A Multiscale Finite Element Modeling Approach for Thermal Management in Heterogeneous Integrated Circuits

Thesis

Presented in Partial Fulfillment of the Requirements for the Degree Master of Science in the Graduate School of the Ohio State University

By

Peter J Bonavita, B.S.M.E

Graduate Program in Mechanical Engineering

The Ohio State University

2019

Thesis Committee:

Dr. Rebecca B. Dupaix, Advisor

Dr. Waleed Khalil

Abstract

Modern radio frequency (RF) microsystems are challenged to deliver improved performance across an ever-changing landscape of applications and requirements. As the demand for high-power and high-frequency systems grows, heterogeneous integration of new, high-power compound semiconductor (CS) technologies with existing silicon-based circuitry may lead to a paradigm shift in the field of RF electronics. Intimate integration of high-power technologies leads to increased power densities which forces thermal management considerations to the forefront of design. At the moment, the availability thermal analysis tools for heterogeneously integrated technologies are limited, and thermal considerations are often relegated to the back end of the design cycle.

In this work, a multiscale finite element approach is developed for thermal management of heterogeneous integrated circuits. The proposed method is capable of simulating heat flow at multiple length scales using submodeling techniques which incorporate high spatial resolution near the active region while including realistic approximations of global boundary conditions.

Thermal simulations are presented here for a device implemented using DARPA's Diverse Accessible Heterogeneous Integration (DAHI). The device's thermal behavior is explored for a variety of possible configurations and operating conditions. To better inform future circuit designers, the device's primary thermal bottlenecks are identified and quantified in terms of their influence on temperatures within the device.

Acknowledgements

First and foremost I would like to thank my parents, Lorraine and Matthew Bonavita, for their endless love and support. I love you both with all my heart. Mom, thank you for our countless phone calls whenever I needed a sounding board or a soothing word of encouragement; you are the rock of our family; thank you for believing in me. Dad, at age twelve you taught me how to pitch: step straight towards the plate, and remember to follow through. I never made it to the majors, but the moral behind that pitching lesson helped me make it through the past year.

Thank you to my siblings Daniel, Anna, and Andrew, for all the laughter and memories we've shared together. Thank you to my friends David and Colin, for keeping me sane day-to-day in undergrad, and continuing that from afar during my graduate school. Thank you to Chad and Susie McGarry for sharing their home with me, *except* for on Sunday afternoons.

I would like to thank my advisor and mentor, Prof. Rebecca B. Dupaix for her guidance and support throughout my career at Ohio State. She has always inspired me when the path forward was unclear, and met me with understanding when I encountered turbulence. I have greatly enjoyed being her student.

I would like to thank Prof. Waleed Kahlil, Trevor Dean, Gus Fragasse, and Matt LaRue at the Electroscience Lab for their help answering my questions. Likewise, I would like to thank all the folks from the Sensors Directorate at AFRL who provided critical insights which bridged gaps in my knowledge. I would also like to thank my funding sources at the Dayton Area Graduate Studies Institute for graciously providing me with a graduate fellowship.

Vita

May 2013……....……Bellbrook High School, Bellbrook, OH

August 2014……..….Design Intern, Betts Industries, Warren, PA

December 2015……Assembly Intern, BMW Manufacturing, Spartanburg, SC

August 2016…..……....Vehicle Pre-Design Intern, Raytheon Missile Systems, Tucson, AZ

August 2017……Modeling and Simulation Intern, Booz Allen Hamilton, Dayton, OH

December 2018…….B.S. Mechanical Engineering, The Ohio State University

January 2018 – May 2019…….Graduate Research Assistant, The Ohio State University

Fields of Study

Major Field: Mechanical Engineering

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 Motivation

Future wireless communication applications require the development of agile and high-performance electronic systems. Modern radio frequency (RF) systems must meet steep performance demands, such as providing increased power capability within decreased device footprint [1]. DARPA's Diverse Accessible Heterogeneous Integration (DAHI) program investigates the performance benefits of heterogeneous integration between traditional complementary metal-oxide-semiconductor (CMOS) circuitry and emerging compound semiconductor (CS) technologies. When married together, these two technologies promise to deliver the high-power advantages of CS technologies, such as gallium nitride (GaN), along with the existing complexity and time accuracy available in CMOS. Such combinations offer breakthrough performance in next-generation microwave systems for both defense and commercial applications [1].

Previous heterogeneous integration efforts have focused on incorporating diverse technologies in the form of multi-chip modules; however, large separations between devices leads to delays which hinder system performance [2]. To address these limitations, single-chip integration has been proposed, resulting in the rise of thermal management as a key design challenge [2] [3] [4]. Due to the high-power capability of CS technologies, both CS and CMOS technologies on the same chip may be exposed to large temperature increases which exacerbate thermal failure mechanisms in the device [5]. Thermal challenges associated with multi-level 3D

integration are well-documented; lack of 3D-capable design tools [6], increased density of heat sources [7], and difficulty obtaining measured data [8] pose significant barriers to overcoming thermal problems at the design stage.

Simultaneously, costly and lengthy development cycles have driven efforts toward reducing DAHI's cost as a key aspect of accessibility [2]. Heterogeneous integration's limitless possible configurations and high power density require fast thermal simulation as an integral part of development [6] [9]. Melamed et al. [3] and Harris et al. [10] have developed techniques to create simulations for thermal analysis directly from layout, achieving a reasonable degree of accuracy, but with significant computational cost. Other tools which sacrifice accuracy for speed use simplified assumptions and abstractions to produce thermal profiles for first-level analysis. While fast and easy to use, the utility of simplified approaches is limited in that they do not provide accuracy necessary to maximize thermal performance [9].

In this thesis, a high-fidelity finite element modeling approach has been proposed for thermal management of a family of heterogeneous integrated circuits developed using the DAHI process. A multiscale, submodeling-based technique has been developed to allow high spatial resolution close to the heat source, while minimizing overall simulation times. The benefits of this technique include the ability to obtain accurate thermal maps in the vicinity of high-power devices, enabling thermal management to become an integral part of the circuit design process. Furthermore, this technique combines high-resolutions required for accuracy with realistic global boundary conditions, thereby eliminating the use of simplified abstractions employed previously. Outcomes of this work include predictions of the device's thermal characteristics across a range of operating conditions, and detailed investigations of critical thermal bottlenecks meant to inform future circuit designers.

1.2 Outline of Following Chapters

Chapter 2 contains an overview of literature topics relevant to this research, including background on heterogeneous integration, an overview of heat transfer, a description of the finite element method, and prior thermal modeling efforts. The information is intended to be pertinent and accessible for readers with various multidisciplinary backgrounds, assuming a basic level of technical familiarity.

Chapter 3 describes the function and structure of the device being simulated, and the thermal model developed in this thesis. Chapters 4 and 5 present results from the model under the assumption of single-channel operation. Initial exploration of the device's thermal characteristics, and a number of trade studies describing the influence of various thermal design factors are presented here. In Chapter 6, the model is extended to encompass multichannel operation and thermal co-interaction between two active devices. Chapter 7 summarizes the findings of the previous chapters, and provides an outline for future work related to this research effort.

2 Literature Review

2.1 History of Heterogeneous Integration Technology

Heterogeneous integration technology traces its roots back to the development of three-dimensional integrated circuit (3DICs) technology. In the early 1980s, 3DICs were proposed as a pathway toward improved packing efficiency [11]. Even at that time, thermal management of multi-level devices was perceived as a significant challenge which led to the pursuit of wafer-scale integration techniques perceived as "more amenable to solution" [11]. In the early 2000s, 3DIC efforts were revived to increase device density and reduce interconnect delays [12]. Once more, this brought thermal considerations to the forefront of design. Rahman and Reif examined power dissipation within 3DICs using finite element simulations of multi-strata devices and found that significant reductions in device thermal resistance were required for feasible implementation of 3DICs [13]. Other modeling efforts which followed explored early heterogeneous integration and focused on various aspects of thermal design, including dissipation through inter-layer vias [14], automated thermal simulation from layout [6], influence of device layout and substrate characteristics [15], and flip-chip.

In the mid-2000s, DARPA's Compound Semiconductor Materials on Silicon (COSMOS) program began to explore heterogeneous integration of indium phosphide (InP) on silicon CMOS [16]; this project later became the forerunner to the DAHI effort. Around 2012, the DAHI program extended its heterogeneous integration efforts to include GaN-on-Si devices [2]. These

projects sought to extend next-generation RF performance by leveraging the material advantages of GaN technology, including its large bandgap and high-power capability, with existing high-density CMOS technology [17]. The intervening years have seen the GaN-based DAHI efforts mature significantly. In 2017, LaRue et al. demonstrated the first fully-integrated transmitter combining both GaN and CMOS [18]. This DAHI-integrated transmitter is capable of producing high output powers which lead to large temperature increases within the device; the focus of this thesis is thermal modeling of this particular device.

2.2 Intro to GaN Devices and Heat Transfer

As noted above, GaN's large bandgap is desirable for implementation in high voltage devices, and demonstrated good performance in the development of both light-emitting diodes and RF power amplifiers [19] [20]. The backbone of the DAHI-integrated transmitter modeled in this thesis is an eight-finger GaN high-electron mobility transistor (HEMT). A conceptual illustration of an HEMT is given in Figure 1.

At the interface between AlGaN and GaN, carriers gather in high concentration due to the differences in the two materials' electron bands [19] [20]. This region is known as the channel, or two-dimensional electron gas (2DEG). Flow of electrons through the 2DEG can be controlled by varying voltage applied to the gate, which allows the device to function as an electrical switch. While operating as an RF power amplifier (PA), the gate receives an input signal v_{in} which is amplified into a larger AC signal v_{out} across the drain and source terminals [20]. Additional DC power is supplied to the device, some of which is lost as heat due to Joule heating inefficiencies.

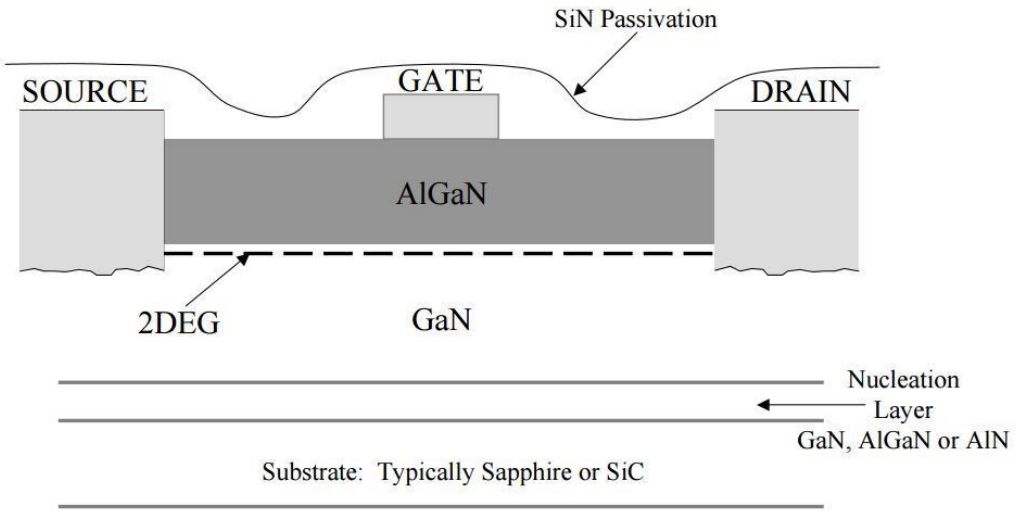


Figure 1: Conceptual Illustration of GaN HEMT, from [19]

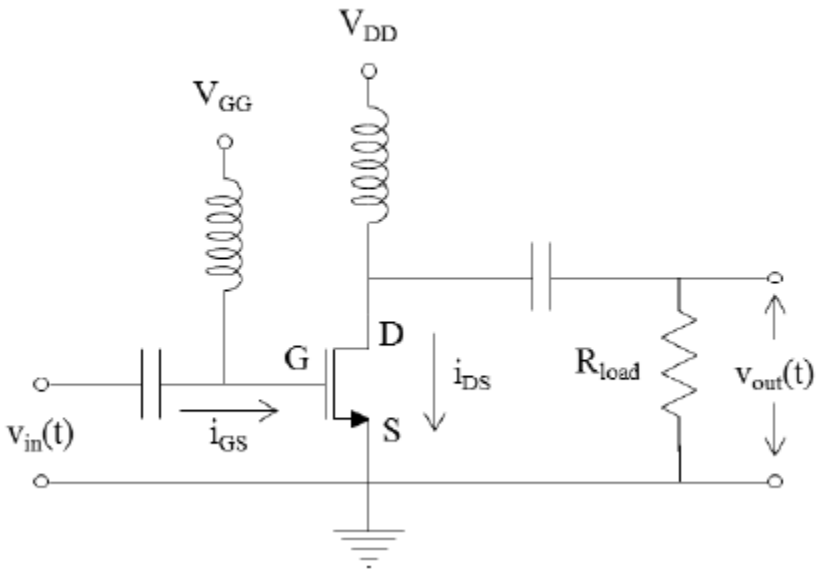


Figure 2: Basic RF PA Circuit, from [20]

Heat generation within semiconductor devices is primarily governed by joule heating as electrical current passes through the resistance of the channel [21]. The net heat generation due to Joule heating is generally taken to be the dot product of the current density $\bar{J}$ and electric

field $\overline{E}$, as given below. For the devices considered here, maximum heat generation occurs in the channel region where current and field densities are highest [20].

$$H = \bar{J} \cdot \overline{E}$$

In the case of the GaN PA, heating losses cause the device's temperature to rise significantly, hence the need for thermal management within the device. Once heat is generated at the PA, conservation of energy dictates the heat must be dissipated away by one of three modes of heat transfer: conduction, convection, or radiation. In general, all three modes of heat transfer are relevant, though radiation has been neglected in this analysis. The temperature of the device is then governed by the classical heat transport equation given below. In the form shown, ρ is density, C is specific heat, and H is the rate of heat generation as defined above.

$$\rho C \frac{dT}{dt} = -\nabla \cdot \overline{q_c} + H$$

Fourier's law of conduction describes the conductive heat flux $\overline{q_c}$ through a solid in terms of its thermal conductivity k. In general, heat transfer via conduction refers to the combined effects of energy transferred through the motion of electrons, and vibrational waves called phonons which are translated through the atomic lattice of the material [22]. Plentiful electrons are present in metals; the energy transfer associated with their many collisions is responsible for most conductive heat transfer. For semiconductors, relative lack of electrons means that phonon conduction through the lattice dominates conduction. In either case, thermal conductivity is the phenomenological transport property which describes how easily heat flows through the solid medium.

$$\overline{q_c} = -k\nabla T$$

Likewise, convection refers to heat exchange between an object's surface and its surrounding fluid. In general, convective heat flux q_h is determined by the combined effects of

conduction at the solid-fluid interface, and advective mixing due to the fluid's motion. Newton's Law of cooling describes convective heat exchange between an object and its surrounding fluid (most often air) in terms of their temperature difference $T - T_\infty$, and a convective heat exchange coefficient h. In most cases, determination of the heat transfer coefficient is difficult because the underlying interactions are quite complex, but correlations are widely available corresponding to a wide range of fluids and flow regimes [22]. Unlike the Fourier Law, Newton's Law of Cooling cannot be explicitly stated in vector form, but the general form is given below.

$$q_h = h(T - T_\infty)$$

In steady state form, the governing equation for heat transfer within the device reduces to the form given below. Recognizing convection occurs only at the outer surfaces, convection is treated as a boundary condition, so it has not been explicitly shown in the governing equation. The right-hand side of the equation equals zero, which reflects a re-statement of conservation of energy. Heat within the device cannot be created or destroyed; rather, it must conduct outward from the heat source until it arrives at the outer surfaces to be dissipated away to the surroundings via convection.

$$\nabla \cdot (k\nabla T) + H = 0$$

In one dimension, assuming no heat generation, the heat equation simplifies to the form shown below. For this case, the equation can be solved to describe the temperature distribution of a plane wall of length L with prescribed temperatures T_1 and T_2 on either side. The net heat flux per unit area (q") can then be determined, as shown in Figure 3. Assuming constant cross-sectional area, the total heat transfer rate q through the wall is obtained from the product of the heat flux q" and the wall area A. The negative sign in the one dimensional heat equation is

chosen to satisfy the second law of thermodynamics, such that a positive heat flux flows in the direction of decreasing temperature.

$$\nabla \cdot (k\nabla T) = \frac{d}{dx}\left(k\frac{dT}{dx}\right) = 0$$

$$q'' = -k\nabla T = -k\frac{dT}{dx} = -k\frac{(T_2 - T_1)}{L}$$

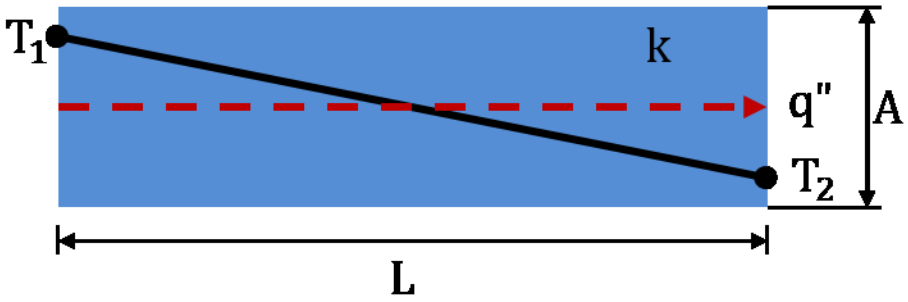


Figure 3: Heat Transfer through a Plane Wall

In practice, this 1D scenario is commonly described using an analogy borrowed from electronic circuits. The temperature drop across the wall and the corresponding heat transfer rate are related via the wall's thermal resistance R_{1D}, which is a function of the wall's geometry (L, A) and its thermal conductivity (k) as shown in Figure 4.

$$R_{1D} = \frac{T_1 - T_2}{q} = \frac{L}{kA}$$

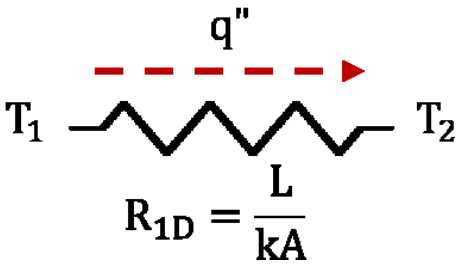


Figure 4: 1D Thermal Resistance

The 1D analogy is valid only for scenarios which conform to the simplifying conditions explained above. In the context of modeling GaN devices, the size of the heat source is small relative to the overall size of the device, which introduces an additional term conceptually known as thermal spreading resistance R_s [23]. Here, the increase in overall thermal resistance is attributed to the generation of heat in a small area, which generally increases the difficulty of removing heat and leads to greater temperature increases at the heat source. For instance, consider the example described in Figure 5. Here a 1000 W uniform heat flux is applied to one side of a 1D bar while the other side's temperature is fixed to 0°C. In the first case, corresponding to the 1D analogy above, the heat flux is applied to the entire 4 m^2 area of the bar's end surface. In the second case, the same 1000 W heat flux is applied to a smaller 0.5 m^2 area. Simulations of each case shown in Figure 6 reveal the peak temperature of Case 2 is 64% higher than Case 1 due to the additional spreading resistance, found to be $R_s = 0.008\,°\frac{C}{W}$. The simulations themselves do not include a spreading resistance term, so to speak. Rather they are full 3D simulations of the conduction equation, here intended to illustrate the concept of spreading resistance.

$$R_T = R_{1D} + R_s$$

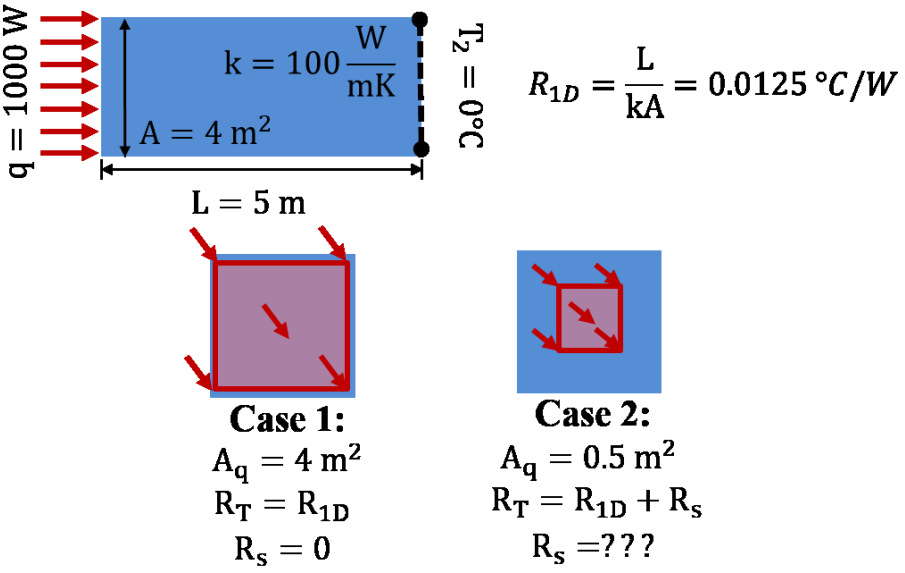


Figure 5: Spreading Resistance Example

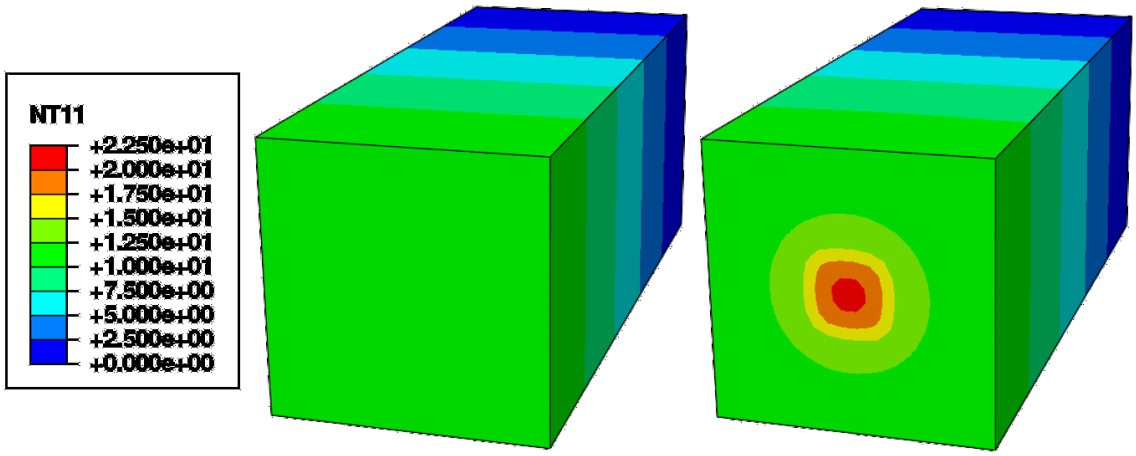


Figure 6: Spreading Resistance Example Simulation Results

2.3 Introduction to Finite Element Method

Analytical solutions are available for various basic cases of 2D and 3D heat transfer problems, but in general the complex geometries of multidimensional heat transfer makes these

solutions quite complex or limited to specific applications. Instead, thermal modeling is commonly done using numerical solution techniques such as the finite difference method, or the finite element method (FEM). In this thesis, thermal modeling is performed using the finite element method; this section provides a brief introduction for those unfamiliar with the subject of FEM.

The finite element method is a numerical analysis technique commonly used to solve problems in engineering and physics. Most commonly, FEM is used for structural analysis, such as analyzing the deflection of a bridge under load, but the technique is generally applicable to a wide variety of problems including dynamics, fluid mechanics, and heat transfer [24]. In general, solving the governing equations of large structural, fluid, or thermal problems requires solving a partial differential equation or boundary value problem, which is often difficult. Instead, FEM works by breaking a complex problem into small, discrete pieces called elements, then solving the governing equations for each element using a large, interconnected system of equations. In this way, FEM solves an approximated form of the problem being investigated, but solves the problem in a way which can be solved by computers *very* quickly.

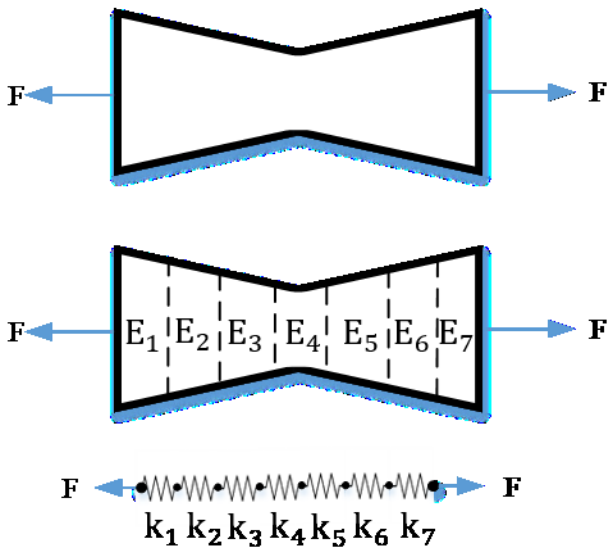


Figure 7: Pulling on a Plate, FEM Example

For example, Figure 7 illustrates a plate being pulled by force (F) on either end. Normally, the irregular geometry of the plate would make it difficult to analytically determine the how far the plate will stretch. Instead, the plate can be broken up into several small elements which can be thought of as a series of mechanical springs. The stiffness of each spring element (k_x) is independently determined by the plate's material properties, and the element geometry. In this manner, FEM software assembles what's known as a stiffness matrix representing the relationship between each individual element and its neighbors. The software then solves the corresponding system of equations to determine the forces and displacements throughout the structure for any specified combination of loads.

The example above is easily extended to multiple dimensions, or to thermal analysis using the thermal resistance analogy described in the previous section. The underlying physics differ, but the overall goal remains the same: FEM assembles a system of discretized equations describing the problem, then solves it. The challenge of performing finite-element analysis is in the proper design of the model to suit the scenario being studied. For example, the solution's accuracy is determined by the number of elements to discretize the problem. This is largely due

to the use of numerical integration techniques whose accuracy is sensitive to element size, as in the approximation of an integral using Riemann sums shown in Figure 8.

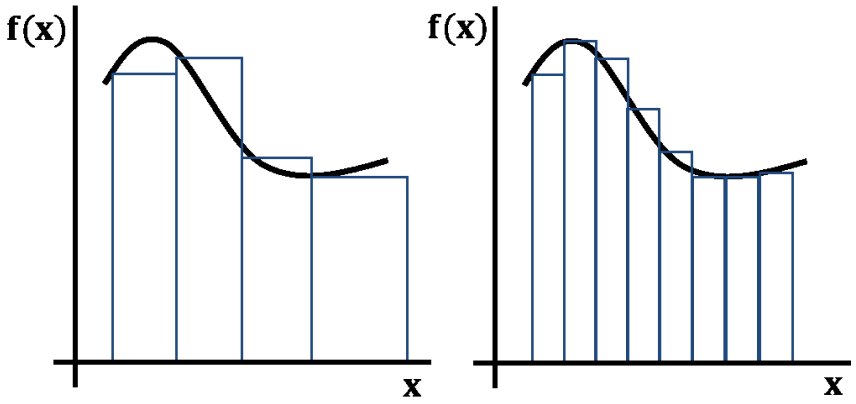


Figure 8: Riemann Sum Integration Illustration

However, a tradeoff exists between accuracy of results and simulation time. As element count is increased, the computational speed of finite element simulations begins to increase nonlinearly. This is caused by scaling of the stiffness matrix, which requires many more operations to solve each additional equation in large, multidimensional problems. Modern solvers generally have sufficient memory to tackle large problems, but in certain scenarios the size of the stiffness matrix becomes prohibitively large for fast solutions.

Such is often the case in finite element modeling of high-power transistors. A length scale problem emerges when heat transfer must be modeled at high resolution near active regions, with feature sizes on the order of nanometers, and across the global boundaries of the device's chiplet, which may typically measure on the order of a millimeter. In this scenario, the large element counts required for accuracy near the heat source often render the problem intractably large, leading designers to pursue simplified analysis techniques in the form of approximated boundary conditions or 2D analysis. In Chapter 3 a multiscale, submodeling-based approach will be introduced to alleviate some of the length scale problems associated with the

devices modeled in this thesis. This method combines results between two solution sets: one simulated using coarse detail at the global scale, and another encompassing only a smaller region of interest modeled at high resolution. In latter chapters of this work, this technique will be demonstrated to be adept at exploring the device's performance through a range of simulation-based investigations.

2.4 Thermal Modeling & Analysis of GaN Devices

Owing to the recent development and wide variety of heterogeneous integration techniques, there are relatively few thermal modeling studies available specifically dedicated to heterogeneous devices. Instead, this section presents literature related to thermal analysis of GaN devices, which serves as the foundation for much of the research developed in this thesis.

In the realm of GaN devices, Darwish et al. have developed closed-form analytical expressions to predict the thermal resistance and temperature of field-effect transistors using solutions of the Laplace equation [25]. Their method predicts peak channel temperature using an infinite array of equally-spaced gate fingers, and agrees well with measured data obtained from GaAs-based devices.

More commonly, thermal modeling is accomplished using numerical techniques such as the finite-difference method, or using finite element method solvers. Garven and Calame [26] developed a multiscale thermal model for GaN devices using a finite difference code. Their model divides the device into a set of geometric parameters, and discretizes the domain across several length scales; in addition to peak temperature, the model reports global device temperatures as well. Interestingly, theirs is one of the few GaN thermal models which incorporates a die-attachment boundary condition.

One topic of importance in the literature is the proper level of inter-combination required between electrical and thermal simulations of the device. Multiphysics simulations have been explored to various extents, often with some tradeoff in detail between the electrical and thermal domains. Turin and Baladin [27] developed a model for GaN HEMTs which coupled analytical thermal solutions from the method of images with a 2D electrical simulations using the drift-diffusion model. Their model includes a 2D representation of global device temperatures which was used to compare the thermal performance of devices on different substrate materials. Likewise, Sodan et al. [28] coupled detailed 2D power profiles from electrical TCAD simulations with 3D finite element model simulations. Their work highlighted the importance of obtaining accurate power profiles when estimating peak temperature.

Significant efforts have been made in thermal modeling of GaN using finite element based tools to investigate various aspects of device design [29] [30] [31] [32]. Bertoluzza et al [29] illustrated the importance of including 3D thermal effects rather than relying on 2D approximations. The authors also demonstrated the effects of varying gate pitch, and width under the assumption of a uniform heat flux at the gate edge. Their model compares thermal performance of three typical substrates, and shows that predictions of peak channel temperature differ significantly from other methods which predict an average temperature in the channel. Douglas et al. [30] developed a similar model which corroborated differences seen between 2D and 3D thermal simulations and the improved performance of silicon carbide (SiC) and silicon (Si) substrates relative to sapphire. Both models listed above assumed perfect heat-sink boundary conditions at the base of the substrate, choosing to neglect thermal effects from packaging.

Heller and Crespo [31] developed an electro-thermal model of an AlGaN/GaN device which detailed its thermal performance for a variety of changing bias conditions. Their

simulations found that device thermal resistance varied significantly with bias conditions. Similarly, Venkatachalam et al. [32] used a series of coupled electro-thermal simulations to study the effect of device bias on temperature. Their simulations included thermo-mechanical effects to study thermal stresses introduced at the GaN-SiC interface, and found that increasing the device bias resulted in increasing compressive stress in GaN owing to the mismatch between thermal expansion coefficients of GaN and SiC.

Bagnall [20] developed a comprehensive framework for electrothermal modeling of GaN HEMTs. A thermal finite element model was used to investigate five primary factors most significant to heat generation within a GaN device: near-junction spreading resistance, substrate material conductivity, GaN-substrate boundary resistance, heat source size, and substrate base boundary condition. In particular, die-attachment adhesive was investigated in the form of a base boundary condition, but only for one idealized conductance case. In [33], Bagnall's work was extended to encompass experimental validation of the model using microRaman thermal measurements. There, characterizations of the device's mechanical stress and electrical characteristics were developed for a variety of thermal conditions.

In each of the studies noted above, simplifying assumptions were used to reduce the computational complexity of thermal modeling, which leads to variation of accuracy seen in each case. Packaging, device boundary conditions, and radiative heat transfer are three categories of assumptions made which have been recognized as significant factors influencing the accuracy of current thermal models [34]. Such variation has led others to view experimental measurement as the preferred method of thermal analysis for GaN devices. Micro-Raman [35] and infrared [36] measurement techniques have both been widely used in thermal measurements of GaN devices, and used in validation studies of the models presented above. Again, owing to the state of

development of heterogeneously integrated GaN, few operational devices are available for thermal experiments, so obtaining measured data has not been a focus in this project. The main goal of this effort has been to identify which aspects of the device are primary thermal bottlenecks; simulations are well-suited for analysis of multiple device configurations found in the later chapters of this thesis.

2.5 Thermal Modeling & Analysis in Heterogeneously-Integrated Devices

As noted above, there are fewer available sources of thermal analysis dedicated to heterogeneous devices, specifically GaN. This section will continue to focus on recent heterogeneous integration efforts from the DAHI program and related efforts which have performed some level of thermal analysis.

In 2006, Ji et al. [37] demonstrated thermal analysis of a flip-chip mounted GaN device using micro-Raman spectroscopy. While not a direct example of heterogeneous integration, the device in question featured tin bump interconnects used in the flip-chip mounting process. This multilevel integration is similar to the techniques employed by the device modeled in this thesis. Their confocal spectroscopy technique allowed for thermal measurements taken from within the substrate material, which enabled analysis of the heat-sinking capability of the flip-chip tin bumps. Thermal simulations under-predicted the measured temperature data, which was attributed to the possibility of imperfect attachment at the bump-substrate interfaces.

More recently, Choi et al [38] demonstrated a flip-chip heterogeneously integrated InGaP/GaAs HBT device which included integrated on-chip thermal management. In their device, indium bump heat sinks were placed near the active region to facilitate heat removal. Overall device thermal resistance was cut in half compared to previous devices. Three-dimensional thermal simulations were developed which demonstrated varying levels of

agreement with experimentally measured thermal data. Several experimental methods were used in their study to varying degrees of success based on uncertainties associated with the device's material properties.

From within the DAHI program, Harris et al. [10] [8] developed thermal simulations for DAHI-integrated InP and GaN devices. Their approach, similar to the power blurring technique developed in [9], extracts thermal model data from the layout and was used to simulate both InP and GaN devices for low power levels. Their results report only peak temperatures from the active region; temperatures of lower substrates are not given. Their work also included preliminary analysis of the effects of heterogeneous interconnect (HIC) placement on device peak temperature. Later, their model results were compared to experimental measurements obtained using infrared and micro-Raman microscopy; good agreement was observed between the simulation and the infrared results, with micro-Raman measurement reporting slightly higher temperatures than either other method.

Recently, McCluskey [39] studied thermal isolation between heterogeneously-integrated devices. His work included modeling of thermal dissipation in through-silicon via (TSV) arrays and experimental characterization of device dissipation using local laser spotting to simulate device heating patterns. Interestingly, they suggest leveraging the lateral thermal resistance of a low-k interposer to thermally isolate regions of the circuit with disparate thermal operating conditions. In the fashion later demonstrated by Fish et al. [40], microfluidic cooling arrays can be designed around local device hotspots to provide localized cooling where it is needed. To be feasible, the localized microfluidic cooler must operate more efficiently than a larger system designed to uniformly cool the entire device.

The efforts listed above describe the variety of approaches which have attempted to solve some aspect of thermal management for heterogeneous integrated circuits. Most of the efforts focus on experimental characterization of a few unique devices, others developed simulation techniques relevant to a larger number of systems. In any case the complex, multi-disciplinary nature of heterogeneous integration means a universal thermal management approach likely does not exist. Instead of seeking standardized solutions, it may be most helpful for circuit designers to have some insight into the most significant thermal bottlenecks within each particular family of devices. Therefore, the goal of the thermal modeling performed in this thesis is to characterize those factors which most significantly affect device heating and are most amenable to solution in the next iteration of heterogeneously integrated devices.

2.6 Die-Attachment Literature Review

As noted in prior sections, the effects of device packaging and global boundary condition are often ignored by simplified approaches typically used in thermal models of GaN devices. In the context of the devices modeled here, there is a large uncertainty associated with the role of the die-attach adhesive used to bond the DAHI-integrated die to its circuit board. A large portion of this thesis will be focused on investigating the device's performance under a range of adhesive conditions to better inform circuit designers of its significance.

During packaging and assembly of the device modeled in this thesis, a silver-filled die-attach adhesive was used to structurally bond the DAHI-integrated chiplet to a printed circuit board. This adhesive falls under a larger category of epoxy adhesives consisting of metal particles suspended in an adhesive carrier solution [41]. The role of the carrier solution is to form a high-strength bond between the chiplet and board, capable of withstanding mechanical stresses experienced by the device during operation. Metallic additives with high thermal conductivity,

such as silver or gold, are then mixed in to increase the conductivity of the composite mixture and allow for more efficient heat removal from the bottom of the die. However, metal filler's performance at high temperatures has been questioned due to interruptions in the thermal pathway created by isolations between individual metal flakes or particles [41].

Examples of similar silver-filled epoxy adhesives include [42] [43] [44]. Their stated conductivities range from 2-7 W/mK for bondline thicknesses on the order of 1 to 2 mil (0.025 to 0.05 mm). Assuming uniform application and neglecting voiding or contact resistances at the interface, the manufacturer's stated properties yield an upper limit on the equivalent boundary conductance on the order of $k'' = 2.56e5 \frac{W}{m^2 K}$ [42], as described below in Figure 9.

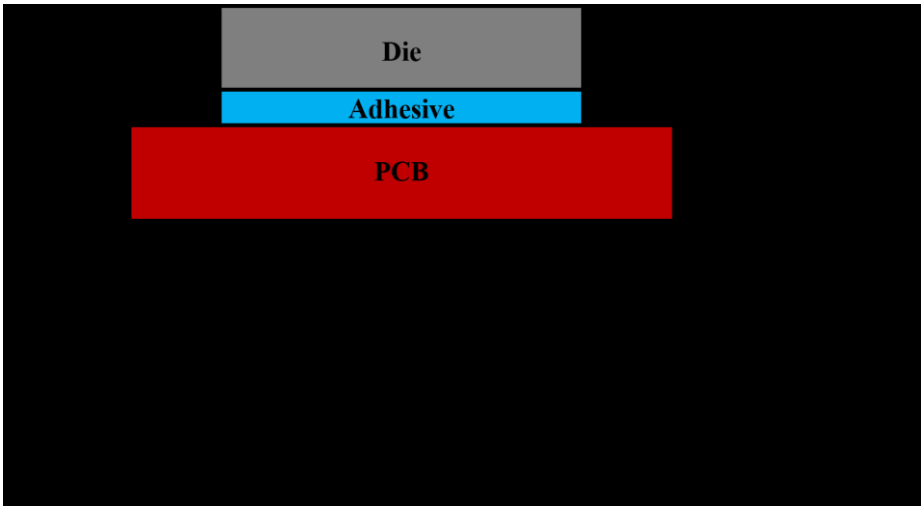


Figure 9: Die-Attach Conductance Illustration, Equivalent Conductance Calculation

However, several experimental studies of similar die-attach materials suggest that actual interfaces' conductances may be orders of magnitude smaller than the value given above, suggesting that bulk material data is not suitable for describing heat flow through the die-attachment interface [45]. Kurabayashi and Goodson developed a laser-reflectance thermometry technique to characterize the interfacial conductance of silver-filled epoxies and thermoplastic adhesives under a variety of conditions. They reported measurements of interfacial conductances

varying from $k'' = 1e3$ to $4e3 \frac{W}{m^2K}$. Furthermore, their results demonstrated strong dependence on the presence of voids or delamination within the sample interface, which have previously been shown to worsen at high temperatures [46].

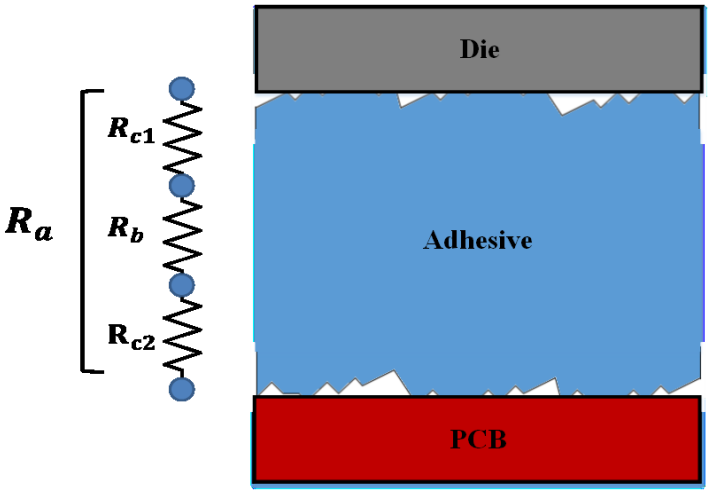


Figure 10: Die-attachment resistance model illustration, from [47]

Teertstra demonstrated a more complex measurement technique which further subdivided the attachment interface into a networked chain of series resistances, illustrated in Figure 10 [47]. In his work, the total measured die-attachment resistance (R_a) consisted of the conductive resistance of the adhesive (R_b), and two contact resistance terms (R_{c_1} and R_{c_2}) accounting for micro-scale roughness at the adhesive's interface surfaces. He also suggests the contact resistance terms were sensitive to settling of the filler material encountered while curing certain two-part epoxies. Teerstra's conductance measurements for silver-filled epoxy were similar to those published by Kurabayashi and Goodson, ranging from $k'' = 1.8e3$ to $5.6e3 \frac{W}{m^2K}$. The largest conductances measured in Teerstra's study were from a silver-filled thermoplastic material, ranging from $k'' = 1.4e4$ to $2.8e4 \frac{W}{m^2K}$. Both studies' conductance ranges fall orders of

magnitude lower than the value predicted from first-principles using bulk conductivity data from the manufacturer.

The findings of both studies described above illustrate the difficulty of modeling the die-attachment interface's thermal resistance. Die interfaces' thermal resistances have been shown to differ significantly from values predicted using first principles and bulk conductivity data. Furthermore, the resistance of individual adhesives have been found to be sensitive to the application procedure. Later chapters of this thesis will explore how changes in the behavior of this interface may influence device temperature in the DAHI-integrated device.

3 Methodology – Description of Device/Model

3.1 Device Overview & Structure

The device modeled throughout this thesis consists of a DAHI-integrated, GaN-on-SiC transmitter consisting of a phase modulator, CMOS/GaN buffer, and a three-stage GaN power amplifier (PA). The transmitter operates across the 2-5 GHz range with peak RF power output of 32.93 dBm and peak total transmitter efficiency of 41.32% [18]. The Stage 3 PA consists of an eight finger GaN HEMT, indicated in Figure 11, which is of primary interest as it serves as the main heat source on the chip. The PA consists of an 8 x 100um GaN transistor with gate length of 1um. Dissipation of the intermediate PA stages, phase modulator, and other passive components amount to less than 6% of total device heat dissipation and thus have less effect on the global temperature of the device [48]. Figure 12 contains results published previously from device measurements. During testing, peak RF output power reached 4.45W with a third-stage efficiency of <40% across all operating frequencies. Accounting for third-stage losses and assuming worst-case efficiency performance, the device dissipates as much as 5W of heat during normal operation [48]. For most simulations and explorations in this thesis, 5W heat dissipation has been assumed for ease of comparison between simulations. In Section 5.6, the thermal performance of the device will be explored over its full feasible range of operation at multiple heating & efficiency points.

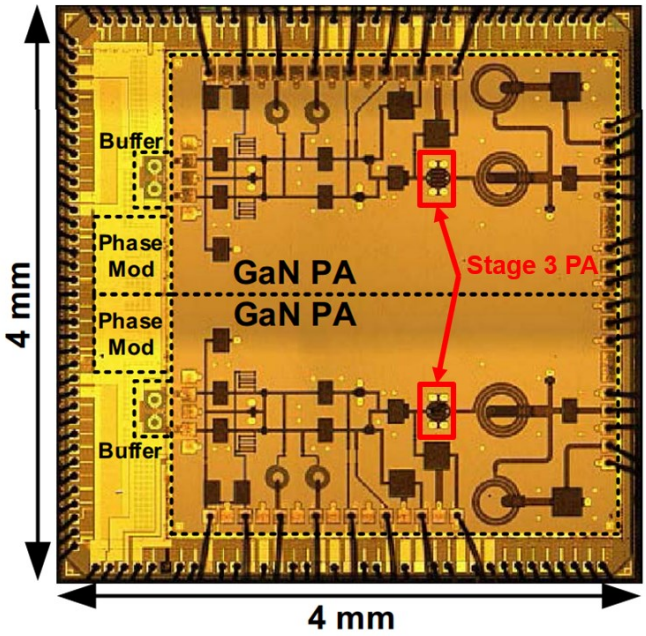


Figure 11: Top view image of the three-stage GaN PA, two channels shown [18]

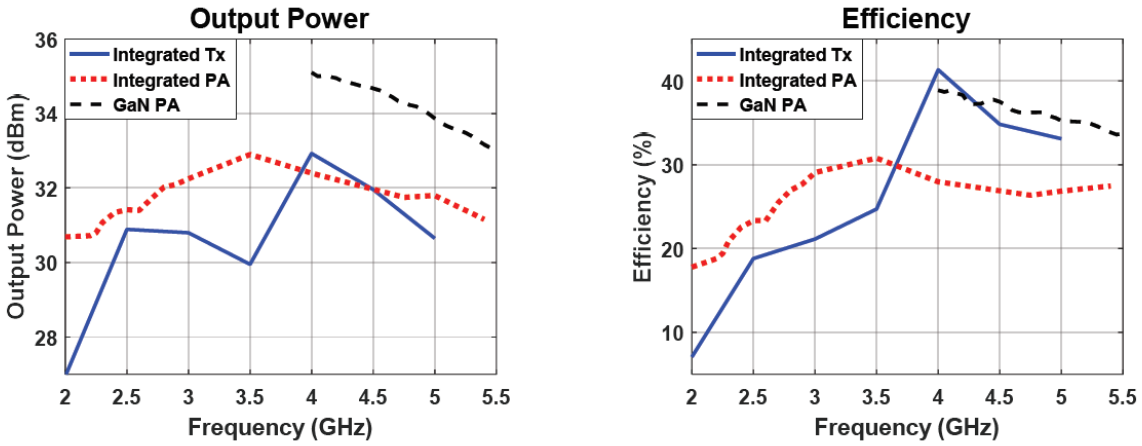


Figure 12: Measured output power of the GaN PA [48]

Figure 13 illustrates the device's cross-section and enumerates the nominal thickness of each layer. Only primary substrate layers are illustrated below; various thin nucleation and buffer layers deposited in the fabrication process not shown as they are neglected in this analysis. The device consists of two main epitaxial substrates consisting of several material layers, collectively

referred to as the ‘GaN substrate’ and ‘CMOS substrate’. The GaN substrate consists of a GaN layer of roughly 2um thickness; this layer is grown at high temperature using the standard process of metal-organic chemical vapor deposition (MOCVD) [49]. As is common in similar devices, the GaN layer was grown on top of a 50 um SiC substrate [50]. At the transition between SiC and GaN, a thin AlN buffer layer is introduced to aid nucleation and deposition of the GaN layer. Phonon mismatch at the AlN interfaces introduces a thermal boundary resistance between GaN and SiC which has been extensively studied and is known to cause significant increases in device temperatures during operation [51] [52]. Following growth, the PA’s ohmic contacts were then deposited onto the GaN surface. A gold metal layer below SiC to serves as a ground plane and mounting surface for vias which run through the GaN substrate. Each material’s basic thermal properties are recorded in Table 1 for reference.

Table 1: Thermal Material Properties of the Device’s Materials & Interfaces at 300K

Material	Thermal Conductivity (W/mK)
GaN	160 [29]
SiC	390 [29]
Gold	314
CMOS	1 [53] to 400 [54]
Bulk Si	150 [29]
PCB	68

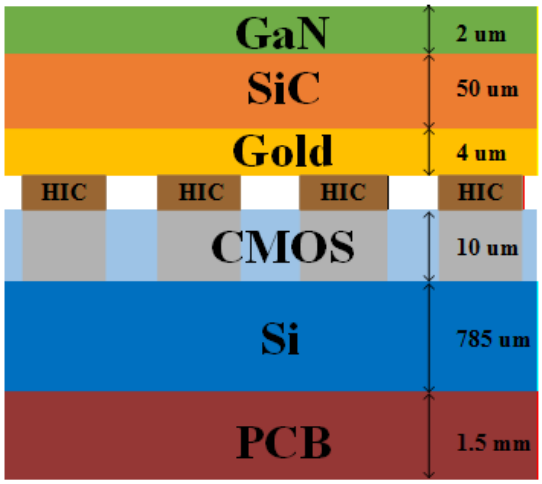


Figure 13: Device Substrate Cross-Section with Materials and Thicknesses

During the DAHI process, the GaN and CMOS substrates are combined via an array of heterogeneous interconnects (HICs) [55]. The HICs consist of gold microbumps electrically connected to the backside of the GaN substrate. Two types of HICs exist for different functions: electrical HICs (eHICs) serve as the primary electrical vein between GaN & CMOS, while thermal HICs (tHICs) serve only as thermal pathways from the GaN backside to the CMOS layer. Both types of HICs connect to the top metal layers of the CMOS process; below eHICs the metallization continues only partway through the CMOS layer, while in tHICs the metallization extends through the full thickness of the CMOS layer to the bulk silicon substrate on the backside. Elsewhere throughout the device, multifunctional through-silicon slot vias (TSVs) also run through the GaN substrate. An illustration of the two HIC variants and a TSV are illustrated in Figure 14.

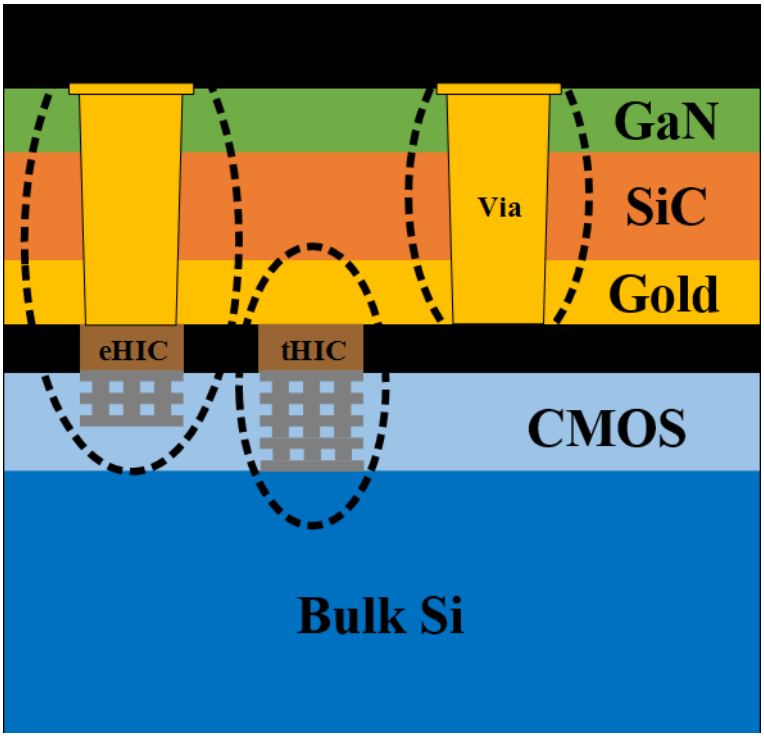


Figure 14: Illustration of eHIC, tHIC, and TSV features

In the immediate vicinity of the PA, HICs are arranged in an alternating checkerboard pattern illustrated below. Here, near the heat source, HICs are spaced closely together at the maximum allowable density allowed by metallization design rules which govern the maximum amount of metal in each layer of substrate. Farther from the heat source, the HIC density decreases to a uniform rectangular grid of HICs spaced farther apart. To differentiate between HICs in either region, latter sections of this report will refer to HIC instances in the checkerboard or rectangular grid as 'near-field' or 'far-field' HICs, respectively. As illustrated in Figure 15, eHICs generally exist only in the far-field of the PA, thus their contribution to thermal behavior is expected to be small. In this study, all eHICs have been modeled as tHICs and any differentiation has been neglected.

The top layer of the CMOS substrate consists of a mixed array of metal layers surrounded by silicon field oxide. Here, the metallization is used as the primary pathway for heat dissipation

from the bottom of the GaN substrate. Far from the Stage 3 PA, the CMOS layer includes low-power active components and some driving logic. During operation, the temperature of the GaN PA exceeds acceptable limits for the active CMOS components, which requires sacrificially increasing the size of the die to afford physical separation between the two. Low-power CMOS circuitry has been largely disregarded in this thermal analysis [18], however latter sections of this thesis will explore the effect of multichannel spacing on the device's self-heating within CMOS. Below the CMOS layer, roughly 800um of bulk silicon serves as an anchor and heatsink for the entire chiplet.

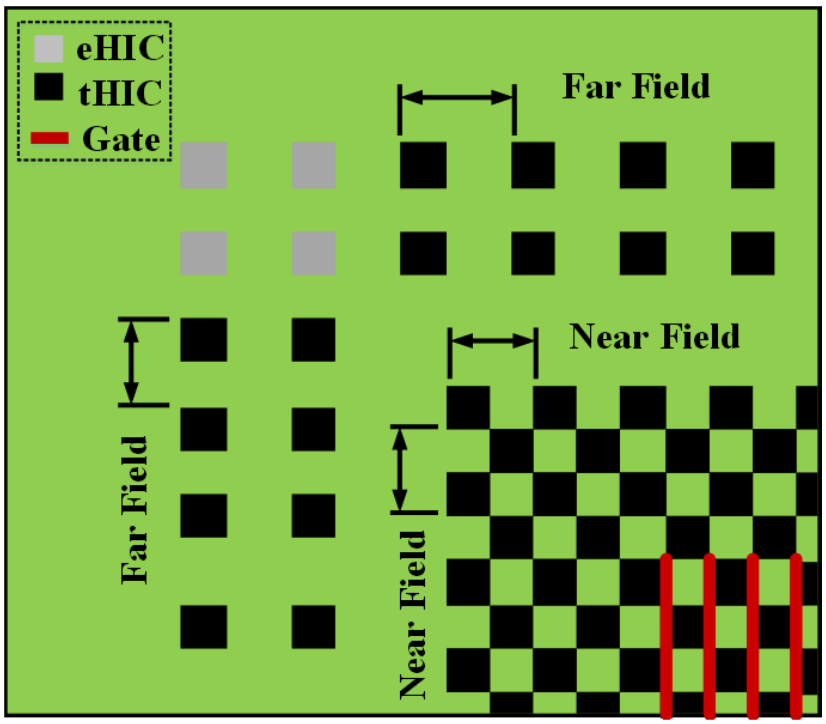


Figure 15: HIC Arrangements around Stage 3 PA (Quarter Symmetry View)

Accurate determination of the exact heating profile in multi-finger GaN devices remains a difficult task, but thermal models used in studies of similar devices typically assume constant heat flux along the profile of the gate [31]. In reality some asymmetry is expected, with more heat being dissipated toward the drain side near the location of highest current density.

Furthermore, it has been shown that the heating profile is heavily dependent on the bias conditions of both drain and source as well as the shape of the device [33]. For this study, simulations assume constant heat flux across the gate profile given in Figure 16 unless otherwise noted. Further investigation of this assumption will be presented in Section 5.1.

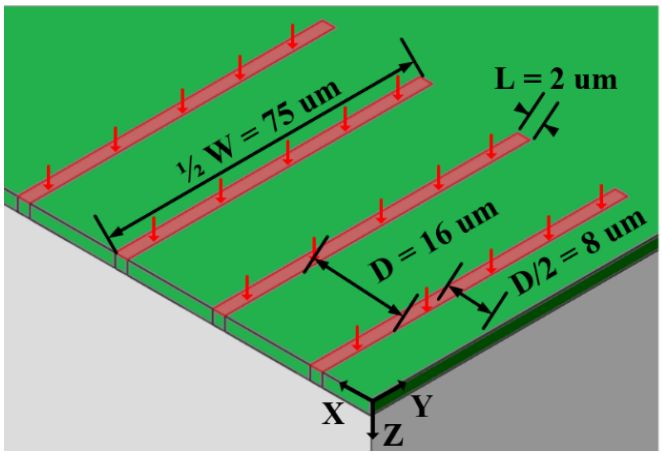


Figure 16: Stage 3 PA Gate Heating Region Dimensions (Quarter Symmetry View)

Following integration of the GaN/CMOS substrates using the DAHI process, the completed chiplets were bonded to a printed circuit board using silver-based die attachment epoxy. Sections 5.5 and 6.3 of this thesis will present in-depth analyses of the significance of the die attachment material. A number of wirebonds along the exterior of the chiplet were used to make electrical connections during testing. Before operation, an off-the-shelf heatsink was applied to the backside of the PCB to increase heat removal from the system [56].

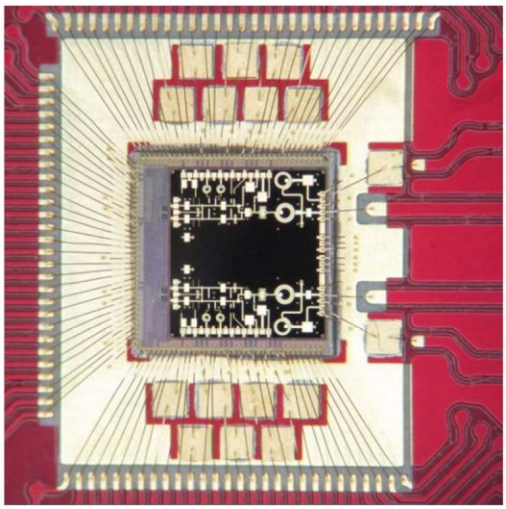


Figure 17: DAHI-Integrated Chiplet Die-Attached to PCB with External Wirebonds [48]

3.2 Submodeling Overview

Submodeling is a finite element modeling technique which leverages two linked simulation sets to achieve increased accuracy and resolution in regions where fine detail and accuracy are required. In situations where large ranges of length scales are present in the same system, appropriate boundary conditions are often only applicable away from the region of interest. On the other hand, simulation accuracy relies on mesh refinements which rapidly increase element count and lead to excessive simulation times. Analysts and designers are thus forced to weigh a tradeoff between gross assumptions of boundary conditions, or poor spatial resolution of results.

In the submodeling technique, two semi-independent simulation sets, known as the 'global' model and the 'submodel' are defined. The global model contains a rough description of the device being studied, with physically realistic boundary conditions imposed on a coarse-grained approximation of the device. An initial simulation, perhaps one with a simplified loading profile, is run on the global model, and the results of this simulation are then passed to the 'submodel'. To obtain boundary conditions for the submodel, the global model's results are

mapped to their equivalent locations at the submodel boundaries. This process is controlled using a node-to-node interpolation scheme determined by the FEM software's capabilities and the scenario being modeled. For example, structural submodeling can map either forces or displacements to the submodel, or in the case of thermal analysis, temperatures and heat fluxes can be used.

Figure 18 illustrates an example of a submodeling approach applied to structural analysis of a bicycle frame. Here, the global model contains a simplified representation of the entire frame modeled with beam elements; loads and boundary conditions have been applied to simulate the weight of a rider. The submodel, shown in the inset image, then represents a small portion of the lower frame which has been modeled to a higher level of detail. Reaction forces from the global simulation results are imported to the cut boundaries of the frame in the submodel. Fine features such as the hub weld geometry can be modeled, and refined meshes can be applied throughout the submodel region to obtain a "zoomed-in" depiction of the frame's behavior under load.

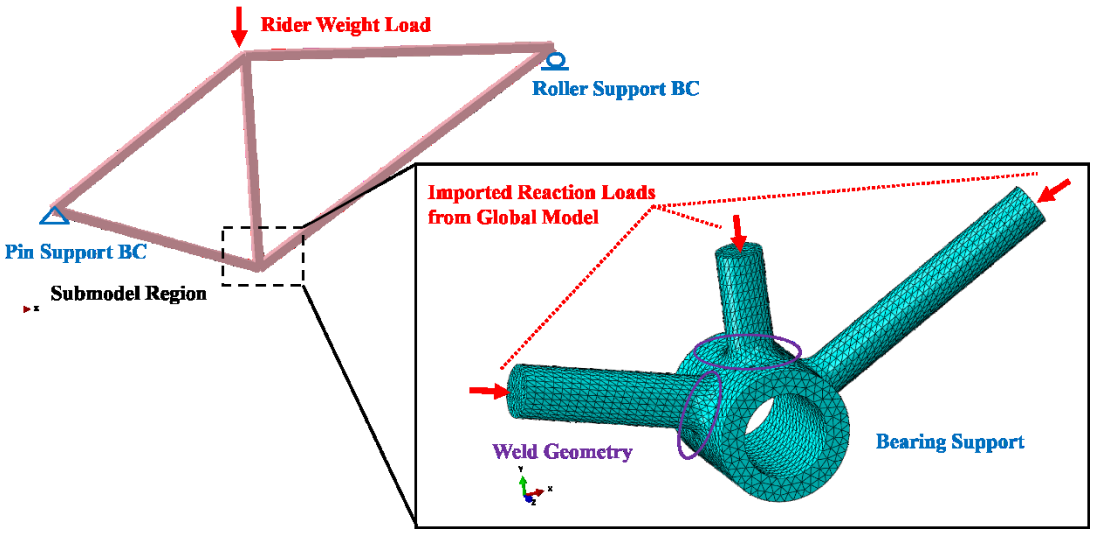


Figure 18: Example of Structural Submodeling for a Bike Frame

Likewise, Figure 19 illustrates a submodeling scheme employed for thermal analysis of an integrated circuit, as is done throughout the entirety of this thesis. Here, a global model is used to coarsely model the IC chip integrated on a circuit board using a relatively coarse mesh. Global loads and boundary conditions describe the estimated heat dissipation within the device, and convective cooling from ambient airflow across the chiplet surface. The global model can also include details such as contact resistance between the chiplet and circuit board surfaces. The segment of the chiplet closest to the heat source is defined as the region of interest for the submodel, which can be modeled with a high degree of mesh refinement to include micro-scale features. Simulated global temperatures at the submodel interfaces are mapped as a temperature boundary condition to the exterior of the submodel.

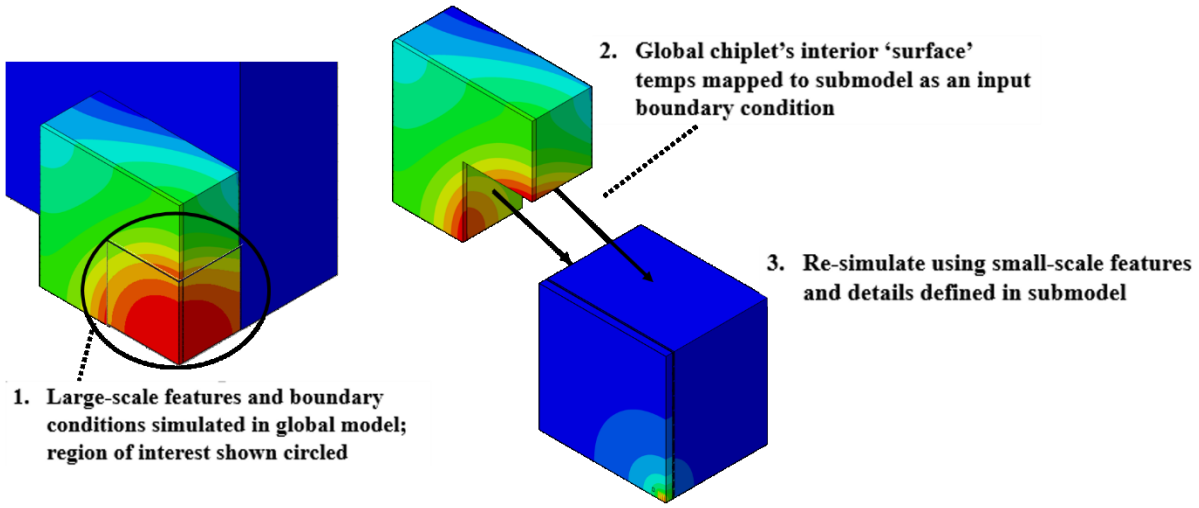


Figure 19: Example of Thermal Submodeling for an Integrated Circuit

3.3 Meshing Techniques

All parts and substrates within the model were meshed with linear hex heat transfer elements (DC3D8) from the standard element library in ABAQUS CAE [57]. In general, mesh sizes were chosen to suit part geometry and refined to capture detail in critical areas, such as within the GaN finger region. Convergence was verified by repeating refinement until subsequent refinements produced negligible change in peak temperature.

Element sizes used in simulation range from ~0.5um in the submodel's GaN fingers to ~500um in the global model's printed circuit board. Total element counts in the submodel and global models are approximately 1.0M and 1.4M, respectively. Typical simulation runtimes for the global and submodels are between 7-10 minutes each. The time-saving advantage of submodeling becomes apparent when comparing the combined global/sub-model runtime to that of an equivalent simulation encompassing submodel-scale detail throughout the entirety of the chiplet. For example, when the global model's abstractions are removed and replaced with instances of substrate layers and additional HICs, element count quickly increase to ~5 million and single simulation runtimes bloom upwards to 50-60 minutes.

Furthermore, submodeling reduces the upfront time required to build and modify a detailed finite element model, especially when exploratory studies require changes to many different parameters in the model. By zooming in on a particular region of interest, changes can be made within the submodel region without requiring changes to the definition of the global model. The global model can then be re-used for further simulations in the submodel. For example, Chapter 5 of this thesis contains several trade studies which examine the effects of changing feature sizes and properties within the device. By leveraging the same global model, continuity is maintained between each family of models while allowing for fast interchange between unique configurations of the submodel.

4 Single Channel Device Model

A finite element model representing the device was prepared in ABAQUS CAE to characterize the device's baseline thermal performance. A pair of high level abstractions were applied to this stage of modeling. Namely, the model included only a single channel, and shifts the location of the Stage 3 PA at the center of the 4 x 4 mm chiplet's top face, as illustrated in Figure 20. The effects of multichannel operation, including increased temperatures and thermal crosstalk, have been investigated and will be discussed later in Chapter 6. Shifting the PA allows the model to utilize quarter symmetry to significantly reduce computational cost without significantly altering the behavior of the hottest regions at the center of the PA. Furthermore, the model neglects surface metallization at the PA, and does not include dissipation from passive components or other active components besides the third-stage PA.

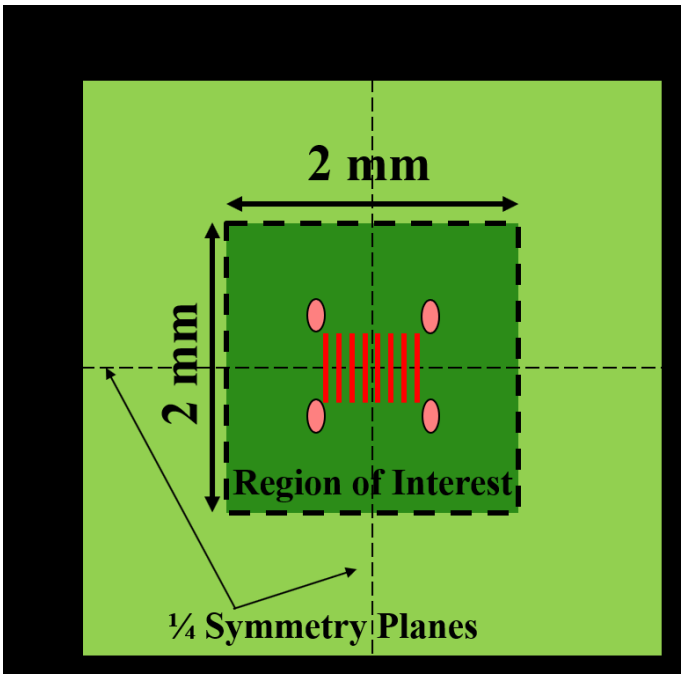


Figure 20: Chiplet Region of Interest Definition (compare to Figure 11)

4.1 Device Submodel Description

Submodeling techniques were used to obtain a highly detailed model of the device in the region of interest surrounding the Stage 3 PA. In this case, the "region of interest" for the submodel refers to the 2 mm by 2 mm area of the chiplet immediately surrounding the PA, from the top of the GaN layer through to the bottom of the bulk silicon substrate. Figure 21 below illustrates the contents of the submodel region corresponding to a 1 mm by 1 mm area after quarter symmetry is applied. In this region the device has been modeled to include all relevant substrate layers at a high spatial resolution, including features at sub-micrometer length scales. As will be discussed below, the HIC & metal stack structures have been substituted for an equivalent material model derived from separate one-dimensional simulations, illustrated below in Figure 22. Both near and far field HICs are included, along with relevant interfacial thermal resistance terms. This allows the model to accurately describe the peak temperatures across

extremely thin features such as the gate fingers, and to illustrate complex thermal interactions between neighboring HICs.

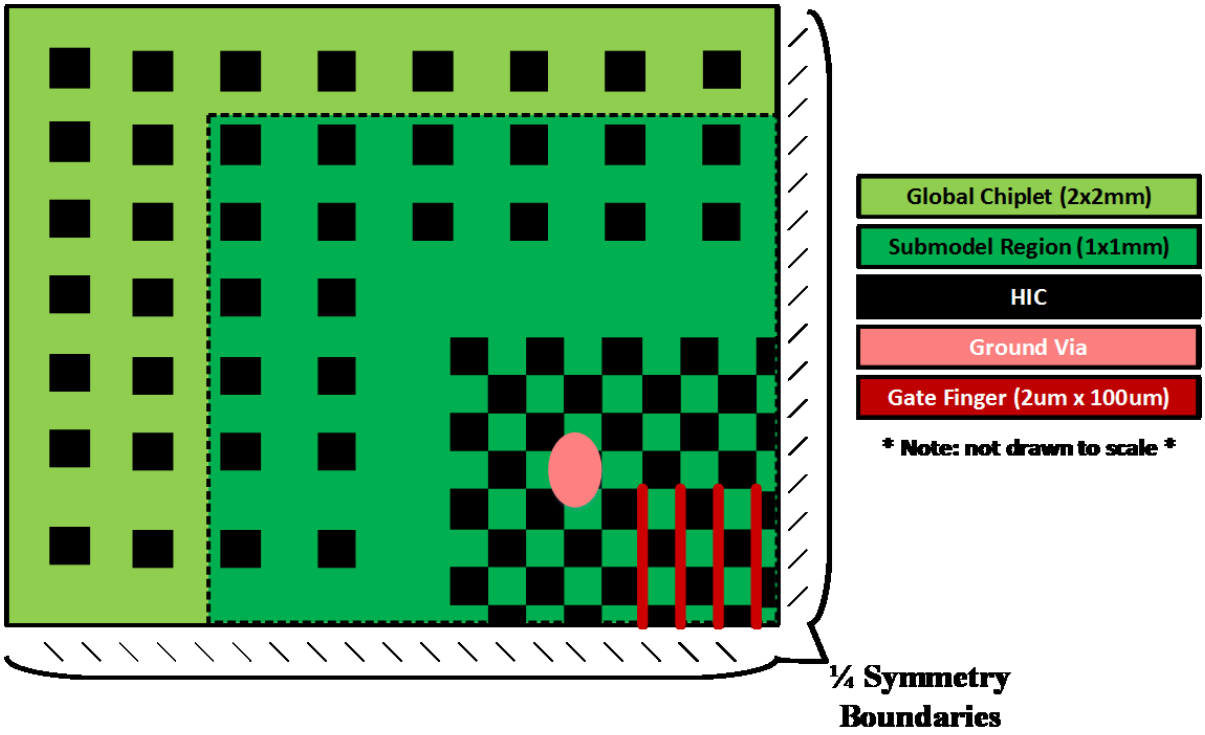


Figure 21: Global and Submodel Illustration

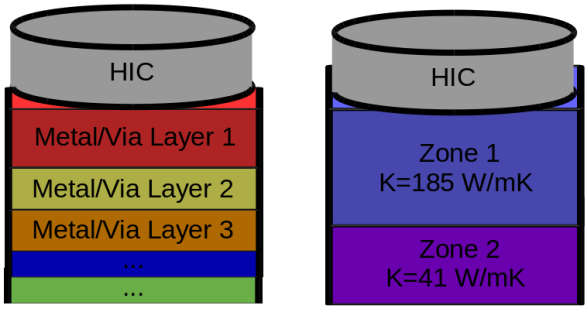


Figure 22: Conceptual Illustration of 2-Zone Model Applied to Metal Stack Structures in Submodel, Discussed Below

The global model includes the remaining chiplet volume, plus printed circuit board to which the chiplet is mounted, and heatsink adhered to the underside of the PCB. The remaining

global chiplet volume outside the submodel region was approximated using equivalent materials meant to describe the bulk behavior of the CMOS and GaN substrates, as depicted in Figure 23. Likewise, the global simulation's HIC interfaces were modeled using 1D thermal boundary conductance values derived for both the near and far field HIC arrays, as will be discussed later. The following two sections will detail two sets of simulations incorporating one-dimensional abstractions used to derive homogenized parameters for use in the submodel, namely metal stack material modeling and global model equivalent properties.

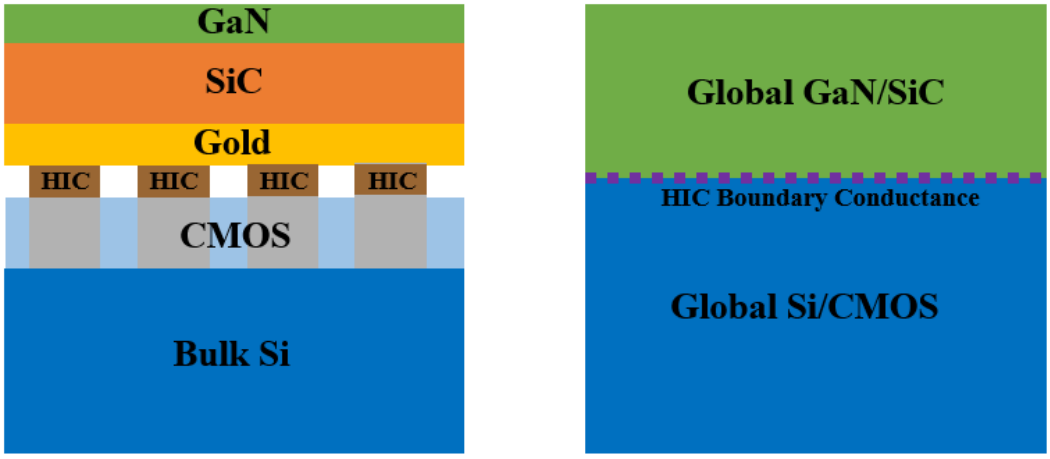


Figure 23: Conceptual Illustration of Abstractions Applied to Global Model's Material Properties

4.2 Metal Stack Simulations, Two-Zone Conductivity Model

Beneath each HIC, a layered network of interlinked copper metal runs through the top ~10um of the CMOS substrate. The primary function of this high-conductivity metal stack, is to remove heat vertically downward and away from the PA's heat source through the low-conductivity field oxide. The thermal conductivity of copper is over a hundred times higher than that of the field oxide, so the thermal pathway through the CMOS layer funnels primarily through the metallic structures, as illustrated in Figure 24 below.

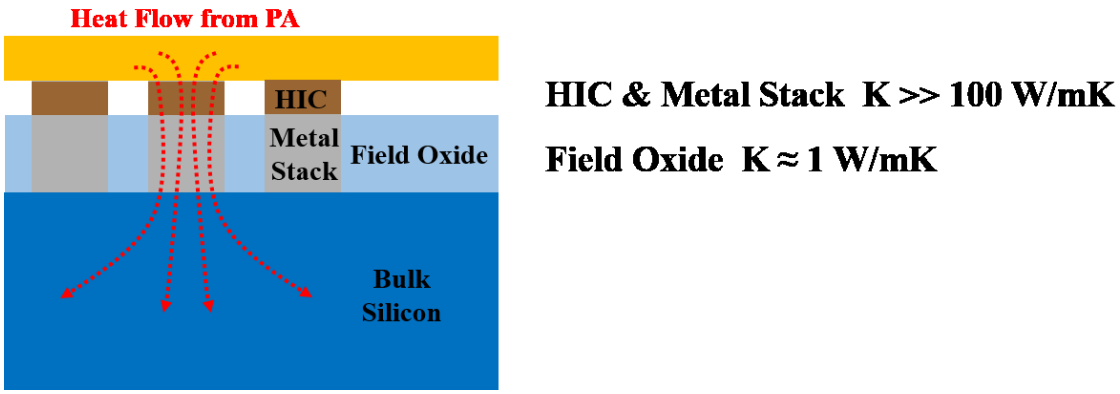


Figure 24: Metal Stack Heat Funnel Illustration

Layers in the stack are comprised of metal strips embedded within field oxide. Strips within each layer are aligned along their long axis, with the orientation of the axis alternating between vertical and horizontal in every subsequent layer. Each layer is connected to its vertical neighbor through a repeating grid of rectangular metal vias. In general, the thickness and cross-sectional size of features on each metal/via layer decrease from top to bottom through the stack, ranging from hundreds to tens of nanometers [58]. Element sizes of 2-5nm in the lower layers of the stack were required to capture the behavior of the smallest via features. When modeled at this scale, each stack contains roughly 1 million elements. If this level of detail were to be maintained across hundreds of HIC instances in the model, the simulation would become intractably large.

Due to the small size and large feature count in the lower via layers, simulations of the device incorporating hundreds of metal stacks are impossibly large and complex. As noted above, low-level vias require element sizes on the order of nanometers, while the global chiplet and board are tens of millimeters in size. Full characterization of individual stacks is not computationally possible, so some homogenization of the metal stack is desirable. Furthermore its gridded via structure contains restrictions and isolations between subsequent layers whose compounded thermal spreading resistances present a tortuous thermal pathway with reduced total conductivity. Thus, an initial 1D study was conducted looking at the HIC & metal stack

structures alone to characterize a simplified but equivalent material in the submodel, as illustrated in Figure 22 above.

In this study, a representative segment of the HIC/stack structure was modeled at full resolution, incorporating nanometer scale features in its geometry. Properties assumed for copper and field oxide within the CMOS layers are given in Table 2. Thin metal layers below 200nm thickness are known to exhibit reduced conductivity, thus the lower layers were assigned the reduced conductivity value shown in the table [54]. Adiabatic boundaries were assumed to correspond with the low-conductivity field oxide which surrounds and isolates the outside of the stack, while the base of the structure was fixed to a constant temperature, as shown in Figure 25. A unit heat flux was applied to the top of the structure, and the resulting temperature profile is plotted below in Figure 26.

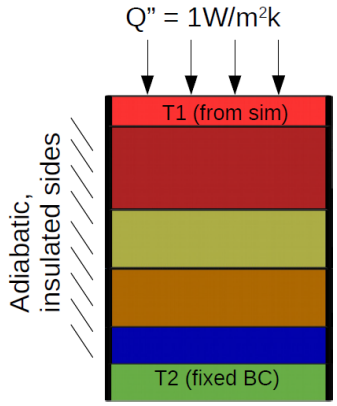


Figure 25: Metal Stack Simulation Boundary Conditions

Table 2: Material Properties Used in Metal Stack Simulations

Material	Thermal Conductivity (W/mK)
Copper (Upper Layers)	400 [54]
Copper (Lower Layers)	220 [54]
Field Oxide	1 [53]

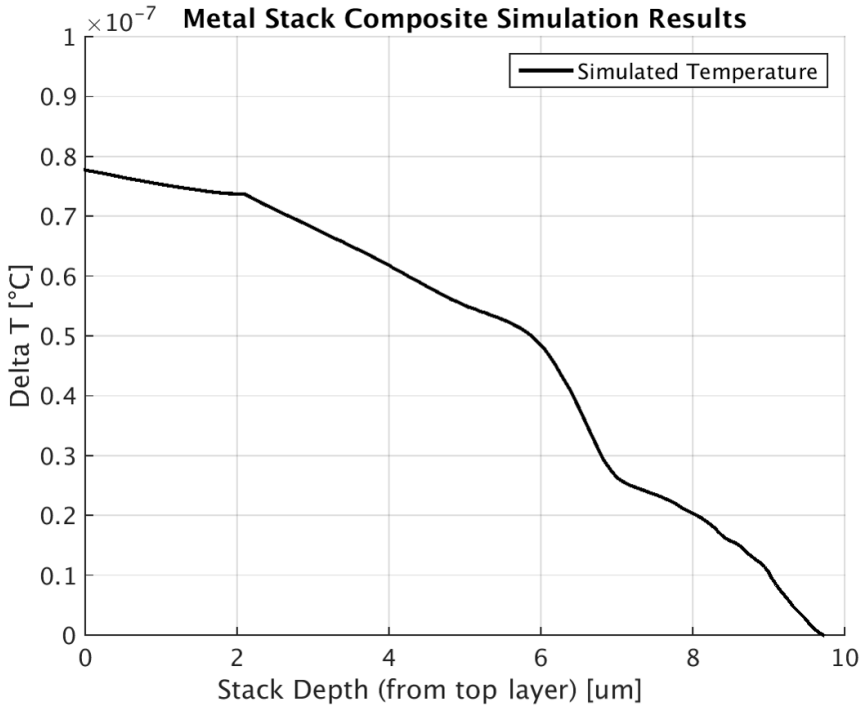


Figure 26: Metal Stack 1D Composite Simulation Temperature Profile

Please note the temperature scale on the Y-axis of Figure 26 is correct given the properties and size of the metal stack under the assumed unit heat flux. The overall length of the metal stack is on the order of tens of micrometers, while its thermal conductivity is on the order of k = 100 W/mK. Likewise, the cross-sectional area of the stack is small ($A \approx 1E^{-12}\ m^2$) compared to the magnitude of the assumed heat flux ($q'' = 1\ W/m^2$) so the total amount of heat dissipated in the stack simulation is very small. Thus, the temperature rise within the stack is expected to be small, on the order of 1e-7 as described below.

$$q'' = -k\frac{dT}{dx} \approx -k\frac{\Delta T}{L} \quad \text{thus, } O(\Delta T) \approx \frac{O(q'')\,O(L)}{O(k)} \approx \frac{(1)(10e-6)}{(100)} \approx 1e-7$$

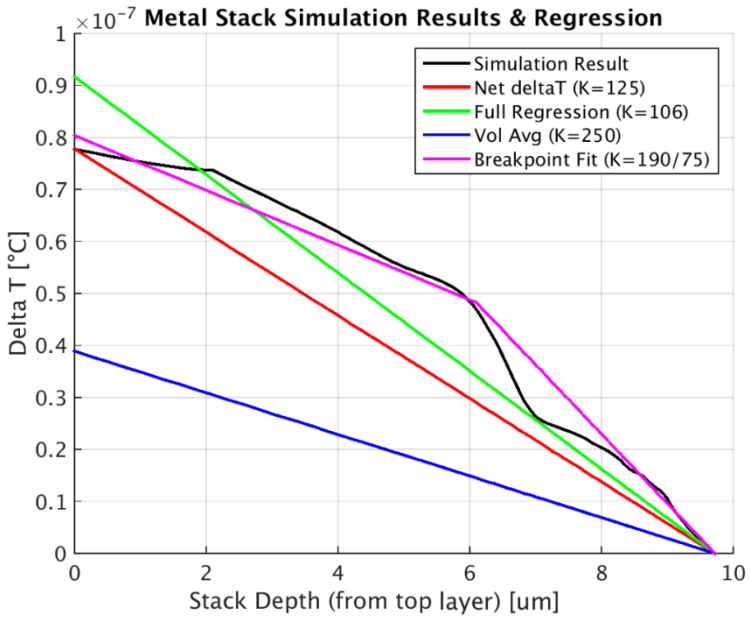


Figure 27: Metal Stack 1D Simulation Results & Regression Modelling

Several models were initially proposed to fit the simulated metal stack temperature profile. Each model was used to compute a set of conductivity parameters for use in a simpler, homogenized representation of metal stack. Stack simulations using each proposed model were then repeated to verify the stack model's fit to the original composite simulation. The re-simulated models' temperature profiles are illustrated in Figure 27.

The first homogenized model, shown in blue, incorporated a volume averaged conductivity obtained by taking the volume fractions of copper and field oxide in the stack, and computing a weighted average of the two materials' conductivities. This method fails to account for the increased spreading resistances incurred at via intersections, and yields a high conductivity value which underestimates the thermal resistance of the stack.

The second metal stack model, shown in green, fitted a pure least-squares linear regression to the entire temperature profile. Per Fourier's conduction law, the slope of the simulated stack temperature profile corresponds to the inverse of its effective thermal conductivity. Thus, the best-fit slope from a linear regression can be used to compute an equivalent best-fit conductivity for the entire structure. As illustrated in Figure 28, the peak temperature predicted by the regression model is significantly higher than the original composite simulation result, indicating that this model generally overestimated the stack's 1D thermal resistance.

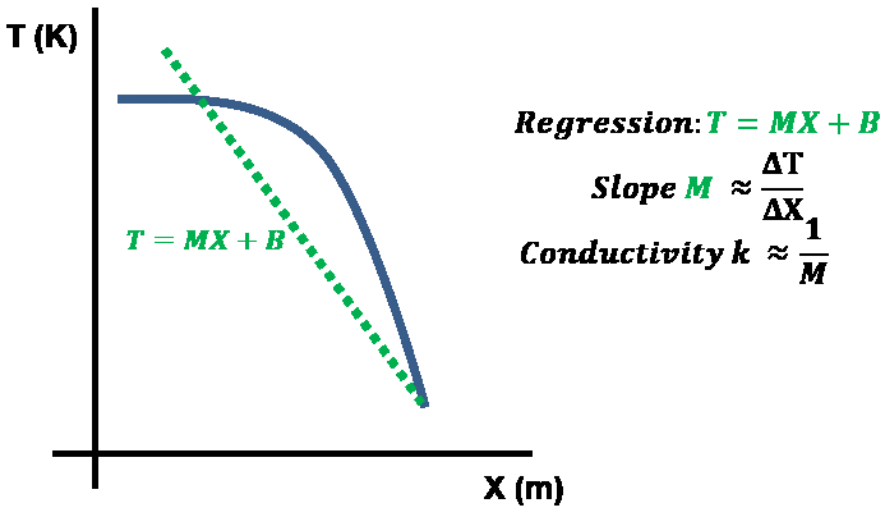


Figure 28: Illustration of Linear Regression-Based Conductivity Model

The third model, shown in red, computed a conductivity according to the simplified one-dimensional Fourier conduction law, which replicates peak temperature but does not match the composite simulation in the middle of the stack. Here, a two-point slope is calculated by subtracting the temperatures of the top and bottom of the stack; the slope is then further manipulated to yield an equivalent two-point conductivity, as noted in Figure 29. Re-simulation of this model reproduces the original peak temperature exactly but otherwise under-predicts temperature in the remainder of the stack.

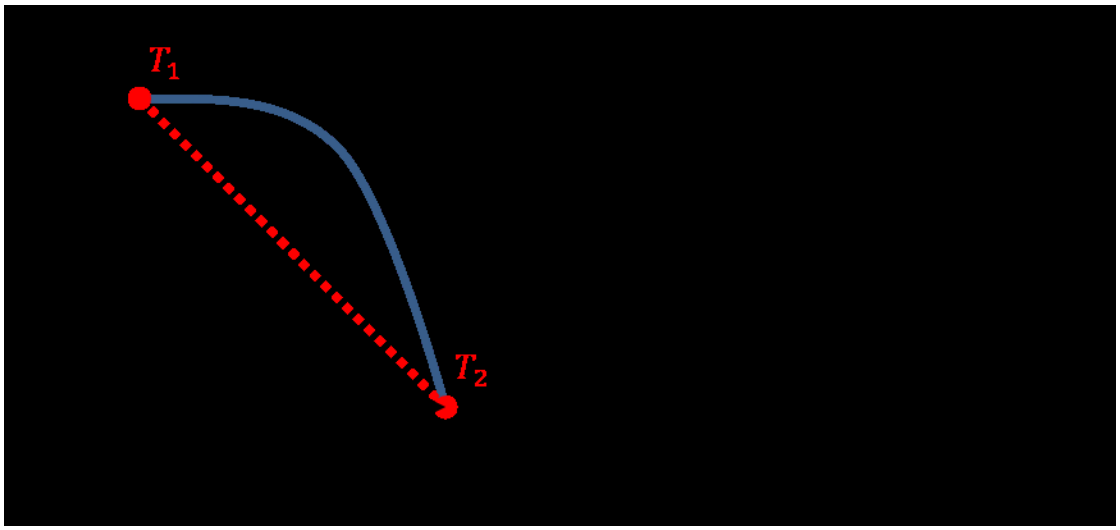


Figure 29: Illustration of Two-Point Slope Conductivity Model

The fourth model, shown above in pink, takes into account that the top three layers of the stack contain much more metal than the bottom three. Examination of Figure 27 reveals that the temperature profile's slope decreases dramatically at an approximate depth of 6um. Here, a two-zone model is proposed to distinguish between the unique behaviors of each region; 'zone 1' corresponding to the top 6um of substrate, and 'zone 2' corresponding to the remaining lower layers. Two separate least-squares regressions were performed for layers in zones 1 and 2, producing net conductivities of K1 = 185 W/mK and K2 = 41 W/mK respectively. As shown below in Figure 30, the two zone model agrees well with the results of the 1D metal stack simulation, and will be used in the rest of the thesis to approximate the properties of the metal stack.

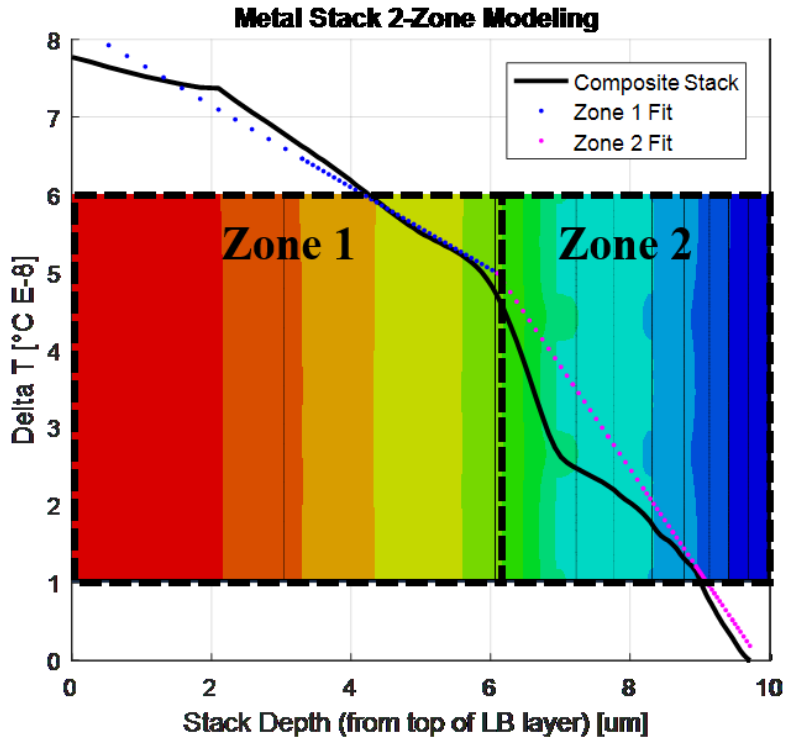


Figure 30: Metal Stack Model 2-Zone Fitting

Deposition of CMOS metal layers during the MOCVD process imparts some uncertainty to the actual thickness of individual layers. Thus, each layer is specified with a max/min thickness tolerance range. A final series of metal stack simulations were performed to compare the performance of the metal stack at its upper and lower tolerance extremes, corresponding to the maximum and minimum metal deposition cases. In either extreme case, the overall thickness of the metal stack changes by +/-2% of stack height. Figure 31 below illustrates the results of the sizing tolerance study following re-simulation of the regression models. In the figure, solid lines indicate the composite simulation results under each of the min/nominal/max metal assumptions. Dashed lines are used to indicate the parameterized model's simulation results after performing the 2-zone regression spelled out above. Here, tolerance variations produce seemingly significant changes to the stack's thermal conductivity; however, when mapped to the full-scale device simulation these uncertainties amount to deviations of <1K variation in peak and local device

temperatures within the CMOS layer. Moving forward, the metal stack tolerances were ignored and the nominal two-zone model conductivities were used to simplify the representation of the metal stack structures in the detailed device submodel.

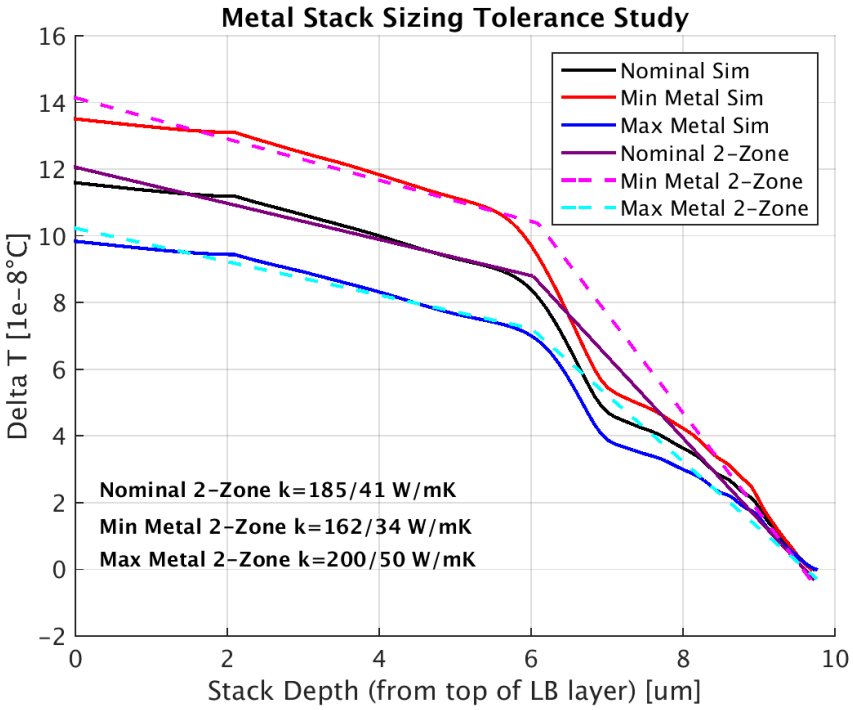


Figure 31: Metal Stack Sizing Tolerance Study Results

4.3 Global Model Material Properties

In the submodeling technique, the intent of the global model is to provide approximate, 'rough-cut' results for inclusion in the detailed submodel. Thus, it is desirable to simplify the properties and geometry of the submodel using some level of abstraction or equivalent modeling. For example, early-stage submodeling efforts initially used in this research assumed a homogenous volume-averaged conductivity for the global model's chiplet. As in the metal stack study discussed above, the volume-averaged assumption was found to be invalid as it significantly underestimates the thermal spreading resistance at the HIC interface. A short series of one-dimensional simulations featuring a simplified representation of the device's through

thickness layers and both near and far field HIC arrays was used to distill the global chiplet's behavior into three parameters for use in the global model, illustrated in Figure 32.

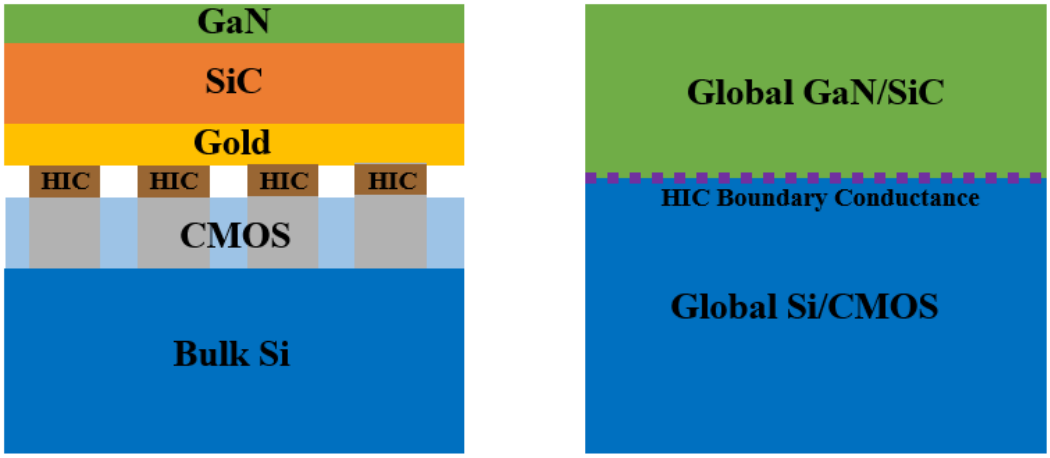


Figure 32: Conceptual Illustration of Abstractions Applied to Global Model's Material Properties

As in the metal stack simulations detailed above, the chiplet was 'sliced' into a representative segment large enough to contain several HIC instances. The near and far field arrays' HIC spacing was then replicated in a separate model for each case. The sides of the stack were given insulated boundary conditions to enforce a one-dimensional heat conduction condition. The base of the bulk silicon substrate was prescribed a fixed temperature boundary condition, and a uniform unit heat flux was applied to the top of the GaN layer. Figure 33 illustrates the rough structure and boundary conditions of the model along with an illustration of both HIC array patterns.

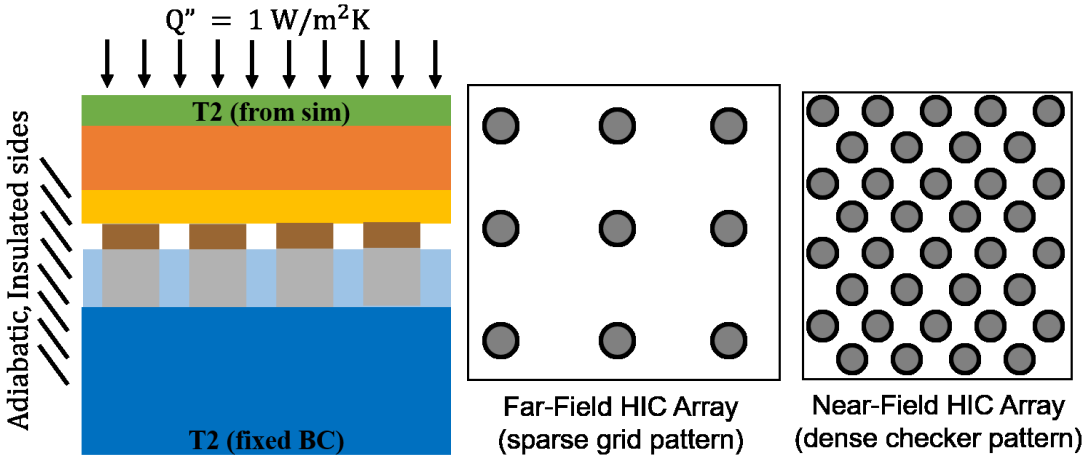


Figure 33: Global Model 1D Simulation Illustration and Near/Far Field HIC Arrays
Note: not drawn to scale

Figure 34 below illustrates the results of the global chiplet simulation and the fitted 3-parameter models for both near and far field HIC arrays; the parameters are listed in Table 3. Here, the blue and red lines plot the temperature results directly from the simulation; note that only the top 100um of substrate below GaN is shown in the plot. The pink and cyan dotted lines plot the results of the simplified chiplet model following regression and re-simulation. Equivalent conductivities for substrates between GaN and the HIC interface, totaling ~55um in thickness, are parameterized via the 'GaN/SiC Equivalent Conductivity'. Those substrates below the HIC interface, roughly ~800um in thickness, are lumped into the 'Si/CMOS Equivalent' parameter. Both conductivity parameters were computed in a manner similar to the two-zone breakpoint model employed above in the metal stack study; as such they are identical for both near & far field cases. The HIC interface creates a thermal discontinuity between the two substrates which is modeled as a thermal conductance $q" = k"(T_1 - T_2)$. Figure 34 shows that the discontinuity is much larger for the far field due to the sparse placement of HICs, whose localizations create a convoluted path for heat transfer between the substrates. In the near field, where HICs are spaced roughly half as far apart in either direction, the substrate temperature

discontinuity is significantly smaller. The reduction in HIC density from near to far field reduces the conductance of the interface by roughly 80 percent. By inspection of the temperature rise in the substrates, this reduction in HIC conductance alone accounts for a doubling of the overall chiplet's one-dimensional thermal resistance.

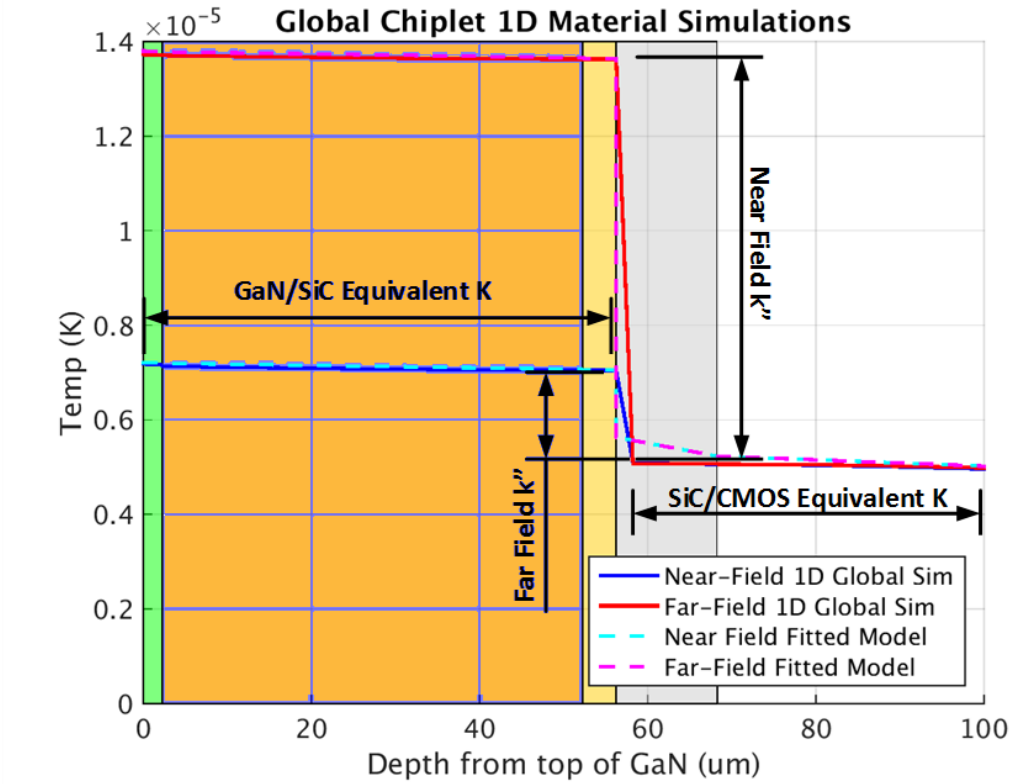


Figure 34: Global Chiplet 1D Simulation Results
Note: GaN/SiC/HIC/CMOS layers overlaid as colors drawn to scale.

Table 3: Three-Parameter Model for Global Model Chiplet

Material/Interface	Value
GaN/SiC Equivalent Conductivity (W/mK)	358
Si/CMOS Equivalent Conductivity (W/mK)	137
Near-Field HIC Interface Conductance (W/m^2K)	7.07e5
Far-Field HIC Interface Conductance (W/m^2K)	1.25e5

4.4 Global & Submodel Boundary Conditions, Loads, Constraints

Figure 35 illustrates the global model in quarter symmetry view; the inset image of the chiplet further demarcates between the global and submodel regions of the chiplet. Global simulations include the entire chiplet region, while detailed submodel simulations include only the dark green region. According to the quarter symmetry assumption, the symmetry planes were assigned adiabatic boundary conditions, and one quarter of the total PA heat dissipation was applied. In the worst case heating scenario of 5W total dissipation, the simulation receives 1.25W as its input. Heat is applied as a uniform surface heat flux to the model. In reality, heat generation occurs throughout volumetrically within the device channel region [31], but due to the thinness of the heated region, surface heating assumptions are commonly employed in similar GaN models [29]. Thus, heat is applied to the gate finger regions in the submodel, while the global model relies on a coarser approximation of a rectangular 50x75um heating profile corresponding to the outer profile of the Stage 3 PA, as represented in Figure 36.

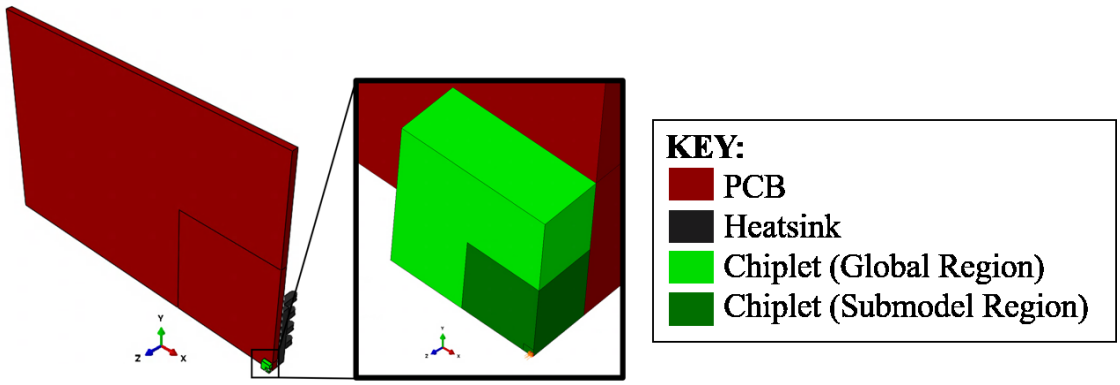


Figure 35: Global Model (left: full global view, right: zoomed on submodel region)

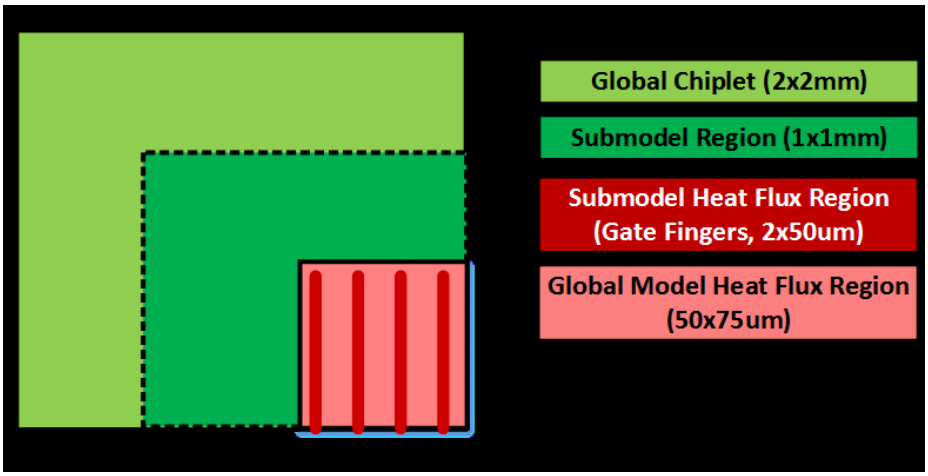


Figure 36: Global and Submodel Heat Flux Regions

In both simulation sets, natural convection conditions in ambient air (film coefficient $h = 10 \frac{W}{m^2 K}$, temperature $27°C$) were assumed for all the external surfaces of the chiplet, circuit board, and heatsink. Similar conditions have been applied frequently in thermal models for other GaN devices [30] [15]. A submodel boundary condition is used to map temperatures to surfaces where the submodel intersects with the global model. Adjacent surfaces' temperatures have all been linked via tie constraints except where significant thermal boundary resistances are expected. In ABAQUS, surface-to-surface tie constraints are akin to fusing the two surfaces' temperatures together, simulating perfect conduction at the interface [57].

Thermal boundary resistances are included in the model at interfaces not suited for the assumption of a tie constraint. At these locations, conduction of heat through the interface is blocked by the presence of a thermal barrier. For instance, the AlN nucleation layer between GaN and SiC is widely-known to act as a thermal boundary [51]. The die-attachment epoxy [45] and heatsink adhesive [56] used to assemble the chiplet and circuit board also function as thermal resistances between interfaces. Furthermore, the global model replaces HIC instances for an equivalent resistive parameter as spelled out in Section 4.3.

In ABAQUS, these thermal boundary resistances are modeled as an equivalent gap conductance interaction between the surfaces indicated in Figure 37. Gap conductance and thermal boundary resistance are two terms for the same phenomena whose parameter values are reciprocals of each other, as described in Figure 38. Values for each interface parameter are given in Table 4.

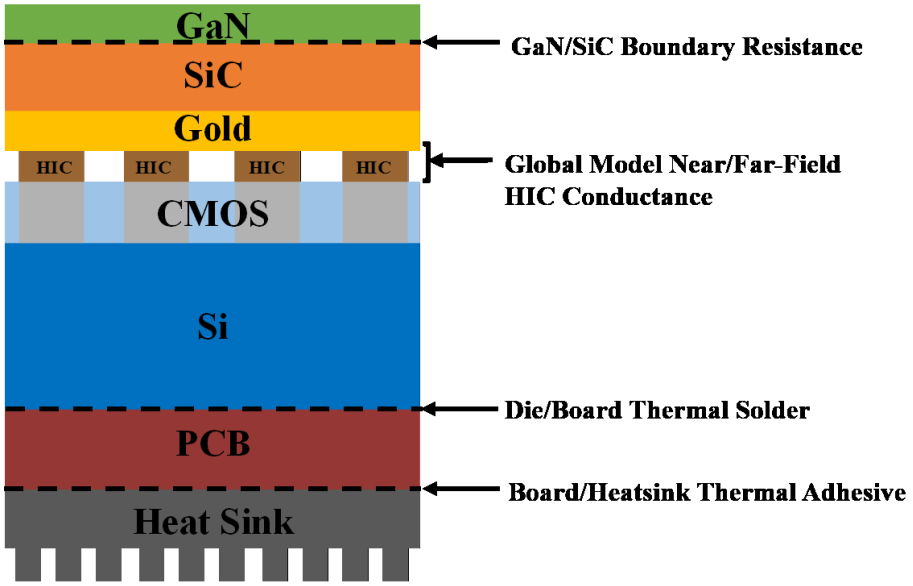


Figure 37: Illustration of Interface Conductance Locations in Device

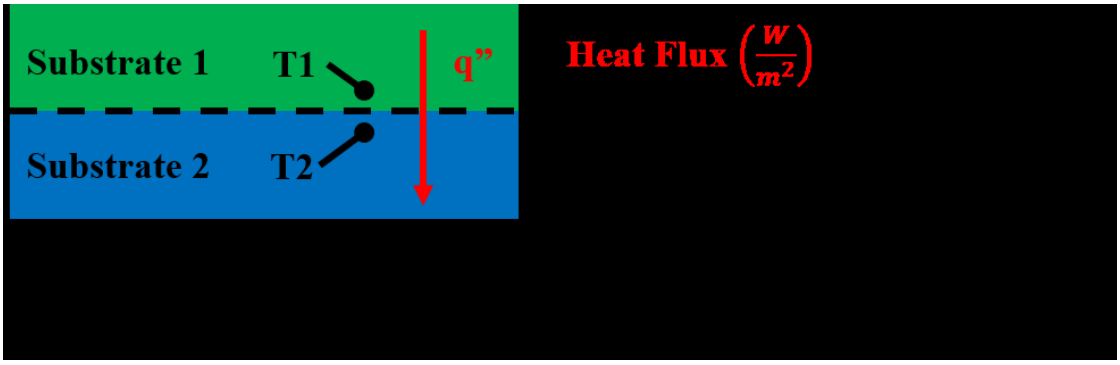


Figure 38: Explanation of Gap Conductance and Thermal Boundary Resistance

Table 4: Interface Conductance Parameters Assumed in the Model

Interface	Interface Conductance (W/m^2K)
GaN/SiC Boundary Resistance	3e7 [51]
Die/Board Thermal Solder	1e3-4e3 [45]
Board/Heatsink Thermal Adhesive	1.40e3
Near-Field HIC Conductance *	7.07e5
Far-Field HIC Conductance *	1.25e5

* Equivalent property derived in Section 4.3 for use in global model only

4.5 Single Device Model Results

Results of the single device model are presented in this section under the assumption of worst-case 5W dissipation and temperature-constant thermal properties. Salient points of the global model are presented first, followed by a discussion of the submodel results.

4.5.1 Single Device Model – Global Results & Discussion

Global simulation of the single-channel device was carried out using the homogenized three-parameter model developed above, and incorporating the board/heatsink configuration deployed during physical testing. Figure 39 depicts the temperature contour of the assembled system. At first glance, the board and heatsink temperatures appear uniform because the regions of elevated temperature are concentrated within a small area on the chiplet. Figure 40 and Figure 41 present the temperature contours of the 4 mm x 4 mm chiplet regions by themselves in zoomed detail with the board and heatsink removed.

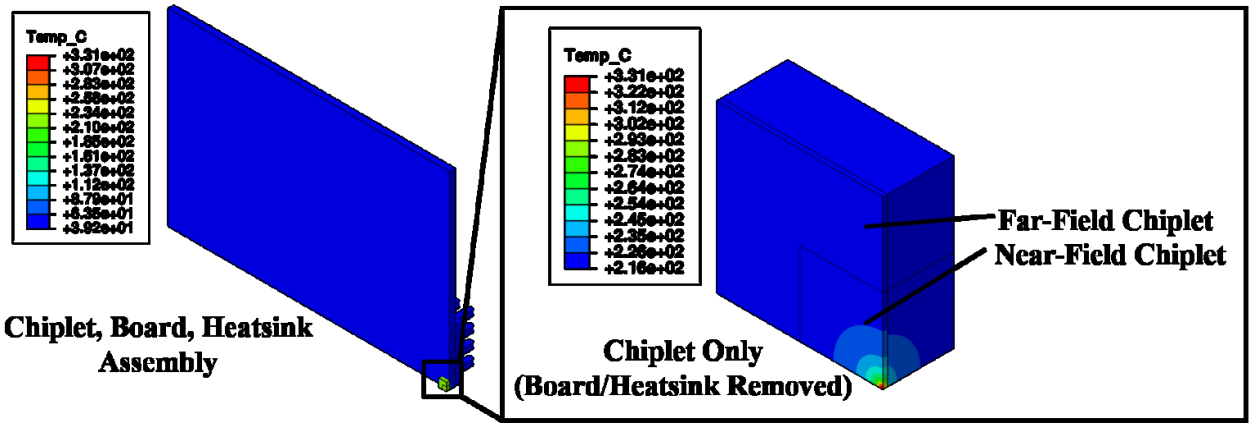


Figure 39: Global Model Temperature Contours (left: assembly view, right: chiplet only)

As indicated in Figure 39, the global model predicts a peak temperature of 331° within the chiplet. Intense variations in temperature are concentrated within small areas of the chiplet. For example, temperature variations >100°C occur in the near-field region dedicated to the submodel, which highlights the need for additional resolution to accurately capture this behavior. Strictly speaking, the global model's peak temperature result should not be taken as a direct estimate of the device's peak temperature in light of the abstractions employed in the global model. Rather, closer attention should be paid to the results shown in Figure 40 which illustrate the behavior of the chiplet in the 'far-field', i.e. far away from the heat source.

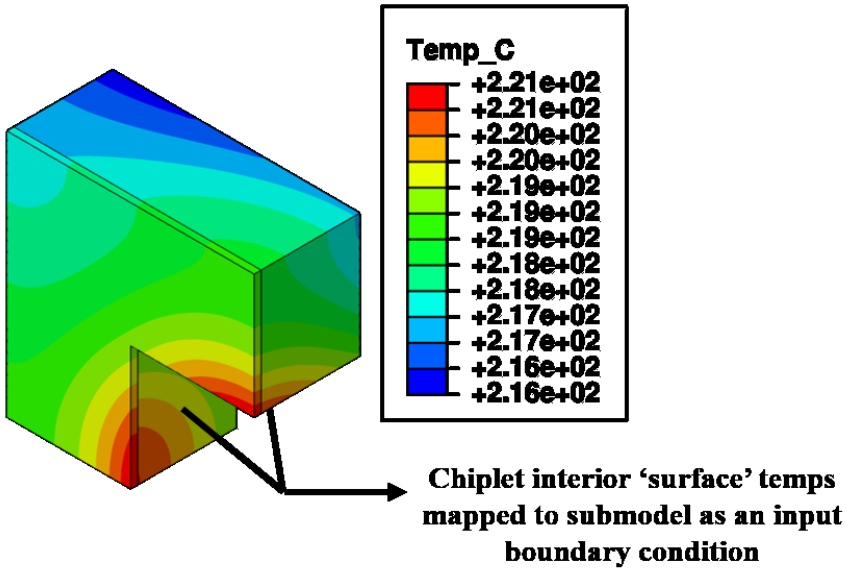


Figure 40: Global Model 'Far-Field' Chiplet Temperature Contour

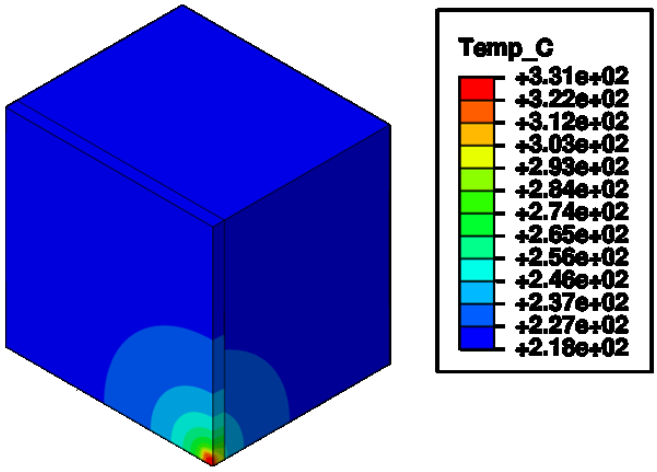


Figure 41: Global Model 'Near-Field' Chiplet Temperature Contour (corresponds to Submodel Region)

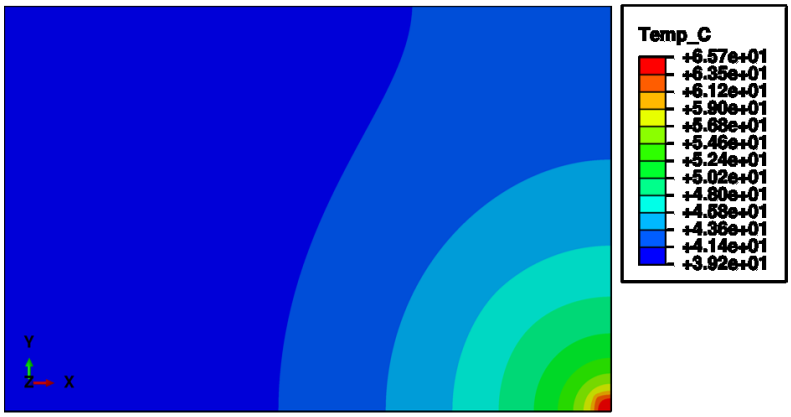


Figure 42: Global Model Board Temperature Contour Detail (Chiplet removed to show board temperatures under die-attachment interface)

In the far-field, the model predicts temperatures of approximately ~220°C at the surfaces corresponding to the interface with the submodel region. Except for regions near the heat source, the entire global chiplet is heated uniformly to temperatures above 200°C. Uniform heating experienced by regions of the chiplet far from the heat source indicate that heat is being trapped within the chiplet. Meanwhile, the majority of the board stays relatively cool at <45°C with a peak temperature of 66°C directly under the chiplet, as shown in Figure 42. The large temperature difference between backside of the die and the PCB is a direct result of the die-attach conductance term ($k'' = 2000\ W/m^2K$). In this case, the die-attachment serves as a high-resistance barrier between the chiplet and the board/heatsink; heat generated within the chiplet therefore cannot dissipate outwards to escape via convective cooling.

One of the arguments for developing a multiscale, submodel-based approach is that global boundary conditions can be more readily adjusted to correspond with the physical reality of the device in operation. For simplicity, similar finite element modeling efforts for GaN-based HEMTs have commonly disregarded global boundary conditions such as die-attachment and employed the assumption of a constant ambient temperature or perfect heatsink to the backside of the die [30] [29]. Inspection of Figure 42 shows that in this case the board temperature

remains relatively constant underneath the die. However, the temperature at this location remains significantly above ambient, which indicates any assumption of ambient temperature conditions for the die backside may not correspond with physical reality.

Furthermore, inspection of the heat flux at the die-board attachment interface sheds light on the influence of convective cooling terms within the model. Figure 43 contains a plot of the heat flux (units: watts) across the die-attachment interface. The values shown in the plot have been weighted by nodal areas such that a nodal summation for the die-attach interface yields the total heat dissipated through the adhesive boundary. Out of the 5W simulated heat generation within the model, 4.94W total dissipation occurs through the die-attachment interface. Less than 1% of total heat dissipates due to convection from the chiplet surfaces; instead, the remaining 99% flows into the board and heatsink to be dissipated via convection from those surfaces.

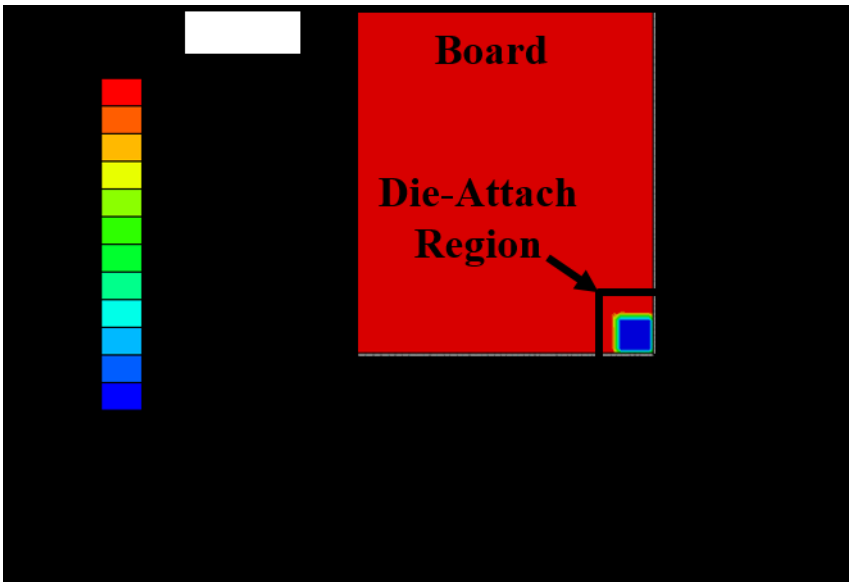


Figure 43: Die-Board Heat Flux

4.5.2 Single Device Model – Submodel Results & Discussion

The global model results presented in the previous section were then supplied as boundary conditions to the submodel simulation detailed in this section. Figure 44 illustrates a 3D view of temperature distribution in the submodel; detail insets to Figure 45 and Figure 46 further illustrate heat flow patterns in the GaN and CMOS substrates. At these conditions, under 5W heat dissipation to an ambient temperature of 27°C, the total device thermal resistance is found to be 67.6 °C/W. Peak temperature of 365°C occurs in the GaN substrate at the center of the innermost gate finger. Here, localized temperature spikes are observed due to the spreading resistance of the finger geometry and its proximity to the GaN/SiC boundary resistance. Temperatures above 325°C do not penetrate into the SiC layer. In fact, the temperature distribution at the top of the SiC substrate appears uniform, such that the SiC substrate experiences heat fluxes from individual gate fingers as if they were 'blurred' together by spreading out within the GaN layer before. Figure 48 contains a plot of temperature taken along the channel symmetry axis which illustrates this observation. In contrast to the sharp temperature spikes within the PA itself, the smooth SiC temperature profile along the symmetry axis indicates a uniform heating pattern at the GaN/SiC interface.

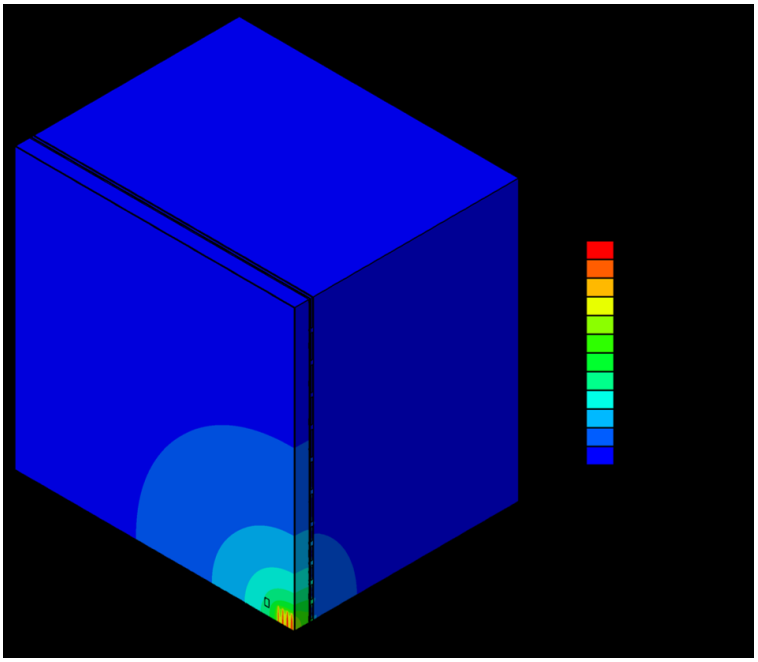


Figure 44: Single Device Model Results, Constant Properties

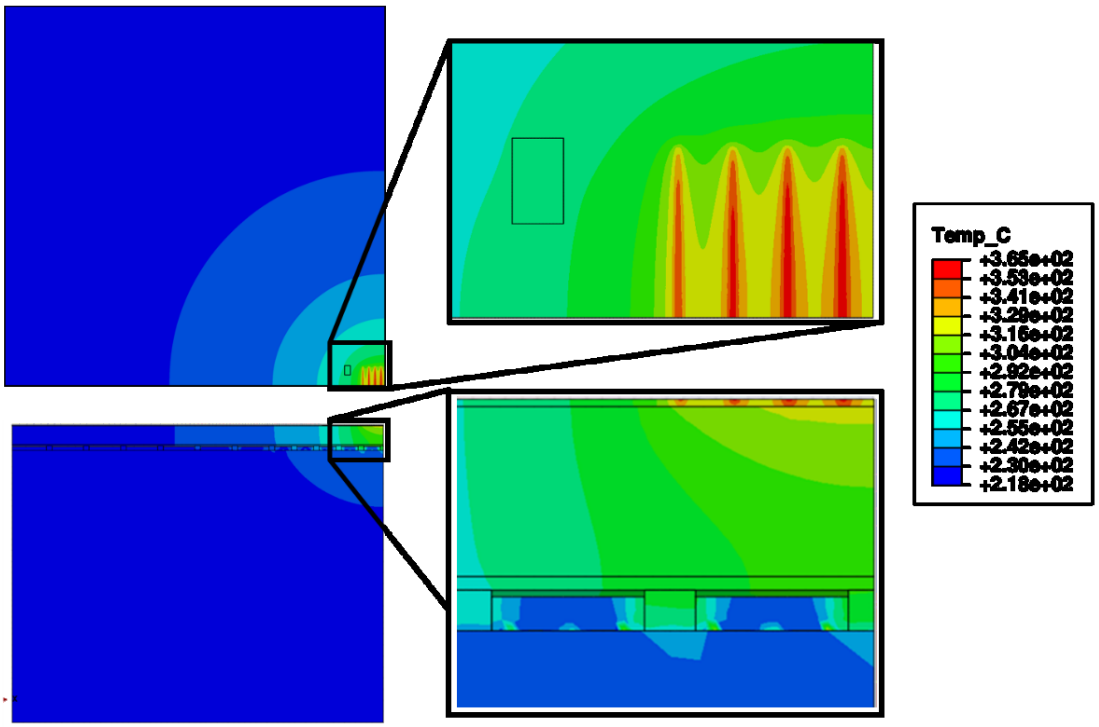


Figure 45: Single Device Model Result Detail View: GaN (top) and Substrate (bottom)

The CMOS substrate, as shown in Figure 46, experiences a reduced peak temperature of 294°C due to its significant thermal isolation from the heat source. Despite the placement of HICs at their maximum allowable density around the PA, only a handful of HICs bear a large portion of the heat dissipation from the PA. These instances experience temperatures above 275°C, while all others experience temperatures between ~250-260°C. This is a direct result of the relatively small size of the PA heat source with respect to the spacing of the surrounding HICs. Under the stated conditions, the temperature of the CMOS substrate can be used to compute the effective thermal resistance of 54 °C/W between the CMOS layer and ambient air.

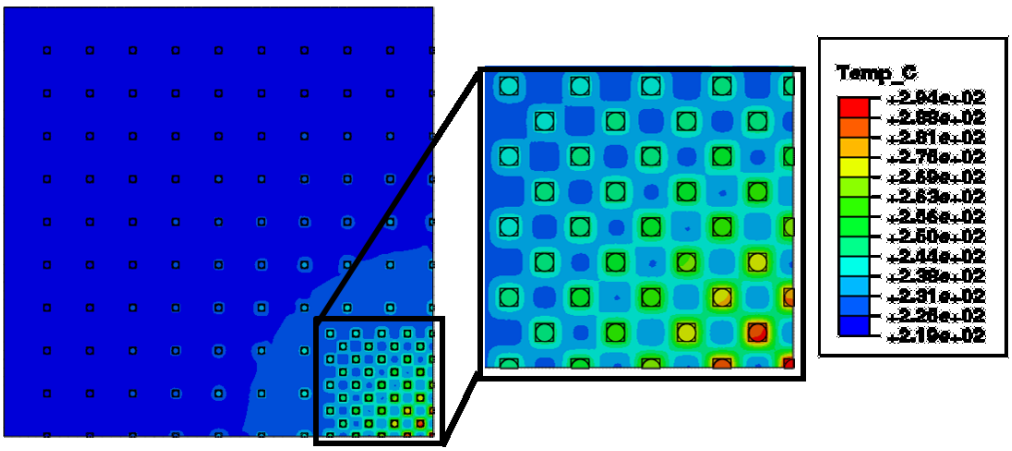


Figure 46: Single Device Model CMOS Detail View

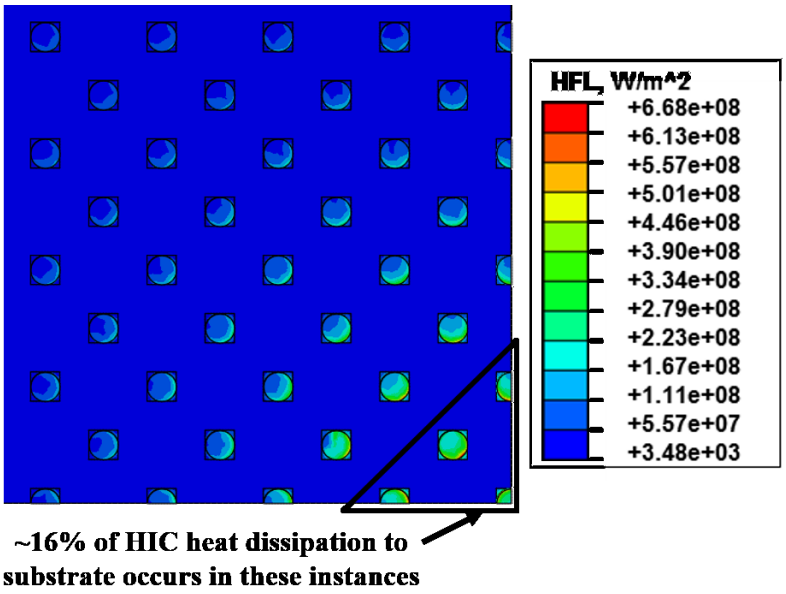


Figure 47: Single Device HIC Heat Flux Distribution

Figure 47 illustrates the distribution of heat fluxes dissipated through the near-field HICs. The highlighted ¼-symmetry region corresponds to nine HIC instances who lie directly under the PA. Due to their proximity to the PA, these nine instances experience disproportionately high heat fluxes relative to the other HICs; out of hundreds of HICs in the device, these nine are responsible for dissipating roughly 16% of heat generated by the device out of the GaN substrate. The small cross-sectional area of an individual HIC's interface is the limiting factor which essentially chokes heat removal from the upper substrate.

Figure 48 again serves as further illustration of the relative isolation of individual HICs. Inspection of the CMOS temperature curve illustrates that while the temperature rise of the central HIC indicates it is conducting away from the PA, temperature quickly drops off between each HIC. Localization of the PA heat source with respect to the HICs thus significantly insulates the CMOS substrate. From the standpoint of a designer seeking to dissipate heat from the GaN PA to lower substrates via the HICs, this thermal isolation is a primary concern as it significantly increases peak temperature. However, as GaN is capable of sustaining higher-

temperature operation than CMOS components employed elsewhere on the chip [59] [60], this isolation could be leveraged to shrink the device size. For example, thermally-sensitive CMOS components could theoretically be placed in closer proximity to the Stage 3 PA without overheating.

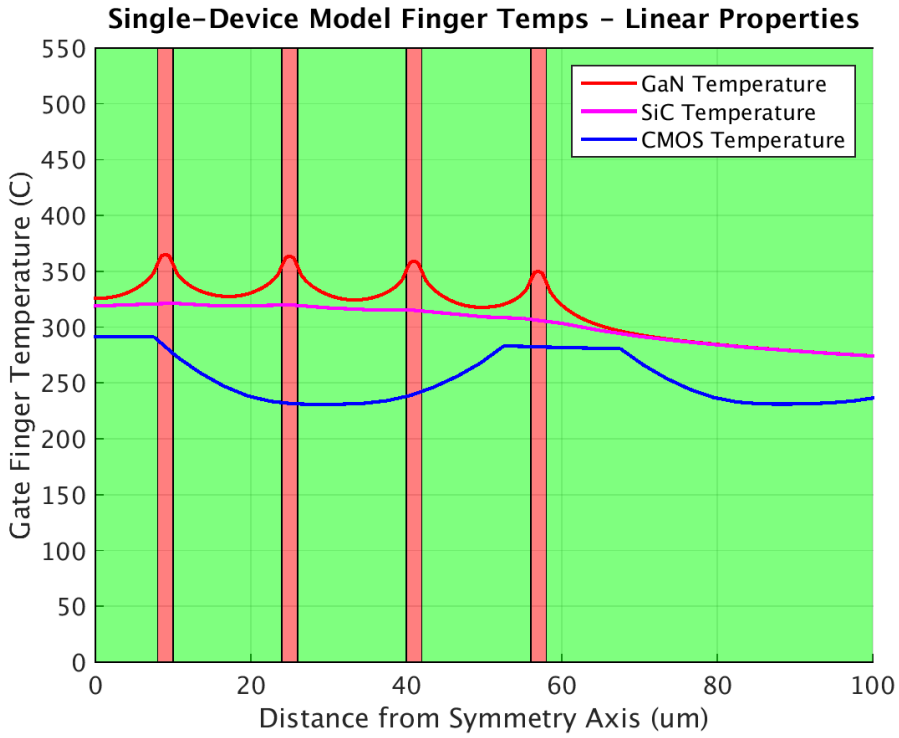


Figure 48: Single Device Model GaN/CMOS Finger Channel Temperature Profile
Note: GaN and Gate Finger regions overlaid as colors drawn to scale.

Figure 49 illustrates the temperature profile taken through the thickness of each substrate layer along the ¼-symmetry axis; only the top 100um of depth is shown. The step discontinuity at ~70um depth occurs at the interface between CMOS field oxide (k = 1 W/mK) and bulk silicon (k = 150 W/mK). At the base of the bulk silicon substrate, the device temperature approaches the far-field chiplet temperature (200-220°C) from the global simulation given above.

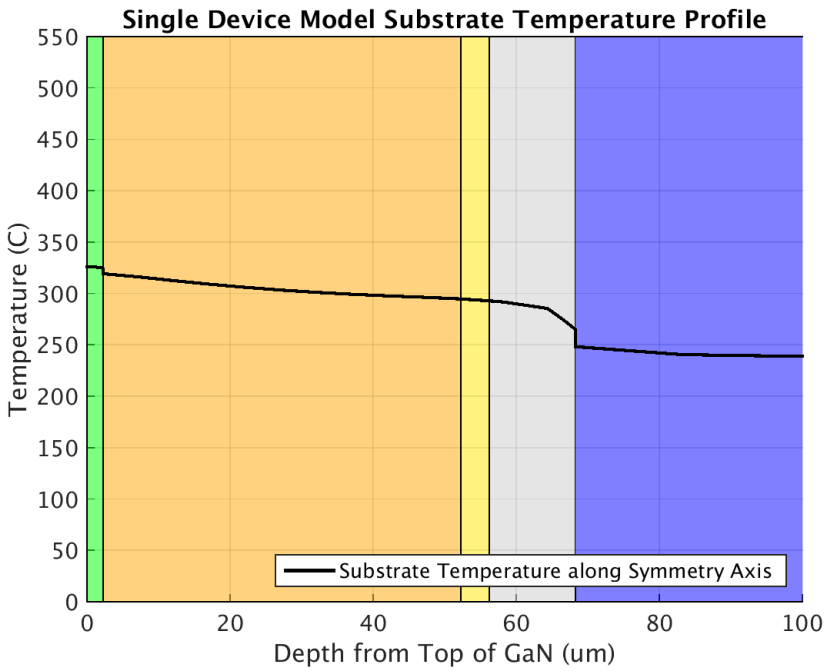


Figure 49: Single Device Model Through-Substrate Temperature Profile
Note: GaN/SiC/HIC/CMOS layers overlaid as colors drawn to scale.

Figure 50 illustrates a direct comparison between the predictions of the global model and submodel. Outside the immediate region of the PA, the global model serves as a good approximation for temperatures in the device. However, it under predicts peak temperature by almost ten percent. Likewise, thermal discontinuity at the HIC interface is not entirely replicated by the simplifying abstractions applied to the global model. Here, temperature is distributed more uniformly throughout the entire CMOS substrate, without localized effects of the HICs. That being said, ten percent disagreement between global and submodels still represents a good approximation, especially since the far field results of both models have converged together.

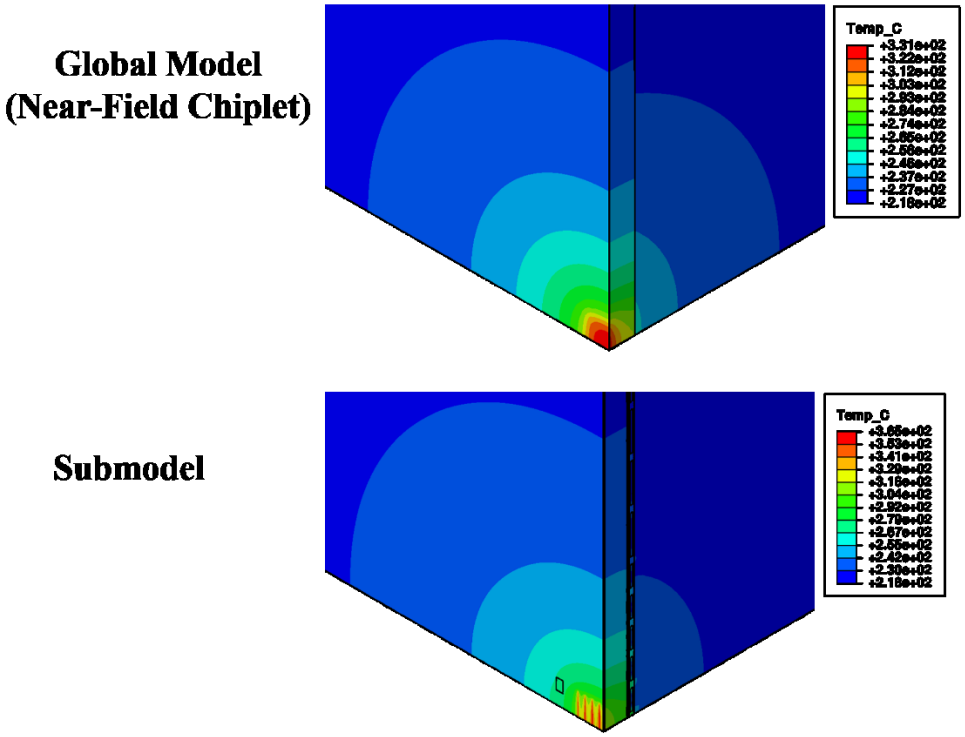


Figure 50: Global Model and Submodel Results Comparison

4.5.3 Aside: Discussion of Assumptions to be Explored in Other Sections

Thermal measurement data is not yet available for the devices being studied in this thesis; specifically, the performance of the die-attachment adhesive in use currently has not been quantified. Interpretation of the single-channel results presented above should carefully consider the significance of the die-attach conductance.

Section 5.5 of this thesis will further investigate the model's sensitivity to this parameter. The particular 'baseline' value assumed here is taken from the middle of a range of measurements performed by Kurabayashi et al. [45]. They studied the performance of various adhesives under of perfect and imperfect attachment conditions for chips of similar size to those being modeled here. Other studies have performed further experiments and modeling on various adhesives and processing conditions, reporting a range of conductances which vary from $1e3 < k'' < 3e5$ $[W/m^2K]$ [47] [61] [62]. The value chosen here falls in the middle of the operating

range for silver-filled epoxy adhesives with moderate voiding. Based on the available data, this assumption serves as a conservative estimate of the adhesive's properties; as a result, the device temperatures shown above are likely over-predicted.

The die-attach conductance parameter becomes particularly important when examining the assumption of temperature-independent material properties. Section 5.4 of this thesis will show that the substrate materials within the device exhibit significantly reduce thermal conductivities at larger temperature variations induced by certain values of the die-attach resistance. Reduced thermal conductivity leads to further temperature increases within the substrate, especially in the case of multichannel operation. The assumption of temperature-independent properties presented above remains useful as a baseline prior to investigating other details of the device's behavior in the following chapter.

5 Further Single-Channel Exploration

This chapter contains a series of short simulation studies conducted to examine the significance of various features and assumptions within the single channel model. These studies are presented in detail here as they provide justification for simplifying assumptions incorporated in more complex multichannel models presented in Chapter 6.

5.1 Gate Finger Heating Region

In the model presented above, heat generation has been assumed to occur over evenly across the surface area of each gate finger; thermal models for GaN devices have traditionally relied upon this assumption. For example, Darwish et al. developed a well-known analytical model for multifinger GaN devices which suggests that reductions in gate length strongly influence increased thermal spreading resistances and corresponding device temperatures [25]. However, other experimental and simulation studies have demonstrated weaker correlation between gate sizing and thermal resistance.

The 3D thermal FEM model developed by Bertoluzza et al. found that varying gate length from 1.0 um to 0.5 um produced marginal 5% variation in peak temperature [29]. The electro-thermal model presented by Heller et al. extended beyond assuming uniform heat flux and includes simulations of Joule heating within the gate region [31]. Results of these simulations were coupled to a 3D finite element model of a multifinger GaN device which showed that device thermal resistance was largely independent of gate length. Moreover, these simulations

found power density within the channel was concentrated toward the drain side, with distribution which varied significantly at different device bias conditions. Bagnall developed a similar series of multiphysics models for GaN devices which again demonstrated a complex relationship between power distribution and device bias conditions [33].

Though the GaN device itself is similar to those studied by Bertoluzza et al., Heller et al., and Bagnall, this device is heterogeneously integrated. Furthermore, the feature size of individual HICs near the PA are on the same order as the PA's dimensions. Thus, minute changes to the assumed heating profile may have a more significant influence on device temperature than was observed in the GaN-only models mentioned above. A series of simulations were developed to explore the effect of varying gate length in our model. Within the submodel, gate finger spacing was left constant while varying gate finger length from the 2 um baseline down to 0.5 um. At each point, the area of the gate was reduced by removing area from the source side of the channel, so as to replicate the effect of concentrating device power distribution toward the drain side, as illustrated in Figure 51.

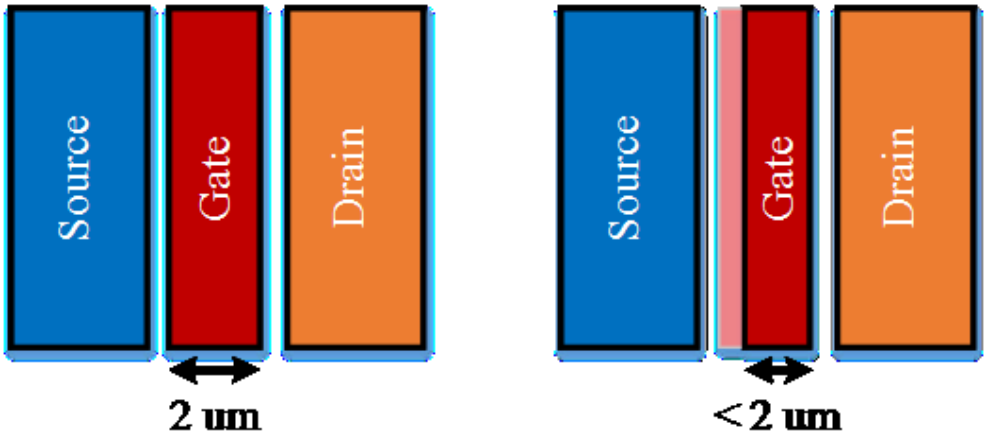


Figure 51: Gate Finger Sizing Illustration

The results of these simulations are illustrated in Figure 52. Reducing the finger length from 2 um to 0.5um increased GaN peak temperatures from 365°C to 384°C, amounting to a 5.6% increase in global device temperature rise above ambient. Peak temperatures within the CMOS layer varied by less than a degree, indicating thermal spreading through the SiC layer largely blurs out the influence of the increased power density at the point heat flow enters lower substrates. Furthermore, the overall temperature distribution of the device was largely unaffected, save for those regions immediately adjacent to the fingers, as illustrated in Figure 53.

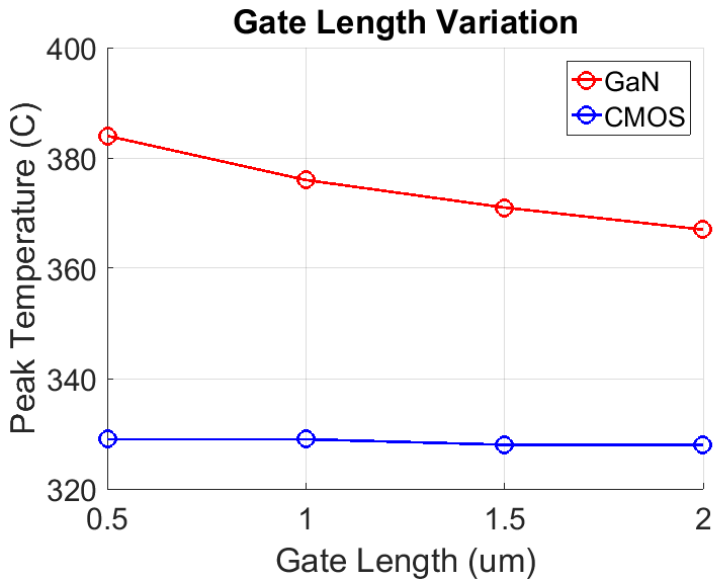


Figure 52: Gate Finger Length Variation Study Results

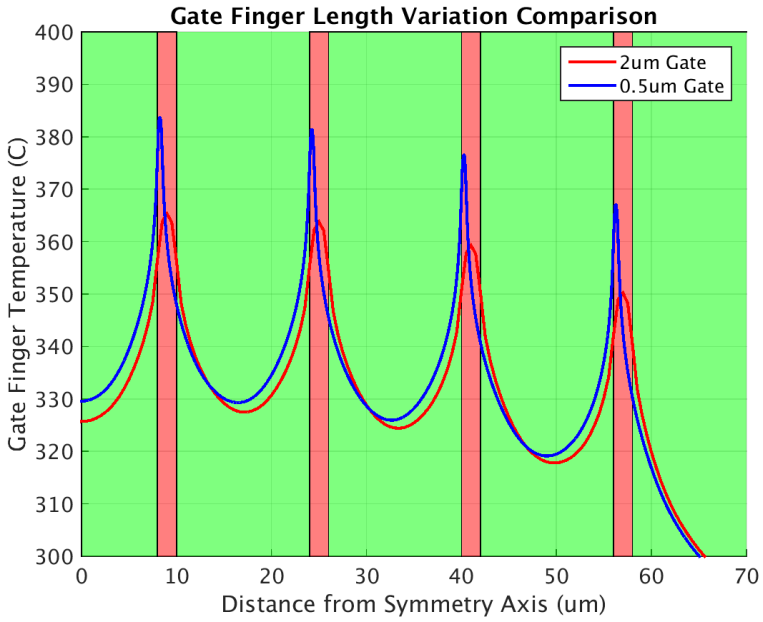


Figure 53: Gate Finger Length Variation Comparison

The GaN temperature results of this study largely agree with those presented by Bertoluzza et al [29]. Thermal isolation introduced by heterogeneous integration of GaN-on-CMOS appears not to influence the behavior of the device with respect to variation of the

assumed heat flux region. For this finding to hold, further testing at multiple bias operation points would be required to confirm that the device's power distribution generally matches the envelope explored in this study. Moving forward, all further simulations in this thesis will assume the nominal 2 um gate finger length and ignore the possible effects of bias conditions of heat generation.

5.2 Ground Via Investigation

At numerous locations throughout the device, through-silicon ground vias serve as electrical interconnections through the thickness of the GaN/SiC substrates [55]. The majority of these instances occur paired with electrical-HIC structures located around the outside of the device; as noted above, these instances exist in low-temperature regions which are not considered thermally significant and have been ignored in this analysis. However, there are four instances of ground vias at each of the corners of the Stage 3 PA which lie within the high-temperature region of the device. These vias consist of hollow gold metal, elliptical in cross-section, roughly 15 by 25 um in size, connecting directly to the ground plane at the bottom of the SiC substrate. The thermal conductivity of the gold via (k = 314 W/mK) is relatively high, and its placement and interconnection to the backplane suggests that it serves as a significant thermal pathway for heat to dissipate through the device. The single-channel device model presented above in Section 4.5.2 included a representation of the gold via modeled as a rectangular cross-section of roughly equivalent size, depicted in Figure 54. The plating thickness inside the via was taken to be 4 um, same as the thickness of the gold backplane.

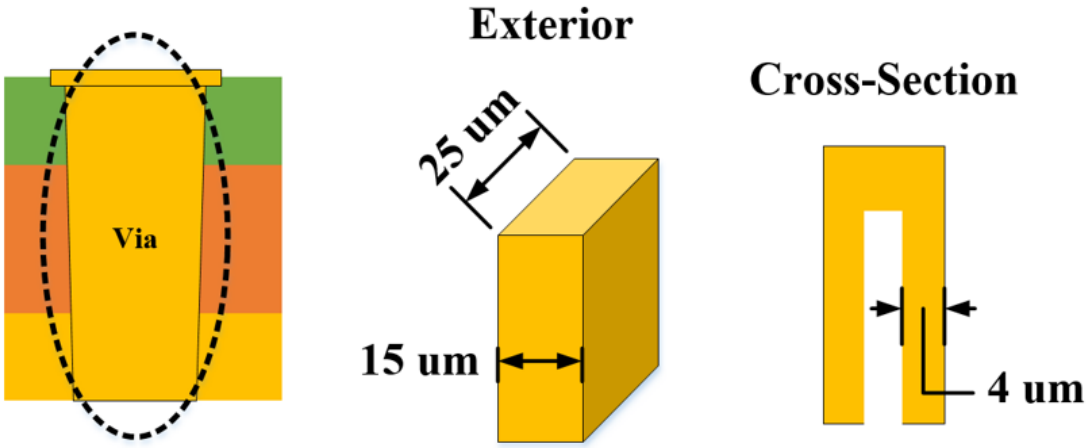


Figure 54: Gold Via Dimensions (as modeled)

Including the via in the submodel adds complexity as numerous tie constraints are required to link the via surfaces to each of their neighboring substrates. In large, multichannel device simulations, it would be advantageous to reduce complexity by altogether ignoring the vias near the PA. Furthermore, while gold's thermal conductivity (k = 314 W/mK) is relatively high compared to GaN (k = 160 W/mK) or silicon (k = 148 W/mK), the gold via primarily passes through SiC (k = 400 W/mK). While at first glance inclusion of the gold via may be thought to yield better heat dissipation near the PA, its conductivity is ~30% lower than the surrounding SiC instead appears to *reduce* heat dissipation from the GaN layer. Bearing these findings in mind, a set of simulations were developed to examine the influence of the gold via on the device's temperature distribution, and to determine whether or not it can be neglected in further analyses. After removing the via from the submodel and 'filling in' the resulting void with the appropriate substrate materials, the model was re-simulated for comparison against the baseline single-device model results presented in Section 4.5.2.

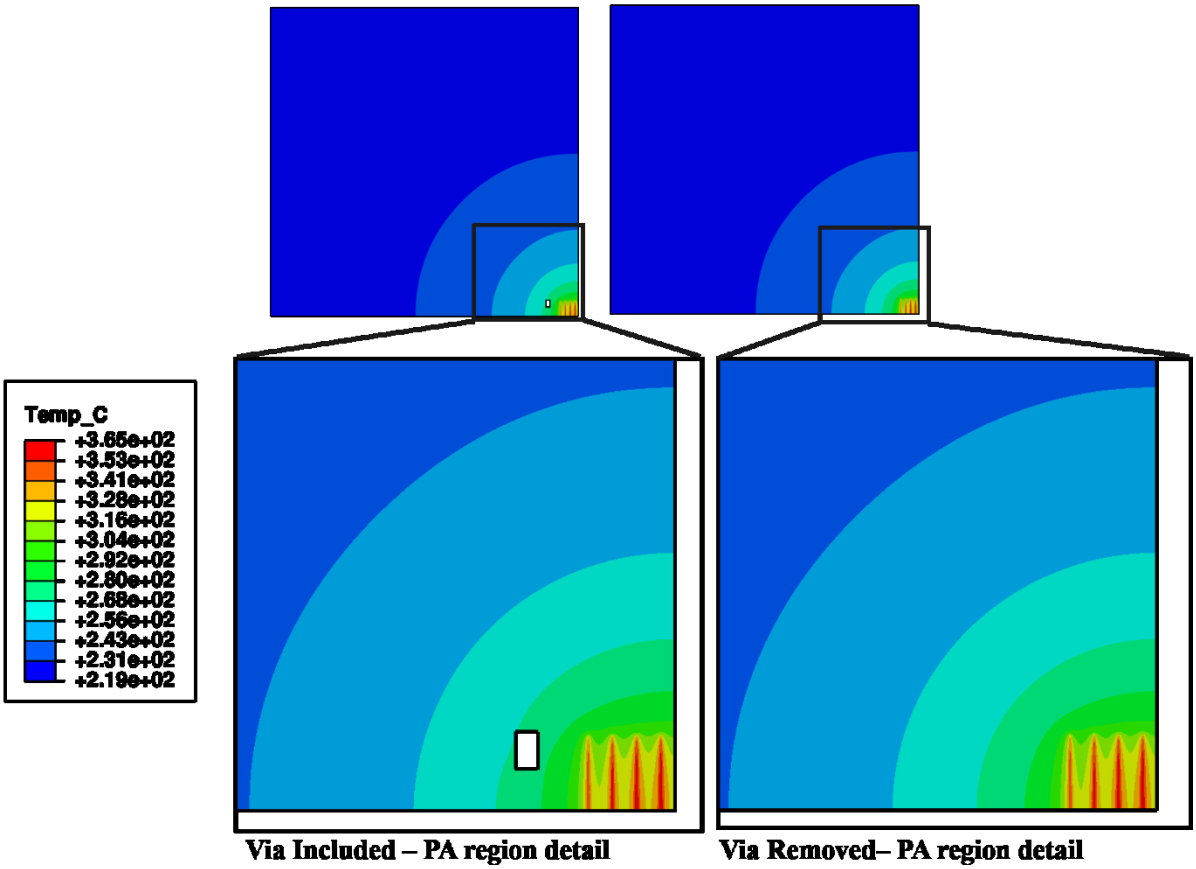


Figure 55: Via/No-Via Simulation Result Comparison

Figure 55 above illustrates that removal of the ground via has negligible effect on the results of the simulation. Peak temperature of 365°C is predicted in both cases to within << 1°C total deviation. At left, the single-device model presented in Section 4.5.2 predicts temperatures near the via of 280-300°C. At the corresponding location in the no-via simulation, local temperatures match to within 3-5°C. Temperatures in substrates below GaN are also largely unaffected by removing the via. As suggested above, the thermal conductivity of the gold via and its surrounding SiC substrate are not different enough to impact the device's overall heating behavior. The via is thus shown to be largely ineffective at dissipating additional heat through the substrate; the substrate itself dissipates just as much by itself. Moving forward, subsequent thermal simulations presented in this thesis will neglect to include the gold ground via.

5.3 Material Sensitivity Study

Uncertainties in the assumed values of thermal conductivity are one of the primary sources of possible error within the models presented in this thesis. Measurements of thermal conductivity in GaN [63] and SiC [64] substrates vary significantly from source to source, and at best carry 20% and 10% experimental uncertainties, respectively. Furthermore, the thermal conductivity of molecularly-grown GaN has been found to be strongly dependent on the parameters of the growth process used [65]. Like the GaN substrate conductivity, the thermal boundary resistance of the GaN/SiC interface is sensitive to growth parameters [51]. The combined effects of these uncertainty terms may lead to thermal simulation seeming less reliable than measurement, as suggested by Kuball et al. [35]. Their study compared micro-Raman measurements of device temperature in GaN HFETs with results from a 3D finite element model, finding only fair agreement between the two. For various configurations of finger lengths and spacings, the FE model either over or under-predicted the device's temperatures and thermal resistances.

Such findings might seem to encourage experimental measurement in lieu of developing thermal simulations for GaN devices until the device's material properties are better understood. However, at this time there are a limited number of functional devices available for testing; heating one up and risking its failure during a thermal test carries significant risk. As such, experimental capability is limited, yet some effort must be made to understand the uncertainties present in the model. This section describes a series of simulations which vary the assumed properties of primary substrates GaN, and SiC in in order to quantify the model's built-in temperature uncertainty.

The assumptions of these models were maintained from the simulation sets described in prior sections, except for the changes described below. In separate simulations, the thermal conductivity of GaN and SiC were independently varied by +/- 20% to approximate the experimental uncertainty ranges noted above. Likewise, the thermal boundary resistance at the GaN/SiC interface was varied by +/-20%. In addition, a simulation incorporating orthotropic properties was included for the SiC substrate. Due to its hexagonal crystalline structure, the thermal conductivity of SiC is known to be orthotropic, with increased conductivity in the cross-plane (k_z) direction [64]. Finally, two scenarios combining the low and high conductivity values of all three materials were simulated to estimate the total temperature uncertainty built into the model. For reference, Figure 56 re-illustrates the substrate temperature profile originally presented in Section 4.5.2 for the nominal material case. Temperature profiles in the following figures are plotted through the substrate layers along the ¼ symmetry axis depicted in the inset image.

Table 5: Material Property Variations Included in Sensitivity Study

	Thermal Conductivity (W/mK)		
Material	**Nominal k**	**-20% k**	**+20% k**
GaN	160 [29]	128	198
SiC	380 [29]	312	468

	Interface Conductance (W/m^{2}K)		
Interface	**Nominal k"**	**-20% k"**	**+20% k"**
GaN/SiC TBR	3e7	2.4e7	3.6e7

	Thermal Conductivity (W/mK)		
Material	**Kx**	**Ky**	**Kz**
Ortho SiC [66]	390	390	490

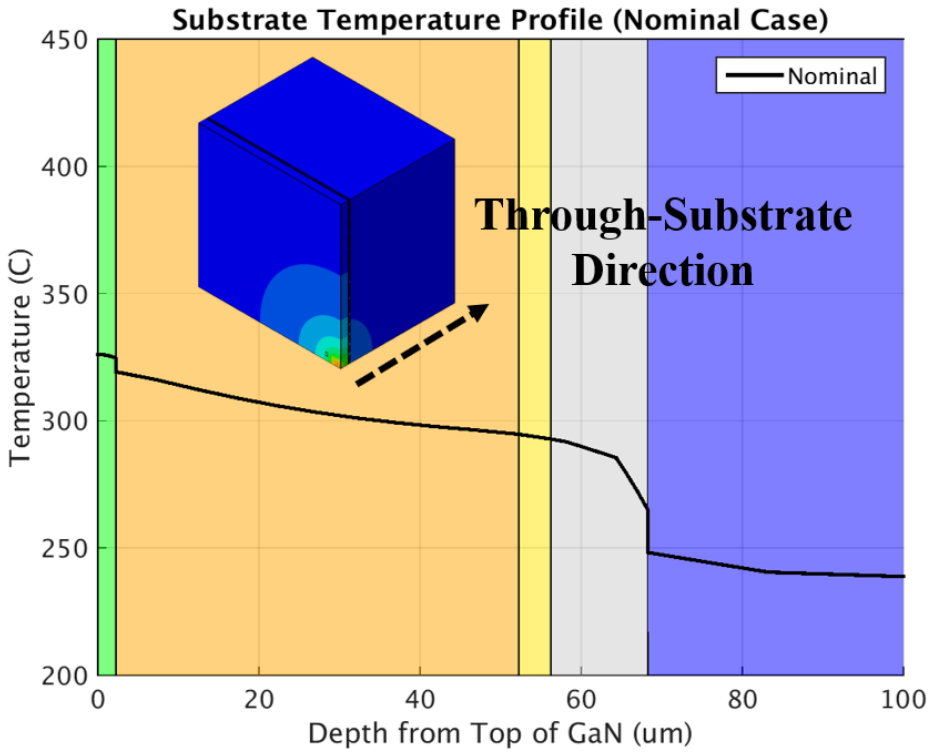


Figure 56: Substrate Temperature Profile, Nominal Material Case (repeated from Section 4.5.2, with notation of through-substrate direction)

Figure 57 illustrates the re-simulated results of varying GaN's conductivity. Here the nominal case is depicted as a black line, while red and blue lines correspond to the -20% reduced and +20% increased conductivity cases. Here, both variations change the local temperatures of the GaN layer by <1°C without changing temperatures in lower substrates. Similiarly, Figure 58 illustrates that variation of the GaN/SiC interfacial resistance term changes the local GaN temperature by a few degrees without changing temperatures in the lower substrates. Changes on the order of one or two degrees amount to less than one percent uncertainty in temperature rise within the device; in this particular case, GaN-related property uncertainties is not likely to be a significant source of error in the model.

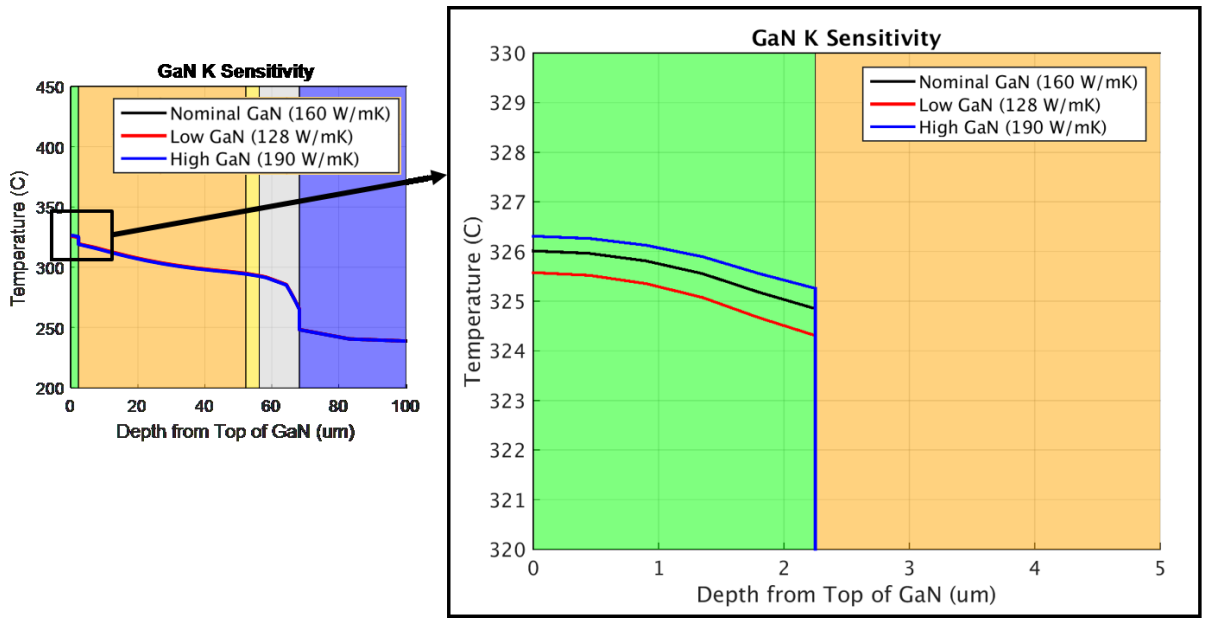


Figure 57: Substrate Temperature Profile, GaN Sensitivity Study

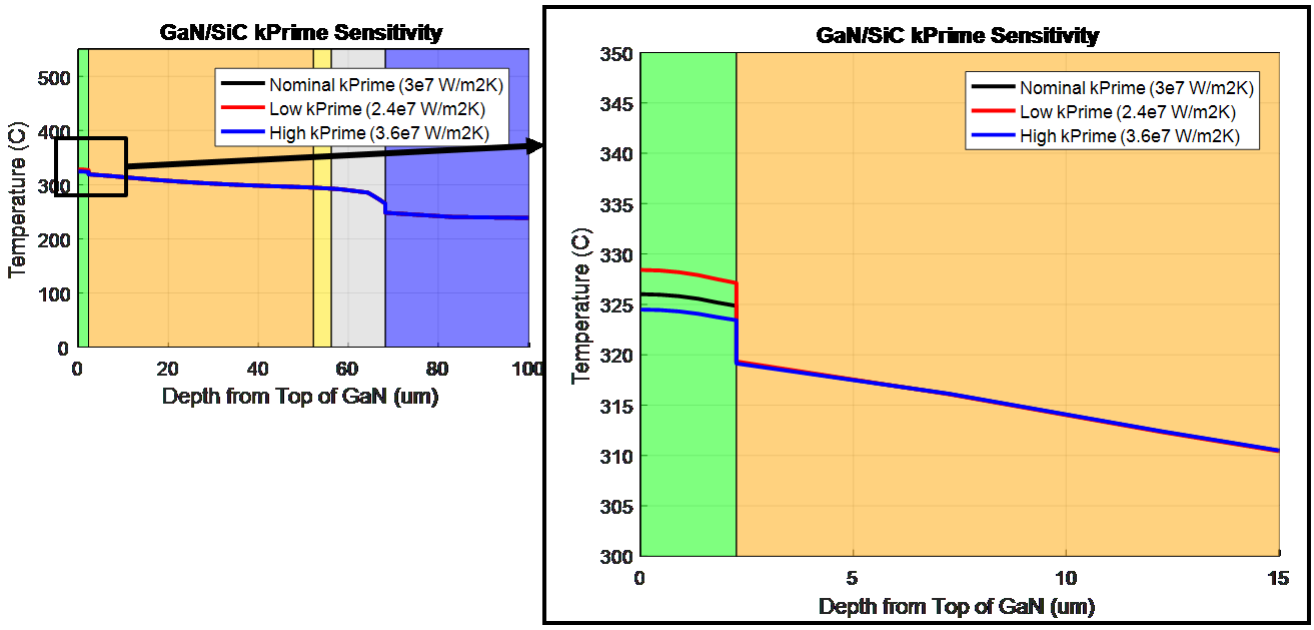


Figure 58: Substrate Temperature Profile, GaN/SiC TBR Sensitivity Study

The lack of change in lower substrates is best explained with an analogy to electronic circuits depicted in Figure 59. In this device the GaN layer is thin, so heat cannot spread out laterally in the layer as it travels through its thickness. Thus, the GaN layer appears to behave as

a one-dimensional thermal resistance between the PA heat source and the top of the SiC layer. In this manner, the device substrates are analogous to as a chain of resistors linked in series, with the thin GaN layer forming the first resistor in the chain.

Each substrate's thickness and thermal conductivity (k) determine its thermal resistance. A final resistance determined by the convective cooling term (h) links the base of the chain to ambient temperature, akin to electrical ground. The heat flux being dissipated through the device is analogous to an input current flowing through the substrate chain; as current is conserved in an electrical circuit, heat is conserved in this mechanical system. Because the current is held constant in this example, increasing resistance in the top link raises the voltage level of only the top link and does not affect those links further down the chain. Similarly, decreasing GaN's conductivity is expected to increase temperature in only the GaN layer.

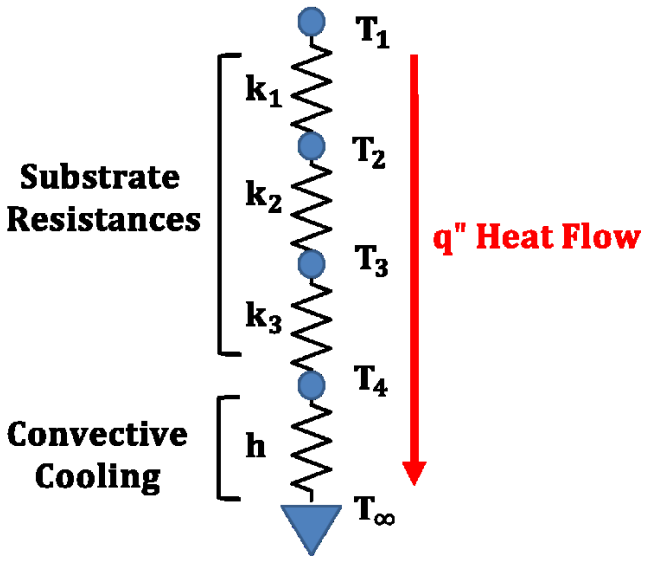


Figure 59: Substrate Temperature Circuit Analogy

This linear, one-dimensional example is not a perfect analogy for lower substrates with larger thicknesses where thermal spreading becomes three-dimensional. For example, the patterning of HICs causes heat to be removed non-uniformly from the base of the substrate.

Spreading out of heat within the SiC layer to multiple HICs renders the problem nonlinear, and the one-dimensional analogy quickly breaks down when the properties of the SiC layer change, as will be explained in the following paragraphs.

Figure 60 presents the results of varying SiC conductivity. As before, red and blue lines indicate the reduced and increased conductivity cases. Here, SiC's uncertainty causes the model's peak temperature to vary by approximately -10 to +20°C from the baseline case, amounting to approximately 7% model uncertainty. In each case, temperatures in the top 100 um of substrate are noticeably changed, including in the CMOS and Si substrates below SiC. This conflicts with the behavior suggested by the 1D circuit analogy explained earlier, primarily due to the location of the HICs beneath the PA, described later.

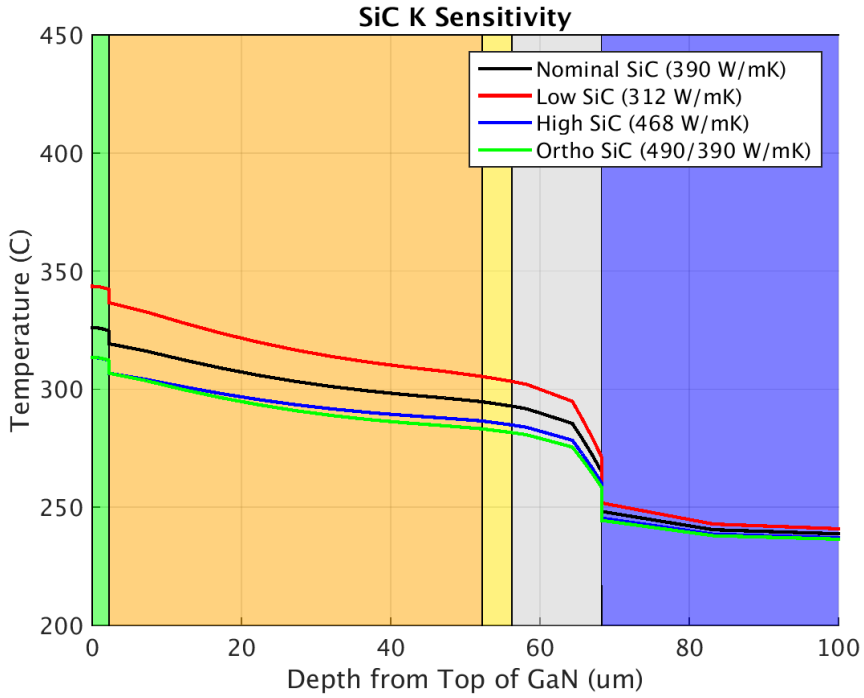


Figure 60: Substrate Temperature Profile, SiC Sensitivity Study

The green line in Figure 60 indicates the temperature profile of the device assuming orthotropic SiC with increased conductivity in the through-substrate (k_z) direction. The model

behavior is nearly identical in both the orthotropic and high-k cases. There is less than 5% difference between the through-substrate conductivities, which accounts for most of the similarity in performance. Interestingly, comparison of the two cases' temperatures in the below the PA shows the reduced in-plane conductivity (k_{xy}) of the orthotropic case appears to have negligible effect on substrate temperatures directly under the PA. Figure 61 illustrates the temperature profile in the portion of the SiC layer directly underneath the PA is nearly identical for both cases. This finding indicates that in-plane thermal spreading through this portion of the SiC substrate is less significant than heat flow through the plane of the substrate. Because the behavior of the high-k and orthotropic cases are roughly equivalent, only the orthotropic case will be presented in the following figures.

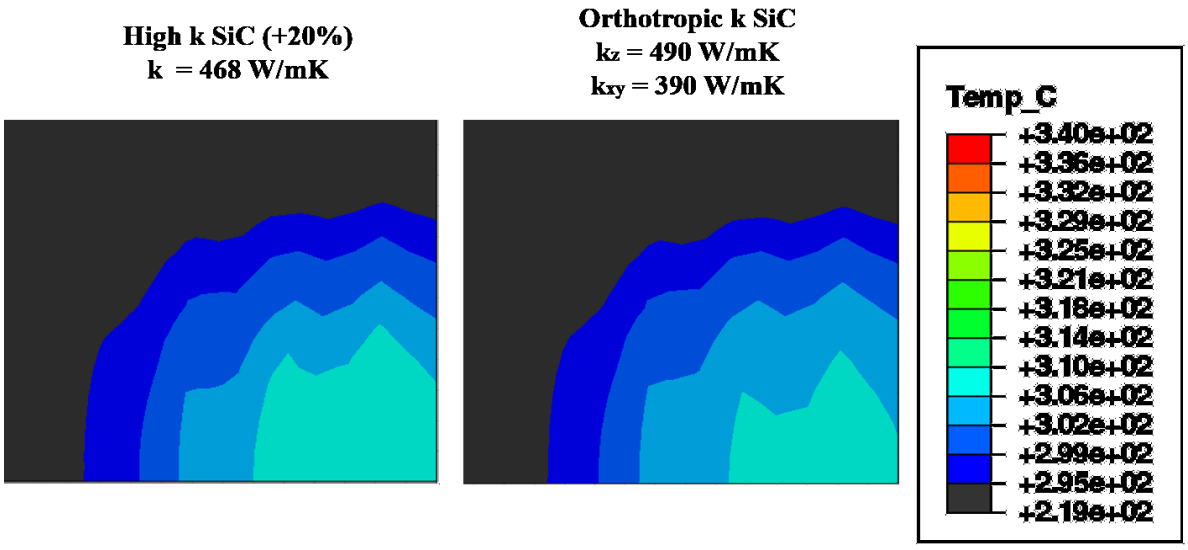


Figure 61: SiC Temperature Contour Comparison for High k (+20%) and Orthotropic Cases

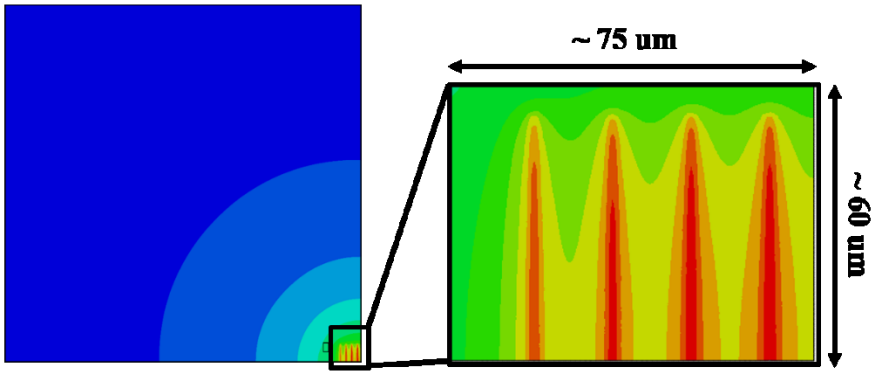


Figure 62: Detail View Region for Temperature Contours Shown in Figure 63-Figure 65

Figure 63-through Figure 65 further illustrate the performance of each SiC conductivity case with temperature contours from the surfaces of the GaN, SiC, and CMOS substrates. Each figure shows a zoomed view of an area roughly 75 um by 60 um surrounding the PA, as depicted in Figure 62. Furthermore, each figures' color contour scale has been individually adjusted to better describe heat flow through the substrate.

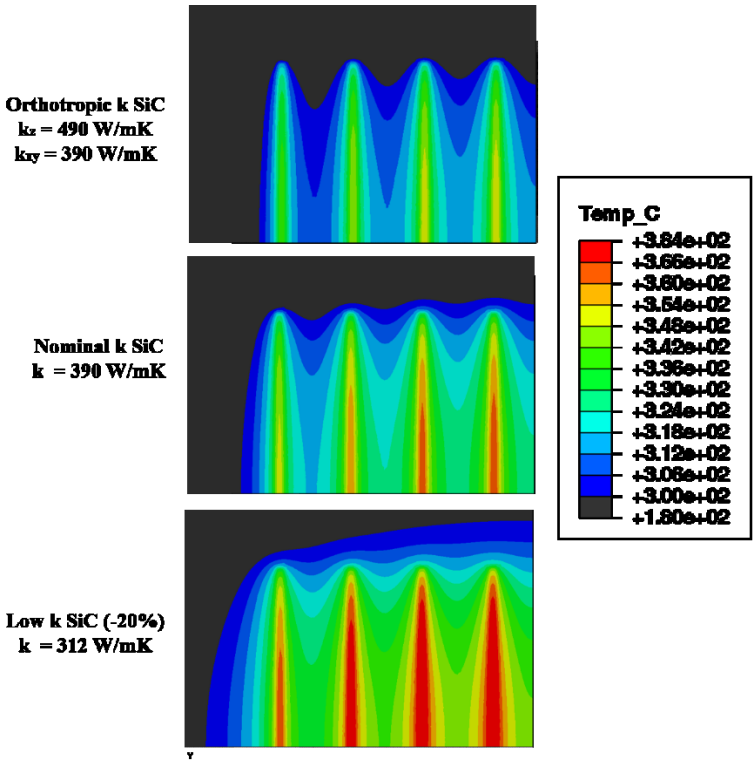


Figure 63: GaN Temperature Contours for Varied SiC Conductivity Cases

In the GaN region of Figure 63, reduced SiC conductivity is reflected by corresponding increases in the temperature of the GaN fingers. The heated region appears to grow in size as the reduced conductivity of SiC allows less heat to sink down and away from the PA. This behavior is reflected again in the SiC regions shown in Figure 64; at lower conductivities, high-temperature regions develop below the PA and continue to spread out over larger areas as the conductivity is further reduced. This trend does not continue into the CMOS layer, discussed below.

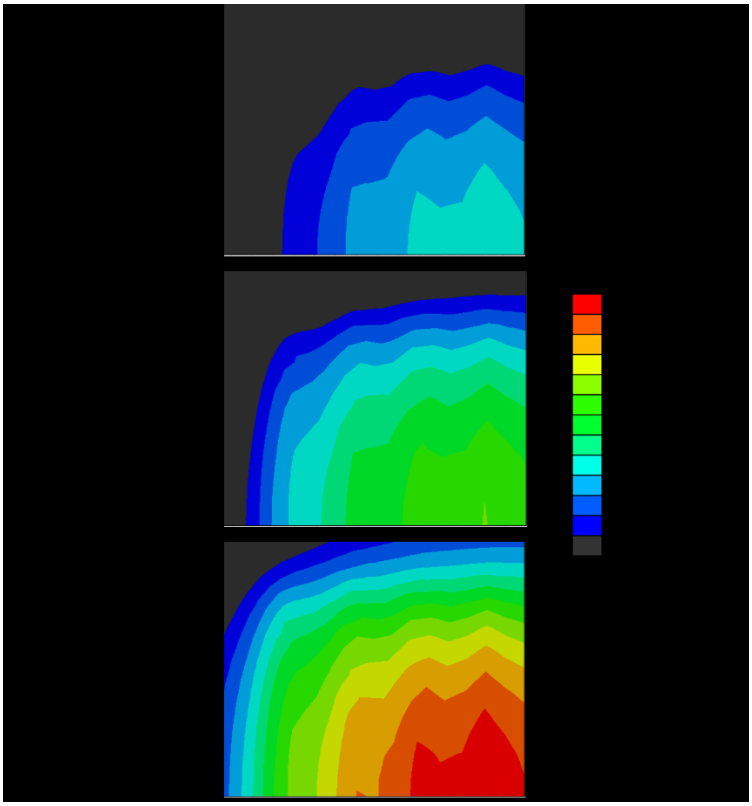


Figure 64: SiC Temperature Contours for Varied SiC Conductivity Cases

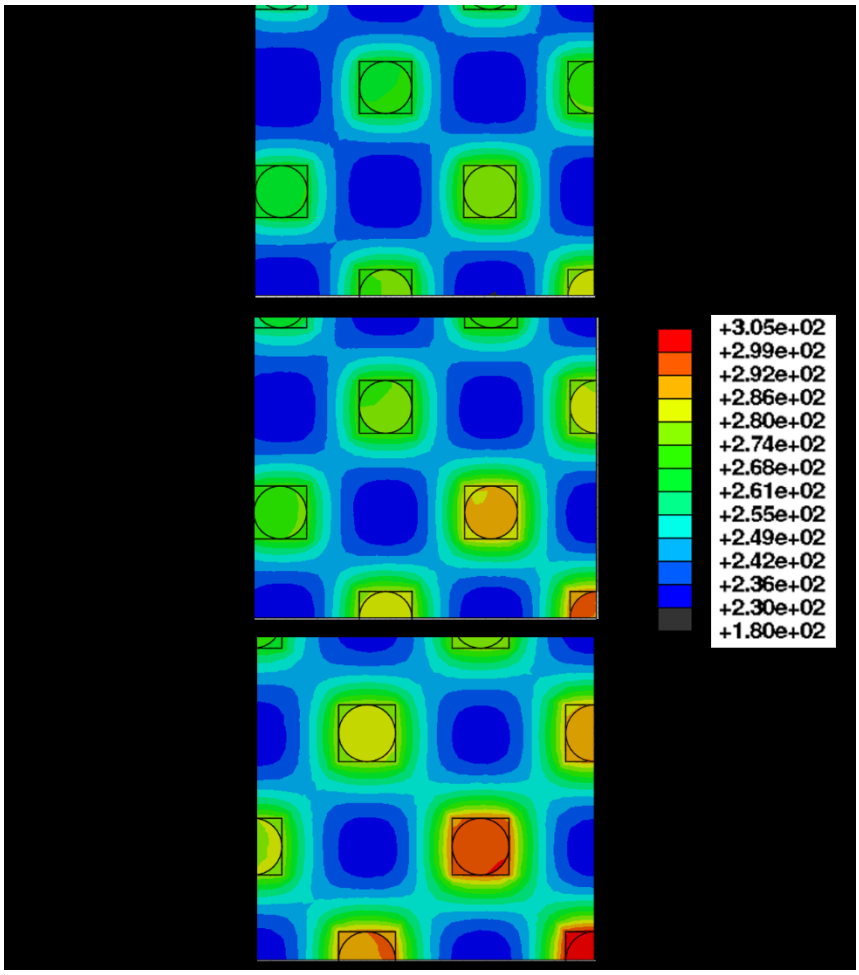


Figure 65: CMOS Temperature Contours for Varied SiC Conductivity Cases

Figure 65 illustrates the temperature of the CMOS layer as SiC's conductivity is varied. Physical isolation between individual HICs disrupts the trend seen above of a growing, high-temperature region. Instead, the temperature increases seen in this layer are primarily confined to only a few HICs located directly beneath the PA. As thermal resistance increases throughout the upper substrate, the increased temperatures experienced by SiC do not translate directly to uniform increases in the CMOS layer. Instead, heat flow is constrained to funnel downward through a small number of HIC instances, encountering strong spreading resistances along the way which lead to localized heating effects within the CMOS layer. Because heat flow is not spread uniformly across the top of the CMOS layer, significant changes to the conductivity of

SiC, and the corresponding temperature rise in SiC, will lead to increased peak temperatures of the CMOS layer. At this point, it becomes clear that the HIC geometry renders the problem more complex than the 1D circuit model introduced above, and the analogy begins to break down.

A final set of two simulations were run to determine the effects of combining the various uncertainty terms outlined above. From the standpoint of a circuit designer, the best-case scenario involves combining all of the upper conductivity values assumed earlier, which will lead to reductions in simulated device temperature; similarly, the worst-case scenario contains a combination of all the lower conductivity values assumed earlier. The resulting combinations were re-simulated, and their substrate temperature profiles are shown in Figure 66. Because the conductivities assumed in this section are temperature-independent, net changes to individual assumptions can be added linearly; thus, the combined cases' net temperature changes could also be found by simply adding the relative changes from each of the cases described above. The best-case scenario results in a net peak temperature change of -10°C, while the worst-case scenario sees temperature increases of up to +25°C above the nominal baseline. Relative to the temperature rise of the baseline case, the stated material properties used in this model introduce a 35°C uncertainty in peak temperature, which amounts to about 12% of the device's temperature rise above ambient.

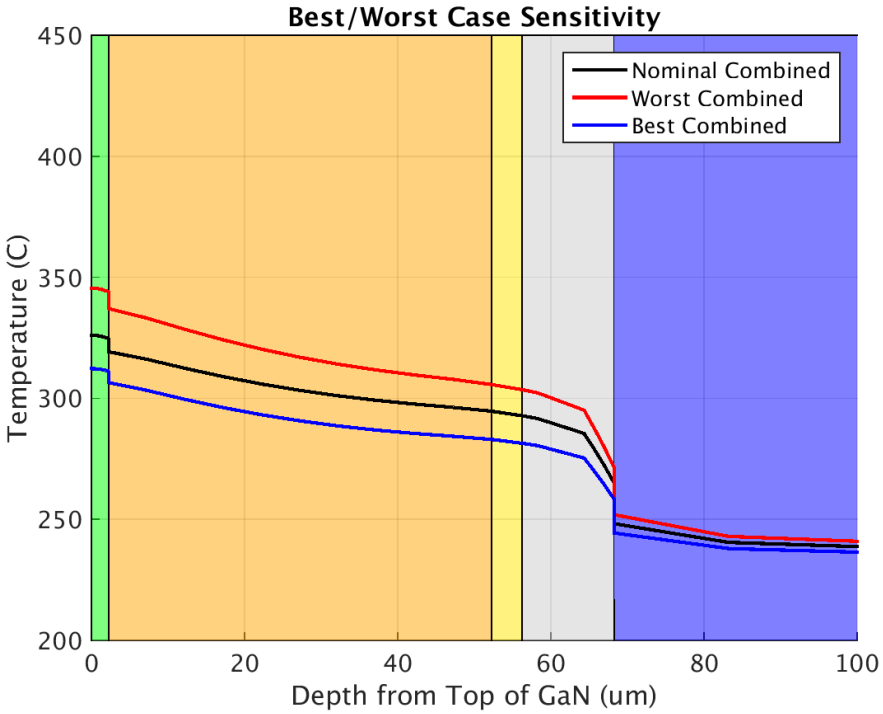


Figure 66: Substrate Temperature Profile, Combined Best/Worst Case Sensitivity Study

5.4 Temperature-Dependent (Nonlinear) Conductivities

Up to this point, all of the models and results presented in this thesis have assumed temperature-independent thermal conductivities for all materials in the device. However, real materials' thermal conductivities typically exhibit some degree of temperature dependence. Across the large temperature ranges seen by the different substrates in this device, conductivities are expected to vary enough to impact simulation results. Similar finite element studies of GaN HEMTs have demonstrated this and included such parameters in their thermal models [29] [30].

Thermal conductivities of metallic solids have been measured and reported many times in literature; various fitted correlations are available to describe conductivity as a function of temperature [67]. While the underlying mechanisms of conduction differ, similar correlations exist for nonmetals as well. Conductivities of semiconductors are typically fitted to a

temperature-varying power laws of the form given below, where temperature is measured on an absolute scale.

$$k = k_0 \left(\frac{T}{T_0}\right)^{\alpha} \qquad [67]$$

Conductivity correlations for primary device materials are given below in Table 6, and plotted in Figure 67 across a representative temperature range. At temperatures of 500°C, the conductivities of GaN, SiC, and Si substrates are all cut in half from their bulk values at room temperature. In this same range, gold's thermal conductivity drops by less than ten percent.

Table 6: Temperature-Dependent Thermal Conductivity Models for Device Materials

Material	**Thermal Conductivity (W/mK, T in Kelvin)**
GaN	$160\left(\frac{300}{\mathrm{T}}\right)^{1.4}$ [29]
SiC	$400\left(\frac{300}{\mathrm{T}}\right)$ [29]
Si	$148\left(\frac{300}{T}\right)^{1.3}$ [29]
Gold	$91.3T^{0.22}exp\left(\frac{36.2}{T}\right)$ [67]

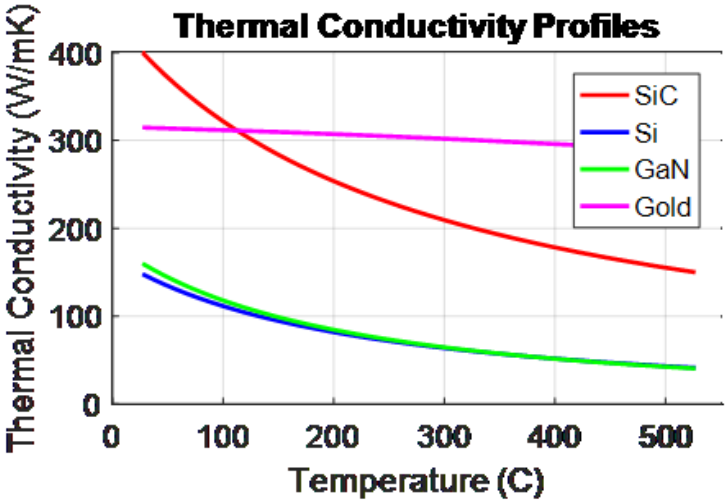


Figure 67: Temperature-Dependent Conductivity of Materials at Different Operating Temperatures

The finite element model parameters were modified to incorporate the temperature-dependence listed above for GaN, SiC, and Si. As previously mentioned, gold's conductivity does not change much, so its conductivity was not changed. In addition to the parameters applied to the submodel from Table 6, the global model's three-parameter description of the global chiplet was updated to incorporate temperature-dependence. Namely, the power law dependence of SiC and Si were applied to the conductivity of GaN/SiC and Si/CMOS in the global chiplet. The model was then re-simulated using these updated parameters for comparison with the temperature-independent results presented in Section 4.5. Introduction of conductivity's temperature-dependence causes the simulation to behave nonlinearly; that is, a nonlinear relationship exists between input power and device peak temperature, and the device's total thermal resistance is no longer a constant at all power levels. Simulations including this assumption will be referred to as nonlinear throughout the remainder of this thesis.

Table 7: Temperature-Dependence Applied to Global Model Chiplet

Material	Thermal Conductivity (W/mK, T in Kelvin)
GaN/SiC	$358\left(\frac{300}{T}\right)$
Si/CMOS	$137\left(\frac{300}{T}\right)^{1.3}$

5.4.1 Temperature-Dependent Properties – Global Model Results & Discussion

The results of the Figure 68 and 69 contain temperature contours from each of the parts in the nonlinear global model. With the exception of regions close to the heat source in the near-field chiplet, results of the nonlinear global simulation are virtually identical to those presented in Section 4.5.2 for the temperature-independent (linear) case. As such, the figures are not described in great detail here except for when they differ from the linear case's results.

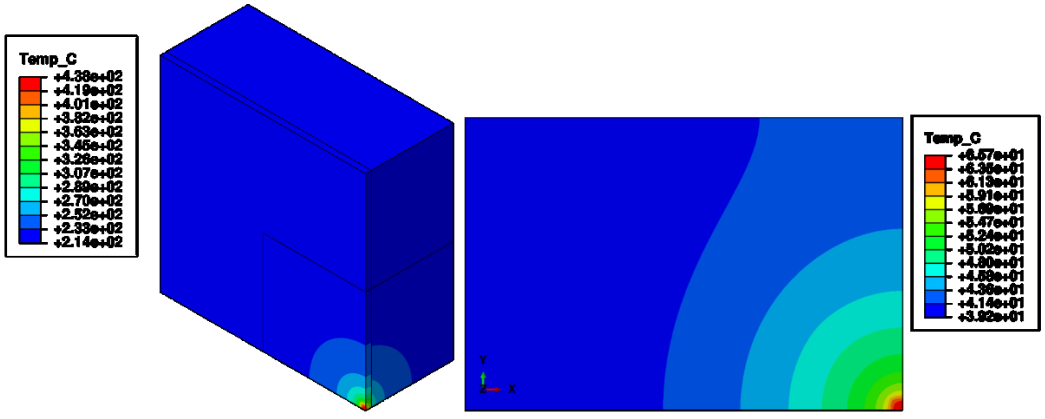


Figure 68: Global Model Chiplet & Board Temperature Contours (Nonlinear Material Properties)

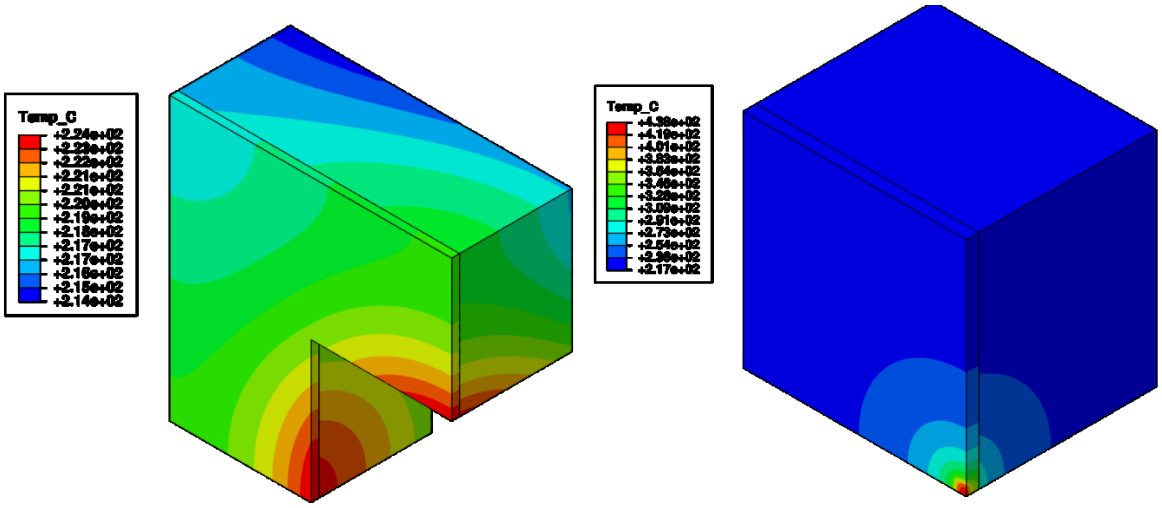


Figure 69: Global Model Far-Field Chiplet Temperature Contour (Nonlinear Material Properties)

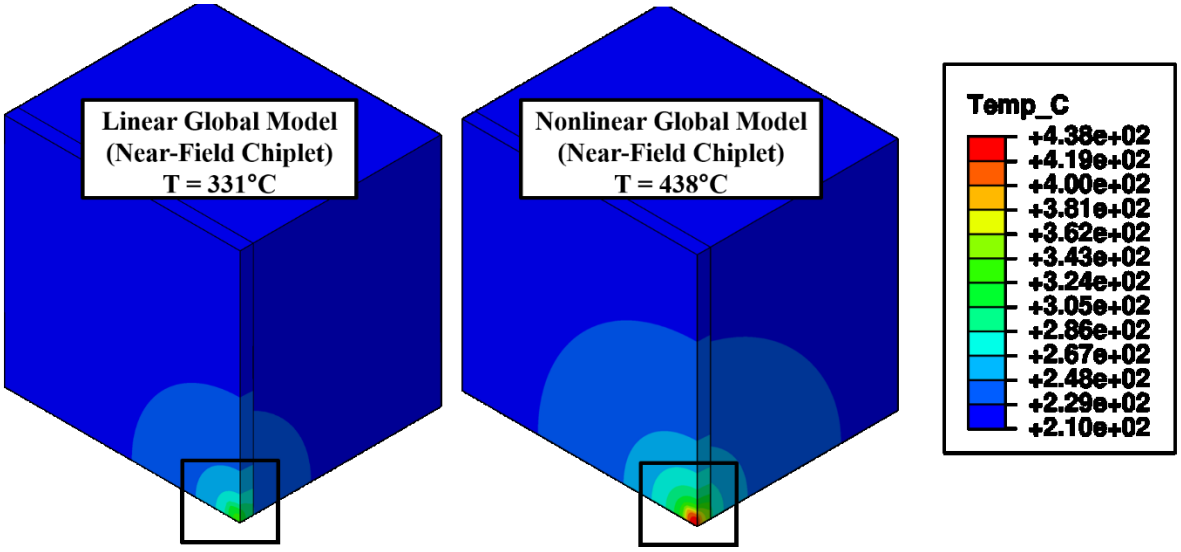


Figure 70: Comparison of Linear and Nonlinear Global Simulation Results

Table 8: Tabulated Comparison of Linear and Nonlinear Global Simulation Results

Location	Temperature (°C)		Rel. Change
	Linear	Nonlinear	
Board Peak Temp	66	66	0 %
Near-Field Chiplet Max Temp	331	438	32 %
Far-Field Chiplet Max Temp	221	224	1.5 %
Far-Field Chiplet Min Temp	216	214	1.0 %
Boundary Temps for Submodel	219 – 220	220 - 223	~ 1.5 %

Figure 70 contains a side-by-side comparison of the linear and nonlinear global simulation results; likewise, a summary of key temperature results compared between the two cases are presented in Table 8. As shown above, there is less than 2% change between the linear and nonlinear global models' temperatures for regions far from the PA. In fact, almost all of the relevant change occurs in the close vicinity of the PA. The interfacial temperatures between the near-field and far-field chiplet change by less than 3°C in either case. . However, the peak temperature predicted by the nonlinear simulation is 32% higher than the linear simulation's baseline. This indicates that temperature-dependence of thermal conductivity is only relevant within the submodel's near-field region of interest, close to the heat source As noted in prior sections, the global model's peak temperature should not be used as a direct estimate of actual device peak temperature; however, the discrepancy between linear and nonlinear models at the global scale foreshadows the importance of including temperature-dependent properties in the detailed submodel.

5.4.2 Temperature-Dependent Properties –Submodel Results & Discussion

Figure 71 illustrates the temperature contour of the temperature-dependent submodel simulation. Peak temperature in the nonlinear simulation's GaN layer was 520°C, corresponding

to a 155°C increase over the baseline from the linear simulation presented in Section 4.5.2. As suggested by the global model comparison given above, Figure 72 illustrates that the linear and nonlinear simulations' differences appear subtle when plotted on the same color scale. At first glance, Figure 72 might suggest that far-field temperatures of the device are unaffected by assuming nonlinear thermal conductivity, however this is not the case. Closer inspection of the gate finger temperature profile given in Figure 73 further illustrates the linear and nonlinear simulations' striking differences.

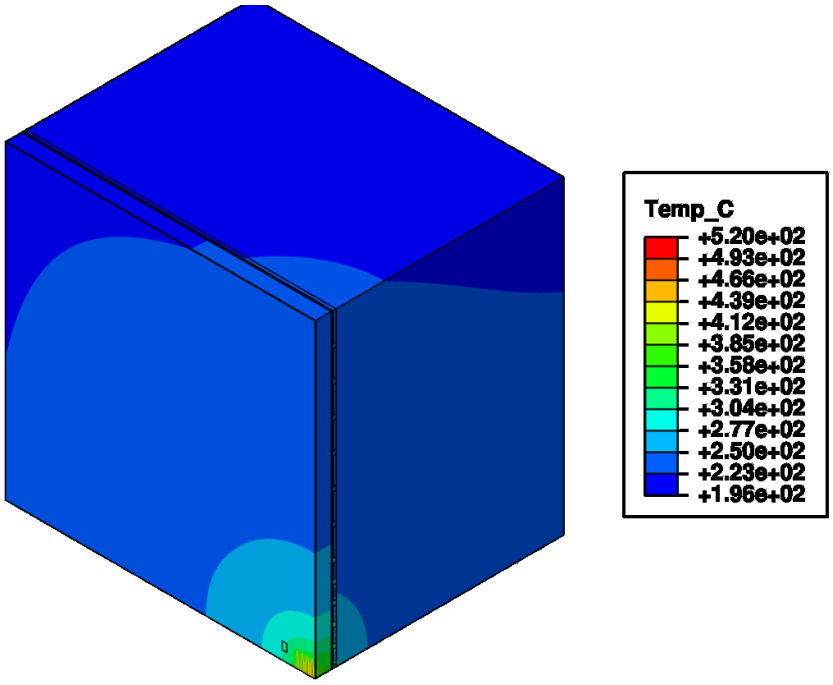


Figure 71: Temperature-Dependent Submodel Simulation Temperature Contour

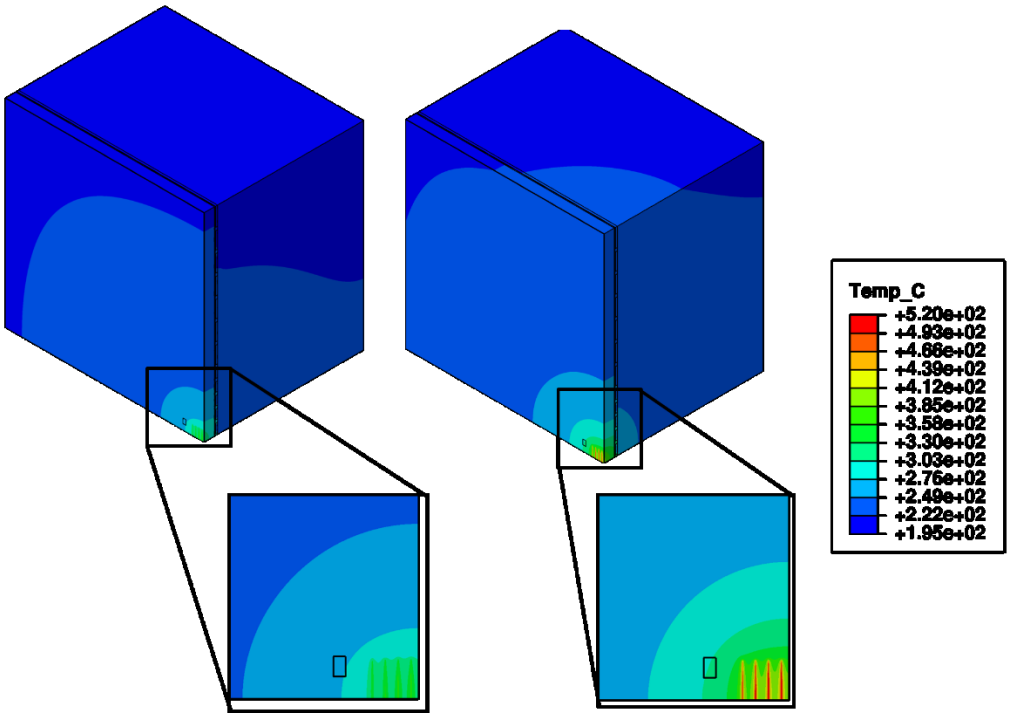


Figure 72: Submodel Temperature Contour Comparison: linear (left), nonlinear (right)

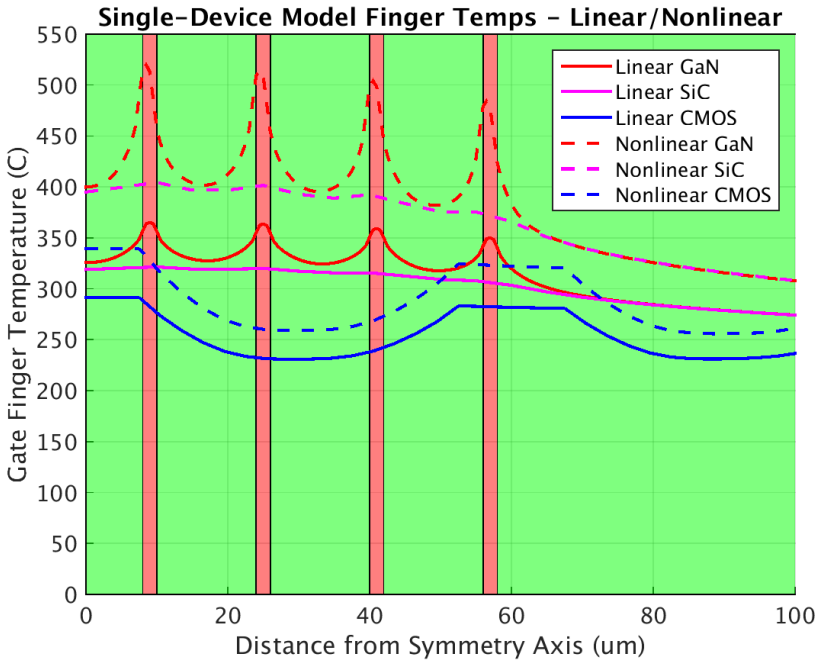


Figure 73: Submodel Gate-Finger Temperature Profiles, Linear/Nonlinear Comparison

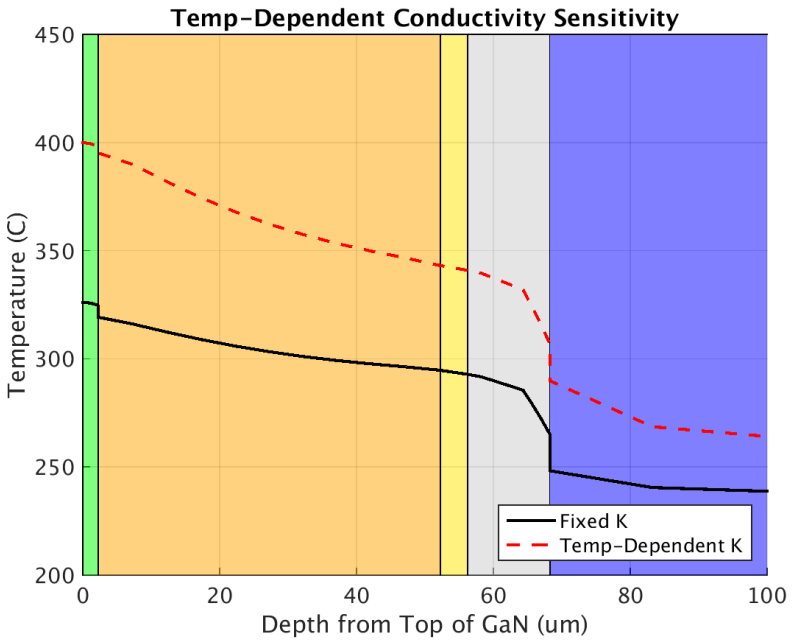


Figure 74: Submodel Substrate Temperature Profiles, Linear/Nonlinear Comparison

Figure 73 illustrates the intense temperature spikes seen within the gate finger regions in the nonlinear simulation. Here, the GaN temperatures are >>100°C hotter than those predicted in the linear simulation, and these temperature increases propogate downward to significantly raise the temperatures of both SiC and CMOS substrates. Figure 74 depicts the through-substrate temperature profile under the PA. Even at 100 um below the surface of the GaN PA, the simulation's temperatures are still grossly affected by the devices' reduced thermal conductivity at high temperature.

The results presented in this section underscore the significance of including temperature-dependent material properties in the finite element model definition. While temperature-independent results are useful as first approximations, localization of the PA heat source drives temperature rises which cannot be predicted accurately without incorporating the substrates' reduced dissipation at high temperatures. Therefore, the remaining investigations and results presented in this thesis all include this assumption.

5.5 Die-Attach Resistance Study

Prior sections of this thesis have mentioned the significance of the die-attachment resistance parameter. This section will detail the reasons for its inclusion, and investigate its significance on the uncertainty of model results presented in this thesis.

As described in Section 3.1, a silver-filled die-attach adhesive was used to structurally bond the DAHI-integrated chiplet to a printed circuit board. In experimental studies published by Kurabayashi & Goodson [45] and Teerstra [47], conductances varied across a wide range and were found to be sensitive to a number of factors involved in the application process. As noted above, the die-attachment parameter assumed thus far in this thesis ($k'' = 2e3 \frac{W}{m^2 K}$) is a conservative estimate of silver-filled epoxy with moderate voiding. However, initial investigations of this parameter revealed simulation results were extremely sensitive to the choice of an assumed value. At this time, no thermal measurements of the DAHI-integrated device are available with which to calibrate this choice. To illustrate the large range of uncertainty built into this assumption, and to highlight the need for experimental calibration and validation of this model, a series of simulations were performed. In this study, the global model's die-attach parameter was varied and re-simulated across the range of possible values noted in Table 9; these values roughly correspond to the performance range of available silver-based adhesives [45] [47].

Table 9: Die-Attach Study, Assumed Parameter Values

K" Value ($\frac{W}{m^2K}$)	Corresponding Case
1e3	Silver epoxy, incomplete attach, major delamination
2e3	Silver epoxy, complete attach, nominal voiding
4e3	Silver epoxy, complete attach
8e3	Silver epoxy, optimistic
12e3	Low-performing silver thermoplastic
18e3	Nominal silver thermoplastic
25e3	High-performing silver thermoplastic

5.5.1 Die-Attachment Study, Results and Discussion

Figure 75 plots the peak temperature of both the global model and submodel for each die-attachment case; the baseline $k'' = 2e3\frac{W}{m^2K}$ result is highlighted by the black box. In addition, the dotted lines represent simulation results obtained by assuming a perfectly-conducting tie-constraint at the adhesive interface, equivalent to $k'' = \infty$. In the performance range of silver-filled epoxies ($k'' = 1e3$ to $8e3\frac{W}{m^2K}$), the simulation's peak temperatures are highly sensitive to changes in k". Die-attach resistance dominates the device's thermal behavior in this regime. At the other end of the spectrum, peak temperatures begin to converge toward the tie-constraint limit, which reflects that the substrates' conductivities and spreading resistances are primarily governing the device's temperature.

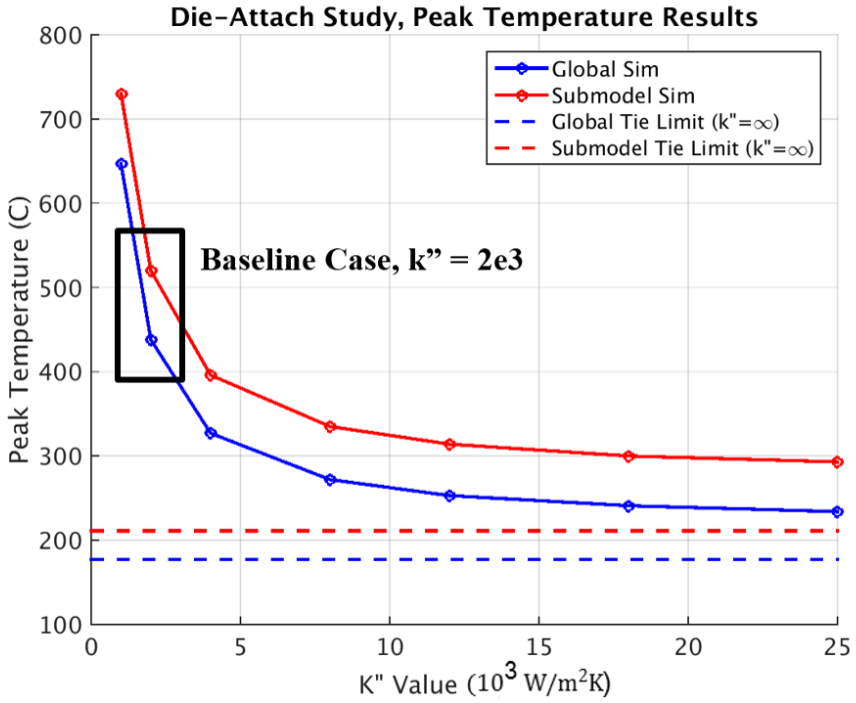


Figure 75: Die-Attach Study Peak-Temperature Curves

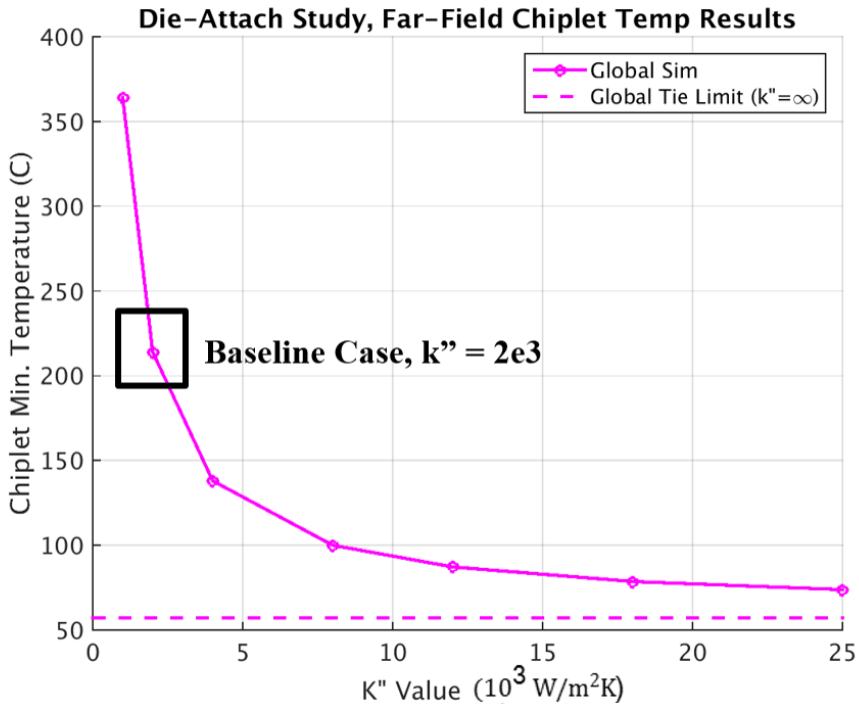
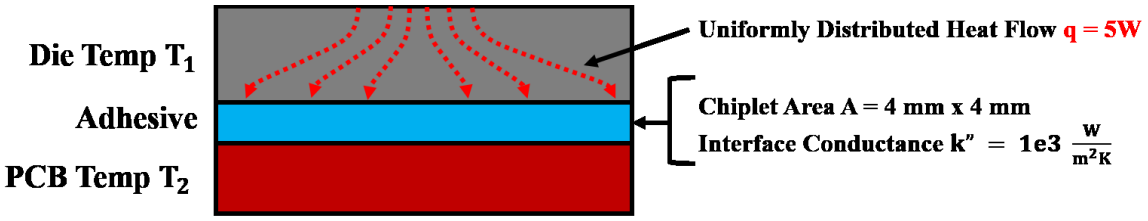


Figure 76: Die-Attach Study Chiplet Minimum Temperature Curve

Furthermore, the adhesive resistance directly governs uniform heating in regions far from the PA. Figure 76 plots the minimum temperature seen anywhere on the 4 mm by 4 mm chiplet as a function of die-attach conductance. These values are taken directly from the global simulation and indicate the severity of the thermal bottleneck at the adhesive interface. In the high-resistance regime, the *entire* chiplet is warmed to temperatures greater than 200°C; in this case, there is no location which remains below CMOS's thermal failure limits. At first glance, those figures may seem impossibly hot, even in the worst-case scenario of 5W heat dissipation. However, some simple math can confirm these ominously high temperatures.

Assume, for example, that free convection governs cooling across the surface area of the chip. Because the chip is so small, any resulting heat flux from its surfaces can be assumed negligible. Furthermore, the heat flux has likely had ample room to spread out across the entire base of the substrate by the time it reaches the backside of the die, thereby approximating a uniformly distributed heat flux. Therefore, one might assume that all 5W of heat dissipated from the chiplet must pass by conduction through the 4 mm x 4 mm area at the base of the chip. As described in Figure 77, an adhesive conductance of $k'' = 1e3 \frac{W}{m^2 K}$ would thus be expected to produce a temperature rise of over 300°C across the interface.



$$\text{Heat Flux } q'' = \frac{q}{A} = \frac{5W}{1.6e{-}5\ m^2} = 3.13e5\ \frac{W}{m^2}$$

$$\text{Temperature Rise } (T_1 - T_2) = \frac{q''}{k''} = \frac{3.13e5\frac{W}{m^2}}{1e3\frac{W}{m^2K}} \approx \boxed{313°C}$$

Figure 77: Die-Attach Temperature Rise Illustration

The primary conclusion drawn from these results is that die-attachment uncertainty is the largest source of possible error built into the current finite element model. While every effort has been made to establish a realistic, experimental basis for the chosen simulation parameters, further testing and characterization of the die-attachment interface is required for the DAHI-integrated devices before the results of this model should be broadly accepted.

Bearing in mind the need for experimental validation, all the effort spent preparing this model has not been wasted. Rather, the findings presented thus far provide insights which should be useful for circuit designers during the next design cycle. Care must be taken to avoid assuming non-physical boundary conditions when performing thermal modeling. Assumption of perfect heat sinks, or isothermal backside boundary conditions will lead to inaccurate models and spurious simulation results. Though it may add a degree of complexity to some areas of the model, the inclusion of realistic global boundary conditions should be practiced; even coarse approximations similar to those obtained here using submodeling will prove more useful to designers.

Furthermore, this section illustrates the need for careful selection of die-attachment adhesives and processes. The interface's sensitivity to voiding and delamination makes the

assembly process critical. Every effort must be made to choose materials with interfacial conductance high enough to maintain survivable temperatures within the device. In the case of GaN devices studied here, thermal failure mechanisms become significant at channel temperatures exceeding 500°C [59]. For the devices simulated in this study, a minmum interface conductance of $k'' = 4e3 \frac{W}{m^2 K}$ satisfies this thermal limit. Ideally, conductance in the range of $k'' = 10e3 - 20e3 \frac{W}{m^2 K}$ is preferable, as above this range the conductive resistance of the DAHI-integrated substrate begins to govern device heating.

5.6 Single-Device Power Study

Simulations presented in prior sections of this thesis all assumed 5W total heat dissipation in accordance with the Stage 3 PA's worst-case measured efficiency [48]. In reality, the PA's efficiency and power output both with operating frequency and bias condition, which means its actual heat dissipation may vary during operation. In the case of temperature-independent material properties, thermal resistance of the device would remain constant for all conditions and temperature would be linearly related to dissipated power. However, as described in Section 5.4 the model now includes nonlinear material properties which cause conductivity to decrease and device thermal resistance to increase with temperature. Thus, overall thermal resistance is expected to decrease for higher efficiency, lower heat-dissipation cases.

This section presents a series of simulations which simulate lower power levels corresponding to 4W, 3W, 2W, and 1W of heat dissipation. Initially, simulations were performed using the same baseline die-attachment conductance of $k'' = 2e3 \frac{W}{m^2 K}$ assumed previously. Due to the large temperature uncertainties associated with the die-attachment interface detailed above in Section 5.5, these power level simulations were repeated for alternate conductance levels of

$k'' = 8e3$ and $25e3 \frac{W}{m^2K}$, corresponding to lower resistance, better-performing thermal adhesives. In each case, both the global and submodels were re-simulated after scaling the input power level to each.

Figure 78 illustrates the device temperatures as a function of heat dissipated; each line corresponds to a die-attach conductance case. The simulation temperatures are seen to increase nonlinearly with increasing power; curvature in these profiles is a direct result of reduced substrate conductivity at higher temperatures. For comparison, the three dashed lines indicate the temperature-independent simulation results; here, constant slope indicates the device's thermal resistance remains constant. Significant deviation between the solid and dashed lines indicates the temperature error that would occur if a low-power simulation result were linearly extrapolated to a higher power case without accounting for temperature-dependence within the model's properties.

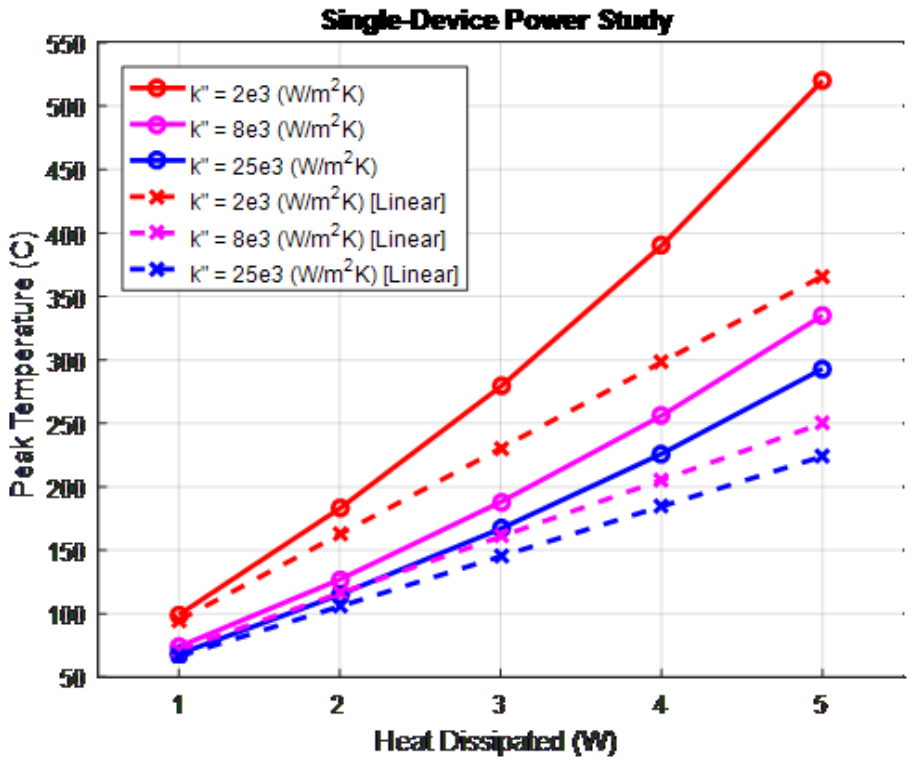


Figure 78: Single-Device Power Study, Peak Temperature Results

This result is communicated in different form by Figure 79, which contains a plot of overall device thermal resistance as a function of dissipated power. Here, thermal resistance has been calculated by the device's temperature rise above ambient over the total input power, as described below. The dashed lines indicate the device's thermal resistance calculated from the temperature-independent simulations; for die-attach conductances $k'' = 2e3, 8e3, \text{and } 25e3 \frac{W}{m^2K}$, total thermal resistance comes out to 67.8, 44.6, and 39.4 °C/W in each linear case. As previously indicated by the constant slope in Figure 78, lack of temperature dependence means that thermal resistance remains constant for all power levels, hence the linear simulations' resistances appear horizontally flat in Figure 79. The simulation results indicate the device's temperature-dependent thermal resistance exceeds the resistance predicted by the linear simulation, and increases by about ~30% for each die-attach case as dissipated power increases from one to five watts.

$$R_T = \frac{T_{max} - T_{amb}}{q_{dissipated}}$$

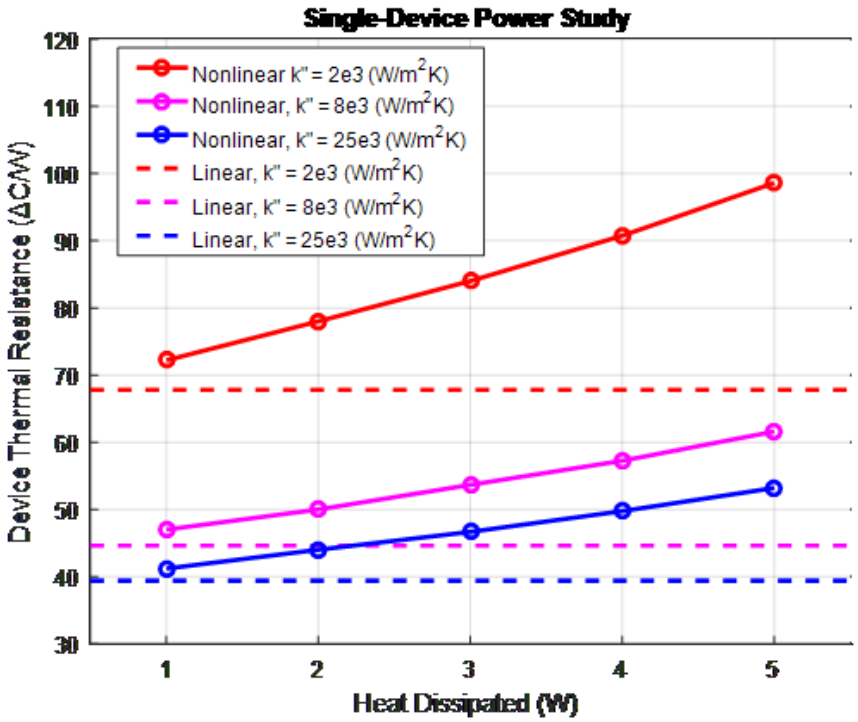


Figure 79: Single-Device Power Study, Thermal Resistance Results

The large gap between the k" = 2e3 and 8e3 $\frac{W}{m^2K}$ cases in Figure 79 is indicative of the need for proper application of the die-attachment adhesive. These two cases correspond roughly to silver-filled epoxy under partially voided (k" = 2e3), and complete attachment (k" = 8e3) conditions. Recalling the result of Figure 75 in Section 5.5.1, the latter case corresponds to the region where the temperature/die-conductance curve 'bottoms out' and substrate conductivities begin to dominate heat dissipation. This case represents the desirable range of adhesive conductance; here, the device's thermal resistance is low enough to allow safe operation well below GaN's thermal failure limits.

As noted in Chapter 3, the devices being studied relegate active CMOS components to the cooler regions at the exterior of the device. Thus, peak temperature is not a useful metric for

predicting thermal failure in the CMOS region as the relevant temperatures are dictated by the CMOS components' proximity to the PA heat source. However, for any active CMOS to survive in the device, some region of the CMOS substrate must remain below its thermal failure limit. Thus, the minimum temperature of the CMOS substrate must be lower than 150°C for the device to be thermally feasible [60]. Figure 80 illustrates the minimum temperature *anywhere* in the chiplet as a function of dissipated power and die-attach conductance. At high power and low conductance, temperatures in the entire device exceed the CMOS thermal limit. This result foreshadows the limited range for safe CMOS operation which will be explored in more detail as part of Section 6.5.

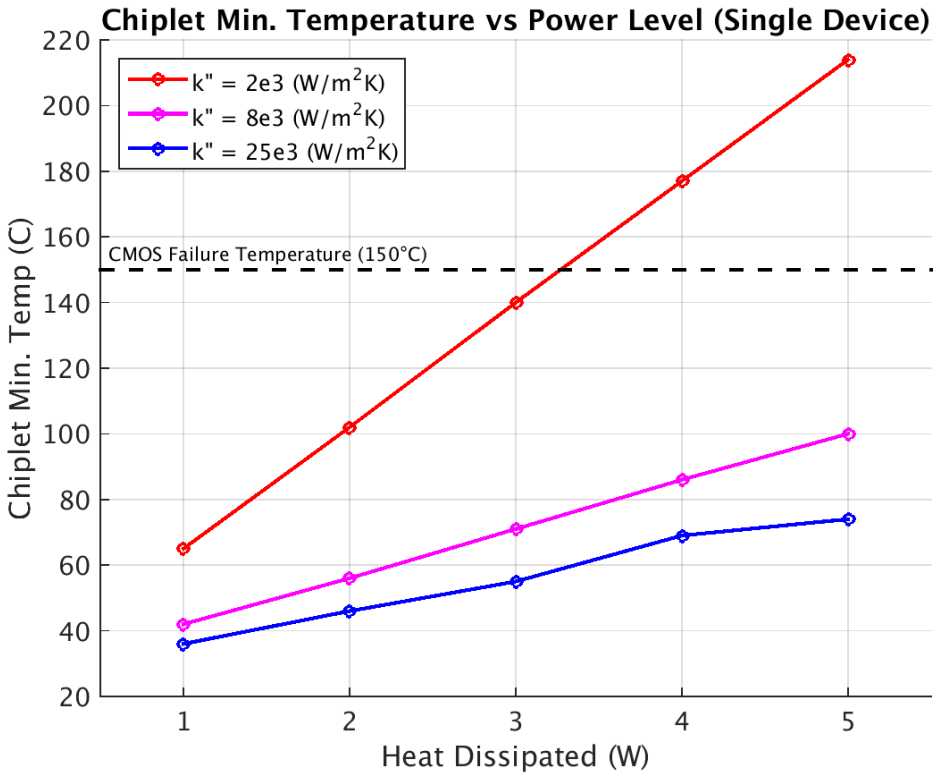


Figure 80: Single-Device Power Study, Far-Field Chiplet Minimum Temperatures

6 Multichannel Device Model

This chapter presents a finite element model of the device which has been extended to include multichannel operation of two Stage 3 PAs on the same chiplet. In reality, the two devices are located slightly off-center on the chiplet, spaced roughly 1.6 mm apart as indicated previously in Figure 11. As in the single-device case, the Stage 3 PAs were shifted to the centerline of the chiplet to simplify the model using ¼ symmetry; this configuration is illustrated in Figure 81 below. In general, the multi-device model maintains the same set of assumptions and parameters defined previously in the single-device model. However, the "region of interest" used to delineate between the global and submodels has been modified to better suit the multichannel geometry, as will be described below.

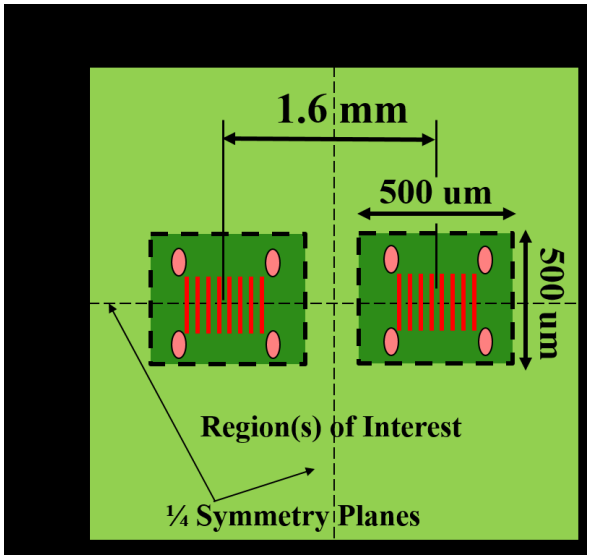


Figure 81: Multichannel Chiplet Region of Interest Definition (compare to Figure 11)

6.1 Multichannel Device Model Description

In the multichannel case, the submodel region of interest corresponds to the 500 um by 500 um area of the chiplet, shown in dark green in Figure 81. This multichannel submodel region includes only the near-field HIC checkerboard surrounding either PA, hence it is smaller than the 1 mm by 1 mm area used in the single-device submodel which included both near and far-field HICs. However, the ¼-symmetry model of each individual PA in the multichannel case now contains all eight gate fingers, and twice as many near-field HIC instances, as indicated in Figure 82. To reduce the submodel's complexity, far field HICs were removed from the detailed region and instead included as an equivalent conductance within the global model. In Section 4.5.2, good agreement was found between the global and submodel simulations for the far-field HIC region, so this change was not expected to influence the model's results in the region of interest near the PA.

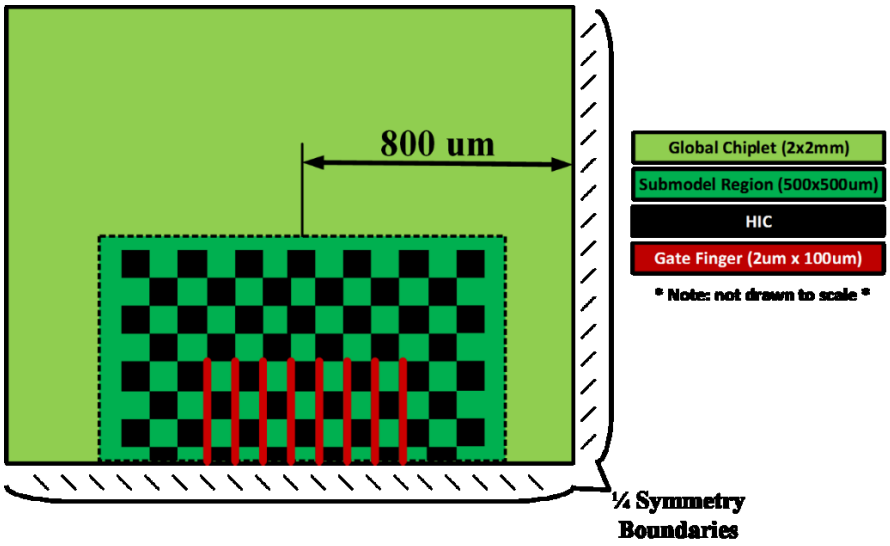


Figure 82: Multichannel Global and Submodel Illustration

Furthermore, the multichannel PA contains eight gate fingers rather than four. As each gate finger experience large, localized temperature rises, each finger requires a fine mesh which

quickly increases the submodel's total element count. To further simplify the multichannel submodel and reduce total element count, only the top 50 um of the bulk silicon substrate was included within the detailed region, as indicated in Figure 83. In the single-device model, the full depth of bulk silicon was included in the submodel, which was permissible because that simulation included significantly fewer elements. In this case, the bulk silicon substrate was segmented such that the lower portion was included in the global model's simulation. As indicated in Figure 83, the structure of the simulated device remains fundamentally the same, only the location of the submodel boundary condition changes.

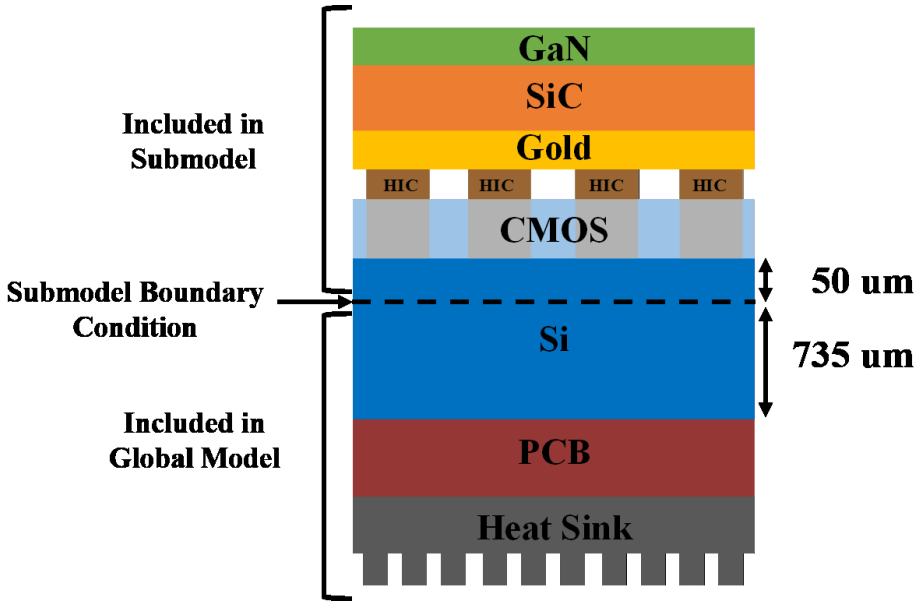


Figure 83: Bulk Silicon Division between Global and Submodels

In accordance with the findings of Section 5.4, the multichannel submodel includes the same temperature-dependent material properties assumed above. Likewise, the same three-parameter scheme described in Section 4.3 is used to define the global model. For this multichannel case, worst-case heat dissipation of 5W per device is assumed unless otherwise noted. As in the single-channel case, the heat flux is assumed to occur uniformly across each gate finger; however the total heat input to both the global and submodel is doubled due to the

doubling of the gate area from four to eight fingers in the multichannel case. For example, ¼ symmetry now dictates that the model receives 2.5W of heat to the red regions indicated in Figure 84. Furthermore, the exterior surfaces of both models are again exposed to natural convection conditions $\left(\text{film coefficient h} = 10\frac{\text{W}}{\text{m}^2\text{K}}\right)$ with an ambient temperature of 27°C. In keeping with assumptions made earlier, the same interfacial conductances are assumed at the GaN/SiC boundary $\left(\text{k}'' = 3\text{e}7\frac{\text{W}}{\text{m}^2\text{K}}\right)$ and heatsink adhesive $\left(\text{k}'' = 1.4\text{e}3\frac{\text{W}}{\text{m}^2\text{K}}\right)$ interfaces. Unless otherwise noted, the baseline die-attach conductance of $\text{k}'' = 2\text{e}3\frac{\text{W}}{\text{m}^2\text{K}}$ is again assumed for the multichannel case. Section 6.3 will present the device's performance across the same range of conductance values shown in Section 5.5.

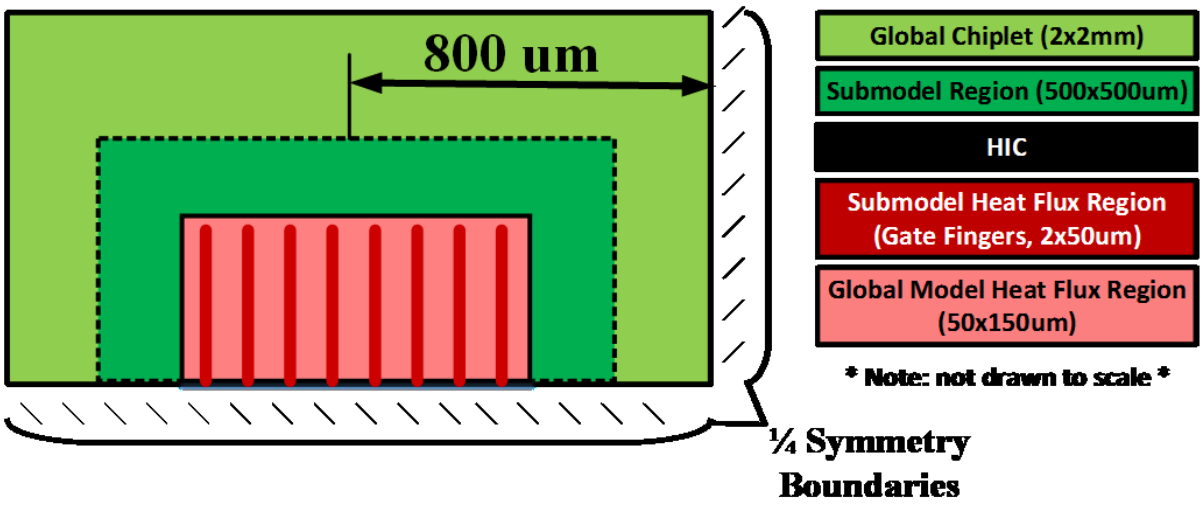


Figure 84: Multichannel Global and Submodel Heat Flux Regions

The submodel simplifications described at the beginning of this section are particularly useful for examining the effect of varying spacing within the device, as will be discussed further in Section 6.4. In particular, removal of the far-field HICs makes the global model's inter-channel spacing dimension easily reconfigurable. Different inter-channel spacings can be assigned by changing one dimension within the global model, re-meshing, and re-simulating.

The submodel remains unchanged in each case; it simply snaps in to the correct region via the use of the submodel boundary condition, which automatically updates when the global model definition changes. Figure 85 illustrates in 3D the re-configuration of the global model to multiple inter-channel spacings.

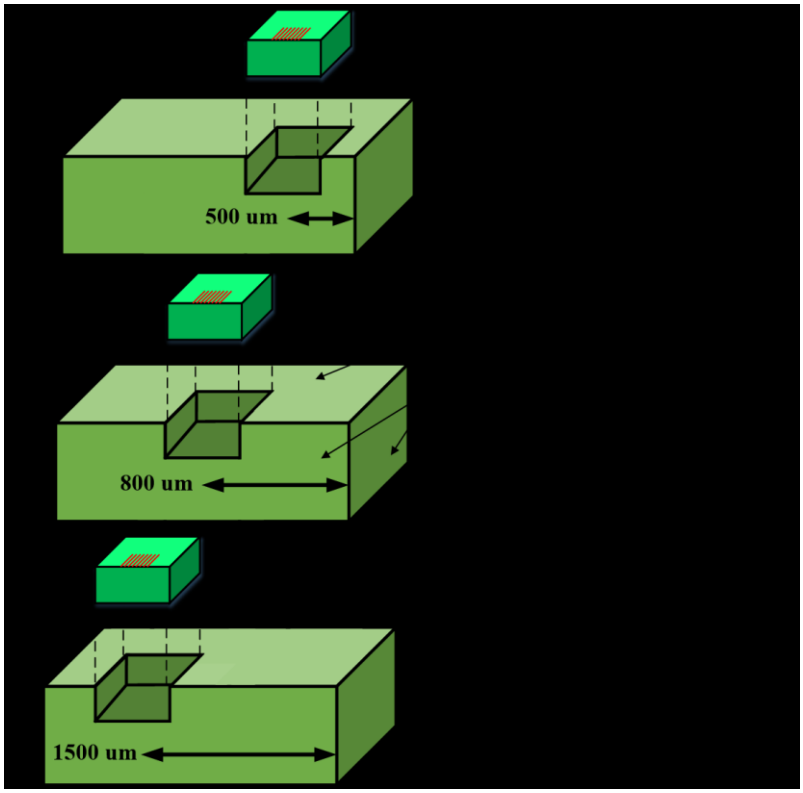


Figure 85: Illustration of Multichannel Global Model Re-Configuration; submodel snaps in at each spacing

6.2 Multichannel Model Results & Discussion

The simulation results presented in this section correspond to the baseline case of two devices spaced 1.6 mm apart, each dissipating 5W of heat, with die-attachment conductance of $k'' = 2e3 \frac{W}{m^2K}$. As stated previously, all multichannel simulations in this chapter include temperature-dependent (nonlinear) material properties, so comparisons will be made to the nonlinear, single-channel model results presented in Section 5.4.2.; for reference, the peak

temperature of the single-channel global and submodel simulations were found to be 438°C and 520°C respectively.

6.2.1 Multichannel Model Global Simulation Results & Discussion

Figure 86 illustrates the temperature contour of the multichannel global chiplet. The global simulation predicts a peak temperature of 685°C, which is significantly hotter than the 438°C peak global temperature encountered in the single-device case under equivalent conditions. Furthermore, the far-field of the chiplet is heated uniformly to ~400°C, which indicates the presence of the thermal bottleneck effect caused by the die-attach resistance. The sharp temperature increase between single channel and multichannel cases is a function of both die-attach resistance and inter-device spacing. In this particular case, the temperature increase is primarily governed by die-attach resistance, as will be described in more detail in Section 6.3.

For the quarter-symmetry configuration shown, a second device exists 1.6 mm away from the first device, to the right across the model's vertical symmetry plane. Slight temperature asymmetry between the chiplet's inner and outer sides (right and left sides in the figure) is indicative of increased temperatures in the area between the two devices. In this central region, the other device is represented by an adiabatic boundary which forces heat to dissipate away from the symmetry plane, thereby increasing local temperatures. Figure 87 illustrates a zoomed temperature contour for the 250 um by 500 um region of interest corresponding to the submodel. For this region close to the PA, asymmetry between inside and outside of the chiplet is almost imperceptible.

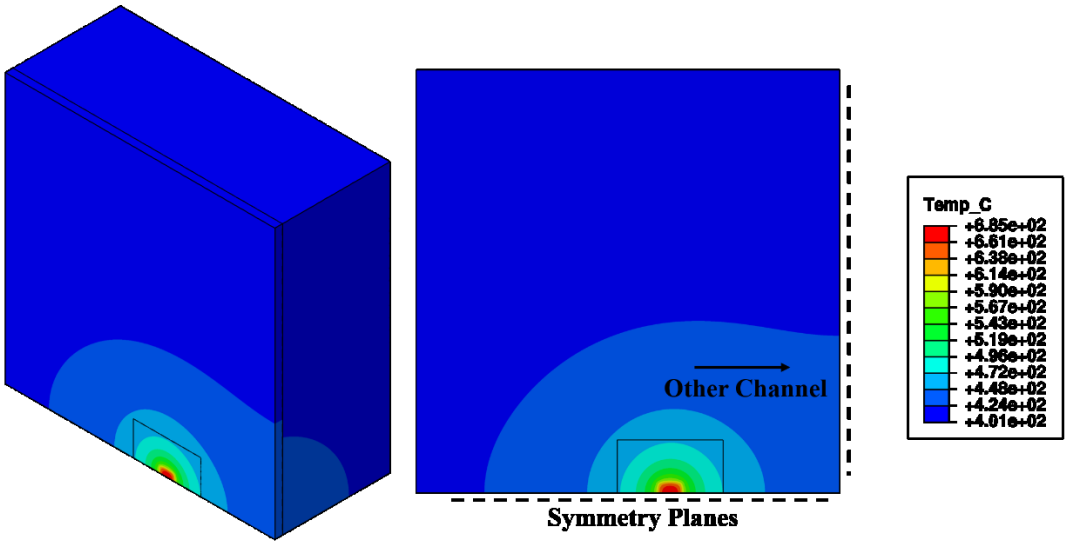


Figure 86: Multichannel Global Model Chiplet Temperature Contour (Nominal 1.6 mm spacing, k"=2e3 W/m^{2}K)

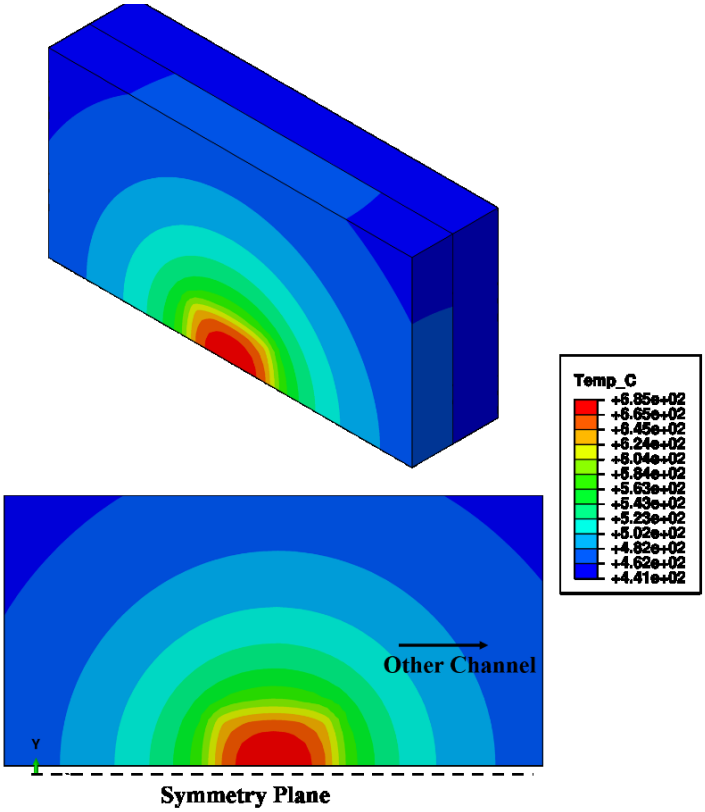


Figure 87: Multichannel Global Model Chiplet Region of Interest Temperature Contour (Nominal 1.6 mm spacing, k"=2e3 W/m^{2}K)

Figure 88 illustrates the board's temperature contour from the global simulation. The circuit board sees a peak temperature of 105°C during multichannel operation, compared to the single-channel case's temperature of 66°C. Accounting for the assumed ambient temperature of 27°C, the board's temperature rise has doubled from 39°C to 78°C. This finding serves as a sanity-check on the multiscale simulation's behavior. As the model scales from one to two channels, total chiplet heat dissipation doubles from 5W to 10W while the board's convective area remains constant, so the board temperature must also double.

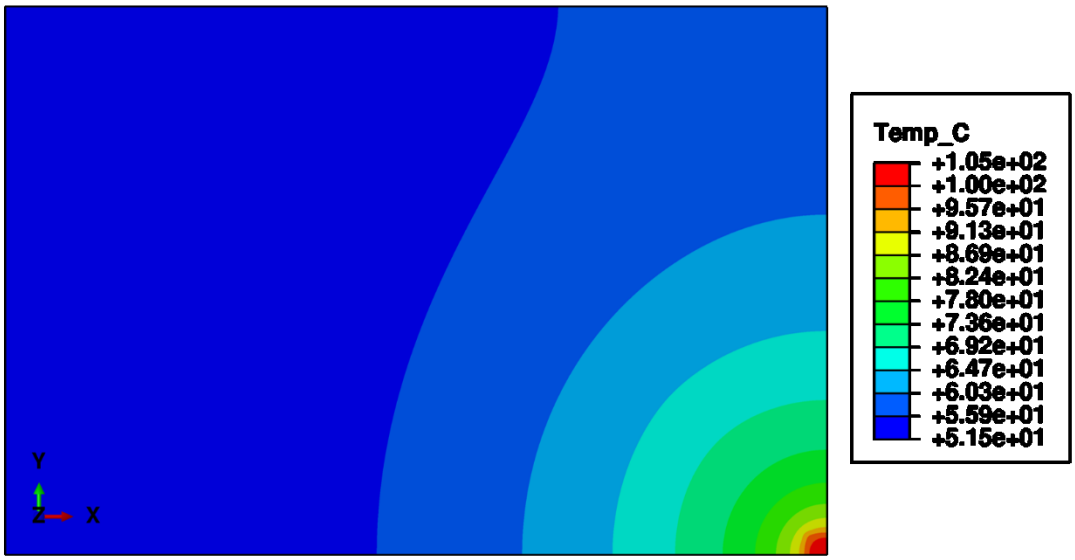


Figure 88: Multichannel Global Model Board Temperature Contour

6.2.2 Multichannel Submodel Simulation Results & Discussion

Figure 89 contains the temperature contour of the multichannel submodel region. The detailed simulation predicts a peak temperature of 784°C; 99°C hotter than the temperature predicted by the global model in the previous section. Such a discrepancy supports the need for high-resolution detail near the PA. At this larger scale, such detail would not be possible without the advantages of the multiscale submodeling approach deployed throughout this thesis. The device's temperature rise corresponds to thermal resistance of 75.7°C/W, higher than the value

encountered in the single channel case. The increased thermal resistance of the device is attributed to the behavior of the die-attach interface in the multichannel case, as will be explained further in Section 6.3 Figure 90 shows the same submodel region viewed from the top face of the GaN substrate. Here, close observation shows evidence of asymmetric heating at the inside and outside of the submodel region.

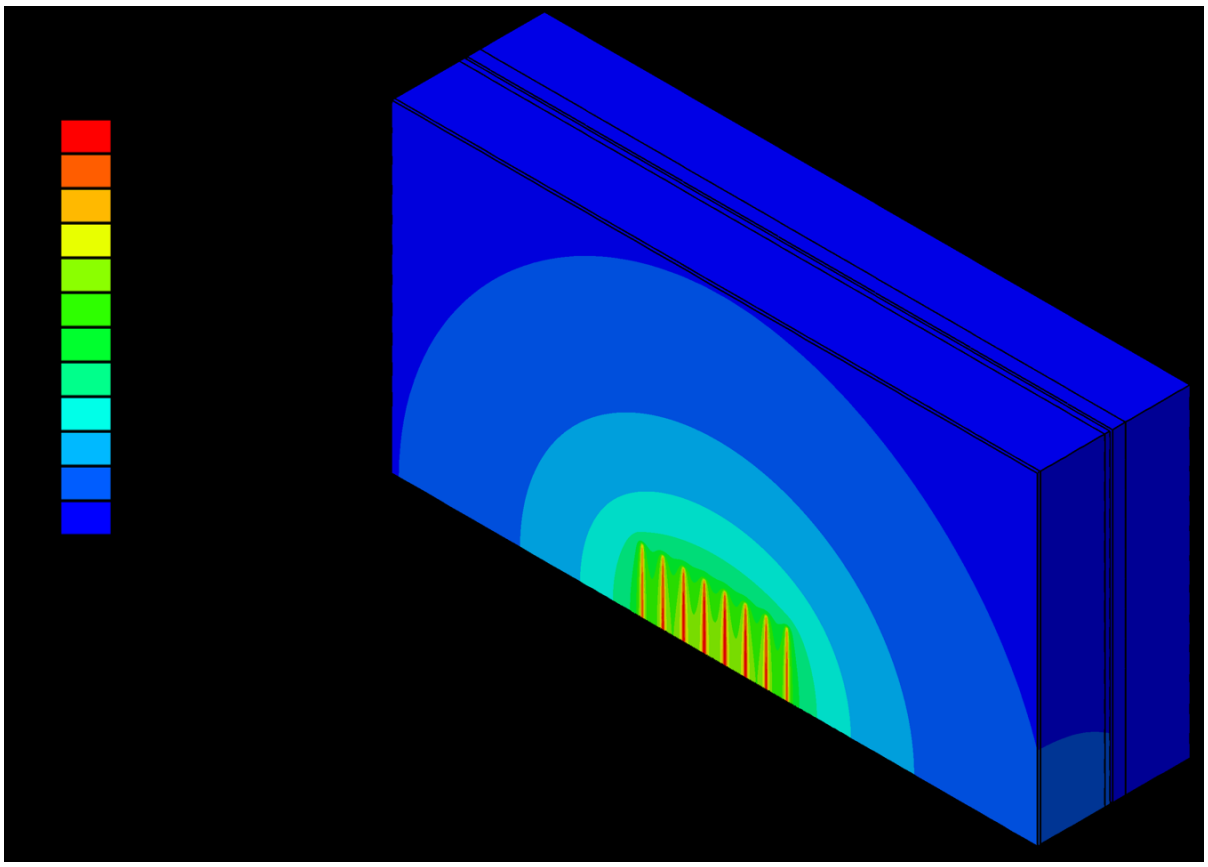


Figure 89: Multichannel Submodel Temperature Contour

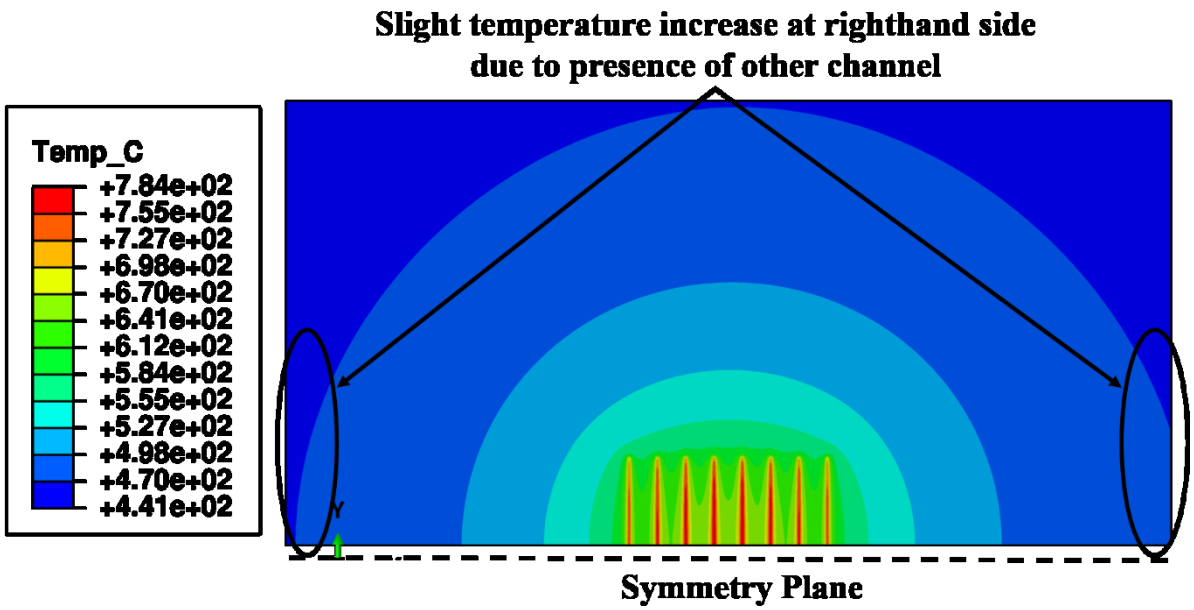


Figure 90: Multichannel Submodel GaN Temperature Contour

As noted above, the multichannel submodel includes only 50 um of bulk silicon to reduce the overall complexity of the model. Inspection of Figure 89 shows that the lower substrates' temperatures are uniformly heated to ~441°C in the submodel. Temperature uniformity at this interface shows most thermal activity of interest occurs far away from the submodel boundary conditions, and suggests that inclusion of more silicon substrate material in the detailed region would not significantly alter the simulation's results.

Figure 91 illustrates the temperature contour along the top of the CMOS face. As indicated by the triangular region, only nine total HICs experience temperatures in excess of 540°C. As encountered in the single-device simulations, thermal isolation of individual HICs causes those nine central HICs to carry most of the heat dissipation from the upper GaN/SiC substrates. The quick dropoff in temperatures seen in the outer regions of the near-field HIC array suggest that many of these HIC instances carry less heat individually, and thus are less critical in terms of reducing the PA's peak temperature.

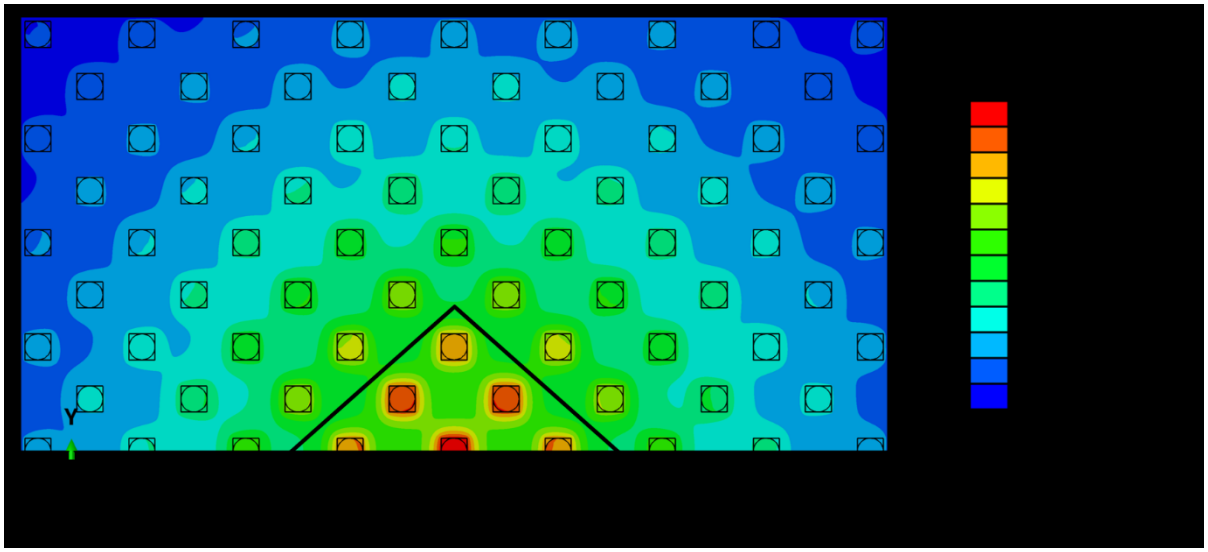


Figure 91: Multichannel Submodel CMOS Temperature Contour

Figure 92 illustrates the distribution of heat fluxes across the HIC interface surfaces. As in the single device case, central HICs in the indicated region carry a disproportionately high percentage of total heat dissipation due to their proximity to the PA. In this case, the result shown below indicates that these nine HICs account for 23% of the total heat dissipation from HICs into the lower substrate.

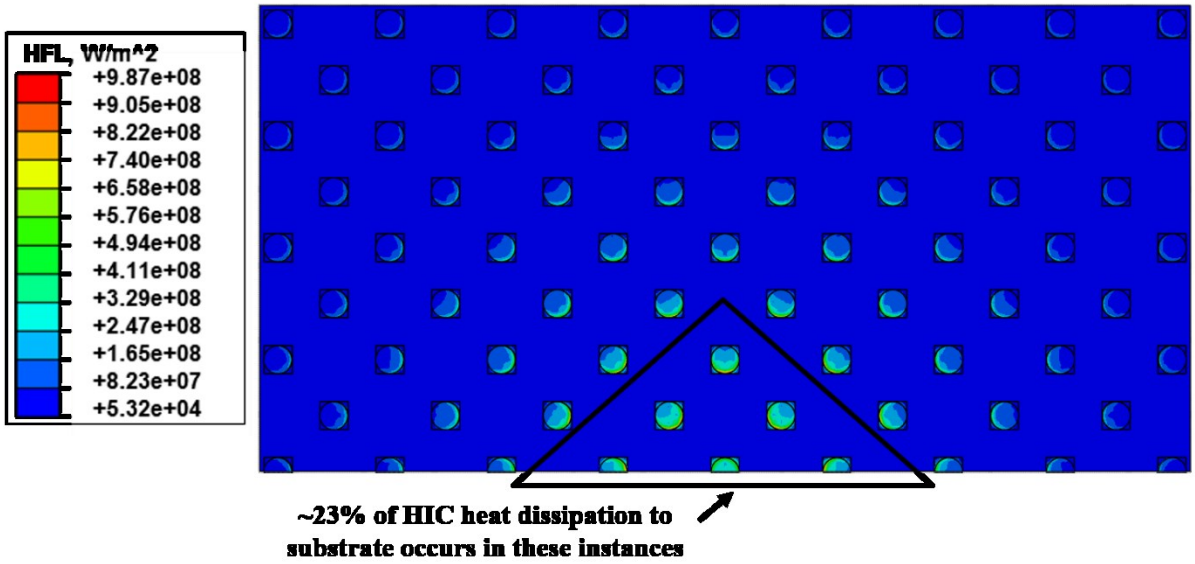


Figure 92: Multichannel HIC Heat Flux Distribution

6.3 Multichannel Model Die-Attach Resistance Study

This section includes a series of simulations which describe the effect of varying the die-attach conductance parameter within the multichannel model, as was done in Section 5.5 for the single channel model. For these simulations, inter-device spacing was held constant at its nominal value of 1.6 mm. The same range of representative die-attach conductances listed in Table 9 were re-simulated in the multichannel model for the worst-case 5W heating scenario.

Figure 93 illustrates the global and submodel peak temperatures from the multichannel die-attach simulations. For comparison, the dotted lines indicate the global and submodel simulation results assuming a zero-resistance tie constraint at the die interface. In general, the curves exhibit the same behaviors described previously in Section 5.5: the coarse global model generally under-predicts submodel peak temperatures by ~100°C, and die-conductance values below $k'' = 8e3 - 10e3 \frac{W}{m^2K}$ lead to temperatures approaching the GaN device's 500°C failure limit. Furthermore, Figure 94 plots the multichannel simulation's minimum chiplet temperature as a function of die-attach conductance. For conductances below $k'' = 12e3 \frac{W}{m^2K}$, the entire 4 mm x 4 mm chiplet is

heated to temperatures above the 150°C failure limit for CMOS components. The multichannel model's far-field temperatures are all significantly hotter than the equivalent single channel cases because total heat dissipation has doubled; this fact alone greatly increases the likelihood of the device exceeding the CMOS failure limits during multichannel operation.

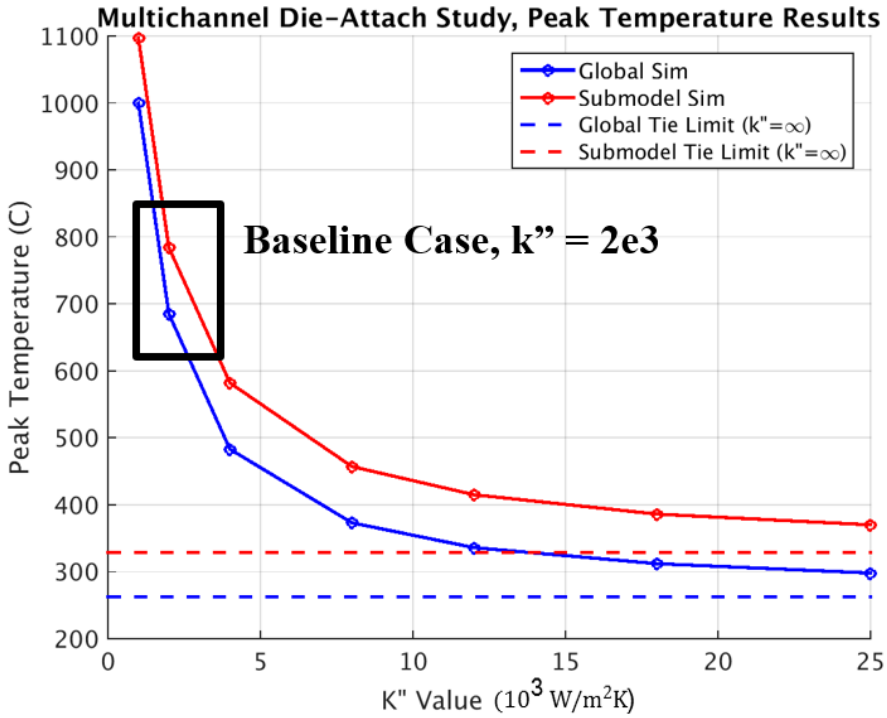


Figure 93: Multichannel Die-Attach Study Results

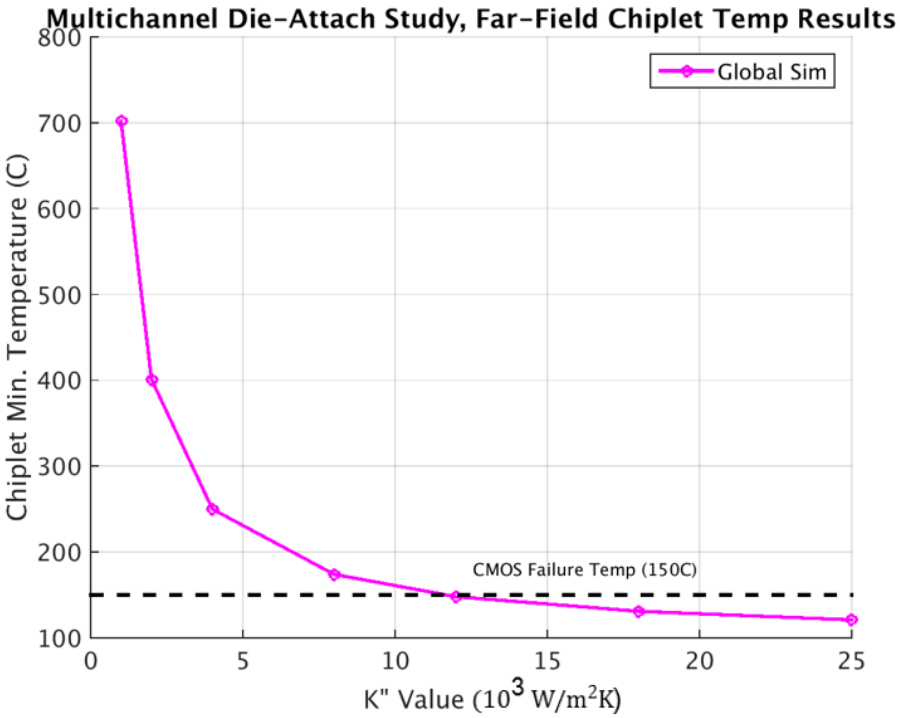


Figure 94: Multichannel Far-Field Chiplet Temperature Results

Comparison between single channel and multichannel die-attachment studies is of particular interest; Figure 95 compares the submodel peak temperature results from the single and multichannel simulations for each die-attach conductance case. At high conductances, the multichannel simulations see temperature increases on the order of ~100°C larger than those encountered in single-device simulations. Despite the doubling of total heat dissipation experienced during multichannel operation, temperature deltas of 100°C correspond to only 25-30% increase from the single channel case, as indicated in Figure 96. However, at die-attach conductances below $k'' = 8e3 \frac{W}{m^2K}$, these relative increases grow to 40-50%.

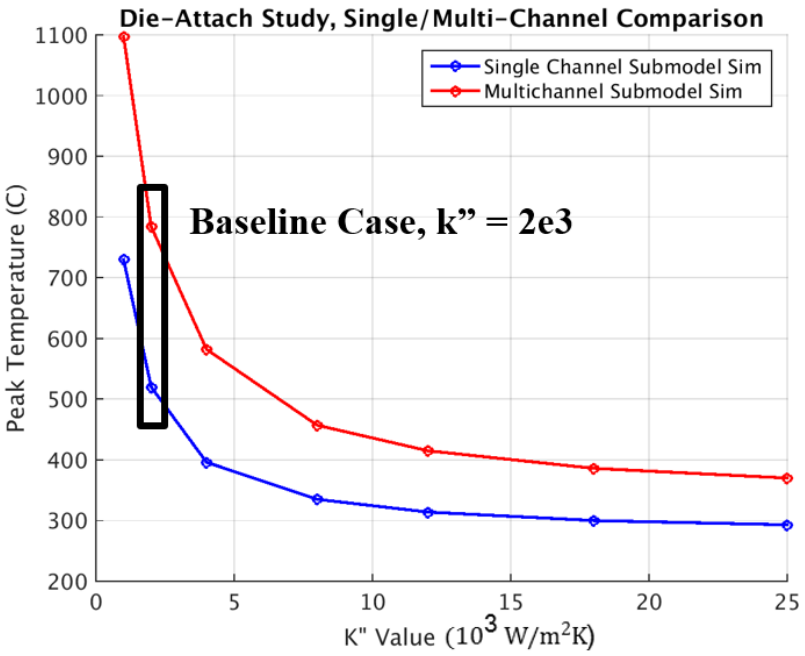


Figure 95: Single/Multichannel Die-Attach Study Comparison

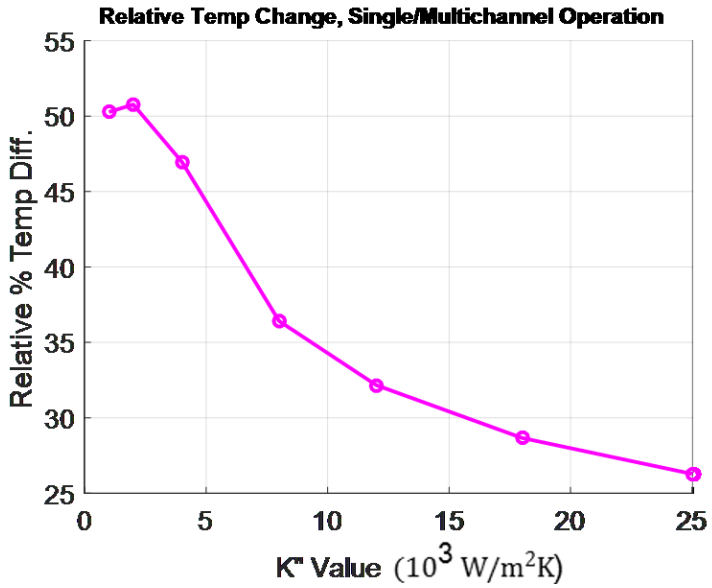


Figure 96: Relative Temperature Increase from Single-Channel to Multichannel Operation

The change in behavior at low and high conductance regimes indicates a transition between two resistive terms governing the device's overall thermal resistance. At low

conductance, the die-attach adhesive dominates device heating. In this state, heat remains trapped within the confines of the chiplet substrates. Such conditions cause the two PA heat sources to appear as if their net thermal effects are additive; that is, the lack of isolation between the devices causes their influence to blend together. At high die-conductances, heat escapes readily through the bottom of the chiplet to be dissipated via convection in the board. In this case the lack of thermal resistance at the interface means the lower substrates remain cool as heat flows downhill away from the PA. This efficient wicking of heat away from the PA means the two channels' thermal effects are less additive.

As suggested above, the doubling of heat flux across the die-attachment interface is suggested to drive most of the temperature increases associated with multichannel operation. According to the equation shown below, the additional 5 watts dissipated by the multichannel case are associated with temperature rises at the adhesive interface ($\Delta T_{k''}$) which vary with the magnitude of the assumed conductance.

$$\Delta T_{k''} = \frac{\Delta q_{heat}}{k''A} = \frac{5W}{k'' \ (16e-6 \ m^2)}$$

Figure 97 illustrates this behavior. The black line plots this calculated temperature rise ($\Delta T_{k''}$) associated with the additional multichannel heat flux. The magenta line plots peak temperature difference between the single and multichannel cases, taken directly from the simulation result. For most die-conductances the trend between the two lines is similar, which indicates that $\Delta T_{k''}$ is primarily responsible for the device's temperature trend as k″ varies. The *peak temperature* overshoots the die-attachment $\Delta T_{k''}$ prediction line by about 80-100°C, but this discrepancy is likely associated with the presence of material nonlinearities close to the PA which exacerbate temperature rises in that region.

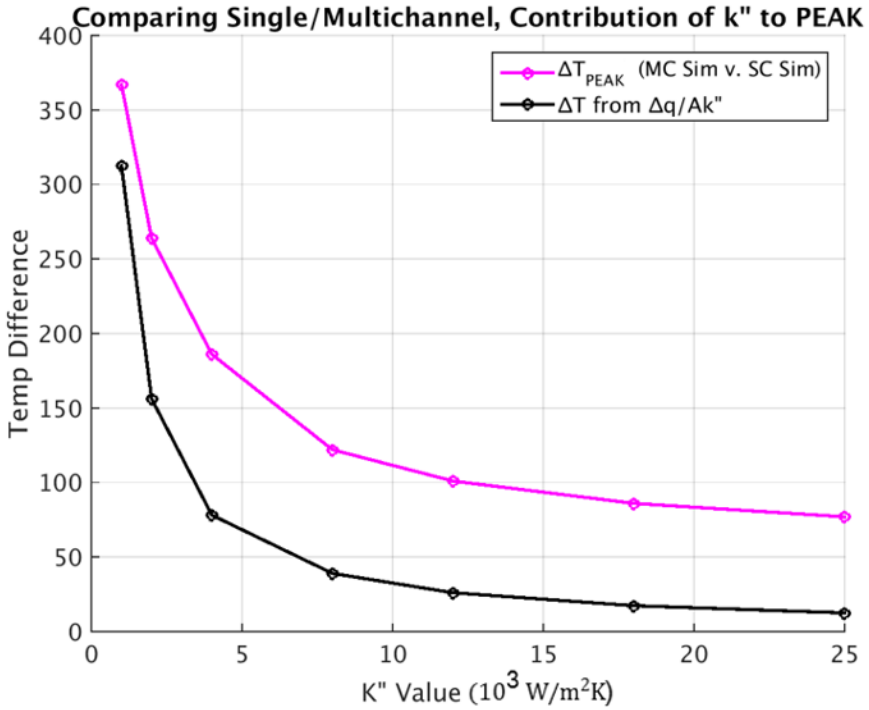


Figure 97: Conductance Term's Contribution to Differences between Single and Multichannel Cases; peak temperature comparison

The $\Delta T_{k''}$ prediction described above is more applicable if far-field chiplet temperatures are considered rather than overall peak temperatures. In Figure 98, the magenta line indicates the same comparison between single and multichannel temperature, but for the minimum temperature at the backside of the chiplet rather than peak temperature. The black line plots the same $\Delta T_{k''}$ described above; here, there's a more direct correlation between temperature rise at the interface and far-field chiplet temperature. This finding reinforces the belief that die-attach thermal resistance primarily drives device heating in the multichannel case.

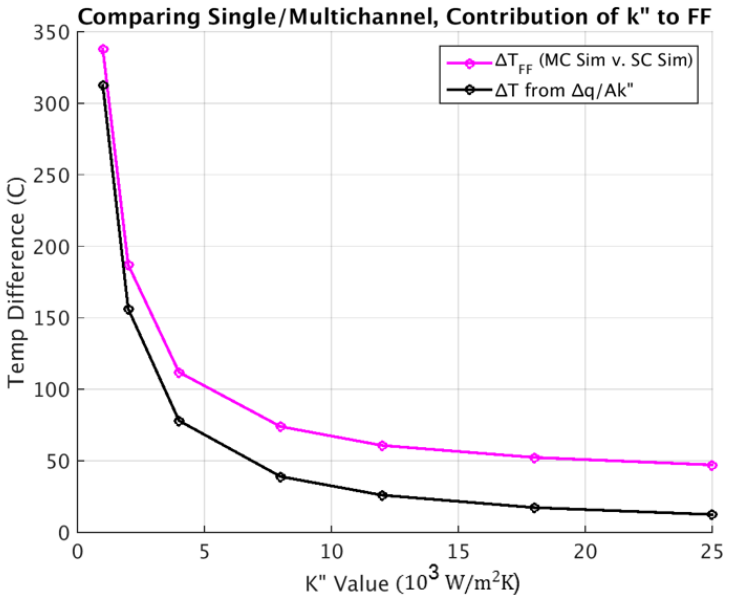


Figure 98: Conductance Term's Contribution to Differences between Single and Multichannel Cases; far-field minimum temperature comparison

6.3.1 Generalized Multichannel Conductance Study

Prior to studying the multichannel model's results, significance of the die-attachment interface in the multichannel case was underestimated. Thus, the dramatic temperature increases associated with the low conductance cases presented above were initially surprising. A series of simplified simulations were developed to quickly verify the model's behavior, and further illustrate the significance of the die-attachment interface.

In these simulations, simplified representations of the DAHI chiplet substrate and its PCB base are linked via a die-attach conductance as shown in Figure 99. At the top of the substrate, heat sources replicate the location of two devices operating simultaneously on the same chiplet, spaced 1 mm apart. For comparison, each multi-channel case was compared to a single-channel baseline with only one heat source. Conductivities of the substrate $\left(k_1 = 150 \frac{W}{mK}\right)$ and base $\left(k_2 = 100 \frac{W}{mK}\right)$ materials were chosen to approximate the device itself; in this case the conductivities were assumed temperature-independent for simplicity. Five watts of heat

dissipation were applied to a small area at the location of each heat source, and the bottom of the base substrate was set to a prescribed boundary temperature of 27°C. As described in previous chapters, convection from the chiplet plays a negligibly small role in device heating, so convection was not included in these simulations. Three die attach conductance values of $k'' = 1e3, 1e4, \text{and } 1e5 \frac{W}{m^2K}$ were chosen to conceptually represent poor, fair, and excellent die-attachment adhesives.

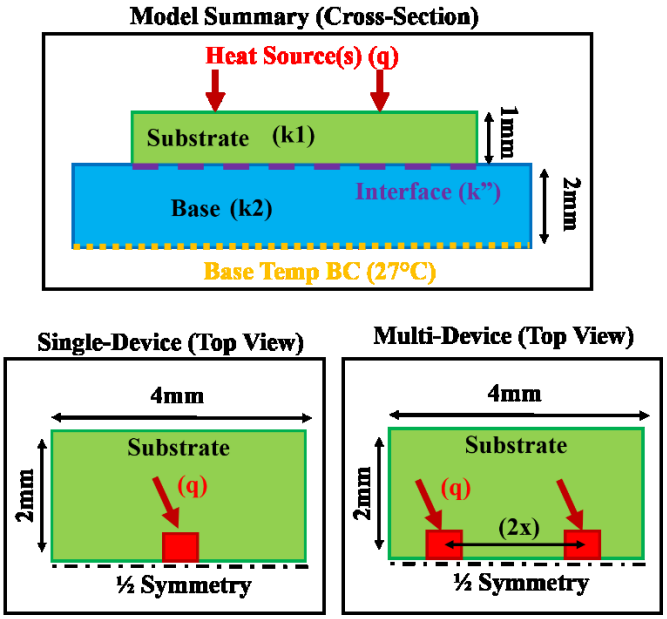


Figure 99: Illustration of Simplified Model Used in Generalized Conductance Study

Figure 100 contains the simulated temperature contours for each conductance case; similarly, Figure 101 enumerates the simulations' peak temperatures for side-by-side comparison. In the two higher conductance cases $\left(k'' = 1e4 \text{ and } 1e5 \frac{W}{m^2K}\right)$, multi-channel operation increases peak device temperature by 30% and 10% respectively. Here, the multi-channel devices dissipate twice as much total heat, yet remain partially isolated from each other such that temperature increases by only a fraction compared to the single-device case. However, in the low-conductance case temperature almost doubles, increasing by 75% over the single-

device baseline. This ‘doubling’ of peak temperature indicates that heat is being trapped within the chiplet. At this low die-conductance, the thermal bottleneck of the die-attach adhesive causes the two devices to co-interact enough that they almost appear as a single, magnified heat source.

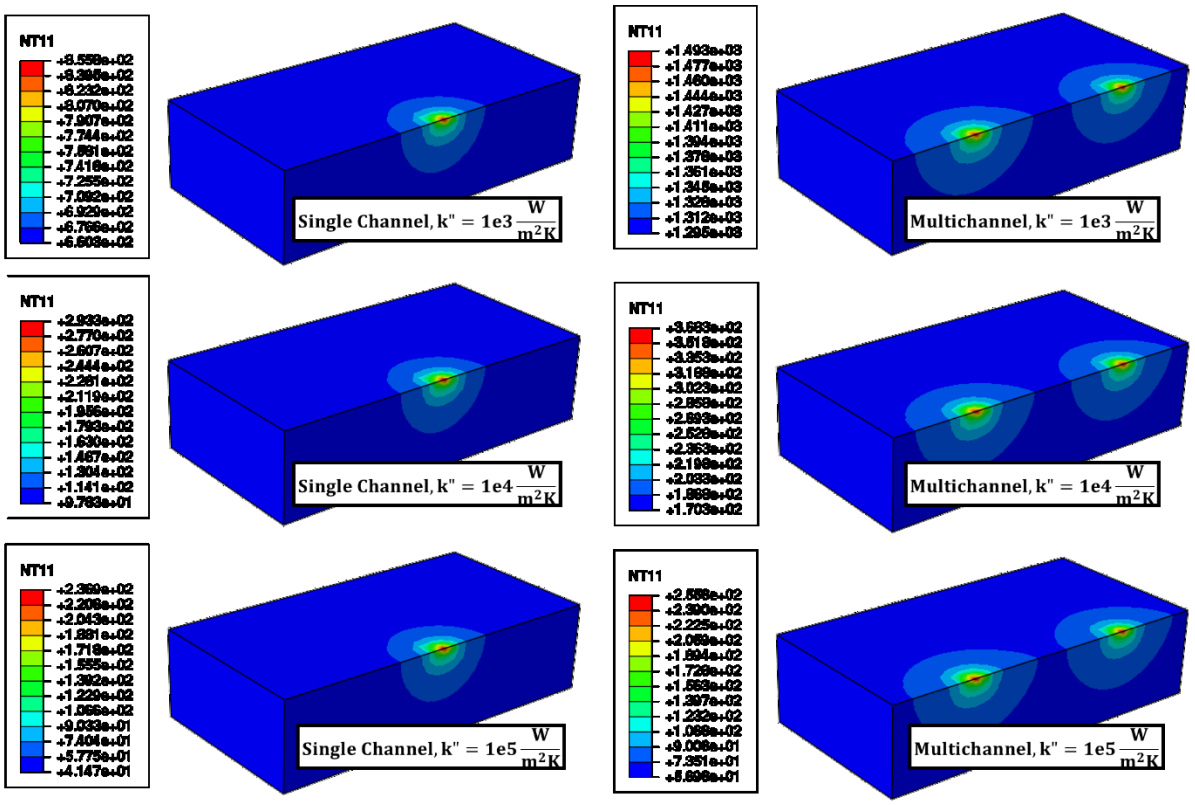


Figure 100: Generalized Conductance Simulation Temperature Contours

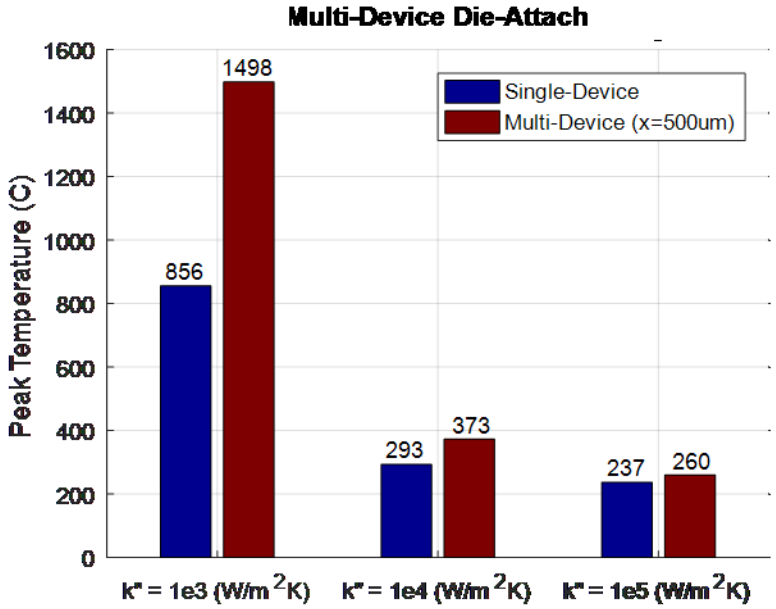


Figure 101: Generalized Conductance Study Results Summary

While these simulations are only abstractions of the actual device, this investigation further illustrates the added significance of the die-attachment interface in the multichannel case. Because little heat escapes the chiplet via convection, almost all of the additional heat introduced during multichannel operation must pass through the die-attachment resistance. If this interface is not well-conditioned, its resistance alone could be responsible for nearly doubling temperature within the device.

6.4 Multichannel Device Spacing Study

In multichannel operation, device temperatures also depend on the spacing between individual heat sources. From a thermal perspective, increasing spacing between channels allows more conductive area for heat to dissipate through, which in turn reduces temperatures. However, from a circuit designer's perspective, increased spacing comes at a price of increased device size/weight/cost, and interconnect delays. The goal of the DAHI program's heterogeneous

integration effort is to "intimately" combine different technologies, so sprawling out over large area defeats the purpose of heterogeneous integration.

This section presents a study in which inter-channel spacing is varied within the device. By determining to what extent channel spacing governs overall device temperatures, circuit designers will be better informed when weighing thermal considerations against physical size constraints. As noted above in Section 6.1, the multichannel finite element model was designed to be easily re-configurable to any inter-channel spacing between the two Stage 3 PAs. In these simulations, the device was studied for spacings between 0.5 and 2 millimeters. Re-simulation required updating the global model's geometric definition to account for changing spacing in each case. As explained previously in Figure 85, the submodel's geometric definition remains unchanged between simulations. As suggested by the previous section, the device's multichannel behavior varies significantly at different die-attach regimes, so the spacing study was repeated for three representative conductance cases k" $= 2e3, 8e3, \text{and } 25e3 \frac{W}{m^2K}$.

Figure 102 illustrates the submodel simulation's peak temperature as a function of inter-channel spacing for three different die-attachment conditions. In each curve, peak temperature increases by a few degrees as spacing decreases. Temperature remains constant for spacings between 1800-2200 um, indicating that the two channels are spaced sufficiently to insulate them from one another. At the maximum spacing of 2400 um, each device is approaching the outside edge of the chiplet, which reduces the area available for heat dissipation through the substrate and causes a slight increase in peak temperature.

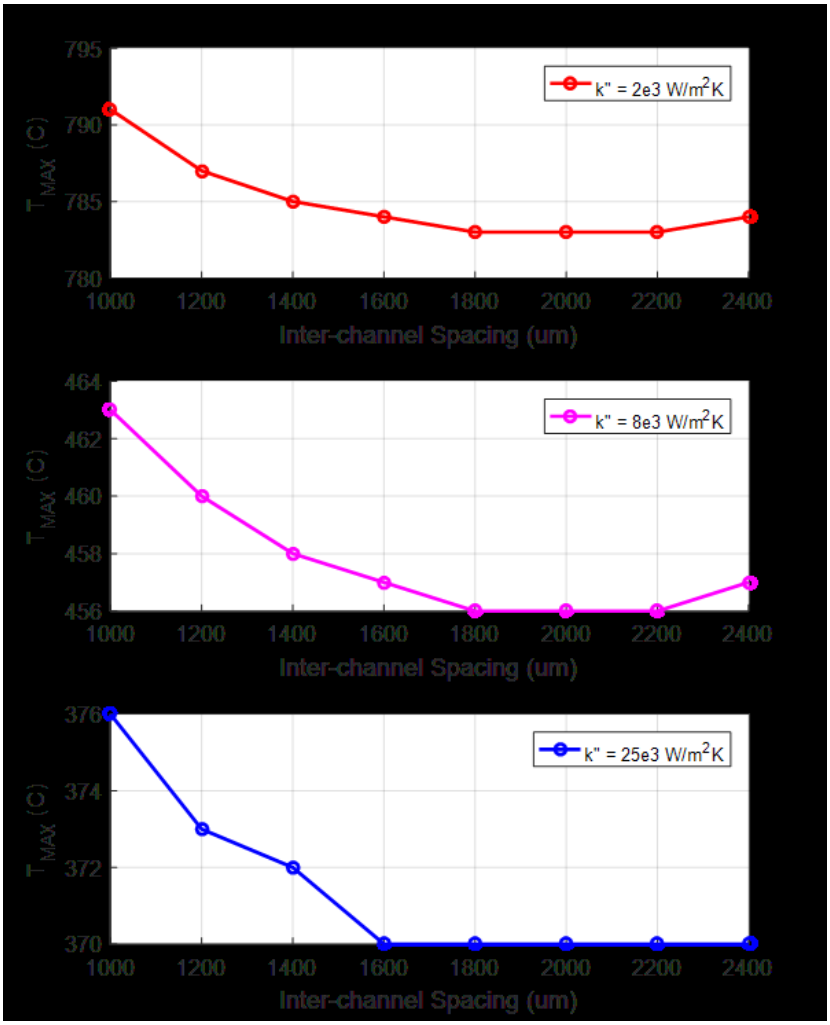


Figure 102: Multichannel Spacing Study Results for Three Die-Attach Cases

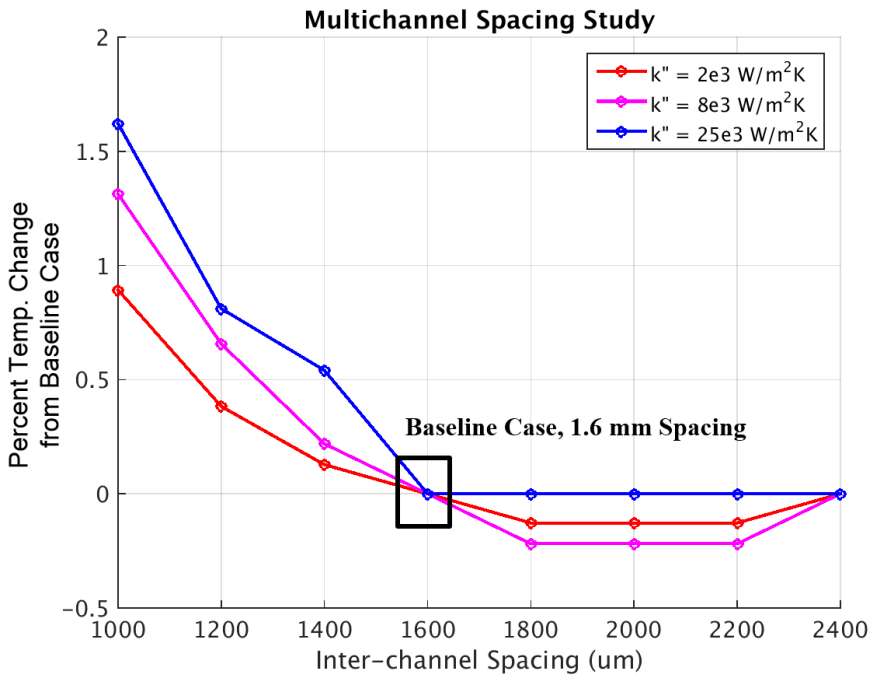


Figure 103: Multichannel Spacing Study, Relative Change

Figure 103 illustrates these same results in terms of percent change from the baseline 1.6 mm spacing condition. Varying device spacing leads at most a 2% relative change in device peak temperature, which suggests that designers' placement of channels does not affect device heating substantially. However, peak temperature is not the only metric appropriate for consideration in this study; inter-channel spacing strongly affects the device's temperatures in the regions far from the PA.

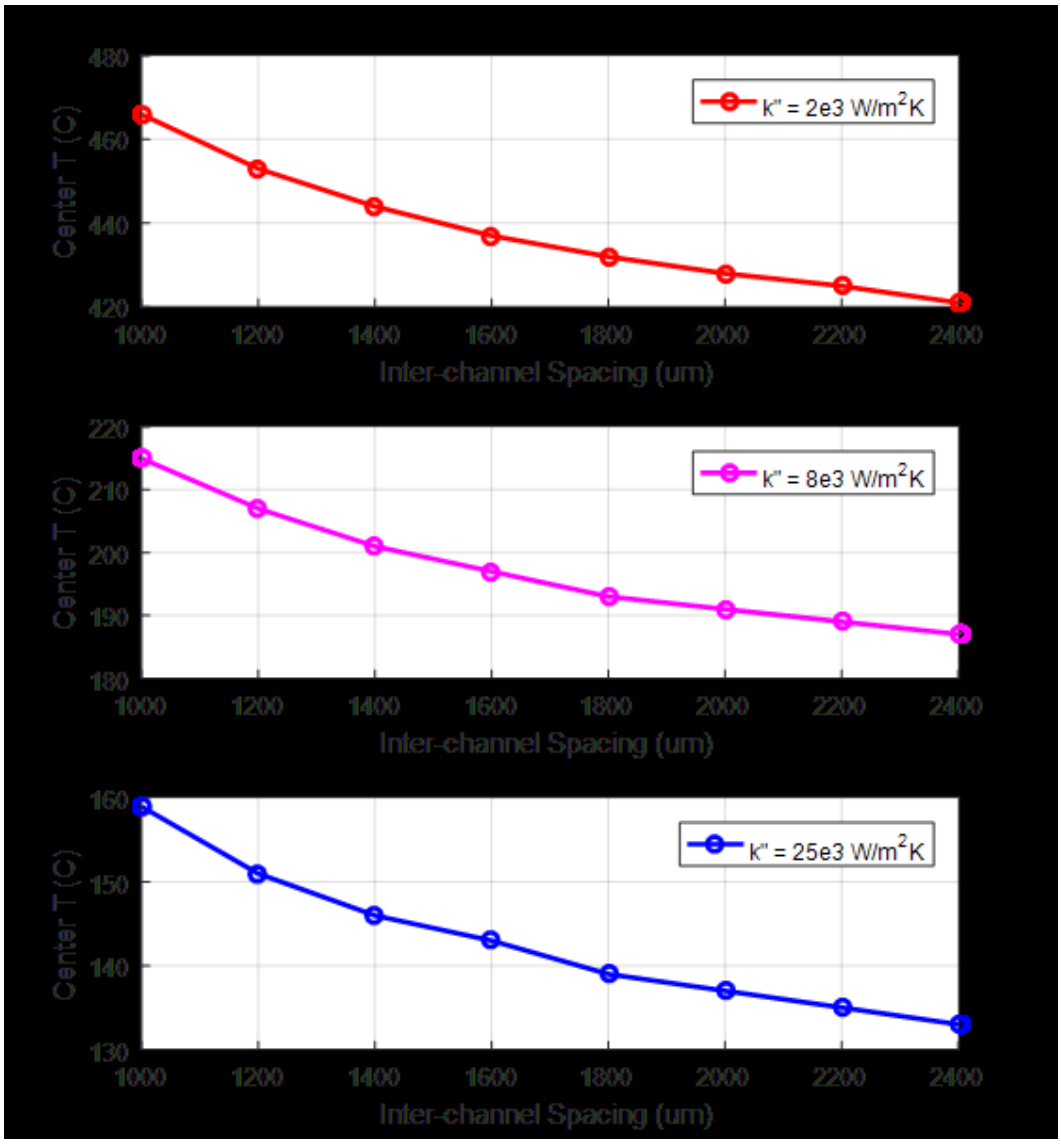


Figure 104: Multichannel Study, Centerline CMOS Temperatures

For example, Figure 104 illustrates spacing's effect on device temperatures measured in the CMOS substrate at the center point between the two PAs for three different values of die-attach conductance. At this location, temperatures vary by ~30°C across the configurations being studied, which amounts to roughly 20% possible temperature variation in this region. The two high conductance cases shown are particularly interesting; at large inter-channel spacings, the device's center temperature begins to decrease to the upper temperature limit for CMOS. This suggests that, if a suitably conductive adhesive were used, the two channels may be spaced such

that active CMOS components could safely be placed in the regions between the two channels, even at the worst-case 5W heating scenario. Safe placement of CMOS circuitry is explored in more detail for various power levels in the next section.

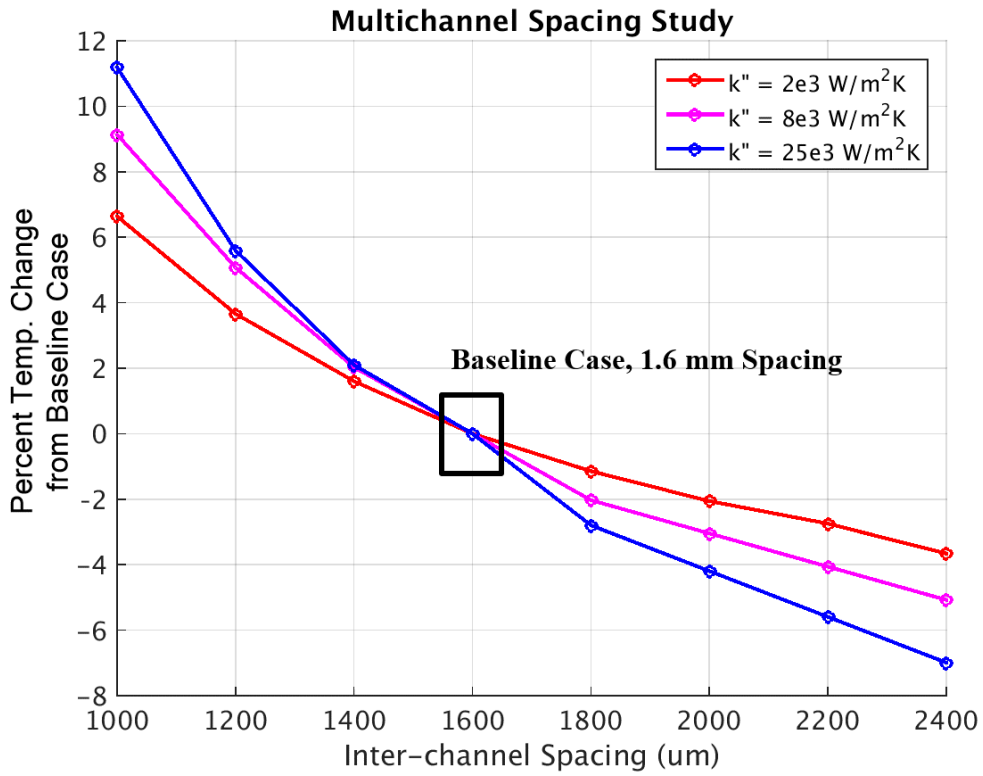


Figure 105: Multichannel Study, Centerline GaN Relative Temp. Change from Baseline

6.5 Multichannel Device Power Study

This section describes simulations of the multichannel device at power levels from 1W to 5W similar to the simulations presented in Section 5.6. These simulations were repeated for three representative die-attachment conditions. Based on the previous section's conclusion that inter-channel spacing minimally impacts device peak temperature, all simulations in this section assume the device's nominal 1.6 mm inter-channel spacing.

Figure 106 illustrates the simulated peak temperature curves for each die-attachment case as a function of dissipated heat. As in the single device power study illustrated in Figure 78,

temperatures in the low conductance case are dramatically higher than in the other two cases at all power levels. Temperatures between the k″ = 8e3 and k″ = 25e3 cases match more closely, differing by <100°C at the 5W dissipation level. Compared to the single-channel simulations, the multichannel simulation's temperature appears to increase more linearly as a function of dissipated power. However, the plot of overall device thermal resistance shown in Figure 107 indicates that device thermal resistances increase at higher power levels; this indicates that the material nonlinearities introduced in Section 15.4 are still influencing the simulation results.

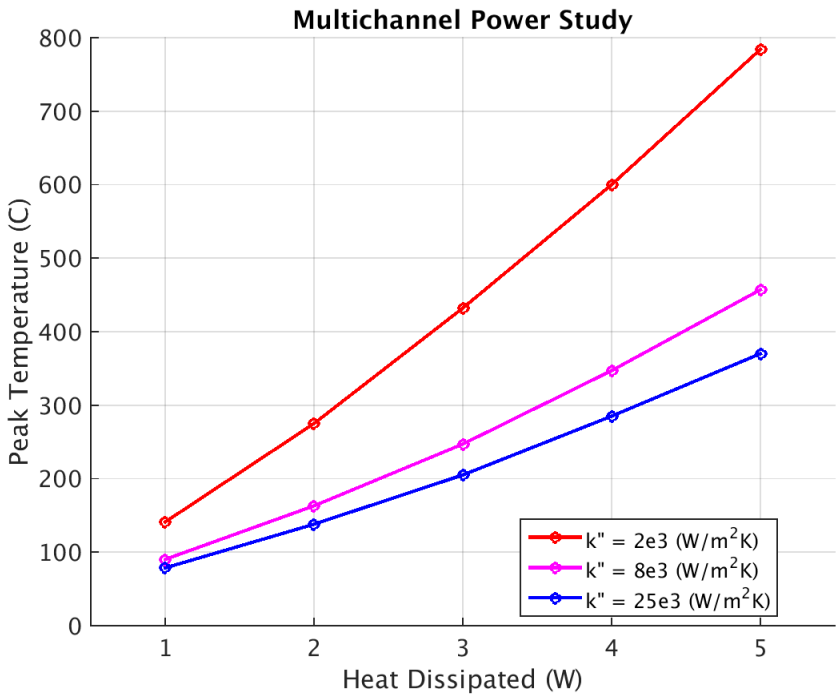


Figure 106: Multichannel Power Study, Temperature Results

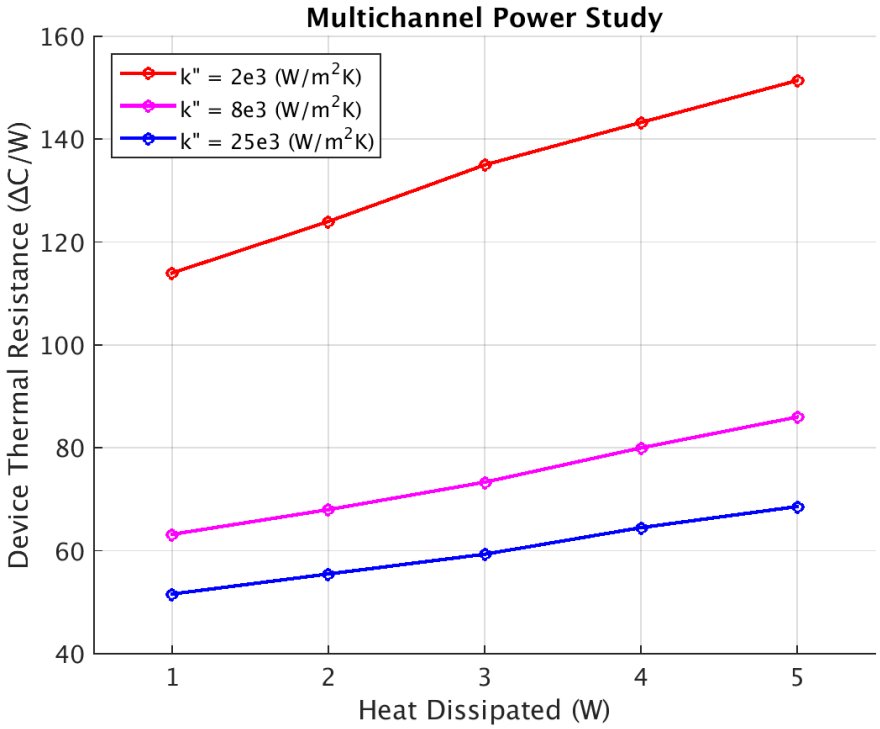


Figure 107: Multichannel Power Study, Thermal Resistance Results

The lack of curvature in the trends shown in Figure 106 likely reflects a tradeoff between thermal resistances attributed to nonlinear material properties versus thermal resistance of the die-attachment interface. Previous chapters explained in detail that the thermal bottleneck at the base of the device dominates the device's thermal behavior; this is especially true in the multichannel scenario, where twice as much heat dissipation occurs across the same fixed interfacial area. The conductance parameter used in these simulations has been assumed constant due to the lack of temperature-dependent experimental data available in literature. Thus, the temperature increase associated with die-attachment resistance is two times larger in the multichannel case than the single-channel case, and scales linearly with temperature. For these two reasons, the linearity of the die-attachment interface appears to mute the curved nonlinearities previously associated with the device's temperature profile at high power levels.

6.6 CMOS Safe Keep-Out Distance

In the GaN substrate, thermal failure is primarily determined by the PA's peak temperature as the heat source corresponds to the location of the PA. However, in the CMOS substrate critical components are placed far away from the GaN PA, i.e. far from the location of maximum CMOS temperature, so peak temperature is less useful for failure predictions. Rather, because CMOS components' temperatures might seem to depend strongly on their physical proximity to the PA, evaluation of the design's thermal characteristics should take the layout geometry into account and focus on local temperatures at specific locations away from the PA. In this section, CMOS thermal results are presented at various locations across the chiplet to reflect the design intent of thermally insulating the sensitive CMOS components.

A primary goal of heterogeneous integration is to shrink the overall size of the device, so closely packing CMOS components near the PA is of interest for designers. Thus, the results of the finite element simulations may be used to identify locations suitably far from the PA whose temperatures satisfy CMOS thermal failure limits. A minimum keep-out radius around the PA may serve as a useful parameter for determining safe placement of CMOS components within the device, as illustrated in Figure 108. For the multiscale simulations developed above, distance between the center of the PA region and the temperature contours corresponding to failure limits for CMOS were recorded for a variety of heating and die-attachment scenarios.

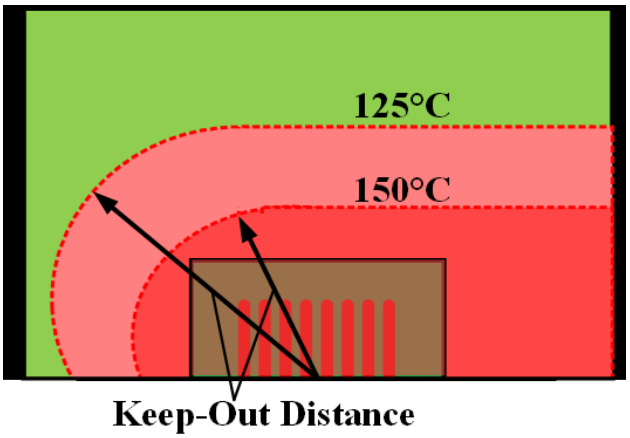


Figure 108: CMOS Keep-Out Distance Illustration

The first set of results given below in Table 10 list the safe keep-out distances reported for CMOS temperature limits of 125°C and 150°C respectively. Red cells in the table represent cases where the entire chiplet is too hot for CMOS; similarly, green cells indicate cases whose chiplet temperatures are all safely below the CMOS thermal limits. The few yellow highlighted cells in the table indicate cases where the CMOS keep-out radius is applicable. In these few cases, CMOS components can be safely placed anywhere outside the specified radius from the Stage 3 PA. The data presented in the table is mostly binary; depending on the combination of die-attachment condition and heat dissipation, most cases are either entirely too hot or completely safe for CMOS operation. Table 11 presents the same results for the $k'' = 8000 \frac{W}{m^2K}$ case at smaller power intervals.

The binary nature of these results illustrates the sensitivity of the CMOS substrate to changes in the device's power dissipation or die-attachment conditions. At high power levels the CMOS substrate is entirely too hot for safe operation, regardless of die-attachment condition. In terms of survivable operation for reasonable die-attachment conditions, the device's heat dissipation must be kept low by maintaining high efficiency within the Stage 3 PA. In operation, this could be accomplished by constraining the device to a smaller range of operational

frequencies, or possibly by adjusting the bias conditions used during testing. In either case, any effort toward improving the PA's efficiency would require a fundamental change to the device's intended capabilities. However, designers should consider that the upper limit of 5W heat dissipation was only encountered in a few tests; the majority of tested frequencies produced heat dissipations on the order of 2-3.5W. At these dissipation levels, CMOS components will become survivable if careful attention is paid to the conductance of the die-attachment interface. Instead of re-designing the device, survivable operation may be achieved if the interfacial conductance is confirmed to be suitably high.

Table 10: Safe CMOS Keep Out Distance Results for Power and Die-Attachment Cases

	CMOS Keep Out Distance (Multi Channel)						
	K"	2000 W/m^{2}K		8000 W/m^{2}K		25000 W/m^{2}K	
	Limit	125°C	150°C	125°C	150°C	125°C	150°C
Power	5W	X	X	X	X	X	250 um
	4W	X	X	X	1350 um	260 um	✓
	3W	X	X	500 um	✓	✓	✓
	2W	X	X	✓	✓	✓	✓
	1W	✓	✓	✓	✓	✓	✓

KEY	
X	Too hot; no safe distance exists on chiplet
✓	Suitably cool; temperatures on entire chiplet do not exceed CMOS limits
### um	Specifies minimum safe distance from PA to satisfy CMOS temperature limit

Table 11: Safe CMOS Keep-Out Distance Results, K"=8000 W/m^{2}K, Intermediate Power Levels

	CMOS Keep Out Distance (Multichannel)		
	KP	8000 W/m^{2}K	
	Limit	125°C	150°C
Power	5W	✗	✗
	4W	✗	1350 um
	3.75W	✗	530 um
	3.5W	✗	260 um
	3.25W	220 um	150 um
	3W	500 um	✓
	2.75W	✓	✓
	2.5W	✓	✓
	2.25W	✓	✓
	2W	✓	✓
	1W	✓	✓

KEY	
✗	Too hot; no safe distance exists on chiplet
✓	Suitably cool; temperatures on entire chiplet do not exceed CMOS limits
### um	Specifies minimum safe distance from PA to satisfy CMOS temperature limit

In the device being modelled, the primary CMOS components of interest include a phase modulator and CMOS-to-GaN buffer for each of the two channels on the chip [18]. As indicated in Figure 109, these two regions are separated by roughly 1800 um on the chip. Thus, CMOS temperatures at locations ~1800 um are of primary interest with respect to the device's operation. Figure 110 recasts the CMOS temperature results from the multichannel simulations in slightly different form; here, the multichannel simulation's temperature is plotted for various heating and die-attachment conditions at locations 1800 um and 900 um away from the PA in the direction of the CMOS buffer. Three representative die-attachment conditions are studied for power dissipations ranging from 1 to 5 watts. In the figure, solid lines indicate the location corresponding to the buffer 1800 um from the PA, while the colored, dashed lines indicate temperatures for locations 900 um from the PA, primarily to illustrate the effect of spacing on CMOS temperatures. Temperatures of 125°C and 150°C corresponding to CMOS thermal limits are overlaid as black dashed lines for reference.

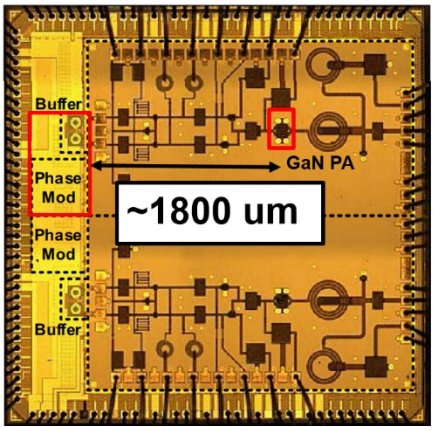


Figure 109: Distance between CMOS Buffer and Main PA, from [18]

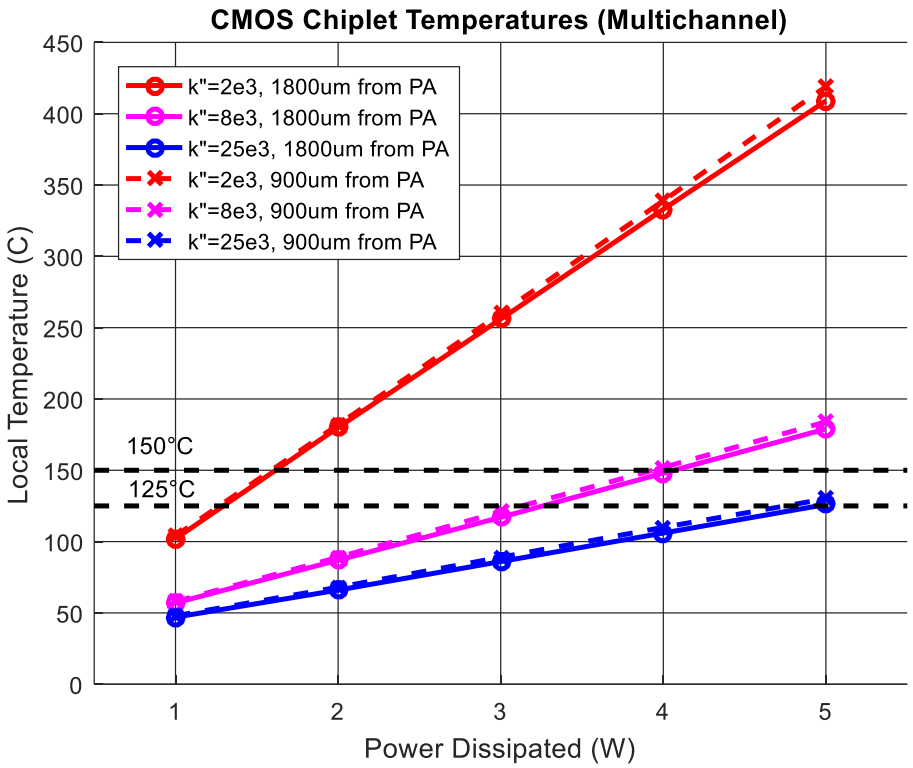


Figure 110: CMOS Chiplet Temperatures at Various Locations

These results indicate less than 5°C difference in CMOS temperature between the 1800 and 900 um curves; in every case, this illustrates that spacing of the CMOS components is not likely to drive their thermal failure. Instead, heating of the CMOS substrate again appears to be

primarily driven by the device's total heat dissipation, and the die-attachment condition. In particular, the low die-attachment case corresponding to a lower-performing silver epoxy is survivable for dissipations below 2 watts. The magenta and blue curves corresponding to moderate and high die-attachment conductances of $k'' = 8e3 \text{ and } 25e3 \frac{W}{m^2K}$ achieve survivable CMOS temperatures for powers between 3-4 watts, much closer to the actual dissipations measured during testing. For the circuit layout designer, this suggests the most critical factors for reducing CMOS temperatures include improving device efficiency and specifying a suitably-conductive die-attachment adhesive. As mentioned above, re-design to improve device efficiency is costly and self-defeating; rather, this finding suggests that better adhesives may be a low-hanging fruit which lead to immediate thermal improvements in the CMOS region.

7 Conclusions and Future Work

Heterogeneous integration will soon deliver performance breakthroughs in next-generation microwave systems. The foundational work of the DAHI program thus far has demonstrated the promise of heterogeneously-integrated GaN-on-silicon technologies, and identified areas of improvement in manufacturing and survivability required to deliver heterogeneous integration from conceptual development to the state of the art in RF systems.

The work presented here is a step toward alleviating thermal challenges which currently limit the functionality of heterogeneously integrated devices. In particular, this thesis identifies which aspects of the device's design are most critical to its self-heating behavior, and seeks to provide fundamental knowledge which will guide future circuit designer's decisions for the next generation of heterogeneous devices. The multiscale modeling approach deployed in this thesis includes high-fidelity resolution required for simulation accuracy, and demonstrates applicability in studying the device's thermal behavior across a range of operating conditions and configurations. The use of submodeling techniques allowed for easy reconfiguration of the model to capture combinations of multiple simulation parameters for a number of design trade studies described in the chapters above. Furthermore, submodeling allowed for the introduction of global, reality-based boundary conditions whose significance was underscored throughout this thesis by demonstrating the significance of the die-attachment adhesive on the device's thermal operation. Uncertainties associated with the assumption of material properties, geometry, and boundary conditions assumed in the model were all examined and quantified in terms of their

influence on the model's results. Finally, the safe operating range of the device was identified in terms of thermal failure limits for both CMOS and GaN.

Moving forward, many aspects of this model require further investigation; in particular, this work sheds light on the significant uncertainty associated with the die-attachment interfacial conductance term. Above all else, this parameter greatly affected the temperatures seen in simulation. In the short term, experimental measurement of the die-attachment adhesive's conductance is required to calibrate the value assumed in the finite element model before the thermal simulation will be capable of precisely predicting temperature rise. Thermal measurements of the adhesive's resistance on a metal carrier by itself, and integrated on a functioning device should be conducted to narrow the uncertainty associated with the adhesive's thermal resistance. At the current time, published literature suggests that available adhesives' resistances fall within up to two orders of magnitude of uncertainty, corresponding to temperature ranges on the order of hundreds of degrees Celsius in the devices considered here. Adhesive measurements and characterization will aid future thermal modeling efforts and will inform circuit designers on selection of die-attachment adhesives suitable for operation in high-power applications. In addition, the following areas may require further investigation based on the initial results explored herein:

- Thermal conductance of the bonded substrate interfaces at the HICs, including any imperfections introduced during the manufacturing process
- Long-term behavior and possible failure of the device at sustained high temperatures
- Optimization of the device's operating range to those frequencies associated with better PA efficiency, which will decrease heat dissipation and device temperatures
- Role of device bias conditions on heat generation within the PA

- Thermal runaway resulting from co-interaction of the device's thermal and electrical behaviors at high temperatures

In the future, adequate thermal management will prove critical to the success of forthcoming high-power devices developed using heterogeneous integration. . The task of solving thermal problems at the layout design stage will require innovative, multidisciplinary solutions leveraging both simplicity required for real-time analysis against the complexity required for accuracy. As the integration technology and thermal analysis techniques both mature, they will improve device reliability and maximize the potential impact of heterogeneous integration in both commercial and military applications.

References

[1] D. Green and e. al, "A Revolution on the Horizon from DARPA: Heterogeneous Integration for Revolutionary Microwave/Millimeter-wave Circuits at DARPA," vol. 18, no. 2, 2017.

[2] S. Raman, C. Dohrman and T. H. Chang, "The DARPA Diverse Accesible Heterogeneous Integration (DAHI) Program: Convergence of compound semiconductor devices and silicon-enabled architectures," in *IEE. Int. Symposium on RFIT*, 2012.

[3] S. Melamed and T. Thorolfsson, "Junction-Level Thermal Extraction and Simulation of 3DICs," in *IEEE Inter. Conf. on 3D Sys. Int.*, 2009.

[4] S. Choi, G. M. Peake, G. A. Keeler and K. M. Geib, "Thermal Design and Characterization of Heterogeneously Integrated InGaP/GaAs HBTs," *IEEE Trans. on Comp. Pack. and Manf. Tech,* vol. 6, no. 5, 2016.

[5] A. D. Kraus and A. Bar-Cohen, Thermal Analysis and Control of Electronic Equipment, Washington, DC: Hemisphere, 1983.

[6] P. Wilkerson, A. Raman and M. Turowski, "Fast, automated thermal simulation of three-dimensional integrated circuits," in *Intersoc. Conf. on Therm.and Thermomech. Phen.in Elec. Sys.*, 2016.

[7] D. Choudhury, "3D Integration Technologies for Emerging Microsystems," in *IEEE MTT-S International Microwave Symposium*, 2010.

[8] T. R. Harris, G. Pavlidis, E. J. Wyers, D. M. Newberry, S. Graham, P. Franzon and W. R. Davis, "Thermal Raman and IR Measurement of Heterogeneous Integration Stacks," in *IEEE Intersoc. Conf. on Therm. and Thermomech. Phen. in Elec. Sys.*, 2016.

[9] S. Melamed, T. Thorolfsson, T. Harris, S. Priyadarshi, P. Franzon, M. B. Steer and W. Davis, "Junction-Level Thermal Analysis of 3D Integrated Circuits Using High Definition Power Blurring," *IEEE Trans. on Comp-Aid Des. of Int. Circ. and Sys.,* vol. 31, no. 5, 2012.

[10] T. R. Harris, E. J. Wyers, L. Wang, S. Graham, G. Pavlidis, P. D. Franzon and], "Thermal simulation of heterogeneous GaN/InP/silicon 3DIC stacks," in *Inter. 3D Sys. Int. Conf.*, 2015.

[11] J. D. Meindl, "Ultra-Large Scale Integration," *IEEE Trans. on Elec. Dev.,* vol. 31, no. 11, 1984.

[12] A. Raman, "System-Level Performance Evaluation of Three-Dimensional Integrated Circuits," *IEEE Trans. on VLSI Sys.,* vol. 8, no. 6.

[13] A. Raman and R. Reif, "Thermal Analysis of Three-Dimensional Integrated Circuits," in *IEEE Interconnect Tech. Conf.*, 2001.

[14] T. Y. Chiang, S. J. Souri, C. O. Chui and K. C. Saraswat, "Thermal analysis of heterogeneous 3D ICs with various integration scenarios," in *IEEE Electron Dev. Mtg.*, 2001.

[15] T. Anderson, F. Ren, L. Covert, J. Lin and S. Pearton, "Thermal Considerations in Design of Vertically Integrated Si/GaN/SiC Multichip Modules," vol. 153, no. 10, 2006.

[16] S. Raman, T. Chang, I. Abdomerovic, C. L. Dohrman, C. Maxey and M. Rosker, "The DARPA COSMOS and ELASTx Programs: Toward Next Generation Linearized Microwave/Mm-Wave Transmitters," in *IEEE Comp. Semi. Int. Circ. Symp.*, 2010.

[17] D. S. Green, C. L. Dohrman, A. S. Kane and T. Chang, "Materials and Integration Strategies for Modern RF Integrated Circuits," in *IEEE Comp. Semic. Int. Circ. Symp.*, 2014.

[18] M. LaRue, B. Dupaix, S. Rashid, T. Barton, T. James, W. Gouty, P. Watson, T. Quach and W. Khalil, "A fully-integrated S/C band transmitter in 45nm CMOS/ 0.2gm GaN heterogeneous technology," in *IEEE Compound Semiconductor Integrated Circuit Symposium (CSICS)*, Miami, FL, 2017.

[19] U. Mishra, P. Parikh and Y. F. Wu, "AlGaN/GaN HEMTs-an overview of device operation and applications," in *Proceedings of the IEEE*, 2002.

[20] K. R. Bagnall, "Device-Level Thermal Analysis of GaN-based Electronics," 2013.

[21] U. Lindefelt, "Heat generation in semiconductor devices," *J. of Appl. Phys,* vol. 75, 1994.

[22] T. L. Bergman, A. S. Lavine, F. P. Incropera and D. P. Dewitt, Introduction to Heat Transfer, Hoboken: Wiley, 2011.

[23] Y. S. Muzychka, M. R. Sridhar and M. M. Yovanovich, "Thermal Spreading Resistance in Multilayered Contacts: Applications in Thermal Contact Resistance," *J. of Therm. and Heat. Trans.,* vol. 13, no. 4, 1999.

[24] R. B. Dupaix, Becoming a Finite-Element Analyst: A Design-Model-Verify Approach, Cognella, 2015.

[25] A. M. Darwish, A. J. Bayba and H. A. Hung, "Accurate Determination of Thermal Resistance of FETs," vol. 53, no. 1, 2005.

[26] M. Garven and J. P. Calame, "Simulation and Optimization of Gate Temperatures in GaN-on-SiC Monolithic Microwave Integrated Circuits," *IEEE Trans. on Comp. and Pack. Tech.,* vol. 32, no. 1, 2009.

[27] V. O. Turin and A. A. Balandin, "Electrothermal Simulation of the self-heating effects in GaN-based field-effect transistors," *J. of App. Phys.,* vol. 100, 2006.

[28] V. Sodan, S. Stoffels, H. Oprins, M. Baelmans, S. Decoutere and I. De Wolf, "A Modeling and Experimental method for Accurate Thermal Analysis of AlGaN/GaN Powerbars," in *27th Inter. Symp. on Pow. Semic. Dev. and ICs*, 2015.

[29] F. Bertoluzza, N. Delmonte and R. Menozzi, "Three-dimensional finite-element thermal simulation of GaN-based HEMTs," vol. 49, no. 5, 2009.

[30] E. Douglas, F. Ren and S. Pearton, "Finite-element simulatoins of the effect of device design on channel temperature for AlGaN/GaN high electron mobility transistors," vol. 29, no. 2, 2011.

[31] E. R. Heller and A. Crespo, "Electro-thermal modeling of multifinger AlGaN/GaN HEMT devcie operation including thermal substrate effects," vol. 48, 2008.

[32] A. Venkatachalam, W. T. James and S. Graham, "Electro-thermo-mechanical modeling of GaN-based HFETS and MOSHFETs," *Semicond. Sci. Techno.,* vol. 26, 2011.

[33] K. R. Bagnall, *Multiphysics characterization of GaN HEMTs via micro-Raman spectroscopy,* 2017.

[34] M. Azarifar and N. Donmezer, "A Roadmap for Building Thermal Models for AlGaN/GaN HEMTs: Simplifications and Beyond," in *ASM Heat Trans. Summer Conf.*, 2016.

[35] M. Kuball, S. Rajasingam, A. Sarua, M. J. Uren, T. Martin, B. Hughes, K. P. Hilton and R. S. Balmer, "Measurement of temperature distribution in multifinger AlGan/GaN heterostructure field-effect transistors using micro-Raman spectroscopy," *Appl. Phys. Lett.,* vol. 82, no. 124, 2003.

[36] A. M. Conway, P. M. Asbeck, J. S. Moon and M. Micovic, "Accurate thermal analysis of GaN HFETs," *Solid State Electronics,* vol. 52, no. 5, 2008.

[37] H. Ji, M. Kuball, A. Sarua, J. Das, W. Ruythooren, M. Germain and G. Borghs, "Three-Dimensional Thermal Analysis of a Flip-Chip Mounted AlGaN/GaN HFET Using Confocal Micro-Raman Spectroscopy," *IEEE Trans. on Elec. Dev.,* vol. 53, no. 10, 2006.

[38] S. Choi, G. M. Peake, G. A. Keeler, K. M. Geib, R. D. Briggs, T. E. Beechem, R. A. Shaffer, J. Clevenger, G. A. Patrizi, J. F. Klem, A. Tauke-Pedretti and C. D. Nordquist, "Thermal Design and Characterization of Heterogeneously Integrated InGaP/GaAs HBTs," *IEEE Trans. on Comp. Pack. and Manf. Tech.,* vol. 5, p. 6, 2016.

[39] P. McCluskey, "Thermal Isolation and Differential Cooling of Heterogeneously Integrated Devices," Air Force Research Laboratory, 2016.

[40] M. Fish, B. Martinis, P. McCluskey and A. Bar-Cohen, "Embedded Fluid Coolling of Close-Packed Via Arrays in Glass," in *Conf. on Therml and Thermomech. Phen. in Elec. Sys.*, 2018.

[41] V. R. Manikam and K. Y. Cheong, "Die Attach Materials for High Temperature Applications: A Review," *IEEE Trans. on Comp. Pack. and Manf. Tech.,* vol. 1, no. 4, 2011.

[42] Henkel Technologies, "QMI 529 HT Datasheet," April 2004. [Online]. Available: http://www.icproto.com/pdf/H-QMI529HT-EN.PDF. [Accessed 1 November 2018].

[43] Lord Corporation, "CoolTherm MT-322 Thermally Conductive Adhesive Datasheet," January 2018. [Online]. Available: https://lordfulfillment.com/pdf/44/DS3942_CoolThermMT-322.pdf. [Accessed 14 January 2019].

[44] Epoxy Technology, "EPO-TEK H20E Datasheet," June 2018. [Online]. Available: http://www.epotek.com/site/administrator/components/com_products/assets/files/Style_Uploads/H20E.pdf. [Accessed 14 January 2019].

[45] K. Kurabayashi and K. Goodson, "Precision measurement and mapping of die-attach thermal resistance," vol. 21, no. 3, 1998.

[46] A. S. Fleischer, L.-h. Cheng and B. C. Johnson, "The effect of die attach voiding on the thermal resistance of chip level packages," *Microelectronics Reliability,* vol. 46, 2006.

[47] P. Teertstra, "Thermal Conductivity and Contact Resistance Measurements for Adhesives," Vancouver, Canada, 2007.

[48] M. LaRue, *A Fully-Integrated Four-way Outphasing Architecture in Heterogeneously Integrated CMOS/GaN Process Technologies,* 2018.

[49] R. Quay, Gallium Nitride Electronics, Berlin, Germany: Springer, 2008.

[50] D. F. Brown, S. Keller, F. Wu, J. S. Speck, S. P. DenBaars and U. K. Mishra, "Growth and characterization of N-polar GaN films on SIC by metal organic chemical vapor deposition," vol. 104, 2008.

[51] A. Sarua, H. Ji, K. P. Hilton, D. J. Wallis, M. J. Uren, T. Martin and M. Kuball, "Thermal Boundary Resistance Between GaN and Substrate in AlGaN/GaN Electronic Devices," vol. 54, no. 12, 2007.

[52] A. Manoi, J. Pomeroy, N. Killat and M. Kuball, "Benchmarking of Thermal Boundary Resistance in AlGaN/GaN HEMTs on SiC Substrates: Implications of the Nucleation Layer Microstructure," vol. 31, no. 12, 2010.

[53] M. von Arx, O. Paul and H. Baltes, "Process-dependent thin-film conductivities for thermal CMOS MEMS," vol. 9, no. 1, 2000.

[54] N. Sojanovic, J. Yun, E. Washington, J. Berg, M. Holtz and H. Temkin, "Thermal Conductivity Measurement Using Microelectrothermal Test Structures and Finite-Element Model-Based Data Analysis," vol. 16, no. 5, 2007.

[55] A. Gutierrez-Aitken, B. Y.-C. Wu, D. Scott, K. Sato, B. Poust, M. Watanabe and e. al, "A Meeting of Materials," no. March/April, 2017.

[56] Wakefield-Vette, "Heatsink Datasheet 658-25AB," [Online]. Available: http://www.wakefield-vette.com/Portals/0/ resources/datasheets/658.pdf. [Accessed 1 October 2018].

[57] M. Smith, ABAQUS/Standard User's Manual, Version 6.14, Providence, RI: Simulia, 2014.

[58] Global Foundries, *CMOS 12S0 Technology Design Manual,* vol. October 2016, Global Foundries.

[59] D. Maier, M. Alomari, N. Grandjean and e. al., "Testing the Temperature Limits of GaN-Based HEMT Devices," vol. 10, no. 4, 2010.

[60] O. Semenov, A. Vassighi and M. Sachdev, "Impact of Self-Heating Effect on Long-Term Reliability and Performance Degredation in CMOS Circuits," vol. 6, no. 1, 2006.

[61] M. Beyfuss, J. Baumann and R. Tilgner, "Photothermal Imaging of Local Thermal Resistances," vol. 62, 1990.

[62] G. Pfannschmidt, "Ultrasonic Microscope Investigations of Die Attach Quality and Correlations with Thermal Resistance," vol. 8, 1992.

[63] J. Cho, D. Francis, D. H. Altman, M. Asheghi and K. E. Goodson, "Phonon Conduction in GaN-diamond composite substrates," *J. Appl. Phys.,* vol. 121, 2017.

[64] X. Qian, P. Jiang and R. Yang, "Anisotropic thermal conductivity of 4H and 6H silicon carbide measured using time-domain thermoreflectance," *Materials Today Physics,* vol. 3, 2017.

[65] D. I. Florescu, V. M. Asnin, F. H. Pollak, A. M. Jones, J. C. Ramer, M. J. Schurman and I. Ferguson, "Thermal conductivity of fully and partially coalesced lateral epitaxial overgrown GaN/saphhire (0001) by scanning thermal microscopy," *Appl. Phys. Lett.,* vol. 77, 2000.

[66] Cree, Inc., "Cree Silicon Carbide Substrates and Epitaxy," 2013. [Online]. Available: https://www.macrogroup.ru/sites/default/files/uploads/catalog/silovaya_electronika/CREE/matcatalog.pdf. [Accessed 1 July 2017].

[67] S. I. Abu-Eishah, "Correlations for the Thermal Conductivity of Metals as a Function of Temperature," *International J. of Thermophys.,* vol. 22, no. 6, 2001.

[68] A. A. Bajwa and e. al, "Heterogeneous Integration at Fine Pitch using Thermal Compression Bonding," 2017.

[69] E. Beyne, "The 3D Interconnect Technology Landscape," vol. 33, no. 3, 2016.

[70] B. Poust, "DAHICATS Thermal Tool User Documentation," 2015.

[71] E. J. Wyers, T. R. Harris, W. S. Pitts, J. E. Massad and P. D. Franzon, "Characterization of the mechanical stress impact on device electrical performance in the CMOS and III–V HEMT/HBT heterogeneous integration environment," in *Inter. 3D Sys. Int. Conf.* , 2015.