Monolithic integration of thin-film coolers with optoelectronic devices

Christopher LaBounty, MEMBER SPIE
Department of Electrical and Computer Engineering
University of California–Santa Barbara
Santa Barbara, California 93106-9560
E-mail: labounty@opto.ucsb.edu

Ali Shakouri
Jack Baskin School of Engineering
University of California–Santa Cruz
Santa Cruz, California 95064-1077

Patrick Abraham
John E. Bowers
Department of Electrical and Computer Engineering
University of California–Santa Barbara
Santa Barbara, California 93106-9560

Abstract. Active refrigeration of optoelectronic components through the use of monolithically grown thin-film solid-state coolers based on III-V materials is proposed and investigated. Enhanced cooling power compared to the thermoelectric effect of the bulk material is achieved through thermionic emission of hot electrons over a heterostructure barrier layer. These heterostructures can be monolithically integrated with other devices made from similar materials. Experimental analysis of an InP *pin* diode monolithically integrated with a heterostructure thermionic cooler is performed. Cooling performance is investigated for various device sizes and ambient temperatures. Several important nonideal effects are determined, such as contact resistance, heat generation and conduction in the wire bonds, and the finite thermal resistance of the substrate. These nonideal effects are studied both experimentally and analytically, and the limitations induced on performance are considered. Heterostructure integrated thermionic cooling is demonstrated to provide cooling power densities of several hundred W/cm^2. These microrefrigerators can provide control over threshold current, power output, wavelength, and maximum operating temperature in diode lasers. [S0091-3286(00)02511-3]

Subject terms: refrigeration; integrated; thermionic; cooling.

Paper 200084 received Mar. 3, 2000; revised manuscript received June 30, 2000; accepted for publication June 30, 2000.

1 Introduction

Optoelectronic devices such as laser sources, switching/routing elements, and detectors require careful control over operating temperature. This is especially true in current high-speed and wavelength division multiplexed (WDM) optical communication networks. Long-haul optical transmission systems operating around 1.55 μm typically use erbium-doped fiber amplifiers (EDFAs) and are restricted in the wavelengths they can use due to the finite bandwidth of these amplifiers. As more channels are packed into this wavelength window, the spacing between adjacent channels becomes smaller and wavelength drift becomes very important. Temperature variations are the primary cause of wavelength drift, and also affect the threshold current and output power in laser sources. Distributed feedback (DFB) lasers and vertical-cavity surface-emitting lasers (VCSELs) can generate large heat power densities, on the order of kW/cm^2, over areas as small as 100 μm^2.[1] Typical temperature-dependent wavelength shifts for these laser sources are on the order of 0.1 nm/°C. Therefore a temperature change of only a few degrees in a WDM system with a channel spacing of 0.2 to 0.4 nm would be enough to switch data from one channel to the adjacent one, and even less of a temperature change could dramatically increase the crosstalk between two channels. In many optoelectronic applications this temperature dependence is used to actively control the characteristics of the device, as in optical filters[2] or switches.[3,4] In other instances large absolute cooling is desired, as in IR photodetectors.[5] Temperature stabilization or refrigeration is commonly performed with conventional thermoelectric (TE) coolers; however, since integration with optoelectronic devices is difficult,[6] component cost is greatly increased because of packaging. The reliability and lifetime of packaged modules is also usually limited by the TE cooler.[7]

An alternative solution to thermal management needs is to incorporate heterostructure integrated thermionic (HIT) refrigerators with optoelectronic devices.[8,9] These thin-film coolers use the selective thermionic emission of hot electrons over a heterostructure barrier layer. This evaporative cooling occurs because the hot electrons that are on one side of the Fermi energy are emitted. So as to maintain the quasiequilibrium Fermi distribution, lower-energy electrons absorb thermal energy from the lattice at the junction. The emitted electrons then redeposit their energy after passing over the barrier. Since these thin-film coolers can be made with conventional III-V semiconductor materials, monolithic integration with optoelectronics is possible. The result of this integration is an extended component lifetime and a very fast cooling response due to the small thermal mass of the cooler. Furthermore, standard integrated-circuit batch fabrication techniques can be used to manufacture these coolers, whereas TE coolers use a bulk fabrication process. Another advantage of a HIT cooler is the dramatic gain in cooling power density.

To demonstrate monolithic integration of a device with a HIT cooler, a simple InP *pin* diode was grown on top of a 1-μm-thick InGaAs/InGaAsP HIT cooler. Cooling performance was investigated for various currents through the diode, which corresponds to effectively changing the ther-

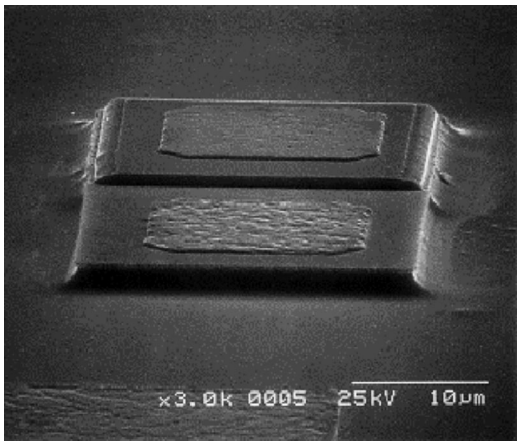


Fig. 1 Scanning electron micrograph (SEM) of a processed device showing the *pin* diode on top of the 1-μm-thick HIT cooler.

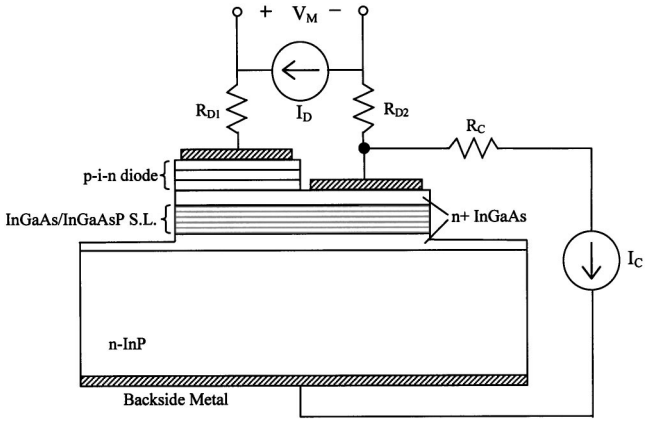


Fig. 2 Setup for measuring the temperature of the integrated diode versus the current through the cooler (I_C) and the diode (I_D). Here R_{D1}, R_{D2}, and R_C are the wire bond resistances.

mal load. After discussing the measurements, the nonideal parasitic effects are identified and examined. These results are then used to predict the effectiveness of HIT-cooler integration with diode lasers.

2 Device Structure

The tested HIT cooler structure consisted of a 1-μm-thick superlattice barrier (25 periods of 10-nm InGaAs and 30-nm InGaAsP, $\lambda = 1.3\,\mu$m) surrounded by n^+ InGaAs cathode and anode layers grown by metal-organic chemical vapor deposition (MOCVD). The cathode and anode layers were 0.3 and 0.5 μm thick, respectively. On top of these layers, a 0.85-μm-thick *pin* diode was grown during the same MOCVD growth. In two wet etching steps, two stacked mesas are defined corresponding to the diode on top of the cooler, as shown in Fig. 1. The cooler mesas ranged in size from 20×40 to $100\times 200\,\mu\text{m}^2$, and the diode size was one-half the cooler size. Ti/Pt/Au was used to make ohmic contacts to both the *p*- and *n*-type material. The substrate was thinned to approximately 125 μm before the back-side metal was deposited. The integrated devices were then cleaved, packaged, and wire-bonded for testing.

3 Measurement

The diode serves two purposes in the measurement. By changing the current through the diode we can effectively change the heat load of the cooler. Also, by monitoring the voltage across the diode, we can measure the temperature[10] on the cold side of the cooler, T_C. The temperature sensitivity of the diode near room temperature was determined to be 1.936 mV/°C at a bias of 1 mA. The measurement setup is illustrated in Fig. 2. A constant current (I_D) was sent through the diode, and the diode voltage (V_D) was monitored as the cooler current (I_C) was varied. The resistors of Fig. 2 represent the parasitic wire bonds, which add to the heat load on the device. The measured voltage can be expressed as

$$V_M = (R_{D1}+R_{D2})I_D + V_D + \frac{\rho_c}{A_c}(I_D+I_C), \tag{1}$$

where R_{D1} and R_{D2} are the wire bond resistance, ρ_c is the contact resistivity, and A_c is the area of the metal contact on the cooler. If Eq. (1) is rearranged to solve for V_D as a function of I_C, then the resulting expression is equal to the measured voltage minus some constant values and minus a value that changes with respect to I_C, that is, $\rho_c I_C/A_C$. Therefore, to correctly measure the temperature on top of the device, the contact resistivity must be determined. Once V_D is known, the temperature can be calculated using the temperature dependence of the diode voltage at a constant current as given above.

The thermionic heating and cooling are proportional to the current and can be expressed (in the limit of Boltzmann statistics) as

$$Q_{\text{TI}} = \left(\phi_B + \frac{2k_BT}{e}\right)\cdot I, \tag{2}$$

where ϕ_B is the heterojunction barrier height, k_B is Boltzmann's constant, and e is the charge of an electron.[9] Since cooling is linearly proportional to current and Joule heating is proportional to the square of current, the temperature versus current is described by a second-order polynomial. This can best be seen from the expression for overall cooling power:

$$Q_1 = Q_{\text{TI}} - \frac{\rho d}{A_c}I_C^2 - \frac{\beta}{d}\Delta T, \tag{3}$$

where ρ is the electrical resistivity of the cooler, d is its thickness, β is the thermal conductivity between hot and cold junctions, and ΔT is the temperature difference across the cooler. When this expression is solved for ΔT, the terms in the second-order polynomial are apparent.

4 Results

A direct way of measuring the temperature of the device is to use a microthermocouple. Device A from Fig. 3 shows the temperature versus current for the cold side of a $70\times 140\,\mu\text{m}^2$ cooler that had the integrated diode removed by selective wet etching. The temperature data was curve-

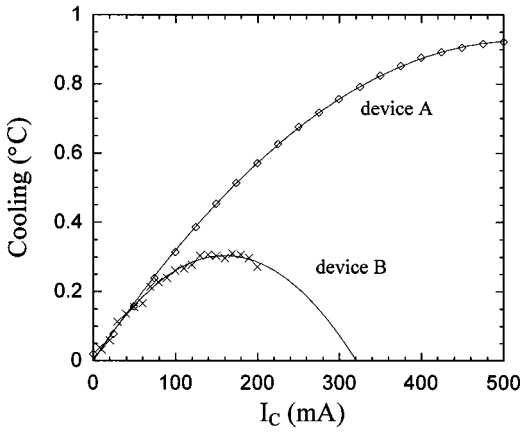


Fig. 3 Measured cooling versus cooler current using a microthermocouple for a well-packaged device (curve 1), and using the integrated *pin* diode with poorer packaging (curve 2). Both device areas were $70\times140\ \mu m^2$.

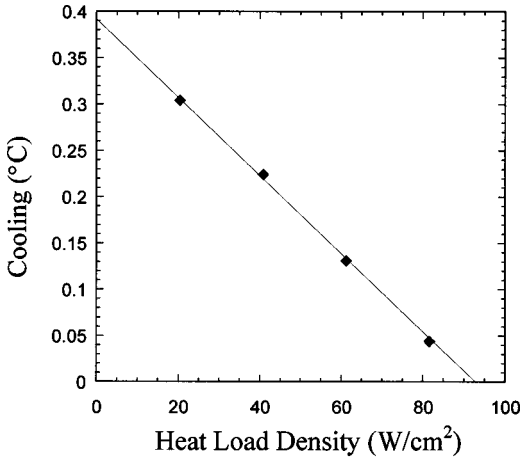


Fig. 4 Cooling versus excess heat load density. The points are experimental data, while the linear curve fit corresponds to Eq. (3).

fitted with a second-order polynomial, resulting in a linear coefficient of 3.53°C/A and a quadratic coefficient of 3.37°C/A^2. Since this cooler structure was identical to the one with the integrated diode (device B), the two devices should have the same linear coefficient. The quadratic coefficient, on the other hand, may differ, since Joule heating is dependent on processing variations such as contact resistance, packaging, and wire bonding. Consequently the contact resistance (ρ_C) can be determined from Eq. (1) by adjusting its value until the linear coefficients match. Using this method, device B from Fig. 3 also shows the temperature versus current for the cold side of a cooler with an integrated diode biased at 1 mA. The performance is much worse, due to poorer packaging and a higher contact resistance. For device B the contact resistance was determined to be $4.4\times10^{-6}\ \Omega\ cm^2$. The contact resistance for device A was measured by the transmission-line model[11] and was estimated to be roughly $5\times10^{-7}\ \Omega\ cm^2$. Knowing the contact resistance, the current through the diode was increased, and the temperature versus cooler current was again measured. Each time the diode current was changed, the temperature sensitivity was recalibrated. By changing the diode current, the heat load is changed. Figure 4 shows the maximum cooling on top of the device as a function of heat load. The dependence of cooling on the heat load density can be described in a similar manner to that of a thermoelectric device,[12]

$$T_C = T_{C\,\max}\cdot\left(1-\frac{Q_l}{Q_{l\,\max}}\right), \tag{3}$$

where T_C is the cold-side temperature, Q_l is the heat load density, and $T_{C\,\max}$ and $Q_{l\,\max}$ are the corresponding maximum values. From Fig. 4 we find that the maximum cooling was 0.39°C, and the maximum heat load density was 93 W/cm^2, considering only heat generation by the diode. In addition to the diode there is a constant heat load due to the two wire bonds attached to the diode (R_{D1} and R_{D2}), and another heat load due to the wire bond to the cooler (R_C), which changes with changing cooler current. In fact, much of the improvement from device B to device A in Fig. 3 can be attributed to the shorter wire bonds. The wire-bond length should not be too short, however, or else the heat conduction through the wire from the cold side to the heat sink will begin to become an issue. An optimum wire length can be determined from the wire's diameter, its electrical and thermal conductivities, and the temperature difference across the wire bond. Figure 5 shows this analysis for a gold wire (25-μm diameter) with various values of current and cooling temperature, assuming that the package side of the wire bond is well heat-sunk. For long wire bonds the Joule heating dominates the cooling power loss, and curves with a similar current approach the same value. For short wire bonds the heat conduction dominates, and curves with a similar temperature difference approach the same value. In between these two extremes there is an optimum wire bond length for a given current and temperature difference.

It is useful to examine the magnitudes of the various sources of heat that are contributing to the total applied thermal load. Figure 6 shows the heat load density versus

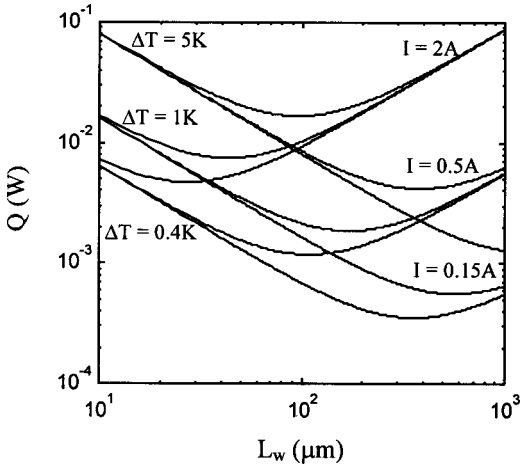


Fig. 5 Cooling power lost versus gold-wire-bond length for various bias currents and temperature differences. The electrical and thermal conductivities of gold were assumed to be $45.5\times10^4\ \Omega^{-1}\ cm^{-1}$ and 3.17 W/cm K, respectively.

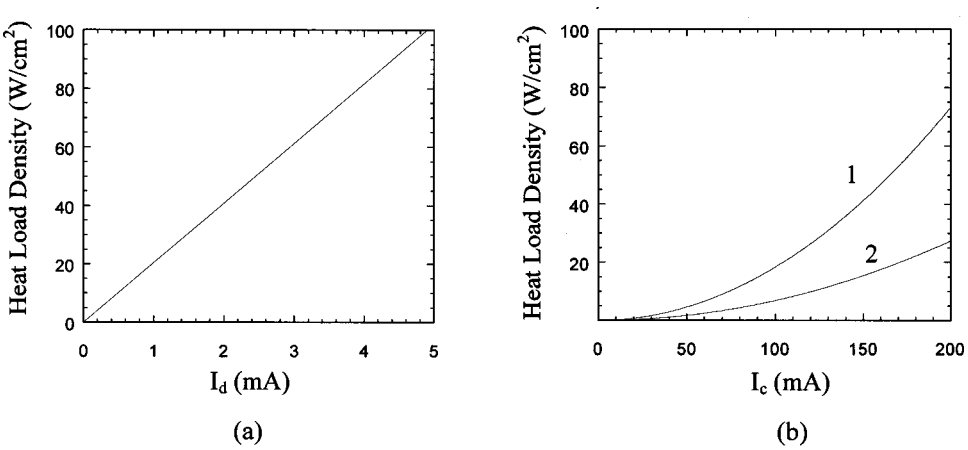


Fig. 6 (a) Heat load density for the diode versus diode current. (b) Heat load density from the wire bond (curve 1) and from the contact resistance (curve 2) versus cooler current.

the respective current bias for the dominating sources of heat. For small temperature differences, it was assumed that approximately one-half the heat generation in the wire bonds arrives at the device, while the other half goes to the heat sink. The heat generation due to the diode contact resistance and wire bonds connected to the diode is not shown, since it was several orders of magnitude smaller due to the small value of the diode current. At the optimum cooler current bias of 160 mA (device B from Fig. 3), there is 47 W/cm^2 of heat load density due to the wire bond connected to the device, and an additional 18 W/cm^2 due to the contact resistance. Using these values, the actual maximum heat load density was 158 W/cm^2.

The smallest coolers yielded the largest absolute cooling. Since a smaller device requires less current to achieve the same temperature gradient, there is less current for parasitic Joule heating from the wire bonds and contact resistance, and the overall cooling is larger. Increasing the heat-sink temperature resulted in a further increase in cooling. Figure 7 shows the maximum cooling increasing from 1.15 to 2.25°C when the heat-sink temperature was raised from 20 to 90°C. One reason for the improvement is the larger thermal spread of carriers at higher heat-sink temperatures, which allows for more carriers to pass over the heterojunction barrier. The other reason is the reduced thermal conductivity of the barrier material, cutting the amount of heat that returns to the cold junction.

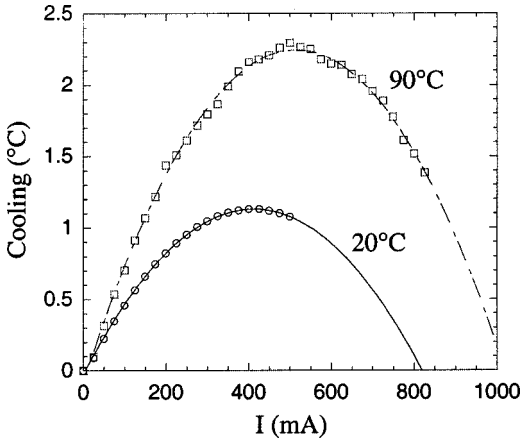


Fig. 7 Measured cooling at heat-sink temperatures of 20 and 90°C.

5 Application to Optoelectronics

The benefits of integrated thin-film cooling for optoelectronic devices can be discussed in three regimes, described as temperature tuning, heat pumping, and absolute cooling. In the first regime the temperature is to be tuned by a moderate amount to control the operating characteristics such as the emission wavelength in a laser diode. In a laser diode operating near or above lasing threshold, most of the heat is generated in the vicinity of the active region due to nonradiative recombination and absorbed radiative recombination.[13,14] Since there is a finite thermal resistance between the active region and heat sink, the temperature of the active region will be greater than that of the heat sink. This thermal resistance is mostly due to the substrate underneath the laser for in-plane geometries, and a combination of the substrate and bottom mirror for vertical cavity geometries. By using the integrated cooler, the active-region temperature may be controlled more precisely and rapidly, as the distance to the cooler can be as small as a few microns and the mass of the cooler is comparable to that of the laser. This is also true in other optoelectronic devices such as filters and switches. Many of these devices make use of Bragg gratings or nonlinear waveguides, where it is important to maintain the temperature at a certain value to achieve the desired phase-matching condition. In many of these cases, the cooler would often be operated in reverse bias to heat as well as cool the device.

The active region in a typical diode laser can reach temperatures more than 70°C above the heat-sink temperature. In this regime the microrefrigerator does not need to cool the laser below the heat-sink temperature, but only needs to provide high heat pumping densities. If the conventional cold side of the cooler is actually hotter than that of the heat sink, then the heat conduction [third term in Eq. (3)] is in the opposite direction and larger cooling power densities are possible. For example, if the active region of a DFB laser with a temperature-dependent wavelength shift of 0.1 nm/°C can be cooled to the heat-sink temperature, then a

tuning of 7 nm should be possible. This wavelength shift could also be used to monitor the temperature of the active region, providing yet another way to characterize the thin-film cooler's performance. Similarly, the output power for a typical DFB laser changes by approximately 0.4 dB/°C. Temperature sensitivity in other kinds of laser diodes is typically much greater. Broad-area Fabry-Perot devices operating in the red and near-IR spectral regions typically have wavelength sensitivities of 0.5 nm/°C,[15] resulting in a larger tunability. Furthermore, if the temperature of the active region could be kept close to that of the heat sink, then the maximum operating temperature of the laser would be equal to that of the heat sink. Keeping the temperature of the active region close to that of the heat sink would also inevitably extend the lifetime of the laser. Finally, with a cooled active region, the carrier leakage out of the separate confinement heterostructure layers would be reduced, resulting in a lower threshold current.

The last regime considered is that of large absolute cooling. If substantial cooling below the heat-sink temperature is desired, simulations with better-optimized structures and packaging have predicted single-stage cooling of 20 to 30°C with cooling power densities of several thousand W/cm^2.[16] This better performance would not only enhance their utility in the above-mentioned applications, but also allow for use in such applications as IR lasers and photodetectors.

6 Conclusions

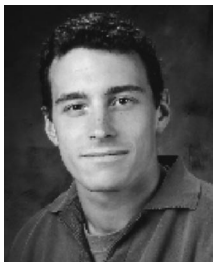
Monolithically integrated cooling of optoelectronic devices with thin-film solid-state coolers has been demonstrated. Cooling power densities of several hundred W/cm^2 were measured with an integrated InP *pin* diode and a 1-μm-thick InGaAs/InGaAsP HIT cooler. The wire bonds and contact resistance were determined to add significantly to the heat load density and must be minimized in an optimally packaged device. Integration of coolers with diode lasers grown from similar materials is possible, and it should be suitable for significantly controlling the characteristics of the laser sources by manipulating the cooler current. More generally, these monolithically integrated thin-film coolers should have wide applications in any optoelectronic device where it is beneficial to control the operating temperature.

Acknowledgments

This work was supported by the Army Research Office through the DARPA/HERETIC program and by the Office of Naval Research.

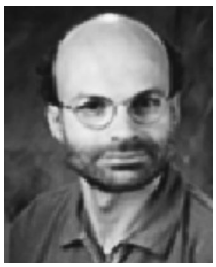
References

1. J. Piprek, Y. A. Akulova, D. I. Babic, L. A. Coldren, and J. E. Bowers, ''Minimum temperature sensitivity of 1.55 μm vertical-cavity lasers at 30 nm gain offset,'' *Appl. Phys. Lett.* **72**(15), 1814–1816 (1998).
2. J. Hukriede, D. Kip, and E. Kratzig, ''Thermal tuning of a fixed Bragg grating for IR light fabricated in a $LiNbO_3$:Ti channel waveguide,'' *Appl. Phys. B: Lasers Opt.* **70**, 73–75 (2000).
3. Y. Baek, R. Schiek, and G. I. Stegeman, ''All-optical switching in a hybrid Mach-Zehnder interferometer as a result of cascaded second-order nonlinearity,'' *Opt. Lett.* **20**(21), 2168–2170 (1995).
4. X. Lu, D. An, L. Sun, Q. Zhou, and R. T. Chen, ''Polarization-insensitive thermo-optic switch based on multimode polymeric waveguides with an ultralarge optical bandwidth,'' *Appl. Phys. Lett.* **76**(16), 2155–2157 (2000).
5. J. Piotrowski, C. A. Musca, J. M. Dell, and L. Faraone, ''Infrared photodetectors operating at near room temperature,'' in *Proc. Conf. on Optoelectronic and Microelectronic Materials and Devices*, pp. 124–127, Perth, WA, Australia (1998).
6. L. Rushing, A. Shakouri, P. Abraham, and J. E. Bowers, ''Micro thermoelectric coolers for integrated applications,'' in *Proc. 16th International Conf. on Thermoelectrics*, pp. 646–649, Dresden, Germany (1997).
7. T. A. Corser, ''Qualification and reliability of thermoelectric coolers for use in laser modules,'' in *41st Electronic Components and Technology Conference*, pp. 150–165, Atlanta, GA, USA (1991).
8. A. Shakouri, C. LaBounty, J. Piprek, P. Abraham, and J. E. Bowers, ''Thermionic emission cooling in single barrier heterostructures,'' *Appl. Phys. Lett.* **74**(1), 88–89 (1999).
9. A. Shakouri, E. Y. Lee, D. L. Smith, V. Narayanamurti, and J. E. Bowers, ''Thermoelectric effects in submicron heterostructure barriers,'' *Microscale Thermophys. Eng.* **2**(1), 37–48 (1998).
10. J. M. Swartz and J. R. Gaines, ''Wide range thermometry using gallium arsenide sensors,'' in *Measurement & Control in Science & Industry*, Vol. 4, Harmon H. Plumb, Ed., pp. 1117–1124, Instrument Society of America, Philadelphia, PA (1972).
11. G. K. Reeves and H. B. Harrison, ''Obtaining the specific contact resistance from transmission line model measurements,'' *IEEE Electron Device Lett.* **EDL-3**(5), 111–113 (1982).
12. D. M. Rowe, *CRC Handbook of Thermoelectrics*, CRC Press, New York (1995).
13. N. K. Dutta, T. Cella, R. L. Brown, and D. T. C. Huo, ''Monolithically integrated thermoelectric controlled laser diode,'' *Appl. Phys. Lett.* **47**(3), 222–224 (1985).
14. P. R. Berger, N. K. Dutta, K. D. Choquette, G. Hasnain, and N. Chand, ''Monolithically Peltier-cooled vertical-cavity surface-emitting lasers,'' *Appl. Phys. Lett.* **59**(1), 117–119 (1991).
15. T. S. El-Bawab, C. H. Vaishnav, A. P. Jayasumana, H. Temkin, J. R. Sauer, and H. A. Willebrand, ''Robust wavelength division multiplexed local area networks,'' *Fiber Integr. Opt.* **16**, 237–260 (1997).
16. C. LaBounty, A. Shakouri, G. Robinson, P. Abraham, and J. E. Bowers, ''Design of integrated thin film coolers,'' in *Proc. 18th International Conf. on Thermoelectrics*, Baltimore, MD, USA, 1999.

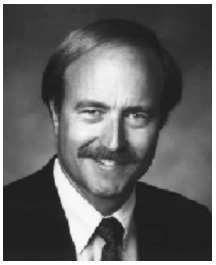


Christopher LaBounty received his BS in electrical engineering from the University of Vermont in 1997 and his MS in 1999 from the University of California, Santa Barbara. He is currently pursuing his PhD at UCSB, working on heterostructure integrated thermionic coolers and their integration with optoelectronics. His research interests also include optical WDM communication systems and thermal management in optoelectronic components.

Ali Shakouri received his Engineer degree from Ecole Nationale Superieure des Telecommunications de Paris, France, in 1990, and PhD degree from California Institute of Technology in 1995. He is an assistant professor of electrical engineering at the University of California, Santa Cruz. From 1995 to 1998 he was with the Ultrafast Optoelectronics Group at UCSB. His present research is in integrated cooling of optoelectronic devices and photonic switching.



Patrick Abraham received the MS and PhD degrees in material science from the University C. Bernard Lyon 1, France, in 1984 and 1987, respectively. He was a researcher at Centre National de la Recherche Scientifique (CNRS), France, and worked until 1988 at the Laboratoire de Physico-Chimie Minerale, France. He is currently a research engineer at the University of California at Santa Barbara, where his research interests include the study of new active regions for long-wavelength lasers, their growth by MOCVD, and their growth on compliant universal substrates.



John E. Bowers received the MS and PhD degrees from Stanford University, Palo Alto, California. He is the director of the Multidisciplinary Optical Switching Technology Center (MOST), the executive director of the Center for Entrepreneurship and Engineering Management, and a professor in the Department of Electrical and Computer Engineering at the University of California, Santa Barbara (UCSB). He is a member of the Optoelectronics Technology Center and the NSF Science and Technology Center on Quantized Electronic Structures. His research interests are primarily concerned with high-frequency optoelectronic devices and physics. He worked for AT&T Bell Laboratories and Honeywell before joining UCSB in 1987.