Abstract

Title of Thesis: Thermal Modeling and Analysis of Three-Dimensional (3D) Chip Stacks

Christopher Bachmann, Master of Science, 2007-09-11

Thesis Directed by: Avram Bar-Cohen and Jungho Kim
Mechanical Engineering Department

Three-dimensional (3D) chip architectures have garnered much research interest because of their potential to alleviate the interconnect delay bottleneck that is expected to limit the traditional progression of Moore's law through device scaling in planar chips. While the benefits of 3D chip integration are clear, there are several obstacles to its broader implementation. In particular, the issue of power dissipation is a major challenge to the development of high performance 3D chip stacks. The well-documented difficulties in cooling future 2D chips will only be exacerbated by 3D architectures in which volumetric power density is increased and non-uniform power dissipation is more severe. This thesis focuses on three relevant topics in the cooling of 3D chip stacks: 1) the determination of effective thermal properties for use in compact thermal models, 2) single phase internal liquid cooling, and 3) hot spot remediation with anisotropic thermal interface materials.

THERMAL MODELING AND ANALYSIS OF THREE-DIMENSIONAL (3D) CHIP STACKS

by:

Christopher Bachmann

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Master of Science
2007

Advisory Committe:

Professor Avram Bar-Cohen, Chair and Advisor
Professor Jungho Kim, Co-Advisor
Professor Bao Yang

Acknowledgements

The completion of the work represented in this thesis is largely attributable to the guidance, inspiration, and help of others. I would like to take this opportunity to acknowledge the many contributors now.

I am in debt to my advisor, Dr. Avram Bar-Cohen, and co-advisor, Dr. Jungho Kim, who identified the need for this research and guided me along the way. Their wisdom and insight is reflected in many places throughout the text. I am continually humbled by their seemingly effortless recollection and understanding of all things “heat transfer.” Also, Dr. Bao Yang provided valuable insight in reviewing the thesis as a committee member.

My labmates – Dae Whan Kim, Peng Wang, Kyoung Joon Kim, Emil Rahim, Ilai Sher, Juan Cevallos, Slava, Jeff Kahn, and Josh Gess – have been a good source of conversation about classwork, heat transfer, and everyday life. It is nice to know that I’m not the only one who is at work more than he should be.

Finally, my mother, Joanne, best friend, Jason, and love of my life, Kristin, have provided mental and emotional support through the many ups and downs of life in graduate school. I often sacrificed time with them in order to complete my work and I greatly appreciate their patience. In particular, Kristin’s encouragement, dedication, and sacrifices were instrumental in both my completion of the thesis and the retention of my sanity.

Table of Contents

CHAPTER 1 : INTRODUCTION 1

MOTIVATION FOR 3D CHIP STACKS 1

THERMAL CHALLENGES IN 3D CHIP STACKS 7

THESIS ORGANIZATION 12

CHAPTER 2 : EQUIVALENT THERMAL CONDUCTIVITY DETERMINATION 15

INTRODUCTION..15

A REPRESENTATIVE LOCAL MODEL..17

APPROXIMATE ANALYTICAL DETERMINATION OF EQUIVALENT THERMAL CONDUCTIVITY 20

NUMERICAL DETERMINATION OF EQUIVALENT THERMAL CONDUCTIVITY 27

CONCLUSION..32

CHAPTER 3 : INTERNAL LIQUID COOLING WITH SINGLE PHASE MICROCHANNELS....34

INTRODUCTION..34

NOVEL 3D CHIP STACK MODEL..36

A REVIEW OF ANALYTICAL BASICS 40

Infinitely Wide Parallel Plates..41

Developing Laminar Flow..43

A TWO-DIMENSIONAL APPROXIMATION OF FLOW IN THE BCS 46

General CFX Validation..50

2D Thermal and Hydrodynamic Profile Validation....................................52

A Note About Turbulence..55

RESULTS OF 2D MODEL SIMULATIONS..57

Percentage of TCS Heat Applied..57

Inlet Velocity..59

Thermal Conductivity of BCS Slices..60

Summary *63*

A THREE-DIMENSIONAL MODEL OF FLOW IN THE BCS 64

RESULTS OF 3D MODEL SIMULATIONS 66

Inlet Velocity *67*

Thermal Conductivity of BCS Slices *70*

Summary *72*

CONCLUSION 73

CHAPTER 4 : ANISOTROPIC TIMS/SPREADERS 76

INTRODUCTION 76

ANALYTICAL SOLUTION 82

VARIATION OF IN-PLANE SPREADER THERMAL CONDUCTIVITY 91

VARIATION OF SPREADER THICKNESS 97

VARIATION OF FLUX-SPOT SIZE AND HEAT TRANSFER COEFFICIENT 108

NUMERICAL SIMULATIONS AND CONTACT RESISTANCE VARIATION 113

CONCLUSION 125

CHAPTER 5 : CONCLUSIONS 129

APPENDIX A: DETAILED LOCAL MODEL DRAWINGS FOR CHAPTER 2 134

APPENDIX B: MATLAB CODE FOR DETERMINING HOTSPOT TEMPERATURE 137

APPENDIX C: MATLAB CODE FOR VARYING SPREADER THICKNESS 140

REFERENCES 143

List of Figures

Figure 1-1: Planar IC technology – Intel Pentium 4 processor [3] 2

Figure 1-2: (a) Schematic of a 3D chip stack concept (adopted from [13]); (b) a two-dimensional representation of the layers and interconnects in a 3D chip stack [14]..5

Figure 1-3: Reduction of maximum wire length due to chip stacking (adapted from [12])..5

Figure 1-4: Schematic representation of 3D SOC fabrication (adopted from [19]) 7

Figure 1-5: Traditional air-cooled heat sink applied to a 3D stack (adapted from [20])...9

Figure 1-6: The heat transfer coefficients provided by different cooling technologies 12

Figure 2-1: (a) Five-layer 3D chip stack; (b) Top view of a single layer with dielectric material removed to expose electronics and interconnects; (c) a single 'unit cell' of the layer; (d) an individual heat source within one unit cell 16

Figure 2-2: (a) Interconnections between three cells of a 3D memory device [31]; (b) Vertical interconnects between a single active device on the lowest layer and various active devices in higher layers [32] 18

Figure 2-3: (a) an isometric view of the 3D local model, (b) with the BCB removed, (c) with BCB and InP substrate removed, (d) a top view of the model with BCB removed and details of a single heat source (*i.e.* active device)...20

Figure 2-4: Classic one-dimensional conduction of heat in a plane wall 21

Figure 2-5: The local model with outer dimensions and coordinate system labeled.........22

Figure 2-6: Top and front views of the local model with BCB removed to show the significance of the embedded interconnect gold.................................23

Figure 2-7: Resistance network for heat conduction in the z-direction of the local model..24

Figure 2-8: Resistance network for heat conduction in the xy-plane of the local model..26

Figure 2-9: Method for determining the equivalent thermal conductivity in ANSYS 29

Figure 2-10: Verification model – 5 x 5 array of gold cylinders embedded in an InP slab....30

Figure 3-1: Notional implementation of microchannels in a 3D chip stack [20] 36

Figure 3-2: Schematic of hybrid chip stack examined in this chapter....37

Figure 3-3: The hybrid chip stack to be analyzed in this chapter....39

Figure 3-4: Developing laminar flow temperature and velocity profiles for isoflux channels....44

Figure 3-5: Centerline velocity development, laminar flow in a parallel plate channel [47] 46

Figure 3-6: 2D flow configuration assumed in the BCS....48

Figure 3-7: 2D model mesh with boundary conditions and details shown....49

Figure 3-8: CFX verification model 50

Figure 3-9: Symmetric Heating 51

Figure 3-10: Asymmetric Heating 51

Figure 3-11: (a) Velocity profile in the 2D model at 1.5 mm for $V_{inlet} > 1$ m/s, (b) Comparison of centerline velocity to the expected values reported in [47]....53

Figure 3-12: Temperature profiles at different axial positions in the channel....54

Figure 3-13: Boundary Layer Development over a Flat Plate....55

Figure 3-14: Effect of TCS Heat Load on Maximum Silicon Temperature....58

Figure 3-15: Effect of Inlet Velocity on Maximum Silicon Temperature....59

Figure 3-16: Heat flux variation along the length of the 2D channel at the interface of the silicon and cooling fluid (FC-72) 61

Figure 3-17: Wall conductivity varied....62

Figure 3-18: Hypothetical configuration to increase spreading....63

Figure 3-19: Schematic of the 3D model configuration 66

Figure 3-20: Temperature contours in the silicon with V_{in} = 20 m/s, $k_{silicon}$ = 124 W/m-K (the figure shows a side view of one slice at the silicon/fluid interface) 68

Figure 3-21: Temperature contours in slice with V_{in} = 20 m/s, k_{eff} = 600 W/m-K, (the figure shows a side view of one slice at the silicon/fluid interface) 71

Figure 3-22: Varying the effective conductivity of the BCS slices in the 3D model....72

Figure 4-1: The 2005 ITRS forecast for high performance microprocessors [49] 76

Figure 4-2: On-chip heat flux (left) and temperature map (right) (adapted from [51])....77

Figure 4-3: Potential implementation of a silicon TEC (adapted from [63]) 80

Figure 4-4: Potential implementation of an anisotropic TIM/Spreader in a 3D stack up 81

Figure 4-5: Schematic of compound chip/spreader system 83

Figure 4-6: Schematic of the single layer system from which the bi-layer solution is found....84

Figure 4-7: Excess temperature profiles taken through the middle of the active chip surface for various k_{xy} 93

Figure 4-8: Hotspot cooling compared to an isotropic spreader....95

Figure 4-9: Temperature rise on the back of silicon, copper, and APG spreaders 97

Figure 4-10: Typical thermal resistance trends for various spreader thicknesses 99

Figure 4-11: R_s variations for spreaders with 'low' and 'intermediate' thermal conductivity....100

Figure 4-12: Variation of R_T and average hotspot excess temperature for increasing spreader thickness (k_z = 5 W/m-K) 102

Figure 4-13: Optimum Spreader thickness varies with in-plane thermal conductivity....103

Figure 4-14: Variation of R_T for increasing spreader thickness for alternative spreaders 104

Figure 4-15: Effect of chip thickness of spreader thickness variation for an orthotropic spreader with k_z = 5 W/m-K and k_{xy} = 350 W/m-K....106

Figure 4-16: Normalized total resistance for different chip thicknesses and an orthotropic spreader with k_z = 5 W/m-K and k_{xy} = 350 W/m-K.............107

Figure 4-17: Change in optimum spreader thickness for various k_{xy} and t_1 108

Figure 4-18: Generic 3D plot of the temperature spike on the active side of the chip...109

Figure 4-19: Hotspot temperature rise for escalating flux-spot size...............................111

Figure 4-20: Sensitivity of the thermal resistances to varying heat transfer coefficient 113

Figure 4-21: Mesh used to evaluate the effects of contact resistance in ANSYS............115

Figure 4-22: Model used to investigate the thermal contact resistance in ANSYS........117

Figure 4-23: ANSYS temperature profile through the thickness of the model in Figure 4-22...118

Figure 4-24: Comparison of numerical and analytical temperature profiles on the active side of the chip for various for k_{xy} = 50 and k_{xy} = 1800...............119

Figure 4-25: Surface temperature plots show good matching for k_{xy} = 100 W/m-K.......119

Figure 4-26: ANSYS results approach the analytical predictions for progressive node count...121

Figure 4-27: Numerical results under-predict hotspot cooling for larger and larger k_{xy}/k_z...121

Figure 4-28: Heat flux plots in the silicon for perfect and poor thermal contact............123

Figure 4-29: Hotspot vs. contact resistance for various k_{xy} 124

Chapter 1 : Introduction

Motivation for 3D Chip Stacks

Since the introduction of the integrated circuit (IC) in 1958, progress in the semiconductor industry has been characterized by a doubling of chip transistor count roughly every two years. Gordon Moore, a co-founder of Intel Corporation, is credited as being the first to recognize this trend in 1965 [1] [2]. Moore famously inferred that the trend would persist for the near future, thus introducing a precept – "Moore's Law" – that has shaped and guided the IC industry for the past four decades. During this time period, the prevailing chip architecture has been planar in nature with functional blocks laid out in the familiar "tile" pattern seen in Figure 1-1. Traditionally, the Moore's Law progression in transistor count has been accomplished through continual downscaling of chip feature sizes within the confines of this 2D architecture. However, as industry capabilities push beyond the 45 nm technology node – a milestone that is expected to be ushered in at the end of this year with the commercial introduction of Intel's Penryn microprocessor [4] – it is unclear whether further scaling can be supported in the traditional planar chip architecture.

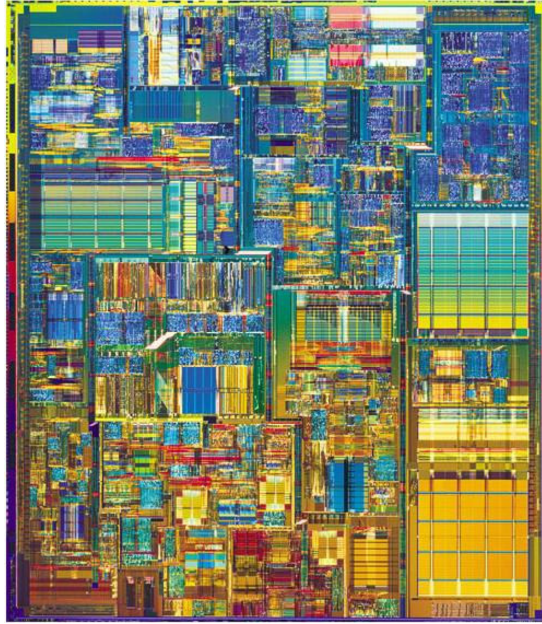


Figure 1-1: Planar IC technology – Intel Pentium 4 processor [3]

One of the largest hurdles facing future 2D scaling efforts is the issue of global interconnect delays [5]. Signals are transmitted to different portions of a chip along thin interconnects that are traditionally routed around the chip periphery. Unlike MOSFET transistors, whose gate switching speeds benefit from diminishing feature size, interconnect wires degrade in performance with progressive miniaturization [6]. Aggressive feature scaling results in more tightly packed wires of smaller cross-sectional area; however, since chip footprints have remained relatively constant for recent technology nodes, global interconnect wires are required to traverse the same intra-chip distance. The resulting increase in aspect ratio (length/diameter) leads to progressively larger electrical resistance and capacitance in the wire which, in turn, increase signal propagation delay ('RC delay'). At larger technology nodes, this interconnect delay was generally considered a non-issue since transistor gate delays were considerably longer than interconnect delays [6]. As scaling continues, however, it is expected that

interconnect delays will compete with gate switching delays to become the main bottleneck in on-chip data transmission [7].

The impending interconnect bottleneck is underscored in [6] where it is noted that RC wire delays could exceed gate switching delays by two orders of magnitude at the 35 nm technology node. One potential solution to this problem is to increase the cross-sectional area of global interconnects over the main portion of their length so as to reduce electrical resistance [11]. However, "fattening" the wires in this way significantly increases the amount of chip area needed for interconnect routing and thus increases the cost of the chip. Another potential solution is to intermittently place so-called 'repeaters' along the length of wire; these repeaters reduce the effective resistance of the path by regenerating the signal and sending it further down the line, either to the next repeater or to the final destination [11]. While repeaters offer lower effective resistance, their implementation comes at the expense of increased circuit complexity and power consumption. Alternatively, novel chip architectures could address the RC delay issue with global interconnect routing schemes that significantly reduce the maximum distance that signals must travel. Three-dimensional (3D) chip packages are an emerging technology that offer great potential in this regard as well as numerous other benefits (e.g. a smaller footprint, higher yield, and the possibility of heterogeneous integration) [12]. The remainder of this section will serve to introduce 3D chip architectures and identify the benefits of three-dimensional system integration.

In contrast to the planar architecture shown in Figure 1-1, three-dimensional system integration is achieved by vertically stacking the functional elements of a circuit (e.g., logic and cache) as shown in Figure 1-2. A vital attribute of this layered

architecture is the reduced distance between communicating portions of the circuit. The increased device proximity can be leveraged to reduce global wire length through the implementation of vertical inter-layer interconnects as shown in Figure 1-2b. The schematic in Figure 1-3 qualitatively shows the reduced maximum distance that a signal would have to travel in a 3D system compared to the same circuit implemented in a 2D fashion. Yu [16] indicates that combining the elements of a 1 cm square microprocessor into a 3D chip stack can reduce the maximum length of a global interconnect from 20,000 µm to 10 µm. This dramatic reduction in wire length implies that 3D architectures can offer relief to the interconnect delay problem discussed above. Using detailed analytical wiring models, Banerjee *et al.* were able to quantify the magnitude of interconnect delay reduction for a particular case [7]. Banerjee *et. al.* showed that a 63% reduction in interconnect delay time can be had at the 50 nm technology node by arranging a planar microprocessor into a two-layer 3D chip stack with the same footprint area. Furthermore, this 63% reduction in delay time allows the 3D chip stack to operate at double the frequency of its 2D counterpart (6 GHz versus 3 GHz). Similar reports by other authors [8]-[10] further indicate the potential for 3D chip architectures to extend the Moore's Law progression in transistor count and processing speed beyond the interconnect delay bottleneck that is expected to occur in traditional planar architectures.

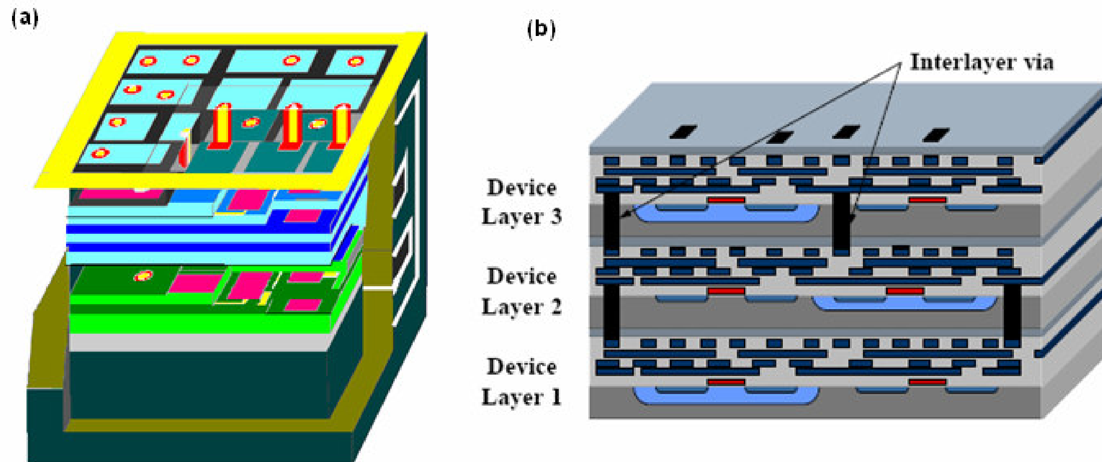


Figure 1-2: (a) Schematic of a 3D chip stack concept (adopted from [13]); (b) a two-dimensional representation of the layers and interconnects in a 3D chip stack [14]

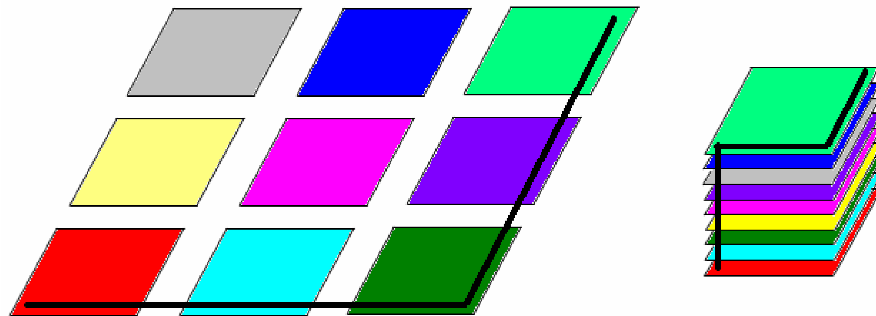


Figure 1-3: Reduction of maximum wire length due to chip stacking (adapted from [12])

Three-dimensional chip stacks have garnered considerable attention not only for their potential to extend Moore's Law for microprocessors, but also for their prospective enablement of high-end system-on-a-chip (SOC) designs [18]. Conceptually, SOC designs involve the convergence of disparate system technologies onto a single chip with the goals of increasing performance, minimizing volume, and reducing cost. It is

anticipated that diverse analog and digital technologies (*e.g.*, logic, memory, RF, bioelectronics, MEMS, etc.) will be combined in SOC designs for numerous applications including wireless multimedia communication, aircraft control and safety, chemical and biohazard detection, and various medical applications [7]. Planar implementation of high-performance SOC designs with large-scale integration will likely yield oversized chips with an abundance of excessively long interconnects [7]. As previously discussed, these conditions conspire to generate parasitic interconnect delays which can, again, be mitigated with vertical interconnections in 3D chip stacks. Furthermore, the wafer stacking method that is commonly used to manufacture 3D chip stacks naturally lends itself to the integration of dissimilar technologies (see Figure 1-4 and the accompanying caption). Technologies with dissimilar operational constraints or incompatible wafer processing methods can be placed on separate but interconnected layers so as to allow heterogeneous integration. For instance, digital and analog technologies can be created on different layers of the chip stack in order to minimize electromagnetic noise that might interfere with their normal operation [7]. Taken together, the reduced interconnect delays and "technology shelving" afforded by 3D chip stacks make them a promising architecture for the realization of highly integrated systems-on-a-chip.

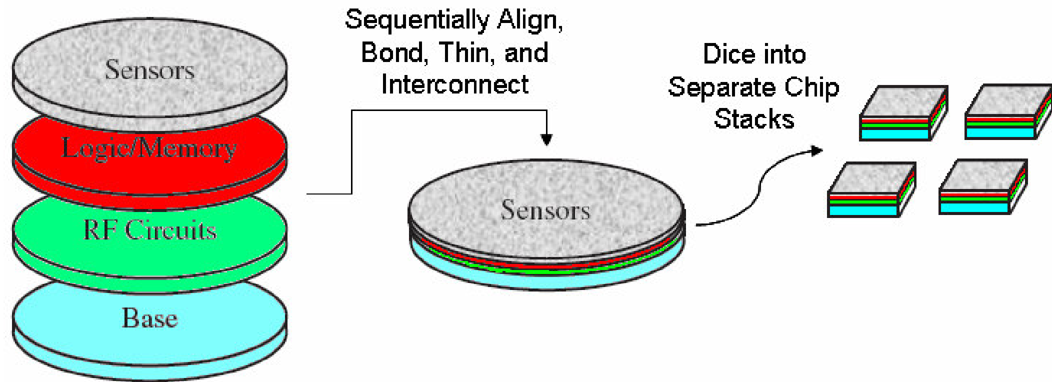


Figure 1-4: Schematic representation of 3D SOC fabrication (adopted from [19])

Thermal Challenges in 3D Chip Stacks

While the previous section serves to elucidate the potential for 3D chip architectures to meet future performance and integration demands, it should be noted that true 3D chip stacking is a relatively immature technology that faces several obstacles to its broader implementation. The commercial introduction of highly-integrated 3D systems will depend on the development of several key enabling technologies such as: 1) reliable vertical inter-layer interconnects, 2) precise alignment methods for wafer stacking, 3) refined testing procedures that can accurately assess the integrity of internal connections, 4) design software for automated routing of complex 3D interconnect paths, and 5) novel thermal management techniques [19]. The latter of these is particularly critical since severe thermal conditions are expected to develop in power-dense 3D chip stacks [7]. Novel cooling techniques and thermally-aware circuit designs need to be developed and implemented to sufficiently limit chip temperatures. This section will discuss the potential for acute thermal problems to arise in 3D chip stacks and briefly present examples of thermal management approaches.

The thermal management challenges facing traditional high performance planar microprocessors have been well documented [17]. It is noted that the ever-present demand for greater chip performance is accompanied by an attendant rise in maximum on-chip power dissipation and peak heat flux. Roadmap projections by the International Electronics Manufacturing Initiative (iNEMI) indicate that these thermal metrics will approach 350 W and 190 W/cm^2, respectively, for high-performance 2D microprocessors within the next decade [50]. Furthermore, increasingly non-uniform on-chip power dissipation is leading to the development of localized regions of high heat flux, or 'flux-spots,' which can significantly elevate local temperatures and cause extreme thermal gradients. In the absence of sufficient cooling, the on-chip temperatures associated with these severe thermal loads can degrade processor performance and reduce reliability [51]. Unfortunately, the reduced transistor size, increased processing speed, and larger-scale integration afforded by 3D chip stacks only serve to exacerbate the thermal issues found in planar chips [7]. As such, traditional air-cooled heat sinks may not be able to sufficiently cool high-performance 3D chip stacks without significant weight and volume penalties.

It is immediately evident that 3D integration of an otherwise planar chip will yield a sharp increase in volumetric power density due to simultaneous performance enhancement and reduction of exposed area available for cooling.. The increased power density in high-performance 3D chip stacks renders traditional air cooling schemes, developed for planar chips, inadequate. These shortcomings are made evident in Figure 1-5, which depicts the attachment of a traditional air-cooled heat sink to a 3D SOC. It can be seen that stacking the active layers requires internally generated heat to be

conducted through adjacent layers before arriving at the interface with heat sink. For instance, the conduction path from the logic layer to the heat sink in Figure 1-5 is extremely inefficient since heat must be conducted through multiple device layers and low thermal conductivity inter-layer dielectric materials (typical dielectric materials have thermal conductivities on the order of 0.3 W/m-K [23]). Therefore, the top layer must be overcooled in order to sufficiently reduce device temperatures deep within the stack.

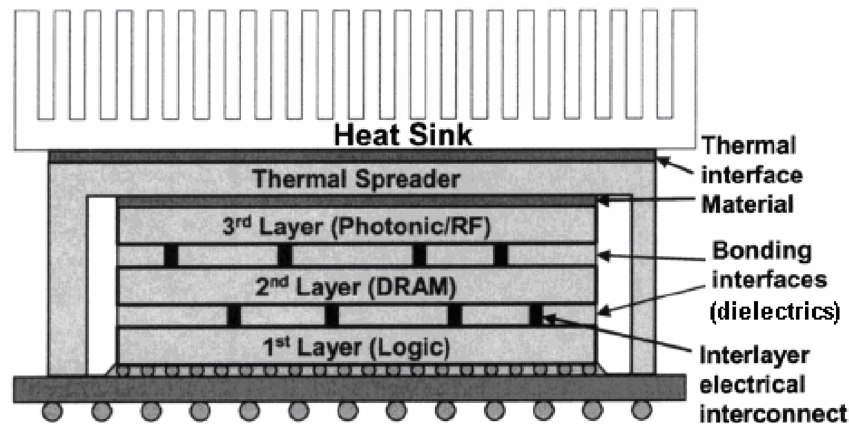


Figure 1-5: Traditional air-cooled heat sink applied to a 3D stack (adapted from [20])

Loi *et al.* [15] analytically compared the performance of 2D and 3D implementations of a processor/cache-memory system acted on by a traditional air-cooled heat sink and subject to a 100C maximum operating temperature constraint (the thermal resistance from the package and heat sink is quoted as 0.7 K/W with a 45C ambient). The authors found that the temperature constraint consistently limited the 3D chip stack to operate at a lower clock frequency than the planar system. Another comparison between planar and 3D chip architectures [22] reveals that a traditional air-cooled heat sink applied to a 4-layer chip stack variant of the Alpha 21364 processor yields a 33C

increase in maximum temperature over its planar counterpart. These examples serve to illustrate that the performance gains made possible by the use of 3D chip stacks may not be fully realized if traditional thermal management approaches are applied.

Potential methods for reducing high temperatures and large temperature gradients in 3D chip stacks can be classified into two broad categories: 1) thermal-aware circuit design and 2) enhancement of thermal transport. In the former approach, electronic floorplanning and processor resource allocation techniques are employed to reduce device temperatures, while the latter approach involves novel application of conductive, convective, and radiative heat transfer mechanisms. The most successful thermal designs for high performance 3D chip stacks will likely concentrate on thermal transport enhancement because, despite the benefits provided by thermal-aware circuit design methods – such as partition-driven [24] and force-directed [25] standard cell placement, dynamic voltage scaling [26], and global clock gating [27] – their implementation can cause up to a 36% slowdown in operating speed [27]. Alternatively, thermal transport enhancement methods can deliver effective and targeted cooling to critical portions of the circuit without causing slowdown.

The removal of internally generated heat is a critical challenge facing the successful implementation of 3D chip stack cooling methods. In a purely external cooling approach, one might attempt to enhance the conductive path between the internal device layers and system envelope while simultaneously applying a high external heat transfer coefficient. In contrast, an internal cooling approach might attempt to shorten the conductive path by bringing the cooling mechanism into closer contact with internal heat sources. A survey of the heat transfer coefficients provided by various cooling

methods is shown in Figure 1-6. The high heat transfer coefficients associated with advanced liquid cooling techniques – such as cooling by immersed pool boiling, microjet impingment, and spray cooling – make such methods prime candidates for external cooling of high performance 3D chip stacks. An externally applied advanced liquid cooling method, combined with the use of dedicated thermal vias [38] or interleaved diamond spreaders [39] to enhance conduction in the stack, could act to sufficiently reduce internal device temperatures.

In contrast, relatively low heat transfer coefficients can be afforded through the use of internal cooling – with imbedded microchannels, for instance [20] – since the area lost during chip stacking is regained by introducing liquid internally. Internal cooling schemes have the advantage of leaving envelope area open for signal transmission and reception, a function that is becoming increasing prevalent as RF and optical devices are integrated in 3D SOC packages [7]. Also, since internal cooling approaches tend to rely less on conduction in the stack, they are typically more scalable than external cooling methods. However, implementing the manufacture of small microchannels into the already demanding process flow for 3D chip stacks is considerably more difficult than introducing thermal vias to aid the performance of externally applied cooling. The benefits and drawbacks of internal and external cooling methods should be evaluated when developing a 3D chip stack cooling scheme so that the most effective thermal enhancement approach is implemented and the use of performance draining 'thermal-aware circuit design' can be kept at a minimum.

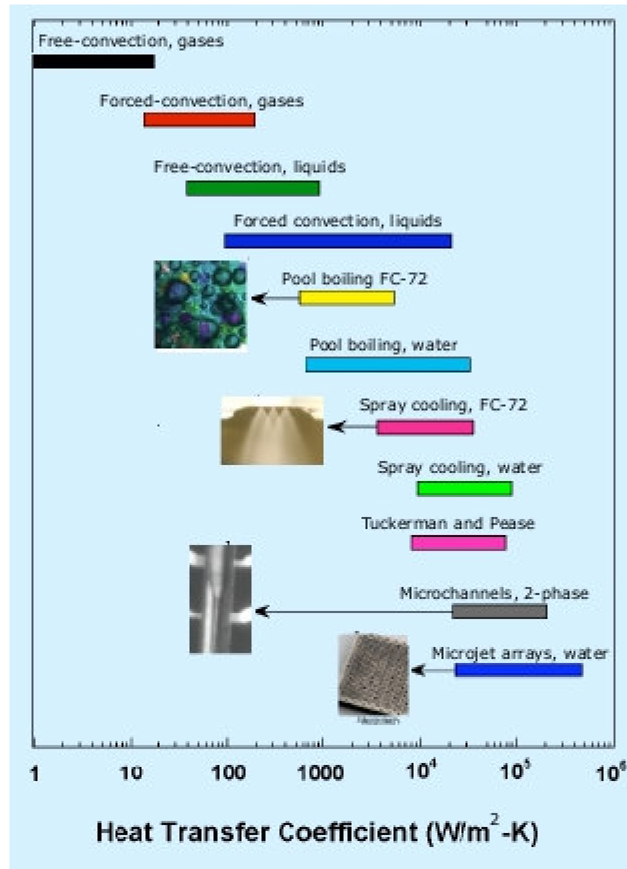


Figure 1-6: The heat transfer coefficients provided by different cooling technologies

Thesis Organization

The previous section discussed the potential for acute thermal problems to occur in 3D chip stacks. In the face of these thermal issues, lack of sufficient cooling methods could threaten the broader implementation of high performance 3D integrated circuits. Therefore, it is necessary create and explore novel thermal management techniques through the development of accurate thermal models. To this end, this thesis will focus on three separate, but sometimes related, facets of cooling 3D chip architectures:

- The highly complex nature of three-dimensional circuitry can make numerical modeling difficult due to the large number of elements that are

needed to explicitly resolve all geometric features. The development of compact models is thus important to the efficient assessment of potential cooling methods. One aspect of compact model development is the determination of equivalent thermal conductivity. In Chapter 2, a representative model of a layer in a 3D chip stack will be introduced and its equivalent thermal conductivity will be determined by both analytical and numerical methods.

- In Chapter 3, application of direct single phase internal liquid cooling with the dielectric liquid, FC-72, will be explored for a novel hybrid 3D chip stack. Parametric sensitivities to system geometries, heat generation, and fluid inlet conditions will be explored through the use of ANSYS CFX, a commercial computational fluid dynamics (CFD) simulation package. Also, analytical methods will be introduced throughout the chapter and used to validate the CFD models.
- Chapter 4 will explore the potential for orthotropic spreading materials to be implemented in a 3D chip stack and reduce the detrimental high temperatures and sharp gradients that are associated with localized 'flux-spots.' First, an analytical solution found in the literature will be presented and discussed. Then, use will be made of this analytical solution to explore parametric sensitivities of in-plane conductivity, spreader thickness, chip thickness, flux-spot size, and heat transfer coefficient. Finally, the detrimental effects of an interfacial contact

resistance will be investigated through the use of a validated finite element model in ANSYS.

Detailed conclusion sections will be provided at the end of each Chapters 2, 3, and 4 in addition to the broad conclusions presented in final chapter.

Chapter 2 : Equivalent Thermal Conductivity Determination

Introduction

Three-dimensional circuit integration provides an extra degree of freedom in system design which can lead to a substantial increase in overall system complexity. In particular, the intricate three-dimensional routing of interconnect wires is a primary source of added complexity and feature crowding in 3D chip stacks [28]. As feature sizes shrink and system complexity escalates, the use of compact thermal models becomes critical to the efficient evaluation of a potential cooling scheme. Resistance networks based on electrothermal analogies and numerical finite element models benefit greatly from the use of simplified geometries with equivalent thermal properties.

For example, consider the five-layer 3D chip stack in Figure 2-1a, where each discrete layer is made up of an array of similar electronics. Figure 2-1b depicts a top view of a single layer in the chip stack where the gray dielectric material has been removed to expose the details of the circuitry. It can be seen from the enlarged pictures in Figure 2-1c and Figure 2-1d that the layer contains numerous small-scale electronic components, vias, and interconnects. These small features can complicate the detailed analysis of a global cooling scheme. Suppose, for instance, that the finite element method is used to explore the effectiveness of an envelope spray cooling approach for this chip stack. Under such circumstances, it would be ideal to generate a global model of the 3D stack with all circuit details fully modeled; however, the abundance of small features makes explicit discretization of the entire model problematic without significant computational resources. Alternatively, a compact thermal model could be developed in

which homogeneous blocks with equivalent thermal conductivity are used to represent the presence of a detailed layer. By removing the geometric details of the system and accurately capturing their affect on heat transfer through the use of equivalent thermal conductivities, one can save considerable computational time and sacrifice little in the way of accuracy.

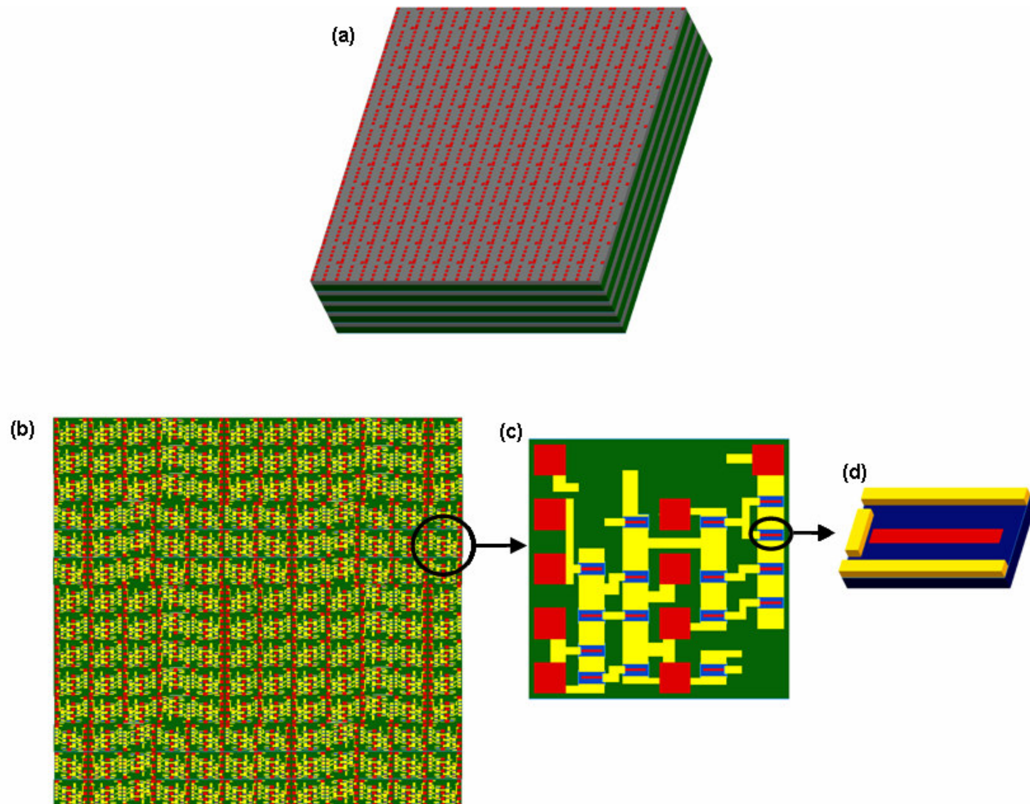


Figure 2-1: (a) Five-layer 3D chip stack; (b) Top view of a single layer with dielectric material removed to expose electronics and interconnects; (c) a single 'unit cell' of the layer; (d) an individual heat source within one unit cell

Several compact models for 3D chip architectures have been reported in recent years. Palacin *et al.* were able to develop a dynamic compact thermal model to study the transient thermal response of a 3D chip stack with multiple power sources [29]. Chiang

et al. created a compact thermal model to account for both heat generation in active device layers and interconnect joule heating [23]; also, the authors were able to corroborate the results of their compact model with the results of a detailed finite element analysis. Most interestingly, Arik, Garg, and Bar-Cohen demonstrated that use can be made of equivalent thermal conductivities to develop a compact and efficient numerical model for a three-dimensional high heat flux system-on-package design [30]. By replacing complex interconnect structures with homogenous blocks of equivalent thermal conductivity, a compact finite element model requiring only 36% of the elements of a detailed model was shown to predict the maximum chip and substrate temperatures to within 1.6% and 3%, respectively. This analysis points to the utility of equivalent thermal conductivities in the development of compact thermal models. In this chapter, approximate analytical methods and numerical simulations will be used to find the equivalent thermal conductivity for a representative local model of a single layer in a 3D chip stack. The analytically determined thermal conductivities will be compared to the values provided by numerical simulation, after establishing the mesh-insensitivity of the numerical model.

A Representative Local Model

The intent of this chapter is to explore and compare analytical and numerical methods for determining the equivalent thermal conductivity of a nominal device layer in a 3D chip stack. To the extent that the layer of interest exhibits spatially similar electronic circuitry, a small representative portion of the layer – a 'local model' – can be used to capture the

equivalent thermal properties of the entire layer. For instance, the properties of the layer in Figure 2-1b can be found from analysis of the local model shown in Figure 2-1c.

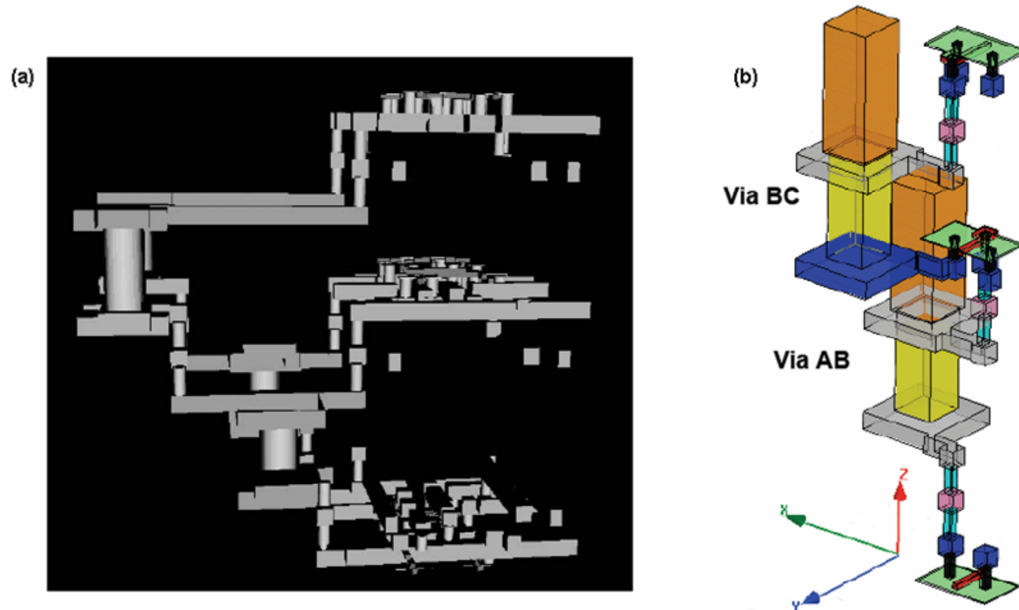


Figure 2-2: (a) Interconnections between three cells of a 3D memory device [31]; (b) Vertical interconnects between a single active device on the lowest layer and various active devices in higher layers, with capture pads between layers [32]

The open literature was consulted in order to arrive at a representative local model for use in this chapter. As noted in Chapter 1, the layers of a 3D chip stack are commonly composed of a semiconductor material – such as silicon, indium phosphide, or gallium arsenide – with a dielectric material separating adjacent layers [7]. The dielectric layer is commonly benzocyclobutene (BCB), but can also be methylsilsesquioxane or Parylene-N [19]. The literature search also revealed that metallization and vertical interconnects generally exhibit the patterns shown in Figure 2-2, where vertical connections have either cylindrical or square crossections and are capped by rectangular 'capture pads' on either end. In the commonly used wafer bonding manufacturing

process for 3D chip stacks, the wafer alignment precision constrains vertical inter-layer interconnect sizes to be at least 5 μm [33]. Also, the pitch between layers in the wafer bonding process can be as small as 25 μm [29].

The above characteristics were used as guidelines in developing the representative local model seen in Figure 2-3. This 50 μm x 50 μm local model consists of a 20 μm thick indium phosphide (InP) substrate, a 13 μm thick dielectric layer of BCB, 10 gold vertical interconnects (called 'vias' for brevity), and 16 InP active devices. The quantity and placement of both the vias and active devices were chosen arbitrarily. It can be seen that the vias possess a square crossection in the BCB layer (6 μm on a side) and a circular crossection in the InP substrate (6 μm diameter); also, each via has a 'capture pad' on the top and bottom for mating to vias in adjacent layers. The active devices have 'first interconnect' (FIC) gold metallization attached to them in three locations as shown in Figure 2-3d. The horizontal gold traces throughout the local model were created somewhat arbitrarily, with the only constraint being that 25% area coverage must be achieved (*i.e.* about 600 μm^2). The thermal conductivities of the three materials used in the local model – BCB, InP, and gold – were found in sources [34], [35], and [36], respectively, and are reported in Table 2-1. Larger and more detailed pictures of the local model with critical dimensions labeled can be seen in Appendix A. The local model reported in this section will be used throughout the remainder of the chapter to demonstrate the determination of equivalent thermal conductivity.

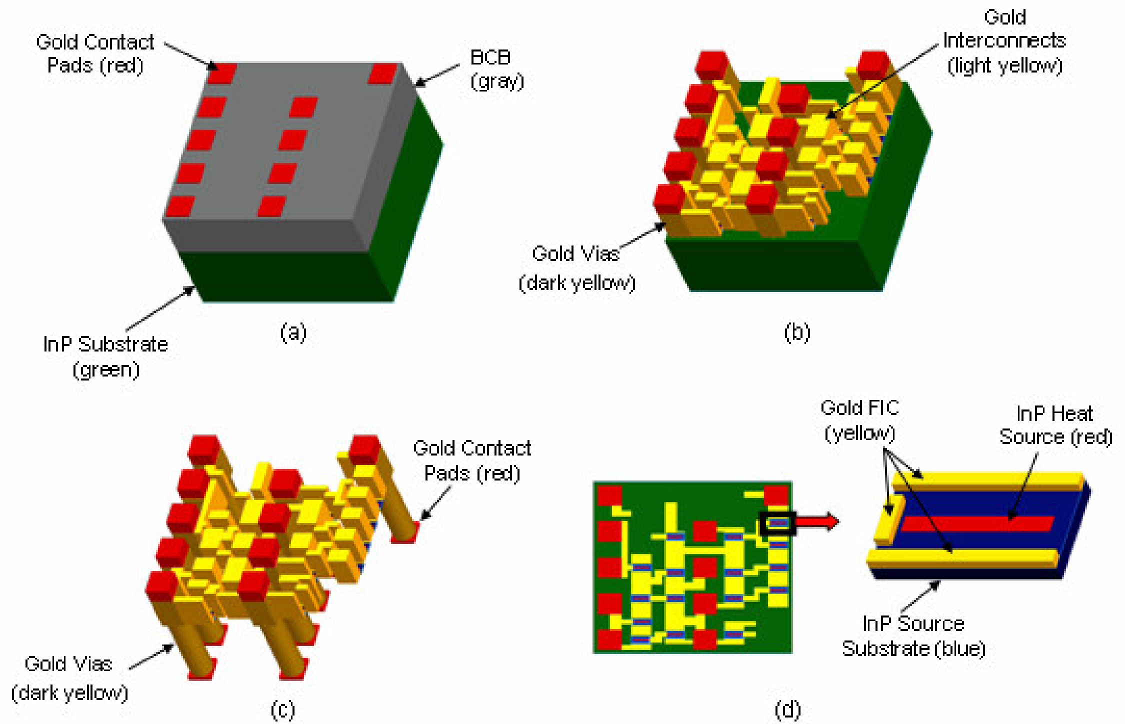


Figure 2-3: (a) an isometric view of the 3D local model, (b) with the BCB removed, (c) with BCB and InP substrate removed, (d) a top view of the model with BCB removed and details of a single heat source (*i.e.* active device)

Table 2-1: Thermal conductivities of the three materials used in the local model [34]-[36]

Material	k (W/m-K)
BCB	0.29
InP	68
Gold	317

Approximate Analytical Determination of Equivalent Thermal Conductivity

The well known Fourier Law for heat conduction in one dimension is given by

$$\dot{q} = -kA\frac{dT}{dx} \tag{1}$$

where $\dot{q}$ is the applied heat rate (W), k is the thermal conductivity of the medium (W/m-K), A is the crossectional area perpendicular to the direction of heat flow (m^2), and dT/dx is the temperature gradient (K/m). Considering one-dimensional heat flow in the classic plane wall shown in Figure 2-4, Equation 1 can be rewritten as

$$\frac{T_1 - T_2}{\dot{q}} = \frac{L}{kA} \tag{2}$$

where the negative sign has been dropped due to the indicated direction of heat flow. Using analogies to basic electronic circuit theory, one can understand that the temperature difference is like voltage, V, in that it provides the driving potential for heat flow. From this perspective, the quantity $\dot{q}$ is analogous to electric current, I, and knowledge of Ohm's Law, V/I = R, reveals that the right hand side of Equation 2 must be like a resistance to thermal conduction. Therefore, we can define the thermal conduction resistance as

$$R = \frac{L}{kA} \tag{3}$$

where L is the one-dimensional conduction length.

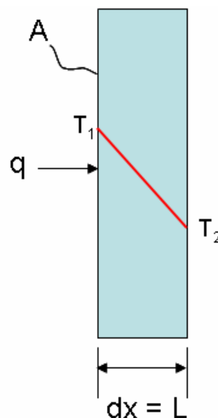


Figure 2-4: Classic one-dimensional conduction of heat in a plane wall

This interpretation of heat conduction is of great utility when attempting to determine the equivalent thermal conductivity of a composite domain with multiple conduction paths. In many cases, the conduction paths in the model can be represented by a network of thermal resistances, and this resistance network can be collapsed into a single equivalent thermal resistance, R_{eq}, using standard rules for combining resistances in series and parallel. Then, knowing the total conduction length, L, and crossectional area, A, from model dimensions, one can solve for the equivalent thermal conductivity of the model, k_{eq}. This method will be used in this section to approximately determine the equivalent thermal conductivity of the local model in Figure 2-3.

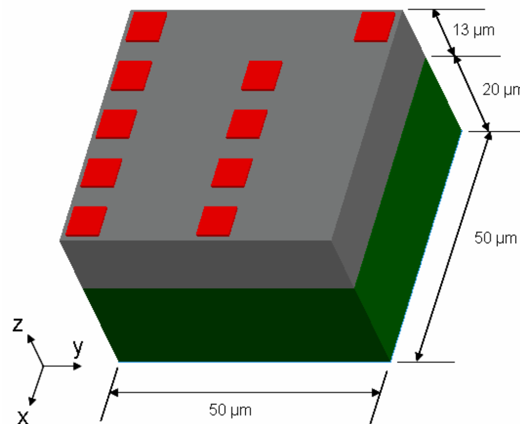


Figure 2-5: The local model with outer dimensions and coordinate system labeled

It is desirable to determine the equivalent conductivities of the local model in each of the three coordinate directions shown in Figure 2-5. First, conduction in the z-direction will be considered. Heat being conducted in the z-direction can take one of two main paths: 1) directly through the gold vias, or 2) through the BCB and InP substrate. Given the relatively low thermal conductivity of the BCB material, one might expect that

very little heat will flow through the latter path. However, as shown in Figure 2-6, there is a significant amount of highly conductive gold embedded in the bulk BCB material (due to the presence of horizontal interconnects). Therefore, heat being conducted through the upper portion of the model actually takes parallel paths through the BCB material and high conductivity, low resistance gold. The resistance network resulting from these idealized one-dimensional conduction paths is shown in Figure 2-7 and Table 2-2 describes each resistance in detail.

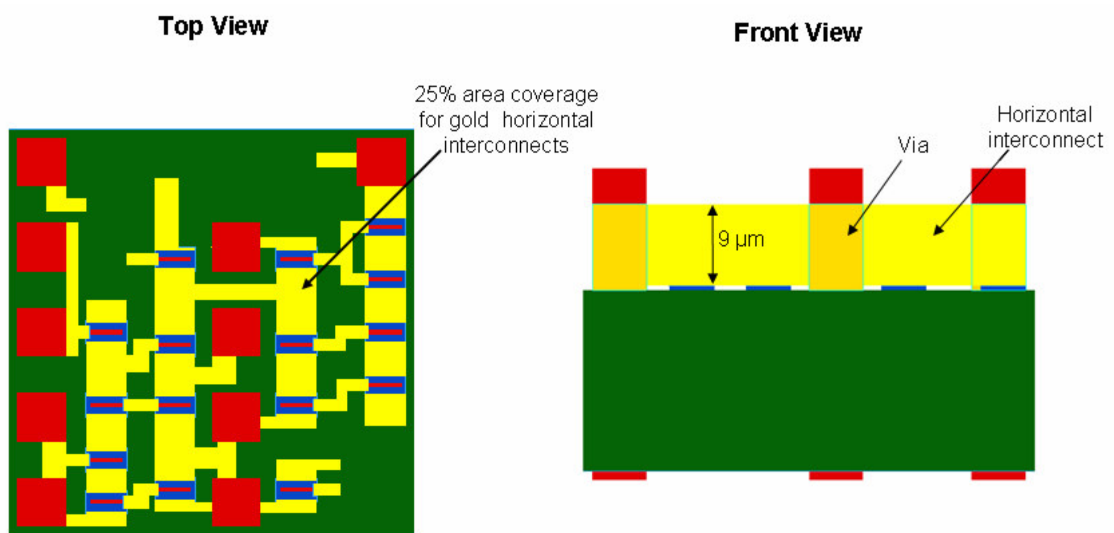


Figure 2-6: Top and front views of the local model with BCB removed to show the significance of the embedded interconnect gold

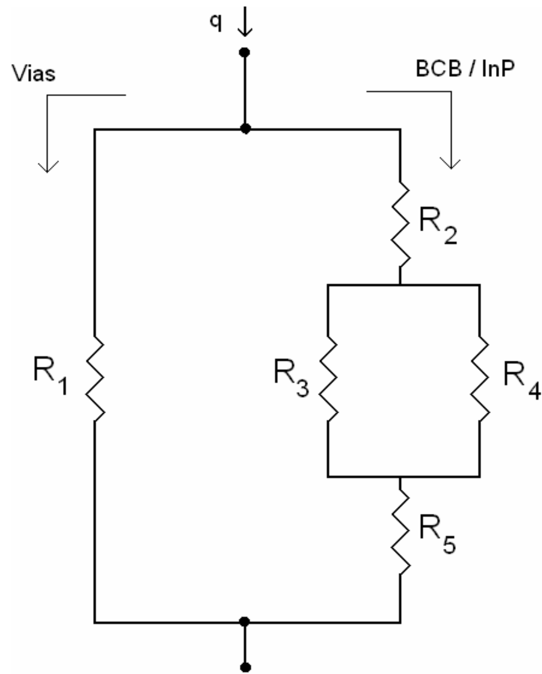


Figure 2-7: Resistance network for heat conduction in the z-direction of the local model

Table 2-2: Detailed description of the thermal resistances in Figure 2-7

Resistance #	**Path Represented**	**L [µm]**	**k [W/m-K]**	**A [m^2]**	**$R = L/kA$ [K/W]**
R_1	Thru vias	33	317	3.6×10^{-10}	289.17
R_2	Thru solid BCB	4	0.29	2.14×10^{-9}	6445.38
R_3	Thru BCB adjacent to horizontal i.c. gold	9	0.29	$0.75 \times 2.14 \times 10^{-9}$	19336.10
R_4	Thru horizontal i.c. gold in BCB	9	317	$0.25 \times 2.14 \times 10^{-9}$	53.07
R_5	Thru InP substrate	20	68	2.14×10^{-9}	137.44

Using the rules for combining resistances in series and parallel, it can be shown that an equivalent resistance for the system in Figure 2-7 is given by

$$R_{eq,z} = \frac{R_1\left[R_2 + \left(\frac{R_3 R_4}{R_3 + R_4}\right) + R_5\right]}{R_1 + R_2 + \left(\frac{R_3 R_4}{R_3 + R_4}\right) + R_5} = 277.1 \ K/W \tag{4}$$

Thus, the equivalent thermal conductivity in the z-direction, $k_{eq,z}$, can be calculated as

$$k_{eq,z} = \left(\frac{R_{eq,z} A_{base}}{L_z}\right)^{-1} = \boxed{47.64 \ W/mK} \tag{5}$$

where A_{base} is the footprint area of the local model (A_{base} = 50 x 50 μm^2), and L_z is the length of the model in the z-direction (L_z = 13 μm + 20 μm = 33 μm).

We now seek the equivalent conductivity in the x- and y- coordinate directions. Any heat applied in these directions will flow through the network of horizontal gold interconnects in the BCB layer. As seen in the top view of Figure 2-6, the path traced by this gold is somewhat complicated, with various crossectional areas and branching pathways. Also, the horizontal interconnects are typically only connected to one another through the FIC gold on the active devices (see Figure 2-3d). This FIC gold does not maintain continuity and is only a fraction of a micron tall, while the horizontal interconnects are 9 μm tall (see Appendix A for dimensions); thus, each interface between a gold trace and an active device represents a major constriction in the conduction path. Since the horizontal gold interconnects do not create a robust thermal path, one might anticipate that the low conductivity BCB material will dominate the conduction of heat in the upper layer of the model. Similarly, it is expected that InP will dominate the conduction of heat in the lower layer of the model, despite the presence of

the gold vias. Therefore, to a first approximation, the resistance network for both the x- and y- directions is simply given by Figure 2-8, where R_6 and R_7 are based on conduction through solid blocks of BCB and InP, respectively. Detailed descriptions of each resistance in this network are shown in Table 2-3.

The parallel combination of R_6 and R_7 is used to find $R_{eq,xy}$ as the following

$$R_{eq,xy} = \frac{R_6 R_7}{R_6 + R_7} = 733.26 \ \ K/W \tag{6}$$

and the equivalent thermal conductivity in the x- and y- directions, $k_{eq,xy}$, can be calculated as

$$k_{eq,xy} = \left(\frac{R_{eq,xy} A_{side}}{L_{xy}} \right)^{-1} = \boxed{41.33 \ \ W/mK} \tag{7}$$

where A_{side} is the crossectional area of the layer as seen from one of its sides (A_{side} = 50 µm x 33 µm = 1650 µm^2) and L_{xy} is the conduction length in either the x- or y- coordinate direction (L_{xy} = 50 µm).

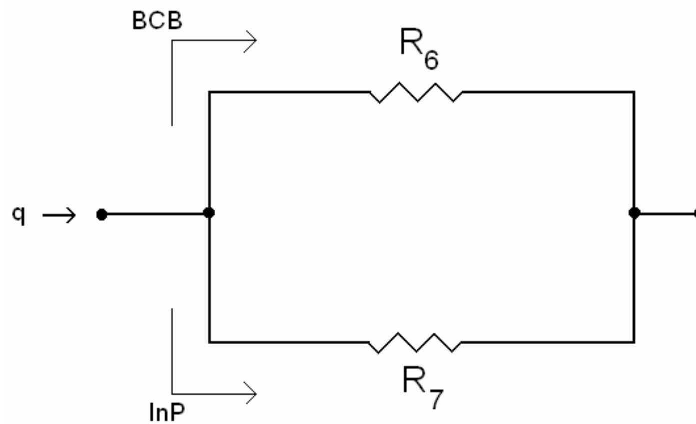


Figure 2-8: Resistance network for heat conduction in the xy-plane of the local model

Table 2-3: Detailed description of the thermal resistances in Figure 2-8

Resistance #	**Path Represented**	**L [µm]**	**k [W/m-K]**	**A [m^2]**	**$R = L/kA$ [K/W]**
R_6	Thru BCB upper layer of model	50	0.29	6.5×10^{-10}	265252
R_7	Thru InP lower layer of model	50	68	1×10^{-9}	735.30

Comparison of Equations 5 and 7 reveals that the equivalent thermal conductivity in the z-direction is higher than that in the xy plane. This difference is predominately attributable to the orientation of the gold vias. The high conductivity gold vias run vertically through the layer and thus provide a robust thermal path in the z-direction.

Numerical Determination of Equivalent Thermal Conductivity

As an alternative to the approximate analytical methods described in the previous section, use can be made of the finite element method (FEM) to determine the equivalent thermal conductivity of the local model. In this approach, the model geometry is created virtually in a three-dimensional computer aided design (CAD) package and then imported into finite element analysis software where it is discretized with many small elements. Each finite element is comprised of several nodes and the heat conduction equation is linearized over this collection of nodes in order to arrive at an approximate temperature distribution. When appropriate boundary conditions are applied to the local model, the resulting temperature field can be used to determine the equivalent thermal conductivity of the model.

The local model in Figure 2-3 was created in the commercial CAD program, Pro/Engineer; the Pro/Engineer geometry was then imported to the commercial FEM

package, ANSYS, for model descretization. A ten-node tetrahedral element called 'Solid87' was used to discretize the model with varying degrees of resolution. The simulation method and boundary conditions used to determine the equivalent thermal conductivity of the model will now be discussed; please refer to Figure 2-9 as the discussion proceeds.

First, high-conductivity blocks (of essentially infinite thermal conductivity) were added to opposing sides of the local model in whatever direction the equivalent conductivity was desired. These blocks are only simulation features and are not part of the true system geometry. For instance, in Figure 2-9, the equivalent conductivity is being determined along the z-axis (as indicated by the coordinate system in the figure), so the high-conductivity blocks are placed on the top and bottom of the local model. Second, boundary conditions were applied to the model. An arbitrary heat flux was applied to one of the high-conductivity blocks and an arbitrary temperature was specified on the other block. The boundary conditions on the remaining outer surfaces of the model were made adiabatic to ensure that the bulk flow of heat would be from one high-conductivity block to the other. Finally, upon running the simulation, the block with the heat flux boundary condition attains some constant temperature (since the conductivity of the block is essentially infinite). The temperature difference, ΔT, between the high-conductivity blocks was then used in Fourier's law to determine the equivalent conductivity of the model (the conduction distance, Δx, and the heat flux, q'', are already known from model dimensions and the arbitrarily applied heat flux, respectively).

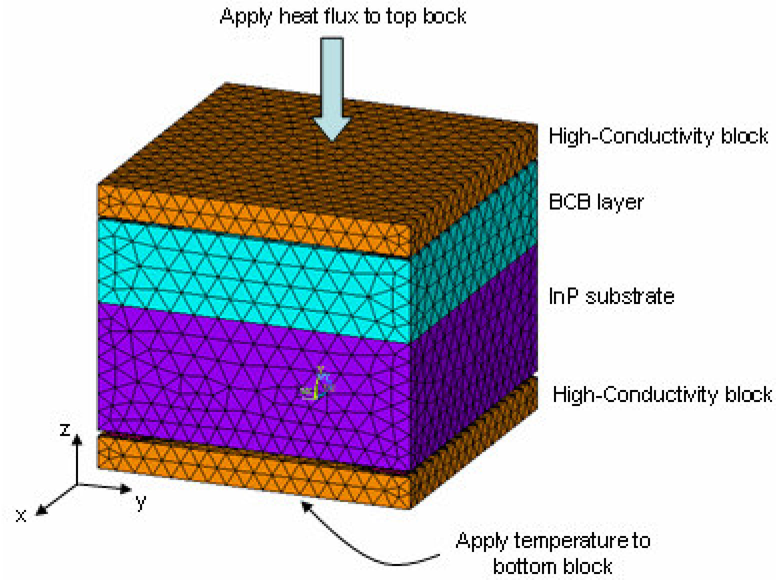


Figure 2-9: Method for determining the equivalent thermal conductivity in ANSYS

The approach outlined above for finding the equivalent thermal conductivity was verified in a general way by using the simple model shown in Figure 2-10. The validation model consists of an array of gold cylinders that run completely through the thickness of a solid indium phosphide slab (cylinder diameter: 5 μm; slab dimensions: 50 x 50 x 20 μm). This configuration was chosen because a simple analytical calculation can be used to determine the equivalent thermal conductivity in the z-direction (the coordinate system is indicated in Figure 2-10). The analytical expression for the equivalent conductivity is given by the following:

$$k_{eq} = k_{gold}\left(\frac{A_{gold}}{A_{total}}\right) + k_{InP}\left(\frac{A_{InP}}{A_{total}}\right) \tag{8}$$

where A_{total} is the total cross-sectional area in the xy-plane ($A_{total} = A_{gold} + A_{InP}$), and the conductivities, k_{gold} and k_{InP}, are given in Table 2-1. Upon substitution of values, the analytical expression indicates that the equivalent thermal conductivity should be 116.9

W/m-K. To check this in ANSYS, two high-conductivity blocks were placed on the top and bottom of the model (analogous to what is shown in Figure 2-9). The boundary conditions were imposed as described in the previous paragraph and Fourier's Law was used to determine the equivalent conductivity. The value of k_{eq} found through the ANSYS simulation was 115.7 W/m-K, which is only a ~1% error from that predicted by the analytical expression. This provides first-order verification that the method used to determine the equivalent thermal conductivity is accurate.

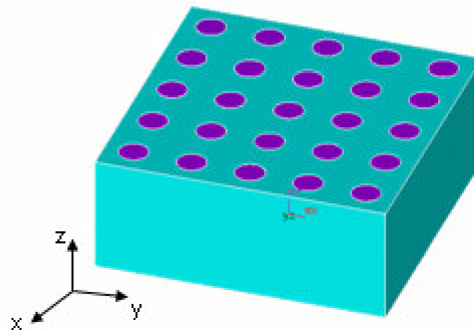


Figure 2-10: Verification model – 5 x 5 array of gold cylinders embedded in an InP slab

With the understanding that the numerical approach outlined above can produce accurate estimates for the equivalent thermal conductivity, an initial ANSYS simulation was run to determine $k_{eq,x}$, $k_{eq,y}$, and $k_{eq,z}$ for the local model. In this initial simulation, the model was discretized into 72,661 elements and the resulting equivalent thermal conductivities are reported in Table 2-4. In congruence with the results of the previous section, the highest equivalent conductivity was found to occur along the z-axis and the equivalent conductivities in the x- and y- directions were found to be in close proximity.

However, the numerically determined conductivities are seen to be 15% to 24% larger than the approximate analytical values determined in the previous section.

Table 2-4: Numerically determined equivalent conductivities for the local model

Equivalent Conductivity	**Value [W/m-K]**	**% greater than analytical**
$k_{eq,x}$	49.9	20.7
$k_{eq,y}$	47.8	15.7
$k_{eq,z}$	59.0	23.8

For fear that the initial discretization was too coarse, a mesh sensitivity analysis was conducted for all ANSYS simulations. In these sensitivity analyses, the variation of the temperature field was observed for different levels of mesh refinement. Each group of components in the local model was given a name and assigned a mesh multiplier as shown in Table 2-5. Between trials of a given analysis, the mesh multipliers were varied in order to change the resolution of the mesh in certain components (a smaller mesh multiplier yields a more refined mesh). Table 2-5 shows representative mesh sensitivity results from simulations designed to find the equivalent conductivity along the z-axis (see Figure 2-9 for the configuration). Comparison of trial 1 and trial 7 in Table 2-5 reveals that a 13 fold increase in the element count – to nearly 1,000,000 elements – yields a difference of less than 1% in the maximum temperature rise. In these simulations, the maximum temperature is achieved in the high-conductivity block where the heat flux is applied, and it is the key value used in determining the equivalent conductivity. Thus, it is found that the numerically determined equivalent conductivities are quite insensitive to mesh refinement beyond the initial discretization. Coupled with the validation model

presented above, this mesh sensitivity analysis serves to verify that the numerically determined equivalent conductivities reported in Table 2-4 are accurate and that three-dimensional conduction effects in the individual layers account for the discrepancy between values obtained from the well discretized numerical models and the first-order analytical approximations based on parallel heat flow paths.

Table 2-5: Mesh sensitivity analysis for k_{eq} along the z-axis

Mesh Sensitivity Analysis							
Entity	**Multiplier**						
	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6	Trial 7
source	10	10	5	2	5	5	2
gold	15	10	10	10	10	5	5
viabcb	15	10	10	10	10	5	5
viasub	15	10	10	10	10	5	5
tpad	10	10	10	10	3	3	4
bpad	10	10	10	10	3	3	4
bcb	20	15	15	15	15	10	12
sub	20	15	15	15	15	10	12
hik	20	15	15	15	15	10	12
Statistics							
Element count	72661	125406	144434	282364	401705	846721	945201
Solve Time (mins)	0.25	0.38	0.88	1.05	1.63	6.47	11
Max Temp	0.233236	0.232971	0.233202	0.233823	0.234611	0.234820	0.235020
% Difference from Trial 1	-	0.11%	0.01%	0.25%	0.59%	0.68%	0.76%

Conclusion

Two different methods of determining the equivalent thermal conductivity of a representative layer in a 3D chip stack were presented in this chapter. These equivalent properties find use in global numerical simulations of 3D chip stacks where explicit modeling of complex geometric features can overwhelm traditional computational platforms. In contrast, homogeneous blocks with equivalent thermal properties can be

used in place of exceedingly complex circuitry to arrive at more efficient global numerical models.

Approximate analytical and numerical methods were employed in this chapter to determine the equivalent conductivities of a representative local model of a single layer in a 3D chip stack. The analytical method used was approximate in that it assumed ideal one-dimensional conduction; also, some potential conduction paths were ignored for the sake of simplicity or because it could be reasoned that inclusion of those paths would not significantly impact the results. The numerical method of determining equivalent conductivities involved the use of ANSYS, a commercial FEM package. A validation model for the numerical approach was developed and a thorough mesh sensitivity analysis was conducted to verify the robustness of the numerical results.

It was found that the numerically determined equivalent thermal conductivities were 15% to 24% larger than the analytically determined values. This difference is largely attributable to the approximate nature of the simple parallel/series heat flow paths used in the analytical procedure. It is concluded that the analytical method can save considerable time in approximating the equivalent properties of a layer, but can become cumbersome for very complex circuits. Alternatively, numerical simulations using FEM software can take a considerable amount of time to set up, run, and validate, but they are likely to yield more accurate results for complex circuits like the representative model considered in this chapter.

Chapter 3 : Internal Liquid Cooling with Single Phase Microchannels

Introduction

As discussed in Chapter 1, the issue of heat dissipation is seen as a major roadblock in the development of high performance 3D chip stacks [37]. In particular, the increased power density created by stacking active device layers is presenting a significant challenge to thermal management engineers. Removal of internally generated heat by way of traditional external cooling is made difficult by the use of low thermal conductivity inter-layer dielectric materials. These low conductivity materials yield inefficient thermal conduction paths to the envelope of the system and thus require external cooling approaches to provide very high heat transfer coefficients at the envelope in order to sufficiently cool internal circuitry. However, some promising methods of spreading internally generated heat to the boundaries of 3D chip stacks – such as the use of high conductivity thermal vias [38] and implementation of diamond heat spreaders [39] – have made the use of external cooling schemes more viable.

In contrast to external cooling methods, internal liquid cooling with perflourinated liquids [40] offers the ability to bring effective cooling directly to internal active devices. The inherent benefit of this approach is that the area for heat transfer lost in the chip stacking process is regained, and perhaps surpassed, by introducing fluid internally. This area increase implies that internal cooling schemes can provide the same heat removal rate as an external approach, with a lower heat transfer coefficient. Implementation of internal liquid cooling also has the added benefit of leaving envelope surface area

available for optical and RF devices to receive and transmit data, which is an increasingly important function as true 3D system-on-a-chip designs are realized [7].

Despite the merits of internal liquid cooling for 3D chip stacks, the concept has received very little direct attention in the literature. Koo *et al.* have published perhaps the only thorough paper on the topic [20], in which the benefits of a liquid cooled integrated system of microchannels like those in Figure 3-1 are explored. The authors use a thermal resistance network and an empirical correlation for flow boiling in horizontal channels – proposed by Kandlikar [41] – to examine the benefit of introducing two-phase flow of saturated water (at sub-atmospheric pressure such that T_{sat} = 70C) into a system of 400 μm by 300 μm rectangular microchannels. A 150 W multi-layer chip stack comprised of memory and logic layers was alternately subjected to the microchannel cooling approach and a traditional air-cooled heat sink with a thermal resistance of 0.25 K/W acting on the first device layer. It was found that the microchannel cooling approach yielded greater temperature uniformity in the stack as compared to the air-cooled design, since the microchannels provide well-distributed cooling throughout the internal regions. This result indicates the scalability of internal cooling approaches and implies that their use may be preferred in systems where numerous device layers are required. Also, the temperature of the logic layers – which generated 90% of the heat in these simulations – was reduced considerably in the microchannel cooling approach when the flow boiling heat transfer mechanism was active. In the best performing configuration studied by the authors, the two-phase microchannel approach was able to remove 68 W/cm^2 at each device layer while keeping the junction temperatures below 85C. The air-cooled heat sink was shown to remove 37.5 W/cm^2 at the top device layer, yet internal temperatures

routinely exceeded 110C in chip stacks with more than two device layers. This theoretical study illustrates the potential for a bank of liquid cooled microchannels to remove considerable heat from a 3D chip stack.

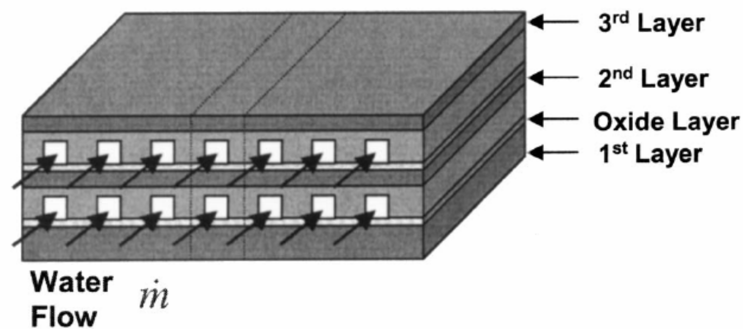


Figure 3-1: Notional implementation of microchannels in a 3D chip stack [20]

In this chapter, application of direct single phase internal liquid cooling with the dielectric liquid, FC-72, will be explored for a novel hybrid 3D chip stack. Parametric sensitivities to system geometries, heat generation, and fluid inlet conditions will be explored through the use of ANSYS CFX, a commercial computational fluid dynamics (CFD) simulation package. Also, analytical methods will be introduced throughout the chapter and used to validate the CFD models.

Novel 3D Chip Stack Model

The hybrid 3D chip stack that will be analyzed in this chapter is shown schematically in Figure 3-2. The model consists of two functional chip stacks as follows: the upper chip stack represents a logic-intensive 3D processor stack with highly dissipative and densely

packed active device layers; meanwhile, the lower chip stack is comprised of more sparse electronic circuitry which represents the presence of low level processing or memory storage. For the sake of brevity, these two distinct portions of the model will be called the 'top chip stack' (TCS) and 'bottom chip stack' (BCS), respectively, from this point forward.

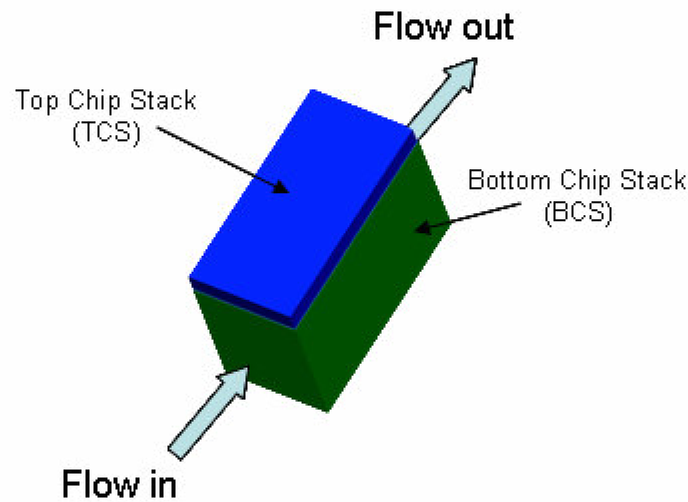


Figure 3-2: Schematic of hybrid chip stack examined in this chapter

The high performance TCS dissipates nearly all of the heat in the hybrid chip stack and thus is a prime candidate for aggressive internal cooling. However, the TCS contains densely packed active devices and an abundance of vertical inter-layer interconnects which may prohibit the use of internal microchannel cooling. Introducing process steps to manufacture banks of microchannels into the already demanding process flow for making high performance 3D chip stacks, such as the TCS, is seen as a major challenge [20]. The implementation of dedicated layers for etching or machining of microchannels can extend the required length of inter-layer interconnects and thus

increase delay times and degrade performance. Also, global interconnect routing is made considerably more complex by the need to funnel all inter-layer interconnects through the available spaces between microchannels.

In contrast to the difficulties presented by the complex nature of the TCS, the BCS contains relatively uncomplicated circuitry and has less stringent performance demands. Thus, the BCS is a more practical and convenient location for the introduction of internal liquid cooling. However, since the BCS produces very little heat in comparison to the TCS, the layers of the BCS ought to be stacked in such a way as to promote heat transfer from the TCS. To this end, one might consider orienting the layers of the BCS so that they run perpendicular to the layers of the TCS, effectively forming a series of 'extended surfaces' below the TCS. Fluid could then be flushed in the gaps in the vertically oriented BCS layers so as to draw heat out of the TCS in a manner analogous to a heat sink. This conceptual design is depicted schematically in Figure 3-3 where dimensions have been included.

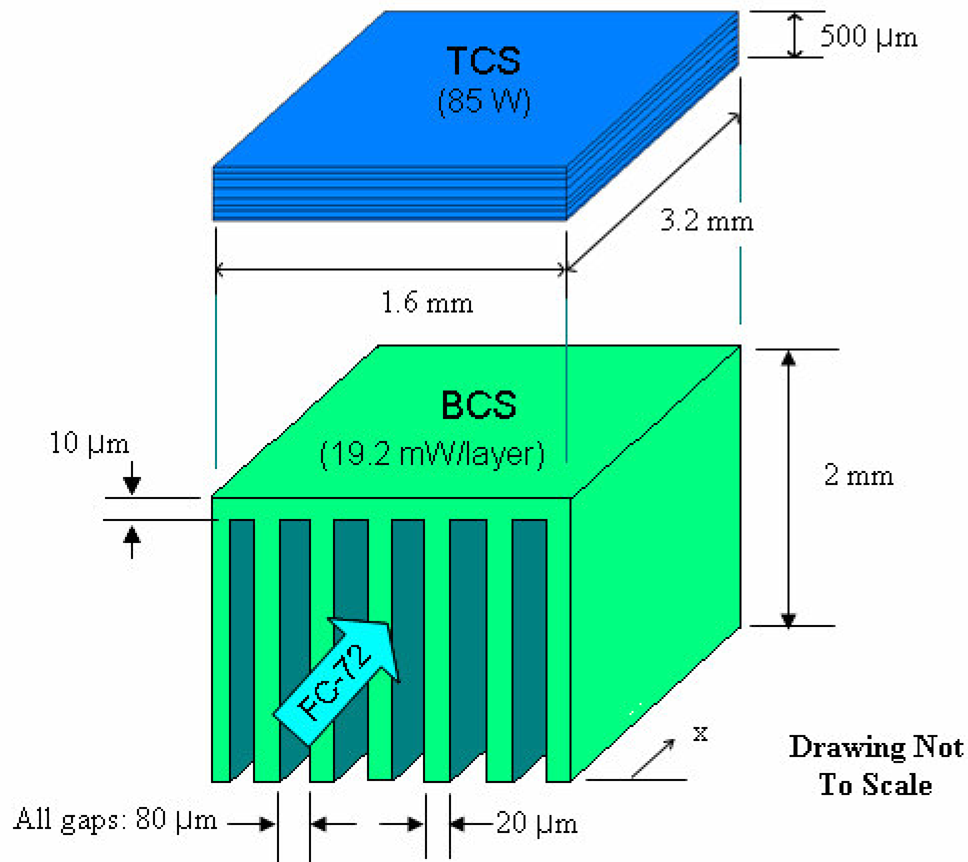


Figure 3-3: The hybrid chip stack to be analyzed in this chapter

As indicated, the system has a small footprint of 1.6 mm x 3.2 mm and the TCS is 500 µm thick while the BCS is 2 mm tall. Both the top and bottom chip stacks are made predominately of silicon, which is assumed to have a thermal conductivity of 124 W/m-K [42]. The silicon layers, or 'slices,' in the BCS are each 20 µm thick, with 80 µm gaps between slices. This thickness and spacing combination allows 16 vertically oriented silicon slices to be placed under the TCS (note that the scale is exaggerated in Figure 3-3 to show detail; not all slices are shown). The BCS dissipates a total of 0.307 W, which translates to an internal heat generation of 19.2 mW in each of the 16 silicon slices. In contrast, the TCS dissipates 85 W, a considerable amount given its size, yielding an

internal heat generation rate of 33.2 kW/cm^3 and imposing a heat flux of 1.66 kW/cm^2 on the BCS.

Throughout this chapter, single phase FC-72 is assumed to flow through the gaps in the BCS. FC-72 is a perfluorinated liquid which is chemically and electrically inert and is thus suitable for direct liquid cooling of electronics. The properties of FC-72 are taken at 25C throughout the chapter and are given by the following [43]:

- $\rho = 1718\ kg/m^3$
- $k = 0.05526\ W/m\text{-}K$
- $\mu = 6.011 \times 10^{-4}\ N{*}s/m^2$
- $c_p = 1196\ J/kg\text{-}K$
- $Pr = 13$
- $T_{sat} = 57\ ^{\circ}C$

The remainder of this chapter will analytically and numerically review the thermo-fluid conditions that arise from application of this liquid cooling method for different system parameters.

A Review of Analytical Basics

All of the numerical simulations that were run in CFX had single phase laminar flow characteristics. As such, it would be beneficial to review some important points from analytical treatments of heat transfer to laminar internal flows.

Consideration of an energy balance leads to the familiar equation:

$$Q = \dot{m} c_p \left(T_{m,o} - T_{m,i} \right) \tag{9}$$

Where Q is the total heat transfer rate of the conduit, m with an overdot is the mass flow rate, c_p is the specific heat capacity of the fluid at constant pressure, and $T_{m,i}$ and $T_{m,o}$ are

the mean inlet and outlet temperatures, respectively. This simple energy balance is true irrespective of the flow conditions and wall heating conditions.

Another familiar equation is the expression for local convective heat transfer to the fluid:

$$q_{w,x}'' = h(T_w - T_{m,x}) \tag{10}$$

where $q_{w,x}''$ is the local wall heat flux, h is the convective heat transfer coefficient, $T_{w,x}$ is the local wall temperature, and $T_{m,x}$ is the local mean temperature of the fluid. The non-dimensional heat transfer coefficient at the wall, or Nusselt number, is defined in the following way:

$$Nu = \frac{h D_h}{k} \tag{11}$$

where h is the heat transfer coefficient, D_h is the hydraulic diameter of the conduit and k is the thermal conductivity of the fluid. A discussion of a basic heat transfer analysis in tubes (in which equations 9-11 are developed) is available in all heat transfer textbooks and given specifically in [44].

Infinitely Wide Parallel Plates

The gaps in the BCS will serve as channels for fluid flow. The ratio of the channel width to the gap height is relatively large, so that analytical treatments of infinitely wide parallel plate channels are applicable to in the present analysis. Useful information on analyses of this type can be found in a textbook written by Kays and Crawford [46]. The authors report a method to determine the Nusselt number at different axial positions in a parallel plate channel heated by a constant wall heat flux on both sides of the channel, q_1'' and q_2'', where the subscript indicates the wall on which the flux is applied.

Solutions exist for walls heated symmetrically ($q_1 = q_2$) or asymmetrically ($q_1 \neq q_2$). The Nusselt number expression for wall 1 is given by

$$Nu_1 = \frac{Nu_{ii}}{1-(q_2"/q_1")\theta_i} \tag{12}$$

To obtain Nu_2 from equation 12, simply switch the subscript on each term from a 1 to a 2, or vice versa. The isolated wall Nusselt numbers, $Nu_{ii,}$ and the influence coefficients θ_i, in equation 12, and are tabulated in Table 3-1 for different values of the non-dimensional distance into the channel, x^+, defined as

$$x^+ = \frac{2(x/D_h)}{\mathrm{Re}\,\mathrm{Pr}} \tag{13}$$

where x is the axial distance into the channel, D_h is the hydraulic diameter, Re is the Reynolds number, and Pr is the Prandlt number. For flow between infinitely wide parallel plates, the hydraulic diameter can be shown to equal $D_h = 2a$ (where a is the spacing between the plates). Also, the well known definitions for the Reynolds and Prandlt numbers are the following:

$$\mathrm{Re} = \frac{\rho V D_h}{\mu} \qquad \mathrm{Pr} = \frac{c_p \mu}{k} \tag{14}$$

where V is the mean fluid velocity, and ρ, μ, and k are the density, dynamic viscosity, and thermal conductivity of the liquid, respectively.

Table 3-1: Nusslet numbers and influence coefficients of interest from [46]

x^+	Nu_{ii}	θ_i
0.0005	23.5	0.01175
0.005	11.2	0.0560
0.02	7.49	0.1491
0.10	5.55	0.327
0.25	5.39	0.346
∞	5.38	0.346

Thus, for a given fluid and a given axial distance into the channel, *x*, one can find the wall temperature by the following process:

1. Use *x*, the fluid properties, and the plate spacing to find x^+ in equation 13.
2. Knowing x^+, use the tables in [46] to find Nu_{ii} and θ_i, interpolating as necessary.
3. Determine Nu_1 or Nu_2 from equation 12 (knowing the heat fluxes).
4. Use the Nusselt number determined in previous step to determine the value of h for that wall from equation 11.
5. Use q_1'', q_2'', *x*, and the plate dimensions to determine Q, the total heat transferred to the fluid up to the point *x*.
6. Then, knowing Q, equation 9 can be used to find $T_{m,o}$ (in this case, $T_{m,o} = T_{m,x}$).
7. Using $T_{m,x}$ and the appropriate h and q_w'' (either q_1'' or q_2''), equation 10 can be used to find the wall temperature, T_w.

Developing Laminar Flow

Later, it will be shown that fully developed thermal and hydrodynamic flow is not achieved in the BCS under reasonable flow conditions; that is, for reasonable working

fluids and inlet velocities. Since developing conditions will dominate the flow, a review of the characteristics of developing profiles would be beneficial at this point.

The general axial development of temperature and velocity profiles for laminar flow is shown in Figure 3-4. The temperature and velocity are assumed to be uniform at the inlet. As axial distance into the channel increases, boundary layers develop on the channel walls (represented by the gray regions). The temperature and velocity of the fluid vary from the wall to the edge of the boundary layer, but the core of the flow maintains the uniform inlet values. Eventually, the boundary layers grow large enough to join in the center of the channel and eliminate this core of uniform inlet temperature or velocity. At this merging point, the profile becomes parabolic and conditions are said to be fully developed. In Figure 3-4, the point at which fully developed thermal and hydrodynamic conditions are reached is marked by $x_{fd,t}$ and $x_{fd,h}$, respectively.

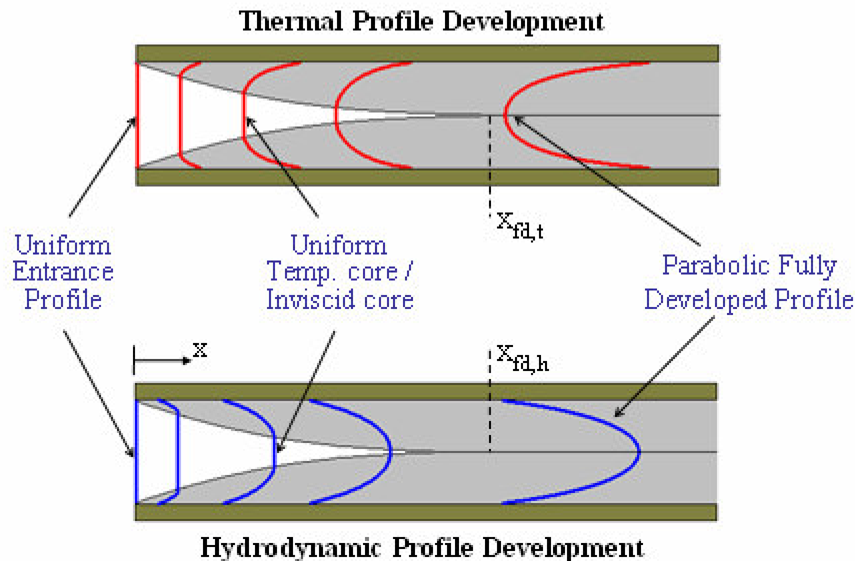


Figure 3-4: Developing laminar flow temperature and velocity profiles for isoflux channels

The values of $x_{fd,t}$ and $x_{fd,h}$ are dependent on the inlet velocity, channel geometry, and fluid properties. Most introductory heat transfer and fluid dynamics texts offer the following approximate expressions for the entry lengths [45]:

$$x_{fd,t} \approx \frac{1}{20} \mathrm{Re}\, D_h \,\mathrm{Pr} \tag{15}$$

$$x_{fd,h} \approx \frac{1}{20} \mathrm{Re}\, D_h \tag{16}$$

As indicated by equations 15 and 16, the difference in the rate of thermal and hydrodynamic profile development depends on the Prandlt number of the fluid. Velocity profiles develop more rapidly than thermal profiles for fluids with high Prandlt numbers (>>1). If the Prandlt number is approximately unity, as is the case for many gases including air, the hydrodynamic and thermal profiles will develop simultaneously. This is generally not the case for liquids and decidedly not the case for the high Pr number FC liquids. It is to be noted that in the developing flow region, higher fluid velocity results in thinner boundary layers and higher heat transfer coefficients.

Due to the difficulty in graphically determining the fully-developed state when investigating profiles generated from numerical models, use can be made of the changing ratio of the centerline velocity to the inlet (or average) velocity as the fully-developed profile is approached. Figure 3-5 displays this ratio for developing laminar flow in parallel plate channels. The vertical axis in Figure 3-5 is the ratio of the centerline velocity to the uniform inlet velocity, while the horizontal axis is a non-dimensional axial distance into the channel. As shown, the ratio of centerline velocity to inlet velocity approaches 1.5 as the flow becomes fully developed. This fact will be used in later discussions to help assess the validity of the profiles generated by ANSYS CFX.

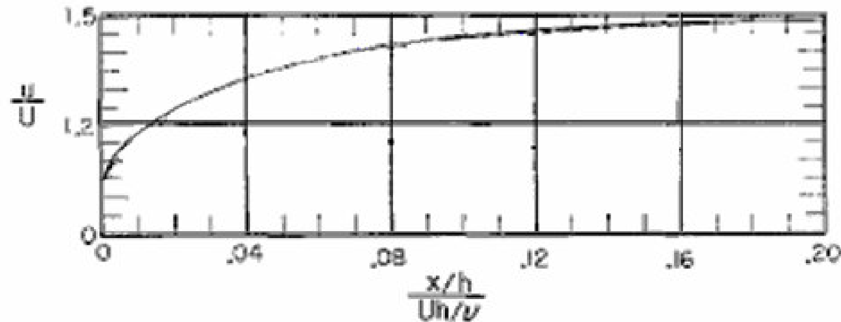


Figure 3-5: Centerline velocity development, laminar flow in a parallel plate channel [47]

A Two-Dimensional Approximation of Flow in the BCS

Although the complexity of the thermo-fluid conditions in the BCS make it desirable to perform a three-dimensional simulation of the velocity and temperature fields, the large aspect ratios of the BCS channels necessitate a very large number of elements and make such simulations difficult. Initial simulations were, therefore, performed with a two-dimensional configuration in which the fluid inlet is at the bottom of the BCS and the outlet is at the top (where the TCS sits). While this configuration departs from the actual flow distribution pattern, it does make it possible to study the salient parametric trends and obtain a first-order approximation of the BCS temperature distribution. The simplified 2D flow pattern also facilitates validation of the numerical results with the analytical results discussed in the previous section. Execution of the more resource intensive 3D simulations was limited to only a few cases, which will be presented in more detail later.

A schematic of the 2D model created in CFX is shown in Figure 3-6 (the model is a single gap of the BCS). The fluid inlet is assumed to be at the bottom of the BCS, and

the fluid outlet at the top. Each 20 μm thick silicon slice generates heat internally, and heat flux (resulting from heat generation in the TCS) is supplied to the tops of the silicon slices. The 80 μm gap between silicon slices allows 16 slices to fit under the footprint of the TCS. Referring to the dimensions in Figure 3-3, the area of a single silicon tip at the interface with the TCS can be calculated in the following way:

$$A = 20\ \mu\text{m} \times 3.2\ \text{mm} = 6.4 \times 10^{-8}\ \text{m}^2 \quad (17)$$

Also, the total heat coming down from the TCS is 85 W; thus, assuming 100% of the heat generated in the TCS is coming down to the BCS, the heat flux applied to the top of a single slice is given by

$$q'' = \frac{85\ \mathbf{W}}{16\ \mathbf{slices} \times \mathbf{A}} = 8.3 \times 10^{7}\ \ \mathbf{W/m^2} \quad (18)$$

In addition to the heat flux in equation 18, each slice of silicon generates heat internally. As reported earlier, a total of 0.0192 W is dissipated in each slice, while approximately 5.3 W is applied to each silicon slice by the TCS. Thus, the heat supplied by the TCS can be expected to play the dominant role in determining the temperature distribution in the silicon slices.

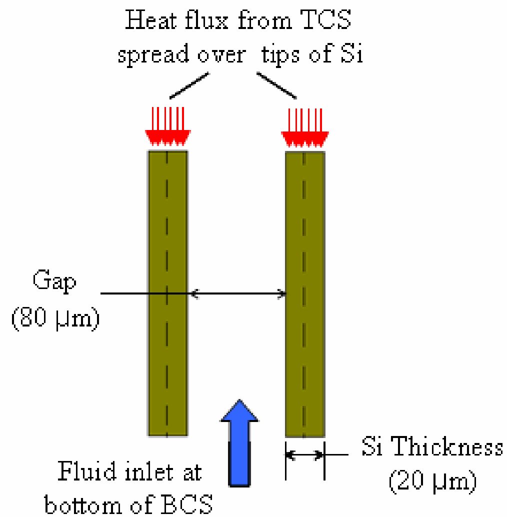


Figure 3-6: 2D flow configuration assumed in the BCS

The two dimensional geometry seen in Figure 3-6 was created in ANSYS CFX, a commercial computational fluid dynamics analysis package. The simulation process flow in CFX proceeds as follows: first, the model geometry is created using dedicated CAD modeling software called DesignModeler; then, the CAD geometry is imported into CFX-Mesh, a meshing program which provides a means to discretize the geometry with finite elements; the resultsing mesh is imported to CFX-Pre, a pre-processing program which allows the definition of material/fluid properties and boundary conditions; finally, the mesh information and boundary conditions are communicated to the CFX solver, which discretizes the Navier-Stokes equations over the computational grid in the fluid domain and accounts for conjugate effects to arrive at a temperature distribution in the solid domain. The ability to monitor convergence criteria and compute a system energy balance in the CFX solver was used to test different levels of model discretization for various system parameters and arrive at mesh-independent results. The resulting mesh and boundary conditions used in the 2D CFX simulations are shown in Figure 3-7 (the

model has been turned on its side in the figure). The model was meshed with predominately tetrahedral elements (58,970 elements); however some hexahedral "inflation" elements (10,240) were used near walls to better resolve the large temperature and velocity gradients that are expected in these regions. Due to symmetry, the silicon slices are one-half their usual thickness (10 μm) and have insulated boundary conditions on one side. The working fluid is assumed to be FC-72, and the inlet temperature is always set at a constant 25C (298K).

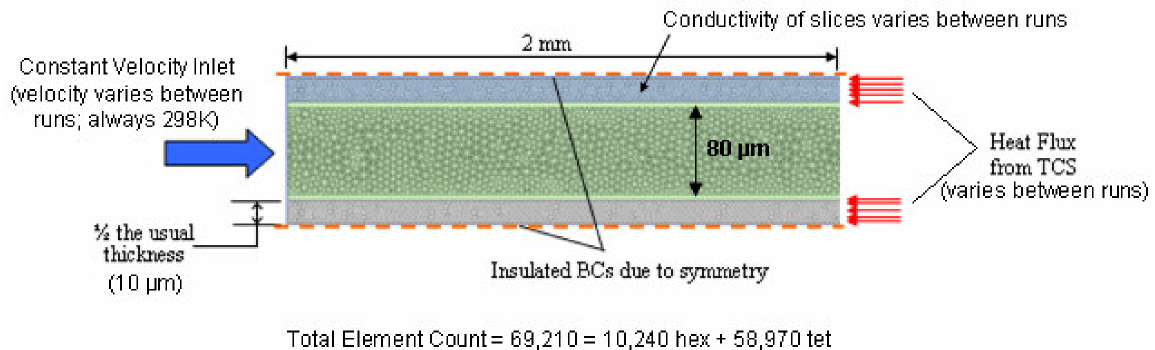


Figure 3-7: 2D model mesh with boundary conditions and details shown

Approximately 40 simulations were performed with this mesh for various scenarios where the fluid inlet velocity, the heat flux imposed by the TCS, and the conductivity of the BCS 'slices' was varied parametrically between runs. The computational platform used during these simulations was a Dell PC with a Windows based operating system, 2 gigabytes of random access memory, and a Pentium 4 processor running at 2.66 GHz. To ensure proper convergence and accurate results, the root mean square residual target was set to 10^{-7} on the iterative solver and the computed energy balance was required to be

within 1% accuracy. The results of these 40 parametric variations will be discussed shortly, but first validation of the CFX and the 2D mesh will be presented.

General CFX Validation

In the early stages of dealing with CFX, the model shown in Figure 3-8 was created and checked against analytical predictions. As shown, the model consisted of a single 150 x 10 x 50 μm fluid domain with uniform and constant heat flux (q'' = 2.88 x 10^6 W/m^2) applied to the areas of the bounding walls. Water was chosen as the working fluid with an inlet velocity of 1 m/s.

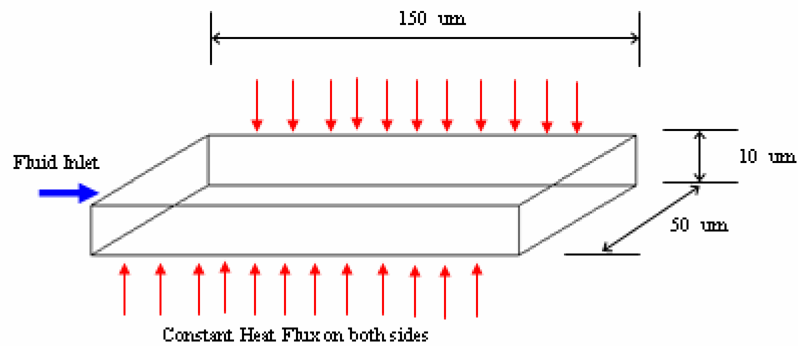


Figure 3-8: CFX verification model

The wall temperatures for this configuration can be determined analytically from equations 9-14 via the method laid out in the *Infinitely Wide Parallel Plates* section. Two cases were run in CFX. The heat loading was symmetric in the first case (q'' applied on both sides), while asymmetric heating (q'' applied on the bottom and zero flux on the top) was applied in the second case. The results of the CFX runs are compared to the analytically determined wall temperatures in Figure 3-9 and Figure 3-10. As shown, the analytical and CFX results match quite well (to within 0.5 %).

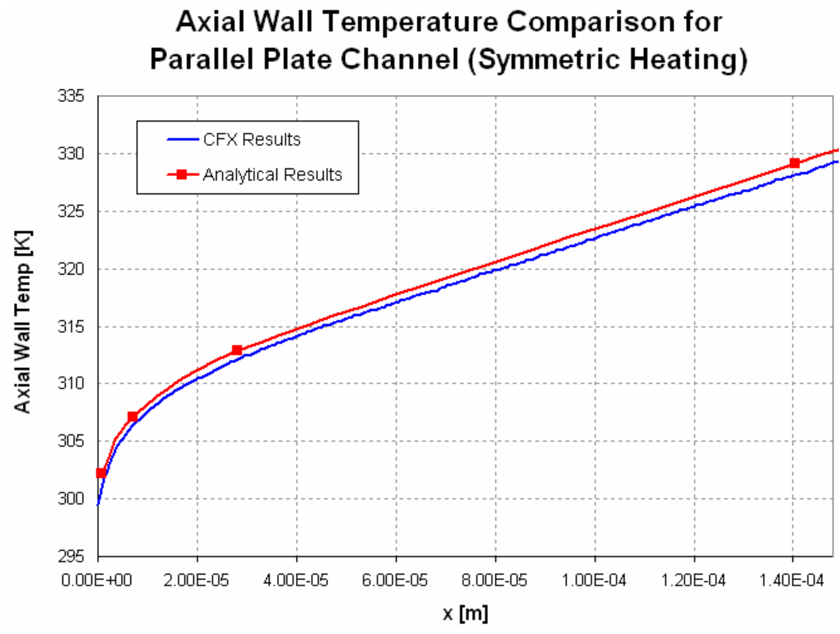


Figure 3-9: Symmetric Heating

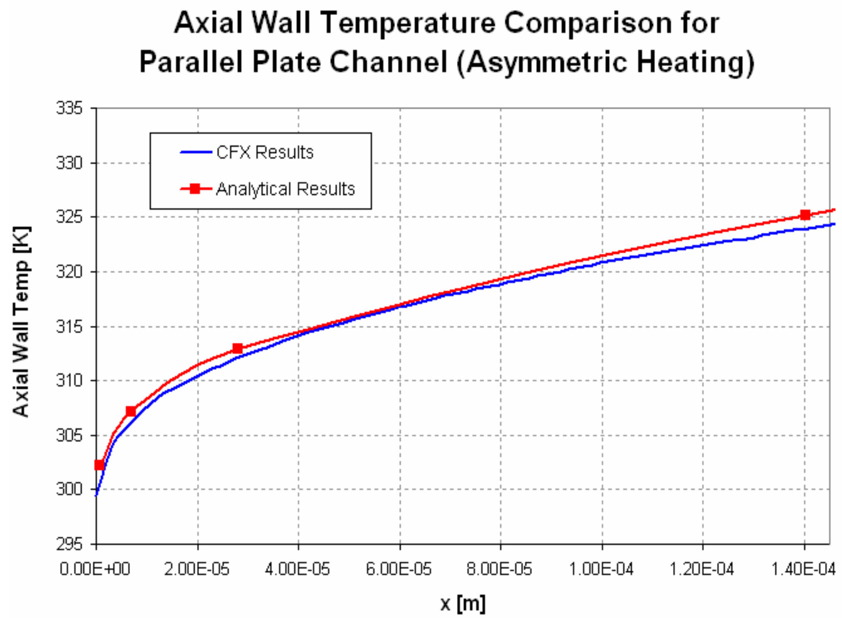


Figure 3-10: Asymmetric Heating

2D Thermal and Hydrodynamic Profile Validation

Before examining the results of the 2D simulations, it will be instructive to examine what type of flow conditions should be expected in the 2D model.

The thermal and hydrodynamic entry lengths are known from equations 15 and 16, respectively. Using the fluid properties of FC-72, assuming the gap is 80 µm wide, and taking a conservatively small inlet velocity of 1 m/s, we have the following from equations 14 and 16:

$$x_{fd,h} = \frac{1}{20}\frac{\rho \mathrm{V}(\mathrm{D_h})^2}{\mu} \approx 3.7\,mm \tag{19}$$

Since the thermal entry length is just the hydrodynamic entry length multiplied by the Prandlt number, and the Prandtl Number for FC-72 is approximately 13, we have:

$$x_{fd,t} = x_{fd,h}\,\mathrm{Pr} = 13x_{fd,h} \approx 48\,mm \tag{20}$$

Equation 19 clearly shows that even for the relatively small inlet velocity of 1 m/s, the hydrodynamic entry length is longer than the axial length of the 2D channel (2 mm). The thermal entry length is even longer, stretching well beyond the axial length of the 2D channel. Therefore, we expect the thermal and hydrodynamic profiles to be developing throughout the entire length of the channel for an inlet velocity of 1 m/s or greater. Furthermore, it will be shown later that fluid inlet velocities greater than 1 m/s are needed to effectively cool the BCS; thus, the profiles will be developing for all effective inlet velocities.

The relevant characteristics of developing hydrodynamic and thermal profiles were reviewed above in the *Developing Laminar Flow* section. Profiles extracted from

the 2D CFX simulation results will now be compared to the expected profile characteristics for validation purposes.

Visual inspection of the velocity profiles produced by CFX – for example the velocity profile shown in Figure 3-11a at an entry length of 1.5mm – reveals an approach to the parabolic, fully developed profile, though the centerline region is seen to still reflect the inlet condition. Values of the centerline velocity from multiple CFX runs (at different inlet velocities) were compared to the plot shown in Figure 3-5. The results of this comparison are shown graphically in Figure 3-11b (the blue diamonds are the results from CFX). As shown, the CFX results agree well with the analytical prediction of the centerline velocities, so the CFX velocity profiles can be assumed to be within engineering accuracy.

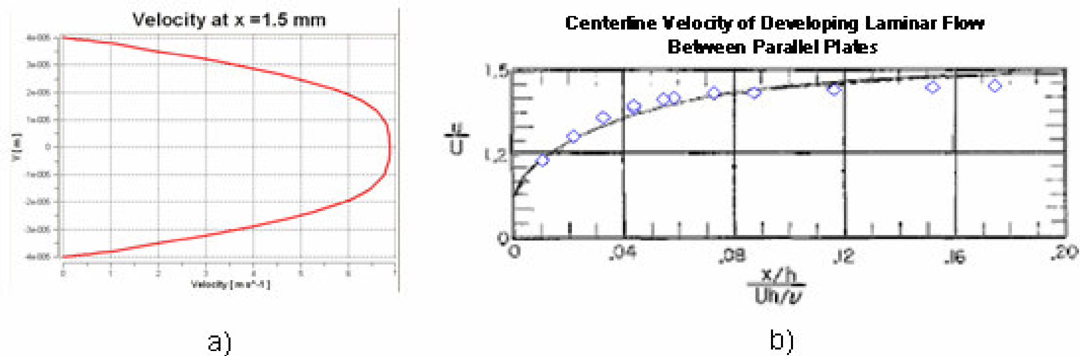


Figure 3-11: (a) Velocity profile in the 2D model at 1.5 mm for $V_{inlet} > 1$ m/s, (b) Comparison of centerline velocity to the expected values reported in [47]

The thermal profiles produced by CFX are somewhat more easily verified. The developing nature of the thermal profiles is easier to observe since flow within the small, 2 mm channel leaves us deep within the thermally developing range. A plot of the

thermal profiles from a typical 2D model simulation is shown in Figure 3-12. The plot lines in Figure 3-12 are color-coded; each color corresponds to the temperature profile at a different axial location (see the legend at the bottom of the figure). These profiles qualitatively exhibit the characteristics that are displayed in Figure 3-4. Near the channel inlet, the flow is mostly a constant temperature; then, at greater entry lengths, the size of the constant temperature core shrinks as the boundary layer grows. Clearly the CFX-computed temperature field is still far removed from being thermally fully developed even at the outlet (the red profile in Figure 3-12), reflecting the expectation that nearly a 50mm flow length would be required to achieve the thermally fully-developed state.

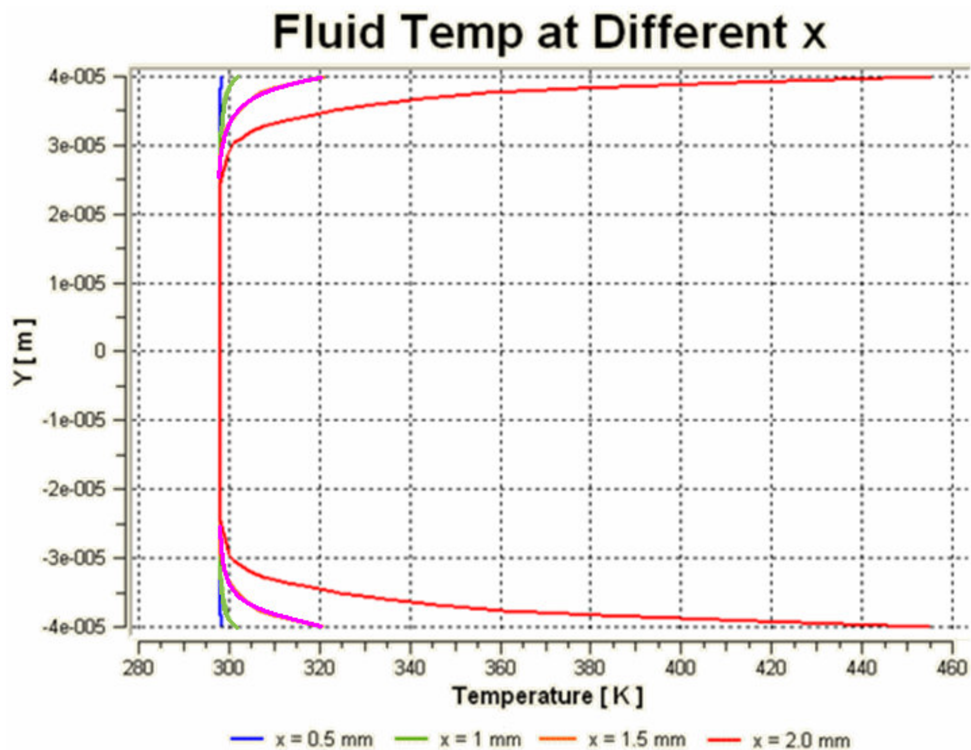


Figure 3-12: Temperature profiles at different axial positions in the channel

A Note About Turbulence

The hydrodynamic entry length for turbulent channel flow is given by equation 21 [48]. For example, consider FC-72 flowing into the 80 μm gap between silicon slices (D_h = 160 μm) with an inlet velocity of 20 m/s. Equation 21 reveals that the turbulent hydrodynamic entry length is ~3.2 mm under these conditions.

$$x_{fd,h,turb} = 4.4 \mathrm{Re}^{1/6} \mathbf{D_h} \quad (21)$$

Since 3.2 mm is greater than the 2 mm channel length, the turbulent profile does not achieve a fully developed state in the channel. Further, since the boundary layers do not merge in the channel, one can consider each side of the channel as an isolated flat plate with a free stream of liquid flowing over it. This familiar situation is shown in Figure 3-13. As labeled, x_c is the critical distance at which the laminar-to-turbulent transition begins. For flow over a flat plate this transition occurs roughly around $Re_{x,cr} = 5 \times 10^5$, where the length scale on the Reynolds number is the distance from the leading edge, x [45].

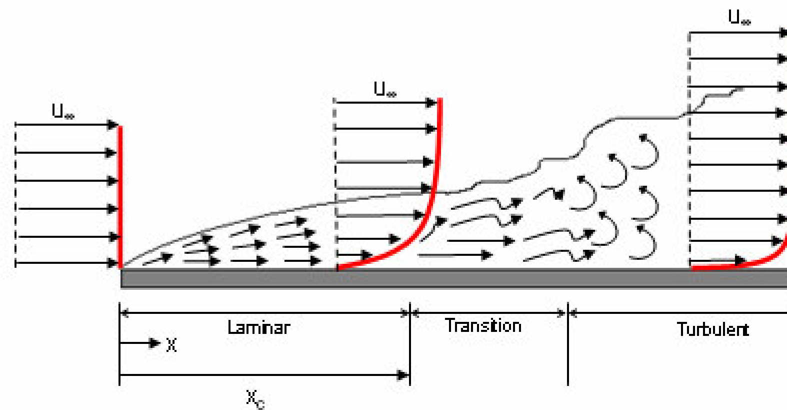


Figure 3-13: Boundary Layer Development over a Flat Plate

Several CFX simulations were run on the 2D configuration for different FC-72 inlet velocities. One can compute the value of Rex at a given x location by taking the centerline velocity at x as the equivalent free stream velocity. The results of such a calculation for inlet velocities of 10, 15 and 20 m/s are shown in Table 3-2 for different x locations. As shown, the critical laminar-to-turbulent value of $Re_{x,cr} = 5 \times 10^5$ is not met or exceeded anywhere in the channel for the velocities considered. Thus, one would not expect a turbulent boundary layer to be fully realized in the channel, even at the highest velocity considered.

Table 3-2: Flat plate Reynolds Numbers for "isolated" plates subject to the channel centerline velocity

	Re_x at Different x [mm] for Turbulent Simulations				
Inlet Velocity [m/s]	x = 0.25	x = 0.5	x = 1.0	x = 1.5	x = 2.0
10	7.28E+03	1.48E+04	3.05E+04	4.65E+04	6.23E+04
15	1.09E+04	2.23E+04	4.63E+04	7.10E+04	9.53E+04
20	1.46E+04	2.98E+04	6.21E+04	9.54E+04	1.28E+05

Recall that an inlet velocity of 20 m/s was used in the calculation of $x_{fd,h,turb}$ to justify the isolated plate analysis. However, the isolated plate analysis is still valid for the 10 and 15 m/s inlet velocities considered in Table 3-2. The smaller inlet velocities do not significantly impact the value of $x_{fd,h,turb}$ in equation 21 due to the weak dependence of $x_{fd,h,turb}$ on Reynolds Number ($Re^{1/6}$). It can easily be shown that $x_{fd,h,turb}$ = 2.8 mm for an inlet velocity of 10 m/s, which is still well beyond the end of the 2D channel.

The parametric range of inlet velocity was capped at 20 m/s because a pressure drop of 1.8 atm was necessary to sustain such a flow (pressure drops in the range of 1-2

atm are considered acceptable for this application). It is conceivable that one could accept pressure drops slightly higher than 2 atm and, thus, be able to afford a higher inlet velocity. As shown in Table 3-2, Re_x is approaching $Re_{x,cr}$ at the end of the channel for V_{in} = 20 m/s. Therefore, if pressure drops greater than 2 atm are attainable, then inlet velocities greater than 20 m/s could be sustained and the boundary layer might transition to turbulence near the end of the channel. However, the vast majority of the channel boundary layer would still be in the laminar regime and one would expect a laminar numerical model to approximate the thermal and flow fields well. Since laminar conditions are encountered for most reasonable inlet velocities and laminar analyses lead to more conservative results, only laminar analyses will be carried out and discussed in this chapter.

Results of 2D Model Simulations

The following sub-sections will show the results of the 2D CFX simulations for varying model parameters. Reliability concerns typically impose a nominal chip temperature limit of 90-110C for microelectronic applications; therefore, model parameters that yield maximum silicon temperatures in this range will be sought in each sub-section.

Percentage of TCS Heat Applied

The amount of heat applied to the 2D model by the flux from the TCS was varied in order to determine the effect on the surface temperature of the silicon. Recall that 5.3 W is supplied to the two half-slices represented in the 2D model when 100% of the TCS heat is applied as a flux. Also recall that the heat generated in each silicon slice is only 0.0192

W, meaning the TCS heat flux is dominant. One can understand that if a large portion of the TCS heat were intercepted before entering the BCS, then cooling the BCS by internal flow would be made easier. The simulations were, therefore, run for varying TCS heat loading, representing the possibility that a certain percentage of the TCS heat could be removed (perhaps by external cooling) before reaching the BCS.

Simulations were run with a 5.029 m/s inlet velocity of FC-72 and for 30 % – 100 % TCS heat flux (where at 100 % the TCS heat flux is given by equation 18). The maximum silicon slice temperature is seen in Figure 10 to reach 200C (for an inlet temperature of 25C) for the 100% TCS heat load and to decrease linearly to 75C for a 30% TCS load. Changes in the fluid inlet temperature can be expected to translate directly into changes in the maximum silicon temperature.

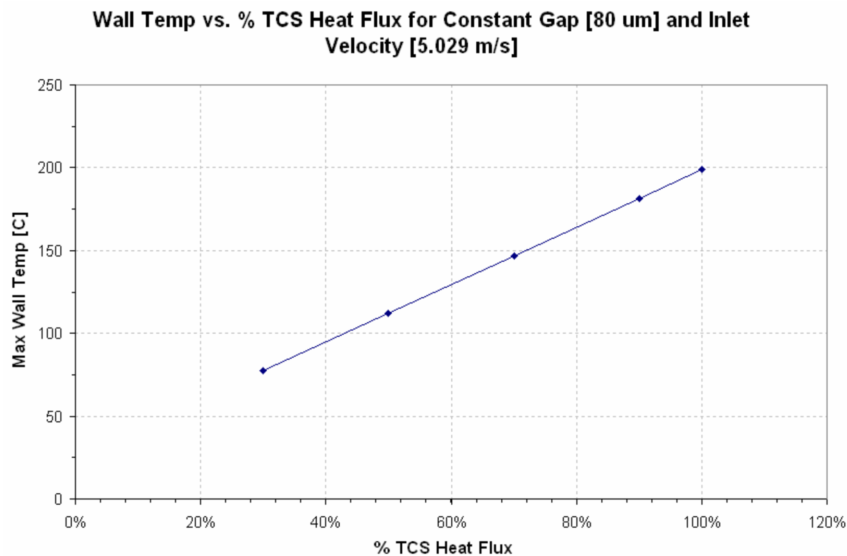


Figure 3-14: Effect of TCS Heat Load on Maximum Silicon Temperature

Inlet Velocity

As discussed above, the thermal profile is developing throughout the entire 2D channel. Thus, one would expect that increasing the fluid velocity would yield a considerable decrease in the wall temperature of the silicon. To investigate this, the inlet velocity was varied from 0.3 m/s to 20 m/s for a constant TCS heat load (100%) and fixed gap height (80μm). The results, along with the pressure drop at each velocity, are shown in Figure 3-15.

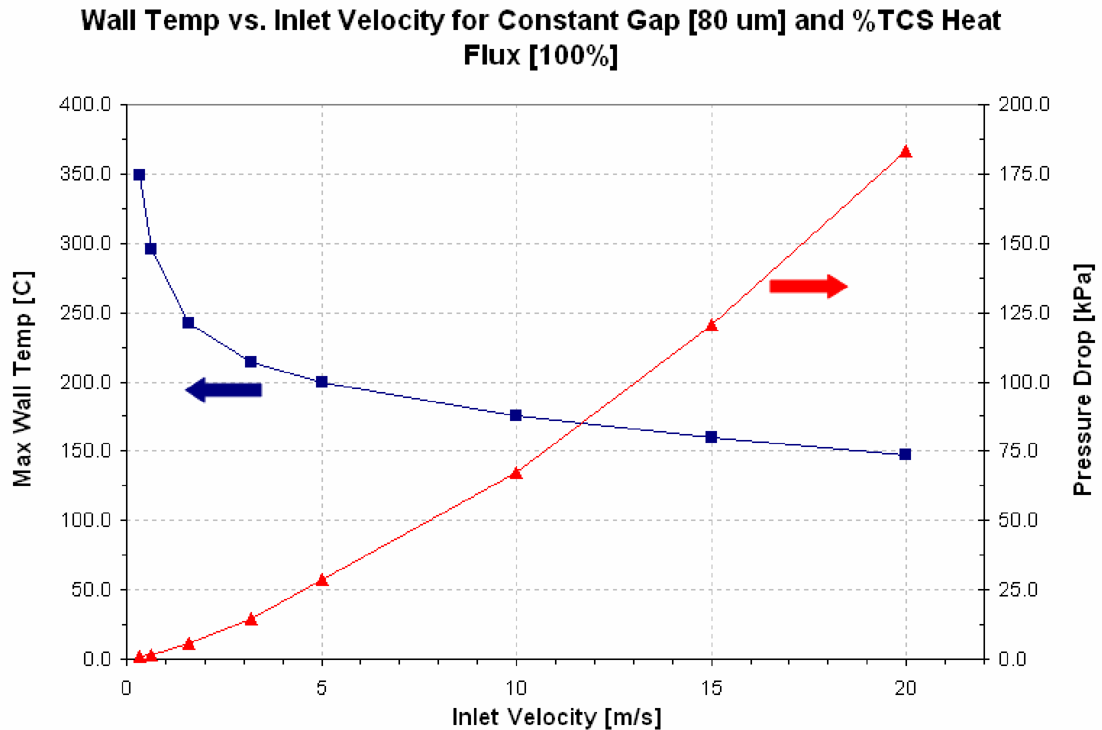


Figure 3-15: Effect of Inlet Velocity on Maximum Silicon Temperature

As shown, increasing the velocity over this range lowers the wall temperature by about 62% (this is calculated as a percentage of the temperature rise above the 25 C inlet).

However, maximum wall temperatures are still unacceptably high (i.e. >110C), even for the highest inlet velocity considered (T_{max} = 150C at V_{in} = 20 m/s). These simulations were run with the assumption that 100 % of the TCS heat is applied to the BCS. If some of the BCS heat were intercepted by external cooling methods and/or if the inlet temperature were lowered, the wall temperature curve in Figure 3-15 would shift downward (due to the linear relationship described in the previous sub-section) and reduce the maximum wall temperature to more acceptable levels.

It is noteworthy that although the maximum silicon temperature is 200C for an inlet velocity of ~5 m/s, the pressure drop is only about one quarter of an atmosphere (~28kPa). Given that pressure drops of 1-2 atmospheres are probably acceptable for this application, it stands to reason that the velocity could be increased considerably before pressure drop becomes a serious issue. As noted earlier, an inlet velocity of 20 m/s yields a pressure drop around 1.8 atm. However, the four fold increase in velocity from 5 m/s to 20 m/s yields only a 50C reduction in the maximum wall temperature. At larger velocities, prohibitively large pressure drops and diminishing improvements in silicon temperatures can be expected.

Thermal Conductivity of BCS Slices

Due to the inherently low internal heat generation in the BCS slices and the dominance of heat conducting into the BCS from the upper chip stack, the silicon slices in the BCS essentially serve the same role as fins of a heat sink. In this analogy, the TCS can be thought of as the 'base' of the heat sink; heat is conducted from this 'base' into the slices – or 'fins' – which create more surface area for convective heat transfer. It is evident that the heat flux entering the fluid through the 'fin' walls will be highest near the TCS/BCS

interface and diminish along the length of the fin. This heat flux variation is shown in Figure 3-16 for silicon slices with k = 124 W/m-K and an inlet velocity of 5.029 m/s. It is clear that the majority of the heat enters the fluid near the TCS/BCS interface while almost no heat enters the fluid near the inlet at the bottom of the BCS. However, this heat flux variation is heavily dependent on the thermal conductivity of the slices or 'fins.' Higher thermal conductivity in the slices, achieved through replacement of the silicon with a higher conductivity material or coating the silicon with a much more conductive material (e.g. diamond), would yield a more uniform heat flux variation and enhance thermal performance. The results of varying the effective wall conductivity from that of silicon (124 W/m-K) to that of high thermal conductivity silicon-diamond composite (~1000 W/m-K) are shown in Figure 3-17.

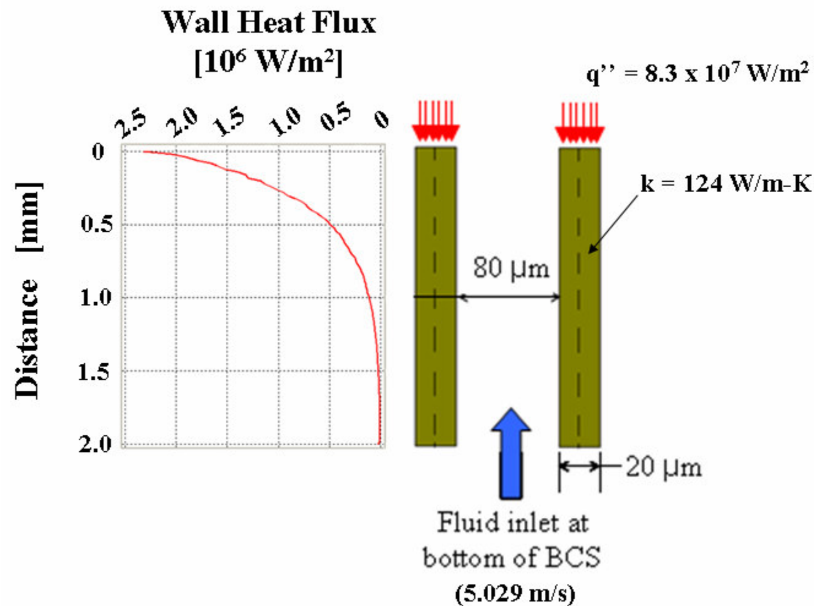


Figure 3-16: Heat flux variation along the length of the 2D channel at the interface of the silicon and cooling fluid (FC-72)

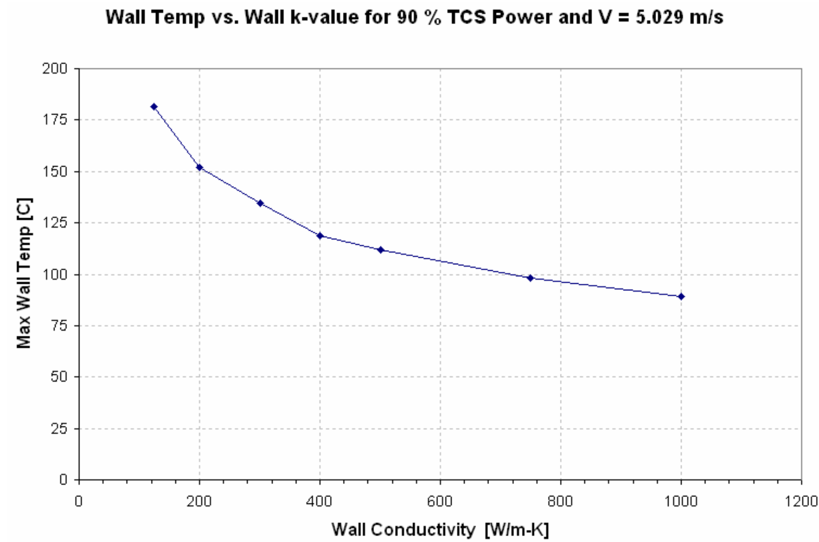
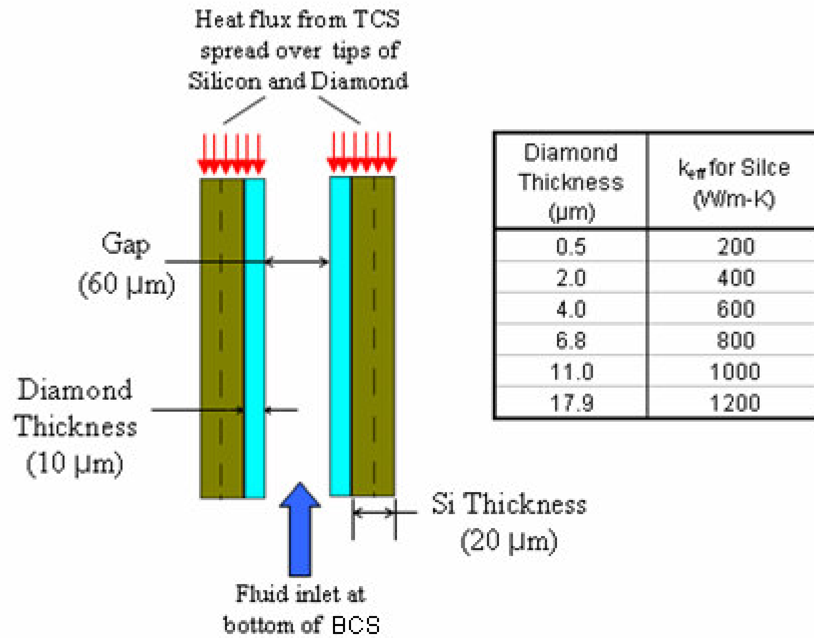


Figure 3-17: Wall conductivity varied

The maximum wall temperatures start to approach the target levels (<110C) as the wall conductivity approaches 600 W/m-K. These results indicate that the effectiveness of this internal cooling technique is heavily dependent on the spreading resistance in the slices. If a diamond spreader could be implemented as shown in Figure 3-18, then temperatures in the silicon would certainly be reduced. The effective thermal conductivity of the silicon/diamond composite slices shown in Figure 3-18 is reported in the inlayed table for different diamond thicknesses. These effective thermal conductivities were determined using area ratios and assuming that the thermal conductivity of the silicon and diamond are 124 W/m-K and 1800 W/m-K, respectively. One can use the table in Figure 3-18 along with the plot in Figure 3-17 to determine the thickness of diamond needed to achieve some prescribed maximum wall temperature.



Diamond Thickness (μm)	k_{eff} for Silce (W/m-K)
0.5	200
2.0	400
4.0	600
6.8	800
11.0	1000
17.9	1200

Figure 3-18: Hypothetical configuration to increase spreading

Summary

Superposition of the individual parameter variations presented in the preceding sub-sections reveals the conditions for which the target silicon operating temperatures (90-110C) can be met. Increasing the inlet velocity is a beneficial, but this approach alone can not sufficiently cool the BCS under the stated conditions since diminishing returns are seen for escalating inlet velocities and pressure drop can become an issue. For an inlet velocity of ~5 m/s, increasing the thermal conductivity of the BCS slices beyond 600 W/m-K reduces device temperatures to target levels. Alternatively, target temperatures can be achieved at a 5 m/s inlet velocity by intercepting ~50% of the TCS heat by alternate cooling methods. In a combined approach, a lower enhanced BCS slice conductivity of 250 W/m-K could be afforded if only 25% of the TCS heat was removed by other means. If it is not possible to intercept some of the TCS heat through other

cooling methods, then a combination of high inlet velocity and enhanced spreading in the BCS might be used. If a pressure drop of up to 2 atm can be afforded in the BCS, then an inlet velocity of ~20 m/s could be used in conjunction with an effective BCS slice conductivity of ~275 W/m-K to reduce the maximum BCS temperatures lower than 110C.

It should again be noted that the above results were developed under the simplified 2D flow approximation. While these results indicate the salient parametric sensitivities, a true three dimensional model is necessary to most accurately capture the performance of the proposed liquid cooling approach. Thus, a 3D model was developed and will be discussed in the following section.

A Three-Dimensional Model of Flow in the BCS

The 3D model is representative of a more practical flow configuration in the BCS in which the flow enters from one of the lateral sides of the BCS and exists from the opposite lateral side. A schematic of the 3D model is shown in Figure 3-19. All of the dimensions and materials are the same as the 2D configuration shown in Figure 3-6. Some key differences are that length of the channel in the flow direction is now 3.2 mm instead of 2 mm. Also, the manner in which the heat flux is applied is different. The 2D flow simplification required that the fluid exit from the "top" of the BCS. As a result, a heat flux boundary condition could not be applied to the entire top of the 2D model; rather, the flux was artificially applied only to the tips of the silicon (see Figure 3-6). In the more realistic 3D model, a uniform heat flux is applied to a silicon 'base.' Heat conducts to the fluid stream either directly through the base or by flowing down the silicon slice and into the fluid. Since the heat flux is applied over the silicon base in the

3D case—instead of just the silicon tips as in the 2D case—the applied heat flux must be altered. Altering equations 17 and 18 to account for the base area yields equations 22 and 23:

$$\mathbf{A = 100 \times 10^{-6}\,m \times 3.2 \times 10^{-3}\,m = 3.2 \times 10^{-7}\,m^2} \tag{22}$$

$$q'' = \frac{85\,W}{16\,slices \times \mathbf{A}} = 1.66 \times 10^{7}\,\frac{W}{m^2} \tag{23}$$

The heat flux in equation 23 assumes that 100 % of the heat generated in the TCS is entering the BCS. Comparing equation 18 to equation 23 reveals that the incident heat flux has been reduced by ~80 % due to the added area, but at 3.2kW/cm^2 is still approximately an order of magnitude above the heat fluxes encountered in state-of-the-art electronic cooling systems.

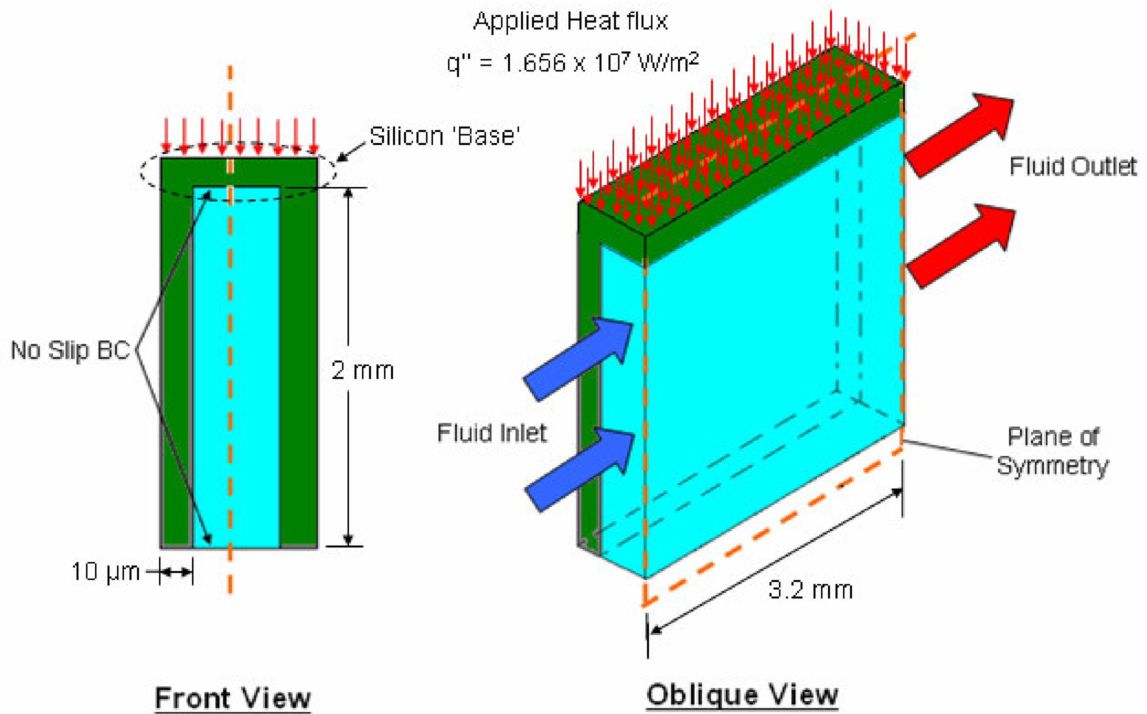


Figure 3-19: Schematic of the 3D model configuration

Results of 3D Model Simulations

The same computational platform and CFX simulation process flow used to develop the mesh for the 2D case was also used to create the mesh for the 3D flow configuration. The mesh in the fluid domain in the 3D simulations consists of 399,300 elements, about 85% of which are hexahedral elements along the fluid/solid interfaces (342,200 hexahedral elements). The mesh in the solid domain is comprised completely of tetrahedral elements (955,500), bringing the total number of elements in the 3D model to over one million (1,354,800). Due to the size and complexity of the 3D model, complete convergence of the numerical analysis takes over 8.5 hours. This considerable runtime prohibits the extensive parametric exploration conducted for the 2D case. However, the

results of the 2D model indicate that increasing inlet velocity and enhancing the thermal conductivity of the BCS slices can significantly reduce temperatures in the BCS. Therefore, the effectiveness of these two approaches will be examined for the 3D flow configuration, albeit with a more limited scope.

Inlet Velocity

The results of the 2D simulations revealed that inlet velocities beyond 5 m/s were needed to yield BCS temperatures approaching acceptable limits. Therefore, the 3D simulation was only run for inlet velocities greater than 5 m/s. Three different FC-72 inlet velocities – 5, 15, and 20 m/s – were explored with the laminar fluid flow model in CFX.

A plot of the typical temperature contours (shown in Kelvins) of the silicon at the fluid/silicon interface is shown in Figure 3-20 for an inlet velocity of 20 m/s. This temperature variation qualitatively makes sense as the hottest part of the wall is near the outlet at the TCS interface where the heat is being supplied. Similarly, the temperature is lowest at the bottom of the slice, far away from the applied heat flux and toward the inlet. It is seen that the pure silicon thermal conductivity of 124 W/m-K yields sharp temperature gradients in the upper portion of the slice and a relatively uniform temperature along the bottom half of the slice.

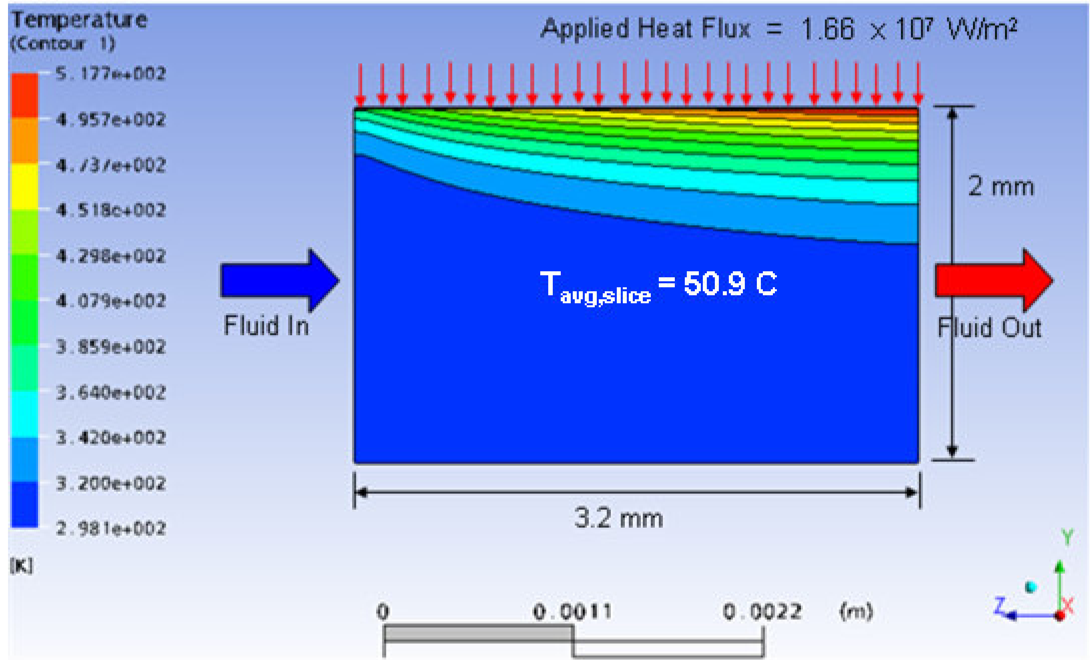


Figure 3-20: Temperature contours in the silicon with V_{in} = 20 m/s, $k_{silicon}$ = 124 W/m-K (the figure shows a side view of one slice at the silicon/fluid interface)

Table 3-3: 3D simulations results (laminar flow)

Inlet Velocity [m/s]	Max Si Temp [C]	Pressure Drop [kPa]
5	334	33.9
15	271	159.4
20	256	241.8

The maximum silicon temperature and pressure drop from the 3D simulations are displayed in Table 3-3. The results for the 3D simulation with a 5 m/s inlet velocity can be compared to the data point at 5 m/s in Figure 3-15 which pertains to the 2D simulation. Two important differences between the 2D and 3D model can be determined from comparison of the two results with V_{in} = 5 m/s:

1. The 2D model clearly underestimates the maximum silicon temperature. This is because the flow configuration in the 3D model requires that a portion of the fluid is exposed to the highest applied heat flux along the entire length of the channel. In the 2D model, the direction of fluid flow and the direction of heat flow oppose one another so that cold fluid is continually washing over the region of the slice where heat is applied. The maximum temperature in the silicon is under-predicted by 134C, or 40%, in the 2D case because of this unrealistic flow pattern. Similarly, the 2D model under-predicts the average temperature at the interface of the silicon and FC-72 by 35% ($T_{avg,3D}$ = 72.6C and $T_{avg,2D}$ = 47.8C).
2. The 2D model also under-predicts the pressure drop because of the shortened channel length. In the more realistically oriented 3D flow pattern, the channel length is 1.6 times that of the 2D model. Thus, the pressure drop across the channel is somewhat larger in the 3D simulation. Note that earlier it was stated that a 20 m/s inlet velocity resulted in at 1.8 atm pressure drop for the 2D simulation. However, Table 3-3 reveals that the 3D simulation indicates a pressure drop of 2.4 atm. This is an important difference, as the 2D simulation might lead one to believe that such an inlet velocity is achievable, while the 3D simulation might lead one to reconsider whether 20 m/s can be attained.

The average temperature of the slices in the BCS is found to be 72.6C for the case with an inlet velocity of ~5 m/s. This average temperature is relatively close to the 25C fluid inlet temperature and is small in comparison to the maximum BCS temperature of 334C. The large difference between maximum and average BCS temperatures indicates

the need for greater thermal spreading in the BCS. Implementation of a high conductivity spreader on the walls of the 3D channel – as indicated in Figure 3-18 for the 2D case – will act to soften the sharp thermal gradients seen in Figure 3-20 and reduce the maximum BCS temperatures to more acceptable levels. The following section will serve to explore the effectiveness of this approach for the 3D model.

Thermal Conductivity of BCS Slices

Parametric exploration with the two-dimensional model, reported in previous sections of this chapter, revealed that increasing the effective thermal conductivity of the 'slices' in the BCS can significantly improve thermal spreading in the slices and thus reduce the maximum temperatures in the BCS. Higher effective thermal conductivity in the slices can be realized through the use of a more highly conductive material than silicon or, more likely, through the implementation of a separate high conductivity spreader. For example, a diamond spreader might be used in an analogous manner to that depicted in Figure 3-18 in order to better facilitate spreading in the BCS. The thermal conductivity of the slices was varied from 124-1200 W/m-K in several CFX simulations to represent the use of a more highly conductive material or the presence of a spreader in the BCS. The temperature profile on the wall of the slice (*i.e.* at the interface of the FC-72 and the slice) is shown in Figure 3-21 for a representative case where the effective thermal conductivity of the slice is assumed to be 600 W/m-K and the inlet velocity is 20 m/s. Comparison of this temperature profile with that shown in Figure 3-20 for a silicon slice (k = 124 W/m-K) reveals that the increasing the effective conductivity of the slice to 600 W/m-K raises the average temperature by 2.4C, but lowers the maximum temperature by 117C and greatly softens the temperature gradient in the upper half of the slice.

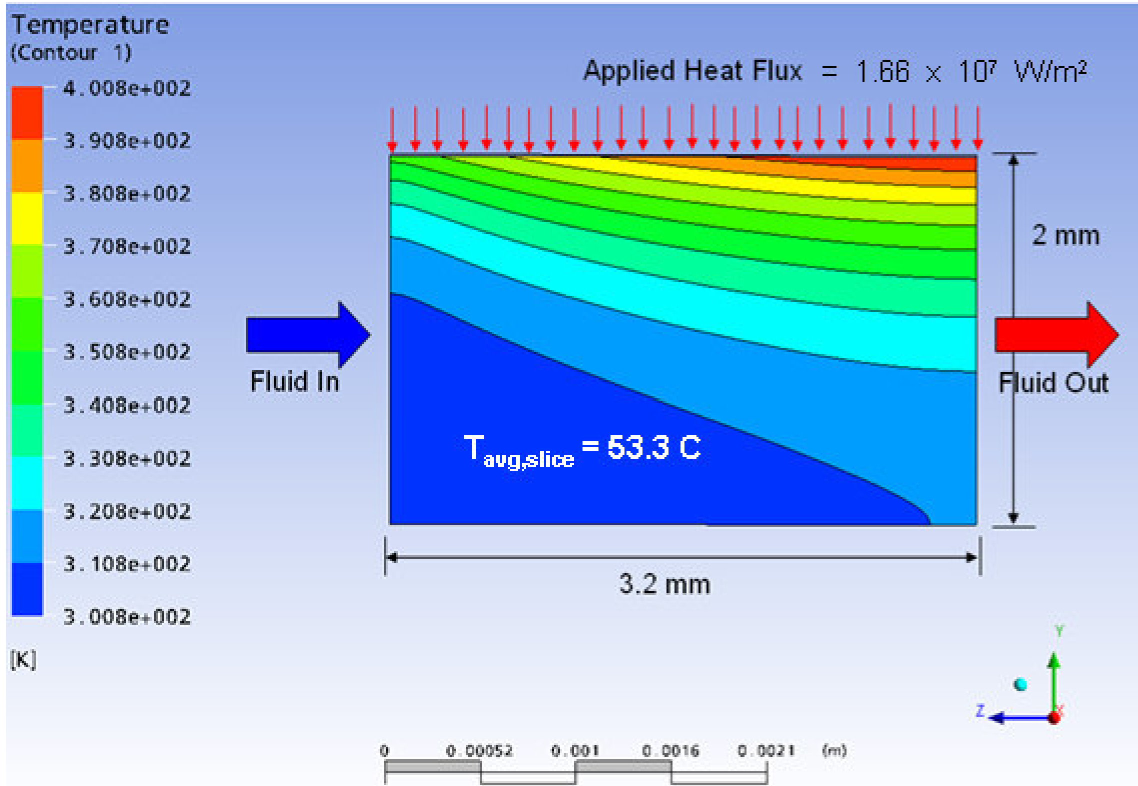


Figure 3-21: Temperature contours in slice with V_{in} = 20 m/s, k_{eff} = 600 W/m-K, (the figure shows a side view of one slice at the silicon/fluid interface)

The results of the parametric variation of effective slice conductivity from 124 W/m-K to 1200 W/m-K are shown in Figure 3-22 for high inlet velocity of 20 m/s. As expected, large reductions in the BCS temperature are seen for increasing effective wall conductivities. It is seen that effective conductivities greater than 800 W/m-K are needed to reduce the maximum device temperatures below 110C. The thickness of diamond needed to achieve this 800 W/m-K conductivity is 6.8 μm as indicated in Figure 3-18.

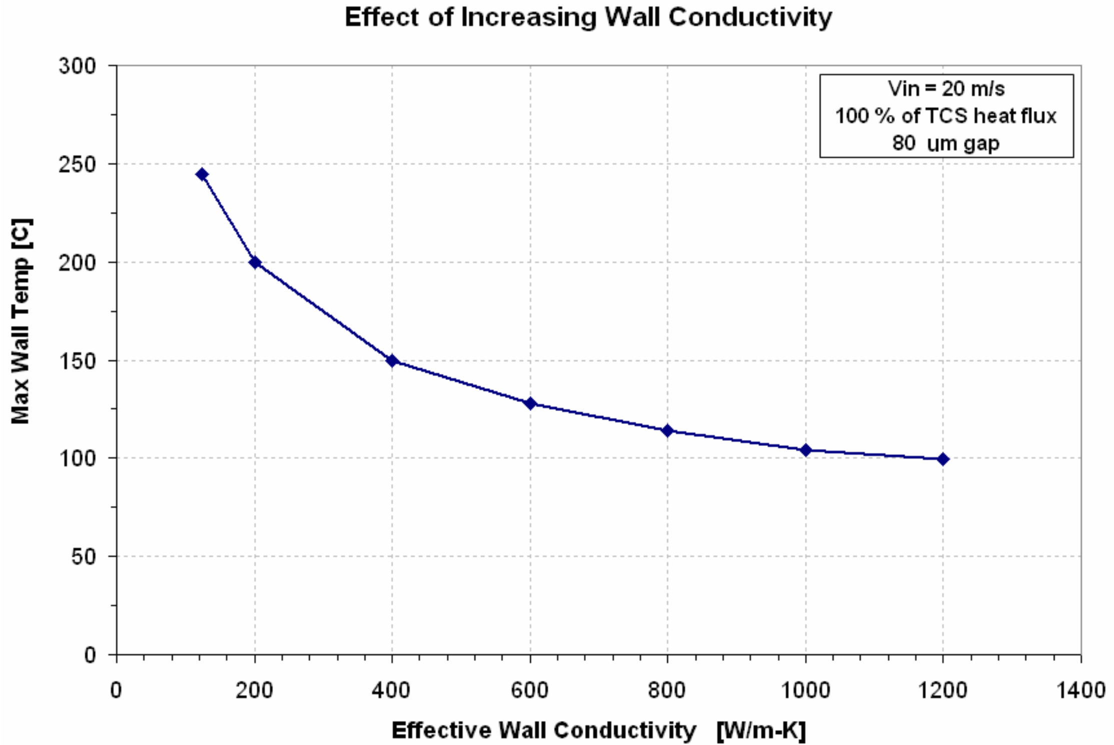


Figure 3-22: Varying the effective conductivity of the BCS slices in the 3D model

Summary

The resource intensive 3D simulations were only run for the two most promising approaches for reducing BCS temperatures found during parametric exploration of the 2D model: 1) increasing inlet velocity and 2) enhancing spreading in the slices. It was found that increasing the inlet velocity to 20 m/s by itself was insufficient to cool the BCS. For an inlet velocity of 20 m/s, the peak BCS temperature was found to be ~160C above nominal acceptable maximum electronic device temperatures (90C-110C), indicating that inlet velocities well beyond 20 m/s would be necessary to cool the BCS if no spreading enhancement was used. In order to reduce temperatures to target levels, a combined approach with a high inlet velocity and large effective thermal conductivity in the BCS slices must be used. With a 20 m/s inlet of FC-72, the effective thermal

conductivity of the slices must be 800 W/m-K or greater to reduce the maximum BCS temperature below 110C, with an effective conductivity of 1200 W/m-K yields a maximum temperature of 100C.

Conclusion

In this chapter, application of direct single phase internal liquid cooling with the dielectric liquid, FC-72, was explored for a novel hybrid 3D chip stack through numerical simulations with ANSYS CFX. The hybrid 3D chip stack is comprised of upper and lower portions which are respectively referred to as the top chip stack (TCS) and bottom chip stack (BCS). In the proposed approach, cooling is achieved by flushing FC-72 through channel-like openings in the BCS. Relevant analytical treatments of single phase channel flow were reviewed in beginning of this chapter, and these analytical methods were later employed to validate numerical models and establish mesh independent simulation results.

A simplified two-dimensional model of the thermo-fluid characteristics in the BCS was developed in ANSYS CFX in order to quickly explore parametric sensitivities and validate CFX results with respect to hydrodynamic and thermal profile development. The small dimensions of the flow path in the BCS result in developing flow conditions throughout the channels. Furthermore, the small channel dimensions delay the onset of turbulence in the channel until relatively high inlet velocities of >20 m/s are introduced. Therefore, all simulations were conducted assuming developing laminar flow conditions in the BCS. Over 40 numerical simulations were run with the 2D model for various combinations of applied TCS heat flux, fluid inlet velocity, and thermal conductivity of

the BCS slices. These simulations revealed that increasing the latter two parameters provided the most promising results for reducing BCS device temperatures. It was found that a combined approach of increasing inlet velocity to 20 m/s and enhancing the thermal conductivity of the BCS slices beyond ~275 W/m-K reduces maximum temperatures into the nominal target range of 90C-100C for microelectronics applications. The enhanced BCS conductivity might be achieved through the implementation of a high conductivity spreading material, such as diamond. The chart in Figure 3-18 can be used to infer that a diamond layer only 1.5 μm thick is needed to achieve an effective slice conductivity of ~275 W/m-K. However, these results are relevant only to the simplified 2D flow model.

In order to more accurately resolve thermofluid conditions in the BCS, a true three-dimensional numerical model was developed. The 3D model required significantly more computational resources than the 2D model, and the 8 ½ hour runtime for the 3D model prohibited detailed exploration of the parametric space. Therefore, with the results of the 2D simulations as a guide, the 3D model was run for about 10 different combinations of inlet velocities and BCS slice thermal conductivities. The results of the 3D model reveal that the simplified 2D model significantly under-predicts the severity of the thermal conditions in the BCS (maximum and average temperatures in the slices are under-predicted by 40% and 35%, respectively, with the 2D model). The more realistic 3D simulations show that large inlet velocities of 20 m/s and very highly conductive slices with $k_{eff} \approx 800$ W/m-K are needed to reduce maximum BCS temperatures lower than 110C, and $k_{eff} = 1200$ W/m-K is necessary to achieve $T_{max} = 100$C. It should be noted that all simulations conducted in this chapter are assumed to have an inlet

temperature of 25C, and reduction of this inlet temperature can be expected to translate directly into reduction of the maximum device temperatures.

An important conclusion from this chapter is the beneficial effect that introduction of a thermal spreader has on reducing the maximum BCS temperatures. It is seen that a highly conductive material attached to the side of the silicon BCS slices can significantly reduce temperature gradients in the silicon near the applied heat flux. One can imagine that the implementation of such a spreader would also serve to lessen the severity of non-uniform power dissipation within the slices themselves. Thus far, only spreading materials of isotropic thermal conductivity have been considered, but the following chapter will focus on the ability for orthotropic spreading materials to reduce the detrimental temperature rise associated with non-uniform power dissipation on a chip.

Chapter 4 : Anisotropic TIMs/Spreaders

Introduction

Technology scaling, fueled by consumer and industry demands for higher performance microprocessors, is leading to nanoscale chip feature sizes and significant increases in on-chip power dissipation. As indicated in Figure 4-1, The International Technology Roadmap for Semiconductors (ITRS) forecasts that high performance microprocessors will exhibit feature sizes as small as ~15 nm and transistor densities approaching ten billion transistors per square centimeter by the end of the next decade [49]. Also, the International Electronics Manufacturing Initiative (iNEMI) Technology Roadmap predicts that maximum on-chip power levels will exceed 350 W over the same period of time, presenting a significant challenge to thermal management engineers [50].

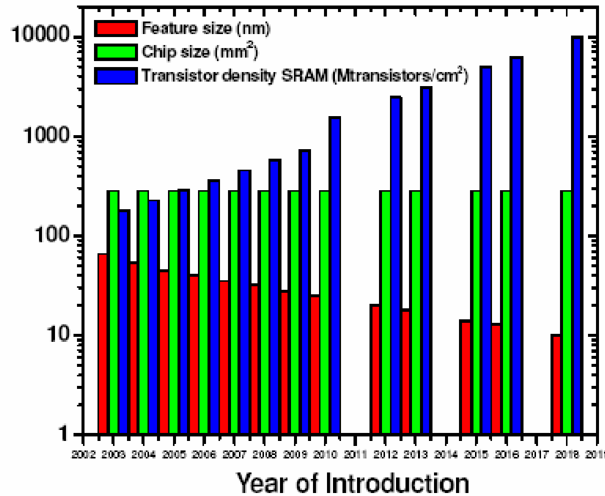


Figure 4-1: The 2005 ITRS forecast for high performance microprocessors [49]

Moreover, efforts to reduce delay times and increase performance are forcing highly dissipative elements of the chip into closer proximity, thus contributing to severe non-uniformity in on-chip power dissipation. The segregation of logic and memory cells and the desire to reduce on-chip communication delays have led the logic portions of the chip, which dissipate up to 90% of the total power, to occupy only 25%-50% of the total chip area [49]. This allocation of processing resources facilitates the development of sub-millimeter high heat flux regions, or "flux-spots," that may exceed the average chip flux by a factor of six to ten, with peak fluxes approaching 1000 W/cm^2 [49] [52]. In the absence of an adequate cooling solution, these flux-spots produce locally high temperatures and extreme thermal gradients which can degrade processor performance and compromise reliability through the acceleration of thermally activated failure mechanisms [51]. While the roadmap projections and non-uniform power dissipation trends cited here relate to planar IC technology, 3D chip stacks are expected endure flux-spots of similar size and severity [53]. Figure 4-2 depicts the non-uniform power dissipation for a representative chip and the corresponding temperature non-uniformity.

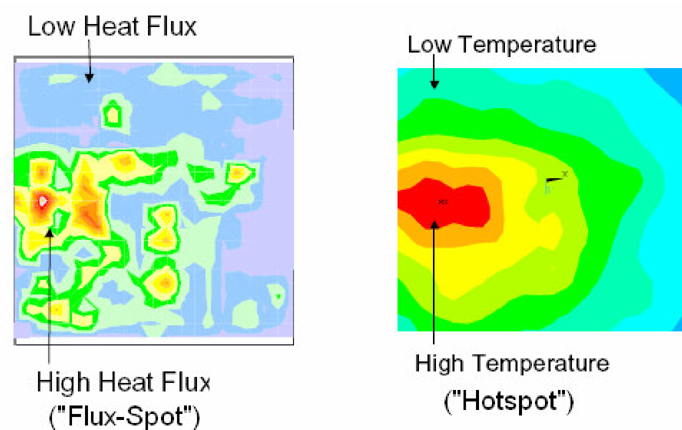


Figure 4-2: On-chip heat flux (left) and temperature map (right) (adapted from [51])

In this Chapter, the term "flux-spot" is used in reference to the actual region of elevated chip power dissipation, while the term "hotspot" refers to the temperature rise associated with a given flux-spot. This distinction is useful since there are two primary approaches to alleviating high temperatures that result from non-uniform power dissipation: The first approach is to reduce the severity of the flux-spot by altering its fundamental characteristics (e.g. reducing its size and/or associated power dissipation); the second approach is to leave the flux-spot characteristics unaltered, but implement a thermal management solution that delivers more targeted and effective cooling to the flux-spot. The former approach, which might be termed 'flux-spot mitigation,' is commonly achieved by altering floorplanning at the microarchitectural level [54] and/or reallocating computational tasks in real time [55] [56] in an attempt to change spatial or temporal non-uniformities in power dissipation. While flux-spot mitigation techniques are sometimes successful, the associated temperature reductions often come at the expense of processor slowdown [57]. Alternatively, the second approach, or 'hotspot remediation,' which is the focus of this chapter, requires the use of novel cooling techniques to selectively cool sub-millimeter flux-spots

In light of the deleterious effects associated with severe flux-spots, recent thermal management designs have diverted focus from uniform reduction of die temperature to site-specific flux-spot cooling. Advanced liquid cooling techniques – such as single and two-phase flow in microchannels [58], thin-film microgap cooling [59], microjet impingement [60], and spray cooling [61] – offer high heat transfer coefficients that may meet the local cooling needs imposed by the presence of flux-spots on planar chips.

However, 3D chip architectures promote the development of flux-spots on internal device layers, which may be inaccessible by proposed applications of microjet and spray cooling technology. Thus, microchannel and microgap cooling approaches, which were outlined in the previous chapter, will likely be the favored methods of delivering fluid to internal flux-spots in 3D chip stacks. While these approaches provide very high heat transfer coefficients, their cooling performance is directly tied to channel/gap size and mean fluid velocity. Therefore, global application of a microchannel or microgap system that is designed to sufficiently cool severe flux-spots will result in overcooling of the remainder of the chip at the expense of increased pressure drop and required pumping power. Novel cooling techniques with the ability to provide targeted flux-spot cooling could be used in conjunction with microchannel and microgap cooling to reduce the required heat transfer coefficient and thus decrease the required pumping power.

As reported in [62], mini-contact enhanced bismuth telluride thermoelectric coolers (TEC) offer promising hotspot remediation in the presence of a global cooling scheme. These thermoelectric coolers are solid state refrigeration devices that – upon the application of electric current – take advantage of the Peltier effect to provide a local cooling flux on the back of the chip. It is reported that hotspot temperature reductions as large as 17 C are achievable with an optimized mini-contact pad and low thermal contact resistances [62]. Thus, use of thin-film TECs could be an attractive flux-spot cooling approach, provided that the TECs can be made small enough, and that they can be implemented in the processing of 3D chip stacks without the development of significantly detrimental thermal contact resistances. Alternatively, it is reported in [64] that the thermoelectric properties of the silicon chip itself can be leveraged to yield significant

hotspot remediation. In this approach, electric current is delivered to the silicon via a metal post that might be monolithically grown on the back of the silicon to minimize thermal contact resistance. Current flow proceeds from this post, through the silicon, and finally to a ring electrode where Peltier heating takes place (far from the flux-spot). Figure 4-3 indicates how a silicon thermoelectric cooler might be implemented in a 3D chip stack. This approach has a distinct advantage over separate thin film TECs in that it avoids the significant thermal contact resistance likely to be associated with the attachment of thin film TEC's to the silicon chip and occupies very little additional volume; also, the necessary metallization could be added to 3D stack processing with relative ease.

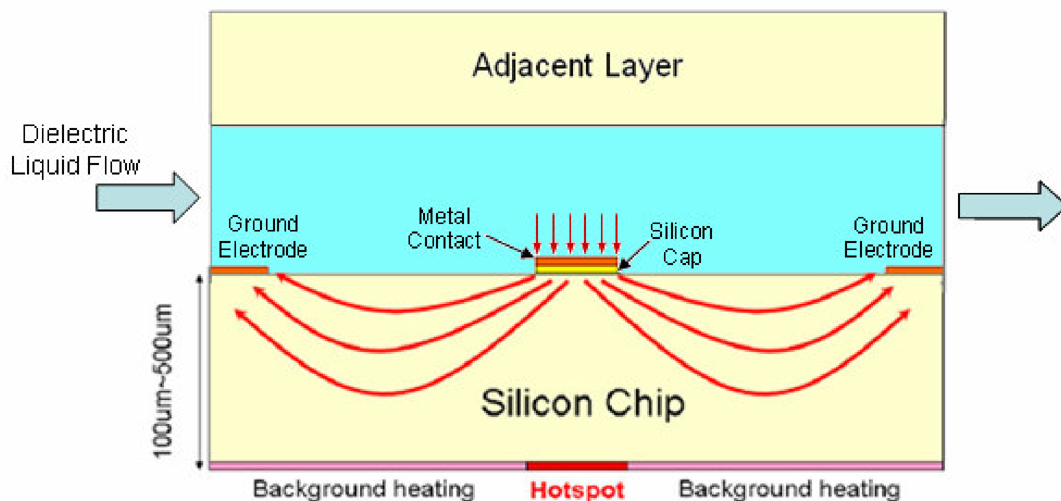


Figure 4-3: Potential implementation of a silicon TEC (adapted from [63])

In contrast to the active thermoelectric cooling techniques discussed above, anisotropic TIMs/spreaders provide a passive approach to hot spot remediation. When

used together with an existing global cooling solution, such materials, bonded directly to the silicon chip as shown in Figure 4-4, can conduct heat laterally away from the flux-spot and towards cooler areas of the chip that are subjected to lower heat flux. These anisotropic TIM/spreaders can thus substantially reduce the heat flux variations imposed at the interface with the global cooling scheme and lower the temperature rise associated with a severe flux-spot. This approach has the usual benefits associated with being a passive technique: 1) it consumes no extra energy and 2) it is reliable (assuming that thermal integrity of the chip/spreader interface is maintained). Additionally, no a priori knowledge of the flux-spot location is required, provided that the TIM/spreader blankets the entire chip area (the performance of the anisotropic TIM/spreader is dependent on the location of the flux-spot, but it is not a "hit or miss" approach like some other flux-spot cooling techniques).

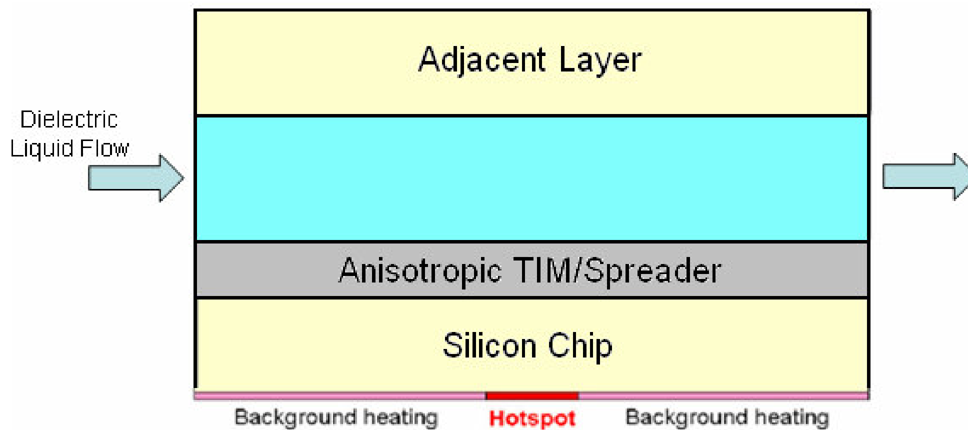


Figure 4-4: Potential implementation of an anisotropic TIM/Spreader in a 3D stack up

The effectiveness of attaching an anisotropic TIM/spreader will be reviewed in this chapter. First, an analytical solution found in the literature will be presented and

discussed. Then, use will be made of this analytical solution to explore parametric sensitivities of in-plane conductivity, TIM/spreader thickness, chip thickness, flux-spot size, and heat transfer coefficient. Finally, the detrimental effects of an interfacial contact resistance will be investigated through the use of a validated finite element model in ANSYS.

Analytical Solution

The bi-layer compound slab shown in Figure 4-5 was used to investigate the hotspot remediation provided by an orthotropic spreader attached directly to the back of a square chip with a single, centrally located, square flux-spot. All external surfaces are assumed adiabatic, except for a heat flux boundary condition at the flux-spot and a convective boundary condition on the back of the orthotropic spreader. Background heating on the active side of the chip is forgone because only the hotspot temperature rise is sought. The boundary conditions are such that inclusion of background heating would simply elevate the entire temperature field. This effect becomes non-trivial if temperature dependent material properties and heat transfer coefficients are employed, but all properties and heat transfer coefficients are taken to be constant in this analysis. The convective boundary condition represents the influence of a global cooling scheme and is modeled by a uniform heat transfer coefficient on the back of the spreader.

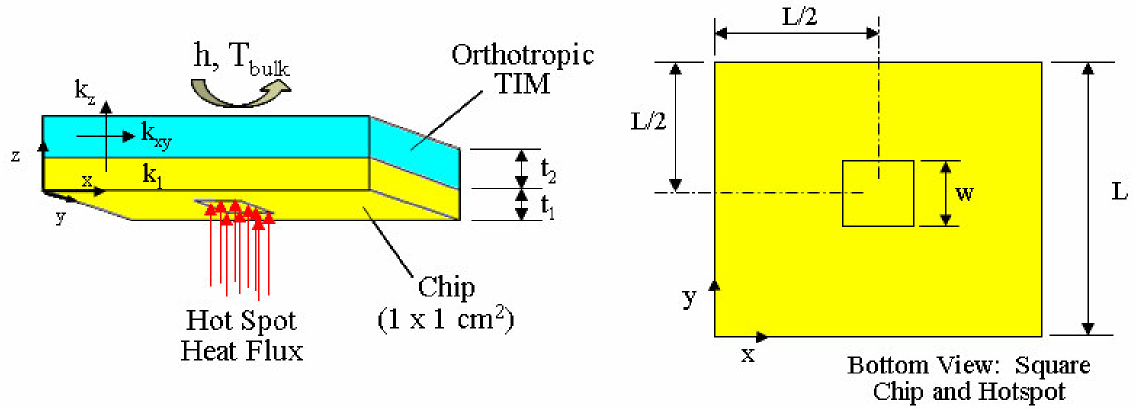


Figure 4-5: Schematic of compound chip/spreader system

Using the separation of variables method, Muzychka, *et al.* [65] have solved the heat conduction equation for the system and boundary conditions shown in Figure 4-5. Their final solution is developed from successive modifications to the solution obtained for an isotropic single-layer system, shown in Figure 4-6. It is reported in [65] that simple alterations can be made to certain parameters in the single-layer solution to arrive at the solution for a bi-layer system with isotropic thermal conductivity in each of the specified layers. The authors further indicate that orthotropic conductivity in the second layer can be accounted for through the use of thickness and conductivity transformations. The intent of this section is to present an abridged derivation of the temperature field solution for the isotropic layer shown in Figure 4-6, discuss the modifications used to account for the presence of a second orthotropic layer as in Figure 4-5, and finally, to recast these results in terms of a useful spreading resistance.

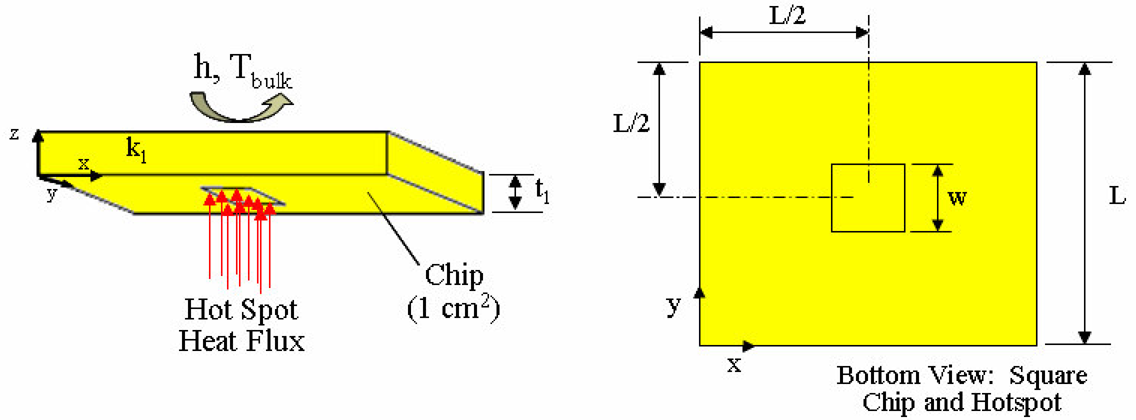


Figure 4-6: Schematic of the single layer system from which the bi-layer solution is found

The governing equation for heat diffusion in an isotropic solid is the so called Laplace Equation in three dimensions and is given by Equation 24:

$$\frac{\partial^2 T}{\partial x^2}+\frac{\partial^2 T}{\partial y^2}+\frac{\partial^2 T}{\partial z^2}=0 \tag{24}$$

where T is the temperature in the solid. The boundary conditions in Figure 4-6 are represented mathematically by the following:

$$\frac{\partial T}{\partial z}=-\frac{Q/A_s}{k_1} \quad for \quad \frac{L}{2}-\frac{w}{2}<x<\frac{L}{2}+\frac{w}{2};\ \frac{L}{2}-\frac{w}{2}<y<\frac{L}{2}+\frac{w}{2};\ z=0 \tag{25}$$

$$\frac{\partial T}{\partial z}=-\frac{h}{k_1}\left[T(x,y,z)-T_{bulk}\right] \quad for \quad 0\le x\le L;\ 0\le y\le L;\ z=t_1 \tag{26}$$

In Equation 25, Q and A_s are the total heat added at the flux spot (in Watts), and the area of the flux-spot, respectively. The heat transfer coefficient in Equation 26 can be either a direct heat transfer coefficient or an effective heat transfer coefficient achieved through the use of extended surfaces. All other outer surfaces are assumed adiabatic.

The essence of the separation of variables method is the assumption that the solution to the problem posed in Equations 24-26 is expressible as a linear combination

of functions that each depend on only one variable, and thus are more easily obtained. In this situation, we seek a separated solution of the form

$$T(x, y, z) = X(x) * Y(y) * Z(z) \tag{27}$$

where the independent variables (i.e. the coordinate directions) are written as lower case letters and the functions are written as capital letters. Assuming a solution of the form in Equation 27 and substituting it into Equation 24 yields the following

$$X''YZ + XY''Z + XYZ'' = 0 \tag{28}$$

where each prime denotes differentiation of the function with respect to its independent variable. Upon dividing both sides of Equation 28 by *XYZ* we arrive at the following

$$\frac{X''}{X} + \frac{Y''}{Y} + \frac{Z''}{Z} = 0 \tag{29}$$

In order for Equation 29 to be satisfied, each of the quotients X''/X, Y''/Y, and Z''/Z must be constant. Therefore, each pair of quotients and constants provides a separate ordinary differential equation that may be solved independently using the stated boundary conditions.

For instance, if we assume that the quotient X''/X is equal to a constant, $-\lambda^2$, then we have the following ordinary differential equation

$$X'' + \lambda^2 X = 0 \tag{30}$$

We know that the walls at $x = 0$ and $x = L$ are adiabatic, so we have

$$\frac{\partial T}{\partial x} = X'YZ = 0 \quad at \quad x = 0 \;\; and \quad x = L \tag{31}$$

Wishing to ignore the trivial case where either, or both, of the funtions *Y(y)* and *Z(z)* vanish, we set $X' = 0$ at $x = 0$ and $x = L$. This provides the boundary conditions from which to solve the ODE in Equation 30. Solutions for the functions *Y(y)* and *Z(z)* are

found in an analogous way. The details of this process are long and beyond the scope of this document. Books by Carslaw and Jaeger [66] and Strauss [67] will provide a sufficient overview of the method of separation of variables for the interested reader.

The final solution for the temperature field in the system shown in Figure 4-6 is reported in [49] to be:

$$\begin{aligned}\Delta T_{bulk} = T(x,y,z) - T_{bulk} = A_0 + B_0 z + &\sum_{m=1}^{\infty} \cos(\lambda_m x)\left[A_m \cosh(\lambda_m z) + B_m \sinh(\lambda_m z)\right]... \\ +&\sum_{n=1}^{\infty} \cos(\delta_n y)\left[A_n \cosh(\delta_n z) + B_n \sinh(\delta_n z)\right]... \\ +&\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \cos(\lambda_m x)\cos(\delta_n y)\left[A_{mn} \cosh(\beta_{mn} z) + B_{mn} \sinh(\beta_{mn} z)\right]\end{aligned} \quad (32)$$

Notice in Equation 32 that the bulk temperature of the fluid has been placed on the left hand side of the equation, yielding an expression for the temperature excess, ΔT_{bulk}, in the solid. The eigenvalues are given by $\lambda_m = \frac{m\pi}{L}$, $\delta_n = \frac{n\pi}{L}$, and $\beta_{mn} = \sqrt{\lambda_m^2 + \delta_n^2}$ and the Fourier coefficients are found from application of the boundary conditions in the z-direction. Imposing the boundary conditions at $z = 0$ results in expressions for the 'A' coefficients as follows:

$$A_0 = \frac{Q}{L^2}\left(\frac{t_1}{k_1} + \frac{1}{h}\right) \quad (33)$$

$$A_m = \frac{2Q\left[\sin\left(\frac{L+w}{2}\lambda_m\right)-\sin\left(\frac{L-w}{2}\lambda_m\right)\right]}{L^2 w k_1 \lambda_m{}^2 \phi(\lambda_m)}$$

$$A_n = \frac{2Q\left[\sin\left(\frac{L+w}{2}\delta_n\right)-\sin\left(\frac{L-w}{2}\delta_n\right)\right]}{L^2 w k_1 \delta_n{}^2 \phi(\delta_n)} \tag{34}$$

$$A_{mn} = \frac{16Q\cos\left(\frac{\lambda_m L}{2}\right)\sin\left(\frac{\lambda_m w}{2}\right)\cos\left(\frac{\delta_n L}{2}\right)\sin\left(\frac{\delta_n w}{2}\right)}{L^2 w^2 k_1 \beta_{mn}\lambda_m \delta_n \phi(\beta_{mn})}$$

Meanwhile, the '*B*' coefficients are found from application of the boundary condition at $z = t_1$. The coefficient B_0 is found to be $-Q/k_1L^2$ and the remaining '*B*' coefficients are related to those in Equation 34 by a parameter, φ, in the following way

$$B_{m,n,mn} = -\varphi A_{m,n,mn} \tag{35}$$

where the spreading parameter, φ, is a function of a dummy variable ζ, with $\varphi(\zeta)$ given by

$$\varphi(\zeta) = \frac{\zeta\sinh(\zeta t_1) + h/k_1\cosh(\zeta t_1)}{\zeta\cosh(\zeta t_1) + h/k_1\sinh(\zeta t_1)} \tag{36}$$

and ζ is replaced by λ_m, δ_n, or β_{mn} in Equation 36 as appropriate.

An expression for the excess temperature on the active side of the chip can be found by substituting $z = 0$ into Equation 32. The nature of the hyperbolic sine and hyperbolic cosine functions is such that $sinh(0) = 0$ and $cosh(0) = 1$. This conveniently allows us to drop some terms in Equation 32 to arrive at a simpler expression for the excess temperature on the heated side of the solid

$$\begin{aligned}\Delta T_{bulk}(x,y,z=0) = A_0 + \sum_{m=1}^{\infty} A_m\cos(\lambda_m x) + \sum_{n=1}^{\infty} A_n\cos(\delta_n y)\ldots \\ + \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} A_{mn}\cos(\lambda_m x)\cos(\delta_n y)\end{aligned} \tag{37}$$

The authors of [65] report that the excess temperature on the active side of the chip (z = 0) in a bi-layer system can be obtained through use of Equation 37 when some simple substitutions are made for the coefficient A_0 (Equation 33) and the parameter $\varphi(\zeta)$ (Equation 36). The proposed expressions for A_0 and $\varphi(\zeta)$ for a bi-layer system are given by Equations 38 and 39, respectively.

$$A_0 = \frac{Q}{L^2}\left(\frac{t_1}{k_1}+\frac{t_2}{k_2}+\frac{1}{h}\right) \tag{38}$$

$$\varphi(\zeta) = \frac{\left(\alpha e^{4\zeta t_1} - e^{2\zeta t_2}\right)+\psi\left[e^{2\zeta(2t_1+t_2)} - \alpha e^{2\zeta(t_1+t_2)}\right]}{\left(\alpha e^{4\zeta t_1} + e^{2\zeta t_2}\right)+\psi\left[e^{2\zeta(2t_1+t_2)} + \alpha e^{2\zeta(t_1+t_2)}\right]} \tag{39}$$

The new parameters in Equation 39 are to be evaluated as $\psi = \frac{\zeta + h/k_2}{\zeta - h/k_2}$ and $\alpha = \frac{1-k_2/k_1}{1+k_2/k_1}$ where ζ is again replaced by λ_m, δ_n, or β_{mn} as appropriate.

Bear in mind that the conductivity of the second layer, k_2, is assumed to be isotropic in the above expressions. However, it is revealed in [65] that if either of the layers exhibits orthotropic conductivity, the solution for purely isotropic layers can be used when the following length scale and conductivity transformations are employed in the subject orthotropic layer:

$$\begin{aligned} k &\rightarrow k_{eq} = \sqrt{k_{xy}k_z} \\ t &\rightarrow t_{eq} = \frac{t}{\sqrt{k_z/k_{xy}}} \end{aligned} \tag{40}$$

Thus, k_2 and t_2 in Equations 38-39 can be replaced by the transformations in Equation 40 to account for orthotropicity in the spreader.

The collection of Equations 37-40, along with the Fourier coefficients in Equation 34, provide a full solution for the excess temperature on the active side of the compound

structure shown in Figure 4-5. The overall resistance to heat transmission for the system in Figure 4-5 is comprised of 1) the resistance to one-dimensional heat flow, and 2) the spreading resistance. Each of the four terms in Equation 37 can be traced to one of these two resistances. The first term in Equation 37 is the Fourier coefficient, A_0, which is given by Equation 38 and is attributable to uniform one-dimensional conduction through the compound system. The three remaining terms are related to thermal spreading and thus vanish as the hotspot size approaches the chip size. It will prove useful to relate parametric trends to the concepts of one-dimensional and spreading resistances in later discussions. Therefore, the above results will now be recast in terms of thermal resistance following the work of Muzychka, *et al.* [65].

The total thermal resistance can be related to the average excess temperature at the hotspot through the following definition

$$R_T = \frac{\overline{\Delta T_{bulk}}}{Q} \tag{41}$$

where R_T is the total thermal resistance, including thermal transport by both conduction and convection. The term $\overline{\Delta T_{bulk}}$ in Equation 41 is found by integrating Equation 37 over the hotspot region and dividing by the hotspot area, or, expressed mathematically:

$$\overline{\Delta T_{bulk}} = \frac{1}{A_s} \iint_{A_s} \Delta T_{bulk}(x, y, 0)\, dA_s \tag{42}$$

As noted, the total thermal resistance is also the sum of the one-dimensional resistance and the spreading resistance as follows

$$R_T = R_{1D} + R_s \tag{43}$$

The one-dimensional resistance to heat conduction is easily found to be

$$R_{1D}=\frac{t_1}{k_1L^2}+\frac{t_2}{k_zL^2}+\frac{1}{hL^2} \tag{44}$$

Meanwhile, the spreading resistance, R_s, can be found by substituting Equations 41, 42, and 44 into Equation 43:

$$R_s=\frac{1}{A_sQ}\iint_{A_s}\Delta T_{bulk}(x,y,0)dA_s-\left(\frac{t_1}{k_1L^2}+\frac{t_2}{k_zL^2}+\frac{1}{hL^2}\right) \tag{45}$$

The integration in Equation 45 was carried out by Yovanovich *et al.* in [68] and was found to yield the following expression for the spreading resistance in a bi-layer structure with a centrally located flux-spot

$$\begin{aligned}R_s=&\frac{1}{2\left(w/2\right)^2\left(L/2\right)^2k_1}\left[\sum_{m=1}^{\infty}\frac{\sin^2\left(w\delta_m/2\right)}{\delta_m{}^3\varphi(\delta_m)}+\sum_{n=1}^{\infty}\frac{\sin^2\left(w\lambda_n/2\right)}{\lambda_n{}^3\varphi(\lambda_n)}\right]\cdots\\&+\frac{1}{\left(w/2\right)^4\left(L/2\right)^2k_1}\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\frac{\sin^2\left(w\delta_m/2\right)\sin^2\left(w\lambda_n/2\right)}{\delta_m{}^2\lambda_n{}^2\beta_{mn}\varphi(\beta_{mn})}\end{aligned} \tag{46}$$

where the eigenvalues are the same as those for Equation 32 and the parameter φ is given by Equation 39. The reader is reminded that the thickness and conductivity transformations given in Equation 40 can be used in Equation 46, if either of the layers exhibit orthotropic conductivity.

The analytical solutions developed in this section allow for rapid exploration of the parametric trends in hotspot remediation with orthotropic TIMs/spreaders. MATLAB codes were developed to aid in the evaluation and graphical representation of the series solutions presented above (see Appendix B). Input parameters to the MATLAB program were varied to determine the parametric sensitivities of hotspot temperature and overall thermal resistance to 1) in-plane spreader conductivity, 2) spreader thickness, 3) hotspot

size, and 4) heat transfer coefficient. The results of these parametric explorations are discussed in the following sections.

Variation of In-Plane Spreader Thermal Conductivity

Successful hotspot remediation via implementation of an orthotropic spreader depends on the ability of the spreader to conduct heat away from local regions of high heat flux to other parts of the chip with lower thermal loads. Therefore, one might anticipate that the hotspot remediation provided by an orthotropic spreader is directly tied to its in-plane thermal conductivity, k_{xy}. This conclusion is supported by the thickness and conductivity transformations in Equation 40, which clearly show that any increase in k_{xy} for fixed values of k_z and t will result in larger values k_{eq} and t_{eq}. Both of these effects reduce the spreading portion of the overall thermal resistance, while the resistance to one-dimensional conduction remains constant (since $t_{eq}/k_{eq} = t/k_z$). Therefore, an increase in the conductivity ratio, k_{xy}/k_z, leads to an attendant decrease in the overall thermal resistance, and thus a decrease in the average hotspot temperature.

A representative chip/spreader system with the parameter settings listed in Table 4-1 was used to determine the magnitude of the benefits of increasing k_{xy} (see Figure 4-5 for parameter definitions). Please note that the parameters in Table 4-1 will be used throughout this section and can be assumed unless otherwise specified. The thru-plane conductivity of the spreader, k_z, was chosen to be 5 W/m-K because this is a nominal thru-plane conductivity for some natural graphite materials as well as pyrolytic graphite [69]-[70]. A heat transfer coefficient of 10,000 W/m^2-K was applied to represent the presence of an aggressive cooling approach (e.g. pool boiling or a microchannel cold

plate); also, and average heat transfer coefficient of 10,000 W/m^2-K can be achieved in the internal cooling approach outlined in the previous chapter with a fluid inlet velocity of 0.55 m/s and a gap of 500 µm. The in-plane conductivity was varied between 5 and 1800 W/m-K in order to determine the effect that the degree of anisotropy had on hotspot remediation.

Table 4-1: Parameter Settings for k_{xy} variation

Parameter	Description	Value	Units
t_1	chip thickness	250	µm
t_2	spreader thickness	500	µm
k_1	isotropic chip conductivity	163	W/m-K
k_z	thru-plane spreader conductivity	5	W/m-K
k_{xy}	in-plane spreader conductivity	varies	W/m-K
q"	hotspot heat flux	1.4	kW/cm^2
h	effective heat transfer coefficient	10000	W/m^2-K
T_{bulk}	ambient temperature for convective transfer	25	C
L	chip size, square	1	cm
w	hotspot size, square	500	µm

The excess temperature profiles on the active side of the chip, subjected to a 1.4kW/cm^2, 0.5mm flux spot, are shown for several different in-plane conductivities in Figure 4-7. It is found that the isotropic spreader ($k_z = k_{xy} = 5$ W/m-K) has a hotspot temperature that is nearly 47.5 C above ambient. As expected, increasing the in-plane conductivity, k_{xy}, decreases this temperature excess. For $k_z = 5$ and $k_{xy} = 350$, which is representative of natural graphite sheets [69], the hotspot temperature is ~9.3 C below that obtained through use of the isotropic spreader. If the in-plane conductivity is further increased to 1800 W/m-K, a hotspot suppression of ~14.3 C is attained. It is noteworthy

that while high in-plane conductivities yield substantial reductions in the maximum hotspot temperature, which occupies 0.16mm^2, much of the chip area and most notably the edges of the chip experience very modest increases in temperature.

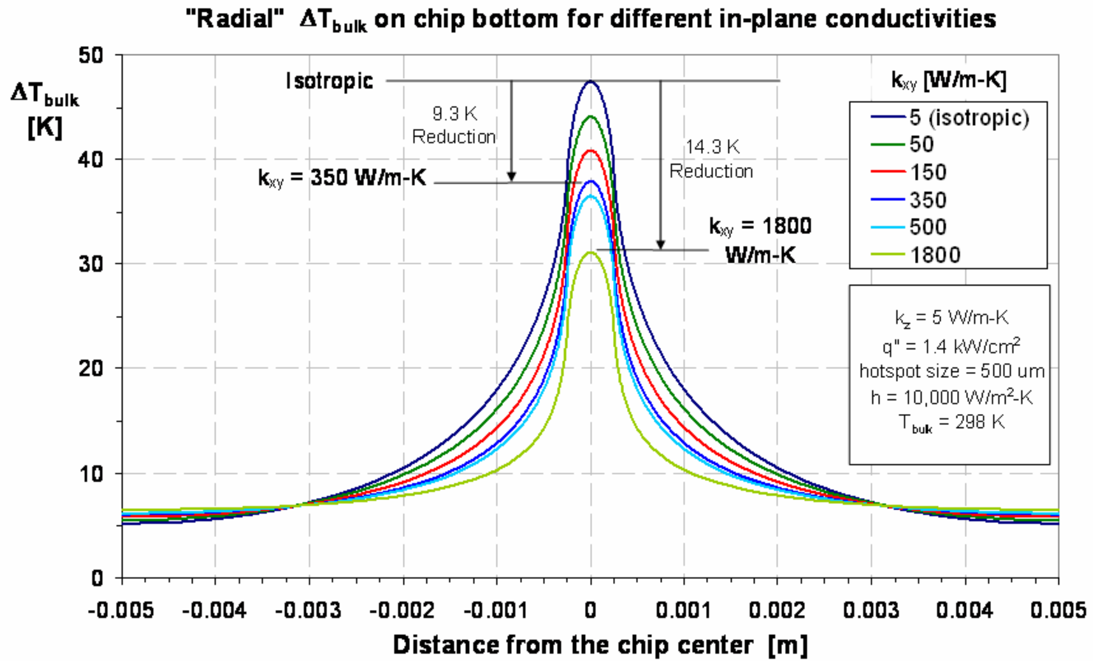


Figure 4-7: Excess temperature profiles taken through the middle of the active chip surface for various k_{xy}

Hotspot remediation relative to the temperature rise resulting from a low conductivity isotropic spreader is more clearly shown for each k_{xy} on the left side of Figure 4-8. In order to put these results into context, the performance of several alternative spreaders with varying conductivities were evaluated subject to the same geometric parameters and thermal boundary conditions listed in Table 4-1. With the isotropic spreader (k = 5 W/m-K) as a baseline, the hotspot cooling achieved by each alternative spreader is shown on the right hand side of Figure 4-8. The first data point corresponds to an isotropic spreader with the same conductivity of the silicon chip, k =

163 W/m-K; this scenario represents the possibility of simply extending the silicon chip to provide better cooling. It is found that extending the silicon in this way provides about 18.3 C of cooling, which exceeds by 4 C the hotspot suppression achieved through use of the orthotropic spreader with k_z = 5 W/m-K and k_{xy} = 1800 W/m-K. However, it can be seen in Figure 4-8 that an orthotropic spreader with k_z = 10 W/m-K and k_{xy} = 1700 W/m-K – which is characteristic of annealed pyrolytic graphite (APG) [70] – is found to provide the same hotspot suppression as the extended silicon chip. Implementation of a copper spreader with isotropic thermal conductivity of 400 W/m-K provides further hotspot suppression of ~23.0 C. Of the alternative spreaders considered, the best hotspot suppression of ~27.0 C was provided by the orthotropic CVD diamond film with k_z = 1450 W/m-K and k_{xy} = 2000 W/m-K [71]. In the case of the diamond it was assumed, perhaps unrealistically, that the larger of the diamond's conductivities could be oriented to correspond with the plane of the chip. Regardless, the exceptional performance of the diamond film is most readily attributed to its large average thermal conductivity, with only modest enhancement coming from its relatively small degree of anisotropy. For instance, the orthotropic diamond considered here only provides 0.4 C better cooling than an isotropic spreader with $k_z = k_{xy}$ = 1450 W/m-K. Also, orienting the conductivities in a more realistic fashion, with k_z = 2000 W/m-K and k_{xy} = 1450 W/m-K, reduces the hotspot temperature by only 0.3 C. Thus, the anisotropy of the diamond plays only a minor role in its effectiveness as a spreader.

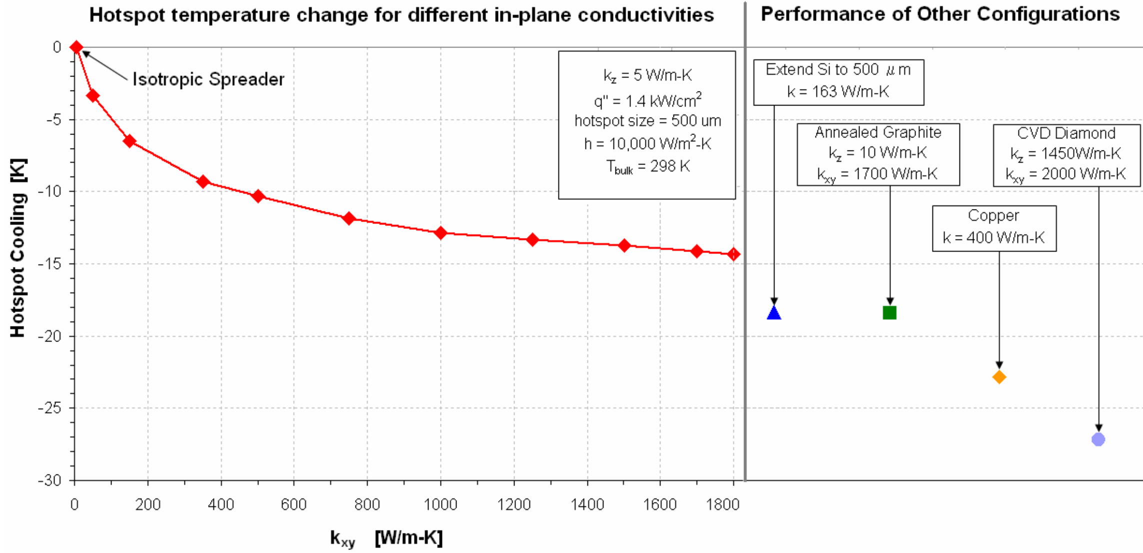


Figure 4-8: Hotspot cooling compared to an isotropic spreader

The above results show that increasing the in-plane thermal conductivity of an orthotropic spreader can provide substantial hotspot temperature reduction. For the parameters considered, the cooling performance of a highly orthotropic APG spreader was able to match the cooling provided by an equal thickness of pure silicon. This is an important result given the general reluctance of chip manufacturers to allocate valuable silicon for thermal management that might otherwise be used to produce more chips. Despite the good performance of the best highly orthotropic spreaders, an equally sized copper spreader provides about 4 C better hotspot remediation for the conditions examined. However, natural graphite orthotropic spreaders may have a practical advantage over copper spreaders in space-constrained 3D chip stacks since they provide respectable hotspot cooling, yet can be made extremely pliable and thin (one company supplies 50 µm thick graphite sheets [72]). Also, graphite spreaders offer a weight advantage over copper spreaders of equal size since the density of natural graphite is

approximately 4 times lower than that of copper. This weight difference could be significant in mobile applications where portability is a top concern. Furthermore, highly orthotropic spreaders with low thru-plane conductivity have the ability to reduce hotspot temperatures while simultaneously insulating adjacent layers of a 3D chip stack.

A more subtle advantage of highly orthotropic TIMs/spreaders is that they tend to reduce surface temperature variation at the interface with the cooling scheme. Figure 4-9 shows the temperature rise on the back of the silicon, copper, and APG spreaders as a function of location on the back of the spreader. The ΔT displayed in Figure 4-9 is the difference between the local temperature and the edge temperature; thus, the value of ΔT vanishes for all profiles as the edge of the spreader is approached. For the parameters considered it can be seen that the silicon and copper spreaders allow a maximum temperature variation of 8 C and 4.3 C on the back of the spreader, respectively, compared to a maximum variation of 0.05 C for the APG spreader. As such, implementation of an APG spreader would yield a system that is less susceptible to local dryout or critical heat flux under the influence of a directly applied two-phase cooling scheme (e.g. pool boiling or flow boiling).

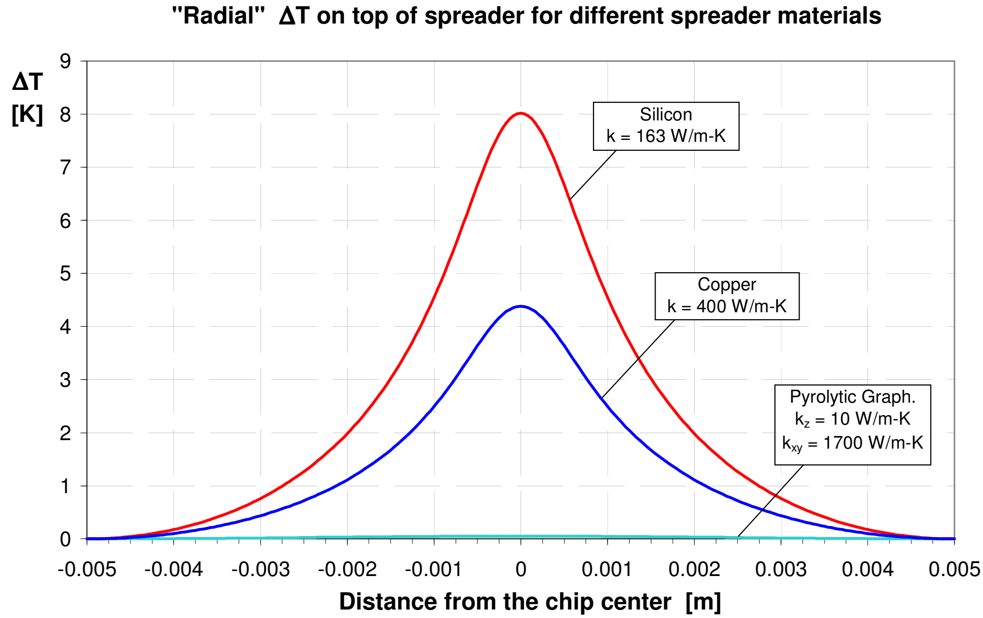


Figure 4-9: Temperature rise on the back of silicon, copper, and APG spreaders

Variation of Spreader Thickness

The thickness of the TIM/spreader in Figure 4-5 will play a direct role in not only hotspot cooling performance, but also in the realistic implementation of orthotropic TIMs/spreaders in 3D chip stacks where persistent miniaturization trends set practical limits on the space claimed by thermal management solutions. As such, it is important to explore the tradeoffs between spreader thickness and cooling performance when assessing the merits of an anisotropic spreader. The impact of spreader thickness on hotspot remediation is best understood in terms of the overall thermal resistance of the system, R_T, which is the sum of the spreading resistance, R_s, and the resistance to one-dimensional conduction and convection, R_{1D} (see Equation 43). MATLAB codes were developed to aide in the evaluation of R_{1D} (Equation 44) and R_s (Equation 46) for various spreader thicknesses (see Appendix C).

Using the model in Figure 4-5 and the parameter settings in Table 4-1, the thickness of each of the spreaders represented in Figure 4-8 was varied to determine the effect on R_{1D}, R_s, and hence R_T. Typical trends seen during this analysis are depicted in Figure 4-10 where it is clear that the total thermal resistance experiences a minimum for some critical value of the spreader thickness. This can be explained by monitoring the variations of R_s and R_{1D} for increasing spreader thickness. Near a spreader thickness of zero, the total thermal resistance of the system approaches that of a single layer of silicon (see Figure 4-6). As thickness increases, the one-dimensional resistance grows linearly and the spreading resistance deteriorates monotonically. Initially, R_s deteriorates at a greater rate than R_{1D} increases so that the total thermal resistance declines. However, R_s experiences a diminishing rate of decline for larger and larger spreader thickness such that the negative slope of R_s eventually equals the positive slope of R_{1D} in magnitude. The thickness at which this occurs is the optimum spreader thickness for minimization of the total thermal resistance (and, thus, minimization the average hotspot temperature). For any increase in spreader thicknesses beyond this optimum, the linear rise in R_{1D} is greater than the decrease in R_s and the total thermal resistance increases.

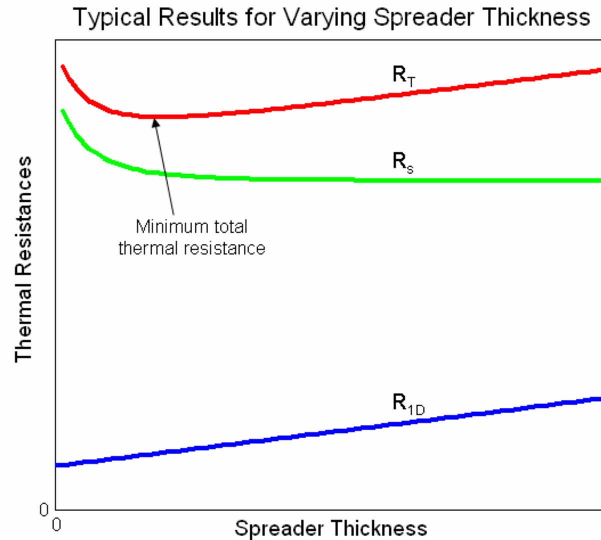


Figure 4-10: Typical thermal resistance trends for various spreader thicknesses

It should be noted that the spreading resistance trend shown in Figure 4-10 is only observed for spreaders of sufficiently high conductivity. If spreader conductivities become low enough, the variation of spreading resistance with thickness can change significantly, causing the variation of R_T to diverge from the trend observed in Figure 4-10. For instance, consider again the model in Figure 4-5 with the settings in Table 4-1 and an isotropic spreader of $k_z = k_{xy} = 5$ W/m-K. The variation of spreading resistance with increasing thickness is shown for this case in Figure 4-11 (All of the spreading resistance data in Figure 4-11 is normalized by the spreading resistance that would be achieved if no spreader were present and the bare chip was cooled directly by the same heat transfer coefficient; this resistance is called R_{bare} and $R_s/R_{bare}\rightarrow 1$ as $t_2\rightarrow 0$). It can be see that R_s persistently escalates for increasing values of t_2; thus, no minimum total thermal resistance exists and the addition of any size spreader, in this case serving as an "insulator," will only increase the average hotspot temperature as compared to the case of

a directly cooled bare chip. A more complex spreading resistance variation is seen in Figure 4-11 for an orthotropic spreader with k_z = 5 W/m-K and k_{xy} = 20 W/m-K. The spreading resistance first decreases, then achieves a minimum near 130 μm, and finally increases for all larger thicknesses. A similar undulating pattern is seen for an isotropic spreader with $k_z = k_{xy}$ = 10 W/m-K. It is possible for the total thermal resistance to undergo a minimum for these more complex R_s patterns, but no minimum is guaranteed; rather, the existence of a minimum will depend on the rate at which the one-dimensional resistance increases.

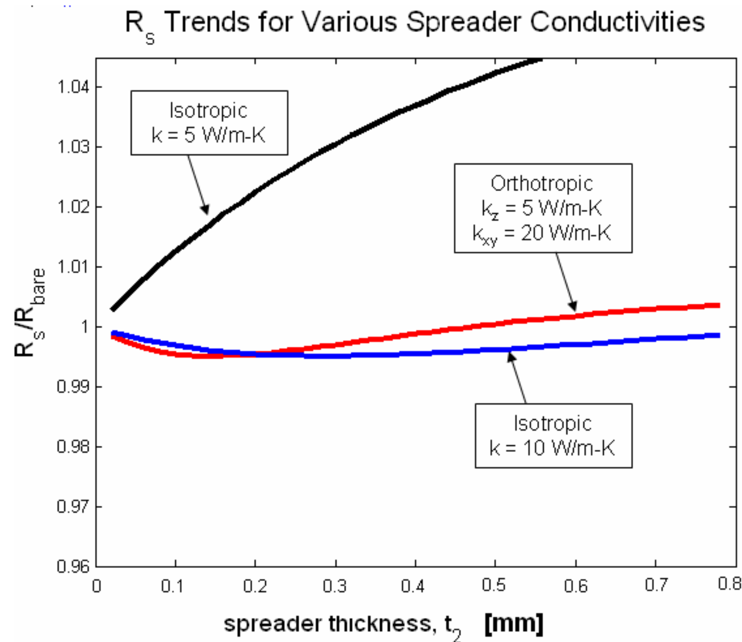


Figure 4-11: R_s variations for spreaders with 'low' and 'intermediate' thermal conductivity

Analytical determination of the conductivity marking the transition from the spreading resistance variations shown in Figure 4-11 to the monotonically decreasing trend shown in Figure 4-10 is made difficult by the mathematical complexities of the

equation for spreading resistance (Equation 46). Therefore, the developed MATLAB code was used to approximately determine this transition point by trial and error. For the parameters considered, it was found that the development of a monotonically decreasing spreading resistance occurs in the vicinity of k_{eq} = 14.5 W/m-K for either an isotropic or orthotropic spreader, where $k_{eq} = k_z = k_{xy}$ for an isotropic material and k_{eq} for an orthotropic spreader is given by Equation 40. Solution of Equation 40 for an orthotropic spreader with k_z = 5 W/m-K reveals that an in-plane conductivity greater than 42 W/m-K will yield $k_{eq} \geq 14.5$ W/m-K, and thus, a spreading resistance trend like that shown in Figure 4-10.

In the previous section, the sensitivity of hotspot temperature was explored with respect to varying in-plane conductivity for TIMs/spreaders of fixed thickness with k_z = 5 W/m-K (see Figure 4-8, left side). The same range of spreader conductivities will now be examined for varying spreader thickness. Figure 4-12 – which shows the variation of total thermal resistance against spreader thickness for a variety of k_{xy} – confirms the expectations set forth in the preceding paragraphs. First, the total thermal resistance for each of the k_{xy} curves approaches the resistance of a directly cooled chip as spreader thickness approaches zero ($R_T \rightarrow 10.84$ K/W as $t_2 \rightarrow 0$). Second, corroborating the results of the previous section, the total thermal resistance and average hotspot excess temperature decrease for increasing k_{xy}. Third, for low in-plane conductivities – as seen for the dashed lines where k_{xy} = 5 and 25 W/m-K – there is no optimum thickness and R_T is strictly increasing. Finally, high in-plane conductivities yield a trend like that in Figure 4-10 where R_T is minimized for some optimum spreader thickness. The optimum

thickness varies with in-plane conductivity as shown in Figure 4-13, exhibiting a peak optimum thickness of ~160 µm in the vicinity of k_{xy} = 200 W/m-K.

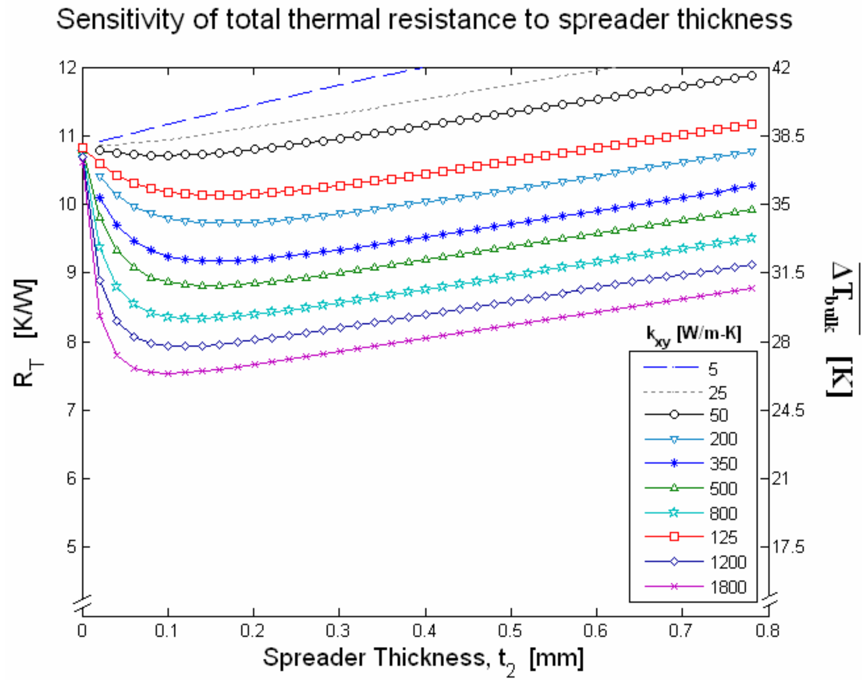


Figure 4-12: Variation of R_T and average hotspot excess temperature for increasing spreader thickness (k_z = 5 W/m-K)

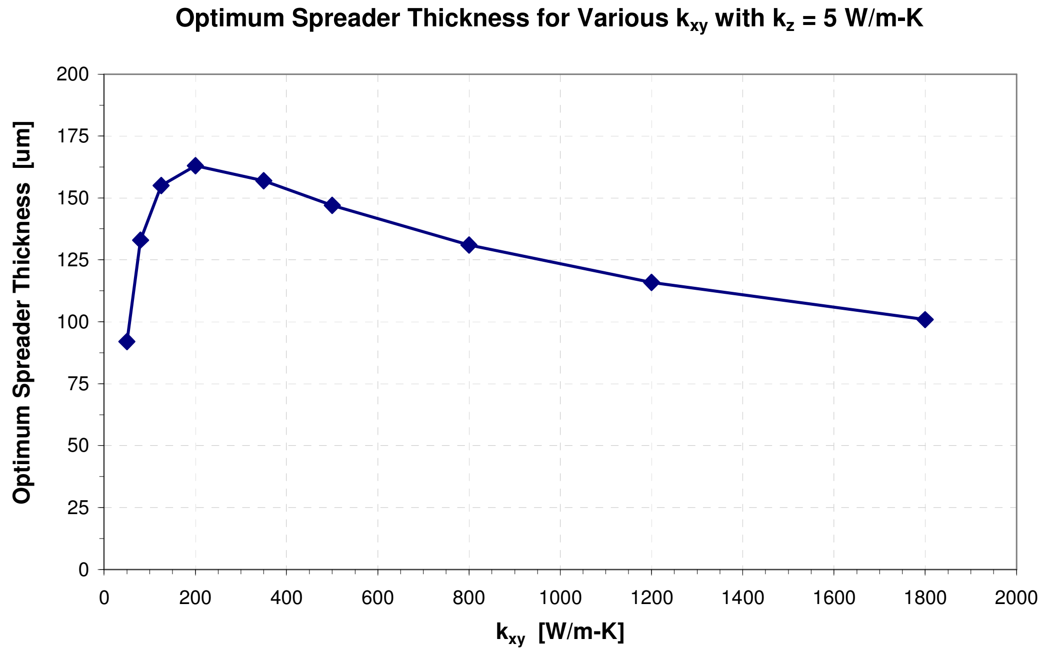


Figure 4-13: Optimum Spreader thickness varies with in-plane thermal conductivity

Recall that several alternative spreaders were examined in the previous section, namely annealed pyrolytic graphite (APG), silicon, copper, and CVD diamond (see Figure 4-8, right side). The variation of R_T with t_2 for each of these alternative spreaders can be seen in Figure 4-14. It is found that the highly orthotropic APG exhibits similar behavior to that shown in Figure 4-12, with a distinct minimum occurring at a spreader thickness of 157 μm. However, the silicon, copper, and diamond spreaders all exhibit a broad 'plateau' for which the thermal resistance remains relatively constant with thickness (these spreaders do indeed have minimum values of $R_{T,}$ but the minima occur beyond the 2 mm thickness at which plotting was stopped in Figure 4-14). The appearance of the 'plateau' for in total thermal resistance for silicon, copper, and diamond is attributable to the high thru-plane conductivities of these materials. Recall from the discussion of Figure 4-10 that a minimum in R_T occurs when the downward

slope of R_s is equal to the upward slope of R_{1D}. The one-dimensional resistance of a spreader with high k_z is relatively insensitive to changes in t_2; therefore, the negative slope of R_s must decrease significantly in magnitude before it offsets the positive slope of R_{1D}. For silicon, copper, and diamond, the slope of R_s approaches that of R_{1D} quite slowly, leading to the formation of the 'plateau' seen in Figure 4-14. The thicknesses at which the total resistance of silicon, copper, and diamond are within about 1% of their minima are 1400 µm, 1300 µm, and 800 µm, respectively.

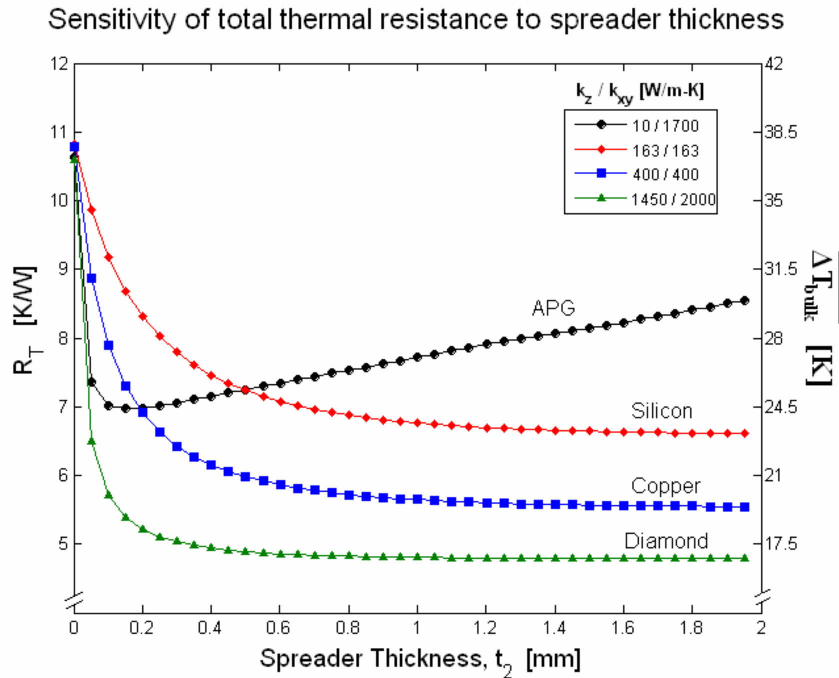


Figure 4-14: Variation of R_T for increasing spreader thickness for alternative spreaders

Figure 4-13 and Figure 4-14 reveal that highly orthotropic graphite TIMs/spreaders yield optimum hotspot cooling performance at relatively low thicknesses – under 165 µm – for the conditions examined. Furthermore, it is interesting to note in Figure 4-14 that the orthotropic APG spreader yields lower average hotspot temperatures

than copper for thicknesses up to ~200 μm and lower temperatures than silicon up to ~500 μm. Also, the minimum average hotspot excess temperature for APG is 24.4 K at 157 μm, which is only 5.0 K and 1.3 K hotter than that provided by nine times the thickness of copper and silicon (~1.4 mm), respectively. The exceptional performance of highly orthotropic TIMs/spreaders at low thickness may lead them to be favored over conventional heat spreading materials in space constrained 3D chip stacks.

Up to this point, the thickness of the silicon chip in Figure 4-5 has been fixed at 250 μm for all cases. One can understand that the silicon chip may indeed exhibit different thicknesses depending on wafer processing, and that these different chip sizes will affect the overall thermal resistance of the chip/spreader system. The variation of R_T with spreader thickness is shown in Figure 4-15 for a spreader with k_z = 5 W/m-K and k_{xy} = 350 W/m-K, where each of the plotted line represents a different chip thickness (all other parameters remain unchanged, see Table 4-1). The total thermal resistance is seen to decrease with increasing chip thickness. However, it is clear that the additional cooling provided by an optimally thick spreader becomes less dramatic for greater chip thicknesses. This is more clearly shown in Figure 4-16 where the R_T data for each plotted line in Figure 4-15 has been normalized by the total thermal resistance the would exist if the spreader were removed and the bare chip were cooled directly (this resistance is called $R_{T,bare}$). It is seen that an optimally thick spreader reduces $R_{T,bare}$ by ~43% when the chip is 75 μm thick but only reduces $R_{T,bare}$ by ~7% when the chip is 400 μm thick. The reduction in spreader effectiveness for increasing chip size is the result of the silicon bearing more of the burden of spreading heat, and thus is not exclusive to orthotropic spreaders (i.e. highly conductive isotropic spreaders suffer an analogous reduction in

effectiveness). As chip thicknesses shrink – which is a likely scenario as chip manufactures strive to more efficiently utilize expensive silicon ingots – the inherent spreading provided by the chip is reduced. The results in Figure 4-16 indicate that the implementation of orthotropic TIM/spreaders can compensate for the loss of inherent spreading in thinner silicon chips of the future.

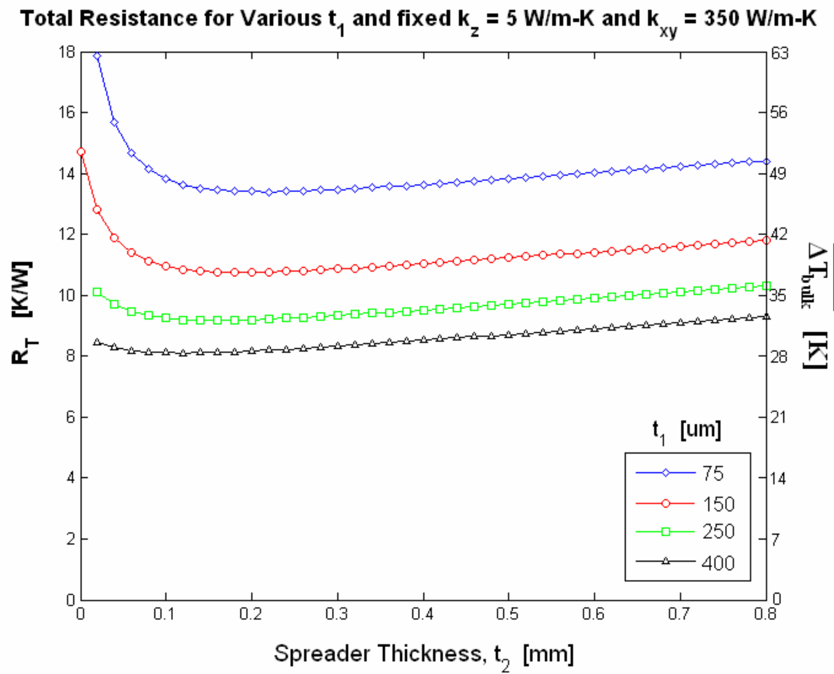


Figure 4-15: Effect of chip thickness of spreader thickness variation for an orthotropic spreader with k_z = 5 W/m-K and k_{xy} = 350 W/m-K

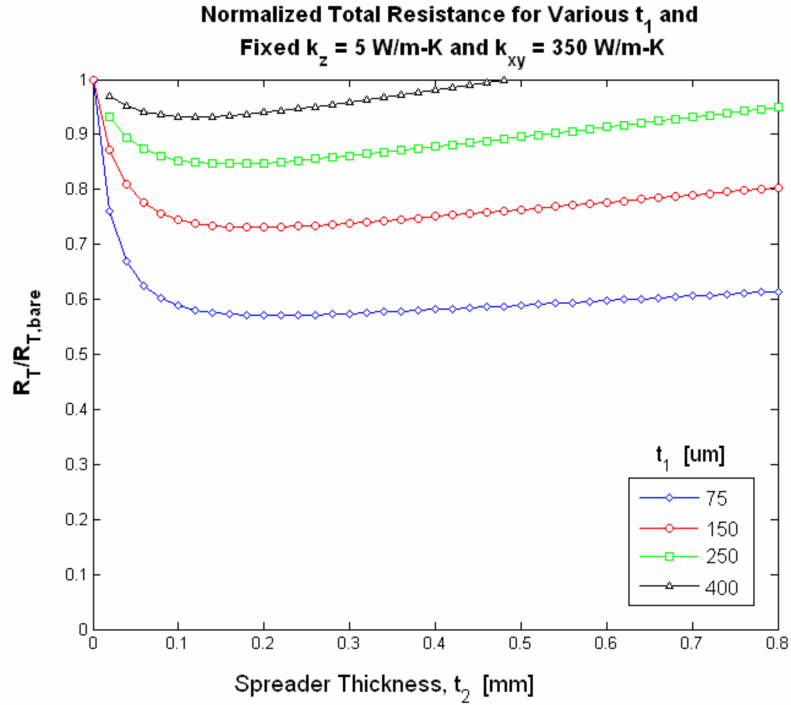


Figure 4-16: Normalized total resistance for different chip thicknesses and an orthotropic spreader with k_z = 5 W/m-K and k_{xy} = 350 W/m-K

Figure 4-16 also reveals that increasing chip thicknesses are accompanied by a steady decrease in the optimum spreader thickness for a given k_{xy}. In order to better understand this variation, the plot in Figure 4-13 was reproduced for different values of t_1. The results can be seen in Figure 4-17, and it is clear that smaller chip thicknesses yield a larger optimum spreader thickness for a given k_{xy}. Also, thicker chips yield lower sensitivity of optimum thickness to in-plane conductivity, as evidenced by the suppression of the peak optimum thickness in Figure 4-17 for larger values of t_1.

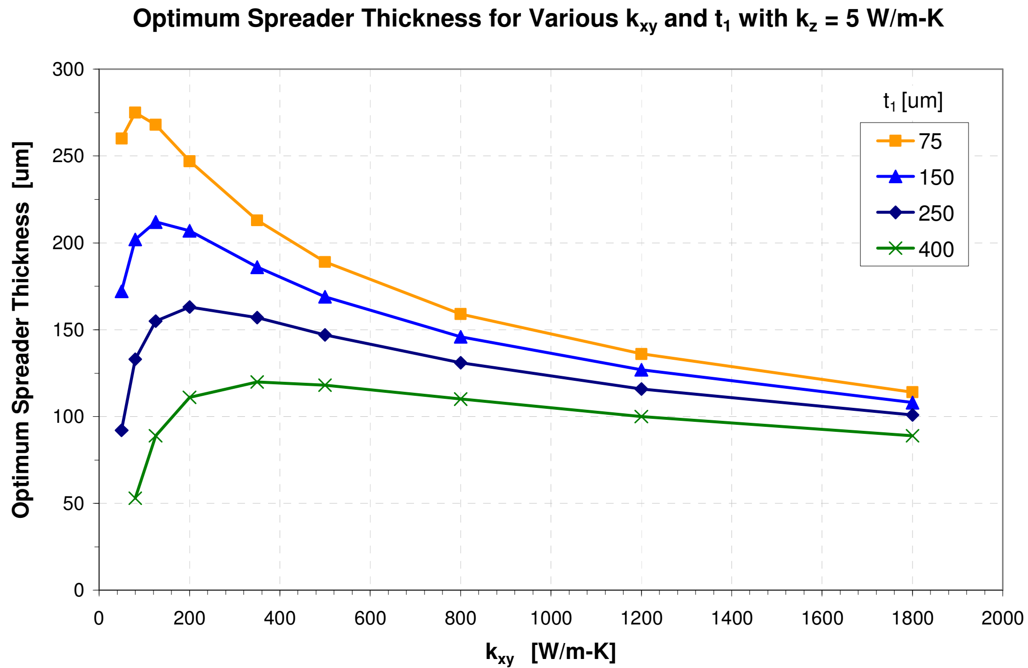


Figure 4-17: Change in optimum spreader thickness for various k_{xy} and t_1

Variation of Flux-Spot Size and Heat Transfer Coefficient

The size and overall power dissipation of a flux-spot are dependent on the intensity of the computational task being performed and the allocation of processing resources on the chip. Thus, the wide variability of 3D chip architectures implies the potential for a myriad of flux-spot sizes and power dissipations to be encountered. Also, the assortment of available cooling approaches – from high end liquid cooling to traditional heat sink/fan combinations – supply a wide range of potential heat transfer coefficients. Both of these parameters – flux-spot size and heat transfer coefficient – have an immediate effect on the temperature achieved at the hotspot. This section investigates the sensitivity of

hotspot temperature to flux-spot size and heat transfer coefficient for the system in Figure 4-5.

Figure 4-18 depicts an example of the temperature rise associated with a 500 x 500 μm hotspot under the influence of a 1.4 kW/cm^2 heat flux on the 1 x 1 cm chip/spreader system of Figure 4-5. For progressively larger flux-spots, the temperature spire will become broader and exhibit flatter and flatter peaks. Eventually, the active chip temperature will approach the uniform value associated with one-dimensional conduction in the limit that the flux-spot size approaches the size of the chip. Alternatively, the temperature spike is expected to vanish as the flux-spot size approaches zero. These two limiting cases provide upper and lower bounds for the hotspot temperature and give a sense of the general shape and magnitude of the temperature rise for different flux-spot sizes.

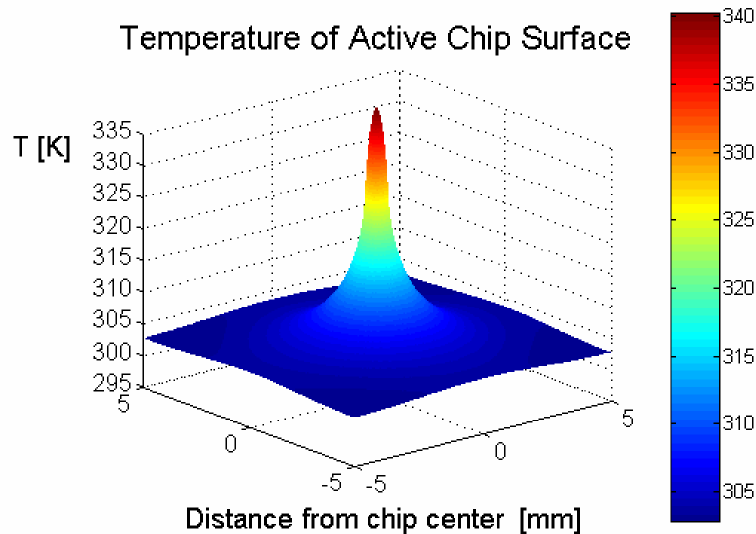


Figure 4-18: Generic 3D plot of the temperature spike on the active side of the chip

The temperature rise associated with very small flux-spots (i.e. those that are small in comparison to the chip dimensions) is dominated by thermal spreading in the silicon, and thus is relatively insensitive to the presence of the spreader and the nature of the applied cooling scheme. Therefore, it can be expected that very small flux-spots will yield thermal behavior akin to that of a flux-spot acting on a semi-infinite slab of silicon. For larger flux-spots, the TIM/spreader characteristics and the applied heat transfer coefficient will exert increasing influence on the hotspot temperature, causing the hotspot temperature to diverge from that predicted by the semi-infinite slab model. The effects of varying spreader and chip properties were examined in the previous section and the effects of varying heat transfer coefficient will be discussed later in this section.

The peak temperature for the classical case of a flux-spot acting on a semi-infinite medium of isotropic conductivity, k_1, is given by the following expression [73]

$$\Delta T_{edge} = \frac{q"w}{k_1\sqrt{\pi}} \quad (47)$$

where q’’ is the applied heat flux, *w* is the length of one side of the square flux-spot and ΔT_{edge} is the difference between the peak temperature and the edge temperature (for a semi-infinite solid the “edge” is assumed to be infinitely far away and have a temperature approaching 0 K). Equation 47 is plotted in Figure 4-19 for flux-spots that are 10 μm – 500 μm one a side with an applied flux of 1.4 kW/cm^2. The peak-to-edge temperature difference is also plotted in Figure 4-19 against *w* for three alternate chip thicknesses in the chip/spreader system defined by Figure 4-5 and Table 4-1 (where k_{xy} = 100 W/m-K for the orthotropic spreader). The edge temperature used in the definition of ΔT_{edge} for the finite case is found by substituting x = 0 and y = L/2 into Equation 37 and adding the bulk fluid temperature (see Figure 4-5 for the coordinate system). As expected, the peak

temperature rise asymptotically approaches the semi-infinite case as the flux-spot size tends toward zero. For larger flux-spots, the finite nature of system, the spreader characteristics, and the applied heat transfer coefficient all exert greater influence on the temperature profile and cause a divergence of ΔT_{edge} from the semi-infinite prediction. Figure 4-19 also confirms the elementary result that, for systems with thicker chips, the temperature rise is well-approximated by the semi-infinite model over a wider range of flux-spot sizes.

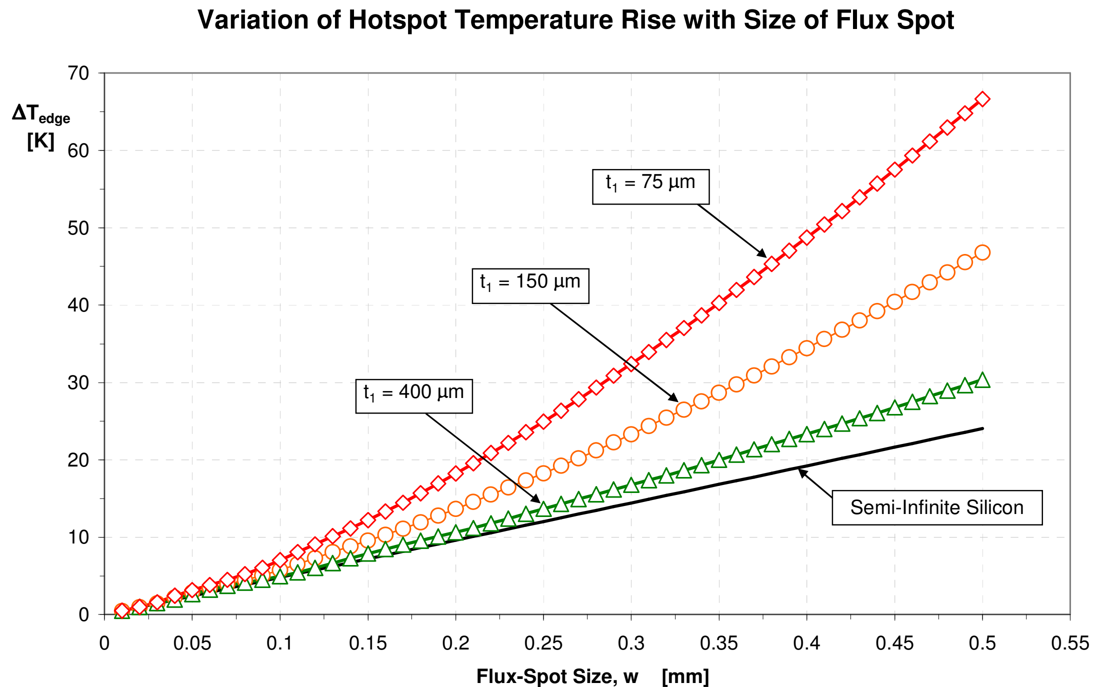


Figure 4-19: Hotspot temperature rise for escalating flux-spot size

The convective boundary condition applied to the back of the spreader in Figure 4-5 represents the presence of a global cooling solution. The magnitude of the effective heat transfer coefficient depends on the nature of the convective heat transfer and

whether or not extended surfaces or other area enhancements are present in the cooling system. It is thus desirable to determine the influence that a change in heat transfer coefficient has on hotspot temperature. Given the form of the expression for one-dimensional resistance in Equation 44, one can easily anticipate that R_{1D} will vary as 1/h ($R_{1D} \rightarrow \infty$ as $h \rightarrow 0$ and $R_{1D} \rightarrow t_1/(k_1 A) + t_2/(k_z A)$ as $h \rightarrow \infty$). However, the dependence of the spreading resistance to a change in heat transfer coefficient is less easily predicted. Therefore, the developed MATLAB codes were used to explore the variation of R_s for heat transfer coefficients ranging from 1,000 to 10,000 W/m^2K. For the conditions in Table 4-1 with k_{xy} = 350 W/m-K and the stated range of heat transfer coefficient, the spreading resistance is found to remain relatively constant, with only a modest decline for increasing heat transfer coefficient. Figure 4-20 shows the total thermal resistance and its decomposition into R_{1D} and R_s for this representative case. As noted, R_s is seen to remain relatively constant wile R_{1D} follows the expected 1/h variation. Given the relative invariance of R_s with h, the variation of the total thermal resistance, and thus the average hotspot temperature, primarily exhibits a 1/h dependence on heat transfer coefficient. Thus, the assumption that R_T varies as 1/h can be used as an excellent first approximation.

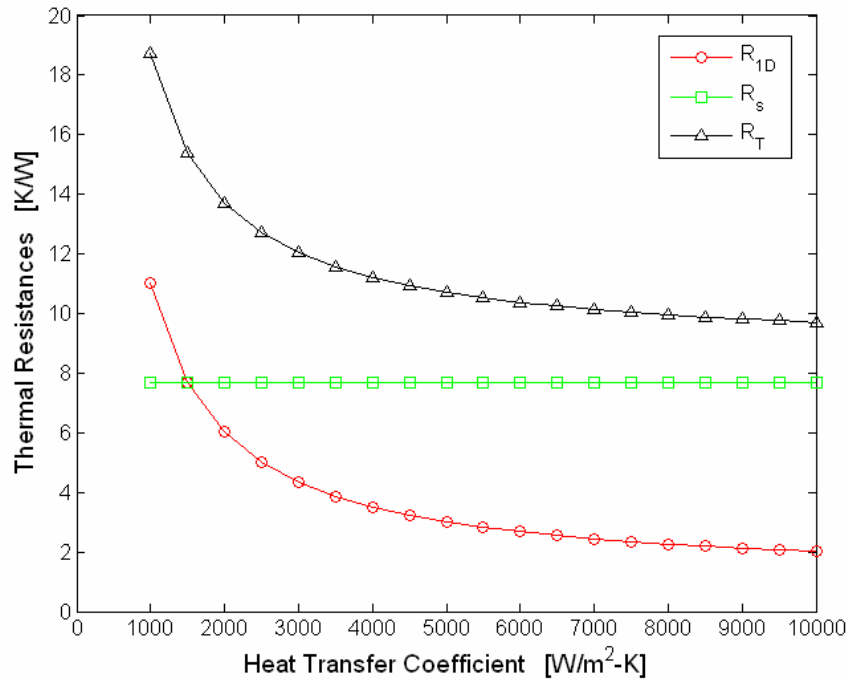


Figure 4-20: Sensitivity of the thermal resistances to varying heat transfer coefficient

Numerical Simulations and Contact Resistance Variation

The parametric results above indicate that use of an orthotropic spreader is a promising approach to reducing hotspot temperatures. However, it may be anticipated that the physical attachment of the orthotropic material to the back of the chip, as depicted in Figure 4-5, may well result in the creation of a potentially significant and deleterious thermal contact resistance. Typical contact resistances for electronic packaging applications are reported to be in the range of 10^{-3} to 1.0 K-cm^2/W [75]. Resistances of ~10^{-3} K-cm^2/W are representative of an excellent interface achieved by monolithic growth or eutectic interface attachment, while resistances of ~1.0 K-cm^2/W represent a relatively poor thermal interface achieved through the use of phase change

materials and elastomeric pads [51]. It is desirable to model the contact resistance at the chip/spreader interface and determine its effect on hotspot remediation; however, to the author's knowledge, there is no analytical solution that explicitly accounts for the contact resistance in a layered structure. Consequently, the parametric sensitivity of hotspot temperature to contact resistance is explored through numerical simulations with ANSYS for contact resistances varying from 0 to 1.0 K-cm^2/W.

A one-quarter symmetry model of the chip/spreader structure shown in Figure 4-5 was created in ANSYS using the parameters listed in Table 4-1. The tetrahedral version of 'SOLID87' – a ten node thermal element with a single degree of freedom (temperature) – was used to discretize both the chip and spreader domains as shown in Figure 4-21. The mesh was refined significantly in the vicinity of the flux-spot in order to better resolve the large thermal gradients that are anticipated in this area. For boundary conditions, an inward heat flux was applied at the flux-spot area and an effective heat transfer coefficient was applied to the top surface of the spreader (all remaining boundaries are adiabatic). The thermal contact resistance enters into the numerical modeling through the use of so-called 'Contact' and 'Target' surface-to-surface contact elements ('TARGE170' and 'CONTA174' were used in this ANSYS model). One side of the chip/spreader interface is assigned the target surface, while the opposing side is declared a contact surface. The appropriate elements are meshed onto the contacting surfaces of each volume and the combined presence of Contact and Target elements forms a so-called 'contact pair.' The nature of the contact between the elements is defined through the specification of certain 'key options' and 'real constants' that toggle particular interface handling characteristics for each of the TARGE170 and

CONTA174 element types [74]. In this case, all mechanical properties of the interface are turned off, and only thermal characteristics are considered. A symmetric surface-to-surface interface with a specified thermal contact conductance was thus created between the chip and spreader volumes. The final mesh consists of 298,645 elements (120,654 elements in the chip volume, 160,654 in the spreader volume, 8,210 target elements, and 9,136 contact elements). Approximately 20 minutes are needed to run a single case on a Windows based operating system with 2 Gb of RAM and a Pentium 4 processor running at 2.66 GHz.

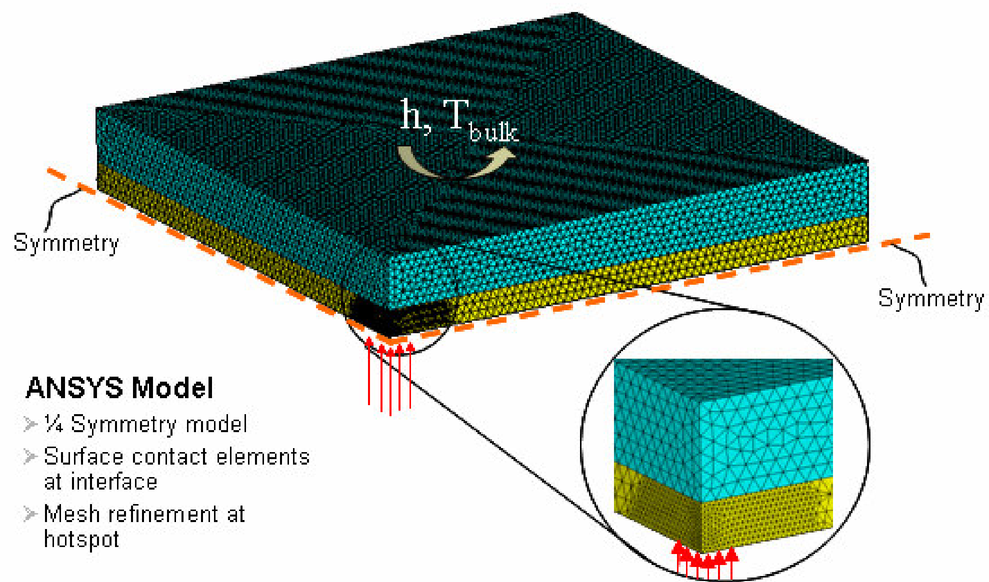


Figure 4-21: Mesh used to evaluate the effects of contact resistance in ANSYS

The ANSYS numerical model is valuable because it makes possible the analysis of a physical situation for which there is no analytical solution. Naturally, the lack of an analytical solution complicates direct validation of the full ANSYS model; however, one might reasonably assume that the full model is accurate if the accuracy of each salient

model feature can be independently validated. The two major features of the ANSYS model in Figure 4-21 are the performance of the contact resistance and the resolution of complex thermal conduction from the flux-spot to the ambient. Analytical solutions are available for each of these salient features and the following paragraphs will be devoted to independently verifying that the contact resistance and thermal spreading portions of the numerical model agree with analytically determined results.

Figure 4-22 depicts the representative two-layer system that was used to validate contact resistance modeling in ANSYS. This model is nearly identical to that of Figure 4-5, except a contact resistance is included between the chip and spreader, and the heat flux is applied uniformly over the entire active surface of the chip. Since the heat flux acts on the entire chip area and the edges are adiabatic, no lateral thermal conduction takes place and the system can be modeled by the simple one-dimensional resistance network shown in Figure 4-22. The resistance network can be solved to find the contact resistance, R_c'', and heat flux, q'', if all other parameters are fixed. The expressions for R_c'' and q'' are given by Equations 48 and 49, respectively, where δT is the interfacial temperature jump:

$$R_c" = \frac{\frac{\delta T\ t_2}{k_2} + \frac{\delta T}{h}}{T_1 - T_{bulk} - \delta T} \tag{48}$$

$$q" = \frac{\delta T}{R_c"} \tag{49}$$

With knowledge of the above expressions for contact resistance and heat flux, one can choose values for δT and T_1 such that it is convenient to verify ANSYS results.

Choosing the convenient values of $\delta T = 10$ K and $T_1 = 330$ K and substituting them into Equations 48 and 49 reveals that a contact resistance of 0.606 K-cm^2/W and a heat flux of 165 kW/m^2 are needed to achieve the stated δT and T_1.

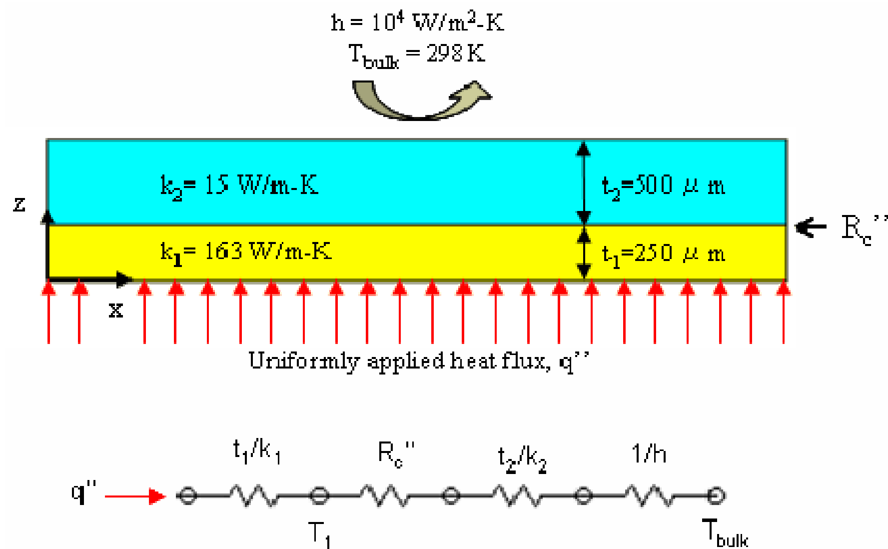


Figure 4-22: Model used to investigate the thermal contact resistance in ANSYS

An ANSYS model was run for the system arrangement in Figure 4-22 with $R_c'' = 0.606$ Kcm^2/W and $q'' = 165$ kW/m^2 to determine if the designed thermal characteristics would appear in the numerical results. It can be seen from the resulting through-thickness temperature distribution in Figure 4-23 that, as designed, the temperature on the chip side of the interface is 330 K and a 10 K temperature jump exists at the interface ($z = 0.25$ mm). The matching between analytical and numerical results for this representative case provides validation that the contact resistance feature in ANSYS performs as expected.

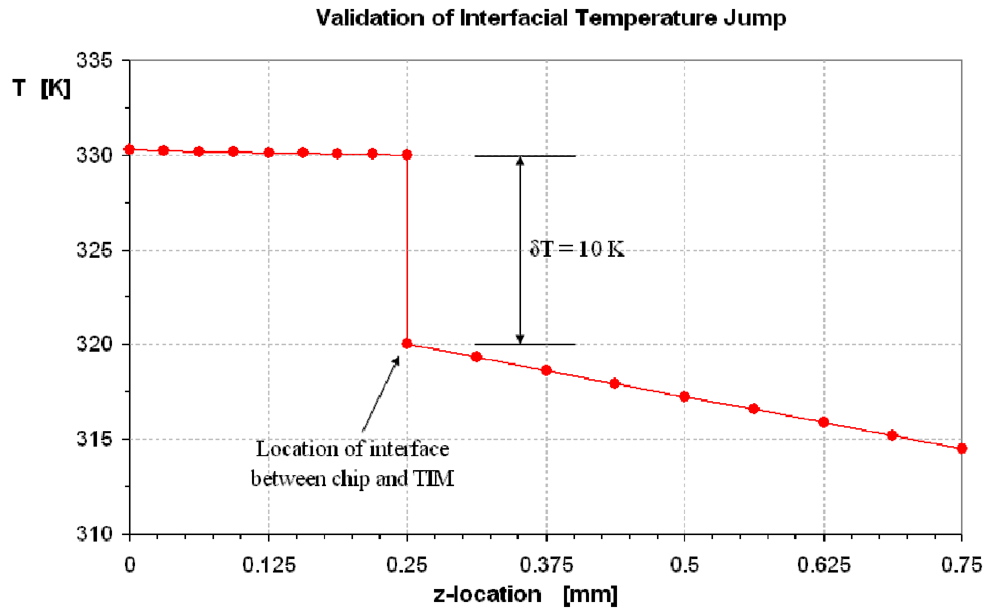


Figure 4-23: ANSYS temperature profile through the thickness of the model in Figure 4-22.

With the knowledge that a contact resistance can be successfully modeled in ANSYS, attention now turns to validation of thermal spreading in the bi-layer chip/spreader system. The analytical solutions discussed in the beginning of this chapter will be used as a benchmark for ANSYS results. The usual baseline model defined by Figure 4-5 and Table 4-1 was created in ANSYS without a thermal contact resistance and the in-plane thermal conductivity of the spreader was varied in an attempt to reproduce the analytical results found on the left-hand side of Figure 4-8.

For low and moderate degrees of spreader anisotropy, excellent matching was found between the chip temperature profiles provided by ANSYS and the temperature profiles determined from the analytical expression in Equation 37. For instance, Figure 4-24 shows good agreement between the numerical and analytical temperature profiles for a spreader with k_z = 5 W/m-K and k_{xy} = 50 W/m-K. Furthermore, the 3D surface

temperature plots in Figure 4-25 show that analytically and numerically determined chip temperatures are in good agreement (to within 0.5%) over the entire active surface of the chip for a spreader with $k_z = 5$ W/m-K and $k_{xy} = 100$ W/m-K.

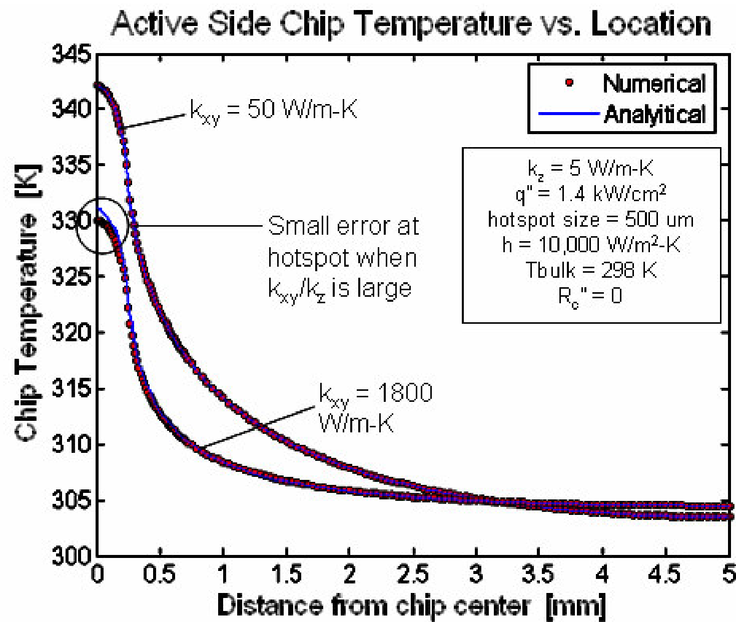


Figure 4-24: Comparison of numerical and analytical temperature profiles on the active side of the chip for various for $k_{xy} = 50$ and $k_{xy} = 1800$

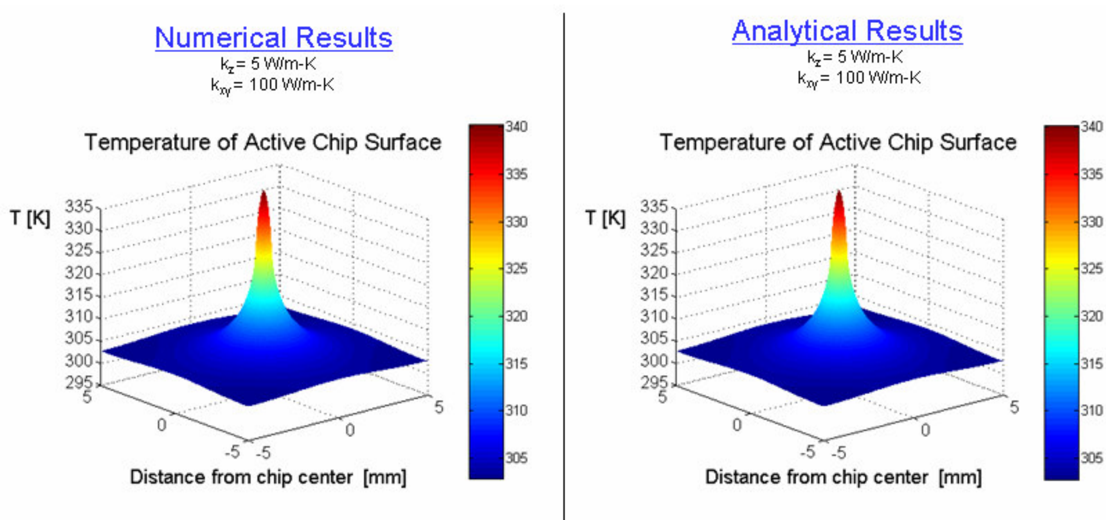


Figure 4-25: Surface temperature plots show good matching for $k_{xy} = 100$ W/m-K

In contrast to the excellent agreement seen at low degrees of anisotropy, highly anisotropic spreaders yield some discrepancy between analytical and numerical results. Referring back to Figure 4-24, the case where k_{xy} = 1800 W/m-K shows some inconsistency between analytical and numerical results, particularly in the vicinity of the hotspot where the ANSYS model tends to underpredict the analytically-derived hotspot temperature. Due to limitations on node count in the version of ANSYS being used (512,000 node maximum), the model's discretization could not be refined well enough to yield complete matching for high degrees of anisotropy. Figure 4-26 shows the peak hotspot temperature predicted by ANSYS approaching the analytical peak hotspot temperature for progressive mesh refinement (k_{xy} for the spreader is 500 W/m-K in this case). The increase in node count was halted at ~415,000 in Figure 4-26 because further mesh refinement pushed the number of nodes beyond the limits of the software. This ~415,000 node mesh is depicted in Figure 4-21 and was used in all subsequent simulations since it provides the nearest agreement to analytical results. Finally, Figure 4-27 shows the results of Figure 4-8 with analytical results overlaid. This further illustrates that the hotspot temperatures predicted by analytical and numerical methods diverge as in-plane spreader conductivity is increased.

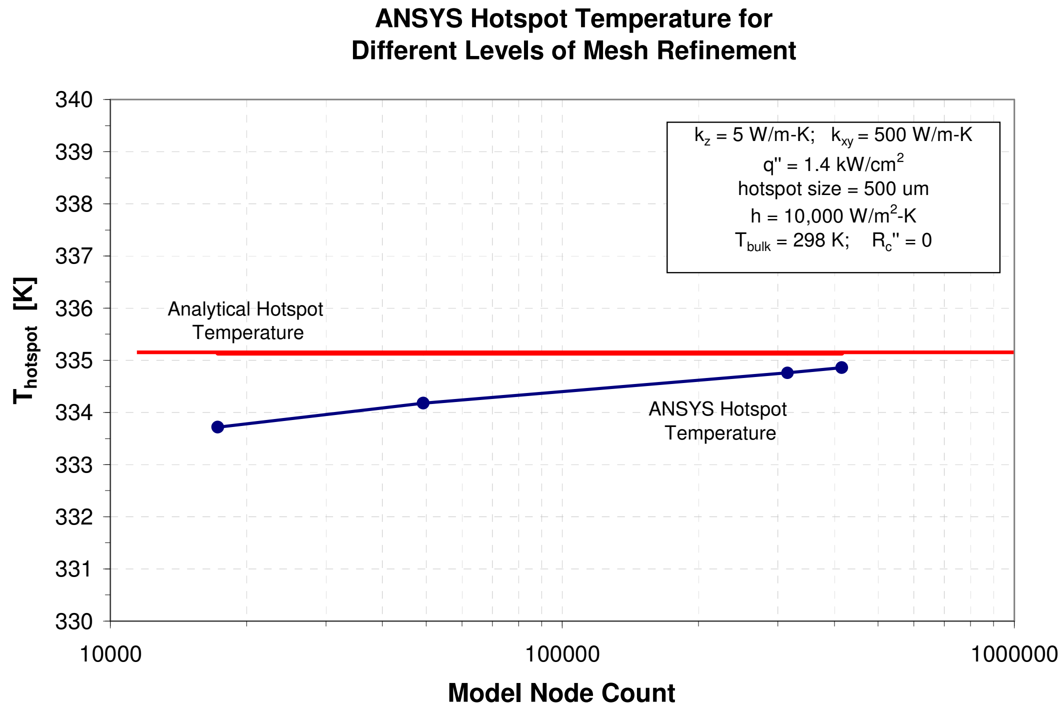


Figure 4-26: ANSYS results approach the analytical predictions for progressive node count

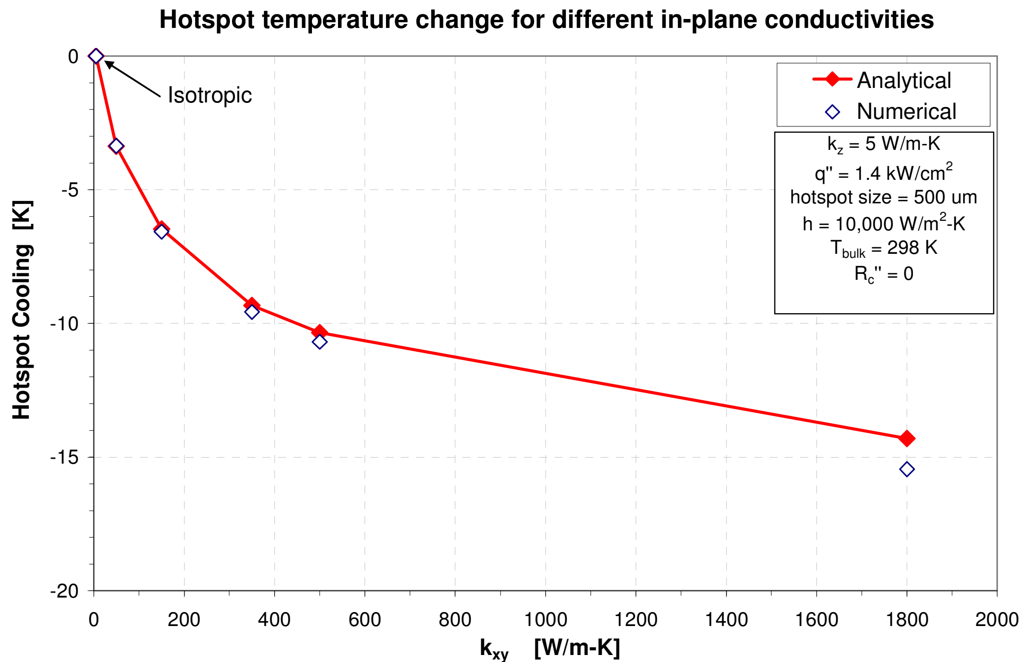


Figure 4-27: Numerical results under-predict hotspot cooling for larger and larger k_{xy}/k_z

With validation that both contact resistance and thermal conduction can be accurately modeled in ANSYS for all but the most extreme cases of anisotropy, one might reasonably assume that the full ANSYS model – which accounts for spreading in the presence of a contact resistance – will be accurate under similar circumstances. In order to obtain a general understanding of how the presence of a contact resistance affects the flow of heat in the system, the full ANSYS model was run for two extreme cases of contact resistance – *i.e.* perfect thermal contact (R_c'' = 0 K-cm^2/W) and very poor thermal contact (R_c'' = 6 K-cm^2/W). The system parameters listed in Table 4-1 were used and the spreader conductivities were taken as k_z = 5 W/m-K and k_{xy} = 500 W/m-K. Figure 4-28 shows the resulting heat flux vectors near the flux-spot in the silicon chip for each contact resistance, with the results for perfect thermal contact on the top and poor thermal contact on the bottom. Comparing the two heat flux plots reveals some subtle differences that indicate the influence of contact resistance on the flow of heat in the system. In the case of perfect thermal contact, the heat flux vectors generally have a larger vertical component, representing the expectation that heat will more readily flow across the interface and into the spreader. Similarly, the heat flux vectors in the case of poor thermal contact exhibit a greater component in the horizontal direction since, under the influence of the contact resistance, heat is required to spread more in the silicon before reaching the orthotropic TIM/spreader. The contact resistance thus acts to more evenly distribute the heat flux imposed on the spreader, thereby reducing the spreader's effectiveness. Ultimately, the presence of the contact resistance results is larger peak and average temperature rises at the flux-spot (these peak and average are 61.9C and 57.7C

for the perfect interface, respectively, with the poor thermal interface resulting in 96.2C and 91.6C peak and average temperatures, respectively).

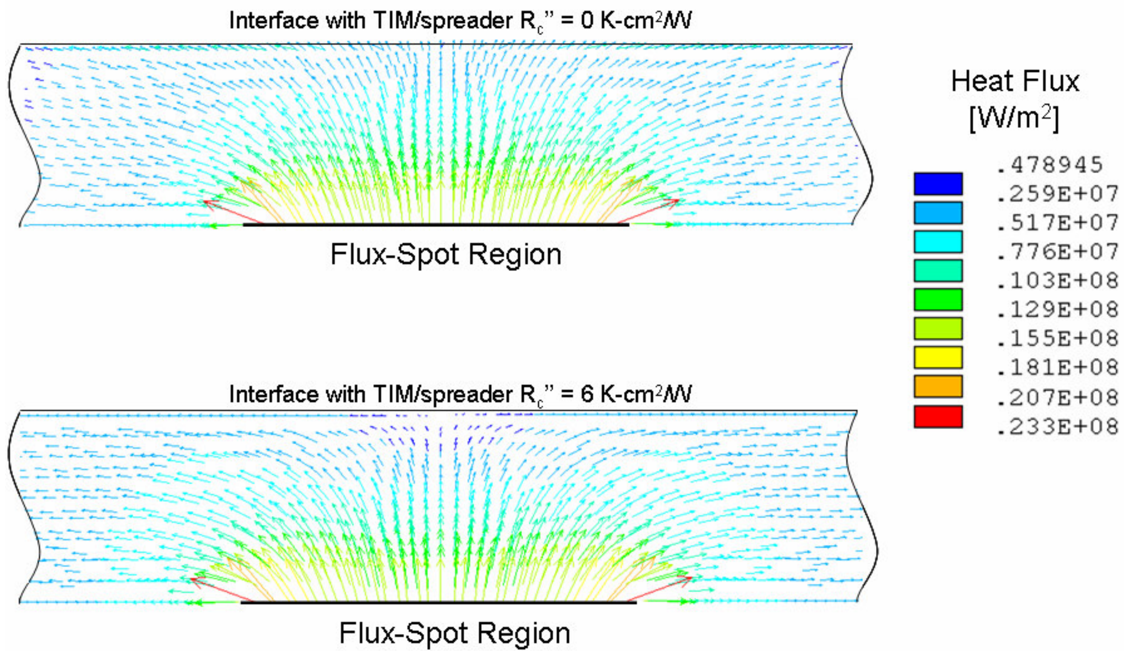


Figure 4-28: Heat flux plots in the silicon for perfect and poor thermal contact

With the expectation that hotspot temperatures should increase for escalating contact resistance, a total of 78 ANSYS simulations were run for contact resistances ranging from 0 K-cm^2/W to 1 K-cm^2/W, with model parameters defined in Table 4-1. Figure 4-29 depicts the increase in hotspot temperature over the stated contact resistance range for differing degrees of spreader anisotropy (the in-plane conductivities shown here are the same as those tested in Figure 4-7). The reader is reminded that 10^{-3} K-cm^2/W is a very low contact resistance that may be achieved by way of monolithic growth on the back of the chip or through the use of soldered interface; alternatively, a poor interface,

such as a loosely pressed interface with a phase change material or elastomeric pad, is represented by a contact resistance near 1 K-cm^2/W.

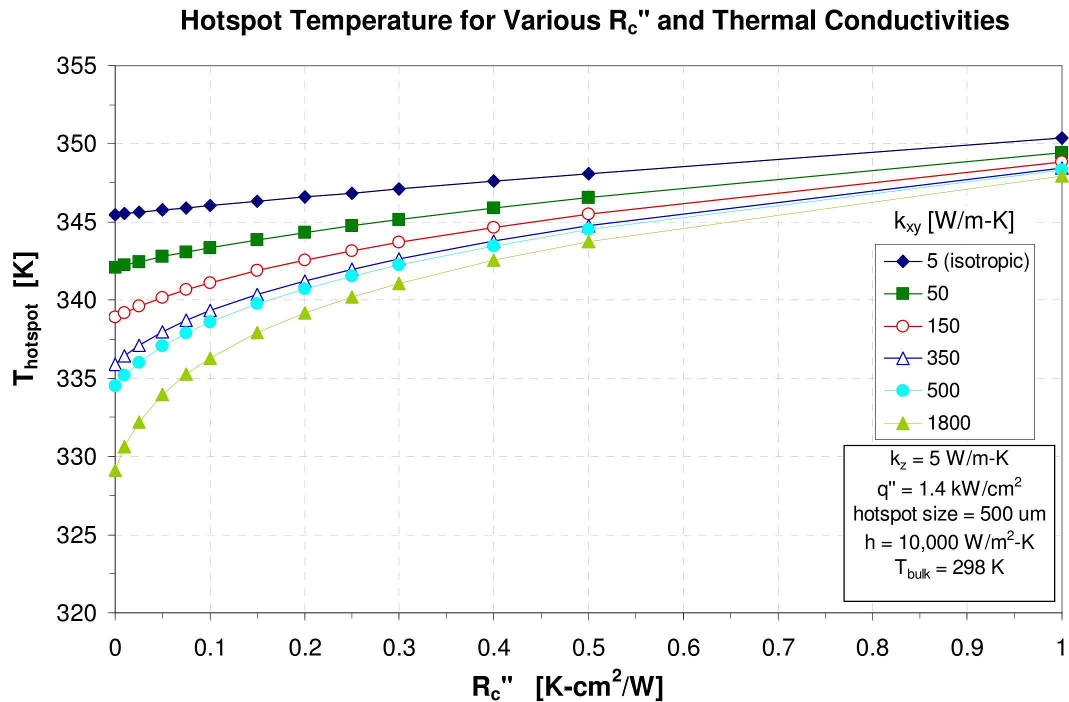


Figure 4-29: Hotspot vs. contact resistance for various k_{xy}

For the conditions studied, it is found that the contact resistance has a significant effect on hotspot temperature, particularly when extreme anisotropy is present in the spreader. In order for a spreader with k_z = 5 W/m-K and k_{xy} = 1800 W/m-K to provide at least 10 K better hotspot cooling than the isotropic spreader, the contact resistance must be made lower than 0.1 K-cm^2/W. However, even for a contact resistance of 0.5 K-cm^2/W, a nearly 5 K temperature reduction can be achieved by the best orthotropic material and 3 or 4 K for more commonly available graphite TIM/spreaders.

It was noted earlier in this chapter that the exceptional hotspot remediation provided by thin CVD diamond films is more attributable to their high average thermal conductivity than to their orthotropic nature. Since it is generally beneficial to implement a spreader with a high average conductivity, and the aim of this chapter is to investigate the effects of orthotropic properties on hotspot remediation, the performance of diamond spreaders has been largely ignored thus far. However, it is worth mentioning here that slightly orthotropic polycrystalline CVD diamond films offer a significant advantage in terms of contact resistance due to their ability to be monolithically grown on the back of the silicon chip. The extremely high thermal conductivity of CVD diamond films, coupled with the potential to minimize contact resistance through monolithic growth on silicon, make diamond films an extremely attractive thermal spreader for this application for hotspot remediation. However, the practical integration of CVD diamond growth into silicon wafer processing is a difficult since high temperatures (~900C) are typically needed. If diamond is grown on a wafer as a final step, the high temperatures could damage existing circuitry and reduce yield. Also, the cost of CVD diamond is often prohibitively high for all but the most advanced commercial and military applications.

Conclusion

The opening portion of this chapter introduced the growing problem of non-uniform on-chip power dissipation in 3D microelectronic systems and reviewed potential methods for cooling the resulting ‘flux-spots.’ When used in the presence of a global cooling solution, anisotropic TIMs/spreaders, bonded directly to the silicon chip, can lower the temperature rise associated with a severe ‘flux spot’ by preferentially conducing heat

laterally away from the flux-spot, toward portions of the chip that are subjected to lower heat flux. The bulk of this chapter served to explore the potential for such anisotropic materials to mitigate hot spots for different geometric parameters and varying degrees of anisotropy. An existing analytical solution for a flux spot on a bi-layer slab was introduced and subsequently used to explore parametric sensitivities of in-plane thermal conductivity, TIM/spreader thickness, chip thickness, flux-spot size, and heat transfer coefficient.

Increasing the in-plane conductivity of an orthotropic spreader with k_z = 5 W/m-K revealed that hot spot temperatures can be suppressed 9.3C and 14.3C below the peak temperature attained for the isotropic spreader ($k_z = k_{xy}$ = 5 W/m-K) with in-plane conductivities of 350 W/m-K and 1800 W/m-K, respectively. Comparison of these results – which are representative of orthotropic natural graphite sheets – to common isotropic spreaders (*e.g.* silicon and copper) showed that better performance can sometimes be attained with the isotropic spreaders. However, the natural graphite orthotropic TIMs/spreaders do provide respectable hot spot remediation and have added the benefits of being relatively low cost and low density. Additionally, highly orthotropic spreaders suppress the temperature variation at the interface with the cooling scheme, which can delay the onset of dryout when direct two-phase cooling is applied.

The interplay between the thickness of othrotropic TIMs/spreaders and the hot spot remediation that they provide is critical in space-constrained 3D chip stacks. As such spreader thickness was varied in order to determine its effect on hot spot reduction. It was found that an optimum thickness, which minimizes the overall thermal resistance, exists above modest degrees of anisotropy. For the conditions examined, highly

orthotropic spreaders are seen to exhibit relatively low optimum thicknesses in comparison to isotropic silicon and copper spreaders. Furthermore, an annealed pyrolitic graphite spreader with k_z = 10 W/m-K and k_{xy} =1700 W/m-K was found to provide better hotspot mitigation than a silicon spreader for thicknesses less than 500 μm and better than a copper spreader for thicknesses less than 200 μm. The exceptional performance of highly orthotropic spreader at very small thicknesses provides advantage over traditional isotropic materials in space-constrained 3D chip stacks.

The variation of chip thickness, hot spot size, and applied heat transfer coefficient were also studied. Parametric variation of the chip thickness revealed that the introduction of an optimally thick orthotropic spreader is less effective for thicker silicon chip since the silicon itself bears more of the spreading burden in this case. In contrast, an optimally thick spreader offers a pronounced reduction of hot spot temperature when implemented on a relatively thin chip. Orthotropic TIMs/spreaders can thus compensate for the loss of inherent thermal spreading that will occur in thinner chips of the future. Varying the flux-spot size, *w*, showed that the temperature rise associated with a heat flux acting on a semi-infinite silicon slab is approached as *w* vanishes. The intuitive result that the hot spot temperature rise approaches the semi-infinite limit more rapidly for increasing chip thickness was also confirmed. Variation of the applied heat transfer coefficient showed that the spreading resistance is relatively insensitive to variation of *h* from 1,000 to 10,000 W/m^2-K. It is therefore concluded that the variation of the total thermal resistance, and thus the average excess hotspot temperature, will very nearly exhibit a 1/*h* variation for the geometric parameters considered.

Finally, the detrimental effects of an interfacial contact resistance between the chip and spreader were evaluated through the use of a validated finite element model. The contact resistance has a deleterious effect on hot spot temperature because it requires that heat be spread more in the silicon chip before reaching the orthotropic TIM/spreader. Simulation results showed that an extremely robust thermal interface (*i.e.* one with contact resistance less than or equal to 10^{-3} K-cm^2/W) had very little effect on the thermal performance of the considered orthotropic TIMs/spreaders. However, as the thermal robustness of the interface deteriorates, so does spreader performance. The deterioration in performance is particularly rapid when the spreader has a high degree of anisotropy. For contact resistances beyond 0.7 K-cm^2/W, a highly orthotropic spreader with $k_z = 5$ W/m-K and $k_{xy} = 1800$ W/m-K provided only slightly better hotspot reduction than an isotropic spreader with $k_z = k_{xy} = 5$ W/m-K. Thus, efforts should be made to reduce the deleterious interfacial contact resistance that arises upon physical attachment of an orthotropic spreader onto the back of a silicon chip.

Chapter 5 : Conclusions

The opening chapter of this thesis served to introduce the interconnect delay bottleneck that is expected to limit the traditional progression of Moore's law through device scaling in planar chips. An alternative three-dimensional chip architecture has been identified which can potentially alleviate these interconnect delay issues while simultaneously allowing for the integration of heterogeneous technologies into a single 3D microsystem. While the benefits of 3D chip integration are clear, there are several obstacles to its broader implementation. In particular, the issue of power dissipation is a major challenge to the development of high performance 3D chip stacks. The well-documented difficulties in cooling future 2D chips will only be exacerbated by 3D architectures in which volumetric power density is increased and non-uniform power dissipation is more severe. Traditional external cooling approaches using heat sink and fan combinations are unlikely to meet the needs of 3D chip stacks since high thermal resistance between internal device layers and the heat sink exist under traditional configurations. The development of novel cooling approaches and accurate thermal modeling procedures is necessary to overcome the acute thermal challenges forecasted for 3D chip stacks. Therefore, this thesis focuses on three relevant topics in the cooling of 3D chip stacks: 1) the determination of effective thermal properties for use in compact thermal models, 2) single phase internal liquid cooling, and 3) hot spot remediation with orthotropic thermal interface materials. It should be noted that this concluding chapter will discuss the results of these analyses in a general way. More thorough and specific discussions can be found in the conclusions at the end of each of chapters 2, 3, and 4.

In light of the extremely complex electronic circuitry engendered by three dimensional system integration, compact thermal models are needed to efficiently assess the viability of different global cooling schemes. The determination of equivalent thermal conductivity is often critical to the development of accurate compact thermal models. A representative 3D chip stack was presented in Chapter 2 along with numerical and approximate analytical methods of determining its effective conductivity. The analytical method consisted of simple one-dimensional conduction paths through different portions of the model, while the numerical simulations involved a more robust process of modeling the 3D circuitry in detail and performing a mesh sensitivity analysis to ensure the mesh-independence of the final results. It was found that the approximate analytical method under-predicted the numerically determined equivalent thermal conductivities considerably. It is thus concluded that a more thorough analytical approach, with a more detailed resolution of the relevant conduction paths, would be necessary to yield accurate equivalent thermal conductivity results. However, analytical resistance networks can become cumbersome to use and time consuming to develop when the circuitry being represented exhibits significant complexity. Numerical simulations also take considerable time to set up, run, and validate, but they will likely yield more accurate results than approximate analytical methods for complex circuits like the representative model considered Chapter 2. Once in hand, these equivalent thermal conductivities, whether analytically or numerically determined, can be applied in compact global models to more efficiently evaluate the performance of various cooling approaches.

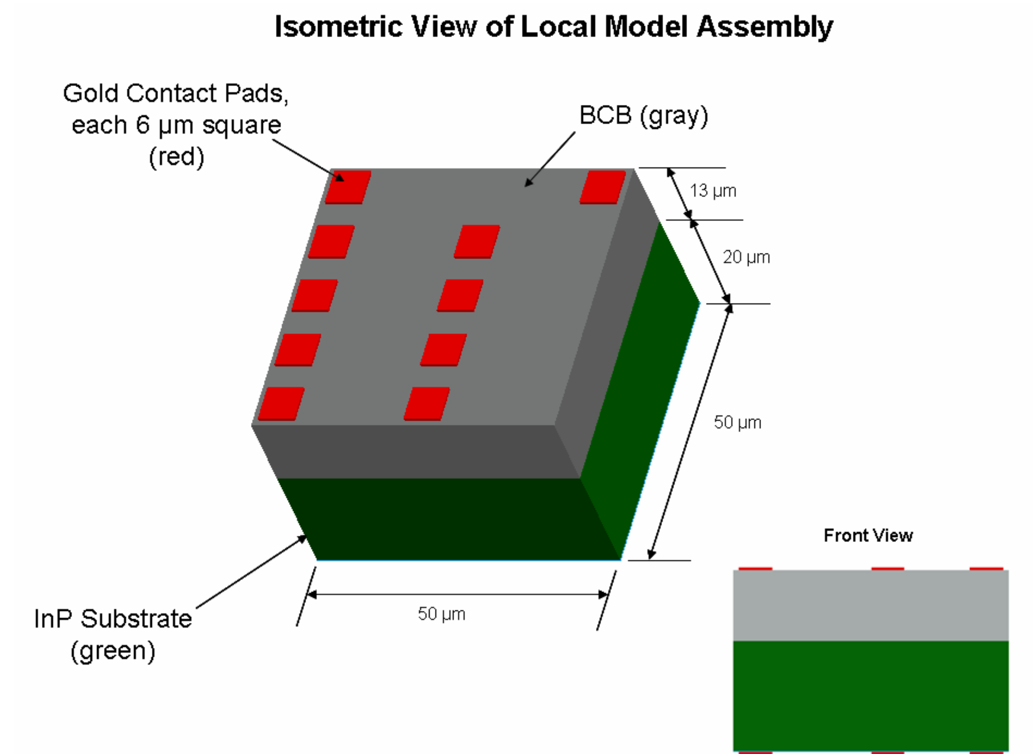
In Chapter 3, application of direct single phase internal liquid cooling with the dielectric liquid, FC-72, was explored for a novel hybrid 3D chip stack through numerical simulations with ANSYS CFX. The hybrid 3D chip stack is comprised of upper and lower portions which are respectively referred to as the top chip stack (TCS) and bottom chip stack (BCS). The TCS portion of the model imposes a significant heat load on the low power dissipation BCS. In the proposed approach, cooling is achieved by flushing FC-72 through channel-like openings in the BCS. Chapter 3 serves to 1) review relevant analytical methods for determining thermal and hydrodynamic behavior in the BCS, 2) develop and validate an approximate 2D model of the thermofluid conditions in the BCS for rapid parametric exploration, and 3) to develop a computationally demanding 3D model of the BCS in order to determine the flow conditions under which device temperatures could be reduced to the nominal 90-110C maximum operating temperature range for microelectronics. It was shown developing flow conditions will prevail in the BCS for reasonable inlet velocities (*i.e.* above 1 m/s) due to the small channel dimensions. The 2D model simulations – which represented an unrealistic flow pattern in the BCS but allowed for rapid variation of system parameters – revealed that increasing the fluid inlet velocity and enhancing the thermal conductivity of the BCS material are two beneficial approaches to reducing device temperatures. With these results as a guide, more accurate and resource intensive 3D simulations were run to determine the true BCS temperatures. It was found that inlet velocities of 20 m/s and an equivalent BCS 'slice' thermal conductivity of greater than 1200 W/m-K are needed to reduce device temperatures to less than 100C. An inlet velocity of 20 m/s results in a pressure drop of ~2.4 atm, and the

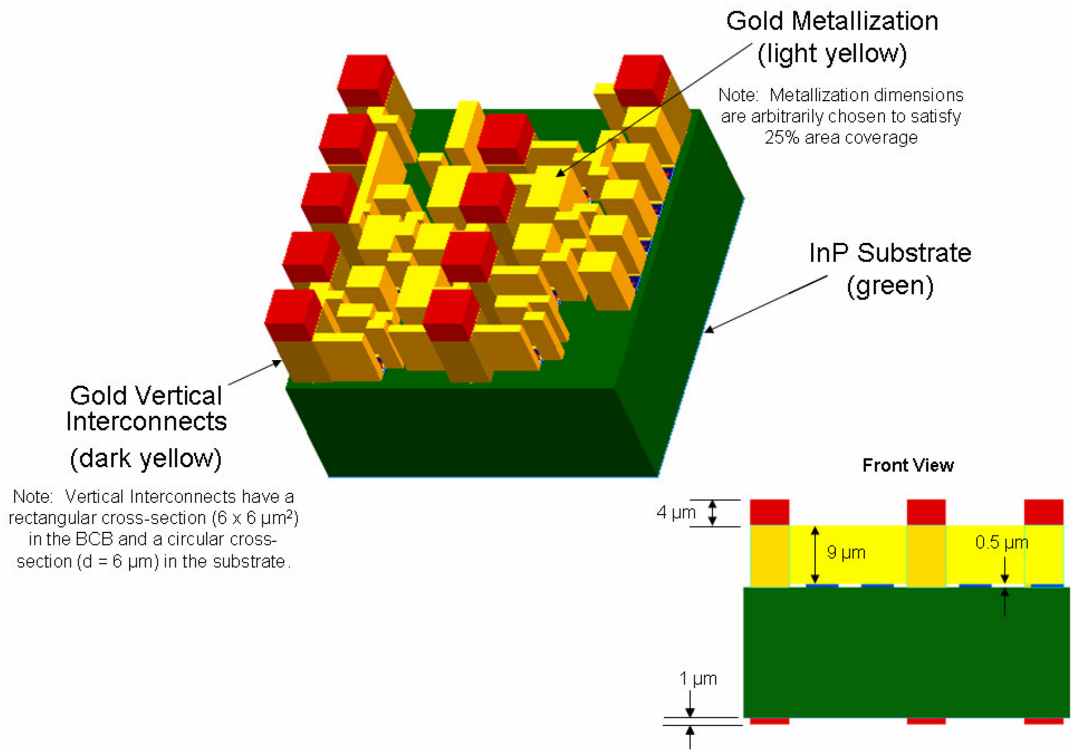
enhanced BCS thermal conductivity of the can be achieved through the implementation of a high conductivity material, such as CVD diamond, on the silicon slices in the BCS.

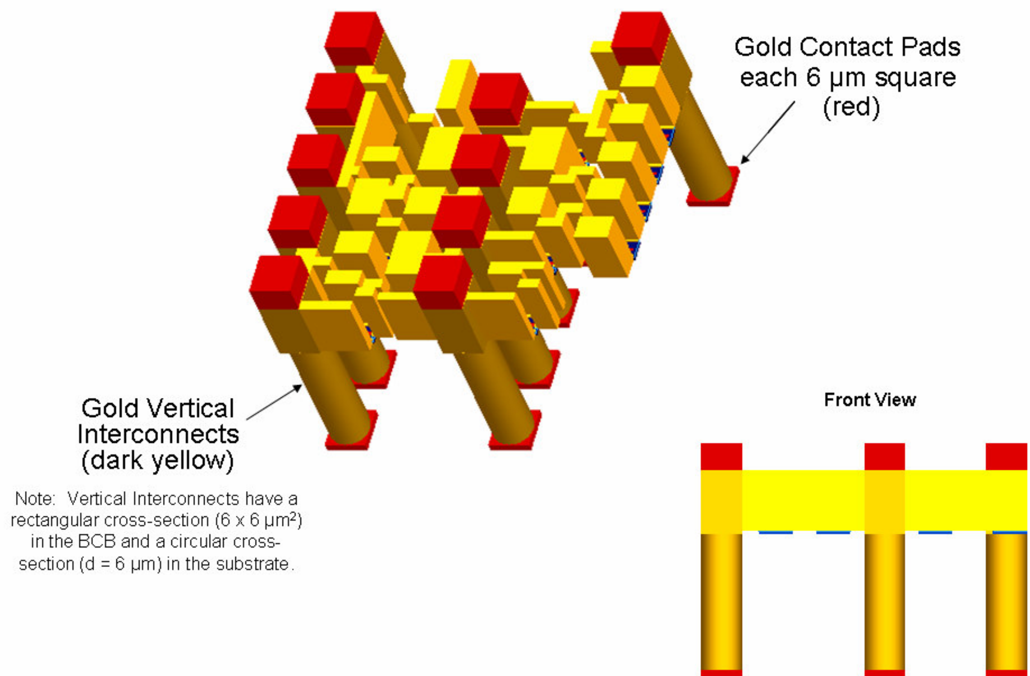
The concept that enhanced spreading in the BCS could reduce the detrimental effects of a localized high heat flux was continued in Chapter 4, in which the potential for anisotropic spreading materials to mitigate the detrimental high temperatures associated with non-uniform on-chip power dissipation (*i.e.* 'flux-spots') was explored. When used in the presence of a global cooling solution, anisotropic TIMs/spreaders, bonded directly to the silicon chip, can lower the temperature rise associated with a severe 'flux spot' by preferentially conducing heat laterally away from the flux-spot, toward portions of the chip that are subjected to lower heat flux. In Chapter 4 the hotspot remediation provided by anisotropic spreaders was quantified for different geometric parameters and varying degrees of spreader anisotropy. An existing analytical solution for a flux spot on a bi-layer slab was introduced and subsequently used to explore parametric sensitivities of in-plane thermal conductivity, TIM/spreader thickness, chip thickness, flux-spot size, and heat transfer coefficient. Later, a validated numerical model was used to determine the effects of a contact resistance at the chip/spreader interface. It was found that highly orthotropic natural graphite spreaders (with k_{xy} = 1800 W/m-K and k_z = 5 W/m-K) can achieve 14.3C more hot spot cooling than an isotropic spreader ($k_z = k_{xy}$ = 5 W/m-K) of the same thickness for a chip subjected to a 500 x 500 μm flux-spot dissipating 1.4 kW/cm^2. Comparison of orthotropic spreaders and high conductivity isotropic spreaders such as silicon and copper revealed that highly orthotropic spreaders can yield better hot spot remediation, particularly at low spreader thicknesses. The orthotropic graphite materials have the additional benefits of being more cost effective than increasing silicon

chip thickness, and being far less dense than both silicon and copper. It was also found that highly orthotropic spreaders act to isothermalize the temperature at the interface of the spreader and the global cooling scheme. These benefits along with the exceptional performance of highly anisotropic spreaders at small thicknesses may lead anisotropic spreaders to be favored over traditional isotropic spreaders in space-constrained 3D chip stacks where portability is a top concern. However, as with any spreader, the physical attachment of an orthotropic TIM/spreader to a silicon chip can result in the formation of a deleterious interfacial contact resistance. This effect of various contact resistances was explored with a numerical model created in ANSYS. For the conditions considered, it was found that contact resistances below 0.05 K-cm^2/W yield small performance losses, while a contact resistance of >0.7 K-cm^2/W significantly reduce the hotspot remediation provided by even highly orthotropic spreaders. Efforts should thus be made to reduce the severity of interfacial thermal contact resistances during assembly.

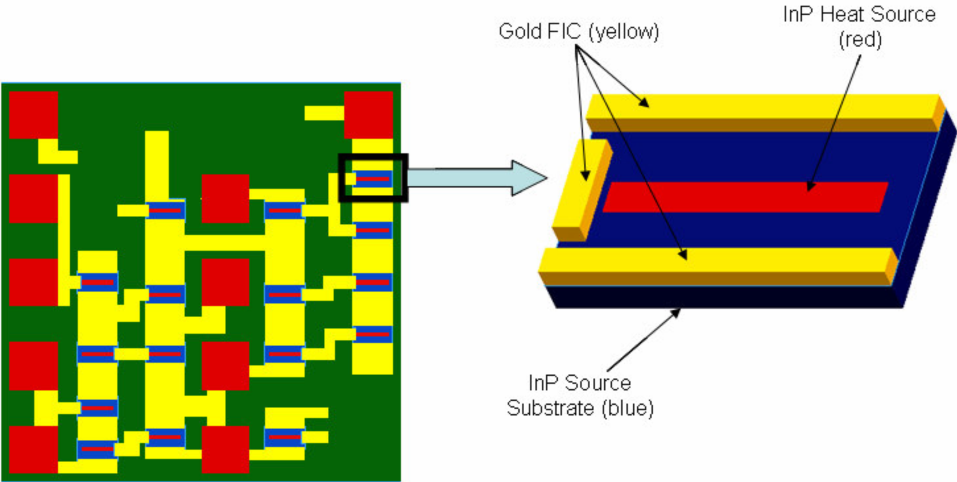
Appendix A: Detailed Local Model Drawings for Chapter 2

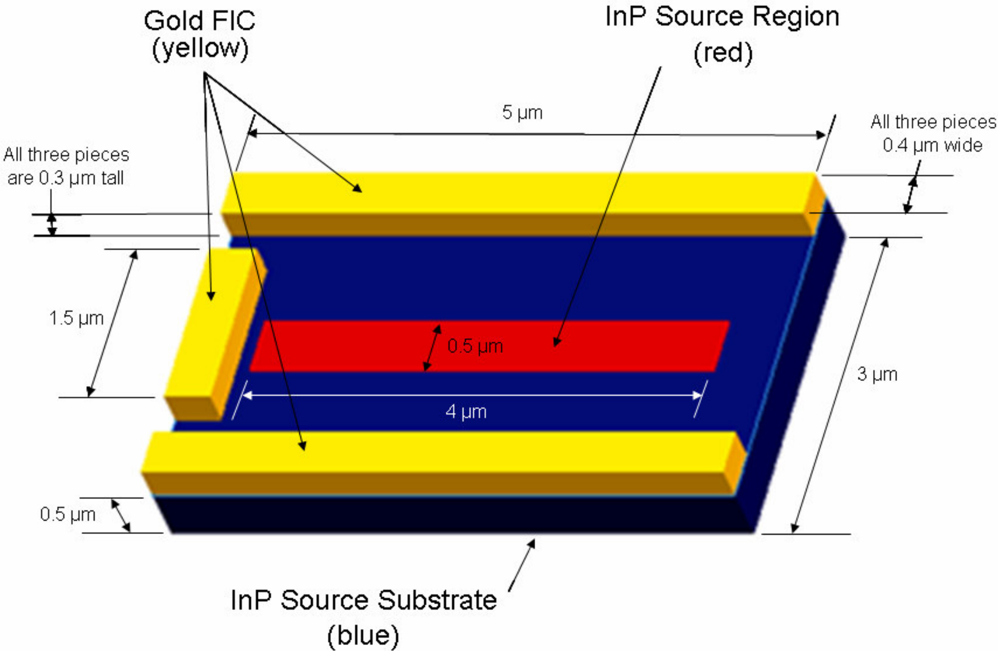



View with BCB Hidden
Gold Metallization
(light yellow)
Note: Metallization dimensions
are arbitrarily chosen to satisfy
25% area coverage
InP Substrate
(green)
Gold Vertical
Interconnects
(dark yellow)
Note: Vertical Interconnects have a
rectangular cross-section (6 x 6 μm2)
in the BCB and a circular cross-
section (d = 6 μm) in the substrate.
Front View
4 μm
9 μm
0.5 μm
1 μm


View with BCB & Substrate Hidden
Gold Contact Pads
each 6 μm square
(red)
Gold Vertical
Interconnects
(dark yellow)
Note: Vertical Interconnects have a
rectangular cross-section (6 x 6 μm2)
in the BCB and a circular cross-
section (d = 6 μm) in the substrate.
Front View

•


Top View with BCB Hidden
Gold FIC (yellow)
InP Heat Source
(red)
InP Source
Substrate (blue)


Detailed View of an Active Device
Gold FIC
(yellow)
InP Source Region
(red)
5 µm
All three pieces
0.4 µm wide
All three pieces
are 0.3 µm tall
1.5 µm
0.5 µm
4 µm
3 µm
0.5 µm
InP Source Substrate
(blue)

Appendix B: MATLAB Code for Determining Hotspot Temperature

```
clear
clc

%-------------------------------------------
% Chip Dimensions and Thermal Characteristics
%-------------------------------------------
a = 1e-2;                          %Length of a side of the rectangular system, m
b = a;                             %Length of the other side, m
t_c = 250e-6;                      %Thickness of the chip, m
k_c_perp = 163;                    %Thermal conductivity perpendicular to the plane of the chip, W/m-
K
k_c_par = 163;                     %Thermal conductivity parallel to the plane of the chip, W/m-K
k_c_e = (k_c_perp*k_c_par)^(1/2);    %Effective thermal conductivity of the chip, W/m-K
t_c_e = t_c/(k_c_perp/k_c_par)^(1/2);    %Effective thickness of the chip, W/m-K



%------------------------------------------------
% Spreader Dimensions and Thermal Characteristics
%------------------------------------------------
t_s = 500e-6;                      %Thickness of the spreader, m
k_s_perp = 5;                      %Thermal conductivity perpendicular to the plane of the spreader,
W/m-K
k_s_par = 350;                     %Thermal conductivity parallel to the plance of the spreader, W/m-
K
k_s_e = (k_s_perp*k_s_par)^(1/2);    %Effective thermal conductivity of the spreader, W/m-K
t_s_e = t_s/(k_s_perp/k_s_par)^(1/2);    %Effective thickness of the spreader, W/m-K



%---------------------------------------------------------
% Hotspot Location, Dimensions, and Thermal Characteristics
%---------------------------------------------------------
%% Please note that the origin is located at the bottom left corner of the
%% rectangle.  The dimension, b, extends in the y-direction and the
%% dimension, a, extends in the x-direction

y_c = b/2;                         %y-location of hotspot, m
x_c = a/2;                         %x-location of the hotspot, m
c = 500e-6;                        %Extent of rectangular hotspot in the x-direction, m
d = c;                   %Extent of rectangular hotspot in the y-direction, m
q_flux = 1.4e7;                    %Heat flux at the hotspot, W/m^2
Q = q_flux*(c*d);                  %Heat rate at hotspot, W

%------------------------------------------------
% System Constants (heat transfer coeff included)
%------------------------------------------------
x = x_c;                           %x-location at which the temperature is desired, m
y = y_c;                           %y-location at which the temperature is desired, m
T_f = 298;                         %Bulk temperature of the convective fluid, K
h = 10000;                         %Heat transfer coefficient on the back of the spreader, W/m^2-K
kappa = k_s_e/k_c_e;                    %Parameter used to simplify later expressions, dimensionless
```

```
alpha = (1-kappa)/(1+kappa);          %Parameter used to simplify later expressions,
dimensionless
A_o = (Q/(a*b))*(t_c_e/k_c_e+t_s_e/k_s_e+1/h);  %The first Fourier Coefficient, based on 1-D
conduction, K
term1 = A_o;                          %This is the first term of the solution



%---------------------------
% Iterations for Summations
%---------------------------
%% Iteration control
m_max = 1000;                         %Number of iterations done on summations over m
n_max = m_max;                          %Number of iterations done on summations over n

%% Second Term of Solution
term2m = 0;
for m = 1:m_max
   term2=term2m;
   lambda2 = m*pi/a;                           %The eigenvalue, lambda
   stigma2 = (lambda2+h/k_s_e)/(lambda2-h/k_s_e);   %greek letter used by Yovanovich, et al.

   phi2 = (alpha*exp(4*lambda2*t_c_e) -
exp(2*lambda2*t_c_e)+stigma2*(exp(2*lambda2*(2*t_c_e+t_s_e)) -
alpha*exp(2*lambda2*(t_c_e+t_s_e))))/...
        (alpha*exp(4*lambda2*t_c_e) +
exp(2*lambda2*t_c_e)+stigma2*(exp(2*lambda2*(2*t_c_e+t_s_e)) +
alpha*exp(2*lambda2*(t_c_e+t_s_e))));

   A_m2 = (2*Q*(sin((2*x_c+c)/2*lambda2)-...
             sin((2*x_c-c)/2*lambda2)))/...
            (a*b*c*k_c_e*lambda2^2*phi2);       %Fourier Coefficients for the second term
          if isnan(A_m2)
            term2m = term2;
            break
          end
  term2m = A_m2*cos(lambda2*x)+term2;
end

%% Third Term of the Solution
term3n = 0;
for n = 1:n_max
   term3 = term3n;
   delta3 = n*pi/b;                           %The eigenvalue, delta
   stigma3 = (delta3+h/k_s_e)/(delta3-h/k_s_e);    %greek letter used by Yovanovich, et al.

   phi3 = (alpha*exp(4*delta3*t_c_e)-
exp(2*delta3*t_c_e)+stigma3*(exp(2*delta3*(2*t_c_e+t_s_e))-
alpha*exp(2*delta3*(t_c_e+t_s_e))))/...

(alpha*exp(4*delta3*t_c_e)+exp(2*delta3*t_c_e)+stigma3*(exp(2*delta3*(2*t_c_e+t_s_e))+alpha*
exp(2*delta3*(t_c_e+t_s_e))));

   A_n3 = (2*Q*(sin((2*y_c+d)/2*delta3)-...
             sin((2*y_c-d)/2*delta3)))/...
```

```
            (a*b*d*k_c_e*delta3^2*phi3);       %Fourier Coefficients for the third term
          if isnan(A_n3)
            term3n = term3;
            break
          end
   term3n = A_n3*cos(delta3*y)+term3;
end

%% Fourth Term of the Solution
term4mn = 0;
for m = 1:m_max
   term4n = 0;
   term4m = term4mn;
   lambda4 = m*pi/a;                             %The eigenvalue, lambda
      for n = 1:n_max
         term4 = term4n;
         delta4 = n*pi/b;                        %The eigenvalue, delta
         beta4 = (lambda4^2+delta4^2)^(1/2);     %The eigenvalue, beta
         stigma4 = (beta4+h/k_s_e)/(beta4-h/k_s_e);    % greek letter used by Yovanovich, et al.

         phi4 = (alpha*exp(4*beta4*t_c_e) -
exp(2*beta4*t_c_e)+stigma4*(exp(2*beta4*(2*t_c_e+t_s_e)) -
alpha*exp(2*beta4*(t_c_e+t_s_e))))/...
               (alpha*exp(4*beta4*t_c_e) +
exp(2*beta4*t_c_e)+stigma4*(exp(2*beta4*(2*t_c_e+t_s_e)) +
alpha*exp(2*beta4*(t_c_e+t_s_e))));

         A_mn4 = (16*Q*cos(lambda4*x_c)*sin(lambda4*c/2)*cos(delta4*y_c)*sin(delta4*d/2))/...
               (a*b*c*d*k_c_e*beta4*lambda4*delta4*phi4);      %Fourier Coefficients for the third
term
         if isnan(A_mn4)
            term4n = term4;
            break
         end
         term4n = A_mn4*cos(lambda4*x)*cos(delta4*y)+term4;
      end
   term4mn = term4n+term4m;
end
term1;
term2m;
term3n;
term4mn;
sum = term1+term2m+term3n+term4mn;
T_hot = sum+T_f
```

Appendix C: MATLAB Code for Varying Spreader Thickness

```
clear
clc

%-------------------------------------------
% Chip Dimensions and Thermal Characteristics
%-------------------------------------------
a = 1e-2;                          %Length of a side of the rectangular system, m
b = a;                             %Length of the other side, m
t_c = 400e-6;                      %Thickness of the chip, m
k_c_perp = 163;                    %Thermal conductivity perpendicular to the plane of the chip, W/m-
K
k_c_par = 163;                     %Thermal conductivity parallel to the plane of the chip, W/m-K
k_c_e = (k_c_perp*k_c_par)^(1/2);    %Effective thermal conductivity of the chip, W/m-K
t_c_e = t_c/(k_c_perp/k_c_par)^(1/2);    %Effective thickness of the chip, W/m-K



%------------------------------------------------
% Spreader Dimensions and Thermal Characteristics
%------------------------------------------------
t_s = [0e-6:40e-6:800e-6];           %Thickness of the spreader, m [1e-6:100e-6:5000e-6]
k_s_perp = 400;                    %Thermal conductivity perpendicular to the plane of the spreader,
W/m-K
k_s_par = 400;                     %Thermal conductivity parallel to the plance of the spreader, W/m-
K
k_s_e = (k_s_perp*k_s_par)^(1/2);    %Effective thermal conductivity of the spreader, W/m-K
t_s_e = t_s/(k_s_perp/k_s_par)^(1/2);    %Effective thickness of the spreader, W/m-K



%------------------------------------------------
% Hotspot Dimensions and Thermal Characteristics
%------------------------------------------------
c = 500e-6;                        %Extent of rectangular hotspot in the x-direction, m
d = 500e-6;                        %Extent of rectangular hotspot in the y-direction, m
q_flux = 1.4e7;                    %Heat flux at the hotspot, W/m^2
Q = q_flux*(c*d);                  %Heat rate at hotspot, W

%------------------------------------------------
% System Constants (heat transfer coeff included)
%------------------------------------------------
T_f = 298;                         %Bulk temperature of the convective fluid, K
h = 10000;                         %Heat transfer coefficient on the back of the spreader, W/m^2-K
kappa = k_s_e/k_c_e;               %Parameter used to simplify later expressions, dimensionless
alpha = (1-kappa)/(1+kappa);          %Parameter used to simplify later expressions,
dimensionless

%--------------------------
% Iterations for Summations
%--------------------------
%% Iteration control
m_max = 1000;                      %Number of iterations done on summations over m
```

```
n_max = m_max;                      %Number of iterations done on summations over n

for i = 1:length(t_s_e)
%% First Term of Solution
term1m = 0;
for m = 1:m_max
   term1=term1m;
   delta1=m*pi/(a/2);
   stigma1 = (delta1+h/k_s_e)/(delta1-h/k_s_e);   %greek letter used by Yovanovich, et al.

   phi1 = (alpha*exp(4*delta1*t_c_e) +
exp(2*delta1*t_c_e)+stigma1*(exp(2*delta1*(2*t_c_e+t_s_e(i))) +
alpha*exp(2*delta1*(t_c_e+t_s_e(i)))))/...
        (alpha*exp(4*delta1*t_c_e) -
exp(2*delta1*t_c_e)+stigma1*(exp(2*delta1*(2*t_c_e+t_s_e(i))) -
alpha*exp(2*delta1*(t_c_e+t_s_e(i)))));

   term1m = (1/(2*(c/2)^2*(a/2)*(b/2)*k_c_e)*((sin(c/2*delta1))^2)*phi1/delta1^3)+term1;
   if isnan(term1m)
      term1m = term1;
      break
   end
end



%% Second Term of Solution
term2m = 0;
for m = 1:m_max
   term2=term2m;
   lambda2 = m*pi/(b/2);                   %The eigenvalue, lambda
   stigma2 = (lambda2+h/k_s_e)/(lambda2-h/k_s_e);   %greek letter used by Yovanovich, et al.

   phi2 = (alpha*exp(4*lambda2*t_c_e) +
exp(2*lambda2*t_c_e)+stigma2*(exp(2*lambda2*(2*t_c_e+t_s_e(i))) +
alpha*exp(2*lambda2*(t_c_e+t_s_e(i)))))/...
        (alpha*exp(4*lambda2*t_c_e) -
exp(2*lambda2*t_c_e)+stigma2*(exp(2*lambda2*(2*t_c_e+t_s_e(i))) -
alpha*exp(2*lambda2*(t_c_e+t_s_e(i)))));

   term2m = (1/(2*(d/2)^2*(a/2)*(b/2)*k_c_e)*((sin(d/2*lambda2))^2)*phi2/lambda2^3)+term2;
   if isnan(term2m)
      term2m = term2;
      break
   end
end

%% Third Term of the Solution
term3mn = 0;
for m = 1:m_max
   term3n = 0;
   term3m = term3mn;
   delta3 = m*pi/(a/2);                    %The eigenvalue, lambda
      for n = 1:n_max
         term3 = term3n;
```

```
        lambda3 = n*pi/(b/2);                    %The eigenvalue, delta
        beta3 = (delta3^2+lambda3^2)^(1/2);          %The eigenvalue, beta
        stigma3 = (beta3+h/k_s_e)/(beta3-h/k_s_e);   %greek letter used by Yovanovich, et al.

        phi3 = (alpha*exp(4*beta3*t_c_e) +
exp(2*beta3*t_c_e)+stigma3*(exp(2*beta3*(2*t_c_e+t_s_e(i))) +
alpha*exp(2*beta3*(t_c_e+t_s_e(i)))))/...
                (alpha*exp(4*beta3*t_c_e) -
exp(2*beta3*t_c_e)+stigma3*(exp(2*beta3*(2*t_c_e+t_s_e(i))) -
alpha*exp(2*beta3*(t_c_e+t_s_e(i)))));

        term3n =
(1/((c/2)^2*(d/2)^2*(a/2)*(b/2)*k_c_e)*(sin(c/2*delta3))^2*(sin(d/2*lambda3))^2*phi3/(delta3^2*lam
bda3^2*beta3))+term3;
        if isnan(term3n)
           term3n=term3;
           break
        end
     end
   term3mn = term3n+term3m;
end
term1;
term2m;
term3mn;
k_s_e;
t_s_e;
R_s(i) = term1+term2m+term3mn;
R_1D(i) = t_c/(k_c_perp*a*b)+t_s(i)/(k_s_perp*a*b)+1/(h*a*b);
R_total(i) = R_s(i)+R_1D(i);
T_bar(i) = R_total(i)*Q+T_f-273;
combine(i,1) = t_s(i);
combine(i,2) = R_total(i);
combine(i,3) = R_total(i)*Q;
end
R_bare = 8.69509;
plot(1e3*t_s,R_total,'color','g','Marker','o')    %,'Marker','o'
hold on
sorted = sortrows(combine,2);
1e6*sorted(1,1)
sorted(1,3)
```

References

[1] Meindl, James, "Low Power Microelectronics: Retrospect and Prospect," Proceedings of the IEEE, Vol 83, No. 4, April 1995.

[2] "Moore's Law: Made Possible by Intel Technology", website: http://www.intel.com/technology/mooreslaw/index.htm, last visited: 7/26/07.

[3] "Moore's Law Turns 40", website: http://news.com.com/Images+Moores+Law+turns+40/2009-1041_3-5649019-6.html?tag=st.num, Last accessed: 7/26/07.

[4] "Meet the World's First 45 nm Processor," website: http://www.intel.com/technology/silicon/45nm_technology.htm, last accessed: 7/26/07.

[5] Bohr, Mark. "Interconnect Scaling – The Real Limiter to High Performance ULSI," IEEE International Electron Devices Meeting, pp. 241-244, 1995.

[6] Davis, Jeffrey, R. Venkatesan, et al., "Interconnect Limits on Gigascale Integration (GSI) in the 21st Century," Proceedings of the IEEE, Vol. 89, No. 3, March 2001.

[7] Banerjee, K, S. Shukri, et al., "3-D ICs: A Novel Chip Design for Improving Deep Submicrometer Interconnect Performance and Systems-on-Chip Integration," Proceedings of the IEEE, Vol. 89, No. 5, pp 602-633, May 2001.

[8] Meindl, J.D., "Interconnect Opportunities for Gigascale Integration," IEEE Micro, Vol. 23, Issue 3, pp. 28-35, May-June 2003.

[9] Venkatesan, R. et al., "Optimal n-tier Multilevel Interconnect Architectures for Gigascale Integration (GSI)," IEEE Transactions on Very Large Scale Integration Systems, Vol. 9, No. 6, pp. 899-912, December 2001.

[10] Pavlidis, V. and E.G. Friedman, "Interconnect Delay Minimization Through Interlayer Via Placement in 3-D ICs," Proceedings of the 15th ACM Great Lakes Symposium on VLSI, Chicago, Illinois, USA, pp 20-25, 2005.

[11] Sai-Halasz, Gerorge, "Performance Trends in High-End Processors," Proceedings of the IEEE, Vol. 83, No. 1, January 1995.

[12] Garrou, P, "Future ICs Go Vertical," Semiconductor Packaging Magazine, pp. 10-16, Feb. 2005.

[13] Niedrmayer, M, S. Guttowski, et. al., "Miniaturization Platform for Wireless Sensor Nodes Based on 3D-Packaging Technologies," Proceedings of the Fifth International Conference on Information Processing in Sensor Networks, Nashville, Tennessee, USA, pp. 391-398, 2006.

[14] Mondal, Mosin, et al., "Thermally Robust Clocking Schemes for 3D Integrated Cicuits," EDAA, pp. 1206-1211, 2007.

[15] Loi, G.L., B. Agrawal, et al., "A Thermally-Aware Performance Analysis of Vertically Integrated (3-D) Processor-Memory Hierarchy," DAC 2006 San Francisco, CA, USA, pp 991-996, July 24-28, 2006.

[16] Yu, C.H., "The 3rd dimension—More Life for Moore's Law," IEEE INTERPACK Conference, 2006.

[17] Garimella, S. et al., "Thermal Challenges in Next Generation Electronic Systems – Summary of Panel Presentations and Discussions," IEEE Transactions on Components and Packaging Technologies, Vol. 25, No. 4, December 2002.

[18] Knickerbocker, J.U., et. al., "Development of next-generation system-on-package technology based on silicon carriers with fine-pitch chip interconnection," IBM

Journal of Research and Development, Vol. 49, No. 4/5, pp 725-753, July/September 2005.

[19] Kwon, Yongchai and Jongwon Seok, "An Evaluation Process of Polymeric Adhesive Wafer Bonding for Vertical System Integration," Japanese Journal of Applied Physics, Vol. 44, No. 6A, pp. 3893-3902, 2005.

[20] Koo, J.M., Sungjun Im, et. al., "Integrated Microchannel Cooling for Three-Dimensional Electronic Circuit Architectures," Journal of Heat Transfer, Vol. 127, pp 49-58, January 2005.

[21] Wilkerson, P., et al., "Fast, Automated Thermal Simulation of Three-Dimensional Integrated Circuits," Inter Society Conference on Thermal Phenomena, pp 706-713, 2004.

[22] Puttaswamy, K., and G Loh, "Thermal Analysis of a 3D Die-Stacked High-Performance Microprocessor," GLSVLSI'06, Piladelphia, PA, USA, pp 19-24, April 30-May 2, 2006.

[23] Chiang, T.Y., et al., "Thermal Analysis of Hetergeneous 3D ICs with Various Integration Scenarios," IEDM'01, pp. 681-684, 2001.

[24] Chen, G., and S Sapatnekar, "Partition-Driven Standard Cell Thermal Placement," ISPD'03, pp. 75-80, Monterey, California, USA, April 6-9, 2003.

[25] Tsai, J.L., et al., "Temperature-Aware Placement for SOCs," Proceedings of the IEEE, Vol. 94, No. 8, pp 1502-1518, August, 2006.

[26] Yuan, L., S. Leventhal, and G. Qu, "Temperature-Aware Leakage Minimization Technique for Real-Time Systems," ICCAD'06, pp 761-764, November 5-9, 2006.

[27] Skadron, K., et al., "Temperature-Aware Microarchitecture: Modeling and Implementation," ACM Transactions on Architecture and Code Optimization, Vol. 1, No. 1, pp. 94-125, March 2004.

[28] Balakrishnan, K., et al., "Wire Congestion and Thermal Aware 3D Global Placement," ASP-DAC 2005, pp. 1131-1134.

[29] Palacin, J. M. Salleras, et al., "Dynamic Compact Thermal Models with Multiple Power Sources: Application to an Ultrathin Chip Stacking Technology," IEEE Transactions on Advanced Packaging, Vol. 28, No. 4, pp. 694-703, November 2005.

[30] Arik, M., J. Garg, and A. Bar-Cohen, "Thermal Modeling and Performance of High Heat Flux SOP Packages," IEEE Transactions on Advanced Packaging, Vol. 27, No. 2, May 2004.

[31] Jacob, P., et al., "Predicting the Performance of a 3D Processor-Memory Chip Stack," IEEE Design & Test of Computers, pp 540–547, November-December, 2005.

[32] Davis, W.R., et al., "Demystifying 3D ICs: The Pros and Cons of Going Vertical," IEEE Design and Test of Computers, Vol. 22, No. 6, pp 298-510, Nov/Dec 2005.

[33] Benkart, P, et al., "3D Chip Stack Technology Using Through-Chip Interconnects," IEEE Design & Test of Computers, pp. 512–518, November-December, 2005.

[34] Im, Sungjun and Kaustav Banerjee, "Full Chip Thermal Analysis of Planar (2-D) and Vertically Integrated (3-D) High Performance ICs," IEDM Technology Digest, pp. 727-730, 2000.

[35] http://www.ioffe.rssi.ru/SVA/NSM/Semicond/InP/thermal.html, website, last accessed: August 4, 2007.

[36] Incropera, Frank and David DeWitt. Fundamentals of Heat and Mass Transfer, 5th Edition. New York: John Wiley & Sons, 2002; Appendix A

[37] Tuckerman, David, "5 Critical Challenges for 3D Packaging," Chip Scale Review, pp. 62, March 2007. Available online at www.ChipScaleReview.com

[38] Goplen, B. and S. Sapatnekar, "Thermal Via Placement in 3D ICs," Proceedings of the 2005 International Symposium on Physical Design, San Francisco, California, USA, pp. 167-174, 2005.

[39] Ozguz, V., et al., "High Power Chip Stacks with Interleaved Heat Spreading Layers," IEEE Electronic Components and Technology Conference, Las Vegas, Nevada, USA, pp 1467-1469, May 21-24, 2000.

[40] 3M Dielectric Fluids, website: www.3m.com, last accessed August 5, 2007.

[41] Kandlikar, S.G., "A General Correlation for Saturated Two-Phase Flow Boiling Heat Transfer Inside Horizontal and Vertical Tubes," ASME Journal of Heat Transfer, Vol. 112, pp. 219-228, 1990.

[42] CRC Handbook of Chemistry and Physics, David R. Lide, Ed. 79th Edition, CRC Press, Boca Raton, FL, 1998.

[43] 3M data sheet, available online at the following web address: http://products3.3m.com/catalog/us/en001/oil_gas/specialty_materials/node_2KJ8NGV19Gbe/root_GST1T4S9TCgv/vroot_G1F6DNZDBVge/gvel_X65PZNZ1SHgl/theme_us_oilgas_3_0/command_AbcPageHandler/output_html. Last accessed: 8/17/07.

[44] Incropera, Frank and David DeWitt. *Fundamentals of Heat and Mass Transfer*, 5th Edition. New York: John Wiley & Sons, 2002; Chapt 8.

[45] Incropera, Frank and David DeWitt. *Fundamentals of Heat and Mass Transfer, 5th Edition.* New York: John Wiley & Sons, 2002; pp 337, pp 467-472.

[46] Kays, William, Michael Crawford and Bernhard Weigand. *Convective Heat and Mass Transfer, 4th Edition.* New York: McGraw-Hill, 2005. pp 92-94; 107-108.

[47] Sparrow, E.M and S.H. Lin. Physics of Fluids v. 7 iss. 3, 1964. pp 338-347.

[48] Munson, Bruce, Donald Young, and Theodore Okiishi. *Fundamentals of Fluid Mechanics, 4th Edition.* New York: John Wiley & Sons, 2002; pp 448.

[49] ITRS, The International Technology Roadmap for Semiconductors, 2005, Semiconductor Industry Association, http://www.ITRSnemi.org.

[50] iNEMI, Electronics Manufacturing Initiative Technology Roadmap, 2004, http://www.nemi.org.

[51] Viswanath, R., V. Wakharkar, A. Watwe, and V. Lebonheur, "Thermal Performance Challenges from Silicon to Systems," Intel Technology Journal, Quarter 3, 2000.

[52] Shakouri, A., and Zhang, Y., "On-chip solid-state cooling for integrated circuits using thin-film microrefrigerators," IEEE Transactions on Components and Packaging Technologies, Part A: Packaging Technologies, Vol. 28, No.1, pp. 65 – 69, 2005.

[53] Wilkerson, P., A. Raman, and M. Turowski, "Fast, Automated Thermal Simulation of Three-Dimensional Integrated Circuits," IEEE Inter Society Conference on Thermal Phenomena, pp. 706-713, 2004.

[54] Ekpanyapong, Mongkol, Michael Healy, et al., "Thermal-Aware 3D Microarchitectural Floorplanning," Technical Report GIT-CERCS-04-37, Georgia

Institute of Technology Center for Experimental Research in Computer Systems, 2004.

[55] Chun Ku, Ja, Serkan Ozdemire, Gokhan Memik, and Yehea Ismail, "Thermal Management of On-Chip Caches Through Power Density Minimization," IEEE Transactions on VLSI Systems, Vol. 15, No. 5, pp. 592-604, May 2007.

[56] Link, G.M. and N. Vijaykrishnan, "Hotspot Prevention Through Runtime Reconfiguration in Network-On-Chip," IEEE Proceedings of the Design, Automation and Test In Europe Conference and Exhibition, 2005.

[57] Chaparro, Pedro, et al. "Thermal-Aware Clustered Microarchitectures" Proceedings of the IEEE International Conference on Computer Design, 2004.

[58] Koo, J.M. et al., "Convective Boiling in Microchannel Heat Sinks with Spatially-Varying Heat Generation," IEEE Inter Society Conference on Thermal Phenomena, pp. 341-346, 2002.

[59] Bar-Cohen, A., P. Weng, C. Bachmann, E. Rahim, "Thermal Analysis and Control of On-Chip Hot Spots in Nanoelectronic Devices," submitted for publication at the 18th International Symposium on Transport Phenomena, 27-30 August, 2007, Daejeon, Korea.

[60] Wang, E. et al., "Micromachined Jets for Liquid Impingement Cooling of VLSI Chips," Journal of Microelectromechanical Systems, Vol. 13, No. 5, pp 833-842, 2004.

[61] Cader, T. and D. Tilton, "Implementing Spray Cooling Thermal Management in High Heat Flux Applications," IEEE Inter Society Conference on Thermal Phenomena, pp. 699-701, 2004.

[62] Yang, B., P. Wang, and A. Bar-Cohen, "Thermoelectric Mini-Contact Cooler for Hot-Spot Removal in High Power Devices," IEEE Electronic Components and Technology Conference, pp 997-1002, 2006.

[63] Wang, Peng, "On-Chip Thermoelectric Cooling of Semiconductor Hotspot," PhD Thesis, University of Maryland, June 2007.

[64] Wang, P., A Bar-Cohen, B Yang, "Analytical Modeling of On-Chip Hotspot Cooling Using Silicon Thermoeletric Microcoolers," Proceedings of the IMECE 2006, Paper ID: IMECE 2006-14383.

[65] Muzychka, Y.S., J.R. Culham, M.M. Yovanovich, "Thermal Spreading Resistance of Eccentric Heat Sources on Rectangular Flux Channels," Journal of Electronic Packaging, vol. 125, pp 178-185, June 2003

[66] Carslaw, H.S. and J.C. Jaeger. *Conduction of Heat in Solids*, 2nd Edition. Oxford, England: OxfordUniversity Press, 1946.

[67] Strauss, W.A. *Partial Differential Equations, An Introduction*. Canada: John Wiley and Sons, Inc., 1992.

[68] Yovanovich, M. M., Muzychka, Y.S. and Colham, J.R., "Spreading Resistance of Isoflux Rectangles and Strips on Compound Flux Channels," Journal of Thermophysics and Heat Transfer, vol. 13, pp. 495-500, 1999.

[69] Smalc, Martin et al, "Thermal Performance of Natural Graphite Heat Spreaders," Proceedings IPACK2005, San Francisco, California USA, July 17-22, Paper Number: IPACK2005-73073.

[70] Montesano, Mark, "Annealed Pyrolytic Graphite," Advanced Materials and Processes, June 2006, pp. 1-3.

[71] Ivakin, Sukhodolov, Ralchenko, *et al.* "Measurement of thermal conductivity of polycrystalline CVD diamond by laser-induced transient grating technique," Quantum Electronics, vol. 32 iss. 4, pp 367-372, 2002.

[72] http://www.graftechaet.com/Technical-Documents/Technical-Bullitin/TB321-EG-SS-TYP.pdf Last accessed: July 10, 2007.

[73] Madhusudana, C. V., *Thermal Contact Conductance*, Springer-Verlag, New York, 1996.

[74] ANSYS Release 9.0 Documentation from ANSYS Incorporated; Chapter 1: Contact Technology; Chapter 5.1: Modeling Thermal Contact; TARG170 in element reference; and CONTA174 in element reference.

[75] Mahajan, R., et al., "Thermal Interface Materials: A Brief Review of Design Characteristics and Materials," Electronics Cooling Magazine, vol. 10, no. 1, 2004.