

Application of Machine Learning for Optimization of 3-D Integrated Circuits and Systems

Sung Joo Park, *Senior Member, IEEE*, Bumhee Bae, *Member, IEEE*, Joungho Kim, *Senior Member, IEEE*, and Madhavan Swaminathan, *Fellow, IEEE*

Abstract—The 3-D integration helps improve performance and density of electronic systems. However, since electrical and thermal performance for 3-D integration is related to each other, their codesign is required. Machine learning, a promising approach in artificial intelligence, has recently shown promise for addressing engineering optimization problems. In this paper, we apply machine learning for the optimization of 3-D integrated systems where the electrical performance and thermal performance need to be analyzed together for maximizing performance. In such systems, modeling can be challenging due to the multiscale geometries involved, which increases computation time per iteration. In this paper, we show that machine learning can be applied to such systems where multiple parameters can be optimized to achieve the desired performance using the minimum number of iterations. These results have been compared with other promising optimization methods in this paper. The results show that on an average, 4.4%, 31.1%, and 6.9% improvement in temperature gradient, CPU time, and skew are possible using machine learning, as compared with other methods.

Index Terms—3-D IC, Bayesian optimization (BO), electrical-thermal simulation, machine learning, temperature gradient, thermal-induced skew.

I. INTRODUCTION

CONTINUOUS growth in operational speed and circuit density of electronic systems has resulted in new technologies to achieve these goals. Three-dimensional integration technique, an innovative technique in systems packaging, provides solutions for improving performance and density of electronic systems [1]. However, improved circuit and power density also increases the heat flux, which can increase temperature and cause thermal related reliability problems [2]. Increased temperature and their gradients can degrade electrical performance, since it can have a direct impact on clock skew. Since electrical performance and thermal performance are related through joule heating, their combined

analysis is required for accurately predicting the temperature distribution [3]. In addition, the clock tree needs to be modeled in the presence of the temperature distribution for estimating clock skew [3]. Since several parameters such as physical geometries, interface materials, fan speed, and so on have a direct influence on the temperature profile, these parameters need to be suitably tuned to achieve the desired electrical performance. Hence, this translates into 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 reaching optima due to the multiphysics and multiscale modeling required that causes an increase in computational time.

Several studies have proposed statistical methods 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 minimal and 2) the number of levels is normally limited to below three. Moreover, these methods can lead to quantization error during optimization, due to their implementation at discrete points in the problem space.

Other approaches for global optimization are also available, as discussed in [6]. As an example, global optimization algorithms typically require large computing resources due to a combination of large number of data sets and compute time for each data set. 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 comparison, 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 nonconvex, black box optimization problems as well, which was another requirement.

Machine learning has three components, task, experience, and performance, which consists of two phases, training,

Manuscript received July 21, 2016; revised November 14, 2016; accepted December 17, 2016. This work was supported by the Center for Advanced Electronics through Machine Learning.

S. J. Park was with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA. He is now with Samsung Electronics, Co., Ltd., Hwaseong 18448, South Korea (e-mail: sjoo@samsung.com).

M. Swaminathan is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: madhavan@ece.gatech.edu).

B. Bae and J. Kim are with the Department of Electrical Engineering and Computer Science, Korea Advanced Institute of Science and Technology, Daejeon 305701, South Korea (e-mail: bhbae@kaist.ac.kr; joungho@ee.kaist.ac.kr).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TVLSI.2017.2656843

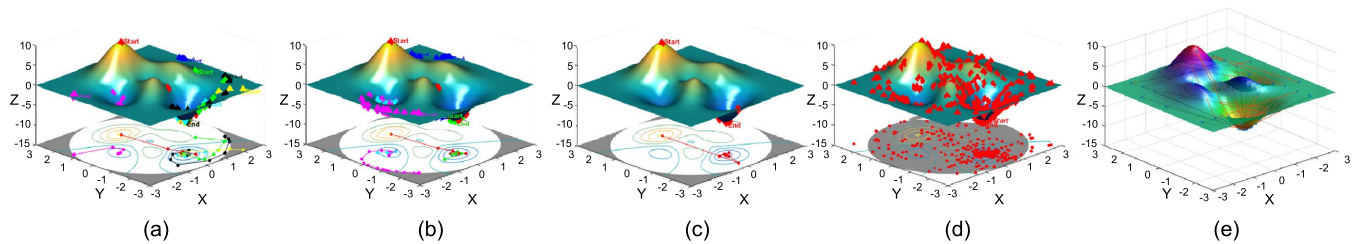


Fig. 1. Optimization methods applied to a predefined function. (a) Multistart (function counts = 345). (b) Global search (function counts = 273). (c) Pattern search (function counts = 272). (d) Genetic algorithm (function counts = 2650). (e) BO (function counts = 100) [8].

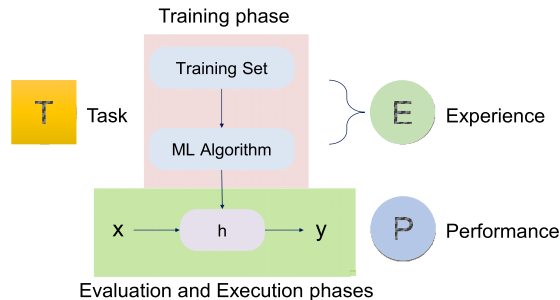


Fig. 2. Concept of machine learning consists of training and evaluation/execution phases [9], [10].

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, we apply machine learning for the optimization of 3-D ICs and systems, to minimize temperature and temperature gradients.

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: Section II describes the problem in the context of 3-D integration and discusses a test chip for validating the solver with measurements; Section III discusses system optimization using machine learning with results provided in Section IV; followed by conclusions in Section V.

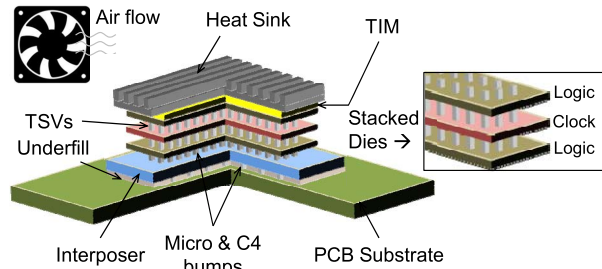


Fig. 3. Configuration of a 3-D system for optimization.

II. PROBLEM DEFINITION

A. 3-D Integrated System

Our objective in this paper is to minimize the global skew caused by temperature and temperature gradients in 3-D systems. We rely on 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 address the multiscale geometries and accounts for multiple materials in the structure related to die, interposer, and PCB by enforcing 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 using 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

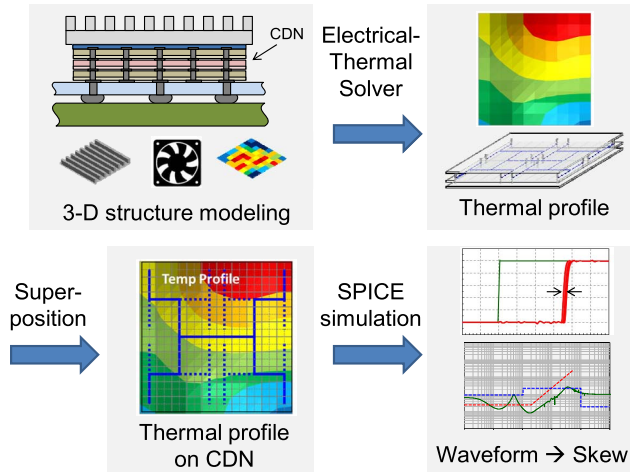


Fig. 4. Flow of electrical–thermal simulation for 3-D system design.

clock distribution network (CDN). It is important to note that temperature gradient has a significant impact on clock skew [3].

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 temperature generation and monitoring circuits. Though the test vehicle did not contain a 3-D stack with TSVs due to limited availability, the authors 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 process was used to design and fabricate the prototype, since it was part of a low-cost multiproject wafer. Since the objective was to validate the models and modeling process, the 180 nm technology node was chosen. 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).

C. Validation of Electrical–Thermal Solver

On-chip temperature gradients were measured using temperature generating and monitoring blocks. We induced variable current to each heater with resistor networks built on test board by varying resistance and input voltage. To measure the local temperature, we used temperature monitoring circuits with diodes. 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 I
FABRICATION PROCESS SPECIFICATIONS

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

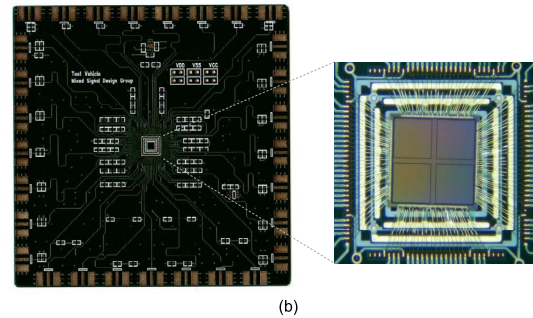
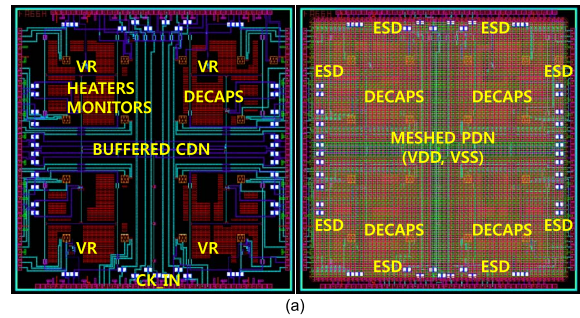


Fig. 5. (a) Chip layout. (b) Fabricated PCB and wire-bonded chip.

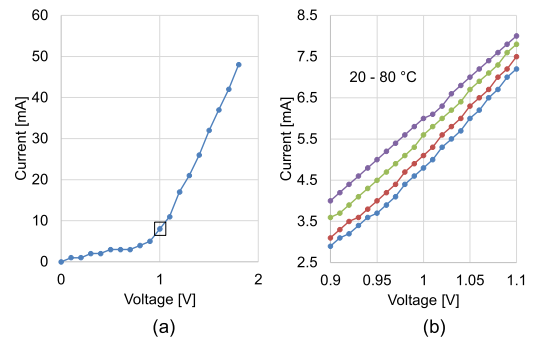


Fig. 6. (a) Measured I – V profile of temperature monitoring circuits. (b) I – V profile with temperature variations.

The power consumed by each heater was calculated using 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 $5\ \text{W}/(\text{m}^2 \cdot \text{K})$ [19], a heat transfer coefficient of $4.0\ \text{W}/(\text{m}^2 \cdot \text{K})$ 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

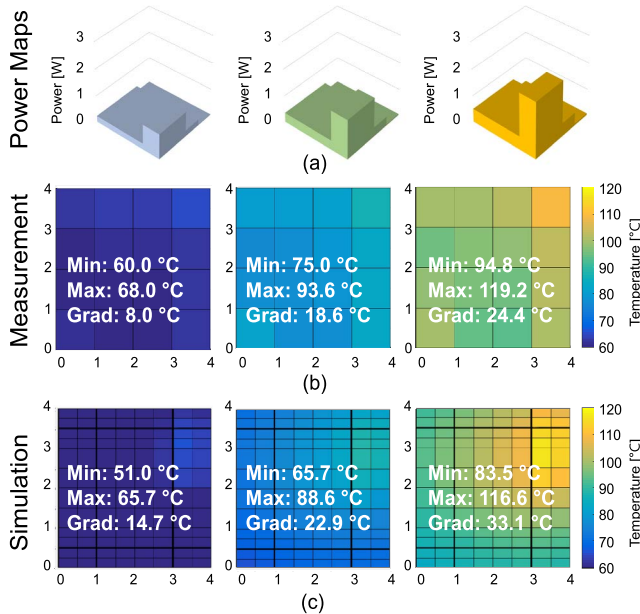


Fig. 7. (a) Power maps used for simulation and measurement. (b) Measured temperature profiles. (c) Simulated temperature profiles.

with the electrical–thermal simulations. From Fig. 7, it can be seen that the correlation is good 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 due to the smaller values involved. Nevertheless, these correlations provide a reasonable degree of confidence in the simulated temperatures, since 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. SYSTEM OPTIMIZATION

In this section, we briefly discuss the BO using GP algorithm in the context of 3-D system optimization. We also provide some basic theory and define the system parameters used as part of the optimization process.

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

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

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.

Equation (1) interprets Bayes' rule regarding possibilities of multiple events, before (prior to) and after (posterior to) event

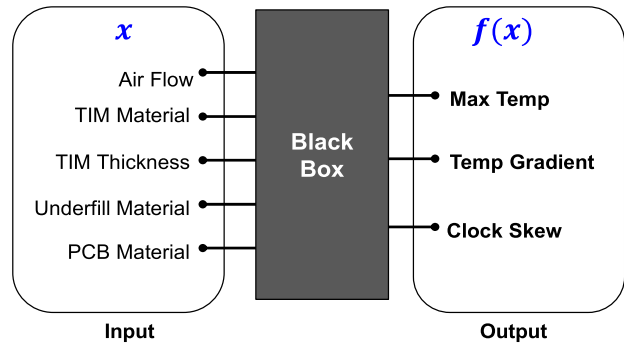


Fig. 8. Black box function with multivariable for 3-D system design.

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 ” to be optimized to achieve a target output “ $f(x)$ ” are shown in Fig. 8. The black box function $f(x)$ is obtained using the electrical–thermal solver described earlier. Based on sensitivity analysis, we picked five input variables for optimization namely, 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 $f(x)$ using weights for each of the parameters, as explained in a later section.

C. Bayesian Optimization With Gaussian Process

In Bayesian statistics, we model our uncertainty with a prior probability distribution. In other words, we estimate the distribution and use this information 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 $f(x_{1:M})$ for each variable is defined as a GP given by

$$f(x_{1:M}) = N(\mu(x_{1:M}), k) \quad (3)$$

where $x_{1:M}$ represent the “ M ” observation points for each input variable, $\mu(x_{1:M})$ is the corresponding mean vector

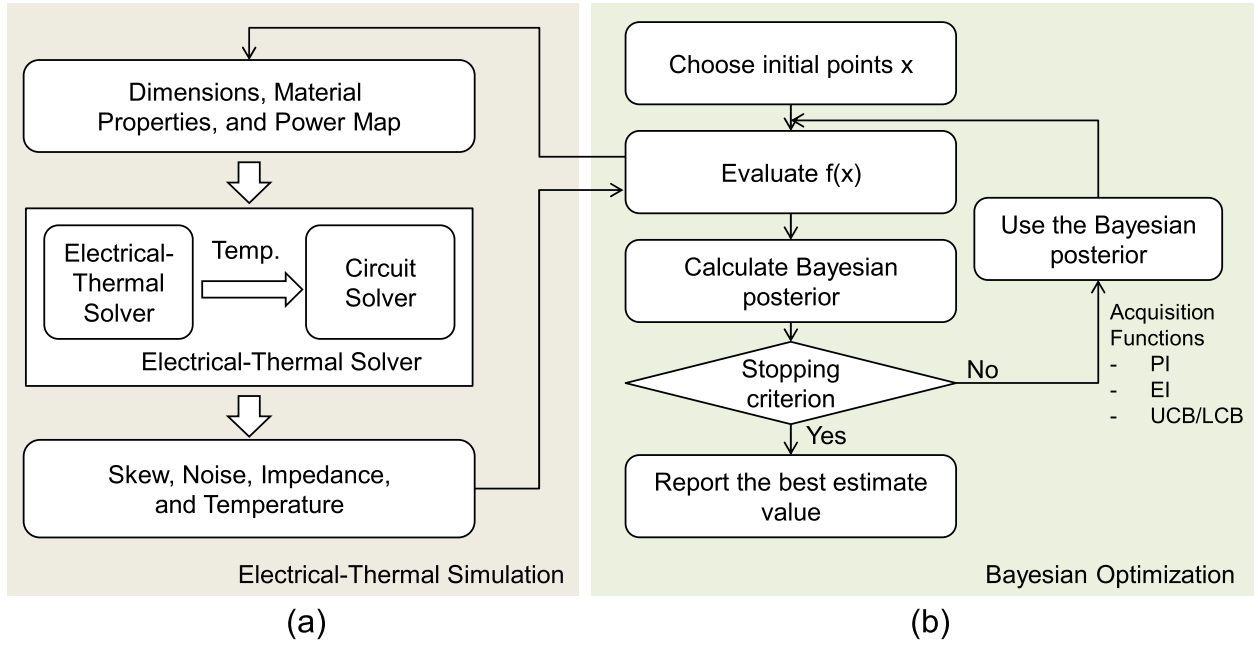


Fig. 9. Proposed flow for electrical-thermal simulation using BO. (a) Electrical-thermal simulation. (b) BO.

and k (also called the kernel) is the corresponding covariance matrix given by [21]

$$\boldsymbol{\mu}(\mathbf{x}_{1:M}) = [\boldsymbol{\mu}(\mathbf{x}_1) \boldsymbol{\mu}(\mathbf{x}_2) \cdots \boldsymbol{\mu}(\mathbf{x}_M)]^T \quad (4)$$

$$\mathbf{K}(\mathbf{x}_{1:M}) = \begin{bmatrix} \mathbf{k}(\mathbf{x}_1, \mathbf{x}_1) & \cdots & \mathbf{k}(\mathbf{x}_1, \mathbf{x}_M) \\ \vdots & \ddots & \vdots \\ \mathbf{k}(\mathbf{x}_M, \mathbf{x}_1) & \cdots & \mathbf{k}(\mathbf{x}_M, \mathbf{x}_M) \end{bmatrix} \quad (5)$$

where the covariance is defined by

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

To predict $f(\mathbf{x}_{M+1})$ 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) $f(\mathbf{x})$ subject to \mathbf{x} where $f(\mathbf{x}_{M+1})$ can be a nonconvex black-box function defined by

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

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

From [21], the mean and variance of $f(\mathbf{x}_{M+1})$ can be computed as

$$\boldsymbol{\mu}(\mathbf{x}_{M+1}) = \mathbf{k}^T \mathbf{K}^{-1} \mathbf{f}_{1:M} \quad (8)$$

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

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

This approach provides 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).

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

- 1) Choose initial points of N input variables \mathbf{x} and evaluate $\mathbf{f}(\mathbf{x})$ including error (with regard to the target value desired).
- 2) While $[\mathbf{f}(\mathbf{x}) - \text{target}] \leq \text{error}$, calculate Bayesian posterior distribution on " \mathbf{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

$$\mathbf{x}_{M+1} = \arg \min_i [\boldsymbol{\mu}(\mathbf{x}_i) - \kappa \sigma(\mathbf{x}_i)] \quad (10)$$

where $\kappa \geq 0$ and $\kappa = (2 \log \pi^2 x^2 / 12\nu)^{1/2}$, (where ν equals 0.05), and $\boldsymbol{\mu}(\mathbf{x}_i)$ and $\sigma(\mathbf{x}_i)$ 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 $\mathbf{f}(\mathbf{x})$, since (10) is computed only based on the previous results, which minimizes computational time. Since the entire

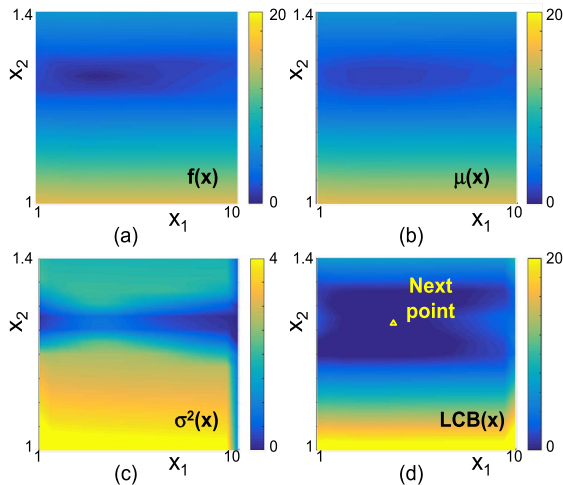


Fig. 10. Distribution plots of (a) function, (b) posterior mean, (c) posterior variance, and (d) LCB acquisition function for optimization of the 3-D system.

procedure minimizes the number of points at which $\mathbf{f}(\mathbf{x})$ 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 $\mathbf{f}(\mathbf{x})$ described in a later section.

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

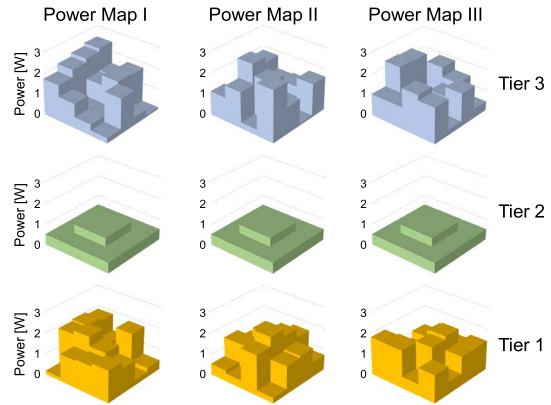


Fig. 11. Power maps used for optimization.

TABLE II
INPUT VARIABLES FOR OPTIMIZATION

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

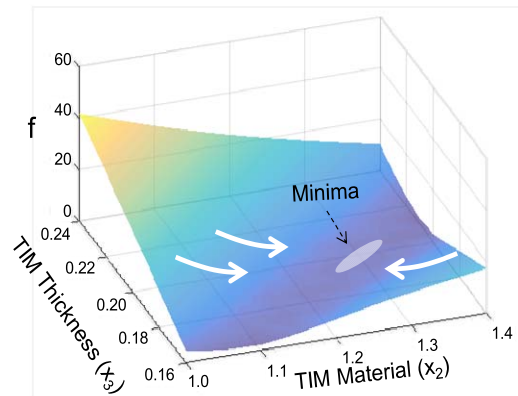


Fig. 12. Response surface with a target value.

their respective range. These parameters are air flow velocity or heat transfer coefficient, thermal conductivity of the TIM, thermal conductivity of UF material, thermal conductivity of PCB, and thickness of TIM. The range for these parameters were chosen based on manufacturability.

C. Multiobject Optimization Using Target Function

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

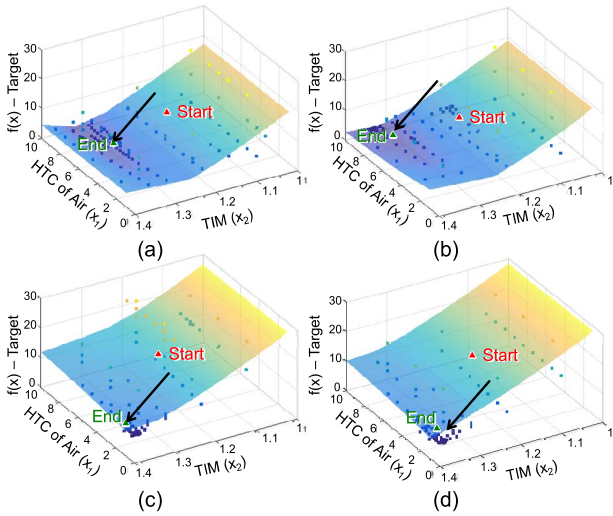


Fig. 13. Optimization results for heat transfer coefficient (X_1) and TIM thermal conductivity (X_2) showing convergence; TIM thickness (X_3) is not plotted ($N = 3$).

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

$$f(y_1, y_2) = \sum_{i=1}^2 w_i \times y_i \quad (11)$$

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

$$y_1 = \text{maximum temperature } T_{\text{MAX}}$$

and

$$y_2 = \text{temperature gradient } T_{\text{GRAD}}.$$

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).

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

TABLE III
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	0.200	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

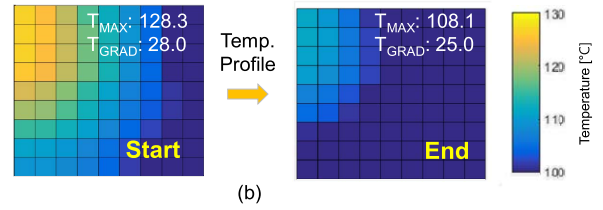
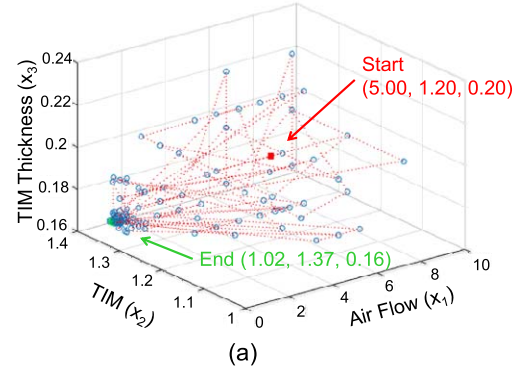


Fig. 14. Optimization results with target value of T_{MAX} : 120.0 and T_{GRAD} : 25.0. (a) Found X_s and (b) temperature.

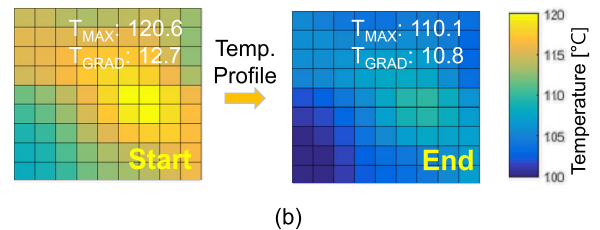
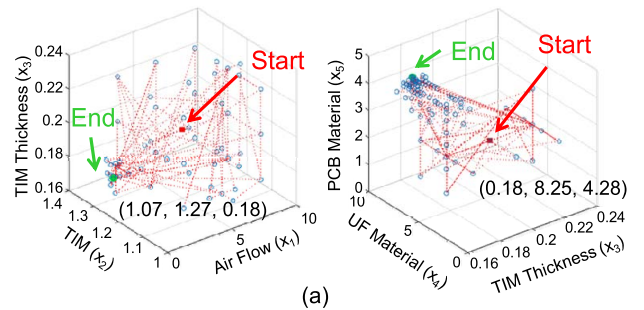


Fig. 15. Optimization with power map II. (a) Iterations shown as a function of three parameters only. (b) Temperature distribution.

value (indicated as “End”) in the figure. As can be noted from Fig. 14, the sampling is nonuniform. The maximum temperature and temperature gradient before and after optimization on the center die containing the CDN are shown in Fig. 14(b).

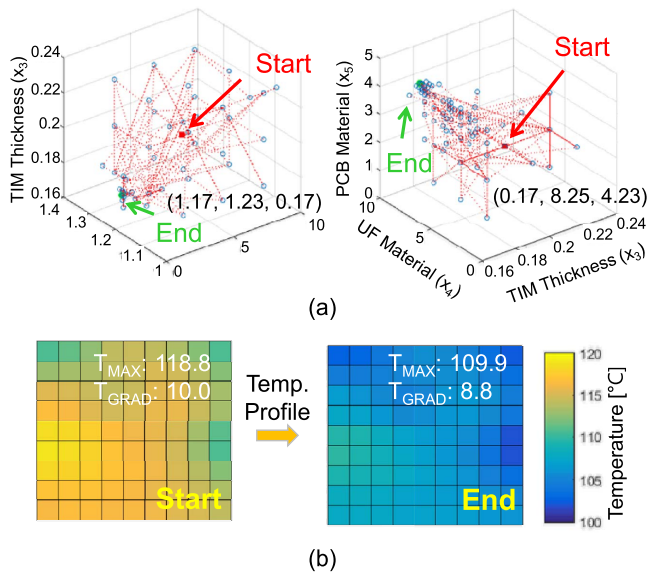


Fig. 16. Optimization with power map III. (a) Iterations shown as a function of three parameters only. (b) Temperature distribution.

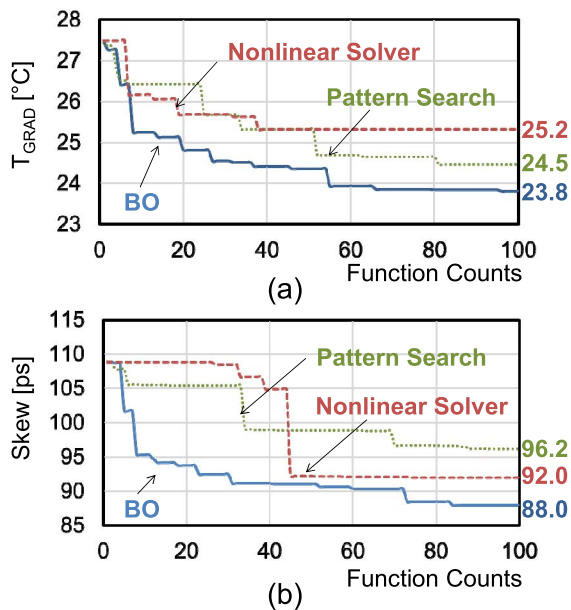


Fig. 17. Comparison of convergence between pattern search, nonlinear solver, and BO (a) temperature gradient and (b) thermal skew.

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 °C and 11 °C for power map II and 110 °C and 9 °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. After optimization, power map II and power map III resulted in a

TABLE IV
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 V
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

clock skew of 44.2 and 33.0 ps, respectively. The optimization results are shown in Table IV.

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.

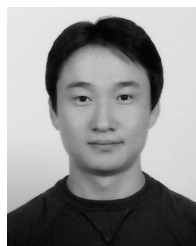
V. CONCLUSION

This paper presented machine learning combined with BO, for optimizing the electrical–thermal performance of 3-D integrated circuits and systems. Optimization results and comparison to other techniques show several advantages with the approach proposed. Our conclusion is that the method described is suitable for optimizing system-level electrical–thermal cosimulation problems, (which often require long simulation time and a large number of simulation cases), is accurate and requires lower computational cost (−31.1% as

CPU time) as compared with other conventional design optimization methods. This approach also showed the capability of handling a large number of input parameters with fast convergence and flexibility. The 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, we believe that the efficiency of the optimization described in this paper can be increased further.

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/gads_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] 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] E. 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] A. 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," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 26, no. 1, pp. 166–176, Jan. 2007.
- [14] 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.
- [15] H. Lin and P. Li, "Circuit performance classification with active learning guided sampling for support vector machines," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 34, no. 9, pp. 1467–1480, Sep. 2015.
- [16] S. Das, J. R. Doppa, P. P. Pande, and K. Chakrabarty, "Design-space 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.
- [17] 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.
- [18] 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.
- [19] J. H. Whitelaw, *Convective Heat Transfer*, accessed on Apr. 12, 2016. [Online]. Available: <http://www.thermopedia.com/content/660/>.
- [20] A. Gelman, *Bayesian Data Analysis*. Boca Raton, FL, USA: CRC Press, 2008.
- [21] E. 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>
- [22] W. B. Powell and P. I. Frazier, "Optimal learning," in *Proc. Tuts. Oper. Res. INFORMS*, 2008, pp. 213–246.
- [23] 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.
- [24] 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.
- [25] H. J. Kushner, "A new method of locating the maximum point of an arbitrary multipeak 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>



Sung Joo Park (M'09–SM'16) received the B.S. and M.S. degrees in electronics engineering from Sogang University, Seoul, South Korea, in 1998 and 2000, respectively, and the Ph.D. degree from the Georgia Institute of Technology, Atlanta, GA, USA, in 2016.

He has been a Principal Engineer with the DRAM Development and Technology Department, Samsung Electronics Company, Ltd., Hwasung, South Korea, where he is involved in the design and development of memory systems. He holds 14 issued patents in memory modules and systems and several patents pending. His current research interests include high-speed interfaces and signal, power, and thermal integrity in the designs of 3-D ICs and systems.

Dr. Park is a member of the JEDEC Solid State Technology Association, where he has been developing industry standards since 2006. He was a recipient of the JEDEC Chairman's Award in 2010, the Samsung Fellowship in 2011, and the Best Student Paper Award at the IEEE International Symposium on Electromagnetic Compatibility in 2013.



Bumhee Bae (S'11–M'15) received the B.S., M.S., and Ph.D. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2009, 2011, and 2014, respectively.

He is currently a Post-Doctoral Researcher with KAIST. He has research experience of more than five years in the field of electromagnetic noise effects on integrated circuit. He has authored or co-authored over 30 papers (journal and conference). His current research interests include SI/PI/EMC of mixed-mode system with chip-package-PCB hierarchical structures.



Joung-ho Kim (A'04–M'08–SM'14) received the B.S. and M.S. degrees in electrical engineering from Seoul National University, Seoul, South Korea, in 1984 and 1986, respectively, and the Ph.D. degree in electrical engineering from the University of Michigan, Ann Arbor, MI, USA, in 1993.

In 1994, he joined the Memory Division of Samsung Electronics, Suwon, South Korea, where he was involved in Gbit-scale DRAM design. In 1996, he joined the Korea Advanced Institute of Science and Technology, Daejeon, South Korea, where he is currently a Professor with the Electrical Engineering Department. He is also the Director of the 3-DIC Research Center supported by SK Hynix, Inc., and the Smart Automotive Electronics Research Center supported by KET, Inc. He has authored or co-authored over 404 technical papers published at refereed journals and conference proceedings, and also a book, *“Electrical Design of Through Silicon Via;”* (New York, NY, USA: Springer, 2014). He has given more than 219 invited talks and tutorials at the academia and the related industries. His current research interests include EMC modeling, design and measurement methodologies of 3-D IC, TSV, Interposer, System-in-Package, multilayer PCB, wireless power transfer technology for 3-D IC, on chip-package-PCB codesign and cosimulation for signal integrity, power integrity, ground integrity, timing integrity, radiated emission in 3-D IC, TSV, and Interposer.

Dr. Kim was a recipient of the Outstanding Academic Achievement Faculty Award of KAIST in 2006, the KAIST Grand Research Award in 2008, the National 100 Best Project Award in 2009, the KAIST International Collaboration Award in 2010, and the KAIST Grand Research Award in 2014, respectively. He received the Technology Achievement Award from the IEEE Electromagnetic Society in 2010. He was the Conference Chair of the IEEE Wireless Power Transfer Conference 2014, held at Jeju Island, South Korea, the symposium Chair of the IEEE EDAPS Symposium 2008, and the TPC Chair of the Asia-Pacific International Symposium on Electromagnetic Compatibility 2011. He was appointed as the IEEE Electromagnetic Compatibility Society Distinguished Lecturer in a period from 2009 to 2011. He served as a Guest Editor of the special issue in the IEEE TRANSACTIONS OF ELECTROMAGNETIC COMPATIBILITY for PCB level signal integrity, power integrity, and EMI/EMC in 2010, and also as a Guest Editor of the special issue in the IEEE TRANSACTIONS OF ADVANCED PACKAGING for Through-Silicon-Via in 2011. He is the TPC Member of the Electrical Performance of Electronic Packaging and System. He is also an Associated Editor of the IEEE TRANSACTIONS OF ELECTROMAGNETIC COMPATIBILITY.



Madhavan Swaminathan (M'95–SM'98–F'06) is the John Pippin Chair in Microsystems Packaging & Electromagnetics in the School of Electrical and Computer Engineering (ECE) and Director of the Center for Co-Design of Chip, Package, System (C3PS), Georgia Tech. He formerly held the position of Joseph M. Pettit Professor in Electronics in ECE and Deputy Director of the NSF Microsystems Packaging Research Center, Georgia Tech. Prior to joining Georgia Tech, he was with IBM working on packaging for supercomputers. He is the author of

450+ refereed technical publications, holds 29 patents, primary author and co-editor of 3 books, founder and co-founder of two start-up companies (E-System Design and Jacket Micro Devices) and founder of the IEEE Conference Electrical Design of Advanced Packaging and Systems (EDAPS), a premier conference sponsored by the CPMT society on Signal Integrity in the Asian Region. He is an IEEE Fellow and has served as the Distinguished Lecturer for the IEEE EMC society. He received his MS and PhD degrees in Electrical Engineering from Syracuse University in 1989 and 1991, respectively.