Improved thermoelectric cooling based on the Thomson effect

G. Jeffrey Snyder,[1,*] Eric S. Toberer,[2] Raghav Khanna,[1] and Wolfgang Seifert[3]

[1]*Materials Science, California Institute of Technology, 1200 E. California Boulevard, Pasadena, California 91125, USA*
[2]*Department of Physics, Colorado School of Mines, Golden, Colorado 80401, USA*
[3]*Institute of Physics, University Halle-Wittenberg, D-06099 Halle, Germany*

Traditional thermoelectric Peltier coolers exhibit a cooling limit which is primarily determined by the figure of merit, zT. Rather than a fundamental thermodynamic limit, this bound can be traced to the difficulty of maintaining thermoelectric compatibility. Self-compatibility locally maximizes the cooler's coefficient of performance for a given zT and can be achieved by adjusting the relative ratio of the thermoelectric transport properties that make up zT. In this study, we investigate the theoretical performance of thermoelectric coolers that maintain self-compatibility across the device. We find that such a device behaves very differently from a Peltier cooler, and we term self-compatible coolers "Thomson coolers" when the Fourier heat divergence is dominated by the Thomson, as opposed to the Joule, term. A Thomson cooler requires an exponentially rising Seebeck coefficient with increasing temperature, while traditional Peltier coolers, such as those used commercially, have comparatively minimal change in Seebeck coefficient with temperature. When reasonable material property bounds are placed on the thermoelectric leg, the Thomson cooler is predicted to achieve approximately twice the maximum temperature drop of a traditional Peltier cooler with equivalent figure of merit (zT). We anticipate that the development of Thomson coolers will ultimately lead to solid-state cooling to cryogenic temperatures.

I. INTRODUCTION

Peltier coolers are the most widely used solid-state cooling devices, enabling a wide range of applications from thermal management of optoelectronics and infrared detector arrays to polymerase chain reaction (PCR) instruments. Thermoelectric coolers have been traditionally understood by means of the Peltier effect, which describes the reversible heat transported by an electric current. This effect is traditionally understood in terms of absorption or release of heat at the junction of two dissimilar materials. The conventional analysis of a Peltier cooler approximates the material properties as independent of temperature [constant property model (CPM)]. This results in a maximum cooling temperature difference ΔT_{max} for a CPM cooler, which is dependent on the figure of merit ZT of the device,[1,2]

$$\Delta T_{max} = \frac{ZT_c^2}{2}. \qquad (1)$$

For the best commercial materials this leads to a ΔT_{max} of 65 K (single stage),[3] which translates to a device ZT at 300 K of 0.74. In the CPM the device ZT is equal to the material zT. Material zT depends on the Seebeck coefficient (α), temperature (T), electrical resistivity (ρ), and thermal conductivity (κ), $zT = \frac{\alpha^2 T}{\rho \kappa}$. In the CPM, the only way to increase ΔT_{max} for a single stage is to increase zT, leading to the focus of much thermoelectric research on improving zT. It is well known that even further cooling to lower temperatures can be achieved using multistage Peltier coolers.[1,2] In principle, each stage can produce additional cooling to lower temperatures, regardless of the zT of the thermoelectric material in the stage. In practice, the thermal losses and complications of fabrication limit the performance of such devices. The six-stage cooler of Marlow achieves a ΔT_{max} of 133 K; this doubling of ΔT_{max} compared to a single-stage cooler is achieved despite using materials with

similar zT.[3] Alternatively, such ΔT_{max} with a single-stage CPM cooler would require ZT to be 2.5.

The transport properties across a single thermoelectric leg can be manipulated to improve cooling performance, although it has been less effective in reducing ΔT_{max} than a multistage approach. One common strategy is to engineer a change in extrinsic dopant concentration across a thermoelectric element which can significantly alter α, ρ, and even κ. For example, this has been demonstrated for thermoelectric generators in n-type PbTe doped with I.[4] Similar efforts have been done with cooling materials, as has been reviewed in Ref. 5. The simplest explanation for an improvement is an increase in the local zT at some temperatures by spatially adjusting the dopant composition within a material.[6]

Early theoretical work by Sherman *et al.* for thermoelectric coolers (TECs) found that different ΔT_{max} could be predicted from materials that have the same or similar average zT but different temperature dependence of the individual properties α, ρ, κ.[7] This demonstrated that optimizing cooler performance is significantly more complex than simply maximizing zT. More recently, Müller and co-workers[8–10] and Bian and co-workers[11,12] used different numerical approaches to predict substantial gains in cooling to ΔT_{max} from functionally grading where an average zT remains constant in an effort to determine the best approach to functionally grading.

Different material classes optimized for different temperatures can also be segmented together to improve performance of thermoelectric generators but the current must also be matched.[13] The analysis of segmentation strikingly demonstrates that increasing the average zT does not always lead to an increase in overall thermoelectric efficiency and so an understanding of the thermoelectric compatibility factor is needed to explain device performance.[14]

This paper derives the cooling limit for a single-stage, fully optimized (self-compatible) TEC that functions as an

infinitely staged cooler. The Fourier heat divergence in such an optimized cooler is found to be dominated by the Thomson effect rather than the Joule heating as in traditional Peltier coolers. This new opportunity presents a new challenge for material optimization based on compatibility factor rather than only zT.

II. THEORY

Coolers are characterized by the coefficient of performance ($\phi = Q_c/P$), which relates the rate of heat extraction at the cold end Q_c to the power consumption P in the device.[15] For simplicity, but without loss of generality, a single thermoelectric element can be considered rather than a complete device. A TEC leg can be treated as an infinite series of infinitesimal coolers, each of which is operating locally with some coefficient of performance (COP). Scaling this COP to the local Carnot COP (T/dT) yields the local reduced coefficient of performance ϕ_r.[16] This relationship between local performance across the leg and global COP, ϕ, given in Eq. (2), is derived in the Appendix based on Refs. 7 and 17,

$$\frac{1}{\phi} = \exp\left(\int_{T_c}^{T_h} \frac{1}{T}\frac{1}{\phi_r(T)}dT\right) - 1. \tag{2}$$

While TECs are traditionally analyzed using a global approach, we have previously shown the utility of a local approach.[14,15,18] This local approach leads to a consideration of material "compatibility".

The compatibility approach to optimizing thermoelectric cooling arises naturally from an analysis of the thermal and electric transport equations. This method has been described in detail for thermoelectric generators[15] and coolers[19] and are reproduced here for TEC. The method has been experimentally verified[20] and shown to reproduce results using a more traditional finite-element result but with less computational complexity. This method has been incorporated into several engineering models such as those used by NASA for radioisotope thermoelectric generators[21] and Amerigon/BSST for automotive applications.[20,22] Consider an infinitesimal section of thermoelectric leg in a temperature gradient and an electric field. The temperature gradient will induce a Fourier heat flux ($\mathbf{q}_\kappa = -\kappa\nabla T$) across this segment. The divergence of this heat [Eq. (3)] is equal to the source terms: irreversible Joule heating (ρj^2) and the reversible Thomson heat ($T\frac{d\alpha}{dT}j\nabla T$), both of which depend on the electric current density (j). From these two effects, the governing equation for heat flow in vector notation is

$$\nabla \cdot \mathbf{q}_\kappa = \nabla \cdot (-\kappa\nabla T) = \rho j^2 - \tau\,\mathbf{j}\cdot\nabla T \tag{3}$$

with Joule heat per volume ρj^2, Thomson coefficient $\tau = T\frac{d\alpha}{dT}$, and Thomson heat per volume $\tau\,\mathbf{j}\cdot\nabla T$. The Peltier, Seebeck, and Thomson effect are all manifestations of the same thermoelectric property characterized by α. The Thomson coefficient ($\tau = T\frac{d\alpha}{dT}$) describes the Thomson heat absorbed or released when current flows in the direction of a temperature gradient.

Restricting the problem to one spatial dimension, Eq. (3) is typically examined assuming the heat flux and electric current are parallel.[15] In the typical CPM model used to analyze Peltier
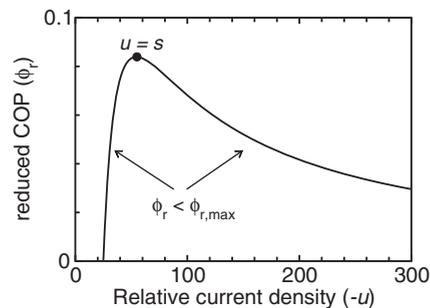

FIG. 1. The local reduced coefficient of performance ϕ_r is optimized at a specific reduced current density, termed s. If $u \neq s$, the ϕ_r is less than that predicted by the material zT. Here, $z = 0.002\,\mathrm{K}^{-1}$, $\alpha = 200\,\mu\mathrm{V\,K}^{-1}$.

coolers, the Thomson effect is zero because α is constant along the leg ($\frac{d\alpha}{dT} = 0$).

The exact performance of a thermoelectric leg with $\alpha(T)$, $\rho(T)$, and $\kappa(T)$ possessing arbitrary temperature dependence can be straightforwardly computed using the reduced variables: relative current density (u) and thermoelectric potential (Φ).[18] The relative current density u, given in Eq. (4), is primarily determined by the electrical current density j, which is adjusted to achieve maximum global COP. The thermoelectric potential Φ is a state function which simplifies Eq. (2) to Eq. (6),[15]

$$u = \frac{-j^2}{\kappa\nabla T \cdot \mathbf{j}}, \tag{4}$$

$$\Phi = \alpha T + 1/u, \tag{5}$$

$$\phi = \frac{\Phi(T_c)}{\Phi(T_h) - \Phi(T_c)}. \tag{6}$$

Changing variables to T via the monotonic function $x(T)$, Eq. (3) simplifies to the differential equation in $u(T)$,

$$\frac{du}{dT} = u^2\left(T\frac{d\alpha}{dT} + \frac{\alpha^2}{z}u\right). \tag{7}$$

Using this formalism, the reduced coefficient of performance (ϕ_r) can be simply defined for any point in the cooler [Eq. (8)]. Figure 1 shows this relationship between u and ϕ_r. From Eq. (2), it can be shown that ϕ is largest when ϕ_r is maximized for every infinitesimal segment along the cooler. Hence, global maximization can be traced back to local optimization,[23]

$$\phi_r = \frac{u\frac{\alpha}{z} + \frac{1}{zT}}{u\frac{\alpha}{z}\left(1 - u\frac{\alpha}{z}\right)} = \frac{u\alpha + \frac{1}{T}}{u(\alpha - u\rho\kappa)} \tag{8}$$

The optimum u which maximizes ϕ_r ($\frac{d\phi_r}{du} = 0$) can be expressed solely in terms of local material properties [Eq. (9)]. This optimum value of u is defined as the thermoelectric compatibility factor s_c for coolers,

$$s_c = \frac{-\sqrt{1+zT}-1}{\alpha T}. \tag{9}$$

As this paper strictly focuses on coolers, we will refer to s_c as simply s.

The maximum local ϕ_r, denoted $\phi_{r,\max}$, occurs when $u = s$. The expression for $\phi_{r,\max}$ [Eq. (10)] is an explicit function of

the material zT and is independent of the individual properties α, ρ, κ. This maximum allowable local efficiency provides a natural justification for the definition of zT as the material's figure of merit,

$$\phi_{r,\max} = \frac{\sqrt{1+zT}-1}{\sqrt{1+zT}+1}. \tag{10}$$

One thus wishes to construct devices where, locally, each segment has "$u = s$" and thus $\phi_{r,\max}$ is obtained. Globally, maximum ϕ is found when the entire cooler satisfies $u = s$.

III. COOLING PERFORMANCE

To compare the cooling performance of traditional Peltier coolers and $u = s$ coolers, we consider coolers with equivalent z. Traditional Peltier coolers have typically been analyzed with the constant property model (CPM), yielding a constant z (where zT is linearly increasing with temperature). We will show that constant z, but allowing α, κ, ρ to vary with T, can lead to substantial improvement in cooling. At the limit of this variation, we will assume that the properties can be varied to satisfy $u = s$.

Performance of a CPM cooler. CPM coolers have been extensively studied, typically using a global approach to the transport behavior. The ϕ for a CPM cooler (operated at optimum j) is given by Eq. (11).[2] Figure 2(a) shows that the ϕ of a CPM cooler decreases with increasing ΔT. With increasing cooling, this ϕ decreases and reaches zero at $\Delta T_{\max}$

(a)

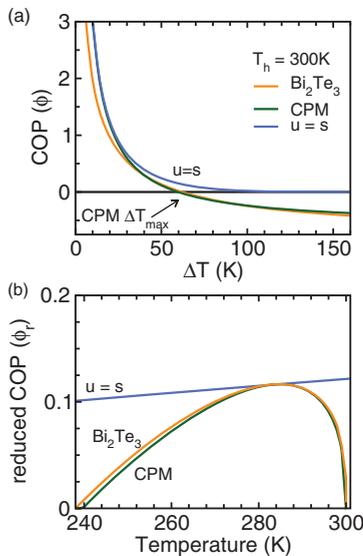FIG. 2. (Color online) (a) The CPM Peltier cooler and $u = s$ Thomson cooler are compared using the same constant $z = 0.002$ K^{-1}. The overall device ϕ of a CPM cooler crosses zero at a finite temperature, indicating $\Delta T_{\max}$ is reached, while ϕ remains positive for all temperatures for the $u = s$ cooler. (b) The local performance of a CPM cooler (ϕ_r) is significantly compromised at both the hot and cold ends. In contrast, $\phi_{r,\max}$ is achieved at all temperatures when $u = s$. In both panels, the performance calculated for an actual Bi$_2$Te$_3$ Peltier cooler leg is similar to the CPM.

[Eq. (1)],

$$\phi^{CPM} = \left(\frac{T_c}{\Delta T}\right)\left(\frac{\sqrt{1+zT_{\text{avg}}}-\frac{T_h}{T_c}}{\sqrt{1+zT_{\text{avg}}}+1}\right). \tag{11}$$

To understand what is limiting the CPM cooler at $\Delta T_{\max}$, we derive the local reduced coefficient of performance $\phi_r^{CPM}(T)$. To obtain ϕ_r^{CPM} we need u as a function of T. The solution to differential equation (7) for CPM is

$$\frac{1}{u(T)^2} = \frac{1}{u_h^2} + \frac{2\alpha^2}{z}(T_h - T), \tag{12}$$

where the value of u at $T = T_h$ (u_h) serves as an initial condition. This expression allows $u(T)$ to be determined for any CPM cooler, regardless of temperature drop ($\Delta T \leqslant \Delta T_{\max}$) and applied current density (**j**). The global maximum COP (ϕ) is obtained when the optimum u_h from Eq. (13) is employed,

$$\frac{1}{u_h} = \frac{-\alpha}{z}\frac{zT_c^2 - 2(T_h - T_c)}{T_h + T_c\sqrt{z\left(\frac{T_h+T_c}{2}\right)+1}}. \tag{13}$$

Consideration of Eq. (13) reveals that the maximum T_c is obtained when $1/u_h$ approaches zero. Figure 3 shows $|u|$ becoming infinite at T_h for the CPM cooler. In this limit, Eq. (13) can be simplified to give Eq. (1) with $Z = z$. Thus, a local approach to transport yields the classic CPM limit typically obtained through an evaluation of global transport behavior.

Combining Eqs. (8), (12), and (13) results in $\phi_r(T)$ at $\Delta T_{\max}$ for the CPM Peltier cooler [Eq. (14)]. This expression reveals that ϕ_r drops to zero at both ends of the CPM cooler leg, as shown in Fig. 2(b). This prohibits additional cooling and sets $\Delta T_{\max}$,

$$\phi_{r,\Delta T_{\max}}^{CPM} = \frac{\sqrt{2z(T_h - T)} - 2\frac{T_h-T}{T}}{1 + \sqrt{2z(T_h - T)}}. \tag{14}$$

To achieve cryogenic cooling ($T_c \to 0$) within the CPM, zT must approach infinity [Eq. (1)]. For example, cooling with a single-stage CPM cooler to 10 K would require zT to
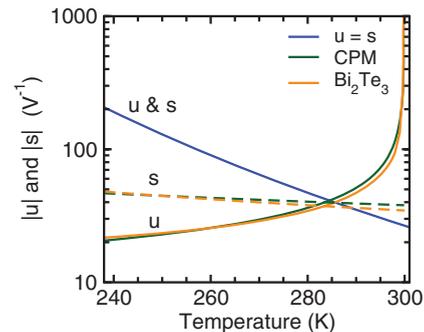

FIG. 3. (Color online) Both the CPM and Bi$_2$Te$_3$ coolers have $u = s$ at one point along the leg. By definition, the $u = s$ is self-compatibile along the entire leg. The different slope signs of u for CPM and $u = s$ reveals that these coolers are fundamentally distinct. The curves were generated for an optimized cooler at $\Delta T_{\max}$ with $z = 0.002$ K^{-1}, $T_h = 300$ K.

be over 1000 if the hot side is 300 K. When ϕ is negative, the net effect of the thermoelectric device is to supply heat, rather than remove heat, from the cold side. For negative ϕ values for the CPM cooler to be obtained requires certain parts of the cooler to locally possess $\phi_r < 0$. Such a result may be surprising at first as this $\phi_r < 0$ region is made from material possessing positive zT. This seems particularly odd when compared to the behavior of staged generators, discussed above. Clearly, single- and multistaged CPM legs exhibit fundamentally different behavior, despite being composed of exactly the same material. Such behavior can be rationalized using the thermoelectric compatibility concept.

Figure 3 shows that the compatibility condition ($u = s$) is maintained at only one point in the CPM cooler. Consequently, CPM coolers operate inefficiently ($u \neq s$) at both the hot and cold ends. This is demonstrated in Fig. 2(b), where $\phi_r < \phi_{r,\max}$ for all but one point. Once ϕ_r goes below zero at low temperature, the thermoelectric device is no longer cooling the cold end and $\Delta T_{\max}$ is reached [Fig. 2(a)].

While real coolers do not possess temperature-independent properties, the qualitative results for CPM translate well to traditional Peltier coolers due to their weak material gradients. Considering a Bi_2Te_3 leg with temperature-dependent properties described in Ref. 24, we find u and s to be quite close to a z-matched CPM cooler (Fig. 3). Like the CPM cooler, $u = s$ at only one temperature along the leg. This leads to similar $\phi_r(T)$ for the Bi_2Te_3 and CPM coolers, shown in Fig. 2(b).

Within the CPM, large zT results in a high upper limit to ϕ_r but does not ensure that this $\phi_{r,\max}$ is achieved. Generally, commercial cooling materials such as Bi_2Te_3 and any material that can be described by the CPM model will be operating significantly below the ϕ_r predicted by the zT they possess [Eqs. (8) and (10)].

Performance of a $u = s$ cooler. We now consider an idealized cooler which maintains $u = s$ across the entire leg. ϕ_r for this cooler is simply given by Eq. (10). This ϕ_r is found to be positive for all T, as z is always a positive real number. Globally, this translates to the analytic maximum for ϕ for a cooler where z is defined and limited.

To facilitate comparison with CPM, we consider a constant z model where the individual properties are adjusted to maintain $u = s$. The constant z approach yields vanishing zT at low T, consistent with real materials. Evaluating ϕ [Eq. (2)] for a $u = s$ cooler and the assumption of constant z, one obtains Eq. (15), where $M_i = \sqrt{1 + zT_i}$ with $T_i = T_h, T_c$,

$$\frac{1}{\phi^{u=s}} = \left(\frac{M_h - 1}{M_c - 1}\right)^2 \exp\left(\frac{2(M_h - M_c)}{(M_h - 1)(M_c - 1)}\right) - 1. \quad (15)$$

Inspection of Eq. (15), where $M_h > M_c > 1$, reveals that ϕ is always greater than zero for a $u = s$ cooler.

The difference between CPM and $u = s$ coolers can be visualized in Fig. 2(a), with the ϕ of the Thomson cooler asymptotically approaching zero with increasing ΔT. Figure 2(b) shows that ϕ_r for a self-compatible cooler with constant z remains finite and positive throughout the device. In contrast, the CPM cooler is operating inefficiently at both the hot and cold ends, limiting its temperature range.

In principle, if $u = s$ can be maintained, the idealized $u = s$ cooler can achieve an arbitrarily low cold side temperature as

long as all of the materials have a finite zT. However, the material requirements to maintain $u = s$ become exceedingly difficult to achieve as the cooling temperature is reduced and the ultimate cooling will be finite, yielding $T_c > 0$.

IV. MATERIAL REQUIREMENTS

A CPM cooler has a fixed z and performance which is independent of the ratio of individual properties as long as they are constant with respect to temperature [Eq. (1)]. In contrast, a $u = s$ cooler requires dramatic changes in properties with temperature to maintain self-compatibility. Within the constraint of constant z, consideration of Eq. (9) suggests that the Seebeck coefficient must be varied across the device to maintain $u = s$. Additionally, as $\alpha(T) = \sqrt{z\rho(T)\kappa(T)}$ within a constant z model, the product $\rho(T)\kappa(T)$ must also vary across the device.

The Seebeck coefficient profile $\alpha(T)$ for a $u = s$ cooler with constant z can be solved analytically. Combining Eq. (7) and $u = s$ yields the simple differential equation of $\alpha(T)$:

$$\frac{d}{dT}\left(\frac{\alpha T}{1 + \sqrt{1 + zT}}\right) = T\frac{d\alpha}{dT} - \frac{\alpha}{z}\frac{1 + \sqrt{1 + zT}}{T}. \quad (16)$$

Solving this equation yields

$$\alpha(T) = \alpha_0 \frac{\sqrt{1 + zT} - 1}{\sqrt{1 + zT}} \exp\left(\frac{-2}{\sqrt{1 + zT} - 1}\right). \quad (17)$$

With this expression for $\alpha(T)$, it is possible to evaluate $s(T)$ with Eq. (9). Figure 3 shows the variation in s required for a $u = s$ cooler with constant z. The self-compatible cooler modeled in Fig. 3 has $z = 0.002$; 60 K of cooling results in a change in s of one order of magnitude.

The approximation for small zT yields a simple expression for $\alpha(T)$, given by Eq. (18),

$$\frac{d}{dT}[\ln \alpha(T)] = \frac{4}{zT^2} \longrightarrow \alpha(T) \propto \exp\left(\frac{-4}{zT}\right). \quad (18)$$

This reveals that α should be very large at the hot end and must decrease to a low value at the cold end. This exponentially varying $\alpha(T)$ required to maintain $u = s$ for constant z is anticipated to be the limiting factor in real coolers and place bounds on the maximum cooling obtainable. We consider the realistic range of α below.

Large values of α are found in lightly doped semiconductors and insulators with large band gaps (E_g) that effectively have only one carrier type, thereby preventing compensated thermopower from two oppositely charged conducting species. Using the relationship between peak α and E_g of Goldsmid [Eq. (19)] allows an estimate for the highest $\alpha(T_h)$ we might expect at the hot end, α_h.[25] Good thermoelectric materials with a band gap of 1 eV are common while 3 eV should be feasible. For a cooler with an ambient hot side temperature, this would suggest that α_h should be ~1–5 mV/K. Maintaining zT at such large α will require materials with both extremely high electronic mobility and low lattice thermal conductivity,

$$\alpha_h = E_g/(2eT_h). \quad (19)$$

A lower bound to α_c also arises from the interconnected nature of the transport properties. We require zT to be finite; thus the electrical conductivity σ must be large as α_c
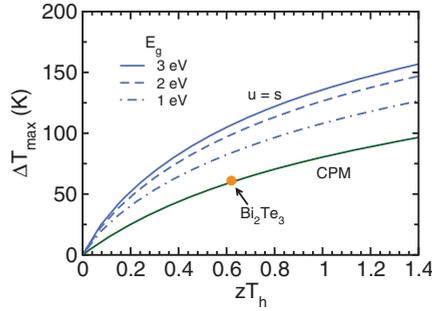
FIG. 4. (Color online) The maximum temperature drop ΔT_{max} of a $u = s$ Thomson cooler exceeds that of a Peltier cooler with the same z. Large band gap (E_g) thermoelectric materials are necessary at the hot junction to improve the performance ($T_h = 300$ K).

tends to zero. In this limit, the electronic component of the thermal conductivity (κ_E) is much larger than the lattice (κ_L) contribution and $\kappa \sim \kappa_E$. To satisfy the Wiedemann-Franz law ($\kappa_E = L\sigma T$ where $L = \frac{\pi^2}{3}\frac{k^2}{e^2}$ is the Lorenz factor in the free electron limit), α_c has a lower bound given by Eq. (20). For example, a $z = \frac{1}{300}$ K^{-1} and $T_c = 175$ K results in a lower bound to α_c of 119 μV/K,

$$\alpha_c^2 = LzT_c = \frac{\pi^2}{3}\frac{k_B^2}{e^2}zT_c. \tag{20}$$

The maximum cooling temperature T_c can be solved as a function of z, E_g, and T_h from Eqs. (17), (19), and (20). For small z the approximate solution

$$\Delta T \approx \frac{z}{8}\,T_h^2\ln\left(\frac{E_g^2}{\frac{4}{3}\pi^2 k_B^2\,z\,T_h^3}\right) \tag{21}$$

gives an indication of the important parameters but quickly becomes inaccurate for zT above 0.1.

Material limits to performance. With these bounds on material properties, we consider the ΔT_{max} of a $u = s$ cooler. Figure 2 suggests that the ϕ of a $u = s$ cooler remains positive for all temperature. However, obtaining materials with the required properties limits ΔT_{max} to a finite value. Figure 4 compares the ΔT_{max} solution for $u = s$ and CPM coolers with the same z. Here, the maximum Seebeck coefficient is set by the band gap ($E_g = 1$–3 eV), per Eq. (19). The $u = s$ cooler provides significantly higher ΔT_{max} than the CPM cooler with the same zT, nearly twice the ΔT_{max} for $E_g = 3$ eV.

Spatial dependence of material properties. These analytic results are possible because the compatibility approach does not require an exact knowledge of the spatial profile for the material properties. Nevertheless, it is possible to solve for the spatial dependence of the $u = s$ cooler, given some material constraints. To determine $x(T)$, we integrate Eq. (4), recalling that we have assumed constant cross-sectional area [$j(x) = $ const.], obtaining Eq. (22),

$$x(T) = \frac{-1}{j}\int_T^{T_h} u\kappa\, dT. \tag{22}$$

Thus, the natural approach to cooler design within the $u = s$ approach is to determine the temperature dependence of the material properties, and then determine the required spatial dependence from the resulting $u(T)$ and $\kappa(T)$.

Figure 5(a) shows an example of the Seebeck distribution $\alpha(x)$ along the leg that will provide the necessary $\alpha(T)$, where a constant $\kappa_L = 0.5$ W/mK is assumed. The α of Fig. 5(a) spans the range permitted by Eqs. (19) and (20).

In a real device the spatial profile of thermoelectric properties will need to be carefully engineered. If this rapidly changing $\alpha(x)$ is achieved by segmenting different materials, low electrical contact resistance is required between the interfaces. We anticipate that such control of semiconductor materials may require thin-film methods on active bulk thermoelectric substrates.

V. COOLER PHASE SPACE

The improved performance of a $u = s$ cooler is not simply an incremental improvement, but rather we find that CPM and $u = s$ coolers operate in fundamentally different phase spaces. Here, by phase space we refer to the class of solutions defined by the sign of the Fourier heat divergence [$\nabla \cdot \mathbf{q}_\kappa$ in Eq. (3)]. The Fourier heat divergence in a cooler contains both the Joule (ρj^2) and Thomson ($\tau \mathbf{j} \cdot \nabla T$) terms.

We begin by considering the Fourier heat divergence in CPM and $\mathrm{Bi_2Te_3}$ coolers and then compare this behavior to $u = s$ coolers. In the typical CPM model to analyze Peltier coolers, $\tau = 0$ as there is no variation in α. In a CPM cooler, $\nabla \cdot \mathbf{q}_\kappa$ is thus greater than zero. This can be seen by the downward concavity of the temperature distribution in Fig. 5(b). In a
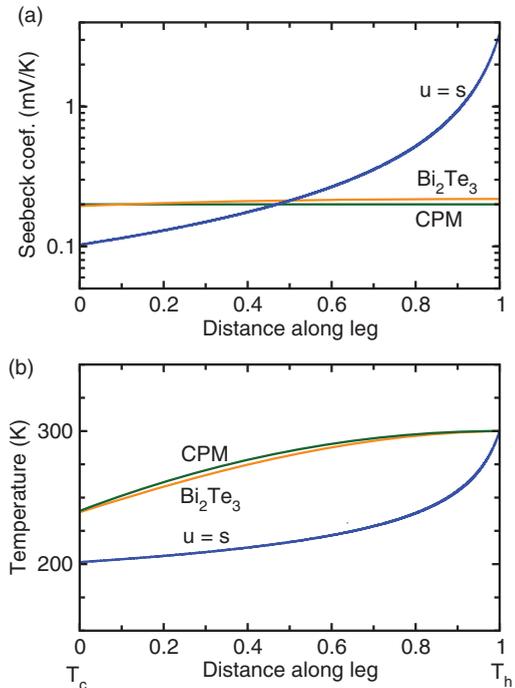

FIG. 5. (Color online) (a) The Seebeck coefficient of the $u = s$ Thomson cooler varies exponentially, while it is by definition constant for the CPM cooler. As a degenerately doped semiconductor, the Seebeck coefficient of commercial $\mathrm{Bi_2Te_3}$ increases gradually with temperature. (b) The curvature of $T(x)$ for the CPM Peltier cooler temperature profile is opposite that of the $u = s$ cooler because of the different sign of the Fourier heat divergence. Again, similar behavior is found between the CPM and commercial $\mathrm{Bi_2Te_3}$ cooler.

typical Peltier cooler (e.g., Bi_2Te_3), the concavity is the same as the CPM cooler and thus the divergence is likewise positive. This is because the Thomson term is always less than the Joule term in a conventional thermoelectric cooler.

In contrast, a $u = s$ cooler changes the sign of the Fourier heat divergence such that $\nabla \cdot \mathbf{q}_\kappa$ is less than zero. This can be readily visualized in Fig. 5(b), where the concavity of the $u = s$ cooler is opposite the CPM and Bi_2Te_3 coolers. This difference in concavity must come from the Thomson term being positive and greater than the Joule heating term. The large magnitude of Thomson term is understandable with the exponentially rising Seebeck coefficient seen in Fig. 5(a). The reversibility of the Thomson effect requires that for $\nabla \cdot \mathbf{q}_\kappa$ to be less than zero, the hot end must have a high $|\alpha|$ relative to the cold end, and not vice versa. This translates to a requirement for τ such that $\tau \mathbf{j} \cdot \nabla T > \rho j^2$.

We can also express the Fourier heat divergence in terms of reduced variables,

$$\nabla \cdot \mathbf{q}_\kappa = \mathbf{j} \cdot \nabla \frac{1}{u} = \frac{-1}{u^2} \frac{du}{dT} \mathbf{j} \cdot \nabla T. \tag{23}$$

Manipulation with Eq. (4) produces a form where the sign of u and directions of j and ∇T are irrelevant,

$$\nabla \cdot \mathbf{q}_\kappa = \frac{j^2}{2\kappa u^4} \frac{d}{dT} u^2. \tag{24}$$

Thus the sign of $\nabla \cdot \mathbf{q}_\kappa$ is determined by the sign of $\frac{d|u|}{dT}$, which is valid for both p- and n-type elements regardless of the sign of u.

The Fourier heat divergence criterion is a convenient definition to distinguish these two regions of thermoelectric cooling in experimental data. The Peltier cooling region, defined by $\nabla \cdot \mathbf{q}_\kappa > 0$, is found in the phase space where $\frac{d|u|}{dT} > 0$. Likewise, the Thomson cooling region defined by $\nabla \cdot \mathbf{q}_\kappa < 0$ is the phase space where $\frac{d|u|}{dT} < 0$. The constant relative current $u(T) = \text{const.}$ separates the Thomson-type and from the Peltier-type solutions to the differential equation. In Fig. 3, the CPM and $u = s$ cooler have opposite slopes, indicating that these coolers exist in separate regions of the cooling phase space. This result is consistent with our discussion above concerning the concavity of $T(x)$ in Fig. 5.

For clarity, we suggest that coolers which are predominately in the Thomson phase space, $\nabla \cdot \mathbf{q}_\kappa < 0$, but may not have $u = s$ be referred to as "Thomson coolers." Similarly, "Peltier coolers" should refer to coolers operating in the usual $\nabla \cdot \mathbf{q}_\kappa > 0$ Fourier heat divergence phase space where Joule heating dominates.

This understanding of phase space for $u = s$ and CPM coolers enables us to hypothesize that the performance advantages of $u = s$ coolers extends to imperfect Thomson coolers. We expect that such coolers possess two primary advantages over traditional Peltier coolers. First, for a given material zT, performance (ΔT_{max} and ϕ) of the Thomson cooler is greater (Fig. 2). The ΔT_{max} solution for the $u = s$ cooler is compared to a Peltier cooler with the same material assumption for z in Fig. 4. Here, the maximum Seebeck coefficient is set by the band gap ($E_g = 1$–3 eV), per Eq. (19). The Thomson cooler provides significantly higher ΔT_{max} than the Peltier cooler with the same zT, nearly twice the ΔT_{max} for $E_g = 3$ eV. Second,

in a Thomson cooler, the temperature minimum is not limited by zT explicitly like it is in traditional Peltier coolers.

VI. DISCUSSION

Efficiency improvements from staging and maintaining $u = s$ also exists for thermoelectric generators, but the improvement is small (<10% compared to CPM). This is because the u does not typically vary by more than a factor of 2 across the device. However, in a TEC the compatibility requirement is much more critical. When operating a TEC to maximum temperature difference, the temperature gradient varies from zero to very high values, which means u will have a much broader range (Fig. 3) in a TEC than in a generator. Thus, unless compatibility is specifically considered, the poor compatibility will greatly reduce the performance of the thermoelectric cooler, and this results in the ΔT_{max} limit well known for Peltier coolers.

In real materials, changing material composition also changes zT so the effect of maximizing average zT is difficult to decouple from the effect of compatibility. As such, efforts which are focused on maximizing zT will generally fail to create a material with $u = s$ and may only marginally increase ΔT_{max}. Conversely, focusing on $u = s$ without consideration of zT could rapidly lead to unrealistic materials requirements.

In this new analysis we have focused on the compatibility criterion, $u = s$, with constant z [as opposed to zT (Ref. 16)] to demonstrate the differences between a Thomson and a Peltier cooler typically analyzed with the CPM model. Generally, achieving $u = s$ in a material with finite zT is more important to achieve low-temperature cooling than increasing zT.

Minor improvements in thermoelectric cooling beyond increasing average zT by increasing the Thomson effect in a functionally graded material were predicted as early as 1960.[7] Similarly Müller *et al.* describe modest gains in cooling from functionally grading[8–10] where material properties are allowed to vary in a constrained way such that the average zT remains constant. Such additional constraints can keep the analysis within the Peltier region, preventing a full optimization to a $u = s$ solution.

Bian and co-workers[11,12] propose a thermoelectric cooler with significantly enhanced ΔT_{max} using a rapidly changing Seebeck coefficient in at least one region. The method of Bian and co-workers focuses on the redistribution of the Joule heat rather than a consideration of the Thomson heat or the effect of compatibility. Nevertheless, the region of rapidly changing Seebeck coefficient would also create a significant Thomson effect and likely place that segment of the cooler in the Thomson cooler phase space while other segments would function like a CPM Peltier cooler.

In a traditional single-stage (or segmented) thermoelectric device, the current flow and the heat flow are collinear and flow through the same length and cross-sectional area of thermoelement. This leads to the compatibility requirement between the optimal current density and optimal heat flux to achieve optimal efficiency. In a multistage (cascaded) device the thermal and electrical circuits become independent and so the compatibility requirement is avoided between stages.

A transverse Peltier cooler also decouples the current and heat flow by having them transport in perpendicular directions.

Gudkin showed theoretically that the transverse thermoelectric cooler could function as an infinite cascade by the appropriate geometrical shaping of the thermoelement.[26] The different directions of heat and current flow enable, in principle, an adjustment of geometry to keep both heat and electric current flow independently optimized. Cooling of 23 K using a rectangular block was increased to 35 K using a trapezoidal cross section.[27]

VII. CONCLUSION

Here, we compare self-compatible coolers with CPM and commercial Bi_2Te_3 thermoelectric coolers. Significant improvements in cooling efficiency and maximum cooling are achieved for equivalent z when the cooler is self-compatible. Such improvement is most pronounced when the goal is to achieve maximum temperature difference, rather than high coefficient of performance at small temperature difference. Optimum material profiles are derived for self-compatible Thomson coolers and realistic material constraints are used to bound the performance. Self-compatible coolers are found to operate in a fundamentally distinct phase space from traditional Peltier coolers. The Fourier heat divergence of Thomson coolers is dominated by the Thomson effect, while this divergence in Peltier coolers is of the opposite sign, indicating the Joule heating is the dominant effect. This analysis opens a new strategy for solid-state cooling and creates new challenges for material optimization.

APPENDIX

The metric for summing the efficiency and coefficient of performance of these thermodynamic processes is not a simple summation because energy is continuously being supplied or removed so that neither the heat nor energy flow is a constant. The derivation, attributed to Zener,[17] of the coefficient of the performance Eq. (2) and efficiency summation metric for a continuous system in one dimension given here is based on Zener and similar derivations given in Refs. 7, 28, 29, and 19.

Consider n heat pumps (or heat engines) connected in series such that the heat entering the ith pump, Q_i, is the same as the heat exiting the $i - 1$ pump, namely Q_{i-1}. Then by conservation of energy

$$Q_i = Q_{i-1} + P_i, \tag{A1}$$

where P_i is the power entering the pump i. The coefficient of performance ϕ_n of the pump n is defined by

$$\frac{1}{\phi_i} = \frac{P_i}{Q_{i-1}}. \tag{A2}$$

This gives the recursive relation

$$Q_i = Q_{i-1}\left(1 + \frac{1}{\phi_i}\right), \tag{A3}$$

which can be solved for Q_n as

$$Q_n = Q_0 \prod_{i=1}^{n}\left(1 + \frac{1}{\phi_i}\right). \tag{A4}$$

The total power added to the system P from the n pumps is

$$P = \sum_{i=1}^{n} P_i. \tag{A5}$$

The coefficient of performance for the n pumps, ϕ, with heat $Q_c = Q_0$ entering from the cold side is defined by

$$\frac{1}{\phi} = \frac{P}{Q_0} \tag{A6}$$

By conservation of energy [or recursive relation Eq. (A1)], the heat exiting the n pumps, Q_n, is the heat pumped by the first pump plus the total power, Eq. (A5),

$$Q_n = Q_0 + P. \tag{A7}$$

Combining Eqs. (A6), (A7), and (A4), it is straightforward to show that

$$\left(1 + \frac{1}{\phi}\right) = \prod_{i=1}^{n}\left(1 + \frac{1}{\phi_i}\right). \tag{A8}$$

This can be transformed into a summation by use of a natural logarithm,

$$\ln\left(1 + \frac{1}{\phi}\right) = \sum_{i=1}^{n} \ln\left(1 + \frac{1}{\phi_i}\right). \tag{A9}$$

While $1/\phi_i$ should become very small with small $\delta T_i = T_i - T_{i-1}$ the reduced coefficient of performance $\phi_{r,i}$ should remain finite,

$$\frac{1}{\phi_i} = \frac{1}{\phi_{r,i}}\frac{\delta T_i}{T_i}. \tag{A10}$$

Assuming a monotonic temperature distribution (for a simple TEC $T_i > T_{i-1}$ for all i) the sum in Eq. (A9) can be converted to an integral in the limit that $n \rightarrow \infty$, where $\delta T_i = (T_h - T_c)/n$ and $\ln(1 + x) \rightarrow x$ for small x,

$$\ln\left(1 + \frac{1}{\phi}\right) = \int_{T_c}^{T_h} \frac{1}{T\phi_r(T)}dT. \tag{A11}$$

If the temperature distribution is not monotonic but can be divided up into monotonic segments, each of these monotonic segments can be individually transformed into integrals.

For a generator (as opposed to a cooler or heat pump) power is extracted rather than added at each segment in the series. Then the sign of the power in Eq. (A1) is negative for a generator with the efficiency given by $\eta = -P/Q_0 = -1/\phi$ and reduced efficiency $\eta_r = -1/\phi_r$. Thus the above method can be used to derive the analogous equation for generator efficiency:

$$\ln(1 - \eta) = -\int_{T_c}^{T_h} \frac{\eta_r}{T}dT. \tag{A12}$$

*jsnyder@caltech.edu

[1]H. J. Goldsmid, *Introduction to Thermoelectricity* (Springer-Verlag, Berlin, 2010).

[2]R. R. Heikes and R. W. Ure, *Thermoelectricity: Science and Engineering* (Interscience, New York, 1961).

[3]Marlow Industries, Inc., Six stage–thermoelectric modules, [http://www.marlow.com, 2011].

[4]Y. Gelbstein, Z. Dashevsky, and M. P. Dariel, Physica B **363**, 196 (2005).

[5]V. Kuznetsov, in *CRC Handbook of Thermoelectrics*, Chap. 38 (CRC, Boca Raton, 2006).

[6]G. D. Mahan, J. Appl. Phys. **70**, 4551 (1991).

[7]B. Sherman, R. R. Heikes, and R. W. Ure, J. Appl. Phys. **31**, 1 (1960).

[8]E. Müller, Č. Drašar, J. Schilz, and W. A. Kaysser, Mater. Sci. Eng. A **362**, 17 (2003).

[9]E. Müller, S. Walczak, and W. Seifert, Phys. Status Solidi A **203**, 2128 (2006).

[10]E. Müller, G. Karpinski, L. M. Wu, S. Walczak, and W. Seifert, in *25th International Conference on Thermoelectrics*, edited by P. Rogl (IEEE, Piscataway, 2006), pp. 204–209.

[11]Z. Bian and A. Shakouri, Appl. Phys. Lett. **89**, 212101 (2006).

[12]Z. Bian, H. Wang, Q. Zhou, and A. Shakouri, Phys. Rev. B **75**, 245208 (2007).

[13]B. Moizhes, Y. P. Shishkin, A. Petrov, and L. A. Kolomoet, Sov. Phys. Tech. Phys. **7**, 336 (1962).

[14]G. J. Snyder Appl. Phys. Lett. **84**, 2436 (2004).

[15]G. J. Snyder in *CRC Handbook of Thermoelectrics* (Ref. 5), Chap. 9.

[16]W. Seifert *et al*., J. Mater. Res. **26**, 1933 (2011).

[17]C. Zener, in *Thermoelectricity*, edited by P. Egli, Chap. 1 (Wiley, New York, 1960), pp. 3–22.

[18]G. J. Snyder and T. S. Ursell, Phys. Rev. Lett. **91**, 148301 (2003).

[19]W. Seifert, E. Müller, and S. Walczak, in *25th International Conference on Thermoelectrics* (Ref. 10), pp. 714–719.

[20]D. T. Crane and L. E. Bell, J. Energy Res. Technol. **131**, 012401 (2009).

[21]E. G. Wood, R. C. Ewell, J. Patel, D. R. Hanks, J. A. Lozano, G. J. Snyder, L. Noon, NASA Tech Briefs **45252**, 51 (2008).

[22]D. T. Crane, J. Electron. Mater. **40**, 561 (2011).

[23]W. Seifert, K. Zabrocki, G. Snyder, and E. Müller, Phys. Status Solidi A **207**, 760 (2010).

[24]W. Seifert, M. Ueltzen, and E. Müller, Phys. Status Solidi A **194**, 277 (2002).

[25]H. J. Goldsmid and J. W. Sharp, J. Electron. Mater. **28**, 869 (1999).

[26]T. S. Gudkin, E. K. Iordanishviliu, and E. E. Fiskind, Sov. Phys. Semicond. **11**, 1048 (1977).

[27]T. S. Gudkin, E. K. Iordanishviliu, and E. E. Fiskind, Sov. Phys. Tech. Phys. Lett. **4**, 844 (1978).

[28]S. I. Freedman in *Direct Energy Conversion*, Inter-University Electronic Series Vol. 3 (McGraw-Hill, New York, 1966), Chap. 3, pp. 105–180.

[29]T. C. Harman and J. M. Honig, *Thermoelectric and Thermomagnetic Effects and Applications* (McGraw-Hill, New York, 1967).