Supplementary Information

Ab initio optimization of phonon drag effect for lower-temperature thermoelectric energy conversion

Jiawei Zhou[1], Bolin Liao[1], Bo Qiu[1], Samuel Huberman[1], Keivan Esfarjani[2,3], Mildred S. Dresselhaus[4,5], and Gang Chen[1]

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

2. Department of Mechanical and Aerospace Engineering, Rutgers University, Piscataway, New Jersey, 08854, USA

3. Institute for Advanced Materials, Devices and Nanotechnology, Rutgers University, Piscataway, New Jersey, 08854, USA

4. Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

5. Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Supplementary Note 1. Coupled electron-phonon Boltzmann equation

The analysis based on coupled electron-phonon Boltzmann equation for the phonon drag effect has been given in previous work(1–3). We do not repeat the derivations but provide some necessary details to complete the computational formalism we propose. The coupled electron-phonon Boltzmann equation has been shown in the main text:

$$\begin{cases} \mathbf{v}_\alpha(\mathbf{k})\cdot\dfrac{\partial f_\alpha(\mathbf{k})}{\partial T}\nabla T - e\mathbf{v}_\alpha(\mathbf{k})\cdot\dfrac{\partial f_\alpha(\mathbf{k})}{\partial E}\nabla\varphi = -\dfrac{f_\alpha(\mathbf{k})-f_\alpha^0(\mathbf{k})}{\tau_\alpha^*(\mathbf{k})} + \left(\dfrac{\partial f_\alpha(\mathbf{k})}{\partial t}\right)_{e-ph} \\ \mathbf{v}_\lambda(\mathbf{q})\cdot\dfrac{\partial n_\lambda(\mathbf{q})}{\partial T}\nabla T = -\dfrac{n_\lambda(\mathbf{q})-n_\lambda^0(\mathbf{q})}{\tau_\lambda^*(\mathbf{q})} + \left(\dfrac{\partial n_\lambda(\mathbf{q})}{\partial t}\right)_{e-ph} \end{cases} \quad \text{(S1)}$$

where f and N represent the distribution functions for electrons and phonons respectively, with equilibrium state described by

$$\begin{cases} f_\alpha^0(\mathbf{k}) = \dfrac{1}{e^{(E_\mathbf{k}^\alpha-\mu)/k_BT}+1} \\ N_\lambda^0(\mathbf{q}) = \dfrac{1}{e^{\hbar\omega_\mathbf{q}^\lambda/k_BT}-1} \end{cases} \quad \text{(S2)}$$

For electrons, the external driving forces include the electrochemical potential gradient $\nabla\varphi$ and temperature gradient ∇T, while for phonons the only external driving force comes from the temperature gradient. The right-hand-side of Eq. (S1) describes the various scattering events experienced by electrons and phonons.

In equation (S1), we assume that all other scattering mechanisms except the electron-phonon interaction can be described by the mode-dependent relaxation time model, adding these contributions together according to Matthiessen's rule. For electrons $\tau_\alpha^*(\mathbf{k})$ describes the electron-impurity scattering. For phonons $\tau_\lambda^*(\mathbf{q})$ includes both phonon-phonon scattering and phonon-impurity scattering. The scattering rates due to the electron-phonon interaction can be obtained if we have the knowledge of the interaction matrix elements:

$$\begin{cases} \left(\dfrac{\partial f_\alpha(\mathbf{k})}{\partial t}\right)_{e-ph} = \dfrac{2\pi}{\hbar}\displaystyle\sum_{\mathbf{k}'\beta,\mathbf{q}\lambda} \begin{Bmatrix} -\left|g_{\alpha\beta\lambda}(\mathbf{k},\mathbf{k}',\mathbf{q})\right|^2 n_{\mathbf{q}\lambda} f_{\mathbf{k}\alpha}(1-f_{\mathbf{k}'\beta})\delta(\mathbf{k}'-\mathbf{k}-\mathbf{q})\delta(E_{\mathbf{k}'\beta}-E_{\mathbf{k}\alpha}-\hbar\omega_{\mathbf{q}\lambda}) \\ -\left|g_{\alpha\beta\lambda}(\mathbf{k},\mathbf{k}',\mathbf{q})\right|^2 (n_{\mathbf{q}\lambda}+1) f_{\mathbf{k}\alpha}(1-f_{\mathbf{k}'\beta})\delta(\mathbf{k}'-\mathbf{k}+\mathbf{q})\delta(E_{\mathbf{k}'\beta}-E_{\mathbf{k}\alpha}+\hbar\omega_{\mathbf{q}\lambda}) \\ +\left|g_{\alpha\beta\lambda}(\mathbf{k}',\mathbf{k},\mathbf{q})\right|^2 n_{\mathbf{q}\lambda}(1-f_{\mathbf{k}\alpha}) f_{\mathbf{k}'\beta}\delta(\mathbf{k}-\mathbf{k}'-\mathbf{q})\delta(E_{\mathbf{k}\alpha}-E_{\mathbf{k}'\beta}-\hbar\omega_{\mathbf{q}\lambda}) \\ +\left|g_{\alpha\beta\lambda}(\mathbf{k}',\mathbf{k},\mathbf{q})\right|^2 (n_{\mathbf{q}\lambda}+1)(1-f_{\mathbf{k}\alpha}) f_{\mathbf{k}'\beta}\delta(\mathbf{k}-\mathbf{k}'+\mathbf{q})\delta(E_{\mathbf{k}\alpha}-E_{\mathbf{k}'\beta}+\hbar\omega_{\mathbf{q}\lambda}) \end{Bmatrix} \\ \left(\dfrac{\partial n_\lambda(\mathbf{q})}{\partial t}\right)_{e-ph} = \dfrac{2\pi}{\hbar}\dfrac{1}{2}\displaystyle\sum_{\mathbf{k}\alpha,\mathbf{k}'\beta} \begin{Bmatrix} -\left|g_{\alpha\beta\lambda}(\mathbf{k},\mathbf{k}',\mathbf{q})\right|^2 n_{\mathbf{q}\lambda} f_{\mathbf{k}\alpha}(1-f_{\mathbf{k}'\beta})\delta(\mathbf{k}'-\mathbf{k}-\mathbf{q})\delta(E_{\mathbf{k}'\beta}-E_{\mathbf{k}\alpha}-\hbar\omega_{\mathbf{q}\lambda}) \\ +\left|g_{\alpha\beta\lambda}(\mathbf{k},\mathbf{k}',\mathbf{q})\right|^2 (n_{\mathbf{q}\lambda}+1) f_{\mathbf{k}\alpha}(1-f_{\mathbf{k}'\beta})\delta(\mathbf{k}'-\mathbf{k}+\mathbf{q})\delta(E_{\mathbf{k}'\beta}-E_{\mathbf{k}\alpha}+\hbar\omega_{\mathbf{q}\lambda}) \\ -\left|g_{\alpha\beta\lambda}(\mathbf{k}',\mathbf{k},\mathbf{q})\right|^2 n_{\mathbf{q}\lambda}(1-f_{\mathbf{k}\alpha}) f_{\mathbf{k}'\beta}\delta(\mathbf{k}-\mathbf{k}'-\mathbf{q})\delta(E_{\mathbf{k}\alpha}-E_{\mathbf{k}'\beta}-\hbar\omega_{\mathbf{q}\lambda}) \\ +\left|g_{\alpha\beta\lambda}(\mathbf{k}',\mathbf{k},\mathbf{q})\right|^2 (n_{\mathbf{q}\lambda}+1)(1-f_{\mathbf{k}\alpha}) f_{\mathbf{k}'\beta}\delta(\mathbf{k}-\mathbf{k}'+\mathbf{q})\delta(E_{\mathbf{k}\alpha}-E_{\mathbf{k}'\beta}+\hbar\omega_{\mathbf{q}\lambda}) \end{Bmatrix} \end{cases}$$

with the electron-phonon interaction matrix element

$$g_{\alpha\beta\lambda}(\mathbf{k},\mathbf{k}',\mathbf{q}) = \left(\frac{\hbar}{2m_0\omega_{\mathbf{q}\lambda}}\right)^{1/2} \cdot \left\langle \mathbf{k}'\beta \middle| \partial_{\mathbf{q}\lambda} V \middle| \mathbf{k}\alpha \right\rangle \tag{S3}$$

In equation (S3), $E_{\mathbf{k}\alpha}$ and $E_{\mathbf{k}'\beta}$ describes the energy of electron states, $\omega_{\mathbf{q}\lambda}$ describes the phonon frequency and the two delta functions impose momentum and energy conservation respectively. In the formula for the electron-phonon interaction matrix element, m_0 is the mass of one unit cell, $\left|\mathbf{k}\alpha\right\rangle$ and $\left|\mathbf{k}'\beta\right\rangle$ describe the eigenstates of electrons, while $\partial_{\mathbf{q}\lambda} V$ is the perturbing potential due to the ionic displacement corresponding to a phonon with wave vector $\mathbf{q}$ and branch number λ. Note that for the scattering rate for phonons there is an extra $1/2$. This is because when $\mathbf{k}$ and $\mathbf{k}'$ go over the Brillouin zone each $\mathbf{k}_1 \to \mathbf{k}_2$ process is counted twice (let $\mathbf{k} = \mathbf{k}_1$ and $\mathbf{k}' = \mathbf{k}_2$ or vice versa).

In equilibrium, all distribution functions take their equilibrium values. In this case, the scattering rates as shown in (S3) should vanish, because otherwise the state of the system will move away from the equilibrium (see Eq. (S1)). The lowest order approximation for Eq. (S3) comes by taking the first order deviation of the distribution functions, which gives rise to the widely-used linearized Boltzmann equation:

$$\begin{cases}\left(\dfrac{\partial f_\alpha(\mathbf{k})}{\partial t}\right)_{e-ph} \simeq -\left[\sum\limits_{\mathbf{k}'\beta,\mathbf{q}\lambda} F_{\mathbf{k}\alpha}(\mathbf{k}'\beta,\mathbf{q}\lambda)\right]\cdot\Delta f_{\mathbf{k}\alpha}+\sum\limits_{\mathbf{k}'\beta,\mathbf{q}\lambda}\left[F_{\mathbf{k}'\beta}(\mathbf{k}\alpha,\mathbf{q}\lambda)\cdot\Delta f_{\mathbf{k}'\beta}\right]+\sum\limits_{\mathbf{k}'\beta,\mathbf{q}\lambda}\left[F_{\mathbf{q}\lambda}(\mathbf{k}\alpha,\mathbf{k}'\beta)\cdot\Delta n_{\mathbf{q}\lambda}\right] \\ \left(\dfrac{\partial n_\lambda(\mathbf{q})}{\partial t}\right)_{e-ph} \simeq \sum\limits_{\mathbf{k}\alpha,\mathbf{k}'\beta}\left[G_{\mathbf{k}\alpha}(\mathbf{k}'\beta,\mathbf{q}\lambda)\cdot\Delta f_{\mathbf{k}\alpha}+G_{\mathbf{k}'\beta}(\mathbf{k}\alpha,\mathbf{q}\lambda)\cdot\Delta f_{\mathbf{k}'\beta}\right]-\left[\sum\limits_{\mathbf{k}\alpha,\mathbf{k}'\beta} G_{\mathbf{q}\lambda}(\mathbf{k}\alpha,\mathbf{k}'\beta)\right]\cdot\Delta n_{\mathbf{q}\lambda}\end{cases} \tag{S4}$$

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

$$\begin{cases}F_{\mathbf{k}\alpha}(\mathbf{k}'\beta,\mathbf{q}\lambda)=\left[\left(n^0_{\mathbf{q}\lambda}+f^0_{\mathbf{k}'\beta}\right)\Pi_-+\left(n^0_{\mathbf{q}\lambda}+1-f^0_{\mathbf{k}'\beta}\right)\Pi_+\right] \\ F_{\mathbf{k}'\beta}(\mathbf{k}\alpha,\mathbf{q}\lambda)=\left[\left(n^0_{\mathbf{q}\lambda}+1-f^0_{\mathbf{k}\alpha}\right)\Pi_-+\left(n^0_{\mathbf{q}\lambda}+f^0_{\mathbf{k}\alpha}\right)\Pi_+\right] \\ F_{\mathbf{q}\lambda}(\mathbf{k}\alpha,\mathbf{k}'\beta)=\left[\left(f^0_{\mathbf{k}'\beta}-f^0_{\mathbf{k}\alpha}\right)\Pi_-+\left(f^0_{\mathbf{k}'\beta}-f^0_{\mathbf{k}\alpha}\right)\Pi_+\right] \\ G_{\mathbf{k}\alpha}(\mathbf{k}'\beta,\mathbf{q}\lambda)=\left[-\left(n^0_{\mathbf{q}\lambda}+f^0_{\mathbf{k}'\beta}\right)\Pi_-+\left(n^0_{\mathbf{q}\lambda}+1-f^0_{\mathbf{k}'\beta}\right)\Pi_+\right] \\ G_{\mathbf{k}'\beta}(\mathbf{k}\alpha,\mathbf{q}\lambda)=\left[\left(n^0_{\mathbf{q}\lambda}+1-f^0_{\mathbf{k}\alpha}\right)\Pi_--\left(n^0_{\mathbf{q}\lambda}+f^0_{\mathbf{k}\alpha}\right)\Pi_+\right] \\ G_{\mathbf{q}\lambda}(\mathbf{k}\alpha,\mathbf{k}'\beta)=\left[\left(f^0_{\mathbf{k}'\beta}-f^0_{\mathbf{k}\alpha}\right)\Pi_--\left(f^0_{\mathbf{k}'\beta}-f^0_{\mathbf{k}\alpha}\right)\Pi_+\right]\end{cases} \tag{S5}$$

with

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

denoting processes due to the absorption of a phonon $\omega_{\mathbf{q}\lambda}$ and the emission of a phonon, respectively. The first-order deviations $\Delta f_{\mathbf{k}\alpha}=f_{\mathbf{k}\alpha}-f^0_{\mathbf{k}\alpha}$, $\Delta f_{\mathbf{k}'\beta}=f_{\mathbf{k}'\beta}-f^0_{\mathbf{k}'\beta}$ and $\Delta n_{\mathbf{q}\lambda}=n_{\mathbf{q}\lambda}-n^0_{\mathbf{q}\lambda}$ characterize the non-equilibrium state of electrons and phonons. For normal electrical property calculations, the relaxation time approximation is often used. This approximation naturally arises if we assume that only the distribution of the initial state of the electron deviates from the equilibrium ($\Delta f_{\mathbf{k}\alpha}\neq 0$) and that of the final electron state and of phonons remain at equilibrium ($\Delta f_{\mathbf{k}'\beta}=0,\Delta n_{\mathbf{q}\lambda}=0$), which is essentially the Bloch condition. In this case, the prefactor before $\Delta f_{\mathbf{k}\alpha}$ can be defined as $1/\tau^{e-ph}_\alpha(\mathbf{k})$. The relaxation time $\tau^{e-ph}_\alpha(\mathbf{k})$ is what is usually called the electron-phonon relaxation time (for electrons), which determines the intrinsic

mobility of one material. We should also note that $\Delta f_{\mathbf{k}'\beta}$ is essentially neglected 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, which is called the momentum relaxation time(4). In semiconductors, however, it is proved, based on deformation potential models, that for nearly isotropic scattering, the neglect of $\Delta f_{\mathbf{k}'\beta}$ will not cause much difference(4). It has also been shown(5, 6) that without considering $\Delta f_{\mathbf{k}'\beta}$, good agreement for the electrical properties in silicon with experiments can be achieved, justifying the approximation that terms containing $\Delta f_{\mathbf{k}'\beta}$ can be neglected.

A more important perturbation term from equilibrium comes in the evaluation of the assumption $\Delta n_{\mathbf{q}\lambda}=0$. It is clear that this assumption makes non-equilibrium phonons have no effect on the electron system. When phonons are far away from equilibrium, assuming $\Delta n_{\mathbf{q}\lambda}$ to be zero is no longer valid. These non-equilibrium phonons described by non-zero $\Delta n_{\mathbf{q}\lambda}$ in the electron system (the last term in the first equation of (S4)) are responsible for the phonon drag effect.

The above picture is based on the Seebeck effect, where a temperature gradient induces a phonon heat flow, which delivers part of its momenta to the electron system and gives rise to an extra current. Because of the Kelvin relation $\Pi=TS$, an extra contribution to the Seebeck coefficient also implies an extra Peltier coefficient. We want to further clarify the phonon drag effect in the context of the Peltier effect, which completes the picture and provides a straightforward derivation of equation (2) in the main text. For the Peltier effect, electrons are first driven by the electric field. These non-equilibrium electrons can then deliver their momenta to the phonon system through the first two terms of the right-hand-side in the second equation of (S4), which can be analogously called the "electron drag" effect but is essentially just the manifestation of the "phonon drag" effect in the Peltier effect. Therefore the phonon

system acquires an extra heat flow due to electron motions, which contributes to the Peltier coefficient.

Now we still have one term left (the last term in the second equation of (S4)), which describes the scattering of phonons by equilibrium electrons. The prefactor of $\Delta n_{\mathbf{q}\lambda}$ in the phonon Boltzmann equation can be readily written as $1/\tau_{\lambda}^{e-ph}(\mathbf{q})$ (just as the definition of electron-phonon relaxation time for electrons). Apparently, higher doping concentrations lead to stronger scattering. We show in our paper that this higher doping concentration is responsible for the reduction of the phonon drag effect in heavily-doped samples compared to lightly-doped samples. Besides, this "electron drag" also accounts for some fraction of the reduction of the thermal conductivity in heavily-doped materials(7).

Having discussed the meaning of each term in equation (S1) and (S4), now we want to make the inclusion of these scattering terms more compact by rearranging them. If we incorporate the first term of the right-hand-side in the first line of Eq. (S4) and the last term in the second line of Eq. (S4) into the relaxation times we have in (S1), the coupled Boltzmann equation now becomes

$$\begin{cases} \left(\dfrac{\partial f_{\alpha}(\mathbf{k})}{\partial t}\right)_{drift} = -\dfrac{f_{\alpha}(\mathbf{k}) - f_{\alpha}^{0}(\mathbf{k})}{\tau_{\alpha}(\mathbf{k})} + \sum_{\mathbf{k}'\beta,\mathbf{q}\lambda}\left[F_{\mathbf{k}'\beta}(\mathbf{k}\alpha,\mathbf{q}\lambda)\cdot\Delta f_{\mathbf{k}'\beta}\right] + \sum_{\mathbf{k}'\beta,\mathbf{q}\lambda}\left[F_{\mathbf{q}\lambda}(\mathbf{k}\alpha,\mathbf{k}'\beta)\cdot\Delta n_{\mathbf{q}\lambda}\right] \\ \left(\dfrac{\partial n_{\lambda}(\mathbf{q})}{\partial t}\right)_{drift} = -\dfrac{n_{\lambda}(\mathbf{q}) - n_{\lambda}^{0}(\mathbf{q})}{\tau_{\lambda}(\mathbf{q})} + \sum_{\mathbf{k}\alpha,\mathbf{k}'\beta}\left[G_{\mathbf{k}\alpha}(\mathbf{k}'\beta,\mathbf{q}\lambda)\cdot\Delta f_{\mathbf{k}\alpha} + G_{\mathbf{k}'\beta}(\mathbf{k}\alpha,\mathbf{q}\lambda)\cdot\Delta f_{\mathbf{k}'\beta}\right] \end{cases} \tag{S6}$$

where the relaxation times also include the electron scattering by equilibrium phonons $\tau_{\alpha}^{eph}(\mathbf{k})$ as well as the phonon scattering by equilibrium electrons $\tau_{\lambda}^{eph}(\mathbf{q})$:

$$\begin{cases} \dfrac{1}{\tau_{\alpha}(\mathbf{k})} = \dfrac{1}{\tau_{\alpha}^{*}(\mathbf{k})} + \dfrac{1}{\tau_{\alpha}^{eph}(\mathbf{k})} \\ \dfrac{1}{\tau_{\lambda}(\mathbf{q})} = \dfrac{1}{\tau_{\lambda}^{*}(\mathbf{q})} + \dfrac{1}{\tau_{\lambda}^{eph}(\mathbf{q})} \end{cases} \tag{S7.a}$$

with

$$\begin{cases} \dfrac{1}{\tau_{\alpha}^{eph}(\mathbf{k})} = \sum\limits_{\mathbf{k}'\beta,\mathbf{q}\lambda} F_{\mathbf{k}\alpha}(\mathbf{k}'\beta,\mathbf{q}\lambda) \\ \dfrac{1}{\tau_{\lambda}^{eph}(\mathbf{q})} = \sum\limits_{\mathbf{k}\alpha,\mathbf{k}'\beta} G_{\mathbf{q}\lambda}(\mathbf{k}\alpha,\mathbf{k}'\beta) \end{cases} \tag{S7.b}$$

and the remaining terms describe the coupling between non-equilibrium states in electron and phonon systems.

Supplementary Note 2. Derivation of the phonon drag Seebeck coefficient

We have above derived the coupled electron-phonon Boltzmann equation in a compact form as shown in Eq. (S6). The electron relaxation time $\tau_\alpha(\mathbf{k})$ incorporates electron-impurity scattering and electron scattering by equilibrium phonons, while the phonon relaxation time $\tau_\lambda(\mathbf{q})$ contains phonon-phonon interaction, phonon-impurity scattering and phonon scattering by equilibrium electrons. The coupling through the non-equilibrium distribution is manifested by the collision terms that are not described by the relaxation times in Eq. (S6) and those collision terms are responsible for the phonon drag effect. In the main text we have derived the formalism based on the Seebeck picture (the temperature gradient generates a voltage difference), here we adopt the Peltier picture (the isothermal electric field produces a heat flow), which directly provides the phonon drag contribution from each phonon mode and also shows as an explicit proof of the Kelvin relation for the phonon drag effect.

For the Peltier effect, a non-equilibrium distribution of electrons is generated first by the electric field, which will then drive the phonons away from equilibrium. The induced non-equilibrium phonons will now perturb the electrons in a second-order effect, which can be justified by the fact that the phonon drag phenomenon is found to have a small influence on the electrical conductivity(3). Therefore for the electron system, we can then assume that phonons are at equilibrium ($\Delta n_{\mathbf{q}\lambda}=0$ in the first line in (S6), and note that for the Peltier effect the phonon drag comes in through the last term in the second equation of (S6) instead of the last term in the first equation). This assumption exactly corresponds to the treatment we did in the main text to decouple the phonon transport from the electron system (i.e., assume $\Delta f_{\mathbf{k}\alpha}=0$ in the phonon BTE) when deriving the phonon drag formula in the Seebeck picture. As a result, the electron distribution function can be directly written down using the relaxation time model $\Delta f_{\mathbf{k}\alpha}=e\tau_{\mathbf{k}\alpha}\mathbf{v}_{\mathbf{k}\alpha}\cdot\nabla\varphi\frac{\partial f^0_{\mathbf{k}\alpha}}{\partial E}$ (Here we also assume that the term

$\sum_{\mathbf{k}'\beta,\mathbf{q}\lambda}\left[F_{\mathbf{k}'\beta}(\mathbf{k}\alpha,\mathbf{q}\lambda)\cdot\Delta f_{\mathbf{k}'\beta}\right]$ will vanish, which is a commonly used approximation(4)).

The normal electrical conductivity and Peltier coefficient (related to the normal Seebeck coefficient via the Kelvin relation $\Pi = TS$) can be obtained by looking at the charge current and energy current induced by such a non-equilibrium electron distribution, respectively:

$$\begin{cases} \sigma = \left(\dfrac{e}{3N_{\mathbf{k}}}\sum\limits_{\mathbf{k}\alpha} v_{x,\mathbf{k}\alpha}\Delta f_{\mathbf{k}\alpha}\right)/\left(-\nabla\varphi\right) = \dfrac{e^2}{3N_{\mathbf{k}}}\sum\limits_{\mathbf{k}\alpha}\mathbf{v}_{\mathbf{k}\alpha}^2\tau_{\mathbf{k}\alpha}\left(-\dfrac{\partial f_{\mathbf{k}\alpha}^0}{\partial E}\right) \\ \Pi = \left(\dfrac{e}{3N_{\mathbf{k}}}\sum\limits_{\mathbf{k}\alpha}(E-\mu)v_{x,\mathbf{k}\alpha}\Delta f_{\mathbf{k}\alpha}\right)/\left(\dfrac{e}{3N_{\mathbf{k}}}\sum\limits_{\mathbf{k}\alpha} v_{x,\mathbf{k}\alpha}\Delta f_{\mathbf{k}\alpha}\right) \\ \quad = \dfrac{e}{3\sigma N_{\mathbf{k}}}\sum\limits_{\mathbf{k}\alpha}(E-\mu)\mathbf{v}_{\mathbf{k}\alpha}^2\tau_{\mathbf{k}\alpha}\left(-\dfrac{\partial f_{\mathbf{k}\alpha}^0}{\partial E}\right) \end{cases} \tag{S8}$$

The energy current not only can come from the electrons, but also has its origin in the phonon heat flow. As we have discussed, in the isothermal condition the phonons acquire the momentum via the electron-phonon coupling shown by the last term in the second equation of (S6) and lead to an extra heat flow, which manifests the "phonon drag" effect. Given the electron distribution in (S6), the phonon Boltzmann equation now becomes

$$0 = -\frac{n_\lambda(\mathbf{q}) - n_\lambda^0(\mathbf{q})}{\tau_\lambda(\mathbf{q})} + e\nabla\varphi\cdot\sum_{\mathbf{k}\alpha,\mathbf{k}'\beta}\left[G_{\mathbf{k}\alpha}(\mathbf{k}'\beta,\mathbf{q}\lambda)\tau_{\mathbf{k}\alpha}\mathbf{v}_{\mathbf{k}\alpha}\frac{\partial f_{\mathbf{k}\alpha}^0}{\partial E} + G_{\mathbf{k}'\beta}(\mathbf{k}\alpha,\mathbf{q}\lambda)\tau_{\mathbf{k}'\beta}\mathbf{v}_{\mathbf{k}'\beta}\frac{\partial f_{\mathbf{k}'\beta}^0}{\partial E}\right] \tag{S9}$$

where the drift term vanishes because there is no temperature gradient. It can be readily solved to obtain the phonon distribution function

$$\Delta n_\lambda(\mathbf{q}) = \tau_\lambda(\mathbf{q})e\nabla\varphi\cdot\sum_{\mathbf{k}\alpha,\mathbf{k}'\beta}\left[G_{\mathbf{k}\alpha}(\mathbf{k}'\beta,\mathbf{q}\lambda)\tau_{\mathbf{k}\alpha}\mathbf{v}_{\mathbf{k}\alpha}\frac{\partial f_{\mathbf{k}\alpha}^0}{\partial E} + G_{\mathbf{k}'\beta}(\mathbf{k}\alpha,\mathbf{q}\lambda)\tau_{\mathbf{k}'\beta}\mathbf{v}_{\mathbf{k}'\beta}\frac{\partial f_{\mathbf{k}'\beta}^0}{\partial E}\right] \tag{S10}$$

Considering the heat flow described by $\mathbf{q} = \sum_{\mathbf{q}\lambda}\hbar\omega_{\mathbf{q}\lambda}\mathbf{v}_{\mathbf{q}\lambda}\Delta n_{\mathbf{q}\lambda}$, we finally arrive at the phonon drag Peltier coefficient.

$$\Pi_{ph} = \frac{2e}{3\sigma\Omega N_{\mathbf{k}} N_{\mathbf{q}} k_B T} \sum_{\mathbf{q}\lambda} \hbar\omega_{\mathbf{q}\lambda}\tau_{\mathbf{q}\lambda}\mathbf{v}_{\mathbf{q}\lambda} \cdot \left[\sum_{\mathbf{k}\alpha,\mathbf{k}'\beta} \left(\tau_{\mathbf{k}\alpha}\mathbf{v}_{\mathbf{k}\alpha} - \tau_{\mathbf{k}'\beta}\mathbf{v}_{\mathbf{k}'\beta}\right) \cdot f^0_{\mathbf{k}\alpha}\left(1 - f^0_{\mathbf{k}'\beta}\right) n^0_{\mathbf{q}\lambda} \cdot \frac{2\Pi}{\hbar} \right] \quad \text{(S11)}$$

$$\text{with} \quad \Pi = \pi \left| g_{\alpha\beta\lambda}(\mathbf{k},\mathbf{k}',\mathbf{q}) \right|^2 \cdot \delta\left(E_{\mathbf{k}'\beta} - E_{\mathbf{k}\alpha} - \hbar\omega_{\mathbf{q}\lambda}\right) \cdot \delta\left(\mathbf{k}' - \mathbf{k} - \mathbf{q}\right)$$

The comparison of Eq. (S11) with Eq. (2) in the main text clearly shows the Kelvin relation ($\Pi_{ph} = TS_{ph}$).

Supplementary Note 3. Simulation details

Knowing all the quantities in Eq. (2) of the main text, phonon drag Seebeck coefficient can be calculated on an equal electron and phonon Brillouin zone mesh. The total electron relaxation time $\tau_{\mathbf{k}\alpha}$, as we have elaborated, combines both electron-phonon scattering as well as the electron-impurity scattering, according to Matthiessen's rule. Similarly, the phonon relaxation time $\tau_{\mathbf{q}\lambda}$ includes both phonon-phonon scattering, phonon-electron scattering and phonon-impurity scattering. Supplementary Table S1 shows the scattering mechanisms considered in this work and how they are treated.

Supplementary Table S1. Scattering mechanisms for electrons and phonons

Carrier	Scattering	Method
Electron	Electron-phonon	DFT
	Electron-impurity	Brooks-Herring model(4, 8)
Phonon	Phonon-phonon	DFT
	Phonon-electron	DFT
	Phonon-impurity	Tamura model(9)

The equilibrium properties of electrons and phonons are calculated from first principles using the QUANTUM ESPRESSO package(10) as described in Methods. Apart from these equilibrium properties, the key ingredients towards the first principles result are the quantities describing non-equilibrium properties. The electron phonon interaction matrix element $g_{\alpha\beta\lambda}(\mathbf{k},\mathbf{k}',\mathbf{q}) = \left(\frac{\hbar}{2m_0\omega_{\mathbf{q}\lambda}}\right)^{1/2} \cdot \langle \mathbf{k}'\beta | \partial_{\mathbf{q}\lambda} V | \mathbf{k}\alpha \rangle$ is one of them and leads to the electron relaxation time(11). The stringent convergence demands the knowledge of both wavefunctions and perturbing potentials on an ultra-dense mesh, which only became accessible recently due to a Wannier basis-based interpolation scheme(12), allowing us to interpolate between the matrix

elements from the coarse meshs to produce finer meshes. Fine meshes up to $100\times100\times100$ have been tested for their convergence (for both electrons and phonons). To obtain the electron relaxation time due to the electron-phonon coupling, we use the Gaussian smearing method. The convergence with respect to the Gaussian broadening parameter has also been checked. A band energy cutoff (measured from the band edge) is used to select only electron states near the band edge because states far away from the band edge will not contribute to the transport. Besides, the phonon relaxation times are also required. They are essentially related to the anharmonic force constants(13, 14), which can be obtained from first principles calculations. For heavily-doped samples, electron-phonon scattering of phonons also needs to be considered for calculating the phonon relaxation time. This is solved following our previous paper(7). Knowing all of these quantities, equation (2) in the main text (or Eq. (S11)) can be calculated at points on an equal electron and phonon Brillouin zone mesh, for which the tetrahedra integration method is implemented(15). To further speed up the calculation, based on the knowledge that only phonons that are close to the zone center contribute to the phonon drag effect, we define a wave vector cutoff, above which phonons will not be considered for equation (1). This cutoff has been checked and the change of the result is within 1% of the original value. Supplementary Table S2 lists some of the key parameters we used in this calculation.

Supplementary figure S1 shows the resulting intrinsic mobility and thermal conductivity with respect to temperature as a test to the electron and phonon relaxation times we obtain. We note here that previous first principles calculation have obtained similar agreement with experiments for the electron mobility(5, 6, 16) and the thermal conductivity(13, 14) in silicon. As is seen in supplementary figure S1, overall the results agree well with the experimental data. There is some discrepancy for the hole mobility in p-type silicon near room temperature. This shows that the experimental samples experience more scattering and therefore bear a lower mobility. We speculate this to be a result of the split valence bands due to the spin-orbit coupling at valence band edge. We cannot confirm this point yet because the spin-orbit coupling is not included in this work. However, this discrepancy should not

affect the Seebeck coefficient calculation much, because we know from the Boltzmann description of the diffusive Seebeck coefficient and equation (1) that both of them will not be changed if the electron scattering time is just changed by a single constant factor. Therefore the results on p-type silicon can still give guidance in how to utilize the phonon drag effect in p-type materials.

For heavily-doped silicon, impurity scattering needs to be considered. Due to the lack of accurate and computationally feasible methods for calculating the impurity scattering, the effects from impurities are described using empirical models – the Brooks-Herring model for electron-impurity scattering(4, 8) and the Tamura model for phonon-impurity scattering(9). It was known that the Brooks-Herring model tends to underestimate the electron impurity scattering rate(4). We found from our calculation that the electron-impurity scattering has a small influence on the phonon drag effect. This small influence lies in the fact that the Seebeck coefficient essentially represents the ratio of the temperature-gradient induced current to the electric-field driven current. When the electron relaxation time is reduced, both of them are weakened and therefore the ratio between them is less affected. The Tamura model(9) is used to examine the effect of the ionized impurity scattering on the phonon drag effect, where the mass difference ratio $\Delta M / \bar{M}$ is chosen to be 1 to represent both the mass disorder and strain effect. We found from the calculation that phonon-ionized impurity scattering also only has small influences on the phonon drag effect. Based on the reasons given above, the use of empirical models in our calculation can be justified.

Supplementary Table S2. Parameters used in determining the electron and phonon relaxation times as well as in the calculation of the phonon drag effect. The parameter "a" in the last column is the lattice constant of silicon.

Quantities required	Electron relaxation time	Phonon-phonon relaxation time	Electron-phonon scattering of phonons	Phonon drag effect
k-mesh (electron)	70^3~100^3		70^3	70^3~100^3
q-mesh (phonon)	80^3	70^3~100^3	70^3	70^3~100^3
Integration method (broadening parameter if any)	Gaussian (0.002eV)	Gaussian ($1cm^{-1}$)	Tetrahedra	Tetrahedra
Energy / wavevector ($\mathbf{q}$) cutoff	0.5 eV	$\mathbf{q}:(0.2\sim 2)\frac{2\pi}{a}$	1.0 eV	energy: 0.5eV $\mathbf{q}:(0.2\sim 2)\frac{2\pi}{a}$
Nearest neighbor considered in the force constant fitting(13)		2^{nd} force constant (harmonic): 7 3^{rd} force constant (anharmonic): 1		

Supplementary Note 4. Electron mode contribution to phonon drag

Here we show the mode-specific contributions to the phonon drag Seebeck coefficient as well as the electrical conductivity and diffusive Seebeck coefficient from the electron side. This provides us the knowledge of what portion of electrons contributes to these transport properties and especially the phonon drag effect most notably, which will be used when we discuss the usage of nanoclusters to selectively scatter phonons. Equation (2) of the main text directly presents the phonon mode-specific contribution to the phonon drag (term inside the bracket). For the electrons, we can similarly combine all the terms that are labeled with the same electron wave vector **k** and band number α in the summation. The result is given in supplementary figure S4 for the lightly-doped n-type silicon with a doping concentration of 10^{14} cm^{-3}.

In general, we see that at the same temperature the accumulated contribution curves for the three physical quantities (electrical conductivity, diffusive Seebeck coefficient and phonon drag Seebeck coefficient) almost overlap. This is because the Fermi-Dirac distribution, which modifies the population of the electrons, changes more strongly with the electron states compared to other properties such as scattering rates, and essentially confines the electron states that are important for transport properties to a small region near the band edge. Therefore we see a general monotonically-increasing accumulated contribution curve. To be more specific, we see in supplementary figure S4a that, when temperature decreases, the curves move towards the band edge. This is a result of the temperature characteristics of Fermi-Dirac distribution function, which decreases more rapidly with energy as temperature decreases. As a result, the electrons that participate in the transport are more confined to the band edge at lower temperatures. At 300K, most electron states that are important for the transport properties are located within 0.2eV from the band edge. Because electrons are confined to the band edge, which is made up of six equivalent electron pockets, and the conduction band minimum corresponds to a wavelength of 0.67nm, significant contributions to the transport properties should come from electrons with wavelength

around 0.67nm. In supplementary figure S4b, indeed we see that most of the contributions come from electrons with wavelength between 0.6nm and 0.7nm. At lower temperatures, the curve becomes slightly narrower, and the reason is the same as before: electrons become more confined to the band edge and the reciprocal space they occupy then shrinks. In terms of the electron mean free path, Supplementary figure S4c indicates that, electrons at 300K have mean free paths between 20nm and 80nm. The mean free path increases as temperature decreases, and at 100K the majority of electrons have mean free paths around 100nm~300nm. We should note that these plots are obtained for a lightly-doped silicon. For the heavily-doped silicon, the characteristics of the energy-dependence(5) and wavelength-dependence will remain the same, meaning that electrons involved in the transport process are still confined within ~0.2eV from the band edge and have wavelengths around 0.6nm~0.7nm, because the qualitative argument given above does not change. In comparison, mean free paths of electrons will decrease due to the impurity scattering, and therefore the mean free path accumulated curve will move towards the left. For example, it was known that for the n-type silicon with 10^{19} cm^{-3} doping concentration, the electrons have mean free paths below 20nm within the temperature range of 100K~300K(5).

Supplementary Note 5. Nanocluster scattering as a phonon frequency filter

Here we examine the nanocluster scattering to effectively scatter high-frequency phonons. Nanoclusters are clusters that have impurity atoms different from the host atoms with sizes ranging from sub-nanometer to a few nanometers. One extreme case of the nanocluster is a single impurity atom embedded in the host, for which the theoretical model developed by Tamura can be used to estimate the phonon-impurity scattering(9). For clusters that contain more than one impurity atom, there has been development of first principles approach based on Green's function calculation(17, 18), which can provide more accurate results. For simplicity, we will not use such a rigorous method to describe the nanocluster scattering. Instead, we use an analytical formula(17) generalized from the Tamura model for the description of the nanocluster scattering effect. In the Born approximation, it can be shown that the phonon-nanocluster scattering rate is

$$\tau_{imp}^{-1}(\mathbf{q},\lambda) = \frac{\pi}{12N} f \left(\frac{\Delta M}{\bar{M}} \right)^2 \omega_{\mathbf{q}\lambda}^2 \cdot D^*(\omega_{\mathbf{q}\lambda})$$
$$D^*(\omega_{\mathbf{q}\lambda}) = 6\sum_{\mathbf{q}'\lambda'} \left[\left| \sum_{\sigma} \vec{\mathbf{e}}_{\sigma}^*(\mathbf{q}',\lambda') \cdot \vec{\mathbf{e}}_{\sigma}(\mathbf{q},\lambda) \right|^2 \cdot \left| S_{\Delta\mathbf{q}} \right|^2 \cdot \delta\left(\omega_{\mathbf{q}\lambda} - \omega_{\mathbf{q}'\lambda'} \right) \right] \quad \text{(S12)}$$

where N is the total number of unit cells, f is the volume fraction of the nanoclusters, ΔM is the mass difference of the impurity atom and the host atom, $\bar{M}$ is the average mass of all the atoms, $\omega_{\mathbf{q}\lambda}$ describes the phonon frequency and $\vec{\mathbf{e}}_{\sigma}(\mathbf{q},\lambda)$ is the unit vector along the polarization of the atom labeled by σ in the unit cell. $S_{\Delta\mathbf{q}}$ is the structure factor with the sum includes all the unit cells occupied by one nanocluster. $D^*(\omega_{\mathbf{q}\lambda})$ can be regarded as a generalized phonon density of state. It can be shown that when the nanocluster contains only one impurity atom (the structure factor is one in this case), $D^*(\omega_{\mathbf{q}\lambda})$ reduces to the normal phonon density of state, and as a result, Eq. (S12) is essentially the same as the Tamura model. The unit vectors along the polarizations of the atoms as well as the phonon frequencies are

obtained from first principle calculations. The mass fraction term $\frac{\Delta M}{\bar{M}}$ is chosen to be 1 to represent both mass disorder and force constant disorder. This is a typical number for alloys. For example, if the host is silicon, then the addition of germanium atoms act as impurities with a mass fraction of around 1.6.

We test nanocluster size up to 1nm (this is the equivalent diameter defined through the total volume of the unit cells contained in the nanocluster). For the value of 1nm, electron wavelengths are comparable to the nanoclusters size (see supplementary figure S4). We can use the geometric limit to estimate the upper bound for the electron-nanocluster scattering, which gives a corresponding mean free path of $\Lambda_{nanoparticle} = \frac{4r}{3f} \approx 330\text{nm}$, where r is the characteristic radius of the nanoparticle and chosen to be 0.5nm for estimation (volume fraction f here is chosen to be 0.2%, which is the maximum value we set). For nanoclusters that are smaller, the geometric limit becomes smaller. However, the wavelengths are now large compared to the nanocluster sizes and enter into the Rayleigh scattering regime, where the scattering rate falls below the geometric limit. Therefore $\Lambda_{nanoparticle}$ should be on the order of 300nm or even larger. From previous work on first principles calculation of silicon(5) we know that for the doping concentration of 10^{19} cm^{-3} the electron mean free paths are less than 20nm. The comparison with $\Lambda_{nanoparticle}$ indicates that the dominant scatterings for electrons still come from the phonons and dopants. Therefore the nanocluster scattering for electrons can be neglected and the electrical conductivity is barely affected.

Supplementary Note 6. A unified picture regarding the carrier concentration dependence of phonon drag

For the theoretical understanding of the phonon drag, there have been two theories: one by Herring(1) stating that the phonon drag does not depend on the carrier concentration as long as the concentration is small, and the other provided by MacDonald(19) showing that the phonon drag magnitude should be inversely proportional to the carrier concentration. We think the difference is mainly due to the different carrier concentration regions that are examined. In the following, we will show that in fact these two arguments are consistent with each other and how our results can be put under such a unified picture.

First we give a brief unified discussion on these two theories. Both theories consider the phonon drag as a momentum transfer from phonons to electrons, and consider its contribution to the Seebeck coefficient to be proportional to the extra momentum gain per electron, or $S \propto \frac{\Delta p}{n}$, where Δp is the total momentum gain of the electron system and n is the carrier concentration. Based on this picture, the carrier concentration not only directly modifies the phonon drag, but also changes the phonon drag through affecting the total momentum transfer. It can be further shown that $\Delta p \propto \bar{\tau} \cdot n$, where $\bar{\tau}$ is the averaged relaxation time of phonons that contribute to phonon drag (similar to Herring's formula(1) but here we are considering Seebeck picture instead of Peltier picture). The momentum transfer depends on the carrier concentration mainly because the momentum transfer is limited by the available electron states that can couple to phonons. Higher carrier concentration implies a Fermi level closer to the conduction band (for n-type material) and therefore more electron populations. In fact, both the carrier concentration and the electron population depends on the same exponential term $\exp(E_f / k_B T)$. Therefore we expect the momentum transfer to be proportional to n. Besides, the relaxation times of the phonons that couple to electrons and contribute to phonon drag are of primary importance because longer relaxation times would suggest that phonons can transfer

more momentum to the electrons while shorter relaxation times imply that the phonon momentum is more frequently destroyed before it can be transferred. Considering all these, the carrier concentration dependence of the phonon drag can be simplified to

$$S \propto \bar{\tau}(n) \tag{S13}$$

In lightly-doped semiconductors, $\bar{\tau}$ is mainly determined by intrinsic phonon-phonon scattering and therefore the phonon drag Seebeck coefficient does not depend on the carrier concentration (shown in Fig. S8). As the concentration becomes high, $\bar{\tau}$ starts to decrease due to the scattering of phonons by electrons, leading to the so-called "saturation" effect. The "saturation" effect in this case means the saturation of the total momentum that is transferred from phonons to the electron system ($\Delta p \propto \bar{\tau} \cdot n \rightarrow \text{constant}$). Because the electron concentration is still increasing, the average momentum gain per electron decreases. This leads to the reduction of the phonon drag and can be seen in supplementary figure S8.

MacDonald(19) has used a more phenomenological model to explain the phonon drag, where the pressure on the electrons is generated due to the phonon heat flow. He proposed that the total momentum transfer should be proportional to the variation of the phonon's energy with respect to the temperature: $\Delta p \propto \frac{\partial U}{\partial T} \propto c_v$. As a result, the phonon drag Seebeck coefficient will be inversely proportional to the carrier concentration ($S \propto \frac{\Delta p}{n} \propto \frac{c_v}{n}$). This $1/n$ dependence however inherently has an assumption (19): the phonons are solely scattered by electrons, which makes the result only applicable to metals with high electron concentration. This can be understood by looking at equation (S13). When the carrier concentration is very high, the phonons that contribute to phonon drag (those with long wavelengths) are predominantly scattered by electrons. It has been shown that for electron-phonon scattering the averaged relaxation time of phonons is inversely proportional to the carrier concentration(7, 20) (This is true for nondegenerate case but only an approximation for dengerately-doped semiconductors). As a result, equation (S13) suggests that in the limit of degenerately-doped semiconductors, one has $S \propto \bar{\tau} \propto \frac{1}{n}$. More

specifically, as the carrier concentration increases, the available electrons that can couple to phonons increase, while the phonons are more strongly scattered and have less momentum to transfer, making the total momentum transfer saturate. Therefore, Macdonald's prediction can be viewed as a limiting case of our more general consideration (equation (S13)) when the carrier concentration becomes extremely high, and clearly equation (S13) connects the two theories from low and moderately-high carrier concentration to extremely high carrier concentration region.

In supplementary figure S8 we redraw the phonon drag Seebeck coefficient with both log scales for Seebeck coefficient and carrier concentration. The carrier concentration dependence can be separated into three regions. At low carrier concentrations the phonon drag is a constant (represented by the green dashed line). As the carrier concentration increases, the phonon drag starts to decrease ("saturation" effect). If the carrier concentration further increases beyond 10^{21}cm^{-3}, we would expect the phonon drag starts to follow MacDonald's prediction ($1/n$ dependence, as represented by the blue dashed line) and may not benefit the zT enhancement. However, the maximum carrier concentration in our calculation is still in the transition region and the reduction of the phonon drag is not significant, therefore we observe a weaker carrier concentration dependence than the $1/n$ behavior. For the carrier concentration range we have examined, the MacDonald's theory cannot be applied and this is also the reason why we can optimize the material for maximizing the phonon drag effect.

References

1. Herring C (1954) Theory of the Thermoelectric Power of Semiconductors. *Phys Rev* 96(5):1163–1187.

2. J. M. Ziman (1960) *Electrons and Phonons: The Theory of Transport Phenomena in Solids* (Clarendon Press).

3. Bailyn M (1958) Transport in Metals: Effect of the Nonequilibrium Phonons. *Phys Rev* 112(5):1587–1598.

4. Lundstrom M (2002) Fundamentals of Carrier Transport, 2nd edn. *Meas Sci Technol* 13(2):230.

5. Qiu B, et al. (2015) First-principles simulation of electron mean-free-path spectra and thermoelectric properties in silicon. *EPL Europhys Lett* 109(5):57006.

6. Restrepo OD, Varga K, Pantelides ST (2009) First-principles calculations of electron mobilities in silicon: Phonon and Coulomb scattering. *Appl Phys Lett* 94(21):212103.

7. Liao B, et al. (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(11):115901.

8. Brooks H (1955) Theory of the Electrical Properties of Germanium and Silicon. *Advances in Electronics and Electron Physics*, ed Marton L (Academic Press), pp 85–182.

9. Tamura S (1983) Isotope scattering of dispersive phonons in Ge. *Phys Rev B* 27(2):858–866.

10. Giannozzi P, et al. (2009) QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J Phys Condens Matter* 21(39):395502.

11. Giustino F, Cohen M, Louie S (2007) Electron-phonon interaction using Wannier functions. *Phys Rev B* 76(16):165108.

12. Noffsinger J, et al. (2010) EPW: A program for calculating the electron–phonon coupling using maximally localized Wannier functions. *Comput Phys Commun* 181(12):2140–2148.

13. Esfarjani K, Chen G, Stokes H (2011) Heat transport in silicon from first-principles calculations. *Phys Rev B* 84(8):085204.

14. Ward A, Broido DA (2010) Intrinsic phonon relaxation times from first-principles studies of the thermal conductivities of Si and Ge. *Phys Rev B* 81(8):085205.

15. Lambin P, Vigneron J (1984) Computation of crystal Green's functions in the complex-energy plane with the use of the analytical tetrahedron method. *Phys Rev B* 29(6):3430–3437.

16. Wang Z, et al. (2011) Thermoelectric transport properties of silicon: Toward an ab initio approach. *Phys Rev B* 83(20):205208.

17. Kundu A, Mingo N, Broido DA, Stewart DA (2011) Role of light and heavy embedded nanoparticles on the thermal conductivity of SiGe alloys. *Phys Rev B* 84(12):125426.

18. Katcho NA, Mingo N, Broido DA (2012) Lattice thermal conductivity of (Bi${}_{1$-${}x}$Sb${}_{x}$)${}_{2}$Te${}_{3}$ alloys with embedded nanoparticles. *Phys Rev B* 85(11):115208.

19. D.K.C. MacDonald (1962) *Thermoelectricity: an introduction to the principles* (John Wiley & Sons).

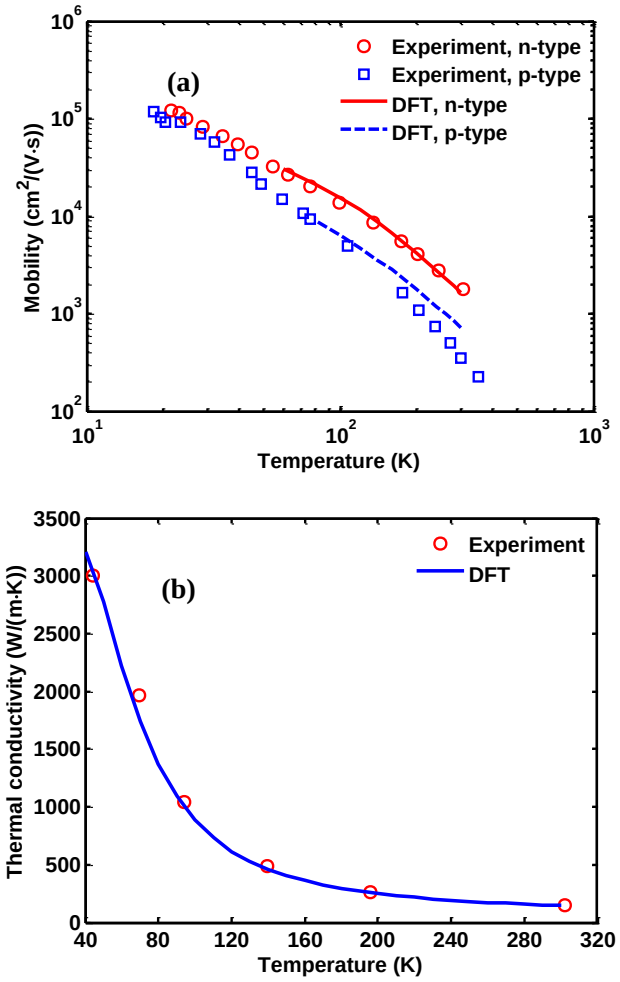
20. Ziman JM (1956) XVII. The effect of free electrons on lattice conduction. *Philos Mag* 1(2):191–198.

21. Logan RA, Peters AJ (1960) Impurity Effects upon Mobility in Silicon. *J Appl Phys* 31(1):122–124.

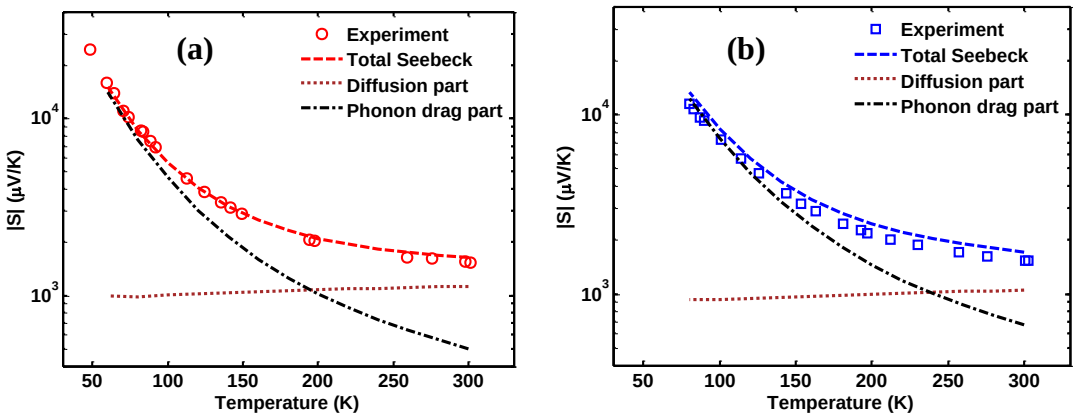
22. Glassbrenner C, Slack G (1964) Thermal Conductivity of Silicon and Germanium from 3 ℃K to the Melting Point. *Phys Rev* 134(4A):A1058–A1069.

23. Geballe TH, Hull GW (1955) Seebeck Effect in Silicon. *Phys Rev* 98(4):940–947.

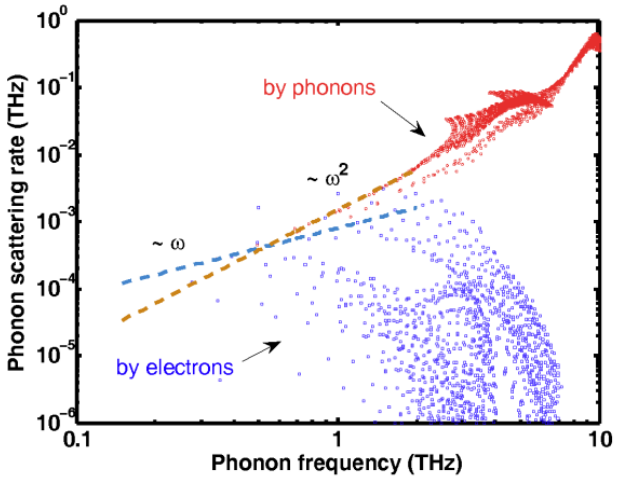
Supplementary Figure Legends



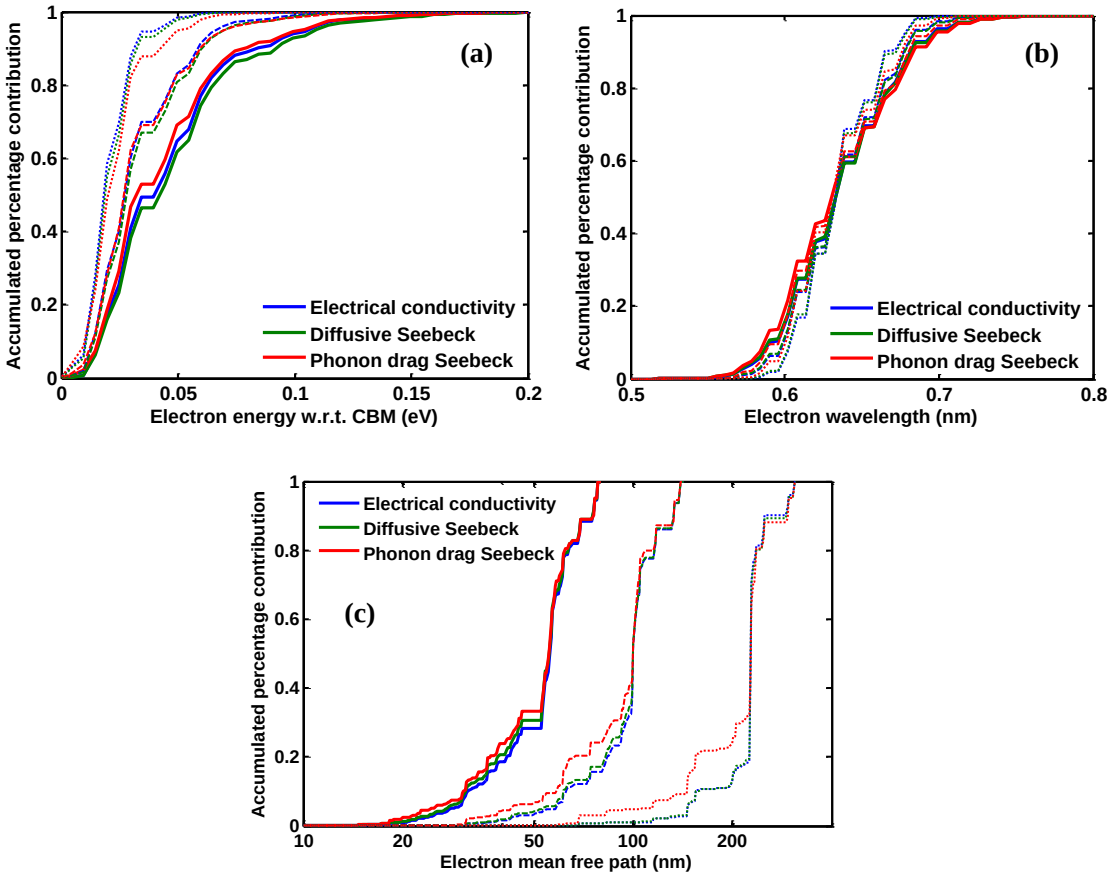
Supplementary Figure S1. (a) Temperature dependence of the intrinsic mobility in n-type and p-type silicon compared with that of sufficiently pure samples(21) as well as (b) the thermal conductivity of pure silicon compared with the experiments(22). The intrinsic mobility is calculated assuming a carrier concentration of 10^{14} cm^{-3}.



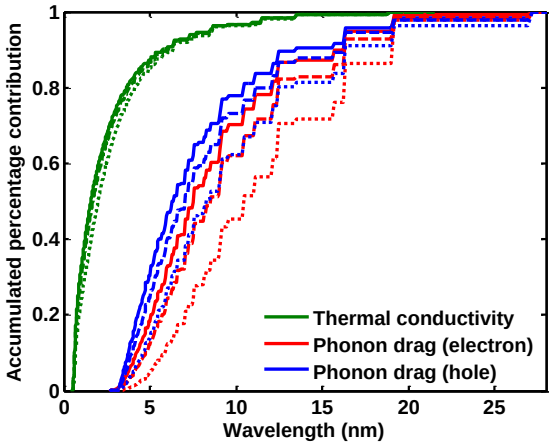
Supplementary Figure S2. Phonon drag Seebeck coefficient for (a) electrons and (b) holes in lightly-doped silicon. The open circles and squares are taken from the experiment(23), with the corresponding net doping concentration of $2.8\times10^{14}\,\mathrm{cm}^{-3}$ for electrons and $8.1\times10^{14}\,\mathrm{cm}^{-3}$ for holes, respectively. Lines are first principles results assuming the same doping concentrations. Dotted lines represent the diffusive Seebeck coefficient while dash-and-dot lines represent the phonon drag contribution with respect to the temperature on a semilog plot. The phonon drag contribution increases dramatically as the temperature decreases and converges to the total Seebeck coefficient, shown by the dashed lines. The experimental data(23) with larger sample size and lower net doping concentration are chosen as comparison.



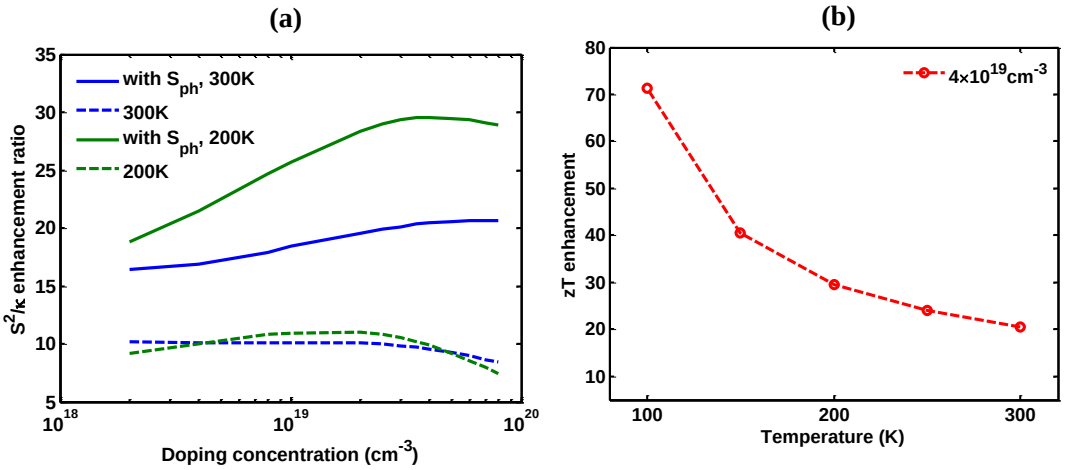
Supplementary Figure S3. Phonon scattering rates by phonons themselves and by electrons. The phonon-phonon scattering rate follows a $1/\tau \sim \omega^2$ trend for low-frequency phonons. The phonon-electron scattering only roughly follows a $1/\tau \sim \omega$ because this scaling behavior is valid only for very low frequency phonons and does not consider the anisotropy of the energy bands(7). Nonetheless, it is clearly seen that below around 1THz, the phonon-electron scattering starts to dominate over the intrinsic phonon-phonon scattering. The calculation is carried out for the n-type silicon with a 10^{19} cm^{-3} doping concentration at 300K.



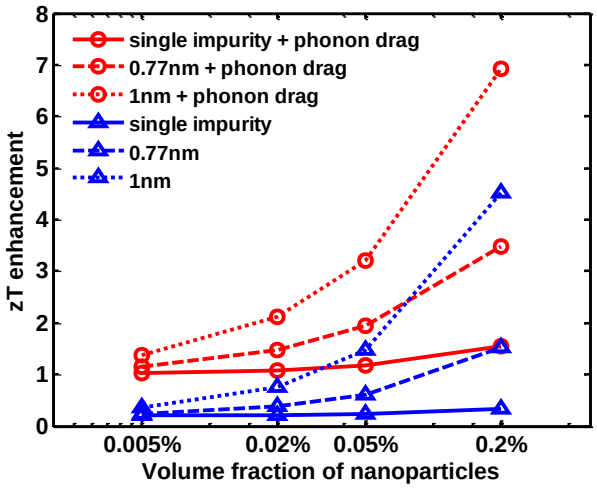
Supplementary Figure S4. Accumulated contribution to electrical conductivity, diffusive Seebeck coefficient and phonon drag Seebeck coefficient, with respect to (a) electron energy (measured from the conduction band edge), (b) electron wavelength and (c) electron mean free path. Physical quantities are labeled with different colors (blue for electrical conductivity, green for diffusive Seebeck coefficient and red for phonon drag Seebeck coefficient). In all three plots, solid curves describe results at 300K, while dashed curves represent 200K and dotted lines 100K. The results are calculated assuming a doping concentration of 10^{14} cm^{-3}.



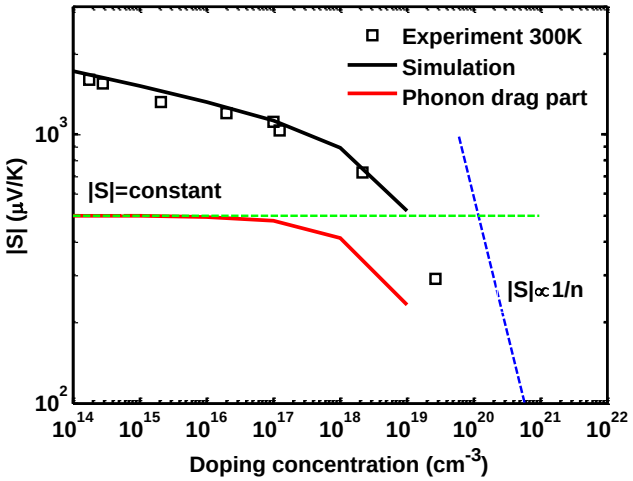
Supplementary Figure S5. Phonon mode-specific accumulated contributions to the phonon drag Seebeck coefficient and the thermal conductivity with respect to the phonon wavelength. Solid lines show the contribution at 300K, while dashed lines are used at 200K and dotted lines at 100K. Green curves show results for the thermal conductivity. Red and blue curves represent results for phonon drag in n-type and p-type silicon, respectively. Clearly phonons that are significant in phonon drag typically have longer wavelengths. Besides, we can see that the spectral difference between the contributions of the phonon modes to the phonon drag Seebeck coefficient and to the thermal conductivity becomes larger at lower temperatures.



Supplementary Figure S6. (a) The enhancement of the factor S^2/κ as a function of doping concentration when phonon modes are selectively scattered and (b) the resulting zT enhancement assuming electrical conductivity is not affected at a doping concentration of $4\times10^{19}\mathrm{cm}^{-3}$ in n-type silicon. In (a), solid lines represent the results where phonon drag is included and preferable phonon modes are chosen, while for the dashed lines it is assumed that phonon drag is neglected. The results examined in (a) assume that the thermal conductivity is reduced to $4\mathrm{W/(m\cdot K)}$ for all the curves. Without the phonon drag effect, the enhancement of S^2/κ is around 10, which barely changes from 300K to 200K, because the diffusion contribution to the Seebeck coefficient decreases while the thermal conductivity actually increases as the temperature decreases, which tend to cancel each other. The situation is different if the phonon drag effect is included. First, because the phonon drag magnitude is comparable to the diffusion contribution, the inclusion of phonon drag with selective phonon modes helps to boost S^2/κ. Furthermore, when the temperature is lowered, the phonon drag effect becomes even more pronounced and therefore makes the enhancement of S^2/κ even larger. The latter point is more clearly seen in (b), where we fix the doping concentration to be $4\times10^{19}\mathrm{cm}^{-3}$ and vary the temperature. The enhancement of S^2/κ directly translates to the zT enhancement compared to the bulk doped silicon at same temperatures, reaching a value as large as ~70 at 100K.



Supplementary Figure S7. The enhancement of the thermoelectric figure of merit zT as volume fraction of nanoclusters using phonon frequency selectivity. The calculation is done for the n-type silicon with the doping concentration of 10^{19} cm^{-3} at 200K. Curves labeled with phonon drag include the phonon drag contribution to the Seebeck coefficient while others neglect the phonon drag effect. It can be seen that, using 1nm size nanoclusters with a volume fraction of 0.2% at 200K, the enhancement can reach above 7, and the neglect of the phonon drag effect will reduce the Seebeck coefficient and therefore the enhancement as well. At the same volume fraction, nanoclusters with larger size scatter phonons more strongly. This fact is known for decades (2) and is utilized here to select low-frequency phonons. However, nanoclusters that are much larger will also significantly reduce the phonon drag effect. Therefore there will be an optimal choice for the size of the nanoclusters and our calculations show that this optimal size is around 1nm for silicon at a doping concentration of 10^{19} cm^{-3}.



Supplementary Figure S8. Carrier concentration dependence of phonon drag Seebeck coefficient in silicon at 300K in a log-log scale. The squares are taken from the experiment(23) while the simulation results are the same with that in Fig. 1. Here we add two dashed lines to represent the limiting cases of the carrier concentration dependence of the phonon drag: the constant behavior (Herring's picture for semiconductors with low carrier concentration (1)) shown by the dashed green line and the $1/n$ behavior (Macdonald's prediction for metals (19)) shown by the dashed blue line.