JOURNAL OF APPLIED PHYSICS VOLUME 89, NUMBER 7 1 APRIL 2001

Design and characterization of thin film microcoolers

Chris LaBounty[a)]
Department of Electrical and Computer Engineering, University of California, Santa Barbara, California 93106

Ali Shakouri
Jack Baskin School of Engineering, University of California, Santa Cruz, California 95046

John E. Bowers
Department of Electrical and Computer Engineering, University of California, Santa Barbara, California 93106

(Received 29 February 2000; accepted for publication 15 January 2001)

Thin film coolers can provide large cooling power densities compared to bulk thermoelectrics due to the close spacing of hot and cold junctions. Important parameters in the design of such coolers are investigated theoretically and experimentally. A three-dimensional (3D) finite element simulator (ANSYS) is used to model self-consistently thermal and electrical properties of a complete device structure. The dominant three-dimensional thermal and electrical spreading resistances acquired from the 3D simulation are also used in a one-dimensional model (MATLAB) to obtain faster, less rigorous results. Heat conduction, Joule heating, thermoelectric and thermionic cooling are included in these models as well as nonideal effects such as contact resistance, finite thermal resistance of the substrate and the heat sink, and heat generation in the wire bonds. Simulations exhibit good agreement with experimental results from InGaAsP-based thin film thermionic emission coolers which have demonstrated maximum cooling of 1.15 °C at room temperature. With the nonideal effects minimized, simulations predict that single stage thin film coolers can provide up to 20–30 °C degrees centigrade cooling with cooling power densities of several 1000 W/cm^2. [DOI: 10.1063/1.1353810]

I. INTRODUCTION

Temperature control of microelectronic and optoelectronic components is typically accomplished with thermoelectric (TE) coolers. TE coolers have become essential in modern optical telecommunications to control the characteristics of laser sources, switching/routing elements, and detectors used in wavelength division multiplexed systems. Cooling requirements in microprocessors and other integrated circuits have also risen dramatically in recent years due to the increase in clock speed and reduction in feature size. Generally, as these devices have become smaller, faster, and more dense, the power density has greatly increased. Conventional TE coolers are incompatible with integrated circuit fabrication processes, and are therefore limited in how small they can be manufactured. This bulk fabrication technology makes integration with microelectronic and optoelectronic devices difficult,[1] resulting in a high cost of packaging. Furthermore, the reliability of packaged modules employing a TE cooler is usually limited by the reliability of the cooler itself.[2] A solution to these problems is to shift from bulk thermoelements to integrated thin film coolers.

The greatest advantage of thin film coolers is the dramatic gain in cooling power density as it is inversely proportional to the length of the thermoelements. Thin films on the order of microns should provide cooling power densities greater than 1000 W/cm^2. In addition, several methods such as thermionic emission in heterostructures[3,4] and decreased thermal conductivity in superlattices,[5,6] are being explored to improve the performance beyond what is possible with bulk thermoelements. Thin film coolers can also be made in large quantities using well known integrated circuit batch fabrication methods resulting in lowered cost and greater reliability. Integration with microelectronic and optoelectronic devices is also possible for active localized cooling.

Many nonidealities become apparent and must be considered when moving from bulk to thin film coolers. Whereas contact resistance, thermal resistance of the heat sink, and heat generation in the current carrying connections are secondary effects in bulk TE coolers, they can all become critical in thin film coolers. In the following, these nonideal effects, as well as heat conduction, Joule heating, thermoelectric and thermionic cooling, are investigated. A three-dimensional (3D) self-consistent thermal/electrical software simulator[7] is used to model these effects (Fig. 1), and the results are compared to experimental measurements. The three-dimensional analysis is necessary to accurately model the electrical and thermal spreading resistance, however simulation of 3D structures is somewhat slow. The 3D electrical and thermal resistances can be determined for various geometries and used in an effective one-dimensional (1D) model[8] to obtain faster results when several parameters are to be varied. Thermionic cooling is considered throughout to make fair comparisons to experimental data, however the results are equally applicable to thermoelectric thin film coolers.

[a)]Electronic mail: labounty@opto.ucsb.edu

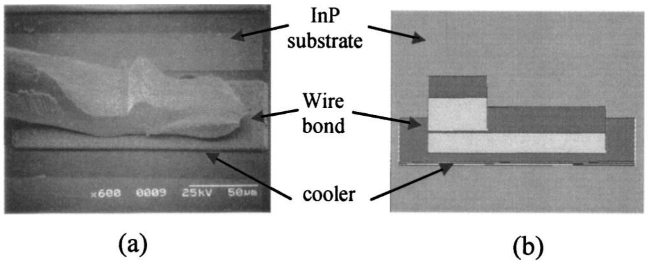

FIG. 1. (a) Scanning electron micrograph of a thin film thermionic cooler (InGaAs/InGaAsP/InGaAs 0.3 μm/1 μm/0.5 μm) and (b) the simulated structure. The uppermost wire bond length is scaled to reduce the element count in the mesh.

II. NONIDEAL EFFECTS

The most evident nonideal effect for thin film coolers is the substrate and package thermal resistance. Qualitatively speaking, if the distance between the cooling and heating regions is several orders of magnitude smaller than the distance between the heating and heat sink regions, most of the heat will flow back to the cold side of the device if the thermal conductivities of the thin film and substrate are comparable. Figure 2(a) shows simulation results of substrate thermal resistance versus thickness for various substrate materials. The simulation is performed assuming a rectangular etched mesa (5000 μm^2) thin film cooler on a semi-infinite plane substrate. The boundary conditions imposed assume the sides and top of the substrate to be adiabatic and the bottom isothermal. A uniform heat load on top of the mesa is assumed. The relatively good fit to a $\ln(x)$ function is indicative of the thermal spreading in two and three-dimensional heat flow.[9,10] Below 15 μm, the heat flow becomes dominantly one dimensional and the thermal resistance scales linearly with substrate thickness. The logarithmic fitting function of the thermal resistance and the corresponding coefficients can be given as

$$R_{\text{thermal}}(h)=a_i \ln(h)+b_i, \tag{1a}$$

$$a_i=\frac{1167}{\beta_i[W/mK]};\quad b_i=\frac{-1445}{\beta_i[W/mK]}, \tag{1b}$$

where R_{thermal} has units of [K/W], h is the substrate thickness in microns, and a_i and b_i are the fitting coefficients corresponding to the thermal conductivity of the substrate (β_i).

Clearly it is beneficial to use substrates with high thermal conductivity and a minimum thickness. However, thin film coolers that are grown on typical substrate materials (InP, Si, etc.) can be lapped only by a limited amount before the material begins to warp and become severely fragile. This warping occurs with InP substrates when they are lapped below 100 μm, which for the device size simulated corresponds to a thermal resistance of 57 K/W. Using a 1 μm film of InGaAsP for the cooler ($\beta=3.3$ W/m K) gives a thermal resistance of 61.6 K/W showing that roughly half the heat flows back to the cold side of the device.

The area of the device also affects the thermal resistance. Figure 2(b) shows simulation results of thermal resistance (points on lower curve) for various device areas assuming a 125-μm-thick InP substrate. The upper curve represents a

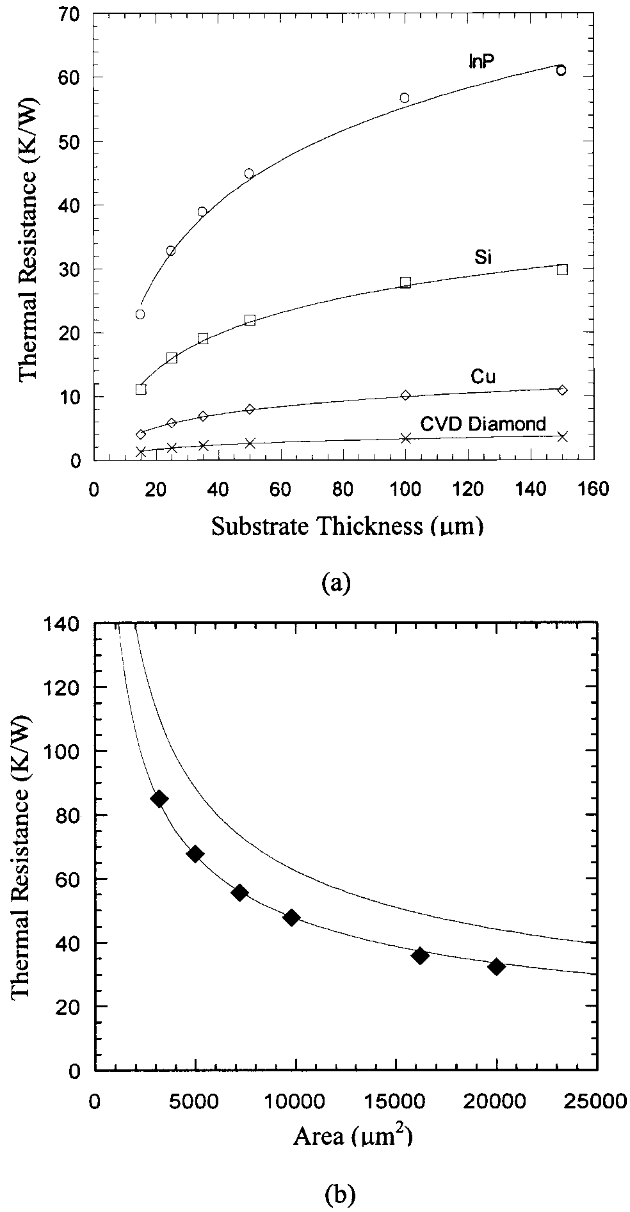

FIG. 2. (a) Thermal resistance vs substrate thickness for a device area of 5000 μm^2 on various substrates: InP($\beta=71$ W/m K), Si($\beta=145$ W/m K), Cu($\beta=398$ W/m K), and chemical vapor deposition diamond($\beta=1200$ W/m K). The points correspond to 3D simulation results, and the solid curves are the theoretical $a_i\cdot\log(x)+b_i$ curve fits. (b) Thermal resistance vs device area. The points are simulation results assuming a 125-μm-thick InP substrate and the upper solid curve is the theoretical plot assuming the substrate is an entire half space.

theoretical expression that assumes a disk heat source on a half space denoting a purely three-dimensional heat flow

$$R_{\text{thermal}}=\frac{\sqrt{\pi}}{4\beta\sqrt{A}}, \tag{2}$$

where β is the thermal conductivity of the half space and A is the area of the disk.[9] The simulation results are fitted with this same expression (solid line), and the resulting expression is equivalent to Eq. (2) multiplied by a reduction factor of 0.761. As the substrate thickness is increased, the simulation results approach that of the theoretical expression.

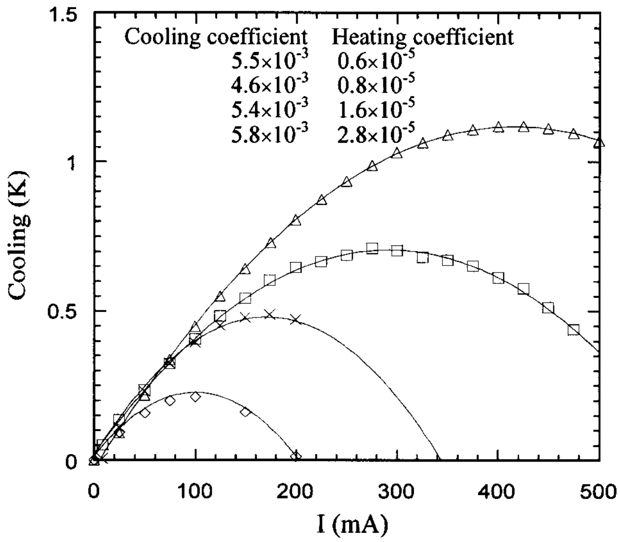

FIG. 3. Measured cooling for a 1-μm-thick cooler (InGaAs/InGaAsP superlattice) with various packages. As the packaging is improved the heating coefficient is reduced resulting in greater absolute cooling. The cooling (linear) and heating (quadratic) coefficients correspond to a second order polynomial fit of ΔT vs I. All temperatures are relative to the value at zero current with a heat sink temperature of 20 °C.

Instead of working with thin substrates that have a low thermal conductivity, switching to a substrate with a very high thermal conductivity would be advantageous. Several methods for transferring thin films to surrogate substrates with better thermal properties are possible, but each has potential problems. Epitaxial lift-off and grafting of epitaxial films has been discussed in the literature[11–14] for various semiconductor material systems, but neither the electrical nor thermal contact resistance of the interface has been characterized. Another possibility is to use flip chip bonding techniques to accomplish the substrate transfer, however thermal expansion mismatch between the thin films and the new substrate can cause problems in material quality.

Packaging of the single element thin film coolers was necessary to attain significant cooling since simply probing the cold side of the device presented too large a heat load. Using wire bonds from the package to the device alleviated some of this problem, but a significant heat load from the wire bond remained as well as the additional thermal resistance of the package between the hot side of the cooler and the heat sink. An optimum wire length can be determined by considering the trade-off between Joule heating in the wire and thermal conduction away from the cold junction. Ultimately when the thin film coolers are integrated with real devices or are packaged in a conventional TE configuration (n- and p-type legs electrically in series, thermally in parallel—see, for example, Ref. 15), the issue of wire bonding will be less of a concern. Figure 3 shows the increase in InGaAsP thin film cooler performance for four generations of packages. The details of the cooler will be discussed later. Generally in each successive package generation the optimum wire bond length is approached and the heat sinking is improved by choosing package materials with higher thermal conductivities. The best cooling shown in Fig. 3 is for a package made from a 350-μm-thick silicon wafer and using a wire bond length of 250 μm. The sample was mounted with a 4-μm-thick tin solder layer. Qualitatively, when the packaging improves the linear coefficient that represents thermionic and thermoelectric cooling effects remains relatively constant while the quadratic coefficient that represents Joule heating is reduced. This results in greater net cooling. One should note that, due to a relatively high contact resistance, the maximum current that can be applied to the thin film device is limited and thus we are still in a linear transport regime. With improvements in nonideal effects and by using asymmetric barriers, one can benefit fully from the thermionic cooling process by going into the nonlinear transport regime and thus increasing the cooling power density.[16]

Another important nonideal effect for thin film coolers is the electrical contact resistance. Because the InGaAsP-based layers have a low electrical resistivity ($\rho = 1.17 \times 10^{-3}\ \Omega$ cm) the contact resistivity usually dominates. This additional Joule heating occurs very close to the cooling region of the device and must be minimized to attain any appreciable cooling. In order to determine the quality of our ohmic contacts, contact resistivity studies were performed with four point probe measurements on transmission line model test patterns.[17] The lowest measured value of specific contact resistance was roughly $5.5 \times 10^{-7}\ \Omega\ \mathrm{cm}^2$ for alloyed Ni/AuGe/Ni/Au contacts to 0.5-μm-thick $n+$InGaAs ($2 \times 10^{19}\ \mathrm{cm}^{-3}$). By studying various annealing conditions, it was determined for the InGaAs alloyed contacts that rapid heating and cooling produced the lowest contact resistance in agreement with previous literature.[18,19] The alloying depth should also be considered when designing thin film coolers. This depth is equal to the amount of semiconductor that intermixes with the contact metal and is roughly 0.1–0.2 μm in most III–V systems.[18–20] In our designed thin film coolers, the top $n+$ or $p+$ contact regions are as short as possible for low resistance but safely greater than the expected alloying depth. Nonalloyed contacts would circumvent this situation, however they typically have a higher contact resistance and the trade-off between the contact resistance and the resistance of a thinner contact region should be considered. The effects of contact resistance on cooler performance are discussed further in the next section.

III. COMPLETE DEVICE SIMULATION

Complete three-dimensional device structures with all nonideal effects included are simulated to fit experimental data and determine which areas of the thin film cooler designs need to be improved. These simulations model self-consistently the thermal and electrical operation. Once the simulation is in agreement with experimentally measured temperature profiles, particular nonideal effects can be removed in succession and the dominating ones determined.

Measured cooling for a 1-μm-thick, $50 \times 100\ \mu\mathrm{m}^2$ thin film cooler (best device from Fig. 3) is shown with the simulated curve in Fig. 4 (curve 1). To check for the presence of thermionic cooling, reference samples with no barrier (InGaAs only) were tested. The reference sample displayed a cooling that was a factor of 2 smaller. Since the reported thermal conductivity of quaternary InGaAsP is only 20%–

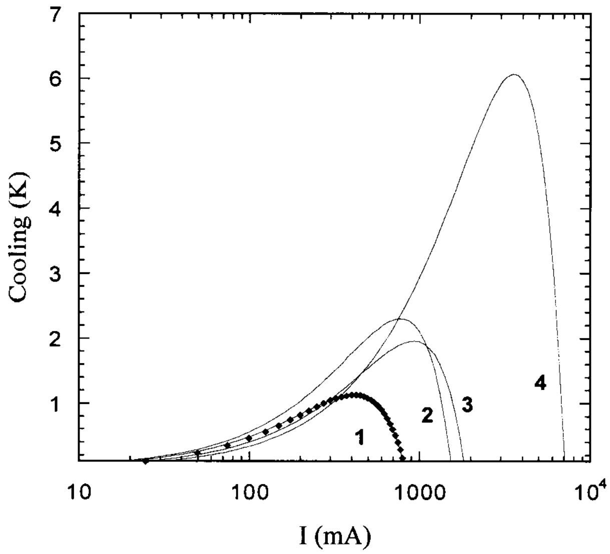

FIG. 4. Experimental and simulation results for the 1-μm-thick cooler shown in Fig. 3. The points and curve 1 correspond to the simulated fit of the experimental data with all nonideal effects. Curves 2 and 3 are the repeated simulation results when the contact resistance and substrate thermal resistance are taken away, respectively. Curve 4 corresponds to both nonideal effects removed.

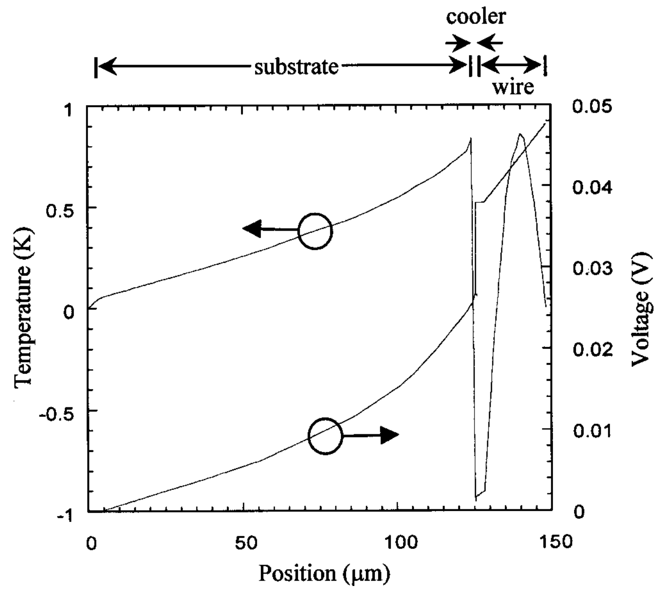

FIG. 5. Simulated temperature and voltage profiles along a path through the substrate, cooler, and wire bond. The actual wire bond length was 400 μm, but was scaled by 20× to reduce the number of elements in the 3D simulation.

30% less than InGaAs, one expects that thermionic cooling plays a significant role in the device operation. Both the thermionic and thermoelectric cooling effects were included in the simulation and the thermal conductivity of the InGaAs/InGaAsP superlattice was taken to be 5 W/m K. With an accurate model of the device in hand, the contact resistance was set to zero and the simulation was repeated (curve 2). The maximum cooling temperature increased from 1.14 to 2.3 °C. This was repeated with the contact resistance reset to its original value and the 120-μm-thick InP substrate replaced with a 10 μm copper substrate resulting in a similar improvement in performance (curve 3). The simulation was then again performed with the top wire bond removed (not shown), resulting only in a very small increase in cooling. Finally, the simulation was repeated once more with both the contact resistance removed and the copper substrate resulting in a maximum cooling temperature of 6.07 °C or a cooling power density of 1821 W/cm^2 (curve 4). Therefore in this particular device structure, both the contact resistance and the substrate thermal resistance will need to be reduced to see substantial improvement in performance.

It is insightful to examine the temperature and voltage profiles along a line through the 3D structure. Figure 5 shows simulation results following a path from the bottom of the substrate, traveling through the cooler, and continuing to the end of the wire bond. The current source is applied between the end of the wire bond and the bottom substrate plane, and an ideal heat sink is assumed at these two points. The wire bond (25 μm diameter) was 400 μm long, but was scaled by a factor of 20 to reduce the number of elements in the 3D simulation. A linear voltage drop and quadratic temperature distribution is observed along the wire bond as expected for one-dimensional behavior. At the cooler, a steep temperature gradient is observed across the 1-μm-thick film indicating a large heat flux. There is also a sharp voltage drop located at the interface of the metal and semiconductor corresponding to the contact resistance which was assumed to be $6 \times 10^{-7}\ \Omega\ \mathrm{cm}^2$. This voltage drop is nearly one quarter the total voltage across the entire structure. Inside the substrate, the voltage and temperature profiles can be seen to drop off nonlinearly due to the three-dimensional spreading. Finally, near the bottom of the substrate there is another change in the slope due to a 4-μm-thick tin solder layer.

From the three-dimensional modeling, it is possible to extract the important thermal and electrical spreading resistances similar to Fig. 2. Using these simulated values it becomes possible to construct an accurate one-dimensional model. Figure 6 shows the 1D model and boundary conditions applied to the heat flux equation

$$\kappa_x A_x \frac{d^2 T}{dx^2} = \frac{I^2}{\sigma_x A_x}, \tag{3}$$

which relates the temperature gradient to the Joule heat generation where σ_x, κ_x, and A_x, are the electrical conductivity, thermal conductivity, and area, respectively, defined in each of the three regions. The right side of Fig. 6 represents the equivalent circuit model with the thermionic (Q_{TI}) and thermoelectric (Q_{TE}) cooling and heating sources as well as contact resistance (Q_C). The thermionic cooling and heating occur on their respective sides of the thin film, however only one thermoelectric source is included at the upper metal-semiconductor interface since the bottom semiconductor-metal interface occurs at the heat sink. Equation (2) is valid in each region, and by integrating twice with respect to position an expression for temperature with two unknown constants results, giving a total of six unknowns. Four boundary conditions exist for the temperature since it must be continuous across all three regions. Here thermal boundary resistances are neglected. The last two boundary conditions are

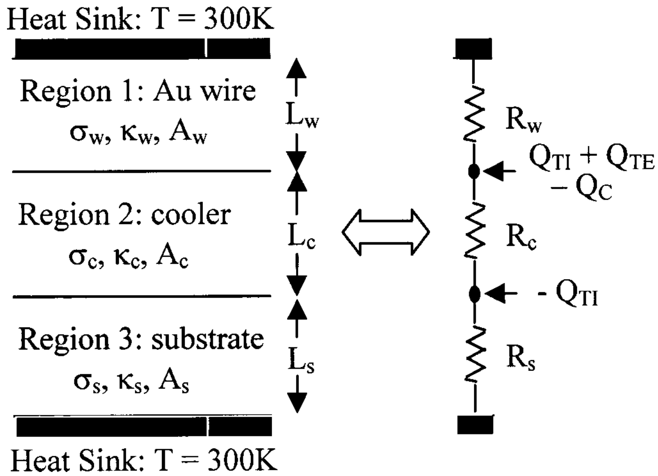

FIG. 6. One-dimensional model and boundary conditions. An electrical conductivity (σ), thermal conductivity (κ), and area (A), are defined in each region. The equivalent circuit model is shown on the right with the arrows indicating sources or sinks of heat flux. Q_{TI} refers to thermionic heating/cooling, Q_{TE} to thermoelectric cooling, and Q_C to heat generation by contact resistance.

obtained from the discontinuity of the heat conduction ($\kappa A dT/dx$) by the heat generation and absorption on either side of the cooler, resulting in six boundary conditions and six unknowns. The solution was then manipulated and plotted in MATLAB.

Figure 7 compares the results of the 3D and 1D models. The two simulations were found to be in good agreement over various changes in parameters. The largest difference can be seen in the temperature profile through the substrate where the 1D and 3D distributions are inherently different. The other minor discrepancy is in the temperature distribution of the wire that arises from the additional thermal spreading at the wire-cooler interface. This effect is more difficult to include in the 1D model.

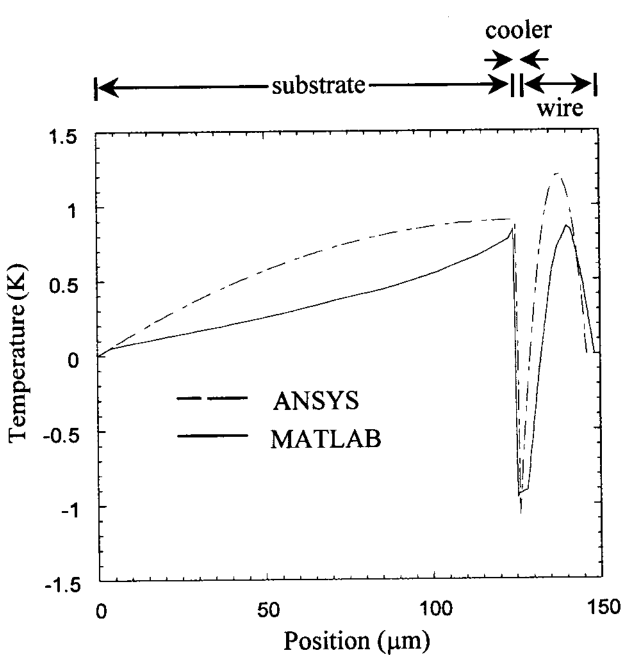

FIG. 7. Comparison of simulated temperature profiles through the substrate, cooler, and wire bond for the three-dimensional (ANSYS) and one-dimensional (MATLAB) models. The zero temperature axis is relative to 300 K.

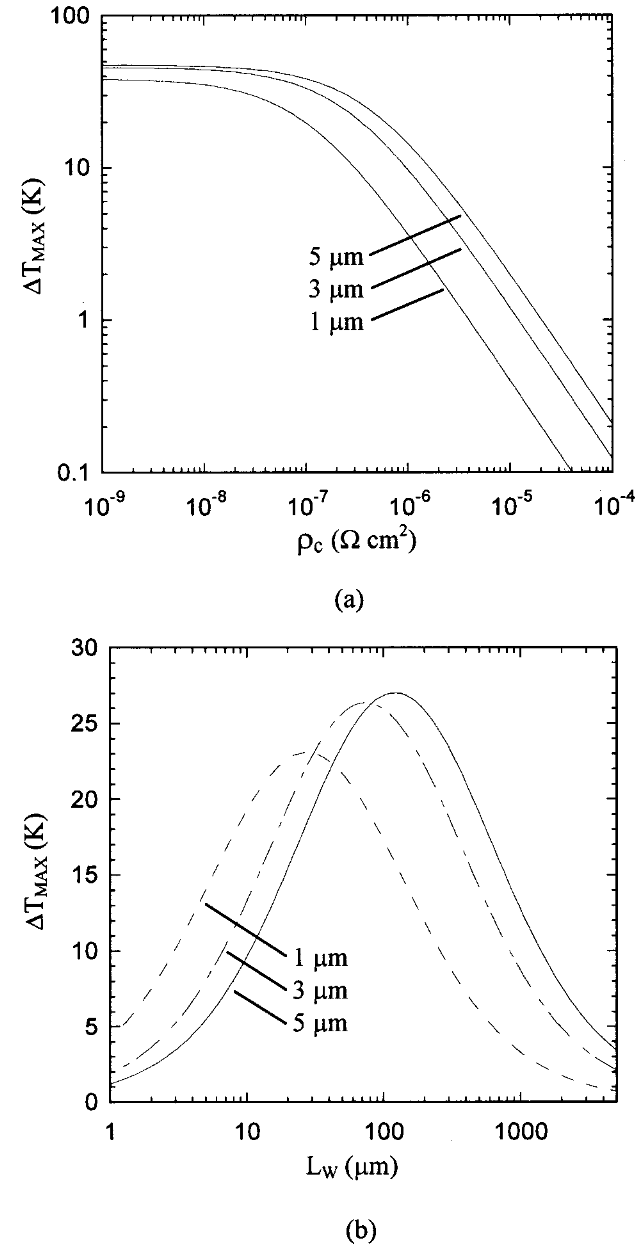

FIG. 8. (a) Simulation of maximum cooling vs contact resistivity for various cooler thicknesses. The Au wire bond was 50 μm long and 50 μm in diameter and zero heat conduction though the wire was assumed as in a p and n cooler configuration. (b) Simulation of single stage maximum cooling vs Au wire bond length (25 μm diameter wire) where now heat conduction through the wire is considered. The contact resistance was assumed to be $10^{-8}\ \Omega\ \mathrm{cm}^2$. The substrate thermal resistance was 4 K/W in each case.

The advantage of the one-dimensional model is the speed with which many device parameters can be varied and their affects determined. Figure 8(a) shows a simulation of maximum cooling versus contact resistivity for various cooler thicknesses. To observe just the limitation of contact resistance, the other nonideal effects were minimized to reasonable values. The substrate thermal resistance was assumed to be 4 K/W which from Fig. 2 could represent either a thin 15 μm Cu substrate or an arbitrarily thick diamond substrate. The wire bond was assumed to be 50 μm long and 50 μm in diameter with no heat conduction through the wire as would be in a *p*- and *n*-type conventional thermoelectric configuration. The maximum cooling drops off significantly for a contact resistivity above $10^{-7}\ \Omega\ \mathrm{cm}^2$. Also, thicker devices cool more when the contact resistance is substantial. The reason being that the optimum current for the cooler performance scales as one over the device thickness. A lower current results in less Joule heating from the contact resistance. Optimum currents for the thicker devices were less than 1 A, while the thinner devices required as much as 5 A. While the thicker devices do attain higher absolute cooling in this case, the cooling power remains approximately unchanged.

In applications for which single element coolers are monolithically integrated with electronic or optoelectronic devices, it is necessary to have an external wire bond connected directly to the cold side of the cooler. The question arises as to whether useful cooling can still be achieved in this configuration. Figure 8(b) shows a simulation of maximum cooling versus wire bond length (25 μm diameter). In this case heat conduction through the wire is considered, and the contact resistivity is assumed to be $10^{-8}\ \Omega\ \mathrm{cm}^2$ in order to study the effects of the wire only. For a given cooler thickness, there exists an optimum wire length resulting from a trade-off between Joule heating in long wires and heat conduction from heat sink to the cold junction in short wires. At longer wire lengths the thicker devices cool better by the same argument made for contact resistance. That is thicker coolers are optimized at a lower current and hence less Joule heating occurs. The cooling power of the thicker devices is reduced somewhat, however. For shorter wire lengths the thinner cooler performs best. Since the thinner cooler has a smaller thermal resistance between cold and hot junctions, more heat conduction occurs across the cooler than through the wire.

The temperature dependence of material properties and cooling amount are not taken into account in the above models. For small values of cooling and heating on the order of 5–10 °C, this is a reasonable assumption. However, these effects will need to be considered in future simulations.

IV. CONCLUSIONS

Important parameters and nonideal effects in thin film coolers have been discussed through experimental and simulation results. A three-dimensional finite element simulation has been developed and used to determine the dominating nonideal mechanisms for thin film coolers and the impact of changing device characteristics. A one-dimensional simulation was also developed using three-dimensional spreading resistance values obtained from the three-dimensional model. Contact resistance, finite thermal resistance of substrate and heat sink, and heat generation in wire bonds have all been identified as limitations in thin film cooler performance. Experimental results in thin film thermionic emission coolers have demonstrated cooling by 1.1–1.2 °C at room temperature with cooling power densities of several 100 W/cm^2, and simulations have predicted cooling of 20–30° with cooling power densities of several 1000 W/cm^2 for more optimized structures and packaging.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research under Contract No. 442530-25845, and the Army Research Office through the DARPA/HERETIC program under Contract No. 442530-23002.

[1] L. Rushing, A. Shakouri, P. Abraham, and J. E. Bowers, Proceedings of the 16th International Conference on Thermoelectrics, Dresden, Germany, August 1997, pp. 646–649.
[2] T. A. Corser, 41st Electronic Components and Technology Conference, Atlanta, GA, May 1991, pp. 150–156.
[3] A. Shakouri, C. LaBounty, P. Abraham, J. Piprek, and J. E. Bowers, Mater. Res. Soc. Symp. Proc. **545**, 449 (1999).
[4] A. Shakouri, C. LaBounty, J. Piprek, P. Abraham, and J. E. Bowers, Appl. Phys. Lett. **74**, 88 (1999).
[5] G. Chen, Phys. Rev. B **57**, 14958 (1998).
[6] P. Hyldgaard and G. D. Mahan, Phys. Rev. B **56**, 10754 (1997).
[7] ANSYS Release No. 5.5.3 (1999) by Swanson Analysis Systems, Inc., Houston, PA.
[8] MATLAB Release No. 5.0.0.4073 (1996) by The Math Works, Inc.
[9] L. Coldren and S. Corzine, in *Diode Lasers and Photonic Integrated Circuits* (Wiley, New York, 1995), pp. 55–57.
[10] F. Masana, IEEE Trans. Compon., Packag. Manuf. Technol., Part A **19**, 539 (1996).
[11] J. C. Fan, C. P. Lee, C. M. Tsai, S. Y. Wang, and J. S. Tsang, J. Appl. Phys. **83**, 466 (1998).
[12] C. B. Morrison, R. L. Strijek, and J. H. Bechtel, Proc. SPIE **3005**, 120 (1997).
[13] A. Yi-Yan, W. K. Chan, T. S. Ravi, T. J. Gmitter, R. Bhat, and K. H. Yoo, Electron. Lett. **28**, 341 (1992).
[14] E. Yablonovitch, D. M. Hwang, T. J. Gmitter, L. T. Florez, and J. P. Harbison, Appl. Phys. Lett. **56**, 2419 (1990).
[15] D. M. Rowe, *CRC Handbook of Thermoelectrics* (CRC, New York, 1995).
[16] A. Shakouri and J. E. Bowers, Appl. Phys. Lett. **71**, 1234 (1997).
[17] G. K. Reeves and H. B. Harrison, IEEE Electron Device Lett. **EDL-3**, 111 (1982).
[18] M. Ogawa, J. Appl. Phys. **51**, 406 (1990).
[19] G. Y. Robinson, Solid-State Electron. **18**, 331 (1995).
[20] G. K. Reeves, P. W. Leech, and H. B. Harrison, Solid-State Electron. **38**, 745 (1995).