



ELSEVIER

Contents lists available at ScienceDirect

INTEGRATION, the VLSI journal

journal homepage: www.elsevier.com/locate/vlsi

Compact thermal modeling for packaged microprocessor design with practical power maps[☆]

Zao Liu^a, Sheldon X.-D. Tan^{a,*}, Hai Wang^b, Yingbo Hua^a, Ashish Gupta^c^a Department of Electrical Engineering, University of California, Riverside, CA, USA^b School of Microelectronics and Solid-State Electronics, University of Electronic Science and Technology of China, China^c Intel Corporation, Chandler, AZ, USA

ARTICLE INFO

Article history:

Received 19 February 2013

Received in revised form

23 July 2013

Accepted 27 July 2013

Available online 12 August 2013

Keywords:

Thermal modeling

Nonlinear modeling

Microprocessor

Package

Power map

ABSTRACT

In this paper, we propose a new behavioral thermal modeling technique for high-performance microprocessors at package level. Firstly, the new approach applies the subspace identification method with the consideration of practical power maps with correlated power signals. We show that the input power signal needs to meet an independence requirement to ensure the model predictability and propose an iterative process to build the models with given error bounds. Secondly, we show that thermal systems fundamentally are nonlinear and then propose a piecewise linear (PWL) scheme to deal with nonlinear effects. The experimental results validated the proposed method on a realistic packaged integrated system modeled by the multi-domain/physics commercial tool, COMSOL. The new piecewise linear models can model thermal behaviors over wide temperature ranges or over different thermal boundary convective conditions due to different fan speeds. Further, the PWL modeling technique can lead to much smaller model order without accuracy loss, which translates to significant savings in both the simulation time and the time required to identify the reduced models compared to the simple modeling method by using the high order models.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Temperature has become a major concern for high performance microprocessor and package design as more devices are integrated on a chip [3,4]. Thermal management and related design problems continue to be identified by the Semiconductor Industries Association Roadmap [5] as one of the five key challenges during the next decade to achieve the projected performance goals of the industry. Thus, accurate and efficient thermal modeling and analysis is vital for the thermal-aware chip and package designs to improve performance, reliability, as well as for efficient online temperature regulation and management [6–8].

The traditional bottom-up approaches including FEM (finite element), FDM (finite difference), and computational flow dynamics (CFD) based methods were widely used for thermal modeling and

analysis in the past [9–11]. For compact modeling, many existing approaches try to use thermal resistors and capacitors with fixed topology networks subject to different thermal boundary conditions [12–14]. However, the accurate RC values of elements, especially for complex geometries and boundary conditions, are difficult to determine, and the calibration against the numerical field solvers or analytical results [15,16] and measured data are usually required [17].

For thermal modeling at the architecture or package levels, many works have been proposed targeting at different applications at different abstract levels in the past. An excellent survey on recent works can be found in [18]. In general, existing approaches can be classified into bottom-up white-box approaches, which are based on the thermal dynamic physics and approximation techniques, the top-down black-box behavioral modeling approaches and the third modeling approaches, which are something in between the two approaches [18].

Existing work on HotSpot [19,8] attempts to solve this problem by generating the compact thermal model in a bottom-up manner based on processor and package structures. However, such white-box models may suffer from accuracy issues for complicated structures and boundary conditions, which are not properly modeled in the starting models. For instance, complicated package design may require exploration of packages with different structures and materials and boundary conditions for their thermal performance in the industry setting.

[☆]Some preliminary results of this paper appeared in *1st International IEEE Workshop on Thermal Modeling and Management: Chips to Data Centers, (TEMM'11)*, [1] and *Proc. Asia South Pacific Design Automation Conference (ASP-DAC13)* [2]. This work is supported in part by NSF Grant under no. CCF-1255899, in part by Semiconductor Research Corporation (SRC) Grant under no. 2013-TJ-2417 and UC Academic Senate Committee on Research (COR) Award.

* Corresponding author. Tel.: +1 951 827 5143.

E-mail addresses: zliu008@ucr.edu (Z. Liu), stan@ee.ucr.edu (S.X.-D. Tan), wanghai@uestc.edu.cn (H. Wang), yhua@ee.ucr.edu (Y. Hua), ashish.x.gupta@intel.com (A. Gupta).

Recently, top-down black-box behavioral thermal modeling methods have been proposed. Many proposed models target at the on-line temperature regulation applications, in which compact thermal models can estimate or predict the thermal behaviors of a real systems and they can be built in a dynamic way based on the online thermal sensor reading and online power estimation techniques [20–22]. Recently a distributed online thermal model was proposed and validated on a realistic many-core system [23–25].

Another important development for the top-down black-box thermal modeling is for building more accurate and even parameterized thermal models for architecture or package level thermal-aware design and optimization. The input power signal and output thermal temperatures for learning or training are assumed to be measured from off-line complicated equipments in the lab. Existing works consist of the matrix pencil method [26] and the subspace identification method [27,1,2]. The major advantage of such pure behavioral modeling methods is their flexibility and easy to use as no physical restrictions and assumptions are made or required for the models. They are also very accurate as the training is based on measured data. However stability and other model properties of thermal systems need to be enforced explicitly.

Recently, Beneventi et al. [18] proposed a hybrid (or *gray-box*) identification method, in which a pre-structured compact model under physical constraints is built via an optimization approach. The main advantage of such models is that many physical properties such as stability and passivity can be satisfied automatically. But such models will be less flexible for different architectures and structures as the thermal models or topologies are based on specific architectures. Also, all the existing thermal behavioral modeling methods assume that the thermal systems are linear, which may not be the case for many practical thermal systems as we show in this work, and they have difficulty to deal with varying thermal boundary conditions.

In this work, we still focus on the black-box based thermal modeling scheme based on the subspace identification method. We consider practical measured power maps, which can be obtained from thermal lab based on the test thermal vehicle (testing thermal chips). We first observe that the subspace identification method may suffer from the lack of predictability problems in general [28,29], especially when the input power is given as series of 2-dimensional power distributions (called *power maps*) in which the input signal is highly spatially correlated. Power map-based thermal characterization is widely used in industry for thermal characterizations of package design as power maps can be easily obtained (measured or computed) practically. However, the spatially correlated power signals in the power map make the system identification process more difficult. The reason is that it is difficult to distinguish the contribution from specific input when all the inputs have the same or similar transient waveforms. Also compact behavioral thermal modeling for changing boundary conditions still remain to be a challenging problem.

In this paper, we present a new behavioral thermal modeling technique considering more practical power maps and nonlinear effects of thermal systems for package level design space exploration of high-performance microprocessors. The new approach consists of two major contributions or improvements over existing approaches:

- First, we observe that the subspace identification method may suffer predictability problem when power maps are given where power inputs are spatially correlated. For instance, the busy ALU will be likely to have frequent memory accesses and many instruction fetching activities. As a result, the corresponding function units will have power increases or decreases at the same or similar times. Such correlated input signals pose

difficulty for the subspace identification method and will easily lead to loss of predictability as it is more difficult to distinguish the contributions from specific inputs when all the inputs have the same or similar transient waveforms. In this paper, we show that the input power signal needs to meet some independence requirements to ensure model predictability (rank of input power maps or their power signal matrix needs to meet certain requirements). A new algorithm, *ThermSubCP*, can generate independent power maps to meet the spatial rank requirement and can also automatically select the order of the resulting thermal models for the given error bounds.

- Second, we show that thermal systems are fundamentally nonlinear. One important example is that thermal conductivities of silicon and package materials are temperature dependent. Another example is the changing thermal boundary conditions due to different fan speeds. To mitigate this problem, we apply the piecewise linear (PWL) scheme to characterize the nonlinear thermal behavior under those conditions. Our experiments show that the nonlinear effects in the thermal systems are typically mild and weak but are still significant enough to warrant the PWL modeling. However, nonlinearity due to boundary conditions can be very significant. PWL can deal with both mild and hard nonlinearities. We observe that the PWL method can lead to smaller models and reduced modeling costs compared to high order model approximation. This is important as the costs of identifying and simulating the reduced models will grow at least quadratically, it is critical to reduce the model order to maintain the efficiency gain from the reduced order modeling. The new modeling algorithm, *ThermSubPWL*, partitions the nonlinear ranges (due to temperature or boundary condition changes) into a number of small ranges and performs the modeling on each range using the previously proposed *ThermSubCP* method. A linear transformation method, which avoids the existing multi-transition requirement, is proposed to transform the identified linear local-models to the common state basis to build the continuous piecewise linear model. To the best knowledge of the authors, the proposed method is the first work addressing the nonlinear thermal modeling problem.

Experimental results validate the proposed method on a realistic packaged integrated system modeled by the multi-domain/physics commercial tool, COMSOL, under practical power signal inputs. The new piecewise linear models can model thermal behavioral over wide temperature ranges and different thermal boundary convective conditions due to different fan speeds. Further, the PWL modeling technique can lead to much smaller model order without accuracy loss, which translates to significant simulation time reduction and about $10 \times$ less time to identify the reduced models compared to the simple modeling method by using the high order models in our examples.

The rest of this paper is organized as follows: [Section 2](#) presents the thermal modeling problem we are trying to solve. [Section 3](#) reviews the subspace identification method. [Section 4](#) presents the new power-map based thermal modeling technique, and [Section 5](#) gives the new nonlinear thermal modeling technique. Finally, [Section 6](#) presents the experimental results of both *ThermSubCP* and *ThermSubPWL*, with [Section 7](#) concluding the paper.

2. Thermal modeling problem considering power maps

We first present how the power inputs are modeled in our problem. A microprocessor chip is partitioned into $p = n \times m$ power grids as shown in [Fig. 1](#), where each square power grid has a power source as an input and the measured temperatures at

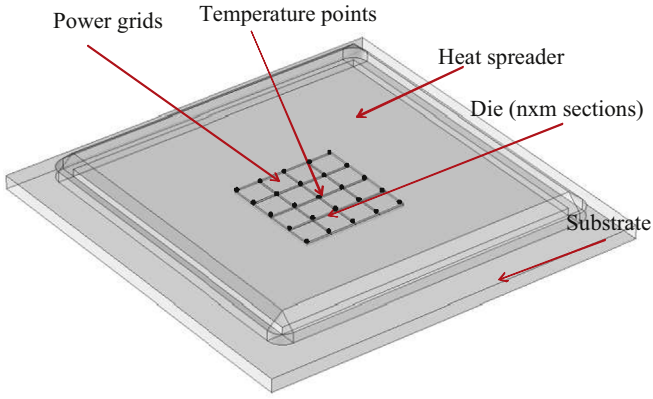


Fig. 1. Meshed chip and package.

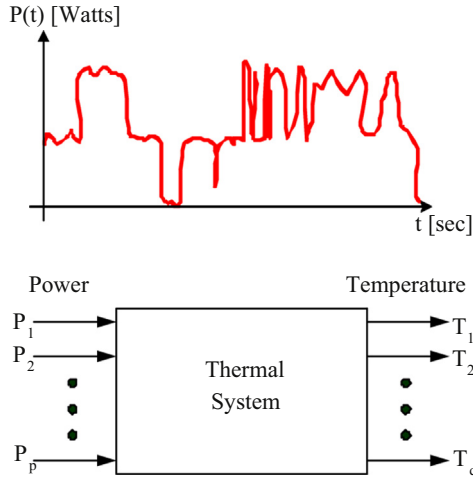


Fig. 2. The abstracted model system and correlated power inputs.

its adjacent 4 corners as outputs. We can abstract this power grid model into a discrete multi-input and multi-output (MIMO) system with $p = n \times m$ power inputs and q temperature outputs as shown in Fig. 2. The $n \times m$ power input distribution at one time instance is defined as a *power map*, which can be measured or obtained by simulations or other methods practically.

In general, the abstracted p -input and q -output thermal system could be represented as

$$\begin{aligned} x(t+1) &= F(x(t), u(t)), \\ y(t) &= G(x(t), u(t)), \end{aligned} \quad (1)$$

where both $F(x)$ and $G(x)$ are nonlinear vector functions of state variable vector $x(t)$ and input signal vector $u(t)$. In our problem, the input vectors $u(t) \in \mathbb{R}^{kp}$ are the measured power input traces and output vectors $y(t) \in \mathbb{R}^q$ are the temperature responses.

Existing approaches typically assume that thermal system in Fig. 1 is linear. As a result, (1) can be rewritten as the standard linear state transition form

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t), \\ y(t) &= Cx(t) + Du(t), \end{aligned} \quad (2)$$

where $A \in \mathbb{R}^{l \times l}$ is a stable matrix, l is the number of states. $B \in \mathbb{R}^{l \times p}$, $C \in \mathbb{R}^{q \times l}$, and $D \in \mathbb{R}^{q \times p}$. With p input $u(t_i)$ and q outputs $y(t_i)$ where $i = 1, 2, \dots, s$, the problem now becomes finding state matrices A , B , C , and D , where D is typically considered as a matrix of zeros. $x(t)$ are the Kalman state vector and it does not have different physical meaning.

Also, the $n \times m$ power inputs may be highly correlated as mentioned before. In an extreme case, all the power input

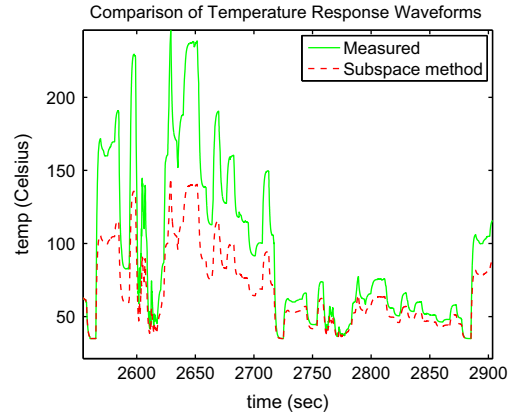


Fig. 3. Transient temperature response of the system identified with one highly correlated power inputs.

waveforms are exactly the same and they are only different in magnitude. Fig. 2 (top) shows a typical power input waveforms. Their spatial difference in magnitude essentially is described by the *power map* of the chips. The magnitude (power map) distribution can be defined by a function and applied in a practical setting to an artificial testing package (called *testing vehicle*). Such magnitude distribution is called *power map configuration* in this paper.

Such highly correlated power inputs, however, will lead to poor predictability when using the subspace identification method [30,29]. Fig. 3 shows the waveforms produced by the model identified with highly correlated power inputs (one power map configuration), where the results from model and from original temperature do not match well. To overcome this problem, subspace identification procedure using multiple power map configurations is proposed to identify the thermal package model [1], and it works well when the system is linear and can be described by (2).

The second issue we are facing is that thermal behavior of packaged electronic systems will show weakly nonlinear effects due to the temperature-dependent properties of the packaging materials [31]. Fig. 4 shows the temperature dependence of thermal conductivity of Si and Cu. Fig. 5 shows if we excite a chip package system shown in Fig. 1 with a sinusoid power input, we can clearly observe the harmonic components in the discrete Fourier transform of the transient temperature response, which evidently indicates the nonlinearity of the underlying thermal system although the nonlinear components are mild and weak. Such mild nonlinear behaviors, however, can still lead to significant loss of accuracy as shown in Fig. 6 if typical low order is used. In addition, the potential change of external cooling condition need to be modeled as variable thermal resistance, and its variation possibly could lead to even stronger nonlinearity, making pure linear modeling approach more difficult.

To mitigate this problem, we propose to use linear models to represent the thermal behavior of the packaged electronic systems under different ranges of thermal conditions (piecewise linear model approach). As a result, significant accuracy improvement is achieved by using just low order models.

3. Review of subspace methods for system identification

Given input $u(t)$ and output $y(t)$, subspace identification method identifies the state matrices A , B , C , and D of (2). The subspace identification basically tries to first identify the system states (Kalman states), then the state matrices will be obtained by the least square based optimization method [29]. There are several implementations such as the Ho-Kalman' method, the MOSEP

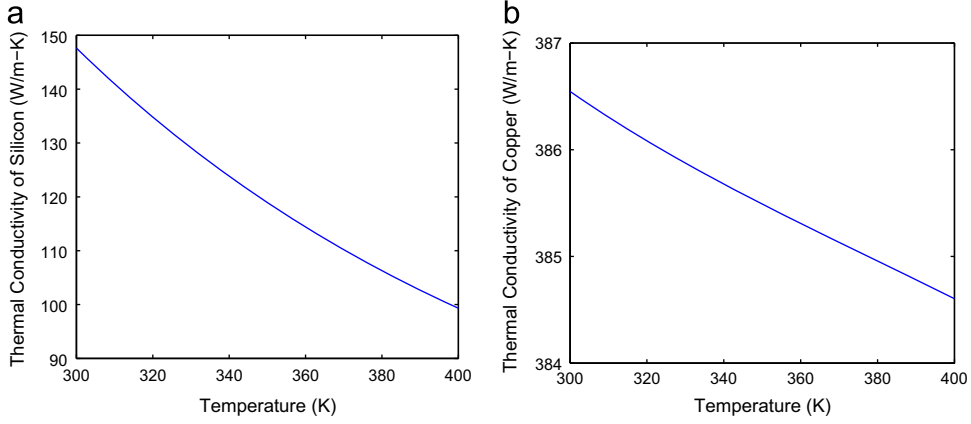


Fig. 4. Temperature dependence of the thermal conductance (a) silicon, (b) copper.

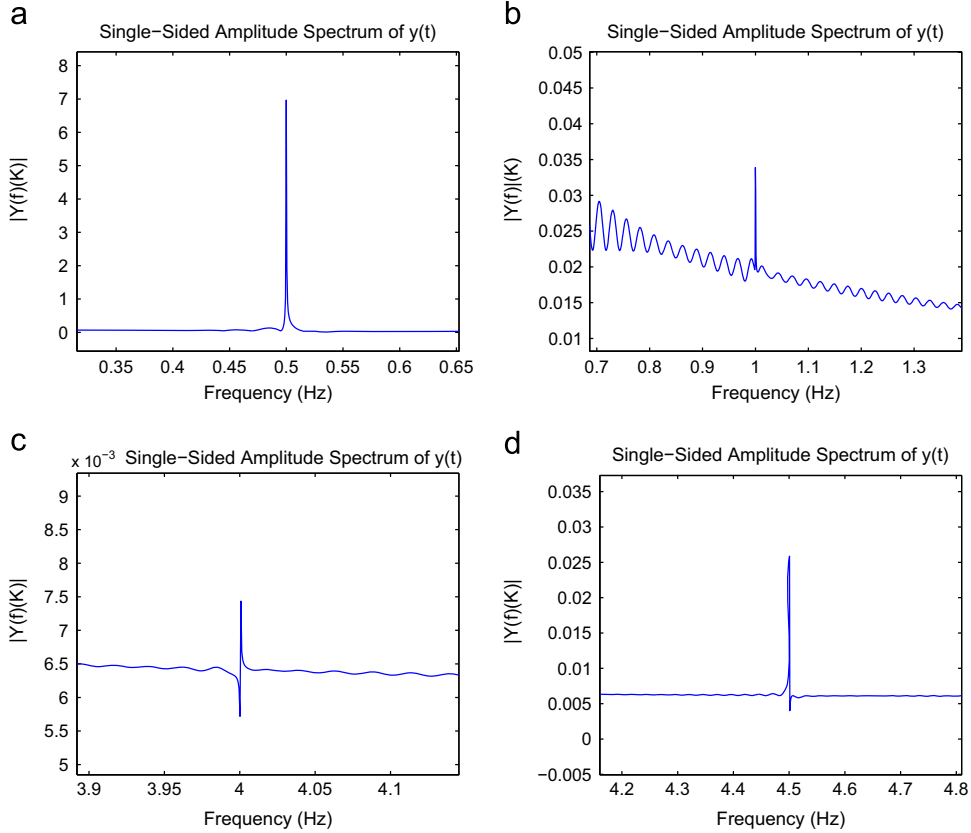


Fig. 5. Frequency domain response of the thermal system under sinusoid input with frequency of 0.5 Hz (a) baseband spectral, (b) 1st order harmonics, (c) 2nd order harmonics and (d) 3rd order harmonics.

method and the N4SID (numerical algorithms for subspace system identification) method [28]. In this paper, we apply the widely used N4SID method for this system identification problem.

3.1. Method flow of N4SID

Defining the input Hankel matrix of an l -order system as

$$U_{a|b} := \begin{bmatrix} u(a) & u(a+1) & \cdots & u(a+N-1) \\ u(a+1) & u(a+2) & \cdots & u(a+N) \\ \vdots & \vdots & \ddots & \vdots \\ u(b) & u(b+1) & \cdots & u(b+N-1) \end{bmatrix} \in \mathbb{R}^{(b-a+1)p \times N}, \quad (3)$$

and the output Hankel matrix $Y_{a|b}$ is defined accordingly. The state sequence is defined according to a given number a and the

arbitrary number N as

$$X(a) := [x(a), x(a+1), \dots, x(a+N-1)] \in \mathbb{R}^{l \times N}. \quad (4)$$

Based on the previous definition, the *past* input, output Hankel matrices and state sequence are defined as

$$U_p := U_{0|k-1}, \quad Y_p := Y_{0|k-1}, \quad X_p := X(0), \quad (5)$$

and the *future* input, output Hankel matrices and state sequence as

$$U_f := U_{k|2k-1}, \quad Y_f := Y_{k|2k-1}, \quad X_f := X(k), \quad (6)$$

The *past* data matrix and *future* data matrix now are defined as

$$W_p := \begin{bmatrix} U_p \\ Y_p \end{bmatrix}, \quad W_f := \begin{bmatrix} U_f \\ Y_f \end{bmatrix}. \quad (7)$$

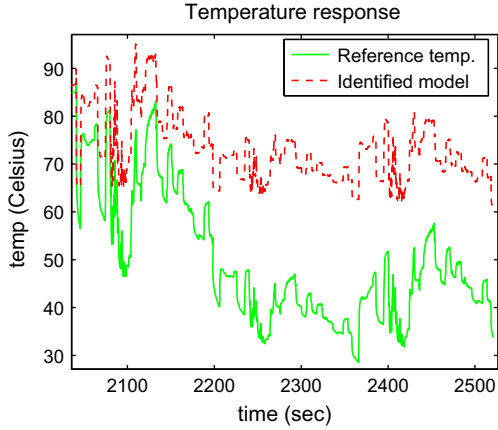


Fig. 6. Accuracy loss of the temperature response of the identified 4th order linear model.

The numbers k and N , which determine the row and column size of the input and output Hankel matrices, are determined by the user according to the number of input samples available along the time axis. Also, $k > l$ should be satisfied given that l is the order of the system.

Additionally, the extended observability matrix \mathcal{O}_k is defined as follows:

$$\mathcal{O}_k := \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{k-1} \end{bmatrix} \in \mathbb{R}^{kq \times l}, \quad (8)$$

where q is the number of output ports.

In N4SID algorithm, an important property which can be proved is

$$\mathcal{P}_{U_f}(Y_f, W_p) = \mathcal{O}_k X_f, \quad (9)$$

where $\mathcal{P}_B(A, C)$ represents an oblique projection of the row space of A onto the row space of C along row space of B [28–30].

By applying Singular Value Decomposition (SVD) on the left hand side of (9), there is

$$\mathcal{P}_{U_f}(Y_f, W_p) = [U_1 \ U_2] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} = U_1 \Sigma_1 V_1^T. \quad (10)$$

From (9) and (10), the extended observability matrix \mathcal{O}_k and the future state sequence X_f are readily identified as

$$\mathcal{O}_k = U_1 \Sigma_1^{1/2}, \quad (11)$$

$$X_f = \Sigma_1^{1/2} V_1^T. \quad (12)$$

Now the state sequence $X_f = X(k) = [x(k), x(k+1), \dots, x(k+N-1)]$ is identified, we can proceed to determine the system matrices A, B, C, D . Specifically, define the following $(N-1)$ -column matrices (compared to the previously defined N -column matrices) as

$$\bar{X}_{k+1} := [x(k+1), x(k+2), \dots, x(k+N-1)], \quad (13)$$

$$\bar{X}_k := [x(k), x(k+1), \dots, x(k+N-2)], \quad (14)$$

$$\bar{U}_{k|k} := [u(k), u(k+1), \dots, u(k+N-2)], \quad (15)$$

$$\bar{Y}_{k|k} := [y(k), y(k+1), \dots, y(k+N-2)]. \quad (16)$$

All the four matrices are known or identified already. Then, the system matrices are solved as a least square problem directly from

$$\begin{bmatrix} \bar{X}_{k+1} \\ \bar{Y}_{k|k} \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} \bar{X}_k \\ \bar{U}_{k|k} \end{bmatrix}. \quad (17)$$

3.2. Persistently exciting condition of subspace identification method

The quality of input signal is very important in subspace methods. In order to get a good system, the input signal should satisfy the so-called *persistently exciting* condition. That is, for an l -order system and $l < k$, $\text{rank}(U_{0|2k-1}) = 2pk$ for a p -port system, assuming that the number of columns of the Hankel matrix N is sufficiently large in the N4SID method [29,30].

Generally speaking, the input signal is *qualified* if a unique system can be identified from the given input and output data. This can be easily explained by following the FIR (finite impulse response) example

$$y(t) = \sum_{i=0}^{2k-1} q_i u(t-i), \quad (18)$$

where $q_0, q_1, \dots, q_{2k-1} \in \mathbb{R}^{p \times p}$ are the system impulse responses. Denote

$$Q_{2k-1} := [q_{2k-1}, q_{2k-2}, \dots, q_0], \quad (19)$$

and the system identification comes down to determining Q_{2k-1} using the input and output data. From (18), we readily get

$$Y_{2k-1|2k-1} = Q_{2k-1} U_{0|2k-1}, \quad (20)$$

where $U_{0|2k-1} \in \mathbb{R}^{2pk \times N}$. This is a least square problem and Q_{2k-1} can be uniquely solved when $U_{0|2k-1}$ has full row rank [32].

$$\text{rank}(U_{0|2k-1}) = 2pk. \quad (21)$$

4. Power map based thermal modeling method for correlated power inputs

4.1. Spatial rank requirement

In general, the *persistently exciting*, or PE condition can be easily satisfied for a MIMO dynamic system when all the transient input signals are uncorrelated spatially. However, if those signals are highly correlated spatially as in the case of power maps obtained from the measurement, the PE condition may not be easily satisfied, which leads to poor predictability of the resulting models as shown in Fig. 3.

Consider a 2-input system as an example. We assume that all the inputs have exactly the same time domain waveform and denote it as $f(t)$. The differences in magnitude are represented by another spatial function $g(x)$ in 1-D space (x -axis) for simplicity, which represents the 1-D power map configuration. The i -th input sample for such 2-input system is

$$u(i) = \begin{bmatrix} g(x_0)f(t_i) \\ g(x_1)f(t_i) \end{bmatrix}. \quad (22)$$

We further define the i -th block row in the input Hankel matrix as

$$U_i = \begin{bmatrix} g(x_0)f(t_i) & g(x_0)f(t_{i+1}) & \dots & g(x_0)f(t_{i+N-1}) \\ g(x_1)f(t_i) & g(x_1)f(t_{i+1}) & \dots & g(x_1)f(t_{i+N-1}) \end{bmatrix}. \quad (23)$$

The input Hankel matrix can be expressed as

$$U_{0|2k-1} = \begin{bmatrix} U_0 \\ U_1 \\ \vdots \\ U_{2k-1} \end{bmatrix}. \quad (24)$$

The *persistently exciting* condition is satisfied when $U_{0|2k-1}$ has full row rank that is $\text{rank}(U_{0|2k-1}) = 4k$ for this 2-input case. However,

it is clear that the two rows in U_i are linearly dependent, which makes $\text{rank}(U_{0:2k-1}) = 2k$ and fails to satisfy the *persistently exciting* condition.

In order to make the input Hankel matrix $U_{0:2k-1}$ full row rank, we need to make the i -th block row U_i full row rank, assuming $N \gg k$. For this 2-input example case, we can achieve this by simply introducing another power map configuration. Now we have two configurations, g_1 and g_2 . The i -th block row U_i is shown in (25), where $i < m < i + N - 1$. The two dimensional case can be generalized into higher dimensions with g_i as a function of two spatial variables x and y . By calling the row rank of i -th block row U_i the *spatial rank* of input signals and assuming that we have sufficient samples (N is large enough), we have the following proposition:

$$U_i = \begin{bmatrix} g_1(x_0)f(t_i), \dots, g_1(x_0)f(t_{i+m}), g_2(x_0)f(t_{i+m+1}), \dots, g_2(x_0)f(t_{i+N-1}) \\ g_1(x_1)f(t_i), \dots, g_1(x_1)f(t_{i+m}), g_2(x_1)f(t_{i+m+1}), \dots, g_2(x_1)f(t_{i+N-1}) \end{bmatrix} \quad (25)$$

Proposition 1. For p -input MIMO dynamic systems with correlated input signals, the spatial rank of input signal matrix, U_i , must be equal to p to satisfy the *persistently exciting* condition in the subspace identification method.

We remark that for real microprocessor with package, if we can feed the microprocessors with different programs and those programs may generate different power maps to meet the PE conditions. For thermal characterization based on *testing vehicle*, we have the flexibility to generate different power maps on the tested chip. In this case, we propose to automatically generate multiple *independent* power map configurations to meet such spatial rank requirement for input signals, in the next part, we show how this can be achieved.

4.2. Orthogonal set of power map configurations

In this subsection, we show how to automatically generate independent power map configurations to meet the PE requirement as mentioned in the previous subsection. This process is useful since the number of inputs can be large.

Take the 1-D example again, if $x \in [0, L]$, a set of orthogonal functions over the interval $[0, L]$ can serve as the systematic solution for the independent and robust configurations. The orthogonal function set $\{g_1, g_2, \dots, g_N\}$ satisfies

$$\int_0^L g_i(x)g_j(x) dx = \begin{cases} 0 & i \neq j \\ \|g_i(x)\|^2 & i = j \end{cases} \quad (26)$$

Note that one is free to choose any set of orthogonal functions for automatic power map generation. In this paper, We arbitrarily choose orthogonal power map configurations generated by the 2-D function set

$$g_{mn}(x, y) = \sin(m\pi x/L_x) \sin(n\pi y/L_y), \quad (27)$$

in which m and n are the indices starting from 1; x and y are the position variables; L_x and L_y are the size of the chip in the direction of x and y axis respectively.

We remark that users may not have the luxury to generate the orthogonal spatial functions as they are limited by specific functional logics in practical chips or artificial power patterns in testing chips. The user could use any power map (spatial configurations) as long as it satisfies the input rank requirement. Nevertheless, the proposed power map generation method can provide some general guidelines for practical model generations.

ALGORITHM: THERMSUBCP

Input: p power map configurations (power inputs and output responses)

Output: thermal model with proper order to meet the error bound

1. Start with order one and use the p model identification configurations to generate thermal models.
2. Use the validation configurations to generate the output of the subspace model.
3. Compute the average error for each output.
4. If the error criteria is not satisfied, increase the model order and goto step 1. Otherwise, return the models obtained so far and stop.

Fig. 7. The new ThermSubCP algorithm.

4.3. Thermal modeling algorithm with automatic order selection – ThermSubCP

Now we are ready to introduce our linear thermal modeling algorithm considering the highly correlated power inputs – *ThermSubCP*, which stands for Thermal modeling using the Subspace Identification method for Correlated Power maps.

Once we generate all the independent power map configurations, we need to generate two transient power sequences – one for model identification and one for validation. For each power map configuration, we basically divide the given transient power input waveform into two parts in this approach. The first part will be used for the model identification and the second part will be used for the validation. To test the predictability of the models, we will also add some additional power maps, which are not used in model identification. For instance, suppose we have a 4-input MIMO system, then we need 4 independent power maps with transient power inputs denoted as P_1, P_2, P_3, P_4 . Then we split $P_1 = [P_{11}, P_{12}]$ into two parts in time scale. We do the same thing for the other 3 power inputs. Then the identification sequence will be $[P_{11}, P_{21}, P_{31}, P_{41}]$, while the validation sequence will be $[P_{12}, P_{22}, P_{32}, P_{42}, P_{a1}, P_{a2}]$, where P_{a1}, P_{a2} are the additional power inputs in power maps not used for identification.

ThermSubCP method also tries to automatically select the proper order of the models to satisfy the given error bounds by gradually increasing the order of the models until the accuracy in the validation phase is met. In our implementation, we use the maximum of the mean errors over all the transient responses for all the outputs as the error criteria. The proposed ThermSubCP flow is presented in Fig. 7.

5. Piecewise linear thermal modeling approach – ThermSubPWL

The second issue we try to mitigate for thermal modeling is to deal with nonlinear effects of packaged microprocessor thermal systems. In this section, we present the new nonlinear thermal modeling technique, *ThermSubPWL*, on top of the linear modeling techniques we have discussed.

5.1. Local models for partitioned linear ranges

As shown in Fig. 5, thermal systems for packaged microprocessors show weakly nonlinearity, which may come from

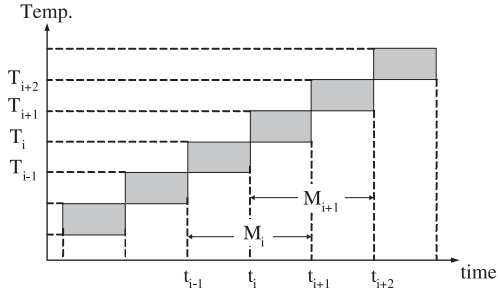


Fig. 8. Identification of linear subsystems for different temperature ranges.

temperature dependent material properties, as well as different thermal boundary conditions such as variant convective coefficients for heat sinks cooled by fans with different speeds. Practically, we may have to build thermal models for different temperature ranges and boundary conditions, which is very important modeling objective for many thermal component modeling problems [17,16]. In this case, the thermal systems become naturally nonlinear as the temperature dependent material properties will lead to variable thermal system matrices, and the variant boundary conditions will lead to variable thermal resistances to the ambient for the thermal circuits.

If we still use the linear models to characterize the system, we observed that we have to use higher order to get good approximation since the nonlinearity is mild in many cases. Of course, such approximate models will not show any nonlinear effects, but they will use more poles or states to emulate the effects of nonlinearity on thermal responses of those systems. So we will end up with much higher orders for the thermal models even for linear approximation, which will hurt the performance of the thermal analysis. In addition, such analysis also loses the nonlinear effects of the original systems.

To mitigate this problem and reduce the model order and modeling errors, we propose to partition the temperature range or each fixed boundary conditions into a number of sub-ranges and then we build the state space linear models for each sub-ranges by the *ThermSubCP* method, and these local models are then used to build piecewise linear thermal model for the whole thermal system.

In the following, we build PWL model for different temperature ranges as a driving example. Similar strategy can be apply to build PWL model for different boundary conditions, with results shown in the experimental result section.

We notice that one issue with such a piecewise linear thermal modeling scheme is that temperature is location dependent across a whole package. There may be temperature gradients among different locations. To mitigate this problem, we use the average temperature of any instance time to guide model switching. The thermal gradients in a well-designed chip are typically well managed and reduced by the online thermal management techniques [33,20]. Even with some degrees of thermal gradients, the local models should still be valid as it is a localized model and should be valid for a temperature range as long as the average temperature is still representative for the overall temperature of the processor die.

In order to obtain local models for different temperature ranges, we use stair-like power-temperature training sets to identify these models as shown in Fig. 8. For example, the model M_i is identified during time interval $[t_{i-1}, t_{i+1}]$, which corresponds to the temperature range of $[T_{i-1}, T_{i+1}]$. Since model M_i is identified with the temperature data ranging from $[T_{i-1}, T_{i+1}]$, the correct using of the subspace identification method guarantees that the identified model is valid for this temperature range.

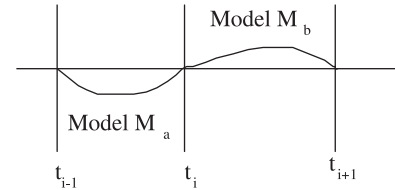


Fig. 9. Abrupt model transition at known time instance.

In order to avoid the predictability issue and improve the accuracy of the subspace identification method, we use independent power map configurations as given by (27) to identify each local model for the corresponding temperature range.

By using the stair-like input–output data, the linear models of the subsystems in different temperature ranges could be accurately identified via the *ThermSubCP* method.

Note that, all the pairs of the two adjacent models, like M_i and M_{i+1} , are identified with a shared portion of data, which makes both models valid for the same temperature range, like $[T_i, T_{i+1}]$ shown in Fig. 8. The reason is that the transition from one thermal model to another thermal model is gradual, and this shared portion can facilitate determination of model transformation matrices as will be discussed below.

5.2. Determination of model transitions

The linear models for each subsystem could not be directly combined to build the piecewise linear model for the thermal system of the microprocessor package because these identified models are not built on the same state variable basis. Hence, it is desirable to have a common state basis for all the local models instead of using different states for different models. This requires linear transformation to transform all these models onto the same basis. In [34], the transitions are assumed to be known at each time instance. Assuming that the model transition is abrupt at the transition time instance t_i as shown in Fig. 9, it can be proved that the state of model M_a and state of model M_b are differed by a linear transformation T_{ba} as

$$x_{M_b}(t_i) = T_{ba}x_{M_a}(t_i), \quad (28)$$

where $x_{M_a}(t_i)$ is the state of model M_a at the transition time instance t_i , and $x_{M_b}(t_i)$ is the state of model M_b at the same transition time. Hence, to determine the linear transformation matrix T_{ba} , multiple transitions are required to solve the linear equations (29) in the sense of least squares as shown in [34].

$$[x_{M_b}(t_1), x_{M_b}(t_2), \dots] = T_{ba}[x_{M_a}(t_1), x_{M_a}(t_2), \dots]. \quad (29)$$

However, in our thermal system modeling, if we have specific temperature value for transitions between two models, we have to excite the states of the two models such that we have many independent states from the two models and transitions happens between the two models with those states. This will lead to much larger training tasks for the modeling process.

To mitigate this problem, we propose a *transition region* concept in this paper. We observe that the temperature transition from one model to another model is in general a gradual process as indicated in the time interval from t_i to t_{i+1} shown in Fig. 10, instead of an abrupt one that happens at a specific time instance. We define a transition region as shown in Fig. 10 in which both local models are valid (in other words, a state of model M_a will switch to a corresponding state of model M_b at any given time in this region).

As discussed before, the subspace identification method guarantees that any two adjacent models are valid for a portion of shared data sets from t_i to t_{i+1} (even though in the simulation we

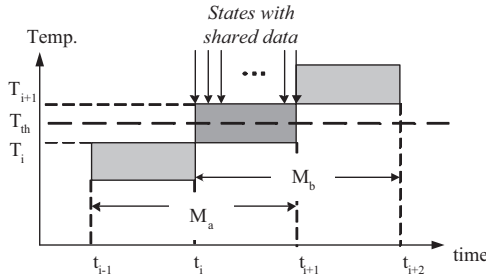


Fig. 10. Model transition from M_a to M_b .

specify arbitrary T_{th} to determine the abrupt transition from M_a to M_b as shown in Fig. 10, thus, the relationship of the states for these two models within the range of the shared data set could be written as

$$x_{M_b}(t_i : t_{i+1}) = T_{ba} x_{M_a}(t_i : t_{i+1}), \quad (30)$$

in which the matlab-like notation $x_{M_a}(t_i : t_{i+1})$ represents the states of model M_a during t_i to t_{i+1} , and $x_{M_b}(t_i : t_{i+1})$ represents the states of model M_b during t_i to t_{i+1} as shown in Fig. 10.

Hence, in this way, instead of requiring multiple model transitions as in [34], as long as we have enough independent states in the gradual transition region to make the state matrix $x_{M_a}(t_i : t_{i+1})$ full row rank, we could compute the transformation matrix T_{ba} by solving (30) in a least square sense. By using T_{ba} , we could transform model M_b to the basis of model M_a through (28).

Following this method, we could calculate the transformation matrices between any two adjacent thermal models using (30), and transform the model basis with (28). In this way, it is straightforward to transform all the identified local linear models to the common basis. We illustrate this by a 3-local system that has models M_a , M_b and M_c . It is straightforward to transform other model states into common model basis of M_a using

$$\begin{aligned} x_{M_b}(t_i) &= T_{ba} x_{M_a}(t_i), \\ x_{M_c}(t_i) &= T_{cb} T_{ba} x_{M_a}(t_i). \end{aligned} \quad (31)$$

In other words, we can just use the common state x_{M_a} as the local model states for all the local models. With the linear local model built from different temperature ranges or different linear ranges onto the same state basis through linear transformations, the piecewise linear model could smoothly switch from one model to another, which benefits the simulation accuracy.

6. Numerical results and discussions

The proposed method has been implemented in Matlab. We show results on a practical package modeled by commercial modeling tool COMSOL V 4.1 [35]. We first use a linear modeling method, then we employ the proposed piecewise linear method to reduce the model order and improve the simulation efficiency.

The packaged microprocessor design used in this study is shown in Fig. 11, where the convective boundary on the top of heat sink models the convective cooling from the fan placed above the processor. The aluminum heat sink is glued to the copper integrated heat spreader (IHS) that is attached to silicon die through a thin layer of thermal interface material (TIM). The materials and geometries of the major parts of the package are shown in Table 1, and we partition the die area into 4×4 power grids as shown in Fig. 12, and each grid represents a different function block.

To model the power consumption of these function blocks, the input power sources are placed in these power grids and we measure the temperature at the adjacent 4 points of each square

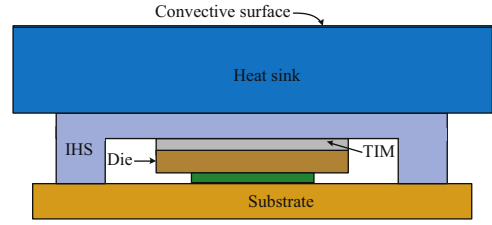


Fig. 11. Microprocessor chip package.

Table 1

Material and geometry of the microprocessor package.

Parts	Material	Dimensions (mm)
Die	Silicon	$10 \times 10 \times 0.7$
TIM	Thermal grease	$10 \times 10 \times 0.2$
IHS	Copper	$31 \times 31 \times 1.5$
Heat sink	Aluminium	$64 \times 64 \times 6.3$
Substrate	FR4	$37.5 \times 37.5 \times 1.3$

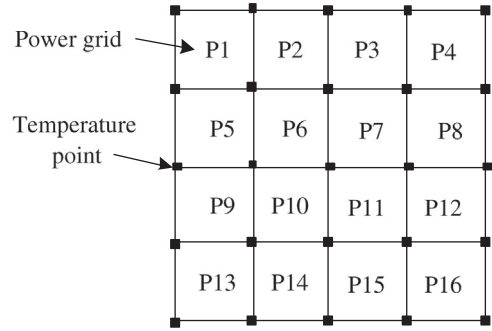


Fig. 12. Top view of partitioned die area with power grids and temperature points (heat sink removed).

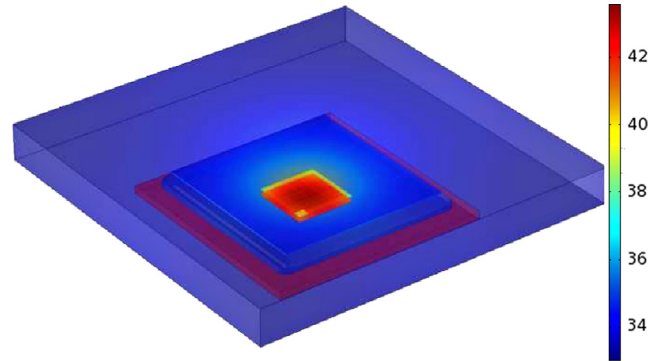


Fig. 13. Steady state temperature distribution simulated by COMSOL 4.1. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

power grid, and thus we could have 25 observable temperature points starting from the up left corner of grid P1 (1st temperature point) to the down right corner of grid P16 (25th temperature point) as shown in Fig. 12. As a result, we end up with 16-input and 25-output thermal systems. The convection coefficient is set to $450 \text{ (W/(m}^2 \text{ K))}$ to model the convective air cooling effect from the cooling fan on top of the chip package.

To build a more realistic package with right dimension and materials, we applied COMSOL V4.1 [35] to build the package structures with on-chip power waveforms as inputs. The time step is set to 0.1 s for the transient simulation, and the thermal response could be obtained by COMSOL V4.1 using the finite element method under the input power maps we generated.

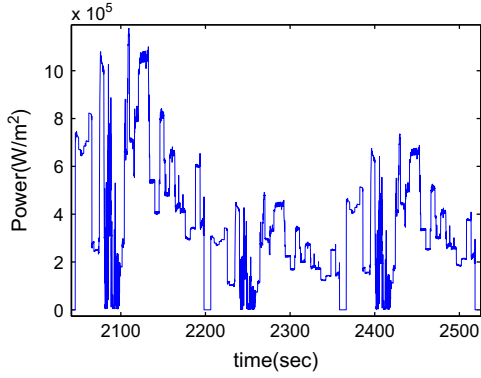


Fig. 14. Transient on-chip power waveform from industry.

Table 2

Additional input power map configurations for validation (range of position variables: $0 < x < 4$ $0 < y < 4$).

Configuration Num.	Spatial distribution of the input
M_1	$P = P_0 \left(\cos \left(\frac{x+y-2}{4} \right) + \frac{xy^3}{\ln(x+2)} + e^{y+2} + \frac{1}{3} \frac{\sqrt{xy} \ln(4y+x-4)}{5x^2 \sqrt{y+1} \sqrt{\sqrt{xy}}} \right)$
M_2	$P = P_0 \left(e^{-2x-y} + \frac{x}{4} \right)$
M_3	$P = P_0 \left(e^{-xy} + \frac{\ln(\sqrt{xy}^2 + 1)}{x} \right)$
M_4	$P = P_0 \frac{(x+y + e^{-x+2} + xy^3 e^{-x} + 1)}{5y\sqrt{x}}$

Fig. 14 shows the transient power waveform from our industry partner (The magnitude is determined by the specific power map.).

Fig. 13 shows the steady state temperature distribution under a given power input on the constructed package and the chip. The colors in Fig. 13 indicate different temperatures, the red color represents hotter part of the package, in this case, it is at the center of the die, and the blue part represents the cooler part of the package like heat spreader. The temperature goes from high to low, and the color turns from red to blue. At the edge of the die, it is hotter than the heat sink and cooler than the die center, so it displays orange or yellow.

6.1. Multi-configuration model identification and validation for the package

Using the proposed method in Section 4.2, 16 orthogonal power map configurations are generated by the 2-D orthogonal sine sets. The system is identified with these 16 input power map configurations, which cover 12 800 transient time steps (about 800 time steps per power map). But we can use less number of time steps per power map. In the validation phase, in addition to the 16 automatically generated power map configurations, new configurations from M_1 to M_4 are introduced and their spatial distributions are defined in Table 2.

A 15th-order thermal model is obtained. The matched time domain response is shown in Fig. 15(a) and frequency domain response is shown in Fig. 16 for this 15th-order thermal model. The baseline results are obtained by using step power input with only one port is excited at a time and it can be viewed as the golden of the transfer function. From those figures, we can clearly see that the proposed method gives very accurate thermal models compared to the golden. We could clearly see the filtering effect of the thermal system, which actually makes good physic sense.

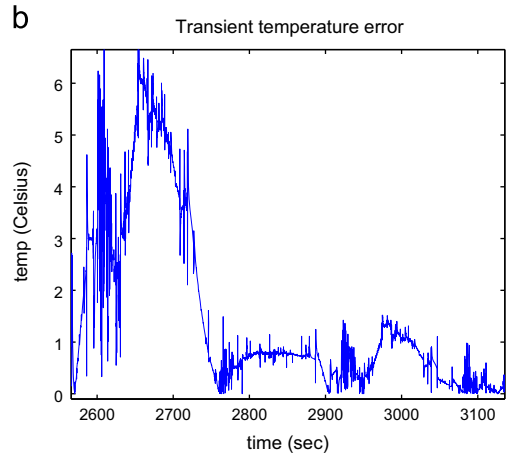
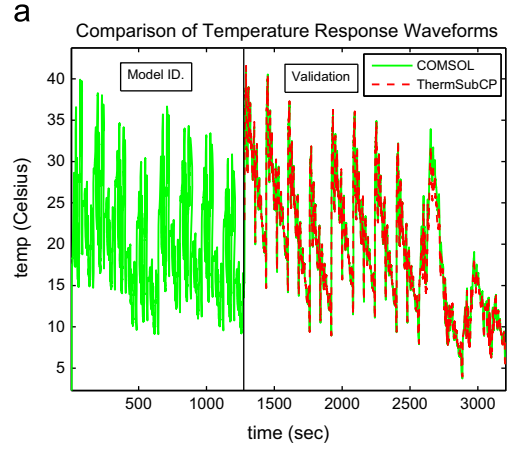


Fig. 15. Transient temperature responses and the temperature errors of the identified model (order:15). (a) On-chip temperature responses at the 1st temperature point. (b) Absolute errors of on-chip temperature responses at the 1st temperature point.

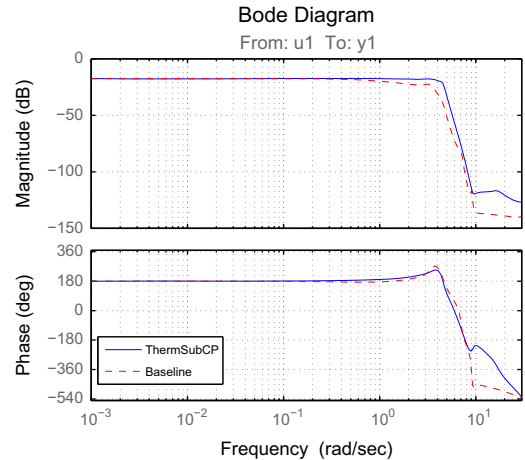


Fig. 16. Bode diagram of the transfer function from input u_1 to output y_1 .

The power, analogous to the electrical current in a resistor, could be changed instantly, while the temperature changes of the system, analogous to the voltage changes on a cap, usually takes some time to happen due to the heat capacitance of the system.

The pole-zero analysis indicates that the system is stable, and Fig. 17 shows that the poles of one transfer function of the 15th-order model ('x' represents poles while 'o' represents zero) are within a unit cycle. We have similar observation for other transfer

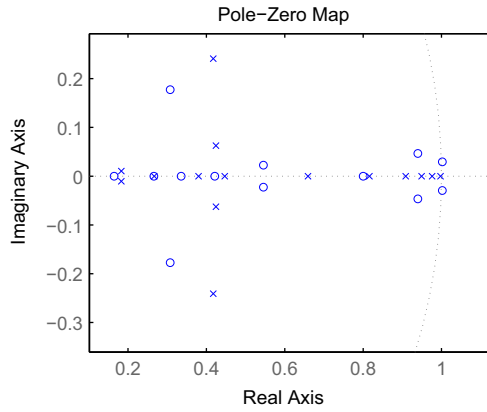


Fig. 17. Poles and zeros distribution of the 15th-order model.

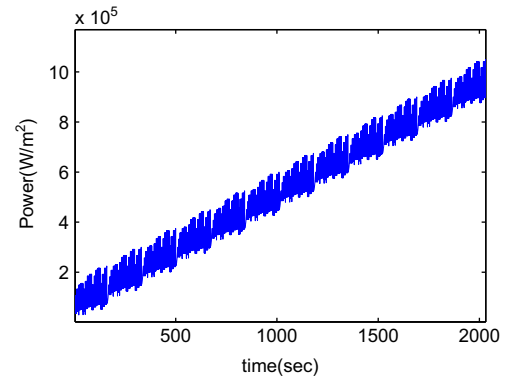


Fig. 18. The stair-like PRBS input signal in model identification phase.

Table 3
Model identification CPU times and model errors.

ThermSubCP order	35	25	15
Model ID time (min)	25.0	15.5	4.83
Max mean error (%)	1.17	2.58	4.91

functions and model of different orders. Also, Fig. 15(b) plots the zoomed-in temperature absolute errors in the transient response, which gives the worst errors in entire validation period. The errors at some worst time points are relatively larger (about 15% or 6–7 °C), but over all the time steps, they are quite small. The output error statistics of the identified systems with different order is summarized in Table 3, where we list the maximum of the relative mean errors among all the ports over the entire transient simulation period (*Max Mean error*), and *Model ID time* is the time (in minutes) required to identify the model.

It clearly shows the trade-off between the model accuracy and the identification time for the same amount of identification data.

6.2. PWL-based nonlinear modeling and validation

In this section, we use PWL modeling approach to build a compact thermal model of the packaged microprocessor system. To obtain the identification data of different local models for different temperature ranges, PRBS (Pseudo Random Binary Sequence) signals with stair-like envelopes shown in Fig. 18 are used as inputs to characterize the system parameters; and in the validation phase, the transient waveform in Fig. 14 is used. PRBS signal has the white-noise like spectrum so that it can excite all the thermal system states, and we expect that it could potentially lead to better and more accurate identification results [25].

6.2.1. PWL-based nonlinear modeling for different temperature ranges

As discussed before, temperature dependent material properties could lead to nonlinear behavior of the thermal system, thus we use piecewise linear method to model the thermal behavior in different temperature ranges. In this case, the stair-like envelop contains 12 “steps” that corresponds to 12 different ranges of input power intensity as shown in Fig. 19. We could arbitrarily partition the data and attribute them to different linear models that need to be identified with these data. At the beginning, we use “scheme-1” shown in Fig. 19(a) to identify the linear models. In this scheme, each model is identified based on two consecutive data sets, and the adjacent models are built with one shared data set. In this way,

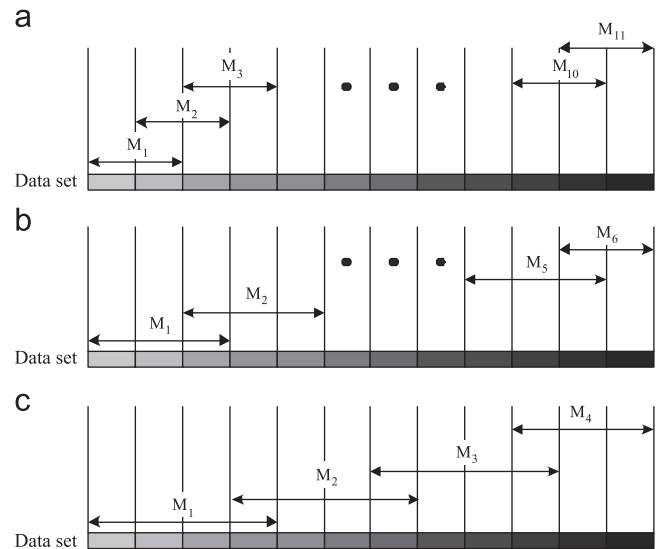


Fig. 19. Data partition schemes for model identification (a) scheme-1 (11 local models), (b) scheme-2 (6 local models), and (c) scheme-3 (4 local models).

11 models will be identified in total, given the 12 data sets, and the piecewise linear model is to be built with these 11 local models. In order to avoid the predictability issue as discussed before, for each range of the input power, 16 orthogonal configurations are generated by $g_{mn}(x, y)$ as defined in (27).

By choosing 4th order model, and using the subspace identification method, all the 11 linear models could be identified. Applying the proposed method to determine the transformation matrices, all the linear models could be transformed to the same basis. Since the piecewise linear model built up in this way contains multiple local models, it is reasonable to partition the overall temperature range into the sub-ranges that the local models correspond to. The simulation result in Fig. 20(a) confirms that the temperature value predicted by the output of the identified piecewise model (dash line) closely matches the reference data (solid line).

In comparison, we also use different schemes of data partition. By using “scheme-2” and “scheme-3”, we end up building the piecewise linear models with 6 local models and 4 local models respectively. From the transient simulation results in Fig. 20 and the temperature error of the transient response in Fig. 21, it clearly demonstrated the performance improvement as more linear local models are used. We could clearly observe that, for the same order, the error reduces as the number of the linear models in use increases, which shows a compelling evidence of using piecewise linear model for compact thermal behavior modeling and

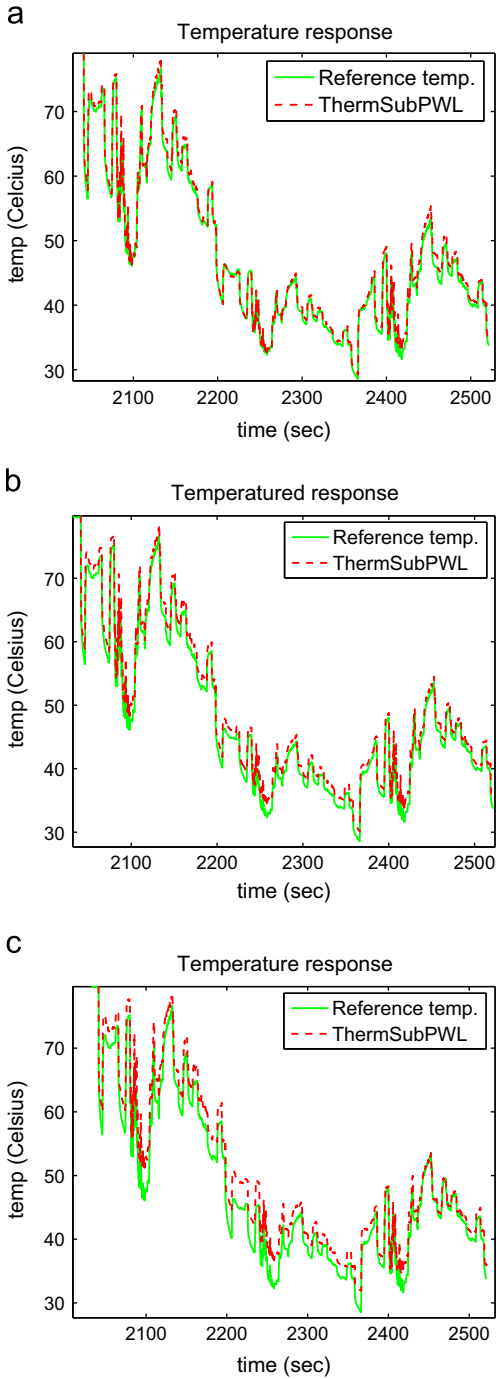


Fig. 20. Transient view of the on-chip temperature response (at the 1st temperature point) of the piecewise linear models (PLM) built with different numbers of local models. (a) PLM built with 11 local models. (b) PLM built with 6 local models. (c) PLM built with 4 local models.

simulation. The output error statistics of the identified system is summarized in Table 4, where we also list the maximum of the mean errors (*Max Mean error*) among all the ports over the entire transient simulation period. We have also confirmed that each local models are stable, and as an example, a pole-zero configuration (poles marked with ‘×’ and zeros marked with ‘○’) of one local linear model is shown in Fig. 22, which shows that all the poles are inside the unit circle.

On the other hand, the linear model with order 4 suffers from huge loss of accuracy as shown in Fig. 6 before. To make the linear model achieve comparable accuracy with the piecewise linear model, high order model needs to be chosen. In our experiment,

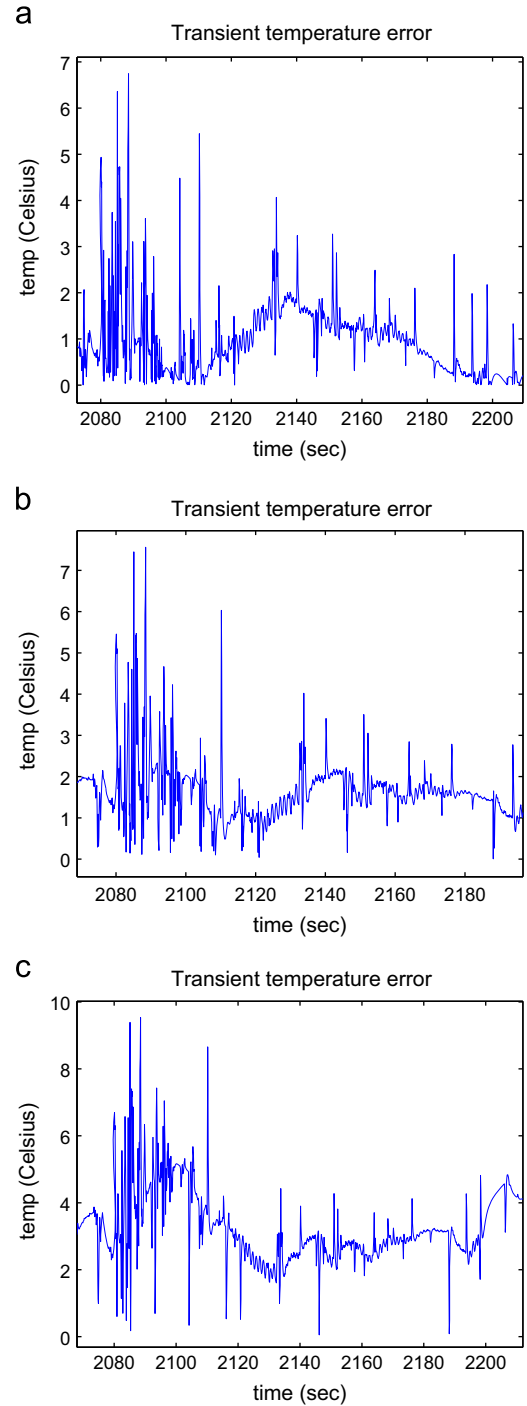


Fig. 21. Absolute errors of the on-chip temperature response (at the 1st temperature point) of the piecewise linear models (PLM) built with different numbers of local models. (a) PLM built with 11 local models. (b) PLM built with 6 local models. (c) PLM built with 4 local models.

Table 4
Model accuracy comparison with different identified models (order: 4).

Num. of linear models in use	11	6	4
Max mean error (%)	2.1	3.9	5.9

we used 20 412 transient time points to identify the model. As summarized in Table 5, the time required to identify (ID time) the high order linear model (LM) is 627.1 s, while on the other

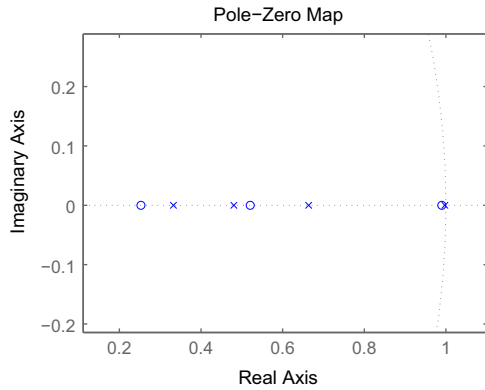


Fig. 22. Pole-zero configuration of a transfer functions in an identified local model.

Table 5

Comparison of model accuracy and CPU times.

Comparison items	Error (%)	ID time (s)	Simulation time (s)
PLM (order:4)	2.1	63.8	7.88
LM (order:15)	2.3	627.1	22.2

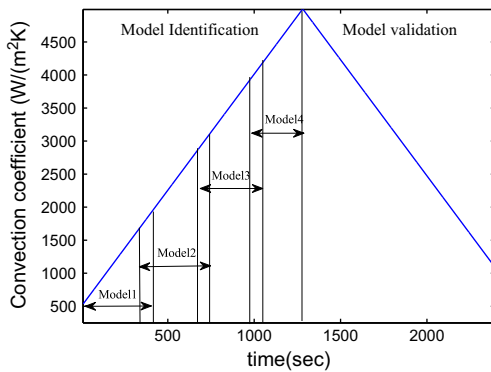


Fig. 23. Different convection rates of the heat sink.

hand, the time required to identify all the piecewise linear models (PLM) is 63.8 s. Hence, the speedup factor for model identification is 9.8 comparing with linear model. Also, we used 25 412 time points in transient simulation, and the high order linear model uses 22.2 s to conclude the simulation, and the piecewise linear model uses only 7.88 s to conclude the simulation, which is approximately 35% of the simulation time of the high order linear model.

We remark that although in our case, order 15 of the linear model is not significantly higher than the order 4 in a sense, yet, the time required to identify the state space model through subspace identification method increases significantly because a large amount of input and output data are required to identify the state space model accurately. As a result, choosing low order model to identify the targeted dynamic system leads to substantial savings in subspace identification method, which is important in the process of building and calibrating a dynamic model in a dynamically changing environment. Also, piecewise linear model achieves substantial savings in simulation time because the lower order model is used in simulation.

6.2.2. PWL-based nonlinear modeling for different convection rates

Another potential thermal nonlinearity could be caused by the changing speed of the cooling fan, which translates to changing

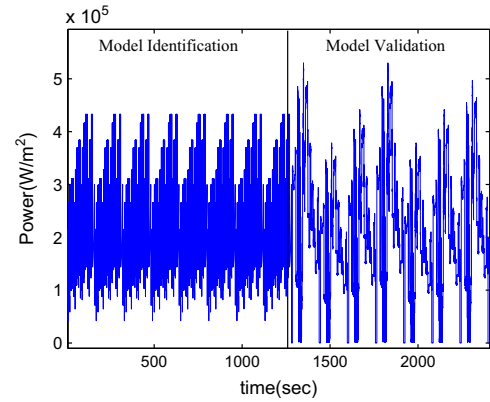


Fig. 24. Power traces used to build and validate the piecewise linear model for different convection rates.

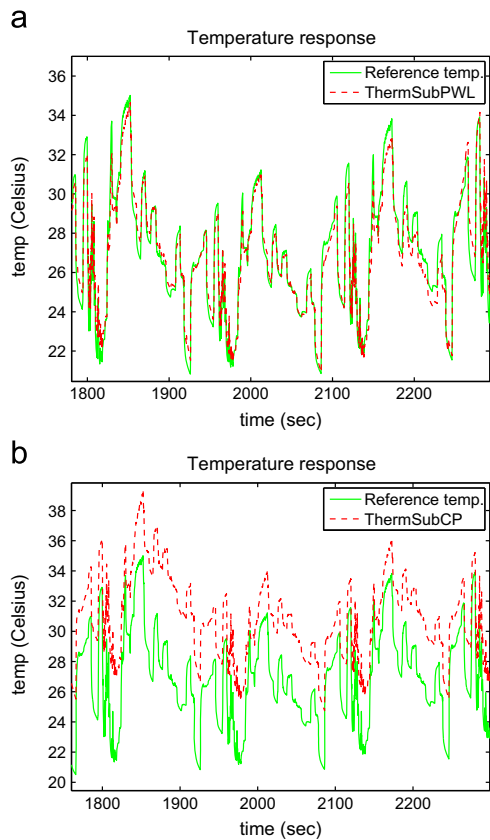


Fig. 25. Comparison of the on-chip temperature response (at the 1st temperature point) predicted by the piecewise linear models (PLM) with the one predicted by a linear model under variant convection rate. (a) PLM built with 4 local 4th order models. (b) A simple linear 4th order model.

(nonlinear) boundary conditions for the thermal circuits. In this case, our proposed piecewise linear model offers a viable solution to build a model working for the changing boundary conditions.

In package level thermal modeling, different fan speeds could be translated to different convection rates at the top surface of the heat sink. Hence, to model various fan speeds, in our experiment, we linearly change the convection rates of the top surface of the heat sink as shown in Fig. 23, and use the model identification routine with 16 orthogonal power map configurations to identify each local model for a certain range of the varying convection rate in Fig. 23. Each two adjacent models covers a common range of the convection rate, which is served as model transition region to

determine the transformation matrices that transform all the local models to a common state base using the procedure in Section 5.2. The identified model is validated under different ranges of convection rate in the validation phase. The power trace in the experiment is shown in Fig. 24, in which PRBS signal is used for model identification and while the power signals from our industry partner are used for model validation.

By choosing 4th order model, and following the same procedure used before, all the 4 local models could be identified. The simulation result in Fig. 25(a) confirms that the piecewise linear model could effectively predict temperature waveforms in different ranges of convection rate. As a comparison, if we use a single

4th order linear model, the predicted transient response sometimes could be significantly deviated away from the reference temperature data as Fig. 25(b) shows. The errors of the transient waveform by the piecewise linear model and the corresponding linear model with the same order are shown in Fig. 26(a) and (b) respectively.

In addition, we could anticipate that the model identified under higher convection rate should be a faster model because the package could reach steady state quicker with higher convection rate, which could be confirmed by investigating the pole-zero configuration shown in Fig. 27. We could find that the dominant pole (0.9879 in z-plane) of one transfer function in Model2 (identified with convection rate from 1625 W/(m² K) to 2750 W/(m² K) is farther away from the unit circle than the dominant pole (0.9920 in z-plane) of the corresponding transfer function in Model1 (identified with convection rate varying from 500 W/(m² K) to 1725 W/(m² K)), and it is also much closer to zero, which partially cancels the effect of the dominant pole. Hence, this result of pole-zero configuration suggests that Model2 has faster response as theoretically anticipated, which also indicates that each local model is identified to take account of the significantly changed convection rate overtime.

Hence, to sum up, we applied the proposed method to model the chip package with variant convection rate, and the experiment result of the piecewise linear model shows significant accuracy improvement over the conventional linear modeling, which confirms the validity of the proposed methodology in thermal modeling. By investigating the resulting pole-zero configuration of the identified model under different convection rates, we also confirmed that each local model is identified to consider the potential huge variation of the fan cooling effects during the chip operation.

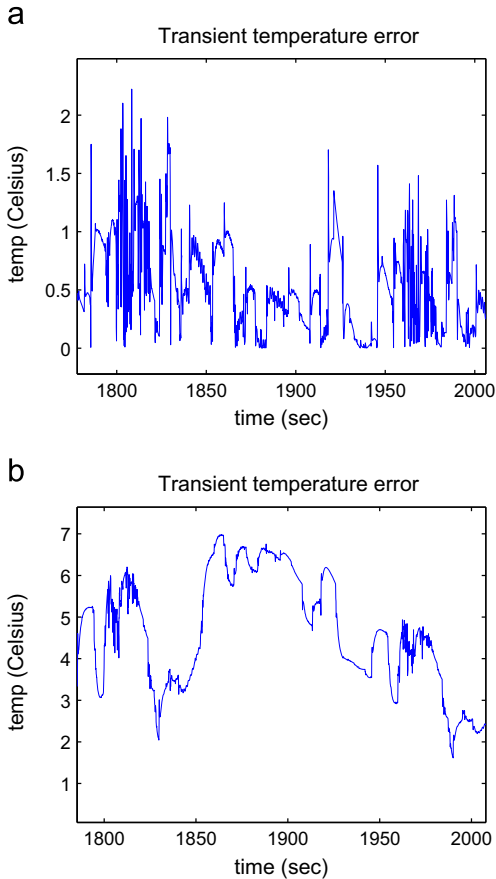


Fig. 26. Absolute error of one on-chip temperature response at the 1st temperature point. (a) Piecewise linear model. (b) Linear model.

7. Conclusions

In this paper, we have first proposed a new thermal modeling technique considering practical power maps with highly correlated input powers. We have proposed to generate independent power maps to meet the spatial rank requirement in the presence of highly correlated input signal. The new method, *ThermSubCP*, can also automatically select the order of the thermal models for the given error bounds. Secondly, we have also proposed piecewise linear (PWL) modeling approach for modeling nonlinear effects and various boundary conditions. The new modeling algorithm, *ThermSubPWL*, partitions the nonlinear ranges (due to temperature or boundary condition changes) into a number of small ranges and performs the modeling on each range. Experimental results have validated the

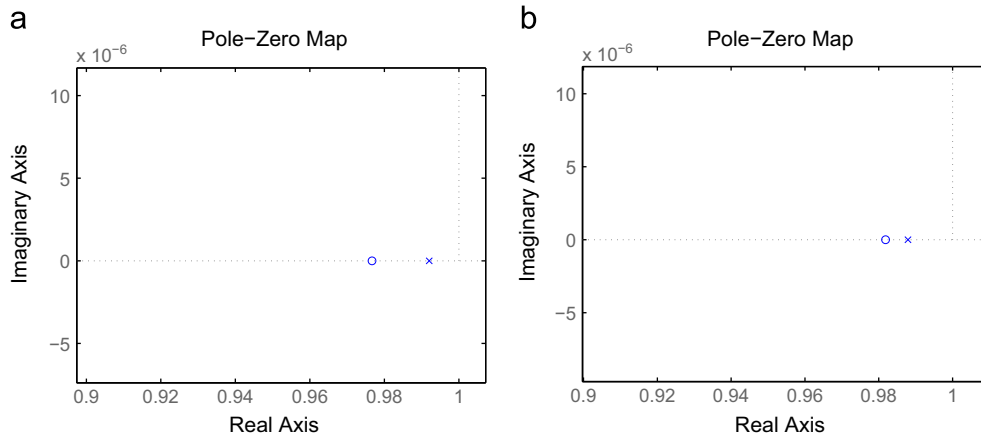
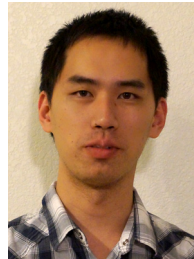


Fig. 27. Comparison of dominant pole-zero of the transfer functions in Model1 and Model2 (from input 1 to output 1). (a) Model1 and (b) model2.

proposed method on a practical microprocessor package modeled by commercial multi-domain/physics tool, COMSOL V4.1, under practical power signal inputs. By using multiple power map configurations in model identification, the predictability of the identified model is significantly improved. The new piecewise linear models can model thermal behavior over wide temperature ranges or over different thermal boundary conditions due to different cooling configurations. Further, the PWL modeling technique can lead to much smaller model order without accuracy loss, which translates to significant simulation time reduction and about $10 \times$ less time to identify the reduced models compared to the simple modeling method by using the high order models in our examples.

References

- [1] Z. Liu, S.X.-D. Tan, H. Wang, R. Quintanilla, A. Gupta, Compact thermal modeling for package design with practical power maps, in: 1st International IEEE Workshop on Thermal Modeling and Management: Chips to Data Centers (TEMM), July 2011.
- [2] Z. Liu, S.X.-D. Tan, H. Wang, A. Gupta, S. Swarup, Compact nonlinear thermal modeling of packaged integrated systems, in: Proceedings of Asia South Pacific Design Automation Conference (ASPDAC), January 2013, pp. 157–162.
- [3] W. Huang, E. Humenay, K. Skadron, M.R. Stan, The need for a full-chip and package thermal model for thermally optimized IC designs, in: Proceedings of the 2005 International Symposium on Low Power Electronics and Design, ACM, New York, NY, USA, 2005, pp. 245–250.
- [4] Y.-K. Cheng, C.-C. Teng, S.-M. Kang, C.-H. Tsai, Electrothermal Analysis of VLSI Systems, Cambridge University Press, New York, NY, USA, 2000.
- [5] International Technology Roadmap for Semiconductors (ITRS), 2010 update (<http://public.itrs.net>), 2010.
- [6] M. Pedram, S. Nazarian, Thermal modeling, analysis, and management in VLSI circuits: principles and methods, Proceedings of the IEEE 94 (August (8)) (2006) 1487–1501.
- [7] D. Brooks, M. Martonosi, Dynamic thermal management for high-performance microprocessors, in: Proceedings of International Symposium on High-Performance Computer Architecture, 2001, pp. 171–182.
- [8] K. Skadron, M.R. Stan, W. Huang, S. Velusamy, K. Sankaranarayanan, D. Tarjan, Temperature-aware microarchitecture, in: Proceedings of International Symposium on Computer Architecture (ISCA), 2003, pp. 2–13.
- [9] S.P. Gurrum, Y.K. Joshi, W.P. King, K. Ramakrishna, M. Gall, A compact approach to on-chip interconnect heat conduction modeling using the finite element method, Journal of Electronic Packaging 130 (September) (2008) 031001.1–031001.8.
- [10] A.M. Sridhar, A. Vincenzi, et al., 3D-ICE: fast compact transient thermal modeling for 3D-ICs with inter-tier liquid cooling, in: Proceedings of International Conference on Computer Aided Design (ICCAD), IEEE Press, 2010, pp. 463–470.
- [11] S. Shidore, Compact thermal modeling in electronics design, Electronics Cooling 13 (2) (2007) 22–25.
- [12] C. Lasance, H. Vinke, H. Rosten, K.-L. Weiner, A novel approach for the thermal characterization of electronic parts, in: Proceedings of the IEEE 11th Annual Semiconductor Thermal Measurement and Management Symposium, 1995, pp. 1–9.
- [13] F. Christiaens, B. Vandeveld, E. Beyne, R. Mertens, J. Berghmans, A generic methodology for deriving compact dynamic thermal models, applied to the PSGA package, IEEE Transactions on Components, Packaging, and Manufacturing Technology, Part A 21 (December (4)) (1998) 565–576.
- [14] A. Augustin, B. Maj, A. Kostka, A structure oriented compact thermal model for multiple heat source ASICs, Microelectronics Journal 36 (August (8)) (2005) 700–704.
- [15] Y.C. Gerstenmaier, G. Wachutka, Rigorous model and network for transient thermal problems, Microelectronics Journal 33 (September) (2002) 719–725.
- [16] H. Pape, D. Schweitzer, J.H. Janssen, A. Morelli, C.M. Villa, Thermal transient modeling and experimental validation in the european project PROFIT, IEEE Transactions on Components and Packaging Technologies 27 (September (3)) (2004) 530–538.
- [17] M. Rencz, G. Farkas, V. Székely, A. Poppe, B. Courtois, Boundary condition independent dynamic compact models of packages and heat sinks from thermal transient measurements, in: Proceedings of the 5th Electronics Packaging Technology Conference, 2003, pp. 479–484.
- [18] F. Beneventi, A. Bartolini, A. Tilli, L. Benini, An effective gray-box identification procedure for multicore thermal modelling, IEEE Transactions on Computers, 2013, p. 1. Available: (<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6381401>) (online).
- [19] W. Huang, M. Stan, K. Skadron, K. Sankaranarayanan, S. Ghosh, S. Velusamy, Compact thermal modeling for temperature-aware design, in: Proceedings of Design Automation Conference (DAC), 2004, pp. 878–883.
- [20] A.K. Coskun, T.S. Rosing, K.C. Gross, Utilizing predictors for efficient thermal management in multiprocessor SoCs, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems 28 (October) (2009) 1503–1516, Available: (<http://dx.doi.org/10.1109/TCAD.2009.2026357>) (online).
- [21] R. Cochran, A.N. Nowroz, S. Reda, Post-silicon power characterization using thermal infrared emissions, in: Proceedings of International Symposium on Low Power Electronics and Design (ISLPED), ACM, New York, NY, USA, 2010, pp. 331–336. Available: (<http://doi.acm.org/10.1145/1840845.1840914>) (online).
- [22] D.-C. Juan, H. Zhou, D. Marculescu, X. Li, A learning-based autoregressive model for fast transient thermal analysis of chip-multiprocessors, in: Proceedings of Asia South Pacific Design Automation Conference (ASPDAC), 2012, pp. 597–602.
- [23] A. Bartolini, M. Cacciari, A. Tilli, L. Benini, A distributed and self-calibrating model-predictive controller for energy and thermal management of high-performance multicores, in: Proceedings of European Design and Test Conference (DATE), 2011, pp. 1–6.
- [24] A. Bartolini, M. Cacciari, A. Tilli, L. Benini, Thermal and energy management of high-performance multicores: distributed and self-calibrating model-predictive controller, IEEE Transactions on Computers 24 (2013) 170–183.
- [25] A.B.A.T.F.B. Roberto Diversi, L. Benini, SCC thermal model identification via advanced bias-compensated least-squares, in: Proceedings of European Design and Test Conference (DATE), 2013, pp. 1–6.
- [26] D. Li, S.X.-D. Tan, E.H. Pacheco, M. Tirumala, Parameterized architecture-level dynamic thermal models for multicore microprocessors, ACM Transactions on Design Automation of Electronic Systems 15 (2) (2010) 1–22.
- [27] T. Eguia, S.X.-D. Tan, R. Shen, D. Li, E.H. Pacheco, M. Tirumala, L. Wang, General parameterized thermal modeling for high-performance microprocessor design, IEEE Transactions on Very Large Scale Integration (VLSI) Systems (2011).
- [28] P.V. Overschee, B.D. Moor, N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems, Automatica 30 (1) (1994) 75–93.
- [29] T. Katayama, Subspace Methods for System Identification, Springer, 2005.
- [30] P.V. Overschee, B.D. Moor, Subspace Identification for Linear Systems, Theory – Implementation – Applications, Kluwer Academic Publishers, 2006.
- [31] M. Rencz, V. Székely, Studies on the nonlinearity effects in dynamic compact model generation of packages, IEEE Transactions on Components, and Packaging Technologies 27 (March (1)) (2004) 124–130.
- [32] G.H. Golub, C.V. Loan, Matrix Computations, 3rd ed., The Johns Hopkins University Press, 1996.
- [33] I. Yeo, C.C. Liu, E.J. Kim, Predictive dynamic thermal management for multicore systems, in: Proceedings of Design Automation Conference (DAC), Series DAC '08, ACM, New York, NY, USA, 2008, pp. 734–739. Available: (<http://doi.acm.org/10.1145/1391469.1391658>) (online).
- [34] V. Verdult, M. Verhaegen, Subspace identification of piecewise linear systems, in: Proceedings of 43rd IEEE Conference on Decision and Control (CDC), 2004, pp. 3838–3843.
- [35] COMSOL Mutiphysics: User Guide, Version 4.1 (www.comsol.com).



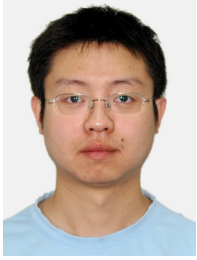
Zao Liu received his MS degree in Electrical and Computer Engineering from Northeastern University (MA) in 2010, and currently he is a Ph.D candidate in University of California, at Riverside. His dissertation research focuses on package level thermal modeling and management for VLSI chip and multi-core processor systems, which is funded by SRC organization



Sheldon X.-D. Tan (S'96-M'99-SM'06) received his B.S. and M.S. degrees in electrical engineering from Fudan University, Shanghai, China in 1992 and 1995, respectively and the Ph.D. degree in electrical and computer engineering from the University of Iowa, Iowa City, in 1999. He is a Professor in the Department of Electrical Engineering, University of California, Riverside, CA. He is the Associate Director of Compute Engineering Program (CEN) at Bourn College of Engineering at UC Riverside since 2009. He also is a cooperative faculty member in the Department of Computer Science and Engineering at UCR. His research interests include

statistical modeling, simulation and optimization of mixed-signal/RF/analog circuits, fast thermal analysis, modeling and dynamic thermal management for microprocessors and platform systems, parallel circuit simulation techniques based on GPU and multicore systems, and embedded system designs based on FPGA platforms. He also co-authored book "Symbolic Analysis and Reduction of VLSI Circuits" by Springer/Kluwer 2005 and "Advanced Model Order Reduction Techniques for VLSI Designs" by Cambridge University Press 2007 and "Statistical Performance Analysis and Modeling Techniques for Nanometer VLSI Designs", Springer 2012. Dr. Tan now is serving as an Associate Editor for three journals: ACM Transaction on Design Automation of Electronic Systems (TODAE), Integration, The VLSI Journal, and Journal of VLSI Design. Dr. Tan received Outstanding Oversea Investigator Award from the National Natural Science Foundation of China (NSFC) in 2008. He received NSF CAREER Award in 2004. Dr. Tan received the Best Paper Award from 2007 IEEE International Conference on Computer

Design (ICCD'07), two Best Paper Award Nomination from 2005 and 2009 IEEE/ACM Design Automation Conferences, the Best Paper Award from 1999 IEEE/ACM Design Automation Conference. He served as a technical program committee member for DAC, ICCAD, ASPDAC, ICCD, ISQED, BMAS, ASICON.



Hai Wang received his B.S. degree from Huazhong University of Science and Technology in 2007, and his M.S. and Ph.D. degree from University of California, Riverside in 2008 and 2012, respectively. After that, he has been an associate professor at University of Electronic Science and Technology of China. His research interests mainly lie in electrical/thermal verification and optimization of VLSI circuits and systems. He serves as TPC member of several international conferences including ASP-DAC and ISQED, and also serves as reviewer of many journals including IEEE TC, IEEE TCAD, and ACM TODEAS.



Yingbo Hua (S'86-M'88-SM'92-F'02) received a B.S. degree (1982) from Southeast University, Nanjing, China, a M.S. degree (1983) and a Ph.D. degree (1988) from Syracuse University, Syracuse, NY. He was a Lecturer (1990-1992), a Senior Lecturer (1993-1995), and a Reader and Associate Professor (1996-2000) with the University of Melbourne, Australia. He was a Visiting Faculty Member with Hong Kong University of Science and Technology (1999-2000), and a Consultant with Microsoft Research, WA (summer 2000). Since 2001, he has been with the University of California at Riverside, where he is a Senior Full Professor.

He has served as Editor, Guest Editor, Member of Editorial Board and/or Member of Steering Committee for IEEE Transactions on Signal Processing, IEEE Signal Processing Letters, EURASIP Signal Processing, IEEE Signal Processing Magazine, IEEE Journal of Selected Areas in Communications, and IEEE Wireless Communication Letters. He has been a Member of IEEE Signal Processing Society's Technical Committees for Underwater Acoustic Signal Processing, Sensor Array and Multichannel Signal Processing, and Signal Processing for Communication and Networking. He has served as member of Technical and/or Advisory Committees for over forty international conferences and workshops. He has authored near three hundreds of articles and coedited three volumes of books, with more than six thousands of citations, in the fields of Sensing, Signal Processing and Communications. He is a Fellow of IEEE and AAAS.



Ashish Gupta, manages the Thermals and Fluids Core Competency Team in Intel's Assembly and Test Technology Development (ATTD) Group in Chandler, Arizona. He holds a Ph.D. degree in Mechanical Engineering from Purdue University. His group is responsible for the thermal design, assembly and technology of IC packages and associated components, with focus on thermal solutions for meeting increased demands for novel solutions across all market segments ranging from hand held devices and small form factor packages with constrained physical and thermal environments to higher power server products. His team is also responsible for providing research and

development of physics based semiconductor assembly, manufacturing and test process fundamentals during the discovery, definition and development stages of technology maturity.