Analytical and Numerical Modeling of the Thermal Performance of Three-Dimensional Integrated Circuits

Ankur Jain, Robert E. Jones, *Member, IEEE*, Ritwik Chatterjee, and Scott Pozder

***Abstract*—Three-dimensional (3D) interconnection technology offers several electrical advantages, including reduced signal delay, reduced interconnect power, and design flexibility. 3D integration relies on through-silicon vias (TSVs) and the bonding of multiple active layers to stack several die or wafers containing integrated circuits (ICs) and provide direct electrical interconnection between the stacked strata. While this approach provides several electrical benefits, it also offers significant challenges in thermal management. While some work has been done in the past in this field, a comprehensive treatment is still lacking. In the current work, analytical and finite-element models of heat transfer in stacked 3D ICs are developed. The models are used to investigate the limits of thermal feasibility of 3D electronics and to determine the improvements required in traditional packaging in order to accommodate 3D ICs. An analytical model for temperature distribution in a multidie stack with multiple heat sources is developed. The analytical model is used to extend the traditional concept of a single-valued junction-to-air thermal resistance in an IC to thermal resistance and thermal sensitivity matrices for a 3D IC. The impact of various geometric parameters and thermophysical properties on thermal performance of a 3D IC is investigated. It is shown that package and heat sink thermal resistances play a more important role in determining temperature rise compared to thermal resistances intrinsic to the multidie stack. The improvement required in package and heat sink thermal resistances for a 3D logic-on-memory implementation to be thermally feasible is quantified. An increase in maximum temperature in a 3D IC compared to an equivalent system-in-package (SiP) is predicted. This increase is found to be mainly due to the reduced chip footprint. The increased memory die temperature in case of memory-on-logic integration compared to a SiP implementation is identified to be a significant thermal management challenge in the future. The results presented in this paper may be useful in the development of thermal design guidelines for 3D ICs, which are expected to help maximize the electrical benefits of 3D technology without exacerbating thermal management issues when implemented in early-stage electrical design and layout tools.**

***Index Terms*—Die stacking, electrical-thermal co-design, junction-to-air thermal resistance, three-dimensional (3D) integrated circuits (ICs), through-silicon via (TSV).**

Manuscript received September 30, 2008; revised January 27, 2009. First published September 18, 2009; current version published nulldate. Recommended for publication by Associate Editor T. Lee upon evaluation of reviewers' comments.

A. Jain is with Molecular Imprints, Inc., Austin, TX 78758 USA.

R. E. Jones and S. Pozder are with Freescale Semiconductor, Austin, TX 78735 USA (e-mail: ankurjain@stanfordalumni.org).

R. Chatterjee is with the Georgia Insititute of Technology, Atlanta, GA 30332 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCAPT.2009.2020916

I. Introduction

THE continued reduction of microelectronics device size over the past few decades has led to unprecedented improvement in performance of electronic products. This miniaturization has been fueled by continued improvements in process technologies. As physical limits for several process technologies begin to be approached, the further reduction of device size is starting to get more and more challenging from both technological and financial perspectives. This has fueled interest in other means of obtaining the benefits offered by dimensional scaling. Examples include subambient microprocessor operation [1], three-dimensional (3D) integrated microelectronics [2]–[4], etc. Vertically integrated 3D circuits have attracted significant attention in the recent past due to several potential benefits. 3D integration is expected to increase device density, reduce signal delay, and enable new circuit and architecture design paradigms [5]. Vertical integration of heterogeneous technologies using 3D integration may offer many advantages over alternatives like system-on-chip (SoC) [6]. There are no mainstream design and layout tools available for 3D technology at present, although there is a significant ongoing effort in this direction.

A number of innovative process technologies contribute toward the realization of a working 3D circuit. These include wafer thinning, etching and filling of high aspect ratio holes in silicon, interstrata bonding, etc. [7]–[9]. A key component of 3D technology is a metal-filled through-silicon via (TSV) that enables communication between the two die as well as with the package. A variety of integration technologies for manufacturing 3D integrated circuits (ICs) have been demonstrated. These include both face-to-face integration [7] and back-to-face integration [9].

While 3D technology has some clearly established benefits in terms of electrical performance, it also exacerbates the already severe challenge of microelectronics cooling [10]. Vertical integration results in increased device density and, hence, higher power density. The thermal characteristics of TSVs and interdie bonding layer are not well known. The implementation of heterogeneous integration must also take into consideration the different operating temperature requirements of individual strata. While at the circuit level, it may be possible to utilize a metal-filled TSV for thermal hotspot management [11], [12], aggressive 3D integration is likely to lead to more challenging thermal management requirements at the package level.

Successful application of 3D integration in products will require analysis of several thermal management problems and establishment of thermal design rules governing the feasibility of various integration options. Papers in the past have individually addressed some of the several thermal issues related to 3D integration. Kleiner, *et al.* presented a simple, one-dimensional (1D) thermal analysis of a three-chip stack [13]. 3D FEM simulation was used in this work to investigate the effect of interconnects on temperature distribution in a two-die stack. The type of thermal boundary conditions assumed in this work is not clear. A thermal analysis of various 3D integration schemes has been presented [14]. Some work has been reported on optimizing the problem of placement of vias for heat dissipation in 3D ICs [11], [12]. Numerical thermal simulations have been carried out to convert power dissipation distribution into a temperature distribution in a 3D IC [15]. The electrical challenge involved in stacking memory on logic has been analyzed, and it has been shown that the increase in peak temperature in this case is expected to be limited to a few °C [16]. In fact, the potential reduction in power consumption in signal transmission due to reduced interconnect length may further diminish the thermal penalty to be paid due to vertical integration. Recently, convective heat transfer from a multidie stack using liquid cooling has been also analyzed [17], [18].

In order to accurately analyze thermal management of 3D ICs, the development of a fundamental analytical model for heat transport in 3D ICs is desirable. Such a model will provide a framework in which to analyze the general problem of heat dissipation in 3D ICs and will offer thermal design guidelines that address the limits of what can be 3D-integrated without becoming thermally infeasible. For example, chip and package designers would be interested in finding out the location of the maximum temperature in a given multidie stack and its sensitivity to various design parameters. Furthermore, there is interest in developing tools for multiphysics thermal-electrical optimization of the stacking arrangement and smart design of the chip-package interface in order to alleviate the thermal dissipation problem. The influence of features inherent to 3D technology such as TSVs, interdie bonding layers, etc., on the temperature distribution also needs to be investigated. While the thermal characteristics of microelectronics have traditionally been represented in terms of junction-to-air thermal resistance and density factors [19], these concepts needs to be expanded in view of the presence of multiple heat sources and multiple junction temperature in a 3D IC.

The current work presents an analytical model for heat transfer in a multisource 3D stack. The model is used to predict the temperature rise in various strata and determine the thermal feasibility limits of multidie integration. Numerical modeling for capturing thermal hotspot effects is also discussed in the paper and is used to compare thermal performance of a 3D IC with a system-in-package (SiP).

II. Heat Transfer Model for 3D Integrated Electronics

A. Analytical Model

In order to develop an analytical model governing the temperature rise in various strata in a 3D IC, one may assume uniform heat generation in the device planes and neglect heat spreading by assuming heat flow only normal to the device planes. In general, a 3D microelectronics system comprises N die electrically and mechanically connected to each other through an appropriate bonding technology. Each die is associated with two significant thermal resistances—that of the silicon/SOI substrate and that of the back-end-of-line (BEOL) metal-dielectric stack. Despite its small thickness, BEOL thermal resistance is becoming increasingly significant due to the low thermal conductivity of low-k dielectrics that are being incorporated in BEOL stacks. The other significant thermal resistances within the die stack are those of the micropad bonding layers between die. A numerical model of the interstrata micropad layer is presented later in order to estimate the value of this thermal resistance.

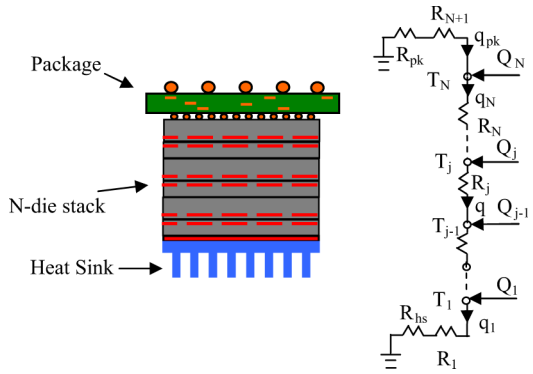

Fig. 1. Schematic of the general N-die stack and the thermal resistance network with multiple heat generating junctions. Dashed lines represent the planes in which heat generation occurs.

The general thermal resistance network, shown in Fig. 1, comprises N heat sources and $(N+3)$ thermal resistances, including $R_{\rm hs}$ and $R_{\rm pk}$, which are the resistances of the heat sink and package, respectively. R_1 is the thermal resistance between the node on the stratum closest to the heat sink and the heat sink, and R_{N+1} is the thermal resistance between the node on the stratum closest to the package and the package. R_j is the thermal resistance between junction j and junction $j-1$ and is the sum of thermal resistances of substrates, BEOL stacks, and interdie micropad layers located between the two junctions. For example, in case of a face-to-face, two-die stack, the thermal resistance R_2 between junctions 1 and 2 comprises the thermal resistance of two BEOL stacks and the micropad layer between the two die. T_j and Q_j represent the temperature and heat generation at node j, respectively. q_j represents the heat flow from node j to node $j-1$. For low-power applications that operate without a heat sink, the thermal resistance of the heat sink may be conveniently removed from this general analysis. This representation of the 3D, multidie stack differs from the two-resistor model used to describe the traditional single substrate in the introduction of several new resistances and multiple heat sources. Physically speaking, heat generated in one stratum must travel through several other strata before dissipating into the package or heat sink, thus causing additional temperature rise at each junction. As a result, temperature rise in one stratum is influenced by heat generation not only in that stratum, but in all other strata as well. In steady state,

temperature distribution in the various strata is governed by the following energy-conservation equations.

For $j = 1, 2 \ldots, N-1$,

$$q_j = q_{j+1} + Q_j. \tag{1}$$

For $j = N$, we have

$$q_N = q_{\rm pk} + Q_N. \tag{2}$$

For $j = 2, 3 \ldots, N$, temperature and heat flow are related to each other as follows:

$$T_j - T_{j-1} = q_j \cdot R_j. \tag{3}$$

Finally, the following equations govern the temperature of nodes in the strata next to the heat sink and package:

$$\begin{aligned} T_N - T_{\rm amb} &= -q_{\rm pk} \cdot (R_{\rm pk} + R_{N+1}) \\ T_1 - T_{\rm amb} &= q_1 \cdot (R_{\rm hs} + R_1) \end{aligned} \tag{4}$$

where $T_{\rm amb}$ is the ambient temperature.

Equations (1)–(4) represent $(2N+1)$ equations that can be easily solved to determine the $(2N+1)$ variables $-T_1, T_2 \ldots, T_N, q_1, q_2 \ldots, q_N$ and $q_{\rm pk}$. The solution is found to be

$$\begin{aligned} \frac{\theta_i}{R_{\rm hs}} = &\sum_{j*=i}^{N} Q_{j*}\,(1-F_{j*}) \left(1+\sum_{m=1}^{i} r_m\right) \\ &+\sum_{j*=1}^{i-1} Q_{j*} \left[(1-F_{j*})\left(1+\sum_{m=1}^{j*} r_m\right) - F_{j*} \sum_{m=j*+1}^{i} r_m\right] \end{aligned} \tag{5}$$

where

$$F_{j*} = \left(\frac{1+\sum_{m=1}^{j*} r_m}{1+\alpha+\sum_{m=1}^{N} r_m}\right) \tag{6}$$

and

$$\alpha = \left(\frac{R_{\rm pk} + R_{N+1}}{R_{\rm hs}}\right). \tag{7}$$

r_m represents the resistance between nodes m and $m-1$ nondimensionalized by $R_{\rm hs}$. θ_i represents the temperature rise above ambient of the ith node. Equations (5)–(7) represent the temperature distribution in the various nodes and help answer some of the key thermal questions in 3D IC design, including the maximum temperature attained, location of the hottest stratum, influence of the heat sink and package thermal resistances, the thermally optimum arrangement of the various strata, and the sensitivity of the maximum temperature to various parameters of the problem.

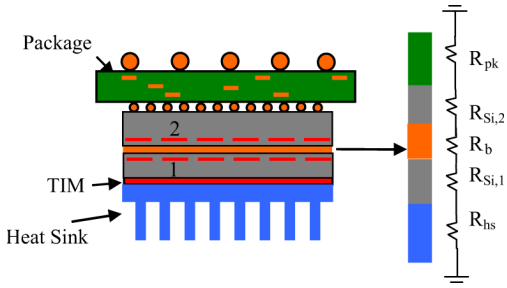

Fig. 2. Schematic of a face-to-face bonded two-die stack with the various thermal resistances between the junctions and ambient.

Note that, in general, the hottest die is not necessarily the one located farthest from the heat sink. Depending on the relative magnitude of heat generation in various die, and the relative magnitude of package and heat sink thermal resistance, the hotspot may lie in any of the die in the N-die stack. The temperature profile increases monotonically with increasing node number starting from node 1 until the hottest node, following which it decreases monotonically.

The general solution represented by (5)–(7) takes on simpler forms for a number of special cases. Two such cases are of particular interest. In the first case, one may consider a two-die stack, shown in Fig. 2, which is the simplest and likely to be the first implementation of 3D technology. In this case, $N = 2$, and the temperature of the two die is given by

$$\begin{aligned} \theta_1 &= \frac{(Q_1(R_2 + R_{\rm pk} + R_3) + Q_2(R_3 + R_{\rm pk})) \cdot (R_1 + R_{\rm hs})}{R_{\rm hs} + R_1 + R_2 + R_3 + R_{\rm pk}} \\ \theta_2 &= \frac{(Q_1(R_{\rm hs} + R_1) + Q_2(R_1 + R_2 + R_{\rm hs})) \cdot (R_3 + R_{\rm pk})}{R_{\rm hs} + R_1 + R_2 + R_3 + R_{\rm pk}}. \end{aligned} \tag{8}$$

For face-to-face 3D interconnect technology, R_1 and R_3 are the silicon thermal resistances of *die1* and *die2*, respectively, denoted by $R_{\rm Si,1}$ and $R_{\rm Si,2}$. R_2 is the thermal resistance between the two junctions, which in this case comprises $R_{\rm b}$, the thermal resistance of the bonding layer between the two die and $2R_{\rm BEOL}$, the thermal resistance of the BEOL of the two die. Assuming $R_{\rm BEOL}$ to be the same for both die, equation (8) may be simplified to equation (9), shown at the bottom of the next page.

The temperature solution for other integration technologies like back-to-face integration may also be derived from equations (5)–(7). These solutions will differ from equation (9) in the arrangement of the silicon, interdie bond, and BEOL thermal resistances.

Note that equation (9) shows that die2 is not necessarily the hotter of the two. In fact, it is easy to show that in spite of being located farther from the heat sink, die2 is the hotter die only if

$$\frac{Q_2}{Q_1} > \frac{(R_{\rm hs} + R_{\rm Si,1})}{(R_{\rm pk} + R_{\rm Si,2})}. \tag{10}$$

Since silicon thermal resistance is usually much smaller than $R_{\rm hs}$ and $R_{\rm pk}$, the right-hand side of the inequality above reduces to $R_{\rm hs}/R_{\rm pk}$. Interestingly, equation (10) shows that a memory die may run cooler than a logic die even if the former is stacked next to the package, provided its heat dissipation is low enough. Note that the result derived in equation (10) is independent of the thermal resistance $R_{\rm b}$ between the dies.

A second interesting special case of the general solution is the one where there is no heat loss through the package. This may be relevant for high-power applications where a heat sink removes most of the heat, and the package end is conservatively assumed to be insulated. In this case, the temperature solution is given by

$$\frac{\theta_i}{R_{\rm hs}} = \sum_{j*=2}^{i} r_{j*} \cdot \left(\sum_{m=j*}^{N} Q_m \right) + (1 + r_1) \cdot \sum_{m=1}^{N} Q_m. \quad (11)$$

For a two-die case, the temperature solution is given by

$$\begin{aligned} \theta_1 &= (Q_1 + Q_2)\,(R_{\rm Si,1} + R_{\rm hs}) \\ \theta_2 &= (Q_1 + Q_2)\,(R_{\rm Si,1} + R_{\rm hs}) \\ &\quad + Q_2\,(R_{\rm b} + 2 \cdot R_{\rm BEOL})\,. \end{aligned} \quad (12)$$

In this case, the die next to the package is always the hottest die. Furthermore, note that this case is similar to a low-power application where no heat sink is provided and all heat loss occurs through the package. Equation (11) applies, with the package resistance replacing the heat sink resistance and with appropriate redefinition of nodes.

As shown in equations (5)–(7), the relative magnitudes of the various thermal resistances in the network play a key role in determining the temperature profile. Consequently, a certain type of 3D integration may be thermally feasible for one packaging technology and not so for another one. In general, the silicon thermal resistances $R_{\rm Si,1}$ and $R_{\rm Si,2}$ are much smaller than other thermal resistances, which leads to some simplification of these equations.

As an example, consider a two-die stack consisting of 300-μm-thick die with a 10 mm $\times$ 10 mm cross section. Heat sink and package thermal resistances are assumed to be 2 and 20 K/W, respectively. These values are usually available from thermal modeling of the heat sink and package. The interstrata bond layer is assumed to be 10 μm thick, with an effective thermal conductivity of 0.1 W/mK. Assuming each die dissipates 10 W, the temperature rise in the two die are determined using the model presented here to be 43.8 and 36.7 K, respectively. If instead of uniform power, *die1* (the die close to the heat sink) and *die2* dissipate 2 and 18 W, respectively, the temperature rise of *die2* increases to 50.7 K, while *die1* temperature rise reduces slightly to 35.3 K. If a finer pad pitch or reduced interface resistance results in an improvement in the interstrata bond layer thermal conductivity to 1 W/mK instead of 0.1 as considered above, the die temperature rise values are 37.4 and 36.6 K, respectively, when each die dissipates 10 W power. Note that the thermal conductance of the interstrata bond layer plays an important role in determining the die temperatures. As this conductance increases, the two die temperatures come closer.

B. Extension of Junction-to-Air Thermal Resistance Concept

The single-valued junction-to-air thermal resistance has traditionally been used to describe the thermal characteristics of a microprocessor die in package. This concept is not sufficient for a complete description of the thermal performance of a 3D IC. Due to the presence of multiple heat sources and multiple internal resistances, representing the junction-to-air thermal resistance in a matrix form is appropriate. In this framework, R_{ij} represents the temperature rise in the ith stratum per unit heat dissipation in the jth stratum. From (5)–(7), it is easy to determine R_{ij} as shown in equation (13) at the bottom of the page, where $q_{ki} = Q_k/Q_i$. In addition to the thermal resistance matrix R_{ij}, designers may also be interested in the thermal sensitivity matrix S_{ij}, which represents the change in temperature of the ith stratum due to an incremental change in heat generation in the jth stratum. S_{ij} is given by equation (14) at the bottom of the next page.

Note the subtle difference between R_{ij} and S_{ij}. While S_{ij} is independent of heat generation terms, and is thus closer in

$$\begin{aligned} \theta_1 &= \frac{(Q_1(R_{\rm b} + 2 \cdot R_{\rm BEOL} + R_{\rm pk} + R_{Si,2}) + Q_2(R_{Si,2} + R_{\rm pk})) \cdot (R_{Si,1} + R_{\rm hs})}{R_{\rm hs} + R_{Si,1} + R_{\rm b} + 2 \cdot R_{\rm BEOL} + R_{Si,2} + R_{\rm pk}} \\ \theta_2 &= \frac{(Q_1(R_{\rm hs} + R_{Si,1}) + Q_2(R_{Si,1} + R_{\rm b} + 2 \cdot R_{\rm BEOL} + R_{\rm hs})) \cdot (R_{Si,2} + R_{\rm pk})}{R_{\rm hs} + R_{Si,1} + R_{\rm b} + 2 \cdot R_{\rm BEOL} + R_{Si,2} + R_{\rm pk}} \end{aligned} \quad (9)$$

$$R_{ij} \equiv \frac{\theta_i}{Q_j} = R_{\rm hs} \left[\begin{array}{l} \displaystyle\sum_{k=1}^{i-1} q_{ki} \left[(1 - F_k) \left(1 + \sum_{m=1}^{k} r_m \right) - F_k \sum_{m=k+1}^{i} r_m \right] \\ \qquad + \displaystyle\sum_{k=i}^{N} \left[q_{ki} (1 - F_k) \left(1 + \sum_{m=1}^{i} r_m \right) \right] \end{array} \right] \quad (13)$$

nature to the traditional junction-to-air thermal resistance for a single die, R_{ij} is dependent on the heat generation itself due to coupling. Chip and package designers must be aware of the importance of both R_{ij} and S_{ij} and the complex coupling of temperature with multiple heat sources in order to come up with smart thermal design of 3D ICs.

C. Thermal Conductance of Interdie Bond Layer

The interdie bonding layer used to establish electrical communication between die offers resistance to heat flow, which has not been characterized in the past. This bonding layer typically consists of metal micropads that are bonded to each other at high temperature and under pressure [8]. The space around the bonding pads may be filled by an underfill material that aids in temporary alignment of the pads during the bonding process, although underfill-free process has also been demonstrated [8]. Depending on the nature of integration used and the power distribution among various die, the interdie bonding layer thermal resistance may play a key role in determining the maximum temperature. The thickness, material, and area density of micropads and the nature of the bonding between strata all influence the interstrata thermal resistance. In order to apply the analytical model results presented in the previous subsection, a numerical model is developed for the determination of the thermal resistance of the interdie bonding layer. Metal bond pads are assumed to be distributed uniformly on the bonding planes of both strata, and a unit cell, shown in Fig. 3, is used for simulations. The top die/wafer is assumed to have a two-metal bonding pad, while the bottom die/wafer is assumed to have a single metal bonding pad. Thermal properties of electroplated Cu and Sn are used for modeling the metal pads. The dimensions of the top and bottom bonding pads are 39 and 54 μm, respectively. The total pad thickness is around 15 μm. Outer dimensions of the unit cell are determined by the micropad area density. A constant temperature difference is imposed between the top and bottom faces of the unit cell. Using finite element simulations, the resultant heat flux between the two faces, and the effective thermal resistance of the bond layer is hence determined.

Fig. 4(a) shows isothermal contours in a cross section of the model geometry. Fig. 4(b) shows the effective thermal resistance for the metal micropad layer for a typical 10 mm $\times$ 10 mm die footprint as a function of the area density of bond pads as well as the thermal conductivity of the filler material. As expected, the effective thermal resistance reduces with a higher metal density and with higher thermal conductivity of the filler material. Even when the filler material thermal conductivity is as poor as that of air, the effective thermal resistance is quite low due to heat conductance through the metal features. The thermal

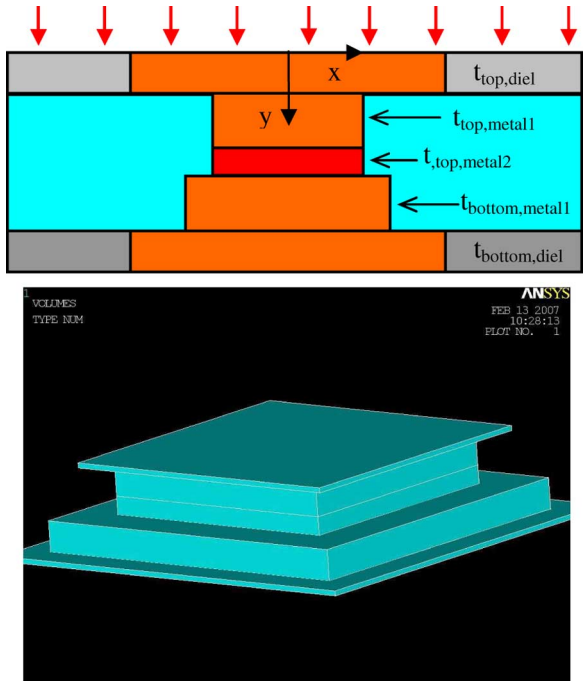

Fig. 3. Schematic of the unit cell used for computing thermal resistance of the interdie bonding layer.

resistance of the back-end metal-dielectric stack certainly dominates over that of the micropad layer. Furthermore, most of the temperature rise still occurs external to the die stack. Since metal micropads do not occupy space on the device layer, it should be possible to cover a large fraction of the die face by "thermal" micropads in order to ensure good thermal and mechanical contact. Improvement in the thermal performance of the interdie bonding layer is not expected to provide much benefit due to the dominance of the heat sink and package thermal resistance.

The next section discusses results obtained from the application of the analytical model developed in this section to a typical multidie stack. In addition, numerical simulation results accounting for heat spreading and comparing 3D technology with the system-in-package (SiP) technology are presented.

III. Results and Discussion

The analytical model presented in the previous section may be used to obtain several practical thermal design guidelines for 3D electronics. These include the optimal arrangement of various strata, the influence of internal thermal resistances on the maximum temperature, the heat sink and package thermal requirements for 3D integration, the thermal feasibility limits of 3D integration, etc. In this section, a generic two-die stack is

$$S_{ij} \equiv \frac{\partial \theta_i}{\partial Q_j} = \begin{cases} R_{\mathrm{hs}}(1-F_j)\left(1+\sum_{m=1}^{i} r_m\right) & i \le j \\ R_{\mathrm{hs}}(1-F_j)\left(1+\sum_{m=1}^{j} r_m\right) - F_j \sum_{m=j+1}^{i} r_m & i > j \end{cases} \tag{14}$$

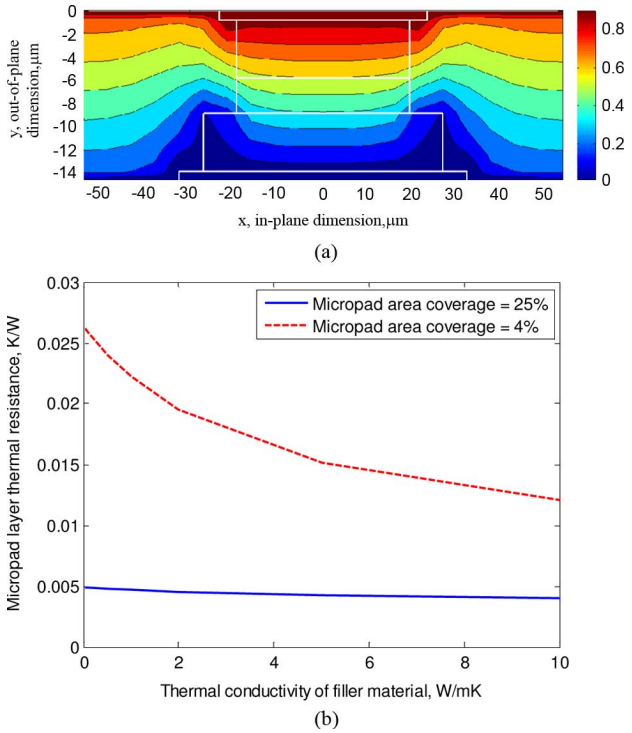

Fig. 4. (a) Isothermal contours in the cross section of the micropad geometry. (b) Variation of the effective thermal resistance of the metal micropad layer for a typical 10 mm × 10 mm die footprint as a function of the thermal conductivity of filler material and the micropad area density.

considered, though the work is easily extendable to stacks with more than two die. For simplicity, the die with the higher power dissipation is referred to as the logic die, and the other die is referred to as the memory die.

Fig. 5 shows the dependence of temperature of the logic die on thermal resistance of the interdie bond layer. As expected, the die temperature exhibits only weak dependence due to the small value of the interdie bond thermal resistance compared to other resistances in the circuit. This effect is prominent regardless of whether logic or memory die is placed closer to the heat sink. Fig. 5 demonstrates that not much thermal benefit may be obtained by improving the thermal resistance of the interstrata bond layer. The influence of the silicon die thermal resistance on the maximum temperatures is also similarly limited. As a result, die thinning, an important process technology that enables 3D integration, does not have much thermal benefit. Die thinning may, in fact, increase hotspot temperature by reducing the heat spreading effect of silicon substrate. It is important to note that metal-filled TSVs and metal contact pads may have a role to play in localized heat dissipation, as discussed in recent papers [11], [12].

Fig. 6 shows the improvement required in the heat sink thermal resistance in order to incorporate an extra die dissipating a given amount of power while maintaining the same maximum temperature. For a memory die dissipating around 10% of the power dissipated by the logic die, an improvement of around 10% in the thermal performance of the heat sink will be necessary to maintain the same maximum temperature. The required improvement is only weakly dependent on the interstrata bond layer thermal resistance, especially if one of the

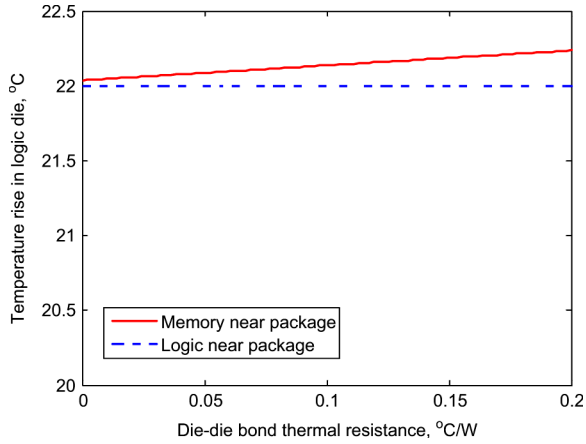

Fig. 5. Logic die temperature as a function of thermal resistance of interdie bond layer. Memory die is assumed to dissipate 10% of the logic die power. A very weak dependence is observed regardless of which die is stacked next to the package.

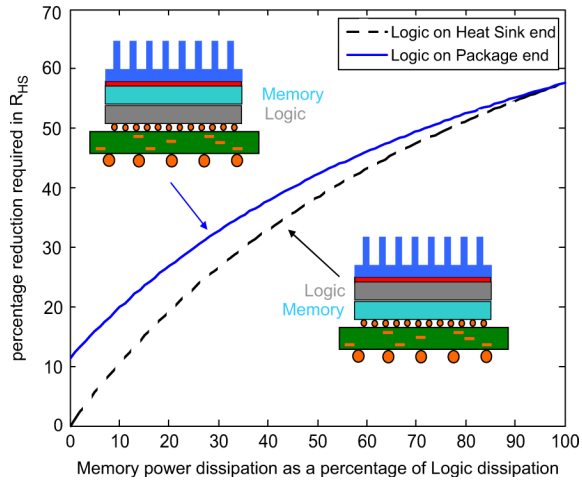

Fig. 6. A plot showing the improvement required in heat sink thermal resistance as a function of power dissipation in the memory die. The baseline case is a single logic die. Note the dependence on the stacking sequence.

die dissipates much lesser power than the other. Fig. 6 provides thermal design guidelines for heat sink design for 3D chips and provides the limits of thermal feasibility of 3D integration.

Another parameter of interest is the power reduction required to facilitate two-die integration in the same package and heat sink used for a single die. For a test case with a memory die dissipating 10% of the power of a logic die, calculations using equation (9) show that the logic die power must be reduced by 16% in order to keep a logic-on-memory die stack at the same maximum temperature as a single logic die, assuming that the package and heat sink remain the same. Similar calculations based on equation (9) can be used to understand the effect of relative changes in the heat sink and package thermal resistances.

Integration of multiple heterogeneous components (for example, memory and logic) inside the same package is common at present. 3D technology enables this integration to occur normal to the package plane, thus improving form factor. The stacking of multiple die in 3D however leads to a different thermal problem that needs to be fully understood before 3D integration may be implemented. When multiple die are

TABLE I
COMPARISON OF THERMAL PERFORMANCE OF LOGIC AND MEMORY DIE IN VARIOUS INTEGRATION SCENARIOS, INCLUDING BASELINE SIP AND VARIOUS 3D INTEGRATION OPTIONS

Scenario	T_Logic (°C)	T_Memory (°C)
(A) Baseline 2D SiP	109.0	98.3
(B.1) 3D with same lid, same power	109.6	109.1
(B.2) 3D with same lid, 7% reduced main power	106.1	105.4
(C.1) 3D with smaller lid, same power	113.0	112.3
(C.2) 3D with smaller lid, 7% reduced main power	109.1	108.4

stacked, a reduction of die footprint results due to the presence of more than one silicon plane on which to place circuits.

While the exact amount of reduction depends on architectural optimization, one may assume for thermal analysis that the total circuit area on the two die stack remains the same as the original two-dimensional (2D) die. Finite element simulations are carried out in order to compare a SiP with logic and memory die with equivalent stacked two-die implementation, where the logic and memory blocks are partitioned and placed on different die. In both cases, the same package and heat sink are used. The effects of reducing the die footprint and changing the lid size over the 3D stack are investigated. The temperature rise in each block is computed as a function of whether the package lid also scales down with the die footprint or not. In addition, simulations are also run to account for power reduction in the global wires due to reduced wire length. Calculations have shown that for typical architecture configurations, the global wire power can be reduced by as much as 30%. Depending on what fraction the global wire power is of the total power consumption, this could represent significant overall power saving. Table I shows the temperature rise in logic and memory for the various cases. While the scaling down of the die footprint results in an increase in maximum temperature, this effect diminishes when the lid size remains the same as the original 2D die. This shows that the lid footprint plays a far more critical role in determining the maximum temperature than the die footprint. While 3D integration by itself does not result in improved thermal performance, it may be possible that, as shown in Table I, the reduction in power consumption due to reduced global wire lengths may result in a 3D system with similar thermal performance as a 2D die.

Note that the memory temperature increases significantly in the 3D implementation even when the lid is not scaled down. This is because the memory die is much closer to the higher heat dissipating logic die in the 3D stack compared to the 2D SiP. In cases where the memory is rated to run at a lower temperature than logic, this may be a significant thermal problem.

Reduction of die footprint is also being driven by other factors, including technology scaling and yield issues. The introduction of 3D technology may accelerate this shrinkage, and thus require new, innovative packaging and cooling solutions.

Finally, it is to be noted that the tradeoff of electrical and thermal considerations is quite inherent in the design of 3D ICs. For example, while it is thermally optimal to place the higher power dissipation logic die that dissipates more power closer to the heat sink rather than the package, this approach is likely to pose greater integration challenges since the logic die typically requires more connections to the package, and thus more TSVs will be needed to be routed through the memory die in order to reach the logic die. As another example, it is electrically advantageous to stack a core block on top of another core block due to the shorter wire lengths obtained. However, this approach results in significant thermal challenge due to the increased local power density. This inherent tradeoff underlines the need for early incorporation of thermal models in design and layout tools so that electrical benefits offered by 3D technology may be fully taken advantage of without significant exacerbation of the thermal management challenge.

IV. CONCLUSION

While 3D ICs offer several electrical benefits, the associated thermal challenges need to be fully understood. The current work offers a framework in which to analyze the thermal performance of 3D ICs through an analytical model for predicting temperature rise in a multidie, multisource thermal resistor network. The model predicts the sensitivity of the maximum temperature on various design parameters such as power distribution among various die, intrastack thermal resistances, etc. The traditional understanding of thermal dissipation in microelectronics in terms of a single-valued junction-to-air thermal resistance is extended to a thermal resistance matrix due to the presence of multiple sources. Numerical simulations performed in this work quantify the thermal performance of the interdie bonding layer and show that reduction in the die footprint area due to vertical integration causes an increase in maximum temperature.

While the 1D heat transfer model presented in this work enhances the understanding of heat transfer issues in 3D microelectronics, there is need for future work in several directions. The effect of localized heating is not captured in the current model. The role of TSVs in heat dissipation from localized hotspots needs to be further investigated. Also, heat transfer issues arising from the integration of unequally sized die also need to be fully understood. Finally, due to the strong coupling between die-level electrical design and thermal management in 3D ICs, reconciliation of thermal and electrical design constraints and objectives is very important in early design stage of 3D ICs. Thermal and electrical co-design of 3D ICs may help alleviate late-stage thermal management problems by thermal-friendly floorplanning. This will require the incorporation of thermal design models like the ones presented here in electrical design and layout tools.

ACKNOWLEDGMENT

Useful discussions with Z. Huang, K. Ramakrishna, B. Joiner, and S. Neelakantan are gratefully acknowledged.

REFERENCES

[1] A. Jain and S. Ramanathan, "Theoretical investigation of sub-ambient on-chip microprocessor cooling," in *Proc. IEEE ITherm*, 2006, pp. 765–770.

[2] K. Banerjee, S. J. Souri, P. Kapur, and K. C. Saraswat, "3-D ICs: A novel chip design for improving deep-submicrometer interconnect performance and systems-on-chip integration," *Proc. IEEE*, vol. 89, pp. 602–633, 2001.
[3] S. List, C. Webb, and S. Kim, "3D wafer stacking technology," in *Proc. Adv. Metallization Conf.*, 2002, pp. 29–36.
[4] R. S. Patti, "Three-dimensional integrated circuits and the future of system-on-chip designs," *Proc. IEEE*, vol. 94, pp. 1214–1224, 2006.
[5] S. M. Alam, R. E. Jones, S. Pozder, R. Chatterjee, and A. Jain, "New design considerations for cost effective three-dimensional (3D) system integration," *IEEE Trans. VLSI Syst.*, 2009, to be published.
[6] G. H. Loh, Y. Xie, and B. Black, "Processor design in 3D die-stacking technologies," *IEEE Micro Mag.*, vol. 27, no. 3, pp. 31–48, 2007.
[7] P. R. Morrow, C.-M. Park, S. Ramanathan, M. J. Kobrinsky, and M. Harmes, "Three-dimensional wafer stacking via Cu-Cu bonding integrated with 65-nm strained-Si/low-k CMOS technology," *IEEE Electron Dev. Lett.*, vol. 27, no. 5, pp. 335–337, May 2006.
[8] S. Pozder, R. Chatterjee, A. Jain, Z. Huang, R. E. Jones, and E. Acosta, "Progress of 3D integration technologies and 3D interconnects," in *Proc. IEEE Int. Interconnect Tech. Conf.*, 2007, pp. 213–215.
[9] K. W. Guarini, A. T. Topol, M. Ieong, R. Yu, L. Shi, M. R. Newport, D. J. Frank, D. V. Singh, G. M. Cohen, S. V. Nitta, D. C. Boyd, P. A. O'Neil, S. L. Tempest, H. B. Pogge, S. Purushothaman, and W. E. Haensch, "Electrical integrity of state-of-the-Art $0.13\ \mu\mathrm{m}$ SOI CMOS devices and circuits transferred for three-dimensional (3D) integrated circuit (IC) fabrication," *IEDM Tech. Dig.*, pp. 943–945, 2002.
[10] R. Mahajan, "Thermal management of CPUs: A perspective on trends, needs and opportunities," presented at the Keynote Presentation at the 8th Int. Workshop on Thermal Investigations of ICs and Syst., 2002.
[11] B. Goplen and S. S. Sapatnekar, "Placement of thermal vias in 3D ICs using various thermal objectives," *IEEE Trans. Comput.-Aided Design of Integr. Circuits Syst.*, vol. 26, no. 4, pp. 692–709, Apr. 2006.
[12] J. Cong and Y. Zhang, "Thermal via planning for 3-D ICs," in *Proc. IEEE/ACM Intl. Conf. Comput. Aided Design*, 2005, pp. 745–752.
[13] M. B. Kleiner, S. A. Kuhn, P. Ramm, and W. Weber, "Thermal analysis of vertically integrated circuits," *IEDM Tech. Dig.*, pp. 487–490, 1995.
[14] T.-Y. Chiang, S. J. Souri, C. O. Choi, and K. C. Saraswat, "Thermal analysis of heterogeneous 3-D ICs with various integration schemes," in *Proc. IEEE Electron Devices Meeting*, 2001, pp. 681–684.
[15] K. Puttaswamy and G. H. Loh, "Thermal analysis of a 3D die-stacked high-performance microprocessor," in *Proc. ACM/IEEE Great Lakes Symp. VLSI*, 2006, pp. 19–24.
[16] B. Black *et al.*, "Die stacking (3D) microarchitecture," in *Proc. IEEE/ACM Int. Symp. Microarchit.*, 2006, pp. 469–479.
[17] T. Brunschwiler, B. Michel, H. Rothuizen, U. Kloter, B. Wunderle, H. Oppermann, and H. Reichl, "Forced convective interlayer cooling in vertically integrated packages," in *Proc. IEEE ITherm*, 2008, pp. 1114–1125.
[18] V. Natarajan, "Convective heat transfer from a die-stacked electronic package," in *Proc. IEEE ITherm*, 2008, pp. 1132–1138.
[19] J. Torresola, C.-P. Chiu, G. Chrysler, D. Grannes, R. Mahajan, R. Prasher, and A. Watwe, "Density factor approach to representing impact of die power maps on thermal management," *IEEE Trans. Adv. Packag.*, vol. 28, no. 4, pp. 659–664, 2005.

Ankur Jain received the B.Tech. degree in mechanical engineering from the Indian Institute of Technology (IIT), Delhi, India, in 2001, and the M.S. and Ph.D. degrees in mechanical engineering from Stanford University, Stanford, CA, in 2003 and 2007, respectively.

Following the completion of the Ph.D. degree, he worked as a Senior Staff Scientist at Freescale Semiconductor, Austin, TX. In 2009, he joined Molecular Imprints, Inc., Austin, TX, as a Senior Research Engineer. His research interests include thermal management of microelectronics and MEMS, thermal-electrical optimization of microelectronics, and advanced microelectronics technologies including 3D integrated circuits.

Robert E. Jones (M'02) received the B.A. in physics from Kansas State University of Pittsburg in 1965 and the Ph.D. degree in physics from Iowa State University, Ames, in 1971.

He taught for a year in Germany before becoming a Research Associate from 1972 to 1975 at Lakehead University, Thunder Bay, ON, Canada, where he later joined the faculty from 1975 to 1978. In 1978, he joined the faculty of the Physics Department at the University of Colorado at Colorado Springs. In 1984, he joined Inmos Corporation, Colorado Springs, CO, where he led interconnect development. In 1988, he joined Motorola SPS, Austin, TX, which became Freescale Semiconductor, where he is a Distinguished Member of the Technical Staff. He has 35 issued U.S. patents and over 100 publications in journals and conference proceedings. His research has focused on new materials technologies including silicides, ferroelectric memories, high-k dielectrics, quantum dot memories, stacked chips, and integrated photodetectors.

Dr. Jones is a Fellow of the American Physical Society and a Member of the Materials Research Society.

Ritwik Chatterjee received the B.S. degree in electrical engineering from Purdue University, West Lafayette, IN, in 1996, and the M.S. and Ph.D. degrees in electrical engineering from the Massachusetts Institute of Technology (MIT), Cambridge, in 1998 and 2002, respectively.

His M.S. research was in the area of linearity of power AlGaAs/GaAs heterojunction bipolar transistors, and his doctoral research was on the use of unsaturated fluorocarbons for dielectric etch applications. He was a MIT Martin Fellow of Environmental Sustainability and an AMD/SRC Fellow. After completing graduate school, he joined Freescale Semiconductor in Austin, TX, in 2002 as an Integration Engineer working on advanced BEOL module development and 3D interconnect integration. In 2008, he joined the Packaging Research Center at Georgia Institute of Technology, Atlanta, as the Program Director for the center's new industry-academia global consortium on 3D All Silicon System Module. He has authored or co-authored 30 publications and has 10 patents pending.

Scott Pozder received the B.S. degree in physics from Montana State University, Bozeman, and the M.S. degree in material science from Colorado School of Mines, Golden.

In his 13 years at Freescale Semiconductor (formerly Motorola), Austin, TX, he has worked on projects in the areas of metal silicide, Cu metal interconnects and the evaluation of package-level reliability of low k dielectric materials. From 2002, he has been a key contributor in the areas of 3D stacked die-to- wafer and wafer-to-wafer bonding projects at Freescale. He has authored papers on 3D stacked strata structures and holds 14 U.S. patents.