

Multi-Chip Power Module Fast Thermal Modeling for Layout Optimization

Brett W. Shook¹, Zihao Gong², Yongfeng Feng³, A. Matthew Francis⁴, and H. Alan Mantooth⁵

¹University of Arkansas, bx003@uark.edu

²University of Arkansas, gong@uark.edu

³University of Arkansas, yfeng@uark.edu

⁴University of Arkansas, amfranci@uark.edu

⁵University of Arkansas, mantooth@uark.edu

ABSTRACT

This paper proposes a method for fast thermal modeling of Multi-Chip Power Modules (MCPMs) for use in layout optimization. MCPMs integrate multiple high power semiconductor devices into a single package and allow for increased power densities and efficiencies critical for next generation alternative energy systems. In a typical MCPM layout, the design is modified iteratively to search for best electrical, thermal, and electromagnetic performance. In order to speed up design optimization, a novel thermal model is developed which can approximate thermal performance of a module under modifications of metal trace layout and power device positioning and quantity. Particle Swarm Optimization is chosen to perform the layout optimization process and is briefly compared against other optimization techniques.

Keywords: Multi-Chip Power Modules, Thermal Modeling, Layout Optimization

1 INTRODUCTION

As alternative energy systems grow in number, power electronics play an increasingly important role in efficiently conveying power to and from, the grid, motors, and other devices. Multi-Chip Power Modules (MCPMs) increase the efficiency and power density of these systems by integrating multiple high power semiconductor devices and control circuitry into one package [6]. A fundamental design objective in power electronics systems is the increase of switching frequency. At higher switching frequencies passive components such as inductors, transformers, and capacitors can be reduced in size leading to large gains in efficiency, size, and cost. Parasitic inductances and capacitances of power electronics modules present a major limiting factor of switching frequency thus a highly integrated and compact system reduces these parasites [8].

As MCPMs become more compact in order to increase switching frequency, it is crucial to improve heat dissipation capability. The reduction of maximum temperature in a module decreases material stress and fatigue increasing the lifetime and reliability of the internal semiconductor devices and increases efficiency [12], [14]. In a power module, die placement and copper/aluminum trace layout (Fig. 1) have a great impact on heat flux and temperature distribution. Thus, the layout of a power module is an important factor in its heat dissipation capability.

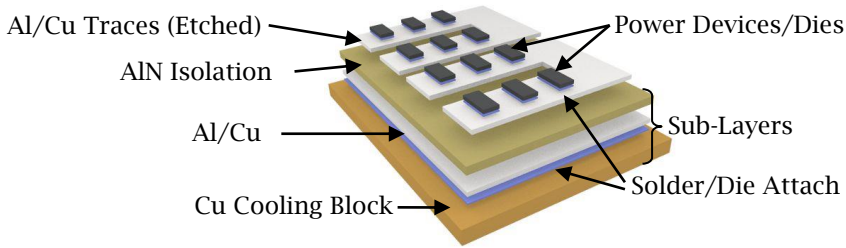


Fig. 1: Simplified structure of an MCPM: The portions of a module considered for layout are the die positions and the shape of the Al/Cu traces directly underneath the dies. The Al/Cu traces are shaped by etching away portions of the original metal.

In order to evaluate the thermal performance of a module quickly, it is modeled as a lumped element heat transfer system with thermal resistances and capacitances characterized via a thermal FEM tool [14]. This is where most thermal modeling methods stop, a characterization for a single design. The novel thermal modeling method developed in this paper provides thermal resistance values based on the layout of the module and number of dies. The key idea to the modeling method is the use of spatial superposition of temperature and heat flux distributions to predict changes in thermal spreading resistance [9]. The initial distributions are acquired through the ANSYS FEM thermal analysis tool [1]. The numerical solution of partial differential equations describing heat transfer is computationally expensive and is reduced to only an initial analysis. A fast thermal model also opens up opportunities for multi-objective optimization when considering electrical parasites and electromagnetic performance.

The closest related work targeting this problem solves the heat equation analytically via separation of variables for a specific type of system. The solution is derived for a system with two material layers which have a thermal spreading resistance between them [9]. The power module systems which are considered in our paper include two levels of spread resistance and multiple material layers, thus the analytical solution proposed in [9] does not apply. The use of spatial superposition of temperature distributions is presented in [9] when solving systems with multiple dies which lays the theoretical basis for the modeling algorithm in this paper. An initial step in developing an optimization methodology for MCPMs is presented in [4], but focuses on the optimization of a specific MCPM design. The thermal model developed in [4] uses a separate lumped element network for each die in the module, so thermal interaction between dies is completely neglected.

Particle Swarm Optimization (PSO) is adopted to perform layout optimization based on performance evaluations from the thermal model. A brief survey of optimization methods is conducted to find which best suits this particular problem.

2 THERMAL MODELING

This section covers the lumped element heat transfer topology, the thermal characterization process, the fast thermal modeling algorithm, and verification of the model. The primary objective of the thermal modeling algorithm is to predict the average temperature of the top and bottom surface of each die. This thermal modeling approach assumes that all of the die in a module are identical, but could easily be extended to work for systems with non-identical dies and devices.

2.1 Lumped Element Heat Transfer Topology

In a power module, heat dissipation takes place in three ways: conduction, convection, and radiation [15]. The primary heat dissipation path in a power module starts from the semiconductor dies (heat sources) down to the copper cooling block via conduction (Fig 2(a)). Next, the cooling block dissipates heat into the ambient environment via convection. Radiation makes up a small contribution to the system and is neglected in the proposed model.

A lumped element heat transfer system models heat flow by analogy to an electrical circuit. Heat sources are represented by current sources, temperature by voltage, and material by impedances. The thermal resistance of a material or composite layer can be calculated by

$$R_{thermal} = \frac{T_i - T_{i+1}}{P} \quad (2.1)$$

where T_i and T_{i+1} are the average temperatures of the start and end of the layer and P is the heat flow through the layer [13], [14]. Each die in a module is modeled to be a constant heat source, based on power losses from the die, with a material resistance and spread resistance in series as can be seen in Fig. 2(b) in the red box. Extra branches can be added depending on how many dies are in the system. Thermal spread resistance exists between two layers with different conduction areas, e.g. between die and trace or trace and isolation [15]. Spread resistance between layers is also found using Eqn. (2.1).

The thermal resistance for each material layer remains constant over layout changes, except the trace layer which small enough to be neglected due to low thermal resistance. However, the spread resistances $R_{sp,n}$ change significantly with respect to die positions and metal trace layout while $R_{sp,trace}$ changes with trace only (Fig. 2(b)). Thus, a thermal modeling technique is proposed to approximate these spread resistances based on layout. The thermal model only approximates steady-state spread resistance; transient behavior would require more advanced techniques [8], [13].

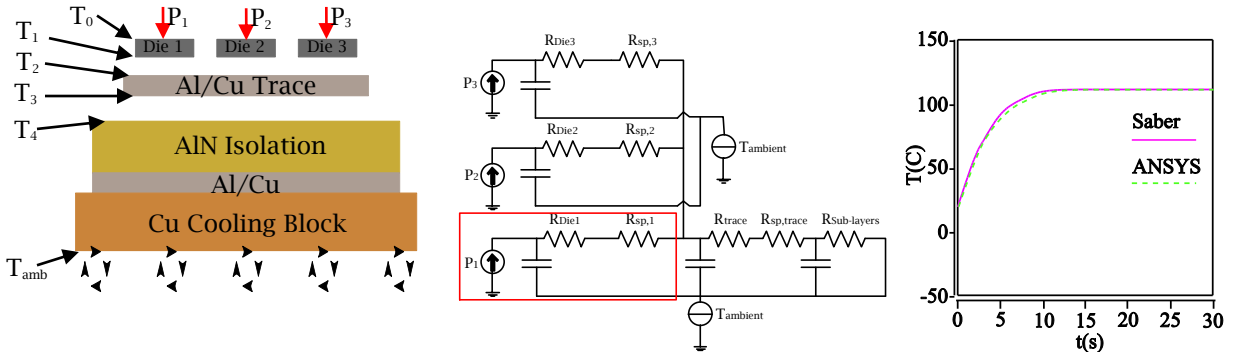


Fig. 2: Heat Transfer Topology: (a) Cross-section of an MCPM, (b) Heat Transfer Topology for 3 Die System, (c) Average Temperature Comparison of Top Surface of Trace Layer (FEM vs. Lumped Model)

In order to verify this heat transfer topology, FEM thermal analysis was performed in ANSYS. The thermal network was extracted from FEM data and constructed in the multi-disciplinary Saber circuit simulator [11]. All dies are turned on simultaneously, this allows the spread resistance of each die to remain constant; otherwise more advanced multi-port modeling techniques would be required [10]. The steady-state and transient data of temperature in each layer from Saber matched the FEM data from ANSYS with high accuracy confirming the topology (Fig. 2(c)).

2.2 Temperature and Heat Flux Distribution Characterization

The goal of the characterization process is to find temperature and heat flux distributions for a single die in an FEM tool and transform this data into rectangular contours (Fig. 3(b) and 3(c)). The thermal modeling algorithm uses this temperature and heat flux distribution data to predict temperatures and spread resistances in multiple die systems. The rectangular contour format allows for compact storage of the distribution data and fast intersection calculations which help speed up the modeling algorithm.

In order to obtain the distributions, a model of a power module is constructed in ANSYS. Rather than creating a complete layout design only a single die is placed in the center of the module and given a constant heat source at its top surface. The value of this heat source should be in the range of the average operating conditions of the die. The trace layer is constructed in a non-etched state, meaning that it has the same planar dimensions as the isolation layer below it. Temperature and heat flux (magnitude) distributions are requested for the top of the isolation layer in a regular grid format. The top of the isolation layer is selected on the basis that the temperature distribution at this layer can be directly mapped up to the top of the trace layer and die bottom. This is due to the high thermal conductivity of the metal trace layer and thus small vertical temperature difference.

The contours for both heat flux and temperature are constructed by the same process. The first step is to take two slices of data through the center of the module along the x and y axis. A set of uniformly spaced points are found along the upper half of the shorter axis of the module and mapped to the longer axis by temperature or flux value (Fig. 3(a)). A corresponding X and Y pair of points is ensured by mapping points on the shorter axis to the longer axis. In the case shown in Fig. 3, the longer axis of the module happens to be the along the x and the shorter along the y, thus a mapping from y to x.

Originally, a derivative based spacing of points was tried which lead to a higher density of points along the higher sloped regions of the slice curves. In practice this lead to poorer optimization results because the finer grained data was focused on the higher temperature points of the distribution instead of the better trade-off regions of the temperature curve. A simple uniform spacing of points along the coordinate axes was found to place a good density of points near the trade-off regions.

Each X-Y pair of points is used to create a set of rectangles R_n that are symmetric across the center of the module (Fig. 3(b)). A set of rectangular contours S_n is formed from R_n by removing the smaller rectangle above from the larger rectangle below:

$$S_n = R_n \ominus R_{n-1} \quad (2.2)$$

where \ominus is the symmetric difference. The rectangular contour regions are basically a set of rectangular rings except for the topmost region S_0 which is a rectangle.

The height or value of each contour is defined by the average value of the distribution inside the contour region which is found by numerical integration and not by the mapping value. Otherwise, the average value of the contour representation of the distribution would be too low. The result of this process is visualized in 3D form in Fig. 3(c). In order to place multiple temperature distributions in superposition, the ambient temperature needs to be subtracted out from the distribution:

$$q(x,y,z = Z_{isolation}) = T(x,y,z = Z_{isolation}) - T_{amb} \quad (2.3)$$

The symbol q represents a temperature that is referenced to the ambient temperature.

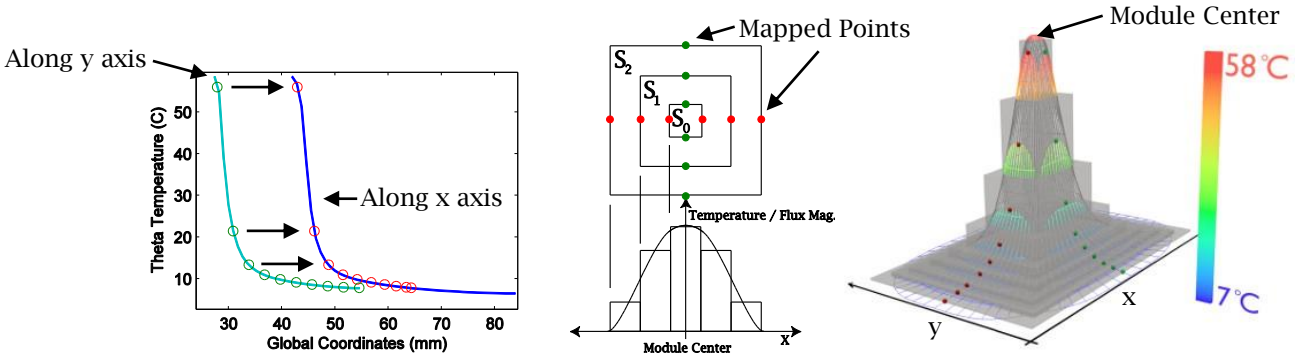


Fig. 3: Temperature and Flux Distribution Characterization: (a) Mapping Points from Y to X, (b) Set of Rectangular Contours from Distribution Data, (c) 3D Visualization of Rectangular Contours Vs. Scalar Temperature Distribution

2.3 Trace Scaling Characterization

The average temperature of the metal trace and die increases as the trace shrinks under the die (less dissipation area). This temperature behavior is hard to predict given only the temperature and flux distribution information for maximized trace. The trace scaling characterization process records the average temperature of the metal trace and bottom of a die as the trace area is decreased. This dataset helps keep the thermal modeling algorithm on track as trace area decreases significantly from the original characterization conditions.

2.4 Thermal Modeling Algorithm

For each die, the die-to-trace spread resistance $R_{sp,n}$ changes significantly with respect to placement and trace layout. $R_{sp,n}$ is the sum of two components

$$R_{sp,n} = R_{C,n} + R_{E,n} \quad (2.4)$$

which we have coined the edge effect resistance $R_{E,n}$ and the thermal coupling resistance $R_{C,n}$.

The calculation of $R_{C,n}$ involves the superposition of neighboring dies' temperature distributions. This is achieved by translating (positioning) the characterized set of rectangular temperature contours to the location of each contributing die. Fig. 4(a) gives an illustration of a die (orange) affected by a single neighboring die. Next, the intersection of these contours with the footprint region of the die is used to find the overall average contribution of temperature to the die being evaluated as follows:

$$q_{die,n} = q_{die_self} + \frac{1}{A_{die}} \sum_{i=0}^{C-1} q_{contrib,i} * A_i \quad (2.5)$$

A_{die} is the area of the footprint of die,n , A_i is the area of the intersection, $q_{contrib,i}$ is the average temperature value of the intersecting contour, and C is the total number of intersecting contours. q_{die_self} is the average temperature of the die by itself and is found by linear interpolation of the trace scaling data based on metal trace area. $R_{C,n}$ is finally calculated using the equation for thermal resistance,

$$R_{C,n} = \frac{q_{die,n} - q_{trace,all}}{P_0} \quad (2.6)$$

where P_0 is the heat flow from die,n and $q_{trace,all} = q_{trace} * N$. q_{trace} is the average temperature of the trace which is found by interpolation of the trace scaling data based on the metal trace area. N is the number of dies in the system. Basically, $q_{trace,all}$ is the average temperature of the trace when all other dies are present and adding their temperature to the trace, thus multiplication by N .

The edge effect spread resistance $R_{E,n}$ is approximated by how much heat conduction ability is lost due to being close to the edge of the trace. As a die moves from the center of a piece of trace and closer to the edge, the die loses cross-sectional area for heat energy to flow through. The heat flux distribution characterized earlier allows for a gauging of heat conduction value around a die. For each die, rectangular heat flux contours are translated to the position of the die and checked for intersection with a set of trace rectangles. This set of trace rectangles describes the layout of trace in a module. The effective heat flow P_E is computed by

$$P_E = \sum_{i=0}^{F-1} f_i * U_i \quad (2.7)$$

where f_i is the flux value of the contour and U_i is the intersection area. Fig. 4(b) shows a die with its set of heat flux contours intersected with the piece of trace it resides on. The edge effect resistance is then calculated by

$$R_{E,n} = \frac{Dq_{DT}}{P_E} - \frac{Dq_{DT}}{P_0} = \frac{Dq_{DT}(P_0 - P_E)}{P_0 * P_E} \quad (2.8)$$

where $Dq_{DT} = q_{die_self} - q_{trace}$ from the earlier trace scaling interpolations. $R_{E,n}$ is formed by realizing the increase in spread resistance is due to the decrease in effective heat flow P_E , thus taking the

difference from the original heat flow yields Eqn. (2.8). Notice that if the die is far from an edge, $P_E = P_0$, so $R_{E,n} = 0$ (no edge effect).

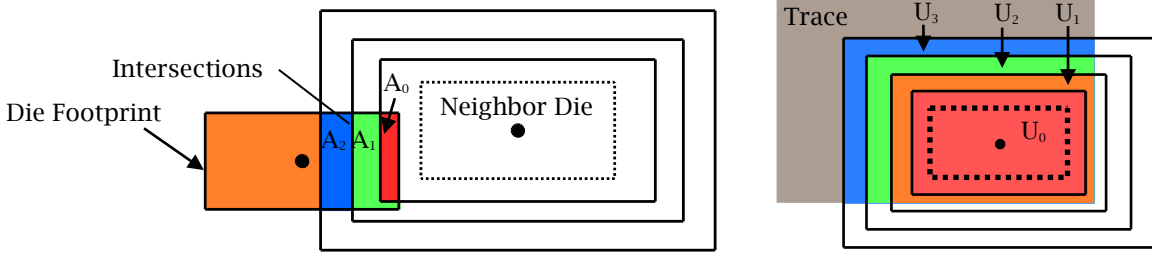


Fig. 4: Thermal Modeling Illustrations: (a) Thermal Coupling Intersections, (b) Edge Effect Intersections

The trace to isolation spread resistance $R_{sp,trace}$ is computed by $\frac{q_{trace} - q_{isolation}}{P_0}$. $q_{isolation}$ is found during the characterization process and is simply the average value of the isolation layer at its top surface for a single die. $q_{isolation}$ does not change significantly with dramatic changes in die position and trace shape and size, so a single average temperature value recorded from the characterization is sufficient. q_{trace} is found via interpolation of the trace scaling data as mentioned earlier.

After calculating the thermal resistances described above, the values are placed in their correct positions in an impedance matrix derived from the topology of the heat transfer circuit. The matrix and a vector of heat flows from the dies are used to solve for average die temperatures using standard matrix solution techniques.

2.5 Thermal Model Verification

The accuracy of the model is verified by placing dies and trace in a variety of situations and comparing temperature and spread resistance to those obtained from ANSYS and the model. The experiment shown in Fig. 5(a) tests both coupling and edge effects simultaneously by keeping Die 1 stationary in the center of the trace and moving Die 2 farther away from the edge of the trace and closer to Die 1. The edge effect dominates on the left side of Fig. 5(b) and thermal coupling on the right. Fig. 5(c) and (d) show an experiment where Dies (0-5) are added to the system. The amount of error does not increase as more dies are added.

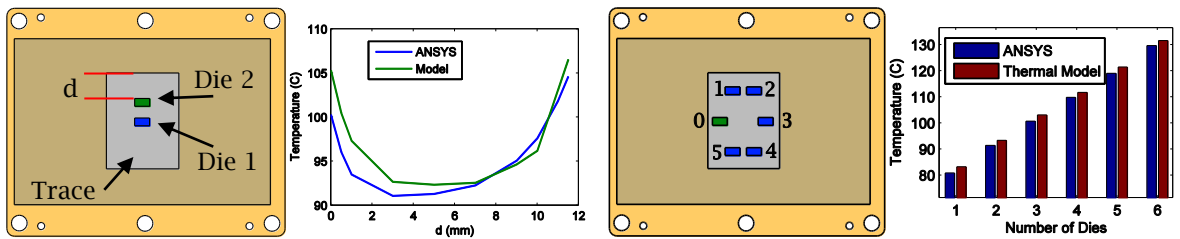


Fig. 5: Model Verification: (a) Edge and Coupling Effect Experiment, (b) Model vs. ANSYS Temperature of Die 2, (c) Die Quantity Experiment Setting, (d) Average Top Surface Temperature of Die 0 from ANSYS and Model with Increasing Die Quantity

The fast thermal model, implemented in Python, was found to run for $116 \mu s$ for 1 die and $1794 \mu s$ for 6 die. The FEM model ran for 13.13 s for 1 die and 18.00 s for 6 die. Both the FEM and thermal model were simulated on an Intel Core i7-870 clocked at 2.93 GHz per core. The asymptotic computational complexity of the thermal modeling algorithm is $O(n^2)$ with n dies. This is due to the evaluation of each die with every other die when computing thermal coupling. The maximum error of the fast thermal model compared to the ANSYS FEM results in Fig. 5(b) and 5(d) are 6.0% and 2.9% respectively.

3 LAYOUT OPTIMIZATION

3.1 Survey of Optimization Algorithms

An evolutionary algorithm optimization approach is selected on the basis that convex optimization and linear programming approaches require fitness evaluation and constraint functions that must conform to a specific mathematical form which our algorithm does not [2].

A statistical comparison of genetic algorithms (GA) and PSO across a broad set of optimization problems is presented in [5], concluding with the claim that PSO has the same ability to find accurate globally optimal solutions but with less computational cost. A comparison between simulated annealing (SA) and PSO is made in [3] and also concludes that PSO requires a smaller number of performance evaluations to find an optimal solution, but is conducted for only one type of problem. Based on the results of these two references, PSO was chosen to perform the thermal layout optimization. A detailed description of the PSO algorithm can be found in [7].

3.2 Optimization Process and Results

The optimization experiment shown below in Fig. 6 consists of three die and a three fingered piece of trace. This experiment tests the thermal model and optimization algorithm by starting the three die in a bad initial configuration (Fig. 6(a)) and verifying whether an optimal configuration is reached (Fig. 6(b)). The precise optimal die placements were not known, but in order to gauge the effectiveness of the optimization process a human solution was conducted (Fig. 6(c)). This experiment optimizes the placement of dies based on minimizing the maximum average die temperature in the module with the goal of reducing peak temperatures.

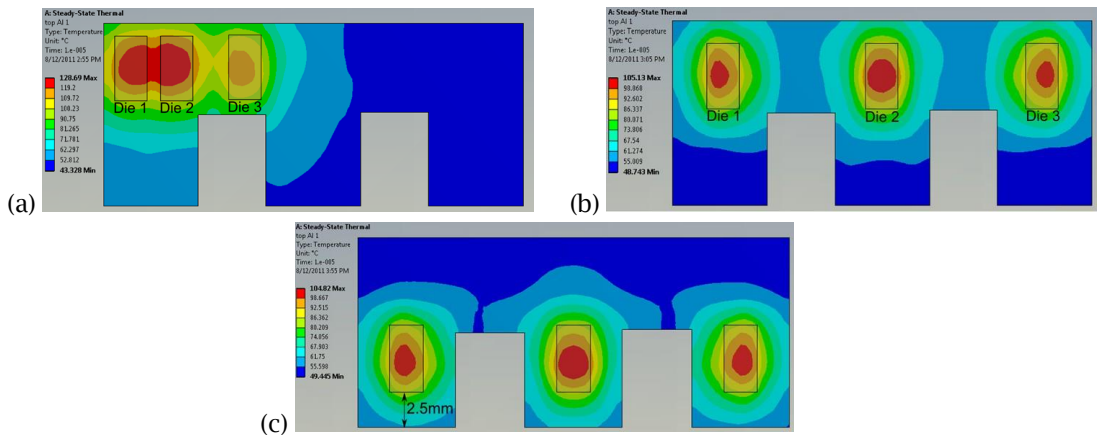


Fig: 6: Three Finger Optimization: (a) Bad Initial Placement, (b) Final Optimization Placement, (c) Human Placement

The final algorithmic placement has a slightly higher peak temperature (105.13 °C) than the human based design (104.82 °C), because the thermal model does not predict peaks in the temperature distribution but average temperature of each die. The thermal model could be modified to identify peak temperatures in regions of the module by considering the intersections of every temperature contour contributed by the dies. The algorithmic placement does achieve lower average maximum temperature for the dies as compared to the human design and significantly improves the design over the initial placement.

An MCPM from a commercial process was characterized using ANSYS and setup in a PSO based optimization experiment where dies were constrained to two fingers of trace (Fig. 7). The MCPM's maximum temperature was reduced by 10.8 °C under steady state conditions. The optimization considered 6 dies and took 120 s for 800 design iterations with a swarm population of 40. The dies in the optimized layout are spaced significantly farther apart and actually bow outward

slightly. A constraint could be placed on the dies to prevent this bowing, but provides for an interesting result.

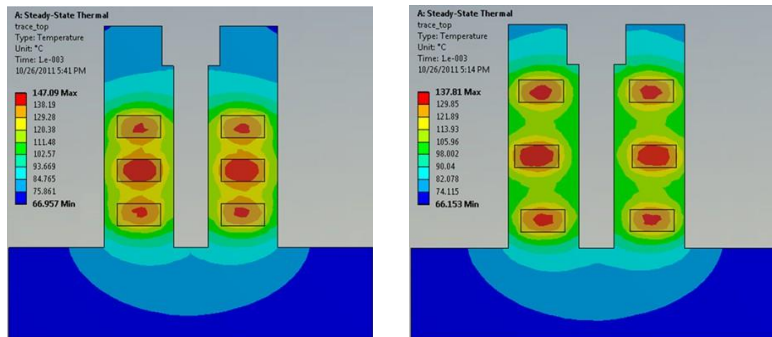


Fig. 7: Optimization Comparison: (a) Commercial Design, (b) Optimized Design

4 CONCLUSIONS

A fast thermal modeling algorithm has been presented which significantly speeds up thermal evaluation of a power module based on layout. This modeling technique allows for fast single or multiple objective MCPM layout optimizations allowing reduced design cost and improved reliability. The model has been verified to provide accurate thermal behavior under a variety of layout modifications allowing a designer to quickly experiment with and optimize layout designs. The model could also be extended for use in fast electro-thermal simulations where the electrical models of transistors or diodes are coupled with the thermal model of a module. Although, the characterization process would need to be extended to capture information about the temperature and flux distributions of dies at different heat flows in order to obtain accurate transient solutions.

The fast thermal model gains significant value when confronted with multi-objective trade-off scenarios. It can be difficult and time consuming for a human to predict and trade-off thermal and electrical characteristics.

5 ACKNOWLEDGEMENTS

This work is supported by the National Science Foundation Industry/University Cooperative Research Center on GRid-connected Advanced Power Electronic Systems (GRAPES) Project GR-10-11. Arkansas Power Electronics International (APEI) provides industrial insight to the project.

REFERENCES

- [1] ANSYS Workbench, ANSYS, <http://www.ansys.com/>, 2010.
- [2] Boyd, S.; Vandenberghe, L.: *Convex Optimization*, Cambridge University Press, Cambridge, New York, 2004.
- [3] Ethni, S.A.; Zahawi B.; Giaouris, D.; Acarnley, P.P.: Comparison of particle swarm and simulated annealing algorithms for induction motor fault identification, 7th IEEE International Conference on Industrial Informatics, 2009, 470–474. [doi:10.1109/INDIN.2009.5195849](https://doi.org/10.1109/INDIN.2009.5195849)
- [4] Hingora, N.; Liu X.; Feng, Y.; McPherson, B.; Mantooth, A.: Power-CAD: A novel methodology for design, analysis and optimization of Power Electronic Module layouts, Energy Conversion Congress and Exposition (ECCE), 2010, 2692–2699. [doi:10.1109/ECCE.2010.5618043](https://doi.org/10.1109/ECCE.2010.5618043)
- [5] Hassan, R.; Cohanin, B.; De Weck, O.; Venter, G.: A comparison of particle swarm optimization and the genetic algorithm, Proceedings of the 1st AIAA Multidisciplinary Design Optimization Specialist Conference. 2005.
- [6] Homberger, J.M.; Mounce, S.D.; Schupbach, R.M.; Lostetter, A.B.; Mantooth, H.A.: High-temperature silicon carbide (SiC) power switches in multichip power module (MCPM) applications, Industry Applications Conference, 1, 2005, 393- 398.

- [7] Kennedy, J.; Eberhart, R.: Particle swarm optimization, IEEE International Conference on Neural Networks, 4, 1995, 1942-1948. [doi:10.1109/ICNN.1995.488968](https://doi.org/10.1109/ICNN.1995.488968)
- [8] Lee, F.C.; van Wyk, J.D.; Boroyevich, D.; Lu, Guo-Quan; Liang, Zhenxian; Barbosa, P.: Technology trends toward a system-in-a-module in power electronics, IEEE Circuits and Systems Magazine, 2(4), 2002, 4- 22. [doi:10.1109/MCAS.2002.1173132](https://doi.org/10.1109/MCAS.2002.1173132)
- [9] Muzychka, Y.S.; Culham, J.R.; Yovanovich, M.M.: Thermal Spreading Resistance of Eccentric Heat Sources on Rectangular Flux Channels, J. Electron. Packag, 125(2), 2003,178. [doi:10.1115/1.1568125](https://doi.org/10.1115/1.1568125)
- [10] Rencz, M.; Szekely, V.: Dynamic thermal multiport modeling of IC packages, IEEE Transactions on Components and Packaging Technologies, 24(4), 2001, 596-604. [doi:10.1109/6144.974946](https://doi.org/10.1109/6144.974946)
- [11] Saber, Synopsys, <http://www.synopsys.com/>, 2003.
- [12] Srihari, T.; Bharath, R.J.: A Novel Component-Level Thermal Modeling of Power Electronic Devices, IjPEE, 01(01), 2010, 1-6.
- [13] Vassighi, A.; Sachdev, M.: Thermal and Power Management of Integrated Circuits, Springer, New York, 2006.
- [14] Wang, H.; Khambadkone, A.M.; Yu, X.: Dynamic electro-thermal modeling in Power Electronics Building Block (PEBB) applications, Energy Conversion Congress and Exposition (ECCE), 2010, 2993-3000. [doi:10.1109/ECCE.2010.5618360](https://doi.org/10.1109/ECCE.2010.5618360)
- [15] Wu, L.K.; Kao B.: A study of thermal spreading resistance, Innovative Technique Center, Cooler Master Co., Ltd 2004.