Minimum length scales for enhancement of the power factor in thermoelectric nanostructures

P. Pichanusakorn and P. R. Bandaru[a)]
Department of Mechanical Engineering, Materials Science Program, University of California, San Diego, La Jolla, California 92093-0411, USA

(Received 25 November 2009; accepted 6 February 2010; published online 6 April 2010)

It is shown, through a comparison of the respective electron density distribution and the density of states (DOS), that there exists an optimal length scale only below which the thermoelectric power factor of nanostructures is enhanced over the bulk value. The comparison is done for an optimal value of the Seebeck coefficient in three, two, and one dimensions for Si/SiGe, Bi_2Te_3, PbTe, and $SrTiO_3$ with various scattering mechanisms. It is then concluded that the increase in the magnitude of the integrated DOS and *not* the change in shape, as is commonly believed, to be most responsible for the increases in the power factor. © *2010 American Institute of Physics*.
[doi:10.1063/1.3359659]

I. INTRODUCTION

The solid state based conversion of heat to electricity, and vice versa, using thermoelectric materials have many tantalizing applications such as waste heat recovery in automobiles or industrial processes, body-heat powered biomedical devices, solar thermal energy generation, spot cooling of microelectronics, optoelectronics, etc. However, the efficiency of traditional bulk thermoelectric materials has been quite low (<10%). Therefore, much excitement was generated when theoretical predictions[1,2] and experiments[3–5] indicated that the efficiency could be greatly increased through the use of low-dimensional nanostructures such as quantum wells (QWs) or nanowires (NWs).

The efficiency of a thermoelectric material is generally indicated by the figure of merit, $ZT=(S^2\sigma/\kappa)T$, where S is the Seebeck coefficient, σ the electrical conductivity, κ the thermal conductivity, and T, the temperature. There are generally two approaches to increasing ZT using nanostructures. First, κ could be significantly reduced through enhanced or modified scattering of phonons by the surfaces of NWs,[6] interfaces of QW superlattices,[7] or the grain boundaries in nanocrystalline materials.[8] Alternatively, the power factor ($S^2\sigma$) could be enhanced through electron confinement.[1,2] In the latter approach, the confinement length, which is nominally the thickness of the QW or the diameter of NW, is the critical parameter that determines whether and how much the $S^2\sigma$ may be enhanced. In this paper, we examine how the confinement length required to achieve an enhancement in power factor depends on innate material properties, such as the effective mass, and the temperature of operation.

II. DESCRIPTION OF DIFFUSIVE ELECTRON TRANSPORT PARAMETERS

The electrical conduction and the Seebeck effect are the manifestations of electron diffusion along a concentration gradient which is established in the presence of potential or temperature gradient, respectively. Other physical effects such as ballistic conduction or phonon-drag Seebeck effect may also be present, but is beyond the scope of this paper. The diffusion of electrons is generally described by the Boltzmann transport equation (BTE), whose solutions are used in conjunction with material-specific density of states (DOS), $g(E)$, and relaxation time, $\tau(E)$, to determine the diffusive Seebeck coefficient (S) and the electrical conductivity (σ).

From the solution of the BTE, the S is described through the following equation:

$$S = \mp \frac{1}{eT}\left(\frac{\int_0^{+\infty} g(E)\tau(E)v^2(E)E\frac{df(E)}{dE}dE}{\int_0^{+\infty} g(E)\tau(E)v^2(E)\frac{df(E)}{dE}dE} - E_F\right) = \mp \frac{1}{eT}\left(\frac{\langle \tau E\rangle}{\langle \tau\rangle} - E_F\right). \quad (1)$$

In the above, E_F is the Fermi energy relative to the band/subband's ground energy level (which is arbitrarily set to be zero here), $v(E)$ is the velocity as a function of energy E, $f(E)$ is the Fermi–Dirac distribution function, and $\pm e$ is the unit charge of holes/electrons, respectively. The expression in the middle of Eq. (1) was derived without assuming any particular relations for the $g(E)$, $\tau(E)$, and $v(E)$. However, if a power law dependence is assumed, then the expression on the right follows, where $\langle x\rangle = \int_0^{+\infty} g(E)x(E)f(E)dE / \int_0^{+\infty} g(E)f(E)dE$ is the average value of the parameter, $x(E)$ at equilibrium. The latter expression in Eq. (1), intuitively describes the S as proportional to the difference between the average energy of electrons weighed by the relaxation time at equilibrium and the Fermi energy.

The DOS and $v(E)$ follow a power law for parabolic energy band dispersion. For a single band, or subband (in a QW/NW) the DOS is

a)Electronic mail: pbandaru@ucsd.edu.

$$g(E) = \frac{N}{g_D a^{3-D}} \left(\frac{2m_d}{\hbar^2} \right)^{D/2} E^{D/2-1}, \tag{2}$$

where $\hbar$ is the Planck constant, m_d is the DOS effective mass, N is the number of degenerate conduction valleys intrinsic to the band, a is the confinement length (nominally equal to the QW thickness or NW diameter), D is the dimensionality factor (=3, 2, and 1 for bulk, QW, and NW, respectively), and g_D is a constant equal to $2\pi^2$ for $D=3$, and equal to $D\pi$ for $D=2$ or 1. The velocity also varies through $v^2(E)=2E/Dm_\sigma$, where m_σ is the conductivity or inertial effective mass. For simplicity, we will assume isotropic effective mass in our subsequent calculations, in which case $m=m_d=m_\sigma$.

The relaxation time may be described through $\tau(E)=\tau_0 E^r$, where r is a characteristic constant related to a particular scattering mechanism and τ_0 is a constant related to both the mechanism and the material properties.[9] Common values of r are $-1/2$ for strongly-ionized impurity and for acoustic and optical phonon mediated deformation potential scattering, or $+3/2$ for weakly-ionized impurity scattering in bulk materials. Occasionally, a constant relaxation time approximation is assumed, with $r=0$.

Incorporating these specific functions for $g(E)$, $\tau(E)$, and $v(E)$ into Eq. (1), the S is explicitly written as

$$S = \mp \frac{k_B}{e} \left[\frac{\left(r + \frac{D}{2} + 1 \right) F_{r+D/2}(\eta)}{\left(r + \frac{D}{2} \right) F_{r+D/2-1}(\eta)} - \eta \right], \tag{3}$$

where k_B is the Boltzmann constant, $\eta=E_F/k_BT$ is the reduced Fermi energy and $F_j(\eta)=\int_0^{+\infty} x^j/[\exp(x-\eta)+1]dx$ is the jth order Fermi integral. It is noticed that in this form, the S is independent of a, T, m, N, and τ_0.

The σ can also be derived from the solution to the BTE. For brevity, we only present the final expressions for the carrier concentration (n) and electron mobility (μ) as

$$n = \frac{N}{g_D a^{3-D}} \left(\frac{2k_B T m_d}{\hbar^2} \right)^{D/2} [F_{D/2-1}(\eta)], \tag{4}$$

$$\mu = \frac{e\tau_o}{m_\sigma} (k_B T)^r \left[\left(\frac{2r}{D} + 1 \right) \frac{F_{r+D/2-1}(\eta)}{F_{D/2-1}(\eta)} \right]. \tag{5}$$

For a constant relaxation time, μ takes a familiar form $\mu=e\tau_o/m_\sigma$. The σ is then a product of n and μ through $\sigma=ne\mu$, and the power factor, $S^2\sigma$, may be calculated from Eqs. (3)–(5).

III. THE OPTIMAL SEEBECK COEFFICIENT FOR MAXIMUM POWER FACTOR

Since $|S|$ is proportional to the difference of the average electron energy, $\langle E \rangle$ and the lowest energy level (*e.g.*, E_F at 0 K)—from Eq. (1), it is increased as η is decreased, i.e., if T is increased (which increases the $\langle E \rangle$) or if E_F is decreased. On the other hand, σ tends to decrease as E_F and η is decreased due to decreasing n, although it may increase as T is increased (i.e., $n \propto T^{D/2}$). Consequently, the power factor ($S^2\sigma$) tends to increase with increasing T assuming a constant E_F, but exhibits a peak at an optimal E_F when constant T is assumed.

TABLE I. The optimal reduced Fermi energy, η_{opt} (first row) and corresponding Seebeck coefficient, S_{opt} (second row), for maximum power factor is peaked, as a function of the scattering constant, r and dimensionality, D. The r value for strongly-screened ionized impurity scattering and acoustic and optical phonons scattering via deformation potential changes with D[9] (Table edited from Ref. 10).

	$r=-1/2$	$r=0$	$r=+1/2$
$D=3$	0.67[a]	2.47[b]	No limit
	167[a]	130[b]	⋯
$D=2$	−0.37	0.67[a,b]	No limit
	187	167[a,b]	⋯
$D=1$	$\eta_{opt}=-1.14$	−0.37[b]	0.67[a]
	$S_{opt}=197$	187[b]	167[a]

[a]Ionized impurity photons.
[b]Constant relaxation time.

The optimal E_F can be easily found, as we have previously shown,[10] through determining the value of the optimal *reduced* Fermi energy (η_{opt}), which was also found to be independent of a and T, and the material dependent parameters such as m, N, and τ_0. This is justified as the expression for $S^2\sigma$, from Eqs. (3)–(5), essentially consists of the product of two functions, *i.e.*, $S^2\sigma=h(a,m,N,\tau_0,T)p(r,D,\eta)$. Subsequently, the peak $S^2\sigma$ must coincide with the maximum of $p(r,D,\eta)$, regardless of m, N, a, and τ_0, which are constants for a given material at T. The η_{opt} is listed in Table I (first row of each entry) for common r and D values. The ionized impurity scattering and phonon scattering via deformation potential both have an r value of $-1/2$ in bulk material. However, since their scattering rate is dependent on the DOS, the r value correspondingly changes to 0 and 1/2 in two and one dimension, respectively. In some cases ($r=1/2$, $D=3$ or 2), the additional increase in μ is sufficiently large to compensate for the corresponding reduction in $|S|$ as the E_F is increased; thus the power factor may increase without limit in these regimes, and there is no optimal E_F.

Since the S also depends solely on r, D, and η, there is a corresponding optimal Seebeck coefficient (S_{opt}) that is independent of m, N, a, τ_0, or T, as well. S_{opt} may be either positive or negative, and is listed in units of μV/K in the second row of each entry in Table I. On the other hand, the optimal carrier concentration (n_{opt}) would not be a constant like η_{opt} or S_{opt}, but could change with T and a, and is material specific. In Fig. 1 (reproduced from Ref. 10), the n_{opt} to achieve peak power factor for PbTe at 300 K is approximately 7×10^{18} cm^{-3}, while that of Bi_2Te_3 is shifted from approximately 4×10^{18} to 9×10^{18} cm^{-3} as the T is increased from approximately 100 to 150 K. However, while n_{opt} may change with material and temperature, the peak $S^2\sigma$ always occurs near the predicted S_{opt} values of 130–167 μV/K, as shown in both the main graph and the inset.

IV. ENHANCEMENT OF THE PEAK POWER FACTOR IN NANOSTRUCTURED THERMOELECTRICS

The peak $S^2\sigma$ can be increased through electron confinement in QW or NW based materials. Such an enhancement

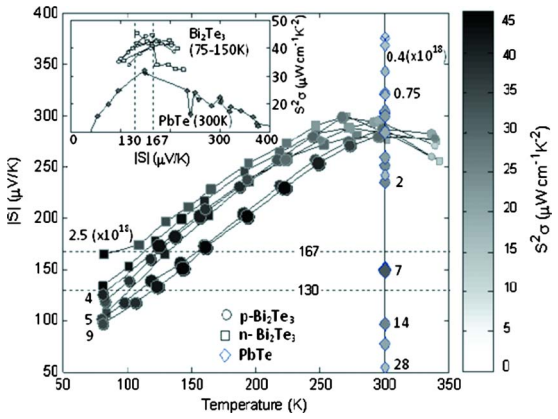

FIG. 1. (Color online) The power factor ($S^2\sigma$) of Bi_2Te_3 and PbTe as a function of temperature. The $S^2\sigma$ is represented by the size and shade of the data points (larger and darker indicating higher values). While the optimal carrier concentration (in 10^{18} cm^{-3}) where the peak $S^2\sigma$ occurs is material and temperature dependent, the optimal Seebeck coefficient, S_{opt}, is always ~130–167 $\mu V/K$, as shown in the *inset*. (Figure reproduced from Ref. 10).

arises mainly from the increase in the magnitude of the DOS, which is illustrated through Fig. 2. In Fig. 2(a), the DOS of bulk Si ($m=0.32m_0$, $N=6$, and m_0 is the free electron mass), at 300 K, is illustrated. Assuming a constant relaxation time ($r=0$), the optimal E_F (with $\eta_{opt}=2.47$, from Table I) is 0.063 eV ($=2.47\times k_B\times 300$ K), and $n=9.7\times 10^{19}$ cm^{-3}. $|S|$ is proportional to the difference between E_F and the $\langle E\rangle$, the latter of which was computed to be 0.102 eV, given an S_{opt} of 130 $\mu V/K$. Since the optimal E_F is assumed, the peak $S^2\sigma$ at 300 K has been reached and cannot be increased any further by doping. However, the peak $S^2\sigma$ may yet be increased by changing the DOS of Si through electron confinement in lower dimensional structures, *e.g.,* NWs.

In Fig. 2(b), the DOS of 2 and 5 nm diameter Si NWs is shown in comparison to that of bulk Si. The shapes of the DOS of both the NWs are equivalent, but very different from that of bulk Si. The optimal E_F for both NWs is also different from the bulk value, and equals −0.010 eV at 300 K (with $\eta_{opt}=-0.37$, from Table I). Since the 2 nm wire has a smaller confinement length, a, its DOS is larger than that of the 5 nm, as can also be deduced from Eq. (2). Given the same E_F,

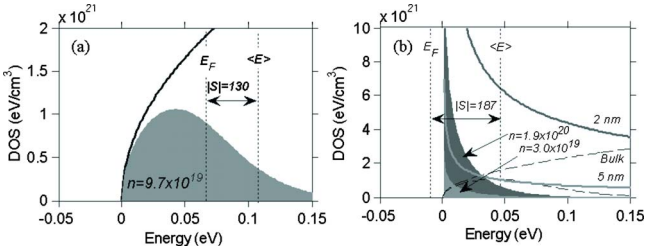

FIG. 2. (Color online) (a) The electron density distribution (shaded area) overlaid on the parabolic variation ($\sim E^{1/2}$) of the DOS (solid line) in the conduction band (CB), of power factor *optimized* bulk Si (at 300 K) $E_F=0.063$ eV and the average electron energy, $\langle E\rangle=0.102$ eV. The energy separation between E_F and $\langle E\rangle$ corresponds to a S_{opt} of 130 $\mu V/K$. (b) Electron density distribution, in the CB of *one-dimensional* Si NWs, of diameter 2 and 5 nm in comparison to the bulk electron distribution. For both the 2 and 5 nm NWs, S_{opt} is ~187 $\mu V/K$, with $E_F=-10$ meV and $\langle E\rangle=4.7$ meV. A larger (/smaller) S^2n, compared to the bulk, is obtained for the 2 nm (/5 nm) NW. The density is expressed per unit volume for comparison with Fig. 2(a).

the optimal carrier concentration for the 2 nm wire ($n_{opt}=1.9\times 10^{20}$ cm^{-3}) is larger than that for the 5 nm wire ($n_{opt}=3\times 10^{19}$ cm^{-3}). However, since the DOS and the distribution of electrons are uniformly scaled through changes in a, the $\langle E\rangle$ and $|S|$ would have the same value of 0.047 eV and 187 $\mu V/K$ for both wires, respectively.

Assuming that the μ is unchanged, the peak $S^2\sigma$ for the 2 nm NW could be four-times *larger* than that of bulk value due to the enhancement in the peak S^2n alone. On the contrary, the peak $S^2\sigma$ of the 5 nm NW would be 36% *smaller* than that of the bulk due to the reduced DOS. However, it was postulated μ could be reduced by 80% due to strong scattering induced by large confinement potential.[11] While an accurate prediction of μ is beyond the scope of the present study, nonetheless we could account for the reduction in μ through a consideration of the gain factor, which will be discussed later in he text.

Figure 2 shows that the $S^2\sigma$ is increased primarily because of the increase in the magnitude of the DOS, and not because of the change in shape. Similar analysis on two-dimensional QW structures, which have a flat DOS, also showed that the maximum $S^2\sigma$ can be either larger or smaller than that of bulk material depending on the confinement length. It can then be inferred that the increase in S^2n is due to an enhancement in n_{opt} brought about by an increase in the magnitude of the DOS, rather than the change in shape of the DOS, as is often believed. The latter conclusion is often drawn through the Mott expression,[12] *viz.,* $|S|=\pi^2k_B/3e\eta[1/n(E)dn(E)/d\ln(E)+1/\mu(E)d\mu(E)/d\ln(E)]$ which suggests that $|S|$ can be increased by increasing the energy dependence of the DOS through the $n(E)$ term, i.e., by imposing a sharply-peaked DOS as was invoked to explain recent observations in Tl doped PbTe.[13] However, such an explanation is contradicted by the fact that the energy dependence of the DOS is actually *reduced* from 1/2 in bulk to 0 and −1/2 in QWs and NWs, respectively. Instead, as we have shown, the $|S|$ enhancement is due to the increase in magnitude of the DOS, which leads to the reduction in E_F and η, at a given carrier concentration. Since the magnitude of the DOS is inversely proportional to the confinement length, from Eq. (2), there is a critical confinement length, a_{min}, only below which the $S^2\sigma$ increase may be observed.

The a_{min} required to achieve an arbitrary gain, in the $S^2\sigma$, by a gain factor G was found through solving $(S^2\sigma)_{r,D}=G(S^2\sigma)_{r,D=3}$ at η_{opt}. For constant relaxation time, this yields

$$a_{min}=\frac{\hbar}{\sqrt{2k_BT}}\times\left[\frac{N_D}{N_3}\frac{1}{\sqrt{\Pi_{i=1}^{3-D}m_{c,i}}}\frac{2\pi}{GD}\frac{F_{D/2-1}(\eta_{opt})}{F_{1/2}(2.47)}\left(\frac{S_{opt}}{130\frac{\mu V}{K}}\right)^2\right]^{1/3-D}. \quad (6)$$

To avoid a reduction in the DOS, the confinement direction should generally be chosen such that the number of conduction valleys, N, is preserved.[11] As an example, the four conduction valleys in PbTe are situated on the [111] axes, which implies that for a QW grown in a particular [111] orientation,

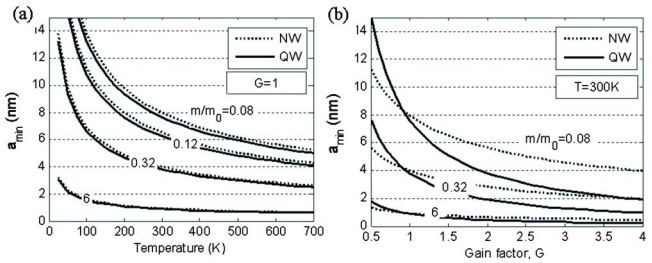

FIG. 3. (Color online) (a) The minimum confinement length (a_{min}) for select thermoelectric materials, at which *one-dimensional* NWs (dotted lines) and *two-dimensional* QWs (solid lines) have a *greater* $S^2\sigma$ compared to bulk thermoelectrics (the gain factor, $G=1$). a_{min} increases with decreasing DOS electron effective mass—Bi_2Te_3 ($m/m_o=0.08$), PbTe ($m/m_o=0.12$), SiGe ($m/m_o=0.32$), and $SrTiO_3$ ($m/m_o=6$) and decreased temperature. (b) The a_{min} as a function of G (the ratio of the power factor of the nanostructure compared to the bulk) for NWs and QWs, decreases much faster for the latter at higher G.

the degeneracy of the four conduction valleys would be broken and yield one longitudinal valley and three oblique valleys. This could, in turn, have the effect of reducing N from four to one. However, if the QW structure is grown along the [100] axes, the valley degeneracy would be preserved. We will assume that such a condition could be met, and that N is the same for both bulk material and nanostructures. a_{min} also depends on the effective mass in the direction/s of confinement (m_c), *e.g.,* there will be two values of m_c in a NW. In our calculations, we assume an isotropic effective mass, where $m_c=m$.

The a_{min} calculation shows that the enhancement becomes increasingly difficult to obtain at higher temperature or with materials with large effective masses, as smaller nanostructures would be needed. In Fig. 3(a), we have plotted the a_{min} required to achieve parity, i.e., with $G=1$, for a number of m associated with thermoelectric materials, *e.g.,* Bi_2Te_3, PbTe, SiGe, and $SrTiO_3$. This figure shows that typically, a QW/ (/NW) thickness (/diameter) under 10 nm is required at 300 K, and this a_{min} decreases as the m is increased. However, as the temperature is decreased, the a_{min} increases, suggesting that nanostructured thermoelectric devices may be best suited for lower temperature applications. Note that our gain factor so far accounts only for the increase due to S^2n, while assuming a constant μ. If μ is somehow reduced by, say, 80%, then a gain factor of $G=1/(1-0.8)=5$ would actually be required for the nanostructures to have the same power factor as the bulk material.

Although the a_{min} for QWs and NWs are similar for parity with bulk i.e., at $G=1$, the decrease is more rapid for QWs than NWs at higher G values, as shown in Fig. 3(b). This is consistent with the predictions of larger $S^2\sigma$ enhancement in NWs compared to QWs.[1,2] The figure also indicates that while a_{min} is larger for material with small effective mass, it also decreases much faster as the G is increased.

If acoustic or optical phonon deformation potential or strongly-screened ionized impurity scattering is assumed where $r=1-D/2$, then

$$a_{min}=\frac{\hbar}{\sqrt{2}}\left[\frac{N_D}{N_3}\frac{1}{\sqrt{\Pi_{i=1}^{3-D}m_{c,i}}}\frac{\tau_{0,D}}{\tau_{0,D=3}}\frac{6\pi}{GD^2}\right]^{1/3-D}. \quad (7)$$

The ratio $\tau_{0,D=2}/\tau_{0,D=3}$ can be quite complex and difficult to determine, and will depend on the specific scattering mechanisms. For example, with acoustic phonon scattering this ratio will involve the electron-phonon interaction matrix elements.[9] However, it is interesting to note that while the dependence on effective mass remains the same, a_{min} would now be temperature independent.

V. CONCLUSIONS

In our previous work, we had shown that for any material that the maximum $S^2\sigma$ at a given temperature could be expected when S is in the optimal range of 130–187 μV/K. With this as the basis, we have now shown that there exists a threshold confinement length only below which lower dimensional nanostructures, such as QWs or NWs, could have a higher power factor compared to bulk materials. For example, Si NWs with diameter below 4 nm would be *required* for an enhanced power factor over bulk values. However, to date, while the characterization of σ has been performed in sub 5 nm wires,[14] S and κ measurements have been reported only in NWs >10 nm in diameter.[6,15,16] Consequently, further experimental work is necessary to probe smaller length scales and verify the results reported in this paper. Additionally, we have shown that it is the increase in magnitude of the DOS and not the change in shape that increases the power factor.

ACKNOWLEDGMENTS

We gratefully acknowledge support from the National Science Foundation (Grant No. ECS-06-43761) and N. B. Elsner of Hi-Z Inc. A refinement to Table I was enabled through communications with Professor M. Lundstrom (Purdue University).

[1] L. D. Hicks and M. S. Dresselhaus, Phys. Rev. B **47**, 12727 (1993).
[2] L. D. Hicks and M. S. Dresselhaus, Phys. Rev. B **47**, 16631 (1993).
[3] L. D. Hicks, T. C. Harman, X. Sun, and M. S. Dresselhaus, Phys. Rev. B **53**, R10493 (1996).
[4] T. C. Harman, D. L. Spears, and M. J. Manfra, J. Electron. Mater. **25**, 1121 (1996).
[5] Y. Mune, H. Ohta, K. Koumoto, T. Mizoguchi, and Y. Ikuhara, Appl. Phys. Lett. **91**, 192105 (2007).
[6] A. I. Boukai, Y. Bunimovich, J. Tahir-Kheli, J.-K. Yu, W. A. Goddard III, and J. R. Heath, Nature (London) **451**, 168 (2008).
[7] R. Venkatasubramanian, E. Siivola, T. Colpitts, and B. O'Quinn, Nature (London) **413**, 597 (2001).
[8] B. Poudel, Q. Hao, Y. Ma, Y. Lan, A. Minnich, B. Yu, X. Yan, D. Wang, A. Muto, D. Vashaee, X. Chen, J. Liu, M. S. Dresselhaus, G. Chen, and Z. Ren, Science **320**, 634 (2008).
[9] M. Lundstrom, *Fundamentals of Carrier Transport*, 2nd ed. (Cambridge University Press, Cambridge, 2000).
[10] P. Pichanusakorn and P. Bandaru, Appl. Phys. Lett. **94**, 223108 (2009).
[11] D. A. Broido and T. L. Reinecke, Phys. Rev. B **64**, 045324 (2001).
[12] D. K. C. MacDonald, *Thermoelectricity: An Introduction to the Principles* (Dover, Mineola, NY, 1962).
[13] J. P. Heremans, V. Jovovic, E. S. Toberer, A. Saramat, K. Kurosaki, A. Charoenpakdee, S. Yamanaka, and G. J. Snyder, Science **321**, 554 (2008).
[14] M. T. Björk, H. Schmid, J. Knoch, H. Riel, and W. Riess, Nat. Nanotechnol. **4**, 103 (2009).
[15] A. I. Hochbaum, R. Chen, R. D. Delgado, W. Liang, E. C. Garnett, M. Najarian, A. Majumdar, and P. Yang, Nature (London) **451**, 163 (2008).
[16] D. Li, Y. Wu, P. Kim, L. Shi, P. Yang, and A. Majumdar, Appl. Phys. Lett. **83**, 2934 (2003).