High-Efficiency Transient Temperature Calculations for Applications in Dynamic Thermal Management of Electronic Devices

Maxat N. Touzelbaev
Eurasian National University,
Astana, Kazakhstan
e-mail: touzelbaev_mn@enu.kz

Josef Miler
Department of Mechanical Engineering,
Stanford University,
Stanford, CA 94305

Yizhang Yang
Apple Inc.
Cupertino, CA 95014

Gamal Refai-Ahmed
PreQual Technologies Corp.,
Markham, ON, Canada

Kenneth E. Goodson
Department of Mechanical Engineering,
Stanford University,
Stanford, CA 94305

The highly nonuniform transient power densities in modern semiconductor devices present difficult performance and reliability challenges for circuit components, multiple levels of interconnections and packaging, and adversely impact overall power efficiencies. Run-time temperature calculations would be beneficial to architectures with dynamic thermal management, which control hotspots by effectively optimizing regional power densities. Unfortunately, existing algorithms remain computationally prohibitive for integration within such systems. This work addresses these shortcomings by formulating an efficient method for fast calculations of temperature response in semiconductor devices under a time-dependent dissipation power. A device temperature is represented as output of an infinite-impulse response (IIR) multistage digital filter, processing a stream of sampled power data; this method effectively calculates temperatures by a fast numerical convolution of the sampled power with the modeled system's impulse response. Parameters such as a steady-state thermal resistance or its extension to a transient regime, a thermal transfer function, are typically used with the assumption of a linearity and time-invariance (LTI) to form a basis for device thermal characterization. These modeling tools and the time-discretized estimates of dissipated power make digital filtering a well-suited technique for a run-time temperature calculation. A recursive property of the proposed algorithm allows a highly efficient use of an available computational resource; also, the impact of all of the input power trace is retained when calculating a temperature trace. A network identification by deconvolution (NID) method is used to extract a time-constant spectrum of the device temperature response. We verify this network extraction procedure for a simple geometry with a closed-form solution. In the proposed technique, the amount of microprocessor clock cycles needed for each temperature evaluation remains fixed, which results in a linear relationship between the overall computation time and the number of temperature evaluations. This is in contrast to time-domain convolution, where the number of clock cycles needed for each evaluation increases as the time window expands. The linear dependence is similar to techniques based on FFT algorithms; in this work, however, use of z-transforms significantly decreases the amount of computations needed per temperature evaluation, in addition to much reduced memory requirements. Together, these two features result in vast improvements in computational throughput and allow implementations of sophisticated runtime dynamic thermal management algorithms for all high-power architectures and expand the application range to embedded platforms for use in a pervasive computing environment.
[DOI: 10.1115/1.4024747]

Keywords: transient thermal response, thermal transfer function, thermal RC network, thermal management, digital filters

1 Introduction

Thermal management of high-performance integrated-circuit chips has become one of the most critical design challenges throughout all integrated-circuit architectural and manufacturing communities. As transistor feature size in semiconductor devices continues to shrink and deliver the potential for reduced gate delay, the corresponding power densities and operating frequencies have increased rapidly. Multicore circuit architectures can further exacerbate the problem by creating highly localized transient heat fluxes. These high heat fluxes can cause temperature excursions that have major adverse impacts on device performance, reliability and power efficiency.

Standard cooling solutions may not be well suited for minimizing these highly localized, transient hotspots. Furthermore, cost constraints in industrial applications imply that cooling alone will not be able to resolve these thermal challenges. Instead, attention has shifted to thermally aware circuit design and dynamic thermal management (DTM). In thermally aware circuit design, accurate evaluations of thermal behavior under long power traces are used to optimize applications during the early stages of architecture-level designs [1,2]. This strategy is effective for system design but cannot be adapted to the chip's changing operating conditions and application demands. To account for these effects, dynamic

Contributed by the Electronic and Photonic Packaging Division of ASME for publication in the Journal of Electronic Packaging. Manuscript received July 3, 2012; final manuscript received April 18, 2013; published online July 24, 2013. Assoc. Editor: Amy Fleischer.

thermal management (DTM) performs run-time control of local chip power in response to temperature measurements from on-chip thermal sensors [3]. To date, DTM schemes have not leveraged transient thermal models because existing modeling techniques are not adequately efficient for runtime applications. Widely available numerical modeling software (e.g., finite element models) is inappropriate for runtime applications due to extensive requirements on computational resources. These calculation tools instead provide an excellent reference for validating alternative modeling techniques and will be used for that purpose in this work. Thermal circuit models offer an alternative modeling technique with reduced computational requirements and are discussed in more detail below.

The methods for constructing dynamic compact thermal models can be divided into two general categories: The thermal *RC* network approach and thermal *RC* ladder approach. The first approach constructs an equivalent thermal *RC* network that accurately describes dynamics of the thermal system. This can be achieved by transforming the spatially discretized system matrices of the governing equation in finite element/volume models, into a thermal circuit network consisting of thermal resistance element interconnecting neighboring nodes and heat capacity element to the reference thermal ground [4]. The thermal circuit network is formulate a system of ordinary differential equations (ODE)

$$\boldsymbol{C}\dot{T} + \boldsymbol{R}^{-1}T = \boldsymbol{F}P \tag{1}$$

where $\boldsymbol{C}$ and $\boldsymbol{R}$ are the thermal capacitance and thermal resistance matrices, T is the vector of node temperatures, $\boldsymbol{F}$ is the input power select matrix that maps the power source vector P onto the nodes. While the ODE system in Eq. (1) can be directly solved in circuit simulation software such as SPICE and HotSpot [5], its dimension is proportional to the number of nodes which makes it poorly suited for runtime applications due to high computational requirements. To improve the computational efficiency, several model order reduction methods are developed to transform the high-dimensional system to a low-dimensional one for a faster calculation [6,7]. Figure 1 shows a schematic of a representative thermal network model.

An alternate approach, the *RC* ladder model, is formulated for better computational efficiency for the location of interest. The thermal response of the system, subjected to a step-function power pulse, is recorded with appropriately resolved timescales. A suitable extraction technique is then used to define an *RC* ladder model with an equivalent response [8]. Typically, the needed step-function response $Z(t)$ can be modeled numerically or obtained directly from a measurement. Assuming model linearity, the temperature response $T(t)$ to an arbitrary power trace input $P(t)$ can be computed by the convolution integral between input power and the time derivative of $Z(t)$.

$$T(t) = \int_0^t P(\tau) \cdot \dot{Z}(t-\tau) d\tau = P(t) \otimes \dot{Z}(t) \tag{2}$$

The *RC* network model can be suited for describing a multi-input-multi-output (MIMO) thermal system. In this modeling approach, obtaining a dynamic thermal response in Eq. (1) requires a spatial discretization of governing equations over a complete model domain. The model complexity often requires extensive linear algebra manipulations for reducing the number of unknowns in the studied system. A model order reduction method must be carried out with caution in order to ensure numerical stability. In addition, this approach relies on the discrete numerical models representing the thermal system and cannot directly be based on the experimental results.

In contrast, the *RC* ladder approach can take either simulation or experimental input for model formulation. The *RC* ladder model resolves only a single conduction path; for modeling a MIMO system such as a chip with multiple hotspots, the results of several models must be linearly superimposed. Therefore, the *RC* ladder approach is more appropriate for thermal systems with a limited number of points of interest, such as hotspots. In this paper, we use the *RC* ladder model.

Within the category of *RC* ladder models, there are two types of models to consider: Foster ladder models and Cauer ladder models, as shown in Fig. 2. Cauer ladder models provide a better physical description the heat flow path in the system, while Foster ladder models only capture the thermal behavior but have no

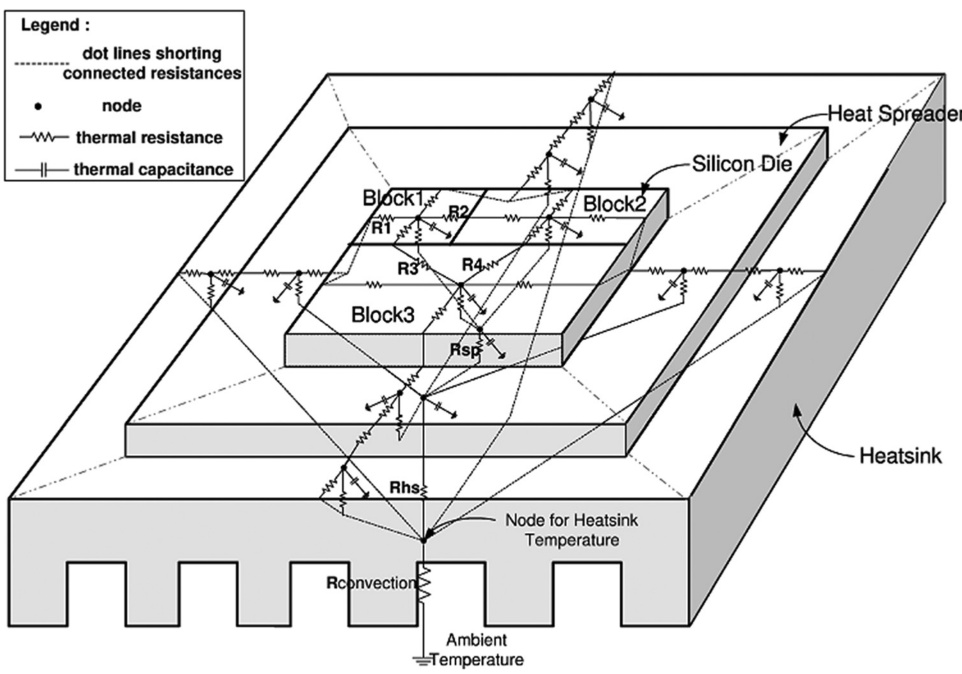

Fig. 1 Example of a network circuit model of a chip die on a heat spreader attached to a heat sink. This type of model [5] is not well-suited for runtime implementations due to a significant computational penalty.

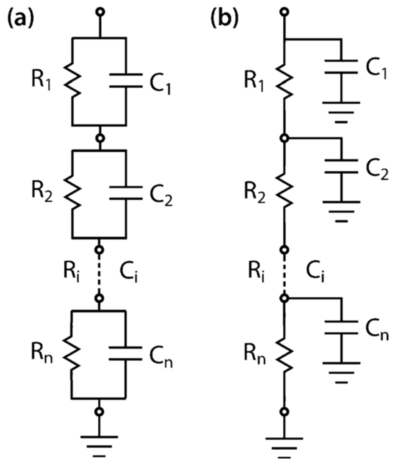

Fig. 2 (*a*) Foster *RC* ladder and (*b*) Cauer *RC* ladder representations of thermal system.

physical equivalent. Use of the Cauer ladder is not straightforward due to its complicated mathematical representation. Fortunately, a Foster ladder can be easily transformed to a Cauer ladder which provides the same step-function thermal response $Z(t)$. For this reason, step-function responses are typically characterized by the Foster-type ladder, and such an approach is taken in this paper.

As mentioned, the improvements in modeling efficiency using *RC*-ladder approach are made by recognizing that the thermal response is only needed for specific regions of the chip, typically at hotspot locations. Still, developing reduced models presents challenges due to strong dependence of the device thermal response on the temporal and spatial pattern of input power, the disparate values of thermal time constants, and variations in boundary conditions. Furthermore, implementing these models for a chip-level runtime temperature regulation means that the model needs to be extremely efficient, in particular when evaluating convolution integral needed for the temperature calculation.

The present work directly addresses the modeling and computational challenges needed for a successful implementation of runtime temperature calculations into a framework of a thermally aware architecture. Section 2 provides a summary of network identification by deconvolution (NID) and discusses numerical aspects of this technique. We then propose an improved method for calculating temperature evolution with arbitrary power traces which bypasses direct evaluation of the convolution integral. It employs a recursive infinite impulse response (IIR) digital filter on sampled power traces, which is a well-developed and commonly used approach in digital signal processing community. A derivation of the IIR digital filter coefficients based on the parameterized thermal *RC* ladder is then presented. Section 3 discusses model validation results and demonstrates the best achievable scaling of the required execution time with the number of runtime computations by comparing the proposed technique to the existing convolution methods. The improvements in computational efficiency proposed in this work make this approach well-suited for implementation in a majority of computing architectures developed to date and ease the adoption for future use.

2 Thermal Modeling Approach

2.1 Determining the Step Response of the Foster RC ladder. Using a Foster *RC* ladder model, the time dependent thermal impedance can be written as

$$Z(t) = \sum_{i=1}^{K} R_i \left(1 - e^{-\frac{t}{R_i C_i}}\right) \tag{3}$$

where R_i (K/W) and C_i (J/K) form the ith stage of *RC* ladder (Fig. 2), with K being the number of stages in the ladder. Various methods have been developed to determine the discrete elements in a ladder shown in Fig. 2, e.g., through fitting of the simulated/measured heating curves in time- [9] and frequency- [10] domains. A method preferred in this work was proposed by Szekely [11] and is based on computing distributed time-constant spectrum from a measured thermal transient response in time domain. For an experimental implementation, this method has been extended by applying a similar identification procedure on the measured impulse response spectrum in the frequency domain [12]. As previously stated, the Foster *RC* ladder can be converted to an equivalent Cauer *RC* ladder to provide increased physical insight, if desired. A brief summary of this technique follows next.

The step response for unit power can be generalized by considering a continuous time-constant spectrum

$$Z(t) = \int_0^\infty R(\tau)\left(1 - e^{-\frac{t}{\tau}}\right)d\tau \tag{4}$$

Introducing variables

$$z = \ln(t); \psi = \ln(\omega); \xi = \ln(\tau) \tag{5}$$

one obtains

$$T(z) = \int_{-\infty}^{\infty} R(\xi)\left(1 - e^{-e^{z-\xi}}\right)d\xi \tag{6}$$

$$Q_t(z) = \frac{dT(z)}{dz} = \int_{-\infty}^{\infty} R(\xi) e^{z-\xi-e^{z-\xi}} d\xi = R(z) \otimes e^{z-e^z} \tag{7}$$

Similarly, in frequency domain, the following expressions are found

$$\begin{aligned} Q_r(\psi) &= -\frac{dRe[\theta(\psi)]}{d\psi} = R(-\psi) \otimes w_r(\psi) \\ Q_i(\psi) &= -Im[\theta(\psi)] = R(-\psi) \otimes w_i(\psi) \end{aligned} \tag{8}$$

Equations (7) and (8) are convolutions of the desired system response with the functions

$$\begin{aligned} w_t(x) &= e^{x-e^x} \\ w_r(x) &= \frac{2e^{2x}}{(1+e^{2x})^2} \\ w_i(x) &= \frac{e^x}{1+e^{2x}} \end{aligned} \tag{9}$$

Szekely [13] gave an overview of the chances of deconvolving time-constant spectrum from the measured or calculated system responses. In this work, we use closed-form expressions for the spectrum of these convolving functions

$$\begin{aligned} W_t(k) &= \Gamma(1 - 2i\pi k) \\ W_r(k) &= \pi^2 k \operatorname{csch}(\pi^2 k) \\ W_i(k) &= \frac{\pi}{2} \operatorname{sech}(\pi^2 k) \end{aligned} \tag{10}$$

where k is the variable in the Fourier domain complementary to the logarithm of the time constant, the gamma function Γ (analytically continued into a complex plane) is used for the time-domain deconvolution method, and hyperbolic functions are used for deconvolution using the associated frequency-domain techniques.

The time constant spectrum can then be found by the inverse transform

$$R_{t,r,i}\begin{pmatrix} z_t \\ -z_r \\ -z_i \end{pmatrix} = \boldsymbol{F}^{-1}\left\langle \frac{G_{t,r,i}(k)}{W_{t,r,i}(k)} \boldsymbol{F}\langle Q_{t,r,i}\rangle \right\rangle \tag{11}$$

where $\boldsymbol{F}$ and $\boldsymbol{F}^{-1}$ represent a Fourier transform pair and $G_{t,r,i}$ is an appropriately chosen filter, typically Gaussian [13], suppressing discretization and numerical round-off errors at high k values.

To uncover the computational aspects of NID methodology, it is useful to examine a system consisting of one single time constant. In this case, a transform of a Q_t function, given by $R_0 e^{z-\xi_0-e^{z-\xi_0}}$, can be written out as

$$\mathbf{Q}(k,\xi_0) = R_0 \times e^{-2\pi k i \xi_0} \times W_t(k) \tag{12}$$

Dividing this expression by $W_t(k)$ in principle yields a spectrum of a single pole located at ξ_0. In practice, $W_t(k)$ decreases sharply with increasing k, and if the transform is calculated numerically, as opposed to the analytical form of Eq. (12), the deconvolution accuracy may be limited by round-off errors in calculating $\mathbf{Q}(k,\xi_0)$. The errors with magnitudes comparable to $|R_0 \times W_t(k)|$ at sufficiently high k-values will render these components unusable. Filtering this part of the spectrum is necessary and leads to broadening of identified peaks. Limiting value of functions in Eq. (10) to above numerical noise threshold was also found effective [13].

2.2 Review of Methods for Computing the Convolution Integral. A direct approach is to compute the convolution integral in Eq. (2) in time domain. For a general case of a piecewise constant power input P_j between t_j and t_{j+1}, a discretized convolution integral becomes

$$T(t) = \sum_{j=1}^{m} P_j\left[Z(t-t_j) - Z(t-t_{j+1})\right] \tag{13}$$

where $t = t_{m+1}$. A similar approach with interpolated $Z(t)$ is proposed in Ref. [14].

A model reduction discussed in Sec. 1 with a piecewise linear approximation of $P(t)$ can be used to obtain a different approximation of the convolution integral [15]

$$T(t) = \sum_{i=1}^{K} R_i \times \sum_{j=1}^{m}\left\{ e^{-\frac{t-t_{j+1}}{\tau_i}} \times \left[P_j + a_j\cdot(t_{j+1}-t_j-\tau_i)\right] - e^{-\frac{t-t_j}{\tau_i}} \times \left[P_j - a_j\cdot\tau_i\right]\right\} \tag{14}$$

where $a_j = P_{j+1} - P_j/t_{j+1} - t_j$ and $\tau_i = R_i C_i$.

These direct time-domain convolution techniques can be effective for short temperature time traces, but due to linearly expanding number of terms with increasing time, the amount of calculations needed for the m-th evaluation scales with m^2.

The convolution integral can also be calculated by a multiplication in frequency domain of a power spectrum and of a thermal transfer function, followed by the inverse operation

$$T(t) = \boldsymbol{F}^{-1}\langle \boldsymbol{F}\langle P(t)\rangle \times \theta(f)\rangle \tag{15}$$

where $\theta(f) = \boldsymbol{F}\dot{Z}(t)$. Equation (15) is a frequency domain equivalent of the convolution integral given by Eq. (2). With the transforms evaluated numerically, this formulation is an equivalent of approach shown in Eq. (13) without a need for a model reduction to an RC ladder once $\theta(f)$ is available. A reduced RC model would lead to an expression of the form

$$T(t) = \sum_{i=1}^{K} R_i \times \left\{ \boldsymbol{F}^{-1}\left\langle \boldsymbol{F}\langle P(t)\rangle \times \frac{1}{1+2\,\pi\, i\, f\tau_i}\right\rangle\right\} \tag{16}$$

where f is frequency. The drawback of this approach is the requirement for the calculation of a power spectrum for all time-scales. As number of temperature calculation increases, Fourier-domain techniques can outperform direct convolution schemes especially if using fast Fourier transforms, which scale as $N_f \times \log(N_f)$, where N_f is the number of points kept in discretized power spectrum. The overall computational penalty still makes its efficient implementation a challenge for a runtime implementation with limited compute and memory resources.

2.3 Recursive Digital Filtering Technique for Computing the Convolution Integral. A technique for computing the convolution integral, presented in this work, is developed by using a form of discrete Laplace transforms (z-transforms) for convolving a sampled physical variable with a known impulse response. The transfer function of a digital filter can be constructed to approximate a response of a modeled thermal system due to its linear time-invariant (LTI) property.

A model reduction is based on Foster RC ladder shown in as a series of n-stage, first-order, low-pass filters in series, acting on input power and outputting temperature in the continuous time domain. These filters in the continuous time domain can be transformed to a digital format in the discrete time domain, which would then be applied to the sampled input power to achieve the desired temperature response.

The use of digital filters instead of continuous-time filters is advantageous for numerous reasons: it is easy to design and implement; it can handle large dynamic range; it has stable performance. The general representation of digital filter in discrete time domain is the difference equation:

$$y(n) = \sum_{i=0}^{M} b_i \cdot x(n-i) - \sum_{i=0}^{N} a_i \cdot y(n-i) \tag{17}$$

where $x(n)$ is the input signal at instant n, $y(n)$ is the output signal at instant n, and constants a_i, $i = 0, 1\ldots N$, and b_j, $j = 0,1\ldots M$, are feed-back and feed-forward coefficients, respectively. Infinite impulse response (IIR) filters have nonzero feedback coefficients ($a_i \neq 0$), as opposed to the finite impulse response (FIR) filters ($a_i = 0$). The IIR filter has an impulse response that is nonzero over an infinite length of time, which is a desirable property for modeling a physical thermal response. In z-domain, a transfer function of the digital filter in Eq. (27) is a ratio of polynomials with powers of z^{-1}, with the feed-back and feed-forward coefficients forming the denominator and numerator polynomials, respectively.

For a system represented by the Foster RC ladder, the impulse response function $\dot{Z}(t)$ has properties of an IIR filter due to its exponentially decaying terms. To determine the feed-back a_i and feed-forward b_j coefficients for the difference equation, the continuous transfer function $H(s)$ of the analog multistage filter, which can also be represented as a ratio of polynomials, is transformed to a discrete transfer function $H(z)$ of its approximate IIR digital equivalent. An inverse z-transform is then applied to calculate filter coefficients for the difference equation. There are numerous available methods to transform from $H(s)$ to $H(z)$; in this work, bilinear transformation method is preferred due to the absence of frequency aliasing distortions.

In the continuous frequency domain, the complex impedance of the Foster thermal RC ladder is given by

$$H(s) = \sum_{i=1}^{K} \frac{R_i}{1+s\cdot\tau_i} \tag{18}$$

where s is the complex frequency. A bilinear transformation is performed by substitution of s in $H(s)$ with

$$s = \frac{2}{\Delta t}\frac{z-1}{z+1} \tag{19}$$

with Δt denoting the sampling interval. Using inverse z-transform one then obtains the desired difference equation

$$T(n) = \sum_{i=1}^{K} \left\{ \frac{2\tau_i - \Delta t}{2\tau_i + \Delta t} T_i(n-1) + \frac{R_{th,i}\Delta t}{2\tau_i + \Delta t}[P(n) + P(n-1)] \right\} \tag{20}$$

where $T(n)$ and $P(n)$ are the discrete temperature output and power input at time $n \times \Delta t$. As readily seen, this method bypasses the direct convolution of the integral in Eq. (2). It recursively calculates the transient temperature response at any time step by using the temperature output at the previous time step while applying trapezoidal rule for integrating power input within the time interval.

This work models a system subjected to a single heat source. For a thermal system subjected to multiple heat sources, the transient temperature response at any location is the superposition of the responses from multiple power excitations. Due to nonlinear frequency mapping, the bilinear transformation method causes frequency warping in $H(z)$ that does not preserve the frequency characteristics of corresponding $H(s)$. This effect can be either eliminated by prewarping before transformation, or minimized by reducing sampling interval Δt according to the following condition:

$$\frac{\Delta t}{\tau_{\min}} \approx \tan\left(\frac{\Delta t}{\tau_{\min}}\right) \tag{21}$$

where $\tau_{\min}$ is the minimum thermal time constant of the stages in Foster RC ladder. For $\Delta t/\tau_{\min} = 0.2$, the amount of frequency warping is approximately 1.4% and has negligible effect on transient response in the time domain.

To summarize, the general procedure for obtaining digital filter coefficients followed in this work involves identification of time constant spectrum, its discretization into Foster's ladder, mapping of a transfer function into z-domain and finally an inverse z-transform to obtain feedback coefficients for the part of the filter acting recursively on the previously calculated temperature and feed-forward coefficients for the part of the filter acting on power input into the identified Foster's ladder.

3 Model Verification and Applications

3.1 Verification. The technique is verified against the analytical solution for one-dimensional conduction in a Cartesian geometry. The solution in complex frequency domain for semi-infinite geometry yields

$$\theta(s) = \frac{1}{\gamma\sqrt{s/\alpha}} \tag{22}$$

while for the finite geometry restricted to a slab of thickness L, with zero temperature boundary condition at the side opposing to the entering heat flux, the solution is

$$\theta(s) = \frac{\tanh(L\sqrt{s/\alpha})}{\gamma\sqrt{s/\alpha}} \tag{23}$$

where $\theta(s)$ is the temperature response per unit heat flux, γ and α are thermal conductivity and diffusivity, respectively. The temperature response at a time constant τ can be directly calculated using real negative axis in s-plane to obtain the spectrum as [16]

$$R(\tau) = \frac{1}{\pi} Im\theta\left(-\tau^{-1}\right) \tag{24}$$

The time-constant spectrum for the distributed case of semi-infinite media is then

$$R(\tau) = \frac{\sqrt{\alpha\tau}}{\pi\gamma} \tag{25}$$

while for the case with restricted geometry the spectrum is discontinuous

$$\tau_n = \frac{4}{(\pi n)^2}\frac{L^2}{\alpha}; \quad R_n = \frac{8}{(\pi n)^2}\frac{L}{\gamma}; \quad n = 1, 3, 5\ldots \tag{26}$$

with zero response at all other time constant values besides τ_n listed above.

Equation (23) provides a solution for a thermal transfer function in complex frequency for a case of restricted geometry domain, while in time domain the solution for unit step in power density is given by [17]

$$Z(t) = \frac{2\sqrt{\alpha t}}{\gamma}\sum_{n=0}^{\infty}(-1)^n\left[ierfc\frac{nL}{\sqrt{\alpha t}} - ierfc\frac{n+1}{\sqrt{\alpha t}}L\right] \tag{27}$$

Figure 3 shows the time constant spectrum identified by the NID procedure with all three deconvolution methods of Eq. (11) using discrete Fourier transforms. The thermal properties are that of silicon ($\gamma = 150\,\mathrm{W\,m^{-1}\,K^{-1}}$, $\alpha = 8.47\ 10^{-5}\ \mathrm{m^2\,s^{-1}}$) and thickness of the slab is $L = 15 \times 10^{-3}$ m. All extracted time constants collapse onto one curve due to the use of an identical Gaussian filter function in the complementary domain. If a definite peak is identified, it is integrated to obtain a lumped value centered at the peak location. First several system poles, given by Eq. (26), are identified accurately by the deconvolution procedure, as shown in Fig. 4. The identified response converges to the distributed semi-infinite limit given by Eq. (26); the time constants at these values can be lumped into a Foster's ladder by integrating sections of a continuous spectrum.

Figure 5 shows transfer functions constructed from the identified ladder. Using a full spectrum of identified response provides near-identical match with the exact form given by Eq. (23) at $s = i \times \omega$. A ladder with seven or more elements provides a good approximation to the system thermal behavior at frequencies with relevant amplitude responses. Figure 6 shows response of the constructed IIR filter to step in power for seven-stage identified ladder with varying time steps. The frequency warping becomes pronounced at large time steps but the filter output is stable and always converges to the expected steady state value.

3.2 Applications. Figure 7 depicts a schematic of a common thermal system for demonstrating the proposed model generation

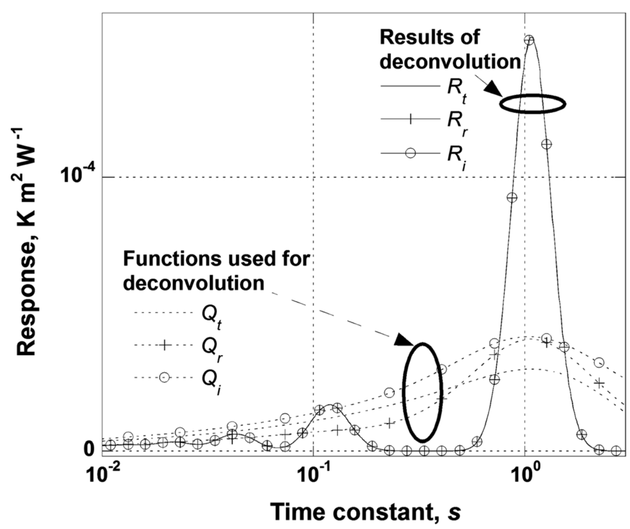

Fig. 3 Network identification by deconvolution (NID) using responses in time and frequency domains for one-dimensional conduction with a restricted geometry discussed in Sec. 3

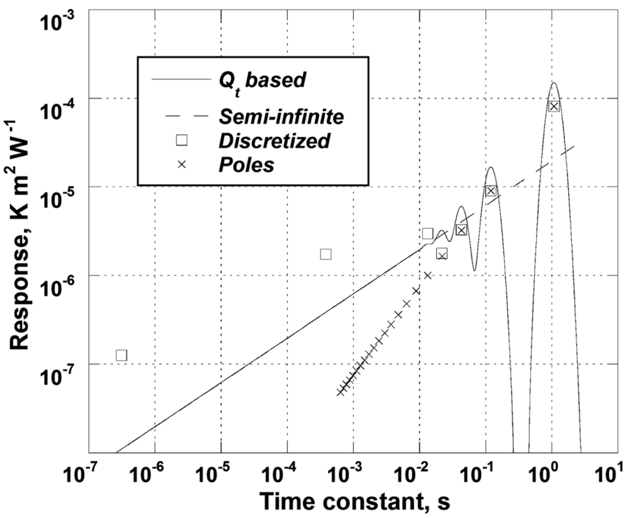

Fig. 4 An identification of system poles and its semi-infinite limit. Refer to Sec. 3 for more detail.

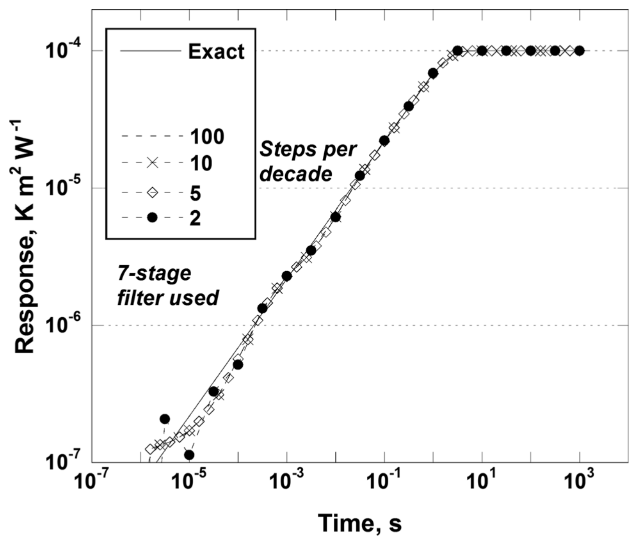

Fig. 6 Time step-size dependence. Technique demonstrates good accuracy with limited time-steps. Slight deviations are observed at very short timescales when using limited number of time-steps.

using a commercially available solver. A representative two-layer configuration consists of a chip with properties of silicon and a heat sink with properties of copper subjected to forced convective cooling. A thickness of 10 mm × 10 mm silicon chip is 0.5 mm and a thickness of 30 mm × 30 mm copper spreader is 1 mm, with a uniform convection coefficient of 10^4 W m^{-2} K^{-1}. A uniform power of 100 W was applied at the top surface of the silicon layer. This representative system was modeled using COMSOL® Multiphysics software. Three different mesh sizes ranged from approximately 1100 to 10,500 mesh points. For all cases, the results are self-consistent.

Figure 8 shows a result of the ladder extraction using a power step response, as discussed earlier, with logarithmically spaced time steps. The IIR filter, generated with 11-element identified ladder, was used to process the same power step at identical time samples. The agreement is excellent with less than 0.4% maximum transient error with respect to the steady-state response value.

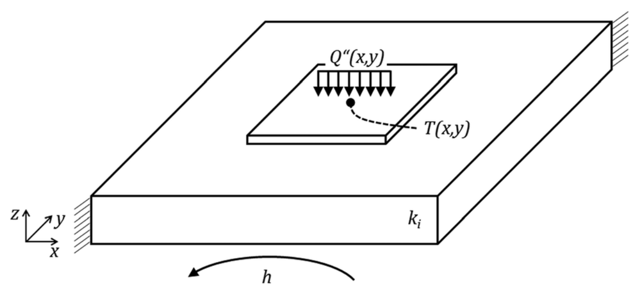

Fig. 7 Schematic of the chip geometry used for the numerical simulation and proposed modeling method.

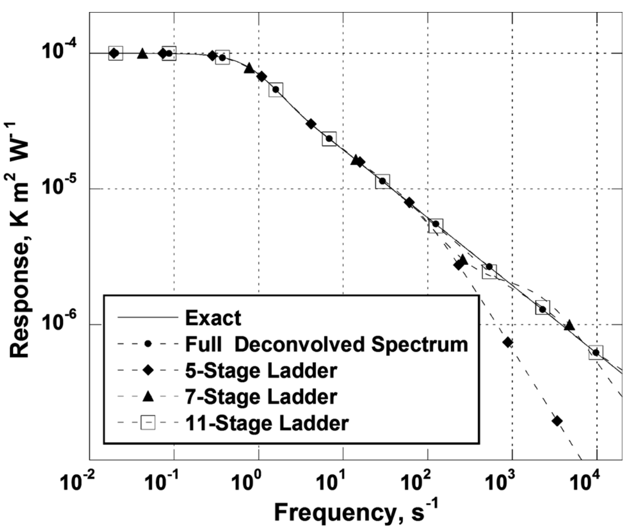

Fig. 5 Transfer function identification. The exact transfer function is well-represented with a limited number of circuit stages using proposed method. Deviations occur at high frequencies depending on the number of latter stages.

Figure 9 compares various convolution methods based on extracted step shown in Fig. 8. The agreement is good between the direct time domain, FFT-based frequency domain and IIR filter convolutions.

Figure 10 compares execution times of the IIR filter with that of the existing techniques. The widths of the summation intervals in Eqs. (13) and (14) are proportional to the time step index m. This results in linear with number of past time steps demand for computation power for each new time increment, which in turn increases the overall computation load proportionally to m^2. Memory requirements for the storage of time/power traces also become a consideration. For the reasons stated, the existing algorithms do not provide efficiencies needed for runtime computational temperature tracing given real-time transient power input. These inefficiencies make implementations particularly difficult in most embedded platforms.

Frequency-domain FFT-based techniques can provide linear scaling but each time step requires significant number of calculations in order to compute the power spectrum and corresponding responses with the sufficient resolution. IIR digital filter based computation also provides linear scaling for the overall computational load, but the number of computation per each time step is reduced significantly. Additionally, memory requirements are much less demanding compared to existing techniques, making the proposed methodology ideally suited for runtime temperature calculations.

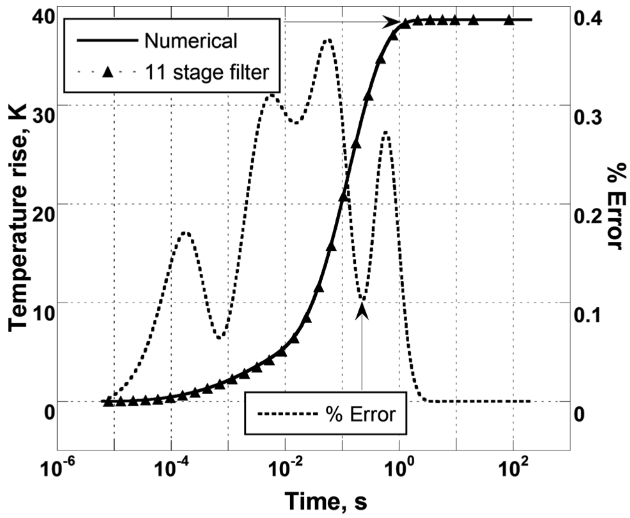

Fig. 8 Thermal modeling of chip-spreader geometry shown on Fig. 7. Compared are the results of simulations using commercial solver with the output of an IIR filter based on 11 stage ladder. The power step is at 100 W. The maximum transient errors are less than 0.4% of the steady state response.

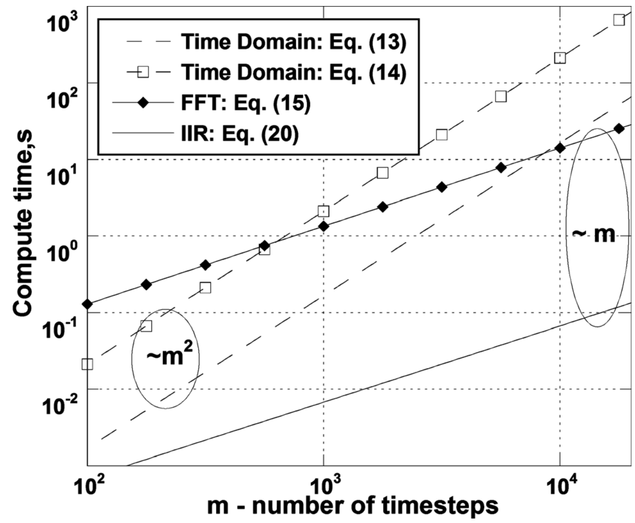

Fig. 10 Comparison in computational efficiency of different methods for evaluation of convolution integrals. The recursive IIR digital filter is superior to other convolution techniques and is best-suited for run-time temperature calculations.

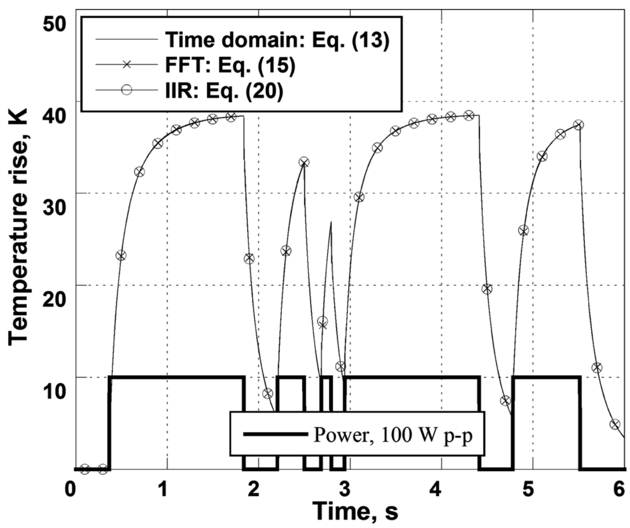

Fig. 9 Comparison between direct time-domain, frequency-domain and IIR digital filter convolution calculations using extracted models from step response given in Fig. 8.

4 Summary and Concluding Remarks

This work presents a novel approach for predicting temperature evolution in electronic devices subjected to transient heat sources. It is based on modeling dynamic behavior of a thermal system with an identified and discretized time-constant spectrum. We use network identification by deconvolution (NID) to obtain a compact thermal model as an *RC* Foster ladder; this form is found particularly useful due to an easy conversion to a digital filter representation using a bilinear transformation. We verify the model extraction procedure using analytical solution and demonstrate correct identification of known system poles and convergence of the extracted time constant spectrum to the limiting case. We then present IIR digital filters suited for run-time evaluation of convolution integral in discrete time-domain. A simple formulation of recursive digital filters makes the algorithm well-suited for run-time temperature predictions. The resulting recursive algorithm yields temperature calculation at a given time instant using just one value of a previous temperature response for each filter stage; the power storage requirements are even less demanding since it's common to all filter stages. A numerical model of semiconductor device is created to generate time-domain temperature responses to step-function power excitation; excellent accuracy of the filter output is confirmed when compared to simulations.

Acknowledgment

The authors gratefully acknowledge support from Advanced Micro Devices (AMD) Inc. as part of the Semiconductor Research Consortium (SRC) Task 1966 and further support from the Stanford Department of Mechanical Engineering Graduate Teaching and Research Fellowship. We would also like to acknowledge discussions with Dr. Sridhar Sundaram of Samsung Austin Research and Development Center and support from Dr. Alexander Glew of Glew Engineering and Dr. Almasadam Satkaliyev of Samruk-Energy, Kazakhstan. This work was supported in part by Advanced Micro Devices (AMD) Inc. as part of the Semiconductor Research Consortium (SRC) Task 1966.

References

[1] Rosinger, P., Al-Hashimi, B. M., and Chakrabarty, K., 2006, "Thermal-Safe Test Scheduling for Core-Based System-on-Chip Integrated Circuits," IEEE Trans. Comput.-Aided Des., **25**(11), pp. 2502–2512.

[2] Liu, C., Iyengar, V., and Pradhan, D. K., 2006, "Thermal-Aware Testing of Network-on-Chip Using Multiple-Frequency Clocking," Proceedings of the 24th IEEE VLSI Test Symposium, Berkeley, CA, April 30–May 4.

[3] Brooks, D., and Martonosi, M., 2001, "Dynamic Thermal Management for High-Performance Microprocessors," Proceedings of the 7th International Symposium on High-Performance Computer Architecture (HPCA), Monterrey, Mexico, January 19–24, pp. 171–182.

[4] Hsu, J. T., and Vu-Quoc, L., 1996, "A Rational Formulation of Thermal Circuit Models for Electrothermal Simulation—Part I: Finite Element Method," IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications, **43**(9), pp. 721–732.

[5] Stan, M. R., Skadron, K., Barcella, M., Huang, W., Sankaranarayanan, K., and Velusamy, S., 2002, "Hotspot: A Dynamic Compact Thermal Model at the Processor-Architecture Level," Microelectron. J., **34**(12), pp. 1153–1165.

[6] Hsu, J. T., and Vu-Quoc, L., 1996, "A Rational Formulation of Thermal Circuit Models for Electrothermal Simulation—Part II: Model Reduction Techniques," IEEE Trans. Circuits Syst., I: Fundam. Theory Appl., **43**(9), pp. 733–744.

[7] Wang, J. M., and Nguyen, T. V., 2000, "Extended Krylov Subspace Method for Reduced Order Analysis of Linear Circuits With Multiple Sources," Proceedings of the 37th Annual Design Automation Conference (DAC), Los Angeles, CA, June 5–9, pp. 247–252.

[8] Sofia, J. W., 1995, "Analysis of Thermal Transient Data With Synthesized Dynamic Models for Semiconductor Devices," IEEE Trans. Compon., Packag., Manuf., **18**(1), pp. 39–47.

[9] Christiaens, F., and Beyne, E., 1996, "Transient Thermal Modeling and Characterization of a Hybrid Component," 46th Electronic Components and Technology Conference (ECTC), Orlando, FL, May 28–31, pp. 154–164.
[10] Christiaens, F., Beyne, E., Vandevelde, B., Roggen, J., Mertens, R., Puymbroeck, J. V., Heerman, M., and Berghmans, J., 1997, "Compact Transient Thermal Models for the Polymer Stud Grid Array," Eurotherm Seminar No. 58, Thermal Management of Electronic Systems, Nantes, France, September 24–26.
[11] Szekely, V., and Bien, T. V., 1988, "Fine Structure of Heat Flow Path in Semiconductor Devices: A Measurement and Identification method," Solid-State Electron., **31**(9), pp. 1363–1368.
[12] Yang, Y., Master, R., Tosaya, E., and Touzelbaev, M., 2012, "Transient Frequency-Domain Thermal Measurements With Applications to Electronic Packaging," IEEE Trans. Compon., Packag., Manuf. Technol., **2**(3), pp. 448–456.
[13] Szekely, V., 1998, "Identification of RC Networks by Deconvolution—Chances and Limits," IEEE Trans. Circuits Syst., I: Fundam. Theory Appl., **45**(3), pp. 244–258.
[14] Schweitzer, D., 2009, "A Fast Algorithm for Thermal Transient Multisource Simulation Using Interpolated Zth Functions," IEEE Trans. Compon. Packag. Technol., **32**(2), pp. 478–483.
[15] Gerstenmair, Y. C., and Wachutka, G., 2000, "Calculation of the Temperature Development in Electronic Systems by Convolution Integrals," 16th Annual IEEE Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM), San Jose, CA, March 21–23, pp. 50–59.
[16] Szekely, V., 1991, "On the Representation of Infinite-Length Distributed RC One-Ports," IEEE Trans. Circuits Syst., **38**(7), pp. 711–719.
[17] Carslaw, H. S., and Jaeger, J. C., 1959, *Conduction of Heat in Solids*, 2nd ed., Oxford University Press, Oxford, UK.