IOP Publishing Semiconductor Science and Technology
Semicond. Sci. Technol. 31 (2016) 043001 (24pp) doi:10.1088/0268-1242/31/4/043001

Topical Review

First-principles calculations of thermal, electrical, and thermoelectric transport properties of semiconductors

Jiawei Zhou, Bolin Liao and Gang Chen

Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, 02139, USA

E-mail: gchen2@mit.edu

Received 7 September 2015, revised 28 December 2015
Accepted for publication 19 January 2016
Published 7 March 2016

Abstract

The transport properties of semiconductors are key to the performance of many solid-state devices (transistors, data storage, thermoelectric cooling and power generation devices, etc). An understanding of the transport details can lead to material designs with better performances. In recent years simulation tools based on first-principles calculations have been greatly improved, being able to obtain the fundamental ground-state properties of materials (such as band structure and phonon dispersion) accurately. Accordingly, methods have been developed to calculate the transport properties based on an *ab initio* approach. In this review we focus on the thermal, electrical, and thermoelectric transport properties of semiconductors, which represent the basic transport characteristics of the two degrees of freedom in solids—electronic and lattice degrees of freedom. Starting from the coupled electron-phonon Boltzmann transport equations, we illustrate different scattering mechanisms that change the transport features and review the first-principles approaches that solve the transport equations. We then present the first-principles results on the thermal and electrical transport properties of semiconductors. The discussions are grouped based on different scattering mechanisms including phonon-phonon scattering, phonon scattering by equilibrium electrons, carrier scattering by equilibrium phonons, carrier scattering by polar optical phonons, scatterings due to impurities, alloying and doping, and the phonon drag effect. We show how the first-principles methods allow one to investigate transport properties with unprecedented detail and also offer new insights into the electron and phonon transport. The current status of the simulation is mentioned when appropriate and some of the future directions are also discussed.

Keywords: mobility, Seebeck coefficient, thermal conductivity, semiconductor, scattering mechanism, mean free path, accumulated contribution

(Some figures may appear in colour only in the online journal)

1. Introduction

Investigating the transport property of solids is an old but enlightening problem. Many insightful ideas have emerged from such studies, involving for instance, superconductivity, localization, etc, which in some sense reflect the modern development of condensed matter physics. It is intriguing also because most of our daily 'machines' rely on such transport, either of charge, mass, or of energy. Examples include transistors (mobility), phase change medium for data storage (thermal conductivity), thermoelectric cooling and power generation devices (electrons and phonons), and so on. Better materials are constantly sought after to improve the overall performance of the working device. The improvement of

existing materials and the discovery of new ones thus require a clear understanding of the underlying physics of transport properties. Theoretical models combined with experiments have offered great insights into the different mechanisms that affect transport properties. However some details have been ignored in order for the theories to be tractable. For example, the simple kinetic formula, $\sigma = \frac{1}{3}c\mathrm{v}\Lambda$, describes the electrical conductivity σ in terms of three averaged parameters of the energy carriers: carrier energy (c), velocity (v) and mean free path (MFP) (Λ). In reality, however, these parameters can vary significantly among different carriers and even the same type of carriers. Such information has been difficult to obtain within traditional methods because experiments normally probe transport properties that are averaged over all the carriers, while the theoretical models are intractable without certain approximations and thus lack accuracy and in many cases, predictive power. The development of first-principles computational techniques in recent decades enables one to study materials' physical properties accurately without adjustable parameters [1]. The merit of first-principles methods is that they are easily justifiable and useful for a wide range of problems. Since their invention [2, 3], they have been widely used for studying and also predicting the physical properties of many materials, such as the equilibrium crystal structure, the electronic band structure, or the phonon dispersion, etc [4]. We will briefly describe how first-principles calculations can be utilized for studying transport properties. We do not discuss how first-principles methods work but refer readers to some of the books that already exist and the references therein [4, 5].

We want to be cautious: 'first principles' is sometimes a misleading name due to the fact that the underlying theory (which should be referred to as the density functional theory (DFT)) also involves approximations to the quantum mechanical governing equation—the Schrödinger equation (including relativistic corrections derived from the Dirac equation). For example, only the ground-state properties are calculated and the strongly-correlated systems cannot be described well without the introduction of some tuning parameters [4]. In this light, there are also other quantum mechanics-based algorithms which are more accurate for such problems but are also much more time-consuming [6]. In the following the main results are obtained using the DFT, which describes many properties of solids very well and currently is sufficient for the transport phenomenon study in commonly-used semiconductors. However, we will also point out that some other formalisms (to more accurately describe the many-body effect of electrons) will be desired along the path of developing first-principles tools for more complex materials.

In this review we focus on the studies of semiconductors, and in particular, their thermal, electrical and thermoelectric properties. Semiconductors, though only a portion of the family of solids, play important roles in various applications such as microelectronics, photovoltaics, optoelectronics and thermoelectrics. Although extensive studies into semiconductors from theoretical models to experiments have revealed many physical mechanisms for describing their transport behaviors, so far it is hard to predict a material's transport property (for example, mobility) without performing experiments. A computational technique that has a predictive power will thus greatly facilitate the optimization of existing materials and the search for better ones [7]. The first-principles methods have been traditionally applied to study electronic structure and phonon dispersion, both of which are eigenstates of the system. In a particle picture, each eigenstate corresponds to one or several particles (electrons or phonons) that carry certain energy with some velocity. To describe a transport property such as mobility and thermal conductivity, one must know how these particles evolve under the influence of other states. This process is formalized by the famous Boltzmann transport equation (BTE) and will be the cornerstone of our following discussions. One of the major challenges of solving the BTE is to know how the particles are scattered by each other (or by the defect and boundary). The scattering process is determined by the coupling between different eigenstates. The calculation of such coupling strengths has been computationally prohibitive and has only become accessible recently thanks to a series of developments for an efficient extraction of the interaction strengths between different eigenstates (electrons and phonons). In the next section, different types of interaction along with their calculation will be discussed under the BTE formalism.

Among all the transport properties, thermoelectric transport presents a particular challenge; firstly this is because it is affected by three transport phenomena as shown by the figure of merit $zT = \sigma S^2 T/\kappa$ which characterizes a material's thermoelectric efficiency, where σ, S, κ, T are the electrical conductivity, Seebeck coefficient, thermal conductivity and absolute temperature, respectively. For semiconductors, the electrical conduction and Seebeck effect are mostly related to the electronic properties while the thermal conduction must consider the phonons. Therefore for the thermoelectric transport as a whole, one must consider both electrons and phonons, as well as the coupling between them. Secondly, thermoelectric materials are often complex due to the alloying and doping procedure involved, which always poses a challenge for first-principles simulations. In this review, we briefly summarize the recent efforts made towards a first-principles calculation of the thermal, electrical and thermoelectric properties of semiconductors, particularly the group IV and III–V semiconductors. In section 2 we introduce the formalism of the coupled electron-phonon BTEs, which serve as the starting point of our discussions. Under the BTE formalism, basic concepts are explained and the corresponding computational techniques are discussed. The following two sections are devoted to the first-principles results of the transport property: section 3 deals with thermal transport and section 4 examines electrical transport. In the final section we briefly discuss the possible future directions that might help to improve first-principles simulations. We anticipate that the development of computational tools will further extend the calculations to more complex materials.

2. Coupled electron-phonon Boltzmann equations

The electron and phonon properties of a solid are normally described by the band structure and phonon dispersion, respectively. For convenience we would like to think of these eigenstates as particles without considering their phase. These particles carry a certain charge or energy dictated by quantum mechanics. This semiclassical picture is not always valid but can be justified if the variation of the external field occurs in a larger scale than the wave packet associated with this particle and the field is relatively weak compared to the periodically varying atomic potential [8]. A particle in an eigenstate with finite group velocity implies that it keeps moving without being deflected under the dynamics of the system, since the Hamiltonian acting on the eigenstate gives itself. In other words, the eigenstate has an infinite lifetime. Practically, this cannot be achieved because there are always perturbations to the system, which causes finite transition probabilities between different states. In a particle picture, the perturbations lead to the scattering of the particle by other particles or impurities, which give rise to the 'resistance' for the charge or energy flow. To describe such a process, one must have a way to treat the great amount of particles that are moving inside the material with the interactions that lead to their scatterings.

Rigorously, a (6N + 1)-dimensional phase space (3N for positions, 3N for momenta and one for time) is used to describe the dynamics of N particles. Such a description is adopted by the molecular dynamics technique, and has been used to calculate phonon properties as well as lattice heat conduction [9, 10]. However, the molecular dynamics method lacks predictive power due to the empirical interatomic potentials involved, while a first-principles calculation with a large system of N particles is formidable. Alternatively, one could use a single-particle distribution function to represent all the particles in the system. This function has only seven independent variables (three for positions, three for momenta and one for time) and describes how many particles stay in the state characterized by a given position and momentum at each time. The crucial step towards describing the transport of the system is to rewrite the dynamics of N particles into a governing equation for the distribution function. This is achieved by the introduction of the BTEs. For the electron and phonon systems in the steady state, the BTEs read as [8, 11–13]

$$\begin{cases} \mathbf{v}_\alpha(\mathbf{k}) \cdot \nabla_{\mathbf{r}} f_\alpha(\mathbf{k}) + \dfrac{\mathbf{F}}{\hbar} \cdot \nabla_{\mathbf{k}} f_\alpha(\mathbf{k}) = \left(\dfrac{\partial f_\alpha(\mathbf{k})}{\partial t}\right)_{\text{coll}} \\ \mathbf{v}_\lambda(\mathbf{q}) \cdot \nabla_{\mathbf{r}} n_\lambda(\mathbf{q}) = \left(\dfrac{\partial n_\lambda(\mathbf{q})}{\partial t}\right)_{\text{coll}} \end{cases} \tag{1}$$

where $\hbar$ is the reduced Planck constant, $\mathbf{F}$ is the external force exerting on this particle (electrostatic field in our case, which affects electrons but not phonons) and the velocity vectors $\mathbf{v}$ for electrons ($\mathbf{v}_\alpha(\mathbf{k})$) and for phonons ($\mathbf{v}_\lambda(\mathbf{q})$) are specified with wave vectors $\mathbf{k}$ (for electrons) and $\mathbf{q}$ (for phonons), as well as band number α and branch number λ. f and n represent the distribution functions for electrons and phonons, respectively, with the equilibrium state described by Fermi–Dirac and Bose–Einstein statistics:

$$\begin{cases} f_\alpha^0(\mathbf{k}) = \dfrac{1}{e^{(E_{\mathbf{k}}^\alpha - \mu)/k_B T} + 1} \\ n_\lambda^0(\mathbf{q}) = \dfrac{1}{e^{\hbar\omega_{\mathbf{q}}^\lambda / k_B T} - 1} \end{cases}. \tag{2}$$

The first equation of equation (1) is for electrons and will be denoted as the electron-BTE while the second equation is for phonons and will be called the phonon-BTE. Here we skip the derivation of equation (1), which can be found in many textbooks [8, 11–13]. We instead want to pay more attention to the meanings of each term and their explicit forms. Briefly speaking, the BTEs describe the balance of the change of the particle number in various available states. There are two major causes for these changes. One comes from the external field exerted on the particles, with examples including the electric field for electrons (second term of electron-BTE) and temperature gradients for both electrons (included in first term of electron-BTE) and phonons (first term of phonon-BTE). The other originates from the scatterings between the particles. For convenience, the terms of the former type in the Boltzmann equation are often called drift terms while those of the latter type are referred to as collision terms [8].

In equilibrium, the distribution functions for electrons and phonons are described by Fermi–Dirac and Bose–Einstein statistics, respectively. As the macroscopic fields start to drive the system, electrons and phonons will move away from the equilibrium, until they are balanced by the collision terms, which rebuild their distribution functions. The degree of the non-equilibrium is thus determined by the strength of the scatterings. For both electrons and phonons, there are various mechanisms that can make transitions from one state to another, such as phonon-phonon interaction, electron-phonon interaction and electron-electron interaction. Besides, due to the impurities involved (for example, by alloying or doping procedure), we can also have impurity scatterings for both electrons and phonons. In general, the collision terms can be described as

$$\begin{cases} \left(\dfrac{\partial f_\alpha(\mathbf{k})}{\partial t}\right)_{\text{coll}} = \left(\dfrac{\partial f_\alpha(\mathbf{k})}{\partial t}\right)_{e-ph} + \left(\dfrac{\partial f_\alpha(\mathbf{k})}{\partial t}\right)_{e-imp} + \ldots \\ \left(\dfrac{\partial n_\lambda(\mathbf{q})}{\partial t}\right)_{\text{coll}} = \left(\dfrac{\partial n_\lambda(\mathbf{q})}{\partial t}\right)_{ph-ph} \\ \quad + \left(\dfrac{\partial n_\lambda(\mathbf{q})}{\partial t}\right)_{ph-e} + \left(\dfrac{\partial n_\lambda(\mathbf{q})}{\partial t}\right)_{ph-imp} + \ldots \end{cases} \tag{3}$$

where each term represents the scattering probabilities into and out of a certain state, added up together according to Matthiessen's rule [8]. We neglect the electron-electron scattering because within most temperature ranges it is hardly as effective as other mechanisms in scattering the electrons [11, 14]. However we will mention the necessity of considering the electron-electron interaction in the transport calculations as a step beyond the current status.

The central problem to the BTEs is to solve the distribution functions. If we know the distribution function, the transport properties can be easily calculated by summing up all the states with corresponding energies and velocities. For this purpose, we need to explicitly describe the collision terms. These scatterings depend on the available states of electrons and phonons, and typically involve the distribution functions in a non-linear way. At moderate external fields, however, the distribution functions only have small deviations from the equilibrium distribution. Consequently, one can expand the distribution function near the equilibrium condition in the BTEs and take the lowest-order terms, which results in the linearized BTEs [11]:

$$\begin{cases} \mathbf{v}_\alpha(\mathbf{k}) \cdot \dfrac{\partial f_\alpha^0(\mathbf{k})}{\partial T}\nabla T - e\mathbf{v}_\alpha(\mathbf{k}) \cdot \dfrac{\partial f_\alpha^0(\mathbf{k})}{\partial E}\nabla\varphi = \left(\dfrac{\partial f_\alpha(\mathbf{k})}{\partial t}\right)_{\text{coll}} \\ \mathbf{v}_\lambda(\mathbf{q}) \cdot \dfrac{\partial n_\lambda^0(\mathbf{q})}{\partial T}\nabla T = \left(\dfrac{\partial n_\lambda(\mathbf{q})}{\partial t}\right)_{\text{coll}} \end{cases} \tag{4}$$

where T is the temperature, e is the electron charge (negative), E is the electronic energy level and φ is the electrochemical potential (incorporating the electrostatic potential ϕ and the chemical potential μ). Note that we have replaced the distribution functions in the left-hand side of equation (1) with the equilibrium values. By doing so, the first term of the electron-BTE is split into two terms including the temperature gradient and the chemical potential gradient, respectively; the latter of which, when combined with the electrostatic force term in the electron-BTE, leads to the second term in the electron-BTE of equation (4), by defining $\varphi = \phi + \mu/e$ [15].

The collision terms, under the linearized BTE formalism, can be generally expressed as

$$\begin{cases} \left(\dfrac{\partial f_\alpha(\mathbf{k})}{\partial t}\right)_{\text{coll}} = A_{\mathbf{k}\alpha} \cdot \Delta f_\alpha(\mathbf{k}) \\ \quad + \left[\displaystyle\sum_{\mathbf{k}'\neq\mathbf{k},\alpha} A_{\mathbf{k}'\alpha}\Delta f_\alpha(\mathbf{k}') + \sum_{\mathbf{q}\lambda} B_{\mathbf{q}\lambda}\Delta n_\lambda(\mathbf{q})\right] \\ \left(\dfrac{\partial n_\lambda(\mathbf{q})}{\partial t}\right)_{\text{coll}} = C_{\mathbf{q}\lambda} \cdot \Delta n_\lambda(\mathbf{q}) \\ \quad + \left[\displaystyle\sum_{\mathbf{q}'\neq\mathbf{q},\lambda} C_{\mathbf{q}'\lambda}\Delta n_\lambda(\mathbf{q}') + \sum_{\mathbf{k}\alpha} D_{\mathbf{k}\alpha}\Delta f_\alpha(\mathbf{k})\right] \end{cases} \tag{5}$$

where the collision terms linearly depend on (with the proportionality factors denoted by A, B, C, D) the deviations of the distribution functions from the equilibrium:

$$\begin{cases} \Delta f_\alpha(\mathbf{k}) = f_\alpha(\mathbf{k}) - f_\alpha^0(\mathbf{k}) \\ \Delta n_\lambda(\mathbf{q}) = n_\lambda(\mathbf{q}) - n_\lambda^0(\mathbf{q}) \end{cases}. \tag{6}$$

Note that the first-order deviation in the drift terms and the second-order deviation in the collision terms have been ignored because they are higher-order contributions to the BTEs. For quasi-ballistic transport, however, these higher-order contributions cannot be neglected because the deviation of the distribution function from the equilibrium is relatively large [15]. In equation (5) we have also separated the collision terms into two parts. One is directly proportional to the distribution deviation of the state of interest (**k** or **q**), and the remaining parts rely on the distribution of other states. The linearized BTE has been the starting point for most of the theoretical work on transport properties. There could be situations where the small deviation of the distribution function is not well justified, which however also poses questions concerning the use of the BTE formalism [8]. Overall the linearized BTE has been shown to be applicable to a wide range of conditions and will be the basis of our following discussion.

A linear system as determined by equation (4) generally implies that the distribution function for the given state depends on those of other states. To obtain exact results, numerical techniques that solve large linear systems are usually applied, such as the direct or iterative methods. One important but not necessary step here is to neglect the dependence of the collision terms for the given state on any other state, which means that the terms in the bracket are ignored in equation (5). In other words, this assumes that when considering the scatterings of a given state, all other states appear to be at equilibrium. By doing so, the only term left in the collision terms is (take electrons as an example)

$$\left(\frac{\partial f_\alpha(\mathbf{k})}{\partial t}\right)_{\text{coll}} = A_{\mathbf{k}\alpha} \cdot \Delta f_\alpha(\mathbf{k}) = -\frac{\Delta f_\alpha(\mathbf{k})}{\tau_{\mathbf{k}\alpha}} \tag{7}$$

where we replace the proportionality factor with the inverse of a characteristic time (the minus sign is for introducing a properly defined positive time constant as will be clear in the explicit forms). Considering the meaning of the collision, this characteristic time represents how fast the non-equilibrium distribution functions return to the equilibrium due to scatterings. For this reason, they have been called the relaxation times [8]. Using equation (7), equation (4) can then be readily solved yielding

$$\begin{aligned} f_\alpha(\mathbf{k}) = f_\alpha^0(\mathbf{k}) - \tau_{\mathbf{k}\alpha}\Bigg[\mathbf{v}_\alpha(\mathbf{k}) \cdot \frac{\partial f_\alpha^0(\mathbf{k})}{\partial T}\nabla T \\ - e\mathbf{v}_\alpha(\mathbf{k}) \cdot \frac{\partial f_\alpha^0(\mathbf{k})}{\partial E}\nabla\varphi\Bigg]. \end{aligned} \tag{8}$$

Such approximation largely simplifies the procedure of solving the BTE and has been called the relaxation time approximation (RTA). However, the crude analysis above has totally removed the coupling between the electron and phonon system, which will be incorrect for certain transport phenomena [11]. We will see that some of the terms inside the bracket in equation (5), depending on different situations, must be taken into account to capture the necessary physics with as little computational difficulty as possible. In those cases, the collision term proportional to the given state can still be described using the relaxation time and we regard this also as an example of using the RTA. In essence, the RTA specifically considers the relaxation of the non-equilibrium

distribution of a given state, or its decay into other available states. In the particle picture, it can be viewed as the averaged time between the two scattering events of a particle at a given state. Its product with the particle's velocity v gives rise to the concept of a MFP $\Lambda = v\tau$, the averaged distance the particle travels before being scattered. In the following we discuss three major scattering mechanisms that contribute to the collision terms and the corresponding numerical framework that solves them.

2.1. Phonon-phonon interaction

A phonon mode describes a collective atomic movement where each atom vibrates according to a certain displacement pattern. Each phonon mode is associated with a wave vector and also a frequency, describing the vibration as a wave propagating in the medium. This classification assumes that the change of the total potential energy with respect to the displacements of the atoms can be written in a homogeneous quadratic form [8]

$$H_{\text{pot}} = \frac{1}{2}\sum_{\mathbf{R}\mathbf{R}'} D_{\alpha\beta}(\mathbf{R}, \mathbf{R}')u_{\alpha}(\mathbf{R})u_{\beta}(\mathbf{R}') \tag{9}$$

where $\mathbf{R}(\mathbf{R}')$ is the equilibrium position of each atom in the crystal, u is the atomic displacement, $\alpha(\beta)$ represents the three directions in the Cartesian coordinate, and D is the dynamical matrix. We note that the first-order derivative of the potential energy with respect to the atomic displacement is the force. The derivatives of the force with respect to the displacement are usually called force constants, which are natural generalizations of spring constants from mechanics. The nth-order derivatives of the energy (n-1th order derivative of force) are denoted as nth-order force constants. The form assumed by equation (9) therefore implies that all the force constants vanish except for the 2nd-order force constants (essentially dynamical matrix). This is the harmonic crystal approximation to real solids [8], which justifies the concepts of phonons. It is valid for studying heat capacity since higher-order effects are often negligible when the atomic displacements are small [8]. However, as we have mentioned, if the eigenstates are the true states of the system, a particle in a state with finite group velocity will keep moving without being deflected, implying an infinite phonon thermal conductivity. It is the higher-order effect that limits the intrinsic thermal conduction [8, 11]. For this reason, we usually call the 2nd-order force constants the harmonic force constants, while referring to the 3rd-order force constants and above as anharmonic force constants.

We can understand the anharmonic force constants that lead to thermal resistance as follows. Since phonon modes are only derived from the harmonic force constants, they are not the exact eigenstates of the system when anharmonic force constants exist. However, one can still use these phonons to describe the dynamics of the system by introducing the transition probabilities between different modes. Classically, it can be understood by a model where spring constants can vary with the displacements, and as a result the original eigenmodes will convert to other modes due to the non-linear vibration dynamics. Quantum mechanically, this is because the higher-order effects introduce the perturbations to the original system, which manifest as the coupling between the eigenmodes that lead to the transition processes (time-dependent perturbation theory) [11].

Among the interactions between phonons, the dominant scattering comes from the three-phonon process [11]. There are mainly two different types of interactions. In one case two phonons merge into one phonon, creating a new particle with higher energy. In another case one phonon transfers its energy into two phonons. Note that in the phonon-phonon scattering process the phonon number is not conserved. There are also higher-order processes such as four-phonon scattering. These typically rarely happen and in most situations can be safely ignored.

For a given phonon state, there are several ways to change its distribution function. This state can decay into two phonons, or can be combined with another phonon and merge into a new phonon. These describe the out-scattering terms. Besides, two other phonons can also scatter and merge into this phonon, resulting in an increase of its distribution function, which are called in-scattering terms. Considering all these processes, one can use the Golden Rule to calculate the transition rates of a given state (corresponding to the proportionality factor $C_{\mathbf{q}\lambda}$ in equation (5)) due to three-phonon scatterings based on the three-phonon interaction matrix elements under the linearized BTE formalism [11, 16, 17]

$$\begin{aligned}\frac{1}{\tau^{ph-ph}_{\mathbf{q}\lambda}} = \frac{\pi}{\hbar^2 N_{\mathbf{q}}}\sum_{\mathbf{q}_1\lambda_1,\mathbf{q}_2\lambda_2} |V_{\mathbf{q}\lambda,\mathbf{q}_1\lambda_1,\mathbf{q}_2\lambda_2}|^2 \\ \times\left\{\begin{array}{l}(1+n^0_{\mathbf{q}_1\lambda_1}+n^0_{\mathbf{q}_2\lambda_2}) \\ \quad\cdot[\delta(\omega_{\mathbf{q}\lambda}-\omega_{\mathbf{q}_1\lambda_1}-\omega_{\mathbf{q}_2\lambda_2})\delta_{\mathbf{q}-\mathbf{q}_1-\mathbf{q}_2,\mathbf{G}}] \\ +(n^0_{\mathbf{q}_2\lambda_2}-n^0_{\mathbf{q}_1\lambda_1}) \\ \quad\cdot\left[\begin{array}{l}\delta(\omega_{\mathbf{q}\lambda}+\omega_{\mathbf{q}_2\lambda_2}-\omega_{\mathbf{q}_1\lambda_1})\delta_{\mathbf{q}-\mathbf{q}_1+\mathbf{q}_2,\mathbf{G}} \\ -\delta(\omega_{\mathbf{q}\lambda}+\omega_{\mathbf{q}_1\lambda_1}-\omega_{\mathbf{q}_2\lambda_2})\delta_{\mathbf{q}+\mathbf{q}_1-\mathbf{q}_2,\mathbf{G}}\end{array}\right]\end{array}\right\}\end{aligned} \tag{10}$$

where the three-phonon interaction matrix is explicitly related to the anharmonic force constants via (in the following we only consider up to the 3rd-order, or cubic, force constant):

$$\begin{aligned}V_{\mathbf{q}\lambda,\mathbf{q}_1\lambda_1,\mathbf{q}_2\lambda_2} = \left(\frac{\hbar}{2}\right)^{3/2}\sum_{\mathbf{R}_i\tau_i\alpha_i}\Psi^{\alpha,\beta,\gamma}_{0\tau,\mathbf{R}_1\tau_1,\mathbf{R}_2\tau_2} \\ \times\frac{e^{i(\mathbf{q}_1\cdot\mathbf{R}_1+\mathbf{q}_2\cdot\mathbf{R}_2)}e^{\tau\alpha}_{\mathbf{q}\lambda}e^{\tau_1\beta}_{\mathbf{q}_1\lambda_1}e^{\tau_2\gamma}_{\mathbf{q}_2\lambda_2}}{[M_\tau M_{\tau_1}M_{\tau_2}\omega_{\mathbf{q}\lambda}\omega_{\mathbf{q}_1\lambda_1}\omega_{\mathbf{q}_2\lambda_2}]^{1/2}}.\end{aligned} \tag{11}$$

In the above expressions, $N_{\mathbf{q}}$ is the number of the discrete points in the reciprocal space mesh ($\mathbf{q}_1$ and $\mathbf{q}_2$ go over the same mesh), $n^0_{\mathbf{q}\lambda}$ is the Bose–Einstein distribution of phonons (equation (2)), $\omega_{\mathbf{q}\lambda}$ is the phonon frequency, $e_{\mathbf{q}\lambda}$ describes the displacement pattern for the phonon eigenmodes and M_τ is the atomic mass on the sublattice site τ. The cubic force constant $\Psi^{\alpha\beta\gamma}_{0\tau,\mathbf{R}_1\tau_1,\mathbf{R}_2\tau_2}$ is defined as the 3rd-order derivative of the total energy with respect to the displacements of three atomic sites (0τ, $\mathbf{R}_1\tau_1$, $\mathbf{R}_2\tau_2$) along the corresponding directions (α, β, γ), for which ($\mathbf{R}\tau$) represents the atom at the

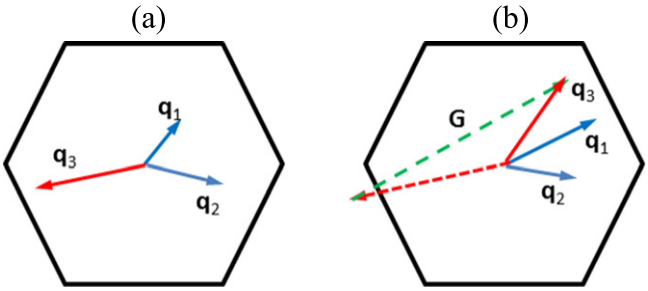

Figure 1. (a) Normal phonon-phonon scattering process where $\mathbf{q}_1 + \mathbf{q}_2 + \mathbf{q}_3 = 0$ and (b) Umklapp process where $\mathbf{q}_1 + \mathbf{q}_2 + \mathbf{q}_3 = \mathbf{G}$. A hexagonal cell is used here as an example for the FBZ.

sublattice site τ in the unit cell $\mathbf{R}$ away from the origin and α_i are in Cartesian coordinates. For phonons, we have intentionally ignored the non-equilibrium distribution of other states when evaluating the scattering of a given state (see equation (5)). As a result, the collision term due to the phonon-phonon interaction in equation (3) can be written as

$$\left(\frac{\partial f_\alpha(\mathbf{k})}{\partial t}\right)_{ph-ph} = -\frac{\Delta n_\lambda(\mathbf{q})}{\tau^{ph-ph}_{\mathbf{q}\lambda}} \tag{12}$$

with the relaxation times given by equation (10). During these scatterings, the energy and crystal momentum conservation need to be satisfied, which impose the conditions on which processes are allowed, as shown by the delta functions in equation (10). The crystal momentum conservation requires that the three-phonon wave vectors involved in the scattering add up to zero or up to an integer number of reciprocal lattice vectors. Phonon wave vectors are usually defined in the first Brillouin zone (FBZ). As shown by figure 1, when two of the phonon wave vectors are small the third wave vector is just their summation and the total crystal momentum is not changed. This is called the normal process. On the other hand, when two of the phonon wave vectors are large, their summation will fall outside the FBZ and reciprocal lattice vectors have to be involved to bring it back. Such case is called the Umklapp process and causes the change of total crystal momentum. It is these Umklapp processes that scatter phonons back to the equilibrium (reduce the net crystal momentum when they are set up by the driving force) and essentially create the thermal resistance [18].

In fact the phonon-phonon interaction is the main contribution to the thermal resistance for the semiconductors in most temperature ranges. Using RTA, we can write the phonon collision term as

$$\left(\frac{\partial n_\lambda(\mathbf{q})}{\partial t}\right)_{\text{coll}} = -\frac{\Delta n_\lambda(\mathbf{q})}{\tau_{\mathbf{q}\lambda}} \tag{13}$$

where the relaxation time includes phonon-phonon scatterings and also other terms due to impurities or electrons when necessary. Combined with equation (1), the phonon distribution function can be readily solved:

$$n_\lambda(\mathbf{q}) = n^0_\lambda(\mathbf{q}) - \tau_{\mathbf{q}\lambda}\mathbf{v}_\lambda(\mathbf{q}) \cdot \frac{\partial n^0_\lambda(\mathbf{q})}{\partial T}\nabla T. \tag{14}$$

The thermal conductivity as the ratio of the heat current and the temperature gradient assuming isotropic property is then obtained

$$\boldsymbol{\kappa}_{ph} = \frac{1}{3\Omega N_{\mathbf{q}}}\sum_{\mathbf{q}\lambda}\mathbf{v}^2_{\mathbf{q}\lambda}\tau_{\mathbf{q}\lambda}\hbar\omega_{\mathbf{q}\lambda}\frac{\partial n^0_{\mathbf{q}\lambda}}{\partial T} \tag{15}$$

where Ω is the unit cell volume and $N_{\mathbf{q}}$ is the number of the discrete points in the reciprocal space. It is clear from the discussion above that the key ingredients for describing the phonon-phonon interaction are the cubic force constants in equation (11). Harmonic force constants involve two independent atomic variables and are routinely calculated to obtain the phonon dispersion by the use of the density functional perturbation theory (DFPT) [19]. Anharmonic force constants, however, involve three independent atomic variables and are much harder to obtain due to their larger number involved. There are mainly two ways to obtain the anharmonic force constants. One is based on a real space approach and the other starts from the reciprocal space. The real space approach starts by creating many different atomic configurations in a supercell (one that includes many primitive cells) where one or more atoms are displaced from the equilibrium positions along certain directions [20]. The first-principles results for the forces on all the atoms are recorded. These forces can be written in an expansion expression in terms of the force constants and the displacements (here only up to 3rd-order contributions are explicitly shown):

$$\begin{aligned} F^\alpha_{\mathbf{R}\tau} = &-\sum_{\mathbf{R}_1\tau_1}\Phi^{\alpha,\beta}_{\mathbf{R}\tau,\mathbf{R}_1\tau_1}u^\beta_{\mathbf{R}_1\tau_1} \\ &-\frac{1}{2}\sum_{\mathbf{R}_1\tau_1,\mathbf{R}_2\tau_2}\Psi^{\alpha,\beta,\gamma}_{\mathbf{R}\tau,\mathbf{R}_1\tau_1,\mathbf{R}_2\tau_2}u^\beta_{\mathbf{R}_1\tau_1}u^\gamma_{\mathbf{R}_2\tau_2} \end{aligned} \tag{16}$$

where $\Phi^{\alpha,\beta}_{\mathbf{R}\tau,\mathbf{R}_1\tau_1}$ is the harmonic force constant and $\Psi^{\alpha,\beta,\gamma}_{\mathbf{R}\tau,\mathbf{R}_1\tau_1,\mathbf{R}_2\tau_2}$ is the cubic force constant; $F^\alpha_{\mathbf{R}\tau}$ describes the force along the α direction exerted on the atom at $(\mathbf{R}\tau)$ and $u^\alpha_{\mathbf{R}\tau}$ is its corresponding displacement. The different real space approaches differ in how the force constants are fitted from the above equation. It has been shown by Esfarjani *et al* [20] that using the symmetry (permutation, translation, rotation and Huang symmetry) the number of force constants that are required to fit can be largely reduced. They have shown the extraction of force constants up to 4th-order that compare reasonably well with the exact results when an artificial interaction potential is used. Other methods such as those based on compressive sensing or slave mode expansion are also employed to recognize the most important anharmonic terms so as to further reduce the number of anharmonic force constants that one has to deal with and simplify the analysis [21–23]. The advantage of the real space approach is that one can treat the DFT package as a black box and only produce his/her own post-processing code, thus reducing the

programming complexity. However, it has been known from the phonon dispersion calculation that the real space approach is not as accurate as the reciprocal space method (DFPT) at a given wave vector **q** because the former one introduces errors when truncating the force constants beyond the size of the supercell while the latter rigorously treats the infinite size of the crystal [19]. The power of DFPT relies on the '2n + 1' theorem [19, 24], which states that in order to know the 2n + 1th-order derivatives of the total energy only the information of the nth-order derivatives of the wavefunctions is required. For phonon dispersion calculation, the first-order derivatives of the wavefunctions are already obtained, which can be used to calculate up to 3rd-order force constants. This formalism directly calculates the phonon modes at a given **q** and does not rely on supercells. However, its programming complexity is greater than the real space method, and up to now the calculations that used this approach are only restricted to single-element materials such as Si, diamond and graphene [17, 25–30]. We want to note that further development on the reciprocal space method towards more complicated materials can provide an evaluation of the accuracy of the first-principles method as we move from well-known materials to unexplored ones.

2.2. Electron-phonon interaction

The BTEs as shown by equation (1) wi + ion between the electron and phonon systems [11]. This interaction behaves as a perturbation to each of them from the other system, and causes scatterings for both electrons and phonons. Electronic band structures are derived assuming atoms are at their equilibrium positions. In practice, the atoms vibrate according to the phonon modes and are seldom at equilibrium positions. The phonon modes change the environment (potential energy) the electrons experience and therefore perturb the electrons' movements. In a particle picture, such coupling means that electrons can collide with phonons, making a transition from one state to another. These processes need to satisfy the energy and crystal momentum conservation. For the electron-phonon interaction, Born–Oppenheimer approximation is often assumed, which states that as the atoms vibrate, electrons respond so fast that they almost see a static atomic configuration and therefore finds the lowest energy corresponding to that configuration. This simplifies the problem because electrons and phonons still have their own eigenstates, and are only coupled to each other via the coupling matrix element. In an intuitive way, Bloch proposes an analytic form for the electron-phonon coupling as follows [31]

$$\delta V = \sum_{\tau,\mathbf{R}} \mathbf{u}_{\tau,\mathbf{R}} \cdot \frac{\partial V}{\partial \mathbf{u}_{\tau,\mathbf{R}}} \tag{17}$$

where **u** is the atom displacement and $\frac{\partial V}{\partial \mathbf{u}_{\tau,\mathbf{R}}}$ is the perturbed potential of electrons due to the atomic movements. In this expression we regard $\mathbf{u}_{\tau,\mathbf{R}}$ as an operator acting on the phonon eigenstates while $\frac{\partial V}{\partial \mathbf{u}_{\tau,\mathbf{R}}}$ is an operator acting on the electron states. A simple explanation of this equation is that, as the atoms in the system move around, the electrons will adjust themselves to find the lowest energy for that atom configuration and therefore the potential energy also changes. It has been shown by Ziman that equation (17) can be derived following a more rigorous quantum mechanical treatment [32]. The resulting matrix element is

$$\langle n_{\mathbf{q}\lambda} \pm 1 | \, \mathbf{u} \, | n_{\mathbf{q}\lambda} \rangle \cdot \langle \mathbf{k}'\beta | \, \partial_{\mathbf{q}\lambda} V \, | \mathbf{k}\alpha \rangle \tag{18}$$

where we have transformed the change of the energy due to the displacement of one single atom $\frac{\partial V}{\partial \mathbf{u}_{\tau,\mathbf{R}}}$ to the potential change $\partial_{\mathbf{q}\lambda} V$ due to a collective atomic motion corresponding to a phonon mode ($\mathbf{q}$, λ). $|\mathbf{k}\alpha\rangle$ and $|\mathbf{k}'\beta\rangle$ are different electron states that are coupled through the perturbation, and $|n_{\mathbf{q}\lambda}\rangle$ represents the phonon state (note that it is characterized by the number of phonons and the final state can only differ by one). The scatterings can be grouped into two main processes. In one case the electron is scattered into another state by absorbing one phonon and thus increasing its energy. Alternatively, the scattering can emit one phonon and the electron energy is lowered. These two processes are distinguished by the first term in equation (18) which, depending on whether the phonon is absorbed or emitted, gives different proportionality factors to the magnitude of the coupling [11].

The second term in equation (18), which includes electron states and perturbed potential, is the key element that is required to calculate the scattering rates and thus the transport properties. For convenience, we define the electron-phonon interaction matrix element as

$$g_{\alpha\beta\lambda}(\mathbf{k}, \mathbf{k}', \mathbf{q}) = \left(\frac{\hbar}{2m_0\omega_{\mathbf{q}\lambda}}\right)^{1/2} \cdot \langle \mathbf{k}'\beta | \, \partial_{\mathbf{q}\lambda} V \, | \mathbf{k}\alpha \rangle \tag{19}$$

which only differs from equation (18) by a prefactor inversely proportional to the square root of the phonon frequency. In equation (19) m_0 is the electron mass. Under the linearized BTE formalism, it can be shown by using the Golden Rule that the collision terms (equation (5)) due to the electron-phonon interaction take the following form [33]

$$\begin{cases} \left(\dfrac{\partial f_\alpha(\mathbf{k})}{\partial t}\right)_{e-ph} \simeq -\left[\displaystyle\sum_{\mathbf{k}'\beta,\mathbf{q}\lambda} F_{\mathbf{k}\alpha}(\mathbf{k}'\beta, \mathbf{q}\lambda)\right] \\ \quad \cdot \Delta f_{\mathbf{k}\alpha} + \displaystyle\sum_{\mathbf{k}'\beta,\mathbf{q}\lambda} [F_{\mathbf{k}'\beta}(\mathbf{k}\alpha, \mathbf{q}\lambda) \cdot \Delta f_{\mathbf{k}'\beta}] \\ \quad + \displaystyle\sum_{\mathbf{k}'\beta,\mathbf{q}\lambda} [F_{\mathbf{q}\lambda}(\mathbf{k}\alpha, \mathbf{k}'\beta) \cdot \Delta n_{\mathbf{q}\lambda}] \\ \quad - \left[\displaystyle\sum_{\mathbf{k}\alpha,\mathbf{k}'\beta} G_{\mathbf{q}\lambda}(\mathbf{k}\alpha, \mathbf{k}'\beta)\right] \cdot \Delta n_{\mathbf{q}\lambda} \\ \left(\dfrac{\partial n_\lambda(\mathbf{q})}{\partial t}\right)_{e-ph} \simeq \displaystyle\sum_{\mathbf{k}\alpha,\mathbf{k}'\beta} [G_{\mathbf{k}\alpha}(\mathbf{k}'\beta, \mathbf{q}\lambda)] \\ \quad \cdot \Delta f_{\mathbf{k}\alpha} + G_{\mathbf{k}'\beta}(\mathbf{k}\alpha, \mathbf{q}\lambda) \cdot \Delta f_{\mathbf{k}'\beta} \end{cases} \tag{20}$$

where the coefficients F and G only depend on the equilibrium distribution functions:

$$\begin{cases} F_{\mathbf{k}\alpha}(\mathbf{k}'\beta, \mathbf{q}\lambda) = [(n^0_{\mathbf{q}\lambda} + f^0_{\mathbf{k}'\beta})\Pi_- + (n^0_{\mathbf{q}\lambda} + 1 - f^0_{\mathbf{k}'\beta})\Pi_+] \\ F_{\mathbf{k}'\beta}(\mathbf{k}\alpha, \mathbf{q}\lambda) = [(n^0_{\mathbf{q}\lambda} + 1 - f^0_{\mathbf{k}\alpha})\Pi_- + (n^0_{\mathbf{q}\lambda} + f^0_{\mathbf{k}\alpha})\Pi_+] \\ F_{\mathbf{q}\lambda}(\mathbf{k}\alpha, \mathbf{k}'\beta) = [(f^0_{\mathbf{k}'\beta} - f^0_{\mathbf{k}\alpha})\Pi_- + (f^0_{\mathbf{k}'\beta} - f^0_{\mathbf{k}\alpha})\Pi_+] \\ G_{\mathbf{k}\alpha}(\mathbf{k}'\beta, \mathbf{q}\lambda) = [-(n^0_{\mathbf{q}\lambda} + f^0_{\mathbf{k}'\beta})\Pi_- + (n^0_{\mathbf{q}\lambda} + 1 - f^0_{\mathbf{k}'\beta})\Pi_- \\ G_{\mathbf{k}'\beta}(\mathbf{k}\alpha, \mathbf{q}\lambda) = [(n^0_{\mathbf{q}\lambda} + 1 - f^0_{\mathbf{k}\alpha})\Pi_- - (n^0_{\mathbf{q}\lambda} + f^0_{\mathbf{k}\alpha})\Pi_+] \\ G_{\mathbf{q}\lambda}(\mathbf{k}\alpha, \mathbf{k}'\beta) = [(f^0_{\mathbf{k}'\beta} - f^0_{\mathbf{k}\alpha})\Pi_- - (f^0_{\mathbf{k}'\beta} - f^0_{\mathbf{k}\alpha})\Pi_+] \end{cases} \tag{21}$$

with $\Pi_\pm$

$$\begin{cases} \Pi_- = \dfrac{2\pi}{\hbar}\,|g_{\alpha\beta\lambda}(\mathbf{k}, \mathbf{k}', \mathbf{q})|^2 \\ \quad \cdot \delta(E_{\mathbf{k}'\beta} - E_{\mathbf{k}\alpha} - \hbar\omega_{\mathbf{q}\lambda}) \cdot \delta_{\mathbf{k}'-\mathbf{k}-\mathbf{q},\mathbf{G}} \\ \Pi_+ = \dfrac{2\pi}{\hbar}\,|g_{\alpha\beta\lambda}(\mathbf{k}, \mathbf{k}', \mathbf{q})|^2 \\ \quad \cdot \delta(E_{\mathbf{k}'\beta} - E_{\mathbf{k}\alpha} + \hbar\omega_{\mathbf{q}\lambda}) \cdot \delta_{\mathbf{k}'-\mathbf{k}+\mathbf{q},\mathbf{G}} \end{cases} \tag{22}$$

denoting the phonon absorption and emission processes, respectively. During these scattering processes, the energy and crystal momentum conservation need to be satisfied, as is clear in equation (22). Similar to the phonon interaction case, the electron-phonon scattering processes can also be classified into normal and Umklapp types. Normal processes occur for electrons and phonons with relatively small wave vectors while Umklapp processes happen for those with large wave vectors [11].

For the phonon-phonon interaction, we have ignored the dependence of the collision terms on the states except for the state of interest; this implies that when considering the scattering of one phonon state, other states are assumed to be at equilibrium. For the electron-phonon interaction, we will however show that the dependence of the electron scattering on other phonon states cannot be neglected in some situations. To simplify equation (20), we see that there are terms directly proportional to the state of interest (just as the case for phonon-phonon interaction), which can be rewritten by defining new relaxation times due to the electron-phonon coupling

$$\begin{cases} \dfrac{1}{\tau^{e-ph}_{\mathbf{k}\alpha}} = \displaystyle\sum_{\mathbf{k}'\beta,\mathbf{q}\lambda} F_{\mathbf{k}\alpha}(\mathbf{k}'\beta, \mathbf{q}\lambda) \\ \dfrac{1}{\tau^{e-ph}_{\mathbf{q}\lambda}} = \displaystyle\sum_{\mathbf{k}\alpha,\mathbf{k}'\beta} G_{\mathbf{q}\lambda}(\mathbf{k}\alpha, \mathbf{k}'\beta) \end{cases}. \tag{23}$$

The first relaxation time $\tau^{e-ph}_{\mathbf{k}\alpha}$ describes the scattering of electrons by equilibrium phonons, and is the dominant scattering mechanism for the mobility of lightly-doped semiconductors near room temperature or above [11, 13]. For a normal electrical property (electrical conductivity and Seebeck coefficient) calculation, one uses the RTA and assumes $\Delta f_{\mathbf{k}'\beta} = 0$ and $\Delta n_{\mathbf{q}\lambda} = 0$ in equation (20); therefore the collision term can be written in the same form as equation (7), where the relaxation time includes the electron-phonon relaxation time defined by equation (23) and also the relaxation time due to impurities when necessary [13]. The treatment of the impurity scattering will be given below. The distribution function is then readily solved giving equation (8). The electrical conductivity is obtained by summing up all the electron states and dividing the total electrical conduction by the electrochemical potential gradient (assume temperature is uniform):

$$\begin{aligned} \sigma &= \left(\frac{e}{3\Omega N_{\mathbf{k}}}\sum_{\mathbf{k}\alpha} v_{x,\mathbf{k}\alpha}\Delta f_{\mathbf{k}\alpha}\right)/(-\nabla\varphi) \\ &= \frac{e^2}{3\Omega N_{\mathbf{k}}}\sum_{\mathbf{k}\alpha}\mathbf{v}^2_{\mathbf{k}\alpha}\tau_{\mathbf{k}\alpha}\left(-\frac{\partial f^0_{\mathbf{k}\alpha}}{\partial E}\right) \end{aligned} \tag{24}$$

where $N_{\mathbf{k}}$ is the number of the discrete points for electrons in the reciprocal space. Note that we have used the deviation of the distribution since at equilibrium the electrical conduction will be zero. The Seebeck coefficient measures the induced voltage difference across a sample in response to a given temperature gradient. In this case we keep the term in $\Delta f_{\mathbf{k}\alpha}$ that is proportional to the temperature gradient and obtain

$$\begin{aligned} S_{\text{diff}} &= \left(\frac{e}{3\Omega N_{\mathbf{k}}}\sum_{\mathbf{k}\alpha} v_{x,\mathbf{k}\alpha}\Delta f_{\mathbf{k}\alpha}\right)/(\sigma\nabla T) \\ &= \frac{e}{3\sigma\Omega N_{\mathbf{k}}T}\sum_{\mathbf{k}\alpha}(E-\mu)\mathbf{v}^2_{\mathbf{k}\alpha}\tau_{\mathbf{k}\alpha}\left(-\frac{\partial f^0_{\mathbf{k}\alpha}}{\partial E}\right). \end{aligned} \tag{25}$$

We call this Seebeck coefficient diffusive because it originates from the diffusion of electrons. This is also distinguished from another contribution to the Seebeck effect which will be clear later. We can also obtain the electronic contribution to the thermal conductivity

$$\kappa_e = \kappa_0 - T\sigma S^2 \tag{26}$$

where

$$\kappa_0 = \frac{1}{3\Omega N_{\mathbf{k}}T}\sum_{\mathbf{k}\alpha}(E-\mu)^2\mathbf{v}^2_{\mathbf{k}\alpha}\tau_{\mathbf{k}\alpha}\left(-\frac{\partial f^0_{\mathbf{k}\alpha}}{\partial E}\right). \tag{27}$$

The second relaxation time $\tau^{e-ph}_{\mathbf{q}\lambda}$ in equation (23) describes the scatterings of phonons by equilibrium electrons [34]. As we have mentioned, typically this is not as significant as the phonon-phonon scattering. However, if we carefully examine its dependence on the electron state, we will see that this type of scattering strongly increases as the Fermi level increases (or the carrier concentration increases). This implies that, in heavily-doped semiconductors, the phonon scattering by electrons might play a non-negligible role. We will clarify this point more when discussing the first-principles results.

In evaluating the electrical property, we have assumed $\Delta f_{\mathbf{k}'\beta} = 0$ in equation (20). We should note that this is neglected essentially because the terms containing $\Delta f_{\mathbf{k}'\beta}$ sum up to approximately zero. In metals and for elastic scattering with impurities, this approximation is not valid and therefore an extra correction term $(1 - \cos\theta)$ is often added to the electron-phonon relaxation time to take into account the terms containing $\Delta f_{\mathbf{k}'\beta}$, leading to the so-called momentum

relaxation time [13, 35]

$$\frac{1}{\tau^{e-ph,\text{momentum}}_{\mathbf{k}\alpha}} = \sum_{\mathbf{k}'\beta,\mathbf{q}\lambda} F_{\mathbf{k}\alpha}(\mathbf{k}'\beta, \mathbf{q}\lambda)(1 - \cos\theta) \quad (28)$$

where θ is the angle between the two velocity directions of the initial and final electron state. In semiconductors, however, it has been proven, based on deformation potential models, that for nearly isotropic scattering, the neglect of $\Delta f_{\mathbf{k}'\beta}$ will not cause much difference [13]. We will also show that without considering terms containing $\Delta f_{\mathbf{k}'\beta}$, good agreement for the electrical properties in silicon can be achieved with experiments [36–38]. In essence, the momentum relaxation time is only an approximation to the full expression in equation (20), and further work should be done to evaluate the accuracy associated with the use of the two relaxation times for calculating the transport property of semiconductors within the first-principles approach.

Another important simplification we have made when evaluating the electrical property is $\Delta n_{\mathbf{q}\lambda} = 0$, for the reason that the frequent phonon-phonon Umklapp scatterings normally restrict the phonon distribution function close to the equilibrium [39]. This assumption makes non-equilibrium phonons have no effect on the electron system. However, below the Debye temperature, phonons have larger deviations from the equilibrium because the phonon scatterings are largely suppressed by the decreased temperature. These non-equilibrium phonons (non-zero $\Delta n_{\mathbf{q}\lambda}$) in the electron system (equation (20)) then contribute to an extra scattering for the electron's collision terms. Under a given temperature gradient, we use the RTA for the phonon system and write $\Delta n_{\mathbf{q}\lambda} = \tau_{\mathbf{q}\lambda}\mathbf{v}_{\mathbf{q}\lambda} \cdot \nabla T \frac{\partial n^0_{\mathbf{q}\lambda}}{\partial T}$. The electron-BTE can then be solved giving the electron distribution function (here we only consider the non-equilibrium part that results from the non-equilibrium phonons):

$$\Delta f_{\mathbf{k}\alpha} = \tau_{\mathbf{k}\alpha}\nabla T \cdot \sum_{\mathbf{k}'\beta,\mathbf{q}\lambda}\left[F_{\mathbf{q}\lambda}(\mathbf{k}\alpha, \mathbf{k}'\beta) \cdot \tau_{\mathbf{q}\lambda}\mathbf{v}_{\mathbf{q}\lambda}\frac{\partial n^0_{\mathbf{q}\lambda}}{\partial T}\right]. \quad (29)$$

Considering the resulting electrical current per temperature gradient, we arrive at an extra contribution to the Seebeck coefficient due to the non-equilibrium phonons

$$S_{\text{ph}} = \frac{2e}{3\sigma\Omega N_{\mathbf{k}} N_{\mathbf{q}} k_B T^2}\sum_{\mathbf{q}\lambda}[\hbar\omega_{\mathbf{q}\lambda}\tau_{\mathbf{q}\lambda}\mathbf{v}_{\mathbf{q}\lambda} \cdot \times\left(\sum_{\mathbf{k}\alpha,\mathbf{k}'\beta}(\tau_{\mathbf{k}\alpha}\mathbf{v}_{\mathbf{k}\alpha} - \tau_{\mathbf{k}'\beta}\mathbf{v}_{\mathbf{k}'\beta})f^0_{\mathbf{k}\alpha}(1 - f^0_{\mathbf{k}'\beta})n^0_{\mathbf{q}\lambda}\Pi_-\right)\right]. \quad (30)$$

This effect describes an extra current generated when there is a phonon heat flow due to the non-equilibrium phonons, as if the electrons were 'dragged' by phonons. Therefore it has been dubbed the phonon drag effect [33, 39, 40]. We will discuss this effect further in section 4. Here we want to mention that the above picture describing the phonon drag effect is based on the Seebeck effect, where a temperature gradient induces a phonon heat flow, which delivers its momenta to the electron system and creates the electrical current. Because of the Kelvin relation $\Pi = TS$ [15, 41, 42], an extra contribution to the Seebeck coefficient also implies an extra Peltier coefficient. This can be understood by the first two terms of the collision terms for the phonon in equation (20), which transfer the momenta of the electron system to phonons when there is an electrical current and thus non-zero $\Delta f_{\mathbf{k}\alpha}$. We will not examine this second case but point out that the derivation based on the Peltier picture will lead to the same result as shown by equation (30).

All of the above transport properties are contained in equation (20), of which the key element is the electron-phonon interaction matrix element given by equation (19). In evaluating the electron-phonon coupling matrix, besides the electron wavefunctions and phonon modes, the change of the electron potential energy with respect to the atomic displacements is required. One way is to use the deformation potential model [11, 13], which connects the magnitude of the perturbed potential to the change of the band edge energy with respect to the lattice dilation. By using the deformation potential, the large amount of the coupling matrix elements are not needed, thereby greatly saving the computational cost. However the use of this parameter ignores the dependence of the electron-phonon coupling on the electron states, and is only valid for band edge states and long-wavelength phonons, which cannot be accurate for materials in general. A more accurate way is to use DFPT, because when calculating the phonon dispersion, DFPT already obtains the perturbed potential and therefore provides the electron-phonon coupling matrix as a by-product. However, in contrast to the use of real space force constants in calculating the phonon-phonon coupling matrix, direct electron-phonon coupling matrix calculations using DFPT are time-consuming. It will thus be favorable to develop methods to reduce the DFPT calculation and use formulas similar to equation (11) to obtain the electron-phonon coupling matrix for any state, since a fine mesh is necessary for the convergence of the transition rates. Such an approach has recently been developed [43, 44], which utilizes the maximally-localized Wannier functions [45] to transform the DFPT results from the reciprocal space to the real space. In the real space, the electron states can be described by orbital-like states while the perturbed potential is centered around the displaced atom, both of which are localized in real space [43]. These essentially serve as the tight-binding basis and can be used to construct the electron and phonon states as well as their coupling matrix for any given states. In fact, this whole process can be understood as a generalization of the tight-binding model. An efficient calculation of the electron-phonon coupling matrix can then be done with DFPT results on a coarse mesh only. This method has allowed the electron-phonon coupling constant to be examined for many materials, especially superconductors [46–48]. We refer readers to the literature for more details on the Wannier states [45] and the extraction of electron-phonon matrix elements [43, 44]. We also want to note that, as pointed out by the original work [43], the Wannier interpolation for the electron-phonon coupling matrix will be less

accurate if there are long-range interactions in the system, such as the electric field accompanied by the longitudinal optical (LO) phonon near the zone center in polar semiconductors. For these materials, these long-range forces need to be separately treated, as has been done for the LO-TO splitting in the phonon dispersion calculation using DFPT [19, 49, 49]. Recently these long-range effects have been incorporated into the Wannier interpolation scheme by using a rigid ion model (assigning effective charges to each atom and considering their Coulomb interactions with electrons) [50] to approximate the electron-phonon matrix element for long-wavelength LO phonons [51, 52]. This development thus further extends the efficient computation of electron-phonon couplings towards polar semiconductors.

2.3. Impurity scattering

In semiconductor engineering, external methods that modify the physical properties of the materials are indispensable, with examples including doping and alloying. For the thermoelectric property in particular, doping changes the carrier concentration and is used to find the optimal carrier concentration for achieving a higher efficiency. On the other hand the alloying provides a way to modify the band structure for a higher Seebeck coefficient [53] or to create more phonon scatterings for lower thermal conductivity [54]. In general, these material processing techniques break the translational symmetry of the periodic crystal, leading to further scatterings of electrons and phonons. In most cases they are treated as perturbations to the pristine crystal, therefore the scattering rates can be studied by a perturbative approach. Generally we regard them as impurity scattering because they introduce irregularities into the host material and have similar treatment for the scattering rate calculation.

First we consider neutral impurities of a single type embedded in a medium. The key idea of the perturbative approach is to replace the real material with a pristine one, therefore justifying the use of the eigenstate [55]. In a more general sense, when considering a disordered crystal, one can replace that with an ordered virtual crystal that has an averaged property (structure, force constant, electron potential, etc), which is also known as the virtual crystal approximation (VCA) [55, 56]. The disorder is then treated as a perturbation to the system. For relatively long-wavelength phonons, this approximation is reasonable because these phonons mainly see an averaged potential [55]. It is questionable when the wavelength of phonons is comparable to the characteristic length of the disorder. It has been shown by using a full-order scattering theory that the VCA fails for high-frequency phonons and predicts an incorrect phonon density of state [57]. However, such high frequency phonons are shown to have large scattering rates [57], which make VCA still a good approximation in evaluating the heat transport of alloys.

The perturbation due to the introduction of the impurities leads to scatterings between the original states. Using the Golden Rule, we can write the scattering rates of phonons due to the impurity scattering as [58]

$$\frac{1}{\tau^{ph-imp}_{\mathbf{q}\lambda}} = f\sum_{\mathbf{q}'\lambda'}\frac{\pi\Omega}{2\omega^2_{\mathbf{q}\lambda}}\frac{1}{V_s}\,|\langle\mathbf{q}\lambda|\ \mathbf{T}\ |\mathbf{q}'\lambda'\rangle|^2\delta(\omega_{\mathbf{q}\lambda}-\omega_{\mathbf{q}'\lambda'}) \tag{31}$$

where f is the volume fraction of the impurity, V_s is the volume of the scatterer, $|\mathbf{q}\lambda\rangle$ describes the phonon eigenmode normalized to the unit cell volume Ω and $\mathbf{T}$ is the matrix that describes the perturbation induced by a single impurity that couples different phonon modes. Note that in equation (31) we have assumed a dilute limit for which the interaction between impurities is ignored. In most cases, the T-matrix is approximated by the perturbation matrix, known as the Born approximation, which can be shown to give rise to a scattering rate proportional to the frequency to the fourth power [59, 60] (Rayleigh scattering). Similarly, for electrons we can write down the scattering rates based on the potential difference ΔV between the pristine material and the one with impurity:

$$\frac{1}{\tau^{e-imp}_{\mathbf{k}\alpha}} = \frac{c}{\Omega N_{\mathbf{k}}}\sum_{\mathbf{k}'\beta}\frac{2\pi}{\hbar}\,|\langle\mathbf{k}'\beta|\ \Delta V\ |\mathbf{k}\alpha\rangle|^2(1-\cos\theta) \tag{32}$$

where c is the density of the impurity, $|\mathbf{k}\alpha\rangle$ describes the electron eigenstates and θ is the angle between the group velocity vectors of the initial and final state. This perturbative approach has been widely applied to study the influence of neutral defects on a material's physical property. The calculation is usually performed in a supercell where an atom is replaced by the defect. By comparing the results with those obtained in the same supercell of pristine crystal, defect-induced perturbation can be extracted and used for understanding how the defect modifies the material's property. A common example is the calculation of formation energy of different substitutes in some host materials [61]. We will see that this method has been combined with the first-principles calculation to study the defect scattering of electrons and phonons. This perturbative thinking can also be applied to the transport studies of alloys using VCA (instead of constructing a supercell) as we have mentioned. We will see that good agreements have been achieved for the electrical and thermal conductivity of the SiGe alloy [62, 63], as well as the thermal transport in some other alloys [64–66], thereby justifying the applicability of VCA. It should be noted that the perturbed potential (either for electrons or phonons) in the above situations is relatively short-ranged and therefore a supercell in which the perturbed potential decays to zero is possible within the DFT capability. In fact, the recovering of bulk potential at distances away from the defect has been used to set the correct energy reference shift in some calculations [62].

However, when the impurities introduced into the material are not charge-neutral (for example ionized dopants), they will create long-range Coulomb interaction, which decays much slower and makes the supercell approach inappropriate due to the interaction between the dopant and its artificial images entailed by the periodic boundary condition. Besides, when modeling the charged impurity in a supercell,

the net charge density would lead to the divergence of the total energy. The way to overcome this difficulty is to introduce a uniform jellium charge background, making the whole system charge-neutral. This trick however introduces a spurious interaction between the electron and the background charge. Though the spurious interactions decay to zero as the supercell size increases to infinity, within the current DFT capability such a large supercell is almost impossible.

For comparison between dopant elements in the same column, a supercell can still be used in spite of the long-range Coulomb potential [67]. This is because the elements have the same charge and therefore an incorrect treatment on the Coulomb interactions will not affect their comparison. To quantitatively evaluate the electron scattering by the charged impurities, one must find ways to extend the screened Coulomb potential beyond the supercell. Rurali *et al* [68] have found in the silicon nanowire transmission calculation that the perturbed potential induced by a single charged impurity converges quickly around the dopant; in addition they discovered that its values in the outer region of the dopant, though varying as the system size increases, can be reproduced by a simple screened electrostatic model. Therefore the potential corresponding to a long nanowire is constructed by connecting the converged potential in the inner region and the screened electrostatic potential in the outer region. Similarly, a screened Thomas–Fermi potential was used by Restrepo *et al* to extend the potential obtained from the first-principles calculation beyond the supercell, and the perturbed potential in the whole space is then used to calculate the electron-impurity scattering rates in silicon [36]. However, we find that there has not been a consensus on which physical model is better suited for extending the screened Coulomb potential beyond the supercell and how accurate these methods are in different situations. Regarding the thermodynamic information (e.g. formation energy) of the point defects in solids, there have been great developments on the correction schemes for the spurious interactions due to the long-ranged forces [69]. It will be desirable to see how similar ideas can be used to facilitate the calculation of the electron scattering rate due to the charged impurities, which has control over the mobility of heavily-doped semiconductors.

3. Thermal transport in semiconductors

From this section on, we discuss the first-principles results on the transport properties of semiconductors. We first examine the thermal transport properties. Some previous review papers have discussed the first-principles calculations of thermal conductivity, mainly to illustrate the power and accuracy of first-principles methods [70, 71], by comparing with the experimental data. In the following our discussions will specifically focus on the column IV and III–V semiconductors. To give a unified picture for both phonons and electrons, we will discuss their relaxation times and MFPs from pure crystals to alloys. We show that the first-principles calculation not only enables us to investigate the thermal transport properties in a variety of materials with unprecedented details, but also provides new insights into some of the old topics, for which traditional perspectives may not be entirely correct.

3.1. Phonon relaxation times

Among all the semiconductors, silicon is probably the one most studied. It can serve as a testbed for first-principles computational tools and also provides insightful details into phonon transport properties in general. However, a first-principles prediction of its thermal conductivity has been challenging, because of the requirement of obtaining accurate harmonic and anharmonic force constants that lead to phonon dispersion and phonon scattering rates. Force constants extracted from the empirical interatomic potential have been shown to give qualitative agreement with experiments but the accuracy is far from being satisfactory [72]. The thermal conductivity result also largely depends on the empirical potential chosen to use, therefore questioning such an approach as a quantitative tool for evaluating thermal transport. By using the BTE formalism with force constants extracted from DFT calculations, Broido *et al* were able to first show that the first-principles approach can predict the thermal conductivity of silicon and germanium within 5% of the experimental value at room temperature without any adjustable parameter [17], where a reciprocal space method is used to extract force constants. The real space approach, developed by Esfarjani *et al* [16, 20], further allows an easy extraction of force constants in more complex materials, and was soon applied by many groups to other materials, including III–V semiconductors [73–75], II–VI binary compounds [76], typical thermoelectric materials [64–66, 77–85] and 2D materials [86–93]. We should note that not all of these works used the RTA. In fact, in materials like graphenewhere normal processes are more frequent compared to Umklapp processes, it was shown that an exact solution of the full BTE is necessary to obtain quantitative agreement with experiments [87, 94], which is also used in the original work on silicon [17]. However for silicon, germanium and III–V semiconductors, Umklapp scatterings are relatively strong and as a result the use of RTA only introduces small errors compared to an exact calculation [16, 74].

The first-principles approach allows us to look into the details of each phonon mode in contributing to the thermal transport. In Si and Ge, it has been found that more than 90% of the heat is carried by acoustic phonons at room temperature [17]. While the optical phonons carry much less heat than acoustic phonons, they provide significant scattering channels for the acoustic phonons. For Si and Ge, it was found that more than 50% of the scattering processes of the acoustic phonons involve at least one optical phonon [26]. Therefore, neglecting the optical branch (in the sense that their scatterings with acoustic phonons are also neglected) will largely overestimate the thermal conductivity. Similar results have also been obtained in GaAs, where over 90% of the heat is carried by acoustic phonons over a temperature range from 100 K to 400 K [75]. These results thus quantified our qualitative understanding based on the theoretical models.

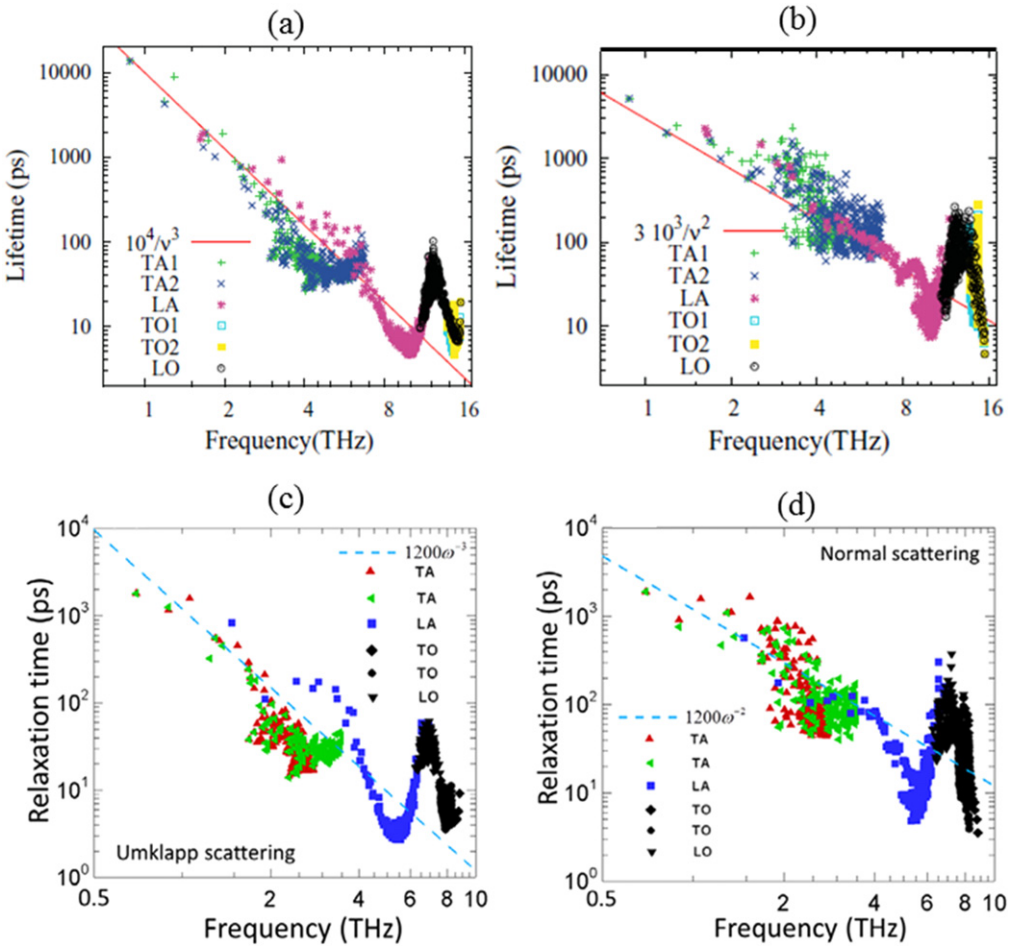

Figure 2. Phonon relaxation times from first principles with respect to the phonon frequency for (a) Umklapp processes in Si, (b) normal processes in Si, (c) Umklapp processes in GaAs and (d) normal processes in GaAs. Note that for the low-frequency phonons, the normal process relaxation time has been fitted with ω^2 while the Umklapp process relaxation time is fitted using ω^3. (Si results reprinted with permission from Esfarjani *et al* [16]. Copyright (2015) American Physical Society; GaAs results reprinted with permission from Luo *et al* [75].)

From the first-principles method, we can further examine the mode-dependent relaxation times. It has been found that in Si and Ge the normal process scattering rate scales with ω^2 while the Umklapp process scattering rate scales with $\omega^{3\sim4}$ [16, 26]. Figures 2(a) and (b) show the frequency dependence of the relaxation times in Si. Normal scattering usually dominates for relatively low-frequency phonons and the squared frequency dependence agrees with the commonly-used form, while the Umklapp process starts to dominate at higher frequencies, giving rise to a higher-order frequency dependence. However such higher-order frequency dependence has been neglected in many recent room temperature RTA calculations. The first-principles results thus highlight the importance of using the correct scaling behavior of the relaxation time, which directly affects the estimation of the thermal conductivity [26]. The relaxation times of GaAs are also shown in figure 2, exhibiting similar frequency scaling behavior [75]. The majority of the relaxation times at high-symmetry points at 300 K are within the range of $3 \sim 10$ ps, while TO modes at the L and X points have longer relaxation times of ~30 ps, as shown in figure 3. The quantitative determination of these relaxation times therefore provides critical information for hot-phonon effects in many optoelectronic devices.

Such a first-principles calculation has also been applied to study the heat conduction in a superlattice [99], where one obtains the force constants from a virtual crystal calculation with averaged masses and the masses of different atoms are included via equation (11). For an infinite GaAs/AlAs superlattice, it has been found that by treating the interface roughness as a random mixing of Ga and Al atoms in a narrow range, the low-frequency phonons dominate the heat conduction and are mainly affected by phonon-phonon scatterings, while higher frequency phonons are also scattered due to the interfacial disorder [99], as shown in figure 4.

The origin of some thermal transport features can also be understood by studying the scattering channels for each phonon branch. Work has been done to systematically examine the temperature-dependent thermal conductivity of III–V semiconductors using first principles [74], including all the binary compounds determined by (Al, Ga, In)–(P, As, Sb). The effects of different elements on the phonon properties can then be conveniently compared. It was found that the gap

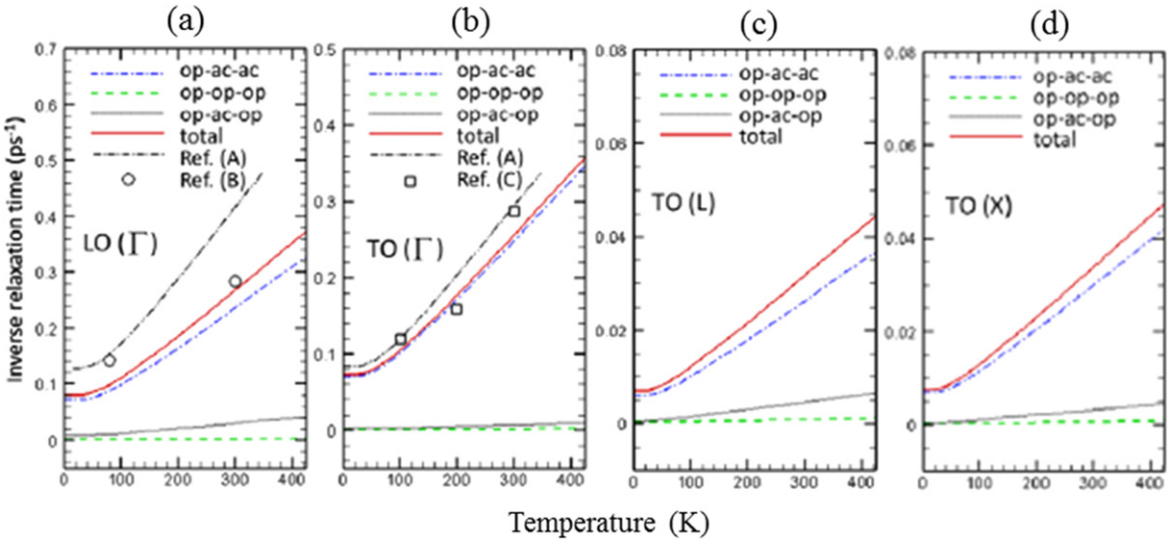

Figure 3. Phonon scattering rates in GaAs as a function of temperature for (a) LO phonon at Γ-point, (b) TO phonon at Γ-point, (c) TO phonon at L-point and (d) TO phonon at X-point. Shown in the figure is the total scattering rate as well as its decomposition into different types of three-phonon scattering processes (op: optical phonon; ac: acoustic phonon). The obtained first-principles results (lines) are compared to previous first-principles calculations (reference (A) from [95] and reference (C) from [96], only performed at Γ-point) as well as experimental data (reference (B) from [97, 98]). (Reprinted figure with permission from Luo *et al* [75].)

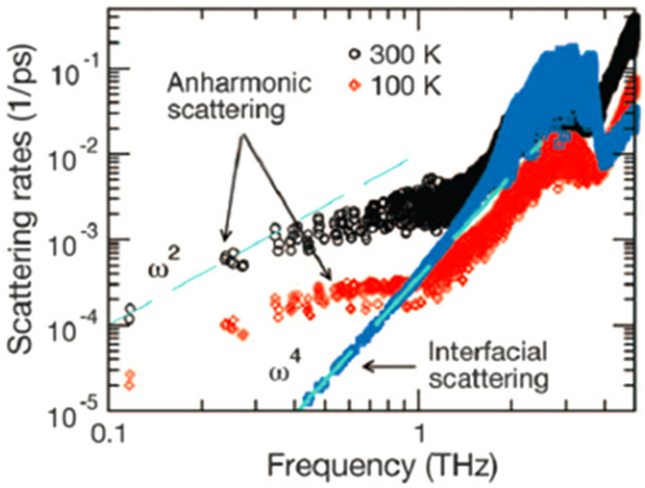

Figure 4. Phonon relaxation times from first principles with respect to the phonon frequency for an infinite GaAs/AlAs superlattice. The plot shows the intrinsic phonon scattering rates at two different temperatures (100 K and 300 K) as well as the temperature-independent interfacial scattering. It is clear that for low-frequency phonons (less than 1 THz) the anharmonic scattering dominates while for higher-frequency phonons the interfacial scattering starts to play an important role. (Figure reprinted from Luckyanova *et al* [99]. Reprinted with permission from AAAS.)

between the acoustic and optical branch will modify the scattering processes of the acoustic phonons (due to the requirement of the energy conservation), and therefore consistently affect the thermal conductivity when the V-element moves downwards in the periodic table. By taking into account the phonon-isotope scattering (the treatment essentially uses equation (31) but with Born approximation, which reduces to the Tamura model [60]), good agreements with the experiments can be achieved in these materials [74]. A further study into GaN reveals that the isotope scattering is relatively strong compared to the intrinsic phonon-phonon scattering, and an isotopic enrichment should increase the thermal conductivity of GaN by ~65% at room temperature [73]. This large 'isotope effect' was claimed to result from (1) a large gap between the acoustic and optical branch and (2) a large phonon energy scale, both of which reduce the phonon-phonon scattering rates and make the isotope scattering relatively more significant [73]. These calculations offered new insights into the isotope effect in semiconducting materials and have implications for engineering the thermal transport properties.

3.2. Phonon MFP distribution

The first-principles method can also provide details into the MFP distribution of phonons, which is important information for nanoscale applications due to the size effect. To clarify the contribution to the heat conduction from different phonon modes, a thermal conductivity accumulation function has been introduced by Dames and Chen [100], which is defined as the contribution to the thermal conductivity from all phonons with MFPs below a certain value. Although a kinetic formula $\kappa = \frac{1}{3} c_v v \Lambda$ typically uses a single MFP, the real crystal can have a MFP distribution spanning across several orders of magnitude from submicron to hundreds of microns. For example, an analysis based on the kinetic formula predicts a MFP of 41 nm in Si [101]. The spectral phonon property analysis based on the molecular dynamics simulation, however, shows that the phonon MFPs in Si span from 30 nm to 100 μm at room temperature [102]. This emphasizes the importance of recognizing the broad MFP distribution when considering the nanostructures where long MFP

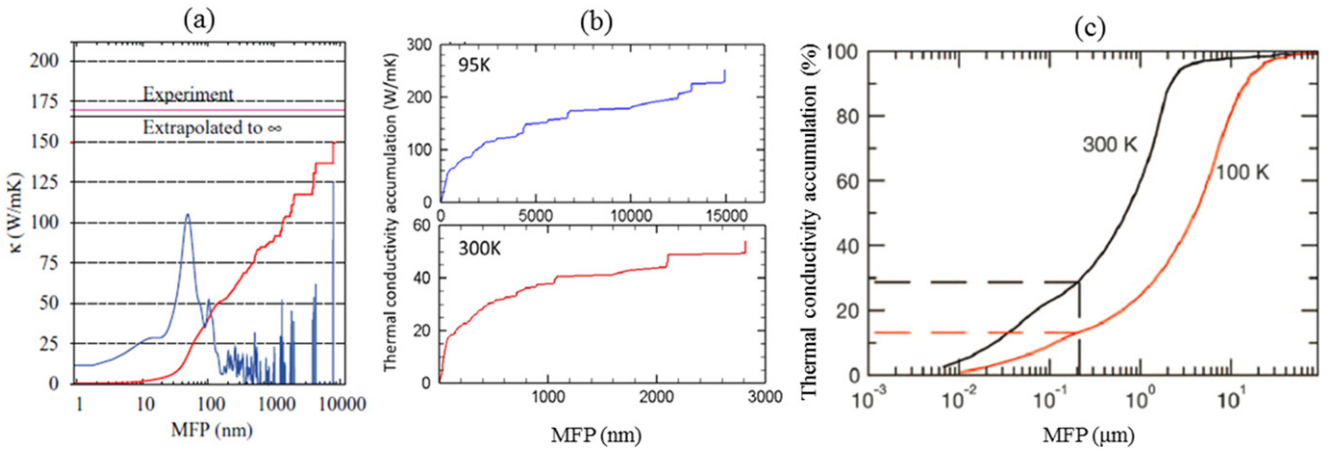

Figure 5. Accumulated contribution to thermal conductivity from first principles with respect to the phonon MFP for (a) Si, (b) GaAs and (c) GaAs/AlAs superlattices. The results for Si are evaluated at 277 K while the other two materials are evaluated at both room temperature and cryogenic temperatures. The 'sawtooth' features for long MFP phonons are due to the finite mesh size used. (Si results reprinted from Esfarjani *et al* [16]. Copyright (2015) American Physical Society. GaAs results reprinted with permission from Luo *et al* [75]. GaAs/AlAs superlattice results reprinted with permission from Luckyanova *et al* [99]. Reprinted with permission from AAAS.)

phonons will be more frequently scattered by interfaces and boundaries, for which a diffusive transport model will break down and ballistic effects must be considered.

Compared to the molecular dynamics simulation, we regard the first-principles calculation as a more accurate way to look into the spectral phonon properties of materials. It has been shown that, at room temperature more than 50% of the heat is carried by phonons with MFPs longer than 1 μm in Si [16]. This fraction will become larger at lower temperatures. It should also be noted that in Si the MFPs of most phonons span almost four orders of magnitude, from 10 nm to 100 μm at room temperature (figure 5(a)). Similarly in GaAs, the MFPs span over three orders of magnitude from 10 nm to more than 30 μm [75]. It was also found that more than 50% of the thermal transport comes from phonons with MFPs longer than 350 nm at room temperature, and at a lower temperature (95 K) 50% of the heat is carried by phonons with MFPs longer than 4 microns (figure 5(b)). In an infinite GaAs/AlAs superlattice, phonons with MFPs longer than 216 nm were found to contribute 87% (at 100 K) and 71% (at 300 K) to the total heat conduction [99] (figure 5(c)). We want to briefly mention that the recent development in the phonon MFP spectroscopy has opened up the possibility of reconstructing such MFP accumulated thermal conductivity information from the experimental data based on quasi-ballistic heat transport [103–105]. Good agreement has been achieved in Si between the reconstructed accumulation curve and the first-principles results [106–108]. Clearly such an understanding of the MFP distribution is crucial for modeling and engineering the nanostructures, and the first-principles calculation provides an accurate evaluation for such information.

3.3. Effect of alloying and doping

First-principles calculation on an alloy system was first performed for a SiGe alloy using the virtual crystal approximation by Garg *et al* [63]. It was uncovered that more than 50% of the heat conduction is carried by phonons with a MFP longer than 1 μm, as shown in figure 6. Besides, the large reduction of the thermal conductivity with only a small fraction of Ge added into Si is well-captured by the first-principles calculation and agrees reasonably well with the experiment at room temperature. Compared to the pristine crystals, the mass disorder more strongly scatters high-frequency phonons and makes low-frequency (long-wavelength) phonons more significant in the thermal transport. Similar materials, including $PbTe_xSe_{1-x}$, $Mg_2Si_xSn_{1-x}$ and BiSb alloy, have also been studied using the same VCA [64–66]. These results provide a deeper understanding of the thermal transport in alloy systems and also guidelines for engineering the thermal performance of an alloy structure.

Doping procedure introduces ionized dopants and electrons into the host material, which also increases the scattering rates for phonons. It has long been thought that the reduction of the thermal conductivity of semiconductors is mainly due to the impurity scatterings from the dopants. The electron scattering of phonons, though formalized and discussed in metal systems [34, 109, 110], was regarded as insignificant for semiconductors. With the first-principles approach, one can now incorporate the phonon scattering rate due to electrons (the second line in equation (23)) into the intrinsic phonon-phonon scattering rate (equation (10)) using Matthiessen's rule $1/\tau_{\mathbf{q}\lambda} = 1/\tau^{ph-ph}_{\mathbf{q}\lambda} + 1/\tau^{e-ph}_{\mathbf{q}\lambda}$. The effect of the electron scattering of phonons can then be evaluated at different temperatures and doping levels. Surprisingly, it has been found in Si that the thermal conductivity starts to decrease as the carrier concentration goes beyond 10^{19} cm^{-3} and the reduction reaches 45% in p-type silicon at around 10^{21} cm^{-3} [111], as shown by figure 7(a). The major scatterings due to electron-phonon coupling are found in low-frequency acoustic phonons and optical phonons, as shown in figure 7(b). In fact, based on the deformation potential model,

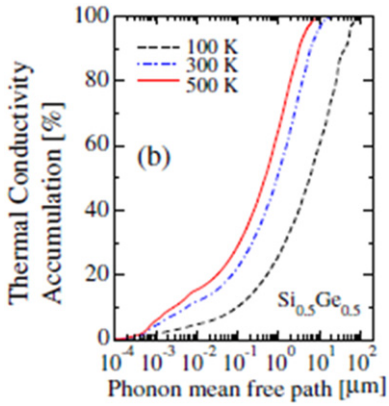

Figure 6. Accumulated thermal conductivity from first principles with respect to the phonon MFP in a $Si_{0.5}Ge_{0.5}$ alloy. (Reprinted figure with permission from Garg *et al* [63]. Copyright (2015) American Physical Society).

it can be found that the phonon scattering rates due to electrons linearly scales with the phonon frequency [111] while, as we have discussed above, the intrinsic phonon-phonon scattering is dominated by the normal process for low-frequency phonons, with a frequency dependence as a frequency to the second power. The latter therefore drops faster than the former and as the carrier concentration increases low-frequency phonons will be more strongly scattered by the electrons. This finding has great implications for thermoelectric applications, which typically involve heavily-doped semiconductors, and also provide a modern understanding of the effect of the electron-phonon coupling on the thermal transport.

4. Electrical transport of semiconductors

While the first-principles calculation for the phonon thermal transport has been largely improved in the past decade, most of the evaluations of the electrical transport properties still require the use of parametrized models such as constant RTA or deformation potential approximation [112]. These parametrized models for the electron scattering process have helped us to understand the electron transport properties of certain thermoelectric materials and give qualitative agreement with the experiment [113–117]. However, such methods lack accuracy if the band structure becomes complicated, in which case the electron-phonon coupling no longer satisfies a simple parametrized model. Therefore it is necessary to have an accurate description for the electron-phonon scattering through the first-principles approach. Only very recently has the first-principles method been applied to study the electrical properties including mobility and the Seebeck coefficient in silicon [36, 37]. For the first-principles calculation of electrical properties using parametrized models for electron scattering, readers can refer to [112]. Here only the work that has applied the full calculation of the electron-phonon scattering rates will be discussed in detail. As a widely-studied material, silicon still presents a challenge to the electrical property calculation and can offer new perspectives to the thermoelectric application as we will show. In the following we will discuss this material and some other efforts on III–V semiconductors, and also explain the difficulties involved in these calculations.

4.1. Electron relaxation times, MFPs and mobility

At the center of the electrical property calculation is the electron relaxation time. For semiconductors there are two main mechanisms that lead to the scatterings of electrons: electron-phonon scattering and electron-impurity scattering. In lightly-doped silicon, the electron-phonon scattering dominates over most of the temperature range. As the doping level increases, the electrons near the band edge are strongly scattered by the ionized dopants, resulting in a decreased mobility. Traditionally, the scattering has been treated by making approximations for the interaction matrix. For example, the electron-phonon interaction matrix element is readily described by the use of the deformation potential while the electron-impurity interaction is approximated using the screened Coulomb potential model [13]. By considering both electron-phonon and electron-impurity scattering in solving the BTE formula, good agreements have been achieved for the carrier concentration dependent mobility compared to the experiment [113]. However, the use of the parametrized model neglects details of the electron transport such as the dependence of the electron-phonon coupling on the electron energy and wave vector, and therefore is not satisfactory as a quantitative predictive tool. The first-principles approach provides such information with no adjustable parameters, and is more accurate and suitable for studying the contribution to the electrical transport from each electron state. The first-principles method was first applied to study the intervalley scatterings between certain electron pockets in Ge, GaAs and GaP [118, 119]. The results obtained agree well with the experiments, justifying the use of the first-principles approach in evaluating the electron-phonon scattering rate. These results also have implications for high-field electrical transport and the relaxation of hot carriers. However, to accurately predict the electrical transport property, we must recognize the difference of the relaxation times between different electron states. Such an analysis based on first-principles calculation was given by Restrepo *et al* [36], who showed that the electron scattering rates in silicon on a large energy scale (several electron volts) follow the shape of the electron density of state. This general feature comes from the fact that at energy levels away from the band edge, the scattering rates are mainly limited by the available states that one electron can be scattered into, and therefore a higher density of state provides more scattering channels and reduces the scattering time. In their original work, only the energy dependence of the scattering rate is explicitly given [36]. A further detailed study into Si and Ge shows that the electron-phonon coupling strength varies significantly along different crystallographic directions, which causes wave-vector dependence of the scattering rates as well [120], as shown in figures 8(a) and (b). We note that such a full examination of

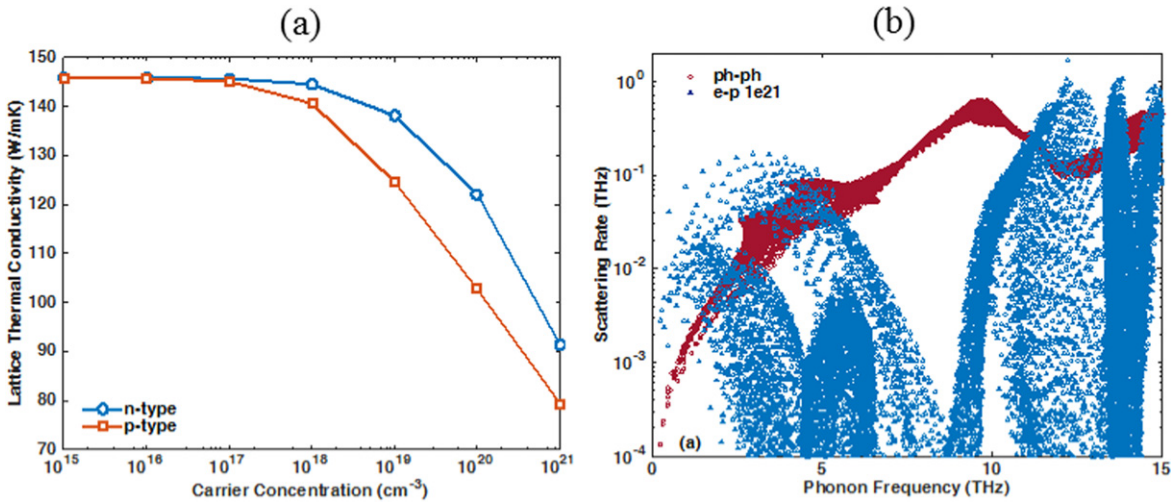

Figure 7. (a) Reduction of lattice thermal conductivity due to the phonon scattering by electrons for both n-type (blue curve) and p-type (red curve) silicon at 300 K, and (b) the comparison between phonon-phonon scattering rates (red dots) and phonon-electron scattering rates (blue dots) for n-type silicon. The carrier concentration for part (b) is assumed to be 10^{21} cm^{-3}. (Reprinted figures with permission from Liao *et al* [111]. Copyright (2015) American Physical Society).

the electron-phonon scattering over the whole Brillouin zone has enabled the analysis of the phonon-assisted optical absorption as well as the hot carrier relaxation process in silicon [14, 121], and is also crucial for a first-principles prediction of the electrical property. Hot carrier relaxation has also been examined in GaAs with the electron scattering times being studied over the whole Brillouin zone [122] (figure 8(c)). It was suggested that the polar optical phonon scattering is not as strong as previously believed and does not dominate over the acoustic phonon scattering for hot electrons.

These calculations based on the first-principles approach have offered insights into the regimes which old models cannot reach or reliably analyze, such as the relaxation of hot electrons with a wide distribution of lifetimes and the transition probabilities between different pockets examined independently. For the transport property such as low-field mobility and the Seebeck coefficient, the electrons near the band edge (within a few k_BT from the band edge) are the dominant players, because only these electrons are notably thermally excited due to the Fermi–Dirac statistics. Their mode-dependent scattering rates are more sensitively affected by the environment such as impurities, and require a closer examination. Figure 9(a) shows the electron scattering rates in silicon close to the band edge [38]. The 'kink' at an energy around 0.06 eV above the conduction band edge is due to the phonon emission process (note that the electron needs to have some energy above the conduction band edge to emit a phonon). It is shown that a fine mesh and an adaptive Gaussian broadening parameter for approximating the delta functions when evaluating the energy conservation are required to resolve this 'kink' feature [38]. As the doping level increases, the band edge electrons will be scattered more strongly by ionized dopants, causing a U-shape scattering rate profile as shown in figure 9(b). Band edge electron relaxation times have also been calculated in GaAs, where 'kinks' in the scattering rate profile can be associated with different electron pockets (figure 9(c)), and a GW calculation (including electron many-body effect) was used to accurately describe the band structure [122].

In addition to the relaxation times, the electron MFP is another important piece of information for nanoscale modeling and engineering. Figure 10(a) shows the electron MFP as a function of the electron energy. At lower doping concentrations, there is a maximal electron MFP with respect to the electron energy. This is because at higher energies electrons suffer more scatterings (figure 9) and have smaller relaxation times, while near the band edge the group velocity tends to be zero. The competence between the relaxation time and the group velocity gives rise to the maximal electron MFP. Figure 10(b) further shows the accumulated contribution to the electrical conductivity with respect to the electron MFP [37]. Compared to the phonon MFP distribution, we see that electron MFPs span over a smaller range than those of phonons, from 10 nm to 100 nm for lightly-doped silicon and from 1 nm to 10 nm for heavily-doped silicon. The sharp cutoff of the accumulation curve for the electrical conductivity is due to the maximal electron MFP, as we have discussed above. The clear difference between the phonon MFP and the electron MFP can be used to quantitatively examine the effect of nanostructuring techniques in reducing the phonon thermal conductivity while maintaining the electrical transport properties [37]. However such detailed MFP information based on first-principles calculations has not been reported for other semiconducting materials. We anticipate that along the path more materials can be studied with detailed examination into their electron scattering mechanisms and MFP distributions.

Given the relaxation time of each electron state, mobility can be calculated based on the BTE solution (equation (24)). An accurate prediction of the transport property will involve integration over the whole Brillouin zone, thereby requiring a large amount of relaxation time data, which in turn depend on

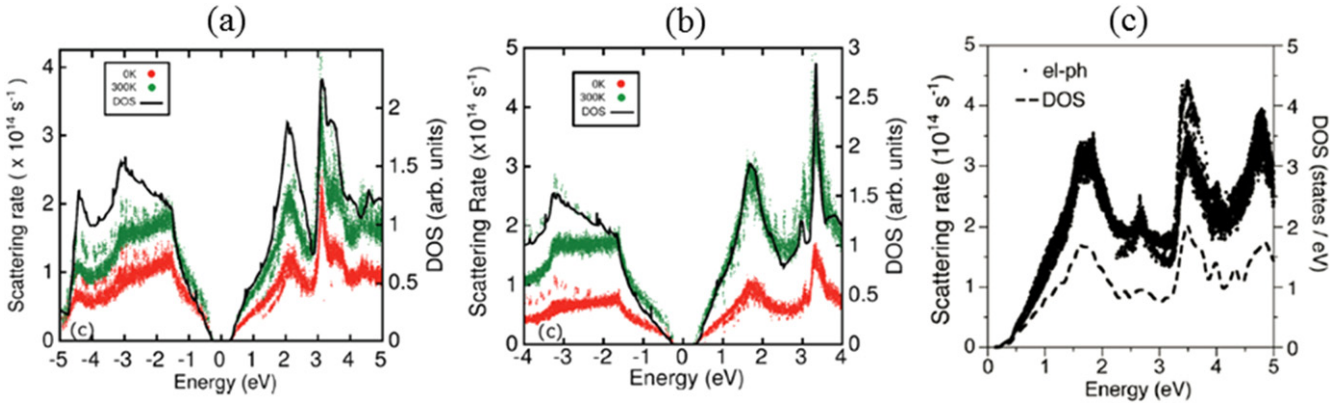

Figure 8. Electron-phonon scattering rates compared with the electron density of state for (a) Si, (b) Ge and (c) GaAs. It is shown that the scattering rate profile in a larger energy scale closely follows the density of state profile. The wave-vector dependence of the scattering rates can be seen by the scatter of the data at the same energy level. The above calculations assume pure crystals and therefore the Fermi level lies inside the band gap. (Si and Ge results reprinted with permission from Tandon *et al* [120]. Copyright (2015) American Institute of Physics. GaAs results reprinted with permission from Bernardi *et al* [122].)

the scattering processes between many electron and phonon states. As a result, scattering rates between certain pockets are not sufficient in describing the transport property and the first-principles calculation of mobility has always been challenging due to the large mesh entailed. Restrepo *et al* first reported the mobility calculation for silicon including both electron-phonon and electron-impurity scattering [36]. The phonon-limited mobility (neglect the impurity scattering) for n-type Si is calculated to be 1970 $cm^2\,V^{-1}\,s^{-1}$ at low carrier concentration, which agrees reasonably well with the experimental mobility (~1700 $cm^2\,V^{-1}\,s^{-1}$ [123]) of lightly-doped silicon but slightly larger. The mesh density was not reported in their work but a more accurate calculation will inevitably require a dense mesh. A linear interpolation scheme [38] that obtains the electron-phonon matrix elements on a fine mesh has been shown to give rise to a mobility value (~1860 $cm^2\,V^{-1}\,s^{-1}$) closer to the experimental result. The linear interpolation works well in this case essentially because the electron-phonon coupling matrices do not vary much in the FBZ for Si. As more complicated materials are considered, however, we consider the Wannier interpolation [43] as a more accurate approach for extracting the electron-phonon coupling matrix.

4.2. Seebeck coefficient

On the other hand, the Seebeck coefficient requires a more careful look. This is because in lightly-doped silicon the phonon drag effect has been found to be significantly strong at low temperatures [39, 124]. Therefore when considering the Seebeck coefficient we should add up both the diffusive contribution (equation (25)) and the phonon drag contribution (equation (30)). We note that in the expression for the phonon drag Seebeck coefficient both relaxation times for electrons and phonons are required, thereby making the calculation even more challenging than the mobility calculation. The calculation using first-principles-extracted deformation potential was able to describe the diffusive Seebeck coefficient [113], which is obtained by removing an empirical phonon drag contribution from the measured data. Recently Mahan *et al* combined first-principles phonon relaxation times with the deformation potential model to describe both the diffusive and phonon drag contribution to the Seebeck coefficient; good agreement with experiments across a wide temperature range was seen [125]. However, it was found that the phonon drag from the simulation increases faster with decreased temperature than the experimental value, which could result from the use of the deformation potential model. In a later study using a full first-principles calculation with both electron and phonon relaxation times extracted from DFT calculations, the Seebeck coefficient including the phonon drag was predicted with excellent agreements compared to the experiment [126], as shown in figure 11(a). This emphasizes the use of a first-principles approach for a better description of coupled electron-phonon transport. These calculations [125, 126] quantitatively investigated the temperature dependence of the phonon drag effect: although it dominates at low temperatures, it has influences even beyond the room temperature. It was shown that at room temperature in lightly-doped n-type (p-type) silicon the phonon drag contributes to 30% (40%) of the total Seebeck coefficient [126]. As we have mentioned, the phonon drag effect can be understood as the momentum transfer from non-equilibrium phonons with long MFPs to the electron system. In such processes, the phonons are the key players and one important aspect to observe will be how much are the contributions from different phonons. Figure 11(b) shows the accumulated contribution to the phonon drag with respect to the phonon MFP. It can be seen that the phonons that contribute to the phonon drag mostly have long MFPs (small wave vectors) compared to those thermal phonons that carry heat. This is essentially because of the momentum and energy conservation for the electron-phonon scattering process, which restricts the wave vector of phonons to be small so that they can scatter band edge electrons (normal process). Clearly, we see from figure 11(b) that there is a spectral difference between the contributions to the thermal conductivity and to

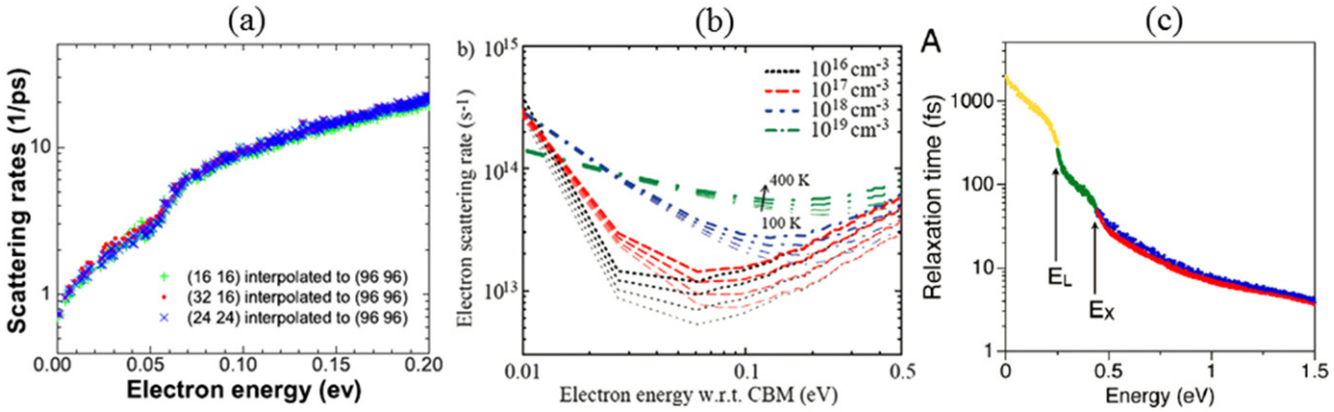

Figure 9. Electron scattering rates for (a) pure Si and (b) doped Si, as well as (c) electron relaxation times for pure GaAs. These plots consider electrons that are close to the conduction band edge. The results for pure Si and pure GaAs are evaluated at room temperature, while those for doped Si are shown from 100 K to 400 K. (Pure Si results reprinted with permission from Li *et al* [38]. Copyright (2015) American Physical Society. Doped Si results reprinted with permission from Qiu *et al* [37]. GaAs results reprinted with permission from Bernardi *et al* [122].)

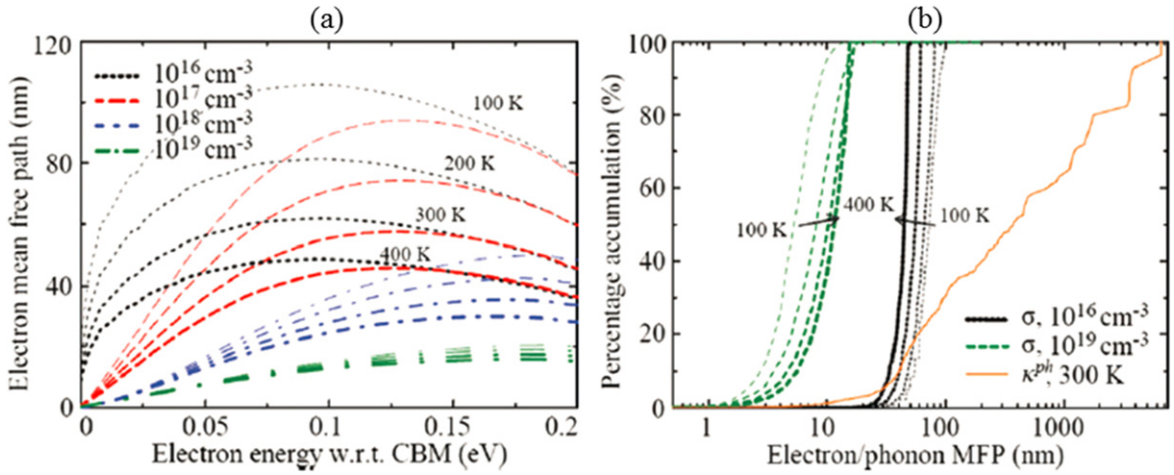

Figure 10. (a) Electron MFP as a function of the electron energy and (b) accumulated contribution to the electrical conductivity with respect to the electron MFP compared to the accumulated thermal conductivity contribution (with respect to phonon MFP) in silicon from 100 K to 400 K. In part (b), two different doping levels (lightly-doped and heavily-doped) are shown. (Reprinted figures with permission from Qiu *et al* [37].)

the phonon drag from phonons. Such quantitative information can be used to engineer the material's property by modifying the phonon spectrum. For example, one can enhance the thermoelectric performance by filtering out short MFP phonons, in which case the phonon drag will be retained but the thermal conductivity can be reduced, a strategy recently proposed in [126].

Furthermore, by considering the phonon scattering by electrons in the phonon-relaxation-time calculation, the reduction of the phonon drag effect at higher carrier concentrations, known as the saturation effect [39], can also be captured by the fully first-principles approach [126], with a good agreement with the experiment shown in figure 12. This reduction comes from the increased scatterings of phonons with long wavelengths by electrons. As shown above, phonons that contribute to phonon drag mostly have small wave vectors and thus low frequencies. For these low-frequency phonons, the electron scattering becomes appreciable as we have discussed in section 3. The impurity scattering is in fact insufficient to strongly affect these phonons despite of their effect on the mobility, because it scales as ω^4 and decreases quickly as the phonon frequency becomes small. It was further uncovered that [126], though the phonon drag is largely reduced due to the saturation effect, it is not negligible for a heavily-doped silicon. In fact its contribution is comparable to the diffusive Seebeck coefficient at 10^{19} cm^{-3} even at room temperature (figure 12). This result challenges the previous belief that the phonon drag effect vanishes in heavily-doped samples, and provides new perspectives into the electron-phonon coupled transport of the decades-old material—silicon.

We have seen that the use of the first-principles method has provided us with insightful details into the electrical transport properties in silicon. However, up to now there are only few materials whose mobility or Seebeck coefficient has been calculated using a parameter-free first-principles

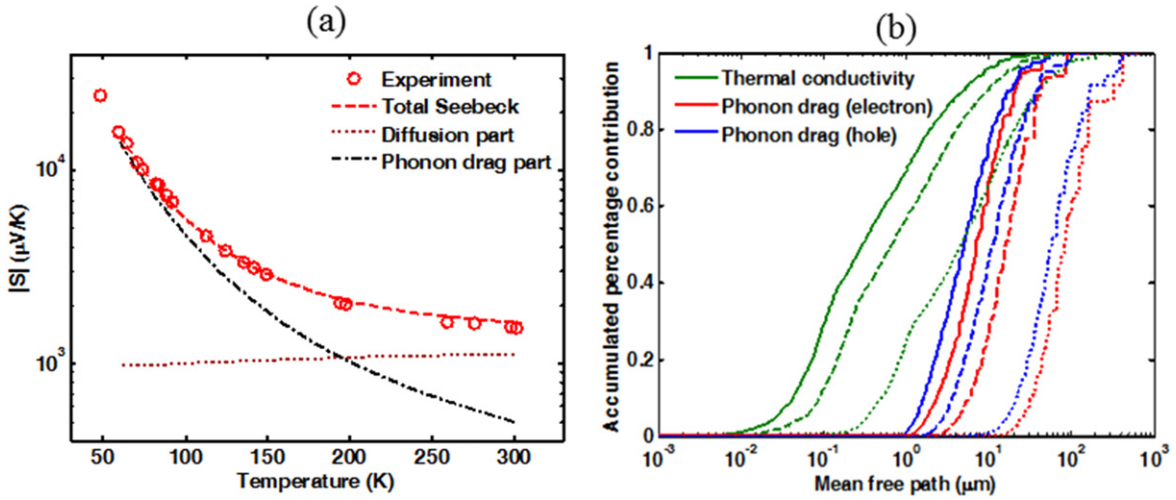

Figure 11. (a) First-principles calculation of the temperature-dependent Seebeck coefficient in pure Si and (b) the accumulated contribution to the phonon drag. In part (a) the total Seebeck coefficient is also decomposed into the diffusion part and the phonon drag part, the latter of which dramatically increases as the temperature decreases. The results agree well with the experimental data [124]. In part (b) the accumulated curves are shown for three temperatures (solid lines for 300 K, dashed lines for 200 K and dotted lines for 100 K). The first-principles results are from [126].

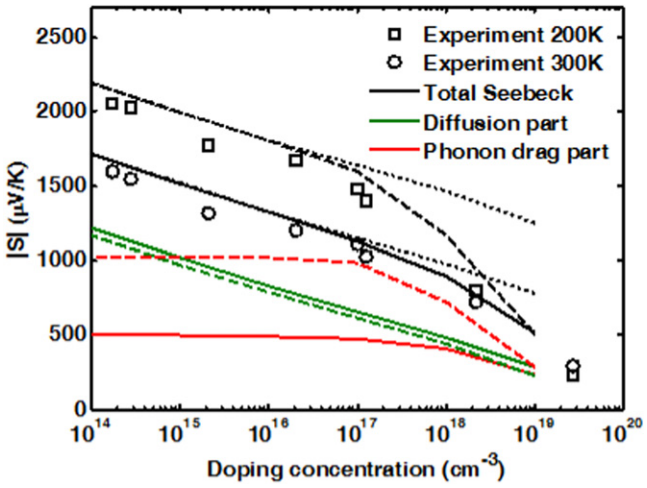

Figure 12. Carrier-concentration-dependent Seebeck coefficient from first principles in doped n-type Si. The results are shown at two different temperatures (solid lines for 300 K and dashed lines for 200 K). The total Seebeck coefficient (black curves) is also decomposed into the diffusion part (green curves) and the phonon drag part (red curves). The phonon drag at low carrier concentration does not depend on the doping level. The dotted lines assume that this value is used throughout the whole doping range. Clearly for high doping concentrations the dotted line overestimates the Seebeck coefficient, and the difference is due to the scattering of phonons by electrons, leading to a reduction of the phonon drag part. The first-principles results are from [126].

approach, with examples including silicon, MoS_2 and phosphorene [36–38, 126, 127]. Here we deem the use of the deformation potential model (even with the deformation potential extracted from the first principles) as an complementary but not a predictive method, because it neglects the electron-state-dependence of the electron-phonon coupling, which is hardly justified for general materials, especially considering the fact that the material search has been going towards more complex ones. Large discrepancy has also been seen between different studies using similar deformation potentials extracted from DFT calculations for the same material. We however believe that a detailed study into the validity of the deformation potential by comparing it with first-principles results will help to elucidate its applicability in the wide range of many new and unexplored materials.

One difficulty associated with the first-principles electrical property calculation of other materials is that the efficient interpolation of the electron-phonon matrix elements from a coarse onto a dense mesh, which is necessary for the integration to obtain the relaxation times, requires the force constant (or the perturbed potential) to be short-ranged in real space [43]. However in materials with more than two elements, accompanied with the LO phonon near the zone center there will be a polarization field, which induces an electric field that can cause scatterings for electrons. This scattering due to the long-range electric field induced by a polar LO phonon is usually called polar optical phonon scattering. Therefore, the interpolation scheme, as we have discussed before, will be less accurate if the transport property of the material is governed by such scatterings. For a finite phonon wave vector, DFPT exactly solves the eigen-equation and therefore automatically includes such a long-range effect. This merit of DFPT is also used in the study of hot electron relaxations in GaAs [122]. The interpolation method for the electron-phonon coupling matrix, however, needs to be modified to correctly capture the long-range effect. Work has been done to use the first-principles band structure combined with scattering time models to evaluate the electrical property of other materials. For example the mobility of $SrTiO_3$ was modeled using the Fröhlich model for the interaction matrix [128]. Such an approach however does not differentiate the importance of different scattering mechanisms in affecting the transport property. Recently Sjakste *et al* and Verdi *et al* incorporated the polar optical phonon scattering into the

Wannier interpolation scheme by using a rigid ion model to treat the long-range effect, where each atom carries a certain effective charge due to their ionic character and the polar scattering mainly comes from the Coulomb interaction between them and the electrons [51, 52]. The interpolated electron-phonon matrix elements agree well with those obtained from direct DFPT calculations [51, 52]. We believe the inclusion of the polar scattering will facilitate the study of the electrical transport of more complicated semiconductors and provide more possibilities for engineering materials' electrical property.

4.3. Effect of alloying

Using a perturbative approach with information extracted from DFT calculations, Murphy-Armando and Fahy examined the alloying scattering for electrons in a SiGe alloy [62]. They quantified the contribution to the electron scattering rates from intervalley scattering and intravalley scatterings, and found that they are comparable with each other, contrary to some previous modeling results which claim that intervalley scattering is negligible. They have also obtained good agreements with the experimental value by incorporating the deformational potential model for describing the electron-phonon scattering. Particularly, the first-principles results correctly captured the abrupt change of the mobility in the $Si_{1-x}Ge_x$ alloy at $x \sim 0.85$, which is due to the $\Delta - L$ band crossing [62]. We note that the perturbed potential ΔV in this case is the potential difference between the pristine Si and pristine Ge, while the electron wavefunctions are obtained in the virtual crystal with averaged properties. Care must be taken to set the reference for the potential energy [62].

The same perturbative approach has been applied to compare the effect of different substitutes on the electron scattering within the same column [67]. As we have mentioned in section 2, the supercell method will inevitably introduce spurious interactions due to the image atoms and background compensating charges, therefore questioning a direct evaluation of the perturbed potential. However, in the same column the element has the same number of valence electrons and the errors thus introduced will be the same. The difference mainly comes from the scattering due to strain-induced disorder. We note that to compare elements in different columns, an accurate description of the perturbed potential will be necessary. One way to achieve this has been illustrated by Restrepo *et al* in their calculation of the electron-impurity scattering rate [36]. Other techniques borrowed from the treatment of calculating formation energy [69] can also be possibly used to develop an efficient method for describing electron scattering by charged impurities.

Besides the alloying scattering for electrons, the alloy structure also modifies the phonon modes and therefore the coupling between electrons and phonons. In the lowest-order approximation, phonon eigenmodes as well as the perturbed potential can be obtained in the virtual crystal. The electron-phonon scattering rates are then calculated based on this virtual crystal perturbation combined with the electron modes of the virtual crystal, assuming the electronic disorder is weak [129]. However the disorder effect on the electron-phonon coupling has been ignored in such approach. We would expect that a larger supercell calculation will help us to understand how the electron-phonon coupling is modified by the alloy structures, with detailed studies into the electron relaxation times and MFP distribution, providing crucial information for evaluating the electron size effect at nanoscale.

5. Summary

We have reviewed the recent development of the first-principles approach to obtain the transport properties (in particular, the electrical conductivity, the Seebeck coefficient and the thermal conductivity) of materials, which provides us with a better understanding of transport features and leads to rational designs of material properties. The coupled electron-phonon BTE has been introduced, discussed and used for relating the transport property to the eigenmodes (electrons, phonons) and their coupling with each other. The relaxation time is introduced as a measure of the decay of non-equilibrium distribution back into an equilibrium state, and is the key variable in calculating the transport property under the so-called relaxation time approximation. Relaxation times derive from the scatterings between the states, which are governed by the coupling due to perturbations introduced to the original system. For phonon-phonon scattering, the perturbations come from the anharmonic interatomic forces. For electron-phonon scattering, the atomic displacements perturb the environments seen by the electrons. Besides, impurities naturally introduce perturbations into the original pristine crystal. All of these can be analyzed using the same framework and we have seen that the crucial ingredient in describing these scattering processes and thus the relaxation times is the coupling matrix that determines the transition probabilities from one state to another.

Methods to extract the electron-phonon coupling matrix based on first-principles calculations are briefly discussed. For the thermal conductivity calculation, the important information is the anharmonic force constant while for the electrical transport properties the electron-phonon coupling matrix is the vital component. An efficient extraction of such information is needed, due to the requirement for the convergence of the relaxation time calculation involving the states across the Brillouin zone.

We have seen that the calculation of thermal transport has been greatly improved in recent years, from simple semiconductors to more complex systems. Some practical systems like alloys can now be calculated with good agreements with experiments. Further research development could involve the calculations of even more complex materials like perovskites and organic semiconductors. These complex materials are difficult for a first-principles calculation mainly because the computational time scales approximately as N^3 (N is the total number of atoms in the unit cell) due to the non-locality inherent in quantum mechanics [4]. An order-N method (computational time scales as N) that uses certain features of

the system to reduce the computational load has great advantages over the traditional method [4], but has not been introduced into the first-principles calculation of thermal conductivity. Another possible strategy for accelerating the calculation is to use effective force constants to represent some atomic clusters in a large unit cell. For example, the inorganic-organic perovskite semiconductor $CH_3NH_3PbI_3$ which has recently gained popularity due to its potential in solar cells [130] has an ABO_3 structure with site-A replaced by an organic group. If the organic group can be represented by a meta-atom with effective force constants (between this meta-atom and other atoms), the total number of atoms can be greatly reduced. This strategy, if possible, can facilitate the first-principles thermal transport study of many inorganic-organic hybrid materials.

In comparison with the first-principles calculation for thermal transport, calculations of electrical transport in semiconductors have only received attention recently and up to now only a few materials (silicon, phosphorene, MoS_2, etc) are calculated within a fully first-principles approach for the electrical transport properties [36–38, 126, 127]. Other methods usually use parameterized models such as the deformation potential model, fitted from experiments or first-principles calculations. The first-principles method can provide insightful details into the relaxation time profile and MFP distribution of the electrons, which are of great significance for modeling and engineering nanoscale devices. Besides, unprecedented details in the transport property have been seen by using the first-principles approach. For example, the phonon drag effect was seen to play a non-negligible role in the Seebeck coefficient even in heavily-doped samples at room temperature. Along this path, we think some of the challenges include the consideration of the spin–orbit coupling for heavy elements and alloy (and doping) effects on the electrical transport. Since the band structure greatly affects the scattering rate profile as we have seen in GaAs [122], a more advanced first-principles method such as GW calculation (a method to more accurately describe the electron-electron interaction with G standing for Green's function and W standing for the screened Coulomb interaction) to take into account the electron many-body effects on the band structure will be necessary for a better description of the electronic properties. In addition, one can develop the theory to recast the transport property calculation based on BTEs using the matrix elements between different eigenstates into one that only uses the matrix elements in the Wannier basis; the latter of which requires less computational work but contains all the essential information, a strategy already mentioned in [43]. The Wannier basis (for electron and phonons) may also serve as good starting points for the study of transport problems in disordered materials (such as polymer or amorphous structures), where the periodic boundary condition breaks down and the dynamics of the system is more often described using spatially-localized modes. Another challenging question is how to concurrently solve the coupled electron-phonon BTEs, putting the effect of the electron on phonons and that of the phonon on electrons on an equal footing. For this, one has to use an iterative solver to deal with the BTEs, which goes beyond the relaxation time model. This further step will provide us with more information for the coupled electron-phonon transport. It is also possible that unique physical behaviors may emerge from such a coupled transport problem. In general, the advancement of the first-principles computational technique, as we believe, extends our toolbox for studying the transport properties in a quantitative way and will finally open up the venue towards a high-throughput material search based on *ab initio* predictions.

Acknowledgments

We thank Mingda Li, Qichen Song and Zhiwei Ding for the helpful discussions. This material is based upon work supported as part of the Solid State Solar-Thermal Energy Conversion Center (S^3TEC), an Energy Frontier Research Center funded by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under Award Number DE-SC0001299/DE-FG02-09ER46577.

References

[1] Jones R O 2015 Density functional theory: its origins, rise to prominence, and future *Rev. Mod. Phys.* **87** 897–923
[2] Hohenberg P and Kohn W 1964 Inhomogeneous electron gas *Phys. Rev.* **136** B864–71
[3] Kohn W and Sham L J 1965 Self-consistent equations including exchange and correlation effects *Phys. Rev.* **140** A1133–8
[4] Martin R M 2004 *Electronic Structure: Basic Theory and Practical Methods* (Cambridge University Press)
[5] Singh D J and Nordstrom L 2006 *Planewaves, Pseudopotentials and the LAPW Method* 2nd edn (Springer: Berlin)
[6] Carter E A 2008 Challenges in modeling materials properties without experimental input *Science* **321** 800–3
[7] Curtarolo S, Hart G L W, Nardelli M B, Mingo N, Sanvito S and Levy O 2013 The high-throughput highway to computational materials design *Nat. Mater.* **12** 191–201
[8] Ashcroft/Mermin 1976 *Solid state physics* (Cengage Learning)
[9] McGaughey A J H and Larkin J M 2014 Predicting phonon properties from equilibrium molecular dynamics simulations *Annu. Rev. Heat Transf.* **17** 49–87
[10] Shiomi J 2014 Nonequilirium molecular dynamics methods for lattice heat conduction calculations *Annu. Rev. Heat Transf.* **17** 177–203
[11] Ziman J M 1960 *Electrons and Phonons: The Theory of Transport Phenomena in Solids* (Clarendon Press)
[12] Hess K 1988 *Advanced theory of semiconductor devices* (Prentice-Hall)
[13] Lundstrom M 2009 *Fundamentals of Carrier Transport* (Cambridge University Press)
[14] Bernardi M, Vigil-Fowler D, Lischner J, Neaton J B and Louie S G 2014 *Ab Initio* study of hot carriers in the first picosecond after sunlight absorption in silicon *Phys. Rev. Lett.* **112** 257402
[15] Chen G 2005 *Nanoscale Energy Transport and Conversion: A Parallel Treatment of Electrons, Molecules, Phonons and Photons* (Oxford University Press)

[16] Esfarjani K, Chen G and Stokes H 2011 Heat transport in silicon from first-principles calculations *Phys. Rev.* B **84** 085204

[17] Broido D A, Malorny M, Birner G, Mingo N and Stewart D A 2007 Intrinsic lattice thermal conductivity of semiconductors from first principles *Appl. Phys. Lett.* **91** 231922

[18] Peierls R 1929 Zur kinetischen Theorie der Wärmeleitung in Kristallen *Ann. Phys.* **395** 1055–101

[19] Baroni S, de Gironcoli S, Dal Corso A and Giannozzi P 2001 Phonons and related crystal properties from density-functional perturbation theory *Rev. Mod. Phys.* **73** 515–62

[20] Esfarjani K and Stokes H T 2008 Method to extract anharmonic force constants from first principles calculations *Phys. Rev.* B **77** 144112

[21] Zhou F, Nielson W, Xia Y and Ozoliņš V 2014 Lattice anharmonicity and thermal conductivity from compressive sensing of first-principles calculations *Phys. Rev. Lett.* **113** 185501

[22] Ai X, Chen Y and Marianetti C A 2014 Slave mode expansion for obtaining *ab initio* interatomic potentials *Phys. Rev.* B **90** 014308

[23] Chen Y, Ai X and Marianetti C A 2014 First-principles approach to nonlinear lattice dynamics: anomalous spectra in PbTe *Phys. Rev. Lett.* **113** 105501

[24] Gonze X and Vigneron J-P 1989 Density-functional approach to nonlinear-response coefficients of solids *Phys. Rev.* B **39** 13120–8

[25] Deinzer G, Birner G and Strauch D 2003 *Ab initio* calculation of the linewidth of various phonon modes in germanium and silicon *Phys. Rev.* B **67** 144304

[26] Ward A and Broido D A 2010 Intrinsic phonon relaxation times from first-principles studies of the thermal conductivities of Si and Ge *Phys. Rev.* B **81** 085205

[27] Ward A, Broido D A, Stewart D A and Deinzer G 2009 *Ab initio* theory of the lattice thermal conductivity in diamond *Phys. Rev.* B **80** 125203

[28] Paulatto L, Mauri F and Lazzeri M 2013 Anharmonic properties from a generalized third-order *ab initio* approach: Theory and applications to graphite and graphene *Phys. Rev.* B **87** 214303

[29] Fugallo G, Cepellotti A, Paulatto L, Lazzeri M, Marzari N and Mauri F 2014 Thermal conductivity of graphene and graphite: collective excitations and mean free paths *Nano Lett.* **14** 6109–14

[30] Cepellotti A, Fugallo G, Paulatto L, Lazzeri M, Mauri F and Marzari N 2015 Phonon hydrodynamics in two-dimensional materials *Nat. Commun.* **6**

[31] Bloch F 1928 *Z. Für. Phys.* **52** 555

[32] Ziman J M 1955 The electron-phonon interaction, according to the adiabatic approximation *Math. Proc. Camb. Philos. Soc.* **51** 707–12

[33] Bailyn M 1958 Transport in metals: effect of the nonequilibrium phonons *Phys. Rev.* **112** 1587–98

[34] Ziman J M 1956 XVII. The effect of free electrons on lattice conduction *Philos. Mag.* **1** 191–8

[35] Grimvall G 1981 *The Electron-phonon Interaction in Metals* North-Holland

[36] Restrepo O D, Varga K and Pantelides S T 2009 First-principles calculations of electron mobilities in silicon: phonon and coulomb scattering *Appl. Phys. Lett.* **94** 212103

[37] Qiu B, Tian Z, Vallabhaneni A, Liao B, Mendoza J M, Restrepo O D, Ruan X and Chen G 2015 First-principles simulation of electron mean-free-path spectra and thermoelectric properties in silicon *EPL Europhys. Lett.* **109** 57006

[38] Li W 2015 Electrical transport limited by electron-phonon coupling from Boltzmann transport equation: an *ab initio* study of Si, Al, and MoS_2 *Phys. Rev.* B **92** 075405

[39] Herring C 1954 Theory of the thermoelectric power of semiconductors *Phys. Rev.* **96** 1163–87

[40] Geballe T H and Hull G W 1954 Seebeck effect in germanium *Phys. Rev.* **94** 1134–40

[41] Onsager L 1931 Reciprocal relations in irreversible processes: I. *Phys. Rev.* **37** 405–26

[42] Onsager L 1931 Reciprocal relations in irreversible processes: II. *Phys. Rev.* **38** 2265–79

[43] Giustino F, Cohen M and Louie S 2007 Electron-phonon interaction using Wannier functions *Phys. Rev.* B **76** 165108

[44] Noffsinger J, Giustino F, Malone B D, Park C-H, Louie S G and Cohen M L 2010 EPW: a program for calculating the electron–phonon coupling using maximally localized Wannier functions *Comput. Phys. Commun.* **181** 2140–8

[45] Marzari N, Mostofi A A, Yates J R, Souza I and Vanderbilt D 2012 Maximally localized Wannier functions: theory and applications *Rev. Mod. Phys.* **84** 1419–75

[46] Giustino F, Yates J R, Souza I, Cohen M L and Louie S G 2007 Electron-phonon interaction via electronic and lattice wannier functions: superconductivity in boron-doped diamond reexamined *Phys. Rev. Lett.* **98** 047005

[47] Giustino F, Cohen M L and Louie S G 2008 Small phonon contribution to the photoemission kink in the copper oxide superconductors *Nature* **452** 975–8

[48] Noffsinger J, Giustino F, Louie S G and Cohen M L 2009 Origin of superconductivity in boron-doped silicon carbide from first principles *Phys. Rev.* B **79** 104511

[49] Gonze X and Lee C 1997 Dynamical matrices, Born effective charges, dielectric permittivity tensors, and interatomic force constants from density-functional perturbation theory *Phys. Rev.* B **55** 10355–68

[50] Vogl P 1976 Microscopic theory of electron-phonon interaction in insulators or semiconductors *Phys. Rev.* B **13** 694–704

[51] Sjakste J, Vast N, Calandra M and Mauri F 2015 Wannier interpolation of the electron-phonon matrix elements in polar semiconductors: polar-optical coupling in GaAs *Phys. Rev.* B **92** 054307

[52] Verdi C and Giustino F 2015 Fr"ohlich electron-phonon vertex from first principles *Phys. Rev. Lett.* **115** 176401

[53] Pei Y, Shi X, LaLonde A, Wang H, Chen L and Snyder G J 2011 Convergence of electronic bands for high performance bulk thermoelectrics *Nature* **473** 66–9

[54] Poudel B *et al* 2008 High-thermoelectric performance of nanostructured bismuth antimony telluride bulk alloys *Science* **320** 634–8

[55] Abeles B 1963 Lattice thermal conductivity of disordered semiconductor alloys at high temperatures *Phys. Rev.* **131** 1906–11

[56] Elliott R J, Krumhansl J A and Leath P L 1974 The theory and properties of randomly disordered crystals and related physical systems *Rev. Mod. Phys.* **46** 465–543

[57] Mendoza J, Esfarjani K and Chen G 2015 An *ab initio* study of multiple phonon scattering resonances in silicon germanium alloys *J. Appl. Phys.* **117** 174301

[58] Kundu A, Mingo N, Broido D A and Stewart D A 2011 Role of light and heavy embedded nanoparticles on the thermal conductivity of SiGe alloys *Phys. Rev.* B **84** 125426

[59] Klemens P G 1955 The Scattering of low-frequency lattice waves by static imperfections *Proc. Phys. Soc. Sect.* A **68** 1113

[60] Tamura S 1983 Isotope scattering of dispersive phonons in Ge *Phys. Rev.* B **27** 858–66

[61] Walle de C G V and Neugebauer J 2004 First-principles calculations for defects and impurities: applications to III-nitrides *J. Appl. Phys.* **95** 3851–79

[62] Murphy-Armando F and Fahy S 2006 First-principles calculation of alloy scattering in Ge_xSi_{1-x} *Phys. Rev. Lett.* **97** 096606

[63] Garg J, Bonini N, Kozinsky B and Marzari N 2011 Role of disorder and anharmonicity in the thermal conductivity of silicon-germanium alloys: a first-principles study *Phys. Rev. Lett.* **106** 045901

[64] Li W, Lindsay L, Broido D A, Stewart D A and Mingo N 2012 Thermal conductivity of bulk and nanowire $Mg_2Si_xSn_{1-x}$ alloys from first principles *Phys. Rev.* B **86** 174307

[65] Tian Z, Garg J, Esfarjani K, Shiga T, Shiomi J and Chen G 2012 Phonon conduction in PbSe, PbTe, and $PbTe_{1-x}Se_x$ from first-principles calculations *Phys. Rev.* B **85** 184303

[66] Lee S, Esfarjani K, Mendoza J, Dresselhaus M S and Chen G 2014 Lattice thermal conductivity of Bi, Sb, and Bi-Sb alloy from first principles *Phys. Rev.* B **89** 085206

[67] Lordi V, Erhart P and Åberg D 2010 Charge carrier scattering by defects in semiconductors *Phys. Rev.* B **81** 235204

[68] Rurali R, Markussen T, Suñé J, Brandbyge M and Jauho A-P 2008 Modeling transport in ultrathin Si nanowires: charged versus neutral impurities *Nano Lett.* **8** 2825–8

[69] Freysoldt C, Grabowski B, Hickel T, Neugebauer J, Kresse G, Janotti A and Van de Walle C G 2014 First-principles calculations for point defects in solids *Rev. Mod. Phys.* **86** 253–305

[70] Tian Z, Lee S and Chen G 2014 Comprehensive review of heat transfer in thermoelectric materials and devices *Annu. Rev. Heat Transf.* **17** 425–83

[71] Esfarjani K, Garg J and Chen G 2014 Modeling heat conduction from first principles *Annu. Rev. Heat Transf.* **17** 9–47

[72] Broido D A, Ward A and Mingo N 2005 Lattice thermal conductivity of silicon from empirical interatomic potentials *Phys. Rev.* B **72** 014308

[73] Lindsay L, Broido D A and Reinecke T L 2012 Thermal conductivity and large isotope effect in GaN from first principles *Phys. Rev. Lett.* **109** 095901

[74] Lindsay L, Broido D A and Reinecke T L 2013 *Ab initio* thermal transport in compound semiconductors *Phys. Rev.* B **87** 165201

[75] Luo T, Garg J, Shiomi J, Esfarjani K and Chen G 2013 Gallium arsenide thermal conductivity and optical phonon relaxation times from first-principles calculations *EPL Europhys. Lett.* **101** 16001

[76] Lindsay L, Broido D A and Reinecke T L 2013 Phonon-isotope scattering and thermal conductivity in materials with a large isotope effect: a first-principles study *Phys. Rev.* B **88** 144306

[77] Shiomi J, Esfarjani K and Chen G 2011 Thermal conductivity of half-Heusler compounds from first-principles calculations *Phys. Rev.* B **84** 104302

[78] Shiga T, Shiomi J, Ma J, Delaire O, Radzynski T, Lusakowski A, Esfarjani K and Chen G 2012 Microscopic mechanism of low thermal conductivity in lead telluride *Phys. Rev.* B **85** 155203

[79] Carrete J, Li W, Mingo N, Wang S and Curtarolo S 2014 Finding unprecedentedly low-thermal-conductivity Half-Heusler semiconductors via high-throughput materials modeling *Phys. Rev.* X **4** 011019

[80] Hellman O and Broido D A 2014 Phonon thermal transport in Bi_2Te_3 from first principles *Phys. Rev.* B **90** 134309

[81] Lee S, Esfarjani K, Luo T, Zhou J, Tian Z and Chen G 2014 Resonant bonding leads to low lattice thermal conductivity *Nat. Commun.* **5**

[82] Li W and Mingo N 2014 Lattice dynamics and thermal conductivity of skutterudites $CoSb_3$ and $IrSb_3$ from first principles: why $IrSb_3$ is a better thermal conductor than $CoSb_3$ *Phys. Rev.* B **90** 094302

[83] Li W and Mingo N 2014 Thermal conductivity of fully filled skutterudites: role of the filler *Phys. Rev.* B **89** 184304

[84] Li W and Mingo N 2015 Ultralow lattice thermal conductivity of the fully filled skutterudite $YbFe_4Sb_{12}$ due to the flat avoided-crossing filler modes *Phys. Rev.* B **91** 144304

[85] Tadano T, Gohda Y and Tsuneyuki S 2015 Impact of Rattlers on thermal conductivity of a thermoelectric clathrate: a first-principles study *Phys. Rev. Lett.* **114** 095501

[86] Bonini N, Garg J and Marzari N 2012 Acoustic phonon lifetimes and thermal transport in free-standing and strained graphene *Nano Lett.* **12** 2673–8

[87] Lindsay L, Li W, Carrete J, Mingo N, Broido D A and Reinecke T L 2014 Phonon thermal transport in strained and unstrained graphene from first principles *Phys. Rev.* B **89** 155426

[88] Lee S, Broido D, Esfarjani K and Chen G 2015 Hydrodynamic phonon transport in suspended graphene *Nat. Commun.* **6**

[89] Li W, Carrete J and Mingo N 2013 Thermal conductivity and phonon linewidths of monolayer MoS_2 from first principles *Appl. Phys. Lett.* **103** 253103

[90] Gu X and Yang R 2014 Phonon transport in single-layer transition metal dichalcogenides: a first-principles study *Appl. Phys. Lett.* **105** 131903

[91] Gu X and Yang R 2015 First-principles prediction of phononic thermal conductivity of silicene: a comparison with graphene *J. Appl. Phys.* **117** 025102

[92] Qin G, Yan Q-B, Qin Z, Yue S-Y, Hu M and Su G 2015 Anisotropic intrinsic lattice thermal conductivity of phosphorene from first principles *Phys. Chem. Chem. Phys.* **17** 4854–8

[93] Jain A and McGaughey A J H 2015 Strongly anisotropic in-plane thermal transport in single-layer black phosphorene *Sci. Rep.* **5** 8501

[94] Lindsay L, Broido D A and Mingo N 2010 Flexural phonons and thermal transport in graphene *Phys. Rev.* B **82** 115427

[95] Debernardi A 1998 Phonon linewidth in III-V semiconductors from density-functional perturbation theory *Phys. Rev.* B **57** 12847–58

[96] Bao H, Qiu B, Zhang Y and Ruan X 2012 A first-principles molecular dynamics approach for predicting optical phonon lifetimes and far-infrared reflectance of polar materials *J. Quant. Spectrosc. Radiat. Transf.* **113** 1683–8

[97] von der Linde D, Kuhl J and Klingenberg H 1980 Raman scattering from nonequilibrium LO phonons with picosecond resolution *Phys. Rev. Lett.* **44** 1505–8

[98] Kash J A and Tsang J C 1988 *Secondary Emission Studies of Hot Carrier Relaxation in Polar Semiconductors* **31** 419–24

[99] Luckyanova M N *et al* 2012 Coherent phonon heat conduction in superlattices *Science* **338** 936–9

[100] Dames C and Chen G 2005 Thermoelectrics handbook: macro to nano

[101] Chen G 1998 Thermal conductivity and ballistic-phonon transport in the cross-plane direction of superlattices *Phys. Rev.* B **57** 14958–73

[102] Henry A S and Chen G 2008 Spectral phonon transport properties of silicon based on molecular dynamics simulations and lattice dynamics *J. Comput. Theor. Nanosci.* **5** 141–52

[103] Minnich A J, Johnson J A, Schmidt A J, Esfarjani K, Dresselhaus M S, Nelson K A and Chen G 2011 Thermal Conductivity Spectroscopy Technique to Measure Phonon Mean Free Paths *Phys. Rev. Lett.* **107** 095901

[104] Johnson J A, Maznev A A, Eliason J K, Minnich A, Collins K, Chen G, Cuffe J, Kehoe T, Sotomayor Torres C M and Nelson K A 2011 Experimental evidence of non-diffusive thermal transport in Si and GaAs *Symp. BB—Nanoscale Heat Transport—From*

Fundamentals to Devices MRS Online Proc. Library vol 1347
[105] Minnich A J 2012 Determining phonon mean free paths from observations of quasiballistic thermal transport *Phys. Rev. Lett.* **109** 205901
[106] Johnson J A, Maznev A A, Cuffe J, Eliason J K, Minnich A J, Kehoe T, Torres C M S, Chen G and Nelson K A 2013 Direct measurement of room-temperature nondiffusive thermal transport over micron distances in a silicon membrane *Phys. Rev. Lett.* **110** 025901
[107] Cuffe J *et al* 2015 Reconstructing phonon mean-free-path contributions to thermal conductivity using nanoscale membranes *Phys. Rev.* B **91** 245423
[108] Hu Y, Zeng L, Minnich A J, Dresselhaus M S and Chen G 2015 Spectral mapping of thermal conductivity through nanoscale ballistic transport *Nat. Nanotechnol.* **10** 701–6
[109] Sommerfeld A and Bethe H 1933 *Handb. Phys.* (Berlin: Springer)
[110] Klemens P 1954 The electrical and thermal conductivities of monovalent metals *Aust. J. Phys.* **7** 70–6
[111] Liao B, Qiu B, Zhou J, Huberman S, Esfarjani K and Chen G 2015 Significant reduction of lattice thermal conductivity by the electron-phonon interaction in silicon with high carrier concentrations: a first-principles study *Phys. Rev. Lett.* **114** 115901
[112] Sjakste J, Timrov I, Gava P, Mingo N and Vast N 2014 First-principles calculations of electron-phonon scattering *Annu. Rev. Heat Transf.* **17** 333–83
[113] Wang Z, Wang S, Obukhov S, Vast N, Sjakste J, Tyuterev V and Mingo N 2011 Thermoelectric transport properties of silicon: toward an *ab initio* approach *Phys. Rev.* B **83** 205208
[114] Yang J, Li H, Wu T, Zhang W, Chen L and Yang J 2008 Evaluation of Half-Heusler compounds as thermoelectric materials based on the calculated electrical transport properties *Adv. Funct. Mater.* **18** 2880–8
[115] Yang J, Qiu P, Liu R, Xi L, Zheng S, Zhang W, Chen L, Singh D J and Yang J 2011 Trends in electrical transport of *p*-type skutterudites RFe_4Sb_{12} (R = Na, K, Ca, Sr, Ba, La, Ce, Pr, Yb) from first-principles calculations and Boltzmann transport theory *Phys. Rev.* B **84** 235205
[116] Babaei H, Khodadadi J M and Sinha S 2014 Large theoretical thermoelectric power factor of suspended single-layer MoS_2 *Appl. Phys. Lett.* **105** 193901
[117] Kutorasinski K, Wiendlocha B, Kaprzyk S and Tobola J 2015 Electronic structure and thermoelectric properties of *n*- and *p*-type SnSe from first-principles calculations *Phys. Rev.* B **91** 205201
[118] Sjakste J, Vast N and Tyuterev V 2007 *Ab initio* method for calculating electron-phonon scattering times in semiconductors: application to GaAs and GaP *Phys. Rev. Lett.* **99** 236405
[119] Tyuterev V G, Obukhov S V, Vast N and Sjakste J 2011 *Ab initio* calculation of electron-phonon scattering time in germanium *Phys. Rev.* B **84** 035201
[120] Tandon N, Albrecht J D and Ram-Mohan L R 2015 Electron-phonon interaction and scattering in Si and Ge: implications for phonon engineering *J. Appl. Phys.* **118** 045713
[121] Noffsinger J, Kioupakis E, Van de Walle C G, Louie S G and Cohen M L 2012 Phonon-assisted optical absorption in silicon from first principles *Phys. Rev. Lett.* **108** 167402
[122] Bernardi M, Vigil-Fowler D, Ong C S, Neaton J B and Louie S G 2015 *Ab initio* study of hot electrons in GaAs *Proc. Natl. Acad. Sci.* **112** 5291–6
[123] Logan R A and Peters A J 1960 Impurity effects upon mobility in silicon *J. Appl. Phys.* **31** 122–4
[124] Geballe T H and Hull G W 1955 Seebeck effect in silicon *Phys. Rev.* **98** 940–7
[125] Mahan G D, Lindsay L and Broido D A 2014 The seebeck coefficient and phonon drag in silicon *J. Appl. Phys.* **116** 245102
[126] Zhou J, Liao B, Qiu B, Huberman S, Esfarjani K, Dresselhaus M S and Chen G 2015 *Ab initio* optimization of phonon drag effect for lower-temperature thermoelectric energy conversion *Proc. Natl. Acad. Sci.* **112** 14777–82
[127] Liao B, Zhou J, Qiu B, Dresselhaus M S and Chen G 2015 *Ab initio* study of electron-phonon interaction in phosphorene *Phys. Rev.* B **91** 235419
[128] Himmetoglu B, Janotti A, Peelaers H, Alkauskas A and Van de Walle C G 2014 First-principles study of the mobility of $SrTiO_3$ *Phys. Rev.* B **90** 241204
[129] Murphy-Armando F and Fahy S 2008 First-principles calculation of carrier-phonon scattering in *n*-type $Si_{1-x}Ge_x$ alloys *Phys. Rev.* B **78** 035202
[130] Kojima A, Teshima K, Shirai Y and Miyasaka T 2009 Organometal Halide Perovskites as visible-light sensitizers for photovoltaic cells *J. Am. Chem. Soc.* **131** 6050–1