

Machine Learning Approach for Optimization of 3D Integrated Circuits and Systems

M. Sandhya, G. Nihar, Dr. Shruti Bhargava Choubey

Sreenidhi Institute of Science and Technology

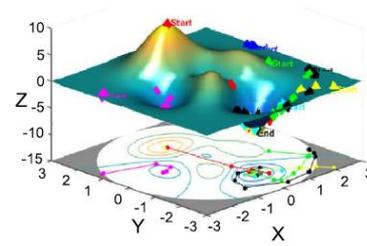
Abstract - The 3D integration is helpful in improving the performance and density of electronic systems. Their codesign is required however, since the thermal and electrical performances for 3D integration are related to each other. To address these optimization problems in engineering, a promising new approach in artificial intelligence, Machine learning is employed. In this paper, for the optimization of 3D integrated circuits, we employ machine learning technique where in the electrical and thermal properties of the design need to be analyzed together to maximize the performance. The modelling of such systems can be challenging due to the multiscale geometries involved, which increase the computation time per iteration. In this paper, we show that machine learning can be applied to such systems to achieve the desired performance where multiple parameters can be optimized with minimum number of iterations. These results have been compared with other promising optimization techniques in this paper. The results show that on an average 4.4%, 31.1%, 6.9% improvement in temperature gradient, CPU time and skew are possible using machine learning, as compared with other techniques.

I. INTRODUCTION

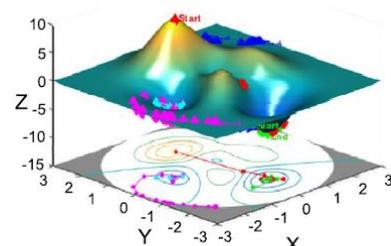
Continuous growth in circuit density and performance of electronic systems has resulted in new technologies to realize these goals. Three-dimensional integration, an innovative technique in systems packaging, provides solutions for increasing performance and density of electronic systems [1]. However, improved power and circuit density also results in the increase of heat flux, which in turn increases temperature and causes thermal related reliability problems [2]. Increased temperature and their gradients will degrade electrical performance, as it can have a direct impact on clock skew. Since electrical performance and thermal performance are related to each other through joule heating, their combined analysis is required for predicting the temperature distribution accurately [3]. Additionally, the clock tree needs to be modeled in the presence of the temperature distribution for estimating clock skew [3]. Since a significant number of parameters such as physical geometries, interface materials, fan speed, and so on have a direct influence on the temperature profile, these parameters must be suitably tuned to achieve the desired electrical performance. Hence, this gives rise to a multivariate system optimization problem. In this paper, the attributes of the problem require: 1) black box optimization since the output function is unknown; 2) possible application to nonconvex response surfaces, since the system behavior is unknown a priori; and 3) minimizing iterations for attaining optima due

to the Multiphysics and multiscale modeling involved which increases the computational time.

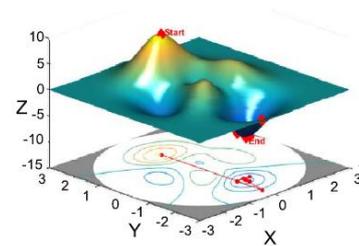
Several statistical methods have been proposed by recent studies such as worst case and Monte Carlo analyses [4] to optimize a large number of design parameters. Due to the large number of simulation cases and expensive computational overhead for these methods, others have proposed approaches that reduce the number of simulations, using design of experiments (DOEs) [5]. However, the DOE approach has restrictions such as: 1) interactions between parameters need to be as small as possible and 2) the number of levels is normally limited to below three. Moreover, these techniques can lead to quantization error during optimization, due to their implementation at discrete points in the problem space.



(a)



(b)



(c)

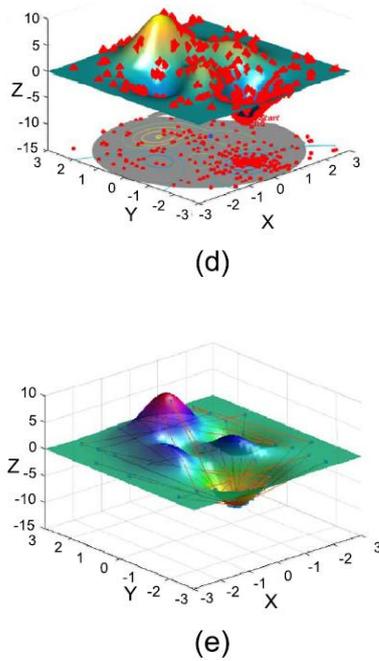


Fig. 1. Optimization methods applied to a predefined function. (a) Multistart (function counts = 345). (b) Global search (function counts = 273). (c) Pattern search (function counts = 272). (d) Genetic algorithm (function counts = 2650). (e) BO (function counts = 100) [8].

Other approaches for global optimization are also available, as discussed in [6]. As an example, global optimization algorithms typically require vast computing resources due to a combination of large compute time for each data set and number of data sets. As shown in Fig. 1(a)–(d), global optimization algorithms when applied to “peaks” function, which is an example function for two variables in MATLAB, required between 272–2650 function counts to converge to the minimum value. In contrast, machine learning based on Bayesian optimization (BO) [7] applied to the same function, required just 100 function counts to converge to the minimum, as shown in Fig. 1(e). Function count represents the number of objective function evaluations during the optimization process, where each iteration can require multiple function counts. This was our main motivation for investigating machine learning methods in this paper. In addition, such methods can be applied to non-convex, black box optimization problems as well, which was another requirement.

Machine learning has three elements, task, experience, and performance, which consists of two phases, training, and evaluation/execution, as shown in Fig. 2. “Task” and “performance” represent training and target respectively, while “experience” is used to improve the target performance [9].

Though there are several algorithms available in the literature for machine learning, our focus in this paper is on BO

due to its capability for handling a large number of input parameters and its quick convergence [7]. Machine learning methods have been applied to electromagnetic problems [11], static timing analysis [12], high-speed interconnect systems [13], and time domain performance estimation [14] in the past. In this paper, to minimize temperature and temperature gradients, we apply machine learning for the optimization of 3-D ICs and systems.

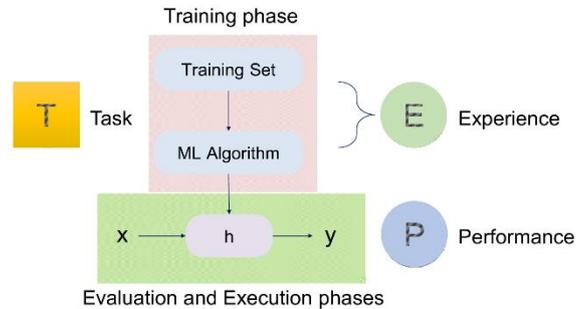


Fig. 2. Concept of machine learning consists of training and evaluation/execution phases [9], [10].

For minimizing the number of training data sets required, we chose BO with Gaussian process (GP), since GP helps improve the performance [7]. Several machine learning methods based on support vector machine and sparse-vertical link have been discussed in the past for optimizing electrical circuits with minimum training data [15] and for optimizing 3-D circuits [16]. However, these methods required expense for problem-dependent hyper-parameter and complex allocation problem, respectively, therefore BO method is more efficient for optimization. In [17], a preliminary application of machine learning for 1-D problems was discussed. In this paper we expand [17] to include multivariable optimization along with correlation of the solver with measurements and convergence study.

This paper is organized as follows: in Section II we describe the problem in the context of 3-D integration and we discuss about a test chip for validating the solver with measurements; in Section III we discuss the system optimization using machine learning with results being provided in Section IV; followed by conclusions in Section V.

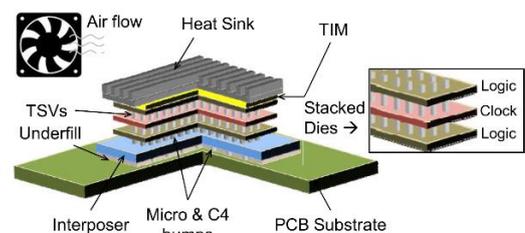


Fig. 3. Configuration of a 3-D system for optimization.

II. PROBLEM DEFINITION

A. 3D Integrated System

Our objective in this paper is to minimize the global skew

caused due to temperature and temperature gradients in 3-D systems. We take the help of simulated temperature profiles superimposed on to temperature-sensitive clock tree to estimate the skew.

An example of a 3-D integrated system comprising of stacked dies, interposer, and printed circuit board (PCB), is shown in Fig. 3. To build the full system model of chip/package/PCB, we rely on an iterative solver [18] based on the finite volume method which numerically solves the coupled thermal and electrical partial differential equations. The solver uses a volumetric cell for discretization and incorporates the user defined conduction and convection boundary conditions.

The solver uses a nonuniform grid and domain decomposition to deal with the multiscale geometries and accounts for multiple materials in the structure associated die, interposer, and PCB by implementing the necessary boundary conditions between cells containing different materials properties.

We assume power maps on the chip [3] and use nonconformal domain decomposition and parameterized model order reduction techniques; as described in [18]; to compute the temperature profiles. Signal and power integrity performance, such as skew, noise, and impedance, are then computed with the help of a circuit solver, which includes temperature gradients and power delivery network response super-imposed on an H tree clock network containing temperature-dependent nonlinear clock buffers and interconnect models [3]. This procedure, as shown in Fig. 4 and discussed in detail in [3] results in the computation of the temperature distribution across the die along with temperature dependent skew, jitter and power supply noise for the clock tree in the center die. In this paper, our main focus is on optimizing the temperature distribution on the center die and computing the resultant skew on the clock distribution network (CDN). It is important to note that temperature gradient has a significant impact on clock skew [3].

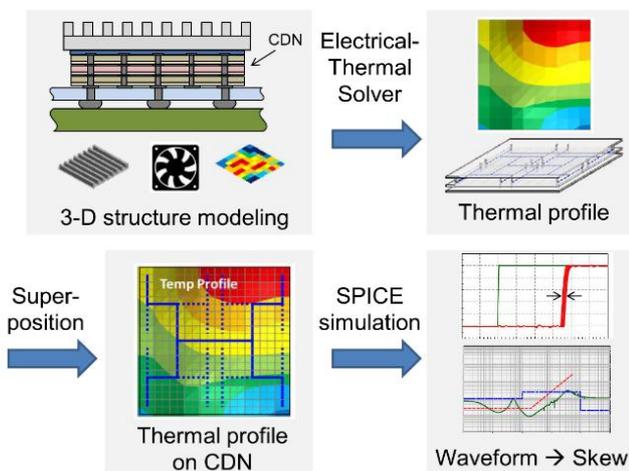


Fig. 4. Flow of electrical–thermal simulation for 3-D system design.

B. Materials and Methods

All computations in this paper are based on the electrical–thermal solver described in [18]. To calibrate the accuracy of the results, a custom IC was designed, and a test vehicle fabricated. The test chip contained both monitoring circuits and temperature generation. Although the test vehicle did not contain a 3-D stack with TSVs due to limited availability, I believe that the test vehicle provides a method for calibration and attests to the accuracy of the simulations.

On-chip heaters implemented using polysilicon resistors were used to generate the heat while MOS diodes were used to record the temperature. The dimension of on-chip heaters was $100\ \mu\text{m} \times 100\ \mu\text{m}$ and these heaters were implemented on the poly-silicon layer. Sixteen pairs of heaters, temperature sensors and MOS diodes were placed on a 4×4 grid on the chip which measured $3.8\ \text{mm} \times 3.8\ \text{mm}$.

The chip consisted of six metal layers and was fabricated using the 180-nm process. This method was used to design and fabricate the prototype, since it was part of a low-cost multiproject wafer. Since our objective was to validate the models and modeling process, we chose the 180 nm technology node. The specifications of the CMOS process are shown in Table I. The layout of the chip is shown in Fig. 5(a). The fabricated chip was directly bonded to a PCB (chip-on-board), which measured $100\ \text{mm} \times 100\ \text{mm}$, as shown in Fig. 5(b).

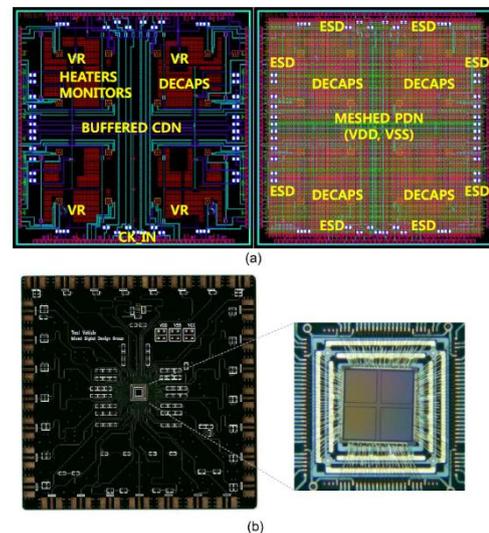


Fig. 5. (a) Chip layout. (b) Fabricated PCB and wire-bonded chip.

C. Validation of Electrical–Thermal Solver

On-chip temperature gradients were measured using monitoring blocks and temperature generating. By varying resistance and input voltage we induced variable current to each heater with resistor networks built on test board. We used temperature monitoring circuits with diodes to measure the local temperature, Fig. 6(a) and (b) shows the measured I–V profile of temperature monitoring circuits

and the measured I–V curves for different temperatures, respectively.

Table 1. Fabrication process specifications

Process	180nm cmos
Layers	1 polysilicon layer, 6 metal layers
Devices	1.8V(thin-oxide)/3.3V(thick-oxide)/5V
Min. gate length	0.18µm for 1.8V, 0.30/0.35µm for 2.5V
Substrate	P Substrate with N wells

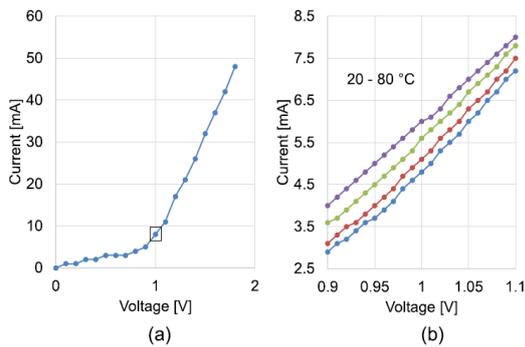


Fig. 6. (a) Measured I–V profile of temperature monitoring circuits.(b) I–V profile with temperature variations.

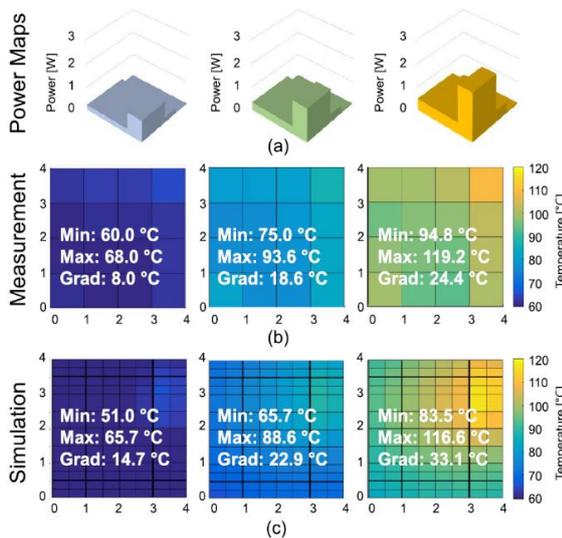


Fig. 7. (a) Power maps used for simulation and measurement. (b) Measured temperature profiles. (c) Simulated temperature profiles.

The power consumed by each heater was calculated by a voltage source and resistor divider resulting in a power map as shown in Fig. 7(a). The electrical–thermal solver was used to compute the temperature distribution on the die for the test vehicle in Fig. 5. Since the typical heat transfer coefficient for natural convection is around 5 W/(m² ·K)

[19], a heat transfer coefficient of 4.0 W/(m² ·K) was used as the convection boundary condition for analysis, which accounts for any radiation effects as well. Fig. 7(b) and (c) shows the simulated and measured temperature distribution for the power map used in Fig. 7(a). The measured results are well correlated with the electrical–thermal simulations. From Fig. 7, the correlation is promising for minimum and maximum temperatures for the three power maps (11.9%–15.0% and 2.2%–3.4% error in simulations) while the error is larger for the temperature gradients because of the smaller values involved. All the same, these correlations provide a reasonable degree of confidence in the simulated temperatures, as there is some inaccuracy in the position of the heaters and monitoring circuits due to the 4×4 grid used for the chip, as opposed to a much finer nonuniform grid used in the simulations.

III. OPTIMIZING THE SYSTEM

A. Bayes’ Theorem

BO originated from a well-known equation in probability theory and statistics, called Bayes’ theorem. Bayes’ theorem [20] can be applied to machine learning using

$$P(h|D) = \frac{P(D|h)}{P(D)} \quad (1)$$

In (1) “P(D)” and “P(h)” are the probabilities of observing “D” and “h,” respectively. They are referred to as the prior over data “D” and hypothesis “h,” respectively. “P(D|h)” is the probability of observing data “D” given a hypothesis “h” and is referred to as the likelihood while P(h|D) is the probability of hypothesis “h” given data “D” also called the posterior.

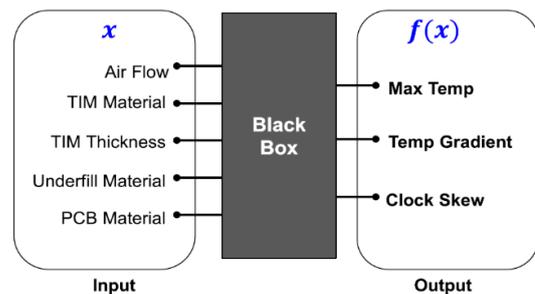


Fig. 8. Black box function with multivariable for 3-D system design.

Equation (1) interprets Bayes’ rule regarding possibilities of multiple events, before (prior to) and after (posterior to) event which can be rewritten in the form

$$P(h|D) \propto \frac{P(D|h)P(h)}{P(D)} \quad (2)$$

where, the proportionality symbol indicates that if “h” varies but keeping “D” fixed, the left-hand side is equal to a constant times the right-hand side. In words, posterior is proportional to prior times likelihood: determined by the

Bayes factor [20]. This forms the framework for BO used in this paper.

B. Black Box Function

For the 3-D system in Fig. 3, the input parameters “x” that needs to be optimized to achieve a target output “f(x)” are shown in Fig. 8. The black box function f(x) is obtained by using the electrical–thermal solver as discussed earlier. Based on sensitivity analysis, we picked five input variables for optimization which are, heat transfer coefficient (determined by the air flow rate), thermal conductivity of thermal interface material (TIM), TIM thickness, PCB and thermal conductivity of under-fill (UF) material, while the target parameters chosen were maximum temperature and temperature gradient. The improvement in clock skew resulting from the temperature distribution was used as a metric for the optimization.

The target parameters namely, maximum temperature and temperature gradient were combined to form the function f(x) by assigning weights for each of the parameters, as explained in the coming section.

C. Bayesian Optimization with Gaussian Process

In Bayesian statistics, the uncertainty is modelled with a prior probability distribution. In other words, we estimate the distribution and this information is used to decide the point evaluated next, which is a key point of BO that differentiates it from other methods.

For GP priors, the model uses a joint Gaussian with the entire set of available observation points. In this optimization, the function “f” is defined as a GP prior with mean function “m” and covariance function “k.” Based on prior observation points “M” for the variable “x,” the prior function f(x1:M) for each variable is defined as a GP given by

$$f(x1:M) = N(\mu(x1:M), k) \quad (3)$$

where x1:M represent the “M” observation points for each input variable, $\mu(x1:M)$ is the corresponding mean vector and k (also called the kernel) is the corresponding covariance matrix given by [21]

$$\mu(x1:M) = [\mu(x1)\mu(x2) \cdots \mu(xM)]^T \quad (4)$$

$$K(x1:M) = \begin{bmatrix} k(x1, x1) & \cdots & k(x1, xM) \\ \vdots & \ddots & \vdots \\ k(xM, x1) & \cdots & k(xM, xM) \end{bmatrix} \quad (5)$$

where the covariance is defined by

$$k(x_i, x_j) = \exp\left(-\frac{1}{2}\|x_j - x_i\|^2\right). \quad (6)$$

To predict f(xM+1) at the next data point, we consider the joint distribution over f of the old data points and new data point, as shown in (7). The optimization problem now relates to maximizing (or minimizing) f(x) subject to x where f(xM+1) can be a nonconvex black-box function defined by

$$\begin{bmatrix} f(x1:M) \\ f(xM+1) \end{bmatrix} \sim N\left(\begin{bmatrix} m(x1:M) \\ m(xM+1) \end{bmatrix}, \begin{bmatrix} K & k \\ k^T & k(x_{M+1}, x_{M+1}) \end{bmatrix}\right) \quad (7)$$

where K is the kernel matrix and k is the kernel function given by (5) and (6).

From [21], the mean and variance of f(xM+1) can be computed as

$$\mu(x_{M+1}) = k^T K^{-1} f1 : M \quad (8)$$

$$\sigma^2(x_{M+1}) = k(x_{M+1}, x_{M+1}) - k^T K^{-1} k. \quad (9)$$

Such an approach can be extended to N independent input variables, where in this paper we use N <= 5.

This technique gives a posterior distribution of the unknown function. We can choose the next value of the function representing the targeted values by either maximizing or minimizing an acquisition function (explained later).

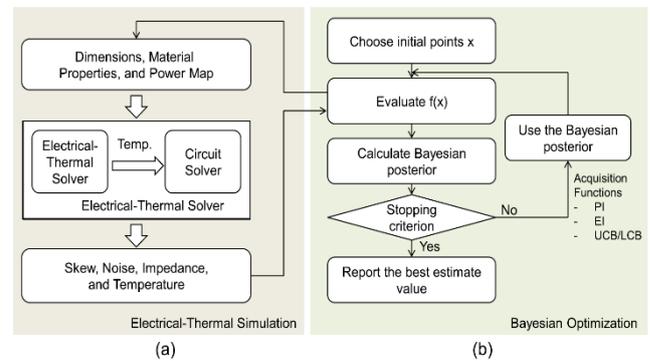


Fig. 9. Proposed flow for electrical–thermal simulation using BO. (a) Electrical–thermal simulation. (b) BO.

The typical flow of BO using GP [22] is as follows.

- 1) Choose initial points of N input variables x and evaluate f(x) including error (with regard to the target value desired).
- 2) While [f(x)–target] ≤ error, calculate Bayesian posterior distribution on “f” from the points observed.
- 3) Using the prior observation points and acquisition function determine the point to evaluate next.
- 4) Stop if the error criterion is met and report the point with the best value.

This approach is based on the infinite-metric GP optimization algorithm presented in [23].

Based on BO with GP, the flow for system optimization is as shown in Fig. 9 where the electrical–thermal simulator is used to compute the black box function. In the flowchart, acquisition functions are used to choose the posterior. In general, three acquisition functions have been widely used in the open literature for GP based optimization, namely [7]: probability of improvement (PI), expected improvement (EI), and upper/lower confidence bound (UCB/LCB)

[24]– [26]. The goals of the first two strategies are to maximize the PI and the EI of the current value, respectively. The third strategy is targeted toward exploiting UCBs/LCBs with high probability using acquisition functions that minimize regret [7]. In this paper, LCB is used, described by

$$\mathbf{x}_{M+1} = \underset{\mathbf{x}}{\operatorname{argmin}} \quad [\mu(\mathbf{x}_i) - \kappa \sigma(\mathbf{x}_i)] \quad (10)$$

where $\kappa \geq 0$ and $\kappa = (2 \log \pi 2x_2/12v)1/2$, (where v equals 0.05), and $\mu(x_i)$ and $\sigma(x_i)$ are determined from (8) and (9) for each input parameter. It is important to note that the selection of the next sample does not require the computation of $f(x)$, since (10) is computed only based on the previous results, which minimizes computational time. Since the entire procedure minimizes the number of points at which $f(x)$ is computed [21], the computational time required for optimization can be reduced significantly. Unlike most optimization techniques, this approach provides a posterior distribution of the unknown function and hence the search involves determining the function (rather than the output itself) that is closer to the targeted goal.

As an example, for choosing the next value, Fig. 10 shows the distribution of the function with two random variables X_1 and X_2 , along with the posterior mean and variance across the input parameter space, the distribution of the acquisition function defined using the LCB and selection of the next point. The minimum of the acquisition function is chosen as the next point in the input parameter space, shown using a triangle marker in Fig. 10. In the figure, X_1 represents the heat transfer coefficient of the air flow in $W/(m^2 \cdot K)$ and X_2 represents thermal conductivity of the TIM material in $W/(m \cdot K)$ for the 3-D system being optimized, with the target function $f(x)$ described in a later section.

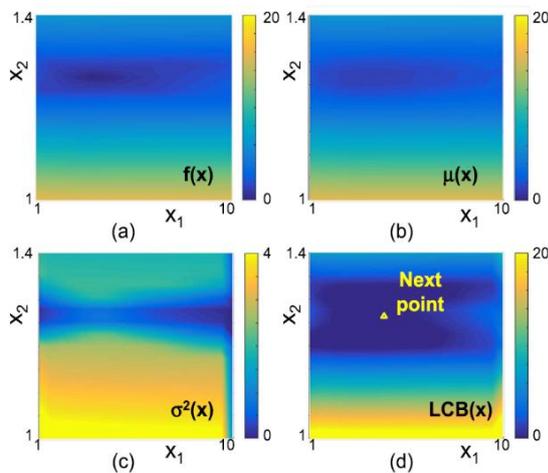


Fig. 10. Distribution plots of (a) function, (b) posterior mean, (c) posterior variance, and (d) LCB acquisition function for optimization of the 3-D system.

IV. RESULTS

A. System Details

A 3-D system for optimization comprises of stacked dies, interposer, and PCB, as shown in Fig. 3 of Section II. We use multiple power maps as described in [17] to simulate the 3-D structure where the power maps are randomly distributed on the top and bottom die. The total power for the three dies was 50 W with 20 W for the bottom and the top die respectively, and 10 W for the center die. The center die incorporates the CDN, which is used to compute the skew. The clock buffers and interconnects used for the CDN were based on the 45-nm process [27], as described in [3]. The three power maps were used to reflect the three different temperature distributions. Fig. 11 shows the power maps used.

B. Input Parameters

As discussed earlier, five input parameters were selected for optimization with details provided in Table II, along with their respective range. The parameters are, thermal conductivity of the TIM, air flow velocity or heat transfer coefficient, thermal conductivity of UF material, thermal conductivity of PCB, and thickness of TIM. The range for these parameters were chosen based on manufacturability.

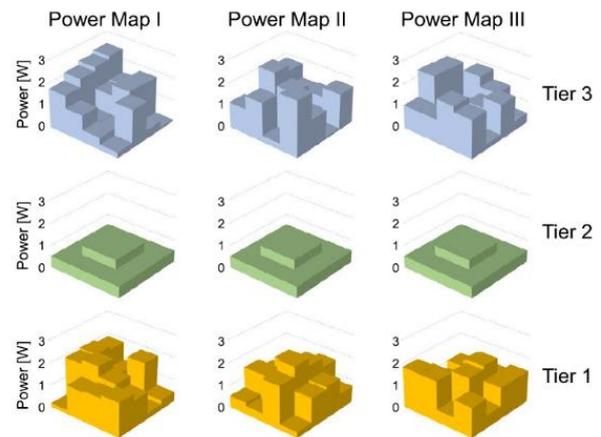


Fig. 11. Power maps used for optimization.

TABLE 2 INPUT VARIABLES FOR OPTIMIZATION

Parameter		Unit	Min	Nom	Max
Heat Transfer Coefficient	x_1	$W/(m^2.k)$	1	5	10
TIM property	x_2	$W/(m.k)$	1	1.2	1.4
TIM thickness	x_3	mm	0.16	0.20	0.24
UF material	x_4	$W/(m.k)$	0.3	4.3	8.3
PCB material	x_5	$W/(m.k)$	0.3	0.3	4.3

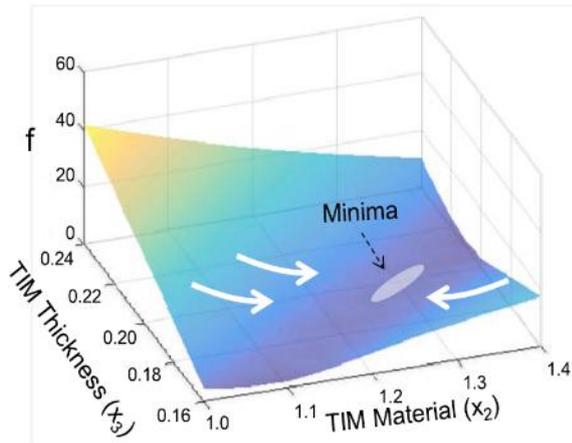


Fig. 12. Response surface with a target value.

C. Usage of Target Function for Multiobject Optimization

There are two target parameters that are important for optimization namely, maximum temperature and temperature gradient on the center die. Limiting the maximum temperature is important to maximize system reliability while minimizing the temperature gradient is required to minimize clock skew. Both these parameters vary with the input parameters.

As an example, the response surface of the function f in (11) is shown as a function of two parameters, namely TIM thickness and thermal conductivity in Fig. 12. The figure shows a surface where a combination of input parameters leads to a minima, where the minima here corresponds to a temperature and temperature gradient less than 120 °C and 25 °C, respectively. Our goal in this paper is to achieve the target value for the maximum temperature and temperature gradient by tuning the input parameters through optimization. Our target function is defined as

$$f(\mathbf{y}_1, \mathbf{y}_2) = \sum_{i=1}^2 w_i \times \mathbf{y}_i \quad (11)$$

where w_i and y_i are the weights and the selected outputs, respectively. In (11)

y_1 = maximum temperature TMAX

and

y_2 = temperature gradient TGRAD.

In this paper, we used weights of $w_1 = 0.34$ and $w_2 = 4.5$ in (11) to define the target function. This was determined based on the importance of reducing the clock skew as opposed to minimizing the maximum temperature, though both are important to ensure a reliable system.

D. Optimization with Multiple Input Parameters

The target parameters defined in Table II were used along with the target function in (11) and three input parameters heat transfer coefficient, TIM thickness and TIM thermal conductivity to perform optimization using power map I in Fig. 11. Fig. 13 and Table III show the optimization results. A total of 100 iterations were used. In Fig. 13,

the sampling points used for each iteration are shown for the four cases evaluated in Table III (plotted only as a function of input variables X_1 and X_2).

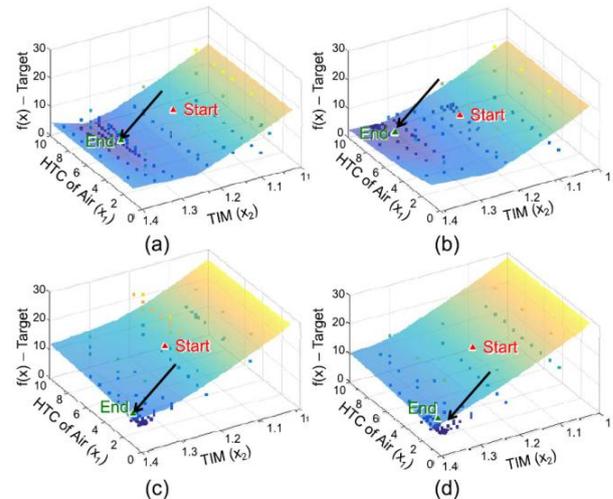


Fig. 13. Optimization results for heat transfer coefficient (X1) and TIM thermal conductivity (X2) showing convergence; TIM thickness (X3) is not plotted (N = 3).

For Case (d), the sampled points are shown as a function of three input parameters in Fig. 14(a). The optimization algorithm “Starts” from an initial value, which represents the median of each parameter, and converges to the optimized value (indicated as “End”) in the figure. As can be noted from Fig. 14, the sampling is non uniform. The maximum temperature and temperature gradient before and after optimization on the center die containing the CDN are shown in Fig. 14(b).

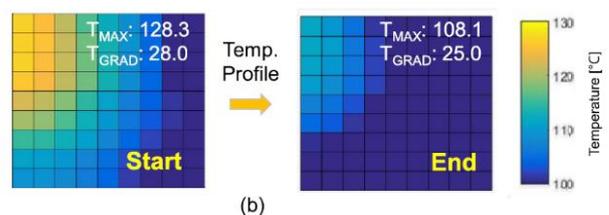
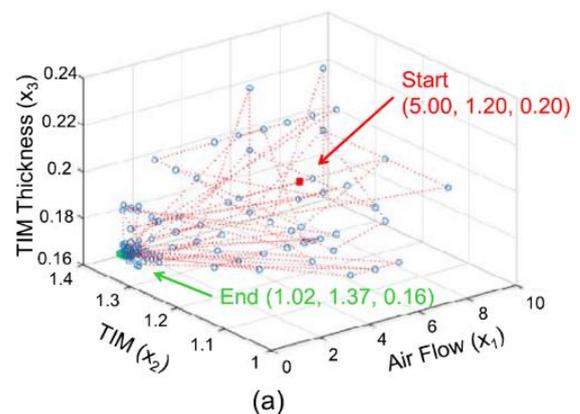


Fig. 14. Optimization results with target value of TMAX: 120.0 and TGRAD: 25.0. (a) Found X_s and (b) temperature.

TABLE 3 OPTIMIZATION RESULTS WITH VARIOUS TARGET VALUES

Case	Target T_{MAX}	Target T_{GRAD}	x_1	x_2	x_3	T_{MAX}	T_{GRAD}
(a)	125.0	27.5	5.72	1.27	0.199	122.9	27.5
(b)	120.0	27.5	8.83	1.30	120.2	120.2	27.5
(c)	125.0	25.0	1.01	1.37	0.164	108.0	25.0
(d)	120.0	25.0	1.01	1.37	0.164	108.1	25.0

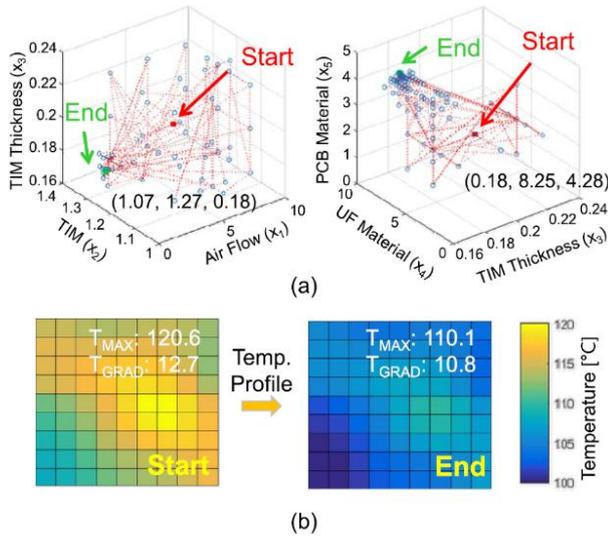


Fig. 15. Optimization with power map II. (a) Iterations shown as a function of three parameters only. (b) Temperature distribution.

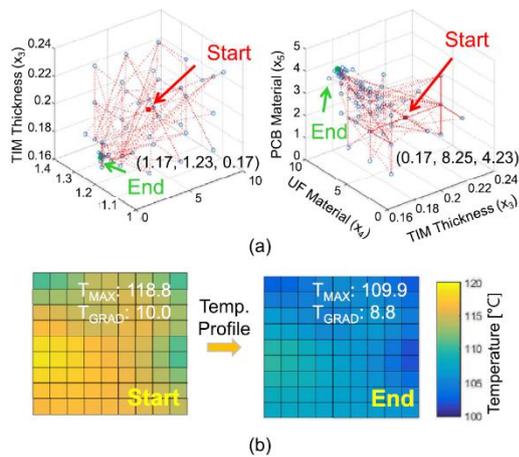


Fig. 16. Optimization with power map III. (a) Iterations shown as a function of three parameters only. (b) Temperature distribution.

To verify the efficiency of the optimization procedure, a case study was performed with various power maps shown in Fig. 11 [17] with more input variables ($N = 5$) and with an iteration number of 200. Figs. 15 and 16 show the results with power map II and power map III, respectively. The target values used for T_{MAX} and T_{GRAD} were $110\text{ }^{\circ}\text{C}$ and $11\text{ }^{\circ}\text{C}$ for power map II and $110\text{ }^{\circ}\text{C}$ and $9\text{ }^{\circ}\text{C}$ for power

map III, respectively. Optimization results show convergence to the target value in Figs. 15 and 16. The temperature distribution before and after optimization are also shown in these figures. Before optimization, power map II and power map III resulted in a clock skew of 51.8 and 39.2 ps, respectively.

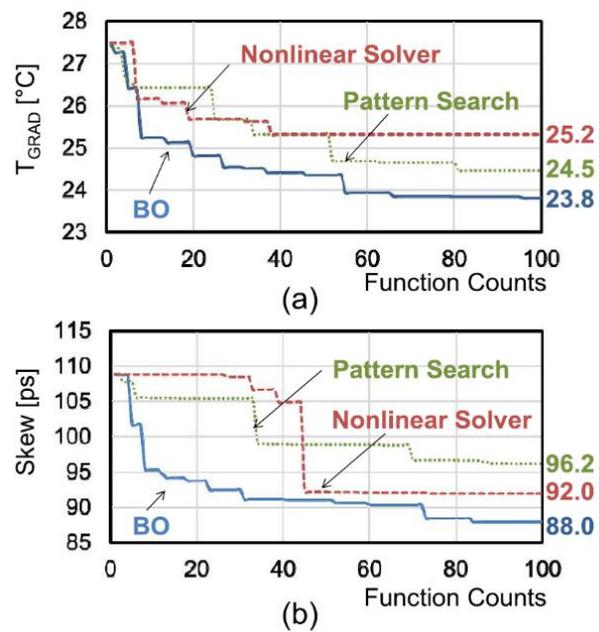


Fig. 17. Comparison of convergence between pattern search, nonlinear solver, and BO (a) temperature gradient and (b) thermal skew.

After optimization, power map II and power map III resulted in a clock skew of 44.2 and 33.0 ps, respectively. The optimization results are shown in Table IV.

TABLE 4 OPTIMIZATION RESULTS WITH VARIOUS POWER MAPS

Power Map	T_{MAX} [°C]		T_{GRAD} [°C]		Skew [ps]	
	Before	After (%)	Before	After (%)	Before	After (%)
I	127.4	115.3 (-9.5%)	27.4	25.0 (-8.8%)	108.9	93.7 (-14.0%)

II	120.6	110.1 (-8.7%)	12.7	10.8 (-15.0%)	51.8	44.2 (-14.7%)
III	118.8	109.9 (-7.5%)	10.0	8.8 (-12.0%)	39.2	33.0 (-15.8%)

TABLE 5 COMPARISON OF OPTIMIZATION PERFORMANCE AFTER 100 FUNCTION COUNTS

	Non-Linear Solver	Pattern Search	BO (This work)
T _{GRAD} [°C]	25.2(+5.9%)	24.5(+2.9%)	23.8
CPU Time (normalized)	1.38(+38.5%)	1.53(+52.6%)	1
Skew [ps]	92.0(+4.5%)	96.2(+9.3%)	88.0

E. Comparison

To compare the optimization performance with existing methods and algorithms, the number of function counts and optimized values were compared. Fig. 17 compares the optimization results, for temperature gradient and the resulting skew for power map I and five input parameters, using BO, “pattern search” (available in MATLAB) and “fmincon,” a constrained nonlinear minimization solver (also available in MATLAB). We chose the “pattern search” and “fmincon” algorithms for comparison since they led to fewer function counts as compared with other methods described in Fig. 1. After 100 function counts BO produced temperature gradient and thermal skew of 23.8 °C and 88.0 ps respectively as compared with 24.5 °C and 96.2 ps using “pattern search” and 25.2 °C and 92.0 ps using “fmincon,” as illustrated in Fig. 17. Fig. 17 also shows a faster convergence rate for BO as compared with “pattern search” and “fmincon” algorithms, especially during the early period.

A comparison of the optimization results including temperature gradient, normalized CPU time for temperature gradient, and optimized thermal skew is shown in Table V.

V. CONCLUSION

In this paper we presented machine learning combined with BO, for the optimization of the electrical–thermal performance of 3-D integrated circuits and systems. The optimization results and comparison with other techniques show several advantages with the proposed approach. Our conclusion is that the described method is suitable for the optimization of system-level electrical– thermal co-

simulation problems, (which often take long simulation time and an even large number of simulation cases), is accurate and demands lower computational cost (–31.1% as CPU time) as compared with other traditional design optimization methods. This method also showed the capability of handling a large number of input parameters with fast convergence and flexibility. This optimization approach using machine learning methods can become useful when system complexity increases along with many input parameters that need to be optimized simultaneously, especially for 3-D applications. Since many BO algorithms have been presented in the open literature, it can be believed that the efficiency of the optimization described in this paper can be increased further.

VI. ACKNOWLEDGMENTS

We are thankful to the IJATER journal for providing us with an opportunity to publish our scientific work in their Journal.

REFERENCES

- [1]. K. Banerjee, S. J. Souri, P. Kapur, and K. C. Saraswat, “3-D ICs: A novel chip design for improving deep-submicrometer interconnect performance and systems-on-chip integration,” *Proc. IEEE*, vol. 89, no. 5, pp. 602–633, May 2001.
- [2]. M. Swaminathan and K. J. Han, *Design and Modeling for 3D ICs and Interposers*. Singapore: World Scientific, 2013.
- [3]. S. J. Park, N. Natu, and M. Swaminathan, “Analysis, design, and prototyping of temperature resilient clock distribution networks for 3-D ICs,” *IEEE Trans. Compon. Packag. Manuf. Technol.*, vol. 5, no. 11, pp. 1669–1678, Oct. 2015.
- [4]. M. D. Meehan and J. Purviance, *Yield and Reliability in Microwave Circuit and System Design*. Boston, MA, USA: Artech House, 1993.
- [5]. G. Taguchi and S. Konishi, *Orthogonal Arrays and Linear Graphs*. Dearborn, MI, USA: Amer. Supplier Inst. Inc., 1987.
- [6]. Global Optimization Tool Boxuser’s Guide in MATLAB, accessed on Mar. 10, 2016. [Online]. Available: http://www.mathworks.com/help/pdf_doc/gads/gas_tb.pdf
- [7]. J. Snoek, H. Larochelle, and R. P. Adams, “Practical Bayesian optimization of machine learning algorithms,” in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 2951–2959.
- [8]. [8] S. DeLand, *Tips and Tricks—Getting Started Using Optimization With MATLAB*. Natick, MA, USA: MathWorks, Nov. 2012.
- [9]. T. Mitchell, *Machine Learning*. New York, NY, USA: McGraw-Hill, 1997.
- [10]. Alpaydin, *Introduction to Machine Learning*, 2nd ed. Cambridge, MA, USA: MIT Press, 2009.
- [11]. Q. J. Zhang and K. C. Gupta, *Neural Networks for RF and Microwave Design*. Norwood, MA, USA: Artech House, 2000.

- [12]. B. Kahng, M. Luo, and S. Nath, "SI for free: Machine learning of interconnect coupling delay and transition effects," in Proc. ACM/IEEE Int. Workshop Syst. Level Interconnect Predict. (SLIP), Jun. 2015, pp. 1–8.
- [13]. W. T. Beyene, "Application of artificial neural networks to statistical analysis and nonlinear modeling of high-speed interconnect systems,"
- [14]. IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst., vol. 26, no. 1, pp. 166–176, Jan. 2007.
- [15]. N. Ambasana, G. Anand, B. Mutnury, and D. Gope, "Eye height/width prediction from S-parameters using learning-based models," IEEE Trans. Compon. Packag. Manuf. Technol., vol. 6, no. 6, pp. 873–885, Jun. 2016.
- [16]. H. Lin and P. Li, "Circuit performance classification with active learning guided sampling for support vector machines," IEEE Trans. Computer- Aided Des. Integr. Circuits Syst., vol. 34, no. 9, pp. 1467–1480, Sep. 2015.
- [17]. S. Das, J. R. Doppa, P. P. Pande, and K. Chakrabarty, "Designspace exploration and optimization of an energy-efficient and reliable 3D small-world network-on-chip," IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst., vol. PP, no. 99, pp. 1–18, Aug. 2016.
- [18]. S. J. Park, H. Yu, and M. Swaminathan, "Preliminary application of machine-learning techniques for thermal-electrical parameter optimization in 3-D IC," in Proc. IEEE Int. Symp. Electromagn. Compat. (EMC), pp. 402–405, July 2016.
- [19]. J. Xie and M. Swaminathan, "Electrical-thermal co-simulation of 3D integrated systems with micro-fluidic cooling and Joule heating effects," IEEE Trans. Compon., Packag., Manuf. Technol., vol. 1, no. 2, pp. 234–246, Feb. 2011.
- [20]. J. H. Whitelaw. Convective Heat Transfer, accessed on Apr. 12, 2016. [Online]. Available: <http://www.thermopedia.com/content/660/>.
- [21]. Gelman, Bayesian Data Analysis. Boca Raton, FL, USA: CRC Press, 2008.
- [22]. Brochu, V. M. Cora, and N. de Freitas. (2010). "A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning." [Online]. Available: <https://arxiv.org/abs/1012.2599>
- [23]. W. B. Powell and P. I. Frazier, "Optimal learning," in Proc. Tuts. Oper. Res. INFORMS, 2008, pp. 213–246.
- [24]. K. Kawaguchi, L. P. Kaelbling, and T. Lozano-Pérez, "Bayesian optimization with exponential convergence," in Proc. Adv. Neural Inf. Process. Syst. (NIPS), 2015, pp. 2809–2817.
- [25]. J. Mockus, V. Tiesis, and A. Zilinskas, "The application of Bayesian methods for seeking the extremum," Towards Global Optim., vol. 2, nos. 117–129, p. 2, 1978. J. Kushner, "A new method of locating the maximum point of an arbitrary multipiece curve in the presence of noise," J. Basic Eng., vol. 86, no. 1, pp. 97–106, 1964.
- [26]. N. Srinivas, A. Krause, M. Seeger, and S. M. Kakade, "Gaussian process optimization in the bandit setting: No regret and experimental design," in Proc. 27th Int. Conf. Mach. Learn. (ICML), 2010, pp. 1015–1022.
- [27]. (2012). 45nm NCSU FreePDK. [Online]. Available: <http://www.si2.org>