# Heterostructure integrated thermionic coolers

Ali Shakouri[a] and John E. Bowers
*Department of Electrical and Computer Engineering, University of California, Santa Barbara, California 93106*

(Received 10 March 1997; accepted for publication 1 July 1997)

Thermionic emission in heterostructures is proposed for integrated cooling of high power electronic and optoelectronic devices. This evaporative cooling is achieved by selective emission of hot electrons over a barrier layer from the cathode to the anode. It is shown that with available high electron mobility and low thermal conductivity materials, and with optimized conduction band offsets in heterostructures, single-stage room temperature cooling of up to 20°–40° over thicknesses of the order of microns is possible. © *1997 American Institute of Physics.* [S0003-6951(97)03935-1]

Coolers are important elements of many electronic and optoelectronic systems. Thermoelectric (TE) coolers are used in most telecommunication lasers because of the need to stabilize the characteristics (threshold, power output, and wavelength) of the laser. TE coolers are essential in many infrared detectors applications because the sensitivity of the imaging array is much higher at low temperatures. The conventional thermoelectric effect is based on bulk properties of materials.[1] 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 will cool down the junction between two materials. In a linear transport regime, the overall performance of TE cooling devices can be expressed as a function of the dimensionless figure of merit $ZT=S^2\sigma T/\beta$, which describes the interplay between the Peltier cooling at the junctions given by the Seebeck coefficient $S$, Joule heating in the material given by electrical conductivity $\sigma$, and heat conduction from the hot to the cold junctions given by coefficient of thermal conductivity $\beta$.[1]

Recently there have been proposals to improve the thermoelectric figure of merit by using lower dimensional semiconductor structures,[2–6] by modifying parallel or perpendicular transport in superlattices,[7] or by using multiple potential barriers.[8,9] In this letter, semiconductor heterostructures are used in a transport regime beyond the linearized Boltzmann equation. Using the analogy with thermionic power generation by vacuum diodes, new heterostructure devices with enhanced cooling power are proposed. More important, this will allow use of more conventional semiconductor materials for fabrication of cooling devices integrated with high power electronic and optoelectronic components.

Thermionic energy conversion is based on the idea that a high work function cathode in contact with a heat source will emit electrons.[10] These electrons are absorbed by a cold, low work function anode, and they can flow back to the cathode through an external load where they perform useful work. Practical thermionic generators are limited by the work function of available metals or other materials that are used for the cathodes. Another important limitation is the *space charge effect*. The presence of charged electrons in the space between the cathode and anode will create an extra potential barrier which reduces the thermionic current. Recently, Mahan[11] proposed these vacuum diodes for thermionic refrigeration. Basically, the same vacuum diodes that are used for generators under applied bias will work as a cooler on the cathode side and a heater on the anode side. Mahan predicted efficiencies of over 80% of the Carnot value, but these refrigerators only work at high temperatures ($>500$ K). We will look at the prospects of heterostructures for thermionic refrigeration in the following.

The precise control of layer thickness and composition using techniques of molecular beam epitaxy (MBE) or metalorganic chemical vapor deposition (MOCVD) in conjunction with band gap engineering enable the design of a new class of semiconductor thermionic emission devices with improved cooling capacities (or increased power generation efficiencies). Using various material systems (such as GaAs/AlGaAsSb, InP/InGaAsP, Si/SiGe, HgCdTe/CdTe, or even more exotic PbTe/PbEuTe and HgCdSe/CdSe) one can produce different barrier heights in the anode and in the cathode (typically 0–0.4 eV). This is determined by the band edge discontinuity between heterolayers. Depending on the growth constraints and the lattice mismatch between materials, one can grade the barrier composition to produce internal fields and to enhance electron transport properties (see Fig. 1 inset). Close and uniform spacing of the cathode and anode is not a problem anymore since it can be achieved with atomic resolution.

To study the advantages and possible limitations of heterostructure integrated thermionic (HIT) cooling more quantitatively, we will consider a barrier material made of $Hg_{1-x}Cd_xTe$ that has optimized conduction band edge discontinuities at the two sides (cathode and anode). This material is a good candidate for HIT cooling because of very low thermal conductivity ($\beta\approx 1$ W/mK at 300 K and 2 W/mK at 100 K for $x\approx 0.2-0.4$), very good electron mobility ($\mu\approx 10^4$ cm$^2$/V s at 300 K and $10^5$ cm$^2$/V s at 77 K), and the fact that the conduction band edge can be easily tailored to create cathode and anode Schottky barriers of 0–0.4 eV.

In cooling by thermionic emission, the amount of heat absorbed in the cathode is the total current times the average energy of carriers that are emitted over the barrier. As the energy distribution of emitted electrons is almost exclusively on one side of Fermi energy, upon current flow, the strong carrier–carrier and carrier–lattice scatterings tend to restore the ''quasiequilibrium'' Fermi distribution in the cathode by

[a]Electronic mail: ali@opto.ucsb.edu

 0003-6951/97/71(9)/1234/3/$10.00 

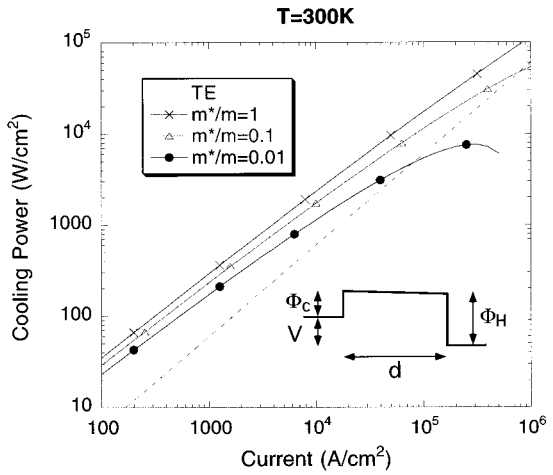


FIG. 1. Comparison of thermionic and thermoelectric cooling powers as a function of current at $T=300$ K. For the thermionic case, different curves correspond to electron effective masses in the cathode layer. For the thermoelectric case, the expression for Peltier cooling is derived in the linearized transport regime and is valid for low and moderate current densities ($<10^3-10^4$ A/cm$^2$). A Skebeck coefficient of 200 mV/K is assumed. The inset shows the conduction band diagram of a HIT cooler under bias.

absorbing energy from the lattice, and thus cooling the emitter junction.

By the use of Boltzmann statistics (which is valid for barrier heights greater than a few $k_B \cdot T$ that correspond to currents smaller than $10^5$ A/cm$^2$ at room temperature), we can calculate the average transport energy of carriers to be $\phi_C+(2k_BT_C/e)$, where $\phi_C$ is the cathode barrier height. We can express the barrier height as a function of current by using Richardson's thermionic emission expression:

$$\phi_C(T,I)=\frac{k_BT}{e}\left[\ln\left(\frac{em^*k_B^2T^2}{2\pi^2\hbar^3}\right)-\ln(I)\right],$$

where $m^*$ is the minimum of electron effective mass in the emitter and barrier regions. The Richardson current is valid as long as electron transport over the barrier is not limiting the current. In the latter case, one should solve the drift-diffusion equation or the Boltzmann equation in the barrier layer, with the appropriate boundary conditions at the cathode and anode junctions. The solution to such a nonisothermal problem, especially for the case of short barrier layers where ballistic transport is non-negligible, is quite involved and may require Monte Carlo simulations. As a first order approximation, we will take the total voltage over the barrier to be such that the voltage drop over a distance of the order of the electron mean free path is a few times the thermal energy. This so-called Bethe criterion[12] will justify the use of Richardson's thermionic emission expression. It should be emphasized that this might not be the optimum operating condition of a HIT cooler. Under this assumption, the Joule heating in the barrier is given by

$$Q_{\mathrm{J,TI}}\approx\frac{1}{2}\left[\frac{2k_B}{e}\left(T_C+\frac{\Delta T}{2}\right)\frac{d}{\lambda}\right]I,$$

where $\lambda$ is the electron mean free path in the barrier layer. This expression of Joule heating, which assumes one-half of the heat generated in the barrier goes to the emitter junction, is valid in a diffusive transport regime. When the barrier

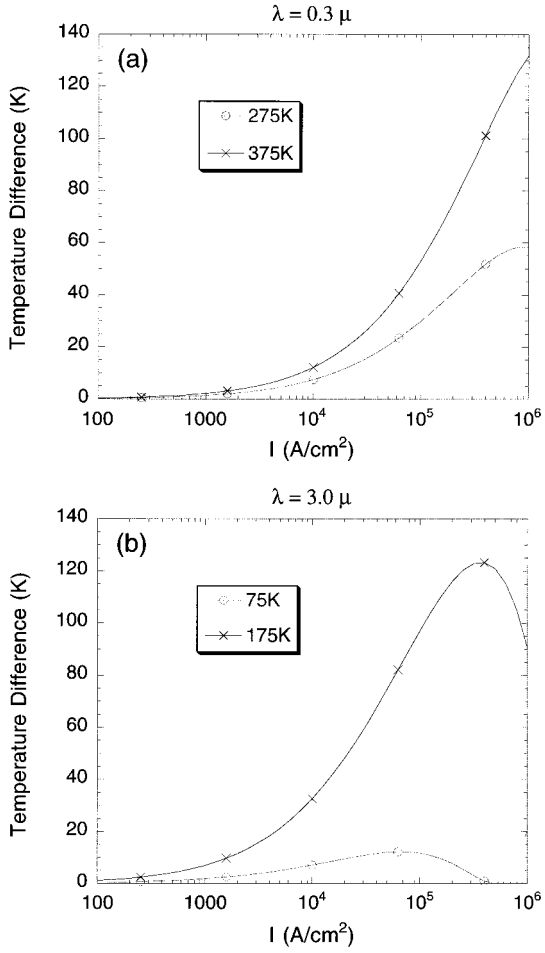


FIG. 2. Maximum temperature difference between the cold and hot sides as a function of current for the HIT cooler. (a) The electron mean free path is 0.3 $\mu$m and $T_C=275$ and 375 K. (b) The electron mean free path is 3 $\mu$m and $T_C=75$ and 175 K.

thickness is on the order of the energy relaxation length for electrons, the amount of heat transported back to the cold junction is only 30%–50% of this value.[13] It is important to note that the above expression of Joule heating is linear in current, in contrast to TE case where it was quadratic in current. The reason is that the supply of the electrons at the cathode is the current limiting process and not the transport over the barrier. So by using band gap engineering and appropriate cathode barrier height, the operating bias is chosen so that the total voltage drop over the barrier is the minimum required by the Bethe criterion. To calculate maximum cooling power, the anode barrier height is chosen to be high enough to suppress the reverse current from the anode to the cathode. The amount of heat conduction from the hot to the cold junction is given by $Q_c=(\beta\Delta T)/d$. From the expression of the overall thermionic cooling capacity one can calculate the maximum achievable temperature difference for an optimized device length as a function of electron mean free path and the coefficient of the thermal conductivity.

$$\Delta T_{\max}=T_C\left[\sqrt{1+\frac{\lambda k_B}{2e\beta}\left(\frac{e\Phi_C(T_C,I)}{k_BT_C}+2\right)^2I}-1\right].$$

Let us first consider the thermionic cooling term. Figure 1 shows that except at very high current densities

($10^6$ A/cm$^2$) or very low electron effective mass in the cathode layer, thermionic cooling is two to five times better than Peltier cooling. The reason for this is the *selective* emission of hot carriers from the cathode to the anode. Now if we consider the total thermionic cooling capacities for two cases of electron mean free path $=0.3$ $\mu$m [Fig. 2(a), $T=275$ and 375 K) and mean free path of 3 $\mu$m [Fig. 2(b) $T=75$ and 175 K), we see that for a current of $10^5$ A/cm$^2$, cooling of $\Delta T=30°$ (at $T_C=275$ K) over very short distance of only 0.6 $\mu$m barrier thickness is possible. Similarly, we can get $\Delta T=10°$ at $T_C=75$ K for $I=10^5$ A/cm$^2$ or $\Delta T=30°$ at $T_C=175$ K for $I=10^4$ A/cm$^2$. These currents are large, but when the thermionic coolers are integrated with small devices, one does not need to cool very large areas and this level of current is possible.

If we compare the maximum achievable TI cooling with the TE expression (which is independent of current): $\Delta T=ZT_C^2/2$,we see that for small currents there is a clear advantage for conventional TE cooling. Thermoelectric devices, however, in addition to using nonconventional semiconductor materials, usually require longer length devices, so for integration with high power components HIT cooling provides an effective alternative.

In the example discussed above, space charge effects were ignored. In an actual device, problems with electrostatic potential barriers can be reduced by modulated doping (on the order of $10^{16}$–$10^{17}$ cm$^{-3}$) or by modification of the band profile in the barrier layer. Another issue is the equilibrium between the electron system and the phonon system and their *effective* temperatures in cathode and anode layers. At very high currents it is possible to have different temperatures for the two systems. Even though the cooling of typical high power electronic and optoelectronic devices ($1-10$ kW/cm$^2$) does not go in this range of operation, the out of equilibrium issues will ultimately put an upper bound to the total cooling power.

One of the main benefits of heterostructure integrated thermionic cooling is the fact that there are fewer constraints on the materials to be used. Thermoelectric cooling using semiconductors in the linearized Boltzmann transport regime requires a material with high electrical conductivity and thermopower but low thermal conductivity $(Z=S^2\cdot\sigma/\beta)$.[14] In the case of thermionic cooling, the barrier material should have reasonably good electrical conductivity and very low thermal conductivity, but there is no requirement for high thermopower; the band edge discontinuities (Schottky barriers) at the anode and cathode will do the job sufficiently.

An interesting question is raised by Fig. 1 (inset) which displays cooling by thermionic emission: Is it necessary to have a hot junction at the anode side of the device? By band gap engineering and appropriate doping it should be possible to enhance the interaction of electrons with, for example, photons so that hot carriers at the anode side lose their energy by emitting light rather than by heating the lattice. The light emission could occur in a conventional $pn$ junction at the anode side or in a more elaborate unipolar quantum cascade laser configuration.[15] In these applications of integrated cooling and light emitting elements, the conventional requirement for a barrier material with low thermal conductivity does not apply anymore and one needs instead a material with very good thermal conductivity. This method of cooling can be viewed as an electrically pumped version of conventional laser cooling which has been used for atom trapping and very recently for cooling of macroscopic objects.[16]

One could combine the advantages of heterostructure thermionic cooling with lower dimensional structures by considering thermionic emission in multiquantum well structures. The change in density-of-states will affect the thermionic current and its cooling capacity. The cascading of heterostructures in order to reach lower temperatures can easily be done during the crystal growth and by fabrication of extra electrical contacts for intermediate layers. With existing sophisticated semiconductor processing techniques, if in some applications the thermal conductivity of the barrier is a limiting factor, one could selectively remove the barrier layers and recover the old vacuum thermionic generator with extremely small and precise cathode–anode separation. This also gives the possibility of using negative electron affinity semiconductors or microfield emission devices[17] as the cathode; or to use dipole doping and band gap engineering to modify the work function of the materials. It is expected that these heterostructure thermionic coolers would require small voltages and high currents. The possibility of using hole thermionic emission opens the way to put many devices thermally parallel and electrically in series, similar to TE cooler devices.

In conclusion, nonisothermal thermionic emission in heterostructures is proposed for integrated cooling of high power electronic components. Room temperature single-stage cooling of about 20°–40° for a barrier thickness of about 1 $\mu$m and currents of $1-2\times10^5$ A/cm$^2$ can be achieved with available high mobility, low thermal conductivity materials.

The authors would like to acknowledge many stimulating discussions with H. Kroemer, V. Narayanamurti, R. Ram, and D. Smith. This work was supported by DARPA and AFOSR under the project MOST.

[1] H. J. Goldsmit, *Electronic Refrigeration* (Pion, London, 1986).
[2] L. D. Hicks and M. S. Dresselhaus, Phys. Rev. B **47**, 12 727 (1993).
[3] L. D. Hicks, T. C. Harman, X. Sun, and M. S. Dresselhaus, Phys. Rev. B **53**, R10 493 (1996).
[4] T. C. Harman, D. L. Spears, and M. J. Manfra, J. Electron. Mater. **25**, 1121 (1996).
[5] J. O. Sofo and G. D. Mahan, Appl. Phys. Lett. **65**, 2690 (1994).
[6] D. A. Broido and T. L. Reinecke, Phys. Rev. B **51**, 13 797 (1995).
[7] D. M. Rowe and G. Min, Thirteenth International Conference on Thermoelectrics, Kansas City, 1994, p. 339.
[8] L. W. Whitlow and T. Hirano, J. Appl. Phys. **78**, 5460 (1995).
[9] B. Moyzhes, Fifteenth International Conference on Thermoelectrics, Pasadena, 1996, p. 183.
[10] G. N. Hatsopoulos and E. P. Gyftopoulos, *Thermionic Energy Conversion* (MIT Press, Cambridge, 1973–79), Vols. I and II.
[11] G. D. Mahan, J. Appl. Phys. **76**, 4362 (1994).
[12] S. M. Sze, *Physics of Semiconductor Devices*, 2nd ed. (Wiley, New York, 1981).
[13] A. Shakouri, E. Y. Lee, D. L. Smith, V. Narayanamurti, and J. E. Bowers, Microscale Thermophysical Engineering (submitted).
[14] C. B. Vining, Twelfth International Conference on Thermoelectrics, Yokohama, 1993, p. 126.
[15] J. Faist, F. Capasso, D. L. Sivco, C. Sirtori, A. L. Hutchinson, and A. Y. Cho, Science **264**, 553 (1994).
[16] C. E. Mungan, M. I. Buchwald, B. C. Edwards, R. I. Epstein, and T. R. Gosnell, Phys. Rev. Lett. **78**, 1030 (1997).
[17] I. Brodie and P. R. Schwoebel, Proc. IEEE **82**, 1006 (1994).