S. Oktay
H. C. Kammerer

A Conduction-Cooled Module for High-Performance LSI Devices

The advent of LSI chip technology makes possible significantly increased performance and circuit densities by means of large-scale packaging of multiple devices on a single multi-layer ceramic (MLC) substrate. Integration at the chip and module levels has resulted in circuit densities as high as 2.5×10^7 circuits per cubic meter, with the necessity of removing heat fluxes on the order of 100 kW/m^2. This paper describes the development and implementation of a novel packaging concept which meets the stringent and highly interactive demands on cooling, reliability, and reworkability of LSI technology. These requirements resulted in an innovative packaging approach, referred to as the thermal conduction module (TCM). The TCM uses individually spring-loaded "pistons" that contact each chip with helium gas, the conducting medium for removing heat efficiently. A dismountable hermetic seal makes multiple access possible for device and substrate rework, while ensuring mechanical and environmental protection of critical components. A wide range of thermal, mechanical, and environmental experiments are described with analytical and computer models. The one-dimensional approach used in the previous paper by Chu et al. is extended to three-dimensional computer modeling. Simulations of expected chip temperature distributions in the IBM 3081 Processor Unit are discussed. Enhanced thermal performance of the advanced packaging concept for future applications is also indicated.

Introduction

The rapid advances in device technology toward large-scale integration (LSI) have often been restricted by the ability to package and cool the devices, which in turn limits the full utilization of the technology. The performance and reliability achieved through circuit speed, circuit density, and reduced interconnections are offset by the necessity to reduce circuit densities in order to permit cooling and interconnection of the highly integrated circuit devices. Cooling of high-performance multi-chip packages during the early 1970s was generally by means of an external chilled-water system in combination with immersion of the chip in a low-boiling-point fluid. The closed-loop concept of *pool boiling* by means of a low-dielectric, stable, inert fluid required repeatable *nucleate boiling* from the chip surface, along with a cooled exterior for fluid recycling. The concept of immersion, as described by Chu *et al.* [1(a)] in a companion paper in this issue, resulted in a number of cooling concerns which dictated the need for an alternative cooling concept. Therefore, a development effort was initiated to establish a packaging concept using conduction-cooling techniques to meet the projected requirements. Although several techniques were considered, the most promising was based on a patent filed by R. C. Chu *et al.* [1(b)]. This paper traces the development and implementation of this novel packaging scheme, which meets the stringent and highly interactive demands of LSI technology without adversely affecting critical reliability and reworkability requirements. These unique cooling and encapsulation concepts rely on the use of individually spring-loaded "pistons" in contact with each chip; these are thermally enhanced with helium gas in the interface gaps and are sealed into the module with a dismountable metal C-ring. The resulting package achieves a high level of circuit integration and performance while possessing the many other product attributes required of LSI technology.

In the following sections, the basic TCM package is described in terms of the product requirements and resulting engineering design description. The thermal characteristics of the TCM are defined and analyzed with the aid of

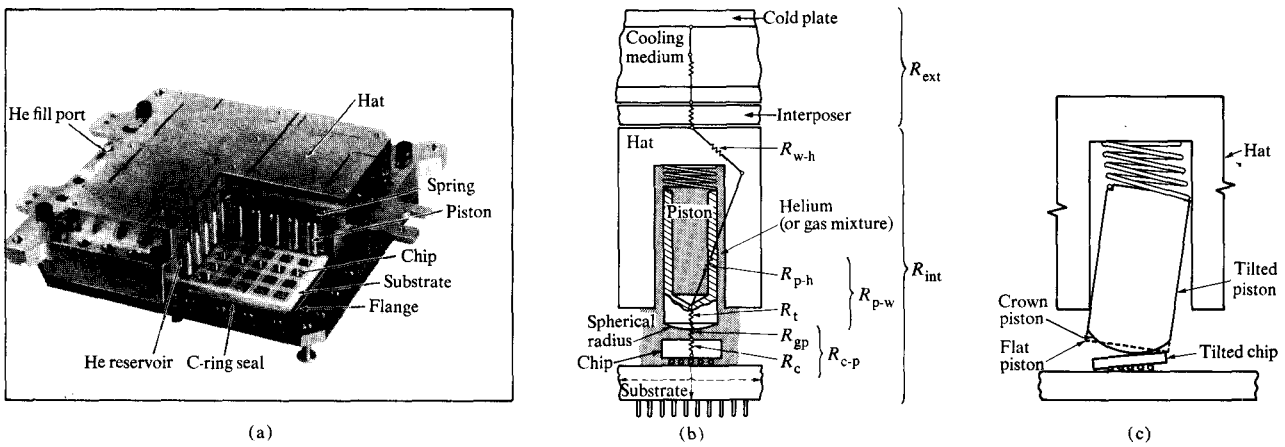


Figure 1 (a) Details of the thermal conduction module (TCM), (b) the thermal resistances used in the text, and (c) the tilting of a flat *vs.* a crown piston.

computer models at the chip and module levels. These models are verified by novel single-chip and multi-chip experiments. The importance of statistical computer simulations under field conditions, especially for multi-chip modules, is illustrated in the discussions involving a set of TCMs for the IBM 3081 system. The encapsulation development is discussed, as well as the extendability of the advanced packaging concept for future applications.

Product requirements

The application requirements imposed on the TCM by use of large-system processors reflect a wide range of operating and environmental machine conditions. These include temperature, humidity, power on/off, machine lifetime, failure rate, and various mechanical requirements. Cooling water is supplied between about 22 and 31°C to provide a heat sink for power dissipation.

The module package, in addition to providing the cooling path, encloses the active circuit devices as mounted to the substrate carrier, and thus must ensure adequate mechanical and environmental protection of the critical elements. To ensure proper circuit performance and maintain high reliability, the logic and memory LSI devices require junction temperatures to be controlled between 40 and 85°C. The specific chip and module power dissipations are nominal maxima of 2.6 and 218 W and design maxima of 4 and 300 W, respectively. The thermal design requirement, as expressed in terms of the maximum thermal resistance between the device layer in the chip and the system cooling water, is 13°C/W.

Because of the complexity of the designs and processes associated with the functional module, key requirements are the capabilities of repairing and updating chips and substrates. Therefore, the module design must permit multiple openings and closures of the hermetic seal for replacement of chips and addition or deletion of overflow wires on the surface of the substrate using the standard manufacturing process. Several general needs involving the module interaction in higher-level assemblies also exist, including the capabilities to affix a detachable cold plate to the upper surface of the module with minimal thermal resistance within the interface, to insert and actuate the module in the system which connects to the I/O pins, and to meet reasonable handling, shipping, environmental, and safety requirements and conditions.

Product description

The TCM (shown in Fig. 1) weighs 1.65 kg and has approximate dimensions of 150 × 150 × 50 mm. The module assembly is a basic field-replaceable unit (FRU) mounted to a board/connector (second-level package) assembly structure. The assembly is composed of the following parts.

Cold plate The heat sink, which provides for total heat dissipation from the circuit chips, is cast or machined from high-thermal-conductivity BeCu and is rigidly attached to the hat by screws. Water flows through the cold plate at approximately 24°C and 4×10^{-5} m^3/s, resulting in a total thermal resistance from the water to the hat surface of 0.02°C/W.

Interposer A ®Lexan [2] thermal insulator [3] is inserted between the cold plate and the hat to provide upward adjustment of chip temperatures. Various thicknesses are utilized depending on the power dissipation of the module.

Hat assembly The key element in the thermal path from the LSI device to the cold plate is the hat assembly, which includes hat pistons, springs, and assembly hardware. The basic hat is fabricated from a high-thermal-conductivity,

machinable, lightweight aluminum alloy and provides primary cooling and mechanical protection for devices mounted on the substrate. The hat is designed to have a He volume of 9×10^{-5} m^3 and contains a fill port/plug to allow pressurization of the He. The pistons are also machined from the aluminum alloy and anodized to provide electrical isolation. The face of the piston that directly contacts the chip is machined to a 150-mm spherical radius (crown radius of curvature $\rho = 150$ mm) [4] to minimize off-center contact due to chip tilt. This provides for more uniform chip temperatures and reduced mechanical stress on the chip.

C-ring A lead-plated ®Inconel C-ring coated with wax is compressed between the hat and substrate frame to form the reworkable hermetic seal.

Substrate assembly This assembly contains all electrical functions and is composed of the multi-layer ceramic (MLC) substrate, semiconductor devices (chips), and discrete wires. The 90 × 90-mm ceramic is brazed to a gold-plated ®Kovar frame, contains up to 33 layers, and has 1800 I/O pins brazed to the bottom surface. The top surface provides mounting sites for the joined chips (with a maximum of either 100 or 118 chips) and each chip site has surface pads for repair and error-correction capabilities. A temperature-sensing thermistor is also mounted on the top surface to monitor and control internal temperature excursions.

Base plate A stainless steel retaining plate is used to provide sufficient clamping load and rigidity for proper C-ring compression and retention. The plate has tapped holes around the periphery to accept screws inserted through clearance holes in the hat; this maintains clamping load on the ring. The base plate also contains precision guide pins and guidance holes for accurate location, insertion, and activation of the module to the next level of assembly.

Cooling development

- *Thermal characteristics of the TCM*

The TCM combines the advantages of conduction cooling used at the chip and module levels with that of the forced convection liquid cooling at the system level. It differs from liquid-encapsulated modules in that the electronic components are not brought into direct contact with the liquid. The components inside the TCM (Fig. 1) are immersed in He gas, which serves as the heat-conducting medium in gaps formed between the contacting surfaces. Helium is used because it has a thermal conductivity approximately six times that of air.

The TCM, the aluminum piston and spring, and the cavity in which the piston rests are shown in Fig. 1(b). For convenience in analyzing the thermal performance of the TCM, the thermal resistance of the heat path from the chip to the cold plate is divided into various segments, as shown; other heat paths are indicated by broken arrows. The associated thermal resistances [5] are now discussed.

- *Internal thermal resistance* R_{int}

The internal thermal resistance of the TCM is defined as $R_{int} = \Delta T_{c\text{-}h}/P_c$, where $\Delta T_{c\text{-}h}$ is the temperature difference between the chip and the top of the hat and P_c is the chip power. R_{int} is divided into three segments: $R_{c\text{-}p}$, the total resistance from the chip surface to the piston, which is made up of the actual chip resistance R_c and the resistance of the interfacial gap R_{gp}; $R_{p\text{-}w}$, the resistance from the piston to the cavity wall, and $R_{w\text{-}h}$, the resistance from the cavity wall to the hat. Note that $R_{w\text{-}h}$ is equivalent to R_h as used in the Chu paper; likewise, $R_{p\text{-}w}$ is similar to Chu's $R_t + R_{p\text{-}h}$.

$R_{c\text{-}p}$ is governed by the metal-to-metal contact between the silicon chip and the aluminum piston and by the gap formed between them due to tilting of one with respect to the other. The literature abounds with studies related to the prediction of contact area and resistance between contacting bodies as a function of pressure [6]. However, experiments conducted with deadweight loads of more than 1 kg on the Al piston against the Si chip (5×10^4 kg/m^2 on a chip-area basis) showed no appreciable change in the contact thermal resistance from a no-load condition.

However, the contribution to thermal resistance of the tilt between the chip and the piston was found to be significant. This tilt results from the tilting of the piston itself in its cavity as it is being pushed by the spring behind it, as well as from tilting of the chip when it is soldered to the substrate. The maximum tilt on any given chip on a module is limited to 0.076 mm with respect to the substrate surface. Thus, it is evident that a piston with a flat end larger than the chip dimensions is expected to catch the edge of the chip and to tilt as shown in Fig. 1(c).

To reduce the gap between the piston and chip, and hence the thermal resistance, the ends of the pistons in the TCM are ground into spherical "crown" shapes. The crown forces the piston to roll away from the edges and towards the center of the chip, thereby resulting in an effectively smaller gap between the chip and piston. Furthermore, by making the piston contact the chip more centrally, the temperature gradients inside the chip are minimized. In fact, the variations of temperatures inside the chip as the piston is made to roll on the back of the chip are easily discernable (as will be shown in a later section).

Figure 2 shows the variation of $R_{c\text{-}p}$ as a function of tilt for both the flat and crown pistons. For tilts <0.051 mm, $R_{c\text{-}p}$ for the flat piston is less than that from the crown piston because the net effect of the crown is to reduce the effective gap

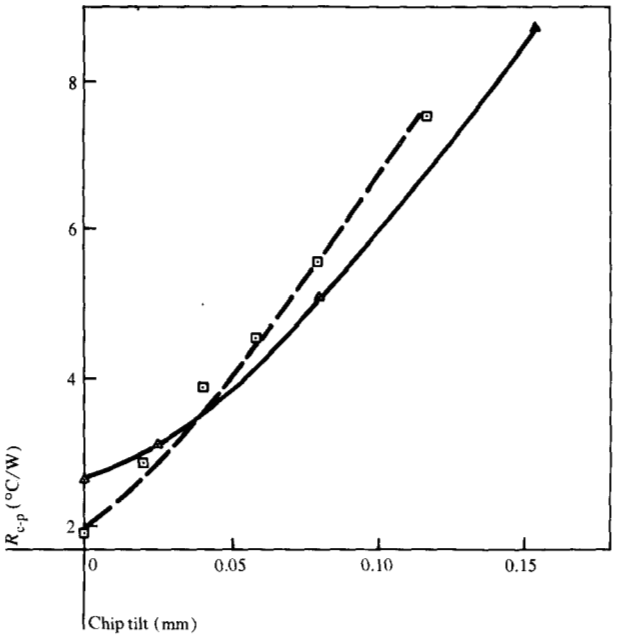


Figure 2 Measured "apparent" $R_{c\text{-}p}$ values of flat (- - -) *vs.* curved (ρ = 150 mm) crown pistons (———). The maximum interface distances are 0.0 mm for flat pistons and 0.017 mm for crown pistons with ρ = 150 mm. The chip power is the same with or without tilt; therefore, heat losses through the C-4 pads differ with different amounts of tilt; hence the "apparent" values.

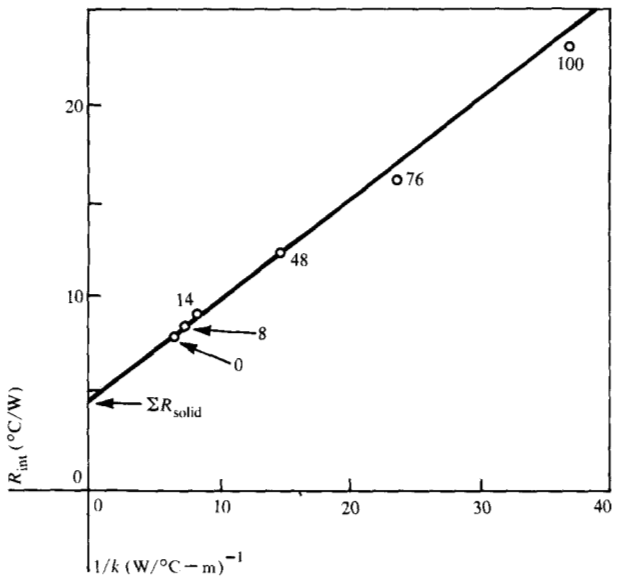


Figure 3 The internal resistance R_{int} as a function of the reciprocal of the thermal conductivity $1/k$ for various He/air mixtures using crown pistons. The numbered data points refer to the percentage of air in the He/air mixture. The intercept is the total thermal resistance of the solid parts R_s (see text); it includes the chip vertical resistance from the device layer to the back surface of the chip.

between the chip and piston at relatively large tilts. The piston tilt also affects $R_{p\text{-}w}$. Because of the large surface area available in the "annulus," the sensitivity of $R_{p\text{-}w}$ to tilt or "eccentricity" is not very significant (see later section on single-chip analysis).

As stated earlier, one of the important features of the TCM is the He gas sealed inside the module. To see this, consider R_{int} of the TCM to be comprised of two parts: R_s, the thermal resistance offered by the solid parts (*i.e.*, the chip, the tip and walls of the hollow piston, the cavity wall, and the top portion of the hat) and R_g, the thermal resistance offered by the gas inside the gaps between the chip and piston and between the piston and cavity wall. (The pistons are made hollow to minimize the impact forces on the chips during shock and vibration.) For the dimensions shown in Fig. 1(c), $R_s \approx 4.4$°C/W. With a nominal tilt in air, where $R_g \approx 17.5$°C/W, R_{int} of the TCM in air is >22°C/W, which is unacceptable. By replacing the air with He, where R_g is now ≈3.5°C/W, R_{int} = 7.9°C/W. To this, we add a chip spreading resistance of ≈0.75°C/W for a total R_{int} of ≈8.7°C/W; this should be compared with the measured value of 8.4°C/W.

Because of the strong dependence of R_g and R_{int} on the thermal conductivity of the encapsulated gas, tests were performed with different air/He mixtures in a TCM. The results yielded measured R_{int} values as shown in Fig. 3. With an expected maximum ingress of 13% air into the module, the thermal resistance degrades by ≈0.7°C/W from an initial value of 8.4°C/W, obtained at the time of initial fill with pure He, to an end-of-life (EOL) value of 9.1°C/W. For convenience, the reciprocal of the thermal conductivity of the various air/He mixtures is plotted along the abscissa.

It is apparent that accurate thermal analysis of a complicated structure, such as that shown in Fig. 1(c) for *one* piston, and in particular of structures containing many pistons, each with their own heat sources, vias, and chips, becomes at best very difficult and occasionally erroneous if simple one-dimensional models are used. In addition, closed-form solutions for predicting temperatures at multiple heat sources (chips) placed on a finite body (substrate) and with nonuniform boundary conditions (cold plate on one side and computer board on the other) are not available. Since the heat flow patterns (flux lines) in the TCM are indeed very complex, a one-dimensional analysis can easily lead to errors in excess of 2°C/W in the overall thermal resistance from the chip to the heat sink. Thus, for a 4-W chip, a 2°C/W error would produce an 8°C error in the chip temperature. The acceptable value for meeting the IBM System 3081 product objectives was ±2°C. To meet this objective, a detailed three-dimensional computer model was required. Furthermore, the computer program had to perform statistical analysis for various dimensional tolerances and parametric distributions. Before presenting the detailed computer

Table 1 Nominal thermal parameters for the TCM. In most cases, a normal distribution is assumed.

Parameters	*Nominal values*
Chip powers (W)	0.4–2.7†
Thermal resistances (°C/W)	
R_{int}-TCM100	all X sites, 8.4
R_{int}-TCM118	X sites, 8.4
	Y sites, 9.3
R_{int}-EOL	0.7
R_{ext}-no interposer	0.02
R_{ext}-0.13-mm interposer	0.07
R_{ext}-0.25-mm interposer	0.11
R_{ext}-1.0-mm interposer	0.39
R_{ext}-EOL	0.005
$R^*_{c\text{-}sub}$	6.5
$R^*_{sub\text{-}b}$	1.0
$R^*_{h\text{-}s}$	1.0
R^*_{subs}	15.0
R^*_{hs}	1.0
Average temperatures (°C)	
T^*_{iw}	24–27†
T^*_{b}	41–47†
T^*_{s}	33–38†

*Subscripts c-sub, sub-b, h-s, subs, and hs stand for chip-to-substrate, substrate-to-board, hat-to-stiffener, substrate spreading, and hat spreading. Subscripts iw, b, and s stand for inlet water, board, and stiffener.
†See Fig. 4.

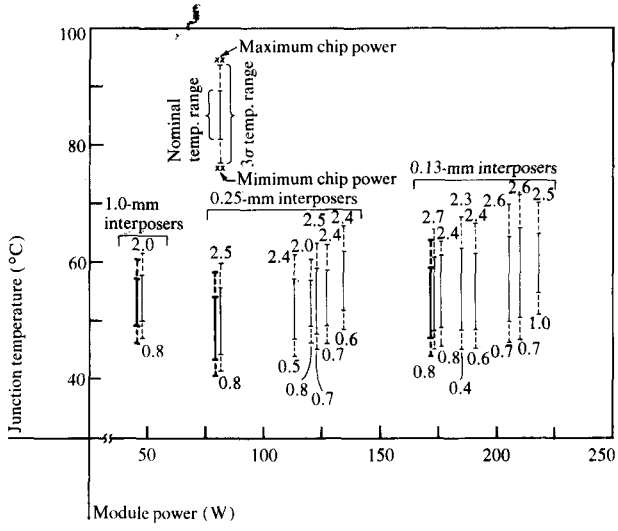


Figure 4 Computer-simulated thermal performance specifications for the IBM 3081 system. The solid bars show the temperature range due to the range of nominal chip powers (W) and other thermal parameters in each module. The dotted extensions bound the 3σ statistical distribution of junction temperatures. Hollow pistons are used; the heavy lines refer to time-zero temperature ranges, while all other lines refer to EOL values. Note that the minimum and maximum junction temperatures T_j do not necessarily correspond to chips with minimum and maximum power, respectively.

and thermal analyses of the TCM and the thermal simulations under various conditions, we discuss the external thermal resistance of the TCM.

• *External thermal resistance* R_{ext}

The external thermal resistance is defined relative to the module; it is obtained by dividing $\Delta T_{h\text{-}iw}$, the temperature difference between the hat and the inlet water, by the module power P_m. R_{ext} is divided into two parts: the interfacial resistance between the hat and mating cold-plate surfaces and the resistance between the cold-plate surface and the circulating water inside the cold plate. The latter is a combination of three resistances through the cold-plate wall, through the convective boundary layer, and between the water inlet and outlet of the cold plate. These, together with a metal-to-metal interfacial resistance of 0.007°C/W and an EOL fouling factor of 0.005°C/W, yield a total R_{ext} of 0.025°C/W.

The chip temperatures are established by the external and internal thermal resistances to heat flow from the chip to the hat, to the cold plate, and to the board on which the modules are held by special connectors. For simplicity, we assume that the rear of the computer board is insulated and that the board temperature is approximately the same as the average temperature of all the chips in all the modules. (In the computer model, we can assign different water and board temperatures for each module position in the machine.) With the above assumptions, one can calculate the expected gross chip temperatures by using the EOL R_{int} and R_{ext} values of 9.1 and 0.025°C/W, respectively. Thus, noting that R_{int} operates on the chip power, and R_{ext} on the module power, the temperature rise ΔT of the chip over the water-sink is given by $R_{int}P_c + R_{ext}P_m$. Thus, the chip temperature T_c is simply $\Delta T + T_{iw}$, where T_{iw} is the inlet water temperature.

It is of interest to calculate the minimum and maximum chip temperatures for minimum and maximum chip-power and sink conditions. For TCM technology, the minimum and maximum chip/module powers are 0.4/40 and 4/300 W, respectively; the minimum and maximum water inlet temperatures are about 22 and 31°C. Using the simplified one-dimensional expression just described, the expected minimum and maximum chip temperatures (T_c^{min}, T_c^{max}) are ≈26 and 77°C. It can be shown that the temperature drop from the power-dissipating junctions to the surrounding bulk silicon is ≈3°C. Thus, the minimum and maximum *junction* temperatures (T_j^{min}, T_j^{max}) are ≈29 and 80°C. It should be noted that in actual modules the chip powers are not uniform. Furthermore, the actual chip temperatures are dependent on the level of thermal interaction among chips, chip sites, stiffeners on the board, etc.

It is clear from these calculations that for the set of 14 modules in the 3081 system (see Fig. 4), each with a different number of chips, chip powers, and module powers,

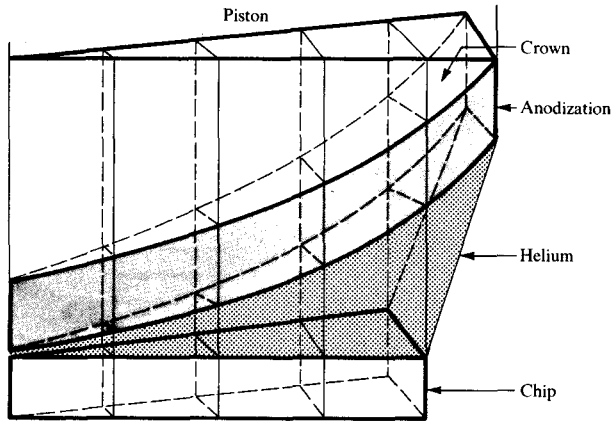
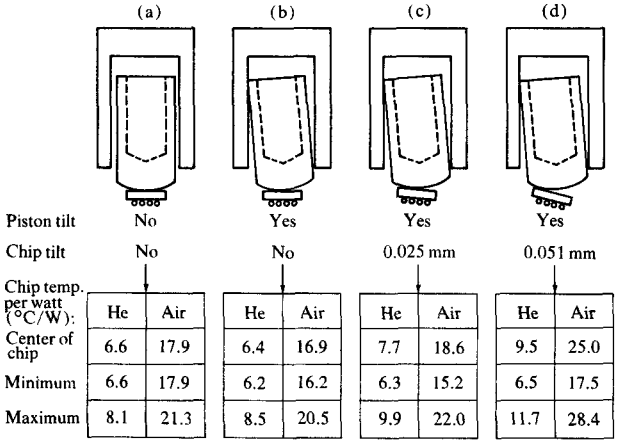


Figure 5 Detail of the chip-to-piston interface connections.



	(a)		(b)		(c)		(d)	
Piston tilt	No		Yes		Yes		Yes	
Chip tilt	No		No		0.025 mm		0.051 mm	
Chip temp. per watt (°C/W):	He	Air	He	Air	He	Air	He	Air
Center of chip	6.6	17.9	6.4	16.9	7.7	18.6	9.5	25.0
Minimum	6.6	17.9	6.2	16.2	6.3	15.2	6.5	17.5
Maximum	8.1	21.3	8.5	20.5	9.9	22.0	11.7	28.4

Figure 6 NASTRAN simulations of the internal thermal resistances for the 3081 system as a function of tilt using He/air mixtures for a 150-mm crown piston.

the chip temperatures will vary considerably, notwithstanding the fact that some of them may be lower than the minimum allowable temperature of 40°C. An important criterion for circuit performance is that chip temperatures lie between 40 and 85°C. Consequently, interposers were added as thermal insulators between the hat and cold plate on judiciously chosen groups of modules so that the average temperature in all modules was approximately the same (55°C), while the individual chip temperatures were within the minimum and maximum specifications.

For brevity, the particulars for the choice of material properties of the interposers are not given here. The needed thermal properties are determined by extensive computer simulations described in the next section. The interposers have the same areas as that of the module hat. The total R_{ext} values for each thickness of interposer are shown in Table 1.

Computerized thermal analysis of TCMs

Although a simplified one-dimensional approach may be used as a guide in the initial stages of module design, it is necessary to develop correct models to accurately predict chip temperatures for various conditions in the field. In the computerized thermal analysis of a TCM, two approaches are used. For parametric study of the heat-transfer characteristics of a particular aspect of the design (such as the change in R_{int} with the radius of curvature ρ of the piston), finite-element techniques are applied with the aid of the NASTRAN program [7]. Once the basic thermal resistances have been determined, an electrical network-analysis program called ASTAP [8] is used in simulating the total module, including all the chips and pistons. Unlike NASTRAN, this program has the ability to perform statistical analysis of chip temperature distributions as a function of various tolerance distributions.

In what follows, we first analyze the thermal characteristics of a single chip and piston in its cavity with the aid of NASTRAN, and then simulate chip temperature distributions in field TCMs with ASTAP. The experimental verification of the computer results is discussed later.

- *Single-chip analysis*

In Fig. 5, detail of the NASTRAN model is shown for the case where the He gap between the flat surface of the chip and the spherical surface of the piston is readily bridged with the application of versatile finite elements. Since both the piston and cavity have circular cross sections, the chip is also circular, which does not introduce any appreciable error into the calculations. The outer surfaces of the cavity (not shown) are prescribed to be adiabatic.

In constructing the three-dimensional model, ≈1000 nodes were used. A special preprocessor program that was developed to enter data into NASTRAN is not described here. The 17.5-level NASTRAN program was run on an IBM System/370; each simulation run took ≈2 min of CPU time. The simulations were performed by arbitrarily setting the temperature at the top of the hat to 0°C and defining this to be the sink temperature, and by assigning appropriate power levels to the elements representing the chip. With NASTRAN, either transient or steady-state temperatures can be computed readily.

It is interesting to simulate temperatures at a given chip power as the piston is allowed to roll on the chip. The temperatures given in Fig. 6 correspond to the various positions of the piston relative to the chip; the chip may also be tilted relative to the piston. The temperatures in the chip change as the rolling contact point moves. As expected, the lowest chip temperature always occurs at the point of contact. As the relative tilt is increased, the overall chip temperatures, the ΔT, and hence the R_{int}, increase.

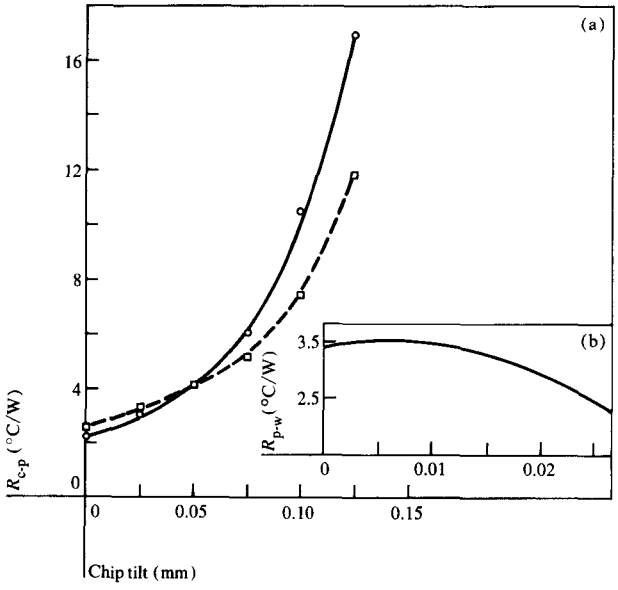


Figure 7 (a) NASTRAN simulation of the change in "intrinsic" $R_{c\text{-}p}$ values as a function of the chip tilt as ρ is increased from 150 mm (- - -) to 200 mm (———). The simulation equation used for calculating $R_{c\text{-}p}$ is $\{[(T_c^{max} + T_c^{min})/2] - T_t\}/P_c$, where the temperatures are obtained from simulations and where T_t is the temperature at the tip of the piston. (b) NASTRAN simulation of $R_{p\text{-}w}$ as a function of the tilt of the piston in the cavity for a 150-mm crown piston.

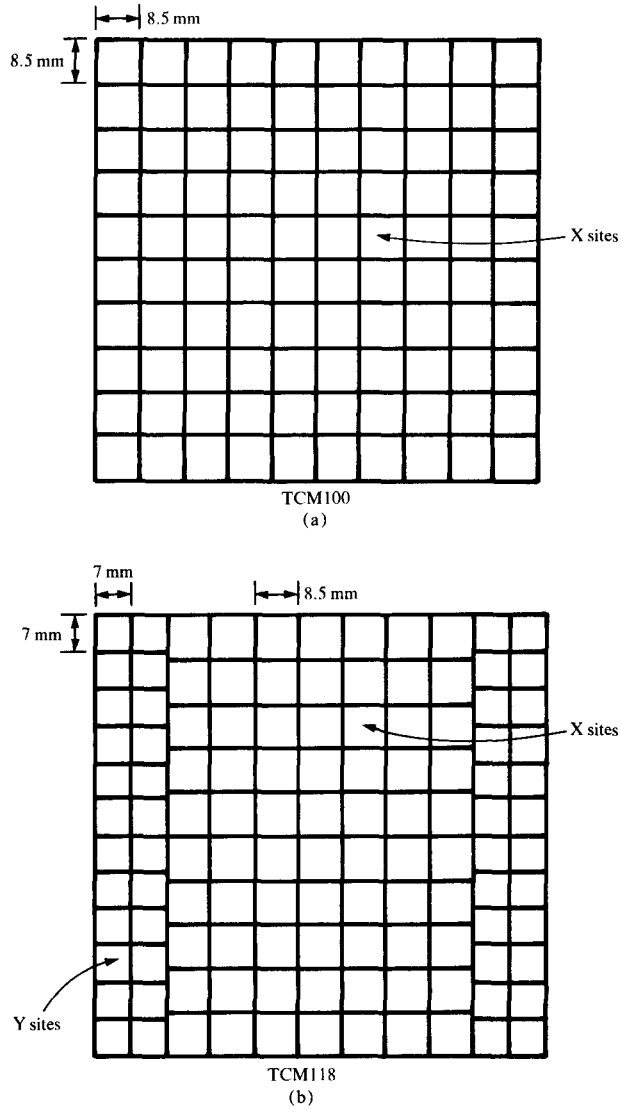


Figure 8 System 3081 chip layouts for the (a) 100-chip (TCM100) and (b) 118-chip (TCM118) modules indicating the X and Y chip sites.

$R_{c\text{-}p}$ also varies as a function of ρ; in the limit as $\rho \rightarrow \infty$, $R_{c\text{-}p}$ approaches the value for contacting flat surfaces. With no tilt, as ρ is increased from 150 to 200 mm, the $R_{c\text{-}p}$ decrease until a chip tilt of ≈0.05 mm is reached, at which point the $R_{c\text{-}p}$ increase as ρ is increased. The choice to use a particular ρ is primarily dependent on the chip tilt specifications. With a specified maximum of 0.076 mm for the 3081 TCM, ρ should be 150 mm. For the dimensions shown in Fig. 1(c), $R_{c\text{-}p}$ = 3.3°C/W for an average tilt of 0.025 mm; see Fig. 7(a). Likewise, analysis by NASTRAN of the resistances $R_{p\text{-}w}$ [Fig. 7(b)] and $R_{w\text{-}h}$ gives expected values of 3.5 and 1.5°C/W, respectively.

When the piston-to-chip tilt is compounded by having the piston tilt in the opposite direction to that of the chip [see Fig. 6(c)], R_{int} is decreased to 8.1°C/W. These resistance values, which include the chip spreading resistance commensurate with the nature of the tilt, should also be compared with the measured value of 8.4°C/W. Recalling the EOL degradation of 0.7°C/W, the total R_{int} becomes 9.1°C/W. These results are for an 8.5 × 8.5-mm chip site, which corresponds to the chip partition scheme on the 100-chip TCM (TCM100). However, in the 3081 system, there are both 100-chip and 118-chip (TCM118) modules on 90-mm substrates. Figure 8 shows the layout of the two modules. The 118-chip module has additional smaller chip sites at the two edges. These two types of chip sites are referred to as X and Y sites, as indicated in the figure. NASTRAN simulations (not given here) show that a change in the chip site does not affect $R_{c\text{-}p}$; however, because of changes in the cavity wall thickness, $R_{p\text{-}w}$ and $R_{w\text{-}h}$ are affected. The total EOL R_{int} for X and Y sites are calculated to be 9.1 and 10.0°C/W, respectively.

These total thermal resistances, which for convenience are called *internal thermal resistances,* were computed assuming that all heat generated by the chip leaves the module via the piston. These would be "true" internal thermal resistances if it were not for other heat paths from the chip to the

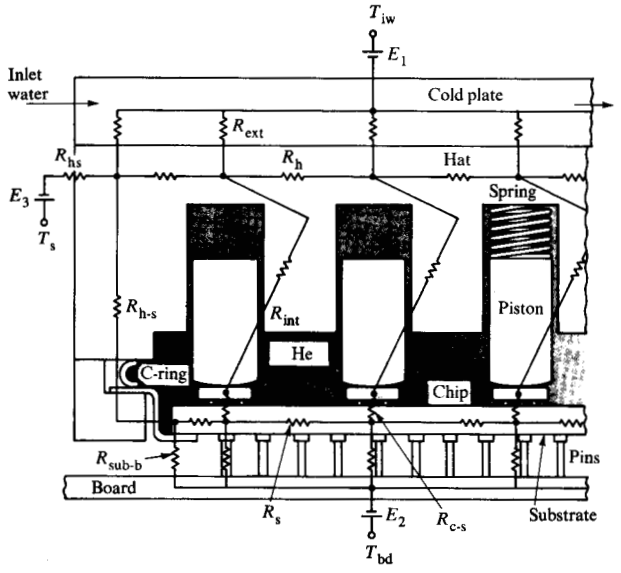


Figure 9 The ASTAP three-dimensional TCM thermal network. Dots (●) represent nodal points; currents into the chips represent chip powers, and convective heat losses from module surfaces to the environment are assumed to be zero.

substrate, to the hat through edges of the substrate, and to the board through the module pins. The various heat paths are shown in detail in the next section. For example, heat flow from the chip through the C-4 (*C*ontrolled *C*hip *C*ollapse *C*onnection) solder pads to the ceramic substrate can be appreciable, depending on the outer boundary conditions on the module (though not as much as it would be with back-bonded chips). A parametric study of temperature profiles in chips joined by C-4 pads was given in [9].

Multi-chip analysis

A schematic multi-chip ASTAP model of a TCM module is shown in Fig. 9. The values of the various resistances are indicated in Table 1. Tolerance limits and their assumed distributions, used to simulate the statistical distribution of chip temperatures in the field, are not given here for brevity.

The water temperatures shown are the average of inlet and outlet temperatures. Since the modules (maximum of three) are connected serially, the cold plates on downstream modules receive heated water (0.006°C/W at 6×10^{-5} m^3/s) from upstream modules. Therefore, water inlet temperatures are a function of module position on the board. Also, board and stiffener temperatures are simulated separately at the system level with given module powers and are then used as input to the ASTAP program.

Simulation of chip temperature distributions in a TCM is shown in Fig. 10. This module has 33 chips, but there can be as many as 118. In each square where there is a chip, the given nominal chip power dissipation P_c for that chip is indicated. The simulated nominal chip and substrate temperatures, T_c and T_{sub}, are given in Fig. 10(a). The lowest chip power is 0.8 W (location D6); the highest is 2 W (C3 and D7). Although the lowest chip temperature in this module (underlined) happens to correspond to the lowest chip power, this need not necessarily be so; this is also true with the highest chip temperature (also underlined) and power. This is because T_c is dependent on its own power as well as the power dissipation of its neighboring chips (*i.e.*, the heat flow is not one-dimensional up through the piston). In fact, the substrate temperature directly below the chip, T_{sub}, can be hotter than the chip itself (T_c), for example, for low-power chips being heated by adjacent regions (D6).

The simulation of field junction temperature T_j distributions is done statistically by varying the parameters listed in Table 1. As seen in Fig. 10(a), without an interposer, the minimum junction temperature in this module is <40°. Since the 3081 minimum junction temperature specification is 40°C, an interposer of appropriate thickness was placed between the module hat and the cold plate to act as a thermal insulator. With the aid of simulations such as those shown in Fig. 10(b), an interposer of 1.00 mm was designed for this module. The ensuing minimum, average (nominal), and maximum junction temperatures T_j^{min}, T_j^{nom}, and T_j^{max} are also shown. Thus, a minimum temperature of ≈47°C is found at J5 for the 1.1-W chip. (Again, this is not necessarily the chip with the lowest power dissipation.) By contrast, the minimum junction temperature for the 0.8-W chip (D6) is ≈48°C.

The junction temperature ranges for this module were plotted in Fig. 4, as were the temperature ranges of all other modules fitted with appropriate interposers. All junction temperatures are above the specified minimum of 40°C, and the average temperature of all the modules is approximately the same, *i.e.*, 55°C.

Experimental thermal analysis of TCMs

Direct verification of the simulation results with an actual product is not possible because there is no one isolated device, such as a diode, on the product chip with which actual temperatures can be measured. Early in the program, however, two special thermal chips were designed in which product chip power levels could be dissipated at the same time that the temperatures were measured on those chips. The thermal chips have a resistor plane which can be powered through the C-4 pads. (The temperature sensing is accomplished by calibrating the forward bias voltage of the twelve diodes in the chips by conventional methods.)

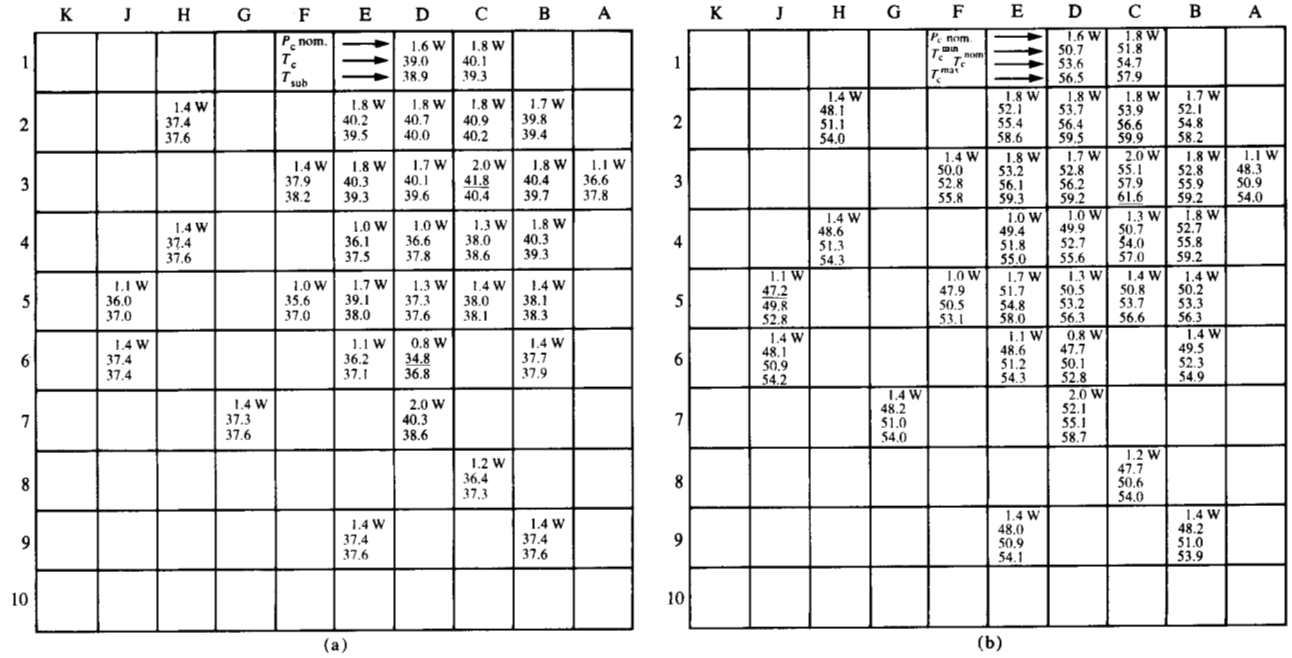


Figure 10 (a) Nominal time-zero chip and substrate temperature distributions in a TCM without interposer, and (b) statistical chip temperature distributions in the same module with a 1.00-mm interposer.

• *Single-chip experiments*

The single-chip tester (SCT), shown in Fig. 11, was designed primarily for the study of the gap thermal resistance between two contacting bodies, and specifically for the study of $R_{c\text{-}p}$ as a function of ρ and the tilt. The SCT chamber can be vacuum sealed so that different cooling media (air, He, and vacuum) can be used. The latter condition is especially useful for studying the nature of heat flow through asperities between contacting bodies as a function of surface characteristics and loading conditions. Most importantly, the direction of the heat flow can be controlled so that heat losses can be eliminated or taken into account for accurate calculation of thermal resistances. The piston block over the chip (see the figure) may be lifted and/or tilted accurately in increments of 2.5 μm by means of a precision motor.

Figure 12 shows a temperature trace of diodes 7 and 2, located at the center and edge of the thermal chip, as the piston is rolled on the chip. The origin of the axes is the center of the chip where the piston rests. With cooling water set at 17°C and chip power dissipation at 2.1 W, center diode 7 reaches 64°C and edge diode 2 reaches 66°C. The environment is He and the maximum vertical gap width at the edge of the chip from the spherical surface of the piston is 0.17 mm (ρ = 150 mm). The module temperature is adjusted to the mean chip temperature (≈65°C) with an auxiliary heater so that the heat flow is up through the piston.

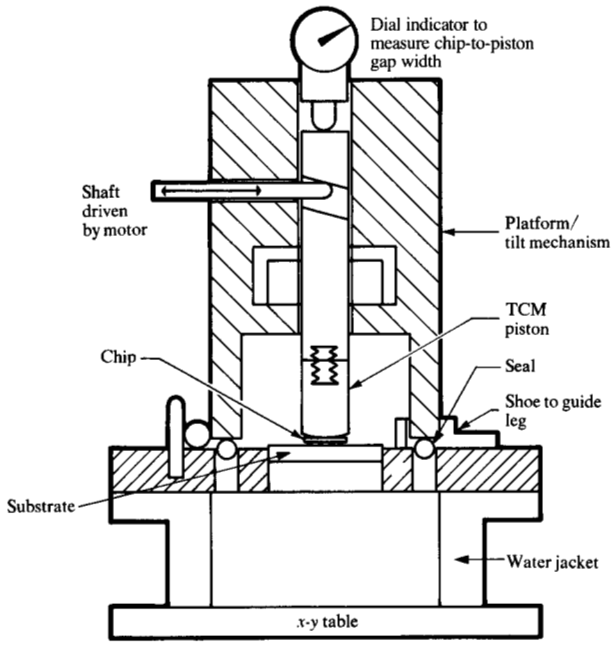


Figure 11 Single-chip tester (SCT).

As the piston rolls away from the center of the chip and towards the edge (defined as the positive direction), the center diode temperature increases while the edge diode

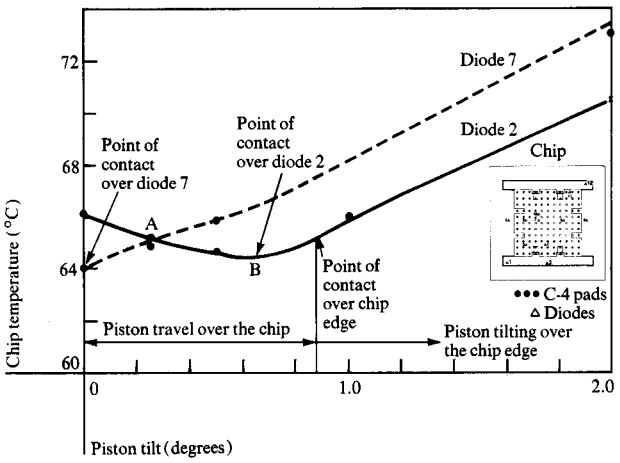


Figure 12 Actual measured chip temperatures as a function of piston tilt. The inset shows the position of the 12 diodes. Curves are for diode 2 (———) and diode 7 (- - -).

temperature decreases, as expected. At point A, the temperatures of both diodes are the same. The temperature at diode 2 continues to decrease until point B, the point of contact over the diode, is reached. Subsequently, both diode temperatures increase, as shown. Note that for diode 2 the piston contact point is closer to the edge of the chip than to the center of the chip (hence, the gap is smaller over diode 2); thus, the smaller temperature values for diode 2 compared to diode 7. With further tilt at the edge, the piston cocks and both diodes register high temperatures as a result of the large tilt.

Changes in T_c are reflected in changes in $R_{c\text{-}p}$. The measured $R_{c\text{-}p}$ for the flat and 150-mm crown pistons as a function of tilt are given in Fig. 2(a). A comparison of $R_{c\text{-}p}$ for the 150-mm crown piston with the NASTRAN simulations in Fig. 7 shows close agreement between the simulated and measured $R_{c\text{-}p}$ as a function of tilt for small tilts.

The chip spreading resistance included in $R_{c\text{-}p}$ depends on the chip-to-piston tilt and on the cooling medium; it is higher in air than in He for the same conditions. The measured ΔT across the chip (within 12 diodes) without tilt is 1°C/W in He and 1.9°C/W in air (150-mm crown piston). (Note that in the multi-chip experiments discussed in the next section, an average chip spreading resistance of 0.75°C/W is used.) The other thermal resistances, $R_{p\text{-}w}$ and $R_{w\text{-}h}$, may also be measured in the SCT by simply replacing the heat exchanging "block" in Fig. 11 with a single-chip section of an Al hat having dimensions appropriate for X or Y sites.

As mentioned earlier, the SCT was designed to measure the various intrinsic thermal resistances and, for this purpose, the heat flow is made unidirectional up through the piston. To measure thermal resistances actually experienced in the module, multi-chip experiments were conducted in product test vehicles.

- *Multi-chip experiments*

The multi-chip experiments were performed with 100- and 118-chip TCM test vehicles. Thus, a large engineering data base exists, ranging from strictly thermal characterizations to the study of the thermal behavior of TCMs as a function of various stress tests (*e.g.*, vibration, shock, thermal shock, and power cycle) to field simulations corresponding to the power map in each module of the 3081.

None of the stress tests adversely affected the failure rate objectives of the 3081 program. We discuss only those experiments directly related to the thermal characteristics of the TCM and the chip temperature distributions in the field. The modules are first calibrated thermally and then subjected to a series of tests. Each test consists of uniformly powering all the chips at 1, 2, 3, and 4 W. For each power level, the chip, hat, and substrate temperatures are measured and entered into a computer. Based on other calibration data already in the computer, average internal thermal resistances for each power level are calculated by plotting the differences between the average chip temperature and the temperature at the center of the hat against the chip power (in this case, a linear plot with slope = 7.16°C/W and an intercept of 0.050°C). An average R_{ext} is also calculated by dividing the difference between the average hat temperature and the inlet water temperature at the 6×10^{-5}-m^3/s flow rate by the total module power. The slope of the curve gives R_{int}. (Note that the intercept must be within certain limits, here ±0.05°C.) The value for R_{int} must be corrected for heat losses to the thermal test system to which the module connector is mounted. With extensive tests, it has been shown that 6% of the power generated by the chips is dissipated through the computer board; 94% finds its way to the cold plate via the piston and substrate and to the hat. With this correction, and including chip spreading resistance of 0.75°C/W, R_{int} = 8.4°C/W, in agreement with the NASTRAN simulation value. With the EOL degradation of 0.70°C/W (discussed earlier), the EOL R_{int} is 9.1°C/W. The overall R_{ext} is computed by noting ΔT and dividing by 94% of the power; this yields 0.020°C/W without the interposer (not including EOL).

Similarly, TCM test vehicles may be powered in different configurations depending on the particular power map used. The modules indicated in Fig. 4 were simulated with chip layouts closely representing actual field conditions. Once these nominal results were verified against the ASTAP runs, all the statistical field temperature distributions were established.

Development of encapsulation

As stated previously, a critical requirement of the thermal design of the TCM is the use of He gas in the interface between the chip and piston, and in the annulus between the piston and wall. Sufficient He levels must be maintained over the expected life of the product to achieve the required thermal resistance values, and the module must be capable of multiple openings and closures.

Although He possesses excellent thermal properties for a gas and is inert, nontoxic, and nonflammable, it is extremely difficult to contain in a sealed enclosure. In fact, it is often used to measure extremely low leak rates in hermetic testing applications. In addition, under molecular flow conditions, air will enter the enclosure at the same time that He is leaking out. Encapsulation ensures a stable thermal performance through control of the internal volume and leak rate of the module, the initial gas fill pressure, the initial He/air mixture, and the required product lifetime. Using a lifetime degradation of 0.7°C/W as the thermal objective, an internal volume of 8.9×10^{-5} m^3, a fill pressure of 1.6×10^5 Pa, an initial 1% air contamination, and a lifetime of 10^5 power-on hours (POH), an initial leak rate of 5×10^{-9} Pa-m^3/s of air is required. This assumes that the average leak rate over the module lifetime does not degrade below the time-zero specification. Based on qualification data (discussed later), this is a valid assumption. Thus, the specific seal design must exhibit an extremely low leak rate over ≈400 mm and must withstand the required mechanical and thermal stresses and the chemical environment.

From a number of potential seal designs, a rectangular ®Inconel C-ring, with a 2.36-mm outside cross-sectional diameter, a 0.38-mm wall thickness, and 103 × 113-mm side lengths, was selected based on its sealability, reworkability, reliability, and availability. A 0.1-mm surface coating of electroplated lead was determined to provide the best overall sealability.

An extensive evaluation program was conducted to assess the initial and time-dependent sealing characteristics under various environmental conditions, including thermal cycling, impact shock, thermal shock, the presences of hostile gases, sliding wear, temperature, and humidity. The key reliability test involved thermal cycling since it appeared to be the most effective for detecting time-dependent sealing problems. The nature of these problems does not permit rigorous analytical testing; thus solutions are often dependent on empirical data. Typically, test periods required 100 days to complete 3600 cycles of a 10°C-to-60°C test. In a standard test chamber, the number of parts that could be cycled was limited to two or three because of the significant thermal mass of the module. Therefore, a water-bath cycling system [10] was designed and built to utilize the higher heat transfer rate of water as compared to air. Using a test cycle of ≈4 min for a 25°C-to-75°C test, the 3600 cycles could be completed in 10 days. In addition, up to 30 modules could be tested simultaneously, thereby further enhancing the evaluation program. In total, over 200 parts were subjected to this test environment. The test was used to establish the seal design and process and to demonstrate repeatability over a range of variables and time spans. This development program resulted in repeatable, reliable hermetic seals which satisfied the encapsulation objectives.

Conclusions

Independent assessment of the module was made in a product qualification step and the product performance was verified against the defined objectives and quantized module functional and reliability characteristics. The final TCM product contains on the average 25 000 logic circuits and 65 000 array bits packaged in 0.001 m^3. The TCM, as mounted to the cold plate, can dissipate up to 300 W and maintain all chip junction temperatures between 40 and 85°C. The TCM development program resulted in a product which met all key performance and reliability objectives, and represented a major breakthrough in the full utilization of advanced LSI devices.

The TCM possesses considerable extendability for future applications, particularly in terms of its cooling capability. Through reduction in internal and external thermal resistances, the allowable power dissipation can be increased. By further dimensional optimization and enhancement of the interface gaps with higher-thermal-conductivity media [11, 12], the internal resistance can be lowered. Similarly, reduction or elimination of external interface resistances [13] can increase thermal capability. Thus, with reasonable design changes at the module and system levels, power dissipation limits can be extended by over 50%.

Acknowledgments

We wish to thank U. P. Hwang for his early analytical calculations and thermal evaluation of the TCMs, M. Mansuria for his extensive computer simulations of chips and modules, and L. Tomaine for his sustained diligence in getting many modules built for thermal and environmental evaluation.

References and notes

1. (a) 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**, 45–54 (this issue, 1982). (b) R. C. Chu, O. R. Gupta, U. P. Hwang, and R. E. Simons, "Gas Encapsulated Cooling Module," U. S. Patent 3,993,123, November 23, 1976.
2. The ®Lexan material is a polycarbonate with a thermal conductivity of 0.238 W/m-°C.
3. R. E. Simons, "Thermal Interface Adapter for a Conduction Cooling Module," U. S. Patent 4,167,771, September 11, 1979.

4. L. C. Tomaine, "Thermal Conduction Piston for Semiconductor Package," *IBM Tech. Disclosure Bull.* **21,** 609–610 (1978).
5. In general, *thermal resistance* R is defined as $R = \Delta T/Q$, where ΔT is the temperature difference and Q is the heat flow between two sections in space. This relationship is a consequence of Fourier's conduction equation, $Q = RA(\Delta T/L)$, where A is the cross-sectional area normal to the heat flow, L is the distance through which the heat flows, and R is the thermal conductivity of the medium.
6. See for example S. S. Burde and M. M. Yovanovich, "Thermal Resistance at Smooth-Sphere/Rough-Flat Contacts: Theoretical Analysis," *AIAA Paper No. 78-871,* presented at the Second Amer. Inst. Astronautics and Aeronautics (AIAA)/ Amer. Soc. Mech. Engineers (ASME) Thermal Physics and Heat Transfer Conference, 1978.
7. See the NASTRAN thermal analyzer manuals, *NASA TMX-3503* and *NASA TMX-3504,* April 1977, available from the National Aeronautics and Space Administration, Washington, DC.
8. See the *ASTAP Reference Manual,* Order No. SH20-1118, available through IBM branch offices.
9. S. Oktay, "Parametric Study of Temperature Profiles in Chips Joined by Controlled Collapse Techniques," *IBM J. Res. Develop.* **13,** 272–285 (1969).
10. L. D. Lipschutz, "Thermal Cycling Tester," *IBM Tech. Disclosure Bull.* **21,** 1565 (1978).
11. U. P. Hwang, S. Oktay, A. L. Pascuzzo, and A. C. Wong, "Conduction Cooling Module," *IBM Tech. Disclosure Bull.* **20,** 4334–4335 (1978).
12. U. P. Hwang and S. Oktay, "Thermal Interface Conduction Pad," *IBM Tech. Disclosure Bull.* **21,** 4028 (1979).
13. U. P. Hwang, K. P. Moran, S. Oktay, and A. L. Pascuzzo, "Liquid Encapsulated Conduction Module," *IBM Tech. Disclosure Bull.* **20,** 4780–4781 (1978).

Received June 5, 1980; revised June 17, 1981

S. Oktay is located at the IBM General Technology Division laboratory, East Fishkill (Hopewell Junction), New York 12533; H. C. Kammerer is located at the IBM Data Systems Division laboratory, Poughkeepsie, New York 12602.