INVITED PAPER

Overview of Solid-State Thermoelectric Refrigerators and Possible Applications to On-Chip Thermal Management

Thermoelectric cooling can be fairly easily added to conventional heat sinks and, with improved materials and processing, very effective hot-spot cooling can be achieved.

By Jeff Sharp, Jim Bierschenk, and Hylan B. Lyon, Jr.

ABSTRACT | The concept of thermal management in microelectronic components is changing, and so is the potential for solid-state cooling to solve emerging problems. There is a qualitative change that differentiates the past from the future. We discuss past practices and future trends in the electronic cooling markets, setting the stage for an outline of designs and processes that provide new and enabling options. We briefly review the science that is empowering these changes, and conclude with some thoughts on the future direction of thermal management of microelectronics.

KEYWORDS | Chip cooling; coefficient of performance (COP); heat sinking; heat spreading; solid-state cooling; spot cooling; thermal management; thermoelectric; *ZT*

NOMENCLATURE

A	area of TE element perpendicular to current flow
α	Seebeck coefficient
COP	coefficient of performance
ΔT	difference between hot side and cold side temperatures
ΔT_{max}	maximum temperature difference between hot and cold side with no heat load
I	electrical current
K	thermal conductance of TE element
λ	thermal conductivity
L	length of TE element, thickness of TE wafer
N	number of P/N thermoelectric couples in TEC
Q_c	net heat pumped by TEC
Q_{max}	maximum heat pumping of TEC at 0 ΔT
q	density of heat pumped
R_c	electrical contact resistance
R	electrical resistance of TE element
ρ	electrical resistivity
R_I	interconnect resistance
S	length of metal interconnect in TEC
T_{avg}	average of hot and cold side temperatures
T_c	cold side temperature
TEC	thermoelectric cooler
TE	thermoelectric
T_h	hot side temperature
T_{sink}	heat sink temperature
TIM1	Thermal interface material contacting the die
t	thickness of metal interconnect in TEC
V	total voltage of TEC
w_s	width of TE element
Z	figure of merit of TE material
ZT	dimensionless figure of merit
θ_{SA}	Sink to ambient thermal resistance

Manuscript received February 1, 2005; revised February 8, 2006.
J. Sharp and **J. Bierschenk** are with Marlow Industries, Inc., Dallas, TX 75238 USA (e-mail: jsharp@marlow.com).
H. B. Lyon, Jr. is with Gamma Design, Richardson, TX 78050 USA.
Digital Object Identifier: 10.1109/JPROC.2006.879795

I. INTRODUCTION

Microelectronic integrated circuits can be cooled below ambient temperatures to increase their clock speed. Historically, this has been the motivation for using thermoelectric coolers (TEC) in certain niches of the computer market. Now, the motivation is different and, with

significant improvements in TEC technology, interest in thermoelectric cooling is more widespread. Computer makers are faced with the limit of forced-air cooling due to the continuing increase in chip power and power density. In order to stay on the path described by Moore's Law, the industry is now considering augmenting or replacing forced-air cooling. In this paper, we describe different ways in which thermoelectric devices can be used to enable chip thermal management.

While the surge in power density has led to interest in thermoelectric chip cooling, it also means that we no longer should consider only approaches based on simple attachment of traditional thermoelectric modules coupled to extruded aluminum heat sinks. The present levels of chip power and power density require innovative thermoelectric cooling solutions, and usually adequate concepts must be challenged in the areas of module design, module dimensions, module fabrication methods, thermal interfaces, and heat exchanger design. We briefly describe two approaches to the problem:

1) management of the entire chip load (thermoelectrically enhanced heat sink);
2) targeted (hot spot) cooling.

These two different approaches encompass more than an order of magnitude change in thermoelectric cooling density. Power density in thermoelectric cooling depends strongly on the thickness of the thermoelectric materials used, and so this dimension is a key scaling parameter. The power densities considered here, from ~3 to ~300 W/cm^2, span the range from nearly routine bulk material thickness down to typical thick-film dimensions. With recent advances, it is conceivable that this entire span can be addressed with bulk materials. Still, film-based thermoelectric devices also may be of interest due to higher ZT or other technical advantage.

The two scenarios for implementation of thermoelectric chip cooling differ in important ways. In the scenario where the entire chip heat load is managed, energy efficiency (COP) is of paramount importance. Computer makers are not receptive to designs that greatly increase the total heat load that must be dissipated by the heat sink. In the case of hot spot cooling, COP is less critical because the thermoelectric devices will be handling only a small fraction of the total chip power. For spot cooling, achieving the necessary power density and adequate heat spreading are key. We address these pivotal differences in the following sections that highlight basic design principles and implementation challenges.

II. BASIC DESIGN PRINCIPLES

Performance of a TEC is governed by a well-known set of equations that are covered in many different references, such as Goldsmid [1]. As such, a detailed derivation of these equations will not be presented, and the equations pertinent to the discussions of this paper will be presented with minimal explanation. The amount of heat that can be pumped by a thermoelectric is the net of three contributions:

1) the Peltier heat pumping, which is opposed by
2) conduction across the thermoelectric elements and
3) resistive heating within the elements.

$$Q_c = N\left[\overbrace{(\alpha_p-\alpha_n)IT_c}^{(a)} - \overbrace{K(T_h-T_c)}^{(b)} - \overbrace{\frac{1}{2}I^2R}^{(c)}\right] \tag{1}$$

where $R=\rho_p(L/A)+\rho_n(L/A)$ and $K=\lambda_p(A/L)+\lambda_n(A/L)$.

The number of p/n couples N and the thermoelectric element length (L) to area (A) ratio characterize the geometry of the TEC. When a current is applied to the TEC, the voltage generated, the sum of the usual IR portion, and an additional component based on the Seebeck coefficient and the temperature difference across the TEC, is given by

$$V = N\left[(\alpha_p-\alpha_n)(T_h-T_c)+IR\right]. \tag{2}$$

Similarly, the input power to the TEC is

$$W = N\left[(\alpha_p-\alpha_n)I(T_h-T_c)+I^2R\right]. \tag{3}$$

The coefficient of performance of the TEC is thus the available heat pumping capacity divided by the input power

$$\mathrm{COP} = \frac{Q_c}{W} = \frac{(\alpha_p-\alpha_n)IT_c - K(T_h-T_c) - \frac{1}{2}I^2R}{(\alpha_p-\alpha_n)I(T_h-T_c)+I^2R}. \tag{4}$$

Overall performance of the TEC is a function of the thermoelectric material properties. The usefulness of a material for thermoelectric refrigeration is based on a quantity referred to as the figure of merit Z, which is given by

$$Z = \frac{\alpha^2}{\rho\lambda}. \tag{5}$$

A cursory examination of (1) would indicate that in order to have the highest heat pumping, one would want to have a high Seebeck coefficient, low electrical resistance, and low thermal conductivity. From (5), the same combination of material properties maximizes the figure of merit.

Thermoelectric materials used for cooling purposes are almost exclusively based on alloys of bismuth telluride (Bi_2Te_3). Historically, bismuth telluride alloys have been grown using directional solidification from a melt. These fabrication processes produce a well-oriented polycrystalline material structure. While this crystalline structure does provide good thermal performance, it also places a limitation on element geometries (length-to-area ratios) because the crystalline materials are fragile.

For a given L/A, an optimum current exists based on the hot and cold side temperatures. This optimum current is the one that maximizes the TEC COP (4)

$$I_{\text{opt}} = \frac{(\alpha_p - \alpha_n)(T_h - T_c)}{R\sqrt{ZT_{\text{avg}} + 1} - 1}. \tag{6}$$

I_{opt} depends on the material properties, along with T_h, T_c, and L/A. Combining (4) and (6) provides the expression for the maximum (optimum) COP

$$\text{COP}_{\text{opt}} = \frac{T_c[\sqrt{1 + ZT_{\text{avg}}} - T_h/T_c]}{(T_h - T_c)[1 + \sqrt{1 + ZT_{\text{avg}}}]}. \tag{7}$$

It is extremely important to use TECs that operate near this optimum COP. Selection and use of a nonoptimum TEC will result in performance significantly poorer than that of an optimum TEC.

Using (7), the theoretical maximum COP can be plotted as a function of the thermoelectric element $\Delta T = (T_h - T_c)$. This is illustrated in Fig. 1 for materials with $ZT = 1$ and $ZT = 2$. As can be seen from the curve, COP values in excess of one are possible with today's best commercially available bulk materials ($ZT = 1$) for ΔT less than about 30 °C. In addition, the COP climbs significantly for a system designed to run optimally at even lower ΔT or with higher ZT materials.

A. Thermoelectrically Enhanced Heat Sinking

As traditional air cooled heat sinks approached their limit, numerous attempts were made to apply thermoelectrics to chip cooling applications. These attempts have largely been unsuccessful because off-the-shelf thermoelectric modules, with inadequate cooling power densities, were used in the analyses and tests [2]–[4]. These thermoelectric devices were not optimally designed for the heat load and ΔT used. When the optimum TEC is used, a net significant improvement in thermal performance can be achieved as reported by Intel [5], [6] and Marlow [7]. In these applications, the thermoelectric is operated over a

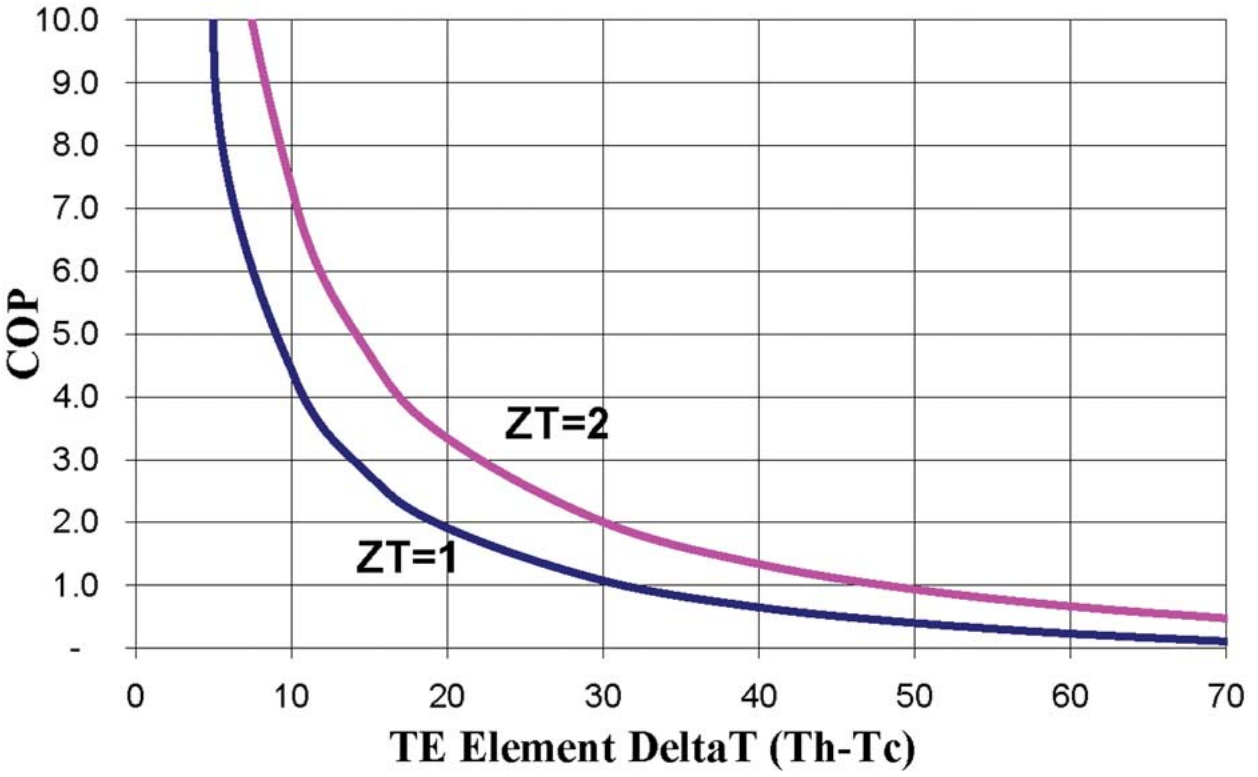

Fig. 1. *Typical device optimum (maximum) COP versus* ΔT.

relatively small ΔT and at a high COP (generally 2–3 or higher). Operating at the theoretical maximum COP is critical to minimize the additional heat generation within the thermoelectric. Any additional heat produced by the TEC must also be dissipated by the heat sink, resulting in larger heat sink temperature rises above the local ambient temperature. Further, the TEC introduces at least one additional thermal interface, and this additional thermal interface receives a larger heat flux due to the additional TEC input power. If, for example, the TEC were operated at a COP of 3, for a $ZT = 1$ material, the TEC ΔT would be around 15 °C. For a 100-W chip, an additional 33 W would flow through the heat sink. This additional rejected heat increases the temperature of the heat sink, partially negating the 15 °C temperature decrease that the TEC generated. If the sink to ambient thermal resistance were 0.25 °C/W, the additional rise would be 8.25 °C (0.25 °C/W × 33 W), leaving a net benefit of about 7 °C. In addition, there will be temperature rises across the thermal interfaces on the hot and cold sides of the TEC. These too would further lessen the 7 °C improvement.

Despite these losses, there is a window where the addition of the TEC results in a net improvement in effective performance of the heat sink as noted in [5]–[7]. In these cases, the TEC operates as an effective negative thermal resistance *even after account is taken of all losses*. Obviously, if the TEC in these cases does not operate at the maximum COP, additional input power will be required to achieve the same ΔT. This excess input power must be dissipated by the heat sink, and the exacerbated heat sink and interface temperature rises will quickly erase any benefit from adding the thermoelectric.

In the concept of the thermoelectrically enhanced heat sink, all of the heat from the CPU is pumped by the TEC. In most of these cases, the preferred embodiment has been to diffuse the heat as much as possible prior to reaching the TECs, meaning the TECs would cover essentially the entire fin area. This minimizes the watt density requirements of the TEC, minimizes the interface and ceramic thermal losses, and confines the highest heat flux (heat dissipated from the TEC) to the hot side interface and heat sink fins. This is the most thermally efficient configuration, but it also is the most expensive as it requires multiple thermoelectric coolers and generally an expensive vapor chamber or heat pipe assembly to spread the heat prior to reaching the TECs. Intel has shown that in an 80 mm × 120 mm TEC-based thermal solution, TECs could reduce the effective sink-to-ambient thermal resistance, θ_{SA} for a 120-W CPU, from 0.18 °C/W without the TECs to approximately 0.11 °C/W at a TEC COP of 3 and approximately 0.08 °C/W at a COP of 2 [8]. These COP values correspond to TEC input powers of 40 and 60 W, respectively, as illustrated in Fig. 2, where we use an effective "sink" temperature based on the TEC cold side. In this manner, an "effective" θ_{SA} can be compared directly to the no-TEC configuration.

An alternative implementation of thermoelectrically enhanced heat sinking is to manage the entire chip load with a single TEC, with minimal heat spreading taking place prior to the TEC and the preponderance of the spreading taking place between the TEC hot side and fins. The smaller TEC surface area minimizes the heat spreading losses between the CPU and the TEC. With most of the spreading occurring at a higher heat flux (because it is downstream from the TEC), this approach may not be as efficient as the multi-TEC approach. In addition, the TEC ceramic losses and TEC interface losses may be higher. *Still, heat sink enhancement with a single TEC may provide a more cost-effective solution and an alternative when one considers cost/performance tradeoffs since the thermoelectric and spreading costs are minimized.*

Regardless of the approach taken, the thermoelectrically enhanced heat sink offers the opportunity for significant reduction of acoustic noise, a key performance parameter in some applications. The TEC acts as a variable, negative thermal resistance. During nonpeak CPU usage (when CPU heat dissipation is reduced), the TEC can maintain a fixed CPU temperature either by operating at very low input power (very high COP) or by operating at normal input power (larger negative thermal resistance), allowing lower fan speed and reduced acoustic noise level. The need exists for a simple methodology to determine whether adding TECs to a thermal solution can improve performance. If thermoelectrics are not used under the right conditions, excess heat will be added to the system and overall system performance may actually be poorer than a system without TECs. It is difficult to generalize the thermal performance of a thermoelectric cooling system, since TEC optimization, heat sink thermal resistance, interface thermal resistance, material ZT, etc., all impact the TEC system performance. Detailed work has been performed showing the sensitivity of system performance as a function of heat sink resistance and other variables [9]. Any such analyses are only good for a given set of assumptions. The following simple approach can be used to determine whether operating conditions are such that adding TECs to a heat sink can make a significant improvement. Fig. 1 is used as the basis for this analysis. The best bulk materials available today are roughly at $ZT = 1$. If we assume that we will always choose thermoelectrics that operate at their maximum COP for the given ΔT (i.e., optimized TECs), then the TEC COP becomes a simple function of ΔT as shown in Fig. 1. For a COP = 2, an optimized TEC will provide approximately a 20 °C ΔT. For COP = 3, the ΔT is approximately 15 °C.

Consider the following example:

$$Q = 120\ \text{W} \quad \theta_{SA} = 0.3\ ^\circ\text{C/W} \quad T_{amb} = 35\ ^\circ\text{C}.$$

Without TECs, the heat sink temperature is 71 °C (35 °C + 0.3 °C/W × 120 W) = 71 °C. If we were to add

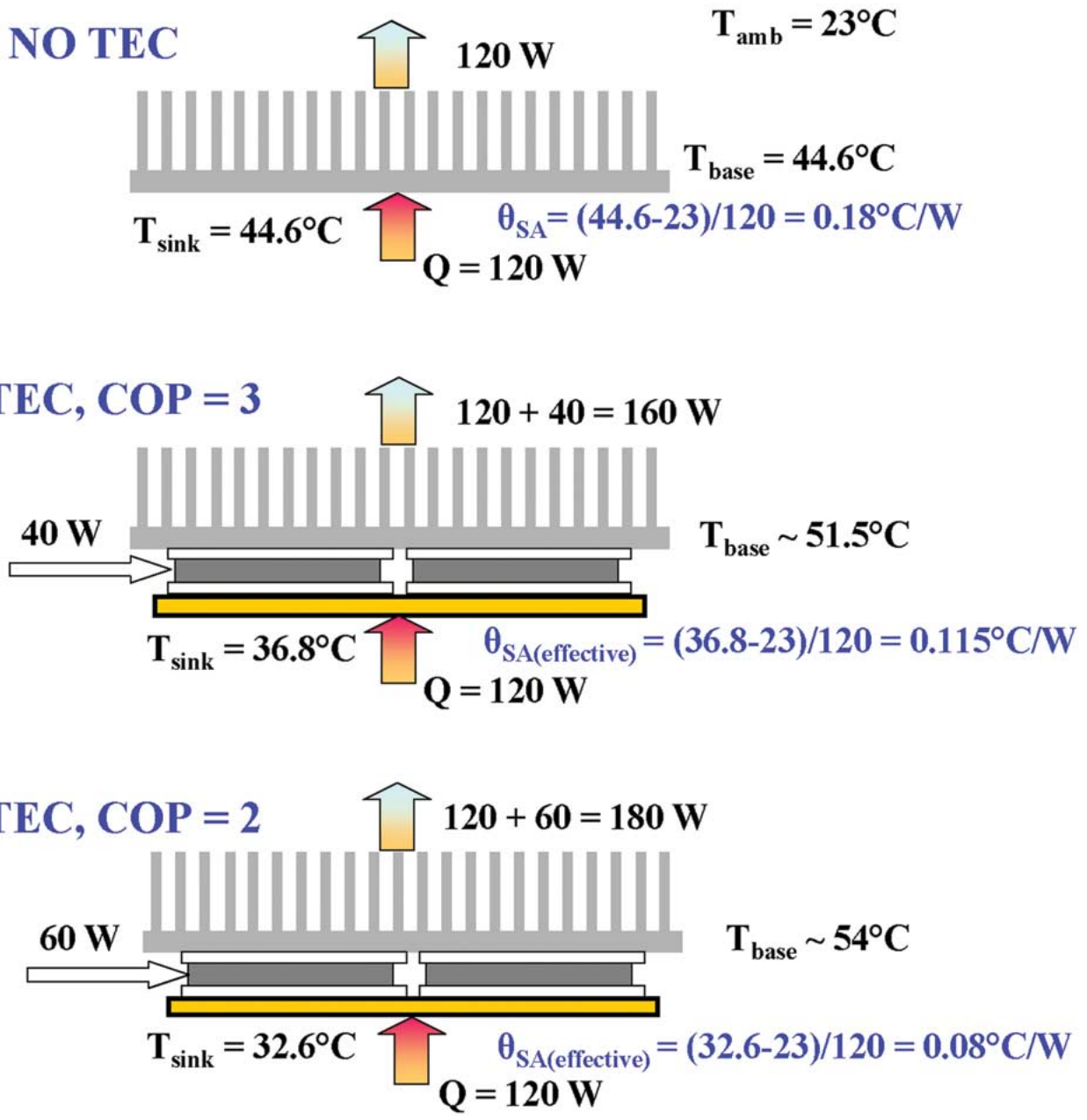

Fig. 2. ***Thermal profile of three different chip-cooling scenarios, showing the impact of TEC COP on the effective heat sink performance.***

thermoelectrics to this thermal solution and have 40 W of power available to the TEC (corresponding to a COP = 3), we have a TEC ΔT of 15 °C for an optimized TEC solution. In addition, the additional 40 W must be dissipated by the heat sink. The additional 40 W of heat dissipation raises the heat sink temperature by 12 °C (40 × 0.3 °C/W). Thus, the effective heat sink temperature (now the TEC cold side) is only improved by about 3 °C (−15 °C from the TEC + 12 °C rise from the heat sink) from 71 °C to 68 °C. The effective θ_{SA} is only marginally reduced from 0.30 to 0.275 ((68 °C − 35 °C)/120 W). If the TEC input power were increased to 60 W (corresponding to a COP = 2), the situation in this case gets worse. The TEC ΔT for an optimized TEC operating at a COP of 2 is approximately 20 °C. The additional heat sink rise jumps to 18 °C (60 W × 0.3 °C/W) and the net gain drops to 2 °C (−20 °C + 18 °C).

While the Intel data reported in [8] showed a significant improvement in heat sink performance, the above example shows when TECs probably should not be used. Both cases had the same heat load (120 W); however, in

the case of the Intel data, the heat sink was significantly better than the heat sink used in this example (0.18 °C/W versus 0.3 °C/W). This example highlights a general rule that TECs can improve the performance of a good heat sink but TECs added to a poor heat sink can actually make performance even worse.

It should be noted that a number of factors are ignored in the above simple analysis. If a more accurate assessment is needed, then the following additional items should be taken into account.

- TEC interfaces—adding the TEC can add an additional one or two thermal interfaces which must be added into the overall thermal resistance.
- Change in heat sink heat input area—many θ_{SA} values for heat sink are defined based on the hottest center temperature created by a concentrated heat source from the CPU. These heat sink numbers can improve significantly based on a more uniform heat input area from several TECs and can be used in the simple analysis methodology.
- Nonuniform temperatures—the previous analysis was based on a uniform temperature one-dimensional (1-D) simplification. Depending on the spreading thermal resistances, this may not be a valid assumption and a more detailed two-dimensional (2-D) or three-dimensional (3-D) analysis may be necessary.
- TEC Optimization—It should also be emphasized that the generalized COP versus ΔT curve in Fig. 1 represents a summary of optimized designs and the resulting ΔT's for COPs between zero and ten. It is not valid to use Fig. 1 to predict thermal performance of a given thermoelectric system operating under varying TEC input power (COP) conditions. In other words, it is not valid to assume that the same TEC designed to achieve a 20 °C ΔT at a COP of 2, will achieve a 15 °C ΔT at a COP of 3. The TEC module(s) required to achieve the latter will be different.

B. CPU Spot Cooling

Another thermal management option is to cool only the hottest spots on the die. In this configuration, not all the heat produced by the CPU must be pumped by the TEC, rather only the hot spots. Current CPUs have average heat fluxes of approximately 10–50 W/cm^2. Peak heat flux, or hot spots, can be up to six times these values, ranging between 100 and 300 W/cm^2. These hot spots drive the thermal design. As the heat load per unit area of the TEC increases, design optimization requires that the thermoelectric elements become increasingly short. As the thermoelectric elements become increasingly short, additional irreversible affects must be taken into account.

To develop a sense of when these effects become significant, it is useful to have a rule of thumb for TEC scaling. In a TEC with $ZT = 1$ materials, operating at maximum COP across a ΔT of 15 °C, the net heat pumping density (q, in W/cm^2) and TE material thickness (L, in cm) are related approximately by $L = 1/(5q)$. Here we have assumed a tightly packed TEC (for operation in air), with 2/3 of the footprint filled with TE material. Fig. 3 summarizes the relationship between TE material thickness and heat pumping density according to the above approximations and assumptions.

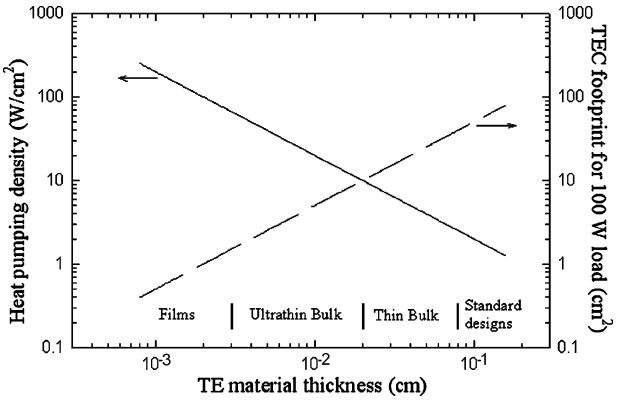

Fig. 3. *Density of heat pumped and required TEC footprint for a benchmark heat load, both as a function of TE material thickness. Assumptions, as described in the text, are $ZT = 1$ material, 2/3 packing fraction, $\lambda \sim 14$ mW/cm-K, $\Delta T = 15$°C, and TEC operating at maximum COP.*

This analysis reveals the point that on-chip spot cooling will require TE materials with thickness in the vicinity of 50 μm or less. This thickness range is a challenge for both bulk TE materials (normally much thicker) and thin-film materials (normally much thinner). In the case of thin-film materials, the packing fraction can be reduced to allow smaller L, but thermal conduction by air ($\lambda = 2.7 \times 10^{-4}$ W/cm-°C at 40 °C) causes a significant decrease in device ZT for a packing fraction less than 20%. Also, as discussed later, interconnect resistance becomes a major challenge in "open" device designs with low packing fraction.

Taking the TE material thickness to be 50 μm leads to further constraints on the TE module design. The thermal conductivity of AlN, the usual substrate choice in high power density applications, is 1.7 W/cm-°C. Because the COP is high, the thermal fluxes through the hot and cold substrates are comparable and we will assume they are equal. To avoid an excessive combined temperature drop across both substrates, the ratio of their thermal conductivity to their thickness should be $\sim$ 20 times the corresponding ratio for the TE layer ($\lambda = 0.014$ W/cm-°C). For AlN, the substrate thickness must be less than about 300 μm.

Likewise, as the thermoelectric element becomes increasingly short, the electrical contact resistance between the Bi_2Te_3 semiconductor and the metal contact

Table 1 Scale, Technology and Challenges of Thermoelectric Cooling as a Function of Watt Density

Thermoelectric Heat Flux (W/cm^2)	Element Length	Bulk	Thin Film	Challenges
1-6	~ 1 mm	X		Space, cost, spreading resistance, weight, supply chain
4-10	250 μm	X		Ceramic conduction resistance, interface resistance, contact resistance
100-300	5-25 μm	X	X	Contact resistance, heat sinking

layers begins to become a significant portion of the overall device resistance. For spot cooling, it will be especially important to minimize this loss.

With spot cooling, not all the heat from the CPU must be pumped by the TEC. As such, the operating COP of the TEC is not as critical since the heat dissipation from the backside of the TEC does impact the ceramic conduction (if present) and the TIM interface losses but it has a smaller impact on the heat sink temperature rise. Heat pumping capacities of 100–300 W/cm^2 are required for the TEC. This creates significant challenges. Table 1 highlights the impact of increasing heat flux on TE element height and challenges associated with scaling down.

III. IMPLEMENTATION CHALLENGES

A. Fundamental Challenges

There are two fundamental challenges involved in using thermoelectric devices at the high COP and/or watt densities typical of one or more of the scenarios we have described: improving material performance (ZT), and achieving low parasitic resistance.

The best ZT values validated in commercial applications belong to bulk materials based on alloys of Bi_2Te_3, Sb_2Te_3, and Bi_2Se_3. Recent advances have taken the ZT of these materials (P and N average) from 0.9 to 1.0 at 300 K. The transition from melt synthesis to solid-state synthesis that enabled this incremental improvement has also enabled the processing of bulk thermoelectric ingots into thin wafers. Consequently, 250-μm scale thermoelectric modules can be made by straightforward methods, and 25–50-μm modules are possible with proprietary innovations in manufacturing processes.

For more dramatic increases in ZT, most research emphasizes nanostructured materials. There are both theoretical motivations and laboratory results to encourage this emphasis [11], [12]. For this paper, it is instructive to examine novel, nanostructured thermoelectric materials from the point of view of how they might be incorporated into devices suitable for chip cooling.

Superlattice films of Bi_2Te_3/Sb_2Te_3 (P-type) and Bi_2Te_3/Bi_2Se_3 (N-type) reportedly possess average ZT values approaching 2.0 for through-plane transport [13]. Device fabrication begins with the deposition of TE superlattice films by metal–organic chemical vapor deposition (MOCVD) [14]. These layers are patterned to make arrays of individual legs, and P and N arrays are joined to form a variation of the usual series circuit. There are thickness limits for both the deposition and patterning processes. So far, MOCVD layers are limited to less than 10 μm in thickness, corresponding to extreme watt density and associated heat spreading and interface losses. Another group has demonstrated reasonably good material quality in sputtered films of thermoelectric materials that are typically 20 μm thick [15], and it may be that superlattice films also can be made by this approach. The ability to make high ZT layers at least 30 μm thick would be an important advance for commercialization.

High ZT values for in-plane transport have been announced for certain quantum well structures based on PbTe and fabricated by molecular beam epitaxy [16]. In this case, though, power densities are miniscule and the results will not impact chip cooling unless through-plane properties (unreported to date) are comparable and other technical and manufacturing challenges are solved. A number of research groups have begun efforts to mimic the observed ZT enhancement in nanostructured materials made by methods more amenable to commercialization, but this research is in the very early stages.

The nanowire array format is another promising approach to high ZT. Theoretical predictions of ZT increases [17] have not been confirmed experimentally, but the laboratory results to-date are very preliminary. Nanowire arrays made by electrochemistry can easily reach the 50–100-μm scale [18] suitable for hot spot cooling, a thickness regime that is challenging for both bulk materials and films deposited by vacuum techniques, but again the issue of time to commercialization emerges.

The second fundamental challenge, parasitic resistance, includes contributions from both interfaces and interconnects. Interfacial contact resistance R_c in commercial (bulk) thermoelectric devices is believed to be $\sim 10^{-6}$ Ω-cm^2. To understand the significance of this value, R_c must be compared to ρL. As a useful guide, we will say that contact resistance is significant when R_c exceeds 10% of $\rho L/2$. (We use $L/2$ rather than L because only half of the Joule heating in the TE legs is a thermal load at the cold side.) ρ is $\sim 10^{-3}$ Ω-cm for present TE materials, meaning that, with present contact technology, R_c becomes significant when L drops below 200-μm. TE materials are small bandgap, heavily doped semiconductors and, in principle, much lower values of R_c are achievable. Groups working with TE films and metal contacts deposited *in situ* report R_c values of 10^{-7} $\Omega - cm^2$ or less [13], but duplicating this result with the less pristine surfaces of bulk material wafers may require changes in current industry practices. Even with lower R_c values, however, TE film-based devices are equally or more challenged by contact resistance due to the limited thickness of the TE layers.

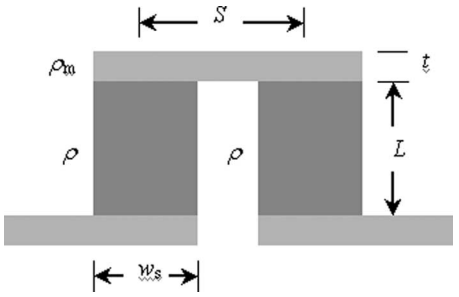

Fig. 4. *Schematic diagram of a single thermoelectric couple noting the parameters relevant to an estimate of interconnect resistance. For simplicity, the p and n TE materials are assumed to have the same resistivity ρ, and the dimension not shown (into the paper) is assumed to be the same for the TE legs and the interconnect. The interconnect resistivity is ρ_m.*

The impact of interconnect resistance R_I can be estimated by considering a single couple (refer to Fig. 4) joined by a metal layer of thickness t, length S (center-to-center of the TE legs) and resistivity ρ_m. R_I becomes a major parasitic loss when $\rho_m S/t$ is significant compared to $\rho L/w_s$. Here we have assumed that the interconnect shares one dimension with the TE legs that it connects, and the other dimension of the TE leg cross section is w_s. In a device with ~50% packing fraction, w_s is ~2/3 of S. Also, copper of ordinary purity offers a resistivity approximately 500X lower than that of TE materials. Combining these two approximations and again using a 10% rule, we see that R_I is significant when $t < S^2/75L$. For traditional scale TE modules with both S and L on the order of 1 mm, the interconnects can be as thin as 20 μm. If the module is scaled down but S^2/L is kept constant, then the constraint on t does not change. However, if L decreases faster than S^2, corresponding to TE legs suited for higher current levels, then the interconnect must be thicker, and the formation of such interconnects could be problematic, especially if they are to be formed by film deposition.

B. Engineering and Manufacturing Challenges

Foreseeable engineering and manufacturing challenges are dependent on whether one utilizes TECs for spot cooling or in a thermoelectrically enhanced heat sinking approach. The multi-TEC approach for a thermoelectrically enhanced heat sink can be implemented using existing thermoelectric technology. Implementation of a multi-TEC thermoelectrically enhanced heat sink may be limited more by economic and logistical challenges, whereas mainly technical challenges are projected for single-TEC enhanced heat sink or spot cooling.

The multi-TEC TE enhanced heat sink can provide a significant performance improvement when used in conjunction with a traditional air-cooled or liquid cooled system. The total cost of the thermal solution will be the single biggest limiting factor for the implementation of the multi-TEC approach. Typical thermal solution costs for present day PCs are around 10 cents per watt of CPU power. A 100-W CPU would generally be cooled by a $10 thermal solution, a 150-W CPU with a $15 thermal solution. The performance and noise benefits of the multi-TEC thermal solution must outweigh the additional TEC, spreader, power supply and temperature control costs. For this reason, these types of thermal solutions will likely find application only in higher performance systems, workstations or gaming PCs or possibly in home entertainment PCs, where the significant noise reduction is highly valued.

One option for creating a wider application of the TE enhanced heat sink solutions would be to reduce costs by driving the thermoelectric solution to a single TEC. In this manner, the cost of the TECs and spreader costs could both be reduced significantly. In order to keep from losing significant thermal performance, technical improvements are needed in order to minimize the larger thermal losses through the interface, and through the ceramics due to the higher heat flux. In addition, standard melt-grown TE materials cannot support the performance-driven thermoelectric dimensions noted in Table 1. Fine grain structured materials, without the inherent cleavage planes found in melt-grown materials, are required in order to produce elements that meet all the L values shown in Table 1. Historically, fine grain structured Bi_2Te_3-based materials suffered from poor thermoelectric efficiency, but breakthroughs have been made in this area. High performance, fine grain structured materials are used by the two largest suppliers of TECs in telecom applications [19], [20]. Scaling the manufacture of these materials to low cost and high volume, with equivalent thermal performance, will be a key step toward market acceptance in microelectronics thermal management.

Moving to the spot cooling approach offers a number of advantages. Spot cooling minimizes thermoelectric material consumption, cost, overall power consumption and additional heat rejection. This approach however, is not without its own set of challenges, including micro TEC fabrication, attachment, passive thermal losses, compatibility with TIM1 thermal interface materials, and power delivery.

One thermoelectric spot cooler manufacturer reported the amount of thermoelectric material would be reduced by a factor of 40 000 using a spot cooler versus a traditional thermoelectric [21]. Minimizing the thermoelectric material consumed should reduce the overall cost. In addition, since only the heat from the individual hot spots is pumped by the TEC, the TEC can be less sensitive to COP, and the additional TEC power can be a small percentage of the overall power consumed by the CPU.

The challenges associated with the spot cooling approach relate to the high heat fluxes on the backside of a CPU, nominally in the 100–300-W/cm^2 range. As discussed earlier, small micro thermoelectric devices that are capable of handling these heat fluxes require very short

thermoelectric elements in tightly packed arrays. One approach to fabricating these micro devices is to make the thermoelectric material fabrication and device structure compatible with normal semiconductor processes [10]. To date, the ΔT_{max} of such devices is small, less than 5–6 °C, which, with extra heat flowing through the remainder of the thermal path, is not enough to achieve a net reduction in hot spot temperature.

Spot coolers based on Bi_2Te_3 alloys must be made separately and attached to the die as discrete components. The advantages of integrated manufacturing are lost, but the superior ZT of these devices more than compensates an additional thermal interface. We now describe two distinct routes for the fabrication of spot coolers based on bismuth telluride.

The first route involves wafers processes, namely deposition of TE material films, deposition of metal layers for contacts, and patterning to form P and N arrays, which are then joined into a series circuit. There are many difficulties associated with this approach. In order to minimize the electrical current requirements, the thermoelectric element cross sections must be kept very small. For example, a thermoelectric element that is 5 μm thick having a 125-μm width, would have a theoretical I_{max} greater than 15 amps. A device with this I_{max}, even if operated at a low percentage of I_{max}, would require copper interconnects many times thicker than the thermoelectric material itself, in order to minimize parasitic electrical resistance that increases electrical power consumption and the thermal load that the TEC must overcome. To minimize these affects requires a shift toward more elements with smaller cross sections, but this direction leads to more difficult joining of the n and p element arrays, requiring micrometer-level precision in locating one array relative to the other. In addition, the substrate material itself, onto which the thermoelectric elements were fabricated can become a significant thermal resistance.

Increasing the TE material thickness can minimize the impacts of both parasitic electrical resistance and excessive thermal flux. This is a difficult and costly requirement for spot coolers based on wafer processes. The second route, namely spot coolers based on ultrathin bulk materials (< 100 μm), offers a way to achieve the optimum material thickness. At the scale required for spot cooling, typical bulk TEC assembly methods (placing individual TE elements in an array) are not attractive. Fabricating devices from cofabricated p-n wafers [22] provides a feasible means of producing tightly arrayed, high watt density TECs for spot cooling. Such devices could be fabricated as discrete components using modified wafer-processing techniques, segmented and then attached to the die in a separate operation. This approach has been used to

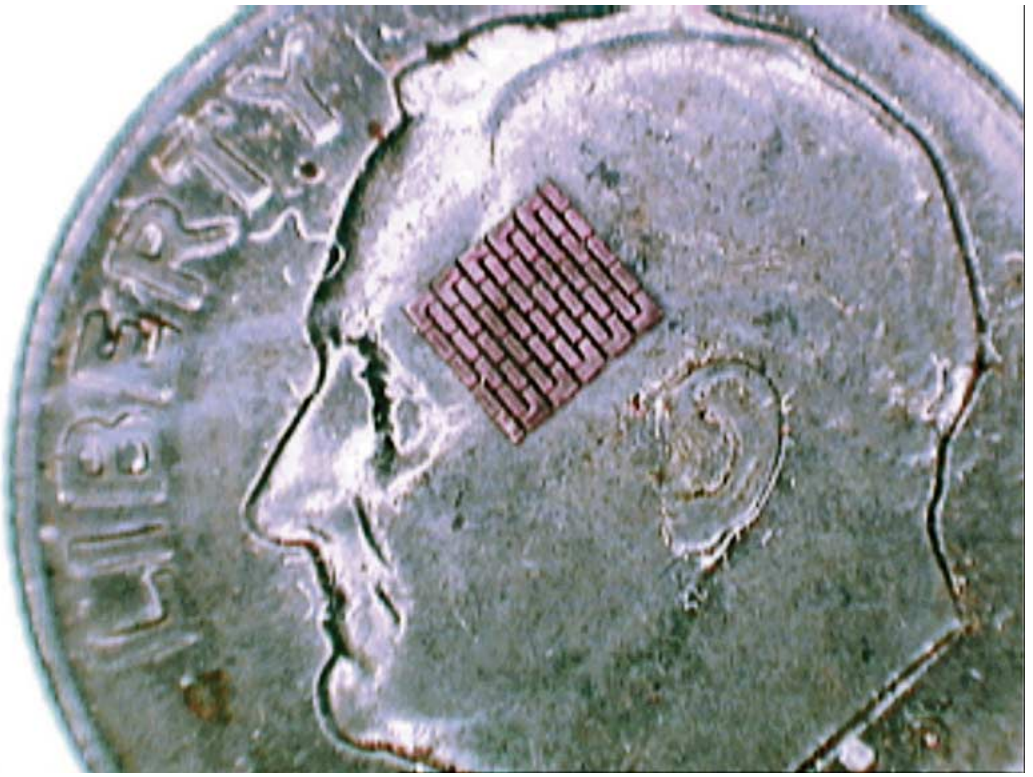

Fig. 5. *50-Couple miniature spot cooling bulk TEC shown on a dime.*

produce prototype 50 couple TECs in a 3.3 mm × 3.3 mm area with thermoelectric elements that are 25 μm tall. A sample device is shown relative to a dime in Fig. 5. Despite relatively poor bulk material electrical contact resistances, the prototypes had ΔT_{max} values approximately 50 °C and Q_{max} values of around 29 W (> 250 W/cm^2) at 100 °C hot side temperatures. Though challenges exist in the areas of wafer thinning and placement, scale down of bulk materials and associated development of new manufacturing processes may provide the most economical approach to reaching optimum TEC watt density for effective spot cooling. In addition, a significant performance improvement can be realized as the bulk material electrical contact resistance values are improved and approach the R_c values already attained in thin-film materials.

IV. CONCLUSION

In this paper we have offered a perspective on the applicability of thermoelectric cooling to the growing thermal management problem in the microelectronics industry. This perspective is rooted in a wealth of experience in thermoelectric applications at Marlow, but further shaped and balanced by our awareness of important changes taking place in the field of thermoelectrics. Recent advances in bulk materials and associated capabilities to shrink thermoelectric modules enable new solutions for chip cooling. The two distinct thermoelectric cooling solutions that we have outlined reflect these new possibilities. Also, hints of higher performance in new, nanostructured materials may lead in due time to large ZT improvements that greatly widen the window of opportunity for using thermoelectric cooling for thermal management of microelectronics.

While we have focused as much as possible on thermoelectric cooling, no component of the thermal path can be optimized without consideration of the overall thermal path from the critical spots on the die to the ambient air stream. There are ongoing advances in "passive" thermal management technologies such as interface materials, heat pipes and liquid cooling. A variety of economic and technical factors will determine, in a given application, whether these advances will enable thermoelectric cooling or make it unnecessary. Finally, these advances in potential microelectronics thermal management technologies are taking place at a time when the microelectronics industry is making a "right-hand turn" toward transistor, chip, and system designs that will lessen the thermal barrier to staying on track with Moore's Law [23]. With many changes occurring simultaneously in microelectronics design and thermal management technologies, and the push for more computing power in smaller volumes continuing, the topic of on-chip cooling is likely to remain an active one for years to come. ■

REFERENCES

[1] H. J. Goldsmid, *Electronic Refrigeration*. London, U.K.: Pion Ltd., 1985.

[2] R. C. Chu and R. E. Simons, "Application of thermoelectrics to cooling electronics: Review and prospects," in *Proc. 8th Int. Conf. Thermoelectrics*, 1999, pp. 270–279.

[3] E. P. Phelan, V. A. Chiriac, and T.-Y. Lee, "Current and future miniature refrigeration cooling technologies for high power microelectronics," *IEEE Trans. Compon. Packag. Technol.*, vol. 25, no. 3, pp. 356–365, Sep. 2002.

[4] ——, "Performance comparison of mesoscale refrigeration technologies for electronics packaging," presented at the Int. Electronic Packaging Technical Conf. and Exhibition (InterPack 2003), Maui, HI.

[5] I. Sauciuc, "Thermoelectric and phase change technology building block with application to CPU cooling," presented at the Intel Technology Symp., San Francisco, CA, 2004.

[6] I. Sauciuc, H. Erturk, G. Chrysler, R. Mahajan, Y. Kimura, H. Noda, and T. Nakamura, "Integration of thermoelectric and phase change (liquid/vapor) devices with application to CPU cooling," presented at the IMAPS Advanced Technology Workshop Thermal Management, Palo Alto, CA, 2004.

[7] J. Bierschenk and D. Johnson, "Extending the limits of air cooling with thermoelectrically enhanced heat sinks," in *9th Intersoc. Conf. Thermal and Thermomechanical Phenomena in Electronic Systems*, 2004, vol. 1, pp. 679–684.

[8] I. Sauciuc, H. Erturk, J. Williams, G. Chrysler, R. Mahajan, and V. Bala, "Thermal devices integrated with thermoelectric modules with applications to CPU cooling," in *Proc. ASME InterPack '05*, p. 73243.

[9] K. Fukutani and A. Shakouri, "Optimization of bulk thermoelectric modules for chip cooling applications," in *Proc. ASME InterPack '05*, p. 73410.

[10] Y. Zhang, J. Christofferson, A. Shakouri, G. Zeng, J. Bowers, and E. Croke, "High speed localized cooling using SiGe superlattice microrefrigerators," in *Proc. 19th Annu. IEEE Semiconductor Thermal Measurement and Management (SEMI-THERM) Symp.*, 2003, pp. 61–65.

[11] G. Chen, M. S. Dresselhaus, G. Dresselhaus, J.-P. Fleurial, and T. Caillat, "Recent developments in thermoelectric materials," *Int. Mater. Rev.*, vol. 48, no. 1, p. 45, 2003.

[12] G. S. Nolas, J. Sharp, and H. J. Goldsmid, *Thermoelectrics: Basic Principles and New Materials Developments*. Berlin, Germany: Springer-Verlag, 2001, ch. 8.

[13] R. Venkatasubramanian, E. Siivola, T. Colpitts, and B. O'Quinn, "Thin-film thermoelectric devices with high room-temperature figures of merit," *Nature*, vol. 413, p. 597, 2001.

[14] R. Venkatasubramanian, T. Colpitts, B. O'Quinn, S. Liu, N. El-Masry, and M. Lamvik, "Low-temperature organometallic epitaxy and its application to superlattice structures in thermoelectrics," *Appl. Phys. Lett.*, vol. 75, no. 8, pp. 1104–1106, 1999.

[15] H. Böttner, J. Nurnus, A. Gavrikov, G. Kühner, M. Jägle, C. Künzel, D. Eberhard, G. Plescher, A. Schubert, and K.-H. Schlereth, "New thermoelectric components using microsystem technologies," *J. Microelectromech. Syst.*, vol. 13, no. 3, p. 414, 2004.

[16] T. C. Harman, P. J. Taylor, M. P. Walsh, and B. E. LaForge, "Quantum dot superlattice thermoelectric materials and devices," *Science*, vol. 297, p. 2229, 2002.

[17] O. Rabin, Y.-M. Lin, and M. S. Dresselhaus, "Anomalously high thermoelectric figure of merit in $Bi_{1-x}Sb_x$ nanowires by carrier pocket alignment," *Appl. Phys. Lett.*, vol. 79, no. 1, p. 81, 2001.

[18] M. S. Sander, R. Gronsky, T. Sands, and A. M. Stacy, "Structure of bismuth telluride nanowire arrays fabricated by electrodeposition into porous anodic alumina templates," *Chem. Mater.*, vol. 15, p. 335, 2003.

[19] H. B. Lyon, Jr., Marlow Industries, Inc. (2005). The Marlow advanced thermoelectric material program. [Online]. Available: http://www.marlow.com/AboutMarlow/advanced_materials.htm

[20] K. Fukuda, Y. Sato, and T. Kajihara, "Thermoelectric semiconductor material, manufacture process therefor, and method of hot forging thermoelectric module using the same," U.S. Patent 6 274 802, Aug. 14, 2001.

[21] R. Venkatasubramanian, "Low-dimensional structures and superlattice thermoelectric technology," presented at the Applications and Commercialization of Next-Generation Thermal Management Materials and Systems Forum, Irving, TX, 2002.

[22] J. Sharp, "Thermoelectric device having co-extruded P-type and N-type materials," U.S. Patent 6 660 925, Dec. 9, 2003.

[23] Intel dual core processors, enabling your PC to think as fast as you can multitask, Intel Corp. (2005). [Online]. Available: http://www.intel.com/technology/computing/dual-core/

ABOUT THE AUTHORS

Jeff Sharp is Chief Scientist at Marlow Industries, Dallas, TX, and Leader of the Materials Research and Development team. He has nearly a decade of experience at Marlow in advanced thermoelectric material research. In recent years he has guided the initiation, development, and scale-to-production of Marlow's proprietary MAM formulation of bismuth telluride alloys while continuing research into new materials.

Jim Bierschenk is Chief Engineer at Marlow Industries, Dallas, TX, and Leader of the Product and Process Research and Development team. He has more than 17 years of experience designing thermoelectric coolers and thermoelectric cooling systems, with a recent focus on technologies and designs that will enable miniaturization and high watt density cooling.

Hylan B. Lyon, Jr. served 13 years with Marlow Industries, Dallas, TX, as Vice President of Research and Development, focusing on the development of new thermoelectric materials, devices and processes. He is currently President and COO of Gamma Design, Richardson, TX.