REPORT

THERMOELECTRIC EFFECTS IN SUBMICRON HETEROSTRUCTURE BARRIERS

Ali Shakouri, Edwin Y. Lee, D. L. Smith, Venky Narayanamurti, and John E. Bowers

Department of Electrical and Computer Engineering, University of California, Santa Barbara, California, USA

Nonisothermal electron transport in thin barrier heterostructures is investigated using Monte Carlo techniques. Particular attention is paid to the energy balance in thermionic emission, and the Joule heating in the barrier region. By introducing an energy relaxation length, an equation for the temperature distribution inside the device is derived. Conditions for creating a steady-state temperature gradient and for integrated cooling of electronic components are examined.

In today's integrated circuits, the need for higher density of devices and higher speeds of operation make the issue of heating and thermal management of prime interest. Optoelectronic devices, such as distributed feedback and vertical cavity surface emitting lasers, are of increasing use in communication systems. In these devices a high density of heat, on the order of kW/cm^2, is generated over very small areas ($100-5{,}000\ \mu m^2$). In addition, the high temperature sensitivity of the device characteristics (such as threshold current, power output, and wavelength) requires means for active temperature control. Thermoelectric (TE) coolers are typically used for temperature stabilization of these high-power and highly sensitive devices. Other components such as HgCdTe-based infrared imaging arrays work best at low temperatures (< 200 K) and require multiple-stage TE coolers. These coolers are based on the Peltier effect. In the 1950s and early 1960s, extensive research was done studying various material systems for TE cooling applications. Almost all of the current commercial room-temperature TE coolers

Received 20 August 1997; accepted 29 August 1997.

This work was supported by DARPA and the Office of Naval Research under contract 442530-25845. The work of D. L. Smith was done at the University of California, Santa Barbara, while on the Los Alamos Visiting Scholar Program, which was conducted under the auspices of the U.S. Department of Energy, supported in part by funds provided by the University of California for the conduct of discretionary research by Los Alamos National Laboratory.

Edwin Y. Lee's present address is Sandia National Laboratories, Livermore, California, USA.

D. L. Smith's permanent address is Los Alamos National Laboratory, Los Alamos, New Mexico, USA.

Address correspondence to Ali Shakouri, ECE Department, UCSB, Santa Barbara, CA 93106, USA. E-mail: ali@opto.ucsb.edu

Microscale Thermophysical Engineering, 2:37–47, 1998
Copyright © 1998 Taylor & Francis
1089-3954/98 $12.00 + .00

are made from the material that emerged at that time, Bi_2Te_3. Improvements in efficiency and cooling power, reduction in size, and development of integrated circuit fabrication techniques [1] are needed. In this article we first review the microscopic origin of the Peltier effect based on linear transport theory. This gives an insight that is then used to look at the prospects of heterostructures for cooling applications. As electron transport in thin barrier heterostructure devices goes beyond the linear transport regime, we use some results from Monte Carlo simulations to obtain an approximate expression for the cooling power of these devices. It is shown that with available high-mobility and low-thermal-conductivity materials, it is possible to create a large temperature gradient over small distances and thus to integrate the cooling elements with high-power electronic components.

The conventional thermoelectric effect is based on *bulk* properties of materials. When electrons flow from a material in which they have an average transport energy smaller than the Fermi energy to another material in which their average transport energy is higher, they absorb thermal energy from the lattice and this cools down the junction between the two materials. Reversing the direction of current instead generates heat and creates a hot junction. At low electric fields, electronic conduction is described using the linearized Boltzmann transport equation. The expressions of the electrical conductivity and the Seebeck coefficient (thermopower) can be written as

$$\sigma = \frac{e^2}{4\pi^3}\iiint \tau(k)v_x^2(k)\left(-\frac{\partial f_{\text{eq}}}{\partial E}\right)d^3k \equiv \int \sigma(E)\left(-\frac{\partial f_{\text{eq}}}{\partial E}\right)dE \tag{1}$$

$$S = \frac{1}{eT}\frac{\iiint \tau(k)v_x^2(k)[E(k)-E_F](-\partial f_{\text{eq}}/\partial E)\,d^3k}{\iiint \tau(k)v_x^2(k)(-\partial f_{\text{eq}}/\partial E)\,d^3k}$$

$$\equiv \frac{k_B}{e}\frac{\int \sigma(E)[(E-E_F)/k_BT](-\partial f_{\text{eq}}/\partial E)\,dE}{\int \sigma(E)(-\partial f_{\text{eq}}/\partial E)\,dE} \propto \langle E-E_f\rangle \tag{2}$$

where we introduce the "differential" conductivity,

$$\sigma(E) \equiv e^2\tau(E)\iint v_x^2(E,k_y,k_z)\,dk_y\,dk_z \cong e^2\tau(E)\bar{v}_x^2(E)\bar{n}(E) \tag{3}$$

Here $\tau(E)$ is the energy-dependent relaxation time, $\bar{v}_x(E)$ is the average velocity of the carriers with an energy between E and $E+dE$ in the direction of current flow, and $\bar{n}(E)$ is the number of electrons in this energy interval. From Eq. (1) it can be seen that electrical conductivity is the sum of the contribution of electrons with various energies E [given by $\sigma(E)$, the differential conductivity] within the Fermi window factor $\partial f_{\text{eq}}/\partial E$. The Fermi window is a direct consequence of the Pauli exclusion principle; at finite temperatures, only electrons near the Fermi surface contribute to the conduction process. In this picture, thermopower is the average energy transported by the charged carriers [see Eq. (2)]. This so-called diffusion thermopower may be enhanced by the coupling of electronic motion to other means of transport of energy (e.g., by phonons). For a typical thermoelectric

cooling application, current flows from a metal to a semiconductor and then to a metal again. This creates a cold junction at the first metal–semiconductor junction and heats the second junction (or vice versa). The overall performance of the devices can be expressed as a function of the dimensionless figure of merit ZT:

$$ZT = \frac{S^2 \sigma T}{\beta} \tag{4}$$

which describes the interplay between the Peltier cooling at the junctions (given by the Seebeck coefficient), Joule heating in the semiconductor (given by electrical conductivity), and heat conduction from the hot to the cold junction (β). To increase thermoelectric cooling efficiency and capacity, one has to maximize the Z factor. Since in most semiconductors the coefficient of thermal conductivity at room temperature is dominated by the lattice contribution, maximizing Z requires maximizing the electrical power product $S^2\sigma \approx |\langle E - E_f \rangle|^2 \sigma$. This means that the differential conductivity, $\sigma(E)$, within the Fermi window should be as big as possible and at the same time as asymmetric as possible with respect to the Fermi energy.

Using the above interpretation, the advantage of going to lower-dimension semiconductor structures becomes clear. When the dimensionality is reduced, the density of states (DOS) "accumulates" near the subband edges. With proper doping, the asymmetry of $\sigma(E)$ with respect to the Fermi energy can thus be increased. Recent literature on quantum well and wire thermoelectrics [2–7] emphasizes the increased density of states, but the symmetry is not mentioned and its consequences are buried in the calculations of optimum doping in these structures. The symmetry of $\sigma(E)$ is the main cause of low thermopower in metals, even though they have a very large DOS.

For practical cooling applications, the advantages of using lower-dimensional semiconductors is mitigated by the thermal conductivity of inactive barrier layers and other nonideal effects [5–7]. Using bandgap engineering and heterostructures one can modify not only DOS but also electron velocity $\overline{v}_x(E)$ and relaxation times $\tau(E)$. Based on these ideas, electron transport perpendicular to the quantum well layers was proposed to reduce the mobility of "cold" electrons and to increase the thermopower [8, 9]. These methods, however, show only a modest improvement of the thermoelectric figure of merit. All of these concepts and primary calculations are based on the linearized Boltzmann transport equation, which is valid in the band conduction regime and when the electronic distribution function is not considerably changed with respect to the Fermi distribution.

The application of heterostructures for thermoelectric cooling goes beyond the linear transport regime. Using conduction or valence band offsets at heterostructures, one can control the energy distribution of emitted electrons precisely. For the single heterostructure barrier of Figure 1, under an applied bias, hot electrons are emitted from the cathode side, and the reverse current is suppressed by a larger anode conduction band offset. By minimizing the amount of Joule heating in the barrier and heat conduction from the hot to the cold side, one can

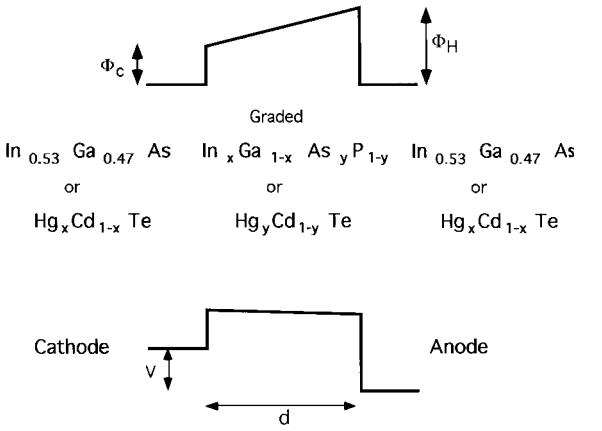


Figure 1. Conduction band diagram of heterostructure integrated thermionic cooling.

create a steady-state temperature gradient and cool down the cathode junction. High-precision epitaxial growth techniques allow the design of the optimum cathode and anode barrier heights in a wide range (0 to 0.4 eV). Depending on the growth constraints and lattice mismatch between materials, one can grade the barrier composition to produce internal fields and to enhance electron transport properties. The problem of space-charge limited current can be controlled by modulation doping or bandgap engineering in the barrier region.

In order to analyze more quantitatively the prospect of thermionic emission in heterostructures for cooling applications, we study the energy balance equation. First, let us look at a conventional TE cooler. The cooling capacity per unit area at the cold junction, Q_{TE}, is given by the Peltier effect ($Q_P = ST_C I$, where S is the Seebeck coefficient, equal to 200 μV/K for Bi_2Te_3), minus the Joule heating generated in the branch ($Q_{J,\mathrm{TE}} = \frac{1}{2}(d/\sigma)I^2$, where d is the thickness of the material and σ is the electrical conductivity, which is equal to 10^5 $\mathrm{m}^{-1}\Omega^{-1}$ for Bi_2Te_3 at dopings typically used for thermoelectric cooling applications), and minus the amount of heat conduction from the hot to the cold junction [$Q_C = (\beta\Delta T)/d$, where β, the thermal conductivity of Bi_2Te_3, is 1–2 W/m K, and ΔT is the temperature difference between the hot and the cold junctions]. Thus:

$$Q_{\mathrm{TE}} = ST_C I - \frac{1}{2}\frac{d}{\sigma}I^2 - \frac{\beta}{d}\Delta T \tag{5}$$

To estimate cooling by thermionic emission, one should calculate the total energy of electrons emitted. As the energy distribution of these electrons is almost exclusively on one side of the Fermi energy, upon current flow, strong carrier–carrier and carrier–lattice scatterings tend to restore the quasi-equilibrium Fermi distribution in the cathode by absorbing energy from the lattice, and thus cooling this junction. Using Boltzmann statistics (which are valid for barrier heights greater than a few $k_B T$, which correspond to currents of about 10^5 A/cm^2 at room

temperature), we can calculate the average transport energy of carriers to be $\phi_C + (2k_B T_C/e)$, where ϕ_C is the cathode barrier height. The cooling power is thus

$$Q_{TI-\text{cooling}} = \left(\Phi_C + 2\frac{k_B T}{e}\right) I \tag{6}$$

The emitted electrons lose their energy and generate heat in the barrier region and at the anode side. To find the net amount of cooling we need to calculate how much of the Joule heating in the barrier is transported back to the cathode, and what is the amount of heat conduction from the hot to the cold junction. As a first-order estimate, we assumed voltage drop over the barrier to be the minimum required by the Bethe criterium for Richardson's thermionic emission current to be valid and that half of the Joule heating goes back to the cold side [10]. In order to minimize heating effects, these preliminary calculations suggested that the barrier region should be very thin, on the order of microns. Clearly, for these device dimensions, the assumptions for pure diffusive transport are hard to justify and a more accurate analysis should take into account the finite energy relaxation length for carriers. Figure 2 shows the result of Monte Carlo simulations for carriers injected into a 3-μm-thick GaAs barrier, under an applied bias of 5 kV/cm. It is assumed that the cathode barrier height is 100 meV. It can be seen that it takes a distance of the order of 1 μm before electrons reach equilibrium and lose energy equal to what they gain at steady state from the external electric field. Fitting the result by an exponential expression that includes

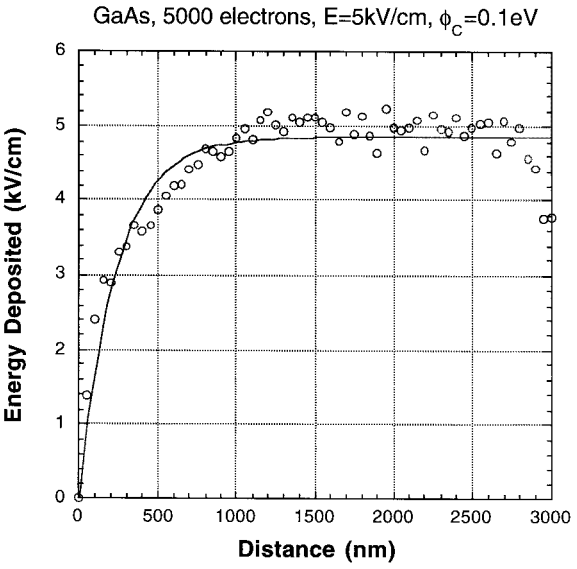


Figure 2. Joule heating generated in the barrier as a function of distance.

a bias-dependent energy relaxation length, λ_E, we get

$$\frac{dQ_{\text{Joule}}}{dx} = I\frac{V}{d}(1 - e^{-x/\lambda_E}) \tag{7}$$

Here, for simplicity, we did not include a heat transport equation for various phonons involved in the energy relaxation mechanism. It is assumed that all of the energy lost by the electronic system is a source for local heat generation. One can obtain an analytic solution for the temperature distribution equation,

$$\beta\frac{d^2T}{dx^2} = -I\frac{V}{d}(1 - e^{-x/\lambda_E}) \tag{8}$$

with the following boundary conditions:

$$\begin{cases} \beta\dfrac{dT}{dx}\Big|_{x=0} = \left(\phi_C + \dfrac{2k_BT_C}{e}\right)I - Q_{TI} \\ T|_{x=d} = T_H \end{cases} \tag{9}$$

These boundaries specify the thermionic cooling minus a possible heat load, Q_{TI}, on one side of the device and a heat sink with temperature T_H on the other side. The expression for overall cooling power is thus

$$Q_{TI} = \left(\Phi_C + 2\frac{k_BT_C}{e}\right)I - IV\left[\left(\frac{1}{2} - \frac{\lambda_E}{d}\right) - \frac{\lambda_E^2}{d^2}(e^{-d/\lambda_E} - 1)\right] - \frac{\beta}{d}\Delta T \tag{10}$$

This equation is similar to the conventional thermoelectric case. It has three terms: thermionic cooling, "Joule" heating, and heat conduction. The Joule heating term is IV (total voltage drop over the barrier times the current) times a coefficient which takes into account the finite electronic energy relaxation length. Plotting this coefficient in Figure 3 shows that in the limit of very thick devices ($d \gg \lambda_E$), it reduces to $\frac{1}{2}$, which is the result for pure diffusive transport. In the other limit of very short devices ($d \ll \lambda_E$), this term gives zero contribution. In this case of ballistic transport, all of the electron's energy is deposited at the anode side.

In the above expression, the coefficient of thermal conductivity to use is not necessarily the sum of lattice and electronic contributions (as is the case for thermoelectrics). The net electric current is from the cold to the hot junction, and only in a pure diffusive conduction, where the electrons have many collisions, can electronic and lattice thermal conductivities be added independently of the net current. Again in the limit of ballistic transport from cathode to anode, there is no contribution of electronic thermal conductivity to the heat flow from anode to cathode (if the reverse current is suppressed by an appropriate band edge discontinuity). As the thermal conductivity of most semiconductors is dominated by the lattice contribution, we ignore the electronic component at this time. The lattice thermal conductivity of a thin film is not equal to the bulk value. There are quite a

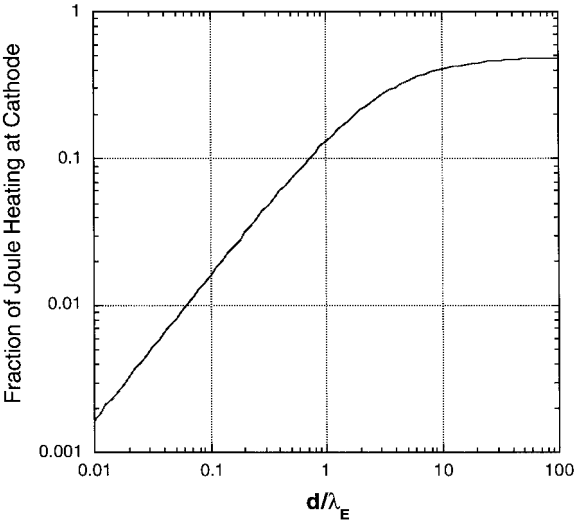


Figure 3. Fraction of the total Joule heating generated in the barrier that is transferred to the cathode, as a function of the barrier thickness over energy relaxation length for carriers. The anode side is assumed to be in contact with a heat sink.

few excellent recent studies of thin-film and superlattice thermal conductivities [11], but in the absence of a comprehensive theory we will use the bulk values here. In fact, from the steady-state temperature gradient in a thin barrier under an applied bias, one can calculate its thermal conductivity based on the electronic transport parameters. This method is already used for thermoelectric materials and is known as the Harman method [12].

Finally, we should consider the bias dependence of the thermionic emission current. Richardson's expression is valid only at high electric fields, when the transport of the carriers over the barrier is not a limiting factor. From more elaborate calculations combining thermionic emission and drift-diffusion theories [13], one can approximately write the overall effect of these two limiting processes ("supply'' and "transport") as follows:

$$\frac{1}{I} \approx \frac{1}{I_{\text{thermionic}}} + \frac{1}{I_{\text{drift-diffusion}}} \tag{11}$$

$$I_{\text{drift-diffusion}} \approx en_0\mu E + eD\frac{n_0 - 0}{d} = en_0\mu\left(E + \frac{k_B\overline{T}}{ed}\right) \tag{12}$$

$$\frac{I}{I_{\text{thermionic}}} \approx \frac{\mu(2m_{\text{eff}}\pi/k_BT_C)^{1/2}\left(E + k_B\overline{T}/ed\right)}{1 + \mu(2m_{\text{eff}}\pi/k_BT_C)^{1/2}\left(E + k_B\overline{T}/ed\right)} \tag{13}$$

where μ is the mobility of the barrier layer and m_{eff} is the effective mass that enters in the thermionic emission expression. Grinberg [14] used arguments based

on the conservation of lateral momentum in thermionic emission to show that the correct effective mass to use is the *minimum* of emitter and barrier regions. The simple estimate [Eq. (13)] and the Monte Carlo calculation are shown in Figure 4. Without any fitting parameters, using a mobility of 8,000 $cm^2/V\,s$ and $m_{eff}/m_0 = 0.067$, the two results match quite well. One can get a better fit of the Monte Carlo result by using a slightly higher mobility value ($\sim$ 10,400 $cm^2/V\,s$).

In order to calculate thermionic cooling power accurately, one should perform Monte Carlo simulations for each material system under study. After deriving parameters such as bias-dependent energy relaxation length, the above equations can be used to find optimum device design (barrier thickness, cathode and anode conduction band offsets, etc.). To get an idea of the net amount of cooling that can be achieved using this method, we consider the case of an InGaAs barrier material lattice matched to InP. This material is a good candidate for thermionic cooling because of its low thermal conductivity, high electron mobility, and not too small electron effective mass. InGaAs is a III–V compound similar to GaAs, so the major energy relaxation mechanisms are expected to be similar in nature [15]. Using various material parameters of InGaAs, and taking 0.4 μm for the energy relaxation length, we can calculate the net cooling power at the cold side as a function of the thickness of the barrier for different cathode barrier heights (Figure 5*a*). The anode barrier height does not enter directly into the calculation of maximum cooling power. It is assumed to be high enough to suppress the reverse

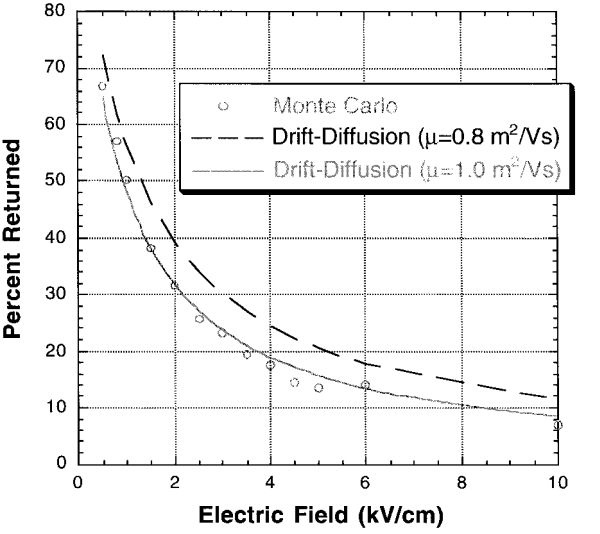
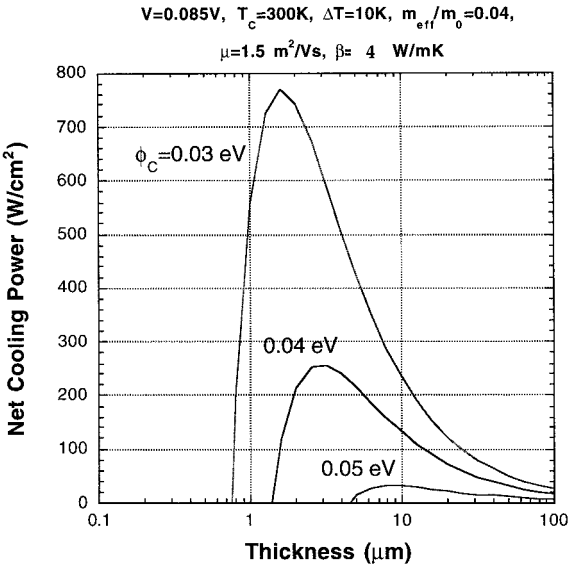
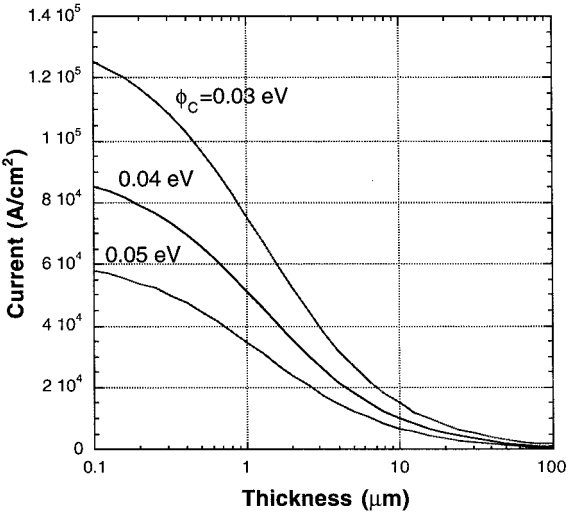


Figure 4. Fraction of the carriers returning back to the cathode after being emitted (Monte Carlo results: circles). This gives the deviation of the total current from Richardson's expression. A simplified drift-diffusion model, which assumes that $m_{eff}/m_0 = 0.067$ and $\mu = 8{,}000\ cm^2/V\,s$, is shown by the dashed line. The solid line is the fit of Monte Carlo result using the drift-diffusion model with a mobility of 10,400 $cm^2/V\,s$.



(*a*)



(*b*)

Figure 5. (*a*) The net cooling power of a HIT cooler made of InGaAs as a function of the thickness of the barrier layer for different cathode-side conduction band offsets. (*b*) The total current as a function of the same parameters.

current form the hot to the cold side. The anode barrier, however, affects the efficiency of the device directly. It can be seen that at room temperature, thermionic cooling can maintain a temperature gradient of 10° over a distance of 2 μm and provide a net cooling power of couple of 100 W/cm^2. The currents required for this cooling are on the order of 50 kA/cm^2 (Figure 5*b*).

It can be seen from Figure 5 that decreasing the cathode barrier height will increase the cooling power. This is true only down to barrier heights on the order of $k_B T$. Further reduction of the barrier height will increase the current substantially (increasing the Joule heating), without much increase in the cooling power. There is thus an optimum barrier height for maximum cooling. This optimum barrier falls outside the range where the assumption for Boltzmann distribution in Eq. (6) is valid. To derive the optimum value of the cathode barrier, one should use the full heat flux equation with a Fermi-Dirac distribution function.

Figure 6 shows the net cooling power as a function of temperature difference between the hot and cold sides for the 2-μm InGaAs barrier (solid line, filled circles). Other curves correspond to the cases where thermal conductivity is reduced (2 and 1 W/m K), mobility is increased (30,000 cm^2/V s), and effective mass is increased ($m_{\text{eff}}/m_0 = 0.08, 0.12$). Of course, most of these parameters are interdependent, and it is impossible to change one without changing others. This highlights, however, the importance of various parameters in the performance of the device and indicates the direction for improvements.

In conclusion, electron transport in single barrier heterostructure devices is investigated using Monte Carlo techniques. Studying the balance among cooling at

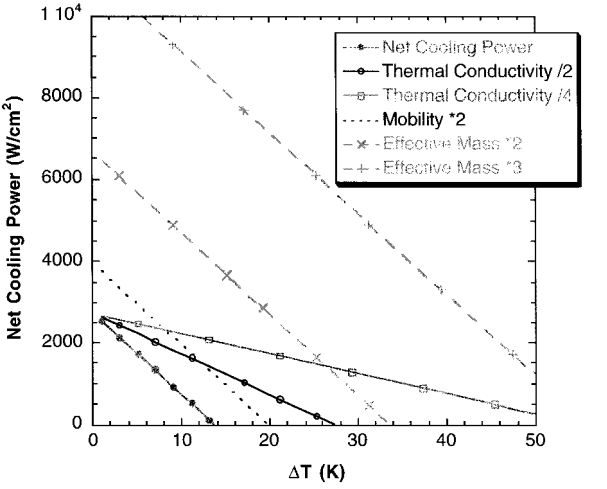


Figure 6. The net cooling power of an InGaAs HIT device as a function of the temperature difference between the hot and cold sides (solid curve with dark circles). The other curves show the prospects of HIT cooling by decreasing the thermal conductivity of the barrier, increasing the mobility, or increasing electron's effective mass (i.e., increasing Richardson's thermionic emission current).

the cathode junction due to the selective emission of hot electrons, Joule heating in the barrier due to various energy relaxation mechanisms, and heat conduction from the hot to the cold sides, conditions for the net cooling of the cathode layer below room temperature are derived and important material parameters are discussed.

REFERENCES

1. L. Rushing, A. Shakouri, P. Abraham, and J. E. Bowers, Micro Thermoelectric Coolers for Integrated Applications, *Proc. Sixteenth Int. Conf. on Thermoelectrics*, Dresden, Germany, August 1997, in press.
2. L. D. Hicks and M. S. Dresselhaus, Effect of Quantum-Well Structures on the Thermoelectric Figure of Merit, *Phys. Rev. B* vol. 47, pp. 12727–12731, 1993.
3. L. D. Hicks and M. S. Dresselhaus, Experimental Study of the Effect of Quantum-Well Structures on the Thermoelectric Figure of Merit, *Phys. Rev. B*, vol. 53, pp. 10493–10496, 1996.
4. T. C. Harman, D. L. Spears, and M. J. Manfra, High Thermoelectric Figures of Merit in PbTe Quantum Wells, *J. Electron. Mater.*, vol. 25, pp. 1121–1127, 1996.
5. J. O. Sofo and G. D. Mahan, Thermoelectric Figure of Merit of Superlattices, *Appl. Phys. Lett.*, vol. 65, pp. 2690–2692, 1994.
6. D. A. Broido and T. L. Reinecke, Effect of Superlattice Structure on the Thermoelectric Figure of Merit, *Phys. Rev. B*, vol. 51, pp. 13797–13800, 1995.
7. D. A. Broido and T. L. Reinecke, Thermoelectric Transport in Quantum Well Superlattices, *Appl. Phys. Lett.*, vol. 70, pp. 2834–2836, 1997.
8. D. M. Rowe and G. Min, Multiple Potential Barriers as a Possible Mechanism to Increase the Seebeck Coefficient and Electrical Power Factor, *Proc. Thirteenth Int. Conf. on Thermoelectrics*, pp. 339–342, Kansas City, MO, 1994.
9. L. W. Whitlow and T. Hirano, Superlattice Applications to Thermoelectricity, *J. Appl. Phys.*, vol. 78, pp. 5460–5466, 1995.
10. A. Shakouri and J. E. Bowers, Heterostructure Integrated Thermionic Coolers, *Appl. Phys. Lett.*, vol. 71, pp. 1234–1236, 1997.
11. D. G. Cahill, Heat Transport in Dielectric Thin Films and at Solid–Solid Interfaces, *Microscale Thermophys. Eng.*, vol. 1, pp. 85–110, 1997.
12. T. C. Harman, J. H. Cahn, and M. J. Logan, Measurement of Thermal Conductivity by Utilization of the Peltier Effect, *J. Appl. Phys.*, vol. 30, pp. 1351–1359, 1959.
13. S. M. Sze, *Physics of Semiconductor Devices*, 2d ed., John Wiley, New York, 1981.
14. A. A. Grinberg, Thermionic Emission in Heterosystems with Different Electronic Effective Masses, *Phys. Rev. B*, vol. 33, pp. 7256–7258, 1986.