Materials Science and Engineering R 67 (2010) 19–63

Contents lists available at ScienceDirect

Materials Science and Engineering R

journal homepage: www.elsevier.com/locate/mser

Nanostructured thermoelectrics

Paothep Pichanusakorn, Prabhakar Bandaru *

Department of Mechanical & Aerospace Engineering (Materials Science program), 9500 Gilman Drive, MC 0411 UC, San Diego, La Jolla, CA 92093, United States

ARTICLE INFO

Keywords:
Thermoelectrics
Nanostructures
Energy conversion
Seebeck coefficient
Power factor
Thermal conductivity

ABSTRACT

Thermal to electrical energy conversion, through thermoelectric and thermionic materials, has been proposed to be much more efficient in lower dimensional materials at the nanoscale. In this paper, we review the underlying materials physics of nanostructured thermoelectrics which gives rise to such enhanced efficiency. We first study the basic phenomenology of the contributing terms to the power factor in the thermoelectric figure of merit, i.e., the Seebeck coefficient (S) and the electrical conductivity (σ), which are analyzed through the Boltzmann transport formalism and then thoroughly compared to recent experiments in nanostructures. Additional factors, hitherto not given much consideration, such as carrier scattering time approximations vis-à-vis dimensionality and the density of states (DOS) are also studied. Through such a study, we postulate that it is the sheer magnitude and not the specific shape of the DOS that is important in enhancing the thermoelectric power factor. We then see that most of the understood increase in the figure of merit of nanostructured thermoelectrics has been accomplished through a drastic reduction of the lattice thermal conductivity (κ_L)—by two orders of magnitude from the bulk values through the introduction of scattering at different length scales. Such a reduction has provided a large impetus for the use of nanostructures and will be reviewed. We next consider solid state implementations of thermionic structures, which seem to be converging towards thermoelectric devices and hence can be described by a similar figure of merit. Approaches for energy filtering and further increasing efficiencies are also described.

Contents

* Corresponding author. Tel.: +1 858 534 5325; fax: +1 858 534 5698.
E-mail address: pbandaru@ucsd.edu (P. Bandaru).

0927-796X/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.mser.2009.10.001

1. Introduction

The problem of *waste heat recovery*, e.g., from transportation vehicles and oil refineries along with *heat dissipation*, e.g., in microelectronics – where leakage currents in transistors exponentially increase with temperature, is of considerable relevance to energy generation and conservation. Thermoelectric materials can be used to convert heat to electricity, through the Seebeck effect (say, in thermocouples) or can be used for cooling or refrigeration through the converse Peltier effect (say, for picnic coolers). While the basic principles underlying these processes are well known, the use of these materials is not very widespread due to low efficiency. However, we could be at the threshold of drastic improvement. Theoretical calculations and laboratory demonstrations over the past few years suggest that size effects in nanostructures such as quantum wells, nanowires, and quantum dots, significantly increase the efficiency and figures of merit of thermoelectric materials. If nanostructured thermoelectrics could be practically implemented, such an approach could be of wide practical import. For example, mechanical cycle-based systems could be replaced by solid state based thermoelectric structures with advantages such as components with no moving parts along with a rapid response time. Moreover, the latter technologies are form-factor adaptable. A brief glimpse of the possibilities is illustrated in Fig. 1.

1.1. The figure of merit

In the most basic form, as practically implemented for energy conversion – see Fig. 2(a) – a thermoelectric device consists of an element placed between a *heat source*, e.g., corresponding to waste heat generation and the ambient (*heat sink*). The transfer of heat from the source to the sink is either through the motion of the carriers (electrons/holes) or through the lattice (through collective lattice vibration modes/phonons). The carrier transport results in a development of a potential difference – the Seebeck voltage (ΔV). The thermopower/Seebeck coefficient S is then the ratio of ΔV to the temperature difference (ΔT). It is noted that the ΔV and S could arise from at least three sources [1], which includes contributions from (1) the work function difference between the contact and the thermoelectric, S_ϕ (2) the diffusive current of the carriers, S_d, and (3) the action of the carriers on the phonons, the *phonon drag* component, S_{ph} i.e., $S = S_\phi + S_d + S_{ph}$. Typically, S_d is the most dominant component, of practical interest, for thermoelectrics and will be extensively discussed in Section 2. While S_{ph} has traditionally been considered at low temperatures [2,3] and due the presence of impurities [2], recent studies in silicon nanowires [4] have interpreted increased values through an increased S_{ph} (Section 3.5) While S_ϕ could be important and is considered in thermionic applications, the detailed effects are still being extensively investigated. As the diffusive effect usually predominates, in the paper, S_d will often be referred to as S in the paper.

However, the efficiency of a thermoelectric device depends on more than just S. Consideration of electrical and heat current in a typical device, as shown in Fig. 2, reveals that both the power generation efficiency and the refrigeration's coefficient of

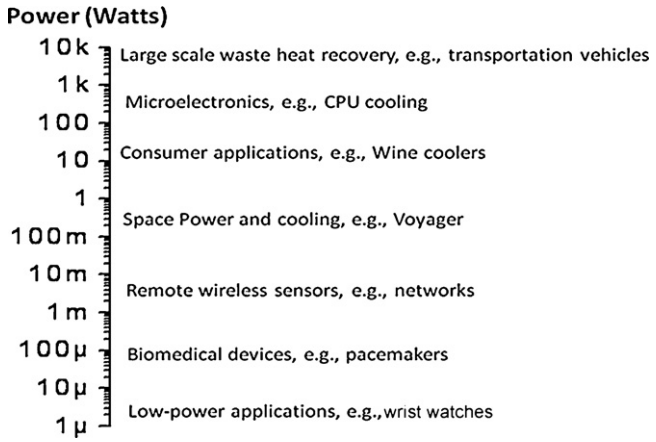

Fig. 1. Thermoelectric materials can be put to use in various energy conversion applications, encompassing ten orders of magnitude in power, as illustrated above.

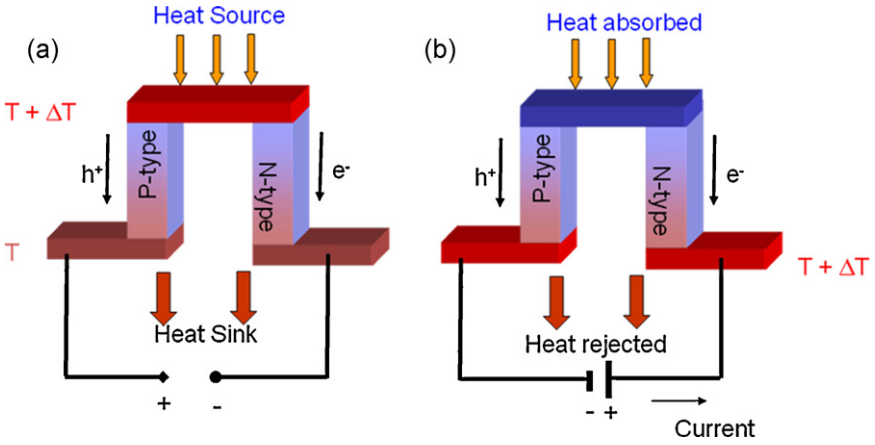

Fig. 2. Basic thermoelectric device for (a) heat-to-electrical energy conversion, and (b) heat pumping. The couple is comprised of p- and n-type semiconductors. In practice, many such elements are chained in series to yield desired power levels.

performance (COP) increases with a, temperature dependent, material's dimensionless thermoelectric figure of merit [5] defined as

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

where T is the average temperature of the hot and cold sides, and σ and κ are the electrical and thermal conductivity of the material. It is to be noted that κ is the direct sum of the contributions from both the carriers (κ_e) *and* the lattice (κ_L). From this, it is clear that a large $|S|$ and σ is desirable along a small κ. Indeed, a large majority of the understood increase in the figure of merit, ZT, of nanostructured thermoelectrics has been facilitated through a decreased κ_L. Common thermoelectric materials used today, e.g., bulk Bi_2Te_3, PbTe, SiGe etc., generally have a $ZT \sim 1$ which corresponds to ~10% efficiency of heat to energy conversion, with a heat source at ~525 K and a heat sink at room temperature, ~300 K. We will see how the advent of nanotechnology along with better materials fabrication tools have shown promise of greatly increasing the ZT.

An alternative to the use of thermoelectric materials, for cooling applications, utilizes thermionic effects that are typified in vacuum diode-based devices [6]. In a vacuum diode, the emission of electrons from the cathode to an anode (Fig. 3) can result in a concurrent cooling of the cathode. Such an idea was proposed by Ioffe [5] as a means to *reduce* the effects of carrier scattering (which reduces the mobility) and lattice thermal conduction (which detracts from heat transfer through carriers) in solid state materials. Ioffe surmised that an electrical energy generator could be obtained from a system comprised of "a number of plates kept at a high temperature T_1 and separated by vacuum from another set of plates maintained at a lower temperature T_2". For a temperature difference, $T_1 - T_2 > 200\,^{\circ}C$, the difference of the kinetic energies could be used for creating an electric potential difference. Such ideas can be translated to both *energy conversion* and *refrigeration*.

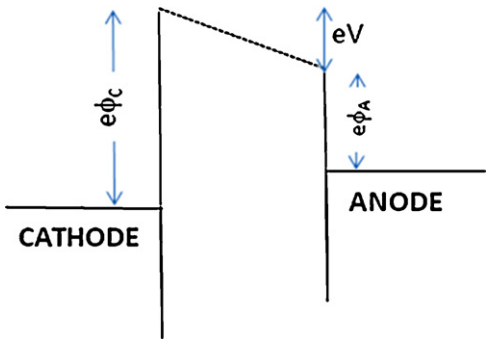

Fig. 3. The principles of thermionic refrigeration. The material at the left (cathode) is cooled through electron emission over a potential barrier ($e\Phi_C$) – constituted by the interface between the metal and vacuum/semiconductor material – into the anode. An applied potential (eV) compensates for the increased temperature of the anode.

While the originally proposed vacuum based thermionic devices had the advantages of low parasitic heat transport, major difficulties in fabrication and poor performance have precluded their large scale implementation. Consequently, solid state implementations where the vacuum has been replaced by a solid state barrier material are currently in vogue. It will be seen that when the barrier layer thickness is of the order of a mean free path, thermionic device efficiencies can also be understood in terms of the thermoelectric figure of merit $-ZT$.

1.2. Scope of the article

In this article, we aim to understand the unifying principles that are relevant to nanostructures in various forms, instead of a more exhaustive survey for which the reader could consult the CRC Handbook on Thermoelectrics [7]. Additionally, for a compilation of the variety of materials that have been suggested for possible use as thermoelectrics, the reviews by Wood [1] and Mahan [8] are excellent resources. In this review, we mainly consider the fundamental ideas along with *commonly accepted* metrics of efficiency for *thermoelectric* and the closely associated *thermionic* devices, and apply them to their nanostructured forms. For the metrics, we closely follow accepted convention in extant literature, e.g., most of the scientific work in thermoelectric seems to pertain to the methodology of increase of the figure of merit (ZT), while thermionics seems to be oriented towards devices, and more closely allied to the overall energy conversion efficiency.

It is then pertinent, in considering the possible improvements or novel features in nanostructured devices, to understand the basic phenomenology which has been described for three-dimensional bulk structures. This includes a comprehension of fundamental aspects such as carrier band structure and the relevant derivatives such as the Seebeck coefficient, the electrical conductivity, and the electronic contribution to the thermal conductivity. We then adapt such basics to lower dimensional structures, such as thin films/superlattices and nanowires where important conclusions on the influence of the carrier concentration and its implications will be drawn.

In reviewing the improvements of the figure of merit (ZT) for nanostructured thermoelectrics, it is seen that a large majority of the understood increase is through a decreased contribution of the

lattice thermal conductivity, κ_L. A decreased κ_L improves the thermoelectric performance by minimizing parasitic heat transport along the device and consequently permitting more efficient conversion of heat to electricity, predominantly through electronic carrier mediated transport. On the other hand, the electronic contribution to the thermal conductivity, κ_e is considered to be proportional to the electrical conductivity, through the Wiedemann–Franz law and hence immutable (exceptions due to temperatures [9], and nanostructuring [10] are noted in Section 2.3) Consequently, it will be seen how κ_L has been regarded as amenable to be tailored at different length scales, ranging from atomic placement in crystal structures to interfaces and surfaces. This part of the review will consider and interpret some of the fundamental principles involving the reduction of κ_L. Additionally, we will seek to understand recent experimental manifestations, e.g., through nanostructured bulk (“nano-bulk”) alloys [11], phonon blocking/electron – transmitting thin film superlattices [12], and sub-100 nm nanowires [13].

Subsequently, we consider theoretical and experimental details concerning carrier emission and transport under the topic of thermionic devices. The article will conclude with examining the prospects and an evaluation of the promise of nanostructured thermoelectrics.

1.3. The promise of nanostructures

While the thermoelectric effects were first discovered in metals, the larger magnitude of the effects in semiconductors facilitated their practical application and use since the 1950s. Initially, there were not many ways, e.g., in addition to doping, that were available to manipulate the intrinsic *ZT* of a given semiconductor material composition. From a practical perspective, the use of nanostructures provides a method for tuning the *ZT* through new methods such as quantum confinement, modulation doping, and the increased influence of interfaces and surfaces peculiar to such scales and dimensions. It has been found recently that dimensional restriction can lead to a much enhanced efficiency over traditionally used bulk thermoelectrics. Experiments, performed over the last decade, on these structures have shown a steadily increasing *ZT*.

Correspondingly, while the *ZT* of commercial, bulk, thermoelectrics such as Bi_2Te_3, PbTe and SiGe alloys hovered ~1 for a long time, theoretical predictions [14] in the mid-90s coupled with exciting theoretical and experimental results in compound semiconductor quantum wells [15] and superlattice [12] structures, Si nanowires [4,13], quantum dot [16] and other nanostructure [11,17] incorporated thermoelectrics have indicated the promise of a higher *ZT* and energy conversion efficiency. A better appreciation of the influences of the nanoscale, as manifested in nanostructures will be obtained through a re-examination of the fundamental aspects.

2. Electrical transport in thermoelectric materials

In this section, the electron transport coefficients relevant to thermoelectric material, namely the diffusive Seebeck coefficient (S), the electrical conductivity (σ), and the electronic thermal conductivity (κ_e), are explained through expressions derived from the Boltzmann transport equation (BTE). The discussion will entail the physical origin of each transport coefficient, their evolution with respect to temperature, carrier concentration, and change in device structure (e.g., from bulk to quantum well), and their interrelationship. With this understanding, the optimal conditions under which the power factor ($S^2\sigma$) is maximized in a material is found, and various methods for further enhancement of the power factor via electron confinement effects are elaborated. Relevant experiments in quantum well superlattices and nanowires are then reviewed.

2.1. Boltzmann transport equation

The electron transport coefficients can be derived from the solution to the BTE. The formal expression for the BTE [18] is

$$\left(\frac{df}{dt}\right)_{sc} = \frac{df}{dt} + \frac{d\mathbf{k}}{dt}\nabla_{\mathbf{k}} f + \frac{d\mathbf{r}}{dt}\nabla_{\mathbf{r}} f, \tag{2}$$

where t is time, $\mathbf{k}$ and $\mathbf{r}$ are the wave and position vectors of electrons, and f is the non-equilibrium distribution function. The BTE describes the evolution of particles in a system under a perturbation which redistributes the particles' momentum and position. For electrons, such perturbations can arise from the force exerted by an electric field (ε) or the temperature gradient (say, along the x-direction) through $(d\mathbf{k}/dt) = (-e\varepsilon/\hbar)$, where $-e$ is the unit charge of electrons, and $\nabla_{\mathbf{r}} f = (\partial f/\partial T)(dT/dx)$, respectively. We wish to solve the BTE for f, which often requires the *relaxation time* ansatz, where electrons are said to equilibrate through effect of random scattering (hence the subscript “*sc*” for the left hand term) within a relaxation time (τ), i.e., $(df/dt)_{sc} = -(f - f_0)/\tau$. The solution is then stated in term of the equilibrium distribution function (f_0).

At equilibrium, electron distribution follows the Fermi–Dirac statistics through:

$$f_0(E) = \frac{1}{\exp((E - E_F)/k_B T) + 1}, \tag{3}$$

where E is the electron's energy level, E_F is the “Fermi energy”, and k_B is the Boltzmann constant. To be precise, the Fermi–Dirac function should be described in term of the chemical potential [18], but we shall insist on using the Fermi energy, which may also be called the Fermi potential or the Fermi level. If the electron concentration is low, its equilibrium distribution may also be approximated by the Maxwell–Boltzmann function, which leads to a much simpler mathematical treatment [19]. However, obtaining accurate results for thermoelectric materials inevitably requires the use of Fermi–Dirac statistics. Then, the first-order, steady state ($(df/dt) = 0$) solution to the BTE may now be written as

$$f(E) - f_0(E) = -\tau(E)v(E)\frac{df_0}{dE}\left[\mp e\varepsilon - \left(\frac{E - E_F}{T}\right)\frac{dT}{dx}\right] \tag{4}$$

Subsequently, Eq. (4) is used to determine the magnitude of charge and heat current density (J and Q, respectively)

$$J \equiv \mp nev = \mp e\int_{-\infty}^{+\infty} g(E)v(E)[f(E) - f_0(E)]dE \tag{5}$$

$$Q \equiv n(E - E_F)v = \int_{-\infty}^{+\infty} g(E)(E - E_F)v(E)[f(E) - f_0(E)]dE, \tag{6}$$

where n is the carrier concentration, v is the carrier velocity, and $g(E)$ is the material's density of states (DOS). Note that while an electron always carries the same amount of charge, e, it may carry any amount of thermal energy, which is given by $(E - E_F)$. The transport coefficients, S, σ, and κ_e, are then defined in term of J and Q (discussed in Section 2.3).

The characteristic feature of the Fermi–Dirac distribution function, as illustrated in Fig. 4, is that the probability of finding an electron with E far below/above the E_F is essentially 100% and 0%, respectively. More importantly, while the change in distribution (df_0) over an incremental change in energy (dE) for electrons near E_F is quite high, it is virtually zero for those electrons with energy far greater or lesser than E_F, i.e., $(df_0/dE) = 0$ for $|E| \gg E_F$. It is then obvious from Eqs. (4)–(6) that only those electrons close to Fermi level are sensitive to external disturbances and contribute to the electrical conduction.

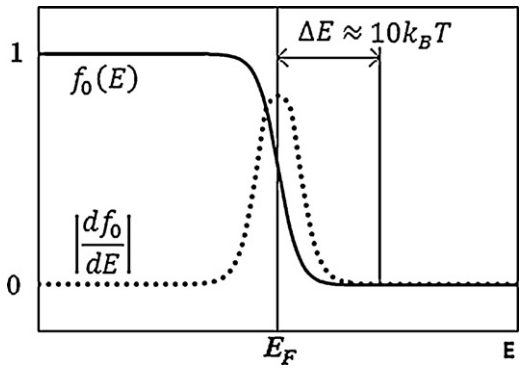

Fig. 4. The Fermi–Dirac distribution function ($f_0(E)$, solid line) and its first derivative (df_0/dE, dotted line). The implication of this diagram is that only electrons with energy close to E_F contribute to electrical/thermal transport.

The semi-classical BTE approach is widely used in the field of thermoelectrics, and continues to provide satisfactory agreement with experiment, even in nanostructured materials that exhibit quantum mechanical effects. The quantum effects such as electron confinement and tunneling are accounted mainly through the modification of the DOS. More modern approaches, not based on the BTE, e.g., using non-equilibrium Green's function (NEGF) methods [20,21] which seek the direct (instead of statistical) solution of the Schrodinger equation have been studied in recent years, but do not seem to show major deviations from the predictions of the simpler BTE-based methods.

2.2. Electronic band structure

Several important entities that determine *J* and *Q*, such as the density of states and carrier velocity, are determined by the material's electronic band structure. Conversely, it is the unique characteristic of each material's electronic band structure which determines the transport coefficients for that material.

The electronic band structure for any material in its entirety is very complex. Fortunately, only portions of the band structure, such as that for Si shown in Fig. 5, close to the E_F will significantly influence the thermoelectric properties. For example, the thermoelectric properties of p-type Si are determined by the characteristics of the local band structure of the valence band (VB) maximum at the Γ-point, while those of n-type Si depend instead on the conduction band (CB) minimum which is located about 80% of the way from the Γ- to the X-points. These electrically active portions of the band that are close to the E_F, i.e., the Γ-point VB maximum in p-type Si and/or the CB minimum in n-type Si, are frequently referred to as *conduction valleys* or *carrier pockets*.

The location in **k**-space, in conjunction with the material's crystal structure, then determines the degeneracy factor (N) of the conduction valleys. For example, since Si has an *fcc* crystal structure with 6-fold symmetry the CB minimum also has a 6-fold degeneracy ($N = 6$), as shown in Fig. 5(b). On the other hand, since the Γ-point is the center of the Brillouin zone, the VB maximum is singly degenerate ($N = 1$). Common thermoelectric materials, such as Bi_2Te_3, SiGe, and PbTe have such "multivalley" band structure (i.e., $N > 1$).

The transport coefficients may be determined for *each* conduction valley. In particular, each valley's transport coefficient is distinguished by the curvature of the band, which is inversely proportional to the effective mass, i.e., $m^* = \hbar^2(d^2E/dk^2)^{-1}$. For parabolic band, the curvature and the effective mass will be a constant. On the contrary, the curvature in non-parabolic bands is not constant, and the effective mass changes with energy and is more complicated to define. Furthermore, there will be multiple distinct effective masses for anisotropic band. The anisotropy of effective masses consequently leads to anisotropic transport properties.

Despite the fact that degenerate valleys have the same band structure, they may yet have different orientations with respect to the direction of electrical conduction and electron confinement (in case of nanostructures). This asymmetry then leads to the breaking of degeneracy. In the first case, the asymmetry is easily eliminated by appropriate averaging of the effective masses in the relevant direction (see "conductivity effective mass" in Section 2.2.2). However, in the latter case, such simple method may be insufficient, and it is more clarifying to treat these valleys as distinct and possessing distinct transport coefficients.

The transport coefficients in a *band* can then be summed from those of the constituent *conduction valleys* by assuming a parallel conductor model. The model simply treats each conduction valley as non-interacting, and yields the following summation rules

$$S = \frac{\Sigma_{i=1}^{N} S_i \sigma_i}{\Sigma_{i=1}^{N} \sigma_i} \tag{7}$$

$$\sigma = \Sigma_{i=1}^{N} \sigma_i \tag{8}$$

$$\kappa_e = \Sigma_{i=1}^{N} \kappa_{e,i}, \tag{9}$$

where the transport coefficient of a valley is denoted by the i^{th} term. Likewise, the transport coefficient of the *material* can then be summed from those of the *band* in the exact same manner. However, as the electrical conductivity of the bands far from the

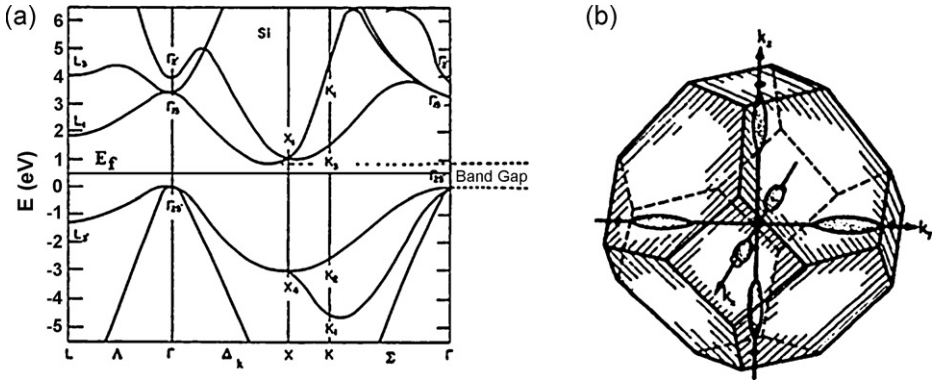

Fig. 5. (a) Si electronic band structure (reprint permission from [164]). (b) The constant energy surfaces at the six Δ-point valleys in the first Brillouin zone of Si (reprint permission from [266]).

Fermi level will be many orders of magnitude smaller than that of the band closest to the Fermi level, it is unnecessary to invoke such rules for effectively single-band materials such as a metal or degenerate semiconductor. On the contrary, the properties of semimetal and intrinsic semiconductor, where conduction valleys in both VB and CB are influential, can only be understood through the use of Eqs. (7)–(9). For example, the tendency for small $|S|$ in semimetals is due to the opposite sign of S for conduction valleys in VB and CB, which subtract from each other according to Eq. (7).

Also, one should note that while the size and alignment of the band gap (i.e., direct or indirect) does not directly affect the conduction valleys' transport coefficient, it can play some part in the overall behavior of a material. For example, materials with small, direct, band gaps such as Bi_2Te_3 and PbTe, tend to have a strong VB–CB coupling, which results in non-parabolic conduction and valence bands with small effective masses. On the other hand, indirect band gap material, such as SiGe, may have parabolic and broader conduction band with larger effective masses. However, the valence bands still tend to be non-parabolic due to heavy- and light-hole band interaction. Furthermore, the size of the band gap will determine the temperature limit where the doped semiconductor becomes intrinsic again due to thermal excitation of minority carriers.

Ultimately, the effective masses (m^*) and the degeneracy factor (N) are the two important parameters derived from the electronic band structure and they both determine the density of states and carrier velocity. These are then used in conjunction with the relaxation time to determines the charge and heat current density (Eqs. (4)–(6)), and, subsequently, the transport coefficients.

2.2.1. Density of states

The *local* band structure near the maxima or minima can often be simply represented by a parabolic dispersion function, which defines the energy–momentum (E–$\mathbf{k}$) relationship of electrons in that band. In bulk material (three dimensions: 3D, represented by x, y, and z), the energy dispersion function at the Brillouin zone center is represented by:

$$E_{3D}(\boldsymbol{k}) = \frac{\hbar^2}{2}\left(\frac{k_x^2}{m_x} + \frac{k_y^2}{m_y} + \frac{k_z^2}{m_z}\right) \tag{10}$$

where ħ is the reduced Planck constant, $k_{(x,y,z)}$ denotes the wave number and $m_{(x,y,z)}$ the principal effective masses in x-, y-, and z-directions, respectively.

A dispersion function of a similar form can also be used for holes in the valence band, provided that a consistent energy scale is used. The energy scale for holes is inverted with respect to that of electrons, as depicted in Fig. 6. For consistency, the E_F must always be measured with respect to the ground state defined in each case; we prefer to set the ground state energy level for both the CB and the VB to zero, i.e., $E_{3D}(k = 0) = 0$.

In quantum wells, this dispersion is modified due to the discretization of carrier momentum in the quantum confined directions. For example, if the confinement is in the z-direction, the electron dispersion function for the n^{th} sub-band can be written as

$$E_{2D}(\boldsymbol{k}) = \frac{\hbar^2}{2}\left(\frac{k_x^2}{m_x} + \frac{k_y^2}{m_y}\right) + E_n, E_n = \frac{\hbar^2\pi^2 n^2}{2a^2 m_z}, \quad n = 1, 2, 3\ldots \tag{11}$$

where a is the quantum well thickness and E_n is the confinement energy. Note that the ground state energy in this band has now increased by E_1 *relative* to the same band in a bulk material. Likewise, the dispersion function for nanowires is modified by the carrier confinement in two directions (i.e., y and z) such that, for

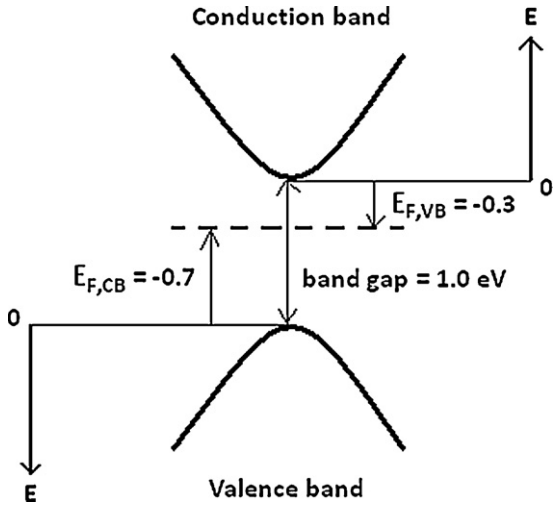

Fig. 6. The energy dispersion and scale as defined for electrons in the upper conduction band (CB) and holes in the lower valence band (VB). The E_F is measured from the minima of the CB and the maxima of the VB. As an example, the E_F is equal to −0.3 and −0.7 eV, when considering carriers in the CB/VB, respectively.

the nm^{th} sub-band, it is given by

$$E_{1D}(\boldsymbol{k}) = \frac{\hbar^2}{2}\left(\frac{k_x^2}{m_x}\right) + E_{nm}, E_{nm} = \frac{\hbar^2\pi^2}{2a^2}\left(\frac{n^2}{m_y} + \frac{m^2}{m_z}\right), \quad n, m = 1, 2, 3\ldots \tag{12}$$

where a is the width of the wire (assumed to have a square cross-section).

In addition to the physical length scales (a), the confinement potential energies for sub-bands in both quantum wells (E_n) and nanowires (E_{nm}) are dependent on the principal effective masses in the directions of confinement. As such, we distinguish the effective masses in this role as the "*confinement effective mass*" (m_c), i.e., $m_c = m_z$ for quantum wells, and $m_{c1} = m_y$ and $m_{c2} = m_z$ for nanowires.

The density of states (DOS) per *unit volume* for a single band/sub-band in three-, two- and one-dimensional systems, including a spin-degeneracy factor of two, are then derived from the dispersion functions in Eqs. (10)–(12) as [22]

$$g_{3D}(E) = \frac{1}{2\pi^2}\left(\frac{2m_d}{\hbar^2}\right)^{3/2} E^{1/2}, \quad E \geq 0, \ m_d = \sqrt[3]{m_x m_y m_z} \tag{13}$$

$$g_{2D,n}(E) = \frac{m_d}{a\pi\hbar^2}, \quad E \geq E_n, \ m_d = \sqrt[2]{m_x m_y} \tag{14}$$

$$g_{1D,nm}(E) = \frac{1}{a^2\pi}\left(\frac{2m_d}{\hbar^2}\right)^{1/2} (E - E_{nm})^{-1/2}, \quad E \geq E_{n,m}, \ m_d = m_x \tag{15}$$

The total DOS is then the aggregate of all individual sub-bands, i.e., $g(E) = \sum_{n=1}^{+\infty} g_n(E)$. Taking into account multiple sub-bands, the staggered pattern for the DOS can be seen in Fig. 7.

However, as it is usually sufficient to consider only the lowest sub-band ($n = 1$), which is closest to the E_F, we often allude to only the lowest/single sub-band. In such case, the total DOS for N number of degenerate conduction valleys is described solely by the power law function:

$$g(E) = \frac{N}{g_D a^{3-D}}\left(\frac{2m_d}{\hbar^2}\right)^{D/2} (E - E_0)^{D/2-1}, \quad E \geq E_0, \ g_D = \begin{Bmatrix} 2\pi^2, & D = 3 \\ D\pi, & D = 2, 1 \end{Bmatrix} \tag{16}$$

where E_0 is the ground state energy level, and the exponent $D = 3$, 2, 1 is termed the "dimensionality" factor for bulk material,

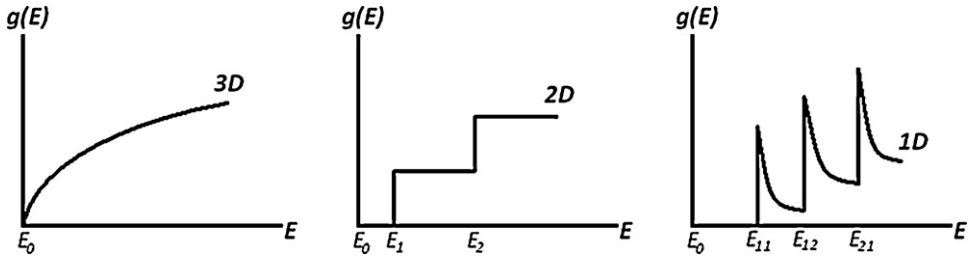

Fig. 7. Density of states, $g(E)$, for 3D, 2D, and 1D electrons gas in bulk material, quantum well, and quantum wire, respectively. The staggered pattern is due to the superposition of multiple sub-bands, each of which is shifted up from the bulk ground state energy (E_0) by the confinement energy (E_1, E_2, etc.).

quantum well, and nanowire, respectively. In the above expressions, we have also introduced the *density of states effective mass* (m_d), which is the geometric mean of the principal effective masses, i.e., $m_d = \sqrt[D]{\Pi_{i=1}^{3-D} m_i}$. Note that in a lower dimensional system, m_d does not include the effective mass in the direction of confinement. Since the conduction valleys can have different orientations with respect to the confinement, they can have both different m_d and confinement potentials/energies.

Due to the critical importance of the DOS and electron confinement in determining transport parameters, we state the following aspects pertinent to nanostructures:

(1) *Change in the energy dependence* of the DOS (i.e., $D/2 - 1$) from +1/2 (bulk) to 0 (quantum well) and −1/2 (quantum wire).
(2) An inverse proportionality of the DOS to the quantum well/wire thickness, where the DOS can be increased by decreasing a.
(3) *Increase in the effective band gap* due to upward (/downward) shift of the CB (/VB) ground states, corresponding to the confinement potential energy which is inversely proportional to m_c and a.

Such changes, all evident at reduced dimensionality, can be exploited for the enhancement of power factor in nanostructured thermoelectrics, as will be described in Section 2.4.

2.2.2. Carrier velocity and relaxation time

In the low electric field approximation, assuming diffusive carrier motion along the direction *perpendicular* to the direction of confinement, the carrier velocity is given by

$$v^2(E) = \frac{2(E - E_0)}{Dm_\sigma} \tag{17}$$

where m_σ is the *conductivity/inertial effective mass*, defined through the *harmonic average* of the effective masses of degenerate conduction valleys in the direction of conduction, i.e., $1/m_\sigma = \sum_{i=1}^{N} 1/m_{\sigma,i}$. For example, the conductivity effective mass for the Δ-point valleys in n-type Si for conduction along ⟨1 0 0⟩ is $m_\sigma = 3(2/m_t + 1/m_t)^{-1}$, where m_t and m_l are the transverse and longitudinal effective mass, respectively. The use of the conductivity effective mass allows us to eliminate the asymmetry in each valley's orientation with respect to the direction of conduction. The dimensionality factor, D, in Eq. (17) accounts for the equi-partition of energy among the available degrees of freedom.

The type and magnitude of carrier scattering within the material determines the relaxation time. The carrier relaxation time, $\tau(E)$, is difficult to determine from first principles as it involves simultaneous carrier interaction with various scatterers such as acoustic and optical phonons, plasmons, defects, interfaces, and even other carriers. In most cases, the relaxation time is modeled through the power law function:

$$\tau(E) = \tau_0 (E - E_0)^r \tag{18}$$

where the exponent r is the characteristic "scattering constant" (not to be confused with the position vector $\mathbf{r}$) peculiar to a particular scattering process, and τ_0 is a constant that depends on both scattering processes and material properties. The value of r can be determined experimentally through (i) mobility–temperature measurements, (ii) the variation of the Seebeck coefficient with carrier concentration, and (iii) the measurement of the Nernst coefficient, which can also determine the sign of r [23].

In some cases, r can be analytically determined [24] (e.g., for scattering by weakly screened ionized impurities, $r = 3/2$, and for neutral impurity scattering, $r = 0$). However, for lattice deformation potential scattering by acoustic and optical phonons (ADP and ODP, respectively) or strongly screened ionized impurity scattering, r is variable, as τ is inversely proportional to the DOS. r may also be affected by multiple scattering processes. Examples of the relation between the scattering processes and r, as a function of dimensionality can be seen in Table 1, in Section 2.4.1.

In some cases, more than one scattering process may be dominant, which complicates the calculation of the relaxation time. The Matthiessen's rule [18] is often invoked in this case, which is equivalent to treating scattering processes as being independent of one another. This then implies that the total relaxation time can be summed from that of each scattering processes as $(1/\tau) = (1/\tau_1) + (1/\tau_2) \ldots$. However, this rule will give inaccurate results if these processes have disparate values of r [24].

2.3. Formulation and analysis of the transport coefficients: σ, S, and κ_e

From Eqs. (4)–(6) for the charge and heat current density, and the standard definitions, we arrive at the following expressions for the transport coefficients for conduction perpendicular to directions of confinement

$$S \equiv \left.\frac{\varepsilon}{\frac{dT}{dx}}\right|_{J=0} = \mp \frac{1}{eT}\left[\frac{X_1}{X_0} - E_F\right] \tag{19}$$

$$\sigma \equiv \left.\frac{J}{\varepsilon}\right|_{dT/dx=0} = e^2 X_0 \tag{20}$$

$$\kappa_e \equiv -\left.\frac{Q}{\frac{dT}{dx}}\right|_{J=0} = \frac{1}{T}\left[X_2 - \frac{(X_1)^2}{X_0}\right] \tag{21}$$

where $X_i = -\int_{E_0}^{+\infty} g(E)\tau(E)v^2(E)E^i(d f_0/dE)dE$. Now, incorporating Eqs. (16)–(18), which assume a single band/sub-band with parabolic energy dispersion, and a power law relaxation time,

the integral X_i can be explicitly written as

$$X_{i,D} = \frac{N\tau_0}{m_\sigma g_D a^{3-D}} \left(\frac{2m_d}{\hbar^2}\right)^{D/2} (k_B T)^{r+D/2+i} \left(\frac{2(r+i)}{D} + 1\right) F_{r+D/2+i-1}(\eta) \tag{22}$$

where $F_j(\eta) = \int_0^{+\infty} x^j/(\exp(x-\eta)+1)dx$ is the jth order Fermi integral and $\eta = (E_F - E_0)/k_BT$ the *reduced Fermi level*. The constant g_D has the value of $2\pi^2$, 2π, and π for D = 3, 2, and 1, respectively.

The reduced Fermi level is a convenient variable since it can represent both changes in the material's temperature and the Fermi level, and also allows X_n to be expressed in terms of Fermi integrals. With the exception of the zeroth-order integral, where $F_0(\eta) = \ln(1+e^\eta)$, $F_j(\eta)$ can only be evaluated numerically. We have extended the published algorithms [25] for half-integer orders of j, in the range of $\eta = [-5, 20]$ which proved to be more than sufficient for values of η relevant for thermoelectric materials. While Eqs. (19)–(22) are sufficient for the evaluation of the transport coefficients, in the next few sections we provide an overview of their interpretation, relevance, and inter-relationship through η.

2.3.1. Electrical conductivity (σ)

The electrical conductivity is a measure of how many (n) carriers (electrons/holes) can contribute to the electrical current along with their mobility (μ) and is given by

$$\sigma = ne\mu \tag{23}$$

Both n and μ can vary as a function of temperature, T, and can be measured experimentally through resistivity and Hall coefficient determination. The magnitude of the power factor is very sensitive to n and finding the suitable concentration is crucial for thermoelectric optimization. As will be discussed later, the relation between n, E_F, and the DOS solely determines the enhancement of power factor in nanostructures.

Assuming parabolic energy dispersion with power law relaxation time, the equilibrium carrier concentration, n, in a set of degenerate conduction valleys is written as

$$n \equiv \int_{E_0}^{+\infty} g(E) f_0(E) dE = \frac{N}{g_D a^{3-D}} \left(\frac{2k_B T m_d}{\hbar^2}\right)^{D/2} F_{D/2-1}(\eta) \tag{24}$$

Occasionally, the degeneracy factor (N) is combined with the density of states effective mass, m_d, to yield the *density of states effective mass for all valleys* as $m_{d,total} \equiv N^{2/D} m_d$. However, we prefer to list N and m_d explicitly. As n is proportional to the material parameters and T, we extract these factors out to plot the normalized carrier concentration, $\tilde{n} = n/\theta$, in Fig. 8(a). The normalization factor is given by

$$\theta = \frac{N}{(a')^{3-D}} \left(\frac{m_d}{m_o} \frac{T}{300\,K}\right)^{D/2} \tag{25}$$

where a' is equal to the thickness of the quantum well/wire *in nm*; $\theta = 1$ for $m_d = m_0$, $N = 1$, $a = 1$ nm, and $T = 300$ K. Due to the normalization. Fig. 8(a) is essentially applicable to *any material and temperature*, provided basic assumptions (i.e., parabolic band, etc.) are satisfied, and can be used to determine n at a given E_F.

Conversely, the E_F is also a function of n and T, and may be variable or a constant depending on the material. For example, since n may not increase significantly with T in extrinsic semiconductors, E_F must decrease with rising T to satisfy Eq. (24). On the contrary, in an intrinsic semiconductor, E_F is in the middle of the band gap and remains relatively constant. Therefore, increasing T would increase n in this case.

Changes in DOS can also affect this relationship. For instance, since a 2 nm quantum well would have larger DOS than a 5 nm quantum well by virtue of having smaller a, (which appears in the denominator of $g(E)$), the former would have smaller E_F compared to the latter, given the same n and T. Likewise, given two different

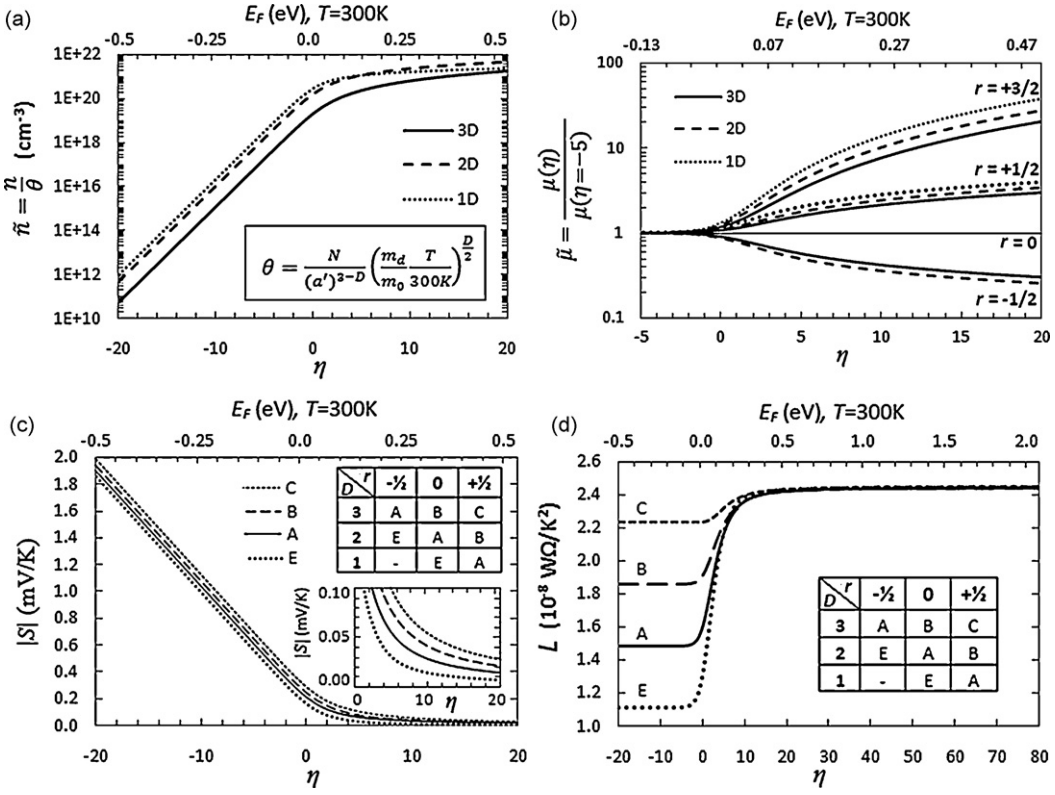

Fig. 8. (a) Normalized carrier concentration, (b) normalized mobility, (c) $|S|$, and (d) L as a function of η. Equivalent value of E_F at 300 K is labeled on the top horizontal axis. n was normalized to the constant θ as defined in the inserted box, where T is temperature in Kelvin, and a' is the width of quantum well or wire in nm. $|S|$ and L are degenerate for several combinations of the r and D, as listed in the inserted table. All plots assume a single sub-band with parabolic energy dispersion, and a power law variation of the relaxation time.

materials with the same n and T, the one with larger m_d and N, and hence larger DOS, would have smaller E_F.

On the other hand, in many cases it is very difficult to model how the mobility, μ, in nanostructured materials differs from bulk material. From the previous expressions for σ and n, the following expression was deduced

$$\mu = \left(\frac{e}{m_\sigma}\right)\frac{\int_{E_0}^{+\infty} g(E)\tau(E)v^2(E)(df_0/dE)dE}{\int_{E_0}^{+\infty} g(E)v^2(E)(df_0/dE)dE} = \frac{e\langle\langle\tau\rangle\rangle}{m_\sigma} \tag{26}$$

where $\langle\langle\tau\rangle\rangle$ denotes an average of the relaxation time involving the *non-equilibrium* terms of the distribution (df_0/dE) as also discussed by Lundstrom [24]. Since $v^2(E) \propto E$, we can use integration-by-parts to simplify $\langle\langle\tau\rangle\rangle$ to

$$\langle\tau\rangle = \frac{\int_{E_0}^{+\infty} g(E)\tau(E)f_0(E)dE}{\int_{E_0}^{+\infty} g(E)f_0(E)dE} = \frac{\langle\langle\tau\rangle\rangle}{((2r/D)+1)}$$
$$= \tau_0(k_BT)^r\frac{F_{r+D/2-1}(\eta)}{F_{D/2-1}(\eta)} \tag{27}$$

where $\langle\tau\rangle$ is now the average over the *equilibrium* carrier distribution.

While it is a simple matter to calculate how μ may change with n through their mutual dependence on η, it is more difficult to predict which scattering processes will become dominant as the material's condition changes. Furthermore, τ_0 is product of many parameters [24], the study of which is beyond the scope of this paper. Subsequently, we plot the normalized mobility, $\tilde{\mu}(\eta) = \mu(\eta)/\mu(\eta = -5)$, in Fig. 8(b) to simply demonstrate the ideal relationship between μ and η. The figure shows that μ increases with η if $r > 0$, and vice versa. If $r = 0$, then both $\langle\langle\tau\rangle\rangle = \langle\tau\rangle = \tau_0$ and $\mu = e\tau_0/m_\sigma$ are constants.

Since μ may change either way with increases in η, σ may also be variable. For example, σ in metal tends to decrease with increasing η because the reduction in μ due to acoustic phonon scattering ($r = -1/2$) dominates the smaller increase in n (when η is large). On the contrary, σ in non-degenerate semiconductors tends to increase with increasing η due to a large increase in n (when η is small).

2.3.2. Seebeck coefficient (S)

Compared to electrical conductivity, the Seebeck coefficient is not as well understood. Often, the Seebeck coefficient is described incompletely, and non-quantitatively as being dependent on the symmetry/asymmetry of the electronic band [26] or the energy dependence of the DOS in the range of the Fermi level [27], or that "high and steep" DOS is required for large Seebeck [28]. In this section, we hope to give the reader a complete and intuitive understanding of the Seebeck coefficient adhering strictly to its physical and mathematical definitions.

Firstly, there is more than one distinct physical process that gives rise to the net Seebeck effect related voltage (also see Section 3.5). The most prominent and common process arises from the diffusion of electrons (also see Section 3.5). Hence, we focus exclusively on deriving the diffusive Seebeck coefficient, which has been described in its most general form in Eq. (19). Assuming that $g(E)$ and $\tau(E)$ are power law functions, the expression for *S of a single band* is simplified to

$$S = \mp\frac{1}{eT}\left(\frac{\int_{E_0}^{+\infty} g(E)\tau(E)Ev^2(E)\frac{df_0}{dE}dE}{\int_{E_0}^{+\infty} g(E)\tau(E)v^2(E)\frac{df_0}{dE}dE} - E_F\right) = \mp\frac{1}{eT}\left(\frac{\langle\tau E\rangle}{\langle\tau\rangle} - E_F\right) \tag{28}$$

where $\langle x\rangle = \int_{E_0}^{+\infty} g(E)x(E)f_0(E)dE / \int_{E_0}^{+\infty} g(E)f_0(E)dE$ is the average value of an arbitrary function $x(E)$ over the equilibrium carrier distribution. If $g(E)$ or $\tau(E)$ are not power law functions, as is the case for non-parabolic bands, then it is necessary to revert back to the more complicated average over *non-equilibrium* carrier distribution seen in Section 2.3.1. Nonetheless, we can persist with the following discussion.

Eq. (28) clearly indicates that the S of a single band is proportional to the difference between the average energy of electrons weighted by the relaxation time and the E_F. In the case of constant relaxation time, $\langle\tau E</>\tau\rangle$ simplifies to $\langle E\rangle$. An intuitive narrative is easily observed: the diffusive Seebeck effect originates from the tendency of electrons in that band, which have an average energy, $\langle E\rangle$, to return to the lowest energy state represented by the Fermi level, E_F. The behavior of the Seebeck coefficient is then easily understood if one always considers how the average electron energy changes. Using the explicit expression in Eqs. (19) and (22), the diffusive Seebeck coefficient can be computed through the following equation

$$S = \mp\frac{k_B}{e}\left(\frac{(r+(D/2)+1)F_{r+D/2}(\eta)}{(r+(D/2))F_{r+D/2-1}(\eta)} - \eta\right) \tag{29}$$

The behavior of $|S|$ versus the reduced Fermi level, η, for different energy dependence of the DOS and relaxation time (D and r, respectively) is then plotted in Fig. 8(c).

Without exception, $|S|$ always decreases as η is increased, i.e., as E_F is increased or T is decreased. Since an increase in Fermi level would increase carrier concentration there would be an increase in average energy as well. However, as the average energy increases relatively slowly, $|S|$ always decreases as the Fermi level is increased. An increase in temperature can also increase carrier concentration and average energy, which means that $|S|$ will increase and vice versa. In an extrinsic semiconductor, the increase in temperature may not lead to significant increase in the carrier concentration, which is fixed to the dopant concentration, and does not increase the average energy. However, in this case, the Fermi level is actually reduced, thus leading to increased $|S|$ anyway. At sufficiently high temperature, the semiconductor may become intrinsic and exhibits a decrease in $|S|$ as temperature is increased. This is because both conduction and valence bands are now relevant, and the Seebeck coefficient of opposing sign in each band leads to a smaller net value, according to Eq. (7).

Assuming a non-changing Fermi level, the $|S|$ also *decreases* as the energy dependence of the DOS (as represented by the exponent D) is decreased. Often, the shape or profile of the DOS is thought of as having a significant influence on the Seebeck effect, i.e., it is commonly believed that a sharp peak in the DOS, such as those found in nanowires, yields a large Seebeck coefficient. This is not necessarily true. The influence of the DOS shape on the $|S|$ is easy to understand if we consider its effect on the average energy. In fact, the parabolic profile of the DOS in bulk material actually gives larger $|S|$ than the step/peak-like profile of the DOS in quantum well/nanowire, given the same Fermi level for all three cases. As shown in Fig. 7, the bulk DOS favors high energy states, while the nanowire DOS is skewed toward lower energy states, which means the average electron energy will be slightly larger in the bulk material than that of the nanowire, if both have the same Fermi level. Friedman had earlier on proposed the measurement of $|S|$, and the determination of the DOS energy dependence as a mean of verifying the confinement effect in two-dimensional superlattices [29].

As the average energy should be weighted by the relaxation time, according to Eq. (28), the increase in the scattering constant r will similarly increase $|S|$. Hence, $|S|$ would be larger if, say, weakly ionized impurity scattering ($r = +3/2$) rather than phonons scattering ($r = -1/2$) dominates. One way of manipulating the relaxation time was through the use of resonant scattering. According to Ravich et al., resonant scattering occurs when electrons in the allowed

energy band of the host material are trapped in the quasi-local states of impurity atoms before they are ejected after a finite period of time [30]. The total relaxation time near the resonant energy range can subsequently be affected, leading to a possible increase *or decrease* in the energy dependence of the total relaxation time and $|S|$. For example, Ravich et al state that if the resonant states are *below* the Fermi level (i.e., low energy states), then an increase in $|S|$ is likely. By the same token, a decrease in $|S|$ would be plausible if the resonant state is above the Fermi level (or at higher energy). In practice, the effect of resonant scattering could be detected by the measurement of r. If resonant scattering was responsible for the increase in $|S|$, then an increase in r should be detected. This was not the case in Tl-doped PbTe (see Section 2.6.2), where resonant scattering was suspected as the reason for enhanced $|S|$, but which was ultimately ruled out after measurement revealed no change in r [27].

Most crucially, $|S|$ is independent of effective masses, the valley degeneracy, quantum well/nanowire thickness, and τ_0, from Eq. (29). For example, we may consider a case where the Fermi level is a constant, and effective mass is increased. The increase in effective mass would increase the magnitude of the DOS, and carrier concentration. However, since the DOS, under the parabolic band assumption, is a power law function and is scale invariant with respect to effective mass (see Eq. (16)), there will be no change in the average energy and $|S|$. This feature crucially led to the proposal of the existence of an optimal Seebeck coefficient where the power factor is maximized (see Section 2.4.1).

Occasionally, approximate expressions for the Seebeck coefficients are used, and one should be aware of their limitations. From Fig. 8(c), three regimes can be defined according to the behavior of S. In the *non-degenerate* semiconductor regime ($\eta < -2$), $|S|$ is large, decreases linearly with η, and can be approximated using classical Maxwell–Boltzmann statistics [19] as

$$S_{nd-SC} \cong \mp \frac{k_B}{e}\left(r + \frac{D}{2} + 1 - \eta\right), \quad \eta < -2 \tag{30}$$

In the *metallic* regime ($\eta \gg 0$), a smaller $|S|$ is obtained, which is inversely proportional to η, and is approximated by the Mott formula [31]

$$S_{\text{metal}} \cong \mp \frac{\pi^2 k_B}{3e\eta}\left(\frac{\partial \ln(g(E))}{\partial \ln(E)} + \frac{\partial \ln(\tau(E))}{\partial \ln(E)} + \frac{\partial \ln(v^2(E))}{\partial \ln(E)}\right) = \mp \frac{\pi^2 k_B}{3e\eta}\left(r + \frac{D}{2}\right), \quad \eta > 30 \tag{31}$$

Note that the η limits of -2 and 30 are loosely defined, and should not be taken as absolute or fundamental. Since Eqs. (30) and (31) are derived from Eq. (29), they share the assumptions of parabolic band, power law relaxation time, etc. However, they cannot be used to accurately describe degenerate semiconductors. The typical value of $|S|$ ranges from around 1 mV/K for non-degenerate semiconductors to ~100 μV/K for degenerate semiconductors, and ~10 μV/K for metals.

It should also be noted that the expressions found in this section alone are not sufficient in describing the Seebeck coefficient of materials where multiple bands are relevant, such as in intrinsic semiconductors and semi-metals. In such materials, the Fermi level is sufficiently close to both the conduction and valence bands, giving rise to significant conductivity in both bands. It is necessary then to use Eq. (7) to compute the total S from the individual S of the conduction and valence bands, which can be computed separately through Eq. (29) in this section.

2.3.3. Electronic thermal conductivity (κ_e) and the Lorentz number (L)

The electronic thermal conductivity (κ_e) is very similar to electrical conductivity except it accounts for the thermal energy, instead of charge, carried by electrons. Understandably, κ_e also increases with n and μ, and it is possible to define the ratio of κ_e to σ, which is known as the Lorentz number (L), as

$$L = \frac{\kappa_e}{\sigma T} = \left(\frac{k_B}{e}\right)^2 \left[\frac{(r + \frac{D}{2} + 2)F_{r+D/2+1}(\eta)}{(r + \frac{D}{2})F_{r+D/2-1}(\eta)} - \left(\frac{(r + \frac{D}{2} + 1)F_{r+D/2}(\eta)}{(r + \frac{D}{2})F_{r+D/2-1}(\eta)}\right)^2\right] \tag{32}$$

In the metallic regime ($\eta \gg 0$), the Lorentz number approaches the value specified by the empirical Wiedemann–Franz law $L = \pi^2/3(k_B/e)^2 = 2.44 \times 10^{-8}(W\Omega/K^2)$, as shown in Fig. 8(d). L varies strongly with D, r, and η in the *degenerate* semiconductor regime ($\eta \approx 0$), and again approaches a constant in the *non-degenerate* semiconductor regime ($\eta < 0$). Like S, L is also independent of the material parameters.

The dependence of L on D and r in the *non-degenerate* and *degenerate* semiconductor regimes reflects the greater sensitivity of κ_e to the energy dependence of the DOS and $\tau(E)$, respectively. Physically, the sensitivity is due to the fact that *hot* electrons from higher energy states carry more thermal energy than *cold* electrons, while the σ is rather insensitive to the energy distribution of the carriers since all electrons, regardless of energy, carry the same amount of charge. For example, given the same n and σ, κ_e is smaller in quantum wire than in bulk material as a greater proportion of the electrons in the quantum wire belong to lower energy states and are "*cold*". Similarly, κ_e is reduced for $r < 0$ because the *hot* electrons would have a smaller $\tau(E)$, implying more frequent scattering compared to *cold* electrons.

2.4. Increasing power factor for higher Z

It can be shown that Z is approximately maximized when the power factor $S^2\sigma$ is maximized. For example, the tentative variations of $S^2\sigma$, κ_e, κ_L, and Z with η are shown in Fig. 9. The power factor exhibits a peak due to the opposing variation of S and σ, while κ_e increases monotonically as η is increased. Since κ_L is generally assumed to be unrelated to electrical properties, it can be treated as a constant independent of η.

In the metallic regime where the η is very large, Z tends to be small because of (1) inherently small $|S|$, and (2) large κ_e due to large σ. Since κ_e is typically much larger than κ_L in metal, we may state that, for metals, $\kappa \sim \kappa_e$ and $Z = S^2/LT$. Evidently, as both $|S|$ and $1/L$ increases as η is decreased (see Fig. 8), a higher Z is always obtained if η *is reduced*, i.e., if the material is made *less metallic*. Interestingly, a very large σ could be detrimental to Z, as κ_e will also then be very large.

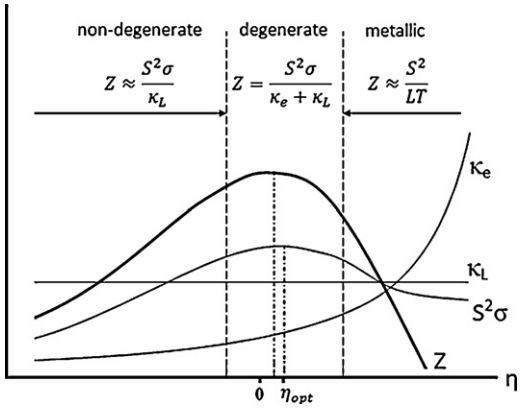

Fig. 9. In the non-degenerate semiconductor regime, κ_e is negligible, and Z increases with η due to increase in σ. In the degenerate semiconductor regime, Z reaches a peak when $\eta \approx \eta_{opt}$, where $S^2\sigma$ is maximized. In the metallic regime, κ_L is negligible, and Z declines with increasing η due to reduction in S, and increases in κ_e.

Table 1
Optimum reduced Fermi level (in eV, top line) and the corresponding magnitude of the Seebeck coefficient (in μV/K, bottom line) for maximum power factor, $S^2\sigma$. For $r = +3/2$ ($D = 3, 2, 1$) and $r = +1/2$ ($D = 3$), $S^2\sigma$ increases without limit as η is increased.

D	r = -½	0	+½	+3/2
3	0.67 167	2.47 130	no limit -	no limit -
2	-0.37 187	0.67 167	2.47 130	no limit -
1	-1.14 197	-0.37 187	0.67 167	no limit -
		Neutral Impurity	ADP ODP	Ionized Impurity (weakly-screened)
		Ionized Impurity (strongly-screened)		

At the other extreme, Z is again small for insulating material such as non-degenerate semiconductors due to very small σ, in spite of a large $|S|$. κ_e is also very small, such that $\kappa \sim \kappa_L$ and $Z = S^2\sigma/\kappa_L$. Consequently, Z would increase with $S^2\sigma$. We then show in the following section that the *optimal reduced Fermi potential*, η_{opt}, where the power factor is maximized, is *always* in the degenerate semiconductor regime. It is not surprising then that good thermoelectric materials such as Bi_2Te_3, PbTe, or SiGe are all heavily doped, degenerate, semiconductors.

2.4.1. Optimization of power factor

The $S^2\sigma$ varies drastically as the carrier concentration and η are changed. Since increasing η tends to reduce $|S|$ but increases σ, and vice versa, the product $S^2\sigma$ is expected to have a maximum at some optimal value of η, which could be determined through Eqs. (19) and (20), given the material parameters and temperature. However, having noted that S is independent of η, and σ can be normalized, we may write

$$S^2\sigma = [BT^{r+D/2}]S^2\tilde{\sigma}(\eta, r, D), \tag{33}$$

where $\tilde{\sigma}$ is the normalized conductivity, and $B = \tau_0 N(2m_d)^{D/2}/m_\sigma a^{3-D}$ is similar to the oft-quoted material figure of merit [32]. From this expression, we have recently shown that there exist an optimal reduced Fermi level and Seebeck coefficient that always gives the maximum power factor, and which is independent of the material's effective masses (m_d and m_σ) and degeneracy factor (N), quantum well/nanowire thickness (a), τ_0, and temperature [33].

Eq. (33) essentially indicates that the power factor is scale invariant with respect to the common factor $BT^{r+D/2}$, and therefore its maxima would be only a function of η, r, and D. Consequently, there exists a universal and optimal reduced Fermi level (η_{opt}), and a corresponding optimal Seebeck coefficient (S_{opt}) as determined from Eq. (29). The optimal carrier concentration also corresponds to η_{opt}, but its value varies according to the material's properties (m_d, N, a) and temperature. The values for η_{opt} and S_{opt} are listed in Table 1. Identical values were also independently reached by Kim et al. [34].

Note that when $r > 0$, i.e., when weakly screened ionized impurity scattering is dominant, there is a possibility of an unlimited increase in power factor due to the combined increase in n and μ being larger than the reduction in S^2, as η is increased. However, this is not observed in practice as strongly screened scattering (say, $r = -1/2$) dominates at high value of η and n. Since the scattering rates of carriers through the acoustic and optical phonon deformation potential (ADP and ODP, respectively) and strongly screened ionized impurities are proportional to the DOS [24], r is not a constant and changes with D in these cases. When multiple scattering processes are concurrent, r may take intermediate values to those listed in Table 1. Generally, as the Fermi integral changes gradually with incremental changes in its indices, the value of η_{opt} will also change gradually, and should follow the trend set forth in the table. Note that the values of η_{opt} indicate that the maximum $S^2\sigma$ always occurs in the *degenerate semiconductor* regime (where, $\eta \approx 0$).

Our calculations show that for any material, the maximum $S^2\sigma$ at any temperature is expected when the $|S|$ is in the optimal range of 130–187 μV/K. The maximum power factor was indeed observed at the predicted optimal value, $|S| \sim 0.167$ mV/K, for a wide variety of materials and temperature including bulk PbTe at 300 K [35], p-type $Si_{0.7}Ge_{0.3}$ at 300 K [36], $Sr_{1-x}La_xTiO_3$ at 300 K [37], and $SrNbTiO_3$ at 1000 K [38]. Consequently, the calculated optimal S could be used for thermoelectric material optimization. For example, a bulk material should be doped until the value of $|S| \sim 0.130$–0.167 mV/K is reached. As a corollary, an observation of $|S|$ much larger than the above could indicate a material that has not yet been optimized for maximum power factor, and *not* as a sign of a good thermoelectric material!

Though the $S^2\sigma$ is always maximized at η_{opt}, the magnitude of $S^2\sigma$ can freely increase according to material parameters and temperature. The generally accepted notion that good thermoelectric materials should have multi-valley band structure (large N) with large effective masses (m_d), and high mobility (large τ_0 and small m_σ) is indeed seen through Eq. (33). The $S^2\sigma$ can also be increased by increasing the operating temperature, T. However, this requires that E_F would also have to be increased in order to maintain the η_{opt}, which is a constant. Since E_F tends to *decrease* with increasing T, the optimal carrier concentration must increase with temperature.

The $S^2\sigma$ is eventually limited by the onset temperature of bipolar conduction. It was estimated that the maximum temperature beyond which Z is reduced is $T_{max} \sim E_g/10k_BT$, where E_g is the band gap [8]. As a result, the absolute maximum power factor of a material is expected at T_{max} and η_{opt}. In bulk material, the maximum power factor is then limited by the fixed, intrinsic properties of material (i.e., m_d and N) at η_{opt} and $T_{\max}$.

However, theoretical and experimental work has subsequently shown that quantum confinement in nanostructured materials could enhance $S^2\sigma$ even beyond this bulk value. The presence of the quantum well/nanowire thickness, a, in Eq. (33) through the B factor, in Eq. (33), hints at the possibility of increasing $S^2\sigma$ by reducing a. In general, requirement for large N, m_d, and small a are all equivalent to having a large magnitude of the DOS. Most methods to increase Z via electron confinement are aimed at increasing the DOS. We elaborate on such issues in the following sections.

2.4.2. Enhancement of the power factor beyond bulk value

The $S^2\sigma$ can be increased by tuning the Fermi level (through changing the carrier concentration) and increasing the temperature, but there is an inherent limit that is achieved at η_{opt} and $T_{\max}$. However, in 1993, Hicks and Dresselhaus proposed another way to increase $S^2\sigma$ through the use of confinement effects in nanostructures such as quantum wells [39] and nanowires [40]. For clarity, we first write $S^2\sigma = (S^2n)e\mu$. The enhancement of $S^2\sigma$ by confinement effect is due completely to the increase in the *magnitude* of the DOS, which then enhances S^2n. This concept is illustrated in the case of Si nanowires in the following.

First, we illustrate the increase in S^2n as a result of the increase in E_F in bulk material. Fig. 10(a) shows the carrier distribution in

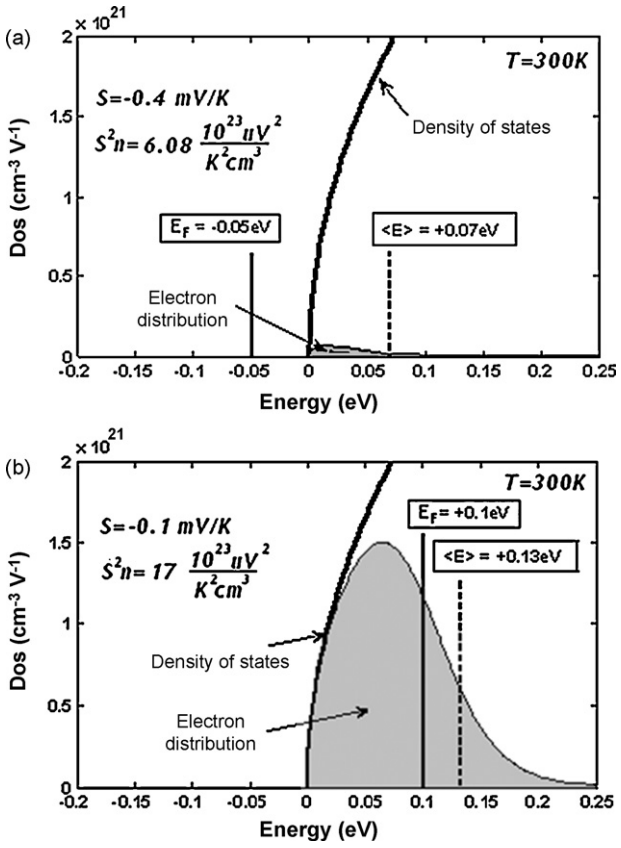

Fig. 10. Electrons density distribution (grey area) in the conduction band of (a) moderately doped ($E_F = -0.05$ eV, $n = 3.8 \times 10^{18}/\text{cm}^3$), and (b) highly doped ($E_F = +0.1$ eV, $n = 1.7 \times 10^{20}/\text{cm}^3$) bulk Silicon. S is proportional to $\langle E \rangle - E_F$. The DOS is plotted in bold line. Both plots are drawn to scale, and with the band's ground state energy at zero.

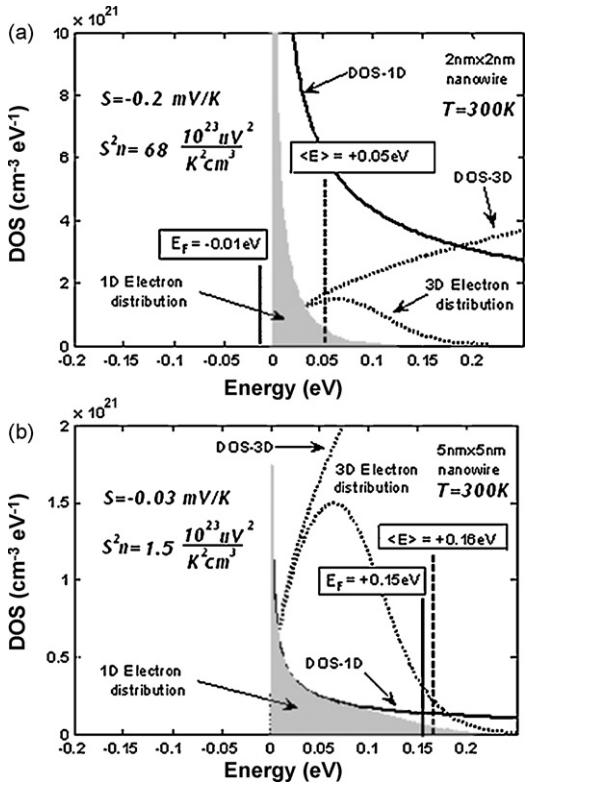

Fig. 11. Electrons density distribution (gray area) in the conduction band of (a) 2 nm, and (b) 5 nm Si nanowires. Dopant/carrier concentration in both cases is the same as that of bulk Si in Fig. 10(b), i.e., $n = 1.7 \times 10^{20}/\text{cm}^3$. Bulk DOS and electron density distribution is plotted (dotted line) for comparison alongside the DOS of the nanowires (solid line). Ground state energy of the band in quantum wire is also set to zero; therefore the relative increase in ground state energy due to confinement potential is not shown in this figure.

the conduction band of bulk Si ($m_d = 0.32m_0$, $N = 6$) with an arbitrary $E_F = -0.05$ eV, which equates to $n = 3.8 \times 10^{18}/\text{cm}^3$ and $\eta = (-0.05\ \text{eV}/8.6 \times 10^{-5}\ \text{eV}/k)(300k) = -1.93$ at 300 K. Assuming that $r = 0$, Eq. (29) indicates that $\langle E \rangle = +0.07$ eV, and $S = (1/e)/(0.07 + 0.05)\ \text{eV}/300k = -0.4$ mV/K. If the dopant concentration is increased such that $E_F = +0.1$ eV ($\eta = 3.86$), then a larger $n = 1.7 \times 10^{20}/\text{cm}^3$ is observed, as shown by larger distribution in Fig. 10(b). Since the corresponding increase in $\langle E \rangle$ (to +0.13 eV) is smaller that of the E_F increase, there is a four-fold *reduction* of $|S|$ to ~0.1 mV/K. Nevertheless, S^2n has increased by a factor of three from 6.1×10^{23} to $17 \times 10^{23}\ \mu\text{V}^2\ \text{K}^{-2}\ \text{cm}^{-3}$. Since we have assumed that μ is a constant ($r = 0$) with respect to η, the $S^2\sigma$ will increase by factor of three as well. As the η in both cases are smaller and greater than the predicted optimal value of 2.47, respectively, we may also infer that the optimal n is in between 3.8×10^{18} and $1.7 \times 10^{20}/\text{cm}^3$.

Now, we examine the enhancement of power factor by quantum confinement. Fig. 11(a) shows that the DOS of a 2 nm diameter Si nanowire is much larger than that of the bulk Si over the range of occupied energy states (up to ~0.1 eV). Given the same $n = 1.7 \times 10^{20}/\text{cm}^3$, the E_F in the nanowire would be reduced to −0.01 eV (as measured from lowest sub-band edge of the nanowire) from +0.10 eV in bulk Si. The electron distribution in the nanowire is also skewed towards the lower energy states due to the shape of the DOS, which results in a smaller average energy, $\langle E \rangle$, of +0.05 eV compared to +0.13 eV in bulk Si. Nevertheless, S has doubled in the nanowire to 0.2 mV/K because of large decrease in E_F. This yields a four-fold increase in S^2n.

Subsequently, the increase in S^2n will translate to increase in power factor if mobility remains the same or is not significantly reduced. It is difficult, however, to predict how the mobility in quantum well or nanowire may change from that in bulk material as there are many contributing factors. Firstly, the DOS is altered, which may leads to increase or decrease of the Fermi level. The dominant scattering mechanism may also change. For example, interface scattering, which is not prevalent in bulk material, may dominate in nanostructures that have a high surface-to-volume ratio. As shown by Fig. 8(b), these changes (to η, r, and D) will affect μ. The precise calculation of μ requires an in-depth knowledge of the interaction between electrons and device features, the investigation of which is beyond the scope of this review. In Section 2.6, we look at the experimental results, some of which show that there is no significant change in μ, whereas other results indicate that reduction in μ can compromise the gain from S^2n.

The enhancement of S^2n is primarily due to increase in the *magnitude* of the DOS, and not the change in *shape* of the DOS. Fig. 11(b) shows while the DOS of a 5 nm nanowire is similarly shaped, it has smaller magnitude compared to that of bulk Si, except for a small range of energy below ~0.01 eV. Therefore, given the same $n = 1.7 \times 10^{20}/\text{cm}^3$, the Fermi level of the quantum wire (+0.15 eV) has to be higher than that of bulk material (+0.1 eV), which leads to a much smaller $|S| \sim 0.03$ mV/K. Consequently, S^2n is reduced by almost an order of magnitude from the bulk value. The implication is then obvious: there is an enhancement of S^2n only when the wires are sufficiently small that the DOS becomes larger than that in bulk material.

The enhancement of S^2n through confinement effects is summarized in Fig. 12. The width of each shape in the figure represents the DOS, and follows intuitively from Fig. 7, where a larger magnitude of the DOS is obtained at the bottom of the band

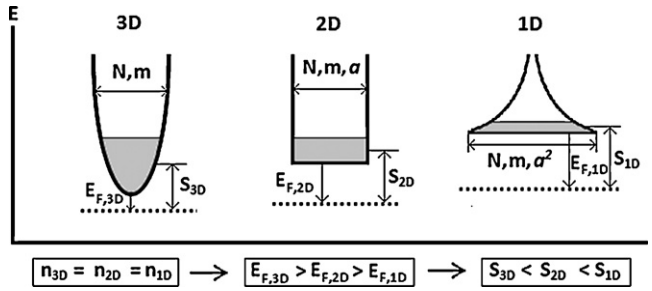

Fig. 12. Given the same n, E_F is reduced as magnitude of DOS is increased. The shape of the carrier pocket is reminiscent of the DOS in Fig. 7 the width of the shape represents the magnitude of the DOS, which is proportional to N, m, and (in 2D and 1D cases) a. $|S|$ subsequently increased for quantum well and nanowires as E_F is decreased.

with decreased dimensionality. The diagram also illustrates increased ground state energy due to an increased confinement potential at lower dimensions. The increase in ground state energy has no effect on S, σ, or κ_e when we consider only a single band. Consequently, given the same n, the E_F (relative to the bottom of the band) may progressively decrease from 3D to 1D. While the average energy of the electrons, $\langle E\rangle$ also decreases, the decrease is relatively smaller compared to the E_F, which then leads to increase in $|S|$ for the same n, as we move down in dimensionality, from three- to one-dimensions. A greater than ten-*fold* enhancement in the power factor, assuming no reduction in μ, via can then be predicted [39,40].

It may also be helpful to recognize the maximum thickness, $a_{\max}$, below which an enhancement of $S^2\sigma$ over bulk value can be expected. This thickness is found by solving the inequality: $S^2\sigma(\eta_{opt,r,D=2,1}) > S^2\sigma(\eta_{opt,r,D=3})$. Assuming a constant relaxation time ($r = 0$, and τ_0 a constant regardless of D), this critical thickness for quantum well or nanowire would be given by

$$a_{\max} = \sqrt{\frac{\hbar^2}{2k_BT}}\left(\frac{1}{\sqrt{\Pi_{i=1}^{3-D} m_{c,i}}}\frac{2\pi}{D}\frac{F_{(D/2)-1}(\eta_{opt,D})}{F_{1/2}(2.47)}\left(\frac{S_{opt,D}}{130\,\mathrm{mV/K}}\right)^2\right)^{1/3-D}, \quad (34)$$

where $m_{c,i}$ is the effective mass in the direction of confinement. Table 2 lists the estimates of $a_{\max}$ for a few materials, assuming isotropic effective mass. Generally, it would be more difficult to detect a power factor enhancement in materials with larger effective mass, as fabrication of a structure with smaller length scales would be required. This also suggests that the confinement direction should be that which has the smallest effective mass, which conversely, would ease the requirement for fabrication. Additionally, this ensures that m_d would be largest, which leads to greater $S^2\sigma$.

Table 2
Calculated $a_{\max}$ for Bi_2Te_3, PbTe, and SiGe quantum wells and nanowires. Enhancement of $S^2\sigma$ is expected when $a < a_{\max}$ (at 300 K).

	m/m_0	Quantum well, $a_{\max}$ (nm)	Nanowire, $a_{\max}$ (nm)
Bi_2Te_3	0.08	7.6	8.0
PbTe	0.12	6.2	6.5
SiGe	0.32	3.8	4.0

The location of the conduction valleys should also be taken into account, when picking the optimal orientation for the highest $S^2\sigma$. It was noted that confinement could break the symmetry of previously degenerate conduction valleys [41]. For example, the conduction valleys in PbTe are located at the L-point along the ⟨1 1 1⟩ axes with the degeneracy factor of 4, and have anisotropic longitudinal and transverse effective masses, m_l and m_t, respectively. In PbTe quantum well based structures grown along the [1 1 1] direction, the three oblique (to the growth/confinement direction) valleys experience a greater confinement potential than the single longitudinal valley, whose $m_c = m_l$ is largest, and which leads to smallest confinement energy. The lifting of degeneracy of the three oblique valleys from the single longitudinal valley leads to reduced DOS at the bottom of the conduction band, and subsequent reduction of $S^2\sigma$. To preserve the valley degeneracy, the PbTe quantum well structures should instead be grown in the [1 0 0] orientation [28]. Conversely, for n-type Si quantum wells, whose conduction valleys are located in the ⟨1 0 0⟩, the best growth orientation may be along the ⟨1 1 1⟩ orientation [42].

2.4.3. Carrier pocket engineering

The concept of *carrier pocket engineering*, first proposed by Koga et al. [43], takes further advantage of confinement effects by exploiting the shifting of carrier pocket/conduction valleys' energy levels. This method distinctively aims to enhance the $S^2\sigma$ by increasing the number of active conduction valleys. This is accomplished by utilizing the differences in confinement energy experienced by different carrier pockets, to shift the carrier pocket energies to similar levels.

To clarify these issues, Fig. 13(a) shows the carrier pockets in a Si/Ge multilayer structure, with the X-valley related to Si and the L-valley to Ge. In the first case, confinement is not present, and all

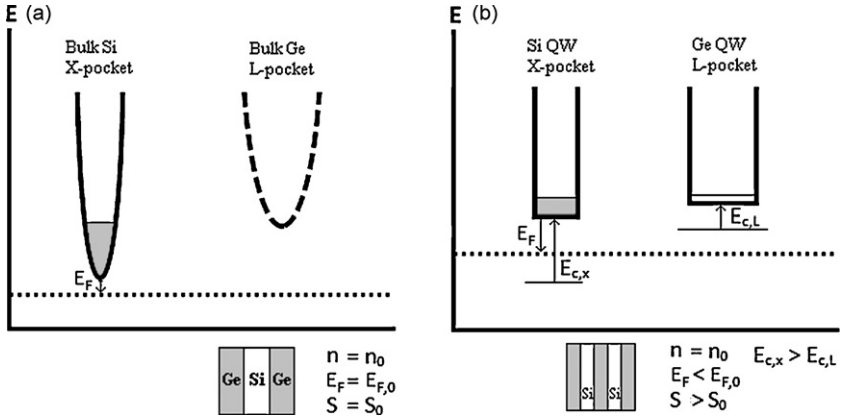

Fig. 13. (a) Carrier pocket configurations in a multilayer structure with bulk-like energy dispersion. Most electrons occupy the lower energy X-pocket. (b) Greater confinement energy in X-pocket ($E_{c,x} > E_{c,L}$) reduces energy separation to L-pocket in quantum well superlattices. Electrons now occupy both pockets, leading to increased DOS, reduced E_F, and increased $|S|$.

electrons occupy the X-valleys, which are at a lower energy compared to the L-valleys. The properties of this device would then be dominated by the characteristics of Si, as the Ge layers do not contribute electrically. If both the Si and Ge layer thicknesses were reduced sufficiently for quantum wells to form, with electrons in Si confined to the X-pockets, and those in Ge confined to the L-pockets, then the DOS in both valleys would change to those shown in Fig. 13(b). Since the L-pockets experience a smaller shift due to their larger effective mass compared to the X-pockets, both X- and L-pockets now have similar ground state energy levels and can be both heavily populated with electrons. Electrons are confined as the intervalley transfer of electrons between X- and L-valleys, would require substantial change in momentum. Both X- and L-pockets would contribute significantly to the overall properties of the superlattices according to the parallel conductor model outlined in Section 2.2.

An enhancement of $S^2\sigma$ then occurs due to both (1) the *local* increase in DOS for both X- and L-pockets, as discussed in the previous section, and (2) the *global* increase in DOS due to the addition of the L-pockets, which were previously inactive. Assuming that the mobility does not change, the power factor of the superlattices could then be larger than that of the bulk materials.

Carrier pocket engineering requires a rather exquisite matching of materials, which must have compatible crystal structure to enable defect-free growth, and electronic band structures with favorable energy configuration, e.g., GaAs/AlAs [43] and Si/Ge [42]. The band edges of Γ- and L-valleys in bulk GaAs, and X-valleys in bulk AlAs are within 0.28 eV of each other as shown in Fig. 14(a). In a GaAs (2 nm)/AlAs (2 nm) superlattices, these valleys can be brought much closer with an energy spread of just 0.05 eV for the ⟨1 1 1⟩ orientation, as shown in Fig. 14(b). The *ZT* of the GaAs/AlAs superlattices is expected to increased from the bulk value of ∼0.01 to ∼0.1 by virtue of the confinement effect on the Γ-pocket alone, and again from 0.1 to 0.4 by the addition of L- and X-pockets through carrier pocket engineering as shown in Fig. 14(c). However, since bulk GaAs and AlAs intrinsically have small *Z*, their superlattices, in spite the enhancement, would still have relatively small *Z* as well.

The Si/Ge superlattice structure holds more promise, since bulk n-type Si has intrinsically large power factor. Furthermore, there is opportunity to use the strain due to the lattice parameter mismatch of Si and Ge to manipulate the band structure. For example, an in-plane compressive strain present when Ge is grown on Si substrates, induces an energy splitting causing an increase (/decrease) of the longitudinal (/transverse) carrier pocket energies. For Si (2 nm)/Ge (2 nm) superlattices grown on (1 1 1) $Si_{0.5}Ge_{0.5}$ substrate, a *ZT* of ∼0.96 at 300 K and increasing 1.75 at 700 K, was predicted (cf., bulk SiGe has a $ZT \sim 1$ at 1000 K [8]).

In summary, carrier pocket engineering entails the selection of favorable growth orientations to increase the number of conduction valleys. However, this method requires meticulous materials design and may not be applicable in the as-formulated form to materials with conduction valleys solely at one point in the Brillouin zone, e.g., for p-type Si and Ge, which have the valence band maximum at the Γ-point.

2.4.4. Semimetal–semiconductor transition for increased power factor

Elemental Bi possesses two attributes of good thermoelectric materials: (1) a large μ due to large relaxation time and low effective mass (μ = 3.5 m^2/V s, compared to 0.14 m^2/V s for Si), and (2) low κ_L due to heavy atomic mass ($\kappa_L \sim 0.9$ W/mK) [44]. However, since Bi is a semimetal, its CB and the VB overlap slightly, and $|S|$ is small due to competing contributions from both bands.

However, if the competing contributions can be reduced, then Bi could exhibit large power factor. It was then pointed out that the confinement potential could be used to shift the CB minimum up and the VB maximum down, thus creating a band gap in Bi quantum wells [45], as illustrated in Fig. 15. The quantum well could subsequently be doped p- or n-type to obtain maximum $S^2\sigma$. The induction of the semimetal (SM)–semiconductor (SC) transition [44] to achieve higher $S^2\sigma$ has been studied in Bi, both theoretically and then experimentally (see Section 2.6.5), and in Bi-Sb alloys [46,47]. It may be also possible to induce SM–SC transitions in graphene through the formation of carbon nanotubes [48].

It was suggested that large power factor could be obtained from semimetals without SM-SC transitions, as well. Thonhauser et al. suggested that Bi can have larger power factor compared to Bi_2Te_3 if the dopant concentration is sufficiently large to push the Fermi level deep into the valence or the conduction band (i.e., E_F = 0.25 eV from band edge) [49]. This effectively makes the Bi a single band

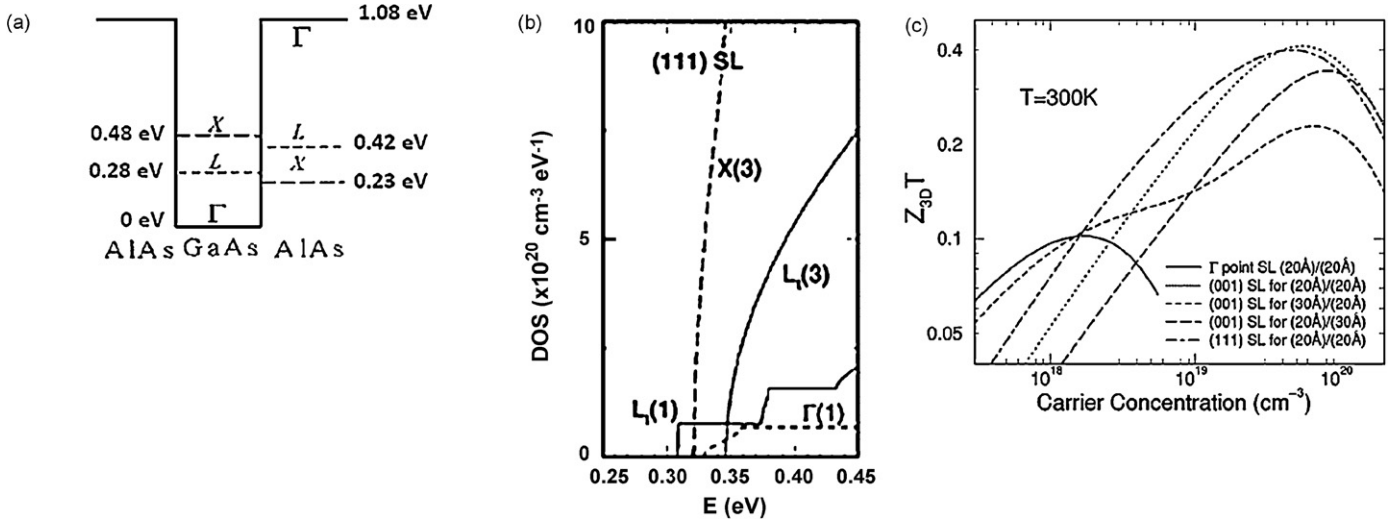

Fig. 14. (a) Relative conduction band edges at Γ-, L- and X-point in bulk GaAs and bulk AlAs heterostructures. (b) Density of states for [1 1 1] GaAs (2 nm)/AlAs (2 nm) superlattices. The energy scale is relative to the Γ-point of bulk GaAs, and is equivalent to that in (a). (c) The calculated *ZT* of GaAs/AlAs quantum well superlattices is increased when contribution from X- and L-pockets (lines except solid) are added that of Γ-pocket (solid line) (reprint permission from [43]).

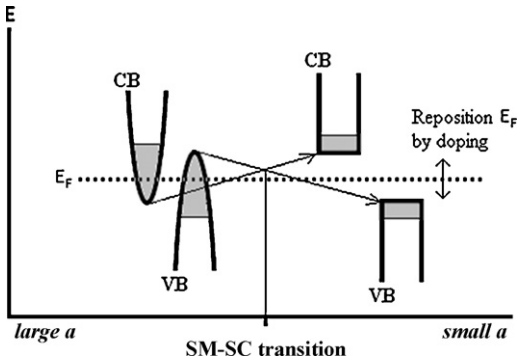

Fig. 15. The valence and conduction band (VB and CB) overlap in bulk Bi is reduced through using Bi quantum well structures by the cross-movement of the bands due to confinement energy. At the semimetal–semiconductor transition point, a band gap is induced. Subsequent doping would be required to make the material p- or n-type.

material. However, we have not seen supporting experimental evidence that confirms the calculations.

2.5. Consideration of barrier layers in superlattices

A common realization of quantum-confined devices is in the form of superlattices, where multiple quantum wells could be separated by a wide-bandgap barrier material, or where nanowires are embedded in a matrix, as depicted in Fig. 16. The transport coefficients of superlattices can be described by treating the quantum well/nanowire and barrier materials as parallel conductors. For example, the properties of quantum well superlattices (subscript "*SL*") would be the sum of the properties of the quantum well (subscript "*a*") and barrier (subscript "*b*") materials as

$$S_{SL} = \frac{S_a\sigma_a a + S_b\sigma_b b}{\sigma_a a + \sigma_b b} \tag{35}$$

$$x_{SL} = \frac{x_a a + x_b b}{a + b}, \quad x = \sigma, k_e, \text{ or } k_L \tag{36}$$

$$Z_{SL} = \frac{S_{SL}^2\sigma_{SL}}{\kappa_{e,SL} + \kappa_{L,SL}} \underset{\kappa_{e,a}a \gg \kappa_{e,b}b}{\overset{\sigma_a a \gg \sigma_b b}{=}} \frac{S_a^2\sigma_a}{\kappa_{e,a} + \kappa_{L,a} + \kappa_{L,b}^{b/a}} \tag{37}$$

The barrier material would ideally be electrically insulating since charge carriers are confined in the quantum well layer, and we may assume that $\sigma_a a \gg \sigma_b b$ (i.e., the sheet resistance of the barrier layers is much greater than that of the quantum well layers) and $\kappa_{e,a}a \gg \kappa_{e,b}b$. However, we could not always assume that $\kappa_{L,a}a \gg \kappa_{L,b}b$. With these assumptions, the figure of merit *of the superlattices* (Z_{SL}) would depend on the lattice properties of both materials, but would include the electrical properties of just the quantum well material, as indicated through the right hand side of

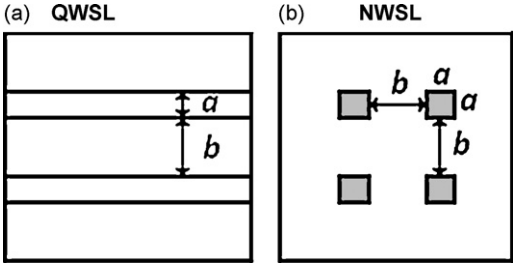

Fig. 16. Schematic cross-section of (a) quantum well and (b) nanowire superlattices.

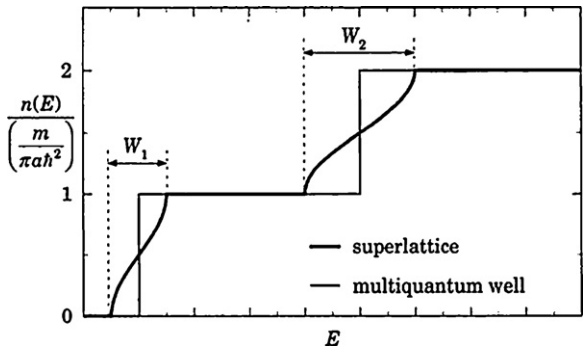

Fig. 17. Multi-quantum well (MQW) has an "ideal" 2D density of states, $g(E)$, with a sharp, step-like profile, while the density of states of a quantum well superlattice (SL) is "broadened" due to effect of electron tunneling. a is the period of the superlattice (reprint permission from [66]).

Eq. (37). In superlattices, there are several issues concerning the effect of barrier material [41,50–53].

While the quantum well/nanowire thickness (a) should still be as small as possible to enhance the power factor of the quantum well/nanowire ($S_a^2\sigma_a$), the barrier layer thickness (b) must be optimized to balance the effect of electron tunneling and parasitic heat conduction on *Z*. Tunneling is a quantum mechanical effect whereby electrons can penetrate the potential barrier, thus escaping confinement in the quantum well/nanowire. However, since the tunneling electrons never *exist* within the barrier layers, they do not increase carrier concentration or electrical conductivity of the barrier material itself. The coupling of nearby quantum wells/nanowires through tunneling introduces additional energy states, which broadens and reduces the DOS in the quantum wells/nanowires as shown in Fig. 17. The reduction in DOS caused a reduction in power factor and *ZT* [51], as shown in Fig. 18 when $a = b = 2.5$ nm. To prevent tunneling, it is necessary to have either thick barrier layer or large band offset between the quantum well/wire and barrier materials. Either solution unfortunately comes with a problem of its own.

The use of thick barrier layer increases parasitic heat conduction, which reduce *ZT* substantially. Eqs. (36) and (37) show that increasing b (or b/a) will essentially increase $\kappa_{L,SL}$ and decrease Z_{SL}. Parasitic heat conduction entails that the barrier layers can conduct much heat away from the source to the sink without

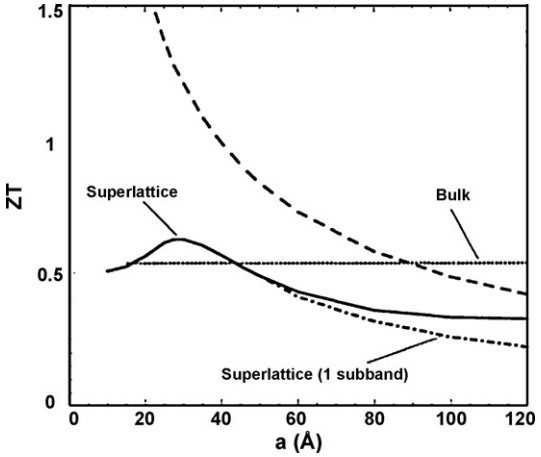

Fig. 18. The *ZT* of Bi_2Te_3 quantum well superlattice (solid line) with equal quantum well layer and barrier layer thickness. The dashed line represents the *ZT* of ideal multi-quantum well structures. The decrease in *ZT* at a ~2.4 nm is due to electron tunneling. The dash-dotted line is similar to the solid line except that only 1 subband was included in the calculation (reprint permission from [51]).

contributing to electrical power generation. Likewise, parasitic heat flow counters the heat-pumping action in a refrigeration device, leading to reduced efficiency. Since materials similar to the quantum well/nanowire are often chosen for the barrier material as well, to facilitate heterostructure growth, the thermal conductivities are comparable and both must be considered. The significant reduction from the ZT value of the dashed line (which assumes $b = 0$) to that of the solid line ($b = a$) shown in Fig. 18 is attributed to such parasitic heat conduction.

The tunneling probability can also be reduced by increasing the band offset through the use of wide band gap material as a barrier [28]. However, large band offset can enhance electron scattering at the interface and reduces mobility [41]. Fig. 19 shows that for a band offset of 0.2 eV, the mobility is essentially unaffected and may even be increased. However, a band offset of 1 eV can cause up to 40% loss of mobility.

Some confusion has also arisen from the preferences of different authors in how the $S^2\sigma$ or Z in superlattices are reported. For example, while some stated the carrier concentration or electrical conductivity per *volume of the quantum well* (i.e., σ_a), others state these quantities per *volume of the superlattices* [50–52] (i.e., $\sigma_{SL} = \sigma_a(a/a + b)$, which is equivalent to that of Eq. (36) if one assumes $\sigma_a a \gg \sigma_b b$). For example, while Hicks et al. have shown that S^2n in PbTe *quantum wells in superlattices* is *larger* than that of bulk PbTe [15], Ishida et al's interpretation of that same data showed that the S^2n of the *quantum well superlattices* is *smaller* than that of bulk PbTe [28]. The subtle difference here is that while Hicks et al. normalized the carrier concentration over the volume of *just the quantum wells*, Ishida et al take into account the volume of *the entire superlattice*, not excluding the volume of the barrier layers. Their respective conclusions then show that while Hicks et al's PbTe quantum wells may be superior to that of bulk PbTe, the entire superlattice as a device may be inferior.

This then suggest that having a thick barrier is undesirable since increasing b reduces S^2n of the superlattices (S^2n_{SL}). However, this is not the case, since the apparent reduction in S^2n_{SL} is compensated by the same reduction in $\kappa_{e,SL}$. Indeed, Eq. (37) shows that Z_{SL} would be equal to $Z_a = (S_a^2\sigma_a/\kappa_{e,a} + \kappa_{L,a})$, which is completely *independent of* b, if $\kappa_{L,b}$ was ignored. Therefore, the only true disadvantage of having thick barrier is the increase in parasitic heat conduction, which increases with barrier thickness as $\kappa_{L,b}(b/a)$.

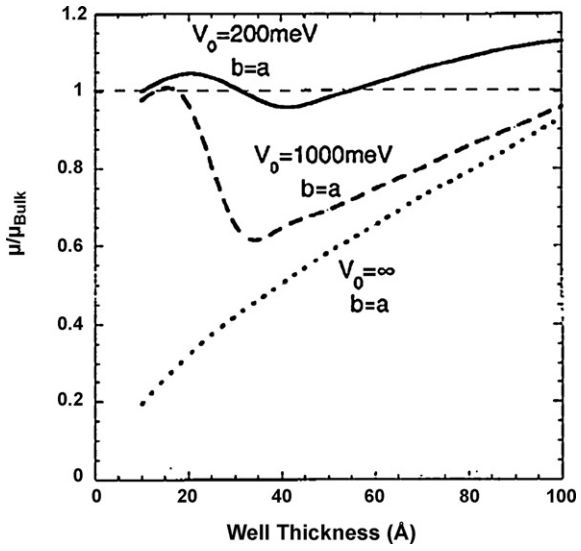

Fig. 19. The carrier mobility, plotted as a function of the [1 0 0] PbTe quantum well thickness, is diminished with increased height of potential barrier, V_0, due to enhanced carrier scattering at the interface (reprint permission from [41]).

In summary, the barrier thickness in the quantum well/nanowire should be carefully optimized to balance the adverse effects of tunneling, parasitic heat conduction, and loss of carrier mobility due to interface scattering.

2.6. Experiments in the enhancement of power factor in nanostructured thermoelectrics

Experiments in thermoelectric research are generally geared towards either measuring the enhancement in power factor ($S^2\sigma$) or the reduction in thermal conductivity (κ). In this section, we review experiments on PbTe-, Si-, and $SrTiO_3$-based quantum wells, and Bi nanowires, which are related to the first approach, while those related to the latter approach are discussed in Section 3.

2.6.1. Characterization method for quantum wells (QWs)

The ability to fabricate high-quality quantum well superlattices with atomic scale control, through techniques such as molecular beam epitaxy (MBE), has enabled two-dimensional layered thermoelectric materials. The earliest material investigated was PbTe, which was promptly followed by studies of SiGe. Recently, quantum well structures made of $SrTiO_3$ have also been studied.

Quantum well superlattices were relatively easy to characterize since the measurements are often done in the in-plane direction, which allow for larger sample size along with a large temperature gradient. This eases the handling, and measurement of S and σ. Furthermore, the thin film geometry of such superlattices enables the use of the established van der Pauw technique for determining n and μ [54] from σ and the Hall coefficient (R_H). These measurements allow for a comprehensive characterization as shown in Fig. 20. The Fermi energy (E_F) can then be estimated from n, if the number of conduction valleys (N), and the density of states effective mass (m_d) are known. Subsequently, S can be calculated from E_F and compared to experimentally measured values. Both the relaxation time parameters, τ_0 and r, can be deduced from measurement of the temperature (T) variation of either μ or S. Additionally, the Nernst coefficient (Q) could be determined as well, and gives direct information on the values of r. The measurement of S, σ, R_H, and Q and subsequent determination of material parameters are accomplished by the *method of four coefficients* [23,55].

In heterostructure devices such as the quantum well superlattices, the measured transport coefficients are a composite of the quantum well and barrier material as described through Eqs. (35)–(37). It is then necessary to perform additional measurements to determine the individual characteristics of the quantum well or barrier layers and extract the layers' properties. Of course, if the barrier layers are insulating, then the electrical properties of the quantum well can be more easily extracted from those of the superlattices.

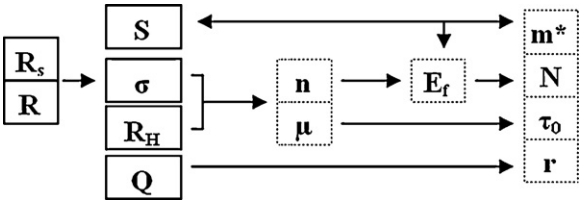

Fig. 20. Flow chart showing how the measured resistance (R) or sheet resistance (R_s) or electrical conductivity (σ), along with the Hall (R_H) coefficient can be used to determined carrier concentration (n) and mobility (μ). If the Seebeck (S) and Nernst (Q) coefficients are also measured, then all relevant parameters such as the effective masses (m^*) and number of conduction valleys (N), the scatterings constants (r and τ_0), and the Fermi level (E_F) can be determines as well.

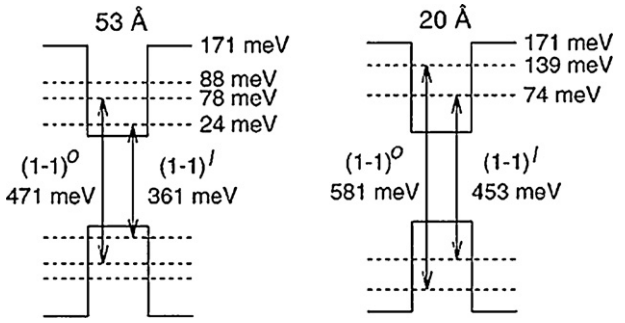

Fig. 21. Measured band gap of bulk PbTe and $Pb_{0.927}Eu_{0.073}Te$ and sub-band energy levels of longitudinal (l) and oblique (o) valleys in 5.3 and 2 nm [1 1 1] oriented PbTe quantum wells. The energy level reference is with respect to the bulk conduction band edge of PbTe (reprint permission from [15]).

2.6.2. PbTe-based nanostructures

Bulk PbTe is a direct band gap ($E_g \sim 0.31$ eV) semiconductor with both the conduction band minimum, and the valence band maximum at the L-point of the Brillouin zone. Its crystal structure is face-centered cubic (fcc), and there are eight equivalent L-point in the 1st Brillouin zone. Subsequently, both p-type (Bi doping [15]) and n-type (Na doping [27]) PbTe have four full (*or* eight half-full valleys, which straddle the edge of the first Brillouin zone) conduction valleys at the L-point ($N = 4$). The CB and VB coupling induces non-parabolicity in the band resulting in small, anisotropic, temperature-dependent effective masses (i.e., $m_L = 0.35m_0$, and $m_T = 0.034m_0$ at 300 K [41]). The large difference between the m_L and m_T induce strong anisotropy in the transport coefficients. Additionally, PbTe has a large μ, and small κ_L, which contribute to a high ZT of ~0.6 at the temperature of 700–800 K. Doping PbTe with Sn (n-type) could further reduce κ_L and increase the ZT to greater than 1 [8] (Figs. 21 and 22).

The enhancement of $S^2\sigma$ in nanostructures was first demonstrated in modulation-doped n-type [1 1 1] PbTe/$Pb_{0.927}Eu_{0.073}Te$ multi-quantum wells (MQW) fabricated via MBE by Hicks and Harman in 1996 [15,35]. As the bandgap and effective mass of the PbEuTe barrier layer (45 nm thickness) increased with Eu concentration, electrons are confined in PbTe layers. At 7.3% Eu concentration, the conduction band offset is estimated to be 171 meV, as shown in Fig. 23. The quantum confinement effects were verified through infrared transmission analysis, which showed the presence of energy levels consistent that of the oblique and longitudinal sub-bands. In [1 1 1] PbTe quantum wells, the degeneracy of the single longitudinal valley is lifted from the three oblique valleys because the oblique valleys have smaller confinement effective mass and thus experience larger confinement potential.

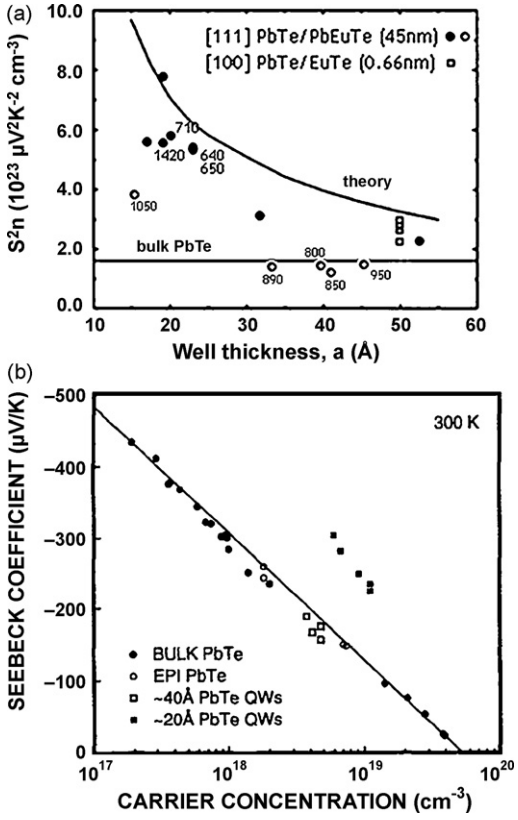

Fig. 22. (a) S^2n for [1 1 1] PbTe/PbEuTe MQW [15,35] (filled and open circles) and [1 0 0] PbTe/EuTe superlattices [28] (open square) increases as a function of decreasing quantum well thickness. Mobility values (in cm^2/V s) are labeled above data points (reprint permission from [15]) (b) 2 nm quantum wells showing larger $|S|$ for the same carrier concentration as 4 nm quantum wells and bulk PbTe (reprint permission from [35]).

An enhancement in the S^2n over the best bulk value was observed in MQWs up to a thickness of at least 5 nm by Hicks et al. [15], as shown in Fig. 22. Harman et al. also reported an enhancement of S^2n, but only for 1.7 nm MQWs [35]. There is a slight discrepancy in the reported data and the theoretical model (solid line), which assumed constant relaxation time ($r = 0$) and the $m_l = 0.620m_0$ and $m_t = 0.053m_0$. These effective mass values are

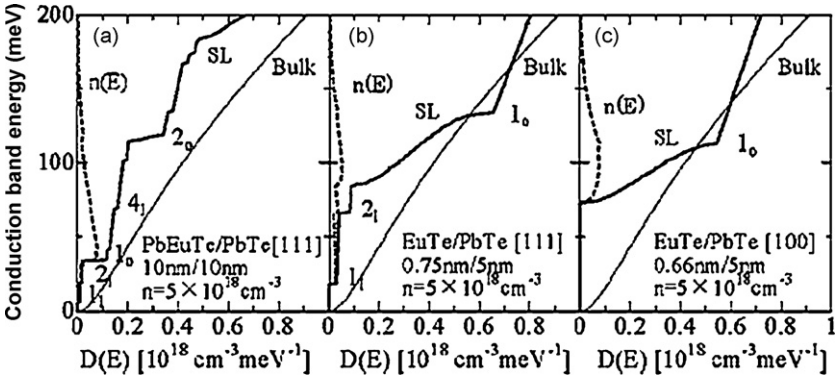

Fig. 23. The DOS of PbTe quantum wells superlattices (heavy solid line), and the bulk PbTe (light solid line). The subscripts l and o refer to the longitudinal and oblique sub-bands, while 1, 2, and 4 refer to the 1st, 2nd, and 4th lowest sub-bands. Electron density distribution is also shown by the dotted line, and corresponds to a total carrier concentration of 5×10^{18}/cm^3 (reprint permission from [28]).

considerably larger than the bulk values, and were extracted from optical measurements [56]. It was later thought that the increase in effective mass is due to penetration of electron wave function into the PbEuTe barrier layer, which has a larger effective mass compared to that of PbTe [57]. We speculate that a better fit may be attained if $r < 0$ (i.e., $r = -1/2$ for phonon scattering) is assumed, and which would also lowers the value of $|S|$.

The enhancement of $|S|$ is also clearly shown to be a result of confinement rather than a simple variation in n and E_F in Fig. 22(b). For equivalent n, the $|S|$ of the 2 nm MQW samples are clearly much larger than that of bulk samples, while those of the 4 nm MQW were comparable. Note that both bulk and quantum well samples exhibit a linear relationship between $|S|$ and $\log_{10}(n)$.

The μ of the PbTe QWs (~640–1420 cm^2/V s, Fig. 22) is similar to that of comparatively doped PbTe thin film (1200 cm^2/V s) and bulk PbTe (1000–1700 cm^2/V s), but much greater than that of PbEuTe (45 cm^2/V s). The growth conditions may explain the large variation in the QW mobility values while interface scattering may explain the reduced μ for the QW and the thin films, compared to the bulk samples. However, as discussed in Section 2.5, a potential barrier of 171 meV is not expected to provide a strong enough interface scattering to significantly affect μ [35]. A large μ of 1420 cm^2/V s for the QWs, which exceeds that of doped PbTe thin film, may also suggest that modulation doping (through carrier transfer from the barrier to the QW) is effective in increasing μ for a number of samples.

Assuming the bulk value of κ for PbTe, the concomitant enhancement of the power factor should yield a ZT of 1.2–2 for the PbTe *quantum well* at 300 K. However, it should be noted that the effective figure of merit for the entire structure (ZT_{SL}) would be much lower due to severe parasitic heat conduction through the thick PbEuTe layers as shown in Table 3.

In 2008, Ishida et al. reported a similar enhancement of S^2n in n-type [1 0 0] PbTe quantum well (5 nm)/EuTe barrier layer (0.66 nm) quantum well superlattices [28]. The larger band gap of EuTe compared to PbTe produced a large 1 eV band offset, which allowed the barrier to be much thinner while preventing excessive tunneling. The growth along the [1 0 0] should offers a more favorable density of states than the [1 1 1] direction, as the valley degeneracy (N) is preserved, which is shown through the comparison of the DOS in Fig. 23. Consequently, given the same n and a, the [1 0 0] QW structures should have a lower E_F and larger $|S|$ than the [1 1 1] oriented QWs. S^2n measured by Ishida et al for the [1 0 0] quantum wells has been renormalized to the volume of the quantum well alone (assuming negligible barrier layer conduction), and plotted alongside Hick and Harman's data in Fig. 22(a). Slightly larger S^2n was observed in the case of the [1 0 0] superlattices.

However, as the value of μ was not reported, it is not known whether the power factor was enhanced or not. While the use of very thin EuTe barrier layers significantly reduces parasitic heat conduction, it may also reduce μ due to the large potential barrier offset (see Section 2.5, Fig. 19). Such a large barrier (1 eV) could reduce the μ in 5 nm PbTe QWs by up to 40% from bulk values due to increased interface scattering [41]. For comparison, we assume bulk values for μ and κ, and then calculate that the PbTe/EuTe [1 0 0] superlattices would have ZT of 0.66 at 300 K.

Harman et al. also explored PbTe/Te quantum well and PbTe–PbSe quantum dot superlattices. However, confinement effects were not observed. Although an enhancement of S^2n for both systems was initially reported in the year 2000 [58,59], the results were later rejected in 2008 and considered to be a result of carrier concentration measurement error due to non-ideal configuration of the van der Pauw setup [60]; the PbTe/PbEuTe quantum well superlattices' measurement results (also by Harman et al. [35]) were not mentioned in the errata. X-ray diffraction analyses also confirmed that the PbTe/Te superlattices "were essentially just PbTe". Thus, unlike PbEuTe or EuTe, Te does not appear to provide a sufficient potential barrier for electron confinement. There also appears to be no electron confinement and no enhancement of the power factor in PbSe nanodots superlattices. Additionally, the mobility of the PbTe–PbSe superlattices actually decreased from that of bulk PbTe, resulting in smaller powers factors for the superlattices. The advantage of PbTe–PbSe quantum dot superlattices over bulk PbTe then appears limited to just the reduced lattice thermal conductivity due to phonon scattering by PbSe nanodots [60,61].

An interesting enhancement of S^2n was recently reported in *bulk* p-type PbTe doped with Tl by Heremans et al. [27]. In this case, instead of quantum confinement effects, *resonant states* [62] formed due to the interaction of the Tl with the valence band of PbTe leads to an increase in the DOS near the band edge, as illustrated in Fig. 24(a). Optical measurements on 1.5% Tl–PbTe sample indicated that the resonant states were located ~0.06 eV below the band edge, and increase the DOS by a factor of 2.6 above the bulk value. Electrical and thermal measurements showed a higher $S^2\sigma$ for the 2% Tl–PbTe sample, but with the same κ as a Na–PbTe control sample, which led to the doubling of ZT near 800 K as shown in Fig. 24(b). Enhancement of S^2n was also observed at lower temperature of 300 K; however, the ZT did not seem to be enhanced at this temperature. The increase in S^2n with increased DOS is consistent with the explanations in Section 2.4.2. The use of transition metal doping to increase the DOS and power factor in bulk material is attractive as such materials may be easier to characterize and manufacture. Extensive review of transition metal-doped lead chalcogenide can be found in [62].

2.6.3. Si/SiGe-based QWs

Bulk $Si_{0.7}Ge_{0.3}$ has a ZT close to ~1 at 1100 K, and was pioneered in the early 1960s for use in radioisotope thermoelectric

Table 3
Projected ZT values for bulk PbTe, and PbTe-based multiple quantum well (MQW) samples at 300 K. ZT_{QW} was calculated from the reported value of S^2n, assuming bulk value of PbTe for κ_L and μ. For the estimation of ZT_{SL}, we assume the same κ_L for both well and barrier material, and $\kappa_L \gg \kappa_e$, i.e. $ZT_{SL} = ZT_{QW}(a/(a+b))$.

PbTe QW thickness	Orientation	Barrier material	Thickness	ZT_{QW}	ZT_{SL}
2 nm	[1 1 1]	PbEuTe	45 nm	1.2–2	0.05–0.09
5 nm	[1 0 0]	EuTe	0.66 nm	0.75	0.66
Bulk PbTe					
ZT = 0.10 at 300 K.					

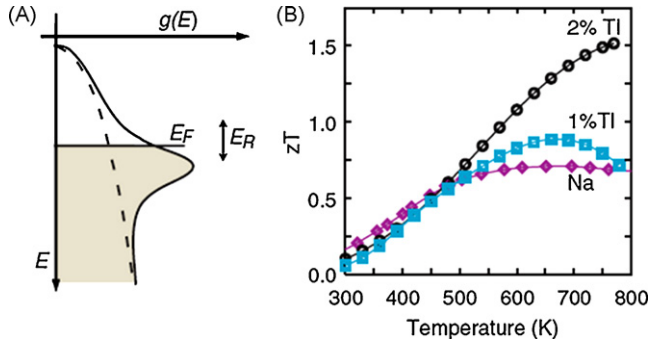

Fig. 24. (a) A spike in the density of states (solid line) above the bulk value (dashed line) occurs due to resonant states in Tl-doped PbTe. (b) The measured ZT of Tl-PbTe and Na-PbTe samples for 300–800 K indicates an improvement due to the addition of Tl (reprint permission from [27]).

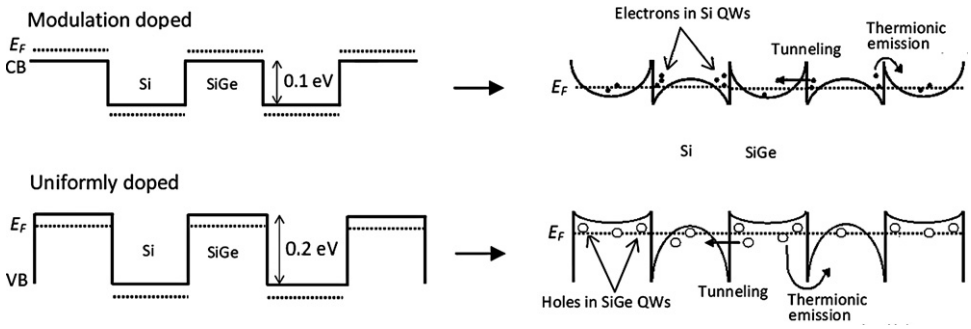

Fig. 25. Band bending in the conduction band (CB) of modulation-doped SiGe/Si superlattices (top figure, Sun et al. [65]) and in valence band (VB) of uniformly-doped Si/SiGe superlattices (bottom figure, Kato et al. [64]) may reduce the potential barrier thickness, allowing for significant tunneling by confined electrons and holes. Thermionic emission also possible given the small barrier height (~0.1–0.2 eV).

generators (RTG) for deep-space satellites. In contrast to PbTe, SiGe is an indirect band gap semiconductor with large m_d but smaller μ. Generally, the band structure [63] and electronic properties of $Si_{1-x}Ge_x$ alloy with $x < 0.8$ are Si-like with the lowest conduction minima near the X-point ($N = 6$) in the Brillouin zone. For $x > 0.8$, $Si_{1-x}Ge_x$ can be modeled through a Ge-like band structure with conduction band minima at the L-point ($N = 4$). The alloying of Si with Ge takes advantage of the high $S^2\sigma$ intrinsic to Si while drastically reducing the κ_L from 133 W/mK for Si, to ~10 W/mK for $Si_{0.7}Ge_{0.3}$ alloy at 300 K. At 1100 K, κ_L is ~4 W/mK for $Si_{0.7}Ge_{0.3}$.

$Si/Si_{1-x}Ge_x$ superlattices were studied by (1) Kato et al. [64] and (2) Sun et al. [65]. In the first study, $Si_{0.8}Ge_{0.2}$/Si superlattices, fabricated by MBE on a [1 0 0] Si substrate, were uniformly-doped p-type, with the $Si_{0.8}Ge_{0.2}$ as the quantum well (QW) layer and Si as the barrier layer. In the second study, $Si/Si_{0.7}Ge_{0.3}$ superlattices were modulation-doped n-type, with Si as the QW layer and $Si_{0.7}Ge_{0.3}$ as the barrier layer. Similar carrier concentrations ($n \approx 10^{19}/cm^3$), quantum well thicknesses ($a \approx 1$–5 nm), barrier thickness ($b \approx 20$–30 nm), and number of periods (10–15), were used in both studies. Si/SiGe heterostructures have a Type-II band alignment [66], where holes and electrons may be confined in the SiGe and Si layers, respectively, as illustrated in Fig. 25.

Significant barrier conduction was expected by both authors. For example, Kato et al. estimated that ~70% of the electrical conduction occurs in the Si barrier layers. This is plausible since the barrier layers are heavily doped in both studies. Additionally, the potential barrier is relatively low (due to a small band offset of 0.1–0.2 eV) and thin (due to significant band bending as a result of heavy doping in the barrier layer), which can lead to both thermionic emission over the barrier and tunneling as shown in Fig. 25. Both thermionic and tunneling currents would reduce quantum confinement effects. It should be noted that the former is a classical effect where thermally excited electrons gain adequate energy to surmount the potential barrier, while the latter is a quantum mechanical effect, where the tunneling electrons penetrate through the barrier in spite of having lower energy than the barrier.

The combination of barrier conduction and lack of detailed analysis and additional evidence of confinement effect raises doubt in regard to the observed increase of S^2n, as shown in Fig. 26. For example, Kato et al. hypothesize that the enhancement of S^2n (as a is decreased) observed in the superlattices could simply be due to the increased proportion of Si (the barrier thickness, b, was the same for all samples), which was thought to have a higher $|S|$ compared to the $Si_{0.8}Ge_{0.2}$. Since measurements on individual layers were not performed, it was not possible to explicitly determine the transport coefficients of the quantum wells. The theoretical model by Sun et al., also plotted (as a solid line) in the figure, was not explained in detail but supposedly assumes optimal dopant concentration and constant relaxation time. The overestimation of the model was attributed to the variation of the actual layers thicknesses due to strain caused by lattice mismatch.

Nevertheless, the comparison to the best bulk value of the S^2n for SiGe appears to show an enhancement. The best bulk values were not stated by Kato et al. or Sun et al., and consequently were compiled from [36]. In both cases, the S^2n values for the smallest superlattices are up to 2–3 times greater than the best bulk value, which indicates that the enhancement is likely due to the increase in DOS, rather than the increase in n (note that n coincidentally increases as a decreases - see Eq. (24)). It is also interesting to note that the increase in the S^2n for superlattices measured by Kato et al. appears weaker than that of Sun et al. This could perhaps be due to the larger barrier conduction in the former, as explained previously.

It was also noted that while the μ of the n-type $Si/Si_{0.7}Ge_{0.3}$ superlattices decreased with a, perhaps due to increased interface scattering, it was much higher compared to bulk Si or $Si_{0.7}Ge_{0.3}$ samples with smaller n and a. For example, $\mu \sim 535\ cm^2/V\ s$ for superlattices with $n = 1.8 \times 10^{19}/cm^3$, while μ of bulk n-Si with similar n is only ~100 $cm^2/V\ s$ [67]. As n is increased to $6.1 \times 10^{19}/cm^3$, the μ of the $Si/Si_{0.7}Ge_{0.3}$ QW superlattice, now measured to be ~105 $cm^2/V\ s$, approaches the value for Si. The higher mobility values might be indicative of the efficacy of modulation doping.

An indication of carrier pocket engineering was reported in n-type Si/Ge superlattices by Venkatasubramanian et al. [68], and is shown in Fig. 27 through the variation of the $S^2\sigma$ with the

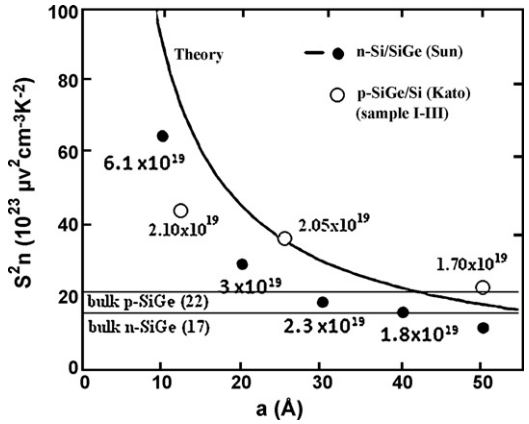

Fig. 26. S^2n of n-type Si/SiGe superlattices (●) from Sun et al. [65] and p-type SiGe/Si superlattices (○) from Kato et al. [64]. Carrier concentration per superlattices' volume is labeled next to data point. Theoretical calculation (solid line) is for the n-type superlattices. The maximum value of S^2n for bulk p- and n-type $Si_{0.7}Ge_{0.3}$ at 300 K were estimated from Ref. [36] (reprint permission from [65]).

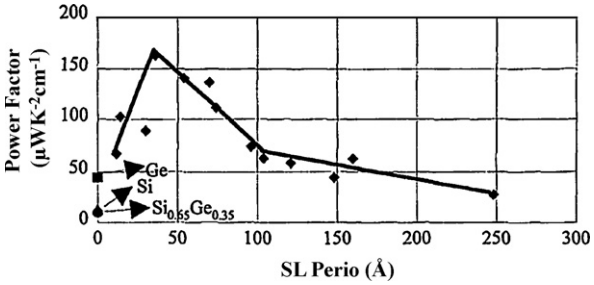

Fig. 27. The measured $S^2\sigma$ vs. the superlattices period in n-type Si/Ge superlattices (reprint permission from [68]).

superlattice period, d. For $d > 10$ nm, the slight increase over bulk material values was attributed to the increase in the DOS due to confinement. For $4\text{ nm} < d < 10\text{ nm}$, the sharp increase was attributed to the transfer/redistribution of electrons from *lighter* m_d L-conduction valleys to *higher* m_d X-conduction valleys, induced by the confinement potential, as discussed in Section 2.4.3. The maximum $S^2\sigma$ of 163 μW/cm K^2 was observed at $d = 4$ nm. For $d < 4$ nm, the decrease in $S^2\sigma$ was possibly due to the reduction in μ due to increased interface scattering. A lower *in-plane* thermal diffusivity (~0.02 cm^2/s), compared to the bulk value (~0.055 cm^2/s), was also measured in the superlattices. The highest ZT, of 0.75 at 300 K, was obtained at $d = 7.5$ nm. However, there is again a lack of detailed modeling and additional evidence of confinement effects to support the conclusion that the observations are due to confinement alone.

Carrier pocket engineering was also investigated on [1 0 0] oriented n-type Si (2 nm)/Ge (2 nm) superlattices [69]. While the ⟨1 1 1⟩ orientation was deemed the best growth orientation [42], fabrication of high-quality defect-free [1 1 1] superlattices could not be achieved. The magnitude of S indicated that ionized impurity scattering was dominant. However, the assumption of such a scattering mechanism overestimated the μ. The discrepancy between the measured and expected $S^2\sigma$, as shown in Fig. 28, was explained a result of an additional contribution from acoustic phonon and defect scattering.

The superlattices would have a ZT of ~0.1 at 300 K, assuming a κ_L ~5 W/mK. The 7-fold increase in ZT compared to the expected nominal value of 0.014 for *bulk Si*, was posited to be a successful validation of the conduction valley/carrier pocket engineering concept. However, since the κ_L for Si (~133 W/m K) is actually about twenty-six times bigger than that assumed value for the superlattices, this comparison actually suggests that the $S^2\sigma$ of the

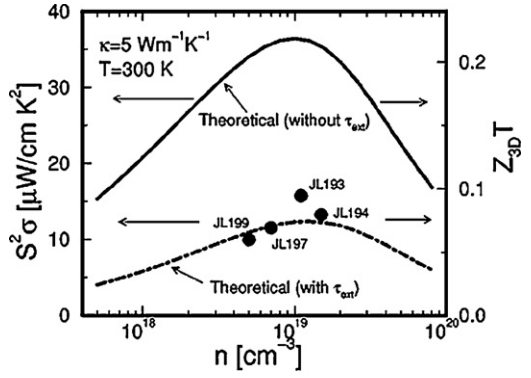

Fig. 28. The measured $S^2\sigma$ of n-type Si (2 nm)/Ge (2 nm) superlattices was lower than expected, and was explained as a result of additional scattering by superlattices interfaces (reprint permission from [69]).

superlattices must be much smaller! In fact, the maximum $S^2\sigma$ measured here (~15 μW/cm K^2) is an order of magnitude smaller than that observed by Venkatasubramanian et al. (~160 μW/cm K^2). Additionally, Koga et al.'s calculations did in fact suggest that the lowest L-valley sub-bands in Ge would be ~0.5 eV higher than the lowest X-valley sub-bands in Si, which implies a minimal contribution from the former [42]. Consequently, the validity of carrier pocket engineering seems to be questionable in this case.

In summary, while some experimental work seems to show an enhancement of S^2n over bulk value, the lack of detailed theoretical calculations that take into account the influence of the barrier conduction makes clear interpretation difficult. Consequently the claims of $S^2\sigma$ enhancement due to quantum confinement, and evidence of carrier pocket engineering is still unclear. It should also be noted that the optimal temperature of SiGe based alloys is over 1000 K, and the ZT at room temperature is much smaller than that of Bi_2Te_3. Hence, SiGe should be operated/characterized at much higher temperature. However, the possible degradation of the nanostructures due to atomic diffusion may render operation at such high temperatures difficult or impractical.

2.6.4. Metal oxide-based QWs

While materials with smaller band gaps (e.g., ~1 eV) are limited to lower temperature because of the earlier onset of bipolar conduction, metal oxide-based thermoelectrics with larger band gaps are suitable for high temperature operation in applications such as waste heat recovery from industrial processes, due to their chemical and thermal stability. We focus on $SrTiO_3$ in particular, which has received much attention lately. It should be mentioned at the very outset, that generally the growth of stoichiometric metal oxide thin films is challenging due to the issues associated with oxygen vacancies.

$SrTiO_3$ has a large band gap ($E_g = 3.5$ eV), which from the $10k_BT$ rule suggests that bipolar conduction would not be an issue for temperature less than 4000 K. The conduction band of $SrTiO_3$ has a minimum at the Γ-point of the Brillouin zone ($N = 1$), formed by triply degenerate Ti *d-orbitals*, with very large and anisotropic effective masses ($m_t = 1.1m_0$ and $m_L = 20m_0$ to $50m_0$) yielding $m_d = 6m_0$ to $13m_0$ [70]. Due to the large m_d and DOS, the carrier concentration can be very high while the $|S|$ is reasonably large. Indeed, the optimal carrier concentration in both bulk Nb-doped [71] and La-doped [37,71] $SrTiO_3$ (both n-type) were found to be approximately 3.3×10^{20} and 10^{21}/cm^3, respectively, while the $|S|$ is close to the predicted optimal value of 167 μV/K (Section 2.4.1).

The optimal n increases with temperature, and was found to be $\sim 4 \times 10^{21}$/cm^3 for Nb-doped samples grown by pulsed-laser deposition (PLD). While cumbersome and inefficient for mass production, PLD was necessary to overcome the thermodynamic dopant solubility limit of 4×10^{20}/cm^3 [38]. The ZT was shown to peak at ~0.4 at 1000 K. Research into new synthesis methods that can overcome these thermodynamic limits, or the use of multiple-electron donors to enhance carrier concentrations, could improve the prospects of $SrTiO_3$ in high temperature thermoelectric application.

A large enhancement in the $S^2\sigma$ as a result of quantum confinement effects was observed in $Sr(Ti_{0.8}Nb_{0.2})O_3/SrTiO_3$ multi-quantum wells (MQW) [72,73]. The QW formation in doped $Sr(Ti_{0.8}Nb_{0.2})O_3$ layers was due to the localization of electrons from the Nb donors, as confirmed through electron energy loss spectroscopy (EELS) and optical absorption measurements [72]. As shown in Fig. 29, a very large (dotted lines) twenty-seven-fold enhancement of S^2n over that of comparatively doped bulk $SrTiO_3$ was observed. A more moderate increase of ZT from 0.09 in bulk to 2.4 in quantum well was seen (solid lines) for the sample with smaller barrier thickness ($b = 9$ unit cells), presumably due to tunneling effects. However, when parasitic heat conduction is

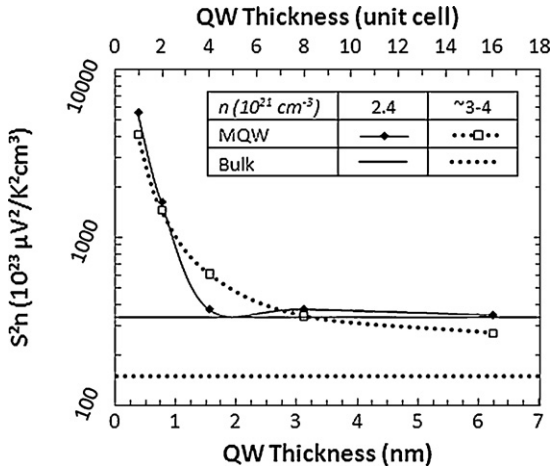

Fig. 29. The S^2n of $Sr(Ti_{0.8}Nb_{0.2})O_3/SrTiO_3$ superlattices vs. quantum well thickness. Both Ohta et al. [72] and Mune et al. [73] data are presented. The barrier thicknesses, in unit cell, are $b = 9$ for Ohta et al. and $b = 16$ for Mune et al.

considered, the effective ZT for this MQW would be reduced to ~0.24.

Although the S^2n of $SrTiO_3$ is much larger than that of PbTe and SiGe, the $S^2\sigma$ is still comparable due to very low μ, as shown in Table 4. The low μ is a result of the localized nature of the d-orbital electrons and very large effective masses. Otherwise, $SrTiO_3$ has relatively comparable $\kappa_L \sim 3$ W/mK at 300 K, which increases to 10 W/mK at 1000 K [71]. The increasing κ_L (with temperature) is due to an increased number of active phonon modes associated with the oxygen atoms, and could lower ZT at high temperature, which would unfortunately diminish the high temperature advantages of $SrTiO_3$.

2.6.5. Bi nanowires

As discussed in Section 2.4.4, Bi has potential for high Z due to high μ and low κ. However, Bi must first be converted from semimetal (SM) in bulk form to a semiconductor (SC), which can be accomplished through quantum confinement effects. This provides the motivation for the study of Bi quantum wells and nanowires. Furthermore, since Bi has small effective mass and large carrier mean free path, the quantum confinement effects can be observed at much larger length scales. For example, a ZT of 3 was predicted for a 10 nm Bi quantum well [44], whereas a similar value of ZT would be expected for a PbTe quantum well of 1 nm thickness [39].

Bi nanowires are easier to fabricate than quantum wells as Bi has a very low melting point (~271 °C), which makes the synthesis amenable to melt-injection and vapor-deposition techniques. Bi nanowires with diameters around 7 nm have been fabricated in porous alumina templates, which can be easily prepared with good control over pore size, via electrolytic processes. Moreover, alumina is an electrical insulator, and will not interfere with electrical measurement of the embedded Bi nanowires and could be selectively etched away leaving behind Bi nanowires [74].

Subsequently, there has been much theoretical and experimental work to determine the properties of Bi nanowires, with the emphasis on identifying the critical wire diameter, d_c, where the SM–SC transition occurs. Early calculations assumed parabolic bands at the L-point of the conduction band and the T-point of the valence band, relied on the cyclotron effective masses, and modeled the nanowire cross-section as a square to determine d_c [75]. These early values of d_c were subsequently refined through inclusion of non-parabolic band model for the conduction band L-point [76,77], assumption of *circular* instead of square nanowires, and using calculated effective masses for the individual conduction

Table 4
Materials with large effective mass tend to have large S^2n but small μ (the best S^2n values which were taken from Figs. 22, 26, 29).

	S^2n (10^{23} μV²/K² cm³)		Bulk, μ (cm²/V s)
	Bulk	QWs	
PbTe [from Fig. 22]	1.6	8	1400
n-SiGe [from Fig. 26]	15	65	100
$SrTiO_3$ [from Fig. 29]	100	5000	6

valleys in place of the cyclotron effective mass, which is essentially an average of effective masses of multiple valleys [78]. This latter effort indicated that, at 77 K, the d_c should equal 55.1, 39.8, 48.5, 48.7, and 49 nm for circular Bi wires orientated along the trigonal, binary, bisectrix, $[1\,0\,\bar{1}\,\bar{1}]$, and $[0\,1\,\bar{1}\,\bar{2}]$ directions, respectively. X-ray diffraction analysis has revealed that $[1\,0\,\bar{1}\,\bar{1}]$ is preferred growth direction for larger diameter nanowires ($d > 60$ nm), while $[0\,1\,\bar{1}\,\bar{2}]$ is preferred by smaller nanowires ($d < 50$ nm) [79]. Nevertheless, both orientations are of similar symmetry as that of the bisectrix direction, and thus the band structure and d_c are similar in value for all 3 directions.

In addition to the dependence on crystal orientation, the d_c also depends on temperature. As illustrated in Fig. 30, the d_c *decreases* to 15.4, 11.2, 13.6, 13.6, and 14.0 nm for the aforementioned orientations as temperature rises to 300 K. Note that most measurements and calculation results are often stated for 77 K, which is deemed the suitable cryogenic operating temperature for a Bi nanowire-based cooler.

It was also theoretically shown by Lin et al. [78] that both p- and n-type Bi nanowires can have the exact same ZT when $d < 5$ nm. Normally, large-diameter p-type Bi nanowires are expected to be inferior to n-type wires due to unfavorable orientation of the dominant T-point valence band valley. However, when $d < 5$ nm, the T-valley is pushed to below the L-point valence band valley by confinement effects. Since the L-point valence band valley is a mirror image of the L-point conduction band valley, both p- and n-type 5 nm Bi nanowires are expected to have identical electronic properties with a predicted ZT of ~6 at 77 K.

Experimentally, the behavior of resistance versus temperature, $R(T)$, is often used to demonstrate the SM–SC transition [74,77,80]. Since it is not possible to identify the number of wires in a typical Bi

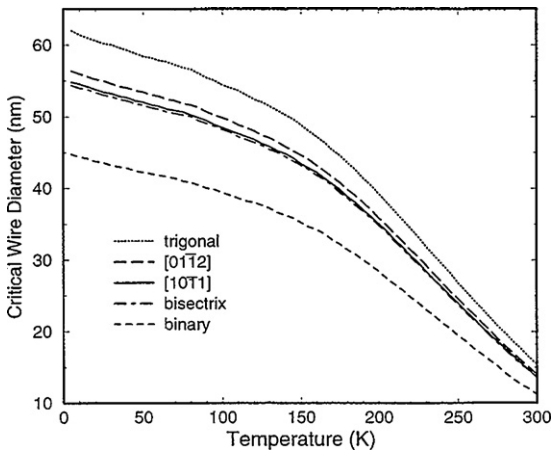

Fig. 30. The critical length scale, d_c for the SM–SC transition in Bi nanowires increases as temperature is decreased, and differs according to crystal orientations. The binary (x), bisectrix (y), and trigonal (z) directions are parallel to the principal crystallographic axes, while the $[1\,0\,\bar{1}\,1]$ and $[0\,1\,\bar{1}\,2]$ are experimentally observed growth orientations for large (>50 nm) and small (<50 nm) diameter wires (reprint permission from [78]).

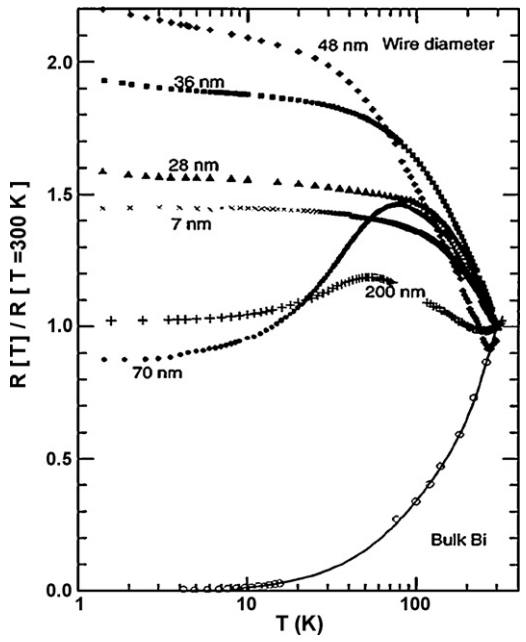

Fig. 31. The variation of the normalized resistance with temperature for Bi nanowires indicates a semimetal–semiconductor transition at ~50 nm (reprint permission from [80]).

nanowire array, normalized resistance, $R_n(T) = R(T)/R(300\ \mathrm{K})$, as indicated in Fig. 31 [80], is used instead. It is observed that R_n increases with T for bulk Bi and Bi nanowires with $d > 70$ nm, consistent with metallic behavior, whereas nanowires with $d < 50$ nm exhibit semiconductor-like behavior where R_n decreases with T. This behavior was subsequently cited as evidence of a SM–SC transition. However, as the absolute value of σ and μ was not determined in these nanowire arrays, it is arguable whether the observed $R_n(T)$ is truly indicative of the SM–SC transition.

At least a few other studies have found contradictory results. Firstly, theoretical calculations had earlier predicted $d_c \sim 15$ nm at 300 K. Therefore, wires with ~50 nm diameter are expected to be metallic at room temperature, in contrast to the conclusions reached through Fig. 31. Furthermore, measurement of S and σ of 40 nm × 15 nm rectangular, polycrystalline Bi nanowires as a function of gate voltage at 20 K showed semi-metallic behavior [81]. Magnetoresistance measurements of arrays of 80 nm and 30 nm diameter Bi nanowires at 5 K also indicated a metallic state [82].

The measurements of the S of Bi nanowires also have not been very informative due to the large variation in experimental variables among various studies, i.e., variability in wire diameter, crystal orientation, dopant concentration and type (n- or p-), temperature of measurement, etc. All these variables affect S, and thus a comparison of data reported by different research groups is not easy. Likewise, there has not been a very systematic study of the effects of n vis-à-vis S in Bi nanowires to show conclusive enhancement due to the band separation or DOS increase.

A few related studies will be presented. Cronin et al. [74] measured $S \sim -60\ \mu$V/K at 300 K for an array of 40 nm wires orientated along the $[10\bar{1}\bar{1}]$ direction. For comparison, the reported S for bulk Bi at 300 K in the direction parallel and perpendicular to the trigonal direction were −100 and −50 μV/K, respectively [45]. Huber et al. report a smaller $S \sim 15\ \mu$V/K for an array of 200 nm undoped nanowires (with unknown orientation) in the temperature range of 77–300 K [83]. Boukai et al. were able to measure the S of a single undoped Bi nanowire, using gate bias to control the carrier concentration, for 28 nm $< d <$ 72 nm, and found a small $S \sim -10\ \mu$V/K [81]. However, since these nanowires were not doped, an increase in $|S|$ may not be seen even if a bandgap was induced, as neither holes nor electrons dominate the conduction process. Meanwhile, Heremans et al. claimed that the observation of very large $|S| \sim 300$ mV/K in array of 9 nm nanowires was due to a SM–SC transition [84]. However, such a value is unbelievably large for an electrical carrier based diffusion process. According to Eq. (28), such value suggests an energy separation between $\langle E \rangle$ and E_F of 30 eV which seems quite improbable. Hence, this observation cannot be explained by conventional electron diffusion theory. To our knowledge, the only other comparably value of $|S|$ seen in literature is perhaps the "giant" Seebeck coefficient of ~30 mV/K that was predicted in graphene, which was attributed to its peculiar band structure with corresponding spikes in the DOS (see Section 2.6.2) [85].

In summary, it was seen that Bi nanowire based thermoelectrics research has been plagued by experimental uncertainty in measurement and precludes definitive conclusions. Many of the problems can be attributed to the paucity of measurement methods that could individually determine n and μ in the nanowires. The use of magnetoresistance measurements, capable of determining a complete set of relevant parameters (i.e., S, n, μ, and σ), along with the characterization of the energy levels/subbands [86,87] could help in this regard.

3. The effect of the lattice—thermal conductivity

Of the total thermal conductivity, constituted from the contribution from the carriers/electrons (κ_e) and the lattice (κ_L), the former is directly proportional to the electrical conductivity (σ) through the Wiedemann–Franz law (Section 2.3.3) and is generally small in magnitude compared to κ_L. Consequently, in this section we mainly address κ_L. As previously mentioned, a large majority of the understood increase in the figure of merit, ZT, of nanostructured thermoelectrics has been facilitated through a decreased contribution of the lattice thermal conductivity. A lower κ_L reduces parasitic heat transport by the lattice and the consequent increased proportion of heat transport by electronic carriers leads to a more efficient heat-to-electrical energy conversion. In this section, we will first review the phenomenology behind κ_L in order to get a better appreciation of the extant ideas and experimental efforts aimed at reducing κ_L.

3.1. Semi-classical theory of thermal conductivity

It can be derived from the elementary principles of the kinetic theory of gases [88], considering energy transport through *non-interacting* particles [89], that the lattice thermal conductivity is $k_L = C_L v l$, where C_L is the heat capacity of the lattice per unit volume, v is the average particle velocity, and l the particle mean free path. The interested reader could refer to Kittel [88] for an elementary exposition of this formula and its consequences. In a solid, the "particles" involved in heat transport would refer to the lattice modes/phonons. The same formula for κ_L is also obtained through the semi-classical form of the Boltzmann transport equation (Eq. (2) of Section 2.1), which is reproduced below:

$$\left(\frac{df}{dt}\right)_{sc} = \frac{df}{dt} + \frac{d\mathbf{k}}{dt}\nabla_{\mathbf{k}} f + \frac{d\mathbf{r}}{dt}\nabla_{\mathbf{r}} f \tag{2}$$

$\mathbf{k}$ and $\mathbf{r}$ are now the wave and position vector of the phonons (instead of electrons), and f is the Planck distribution function: f_P (instead of Fermi–Dirac function). The left hand side of the equation is again approximated by the *relaxation time* ansatz,

where the perturbed distribution function (f_P) relaxes back to the equilibrium distribution ($f_{P,o}$) within the relaxation time (τ), i.e., $(df/dt)_{sc} = (f_p - f_{p,o})/\tau$. At steady state ($df_P/dt = 0$) and neglecting the momentum dispersion of the phonons, i.e., $\nabla_k f = 0$, we get from the *ansatz* that $f_P = f_{P,o} - \tau v \nabla_r f$, where $v = dr/dt$. Since the heat flux $Q(r,t) = \sum_k v f_p E(k)$, we can now write:

$$Q(r,t) = \int v[f_{P,o} - \tau v \nabla_r f_p] E(k) g(E) dE, \tag{38}$$

where the summation over all the modes is replaced by an integral involving a continuous density of states (DOS): $g(E)$. The integral over $f_{P,o}$ vanishes, and writing $\nabla_r f_p = \nabla_T f_p \nabla T$, we get:

$$Q(r,t) = \nabla T \int \tau v^2 \nabla_T f_p E(k) g(E) dE \tag{39}$$

The κ_L is defined as the *ratio* of the $Q(r,t)$ to the temperature gradient (∇T), i.e.,

$$\kappa_L = \int \tau v^2 \nabla_T f_p E(k) g(E) dE \tag{40}$$

This expression can be simplified to $\kappa_L = \tau v^2 C_L$ by invoking the definition of the lattice heat capacity $(C_L) = dU/dT = (d[\int f_p E(k) g(E) dE])/dT$ as the temperature derivative of the total energy (U). If the notion of a phonon mean free path, $\boldsymbol{l}$ $(= v\tau)$ is introduced, then,

$$k_L = C_L v l \tag{41}$$

The crux of κ_L reduction, for obtaining a higher figure of merit in thermoelectrics, is then embedded in ways and means to reduce the individual components, i.e., C_L, v, and l. Such aspects will be considered next through an exposition of basic phenomenology.

3.2. The energy dispersion of phonons

In the previous section, two important assumptions were made, i.e., (i) the particles/phonons were considered to be non-interacting, and (ii) their momentum dispersion was neglected. While these aspects do not qualitatively change the dependence of κ_L on C_L, v, and l, exemplified in Eq. (41), a deeper understanding of the phonon interactions and dispersion would enable a better insight into κ_L and is the purpose of this section. A simple model of the influence of the interaction between atoms on discrete lattice points in response to an applied force is implemented by considering harmonic interactions between nearest neighbors, through Hooke's law, where the force is linearly proportional to the displacement. The resulting modes of vibration of the lattice – *phonons*, can be correlated with the quanta of vibrational energy. It should be noted that there is no longer an analogy to non-interacting particles as number and momentum conservation laws are not strictly valid in the latter case [18]. Consequently, phonon dispersion (phonon frequency – ω vs. wavevector – $\mathbf{k}$) relationships of the form $\omega = \sqrt{4K/m}\sin(ka_o/2)$ may be derived, where a_0 is the lattice point spacing and K is an indicator of the strength of coupling (related to a spring constant like parameter) between the points.

One can then determine the phonon group velocity, v, which enters into the expression for κ_L through $v = d\omega/dk$. It is to be noted that the continuum mechanics formula for v $(= \sqrt{Y/\rho})$, with Y being the elastic modulus and ρ the material density) is strictly valid for $\lambda \to \infty$ or $k \to 0$, e.g., for Si, with $Y \sim 180$ GPa and $\rho \sim 2330$ kg/m^3, $v \sim 9 \times 10^3$ m/s. It can also be deduced that the number of phonons per unit energy – the density of states (DOS) $= (dN/d\omega) = (dN/dk)(dk/d\omega)$ is inversely proportional to v.

A single or multi-valued basis at a lattice point gives rise to low energy acoustic phonon modes and higher energy optical phonon modes named in relation to their response to acoustic (sound) disturbances and electromagnetic radiation, respectively. Generally, the acoustic phonons describe the *inter*-unit cell response of the lattice to external forces, e.g., temperature or voltage while the optical phonons describe *intra-cell* interactions [90]. The former have a larger spread of energy/greater energy dispersion, which from $v = d\omega/dk$, implies a greater distribution of phonon velocities, while the energy dispersion of the optical phonon modes is quite weak, i.e., with small $d\omega/dk$ and low propagation velocity. Since a low v is important for a reduced κ_L, it is important to consider features in the ω vs. k dispersion/phonon band structure which have a small slope. The presence of internal structure (i.e., charge or mass) at a lattice point would also yield additional optical modes, with the implication that such modes have low group velocity and do not contribute to the κ_L. Generally, the increased entropy due to such additional modes could also result in a lower κ_L.

Each acoustic and optical mode can further be decomposed into one longitudinal and two transverse branches of polarization. Extrapolating the phenomenology of lattice points to unit cells, it is seen that for p number of primitive cells with q atoms/unit cell, the total number of degrees of freedom is $3pq$ out of which only $3p$ degrees can be accommodated through the acoustic modes. Consequently, the remaining $3p(q-1)$ degrees of freedom are accommodated through the optical branches. The implication is that structures with a larger number of unit cells (larger p) and with more complex crystal structures (larger q) have a greater number of optical modes and a reduced κ_L. These ideas lead to the notion that an increased crystal complexity decreases κ_L. Such structural characteristics are quantified through a complexity factor, CF (defined as the number of atoms/primitive unit cell) where κ_L is proportional to $CF^{-2/3}$ [91,92]. A large span of CF was postulated, e.g., $CF = 1$ (rare gas crystals, viz., Ar, Kr, etc.), $CF = 2$ (e.g., Si, Ge, GaAs, PbTe), $CF = 14$ (e.g., $MgAl_2O_4$), $CF = 105$ (e.g., β-boron), and $CF = 414$ (e.g., YB_{68}). As the κ_L varies over five orders of magnitude ($\sim 10^3$ to 10^{-2} W/mK), the minimum κ_L would correspond to a maximum CF of $\sim 10^7$ [93].

The crystal structure, through the acoustic and optical phonon energies, also contributes to the lattice heat capacity, C_L. We first consider the contributions from the former. In the Debye model approximation, the energy (U) is evaluated by considering the velocity, v, of the acoustic modes to be a constant, independent of polarization. The C_L is estimated to be proportional to $\left(\frac{T}{\Theta_D}\right)^3$ where Θ_D is the Debye temperature. Θ_D has the same significance for the phonons as E_F does for electrons, in that it can be used to describe the energy through $k_B\Theta_D$, of the acoustic phonons. Θ_D is also inversely proportional to the material density/lattice point spacing [18] which in turn could be influenced through tensile and compressive strain. However, scant attention seems to have been paid in literature for reducing C_L through the manipulation of Θ_D. The contribution of the optical phonons to the C_L is evaluated next through the Einstein model which assumes independent harmonic oscillators [18] at a constant frequency, ω_o. The C_L evaluated using this model tends to the Dulong–Petit value (=24.9 J/mol K) at high temperatures. As the latter was derived from the kinetic theory of gases it may be surmised that the assumption of non-interacting oscillators is more valid at higher temperatures.

The temperature variation of κ_L generally follows the variation of C_L, i.e., as T^3 and such a dependence is indeed found at low temperatures, e.g., at $T < 20$ K, for Si [94] and Ge [95]. At higher temperatures, anharmonic interactions, e.g., phonon–phonon/carrier/defect interactions, which involve a redistribution of energy and momentum are necessary to understand the variation of κ_L. We will first consider phonon–phonon interactions. It has been found that the dominant interactions are of the three-phonon collision type [96,97], which can be considered through an N-process (Normal) and U-process (Umklapp). An example of a

momentum conserving N-process involves a phonon–phonon collision of the type: $\mathbf{k}+\mathbf{k}'\rightarrow\mathbf{k}''$, where $\mathbf{k}$ and $\mathbf{k}'$ are the wavevectors of the incident phonons, while $\mathbf{k}''$ is the wavevector of the resultant phonon. In the N-process, the phonon energy is only *redistributed* from the incident phonons to the transmitted phonons and an infinite κ_L is implied. On the other hand, momentum *non-conserving* U-processes, represented through: $\mathbf{k}+\mathbf{k}'\rightarrow\mathbf{k}''+\mathbf{G}$, result in excess momentum (represented through $\mathbf{G}$ – the reciprocal lattice wavevector) from the phonon collisions. The presence of $\mathbf{G}$ ($=2\pi/a$) where a is the lattice parameter, implies a net phonon *backscattering* and a resistance to phonon and thermal transport resulting in a finite κ_L. The tendency for the occurrence of U-processes is found to be temperature dependent, scaling as $e^{-\Theta_D/2T}$, and is the basis of Eucken's law [6], where κ_L decreases with increasing temperature above Θ_D. For thermoelectric materials, operated at room temperature and beyond, the latter relationship, i.e., κ_L scaling as $1/T$ generally holds and has to be considered in the evaluation of the figure of merit.

Finally, a reduction of the κ_L could also be accomplished through a reduction of the phonon mean free path l ($=v\tau$). This provides a good motivation for the use of nanostructured thermoelectrics, such as thin films and nanowires. While both elastic and inelastic scattering processes are possible, typically the inelastic scattering length is much greater than the elastic scattering length and consequently we only consider the latter limiting process. The length scales involved in phonon transport can be classified in relation to the phonon wavelengths. As we have seen from the phonon energy dispersion, there is a wide range of wavelengths and wavevectors (λ and k), ranging from the spacing between individual lattice points ($\lambda=2a$, $k=\pi/a$) to the continuum limit ($\lambda=\infty$, $k=0$). Additionally, important phonon wavelengths (λ_{ph}) can be associated with dominant interaction forces, e.g., when the temperature is taken as a generalized force [98] driving the lattice vibrations, then the corresponding λ_{ph} can be calculated through $E=\frac{hv}{\lambda_{ph}}=k_BT$, which at 300 K, is ~1 nm. It is interesting to note that as the corresponding wavevector is close to the edge of the Brillouin zone, there are a large number of such *thermal phonons* due to the increased DOS. When l is of the order of the smallest λ_{ph}, i.e., ~$2a$, the *minimum value* of κ_L would be reached, for a given C_L and v. Yet another interpretation of the minimum κ_L was attempted following an early proposal by Einstein [99], where random transport of thermal energy from one oscillator/atom to another throughout the solid over an average time t ($=1/f$), implies $l=v/f=\lambda_{\mathrm{ph}}$. Consequently, a minimum value for the κ_L was derived to be:

$$\kappa_{L,\min}=\left(\frac{\pi}{6}\right)^{1/2}k_Bn^{2/3}\sum_i v_i\left(\frac{T}{\Theta_D}\right)^2\int_0^{\Theta_D/T}\frac{x^3e^x}{(e^x-1)^2}dx \tag{42}$$

Eq. (42) has been used for predicting $\kappa_{L,\ min}$ for a given mode velocity, v_i and number density, n, e.g., $\kappa_{L,\ min}$ for Si is ~1 W/mK [100]. The above model has been justified from the fact that values below $\kappa_{L,\ min}$ have not yet been clearly discerned. The notion of $\kappa_{L,\ min}$ is very useful from the point of view of designing thermoelectrics for an increased figure of merit, ZT. For example, if it is definitely understood that values below $\kappa_{L,\ min}$ are impossible, then engineers should look to enhancement of the power factor as the only option. This is especially relevant as κ_L values very close to the theoretical minimum have been observed recently in silicon nanowires [13].

However, the presence of nanoscale structures (i.e., nanoparticles, nanorods, etc.) of size comparable to or smaller than l could imply that a temperature gradient cannot be well defined across the nanostructure and necessitate a modification of basic Fourier heat conduction theory [101,102]. When the gradient cannot be well defined [103], the heat flux (Q) between two temperatures, T_1 and T_2 and across an area (A) should be considered in terms of radiative transport [104], through a Stefan–Boltzmann type relationship of the form: $Q=A\sigma_B(T_1^4-T_2^4)$, where σ_B is a constant ($=(\pi^2/120)(k_B^4/\hbar^3)\sum_i 1/c_i^2$). The thermal conductivity is then derived from the concomitant equation of phonon radiative transport (EPRT) [104], where the l for determining κ_L is replaced by an effective mean free path, $l_{eff}=l/(1+(4l/3d))$, whereby a reduced κ_L is predicted for an increasing ratio of l to the sample size (d). The experimental verification of this formula would be useful in verifying whether $\kappa_{L,\ min}$ is indeed an absolute lower limit.

On the other hand, the upper limit of λ_{ph} could correspond to the sample size, d, which, for large d, can be related to the continuum limit and the center of the Brillouin zone. Other length scales involved in phonon transport include the coherence length (ξ_{coh}), below which phonon wave interference effects must be considered and above which diffusive, particle-like, motion is dominant [105]. It has been estimated that ξ_{coh} is ~1 nm for the transverse phonons and ~2 nm for the longitudinal phonons in bulk GaAs [106]. Consequently, when $l<\xi_{\mathrm{coh}}<d$, phonons can be treated as particles and when $l<d<\xi_{\mathrm{coh}}$, enhanced phonon wave interference in the sample could induce the onset of localization leading to phonon non-propagation and reduced thermal transport. Alternately, phonon localization occurs when the λ_{ph} is comparable to l, i.e., $\mathbf{k}\cdot l\sim 2\pi$. A wave to particle crossover was indeed observed [107] computationally and seen to correspond to a reduced κ_L. The treatment of phonons as waves also brings forth the possibility of trying to understand the minimum thermal conductance ($\mathbf{K}=\kappa_LA/L$, for heat transport through a length L and cross-sectional area A) due to phonons. This would help in understanding the smallest quantity of heat transport possible in a material and may provide additional insight into the lowest possible κ_L. Such considerations then lead to a comprehension of thermal transport through a quantum mechanical point of view.

3.3. The thermal conductance quantum

Hitherto, classical considerations through kinetic theory and the Boltzmann transport equation have been used so as to define κ_L as a macroscopically determined quantity. One can also take an alternate point of view where the total thermal conductivity/conductance in a sample is the sum of the contribution from several (say, M) discrete phonon modes. The maximum amount of heat that can be transported by *any one* given mode, at a certain temperature, T, is then represented [108] by the quantum of thermal conductance: g_o. The overall thermal conductance of a material would then be proportional to Mg_o. Considering the propagation of phonons (of energy $\hbar\omega$) along a one-dimensional wire, for i longitudinal and transverse branches, the total energy flux is represented through [108]:

$$J_{th}=\sum_i\int_0^\infty\frac{dk}{2\pi}\hbar\omega n(\omega)v \tag{43}$$

The thermal conductance, G_{th} is then defined as the ratio of J_{th} to ΔT and is given by:

$$G_{th}=\sum_i\frac{k_B^2T}{h}\int_{x_m}^\infty\frac{x^2e^x}{(e^x-1)^2}dx,\quad\text{where}\quad x=\frac{\hbar\omega}{k_BT} \tag{44}$$

For the modes of the smallest energy, i.e., $\hbar\omega(k=0)=0$, and with $k_BT<\hbar\omega$, $G_{th}=g_o=\pi^2k_B^2T/3h=9.456\times10^{-13}\ \mathrm{W/K^2}$ (T). The g_o then represents the maximum possible value of the thermal conductance for a given mode at a given temperature [109]. Such a value for g_o has indeed been experimentally determined [110]

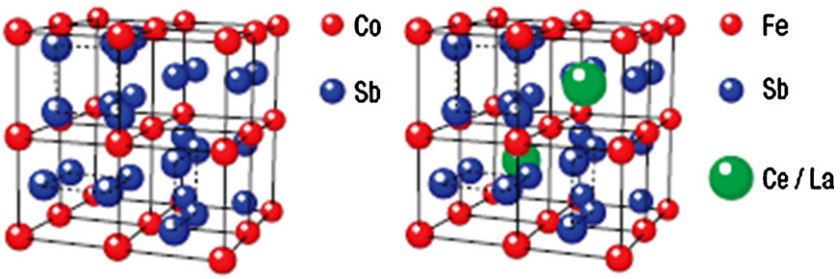

Fig. 32. Skutterudite crystal structures (a) $CoSb_3$, and (b) La/Ce doped $FeSb_3$ compounds (reprint permission from [118]).

under carefully controlled geometries in one-dimensional, ballistic, phonon channels.

3.4. Principles and experimental implementation of nanostructures with reduced κ_L

One of the goals of the research in nanostructured thermoelectrics, is to obtain a κ_L less than that in bulk thermoelectric materials. Such a reduction does seem to have been achieved in many cases with $\kappa_L \sim 1$ W/mK, and has contributed to the spurt in the figure of merit. We will initially discuss general principles commonly relevant to a decrease of κ_L.

3.4.1. Obtaining reduced thermal conductivity in bulk materials

Since κ_L is a property of the lattice, efforts have been made to relate it to other lattice based characteristic properties such as the compressibility (χ), thermal expansion coefficient, melting point (T_m), density (ρ), and the atomic weight (A) of the constituent atoms [111]. The Keyes' relationship, where $\kappa_L T = B(T_m^{3/2}\rho^{2/3}/A^{7/6})$, encompasses the mutual relationship between these various properties through a unified empirically derived expression and seems to be in remarkable agreement with experimental values. The constant of proportionality, B, is related to the Grüneisen parameter, which is a measure of the change in the phonon dispersion spectrum for a given dilatation and is inversely related to the χ [112]. The reader should refer to the original papers by Lawson [112], Dugdale and Macdonald [111], and Keyes [113] for the underlying assumptions. The Keyes' relationship translates to the physical insight that soft materials (low density – ρ) with a high atomic weight (A), and low melting points (T_m) have a lower κ_L. It can also be deduced that κ_L is smaller for ionic bonded materials (lower T_m) compared to covalently bonded structures. The presence of charged ions was expected to yield additional scattering mechanisms through the positive and negative ion sub-lattices contributing to additional optical modes, and it is observed that κ_L decreases with increasing cation/anion mass ratio.

On the basis that low density materials, with relatively open crystal structures, could have lower κ_L, skutterudite thermoelectrics [114], e.g., $IrSb_3$, $CoSb_3$, $RhSb_3$ based compounds represented in Fig. 32, were investigated [115]. In these, loosely bound metal *guest* atoms can be interspersed into the lattice and act as *rattlers* perturbing phonon and thermal energy transport. Consequently, the phonon scattering is enhanced at the atomic level while the periodicity of the native lattice is unchanged. The material is now *glass like* with respect to phonon transport while *crystalline* with respect to electrical carrier transport, with a maximal interaction and resonance scattering of the propagating acoustic phonons when the phonon frequency equals the rattler frequency.

Such an idealization, in terms of a phonon glass electron crystal (PGEC) [116] can also be applied to other *open structured* compounds such as clathrates [117], and defect structured compounds [99]. It is important to realize that such modulations of phonon transport cannot simply be produced by simple monatomic substitutions but require "random, non-central distortions of the lattice" [99]. Other enumerated criteria [6], for PGEC behavior, include: (1) The *rattler* atoms should be uncorrelated with each other and the lattice, with no well defined position, i.e., no long range order and phase coherence with respect to the vibrational motion [118], and (2) the total mass of the *rattler* atoms should be at least 3% of the total mass of the crystal.

Experimental evidence for the PGEC hypothesis has been obtained through the measurement of κ_L, specific heat, and electrical resistivity of the Tl filled skutterudites [119] (e.g., $Tl_xCo_{4-y}Fe_ySb_{12}$ and $Tl_xCo_{4-y}Sn_ySb_{12-y}$). A decrease in the κ_L by a factor of 5 (to ~1 W/mK) through the addition of Tl is shown in Fig. 33. Specific heat measurements [120] indicated that the Tl atoms behave as independent harmonic oscillators [121]. The phonon mean free path (l) would then correspond to the average distance, between individual Tl atoms, which is proportional to $x^{1/3}$. Crystallographic analysis and powder diffraction measurements also showed that Tl could fill as much as 80% of the voids in $CoSb_3$, as was verified through a large atomic displacement parameter (ADP) of 0.05 Å^2. The ADP, which measures the mean square displacement amplitude of an atom about its mean position, is approximately an order of magnitude larger than that accounted for by nominal static disorder, and hence supports the rattler hypothesis.

However, recent neutron spectroscopy studies [118] on La- and Ce-filled Fe_4Sb_{12} skutterudites (Fig. 34) indicated that a quasi-harmonic coupling actually does exist between the inserted *rattler* and the host lattice. This implies that the complexity of the filled crystals (i.e., through the *CF* factor introduced in Section 3.2) and the preponderance of low group velocity optical phonon modes, rather

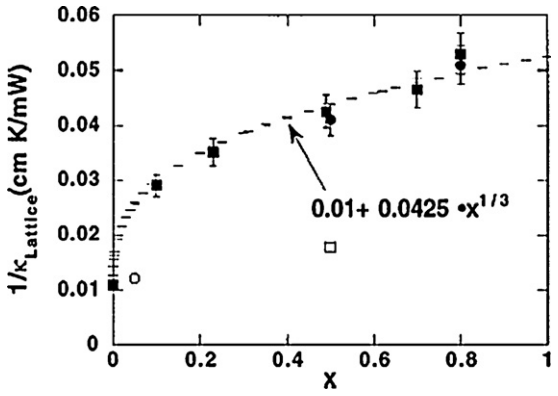

Fig. 33. The lattice thermal conductivity ($\kappa_{Lattice}$) decreases (i.e., $1/\kappa_{Lattice}$ increases) with addition of Tl (x, in $Tl_xCo_{4-y}Sn_ySb_{12-y}$). The dependence on x follows from the average spacing of the Tl atoms (reprint permission from [119]).

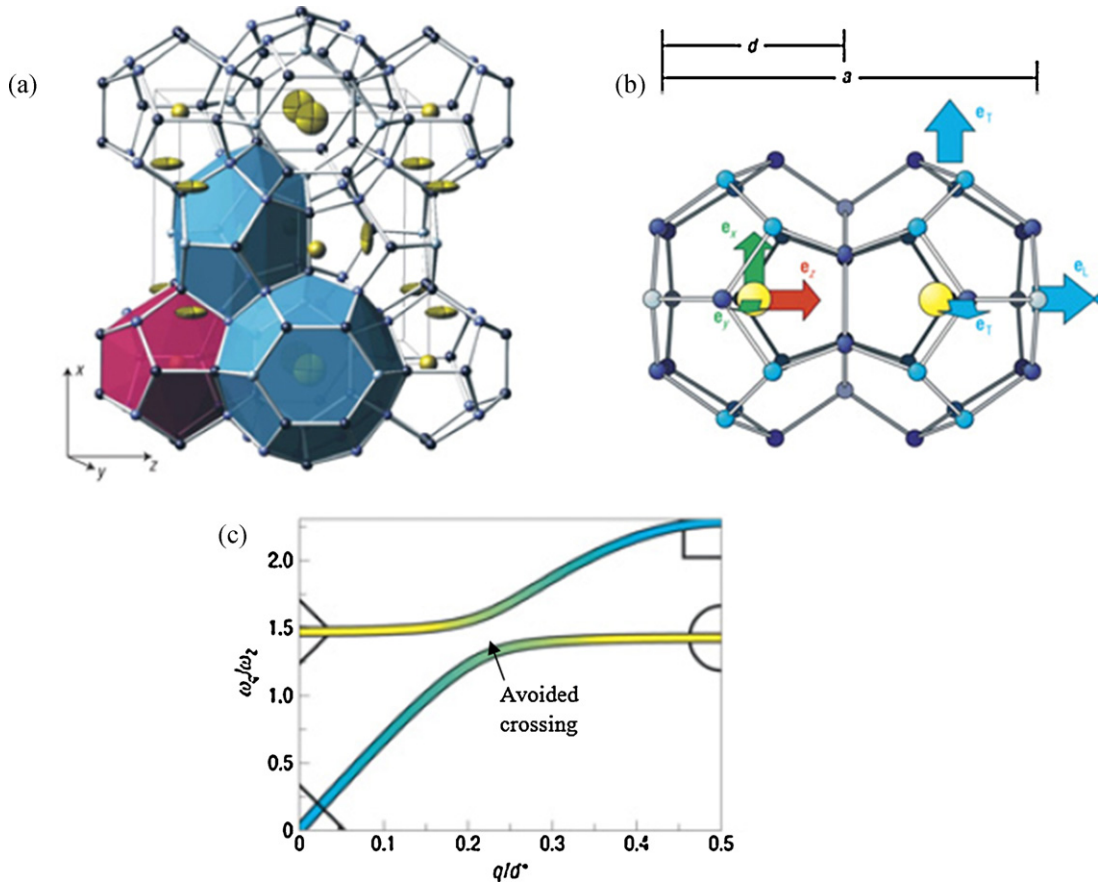

Fig. 34. (a) Illustration of a typical open cage like clathrate structure ($Ba_8Ga_{16}Ge_{30}$) with guest (rattler atom: Ba)–host (cage framework: Ga and Ge atoms) interactions. (b) The rattling motion of the guest can be decomposed into longitudinal (e_L) and transverse (e_T) modes. (c) The interaction of the guest mode phonon dispersion with the host mode is postulated to result in an avoided band crossing, at which the phonon lifetimes are increased (reprint permission from [122]).

than the rattling, could be more responsible for the reduced κ_L. Additional evidence for such coupled modes arises from a study [122] of clathrate thermoelectrics, such as $Ba_8Ga_{16}Ge_{30}$ (Fig. 34), where it was estimated through neutron triple axis spectroscopy that the phonon lifetimes are ~2 ps, an order of magnitude larger than the 0.18 ps expected from inelastic scattering of the acoustic phonons with the rattler atoms. Such enhanced lifetime was postulated as due to the absence of scattering brought about by an avoided band crossing between the acoustic phonon mode of the host lattice and a flat, quasi-localized, low energy mode due to the rattler atom.

In summary, it is probable that the crystal complexity, flat phonon dispersion, and dynamical Umklapp processes are all important for the reduced κ_L and there are some suggestions that the PGEC type of mechanism may have to be re-evaluated.

3.4.2. Embedding nanostructures in bulk materials: nanocomposite thermoelectrics

In this section, we will review how nanoscale features could contribute to the reduction of κ_L. While lower dimensional thermoelectrics, such as superlattices and nanowires, do promise a large reduction in κ_L – mainly through a reduction of the mean free path, l, problems in large scale and reliable synthesis preclude their wide-scale use. Bulk materials with embedded nanoscale features may be more immediately suitable for practical application.

It was mentioned earlier that the radiative transport of phonons across nanostructures with length scales smaller than l can considerably reduce the κ_L. Preliminary investigations of thermal transport in nanocomposite materials (with nanoparticles embedded in a matrix) through Monte–Carlo simulations [123] did indicate a reduction of κ_L to below that of the matrix. However, these simulations used bulk phonon dispersion parameters and did not consider phonon polarization [124] or phonon dispersion and also assumed an l independent of phonon frequency – a *gray-media* approximation [125]. While such an approximation places less computational demands in comparison to a full-fledged frequency dependent modeling of l, the agreement with experiment is poor [106]. With the above caveats, it was then shown that the orientation and shape of the nanoparticles (Fig. 35) modulates the phonon transport and reduces κ_L. For example, at a given alloy composition, a random orientation with a staggered distribution of the nanoparticles was modeled to be more beneficial for minimizing κ_L. However, the degree of randomness and the effects of interfaces still need to be considered and quantified.

It has been experimentally shown that embedding nanostructural features, through suitable materials processing techniques, in conventional bulk thermoelectric materials such as PbTe or Bi_2Te_3, could reduce κ_L and improve ZT. The nanoscale features may be thought to mimic lower dimensional structures, e.g., zero-dimensional quantum dots and thin film structures. A few examples of this approach will be discussed.

A seminal effort in reducing the κ_L was implemented in the class of LAST (for Lead, Antimony, Silver and Tellurium) alloys of composition $AgPb_mSb_nTe_{2+m}$ [126]. These materials satisfy many of the attributes of good thermoelectrics including (i) complex crystal structure – Fig. 36(a), (ii) *low intrinsic* κ_L (~1.3 W/mK) – Fig. 36(b), and (iii) ability to tune electrical carrier concentrations and electrical conductivity (σ) over a large range, through the

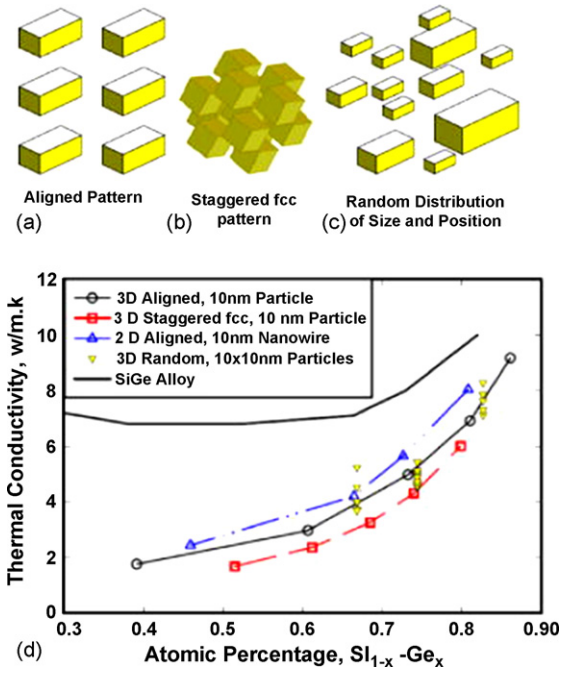

Fig. 35. (a) Nanocomposites constituted of Si particles arranged in various configurations in a Ge matrix [123]. (b) A reduction in the thermal conductivity below that of the bulk materials, as a function of the arrangements in (a) was calculated through Monte–Carlo simulations (reprint permission from [123]).

generation of a large number of compositions (of different m and n). The crystal structure of a LAST alloy is based on the NaCl lattice, where the Ag, Pb, and Sb atoms are spread over the "Na" sites while the Te atoms are arranged over the "Cl" sites. The isoelectronic substitution of Pb^{2+} ions for the Ag^{+} and Sb^{3+} are thought to generate structural and electronic distortions which could both reduce the κ_L and enhance the Seebeck coefficient. Compositional modulations arising from the proximity and high concentrations of the cations were observed in these alloys as illustrated in Fig. 36(c) and (d). Such morphology was thought to mimic features of two- and quasi-zero-dimensional nanostructures, respectively.

The nucleation of such nanostructural features can also be facilitated through spinodal decomposition [127] in $Pb_{1-x}Sn_xTe$-based systems, where Sn was added to enhance the σ. As an example, the spontaneous separation of PbS and PbTe phases can cause structural modulations around 2 nm in wavelength (Fig. 37(a)) in $(Pb_{0.95}Sn_{0.05}Te)_{1-x}(PbS)_x$ ($x = 0.04$–0.3) mixtures. Additionally, nucleation and growth of nanocrystals, 3–10 nm in size, were also seen at lower values of x (Fig. 37(b)). High resolution electron microscopy investigations have indicated that while the nucleated nanocrystals/nanodots have an intrinsic lattice parameter mismatch (~2–5%) with the surrounding matrix [128], they are endotaxially oriented. Endotaxy refers to a process of crystallographically oriented segregation of a phase, e.g., Ag–Sb rich nanocrystal, relative to Ag–Sb poor matrix – predominantly PbTe, and could occur due to the disintegration of the original native solid solution [129]. Dislocations at the nanocrystal–matrix interface and the large number of interfaces serve to impede phonon transmission and reduce thermal conductivity. On the other hand, the σ is negligibly affected at relatively high doping levels. Similar features with reduced κ_L, which was measured through a flash diffusivity method, were seen in other systems such as [130] $AgPb_mSb_nTe_{2+m}$ solid solutions ($\kappa_L \sim 0.43$ W/mK yielding a ZT of 1.45, at 620 K), $Pb_{9.6}Sb_{0.2}Te_{10-x}Se_x$ bulk materials [131] ($\kappa_L \sim 0.4$ W/mK and a ZT of 1.2, at 650 K), and in Ag free $Na_{1-x}Pb_mSb_yTe_{m+2}$ [132] ($\kappa_L \sim 0.85$ W/mK and a ZT of 1.7, at

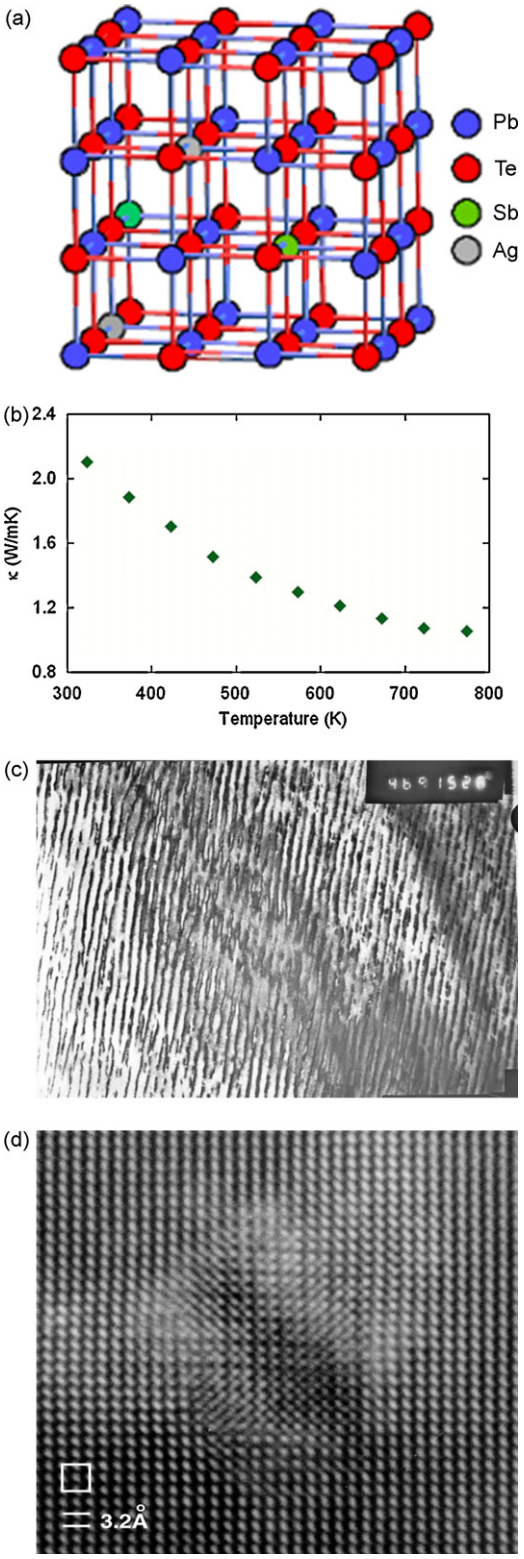

Fig. 36. (a) Crystal structure of $AgPb_mSbTe_{2+m}$. (b) The variation of the thermal conductivity (κ) with temperature. (c) The compositional modulation due to ion (Pb^{2+}, Ag^{+}, Sb^{3+}) induced distortions result in structural modifications at the 20 nm scale. The layered periodicity may be thought to be akin to superlattice morphology. (d) High resolution TEM image of an AgSb nanodot coherently attached to a PbTe matrix. Such nanoscale features would be important for a reduced thermal conductivity, which were measured to be of the order of 1–2 W/mK, while the electrical conductivity is unchanged due to the coherent attachment (reprint permission from [126]).

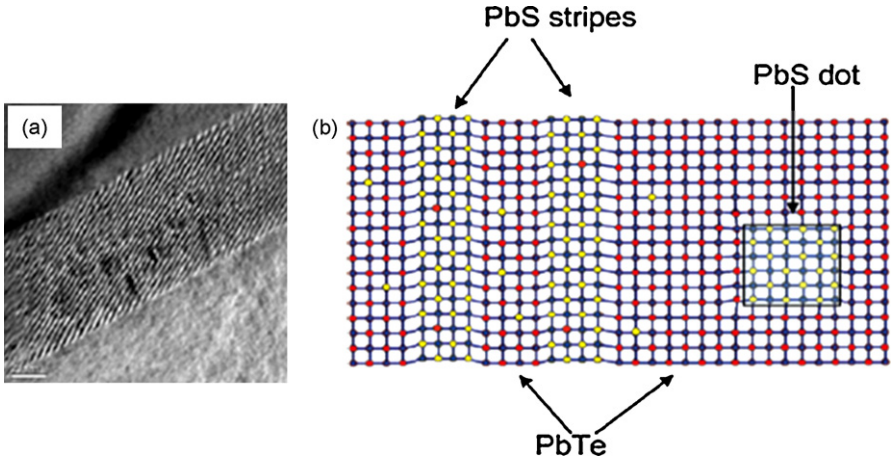

Fig. 37. (a) Spinodal decomposition of $(Pb_{0.95}Sn_{0.05}Te)_{0.84}(PbS)_{0.16}$ solid solutions into PbS (bright) and PbTe (dark) phases. (b) Nanoscale features at the one-dimensional (PbS stripes) and quasi-zero-dimensional level (PbS dots) can be induced to form through phase transformations, such as spinodal decomposition (reprint permission from [17]).

650 K). Increasing [17] the PbS content in the range of $0 < x < 0.3$, for $(Pb_{0.95}Sn_{0.05}Te)_{1-x}(PbS)_x$ enhanced the one-dimensional spatial modulations at the expense of the nanocrystals' nucleation. The consequence was a 3-*fold reduction* in the κ_L compared to the unmodified solid solutions. Lower values of κ_L were also observed for the $x = 0.08$ (~0.38 W/mK) and $x = 0.16$ (~0.4 W/mK) specimens compared to the $x = 0.3$ (~1 W/mK) sample, which implies that zero-dimensional features are *more effective* in reducing the κ_L compared to the one-dimensionally oriented stripes. Combined with a $\sigma > 30\ \Omega^{-1}\ cm^{-1}$ and S (~−220 μV/K), a ZT of ~1.5 was estimated at 642 K, for $x = 0.08$.

The introduction of nanostructural [11] morphologies in bulk p-type $Bi_xSb_{2-x}Te_3$ thermoelectrics also seemed to be successful in reducing the κ_L to ~1.1 W/mK from ~1.4 W/mK for a bulk ingot without the nanoscale features, and concomitantly increasing the ZT from a value of 1 to 1.2 (300 K) and 1.4 (373 K). In this study, the composites were synthesized from ball milling powders from $Bi_{0.5}Sb_{1.5}Te_3$ alloy ingots and subsequent hot pressing. The microstructure then consisted of grains and Te precipitates ~10–50 nm in size, with embedded Sb rich nanodots ~2–10 nm. Consequently, the phonon scattering at various length scales in these "nano-bulk" materials may contribute to the reduced κ_L. However, the σ is surprisingly enhanced, for which the effect of the grain boundaries and interfaces in these materials on electrical conduction would have to be considered.

The enhanced σ could be explained on the hypothesis that grain boundaries are devoid of charge and act as Schottky barriers [133] (Fig. 38). At equilibrium, as the Fermi energy (E_F) is constant through the material, the grain boundary serves as a source of electrical holes which are then injected into the grains and hence enhance the σ. Consequently, grain boundaries which could serve as interfaces for enhancing phonon scattering could also influence

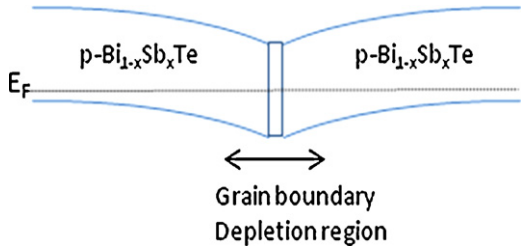

Fig. 38. A depletion region, devoid of holes at the interface of two p-$Bi_{1-x}Sb_xTe$ grains. This has the consequence of increasing the carrier conductivity in the grains while lowering the thermal conductivity.

the σ. Such hypotheses could be confirmed through materials processing to tune the grain structure and subsequently through a study of the relative magnitudes of κ_L and σ with respect to optimization of thermoelectric properties.

3.4.3. The influence of interfaces

At length scales below/approaching the phonon mean free path, l, the thermal boundary resistance (TBR) due to interfaces could limit the thermal conductivity/conductance [134]. An additional thermal resistance per unit area, R_{TBR} is now added to the intrinsic resistance, R_{int} (=L/κ_L) and the net thermal resistance, R_{net} (=L/κ_{net}) = $R_{TBR} + R_{int}$. The heat transport across the interface of two materials could be significantly affected by the TBR which mainly arises due to the reflection, transmission, or absorption of phonons at the interface. The differing phonon distributions on either side results in a temperature drop across the interface with an associated thermal resistance given by the ratio of the temperature drop to the transmitted heat flux. Such phenomena were first discussed in detail for solid–liquid helium interfaces which were posited to have an associated Kapitza resistance [135]. The corresponding Kapitza resistance/TBR between materials was then found to depend on the ratio of the materials' Debye temperatures [136]. It was proposed that a ratio close to one, e.g., ~1.7 for a Si/Ge interface could result in a closer match of the phonon wavelengths and enhanced phonon transmission, compared to when the ratio was much different than unity, e.g., 0.05 for a Pb/diamond interface.

In this regard, two basic models, i.e., (1) the acoustic mismatch model (AMM) and, (2) the diffuse mismatch model (DMM) have been used [137] to understand the thermal conduction across an interface/boundary. In the AMM, the constituents on either side of the interface are treated as a continuum, each with an acoustic impedance, A (=ρv), where ρ is the bulk material density and v the associated acoustic velocity. The heat flux across the interface due to longitudinal phonon waves, from medium 1 to medium 2, is then determined [138] by the product of the incident number of phonons and the transmission probability, $T_{1\to2}$. For an angle of incidence θ_i and angle of refraction θ_t, it can be derived [138] that $T_{1\to2} = 4(A_2/A_1)(\cos\theta_t/\cos\theta_i)/((A_2/A_1) + (\cos\theta_t/\cos\theta_i))^2$. It can then be inferred that the maximum transfer of heat occurs when the two materials have similar values of the mass density *and* specific heat [138]. For layers with identical acoustic impedances, i.e., $A_1 = A_2$, the AMM predicts for normal phonon incidence ($\theta_i = \theta_t = 0^\circ$) a $T_{1\to2}$ of unity. The wave-like propagation also implies a critical angle (θ_c) above which all the incident phonons are totally internally reflected and do not contribute to the heat transfer yielding a reduced TBR. Due to the materials being modeled as

continua, the detailed nature of the interface is ignored in the AMM approximation, i.e., the $\lambda_{ph} >$ interface roughness. However, at increasing phonon energy and decreasing λ_{ph}, the l could be comparable to the scale of surface roughness and imperfections and the AMM would be inappropriate [137,138]. Alternately, the *diffuse mismatch model* (DMM) implicitly takes into account the details of the interface by considering individual phonon traversal. The phonons impinging on the interface lose their original state (direction, polarization, etc.) subsequent to scattering. The transmission of any phonon is now determined by whether there is a corresponding phonon, on the other side, of the same energy to which scattering can occur, i.e., by the phonon density of states $g(\omega)$. While the phonon dispersion is taken into account, *all* the incident phonons now scatter elastically from the interface with the $T_{1\rightarrow 2}$ determined by the mismatch of the phonon DOS on either side. For both longitudinal and/or transverse mode phonons (given through the index j) the transmission is given by the ratio of the in- and out-fluxes (product of the phonon velocities and number), as $T_{1\rightarrow 2}(\omega) = \sum_j \nu_{2,j} g_{2,j}(\omega, T) / \sum_j \nu_{1,j} g_{1,j}(\omega, T)$.

Both the AMM and the DMM have been applied to simulate experimental situations, but their simplified formulation generally precludes universal agreement [137]. For example, in SiO_2 and SiN_x films deposited on Si substrates, the DMM was shown to yield good agreement at low temperatures (<20 K) while it differed from experimental observations by an order of magnitude at room temperature [139]. Typically the DMM, which considers a greater number of phonons elastically scattering would be thought to predict a higher thermal conductance/conductivity compared to the AMM [140]. Indeed, the highest K, corresponding to the phonon radiation limit, is manifested through the DMM [137]. However, inelastic scattering processes due to point defects [141], interfacial roughness, tunneling, excitation of surface phonons, phonon down-conversion [142], influence of electrical carriers, etc., not considered in either model make detailed prediction difficult. Additionally, strain at the interface, which in a coherent superlattice would increase with the number of interfaces, has not been considered in detail. In an attempt to understand the interface scattering in more detail, molecular dynamics (MD) simulation, which serve to check the diffusive approximations inherent in the Boltzmann transport formalism, have been attempted [143]. In MD, mutual atom–atom interactions modeled through Lennard–Jones [88,143] or other inter-atomic potentials [144], are considered and the heat flux computed from the product of the atomic forces and velocities. MD simulations have indicated two orders of magnitude reduction of thermal conductivity in superlattices [145] and nanowires [144] compared to the bulk. The reduction seems to be quite dependent on the boundary scattering and specularity (see Section 3.4.4 for a definition) of the surface.

The TBR has also been invoked in modeling the reduction of the κ_L with decreasing grain size of polycrystalline materials, through the assignment of a Kapitza thermal resistance to the grain boundaries [146], i.e., R_{gbdy} (=R_{TBR}). The net thermal conductance can then be calculated through assuming two parallel conduction processes through the (1) grain, and (2) grain boundary, respectively. It is then expected that decreasing grain size in addition to implying a reduced l, increases the R_{gbdy} contribution. Indeed, the lack of a direct proportionality between the grain size and the measured thermal resistance [147]/conductivity, κ_{net} may be taken as evidence of R_{gbdy}, as borne out in studies on 0.1–0.7 μm grain sized $CoSb_3$ [148]. On the other hand, TBR models not considering mean free path effects and assigning a Kapitza resistance to the grain itself are not well understood [148,149].

3.4.4. Thermal conductivity reduction in superlattices

A reduced l in nanostructured thermoelectrics such as thin films and nanowires provides a straightforward motivation for a reduced κ_L. However, the κ_L values are anisotropic as were seen through thermal diffusivity measurements [150] on GaAs/AlGaAs thin film superlattice structures, where a smaller l in the superlattice growth direction implied a smaller value of κ_L in that direction compared to the value in the plane of the film. A number of more direct experiments probing in-plane and cross-plane thermal conductivity have also been carried out in GaAs/$Al_{1-x}Ga_xAs$-based superlattices. While these material systems are not considered favorably for thermoelectrics, due to intrinsically low figures of merit, the motivation was probably the superior interface quality of lattice matched GaAs and AlGaAs [151,152] fabricated through MBE and MOCVD. It was seen that the κ_L of AlAs/GaAs superlattices [153], comprised of an equal thickness of AlAs and GaAs indeed decreases with a decreasing layer thickness *and* superlattice period (d). However, the measured κ_L was *greater* than that of $Al_{0.5}Ga_{0.5}As$ bulk alloy – it was noted that the κ_L of $Al_{0.5}Ga_{0.5}As$ is smaller compared to that of the AlAs and GaAs individually. The κ_L of the superlattices seemed to be lowered to that of the alloy only in the limit of monolayer thick AlAs and GaAs, from which it can be surmised that alloy scattering is possibly the stronger κ_L suppression mechanism. In two samples of GaAs/AlAs monolayer (ML) superlattices [154], viz., (i) $[GaAs\ (3\ ML)]_{159}/[AlAs\ (3\ ML)]_{159}$, and (ii) $[GaAs\ (12\ ML)]_{27}/[AlAs\ (14\ ML)]_{27}$, κ_L values of 3.1 and 5.8 W/mK were measured, using a picosecond optical technique, which are smaller in comparison to the bulk GaAs value of ~46 W/mK. The lower κ_L for sample (i) compared to (ii) seems to indicate that a larger number of interfaces could also be important.

A decrease in κ_L to ~9.5 W/mK (the value for the bulk alloy) with increased number of interfaces, was measured by the 3ω technique [155] in $Si/Si_{0.7}Ge_{0.3}$ superlattices by Huxtable et al. [156]. Concomitantly, a decrease in the d, which is related to an increased number of interfaces/unit length, decreased the κ_L. In measurements on $Si_{0.84}Ge_{0.16}/Si_{0.74}Ge_{0.26}$ superlattices, the variation of κ_L with d was not as drastic. Such a variation was explained by the authors as due to the smaller acoustic impedance mismatch, which was larger (~1.15) for the $Si/Si_{0.7}Ge_{0.3}$ superlattices compared to the alloy superlattice (~1.03). However, a value below that of the bulk alloy κ_L and above that of amorphous Si was measured [140] for Si_m–Ge_n superlattices (m and n refer to the number of monolayers), where for $3\ nm < d < 7\ nm$, the κ_L decreases with decreasing d. Surprisingly, measurements on structures with $d > 13$ nm indicated an even smaller κ_L. While these observations may rule out the role of *mini*-Umklapp scattering (explained later in this section, see Fig. 41) in modulating the thermal conductivity, detailed information on the number and nature of interfaces for the presented data was not elucidated in the latter study. Given that the critical thickness for the pseudomorphic growth [157] of Ge on Si is ~1–3 nm, attention has to be paid to the possibility of morphology variations, which can introduce additional phonon scattering from the large concentrations of static imperfections [158] such as dislocations (estimated at $\sim 10^{11}/cm^2$).

In a study on symmetrically strained Si (2 nm)/Ge (2 nm) superlattices by Borca-Tasciuc et al. [159] values of κ_L in the range of 2.9–4.0 W/mK were obtained, through the 3ω technique, for the cross-plane κ_L, which are *below* the values (~8.8 W/mK at 300 K) corresponding to equivalent composition bulk $Si_{0.5}Ge_{0.5}$ alloys. These samples were synthesized by MBE on a 1–2 μm thick buffer layer with a lattice parameter equal to the average lattice parameters of Si and Ge, to compensate the strain in the grown superlattice. Transmission electron microscopy examination revealed dislocation densities of $\sim 10^4/cm^2$, which are too low to significantly influence the κ_L. Additionally, a κ_L reduction was seen with increased carrier concentration ($\sim 10^{18}$ to $10^{19}/cm^3$) in p- and n-doped superlattices and with increased period, i.e., at 300 K, the κ_L

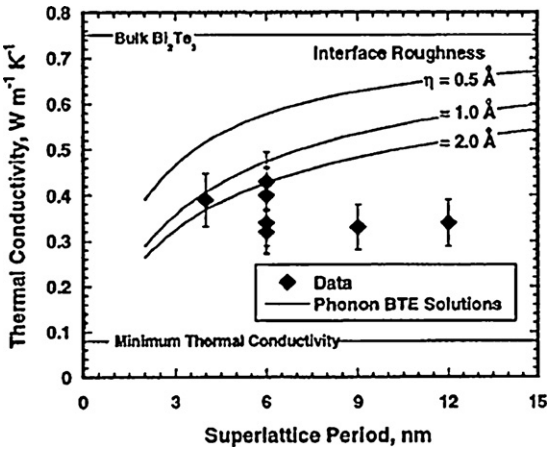

Fig. 39. The measured value of κ_L for Bi_2Te_3/Sb_2Te_3 superlattices is less than that of bulk Bi_2Te_3 and higher than $\kappa_{L,min}$ minimum thermal conductivity. The mean interface roughness (η) can be used to model the reduced κ_L for small superlattice period, d (reprint permission from [161]).

decreases from 4 W/mK at $d = 4$ nm to ~3 W/mK at 9.4 nm and ~2.9 W/mK at 14 nm. However, the increase of the κ_L with increased doping concentration, in n-doped samples, is unexpected and such a variation along with the small dependence of κ_L on $d > 9$ nm demands more critical study.

3ω measurements [160] for the cross-plane κ_L on both Bi_2Te_3 (1 nm)/Sb_2Te_3 (5 nm) and Bi_2Te_3 (3 nm)/Sb_2Te_3 (3 nm) superlattices indicated a κ_L of ~0.25 W/mK. A reduction of κ_L to ~0.3 W/mK, below the value of ~0.75 W/mK for bulk Bi_2Te_3 was also estimated [161] from thermoreflectance measurements (Fig. 39) for Bi_2Te_3/Sb_2Te_3 superlattices, with a pronounced decrease for $d < 6$ nm. The effects of the mean interface roughness (η) on a diminishing κ_L, due to additional scattering of phonons with the concomitant wavelengths seem to be valid for smaller d. It has been postulated [162] through Monte–Carlo simulations, that random roughness can cause phonon backscattering and suppress κ_L. The effective mean path (l_{eff}) in this case is given by: $l_{\text{eff}} = ((1+p)/(1-p))d$, where p is a specularity parameter $p = \exp((-16\pi^3/\lambda_{ph}^2)\eta^2)$ with a Gaussian distribution for the surface roughness [90]. $p = 1$ with purely specular scattering while $p = 0$ represents purely diffusive scattering and yields the smallest κ_L.

The effect of zero-dimensional features on reducing the κ_L can be tested through the formation of quantum dots (QD) which could be made by alternate deposition of two thin films of different lattice parameter. When a critical thickness [157] for maintaining registry of a thin film with an underlying layer is exceeded, the strain energy minimization would induce the spontaneous break down of the overlaid film into randomly distributed islands or QDs. Such QD formation was shown in thermoelectric PbSe/PbTe superlattices, due to the ~5% lattice parameter mismatch [163] of PbSe with PbTe (Fig. 40) where measurements were carried out on a 0.1 mm thick sample constituted of 8000 periods of 13 nm thick $PbTe/PbSe_{0.98}Te_{0.02}$ QDs.

A low value of $\kappa_L \sim 0.33$ W/mK, accrued through a large interface density and enhanced scattering from both the thin film and QD interfaces, was estimated after subtracting the electronic contribution to the thermal conductivity by invoking the Wiedemann–Franz relationship. Consequently, the observed increase in the figure of merit (with a measured ZT of 1.6 at 300 K and 3.5 at 570 K) for the QD incorporated PbSe/PbTe superlattices was mainly attributed to the reduced κ_L. Even lower κ_L values could be harnessed through using quaternary alloys, such as lattice mismatched $PbTe/Pb_{1-x}Sn_xSe_yTe_{1-y}$ QD superlattice structures, where the addition of Sn was used to enhance the σ and the S, through a larger DOS effective mass (Section 2.4).

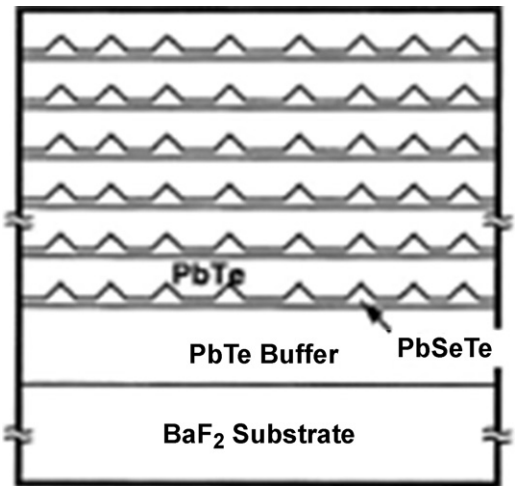

Fig. 40. Zero-dimensional structures (quantum dots) can be in situ synthesized on thin film structures through lattice mismatch and strain induced effects. The enhanced interface density and dimensional confinement can be used to lower the thermal conductivity and increase the effective Seebeck coefficient (reprint permission from [163]).

We will next consider the possibilities of engineering the phonon dispersion in superlattices aimed at modulating/reducing the κ_L. The main ideas involve (i) increased phonon interactions leading to phononic bandgaps, which serve to modulate phonon and thermal transport, and (ii) increasing the number of phonon modes and flattening the phonon energy dispersion, which as seen from Section 3.2, reduces the overall group velocity.

In a superlattice, composed of thin films with spacing d (=na, where n is an integer and a is the spacing of the lattice points) a new periodicity [152] corresponding to d is introduced, in the direction perpendicular to the growth. Translated to reciprocal space, the imposed periodicity intersperses a *mini*-Brillouin zone of extent $2\pi/d$ into the original Brillouin zone of span $2\pi/a$. A larger (/smaller) d corresponds to smaller (/wider) *mini*-Brillouin zones, as

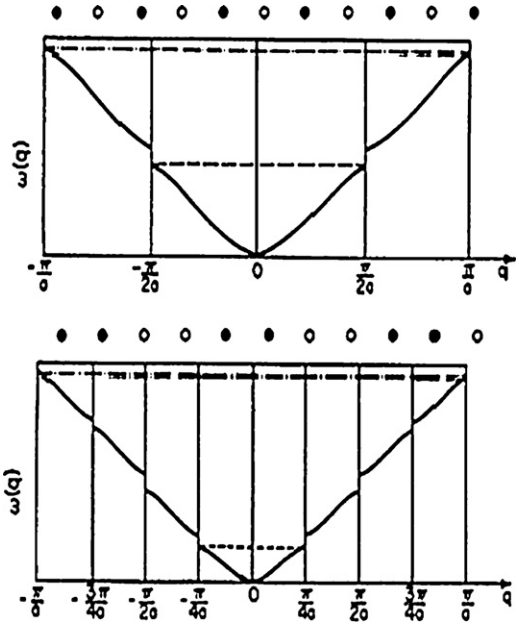

Fig. 41. The occurrence of mini-Brillouin zones in (a) 1×1 and (b) 2×2 superlattices. The dash-dotted line represents the reciprocal lattice wave vector associated with the underlying crystalline lattice while the dashed line depicts the wave vector associated with the superlattice period ($d = 2a$, and $d = 2a$) (reprint permission from [267]).

illustrated in Fig. 41 for a one-dimensional distribution of lattice points in a superlattice [164] corresponding to (a) one unit cell/layer, and (b) two unit cells/layer. The periodic interruption of the band structure with a flattening of the bands/energy dispersion at the edges of the *mini*-Brillouin zones reduces the phonon group velocity and concomitantly the κ_L. Additionally, enhanced scattering through *mini-Umklapp* processes [90] made possible through a smaller backscattering mini-reciprocal lattice vector could also contribute to increased thermal resistance. For a superlattice of two constituents with masses m_A and m_B, it was predicted that κ_L could be lowered by as much as 25% for a $\Delta m/m_{avg} = 1$, where $\Delta m = m_A - m_B$, and m_{avg} is the mass average.

While phonons may be treated as particles, for $d > l$, wave propagation must be considered when $d < l$. In the latter case, the interaction of phonons (of wavelength λ_{ph}), of arbitrary propagation direction and polarization, with a superlattice could be governed through a form of Bragg's law:

$$n\lambda_{ph} = 2d\cos\theta_k \tag{45}$$

θ_k is the angle between the superlattice growth direction and the phonon propagation direction, e.g., when a phonon travels perpendicular to an atomically perfect interface, complete backscattering due to constructive interference [165] occurs when $\lambda_{ph} = 2d$.Such phonon wave interference effects will be clearly manifested if the interface roughness is less than the phonon coherence length (ξ_{coh}) – Section 3.2 – of ~1 nm. Such interferences can be used for selectively backscattering or transmitting phonons and gives rise to the possibilities of phonon filters [166]. Experimentally, such transmission and filtering effects have been observed through phonon spectroscopy [167] and phonon imaging [168] techniques, e.g., through a sharp dip in the transmission spectrum [166] of longitudinal phonons in a (1 1 1) GaAs/AlGaAs superlattice corresponding to a *phonon* filtering mechanism.

In a binary multilayer system (e.g., 1 – GaAs and 2 – AlGaAs, where $d = d_1 + d_2$ is the sum of the respective layer thicknesses) wave interference between different phonon propagation velocities can also induce a *phononic band gap* of magnitude corresponding to the difference in the acoustic impedances, A ($=\rho v$) of the layers. Such a phenomena can be interpreted in analogy with *electronic band gaps* formed due to electron interference [88] (with magnitude proportional to the Fourier component of the crystal potential) or *photonic band gaps* (with magnitude corresponding to the dielectric constant contrast [169]) formed due to electromagnetic wave reflection at the edges of the Brillouin zone.

For InGaAs/AlAs superlattices, when the complete interaction between all the phonon polarization modes, i.e., high energy longitudinal mode and two lower energy transverse modes was considered, *intra-zone* frequency gaps were experimentally observed in the phonon dispersion spectrum and physically ascribed to the mixing and subsequent anti-crossing of phonon energies. A typical spectrum illustrating such aspects is illustrated in Fig. 42, for a (0 0 1) oriented $In_{0.15}Ga_{0.85}As$/AlAs superlattice. Such an energy level mixing and the consequent flattening of the phonon dispersion, leading to a reduced group velocity, is especially pronounced for phonon modes propagating oblique to the superlattice growth direction. For example, the *slow transverse* (ST) and the *longitudinal acoustic* (L) phonon modes, being in the same plane, are coupled to each other and propagate independently of the orthogonally polarized *fast transverse* (FT) mode. For increased phonon frequencies, near the edge of the Brillouin zone, the ST mode is folded back [165] into the main Brillouin zone and through an avoided crossing type interaction with the L mode induces the formation of a forbidden frequency gap. Quantitatively, the band gap arises through satisfying an *inter-mode* Bragg reflection with wavevector conservation, i.e., $\mathbf{k^L}_{\perp} + \mathbf{k^{ST}}_{\perp} = 2n\pi/d$ – for the normal component, and $\mathbf{k^L}_{||} = \mathbf{k^{ST}}_{||}$ – for the parallel component, along with $\omega_L = \omega_{ST}$. The formation of band gaps also gives rise to the possibility of an $L \leftrightarrow T$ mode conversion, which implies that at an interface the transmitted and reflected waves could carry equivalent amounts of energy flux yielding no net energy transport and *minimizing* the κ_L.

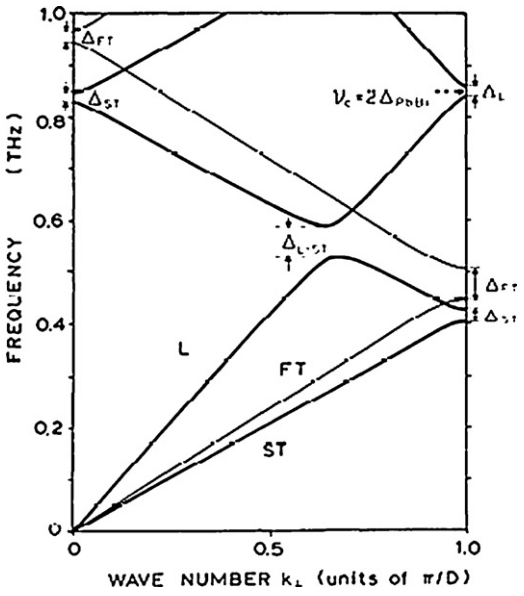

Fig. 42. For phonons propagating in arbitrary directions, a phonon dispersion spectrum exhibiting frequency gaps (Δ) corresponding to the superlattice periodicity and inter-mode interaction are observed. The symbols Δ_L, Δ_{FT}, and Δ_{ST} refer to the band gaps of the longitudinal (L), fast transverse (FT), and slow transverse (ST) modes, respectively. The coupled mode band gap is denoted by $\Delta_{L\text{-}ST}$. A reduced κ could be obtained due to non-propagating phonons (reprint permission from [168]).

As structures with phononic bandgaps allow for selective phonon frequency transmission, i.e., no phonon propagation in the bandgap, devices such as phonon mirrors and phonon resonators have been proposed [170] and could be implemented for regulating the κ_L through changing the (i) *layer thickness*, to modulate the frequency of transmission, and the (ii) *constituent materials*, for varying the acoustic impedance. Another intriguing possibility includes phonon localization [171] near the phononic band edges where a reduced or zero phonon group velocity could be obtained [172] with a corresponding reduction in the κ_L. The increased confinement, due to the localization, could also increase the scattering rates for Umklapp, boundary, and impurity scattering processes, e.g., in simulations [173] on 10 nm Si free standing quantum wells, a strong reduction of the group velocity mainly due to increased impurity scattering of the confined acoustic phonons was observed yielding a seven-fold decrease in the κ_L.

The onset of phonon localization can be experimentally inferred through a reduced value of kl, approaching 2π, where k is the magnitude of the wavevector ($=2\pi/\lambda_{ph}$). A $kl < 1$, was observed through measurements on atomically smooth p-type Bi_2Te_3/Sb_2Te_3 superlattices (Fig. 43(a)). For this data, the l ($\equiv l_{mfp}$) was calculated from κ_L ($=C_L v l$), assuming bulk values for C_L and v while the k was estimated from the phonon phase velocity. Concomitantly, lower κ_L values were also obtained [160] and are illustrated in Fig. 43(b). The reduced κ_L provided the basis for a figure of merit (ZT) of ~2.4, the largest ever obtained in thin films [12]. Additionally, since the measurements were carried out at 300 K, which is much higher than the individual Debye temperatures of Bi_2Te_3 and Sb_2Te_3, high frequency/low wavelength phonons would indeed be the major contributors. The superlattice period, d, was deemed to be more critical than the individual layer thicknesses of the Bi_2Te_3 and Sb_2Te_3, e.g., in Fig. 43(b), the κ_L values for the 6 nm

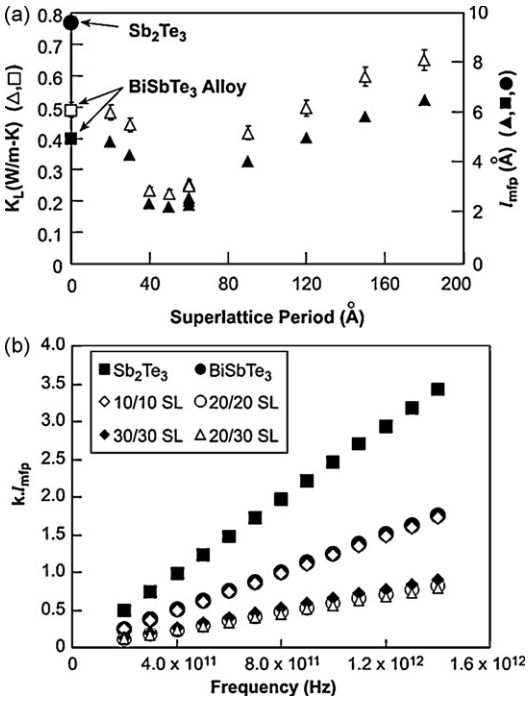

Fig. 43. (a) The lattice thermal conductivity (κ_L) is minimized by a factor of two, compared to the bulk, for a Bi_2Te_3/Sb_2Te_3 superlattice with a period ~5 nm. (b) A coherent backscattering of phonons, indicating localization-like behavior, is manifested through a low value of $\mathbf{k}\cdot l < 1$ (reprint permission from [160]).

period superlattice with differing individual layer thicknesses (3 nm/3 nm, 1 nm/5 nm, and 2 nm/4 nm for the Bi_2Te_3/Sb_2Te_3) is nominally the same. While a minimum in the κ_L corresponds to a similar *d*, the details of the reduction vis-à-vis the individual layers need to be further understood.

A related effect, for phonon channeling, exploits elastic anisotropy due to which the thermal energy transport is not in the same direction as the incident wave vector. Instead, phonon transport occurs preferentially in directions with smaller elastic constants (which imply smaller velocities)—motivated through lower energies—resulting in phonon focusing [174]. While such phenomena have been proposed to lead to a variation in the κ_L for silicon [175] by 90%, they have not been extensively studied in superlattices.

In summary, whether a reduction from the bulk value of κ_L occurs due to length scales involved in superlattices still seems to be a topic subject to extensive experimental investigation. The variation of experimental variables, such as strain, impurities and defects at the interface, and compositional uniformity from one particular investigation to another preclude a complete understanding. Subsequent to such control it would be profitable, from the viewpoint of reducing κ_L, to explore methods to engineer the phonon dispersion and transport through the creation of phononic bandgap structures. Detailed characterization of the predicted modification to the κ_L from phonon engineering has not been attempted and deserves further study.

3.4.5. Lowering thermal conductivity in one- and quasi-one-dimensional structures

It has been proposed that a greater degree of confinement, in systems comprising nanowires (NWs) and nanotubes (NTs) leads to a pronounced modulation of the phonon energy dispersion [176,177]. While group velocity, *v*, in NWs and NTs was predicted to be smaller, as in the case of superlattices (Section 3.4.4), a reduced phonon mean free path in two dimensions and enhanced phonon–surface, phonon–phonon, and phonon–carrier scattering rates would again contribute to an anisotropic and *lower* κ_L. The thermal boundary resistance [137] could also be modified by such scattering, e.g., as observed experimentally in single-walled [178,179] and multi-walled carbon nanotube (CNT) bundles which were seen to have a lower κ_L compared to a single CNT [180]. Considerations similar to those outlined earlier, i.e., roughness, degree of specular or diffuse scattering at the interfaces, etc., can also exert a great influence. For example, it was seen [144] that the κ_L for Si NWs synthesized with a saw tooth morphology [181] (with cross-sectional areas in the range of 2.58–28.62 nm^2) and core–shell nanowires [182] could be two orders of magnitude smaller than that of bulk Silicon, due to an increased number of diffuse reflections and inelastic scattering. For layered nanocomposites (Fig. 44(a)) it was shown that for a given volume fraction of embedded Si NWs (Φ_{Si}) in a Ge bulk matrix, a greater κ_L reduction could be obtained at smaller diameters [183]. The decrease is mainly due to enhanced phonon scattering at the Si–Ge interface, the area of which increases with decreasing NW diameter.

It was also suggested that the κ_L of NWs could be minimized through their embedding in a matrix with smaller acoustic impedance, *A*. On the basis of elastic energy minimization, it was noted [184] that a redistribution of the phonon modes from the acoustically harder nanowire *to* the acoustically softer matrix, could result in *phonon depletion* in the nanowire, and could diminish thermal energy transport. It would be interesting to examine such a principle [182] experimentally in nanocomposites [183] and in Si/SiO_2 or Si/Ge core–shell nanowires [185] – where the core and the shell are of different materials and *A*. However, the precision of the presently used thermal conductivity measurements may not yield valid confirmation.

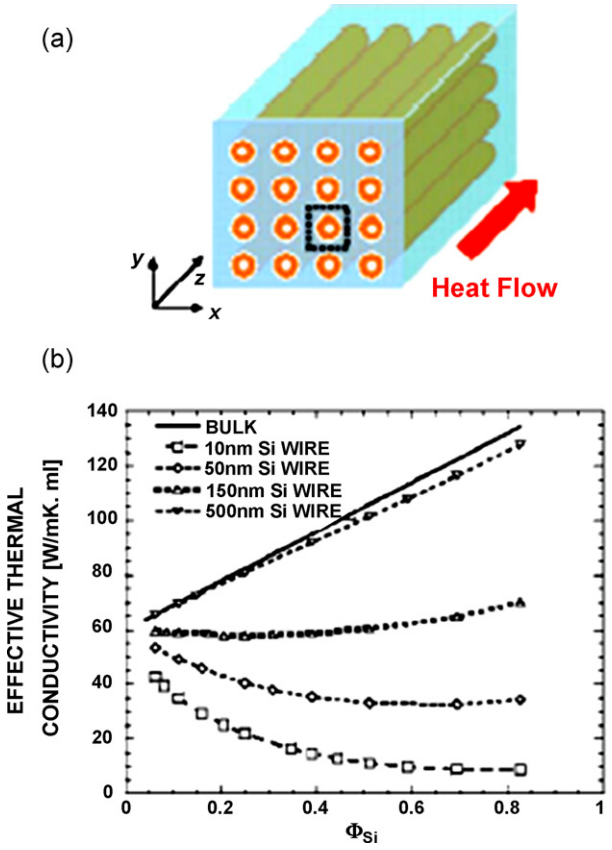

Fig. 44. (a) Periodic two-dimensional nanocomposites with tubular nanowire inclusions. (b) A thermal conductivity reduction below the bulk value as a function of the silicon nanowire radius (10, 50, 150, and 500 nm) placed in a Ge matrix, as a function of the volumetric ratio (Φ_{Si}) of the silicon nanowires (reprint permission from [183]).

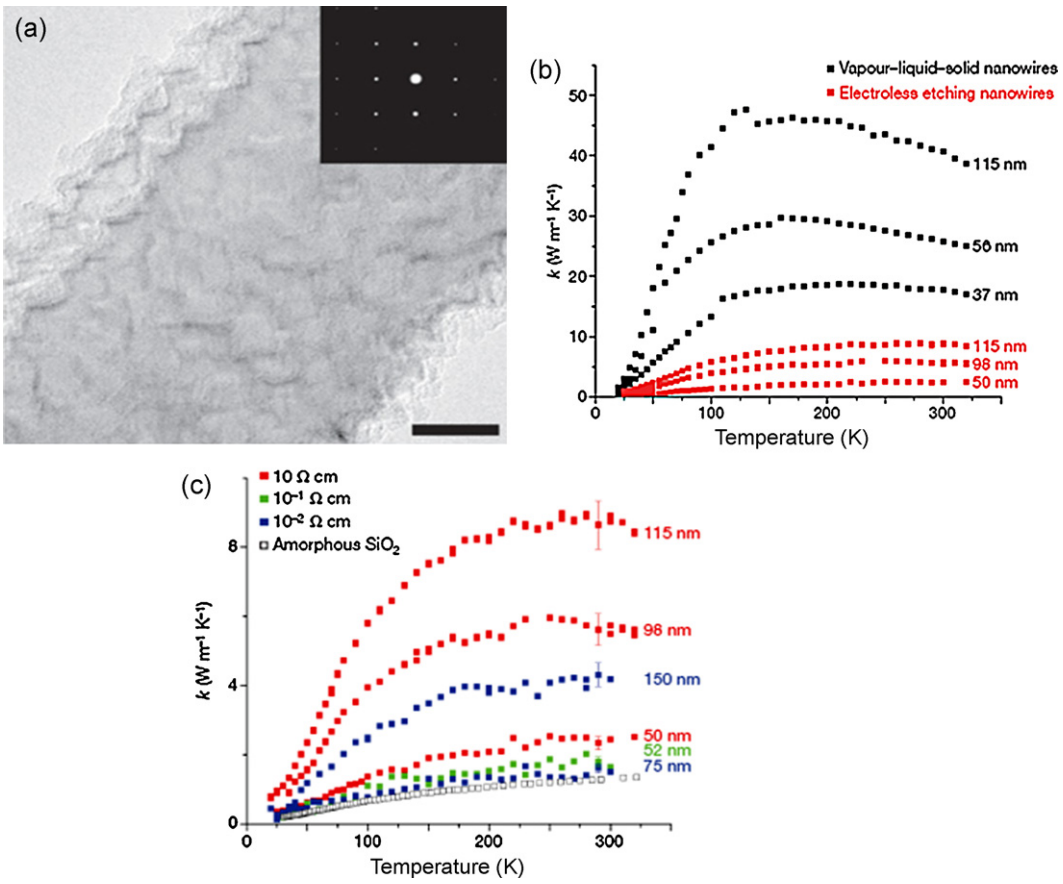

Fig. 45. (a) Rough Si nanowires fabricated through an electroless etching (EE) method. (b) The κ decreases with diameter and seems to be a strong function of the nanowire roughness, which is greater when synthesized through EE. (c) An increased dopant-phonon scattering (at lower wafer resistivity) is more effective for a reduced κ (reprint permission from [13]).

Controlled roughness seems to be a major contributing factor for drastically reducing the κ_L of Si NWs [13] to ~1.2 W/mK, very close to the value for amorphous SiO_2 and to the $\kappa_{L,\,min}$ (Section 3.2) of Si, which is ~1 W/mK. The Si NWs, in one case (20–300 nm diameter range) were fabricated through an electroless etching (EE) method where a Si wafer was locally oxidized using an aqueous solution of $AgNO_3$ and HF. The mean roughness of the synthesized NWs was ~1–5 nm (Fig. 45(a)). The measured κ_L was seen to be dependent on the (i) roughness, (ii) method of synthesis, and (iii) resistivity of the initial Si wafer, as indicated in Fig. 45(b) and (c). As expected, the κ_L decreases with reduced NW diameter presumably due to a reduced l and enhanced boundary scattering.

A reduction of κ_L by a factor of five was obtained for the EE NWs compared to the relatively smooth Si NWs synthesized by the vapor liquid solid (VLS) method. Interestingly, it was seen that the peak of the κ_L–T curve shifts with increasing roughness, from bulk Si (~25 K) → VLS fabricated NWs → EE fabricated NWs (Fig. 46). Since the peak is indicative [186] of the magnitude of the Debye temperature, Θ_D and onset of Umklapp scattering, it can be surmised that (i) the relative stiffness is increased in nanostructures, and that, (ii) the roughness promotes increased scattering. The key role of roughness was also seen through a comparison of the ratio of the bulk and NW κ_L values, i.e., ~100 at 300 K, and ~25,000 at 25 K (Fig. 47). With decreasing temperature, phonons of increasing wavelength contribute and are subject to scattering. A quick calculation from $\lambda_{ph} = h\nu/3k_BT$, with $\nu = 9 \times 10^3$ m/s at $T = 25$ K, yields a dominant phonon wavelength of ~6 nm, in correspondence with the observed roughness on the Si NWs.

The above study has brought forward qualitatively the concept of reducing κ_L through phonon scattering at different length scales, i.e., due to (1) *nanowire diameter*, inducing boundary scattering, (2) *surface roughness*, for longer wavelength phonon scattering, and (3) *atomic scattering*, for short wavelength phonons. Again, the drastic reduction in κ_L was deemed the primary contributor for the 100-fold increase in ZT of the Si NWs compared to bulk Si. However, the enhanced ZT of ~1 is still comparable to currently used thermoelectric materials and methods for reducing the κ_L may no longer be effective for further maximization of ZT, as values close to $\kappa_{L,\,min}$ seem to have been reached.

3.5. The phenomenon of phonon drag

It was mentioned in Section 1.1 that the Seebeck coefficient, S, includes contributions from both the diffusion of the carriers and

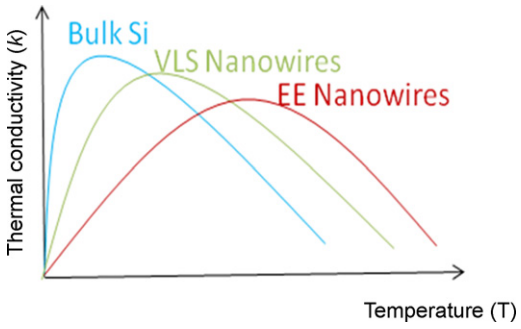

Fig. 46. Schematic variation of the thermal conductivity with temperature, as a function of size and increasing roughness (EE > VLS > Bulk).

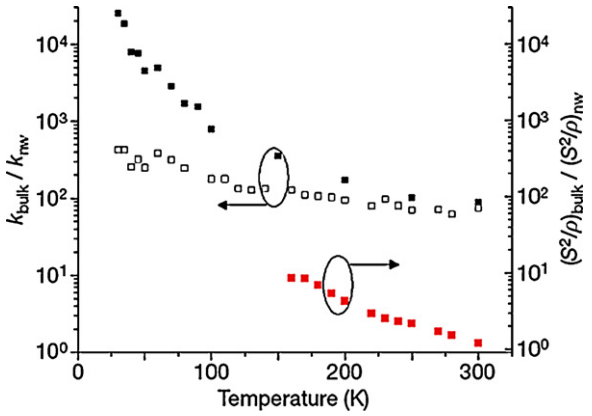

Fig. 47. A sharp increase in the ratio of the bulk to nanowire thermal conductivity with decreasing temperature is ascribed to the increasing influence of the nanowire surface roughness. On the other hand, the $S^2\sigma$ seems to be adversely affected by the roughness (reprint permission from [13]).

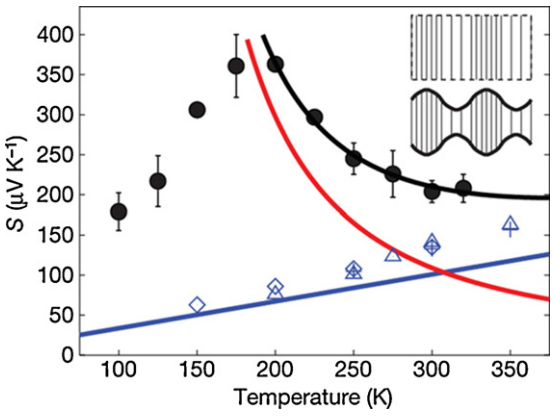

Fig. 48. The variation of the Seebeck coefficient (S) with temperature, invoking both the phonon drag component ($S_{ph\text{-}e}$) and due to the diffusive electron transport (S_e). While S_e is directly proportional to the temperature T (through Eq. (31)), $S_{ph\text{-}e}$ varies as $\exp(\Theta_D/T) - 1$, where Θ_D is the Debye temperature. The $S_{ph\text{-}e}$ component relies on the one-dimensional propagation of phonons down the nanowire, which is also favored by a reduced thermal conductivity (reprint permission from [4]).

due to the influence of carriers on the phonons, i.e., a *phonon drag* component. Typically, the diffusive component dominates the measured S at room temperature as the latter is considered to be a low temperature, impurity, and size dependent phenomena [2]. However, recent experimental results in Si nanowires were modeled to indicate that phonon drag could contribute to an enhanced S and a reduced κ_L, yielding values even below $\kappa_{L,\,min}$ [4]. Such phenomena, if proved valid, promise a new direction to increase the thermoelectric figure of merit. The purpose of this section then is to first outline the underlying principles behind phonon drag and how it could be used to increase S and reduced κ_L of one-dimensional nanostructures.

The original treatment of phonon drag [187] involved the creation of a thermoelectric voltage due to the sweep of the electrons by the phonons, which is increased when the respective heat flow contributions/energies are comparable [2]. Additionally, the phonon drag has a greater contribution from the longitudinal phonon modes compared to the transverse modes due to the greater group velocities in the former case. While the magnitude of the voltage due to phonon drag is small in metallic materials (<5 μV/K) and is most evident at $T < \Theta_D$, orders of magnitude larger contributions can be obtained in semiconductors [3], typically at doping levels $\leq 10^{15}$/cm^3, e.g., 5–20 mV/K at $T < 20$ K in Ge [3,188] and Si [189]. Such an increase was first measured in Ge [188] as deviations from the expected variation of the diffusive thermopower, at low temperatures (~15 K), and correlated to the phonon distribution non-equilibrium and the ratio of the phonon and electron mean free paths. Generally, the contribution to the phonon drag thermopower, $S_{ph\text{-}e}$, is directly proportional [3] to the ratio of the phonon relaxation time ($\bar{\tau}$) and the carrier relaxation time ($\bar{\tau}_e$) along with the fraction (fr) of electronic carrier momentum lost to the phonons through $S_{ph\text{-}e} = \pm m^* v^2 (fr \cdot \bar{\tau}/\bar{\tau}_e)$, and decreases with increasing carrier concentration and temperature [190]. In the above expression, m^* and v refer to the effective mass and acoustic velocity, respectively. A *saturation* of the $S_{ph\text{-}e}$ occurs with increased carrier concentration, e.g., at ~10^{18}/cm^3 in p- and n-Si, due to the increased interactions of the electrical carriers with the phonons, and a consequent decrease of the mutual drag. Consequently, below a critical phonon mean free path $l_p = v\bar{\tau}$, which for Si was estimated to be ~1 μm, $S_{ph\text{-}e}$ is drastically reduced/quenched [191].

3.5.1. Phonon drag in nanostructures

A κ_L value of 0.76 ± 0.15 W/mK was measured for a 10 nm Si NW, which was found to be below the *theoretical* limit (Section 3.2) of $\kappa_{L,\,min}$ ~ 1 W/mK. It was then postulated [4], that enhanced phonon drag phenomena in Si NWs ~10–30 nm in diameter was probably responsible. It was argued that the original definition of $\kappa_{L,\,min}$ assumes a *minimum* phonon mean free path of $\lambda_{ph}/2$, which could be pertinent for three-dimensional transport but may not be valid in one-dimension. A mechanism involving one-dimensional phonon propagation was invoked, where the NW boundaries were incorporated into the phonon mode (see Fig. 48 inset) promoting a new mode of heat transport. The l_p, below which $S_{ph\text{-}e}$ is reduced is now extended to the length of the NW instead of corresponding to the diameter and phonon drag phenomena could again be important.

Consequently, a phonon mode propagating down the nanowire, with positive and negative dilatations (as shown in the inset to Fig. 48 corresponding to alternate heating and cooling and heat transport down the wire is relevant [192]). For longitudinal acoustic (LA) modes propagating along the axis of the NW which contribute to most of the $S_{ph\text{-}e}$, a large λ/high ω corresponds to a large distance between temperature fluctuations and an adiabatic situation while the converse case (small λ, large ω) implies isothermal behavior. As the phonon drag is maximized when the phonon energy (/wavelength – λ_{ph}) is comparable to the electrical carrier energy (/wavelength – λ_e), the dominant λ_{ph} can be calculated from a given carrier concentration, n. With a three-dimensional carrier distribution we can estimate λ_{ph} ($\lambda_e = 2\pi/\sqrt[3]{3\pi^2 n}$) which for $n = 3.10^{19}$/cm^3, is ~6 nm, which is greater than the thermal phonon wavelength ($\lambda_{th} = hv/k_BT$) of ~1.5 nm, assuming a bulk $v \sim 10^4$ m/s and $T = 300$ K, and hence an adiabatic condition could be achieved. In the adiabatic case, lack of contact with the ambient yields a longer phonon relaxation time – $\tilde{\tau}$, and a larger $S_{ph\text{-}e}$. The transverse acoustic modes, which undergo less scattering compared to the LA modes due to the confinement perpendicular to the length of the nanowire were also considered to be adiabatic in character [193] and could contribute to the thermopower. The total S was then equal to the sum of the contributions from the carriers (S_e) and $S_{ph\text{-}e}$. As the scattering rate, related to $(\bar{\tau})^{-1}$, depends on the number of phonons through the Bose–Einstein distribution function, $f_{BE} = 1/(\exp(\Theta_D/T) - 1)$, the $S_{ph\text{-}e}$ was modeled as being proportional to $\exp(\Theta_D/T) - 1$. The S_e is proportional to the temperature through the Mott formula, $S_e \sim aT$. A good data fit to the expression, $S = S_e + S_{ph\text{-}e}$ ($=aT + b[\exp(\Theta_D/T) - 1]$), where a and b are constants of proportionality (Fig. 48) was used to justify the above hypotheses.

It was then remarked that the larger $\tilde{\tau}$ only applies only to phonons participating in the phonon drag. Consequently, the κ_L which is due to the averaged contributions of *all* the phonon modes would be unaffected. However, a reduction in the κ_L can be brought

about by a reduced phonon group velocity peculiar to one-dimensional phonon propagation, as outlined in the inset to (Fig. 48) While this study is interesting in that values below the $\kappa_{L,min}$ were obtained, many details regarding phonon propagation and their contribution to a reduced κ_L still have to be elucidated. For example, under what specific geometries would a three-dimensional to one-dimensional crossover of phonon propagation and the incorporation of the NW boundaries into the propagation mode be possible? The spectral distribution of the phonons contributing to the $S_{ph\text{-}e}$ must also be rigorously clarified along with the range of carrier concentrations for which $S_{ph\text{-}e}$ can be tailored through an increased $\bar{\tau}$. It is also of interest to examine the validity of this model for different materials. From an experimental point of view, the lack of a pronounced thermal isolation, e.g., between the heating and sensing islands, could contribute to errors in the measurement and reproducibility of the ultra-low κ_L values is necessary.

So far, we have considered the electrical and thermal characteristics of nanostructured thermoelectrics with an aim of increasing their figure of merit. In the next section, we will review an alternative method of thermal to electrical energy conversion based on electron emission in a temperature gradient.

4. Thermionics at the nanoscale

The basic aspects of thermionic emission were first introduced in Section 1.2 and through Fig. 3. In this section, we elucidate how thermionics has been implemented in nanostructured materials for energy conversion and refrigeration. Most thermionic applications are, at present, focused on the latter issue. A powerful motivation for thermionic refrigeration is provided by the miniaturization of electronic devices. Intrinsic to the devices, it is seen that high temperatures are prevalent at certain spots/locations resulting in a non-homogeneous temperature distribution (with a ΔT of 5–30 K above the average) and heat fluxes (in the range of 10–50 W/cm^2). Concomitantly, with miniaturization now approaching the nanoscale, module and chip heat fluxes could reach 15 and 250 W/cm^2, respectively [194]. Hence, thermal management becomes critical as a 10 K rise in semiconductor junction devices could shorten their lifetime by a factor of two [195], catalyzed by an exponential dependence on the temperature rise. Consequently, thermionics integrated with electronic devices, which could achieve heat pumping/cooling power densities in the range of 10 W/cm^2 or greater could be very useful. We first consider traditional vacuum-based thermionics and subsequently solid state approaches which could be easier for integration with electronic devices. Much work has been done in arranging [196] solid state superlattice-based structures incorporating thermionic emission based cooling for heat pumping, temperature stabilization, and temperature tuning in microelectronics.

4.1. Evaluation of thermionic device efficiency and challenges in application

The current (J_{th}), from the cathode to the anode in a typical vacuum diode, in the absence of electron interactions, is given by the Richardson equation [18] through $J_{th} = A_0T^2 \exp(e\Phi/k_BT)$. $A_0 (= emk_B^2/2\pi^2\hbar^3)$ is a constant ~120 A/cm^2 K^2, Φ is the work function, as defined in Fig. 3 is typically appropriate for bulk emission—considering a three-dimensional density of states charge distribution at the emitter. It should also be noted that this equation is rigorous only when the electron mean free path is greater than the distance between the cathode and the anode [197], i.e., the electrons do not scatter during transit. Such *ballistic* transport distinguishes thermionic devices from traditional thermoelectrics where electron transport is mostly diffusive. From the original theory [198], additional assumptions underlying the Richardson equation include that (1) the barrier height is much larger than k_BT, (2) thermal equilibrium is established at the plane of emission, and that (3) the net current flow does not affect the thermal equilibrium, and is a superposition of two currents (one from the cathode to the anode: J_C and the other from the anode to the cathode: J_A).

For use of thermionic emission in *energy conversion*, the cathode with a work function Φ_C in contact with a hot reservoir (at temperature, T_H) emits electrons into vacuum which are then attracted to a cold anode (at temperature, T_C) with a work function Φ_A. It is desirable for $\Phi_C < \Phi_A$ to increase forward electron emission relative to electron back flow. The net current (J_{th}), passing through an external circuit, would then be:

$$J_{th} = J_C - J_A = A_0\left\{T_H^2\exp\left(-\frac{\Phi_c}{k_BT_H}\right) - T_C^2\exp\left(-\frac{\Phi_A}{k_BT_C}\right)\right\} \quad (46)$$

To obtain a large power density, the thermal to electrical energy conversion would have to be maximized, i.e., through a large J_{th}, implying a large J_C and a small J_A. In addition to intrinsic factors, such as Φ, T_H, T_C, etc., one must also consider the formation of space charge in between the electrodes, due to finite electron transit times, which could reduce J_{th} [199]. The value of V (=V_0) necessary to reduce the J_{th} to zero (i.e., an open circuit condition) could be *analogous* to the Seebeck voltage [200].

Similar considerations hold for the use of thermionic emission in refrigeration. Here, a hot surface, in contact with the cathode, can be cooled through electron flow from the cathode to the anode. A voltage bias (V) relative to the cathode is applied on the anode (Fig. 3), which increases the Fermi level (E_F) and the effective work function of the anode, i.e. Φ_A is replaced by Φ_A + eV in Eq. (46). This then reduces the counter electron emission current from the anode resulting in a net increase of J_{th} and a reduction of the cathode temperature, T_H. Cooling occurs as the emitted electrons, from the cathode, are of higher energy than the Fermi energy (E_F) of the cathode. Given that, as discussed in Section 2.1, the excess thermal energy of electrons is equal to ($E - E_F$), the removal of such "hot" electrons implies cooling. Conversely, since the anode's emission current is less than that of the cathode's, the anode temperature, T_C, would be raised by the received electrons. This phenomenon would then be analogous to the Peltier effect in a thermoelectric.

For thermionic devices that are designed for refrigeration, the efficiency is usually rated through the coefficient of performance (COP). The COP is given by the ratio of the *net* thermal energy transported to the electrical power input (P_{input}), as

$$\mathrm{COP} = \frac{Q_{TI} + Q_{Rad} - Q_R - Q_K}{J_{th}\cdot V + Q_R} \quad (47)$$

We first consider the numerator, composed of the thermal energy transported from the cathode to the anode by the thermionic current (Q_{TI}) and due to radiative heat transfer (Q_{Rad}). Heat transfer by conduction or convection is avoided through the use of vacuum and is not considered for the COP calculation. Q_{TI} is given by the product $J_{th}\cdot V$, where V includes the overall potential barrier (Φ) that has to be overcome in addition to the average kinetic energy of the electrons as they cross the cathode/barrier interface (=$2k_BT$, assuming a Maxwellian electron distribution [201]), i.e., eV = $\Phi + 2k_BT$, and $Q_{TI} = J_{th}(\Phi/e + (2k_BT/e))$. $Q_{Rad} = -e\sigma_{SB}\,(T_H^4 - T_C^4)$, where e is the thermal emissivity of the electrodes and σ_{SB}, the Stefan-Boltzmann constant (5.67×10^{-8} W/m^2 K^4). While Q_{Rad} is negligible (~0.04 W/cm^2) at T = 300 K with a temperature difference ($\Delta T = T_H - T_C$) of 100 K, it increases rapidly at higher temperatures, e.g., at T = 1000 K and with a ΔT of 100 K, Q_{Rad} is ~2 W/cm^{2}which is significant and could reduce thermionic conversion efficiency. Additionally, heat generation from a finite

device resistance (R) and the thermal conductance (K) leads to both Joule heating at the cathode, Q_R (=$0.5J^2R$) and parasitic heat backflow Q_K, which must be subtracted from the net forward transport of the thermal energy.

For the denominator, the P_{input} to facilitate the electron emission is comprised of the electrical input, $J_{th}{\cdot}V$ and the heating Q_R. Eq. (47), and the COP is derived by considering these quantities. It can also be understood [5] the COP must then be optimized, for particular device characteristics (defined by R, Φ, T, etc.) at a particular J_{th}. Consequently, both the J_{th} and the COP are used for evaluating thermionic device performance.

However, the main challenge in thermionics research seems to be in obtaining a large enough J_{th} for practical cooling applications through decreasing the Φ of the cathode material, rather than the maximization of COP. Since cooling power densities, $\sim Q_{TI} = J_{th}(\Phi/e + (2k_BT/e))$ exceeding 10 W/cm^2 would be required for large scale application of thermionics (at room temperature) typical metals with $\Phi > 2$ eV (e.g., Li: 2.3 eV, Cu: 4.4 eV) would not be able to supply any reasonable J_{th}. Indeed, with $\Phi \sim 2$ eV, a temperature in excess of 1300 K would be necessary to obtain a Q_{TI} of 10 W/cm^2 (with a corresponding J_{th} of ~4.5 A/cm^2, which seems impractical). For room temperature operation, the Φ would have to be reduced to ~0.33 eV, which has not yet been achieved for any material. While Cs–O coated surfaces of doped semiconductors, could have a lower $\Phi \sim 1$ eV [202], such efforts seem to have received less attention (Section 4.4). In the following sections, we look at various other methods to overcome the high work function problem through the use of nanostructures. We will first consider experiments based on the original formulation of thermionics, i.e., where a vacuum separates the cathode and the anode.

4.2. Applications of vacuum-based thermionics through nanofabrication

To overcome the problem with an intrinsically large Φ, thermionic emission across nanometer sized gaps was considered for obtaining large J_{th}. In this case, for a given applied voltage, the electric field (≡ spatial voltage gradient) are enhanced, due to the small gap spacing. An additional contributing factor, relevant at such scales, is that electrons in the electrodes induce positive image charges [203] in the surrounding vacuum and the resulting electric field could supplement the *applied* field, reducing the net barrier to electron emission to below the nominal Φ. The magnitude of the barrier reduction is directly proportional to the total electric field with the result that cooling power densities could be increased by orders of magnitude (Fig. 49). It can also be seen from this figure that at gap spacing less than 10 nm, cooling power densities greater than 1000 W/cm^2 are possible at room temperature [204]. The use of tapered electrodes [205], such as nanotubes/nanowires which have a smaller radii of curvature, which enhance the electric field was also proposed for increasing cooling power.

Cooling from thermionic emission was experimentally investigated using a pair (i.e., cathode and anode pair, with a gap spacing ~1–2 nm) of micro-fabricated tips on cantilevers similar to those used at the ends of an Atomic Force Microscope (AFM) (Fig. 50). While a temperature change ~0.1 to 1 mK, corresponding to nW of cooling was measured, the absence of further characterization in terms of the COP/J_{th} precludes detailed understanding of the efficacy of such an approach.

Moreover, the emission of electrons at large electric fields can also result in undesirable heating. For example, a net cooling of the cathode/emitter will occur when electrons [206] from above the Fermi energy (E_F) are removed. On the other hand, a heating of the cathode occurs for electron emission from energy levels [207] below the E_F; this phenomena has been referred to as the Nottingham effect [208]. The probability of such emission increases under large electric fields with the consequence of a smaller than expected reduction in the emitter temperature. Such effects have been observed experimentally [209]. Other factors that would have to be taken into account, in such experiments, include the possibility of electron tunneling between the electrodes (Fowler–Nordheim type field emission), which could occur if the spacing between the electrodes was reduced adequately to permit such type of field emission. In situations where both thermionic and tunneling effects are prevalent, phenomenology related to *thermo-tunneling* is used [206].

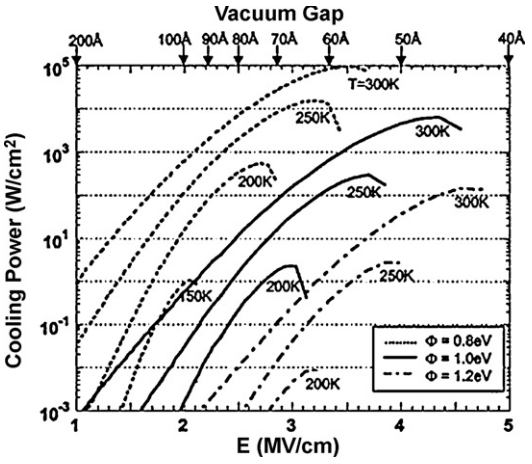

Fig. 49. Cooling powers greater than 100–1000 W/cm^2 can be achieved through vacuum-based thermo-tunneling (incorporating both thermionic emission and tunneling) methods, using low work function (Φ) emitter electrodes. However, heating through electrons emitted from below the E_F (known as the Nottingham effect) could be harmful (reprint permission from [204]).

Whether such arrangements can yield practical cooling powers over a large area is generally unclear due to several reasons. First, a

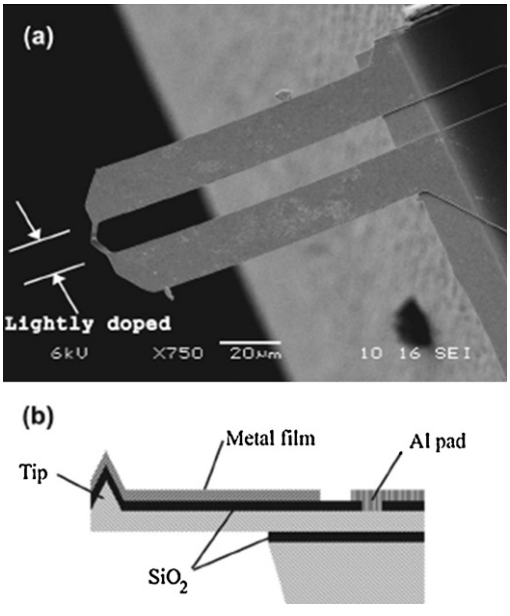

Fig. 50. Arrangement for probing thermo-tunneling phenomena for thermionic emission mediated refrigeration. A suspended cantilever tip is used as an emitter while a collector is placed ~2 nm away, and is controlled through a piezoelectric bimorph actuator. A temperature change/refrigeration of 0.28 mK was measured through monitoring the change of resistance of an integrated resistive thermometer (reprint permission from [210]).

well defined gap must be maintained and second, for the Φ to be reduced to around ~0.33 eV by electric field induced barrier lowering, the initial Φ of the electrode material surface itself should be <1 eV. Both these are difficult to achieve. While it was suggested [210] that increasing the effective emitter area could be used to increase the cooling current density, the maintenance of a small gap (with a maximum spacing ~5 nm) for enhanced electric fields [211] would be quite challenging for large scale fabrication and additionally, prompt a reconsideration of parasitic radiative heat flow from the emitter/cathode to the anode.

4.3. Energy filtering: achieving maximum possible efficiency

In the previous two sections, it was seen that using a purely vacuum-based approach for thermionics, as originally envisaged, presents problems due to the (i) large work function of the emitter, along with the (ii) possibility of the Nottingham effect mediated heating at high electric fields. To control and limit the energy range of the emitted electrons, especially for (ii) above, *resonant tunneling* of electrons along with electron energy filtering has been proposed [212]. Such energy selective emission of electrons requires a greater consideration of the details of their spectral dispersion [213]. This could be done through employing an idealization and considering the optimal conditions for maximizing the efficiency of a thermionic/thermoelectric heat engine using electrons as the "*working gas*". It is well known, for example, that the maximum thermodynamic efficiency is given by the Carnot efficiency, η_{Carnot} ($=1 - T_C/T_H$) where T_C is the temperature of the cold anode, and T_H the temperature of the hot cathode [214]. Additional requirements for achieving η_{Carnot} include perfectly reversible carrier transport at a single energy [215], say, ε_R. Now, for a thermionic *heat engine*/power generator the efficiency, $\eta_{TI}^{PG} = W/Q_H$ is defined as the ratio of the work (W) done per electron transmitted from the hot to the cold reservoir to the heat removed from the hot reservoir (Q_H). Referring to Fig. 51, $W = eV_0 = (\varepsilon_C - \varepsilon_H)$, where ε_C and ε_H are the electrochemical potentials of the cold and hot reservoirs, respectively and $Q_H = (E_R - \varepsilon_H)$. This implies that $\eta_{TI}^{R} = (\varepsilon_C - \varepsilon_H)/(E_R - \varepsilon_H)$. For transmission to occur exactly at ε_R, the electronic probabilities, manifested through the Fermi–Dirac distribution (f_{FD}) at both the HOT (f_{FD}^H) and the COLD (f_{FD}^C) sides, should be equal, as stipulated through:

$$f_{FD}^{H}\left(=\frac{1}{[1+\exp(E_R-\varepsilon_H)/k_BT_H]}\right) = f_{FD}^{C}\left(=\frac{1}{[1+\exp(E_R-\varepsilon_C)/k_BT_H]}\right) \quad (48)$$

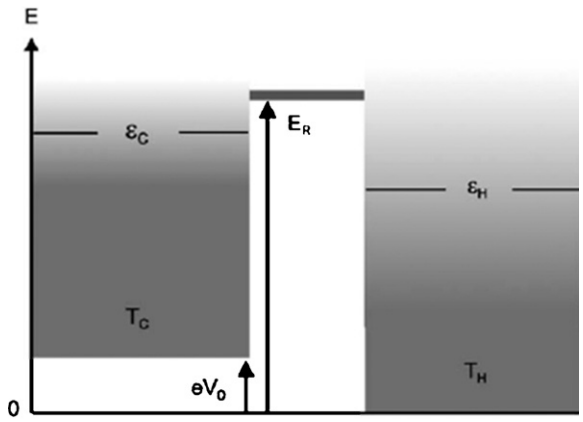

Fig. 51. A thermionic device can be considered as a heat engine operating between a hot reservoir (T_H) and a cold reservoir (T_c), and could achieve Carnot efficiency if reversible electron transport occurred between the reservoirs through a single energy level (E_R) (reprint permission from [215]).

From substituting the above equality in the expression for η_{TI}^{PG}, it can be inferred that η_{TI}^{PG} approaches η_{Carnot} when electron transport indeed occurs through a single energy level. In an analogous manner, the efficiency of thermionic refrigeration (η_{TI}^{R}) can also be derived and equated to the equivalent Carnot efficiency for refrigeration, as:

$$\eta_{TI}^{R} = \frac{E_R - \varepsilon_C}{\varepsilon_c - \varepsilon_H} = \frac{T_H}{T_H - T_C} \quad (49)$$

In practice, such a specific energy level, ε_R, can be obtained in nanostructured devices, e.g., through resonant tunneling effects as outlined in the next section. However, electron emission through such discrete energy levels/narrow energy range would also limit the total available power. Increased power, at the expense of irreversible electron transport and reduced efficiency, may then be obtained by broadening the range of energy levels, above ε_R.

4.4. Vacuum-based resonant tunneling approaches

The electrons from a bulk emitter electrode can be selectively transmitted, through resonant energy levels in the semiconductor, into vacuum with a resultant cooling of the emitter. The principle [212] behind this approach is illustrated in Fig. 52, and primarily involves coating an emitter surface with a thin (<5 nm) wide band gap semiconductor. Application of an electric field, then creates a triangular shaped potential well with quantized energy levels in the semiconductor layer, in the perpendicular direction. Electron emission from the electrode into vacuum could now occur resonantly through these energy levels. As the electric field is increased, the alignment of the semiconductor sub-bands with the emitter *hot electron* energy levels is also enhanced resulting in greater electron emission and cooling. The extent of cooling would also be proportional to temperature, due to the energy level broadening of the sub-bands, which allows for a greater amount of electron flow. A cooling power greater than 1000 W/cm^2, at electric fields ~7 MV/cm, was predicted at $T = 300$ K using such a scheme. Off resonance, the electron emission would be decreased and when the sub-band is of energy lower than the Fermi energy of the emitter, e.g., at very large electric fields, there is the possibility of a net heating of the emitter due to emission of electrons from states below the E_F via the Nottingham effect [216]. The uniformity of

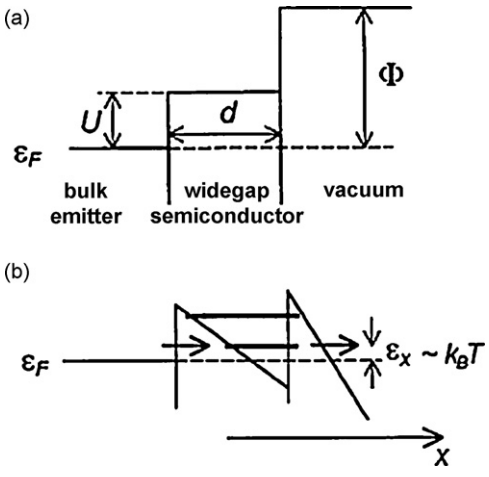

Fig. 52. (a) To overcome the large work function typical of most metal electrodes, the emitter electrode is coated with a, very smooth (less than 20% variation in thickness), thin layer of wide band gap semiconductor. (b) In the presence of an electric field, electrons can resonantly tunnel through the energy levels in the semiconductor into the vacuum thus affecting a cooling of the emitter (reprint permission from [212]).

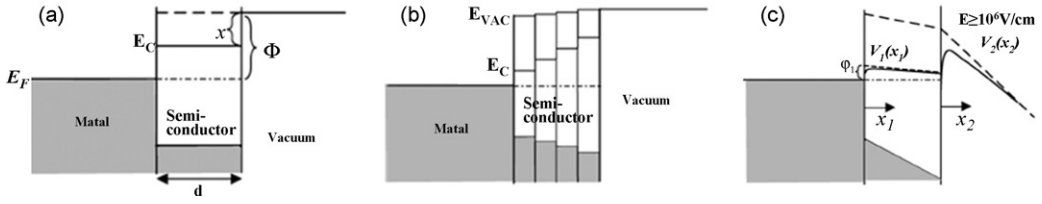

Fig. 53. (a) Coating an electrode emitter with a semiconductor achieves the replacement of a large work function (Φ) with a smaller Schottky barrier ($E_c - E_F \sim 0.5$ eV). Here the electrons traverse the semiconductor layer through internal field emission before being emitted into vacuum. (b) The possibility of carrier heating during their traversal in the semiconductor can be diminished through the deposition of semiconductor layers exhibiting a band gap gradient. (c) On the application of an electric field the conduction band is flattened, reducing the intrinsic Joule heating (reprint permission from [218]).

emission due to the film thickness uniformity and the heating of the anode, at large current densities, are of concern in this scheme.

An alternate method to effectively decrease the barrier for electron emission also involves the coating of the cathode/emitter with a semiconductor layer [205] (Fig. 53(a)). The electron emission is now regulated by the electron affinity (χ) of the semiconductor and the Schottky barrier between the metal cathode and the semiconductor, both of which can be made much smaller than Φ, e.g., by using a Cs–O coated surface with a work function smaller than the semiconductor bandgap [217]. On the application of an electric field, the electrons are emitted from the metal into the semiconductor and then drawn into the vacuum. The Joule heating of the semiconductor layer due to electron transit could be avoided through the use of a series of semiconductor layers with graded energy bandgaps [218] (Fig. 53(b)). On the application of an electric field, the band flattening in the semiconductor layers (Fig. 53(c)) facilitates isothermal electron transfer through the semiconductor. The effective barrier for electron emission is then the modified electron affinity i.e., V_2 in Fig. 53(c), which could be smaller than Φ. Generally, at low electric fields the emission is limited by Φ while the Schottky barrier height V_1 influences the emission at high electric fields, with the highest efficiency obtained when V_1 equals V_2. It was calculated that large emission currents corresponding to cooling power densities of the order of 10–100 W/cm^2 could be achieved in this configuration. While such an idea does not seem to have been experimentally demonstrated, it is pertinent to note that χ is generally easier to manipulate compared to Φ. For instance, certain oxide surfaces (i.e., CaO and MgO) could be formed with a negative Φ, when χ is smaller than the energy band gap.

4.5. Solid state approaches

To date, the experimental implementation of vacuum-based thermionics seems quite restricted by the (1) large Φ, (2) Nottingham effects, and (3) unwieldy demands placed on the experimental apparatus, i.e., small gap spacing, etc. A solid state thermionic approach, where the vacuum would be replaced by a semiconductor barrier layer, was then proposed [219] based on the analogy of the work function/barrier height to the *band offset* between the chemical potentials/E_F of the constituent semiconductors, e.g., in a heterostructure [220]. The proposed solid state approach exploits semiconductor band engineering to tune the effective Φ through combinations of lattice matched semiconductors, e.g., in the GaAs/$Al_{1-x}Ga_xAs$ system where the band offset is dependent on x. Additionally, the spacing between the cathode and the anode can be very precisely and effectively controlled through thin film deposition using MBE or MOCVD. However, the replacement of vacuum with a barrier material of finite thermal conductivity would now allow heat backflow (increased Q_K in Eq. (47)) and reduce thermionic cooling efficiency. Some of the basic principles behind present research and applications focus on the use of multilayers and superlattices, which promise (i) enhanced cooling power densities to reduce module and chip heat fluxes, and (ii) faster transient response for overcoming sudden temperature excursions [221], and will now be discussed.

4.5.1. Solid state thermionic devices

A solid state thermionic design is based on principles similar to those enunciated earlier for vacuum-based systems. The band offset between the emitter and the barrier layer would accomplish energy filtering whereby only electrons higher in energy than the barrier (of energy E_b) would be transported. The separation of high energy electrons could then yield an enhanced effective Seebeck coefficient (S) albeit with diminished electrical conductivity (σ), due to a reduction of the total carrier concentration. The increase of S could also be inferred through its proportionality to the ratio $(E_F - E_b)/k_BT$ in the direction of electron transport over a barrier, where a larger separation of E_b from E_F enhances S (see Section 2.3.2).

To prevent heat backflow, the barrier layer must have the lowest possible κ_L. The thickness of the barrier (L) is then subject to the limits of heat carrier mean free path, l_m (upper limit) and tunneling (lower limit). In analogy to vacuum-based thermionics, where the current flow from the hot cathode to the cold anode is ballistic, the L should be smaller than the effective electron mean free path in the barrier material. At the other limit, electron tunneling is promoted by a very small L. Now, the propensity for thermionic emission through the barrier is proportional to $\exp(-\Phi/k_BT)$ while the extent of tunneling is proportional to $\exp(-\sqrt{2m\Phi/\hbar^2}L)$. Consequently, thermionic emission would dominate if $L > (\Phi/k_BT)\sqrt{\hbar^2/2m\Phi}$ which then implies that below a certain length, $L_t (= \sqrt{(\Phi/2m)}\hbar/k_BT)$ tunneling is dominant. It can then be calculated that for an effective barrier height, i.e., $\Phi \sim 1$ eV and at $T = 300$ K, $L_t \sim 3.7$ nm. The upper limit on $L \sim v_F\tau$, where v_F is the Fermi velocity ($= \hbar k_F/m$) and τ, is the mobility relaxation time ($\sim \mu m/e$) – see Section 2.3.1. With an electron concentration $\sim 10^{22}$/cm^3 and a mobility of ~ 0.1 m^2/V s, an l_m of ~ 500 nm was estimated. Therefore, the barrier layer thickness would have to be in a wide range, between 3.7 and 500 nm. The width of the emitter is normally chosen by considering the electron energy relaxation length and is such that it enables carrier traversal without being subject to electron–phonon scattering [222].

Based on such ideas, a heterostructure based thermionic cooler was theoretically proposed [219] with $Hg_{1-x}Cd_xTe$ as the barrier layer, using optimized band offsets from the anode and cathode. $Hg_{1-x}Cd_xTe$ was chosen due to its low thermal conductivity (~ 1 W/mK) and barrier height of ~ 0.4 eV [223]. The J_{th} through the device was modeled by the Richardson equation to be $\sim 10^4$ A/cm^2 for an electron mean free path of ~ 0.3 μm at $T = 300$ K. While such current densities were deemed to yield cooling power densities on the order of 1000 W/cm^2, it was intriguing to note that an effective Seebeck coefficient, $S \sim 200$ mV/K, was assumed. While electron

transmission over the barrier could possibly yield enhanced S, such an enormous value and its role in obtaining such high cooling powers needs to be further clarified. This is especially important to justify the claim that the thermionic cooling could be two to five times *better* than Peltier cooling. Experimental investigation of such an approach was implemented in a MOCVD fabricated device, comprised of a n+ InGaAs cathode/InGaAsP barrier (1 μm thick)/n+ InGaAs anode patterned into 90 μm × 180 μm mesas [224]. With a barrier height of ~0.1 eV, a cooling of ~0.5 K (at a heat sink temperature of 20 °C) and ~1 K (at 80 °C) was measured and stated to correspond to a cooling power of 200–300 W/cm^2 (Fig. 54). While the larger cooling at higher temperatures could be due to both the increased current density and reduced barrier κ_L, further modeling seems to be necessary to understand the individual contributions. Additionally, since the current density was of the order of 100 A/cm^2, it would have been useful to delineate how the particular values for the cooling power were obtained. Thermal modeling indicated that parasitics (i.e., Joule heating in the wire bonds and at the contacts, heat conduction from the package, thermal conductance of the barrier layers, etc.) contributed to the lower than expected cooling. Such issues seem to be quite common in experimental investigations of thermionic cooling and much attention then has to be paid to packaging. Parasitic heat conduction may be further reduced through the use of multilayer [225]/superlattice [26] structures, which provide additional resistance to heat flow, as outlined in Section 3.4.4. While the presence of band energy offset at each electrode-barrier layer junction layer [226] results in an incremental cooling, it is unclear whether the net cooling obtained would be much different than that obtained by attached single-barrier structures.

It would be interesting to consider the differences and correspondences of the thermionic multilayer structures/superlattices to what has been proposed for thermoelectrics. The thermionic equations are nonlinear with respect to the applied voltage and temperature (e.g., the Richardson equation) and are relevant strictly for the ballistic propagation of electrons. Consequently, the Joule heating in the barrier layer is not of concern in an *ideal* thermionic device [220]. However, the basic mechanism in both *solid state* thermionics and thermoelectrics seems to be related to the Peltier effect, say at the electrode-semiconductor barrier interface. For devices with length scales equal to a electron mean free path it has then been shown that the figures of merit of thermionic and thermoelectric devices converge to a common expression [227]. While it was thought earlier [200,228] that thermoelectric-based devices would be more efficient than those based strictly on thermionic phenomena, the mutual equality can now be understood more intuitively [220]. When a solid state material replaces the vacuum in a thermionic device, the limiting factors for maximum efficiency would indeed be similar to that of a thermoelectric.

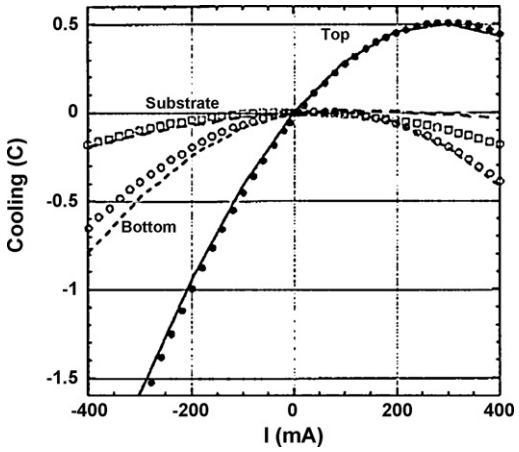

Fig. 54. A cooling of ~0.5 °C was observed by thermionic emission with an n+ InGaAs cathode/InGaAsP barrier/n+ InGaAs anode configuration. Higher cooling efficiencies could be obtained at higher temperatures, due to the waning influence of the parasitics (reprint permission from [224]).

4.5.2. Energy/momentum filtering in superlattice devices

In the heterostructure based coolers considered in the previous section, energy filtering which could yield larger S values along with enhanced efficiencies (see Section 4.3) was postulated to occur due to the selective transmission of electrons. In superlattices, the relevant electron energies from which emission occurs are quantized in the direction perpendicular to growth (say, the z-direction) and only energies/momenta in that direction are relevant. In such devices, assuming atomically smooth interfaces between the constituent thin films, the lateral momentum (k_z) of the carriers is discrete and *conserved*. Additionally, only electrons with energy E ($>E_b$), corresponding to momenta greater than a particular k_z ($>k_b$, corresponding to that of the barrier) surmount the barrier. However, generally, for a given energy E there are a larger number of electrons with a continuous distribution of momenta (k_x, k_y, and k_z) in comparison to when k_z is discrete (and k_x and k_y have a continuous spread). Consequently, a greater current density can be obtained through carrier emission in the former case through a continuous distribution of k_z, i.e., through a *non-conservation* of the lateral momentum [26]. The underlying idea [26] is illustrated in Fig. 55.

When all electron energies, greater than E_b regardless of the momenta, are involved in thermionic emission, k_r filtering is said to have occurred. Referring to Fig. 55, a larger magnitude of electrons in $\mathbf{V_2}$ compared to those in $\mathbf{V_1}$—obtained through k_r filtering, increases emission current densities [229] and concomitantly the power generation or extent of refrigeration.

A possible basis for the expectation of a larger magnitude of currents through such processes is manifested through Ballistic Electron Emission (BEEM) spectroscopy [230] on *non-epitaxial* metal/semiconductor interfaces [231], e.g., Au/GaAs. The emission of ballistic electrons into a solid through vacuum tunneling, from a Scanning Tunneling Microscope (STM), is based on the conservation of the transverse momentum [232] and was expected to populate the Γ valleys, in the GaAs. However, experimental results indicated a considerable electronic current also for the L and the X valleys and were interpreted as due to interfacial electron scattering and non-transverse momentum conserving processes [233], resulting in a redistribution of electrons away from the primary Γ channel.

The conversion of a discrete k_z to a continuous distribution of k_z could then be obtained through breaking the translational

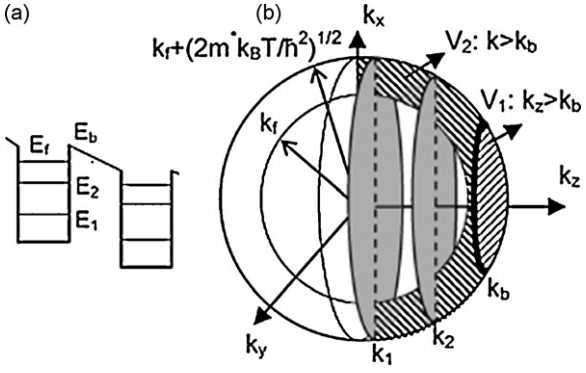

Fig. 55. Electrons of a narrow energy range in the Fermi sphere, corresponding to either a fixed direction (k_x) or a fixed magnitude (k_r) are considered for restricting the energy range of the transmitted electrons (reprint permission from [26]).

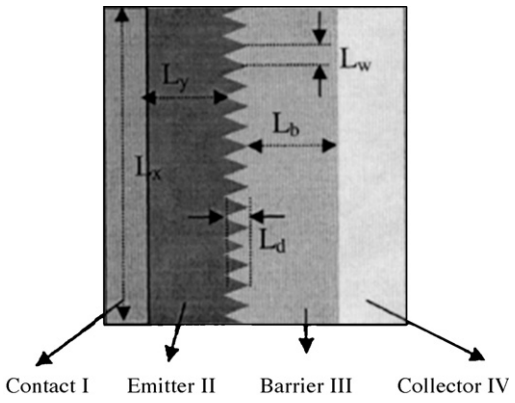

Fig. 56. A non-planar interface can help in breaking the translational invariance of the lateral momentum, k_z, and contribute to greater electronic current transit. In this context, a enhanced degree of roughness, through a smaller L_w and larger L_d, could contribute to enhanced electron emission. The length scales involved in the momentum randomization, at the interface, are intermediate between the electron mean free path and the electron de Broglie wavelength [268]. Such pyramidal structures can be found in lattice mismatched systems such as GaInAs/AlGaAs [269] as self-aligned quantum dots (reprint permission from [234]).

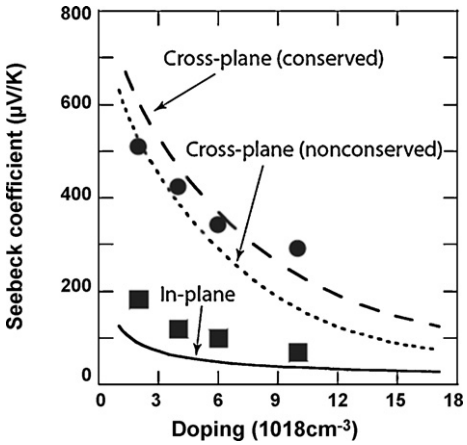

Fig. 58. An enhancement of the Seebeck coefficient in the cross-planar configuration is observed presumably due to enhanced thermionic emission over the barriers [237]. However, the role of ErAs for momentum randomization has not been elucidated (reprint permission from [236]).

invariance at the emitter–barrier interface. This could be accomplished, for example, through the introduction of symmetry breaking non-planar barrier interfaces [234], as illustrated in Fig. 56. The introduction of *controlled roughness* to thus increase electron emission current density, and hence thermionic device efficiency, must be carefully analyzed in view of factors such a reduced thermal conductivity and diminished carrier mobility, along with the possibility of Joule heating. It is also to be noted that the distinction between momentum *conserving* and *non-conserving* processes could be less significant at increased temperatures [235], due to the randomizing effects of thermal energy.

Experimental results in thermionic superlattices, to date, have not demonstrated a very clear efficiency enhancement. In this context, non-planar interfaces for randomizing electron momenta, and increasing total electron current density, were implemented through the growth of sub-monolayer thin films and/or nanoparticles [236], as shown in Fig. 57, e.g., using ErAs based interfaces in InGaAs quantum well/InGaAlAs barrier thermionic superlattices with a conduction band offset of ~0.2 eV. In this case, an enhancement [237] of the S in the *cross-plane* direction (perpendicular to the growth direction) over that in the *in-plane* direction seems to have been observed [236] (Fig. 58). While the enhancement was attributed to electron filtering such a result can also be understood as due to the mode of carrier transport. In the *in-plane* direction, the Seebeck coefficient is the weighted sum of the individual layers' S (from Eq. (35)) and would necessarily be smaller than that for the *cross-plane* direction where the net S is simply additive with the individual S contributions. Additionally, the contribution from enhanced electronic emission is ambiguous between the *conserved* and the *non-conserved* cases. While momentum *non-conservation* along with the possibility of ErAs acting as an electron donors [238] to increase the σ were both considered to be effective, quantitative agreement of the increase of the $S^2\sigma$ from these individual factors is unclear from the presented results [236]. However, an enhanced figure of merit due to a κ_L reduction brought about by ErAs contributing to phonon scattering seems plausible [239]. It was postulated that while alloying was successful in scattering (Rayleigh scattering) the short wavelength/Brillouin zone edge phonons ($\lambda_{ph} \sim 0.1$ nm), ErAs nanoparticles, ~1–4 nm in size, could scatter medium/longer wavelength phonons and hence diminish κ_L. The reduction of the κ_L did seem to be proportional to the (1) amount of introduced ErAs, (2) ErAs particle size, and (3) the randomness of distribution.

Thin film power generators constituted of 400 p- and n-type ErAs:InGaAs/InGaAlAs superlattice (5 μm total thickness, fabricated through MBE and comprising 70 periods of 10 nm InGaAlAs and 20 nm InGaAs) thermoelectric elements were fabricated [240]. For the n-doped material, approximately 0.3–3 vol% of ErAs was randomly distributed in the InGaAs layers, in addition to Si

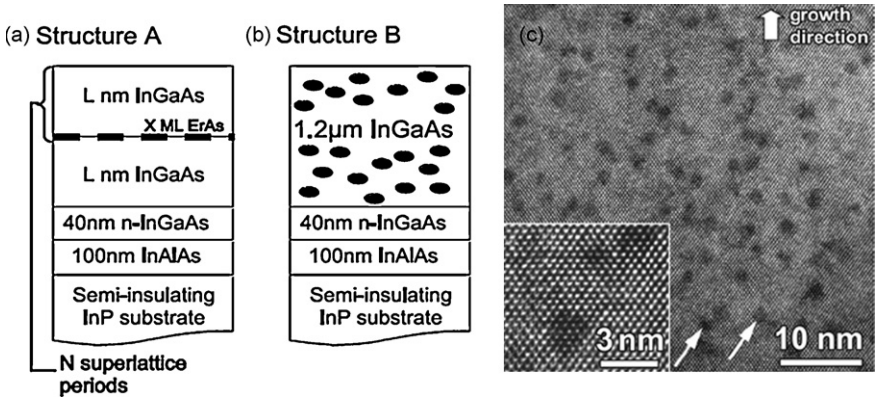

Fig. 57. Multilayer structures, incorporating ErAs monolayers – structure A and, (b) ErAs nanoparticles – structure B, embedded in InGaAs used to probe the effects of electron filtering induced enhancement in the Seebeck coefficient, to explore the effects of non-planar interfaces. (c) The random distribution of the nanoparticles in (b) is illustrated along with an inset of the coherency between the ErAs and the InGaAs matrix (reprint permission from [238]).

co-doping to yield a total carrier concentration ~10^{18} to 10^{19}/cm^3. For the p-type elements, Be was used as a dopant. The InGaAlAs layer was composed of a digital alloy of 60% InGaAs and 40% InAlAs. While it may be surmised that the InGaAs functions as a quantum well, the precise role of the individual layers in analogy to quantum well superlattices is unclear in these experiments. While the COP was not reported, a power density ~2 W/cm^2 was noted (corresponding to a power output of ~0.7 mW from a 0.04 mm^2 device area) with a 30 K temperature difference. Based on thermal simulations, parasitic temperature drop (~85%) across the heater plates, rather than the device, was thought to be responsible for the lower power output. Additionally, as it was seen that the power output increases with temperature, the role [240] of ErAs in reducing the κ_L and improving thermionic efficiency also needs to be elucidated.

When inclusions, e.g., nanoparticles [241], or barriers, e.g., grain boundaries [242], are introduced into the matrix the contrast in the dielectric constant can induce electron scattering at the interfaces. In the case of metallic particles embedded in a semiconductor host, carrier scattering from the interface potential – proportional to the difference of the metal and the semiconductor work functions – was posited to influence the $S^2\sigma$. The S could be enhanced due to a stronger dependence of the relaxation time on energy, e.g., through the $\partial\ln(\tau(E))/\partial\ln(E)$ term in Eq. (31) due to an increase of the r – see Eq. (18). For example, in the case of high energy electrons weakly scattering off the nanoparticles, the Born approximation [243] posits that $\tau(E) \sim E^{3/2}$. The lower energy electrons interact more strongly and could even be backscattered, resulting in a net forward filtering of the electrons. It is expected that such filtering, which could presumably enhance the average electron energy $\langle E \rangle$ and the S (as in Eq. (28)), could be more influential in one-dimensional structures. However, for larger interface potentials, e.g., with Schottky barrier heights >0.3 eV, the Born approximation is not justified and r may not be as large. Additionally, interfacial strain could also influence the energy dependence of the electron scattering and change r.

Experimentally, an enhancement of r from ~1 to ~1.5 was measured in bulk PbTe thermoelectrics with Pb precipitates (20–40 nm) [55,244] through the method of four coefficients (see Section 2.6.1 and Fig. 20), *cf.* in bulk PbTe only, the r was measured to be ~1. While it was shown that the S of the bulk PbTe was almost doubled, at 300 K, from ~250 to ~450 μV/K presumably due to the precipitates, the σ decreased by almost two orders of magnitude from ~10^5 to ~10^3 Ω^{-1} m^{-1}. A similar increase in the S due to a decreased grain size, of ~100 nm, was measured in n-PbTe [242] and ascribed to scattering from the grain boundary interface potential, of ~0.1 eV. Again, a drastic reduction in the σ questions the validity of such approaches in increasing $S^2\sigma$. In the study on InGaAlAs with embedded ErAs nanoparticles (2–3 nm), the S was not increased [245] compared to the samples without the nanoparticles. However, an enhancement in the $S^2\sigma$, up to 10% at 300 K, was attributed to an increased σ due to electron donation from the ErAs to the matrix, as previously noted. An interesting observation was that an increased surface area, corresponding to a larger volume fraction of ErAs particles, could have the same role as the interfacial potential [245].

4.5.3. Influence of interfaces in thermionic devices

The consideration of interfaces is especially important for nanostructures, where the large surface area to volume and contact resistances often dominate electrical properties. It was shown that very small contact resistances (even of the order of 10^{-7} Ω cm^2) can have a deleterious effect on the cooling and the COP of thermionic devices [246], reducing the COP by as much as a factor of 20—*much more significant* than in the case of thermoelectrics. As discussed in Section 3.4.3, any thermal boundary resistance at the emitter–barrier interfaces [247] would reduce the thermal flux by $\delta T/R_B$ where δT is the temperature drop across the interface and R_B is the equivalent resistance of the barrier. An equivalent boundary [248] Seebeck coefficient (S_B) may also be defined, related to the entropy gain of the carriers at the interface, which could add to the bulk/diffusive S. Alternatively, the boundary and the bulk may be considered as parallel conductors and the effective S determined [249] from the formalism outlined earlier in Section 2.5.

4.5.4. Experimental implementations of thermionic cooling

Other superlattice systems, incorporating GaAs/AlGaAs [250], InGaAs/InGaAsP, and Si/Si$_{1-x}$Ge$_x$, have been investigated for thermionic cooling applications. The choice of these material systems is motivated by the possibility of their integration with electronics [198,251] and optoelectronics [252]. A few case studies are presented.

Thin films of Si-doped $Al_{0.1}Ga_{0.9}As$ (10 nm) quantum well—$Al_{0.2}Ga_{0.8}As$ (10 nm) barrier layers were alternately grown as a superlattice to a total thickness of 1 μm on GaAs substrates (with a thermionic barrier height ~85 meV) using MOCVD (Fig. 59). While the COP or power output density was not reported, a temperature reduction of approximately 0.8–2 K was observed [253] for device areas ~3600 μm^2. However, parasitic heating from the contacts needed to be better understood. Additionally, the large current densities, ~10^4 A/cm^2, would imply large Joule heating and contact resistance becomes even more important (see Section 4.5.3). While a number of theoretical proposals exist for using GaAs/AlGaAs layers, e.g., incorporating injection of low energy electrons via resonant tunneling and withdrawal of the high energy electrons through thermionic processes [254], the expected cooling ($\Delta T < 5$ K) seems to be quite low. It would be useful, and perhaps necessary, to estimate the value of device efficiency/COP to enable easy comparison across a variety of theoretical proposals. This is especially important as the cooling, in terms of a temperature drop or cooling power density, is dependent on specific materials and geometry.

Micro-coolers comprised of 25 periods of $In_{0.53}Ga_{0.47}As$ quantum well (5 nm, Si doped)/$Al_{0.52}Ga_{0.48}As$ (3 nm, undoped) barrier layer superlattices were grown through MBE on InGaAs buffer layer/InP substrates and were then fused, through eutectic bonding, with an optoelectronic module on a chip [255]. A cooling of ~0.8 K was obtained in a device of size 1600 μm^2, with smaller cooling observed in larger devices. The cooling power density (Q_c) was then calculated, from measured values of the current (I) and the S, along with estimated values of the heat backflow (=0.5I^2R) for a given electrical resistance (R) and thermal conductance (K), through $Q_c = ST_cI - (1/2)I^2R - K\Delta T$, to be of the order of 50 W/cm^2

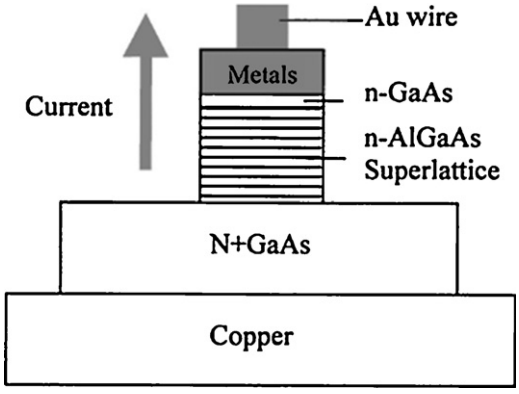

Fig. 59. A superlattice cooler, comprised of alternate layers of n-$Al_{0.1}Ga_{0.9}As$ (10 nm) and n-$Al_{0.2}Ga_{0.8}As$ (10 nm), was used to obtain a maximum cooling of 0.8 K at 25 °C and 2 K at 100 °C (reprint permission from [250]).

Table 5
A comparison of microrefrigerators fabricated from $Si_{0.8}Ge_{0.2}$ alloy and a $Si/Si_{0.75}Ge_{0.25}$ superlattice indicates comparable figures of merit (reprint permission from [260]).

Material	Seebeck coefficient, S (μV/K)	Electrical conductivity, σ $(\Omega\,\mathrm{cm})^{-1}$	Thermal conductivity, β (W/mK)	Power factor, $S^2\sigma$ (10^{-3} W/km)	Figure of marit, $ZT = S^2\sigma T/\beta$
$Si_{0.8}Ge_{0.2}$alloy (microrefrigerator ΔT_{max} = 40 K)	210 210[a]	367 375	5.9 6[a]	1.6 1.7[b]	0.08 0.083[b]
Superlattice $Si/Si_{075}Ci_{0.25}$ (3 mm/12 nm) (microrefrigerator ΔT_{max} = 4.2 K)	200–200(⊥) 235[a] 180(||)	384[a] 384(||)	6.8–8.7(⊥) 8[a]	228(⊥, estimated) 1.2(||) 2.1	0.085(⊥, estimated) 0.080[b]

[a]Values used to get the best fit for all microrefrigerator sizes.
[b]Calculated power factors and ZTs based on the best fit parameters.
||Refers to in-plane material properties, while.
⊥Refers to cross-plane material properties.

with a COP of ~0.3. It was then predicted that reducing the parasitics along with an increased superlattice thickness of ~10 μm could increase the cooling power densities to >200 W/cm^2, with the COP approaching 5. In another study [256], a cooling of ~0.5 K at 25 °C was observed in a single-stage p+ InGaAs *cathode*/p InGaAsP (1 μm)/p+ InGaAs *anode* structure in a 150 μm × 150 μm mesa structure. The COP or cooling powers were not reported in the latter instance.

$Si/Si_{1-x}Ge_x$ superlattice-based thermionic devices have an obvious advantage in that they can be seamlessly integrated with Silicon based electronics and optoelectronics, and are amenable for higher temperature (>1000 K) operation [257]. A 3 μm thick 200 period (3 nm Si/12 nm $Si_{0.75}Ge_{0.25}$) MBE grown superlattice integrated, through a buffer layer, onto a Si substrate was experimentally shown to exhibit a cooling of up to 3.5 K in a 3600 μm^2 device [258]. A transient response of ~20–30 μs (as measured through the rise/fall of the Seebeck voltage generated through the application of a square wave heating pulse to the thermionic device) was measured to indicate quick heat removal and could also presumably be due to the small device size. The maximum Q_c and the COP were estimated [221] at 600 W/cm^2 and 0.36, respectively. In the $Si/Si_{1-x}Ge_x$ devices discussed thus far, to avoid the lattice parameter mismatch between the Ge and Si and consequent dislocation formation, the growth of additional $Si_{1-x}Ge_x$ buffer layers is necessary. This additional growth step was sought to be eliminated through the synthesis of a lattice matched $Si/Si_{1-x-y}Ge_xC_y$ superlattice on a Si substrate. In comparison to $Si/Si_{1-x}Ge_x$, the $Si/Si_{1-x-y}Ge_xC_y$ interface has *both* an increased conduction band and a valence band offset, with the implication that both hole and electron emission could be used for thermionic cooling. Thermal measurements on a 3600 μm^2 area device comprised of 2 μm thick (d) 100 period MBE grown superlattice (10 nm Si/10 nm $Si_{0.89}Ge_{0.10}C_{0.01}$) indicated a cooling of ~2.5 K, at 298 K (and ~6.9 K, at 373 K) [259]. The cooling densities were estimated, by the equation, $Q_c = \kappa(\Delta T_{max} - \Delta T)/d$, to be of the order of 1000 W/cm^2. While the COP was not indicated, the low thermal conductivity of the superlattice (~8.5 W/mK) could also have been important for the enhanced Q_c.

In the above studies, it was noticed that a clear appraisal of the effects of the quantum well and the barrier layers were not often made. Consequently, it is not easy to understand the individual and relative contributions to the device efficiency/cooling power density from, (1) Peltier cooling at the superlattice/contact or superlattice/substrate interfaces, (2) thermionic carrier emission, (3) thermoelectric effects, e.g., an enhanced power factor due to contributions from an increased density of states, or (4) from the reduced thermal conductivity, which prevents heat backflow. A case in point is a recent comparison [260] of the cooling performance of $Si/Si_{0.75}Ge_{0.25}$ superlattice based devices with bulk $Si_{0.8}Ge_{0.2}$ alloy, which seemed to indicate comparable efficiencies, for similar device sizes (Table 5).

It can also be concluded that non-ideal factors, such as contact and thermal resistances, Joule heating, heat backflow, etc., are especially pernicious in thin film based devices [261]. Much work then needs to be done in overcoming such parasitic effects to yield reliable and enhanced cooling power densities in thermionic devices [262].

5. Conclusions

In this article, we have sought to understand how thermoelectric and thermionic nanostructures could yield higher energy conversion efficiencies. A thorough investigation, using the semi-classical Boltzmann transport equation as a basis, was employed to understand the influences of both the carriers (electrons) and the lattice (phonons) in various dimensions and length scales. It was seen that the variation of the electrical conductivity (σ) and the Seebeck coefficient (S) as a function of the reduced Fermi energy (η) serves as a metric for the optimization of the power factor, $S^2\sigma$. It is often easier to understand the variation of S^2n, which can be experimentally measured, as the mobility (μ) is a function of the specific material configuration which is quite difficult to control at the nanoscale. By considering the influence of quantum confinement in enhancing the density of states (DOS) and the power factor, we have postulated that *it is the magnitude and not the specific shape of the DOS that is important*. In this context, the dimensionality and particular carrier scattering mechanisms are very relevant and could cause deviations from paradigms such as the Wiedemann–Franz type relations. We have also noted that maximization of S^2n involves an optimization of both S and n, which vary oppositely with η and implied an optimal S. A survey of experimental results in low dimensional materials systems, e.g., superlattices and quantum wells, indicated that the experimental conditions were far from optimal. It was concluded that there is a paucity of detailed experimental data and interpretation. Lower dimensional systems, such as nanowires or quantum dots, are just beginning to be investigated but much of the improvement, especially in the former, seems to arise due to a contribution from a reduced lattice thermal conductivity, κ_L.

A consideration of the classical and quantum mechanical aspects of the κ_L yields much insight into reducing the thermal conductivity from the purpose of minimizing parasitic heat transport and increasing the figure of merit, ZT. Experimentally, the reduction of the mean free path seems to be a widely used concept in nanostructured thermoelectrics. Another interesting idea deals with reducing κ_L in nanostructures, up to two orders of magnitude, close to the lowest possible value of $\kappa_{L,\,min}$ through the introduction of scattering at different scales, e.g., atomic scattering of short wavelength phonons by compositional fluctuations, surface roughness for longer wavelength phonons, and boundary scattering. However, for further maximization of the ZT it is not clear whether this route is promising as limiting values of κ_L seem

to have been reached. Roughness, for example, could also deteriorate the electrical conductivity. Avenues such as phononic bandgaps and phonon localization which may not adversely affect electrical parameters should be investigated.

Vacuum-based thermionics, which has the merit of ideally obtaining close to zero thermal conductances for reduced heat backflow, does not seem to be feasible for large scale usage due to major difficulties in fabrication coupled with fundamental limitations such as large work function for the emitter. Solid state thermionics, using a low thermal conductivity barrier layer instead of vacuum, have been implemented in compound semiconductor/heterostructure superlattices. However, such implementations seem to be closely allied to thermoelectric devices and have recently begun to be described in literature using similar terminology. Many of the theoretical predictions on enhancing performance, incorporating multilayer refrigeration and energy filtering, still await experimental validation.

In summary, nanostructures offer a means of tuning the energy conversion efficiency beyond those offered by bulk materials. The promise of approaching efficiencies comparable to those in mechanical cycle based engines along with possible applications to solid state based waste heat recovery, power generation, and to the reduction of global warming [263] makes nanostructured thermoelectrics an exciting field of study. For example, it has been estimated that ~500 million barrels of oil can be saved per year through the use of thermoelectrics. In the United States, thermoelectric devices are considered to be a $5 billion industry with a potential worldwide market of around $50 billion [264], which provides sufficient economic imperative. However, it seems that the strategies for the widespread utilization of nanostructured thermoelectrics are at quite a seminal stage and considerable experimental effort is necessary. In this context, a cause for concern in the ultimate application of stand-alone nanostructured thermoelectrics is the minimum volume of material needed for sustained temperature gradients to be maintained. In this connection, *nano-bulk* thermoelectrics, incorporating nanoscale features in traditional bulk materials, could be a viable short-term alternative. Efforts to integrate thermoelectrics with areas of non-traditional usage, e.g., with photovoltaics for harnessing the IR and far-IR regions of the solar spectrum [265] might be attractive on a longer term.

References

[1] C. Wood, Reports on Progress in Physics 51 (1988) 459–539.
[2] D.K.C. MacDonald, Thermoelectricity: An Introduction to the Principles, Dover Press, Mineola, NY, 1962.
[3] C. Herring, Physical Review 96 (1954) 1163–1187.
[4] A.I. Boukai, Y. Bunimovich, J. Tahir-Kheli, J.-K. Yu, W.A. Goddard III, J.R. Heath, Nature 451 (2008) 168–171.
[5] A.F. Ioffe, Semiconductor Thermoelements and Thermoelectric Cooling, Infosearch Ltd., London, 1957.
[6] G.S. Nolas, J. Sharp, H.J. Goldsmid, Thermoelectrics—Basic Principles and New Materials Development, Springer, New York, 2001.
[7] D.M. Rowe (Ed.), Thermoelectrics Handbook: Macro to Nano, Taylor & Francis, 2006.
[8] G.D. Mahan, Good Thermoelectrics, Solid State Physics, vol. 51, Academic Press, 1997, pp. 82–157.
[9] J. Singleton, Band Theory and Electronic Properties of Solids, Oxford University Press, New York, 2004.
[10] T.E. Humphrey, H. Linke, Physical Review Letters 94 (2005) 096601.
[11] 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, Z. Ren, Science 320 (2008) 634–637.
[12] R. Venkatasubramanian, E. Siivola, T. Colpitts, B. O'Quinn, Nature 413 (2001) 597–602.
[13] A.I. Hochbaum, R. Chen, R.D. Delgado, W. Liang, E.C. Garnett, M. Najarian, A. Majumdar, P. Yang, Nature 451 (2008) 163–167.
[14] L.D. Hicks, T.C. Harman, M.S. Dresselhaus, Applied Physics Letters 63 (1993) 3230–3232.
[15] L.D. Hicks, T.C. Harman, X. Sun, M.S. Dresselhaus, Physical Review B 53 (1996) R10493–R10496.
[16] T.C. Harman, P.J. Taylor, M.P. Walsh, B.E. La Forge, Science 297 (2002) 2229–2232.
[17] J. Androulakis, C.H. Lin, H.J. Kong, C. Uher, C.I. Wu, T. Hogan, B.A. Cook, T. Caillat, K.M. Paraskevopoulos, M.G. Kanatzidis, Journal of the American Chemical Society 129 (2007) 9780–9788.
[18] N.W. Ashcroft, N.D. Mermin, Solid State Physics, Saunders College, Orlando, FL, 1976.
[19] H.J. Goldsmid, Applications of Thermoelectricity, Butler & Tanner Ltd., 1960.
[20] A. Bulusu, D.G. Walker, IEEE Transaction on Electron Devices 55 (2008) 423.
[21] A. Bulusu, D.G. Walker, Journal of Heat Transfer 129 (2007) 492–499.
[22] M.J. Kelly, Low-Dimensional Semiconductors: Materials, Physics, Technology, Devices (Series on Semiconductor Science and Technology), Oxford University Press, 1996.
[23] D.L. Young, T.J. Coutts, V.I. Kaydanov, A.S. Gilmore, W.P. Mulligan, Journal of Vacuum Science & Technology A-Vacuum Surfaces and Films 18 (2000) 2978.
[24] M. Lundstrom, Fundamentals of Carrier Transport, Cambridge University Press, Cambridge, 2000.
[25] N. Mohankumar, A. Natarajan, Physica Status Solidi B: Basic Research 188 (1995) 635–644.
[26] D. Vashaee, A. Shakouri, Physical Review Letters 92 (2004) 106103.
[27] J.P. Heremans, V. Jovovic, E.S. Toberer, A. Saramat, K. Kurosaki, A. Charoenpakdee, S. Yamanaka, G.J. Snyder, Science 321 (2008) 554.
[28] A. Ishida, D. Cao, S. Morioka, M. Veis, Y. Inoue, T. Kita, Applied Physics Letters 92 (2008) 182105.
[29] L. Friedman, Journal of Physics C 17 (1984) 3999.
[30] Y.I. Ravich, in: D.M. Rowe (Ed.), CRC Handbook of Thermoelectrics, CRC Press, Boca Raton, FL, 1995.
[31] N.F. Mott, H. Jones, The Theory of The Properties of Metals and Alloys, Dover Publications, New York, 1936.
[32] G.D. Mahan, J.O. Sofo, The Best Thermoelectric, vol. 93, National Academy of Science, 1996,, p. 7436.
[33] P. Pichanusakorn, P. Bandaru, Applied Physics Letters 94 (2009) 223108.
[34] R. Kim, S. Datta, M.S. Lundstrom, Journal of Applied Physics 105 (2009) 034506.
[35] T.C. Harman, D.L. Spears, M.J. Manfra, Journal of Electronic Materials 25 (1996) 1121–1127.
[36] J.P. Dismukes, L. Ekstrom, E.F. Steigmeier, I. Kudman, D.S. Beers, Journal of Applied Physics 35 (1964) 2899–2907.
[37] T. Okuda, K. Nakanishi, S. Miyasaka, Y. Tokura, Physical Review B 63 (2001) 113104.
[38] S. Ohta, T. Nomura, H. Ohta, M. Hirano, H. Hosono, K. Koumoto, Applied Physics Letters 87 (2005) 092108.
[39] L.D. Hicks, M.S. Dresselhaus, Physical Review B 47 (1993) 12727–12731.
[40] L.D. Hicks, M.S. Dresselhaus, Physical Review B 47 (1993) 16631–16634.
[41] D.A. Broido, T.L. Reinecke, Physical Review B 64 (2001) 45324.
[42] T. Koga, X. Sun, S.B. Cronin, M.S. Dresselhaus, Applied Physics Letters 75 (1999) 2438.
[43] T. Koga, X. Sun, S.B. Cronin, M.S. Dresselhaus, Applied Physics Letters 73 (1998) 2950.
[44] L.D. Hicks, T.C. Harman, M.S. Dresselhaus, Applied Physics Letters 63 (1993) 3230.
[45] C.F. Gallo, B.S. Chandrasekhar, P.H. Sutter, Journal of Applied Physics 34 (1963) 144–152.
[46] M.V. Vedernikov, O.N. Uryupin, B.M. Goltsman, Y.V. Ivanov, Y.A. Kumzerov, in: Proceedings of the International Conference on Thermoelectric, vol. 19, 2001, pp. 361–363.
[47] Y.M. Lin, O. Rabin, S.B. Cronin, Y.Y. Jackie, M.S. Dresselhaus, Applied Physics Letters 81 (2002) 2403–2405.
[48] P.R. Bandaru, Journal of Nanoscience and Nanotechnology 7 (2007) 1239–1267.
[49] T. Thonhauser, T.J. Scheidemantel, J.O. Sofo, Applied Physics Letters 85 (2004) 588–590.
[50] J.O. Sofo, G.D. Mahan, Applied Physics Letters 65 (1994) 2690.
[51] D.A. Broido, T.L. Reinecke, Physical Review B 51 (1995) 13797–13800.
[52] D.A. Broido, T.L. Reinecke, Applied Physics Letters 67 (1995) 100–102.
[53] D.A. Broido, T.L. Reinecke, Applied Physics Letters 70 (1997) 2834.
[54] Annual Book of ASTM Standards, American Society of Testing and Materials, 1996, pp. 70–82.
[55] J.P. Heremans, C.M. Thrush, D.T. Morelli, Physical Review B 70 (2004) 115334.
[56] S. Yuan, G. Springholz, G. Bauer, M. Kriechbaum, Physical Review B 49 (1994) 5476.
[57] A. Casian, I. Sur, H. Scherrer, Z. Dashevsky, Physical Review B 61 (2000) 15965.
[58] T.C. Harman, D.L. Spears, M.P. Walsh, Journal of Electronic Materials 28 (1999) L1.
[59] T.C. Harman, P.J. Taylor, D.L. Spears, M.P. Walsh, Journal of Electronic Materials 29 (2000) L1–L2.
[60] C.J. Vineis, T.C. Harman, S.D. Calawa, M.P. Walsh, R.E. Reeder, R. Singh, A. Shakouri, Physical Review B: Condensed Matter and Materials Physics 77 (2008) 235202.
[61] T.C. Harman, P.J. Taylor, M.P. Walsh, B.E. LaForge, Science 297 (2002) 2229.
[62] S.A. Nemov, Y.I. Ravich, Physics-Uspekhi 41 (1998) 735–759.
[63] F. Schaffler, in: M.E. Levinshtein, S.L. Rumyantsev, M.S. Shur (Eds.), Properties of Advanced Semiconductor Materials, John Wiley & Sons, Inc., 2001 , pp. 149–186.
[64] H. Kato, A. Yamamoto, M. Takimoto, T. Ohta, K. Sakamoto, K. Miki, L. Whitlow, K. Kamisako, T. Matsui, in: Proceedings of the 17th International Conference on Thermoelectrics, 1998, pp. 253–256.

[65] X. Sun, S.B. Cronin, J. Liu, K.L. Wang, T. Koga, M.S. Dresselhaus, G. Chen, in: Proceedings of the 18th International Conference on Thermoelectrics, IEEE, 1999.
[66] J.H. Davies, The Physics of Low-Dimensional Semiconductors, Cambridge Press, Cambridge, 1998.
[67] W.R. Thurber, R.L. Mattis, Y.M. Liu, J.J. Filliben, The Relationship Between Resistivity and Dopant Density for Phosphorous and Boron-Doped Silicon, N.B.O.S. Department of Commerce, U.S. Government Printing Office, 1981.
[68] R. Venkatasubramanian, E. Siivola, T.S. Colpitts, in: Proceedings ICT 98. XVII International Conference, 1998, pp. 191–197.
[69] T. Koga, S.B. Cronin, M.S. Dresselhaus, J.L. Liu, K.L. Wang, Applied Physics Letters 77 (2000) 1490.
[70] H.P.R. Frederikse, W.R. Thurber, W.R. Hosler, Physical Review 134 (1964) A442.
[71] S. Ohta, T. Nomura, H. Ohta, K. Koumoto, Journal of Applied Physics 97 (2005) 034106.
[72] H. Ohta, S. Kim, Y. Mune, T. Mizoguchi, K. Nomura, S. Ohta, T. Nomura, Y. Nakanishi, Y. Ikuhara, M. Hirano, H. Hosono, K. Koumoto, Nature Materials 6 (2007) 129.
[73] Y. Mune, H. Ohta, K. Koumoto, T. Mizoguchi, Y. Ikuhara, Applied Physics Letters 91 (2007) 192105.
[74] S.B. Cronin, Y.M. Lin, T. Koga, X. Sun, J.Y. Ying, M.S. Dresselhaus, C. Mit, in: Proceedings of the International Conference on Thermoelectrics, vol. 18, 1999, pp. 554–557.
[75] Z. Zhang, X. Sun, M.S. Dresselhaus, J.Y. Ying, J.P. Heremans, Applied Physics Letters 73 (1998) 1589.
[76] X. Sun, Z. Zhang, M.S. Dresselhaus, Applied Physics Letters 74 (1999) 4005–4007.
[77] Z. Zhang, X. Sun, M.S. Dresselhaus, J.Y. Ying, J. Heremans, Physical Review B 61 (2000) 4850.
[78] Y.M. Lin, X. Sun, M.S. Dresselhaus, Physical Review B 62 (2000) 4610–4623.
[79] Y.M. Lin, S.B. Cronin, J.Y. Ying, M.S. Dresselhaus, J.P. Heremans, Applied Physics Letters 76 (2000) 3944–3946.
[80] J. Heremans, C.M. Thrush, Y.-M. Lin, S. Cronin, Z. Zhang, M.S. Dresselhaus, J.F. Mansfield, Physical Review B 61 (2000) 2921.
[81] A. Boukai, K. Xu, J.R. Heath, Advanced Materials 18 (2006) 864–869.
[82] T.E. Huber, A. Nikolaeva, D. Gitsu, L. Konopko, C.A. Foss Jr., M.J. Graf, Applied Physics Letters 84 (2004) 1326.
[83] T.E. Huber, M.J. Graf, C.A. Foss Jr., P. Constant, in: Proceedings of the Material Research Society Symposium, vol. 626, 2000, p. Z14.12.11.
[84] J.P. Heremans, C.M. Thrush, D.T. Morelli, M.-C. Wu, Physical Review Letters 88 (2002) 216801.
[85] D. Dragoman, M. Dragoman, Applied Physics Letters 91 (2007) 203116.
[86] M.R. Black, M. Padi, S.B. Cronin, Y.M. Lin, O. Rabin, T. McClure, G. Dresselhaus, P.L. Hagelstein, M.S. Dresselhaus, Applied Physics Letters 77 (2000) 4142.
[87] M.R. Black, P.L. Hagelstein, S.B. Cronin, Y.M. Lin, M.S. Dresselhaus, Physical Review B 68 (2003) 235417.
[88] C. Kittel, Introduction to Solid State Physics, John Wiley, New York, 1996.
[89] H. Kroemer, C. Kittel, Thermal Physics, W.H. Freeman and Co., New York, 1980
[90] J.M. Ziman, Electrons and Phonons: The theory of transport phenomena in solids, Clarendon Press, Oxford, 1960.
[91] D.W. Oliver, G.A. Slack, Journal of Applied Physics 37 (1966) 1542–1548.
[92] M. Roufosse, P.G. Klemens, Physical Review B 7 (1973) 5379.
[93] G.A. Slack, Solid State Physics 34 (1979) 1–70.
[94] Y.P. Joshi, G.S. Verma, Physical Review B 1 (1970) 750–755.
[95] R.A.H. Hamilton, J.E. Parrott, Physical Review 178 (1969) 1284–1292.
[96] I. Pomeranchuk, Physical Review 60 (1941) 820–821.
[97] C. Herring, Physical Review 95 (1954) 954–965.
[98] H.B. Callen, Thermodynamics, John Wiley Inc., New York, 1960.
[99] D.G. Cahill, S.K. Watson, R.O. Pohl, Physical Review B: Condensed Matter 46 (1992) 6131–6140.
[100] C. Kittel, Physical Review 75 (1949) 972–974.
[101] G.D. Mahan, F. Claro, Physical Review B 38 (1988) 1963.
[102] G. Chen, Journal of Heat Transfer 118 (1996) 539–545.
[103] A.A. Joshi, A. Majumdar, Journal of Applied Physics 74 (1993) 31–39.
[104] A. Majumdar, Journal of Heat Transfer 115 (1993) 7–16.
[105] S. John, Physics Today (May) (1991) 32–40.
[106] G. Chen, Transactions of the ASME 119 (1997) 220–229.
[107] M.V. Simkin, G.D. Mahan, Physical Review Letters 84 (2000) 927–930.
[108] L.G.C. Rego, G. Kirczenow, Physical Review Letters 81 (1998) 232–235.
[109] K. Schwab, E.A. Hendriksen, J.M. Worlock, M.L. Roukes, Nature 404 (2000) 974–977.
[110] K. Schwab, E.A. Henriksen, J.M. Worlock, M.L. Roukes, Nature 404 (2000) 974–977.
[111] J.S. Dugdale, D.K.C. MacDonald, Physical Review 98 (1955) 1751.
[112] A.W. Lawson, Journal of the Physics and Chemistry of Solids 3 (1957) 155–156.
[113] R.W. Keyes, Physical Review 115 (1959) 564.
[114] C. Uher, Semiconductors and Semimetals 69 (2001) 139–253.
[115] G.A. Slack, Journal of Applied Physics 76 (1994) 1665–1671.
[116] G.A. Slack, in: D.M. Rowe (Ed.), CRC Handbook of Thermoelectrics, 1995, 1–34.
[117] C.B. Vining, Nature Materials 7 (2008) 765–766.
[118] M.M. Koza, M.R. Johnson, R. Viennois, H. Mutka, L. Girard, D. Ravot, Nature Materials 7 (2008) 805–810.
[119] B.C. Sales, B.C. Chakoumakos, D. Mandrus, Physical Review B 61 (2000) 2475.
[120] V. Keppens, D. Mandrus, B.C. Sales, B.C. Chakoumakos, P. Dai, R. Coldea, M.B. Maple, D.A. Gajewski, E.J. Freeman, S. Bennington, Nature 395 (1998) 876–878.
[121] R.P. Hermann, F. Grandjean, G.J. Long, American Journal of Physics 73 (2005) 110–118.
[122] M. Christensen, A.B. Abrahamsen, N.B. Christensen, F. Juranyi, N.H. Andersen, K. Lefmann, J. Andreasson, C.R.H. Bahl, B.B. Iversen, Nature Materials 7 (2008) 811–815.
[123] M.-S. Jeng, R. Yang, R. Song, G. Chen, Journal of Heat Transfer 130 (2008) 042410.
[124] S. Mazumder, A. Majumdar, Journal of Heat Transfer 123 (2001) 749–759.
[125] R.S. Prasher, Transactions of the ASME 125 (2003) 1156–1162.
[126] K.F. Hsu, S. Loo, F. Guo, W. Chen, J.S. Dyck, C. Uher, T. Hogan, E.K. Polychroniadis, M.G. Kanatzidis, Science 303 (2004) 818–821.
[127] D.A. Porter, K.E. Easterling, Phase Transformations in Metals and Alloys, Nelson Thornes, Cheltenham, UK, 1992.
[128] E. Quarez, K.F. Hsu, R. Pcionek, N. Frangis, E.K. Polychroniadis, M.G. Kanatzidis, Journal of the American Chemical Society 127 (2005) 9177–9190.
[129] I. Bonev, Acta Crystallographica A28 (1972) 508–512.
[130] J. Androulakis, K.F. Hsu, R. Pcionek, H. Kong, C. Uher, J. D'Angelo, A.D. Downey, T. Hogan, M.G. Kanatzidis, Advanced Materials 18 (2006) 1170–1173.
[131] P.F.P. Poudeu, J. D'Angelo, H. Kong, A. Downey, J.L. Short, R. Pcionek, T.P. Hogan, C. Uher, M.G. Kanatzidis, Journal of the American Chemical Society 128 (2006) 14347–14355.
[132] P.F.P. Poudeu, J. D'Angelo, A. Downey, J.L. Short, T. Hogan, M.G. Kanatzidis, Angewandte Chemie International Edition 45 (2006) 3835–3839.
[133] C.R.M. Grovenor, Journal of Physics C: Solid State Physics 18 (1985) 4079–4119.
[134] G. Chen, Physical Review B 57 (1998) 14958–14973.
[135] G.L. Pollack, Reviews of Modern Physics 41 (1969) 48–81.
[136] R.J. Stoner, H.J. Maris, Physical Review B: Condensed Matter 48 (1993) 16373–16387.
[137] E.T. Swartz, R.O. Pohl, Reviews of Modern Physics 61 (1989) 605–668.
[138] W.A. Little, Canadian Journal of Physics 37 (1959) 334–349.
[139] S.-M. Lee, D.G. Cahill, Journal of Applied Physics 81 (1997) 2590–2595.
[140] S.M. Lee, D.G. Cahill, Applied Physics Letters 70 (1997) 2957–2959.
[141] Y.K. Koh, Y. Cao, D.G. Cahill, D. Jena, Advanced Functional Materials 19 (2009) 610–615.
[142] M.J. Kelly, Journal of Physics C: Solid State Physics 18 (1985) 5965–5973.
[143] Y. Chen, D. Li, J. Yang, Y. Wu, J.R. Lukes, A. Majumdar, Physica B: Condensed Matter (Amsterdam) 349 (2004) 270–280.
[144] S. Volz, G. Chen, Applied Physics Letters 75 (1999) 2056–2058.
[145] S. Volz, J.B. Saulnier, G. Chen, P. Beauchamp, Microelectronics Journal 31 (2000) 815–819.
[146] L. Braginsky, N. Lukzen, V. Shklover, H. Hofmann, Physical Review B 66 (2002) 134203.
[147] Z. Zhong, X. Wang, Journal of Applied Physics 100 (2006) 044310.
[148] M.S. Toprak, C. Stiewe, D. Platzek, S. Williams, L. Bertini, E. Muller, C. Gatti, Y. Zhang, M. Rowe, M. Muhammed, Advanced Functional Materials 14 (2004) 1189–1196.
[149] C.-W. Nan, R. Birringer, Physical Review B 57 (1998) 8264–8268.
[150] G. Chen, C.-L. Tien, X. Wu, J.S. Smith, Journal of Heat Transfer 116 (1994) 325–331.
[151] C. Weisbuch, B. Vinter, Quantum Semiconductor Structures: Fundamentals and Applications, Academic Press, New York, 1991.
[152] M.J. Kelly, Low-Dimensional Semiconductors, Oxford University Press, New York, 1995.
[153] T. Yao, Applied Physics Letters 51 (1987) 1798–1800.
[154] W.S. Capinski, H.J. Maris, Physica B: Condensed Matter (Amsterdam) 219–220 (1996) 699–701.
[155] D.G. Cahill, Review of Scientific Instruments 61 (1990) 802–808.
[156] S.T. Huxtable, A.R. Abramson, C.-L. Tien, A. Majumdar, C. LaBounty, X. Fan, G. Zeng, J.E. Bowers, A. Shakouri, E.T. Croke, Applied Physics Letters 80 (2002) 1737–1739.
[157] M. Ohring, Materials Science of Thin Films, Academic Press, San Diego, 2002.
[158] P.G. Klemens, Proceedings of the Physical Society (London) A 68 (1955) 1113–1128.
[159] T. Borca-Tasciuc, W. Liu, J. Liu, T. Zeng, D.W. Song, C.D. Moore, G. Chen, K.L. Wang, M. Goorsky, T. Radetic, R. Gronsky, T. Koga, M.S. Dresselhaus, Superlattices and Microstructures 28 (2000) 199–206.
[160] R. Venkatasubramanian, Physical Review B 61 (2000) 3091–3097.
[161] M.N. Touzelbaev, P. Zhou, R. Venkatasubramanian, K.E. Goodson, Journal of Applied Physics 90 (2001) 763–767.
[162] A.L. Moore, S.K. Saha, R.S. Prasher, L. Shi, Applied Physics Letters 93 (2008) 083112–083113.
[163] T.C. Harman, P.J. Taylor, D.L. Spears, M.P. Walsh, Journal of Electronic Materials 29 (2000) L1–L4.
[164] R. Hummel, Electronic Properties of Materials, Springer, New York, 2000.
[165] K. Imamura, Y. Tanaka, S. Tamura, Physical Review B 65 (2002) 174301–174310.
[166] V. Narayanamurti, H.L. Stormer, M.A. Chin, A.C. Gossard, W. Wiegmann, Physical Review Letters 43 (1979) 2012–2016.
[167] D.C. Hurley, S. Tamura, J.P. Wolfe, K. Ploog, J. Nagle, Physical Review B 37 (1988) 8829–8836.
[168] D.C. Hurley, S. Tamura, J.P. Wolfe, H. Morkoc, Physical Review Letters 58 (1987) 2446–2449.
[169] J.D. Joannopoulos, R.D. Meade, J.N. Winn, Photonic Crystals: Molding The Flow of Light, Princeton University Press, Princeton, NJ, 1995.
[170] S. Mizuno, S. Tamura, Physical Review B 45 (1992) 734–741.
[171] R. Orbach, Journal of Non-Crystalline Solids 164–166 (1993) 917–922.
[172] A. Majumdar, Microscale Thermophysical Engineering 2 (1998) 5–9.
[173] A.A. Balandin, K.L. Wang, Physical Review B 58 (1998) 1544–1549.
[174] B. Taylor, H.J. Maris, C. Elbaum, Physical Review Letters 23 (1969) 416–419.
[175] Y.P. Joshi, Pramana 18 (1982) 461–472.

[176] A.A. Balandin, Journal of Nanoscience and Nanotechnology 5 (2005) 1015–1022.
[177] X. Lu, Applied Physics Letters 104 (2008) 054314.
[178] P. Kim, L. Shi, A. Majumdar, P.L. McEuen, Physical Review Letters 87 (2001) 215502.
[179] H. Maune, H.-Y. Chiu, M. Bockrath, Applied Physics Letters 89 (2006) 013109.
[180] R.S. Prasher, Physical Review B 77 (2008) 075424.
[181] F.M. Ross, J. Tersoff, M.C. Reuter, Physical Review Letters 95 (2005) 146104.
[182] R.S. Prasher, Applied Physics Letters 89 (2006) 063121.
[183] R. Yang, M.S. Dresselhaus, G. Chen, Physical Review B 72 (2005) 125418.
[184] E.P. Pokatilov, D.L. Nika, A.A. Balandin, Applied Physics Letters 85 (2004) 825–827.
[185] L.J. Lauhon, M.S. Gudiksen, D. Wang, C.M. Lieber, Nature 420 (2002) 57–61.
[186] P. Carruthers, Reviews of Modern Physics 33 (1961) 92–138.
[187] L. Gurevich, Journal of Physics IX (1945) 477–488.
[188] H.P.R. Frederikse, Physical Review 92 (1953) 248–252.
[189] E. Behnen, Journal of Applied Physics 87 (1990) 287–292.
[190] T.H. Geballe, G.W. Hull, Physical Review 94 (1954) 1134–1140.
[191] R. Trzcinski, E. Gmelin, H.J. Queisser, Physical Review Letters 56 (1986) 1086–1089.
[192] C. Zener, Physical Review 52 (1937) 230–235.
[193] R. Lifshitz, M.L. Roukes, Physical Review B 61 (2000) 5600–5609.
[194] R.C. Chu, in: Proceedings of the Rohsenow Symposium, National Electronics Manufacturing Initiative (NEMI), Cambridge, MA, 2003.
[195] S.S. Tonapi, R.A. Fillion, F.J. Schattenmann, H.S. Cole, J.D. Evans, B.G. Sammakia, in: Proceedings of the IEEE/SEMI Advanced Manufacturing Conference and Workshop, IEEE, Munich, Germany, (2003), pp. 250–254.
[196] C. LaBounty, A. Shakouri, P. Abraham, J.E. Bowers, Optical Engineering 39 (2000) 2847–2852.
[197] J.M. Houston, Journal of Applied Physics 30 (1959) 481–487.
[198] S.M. Sze, K.K. Ng, Physics of Semiconductor Devices, Wiley-Interscience, 2006.
[199] G. Mahan, Journal of Applied Physics 76 (1994) 4362–4366.
[200] G.S. Nolas, H.J. Goldsmid, Journal of Applied Physics 85 (1999) 4066–4068.
[201] C. Herring, M.H. Nichols, Review of Modern Physics 21 (1949) 185–270.
[202] B.F. Williams, Applied Physics Letters 14 (1969) 273–275.
[203] J.D. Jackson, Classical Electrodynamics, John Wiley, New York, 1999.
[204] Y. Hishinuma, T.H. Geballe, B.Y. Moyzhes, T.W. Kenny, Applied Physics Letters 78 (2001) 2572–2574.
[205] N.M. Miskovsky, P.H. Cutler, Applied Physics Letters 75 (1999) 2147–2149.
[206] E.T. Enikov, T. Makansi, Nanotechnology 19 (2008) 075703.
[207] G.M. Flening, J.E. Henderson, Physical Review 58 (1941) 907–908.
[208] W.B. Nottingham, Physical Review 59 (1941) 906.
[209] L.W. Swanson, L.C. Crouser, F.M. Charbonnier, Physical Review 151 (1966) 327–340.
[210] Y. Hishinuma, T.H. Geballe, B.Y. Moyzhes, T.W. Kenny, Journal of Applied Physics 94 (2003) 4690–4696.
[211] T.L. Westover, T.S. Fisher, Physical Review B 77 (2008) 115426.
[212] A.N. Korotkov, K.K. Likharev, Applied Physics Letters 75 (1999) 2491–2493.
[213] M.F. O'Dwyer, R.A. Lewis, C. Zhang, T.E. Humphrey, Physical Review B 72 (2005) 205330.
[214] M.W. Zemansky, R.H. Dittman, Heat and Thermodynamics, McGraw Hill, London, 1981.
[215] T.E. Humphrey, H. Linke, Physica E: Low-dimensional Systems and Nanostructures 29 (2005) 390–398.
[216] R. Tsu, R.F. Greene, Electrochemical and Solid State Letters 2 (1999) 645–647.
[217] J.I. Pankove, Optical Processes in Semiconductors, Dover Publications Inc., New York, 1971.
[218] Y. Hishinuma, B.Y. Moyzhes, T.H. Geballe, T.W. Kenny, Applied Physics Letters 81 (2002) 4242–4244.
[219] A. Shakouri, J.E. Bowers, Applied Physics Letters 71 (1997) 1239–1241.
[220] M.D. Ulrich, P.A. Barnes, C.B. Vining, Journal of Applied Physics 90 (2001) 1625–1631.
[221] Y. Zhang, J. Christofferson, A. Shakouri, G. Zeng, J.E. Bowers, E.T. Croke, IEEE Transactions on Components and Packaging Technologies 29 (2006) 395–401.
[222] M. Zebarjadi, A. Shakouri, K. Esfarjani, Physical Review B 74 (2006) 195331.
[223] G. Chen, A. Shakouri, Transactions of the ASME 124 (2002) 242–252.
[224] A. Shakouri, C. LaBounty, J. Piprek, P. Abraham, J.E. Bowers, Applied Physics Letters 74 (1999) 88–89.
[225] G.D. Mahan, L.M. Woods, Physical Review Letters 80 (1998) 4016–4019.
[226] G. Mahan, J.O. Sofo, M. Bartkowiak, Journal of Applied Physics 83 (1998) 4683–4689.
[227] T.E. Humphrey, M.F. O'Dwyer, C. Zhang, R.A. Lewis, Journal of Applied Physics 98 (2005) 026108.
[228] C.B. Vining, G. Mahan, Journal of Applied Physics 86 (1999) 6852–6853.
[229] M.F. O'Dwyer, R.A. Lewis, C. Zhang, Journal of Physics D: Applied Physics 40 (2007) 1167–1174.
[230] W.J. Kaiser, L.D. Bell, Physical Review Letters 60 (1988) 1406–1409.
[231] M. Kozhevnikov, V. Narayanamurti, C. Zheng, Y.-J. Chiu, Physical Review Letters 82 (1999) 3677–3680.
[232] L.D. Bell, W.J. Kaiser, Physical Review Letters 61 (1988) 2368–2371.
[233] D.L. Smith, E.Y. Lee, V. Narayanamurti, Physical Review Letters 80 (1998) 2433–2436.
[234] Z. Bian, A. Shakouri, Applied Physics Letters 88 (2006) 012102.
[235] D. Vashaee, A. Shakouri, Journal of Applied Physics 101 (2007) 053719.
[236] J.M.O. Zide, D. Vashaee, Z. Bian, G. Zeng, J.E. Bowers, A. Shakouri, A.C. Gossard, Physical Review B 74 (2006) 205335.
[237] G. Zeng, J.M.O. Zide, W. Kim, J.E. Bowers, A.C. Gossard, Z. Bian, Y. Zhang, A. Shakouri, S.L. Singer, A. Majumdar, Journal of Applied Physics 101 (2007).
[238] J.M.O. Zide, D.O. Klenov, S. Stemmer, A.C. Gossard, G. Zeng, J.E. Bowers, D. Vashaee, A. Shakouri, Applied Physics Letters 87 (2005) 112102.
[239] W. Kim, J.M.O. Zide, A.C. Gossard, D.O. Klenov, S. Stemmer, A. Shakouri, A. Majumdar, Physical Review Letters 96 (2006) 045901.
[240] G. Zeng, J.E. Bowers, J.M.O. Zide, A.C. Gossard, W. Kim, S. Singer, A. Majumdar, R. Singh, Z. Bian, Y. Zhang, A. Shakouri, Applied Physics Letters 88 (2006) 113502.
[241] S.V. Faleev, F. Leonard, Physical Review B 77 (2008) 214304.
[242] K. Kishimoto, K. Yamamoto, T. Koyanagi, Japanese Journal of Applied Physics 42 (2003) 501–508.
[243] L.I. Schiff, Quantum Mechanics, McGraw Hill, New York, 1968.
[244] J.P. Heremans, C.M. Thrush, D.T. Morelli, Journal of Applied Physics 98 (2005) 063703.
[245] M. Zebarjadi, K. Esfarjani, A. Shakouri, J.-H. Bahk, Z. Bian, C. Zeng, J.E. Bowers, H. Lu, J.M.O. Zide, A.C. Gossard, Applied Physics Letters 94 (2009) 202105.
[246] M.D. Ulrich, P.A. Barnes, C.B. Vining, Applied Physics Letters 92 (2002) 245–247.
[247] J. Tauc, Czechoslovak Journal of Physics 3 (1953) 282–302.
[248] G.D. Mahan, M. Bartkowiak, Applied Physics Letters 74 (1999) 953–954.
[249] Y.S. Ju, U. Ghoshal, Journal of Applied Physics 88 (2000) 4135–4139.
[250] J. Zhang, N.G. Anderson, K.M. Lau, Applied Physics Letters 83 (2003) 374–376.
[251] F. Ali, A. Gupta, HEMTs and HBTs: Devices, Fabrication and Circuits, Artech House, Boston, MA, 1991.
[252] J. Singh, Semiconductor Optoelectronics: Physics and Technology, McGraw Hill, New York, 1995.
[253] J. Zhang, N.G. Anderson, K.M. Lau, IEEE Electron Device Letters 25 (2004) 345–347.
[254] K.A. Chao, M. Larsson, A.G. Mal'shikov, Applied Physics Letters 87 (2005) 022103.
[255] Y. Zhang, G. Zeng, J. Piprek, A. Bar-Cohen, IEEE Transactions on Components and Packaging Technologies 28 (2005) 658–664.
[256] D. Vashaee, C. LaBounty, X. Fang, G. Zeng, P. Abraham, J.E. Bowers, A. Shakouri, Proceedings of the SPIE 4284 (2001) 139–144.
[257] G. Zeng, A. Shakouri, C. LaBounty, G. Robinson, E.T. Croke, P. Abraham, X. Fan, H. Reese, J.E. Bowers, Electronics Letters 35 (1999) 2146–2147.
[258] A. Shakouri, Y. Zhang, IEEE Transactions on Components and Packaging Technologies 28 (2005) 65–69.
[259] X. Fan, G. Zeng, C. LaBounty, J.E. Bowers, E.T. Croke, C.C. Ahn, S.T. Huxtable, A. Majumdar, A. Shakouri, Applied Physics Letters 78 (2001) 1580–1582.
[260] Y. Ezzahri, G. Zeng, K. Fukutani, Z. Bian, A. Shakouri, Microelectronics Journal 39 (2008) 981–991.
[261] C. LaBounty, A. Shakouri, J.E. Bowers, Journal of Applied Physics 89 (2001) 4059–4064.
[262] D. Vashaee, J. Christofferson, Y. Zhang, A. Shakouri, G. Zeng, C. LaBounty, X. Fan, J. Piprek, J.E. Bowers, E.T. Croke, Microscale Thermophysical Engineering 9 (2005) 99–118.
[263] C.B. Vining, Nanotechnology and New Materials, Al Gore's Solutions Summit, New York, NY, 2008.
[264] Market Potentialities for Thermoelectrics, http://www.tetech.com/.
[265] T.M. Tritt, H. Bottner, L. Chen, MRS Bulletin 33 (2008) 366–368.
[266] K. Seeger, Semiconductor Physics: An Introduction, Springer, 2004.
[267] S.Y. Ren, J.D. Dow, Physical Review B: Condensed Matter 25 (1982) 3750–3755.
[268] D. Vashaee, A. Shakouri, Journal of Applied Physics 95 (2004) 1233–1245.
[269] E. Kapon, E. Pelucchi, S. Watanabe, A. Malko, M.H. Baier, K. Leifer, B. Dwir, F. Michelini, M.A. Dupertuis, Physica E: Low-dimensional Systems and Nanostructures 25 (2004) 288–297.