From thermoelectricity to phonoelectricity

Corey Melnick1,a) and Massoud Kaviany2,b)

AFFILIATIONS
1Brookhaven National Lab Condensed Matter Physics and Material Science Division, Upton, New York 11973-5000, USA
2Heat Transfer Physics Laboratory, Department of Mechanical Engineering, University of Michigan, Ann Arbor, Michigan 48109-2125, USA

a)Electronic mail: cmelnick@bnl.gov
b)Electronic mail: kaviany@umich.edu

ABSTRACT
Here, we review the history and progress of solid-state heat harvesting. Efficiency models, material metrics, and current research are discussed for thermoelectrics, thermionics, laser-coolers, and modern technologies. Then, we discuss nonequilibrium optical phonon harvest in the phonovoltaic cell. We make an effort to distinguish between the harvest of equilibrium heat, or thermal-electricity, and the targeted harvest of a particular nonequilibrium phonon mode, or phonoelectricity. We survey the state of phonovoltaic research, focusing on phonovoltaic materials, the electron-phonon coupling, and entropy production in a phonovoltaic cell. Throughout this review, discussions are connected to the electron and phonon structures, interactions, and transport. The modern thermal-electric harvesters are shown to reshape broad-spectrum, high-entropy heat into a narrow-spectrum of low-entropy emissions in order to efficiently generate thermal-electricity. Phonoelectricity, in contrast, intervenes before a low-entropy population of nonequilibrium optical phonons becomes high-entropy heat. In particular, the phonovoltaic cell generates phonoelectricity by utilizing the nonequilibrium, low-temperature optical phonon population produced by, e.g., the relaxation of electrons excited by an electric field. A phonovoltaic material has an ultranarrow electronic bandgap, such that the hot optical phonon population can relax by producing electron-hole pairs (and power) instead of multiple acoustic phonons (and entropy). The phonovoltaic device has an electric diode, e.g., a p-n junction, such that the internal electric field of the diode splits the electrons and holes in order to produce an electric current, and thus, power. The low entropy and high-temperature of the nonequilibrium optical phonon population enable efficient, in-situ heat harvesting. Bilayer graphene, for example, can theoretically convert the nonequilibrium population of optical phonons generated in a field-effect transistor into electricity with an efficiency exceeding 70% of the Carnot limit. The thermal-electric devices, in contrast, are either inefficient or restricted to ex-situ power generation. Furthermore, as the phonoelectric Carnot limit is defined by the local nonequilibrium between electron and optical phonon populations, rather than the spatial nonequilibrium across the device, the Carnot limit can be very large without inducing melting.

Published under license by AIP Publishing. https://doi.org/10.1063/1.5031425

TABLE OF CONTENTS
I. INTRODUCTION ............................................. 1
II. SOLID-STATE HEAT HARVESTING AND THERMAL-ELECTRICITY .............................. 3
   A. Thermoelectrics ........................................... 4
      1. Efficiency and material metrics .................. 4
      2. Materials .................................................. 5
   B. Thermionics .................................................. 6
      1. Efficiency and material metrics .................. 7
      2. Materials .................................................. 8
   C. Laser cooling ............................................... 8
      1. Efficiency and material metrics .................. 9
      2. Materials .................................................. 10
   D. Modern explorations ..................................... 11
   E. Summary of thermal-electric devices and their performance .................................. 15
      1. TE ..................................................... 15
      2. TI ..................................................... 16
      3. PETE .................................................. 16
      4. TPV .................................................. 16
III. PHONON HARVEST, PHONOELECTRICITY, AND THE PHONOVALETAIC CELL .................. 16
    A. Efficiency and figure of merit ...................... 17
    B. Materials .................................................. 20
    1. Tuning the electronic structure of graphene .... 22

1. Thermophotovoltaics .................................. 11
2. Photon-enhanced thermionics ...................... 12
3. Hot phonon absorption barrier ..................... 14
I. INTRODUCTION

Modern electronic, energy conversion, and energy storage devices release massive amounts of energy which quickly becomes heat. If this heat is not properly managed, the temperature of the device increases drastically, and its performance and reliability suffer. Indeed, modern devices are often limited by our ability to remove the heat they generate. Before the energy they produce becomes heat, however, it first excites the electron population. The resulting nonequilibrium population of hot electrons relaxes by colliding with the crystal lattice and releasing a narrow spectrum of high-frequency vibrational quanta (optical phonons). Then, this nonequilibrium (hot) population of optical phonons down-converts into multiple low-frequency acoustic phonons, as shown in Fig. 1(a). That is, the nonequilibrium optical phonon population thermalizes and becomes equilibrium heat. The resulting, broad-spectrum of acoustic phonons has substantially more entropy than its precursor, a narrow-spectrum of nonequilibrium optical phonons. Therefore, intervening before the hot optical phonon population thermalizes and harvesting it to generate electricity should enable a substantially higher conversion efficiency than that achieved in conventional solid-state heat harvesters like the thermoelectric (TE) generator.

Examples of prethermalization intervention and heat harvesting are shown in Fig. 1(a), the related coupling mechanisms are shown in Fig. 1(b), and the associated energy barrier and excitation are shown in Fig. 1(c). While high-entropy heat can be reshaped, as in the thermophotovoltaic (TPV) shown in Fig. 2(a), prethermalization intervention avoids the formation of entropy in the first place, as in the phonovoltaic (pV) shown in Fig. 2(b). However, the thermalization process typically occurs on the picosecond and nanometer scale, molecular vibrations have an energy that is limited to below $600 \text{ meV}$, and typical optical phonon energies are in the $25–75 \text{ meV}$ range. In combination, these factors make targeting and harvesting a hot optical phonon population very challenging. Thus, the harvest of heat (thermal-electricity), rather than a nonequilibrium optical phonon population (phonon-electricity or phonoelectricity), is much more common. Indeed, the first promising phonoelectric device, the phonovoltaic, was

![Image](image-url)
The conventional thermal cycles, e.g., the Rankine and Brayton cycles, dominate in megawatt-scale power generation. However, solid-state heat harvesting with, e.g., thermoelectric or thermionic power generators, has attracted a lot of attention. Their lack of moving parts, small size, and reliability make them attractive alternatives. Improving upon the efficiency of a modern Rankine- or Brayton-cycle power plant, however, has eluded researchers. Thus, these technologies have often been relegated to remote and space applications, where size, weight, and reliability are crucial; to waste-heat recovery, where size and weight are important; and to laboratories, where the search for improved materials, manufacturing techniques, and thermal-electric processes continues in the hope of realizing the promise of competitive and efficient solid-state heat harvesting. There have been impressive advances on these fronts, and researchers have begun to achieve this goal after a few decades during which thermal-electric research looked like it might have plateaued.

In this review, we will discuss the history and outlook of the conventional solid-state heat-harvest and cooling technologies, i.e., the thermoelectric, thermionic, and laser-cooling processes. Then, we will discuss the more recent explorations of solid-state heat harvesting and cooling. In particular, we will discuss the thermophotovoltaic, the photon-enhanced thermionic, and the hot-phonon absorption barrier. Finally, we will overview phonoelectric power generation in the phonovoltaic cell.

The phonovoltaic, in contrast to these devices, does not harvest photons. Instead, it is driven by a nonequilibrium population of optical phonons. The device uses the low entropy of this population in order to generate phonoelectricity with staggering efficiency: A bilayer-graphene phonovoltaic, for example, can harvest optical phonons with an efficiency of up to 70% of the Carnot limit. Notably, the Carnot efficiency is not defined by the nonequilibrium which develops spatially across the device, but it is instead defined by the local nonequilibrium between the optical phonon and electron populations. Thus, a phonovoltaic can sustain the substantial nonequilibrium required to achieve a large Carnot limit without melting. Additionally, it must function within a device, e.g., it recycles the phonons produced by a graphene field-effect transistor (FET) in-situ. Thermal-electric generators, in contrast, must harvest the waste-heat ex-situ. Therefore, it can prevent a device from producing massive quantities of heat, it recycles the optical phonons before they ever become heat, and it reduces the temperature reached by the device.

Before proceeding to our discussion of the thermal-electric and phonoelectric devices, let us first discuss the physics which allow them to operate. These theoretical discussions will provide the foundation of our review, while the thermal-electric review and history will provide the framework for our discussion of phonoelectricity, phonovoltaic cells, and the advantages of prethermalization intervention.

II. SOLID-STATE HEAT HARVESTING AND THERMAL-ELECTRICITY

Before proceeding to discuss modern heat harvest technologies, let us discuss the history of solid-state heat harvesting and the generation of thermal-electricity. In particular, we will discuss thermoelectric generators, thermionic converters, optical refrigeration, and some modern advances in thermal-electricity, i.e., the thermophotovoltaic, photon-enhanced thermionic converter, and the hot-phonon absorption barrier. Let us begin chronologically, with the discovery of the
thermoelectric effects and the invention of the thermoelectric cooler and generator. Note that more in-depth theory related to heat and phonon harvest is presented in the Appendices.

A. Thermoelectrics

Thermoelectricity has a long history: The first thermoelectric effect, the Seebeck effect, was discovered in 1787 by Volta and independently rediscovered in 1821 by Seebeck, for whom the effect is named. The Seebeck effect, like the Peltier effect, describes the diffusion of charge due to a temperature gradient. Seebeck discovered that two metals connected in a closed loop would deflect a needle when a temperature difference was applied between the two joints. While he did not recognize that this was due to the formation of an electric current, Örsted did, and he coined the term thermoelectricity. Since these early discoveries, the Seebeck effect has been quantified as an electromotive force, $e_s$, which is proportional to the temperature gradient. That is

$$e_s = -x_s \nabla T,$$

where $x_s$ is the Seebeck coefficient.

The Peltier effect is a manifestation of the same phenomena which describes the absorption or release of a heat flux, $q_p$, at the junction of two dissimilar materials as a current flux flows across that junction. That is

$$q_p = \Delta s \rho j_k = T \Delta x j_k,$$

where $\Delta s$ and $\Delta x$ are the difference in the Peltier and Seebeck coefficients between the two materials, and $j_k$ is the electric current density. The Thomson effect describes the same phenomena, but in a bulk material subjected to a temperature gradient. As the Seebeck coefficient is a function of the temperature, there is a corresponding Seebeck gradient in the bulk material. This, in turn, leads to volumetric heat absorption (or release), $q_p = T \delta x / \partial T$.

These effects are fully described in the Boltzmann transport. This allows us to derive the thermoelectric effects and their transport coefficients from the Boltzmann transport equation (BTE). Indeed, if the force acting on the electronic population is the gradient in the potential energy, $E_e$, i.e., $F_e = \nabla E_e$, we can arrive at a description of the thermoelectric force, $F_e$. That is

$$F_e = -\nabla E_e + T(E_{e, p} + E_{e, n} - E_e) \nabla \frac{1}{T},$$

where $E_{e, p}$ is the kinetic energy of the electron. (Note that the potential energy of an electron is used here to describe the difference between the least energetic conduction state and the Fermi level.) The first term is the electrochemical force, while the second term is the thermal force. Thus, a temperature gradient must induce an electrochemical gradient in order to balance the forces acting on the electronic population. Onsager described the resulting transport coefficients, which can be re-arranged to describe the more typical Fourier, Seebeck, and Ohm’s law coefficients (thermal conductivity, $\kappa$, and Seebeck coefficient, $x_s$, and electrical conductivity, $\sigma$) based on $\Delta s$ and $\nabla E_e$ rather than on $\nabla (1/T)$ and $\nabla E_e$.

The discovery of the thermoelectric effects led to the development of thermoelectric applications in the mid twentieth century. Principal among these are temperature measurement using thermocouples (in an open circuit, the Seebeck effect creates a voltage difference between the dissimilar metals), temperature management and cooling using thermoelectric modules or coolers, and electricity generation using thermoelectric generators, as shown in Fig. 3. Initially, there was hope that thermoelectric generators might replace conventional heat engines. Quickly, however, it became apparent that there were limits to thermoelectric “heat engines” which restrict their efficiency to small fractions of the Carnot limit.

1. Efficiency and material metrics

Let us derive the thermoelectric efficiency and the material figure of merit for a thermoelectric generator in order to show how these limits arise. A thermoelectric generator is composed of a series of alternating $n$- and $p$-type semiconductors connected by a metallic junction. When a temperature gradient is imposed across the device, complementary electric fields develop in the semiconductors. (Note that $n$- and $p$-type semiconductors have Seebeck coefficients of opposite sign.) Thus, a thermoelectric current can be extracted from the device across this electric potential in order to generate power.

To begin our derivation, consider the heat equations at the hot and cold sides of the thermoelectric generator. These equations consider the Peltier heat absorbed or released at the $p$-metal-$n$ junctions, the Joule heat generated by the current, and the conduction of heat across the thermoelectric. That is

$$Q_h = J_e T_h \Delta x_s - J^2_e R_e / 2 + G(T_h - T_c),$$

$$Q_c = J_e T_c \Delta x_s + J^2_e R_e / 2 + G(T_c - T_h),$$

where $T_h$ and $T_c$ are the temperatures at the hot and cold sides, $\Delta x_s = x_p - x_n > 0$, where $x_n$ and $x_p$ are the Seebeck coefficients of the $n$ and $p$ type semiconductor. $J_e$ and $R_e$ are the electric current and resistance, and $G$ is the overall thermal conductance across the device. The Power output, $P_e$, equals the difference between $Q_h$ and $Q_c$. That is

$$P_e = J_e \Delta T \Delta x_s - J^2_e R_e,$$

where $\Delta T = T_h - T_c$, and the efficiency, $\eta_{TE}$, is

$$\eta_{TE} = \frac{P}{Q_h} = \frac{J_e \Delta T \Delta x_s - J^2_e R_e}{J_e T_h \Delta x_s - J^2_e R_e / 2 + G(T_h - T_c)}.$$
(Note that we have transitioned from discussing fluxes $j_a$ and $q$ to flows $j_e$, $Q$, and $P_r$ in order to avoid discussing the specific geometry of the TE cell.)

In a short-circuited thermoelectric generator, the current is maximized, $j_e = \Delta T x_0/R_a$, but no electric potential develops and the power vanishes. Conversely, in an open-circuited thermoelectric generator, the electric potential is maximized and the current and power vanish. Somewhere between the open-circuit and short-circuit conditions, then, the power and efficiency are maximized. Note, however, that they are not simultaneously maximized. Let $j_s = j/j_{ac}$. Then, the maximum power condition occurs at $j_s = 1/2$, as the current-voltage curve is linear, with $P_{r,\text{max}} = \Delta T^2 \Delta x_0^2/4R_a$.

Maximizing the efficiency is not as trivial. However, it can be analytically maximized to give

$$\eta_{\text{TE, max}} = \eta_c \frac{(Z_{\text{TE}} T + 1)^{1/2} - 1}{(Z_{\text{TE}} T + 1)^{1/2} + 1 - \eta_c}, \quad (8)$$

where $Z_{\text{TE}} T$ is the dimensionless figure of merit and $\eta_c = 1 - T/T_h$. 64 That is

$$Z_{\text{TE}} T = \frac{x_0^2}{R_c} \langle T \rangle, \quad (9)$$

where $\langle T \rangle = (T_h + T_c)/2$. This device figure of merit translates into an analogous material figure of merit, i.e.,

$$Z_{\text{TE}} T = \frac{x_0^2 \sigma_e \kappa_e} \kappa (T), \quad (10)$$

where $\sigma_e$ is the electrical conductivity and $\kappa$ is the thermal conductivity of the material, i.e., the n- or p-type semiconductor.

Figure 4 shows the predictions of Eq. (8) for variations in $T_h$ and $Z_{\text{TE}} T$. As is shown, a figure of merit around 10 is required to approach the efficiency of conventional heat engines. Still, improving upon the common $Z_{\text{TE}} T \approx 1$ and reaching an achievable value near $Z_{\text{TE}} T = 3$ would allow for reasonably efficient heat harvesting. Note that this derivation assumes that there is no thermal resistance between the metal and semiconductor, i.e., that the junctions are at a single temperature. Moreover, it assumes that there is equilibrium between and within the electron and phonon populations and that recombination or generation events do not occur within the thermoelectric. In general, Eq. (8) provides the maximum efficiency of the ideal thermoelectric generator, and many other effects reduce the efficiency of a real TE device. Indeed, Fig. 4 shows the efficiency of various TE devices, and the efficiency of these devices rarely exceeds 10%, and often falls well below the efficiency limit set by the $ZT_{\text{TE}}$ of the TE material.

2. Materials

The $Z_{\text{TE}} T$ can in principle approach infinity, which enables a thermoelectric efficiency approaching the Carnot limit. However, the electrical and thermal conductivity are not independent material properties. In the Wiedemann-Franz law, for example, the electronic contribution to the thermal conductivity, $\kappa_e$, is proportional to the electronic conductance. That is

$$\kappa_e = \frac{\pi^2 k_B^2}{3} T_e \sigma_e, \quad (11)$$

Similarly, the Seebeck coefficient and transport coefficients are related. Consider an n-type semiconductor, wherein the Seebeck coefficient grows as the Fermi level and conduction band edge separate. However, this exponentially reduces the number of conduction electrons, which in turn reduces the transport properties $\sigma_e$ and $\kappa_e$. As the phonon system always conducts some heat ($\kappa = \kappa_e + \kappa_p$), the connection between the electron thermoelectric properties precludes a large thermoelectric figure of merit. Indeed, the thermoelectric figure of merit for a p- and n-type semiconductor pair was limited to near or below one ($Z_{\text{TE}} T \ll 1$) for many years. From Eq. (8), this corresponds to heat harvesting with a limiting efficiency of around 10% when $\eta_c = 0.5$, only 20% of the Carnot limit. Only recently have scientists recorded or predicted substantially larger figures of merit (reaching $2 < Z_{\text{TE}} T < 3$). A $ZT$ of 3 would double the efficiency of a thermoelectric generator, enabling it to reach an impressive 40% of the Carnot limit.

Due to the interdependence of the power factor, $x_0^2 \sigma$, and thermal transport $\kappa$, this improvement has largely been based on the suppression of the phonon thermal conductivity, e.g., through the nanostructuring or filling of thermoelectric materials. 89 For thermoelectric power generation applications, however, the power factor can be crucial, as it limits the power generation density. In many applications, this is as important as the efficiency. Thus, some researchers have focused and succeeded in improving the power factor. 90,96,97,98,99

Additionally, researchers have proposed the use of the spin Seebeck effect (SSE) to create novel thermoelectric materials. 70 In combination, these efforts have led to drastic improvements in the thermoelectric figure of merit of a number of thermoelectric materials as well as in the discovery and optimization of metal thermoelectric materials, as shown in Fig. 5a. Indeed, a $Z_{\text{TE}} T \rightarrow 3$ now seems feasible, whereas drastically exceeding $Z_{\text{TE}} T = 1$ seemed impossible for many years.
However, the thermoelectric figure of merit depends substantially on the operating temperature, as shown in Fig. 5(b). Larger temperature differences are required for efficient heat harvesting (Fig. 4), this often leads to an effective figure of merit which is drastically reduced from that shown in Fig. 5. In particular, the large figure of merit of Cu2Se is misleading, as this figure of merit arises very sharply during a phase transition. Moreover, the record \( Z_\text{TE} T \) is often reported for only a single, \( n \)- or \( p \)-type semiconductor, not for the thermoelectric package which includes both semiconductor types.

While thermoelectric generators face inherent limitations, improving thermoelectric materials and enhancing thermoelectric power generation continue to receive substantial effort and attention from the community. Moreover, thermoelectrics remain a major technology for precise temperature control, enhanced electronics cooling, and radioisotopic thermal power generation in space applications. If the power factor and figure of merit of thermoelectric materials continue to climb, thermoelectric power generation may become attractive in additional waste-heat recovery applications.

The field of thermionic energy conversion is a technology with a similarly long history. The thermionic generator is also beginning to attract renewed attention, much like the thermoelectric generator. Theoretically, the thermionic converter can achieve a much more impressive conversion efficiency. However, it also faces major obstacles. Let us discuss this alternate solid-state heat harvesting technology.

B. Thermionics

The thermionic generator directly converts heat into electricity, as shown in Fig. 6. However, the energy barrier is the work function of the material, \( \varphi_w \), rather than the change in the chemical potential induced by a temperature gradient, \( \Delta \mu \Delta T \). That is, heat in a thermionic is used to excite electrons in the emitter to energies above its work function such that they evaporate, i.e., escape the emitter surface, travel across a vacuum, and then reach a cold collector which captures them. The resulting thermionic current density can be extracted across some applied electrical potential, \( \Delta \varphi/a \), in order to generate power.

Much like the thermoelectric effects, the thermionic effect has a long history, originating with the experiments of Elster and Geitel in 1882. They heated a platinum wire using a current and then recorded the accumulation of some charge on an opposing electrode.

FIG. 5. (a) Improvement in the maximum thermoelectric figure of merit for representative classes of thermoelectric materials. (b) Thermoelectric figure of merit for variations in the temperature. Data from following Refs.: PbTe,15,56,74,150 PbSe,217,240 PbS;77,218,241 Cu2Se,109,110,147,242 Cu2S;70,235 SnSe,40,151,238,239 Mg3Sb2,233 SiGe,7,82,228 BiCuSeO,236,237 Bi2Te3,94,157,222 Cu(In,Ga)Te2,118,156,232 CoSb3 (Skutterudites),165,174,191 InSb,27,223 and half-Heusler compounds;55,83,225 Review.68,190 (a) Reproduced with permission from J. He and T. M. Tritt, Science 357, 1369 (2017). Copyright 2017 American Association for the Advancement of Science.

FIG. 6. (a) An illustration of the thermionic cell and temperature profile, which steps between the cathode and anode. (b) An energy diagram of the thermionic cell and the electronic populations within it. High temperatures in the cathode excite electrons above its work function. These electrons escape and flow into the anode, where they are captured (as it has a less energetic work function). These electrons carry current and heat. Additional heat is transferred via the thermal radiation of photons. (c) Vacuum thermionics suffer from the accumulation of a space charge which inhibits thermal emission (\( \varphi_{sc} \)), while (d) ion-plasma thermionics suffer from the voltage drop across the ignited arc (\( \Delta \varphi_{arc} \)).

Published under license by AIP Publishing

Appl. Phys. Rev. 6, 021305 (2019); doi: 10.1063/1.5031425
This quickly led to the development of the first thermionic circuits by Edison.\textsuperscript{43} (Indeed, the thermionic effect described above is called the Edison effect.) This ultimately led to the identification of the electron as a negatively charged particle by Thomson in 1897.\textsuperscript{207,208} Soon, an equation for the thermionic current density was derived by Richardson using the kinetic theory of gases.\textsuperscript{162} Dushman would later derive the modern form of the Richardson-Dushman equation.\textsuperscript{41} That is

\begin{equation}
\begin{aligned}
j_e &= A' \cdot A \cdot T^2 \exp \left( - \frac{q_w}{k_B T} \right), \quad (12)
\end{aligned}
\end{equation}

where the Richardson constant \( A' \) is

\begin{equation}
A' = A' \cdot m_e \frac{k_B^3 e}{2 h^3}, \quad (13)
\end{equation}

and \( m_e \) is the electron mass and \( A' \) is a material and surface dependent modification of the Richardson constant, hereafter called the electron emissivity. (This constant has its own history, beginning with Richardson in 1914.\textsuperscript{163}) Using the thermionic effect to harvest heat in a thermionic generator was first proposed in 1915 by Schlichter.\textsuperscript{178} Let us model its operation and derive its efficiency.

### 1. Efficiency and material metrics

The thermionic conversion efficiency may be easily derived under the following assumptions. First, the only mechanisms through which heat leaves the cathode are the thermal emission of photons (thermal radiation, \( Q_r \)) and electrons (thermionic emission, \( Q_e \)). Second, the anode absorbs all of the incoming electrons and photons. Third, the anode emits a negligible quantity of heat and current in comparison to the cathode. Consider a thermionic with a (hot) cathode at temperature \( T_c \) and a (cold) anode at temperature \( T_a \). Under our assumptions, the net current equals the current leaving the cathode, i.e., that given in Eq. (12). Each electron carries, at minimum, kinetic energy equal to the work function of the cathode, \( \varphi_{w,c} \). Thus, the heat flux associated with the current is

\begin{equation}
q_e = e_c \varphi_{w,c} j_e = e_c \varphi_{w,c} A' \cdot A \cdot T^2 \exp \left( - \frac{\varphi_{w,c}}{k_B T} \right), \quad (14)
\end{equation}

where \( A' \) is the electron emissivity of the cathode. Thermal radiation increases the total energy flux. Under our assumptions, the net radiative heat flux, \( \omega_r \), is equal to the radiative heat flux leaving the cathode. That is

\begin{equation}
\omega_r = e_c \cdot \varepsilon_s \cdot \sigma_B \cdot T_0^4, \quad (15)
\end{equation}

where \( \varepsilon_s \) is the emissivity of the cathode and \( \sigma_B \) is the Stefan-Boltzmann constant.

In an ideal thermionic generator, the largest electric potential which can be extracted per charge is the difference between the cathode and anode work functions, \((\varphi_{w,c} - \varphi_{w,a}) / e_c \). Indeed, this difference is the driving electric potential of the thermionic. (In practice, this is reduced substantially by the accumulated space charge unless a specific strategy is used to mitigate this phenomenon.) Thus, the maximum power density produced by a thermionic cell is

\begin{equation}
p_e = e_c (\varphi_{w,c} - \varphi_{w,a}) = e_c A' \cdot A \cdot T^2 \exp \left( \frac{\varphi_{w,c}}{k_B T} \right) \left( \frac{\varphi_{w,c} - \varphi_{w,a}}{k_B T} \right), \quad (16)
\end{equation}

where \( A' \) is the electronic emissivity of the cathode. Thus, the maximum efficiency, \( \eta_{TI} \), is

\begin{equation}
\eta_{TI} = \frac{P}{Q_e + Q_r} = \frac{e_c A' \cdot A \cdot T^2 \exp \left( \frac{\varphi_{w,c}}{k_B T} \right) \left( \frac{\varphi_{w,c} - \varphi_{w,a}}{k_B T} \right)}{e_c A' \cdot A \cdot T^2 \exp \left( \frac{\varphi_{w,c}}{k_B T} \right) + \varphi_{w,c} \cdot \exp \left( \frac{\varphi_{w,c}}{k_B T} \right)}, \quad (17)
\end{equation}

This expression identifies three regimes: (i) thermal radiation dominates \((\varphi_{w,a} > k_B T)\), and the efficiency is negligible; (ii) thermal radiation and electron emission are comparable \((\varphi_{w,a} < k_B T)\), and the efficiency depends strongly on \( T_c \); and (iii) electron emission dominates \((\varphi_{w,a} < k_B T)\), and the efficiency saturates at \( \eta_{TI} \rightarrow 0 \) as \( \varphi_{w,a} \rightarrow 0 \). Second, if \( \varphi_{w,a} \approx 0 \), then the adverse, anode-to-cathode current becomes substantial and our approximations do not hold. Third, if \( \varphi_{w,a} \) becomes too large, the material will melt before the thermionic can reach regime (iii). Instead, it will remain in regime (i) and convert most of the heat into thermal radiation. Thus, \( \varphi_{w,a} \) cannot in practice be much smaller than \( \varphi_{w,c} \).

In practice, \( \varphi_{w,a} \) is the limiting material metric. Let us derive the maximum efficiency for a given anode work function and cathode temperature. That is, let us maximize \( \eta_{TI} \) with respect to \( \varphi_{w,c} \). This maximization gives the cathode work function which maximizes the efficiency, \( \varphi_{w,c,max} \), as well as the maximum efficiency, \( \eta_{TI,max} \). That is

\begin{equation}
\varphi_{w,c,max} = \varphi_{w,a} + k_B T_e [1 + W(z)], \quad (18)
\end{equation}

\begin{equation}
\eta_{TI,max} = 1 - \frac{\varphi_{w,a}}{\varphi_{w,a} + k_B T_e W(z)}, \quad (19)
\end{equation}

where

\begin{equation}
z = \frac{A' \cdot A}{e_c \cdot \varepsilon_s \cdot \sigma_B \cdot T_0^4} \cdot \frac{\varphi_{w,a}}{k_B T_e}, \quad (20)
\end{equation}

and \( W(z) \) is the principal solution for \( w \) in \( z = \exp(w) \).

Note that \( z \) is not a material figure of merit; again, \( \varphi_{w,a} \) is essentially the material figure of merit, with \( \varphi_{w,a} \approx 0.5 \) eV maximizing conversion efficiency in a thermionic converter for a room temperature anode.\textsuperscript{66,67} As shown in Fig. 7, this enables heat harvesting at large fractions of the Carnot limit. Additionally, an ideal thermionic material should have good heat and electron transport and a high melting temperature, such that the electrons can overcome the cathode work function.\textsuperscript{66,67}

Figure 7 also shows that experimentally, TI devices do not harvest energy efficiently. This can largely be traced to few major factors: The major application considered has so far been for space-based generation, wherein efficiency is not the concern (the power-to-weight considerations are most important) and reliability is of the utmost importance; it is difficult to manufacture nanometer-size vacuum gaps (necessary to overcome the space charge problem); and it is extremely...
difficult to manufacture these gaps over a large area and ensure that the gap is not compromised over time (during which the device may undergo massive temperature changes). For these reasons, among others, TI devices have historically relied upon higher work-function materials which overcome the space charge issue by using cesium as the working vacuum gap. Before proceeding to survey some of the material candidates which may be used as low work-function anodes in the TI devices of the future (wherein manufacturing challenges have been overcome).

2. Materials

For a long time, the leading method used to create a thermionic anode was to cover a refractory metal which can safely reach high temperatures in an alkali-metal atom (typically cesium) which reduces the work function of that metal. When the Cs atoms become an ignited plasma, these metals can reach a work function around 1.5 eV.\(^1,86,144,221\) While low, this large of a work function limits the thermionic conversion efficiency to small fractions of the Carnot limit. Experimental TI device efficiencies are shown for the following TI devices: TI/AMTEC (alkali metal thermal-to-electric converter),\(^88\) SET (Solar Energy Thermionic), and HPALM (High Power Advanced Low Mass).\(^3\) The TI/AMTEC device uses a secondary heat-engine (the AMTEC) to achieve the highest efficiency of the three devices.

For example, diamond surfaces terminated with hydrogen show a negative electron affinity.\(^38,134,205\) This electron affinity and the associated work function are strongly correlated with the donor levels of the donor or acceptor level of the dopant. This indicates that electrons moving from the donor level to the conduction band are emitted from a properly passivated surface. As shallow donor levels can exist near the conduction band, a very small work function may be created through the use of appropriate dopants. Indeed, phosphorous-doped diamond with a hydrogen passivated surface exhibits a work function of around 0.67 eV, only slightly larger than the gap between the phosphorous donor level and the conduction band (\(\sim 0.6\) eV).\(^96\)

This ultralow work function is very exciting for thermionic converters, and appears to be stable up to 800 °C.\(^50\) At 1500 K, an ideal thermionic converter utilizing a phosphorous-doped diamond anode can reach an efficiency of up to 58% and 74% of the Carnot limit. This is a substantial improvement upon the current efficiency of thermoelectric generators. However, it has not been realized experimentally, as of yet. Indeed, one of the advantages of a thermoelectric generator is its simplicity. In comparison, an efficient thermionic converter requires much more than a well-tuned material. With the thermionic and thermoelectric generators briefly reviewed, let us discuss a more quantum-mechanical and nanoscale device: the laser cooler, which provides in-situ cooling rather than ex-situ power generation.

C. Laser cooling

In laser cooling, or optical refrigeration, anti-Stokes fluorescence is used to cool the target. In contrast with the thermionic and thermoelectric energy converters, which can be understood from a macroscopic perspective, laser cooling is a phenomenon which requires an atomistic, quantum-mechanical perspective, as shown in Fig. 8. Unsurprisingly, then, its history is more recent. Still, the process was
conceived of early in the twentieth century, well before the laser was invented.\textsuperscript{138} Initially, it was believed to violate the second law of thermodynamics.\textsuperscript{212} Quickly, however, Landau assigned an entropy to both the bandwidth and solid angle at which light was emitted.\textsuperscript{29} Nearly 50 years later, Epstein would demonstrate the first net-cooling, anti-Stokes effect in a solid, by illuminating rare-earth ion doped glass with a laser.\textsuperscript{54} Slightly earlier the first laser cooling was achieved through the Doppler cooling of dilute gasses.\textsuperscript{85}

These initial Doppler cooling experiments were conducted at very low temperatures and densities, where the kinetic energy of the gas was stored in its translation. The Doppler-shifted atomic transitions in such a gas interact with the laser, such that the laser exerts an effective drag on the atoms.\textsuperscript{29} Through this mechanism, researchers have been able to cool dilute gasses into the $\mu$K range, create Bose-Einstein condensates,\textsuperscript{54} and study fundamental atomic physics. However, Doppler cooling has limited practical applications.\textsuperscript{183}

The cooling of interacting media, e.g., a dense gas, is similar in nature but with some crucial differences: The laser direction is no longer important (i.e., the counter-propagating requirement of Doppler cooling), and the interacting media re-emit photons. These differences arise because the laser does not interact directly with the atoms. Instead, it couples with the interatomic interactions where the thermal energy is largely stored, e.g., their collision\textsuperscript{11,214} or vibration.\textsuperscript{51}

In a solid between 10 and 400 K, for example, the kinetic energy is stored in the vibrational degrees of freedom, and the basic, laser cooling scheme is as follows. A host medium is doped with ions exhibiting a ground-state dipole transition of energy, $E_g$. A laser is tuned to excite an electron using this transition, $h\omega_p = E_g$, whereupon the electron absorbs a vibrational quanta (phonon) with energy $h\omega_p$. Finally, the excited electron emits a higher energy photon, $h\omega_y = E_y + h\omega_p$, as it returns to the ground state.\textsuperscript{51} This process, anti-Stokes fluorescence, is shown in Fig. 8(b).

While anti-Stokes cooling is possible in liquids\textsuperscript{31} and gasses\textsuperscript{39} (the first experimental demonstration), the laser cooling of solids is of primary interest. Indeed, solid-state devices represent the majority of practical applications. While inefficient, laser cooling should outperform thermoelectric cooling in the nanoscale regime (where contact resistances and nonequilibrium phenomena prevent efficient operation) and the low-temperature, low-power regime (where efficient thermoelectric materials do not yet exist).\textsuperscript{134} Let us discuss why optical refrigeration is inefficient.

1. Efficiency and material metrics

In optical refrigeration, the input energy is $\hbar \omega_{ph} + \hbar \omega_y$, while the heat removed is only $\hbar \omega_y$.\textsuperscript{54} Thus, the efficiency, $\eta_{opt}$, is limited by

$$\eta_{opt} \leq \frac{\hbar \omega_y}{\hbar \omega_{ph} + \hbar \omega_y}. \quad (21)$$

We can also write the coefficient of performance, $\eta_{CP}^{\omega_{ph}}$, which corresponds to this efficiency for a reversible optical refrigeration process,

$$\eta_{CP}^{\omega_{ph}} = \frac{\hbar \omega_y}{\hbar \omega_{ph}}. \quad (22)$$

For a formal derivation of these equations, we refer the interested reader to Ref. 137. Note that at a given temperature, the highest frequency, occupied phonon modes will have an energy comparable to the thermal energy, $k_B T$. Thus, we can write\textsuperscript{51}

$$\eta_{opt} \approx \frac{k_B T}{E_g}. \quad (23)$$

At moderate temperatures, the thermal energy tends to be much smaller than a ground-state dipole transition ($\sim 1$ eV). Thus, the cooling efficiency is limited to only a few percent. Still, this enables the use of optical refrigeration in practical cooling applications.\textsuperscript{183}

There are a number of losses which greatly reduce this efficiency if they are not properly managed. These losses include the nonradiative relaxation of an excited electron, as shown in Fig. 8(c), Stokes fluorescence, as shown in Fig. 8(d), the re-emission of a photon before the excited electron absorbs a phonon, and the re-absorption of the anti-Stokes fluorescence (which increases the chance that the other losses occur). Often, the efficiency is reformulated to reflect these inefficiencies\textsuperscript{183} using a simple four-level model of the energies (e.g., to represent a rare-earth doped crystal). The resulting efficiency equation is

$$\eta_{opt} = \eta_{QF} \frac{h \omega_y}{E_g} - 1, \quad (24)$$

where $\eta_{QF}$ is the quantum efficiency, i.e., the fraction of excited electrons which absorb a phonon before relaxing to the ground states, and $\eta_{abs}$ is the absorption efficiency, i.e., the fraction of incoming photons from the laser which are absorbed. Thus, the ideal efficiency is reached when $\eta_{QF} \eta_{abs} \rightarrow 1$. This model also holds for semiconductor targets, provided the quantum efficiency accounts for the additional recombination mechanisms and the trapping of fluorescence within the semiconductor.

In general, as the laser pump becomes more intense (populating the excited electron states and reducing the population of the ground-state electron states), and the device becomes cooler (reducing the population of phonons), these losses begin to dominate and $\eta_{opt} \rightarrow 0$. Thus, the quantum efficiency limits the temperature to which a material can be cooled. Conversely, at high temperatures the nonradiative decay processes [Fig. 8(c)] will begin to dominate, the quantum efficiency will fall, and laser illumination will heat the target. Indeed, a simple model of the total (nonradiative and radiative) relaxation rate is

$$\frac{\gamma^{\omega_{ph}}(T, N_p)}{\gamma^{\omega_{ph}}(T = 0)} \approx \frac{\gamma^{\omega_{ph}}(T = 0)}{1 - \exp(-\hbar \omega_p/k_B T)} \right)^{-N_p}. \quad (25)$$

where $N_p \approx E_p/h \omega_p$ is the number of phonons required to relax from an excited state to one of the ground states.\textsuperscript{87} This equation describes a transition from radiative to nonradiative decay. Thus, laser cooling does not become efficient at high temperatures, as Eq. (23) suggests. Instead, it vanishes.

Let us briefly describe a simple model to illustrate the temperature dependence of optical refrigeration. Let the antistokes process occur at a rate proportional to the phonon occupancy, i.e.,

$$\frac{\gamma^{\omega_{ph}}(T = 0)}{\gamma^{\omega_{ph}}(T)} = \exp(-\hbar \omega_p/k_B T). \quad (26)$$

where we have assumed nondegenerate statistics for simplicity and to match the approach of Eq. (25). Then, the quantum efficiency (the fraction of absorbed photons which contribute to the antistokes cooling) is...
The efficiency of optical refrigeration in a toy material with nondegenerate system, both temperature of the fluorescence (the hot reservoir). In a reversible, however, the “Carnot” limit to the coefficient of performance is which reduces the high temperature efficiency.

For the heating effect of the nonradiative relaxation mechanisms, the resulting efficiency is shown in Fig. 9. Note that we do not account for the nonradiative relaxation mechanisms when the phonon mode is populated.

Additionally, it can be shown that optical refrigeration is limited by something resembling a Carnot efficiency. In laser cooling, however, the “Carnot” limit to the coefficient of performance is given by

\[
\eta_{\text{COP,LC}} \leq \frac{T_c}{\frac{k_B}{h} - \frac{T_c}{T_h}}.
\]

where \(T_c\) is the temperature of the antistokes transition and \(T_h\) is the temperature of the fluorescence (the hot reservoir). In a reversible, nondegenerate system, both \(T_c\) and \(T_h\) can be replaced by the Bose occupation statistics in order to recover Eq. (22). That is, \(T_c = \frac{\hbar \omega_p}{k_B}\) and \(T_h = \frac{\hbar (\omega_p + \omega)}{k_B}\). (In a more general optical refrigeration process, flux temperatures are used to quantify the ratio of energy to entropy generation in the laser and target and derive a Carnot limit.)

This perspective reveals the source of the inefficiency in optical refrigeration: The waste heat, i.e., the fluorescence, is emitted with a temperature in the thousands of Kelvin. This contrasts with a typical thermal cycle, which uses most of the incoming energy and rejects it as waste-heat near 300 K. As substantial energy remains in the fluorescence, a few schemes have been proposed which recovered it.

The efficiency equations discussed above do not clearly indicate the material metrics which describe an effective laser cooling medium. However, researchers have derived or investigated the material properties which determine the ability of a material to succeed in a laser cooling application. Often, these are differentiated in their focus on Rare-earth doped crystals or on semiconductors. However, these systems are much more complex than the thermionic or thermoelectric generators. Thus, definitive and simple expressions for a material figure of merit do not exist. Still, we can present general material properties required to maximize \(\eta_{\text{COP,LC}}\).

(i) The electron-phonon coupling (EPC) must be strong, (ii) the phonons which couple should have an energy comparable to \(k_B T\) at the target cooling temperature, and (iii) the energy spread of the ground and excited states should be small in a rare-earth material, or, similarly, there should be a high density of low-energy valence and conduction states. (i) and (ii) ensure that electrons absorb phonons quickly or phonon-assisted photon absorption is fast, while (iii) ensures that the ground states are saturated at low temperatures (rare-earth) or that the excited states are not saturated under strong illumination (semiconductor). Additionally, (iii) increases the absorptivity of a semiconductor system.

2. Materials

Many rare-earth doped materials have been successfully cooled. However, the Yb\(^{3+}\) doped crystal YLF (YLiF\(_4\)) has arisen as the premier rare-earth system, supplanting the Yb\(^{3+}\) doped glass ZBLANP (ZrF\(_4\)-BaF\(_2\)-LaF\(_3\)-AlF\(_3\)-NaF-PbF\(_2\)), wherein the first net-cooling of a solid was achieved. Indeed, YLF has now been cooled well into the cryogenic temperature range.

In contrast, semiconducting materials have only very recently been successfully cooled. The large parasitic absorption in, e.g., GaAs heterostructures, has largely prevented their success. In 2013, however, researchers cooled CdS nanostructures at room temperature by 40 K, a remarkable advance in the field. Theoretical investigations of CdS, GaN, and GaAs suggest that CdS succeeds where GaAs and GaN fail for two reasons. First, the electronegativity of the constituent elements differs more substantially in a II–IV semiconductor than in a III–V semiconductor, given the same average period of those elements. Thus, the atomic displacements caused by a phonon polarize the crystal more substantially, and (ii) the EPC is therefore stronger in a II–IV semiconductor than in a III–V semiconductor. Note that the EPC in GaN is larger than that in CdS, because it includes Nitrogen, a second-period, high-electronegativity element. Indeed, the EPC (and electronegativity difference) correlates not only with the separation of the element families, but also with the average period of the constituent elements, as shown in Fig. 10. GaN, however, is predicted to produce a worse target than CdS for laser-cooling applications. This is because (ii) the optical phonon energy of GaN (~90 meV\(^{5,103}\)) is quite large.
compared to that of CdS (~38 meV) and the thermal energy at room temperature (~26 meV). Indeed, the optical phonon energy is also strongly correlated with the average period of the constituent elements, as shown in Fig. 10. Thus, II–IV semiconductors are predicted to be much more efficient targets than III–V semiconductors at room temperature. Still, substantial work remains to discover and optimize semiconductor targets and nanostructures for efficient laser cooling.

D. Modern explorations

In addition to the heat harvesters discussed above, there have been a number of modern explorations of heat harvesting. The goal, as it has been since the early twentieth century and the invention of the thermoelectric generator, is to create solid-state heat engines, to recycle waste heat, and to cool devices. In the modern context, we can add that nanoscale thermal management is a new goal which reflects the rapidly shrinking devices which produce staggering heat fluxes across nanoscale heterostructures. While optical refrigeration can be used to achieve nanoscale cooling, it is not efficient. Here, we will survey a few of the modern devices which address these goals.

1. Thermophotovoltaics

The thermophotovoltaic (TPV) is an example of a more modern, solid-state heat harvesting technology. While it was proposed in 1979, recent manufacturing, computational, and theoretical advances have spurred renewed interest in the twenty-first century. In the TPV cell, a hot emitter converts heat into the selective thermal radiation of short-wavelength photons. Then, a photovoltaic cell converts these photons into electricity, as shown in Fig. 11.

While solar-photovoltaics (SPV) have a larger Carnot limit than the TPV, as the sun radiates at 5778 K and emitter temperatures are limited to below a few thousand K, much of the black-body radiation does not generate electricity (\( h\nu < \Delta E_g \)) or also generates heat (\( h\nu > \Delta E_g \)). In a TPV, however, the thermal radiation of the emitter can be tuned such that long-wavelength radiation is suppressed (selective emission), and the photovoltaic can be coated such that those long-wavelength photons which are emitted are reflected back to the emitter. In these ways, the TPV efficiency can surpass the SPV efficiency, despite the smaller Carnot limit. Indeed, a TPV can surpass the...
should be minimized. Additionally, the emitter should not oxidize at high temperatures, as shown in Fig. 12.

Shockley-Queisser limit on the conversion efficiency of a SPV (33%), as shown in Fig. 12. Unfortunately, it is difficult to derive a simple analytical model of the efficiency of a thermophotovoltaic cell, $\eta_{TPV}$, as the emissivity of the emitter and the reflectivity of the absorber depend strongly on the photon wavelength. We can make general observations, however, which guide the design of TPV materials. First, the emitter must withstand very high temperatures, such that the Carnot limit is maximized and short-wavelength photons are preferentially generated. Second, the emissivity should be as small as possible for photons less energetic than the band gap. Third, the photovoltaic band gap should be approximately equal to 4\(k_B T_{\text{emitter}}\) where \(T_{\text{emitter}}\) is the emitter temperature.

The emitter materials are usually limited to those which can withstand very high temperatures, i.e., refractory materials, which do not oxidize. This leaves W, Cr, Ta, and Mo. Other refractory materials which have been proposed include ZrC, TiC, VC, TaSi_2, and NbSi_2. The selective emission of small-wavelength photons from these emitters can be realized as follows. First, the photovoltaic and emitter elements are brought so close that near-field radiation dominates. That is, the surface-plasmon-polaritons (SPP) of the emitter and photovoltaic coating interact with each other in order to transfer heat. This SPP interaction can be tuned in order to selectively emit a narrow spectrum of energies. Then, the band gap of the photovoltaic is tuned to absorb the narrow spectrum of interacting SPP modes without losing substantial energy. Alternatively, the emitter may be nanostructured as shown in Fig. 11(a) in order to create a band gap in the SPP structure, such that the emissivity vanishes within this gap. Substantial interest in such strategies has led to a proliferation into the design of a selective emitter (or absorber). References 8, 30, 244, and 245 present a more comprehensive review of these methods and their theory.

If the sun is used to heat the emitter, these mechanisms can be said to reshape the solar spectrum in order for a photovoltaic to surpass the Shockley-Queisser limit. Alternatively, the emitter may be heated by a radioactive isotope (in space applications) or by a furnace. In any of these cases, impressive efficiencies are achievable. Indeed, a SPP “squeezed” TPV can, in the ideal and optimized case, nearly reach the Carnot limit, as shown in Fig. 12. Experimentally, devices have reached nearly a third of the ideal TPV limit. Numerous inefficiencies explain this discrepancy, e.g., the absorption of free-carriers in the electrode or heat losses through other channels. Still, these experimental results show that the TPV substantially outperforms current TI and TE generators.

2. Photon-enhanced thermionics

The photon-enhanced thermionic uses photon-enhanced thermionic emission (PETE) in order to increase the work function of the cathode (and thus the operating voltage) without eliminating the current. In PETE, incoming solar photons are used to boost electrons over one potential energy barrier, the band gap, while the thermal energy is used to excite electrons over the remaining work function, i.e., the electron affinity $\phi_{\text{ek}} = \phi_{\text{ek}} - \Delta E_{\text{tg}}$ as shown in Fig. 13. Thus, the work function of the cathode can be substantially larger than in a conventional thermionic converter without a corresponding drop in the current one would expect from Eq. (12).

Before continuing, let us take a moment to introduce the key features of semiconductor devices which we have not had to delve into until now. First, materials have a Fermi energy, or Fermi level, below which all electronic states are filled at 0 K. In an intrinsic (undoped) semiconductor, the Fermi level lies just above the most energetic valence state, within the electron bandgap. As the temperature rises, electrons begin to populate the conduction states above the bandgap, and the chemical potential moves into the bandgap such that the Fermi-Dirac statistics predicts that an equal density of valence holes, $n_v$ and conduction electrons, $n_c$, exists. (Often, these terms, the Fermi level, Fermi energy, or chemical potential, are used interchangeably.)

![FIG. 12. The efficiency of thermo-photovoltaic cells optimized using surface-plasmon-polaritons (SPP-TPV) which couple to the absorber band gap. Free-carrier absorption in the contact (e.g., by an Ag electrode) can lower the efficiency. The Carnot limit, experimental device efficiencies, and efficiency of other thermal cycles are shown for comparison.](image)

![FIG. 13. (a) A photon-enhanced thermionic generator and (b) its energy diagram. Strong illumination generates a large population of conduction electrons which increases the quasi-Fermi level and enhances the thermionic current. This effect enables the use of emitters with a large work function without drastically reducing the thermionic current density.](image)
When a semiconductor is illuminated by photons with an energy greater than the bandgap, additional conduction electrons and valence holes are created. We can define quasi-Fermi levels for the electrons, $E_{F,n}$ and holes, $E_{F,p}$ which allow us to calculate the new electron and hole densities, $n'_e$ and $p'_h$, using Fermi-Dirac statistics. Note that to use Fermi-Dirac statistics, we must assume that the new charge carriers equilibrate with the existing electron and hole populations. Generally, the electron-electron and electron-phonon interactions are sufficiently fast to ensure that this is an accurate approximation on the scale of a device. As the quasi-Fermi energy levels separate ($\Delta E_{F,n-p} = E_{F,n} - E_{F,p}$), i.e., as the electrons accumulate, so too does the net rate of recombination. Indeed, as illumination continues, there are more conduction electrons which want to fall back to the valence band and more empty states in the valence band into which these conduction electrons can fall. Consider, the band-to-band model for the net recombination, $-\dot{n}_i = n_i < n_p > - n'_e p'_h$. Eventually, the recombination will balance with the generation caused by the illumination and $\Delta E_{F,n-p}$ will reach its steady-state value. Note that the stronger the illumination, the larger $\Delta E_{F,n-p}$ can grow.

In a photovoltaic or PETE device, the goal is to separate and then extract the photogenerated carriers before they recombine and are lost as heat or fluorescence. Typically, this is accomplished by doping the semiconductor so that the photon absorbing surface is p- or n-type (i.e., the Fermi level lies near the valence or conduction bands, respectively, such that there is an excess of positive or negative charge carriers) and the remainder of the device is of n- or p-type (whichever has better transport properties). An electric field then develops between these regions as the excess electrons (holes) diffuse from the n (p) side to the p (n) side, leaving behind the positively (negatively) charged impurities which donated (accepted) electrons to (from) the semiconductor. This field separates the photogenerated carriers and a metal contact is used to collect the electrons (holes) generated nearby in the n-type (p-type) photon absorber. On the other side of the diode, the holes (electrons) must travel through the p-type (n-type) semiconductor until it is emitted (in a PETE) or collected (in a PV). Note that if these carriers recombine before reaching the emitter or collector, they are lost as heat.

Returning to our focus on the PETE, consider the strong illumination of a nondegenerate p-type semiconductor. In this situation, a large number of conduction electrons are generated, increasing the electron density from the equilibrium value, $n_e$, to $n'_e$. Equivalently, we can say that the quasi-Fermi level rises drastically, $E_{F,n} = E_0 + k_B T_e \ln (n'_e/n_e)$. If we follow a derivation of the Richardson-Dushman equation using this quasi-Fermi level, we arrive at a current which quantifies the effects of illumination on the thermionic current density.179 That is

$$j_{PETE} = A^* A_i T_e^2 \exp \left\{ \frac{-[\phi_e - (E_{F,n} - E_F)]}{k_B T_e} \right\}$$

$$= e n'_e \langle u_e \rangle \exp \left\{ \frac{-F_e}{k_B T_e} \right\}$$

(29)

where $\langle u_e \rangle$ is the average electron velocity perpendicular to the cathode surface. That is, the work function is effectively reduced by the increase in the quasi-Fermi level. Alternatively, the electron affinity becomes the new work function if we alter the traditional Richardson constant. Note, however, that the work function is only effectively reduced when we consider the current. The voltage is still limited by the difference between the work function of the cathode and anode.176

Thus, the operating voltage of a PETE can be increased without reducing the thermionic current density.

One might ask why the work function is not reduced by the energy of the incoming photons or by the bandgap, rather than by the quasi-Fermi level. To answer this, we must return to our definition of the quasi-Fermi level, wherein we assumed that the photogenerated electrons quickly returned to a quasi-equilibrium state that could be described by the Fermi-Dirac statistics. If the generated, nonequilibrium electrons were to be collected before they could relax and equilibrate, it is possible to extract the energy of the initial photon. The hot-carrier photovoltaic cell was a proposal by Nozik in 1982 to do just this in a photovoltaic cell.171,172 It has proven very difficult to realize the benefits of this scheme, as the hot electrons tend to relax very quickly. Recently, however, progress has been made in frustrating the relaxation and extracting hot carriers. Thus, hot-carrier photovoltaics are receiving renewed interest.178

In a PETE, if the photogenerated carriers can reach the thermionic emitter before relaxing, a similar benefit can be seen and the work function may be reduced by the photon energy rather than the quasi-Fermi level. However, this optimization has yet to be explored as the PETE is a very new technology.

Indeed, Schwede et al. first described and simulated a concentrated-solar PETE system in 2010.179 Their simulations predicted that a PETE could reach a conversion efficiency of nearly 40% at $\times 1000$ suns concentration for an anode work function of 0.9 eV. Furthermore, they suggested that the waste-heat from this system could be harvested to boost the overall efficiency over 50%.179 Both these figures are substantial improvements upon a typical thermionic converter with a similar anode work function. (In PETE, as in TI, the efficiency increases drastically if the anode work function is lowered.)179

The PETE efficiency is reduced by two competing mechanisms: The radiative recombination of excited electrons before they reach the surface of the cathode, and the transmission of photons through the cathode before they can be absorbed. Thus, strong absorptivity and good electron transport are required, while the thickness of the cathode must be optimized. In subsequent studies, Schwede et al. suggested that a photon-absorber/electron-emitter heterostructure can be built in order to limit the competition between these mechanisms.180

Since these foundational studies, a number of theoretical PETE studies have been presented.180,182,191 These have investigated the thermodynamics,196 theoretical maximum conversion efficiency196 (70% for a combined cycle), and PETE materials.191 Varpula et al. found that InP and GaAs can reach an efficiency between 20% and 25%, leaving substantial room for improvement.190 Si is predicted to perform substantially worse, as its indirect band gap leads to poor photon absorptivity. Indeed, this appears to be the crucial parameter190 and strategies to enhance absorption are key to the success of PETE technologies.

A collection of efficiency predictions are presented in Fig. 14. There are a few noteworthy differences between the TI and PETE efficiency curves. First, a PETE can reach a substantial efficiency at relatively low cathode temperatures, unlike a TI, which requires extremely high temperatures. Second, the PETE exhibits a sharp increase in the efficiency as the thermal energy approaches the effective energy barrier, $\gamma$, and thermionic emission is unlocked. (In a TI, the optimal efficiency is reached at much higher temperatures).
energy barrier is substantially larger, $\phi_{\text{PETE}} \gg \chi$, which leads to a slow increase in the efficiency.) Third, the efficiency begins to fall after this initial, sharp increase. This phenomenon arises as more phonon modes become populated, nonradiative recombination mechanisms become stronger, electronic transport suffers, and more photoexcited carriers recombine before they can reach the cathode surface and contribute to the thermionic current density.

Still, this figure shows that photon-enhanced thermionic emission is an exciting approach for efficient photon harvest. Indeed, the PETE process can be thought of as an improvement upon the photovoltaic as well as the thermionic. Like the thermophotovoltaic, the PETE reshapes the black-body radiation of the sun into a narrow spectrum of emissions, and this reduction in entropy enables the PETE to exceed the Shockley-Queisser limit and approach the Carnot limit. Whether this young technology ends up showing more promise than the theromophotovoltaic remains to be seen. The ability of the device to remain efficient with a relatively hot anode may end up being its crucial advantage: It allows for combined PETE-heat engine cycles with the thermionic current flowing across the heterostructure must absorb some heat in order to do so. In contrast with the thermophotovoltaic and the photon-enhanced thermionic, the hot-phonon absorption barrier is focused on the in-situ cooling of a device rather than the generation of thermal-electricity.

This scheme traces its origin to Shakouri et al. and their proposal for thermionic cooling using an (ungraded) heterobarrier. They hypothesized that a heterostructure could be placed between a hot cathode and cold anode such that hot electrons would evaporate across the heterobarrier (rather than the work function). This would cool the absorber and emitter may be separated by a heterostructure in order to maximize the internal PETE physics and the external thermionic emission and surface recombination physics.

Still, the experimental realization of an efficient PETE device remains a crucial but difficult goal.

3. Hot phonon absorption barrier

The hot-phonon absorption barrier (HPAB) uses a graded heterobarrier in order to recycle Joule heat, as shown in Fig. 15. In the HPAB, energetic electrons flow across the heterobarrier (with potential energy $\Delta \phi_h$), cooling the electron population. Then, a grading accelerates the electrons that pass over the barrier, minimizing the adverse current, and reducing the electrical resistance of the structure. The grading accomplishes this by inducing an electric field, $E$, on the electronic population, which causes some potential energy loss, $\Delta \phi_j = e_i E l_j$, where $l_j$ is the length of the graded structure. Some excess potential energy remains, $\Delta \phi_{\text{HPAB}} = \Delta \phi_h - \Delta \phi_j$, such that a current flowing across the heterostructure must absorb some heat in order to do so. In contrast with the thermophotovoltaic and the photon-enhanced thermionic, the hot-phonon absorption barrier is focused on the in-situ cooling of a device rather than the generation of thermal-electricity.

Early experiments with GaAs did not achieve high efficiency. This is largely attributed to the low cathode temperatures (limited by the stability of GaAs surfaces coated with Cs) and the large rates of surface recombination. However, these first experiments also showed that the theoretical model of PETE operation is quite accurate and that PETE can succeed in practical applications, provided that the cathode material, surface, and device are optimized significantly from their current state. Furthermore, Schwede et al. noted that the...
cathode as it does in a conventional thermionic. Initial predictions showed that cooling by 20–40 °C is possible. A number of theoretical analyses followed\(^{12,185,194,229}\) many of which included the grading\(^{185,194,229}\) or other optimizations\(^{12}\) of the heterobarrier to enhance the thermionic current density and suppress the adverse current. There was additional focus on the microscale regime\(^{185,194}\) and the thermodynamics of this heat harvesting technology. Throughout this study, the scheme has received a number of different names which showcase its similarity with existing technologies: These include thermionic, thermoelectric, and hot-phonon absorbing barriers and tend to reflect the focus of the analysis. Shakouri et al., for example, have noted that the thermionic cooler (submicrometer) behaves much like a thermoelectric as the heterostructure enters the μm regime.\(^{229}\) Conversely, Shin and Kaviany have focused primarily on the nonequilibrium which arises near these barriers and thus renamed a similar cooling scheme the HPAB.\(^{193}\) Here, we use this name as we are focused on the use and manipulation of nonequilibrium optical phonon populations, and the HPAB studies are some of the first to describe such a mechanism. Let us derive the efficiency of the HPAB in order to delve into this discussion.

If an HPAB is inserted into a channel between the source and the drain, the phonon recycling efficiency is defined as the total heat generation within the channel as compared to the heat generation without a HPAB. Assuming that heat generation is due to the Joule heating, we write

$$\eta_{\text{HPAB}} = \frac{\int_{0}^{l} \left( j_{\text{l}} - \dot{\phi}_{\text{HPAB}} dz \right)}{\int_{0}^{l} j_{\text{l}} dz},$$  

(30)

where \(l\) is the length of the channel, \(j_{\text{l}}\) is the Joule heating expected without the HPAB, and \(\dot{\phi}_{\text{HPAB}}\) is the net energy released by phonon emission within the HPAB structure.\(^{192}\) (Note that \(\dot{\phi}_{\text{b}} = \dot{\phi}_{\text{l}}\) if there is no HPAB.) An analysis of the change in the heat production, \(\dot{\phi}_{\text{l}} - \dot{\phi}_{\text{b}}\), assumes that the barrier does not increase the electrical resistance, e.g., due to the grading shown in Fig. 15. In this case, \(\int dz (\dot{\phi}_{\text{l}} - \dot{\phi}_{\text{b}})\) is given by the energy required for a current flux to gain some potential energy \(\Delta \phi_{\text{HPAB}}\) as it flows across an HPAB of length \(l_{\mu}\) i.e., \(\int l_{\mu} \dot{\phi}_{\text{HPAB}}\). Thus, we can then write

$$\eta_{\text{HPAB}} = \frac{\int_{0}^{l} \dot{\phi}_{\text{HPAB}} dz}{\int_{0}^{l} j_{\text{l}} dz / \sigma_{\text{c}}},$$  

(31)

Note that multiple HPAB can be placed in series in a channel, such that \(l\) represents the length of the channel between HPAB.

From Eq. (31), the efficiency is maximized when the current vanishes, the conductivity is high, the effective HPAB barrier height is large, and the channel is small. Note, however, the assumption that the barrier does not increase the resistance between source and drain fails as the barrier grows larger, the current density grows, and the channel length vanishes.\(^{192}\)

In contrast to the above analysis by Shin and Kaviany,\(^{192–194}\) Shakouri et al. discussed the HPAB in terms of creating a thermionic or thermoelectric effect.\(^{185,194,229}\) For example, the HPAB potential can be regarded as a quasi-Peltier potential, \(\dot{P}_{a} \dot{\phi}_{\text{HPAB}} = \dot{P}_{a} T_{\text{c}}\), where in the thermionic (nanoscale) limit they give

$$\dot{P}_{a} = \frac{k_{B}}{e} \left( \frac{\dot{P}_{a} \dot{\phi}_{\text{HPAB}}}{k_{B} T_{\text{c}}} + 2 \right),$$  

(32)

whereas it returns to the thermoelectric definition in the long-length limit.\(^{229}\)

Shin and Kaviany showed that the thermodynamic limit to the HPAB efficiency is approximately 40% at 300 K.\(^{192}\) However, this occurs as the current density (and cooling power) vanishes. Furthermore, ensemble Monte-Carlo Boltzmann transport investigations of Al\(_{x}\)Ga\(_{1-x}\)As HPAB achieve, at most, only half of the efficiency limit,\(^{12,192,229}\) and this efficiency drops rapidly as the field strength increases,\(^{193}\) as shown in Fig. 16. While much less efficient than, e.g., the thermophotovoltaic, photon-enhanced thermionic, and thermoelectric devices, the HPAB enables integrated micro and nanoscale cooling with potentially greater efficiency than optical refrigeration.

That is, while HPABs could be integrated into a macroscale device, they are outperformed by the heat-harvesting technologies which work at the macroscale. At the nano- and microscales, however, HPABs do not suffer from, e.g., the electrical contact resistances which prohibit the use of microthermoelectric coolers. Thus, they find a potential niche as a nanoscale, in-situ heat harvester. However, this device has not been realized experimentally yet, and this remains the most import step in its evaluation as a potential cooling technology.

E. Summary of thermal-electric devices and their performance

With a number of heat harvesting options presented, it is perhaps useful to discuss where each option has found success or garnered excitement, to briefly contrast these myriad options, and to discuss why they are or are not successful in achieving their promise. We will restrict ourselves to those devices which might be used for power generation: the TE, TI, PETE, and TPV.

1. TE

Thermoelectric generators use temperature gradients to induce an electric current via the Seebeck effect. Thermoelectric devices are
limited by their low efficiency. Still, they have found success in space-based power generation, where the lack of moving parts and reliability are attractive. TE devices continue to work at low temperatures for small temperature differences. Thus, the TE can also recover waste-heat, e.g., from the exhaust of an automobile. This is something that neither TI, PETE, nor TPV can do. At some point, however, these other technologies could supplant the TE in space-based applications.

2. TI

Thermionic generators heat a low-work function cathode so that it emits electrons across a narrow vacuum gap, where after they are collected by the anode. Theoretically, the TI can be extremely efficient. Experimentally, this has never been realized, as it is very challenging to manufacture sufficiently small (nanometer) vacuum gaps. Indeed, it is extremely challenging to manufacture nanometer sized gaps (i) over a wide area (ii) that will be subjected to extreme heat fluxes and (iii) large temperature fluctuations (iv) over decades, without that gap deteriorating. Secondarily, the device requires very high temperatures, which precludes it from waste-heat recovery applications. Thus, it must succeed in the much more competitive primary-power generation space.

3. PETE

The PETE is an advanced thermionic which uses concentrated solar to shift the Fermi-level in a thermionic emitter and enhance the thermionic emission. This photon enhancement reduces the stringent temperature requirements of the TI and thus helps with that particular material and manufacturing challenge. However, it adds to the device complexity by requiring that one not only design and manufacture an efficient TI, but also an efficient photon absorber and then optimally combine both these technologies. At this point, experimental results show that the effect exists, but the efficiency achieved remains negligible. Like the TI, the PETE must compete in the competitive primary-power generation space. It is further restricted by the need for a concentrated photon source, and thus competes directly with the TPV, TI, and PV.

4. TPV

The TPV is an advanced photovoltaic which uses high temperatures and selective emission to reshape the solar spectrum into a narrow band of photon emissions which can be efficiently collected by a PV. Thus, it benefits greatly from the wealth of photovoltaic research. Indeed, TPV efficiencies have already surpassed 20%, which is remarkable for such a new technology. Like the TI and PETE, it requires high temperatures and competes in the primary power-generation space. The major advantage of the TI is that it allows for secondary power generation cycles. The major advantage of the PETE is that it allows for lower operational temperatures. The major advantage of the TPV is that it is currently the most promising and there still is substantial room for improvement. Note that it still must compete with the PV, which is currently both simpler and cheaper. If the TPV efficiency begins to reach its ideal limit, however, then it may find a significant place in concentrated solar applications, even though it will almost certainly be more expensive.

III. PHONON HARVEST, PHONOELECTRICITY, AND THE PHONOVALOTIC CELL

The modern explorations of heat harvesting excel by shaping the broad thermal spectrum and limiting entropy. For example, the thermophotovoltaic uses surface-phonon-polaritons, surface nanostructures, and near-field radiation phenomena in order to squeeze the broad spectrum of blackbody radiation into a narrow spectrum of radiation near $4k_B T$. Similarly, the photon-enhanced thermionic generator converts the broad spectrum of blackbody radiation and thermal energy into the emission of electrons with a narrow spectrum of energies greater than the electron affinity of the cathode.

In phonoelectricity, this sort of reshaping is avoided. Instead, a phonoelectric device intervenes before a broad spectrum of thermal energy forms and harvests the already narrow spectrum of nonequilibrium optical phonons produced during electronic relaxation events. While thermal energy and heat are contained in a broad spectrum of energies with a peak intensity near $\sim k_B T$, a phonon mode has a discrete energy and a nonequilibrium population of them can be contained in a small number of states with a narrow spectrum of frequencies. Indeed, a nonequilibrium phonon population has substantially less entropy and more energy per excitation than heat does. Thus, it should be possible to convert a nonequilibrium population of optical phonons into electricity much more efficiently than one can convert heat into electricity. Not only can phonoelectricity achieve the high efficiency of the advanced thermal-electric power generators (e.g., the TPV and PETE), but it can also achieve the nanoscale, in-situ phonon recycling of the thermal-electric coolers (e.g., the laser cooler and HPAB). Moreover, phonoelectricity does not require extreme temperatures in order to operate efficiently. Instead, it relies upon a highly nonequilibrium optical phonon population. While a large temperature can be assigned to this population, $T_{\text{ph},O}$, using the Bose-Einstein statistics in Eq. (B3), the total thermal energy in the lattice and the effective lattice temperature, $T_{\text{eff}}$, can be substantially smaller, i.e., $T_{\text{ph},O} \gg T_{\text{eff}}$. Thus, extremely large optical phonon temperatures do not necessarily induce melting.

Currently, the singular example of phonoelectricity is the photovoltaic cell which was first proposed in 2016. The phonovoltaic cell harvests optical phonons much like how a photovoltaic harvests photons, as shown in Fig. 17. That is, a nonequilibrium (hot) population of optical phonons more energetic than the electronic band gap ($E_{\text{ph},O} > \Delta E_{\text{ph}}$) relaxes by generating electron-hole pairs in a diode (e.g., a pn junction), which splits them to produce a phonoelectric current, voltage, and power. This must occur faster than the hot optical phonon population down-converts into the acoustic phonon modes, producing heat rather than electricity.

Note that if the phonovoltaic cell becomes sufficiently large, the electrons and acoustic phonons it produces will accumulate, inhibiting generation and down-conversion. This drastically reduces the local nonequilibria and enhances the spatial nonequilibrium. In the long-length limit, the local nonequilibrium vanishes. If the contacts are maintained at a cold temperature, $T_c$, and the active junction reaches a hot temperature, $T_p$, then the Seebeck effect will induce a current and/ or form an electric potential ($\sigma_0 \Delta T$) due to the temperature gradient. Generation events in the $p$-n junction will supply this current and maintain the quasi-equilibrium condition. However, the generated electrons will equilibrate with their surrounding environment before collection. Thus, they will not remove heat equal to the bandgap or
cell. Thus, the phonovoltaic can trace its history back to Span et al., who in 2007 proposed the thermovoltaic cell as a variation on the thermoelectric cell. The advantage they saw in the thermovoltaic was its removal of the p-metal-n junction and the associated contact resistances at the hot side of the thermoelectric generator. However, the idea to use a nanoscale device in order to exploit the local nonequilibrium between the optical phonons and electrons did not arrive until 2016.

In that foundational study, the phonovoltaic (pV), thermovoltaic (TV), and conduction hindered (CH) operational regimes were identified. In the first regime, pholectricity is generated; in the second regime, thermal-electricity is generated; and in the third regime, negligible electricity is generated. As shown in Fig. 18, pholectricity generation only occurs when the cooling length, \( \delta_{e-p} \), is large compared to the length of the phonovoltaic, \( L \), and the hot optical phonon population relaxes by generating electrons rather than heat (\( \gamma_{e-p}^{c} > 0.5 \)). As we will show in Sec. III A, a phonovoltaic operating in the phonovoltaic regime operates much differently than a phonovoltaic operating in the thermovoltaic regime. That is, it does not operate like a thermoelectric. Instead, it operates like a photovoltaic.

### A. Efficiency and figure of merit

Let us derive a simple analytical model of the phonovoltaic in order to (i) develop a material figure of merit, (ii) provide accurate predictions, and (iii) fundamental insights into its function. Here, we assumed that the nonequilibria are as follows: A single optical phonon mode is excited, reaching temperature \( T_{p,v} \), and optical phonon energy. Instead, they will remove the Peltier heat, \( Q_{o} = \gamma_{e-p}^{c}E_{p}A_{T} \). That is, a macroscopic photovoltaic generates thermoelectricity (and thermal-electricity) in a device called the thermovoltaic cell.8,21,93,215 Thus, the phonovoltaic can trace its history back to
nonequilibrium in the electron population is limited to the change in
carrier concentration, \(n_e\) and \(n_h\), and the Quasi-Fermi level, \(E_{F,e}\) and 
\(E_{F,h}\) of the electron and hole populations at temperature \(T_e\).
Furthermore, we assume \(T_{f,e} = T_e = T_p\).

The first assumption can be restated as follows: The non-
equilibrium optical phonon population has the least entropy possible.
Therefore, this assumption leads to an efficiency limit which should
decrease as entropy is added to the optical phonon population, i.e.,
if multiple optical phonon modes are heated by the optical phonon
source. The remaining assumptions hold if the electronic and thermal
transport is fast, the device is small, and the electron-electron scatter-
ing rate is much faster than the electron-phonon scattering rate. Let us
continue.

The ratio of the power produced (\(P_e = \epsilon_j \Delta \phi_{\alpha e}\)) where \(\epsilon_j\) is
the current extracted across the applied potential energy \(\Delta \phi_{\alpha e}\) to
the energy supplied (\(Q_{in} = E_{p,0} S_{\alpha e}\), where \(S_{\alpha e}\) is the rate of optical
phonons generated by an arbitrary optical phonon source) determines the
PV efficiency (\(\eta_{PV} = P_e/Q_{in}\)). In an ideal junction (i.e., one with no
excess current loss, e.g., that from surface recombination or junction
 tunneling) with negligible internal resistance, the current produced is
proportional to the net-generation (\(\dot{n}_i\)) within the cell volume (V), i.e.,
\(I_e = e \dot{n}_e V\). The energy flow required to produce this current is \(Q_{e}\).
However, additional heat flows from the optical phonon population into the
acoustic branches due to the net-down-conversion rate (\(\dot{n}_{p,A}\)) within the cell volume, i.e., \(Q_{p,A} = E_{p,0} \dot{n}_{p,A} V\). Assuming
that the interband electron-phonon and phonon-phonon interactions dominate
the optical phonon lifetime, the controlling equations become

\[
P_e = \dot{n}_e V \Delta \phi_{\alpha e}, \tag{33}
\]
\[
Q_{in} = Q_e + Q_{p,A} = E_{p,0} V (\dot{n}_e + \dot{n}_{p,A}), \tag{34}
\]
\[
\eta_{PV} = \frac{P_e}{Q_{in}} = \frac{\epsilon_j \Delta \phi_{\alpha e}}{E_{p,0} \dot{n}_e + \dot{n}_{p,A}}. \tag{35}
\]

Thus, determining the efficiency and power output of a PV cell
requires the relationship between the net-generation and voltage.

To derive this relationship, we assume that the electron and hole
populations are in equilibrium with the metallic contacts on the
and \(p\)-sides of the cell. Thus, the applied voltage affects the electron
and hole population by shifting their quasi-Fermi levels. That is

\[
f_e = \left[ \frac{e (E_f - E_e - \Delta \phi_{\alpha e}/2)}{k_B T_e} + 1 \right]^{-1}, \tag{36}
\]
\[
f_h = \left[ \frac{e (E_f - E_h - \Delta \phi_{\alpha e}/2)}{k_B T_e} + 1 \right]^{-1}. \tag{37}
\]

Next, consider the interband electron-phonon scattering rate presented
in Eq. (A5). For a given optical phonon mode, the scattering rate depends
on the occupation term, \(\Omega\). For the net rate of generation, this term includes all of the relevant population factors, i.e., \(\Omega = \Omega_f (1 - \Omega_f) (1 + f_p) (f_h)\). This equation provides the relationship
between the applied voltage and the net-generation rate. That is

\[
\dot{n}_e \propto \Omega = \dot{n}_{p,0} (1 - \dot{n}_f) (1 - \dot{n}_h) - (1 + \dot{n}_{p,0}) (f_p f_h). \tag{38}
\]

This final term can change in a number of ways depending on the
approximations taken. For example, \(\Omega \approx \dot{n}_{p,0} - f_p f_h\) in a nondegen-
erate semiconductor with \(f_p \rightarrow 0\). This is not, however, a reasonable
approximation as the quasi-Fermi energy approaches the band edge or
as the optical phonon temperature rises. Indeed, it has a large impact
on the predicted current-voltage curve, efficiency, and power of a pho-
novoltaic cell.

Here, we present two descriptions of \(\Omega\): (i) Maxwell occupations
[Eq. (B4)] and approximated \(\Omega = f_p - f_p f_h\) and (ii) exact occupations
(Fermi-Dirac and Bose-Einstein) without approximation of \(\Omega\). In these
descriptions, we assume that the valence and conduction bands are
symmetric and consider a zone-center optical phonon mode so that
the electron (hole) energy in excess of the conduction (valence) ban-
dge is \(E_p = E_h = (E_{p,0} - \Delta E_{\alpha,0})/2\). Under our current assumptions, we
can simplify \(\Omega\) as follows:

\[
(i) \quad \Omega = \exp \left( \frac{E_{p,0}}{k_B T_{p,0}} \right) - \exp \left( \frac{-\Delta E_{\alpha,0} - \Delta \phi_{\alpha e}}{k_B T_e} \right), \tag{39}
\]
\[
(ii) \quad \Omega = -\frac{1}{4} \cosh \left( \frac{E_{p,0}}{2k_B T_{p,0}} \right) \left( \frac{E_{p,0} - \Delta \phi_{\alpha e}}{4k_B T_e} \right) \times \sinh \left( \frac{\Delta \phi_{\alpha e} - \eta C E_{p,0}}{2k_B T_e} \right). \tag{40}
\]

where \(\cosh, \sinh, \) and \(\sin\) are the hyperbolic cosecant, secant, and sin, and
\(\eta C = 1 - T_e/T_{p,0}\) is the local Carnot limit. The resulting curent-
flowmages are shown in Fig. 19(b).

As expected, these two results become identical when \(E_{p,0}/k_B T_{p,0} \ll 1\)
and the nondegenerate statistics are accurate. As \(E_{p,0} \rightarrow k_B T_{p,0}\), however, two notable differences arise. First, the Bose-Einstein phonon
occupancy is much larger than the Maxwell occupancy, driving much faster
generation rates. Second, as the open-circuit voltage grows and the
the quasi-Fermi level approaches the band edge, there are fewer electrons
in the valence band with which phonons may scatter \((1 - f_p)\) and more electrons in the conduction band to block generation \((1 - f_h)\). Thus, the approximated version under-predicts the short-circuit
current and the drop in current as the applied voltage grows. Interestingly, the two approaches yield the same open-circuit voltage, \(\Delta \phi_{\alpha}^c\). That is, if we solve for the applied potential energy for which \(\Omega\) vanishes, we get

\[
(i, ii) \quad \Delta \phi_{\alpha}^c = E_{p,0} \eta C, \tag{41}
\]

where

\[
\eta C = 1 - \frac{T_e}{T_{p,0}}. \tag{42}
\]

Restated, the open-circuit voltage achieved by the phonovolta-
tics is a fraction of the excitation energy, \(E_{p,0}\), and that fraction is
given by the Carnot limit given by the local nonequilibrium
between the electron and optical phonon populations. This is a nice, intuitive result. However, it is generally believed and observed
that the band gap of a system limits its operating voltage; that moving
the quasi-Fermi level into the conduction or valence bands is prohibited by the ability of the diode to function. For these reasons, we
suggest the replacement of the excitation energy in Eqs. (39)
and (40) with the band gap. This leads to the more conservative
open-circuit voltage of \(\eta C \Delta E_{\alpha,0}\). That is

\[
(i) \quad \Omega_{adj} = a^{(i)} \exp \left( \frac{\Delta E_{\alpha,0}}{k_B T_{p,0}} \right) - \exp \left( \frac{-\Delta E_{\alpha,0} - \Delta \phi_{\alpha e}}{k_B T_e} \right), \tag{43}
\]
The current-voltage curve which reaches the limiting fill-factors are shown in Fig. 19(a): An arc curve achieves FF = 1/4 when maximizing the voltage and maximizing the current does not generate power. Similarly, open-circuiting the cell (Je = 0) and maximizing the current density and maximizing the voltage generate no power. The maximum power, where Je ≠ 0, is achieved at the point where the current-voltage curve resembles that of a photovoltaic. In this regime, the nonequilibrium vanishes and the phonovoltaic acts like a thermoelectric (TE) cell and the phonovoltaic cell has a diode-like current-voltage curve which resembles that of a photovoltaic.

In an effort to validate these models, self-consistent hydrodynamic simulations of the electronic transport were coupled to Fourier conduction models of the optical and acoustic phonon transport in a phonovoltaic cell. In these simulations, a band-to-band net-generation model was applied, i.e., \( n = a_{eb}(n_e n_e - n_p n_p) \), where \( n_e, n_p \) are the electron, hole, and intrinsic carrier concentrations, and \( a_{eb} \) is a constant of proportionality which sets \( \gamma_{e-p} = \gamma_{e+p} \) to the desired parameterization. Different boundary, material, and device conditions are set in order to investigate, e.g., the effects of surface recombination or the restriction of the phonon source to a small portion of the phonovoltaic. Our results showed that the hydrodynamic simulations and the analytical models predict a similar efficiency when the assumptions made are reasonable, as shown in Fig. 20. (Surface recombination drastically reduces the efficiency, while doping levels and the active lengths have a notable but less significant effect.)

Let us return to our derivation of the phonovoltaic efficiency, \( \eta_{PV} \), and figure of merit, \( Z_{PV} \). Consider the quantum-efficiency, \( \eta_{QP} \), of the phonovoltaic in short-circuit generating electrons. That is

\[
\eta_{PV} = \frac{1}{Z_{PV}} \Delta \phi_a \rho_e \eta_{Q}\,,
\]

where \( \rho_e \) is the electron density, \( \eta_{Q} \) is the quantum efficiency, and \( Z_{PV} \) is the figure of merit. The results for Eqs. (43) and (44) are presented in Fig. 19(a). These different models of the current-voltage relationship, or curve, have a non-negligible effect on the phonovoltaic efficiency. Indeed, the approximated version, (i), overestimates the efficiency and current-voltage curve resembles that of a photovoltaic; where the thermoelectric (TE) and thermovoltaic (TV) cells have a linear current-voltage curve, the phonovoltaic cell has a diode-like current-voltage curve which resembles that of a photovoltaic.

Published under license by AIP Publishing

Appl. Phys. Rev. 6, 021305 (2019); doi: 10.1063/1.5031425

6, 021305-19


Published under license by AIP Publishing

Published under license by AIP Publishing

\[ F \text{ accounts for the drop in current, as well as the degree of nonequilibrium.} \]

\[ Z_{np}^{\text{eff}} \]

\[ \eta_p \eta_{Qn} F_p = \eta_c Z_{np} F_p, \quad F_p \in (1/4, 1), \quad F_p > F_p^{*}, \] (49)

where the photovoltaic figure of merit, \( Z_{np} \), is

\[ Z_{np} = \Delta E_{eg} / k_B T_c \]

\[ \gamma_{e-p} \]


substantially reducing the electron-phonon scattering rate and \( \gamma_{e-p} \).

\[ \text{Indeed, the most important condition to evaluate before exploring a material candidate is as follows:} \]

\[ \Delta \Phi_{\mu} = \Delta \Phi_{\text{opt}} \] using the approximated current-voltage relationship, Eq. (44), the result is very complex and not particularly revealing. The analytical equation of \( F_p \) in the maximum efficiency condition is even more complicated, and the exact current-voltage relationship does not allow for an analytically derived \( F_p \). However, the unadjusted fill factor can be approximated as \( F_p \approx 1 - \frac{1}{4} \exp \left( -0.1 \eta_c \Delta E_{eg} / k_B T_c \right) \) using Eq. (44).

With this result, model (ii), we can describe some of the important trends hidden within Eq. (49). These trends capture the secondary or tertiary importance of the nonequilibrium and the material metrics. In particular, the larger the band gap becomes compared to the thermal energy, the better the diode will function and the larger the fill factor will become, as shown in Fig. 21(b). Substantial nonequilibrium will enhance this trend.

While \( \Delta E_{eg} / k_B T_c \) and \( \eta_c \) have a large influence on \( F_p \) and thus \( F_p^{*} \), \( \gamma_{e-p} \) also has a large effect on \( F_p \), as shown in Fig. 21(b). Considering, for example, \( \gamma_{e-p} = 1 \). That is, a material where the optical phonon population can only relax by generating electrons. In this case, \( Q_n \) is equal to \( E_{AV} \), and the maximum efficiency is \( \eta_{pV} = \eta_p = \eta_c \Delta E_{eg} / E_{AV} \), i.e., \( F_p^{*} \rightarrow 1 \) when \( \gamma_{e-p} \\rightarrow 1 \). However, this corresponds to the open-circuit, no power condition. Regardless of these secondary effects of \( \eta_c \Delta E_{eg} / k_B T_c \) and \( \gamma_{e-p} \), Fig. 21 shows that the photovoltaic material figure of merit provides a good assessment of a photovoltaic material.

Indeed, Fig. 21(c) shows that the photovoltaic can generate photovoltaic efficiency with an efficiency approaching the Carnot limit, compete with the efficiency of the Rankine or Brayton cycles, and greatly exceed the efficiency of a thermoelectric generator and the photovoltaic figure of merit is sufficiently large. However, the photovoltaic material must meet strict criteria in order to do reach such a figure of merit. Let us discuss these criteria while surveying potential material candidates.

B. Materials

The fundamental criterion is as follows: the optical phonon energy must exceed the band gap. If this criterion is not met, multi-phonon processes are required to generate electrons. Such processes are typically much slower than single-phonon processes, particularly when the optical phonon occupation is small. Thus, the hot optical phonon population will primarily down-convert into the acoustic branches and generate heat rather than electricity, and the quantum-efficiency will vanish.

\[ \text{It is possible that at extremely high optical phonon temperatures, i.e., when the phonon occupation number exceeds unity, that multi-phonon electron-phonon scattering can compete with three-phonon down-conversion processes. Indeed, a simple model of multi-phonon processes gives} \]

\[ \gamma_{e-p} (N_p) \approx \gamma_{e-p} (1) \left( N_p \right)^2, \]

where \( N_p \) is the number of photons involved in the scattering event. However, we will limit our discussion to photovoltaics which utilize single-phonon electron-phonon scattering events, as research into a multi-phonon photovoltaics has not been conducted at this point.

In addition to our requirement that \( E_{AV} > \Delta E_{eg} \), we also require that the band gap is substantially larger than the thermal energy, \( k_B T_c \). Otherwise, the diode will not function well, the fill factor will fall,
and generation will be slowed drastically by the thermally excited electrons. \(^{(30)}\) In combination with these fundamental criteria, we can write \(E_{p,\alpha} > \Delta E_{g,d} \gg k_B T_c\).

Very few materials satisfy \(E_{p,\alpha} > S_{k_B} T\) at room temperature, where \(S_{k_B} T \approx 130 \text{ meV}\). Indeed, the most energetic optical phonon mode in a typical semiconductor is less energetic than 80 meV, as shown in Fig. 22.\(^{2,121,132,201}\) Only the first-row elements are light enough and form bonds strong enough to enable optical phonons more energetic than 130 meV. Diamond and cubic BN, for example, have optical phonon cutoff energies of around 160 meV. As discussed in Sec. II C 2 and shown in Figs. 10 and 22, these strong bonds localize the electrons contained within. Thus, the band gap and optical phonon energy are correlated. However, \(\Delta E_{g,d} \gg E_{p,\alpha}\) in most materials. While \(\Delta E_{g,d} \approx E_{p,\alpha}\) is achievable in soft materials with a metal-semimetal-semiconductor transition, e.g., Hg\(_2\)Cd\(_{1-x}\)Te (MCT), such semimetals have extremely weak bonds and exhibit optical phonons with \(E_{p,\alpha} < 25 \text{ meV}\). Thus, a phonovoltaic made out of MCT will only function when \(T < 30 \text{ K}\).

Graphene is a notable exception: It has 200 meV optical phonon modes and an electronic structure which can be tuned from its semimetallic state (pure graphene) to a semiconductive state.\(^{(2,121,132,148,177,204)}\) Both symmetry and hybridization are crucial. For example, hexagonal boron nitride (h-BN) exhibits more energetic optical phonon modes and a less energetic band gap than the sp\(^3\) hybridized cubic boron nitride. Still, it exhibits a large band gap because the \(\pi\)-bound electrons prefer to reside in orbitals surrounding the more electronegative nitrogen atoms.\(^{(132)}\)

In addition to having energetic optical phonon modes and no band gap, graphene is well suited for use in a phonovoltaic cell for two crucial reasons. First, a band gap can be opened and tuned in the metallic state (pure graphene) to a semiconductive state.\(^{(78,155)}\) Its modes and an electronic structure which can be tuned from its semimetallic and BNyne, see Refs.80 and 230. Reproduced with permission from C. Melnick and M. Kaviany, Phys. Rev. B 94, 245412 (2017). Copyright 2017 American Physical Society.

The symmetry of group IV, sp\(^3\) coordinated materials (e.g., graphene) enables them to overcome this trend. For traditional elemental and composite semiconductors, see Refs. 2, 121, and 201. For graphene (h-C:H), see Refs. 131 and 197. For graphene:BN, see Ref. 132. For materials with the sp\(^3\) acetylene bond, graphdiyne and B\(_n\)yne, see Refs. 80 and 230. Reproduced with permission from C. Melnick and M. Kaviany, Phys. Rev. B 94, 245412 (2017). Copyright 2017 American Physical Society.
semimetal, such that its band gap approaches the optical phonon energy. This can be done through its chemical functionalization or doping, through its deposition on an ordered substrate, or through the application of magnetic or electric fields. Second, the optical phonon modes couple more strongly with electrons than they do with the low-energy acoustic phonon modes, such that it reaches $\frac{c}{c_3} = \frac{e}{c_0} = 0.817,53,132$. Indeed, graphene may be a uniquely suitable phonovoltaic material candidate. While we will discuss the possibility of other material candidates, only tuned graphene has been shown to achieve a high $Z_pV$.

Because of its suitability as a phonovoltaic material candidate, we have conducted a few studies on tuned graphene phonovoltaics. In these studies, we have used hydrogenation, boron-nitride doping, and cross-plane electric fields to tune the band gap of graphene[31,32] or bilayer graphene (BG). These tuned graphene structures are shown in Fig. 23. In particular, these studies have focused on the following aspects of tuned graphene: The tuning of the electron and phonon structures, and the calculation of the electron-phonon-, phonon-phonon-, and phonon-defect coupling strength and interband scattering rate. Let us review these aspects of phonovoltaic research.

1. Tuning the electronic structure of graphene

The low-energy electronic structure of graphene or bilayer graphene can be well modeled and understood by using a simple tight-binding (TB) model as presented in Appendix C. The tight binding model of graphene reveals that the valence and conduction bands are degenerate and dispersionless at the $K$ and $K'$ points of the Brillouin zone (BZ) due to the symmetry between its two carbon sublattices. (Electrons near these points behave like massless Dirac fermions, resulting in their name: the Dirac points.) Moreover, tight binding models show that the electronic structure and bandgap in graphene can be tuned by removing its time-reversal symmetry or and creating asymmetry between its two sublattices. In bilayer-graphene suspended in a vacuum and under no electric or magnetic field, for example, the time-reversal symmetry remains and no band gap opens at the Dirac points. (While another asymmetry leads add mass to the electrons, these points are still often called Dirac points.) As a cross-plane electric field is applied, however, the time-reversal symmetry is lifted, and a band gap opens at the Dirac points, as discussed more thoroughly in Appendix C.

As previously discussed, there are numerous other mechanisms through which to induce and tune this asymmetry. Here, we will focus on those mechanisms which have been used in phonovoltaic studies: Hydrogenation, boron-nitride doping, and the application of a cross-plane electric field. The results presented use a combination of $ab$ initio density-functional theory (DFT) simulations and the tight-binding models above. Let us discuss the results of these simulations.

Figure 24 shows features of the electronic structure for these three tuned graphene materials. As is shown, the band gap size is strongly correlated with the degree of perturbation to the pure graphene...
structure, i.e., the strength of the electric field or the doping concentration. Indeed, the electric potential between graphene layers in bilayer graphene equals the band gap for low field-strengths, as predicted by the simple tight-binding model presented in Eq. (C6). In graphene:BN, if the boron is spread between the two sublattices \( h-C_x(BN)_{y,z} \), then the band gap grows much more slowly than if the boron were to be concentrated onto a single sublattice (ordered \( h-C_x(BN)_{o} \)). Thus, these two tuned graphene materials fit well within the symmetry-asymmetry framework established by the tight-binding model. Moreover, the tight-binding models can be fit to the \textit{ab initio} low-energy band structure of both materials.\textsuperscript{133}

Graphene:H, however, is more complicated. In graphene:H, the hydrogen concentration must exceed a threshold before a band gap opens. Low hydrogen concentrations influence the electronic structure primarily by moving the Dirac point within the Brillouin zone (BZ). However, a phonovoltaic material also requires suitable kinetics. In Secs. III B 2 and III B 3, we will discuss the electron-phonon, phonon-phonon, and phonon-defect interactions and their kinetics in tuned graphene.

### 2. Interband electron-phonon scattering in tuned graphene

The interband electron-phonon coupling drives electron generation in a phonovoltaic. In previous studies of tuned graphene, we apply

---

**FIG. 24.** Electronic structure of (a) and (c)–(e) bilayer graphene under a cross plane field,\textsuperscript{133} (b) and (f) graphene:BN,\textsuperscript{132} and (g) graphene:H.\textsuperscript{131} (a) and (b) The band gap as a function of the perturbation (applied electric potential, \( \phi_e \), or BN doping concentration), (c) the low-energy band surface and (d) band structure, and (e)–(g) the band structure along high-symmetry lines. The band gap is correlated with the perturbation (field strength or doping concentration), while the perturbation mechanism can have secondary effects: The low-energy bands in bilayer-graphene gain a trifold hat-like shape under strong fields, while in graphene:BN they only gain mass, and in graphene:H they both gain mass and also move within the BZ. (b) Experimental results from Ref.\textsuperscript{21} and maximum/average lines from Ref.\textsuperscript{140}. (a) and (c)–(e) Reproduced with permission from C. Melnick and M. Kaviany, Phys. Rev. B \textbf{96}, 205444 (2017). Copyright 2017 American Physical Society. (b) and (f) Reproduced with permission from C. Melnick and M. Kaviany, Phys. Rev. B \textbf{94}, 245412 (2017). Copyright 2017 American Physical Society. (g) Reproduced with permission from C. Melnick and M. Kaviany, Phys. Rev. B \textbf{93}, 094302 (2016). Copyright 2016 American Physical Society.

---
Eqs. (A3)–(A5) in order to evaluate the rate at which the optical phonon modes scatter with the valence electrons and produce conduction electrons. Typically, DFT and density-functional perturbation theory (DFPT) are used to collect the phonon structure (ωk,s and vkr,s), electron structure (Ek,i and |kr,i⟩), and the electron-phonon interaction matrix elements [M pref]ij(kr,ki) throughout the BZ.111-113 However, the derivative of the tight-binding matrix may be taken with respect to a particular phonon mode in order to calculate the interaction matrix elements according to Eq. (A3), as in Refs. 133 and 198. In this approach, the tight-binding model is used to supply information about the electronic structure while canonical or DFPT phonon modes may be used for the phononic structure. With this information, the Fermi golden rule (FGR) integration in Eq. (A4) is carried out. Typically, the Dirac δ function is replaced with a Lorentzian δ function in order to enhance the rate of convergence with increasing k, mesh density.

Figures 26(a)–26(c) show the components of this integral, i.e., (a) the Lorentzian δ function, (b) the matrix elements, and (c) the resulting scattering rate in bilayer graphene under a cross-plane field.123 As the electric potential between bilayers grows and the band gap opens, the electron-phonon coupling elements weaken near the Dirac points. However, this effect is compensated by the drastic increase in the density of the electronic states which can interact with the optical phonon, particularly when the band gap approaches the optical phonon energy. This is caused by the change in topology of the electric bands under strong electric fields: They change from a parabolic shape (where the density of states at the band edge vanishes as ΔEp → Ep,0) to a trifold hat like shape (where the density of states at the band edge grows as ΔEp → Ep,C). Thus, the electron-phonon scattering rate increases substantially as ΔEp → Ep,0 for both the A1g and E2g modes. This phenomenon is more pronounced for the higher energy E2g mode, as the electronic bands become even more deformed as the field strength grows. Note that the TB model does not capture the decrease in the electron-phonon coupling as the field strength grows. Thus, it drastically over-predicts the electron-phonon scattering rate as ΔEp → Ep,O.133

In contrast to bilayer graphene under a field, Graphene:BN maintains a parabolic band shape as the band gap approaches Ep,O. Therefore, the electron-phonon scattering rate vanishes as the band gap approaches the phonon energy. Until ΔEp,O ≈ Ep,O, however, the effective mass of the electrons increases such that the density of available states remains constant. This is true in both the simple TB model given in Eq. (C5) and the ab initio simulations of Graphene:BN.132 While the electron-phonon coupling strength is weakened by the inclusion of BN, it is nearly independent of the band gap. Thus, the electron-phonon scattering rate is nearly independent of the band gap until ΔEp,O ≈ Ep,O, at which point it quickly vanishes.132 as shown in Fig. 26(d).132

Graphene:H also shares a relatively parabolic band structure once a band gap opens. However, the interband electron-phonon coupling drops drastically in graphene:H, as shown in Fig. 26(e). Interestingly, the intraband electron-phonon coupling remains nearly the same, as shown in Fig. 26(f). Note that the σ−π’ band coupling is negligible in graphene, and it is the π−π’ band coupling which leads to the short phonon lifetime, as shown in Fig. 26(b). Thus, we hypothesize that the transition from sp2 hybridization and π bands in graphene to sp2 hybridization and σ bands in graphene:H leads to the drastic reduction in the interband coupling.

These results showcase the importance of each parameter in creating large interband electron-phonon scattering rates. Of primary importance is the electron-phonon coupling, which must remain substantial for fast scattering to occur. Secondly, the low-energy electronic topology determines the trend in γp→p as ΔEp → Ep,0: If the set of band edge states are disconnected points, e.g., K and K’, then γp→p → 0 as ΔEp → Ep,0. If the set of band edge states is a ring (2D) or sphere (3D), then γp→p grows as ΔEp → Ep,0. However, the key parameter in photovoltaic performance is not γp→p but γp→p = γp→p/γp, where γp includes all mechanisms through which a hot optical phonon population relaxes. Let us discuss these mechanisms.
3. Entropy production in tuned graphene

A hot optical phonon population can relax through the generation or excitation of electrons due to the electron-phonon coupling ($\gamma_{e-p}$), and by spreading of the hot optical population into nearby, cold optical modes due to defects in the crystal ($\gamma_{p-d}$). The latter mechanisms are unavoidable mechanisms which generate entropy, rather than electricity. In this section, we will discuss the calculation of these quantities in graphene and tuned graphene.

The $E_{3g}$ phonon linewidth in graphene has been measured experimentally at around 13 cm$^{-1}$ at ~0 K. Subsequent DFT and DFPT investigations showed that electron-phonon coupling contributed dominantly to this linewidth (~11 cm$^{-1}$), with three-phonon processes contributing nearly all of the remainder (~2 cm$^{-1}$). This provides $\gamma_{p-d}^* \approx 0.8$ in pure graphene. The same study predicted an even smaller three-phonon contribution to the linewidth for the $A_{1g}$ mode (~1 cm$^{-1}$). As the electron-phonon contribution to the linewidth in the $A_{1g}$ mode is similar to that of the $E_{3g}$ mode, then $\gamma_{e-p}^* \approx 0.9$ should be achievable for the $A_{1g}$ mode in pure graphene. If the tuning of graphene does not affect the phonon-phonon contributions to the linewidth, then we can expect slightly worse $\gamma_{p-d}$ in graphene:BN and better $\gamma_{e-p}$ in bilayer-graphene under a cross-plane field due to the associated behavior of $\gamma_{e-p}$.

However, direct calculation of the anharmonic force constants in graphene:BN is currently impossible due to the computational cost, particularly for dilute concentrations of BN: There are too many atoms in a unit cell. That is, while the third-order force constants of a material may be calculated within DFPT or through a linear-response calculation, the number of perturbations required scales with the cube of the number of atoms, and current supercomputers can only calculate accurate third-order force constants in unit cells with only a few atoms. Thus, the investigation of the phonon-phonon scattering rate has been limited to simple systems.

We have focused on the exploration of the effects of asymmetry on the phonon down-conversion rate using a graphene layer on an h-BN substrate (graphene:BN) where the h-BN layer is compressed so that its lattice constant matches that of graphene. The h-BN layer induces asymmetry through the cross-plane B-C and N-C interactions and opens a band gap between 50 and 150 meV based on the stacking...
configuration. As shown in Fig. 26, this has a negligible effect on the overall rate of phonon down-conversion or the final energy, momentum, or polarization distribution of the acoustic phonons produced. However, there is a slight suppression of the LA-TA down-conversion pathway, as shown in Fig. 27.132

While it is likely that doping graphene with BN will induce additional anharmonicity and thus increase the rate of down-conversion, the asymmetry itself does not appear to have a significant effect on \( \gamma_{p-p} \).

Graphene:BN and other chemically functionalized graphene materials generate entropy through more couplings than just the phonon-phonon coupling. For example, the chemical dopants, e.g., BN, are defects which act as phonon scattering centers as discussed in Appendix A 3. The scattering of the nonequilibrium optical phonon population with these defects can quickly spread out the population within momentum space, particularly as the defect concentration rises, as shown in Fig. 28. This phenomenon not only decreases \( T_{pD} \), the Carnot limit, and the phonovoltaic efficiency, but it can also create optical phonons that cannot satisfy the energy and momentum conservation required for an interband electron transitions.133 For example, \( \gamma_{c} \) phonons in graphene with sufficiently large momentum only couple to intraband electronic transitions. The same is true for \( \lambda_{s} \) phonons in graphene with large \(|k_{p} - k_{s}|\). If we assume that all phonon-defect interactions produce optical phonons which will eventually become heat, we can write 

\[
\gamma'_{e-p} = \frac{\gamma_{e-p}}{\gamma_{e-p} + \gamma_{p-p} + \gamma_{p-d}}.
\]

Fortunately, \( \gamma_{p-d} \ll \gamma_{p-p} < \gamma_{e-p} \) in graphene:BN for the small concentrations of BN required to open the <200 meV band gap required in a phonovoltaic cell. Thus, it has a negligible effect on \( \gamma'_{e-p} \) and the phonovoltaic figure of merit. While isotopic variations in the atomic masses also generate defect scattering, they do not contribute substantially to the phonon linewidth in graphene with an average distribution of isotopes.209 (However, it has been proposed that the isotopic variation could be increased in order to limit the formation of hot optical phonon populations and reduce phonon drag.)209 Let us use the results in Secs. III B 2 and III B 3 in order to evaluate the phonovoltaic figure of merit and performance of the tuned graphene materials.

4. Figure of merit and performance in tuned graphene

With \( \gamma'_{e-p} \), \( \gamma_{p-p} \), and \( \gamma_{d-p} \) calculated for variations in the band gap for our three tuned-graphene materials, we can calculate the
crucial material metrics, $\eta_{pV}$ and $Z_{pV}$, and the resulting efficiency, $\eta_{pV}$. Here, we focus on $\eta_{pV}/\eta_c$, as the efficiency of a phonovoltaic is roughly proportional to the Carnot limit. The results are presented in Fig. 29.

As is shown, hydrogenated graphene cannot succeed as a phonovoltaic material. This is due to its negligible interband electron-phonon coupling. In contrast, the electron-phonon coupling in Graphene:BN does not depend on the BN concentration. However, the coupling is weakened slightly, limiting $\gamma_{e-p}$ to $\gamma_{e-p} < 0.75$. Furthermore, the electron density of states vanishes at the band edge, such that $\gamma_{e-p} \rightarrow 0$ as $\Delta E_{eg} \rightarrow E_{p,O}$. Therefore, the optimum $Z_{pV}/\gamma_{e-p} \Delta E_{eg}/E_{p,O}$ occurs around $\Delta E_{eg}/E_{p,O} = 0.9$. In combination, this limits the figure of merit to $Z_{pV} < 0.6$. Still this enables phonolectricity generation with an efficiency of up to 0.5. Bilayer graphene under a cross-plane field avoids both these issues. Moreover, the electronic density of states increases near the band edge, such that $\gamma_{e-p}$ increases as $\Delta E_{eg} \rightarrow E_{p,O}$. This enables $\gamma_{e-p} > 0.9$ when $\Delta E_{eg}/E_{p,O} \approx 0.95$ for both the $A_1$ and $E_2g$ optical phonon modes. That is, the phonovoltaic figure of merit can reach 0.9. At such high $\gamma_{e-p}$ and $Z_{pV}$, a phonovoltaic can generate phonovoltaic with an efficiency exceeding 0.7.$\eta_c$.

However, one must take these promising results with some skepticism: As with the thermal-electric devices surveyed, approaching the ideal efficiency is difficult. Most devices require that the tuned graphene material is grown or placed upon a substrate. This substrate leads to additional or enhances existing entropy production pathways, alters the phonon and electron structures, and ultimately reduces the $pV$ performance. While some limited results show that a hexagonal boron nitride substrate largely preserves the properties of the graphene, the most common substrate is SiO$_2$, which does reduce the transport properties of graphene. Furthermore, impurities and other defects will hurt performance.

In summary, tuned graphene is identified as an attractive material candidate through a variety of analyses. However, the ultimate success of the phonovoltaic cell depends not only on $Z_{pV}$, but also on the cost, weight, practicality, dimensionality, and toxicity of the material, the device, and their manufacturing processes. Thus, finding additional material candidates greatly increases the chances of its success. In Sec. III B 5, we will discuss Metal organic frameworks (MOFs), organic phonovoltaics, and their potential as a 3D phonovoltaic material.

5. Potential for 3D phonovoltaic materials: MOFs

Metal organic frameworks (MOFs) are a group of materials classified by the arrangement of organic molecules linking metallic components (or ions) into an one, two, or three dimensional structure. For example, Fig. 30(a) shows Zn$_4$(CO$_2$)$_6$ organizes into a cubic, 3D MOF in the IRMOF-20 paradigm by the organic ligand FFDC (furo[3,2-b]furan-2,5-dicarboxylic acid). Metal organic frameworks have typically been investigated for their ability to capture or separate gases, as they have a large specific surface area. Additionally, they have recently attracted attention for use in myriad applications, as their chemical and physical properties can be easily tuned.

Indeed, their wide range of crystal structures and chemical compositions allows for a wide range of electronic structures. While MOFs are typically dielectrics, MOF structures have been reported with a band gap in the visible and IR range. Figure 30(b) shows the wide range of band gaps which have been reported for different MOF paradigms, e.g., IRMOF-20, IRMOF-2, and MOF-5. As shown in Fig. 30(c), the MOF energy gap between valence and conduction bands (band gap) is very close to that between the highest occupied molecular orbit (HOMO) and lowest unoccupied molecular orbit (LUMO) in the organic linker. Indeed, the low-energy valence and conduction electrons are strongly localized to the organic linkers, note, for example, the small group velocities of these electrons. Thus, one can control the electronic structure by varying the chemical composition or size of the organic linker or by altering the metallic component. In combination, these levers enable a large degree of control of the MOF electronic structure.

In contrast to the typical relationship between the band gap and optical phonon energy (Figs. 10 and 22), the optical phonon cutoff energy is insensitive to the MOF paradigm and electronic band gap. Indeed, organic molecules share common structures, all of which include strong bonds between C, N, and O atoms. Thus, the Raman studies of MOFs connected by such organic molecules reveal similar phonon features: Most notably, an approximately 200 meV optical phonon mode arises due to the common sp$^2$ carbon—carbon bonds. (Note that graphene shares this bond and this optical phonon cutoff energy.) The tunable electronic structure and an invariant, energetic optical phonon mode make MOFs an attractive material candidate for phonoelectric applications.

However, MOFs still present notable obstacles: The electrons are highly localized which makes electron extraction and phonoelectric power generation difficult, and the smallest band gap realized is still approximately five times larger than the 200 meV optical phonon. Furthermore, the flat, energetically confined conduction and valence bands will inhibit the creation of an energetic, nonequilibrium optical
phonon population suitable for harvest. Indeed, these problems are symptomatic of organic semiconductors. Still, it is difficult to envision inorganic, nongraphene materials which can achieve $E_{\text{p}}$, $O$, $D$, $E$, $g$ at room temperature. Thus, organic and organic-inorganic materials may be the most attractive target for a nongraphene phonovoltaic material candidate.

C. Phonovoltaic applications

The phonovoltaic does not only need a material candidate with a substantial figure of merit, it also needs an application wherein in-situ cooling is important and a large nonequilibrium optical phonon population is produced. In this section, we will discuss the sole phonovoltaic application which has been proposed, the combined phonovoltaic and field effect transistor (pV-FET). Additionally, we will discuss the entropy produced by a real optical phonon source, i.e., a strong electric field, and the effects this has on phonovoltaic performance.

1. Bilayer graphene field-effect transistor

In a bilayer-graphene field-effect transistor (FET), a strong in-plane electric field excites the electron population. The excited electrons scatter with and emit optical phonons, creating a hot, nonequilibrium optical phonon population in a spread of modes with momenta near the $\Gamma$ ($E_{\text{g}}$ mode) and K or K’ ($A_1$ mode) points. As bilayer graphene exhibits a very high figure of merit for both these optical phonon modes, a phonovoltaic should be able to recycle the optical phonons produced by a bilayer-graphene FET.

In contrast to thermoelectric generators, however, this source of “heat” must be harvested within the FET itself (in-situ), and not outside of it (ex-situ). Indeed, the high-energy optical phonons in graphene cannot be transmitted through a metallic or dielectric interface (unless the metal or dielectric exhibits phonon modes of equal energy to graphene). Moreover, the optical phonon modes are slow and thermalize relatively quickly. Thus, a nonequilibrium population of optical phonons cannot be transmitted long distances or through most interfaces. Due to this limitation, we have proposed the incorporation of a phonovoltaic directly into a FET to create a pV-FET, as shown in Fig. 31.

In a FET, an electric field is used to gate the ballistic transport of electrons across a channel, between the highly doped source and drain. When electrons reach the drain, they tend to collide with the lattice and produce optical phonons before they are collected. The optical phonons accumulate in and near the drain, reducing the electron transport and heating up the device as the down-convert into the acoustic phonon modes and produce heat. In the pV-FET, the idea is to use the nonequilibrium optical phonon population to generate electrons before they can become heat. To accomplish this, a pV is incorporated into the FET structure, near to the drain. This pV uses its built-in field to separate the electron-hole pairs produced by the hot optical phonon population so that they cannot accumulate or recombine. Thus, the optical phonons become a useful current and power...
material.

With a discussion of the graphene as both an attractive FET and pV clings the optical phonons emitted by excited electrons. Let us start considerations before moving onto the theoretical feasibility of recy-tigation shown how best to do it. Let us discuss the material and device realization of this device will not be simple, nor has a theoretical inves-

proven difficult to produce high-quality and low twist-angle bilayer the leading candidates for incorporation into a FET. However, it has packed into the same device volume. Bilayer graphene became one of Moreover, it is 2D, which should allow for more transistors to be should allow it to theoretically surpass silicon-based transistors. 

exceptional transport properties and a high break-down voltage, which

can down-convert and a built-in field sweeps up the generated electrons before they accumulate in order to generate power and limit heating in the FET. However, some optical phonons emitted during back-scattering cannot contribute to electron generation, as shown in (d). Figure reproduced with permission from C. Melnick and M. Kaviany, Phys. Rev. B 96, 205444 (2017). Copyright 2017 American Physical Society.

source rather than a damaging source of intense heat generation. The realization of this device will not be simple, nor has a theoretical investigation shown how best to do it. Let us discuss the material and device considerations before moving onto the theoretical feasibility of recycling the optical phonons emitted by excited electrons. Let us start with a discussion of the graphene as both an attractive FET and pV material.

Since its discovery, the functionalization and subsequent use of graphene in transistor technologies have remained a major goal. It has exceptional transport properties and a high break-down voltage, which should allow it to theoretically surpass silicon-based transistors. Moreover, it is 2D, which should allow for more transistors to be packed into the same device volume. Bilayer graphene became one of the leading candidates for incorporation into a FET. However, it has proven difficult to produce high-quality and low twist-angle bilayer graphene, which has limited the size of the bandgaps achieved experimentally. Still, experimentalists have produced bilayer graphene FETs with $\Delta E_g > 100$ meV. While this is not large enough for a room temperature logic FET, it enables its use in an amplifier or transducer FET. As we have discussed, bilayer graphene is the most attractive pV material candidate. Thus, material considerations do not prevent the realization of a pV-FET. Note, however, that gating a “floating” transistor, e.g., bilayer graphene in a vacuum, is extremely difficult. In all likelihood, a real device will be placed on a substrate. As discussed previously, this will reduce the pV efficiency.

Device-side considerations are more concerning. Indeed, a number of major questions arise when one looks to design a real pV-FET device. In particular, how will the intrinsic pV field affect FET operation (and vise-versa)? Can the optical phonons in the highly doped drain reach the pV before down-converting into the acoustic modes? Can the pV current be extracted through separate contacts to generate power, as depicted in Fig. 31(b), or must it be extracted alongside the FET current solely to reduce heat generation? Furthermore, unlike traditional FETs, graphene based FETs are never off; can the pV be integrated in such a way as to continuously recover heat, or can it only recover heat in specific transistor states? Answering these questions requires difficult simulations of the new device and the 2D electron transport within it using ensemble Monte-Carlo (MC) or hydrody-namic models.

While the realization and design of a real pV-FET remain one of the most important topics remaining in the theoretical exploration of the pV cell and its applications, it is left for future study. Primarily, we use the speculative pV-FET application in order to consider a more realistic optical phonon source than that used to derive the ideal pV efficiency and material figure of merit. Consider, for example, that some of the optical phonon modes excited by the FET current cannot generate electrons, as illustrated in Fig. 31(d). That is, the back-scattering of high energy electrons produces optical phonon modes with large momenta with respect to the $\Gamma$ or K points. If the momentum is sufficiently large, then these modes cannot contribute to electron generation, and they will relax only by producing heat. In Sec. III C 2, we will discuss this phenomenon and evaluate the entropy of the hot optical phonon population produced in a bilayer-graphene FET. Then, we will quantify the effect this has on phonovoltaic performance.

2. Entropy produced by the optical phonon source

In order to model the nonequilibrium electron and phonon dynamics in a bilayer-graphene FET, we performed full-band Monte-Carlo (MC) simulations of the electron BTE, Eq. (B1), where the force, $F_e$ equals the in-plane electric field. Within the MC approach, the BTE is solved through the simulation of a few million discrete electron superparticles (each containing the information of many electrons). Each superparticle is accelerated by the field between scattering events, where the time between scattering events and scattering event which occurs are chosen using two random numbers, a self-scattering mech-nism, and the cumulative probability function representing the real momentum-dependent scattering rates. By matching the superparticle ensemble through time, the corresponding distribution function is computed. Here, we assume that there is no variation in real space so that we can compute $f_e(k_e, t)$. Additionally, the electron dynamics
is connected to the corresponding evolution of the phonon distribution function, $f_p(k_z, t)$, through the scattering term, $k_{e-p}$.

In a past study, we use this approach to model the nonequilibrium electron and phonon dynamics in graphene and bilayer graphene. The former was used to illustrate the fundamental behavior of an electron population under a strong field, and the latter was used to investigate how the unique low-energy structure of bilayer-graphene affects the resulting nonequilibria. The solid-state results are presented in Fig. 32.

As is shown, the electron and phonon populations are confined to small regions of momentum-space. Indeed, the electron population is confined to a narrow stream as electrons are accelerated from the low-energy states to those near the optical phonon energy, whereupon they emit an optical phonon and return to a low-energy state. The nonequilibrium optical phonon population is thus confined to a small region of $k_z$ space based on the limited number of electronic transitions available to those excited, $E_b \approx E_{pO}$ electrons. The stronger the electric field becomes, the more energy in excess of $E_{pO}$ an electron is likely to gain before emitting an optical phonon. The more energy in an excess of $E_{pO}$ and electron has, the more low-energy electron states available to which it can scatter. This in turn leads to a broader distribution of heated optical phonons. Thus, a strong field creates more entropy in the optical phonon distribution than a weak field.

This behavior is well reported for graphene, However, bilayer-graphene exhibits some notable differences. Primarily, both the electron and phonon distributions are broadened in momentum-space. This is due to the broad, hat-like electronic structure which creates a large density of electronic states near the band edge. Indeed, the electron energy distribution shows that the vast majority of electrons exhibit very low energy. This broadens the electron stream. Furthermore, it increases the number of states to which the excited electrons can return when they emit a phonon. In combination, this broadens the distribution of optical phonons emitted as the electrons relax. That is, the entropy of the nonequilibrium optical phonon population is larger in bilayer graphene due to its low-energy band structure. Still, very large nonequilibrium is achieved under steady-state conditions, as shown in Fig. 32.

3. Adjusted figure of merit for bilayer graphene in a pV-FET

As we discussed, some fraction of the optical phonons emitted in bilayer graphene cannot generate electrons. This reduces the effective $\gamma_{e-p}$ and resulting figure of merit. Thus, we define an effective $\gamma_{e-p}$ (and $Z_{pV}$) using the steady-state electron and phonon occupation functions

$$\gamma_{e-p} = \sum_{k_z} \gamma_{e-p}(k_e, k_p)$$

where we assume that the hole population mirrors the electron population, such that

$$\gamma_{e-p}(k_e, k_p) \propto f_b(k_p)[1 - f_b(k_e)]^2 - \left[1 + f_b(k_p)f_b(k_e)\right]^2. \quad (52)$$

That is, we assume that we are simulating intrinsic bilayer-graphene.

The results are shown in Fig. 33. Due to the increased production of high-momentum optical phonons, our simulations show that a phonovoltaic cell becomes inefficient when the electric field becomes too strong. However, the optical phonons produced in a FET-pV can be recycled with up to 50% efficiency. This is a drastic improvement over a typical ex-situ heat recovery device like the TE generator. Moreover, it is a drastic improvement over previously proposed in-situ phonon harvesters, e.g., graded heterostructures in a channel which recycle optical phonons with, at best, 20% efficiency or optical refrigerators, which are limited to an efficiency below 5%.

D. Summary

A number of different results have been presented throughout this section, which involve three different methods for tuning the bandgap of graphene. Of particular importance is the effect this has on the electron-phonon coupling and figure of merit. Here, we present Fig. 34, which shows these predicted, maximum $Z_{pV}$ of hydrogenated graphene (C:H), boron nitride doped graphene (C:BN), and bilayer graphene (BG) under a cross-plane field, in the chronological order of the associated investigation. As is shown, a substantial amount of progress has been made in recent years. The investigations have moved from hydrogenated graphene, wherein the figure of merit is negligible, to bilayer graphene, which can reach a figure of merit in excess of 0.9. Indeed, it is unlikely that a material will be discovered or developed which substantially exceeds the figure of merit of bilayer graphene under a cross-plane field. At the moment, however, only graphene-based materials have shown promise. This greatly reduces...
the chances that phonovoltaic cells will ever be successfully developed, much less see commercial use. It is of great importance that potential pV material candidates are diversified to include a myriad of options which are more easily incorporated into devices, which are more economical to produce, and which are fit specific applications wherein large populations of optical phonons are produced. Beyond the ongoing, theoretical material search, it is of crucial importance that the following research takes place: One: experimental validation of the phonovoltaic effect and two, the theoretical investigation and modeling of a realistic phonovoltaic device, e.g., the pV-FET. Additionally, it would be interesting and revealing to perform a second law analysis of the phonovoltaic, calculating the entropy, energy, and other thermodynamic quantities which determine the performance of the pV. (See for example, the second law analysis of the HPAB192 or laser cooling.174) This could be used to quantify the currently qualitative arguments about the advantages provided by a nonequilibrium optical phonon source (in comparison to quasi-equilibrium heat).

IV. CONCLUSIONS AND OUTLOOK

In this review, we have over-viewed a number of solid-state heat harvesting devices. Recent research has led to drastic improvement on this front, with thermophotovoltaics and (photon-enhanced) thermionics quickly becoming attractive technologies that have elicited a substantial and renewed research effort. Simultaneously, scientists have substantially improved the record thermoelectric material figure of merit, making it an increasingly effective solid-state cooling and waste-heat recovering technology. Very recently, the first reported laser cooling of a semiconductor was achieved. These promising efforts have been driven by the substantial improvements in manufacturing technology, which allows for the creation of micro- and nanoscale device and material features; and in computation power and the associated condensed matter theory, such that much more efficient and informed material and device design can take place, accelerating research efforts.

From this holistic perspective, we see that thermal-electricity is limited by the high entropy of its energy source, heat. Furthermore, we see that the successful thermal-electric cycles are able to overcome this limitation through the shaping of heat into a less entropic, narrow-spectrum thermal emission, e.g., a narrow spectrum of surface-phonon-polaritons in a thermophotovoltaic or a narrow spectrum of thermionic emission in a photon enhanced thermionic converter. Indeed, both these technologies can reshape the broad-spectrum, black-body radiation of the sun at 5778 K sun into a narrow spectrum of electronic or electromagnetic emissions near ~1 eV. This enables heat harvesting at an efficiency approaching 50%. While the Carnot limit of such a cycle is reduced from that of the sun (\(\eta_{C} \approx 95\%\)) to that of the material melting temperature (\(\eta_{C} \leq 0.8\)), the benefits drastically outweigh this drawback.

Phonoelectricity, in contrast, intervenes before substantial entropy forms. This allows a phonoelectric device to sustain extreme nonequilibria without melting, as the hot optical phonon population can be contained in only a few states. Furthermore, it uses a local, rather than spatial, nonequilibrium to drive energy conversion. Thus, it can (and likely must) function as an in-situ optical phonon recycler rather than an ex-situ heat recovery device.
The phonovoltaic cell is currently the sole example of a phonoelectric device, and it requires substantially more study. Most importantly, it must be demonstrated experimentally. Additionally, the likelihood of its success would be greatly increased if non-graphene based materials were found with a large figure of merit and which could be incorporated in 3D devices. Furthermore, we believe that heterostructures could be constructed in order to enhance the electron-phonon coupling and electron-phonon nonequilibrium. Finally, multiphonon absorption processes may be possible at extreme nonequilibria, which could enable additional material candidates and a higher conversion efficiency.

Still, a number of foundational and important studies into phonoelectricity and the phonovoltaic have been conducted. These studies span a number of scales, as shown in Fig. 35. Most importantly, the device has been described, efficiency models have been developed, and these models reveal an informative material figure of merit. Additionally, tuned graphene has been established as an attractive material candidate which can reach large figures of merit. Bilayer-graphene under a strong cross-plane field, in particular, has been shown to reach $Z_{pv} = 0.9$. This allows phonoelectric energy conversion with an efficiency at up to 70% of the Carnot limit. This is a substantial improvement over current thermoelectrics and comparable to or better than next-generation thermionic or thermophotovoltaic converters. Thus, we believe the outlook of phonoelectricity is promising.

ACKNOWLEDGMENTS

This work was supported by the U.S. National Science Foundation program on Thermal Transport and Processes (Award Nos. CTS-0553651 and CBET 1332807), the University of Michigan Center for Solar and Thermal Energy Conversion, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy Science under Award No. DE-SC-0000957, and employed computing resources of the U.S. Department of Energy National Energy Research Scientific Computing Center (Office of Science, Contract No. DE-AC02-05CH11231).

We would like to thank the APR editors, and are very grateful for the careful, detailed, and constructive comments from the referees suggesting changes and a coherent flow to make this review more accessible to the general applied physics reader.

APPENDIX A: CARRIER-CARRIER SCATTERING RATES

Energy carriers, i.e., electrons (e), phonons (p), photons (ph), and fluid particles (f), carry charge, vibration (and heat), electromagnetism, and mass, respectively, and energy, in general. A Hamiltonian, $H$, describing a system of these energy carriers can be...
decomposed into the individual Hamiltonians representing the electronic ($H_e$), phononic ($H_p$), photonic ($H_p$), and fluidic ($H_f$) subsystems, as well as the interactions between them, e.g., the electron-phonon interaction ($H_{e-p}$). That is

$$H = H_e + H_p + H_{p-ph} + H_f + H_{e-p} + H_{e-ph} + H_{p-f} + \cdots,$$  

(A1)

where we have only written the two-carrier interactions between electrons and other energy carriers. Note that additional terms can be added to describe higher order interactions, e.g., the electron-phonon-photon coupling $H_{e-ph}$ or the interactions. In Secs. A1–A4 we will present the important interaction Hamiltonians in the study of solid-state devices, i.e., the electron-phonon, phonon-phonon, phonon-defect, and electron-phonon-photon couplings.

Then, we will use the Fermi golden rule in order to derive the scattering rate associated with each interaction. Let us begin with the electron-phonon coupling.

1. Electron-phonon coupling

An electron in state $|k_e, i\rangle$ (wavevector $k_e$ and band $i$) can be scattered to state $|k_e', j\rangle$ by a phonon with momentum $k_p$ and polarization $\alpha$ due to the electron-phonon coupling, as shown in Fig. 36. This coupling arises due to the change in the ionic potential field seen by an electron. That is

$$H_{e-p} = \sum_{i,k_e} \sum_{k_p} M_{ij}^{(e,p)} (k_e, k_p) d_{k_e}^\dagger d_{k_p} (a_{k_e, i} + a_{k_p, j}^\dagger),$$  

(A2)

where $k_e' = k_e \pm k_p, d_{k_e, i}$ and $d_{k_p, j}$ are the creation and annihilation operators for a phonon in state $|i_k, k_p\rangle$, and $a_{k_e, i}$ and $a_{k_p, j}$ are the creation and annihilation operators for an electron in state $|x, k_e\rangle$, and $M_{ij}^{(e,p)} (k_e, k_p)$ is the electron-phonon coupling element. Within perturbation theory, this coupling is the perturbation theory, this coupling is $^{124}$

![FIG. 36. (a) A physical illustration of the interband electron-phonon coupling. (b) Phonon absorption and emission processes.](image)

where $H_e$ is the electronic Hamiltonian of choice, e.g., the Kohn-Sham potential when evaluating the electron-phonon coupling (EPC) within density-functional theory (DFT), and $\epsilon_{k_p, \alpha}$ is the phonon eigenvector.

The resulting, differential electron-phonon scattering rate, $\frac{\partial j}{\partial x}(k_e, k_p)$, is typically evaluated using the Fermi golden rule. That is

$$\frac{\partial j}{\partial x}(k_e, k_p) \approx \sum_{k_p} M_{ij}^{(e,p)} (k_e, k_p)^2 \delta (E_{k_e, i} - E_{k_e, i} \pm \hbar \omega_{k_p, \alpha}) \Omega,$$  

(A4)

where $\Omega$ accounts for the relevant population factors. Thus, to calculate the rate at which a particular phonon mode scatters with all electron states, $\frac{\partial j}{\partial x}(k_p)$, the differential rate is summed across all electron wavevectors and bands, and $\Omega \approx f_e - f'_e$, where $f_e$ and $f'_e$ are the occupancies of the initial and final electron states. In this case, $\Omega$ prevents intraband events from contributing substantially to the phonon lifetime unless the Fermi level lies within a few $k_BT$ of the valence or conduction bands. $^{124}$ Indeed, in a nondegenerate semiconductor, $\Omega \approx 1$ for transitions from the valence to conduction bands and $\Omega \approx 0$ otherwise. Under this condition, the integral scattering rate for a particular phonon mode is

$$\frac{\partial j}{\partial x}(k_p) \approx \sum_{k_p} M_{ij}^{(e,p)} (k_e, k_p)^2 \delta (E_{k_e, i} - E_{k_e, i} \pm \hbar \omega_{k_p, \alpha}) (f_e - f'_e),$$  

(A5)

where $i = v$ is the valence band and $j = c$ is the conduction band.

2. Phonon-phonon coupling

The phonon-phonon couplings arise due to crystal anharmonicity, e.g., the third and fourth order terms in the crystal Hamiltonian $^{127,200}$

$$\langle \phi \rangle = \varphi_e + \frac{1}{2} \sum_{i \neq j} \Psi_{ij}^\dagger d_i^\dagger d_j + \frac{1}{6} \sum_{i \neq j \neq k} \Psi_{ijk}^\dagger d_i^\dagger d_j^\dagger d_k^\dagger + \cdots,$$  

(A6)

where $\varphi_e$ is the equilibrium crystal potential, $d_i^\dagger$ is the displacement from equilibrium of atom $i$ in the $x$ (Cartesian) coordinate, and $\Psi_{ij}$ and $\Psi_{ijk}$ are the second and third-order force constants. Note that the second-order interaction determines the phonon dynamics, i.e., the phonon frequencies ($\omega_{k_p, \alpha}$) and eigenvectors ($\epsilon_{k_p, \alpha}$), while the anharmonic (third- and higher-order) interactions are responsible for the up and down-conversion of phonons, the thermal expansion of the lattice, and the temperature dependence of the phonon frequencies.

The $p-p$ coupling which arises from the anharmonicity of the crystal is typically dominated by the three-phonon interactions, $^{127}$ up-conversion, where two low-energy phonons combine to create a single high-energy phonon, and down-conversion, where a high-energy phonon scatters with and creates two low-energy phonons, as shown in Fig. 37. The matrix element for the interaction between three phonons is $^{200}$

$$M_{ij}^{(p-p)} (k_e, k_p) = \left( \frac{\hbar}{2 m \omega_{k_p, \alpha}} \right)^{1/2} \left( k_e + k_p, j \right) \frac{\partial H_e}{\partial \epsilon_{k_e, i}} \left( k_e, i \right).$$  

(A3)
ate statistics. Note that occupation terms \( f_p \) are typically random, rather than ordered, such that the momentum conservation \( \delta p \) in the scattering centers for the phonons of the ideal crystal. In general, a phonon interaction follows from the Fermi golden rule. For a zone-center phonon mode, the FGR gives \( \gamma_p \), \( \Gamma \)

\[
\gamma_p = \frac{2\pi}{N_\text{ph}} \sum_{\omega k} \left| \frac{M_{\text{ph}}^{\text{ph}}(\omega k, k'_{\text{ph}})}{\sqrt{\Gamma (\omega k) \Gamma (\omega k')}} \right|^2 \delta(\omega_{k_{\text{ph}}} - \omega_{k_{\text{ph}}'}) \delta(k_{\text{ph}}, k_{\text{ph}}'),
\]

where \( k_{\text{ph}} \) is the z-direction component on atom \( i \) and we have discarded the third order occupation terms, \( f_p f_{k'_{\text{ph}}} \), which are negligible for nondegenerate statistics. Note that \( k_{\text{ph}} = -k_p \) from the conservation of momentum. The rate of up-conversion can be similarly formed with altered momentum conservation \( (k_p = k_p' + k_0' \rightarrow k_p + k_p' = k_0) \), energy conservation \( (\omega_{k_{\text{ph}}} = \omega_{k_{\text{ph}}}') \), and phonon occupation terms \( f_p(f_{p'} + f_{p''} + 1) \rightarrow f_{p'} f_{p''} \), again neglecting third order occupation terms.

3. Phonon-defect coupling

Defects in a crystal, e.g., the B and N atoms in graphene:BN, act as scattering centers for the phonons of the ideal crystal. In general, a defect can have a different mass than the atoms in the ideal crystal, or it can change the interatomic force constants. These variations lead to local changes in the atomic displacement and the frequency of this displacement, creating a scattering center. Moreover, the vibrations are typically random, rather than ordered, such that the momentum conservation enforced in the e-p and p-p couplings do not apply here. That is, only energy is conserved in the phonon-defect coupling.

The matrix element which arises from a change in mass, e.g., when isotopes are distributed throughout the crystal, is

\[
M_{\text{ph}}^{\text{ph}}(\omega k, k_{\text{ph}}) = \left[ x(1 - x) \frac{\delta m}{m} \right]^{1/2} \omega_{k_{\text{ph}}} \sum_i |\xi_{k_{\text{ph}}}^i \cdot \xi_{k_{\text{ph}}'}^i|,
\]

where \( x \) is the disorder parameter, i.e., the probability an atom at site \( i \) has mass \( m + \delta m \), and \( m \) is the average mass. From the FGR, the rate of isotopic scattering becomes

\[
\gamma_p = \frac{\pi}{2N_\text{ph}} \sum_{\omega k} \left| \frac{M_{\text{ph}}^{\text{ph}}(\omega k, k'_{\text{ph}})}{\sqrt{\Gamma (\omega k) \Gamma (\omega k')}} \right|^2 \delta(\omega_{k_{\text{ph}}} - \omega_{k_{\text{ph}}'}) \delta(k_{\text{ph}}, k_{\text{ph}}').
\]

However, this interaction element does not account for changes to the interatomic force constants, and it is difficult to derive a satisfactory and reasonable matrix element which includes the complete effects dopant clusters in a crystal. Thus, instead of the perturbation approach used for the previous scattering mechanisms, a Green’s function approach is taken.

In this approach, the phonon spectral function \( A_{k_{\text{ph}}} (\omega) \) is evaluated, and its full-width at half-maximum is the scattering rate. The spectral function is given by

\[
A_{k_{\text{ph}}} (\omega) = \frac{1}{\pi} \Im \left[ G_{k_{\text{ph}}} (\omega) \right],
\]

\[
G_{k_{\text{ph}}} (\omega) = \left( \frac{2\omega}{(\omega + ik_{\text{ph}})^2 - D} \right) \epsilon_{k_{\text{ph}}},
\]

where \( D \) is the dynamical matrix of, e.g., the graphene:BN crystal and \( \zeta \) is a small, positive number.

4. Electron-phonon-photon coupling

Electrons can simultaneously interact with both a phonon mode and the electromagnetic field of a photon. Consider that the Hamiltonian, \( H \), of a system can be decomposed into its electronic (e), phononic (p), photonic (ph), and interaction \( (i - f) \) components. That is

\[
H = H_e + H_p + H_{ph} + H_{e-p} + H_{ph-e}.
\]

In a simple electron dipole system under laser illumination these Hamiltonians are

\[
H_e = E_{e-d} d^\dagger d,
\]

\[
H_p = \sum_q \hbar \omega_q a_q^\dagger a_q,
\]

\[
H_{ph} = \hbar \omega_c c^\dagger c,
\]

\[
H_{e-p} = g_p e_{e-p} \left( \frac{\hbar}{2\omega_e} \right)^{1/2} (a_q + a_q^\dagger) d^\dagger d,
\]

\[
H_{ph-e} = -\gamma_{ph-e} \left( \frac{\hbar \omega_{ph-e}}{2\epsilon_c V} \right)^{1/2} (c^\dagger + c) d^\dagger d,
\]

where \( E_{e-q} \) is the difference in energy between the excited and unexcited electronic states, \( \omega_q \) is the frequency of phonon mode \( q \) and
where \( \psi_s \) is the electronic wavefunction and \( f_{ph} \) and \( f_s \) are the photon and phonon occupancies.\(^{29}\) These can be inserted into the Fermi golden rule equation in order to evaluate the kinetics of the system. That is

\[
\dot{\gamma}_{e-p-ph} = \frac{2\pi}{\hbar} |M_{e-p-ph}|^2 \delta(E_{e,i} - E_{e,i} - \hbar \omega_{ph} - \hbar \omega_p). \quad (A21)
\]

**APPENDIX B: ELECTRON AND PHONON TRANSPORT AND NONEQUILIBRIA**

While the scattering mechanisms discussed in Appendix A describe the rate at which energy carriers interact with each other, they do not describe the evolution of a population of energy carriers in time, position-space, or momentum-space due to an external force, spatial nonequilibrium, or local nonequilibrium within or between the populations of interacting energy carriers. In this section, we present the transport equations which model this evolution. Here, we limit ourselves to the semiclassical and classical models which enable meso- and macroscale description of the devices and processes discussed throughout the remaining review. Let us begin with the semiclassical Boltzmann transport equation.

1. **Boltzmann transport equation**

The Boltzmann transport equation (BTE) describes the transport of fluid particles in a dilute gas,\(^{127}\) wherein the particles are typically in free-flight and only occasionally collide with each other. These collisions are treated as classical, elastic collisions which change the momentum of the particles involved. Boltzmann derived a statistical description of such an ensemble of particles within the kinetic theory of gases,\(^{127}\) and defined the probability of finding a particle at time \( t \), position \( x \) and momentum \( p \), i.e., \( f(t, x, p) \). This theory has since been extended from a description of a (fluid particle) gas \( f(t) \) to both electron \( f_s \) and phonon \( f_{ph} \) gases. Let us overview the major results and assumptions of the BTE, a semiclassical description of the evolution of a probability function \( f_s \).

Consider an occupancy function \( f(t) \) with some fraction of the population at time \( t \), position \( x \), and momentum \( p \), i.e., with some \( f(t, x, p) \). If no collisions occur over some short time \( \Delta t \), then these particles will move to position \( x + u \Delta t \), where \( u \) is the speed of these particles, and have momentum \( p + F \Delta t \), where \( F \) is the external force acting on these particles between time \( t \) and \( t + \Delta t \). By conserving this function in a differential control volume, we can arrive at\(^{127}\)

Taking a second order expansion, we get the interaction matrix for a phonon assisted photon absorption, as shown in Fig. 38. That is

\[
M_{2nd} = \sum_m \frac{|f[H_{out}(m)]| |m[H_{out}(m)]|}{E_{e,i} - E_{e,m}}, \quad (A19)
\]

\[
\frac{df}{dt} + u \cdot (\nabla f) + F \cdot (\nabla f) = \frac{Df}{Dt} = \sum \dot{s}, \quad (B1)
\]

where we have dropped the explicit independent parameters \( t, x, \) and \( p \) for brevity, \( \nabla_x \) and \( \nabla_p \) are the derivatives with respect to the position- or momentum-space coordinates, and \( D \) is the total derivative within the BTE. Let us discuss the in- and out-scattering term \( \dot{s} \).

First, note that in an equilibrium population \( f = f^0 \) under no external force \( (F = 0) \), the in- and out-scattering balance \( \dot{s} = 0 \). That is, \( \frac{Df}{Dt} = 0 \) and the population does not evolve in time or space. These equilibrium populations are given by the Fermi-Dirac, Bose-Einstein, and Maxwell occupancies for a fermion \( (f_{fd}) \), boson \( (f_b) \), or classical particle \( (f_m) \). That is

\[
f_{fd} = \frac{1}{e^\beta + 1}, \quad (B2)
\]

\[
f_b = \frac{1}{e^\beta - 1}, \quad (B3)
\]

\[
f_m = \frac{1}{e^\beta}, \quad (B4)
\]

where \( \beta = (E - \mu)/k_B T \) is the energy, \( E \), in excess of the chemical potential, \( \mu \), compared to the thermal energy, \( k_B T \), where \( k_B \) is the Boltzmann constant and \( T \) is the temperature. Next, note that we can connect the in- and out-scattering to the carrier kinetics derived in Appendix A. For a fermion, for example, we can write

\[
\dot{s} = f(p')[1 - f(p')]\dot{s}(p', p) - f(p)[1 - f(p')]\dot{s}(p, p'), \quad (B5)
\]

where \( \dot{s}(p', p) \) is the in-scattering from state \( p' \) into state \( p \), and \( \dot{s}(p, p') \) is the out-scattering from state \( p \) into state \( p' \). If we only consider electron-phonon scattering, then these rates are given by

\[
\dot{s}(p, p') = \sum_{j, \alpha, \beta} \gamma_{e-p} |\langle \alpha | \beta \rangle|^2, \quad (B6)
\]

\[
\dot{s}(p', p) = \sum_{j, \alpha, \beta} \gamma_{e-p} |\langle \alpha | \beta \rangle|^2, \quad (B7)
\]

where the electron-phonon scattering rates are discussed Appendix A1 and given in Eq. (A4).

Next, note that an external force will typically cause a small deviation \( f \) from the equilibrium population \( f^0 \), and we can write \( f(p) = f^*(p) + f^+(p) \). Finally, note that the scattering term
drives a return to equilibrium. If we approximate this return to equilibrium as something which occurs with a constant lifetime, we can write \( \dot{s}(p) = f^s(p)/\tau(p) \). This is the relaxation time approximation (RTA), which is often used to simplify BTE analyses. It is generally accurate, although there are notable exceptions. Generally, these fall into two categories: strongly coupled systems (e.g., relaxons in graphene\(^{20} \)) and highly nonequilibrium systems (e.g., high-field electron transport).

In many ways, the BTE describes classical transport. For example, it describes particles with an exact momentum and position (violating the uncertainty principle) and no wave-like properties. Thus, it cannot describe many quantum phenomena which depend on the wave-like behavior of an energy carrier, e.g., electron tunneling. However, it can capture the interaction between energy carriers as evaluated within a quantum framework (scattering) and the velocity of the energy carrier (\( u_p \)) as evaluated using, e.g., condensed matter theories.

### 2. Hydrodynamic transport model

The hydrodynamic model takes the first three moments of the BTE to simulate transport on the device scale\(^{116} \). Here, we provide a brief overview of the model, and the interested reader is referred to Ref. 116 for a thorough derivation. First, consider the density of a particle number, momentum, or energy the associated density is

\[
\rho_i(t, x) = \frac{1}{V_\Omega} \sum_p \phi(p) \rho_i(t, x, p),
\]

where \( \phi(p) = 1, p, \) or \( E(p) \), respectively, and \( V_\Omega \) is the volume of the unit cell. The associated balance equation can be derived by integrating the BTE, Eq. (B1), over \( p \) after multiplying it by \( \phi(p)/V_\Omega \). That is

\[
\frac{1}{V_\Omega} \sum_p \phi(p) \frac{\partial f}{\partial t} + \frac{1}{V_\Omega} \sum_p \phi(p) u \cdot (\nabla f) + \frac{1}{V_\Omega} \sum_{pp'} \phi(p) F \cdot (\nabla f) = \frac{1}{V_\Omega} \sum_{pp'} \phi(p) \dot{s}.
\]

This balance equation can be represented as\(^{116} \)

\[
\frac{\partial \rho_i}{\partial t} = -\nabla \cdot J_i + G_i - R_i + S_i,
\]

where \( J_i = \sum_p \phi(p) u f_i / V_\Omega \) is the flux associated with quantity \( \phi \). These terms correspond, in order, to those in Eq. (B9), with the exception of \( S_i \). They represent the rate of accumulation of a quantity due to a divergence in the incoming flux, the generation of a quantity due to a force, the return to equilibrium due to the in- and out-scattering, and the application of an external source. (Note that an external source can also be added to the BTE.) Explicitly, we can write

\[
G_i = F - \frac{1}{V_\Omega} \sum_p \nabla \cdot \phi(p),
\]

\[
R_i = \langle \frac{1}{\tau_i} \rangle (n_i - n_i^s),
\]

where

\[
\langle \frac{1}{\tau_i} \rangle = \frac{1}{V_\Omega} \sum_p f \phi / \tau_i
\]

is the ensemble average relaxation rate, and we have assumed non-degenerate statistics where \( f \phi \) vanishes quickly as \( |p| \) grows\(^{116} \). Note that no relaxation time approximation has been taken. However, the time constant described above depends on the \( f(p) \). Thus, it is often assumed that the ensemble relaxation rate is constant or only varies with the ensemble energy. That is, an RTA-like approach is taken\(^{116} \).

Each balance equation for a conserved quantity \( (n_e) \) contains the quantity conserved in the next balance equation (\( j_x \)). Thus, the set of balance equations considered must be truncated at some point in order to avoid solving the BTE itself. The hydrodynamic equations are based on the four three balance equations, i.e., those for the carrier density, \( n \), momentum (current) density, \( j_c \), energy density, \( W \), and energy flux, \( w \). (Note that which have switched from the general \( n_j \) terminology, which is helpful in showcasing the general procedure, to the specific terminology which is typical in the literature and useful for simplification of the results.) They are truncated by using a temperature tensor \( T_{ij} \) to describe the thermal component of the energy density. In a parabolic semiconductor, the tensor simplifies into a scalar and we can write

\[
\frac{3}{2} k_B T = \frac{1}{2} m_e (u_i^2),
\]

where \( u_i \) is the thermal, or random, component of the electron velocity in direction \( i \).

From the first four balance equations, we can, through some manipulation, arrive at the hydrodynamic equations in their typical form. For electrons and holes at a steady-state, these equations are\(^{87,116} \)

\[
\nabla \cdot j_e = e_i n_e,
\]

\[
j_e = \mu_e (\pm k_B T_e^2 n_e^2 + k_n n_e^2 \nabla T_e - e_i n_e \nabla \phi_e),
\]

\[
\nabla \cdot w_e = -j_e \cdot \nabla \phi_e + \sum_i \omega_{ei},
\]

\[
w_e = \frac{3}{2} k_B T_e j_u - \kappa_e \nabla T_e,
\]

where \( e_i \) is the electron charge, the \( \pm \) in Eq. (B16) is positive for electrons and negative for holes, \( n_e \) is the net generation rate, \( \phi_e \) is the potential energy, and \( \omega_{ei} \) accounts for an external source or sink term, e.g., energy exchange between different energy carriers as discussed at the end of this section. Finally, the mobility and thermal conductivity are\(^{87} \)

\[
\mu_e = \frac{e_i^2 \langle \tau_{pe} \rangle}{m_e},
\]

\[
\kappa_e = \frac{\pi^2 k_B^2}{3 e_i} m_e
\]

where \( \langle \tau_{pe} \rangle \) is the momentum dependent relaxation lifetime. The conductivity, \( \kappa_e \), relationship given here is the Wiedemann-Franz law\(^{87} \). For different energy carriers (e.g., phonons or fluid particles), the transport coefficients \( \mu_i \) and \( \kappa_i \) and the force terms (the \( e_i \nabla \phi_e \) terms) must be adjusted.

---

\( \text{Applied Physics Reviews} \)

REVIEW

scitation.org/journal/are
Often in device simulation, the Poisson equation is added to hydrodynamic equations in order to create a self-consistent simulation. This equation is
\[
\nabla \cdot \varepsilon_r \varepsilon_0 \nabla \phi \equiv -\epsilon_r (n_x + n_D - n_h - n_A),
\]
where \(\varepsilon_r\) is the vacuum permittivity, \(\varepsilon_r \varepsilon_0\) is the permittivity of the material, and \(n_D\) and \(n_A\) are the density of electron donor and acceptor atoms.

The two temperature model is an extension of the hydrodynamic model wherein two systems, e.g., the electronic and phononic subsystems, exchange energy due to the temperature difference and coupling between them. That is,
\[
\tau_{ij} = \frac{1}{2} k_b \left( \frac{1}{\tau_{ij}} \right)' (T_i - T_j) n_i,
\]
where \(\langle \tau_{ij} \rangle\) is the energy averaged relaxation time between carriers \(i\) and \(j\).

**APPENDIX C: TIGHT BINDING MODEL OF GRAPHENE**

Within the tight-binding model, it is assumed that the electronic wavefunctions, \(\psi_i\), can be built out of the atomic orbitals, \(\varphi_i\), centered on the atoms of the crystal at positions \(R_i\) and with quantum numbers \(m\). The tight-binding Hamiltonian contains the self-energy, \(\epsilon_{mn}\), of these atomic orbitals (i.e., their interaction with the ion on which they are centered) and the hopping interaction, \(\varphi_{int, \alpha} = \langle \varphi_{\alpha \downarrow} | \varphi_{\alpha \uparrow} \rangle\), where \(\varphi_{\alpha \downarrow}|\varphi_{\alpha \uparrow}\rangle\) is the overlap integral between the \(\alpha\)th orbital at \(R_i\) and \(R_j\). In graphene, the set of atomic orbitals can be safely limited to a single \(p_z\) orbital per carbon atom, i.e., \(p_{z,i} = p_z(R_i - r)\), and the hopping interaction can be limited to an interaction between nearest-neighbors. Thus, the tight-binding wavefunctions and Hamiltonian can be written as
\[
\psi_{\alpha} = \frac{1}{N} \sum_i \beta_{\alpha \downarrow} p_z \exp (i k_x \cdot r),
\]
\[
H = \sum \epsilon_k d_+^\dagger d_+ - \sum_{j} \varphi_{ij} d_+^\dagger d_j d_j d_i^\dagger,
\]
where the magnitude of \(N\) is the number of atoms, \(\beta_{\alpha \downarrow}\) is the coefficient used to construct the linear combination of atomic orbitals for wavefunction \(n, d_+^\dagger, d_j\), and \(d_i\) are the creation and annihilation operators for the orbital \(p_z\), and \(j\) indicates that the summation is across nearest-neighbor pairs.

In a crystal, the Hamiltonian and wavefunctions are invariant for translations by the lattice vectors, \(a_i\), with \(a_1 = a(3^{1/2}, -1, 0)/2\), \(a_2 = a(3^{1/2}, 1, 0)/2\), and \(a = 2.42\) Å in graphene. Thus, a Fourier transform can be applied to all of the relevant quantities and the problem can be reduced from a dimension of \(N\) to \(M\), where \(M\) is the number of atoms in the unit cell. After these transformations, the Hamiltonian becomes
\[
H = \sum_{k, m} \varepsilon_k d_+^\dagger n_k m d_{k, m} - \sum_{k, l} \varphi_{kl} (d_+^\dagger k_m n_k m + h.c.),
\]
where \(n\) and \(m\) are restricted to atoms in the unit cell, h.c. denotes the Hermitian conjugate, and the Fourier transformed wavefunctions are built using a basis of \(M p_z\) orbitals. That is,
\[
\psi_{k} = \sum_{m} \beta_{m, k} |k, m\rangle.
\]
It is convenient to rewrite the Hamiltonian as an \(M \times M\) matrix, \(H\), with elements
\[
H_{nm}(k_c, n) = \langle k_c, n | H | k_c, m \rangle.
\]
Assuming that the wavefunctions are orthogonal, the eigenvalues and eigenvectors of this matrix provide the details of the electronic structure, \(E_{k}, \beta_{m, k}\), and wavefunctions, \(\beta_{m, k}\).

In graphene, this matrix is
\[
H = \begin{pmatrix}
\epsilon_1 & \varphi_{\alpha} (g(k_c)) \\
\varphi_{\alpha} (g(k_c)) & \epsilon_2
\end{pmatrix},
\]
where \(\epsilon_m\) is the self-energy of the \(p_z\) orbital on sublattice \(m\), \(\varphi_{\alpha}\) is the in-plane hopping interaction, and \(g(k_c) = \sum_j \exp (-i C_j \cdot k_c)\) is a phase factor arising from the three vectors, \(C_j\), connecting the nearest neighbors in graphene. We take \(C_1 = a(0, 3^{1/2}, 0), C_2 = C_3 = C_1 - a_1, C_3 = C_1 - a_1 - a_2\) without losing generality. This creates a reciprocal lattice with vectors \(b_1 = 2\pi(1/3^{1/2}, -1, 0)/a\) and \(b_2 = 2\pi(1/3^{1/2}, 1, 0)/a\).

If the \(p_z\) orbital self-energy does not depend on the sublattice, as in pure graphene suspended in a vacuum, then the eigenvalues of this lattice are degenerate at the \(K = (b_1 + 2b_2)/3\) and \(K' = (2b_1 + b_2)/3\) points. Moreover, there is no dispersion in the electronic bands near the \(K\) and \(K'\), such that electrons behave like massless Dirac fermions.

Thus, they are named the Dirac points. If the time-reversal symmetry is broken, i.e., \(\epsilon_1 \neq \epsilon_2\), then a band gap opens and the electrons gain an effective mass at the (former) Dirac points.

For example, the boron and nitrogen \(p_z\) orbitals in hexagonal boron-nitride (which shares the same tight-binding matrix) have a large difference in their self-energy. Thus, h-BN has a large band gap.

In AB-stacked bilayer graphene under a cross electric plane field, the Hamiltonian matrix is
\[
H_{\text{TB}} = \begin{pmatrix}
-\varphi_{\alpha}/2 & -\varphi_{\alpha} & \varphi_{\alpha} & \varphi_{\alpha} \\
-\varphi_{\alpha} & -\varphi_{\alpha}/2 & \varphi_{\alpha} & \varphi_{\alpha} \\
\varphi_{\alpha} & \varphi_{\alpha} & -\varphi_{\alpha}/2 & -\varphi_{\alpha} \\
\varphi_{\alpha} & \varphi_{\alpha} & -\varphi_{\alpha} & -\varphi_{\alpha}/2
\end{pmatrix},
\]
where \(\varphi_{\alpha}\) is the electric potential difference which arises between the graphene layers because of the electric field, the \(\varphi_{\alpha}\) are the cross-plane hopping interactions, and we have assumed all self-energies are identical. Reasonable parameters are listed in Table I. For twisted bilayer-graphene and AA-stacked bilayer graphene or further review of tight-binding models of graphene, we refer the reader to Refs. 114.

**TABLE I. Tight binding parameterization for graphene and bilayer-graphene from Ref. 112.**

<table>
<thead>
<tr>
<th>(\varphi_{\alpha})</th>
<th>(eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\varphi_{\alpha})</td>
<td>2.90</td>
</tr>
<tr>
<td>(\varphi_{\alpha})</td>
<td>0.40</td>
</tr>
<tr>
<td>(\varphi_{\alpha})</td>
<td>2.30</td>
</tr>
<tr>
<td>(\varphi_{\alpha})</td>
<td>0.15</td>
</tr>
</tbody>
</table>
REVIEW

Roman

NOMENCLATURE

\[ \begin{align*}
& a_i \quad \text{Lattice vector} \\
& A_i \quad \text{Richardson constant} \\
& A_{k_x} \quad \text{Phonon spectral function} \\
& A' \quad \text{Surface and material influence on the} \\
& \quad \text{Richardson constant} \\
& A'_k \quad K \text{ and } K' \text{-point TO phonon mode in graphene} \\
& a, a' \quad \text{Phonon annihilation and creation operators} \\
& b_i \quad \text{Reciprocal lattice vector} \\
& c, c' \quad \text{Phonon annihilation and creation operators} \\
& C_i \quad \text{Vectors connecting nearest-neighbor atoms in} \\
& \quad \text{a lattice} \\
& D \quad \text{Total derivative} \\
& \Delta D \quad \text{Phonon dynamical matrix} \\
& d^\dagger_i \quad \text{Atomic displacement of atom } i \text{ in the } x \\
& \quad \text{direction} \\
& d, d' \quad \text{Electron annihilation and creation operators} \\
& e_e \quad \text{Electron charge} \\
& e_i \quad \text{Electric field} \\
& E_{e,i} \quad \text{Electric potential energy} \\
& \Delta E_{e,i} \quad \text{Quasi-Fermi energy for electrons } (i = n) \text{ and} \\
& \quad \text{holes } (i = p) \\
& \Delta E_x \quad \text{Electronic transition energy targeted in laser} \\
& \quad \text{cooling} \\
& E_{p,O} \quad \text{Energy of optical phonon mode} \\
& E_{e,n} \quad \text{Energy of electron with quantum numbers } n \\
& \quad \text{and wavevector } k_e \\
& E_{2\Gamma} \quad \text{Gamma-point, LO phonon mode in graphene} \\
& F \quad \text{Force} \\
& F_T \quad \text{Fill factor} \\
& f_i \quad \text{Non-equilibrium occupation function of population} \\
& \quad i \\
& f'_i \quad \text{Equilibrium occupation function of population} \\
& \quad i \\
& f_i^\prime \quad \text{Deviation of the occupation function from} \\
& \quad \text{equilibrium} \\
& G \quad \text{Conductance} \\
& G_{k_x} \quad \text{Phonon Green’s function} \\
& g(k_e) \quad \text{Tight-binding phase factor} \\
& \text{G}(\Delta T) \quad \text{Electron-phonon energy exchange function} \\
& G_i(p) \quad \text{Generation rate of carrier } i \text{ with momentum } p \\
& h \quad \text{Reduced Planck constant} \\
& H_i \quad \text{Hamiltonian for energy carrier } i \\
& H_{i,j} \quad \text{Interaction Hamiltonian between energy carriers} \\
& \quad i \text{ and } j \\
& j_i \quad \text{Current flux} \\
& J_i \quad \text{Current flow} \\
& k_{B} \quad \text{Boltzmann constant} \\
& L \quad \text{Length} \\
& m \quad \text{Mass} \\
& M_{e,e} \quad \text{Electron effective mass} \\
& M_{i,j}^{n,m}(k_i, k_j) \quad \text{Interaction matrix element between energy} \\
& \quad \text{carriers } i \text{ and } j \text{ with discrete quantum numbers} \\
& \quad n \text{ and } m \text{ and with wavevectors } k_i \text{ and } k_j \\
& n_i \quad \text{Density of energy carriers } i \\
& N_i \quad \text{Number of energy carriers } i \\
& n_i \quad \text{Generation rate of energy carrier } i \\
& p \quad \text{Momentum} \\
& P \quad \text{Electric power} \\
& p_{e,i} \quad \text{Electrical power density} \\
& P_{e,i} \quad \text{Electrical power} \\
& p_{s,i} \quad \text{p}_i \text{ orbital centered on atom } i \\
& q \quad \text{Heat flux} \\
& Q \quad \text{Heat flow} \\
& Q_{e,i} \quad \text{Heat from thermonic emission} \\
& Q_{r,i} \quad \text{Heat from conduction} \\
& Q_{x} \quad \text{Heat from thermal radiation} \\
& R_e \quad \text{Electrical Resistance} \\
& s_{ph,i} \quad \text{Photon polarization vector} \\
& S_i \quad \text{Energy source } (S_i > 0) \text{ or sink } (S_i < 0) \text{ due to} \\
& \quad \text{mechanism } i \\
& \dot{s}_{i}(p, p') \quad \text{Rate of transition from momentum } p \text{ to } p \text{ for} \\
& \quad \text{carrier } i \\
& t \quad \text{Time} \\
& T \quad \text{Temperature} \\
& u \quad \text{Velocity} \\
& V \quad \text{VOLUME} \\
& V_{0} \quad \text{Volume of a unit cell} \\
& w \quad \text{Energy flux} \\
& W(z) \quad \text{Principal solution for } w \text{ in the equation } z = \\
& \quad we^z \\
& x \quad \text{Cartesian coordinate, disorder parameter (phonon-defect coupling)} \\
& x \quad \text{Position vector} \\
& Z_i \quad \text{Material figure of merit for device } i \\
& \gamma \quad \text{Peltier coefficient} \\
& \gamma_e \quad \text{Seebeck coefficient} \\
& \beta \quad \text{Inverse thermal energy } 1/k_0 T \\
& \gamma_{e-p} \quad \text{Fraction of optical phonon scattering} \text{ events} \\
& \gamma_{i-j}(k_i) \quad \text{Scattering rate between an energy carrier } i \text{ with} \\
& \quad \text{quantum numbers } n \text{ and wavevector } k_i \text{ and all} \\
& \quad \text{energy carriers of type } j \\
& \gamma_{i-j}^{n,m}(k_i, k_j) \quad \text{Scattering rate between energy carriers } i \text{ and } j \quad \text{with discrete quantum numbers } n \text{ and } m \text{ with} \\
& \quad \text{wavevectors } k_i \text{ and } k_j \\
& \delta_{i-j} \quad \text{Change or difference} \\
& \delta_{i,j} \quad \text{Cooling length of energy carrier } i \text{ due to energy} \\
& \quad \text{carrier } j \\
& \delta_{\alpha \beta} \quad \text{Dirac delta function} \\
& \Delta E_{\alpha \beta} \quad \text{Electronic bandgap} \\
& \Delta \phi \quad \text{Applied electric potential} \\
& \delta(x) \quad \text{Delta function} \\
\end{align*} \]


Thermoelectricity: Science and Engineering
R. R. Heikes and R. W. Ur,


152. J. P. Petersen, G. Balakrishnan, and C. L. Weeks, “Mof crystal growth: Uv res-...


B. Zhong, Y. Zhang, W. Li, Z. Chen, H. Cui, W. Li, Y. Xie, Q. Hao, and Q. He, “High superionic conduction arising from aligned large lamellae and large figure of merit in bulk Cu$_{1.94}$Al$_{0.02}$Se,” Appl. Phys. Lett. 105, 123902 (2014).


