_2}{TG} \tag{28}$$

Particularly, $S_{\rm p}$ enables the possibility of cooling of the heat bath by an electrical current [39], or the cooling of the drain by a hot heat bath [40].

For the double quantum dots system, when inelastic transport dominates, the transport coefficients in Eq. (27) can be written as

$$L_1 = G\frac{\overline{E}}{e}, \quad L_2 = G\frac{\boldsymbol{\omega}}{e}, \quad K_1 = G\frac{\overline{E}^2}{e^2}$$
$$K_{12} = G\frac{\overline{E}\boldsymbol{\omega}}{e^2}, \quad K_2 = G\frac{\boldsymbol{\omega}^2}{e^2}$$
$$\overline{E} \equiv \frac{E_1 + E_2}{2}, \quad \boldsymbol{\omega} = E_2 - E_1 \tag{29}$$

The conductance is $G = \frac{e^2}{k_B T}\Gamma_{1\to 2}$, where $\Gamma_{1\to 2}$ is the inelastic transition rate between the two quantum dots [15,17]. We have assumed here that the coupling between the QD 1 and the source as well as that between the QD 2 the drain is much stronger than the coupling between the two QDs [15,17].

The conventional thermoelectric effect is described by Eq. (25) for $A_3 = 0$, i.e. the temperature of the heat bath is equal to the average temperature of the source and drain [more precisely, when $T_p = 2T_L T_R/(T_L + T_R)$]. Thus, S is the conventional thermopower. In contrast, S_p is closely related to the inelastic thermoelectric transport: the temperature of the heat bath comes into play only for the inelastic thermoelectric effect. Moreover, the direction of the heat flow out of the heat bath, I_Q^p, can be different from that of the heat flow I_Q^e (we shall term this as "electronic heat current").

In the conventional thermoelectric effect, the electrical current flows parallel to the heat current, which leads to a strong correlation between electrical conductivity and thermal conductivity. This has been found to be a key obstacle that limits the figure of merit [4]. More explicitly, the conventional thermoelectric figure of merit here is given by

$$ZT = \frac{GS^2T}{K_1/T - GS^2T} \tag{30}$$

For a slab of material with area A and thickness l, the electrical conductivity is $\sigma = Gl/A$ and the thermal conductivity is $\kappa = (K_1/T - GS^2T)l/A$. Hence the above equation recovers the well-known Ioffe's figure of merit $ZT = \sigma S^2 T/\kappa$.

The correlation between different transport coefficients can be revealed by considering many parallel transport channels. For the present model, these are many double quantum dots systems that conduct electricity and heat between the source and the drain. That is,

$$\begin{aligned}
&G = \int \rho(E_i)\,\mathrm{d}E_i \int \rho(E_f)\,\mathrm{d}E_f\, G(E_i, E_f)\\
&S = \frac{L_1}{TG} = \frac{\langle \overline{E} - \mu\rangle}{eT}, \quad S_p = \frac{L_2}{TG} = \frac{\langle \boldsymbol{\omega}\rangle}{eT}\\
&K_1 = G\frac{\mathrm{Var}(\overline{E} - \mu)}{e^2} + K_1', \quad K_2 = G\frac{\mathrm{Var}(\boldsymbol{\omega})}{e^2} + K_2'\\
&K_{12} = G\frac{\langle(\overline{E} - \mu)\boldsymbol{\omega}\rangle}{e^2} + K_{12}'
\end{aligned} \tag{31}$$

where E_i is the energy of the electron when it leaves the source, E_f is its energy when it enters the drain,

$$\overline{E} = \frac{1}{2}(E_i + E_f), \quad \boldsymbol{\omega} = E_f - E_i \tag{32}$$

$\rho(E)$ is the density of states, $G(E_i, E_f)$ is the (initial and final) energy-dependent conductivity, and the average is defined as

$$\langle \mathcal{O}\rangle = G^{-1}\int \rho(E_i)\,\mathrm{d}E_i \int \rho(E_f)\,\mathrm{d}E_f\, G(E_i, E_f)\mathcal{O} \tag{33}$$

The heat conductivities from non-electronic processes are also included in the above. They are denoted by K_1', K_2', and K_{12}', for diagonal and off-diagonal heat conduction.

We can then express the figure of merit as

$$ZT = \frac{\langle \overline{E} - \mu\rangle^2}{\mathrm{Var}(\overline{E} - \mu) + \Delta_1}, \quad \Delta_1 = \frac{e^2 K_1'}{G} \tag{34}$$

The above equation, although derived from our model, also applies to normal thermoelectric materials where K_1' comes from the parasitic phonon thermal conductivity. Mahan and Sofo have argued that the figure of merit can be significantly improved in materials with very small electronic bandwidth W. In that limit, $\mathrm{Var}(\overline{E} - \mu) \sim W^2 \ll (k_B T)^2$, the figure of merit is expected to be much larger than unity if $\langle \overline{E} - \mu\rangle^2 \sim (k_B T)^2$. However, in the $W \to 0$ limit, the output power vanishes [3]. In a recent study [4], it is argued that for narrow electronic bandwidth materials, electron mobility is suppressed by the large effective mass and strong back scattering. The suppressed electrical conductivity G then makes the factor Δ_1 dominate the denominator. In such a regime, the figure of merit is suppressed, rather than promoted. Realistic optimization of the figure of merit needs to balance the variance and the factor Δ_1 [4]. The correlation between electrical and thermal conductivity limits further improvement of the figure of merit [4].

In our system, besides the conventional thermoelectric effect, there is an additional thermopower due to inelastic transport, S_p. Using this effect, waste heat from the heat bath can be converted into useful electrical power (reversely, electrical power can be used to cool the heat bath without passing electricity to it). The figure of merit of such thermoelectric energy conversion was first derived by Jiang, Entin-Wohlman, and Imry [15,35]:

$$\tilde{Z}T = \frac{GS_p^2T}{K_2/T - GS_p^2T} = \frac{\langle \boldsymbol{\omega}\rangle^2}{\mathrm{Var}(\boldsymbol{\omega}) + \Delta_2}, \quad \Delta_2 = \frac{e^2 K_2'}{G} \tag{35}$$

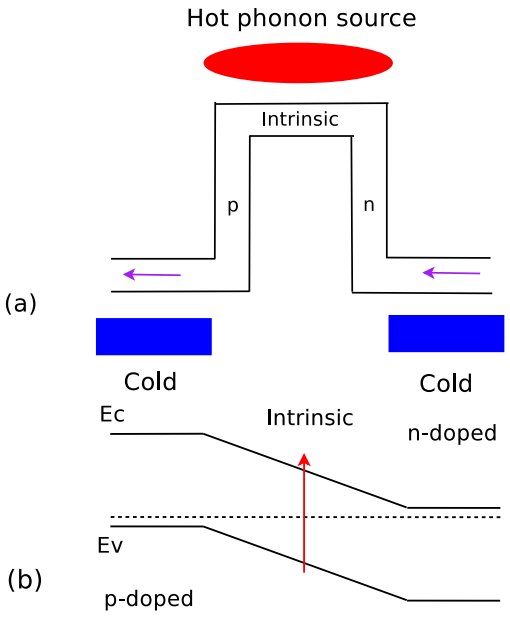

Fig. 4. (a) Structure of a p–n junction three-terminal thermoelectric engine. (b) Energy diagram: absorbing a hot phonon generates an electron–hole pair, which drifts along the built-in electric field of the junction to induce a current, similar to the case of photo-voltaic effects. From Jiang, Entin-Wohlman, and Imry [35].

The advantages of the inelastic thermoelectric effect are manifested in the above equation: (1) high figure of merit demands small variance of $\boldsymbol{\omega}$ (i.e. the energy of collective excitation) in the heat bath, which can be disentangled from electronic band structures! (2) Hence the conductivity G can be uncorrelated with the variance of $\boldsymbol{\omega}$ (this also opens the opportunity for high output power if G is large). (3) Sharp DOS peaks in the spectrum of collective excitations can reduce thermal conductivity due to such excitations (i.e. reduce the denominator in (35)). This also promotes the inelastic scattering rates and hence the electrical conductivity and the output power. (4) Spatial separation of electrical and heat currents enables engineering of electrical and heat transport more independently. For the inelastic transport to give considerable output power and to dominate over the elastic transport, the coupling to bosons has to be very strong and the elastic transport has to be suppressed by an energy barrier.

The advantages of the inelastic thermoelectric effect are explored in Ref. [35] by considering a realistic p–n junction thermoelectric device. As shown in Fig. 4, a pair of an electron and a hole is generated by absorbing a phonon from the hot-phonon reservoir. The device has similarities with a solar cell, except that it exploits heat energy in the form of phonons. The same paradigm can be applied to heat in the form of other bosons, as long as the coupling between the electron and the bosons is strong. For the case of phonons, the coupling between electrons and optical phonons or acoustic phonons near van Hove singularities is very strong [35,47]. Studies in Ref. [35] indicate that, beside material engineering, thermoelectric efficiency and power can be significantly improved via *structure and geometry engineering*. Later, inelastic thermoelectric effects based on quantum dots [16,37] and quantum wells [38] have been studied. Considerably large output power and optimal efficiency have been found with realistic quantum dot/well parameters [16,37,38]. In Ref. [16], the phonon thermal conductivity is included using experimentally measured values. It was found in Ref. [16] that, although the phonon thermal conductivity reduces both figures of merit, ZT and $\tilde{Z}T$, the inelastic thermoelectric figure of merit $\tilde{Z}T$ is more robust than the elastic one ZT. In addition, $\tilde{Z}T$ is usually considerably larger than ZT for the quantum dots array system considered in Ref. [16]. Moreover, the output power for the inelastic thermoelectric effect is significantly greater than that of the elastic thermoelectric effect [16]. The figure of merit can reach a decent value of $\tilde{Z}T \sim 1$ using quantum dot arrays embedded in polymer thin films [16] or using HgCdTe p–n junctions [35]. These results confirm the advantages of the inelastic thermoelectric effect.

5. Rectification and transistor effects in the nonlinear regime for inelastic transport

In this section, we show how inelastic three-terminal transport enhances nonlinear devices like diodes and transistors. Diodes and transistors are crucial elements for electronics and conventional information technology. There is increasing interest to invent diodes and transistors of heat for information technology. Using heat and electricity together to achieve information processing may have energy advantages [48]. This idea was first explored in spin caloritronics, which tried to exploit heat to manipulate magnetic domain walls or induce spin currents, magnon currents, that may help to reduce the energy cost compared to using only electrical currents to achieve the same goal [48,49]. A goal of pursuing the use of heat alone for information technology, namely “phononics”, was also explored in the past decade [50].

Table 1
Functionality of second-order coefficients.

Terms (L_{ijk})	Diode or Transistor effect
L_{111}	charge rectification
L_{222}, L_{333}	electronic and phononic heat rectification
L_{233}, L_{322}	off-diagonal heat rectification
L_{122}, L_{133}	charge rectification by δT or ΔT
L_{211}, L_{311}	heat rectification by voltage
L_{113}, L_{123}	phonon-thermoelectric transistor
L_{212}, L_{112}	other nonlinear thermoelectric effects

In Ref. [17], it has been shown that combining thermal and electrical degrees of freedom together, several new nonlinear functions can be realized where heat and electricity have synergistic effects. For example, charge rectification can be induced by a temperature gradient, heat rectification can be realized by voltage bias. To describe those phenomena, we expand the currents to second order in affinities,

$$I_i = \sum_j M_{ij} A_j + \sum_{jk} L_{ijk} A_j A_k + \mathcal{O}(A^3) \tag{36}$$

The first term on the right-hand side describes the linear responses, whereas the second term gives the lowest-order nonlinear responses. The functionalities represented by various second-order coefficients are summarized in Table 1 [17]. Important examples include charge rectification induced by the temperature difference ΔT, i.e. the difference between phonon and electronic temperature, as well as phonon-thermoelectric transistor effect, i.e. tuning the I_e–V or the I_e–δT curve by the phonon temperature. The M_{ij} and L_{ijk} coefficients for a double quantum-dot system are calculated using realistic material parameters. The results in Ref. [17] indicate that the nonlinear effects are considerable and can be measured using the state-of-the-art experimental techniques. These observations indicate that the synergy of electronics and phononics can offer new platforms and opportunities for high-performance information nanotechnologies.

Beyond those observations, it is found that the inelastic three-terminal transport can induce thermal transistor effect *in the linear-response regime without relying on the onset of negative differential thermal conductance* [17]. The underlying physics can be revealed using thermodynamic arguments, without an explicit microscopic model.

Consider the restrictions on thermal conductance in the linear-response regime imposed by the second law of thermodynamics in a three-terminal set-up for thermal transistor functionality. If the thermal transport mechanisms involve only two terminals (reservoirs) in each microscopic process, then we can write the thermal transport equations as

$$I_{\mathrm{L}\to\mathrm{R}} = K_{\mathrm{LR}}(T_{\mathrm{L}} - T_{\mathrm{R}})/T \tag{37a}$$

$$I_{\mathrm{L}\to\mathrm{p}} = K_{\mathrm{Lp}}(T_{\mathrm{L}} - T_{\mathrm{p}})/T \tag{37b}$$

$$I_{\mathrm{R}\to\mathrm{p}} = K_{\mathrm{Rp}}(T_{\mathrm{R}} - T_{\mathrm{p}})/T \tag{37c}$$

where the subscripts L, R, p denote the source, drain, and heat bath, respectively, and T is the equilibrium temperature. The second law of thermodynamics requires that

$$K_{\mathrm{LR}} \geq 0, \quad K_{\mathrm{Lp}} \geq 0, \quad K_{\mathrm{Rp}} \geq 0 \tag{38}$$

However, due to energy conservation, there are only two independent heat currents, which we can choose as

$$I_{\mathrm{Q}}^{\mathrm{L}} = I_{\mathrm{L}\to\mathrm{R}} + I_{\mathrm{L}\to\mathrm{p}}, \quad I_{\mathrm{Q}}^{\mathrm{p}} = I_{\mathrm{p}\to\mathrm{L}} + I_{\mathrm{p}\to\mathrm{R}} \tag{39}$$

Then we can cast the thermal transport equation into the form

$$\begin{pmatrix} I_{\mathrm{Q}}^{\mathrm{L}} \\ I_{\mathrm{Q}}^{\mathrm{p}} \end{pmatrix} = \begin{pmatrix} K_{\mathrm{L}} & K_{\mathrm{o}} \\ K_{\mathrm{o}} & K_{\mathrm{p}} \end{pmatrix} \begin{pmatrix} \frac{T_{\mathrm{L}} - T_{\mathrm{R}}}{T} \\ \frac{T_{\mathrm{p}} - T_{\mathrm{R}}}{T} \end{pmatrix} \tag{40}$$

where

$$\begin{aligned} K_{\mathrm{L}} &= K_{\mathrm{LR}} + K_{\mathrm{Lp}} \\ K_{\mathrm{p}} &= K_{\mathrm{Lp}} + K_{\mathrm{Rp}}, \quad K_{\mathrm{o}} = -K_{\mathrm{Lp}} \end{aligned} \tag{41}$$

The transistor current amplification factor is defined as

$$\alpha \equiv \left| \frac{\partial_{T_{\mathrm{p}}} I_{\mathrm{Q}}^{\mathrm{L}}}{\partial_{T_{\mathrm{p}}} I_{\mathrm{Q}}^{\mathrm{p}}} \right| \tag{42}$$

In the linear-response regime, one usually finds

$$\alpha = \left| \frac{K_o}{K_p} \right| < 1 \tag{43}$$

Thus no thermal transistor effect exists in the linear-response regime.

However, the above analysis missed an important type of thermal transport mechanism that involves simultaneously *three* terminals (reservoirs) in each microscopic process. A typical example is the inelastic transport process illustrated in Fig. 3. For this type of transport, the thermal transport equation can also be written in the form of Eq. (40). However, the second law of thermodynamics only requires that

$$K_L \geq 0, \quad K_p \geq 0, \quad K_L K_p \geq K_o^2 \tag{44}$$

which does *not* forbid $\alpha = \left| \frac{K_o}{K_p} \right| > 1$. A realistic example that achieves $\alpha > 1$ in the linear-response regime can be found in Ref. [17]. This can also be illustrated by considering the double quantum-dot model, using Eq. (29). Keeping in mind that $I_Q^L = I_Q^e - \frac{1}{2} I_Q^p$ [using Eq. (22) for $\mu_L = \mu_R$], one has

$$\alpha = \frac{|E_1 - \mu|}{|E_2 - E_1|} \tag{45}$$

When $|E_1 - \mu| > |E_2 - E_1|$ (e.g., $E_1 > E_2 > \mu$), then α can be greater than 1 and there can be a heat transistor effect in the linear response regime.

6. Conclusions

Although the development of bulk semiconductor thermoelectric materials may be limited by material parameters, the study of mesoscopic thermoelectric effects is still far from maturity. It has given significant insights for improving thermoelectric efficiency and power in nanostructured thermoelectric materials (one of the main directions in thermoelectric research). We point out several aspects that mesoscopic systems may help with improving thermoelectric performances: inelastic transport, nonlinear effects, and spatial separation of thermal and electrical currents in thermoelectric transport. We also discuss bounds and quantization of thermal and thermoelectric transport coefficients in 1D single channel conductors. Channel number and the dependence on it are considered as well. We then discuss activated transport above a barrier in both the linear and nonlinear regimes and show their relevance for thermoelectric applications. We emphasize, particularly, inelastic thermoelectric effects in three-terminal geometry, which has attracted a lot of research interest recently. The physics and merits of the three-terminal inelastic thermoelectric effect are discussed and demonstrated for both linear and nonlinear transport. This paper attempts to give a pedagogical review of the research frontiers of thermoelectric inelastic transport effects in mesoscopic physics. Recent experimental progress [41–43] is pushing the field forward.

Acknowledgements

We thank Hamutal Bary-Soroker, Ora Entin-Wohlman, Zvi Ovadyahu, Dan Shahar, Moty Heiblum, Eli Zeldov and Dorri Halbertal for useful discussions. JHJ acknowledges support from the faculty start-up funding of Soochow University. YI acknowledges support from the Israeli Science Foundation (ISF) and the US–Israel Binational Science Foundation (BSF). We thank the referee and Robert Whitney for pointing out an error in the previous version of this paper.

References

[1] T.C. Harman, J.M. Honig, Thermoelectric and Thermomagnetic Effects and Applications, McGraw-Hill, New York, 1967;
H.J. Goldsmid, Introduction to Thermoelectricity, Springer, Heidelberg, Germany, 2009.
[2] G.D. Mahan, J.O. Sofo, Proc. Natl. Acad. Sci. USA 93 (1996) 7436.
[3] O. Entin-Wohlman, J.-H. Jiang, Y. Imry, Phys. Rev. E 89 (2014) 012123.
[4] J. Zhou, R. Yang, G. Chen, M.S. Dresselhaus, Phys. Rev. Lett. 107 (2011) 226601;
C. Jeong, R. Kim, M. Lundstrom, arXiv:1103.1274.
[5] E.g. J.-H. Jiang, O. Entin-Wohlman, Y. Imry, Phys. Rev. B 87 (2013) 205420;
R. Bosisio, C. Gorini, G. Fleury, J.-L. Pichard, New J. Phys. 16 (2014) 095005;
R. Bosisio, C. Gorini, G. Fleury, J.-L. Pichard, Phys. Rev. Appl. 3 (2015) 054002.
[6] L.D. Hicks, M.S. Dresselhaus, Phys. Rev. B 47 (1993) 12727;
L.D. Hicks, M.S. Dresselhaus, Phys. Rev. B 47 (1993) 16631.
[7] R. Venkatasubramanian, Phys. Rev. B 61 (2000) 3091;
J.-K. Yu, S. Mitrovic, D. Tham, J. Varghese, J.R. Heath, Nat. Nanotechnol. 5 (2010) 718;
R. Venkatasubramanian, E. Siivola, T. Colpitts, B. O'Quinn, Nature 413 (2001) 597.
[8] A.I. Boukai, et al., Nature 451 (2008) 168;
B. Poudel, et al., Science 320 (2008) 634;
P. Pichanusakorn, P. Bandaru, Mater. Sci. Eng., R Rep. 67 (2010) 19;
A.J. Minnich, M.S. Dresselhaus, Z.F. Ren, G. Chen, Energy Environ. Sci. 2 (2009) 466;
C.J. Vineis, A. Shakouri, A. Majumdar, M.G. Kanatzidis, Adv. Mater. 22 (2010) 3970;
Z.-G. Chen, G. Han, L. Yang, L. Cheng, J. Zou, Prog. Nat. Sci. 22 (2012) 535;
J.-F. Li, W.-S. Liu, L.-D. Zhao, M. Zhou, NPG Asia Mater. 2 (2010) 152.

[9] Kanishka Biswas, Jiaqing He, Ivan D. Blum, Chun-I. Wu, Timothy P. Hogan, David N. Seidman, Vinayak P. Dravid, Mercouri G. Kanatzidis, Nature 489 (2012) 414.
[10] L.D. Hicks, T.C. Harman, M.S. Dresselhaus, Appl. Phys. Lett. 63 (1993) 3230.
[11] M.S. Dresselhaus, G. Chen, M.Y. Tang, R. Yang, H. Lee, D. Wang, Z. Ren, J.-P. Fleurial, P. Gogna, Adv. Mater. 19 (2007) 1043.
[12] M. Cutler, N.F. Mott, Phys. Rev. 181 (1969) 1336.
[13] U. Sivan, Y. Imry, Phys. Rev. B 33 (1986) 551.
[14] R. Whitney, Phys. Rev. Lett. 112 (2014) 130601.
[15] J.-H. Jiang, O. Entin-Wohlman, Y. Imry, Phys. Rev. B 85 (2012) 075412.
[16] J.-H. Jiang, J. Appl. Phys. 116 (2014) 194303.
[17] J.-H. Jiang, M. Kulkarni, D. Segal, Y. Imry, Phys. Rev. B 92 (2015) 045309.
[18] H.G. Bremmerman, in: L.M. LeCam, J. Neyman (Eds.), Proc. Fifth Berkeley Symp. on Math. Statistics and Probability, University of California, Berkeley, CA, USA, 1966;
J.D. Bekenstein, Phys. Rev. Lett. 46 (1981) 623;
P.A. Benioff, Int. J. Theor. Phys. 21 (1982) 177;
J.B. Pendry, J. Phys. A, Math. Gen. 16 (1983) 2161.
[19] Y. Imry, Physics of mesoscopic systems, in: G. Grinstein, G. Mazenko (Eds.), Directions in Condensed Matter, Memorial Volume in Honor of Professor Shang-keng Ma, World Scientific, Singapore, 1986;
Y. Imry, Introduction to Mesoscopic Physics, Oxford University Press, 1997 (out of print). 2nd edition, 2002.
[20] R. Landauer, Philos. Mag. 21 (1970) 863.
[21] B.J. van Wees, H. van Houten, C.W.J. Beenakker, J.G. Williamson, L.P. Kouwenhoven, D. van der Marel, C.T. Foxon, Phys. Rev. Lett. 60 (1988) 848;
D.A. Wharam, T.J. Thornton, R. Newbury, M. Pepper, H. Ahmed, J.E.F. Frost, D.G. Hasko, D.C. Peacock, D.A. Ritchie, G.A.C. Jones, J. Phys. C, Solid State Phys. 21 (1988) L209.
[22] The thermal conductance is defined under the condition of zero current, not zero voltage. This gives a (usually small) thermoelectric correction to the "bare" thermal conductance K^0 that we calculate. We shall ignore this correction in the present section.
[23] J.M. Ziman, Principles of the Theory of Solids, Cambridge University Press, Cambridge, UK, 1969, Chap. 7.
[24] Y. Imry, A. Amir, The localization transition at finite temperatures: electric and thermal transport, in: E. Abrahams (Ed.), 50 Years of Anderson Localization, World Scientific, Singapore, 2010, pp. 191–213, Chap. 9, or arXiv:1004.0966.
[25] K. Schwab, E.A. Henriksen, J.M. Worlock, M.L. Roukes, Nature (London) 404 (2000) 974.
[26] N. Nakpathomkun, H.Q. Xu, H. Linke, Phys. Rev. B 82 (2010) 235428.
[27] G.D. Mahan, J. Appl. Phys. 76 (1994) 4363;
G.N. Hatsopoulos, E.P. Gyftopoulos, Thermionic Energy Conversion, vol. 1, MIT, Cambridge, 1973.
[28] D. Vashaee, A. Shakouri, Phys. Rev. Lett. 92 (2004) 106103.
[29] O. Entin-Wohlman, Y. Imry, A. Aharony, Phys. Rev. B 82 (2010) 115314.
[30] R. Sánchez, M. Büttiker, Phys. Rev. B 83 (2011) 085428.
[31] R. Sánchez, B. Sothmann, A.N. Jordan, M. Büttiker, New J. Phys. 15 (2013) 125001.
[32] B. Sothmann, R. Sánchez, A.N. Jordan, Nanotechnology 26 (2015) 032001.
[33] H.L. Edwards, Q. Niu, A.L. de Lozanne, Appl. Phys. Lett. 63 (1993) 1815;
H.L. Edwards, Q. Niu, G.A. Georgakis, A.L. de Lozanne, Phys. Rev. B 52 (1995) 5714.
[34] J.R. Prance, C.G. Smith, J.P. Griffiths, S.J. Chorley, D. Anderson, G.A.C. Jones, I. Farrer, D.A. Ritchie, Phys. Rev. Lett. 102 (2009) 146602.
[35] J.-H. Jiang, O. Entin-Wohlman, Y. Imry, New J. Phys. 15 (2013) 075021.
[36] B. Sothmann, R. Sánchez, A.N. Jordan, M. Büttiker, Phys. Rev. B 85 (2012) 205301.
[37] A.N. Jordan, B. Sothmann, R. Sánchez, M. Büttiker, Phys. Rev. B 87 (2013) 075312.
[38] B. Sothmann, R. Sánchez, A.N. Jordan, M. Büttiker, New J. Phys. 15 (2013) 095021.
[39] L. Simine, D. Segal, Phys. Chem. Chem. Phys. 14 (2012) 13820;
L. Simine, D. Segal, J. Chem. Phys. 141 (2014) 014704;
L. Arrachea, N. Bode, F. von Oppen, Phys. Rev. B 90 (2014) 125450.
[40] B. Cleuren, B. Rutten, C. Van den Broeck, Phys. Rev. Lett. 108 (2012) 120603.
[41] B. Roche, P. Roulleau, T. Jullien, Y. Jompol, I. Farrer, D.A. Ritchie, D.C. Glattli, Nat. Commun. 6 (2015) 6738.
[42] H. Thierschmann, et al., Nat. Nanotechnol. 10 (2015) 854.
[43] F. Hartmann, P. Pfeffer, S. Höfling, M. Kamp, L. Worschech, Phys. Rev. Lett. 114 (2015) 146805.
[44] B. Sothmann, M. Büttiker, Europhys. Lett. 99 (2012) 27001.
[45] B. Rutten, M. Esposito, B. Cleuren, Phys. Rev. B 80 (2009) 235122;
T. Ruokola, T. Ojanen, Phys. Rev. B 86 (2012) 035454.
[46] L. Onsager, Phys. Rev. 37 (1931) 405;
L. Onsager, Phys. Rev. 38 (1931) 2265.
[47] H. Haug, S.W. Koch, Quantum Theory of the Optical and Electronic Properties of Semiconductors, 4th ed., World Scientific, Singapore, 2004.
[48] J.H. Jiang, Phys. Rev. E 90 (2014) 042126.
[49] G.E.W. Bauer, E. Saitoh, B.J. van Wees, Nat. Mater. 11 (2012) 391.
[50] N. Li, J. Ren, L. Wang, G. Zhang, P. Hänggi, B. Li, Rev. Mod. Phys. 84 (2012) 1045.