Abstract—It is known that 3D-ICs suffer from hot spot temperatures that can reach thousands of degrees if they are not cooled to reasonable operating temperatures. The problem of hot spots is not limited to the high temperatures of the IC; thermal stress can also pose severe problems, even after cooling the chip. This study investigates thermal stress resulting from a 3D-IC hot spot with 20 W power dissipation. The IC is cooled using SiO2 and diamond cooling blocks. The study is performed using three coolants: water, Freon (R22), and Liquid Nitrogen (LN). As expected, the study shows that metal layers on the chip suffer from high thermal stress due to rising of the chip temperature to values higher than the room temperature. It is also noticed that the stress becomes more severe if cooling is done using LN. In fact, the stress exceeded the maximum tensile strength of aluminium, which might cause failure of the chip. This indicates that cooling 3D-IC may not ensure acceptable operation or reliability. Thermal stress must be investigated at both high and low temperatures to ensure high performance and acceptable reliability.

Keywords— 3D-IC, Liquid Cooling, COMSOL, Thermal Stress.

I. INTRODUCTION

With the increasing density of ICs, two challenges have risen: the first is reaching the physical limit of the device size (reaching the end of Moore’s law), and the second is the increase of the power density of the ICs to the point that the performance is adversely affected. To solve the first challenge, a new approach of integration where regular 2D ICs are stacked on top of each other forming 3D-ICs [1]. The electrical power connections and signal transmission between the stacked ICs are done using metal vertical connections through holes, or vias. The 3D-ICs technology provides many advantages. First, the integration density can be increased by a factor equal to the number vertical stacks (for example, to 10,000%, if 100 2D-ICs are stacked on top of each other). Second, the total speed of the IC increases due to the short vertical connections between the stacks. Third, 3D-ICs allowed for integration of heterogeneous technology.

Although 3D-IC technology resulted in many advantages, it introduced serious problems. One of the problems is the introduction of power noise resulting from the need for huge power supplies and long signal delays [2]. The second, is the creation of hot spots, due to the high integration density. Hot spots temperatures may reach thousands of degrees, which would destroy the IC in milliseconds. This is a very serious problem, especially for the next generation of IC where a power density reaching 1000 kW/cm2 will be a requirement, according to ITRS [3].

To manage power dissipation from 3D ICs, several techniques have been proposed and implemented. Powerful fans, large heat sinks, Microelectromechanical Systems (MEMS)-based cooling technology, embedded micro channels [4,5], liquid immersion cooling [6], and microfluidic cooling using thermal TSVs [7] are among the utilized cooling techniques. Although the steady state temperature can be reduced to an acceptable operating temperature, there will be still time-dependent local temperature change [8]. Such dynamic change of the temperature can cause thermal stress and strain. The thermal energy produced from local hot spots can cause severe damage to the IC due to the thermal stress and strain. This mechanical stress is expected to affect the reliability and performance of the circuit, due to the likely chance of damage to the metal and insulator layers.

In this study, we investigated the thermal stress resulting from the dynamic temperature change of hot spots in liquid cooled 3D ICs. The study was performed using advanced multi-physics numerical analysis program, ComsolTM.

II. THERMAL STRESS

The relationship between stress and strain is expressed by,  

$$\sigma = E\varepsilon$$  \hspace{1cm}(1)$$

where $\sigma$ is the uniaxial stress, $\varepsilon$ is the strain or deformation, and E is Young’s modulus [9]. Young’s modulus, E, gives the measure of stiffness of a solid material [10].

Thermal stress can be expressed by the following equation,  

$$\sigma = E\alpha\Delta T = E\alpha (T_f - T_0)$$  \hspace{1cm}(2)$$

where $\alpha$ is the thermal expansion coefficient, $T_0$ is initial temperature, $T_f$ is final temperature, and $\Delta T = T_f - T_0$ in kelvins (K). The temperature, $T_f$ varies with respect to time due to subsequent heating and cooling and results in a temperature difference [11]. From equation (2), we can see that, for a specific material, the stress is proportional to the thermal expansion coefficient, young’s modulus, and temperature difference. Since the materials used to build IC do not have the same thermal expansion, significant thermal stress may arise, adversely impacting the device performance and the circuit reliability [12].

III. DESIGN OF THE COOLING SYSTEM

In this analysis, the test chip contains a heater to represent the hot spot in the 3D-ICs. The heat generated from a heater is assumed to be equal to the heat generated from the hot spot of the 3D ICs [8]. The test chip with the heater is shown in Fig. 1. In this chip, the substrate is a 200µm-thick silicon with an area of 5x5 mm²; on top of the substrate, 0.20µm of SiO2 is deposited; on top of the oxide layer, a 0.18µm-thick serpentine-shaped tungsten layer with a total area of 1000 x 2000 µm² is deposited; another 0.15µm layer of SiO2 is deposited to cover the heater, on top of this oxide layer, two aluminum pads, with 125µm x 250µm area and 0.05µm thickness, are deposited and connected to the two ends of the heater using two vias through the SiO2 layer. Finally, a 7000Å SiO2 passivation layer is deposited.

Fig. 2 shows the cooling block structure which consists of blocks having an area of 40x40 mm². The inlet area is 2x38 mm². The thickness of the walls is 250µm for the top and bottom and 1000µm for the sides. The analysis was done
using the two cooling blocks: one built using SiO₂ and the other using diamond. We used three different liquid coolants in the study: water, R22, and Liquid Nitrogen (LN).

![Serpentine shaped tungsten heating wire on top of Si-substrate.](image1.png)

Fig. 1. Serpentine shaped tungsten heating wire on top of Si-substrate.

![Test chip sandwiched between two cooling blocks.](image2.png)

Fig. 2. Test chip sandwiched between two cooling blocks.

IV. SIMULATION AND ANALYSIS

Comsol multiphysics program is used to conduct the thermal analysis. The emissivity values utilized in the simulation are 0.79 for SiO₂ and 0.63 for diamond [13,14]. The convective heat transfer coefficient is assumed to be 50 W/m²K for all the coolants and 10 W/m²K for air. The heater, representing a hot spot, is assumed to dissipate 20 W of heat.

Our analysis predicted that a 20W hot spot would elevate the chip temperature to 2170 K, if no cooling liquid is admitted; see Fig. 3(a). This temperature would produce a thermal stress of 4552 MPa, as shown in Fig 3(b). This temperature is above the melting point of Silicon (1687 K), and the stress is close to the ultimate strength of Si (5000 MPa). This means that the IC will melt and evaporate before reaching this temperature, and no structure will even remain to experience the stress.

![Temperature & Thermal Stress of the IC without Cooling.](image3.png)

Fig.3: (a) Temperature & (b) Thermal Stress of the IC without Cooling.

To have a functioning IC, a cooling mechanism must be employed to cool the chip to the required operating temperature, within a reasonable time. Although, temperature can be reduced, an overshoot may take-place [8]. Overshoot of temperature may cause high thermal stress.

Thermal stress is investigated for a test chip sandwiched between two cooling blocks, as shown in Fig. 2. Temperature and the thermal stress are calculated for the first 15 s, after admitting the liquid coolants. 15s simulation time is considered for achieving temperature and stress saturation. With a 20W hot spot covering an area of 1000 x 2000 μm², the power density reaches 1000W/cm². Having this hot spot on a 25mm² chip results in a power dissipation of 80 W/cm² for the chip. In this analysis, the coolants were admitted at the same time the heater was powered up, where the system temperature was initially set to room temperature, 293K.

For SiO₂ cooling block, liquid was pumped at a velocity of 10 mm/s when the 20W heater is powered on, and temperature and thermal stress were calculated during the first 15 seconds. From Fig.4, we can observe that the temperature reaches an overshoot within 1 s, for water, R22, and liquid nitrogen cooling. The maximum temperature at the tungsten metal (hot spot) for water, R22, and liquid nitrogen reached 432K, 437K, and 375K and the corresponding coolants reached saturation at 432K, 416K, and 293K. Fig. 5 shows the corresponding stress.

![Hot spot temperature using SiO₂ cooling blocks for water, Liquid N₂, and R22 coolant with 10mm/s inlet velocity.](image4.png)

Fig. 4. Hot spot temperature using SiO₂ cooling blocks for water, Liquid N₂, and R22 coolant with 10mm/s inlet velocity.

![Thermal Stress using SiO₂ cooling blocks on the IC for water, Liquid N₂, and R22 coolant with 10mm/s inlet velocity.](image5.png)

Fig. 5. Thermal Stress using SiO₂ cooling blocks on the IC for water, Liquid N₂, and R22 coolant with 10mm/s inlet velocity.

Fig.5 shows the corresponding thermal stress. Within 1s, the thermal stress rises to maximum values of 324 MPa, 335 MPa, and 177 MPa for Water, R22, and Liquid N₂, respectively, then saturates to 324 MPa, 282 MPa, 33 MPa, as the temperature saturates. The figure shows that the stress value follows the temperature difference |TF -T₀| value (T₀=293K) as explained in Eq. (2). At 1s, |TF -T₀| has the values of 139K, 144K, 82K, for Water, R22, and Liquid N₂, respectively. This explains the reason for the higher thermal
stress for R22 compared to the other cooling cases at 1s. But as the temperature saturates, the difference from the original temperature goes down. This difference becomes smaller for R22 than water, leading to a lower stress. Cooling reduced the thermal stress to less than the ultimate tensile stress of tungsten of 980 MPa.

Aluminum 6061 is widely used in IC technology as interconnects and ohmic contacts. The ultimate tensile strength of that alloy is 310 MPa [15]. For cooling using SiO2, the stress exerted on aluminum pads reaches up to 274 MPa for Water, 232 MPa for R22, and 68 MPa for LN. This high stress might lead to gradual breakdown of aluminum pads.

It appears that inlet velocity of 10 mm/s does not pump enough liquid to cool the IC down. Therefore, it is reasonable to assume that if the inlet velocity increased to 100 mm/s, both the temperature and thermal stress will decrease. Fig. 6 shows that for 100 mm/s inlet velocity, the temperature saturates to 416 K and 361 K for water and R22, from initial temperature of 293 K, respectively. No significant overshoot of temperature for the two coolants take place. For LN, the temperature decreased to 207 K within 1 s and saturates. This is due to the fact LN temperature is 77 K, and the velocity was high enough to pump enough LN to bring the temperature to a much lower value than the initial value much quicker.

Fig. 7 shows the thermal stress for an inlet velocity of 100 mm/s. The temperature difference, |Tf -T0| at saturation are 123 K, 68 K, and 86 K for water, R22, and LN, respectively. The stress can be due to tension or compression. The thermal stresses on the IC is related to |Tf -T0| and, accordingly, are found to be saturated at 303 MPa, and 161 MPa, 233 MPa for water, R22 and LN, respectively. The study shows that thermal stress can exist for high as well as low final temperature. Hence, Deep cooling may cause thermal degradation of the IC.

![Fig. 6: Hot spot temperature after cooling using SiO2 cooling blocks with 100 mm/s inlet velocity. Coolant used are water, R22, and LN.](image)

![Fig. 7: Stress on tungsten when SiO2 cooling blocks with 100 mm/s inlet velocity. Coolant used are water, R22, and LN.](image)

Fig. 7 shows that the highest stress for tungsten results from water cooling. This is understood since the temperature difference in this case is the highest. It is also noticed that even though LN results in the lowest operating temperature, the stress is higher than the case with R22 cooling. This can be understood, if it is realized that the resulted temperature difference, |Tf -T0| is higher in case of LN. Because the operating temperature for LN is lower than the initial value, the type of stress is the opposite (compression rather than tension).

Stress on aluminum is also investigated, for inlet velocity of 100 mm/s. The results show that the stress on aluminum pads are 241 MPa, 107 MPa, & 262 MPa respectively for water, R22, and LN. Similar to the case of tungsten hotspot, R22 cooling provided the lowest stress, because the temperature difference is the lowest.

These results signified some issues. The first is that thermal stress exists not only because of higher operating temperatures, but also because of lower operating temperatures than the initial temperature. Therefore, if the intention is to operate at lower temperatures than room temperature, it may be advised to deposit metals at temperatures closer to the operating temperature. Second, the impact of metal elasticity and ultimate tensile stress are critical reliability factors that must be considered when designing 3D-ICs.
water, 87 MPa for R22, and 482 MPa for LN. Similar to the case of 10 mm/s, LN results in high thermal stress, due to the large difference between the initial and final temperature values of the hotspot.

The stress on the aluminum pads for 100 mm/s velocity is found to be 25 MPa for water, 118 MPa for R22, and 483 MPa for LN. Again, the stress due to LN cooling is the highest. The problem is that it is higher than the maximum tensile strength of the metal. This means the aluminum pads may break down and cause problem for the IC reliability, if cooled with LN.

V. CONCLUSION

The stress analysis for next generation 3D-ICs have been conducted using diamond and SiO2 cooling blocks. The result shows that, in order to ensure the stability of the IC thermal stress needs special consideration along with the operating temperature. The results also show that overcooling may cause metal breakdown. Therefore, the level of cooling should be considered, and the strength of materials used to build the IC should be evaluated. This study would not be complete without studying the material deformation due to the cooling process.

REFERENCES

[1] Tu, K. N. (2011). Reliability challenges in 3d IC packaging technology. Microelectronics Reliability, 51(3), 517-523.
[2] G. Huang, M. Baki and A. Naemi, "Power Delivery for 3D Chip stacks: Physical Modeling and Design Implication," IEEE, pp. 205-208, 2007.
[3] 2015 International Technology Roadmap for Semiconductors (ITRS) https://www.semiconductors.org/main/2015_international_technology_roadmap_for_semiconductors_irts/
[4] D. B. Tuckerman and R. F. W. Pease, "High-performance heat sinking for VLSI," in IEEE Electron Device Letters, vol. 2, no. 5, pp. 126-129, May 1981. doi: 10.1109/EDL.1981.25367
[5] C. Mandalapu and I. Motaleb, S. Hong, and R. Patti "Design Fabrication and Testing of a Liquid Cooling Platform for High Power 3D ICs," 2019 8th International Symposium on Next Generation Electronics (ISNE), Zhengzhou, China, 2019, IEEE Xplore. doi: 10.1109/ISNE.2019.8896611
[6] M. Matsuoka, K. Matsuda and H. Kubo, "Liquid immersion cooling technology with natural convection in data center." 2017 IEEE 6th International Conference on Cloud Networking (CloudNet), Prague, 2017, pp. 1-7. doi: 10.1109/CloudNet.2017.8071539
[7] B. Shi, A. Srivastava and A. Bar-Cohen, "Hybrid 3D-IC Cooling System Using Micro-Fluidic Cooling and Thermal TSVs," 2012 IEEE Computer Society Annual Symposium on VLSI, Amherst, MA, 2012, pp. 33-38. doi: 10.1109/ISVLSI.2012.29.
[8] S. Islam and I. Abdel-Motaleb, "Investigation of the Dynamics of Liquid Cooling of 3D ICs," 2019 8th International Symposium on Next Generation Electronics (ISNE), Zhengzhou, China, 2019, pp. 1-3. doi: 10.1109/ISNE.2019.8896425
[9] IUPAC. Compendium of Chemical Terminology, 2nd ed. (the “Gold Book”). Compiled by A. D. McNaught and A. Wilkinson. Blackwell Scientific Publications, Oxford (1997). Online version (2019-) created by S. J. Chalk. ISBN 0-9678550-9-8.
[10] S. Senturia, Microsystem Design, Chapter 8, Kluwer Academic Publishers, 2001.
[11] F., Carter, Giles (1991). Materials science & engineering. Paul, Donald E. [Materials Park, Ohio]: ASM International. ISBN 9780871703996. OCLC 555220209.
[12] The Rational mechanics of Flexible or Elastic Bodies, 1638–1788: Introduction to Leonardi Euleri Opera Omnia, vol. X and XI, Seriei Secundae, Orell Fussli.
[13] P. Siegal, T. Friedmann, and P. Sullivan, “Diamond and diamond-like carbon films for advanced electronic applications, report, March 1, 1996; Albuquerque, New Mexico.
[14] The Engineering Toolbox, https://www.engineeringtoolbox.com
[15] All About 6061 Aluminum, (Properties, Strength and Uses), https://www.thomasnet.com/articles/metals-metal-products/6061-aluminum/