investigation of The Impact of Liquid Cooling on The integrity of Three-Dimensional integrated Circuits (3D-ICs)

A K M Toufiqul Islam
toufiqsakib41@gmail.com

Follow this and additional works at: https://huskiecommons.lib.niu.edu/allgraduate-thesesdissertations

Recommended Citation

This Dissertation/Thesis is brought to you for free and open access by the Graduate Research & Artistry at Huskie Commons. It has been accepted for inclusion in Graduate Research Theses & Dissertations by an authorized administrator of Huskie Commons. For more information, please contact jschumacher@niu.edu.
The thesis investigates the dynamics of liquid cooling of the Three-Dimensional Integrated Circuit (3D-IC). The use of 3D-IC technology is one of the most viable solutions for achieving high integration density of integrated circuits, without facing the unavoidable problems arising from shrinking the device dimensions. However, 3D-ICs suffer from other types of problems. One of these problems is heat generation, creating extremely hot areas called hot spots.

For high hot spots, effective thermal management techniques need to be developed. Conventional thermal management of surface cooling of chips using heat sinks and fans may reduce the surface temperature, but the embedded hot spots may not be cooled to safe operating temperature.

The objective of this thesis is to investigate the impact of hot spots on 3D-ICs and to assess the effectiveness of different cooling techniques in reducing the temperature of the hot spots to acceptable levels. For the thermal management experimentations, a 1000 W/cm² hot spot was created on a 5mmx5mm silicon chip. The chip temperature was simulated without any type of cooling. Then two mechanisms of liquid cooling were employed. The first was using a microchannel cooling block on top and another one at the bottom. Then, three different liquid coolants (water, R22, and liquid nitrogen) were used. The cooling blocks were built using SiO₂ and diamond. The second one was using a chip embedded microchannel inside the 3D-ICs, for
liquid coolants to flow to cool the hot spots. The same three coolants were used with the embedded micro-channel techniques as well. The temperature change of the cooled chip and thermal stress were investigated.

The results show that without cooling, the chip temperature will reach over 2000 K after 1s, leading to the evaporation of the material. Using microchannel cooling block, it has been found that diamond microchannel cooling block ensures safe operating temperature whereas SiO$_2$ microchannel cooling block ensures safe operating temperature for liquid nitrogen cooling only. The stress on the IC while utilizing these microchannel cooling blocks has been observed as well. The study shows that utilizing a rapid cooling technique does not guarantee the safer operation of the IC due to the thermal stress. Rather, the focus should be keeping the temperature difference lower from the initial temperature of the IC. A higher temperature difference makes the IC subject to severe stress. The chip-embedded channel cooling for the 3D-IC has been investigated as well. Furthermore, the adaptability of the cooling blocks for the next-generation integration is discussed. The result shows that even without using the diamond cooling block, safe operating temperature and reduced thermal stress can be ensured using chip-embedded cooling block.
ACKNOWLEDGEMENTS

I would like to thank the Almighty for giving me the opportunity and the strength to pull this off. I also would like to thank my family and my friends for supporting me always. Last but not the least, I would like to thank my adviser, Dr. Ibrahim M. Abdel-Motaleb and Senior Associate Dean, Dr. Mansour Tahernezhadi, for letting me to be a part of NIU and giving me an opportunity to stand strong in life. I will be forever indebted to them. Specially, I would like to thank, Dr. Ibrahim M. Abdel-Motaleb, for being extremely patient with me and helping me in every step of my thesis.
DEDICATION

To my family
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LIST OF FIGURES</strong></td>
<td>vi</td>
</tr>
</tbody>
</table>

**Chapter**

1. THREE-DIMENSIONAL INTEGRATED CIRCUIT (3D-IC) ..............................................1
   1.1 Introduction ..................................................................................................1
   1.2 3D-IC Structure ..........................................................................................2
   1.3 3D-IC: Beyond Moore Scaling .....................................................................3
   1.4 Advantages of the 3D-IC .............................................................................4
   1.5 Thermal Management Problem of the 3D-ICs ..............................................5
   1.6 Literature Review .......................................................................................6
   1.7 Proposed Work ............................................................................................10
   References .........................................................................................................11

2. INVESTIGATION OF THE DYNAMICS OF LIQUID COOLING OF 3D ICS ............16
   2.1 Introduction ..................................................................................................16
   2.2 Heat Transfer Mechanism ..........................................................................17
   2.3 Design of the Cooling System .....................................................................18
   2.4 Simulation and Analysis .............................................................................20
   2.5 Silicon Dioxide Cooling Block ....................................................................21
   2.6 Diamond Cooling Block .............................................................................23
   2.7 Assessment of Cooling Block Efficiency .....................................................24
   2.8 CONCLUSION .............................................................................................27
   References .........................................................................................................28

3. THERMAL STRESS ANALYSIS OF LIQUID-COOLED 3D-ICS ..........................29
   3.1 Introduction ..................................................................................................29
   3.2 Thermal Stress ............................................................................................31
   3.3 Design of the Cooling System .....................................................................31
   3.4 Simulation and Analysis .............................................................................33
   3.5 Conclusion ...................................................................................................41
   References ..........................................................................................................41
4. NUMERICAL ANALYSIS OF LIQUID COOLING OF 3D-ICS USING EMBEDDED CHANNELS ..........................................................................................................................44
  4.1 Introduction ..........................................................................................................................44
  4.2 Heat Transfer and Thermal Stress .........................................................................................45
  4.3 Design of the Cooling System ...............................................................................................47
  4.4 Simulation and Analysis ........................................................................................................48
  4.5 Low Inlet Velocity .................................................................................................................49
  4.6 Inlet Velocity 100 mm/s .......................................................................................................55
  4.7 Assessment of Cooling Using Embedded Microchannels ......................................................58
  4.8 Conclusion ...........................................................................................................................60
References ..................................................................................................................................61
5. VERIFICATION ........................................................................................................................64
  5.1 Conclusion ...........................................................................................................................67
  5.2 Future Works .......................................................................................................................67
References ..................................................................................................................................68
### LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1. 1: 3D-IC consists of 2D-ICs that are thinned, bonded together, and interconnected with TSVs distributed within the planes of the 2D-ICs [6].</td>
<td>1</td>
</tr>
<tr>
<td>Figure 1. 2: 3D-IC structure [11].</td>
<td>3</td>
</tr>
<tr>
<td>Figure 1. 3: 3D IC temperature may rise up to several thousand Kelvins.</td>
<td>6</td>
</tr>
<tr>
<td>Figure 2. 1: The serpentine-shaped tungsten heating block with aluminum contacts at the two ends.</td>
<td>19</td>
</tr>
<tr>
<td>Figure 2. 2: The geometry of the cooling block, with the chip in between the lower and upper channels. Channel dimensions are 40mm x 40 mm x 2mm; wall thickness is 250μm.</td>
<td>19</td>
</tr>
<tr>
<td>Figure 2. 3: COMSOL mesh of the geometry. Size of the meshes change to ensure accurate simulation results.</td>
<td>21</td>
</tr>
<tr>
<td>Figure 2. 4: Hot spot temperature for water, LN, and R22 coolant with 10 mm/s inlet velocity for SiO₂ cooling block.</td>
<td>22</td>
</tr>
<tr>
<td>Figure 2. 5: Hot spot temperature for water, LN, and R22 coolant with 100 mm/s inlet velocity for SiO₂ cooling block.</td>
<td>23</td>
</tr>
<tr>
<td>Figure 2. 6: Hot spot temperature for diamond block when coolant is admitted before powering the chip. Inlet velocity is 10 mm/s for diamond cooling block.</td>
<td>24</td>
</tr>
<tr>
<td>Figure 2. 7: Hot spot temperature for diamond block when coolant is admitted before powering the chip. Inlet velocity is 100 mm/s for diamond cooling block.</td>
<td>24</td>
</tr>
<tr>
<td>Figure 2. 8: SiO₂ cooling block at 10 mm/s inlet velocity observed for test chip at different wattages.</td>
<td>25</td>
</tr>
<tr>
<td>Figure 2. 9 : SiO₂ cooling block at 100 mm/s inlet velocity observed for test chip at different wattages.</td>
<td>25</td>
</tr>
<tr>
<td>Figure 2. 10: Diamond cooling block at 10 mm/s inlet velocity observed for test chip at different wattages.</td>
<td>26</td>
</tr>
<tr>
<td>Figure 2. 11: Diamond cooling block at 100 mm/s inlet velocity observed for test chip at different wattages.</td>
<td>27</td>
</tr>
<tr>
<td>Figure 3. 1: Serpentine-shaped tungsten heating block on top of silicon substrate.</td>
<td>32</td>
</tr>
<tr>
<td>Figure 3. 2: Test chip sandwiched between two cooling blocks.</td>
<td>32</td>
</tr>
</tbody>
</table>
Figure 3.3: (a) Temperature and (b) thermal stress of the IC without cooling. ........................................... 33

Figure 3.4: Hot spot temperature using SiO$_2$ cooling blocks for water, liquid N$_2$, and R22 coolant with 10mm/s inlet velocity. ........................................................................................................... 34

Figure 3.5: Thermal stress using SiO$_2$ cooling blocks on the IC for water, liquid N$_2$, and R22 coolant with 10mm/s inlet velocity. ........................................................................................................... 35

Figure 3.6: Hot spot temperature after cooling using SiO$_2$ cooling blocks with 100 mm/s inlet velocity. Coolant used are water, R22, and LN. ........................................................................................................... 36

Figure 3.7: Stress on tungsten when SiO$_2$ cooling blocks with 100 mm/s inlet velocity. Coolant used are water, R22, and LN. ........................................................................................................... 37

Figure 3.8: Hot spot temperature for water, liquid nitrogen, and R22 coolant at 10 mm/s inlet velocity using diamond cooling block. ........................................................................................................... 38

Figure 3.9: Stress on the IC for water, liquid nitrogen, and R22 coolant at 10 mm/s inlet velocity using diamond cooling block. ........................................................................................................... 39

Figure 3.10: Hot spot temperature for water, liquid nitrogen, and R22 coolant at 100 mm/s inlet velocity using diamond cooling block. ........................................................................................................... 40

Figure 3.11: Stress on the IC for water, LN, and R22 coolant at 100 mm/s inlet velocity using diamond cooling block. ........................................................................................................... 40

Figure 4.1: Serpentine shaped tungsten heating block (yellow) and copper Wire (Red). .......... 47

Figure 4.2: (a) Heater sandwiched between two cooling blocks; (b) cooling block geometry; inlets are 100μm above the bottom of the silicon substrate. Inclined lines are 350μm. ........................................................................................................... 48

Figure 4.3: Water cooling with 10mm/s pumping velocity (a) temperature and (b) thermal Stress map across Cu and tungsten. ........................................................................................................... 50

Figure 4.4: R22 cooling with 10mm/s pumping velocity (a) temperature and (b) thermal Stress map across Cu and tungsten. ........................................................................................................... 51

Figure 4.5: LN cooling with 10mm/s pumping velocity (a) temperature and (b) thermal Stress map across Cu and tungsten. ........................................................................................................... 51

Figure 4.6: Cu wire temperature at 10mm/s inlet velocity. ................................................................. 52

Figure 4.7: Cu wire stress at 10mm/s inlet velocity ............................................................................... 54

Figure 4.8: Water cooling with 100mm/s pumping velocity (a) surface temperature map and (b) thermal stress map across Cu and tungsten. ........................................................................................................... 55
Figure 4.9: R22 cooling with 100mm/s pumping velocity (a) surface temperature map and (b) thermal stress map across Cu and tungsten. .......................................................... 56

Figure 4.10: LN cooling with 100mm/s pumping velocity (a) surface temperature map and (b) thermal stress map across Cu and tungsten. .......................................................... 57

Figure 4.11: Cu wire temperature at 100mm/s inlet velocity. .................................................. 58

Figure 4.12: Cu wire stress at 100mm/s inlet velocity ............................................................ 58

Figure 4.13: Chip-embedded cooling block at 10 mm/s inlet velocity for test chip at different wattages.................................................................................................................... 59

Figure 4.14: Chip-embedded cooling block at 10 mm/s inlet velocity for test chip at different wattages.................................................................................................................... 59

Figure 5.1: Simulated temperature for 20W power. .............................................................. 64

Figure 5.2: (a) Calculated temperature of the IC from reference paper and (b) simulated temperature for the IC. ........................................................................................................ 65

Figure 5.3: (a) Infrared image of the DC/DC module and (b) simulated temperature of the DC/DC module. ........................................................................................................... 66
NOTES ON PREVIOUSLY PUBLISHED MATERIAL


Chapter 1

THREE-DIMENSIONAL INTEGRATED CIRCUIT (3D-IC)

1.1 Introduction

Gordon E. Moore, in 1965 stated that, the packing density on a microchip increases twice about every two years, whereas the price declines by half [1]. It has been more than 50 years since the statement was made. The continuation of “Moore’s Law” has been achieved by scaling the transistors’ and interconnects’ dimensions. However, achieving further dimensional scaling is reaching its physical limit. Power and performance requirements for future ICs make further scaling according to the conventional process almost impossible [2,3]. Since dimensional scaling has become almost unattainable, researchers have realized that one remedy to this problem is the stacking of multiple 2D-ICs on top of each other, allowing vertical integration or so-called 3D integration. This emergence of the 3D integration has resulted in the development of the 3D-IC technology. Figure 1.1 presents a distinction between the regular 2D-IC and the 3D-ICs as well as a generalized fabrication process of the 3D-ICs. Instead of utilizing the traditional integration process, through silicon via (TSV) is utilized for integration.

Figure 1.1: 3D-IC consists of 2D-ICs that are thinned, bonded together, and interconnected with TSVs distributed within the planes of the 2D-ICs [6].
The 3D-IC technology is believed to be the most practical solution to overcome the problems facing ICs. 3D-IC technology ensures that integrated circuit can be scaled beyond Moore’s law [4,5].

1.2 3D-IC Structure

The 3D integration allows the increase of volumetric density of the devices and packages. By allocating dedicated stack-layers for specific technology, such as memories, converters, microprocessors, Microelectromechanical system (MEMS), and microwave, each stack can be optimized separately and later connected with each other by shorter length interconnects [7,8]. A 3D-IC is a number of 2D silicon MOS (metal-oxide semiconductor) integrated circuits (2D-IC) stacked on top of each other and vertically interconnected through Cu-Cu connections formed by through-silicon via (TSV) technology [9]. Thus, more interconnections between several dies are made possible in the same footprint of the conventional two-dimensional process [10].

Hence, 3D-IC can boost up the overall performance of the integrated systems, and 3D integration has become one of the most feasible, economical, and promising technologies to perform unprecedented system-level integration. Figure 1.2 shows a 3D-IC structure. The figure shows that each functional layer of the IC is stacked and later connected through the vias. The heat sink lies at the bottom of the surface.
1.3 3D-IC: Beyond Moore Scaling

The increasing of the packing density requires the shrinkage of both the components and interconnects. Scaling resulted not only in a better performance and reduced power consumption, but also reduced the fabrication cost by almost 30% as well. Usually, this scaling is achieved by shrinking down all the nodes of a circuit by a factor of 0.7. Reducing the length and width by a factor of 0.7 ensures the packing of double the number of the components within the same area. However, for 2D-ICs, interconnect bottleneck started becoming a serious concern with the scaling and started to degrade the chip performance substantially due to the increase of the signal delays [12]. Moreover, short channel effect and gate leakage current continue to keep deteriorating the chip performance. Therefore, the continual reliance on scaling will not be a viable solution in the near future [13, 14].

Figure 1. 2: 3D-IC structure [11].
3D-IC provides a new way of integration by stacking several 2D-ICs and establishing connection between them. The goal of the 3D integration is to shift the focus from device scaling to circuit and system scaling. With an optimized 3D integration, issues related to interconnect latencies, packing density, heterogeneous integration, and others can be addressed. Different components can be fabricated at different wafers and connected utilizing shorter interconnects at the 3rd dimension, which can provide faster on-chip communication [10,14].

1.4 Advantages of the 3D-IC

The primary advantage of the 3D-IC is the reduction of the device footprint and enhancement of integration density. Moreover, the vertical interconnections allow shorter interconnect length between the components of the 3D-IC and reduces the resistive elements in the interconnect. As a result, the 3D-ICs are superior in dealing with the interconnect bottleneck challenge, so-called the RC challenge [15].

The advancement in fabrication technology has made this vertical integration possible. For example, heterogeneous integration of different components and the ICs have been made possible. This can allow for the optimization of each individual technology, which will in turn lead to the enhancement of the whole system performance. Furthermore, the utilization of the vertical (3D) integration has dramatically improved several NAND flash memories, computer processors and mobile devices [16].

In other words, using the 3D IC technology, optimized system with higher speed and lower power consumption can be obtained [9,10]. In addition, a substantial increase in routing and bandwidth, allowing superior performance and capability at reduced power and most importantly
in a smaller footprint compared to the regular 2D-IC technology, can be achieved [17-18]. At a glance, the advantages are:

- Higher Integration Density,
- Reduced Interconnect Wire,
- Superior Processing Speed,
- Less Power Consumption,
- Heterogeneous Integration,
- Logic Redundancy.

Overall, 3D-IC has smaller footprint, is more cost effective, results in faster chip communication and smaller parasitic elements due to smaller interconnect wires and allows for efficient heterogeneous integration along with optimized system integration.

1.5 Thermal Management Problem of the 3D-ICs

Although 3D-ICs have solved many problems related to reaching the physical limits of scaling, it is faced with some challenges. Because 3D-ICs increase the volumetric density of components, it is prone to what is called “hot spots.”

A hot spot is created when an area is exposed to heat dissipated from a large number of elements, placed in a small area, increasing the energy density to very high values in that particular area. Hot spots do not require high energy dissipation across the whole chip, rather it only requires high energy dissipation within a small area. This area can be as small as a few micrometer squares. These hot spots may damage the elements in this area, rendering the IC faulty. Furthermore, the high temperature of the hot spot is responsible for increasing the thermal resistance between the junctions of the 3D-IC, thermal stress, Interconnect delay, etc. [19,20]
Figure 1. 3D IC temperature may rise up to several thousand Kelvins.

In 2D-ICs, devices, interconnects are located at the surface, hence air-cooling using heat sinks can be effective in removing the generated heat. However, for 3D-IC, the hot spots can reside even inside the 3D-IC covered with several layers of IC stacks, rendering air cooling ineffective. Therefore, for 3D-ICs, different techniques of cooling are required to reach the hot spots and cool it down to acceptable operating temperature.

1.6 Literature Review

Several means of cooling mechanism have been employed so far to resolve the thermal issues related to the 3D-IC hot spots. Some of the widely utilized cooling mechanisms employed for 3D IC cooling are- microchannel cooling, thermal TSV cooling, hybrid cooling, MEMS cooling, liquid immersion cooling etc. In this section, these cooling mechanisms will be discussed.

Shi et al. [21] have proposed a hybrid 3D-IC cooling scheme combining microchannel (acting as a heat sink) and thermal TSV (for heat dissipation). The work addresses the issues related to the microchannel and TSV-based cooling. In an effort to ensure an efficient cooling of the 3D-
IC, they have employed both cooling mechanisms at the same time. The focus of the work is to provide adequate cooling utilizing minimum pumping power under a given pressure drop and acceptable number of cooling TSVs and efficient placement of the thermal TSVs. The calculation of the temperature and Pumping power has been measured utilizing algorithm-based modeling and the improvement study has been performed using an iterative approach. With the heuristic approach, co-optimization of microchannel and thermal TSV, TSV placement and sizing of the TSV information are extracted. The result shows a hot spot temperature of 84°C, which is more efficient than utilizing either microchannel or thermal TSVs. However, this high hot spot temperature might result in greater thermal stress for some components such as metal contacts and interconnects in the 3D IC [20]. The work has not considered thermal stress. Furthermore, the cooling dynamics of the proposed cooling block has not been addressed [22].

Sekar et al. [23] have utilized an aggressive pumping technology, incorporating high pressure drop and flow rates for the cooling of the 3D-ICs. Furthermore, it discusses some of the trade-offs related to TSV diameter and junction-to-ambient thermal resistance. The experiment is performed on different operating frequency while initiating the microchannel cooling block. The temperature varies from 47°C to 88°C depending on the operating frequency. This rapid cooling enabled by the rapid pumping ensures an acceptable operating temperature. However, the performance of the pump degrades with time due to this rapid pumping activity. Moreover, the system is expected to be expensive according to the work presented by Garimella et al. [24]. But the biggest problem is the reliability issues associated with the pumping technology.

The micropin-based microfluidic cooling block module proposed by Wan et al. [25] has experimented the effect of leakage power, which is found to be 55.8% of the dynamic power of the chip. They also have found that ambient heat transfer coefficient plays a very trivial role in
terms of thermal and electrical performance of the 3D-ICs and the performance of the chip strongly depends on the ambient temperature. The study also shows that by increasing the height of the micropin and transversal & longitudinal spacing between the pins, flow resistance of the coolants can be reduced, and the mass flow rate can be increased as a result. Spacing ranging from 150um to 225um, both transversally and longitudinally, results in a chip operating temperature varying from 86°C to 111°C. However, the study does not concern itself with the effect of thermal stress even though the temperature indicates a significant rise.

Kudo et al. [26] proposed a MEMS fabrication technology-based closed channel cooling system. The cooling system utilizes electro-osmotic flow (EOF) pump. In order to imitate the heat dissipation of the 3D-IC, they have employed titanium heaters connected with aluminum contact pads for supplying the voltage. The utilized coolant is isopropyl alcohol. The driving capability of the EOF pump is measured out to be 1 \times 10^4 \text{ Pa} with a mass flow rate of 38 \text{ µl/min}. The cooling block resolves IC heat at a rate of 140 \text{ W/cm}^2. The temperature of the chip at the front side varies from 50°C to 100°C and later gets saturated to 50°C. The temperature of the chip at the back side gets heated up to 120°C and gets saturated to 100°C. This huge change in temperature will result in even greater stress, as shown in [20]. Furthermore, isopropyl alcohol is extremely flammable in the presence of a heated interface [27]. Therefore, this coolant might not be appropriate for 3D-IC environment. Another advantage of the EOF pump is that it requires very high voltage to be operated. The minimum requirement for the EOF pump to be operated is 650V [28].

An et al. [29] present a two-phase liquid immersion cooling channel model, developed on ANSYS Fluent. On top of a computing board having an area of 7cm (width) \times 20 \text{ cm (height)}, square heat sources having an area of 5cm\times5cm are designed. These square heat sources denote the CPU package of the 3D-ICs. Two different arrangements of the processors are investigated in
the study. The purpose of performing the study on these two arrangements, is to ensure that the study considers not only about a single hot spot but also about several hot spots created due to the high packing density of the transistors. In order to imitate the high packing density of the transistors, two heaters are mounted vertically, whereas the single hot spot is defined by just a single heater. The coolant utilized for the simulation is 3M Novec7000, a dielectric coolant with low boiling point (34°C). The coolant also has a high latent heat of vaporization (142 kJ/Kg) [30]. The cooling block does not employ any heat sink or cooling extensions along with the socket. The extensions are not needed as all the sockets are fully immersed in the coolant. To perform the adaptability test of the cooling block for the future integration, the heater power is varied from 50W to 300W. A gap of 4cm between each computing board is maintained to leave space for boiling of the coolants. Eulerian model is utilized to solve multiphase flow and non-equilibrium boiling equations and k-epsilon model is utilized for turbulent equations. A transient of 50ms is utilized with phase-coupled study to consider the evaporation of the coolant. The simulation is run until the temperature reaches a steady state. For the heating power of 50W to 300W, the temperature varies from 37°C to 119°C. The greater substrate area of the computing board (7cmX20cm) increases the contact area for convection, which contributes to much better operating temperature of the IC. However, the study lacks the consideration of thermal stress at the transient, although it encourages to further research on 3M Novec7000.

Mandalapu et al. [31] have designed a liquid cooling block that ensures an effective operating temperature of a chip having a hot spot of 65 W/cm². The test chip is based on an 8-inch silicon wafer with SiO₂ grown on its surface. 32 test chips are fabricated on top of the substrate where each chip has an area of 25.6mmX26mm. Each chip consists of thirty 10W heaters (1X2mm²) and six 20W (2X2mm²) heaters; 42 temperature sensors (200X200 um²) are employed
along with aluminum pads (50um wide, 1um thick) to reduce joule heating. The sensors are placed on the top of the 10W heaters and in between the 20 W heaters. The total power consumption of the chip is 420W; 36 cooling blocks are utilized for these 36 heaters. The work employs both single-phase (water cooling) and double-phase cooling technique. For double-phase cooling, R22 has been utilized. The temperature of the chip, while varying power from 0 to 21.6W ranges from 27°C to 204°C, utilizing single phase cooling. For the double phase cooling, temperature across the 10W heater varies from 41°C to 43°C. For 20 W heater, the temperature saturates at 52°C. Moreover, the greater dimension of the wafer has attributed to greater heat transfer by acting as a huge heat sink. This work has not considered the thermal stress during the transient and steady state.

The aforementioned works describe several cooling solutions for the 3D-ICs. However, the mere consideration of 3D-IC cooling may not completely solve the issue of hotspots. This is because heating/cooling process is dynamic, where the temperature may increase before it decreases, and the temperature of different regions and materials may be different. The variation of temperature, dynamic or static, may cause stress, which may lead to material fatigue or structural damage. Therefore, without investigating the impact of cooling on the integrity of the ICs, it is hard to ensure the acceptable reliability of 3D-ICs.

1.7 Proposed Work

The objective of this thesis is to provide an attempt towards devising an efficient cooling mechanism and address several issues related to the 3D-IC cooling, such as cooling dynamics, the effect of thermal stress, temperature and stress during the transient period, the problems related to the deep cooling and cooling blocks’ efficiency for the next-generation integration. The aforementioned issues are discussed utilizing two different liquid cooling methods: microchannel
cooling and embedded-channel cooling. Using the multiphysics numerical simulation program, COMSOL, these techniques for cooling hots spots with a power density to up to 1000 W/cm², where 20 W is dissipated in 1x2 mm², are developed and tested. Different cooling block materials and different liquid coolants are used to assess the effectiveness of each cooling technique.

Microchannels that carry the coolant are built on top and bottom of the 3D-IC, where different coolants are used, and the resultant temperature values are obtained. Water, R22, and liquid nitrogen are used as liquid coolants. In Chapter 2, the investigation of the dynamics of liquid cooling are presented. In Chapter 3, the results from the investigation of thermal stress due to thermal dissipation and cooling are shown. In Chapter 4, cooling using embedded microchannels is presented. Chapter 5 shows how the accuracy of COMSOL simulation is verified. This is followed by the conclusion and future work.

References


[12] “3D ICs The Industry’s Only Homogeneous and Heterogeneous 3D ICs.”

[13] “The Monolithic 3D Advantage Monolithic 3D is far more than just an alternative to 0.7x scaling!!!”
http://www.monolithic3d.com/3d-ic-edge1.html


https://nanoscalereslett.springeropen.com/articles/10.1186/s11671-017-1831-4

[18] "Is 3D IC The Next Big Profit Driver?" https://www.eetimes.com/is-3d-ic-the-next-big-profit-driver/


[27] “Isopropyl Alcohol Safety Tips from MSDSonline.”


Chapter 2

INVESTIGATION OF THE DYNAMICS OF LIQUID COOLING OF 3D ICS

Although 3D-IC technology can provide very high integration density, it suffers from having hot spots that may reach thousands of degrees. To manage this heat, it is necessary to study the dynamics of cooling and thermal behavior of the ICs. In this study, we report on the dynamics of microchannel liquid cooling using water, R22, and liquid nitrogen. The study shows that using diamond cooling blocks ensures normal operating temperature of 60°C or less, using any of the above coolants. Using SiO₂ blocks, only liquid nitrogen can provide acceptable operating temperatures. The study also shows that liquid latent energy and inlet velocity play a major role in the cooling dynamics.

2.1 Introduction

3D-ICs provide a way to achieve the high integration density craved for by industry. But such high density produces local hot spots with temperatures reaching thousands of Kelvins. Hot spots may cause temperature variations across the chip that may affect the overall performance of the IC. Hot spots may also affect the resistance of the interconnects, slowing the communication between the different parts of the IC. This situation gets even more critical, since according to the International Technology Roadmap for Semiconductors (ITRS), for a single package, the power density may need to increase to up to 10⁶W/cm² to meet future applications [1]. To address these issues, thermal management techniques need to be developed to ensure that no hot spot can exceed the IC operating temperature.
Conventional 2D-IC cooling methods employ heat sinks on the face of the IC to transfer heat from the ICs to the outside environment. Because the power dissipation is not excessively high, this technique can maintain chip temperature within the operating temperature range. However, in the case of the vertically stacked 3D-IC chips, there are multiple heated surfaces, and it is almost impossible to place those bulky heat sinks on top of each of the stacked layers. Cooling from the surface will not be as efficient, since heat coming from the inside layers will have a long heat path that complicates the cooling process, unless fast heat removal techniques can be employed [2].

To solve the cooling issues of the 3D-ICs, other cooling techniques have been proposed. Among them are microelectromechanical systems (MEMS) technology, embedded microchannels [3], liquid immersion cooling [4], and microfluidic cooling using thermal TSVs [5].

In this study, we report the results of investigating the dynamics of cooling a 1000 W/cm² local hot spot, using SiO₂ and diamond cooling blocks, and employing water, liquid nitrogen (LN), and R22 as coolants. The study is conducted using the advanced multiphysics numerical analysis program, Comsol™.

2.2 Heat Transfer Mechanism

Heat transfer between the cooling blocks and the 3D-IC can be due to a combination of conduction, convection, and radiation. Heat conduction describes the flow of heat through materials. In general, conduction is governed by

\[
\frac{Q}{t} = k A \frac{\Delta T}{l} \ldots (1)
\]

Here, Q is the heat energy, t is the time, k is the thermal conductivity (W/mK), A is the cross-sectional area of the flowing path, l is the length of the heat flowing path, and \( \Delta T \) is the temperature difference between the beginning and end of the path in Kelvins (K).
Convection heat flow is a result of the transport of heat by the motion of gas or fluid molecules. Heat convection can be described by the following equation:

\[
Q = h A \Delta T \quad \text{(2)}
\]

In this equation, \( h \) is the convective heat transfer coefficient \((\text{J}/\text{m}^2\text{K})\) [6], \( A \) is the object’s cross-sectional area that carries the fluid, and \( \Delta T \) is the temperature difference between fluid and heated surface.

Radiation is the emission of heat from hot body to the ambient. Radiation power can be obtained from the relation

\[
P = \varepsilon \sigma A T^4 \quad \text{(3)}
\]

Here, \( \varepsilon \) is the emissivity factor, \( \sigma \) is Stefan-Boltzmann constant \((=5.67 \times 10^{-8} \text{ W} \cdot \text{m}^2 \cdot \text{K}^{-4})\), \( A \) is the surface area of the radiating body and \( T \) is the temperature.

### 2.3 Design of the Cooling System

The 3D-IC structure is represented by a 1000x2000 \( \mu \text{m}^2 \) tungsten heater on 5x5 mm\(^2 \) Si wafer. The heater is made of 200\( \mu \text{m} \) Si substrate with a 3000 Å of SiO\(_2\) on its top. Next, a 0.18\( \mu \text{m} \) serpentine-shaped tungsten heater is deposited and covered with 1500 Å of SiO\(_2\). Aluminum pads of an area of 125x250\( \mu \text{m}^2 \) and thickness of 0.05\( \mu \text{m} \) are deposited and connected to the ends of the heater using vias in the SiO\(_2\) layer. Finally, a 7000 Å SiO\(_2\) layer is deposited for passivation. The area of the silicon chip is 5mmx5mm and the tungsten heater has the shape shown in Figure 2.1, which shows the heater structure on top of the silicon wafer. The widths of the wire of the heater is 120\( \mu \text{m} \) and the separation between the wires is 115\( \mu \text{m} \). The aluminum pads are shown at the ends.
Figure 2.1: The serpentine-shaped tungsten heating block with aluminum contacts at the two ends.

Figure 2.2 shows the cooling block, where it has two microchannels with the heater chip sandwiched between them. Both microchannels’ length and width are 40 mm, and the openings are 2x40 mm². The walls of the microchannels are 250μm thick. The liquid enters from the inlet and exits from the other side.

Figure 2.2: The geometry of the cooling block, with the chip in between the lower and upper channels. Channel dimensions are 40mm x 40 mm x 2mm; wall thickness is 250μm.
Two cooling blocks are used; the first one is made of SiO\textsubscript{2} and the second one is of diamond. Three liquid coolants are used: water, R22, and liquid nitrogen (LN). The size of the block is made large to ensure laminar flow of the coolant at the inlet. Turbulent flow is left for another study.

### 2.4 Simulation and Analysis

Figure 2.3 shows Comsol meshing structure for the cooling blocks, where the size of the mesh decreases at the edge to ensure accurate simulation. In this simulation, we used emissivity value of 0.79 for SiO\textsubscript{2} [6] and 0.63 for diamond [7]; the thermal conductivity used are 1.4 W/m.K for SiO\textsubscript{2}, 990 W/m.K for diamond, 130 W/m.K for Si, 174 W/m.K for tungsten and 238 W/m.K for aluminum. The convective heat transfer coefficient is assumed to be 50 W/m\textsuperscript{2}K for all the coolants and 10 W/m\textsuperscript{2}K for air.

The hot spot temperature is simulated without coolant, with just natural air convection and radiation. The results show that with 20W, temperature reached 2170 K or 1896°C. In reality, the chip would be evaporated before reaching this temperature.

Next, temperature is simulated when liquid coolants are admitted. The hot spot temperature is then monitored for 15s. The heater power is controlled by the input voltage. Here, the maximum power used is 20 W, and this resulted in 1000 W/cm\textsuperscript{2} hot spot, or 80W/cm\textsuperscript{2} across the 5mmx5mm chip. The simulation is performed for the SiO\textsubscript{2} and diamond cooling blocks. For each block, water, R22, and LN are used.
Figure 2. 3: COMSOL mesh of the geometry. Size of the meshes change to ensure accurate simulation results.

2.5 Silicon Dioxide Cooling Block

The thermal analysis of the SiO$_2$ cooling block is studied, where the coolants are admitted at the same time the heater gets powered up. The initial temperature of all regions and materials, including the ambient, are set to the room temperature, 293 K. Inlet velocities of 10 mm/s and 100 mm/s are used to investigate the effect of the velocity on the dynamics of the cooling; see Figure 2.4 and Figure 2.5, respectively.

For 10 mm/s velocity, it is noticed that the hot spot temperature overshoots. For LN, temperature overshoot reaches to 375 K before settling to 293 K. For R22, temperature increases to 438 K before it decreases to a constant value of 414K. For water, the temperature increases to 436 K before saturating to 434 K.

This result appears to be for a single-phase cooling, where the cooling is done using liquid. If we have two-phase cooling, where both liquid and gas contribute to the cooling process, we expect to have lower temperature. This in fact what has been noticed in [8]. This decrease in temperature can be attributed to the high latent energy of liquid, or to the high energy required to
evaporate the liquid. It is known that water has high latent energy of 2230 J/gm. For R22, the latent energy is 233 J/gm and for LN it is 200 J/gm. To convert liquid to gas, water needs 10 times the energy needed for R22 or LN. Therefore, water cooling may be more efficient, but to reach the stage of evaporation, temperature should increase to 100°C, which may not be acceptable for IC operation.

![Figure 2.4: Hot spot temperature for water, LN, and R22 coolant with 10 mm/s inlet velocity for SiO₂ cooling block.](image)

When the velocity increases to 100 mm/s, overshoots are almost eliminated; temperature rises only from 293 K to 304 K. For LN, the temperature is reduced to a saturation value of 207K in 1 second. For R22, the temperature does not decrease; instead, it increases to 362 K. For water, the temperature saturates at 416 K. As can be seen, cooling using water or R22 is not acceptable, since the saturation temperatures are higher than the maximum acceptable operating temperature of 60°C, or 333K. This, however, can be solved by doing the following: (a) admitting the liquid coolant before powering the circuit and (b) using high thermal conductivity materials, such as diamond, to build the cooling block.
2.6 Diamond Cooling Block

The diamond cooling block is simulated next for inlet velocity of 10 mm/s and 100 mm/s. The results show that the maximum (overshoot or saturation) temperature has not exceeded 333K, or 60°C, for each of the three coolants at any of the velocities. The enhanced cooling efficiency is attributed to the higher thermal conductivity of diamond than SiO₂. This means all three coolants can be used to provide the required operating temperature.

It appears that the reason for the overshoot, shown in Figure 2.4, is due to the powering of the chip at the same time the coolant is admitted. In this case, temperature increases fast because the initial temperature of the block and the ambient are assumed to be 293K. To determine if this assumption is correct or not, coolant is admitted first until the temperature of the block and the chip reached the coolant temperature, which is 293K for water, 232.2 K for R22, and 77K for LN. The ambient temperature outside the block has remained at room temperature.

The results shown in Figure 2.6 confirm that the overshoot has been eliminated for inlet velocity of 10 mm/s. The behavior remains the same when the velocity is increased to 100 mm/s. In this case, for 10 mm/s velocity, a saturation temperature of 123 K, 272K, and 325 K for LN, R22, and H2O, respectively, are reached. But, as shown in Figure 2.7, these temperatures decrease
to 106 K, 260 K, and 318 K from 77, 232.2, and 293 K, respectively, when the velocity is increased to 100 mm/s. Therefore, it is believed that high-conductivity materials, such as diamond, can provide cooling systems that ensure acceptable operating temperatures for 3D ICs.

![Graph](image)

**Figure 2. 6:** Hot spot temperature for diamond block when coolant is admitted before powering the chip. Inlet velocity is 10 mm/s for diamond cooling block.

![Graph](image)

**Figure 2. 7:** Hot spot temperature for diamond block when coolant is admitted before powering the chip. Inlet velocity is 100 mm/s for diamond cooling block.

### 2.7 Assessment of Cooling Block Efficiency

To assess the effectiveness of each technique for the next generation integration, the heating power is increased from 20W to 100W and the resulting temperature are simulated. Figure 2.8 shows the chip temperature of a SiO$_2$ cooling block, when the power is increased from 20 W to 100W, using water inlet velocity of 10 mm/s. We can see that, when the test heater is supplied
with a power beyond 20W (1000 W/cm² hot spot), the temperature increases to unacceptable values.

Figure 2.8: SiO₂ cooling block at 10 mm/s inlet velocity observed for test chip at different wattages.

Figure 2.9 shows the chip temperature when the inlet speed increased to 100 mm/s. The figure shows that water has not been able to reduce the temperature to an acceptable level, even at 20 W, as the case for R22 coolant. Both of the coolants have failed the test for power beyond 20W. For liquid nitrogen, an acceptable operating temperature can be obtained for power up to 40 W or a power density of 2000W/cm². This shows that SiO₂ is a poor cooling material due to its low thermal conductivity of 1.3 W/mK [9].

Figure 2. 9 : SiO₂ cooling block at 100 mm/s inlet velocity observed for test chip at different wattages.
Figure 2.10 shows the chip temperature if a diamond cooling block is used when the power is changed from 20-100 W and the inlet velocity is 10 mm/s. The figure shows that water can be used for power up to 40W or power density of 2000/cm², R22 for power up to 60 W or power density of 3000 W/cm², while LN can cool the chip even if 100W or power density of 5000 W/cm². This is possible due to the higher thermal conductivity of diamond of 1000 W/m. K [10].

![Graph showing the temperature of different coolants at various wattages](image)

Figure 2.10: Diamond cooling block at 10 mm/s inlet velocity observed for test chip at different wattages.

The increase in coolant velocity to 100mm/s substantially enhances the cooling efficiency of diamond cooling blocks as shown in Figure 2.11. The figure shows that LN and R22 cooling provides acceptable operating temperature even when power is 100W. For water cooling, temperature can be within an acceptable range for power less than 45 W.
Although SiO$_2$ is a poor thermal conductor, it is cheaper than diamond. Therefore, for lower power density applications, SiO$_2$ cooling blocks can be used. For higher power density, diamond cooling blocks can be used instead. If diamond cannot be used, then silicon carbide (SiC) can be utilized instead.

### 2.8 CONCLUSION

We designed cooling blocks to study the dynamics of liquid cooling of 3D-ICs. The results show that an overshoot takes place if chip powering and coolant admission are done at the same time. The results also show that the use of high thermal conductivity material enhances the cooling process. To further understand the dynamics of the cooling process, the internal pressure and velocity of the microchannels should be investigated. To evaluate the potential of product failure due to the cooling cycles, the stress and strain of the chip may be studied. To reduce cost, one should experiment with cooling blocks made of SiC, aluminum, or steel.
References

https://www.semiconductors.org/resources/2015-international-technology-roadmap-for-semiconductors-itr/


Chapter 3

THERMAL STRESS ANALYSIS OF LIQUID-COOLED 3D-ICS

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 the thermal stress resulting from a 3D-IC hot spot with 20W power dissipation. The IC is cooled using SiO$_2$ 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 exceeds the maximum tensile strength of aluminum, 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.

3.1 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 to integration is proposed, 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-
IC 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 allow for integration of heterogeneous technology.

Although 3D-IC technology has 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 ICs 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 microchannels [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 have investigated the thermal stress resulting from the dynamic temperature change of hot spots in liquid-cooled 3D ICs. The study is performed using the advanced multi-physics numerical analysis program, Comsol\textsuperscript{TM}.

3.2 Thermal Stress

The relationship between stress and strain is expressed by

\[ \sigma = E \varepsilon \quad \ldots \quad (1) \]

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:

\[ \sigma = E \alpha \Delta T = E \alpha (T_F - T_0) \quad \ldots \quad (2) \]

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 ICs do not have the same thermal expansion, significant thermal stress may arise, adversely impacting the device performance and the circuit reliability [12].

3.3 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 Figure 3.1. In this chip, the substrate is a 200\( \mu \)m thick silicon with an area of 5x5 mm\(^2\); on top of the substrate, 0.30\( \mu \)m of SiO\(_2\) is deposited; on top of the oxide layer, a 0.18\( \mu \)m-thick serpentine-shaped tungsten layer with a total area of 1000
x 2000 μm² is deposited, and 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.

Figure 3.2 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 is 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).

Figure 3. 1:Serpentine-shaped tungsten heating block on top of silicon substrate.

Figure 3. 2: Test chip sandwiched between two cooling blocks.
3.4 Simulation and Analysis

Comsol multiphysics program is used to conduct the thermal analysis. The emissivity values utilized in the simulation are 0.79 for SiO$_2$ and 0.63 for diamond [13,14]. The convective heat transfer coefficient is assumed to be 50 W/m$^2$K for all the coolants and 10 W/m$^2$K 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, shown in Figure 3.3(a). This temperature produces a thermal stress of 4552 MPa, as shown in Figure 3.3(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.

![Figure 3.3: (a) Temperature and (b) thermal stress of the IC without cooling.](image)
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 Figure 3.2. Temperature and the thermal stress are calculated for the first 15 s, after admitting the liquid coolants. A 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 are admitted at the same time the heater gets powered up, where the system temperature is initially set to room temperature, 293K.

For SiO₂ cooling block, liquid is pumped at a velocity of 10 mm/s when the 20W heater is powered on, and temperature and thermal stress are calculated during the first 15 seconds. From Figure 3.4, we can observe that the temperature reaches an overshoot within 1s for water, R22, and liquid nitrogen cooling. The maximum temperature at the tungsten metal (hot spot) for water, R22, and liquid nitrogen reaches 432K, 437K, and 375K and the corresponding coolants reach saturation at 432K, 416K, and 293K.

![Figure 3.4: Hot spot temperature using SiO₂ cooling blocks for water, liquid N₂, and R22 coolant with 10mm/s inlet velocity.](image-url)
Figure 3.5 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 N\textsubscript{2}, 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 $|T_f - T_0|$ value ($T_0=293$K) as explained in Eq. (2). At 1s, $|T_f - T_0|$ has the values of 139K, 144K, 82K for Water, R22, and Liquid N\textsubscript{2}, respectively. This explains the reason for the higher thermal stress for R22 compared to the other cooling cases at 1s. But as the temperature starts to saturate, the difference from the initial temperature for the coolants starts decreasing. This difference becomes smaller for R22 than water, leading to a lower stress. Cooling has reduced the thermal stress to less than the ultimate tensile stress of tungsten of 980 MPa.

![Thermal stress graph](image)

Figure 3.5: Thermal stress using SiO\textsubscript{2} cooling blocks on the IC for water, liquid N\textsubscript{2}, and R22 coolant with 10mm/s inlet velocity.

Aluminum 6061 is widely used in IC technology as interconnects and ohmic contacts. The ultimate tensile strength of that alloy is 310MPa [15]. For cooling using SiO\textsubscript{2} cooling block and velocity of 10mm/s, the thermal 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. Figure 3.6 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 (R22 and liquid nitrogen) takes place. For LN, the temperature decreases to 207 K within 1s and saturates. This is due to the fact LN temperature is 77 K, and the velocity is high enough to pump enough LN to bring the temperature to a much lower value than the initial value much quicker.

Figure 3.6: Hot spot temperature after cooling using SiO₂ cooling blocks with 100 mm/s inlet velocity. Coolant used are water, R22, and LN.

Figure 3.7 shows the thermal stress for an inlet velocity of 100 mm/s. The temperature differences |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, 161 MPa, and 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 case thermal degradation of the IC.
Fig 3.7 also 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).

Figure 3. 7: Stress on tungsten when SiO₂ cooling blocks with 100 mm/s inlet velocity. Coolant used are water, R22, and LN.

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, and 262 MPa respectively for water, R22, and LN. Similar to the case of tungsten hot spot, R22 cooling provided the lowest stress, because the temperature difference is the lowest.

These results have identified 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 impacts of metal elasticity and ultimate tensile stress are critical reliability factors that must be considered when designing 3D-ICs.
Since SiO$_2$ is a poor thermal conductor, it is reasonable to assume that SiO$_2$ is responsible for the high temperature of the chip. For this reason, diamond, a highly thermal conductive material, is used instead to build the cooling block. Figure 3.8 shows the hot spot temperature for coolants pumped at 10mm/s velocity. The saturation temperatures are found to be 325 K, 272 K, and 143 K for water, R22, and LN, respectively. As can be seen from the figure, the saturation temperatures are much lower than when SiO$_2$ cooling blocks were used. The reduced temperature for water can be attributed to the higher thermal conductivity of diamond, which has resulted in a faster dissipation of heat energy to the outside. For R22 and LN, the temperature is lower than the ambient temperature. This means, heat transfer should be from outside to inside, which in turn should raise the inside temperature for diamond blocks compared with SiO$_2$ block. But the opposite happens. The failure of heat to transfer from outside to inside is believed to be due to the creation of a cold domain outside the block. This domain acts as an isolating region that prevents heat from reaching the inside with sufficient rate to elevate the inside temperature.

![Figure 3.8](image)

Figure 3.8: Hot spot temperature for water, liquid nitrogen, and R22 coolant at 10 mm/s inlet velocity using diamond cooling block.

The stress analysis for diamond blocks with 10 mm/s velocity is shown in Figure 3.9. The thermal stress on tungsten is found to be 77 MPa for water, 58 MPa for R22, and 440 MPa for LN.
As can be seen, the stress for water and R22 is low, since the temperature differences, |TF - T0|, are 32K and 21 K, respectively. For LN, the difference is 150K. This explains why the stress in case of LN is much higher. In all three cases, the stress is lower than the maximum tensile strength of tungsten hot spot. Similarly, for aluminum, the stress is found to be 42 MPa for water, 86 MPa for R22, and 443 MPa for LN. Again, the thermal stress for LN is high, but in this case, it is higher than the maximum tensile strength of aluminum, which may cause metal fatigue.

![Graph showing stress over time for H2O, Liquid N2, and R22](image)

Figure 3.9: Stress on the IC for water, liquid nitrogen, and R22 coolant at 10 mm/s inlet velocity using diamond cooling block.

When the inlet velocity increased to 100 mm/s, the temperature of the hot spot changed to 315 K for water, 260 K for R22, and 106 K for LN, as shown in Figure 3.10. Figure 3.11 shows the thermal stress which is found to be 60 MPa for 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 hot spot.
Figure 3. 10: Hot spot temperature for water, liquid nitrogen, and R22 coolant at 100 mm/s inlet velocity using diamond cooling block.

![Temperature Graph](image1)

**Figure 3. 11: Stress on the IC for water, LN, and R22 coolant at 100 mm/s inlet velocity using diamond cooling block.**

![Stress Graph](image2)

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 problems for the IC reliability, if cooled with LN.
3.5 Conclusion

The stress analysis for next-generation 3D-ICs has 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


https://www.semiconductors.org/main/2015_international_technology_roadmap_for_semiconductors_itrs/


[14] “All About 6061 Aluminum, (Properties, Strength and Uses).”

https://www.thomasnet.com/articles/metals-metal-products/6061-aluminum/
Chapter 4

NUMERICAL ANALYSIS OF LIQUID COOLING OF 3D-ICS USING EMBEDDED CHANNELS

Hot spots are considered among the unavoidable consequences of the high integration density of 3D-ICS. Eliminating hot spots requires employing efficient cooling techniques. Using embedded channels, liquid cooling systems can be designed to deliver the right amount of coolant to each spot of the chip. In this study, numerical analysis is used to investigate the cooling of a 20 W hot spot using embedded channels employing three coolants: water, Freon (R22), and liquid nitrogen (LN). The investigation of thermal management and stress shows that, although LN provides the lowest operating temperature (164 K), it causes the highest stress (355 MPa) at 100 mm/s inlet velocity. The study also shows that coolant delivery using parallel channels results in a wide variation of local temperatures and stress. This stress variation may form “high-stress spots,” which may cause circuit failure, performance degradation, or yield reduction. Therefore, cooling systems and chip fabrication should be designed to ensure the elimination of high-stress hot spots.

4.1 Introduction

Increasing the packing density requires the reduction of the device size. Published reports show that device size has reached its physical limit and Moore’s law is almost reaching its end. Although several technologies, such as quantum computing or single electron transistor, have been proposed to meet the increasing demand for high integration density, the 3D-IC technology is believed to be the most practical one. 3D-ICs are built by stacking several 2D-ICs on top of each
other and connecting them using metal interconnects through vias [1]. Such arrangement increases the integration density by as many folds as the number of the stacked ICs. 3D-IC technology can also provide reduced latency, multitasking, high-speed processing, and heterogeneous integration [2]. The main problem with 3D-ICs is the creation of the hot-spots, with local temperatures reaching very high temperatures. These hot spots may result in dysfunctional operation, component failure [3], thermal stress, signal delays or permanent degradation of the chip. Although several cooling measures, such as microelectromechanical Systems (MEMS) based technology, embedded microchannels [5,6], liquid immersion cooling [7], and microfluidic cooling using thermal through-silicon vias (TSVs) [8] have been proposed, they may not be able to eliminate these thermal challenges completely.

In this study, thermal stress analysis is conducted for a chip with a 20W hot spot that is liquid cooled through parallel embedded channels. The study was conducted using Comsol, a multiphysics numerical analysis program. Water, freon (R22), and liquid nitrogen (LN) are used as coolants.

4.2 Heat Transfer and Thermal Stress

The heat transfer in any system can be a result of a combination of one or more of the following mechanisms: conduction, convection, and radiation. Heat conduction through a material with length \( l \) and cross-section \( A \) can be expressed by the following equation:

\[
\frac{Q}{t} = kA \frac{dT}{l}
\]  

(1)

Here, \( Q \) is the transferred heat, \( t \) is the time, \( k \) is the thermal conductivity (W.m\(^{-1}\).K\(^{-1}\)) and \( \Delta T \) is the temperature gradient, or the temperature difference between the two ends of the object in Kelvin (K). \( Q/t \) is the power in watts.
Convection is the heat flow carried by the motion of fluid or gas. Heat convection can be expressed by the following equation:

\[ Q = hA\Delta T \]  

(2)

In this equation, \( h \) is the convective heat transfer coefficient (J.m\(^{-2}\).K\(^{-1}\)), \( A \) is the object’s cross-sectional area that carries the fluid, and \( \Delta T \) is the temperature difference between fluid temperature, \( T_F \), and heated surface temperature, \( T \).

Radiation is a method of heat transfer where the exchange of heat takes place through heat emission to the ambience. Radiation can be described by the following equation:

\[ P = \varepsilon\sigma AT^4 \]  

(3)

Here, \( P \) is the power, \( \varepsilon \) is the emissivity factor, \( \sigma \) denotes Stefan-Boltzmann constant (=5.67 X 10\(^{-8}\) W.m\(^{-2}\).K\(^{-4}\)), \( A \) is the surface area of the radiating body and \( T \) is the temperature of the radiating body [9].

The relationship between stress and strain is bound by the following equation,

\[ \sigma = \varepsilon E \]  

(4)

In this equation, \( \sigma \) is the uniaxial stress, \( \varepsilon \) is the strain or deformation, and \( E \) is Young’s modulus, which measures the stiffness of a solid material [10,11].

Heat may cause thermal stress. Thermal stress can be expressed by the following equation,

\[ \sigma = E\alpha\Delta T = E\alpha(T_F - T_o) \]  

(5)
where $\alpha$ is the thermal expansion coefficient, $T_0$ is the initial and $T_F$ is the final temperature of the object, and $\Delta T = T_F - T_0$ in Kelvins (K). The object temperature, $T_F$, varies with respect to time due to heat transfer, resulting in temperature difference [12].

4.3 Design of the Cooling System

For this study, a 20 W hot spot is represented by a 20 W tungsten heater [13]. Figure 4.1 shows the heater and the Cu wire used in this study. The heater is built on a 5x5 mm$^2$ silicon substrate with 200 μm thickness. On top of the wafer, a 3000 Å layer of SiO$_2$ is grown, on top of which a 0.18μm-thick serpentine-shaped tungsten heater and Cu wires are deposited, separated by a distance of 450μm. The area of the rectangle, representing the outer boundaries of the heater and the Cu wire, is about 1000x2000 μm$^2$. On top of the metals, 0.18μm of SiO$_2$ is deposited. Two 0.18μm aluminum pads with an area of 125x250 μm$^2$ are deposited at the ends of the of the heater and connected through the SiO$_2$ layer to the heater vias. Similar pads of Cu are deposited at the end of the Cu wire and connected through vias too. Finally, a 7000 Å SiO$_2$ passivation layer was deposited on the top.

Figure 4. 1: Serpentine shaped tungsten heating block (yellow) and copper Wire (Red).
The cooling block is made using two bonded Si substrates with a total width of 600μm and an area of 5x5 mm². One 2D-IC is sandwiched between two blocks, as shown in Figure 4.2 (a). To build a 3D-IC, the stacks and the blocks alternate one on top of the other, terminated with a cooling block on top and one at the bottom. The cooling block has embedded channels with the dimensions shown in Figure 4.2(b).

Figure 4.2: (a) Heater sandwiched between two cooling blocks; (b) cooling block geometry; inlets are 100μm above the bottom of the silicon substrate. Inclined lines are 350μm.

4.4 Simulation and Analysis

The thermal analysis is done using the physical material parameters provided by the Comsol material library unless it is otherwise stated. The emissivity value used in the simulation for Si is 0.93 [14]. The convective heat transfer coefficient is set to 50 W/m²K for all the coolants and 10 W/m²K for air. If the heater was turned on without admitting any coolant, the chip temperature would reach 2170K and the thermal stress would reach 4552 MPa within a few
seconds, as reported in [15]. It should be noted that without cooling, the chip will evaporate, and nothing will be left to talk about. Therefore, we need an efficient cooling mechanism to ensure that both temperature and stress do not exceed the safety limit. Embedded liquid cooling channels may present an efficient way of cooling if the appropriate inlet velocity is used to optimize the cooling process.

4.5 Low Inlet Velocity

First, we set the inlet velocity into the embedded channels, shown in Figure 4.2, to a low value of 10 mm/s to cool the chip. The corresponding temperature and stress are simulated. Figure 4.3(a) shows the map of the resulting temperature for water cooling. The figure shows that the temperature of the chip is ranging from about 293 K to 420 K or 20 °C to 147 °C. When R22 is used instead, the temperature ranges from 240K to 380K or -33°C to 107°C; see Figure 4.4(a). Although R22 reduces the temperature compared with cooling with water, the hot spot temperatures are still unacceptable for both cooling techniques. To reduce the temperature to an acceptable value, we have two choices: either to use a different coolant or to increase the inlet velocity. We started with using a different coolant, LN. For LN, temperature is reduced ranging from 77 K to 220 K or -196 °C to -73 °C; see Figure 4.5(a).

As can be seen from Figure 4.3(a), 4.4(a), and 4.5(a), the chip temperature at the inlet is the lowest and is equal to the temperature of the coolant liquid itself. As the liquid flows inside the block, it cools the IC by absorbing the heat from the chip. At the same time, the liquid temperature increases and its efficiency in cooling decreases. Hence the chip cooling rate slows down, and the temperature grows above the inlet temperature. The chip temperature increases as we go along the channels, and it may reach unacceptably high values, forming hot spots. These high temperatures
are unacceptable because of their value and non-uniformity. The high temperature values may cause failure; and the non-uniformity may cause performance degradation of the chip.

Figure 4. 3: Water cooling with 10mm/s pumping velocity (a) temperature and (b) thermal Stress map across Cu and tungsten.
Figure 4.4: R22 cooling with 10mm/s pumping velocity (a) temperature and (b) thermal Stress map across Cu and tungsten.

Figure 4.5: LN cooling with 10mm/s pumping velocity (a) temperature and (b) thermal Stress map across Cu and tungsten.
To probe the temperature variation more closely, the temperature of the copper wire is observed over 15 seconds. Figure 4.6 shows that the temperature rises as an exponential function, then saturates within 2s. The temperature saturates to 415K, 367K and 213K for water, R22, and LN, respectively. These values are close to the maximum temperature observed in Figures 4.3(a), 4.4(a), and 4.5(a).

One observation from the figure is that there is no overshooting, as the case of [13]. The disappearance of temperature overshoot is believed to be due to the higher rate of cooling. Higher rate of cooling can be achieved either by increasing heat convection (by increasing liquid velocity), increasing heat conduction, or increasing radiation. For similar liquid velocities and outside atmospheres, it appears that the high thermal conductivity of Si (130 Wm\(^{-1}\)K\(^{-1}\)) is responsible for the disappearance of overshooting. To the contrary, in [13] conduction takes place through a low-thermal conductive SiO\(_2\) layer of 1.5 Wm\(^{-1}\)K\(^{-1}\). Therefore, using materials with higher thermal conductivities will result in increasing heat dissipation.

Figure 4. 6: Cu wire temperature at 10mm/s inlet velocity.

Temperature non-uniformity should raise serious concern since it is expected to result in the non-uniformity of the material properties across the chip, which would lead to the non-
uniformity of the characteristics of identical devices across the chip, and this leads to an added level of degradation. The problem is not only related to the electrical or optical performance of the devices. The problem extends to the physical structure of the chip. Among these problems, the creation of high thermal stress at the hot spots is one of the biggest concerns, which may lead to fatigue, fracture, or bulging of the films.

Stress on thin film metals is investigated. Figures 4.3(b), 4.4(b), and 4.5(b) show maps for the von Mises stress on the hot spot (tungsten heater) and the copper wire when the chip is cooled using H₂O, R₂₂, and LN. For all coolants, it is noticed that the stress on the Cu wire is higher than on tungsten. This is due to two factors. The first is the thermal expansion for copper is higher than tungsten, 16.5x10⁻⁶/K compared with 4.5x10⁻⁶/K. The second, is the fact that, in general, tungsten temperature is lower because it is closer to the inlet.

For water cooling, the highest temperature difference of the Cu is 122 °C (415 K-293 K). For R₂₂ cooling, that difference is 74 °C, and for LN the difference is -80 °C. Therefore, the von Mises stress on Cu is 337 MPa for water, 200MPa for R₂₂, and 211MPa for LN; see Figure 4.7. For water cooling, the resulting maximum stress is beyond both the tensile yield strength (33.3 MPa) and the ultimate tensile strength (of 210 MPa) for pure copper [16]. For electroplated copper, with Poisson’s ratio of 0.35 and thickness less than 0.885μm, the yield strength reaches 300MPa [17]. This higher yield strength is not high enough to prevent failure or fracture of the copper [18]. For cooling with R₂₂ or LN, pure copper will deform or fracture under the resulting stress, but electroplated Cu can withstand the stress without a problem.
The maps for the local stress are shown in Figures 4.3(b), 4.4(b) and 4.5(b), for water, R22, and LN cooling, respectively. From Figure 4.3(b) and 4.4(b), it can be realized that the stress for the tungsten increases as we move far from the inlet, while it decreases on the Cu wire. This is because, for tungsten, the side closer to the inlet is exposed to colder liquid and has lower temperature whereas, for Cu, the side closer to the inlet is closer to the heater and has higher temperature. This remains valid, as long as the stress is tensile, and the temperature is above the room temperature. For LN cooling, Figure 4.5(b), the stress behavior is the opposite, even though the temperature behavior is the same as the other two cases with R22 and water. This can be explained by the fact that LN results in a compressive stress that increases as the temperature decreases. Hence, it is expected that the colder side will have higher stress and the warmer side will have lower stress, as long as the temperature is lower than the room temperature.

For tungsten heater, the stress is tensile when coolant is H₂O or R22. For water cooling, the highest stress is found to range from 200 MPa at the sides to about 320 MPa at the middle. For cooling using R22, the stress ranges from about 100 MPa to 220 MPa. For LN, the stress is compressive, ranging from about 80 MPa to 156 MPa. In all cases, the stress on tungsten is much

![Figure 4. 7: Cu wire stress at 10mm/s inlet velocity.](image-url)
less than the yield strength of 750 MPa and the ultimate tensile strength of 980 MPa [19]. In fact, tungsten yield can reach 1670 MPa and an ultimate strength to 3900 MPa [20].

The stress observed on the aluminum pads are tensile for water and R22, reaching 318 MPa and 206 MPa, respectively. For LN, the stress is compressive, reaching 148 MPa. Aluminum 6061 has a tensile yield strength of 267 MPa and maximum tensile strength of 310 MPa [21]. The results indicate that, except for water cooling, Al 6061 will not deform under cooling with R22 and LN. Stress on Al is determined by its thermal expansion (23x10^{-6} K^{-1}) and temperature difference. Since Al pads are located at the sides of the heater, they have lower temperatures.

4.6 Inlet Velocity 100 mm/s

The inlet velocity was increased to 100 mm/s. The temperature and von Mises stress are mapped in Figure 4.8, 4.9 and 4.10 for water, R22 and liquid N₂ cooling, respectively.

![Figure 4.8: Water cooling with 100mm/s pumping velocity (a) surface temperature map and (b) thermal stress map across Cu and tungsten.](image)
Figure 4.11 shows that, after 1 s, the temperature values are reduced to 345K, 321K and 164K for water, R22 and LN, respectively. This results in a temperature difference from room temperature of 52K, 28K, and -129K, for water, R22, and LN, respectively. Proportional to this temperature difference, Figure 4.12 shows that the stress on the copper wire reaches saturation at 144MPa, 78MPa and 355MPa respectively for water, R22 and LN. Bulk Cu will yield and deform at these stresses, because the yield strength is 33.3 MPa. If electroplated Cu is used, metal will not deform for water cooling, although LN cooling would result in deformation of Cu.

Figure 4.9: R22 cooling with 100mm/s pumping velocity (a) surface temperature map and (b) thermal stress map across Cu and tungsten.
Our simulation shows that, at the tungsten hot spot, the temperature saturates at 364 K for water, 336 K for R22, and 179 K for LN. The stress observed at the hot spot are 150 MPa, 94 MPa and 222 MPa for water, R22 and liquid nitrogen, respectively. For tungsten hot spot, the stress values are far below the tensile and compressive yield stress. The stress on the aluminum pads is 158 MPa, 104 MPa, and 242 MPa for water, R22, and LN. Since the tensile yield strength of Al 6061 is 267 MPa, Al pads will be resistant to deformation at these stresses [21].

Figure 4. 10: LN cooling with 100mm/s pumping velocity (a) surface temperature map and (b) thermal stress map across Cu and tungsten.

It is quite normal for a processor to have an idle temperature around 318K to 323K (25 °C to 30 °C) and full load temperature (50 °C to 60 °C) [22]. Therefore, all cooling techniques at 100
mm/s velocity can satisfy the temperature requirement for this case. However, electroplated Cu must be used to ensure higher yield strength, or other metal structures with higher yield need to be used.

Figure 4.11: Cu wire temperature at 100mm/s inlet velocity.

Figure 4.12: Cu wire stress at 100mm/s inlet velocity.

4.7 Assessment of Cooling Using Embedded Microchannels

There are several reasons for utilizing the silicon-chip-embedded cooling system. Silicon is a commercially available material, and the cooling block is quite easy to fabricate. Using embedded channels for cooling has the advantages of the ability to reach the location of the hot spots inside the 3D-ICs. Hence the channel design can be optimized to obtain the best results. In addition, silicon has a good thermal conductivity of 148 W/mK [5]. Furthermore, the high tensile
strength of Silicon (7000 MPa) can withstand the thermal stress during cooling [6]. Figure 4.13 shows that at 10mm/s velocity, cooling using water is not acceptable at any power, and R22 cooling cannot achieve acceptable temperature beyond the power of 20 W. Liquid nitrogen cooling, on the other hand, can provide acceptable operating temperature up to 42W (2100 W/cm²).

![Figure 4.13: Chip-embedded cooling block at 10 mm/s inlet velocity for test chip at different wattages.](image)

Figure 4.14 shows the embedded cooling block’s performance at 100mm/s velocity. The figure shows that water can provide acceptable operating temperature at 20 W. R22 can be an efficient coolant up to 30 W, while liquid nitrogen can handle dissipated heat up to a power of 60W.

![Figure 4.14: Chip-embedded cooling block at 10 mm/s inlet velocity for test chip at different wattages.](image)
4.8 Conclusion

The study investigates the temperature and stress on the different metal wires and pads when a 20 W hot spot is cooled using embedded channels. The study considered three different coolants: water, R22, and LN with admission velocities of 10 mm/s and 100 mm/s. The study revealed important facts. First, water may not be able to bring temperature to an acceptable level to avoid Cu and Al deformation, especially at lower fluid velocity of 10 mm/s. Second, bulk copper cannot withstand the stress using any coolant due to the low yield strength of 33.3 MPa. Third, if electroplated Cu is used, the yield strength increases to 300 MPa, and the metal can withstand the stress resulting from cooling using R22 and LN at 10mm/s velocity. Fourth, if the velocity increases to 100 mm/s, metal can withstand stress when water and R22 are used, but LN cooling causes unacceptable stress. In other words, R22 is the most suitable coolant for cooling 20 W hot spot. Since Al 6061 has a yield of 267 MPa, what is said about electroplated Cu can be said about Al alloy. For tungsten, with a yield strength of a minimum 750 MPa, no deformation will take place under any of the cooling mechanisms. The study also shows that admitting the fluid from one side results in a large temperature and stress gradient that can reach 140 °C and 100 MPa, respectively. This creates local hot-spots and high-stress spots. These spots may not damage the IC but may make the performance unacceptable. The study shows the need for redesigning the embedded channel system in order to optimize the cooling process and eliminate both hot spots and high-stress spots.
References


http://coolcosmos.ipac.caltech.edu/cosmic_classroom/light_lessons/thermal/transfer.html


[16] “Copper Cu Annealed,”

http://matweb.com/search/datasheet_print.aspx?matguid=9aebe83845c04c1db5126fada6f76f7e


[22] “What should be the temperature of my computer processor?”, https://www.computerhope.com/issues/ch000687.htm
Chapter 5

VERIFICATION

To verify that our simulation is correct and there is no unknown mistake, simulation results are compared with analytical results. A power of 20W is applied to the chip without admitting cooling liquids. With emissivity factor of 0.66 for Si, the temperature was calculated using Comsol. Figure 5.1 shows the simulation results, where the heater temperature reached 2160K.

![Simulation Results](image)

Figure 5. 1: Simulated temperature for 20W power.

This is a case of a heated chip cooled by radiation to the atmosphere. The temperature, in this case is governed by Equation 1:

\[ T = \sqrt[4]{\frac{P}{\varepsilon \sigma \text{Area}_{\text{substrate}}}} \quad \text{..... (1)} \]

Here, \( P \) is the applied power=20W, \( \varepsilon \) is the emissivity factor=.66, \( \sigma \) is the Stefan-Boltzman constant= \( 5.6704 \times 10^{-8} \) W/m\(^2\)-K, and \( \text{Area}_{\text{substrate}} \) defines the area of radiation=25mm\(^2\). The
temperature according to Equation 1 is 2150K. This shows that the simulated result clearly falls within the 1% from the calculated value.

The way Comsol simulation is used is compared to other published simulations; Michael et al. [1] is used as reference for the comparison. The comparison between the reference work and the simulated work is presented in Figure 5.2. The temperatures obtained for both cases are almost the same.

![Graph showing temperature comparison](attachment:image.png)

**Figure 5.2:** (a) Calculated temperature of the IC from reference paper and (b) simulated temperature for the IC.
In order to examine the integrity of the simulation further, an experiment related to the thermal analysis of DC/DC module has been simulated as well. The simulated surface temperature shows almost the exact same surface temperature obtained from the experimental approach described by Wang et al [2]. Therefore, it is safe to assume that the understanding of the simulation is correct. Figure 5.3 shows the comparison.

Figure 5.3: (a) Infrared image of the DC/DC module and (b) simulated temperature of the DC/DC module.
5.1 Conclusion

3D-IC stands out to be the most promising solution for the next-generation applications. Over the thesis, several dynamics of the 3D-IC cooling have been studied. The study shows that achieving mere cooling of the 3D-ICs cannot be the only solution for higher integration. Thermal stress needs to be considered as well. Thermal stress exists for high as well as low final temperature. Temperature difference plays an important role in generating this stress. Rather than focusing on the operating temperature of the IC, the goal of the cooling should be keeping the temperature difference minimum because deep cooling might cause more stress on the IC. Utilizing efficient heat sink module and fans, the dependence on the coolants can be reduced. The impact of metal elasticity and ultimate tensile stress are critical reliability factors for 3D-IC design. Stronger metal alloy needs to be implemented for the safety of the contact pads and interconnects. Out of these cooling systems, embedded cooling turns out to be more efficient, cost effective, and reliable for 1000W/cm² hot spot.

5.2 Future Works

To make this thesis more meaningful, fabrication of the designed cooling blocks is required to verify the results of the obtained simulations. Also, it will be helpful for further optimization of the cooling blocks. A proper cooling mechanism for the 3D-ICs requires more research on dielectrics, and IC encapsulation. Hybrid cooling system can be incorporated for utilizing more efficient cooling. The cooling block needs to be tested for junction-ambient thermal resistance. Utilizing aggressive pumping technology can contribute to faster convection of the IC temperature. Therefore, more research needs to be carried out for EOF pumping technology. For enabling liquid immersion cooling, 3M Novec7000 engineered fluid, deionized water seems to be some attractive options.
References
