Dual-tapered-piston (DTP) module cooling for IBM Enterprise System/9000 systems

by G. F. Goth
M. L. Zumbrunnen
K. P. Moran

The water-cooled thermal conduction modules (TCMs) in the IBM Enterprise System/9000™ (ES/9000™) systems require a fourfold thermal improvement over TCMs in the 3090™ system. An examination of the thermal/mechanical tolerance relationships among the chips, substrate, and cooling hardware showed that a cylindrical piston would not meet this requirement. The piston was redesigned with a cylindrical center section and a taper on each end. This shape minimizes the gap between the piston and "hat" while retaining intimate contact between the piston face and chip surface during all assembly conditions. Numerical and analytical models demonstrate that this new piston shape, coupled with improved conductivity of the cooling hardware materials, exceeds ES/9000 system needs. These models were verified by tests conducted on single-site and full-scale modules in the laboratory and by tests on actual ES/9000 systems.

Introduction

Cooling of silicon circuit chips by conducting their heat to a water-cooled cold plate via a contacting piston was first developed for the IBM 308X mainframe computers and was enhanced for use in the IBM 3090™ systems [1, 2]. The chips are cooled within a thermal conduction module (TCM), which contains the contacting pistons and a multilayer substrate with about 100 chips mounted on one side and I/O pins attached on the other side. The TCM is attached to a mating cold plate receiving water from a dedicated coolant distribution system [3, 4]. In the manufacturing of the chips, substrates, cooling hat, and pistons, many tolerances are introduced which must be controlled for effective heat transfer. The fundamental thermal requirement is to provide efficient cooling of all chips while accommodating these tolerances. Additionally, the TCM is required to open and close easily for upgrades or rework and to be suitable for shipping and handling.

Significant improvements to the thermal/mechanical package were necessary to provide cooling to the new generation of modules used in the IBM Enterprise System/9000™ (ES/9000™) processor. These modules have

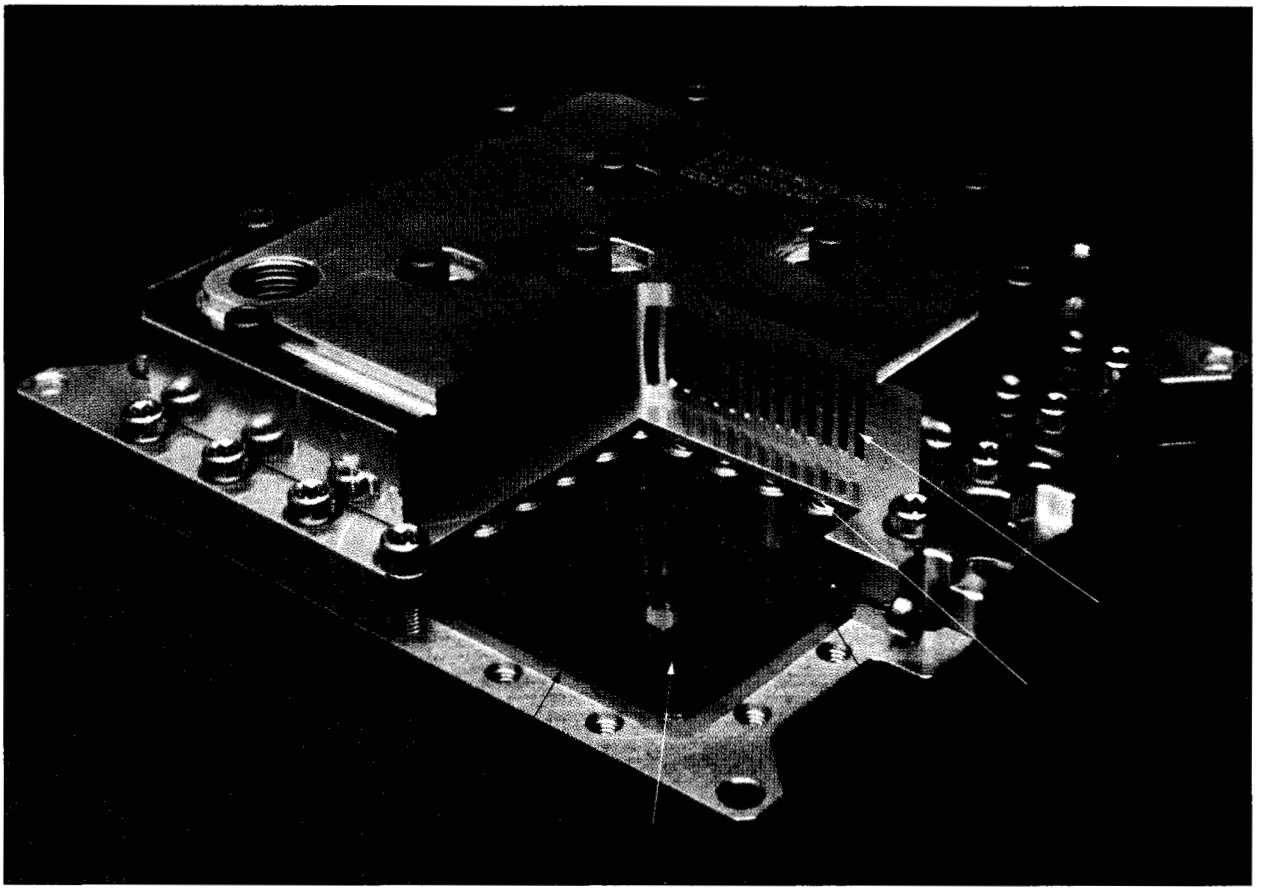

Figure 1

Cutaway view of a 121-site ES/9000 TCM.

either 100 or 121 chip sites at 10.8-mm and 9.9-mm spacing, respectively. Any site may have a maximum power of 27 W, and the maximum module power is 2000 W. To enhance reliability, a reduction in average and maximum device temperatures was required. The internal and external thermal resistances had to be reduced by a factor of 4 from the 3090 TCM levels to satisfy the power, temperature, and reliability objectives. That is, the internal resistance between the heat-generating circuit and the cold-plate mounting surface on the TCM cover or "hat" had to be reduced from about 5.0°C/W/chip to an ES/9000 average resistance of 1.21°C/W/chip. Similarly, the external resistance from the mounting surface to the cold-plate water had to be improved from 2.8°C/W/chip to about 0.7°C/W/chip. The ES/9000 system would consequently have an average device temperature near 50°C and a worst-case circuit temperature less than 75°C. To put the ES/9000 thermal demand in perspective, *the entire thermal resistance from a heat-generating circuit to water is less than the resistance across the gap from the chip to the anodized piston face in the 3090 system TCM!*

The first efforts to extend conduction cooling to the ES/9000 glass-ceramic modules concentrated on changing the heat-transfer medium between the piston and hat from helium to a high-conductivity paste [5]. These modules initially performed adequately, but degraded during accelerated life tests as the paste mixture separated into regions of solids and liquids. Since the required cooling was not achievable with a cylindrical piston, efforts focused on reducing the gap resistance by increasing the amount of heat-transfer surface area engaged between the piston and the hat hole. This was done either by using many small pistons/pins [6] or by using several narrow rectangular fins per chip site [7]. The multiple pistons/fins were joined to a common plate which contacted a chip; they were engaged into a hat having either round holes or rectangular channels. The multiplicity of components per chip site introduced additional locational tolerances which

required increasing the gap between the fins or pistons and the hat. As a result, some of the thermal benefit gained by adding surface area was lost by widening the gap. These multiple-surface designs introduced a higher level of manufacturing complexity. Hence, a study of the thermal/mechanical tolerances was continued, eventually leading to the development of the dual-tapered-piston (DTP) TCM concept [8].

Dual-tapered-piston design

A cutaway view of an ES/9000 TCM, exposing the substrate, chips, base plate, DTP cooling hardware, and cold plate, is shown in **Figure 1**. The dual-tapered pistons are held in contact with the chips by low-force conical springs located behind the pistons within the hat holes. Each piston has a small axial vent machined into its side to facilitate oil fill and disassembly. The function of dual tapers is shown in **Figure 2**, with the tapers exaggerated for clarity. The left chip site illustrates a condition of zero chip tilt (i.e., perfect assembly of the parts), and the right depicts process conditions for the worst case. Also shown (dashed outline) are cylindrical pistons which have larger gaps between the pistons and hat holes. The cylindrical piston face in previous TCMs was unable to fully accommodate the tilt caused by the chip and other mechanical tolerances, whereas the DTP is designed with an upper tapered section, a middle cylindrical section, and a lower tapered section for tilt accommodation. The face of each dual-tapered piston contacting the chip has a large spherical crown that creates a submicron gap across the chip back-side surface. Since practically all of the chip power passes through the small surface area of the chip–piston interface, a thin gap must be maintained under all conditions to achieve high thermal conductance. The length of each section of the piston and the angles of the tapers are determined from a tolerance analysis that is described below. The maximum mid-section diameter provides adequate clearance for the piston to translate freely without sticking in even the smallest hat hole, while the piston length is determined by the amount of chip tilt that must be accommodated. By accommodating all chip and hardware tilts, the DTP retains intimate contact between the piston face and the chip surface under all conditions. The diameter at the piston ends is equal to the diameter of a conventional cylindrical piston that could also accommodate these tilt tolerances. The DTP shape displaces a larger portion of the low-conductivity gap-filling material with high-conductivity piston material, resulting in a 50% lower gap thermal resistance. The improved heat transfer from the chips is a consequence of small radial gaps plus intimate contact with the chip surface.

Additional features of the TCM that have been changed from TCMs in the 3090 system include the interstitial heat-transfer medium, which was helium and is now a highly conductive oil; the piston, which was aluminum and is now made of copper; the aluminum used in the hat, which has a higher conductivity than in the 3090 system; and the hat holes, which are now anodized to ensure electrical isolation between chips. The ES/9000 Type 9121 air-cooled processors also used these features in their 121-chip-site TCMs, although the pistons were cylindrical rather than tapered, thereby resulting in an internal resistance of 1.7°C/W [9]. The oil has two thermal advantages over helium. In contrast to helium, long-term leakage of oil is not significant, nor has it a conductivity reduction in the very narrow chip-to-piston-interface gap [10]. The copper has a conductivity more than double that of the aluminum piston in the 3090 system. Finally, by anodizing the hat hole, the large anodization thermal resistance on the crown of the 3090 piston is eliminated, and only a small thermal penalty remains.

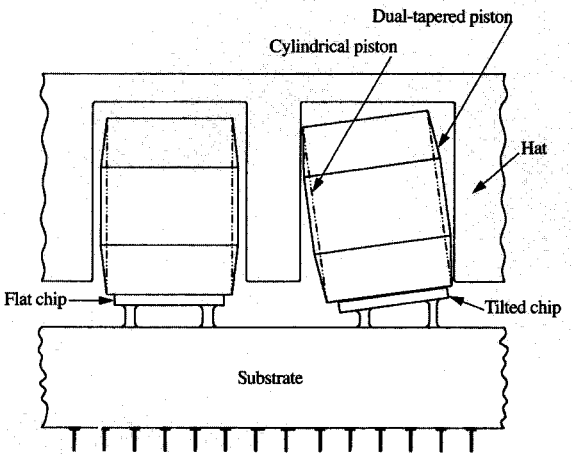

Figure 2

DTP in contact with flat chip (left) and with tilted chip (right).

- *Mechanical design*

Key dimensions affecting the piston/hat assembly are shown in **Figure 3**. Each dimension has an associated machining tolerance, $\pm \delta X_{xx}$, where X_{xx} is any dimensional parameter illustrated in the figure. These dimensions and tolerances are selected so that the tightest clearance accommodates all tilt conditions and is still easily assembled, yet the loosest fit provides adequate thermal performance. A mechanical analysis coupled with thermal modeling and supplier feedback was used to optimize ease of manufacture and thermal performance for each tolerance.

To begin the TCM design, a clearance between the piston and hat was chosen within which the pistons could

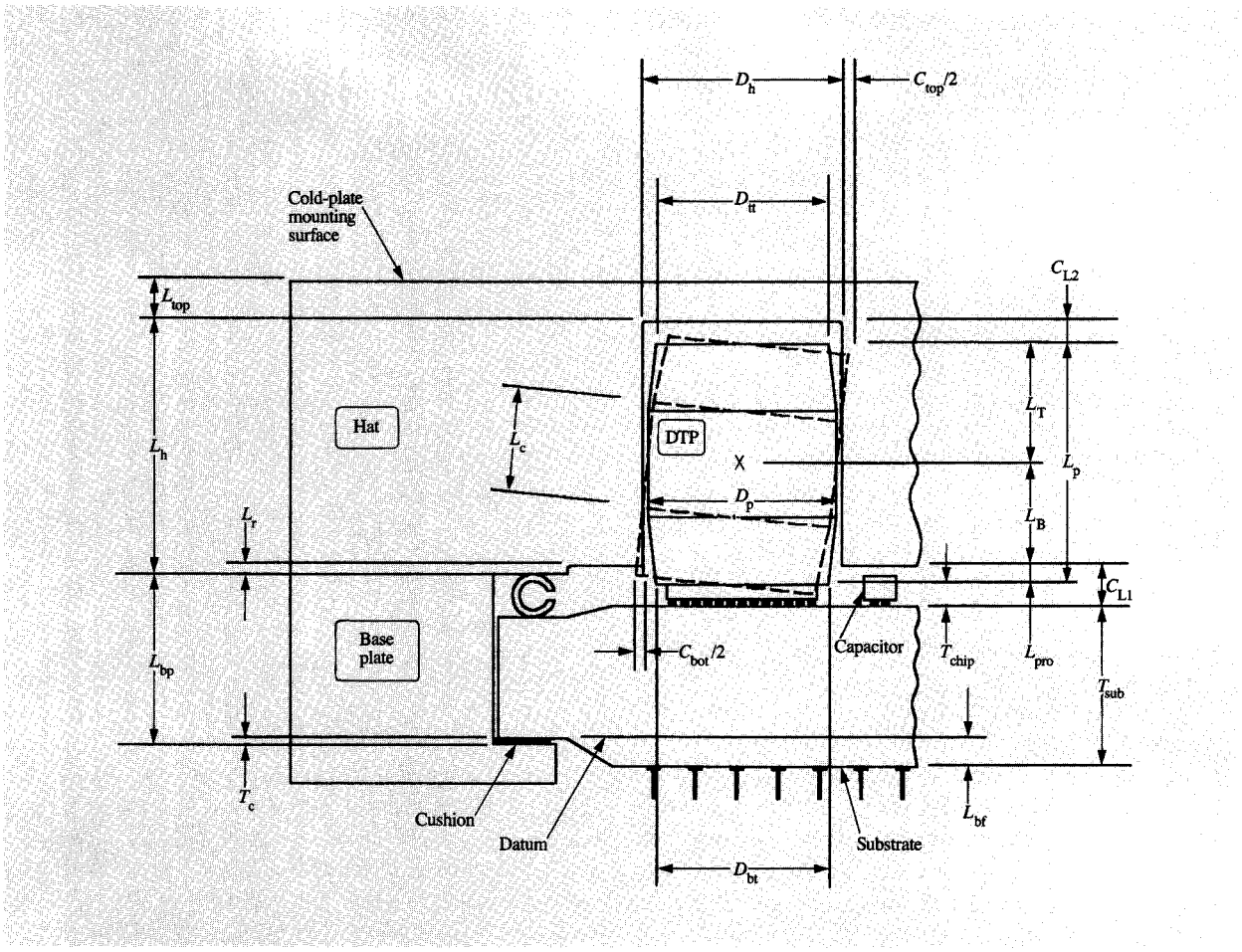

Figure 3

Key dimensional parameters of ES/9000 TCM cooling hardware.

be freely assembled and disassembled. While tighter clearances were tested successfully, DTP currently retains the minimum clearance used in 3081 and 3090 TCMs, which is $(D_h - \delta D_h) - (D_p + \delta D_p) = 0.02$ mm, where D_h is the hat hole diameter and D_p is the diameter of the middle section of the tapered piston. Two other critical clearance conditions required at the start of the analysis are $C_{L1,min}$ and $C_{L2,min}$. These represent the minimum clearances between the substrate and hat and between the piston and hat hole bottom, respectively. The first clearance condition is governed by the heights of capacitors shown in Figure 3 which are located on the top surface of the substrate; the second condition is related to the fully compressed dimension of the spring which resides behind the piston in the hole (not shown). The minimum clearances are obtained when the piston is at its maximum extension into the hat hole and must conform to a tilted chip. If we define $L_{e,max}$ as this condition, the nominal piston length L_p and nominal hole depth L_h are

$$L_p = L_{e,max} - \delta L_p + L_{pro} - \delta L_{pro}, \quad (1)$$

$$L_h = C_{L2} + L_p + T_{chip} + T_{sub} - L_{bf} + T_c - L_{bp}, \quad (2)$$

where L_{pro} is the length of the piston protruding out of the hat hole, T_{chip} is the thickness of the chip, T_{sub} is the overall substrate thickness, L_{bf} is the depth of the flange of the substrate ground from the pin surface to the reference datum, and L_{bp} is the lip length of the base plate.

Both L_p and L_h are dependent on tolerances and fit conditions which require further definition. For instance, if the lower flange surface of the substrate is the reference datum, as shown in Figure 3, the vertical assembly clearances are

$$C_{L1} = C_{L1,min} + \delta T_c + \delta L_{bp} + \delta L_r + \delta L_{bf} + \delta T_{sub}, \quad (3)$$

$$C_{L2} = C_{L2,min} + \delta T_c + \delta L_{bp} + \delta L_r + \delta L_{bf} + \delta T_{sub} + \delta T_{chip} + \delta L_p, \quad (4)$$

and the remaining undefined parameter is the tolerance of the hat recess δL_r. Once the substrate thickness is known, the base-plate thickness can be determined for proper C-ring compression. Then L_r is specified from

$$L_r = -L_{bf} + T_{sub} + C_{L1} + T_c - L_{bp}. \quad (5)$$

Similarly, the distance the piston protrudes out of the hat, L_{pro}, and its tolerance are determined from

$$L_{pro} = -T_c + L_{bp} + L_r + L_{bf} - (T_{sub} + T_{chip}), \quad (6)$$

$$\delta L_{pro} = \delta T_c + \delta L_{bp} + \delta L_r + \delta L_{bf} + \delta T_{sub} + \delta T_{chip}. \quad (7)$$

The thickness between the bottom of the hole and the cold-plate surface, L_{top}, is chosen to provide sufficient bulk material for the heat to travel around the hole to the cold plate with minimal constriction.

Other tolerances that affect the clearance between the piston and hat hole not shown in Figure 3 are the base-plate parallelism, $B_{\|}$, between the hat and substrate mating surfaces; the cushion parallelism, $C_{\|}$, between substrate and base-plate mating surfaces; the perpendicularity of the piston face to its axis; and the perpendicularity of the hat hole relative to the hat surface which mates with the base plate. Cumulatively, these tolerances contribute significantly to increasing the piston-to-hat-hole clearance. For ease of calculation, the first two tolerances are converted to an "equivalent" chip tilt and simply added to the angle of chip tilt, θ_{tilt}, as follows:

$$\theta_{Tot} = \theta_{tilt} + \arctan(C_{\|}/125) + \arctan(B_{\|}/125), \quad (8)$$

where the 125 approximates the substrate length to which the parallelism of the parts applies. Thus, the piston diameter at the uppermost portion of the engaged piston, D_{tt}, is determined from

$$(D_h - \delta D_h) - (D_{tt} + \delta D_{tt}) = C_{top} = 2L_T \sin(\theta_{Tot}), \quad (9)$$

where C_{top} is the minimum clearance between the top piston taper and the hole, and L_T is the length of the upper engaged half of the DTP. Similar derivations can be applied to determine the bottom piston diameter and the length of the center cylindrical section. Once the piston dimensions are determined, the nominal oil gaps are calculated as a function of piston engagement for each set of diametrical tolerances of δD_p, δD_{tt}, δD_{bt}, and δD_h. Once the gaps are known, the thermal performance of the assembly can be determined from the thermal models and compared to the difficulty in obtaining these tolerances.

- *Thermal modeling*

Models were developed to estimate the resistance of single and multiple chip sites from the device to the cold-plate mating surface. The most difficult portions to model accurately are the engaged sections of the DTP, because of annular gap variation along its axis. The overlapping portion of the piston and hat hole separated by a narrow gap filled with a relatively poorly conducting material can be represented by a pair of conduction-coupled fins [11]. Heat is transferred from the base of one fin (the piston face), is conducted across the narrow gap, and exits the assembly at the base of the second fin (the hat at the bottom of the hole). If T_p and T_h are defined as the piston and hat temperatures respectively, the following differential equations may be derived from an energy balance on the conduction-coupled fin assembly, with x representing the location along the piston axis "engaged" in the hat hole:

$$\frac{d^2T_p}{dx^2} = \frac{k_g \pi D_p}{k_p S(x) A_p}(T_p - T_h) \quad (10)$$

and

$$\frac{d^2T_h}{dx^2} = \frac{k_g \pi D_p}{k_h S(x) A_h}(T_h - T_p). \quad (11)$$

The subscripts h, p, and g refer to the hat, piston, and gap, respectively, k is thermal conductivity, A is cross-sectional area, D is diameter, and $S(x)$ is the axial variable annular gap between the piston and hole. A solution to these equations was presented by Chu et al. [1] for the case in which $S(x)$ was a constant (i.e., a cylindrical piston). For the DTP, $S(x)$ varies with axial engagement. Three approaches were used for the DTP geometry to solve these equations: a resistor network, as shown in **Figure 4**, an equivalent gap method, and numerical modeling.

The resistor and equivalent gap models were used to determine the sensitivity of heat transfer to various geometric parameters. Finite element models (FEM) were used to examine specific geometries in detail. Solving the thermal/mechanical relations consisted of iteratively selecting machining tolerances, determining TCM dimensions, and then evaluating the analytical and numerical models.

The models used the following thermal conductivity values, k:

Material	k (W/mK)
Copper (piston)	391
Aluminum (hat)	201
Oil	0.17
Silicon (chip)	129

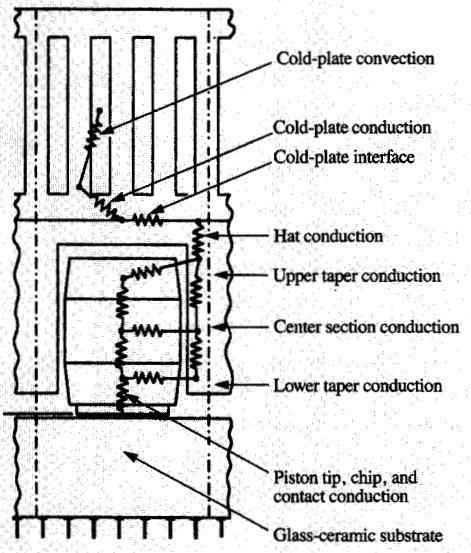

Figure 4

Partial ES/9000 TCM cross section with thermal resistance network.

Chips were typically 6.5 mm square on 10.8-mm centers, although rectangular chips and 9.9-mm spacings were also modeled.

Analytical models

One analytical model adjusted the solution of the cylindrical conduction-coupled fin differential equations to account for the $S(x)$ variation of the gap. An effective cylindrical piston, D_{eff}, was defined which is thermally equivalent to a tapered piston but is mechanically unacceptable because the clearance between parts is insufficient for tolerance accommodation. The low conductivity of the oil relative to the hat and piston metals permitted D_{eff} to be defined as

$$D_{eff} = \frac{D_p L_c + 0.5(D_{tt} + D_p)(L_T - L_c/2) + 0.5(D_{b,eng} + D_p)L_{b,eng}}{L_p - L_{pro}}, \quad (12)$$

where the subscript b,eng represents the dimensions of the engaged portion of the bottom taper of the DTP. Finite difference modeling was included in this analysis to account for the actual chip and piston surface contours and tolerances and to enable a parametric study to minimize the chip-to-piston-interface thermal resistance. The model also incorporated the effects of eccentricity of an off-centered piston located in a hole, as described in [11]. An improvement due to eccentricity occurs because the overall gap resistance is composed of many heat-transfer paths in parallel, radially connecting the piston to the hat hole as shown in Figure 4 by the resistor elements. With the piston tilted and offset, some paths increase in resistance, while others decrease by amounts that provide a lower overall gap resistance. The offset resistance reduction can be approximated from

$$\Delta R_{offset} = \frac{D_h - D_{eff}}{2k_g \pi D_{eff}(L_p - L_{pro})} - \left[\int_0^{\pi} \frac{k_g(L_p - L_{pro})D_{eff}}{r(\theta) - D_{eff}/2}\, d\theta\right]^{-1}, \quad (13)$$

where $r(\theta)$ defines the hat hole relative to the centerline of an offset piston. Finally, the mechanical tolerances of the dimensions illustrated in Figure 3 were put directly into the thermal model to obtain a parametric evaluation of the effect of changes in tolerance on TCM performance.

Numerical models

Building on the findings of the analytic models, extensive finite element modeling was used to understand DTP performance details for various cooling situations and to optimize its design. Single-site models, generally consisting of over 1000 three-dimensional isoparametric thermal solid elements with eight nodal points [12], were utilized. Symmetry was frequently used to reduce model size. **Figure 5** shows one application where half model symmetry is used to assess the impact of different tilt conditions. Shown are temperature contours corresponding to the flat and tilted conditions shown in Figure 2, with the chip powered at 15 W and the cooling hardware at nominal dimensions. These models included the unpowered edge of the chip and the exact interface gap geometry for added accuracy. The cold-plate contribution is included by introducing a convective coefficient boundary condition at the mounting surface on the hat equivalent to the cold-plate resistance. From the figure, the tilted piston case has a maximum chip temperature of 49.4°C, whereas the concentric piston on a flat chip is at 50.9°C. This 1.5°C reduction with the piston tilted against the hat hole represents a 0.1°C/W improvement in internal resistance. As determined from analytical studies, similar effects were found with an off-centered, vertical piston. The DTP is seldom perfectly centered or fully tilted; hence, test results lie between these predictions.

With the use of FEM, the data from the TCM test vehicles for uniformly powered square chips on 10.8-mm spacing were correlated with product modules that had square or rectangular-shaped chips, often nonuniformly powered and on 9.9-mm as well as 10.8-mm spacing.

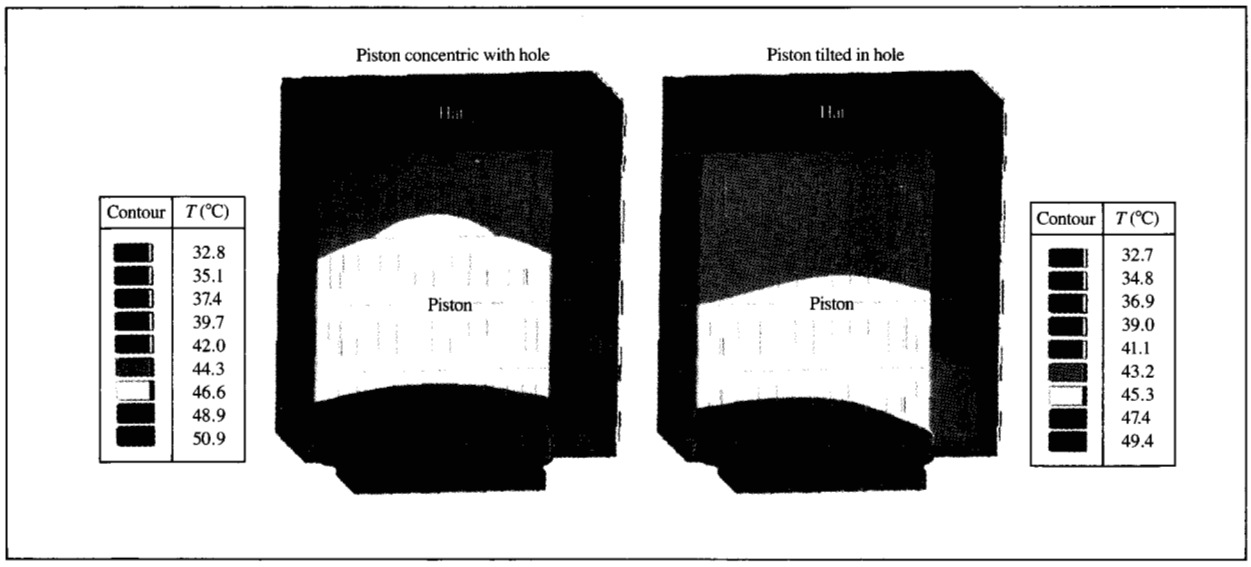

Figure 5

Temperature contours of DTP assembly shown in Figure 2.

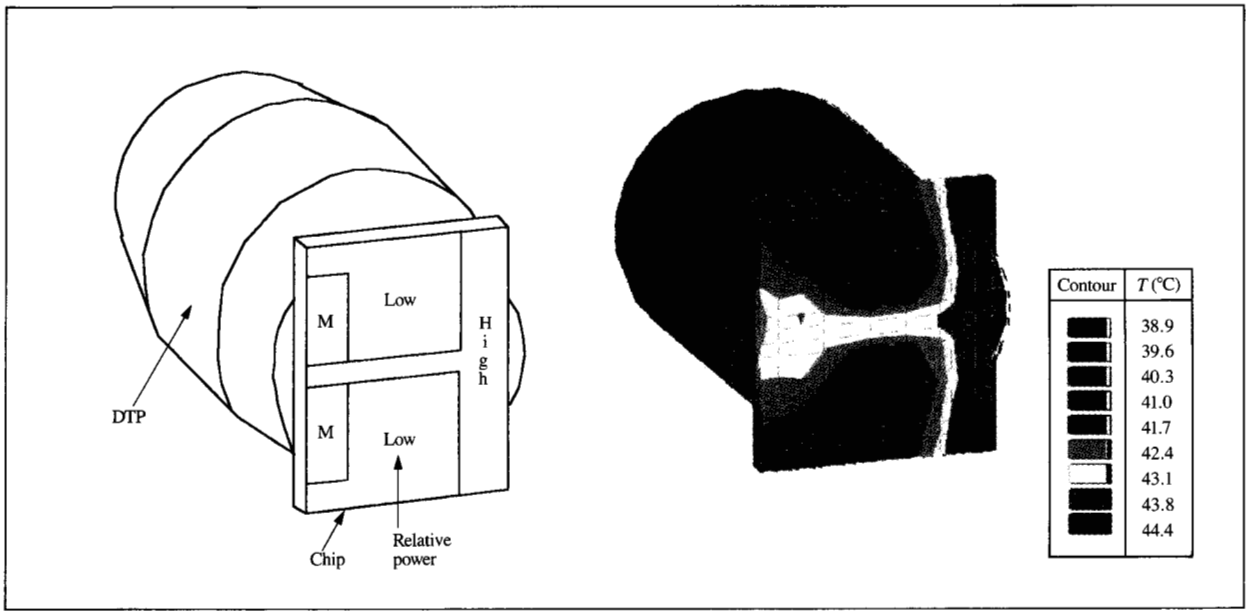

Figure 6

Nonuniformly powered rectangular chip (left) and the corresponding temperature contour (right).

Figure 6 shows the results of an analysis of a rectangular, nonuniformly powered chip on a 9.9-mm spacing from the device side of the chip. Low-, medium-, and high-powered regions are identified at the left, and the resulting temperature contours in the chip and piston are shown at the right.

The DTP module was modeled using two methods. Statistical resistance network models were generated by

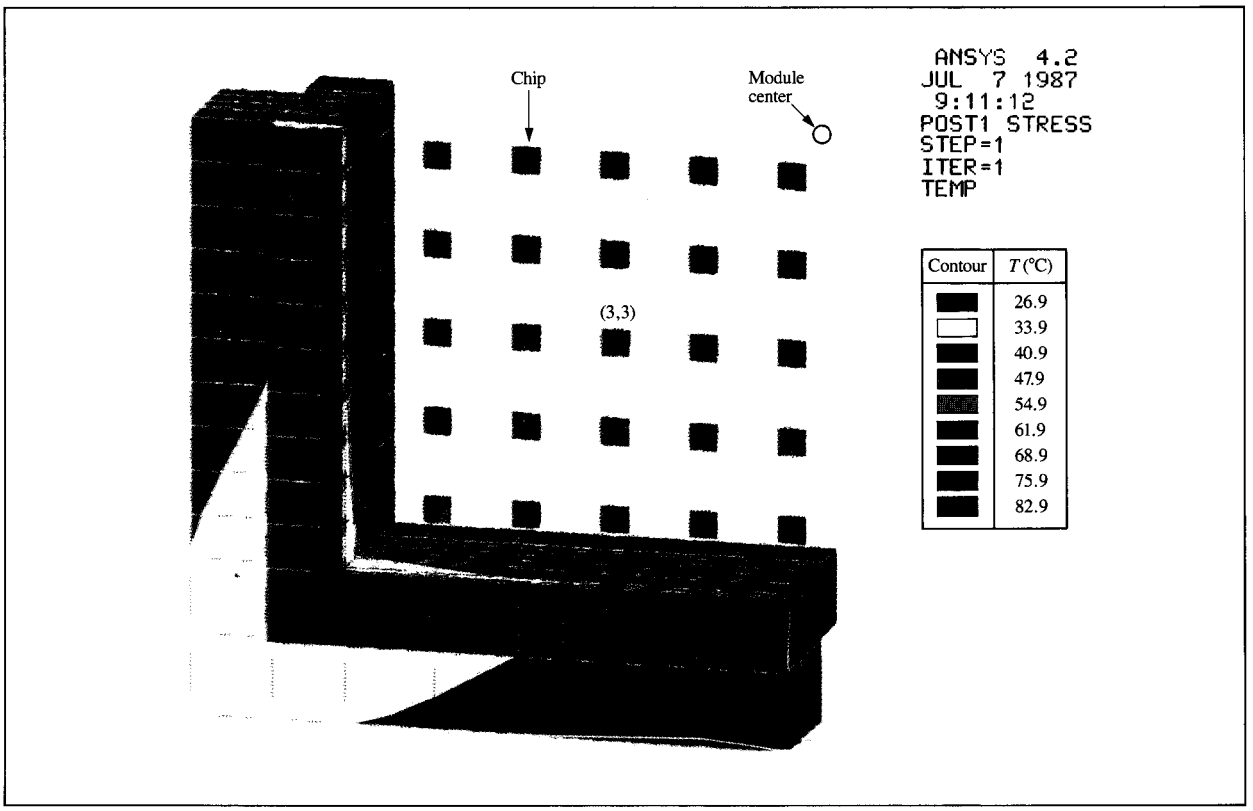

Figure 7

Early projections of temperature contours of one-quarter of a 100-site worst-case module. The pistons are undersized, the chip-to-piston-interface gaps are at worst case, and the external resistance boundary condition is 1.5 times the actual product value of the piston.

connecting the single-site models shown in Figure 4 with appropriate boundary resistors. The network was refined from test data and was used to assess the statistical minimum and maximum temperatures of each chip. These models combine the thermal and power variations in a Monte Carlo simulation and are comparable to the earlier modeling performed on 3081 and 3090 TCMs [2].

The need for accurate module models was apparent prior to generation of the network models. FEM was therefore applied to quantify the benefits of using a high-conductivity honeycomb-like hat structure which efficiently displaces heat around the holes in the hat from the hot to cool regions of the TCM, thereby reducing the temperatures of the high-powered chips. A direct, single-pass detailed FEM was practical for only one-eighth of the module because of the model size. To include more sites, substructuring was needed [12]. The 1300 solid elements of each chip site become a single "superelement." The edges and certain interior nodal points defining these elements, including representative chip nodes, become degrees of freedom in the superelement. A second model is created by using 3D isoparametric solid elements for the TCM perimeter, with each chip site represented as a "superelement" that can be individually powered. **Figure 7** shows temperatures for the 25 chip sites of a quarter-module model and for the perimeter of the substrate, hat, and cold plate. Here the (3,3) chip position is powered to 25 W and the other 24 sites are powered to 18.8 W each. To determine a full set of temperatures for every chip site, the method of substructuring requires an additional FEM solution iteration; however, with our interest focused on the subset describing the chip temperatures, considerable time was saved by selecting those nodes at each chip site. The temperatures at these nodes are displayed directly from the module analysis, as depicted in the figure. The temperature contours show that with a worst-case piston-to-chip-interface gap and an external resistance 50% higher than in the current product, the 25-W site has a temperature of 82.9°C. From these analyses, it was apparent that a 10% reduction of resistance could be realized at the high-powered sites because of the additional heat paths in the hat created from nonuniform chip powers. TCM test data confirmed these projections. Optimization studies from the TCM models yielded longer

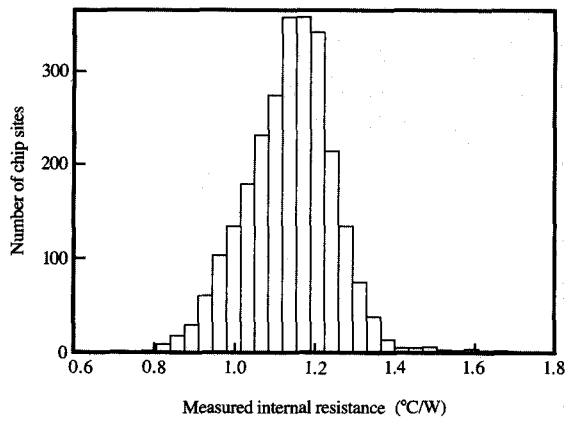

Figure 8

Measured chip-site internal resistance (°C/W) from 30 modules having 100 chips each.

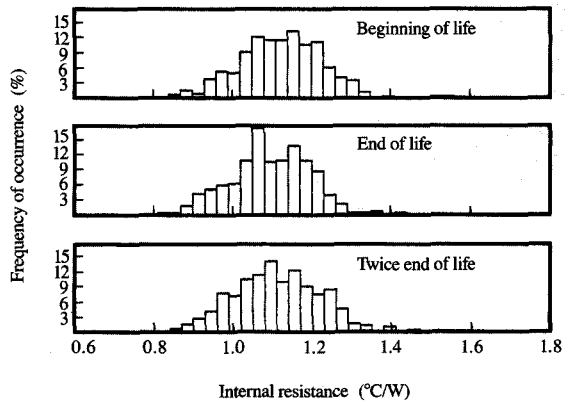

Figure 9

Measured chip-site internal resistance (°C/W) of modules stressed on accelerated life testers.

pistons in comparison with results from the single-site model.

• *Test verification*

The multiplicity of symmetric chip sites permitted the initial tests to be conducted on a single-site basis. The single-site test fixture, which enabled several design permutations to be tested rapidly, consisted of a "hat" of one chip site, a piston, a silicon chip with heating elements, and the peripheral equipment needed to power the chip, cool the hat, and record the data. In addition to verifying the analytical and numerical models, single-site tests were used to measure the decreased heat transfer due to the anodization coating on the hat hole surface at 0.06°C/W.

Full-scale test verification started in 1988 and has continued to the present time. Early development testing evaluated both aluminum and copper DTPs, as well as variations in oil vent size and shape, and supplemented the analyses. The oil vent adds additional manufacturing steps as well as a small thermal penalty. An axial vent located on the surface of the piston was chosen as the best compromise between performance and manufacturability. Chip temperature measurements were performed on thermal test vehicles with 100 chips having resistive heating elements and temperature-sensitive diodes. The TCMs were built in a clean-room environment and instrumented with thermocouples on the hat and substrate surfaces. The module test fixture powered the chips in a uniform manner and recorded 600 temperature and 200 power measurements. The data were reduced to internal resistance values at each chip site.

A database of more than 30 full-scale test vehicles was established to verify the DTP performance. **Figure 8** shows a distribution of measured chip-to-hat internal resistances for equally powered chips. Modeling identified four causes for the variation shown:

Source of variation	*Maximum variation* (°C/W)
Piston and hat tolerances	±0.2
Piston position in hat hole	−0.1
Module edge cooling effects	−0.15
Interface contaminants	+0.4

The test data verified the internal resistance to be 1.21°C/W for the 100-chip-site TCMs, as predicted by the analytic and numerical models. For the 121-chip-site TCMs, the internal resistance is increased to 1.36°C/W because of its smaller chip-site size. Single-site experiments in addition to the modeling efforts were used to determine the effect of the smaller site size.

In parallel with the establishment of a "beginning of life" database, several TCMs underwent a vigorous accelerated stress test to simulate environmental conditions during and after the calculated lifetime of the product. **Figure 9** shows that the TCM performance is slightly improved at the end of one life and is equivalent to new at the end of two expected lives. As the machine is

	L	K	J	H	G	F	E	D	C	B	A
1	8.9 W 40.0°C	8.9 W 39.4°C	8.9 W 40.3°C	8.9 W 40.5°C	8.9 W 39.5°C		8.9 W 39.4°C	8.9 W 39.1°C	8.9 W 39.4°C	8.9 W 38.5°C	8.9 W 36.5°C
2	8.9 W 39.6°C	8.9 W 41.5°C	8.9 W 41.5°C	8.9 W 41.4°C	8.9 W 41.8°C		8.9 W 41.4°C	8.9 W 40.9°C	8.9 W 41.3°C	8.9 W 39.9°C	8.9 W 38.5°C
3	8.9 W 40.1°C	8.9 W 40.9°C	16.9 W 47.4°C	8.9 W 44.8°C	8.9 W 43.8°C	10.0 W 43.7°C	8.9 W 43.7°C	8.9 W 43.2°C	16.9 W 47.9°C	8.9 W 41.4°C	8.9 W 38.0°C
4	8.9 W 39.2°C	8.9 W 40.9°C		20.0 W 56.7°C	20.0 W 59.3°C	20.0 W 56.6°C	20.0 W 57.2°C	20.0 W 54.3°C		8.9 W 40.2°C	8.9 W 38.5°C
5	8.9 W 39.5°C	8.9 W 41.0°C	16.9 W	20.0 W 54.2°C	20.0 W	14.3 W	20.0 W 61.7°C	20.0 W 57.4°C	16.9 W	8.9 W 41.7°C	8.9 W 37.6°C
6		15.0 W	10.9 W	19.5 W	19.5 W	13.6 W	19.5 W	19.5 W	13.8 W	15.0 W	
7	8.9 W 38.0°C	8.9 W 40.5°C	16.9 W	20.0 W 56.9°C	20.0 W 59.1°C	13.9 W	20.0 W 54.8°C	20.0 W 57.6°C	16.9 W	8.9 W 40.9°C	8.9 W 37.1°C
8	8.9 W 37.1°C	8.9 W 39.2°C		20.0 W 53.6°C	20.0 W 57.5°C	20.0 W 57.4°C	20.0 W 57.3°C			8.9 W 39.7°C	8.9 W 37.1°C
9		8.9 W 39.2°C	16.9 W	8.9 W 42.9°C	8.9 W 42.6°C	17.3 W	8.9 W 42.5°C	8.9 W 43.2°C	16.9 W 46.0°C	8.9 W 39.4°C	8.9 W 36.3°C
10	8.9 W 36.6°C	8.9 W 37.5°C	8.9 W 39.7°C	8.9 W 39.4°C	8.9 W 39.7°C	14.7 W	8.9 W 39.6°C	8.9 W 39.7°C	8.9 W 39.0°C	8.9 W	8.9 W 36.2°C
11	8.9 W 35.0°C	8.9 W 36.1°C	8.9 W 37.8°C	8.9 W 37.7°C	8.9 W 37.3°C		8.9 W 36.7°C	8.9 W 38.2°C	8.9 W 37.1°C	8.9 W 37.1°C	8.9 W 35.3°C

Figure 10

Projected nominal chip power (upper number) and measured chip temperature (lower) during operation of a 121-chip-site module in an ES/9000 uniprocessor.

powered on and off, thermal expansion moves the piston relative to the chip. The cycling tends to improve the chip interface gap by embedding asperities into the soft copper piston. This was particularly noticeable on those sites that had high initial resistance values. After end of life, a few sites had an increase in thermal resistance which was attributed to this relative motion scratching the piston face and increasing the interface gap. The maximum resistance at twice end of life was less than 1.4°C/W at these sites, and the TCM average was equal to its initial average.

The TCMs in the ES/9000 systems were the first modules to be thermally tested using actual product. Two testing procedures were used. Upon assembly, the TCMs were thermally tested in a post-encapsulation thermal tester (PETT) by monitoring temperature-sensitive diodes located on the chips. Most of these diodes were accessible through the module I/O pins. Several TCMs were also tested in ES/9000 systems by measuring the actual chip temperatures through wiring on the back side of the TCM board, whereas prior TCM developments relied on statistical resistor network projections. The data from actual product chips were used to refine network models and improve their accuracy. **Figure 10** shows the measured chip temperatures and the estimated nominal chip power *during operation* of a typical 121-chip-site module in an ES/9000 uniprocessor. The center sites in Figure 10 are powered to 20 W, whereas the perimeter sites are at about 9 W. A total resistance from electronic device to water can be calculated from the temperature data for each chip site. The interior sites have a total resistance 0.2°C/W less than the perimeter sites, illustrating the effect of heat spreading from the "hot" sites to the "cool" sites due to the inherently low thermal resistance paths in the hat structure between the chip sites.

Manufacturability

The DTP design retains the manufacturability inherent in prior TCM cooling technology while delivering four times the thermal performance. Despite its simplicity, its cooling capability is as good as those of multi-fin approaches. The

machining of a cylindrical hole in aluminum hats is well established. The piston tapers, with looser tolerances than the center cylindrical section, are readily machinable. The piston crown, which is the most tightly toleranced and critical dimension, has been successfully fabricated by several different processes. The other design attributes, including the hat anodization, conical spring, and oil vent, each have proven processes. The DTP parts are readily fabricated and assembled. Its low part number count minimizes assembly time and operator errors, and helps to ensure the reliability of the modules.

Extendibility

The thermal/mechanical models permitted a broad range of mechanical tolerances, material properties, and boundary conditions to be explored. The extent to which further cooling performance is implemented depends on the trade-offs to be accepted. Three possible thermal improvements include a further reduction in the annular gap, copper hats and pistons, and closer proximity of the coolant to the chips. In the first case, the gap may be reduced by taking advantage of statistical process controls (SPC) used in the manufacture of the hardware [13]. Whereas the ES/9000 TCM is designed to meet a worst-case addition of the many tolerances involved without disturbing the chip-to-piston interface, reduced annular gap width and improved performance are possible with the use of SPC. The second improvement is the performance gained from the high-conductivity materials. The difficulty here is achieving electrical isolation between the chips without the presence of aluminum. The third improvement involves making a combined hat and cold plate with the water channels located between pistons [14]. In this case the hat conduction resistances are eliminated. These extensions taken together can extend the dual-tapered-piston cooling technology on 6.5-mm chips to over 50 W.

Summary

The ES/9000 TCM demands for cooling high-powered chips to lower temperatures required a fourfold decrease in TCM thermal resistance from that of the 3090 system. Replacing helium in the interstitial piston-to-hat gap with highly conductive pastes was not successful, and the more complex designs using multiple pistons/fins per chip increased manufacturing costs. Further examination of the thermal/mechanical tolerance relationship among the chips, substrate, and cooling hardware led to development of the dual-tapered-piston design, which consists of an upper tapered section, a middle cylindrical section, and a lower tapered section. The DTP shape provides a means of displacing relatively low-conductivity oil with high-conductivity copper while retaining intimate contact between the piston face and chip surface under all tolerance conditions.

Analytical and numerical models were developed that provide estimates of the sensitivity of TCM performance to geometric parameters to optimize the TCM design. The single- and multi-site (100 and 121 chip sites) TCM models related experimental results of uniformly powered square chips to those of nonuniformly powered rectangular chips. The models demonstrated that heat transfer can be more efficient with the piston offset and/or tilted in its hat hole. The TCM models also quantified the spreading effects that cooler chips have on hotter chips. The models were consistent with each other and with experimental test data. Test verification was performed on single chip sites, on multi-chip thermal test modules, and on product modules in testers and in ES/9000 machines. Approximately 40 thermal test modules were tested, with several stressed well beyond the expected product lifetimes. Excellent agreement was obtained between the analyses and thermal data, which include data obtained from modules tested on the ES/9000 systems. The ranges of tolerance conditions, material properties, and boundary conditions studied indicate that DTP cooling is extendable to future product.

Acknowledgments

The authors wish to thank IBM East Fishkill Development Engineering and Product Assurance for their thermal and environmental qualification of the full-scale thermal test modules. Special thanks to R. Hegi for the outstanding job he did in developing vendor capabilities and capacity. N. Sandhu's implementation leadership, R. Shepheard's mechanical design contribution, K. Singh's finite element modeling, S. Song's crown development, J. Bowne's encapsulation work, D. Gravel's quality efforts, T. Quinn's and R. Kemink's system measurements, and R. Massey's planning were all vital and greatly appreciated.

References

1. R. C. Chu, U. P. Hwang, and R. E. Simons, "Conduction Cooling for an LSI Package: A One-Dimensional Approach," *IBM J. Res. Develop.* **26**, No. 1, 45–54 (1982).
2. S. Oktay and H. C. Kammerer, "A Conduction-Cooled Module for High-Performance LSI Devices," *IBM J. Res. Develop.* **26**, No. 1, 55–66 (1982).
3. R. C. Chu, R. E. Simons, and K. P. Moran, "System Cooling Design Considerations for Large Mainframe Computers," presented at the International Symposium on Cooling Technology for Electronic Equipment, Honolulu, HI, March 1987.
4. U. P. Hwang and K. P. Moran, "Cold Plates for IBM Thermal Conduction Electronic Modules," *Heat Transfer in Electronic and Microelectronic Equipment*, A. E. Bergles, Ed., Hemisphere Publishing, New York, 1990.
5. C. D. Ostergren and J. A. Paivanas, "Thermal Conduction Disc-Chip Cooling Enhancement Means," U.S. Patent 4,639,829, January 27, 1987.
6. R. C. Chu, J. C. Eid, and M. L. Zumbrunnen, "Circuit Module with Pins Conducting Heat from Floating Plate

Contacting Heat Producing Device," U.S. Patent 4,765,400, August 23, 1988.
7. R. Biskeborn, J. Horvath, and J. Harvlichuck, "High Conduction Cooling Module Having Internal Fins and Compliant Interfaces for VLSI Chip Technology," U.S. Patent 5,052,481, October 1, 1991.
8. G. F. Goth, K. P. Moran, and M. L. Zumbrunnen, "Thermal Conduction Module with Barrel Shaped Piston for Improved Heat Transfer," U.S. Patent 5,005,638, April 9, 1991.
9. J. U. Knickerbocker, G. B. Leung, W. R. Miller, S. P. Young, S. A. Sands, and R. F. Indyk, "IBM System/390 Air-Cooled Alumina Thermal Conduction Module," *IBM J. Res. Develop.* **35**, No. 3, 330–341 (1991).
10. M. L. Zumbrunnen, "Materials and Processing Approaches for High Performance Electronic Cooling," *Proceedings of the Fourth Electronic Materials and Processing Congress*, ASM International, Materials Park, OH, 1991, pp. 399–403.
11. J. C. Eid and M. L. Zumbrunnen, "A One-Dimensional Model and Optimization of Conduction Coupled Fins," *Proceedings of the International Electronic Packaging Society 7th Annual Packaging Conference*, Boston, MA, November 1987, pp. 282–292.
12. G. J. Desalvo and J. A. Swanson, *ANSYS Engineering Analysis System Users Manual*, Vol. 2, June 1, 1985, pp. 4.70.1–4.70.5; available from Swanson Analysis Systems Incorporated, P.O. Box 65, Houston, PA 15342.
13. L. Danziger, "The Statistical Aspects of Capability Indices for Six-Sigma Quality," *Technical Report TR-00.3610*, IBM Data Systems Division, Poughkeepsie, NY, March 26, 1991.
14. G. F. Goth, U. P. Hwang, R. F. Martin, K. P. Moran, and M. L. Zumbrunnen, "Oil-Filled Thermal Conduction Module with Tapered Pistons and Enhanced Side Channels," *IBM Tech. Disclosure Bull.* **33**, 181–184 (December 1990).

Received July 1, 1991; accepted for publication December 10, 1991

Gary F. Goth *IBM Enterprise Systems, P.O. Box 950, Poughkeepsie, New York 12602 (GFGOTH at POKADD6).* Mr. Goth earned a B.S.E. degree from Princeton in 1971, an M.S.M.E. degree from Union College in 1975, and an M.S. degree in statistics and operational research from Rensselaer Polytechnic Institute in 1978. He joined IBM in 1979, was manager of Packaging Application from 1983 to 1985, and was manager of the Power/Mechanical/Thermal Design Verification group in 1986. Mr. Goth joined the Mid-Hudson Valley Thermal Laboratory in 1987; his major responsibility is TCM cooling design. He has received an IBM Outstanding Innovation Award.

Michael L. Zumbrunnen *IBM Enterprise Systems, P.O. Box 950, Poughkeepsie, New York 12602 (ZUMBRUNN at PKEDVM9).* Mr. Zumbrunnen is an advisory engineer in the area of Future Systems Technology, working on the evaluation and implementation of packaging technologies for future large-scale computers. He joined IBM in 1984 and was initially involved in research and development of advanced electronic cooling technologies. He received his B.S. and M.S. degrees in mechanical engineering from the University of Minnesota. Mr. Zumbrunnen has received five IBM Invention Achievement Awards and an IBM Outstanding Innovation Award. He holds four U.S. patents and is a member of the Institute of Electrical and Electronics Engineers.

Kevin P. Moran *IBM Enterprise Systems, P.O. Box 950, Poughkeepsie, New York 12602 (MORAN at POKADD6).* Mr. Moran received his B.E. degree in mechanical engineering from the City College of New York in 1965 and his M.S. in mechanical engineering from Syracuse University in 1971. He joined IBM at the Poughkeepsie Laboratory in 1965 and currently manages the ES Power/Thermal Development group. Mr. Moran has received five IBM Invention Achievement Awards and an IBM Outstanding Innovation Award. He holds five U.S. patents and has two patents pending.