

Thermal Analysis of Heterogeneous 3-D ICs with Various Integration Scenarios

Ting-Yen Chiang, Shukri J. Souri, Chi On Chui, and Krishna C. Saraswat

Department of Electrical Engineering, Stanford University, Stanford, CA 94305

tychiang@stanford.edu, 1-650-725-7062

Abstract

This paper presents detailed thermal analysis of high performance three dimensional (3-D) ICs under various integration schemes. The model incorporates the effect of vias and power consumption due to both devices in active layers and interconnect joule heating. The results show excellent agreement with the 3-D finite element simulations using ANSYS. It is shown that under certain scenarios, 3-D ICs can actually lead to better thermal performance than planar (2-D) ICs. With the effect of vias, as efficient heat dissipation paths, taken into account, our model provides more realistic temperature rise estimation for 3-D ICs. Furthermore, tradeoffs among power, performance, chip real estate and thermal impact for 3-D ICs is evaluated. Finally, the thermal influence from incorporating RF circuits and optical interconnect on 3-D ICs has been discussed.

Introduction

The 3-D IC architecture has emerged as a unique solution to alleviate the interconnect delay problem [1]. Additionally, 3-D ICs provide a vehicle for heterogeneous integration to realize System-on-a-Chip (SoC) designs, e.g., to incorporate RF ICs and optical interconnects. However, there is concern about poor heat dissipation and consequent chip temperature rise due to increased power density and lower thermal conductivity of inter layer dielectrics (ILD) [2]. As 3-D ICs contain numerous vias, which can serve as efficient heat dissipating paths, their effect on the thermal characteristics should be included for realistic thermal analysis. These vias have much higher thermal conductivity and hence can effectively reduce the thermal resistance caused by the ILD layers. Ignoring the effect of these structures can result in overly pessimistic estimations predicting unacceptably high 3-D chip temperatures. Recently, we have developed a model to quantify the via thermal effect in 2-D structures [3, 4]. In this paper a compact analytical model to evaluate temperature rise in 3-D structures incorporating the effect of vias is presented.

Thermal Modeling Incorporating Via Effect

According to ITRS [5], although the average power density for high performance microprocessor will remain relatively constant throughout the technology nodes, current density in the wires will rise significantly (Fig. 1). Furthermore, Cu resistivity will increase due to barriers, surface scattering and skin effect. Thus interconnect joule heating will become significant. In addition, low-k dielectrics with poor thermal conductivity (Fig. 2), will not only lead to higher interconnect temperature in 2-D ICs but also impact the device temperature in various active layers in 3-D ICs (Fig. 3). As seen in Fig. 4, even for the case of two active layers, the ratio of thermal resistance caused by ILD

layers (R_{ILD}) to required package (including glue layers and heat sink) thermal resistance (R_{pkg}) increases rapidly for future technology nodes. The required R_{pkg} is the maximum allowed value which gives the maximum junction temperature specified in ITRS. With multiple active layers, R_{ILD} could become the dominant factor to determine temperature rise in 3-D ICs. The analytical expression, based on first principles, to evaluate temperature rise in 3-D structure is shown in Fig. 5. Via effect is incorporated in the expression by the via correction factor η ($0 \leq \eta \leq 1$), with $k_{ILD,eff} = k_{ILD}/\eta$, where $k_{ILD,eff}$ is the effective thermal conductivity of ILD with the help of via effect and k_{ILD} is the nominal thermal conductivity with via effect ignored [3]. Power consumption due to both active (device) layers and interconnect joule heating are included. This expression can be better understood by comparing it with the Elmore-delay model following an electrical-thermal analogy (Fig. 5). The model has been validated by comparing with full chip thermal simulations done using ANSYS [2]. The two-layer 3-D structure with wafer bonding technique used for the validation is shown in Fig. 6. Since it is too complicated to construct the 3-D structure with thousands of vias in ANSYS, the validation is done for the case where via effect is ignored ($\eta = 1$). The results from analytical expression show excellent agreement with ANSYS (Fig. 7). However, the analytical model takes much less computation time and provides better insight. Furthermore, as shown in Fig. 7, via effect greatly helps heat dissipation and the resultant temperatures are much lower. Therefore, it is crucial to include via effect for thermal analysis of 3-D ICs.

Power Analysis of 3-D ICs

A power dissipation comparison between wire-pitch limited 2-D and 3-D ICs for the 50 nm technology node is presented in Table I. Thermal resistance of the package is assumed to be $2.15 \text{ cm}^2 \text{ }^\circ\text{C/W}$ from ITRS for 2-D ICs. The dynamic power dissipation components considered here are due to logic; interconnect; clock distribution and repeaters and are calculated using $P_{Dynamic} = 1/2 \alpha C V_{dd}^2 f_c$ where α is the activity factor (assumed to be 0.1), V_{dd} is the supply voltage obtained from ITRS, f_c is the operating frequency and C is capacitance. Other power dissipating components include memory, I/O pads and static components such as leakage and short-circuit currents (which are considered negligible), are all combined under P_{Other} .

C is calculated for each component to determine the associated power dissipated. For P_{Logic} , the device capacitance is calculated by considering gate oxide capacitance, overlap capacitance, and junction capacitance all of which are calculated from ITRS. Interconnect capacitances are found from the wire-length distribution and the dimensions of the wire pitches for each tier [1]. Clock distribution capacitances are calculated using

the BACPAC model proposed in [6] by considering a buffered H-Tree model. Power dissipated by repeaters is calculated based on the driver capacitances and the number of repeaters modeled in [1]. P_{Other} is determined in the 2-D case to be the sum of remaining components to achieve the ITRS projected total power dissipation for this generation. Since this component is assumed dominated by dynamic dissipation it is considered linearly dependent on the operating frequency for all 3-D cases.

In 3-D case 2, the total power dissipation is seen to decrease primarily due to the reduction in the wiring requirement thus reducing the interconnect power dissipation, reducing number of required repeaters and reducing the clock distribution network. 3-D case 3 is associated with a larger chip area which requires longer interconnect lines, a larger number of repeaters and clock-distribution network all of which increase the power dissipation. 3-D case 4 shows a dramatic increase in the power dissipated primarily due to the significant increase in operating frequency. 3-D case 5 illustrates the increase in the operating frequency if the chip area and the power dissipation requirements are maintained constant to 2-D.

Although the total power consumption as shown in Table I is reduced by going from 2-D to 3-D ICs due to the reduction in the interconnect and the clock network related capacitance, the heat removal capability could deteriorate as the upper active layers experience longer heat dissipation path to heat sink. Fig. 8 compares the temperature rise for different 3-D integration scenarios. It shows that, in most of the cases, 3-D ICs have similar temperature rise but have the advantage of either reduced chip area (case 2) or increased operating frequency (case 5). In the case of equal chip area and operating frequency (case 3), lower temperature than 2-D ICs can be achieved. Note that to double the operating frequency (case 4), temperature will invariably increase. Nevertheless, the temperature is still much lower than that estimated with via effect ignored [2].

Furthermore, it is desirable to put memory in close proximity to logic circuitry to reduce latency in high performance microprocessors. 3-D ICs provides an excellent opportunity to stack memory and logic. The power consumption in on-chip memory is generally less than 10% of total power consumption and the area occupied by memory and logic are comparable at 50 nm node. With these assumptions, Fig. 9 shows four 3-D stack schemes along with their temperature performance. It can be observed that with logic in the bottom active layer and memory in the upper layer (scheme 1), the resultant temperature rise is the lowest. On the other hand, stacking logic parts back to back will experience much higher temperature rise.

Thermal Impact on Heterogeneous 3-D Integration

Heterogeneous integration of RF circuit in 3-D ICs, e.g., with a microprocessor to facilitate wireless communication, has attracted major interest with the intent to isolate substrate-coupled noise between digital and analog components [1]. Evaluated in Fig. 10 is the noise figure, indexing RF performance, which deteriorates rapidly with temperature rise, imposing more stringent design requirements for 3-D integration.

Employing optical interconnect in global signaling or clocking could eliminate many problems associated with large multi-GHz chips like reducing timing skew, power and area for clock distribution [7]. In an on-chip optical interconnect system, photodetectors are responsible for optical-to-electrical signal conversion. Usually, the detector photocurrent is $\sim 100\mu\text{A}$. Even the dark current (noise) increases rapidly with temperature (Fig. 11), a photo-to-dark current ratio of 100 is still retained, which is enough for reliable communications. In a dense optical interconnect, the receivers are not noise-limited, but indeed gain-limited aiming at low power. Focusing on the transmitting components like lasers and modulator diodes, the three major temperature-induced changes in performance are drop in quantum efficiency, degradation of laser threshold current, and shift in transmitter wavelength. Making use of the temperature-dependent hydrostatic stress [8] and a bimetallic heatsink [9] could get around the detrimental wavelength shift, in addition to the efficiency drop and threshold degradation. Recently, a quantum-dot laser with a maximum operating temperature of 160°C in pulse mode [10] has been demonstrated. The last major components are the waveguide-related devices like modulators, and optical attenuators, etc. For a waveguide (attenuator) with an InGaAsP MQW active layer and InP claddings, a negligible change in absorption coefficient (an index for fiber loss) has been demonstrated from about 20 to 120°C [11]. Even up to 170°C , the absorption loss is still tolerable. From the above analysis, the incorporation of optical interconnect into the 3-D architecture is viable, but may require some more device design efforts for higher temperature operation.

Summary

In conclusion, a compact analytical thermal model to evaluate temperature distribution in 3-D ICs including via effect is presented. It is demonstrated that via effect must be considered to evaluate the thermal capability of 3-D ICs. Temperature performance of various 3-D integration schemes has been examined thoroughly. It is shown that with careful thermal designs, 3-D ICs can have the same thermal capability as that of 2-D ICs. For the high performance case where higher temperature is not avoidable, better circuit design and advanced package solution will be necessary. Finally, examples of 3-D heterogeneous integration of RF circuits and optical interconnects have been explored from the thermal view point.

Acknowledgements

This work was supported by the MARCO Interconnect Focus Center and DARPA HGI Program. We would like to thank Dr. Kaustav Banerjee for his valuable suggestions.

References

- [1] K. Banerjee *et al.*, *Proc. IEEE*, Vol. 89, No. 5, pp. 602-633, 2001
- [2] S. Im *et al.*, *IEDM Tech. Dig.*, pp. 727-730, 2000.
- [3] T-Y, Chiang *et al.*, *IITC*, pp. 92-94, 2001.
- [4] T-Y, Chiang *et al.*, *VLSI Tech. Symp.*, pp. 141-142, 2001
- [5] International Technology Roadmap for Semiconductors, 1999.
- [6] D. Sylvester *et al.*, *Proc. IEEE*, pp. 634-664, May 2001.
- [7] D. A. B. Miller, *IEEE JSTQE*, vol. 6, no. 6, pp. 1312-1317, 2000
- [8] D. A. Cohen *et al.*, *APL*, vol. 69, no. 4, pp. 455-457, 1996.
- [9] D. A. Cohen *et al.*, *OFC '97 Tech. Dig.*, pp. 303-304, 1997.
- [10] F. Klopff *et al.*, *IEEE PTL*, vol. 13, no. 8, pp. 764-766, 2001.
- [11] M. Allard *et al.*, *IEEE JLT*, vol. 18, no. 6, pp. 813-818, 2000.

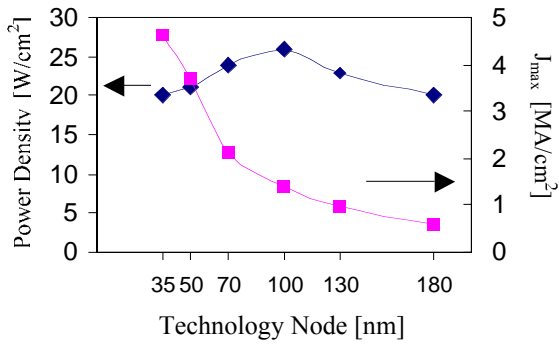


Figure 1. Trends of chip power density and interconnect J_{max} along technology nodes suggested by ITRS [5]. Chip power density is calculated by total power of the chip divided by chip size.

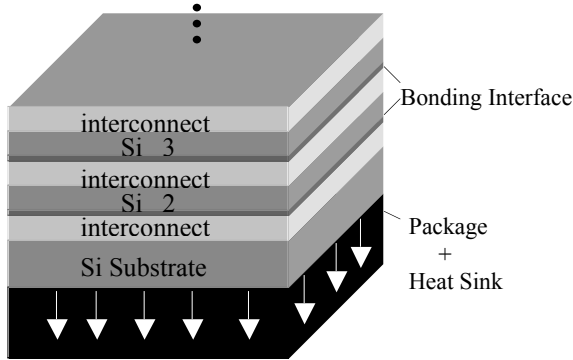


Figure 3. Schematic of multi-level 3-D IC with a heat sink attached to Si substrate.

$$\overline{T_{Si_N}} = T_{amb} + \underbrace{\sum_{m=1}^{N-1} \left[\sum_{n=1}^{N_m} \frac{t_{ILD,mn}}{k_{ILD,mn} s_{mn}} \eta_{mn} \left(\sum_{i=n}^{N_m} j_{rms,mn}^2 \rho H_{mn} + \sum_{j=m+1}^M \Phi_j \right) \right]}_{\text{Temperature rise caused by ILDs}} + \underbrace{\sum_{m=1}^M R_m \left(\sum_{k=m}^M \Phi_k \right)}_{\text{Temp. Rise caused by PKG, glue layer, Si sub.}}$$

where T_{amb} : ambient temperature,
 M : number of strata,
 N_m : number of metal levels in the m^{th} stratum,
 mn : the n^{th} interconnect level in the m^{th} stratum,
 t_{ILD} : thickness of ILD,
 k_{ILD} : thermal conductivity of ILD materials,
 s : heat spreading factor [3],
 η : via correction factor, $0 \leq \eta \leq 1$ [3],
 j_{rms} : root-mean-square value of current density flowing in the wires,
 ρ : electrical resistivity of metal wires,
 H : thickness of metal wires,
 Φ : total power density of m^{th} stratum, including power consumed by active layer and interconnect joule heating,
 R : thermal resistance of glue layer and Si layer for each of the stratum, with R_i represents the total thermal resistance of package, heat sink and Si substrate.

Elmore-Delay Analogy

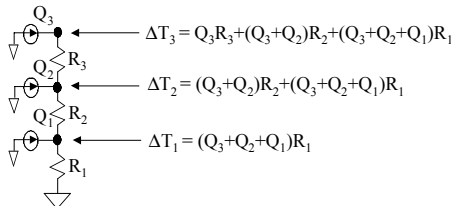


Figure 5. Analytical expression to evaluate average temperature rise on the N^{th} active layer, with total M active layers in the 3-D ICs. The expression can be better understood by comparing the analogy of this thermal model and the Elmore-delay model.

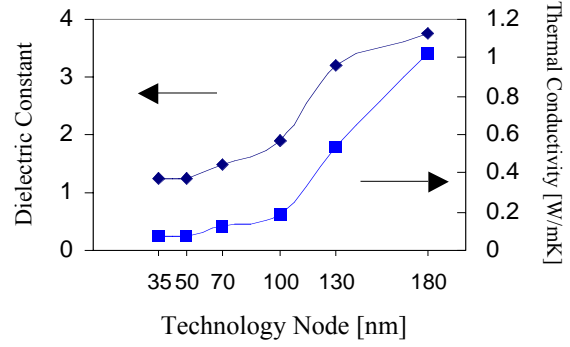


Figure 2. Both dielectric constant and thermal conductivity of ILD materials decrease with advanced technology nodes.

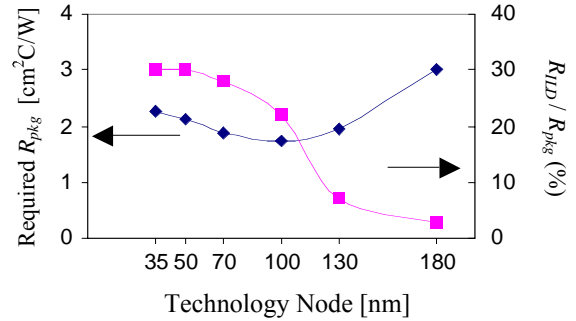


Figure 4. The required package thermal resistance, R_{pkg} , to achieve the maximum junction temperature specified in ITRS and the ratio of R_{ILD} to R_{pkg} vs. technology nodes.

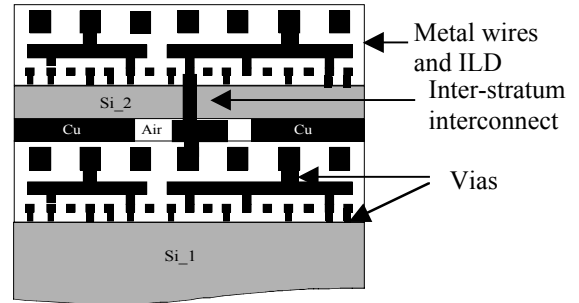


Figure 6. Schematic of 3-D structure fabricated by wafer bonding using Cu pad thermo-compression.

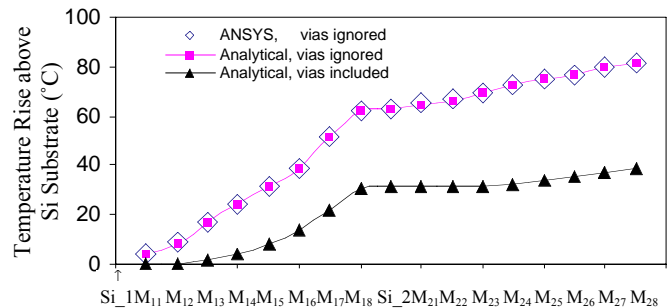


Figure 7. Temperature distribution along the vertical layers from the Si substrate surface (Si_1) to top metal level of the second stratum. The 3-D structure is shown in Fig. 6. M_{11} represents the first metal level in stratum 1, etc. J_{rms} is 4.85×10^5 A/cm² and power density of each active layer is assumed to be 0.615 W/cm². Interconnect parameters are quoted from ITRS 100 nm technology node. For the case of via effect included, the via separations assigned to the metal layers, from 1st to 8th levels, are 1, 5, 15, 30, 50, 80, 200 and 500 μ m, respectively.

	2-D	3-D, Case 1	3-D, Case 2	3-D, Case 3	3-D, Case 4	3-D, Case 5
Active Layers	1	2	2	2	2	2
f_c (MHz)	3000	3000	3000	3000	6000	3559
Feature Size (nm)	50	50	50	50	50	50
Chip Area (cm ²)	8.17	4.25	4.51	8.17	8.17	8.17
Memory Area (cm ²)	3.92	3.92	3.92	3.92	3.92	3.92
Logic Area (cm ²)	4.25	4.25	5.1	12.42	12.42	12.42
P_{Logic} (W)	34.8	34.8	34.8	34.8	69.6	41.28
P_{Local} (W)	17.4	17.4	17.44	20.66	10.44	6.19
$P_{Semi-Global}$ (W)	14.63	14.63	6.89	8.16	30.68	18.2
P_{Global} (W)	6.96	6.96	4.18	5.63	11.78	6.99
P_{Clock} (W)	34.8	34.8	22.97	27.21	56.93	33.76
$P_{Repeaters}$ (W)	45.24	45.24	29.7	35.19	73.6	43.65
P_{Other} (W)	20.17	20.17	20.17	20.17	40.34	23.93
P_{Total} (W)	174	174	136.15	151.82	293.37	174
Power Density Per Active Layer (Wcm ⁻²)	21.30	20.47	15.09	9.29	17.95	10.65

Table I: Comparison of power dissipation due to logic, interconnect, clock distribution and repeaters for 2-D and 3-D ICs with 2 active layers for ITRS 50nm technology node. 3-D IC cases are presented for comparison by varying the chip area, A_c , and operating frequency, f_c , and represent the same 2-D IC (conserving feature size, number of transistors and functionality) converted to 3-D with 2 active layers. 3-D Case 1 is special in that memory and logic are each dedicated to separate active layers without any modifications to wiring. Resulting A_c is determined by larger logic area and power dissipation is unchanged relative to 2-D. The following 4 3-D cases are obtained, and compared to 2-D, by modifying the wiring to achieve, respectively: equal f_c and decreased A_c (Case 2); equal f_c and A_c (Case 3); $2*f_c$ and equal A_c (Case 4); equal A_c with f_c determined by maintaining 2-D P_{Total} (Case 5).

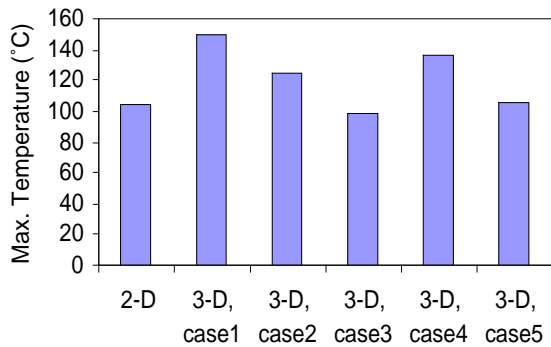


Figure 8. Comparison of temperature performance among 2-D ICs and five different two-active-layer 3-D ICs scenarios.

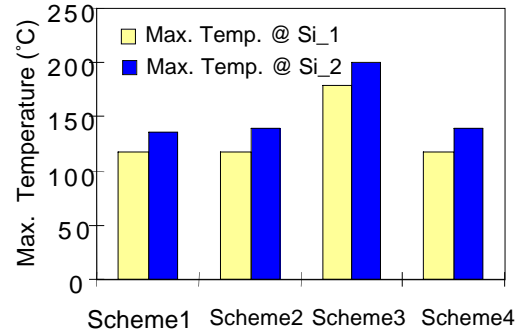
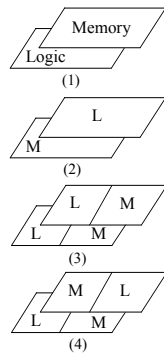


Figure 9. Thermal capability of high performance 3-D ICs (case 4) for four different 3-D logic-memory integration schemes.

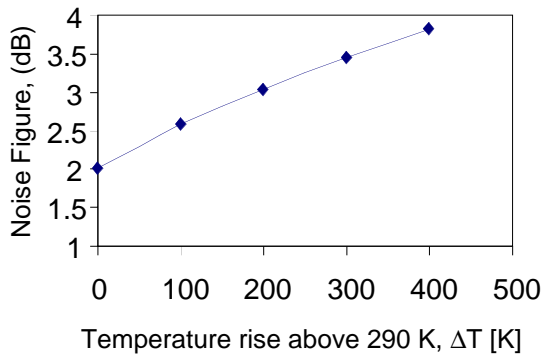


Figure 10. The noise performance of a typical low noise amplifier (LNA) under the temperature effect is evaluated. Noise figure is assumed to be 2dB @ 290K.

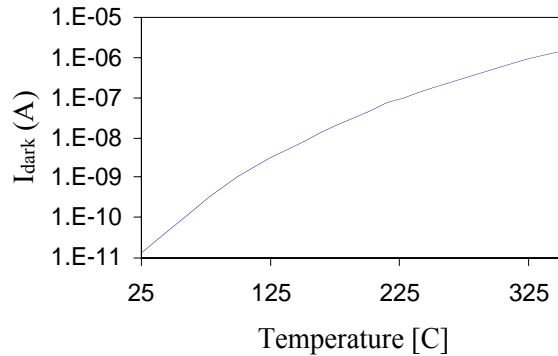


Figure 11. Temperature effect on optical receiver performance is evaluated. A typical example of Ti-Si MSM PD dark current vs. ambience temperature is shown here.