

Chapter 1

Thermal Management Fundamentals and Design Guides in Electronic Packaging

Abstract With increased performance requirements for smaller, more capable, and more efficient electronic systems, thermal challenges have become critical issues in electronic packaging design. Breakthroughs are needed in advanced cooling solutions and pragmatic design at chip, board, and system package levels. In response to these critical needs, advanced materials and process improvements in packaging and cooling technology are required to provide high thermal transfer efficiency, well-controlled thermal transient behavior, environmental compatibility, low weight, and acceptable material and fabrication cost. The lowest device temperature and expected reliability will be achieved when the lowest package thermal resistance is achieved. This chapter provides an outline of thermal challenges in advanced electronic packaging, heat transfer theory, and its application in thermal design, development of thermal management solutions, advanced thermal management materials selection and component design guideline, as well as environmental compatibility requirements in thermal management of electronic packaging.

Rationale of Thermal Management

Electronic packaging of microelectronics can be identified as the art of enclosing, interconnecting, powering, cooling, chip packaging, and protecting chips from the ambient environment as well as protecting the ambient environment from the chips (Rasmussen 2003). It provides signal and power transmission, thermal dissipation, electromagnetic interference (EMI) shielding as well as physical and environmental protection, and act as the bridge that interconnects the integrated circuit (IC) chips and other components into an electronic system to form electronic products. Electronic packaging of a typical electronic system can be divided into several packaging levels from bare chip, packaged chip, printed circuit board assembly, electronic subassembly, and electronic assembly to system or final electronic product. The trend in electronic packaging is to simplify and/or reduce the number of packaging levels.

From a thermal management perspective, electronic packaging is usually classified as chip-level packaging, board-level packaging, and system-level packaging. Each level of the packaging has distinctive interconnection devices associated

with it. Chip level package usually refers to packaged electronic functional devices (e.g., active, passive, or electromechanical devices, etc.) or packaging of silicon chips into dual-in-line packages, small outline ICs, chip carriers, multichip packages, and the chip level interconnects that join the chip to the lead frames, or utilizing tape-automated bonding or chip on board to assemble the chip. The board-level packaging is the assembly of the chip or multichips together with additional components, such as capacitors, resistors, inductors, switches, etc. on a printed circuit board (PCB). Printed conductor paths connect the leads of components to PCBs and to the electrical edge connectors for off-the-board interconnection. System-level packaging is normally the outer enclosure or shell of the electronics, such as the casing of a hearing aid, mobile phone, electronic organizer, etc., with PCB-to-PCB interconnections or card-to-motherboard interconnections, or it can be a rack or frame hold several shelves of subassemblies that are connected together to make up a complete system or a final electronic product. In the past years, increasing frequency and power density coupled with lower product costs have been driving new packaging technologies. Some examples of this are migration from wirebond to flipchip interconnect, higher levels of integration in semiconductors and increased usage of hybrids and multichip modules, such as multicore processors and system-in-package. This trend in microprocessor architecture results in increased heat densities which mandate that thermal management be given a high priority in electronic packaging design so as to maintain system performance and reliability.

With continuous demands of high performance, low cost, and miniaturized microelectronic devices, electronic packaging will continue to undergo evolution at every new generation of microelectronics and product technologies. These changes have been introduced and will continue creating a new set of changes that require advanced packaging technology including innovative thermal management solutions. In addition to the technical challenges, market forces such as declining product prices, increased user experience through miniaturized devices, wireless connectivity, and longer battery life would make these challenges even more complex (Mallik et al. 2005). The challenges in thermal management can be viewed in terms of three different but nonseparable problems (Hannemann 2003). (1) The chip temperature must be maintained at a relatively low level despite high local heat density. (2) High heat loads must be handled at the assembly or module level. (3) The thermal environment of an electronic system, such as the computer machine room, office space, or telecommunications central office must be controlled and the overall rack heat load dealt with. As a result, thermal management would be a serious concern behind any new electronic product designs. Two major objectives should be achieved through thermal management: Prevention of catastrophic thermal failure, and extension of the useful lifetime of the electronic system. Catastrophic thermal failure is usually the result of semiconductor material failure as a result of overheating, or thermal fracture of a mechanical element in an electronic package, such as case or substrate. The failure rate increases exponentially with operating temperature.

Heat Sources and Thermal Effects on Integrated Circuit Operation

The insatiable demand for higher performance processors has led to a steady escalation in power consumption across all the market segments, such as mobile and performance desktops as well as servers and workstations (Viswanath et al. 2000). Increasing power density and current levels in the microprocessors have been main heat sources and cause concerns to do the thermal management of the on-chip hotspots as well as the package and interconnection Joule heating. The result in temperature increasing will affect the operation of the active and passive devices in the integrated circuits. If the temperature increase is high enough, the active or passive devices being heated may either permanently degrade or even totally fail. Such failures include thermal runaway, junction failure, metallization failure, corrosion, resistor drift, and electromigration diffusion (Krum 2004). Therefore, it is crucial to minimize any temperature increase in an electronic package.

Power Density

In general, for microprocessors such as complementary metal oxide semiconductor (CMOS), power dissipation is proportional to the capacitance of the logic elements, the square of the operating voltage swing, and the operating frequency (Hannemann 2003):

$$P \approx NCV^2f, \quad (1.1)$$

where P is power dissipation of CMOS in Watt; N is the number of devices per chip; C is the capacitance of the logic elements in Farad; V is the operating voltage in Volt; and f is the operating frequency in Hz. While the logic element capacitance declines with feature size, and while operating voltages have been significantly reduced, the increase in the number of devices per chip and the operating frequency have driven power levels for the next generation of microprocessors to very high levels (70–200 W). As the frequency scales higher over time, so does the power dissipation of the microprocessors. The improvements in process, such as introduction of multicore processors have been able to hold the power increase to reasonable levels, but it is definitely trended higher. A similar trend is reflected in the average heat flux (power dissipated per unit die area) on the processor, indicating a linear increase over time. This is due to the fact that the power reduction obtained from architecture and process modifications is not commensurate with the scaling in die size, voltage, and frequency to support a cap in power consumption. In addition, the wider range of power and frequency offerings will enable performance and cost trade offs to be made between the various market segments. The need for higher performance and an increased level of functional integration as well as die size optimization on the microprocessor leads to preferential clustering of higher power

units on the processor. In turn, this leads to a higher heat-flux concentration in certain areas of the die and lower heat fluxes in other regions on the die, which manifests itself as large temperature gradients on the die (Viswanath et al. 2000). This thermal non-uniformity typically refers to hotspots, where power densities of 300 W/cm^2 or over are possible. To help quantify the nonuniform power effects, a density factor (DF) has been introduced, which is the ratio of the actual package thermal resistance at the hottest spot to the die-area-normalized uniform power resistance or thermal impedance, and has the unit of inverse area (A^{-1}). DF can be used to quantify the impact of nonuniform die heating on thermal management by the following equation (Mallik et al. 2005):

$$\psi_{\text{jc}} = R_{\text{jc}} \times \text{DF}, \quad (1.2)$$

where ψ_{jc} is the junction-to-case thermal resistance of the package; and the R_{jc} is the thermal interface material resistance. It can be seen that for the same R_{jc} , the power maps with a higher DF will result in a higher package thermal resistance, which in turn requires more effective thermal management solutions. The future trend is that the DF is increasing for emerging generation of microprocessor architectures due to the increasing local power density at the hotspots, therefore, the thermal management solutions have to meet more stringent heat-flux requirements that are significant multiples of the average heat flux at some areas of the chip-package interface.

Joule Heating

Joule heating is generated when current passes through a resistor, following Joule's law: $P = I^2R$, where P is power (W), I is the current (A), and R is resistance (Ω). As current levels continue to rise, Joule heating along the power delivery path becomes significant. Because it is not uncommon for a high-end microprocessor to draw currents in excess of 100 A, for instance, even a resistance of $0.5 \text{ m}\Omega$ along the path will result in a power dissipation of 5 W. In an electronic package, higher current combined with the need to reduce the package size, i.e., thinner and narrower conductors and finer pitch power delivery interconnects, would lead to a high amount of heat generated within the package and electronic interconnection system. This would require thermal management of the entire interconnection system which may include the flip chip joints, the substrate, the socket, and the attaching solder balls (Mallik et al. 2005). For example, one challenging key area is the increasing temperature of the flip-chip die bumps. Without proper attention to thermal design, the bump temperature could be significantly higher than the transistor junction temperature causing bump electromigration problems. The joule heating of interconnects on the package can be effectively managed through careful design, which involves minimizing current concentration and spreading heat through thermal design and thermal management materials selection.

Thermal Failure Induced by Different Coefficient of Thermal Expansions

An electronic package is comprised of various conducting and insulating materials which have different coefficients of thermal expansions (CTE). Figure 1.1 illustrates a cross-sectional view of a ceramic ball grid array package assembly, where a silicon chip is mounted on a multilayer ceramic substrate module through solder joint embedded in epoxy underfill, and the module is attached to a printed circuit board (PCB) through solder ball interconnections to form a final second level assembly. In addition, a metal heat spreader can be attached to the silicon chip through thermal interface material, or a heat sink is attached to the module to dissipate the excessive heat. When the chip is powered up so that the package is subjected to a temperature change, each material with different CTE deforms at a different rate. This nonuniform CTE distribution will produce thermally induced mechanical stresses within the package assembly.

When the assembly is cooled from an assembly temperature, for instance the PCB contracts more than the module. This uneven contraction produces a global bending of the whole assembly as well as relative horizontal displacements between the top and bottom of solder balls. When it is cooled to room temperature, the free thermal contraction of the solder joint at the interfaces is constrained by adjacent materials which have a lower CTE. In general, if the global effect reinforces the local effect at a point in the package, the concentrated strain will be accumulated during thermal cycles, which would result in the premature failure of the device during operation, such as heat sink bending, the thermal interface or solder joint failing, the ceramic substrate warping, or completely failing and cracking. To reduce thermal stress or eliminate thermal failure requires both a selection of the proper materials and a minimization of the temperature changes through thermal management.

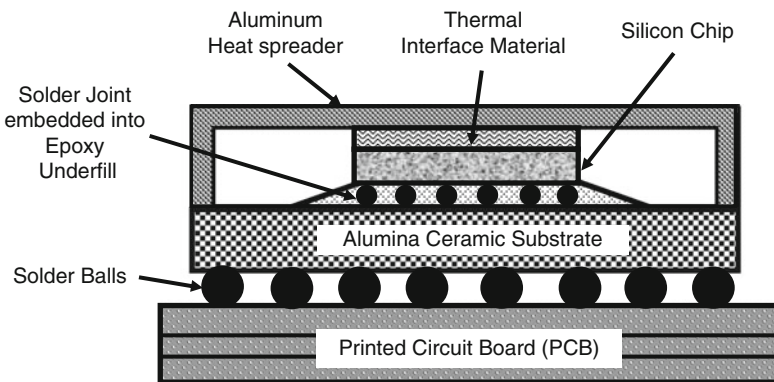


Fig. 1.1 Schematic illustration of a typical ceramic ball grid array package assembly

Thermal Failure Rates

The failure rate of an electronic component can be estimated by the Arrhenius equation (Krum 2004):

$$f = A\varepsilon^{-e_A/kT}, \quad (1.3)$$

where A is Arrhenius constant; f is failure rate, e_A is activation energy in electron volts (eV); k is Boltzmann's constant (8.63×10^{-5} eV/K); and T is junction temperature in K . The activation energies vary for different failure mechanisms, for example, $e_A = 1.3$ eV for charge injection; 1.0 eV for contamination; 0.53–0.70 eV for corrosion; 0.68–0.95 eV for electromigration; 0.3–0.7 eV for gate oxide; 0.55–1.0 eV for gold–aluminum film joint; 0.68–0.95 for hillock; 0.73 eV for aluminum wire–gold film; and 0.69–1.0 eV for gold–aluminum wire couples (Krum 2004). For an activation energy of 0.65, the failure rate would increase by a factor of 2.0 for a temperature increase from 50°C to 60°C. This 0.65 eV value of activation energy is usually used as a rule of thumb, i.e., for every 10°C increase in temperature, the failure rate doubles.

Thermal Management Challenges and Common Concerns

The current trend in microprocessor architecture is to increase the level of integration (higher power), shrink processor size (smaller die), and increase clock speeds (higher frequency). This results in an increase in both the raw power as well as the power density on silicon. The drive to manage yield and reliability is resulting in the need for lower operating temperatures. There are two major reasons to maintain the operating temperature of a device at a certain level. (1) The reliability of circuits (transistors) is exponentially dependent on the operating temperature of the junction. As such, small differences in operating temperature (order of 10–15°C) can result in approximately two times the difference in the lifespan of the devices. (2) The other factor is the speed of the microprocessor. At lower operating temperatures, microprocessors can operate at higher speeds due to reduced gate delay. A secondary effect of lower temperatures is related to a reduction in idle power dissipation (also known as leakage power) of the devices, which manifests itself as reduction in overall power dissipation. These two factors combined dictate the operating temperature of devices as a function of the speed of the device (Viswanath et al. 2000).

This in turn translates to a shrinkage in temperature budgets for thermal design. Careful management of the thermal design space from the silicon to the system level is therefore critical to ensure a viable solution space for succeeding generations of processors (Viswanath et al. 2000). On another hand, the continuing increase in electronic packaging density has resulted in demands for advanced packaging materials with low weight, low cost, high-thermal conductivities, and matched CTEs with

those of ceramic substrates and semiconductors to minimize thermal stresses that can cause component or interconnection failure. Various more comprehensive and accurate modeling, and analysis of the electronic packaging systems are needed to understand the thermal behavior and the behavior for cost efficiency, and optimize the thermal management, which requires identification of critical issues such as shock, vibration, and usage scenarios, considering potential coupling effects, and developing integrated, interdisciplinary solutions. Systematic thermal solutions also need universal design regulations and thermal testing standards.

These challenges have involved all electronic market segments. For instance, the mobile processor market segment is typically constrained by battery life ($\sim 2\text{--}3$ h), and the form-factor must be small and light to allow for portability. If the desktop market is more cost sensitive, the mobile market is more space and weight sensitive. These sensitivities place boundaries on the effective power removal capabilities of the electronic package. The long-term solution would be: (1) The design and architecture of the microprocessor must be such that it optimizes performance and power consumption; (2) New cost-effective technologies in microprocessor and system packaging including effective thermal management solutions must be developed.

As a result, the common concerns of the thermal management for electronic packaging can be summarized as follows.

1. For the current generation of microelectronic products, the two main areas of concern are (a) the thermal control of high-power microprocessors and/or similar ultra-large scale integration components, given the very high local heat density; and (b) the thermal management of the high overall heat loads generated by equipment in racks and machine rooms (Hannemann 2003). For example, telecommunications equipment rooms, whether central offices or smaller remote huts, have reached their limit in terms of power density. Computer facilities have lower overall heat loads, but concentrations of power within these rooms are a serious issue. Office areas (typified by the “workstations” data) have rising heat loads that within an open office environment pose a very significant HVAC (heating, ventilating, and air conditioning) problem. These problems are beginning to be addressed with liquid-cooled equipment frames (allowing heat to be transported to a remote location) and spot cooling units (Hannemann 2003). Furthermore, telecommunications central offices are already stretching power availability and power handling limits. Photonic components used in these systems (as well as data networking systems) have unique and serious microcooling and materials challenges. For example, semiconductor lasers have chip-level heat fluxes on the order of $2,000\text{ W/cm}^2$. These devices also have very narrow thermal control limits, because performance is significantly impacted by operating temperature (e.g., wavelength). Historically, this segment has been a major applications space for thermoelectric cooling—and conventional, cost-effective thermoelectric coolers are reaching their limits in newer applications (Hannemann 2003).
2. For these systems, such as mainframe computer systems, storage arrays, server farms, and network offices, localized thermal control for very high-power chips

is a major issue as chip powers increase beyond 100 W. Ducted forced air cooling and active heat sinks (heat sinks with small, dedicated air movers) will continue to be used, but miniaturized liquid cooling loops and in some cases microelectromechanical systems-based coolers such as embedded droplet impingement for integrated cooling of electronics require development. These devices, however, must meet stringent cost and reliability requirements—no more than \$100 or so as a target cost, and perhaps 1,000–5,000 failures in time (one failure per 10^9 h) for reliability performance. Also, rack-level cooling via liquid cooled frames and compact rack-mountable air conditioners (both meeting size, cost and reliability targets) are of interest. While these cooling approaches may seem straightforward, components such as compressors and extended surface arrays meeting the size, cost, and lifetime constraints are not yet available. Furthermore, equipment rooms also pose HVAC design problems. Improved modeling tools and sensing and control devices are needed (Hannemann 2003).

3. Office systems include high performance workstations, printers, desktop computers, and office- and equipment-closet installed networking and telecommunications products. Cost/performance microprocessors and networking chips will reach 80–100 W. The cooling of these chips will pose unique challenges because some options open to large system designers will be unavailable. Cooling solutions must be extremely compact, quiet, and cost effective. Proven-reliability microheat pipes, breakthrough microsized liquid loops, and especially optimized active heat sinking systems may provide thermal management solutions. Acoustic performance will remain very important in this segment (Hannemann 2003).
4. Other important electronic systems must operate in harsh environments, calling for ruggedized cooling solutions with some unique requirements (such as transient high-heat loads) but no or less pressure on cost and reliability. These applications include automotive, telecommunication, space, and military systems (Hannemann 2003).
5. Another challenge is that the increasing emphasis placed on thermal design at the early design stages has inadvertently increased electromagnetic compliance (EMC) problems. Thermal design often conflicts with EMC design, so fixes that are implemented to address thermal concerns often exacerbate or create EMC problems. The most obvious example is that thermal design requires large holes to enable adequate air flow while EMC design requires small holes to reduce emissions. Another example is that thermal connections (metal studs) are often added to conduct heat away from a hot component. For example, in a small module where forced cooling is not available, it may be necessary to conduct the heat to the enclosure. Switching currents can couple to the stud and cause it to radiate like an antenna. In thermal design, a large surface area is often used to increase convective heat flow. The same large surface areas may also increase capacitive coupling to the enclosure, which enables displacement current to flow from components such as heat sinks to the chassis and then onto cables connected to the chassis (Manning and Johns 2007). Therefore, the concurrent thermal and EMC design process should be implemented.

Consequently, the importance of thermal management has waxed and waned through the technology and product generations of the past decades. Current integrated circuit and photonic technologies as well as the ubiquity of electronic system applications are once again providing a serious challenge to thermal management in terms of basic theory, tools, components, and innovative design. Breakthroughs are needed in advanced cooling and pragmatic design at all levels (Hannemann 2003). The thermal engineering for electronics is on the cusp of a renaissance.

Overall Picture of Thermal Management in Different Packaging Levels

With increasing power densities, the thermal management becomes more and more a central issue in electronic packaging. An overall picture of thermal management in different packaging levels including chip level, board level and system level is shown in Figure 1.2. Chip-level thermal management mainly relates to dissipating and heat spreading within the semiconductor device and chip package. The thermal management technologies include conventional copper, copper-tungsten, or copper-moly heat spreaders as well as advanced thermal design, on-chip thermoelectric cooling, microscale heat pipes, and microchannel cooling, etc. Board-level thermal management indicates heat transfer or spreading through PCB and from the chip

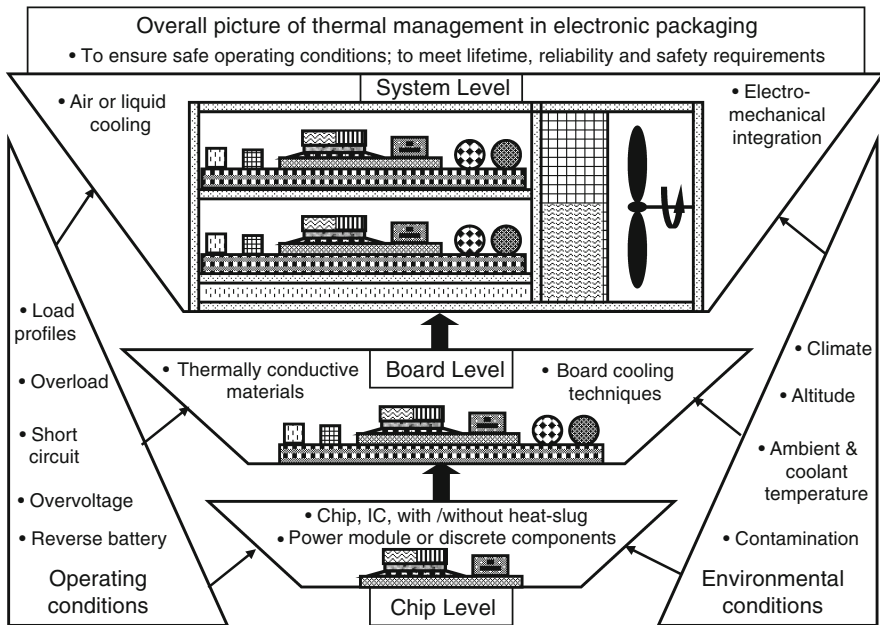


Fig. 1.2 Overall picture of thermal management in different electronic package levels

package to the chassis or electronic system. Conventional technologies include copper and aluminum spreaders. System-level thermal management refers to heat dissipation through system heat exchanger and heat transport from the chassis to the system heat exchanger. Conventional technologies include air-cooled heat sink and radiators, forced air flow, and pumped liquid loops. For some high-power semiconductor devices, a cooling mechanism may directly connect the chip package to the system heat exchanger by passing board level. This reduces the number of thermal interfaces and improves the cooling efficiency, but may be at a price of added costs.

Although the thermal design of the chip, board, and system levels shares the common places, each level has its unique cooling requirements. For example, chip-level cooling designs are dominated by high-heat flux and miniaturization issues. Board level requires transport of large amounts of heat with minimal temperature gradients. System-level heat exchanger designs are required to sink the most heat with minimal volume and mass. In some cases where the temperature of the environment varies greatly, thermal or temperature control mechanisms are needed to protect the electronics by maintaining a relatively constant temperature. The temperature control may be implemented at different levels. For example, variable speed air blowers or liquid bypass can be added to board and system levels to adjust the cooling effectiveness to realize temperature control. With development of the innovative electronic packaging accompanied by advanced thermal management technologies, however, the picture of thermal management in different package levels would undoubtedly change.

Whatever change or modification may occur, the basic target of thermal management is always to ensure safe and reliable thermal rating conditions for all electronic packaging components, including active and passive components, die or chip, PCB, interconnections, subsystem, and system structures. In consideration of the individual system requirements on lifetime, reliability, and safety, this goes far beyond a simple check on compliance with the maximum rating conditions of each component. Special attention must be paid to lifetime data and useful life deratings, which in general are not only dependent on the absolute temperature but also on the temperature cycling during operation. Thermal management starts at the device and chip level, for instance, with or without a heat-slug for an SMT (surface mount technology) device. Such a small technical decision can basically determine the thermal behavior and reliability of the whole system just as many other decisions do on the board and system level. Each decision must take into account the operating and environmental conditions specified for the final system. Even things which at first glance have nothing to do with thermal management can have their impact here, for example, waste or recycling regulations which aim at a better material separability, reduction of encapsulation materials, or ban of certain materials. Because of the far-reaching consequences of many decisions and the high-cost risks of a redesign, it is important that the thermal management is not considered a separate work-package within a design flow, but as a continuous process accompanying the whole system development. This becomes more and more important with the continuously increasing power densities of modern power electronic systems (März 2003).

Chip Level Packaging Thermal Management

The emergence of nanoelectronics has led to localized areas of high-heat flux that are dominating the performance of electronics at the chip level. While the traditional thermal management techniques mainly provided system-level cooling sufficient to meet past electronics cooling requirements, they are facing challenges to meet future chip-level cooling requirements in most of their current forms. Therefore, chip-level thermal management has been explored and focused on advanced chip-level thermal design, and miniaturization of thermal management techniques to the microscale, which is receiving increased attention as a solution to the future thermal management technology shortfall. Microscale thermal management offers several enticing opportunities (Moran 2001): (1) The ability to “spot cool” high-heat flux regions with unparalleled resolution to bring down critical junction temperatures; (2) Potential for macrolevel performance leaps by optimizing microlevel heat transfer; (3) Improved integration of thermal management at the chip level using compatible semiconductor materials and fabrication techniques; (4) Enabling of board-level and system-level miniaturization to support the pervasive trend toward increased capabilities in smaller devices.

There has been a wide range of new technology developments in chip-level microscale thermal management, such as microscale channel cooling, wicking structure, enhanced thermal conduction, and refrigerators or coolers. For instance, microchannels have been utilized to provide increased surface area for improved heat transfer. These devices are generally limited to laminar flow regimens because the increased surface area also increases pressure drop and the subsequent pumping power required. The channels can be micromachined from bulk semiconductor material, or fabricated by other means with a variety of materials. Microtubes, for example, have been developed using copper, silver, platinum, glasses, polymers, alloys, and other materials with diameters ranging from 0.5 to 410 μm (Moran 2001). Used in conjunction with evaporator and condenser regions, microscale wicking structures have been developed and become a critical component of microheat pipes. The wicking structure can be fabricated with a coherent pattern of 2.5 μm diameter pores (channels) which length is equal to 260 μm , utilizing surface tension forces of the working fluid to induce flow without active pumping. The wick is part of a micromachined silicon loop heat pipe concept for removing heat at the chip level with a predicted heat flux cooling capacity greater than 100 W/cm^2 (Moran 2001). Nanomaterials have been explored to offer enhanced microscale thermal conduction. For example, nanocrystalline diamond film from fullerenes resulting in freestanding diamond structures as thin as 0.3 μm has particular interest in high-performance electronics cooling due to their unusual combination of superior thermal conductivity ($\sim 2,000 \text{ W}/\text{m K}$), yet low electrical conductivity. Carbon nanotubes offer the potential for ultrahigh strength structures with outstanding thermal conduction characteristics (e.g., $>5,000 \text{ W}/\text{m K}$). In addition, microrefrigerators and coolers, such as thin-film thermoelectric cooler and miniaturized Stirling cycle refrigerator, have a unique characteristic for chip-level cooling that

differentiates them from all other potential thermal management technologies: the ability to generate cooling temperatures well below the ambient temperature. This key advantage potentially allows junction temperatures to be driven much lower resulting in improved reliability, faster performance, and operation in higher temperature environments (Moran 2001).

On the other hand, optimal thermal design and proper materials selection in chip-level packaging plays an important role in thermal management. In most of advanced electronic devices, the bulk of the heat is concentrated to the active transistors on a semiconductor chip, assembled in some kind of package. In general, this package is mounted on a PCB and attached to electric conductors. The transfer of the heat that is originated on the transistor junction goes to the lead frame and further on to connecting pins of the packaging to some extent via thin bonding wires, but mainly by conduction paths as the adhesive between chip and other parts of the packaging. As the connecting pins are usually soldered to electric conductors of copper, a metallic bridge is formed for the continued heat transfer and the distribution of the heat to the PCB, some of the connecting pins can also have only a heat conducting function. As the dielectric material of the PCB is not a good heat conductor, it is important to include as much metal as possible in the stack-up of the PCB; in this way the heat is spread effectively and hotspots are flattened. Preferably this metal is in most cases copper, as copper is an excellent conductor of both current and heat. As in this interpretation of the heat transport, the temperature gradient is directed from the junction toward the PCB, the PCB must be the colder part. This means the PCB must be cooled by a proper cooling system, which can effectively dissipate the heat through conduction, convection, radiation or combined approaches (Siebert 2005).

Board Level Packaging Thermal Management

Board-level packaging thermal management has been having more and more attention paid to it, starting from the early stage of an IC design. Often, when designing a new board for an electronic device, a thermal simulation can initially be done to identify hotspots. Problems identified at this stage can often be addressed by layout changes that can be made nearly without cost at this stage of the process.

In a typical board-level thermal simulation process flow, the systems architect will develop the initial concept design by creating a functional block diagram. The hardware design engineer then derives the first physical layout directly from the block diagram. At an early stage in the design process, long before the mechanical engineer gets involved, the electrical engineer can use board-level simulation to evaluate the new board design in an existing system. For instance, a three-dimensional (3-D) computational fluid dynamics solver can be used to predict air flow and temperature for both sides of the board, as shown in Figure 1.3 (Petrosjanc 2009). Often the designer will identify hotspots, and cooling management can thus be considered from the earliest stages of the design process. Changes made to the

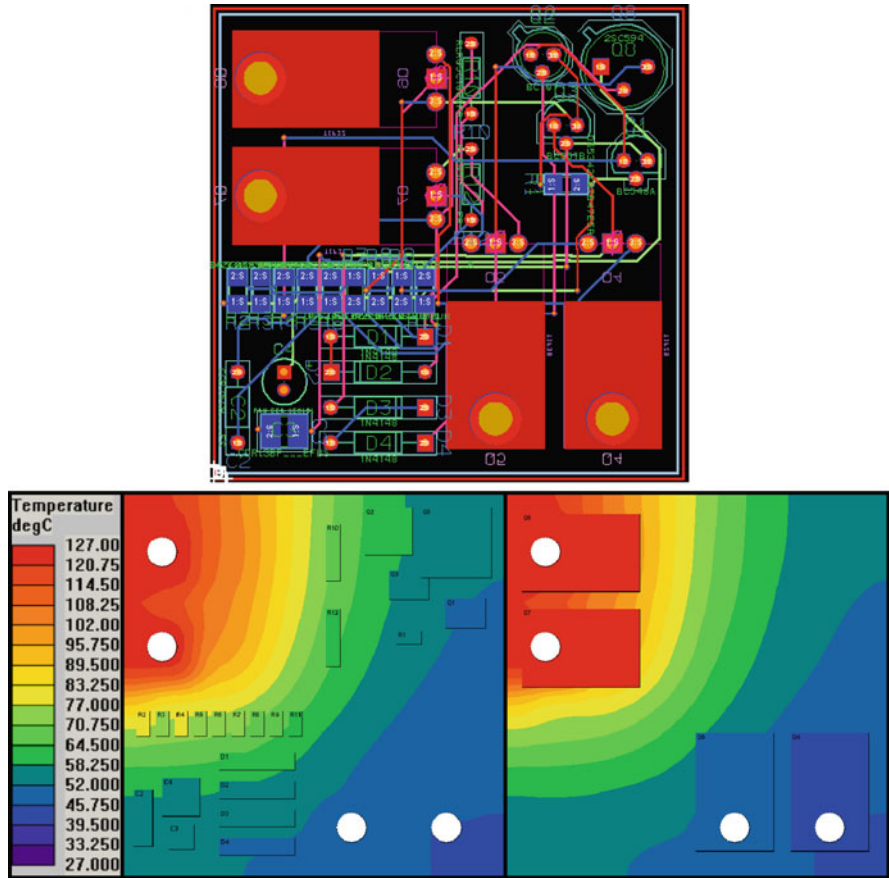


Fig. 1.3 Example of thermal simulation for an audio power amplifier printed circuit board (Petrosjanc 2009)

functional block diagram are instantly reflected in the physical layout and thermal representation. At this stage, many more alternatives exist to deal with thermal problems. Rather than being limited to expensive additions, engineers can consider a wide range of processes such as changing the board layout, adding copper inserts, or changing the package style. The board-level model can also be imported into a system-level thermal model, such as the one that may have been created when the original system was designed. This saves the mechanical engineer time in updating the system-level model, if necessary, while reducing the chance of errors caused by miscommunications. The results from the system-level analysis can also be exported to the board-level simulation, making it possible for the electrical engineer to apply the system-level air flow and temperatures to the board being designed. This approach keeps all team members in sync and enables them to contribute to concept development in real time (Bornoff 2005).

Based on the board-level thermal simulation and thermal design, a board-level packaging system with effective thermal management can be built up for the integration of power semiconductors, sensors, and control electronics. The commonly used techniques include monolithic integration, hybrid integration on ceramic substrates, DCB/DAB (direct copper/aluminum bonded ceramic substrates) technique, lead frame technique (e.g., in combination with ceramic substrates and molded packages), IMS (insulated metal substrate, metal core boards), and PCB. In the low-voltage, low-power range the monolithic integration is well established, e.g., in form of the widespread Smart-Power switches. However, this approach quickly comes to its economic limits with increasing operating voltage and currents because of the large difference in complexity and costs per chip area of modern semiconductor processes for power and logic devices. Considering the mounting techniques for discrete components, the hybrid and DCB technologies can realize the best thermal performance with substrate materials such as Al_2O_3 or AlN ceramic, but also show the highest costs per substrate area. Common to lead frame, DCB and IMS technology is generally the single-layer copper structure which makes it very difficult to route more complex circuits and greatly limits the package density. The comparatively coarse trace widths in DCB and partially in the lead frame technology also limit the integration of more complex signal electronics. On the other hand the thick copper traces (> 0.3 mm), which can be realized in DCB technology easily, allow control of high currents in the range of several hundred amperes. Comparably, the printed circuit board technique is based on the worst heat-conducting substrate material, nevertheless, has the charm to open up the great potential of a modern, very innovative, and very high production volume technology, which by far outperforms the competitive techniques with respect to integration density, flexibility, and costs per substrate area (Moran 2001). The base material of a PCB is typically an epoxy laminated cotton paper or woven fiberglass. The active and passive devices are usually directly soldered on the PCB with copper interconnects. The effective thermal management on a PCB level is typically achieved or enhanced through (1) replacing nonavailable on-board copper area with cooling pins; (2) contacting board to a heat spreader or a heat sink or both; (3) using thermal vias (through-hole, blind or buried) where possible; (4) using maximum copper filing in all available layers acted as the heat spreader; (5) providing a good thermal coupling to components with high thermal conductivity and large surface; (6) using any available forced or convective air flow; (7) introducing emerging thermal management techniques.

In many power electronic systems the integration density in the signal and control electronics parts has reached a level that requires multilayer boards. The integration of the power electronics parts on the same board does not only cause the cancellation of interconnection elements, but also allows new dimensions of multi-functional board integration. This results in many benefits of a multilayer PCB integration of power electronics, such as an SMT-compatible assembly process, integrable heat path with electrical insulation from heat sink, integrable EMI shields and low-noise ground planes, high package density due to multilayer routing, integrable windings for planar magnetics, and cancellation of interconnects

between power and signal electronics. Therefore, advanced board-level packaging can be obtained with optimal cooling of all active and passive power components; very low dead volumes, high power density, very low circuit parasitics, good EMI behavior, and high mechanical stability and functional reliability (März 2003).

However, besides all thermal optimizations on a board level, in many cases money spent on better components or more intelligent circuit topologies is paid back by less expensive heat sinks, improved reliability, and—in the long term—less energy costs (März 2003).

System-Level Packaging Thermal Management

System-level thermal management has been widely used with various techniques which can be conveniently grouped into passive and active methods. Passive methods, which do not require any input power, tend to be very reliable and relatively easy to implement. However, they are also performance limited and therefore inadequate for many high power applications. Typical passive methods include (Moran 2001): (1) natural convection (finned heat sinks, ventilation slots, board/component placement, etc.); (2) radiation (paints, coatings, mechanical surface treatments, component layout, etc.); (3) conduction (heat spreaders, thermal conduction structures, vias, elastomers, pastes, adhesives, pads, chip and board packages, etc.); and (4) emerging thermal management technologies. Strictly speaking, for example, capillary and phase change methods (e.g., heat pipes, wicks, melting wax, boiling, etc.) are also passive methods. However, they generally provide higher performance capabilities and are not as easily implemented as the other passive methods. In addition, phase change devices are limited in terms of their operating temperatures, which are determined by the saturation temperature of the working fluid (or melting temperature of the solid).

Contrastingly, active thermal management techniques require input power to provide increased performance/capacity, but generally at the price of lower reliability and added complexity. Active methods usually include external forced convection (fans, nozzles, etc.), pumped loops (heat exchangers, cold plates, etc.), and refrigerators and coolers (thermoelectric/Peltier, vapor-compression, vortex, gas cycles, etc.). It should be noted that heating (e.g., resistance, induction, etc.) is also an active thermal management tool for applications where a lower temperature boundary must be avoided. Aerospace applications requiring heating are common due to the effective thermal radiation sink temperature of space being a few degrees above absolute zero. On earth, cold region applications may also require heating. In addition, measurement electronics or other systems where a fixed temperature band must be maintained often require an integrated cooling and heating system with temperature feedback control (Moran 2001).

There are quite a few emerging technologies that are under development and evaluation. One area is to enhance convection cooling by improving the heat transfer coefficient through extended heat sink surface and high air flow fans with

a built-in feature for acoustic noise cancellation. Attention has also been paid to the development of heat spreader materials and components, such as carbon fiber, graphite, thermally conductive composites, vapor chambers, heat pipes, and nanomaterials. The integrated solutions also being explored, for instance, are a close-loop liquid cooling system can implement cold plates or microchannels with either single-phase or two-phase liquid cooling. Furthermore, refrigeration to achieve negative thermal resistance has been developed with the focus on reducing the size and cost of the compressor and the heat exchanger. Solid-state refrigeration or thermoelectric cooling is another example for hotspot cooling of devices with highly nonuniform power dissipation or full-chip cooling in conjunction with vapor chamber heat sinks. In addition, emerging nanomaterials hold promise of providing highly conductive thermal interface materials and reducing interconnect Joule heating (Mallik et al. 2005).

Another research area is the system-level dynamic thermal management (DTM), which uses both hardware and software support in a synergistic fashion and hence leads to a significant execution time overhead. For example, a system-level framework for doing fine-grained coordinated thermal management uses a hybrid of hardware techniques (like clock gating) and software techniques (like thermal-aware process scheduling), leveraging the advantages of both approaches in a synergistic fashion. While hardware techniques can be used reactively to manage the overall temperature in case of thermal emergencies, proactive use of software techniques can build on top of it to balance the overall thermal profile with minimal overhead using the operating system support (Kumar et al. 2008). This would build up a new routine for future system-level thermal management of electronic packaging.

Thermal Management Solutions

Effective thermal management of an electronic system requires identifying critical issues such as shock, vibration, and usage scenarios, considering potential coupling effects, and developing an integrated, interdisciplinary solution (Madrid 2000). This usually can be achieved based on understanding the thermal behavior of a system and optimizing that behavior for cost efficiency through modeling and analyzing the system, including for instance (1) system-level cooling and active control; (2) board-level and chip-level thermal design and thermal management; (3) device or die electrothermal modeling and optimal thermal design; and (4) micro- or nanoscale thermal design and engineering or processing. As all different component reliability is inversely proportional to temperature gradients and sustained exposure at the elevated temperatures, power dissipation is a key function of any thermal solution. General thermal management solutions, therefore, typically include hardware assembly, software control, and optimal thermal design as well as the combination of these approaches. System hardware solutions are based on internally distributing and externally dissipating the thermal energy to the ambient environment. System software solutions typically regulate the dissipative power of

the device based on active system feedback controls. Optimal thermal design is required for all level cooling and thermal control, particularly effective for device-level micro- or nanoscale thermal engineering and processing. A well-designed hardware thermal solution combined with software thermal management preserves a system’s functionality and extends the life cycle of an electronic device (Madrid 2000).

Hardware Solutions

Various hardware thermal management devices are able to dissipate a range of heat fluxes over a range of temperature differences. As the package level goes from the chip die to its packaging, to a board, to a motherboard, to a system, the area over which the heat is distributed increases so the heat flux decreases correspondingly. This allows heat removal at one or more packaging levels, where the heat flux and temperature differences are consistent with the available cooling method (Couvillion 2006). As shown in Figure 1.4, such hardware typically consists of the use of natural convection, forced convection, fluid-phase change, thermionic and liquid cooling devices, as well as interface materials, mounting assembly, allocated system, and component and motherboard placement.

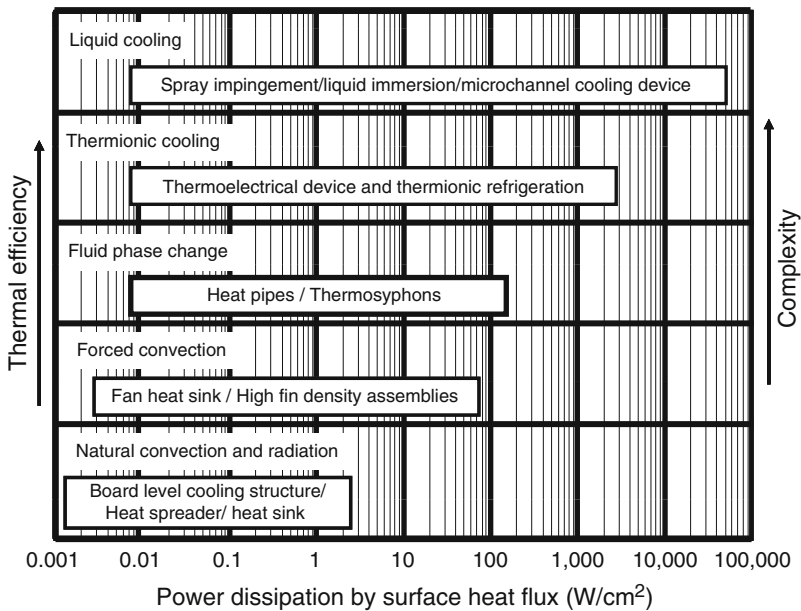


Fig. 1.4 Achievable power dissipation by heat fluxes with various thermal management solutions

Typical natural convection heat sinks are passive in nature and manufactured from copper or aluminum sheet, extruded aluminum, machined, or cast alloys. The heat sink cools a device by expanding the surface area of the part to which it is attached, increasing the amount of heat that can be dissipated by the ambient air.

Forced convection requires forced air velocity generated through the incorporation of either a dedicated or system-level fan in order to increase thermal efficiency. Fan heat sinks, high fin density assemblies, as well as board-level coolers are manufactured and configured for either impingement or cross-flow environments. Fan solutions include passive-active and active approaches. With a passive-active approach, fan heat sink solutions provide air flow and require little or no system air flow, while active fan heat sink solutions incorporate a fan that is attached to the solution. The air flow that a fan produces blows parallel to the fan's blade axis (Madrid 2000).

Fluid-phase change is a recirculating process, typically employs closed loop heat pipes which allow the rapid exchange of heat transfer through evaporation and condensation. Heat pipes are integrated into other heat sink technologies to further increase the thermal efficiency when greater density is required or physical size restrictions exist. Considerations in choosing heat pipes include: single-slot height, power consumption, noise elimination/reduction, low maintenance, sealed enclosure cooling, and extended ambient temperatures.

Thermionic cooling solutions, such as thermoelectrical and thermionic refrigeration, utilize solid-state conversion of low thermal conductivity but good electrical conductivity materials to achieve refrigeration. The electrons in the n-type material and the holes in the p-type material all carry heat away from the top metal-semiconductor junction, which leads to cooling of the junction, by means of a process called the Peltier effect. Conversely if a temperature difference is maintained between the two ends, higher thermal energy electrons and holes will diffuse to the cold side, creating a potential difference that can be used to power an external load. Due to the high dependence of the operation of thermionic cooling on the electronic behavior and properties of the semiconductor materials, it is highly influenced by the presence of magnetic fields (Chen and Shakouri 2002).

A great many liquid cooling solutions have been developed. One of them comprises channeled cold plates along with a heat exchanger and pump system in order to circulate fluids past a heat source. Generally, liquid cooled technologies are reserved for applications containing high heat flux density where forced convection or phase change systems are unable to dissipate the power demands.

Table 1.1 summarizes some typical hardware cooling solutions and their benefits and drawbacks. All of these thermal management solutions must be designed along with the electrical system. However, the thermal resistance between contact interfaces is usually high, and electrical insulating materials used in the electrical system are usually poor thermal conductors. Furthermore, the mismatches in the CTE of materials bonded together create stresses as a package warms and cools when electricity is turned on and off. These conflicts make the choice and design of thermal management methods quite a challenge (Couvillion 2006).

Table 1.1 Typical conventional and emerging hardware cooling solutions

Cooling types	Hardware device	Character description	Advantages	Disadvantages
Air cooling	Heat sink	The simplest of all the technologies and use natural or forced convections as the cooling method. Heat dissipation is directly proportional to available surface area	Extremely simple option; requires no pumping mechanisms; minimal maintenance; systems are inexpensive	Very low heat removal capacity; high thermal resistance; requires cleaning and dehumidification environments; added size and weight; acoustic noise due to fans
	Thermotunneling	Quantum electron tunneling provides cooling rates of about 3–5 W/cm ² , but has potential to achieve around 100 W/cm ² . Electrode work function needed at around 1–1.5 eV to improve tunneling effect	Can be packaged in arrays to cool virtually any size heat load; operates at 70–80% of the maximum theoretical efficiency for cooling; low cost	Very low heat removal capacity; tunneling achieved less than 10% of the total area; tension is introduced during the electroplating process; high work function of electrodes is needed
	Ionization	Ionization sticks to air cooling while providing cooling capacity comparable to liquid cooling at 40 W/cm ² . Management of nanosand, charge insulation and leftover voltage during use is needed	No moving parts, high reliability; best heat flux removal (~40 W/cm ²) achieved in air cooling techniques; small size, possible usage as a layer onto the chip	Possible problems of insulation due to the charge; leftover voltage will be converted to heat which hampers the cooling; possible problems of nanosand (dust, oils) that can inhibit heat-flow
Refrigeration	Vapor compressor	Vapor compressor is more efficient as compared to simple air cooling. The miniaturization of the vapor compression system is the critical success factor	Can generate cooling temperature below the ambient temperature; more efficient as compared to air cooling and cold plates	Difficulty in miniaturizing the system; additional components condenser, compressor and evaporator makes the system bulky
	Stirling cooler	Higher reliability is possible by using diaphragms instead of pistons and displacers. Scalability to micron level at acceptable cost is a key success factor	Higher efficiencies are possible, as compared with vapor compressor; low weight; possible size reduction to 1 cm	Diaphragms or pistons are moving parts, thus reducing reliability; cost is high

(continued)

Table 1.1 (continued)

Cooling types	Hardware device	Character description	Advantages	Disadvantages
Liquid Cooling	Thermoelectric cooler	Thermoelectric Coolers are based on the Peltier Effect and can be scaled to micron level for on-chip cooling. Fabrication at the micron level is a key success factor	Solid state devices with no moving parts; can be made very small; can generate cooling temperature below the ambient temperature	Low efficiencies resulting in high power requirements and added waste heat; low cooling power density as compared to liquid cooling; high cost especially as size decreases
	Hetero-structure cooler	Two stage thermionic/thermoelectric micro-refrigerators may provide large cooling power densities (600 W/cm^2). Cost is a key success factor for commercial applications	Very small size and standard thin film fabrication method; possible to cool hotspots directly; high cooling power density and better transient response time as compared to thermoelectric cooler	Cost is high with packaging issues; high thermal conductivity of materials used; high contact resistance between metal and cap layer
	Superlattice structures	Improved performance over thermoelectric cooler in terms of speed, cooling power density and size	Extremely small size and thin film to provide site-specific microcooling of hotspots; provide best in class thermoelectric materials figure of merit	Lead (Pb) component in the thermoelectric superlattice
Liquid Cooling	Cold plates	Inexpensive as compared to other liquid based technologies and better performance than heat sinks. Power dissipation to weight ratio is a key success factor	Compared to heat sinks, water cold plates have very low thermal resistance and high thermal conductivity; can be effectively used with microchannels and heat pipes to increase the cooling capacity; inexpensive	Low heat removal capacity as compared to other liquid cooling technology; added size and weight; insulation to prevent leakage
	Jet impingement	Higher heat flux in two phase mode, although with increased complexity. Optimized device geometry and flow rates handle high pressure drops is critical	Very high heat fluxes can be achieved; critically designed nozzles not required; flow can be localized if only a single jet is used; can operate in single phase or two phase mode	High pressure drops, which may require larger pumps and optimized device geometry and flow rates; maximum chip heat flux removal is limited by the critical heat flux (2 phase); vapor droplets adherence to chips may hamper cooling

Spray cooling	Provides customized (temperature specific) and localized cooling. Geometry of nozzles and flow rates is a key success factor	Can provide higher heat flux as compared to jet impingement; gives good coverage due to atomizing action; extreme variations in flow rate magnitudes are possible; spray pattern can be flexible allowing for tracking of power density variations with time; localized cooling possible	Requires special spray nozzles; nozzles are sensitive to manufacturing tolerances and quality; nozzle action can change in time due to erosion, corrosion build up, and contaminants; spray velocity and momentum are critical
Two phase cooling device	Two phase liquid cooling provides high flux removal capacity and can cool hotspots directly	High cooling capacity and high heat transfer coefficients; can provide localized on spot cooling and easily be incorporated with microchannels; reduced sensitivity to gravity and acceleration as compared with spray cooling	The system tend to more complex with fabrication and miniaturization of no-moving-part pumps; extra insulation and sealing required; use of dielectrics or liquid metal alloys may lead to compatibility problems and environmental problems

Software Solutions and Software-Based Dynamic Thermal Management

Apart from the hardware cooling solutions, temperature control on a board was also achieved traditionally by using the system BIOS (basic input/output system) and passive cooling functionality. Passive cooling reduces the speed or disables some on-board devices in order to decrease power consumption, thereby reducing overall system temperature. Active cooling as discussed earlier, on the other hand, increases power consumption by activating fans for instance, that increase air flow, thereby reducing the system temperature.

In a typical software-based approach, the system will display central processing unit (CPU) die temperature in real time, allowing the user to read the normal temperature of the system and set accurate values for resume and overheat. With an advanced configuration and power interface (ACPI) solution, temperature control is moved from the BIOS to the operating system. In keeping with more sophisticated temperature control features, the inclusion of the ACPI in operating systems lets the user make more intelligent decisions with a better follow up on the CPU load and applications control. The ACPI thermal design is based on regions called thermal zones. Some systems may have more than one thermal sensor to allow subdividing the system in many more thermal zones. A benefit of ACPI is that it standardizes thermal control methods. In contrast, BIOS control methods are proprietary and do not let applications use them transparently. ACPI control also provides the choice of passive cooling, active cooling, or a mix of both (Madrid 2000).

In fact, both hardware- and software-controlled solutions usually exist for power and thermal management of processor-based systems. Software-based solutions are primarily utilized in connection with mobile platforms. The software-controlled techniques involve an interrupt generated when a processor temperature setting is exceeded. The processor may be throttled after detecting an above temperature condition by polling processor temperature. Generally, the software-controlled solutions have a slower response time than the hardware-controlled solutions. In addition, it tends to overshoot and undershoot problems with software-controlled solutions. The sensors utilized in software-controlled solutions are relatively slow and inaccurate. The on-die sensor (which is normally a diode) is not located on the hottest part of the processor die. The software-controlled solution is based on the premise that the platform exposes a variety of trip points to the operating system. A trip point is a temperature for a particular thermal region when some action should be taken. As the temperature goes above or below any trip point, the platform is responsible for notifying the operating system of this event and the operating system then takes an appropriate action. When a temperature crosses a passive trip point, the operating system is responsible for implementing an algorithm to reduce the processor's temperature. It may do so by generating a periodic event at a variable frequency. The operating system then monitors the current temperature as well as the last temperature and applies an algorithm to make performance changes

in order to keep the processor at the target temperature. Software-controlled throttling is exposed to the operating system, allowing the operating system to know the processor performance at all times. This becomes especially important with future operating systems that guarantee some quality of service based upon the processor performance to the executing applications. This concept is known as guaranteed bandwidth allocation and is based on the processor's current performance level (Cooper 2006).

Another typical software-based solution is DTM. This refers to a range of possible hardware and software strategies which work dynamically, at run-time, to control a chip's operating temperature. In contrast, the packaging and fans for a CPU or computer system were traditionally designed to be able to maintain a safe operating temperature even when the chip was dissipating the maximum power possible for a sustained period of time, and therefore generating the highest amount of thermal energy. This worst-case thermal scenario is highly unlikely, however, and thus such worst-case packaging is often expensive overkill. DTM allows packaging engineers to design systems for a target sustained thermal value that is much closer to average-case for real benchmarks. If a particular workload operates above this point for sustained periods, a DTM response will work to reduce chip temperature. In essence, DTM allows designers to focus on average, rather than worst-case, thermal conditions in their designs. Until now, techniques developed to reduce average CPU power have garnered only moderate interest among the designers of high-end CPUs because thermal considerations, rather than battery life, were their primary concern. Therefore, in addition to reducing packaging costs, DTM improves the leverage of techniques such as clock gating designed to reduce average power (Brooks and Martonosi 2001).

The key goal of DTM is to provide inexpensive hardware or software responses that reliably reduce power, while impacting performance as little as possible. Implementing an effective DTM system involves (1) Selecting simple and effective triggers; (2) Identifying useful response mechanisms; and (3) Developing policies for when to turn responses on and off. For trigger selection, DTM allows arbitrary tradeoffs between performance and savings in cooling hardware. Conservative target selections can still lead to significant cost improvements with essentially zero performance impact, because the trigger point is rarely reached for many applications. In addition, DTM makes other techniques targeting average power more interesting to the designers of high-end CPUs. Effective DTM makes average power the metric of interest even for high-end CPU designers, since packages need no longer be designed for worst-case power. With DTM, lowering average CPU power will reduce the trigger value needed for a particular level of performance, and thus will reduce packaging costs. Not unexpectedly, the triggering delay is a key factor in the performance overhead of DTM. More lightweight, fine-grained policies, such as the microarchitectural techniques often allow the temperature to stay close to the target level with a small performance penalty. Furthermore, the fine-grained policies are less affected by rapid fluctuations in the temperature. Because of these growing opportunities for microarchitectural DTM techniques, a methodology for evaluating new DTM

approaches has been explored. This mechanism correlates power and performance, and looks for “low-hanging fruit;” that is, techniques that can cut power by significantly more than they hurt performance are needed. Identifying these sorts of wasted work, particularly on an application-specific basis, appears to be a promising way of discovering new microarchitectural DTM techniques in the future (Brooks and Martonosi 2001).

In fact, several DTM techniques have been developed. One proposed method is transient thermal management, which involves using heat storage devices to store the heat produced by the processor during power intensive computations and dissipate it gradually over time. This method is combined with a dynamic thermal management technique to lower the power consumption (Cao et al. 1996; Wirth 2004). A system has been developed by monitoring the temperature of the processor and as it reached the set maximum it stalled any power intensive instructions. This system allows the processor to still run normally for low power instruction, which benefited the user over a system which slowed the processor down for all applications (Wirth 2004).

Another DTM technique is to design a system that slows down the entire processor (Brooks and Martonosi 2001). While most of the DTM systems that have been simulated measure the power and temperature and then cause the system to change when the processor is consuming too much power or the temperature is too high, a predictive voltage scaling method uses an algorithm that predicts the power to be used and then sets the voltage and frequency accordingly. This method may have better response time than reactive systems (Srinivasan and Adve 2003). A different idea involving voltage and frequency scaling has been developed with the design that involves separating the chip into different sections in which the voltage and frequency can be scaled independently of the other sections. This design would minimize the performance loss on those tasks that are not power intensive. A similar method uses a performance manager to analyze the power requirements of the next task and then set the voltage and frequency based on values given in a look up table. A performance manager is especially viable in embedded processors because the workload is typically limited to a small range of instructions for which the entire thermal management system can be optimized. In addition, chip architectures have been developed, which have several built-in features to run at lower power during times when the device is not being used and allow for both dynamic voltage and frequency scaling (Clark et al. 2002). Taking the thermal considerations as a part of the design process rather than an afterthought, combining DTM systems to create even more efficient designs for future processors would be a future direction. Still, many of these DTM techniques need to be tested on actual devices to measure their true performance and feasibility. Chip designers will work on developing lower power chips and hardware that has DTM techniques built into the architecture. Meanwhile, better ways of judging performance by developing good, workable thermal requirements have been explored. Whatever direction the future of the industry leads, the problem of heat dissipation has no magic solution. As the processor density continues to increase, so does the need for better methods of heat dissipation (Wirth 2004).

Optimal Thermal Design of a Package

The optimal thermal design is a fundamental approach for future micro- and nanoscale electrothermal modeling, engineering, and processing as well as present packaging thermal management, based on the minimization of thermal resistance. The thermal resistance at device level is associated with conduction of heat from the die to its substrate to the surface of the package. There is often significant thermal resistance in the attachment of the die to the substrate, such as contact or interface resistance. From the chip-level package surface, heat can be removed by a coolant and/or conducted to a board. At the board level, heat is removed by the cooling system or moved to a higher packaging level (Couvillion 2006). The overall thermal resistance of a package is almost entirely determined by (NEC Electronics 2003) package structure, package size, chip dimensions, airflow rate, as well as hardware cooling assemblies.

Different package structures have different thermal resistance characteristics. Packages such as ABGAs (advanced ball grid arrays) and FCBGAs (flip chip ball grid arrays), which feature a copper lid to which the chip is directly attached with thermally conductive paste, offer excellent thermal resistance characteristics. In the case of PBGAs (plastic ball grid arrays), thermal resistance can be lowered by using a four-layer substrate instead of a two-layer substrate, and it can be further lowered by placing solder balls directly underneath the thermal via holes.

In general, the larger the size of the package, the lower its thermal resistance is. This is particularly true for ABGAs and TBGAs (tape ball grid arrays), which have a copper lid offering excellent thermal conductivity characteristics. In the case of packages with lower thermal conductivity such as FPBGAs (flip chip plastic ball grid arrays), there is a weaker correlation between thermal resistance and package size, and there is little variation in thermal resistance among different size packages if the chip dimensions are the same.

Large chip size benefits to the low thermal resistance. The thermal conductivity of silicon (the material chips are made of) is about 100 times higher than that of mold resin and about 10 times higher than that of package substrates; therefore, the surface area of the chip itself greatly contributes to heat dissipation.

Airflow rate and hardware cooling assembly are effective to reduce the thermal resistance. The airflow rate is not directly related to the thermal conductivity of the package itself, but forced air cooling, such as through the use of a fan, efficiently transfers heat from the package surface or printed wiring board to the surrounding atmosphere and thus reduces thermal resistance.

There have been many approaches for optimal thermal design to reduce the thermal resistance and maximize the intrinsic performance and lifespan of semiconductor devices, helping meet market needs for higher performance as well as saving costs by curtailing the need for heat dissipation measures. For example, the thermal resistance of QFPs (quad flat packages) can be reduced through appropriate material selection and lead frame design (NEC Electronics 2003): (1) Employment

of thermally enhanced resin. Because mold resin is the lowest thermally conductive material among package materials, thermal enhancement of the mold resin can significantly lower the thermal resistance of packages. One way to do this is to replace the existing filler with high thermal conductivity filler. Furthermore, the thinner the resin on the chip, the shorter the distance over which heat is transmitted, thus, mounting a heat sink on the surface of the package enhances the heat dissipation effect to a greater extent in the case of tape-bonded quad flat packages, which have a thinner resin thickness than QFPs. (2) Lowering thermal resistance through lead frame design. Thermal resistance can be lowered by changing the material of the lead frame from alloy 42 (Fe–Ni) to a copper alloy; designing the die pad on which the chip is to be mounted as large as possible; and attaching a heat spreader or other cooling system to the die pad so that heat is transmitted almost throughout the entire package.

Similarly, the thermal resistance of BGAs (ball grid arrays) can be lowered by optimal material selection, package structure placement, and substrate design. As in the case of QFPs, the thermal resistance of BGAs can be decreased by using material with a higher thermal conductivity for the filler. For package structure and substrate design, the following measures may be taken (NEC Electronics 2003): (1) Use of thermal balls as a low-cost solution. In terms of package structure, a heat dissipation path is secured from the rear side of the chip to the solder balls immediately beneath the chip by providing a large number of solder balls on the rear side of the chip, and thermally connecting these solder balls to the die pad via through-holes (thermal vias). The balls at the center of a package are electrically grounded and commonly called “thermal balls” as they play a thermal dissipation role by conducting heat to the printed wiring board. This is the cheapest way to dissipate heat. The thermal balls also serve as ground pins and neighboring balls can be assigned as signal pins, meaning that the actual number of pins can be increased. (2) Use of two inner layers of package substrate as ground layers. Generally, a printed wiring board having four or more layers, including power and ground layers, is used as the package substrate for BGAs to ensure satisfactory electrical characteristics. However, the use of a four-layer structure to lower the thermal resistance, rather than improve electrical characteristics, has been increasing. In this case, heat from the chip is transmitted to the inner layers via the die pad’s through-holes, which also serve as grounds, and two out of the four inner layers are used as grounds to secure a heat dissipation path. To further decrease thermal resistance, a substrate with a thick embedded metal core layer has been used. (3) Use of an embedded heat spreader. If the combination of thermal balls and a four-layer package substrate still fails to satisfy the thermal resistance requirement, a heat spreader can be embedded in the package. Such a heat spreader serves to diffuse the heat transmitted through the mold resin to the surface of the package. However, the reduction in thermal resistance that can be achieved this way is limited because the heat spreader is not in direct contact with the die pad and the chip. (4) Modification of printed wiring board design. The thermal resistance of a BGA can also be lowered by modifying the design of the printed wiring board. The thermal resistance changes according to the number of thermal holes, the number of layers of the package substrate, and the presence or absence of a heat

dissipation path. Reassessment of the thermal design of the entire package, including the printed wiring board, can result in a low-cost package that meets thermal resistance requirements. (5) Selection of cavity-down type PBGA. Cavity-down type BGAs in which the chip is flipped and attached to the heat spreader that is exposed at the surface, is the most effective solution for lowering thermal resistance. This also holds true for QFPs. Packages of this type include TBGAs, ABGAs, and FCBGAs. Heat is directly conducted from the chip to the copper plate on the package surface, thereby achieving low thermal resistance.

In addition, measuring the temperature of the chip after it has been installed in a system is very important and can be used to verify or optimize the thermal design. Such measurement data are also useful for estimating the power consumption of devices to be developed in the future. Furthermore, if the power consumption of the system is known to be considerably above the value estimated at the design phase, it is very important to know the junction temperature of the device under actual use in the system.

Fundamentals of Heat Transfer and Thermal Calculation in Electronic Packaging

The objective of thermal management in electronic packaging is to efficiently remove heat from the semiconductor junction to the ambient environment. This process includes: (1) heat transfer within the semiconductor component package; (2) heat transfer from the package to a heat dissipater, such as heat spreader or the initial heat sink; (3) heat transfer from the heat dissipater to the ambient environment through the ultimate heat sink, or other cooling systems. Achieving an efficient thermal connection through these paths requires a thorough understanding of heat transfer fundamentals as well as knowledge of available interface materials and how their key physical properties affect the heat transfer process (Chomerics 1999).

Heat transfer is the movement of heat flow as a result of a temperature difference. Temperature represents the amount of thermal energy available, whereas heat flow represents the movement of thermal energy from place to place. On a microscopic scale, thermal energy is related to the kinetic energy of molecules. The higher a material's temperature, the greater the thermal agitation of its constituent molecules is, which is manifested both in linear motion and vibrational modes. It is a natural process for regions containing greater molecular kinetic energy to pass this energy to regions with less kinetic energy. Several material properties serve to modulate the heat transferred between two regions at differing temperatures, such as thermal conductivities, specific heats, material densities, fluid velocities, fluid viscosities, surface emissivities, and more. Taken together, these properties serve to make the solution of many heat transfer problems an involved process.

The mechanisms of heat transfer depend on the media involved and usually divided into conduction, convection, radiation, and multimode, which is a combination of one

or more of the above. When a temperature gradient exists within a continuous, nonmoving medium, solid or stationary fluid, heat is transferred through the medium via the conduction mode. Convection heat transfer occurs when a surface is in contact with a moving fluid, liquid, or gas at a different temperature. Heat is exchanged between the fluid and the surface. Radiation heat transfer occurs when two surfaces at different temperatures exchange energy in the form of electromagnetic energy emitted by the surfaces. Radiation can occur in a vacuum because no medium between the two surfaces is required as in conduction and convection (Couvillion 2006).

Conduction

Thermal conduction is a process in which heat flows through a solid, liquid, or gas or between two media that are in intimate contact. Conduction is the dominant mechanism for heat transfer within solids, involving the transfer of kinetic thermal energy from one electron to another to cause no visible motion of the particles of the body. Conduction through dielectric solids is almost entirely due to lattice vibrations, while conduction through metallic solids has added energy transport by free electrons. Liquids also conduct thermal energy, but to a significantly lesser extent than solids. When a material changes phase from a solid to a liquid, there is a lessening of its intermolecular bonds and a deterioration of the ordered state of the solid, therefore, more freedom for thermal motion of the molecules makes its thermal conductivity lower. When a liquid changes to a gas, there is a further loosening of molecular bonds that allows random motion of the gas molecules with only restraints of random collisions. As a result, the thermal conductivity of gases is quite low (Krum 2004). Based on Fourier's law of heat conduction, the rate at which heat is conducted through a material is proportional to the area normal to the heat flow, the temperature gradient along the heat flow path, and the thermal conductivity of the material. For a one dimensional, steady state heat flow, as shown in Figure 1.5a, the rate is expressed by Fourier's equation:

$$q = kA \frac{\Delta T}{L}, \quad (1.4)$$

or

$$q'' = \frac{q}{A} = -k \frac{dT}{dx}, \quad (1.5)$$

where q is rate of heat transfer (W); q'' is heat flux (W/m^2); k is thermal conductivity ($\text{W}/\text{m K}$); A is cross sectional transfer area (m^2); ΔT is temperature

difference ($^{\circ}\text{C}$); L is conduction path length. Thermal conductivity, k , is an intrinsic property of a homogeneous material which characterizes the material’s ability to conduct heat. This property is independent of material size, shape, or orientation. For nonhomogeneous materials, however, for instance those having glass mesh or polymer film reinforcement, the term of relative thermal conductivity is usually used because the thermal conductivity of these materials depends on the relative thickness of the layers and their orientation with respect to heat flow (Chomerics 1999).

Another inherent thermal property of a material is its thermal resistance, R , which can be derived from (1.4) and (1.5), and is defined as:

For a slab,

$$R = \frac{\Delta T}{q} = \frac{L}{kA}. \tag{1.6}$$

For a cylindrical shell:

$$R = \frac{1}{2\pi kH} \ln\left(\frac{r_2}{r_1}\right), \tag{1.7}$$

where H is the length of the cylinder; r_1 is the inside radius of the cylinder shell; and r_2 is the outside radius of the cylinder shell.

For a spherical shell:

$$R = \frac{1}{4\pi k} \ln\left(\frac{1}{r_1} - \frac{1}{r_2}\right), \tag{1.8}$$

where r_1 is the inside radius of the spherical shell; and r_2 is the outside radius of the spherical shell.

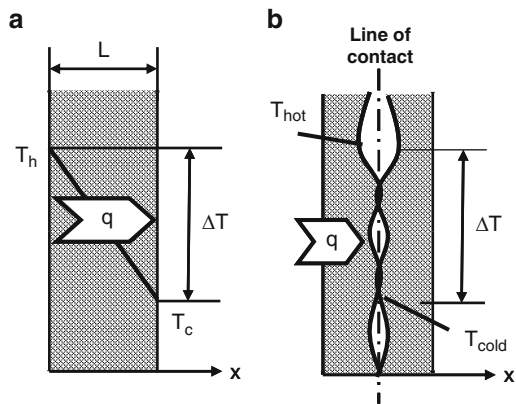


Fig. 1.5 Schematic illustration of (a) one-directional conduction; (b) and contact interface

For a conical frustum:

$$R = \frac{L}{\pi k r_1 r_2}, \quad (1.9)$$

where L is the length of the conical frustum; r_1 and r_2 is the radius of the top and bottom of the conical frustum, respectively.

The thermal resistance is a measure of how a material of a specific thickness resists the flow of heat. For homogeneous materials, thermal resistance is directly proportional to the thickness. For nonhomogeneous materials, the resistance generally increases with thickness but the relationship may not be linear. Thermal conductivity and thermal resistance describe heat transfer within a material once heat has entered the material. When different surfaces contact each other, the contact interface between two surfaces can also produce a resistance to the flow of heat, as shown in Figure 1.5b. Actual contact occurs at the high points, leaving air-filled voids where the valleys align. Air voids resist the flow of heat and force more of the heat to flow through the contact points. This constriction resistance is referred to as surface contact resistance and can be a factor at all contacting surfaces. The total impedance (θ) of a structure is defined as the sum of its material thermal resistance and any contact resistance between the contacting surfaces. Surface flatness, surface roughness, clamping pressure, material thickness, and compressive modulus have a major impact on contact resistance. Because these surface conditions can vary from application to application, thermal impedance of a structure will also be application dependent (Chomerics 1999).

When several materials are stacked in series, such as a die attached with epoxy to a substrate that is soldered to a package base, the total thermal resistance becomes the sum of the individual thermal resistances. For N thermal resistances in series, the total thermal resistance θ is

$$\theta = \theta_1 + \theta_2 + \theta_3 + \cdots + \theta_N. \quad (1.10)$$

The temperature at a particular interface of an electronic package may be calculated as follows (Krum 2004):

$$T_{j, j-1} = T_{\text{hs}} + q \sum \theta_{j-\text{hs}}, \quad (1.11)$$

where $T_{j, j-1}$ is the temperature at interface of layers j and $j-1$; T_{hs} is the contact surface temperature of the heat sink; $\sum \theta_{j-\text{hs}}$ is the sum of thermal resistances from interface of j and $j-1$ to the heat sink.

When there is more than one heat path, for instance N thermal paths, from the dissipating element to ambient, the total thermal resistance is calculated as

$$\frac{1}{\theta} = \frac{1}{\theta_1} + \frac{1}{\theta_2} + \frac{1}{\theta_3} \cdots + \frac{1}{\theta_N}. \quad (1.12)$$

Convection

Convection is the thermal energy transfer between two surfaces as a consequence of a relative velocity between them. It occurs only in fluids wherein the transfer mechanism is the mixing of the fluids. Although each of the surfaces may be a fluid, the most practical application is a solid surface and the other is fluid (Krum 2004). When heat conducts into a static fluid it leads to a local volumetric expansion. As a result of gravity-induced pressure gradients, the expanded fluid parcel becomes buoyant and displaces, thereby transporting heat by fluid motion, i.e., convection, in addition to conduction. Such heat-induced fluid motion in initially static fluids is known as free convection, as shown in Figure 1.6a. For cases where the fluid is already in motion, heat conducted into the fluid will be transported away chiefly by fluid convection. These cases, known as forced convection as shown in Fig. 1.6b, require a pressure gradient to drive the fluid motion, as opposed to a gravity gradient to induce motion through buoyancy. The heat exchange between solid surface and circulating fluid can be described by Newton's law:

$$q = hA\Delta T, \quad (1.13)$$

or

$$q'' = \frac{q}{A} = h\Delta T, \quad (1.14)$$

where q is heat transfer rate (W); A is surface area (m^2); q'' is heat flux (W/m^2); $\Delta T = T_s - T_\infty$, T_s is surface temperature ($^\circ\text{C}$), and T_∞ is fluid temperature ($^\circ\text{C}$); h is convection heat transfer coefficient ($\text{W}/\text{m}^2 \text{K}$). For electronic packaging, the thermal resistance for convection, R (K/W), can be expressed as

$$R = \frac{\Delta T}{q} = \frac{1}{hA}. \quad (1.15)$$

The convection heat transfer coefficient h mainly depends on nature of the fluid motion, fluid properties, and surface geometry, and is usually obtained experimentally. For example, $h = 1.4 (\Delta T/A^{1/2})^{1/4}$ for the nonlinear natural air convection; $h = 4.0 (V_f/L)^{1/2}$ for linear forced air convection, where V_f is air flow velocity, L is the character length in flow direction.

In the electronic packaging thermal management, natural convection can be used for passive cooling, while forced convection mainly for active cooling. Natural convection is caused entirely by differences in density within the fluids resulting from different temperatures and does not use externally forced air movement. Heat flows by conduction or contact from the solid surface to the fluid particles in intimate contact with the surface, and there is a resulting boundary layer of hot

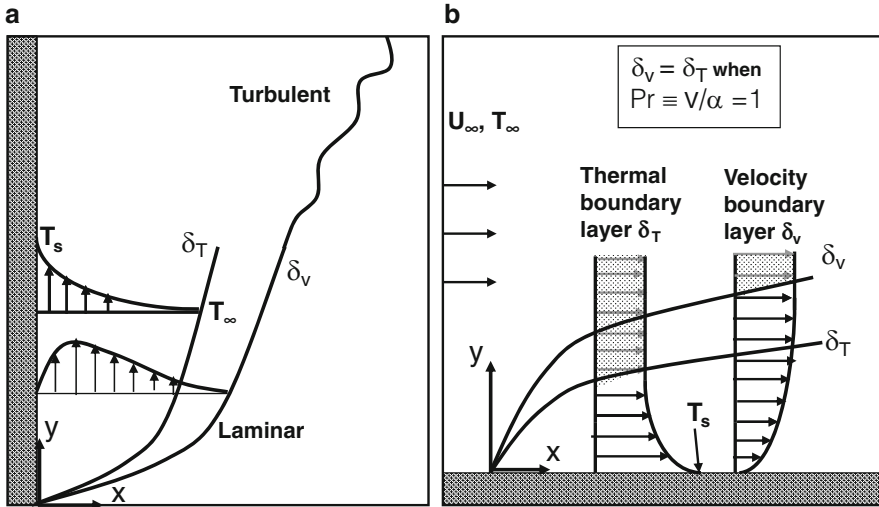


Fig. 1.6 Schematic illustration of (a) natural convection; and (b) forced convection

air immediately adjacent to the surface. In forced convection, thermal energy is transferred from the solid to the adjacent fluid particles in the same manner as in natural convection. However, the subsequent fluid action occurs through artificially induced fluid motion generated by fans, pumps, or blowers. In air cooling, for instance, three types of air moving devices are usually used: centrifugal, propeller, and axial flow. Centrifugal fans are designed to move small volumes of air at high velocities and are capable working against a high resistance. Propeller types are designed to move large volumes of air at low velocities. Axial flow fans are an intermediate type of air mover between the centrifugal and propeller types. In addition, forced convection can be divided into laminar flow and turbulent flow. For air, the transition from laminar to turbulent flow usually occurs at a velocity of 180 linear feet per minute (54.864 m/min). Turbulent flow, characterized by the irregular motion of fluid particles, has eddies in the fluid in which the particles are continuously mixed and rearranged. The heat is transferred from eddies back and forth across the streamlines. Therefore, the greater heat transfer occurs for turbulent flow (Krum 2004). Whatever laminar flow or turbulent flow, the convection cooling method is wholly dependent on the movement of the fluid surrounding the heat dissipating element.

Radiation

All objects with a temperature above 0 K emit thermal radiation. Radiation cooling is the transfer of heat by electromagnetic emission, and is maximized in a complete

vacuum. Radiation from solid objects may be considered to be a totally surface-related phenomenon, and the radiators can be classified as black bodies, gray bodies, and selective radiators. A black body is defined as a surface that absorbs the entire thermal radiation incident upon it, neither reflecting nor transmitting any of the incident radiation. Good absorbing materials are also good emitting materials. The black body, at any given temperature, radiates more energy, both in the total spectrum and for each wavelength interval, than any other temperature radiator and more than any gray body or selective radiator at the same temperature. A gray body is defined as a radiator that has the same spectral emissivity for all wavelengths. A selective radiator is one in which the emissivity varies with wavelength (Krum 2004). Materials used for electronic packaging are usually gray bodies. All materials radiate thermal energy in amounts determined by their temperature, where the energy is carried by photons of light in the infrared and visible portions of the electromagnetic spectrum. When temperatures are uniform, the radiative flux between objects is in equilibrium and no net thermal energy is exchanged. The balance is upset when temperatures are not uniform, and thermal energy is transported from surfaces of higher to surfaces of lower temperature. The transfer of energy by electromagnetic waves can be expressed by Stefan–Boltzmann’s Law, as shown in Figure 1.7, when two surface, gray diffuse enclosure presents with radiatively nonparticipating media,

$$q_1 = -q_2 = \frac{\sigma(T_1^4 - T_2^4)}{\frac{1-\varepsilon_1}{\varepsilon_1 A_1} + \frac{1}{A_1 F_{12}} + \frac{1-\varepsilon_2}{\varepsilon_2 A_2}}. \quad (1.16)$$

For small surface in large surrounds,

$$q_1 = \varepsilon_1 \sigma A (T_{s1}^4 - T_{s2}^4). \quad (1.17)$$

For Infinite parallel plates,

$$q_1 = \frac{\sigma A (T_1^4 - T_2^4)}{\frac{1}{\varepsilon_1} + \frac{1}{\varepsilon_2} - 1}. \quad (1.18)$$

For a three surface gray diffuse enclosure,

$$q_1 + q_2 + q_3 = 0, \quad (1.19)$$

where ε is surface emissivity ($0 < \varepsilon < 1$), for enclosure with black surfaces, $\varepsilon_i = 1$; σ is Stefan–Boltzmann constant, $5.67 \times 10^{-8} \text{ W/m}^2 \text{ K}^4$; T_i is surface temperature (K) of the emission material. The thermal resistance can be expressed as

$$R = \frac{\Delta T}{q} = \frac{1}{h_r A}, \quad (1.20)$$

For electronic devices, $h_r \approx 2\text{--}6 \text{ W/m}^2 \text{ K}$. Radiation cooling is dependent on the temperature difference between objects, their emissivity, and surface area. From a materials standpoint, the thermal design can be optimized with the emissivity parameters.

Space cooling is a particular example of radiation cooling application. In the vacuum of space, radiation plays a major role in the cooling of the electronic systems. Internal to the spacecraft, conduction plays a major role where material interfaces have increased thermal resistance due to the lack of air or any other gas to conduct the heat. In addition, the thermal resistance in high vacuum is further increased by the reduction in joint thermal conductance. For instance, the interface of dry aluminum on aluminum in a vacuum has a thermal conductance approximately half of that seen in air. This degradation varies with each application (Krum 2004).

Multimode Heat Transfer in Electronic Packaging

A major property of packages is how they dissipate the heat generated when a current flows through a resistor in an electric circuit. An electronic system may emit smoke or catch fire if its device generates more heat than anticipated. Excessive heat may also degrade the performance of the device by lowering its operating

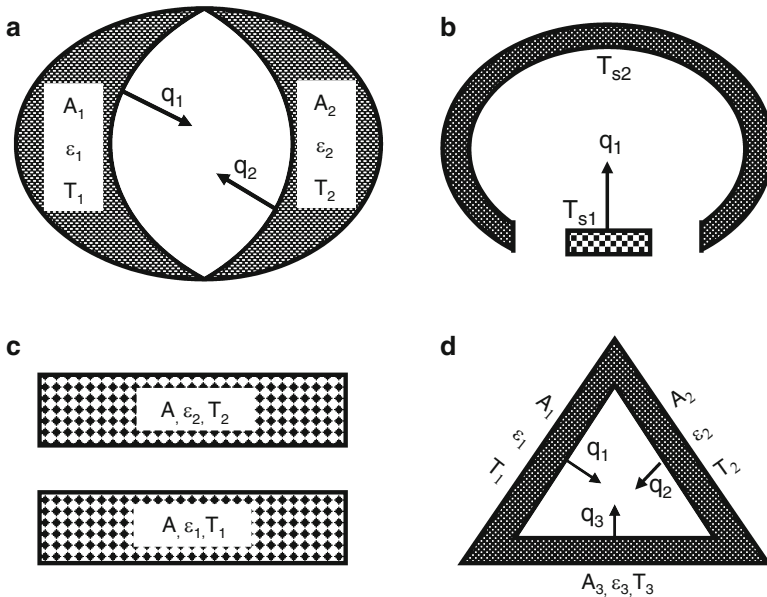


Fig. 1.7 Schematic illustration of thermal radiation: (a) Two surface, gray diffuse enclosure with radiatively nonparticipating media; (b) Small surface in large surroundings; (c) Infinite parallel plates; (d) Three surface gray diffuse enclosure

speed, and in the worst case, damage the device, rendering it inoperable. Even if the worst case can be avoided, reliability is adversely affected through device malfunctions and a shorter system life. Heat transfer in electronic packages is via multiple, coupled modes simultaneously. Figure 1.8 shows heat dissipation paths and causes of thermal resistance, which are influenced by (1) chip area, heat generation (hot-spots), and power consumption; (2) Package materials, structure, dimensions and heat spreader or heat sink; and (3) operating environment, such as cooling conditions, structure of mounting printed wiring board, mounting density, and ambient temperature. Heat dissipation has been designed to be done mostly through the printed wiring board. Because heat radiation effectively occurs only when the surface area of the package is extremely large, heat is actually dissipated from the package via (1) from the surface of the package into the atmosphere; (2) from the external cooling system to the printed wiring board and then into the atmosphere; (3) from the heat source to the sides of the package. Of these three paths, the heat dissipation path via the printed wiring board is the most effective and accounts for 80% of total heat dissipation according to some calculations.

The problem to be solved involves multiple media. Appropriate governing equations must be written based on conservation principles for each medium, along with boundary, interface and initial (if transient) conditions. Traditionally, a single technique is applied at one level of the packaging hierarchy at a time. Multiscale approaches are being developed to analyze across the packaging hierarchy, i.e., modeling performed at the chip-carrier, board, and system levels. The common techniques used include (1) resistor network approach; (2) analytical approach; and (3) numerical solution of governing equations.

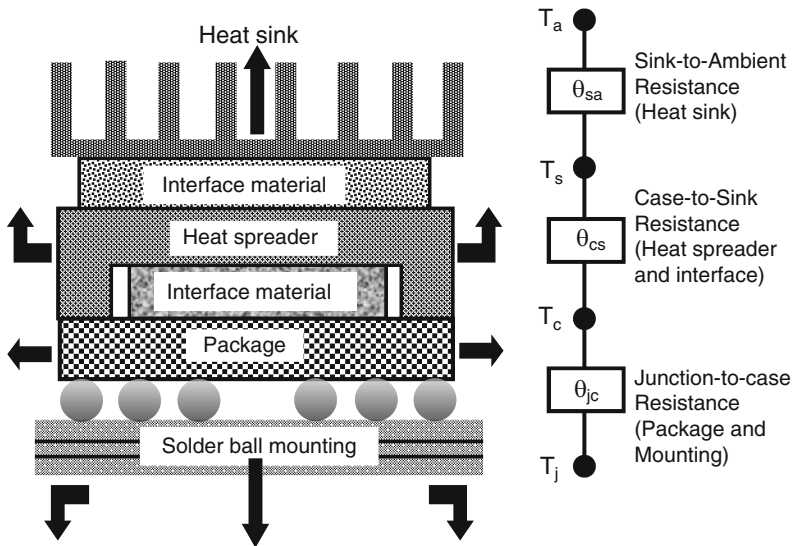


Fig. 1.8 Schematic illustration of heat flow paths in an electronic package

Microscale Heat Transfer

Advances in microfabrication processes of microelectronic devices, such as field effect transistors that contain semiconductor (e.g., silicon), insulator (e.g., silicon dioxide), and metallic (e.g., copper interconnects) layers only a few nanometers thick, have posed a challenge on accurate thermal modeling and design of microelectronic devices and thin film structures at micro- and nanoscales. Microscale heat transfer analysis is unconventional and often challenging because the Fourier heat conduction equation or continuum assumption fails as the characteristic dimension of the structures becomes comparable with the mean free path of energy carriers in phonons of semiconductors and electrons of metals. A phonon refers to the quanta of energy of lattice vibrations in semiconductors such as Si and Ge. Phonons can be treated as particles despite the fact that they are propagating wave packets, which carry energy across the lattice (Asheghi and Liu 2007).

Figure 1.9 gives a schematic hierarchy of microscale heat transfer in electronic packaging. This graph also provides a general guideline for the appropriate treatment of phonon transport in nanostructures. Phonon transport can be predicted using the

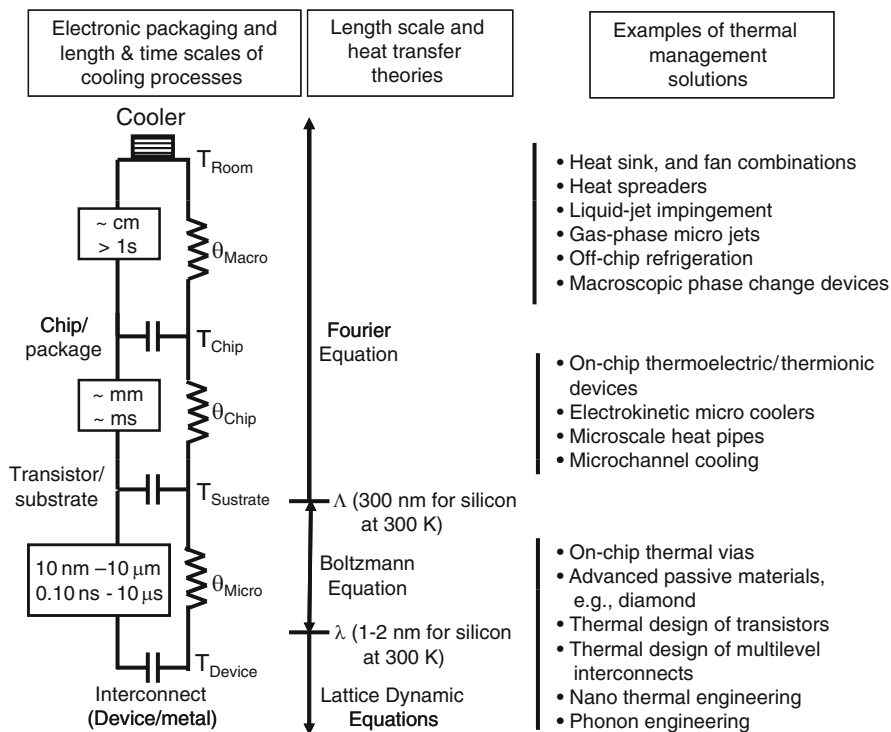


Fig. 1.9 Schematic hierarchy of microscale heat transfer in an electronic package

Boltzmann particle transport equation (BTE), which is required only when the scattering rates of electrons or phonons vary significantly within a distance comparable to their respective mean free paths (Asheghi and Liu 2007). BTE is a general transport equation, and Fourier equation is a special case of simplified BTE. BTE describes particle-like behavior for phonons, photons, and electrons using statistical particle assemblies, which consists of diffusion term, acceleration term, and scattering term (Narumanchi et al. 2006).

$$\frac{\partial f}{\partial t} + v \cdot \nabla f + F \cdot \frac{\partial f}{\partial p} = \left(\frac{\partial f}{\partial t} \right)_{\text{scat}}, \quad (1.21)$$

$$\left(\frac{\partial f}{\partial t} \right)_{\text{scat}} = \frac{f_0 - f}{\tau} + q_{\text{electron-phonon}}, \quad (1.22)$$

$$\frac{1}{\tau} = \frac{1}{\tau_i} + \frac{1}{\tau_u} + \frac{1}{\tau_b}, \quad (1.23)$$

where f is statistical distribution of particles per unit volume and solid angle, depending on time, location and momentum, such as phonon frequency; v is phonon velocity vector, anisotropic and frequency dependent, typically modeled as isotropic for simplicity and lack of comprehensive data; F is external acceleration field and negligible for phonons. Equation (1.22) shows that the scattering term is proportional to difference between phonon equilibrium distribution (f_0) and actual distribution (f) over time to reach equilibrium, i.e., relaxation time τ , which can be obtained from (1.23), where τ_i is due to imperfections; τ_u is due to phonon-phonon scattering; and τ_b is due to phonon-boundary scattering.

The BTE simply deals with the bin counting of the energy carrier particles of a given velocity and momentum, scattering in and out of a control volume at a point in space and time. Analysis of the heat transfer in microelectronic devices, interconnects, and nanostructures using the BTE is very cumbersome and complicated, even for simple geometries, and have been the topic of research and development in the field of micro- and nanoscale heat transfer for the past two decades (Asheghi and Liu 2007).

Lattice dynamic equation can be used to simulate heat conduction across broad length scales with continuum and subcontinuum effects, and solve the BTE by calculating relevant variables on discrete nodal points for both equilibrium and nonequilibrium phenomena. Starting with (1.21) and (1.22), multiplying the equation with $h\omega_p D_p(\omega)$, integrating over frequency, and the energy formulation can be obtained (Narumanchi et al. 2006):

$$\frac{\partial e}{\partial t} + v \cdot \nabla e = \frac{e_0 - e}{\tau} \quad (1.24)$$

Or for one-dimensional space

$$\frac{\partial e}{\partial t} + v_x \cdot \frac{\partial e}{\partial x} = \frac{e_0 - e}{\tau} \quad (1.25)$$

This method can be extended to include electron transport and phonon-electron coupled problems.

For example, the direct simulation of the Boltzmann equation of phonon is being developed and applied to the heat conduction analysis of thin film. However, when dealing with liquid or interphase phenomenon, which is inevitable for phase-change heat transfer, the most powerful tool for the investigation of the microscopic phenomena in heat transfer is the molecular dynamics method. In principal, the molecular dynamics method can be applied to all phases of gas, liquid, and solid and to interfaces of these three phases (Maruyama 2000).

Based on the microscale heat transfer simulation, intelligent electrothermal design along with careful floor planning and microscale thermal management at the device level can largely reduce the temperature rise within a device. This will help to prevent the problem at the early stages and at the device level, rather than to pass the problem to the package-level thermal design (Asheghi and Liu 2007).

Design for Advanced Thermal Management of Electronic Packaging

The thermal design of electronic packaging has become the most important aspect of product design that ensures reliability of the electronic device and thus its ability to compete in a demanding market. The primary objective of the thermal design is to plan the thermal balance of the product so that it operates reliably within the specified environmental conditions throughout its service life (Hienonen et al. 1997). This is generally accomplished by comprehensively considering all factors during thermal design, such as (1) optimizing the heat remove paths from the integrated circuits, i.e., backside of the chip or the substrate for instance; (2) choosing the optimal cooling methods; (3) using highly conductive materials or thermal vias as possible to consume a large of the substrate, reduce interconnect capacity and thermal resistances; (4) minimizing stresses induced in the chips and substrate due to mismatches in thermal coefficients of expansion; and (5) assuring that the set of design goals, the modeling type, the design accuracy, and the test and measurement methods work together so that unintentional over- or underdesign is avoided and that excessively expensive modeling and/or lengthy testing are avoided. All of these considerations will impact performance, cost, and reliability of the product. Furthermore, gathering information for the most critical failure mechanisms of the intended components and materials is necessary to ensure that the main

activities related to the thermal design focus on details where possible failures primarily affect the usability and reliability of the product (Hienonen et al. 1997).

Thermal Design Guidelines

Thermal design is a crucial part of the overall research and development process because thermal control solutions decisively influence the technologies that can or must be used in the electronic product. It is important that thermal, electrical, mechanical, ergonomic, electromagnetic design characteristics are dealt with concurrently with other aspects of the product design. The influence of thermal design on the product is shown in Figure 1.10. Almost all the properties of components and materials change with the temperature and humidity, therefore, controlling these variables is an important part of product design and must be taken into account in thermal design. The thermal control solutions used in the product design invariably limits the scope of design in the areas of mechanics, ergonomics, electromagnetic shielding, and electronic circuits. On the other hand, thermal design solutions can greatly improve the product's functionality, reliability, and corrosion resistance if they are included in the design objectives.

The first task during the thermal design (as shown in Figure 1.10) is to identify the concepts and objectives of the thermal design. The most common objectives of the thermal design in electronics are to (1) keep the junction temperature of the semiconductor devices at or below a given temperature such as 125°C, and keep the temperature of all components inside the product within the limits set by their specifications; (2) level internal temperature differences; and (3) lead excessive heat waste away from the product in a technically and economically sensible way. Prevention of excessive cooling of internal parts of some products in outdoor conditions is also a part of the thermal design. Thermal design is based on the understanding of heat transfer processes from component to unit level through paths of conduction, convection, and radiation. Three general ways of analyzing the design situation include analytical analysis, mathematical modeling, and experiments.

Once the objectives have been set, and when the problem at hand requires the creation of a mathematical model or analytical software, all the boundary conditions influencing the thermal design must be defined, such as (Hienonen et al. 1997): (1) Mechanical constraints, including package or product size and dimensions, and surface treatments of the chosen materials. (2) Electrical constraints, including the power of the unit, the component layout, the PCB mounting layout, electromagnetic compliance requirements for the packaging. (3) Thermal characteristics of components and materials, including maximum allowable operating temperatures of components, reliability target of the unit, temperature dependence of the thermal and other properties of the materials. (4) Environment constraints, including temperature limits of the operating environment, sources of heat in the environment, thermal sinks (mounting platform, ventilation, rain, and wind), and requirements set by the corrosion protection, requirements set by the

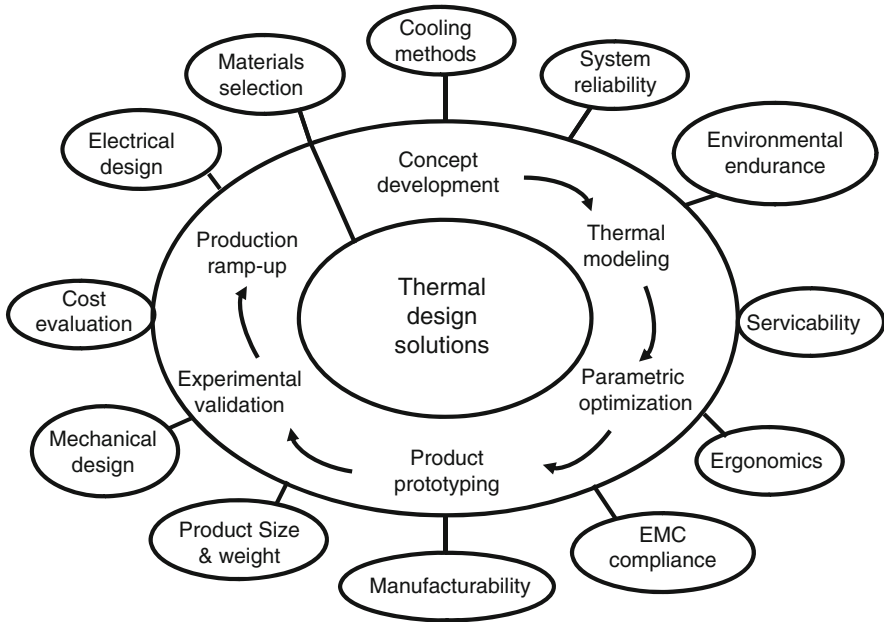


Fig. 1.10 Schematic thermal design guidelines and its influence on the product of electronic packaging

different individual installations. (5) Special requirements by customer or marketing department, such as no fans, no heaters, ergonomics, acoustic noise and appearance.

Once all boundary conditions have been defined based on the product specification, the analysis cases and results sought must be clearly defined. The analysis cases typically include the hottest possible case (such as highest power level, largest heat input from the environment, and highest ambient temperature) and coldest possible case. A case that strongly affects the reliability of the unit is one that includes large power level changes or wide fluctuations of environmental temperature, leading to cyclic temperature changes that stress all interface and materials. Easing these circumstances by thermal design may be the basis of the thermal design process (Hienonen et al. 1997).

Thermal modeling can be performed with analytical analysis or mathematical modeling. For example, one-dimensional or two-dimensional heat equations can be used to evaluate the thermal performance of some electronic components. In most cases, however, numerical modeling is needed to perform thermal analysis for real thermal products, for instance, by solving the flow field and the heat transfer either simultaneously or iteratively. This is because the thermal phenomena in most electronic components and systems are complex and 3-D. In addition, the thermal design in electronic systems not only involves heat transfer but also fluid flow. Furthermore, when mathematical modeling is used as a tool for thermal analysis, experimental validation is always required to validate the accuracy of the mathematical model. The experimental validation involves conducting experiments that

the numerical results can be compared with. With the validated thermal models, various studies can be conducted to study the influence of the design parameters on the thermal performance of the power electronic components and systems (Pang 2005). By identifying the critical design parameters during the parametric studies, the critical design parameter can then be optimized to achieve the best result. Optimization of the design generally concentrates on selectively choosing the best nominal values of design parameters that optimize performance reliability at lowest cost. Then, the best design parameter quantity will be chosen as the output of the design for prototype preparation. Finally, the prototype can be built using the optimized design parameters for further performance evaluation by experimental validation and verification. The verification focuses on the final design and the results of the prototype test and on decisions about possible corrective action if needed before production ramp-up.

With the increase functionality in the power electronics, however, thermal design has become more complex. The need for having a concurrent engineering design process in which electrical, packaging, and thermal designs are performed in parallel is essential in order to achieve a good product design (Pang 2005). In the concurrent engineering design process, the optimized parameters are checked with electrical performance limitations and package manufacturing constraints for optimal performance and manufacturability (Kromann et al. 1998). It involves analyzing all the possible solutions before deciding on the final solution. During the thermal design process, a design guideline had better build up by a standard or principle by which to make a judgment or determine a policy or course of action. Design guidelines can be in the form of theoretic, empirics, or practices that relate the thermal behavior of the power electronics package to the thermal design requirements. By understanding the thermal behavior of the component, the relevant materials, geometries, and cooling strategies can be chosen to optimize the thermal performance of the component. The role of the design guidelines is to provide some insight to the design parameters during the design process. Although the designer may be faced with the available choices for the design, the main point in using the design guidelines is to recognize the available information as well as the known constraints. The design guideline is used to realize common thermal design goals in the electronic components, as mentioned earlier, which typically relate to the thermal performance, reliability, and cost, as well as weight and size reduction. Package design constraints can constitute information on the package manufacturing, assembly, electrical performance, and environment limitations (Pang 2005).

Thermal Modeling and Simulation

The implementation of thermal modeling in the thermal design process can help the designer to understand the thermal behavior of the electronics components and their interrelationship between the electronics components as well as the thermal

dynamics of the system, and thereby to ensure good thermal management for the electronics components in the electronics systems.

In a traditional standard industry semiconductor package, heat is generally conducted from the chip through the wire bonds and solder to the bottom of the package base, and then through an interface material into a heat spreader and heat sink for either air-cooled or water-cooled applications. Heat is also conducted through some thick-compound material or electrical insulated layer to the top of the package. However, almost all of the heat is conducted through the package base rather than to the top of the package due to the higher thermal resistance of the compound material or the insulation layer. In many cases, mathematical models can be used to analyze the thermal behavior of the kind of packages.

On the other hand, the heat transfer in some advanced electronic packages, such as planar multilayer structures, is usually more complicated than in the traditional electronic packages because it is generally 3-D with spatially distributed thermal resistances and capacitances. Therefore, analytical approaches and numerical models are usually required to analyze and simulate the thermal behavior of the planar multilayer structures and to predict the temperature distribution of these structures. In fact, numerical simulation methods for the thermal modeling of electronic packages have routinely been widely used in most design processes. From early spreadsheet-type approaches, the sophisticated finite element analysis or computational fluid dynamics tools are more and more popular, and the mechanical computer aided-design data are directly interfaced into their analysis. Modeling is taken as a necessary step, especially at the early stages of thermal design, during which feasibility studies can narrow the spectrum of possible design choices. Modeling is also used at later stages of the design primarily for verification and design optimization.

Principles of Thermal Modeling

The basic goal of thermal modeling is for cooling electronic packages to improve their reliability. As heat is the major cause of failure of electronic products, thermal modeling is used to find possible means that allow the thermal behavior of the electronic product to stay within the specified ranges at extreme limits of the environmental parameters without the building and testing of expensive prototypes. Also, the thermal model is planned and created to find and screen solutions to thermal problems related to those packaging structures and electronic components that are most critical to the design and of most interest to the designers of the product (Hienonen et al. 1997). For instance, if the operation of the product requires the surface temperature of a microprocessor to be below a certain limit, the model will be constructed so that the best accuracy can be achieved in this particular area and those modes of heat transfer that have the greatest influence on the temperature of this component are modeled accurately. At the same time, the adequate accuracy at the system level should be assured to make the whole model reliable (Hienonen et al. 1997). Therefore, different design targets and objectives will result in different directions of the model modification.

On the other hand, the same thermal model can be used for analysis at different extreme conditions by changing the power levels, environmental boundary conditions, and the thermal characteristics of the packaging structure and materials. One of the most important design criteria for a thermally good structure or system is its insensitivity to changes of individual thermal parameters. As a good practice, sensitivity analysis should always be performed for the thermal model by changing various parameters one by one and recording the subsequent changes of temperature at points of interest; thereby the stability of the system and the inaccuracy caused by incorrect thermal model parameters can be determined, and the model reliability can be increased. For example, the inaccuracy of a good model can be decreased from $\pm 10^{\circ}\text{C}$ up to $\pm 3^{\circ}\text{C}$ (Hienonen et al. 1997).

In addition, if necessary, thermal testing should be used when designing the thermal model. For instance, an accurate description of the behavior of the thermal interface of structures can only be obtained by thermal tests, such as how heat flows and what the real values of the heat transfer coefficient are over the contact surfaces. Using these test results, final fine tuning of the thermal model can be done to reach the real condition. To avoid mistakes and design iterations, however, the test plan must be drawn up and recorded thoroughly, then documented along with the measurement data. The measurement points chosen in the planning of the tests must be simply relatable to the thermal model (Hienonen et al. 1997).

General Approaches

The basic approaches of thermal modeling are generally divided into mathematical model and numerical simulation. Dealing with average thermal parameters such as air temperature, pressure, flow velocity, and overall power, the mathematical model is usually used to simulate the thermal behavior of electronic product in specific situations where the internal and external thermal conditions for the unit vary. Numerical models are usually used to simulate very complicated packages and are full of details ranging from modeled components, printed circuit boards, and board edge guides to the system level. Mathematical or theoretical, numerical and accompanied with experimental validation are equally important and indispensable to the design of a viable, reliable, and cost-effective product. Mathematical modeling of a simplified structure of interest can be especially useful for the selection and mastering of the preprocessing numerical model, while the experimental validation approach enables one to quantitatively assess the role of various uncertainties in the materials properties, geometrical characteristics, and loading conditions.

Predictive modeling is an effective tool for the prediction and prevention of mechanical and functional failure of microelectronic systems, subjected to thermal loading. A very detailed model can predict, for instance, the natural convection from the outer surfaces of the electronic product cabinet, and the radiative heat transfer between component boards and even the conduction of heat in a transistor or a microprocessor along its leads, as well as the effect of the temperature on each component.

Whenever using mathematical model or numerical model, one of the most challenging aspects of thermal modeling is to identify potential problem areas in all kinds of environmental and operational situations, and to understand the methods of heat transfer in such situations in order to increase the geometric accuracy and accuracy of the calculated heat transfer in the right places and in the right way. Often, the places where increased model accuracy is required are not initially known but emerge during the first analysis runs if the modeling work is done carefully. The accuracy of the important parts of the model can be improved by modifying the model iteratively. This entails not only better geometric accuracy but also takes into account all modes of heat transfer. Radiative heat transfer between the internal surfaces of the package or enclosure is often omitted under the impression that only convection is meaningful. In modeling convection, it must be taken into account that the air flow pattern can change drastically over a short distance. Laminar flow can change abruptly to become strongly turbulent, with dramatic impact on the convective heat transfer coefficient of adjacent areas. Such local variations in heat transfer coefficients make modeling work difficult. The accuracy and level of detail of the thermal model determines the quality and usefulness of the results. A simple model-checking routine can easily increase the reliability of the analysis results, including: (1) heat transfer paths must be continuous; (2) thermal balance: in thermal equilibrium, the sum of the power generated in the system and the power absorbed by the system or otherwise received by it must equal the power expelled by the system into its environment (Hienonen et al. 1997).

The complexity of the electronic package being modeled and the purpose of the modeling to a large extent prescribe the selection of modeling tools. For the simplest cases, mathematical calculation and modeling are adequate. As the problems and geometry get more complex, the analysis and numerical tools become more elaborate and powerful. A series of software types have been developed with various numerical programs that solve flow field and radiation simultaneously for time conduction. Whatever tool is used, the prerequisite for successful thermal design is a thorough understanding of heat transfer phenomena and ways to change the relative magnitudes of heat transfer modes in each case, thus guiding the thermal behavior of the product in the desired direction (Hienonen et al. 1997).

Example Methods for Thermal Modeling of Electronic Packaging

The generation of a thermal mathematical model requires a large amount of input data, such as mechanical structure of the analysis object; thermal properties of materials and components; the surface treatments; environmental effects; system overall power level and its distribution; and effects caused by the aging of materials.

Chip Modeling

For a chip with packaging, heat sinks, and cooling systems, the system can be modeled with two parts. First, the temperature distribution in a chip including substrate and interconnects is governed by the heat conduction equation (Wang and Chen 2004):

$$\rho C_p \frac{\partial T(\vec{r}, t)}{\partial t} = \nabla \cdot [\kappa(\vec{r}, T) \nabla T(\vec{r}, t)] + g(\vec{r}, t), \quad (1.26)$$

where T is the time dependent temperature, ρ is the density of the material, C_p is the specific heat, κ is the thermal conductivity, and g is the heat energy generation rate. The physical meaning of (1.20) can be described from the law of energy conservation. For a control volume, the rate of energy stored causing the temperature increase is $dE/dt = \int \rho C_p (\partial T / \partial t) dV$, the rate of heat conduction through surface $d\vec{A}$ is $(\kappa \nabla T \cdot d\vec{A})$ and all surface of the control volume is $Q = \int \kappa \nabla T \cdot d\vec{A} = \int \nabla \cdot [\kappa \nabla T] dV$, and the powered is $Q_p = \int g dV$.

Second, packaging, heat sinks, and cooling systems are modeled as one-dimensional equivalent thermal resistance network. Suppose that the package surfaces are held isothermal. If the package surfaces are not isothermal, a 3-D model is needed for better accuracy to include the contribution due to heat spreading within the package. The effective heat transfer coefficient in the direction of heat flow, \vec{i} , is modeled as $h_i^e = 1/A^i R_{\theta}^i$, where A^i is the effective area normal to \vec{i} and R_{θ}^i are the equivalent thermal resistance. The equivalent convection boundary conditions are

$$\kappa(\vec{r}, T) \frac{\partial T(\vec{r}, t)}{\partial n_i} = h_i^e (T_a - T(\vec{r}, t)), \quad (1.27)$$

where T_a is the ambient temperature and $\partial/\partial n_i$ is the differentiation along the outward direction normal to the boundary surface.

Substrate Modeling

The term $\nabla \cdot [\kappa(\vec{r}, T) \nabla T(\vec{r}, t)]$ in (1.26) can be replaced by $\kappa(T) \nabla^2 T(\vec{r}, t)$ for homogeneous materials, and then a second-order parabolic partial differential equation can be obtained

$$\rho C_p \frac{\partial T(\vec{r}, t)}{\partial t} = k \left[\frac{\partial^2 T(\vec{r}, t)}{\partial x^2} + \frac{\partial^2 T(\vec{r}, t)}{\partial y^2} + \frac{\partial^2 T(\vec{r}, t)}{\partial z^2} \right] + g(\vec{r}, t). \quad (1.28)$$

If the substrate is discretized with size Δx , Δy , and Δz in x , y , and z directions, respectively. Then the temperature $T(x; y; z; t)$ at node $(i; j; k)$ can be replaced by

$T(i\Delta x; j\Delta y; k\Delta z; t)$, which is denoted as $T_{i,j,k}$. To have second order accuracy in space $O[(\Delta x)^2; (\Delta y)^2; (\Delta z)^2]$, the central-difference discretization is applied on (1.22). Therefore, the difference equation at node (i, j, k) can be expressed as

$$\begin{aligned} \rho C_p \Delta V \frac{dT_{i,j,k}}{dt} = & -k \frac{A_x}{\Delta x} (T_{i,j,k} - T_{i-1,j,k}) - k \frac{A_x}{\Delta x} (T_{i,j,k} - T_{i+1,j,k}) \\ & -k \frac{A_y}{\Delta y} (T_{i,j,k} - T_{i-1,j,k}) - k \frac{A_y}{\Delta y} (T_{i,j,k} - T_{i+1,j,k}) \\ & -k \frac{A_z}{\Delta z} (T_{i,j,k} - T_{i-1,j,k}) - k \frac{A_z}{\Delta z} (T_{i,j,k} - T_{i+1,j,k}), \end{aligned} \quad (1.29)$$

where $\Delta V = \Delta x \Delta y \Delta z$ is the control volume of node (i, j, k) , $A_x = \Delta y \Delta z$, $A_y = \Delta x \Delta z$, and $A_z = \Delta x \Delta y$.

Interfaces

The efficiency and quality of heat transfer depend very strongly on the characteristics of the thermal interfaces. One of the most important parts of the modeling is to simulate the heat transfer process at each interface and their proportional shares in the transfer of total waste heat. Thermal interface can be divided into internal and external. External interfaces transfer heat to the environment, and vice versa. Therefore, the characteristics and means of heat transfer at those surfaces determine the average temperature difference between the unit and its environment. Internal interfaces define the heat transfer paths inside the enclosure. Large internal temperature differences can be a sign of poor design of internal heat transfer paths (Hienonen et al. 1997).

Experimental Verification

Experimental verification is used to test and measure product prototypes at various stages of thermal design and product development, complying with the thermal design objectives. Testing includes both functional operation tests and conditioning to different internal operational loads and external environmental loads using calibrated measurement and testing equipment documented and regulated in all quality assurance systems. Although modeling and simulations can give fairly useful information about thermal characteristics when looking for viable options, experimental verification of the performance of the electronic device and cooling system is still imperative in many cases. This is because thermal parameters are not known well enough to verify the system by analytical means only, due to the complexity of materials, components, and constructions used.

Tests and measurements of experimental verification involve various stages of the thermal design and the product development, including (Hienonen et al. 1997): (1) Thermal testing of preliminary structural models when searching for correct

thermal structural solutions. (2) Investigation of functional prototypes to determine the final characteristics before production. (3) Acquisition of field data from production units intended for users by monitoring devices, such as remote measurement technologies, service data from the warranty period. (4) Two technically essential areas: (a) verification of thermal characteristics, such as temperatures, powers, flows, with tests that simulate the operational conditions of a device or a system; (b) verification of the performance and reliability of a device and detection of potential failure mechanisms related to temperature and its variations under the operational conditions of the device.

These fundamental objectives must be taken into consideration when planning verification tests and measurements, which must be designed in close cooperation with the mechanical, EMC, and electrical design as well as other disciplines. Tests must be attached to each device prototype and selected appropriately to each prototype's characteristics. The precursors of the product characteristic during product development, such as cardboard, structure, and thermal models should be tested as appropriate to guide the thermal design and minimize the invariably complicated and time-consuming prototype manufacture (Hienonen et al. 1997).

Materials Selection for Advanced Thermal Management

Advanced electronic systems require improved thermal management to sustain customer expectations of reliability levels. Customer expectations must be satisfied in an environmentally friendly manner and meet the volume, weight, cooling requirements, manufacturability, and reparability of electronic systems. Materials for thermal management in electronics can basically be classified into two main groups which include interface and bulk materials. Interface materials are formulated to provide a low thermal resistance path between the heat-producing device and the second group of materials which move the thermal energy of the device over a larger area and deposit it into a thermal sink. The interface group is usually relatively flexible and is required to overcome the surface irregularities between the device and the heat-sink surfaces for the lowest thermal resistance. In addition to good thermal transfer, these materials may also serve to perform mechanical attachment and offer compliance to provide stress/strain relief, therefore, this group is also considered to include coating and bonding techniques. The materials into which the device thermal energy passes require high thermal conductivity to move the heat power effectively. Other factors which may be significant for these materials include fast thermal response (high thermal diffusivity) to control thermal transient behavior, low weight, and acceptable material and fabrication cost. The lowest device temperature and hence expected reliability will be achieved when the lowest device-to-sink thermal resistance is achieved (Young et al. 2006).

The device-to-sink thermal resistance conventionally consists of a thermal resistance from the active side to the backside of the die, a thermal resistance across the interface between the die and heat spreader, conductive and spreading

thermal resistances from heat spreader, another interface between the heat spreader and heat sink, a thermal resistance associated with the heat sink itself. In many cases, the majority of the device's thermal budget is not taken up by a heat sink, but rather by the interface materials and heat spreaders. Proper selection of thermal interface materials, bulk thermal spreading and dissipating materials, and increasingly, ensuring that these materials interact optimally, is critical to thermal management.

Interface Joining Materials

The interface joining materials are chosen to build up mechanical tolerances in the thermal path with some compliance or flexibility, but at the same time to provide lower thermal interface resistances. Therefore, these joining materials are required that prevent the formation of voids, often a cause of increased thermal resistance. The materials used in assemblies must have similar coefficients of thermal expansion to avoid thermal stress failures. The joining techniques must also take into account the wider range of materials being used and be able to form a low-resistance path between dissimilar materials. There is also pressure to reduce thickness of interface layers in order to reduce thermal resistance. Common materials used for device-to-heat sink interfacing are flexible polymeric films or greases loaded with thermally conductive particles. The filler materials may be metallic or thermally conductive ceramic particle, such as alumina or boron nitride if electrical insulation is required (Young et al. 2006).

In fact, several classes of interface materials have been developed. For most high-power microprocessors, very thin bondlines are desired and grease or grease-like materials are usually used. Phase change materials and thermal gels are thermal grease replacements. Phase change materials have the advantage of being preapplied to a component. Thermal greases, however, remain the standard for thermal resistance performance.

Materials that will form a relatively rigid bond, such as adhesives and solders, also may be used to form a thin bondline. However, these materials are usually used in conjunction with small die, or heat spreaders with a low CTE, to minimize die and bond stress that could lead to die cracking or delamination. The thermal interface materials dominate the total thermal resistance of high power microelectronics.

Tapes typically are used for lower power applications, as are gap-filling pads. The latter also may be used when compliance to variations in component height are required, as the thermal resistance sensitivity to changes in thickness is not high (Dean 2003).

Consequently, interfaces have a significant impact on the thermal impedance of electronic systems and in practice they can be the dominant factor in achieving effective thermal transfer. Interface materials and processes are the methods used to join an electronic device to the thermal transfer medium such as substrate, heat

pipe, and heat sink including coatings and bonding techniques. In this respect they may need to perform the tasks of attachment, stress/strain relief, and thermal transfer over a wide range of temperatures (Young et al. 2006).

Bulk Materials for Heat Spreading and Dissipating

Electronic systems ranging from active electronically scanned radar arrays to web servers all require components made out of bulk materials capable of dissipating heat and maintaining compatibility with the package and die. In response to the needs, many thermal management materials have been developed, such as low-CTE, low-density materials with thermal conductivities ranging between 400 and 1,700 W/m K, and many others with somewhat lower conductivities. Some are low cost; others have the potential to be low cost in high volume.

For instance, the high thermal conductivity materials have been used to make heat spreaders that are attached to the die with the thermal interface material between. The heat spreader has the primary function of spreading the thermal energy from the small footprint of the die to a larger area in order to make more efficient use of convection from the heat sink or package surface.

The surfaces of heat spreaders are often finished to prohibit corrosion or oxidation, improve cosmetics, or allow for marking of some sort. Typical surface finishes are anodizing (aluminum), sputtering, plating, and oxide growth such as black oxide on copper. The effect of the surface finish on thermal properties should be considered. Another key aspect of a heat spreader is the surface roughness and flatness. Both these requirements will have an impact on the average bondline of the thermal interface material and, hence, can significantly impact package thermal performance (Dean 2003).

Another common component is heat sinks usually made of aluminum. Aluminum has the advantage of low cost, easy machining and forming and a corrosion resistant surface which can be further enhanced by anodizing. Its thermal conductivity at around 180 W/m K is low compared with copper (~ 379 W/m K) but aluminum is often preferred on cost and weight factors unless the thermal load is very high.

Other examples are carbon-based materials and active cooling techniques using a liquid. The carbon materials, usually in the form of graphite or diamond, may be combined with metals to give materials that are easier to process into manufactured items. The graphitic materials are lower density than copper or aluminum and can offer higher thermal conductivities. The carbonaceous materials also can be processed to form low-density thermal insulators to protect electronics from heat sources, or to shield parts of the assembly from excessive temperature rise, i.e., in laptop computers and telephone handsets. Phase change is often used in active cooling systems to remove heat from electronics. Almost all of the active cooling systems require a subsequent heat exchanger to transfer the thermal energy to the external environment.

Passive cooling components and materials are generally preferred to active cooling for reasons of cost, complexity, and reliability. The available performance of passive systems is still extending. This is enabled both by improvements in the materials engineering and also by the opportunity for improved design which comes from system-level modeling. However, developments in active cooling are addressing many of the issues which have limited its scope in the past (Young et al. 2006).

In addition, advanced innovative bulk materials must offer substantial improvement relative to mainstream materials in order to have a realistic possibility of being specified. Even if materials do have sufficient technical merit there is a massive investment and lead time required in order to move toward market readiness with appropriate qualification approvals aligning with matched production capacity and downstream integration.

Materials and Components Integration

As requirements on the materials increase, it will no longer be sufficient to choose a good performing material for each component of the heat dissipation path, how these materials interact with each other must be considered. A material that spreads heat very well, but is not wet by an interface material may not perform as well as a lower conductivity material that does wet. Phonon scattering at the interfaces and at plating/surface finish interfaces may need to be optimized in a thermal design. Four approaches have been taken to reduce total thermal resistance (Dean 2003): (1) increase thermal conductivity of the interface materials and bulk heat spreading and dissipating materials; (2) increase wetting or bonding to decrease contact resistance at the surface; (3) increase flatness of the component like spreader to decrease the thickness of the interface to reduce heat transfer path; and (4) eliminate the number of the interfaces in the thermal management package.

Higher Conductivity Materials

While progress has been made, the bulk conductivity of most thermal interface materials is relatively low. Many thermal interface materials now have bulk conductivity in the range of 2–3 W/m K. Improving material conductivity by a factor of 3 to 5 (to ~ 10 W/m K) would result in a reduction in bulk and thus total thermal resistance of approximately $0.10^\circ\text{C cm}^2/\text{W}$. This would drive the bulk resistance down to $0.03\text{--}0.04^\circ\text{C cm}^2/\text{W}$. Further improvements in thermal conductivity would have diminishing returns unless the contact resistance was also improved. Improving the conductivity of the heat spreading and dissipating materials through the use of engineered composite materials or heat pipe/vapor chambers will also be required to drive total device thermal resistance down. Because the spreading

resistance is typically higher than the through-thickness thermal resistance, the use of anisotropic materials is increasing (Dean 2003).

Increasing Wetting or Bonding Forces

Increasing the wetting or bonding to a surface will decrease the contact resistance. For most grease or grease-like interface materials, contact resistance has historically been a fairly low fraction of the overall thermal resistance. However, as the material thermal conductivity has increased and bondline thickness decreased, the contact resistance has begun to become significant. Contact resistance can be reduced through improving the bonding or wetting of the interface material to each surface at the interface. Materials should no longer be chosen individually, but rather the effect of the materials working together will be evaluated. It will be important for suppliers and users of interface materials to work with suppliers and users of heat spreaders to ensure optimal performance. Synergy between metal finishes and interface materials should allow contact resistance to be decreased by 50% (Dean 2003).

Decreasing Interface Thickness

With everything else held constant, a thinner bondline between components will produce a lower interface thermal resistance. This will require flatter components. The tightening of flatness tolerances and shift to die referenced cooling has dropped interface thicknesses by a factor of 2 or more (from ~50 to 25 μm or less). Further improvements in spreader or lid flatness will enable thinner interfaces and better package thermal performance (Dean 2003).

Elimination of the Number of Interfaces

Most high-performance devices use a thermal lid or heat spreader. This creates at least two thermal interfaces in the heat removal path from the die backside to the ambient: the die to spreader and spreader to heat sink interfaces. With current performance levels, eliminating one of the interfaces will eliminate a significant fraction of the overall junction to ambient thermal resistance. Thermal lids are used to protect the die and spread the heat from the concentrated die footprint to a larger area so that the heat sink is more efficient in dissipating heat to the environment. Heat sinks with thicker bases or high performance bases, will help address the latter concern, minimizing the need for a package heat spreader. When a bare die package is used, the interface material between the die and heat sink is softer, or more compliant, to prevent damage during assembly or operation. Improvements in interface materials combined with designs to protect bare die will enable the wider spread use of bare die packaging and its inherent elimination of one of the two interfaces in the heat transfer path (Dean 2003).

As package and heat sink design have improved significantly, the materials used largely determine the thermal performance of electronic packages. Thermal resistance across physical interfaces has progressed from an almost negligible portion of the total junction to ambient thermal resistance to the dominant factor in total thermal resistance. Heat spreading and reduction of hotspots within a die are creating a need for higher conductivity materials. Further increases in device power dissipation will require significant improvements from materials. Increased thermal conductivity ($k \geq 10$ W/m K), improved ability to work synergistically with other packaging components and package designs to allow thinner bondlines, and increased use of bare die packaging will be needed to meet the further challenges (Dean 2003).

With the continuing trend in electronic systems toward higher power and increased packing density, advanced materials for efficient thermal management has become a crucial need. Typical electronic devices and their packages consist of a variety of different types of materials, including metals, semiconductors, ceramics, composites, and plastics. The most important physical properties in the use of materials in thermal management are thermal conductivity and the CTE. It is also quite evident that in general there is an enormous disparity between the CTE of high conductivity metals used for heat sinks (aluminum and copper) and insulators used for electronic substrates (alumina, BeO, AlN, etc.). One of the many challenges of electronic packaging is bridging this thermal expansion gap in a manner that does not compromise the thermal efficiency of the package. To solve this problem, composite materials have been developed and utilized. The printed circuit board is an example of a polymer matrix composite. Metal matrix composites (MMCs) are fabricated using a high thermal conductivity metal matrix such as aluminum or copper, with a low CTE material added to reduce the overall CTE of the composite. By proper adjustment of the relative composition of the composite, the CTE can approach that of silicon and insulating materials while maintaining high thermal conductivity. For instance, the kinetic spray to fabricate MMCs has been developed. Further improvement of the material properties of MMCs for thermal management, and advanced manufacturing techniques that will allow cost-effective MMC fabrication on a production scale.

Environmental Compliance of Thermal Management Materials

Environmental compliance must be evaluated during thermal design of electronic packaging because it has become a key advantage of competitive electronic products. The requirements are regulated by the Restriction of the Use of Certain Hazardous Substances in Electrical and Electronic Equipment (RoHS) and Waste Electrical and Electronic Equipment (WEEE) directives promulgated by the European Union (EU). The directives make it critical for selections of thermal management materials. These environmental compliance regulations provide forceful guidelines for designing electronic products with environmental compliance.

RoHS

RoHS Directive 2002/95/EC, together with WEEE Directive 2002/96/EC became European Law in February 2003, setting collection, recycling and recovery targets for all types of electrical goods. Producers must comply with all provisions of the WEEE directive after August 13, 2005, while the RoHS directive requires that producers prohibit any of the banned substances on the market after July 1, 2006. Producers failing to comply with the RoHS and WEEE directives’ requirements face legal penalties and potential restriction from selling products in the EU.

Table 1.2 lists materials that are currently restricted by the RoHS directive (Tong 2009). These substances are restricted to the ppm (parts per million) threshold level in all applications. All homogeneous material in purchased articles (i.e., materials, components, subassemblies, or products) must be free of the substances or cannot contain higher concentrations than the defined ppm (parts per million) threshold levels as listed in the table. Table 1.3 lists the banned substances in electronic packaging, while the substances must be reported when concentration exceeds the indicated ppm threshold level as listed in Table 1.4 (Tong 2009). Exemptions to the maximum allowed concentrations of restricted materials are identified for cases where technology does not yet allow for substitutions, or where alternatives may have a worse impact on human health and the environment. Some exemptions include mercury in several kinds of fluorescent lamps; lead in steel, copper, and

Table 1.2 Restriction of hazardous substances restricted substances

Material/substance category	Maximum concentration in weight
Asbestos	Not detectable
Certain Azo colorants and Azo dyes	Not detectable
Cadmium/cadmium compounds	75 ppm
Hexavalent chromium/hexavalent chromium compounds	1,000 ppm
Lead/lead compounds	1,000 ppm 300 ppm (PVC cables only)
Mercury/mercury compounds	1,000 ppm
Ozone depleting substances (continuous fiber ceramics, Hydro chlorofluorocarbons, HBFCs (Hydrobromofluorocarbons), carbon tetrachloride, etc.)	Not detectable
Polybrominated biphenyls	1,000 ppm
Polybrominated diphenylethers	1,000 ppm
Polychlorinated biphenyls	Not detectable
Polychlorinated naphthalenes (more than three chlorine atoms)	Not detectable
Radioactive substances	Not detectable
Certain short chain chlorinated paraffins	Not detectable
Tributyl tin and triphenyl tin	Not detectable
Tributyl tin oxide	Not detectable

Table 1.3 Banned substances in electronic packaging

Material/substance category	Maximum concentration in weight
Cadmium, mercury, lead, chromium VI	Total amount 100 ppm
Chlorofluorocarbons	Not detectable
Hydro chlorofluorocarbons	Not detectable

Table 1.4 Substances that must be reported when concentration exceeds the threshold level

Material/substance category	Maximum concentration in weight (ppm)
Antimony/antimony compounds	1,000
Arsenic/arsenic compounds	1,000
Beryllium/beryllium compounds	1,000
Bismuth/bismuth compounds	1,000
Brominated flame retardants (other than PBBs (Polybrominated biphenyl) or PBDEs (Polybrominated diphenylethers))	1,000
Nickel (external applications only)	1,000
Certain phthalates	1,000
Selenium/selenium compounds	1,000
Polyvinyl chloride (disclosure is limited to “is present”/is not present” in amounts that exceed threshold)	1,000

aluminum alloys; lead in some types of solder; and military applications. RoHS Article 3(a) states that RoHS covers electrical and electronic equipment “which is dependent on electric currents or electromagnetic fields in order to work properly and equipment for the generation, transfer, and measurement of such currents and fields falling under the categories set out in Annex IA to Directive 2002/96/EC (WEEE) and designed for use with a voltage rating not exceeding 1,000 V for alternating current and 1,500 V for direct current.” With that said, a microwave oven would not be covered by RoHS because it cannot perform its intended function with the power in the off position. On the other hand, a talking doll can still be used as a doll even when the batteries are removed, therefore, it isn’t covered by RoHS. RoHS does allow for noncompliant components after the July 1, 2006, but only as spare parts for equipment on the market before July 1, 2006.

WEEE

The WEEE directive imposes the responsibility for the disposal of WEEE on the manufacturers of such equipment. Those companies should establish an infrastructure for collecting WEEE in such a way that “Users of electrical and electronic equipment from private households should have the possibility of returning WEEE at least free of charge.” Also, the companies are compelled to use the collected

waste in an ecologically friendly manner, either by ecological disposal or by reuse/refurbishment of the collected WEEE. The WEEE identifies producers as any company that sells electronic and electrical equipment directly or indirectly under its own brand name. The intent of the WEEE directive is to require producers to design products and manufacturing processes which prevent the creation of WEEE and barring that, reuse, recycle, dispose of, or incinerate WEEE. The WEEE directive calls for set percentages of IT (information technology) and telecommunications equipment (Category 3 includes personal computers, wireless devices, and similar devices) to be recovered and reused, or recycled (minimum 65%), incinerated (maximum 10%), or safely disposed of (maximum 25%).

Therefore, the goal of the thermal design and material selection should be compliant with RoHS and WEEE directives in a minimized cost manner. Meanwhile, designing for environmental compliance must assure that time-to-market is minimized to take the product's competitive advantages.

Summary

Escalation of heat flux and power dissipation of electronic chips resulted from the demand for high-performance microprocessors. Meanwhile, the desire for smaller form-factor system and lower semiconductor operating temperatures is compounding the thermal challenge. Thermal design for a microprocessor can no longer be treated in isolation. Power and performance tradeoffs and smart circuit design techniques are required to conserve power consumption. Advanced materials and process improvements in packaging and cooling technology are required to minimize thermal resistance. Therefore, viable thermal design and cooling solutions are critical for development of high performance microprocessors and cost-effective electronic packaging.

In fact, the importance of thermal management has waxed and waned through the technology and product generations of the past decades. Advanced integrated circuit and photonic technologies as well as the ubiquity of electronic system applications are providing a serious challenge to heat transfer design and product development in terms of basic theory, tools, components, and innovative design. Breakthroughs are needed in advanced cooling and pragmatic design at all package levels.

The mechanisms of heat transfer depend on the media involved and are usually divided into conduction, convection, radiation, and multimode, which is a combination of one or more of the above. When a temperature gradient exists within a continuous and nonmoving medium such as solid or stationary fluid, heat is transferred through the medium via the conduction mode. Convection heat transfer occurs when a surface is in contact with a moving fluid, liquid, or gas at a different temperature. Heat is exchanged between the fluid and the surface. Radiation heat transfer occurs when two surfaces at different temperatures exchange energy in the form of electromagnetic energy emitted by the surfaces. Radiation can occur in a

vacuum because no medium between the two surfaces is required as in conduction and convection. The multimode problems present in most electronic packages with various mediums. Appropriate governing equations must be written based on conservation principles for each medium, along with boundary, interface, and initial conditions. Traditionally, a single technique is applied at one level of the packaging hierarchy at a time. Multiscale approaches are being developed to analyze across the packaging hierarchy, i.e., modeling performed at the chip, board, and system levels. The common techniques used include (1) a resistor network approach; (2) an analytical approach; and (3) a numerical solution of governing equations.

General thermal management solutions typically include hardware assembly, software control, and optimal thermal design as well as the combination of these approaches. System hardware solutions are based on internally distributing and externally dissipating the thermal energy to the ambient environment. System software solutions typically regulate the dissipative power of the device based on active system feedback controls. Optimal thermal design is required for cooling and thermal control of all level packages, particularly effective for device-level or chip-level micro/nanoscale thermal engineering and processing.

The thermal design of electronic packaging has become the most important aspect of product design that ensures reliability of the electronic device and thus its ability to compete on a demanding market. The primary objective of the thermal design is to plan the thermal balance of the product so that it operates reliably within the specified environmental conditions throughout its service life. This is generally accomplished by comprehensively considering all factors during thermal design, such as (1) optimizing the heat removal paths from the integrated circuits; (2) choosing the optimal cooling methods; (3) using highly conductive materials or thermal vias as possible to consume a large amount of the substrate and reduce interconnect capacity and thermal resistances; (4) minimizing stresses induced in the chips and substrate due to mismatches in thermal coefficients of expansion; and (5) assuring that the set of design goals, the modeling type, the design accuracy, and the test and measurement methods work together so that unintentional over- or underdesign is avoided and that excessively expensive modeling and/or lengthy testing is avoided. All of these considerations will impact performance, cost, and reliability of the product. Furthermore, gathering information on the most critical failure mechanisms of the intended components and materials is necessary to ensure that the main activities related to the thermal design focus on details where possible failures primarily affect the usability and reliability of the product.

Advanced materials are becoming critical for thermal management of micro-electronic systems. The systems ranging from active electronically scanned radar arrays to web servers all require materials capable of dissipating heat and maintaining compatibility with the package and die. In response to these needs, many thermal management materials have been developed, including low-CTE, low-density materials with thermal conductivities ranging between 400 and 1,700 W/m K. These materials have been used for fabrication of servers, laptops, PCBs, PCB cold plates/heat spreaders, cellular telephone base stations, hybrid electric vehicles, power modules, phased array antennas, thermal interface materials (TIMs), optoelectronic

telecommunication packages, laser diode and light-emitting diode packages, and plasma displays.

In addition, environmental compliance must be evaluated during thermal design of electronic packaging because it has become a key advantage of competitive electronic products. The requirements are regulated by the RoHS and WEEE directives promulgated by the EU. The directives make selections of thermal materials critical. These environmental compliance regulations provide forceful guidelines for designing electronic products with environmental compliance.

References

- Ashghi M, Liu W (2007) Microscale heat transfer. *Electronics Cooling*. <http://www.electronics-cooling.com/articles/2007/feb/a2>. Accessed 16 February 2010.
- Bornoff R (2005) Get on board: solving thermal problems at board level. *Printed Circuit Design & Manufacture* 22:38–41.
- Brooks D, Martonosi M (2001) Dynamic thermal management for high-performance microprocessors. *Proceedings of the 7th International Symposium on High-Performance Computer Architecture*, Monterrey, Mexico. pp. 171–182. IEEE, Piscataway.
- Cao L et al (1996) Transient thermal management of portable electronics using heat storage and dynamic power dissipation control. *Electronic Components and Technology Conference*, Orlando. pp. 205–211. IEEE, Piscataway.
- Chen G, Shakouri A (2002) Heat transfer in nanostructures for solid state energy conversion. *Journal of Heat Transfer* 124:242–252.
- Chomerics X X (1999) Thermal management products. <http://www.chomerics.com/products/documents/thermcat/THERMALMP.pdf>. Accessed 06 February 2010.
- Clark L et al (2002) An embedded 32-b microprocessor core for low-power and high-performance applications. *IEEE Journal for Solid-State Circuits*, 56:1599–1608.
- Cooper B (2006) Operating system coordinated thermal management. US patent 7054720.
- Couvillion R J (2006) Thermal considerations. In Ulrich R K, Brown W D (eds) *Advanced Electronic Packaging*, 2nd edn. John Wiley & Sons, Hoboken.
- Dean N (2003) Materials for thermal management. *Advanced Packaging*. http://ap.pennnet.com/display_article/169366/36/ARTCL/none/none/1/1/Materials-for-Thermal-Management/. Accessed 26 February 2010.
- Hannemann R (2003) Thermal control of electronics: perspectives and prospects. <http://web.mit.edu/hmtl/www/papers/HANNEMANN.pdf>. Accessed 30 January 2010.
- Hienonen R, Karjalainen M, Lankinen R (1997) Verification of the thermal design of electronic equipment. *VTT Automation*. <http://www.vtt.fi/inf/pdf/publications/1997/p.320.pdf>. Accessed 23 February 2010.
- Kromann G, Pimont V, Addison S (1998) CFD and EDA tools the interoperability of flotherm® and board station®/AutoTherm®: Concurrent design of a motorola powerPC™ RISC microprocessor-based microcomputer. Freescale Semiconductor, Inc., <http://www.freescale.com/files/archives/doc/white-paper/603604THERMAL.pdf>. Accessed 26 February 2010.
- Krum A (2004) Thermal management. In Harper C A (ed) *Electronic packaging and interconnection handbook*, 4th edn. McGraw-Hill Professional, New York.
- Kumar A, Li S, Peh L-S, Jha N K (2008) System-level dynamic thermal management for high-performance microprocessors. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 27:96–108.
- Madrid A (2000) Thermal management for applied computing. *Intel Developer Update Magazine* (April). <http://www3.intel.com/technology/magazine/computing/ac04002.pdf>. Accessed 16 January 2010.

- Mallik D et al (2005) Advanced package technologies for high-performance systems. *Intel technology Journal* 9:259–271.
- Manning A, Johns D (2007) Thermal/EMC simulation addresses new mechanical design challenges. http://www.conformity.com/artman/publish/printer_228.shtml. Accessed 03 February 2010.
- Maruyama S (2000) Molecular dynamics method for microscale heat transfer. In Minkowycz W J and Sparrow E M (Eds), *Advances in Numerical Heat Transfer 2*:189–226, Taylor & Francis, New York.
- März M (2003) Thermal management in high-density power converters. *IEEE International Conference on Industrial Technology* 2:1196–1201.
- Moran M (2001) Micro-scale avionics thermal management. NASA/TM-2001-211095. <http://gltrs.grc.nasa.gov/GLTRS>. Accessed 09 February 2010.
- Narumanchi S V J, Murthy J Y, Amon C H (2006) Boltzmann transport equation-based thermal modeling approaches for hotspots in microelectronics. *Heat and Mass Transfer* 42:478–491.
- NEC Electronics (2003) Thermal design of a package. http://www.necel.com/pkg/en/characteristic/heat01_03.html. Accessed 21 February 2010.
- Pang Y F (2005) Assessment of thermal behavior and development of thermal design guidelines for integrated power electronics modules. Ph.D. Dissertation. Virginia Polytechnic Institute and State University, Blacksburg.
- Petrosjanc K O (2009) Thermal design system for chip- and board-level electronic components. <http://ewdtest.com/conf/EWDTS09/57.pdf>. Accessed 12 February 2010.
- Rasmussen F E (2003) Electrical interconnections through CMOS wafers. PhD Thesis. Technical University of Denmark, Hellerup.
- Siebert W P (2005) Alternative electronic packaging concepts for high frequency electronics. PhD Thesis. Royal Institute of Technology, Tellinge (Sweden).
- Srinivasan J, Adve S (2003) Predictive dynamic thermal management for multimedia applications. 17th Annual International Conference of Supercomputing, San Francisco. pp. 109–120. ACM, New York.
- Tong X C (2009) Advanced materials for electromagnetic interference shielding. CRC Press, Boca Raton.
- Viswanath R, Wakharkar V, Watwe A, Lebonheur L (2000) Thermal performance challenges from silicon to systems. *Intel Technology Journal* Q3. ftp://download.intel.com/technology/itj/q32000/pdf/thermal_perf.pdf. Accessed 26 January 2010.
- Wang T, Chen C (2004) SPICE- Compatible thermal simulation with lumped circuit modeling for thermal reliability analysis based on modeling order reduction. http://cc.ee.ntu.edu.tw/~cchen/research/ISQED2004_Wang.pdf. Accessed 26 February 2010.
- Wirth E (2004) Thermal management in embedded systems. BSc Thesis, University of Virginia, Charlottesville.
- Young R et al (2006) Developments and trends in thermal management technologies – a mission to the USA. www.lboro.ac.uk/research/iemrc/documents/.../CB2007.pdf. Accessed 26 February 2010.