

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Three-Dimensional Integrated Circuits Design for Thousand-Core Processors: From Aspect of Thermal Management

Chiao-Ling Lung^{1,2}, Jui-Hung Chien², Yung-Fa Chou²,
Ding-Ming Kwai² and Shih-Chieh Chang¹

¹*Department of Computer Science, National Tsing Hua University*

²*Information and Communications Research Laboratories,
Industrial Technology Research Institute
Taiwan*

1. Introduction

As the performance of a processing system is to be significantly enhanced, on-chip many-core architecture plays an indispensable role. Since there are fast growing numbers of transistors on the chips, two-dimensional topologies face challenges of significant increases in interconnection delay and power consumption (Hennessy & Patterson, 2007; Kurd et al., 2001). Explorations of a suitable three-dimensional integrated circuit (3D IC) with through-silicon via (TSV) to realize a large number of processing units and highly dense interconnects certainly attracts a lot of attention. However, the combination of processors, memories, and/or sensors in a stacked die leads to the cooling problem in a tottering situation (Tiwari et al., 1998). One solution to overcome the obstacles and continue the performance scaling while still is to integrate on chip many cores and their communication network (Beigne, 2008; Yu & Baas, 2006). Through concerted processors, routers, and links, the network-on-chip (NoC) provides the advantages of low power dissipation and abundance of connectivity. Moreover, because of the widespread uses of radio frequency (RF), micro-electro-mechanical systems (MEMS) (Lu, 2009), and various sensors in mobile applications, proposals of three-dimensional integrated circuit (3D IC) with through silicon via (TSV) implementations in a layered architecture have been reported (Lee, 1992; Tsai & Kang, 2000). For interconnection scalability from layer to layer, 3D fabrics are a necessity. Consequently, a thermal solution which has a high heat removing rate seems unavoidable. Since there are fast growing numbers of transistors on the chips, two-dimensional topologies face challenges of significant increases in wire delay and power consumption. The two factors are often regarded as the primary limitations for current processor architectures (Hennessy & Patterson, 2007; Kurd et al., 2001; Tiwari et al., 1998).

On the other hand, the high packing density of the stacked dies also hampers the heat dissipation of the NoC system. Thermal issues arise from increasing dynamic power losses which in turn raise the temperature. Thermal and power constraints are of great concern with 3D IC since die stacking can dramatically increase power density, if hotspots overlap each other, and additional dies are farther away from the heat sink.

Thermal-aware floorplanning is the key in which the inter-layer interconnection plays a role more than just signal transmission or power delivery. Figure 1 depicts the usage of thermal TSV to alleviate the heat accumulation, which is brought from that used in printed circuit boards (PCBs) (Lee et al., 1992). For 3D ICs, the problems of high power/thermal density can be more serious than that in the planar form. Thus, the thermal TSVs become essential for heat dissipation. Of particular interest is the design of an efficient heat transferring path. Some recent works discussed the placement of thermal TSVs. However, not only the routing but also the floorplan may need to be changed substantially after the thermal TSVs are inserted (Tsai & Kang, 2000). This leads to long iterations. Further, as the circuit complexity is increased, to insert the thermal TSVs without largely changing the floorplan is an important technology to be developed (Tsui et al., 2003). In order to keep the original routing and floorplan as much as possible, the temperature-driven design should be brought in early phases of the design procedure.

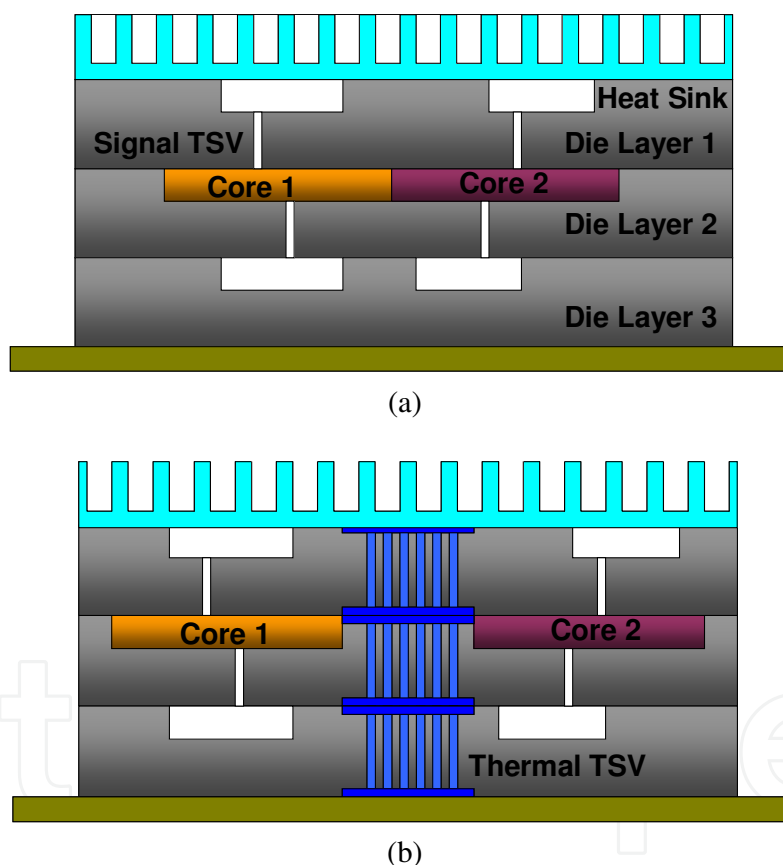


Fig. 1. 3D IC implementations of a multiprocessor system-on-chip (MP-SoC) with (a) a traditional structure and (b) with the insertion of thermal ridges.

2. Design and theoretical analysis of on-chip thermal ridge

2.1 Theoretical analysis

The thermal TSVs are intended to be placed in the inter-CG whitespace, which is called a thermal ridge. In this section, we derive analytical expressions for some key parameters.

2.1.1 Analytical model of the thermal ridge

At the transient state, the heat conduction can be described by the following equation

$$\frac{\partial}{\partial x} k_{xx} \frac{\partial T}{\partial x} + \frac{\partial}{\partial y} k_{yy} \frac{\partial T}{\partial y} + \frac{\partial}{\partial z} k_{zz} \frac{\partial T}{\partial z} + g = \rho C \frac{\partial T}{\partial \theta} \quad (1)$$

where T is the temperature, g is the heat generation rate in W/cm^2 , ρ is the density of the material, C is the thermal capacity of the material, θ is time, and k is the thermal conductivity of the material. This fundamental thermal conduction equation describes that the temperature transmitting through the thermal volume depends on time θ and directional thermal conductivities k_{xx} , k_{yy} , and k_{zz} (Chieh et al., 2010; Lung et al., 2010). The boundary conditions of the top and bottom surfaces of the chip are *adiabatic* and those of the surrounding surfaces are *convective*.

For dissipating the heat into the substrate homogeneously, the inter-core-group thermal ridges are aligned orthogonally in column and in row. The temperature prediction of the many-core system is performed by utilizing CFD-RC which is commercial thermal and fluidic temperature simulation software. However, in order to illustrate the physical phenomenon more intuitively, a simplified one-dimensional conduction equation without taking the transient into consideration is utilized.

$$\frac{\partial}{\partial x} k_{xx} \frac{\partial T}{\partial x} = -g \quad (2)$$

The heat removing rate of the thermal ridge is assumed to be q . Let us consider two CGs. The temperature distribution between CG1 and CG2 can be expressed by

$$T = T_1 + \frac{T_2 - T_1}{w} x - \frac{q}{2k_s} [w - x] x \quad (3)$$

where T_1 and T_2 are the temperatures of CG1 and CG2, respectively, q is the heat conducted to the ambient environment by the thermal ridge, k_s is the equivalent thermal conductivity of the thermal ridge, and w is the width of the thermal ridge. Since T denotes the temperature at the location x , examining the mid-point $T_{1/2}$ by substituting x with $w/2$ into (3), we have

$$w = \left(\frac{8k_s}{q} \right)^{1/2} \left[\frac{T_1 + T_2}{2} - T_{1/2} \right]^{1/2} \quad (4)$$

From (4), it is easy to see that if the mid-point temperature $T_{1/2}$ is targeted to be lower, w needs to be larger.

2.1.2 Effective thermal conductivity of the thermal ridge

The equivalent thermal conductivity k_{szz} of a thermal ridge is decided by the density of the thermal TSVs in the thermal ridge (Chieh et al., 2010; Lung et al., 2010). To determine k_{szz} , the effective thermal conductivity should be taken into account and described as the following equation:

$$k_{szz} = d \cdot k_{emb} + (1 - d)k_{sub} \quad (5)$$

where k_{emb} is the equivalent thermal conductivity of the thermal TSVs, k_{sub} is the thermal conductivity of the silicon substrate, d is the percent contribution of the thermal TSVs in the thermal ridge. Since the orientation of the thermal TSV is longitudinal along the z direction, this effective thermal conductivity cannot be applied to the lateral heat transfer computation. For x and y directional heat transfer, the thermal conductivity should be applied by the following equation.

$$k_{sxx} = (1 - \sqrt{m})k_{sub} + \frac{\frac{\sqrt{m}}{1 - \sqrt{m}}}{\frac{1}{k_{sub}} + \frac{\sqrt{m}}{k_{emb}}} = k_{syy} \quad (6)$$

where m is the percent contribution of the metal lines for thermal conduction in the silicon substrate. In general, the vertical thermal conductivity k_{szz} is much larger than the lateral thermal conductivities k_{sxx} and k_{syy} . By (5) and (6), we can clearly figure out that k_{sxx} is around 10 W/mK and k_{szz} is around 120 W/mK. Thus, the heat flows through the thermal ridge almost dissipates by the heat sink instead of transferring laterally. By substituting the equivalent k_s and the temperature values of T_1 , T_2 and $T_{1/2}$ into (3), we obtained that the widths of the thermal ridge should be $200 \mu\text{m} \sim 400 \mu\text{m}$.

2.2 Design parameters and assumptions

Here, we focus on a mesh-connected NoC with 1,024 cores. A globally asynchronous, locally synchronous (GALS) digital-signal processor (DSP) design is adopted (Tran et al., 2009a, 2009b; Truong et al., 2008). Each DSP, constituting a tile, is composed of a core with an on-chip oscillator for its own clocking and a switch with associated buffers, as shown in Figure 2. The tile allows repetitive, mirrored layout, occupying an area of 0.168 mm^2 ($410 \mu\text{m} \times 410 \mu\text{m}$) (Tran et al., 2009a, 2009b). Consider a simple power map with two major sources in the tile. One is attributed to the computation and the other to the communication. Correspondingly, the average power consumption at the active status is broken down to 17.6 mW and 1.1 mW, respectively (Tran et al., 2009a, 2009b).

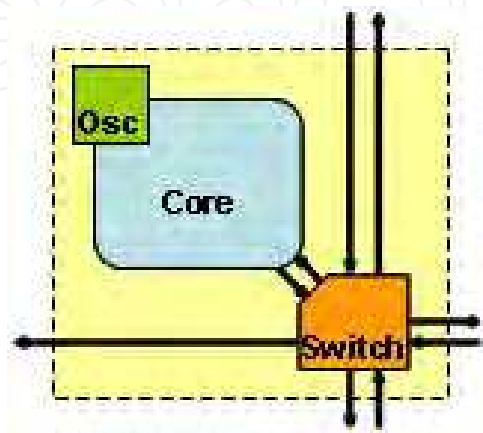


Fig. 2. The DSP element for a GALS many-core system.

The cores are arranged as a 32×32 square mesh. Since the international technology roadmap for semiconductor (ITRS) predicts that the maximum chip size will maintain similar dimensions, we assume $20\text{ mm} \times 20\text{ mm}$ as our upper bound. Under such a constraint, the remaining area not occupied by the tiles is the input/output and peripheral circuits. The total power consumption of the chip is around 20 W , which leads to the average power density of 5 W/cm^2 . Since ITRS also predicts the power density is reasonable up to the level of 100 W/cm^2 , the power density assumed in this chapter is a probable value (Brunschwiler et al., 2009; Xu et al., 2004).

In this chapter, we assumed that there are three layers of the die stack and the many-core NoC is sandwiched in the middle. As mentioned earlier, a commercial tool based on finite element method (FEM) is used. The three-dimensional model of the NoC is created with the widely used package model, in a fashion similar to that shown in Figure 1. However, the heat sink is not modelled and analyzed in our case. Instead, it is simplified to a heat loss, and a proper heat transfer coefficient is applied to the boundary condition on the top surface where the heat sink would have been located originally.

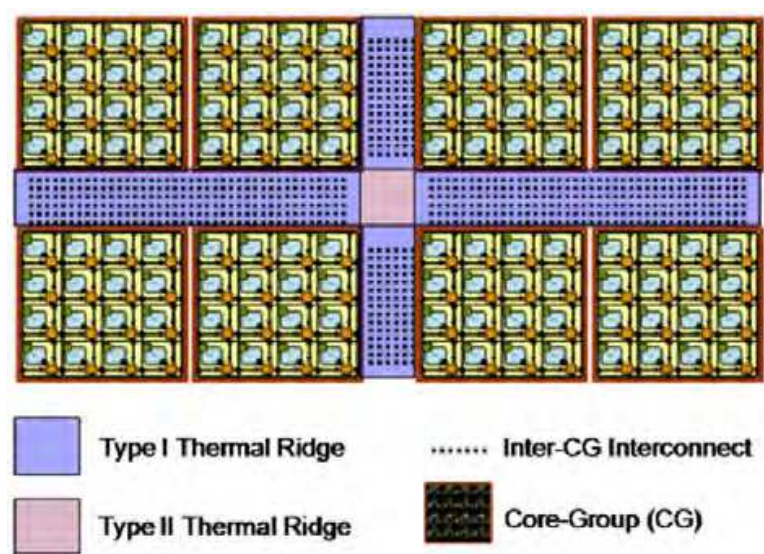


Fig. 3. Insertion of type I and type II thermal ridges into the NoC.

First, the 1,024 cores are divided into 8×8 CGs, each CG consisting of 4×4 cores. As shown in Figure 3, thermal ridges are inserted between the hottest CGs. By the locations where they are inserted, the thermal ridges can be categorized into two types. The type-I thermal ridge has a low density of thermal TSVs and the type-II thermal ridge has a high density of thermal TSVs. This is because the type-I thermal ridge is located between two CGs in which their routing dominates the most of the silicon area, even after the expansion to gain more whitespace. On the other hand, the type-II thermal ridge lies in the intersectional area having no wires passing through, and therefore, a large quantity of thermal TSVs can be planted.

The physical effect of the thermal ridge can be illustrated by using the electrical lumped model as shown in Figure 4. By the duality between electrical and thermal models, the temperature T is substituted by a voltage V , the power P is substituted by a current I , and the thermal resistance R by definition is proportional to the reciprocal of thermal

conductivity k_s . The availability of the thermal ridge can be modelled by the equivalent circuits as follows.

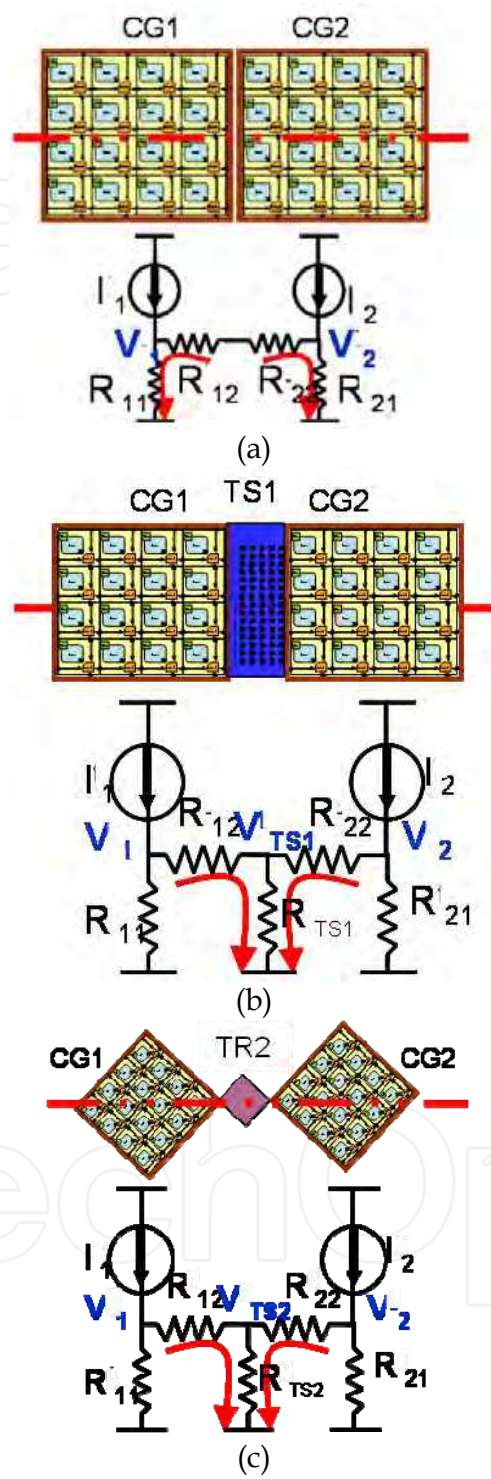


Fig. 4. Resistive thermal models of two adjacent CGs inserted with (a) no thermal ridge, (b) a type-I thermal ridge, and (c) a type-II thermal ridge.

Figure 4(a) shows the case when there is no thermal ridge between CG1 and CG2. It is clear in the schematic that no extra conduction path has been added to the ground. Since the

vertical thermal resistance R_{11} (R_{21}) is much larger than the lateral thermal resistance R_{12} (R_{22}), the voltage V_1 (V_2) keeps at a high value. Figure 4(b) shows the case when a type-I thermal ridge is inserted between CG1 and CG2. Another conduction path is added through the thermal resistance R_{TS1} . As aforementioned, R_{TS1} is inversely proportional to k_s . As long as k_s is much larger than the thermal conductivity k_{sub} of the silicon substrate, R_{TS1} is much smaller than R_{11} (R_{21}); the current I_1 (I_2) goes mostly through R_{TS1} , rather than R_{11} (R_{21}). In addition, by voltage division, V_{TS1} is obviously lower than V_1 (or V_2). In other words, the temperature of the type-I thermal ridge is definitely lower than the temperature of CG1 or CG2. Figure 4(c) shows the case when a type-II thermal ridge is inserted at the intersectional area between the CGs to remove more heat. The value of R_{TS2} depends on that of k_s . Since the thermal TSVs are densely planted on the type-II thermal ridge, R_{TS2} is much smaller than R_{11} (or R_{21}). Compared with CG1 and CG2, the type-II thermal ridge, which has a lower temperature, is designed to be an on-chip heat sink.

2.2.1 Rotation of the hotspots

To verify the feasibility of the proposed scheme for thermal-aware floorplanning, we obtain the temperature distribution of the basic CG first. There are 4×4 cores within a CG as shown in Figure 5. The cores are homogenous, with the hotspot near the lower right corner. It is clear that since the hotspot is not located at the center of the core, when assembled into the CG, the temperature distribution is asymmetric.

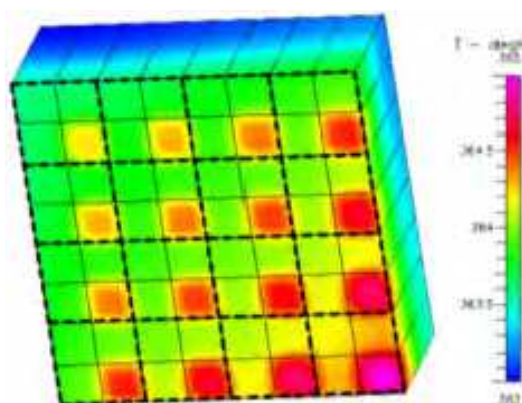


Fig. 5. Temperature distribution of the 16-core CG.

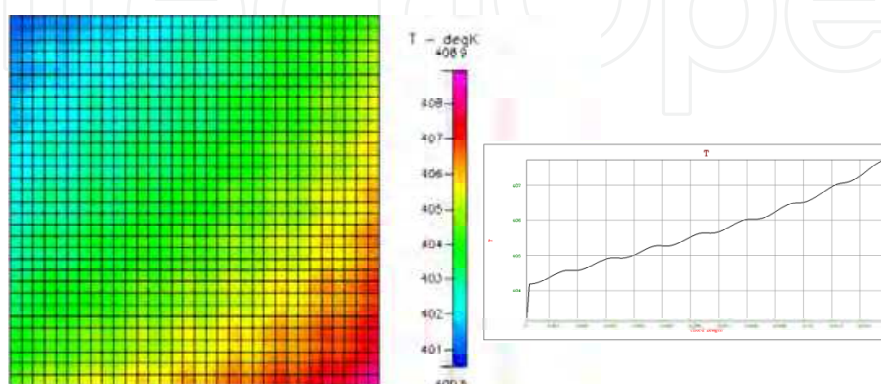


Fig. 6. Temperature distribution of the 1,024-core NoC with the same orientation of each core.

However, the situation becomes worse, when 64 such CGs are put together to construct the 1,024-core NoC. Figure 6 shows a typical layout in which the orientation of each core is kept the same as in the Figure 5, with the hotspot near the lower right corner. Apparently, the design maintains regularity in connectivity with the same routing distance between cores, but unfortunately, it is not thermal-aware. The temperature distribution is still asymmetric and the maximum temperature of the whole chip now rises up to 408.9 K which requires a heat sink. The lack of symmetry leads to that the heat sink cannot be placed at a simple orientation with equal heat dissipation ability.

Let us define the temperature non-uniformity as follows:

$$U = \frac{\Delta T}{\Delta x} \quad (7)$$

where ΔT is temperature difference and Δx is distance between any two points on the single core. Hence, it represents the slope of the temperature gradient per unit length. Clearly, the bigger the value of U , the more severe the temperature difference between neighboring cores. In the case of Figure 6 the maximum U is around 4.1 K/cm the averaged U is around 3.1 K/cm.

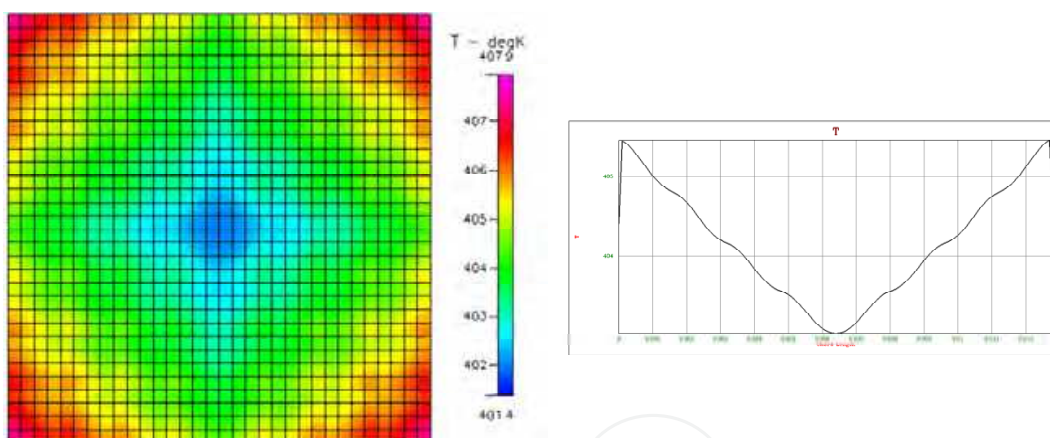


Fig. 7. Temperature distribution of the 1,024-core NoC with the orientation of every quarter of CGs rotated 90 degree.

To mitigate the non-uniformity, we may try to rotate either the cores in the CG or the CGs so as to align the temperature profile symmetrically (Xu et al., 2006). Figure 7 shows the latter approach by dividing the CGs into four quadrants, keeping the orientation of the second quadrant, and rotating the other three quadrants of the CGs to the upper left, upper right, and lower left corners, respectively.

To compare with those attained in Figure 6, the maximum temperature decreases 1 K, but the averaged temperature non-uniformity increases to 3.8 K/cm. If we rotate the cores in the CG in a similar fashion and then assemble such CGs, the result is not much different and hence is not shown here. This illustrates the fact that the rotation of the hotspots cannot reduce the maximum temperature effectively.

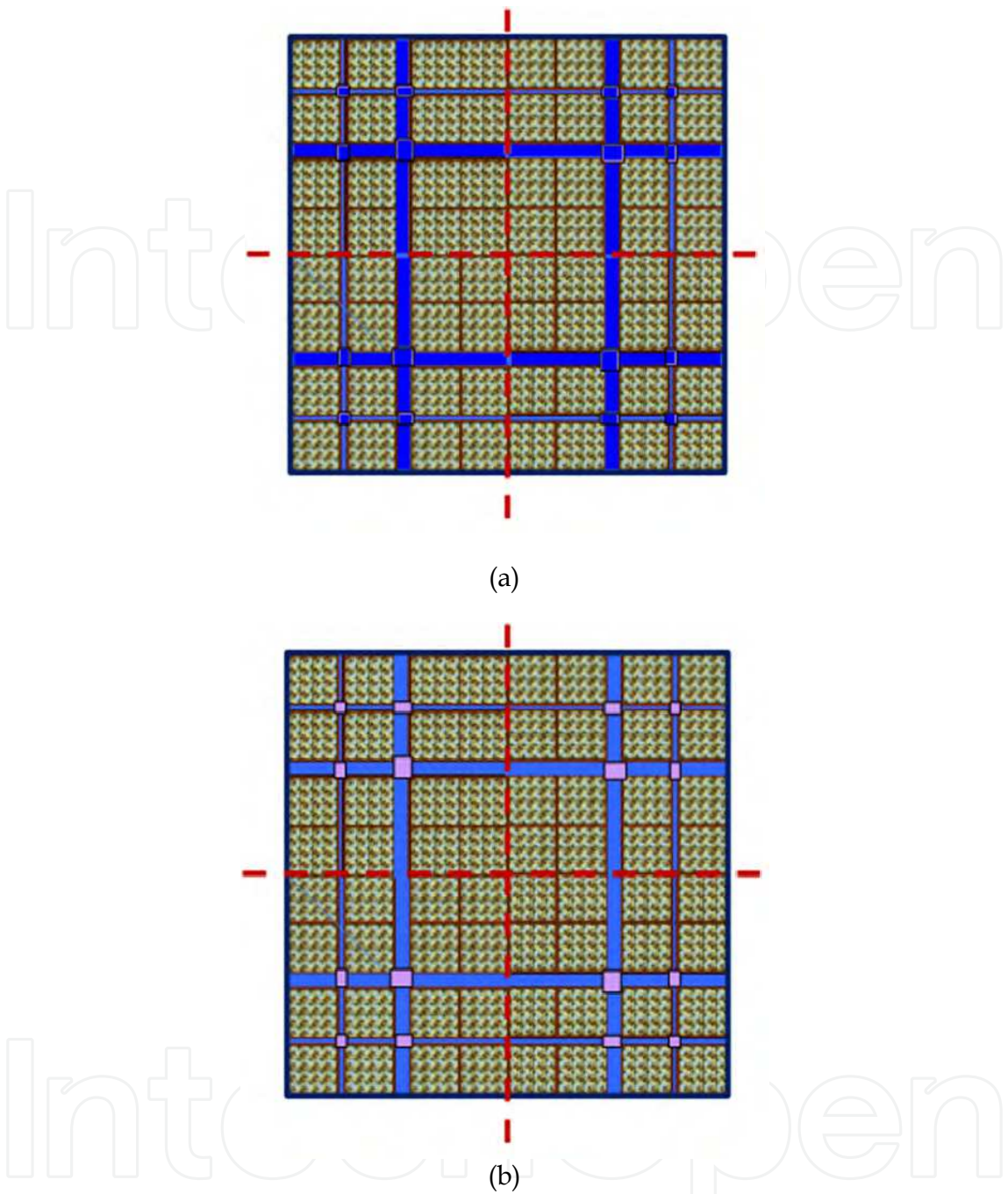


Fig. 8. The insertion places of thermal ridges. (a) Type I only. (b) Type I and Type II.

2.2.2 Insertion of the thermal ridges

The primary objective of the thermal ridges is to reduce the maximum temperature and the temperature non-uniformity at the same time. The thermal ridges are introduced into the design, with the required extra space under the constraint of manufacturing cost. In our case, at most 20% of the chip area is allowed for the thermal ridges and their locations are depicted in the Figure 8. Straits with widths of 400 μm and 200 μm are created by expanding the routing distances between CGs.

2.3 Simulation results of the proposed scheme

First, the type-I thermal ridges are inserted into the straits, except for their intersectional areas as shown in Figure 8(a). The resulting temperature distribution is shown in Figure 9. The maximum temperature is 373.4 K, which occurs in the center of the chip. To compare with the previous solutions, the maximum temperature significantly decreases 35 K by using the thermal ridges. The temperature difference at the center of the chip is about 32 K. Also, the thermal map changes a lot, since the thermal ridges are distributed in the suburb areas.

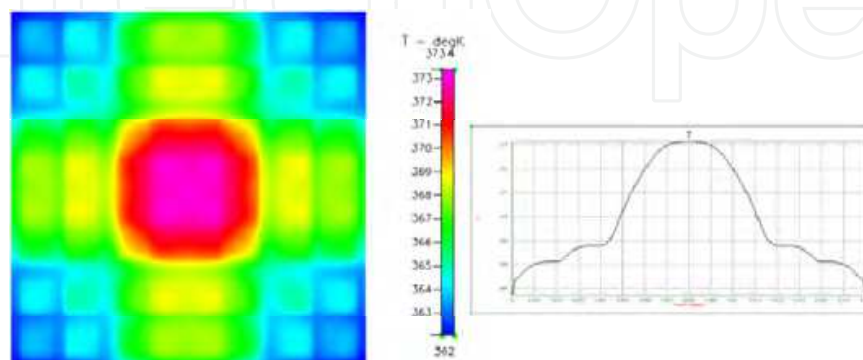


Fig. 9. The temperature distribution of the 1024-core NoC with type I thermal ridge.

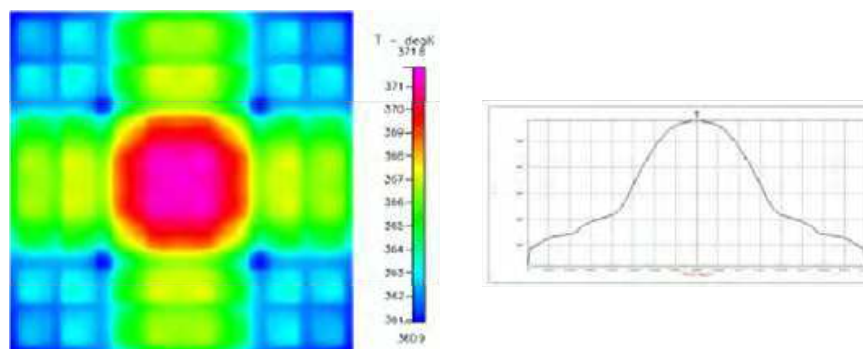


Fig. 10. Temperature distribution of the 1,024-core NoC with type-I and type-II thermal ridges.

Furthermore, the design affects the temperature non-uniformity substantially. In Figure 6 and Figure 7, it is easy to find that the value of U keeps almost constant all around the chip. However, after inserting the thermal ridges, there are several values of U on the chip. The largest U is around 4.6 K/cm, but the average U decreases substantially to 1.5 K/cm. The temperature non-uniformity is largely improved at the center and the suburb areas by the values of 0.5 K/cm and 1.5 K/cm, respectively. About 85% of the chip area is covered in the region. This means that around 850 cores have better temperature non-uniformity. Since the tile size is $410\ \mu\text{m} \times 410\ \mu\text{m}$, the temperature difference between neighboring cores in the region is less than 0.3 K.

In addition, the insertion of the type-II thermal ridge is performed, as shown in Figure 8(b). The temperature profile is shown in Figure 10. The maximum temperature of 371.8 K is

about 1.5 K lower than that shown in Figure 9. It can be further reduced, since the thermal conductivity of the type-I thermal ridge is lower than that of the type-II thermal ridge. The temperature non-uniformity and the temperature profile remain quite similar. Compared with the results from the traditional scheme with mere rotation of the hotspots, the maximum temperature decreases from 408.9 K to 372.8 K, and the temperature non-uniformity decreased from 3.2~4.0 K/cm to 0.5~1.5 K/cm in 80% of the chip area, under the constraint of increasing 20% extra area for the thermal ridges.

3. Chip design and implementation by using metallic thermal skeletons

In this chapter, a realistic thermal dissipation enhancement methodology for NoC system will be introduced. The on-chip virtual 126-core network as the hot-spot dissipates the generated heat through the metallic thermal skeletons. To evaluate the feasibility of the thermal enhancement, 9 arrays of metallic thermal skeletons are designed in the test chip. Essentially, by improving the lateral thermal dissipation path by increasing the thermal metallic skeleton in the back end of line (BEOL) metals, the heat consumed by the virtual core can be conducted into the on-chip heat sink such as the TSVs. The temperature of the hotspot can be lowered substantially if the metallic thermal skeletons arranged properly. In addition, we design thermal sensor-network on chip to facilitate the measurement and evaluation for the capability of heat transfer. Last, some important thermal characteristics of metallic thermal skeleton are listed in this chapter. In order to design a better thermal dissipation path, metallic thermal skeletons can provide alternatives for just increasing the number of thermal TSVs.

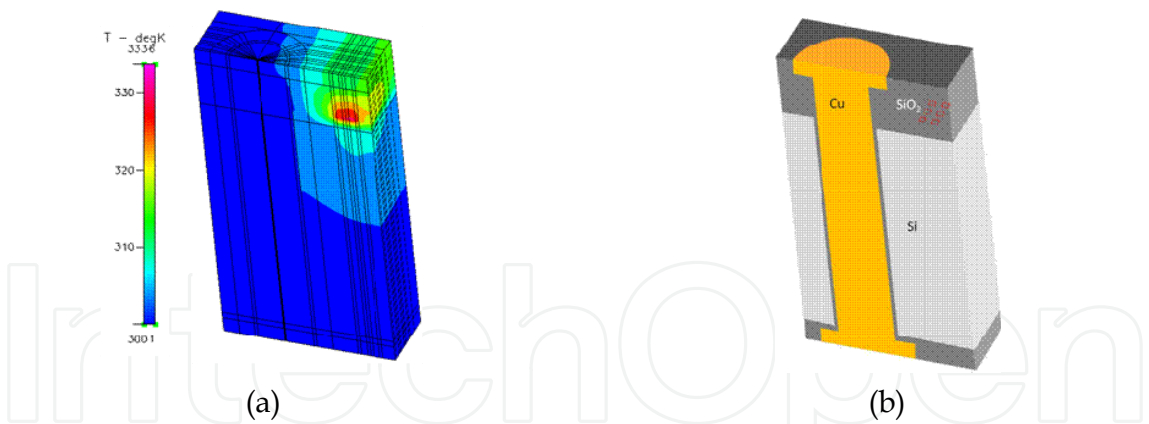


Fig. 11. FEM simulation model and result. (a) Temperature profile. (b) Simulation model.

The FEM simulation is performed by using CFD-RC, based on the following assumptions. As shown in Figure 11, a TSV is on the left, and a heat source is on the right. The other half of the structure is mirrored to the cross section. The heat source consists of 12 squares, each with power of 0.5 mW, and area of 1 $\mu\text{m} \times 1 \mu\text{m}$, which run to the top by local interconnects (not shown in the figure for they are buried in the structure), just shy of the front metal layer at the top. It is seen that the neighboring TSV is unconnected electrically and cold. The simulation assumes a TSV with dielectric thickness of 0.5 μm , diameter of 10 μm , and length of 50 μm .

3.1 Design of the proposed test chip

3.1.1 Overall floorplan of the chip

The floorplan of the proposed test chip is depicted in the Figure 12. The metallic thermal skeletons are arranged and enclosed by the core-sensor blocks. The peripheral area is for input/output and power/ground connections which provide external accesses. The test chip is designed without resorting to a complex control scheme. The virtual cores are arranged in three groups, each consisting of three rows and seven columns. The whole chip can be divided into nine regions. Each region consists of two separate areas which are enclosed by core-sensor block named A1-A7, B1-B7 and C1-C7 respectively and represent 3 types of metallic thermal skeletons. to are identical design of the metallic thermal skeleton, so do the to and to . The major differences among these nine regions are the combinations of , and elements, which are shown in Figure 13. In this design as shown in Figure 13(a), elements , and are different in the distribution densities of metal in the BEOL. For better visualization, Figure 13(b) shows the three-dimensional view of the metallic thermal skeletons. The combinations of TSVs with front metals form the on-chip heat sink, and the BEOL metal 1 to metal 4 form the metallic thermal skeletons.

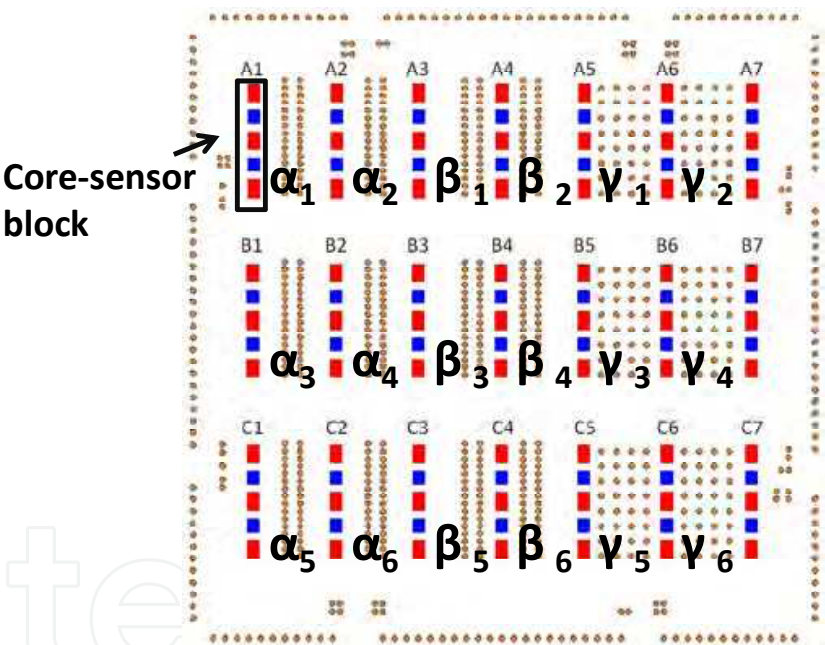


Fig. 12. The floorplan of designed test chip.

In this chapter, the stacking of the identical chips is not included in discussions, only planar die is reported. The future thermal TSV test chip will divide the core area into blocks, each, as shown in Figure 14, consisting of virtual cores, temperature sensors, and a TSV array with metallic thermal skeletons to constructs the on-chip heat sink. The virtual cores and temperature sensors are laid out at the left and right side of the on-chip heat sink. As shown in Figure 14, thermal TSV with front metals will be the on-chip heat sink, and the metallic thermal skeletons play the role as the conduction path for high speed heat transfer. Therefore, the performance of the metallic thermal skeletons are emphasized and compared with each other.

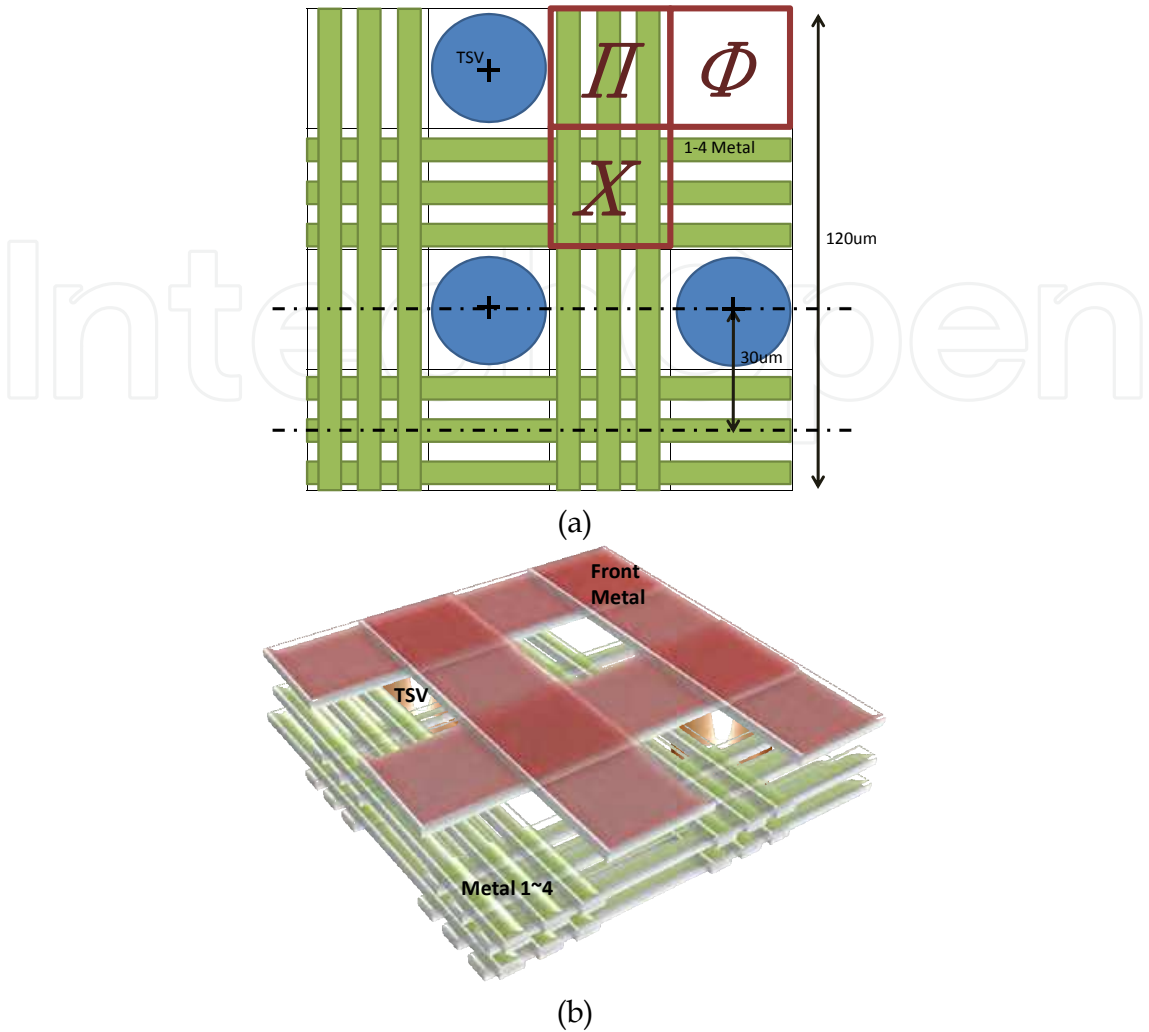


Fig. 13. The design of TSVs with metallic thermal skeletons. (a) The planar floorplan with , , and TSVs. (b) The three-dimensional view of the metallic thermal skeletons.

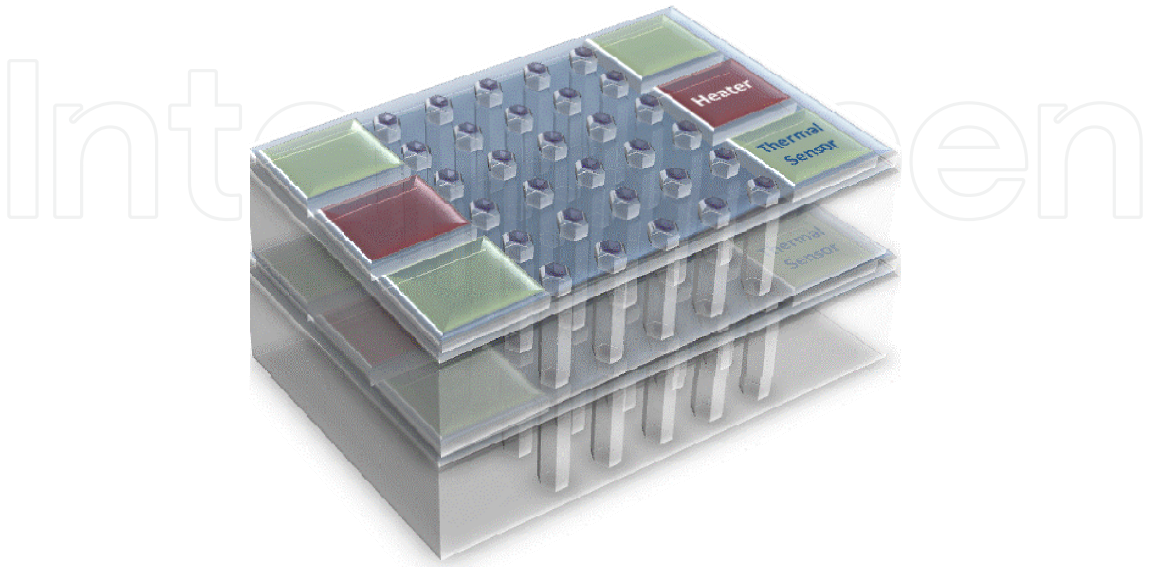


Fig. 14. Concept of virtual block design.

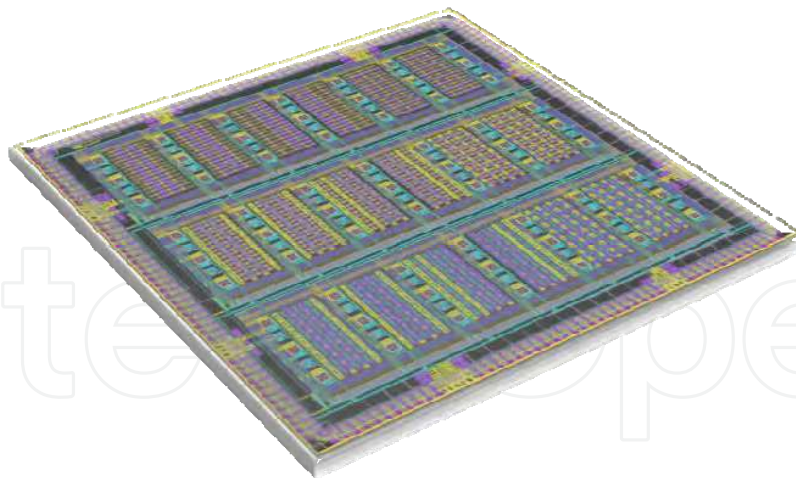


Fig. 15. The layout of the test chip.

In this chapter, to verify the capability of heat conduction, triplet experiments are designed to test the chip. Since A1-A3 is at the corner of the chip, the heat transfers more to the peripherals than to the central area of the chip. Such kind of location factors occur often in the chip measurement of thermal phenomenon. Hence, A1-A3, B1-B3 and C1-C3 are identical combination of the metallic thermal skeletons to avoid the location effects happening. The layout of the designed test chip is shown in Figure 15. The core-sensor blocks, metallic thermal skeletons, peripherals, IOs, and power domains are in one SOC chip as the NoC. The virtual core system composed of on-chip heaters can be operated at the same time. The die size measures $5,040\ \mu\text{m} \times 5,040\ \mu\text{m}$, including the seal ring. There are three voltage levels, four power domains, and nine test regions in this chip. Each voltage level can be separately controlled by the programmable logic analysis instrument. All the cores in the chip can be operated independently through the power gating mechanism. In order to precisely observe the temperature distribution of the chip surface, all sensors on the chip are activated simultaneously, and the measured temperature values can be read out as the matrix data.

3.1.2 Design of the core-sensor block

The temperature sensitive ring oscillator (TSRO) thermal sensor in Figure 16 is based on a ring oscillator whose oscillation frequency is sensitive to temperature, albeit not completely linear. In fact, the ring oscillator is also sensitive to supply voltage. Hence, to minimize power droop is important in improving the accuracy. By establishing the relationship between temperature and frequency, and opting for on-die calibration, the thermal sensor can be quite accurate. The frequency is converted by a counter and read out to a register. Figure 16(a) shows the block diagram. The control unit (CU) accepts a reference clock TS_CK and an input TS_EN which enables the sensing operation when transitioning from 0 to 1. As shown in Figure 16(a), four signals a, b, c, and RDY are generated. When the internal signal a changes from 0 to 1, the counter is reset and the count is cleared. When internal signal b changes from 0 to 1, the ring oscillator is activated and the counter starts; when it changes from 1 to 0, the ring oscillator is deactivated and the counter stops. When the internal signal c changes from 0 to 1, the count is loaded into an output register TS_REG

to be read out. The handshake signal RDY indicates that the count is ready. The physical view of the thermal sensor used in this test chip is shown in Figure 16(b).

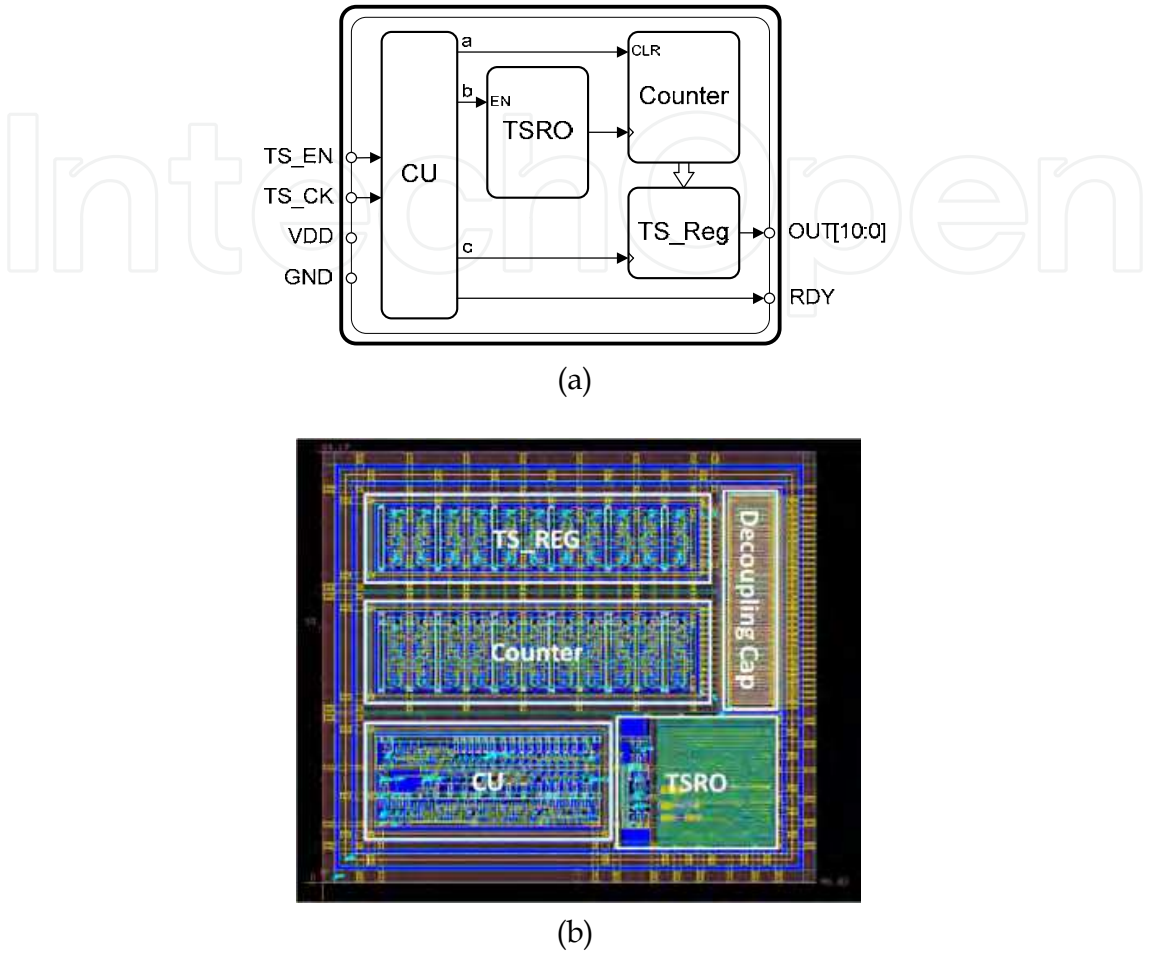


Fig. 16. Thermal sensor design. (a) The block diagram of the thermal sensor. (b) The layout of the thermal sensor, including a regulator, counter and a control unit.

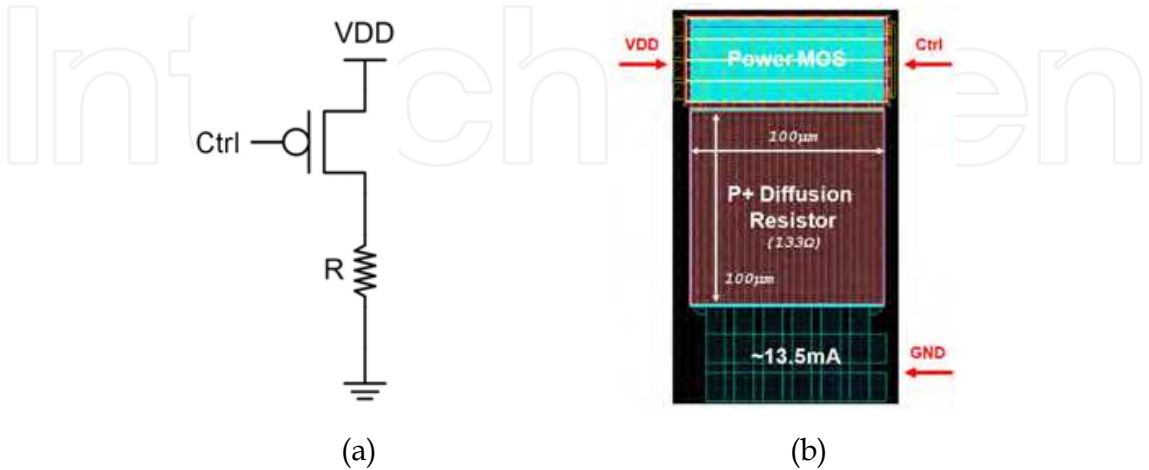


Fig. 17. Power gating design. (a) The schematic diagram of the virtual core circuits. (b) The layout view of the virtual core circuits.

The virtual core circuit is composed of a PMOS switch and a p-type diffusion resistor, as shown in Figure 17. The diffusion resistor is non-silicided and placed in an n-well. Consequently, the n-well becomes hot at first, if the heater in the virtual core is turned on, which is slightly different from a conventional CMOS circuit in that the substrate is more likely to be the heat source. The maximum current flowing into the resistor is regulated below 13.5 mA.

3.2 Thermal property analysis of the metallic thermal skeletons

The metallic thermal skeletons are intended to be placed in the regions enclosed by the core-sensor blocks. In this section, we derive analytical expressions for some key parameters.

3.2.1 Analytical model of the metallic thermal skeleton

It is clear that the heat removing rate of the metallic thermal skeletons is assumed to be q . Let us consider a pair of core-sensor blocks as the heat sources. The temperature distribution on the metallic thermal skeletons between any couple of core-sensor blocks can be expressed by (4), and then can be expressed as the following equation.

$$T_k = T_a + \frac{T_b - T_a}{w}x - \frac{q}{2k_{sk}}[w - x]x \quad (8)$$

As shown in Figure 18, where T_a and T_b are the temperatures of CS1 and CS2, respectively, q is the heat conducted to the ambient environment by the metallic thermal skeletons, k_{sk} is the equivalent thermal conductivity of the metallic thermal skeletons, and w is the width of the metallic thermal skeletons. Since T_k denotes the temperature at the location x , examining the mid-point $T_{1/2}$ by substituting x with $w/2$ into (9), we have

$$w = \left(\frac{8k_{sk}}{q} \right)^{1/2} \left[\frac{T_a + T_b}{2} - T_{1/2} \right]^{1/2} \quad (9)$$

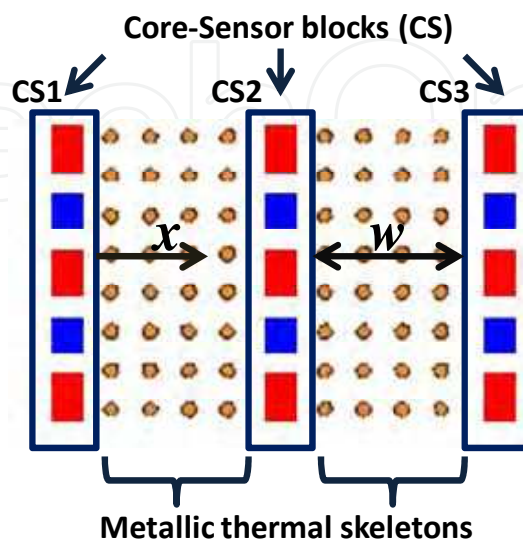


Fig. 18. The theoretical model of the core-sensor blocks with the metallic thermal skeletons.

3.2.2 Effective thermal conductivity of the metallic thermal skeletons

For the die with 9 μm of BEOL and 450 μm of the silicon substrate, we can clearly figure out that k_{sxx} is around 12~68 W/mK and k_{szz} is around 116~147 W/mK, by substituting the thermal conductivities into (6). The variation in the equivalent thermal conductivity depends on the percentage distribution of the metal in BEOL. Thus, the heat flows through the silicon substrate almost dissipates by the metallic thermal skeletons instead of transferring by silicon dioxide in the BEOL. By substituting the equivalent k_{sk} and the temperature values of T_a , T_b and $T_{1/2}$ into (9) we obtained that the widths of the metallic thermal skeleton should be 420 μm. FEM simulations have been performed to see the effectiveness of the proposed metallic thermal skeletons, as shown in Figure 19. For the reason of compatibility, we have combined the simulation results both from CFD-RD and ANSYS, so as to link the design platform for our circuit designers. Hence, to design the metallic thermal skeleton shown in Figure 12, we assumed the type α , β and γ with different distribution densities of metal in the BEOL as following equation.

$$\begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} = D \begin{bmatrix} X \\ \Pi \\ \Phi \end{bmatrix}$$

(10)

where

$$D = \begin{bmatrix} 0.28 & 0.44 & 0.28 \\ 0.20 & 0.52 & 0.28 \\ 0.36 & 0.36 & 0.28 \end{bmatrix}$$

(11)

The matrix D represents the weighting coefficients of the metallic thermal skeletons. The percent contribution of the element Φ is limited by the metal density constraint in the design rule released from the foundry.

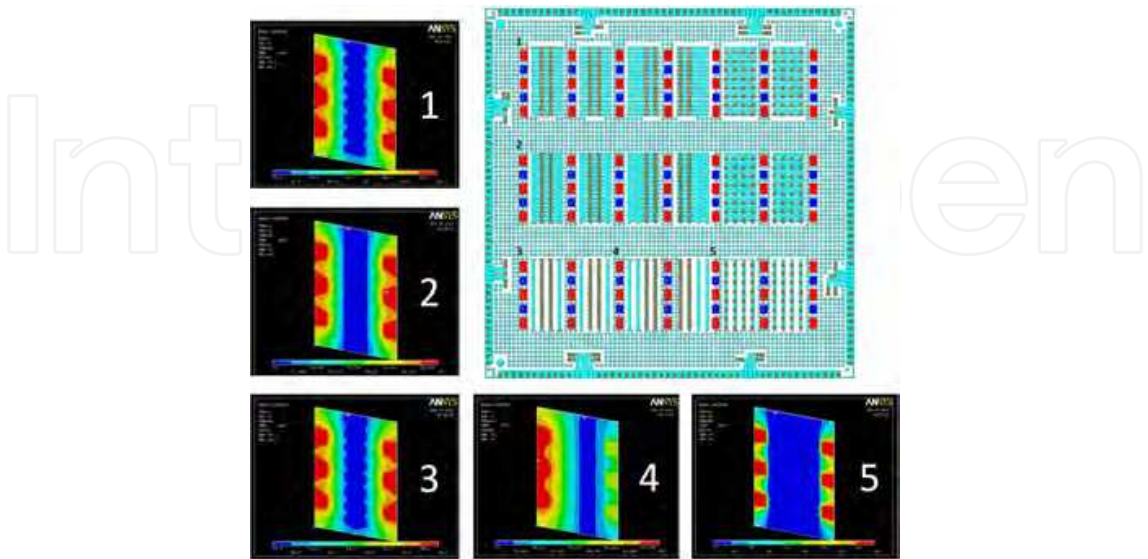


Fig. 19. The simulated results of the selected regions of the proposed architecture are shown. The enable signal H_EN is broadcast to all virtual cores.

3.3 Experimental setup

The die photo of the proposed test chip in this chapter is shown in Figure 20. This chip is fabricated by TSMC in 0.18 μm 1P4M mixed-mode process technology. The package uses 256-pin IST Universal PGA. The front of the chip is covered by the package glue. In order to observe the thermal behavior of the test chip, the back of the chip is exposed to air with a transparent PYREX[®] glass of 120 μm . There is a 6 cm x 6cm open window in the central area of the evaluation board to facilitate the observation on the temperature measurement.

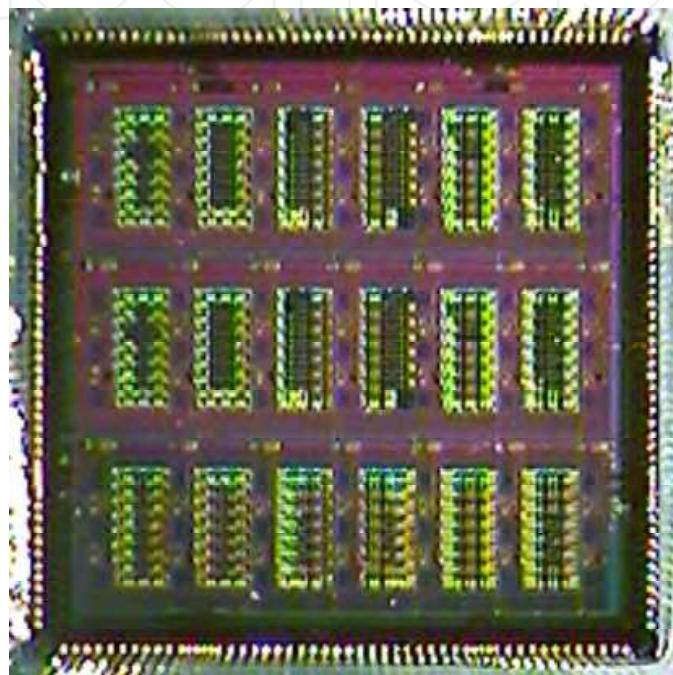


Fig. 20. The die photo of the proposed test chip in this chapter, the dimension of the chip is 5,040 μm \times 5,040 μm , including the seal ring.

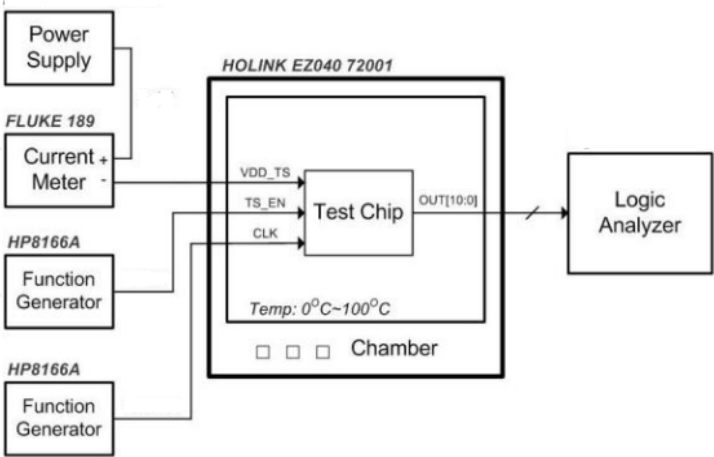
The principle measurement environment setup includes DC power supplier (MOTECH PPS 3210), current meter (FLUKE 189), function generator (HP 8166A), temperature-humidity chamber (HOLINK EZ040-72001), logic analyzer (Agilent N6705A), infrared camera (FLIR SC5700), and thermal management total analysis platform. As shown in Figure 21(a), the FLIR SC5700 with a microscope of three μm resolution is responsible for infrared radiation (IR) inspection. The temperature responses are measured by the thermal management total analysis platform designed by ICL, ITRI as shown in Figure 21(b). It is clear in Figure 21(c), the test environment is controlled at a constant ambient temperature, in which the temperature error varies within ± 0.5 $^{\circ}\text{C}$. The programmable temperature-humidity chamber HOLINK EZ040-72001 is used to control the operation temperature from 0 $^{\circ}\text{C}$ to 100 $^{\circ}\text{C}$. MOTTECH PPS 3210 is the power supply which provides the three voltage levels. The control signals (*TS_EN* and *CLK*) are generated from HP 8166A. The current meter FLUKE 189 is utilized for measuring the current consumption. Last, the output signals are collected and analyzed by Agilent N6705A.



(a)



(b)



(c)

Fig. 21. The testing environment and setup. (a) The test chip is under the measurement environment with the infrared radiation inspection. (b) The naked die with the evaluation board and thermal management total analysis platform. (c) The test chip is placed in the chamber at a nearly constant ambient temperature.

3.4 Results and discussions

The experimental results are shown in Table 1. When the power density of 7.38 W/cm² is applied to the virtual core, each core is operated at the power of 20 mW. To evaluate the thermal conduction capability of the metallic thermal skeleton, the average temperature of the metallic thermal skeleton is an important index. Since the metallic thermal skeletons are employed to conduct the heat flux generated by the virtual cores, the temperature at $w/2$ (referred to Figure 18) especially represents the results of the lateral thermal diffusion. To compare with the experimental steady state data shown in Table 1, it is clear that the virtual cores with the metallic thermal skeleton type γ have better thermal conductive performance. Moreover, $T_{\max}-T_{\min}$ denotes the temperature uniformity in the region. The results show that the metallic thermal skeleton γ has the best performance among these three combinations.

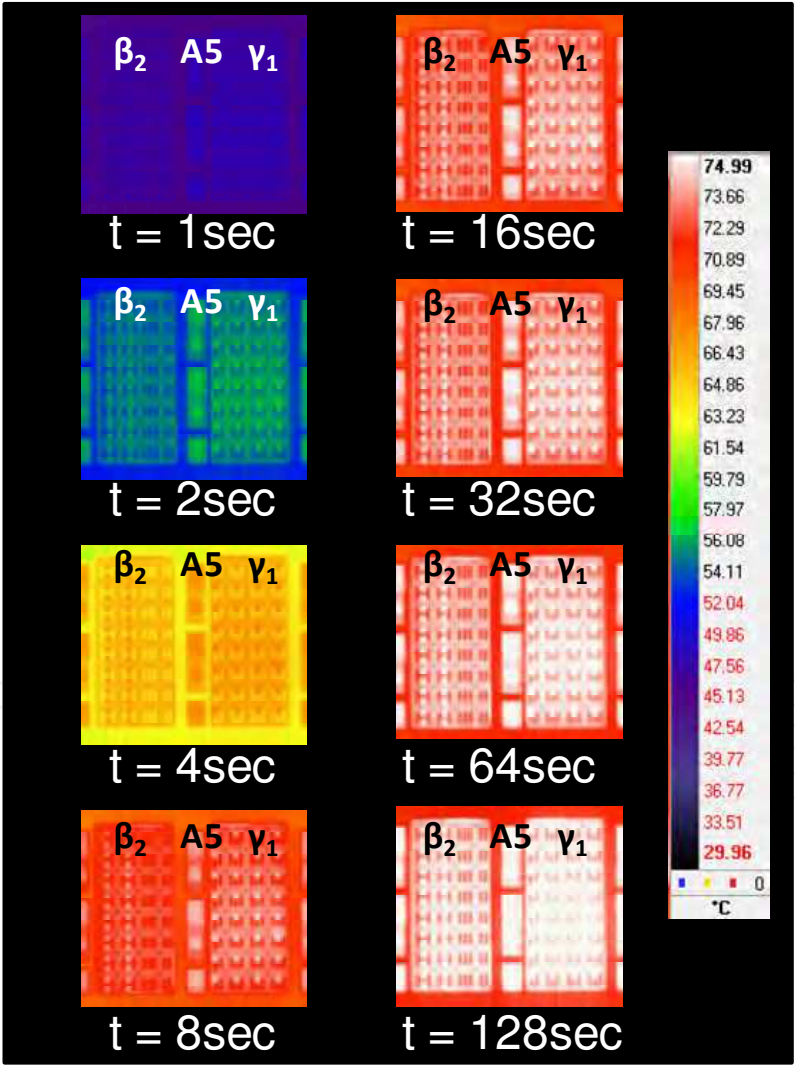


Fig. 22. The transient response of the test chip is taken by the infrared radiation camera when the virtual cores are activated. These are the back views of the test chip.

On the other hand, transient temperature response is recorded by the high speed infrared radiation dynamic photos as shown in Figure 22. Take the region in the photo for example; the β_2 -A5- γ_1 region (referred to Figure 12) includes 2 types of metallic thermal skeletons. It is clear that the temperature of γ_1 is higher than that of β_2 . This results show that the thermal conductive capability of γ_1 is better than that of β_2 . The area of Φ is limited by the metal density constraint in the design rule released from the foundry, therefore no more metal are allowed to be placed. However, the Φ region may be reserved for the placement of thermal TSVs or front metal stripes during the post CMOS process to be the on-chip heat sink.

Type of metallic thermal skeleton	α	β	γ
T_{\max} (at virtual core)	71.00	71.36	70.12
T_{avg} (on the metallic thermal skeleton)	63.13	62.92	63.45
T_{\min} (in the region)	57.48	56.38	57.82
$T_{\max} - T_{\min}$	13.52	14.98	12.30

Temperature in °C

Table 1. The temperature distribution of the test chip.

4. Conclusion

The cost of thermal ridges and metallic thermal skeletons may be compared with the advanced techniques, such as micro-channel liquid cooling or the thermo-electric cooling (TEC). Since by ITRS, the number of stacked dies is expected to increase in the future, the cooling problem of the inter-layer dies will become more challenging. If the heat should be removed by pumping liquid or external energy into the stacked dies, the cooling cost will grow exponentially. The thermal ridges and metallic thermal skeletons proposed in this chapter will be relatively cost-effective and energy-saving. Moreover, this proposed method locally improves the temperature non-uniformity, and the thermal gradient of the most part of the chip also decreases. Nevertheless, the global temperature non-uniformity which affects the chip operations from the electrical perspective deserves more efforts to pursue. Since the 3D IC with TSV now appears as an emerging technology, the early floorplan for the insertion of thermal ridges and metallic thermal skeletons for thermal management will be discussed more and more widespread. The temperature distributions measured by the

infrared radiation and by the thermal sensors are compared in this study. By these results, readers can understand that both of the data could be calibrated with each other if the package of the chip is chosen properly. Meanwhile, authors would like show also that the thermal test chip designed and proposed would be capable to evaluate the thermal properties and thermal characteristics of the packages if desired. In the 3D design of the stacking dies, the thermal measurement and verification are getting much more important. This research may give a direction or inspiration for the engineers to investigate the possibility or feasibility of better thermal designs.

5. References

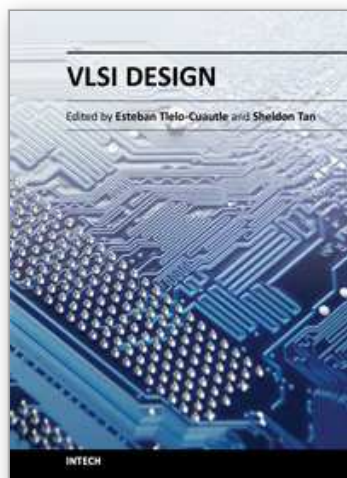
- Beigne, E.; Clermidy, F.; Miermont, S.; and Vivet, P. (2008). Dynamic voltage and frequency scaling architecture for units integration within a GALS NoC, *Proceedings of ACM/IEEE International Symposium on Networks-on-Chip (NoCS)*, ISBN: 0-7695-3098-2, Newcastle upon Tyne, April 2008, pp. 129-138.
- Brunschwiler, T.; Michel, B.; Rothuizen, H.; Kloter, U.; Wunderle, B.; Oppermann H. and Reichl, H. (2007). Interlayer cooling potential in vertically integrated packages, *ACM Journal of Microsystem Technologies - Special Issue on MicroNanoReliability*, Vol. 15, No. 1, October 2008, pp. 57-74, ISSN: 0946-7076
- Chien, J. H.; Lung, C. L.; Tsai, K. J.; Hsu, C. C.; Chen, T. S.; Chou, Y. F.; Chen, P. H.; Chang, S. C. and Kwai, D. M. (2011). Realization of 3-dimentional virtual 126-core system with thermal sensor-network using metallic thermal skeletons, *Proceedings of International Conference on Electronic Components and Technology Conference (ECTC)*, ISBN: 978-1-61284-497-8, Lake Buena Vista, FL., USA., June 2011, pp. 873-879.
- Chien, J. H.; Lung, C. L.; Hsu, C. C.; Chou, Y. F.; and Kwai, D. M. (2010). Floorplanning 1024 cores in a 3D-stacked networkon- chip with thermal-aware redistribution, *Proceedings of 12th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITHERM)*, ISBN: 978-1-4244-5342-9, Las Vegas, NV., USA., June 2010, pp. 1-6.
- Cong, J.; Luo, G.; Wei, J.; and Zhang, Y. (2004). A thermal-driven floorplanning algorithm for 3D ICs, *Proceedings of International Conference on Computer Aided Design (ICCAD)*, ISBN: 0-7803-8702-3, San Jose, CA., USA., November 2004, pp. 306-313.
- Cong, J.; Luo, G.; Wei, J.; and Zhang, Y. (2007). Thermal-aware 3D IC placement via transformation, *Proceedings of Asia and South Pacific Design Automation Conference (ASPDAC)*, ISBN: 1-4244-0630-7, Yokohama, Japan, June 2007, pp. 780-785.
- Hennessy, J. L. and Patterson, D. A. (2007). *Computer Architecture: A Quantitative Approach* (3rd Edition), Morgan Kaufmann, ISBN: 978-1-5586-0596-1, San Francisco, CA., USA.
- Kurd, N. A.; Barkatullah, J. S.; Dizon, R. O.; Fletcher, T. D. and Madland, P. D. (2001). A multigigahertz clocking scheme for the Pentium 4 microprocessor, *IEEE Journal of Solid-State Circuits (JSSC)*, Vol. 36, No. 11, November 2001, pp. 1647-1653, ISSN: 0018-9200
- Lee, S.; Lemczyk, T. F. and Yovanovich, M. M. (1992). Analysis of thermal vias in high density interconnect technology, *Proceedings of Semiconductor Thermal Measurement*

- and Management Symposium (SEMI-THERM)*, ISBN: 0-7803-0500-0, Austin, TX., USA., February 1992, pp. 55-61.
- Lu, J.-Q. (2009). 3-D hyperintergration and packaging technologies for micro-nano systems, *Proceedings of IEEE*, Vol. 97, No. 1, January 2009, pp. 18-30, ISSN: 0018-9219
- Lung, C. L.; Ho, Y. L.; Huang, S. H.; Hsu, C. W.; Liao, J. L.; Huang, S. Y. and Chang, S. C. ; (2010). Thermal analysis experiences of a tri-core SoC system, *Proceedings of International Conference on Green Circuits and Systems (ICGCS)*, ISBN: 978-1-4244-6876-8, Shanghai, China, June 2010, pp. 589-594.
- Lung, C. L.; Ho, Y. L.; Kwai, D. M. and Chang, S. C. (2011). Thermal-Aware On-Line Task Allocation for 3D Multi-Core Processor Throughput Optimization, *Proceedings of Design, Automation & Test in Europe (DATE)*, Grenoble, France, March 2011,. pp. 1-6
- Tiwari, V., Singh, D., Rajgopal, S., Mehta, G., Patel, R., and Baez, F. (1998). Reducing power in high-performance microprocessors, *Proceedings of ACM/IEEE Design Automation Conference (DAC)*, ISBN: 0-89791-964-5, San Fransisco, CA., USA., June 1998, pp. 732-737.
- Tran, A. T.; Truong, D. N. and Baas, B. M. (2009). A GALS many-core heterogeneous DSP platform with source-synchronous on-chip interconnection network, *Proceedings of ACM/IEEE International Symposium on Networks-on-Chip (NoCS)*, ISBN: 978-1-4244-4142-6, San Diego, CA., USA., May 2009, pp. 214-223.
- Tran, A. T.; Truong, D. N. and Baas, B. M. (2009). A low-cost high-speed source-synchronous interconnection technique for GALS chip multiprocessors, *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS)*, ISBN: 978-1-4244-3827-3, Taipei, Taiwan, May 2009, pp. 996-999.
- Truong, D.; Cheng, W.; Mohsenin, T.; Yu, Z.; Jacobson, T.; Landge, G.; Meeuwsen, M.; Watnik, C.; Mejia, P.; Tran, A.; Webb, J.; Work, E.; Xiao, Z. and Baas, B. (2008). A 167-processor 65 nm computational platform with per-processor dynamic supply voltage and dynamic clock frequency scaling, *IEEE Symposium on VLSI Circuits (VLSIC)*, ISBN: 978-1-4244-1804-6, Honolulu, HI., USA., June 2008, pp. 22-23.
- Tsai, C. H. and Kang, S. M. (2000). Cell-level placement for improving substrate thermal distribution, *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD)*, Vol. 19, No. 2, February 2000, pp. 253-266, ISSN: 0278-0070.
- Tsui, Y. K.; Lee, S. W. R.; Wu, J. S.; Kim, J. K. and Yuen, M. M. F. (2003). Three-dimensional packaging for multi-chip module with through-the-silicon via hole, *Proceedings of Electronics Packaging Technology Conference (EPTC)*, ISBN: 0-7803-8205-6, Singapore, Marcg 2003, pp 1-7.
- Yu, Z. and Baas, B. M. (2006). Implementing tile-based chip multiprocessors with GALS clocking styles, *IEEE International Conference on Computer Design (ICCD)*, ISBN: 978-0-7803-9707-1, San Jose, CA., USA., October 2006, pp. 174-179.
- Xu, G.; Guenin, B. and Vogel, M. (2004). Extension of air cooling for high power processors, *Proceedings of 9th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITHERM)*, ISBN: 0-7803-8357-5, Las Vegas, NV., USA., August 2004, pp. 186-193.

Xu, G. (2006). Thermal modeling of multi-core processors, *Proceedings of 10th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITHERM)*, ISBN: 0-7803-9524-7, San Diego, CA., USA., May 2006, pp. 96-100.

IntechOpen

IntechOpen



VLSI Design

Edited by Dr. Esteban Tlelo-Cuautle

ISBN 978-953-307-884-7

Hard cover, 290 pages

Publisher InTech

Published online 20, January, 2012

Published in print edition January, 2012

This book provides some recent advances in design nanometer VLSI chips. The selected topics try to present some open problems and challenges with important topics ranging from design tools, new post-silicon devices, GPU-based parallel computing, emerging 3D integration, and antenna design. The book consists of two parts, with chapters such as: VLSI design for multi-sensor smart systems on a chip, Three-dimensional integrated circuits design for thousand-core processors, Parallel symbolic analysis of large analog circuits on GPU platforms, Algorithms for CAD tools VLSI design, A multilevel memetic algorithm for large SAT-encoded problems, etc.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Chiao-Ling Lung, Jui-Hung Chien, Yung-Fa Chou, Ding-Ming Kwai and Shih-Chieh Chang (2012). Three-Dimensional Integrated Circuits Design for Thousand-Core Processors: From Aspect of Thermal Management, VLSI Design, Dr. Esteban Tlelo-Cuautle (Ed.), ISBN: 978-953-307-884-7, InTech, Available from: <http://www.intechopen.com/books/vlsi-design/three-dimensional-integrated-circuits-design-for-thousand-core-processors-from-aspect-of-thermal-man>

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2012 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

IntechOpen

IntechOpen