

Finding the quantum thermoelectric with maximal efficiency and minimal entropy production at given power output

Robert S. Whitney

Laboratoire de Physique et Modélisation des Milieux Condensés (UMR 5493), Université Grenoble 1 and CNRS, Maison des Magistères, BP 166, 38042 Grenoble, France

(Received 30 October 2014; revised manuscript received 28 January 2015; published 19 March 2015)

We investigate the nonlinear scattering theory for quantum systems with strong Seebeck and Peltier effects, and consider their use as heat engines and refrigerators with finite power outputs. This paper gives detailed derivations of the results summarized in a previous paper [R. S. Whitney, *Phys. Rev. Lett.* **112**, 130601 (2014)]. It shows how to use the scattering theory to find (i) the quantum thermoelectric with maximum possible power output, and (ii) the quantum thermoelectric with maximum efficiency at given power output. The latter corresponds to a minimal entropy production at that power output. These quantities are of quantum origin since they depend on system size over electronic wavelength, and so have no analog in classical thermodynamics. The maximal efficiency coincides with Carnot efficiency at zero power output, but decreases with increasing power output. This gives a fundamental lower bound on entropy production, which means that reversibility (in the thermodynamic sense) is impossible for finite power output. The suppression of efficiency by (nonlinear) phonon and photon effects is addressed in detail; when these effects are strong, maximum efficiency coincides with maximum power. Finally, we show in particular limits (typically without magnetic fields) that relaxation within the quantum system does not allow the system to exceed the bounds derived for relaxation-free systems, however, a general proof of this remains elusive.

DOI: 10.1103/PhysRevB.91.115425

PACS number(s): 73.63.-b, 05.70.Ln, 72.15.Jf, 84.60.Rb

I. INTRODUCTION

Thermoelectric effects in nanostructures [1–4] and molecules [5,6] are of great current interest. They might enable efficient electricity generation and refrigeration [7–9], and could also lead to new types of sub-Kelvin refrigeration, cooling electrons in solid-state samples to lower temperatures than with conventional cryostats [1], or cooling fermionic atomic gases [10–12]. However, they are also extremely interesting at the level of fundamental physics since they allow one to construct the simplest possible quantum machine that converts heat flows into useful work (electrical power in this case) or vice versa. This makes them an ideal case study for *quantum thermodynamics*, i.e., the thermodynamics of quantum systems [13].

The simplest heat engine is a thermocouple circuit, as shown in Fig. 1. It consists of a pair of thermoelectrics with opposite thermoelectric responses (filled and open circles) and a load, connected in a ring. Between each such circuit element is a big reservoir of electrons, the reservoir on the left (L) is hotter than the others $T_L > T_R$, so heat flows from left to right. One thermoelectric's response causes an electric current to flow in the opposite direction to the heat flow (filled circle), while the other's causes an electric current to flow in the same direction as the heat flow (open circle). Thus, the two thermoelectrics turn heat energy into electrical work: a current flow I through the load. The load is assumed to be a device that turns the electrical work into some other form of work: it could be a battery charger (turning electrical work into chemical work) or a motor (turning electrical work into mechanical work).

The same thermocouple circuit can be made into a refrigerator simply by replacing the load with a power supply. The power supply does work to establish the current I around the circuit, and this current through the thermoelectrics can “drag” heat out of reservoir L . In other words, the electrical current

and heat flow are the same as for the heat engine, but now the former causes the latter rather than vice versa. Thus, the refrigerator cools reservoir L , so $T_L < T_R$.

The laws of classical thermodynamics inform us that entropy production can never be negative, and maximal efficiency occurs when a system operates reversibly (zero entropy production). Thus, it places fundamental bounds on heat-engine and refrigerator efficiencies, known as Carnot efficiencies. In both cases, the efficiency is defined as the power output divided by the power input. For the heat engine, the power input is the heat current out of the hotter reservoir (reservoir L) J_L and the power output is the electrical power generated P_{gen} . Thus, the heat-engine (eng) efficiency is

$$\eta_{\text{eng}} = P_{\text{gen}}/J_L. \quad (1)$$

This efficiency can never exceed Carnot's limit

$$\eta_{\text{eng}}^{\text{Carnot}} = 1 - T_R/T_L, \quad (2)$$

where we recall that we have $T_L > T_R$.

For the refrigerator the situation is reversed, the load is replaced by a power supply, and the power input is the electrical power that the circuit absorbs from the power supply P_{abs} . The power output is the heat current out of the colder reservoir (reservoir L) J_L . This is called the cooling power because it is the rate at which the circuit removes heat energy from reservoir L . Thus, the refrigerator (fri) efficiency is

$$\eta_{\text{fri}} = J_L/P_{\text{abs}}. \quad (3)$$

This efficiency is often called the coefficient of performance or COP. This efficiency can never exceed Carnot's limit

$$\eta_{\text{fri}}^{\text{Carnot}} = (T_R/T_L - 1)^{-1}, \quad (4)$$

where we recall that $T_L < T_R$ (opposite of heat engine).

Strangely, the laws of classical thermodynamics do not appear to place a fundamental bound on the power output

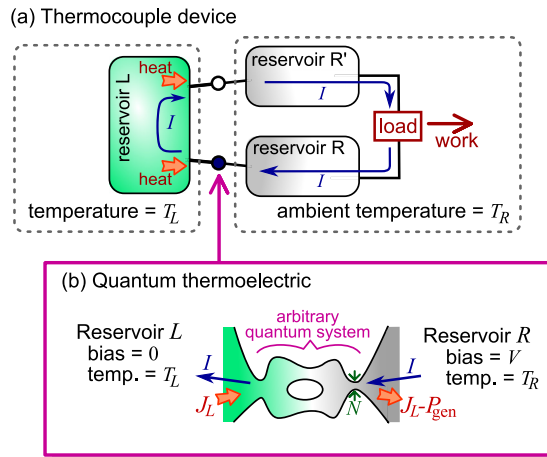


FIG. 1. (Color online) (a) The simplest heat engine is a thermocouple circuit made of two thermoelectrics (filled and open circles). The filled and open circles are quantum systems with opposite thermoelectric responses, an example could be that in (b). For a heat engine, we assume $T_L > T_R$, so heat flows as shown, generating a current I , which provides power to a load (battery charger, motor, etc.) that converts the electrical power into some other form of work. The same thermocouple circuit can act as a refrigerator, if one replaces the load with a power supply that generates the current I . This induces the heat flow out of reservoir L , which thereby refrigerates reservoir L , so $T_L < T_R$. Note that in both cases the circuit works because the two thermoelectrics are electrically in series but thermally in parallel. In (b), N indicates the number of transverse modes in the narrowest part of the quantum system.

associated with reversible (Carnot efficient) operation. Most textbooks say that reversibility requires “small” power output, but rarely define what “small” means. The central objective of Ref. [14] was to find the meaning of “small” and find a fundamental upper bound on the efficiency of an irreversible system in which the power output was *not* small.

Reference [14] did this for the class of quantum thermoelectrics that are well modeled by a scattering theory, which enables one to straightforwardly treat quantum and thermodynamic effects on an equal footing. It summarized two principal results absent from classical thermodynamics. First, there is a quantum bound (qb) on the power output, and no quantum system can exceed this bound (open circles in Fig. 2). Second, there is an upper bound on the efficiency at any given power output less than this bound (thick black curves in Fig. 2). The efficiency at given power output can only reach Carnot efficiency when the power output is very small compared to the quantum bound on power output. The upper bound on efficiency then decays monotonically as one increases the power output towards the quantum bound. The objective of this paper is to explain in detail the methods used to derive these results, along with the other results that were summarized in Ref. [14].

A. Contents of this paper

This paper provides detailed derivations of the results in Ref. [14]. The first part of this paper is an extended introduction. Section II is a short review of the relevant

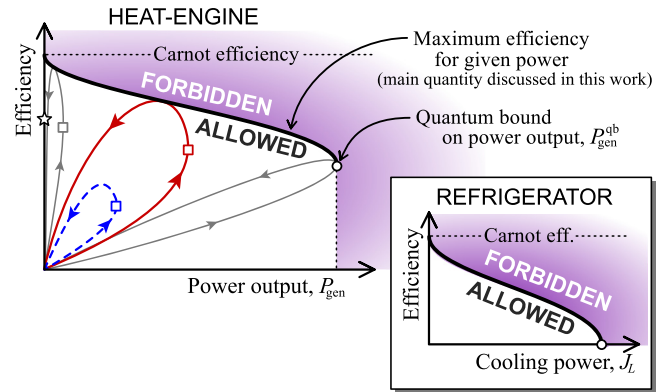


FIG. 2. (Color online) The thick black curves are qualitative sketches of the maximum efficiency as a function of heat-engine power output (main plot), or refrigerator cooling power (inset), with the shaded regions being forbidden. Precise plot of such curves for different temperature ratios T_R/T_L are shown in Fig. 9. The colored loops (red, gray, and blue) are typical sketches of the efficiency versus power of *individual* heat engines as we increase the load resistance (direction of arrows on loop). The power output $P_{\text{gen}} = IV$ vanishes when the load resistance is zero (for which $V = 0$) or infinite (for which $I = 0$), with a maximum at an intermediate resistance (open square). The curves have a characteristic loop form [2], however, the exact shape of the loop depends on many system-specific details, such as charging effects. The dashed blue loop is for a typical nonoptimal system (always well below the upper bound), while the solid red and gray loops are for systems which achieve the upper bound for a particular value of the load. The star marks the Curzon-Ahlborn efficiency.

literature. Section III discusses how we define temperature, heat, and entropy. Section IV recalls the connection between efficiency and entropy production in any thermodynamic machine. Section V reviews the nonlinear scattering theory, which Sec. VII uses to make very simple overestimates of a quantum system’s maximum power output.

The second part of this paper considers how to optimize a system which is free of relaxation and has no phonons or photons. Section VIII gives a hand-waving explanation of the optimal heat engine, while Sec. IX gives the full derivation. Section X gives a hand-waving explanation of the optimal refrigerator, while Sec. XI gives the full derivation. Section XII proposes a system which could in principle come arbitrarily close to the optimal properties given in Secs. IX and XI. Section XIII considers many quantum thermoelectrics in parallel.

The third part of this paper considers certain effects neglected in the above idealized system. Section XIV adds the parasitic effect of phonon or photon carrying heat in parallel to the electrons. Section XV treats relaxation within the quantum system.

II. COMMENTS ON EXISTING LITERATURE

There is much interest in using thermoelectric effects to cool fermionic atomic gases [10–12], which are hard to cool via other methods. This physics is extremely similar to that in this work, but there is a crucial difference. For the electronic

systems that we consider, we can assume the temperatures to be much less than the reservoir's Fermi energy, and so take all electrons to have the same Fermi wavelength. In contrast, fermionic atomic gases have temperatures of order the Fermi energy, so the high-energy particles in a reservoir have a different wavelength from the low-energy ones. Thus, our results do not apply to atomic gases, although our methodology does [12].

A. Nonlinear systems and the figure of merit ZT

Engineers commonly state that wide-ranging applications for thermoelectrics would require them to have a dimensionless figure of merit ZT greater than three. This dimensionless figure of merit is a dimensionless combination of the linear-response coefficients [7] $ZT = TGS^2/\Theta$, for temperature T , Seebeck coefficient S , electrical conductance G , and thermal conductance Θ . Yet, for us, ZT is just a way to characterize the efficiency, via

$$\eta_{\text{eng}} = \eta_{\text{eng}}^{\text{Carnot}} \frac{\sqrt{ZT + 1} - 1}{\sqrt{ZT + 1} + 1},$$

with a similar relationship for refrigerators. Thus, someone asking for a device with a $ZT > 3$ actually requires one with an efficiency of more than one-third of Carnot efficiency. This is crucial because the efficiency is a physical quantity in linear and nonlinear situations, while ZT is only meaningful in the linear-response regime [15–20].

Linear-response theory rarely fails for bulk semiconductors, even when T_L and T_R are very different. Yet, it is completely *inadequate* for the quantum systems that we consider here. Linear-response theory requires the temperature drop on the scale of the electron relaxation length l_{rel} (distance traveled before thermalizing) to be much less than the average temperature. For a typical millimeter-thick bulk thermoelectric between a diesel motor's exhaust system ($T_L \simeq 700$ K) and its surroundings ($T_R \simeq 280$ K), the relaxation length (inelastic scattering length) is of order the mean-free path, typically 1–100 nm. The temperature drop on this scale is tens of thousands of times smaller than the temperature drop across the whole thermoelectric. This is absolutely tiny compared with the average temperature, so linear-response [21] works well, even though $(T_L - T_R)/T_L$ is of order one.

In contrast, for quantum systems ($L \ll l_{\text{rel}}$), the whole temperature drop occurs on the scale of a few nanometers or less, and so linear-response theory is inapplicable whenever $(T_L - T_R)/T_L$ is not small.

B. Carnot efficiency

A system must be reversible (create no entropy) to have Carnot efficiency; proposals exist to achieve this in bulk [21] or quantum [22–24] thermoelectric. It requires that electrons only pass between reservoirs L and R at the energy where the occupation probabilities are identical in the two reservoirs [22]. Thus, a thermoelectric requires two things to be reversible. First, it must have a δ -function-like transmission [21–25], which only lets electrons through at energy ϵ_0 . Second [22], the load's resistance must be such that $e^{-V} = \epsilon_0(1 - T_R/T_L)$, so the reservoirs' occupations are equal at ϵ_0 (see Fig. 4).

By definition, this means the current vanishes, and thus so does the power output P_{gen} . However, one can see how P_{gen} vanishes by considering a quantum system which lets electrons through in a tiny energy window Δ from ϵ_0 to $\epsilon_0 + \Delta$ (see Fig 5). When we take $\Delta/(k_B T_{L,R}) \rightarrow 0$, one has Carnot efficiency, however, we will see [leading-order term in Eq. (48)] that

$$P_{\text{gen}} \propto \frac{1}{\hbar} \Delta^2, \quad (5)$$

which vanishes as $\Delta/(k_B T_{L,R}) \rightarrow 0$.

C. Heat-engine efficiency at finite power output and Curzon-Ahlborn efficiency

To increase the power output beyond that of a reversible system, one has to consider irreversible machines which generate a finite amount of entropy per unit of work generated. Curzon and Ahlborn [26] popularized the idea of studying the efficiency of a heat engine running at its maximum power output. For classical pumps, this efficiency is $\eta_{\text{eng}}^{\text{CA}} = 1 - \sqrt{T_L/T_R}$, which is now called the Curzon-Ahlborn efficiency, although already given in Refs. [27–29]. As refrigerators, these pumps have an efficiency at maximum cooling power of zero, although Refs. [30–33] discuss ways around this.

The response of a given heat engine is typically a “loop” of efficiency versus power (see Fig. 2) as one varies the load on the system [2]. For a peaked transmission function with width Δ (see, e.g., Fig. 5), the loop moves to the left as one reduces Δ . In the limit $\Delta \rightarrow 0$, the whole loop is squashed onto the $P_{\text{gen}} = 0$ axis. In linear-response language, this machine has $ZT \rightarrow \infty$. In this limit, the efficiency at maximum power can be very close to that of Curzon and Ahlborn [34] (the star in Fig. 2), just as its maximum efficiency can be that of Carnot [22] (see previous section). However, its maximum power output is $\propto e^{-V} \Delta/\hbar$ for small Δ (where V is finite, chosen to ensure maximum power), which vanishes for $\Delta \rightarrow 0$, although it is much larger than Eq. (5). Figure 2 shows that a system with larger Δ (such as the red curve) operating near its maximum efficiency will have both higher efficiency and higher power output than the one with small Δ (leftmost gray curve) operating at maximum power.

This paper shows how to derive the thick black curve in Fig. 2, thereby showing that there is a fundamental tradeoff between efficiency and power output in optimal thermodynamic machines made from thermoelectrics [35]. As such, our work overturns the idea that maximizing efficiency at maximum power is the best route to machines with both high efficiency and high power. It also overturns the idea that systems with the narrowest transmission distributions (the largest ZT in linear response) are automatically the best thermoelectrics.

At this point, we mention that other works [36–39] have studied efficiencies for various systems with finite-width transmission functions, for which power outputs can be finite. In particular, Ref. [39] considered a boxcar transmission function, which is the form of transmission function that we have shown can be made optimal [14].

D. Pendry's quantum bound on heat flow

An essential ingredient in this work is Pendry's upper bound [40] on the heat flow through a quantum system between two reservoirs of fermions. He found this bound using a scattering theory of the type discussed in Sec. V. It is a concrete example of a general principle due to Bekenstein [41], and the same bound applies in the presence of thermoelectric effects [42]. The bound on the heat flow out of reservoir L is achieved when all the electrons and holes arriving at the quantum system from reservoir L escape into reservoir R without impediment, while there is no back-flow of electrons or holes from reservoir R to L . The easiest way to achieve this is to couple reservoir L through the quantum system to a reservoir R at zero temperature, and then ensure the quantum system does not reflect any particles. In this case, the heat current equals

$$J_L^{\text{qb}} = \frac{\pi^2}{6h} N k_B^2 T_L^2, \quad (6)$$

where N is the number of transverse modes in the quantum system. We refer to this as the quantum bound (qb) on heat flow because it depends on the quantum wave nature of the electrons; it depends on N , which is given by the cross-sectional area of the quantum system divided by λ_F^2 , where λ_F is the electron's Fermi wavelength. As such, J_L^{qb} is ill defined within classical thermodynamics.

III. UNIQUELY DEFINING TEMPERATURE, HEAT, AND ENTROPY

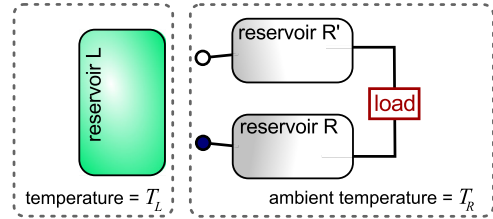
Works on classical thermodynamics have shown that the definition of heat and entropy flows can be fraught with difficulties. For example, the rate of change of entropy cannot always be uniquely defined in classical continuum thermodynamics [43–45]. Here, the situation is even more difficult since the electrons within the quantum systems (circles in Fig. 1) are not at equilibrium, and so their temperature cannot be defined. Thus, it is crucial to specify the logic which leads to our definitions of temperature, heat flow, and entropy flow.

Our definition of heat flow originated in Refs. [46–48], the rate of change of entropy is then found using the Clausius relation [49] (see following). To explain these quantities and show they are unambiguous, we consider the following three-step procedure for a heat engine. An analog procedure works for a refrigerator.

Step 1. Reservoir L is initially decoupled from the rest of the circuit [see Fig. 3(a)], has internal heat energy $Q_L^{(0)}$, and is in internal equilibrium at temperature $T_L^{(0)}$. The rest of the circuit is in equilibrium at temperature $T_R^{(0)}$ with internal heat energy $Q_R^{(0)}$. The internal heat energy is the total energy of the reservoir's electron gas minus the energy which that gas would have in its ground state. As such, the internal energy can be written as a sum over electrons and holes, with an electron at energy ϵ above the reservoir's chemical potential (or a hole at energy ϵ below that chemical potential) contributing ϵ to this internal heat energy. The initial entropies are then $S_i^{(0)} = Q_i^{(0)}/T_i^{(0)}$ for $i = L, R$.

Step 2. We connect reservoir L to the rest of the circuit [see Fig. 3(b)] and leave it connected for a long time t_{expt} . While

(a) Reservoir L decoupled, each dotted box in internal equilibrium



(b) Reservoir L coupled, circuit out of equilibrium & generating work

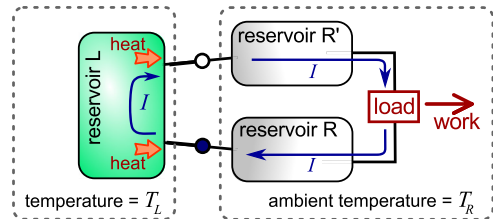


FIG. 3. (Color online) To implement the procedure in Sec. III, one starts with the circuit unconnected, as in (a), one then connects the circuit, as in (b). After a long time t_{expt} , one disconnects the circuit, returning to (a). The circles are the quantum thermoelectrics, as in Fig. 1.

we assume t_{expt} is long, we also assume that the reservoirs are all large enough that the energy distributions within them change very little during time t_{expt} . Upon connecting the circuit elements, we assume a transient response during a time t_{trans} , after which the circuit achieves a steady state. We ensure that $t_{\text{expt}} \gg t_{\text{trans}}$, so the physics is dominated by this steady state. Even then the flow will be noisy [50] due to the fact electrons are discrete with probabilistic dynamics. So, we also ensure that t_{expt} is much longer than the noise correlation time, so that the noise in the currents is negligible compared to the average currents.

Step 3. After the time t_{expt} , we disconnect reservoir L from the rest of the circuit. Again, there will be a transient response, however, we assume that a weak relaxation mechanism within the reservoirs will cause the two parts of the circuit to each relax to internal equilibrium [see Fig. 3(a)]. After this, one can unambiguously identify the temperature T_i , internal energy Q_i , and Clausius entropy $S_i = Q_i/T_i$ of the two parts of the circuit (for $i = L, R$). Since the reservoirs are large, we assume $T_i = T_i^{(0)}$.

Thus, we can unambiguously say that the heat current out of reservoir i averaged over the time t_{expt} is

$$\langle J_i \rangle = (Q_i^{(0)} - Q_i)/t_{\text{expt}}. \quad (7)$$

For the above thermocouple, we treat the currents for each thermoelectric separately, writing the heat current out of reservoir L as $J_L + J_{L'}$, where J_L is the heat current from reservoir L into the lower thermoelectric in Fig. 1 (the filled circle), and $J_{L'}$ is the heat current from reservoir L into the upper thermoelectric in Fig. 1 (the open circle). Treating each thermoelectric separately is convenient, and also allows one to generalize the results to ‘‘thermopiles,’’ which contain hundreds of thermoelectrics arranged so that they are electrically in series, but thermally in parallel.

The average rate of change of entropy in the circuit is $\langle \dot{S}_{\text{circuit}} \rangle = \langle \dot{S} \rangle + \langle \dot{S}' \rangle$, where $\langle \dot{S} \rangle$ is the average rate of change of entropy associated with the lower thermoelectric in Fig. 1, while $\langle \dot{S}' \rangle$ is that for the upper thermoelectric. Then,

$$\langle \dot{S} \rangle = \langle \dot{S}_L \rangle + \langle \dot{S}_R \rangle = -\langle J_L \rangle / T_L - \langle J_R \rangle / T_R, \quad (8)$$

while $\langle \dot{S}' \rangle$ is the same with J_L, J_R, T_R replaced by J_L, J_R, T_R . We neglect the entropy of the thermoelectrics and load, by assuming their initial and final states are the same. This will be the case if they are small compared to the reservoirs, so their initial and final states are simply given by the temperature T_R .

The nonlinear scattering theory in Ref. [51] captures long-time average currents (usually called the dc response in electronics), such as electrical current $\langle I_i \rangle$ and heat current $\langle J_i \rangle$ (see references in Sec. V). It is believed to be exact for noninteracting particles, and also applies when interactions can be treated in a mean-field approximation (see again Sec. V). A crucial aspect of the scattering theory is that we do not need to describe the nonequilibrium state of the quantum system during step 2. Instead, we need that quantum system's transmission function, defined in Sec. V.

In this paper, we will *only* discuss the long-time average of the rates of flows (not the noisy instantaneous flows), and thus will not explicitly indicate the average; so I_i, J_i , and \dot{S}_i should be interpreted as $\langle I_i \rangle, \langle J_i \rangle$, and $\langle \dot{S}_i \rangle$.

IV. ENTROPY PRODUCTION

There are little known universal relations between efficiency, power, and entropy production, which follow trivially from the laws of thermodynamics [52]. Consider the lower thermoelectric in Fig. 1(a) (filled circle), with J_L and J_R being steady-state heat currents into it from reservoirs L and R . Then, the first law of thermodynamics is

$$J_R + J_L = P_{\text{gen}}, \quad (9)$$

where P_{gen} is the electrical power generated. The Clausius relation for the rate of change of total entropy averaged over long times as in Eq. (8) is

$$\dot{S} = -\frac{J_L}{T_L} + \frac{J_L - P_{\text{gen}}}{T_R}, \quad (10)$$

where we have used Eq. (9) to eliminate J_R .

For a heat engine, we take J_L to be positive, which means $T_L > T_R$ and J_R is negative. We use Eq. (1) to replace J_L with $P_{\text{gen}}/\eta_{\text{eng}}$ in Eq. (10). Then, the rate of entropy production by a heat engine with efficiency $\eta_{\text{eng}}(P_{\text{gen}})$ at power output P_{gen} is

$$\dot{S}(P_{\text{gen}}) = \frac{P_{\text{gen}}}{T_R} \left(\frac{\eta_{\text{eng}}^{\text{Carnot}}}{\eta_{\text{eng}}(P_{\text{gen}})} - 1 \right), \quad (11)$$

where the Carnot efficiency $\eta_{\text{eng}}^{\text{Carnot}}$ is given in Eq. (2). Hence, knowing the efficiency at power P_{gen} tells us the entropy production at that power. Maximizing the former minimizes the latter.

For refrigeration, the load in Fig. 1 is replaced by a power supply, the thermoelectric thus absorbs a power P_{abs} to extract heat from the cold reservoir. We take reservoir L as cold ($T_L < T_R$), so J_L is positive. We replace P_{gen} by $-P_{\text{abs}}$ in Eqs. (9)

and (10). We then use Eq. (3) to replace P_{abs} by J_L/η_{fri} . Then, the rate of entropy production by a refrigerator at cooling power J_L is

$$\dot{S}(J_L) = \frac{J_L}{T_R} \left(\frac{1}{\eta_{\text{fri}}(J_L)} - \frac{1}{\eta_{\text{fri}}^{\text{Carnot}}} \right), \quad (12)$$

where the Carnot efficiency $\eta_{\text{fri}}^{\text{Carnot}}$ is given in Eq. (4). Hence, knowing a refrigerator's efficiency at cooling power J_L gives us its entropy production, and we see that maximizing the former minimizes the latter.

Equations (11) and (12) hold for systems modeled by scattering theory because this theory satisfies the laws of thermodynamics [42,53]. The rate of entropy production is zero when the efficiency is that of Carnot, but becomes increasingly positive as the efficiency reduces. In this paper, we calculate the maximum efficiency for given power output, and then use Eqs. (11) and (12) to get the minimum rate of entropy production at that power output.

V. NONLINEAR SCATTERING THEORY

This work uses Christen and Büttiker's nonlinear scattering theory [51], which treats electron-electron interactions as mean-field charging effects. References [17,18,54] added thermoelectric effects by following works on linear response [46–48]. Particle and heat flows are given by the transmission function $\mathcal{T}_{RL}(\epsilon)$ for electrons to go from left (L) to right (R) at energy ϵ , where $\mathcal{T}_{RL}(\epsilon)$ is a *self-consistently* determined function of T_L, T_R , and V . In short, this self-consistency condition originates from the fact that electrons injected from the leads change the charge distribution in the quantum system, which in turn changes the behavior of those injected electrons (via electron-electron interactions). The transmission function can be determined self-consistently with the charge distribution, if the latter is treated in a time-independent mean-field manner (neglecting single-electron effects). We note that the same nonlinear scattering theory was also derived for resonant level models [22,36] using functional renormalization group (RG) to treat single-electron charging effects [37].

The scattering theory for the heat current is based on the observation that an electron leaving reservoir i at energy ϵ is carrying heat $\epsilon - \mu_i$ out of that reservoir [48], where μ_i is the reservoir's chemical potential. Thus, a reservoir is cooled by removing an electron above the Fermi surface, but heated by removing an electron below the Fermi surface. It is convenient to treat empty states below a reference chemical potential (which we define as $\epsilon = 0$) as “holes.” Then, we do not need to keep track of a full Fermi sea of electrons, but only the holes in that Fermi sea. Then, the heat currents out of reservoirs L and R and into the quantum system are

$$J_L = \frac{1}{h} \sum_{\mu} \int_0^{\infty} d\epsilon (\epsilon - \mu e^{-V_L}) \mathcal{T}_{RL}^{\mu\mu}(\epsilon) [f_L^{\mu}(\epsilon) - f_R^{\mu}(\epsilon)], \quad (13)$$

$$J_R = \frac{1}{h} \sum_{\mu} \int_0^{\infty} d\epsilon (\epsilon - \mu e^{-V_R}) \mathcal{T}_{RL}^{\mu\mu}(\epsilon) [f_R^{\mu}(\epsilon) - f_L^{\mu}(\epsilon)], \quad (14)$$

where e^- is the electron charge ($e^- < 0$), so e^-V_i is the chemical potential of reservoir i measured from the reference chemical potential ($\epsilon = 0$). The sum is over $\mu = 1$ for “electron” states (full states above the reference chemical potential), and $\mu = -1$ for “hole” states (empty states below that chemical potential). The Fermi function for particles entering from reservoir j is

$$f_j^\mu(\epsilon) = \{1 + \exp[(\epsilon - \mu e^- V_j)/(k_B T_j)]\}^{-1}. \quad (15)$$

The transmission function $\mathcal{T}_{ij}^{\nu\mu}(\epsilon)$ is the probability that a particle μ with energy ϵ entering the quantum system from reservoir j will exit into reservoir i as a particle ν with energy ϵ . We only allow $\nu = \mu$ here since we do not consider electron to hole scattering within the quantum system (only common when superconductors are present). Interactions mean that $\mathcal{T}_{RL}^{\mu\mu}(\epsilon)$ is a *self-consistently* determined function of T_L , T_R , V_L , and V_R .

The system generates power $P_{\text{gen}} = (V_R - V_L)I_L$, so

$$P_{\text{gen}} = \frac{1}{h} \sum_{\mu} \int_0^{\infty} d\epsilon \mu e^-(V_R - V_L) \mathcal{T}_{RL}^{\mu\mu}(\epsilon) [f_L^\mu(\epsilon) - f_R^\mu(\epsilon)]. \quad (16)$$

It is easy to verify that Eqs. (13)–(16) satisfy the first law of thermodynamics [Eq. (9)]. This theory assumes the quantum system to be relaxation free, although decoherence is allowed as it does not change the structure of Eqs. (13)–(16). Relaxation is discussed in Sec. XV.

We define the voltage drop as $V = V_R - V_L$. Without loss of generality, we take the reference chemical potential to be that of reservoir L , so

$$V_L = 0, \quad V_R = V, \quad (17)$$

then J_L and P_{gen} coincide with Eqs. (8) and (9) in Ref. [14].

Numerous works have found the properties of thermoelectric systems from their transmission functions $\mathcal{T}_{RL}(\epsilon)$. Linear-response examples include Refs. [5,46–48,55–66], while nonlinear responses were considered in Refs. [17,18,36,37,54,67–71] (see Refs. [2–4] for recent reviews). However, here we do not ask what is the efficiency of a given system, we ask what is the system that would achieve the highest efficiency, and what is this efficiency? This is similar in spirit to Ref. [21], except that we maximize the efficiency for given power output.

We need to answer this question in the context of the mean-field treatment of electron-electron interactions [51], in which the transmission function for any given system is the solution of the above-mentioned self-consistency procedure. Despite this complexity, any transmission function (including all mean-field interactions) must obey

$$0 \leq \mathcal{T}_{RL}^{\mu\mu}(\epsilon) \leq N \quad \text{for all } \epsilon, \quad (18)$$

where N is the number of transverse modes at the narrowest point in the nanostructure (see Fig. 1). Let us assume that this is the *only* constraint on the transmission function. Let us assume that for any given T_L , T_R , and V , a clever physicist could engineer any desired transmission function, so long as it obeys Eq. (18). Presumably, they could do this either by solving the self-consistency equations for $\mathcal{T}_{RL}^{\mu\mu}(\epsilon)$ or by experimental

trial and error. Thus, in this work, we find the $\mathcal{T}_{RL}^{\mu\mu}(\epsilon)$ which maximizes the efficiency given solely the constraint in Eq. (18), and get this maximum efficiency. We then rely on future physicists to find a way to construct a system with this $\mathcal{T}_{RL}^{\mu\mu}(\epsilon)$ (although some hints are given in Sec. XII).

VI. FROM THERMOELECTRIC OPTIMIZATION TO THERMOCOUPLE OPTIMIZATION

The rest of this paper considers optimizing a single thermoelectric. However, an optimal thermocouple heat engine (or refrigerator) consists of two systems with opposite thermoelectric responses (full and open circles in Fig. 1). So, here we explain how to get the optimal thermocouple from the optimal thermoelectric.

Suppose the optimal system between L and R (the full circle) has a given transmission function $\mathcal{T}_{RL}^{\mu,\mu}(\epsilon)$, which we will find in Sec. IX. This system generates an electron flow parallel to heat flow (so electric current is antiparallel to heat flow, implying a negative Peltier coefficient). The system between L and R' (the open circle) must have the opposite response. For this we interchange the role played by electrons and holes compared with $\mathcal{T}_{RL}^{\mu,\mu}(\epsilon)$, so the optimal system between L and R' has

$$\mathcal{T}_{R'L}^{\mu,\mu}(\epsilon) = \mathcal{T}_{RL}^{\mu,-\mu}(\epsilon). \quad (19)$$

If the optimal bias for the system between L and R is V (which we will also find in Sec. IX), then the optimal bias for the system between L and R' is $-V$. Then, the heat flow from reservoir L into R' equals that from L into R , while the electrical current from L into R' is opposite to that from L into R , and so P_{gen} is the same for each thermoelectric. The load across the thermocouple (the two thermoelectrics) must be chosen such that the bias across the thermocouple is $2V$. The condition that the charge current out of L equals that into L will then ensure that both thermoelectrics are at their optimal bias.

In the rest of this paper we discuss power output and heat input *per thermoelectric*. For a thermocouple, one simply needs to multiply these by two, so the efficiency is unchanged but the power output is doubled.

VII. SIMPLE ESTIMATE OF BOUNDS ON POWER OUTPUT

One of the principal results of Ref. [14] is the quantum bounds on the power output of heat engines and refrigerators. The exact derivation of these bounds is given in Secs. IX A and XI A. Here, we give simple arguments for their basic form based on Pendry’s limit of heat flow discussed in Sec. II D.

For a refrigerator, it is natural to argue that the upper bound on cooling power will be closely related to Pendry’s bound [Eq. (6)]. We will show in Sec. XI A that this is the case. A two-lead thermoelectric can extract as much as half of J_L^{qb} . In other words, the cooling power of any refrigerator must obey

$$J_L \leq \frac{1}{2} J_L^{\text{qb}} = \frac{\pi^2}{12h} N k_B^2 T_L^2. \quad (20)$$

Now, let us turn to a heat engine operating between a hot reservoir L and cold reservoir R . Following Pendry’s logic, we can expect that the heat current into the quantum system from

reservoir L cannot be more than $J_L^{\text{overestimate}} = \frac{\pi^2}{6h} N k_B^2 (T_L^2 - T_R^2)$. Similarly, no heat engine can exceed Carnot's efficiency [Eq. (2)]. Thus, we can safely assume any system's power output is less than

$$P_{\text{gen}}^{\text{overestimate}} = \eta_{\text{eng}}^{\text{Carnot}} J_L^{\text{overestimate}} = \frac{\pi^2 N k_B^2 (T_L + T_R)(T_L - T_R)^2}{6h T_L}. \quad (21)$$

We know this is a significant overestimate because maximal heat flow cannot coincide with Carnot efficiency. Maximum heat flow requires $\mathcal{T}_{RL}^{\mu\mu}(\epsilon)$ is maximal for all ϵ and μ , while Carnot efficiency requires a δ -function-like dependence on ϵ (see Sec. II B). Nonetheless, the full calculation in Sec. IX A shows that the true quantum bound on power output is such that [72]

$$P_{\text{gen}} \leq P_{\text{gen}}^{\text{qb2}} \equiv A_0 \frac{\pi^2}{h} N k_B^2 (T_L - T_R)^2, \quad (22)$$

where $A_0 \simeq 0.0321$. Thus, the simple overestimate of the bound $P_{\text{gen}}^{\text{overestimate}}$ differs from the true bound $P_{\text{gen}}^{\text{qb2}}$ by a factor of $(1 + T_R/T_L)/(6A_0)$. In other words, it overestimates the quantum bound by a factor between 5.19 and 10.38 (that is, 5.19 when $T_R = 0$ and 10.38 when $T_R = T_L$). This is not bad for such a simple estimate.

VIII. GUESSING THE OPTIMAL TRANSMISSION FOR A HEAT ENGINE

Here, we use simple arguments to guess the transmission function which will maximize a heat engine's efficiency for a given power output. We consider the flow of electrons from reservoir L to reservoir R [the filled circle Fig. 1(a), remembering $e^- < 0$, so electron flow is in the opposite direction to I]. To produce power, the electrical current must flow against a bias, so we require e^-V to be positive, with V as in Eq. (17). Inspection of the integrand of Eq. (16) shows that it only gives positive contributions to the power output P_{gen} , when $\mu[f_L^\mu(\epsilon) - f_R^\mu(\epsilon)] > 0$. From Eq. (15), one can show that $f_L^\mu(\epsilon)$ and $f_R^\mu(\epsilon)$ cross at

$$\epsilon_0 = \mu e^-V / (1 - T_R/T_L) \quad (23)$$

(see Fig. 4). Since e^-V is positive, we maximize the power output by blocking the transmission of those electrons ($\mu = 1$) which have $\epsilon < \epsilon_0$, and blocking the transmission all holes ($\mu = -1$). For $\mu = 1$, all energies above ϵ_0 add to the power output. Hence, maximizing transmission for all $\epsilon > \epsilon_0$ will maximize the power output, giving $P_{\text{gen}} = P_{\text{gen}}^{\text{qb}}$. However, a

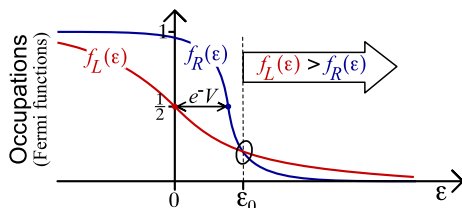


FIG. 4. (Color online) Sketch of Fermi functions $f_L^\mu(\epsilon)$ and $f_R^\mu(\epsilon)$ in Eq. (15), when μe^-V is positive, and $T_L > T_R$. Equation (23) gives the point where the two curves cross ϵ_0 .

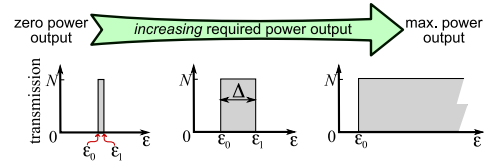


FIG. 5. (Color online) How the optimal “boxcar” transmission changes with increasing required power output. At maximum power output, a heat engine has $\epsilon_1 = \infty$ while ϵ_0 remains finite. At maximum cooling power, a refrigerator has $\epsilon_1 = \infty$ and $\epsilon_0 = 0$. The qualitative features follow this sketch for all T_R/T_L , however, the details depend on T_R/T_L (see Fig. 8).

detailed calculation, such as that in Sec. IX, is required to find the V which will maximize P_{gen} , remembering that P_{gen} depends directly on V as well as indirectly (via the above choice of ϵ_0).

Now, we consider maximizing the efficiency at a given power output P_{gen} , where $P_{\text{gen}} < P_{\text{gen}}^{\text{qb}}$. Comparing the integrands in Eqs. (13) and (16), we see that J_L contains an extra factor of energy ϵ compared to P_{gen} . As a result, the transmission of electrons ($\mu = 1$) with large ϵ enhances the heat current much more than it enhances the power output. This means that the higher an electron's ϵ is, the less efficiently it contributes to power production. Thus, one would guess that it is optimal to have an upper cutoff on transmission ϵ_1 , which would be just high enough to ensure the desired power output P_{gen} , but no higher. Then, the transmission function will look like a “band-pass filter” (the “boxcar” form in Fig. 5), with ϵ_0 and ϵ_1 further apart for higher power outputs. This guess is correct, however, the choice of V affects both ϵ_0 and ϵ_1 , so the calculation in Sec. IX is necessary to find the V , ϵ_0 , and ϵ_1 which maximize the efficiency for given P_{gen} .

IX. MAXIMIZING HEAT-ENGINE EFFICIENCY FOR GIVEN POWER OUTPUT

Now, we present the central calculations of this paper, finding the maximum efficiency of a quantum thermoelectric with *given* power output. In this section, we consider heat engines, while Sec. XI addresses refrigerators.

For a heat engine, our objective is to find the transmission function $\mathcal{T}_{RL}^{\mu\mu}(\epsilon)$ and bias V that maximize the efficiency $\eta_{\text{eng}}(P_{\text{gen}})$ for given power output P_{gen} . To do this, we treat $\mathcal{T}_{RL}^{\mu\mu}(\epsilon)$ as a set of many slices each of width $\delta \rightarrow 0$ [see the sketch in Fig. 6(a)]. We define τ_γ^μ as the height of the γ th slice, which is at energy $\epsilon_\gamma \equiv \gamma\delta$. Our objective is to find

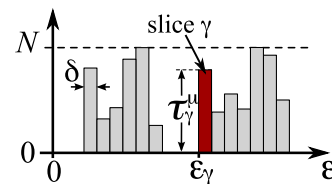


FIG. 6. (Color online) A completely arbitrary transmission function $\mathcal{T}_{RL}^{\mu\mu}(\epsilon)$ (see Sec. IX). We take it to have infinitely many slices of width $\delta \rightarrow 0$, so slice γ has energy $\epsilon_\gamma \equiv \gamma\delta$ and height τ_γ^μ . We find the optimal height for each slice.

the optimal value of τ_γ^μ for each μ, γ , and optimal values of the bias V , all under the constraint of fixed P_{gen} . Often, such optimization problems are formidable, however, this one is fairly straightforward.

The efficiency is maximum for a fixed power P_{gen} if J_L is minimum for that P_{gen} . If we make an infinitesimal change of τ_γ^μ and V , we note that

$$\delta P_{\text{gen}} = \left. \frac{\partial P_{\text{gen}}}{\partial \tau_\gamma^\mu} \right|_V \delta \tau_\gamma^\mu + P'_{\text{gen}} \delta V, \quad (24)$$

$$\delta J_L = \left. \frac{\partial J_L}{\partial \tau_\gamma^\mu} \right|_V \delta \tau_\gamma^\mu + J'_L \delta V, \quad (25)$$

where $|_x$ indicates that the derivative is taken at constant x , and the primed indicates $\partial/\partial V$ for fixed transmission functions. If we want to fix P_{gen} as we change τ_γ^μ , we must change the bias V to compensate. For this, we set $\delta P_{\text{gen}} = 0$ in Eq. (24) and substitute the result for δV into Eq. (25). Then, J_L decreases (increasing efficiency) for an infinitesimal increase of τ_γ^μ at fixed P_{gen} if

$$\left. \frac{\partial J_L}{\partial \tau_\gamma^\mu} \right|_{P_{\text{gen}}} = \left. \frac{\partial J_L}{\partial \tau_\gamma^\mu} \right|_V - \frac{J'_L}{P'_{\text{gen}}} \left. \frac{\partial P_{\text{gen}}}{\partial \tau_\gamma^\mu} \right|_V < 0. \quad (26)$$

Comparing Eqs. (13) and (16), one sees that

$$\left. \frac{\partial J_L}{\partial \tau_\gamma^\mu} \right|_V = \frac{\epsilon_\gamma}{\mu e^{-V}} \left. \frac{\partial P_{\text{gen}}}{\partial \tau_\gamma^\mu} \right|_V, \quad (27)$$

where V is given in Eq. (17). Thus, the efficiency $\eta_{\text{eng}}(P_{\text{gen}})$ grows with a small increase of τ_γ^μ if

$$\left(\epsilon_\gamma - \mu e^{-V} \frac{J'_L}{P'_{\text{gen}}} \right) \left. \frac{\partial P_{\text{gen}}}{\partial \tau_\gamma^\mu} \right|_V < 0, \quad (28)$$

where $P_{\text{gen}}, P'_{\text{gen}}, J_L, J'_L$, and e^{-V} are positive.

For what follows, let us define two energies

$$\epsilon_0 = e^{-V}/(1 - T_R/T_L), \quad (29)$$

$$\epsilon_1 = e^{-V} J'_L/P'_{\text{gen}}. \quad (30)$$

One can see that $(\partial P_{\text{gen}}/\partial \tau_\gamma^\mu)|_V > 0$ when both $\mu = 1$ and $\epsilon > \epsilon_0$, and is negative otherwise. Thus, for $\mu = 1$, Eq. (28) is satisfied when ϵ_γ is between ϵ_0 and ϵ_1 . For $\mu = -1$, Eq. (28) is never satisfied.

A heat engine is only useful if $P_{\text{gen}} > 0$, and this is only true for $\epsilon_0 < \epsilon_1$. Hence, if $\mu = 1$ and $\epsilon_0 < \epsilon < \epsilon_1$, then $\eta_{\text{eng}}(P_{\text{gen}})$ is maximum for τ_γ^μ at its maximum value $\tau_\gamma^\mu = N$. For all other μ and ϵ_γ , $\eta_{\text{eng}}(P_{\text{gen}})$ is maximum for τ_γ^μ at its minimum value $\tau_\gamma^\mu = 0$. Since the left-hand side of Eq. (28) is not zero for any $\epsilon_\gamma \neq \epsilon_0, \epsilon_1$, there are no stationary points, which is why τ_γ^μ never takes a value between its maximum and minimum values. Thus, the optimal $\mathcal{J}_{RL}^{\mu\mu}(\epsilon)$ is a ‘‘boxcar’’ or ‘‘top-hat’’ function

$$\mathcal{J}_{RL}^{\mu\mu}(\epsilon) = \begin{cases} N & \text{for } \mu = 1 \text{ and } \epsilon_0 < \epsilon < \epsilon_1, \\ 0 & \text{otherwise} \end{cases} \quad (31)$$

[see Fig. 6(b)]. It hence acts as a band-pass filter, only allowing flow between L and R for electrons ($\mu = 1$) in the energy window between ϵ_0 to ϵ_1 .

Substituting a boxcar transmission function with arbitrary ϵ_0 and ϵ_1 into Eqs. (13) and (16) gives

$$J_L = N [F_L(\epsilon_0) - F_R(\epsilon_0) - F_L(\epsilon_1) + F_R(\epsilon_1)], \quad (32)$$

$$P_{\text{gen}} = N e^{-V} [G_L(\epsilon_0) - G_R(\epsilon_0) - G_L(\epsilon_1) + G_R(\epsilon_1)], \quad (33)$$

where we define

$$F_j(\epsilon) = \frac{1}{h} \int_\epsilon^\infty \frac{x dx}{1 + \exp[(x - e^{-V_j})/(k_B T_j)]}, \quad (34)$$

$$G_j(\epsilon) = \frac{1}{h} \int_\epsilon^\infty \frac{dx}{1 + \exp[(x - e^{-V_j})/(k_B T_j)]}, \quad (35)$$

which are both positive for any $\epsilon > 0$. Remembering that we took $V_L = 0$ and $V_R = V$, these integrals are

$$F_L(\epsilon) = \epsilon G_L(\epsilon) - \frac{(k_B T_L)^2}{h} \text{Li}_2[-e^{-\epsilon/(k_B T_L)}], \quad (36)$$

$$F_R(\epsilon) = \epsilon G_R(\epsilon) - \frac{(k_B T_R)^2}{h} \text{Li}_2[-e^{-(\epsilon - e^{-V})/(k_B T_R)}], \quad (37)$$

$$G_L(\epsilon) = \frac{k_B T_L}{h} \ln[1 + e^{-\epsilon/(k_B T_L)}], \quad (38)$$

$$G_R(\epsilon) = \frac{k_B T_R}{h} \ln[1 + e^{-(\epsilon - e^{-V})/(k_B T_R)}], \quad (39)$$

for dilogarithm function $\text{Li}_2(z) = \int_0^\infty t dt/(e^t/z - 1)$.

We are only interested in cases where ϵ_0 fulfills the condition in Eq. (29), in this case $(\epsilon_0 - e^{-V})/(k_B T_R) = \epsilon_0/(k_B T_L)$, which means $G_R(\epsilon_0)$ and $F_R(\epsilon_0)$ are related to $G_L(\epsilon_0)$ and $F_L(\epsilon_0)$ by

$$G_R(\epsilon_0) = \frac{T_R}{T_L} G_L(\epsilon_0), \quad (40)$$

$$F_R(\epsilon_0) - \epsilon_0 G_R(\epsilon_0) = \frac{T_R^2}{T_L^2} [F_L(\epsilon_0) - \epsilon_0 G_L(\epsilon_0)]. \quad (41)$$

Equation (30) tells us that ϵ_1 depends on J_L and P_{gen} , but that these depend in turn on ϵ_1 . Hence, to find ϵ_1 , we substitute Eqs. (32) and (33) into Eq. (30) to get a transcendental equation for ϵ_1 as a function of V for given T_R/T_L . This equation is too hard to solve analytically (except in the high and low power limits, discussed in Secs. IX A and IX B, respectively). The red curve in Fig. 7(a) is a numerical solution for $T_R/T_L = 0.2$.

Having found ϵ_1 as a function of V for given T_R/T_L , we can use Eqs. (32) and (33) to get $J_L(V)$ and $P_{\text{gen}}(V)$. We can then invert the second relation to get $V(P_{\text{gen}})$. At this point, we can find $J_L(P_{\text{gen}})$, and then use Eq. (1) to get the quantity that we desire: the maximum efficiency at given power output $\eta_{\text{eng}}(P_{\text{gen}})$.

In Sec. IX A, we do this procedure analytically for high power ($P_{\text{gen}} = P_{\text{gen}}^{\text{qb}2}$), and in Sec. IX B, we do this procedure analytically for low power ($P_{\text{gen}} \ll P_{\text{gen}}^{\text{qb}2}$). For other cases, we only have a numerical solution for the transcendental equation for ϵ_1 as a function of $V, T_R/T_L$, so we must do everything numerically.

Figure 8(a) gives the values of $\Delta = (\epsilon_1 - \epsilon_0)$ and e^{-V} which result from solving the transcendental equation

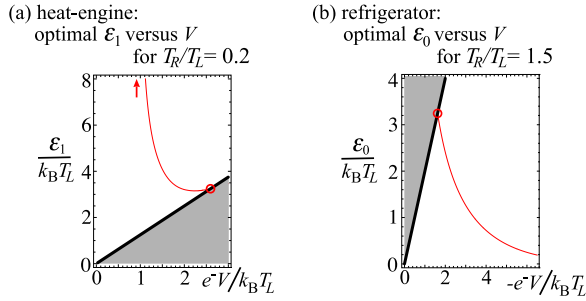


FIG. 7. (Color online) Solutions of the transcendental equations giving optimal ϵ_1 (heat engine) or ϵ_0 (refrigerator). In (a), the red curve is the optimal $\epsilon_1(V)$ for $\epsilon_1 > \epsilon_0$, and the thick black line is ϵ_0 in Eq. (29). The red circle and red arrow indicate the low and high power limits discussed in the text. In (b), the red curve is the optimal $\epsilon_0(V)$ for $\epsilon_0 < \epsilon_1$, and the thick black line is ϵ_1 in Eq. (57).

numerically for a variety of different T_R/T_L . Equation (29) then relates ϵ_0 to e^{-V} . The qualitative behavior of the resulting boxcar transmission function is shown in Fig. 5. This numerical evaluation enables us to find the efficiency as a function of P_{gen} and T_R/T_L , which we plot in Fig. 9(a).

A. Quantum bound on heat-engine power output

Here, we want to find the highest possible power output of the heat engine. In the previous section, we had the power as a function of two independent parameters V and

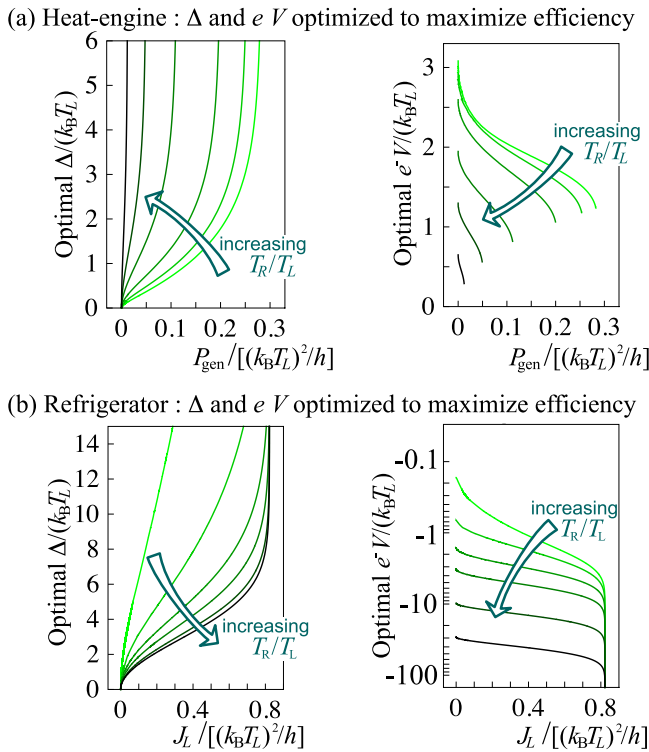
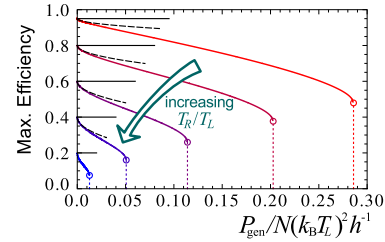


FIG. 8. (Color online) (a) Plots of optimal Δ (left) and e^{-V} (right) for a heat engine with given power output P_{gen} for $T_R/T_L = 0.05, 0.1, 0.2, 0.4, 0.6$, and 0.8 . We get ϵ_0 from e^{-V} by using Eq. (29). (b) Plots of optimal Δ (left) and e^{-V} (right) for a refrigerator with a given cooling power output J_L for $T_R/T_L = 1.05, 1.2, 1.5, 2, 4$, and 10 . We get ϵ_1 from e^{-V} by using Eq. (57).

(a) Heat-engine: maximum efficiency versus power output



(b) Refrigerator: maximum efficiency versus cooling power

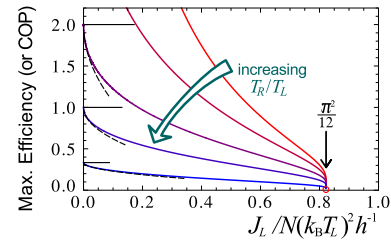


FIG. 9. (Color online) Efficiencies of (a) heat engines and (b) refrigerators. In (a) the curves are the maximum allowed heat-engine efficiency as a function of power outputs for $T_R/T_L = 0.05, 0.2, 0.4, 0.6, 0.8$ (from top to bottom). In (b) the curves are the maximum allowed refrigerator efficiency as a function of cooling power for $T_R/T_L = 1.05, 1.2, 1.5, 2, 4$ (from top to bottom). In both (a) and (b) the horizontal black lines indicate Carnot efficiency for each T_R/T_L , while the dashed black curves are the analytic theory for small cooling power, given in Eq. (51) or (63). The circles mark the analytic result for maximum power output.

ϵ_1 , with ϵ_0 given by Eq. (29). However, we know that Eq. (30) will then determine a line in this two-dimensional parameter space [Fig. 7(a)], which we can parametrize by the parameter V . The maximum possible power corresponds to $P'_{\text{gen}} = 0$, where we recall $P'_{\text{gen}} \equiv dP_{\text{gen}}/dV$. This has two consequences: the first is that from Eq. (29), we see that $P'_{\text{gen}} = 0$ means that $\epsilon_1 \rightarrow \infty$. Thus, the transmission function $\mathcal{T}_{RL}^{\mu\mu}(\epsilon)$, taking the form of a Heaviside step function $\theta(\epsilon - \epsilon_0)$, where ϵ_0 is given in Eq. (29). Taking Eq. (33) combined with Eq. (40) for $\epsilon_1 \rightarrow \infty$ gives

$$P_{\text{gen}}(\epsilon_1 \rightarrow \infty) = N e^{-V} \left(1 - \frac{T_R}{T_L}\right) G_L \left(\frac{e^{-V}}{1 - T_R/T_L}\right).$$

The second consequence of $P'_{\text{gen}} = 0$ is that the V derivative of this expression must be zero. This gives us the condition that

$$(1 + B_0) \ln[1 + B_0] + B_0 \ln[B_0] = 0, \quad (42)$$

where we define $B_0 = \exp[-e^{-V}/(k_B T_L - k_B T_R)] = \exp[-\epsilon_0/(k_B T_L)]$. Numerically solving this equation gives $B_0 \simeq 0.318$. Equation (29) means that this corresponds to $e^{-V} = -k_B(T_L - T_R) \ln[0.318] = 1.146 k_B(T_L - T_R)$, indicated by the red arrow in Fig. 7(a). Substituting this back into $P_{\text{gen}}(\epsilon_1 \rightarrow \infty)$ gives the maximum achievable value of P_{gen} :

$$P_{\text{gen}}^{\text{qb2}} = A_0 \frac{\pi^2}{h} N k_B^2 (T_L - T_R)^2 \quad (43)$$

with

$$A_0 \equiv B_0 \ln^2[B_0]/[\pi^2(1 + B_0)] \simeq 0.0321. \quad (44)$$

We refer to this as the quantum bound (qb) on power output [72] because of its origin in the Fermi wavelength of the electrons λ_F . We see this in the fact that $P_{\text{gen}}^{\text{qb}2}$ is proportional to the number of transverse modes in the quantum system N , which is given by the cross-sectional area of the quantum system divided by λ_F^2 . This quantity has no analog in classical thermodynamics.

The efficiency at this maximum power $P_{\text{gen}}^{\text{qb}2}$ is

$$\eta_{\text{eng}}(P_{\text{gen}}^{\text{qb}2}) = \eta_{\text{eng}}^{\text{Carnot}}/[1 + C_0(1 + T_R/T_L)], \quad (45)$$

with

$$C_0 = -(1 + B_0)\text{Li}_2(-B_0)/(B_0 \ln^2[B_0]) \simeq 0.936. \quad (46)$$

As such, it varies with T_R/T_L , but is always more than $0.3 \eta_{\text{eng}}^{\text{Carnot}}$. This efficiency is less than Curzon and Ahlborn's efficiency for all T_R/T_L (although not much less). However, the power output here is infinitely larger than the maximum power output of systems that achieve Curzon and Ahlborn's efficiency (see Sec. II C).

The form of Eq. (45) is very different from Curzon and Ahlborn's efficiency. However, we note in passing that Eq. (45) can easily be written as $\eta_{\text{eng}}(P_{\text{gen}}^{\text{qb}2}) = \eta_{\text{eng}}^{\text{Carnot}}/[(1 + 2C_0) - C_0 \eta_{\text{eng}}^{\text{Carnot}}]$, which is reminiscent of the efficiency at maximum power found for very different systems (certain classical stochastic heat engines) in Eq. (31) of Ref. [73].

B. Optimal heat engine at low power output

Now, we turn to the opposite limit, that of low power output $P_{\text{gen}} \ll P_{\text{gen}}^{\text{qb}2}$, where we expect the maximum efficiency to be close to Carnot efficiency. In this limit, ϵ_1 is close to ϵ_0 . Defining $\Delta = \epsilon_1 - \epsilon_0$, we expand Eqs. (32) and (33) in small Δ up to order Δ^3 . This gives

$$J_L = \frac{P_{\text{gen}}}{1 - T_R/T_L} + \frac{N \Delta^3 (1 - T_R/T_L)}{3h k_B T_R} g(x_0), \quad (47)$$

$$P_{\text{gen}} = \frac{N \epsilon_0 \Delta^2 (1 - T_R/T_L)^2}{2h k_B T_R} \times \left[g(x_0) + \frac{\Delta (1 + T_R/T_L)}{3 k_B T_R} \frac{dg(x_0)}{dx_0} \right], \quad (48)$$

where Eq. (29) was used to write e^{-V} in terms of ϵ_0 , and we defined $x_0 = \epsilon_0/(k_B T_L)$ and $g(x) = e^x/(1 + e^x)^2$. Thus, for small Δ we find that

$$\eta_{\text{eng}}(\Delta) = \eta_{\text{eng}}^{\text{Carnot}} \left(1 - \frac{2\Delta}{3x_0 k_B T_L} + \dots \right). \quad (49)$$

Equation (30) gives a transcendental equation for x_0 and Δ . However, Δ drops out when it is small, and the transcendental equation reduces to

$$x_0 \tanh[x_0/2] = 3, \quad (50)$$

for which $x_0 \equiv \epsilon_0/(k_B T_L) \simeq 3.24$. Equation (29) means that this corresponds to $e^{-V} = 3.24 k_B (T_L - T_R)$, indicated by the circle in Fig. 7(a). Now, we can use Eq. (48) to lowest order in Δ to rewrite Eq. (49) in terms of P_{gen} . This gives the efficiency

for small P_{gen} as

$$\eta_{\text{eng}}(P_{\text{gen}}) = \eta_{\text{eng}}^{\text{Carnot}} \left(1 - 0.478 \sqrt{\frac{T_R}{T_L} \frac{P_{\text{gen}}}{P_{\text{gen}}^{\text{qb}2}}} + \dots \right), \quad (51)$$

where the ellipsis indicates terms of order $(P_{\text{gen}}/P_{\text{gen}}^{\text{qb}2})$ or higher. Equation (11) then gives the minimum rate of entropy production at power output P_{gen} :

$$\dot{S}(P_{\text{gen}}) = 0.478 \frac{P_{\text{gen}}^{\text{qb}2}}{\sqrt{T_L T_R}} \left(\frac{P_{\text{gen}}}{P_{\text{gen}}^{\text{qb}2}} \right)^{3/2} + \mathcal{O}[P_{\text{gen}}^2]. \quad (52)$$

Thus, the maximal efficiency at small P_{gen} is that of Carnot minus a term that grows like $P_{\text{gen}}^{1/2}$ [dashed curves in Fig. 9(a)], and the associated minimal rate of entropy production goes like $P_{\text{gen}}^{3/2}$.

Note that Eq. (49) shows that Carnot efficiency occurs at any x_0 (i.e., any ϵ_0) when Δ is strictly zero (and so P_{gen} is strictly zero). However, for arbitrary x_0 the factor 0.478 in Eq. (51) is replaced by $\sqrt{8\pi^2 A_0/[9x_0^3 g(x_0)]}$. The value of x_0 that satisfied Eq. (50) is exactly the one which minimizes this prefactor (its minimum being 0.478), and thus maximizes the efficiency for any small but finite P_{gen} .

X. GUESSING THE OPTIMAL TRANSMISSION FOR A REFRIGERATOR

Here, we use simple arguments to guess the transmission function which maximizes a refrigerator's efficiency for given cooling power. The arguments are similar to those for heat engines (Sec. VIII), although some crucial differences will appear.

We consider the flow of electrons from reservoir L to reservoir R [the filled circle in Fig. 1(a), remembering $e^- < 0$ so electrons flow in the opposite direction to I]. To refrigerate, the thermoelectric must absorb power, so the electrical current must be due to a bias; this requires e^{-V} to be negative, with V as in Eq. (17).

Inspection of the integrand of Eq. (13) shows that it only gives positive contributions to the cooling power output J_L , when $[f_L^\mu(\epsilon) - f_R^\mu(\epsilon)] > 0$. Since $T_L < T_R$ and $e^{-V} < 0$, we can use Eq. (15) to show that this is never true for holes ($\mu = -1$), and is only true for electrons ($\mu = 1$) with energies $\epsilon < \epsilon_1$, where

$$\epsilon_1 = -e^{-V}/(T_R/T_L - 1). \quad (53)$$

Thus, it is counterproductive to allow the transmission of electrons with $\epsilon > \epsilon_1$, or the transmission of any holes. Note that this argument gives us an *upper* cutoff on electron transmission energies, despite the fact it gave a *lower* cutoff for the heat engine [see Eq. (23) and the text around it]. All electron ($\mu = 1$) energies from zero to ϵ_1 contribute positively to the cooling power J_L . To maximize the cooling power, one needs to maximize $[f_L^\mu(\epsilon) - f_R^\mu(\epsilon)]$; this is done by taking $e^{-V} \rightarrow -\infty$, for which $\epsilon_1 \rightarrow \infty$. This logic gives the maximum cooling power, which Sec. XI will show equals $\frac{1}{2} J_L^{\text{qb}}$.

Now, we consider maximizing the efficiency at a given cooling output J_L , when $J_L < \frac{1}{2} J_L^{\text{qb}}$. Comparing the integrands in Eqs. (13) and (16), we see that the extra factor of ϵ in

J_L means that allowing the transmission of electrons at low energies has a small effect on cooling power, while costing a similar electrical power at higher energies. Thus, it would seem to be optimal to have a lower cutoff on transmission ϵ_0 , which would be just low enough to ensure the desired cooling power J_L , but no lower. Then, the transmission function will act as a “band-pass filter” (the “boxcar” in Fig. 5), with ϵ_0 and ϵ_1 further apart for higher cooling power. This is correct, however, the choice of V affects ϵ_0 and ϵ_1 , so the calculation in Sec. XI is necessary to find the V , ϵ_0 , and ϵ_1 which maximize the efficiency for cooling power J_L .

XI. MAXIMIZING REFRIGERATOR EFFICIENCY FOR GIVEN COOLING POWER

Here, we find the maximum refrigerator efficiency, also called the coefficient of performance (COP), for given cooling power J_L . The method is very similar to that for heat engines, and here we mainly summarize the differences. The refrigerator efficiency increases for a fixed cooling power J_L if the electrical power absorbed $P_{\text{abs}} = -P_{\text{gen}}$ decreases for fixed J_L . This is so if

$$\left. \frac{\partial P_{\text{abs}}}{\partial \tau_\gamma^\mu} \right|_{J_L} = \left. \frac{\partial P_{\text{abs}}}{\partial \tau_\gamma^\mu} \right|_V - \frac{P'_{\text{abs}}}{J'_L} \left. \frac{\partial J_L}{\partial \tau_\gamma^\mu} \right|_V < 0, \quad (54)$$

where we recall that the primed means (d/dV) . This is nothing but Eq. (26) with $J_L \rightarrow P_{\text{abs}}$ and $P_{\text{gen}} \rightarrow J_L$. Using Eq. (27), we see that $\eta_{\text{fri}}(J_L)$ grows with τ_γ^μ for

$$\left(\frac{-\mu e^{-V}}{\epsilon_\gamma} - \frac{P'_{\text{abs}}}{J'_L} \right) \left. \frac{\partial J_L}{\partial \tau_\gamma^\mu} \right|_V < 0, \quad (55)$$

where P_{abs} , P'_{abs} , J_L , J'_L , and $-e^{-V}$ are all positive.

To proceed, we define the following energies:

$$\epsilon_0 = -e^{-V} J'_L / P'_{\text{abs}}, \quad (56)$$

$$\epsilon_1 = -e^{-V} / (T_R / T_L - 1). \quad (57)$$

Then, one can see that $(\partial J_L / \partial \tau_\gamma^\mu)|_V$ is positive when both $\mu = 1$ and $\epsilon < \epsilon_1^{\text{fri}}$, and is negative otherwise. Thus, for $\mu = -1$, Eq. (55) is never satisfied. For $\mu = 1$, Eq. (55) is satisfied when ϵ_γ is between ϵ_0^{fri} and ϵ_1^{fri} . A refrigerator is only useful if $J_L > 0$ (i.e., it removes heat from the cold reservoir), and this is only true for $\epsilon_0^{\text{fri}} < \epsilon_1^{\text{fri}}$. Hence, if $\mu = 1$ and $\epsilon_0^{\text{fri}} < \epsilon < \epsilon_1^{\text{fri}}$, then $\eta_{\text{fri}}(J_L)$ grows upon increasing τ_γ^μ . Thus, the optimum is when such $\tau_\gamma^\mu = N$. For all other μ and ϵ_γ , $\eta_{\text{fri}}(J_L)$ grows upon decreasing τ_γ^μ . Thus, the optimum is when such $\tau_\gamma^\mu = 0$. This gives the boxcar transmission function in Eq. (31), with ϵ_0 and ϵ_1 given by Eqs. (56) and (57). Comparing with Eqs. (29) and (30), we see these energies are the opposite way around for a refrigerator compared to a heat engine (up to a minus sign).

Substituting Eqs. (32) and (33) into Eq. (56), one gets a transcendental equation for ϵ_0 as a function of V for given T_R / T_L . This equation is too hard to solve analytically (except in the high and low power limits, discussed in Secs. XI A and XI B). The red curve in Fig. 7(b) is a numerical solution for $T_R / T_L = 1.5$.

Having found ϵ_0 as a function of V for given T_R / T_L , we can use Eqs. (32) and (33) to get $J_L(V)$ and $P_{\text{abs}}(V) = -P_{\text{gen}}(V)$. We can then invert the first relation to get $V(J_L)$. Now, we can

find $P_{\text{abs}}(J_L)$, and then use Eq. (3) to get the quantity that we desire: the maximum efficiency (or COP) $\eta_{\text{fri}}(J_L)$ at cooling power J_L .

Figure 8(b) gives the values of $\Delta = (\epsilon_1 - \epsilon_0)$ and e^{-V} which result from solving the transcendental equation numerically. As noted, ϵ_1 is related to e^{-V} by Eq. (57). The qualitative behavior of the resulting boxcar transmission function is sketched in Fig. 5. This numerical evaluation enables us to find efficiency as a function of J_L and T_R / T_L , which we plot in Fig. 9(b).

A. Quantum bound on refrigerator cooling power

To find the maximum allowed cooling power J_L , we look for the place where $J'_L = 0$. From Eq. (56) we see that this immediately implies $\epsilon_0 = 0$. Taking Eq. (32) with $\epsilon_0 = 0$, we note by using Eq. (34) that $F_L(0) - F_R(0)$ grows monotonically as one takes $-e^{-V} \rightarrow \infty$. Similarly, for ϵ_1 given by Eq. (30), we note by using Eq. (34) and $T_R > T_L$ that $F_R(\epsilon_1) - F_L(\epsilon_1)$ grows monotonically as one takes $-e^{-V} \rightarrow \infty$. Thus, we can conclude that J_L is maximal for $-e^{-V} \rightarrow \infty$, which implies $\epsilon_1 \rightarrow \infty$ via Eq. (57). Physically, this corresponds to all electrons arriving at the quantum system from reservoir L being transmitted into reservoir R , but all holes arriving from reservoir L being reflected back into reservoir L . At the same time, reservoir R is so strongly biased that it has no electrons with $\epsilon > 0$ (i.e., no electrons above reservoir L 's chemical potential) to carry heat from R to L .

In this limit, $F_L(\epsilon_1) = F_L(\epsilon_1) = F_R(\epsilon_0) = 0$, so the maximal refrigerator cooling power is

$$J_L = \frac{\pi^2}{12h} N k_B^2 T_L^2, \quad (58)$$

where we used the fact that $\text{Li}_2[1] = \pi^2/12$. This is exactly half the quantum bound on heat current that can flow out of reservoir L given in Eq. (6). The quantum bound is achieved by coupling reservoir L to another reservoir with a temperature of absolute zero, through a contact with N transverse mode. By definition, a refrigerator is cooling reservoir L below the temperature of the other reservoirs around it. In doing so, we show its cooling power is always less than or equal to $J_L^{\text{qb}}/2$. However, it is intriguing that the maximum cooling power is independent of the temperature of the environment T_R of the reservoir being cooled (reservoir L). In short, the best refrigerator can remove all electrons (or all holes) that reach it from reservoir L , but it cannot remove all electrons *and* all holes at the same time.

It is easy to see that the efficiency of the refrigerator (COP) at this maximum possible cooling power is zero, simply because $|V| \rightarrow \infty$, so the power absorbed $P_{\text{abs}} \rightarrow \infty$. However, one gets exponentially close to this limit for $-e^{-V} \gg k_B T_R$, for which P_{abs} is large but finite, and so $\eta_{\text{fri}}(J_L)$ remains finite [see Fig. 9(b)].

B. Optimal refrigerator at low cooling power

Now, we turn to the opposite limit, that of low cooling power output $J_L \ll J_L^{\text{qb}}$, where we expect the maximum efficiency to be close to Carnot efficiency. In this limit, ϵ_0 is close to ϵ_1 .

Defining $\Delta = \epsilon_1 - \epsilon_0$, we expand Eqs. (32) and (33) in small Δ up to order Δ^3 . This gives

$$J_L = \frac{P_{\text{abs}}}{T_R/T_L - 1} - \frac{N \Delta^3 (T_R/T_L - 1)}{3h k_B T_R} g(x_1), \quad (59)$$

$$P_{\text{abs}} = \frac{N \epsilon_1 \Delta^2 (T_R/T_L - 1)^2}{2h k_B T_R} \times \left[g(x_1) - \frac{\Delta (T_R/T_L + 1)}{3 k_B T_R} \frac{dg(x_1)}{dx_1} \right], \quad (60)$$

where Eq. (57) was used to write e^{-V} in terms of ϵ_1 , and we define $x_1 = \epsilon_1/(k_B T_L)$ and $g(x) = e^x/(1 + e^x)^2$. Thus, for small Δ we find that the efficiency is

$$\eta_{\text{fri}}(\Delta) = \eta_{\text{fri}}^{\text{Carnot}} \left(1 - \frac{2\Delta}{3x_1 k_B T_L} + \dots \right). \quad (61)$$

Note that this is the same Eq. (49) for the heat engine at low power output, except that x_0 is replaced by x_1 , and the Carnot efficiency is that of the refrigerator rather than that of the heat engine.

Equation (56) gives a transcendental equation for x_1 and Δ . However, Δ drops out when it is small, and the transcendental equation reduces to

$$x_1 \tanh[x_1/2] = 3, \quad (62)$$

for which $x_1 \equiv \epsilon_1/(k_B T_L) = 3.2436\dots$. Again, this is the same as for a heat engine [Eq. (50)], but with x_1 replacing x_0 . Equation (57) means that this corresponds to $-e^{-V} = 3.2436 k_B (T_R - T_L)$, indicated by the circle in Fig. 7(b). Now, we can use Eq. (59) to lowest order in Δ to rewrite Eq. (61) in terms of J_L . This gives the efficiency (or coefficient of performance, COP) for small J_L as

$$\eta_{\text{fri}}(J_L) = \eta_{\text{fri}}^{\text{Carnot}} \left(1 - 1.09 \sqrt{\frac{T_R}{T_R - T_L} \frac{J_L}{J_L^{\text{qb}}}} + \dots \right), \quad (63)$$

where the ellipsis indicates terms of order (J_L/J_L^{qb}) or higher. Equation (12) gives the minimum rate of entropy generation at cooling power output J_L as

$$\dot{S}(J_L) = 1.09 \frac{J_L^{\text{qb}}}{T_L} \sqrt{1 - \frac{T_L}{T_R} \left(\frac{J_L}{J_L^{\text{qb}}} \right)^{3/2}} + \mathcal{O}[J_L^2]. \quad (64)$$

Thus, we conclude that the maximum efficiency at small J_L is that of Carnot minus a term that grows like $J_L^{1/2}$ [dashed curves in Fig. 9(b)], while the associated minimum entropy production goes like $J_L^{3/2}$.

We note that Carnot efficiency occurs at $J_L = 0$ at any $x_1 = \epsilon_1/(k_B T_L)$. However, then the 1.09 factor in Eq. (63) becomes $\sqrt{4\pi^2/[27x_1^3 g(x_1)]}$. The condition in Eq. (62) minimizes this factor (the minimum being 1.09), and thereby maximizes the efficiency for given J_L .

XII. IMPLEMENTATION WITH A CHAIN OF QUANTUM SYSTEMS

The previous sections have shown that maximum efficiency (at given power output) occurs when the thermoelectric

system has a boxcar transmission function with the right position and width. In the limit of maximum power, the boxcar becomes a Heaviside step function. Here, we give a detailed recipe for engineering such transmission functions for noninteracting electrons, and then discuss how to include mean-field interaction effects.

A Heaviside step function is easily implemented with point contact, whose transmission function is [74]

$$\mathcal{T}_{L,isl}(\epsilon) = \left\{ 1 + \exp \left[-\frac{\epsilon - E(V)}{D_{\text{tunnel}}} \right] \right\}^{-1}, \quad (65)$$

where $E(V)$ is the height of the energy barrier induced by the point contact, and D_{tunnel} is a measure of tunneling through the point contact. A sufficiently long point contact exhibits negligible tunneling $D_{\text{tunnel}} \rightarrow 0$, so the transmission function simplifies to the desired Heaviside step function $\theta[\epsilon - E(V)]$.

For a potential implementation of a boxcar function we consider a chain of sites (quantum dots or molecules) with one level per site, as sketched in Fig. 10(a). The objective is that the hoppings between sites $\{t_i\}$ will cause the states to hybridize to form a band centered at E_0 , with a width given by the hopping [75]. Neglecting electron-electron interactions, the hopping Hamiltonian for five sites in the chain ($k = 5$) can be written as

$$\mathcal{H}_{\text{chain}} = \begin{pmatrix} -ia_0/2 & t_1 & 0 & 0 & 0 \\ t_1 & 0 & t_2 & 0 & 0 \\ 0 & t_2 & 0 & t_3 & 0 \\ 0 & 0 & t_3 & 0 & t_4 \\ 0 & 0 & 0 & t_4 & -ia_0/2 \end{pmatrix}. \quad (66)$$

This is easily generalized to arbitrary chain length k . Here, we treat a_0 as a phenomenological parameter, however, in reality it would be given by $|t_0|^2$ multiplied by the density of states in the reservoir. The fact that particles escape from the chain into the reservoirs means the wave function for any given particle in the chain will decay with time. To model this, the Hamiltonian must be non-Hermitian, with the non-Hermiticity entering in the matrix elements for coupling to the reservoirs (top-left and bottom-right matrix elements). These

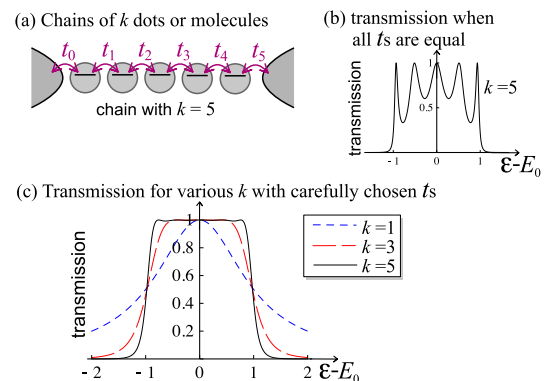


FIG. 10. (Color online) (a) A chain of single-level quantum dots with their energy levels aligned at energy E_0 . (b) Transmission function when all hoppings are equal (note the strong oscillations). (c) Transmission function when all hoppings are carefully chosen (see text). To aid comparison, all bandwidths in the plots have been normalized.

induce an imaginary contribution to each eigenstate's energy E_i , with the wave function of any eigenstate decaying at a rate given by the imaginary part of E_i . The non-Hermiticity of $\mathcal{H}_{\text{chain}}$ also means that its left and right eigenvectors are different, defining $|\psi_i^{(r)}\rangle$ as the i th right eigenvector of the matrix $\mathcal{H}_{\text{chain}}$, and $\langle\psi_i^{(l)}|$ as the i th left eigenvector, we have $\langle\psi_i^{(l)}|\psi_j^{(r)}\rangle = \delta_{ij}$ and $\langle\psi_i^{(l)}|\mathcal{H}_{\text{chain}}|\psi_i^{(r)}\rangle = E_i$. The resolution of unity is $\sum_i |\psi_i^{(r)}\rangle \langle\psi_i^{(l)}| = \mathbf{1}$, where $\mathbf{1}$ is the $k \times k$ unit matrix.

We define $|1\rangle$ as the vector whose first element is one while all its other elements are zero, and $|k\rangle$ as the vector whose last element (the k th element) is one while all its other elements are zero. Then, the transmission probability at energy ϵ is given by

$$\mathcal{T}_{RL}(\epsilon) = |\langle k | [\epsilon - \mathcal{H}_{\text{chain}}]^{-1} | 1 \rangle|^2 a_0, \quad (67)$$

where $[\dots]^{-1}$ is a matrix inverse. To evaluate this matrix inverse, we introduce a resolution of unity to the left and right of $[\epsilon - \mathcal{H}_{\text{chain}}]^{-1}$. This gives

$$\mathcal{T}_{RL} = \sum_i \left| \frac{\langle k | \psi_i^{(r)} \rangle \langle \psi_i^{(l)} | 1 \rangle}{\epsilon - E_i} \right|^2 a_0. \quad (68)$$

For any given set of hoppings a_0, t_1, \dots, t_k , one can easily use a suitable eigenvector finder (we used *Mathematica*) to evaluate this equation numerically, while an analytic solution is straightforward [76] for $k \leq 3$. When all hoppings in the chain are equal, there is a mismatch between the electron's hopping dynamics in the chain and their free motion in the reservoirs. This causes resonances in the transmission, giving the Fabry-Perot-type oscillations in Fig. 10(b) for $k = 5$. However, we can carefully tune the hoppings (to be smallest in the middle of the chain and increasing towards the ends) to get the smooth transmission functions in Fig. 10(c). The $k = 5$ curve in Fig. 10(c) has $t_1 = t_4 = 0.39a_0$ and $t_2 = t_3 = 0.28a_0$, and we choose $a_0 = 1.91$ to normalize the bandwidth to 1. As the number of sites in the chain k increases, the transmission function tends to the desired boxcar function.

The above logic assumes no electron-electron interactions. When we include interaction effects at the mean-field level, things get more complicated. If the states in the chain are all at the same energy E_0 when the chain is unbiased, they will not be aligned when there is a bias between the the reservoirs because the reservoirs also act as gates on the chain states. To engineer a chain where the energies are aligned at the optimal bias, one must adjust the confinement potential of the dots in the chain (or adjust the chemistry of the molecules in the chain) so that their energies are sufficiently out of alignment at zero bias that they all align at optimal bias. In principle, we have the control to do this. However, in practice it would require a great deal of trial-and-error experimental fine tuning. We do not enter further into such practical issues here. Rather, we use the above example to show that there is no *fundamental* reason that the bound on efficiency cannot be achieved.

XIII. MANY QUANTUM SYSTEMS IN PARALLEL

To increase the efficiency at given power output, one must increase the number of transverse modes N . This is because the efficiency decays with the power output divided by the

quantum bounds in Eqs. (43) and (58), and these bounds go like N . However, a strong thermoelectric response requires a transmission function that is highly energy dependent; this typically only occurs when the quantum system (point contact, quantum dot, or molecule) has dimensions of about a wavelength, which implies that N is of order one. Crucial exceptions (beyond the scope of this work) are systems containing superconductors, either SNS structures [1] or Andreev interferometers [77] (see also Ref. [58] and references therein), where strong thermoelectric effects occur for large N .

In the absence of a superconductor, the only way to get large N is to construct a device consisting of many $N = 1$ systems in parallel, such as a surface covered with a certain density of such systems [24,25]. In this case, $P_{\text{gen}}^{\text{qb2}}$ and J_L^{qb} in Eqs. (43) and (58) become bounds on the power per unit area, with N being replaced by the number of transverse modes per unit area. With this one modification, all calculations and results in this paper can be applied directly to such a situation. Carnot efficiency is achieved for a large enough surface area that the power per unit area is much less than $P_{\text{gen}}^{\text{qb2}}$ and J_L^{qb} .

It is worth noting that the number of modes per unit area cannot exceed λ_F^{-2} , for Fermi wavelength λ_F . From this we can get a feeling for the magnitude of the bounds discussed in this paper. Take a typical semiconductor thermoelectric (with $\lambda_F \sim 10^{-8}$ m), placed between reservoirs at 700 and 300 K (typical temperatures for a thermoelectric recovering electricity from the heat in the exhaust gases of a diesel engine car). Equation (43) tells us that to get 100 W of power output from a semiconductor thermoelectric one needs a cross section of at least 4 mm². Then, Eq. (51) tells us that to get this power at 90% of Carnot efficiency, one needs a cross section of at least 0.4 cm². Remarkably, it is *quantum mechanics* which gives these bounds, even though the cross sections in question are macroscopic.

XIV. PHONONS AND PHOTONS CARRYING HEAT IN PARALLEL WITH ELECTRONS

Any chargeless excitation (such as phonons or photons) will carry heat from hot to cold, irrespective of the thermoelectric properties of the system. While some of the phonons and photons will flow through the thermoelectric quantum system, most will flow via other routes (see Fig. 11). A number of theories for these phonon or photon heat currents take the form

$$J_{\text{ph}} = \alpha (T_L^\kappa - T_R^\kappa), \quad (69)$$

where J_{ph} is the heat flow out of the L reservoir due to phonons or photons. The textbook example of such a theory is that of blackbody radiation between the two reservoirs; then, $\kappa = 4$ and α is the Stefan-Boltzmann constant. An example relevant to suspended sub-Kelvin nanostructures is a situation where a finite number N_{ph} of phonon or photon modes carry heat between the two reservoirs [17,40,78,79]; then, $\kappa = 2$ and $\alpha \leq N_{\text{ph}} \pi^2 k_B^2 / (6h)$.

One of the biggest practical challenges for quantum thermoelectrics is that phonons and photons will often carry much more heat than the electrons. This is simply because the hot reservoir can typically radiate heat in all directions as phonons or photons, while electrons only carry heat through

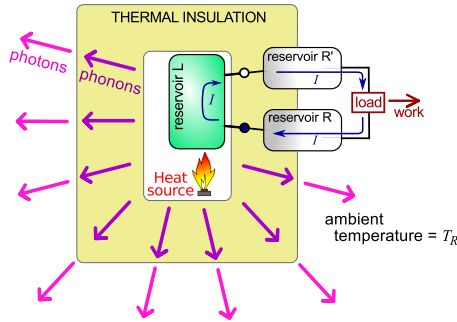


FIG. 11. (Color online) The thermocouple heat engine in Fig. 1, showing the heat flow due to phonon and photons, which carry heat from hot to cold by all possible routes (in parallel with the heat carried by the electrons). This always reduces the efficiency, so it should be minimized with suitable thermal insulation.

the few nanostructures connected to that reservoir. Thus, in many cases the phonon or photon heat flow will dominate over the electronic one. However, progress is being made in blocking phonon and photon flow, by suspending the nanostructure to minimize phonon flow [78] and engineering the electromagnetic environment to minimize photon flow [79], and it can be hoped that phonon and photon effects will be greatly reduced in the future. Hence, here we consider the full range from weak to strong phonon or photon heat flows.

For compactness in what follows, we will only refer to phonon heat flows (usually the dominant parasitic effect). However, strictly one should consider J_{ph} as the sum of the heat flow carried by phonons, photons, and any more exotic chargeless excitations that might exist in a given circuit (mechanical oscillations, spin waves, etc.).

A. Heat engine with phonons

For heat engines, the phonon heat flow is in parallel with electronic heat flow. Thus, the heat flow for a given P_{gen} is $(J_L + J_{ph})$, rather than just J_L (as it was in the absence of phonons). Thus, the efficiency in the presence of the phonons is

$$\eta_{eng}^{e+ph}(P_{gen}) = \frac{P_{gen}}{J_L(P_{gen}) + J_{ph}}. \quad (70)$$

Writing this in terms of the efficiency, we get

$$\eta_{eng}^{e+ph}(P_{gen}) = [\eta_{eng}^{-1}(P_{gen}) + J_{ph}/P_{gen}]^{-1}, \quad (71)$$

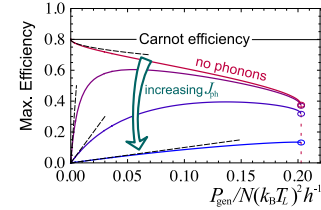
where $\eta_{eng}(P_{gen})$ is the efficiency for $J_{ph} = 0$. Given the maximum efficiency at given power in the absence of phonons, we can use this result to find the maximum efficiency for a given phonon heat flow J_{ph} . An example of this is shown in Fig. 12(a). It shows that for finite J_{ph} , Carnot efficiency is not possible at any power output.

Phonons have a huge effect on the efficiency at small power output. Whenever J_{ph} is nonzero, the efficiency vanishes at zero power output, with

$$\eta_{eng}^{e+ph}(P_{gen}) = P_{gen}/J_{ph} \quad \text{for } P_{gen} \ll J_{ph}. \quad (72)$$

As J_{ph} increases, the range of applicability of this small P_{gen} approximation (shown as dashed lines in Fig. 12) grows

(a) Heat-engine: maximum efficiency for various phonon heat flows



(b) Refrigerator: maximum efficiency for various phonon heat flows

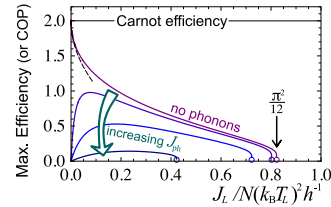


FIG. 12. (Color online) Plots of the maximum efficiency allowed when there is a phonon heat flow J_{ph} in parallel with the heat carried by the electrons. The curves in (a) are for $T_R/T_L = 0.2$, with $J_{ph} = 0, 0.01, 0.1, 1$ (from top to bottom); the curves come from Eq. (71) with $\eta_{eng}(P_{gen})$ given in Fig. 9(a). The curves in (b) are for $T_R/T_L = 1.5$, with $J_{ph} = 0, 0.02, 0.1, 0.4$ (from top to bottom); the curves come from Eq. (73) with $\eta_{eng}(P_{gen})$ given in Fig. 9(b). The maximum cooling power (open circles) is $(\frac{1}{2} J_L^{qb} - J_{ph})$.

towards the maximum power P_{eng}^{qb} (open circles). In contrast, phonon heat flows have little effect on the efficiency near the maximum power output, until these flows become strong enough that $J_{ph} \sim P_{gen}$.

For strong phonon flow, where $J_{ph} \gg P_{gen}$, Eq. (72) applies at all powers up to the maximum P_{gen}^{qb} . Then, the efficiency is maximal when the power is maximal, where maximal power is the quantum bound given in Eq. (43). Thus, the system with both maximal power and maximal efficiency is that with a Heaviside step transmission function (see Sec. XII).

B. Refrigerator with phonons

For a refrigerator to extract heat from a reservoir at rate J in the presence of phonons carrying a back-flow of heat J_{ph} , that refrigerator must extract heat at a rate $J_L = J + J_{ph}$. Note that, for clarity, in this section we take J_{ph} to be positive when $T_L < T_R$ [opposite sign of that in Eq. (69)]. Thus, the efficiency, or COP, in the presence of phonons is the heat current extracted J divided by the electrical power required to extract heat at the rate $J_L = (J + J_{ph})$. This means that

$$\eta_{fri}^{e+ph}(J) = \frac{J \eta_{fri}(J + J_{ph})}{J + J_{ph}}, \quad (73)$$

where $\eta_{fri}(J)$ is the efficiency for $J_{ph} = 0$. We can use this result to find the maximum efficiency for a given phonon heat flow J_{ph} . An example is shown in Fig. 12(b).

Equation (73) means that the phonon flow suppresses the maximum cooling power, so J must now obey

$$J \leq \frac{1}{2} J_L^{qb} - J_{ph} \quad (74)$$

with J_L^{qb} given in Eq. (6). Thus, the upper bound (open circles) in Fig. 12(b) moves to the left as J_{ph} increases.

When the reservoir being refrigerated (reservoir L) is at ambient temperature T_R , then $J_{ph} = 0$ while J_L^{qb} is finite. However, as reservoir L is refrigerated (reducing T_L), J_{ph} grows, while J_L^{qb} shrinks. As a result, at some point (before T_L gets to zero) one arrives at $J_{ph} = \frac{1}{2}J_L^{qb}$, and further cooling of reservoir L is impossible. Thus, given the T_L of J_{ph} for a given system, one can easily find the lowest temperature that reservoir L can be refrigerated to by solving the equation $J_{ph} = \frac{1}{2}J_L^{qb}$ for T_L . To achieve this temperature, one needs the refrigerator with the maximum cooling power (rather than the most efficient one): this is a system with a Heaviside step transmission function (see Sec. XII). Such a system's refrigeration capacities were discussed in Ref. [17].

We also note that, as with the heat engine, phonons have a huge effect on the efficiency at small cooling power, as can be seen in Fig. 12(b). Whenever $0 < J_{ph} < \frac{1}{2}J_L^{qb}$, the efficiency vanishes for small cooling power, with

$$\eta_{fri}^{e+ph}(J) = J \frac{\eta_{fri}(J_{ph})}{J_{ph}} \quad \text{for } J \ll J_{ph}. \quad (75)$$

XV. RELAXATION IN A QUANTUM SYSTEM WITHOUT B FIELD

Elsewhere in this paper, we neglected relaxation in the quantum system. In other words, we assumed that electrons traverse the system in a time much less than the time for inelastic scattering from phonons, photons, or other electrons. We now consider systems in which there is such relaxation, and ask if this relaxation could enable a system to exceed the bounds found above for relaxationless systems. To make progress, we restrict our interest to systems with negligible external magnetic field (B field) [80]. As yet, we have not been able to consider the rich interplay of relaxation and B field [59,61,81].

We use the voltage-probe model [82] shown in Fig. 13(a). A system with relaxation is modeled as a phase-coherent scatterer coupled to a fictitious reservoir M (a region in which relaxation occurs instantaneously). The rate of the relaxation is controlled by the transmission of the lead coupling to reservoir M . We then separate the phase-coherent scatterer into scatterers 1, 2, and 3, as shown in Fig. 13(b), each with their own transmission functions $\mathcal{T}_{ij}(\epsilon)$ with $i, j \in L, M, R$. We assume that the transmission is unchanged under reversal of direction, so $\mathcal{T}_{ij}(\epsilon) = \mathcal{T}_{ji}(\epsilon)$ for all ϵ and i, j . This condition is guaranteed by time-reversal symmetry whenever the B field has a negligible effect on the electron and hole dynamics. However, it also applies for any B field when all particles relax as they traverse the quantum system [then $\mathcal{T}_{LR}(\epsilon) = \mathcal{T}_{RL}(\epsilon) = 0$, which is sufficient to force $\mathcal{T}_{ij}(\epsilon) = \mathcal{T}_{ji}(\epsilon)$ for all i, j].

If the relaxation involves electron-phonon or electron-photon interactions (typically any system which is not sub-Kelvin), the phonons or photons with which the electrons interact usually flow easily between the system and the reservoirs. Thus, these phonons or photons can carry heat current between the fictitious reservoir M and reservoirs L, R (dashed arrows in Fig. 13). The total electrical and heat currents into reservoir M must be zero, and this constraint determines reservoir M 's bias V_M and temperature T_M .

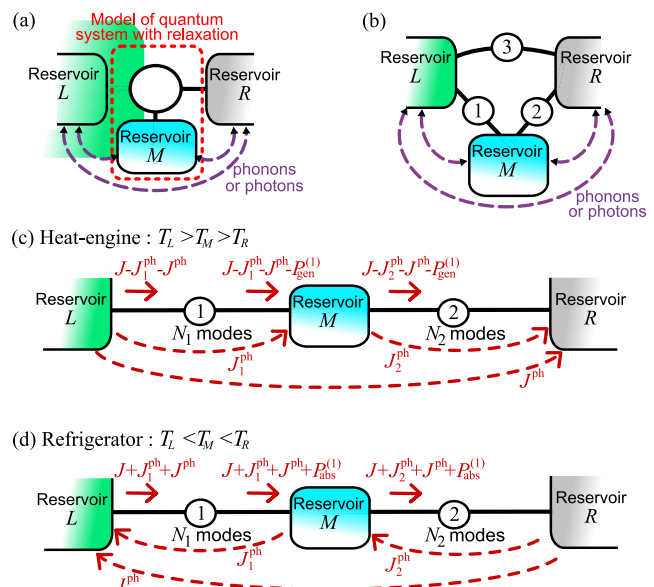


FIG. 13. (Color online) (a) A quantum system in which relaxation occurs is modeled phenomenologically by a coherent quantum system coupled to a third fictitious reservoir M in which the relaxation occurs. (b) The same model after we have separated the system's scattering matrix into three components. The dashed arrows are the exchange of phonons or photons. The arm containing scatterers 1 and 2 is shown in (c) for a heat engine, and in (d) for a refrigerator.

A. Method of overestimation

The optimal choice of \mathcal{T}_{ML} and \mathcal{T}_{RM} depends on T_M , while T_M depends on the heat current, and thus on \mathcal{T}_{ML} and \mathcal{T}_{RM} . The solution and optimization of this self-consistency problem has been beyond our ability to resolve, even though we have restricted ourselves to a simple model of relaxation in a system with negligible B field. Instead, we make a simplification which leads to an *overestimate* of the efficiency. We assume V_M, T_M are free parameters (not determined from \mathcal{T}_{ML} and \mathcal{T}_{RM}), with T_M between T_L and T_R . If we find the optimal \mathcal{T}_{ML} and \mathcal{T}_{RM} for given T_M , and then find the optimal T_M (irrespective of whether it is consistent with \mathcal{T}_{ML} and \mathcal{T}_{RM} or not), we have an overestimate of the maximal efficiency. Even with this simplification, we have only been able to address the low power and high power limits. However, we show in the following that this overestimate is sufficient to prove the following.

(1) At low power, relaxation cannot make the system's efficiency exceed that of the optimal relaxation-free system with N_{\max} modes.

(2) Relaxation cannot make a system's power exceed that of the maximum possible power of a relaxation-free system with N_{\max} modes.

Defining N_L and N_R as the number of transverse modes in the system to the left and right of the region where relaxation occurs,

$$N_{\max} = \max[N_L, N_R], \quad (76)$$

B. Efficiency of heat engine with relaxation

To get the efficiency for our model of a quantum system with relaxation, we must find the efficiency for the system

in Fig. 13(b). This system has two ‘‘arms.’’ One arm contains scatterers 1 and 2, and we define its efficiency as $\eta_{\text{eng}}^{(1\&2)}$. The other arm contains scatterer 3, and we define its efficiency as $\eta_{\text{eng}}^{(3)}$. The efficiency of the full system $\eta_{\text{eng}}^{\text{total}}(P_{\text{gen}})$ is given by

$$\frac{1}{\eta_{\text{eng}}^{\text{total}}(P_{\text{gen}})} = \frac{p_{\text{rel}}}{\eta_{\text{eng}}^{(1\&2)}(p_{\text{rel}}P_{\text{gen}})} + \frac{q_{\text{rel}}}{\eta_{\text{eng}}^{(3)}(q_{\text{rel}}P_{\text{gen}})}. \quad (77)$$

Here, p_{rel} is the proportion of transmitted electrons that have passed through the arm containing scatterers 1 and 2, while $q_{\text{rel}} = (1 - p_{\text{rel}})$ is the proportion that have passed through the arm containing scatterer 3. Physically, p_{rel} is the probability that an electron entering the quantum system relaxes before transmitting, while q_{rel} is the probability that it transmits before relaxing. One sees from Eq. (77) that the maximal efficiency for a given p_{rel} occurs when both $\eta_{\text{eng}}^{(1\&2)}$ and $\eta_{\text{eng}}^{(3)}$ are maximal.

The upper bound on $\eta_{\text{eng}}^{(3)}$ is that given in Sec. IX with $q_{\text{rel}}N_L$ modes to the left and $q_{\text{rel}}N_R$ modes to the right. Our objective now is to find the maximum $\eta_{\text{eng}}^{(1\&2)}$ with $N_1 = p_{\text{rel}}N_L$ modes on the left and $N_2 = p_{\text{rel}}N_R$ modes on the right. More precisely, our objective is to find an *overestimate* of this maximum. For the heat flows indicated in Fig. 13(c), the efficiency is

$$\begin{aligned} \eta_{\text{eng}}^{(1\&2)} &\equiv P_{\text{gen}}^{(1\&2)}/J \\ &= \frac{1}{J} [P_{\text{gen}}^{(1)}(J_1; T_M, T_L) + P_{\text{gen}}^{(2)}(J_2; T_R, T_M)], \end{aligned} \quad (78)$$

where $J_1 = J - J_1^{\text{ph}} - J^{\text{ph}}$ and $J_2 = J - J_2^{\text{ph}} - J^{\text{ph}} - P_{\text{gen}}^{(1)}$. One sees that $\eta_{\text{eng}}^{(1\&2)}$ is maximal for given T_M when $J^{\text{ph}} = J_1^{\text{ph}} = J_2^{\text{ph}} = 0$ (these heat currents cannot be negative because $T_L > T_M > T_R$). Thus, to get our overestimate of the maximal efficiency for given T_M , we assume these phonon and photon heat currents are zero. Then, with a little algebra, one finds that

$$1 - \eta_{\text{eng}}^{(1\&2)}(P_{\text{gen}}^{(1\&2)}) = (1 - \eta_{\text{eng}}^{(1)}(P_{\text{gen}}^{(1)}))(1 - \eta_{\text{eng}}^{(2)}(P_{\text{gen}}^{(2)})),$$

where $P_{\text{gen}}^{(1)}$ and $P_{\text{gen}}^{(2)}$ are related to $P_{\text{gen}}^{(1\&2)}$ by

$$P_{\text{gen}}^{(\mu)} = P_{\text{gen}}^{(1\&2)} \eta_{\text{eng}}^{(\mu)} / \eta_{\text{eng}}^{(1\&2)} \quad (79)$$

for $\mu = 1, 2$. For given T_M , one maximizes $\eta_{\text{eng}}^{(1\&2)}$ by independently maximizing $\eta_{\text{eng}}^{(1)}$ and $\eta_{\text{eng}}^{(2)}$. For low powers, Eq. (51) with $P, N, T_R \rightarrow P_1, N_1, T_M$ gives $\eta_{\text{eng}}^{(1)}$, while with $P, N, T_L \rightarrow P_2, N_2, T_M$ gives $\eta_{\text{eng}}^{(2)}$. In this limit, we can treat efficiencies in Eq. (79) to zeroth order in $P_{\text{gen}}^{(1\&2)}$, taking them to be Carnot efficiencies, so

$$P_{\text{gen}}^{(1)} \simeq \frac{T_L - T_M}{T_L - T_R} P_{\text{gen}}^{(1\&2)}, \quad P_{\text{gen}}^{(2)} \simeq \frac{T_M - T_R}{T_L - T_R} P_{\text{gen}}^{(1\&2)}.$$

Then, some algebra gives the overestimate of efficiency at low powers for given T_M to be

$$\eta_{\text{eng}}^{(1\&2)} \leq \eta_{\text{eng}}^{\text{Carnot}} \left(1 - 0.478 \sqrt{\frac{T_R}{T_L} \frac{P_{\text{gen}} K_{\text{rel}}}{P_{\text{gen}}^{\text{qb}2}(N=1)}} \right), \quad (80)$$

with $P_{\text{gen}}^{\text{qb}2}(N=1)$ given by Eq. (43) with $N=1$, and

$$K_{\text{rel}} = \sqrt{\frac{1}{N_1} \frac{T_R(T_L - T_M)}{T_M(T_L - T_R)}} + \sqrt{\frac{1}{N_2} \frac{T_L(T_M - T_R)}{T_M(T_L - T_R)}}, \quad (81)$$

where $N_1 = p_{\text{rel}}N_L$ and $N_2 = p_{\text{rel}}N_R$ are, respectively, the number of transmission modes in scattering matrices 1 and 2. The overestimate of $\eta_{\text{eng}}^{(1\&2)}$ is maximal when T_M is chosen to minimize K_{rel} . The two minima of K_{rel} are at $T_M = T_R$ and $T_M = T_L$, for which the values of K_{rel} are $1/\sqrt{N_1}$ and $1/\sqrt{N_2}$, respectively. Thus, we have

$$K_{\text{rel}} \geq 1/\sqrt{p_{\text{rel}}N_{\text{max}}}, \quad (82)$$

with N_{max} in Eq. (76). Thus, whatever T_M may be,

$$\begin{aligned} \eta_{\text{eng}}^{(1\&2)}(P_{\text{gen}}^{(1\&2)}) \\ \leq \eta_{\text{eng}}^{\text{Carnot}} \left(1 - 0.478 \sqrt{\frac{T_R}{T_L} \frac{P_{\text{gen}}^{(1\&2)}}{P_{\text{gen}}^{\text{qb}2}(p_{\text{rel}}N_{\text{max}})}} \right). \end{aligned} \quad (83)$$

Since $P_{\text{gen}}^{(1\&2)} = p_{\text{rel}}P_{\text{gen}}$, we can simplify Eq. (83) by noting that

$$\frac{P_{\text{gen}}^{(1\&2)}}{P_{\text{gen}}^{\text{qb}2}(p_{\text{rel}}N_{\text{max}})} = \frac{P_{\text{gen}}}{P_{\text{gen}}^{\text{qb}2}(N_{\text{max}})}, \quad (84)$$

where P_{gen} is the total power generated by the combined system made of scatterers 1, 2, and 3. Then, substituting the result into Eq. (77), we get an overestimate of the efficiency at power output P_{gen} which is equal to the upper bound we found in the absence of relaxation [Eq. (51)].

Thus, we can conclude that for small power outputs, no quantum system with relaxation within it can exceed the upper bound on efficiency found for a *relaxation-free* system with N_{max} transverse modes. Since the proof is based on an overestimate of the efficiency for a system with relaxation, we cannot say if a system with finite relaxation can approach the bound in Eq. (51). Unlike in the relaxation-free case, we cannot say what properties the quantum system with relaxation (as given in terms of the properties of the effective scatterers 1, 2, and 3) are necessary to maximize the efficiency at given power output. We simply know that it cannot exceed Eq. (51).

C. Refrigerator with relaxation

Our objective is to find an overestimate of the maximal efficiency of a refrigerator that is made of quantum systems in which relaxation occurs. The efficiency of the system with relaxation $\eta_{\text{fri}}^{\text{total}}(P_{\text{gen}})$ is given by

$$\eta_{\text{fri}}^{\text{total}}(J_L) = p_{\text{rel}}\eta_{\text{fri}}^{(1\&2)}(p_{\text{rel}}J_L) + q_{\text{rel}}\eta_{\text{fri}}^{(3)}(q_{\text{rel}}J_L), \quad (85)$$

thus we need to find an upper bound on $\eta_{\text{fri}}^{(1\&2)}$. We make an overestimate of this efficiency by taking T_M to be a free parameter between T_L and T_R . For given T_M , the efficiency of the combined systems 1 and 2 is

$$\eta_{\text{fri}}^{(1\&2)}(J) = J / [P_{\text{abs}}^{(1)}(J_1) + P_{\text{abs}}^{(2)}(J_2)], \quad (86)$$

where $J_1 = J + J_1^{\text{ph}} + J^{\text{ph}}$ and $J_2 = J + J_2^{\text{ph}} + J^{\text{ph}} + P_{\text{abs}}^{(1)}$ [see Fig. 13(d)]. This efficiency is maximized when $J_1^{\text{ph}}, J_2^{\text{ph}}, J^{\text{ph}} = 0$ (since $T_L < T_M < T_R$ means these currents are not negative). Then, a little algebra gives

$$1 + \frac{1}{\eta_{\text{fri}}^{(1\&2)}(J)} = \left[1 + \frac{1}{\eta_{\text{fri}}^{(1)}(J)} \right] \left[1 + \frac{1}{\eta_{\text{fri}}^{(2)}(J_2)} \right], \quad (87)$$

where $J_2 = J + P_{\text{abs}}^{(1)} = J[1 + 1/\eta_{\text{fri}}^{(1)}(J)]$. Thus, to maximize $\eta_{\text{fri}}^{(1\&2)}(J)$ for given T_M , one must maximize both $\eta_{\text{fri}}^{(1)}$ and $\eta_{\text{fri}}^{(2)}$. For low power, this can be done using Eq. (63) (much as for the heat engine in Sec. XV B) giving

$$\eta_{\text{fri}}^{(1\&2)} \leq \eta_{\text{fri}}^{\text{Carnot}} \left(1 - 1.09 \sqrt{\frac{T_R}{T_R - T_L} \frac{J_L K_{\text{rel}}}{J_L^{\text{qb}}(N=1)}} \right), \quad (88)$$

where K_{rel} is given in Eq. (81), and $J_L^{\text{qb}}(N=1)$ is given by Eq. (58) with $N=1$. The overestimate of $\eta_{\text{fri}}^{(1\&2)}$ is maximal when K_{rel} is minimal [see Eq. (82)]. Substituting this into Eq. (85), we see that the efficiency with relaxation does not exceed the result in Eq. (63) for a *relaxation-free* system with N_{max} transverse modes.

D. Quantum bounds on power with relaxation

For a heat engine, the arm with scatterers 1 and 2 has a maximum power

$$P_{\text{gen}}^{(1\&2)} \leq A_0 \frac{\pi^2}{h} k_B^2 [N_1(T_L - T_M)^2 + N_2(T_M - T_R)^2].$$

Since $(T_L - T_M)^2 + (T_M - T_R)^2 \leq (T_L - T_R)^2$, the power of the full system cannot exceed the maximum power of a relaxationless system [Eq. (43)] with N_{max} modes.

For a refrigerator, the arm containing scatterers 1 and 2 has a maximum cooling power

$$J \leq \begin{cases} \pi^2 N_1 k_B^2 T_L^2 / (12h), \\ \pi^2 N_2 k_B^2 T_M^2 / (12h) - P_{\text{abs}}^{(1)}, \end{cases} \quad (89)$$

where $P_{\text{abs}}^{(1)}$ is the electrical power absorbed by scatter 1. The upper (lower) term is the limit on the heat flow into scatterer 1 (scatterer 2), noting that the heat flow into scatterer 2 is $J + P_{\text{abs}}^{(1)}$. Unless $N_2 \gg N_1$, the lower limit is the more restrictive one. In any case, the cooling power of the full system can

never exceed the maximum power of a relaxationless system [Eq. (58)] with N_{max} modes.

XVI. CONCLUSIONS

The upper bound on efficiency at zero power (i.e., Carnot efficiency) is classical since it is independent of wavelike nature of the electrons. However, this work on thermoelectrics shows that the upper bound on efficiency at finite power is quantum, depending on the ratio of the thermoelectric's cross section to the electrons' Fermi wavelength. If one thought that electrons were classical (strictly zero wavelength), one would believe that Carnot efficiency was achievable at any power output. Quantum mechanics appears to tell us that this is not so.

However, a crucial point for future work is to discover how universal our bounds on efficiency at finite power are. Our bounds currently rely on the quantum system being (a) well modeled by the nonlinear scattering theory with its mean-field treatment of electron-electron interactions, (b) coupled to only two reservoirs (hot and cold), and (c) relaxation free. Under certain conditions, we have also shown that they apply when there is relaxation in the quantum system. We cannot yet prove that our results are as general as Pendry's bound on heat flow [40], which applies for arbitrary relaxation and for more than two reservoirs [42], as well as for electronic Luttinger liquids [83] and bosons [40]. It also remains to be seen if our bound occurs in systems with strong electron-electron interactions (Coulomb blockade, Kondo physics, etc.). More generally, we wonder whether similar bounds apply to those thermodynamic machines that do not rely on thermoelectric effects, such as Carnot heat engines.

ACKNOWLEDGMENTS

I am very grateful to M. Büttiker for the suggestion which led to the implementation in Sec. XII. I thank P. Hänggi for questions on entropy flow which led to Sec. III and L. Correa for questions which led to a great improvement of Sec. II C. I thank C. Grenier for an analytic solution of Eq. (68) for $k=3$.

-
- [1] F. Giazotto, T. T. Heikkilä, A. Luukanen, A. M. Savin, and J. P. Pekola, *Rev. Mod. Phys.* **78**, 217 (2006); J. T. Muhonen, M. Meschke, and J. P. Pekola, *Rep. Prog. Phys.* **75**, 046501 (2012).
 - [2] G. Benenti, G. Casati, T. Prosen, and K. Saito, [arXiv:1311.4430](https://arxiv.org/abs/1311.4430).
 - [3] B. Sothmann, R. Sánchez, and A. N. Jordan, *Nanotechnology* **26**, 032001 (2015).
 - [4] F. Haupt, M. Leijnse, H. L. Calvo, L. Classen, J. Splettstoesser, and M. R. Wegewijs, *Phys. Status Solidi B* **250**, 2315 (2013).
 - [5] M. Paulsson and S. Datta, *Phys. Rev. B* **67**, 241403(R) (2003).
 - [6] P. Reddy, S. Y. Jang, R. A. Segalman, and A. Majumdar, *Science* **315**, 1568 (2007).
 - [7] H. J. Goldsmid, *Introduction to Thermoelectricity* (Springer, Heidelberg, 2009).
 - [8] F. J. DiSalvo, *Science* **285**, 703 (1999).
 - [9] A. Shakouri and M. Zebarjadi, *Thermal Nanosystems and Nanomaterials*, edited by S. Volz (Springer, Heidelberg, 2009), Chap. 9; A. Shakouri, *Annu. Rev. Mater. Res.* **41**, 399 (2011).
 - [10] C. Grenier, C. Kollath, and A. Georges, [arXiv:1209.3942](https://arxiv.org/abs/1209.3942).
 - [11] J.-P. Brantut, C. Grenier, J. Meineke, D. Stadler, S. Krinner, C. Kollath, T. Esslinger, and A. Georges, *Science* **342**, 713 (2013).
 - [12] Ch. Grenier, A. Georges, and C. Kollath, *Phys. Rev. Lett.* **113**, 200601 (2014).
 - [13] J. Gemmer, M. Michel, and G. Mahler, *Quantum Thermodynamics* (Springer, Berlin, 2010).
 - [14] R. S. Whitney, *Phys. Rev. Lett.* **112**, 130601 (2014).
 - [15] M. Zebarjadi, K. Esfarjani, and A. Shakouri, *Appl. Phys. Lett.* **91**, 122104 (2007); *Mater. Res. Soc. Proc.* **1044**, U10-04 (2008).
 - [16] B. Muralidharan and M. Grifoni, *Phys. Rev. B* **85**, 155423 (2012).
 - [17] R. S. Whitney, *Phys. Rev. B* **88**, 064302 (2013).
 - [18] J. Meair, and Ph. Jacquod, *J. Phys.: Condens. Matter* **25**, 082201 (2013).
 - [19] A. Crépieux and F. Michelini, *J. Phys.: Condens. Matter* **27**, 015302 (2015).
 - [20] J. Azema, P. Lombardo, and A.-M. Daré, *Phys. Rev. B* **90**, 205437 (2014).

- [21] G. D. Mahan and J. O. Sofo, *Proc. Natl. Acad. Sci. USA* **93**, 7436 (1996).
- [22] T. E. Humphrey and H. Linke, *Phys. Rev. Lett.* **94**, 096601 (2005).
- [23] R. Kim, S. Datta, and M. S. Lundstrom, *J. Appl. Phys.* **105**, 034506 (2009).
- [24] A. N. Jordan, B. Sothmann, R. Sanchez, and M. Büttiker, *Phys. Rev. B* **87**, 075312 (2013).
- [25] B. Sothmann, R. Sanchez, A. N. Jordan, and M. Büttiker, *New J. Phys.* **15**, 095021 (2013).
- [26] F. L. Curzon and B. Ahlborn, *Am. J. Phys.* **43**, 22 (1975).
- [27] J. Yvon, *Proceedings of the International Conference on Peaceful Uses of Atomic Energy* (United Nations, New York, 1956), Vol. 2, p. 337.
- [28] P. Chambadal, *Les Centrales Nucléaires* (Armand Colin, Paris, 1957), p. 41.
- [29] I. I. Novikov, *At. Energ.* **3**, 409 (1957) [*J. Nucl. Energy II* **7**, 125 (1958)].
- [30] S. Velasco, J. M. M. Roco, A. Medina, and A. C. Hernández, *Phys. Rev. Lett.* **78**, 3241 (1997).
- [31] C. de Tomás, A. C. Hernández, and J. M. M. Roco, *Phys. Rev. E* **85**, 010104(R) (2012).
- [32] Y. Apertet, H. Ouerdane, A. Michot, C. Goupil, and P. Lecoeur, *Europhys. Lett.* **103**, 40001 (2013).
- [33] L. A. Correa, J. P. Palao, G. Adesso, and D. Alonso, *Phys. Rev. E* **90**, 062124 (2014).
- [34] M. Esposito, K. Lindenberg, and C. Van den Broeck, *Europhys. Lett.* **85**, 60010 (2009).
- [35] This point is easily overlooked in plots where power is in units of each system's maximum power, such as Fig. 4 of Ref. [2]. It is easy to misread such a plot as implying that the system with $ZT \rightarrow \infty$ is the most efficient at all powers.
- [36] N. Nakpathomkun, H. Q. Xu, and H. Linke, *Phys. Rev. B* **82**, 235428 (2010).
- [37] D. M. Kennes, D. Schuricht, and V. Meden, *Europhys. Lett.* **102**, 57003 (2013) (2013).
- [38] M. Leijnse, M. R. Wegewijs, and K. Flensberg, *Phys. Rev. B* **82**, 045412 (2010).
- [39] S. Hershfield, K. A. Muttalib, and B. J. Nartowt, *Phys. Rev. B* **88**, 085426 (2013).
- [40] J. B. Pendry, *J. Phys. A.: Math. Gen.* **16**, 2161 (1983).
- [41] J. D. Bekenstein, *Phys. Rev. Lett.* **46**, 623 (1981); *Phys. Rev. D* **30**, 1669 (1984).
- [42] R. S. Whitney, *Phys. Rev. B* **87**, 115404 (2013).
- [43] W. Kern, *Z. Phys. B* **20**, 215 (1975).
- [44] W. A. Day, *Acta Mech.* **27**, 251 (1977).
- [45] D. Jou, J. Casas-Vázquez, and G. Lebon, *Extended Irreversible Thermodynamics* (Springer, Berlin, 2001).
- [46] H.-L. Engquist and P. W. Anderson, *Phys. Rev. B* **24**, 1151 (1981).
- [47] U. Sivan and Y. Imry, *Phys. Rev. B* **33**, 551 (1986).
- [48] P. N. Butcher, *J. Phys.: Condens. Matter* **2**, 4869 (1990).
- [49] Reference [47] instead considered the von Neumann entropy.
- [50] Ya. M. Blanter, and M. Büttiker, *Phys. Rep.* **336**, 1 (2000).
- [51] T. Christen and M. Büttiker, *Europhys. Lett.* **35**, 523 (1996).
- [52] See, e.g., Eq. (11) of B. Cleuren, B. Rutten, and C. Van den Broeck, *Phys. Rev. Lett.* **108**, 120603 (2012).
- [53] L. Bruneau, V. Jakšić, and C.-A. Pillet, *Commun. Math. Phys.* **319**, 501 (2013).
- [54] D. Sánchez and R. López, *Phys. Rev. Lett.* **110**, 026804 (2013); R. López and D. Sánchez, *Phys. Rev. B* **88**, 045129 (2013).
- [55] L. W. Molenkamp, Th. Gravier, H. van Houten, O. J. A. Buijk, M. A. A. Mabeesoone, and C. T. Foxon, *Phys. Rev. Lett.* **68**, 3765 (1992).
- [56] M. G. Vavilov and A. D. Stone, *Phys. Rev. B* **72**, 205107 (2005).
- [57] D. Nozaki, H. Sevinçli, W. Li, R. Gutiérrez, and G. Cuniberti, *Phys. Rev. B* **81**, 235406 (2010).
- [58] Ph. Jacquod and R. S. Whitney, *Europhys. Lett.* **91**, 67009 (2010).
- [59] K. Saito, G. Benenti, G. Casati, and T. Prosen, *Phys. Rev. B* **84**, 201306(R) (2011).
- [60] K. K. Saha, T. Markussen, K. S. Thygesen, and B. K. Nikolić, *Phys. Rev. B* **84**, 041412(R) (2011).
- [61] D. Sánchez and L. Serra, *Phys. Rev. B* **84**, 201307(R) (2011).
- [62] O. Karlström, H. Linke, G. Karlström, and A. Wacker, *Phys. Rev. B* **84**, 113415 (2011).
- [63] Ph. Jacquod, R. S. Whitney, Jonathan Meair, and M. Büttiker, *Phys. Rev. B* **86**, 155118 (2012).
- [64] S.-Y. Hwang, R. López, M. Lee, and D. Sánchez, *Phys. Rev. B* **90**, 115301 (2014).
- [65] B. Sothmann, R. Sánchez, and A. N. Jordan, *Europhys. Lett.* **107**, 47003 (2014).
- [66] J. Matthews, F. Battista, D. Sánchez, P. Samuelsson, and H. Linke, *Phys. Rev. B* **90**, 165428 (2014).
- [67] M. Galperin, M. A. Ratner, and A. Nitzan, *J. Phys.: Condens. Matter* **19**, 103201 (2007); M. Galperin, A. Nitzan, and M. A. Ratner, *Mol. Phys.* **106**, 397 (2008).
- [68] P. Murphy, S. Mukerjee, and J. Moore, *Phys. Rev. B* **78**, 161406(R) (2008).
- [69] S. F. Svensson, E. A. Hoffmann, N. Nakpathomkun, P. M. Wu, H. Xu, H. A. Nilsson, D. Sánchez, V. Kashcheyevs, and H. Linke, *New J. Phys.* **15**, 105011 (2013).
- [70] F. Battista, F. Haupt, and J. Splettstoesser, *Phys. Rev. B* **90**, 085418 (2014).
- [71] M. A. Sierra and D. Sánchez, *Phys. Rev. B* **90**, 115313 (2014).
- [72] The “2” on $P_{\text{gen}}^{\text{qb}2}$ indicates that this bound is for a quantum system between *two* reservoirs. We do not have the exact bound for a system between three (or more) reservoirs at different temperatures, however, an overestimate is given in Sec. IX of Ref. [42].
- [73] T. Schmiedl and U. Seifert, *Europhys. Lett.* **81**, 20003 (2008).
- [74] M. Büttiker, *Phys. Rev. B* **41**, 7906 (1990).
- [75] This was an idea proposed by M. Büttiker (private communication).
- [76] C. Grenier (private communication).
- [77] J. Eom, C.-J. Chien, and V. Chandrasekhar, *Phys. Rev. Lett.* **81**, 437 (1998).
- [78] J. S. Heron, T. Fournier, N. Mingo, and O. Bourgeois, *Nano Lett.* **9**, 1861 (2009); J.-S. Heron, C. Bera, T. Fournier, N. Mingo, and O. Bourgeois, *Phys. Rev. B* **82**, 155458 (2010).
- [79] D. R. Schmidt, R. J. Schoelkopf, and A. N. Cleland, *Phys. Rev. Lett.* **93**, 045901 (2004); L. M. A. Pascal, H. Courtois, and F. W. J. Hekking, *Phys. Rev. B* **83**, 125113 (2011).
- [80] Reference [14] does not mention this crucial restriction, because I had overlooked its essential role in the proof.
- [81] O. Entin-Wohlman and A. Aharony, *Phys. Rev. B* **85**, 085401 (2012).
- [82] M. Büttiker, *IBM J. Res. Dev.* **32**, 63 (1988).
- [83] C. L. Kane and M. P. A. Fisher, *Phys. Rev. B* **55**, 15832 (1997).