PHYSICAL REVIEW B **66**, 125316 (2002)

# Bias-dependent Peltier coefficient and internal cooling in bipolar devices

K. P. Pipe and R. J. Ram
*Research Laboratory of Electronics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139*

A. Shakouri
*Jack Baskin School of Engineering, University of California Santa Cruz, Santa Cruz, California 95064-1077*
(Received 9 May 2002; published 27 September 2002)

The work described here is an investigation of thermoelectric phenomena in bipolar semiconductors and *p*-*n* junctions. In contrast to majority-carrier semiconductors in which a constant material-dependent Peltier coefficient is defined for a given temperature, bipolar devices can be modeled by introducing a bias-dependent Peltier coefficient at interfaces that takes into account the variation of the carriers' average transport energy. It is shown that this effective Peltier coefficient can vary by orders of magnitude as a function of applied bias, and can give rise to interfacial thermoelectric cooling or heating depending on device parameters. The bias-dependent bipolar Peltier coefficient is modeled analytically for short-length and long-length diodes, and the different regimes of bias for which cooling is achieved are described, as well as the effects of recombination, length, and doping. Analytical expressions to optimize the thermoelectric effect inside an idealized diode cooler are presented with numerical results for several common semiconductors, a figure of merit for internal diode cooling is introduced, and extensions of the model are given for applications such as the internal cooling of a semiconductor laser diode.

DOI: 10.1103/PhysRevB.66.125316 PACS number(s): 73.40.Kp, 72.20.Pa, 42.55.Px, 85.30.−z

## I. INTRODUCTION

Optical refrigeration has been proposed as a cooling mechanism in light-emitting semiconductor devices.[1–3] In this process, heat energy is transported out of the device through the radiative recombination of thermally excited carriers whose energies are greater than the electrical bias energy. While this phenomenon is relevant to optical devices in general, cooling has only been measured to date in optically pumped devices, for which Joule heating is negligible. The dominant heating processes in this case are nonradiative recombination and photon absorption; a high external efficiency is thus required in order to demonstrate overall optical cooling.[4]

Electrical refrigeration has been demonstrated in semiconductor devices through both thermoelectric effects[5] and thermionic emission.[6,7] Thermoelectric effects such as Peltier cooling occur when carriers move between regions in which their near-equilibrium energy distribution changes, while thermionic emission is a nonequilibrium effect analogous to evaporative cooling through which hot electrons are selectively emitted across a heterointerface.

Although many light-emitting devices are electrically pumped, internal thermoelectric effects have generally been neglected as a mechanism for cooling. These effects have been shown to be of comparable magnitude to heating sources in certain homojunction and heterojunction light-emitting diodes and diode lasers,[8,9] and they are relevant to bipolar devices in general.

The nature of bipolar devices leads to a thermoelectric description that is significantly more complicated than the traditional bulk model used for conventional Peltier coolers. Previous work has applied descriptions of thermopower in bulk regions where electrons and holes are not in equilibrium[10,11] to the modeling of heterojunction bipolar transistors[12,13] and photovoltaic devices in reverse bias/low current conditions.[14] Previous thermoelectric analysis of biased diode structures has addressed junction Peltier effects related to minority carrier injection but has not related these effects to a temperature profile or derived a figure of merit.[15] A description of the operating point dependence of the bipolar thermoelectric coefficients under forward bias is useful for examining the regimes for which device cooling is optimized, especially for devices such as lasers that operate exclusively in forward bias. The analytic expressions derived can be used to tailor the design of devices from a thermoelectric perspective. While net cooling (optical refrigeration) may not be easily achieved in electrically pumped light-emitting devices, the distribution of heating and cooling sources can be rearranged to reduce the temperature at certain regions of the device.

Here we derive the relevant parameters through analytic means for ideal short-base and long-base homojunction *p*-*n* diodes; in comparison with conventional Peltier coolers, we then optimize cooling with respect to region width, current density, and doping for several semiconductor materials.

## II. PELTIER HEAT EXCHANGE

The Peltier effect describes heat exchange that takes place at the junction of two different materials when electrical current flows between them. It is caused by the fact that the average energy that electrons transport can vary from material to material; when crossing between two such regions, charged carriers compensate for this energy difference by exchanging heat energy with the surrounding lattice. A material's Peltier coefficient $\Pi$ is related[16] to the average energy (with respect to the Fermi energy) transported by its electrical carriers through $\bar{E}_{tr}=q|\Pi|$; the amount of heat exchanged for a given current $I$ across a junction is equal to

 ©2002 The American Physical Society

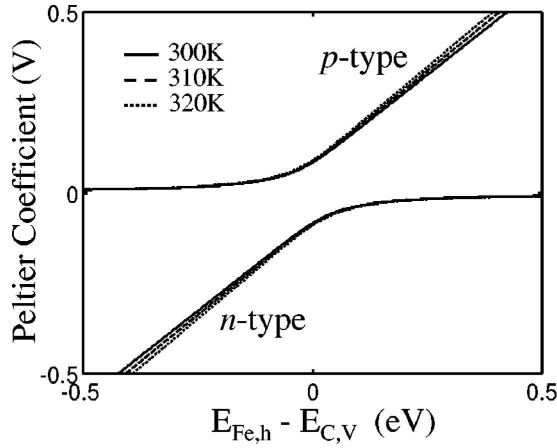


FIG. 1. Variation of the Peltier coefficient with Fermi energy.

$\Delta\Pi * I$. The Peltier coefficient is negative if transport is by electrons and positive if transport is by holes. For a semiconductor with carriers characterized by a Fermi-Dirac distribution $f_{\mathrm{eq}}$ with Fermi energy $E_F$, the Peltier coefficient is given (using the relaxation-time approximation under the assumption of a small bias-dependent perturbation of $f_{\mathrm{eq}}$) by

$$\Pi = \frac{1}{q} \frac{\int \sigma(E)(E_F - E)\left(-\frac{\partial f_{\mathrm{eq}}}{\partial E}\right) dE}{\int \sigma(E)\left(-\frac{\partial f_{\mathrm{eq}}}{\partial E}\right) dE}, \tag{1}$$

where the "differential" conductivity $\sigma(E)$ gives the contribution of a carrier at energy $E$ to the overall conductivity[17,18]

$$\sigma(E) \equiv q^2 \tau(E) \int\int v_x^2(E,k_y,k_z) dk_y dk_z$$
$$\approx q^2 \tau(E) \bar{v}_x^2(E) \bar{n}(E) \tag{2}$$

in terms of the relaxation time $\tau(E)$, average carrier velocity $\bar{v}_x(E)$, and average carrier density $\bar{n}(E)$. From Eq. (1) it is apparent that $|\Pi|$ increases as the carrier distribution becomes more asymmetric with respect to the Fermi energy, since $\partial f_{\mathrm{eq}}/\partial E$ is sharply peaked near $E_F$. The Peltier coefficient thus has an inverse relationship with doping, as shown in Fig. 1, since decreasing doping moves the Fermi energy further into the band gap while the carriers are constrained to stay in the band. The slight dependence of the Peltier coefficient on temperature will be neglected here.

For a nondegenerate semiconductor, the Boltzmann approximation to $f_{\mathrm{eq}}$ is valid, and $\Pi$ can be approximated as[17]

$$\Pi_e \approx -\frac{k_B T}{q}\left(\frac{E_C - E_F}{k_B T} + (2 + r_e)\right), \tag{3a}$$

$$\Pi_h \approx \frac{k_B T}{q}\left(\frac{E_F - E_V}{k_B T} + (2 + r_h)\right), \tag{3b}$$

where $r_e$ is the scattering parameter for electrons defined in terms of the mean free path $l_e$ by

$$l_e = v_e \tau_e = l_0 (E - E_C)^{r_e}. \tag{4}$$

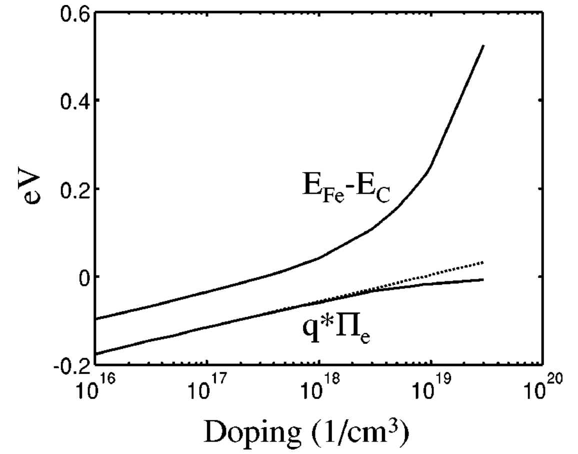


FIG. 2. Variation of the electron Peltier coefficient in $n$-GaAs. The solid line is calculated using Fermi-Dirac statistics [Eq. (1)]; the dotted line is the linear fit of Eq. (3a).

For III-V semiconductors such as GaAs, the dominant scattering mechanism is through optical phonons; for temperatures less than the Debye temperature ($T_D = 344$ K for GaAs), $r_e = r_h = r = \frac{1}{2}$. In some texts on thermolectrics, the scattering parameter is defined in terms of mean free time $\tau(E) = \tau_0 (E - E_C)^{r'}$. Since electron velocity is proportional to $\sqrt{E}$, $r'_e = r_e - \frac{1}{2}$.

The dependence of the Peltier term on doping is further illustrated in Fig. 2, which explicitly shows the electron Peltier term in $n$-GaAs. For $E_{\mathrm{Fe}} < E_C$ (i.e., the Boltzmann limit), the Peltier term is equal to $(E_{\mathrm{Fe}} - E_C)/q$ minus a small offset due to the variation of the average transport energy within the band as described by Eq. (3). As the doping increases beyond the level approximated by the Boltzmann limit $E_{\mathrm{Fe}} - E_C$ increases rapidly, but the Peltier coefficient decreases in magnitude rapidly; the net effect is that the Peltier coefficient increases approximately linearly with the logarithm of the doping. Defining the Peltier coefficient in terms of doping, then, is exact in the Boltzmann limit and also a good approximation in the Fermi-Dirac limit

$$\Pi_e \approx -\frac{k_B T}{q}\left(\ln\frac{N_C}{N_D} + \frac{5}{2}\right), \tag{5a}$$

$$\Pi_h \approx \frac{k_B T}{q}\left(\ln\frac{N_V}{N_A} + \frac{5}{2}\right) \tag{5b}$$

for electrons and holes, respectively, where $N_D$ and $N_A$ are the donor and acceptor concentrations and $N_{C,V}$ is the effective density of states given by

$$N_{C,V} = 2\left(\frac{m_{e,h} k_B T}{2\pi\hbar^2}\right)^{3/2}. \tag{6}$$

In the case of a layer of semiconductor material between two metal contacts, current flow causes heat to be extracted at one metal/semiconductor junction and deposited at the other junction, as shown in Figs. 3(a) and 3(c). In this case, current is carried by majority carriers that are either electrons

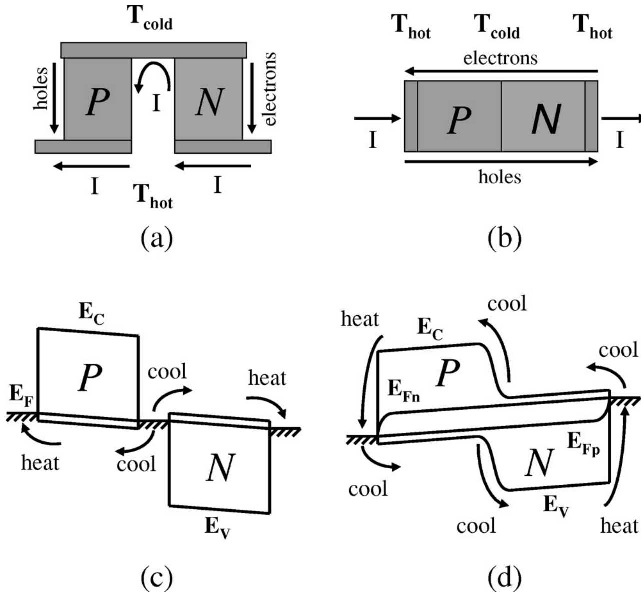


FIG. 3. Illustration and band structure of (a),(c) conventional Peltier cooler and (b),(d) $p$-$n$ junction diode. Note that the direction of current to achieve the same cooling profile is different in the two cases.

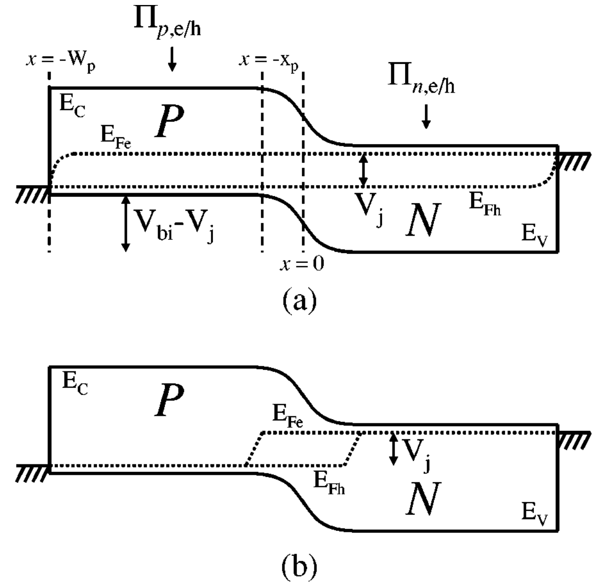


FIG. 4. Homojunction diode in (a) short-length approximation and (b) long-length approximation.

or holes depending on whether the semiconductor is doped with donors ($n$-type doping) or acceptors ($p$-type doping), respectively. Since electrons and holes have opposite charge, carriers in $n$-type and $p$-type regions flow in opposite directions for a given direction of current flux, and heat will be extracted/deposited at the same junctions. If a series of alternating $n$-type and $p$-type layers are connected to each other by metal contacts, the heat exchange at the junctions will also alternate between cooling and heating. By fabricating a structure such that the cooling and heating junctions are on opposite sides of a flat array, an overall heat flux can be achieved from one side of the array to the other. This method of connecting an array of $n$-type and $p$-type materials with metal junctions such that they are electrically in series and thermally in parallel is the traditional scheme for commercial Peltier coolers. The thermal conductivity of the semiconductor regions is balanced with their electrical resistivity [shown in Fig. 3(c) as Joule heating] in order to maximize the value of $T_{\mathrm{hot}} - T_{\mathrm{cold}}$. The operation of a traditional Peltier cooler is intrinsically unipolar since the electrical current that contributes to the thermoelectric effect is carried only by majority carriers. The average energy transported by carriers is constant through each semiconductor region and does not vary with applied voltage bias.

## III. $p$-$n$ DIODE THERMOELECTRIC HEAT EXCHANGE

It is apparent from the position of the Fermi level in Fig. 3(c) that the Peltier coefficients of minority carriers in the semiconductor regions are much larger than the respective majority coefficients; however, the current in this structure is composed exclusively of majority carriers. In order to introduce minority carriers and examine the resulting thermoelectric heat exchange model, we move to the diode structure of Fig. 3(b), in which the $p$-type and $n$-type regions are directly connected.

In the case of an unbiased diode, a large built-in electric field exists at the $p$-$n$ interface; the drift current caused by this field balances the diffusion current out of the doped regions. As the diode is forward biased, the potential barrier seen by carriers at the interface is decreased, and a net current develops according to the standard diode equations.[19] Carriers diffuse against the built-in potential by taking energy from the lattice; this can be seen as bipolar Peltier cooling that depends on the applied bias through the variation of the internal potential.[20] As the minority carrier concentrations increase exponentially in the regions on either side of the junction, the effective Peltier coefficients for electrons and holes decrease as shown in Fig. 1, and the heat exchange at the junction is calculated separately for each kind of carrier.

### A. Short-length approximation

To develop a thermoelectric model of the $p$-$n$ diode, we first examine the idealized structure shown in Fig. 4(a). Under the short-length approximation, all of the injected minority carriers are assumed to recombine at the Ohmic contacts at the boundaries. The injected carrier concentration gradient is assumed to vary linearly between the edge of the depletion region and the ohmic contact, and the Peltier coefficient for injected electrons in the $p$-type region between the contact and the edge of the depletion region ($-W_p < x < -x_p$) is given by applying Eq. (5a) with a spatially varying concentration

$$\Pi_{p,e}(x) = \frac{k_B T}{q}\left\{ \ln\left[ \frac{n_{po}}{N_C(W_p - x_p)} \times [(x+W_p)e^{qV_j/k_BT} - (x+x_p)] \right] - \frac{5}{2} \right\}, \quad (7)$$

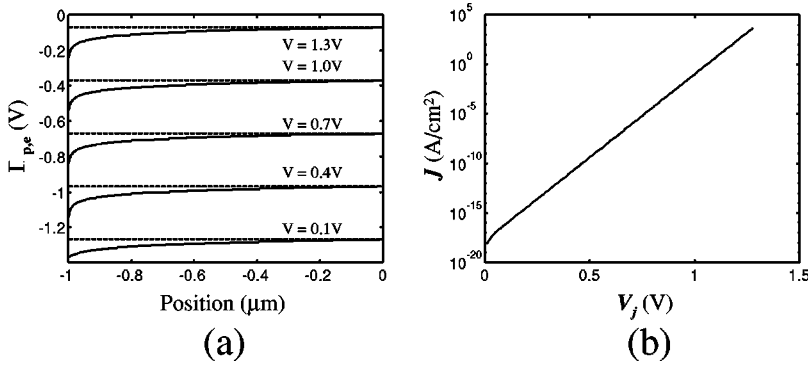


FIG. 5. (a) Variation of $\Pi_{p,e}$ with applied voltage and in the $p$-type quasineutral region $(-W_p<x<-x_p\approx 0)$ of a 2 $\mu$m GaAs diode doped symmetrically at $N_D=N_A=1\times 10^{18}$ cm$^{-3}$. Dotted lines refer to the effective value used for the region as given in Eq. (8). (b) Current-voltage relationship.

where $V_j$ is the junction voltage drop and $n_{po}=n_i^2/N_A$ is the equilibrium electron concentration in the $p$-type region in terms of the intrinsic concentration $n_i$. We neglect for now voltage drops due to resistance in the quasineutral regions that cause the overall bias voltage $V$ to be different from the junction voltage $(IR=V-V_j\approx 0)$ and examine these effects in Sec. IV. As carriers experience diffusion cooling while traversing the quasineutral region, the logarithmic relationship of concentration to the average transport energy causes the Peltier coefficient $\Pi_{p,e}$ to remain relatively constant throughout much of the region while exhibiting a pronounced increase in magnitude near the contact as shown in Fig. 5. We neglect this sharp rise in transport energy due to the fact that carriers will immediately release this same amount of energy as they recombine at the contact; the short spatial scale suggests that a temperature gradient will not develop here.

In order to obtain a simple analytic expression for the bias dependence of $\Pi_{p,e}$, we approximate $\Pi_{p,e}$ by its value at the edge of the depletion region $(x=-x_p)$ and rewrite Eq. (7) as

$$\begin{aligned}\Pi_{p,e}&=\Pi_{n,e}-(V_{bi}-V_j)\\&=V_j-\frac{k_BT}{q}\left(\ln\frac{N_C}{n_{po}}+\frac{5}{2}\right)\\&=-\frac{k_BT}{q}\left(\ln\frac{N_CJ_0}{n_{po}(J_0+J)}+\frac{5}{2}\right),\end{aligned}\tag{8}$$

where $V_{bi}$ is the built-in voltage given by

$$V_{bi}\approx\frac{k_BT}{q}\ln\frac{N_AN_D}{n_i^2}\tag{9}$$

and $J_0$ is the diode saturation current density given in terms of the electron and hole mobilities $\mu_{e,h}$ by

$$J_0=n_i^2k_BT\left(\frac{\mu_e^A}{W_pN_A}+\frac{\mu_h^D}{W_nN_D}\right),\tag{10}$$

where $\mu_e^A$ refers to the electron mobility in a region of doping concentration $N_A$. The diode current density is related to the applied voltage by

$$J=J_0\left(\exp\frac{qV_j}{k_BT}-1\right).\tag{11}$$

Since we are using the linear approximation as shown in Fig. 2, we apply Eq. (9) even for large doping densities. The validity of this approximation will be examined at the end of this section.

The decrease of the bipolar Peltier coefficient within the $p$-type region as the current bias is increased is depicted in Fig. 6. Notice that the majority-carrier component stays roughly constant, while the minority-carrier component does not.

The difference in Peltier coefficient (for both electrons and holes) on either side of the depletion region is equal to $\Delta\Pi=V_{bi}-V_j$. Due to the fact that carriers are symmetrically distributed around the Fermi level in a metal, the Peltier coefficient can be approximated as zero inside the ohmic contacts. Noting that the minority carriers recombine at the contacts and release heat energy $qV_j$, the heat exchange density at the $p$/metal contact (in the absence of other heat sources) can be written as

$$Q_{p/\mathrm{metal}}=-J_e(\Pi_{p,e}-V_j)-J_h\Pi_{p,h},\tag{12}$$

where $J_e$ and $J_h$ are the electron and hole current densities (both taken to be positive) and $\Pi_{p,e}$ and $\Pi_{p,h}$ are the electron and hole Peltier coefficients in the $p$-type region given in Eqs. (8) and (5b). Similar equations can be derived for the $n$-type region; the heat exchange at the junction (again in the absence of other heat sources) is given by

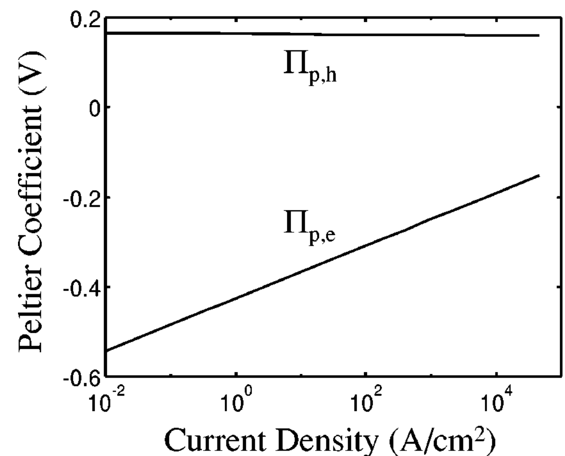


FIG. 6. Bias dependence of the bipolar Peltier coefficient of the $p$-type region.

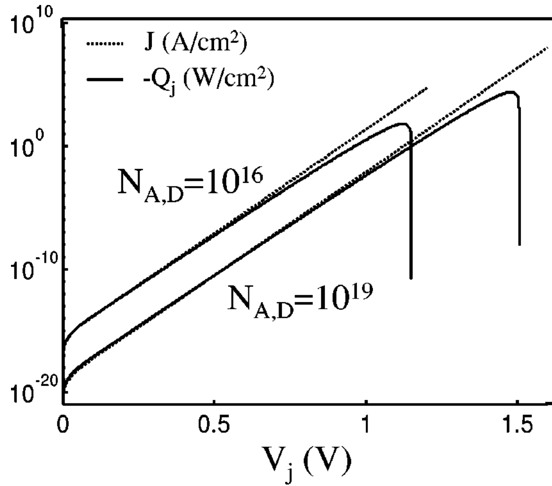


FIG. 7. The variation of $J$ and $-Q_j$ with increasing $V_j$ in a symmetrically doped 1 $\mu$m GaAs diode at two different doping levels.

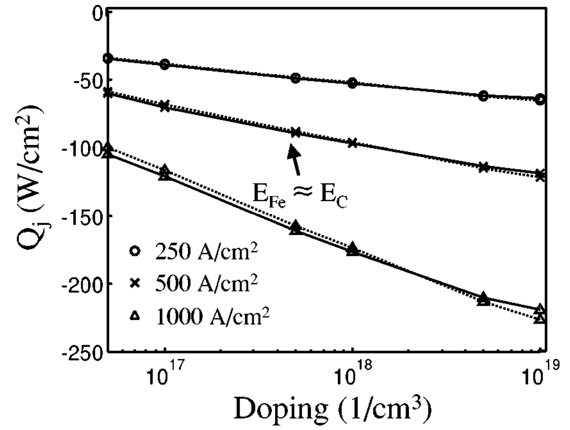


FIG. 8. The variation of $Q_j$ with doping in a symmetrically doped 1 $\mu$m GaAs diode: dotted lines are derived from Eq. (13b); solid lines are derived from Eq. (13a) using drift-diffusion simulations and reflect corrections due to nonzero recombination and Fermi-Dirac statistics at high doping levels.

$$Q_j = J_e(\Pi_{p,e} - \Pi_{n,e}) - J_h(\Pi_{n,h} - \Pi_{p,h}) \tag{13a}$$

$$\approx -J(V_{bi} - V_j), \tag{13b}$$

where $J = J_e + J_h$.

Notice that the total heat exchange given by the sum of the contact terms and the junction term is equal to $JV$, which is the total bias power on the device; the definition of the Peltier coefficient satisfies energy conservation. Since the bias current density $J$ is related exponentially to the voltage $V_j$, the amount of cooling at the junction increases as the device is increasingly forward biased, until $V_j$ approaches $V_{bi}$. By setting $\partial Q_j/\partial V_j = 0$, we see that the cooling power is maximized when $V_{bi} - V_j = k_B T/q$. This value of the junction barrier height is a similar conditions to that cited for heterobarrier coolers utilizing thermionic emission.[21] Using this value in Eq. (13b) gives

$$Q_{j,\mathrm{opt}} = -\frac{1}{e}\left(\frac{k_B T}{q}\right)^2\left(\frac{\tilde{\sigma}_p}{W_p} + \frac{\tilde{\sigma}_n}{W_n}\right), \tag{14}$$

where we define "minority conductivities" $\tilde{\sigma}_p \equiv q\mu_e^A N_D$ and $\tilde{\sigma}_n \equiv q\mu_h^D N_A$; e.g., the minority conductivity for the $p$-type region is the product of the electron mobility in that region times the donor concentration in the $n$-type region. If both $N_D$ and $N_A$ are increased, the minority mobilities decrease, but the minority conductivities show an overall increase; thus, diode coolers typically do not have an optimum doping as majority-carrier coolers do. As shown in Fig. 7, the maximum junction cooling power occurs for the maximum doping.

Figure 8 plots $Q_j$ for a symmetric ($W_n = W_p = 0.5$ $\mu$m) GaAs diode at several different symmetric ($N_D = N_A$) doping levels. Carrier transport energies were calculated by solving the drift-diffusion equations self-consistently with Poisson's equation,[22–24] and the bipolar Peltier coefficients used in Eq. (13a) were calculated by averaging Eq. (1) for carriers in the quasineutral regions outside the diode depletion width (rather than simply taking the value at $x = x_p$), as shown in Fig. 9. There is generally good agreement with Eq. (13b). This can also be seen in Fig. 10, which also compares the analytic expressions derived above to numeric solutions. Notice that the error in the linear approximation depicted in Fig. 2 is small compared to the total difference in Peltier coefficients across the junction $\Delta\Pi \approx V_{bi} - V_j$ shown in Fig. 10. The small error evident in the plots of $Q_j$ at low doping levels in Fig. 8 is due to the resistance of the $n$-type and $p$-type regions that causes an additional voltage drop.

The change in effective Peltier coefficient for electrons and holes at the diode's junction takes place over their respective depletion widths, whose sum is given in terms of the dielectric permittivity $\epsilon_s$ by

$$x_{\mathrm{tot}} = x_n + x_p = \sqrt{\frac{2\epsilon_s}{q}\left(\frac{N_A + N_D}{N_A N_D}\right)(V_{bi} - V_j)}. \tag{15}$$

Notice that increasing the doping level leads to an increased $V_{bi}$ and therefore a larger total cooling at the junction through Eq. (13). For the symmetric case ($N_D = N_A$), assuming $N_D, N_A > n_i$, the cooling region width $x_{\mathrm{tot}}$ also decreases with increased doping; this implies more concentrated cooling at the diode junction. However, calculation of the spatial variation of the cooling term within the depletion region is not described by this thermoelectric analysis due to the comparable magnitudes of the energy relaxation length and depletion width.

### B. Long-length approximation

For the case in which the $p$-type and $n$-type regions are longer than the electron and hole diffusion lengths, all of the injected carriers are expected to recombine in the quasineutral regions rather than at the contacts. This approximation is depicted in Fig. 4(b).

The bias power dissipated in the device is given by

$$JV_j = P_{\mathrm{recomb}} - J(\Pi_{p,h} - \Pi_{n,e}). \tag{16}$$

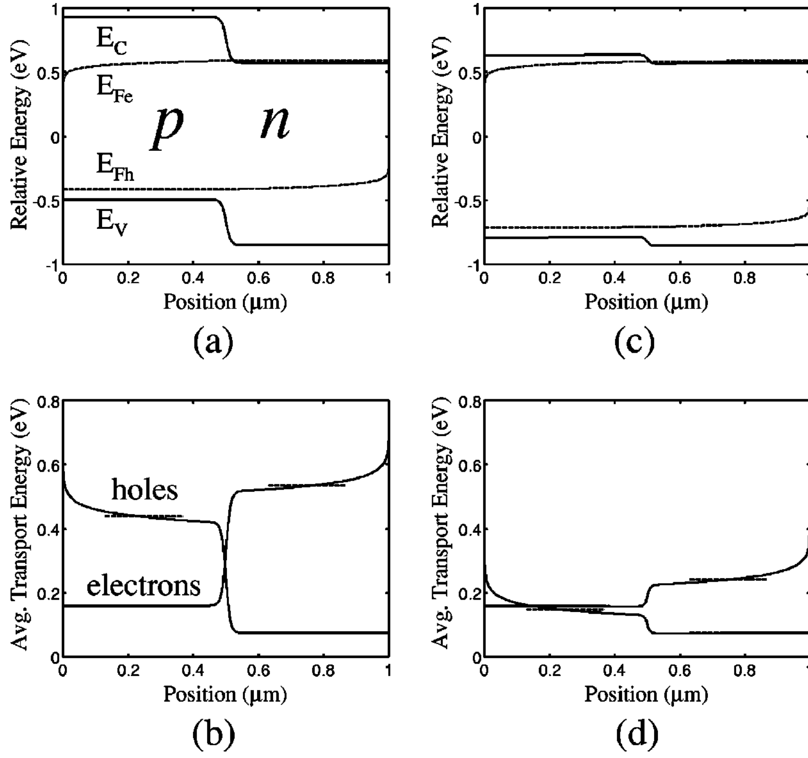


FIG. 9. Homojunction diode: band structure (a),(c) and average transport energy $\bar{E}_{tr}=q|\Pi|$ (b),(d) at 1.00 and 1.30 V respectively, for a 1 $\mu$m GaAs diode doped symmetrically at $N_D=N_A=5\times10^{17}$ cm$^{-3}$. In (b),(d) the dotted lines refer to the averaged Peltier values used to calculate $Q_j$.

We can rewrite the power dissipated through recombination in terms of the band gap $E_g$ as

$$P_{\mathrm{recomb}}=J\frac{E_g}{q}-J(V_{bi}-V_j), \tag{17}$$

where we have made use of the relation

$$\Pi_{p,h}-\Pi_{n,e}=\frac{E_g}{q}-V_{bi} \tag{18}$$

which is valid for homojunction diodes in the Boltzmann approximation. The recombination power given in Eq. (17) is made up of two components $E_g$ and $V_{bi}-V$ that are related to the recombination energy and the energy gain of carriers prior to recombination, respectively.

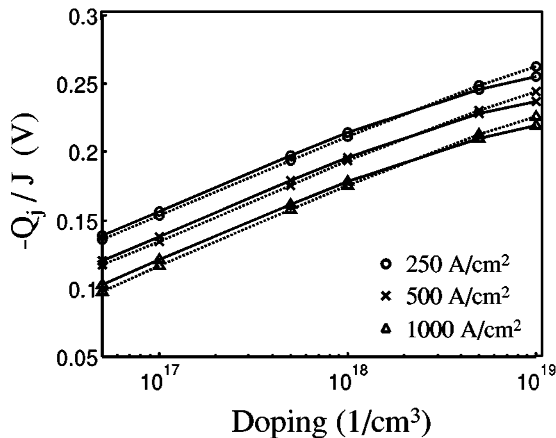


FIG. 10. Comparison of analytic and numeric solutions for 1 $\mu$m GaAs diode $-Q_j/J=V_{bi}-V_j$ computed numerically from Eq. (13a) (solid lines) and approximated by Eq. (13b) (dotted lines).

If a fraction $\eta_{\mathrm{opt}}$ of the recombination power is radiative and leaves the diode, the total heat density (neglecting resistive drops) can be written as

$$Q_{\mathrm{total}}=J\left(V-\frac{E_g}{q}\eta_{\mathrm{opt}}\right), \tag{19}$$

which for $Q_{\mathrm{total}}<0$ yields the traditional constraint for optical refrigeration.[1,2]

## IV. TEMPERATURE

For a layered device structure with thermal conductivity $k(x)$, the heat sources $q(x)$ discussed above produce a steady-state temperature $T(x)$ according to the one-dimensional heat equation $d/dx[k(dT/dx)]=-q$, where $q(x)$ is in units of W/cm$^3$. In order to examine the effects of the rearrangement of the heat sources in a diode, we compare its internal temperature profile to that of the traditional unipolar Peltier cooler shown in Fig. 3(a).

### A. Conventional thermoelectric cooler

The cooling power at the cold junction of the traditional cooler is given in terms of the Seebeck coefficients of the $p$-type and $n$-type regions $\alpha_p$ and $\alpha_n$ by

$$q_c=(\alpha_p-\alpha_n)IT_{\mathrm{cold}}-K(T_{\mathrm{hot}}-T_{\mathrm{cold}})-\frac{I^2R}{2}, \tag{20}$$

where $K$ is the thermal conductance of the $n$-type and $p$-type regions in parallel

$$K=k_p\frac{A_p}{W_p}+k_n\frac{A_n}{W_n} \tag{21}$$

and $R$ is the electrical resistance of the two regions in series

$$R=\rho_p\frac{W_p}{A_p}+\rho_n\frac{W_n}{A_n}. \tag{22}$$

The electrical resistivity of the $n$-type region $\rho_n$ is given by

$$\rho_n=\frac{1}{q\mu_h^{p+n}p+q\mu_e^{p+n}n}\approx\frac{1}{q\mu_e^D N_D}. \tag{23}$$

The Seebeck coefficient $\alpha$ is related to the Peltier coefficient by $\Pi=T\alpha$. By maximizing $q_c$ with respect to current, we find[25] that the maximum temperature difference $T_{\rm hot}-T_{\rm cold}$ for the traditional cooler can be written in terms of the figure of merit $Z$ by

$$(T_{\rm hot}-T_{\rm cold})_{\rm max}=\frac{1}{2}ZT_{\rm cold}^2, \tag{24}$$

where

$$Z=\frac{(\alpha_p-\alpha_n)^2}{KR}. \tag{25}$$

It can be shown that $KR$ is minimized when

$$\frac{W_nA_p}{W_pA_n}=\left(\frac{\rho_pk_n}{\rho_nk_p}\right)^{1/2}. \tag{26}$$

The figure of merit reflects the optimal balance of Joule heating, heat conduction, and Peltier cooling.

### B. Diode cooler

For a diode, we consider the short-length case in which both ohmic contacts are heat sunk at a temperature $T_0$; here we can calculate the maximum temperature difference $(T_0-T_j)_{\rm max}$ that can be achieved between the heat sink and the diode junction. From the perspective of Fig. 3, we connect the $p$-type and $n$-type regions directly together at the "cold" contact and let $T_j=T_{\rm cold}$ and $T_0=T_{\rm hot}$. We neglect thermoelectric voltage drops caused by temperature gradients across the semiconductor regions.

By using Eq. (13b), we can write the cooling power density $Q_j$ at the junction in the short-length limit as

$$Q_j=J(V_{bi}-V_j)-\frac{K}{A}(T_0-T_j)-\frac{J^2RA}{2}, \tag{27}$$

where the nonzero resistance of the quasi-neutral regions is taken into account through $V_j=V-JRA$ and half of the Joule heat is assumed to flow back to the junction.[25] The device is assumed to have uniform cross section $A_p=A_n=A$. From a comparison of Eqs. (20) and (27) we see that the effective Seebeck coefficients for the diode are dependent on bias as well as on geometry and material.

The maximum cooling power density, found by setting $dQ_j/dJ=0$, is given by the condition

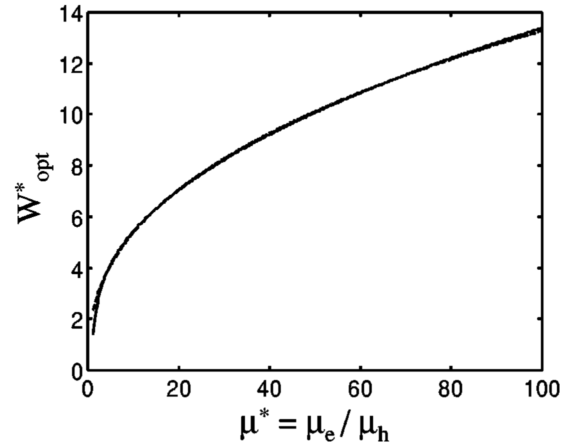


FIG. 11. $W^*_{\rm opt}$ (solid line) and fit $W^*_{\rm opt}=2.17(\mu^*)^{0.39}$ (dotted line).

$$V_{bi}-V_{j,\rm opt}=\frac{k_BT_j}{q}+J_{\rm opt}RA, \tag{28}$$

where $JdV_j/dJ\approx k_BT/q$ for a diode in the short-length approximation [see Eq. (11)]. Equation (28) does not have an analytic solution for the value of $J_{\rm opt}$, but we can arrive at a numeric solution by rewriting it as

$$1+f+\ln f=\ln\left[\left(\frac{\tilde{\sigma}_p}{W_p}+\frac{\tilde{\sigma}_n}{W_n}\right)(\rho_pW_p+\rho_nW_n)\right], \tag{29}$$

where $f\equiv(q/k_BT_j)J_{\rm opt}RA$. Likewise, we can write

$$KR=k\left(\frac{1}{W_p}+\frac{1}{W_n}\right)(\rho_pW_p+\rho_nW_n), \tag{30}$$

where we have assumed $k_n=k_p=k$.

In order to calculate the optimal diode geometry and bias, we first examine the symmetric case ($N_A=N_D$) and later look briefly at the asymmetric case. Equation (14) suggests that both the $p$-type and $n$-type regions should be maximally

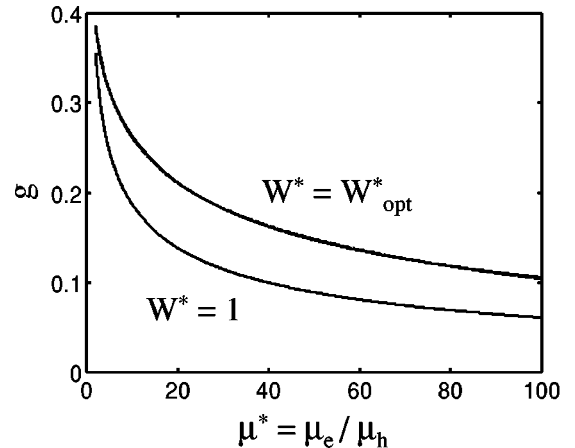


FIG. 12. $g(\mu^*)$ for $W^*=W^*_{\rm opt}$ and for $W^*=1$ (solid lines); fit $g_{\rm opt}=0.42-0.07\ln\mu^*$ (dotted line).

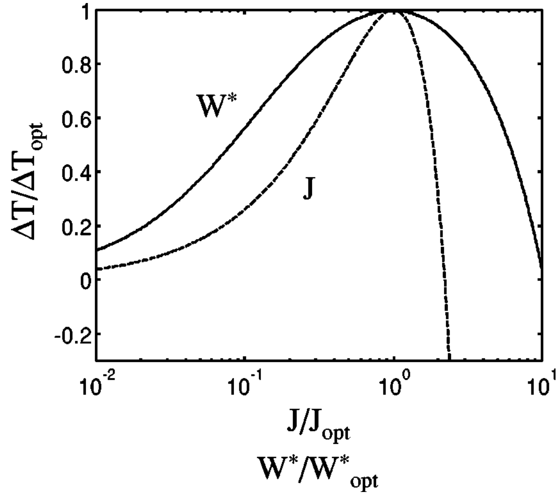


FIG. 13. Variation of $\Delta T$ around the optimality point defined by Eq. (34) for $\mu^*=30$.

doped in order to achieve the most cooling, making the symmetric case especially important. Equation (29) now simplifies to

$$1+f+\ln f=\ln\left(W^*+\frac{1}{W^*}+\mu^*+\frac{1}{\mu^*}\right), \tag{31}$$

where $W^*\equiv W_n/W_p$ and $\mu^*\equiv\mu_e/\mu_h$. Equation (30) can be written as

$$KR=k\rho_n\frac{(W^*+1)(W^*+\mu^*)}{W^*}. \tag{32}$$

The maximum value of the temperature difference $\Delta T = T_0 - T_j$ is then given by setting $Q_j=0$ in Eq. (27) so that the cooling at the junction exactly balances the heat conduction

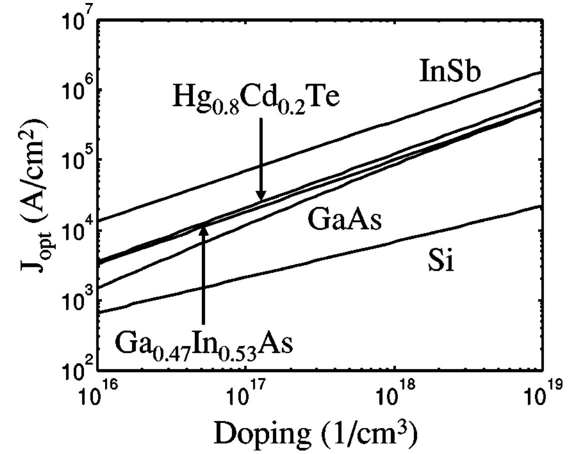


FIG. 15. Optimal value of $J$ for several semiconductor diode materials, assuming $W_n=1$ $\mu$m.

$$\Delta T=\frac{1}{2k\rho_n}\left(\frac{k_BT_j}{q}\right)^2\frac{(f^2+2f)W^*}{(W^*+1)(W^*+\mu^*)}$$
$$=\frac{1}{2k\rho_n}\left(\frac{k_BT_j}{q}\right)^2 g(W^*,\mu^*). \tag{33}$$

The optimal value of $W^*$ can be found in terms of $\mu^*$ by the condition $\partial\Delta T/\partial W^*=0$, as shown in Fig. 11; for typical devices, $\mu^*>1$. The function $g(W^*,\mu^*)$ defined in Eq. (33) can now be expressed as a function of just $\mu^*$, as shown in Fig. 12; $g_{\text{opt}}$ is very closely approximated by a first-order polynomial in $\ln\mu^*$. Also shown is the value of $g$ for a diode with equal region widths ($W^*=1$) for comparison. The optimal value of $\Delta T$ can now be derived from Eq. (33):

$$\Delta T_{\text{opt}}=\frac{1}{2k\rho_n}\left(\frac{k_BT_j}{q}\right)^2\left(0.42-0.07\ln\frac{\mu_e}{\mu_h}\right). \tag{34}$$

From comparison with Eqs. (25) and (24) we see that the effective Seebeck coefficients for the diode after optimizing for current density and geometry are proportional to $k_B/q$ by a factor which is dependent on the ratio of the mobilities.

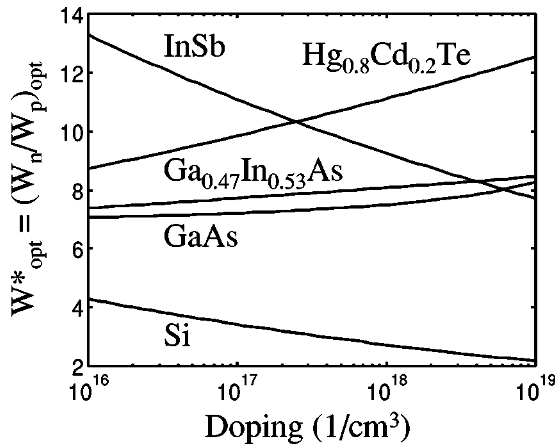


FIG. 14. Optimal value of $W^*$ for several semiconductor diode materials.

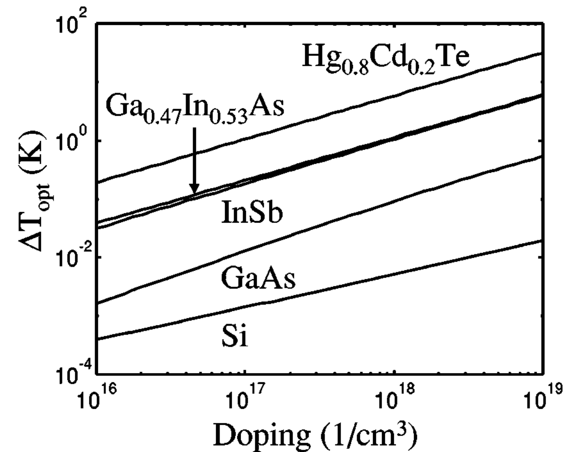


FIG. 16. Optimal value of $\Delta T$ for several semiconductor diode materials.

TABLE I. Material parameters of several semiconductors.

| | $k$ (W/cm K) | $\mu_e$[a] ($cm^2$/V s) | $\mu_h$[a] ($cm^2$/V s) | $\mu^*$[b] |
|---|---|---|---|---|
| Si | 1.31 | 3200–100 | 560–100 | 6–1 |
| GaAs | 0.55 | 8000–3000 | 400–100 | 20–30 |
| InSb | 0.18 | 100 000–10000 | 1000–400 | 100–25 |
| $Ga_{0.47}In_{0.53}As$ | 0.05 | 18 000–3000 | 800–100 | 22–32 |
| $Hg_{0.8}Cd_{0.2}Te$ | 0.01 | 20 000–5000 | 600–60 | 34–86 |

[a]Doping range $10^{16}-10^{19}$.
[b]Calculated.

Notice that increasing the total diode width $W_{tot}=W_n+W_p$ (while keeping $W^*$ constant) has no effect on $\Delta T_{opt}$ but does cause the optimal current $J_{opt}$ to decrease, since $f$ is fixed but $R$ increases.

In order to examine the dependence of $\Delta T$ on current density and width ratio, we plot $\Delta T/\Delta T_{opt}$ versus these variables in Fig. 13 for the value of $\mu^*=30$ (typical for many diodes). As we move the current density away from its optimal value, we keep $W^*$ fixed at its optimal value for $J=J_{opt}$, and vice versa. For $J<J_{opt}$ and for $W^*<W^*_{opt}$, conduction from the side heat sinks dominates, while for $J>J_{opt}$ and for $W^*>W^*_{opt}$, Joule heating dominates.

Values of $W^*_{opt}$, $J_{opt}$, and $\Delta T_{opt}$ are plotted for symmetrically doped diodes of several semiconductor materials in Figs. 14, 15, and 16, respectively. The material parameters used are summarized in Table I. From Fig. 14 we see that $W_p$ must often be kept small in order to ensure that $W_n$ is less than the diffusion length and satisfies the short-length approximation. In Fig. 17 we plot the optimal current as a function of total region width $W_{tot}$ for several materials doped symmetrically at $1\times10^{19}$ $cm^{-3}$. To examine the large internal diode temperature gradients that can occur in a material with very low thermal conductivity and high mobility such as $Hg_{0.8}Cd_{0.2}Te$, we plot the temperature profile of a 2 $\mu$m device doped symmetrically at $1\times10^{19}$ $cm^{-3}$ for several values of $W^*$ in Fig. 18, assuming $J=J_{opt}=3.85\times10^5$ A/$cm^2$. In Fig. 19, we assume $W^*=W^*_{opt}=12.5$ and plot the temperature profile for several values of J.

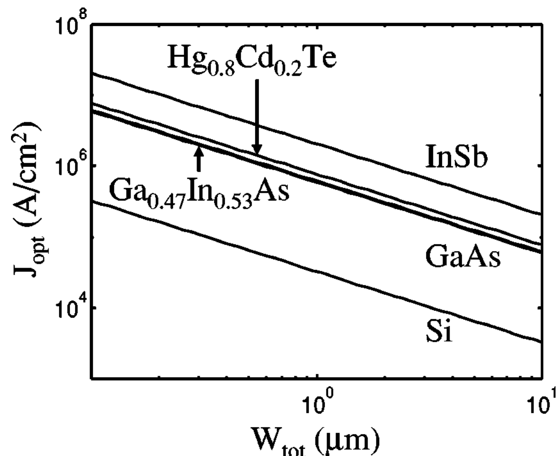


FIG. 17. $J_{opt}$ for several semiconductor diode materials doped symmetrically at $1\times10^{19}$ $cm^{-3}$.

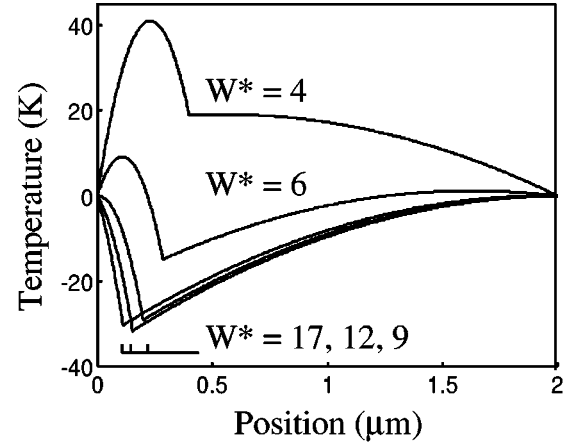


FIG. 18. Internal temperature profile (with respect to $T_0=0$) at several values of $W^*$ for a 2 $\mu$m $Hg_{0.8}Cd_{0.2}Te$ diode doped symmetrically at $1\times10^{19}$ $cm^{-3}$, assuming $J=J_{opt}=3.85\times10^5$ A/$cm^2$. The internal temperature minimum occurs at the junction for $T_{min}<0$.

In the above solution to Eq. (27) we have assumed symmetric doping. To examine the asymmetric case, we plot the optimal temperature and optimal current density for a range of doping profiles for a $Hg_{0.8}Cd_{0.2}Te$ diode in which $W_p=W_n=1$ $\mu$m, as shown in Figs. 20 and 21. As implied by Eq. (14), the best performance is achieved at the highest doping levels (unlike conventional majority-carrier thermoelectrics, in which there is an optimum carrier concentration) provided that the recombination lengths remain in the short length limit. When $W^*$ is set to 1 instead of $W_{opt}$, the optimal current density decreases, but it comes at the cost of a lower optimal temperature.

Moving the junction to the center of the diode can be beneficial if the heat sinks are not ideal. The cooling shown in Figs. 16 and 20 depends on the constraint that both con-

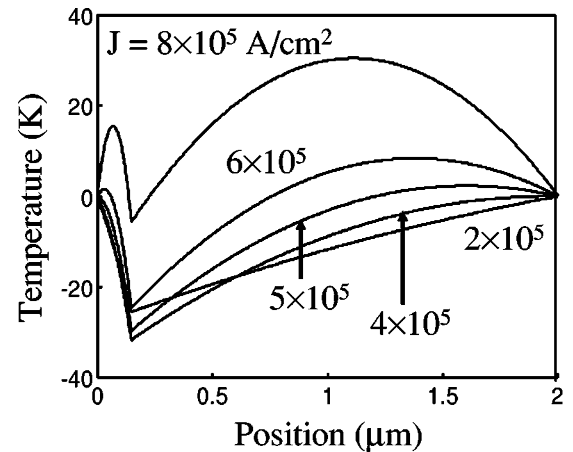


FIG. 19. Internal temperature profile (with respect to $T_0=0$) at several values of $J$ for a 2 $\mu$m $Hg_{0.8}Cd_{0.2}Te$ diode doped symmetrically at $1\times10^{19}$ $cm^{-3}$, assuming $W^*=W^*_{opt}=12.5$. The internal temperature minimum occurs at the junction.

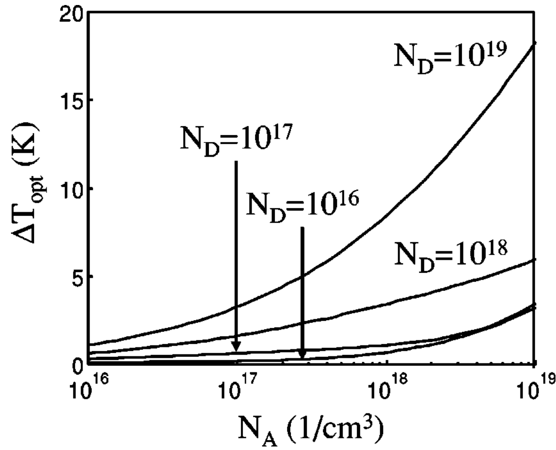


FIG. 20. Optimal temperature for a $Hg_{0.8}Cd_{0.2}Te$ diode with $W_p = W_n = 1$ $\mu$m.

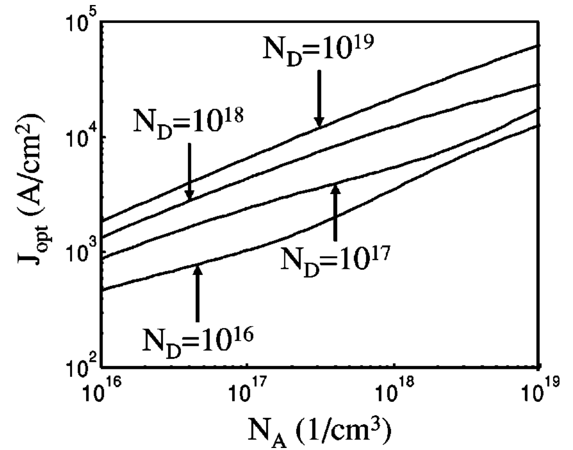


FIG. 21. Optimal current density for a $Hg_{0.8}Cd_{0.2}Te$ diode with $W_p = W_n = 1$ $\mu$m.

tacts are heat sunk at a constant temperature, while a real device can have contact resistance and nonideal heat sinking. For example, if a diode is grown on a thick semiconductor substrate, carriers injected from the top side will recombine over their diffusion length within the substrate; some of this heat will be conducted back to the junction. Depending on the device geometry, this can decrease the cooling effect.

The total amount of power dissipated in the conventional Peltier cooler is equal to $(JA)^2R$, while the diode dissipates $(JA)^2R + JAV_j$. The extra power dissipated in the diode is used to transport heat from the junction to the contacts (where it is often more readily conducted away), producing the thermal gradients discussed above. In the case for which carriers recombine optically, the total power dissipated can be less than $(JA)^2R$, and an analysis similar to that given in Sec. III B can be applied.

The dimensionless figure of merit $ZT$ is defined through Eq. (24) and is given by $2\Delta T_{\mathrm{opt}}/T$. The $ZT$ of most commercial Peltier coolers using bismuth telluride is approximately 1, while the value for a $Hg_{0.8}Cd_{0.2}Te$ diode doped symmetrically at $1\times 10^{19}$ cm$^{-3}$ is approximately 0.2. It is interesting to note that this figure is approximately the same as the bulk ZT derived elsewhere,[26] implying that bipolar thermoelectric effects, although nonlinear and dependent on coefficients that change with voltage, could be ultimately limited by the conventional thermoelectric bulk figure of merit. Based on the analytic expressions for bulk and diode thermoelectric figure of merit, no direct relation between them can be easily seen. This is especially the case since the optimum doping is different for the two cases. The fact that the numerical values of these figures of merit are similar (at the highest dopings that can be easily achieved for a diode) needs further investigation. While the diode is not appropriate as an external cooler, the internal temperature difference that it can sustain is significant. For a device in which external cooling is unsuitable, the internal cooling mechanism of the diode (and of bipolar devices in general) can be optimized.

## V. CONCLUSION

The bipolar Peltier coefficient can be used to model thermoelectric heat exchange as a function of device bias in minority-carrier devices. Analytic expressions can be derived which yield optimal bias conditions for thermoelectric cooling; several material systems show the capability for large internal temperature gradients. This type of internal thermoelectric cooling has many implications for device designs in which heat management is important and for which external cooling is not desirable or not effective.

The mechanism of diode thermoelectric cooling is especially relevant to semiconductor laser diodes, whose performance is strongly dependent on the junction temperature. Since most lasers employ heterostructures for the confinement of carriers and the optical mode, the analysis presented here must be extended in order to account for the traversal of carriers across heterointerfaces. In this case, numerical simulation of carrier transport can be used to derive bias-dependent Peltier coefficients for each layer in a device. This method has been used to show that laser diodes can be optimized through the use of type-II heterointerfaces and appropriate doping such that thermoelectric cooling causes a significant decrease in the temperature of the active region.[27]

## ACKNOWLEDGMENT

A.S. would like to acknowledge support from Grant No. NSF-CTS-0103609.

[1] G.C. Dousmanis, C.W. Mueller, H. Nelson, and K.G. Petzinger, Phys. Rev. **133**, A316 (1964).
[2] W. Nakwaski, Electron Technol. **13**, 61 (1982).
[3] S.R. Bowman and C.E. Mungan, Appl. Phys. B: Lasers Opt. **71**, 807 (2000).
[4] H. Gauck, T.H. Gfroerer, M.J. Renn, E.A. Cornell, and K.A. Bertness, Appl. Phys. A: Mater. Sci. Process. **64**, 143 (1997).
[5] L.D. Hicks, T.C. Harman, X. Sun, and M.S. Dresselhaus, Phys.

Rev. B **53**, R10 493 (1996).

[6] A. Shakouri, P. Abraham, and J.E. Bowers, Appl. Phys. Lett. **71**, 1234 (1997).

[7] C. LaBounty, A. Shakouri, P. Abraham, and J.E. Bowers, Opt. Eng. **39**, 2847 (2000).

[8] K.P. Pipe, R.J. Ram, and A. Shakouri, in *Proceedings of the American Society of Mechanical Engineers Heat Transfer Division*, edited by Y. Jaluria (ASME, New York, 2001), Vol. 7, pp. 219–222.

[9] K.P. Pipe, R.J. Ram, and A. Shakouri, in *OSA Trends in Optics and Photonics (TOPS)*, Vol. 56 of *Conference on Lasers and Electro-Optics* (CLEO 2001), Technical Digest, Postconference Edition (Optical Society of America, Washington, DC, 2001), pp. 8–9.

[10] Yu.G. Gurevich, O.Yu. Titov, G.N. Logvinov, and O.I. Lyubimov, Phys. Rev. B **51**, 6999 (1995).

[11] T. Piotrowski and S. Sikorski, Semicond. Sci. Technol. **16**, 750 (2001).

[12] G.K. Wachutka, IEEE Trans. Comput.-Aided Des. **9**, 1141 (1990).

[13] L.L. Liou, J.L. Ebel, and C.I. Huang, IEEE Trans. Electron Devices **40**, 35 (1993).

[14] G.W. Charache *et al.*, J. Electron. Mater. **27**, 1038 (1998).

[15] O. Yu. Titov *et al.*, in *Proceedings of the 16th International Conference on Thermoelectrics, Dresden, Germany, 1997*, IEEE Catalog No. 97TH8291, edited by A. Heinrich and J. Schumann (IEEE, Piscataway, NJ, 1997), p. 661.

[16] H.J. Goldsmid, *Electronic Refrigeration* (Pion, London, UK, 1986).

[17] W. Nakwaski, Electron Technol. **14**, 81 (1983).

[18] J. Tauc, *Photo and Thermoelectric Effects in Semiconductors* (Pergamon Press, New York, 1956).

[19] S.M. Sze, *Physics of Semiconductor Devices*, 2nd ed. (Wiley & Sons, New York, 1981).

[20] Yu.G. Gurevich and S.I. Shevchenko, Sov. Phys. JETP **35**, 426 (1972).

[21] A. Shakouri *et al.*, Microscale Thermophys. Eng. **2**, 37 (1998).

[22] D.W. Winston and R.E. Hayes, IEEE J. Quantum Electron. **34**, 707 (1998).

[23] J.F. Carlin *et al.*, J. Cryst. Growth **202**, 994 (1999).

[24] H.Z. Fardi, D.W. Winston, R.E. Hayes, and M.C. Hanna, IEEE Trans. Electron Devices **47**, 915 (2000).

[25] D.M. Rowe, *CRC Handbook of Thermoelectrics* (CRC Press, Boca Raton, FL, 1995).

[26] J.O. Sofo, G.D. Mahan, and J. Baars, J. Appl. Phys. **76**, 2249 (1994).

[27] K.P. Pipe, R.J. Ram, and A. Shakouri, IEEE Photonics Technol. Lett. **14**, 453 (2002).