

Supervised Machine-Learning Approach for the Optimal Arrangement of Active Hotspots in 3-D Integrated Circuits

Srikanth Rangarajan¹⁰, Leila Choobineh¹⁰, and Bahgat Sammakia, *Fellow, IEEE*

Abstract—3-D integration is now considered a new paradigm for the semiconductor packaging industry to sustain Moore's law. Vertical stacking of semiconductor chips provides high power density in a given footprint area. However, owing to increased integration, 3-D ICs having multiple core areas (hotspots) on each stack layer can often be prone to thermal interaction between the stack layers (Interlayer) and within the stack layers (Intralayer). In this work, three layers with three core hotspot areas on each layer are considered. This article proposes an optimization methodology to optimally arrange the hotspot active core areas in three layers of 3-D IC. The optimization methodology aims to minimize the maximum core temperatures and maximize the temperature uniformity in the stack. The optimal placement of hotspot core areas in each layer not only aids in reducing the thermal interaction but also aids in improving the temperature uniformity by thermal spreading. A sampling algorithm based on Latin Hypercube Sampling that incorporates the "nonoverlap" constraint is demonstrated in this study. A Genetic algorithm coupled with supervised machine-learning-based artificial neural network is employed as an optimization methodology. The article introduces a unique arrangement parameter for the multiple numbers of hotspots in various layers that could well represent the problem under consideration. The methods and results from this article could be efficiently used to perform a thermal aware core hotspot arrangement of multilayer multihotspot 3-D integrated circuits for any operating conditions.

Index Terms—3-D ICs, active silicon, artificial neural network (ANN), genetic algorithm (GA), machine learning, optimization.

NOMENCLATURE

G	Number of generations.
N	Size of the initial population.
P_c	Probability of crossover.
P_m	Probability of mutation.
λ	Arrangement parameter, mm ²
CV	Coefficient of variation of local arrangement parameters.
SD	Standard deviation.

σ	Standard deviation of temperature (temperature uniformity metric), °C (or) K.
X	Centroid of heat source/substrate/arrangement in X-direction, mm.
Y	Centroid of heat source/substrate/arrangement in Y-direction, mm.
L	Length of the heat source/substrate, mm.
W	Width of the heat source/substrate, mm.
T	Temperature rise, °C (or) K.

Subscripts

a	Arrangement.
i	Layer/substrate.
j	Heat source.
s	Substrate.
c	Centroid.

Superscripts

*	Normalized parameters.
---	------------------------

I. INTRODUCTION

THE 3-D integration is a viable technology that has a proven potential to sustain Moore's Law by vertically integrating multiple layers of active electronic circuits into a single circuit [1]. 3-D ICs facilitate heterogeneous integration of various devices/functionalities in the vertical direction. 3-D integrated circuits offer multiple advantages, including increased device packaging density, shorter interconnect length, low system form factor [2]. Although there are many advantages to 3-D integration, one of the most significant challenges is heat removal owing to the interlayer and intralayer overlap of multiple cores and hotspots. Pangracious *et al.* [3] demonstrated that by replacing the 2-D system on chip with a 3-D system on chip, the Manhattan wire length is significantly reduced for the same total surface area. Geer *et al.* [4] developed a mathematical model to solve steady-state conduction heat transfer in multiple rectangular domains that account for the interactive effects of the heat-generating sources. This model has been employed to solve heat transfer between stacks in 3-D ICs. Choobineh and Jain [5] performed analytical calculations to estimate the 3-D temperature profile in multiple stacks of 3-D ICs. The study

Manuscript received August 14, 2021; accepted August 16, 2021. Date of publication September 1, 2021; date of current version October 28, 2021. This work was supported by the Research Foundation (RF), State University of New York (SUNY). Recommended for publication by Associate Editor S. H. Bhavnani upon evaluation of reviewers' comments. (Corresponding author: Srikanth Rangarajan.)

Srikanth Rangarajan and Bahgat Sammakia are with the Department of Mechanical Engineering, Binghamton University (SUNY), Binghamton, NY 13850 USA (e-mail: srangar@binghamton.edu).

Leila Choobineh is with the Department of Mechanical Engineering, SUNY Polytechnic Institute, Utica, NY 13502 USA.

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TCPMT.2021.3109662>.

Digital Object Identifier 10.1109/TCPMT.2021.3109662

2156-3950 © 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.

See <https://www.ieee.org/publications/rights/index.html> for more information.

was conducted for equal and unequally sized dies. Choobineh and Jain [6] also demonstrated a noniterative approach to analytically calculating the temperature field in 3-D ICs. Using this approach, the authors were able to predict the effect of inter die contact resistance. Jain *et al.* [7] performed a thermal, electrical cooptimization of floor planning of 3-D ICs subject to manufacturing constraints. The authors demonstrated an optimal thermal floorplan has a high penalty concerning the electrical delay. To date, only low-power commercial products have been able to exploit the advantages of the improved performance and increased device-packaging density realized by the 3-D stacking of chips (Apple A7) [8]. As the number of stacks increases and the number of chips/core processors on each layer increases, the passive cooling or conventional air-cooling will not meet the requirements. Although advanced cooling technologies would be a potential solution to mitigate the thermal management issues, this article proposes a thermal aware floor planning of core locations in each layer, intending to minimize the maximum temperature in the stack and to maximize the temperature uniformity in the stacked layers. The maximum temperature and temperature uniformity are critical to the reliability of the device. Kumar Hotta *et al.* [9] demonstrated a heuristic approach to optimize discrete heat sources on a substrate subject to mixed convection. The authors developed an empirical correlation for the maximum chip temperature as a function of a unique geometrical parameter. Madadi and Balaji [10] demonstrated optimization using the micro-genetic algorithm (GA) to optimize the location of multiple discrete heat sources in a ventilated cavity.

Kumar Hotta *et al.* [9] performed optimal floor planning of very large-scale integration (VLSI) using particle swarm optimization. The entire optimization problem was driven by nonoverlap constraints. Ni *et al.* [11] adopted a “Greedy” placement strategy in floor planning of 3-D ICs. The authors concluded that the heat sources with high input are recommended to be placed close to the convective surface.

Rangarajan *et al.* [12] performed optimization of hotspot location for two-layer 3-D ICs with two hotspots on each layer with hotspots of equal size. The optimization methodology, however, lacked a global optimal solution and unique arrangement parameters. The methodology demonstrated in this article would not apply to more number of stacks and more hotspots on each stack.

Wang *et al.* [13] performed an analysis of hotspot distribution and the effect of external cooling on the thermal performance of 3-D ICs. The authors examined the effect of hotspot-targeted through-silicon vias (TSVs) inclusion in 3-D stacks. The work concluded that the core centralized TSV could mitigate the thermal issues owing to the overlap of core hotspots between the layers. Ding *et al.* [14] proposed a novel thermal management scheme for 3-D ICs with multiple cores on each of the multiple layers. A novel interlayer microchannel design was proposed for the efficient thermal management of core hotspots’ temperature. Yazawa *et al.* [15] performed a theoretical investigation and optimization for hotspot-targeted cooling in stacked 3-D ICs. Hotspot size and power were $0.5 \text{ mm} \times 0.5 \text{ mm}$ and 650 W/cm^2 , respectively. The authors concluded that the thermal spreading from each layer of 3-D

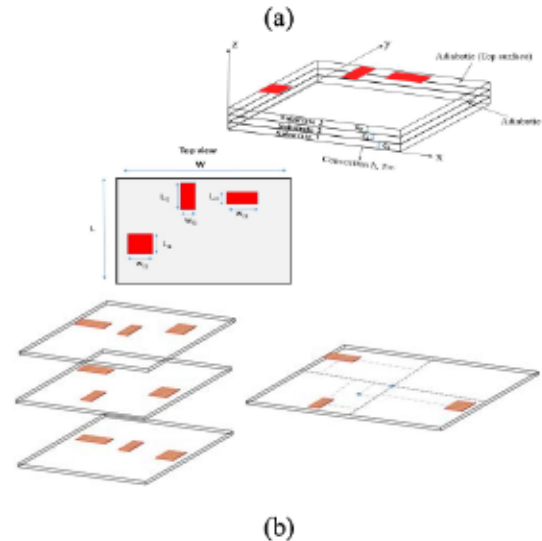


Fig. 1. (a) Schematic of the 3-D IC chip stack. (b) Schematic of the exploded view of 3-D IC chip stack.

ICs plays an important role in mitigating the effect of on-chip hotspots in 3-D ICs. The authors proposed cooling strategies that could maintain hotspot temperature below 85°C . The temperature uniformity in the 3-D stack was seen to influence the microchannel performance. Furthermore, there are numerous works from Shi *et al.* [16] and Bar-Cohen *et al.* [17] on the 3-D IC thermal management by mitigation of hotspot. Although there are numerous studies in the literature on hotspot mitigation using efficient microchannel cooling, few studies on the optimal arrangement of core hotspots could achieve better thermal performance in terms of maximum chip temperature and temperature nonuniformity.

Furthermore, due to high integration, when the number of stacks and the number of core hotspots on each stack increases, efficient implementation of the nonoverlapping constraint becomes critical. This work tries to define and implement the nonoverlapping constraint as a part of the initial sampling and perform an efficient and robust multiobjective optimization. Furthermore, there is an absence of empirical correlations for the maximum temperature as a function of heat source arrangement. This article tries to bridge all the gaps mentioned above in the literature and proposes a simplified approach for core hotspots arrangement of 3-D ICs.

II. HEAT TRANSFER PROBLEM DESCRIPTION

The schematic of the geometry investigated in the current article is as shown in Fig. 1(a). The dimension, aspect ratio of the chips, and the substrate are given in Table I. An effective and quick thermal model is mandatory for an efficient thermal-driven optimal floor planning algorithm to solve many physics-based models that will serve as an input to build the surrogate model for optimization [18]. The heat transfer problems under consideration (governing equation and boundary conditions) are solved analytically, as demonstrated by Choobineh and Jain [6].

The heat transfer study is carried for two different boundary heat transfer scenarios. Study 1 is conducted for constant heat

TABLE I
DIMENSIONS OF THE SUBSTRATES AND THE
HOTSPOT SHOWN IN FIG. 1(A)

S no	Dimension parameter	Value (mm)
1	W	10
2	L	10
3	C1	0.5
4	C2	0.5
5	C3	0.5

transfer coefficient on the convective boundary with different hotspot sizes of the same power. Study 2 is conducted on the convective surface's local heat transfer coefficient with varying hotspot sizes, power density, and power dissipation.

For Study 1, each hotspot generates 4 W heat that corresponds to a heat flux of 4 W/mm². The heater sizes chosen for the first part of the study (Study 1) are 1 mm × 1 mm ($w_{i1} \times L_{i1}$), 1.618 mm × 0.618 mm ($w_{i2} \times L_{i2}$), and 1 mm × 2 mm ($w_{i3} \times L_{i3}$). The aspect ratio of each of the heaters is 1, 2.618, and 0.38, respectively. The heat input received by each of these heaters was 4 W, corresponding to a heat flux of 4 W/mm². A background heat flux of 4 kW/m² was added. The background heat flux is seen to have a negligible or a very monotonic effect on the optimal solution. An interdie resistance of 0.75 K-mm²/W was modeled [19]. The convection heat transfer coefficient on the top wall is assigned to 5000 W/m²K. The size of the hotspots for Study 1 was chosen in such a way that for all three hotspots for a fixed heat input, the heat flux remains the same and only the aspect ratio of the hotspot varies. The methodology and the results from this article are independent of the chosen value of heat inputs and boundary conditions. This is demonstrated through two separate studies Study 1 and Study 2. The designer can choose this methodology for his operating conditions. The schematic of the heaters in the substrate is as shown in Fig. 1(a) and (b). The substrate is assigned thermal properties equivalent to pure silicon. The bottom surface of substrate 1 is exposed to convection and the top surface of substrate 3 and sidewalls are kept adiabatic. The temperature field across the stack is computed by assuming one active heat source [j th heat source $Q_j(x, y)$]. This approach simplifies the governing equation to linear form. The governing equation for the problem is

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} = 0 \quad (1)$$

$$\frac{\partial T}{\partial x} = 0 \text{ at } x = 0, L \quad (2)$$

$$\frac{\partial T}{\partial y} = 0 \text{ at } y = 0, W \quad (3)$$

$$k \frac{\partial T_3}{\partial z} = q_3(x, y) \text{ at } z = c_3 \quad (4)$$

$$k \frac{\partial T_2}{\partial z} \text{ (at } z = c_2) = k \frac{\partial T_3}{\partial z} \text{ (at } z = 0) + q_2(x, y) \quad (5)$$

$$k \frac{\partial T_1}{\partial z} \text{ (at } z = c_1) = k \frac{\partial T_2}{\partial z} \text{ (at } z = 0) + q_1(x, y) \quad (6)$$

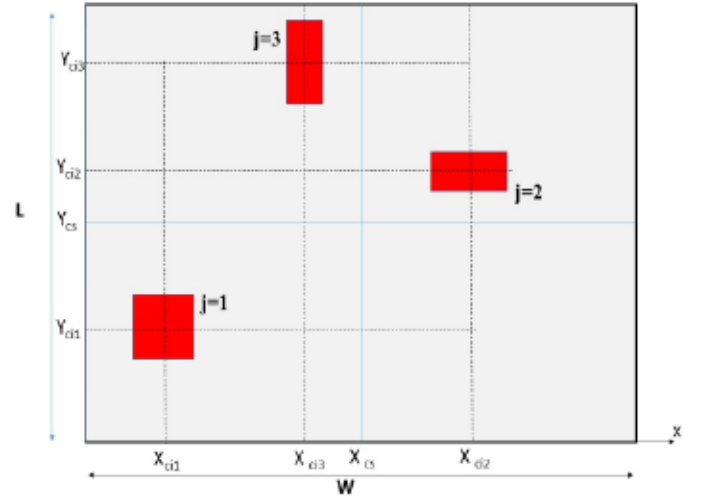


Fig. 2. Schematic showing the arrangement parameter calculation in i th layer.

$$\frac{\partial T_1}{\partial z} \text{ (at } z = 0) = \frac{h}{k} T_1 \text{ (at } z = 0). \quad (7)$$

The detailed procedure of the analytical model and validation is explained in [6]. The results from the analytical model were compared with results from the Ansys [6] and COMSOL multiphysics, and the maximum error was found to be less than 7.4% and 5.2%, respectively. For the second part of the study, COMSOL was employed because it was straightforward to interface the module with MATLAB 2017b.

III. ARRANGEMENT PARAMETER DEFINITION

The arrangement parameters are defined in this article to identify any interlayer and intralayer hotspot arrangement using a single parameter. For this purpose λ_1 and λ_2 are defined.

The centroid of the substrate is chosen as a reference since the heat transfer coefficient is uniform and constant on the convective boundary. λ_1 is defined as the distance between the centroid of the heat sources to the centroid of the substrate. λ_2 is defined as the distance between the centroid of the arrangement and the centroid of the substrate. Both λ_1 and λ_2 are defined globally for the entire stack and locally for each stack.

λ_{1g} and λ_{2g} are the global arrangement parameters. λ_{1l} and λ_{2l} are the local arrangement parameters.

- 1) g indicates global.
- 2) i indicates layer (within each layer).
- 3) j indicates a heat source.

The centroid of the heat sources, arrangement of heat sources, and the substrate are calculated as shown in (8)–(19). The calculation is also depicted in Fig. 2 for any i th layer

$$X_{cij} - \text{Centroid of a } j\text{th heater in } i\text{th layer} = \frac{\sum_{j=1}^{j=n_h} A_{ij} X_{ij}}{\sum_{j=1}^{j=n_h} A_{ij}} \quad (8)$$

$$Y_{cij} - \text{Centroid of a } j\text{th heater in } i\text{th layer} = \frac{\sum_{j=1}^{j=n_h} A_{ij} Y_{ij}}{\sum_{j=1}^{j=n_h} A_{ij}} \quad (9)$$

$$X_{CS} - \text{Centroid of the substrate} = \frac{W}{2} \quad (10)$$

$$Y_{CS} - \text{Centroid of the substrate} = \frac{L}{2} \quad (11)$$

$$X_{CAi} - \text{Centroid of the arrangement in } i\text{th layer} = \frac{\sum A_j x_{cij}}{\sum A_j} \quad (12)$$

$$Y_{CAi} - \text{Centroid of the arrangement in } i\text{th layer} = \frac{\sum A_j y_{cij}}{\sum A_j} \quad (13)$$

The initial studies revealed that λ 's are not enough to characterize a unique arrangement of hotspots in multiple stacked layers. For this purpose we define:

$CV_{\lambda 1}$ – Coefficient of variation λ_{1i} , where $i = 1, 2, \dots, n_l$;

$CV_{\lambda 2}$ – Coefficient of variation λ_{2i} , where $i = 1, 2, \dots, n_l$;

λ_{1i} – Distance between Centroid of Arrangement and Centroid of heat sources;

λ_{2i} – Distance between Centroid of arrangement and Centroid of substrate.

$$\lambda_{1i} = \sum_{j=1}^{nh} (X_{CAi} - X_{Cij})^2 + (Y_{CAi} - Y_{Cij})^2 \quad (14)$$

$$\lambda_{2i} = \sum_{j=1}^{nh} (X_{CAi} - X_{CS})^2 + (Y_{CAi} - Y_{CS})^2 \quad (15)$$

$$\lambda_{1g} = \sum_{i=1}^{nL} \sum_{j=1}^{nh} (X_{CAg} - X_{Cij})^2 + (Y_{CAg} - Y_{Cij})^2 \quad (16)$$

$$\lambda_{2g} = \sum_{i=1}^{nL} \left(\sum_{j=1}^{nh} (X_{CAg} - X_{CS})^2 + (Y_{CAg} - Y_{CS})^2 \right) \quad (17)$$

$$CV_{\lambda 1} = \frac{SD_{\lambda 1}}{\text{Mean}(\lambda_1)} = \frac{\sqrt{\frac{\sum_{i=1}^{n_l} \left(\lambda_{1i} - \frac{\sum_{i=1}^{n_l} \lambda_{1i}}{n_l} \right)^2}{n_l - 1}}}{\sum_{i=1}^{n_l} \frac{\lambda_{1i}}{n_l}} \quad (18)$$

$$CV_{\lambda 2} = \frac{SD_{\lambda 2}}{\text{Mean}(\lambda_2)} = \frac{\sqrt{\frac{\sum_{i=1}^{n_l} \left(\lambda_{2i} - \frac{\sum_{i=1}^{n_l} \lambda_{2i}}{n_l} \right)^2}{n_l - 1}}}{\sum_{i=1}^{n_l} \frac{\lambda_{2i}}{n_l}} \quad (19)$$

n_l is the number of layers and n_h is the number of hotspots in each layer.

The conditions at which the arrangement parameters attain maximum and minimum values are tabulated in Table II.

$(CV)_{\lambda_2}$ helps in identifying variation of local λ_2 . For instance, if the arrangement centroid of a particular layer i highly deviates from the other two layers, then the value of $(CV)_{\lambda_2}$ is very high, indicating a high localized grouping of heat sources in any one of the layers. This contributes to the overall thermal performance of the stack. This is the same case with $(CV)_{\lambda_1}$. High values indicate local grouping (intralayer grouping/overlap) of heat sources in a particular layer. To conclude, $(CV)_{\lambda_1}$, $(CV)_{\lambda_2}$ are defined to account for the tradeoff between the localized grouping of hotspots and the thermal heat spreading for a given boundary condition.

TABLE II
CONDITIONS FOR MAX AND MIN FOR LOCAL AND GLOBAL

Parameter	Condition
Minimum λ_{1l}	$X_{CAi} = X_{Cij}$ $Y_{CAi} = Y_{Cij}$
Maximum λ_{1l}	$X_{CAi} = X_{CS}$ and $(X_{CAi} - X_{Cij})^2 = \left(\frac{W}{2} - \frac{w_{ij}}{2}\right)^2 + \left(\frac{L}{2} - \frac{l_{ij}}{2}\right)^2$
Minimum λ_{2l}	$X_{CAi} = X_{CS}$
Minimum λ_{1g}	$X_{CAg} = X_{Cij}$ $Y_{CAg} = Y_{Cij}$
Maximum λ_{1g}	$X_{CAg} = X_{CS}$ and $(X_{CAg} - X_{Cij})^2 = \left(\frac{W}{2} - \frac{w_{ij}}{2}\right)^2 + \left(\frac{L}{2} - \frac{l_{ij}}{2}\right)^2$
Minimum λ_{2g}	$X_{CAg} = X_{CS}$

These parameters are included in the regression and optimization to account for the effect of the localized grouping of heaters, which will otherwise be not accounted in the equation alongside λ_{1g} and λ_{2g} . The maximum temperature of the stack/layers is only a function of global and local λ_1 , λ_2 $(CV)_{\lambda_1}$, $(CV)_{\lambda_2}$.

The primary objective of defining the arrangement parameters in this study is to arrive at a unique arrangement of heat sources for any given combination of the arrangement parameters. Furthermore, the optimization goal is to minimize the maximum stack temperature rise and maximize temperature uniformity. σ is defined as the temperature uniformity metric, which is calculated as the standard deviation of the temperature in the entire stack volume.

IV. OPTIMIZATION

The optimization problem, in general is driven mainly by constraints than the objective function.

The mathematical statement of the optimization problem is as shown from (20)–(31)

$$F1 : \min T_{\max} = f(\lambda_{1g}, \lambda_{2g}, CV_{\lambda 1}, CV_{\lambda 2}) \quad (20)$$

$$F2 : \min \sigma = f(\lambda_{1g}, \lambda_{2g}, CV_{\lambda 1}, CV_{\lambda 2}). \quad (21)$$

$$\text{s.t. } \lambda_{1g} \min < \lambda_{1g} < \lambda_{1g} \max \quad (22)$$

$$\lambda_{2g} \min < \lambda_{2g} < \lambda_{2g} \max \quad (23)$$

$$(CV_{\lambda 1})_{\min} < CV_{\lambda 1} < (CV_{\lambda 1})_{\max} \quad (24)$$

$$(CV_{\lambda 2})_{\min} < CV_{\lambda 2} < (CV_{\lambda 2})_{\max}. \quad (25)$$

The nonoverlap constraint is stated as

$$\text{s.t. } \lambda_{1i} \neq \lambda_{1i, \min} \quad (26)$$

$$\lambda_{1g} \neq \lambda_{1g, \min} \quad (27)$$

$$\lambda_{2i} \neq \lambda_{2(i+1), \min} \quad (28)$$

$$\lambda_{2i} \neq \lambda_{2g} \quad (29)$$

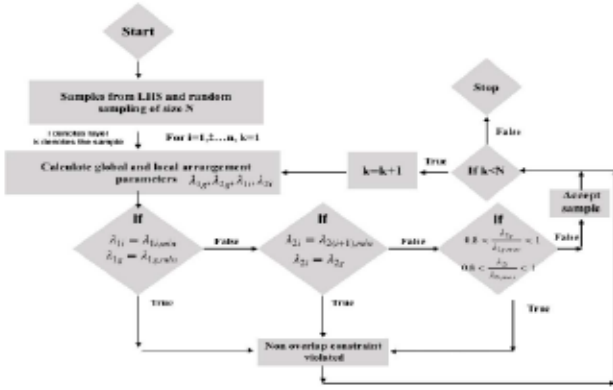


Fig. 3. Flowchart for the constraint incorporated sampling algorithm.

$$\frac{\lambda_{2g}}{\lambda_{2g,max}} < 0.8 \quad (30)$$

$$\frac{\lambda_{2l}}{\lambda_{2l,max}} < 0.8. \quad (31)$$

The purpose of choosing the value 0.8 was to make sure the centroid of arrangement is not too close to the edge/corner of the substrate. When $\lambda_{2g} = \lambda_{2g,max}$ the hotspots are crowded near the edge/corner of the substrate. To allow some clearance between the hotspots and the edge of the substrate a value lower than “1” was chosen (0.8). However, the designer can choose a much lower value based on his design constraint. The first and foremost step in optimization is to generate initial samples for the problem.

A. Learning-Based Sampling

To generate the samples for a different arrangement of hotspots on each layer, Latin hypercube sampling combined with Random sampling was employed with spatial constraints. The most important constraint of nonoverlap in each layer and between the layers was implemented by employing a self-learning sampling algorithm. The overlap in any layer was identified by calculating the centroid of the arrangement and the centroid of the heat sources. If the centroid of the heat sources coincided with the centroid of the arrangement in any particular layer, the nonoverlap constraint is declared violated. The probability of this is high at high values of λ_{2l} the probability of overlap of heat sources between the layers are seen to increase with simultaneously high and equal values of λ_{2g} . When such samples are encountered, the nonoverlap constraint is declared violated.

However, it is challenging to detect the overlap in between the heaters at low values of λ_{2g} . When i is equal and low for all the layers, the probability of nonoverlap still exists. Finally, the chosen samples that do not violate the nonoverlap constraints are employed for the regression analysis. The flowchart of the constraint incorporated sampling algorithm is as shown in Fig. 3. A total number 135 samples that do not violate the overlap constraints were generated.

The initial sampling process observed that the response parameters T_{max} and σ are nonconflicting, as shown in Fig. 4. This implies that optimizing any one of the objectives optimizes the other.

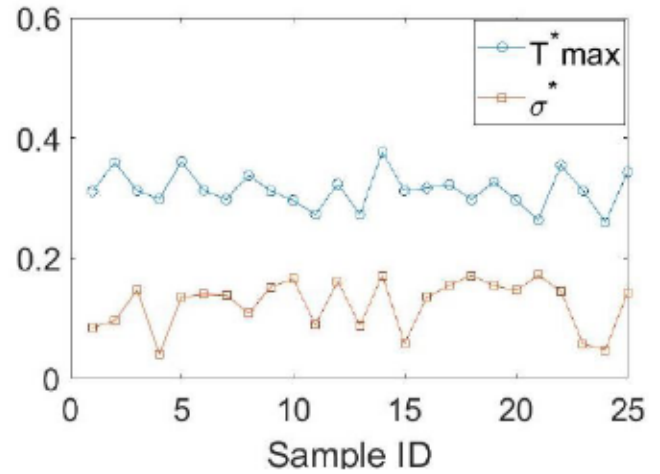


Fig. 4. Nonconflicting nature of the objective functions (nondimensional form).

TABLE III
RESPONSE PARAMETERS FOR RANDOMLY CHOSEN SAMPLES
FROM THE INITIAL SAMPLES

S No	λ_{1g}	λ_{2g}	$CV_{\lambda 1}$	$CV_{\lambda 2}$	$T_{max}(^{\circ}C)$	σ
1	48.52	0.009	0.04	1.41	97.05	4.21
2	48.48	0.58	0.04	0.23	101.75	5.42
3	27.48	0.58	0.06	0.23	106.78	7.02
4	12.48	0.58	0.08	0.23	118.80	8.36
5	12.42	4.61	0.08	0.09	130.94	14.17
6	12.38	12.48	0.08	0.05	148.61	21.16
7	1.43	24.19	0.24	0.04	139.09	20
8	124.71	0.15	0.41	1.6	96.3	2.55

B. Supervised Machine Learning Through Artificial Neural Network

$$T_{max} = f(\lambda_{1g}, \lambda_{2g}, CV_{\lambda 1}, CV_{\lambda 2}).$$

Table III presents the results of the response parameters for a small subset of the initial samples. To have a more accurate regression function as a representative surrogate model for optimization, artificial neural network (ANN), which is a non-linear regression tool, is developed [20] using MATLAB. One of the advantages of adopting an ANN is its efficiency to handle nonlinear trends in initial data. Supervised learning-based static ANN is employed in this study.

The ANN development procedure is as follows.

- 1) Dividing available data collected from the sampling into three groups namely, training, validation, and testing datasets. In total, 135 sets of data collected from the physics-based model provide input for the network: 89% of datasets are allocated to the training, and 11% of which are employed to evaluate the network. About 21 sets of data within the limits of the variables but excluded from training the network are considered to test the robustness of the network.

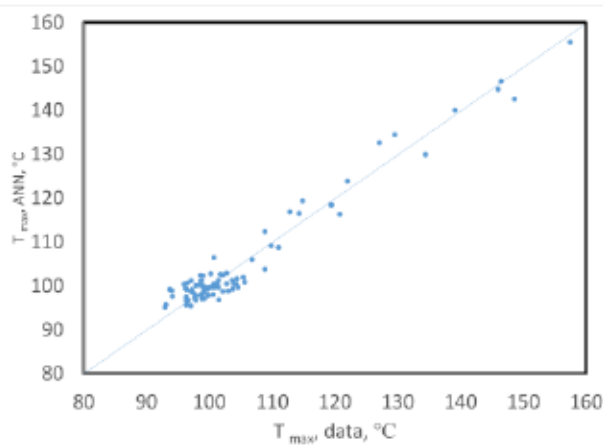


Fig. 5. Parity plot for the ANN-based regression.

- 2) Normalization of the data is done in the range of $[-1, 1]$ as the tan hyperbolic function has its limiting values in the range $[-1, 1]$.
- 3) Determining the optimal number of neurons using the neuron independent study [20]. In this study, the optimal number of neurons is calculated to be 4.

The results of the supervised learning algorithm are shown in Fig. 5. The network's prediction has a good agreement with the actual physics-based data for a different arrangement of the heat sources in the stacked layers. The Parity plot (shown in Fig. 5) obtained from the ANN makes it evident that the curve fitting is accurate for this nonlinear problem. This can be attributed to the regressive sensitivity analysis performed during the training process of ANN. The physics-informed machine-learning algorithm serves the purpose of simplifying a complex problem using an effective surrogate model. The R^2 and the maximum error from ANN-based regression are 0.99 and 4.3%, respectively.

C. Genetic Algorithm

GAs are a class of stochastic optimization algorithms inspired by the process of natural selection. GA is commonly employed for obtaining a global optimal solution for multi-variable multiconstraint optimization problems. The variables considered in GA are represented as binary strings (0s and 1s). The upper bound, lower bound, decimal accuracy required for each variable determine the length of the binary string of that variable, commonly termed as a chromosome. The flowchart of GA is as shown in Fig. 6. The process begins with the generation of the initial population of size N . The initial population generated should satisfy the nonoverlap constraint. The fitness value for each individual of the population is calculated using the objective function derived from regression analysis. GAs aims to improve the fitness value with each iteration/generation. The selection of individuals from the initial population is performed by employing roulette wheel selection. The selected population undergoes crossover and mutation to generate the offspring. The parameters governing the rate of crossover and mutation are the probability of crossover P_c and the probability of mutation P_m .

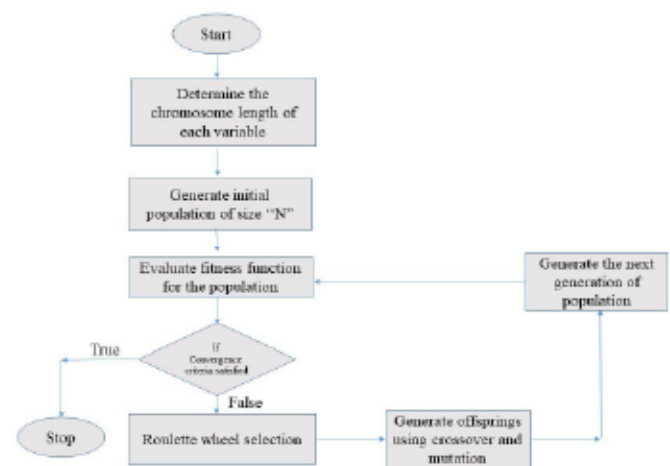
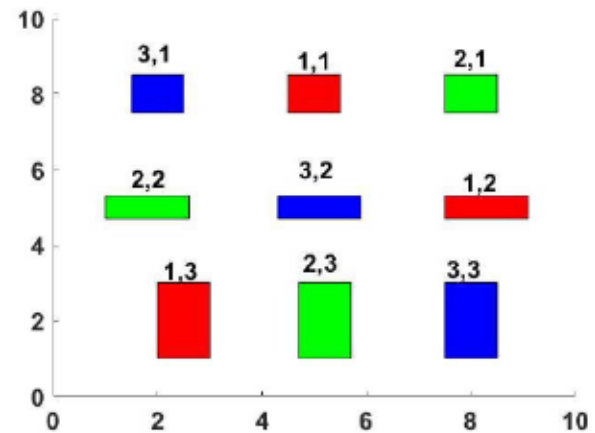


Fig. 6. Flowchart for the GA.

Fig. 7. Schematic of the optimal hotspot arrangement [the numbers above each heat source represent (layer number i , heat source number j)].

A sensitivity analysis is performed [20] to determine the optimal values of P_c and P_m . The fitness values of the new offspring population are calculated, and the process continues until the convergence criteria is satisfied. The results from the sensitivity analysis yielded the following values $P_c = 0.7$ and $P_m = 0.3$.

V. RESULTS AND DISCUSSION

A. Study 1

The optimization results carried out in the article were obtained using the GA coupled with the ANN. The optimal configuration that was obtained is as shown in Fig. 7. The corresponding optimal values of the arrangement parameter are presented in Table IV. It is evident from the optimization that a very low value of λ_{2g} is recommended. However, an optimal value of λ_{1g} is obtained for this particular configuration. Furthermore, the optimal values of $CV_{\lambda 1}$ and $CV_{\lambda 2}$ are near to 0, indicating that a uniform nonoverlapping arrangement of heat sources between the layers is recommended. It is also quite intuitive that the optimal arrangement parameters should facilitate efficient heat spreading [15], thereby reducing the maximum temperature from the heat transfer physics. Furthermore, the value of σ for the optimal configuration

TABLE IV
OPTIMAL VALUES OF THE VARIABLES AND RESPONSE
PARAMETERS IN THIS STUDY 1

Variable	Optimal value
λ_{1g}	108.52
λ_{2g}	0.01
CV_{λ_1}	0.03
CV_{λ_2}	0.51
Response parameter	Optimal Value
T max	89.91
σ	1.75

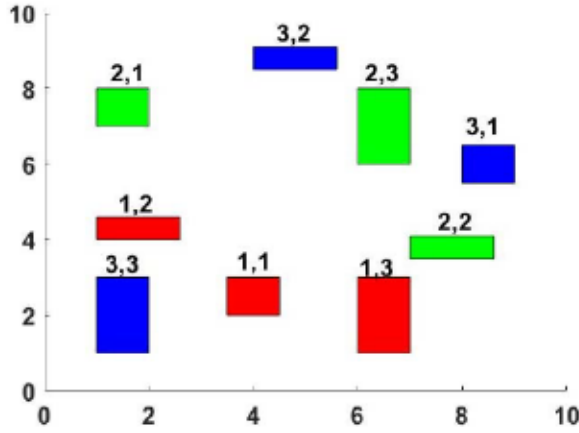


Fig. 8. Schematic of one of the hotspot arrangements with optimal $\lambda_{1g} = 110$, $\lambda_{2g} = 0.15$ and different CV_{λ_1} and CV_{λ_2} [the numbers above each heat source represent (layer number i , heat source number j)].

is 1.75, indicating the best temperature uniformity in the stack volume.

From Table IV, it is evident that the optimal arrangement parameters correspond to a very low value of λ_{2g} and CV_{λ_1} .

From the optimal value of λ_{2g} , which is near to 0, it is understood that the centroid of the arrangement should coincide or have a little offset from the centroid of the substrate. λ_{2g} could be the pivotal point to derive the hotspot arrangement layout from the arrangement parameters. The optimal value of other arrangement parameters could be derived based on a similar approach.

To understand why the arrangement parameters considered could lead to a unique arrangement of heat sources, we considered design guidelines from the optimal values of arrangement parameter pertaining only to $\lambda_{1g} = 110$ and $\lambda_{2g} = 0.15$ (very close to optimal values). The corresponding arrangement is as shown in Fig. 8. However, the response parameters for this case are $T_{max} = 91.7^\circ\text{C}$ and $\sigma = 2.88$, which are highly deviated from the optimal response parameters. The results from this arrangement reinforce the importance of CV_{λ_1} and CV_{λ_2} . The values of CV_{λ_1} and CV_{λ_2} are 0.55 and 1.18, respectively, indicating a local grouping of the hotspots as seen in Fig. 8. The results reiterate that only a unique arrangement could be derived from the four arrangement parameters considered in this study. These arrangement parameters become more critical when the number of stacks and hotspots on each stack increases a forecast trend in the packaging industry [21].

TABLE V
HOTSPOT DIMENSIONS AND CORRESPONDING HOTSPOT
POWER EMPLOYED IN THIS STUDY

Hotspot	Size (mm ²)	Heat flux (W/mm ²)	Power(W)
H1	1 x 1	8	8
H2	2 x 1	5	10
H3	2 x 2	4	16

This reinforces the fact that the primary objective of this study to define unique arrangement parameters was satisfied.

B. Study 2

To test the efficiency of the proposed methodology, we extend the same study to the case with:

- 1) Variable local heat transfer coefficient on the convection surface (with the values obtained from Lee *et al.* [22] for liquid flow through parallel microchannel).
- 2) Variable fluid temperature along the flow direction.
- 3) Three different heater sizes on each layer with different power values as given in Table V.
- 4) Interdie resistance of $0.75\text{ K}\cdot\text{mm}^2/\text{W}$ [19].

Interdie thermal resistance was modeled a thin layer contact resistance characterized by the effective thermal resistance between the stacks and the physics available in COMSOL.

A background heat flux of 4 kW/m^2 was added. The entire model was built with COMSOL Multiphysics 5.6 and interfaced with MATLAB 2017b for generating the initial samples and optimization. The model in COMSOL was built with the physics incorporating “heat transfer in solids,” physics using the governing equation same as the previous Study 1. The active silicon region (hotspots) was created as a boundary heat source in each of the silicon substrate layers. The boundary condition at the convective boundary was incorporated using a function that incorporates the local heat transfer coefficient consistent with measurements reported by Lee *et al.* [22].

The local fluid temperature was calculated using the energy balance equation in the following equation:

$$m \cdot C_p (T_{out} - T_{in}) = q'' A_s. \quad (32)$$

A flow rate of 0.5 lpm was chosen for this study, and the heat transfer coefficient corresponding to the flow was chosen from Lee *et al.* [22]. The heat transfer coefficient was given as continuous function in X -direction in COMSOL

$$\int \frac{q'' A_s}{m \cdot C_p} + T_{fi,x} = T_{fi+1,x}. \quad (33)$$

Since the fluid temperature variation is a function of the boundary heat on the top boundary, we divided the top surface into ten divisions, calculated the fluid temperature at the entrance (i), and exit ($i + 1$) at each of these divisions using (31). An expression was written in COMSOL software for incorporating the variation of fluid temperature as a function of each division's outgoing average heat flux. The variation

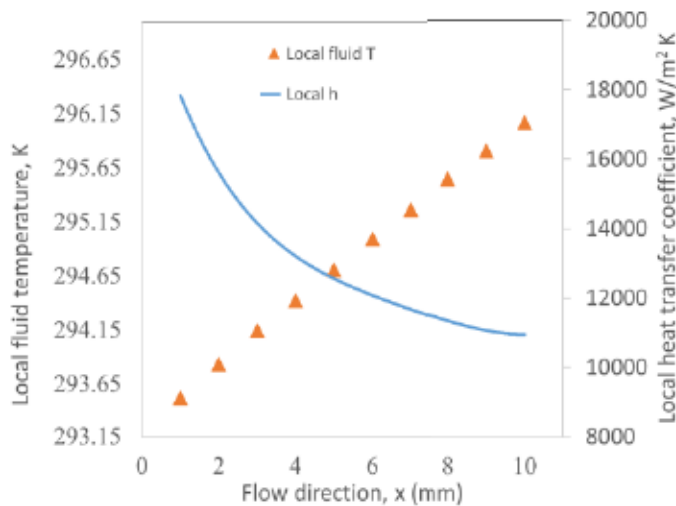


Fig. 9. Variation of local heat transfer coefficient and local fluid temperature boundary condition employed in the model. (Local fluid temperature is a function of local boundary heat flux and the variation shown here is for a specific case.)

of local heat transfer coefficient and fluid temperature is as shown in Fig. 9.

The fluid temperature was coupled with the entire physics and solved.

The grid independent study was performed for the model under consideration. The mesh elements were varied in the range of 17232–292580, and the maximum temperature and temperature uniformity metric were noted. From the results of our grid independence study, it was found that beyond 156420 mesh elements the temperature and uniformity metric changed less than 0.2%, indicating stable mesh. This mesh was employed for the rest of the study. We employed the learning-based sampling algorithm as shown in Fig. 3 to generate initial cases for the optimization study. For this part of the study, a total of 150 samples that does not violate the constraint were generated. Diversified samples are very important for the robust supervised machine-learning algorithm. The response parameters for the randomly chosen diversified samples from this study are given in Table VI.

The samples were solved in the COMSOL Multiphysics and the maximum core temperature, and the temperature uniformity metric was derived from the results. The arrangement parameters as established previously were extracted from the initial sample points. The results from the COMSOL multiphysics model were fed into the supervised learning-based machine-learning algorithm employing the ANN.

Among the 180 data points,

- 1) 110 data point were used for the training purpose.
- 2) 30 data points were employed for validation.
- 3) 40 data points were employed for the testing purpose.

A neuron independent study was performed to decide the optimal number of neurons in the hidden layer. The results indicated that four hidden neurons were optimal at the hidden layer. Fig. 10 shows the Parity plot obtained from the learning algorithm.

The results from Fig. 10 reiterate the fact that the arrangement parameters defined in the article well represent the

TABLE VI
RESPONSE PARAMETERS FOR RANDOMLY CHOSEN DIVERSIFIED SAMPLES FROM THE INITIAL SAMPLES

S No	λ_{1g}	λ_{2g}	CV $_{\lambda 1}$	CV $_{\lambda 2}$	T $_{max}$
1	127	0.22	0.46	1.5	146.41
2	112	1.31	0.29	1.04	139.75
3	50.06	5.43	0.05	0.34	155.98
4	31.8	7.88	0.56	0.32	191.68
5	109.35	0.05	0.07	0.99	134.1

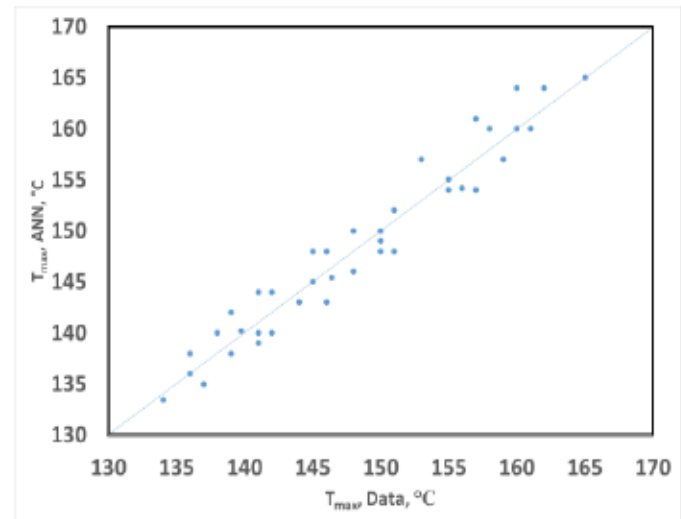


Fig. 10. Parity plot for supervised machine-learning-based ANN using the testing data (40 data points) $R^2 = 0.96$.

TABLE VII
OPTIMAL VALUES OF THE VARIABLES AND RESPONSE PARAMETERS IN THIS STUDY 2

Arrangement Parameter	Value
λ_{1g}	127.54
λ_{2g}	0.83
CV $_{\lambda 1}$	0.49
CV $_{\lambda 2}$	0.73
Response parameters	Value
T $_{max}$	123.08
σ	1.92

problem under consideration. The optimization is performed using the GA as explained in Fig. 6. The optimal arrangement parameter and the response parameter obtained are as given in Table VII.

The schematic of the optimal arrangement for Study 2 is as shown in Fig. 11.

From Fig. 11, it becomes evident that any further increase of λ_{2g} will induce interlayer and interlayer overlap leading to more hot zones and increased maximum temperature. Furthermore, any simultaneous decrease in λ_{2g} and increase in λ_{1g} would move the hotspots away from the entrance, leading to more intense hot regions. Furthermore, the optimal value of CV $_{\lambda 1}$ is not close to zero as it was seen in a previous study

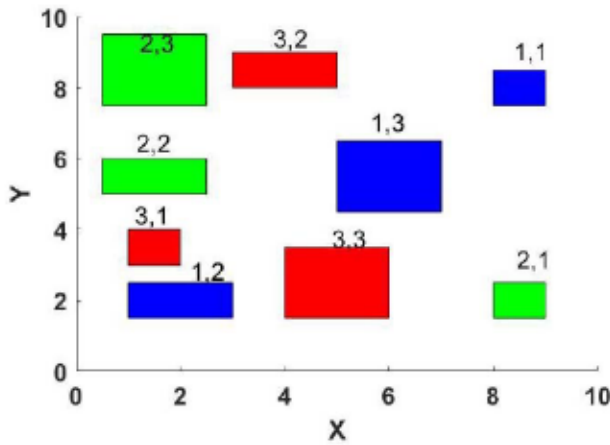


Fig. 11. Schematic of the optimal arrangement [the flow inlet on the top of the package is to the left ($x = 0$) and has the highest heat transfer coefficient]. [The numbers above each heat source represent (layer number i , heat source/hotspot number j).]

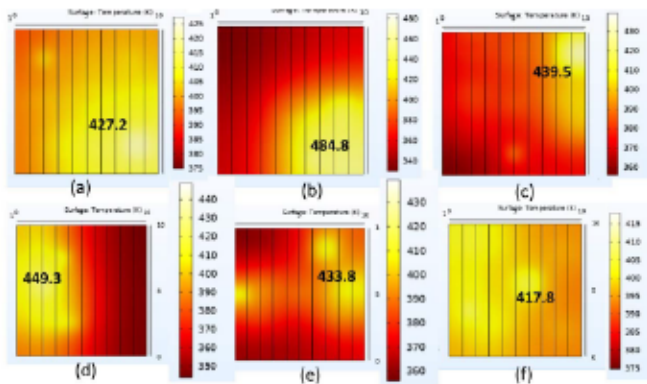


Fig. 12. Contour of absolute chip surface temperature (K) on the convection surface for (a)–(e) nonoptimal and (f) optimal arrangement [the flow inlet is to the left ($x = 0$) and has the highest heat transfer coefficient and decreases downstream]. The vertical lines on the contours represent different work planes created in COMSOL 5.6.

with uniform boundary conditions. This indicates that the optimal arrangement in each layer has become nonsymmetric to adapt to the varying heat transfer boundary condition. The above-stated inferences were a part of the supervised learning algorithm and hence the optimization algorithm was able to converge to the global optimal solution. Therefore, the GA was able to determine the global optimal solution corresponding to the parameters. This study proves that a combination of robust supervised learning-based ANN coupled with a GA can be used as an efficient methodology. For the sake of comparison the nonoptimal case with the optimal case for study 2 is shown in Fig. 12.

The optimal arrangement parameters for a uniform heat transfer coefficient (in Study 1) shown in Fig. 7 is seen to be significantly different from the optimal arrangement parameters shown in Fig. 11 under varying heat transfer coefficient (Study 2).

Owing to the high heat transfer coefficient on one side of the domain the optimal centroid of arrangement is seen to shift and have a nonzero value. However, higher value of λ_{2g} is seen to introduce inter die and intradie partial overlap. To conclude,

there is a unique value of λ_{2g} and λ_{1g} . The arrangement centroid is closer to a high heat transfer coefficient region with minimum overlap and maximum heat spreading. Further increase in λ_{2g} increases the effect of overlap. The optimization algorithm was able to give us the unique optimal arrangement to minimize maximum temperature.

This reiterates the fact that the methodology proposed in the article is still valid for a variable heat transfer boundary condition and can be extended to many such boundary conditions by the designer.

VI. CONCLUSION

The authors performed an efficient multiobjective constrained optimization of the hotspot arrangement for three-layer three hotspots on each layer by employing supervised machine-learning-based ANN combined with a GA. An empirical correlation was developed for the maximum temperature of the 3-D stack as a function of the arrangement parameters. The constraints for the optimization problem were also derived as a function of the arrangement parameters. We considered two objectives: minimizing the maximum temperature and maximizing the temperature uniformity across the stack volume.

The nonconflicting nature of the objective functions helped to reduce the multiobjective optimization problem to a single objective optimization problem. The study was conducted in two parts (Study 1 and Study 2) with constant and varying heat transfer coefficient on the convection surface, respectively. The major conclusions from this study are:

- 1) A constraint incorporated sampling algorithm was developed that could generate initial samples that do not violate the overlap constraint.
- 2) The maximum temperature of the 3-D stack and the temperature uniformity seemed to be nonconflicting objectives.
- 3) There is a unique optimal arrangement of hotspots in the 3-D stack for any boundary condition that could be efficiently represented by the optimal arrangement parameters defined in this article.
- 4) The arrangement parameters defined in this article efficiently recognized the grouping of heaters within the layer and between the layers.
- 5) The methodology demonstrated in this article can be efficiently extended to optimize hotspot location for a 3-D IC stack $n_l \gg 3$ and $n_h \gg 3$ for variable heat transfer boundary conditions.

REFERENCES

- [1] A. Bar-Cohen, "Thermal management of on-chip hot spots and 3D chip stacks," in *Proc. IEEE Int. Conf. Microw., Commun., Antennas Electron. Syst.*, Nov. 2009, pp. 1–8, doi: 10.1109/COMCAS.2009.5385939.
- [2] Y. Li and D. Goyal, *3D Microelectronic Packaging: From Fundamentals to Applications*. Cham, Switzerland: Springer, 2017.
- [3] V. Pangracious, Z. Marrakchi, and H. Mehrez, *Three-Dimensional Design Methodologies for Tree-Based FPGA Architecture*. Cham, Switzerland: Springer, 2015.
- [4] J. Geer, A. Desai, and B. Sammakia, "Heat conduction in multilayered rectangular domains," *ASME J. Electron. Packag.*, vol. 129, no. 4, pp. 440–451, 2007, doi: 10.1115/1.2804094.
- [5] L. Choobineh and A. Jain, "Determination of temperature distribution in three-dimensional integrated circuits (3D ICs) with unequally-sized die," *Appl. Thermal Eng.*, vol. 56, nos. 1–2, pp. 176–184, Jul. 2013.

- [6] L. Choobineh and A. Jain, "An explicit analytical model for rapid computation of temperature field in a three-dimensional integrated circuit (3D IC)," *Int. J. Thermal Sci.*, vol. 87, pp. 103–109, Jan. 2015.
- [7] A. Jain, S. M. Alam, S. Pozder, and R. E. Jones, "Thermal-electrical co-optimisation of floorplanning of three-dimensional integrated circuits under manufacturing and physical design constraints," *IET Comput. Digit. Techn.*, vol. 5, no. 3, pp. 169–178, May 2011.
- [8] W. Chen and B. Bottoms, "Heterogeneous integration roadmap: Driving force and enabling technology for systems of the future," in *Proc. Symp. VLSI Technol.*, Jun. 2019, pp. T50–T51, doi: [10.23919/VLSIT.2019.8776484](https://doi.org/10.23919/VLSIT.2019.8776484).
- [9] T. K. Hotta, C. Balaji, and S. P. Venkateshan, "Optimal distribution of discrete heat sources under mixed convection—A heuristic approach," *J. Heat Transf.*, vol. 136, no. 10, Oct. 2014, Art. no. V003T10A006.
- [10] R. R. Madadi and C. Balaji, "Optimization of the location of multiple discrete heat sources in a ventilated cavity using artificial neural networks and micro genetic algorithm," *Int. J. Heat Mass Transf.*, vol. 51, nos. 9–10, pp. 2299–2312, May 2008.
- [11] T. Ni *et al.*, "Temperature-aware floorplanning for fixed-outline 3D ICs," *IEEE Access*, vol. 7, pp. 139787–139794, 2019.
- [12] S. Rangarajan, Y. Hadad, L. Choobineh, and B. Sammakia, "Optimal arrangement of multiple heat sources in vertically stacked two-layer 3-D IC using genetic algorithm," in *Proc. ASME Int. Tech. Conf. Exhib. Packag. Integr. Electron. Photonic Microsyst.*, Dec. 2019, Art. no. V001T01A003, doi: [10.1115/1.PACK2019-6334](https://doi.org/10.1115/1.PACK2019-6334).
- [13] C. Wang, X.-J. Huang, and K. Vafai, "Analysis of hotspots and cooling strategy for multilayer three-dimensional integrated circuits," *Appl. Thermal Eng.*, vol. 186, Mar. 2021, Art. no. 116336, doi: [10.1016/j.applthermaleng.2020.116336](https://doi.org/10.1016/j.applthermaleng.2020.116336).
- [14] B. Ding, Z.-H. Zhang, L. Gong, M.-H. Xu, and Z.-Q. Huang, "A novel thermal management scheme for 3D-IC chips with multi-cores and high power density," *Appl. Thermal Eng.*, vol. 168, Mar. 2020, Art. no. 114832.
- [15] K. Yazawa *et al.*, "Cooling power optimization for hybrid solid-state and liquid cooling in integrated circuit chips with hotspots," in *Proc. 13th InterSoc. Conf. Thermal Thermomech. Phenomena Electron. Syst.*, May 2012, pp. 99–106, doi: [10.1109/ITHERM.2012.6231419](https://doi.org/10.1109/ITHERM.2012.6231419).
- [16] B. Shi, A. Srivastava, and A. Bar-Cohen, "Co-design of micro-fluidic heat sink and thermal through-silicon-vias for cooling of three-dimensional integrated circuit," *IET Circuits Devices Syst.*, vol. 7, no. 5, pp. 223–231, Sep. 2013, doi: [10.1049/iet-cds.2013.0026](https://doi.org/10.1049/iet-cds.2013.0026).
- [17] A. Bar-Cohen, A. Srivastava, and B. Shi, "Thermo-electrical co-design of three-dimensional integrated circuits: Challenges and opportunities," *Comput. Thermal Sci.*, vol. 5, no. 6, pp. 441–458, 2013.
- [18] J. Cong, J. Wei, and Y. Zhang, "A thermal-driven floorplanning algorithm for 3D ICs," in *Proc. IEEE/ACM Int. Conf. Comput. Aided Design (ICCAD)*, Nov. 2004, pp. 306–313, doi: [10.1109/ICCAD.2004.1382591](https://doi.org/10.1109/ICCAD.2004.1382591).
- [19] L. Choobineh, J. Jones, and A. Jain, "Experimental and numerical investigation of interdie thermal resistance in three-dimensional integrated circuits," *J. Electron. Packag.*, vol. 139, no. 2, Jun. 2017, Art. no. 020908, doi: [10.1115/1.4036404](https://doi.org/10.1115/1.4036404).
- [20] S. Rangarajan and C. Balaji, *Phase Change Material-Based Heat Sinks: A Multi-Objective Perspective*. Boca Raton, FL, USA: CRC Press, 2019.
- [21] *ITRS 2.0 2015 Edition—IEEE Electronics Packaging Society*. Accessed: Jun. 16, 2021. [Online]. Available: <https://eps.ieee.org/technology/heterogeneous-integration-roadmap/itrs-2-0-2015-edition.html>
- [22] P.-S. Lee, S. V. Garimella, and D. Liu, "Investigation of heat transfer in rectangular microchannels," *Int. J. Heat Mass Transf.*, vol. 48, no. 9, pp. 1688–1704, Apr. 2005, doi: [10.1016/j.ijheatmasstransfer.2004.11.019](https://doi.org/10.1016/j.ijheatmasstransfer.2004.11.019).