SPICE-Compatible Thermal Simulation with Lumped Circuit Modeling for Thermal Reliability Analysis based on Modeling Order Reduction

Ting-Yuan Wang  
Electrical and Computer Engineering  
University of Wisconsin-Madison  
Madison, WI 53706, USA  
wangt@cae.wisc.edu

Charlie Chung-Ping Chen  
Graduate Institute of Electronics Engineering & Department of Electrical Engineering  
National Taiwan University, Taipei 106, Taiwan  
cchen@cc.ee.ntu.edu.tw

Abstract

With the growing power dissipation in modern high performance VLSI designs, nonuniform temperature distribution and limited heat-conduction capability have caused thermal induced performance and reliability degradation. However, the problem modeled by finite difference method for interconnect reliability analysis has huge size if we require the resolution with wire width. In addition, the generated lumped circuit has significant number of input sources, and the bottleneck of traditional model reduction methods is the big number of input ports. In this paper, we propose a method of SPICE-compatible thermal simulation for interconnect reliability analysis. The lumped thermal circuit modeling with adaptive approach is used to reduce the problem size. The improved extended Krylov subspace (IEKS) method, independent of the number of input ports, is used for thermal simulation. The experimental results show that our method provides highly accurate results with performance improvement $15\times$ over T-Spice for the problem with node number 72428.

1 Introduction

The ever-increasing demands for more functionality and higher speed have pushed the very large scale integration (VLSI) industry towards more aggressive scaling [1]. This trend has lead to the rapid increase of power consumption and heat generation. As a result, nonuniform temperature distribution on chip has become more and more serious [2] [3] [4]. Without thermal management, thermal problems not only lead to timing failure but also degrade chip reliability.

Electromigration (EM) is the main reliability concern and will become a more limiting factor of IC designs. The power generated by self-heating effect in interconnects only contributes a small part of the chip power consumption. However, the impacts of thermal effects on the reliability of interconnects are serious. Therefore, EM must be addressed together with a thermal reliability modeling and is recognized in the International Technology Roadmap for Semiconductors (ITRS) 2002 update as one of the difficult challenges [1]. In order to diagnose the thermal reliability and improve the design quality, an efficient chip-level three-dimensional (3-D) thermal simulator is crucial to success for the VLSI designs.

There are some challenges for the thermal modeling and simulation. First, the uniform heat distribution in a chip does not guarantee the uniform temperature profile due to the complex 3-D nature of heat spreading and the complicated boundary conditions. Simulation runtime and memory usage are another issues due to the large size of integrated circuit systems. Several approaches have been proposed. A full-chip thermal simulation was presented in [5] solving function-block size problem. The finite difference method (FDM) with equivalent RC model has been presented [6] [7]. However, due to the large size of matrix, the direct matrix-solving algorithms have runtime and memory usage problems for large scale systems. A thermal simulation method based on model reduction was presented to improve the runtime [8]. However, the bottleneck in such method is that the number of input sources can cause big Krylov subspace. This is especially difficult for equivalent thermal circuit which has a large number of equivalent independent current sources. An efficient full-chip thermal simulator based on alternating direction (ADI) method was presented [9] [10]. However, for interconnect related applications such as thermal reliability analysis, the generated problem is huge due to the discretization size of interconnects.

In this paper, we present a SPICE-compatible method of thermal simulation with lumped thermal circuit modeling for interconnect reliability analysis based on the technique of IEKS. First, lumped thermal circuit with adaptive approach is used to model the problem. The complicated packaging, heat sinks, and cooling system are modeled as equivalent thermal resistors. Substrate is discretized according to the temperature gradient and modeled with equivalent thermal elements. Interconnects are also modeled as lumped elements. With this modeling method, the problem size can be reduced and the wire temperature can be obtained. Second, the problem is solved by the IEKS method. Unlike the tradition model reduction method whose runtime is heavily dependent on the number of input ports, the runtime of the proposed method is independent of that. In addition, there is no moment shifting for IEKS method.

The remainder of the paper is organized as follows. The modeling method and problem formulation is presented in Section 2. The multiport model reduction method, IEKS, is presented in
2.2 Substrate Modeling

is governed by the heat conduction equation \[11\]

ture distribution in a chip including substrate and interconnects

can be obtained

where the material,

age \[12\]. The effective heat transfer coefficient in the direction

are not isothermal, 3-D model is needed for better accuracy to

the package surfaces are held isothermal. If the package surfaces

placed by

(3)

Second, packaging, heat sinks, and cooling systems are mod-

ulated as 1-D equivalent thermal resistance network. Suppose that

volume is
generated is

of (1) can be described from the law of energy conservation. For

are a lot of power sources, the equivalent thermal circuit has a large

is equivalent to current flow in electrical circuit. Since there are

the KCL in electrical circuit at node \(i,j,k\),

\(\Delta V = \Delta x \Delta y \Delta z\) is the control volume of node \(i,j,k\),

\(A_x = \Delta y \Delta z\), \(A_y = \Delta x \Delta z\), and \(A_z = \Delta x \Delta y\). The physical meaning of (4) is clear from the law of energy conservation

\[
\frac{dE}{dt} + Q_1 + Q_2 + Q_3 + Q_4 + Q_5 + Q_6 = Q_p,
\]

where \(\frac{dE}{dt} = \rho C_p \Delta V \frac{dT_{i,j,k}}{dt}\), \(Q_1 = \frac{\Delta V}{\Delta y} (T_{i,j,k} - T_{i,j-1,k})\),

\(Q_2 = \frac{\Delta V}{\Delta y} (T_{i,j,k} - T_{i,j+1,k})\), \(Q_3 = \frac{\Delta V}{\Delta y} (T_{i,j,k} - T_{i-1,j,k})\),

\(Q_4 = \frac{\Delta V}{\Delta y} (T_{i,j,k} - T_{i,j+1,k})\), \(Q_5 = \frac{\Delta V}{\Delta y} (T_{i,j,k} - T_{i,j-1,k})\),

\(Q_6 = \frac{\Delta V}{\Delta y} (T_{i,j,k} - T_{i,j+1,k})\), and \(Q_p = \Delta V g_{i,j,k}\). Then (4) can be translated into the equivalent lumped thermal circuit as shown in Figure 1:

\[
C^0 \frac{dT_{i,j,k}}{dt} + T_{i+1,j,k} - T_{i,j,k} + T_{i,j+1,k} - T_{i,j,k} + T_{i,j,k} - T_{i-1,j,k} + T_{i,j,k} - T_{i,j,k} = Q_p,
\]

where \(C^0 = \rho C_p \Delta V\) is the thermal capacitance, \(Q_p\) is the heat flow coming from power generated in \(\Delta V\), and \(R_{i+1,j,k}^0 = \Delta x/(\kappa A_x)\), \(R_{i,j+1,k}^0 = \Delta y/(\kappa A_y)\), and \(R_{i,j,k+1}^0 = \Delta z/(\kappa A_z)\) are the thermal resistances. The analogy between electrical and thermal circuits is shown in Table 1. In fact, (6) is equivalent to the KCL in electrical circuit at node \((i,j,k)\) because heat flow is equivalent to current flow in electrical circuit. Since there are a lot of power sources, the equivalent thermal circuit has a large number of current sources.
of energy stored in a control volume causing the temperature increase is 
\[ \frac{dE}{dt} = \frac{1}{8} \Delta V \rho c p \frac{dT}{dt} + \frac{1}{8} \Delta V \rho c p + \frac{1}{8} \Delta V \rho c p \frac{T^2}{2}. \]

The power generated in control volume is 
\[ Q_p = \frac{1}{8} \Delta V g_1 + \frac{1}{8} \Delta V g_2. \]
The rate of heat conduction from \( T_{k,j,N} \) to \( T_{i-1,j,N} \) and \( T_{i+1,j,N} \) are 
\[ Q_1 = \frac{\rho c p}{\kappa} \frac{\Delta A_z}{\Delta y} (T_{k,j,N} - T_{i-1,j,N}) \]\nand 
\[ Q_2 = \frac{\rho c p}{\kappa} \frac{\Delta A_z}{\Delta x} (T_{i,j,N} - T_{i+1,j,N}), \]
respectively. The other terms can be calculated similarly: 
\[ Q_3 = \frac{\rho c p}{\kappa} \frac{\Delta A_z}{\Delta y} (T_{i,j,N} - T_{i-1,j,N}), \]
\[ Q_4 = \frac{\rho c p}{\kappa} \frac{\Delta A_z}{\Delta y} (T_{i-1,j,N} - T_{i+1,j,N}), \]
\[ Q_5 = \frac{\rho c p}{\kappa} \frac{\Delta A_z}{\Delta x} (T_{i,j,N} - T_{i+1,j,N-1}), \]
The equivalent thermal convection from \( T_{i,j,k} \) to \( T_a \) is 
\[ Q_6 = h_{e+}^c \Delta A_z (T_{i,j,N} - T_a), \]
where \( h_{e+}^c \) is the effective heat transfer coefficient. From energy conservation (5) and equivalent lumped thermal circuit like (6), we can obtain the equivalent thermal circuit as shown in Figure 2.

**Figure 2. The example of non-homogeneous case and boundary conditions.**

### 2.4 Interconnect Modeling

Figure 3 shows a metal wire on an insulator, in which the width, length, thickness, and thermal conductivity are \( w_m, l, t_m, \) and \( \kappa_m \), respectively. The thickness and the thermal conductivity of the underlying insulator are \( t_{ins} \) and \( \kappa_{ins} \). The wire resistivity is temperature dependent and can be described as follows:
\[ \rho_m(T_m) = \rho_o[1 + \beta(T_m - T_o)], \]
where \( \rho_o \) is the wire resistivity at reference temperature \( T_o \), and \( \beta \) is the temperature coefficient of resistivity. Here we suppose that, e.g., C4/CBGA package, the primary heat transfer path is through substrate and heat sinks to the ambient, and the secondary heat transfer path is through the solder balls and PCB to the ambient. The heat conduction from nearby interconnects are ignored. The detailed discussions about the accuracy compared to other models can be found in [13].

**Figure 3. Interconnect lumped model.**

The heat flow generated by self-heating in wire segment \( l \) can be expressed as
\[ Q_p(T_m) = I^2 R = I^2 \rho_m(T_m) \frac{l}{t_m w_m} = Q_{p_e} + Q_{p_e_c}, \]
where \( I \) is the current flowing through the wire, \( Q_{p_e} = I^2 \rho_o(1 - \beta T_o) l / t_m w_m \) is the constant current source, and \( Q_{p_e_c} = (I^2 \rho_0 \beta l / t_m w_m) T_m \) is the voltage-controlled current source due to the temperature dependent resistivity. The thermal resistor along the interconnect has resistance
\[ R_0 = \frac{1}{\kappa_m l t_m w_m}. \]
The primary heat transfer path can be modeled by equivalent thermal resistor with resistance
\[ R_{ins}^0 = \frac{1}{\kappa_{ins} t_{ins} l w_m}. \]
The secondary heat transfer path will be modeled similarly with resistance
\[ R_s^0 = \frac{1}{h_{e+}^c l w_m}. \]
Figure 3 shows the equivalent lumped thermal circuit. The other cases of interconnect modeling can be derived similarly.
2.5 Adaptive Approach

In order to keep the reasonable size of the formulated problem, we will use the adaptive approach to discretize the substrate. In the region of high temperature gradients, finer discretization size is used to reach the accuracy. On the contrary, coarser grid will be used in the region of low temperature gradient to keep the problem size small. To decide the regional discretization size of the chip, we need to know the rough temperature distribution. The temperature profile of the substrate to the internal states. In the thermal circuit, there are no equivalent inductors, but there are a lot of current sources due to the significant number of heat sources. The conductance matrix is the input adjacency matrix mapping the sources to the internal states. In the thermal circuit, there are no equivalent inductors, but there are a lot of current sources due to the significant number of heat sources. The conductance matrix \( G \) and capacitance matrix \( C \) are symmetric and positive definite.

3 Multiport Model Reduction

Model order reduction generates an analytic model which is a compact representation of original circuits by matching their moments or poles. After taking the Laplace transformation of (12), we have

\[
Gx + C \dot{x} = Bu, \tag{12}
\]

where \( G \) and \( C \) represent the conductance and capacitance matrices, \( x \) is the vector of node voltages, \( u \) is the vector of independent current sources, \( B \) is the input adjacency matrix mapping the sources to the internal states. In the thermal circuit, there are no equivalent inductors, but there are a lot of current sources due to the significant number of heat sources. The conductance matrix \( G \) and capacitance matrix \( C \) are symmetric and positive definite.

Apply the Taylor series expansion at zero frequency on both side of (13), we have

\[
(G + sC)(m_0 + m_1 s + m_2 s^2 + \cdots) = B(u_0 + u_1 s + u_2 s^2 + \cdots), \tag{14}
\]

where \( m_i \) and \( u_i \), the coefficients of the \( i_{th} \) term in the Taylor series, are known as the \( i_{th} \) moment of \( x \) and \( u \), respectively. Moment matching is a method to represent the finite unknown moments of the left hand side of (14) in terms of the known moments of the right hand side. In standard method like PRIMA [14], the sources are impulse sources in the moment matching process in order to preserve the input-output transfer characteristics. The impulse sources are constant in the frequency domain and have contribution only in the initial vector for the Krylov iteration. Then (14) can be rewritten as

\[
(G + sC)(m_0 + m_1 s + m_2 s^2 + \cdots) = Bu_0. \tag{15}
\]

This results in an iterative relationship between the moments: \( Gm_0 = Bu_0, \) \( Gm_i + Cm_{i-1} = 0. \) However, there is numerical stability problems in such moment matching, especially in higher order iterations. To avoid the numerical errors, an orthogonal bases \( V \) based on Krylov subspace, which is defined as \( K^r(A, R, q) = \text{colsp}(R, AR, A^2R, \ldots, A^{q-1}R) \) with \( A = -G^{-1}C \) and \( R = G^{-1}B, \) is constructed to span the same subspace spanned by finite moments of \( x(s). \) Thus the order-reduced model can be obtained by projecting the original system onto the Krylov subspace by congruent transformation

\[
\tilde{G} \dot{x} + s \tilde{C} \dot{x} = \tilde{B} u \tag{16}
\]

where \( \tilde{G} = V^T GV, \) \( \tilde{C} = V^T CV, \) and \( \tilde{B} = V^T B. \)

However, the bottleneck in such method is the number of sources in the input vector \( u \) due to the big size of \( B \) causing big Krylov subspace. This is especially difficult for solving the equivalent thermal circuit which has a large number of independent current sources.

3.1 IEKS

The proposed method is inspired by the EKS method [15] which directly calculates the orthogonalized moments of the response when multiple sources are turned on at the same time. Therefore, unlike PRIMA whose runtime is heavily dependent on the port number, the runtime of EKS is independent of that. EKS models the piece-wise-linear (PWL) independent sources as a sum of delayed ramps in frequency domain

\[
u(s) = \frac{1}{s^2} \sum_{i=1}^{N} r_i \exp(-\beta_i s). \tag{17}\]

This expression contains \( 1/s \) and \( 1/s^2 \) terms. In general, the traditional Krylov subspace methods start the moment matching from the \( 0^{th} \) moment. EKS extends the Krylov subspace by shifting the moments in the frequency spectrum. However, the moment shifting in EKS is tedious and error-prone. Therefore, an improved moment calculation method is proposed for thermal simulation, IEKS, which ensures the \( -1^{st} \) and \( -2^{nd} \) order moments are zero for arbitrary finite time PWL waveforms and the moment shifting process can be eliminated. Since only a specific time period is interested for a specific time period, the finite-time assumption is quite general. We believe this procedure is numerically more sound than the original approach.

**Figure 4. PWL waveform of the input source**

For a given finite-time PWL source, the moment representation of IEKS with \(-1^{st}\) and \(-2^{nd}\) order are zero. Suppose that...
the given finite-time PWL source \( u(t) \) is
\[
u(t) = \sum_{i=0}^{N} \left\{ [a_i + \gamma_i(t - t_i)] \mathcal{E}(t-t_i) - [a_{i+1} + \gamma_i(t - t_{i+1})] \mathcal{E}(t-t_{i+1}) \right\}, \quad (18)
\]
where \( \gamma_i = (a_{i+1} - a_i)/(t_{i+1} - t_i) \) and \( \mathcal{E}(t-t_i) \) is the unit-step function with \( t_i \) delay. By taking the Laplace transform and Taylor expansion of (18), we have
\[
u(s) = \frac{1}{s} \sum_{i=0}^{N} \left[ a_i s \sum_{l=0}^{\infty} (-1)^l \frac{t_i^l}{l!} s^l + \gamma_i \sum_{l=0}^{\infty} (-1)^l \frac{t_{i+1}^l}{l!} s^l \right. \\
- \left. a_{i+1} s \sum_{l=0}^{\infty} (-1)^l \frac{t_{i+1}^l}{l!} s^l \right] - \gamma_i \sum_{l=0}^{\infty} (-1)^l \frac{t_{i+1}^l}{l!} s^l \right)  \quad (19)
\]
Let \( \tilde{u}_i \) denote the coefficient of the \( s^l \) term, then (19) can be simplified as
\[
\mathcal{L}(u(t)) = \left\{ \tilde{u}_{-2} s^{-2} + \tilde{u}_{-1} s^{-1} - \tilde{u}_0 + \tilde{u}_1 s + \tilde{u}_2 s^2 + \cdots + \tilde{u}_m s^m + \cdots \right\}. \quad (20)
\]
After the detailed calculation, the first two coefficients, \( \tilde{u}_{-2} \) and \( \tilde{u}_{-1} \), are zero. The process of IEKS to calculate the first \( M \) moments is summarized in Table 2. The process of generating an orthogonal basis \( V \) for the corresponding moments is similar to [15].

**Algorithm Find the first \( M \) moments of PWL source**

Given
- PWL sources \( \{a_0, \ldots, a_{N+1}; t_0, \ldots, t_{N+1}\} \)
- The number of moments to calculate \( M \)

for \( i = 0 \ldots N \)
\[
\gamma_i = \frac{(a_{i+1} - a_i)}{(t_{i+1} - t_i)}
\]
end

for \( m = 1 \ldots M \)

for \( i = 0 \ldots N+1 \)
\[
\beta_i^m = \frac{(-t_i)^m}{m!}
\]
end

\[
\tilde{u}_{m-1} = \left. (a_0 - \gamma_0 \frac{t_0}{m} \beta_0^{(m-1)}) \right|_{m=1} ^{m} - \sum_{i=0}^{N-1} \left( \gamma_i - \gamma_{i+1} \right) \beta_i^{(m-1)} - (a_{N+1} - \gamma_N \frac{t_{N+1}}{m} \beta_N^{(m-1)}) \right|_{m=1} ^{m}
\]
end

**Table 2.**

After generating a system transform matrix \( V \), the original system is transformed into a compact description
\[
\tilde{G} \tilde{X} + \tilde{C} \frac{d}{dt} \tilde{X} = \tilde{B} u. \quad (21)
\]
The compact form can be solved quickly in the time domain by standard integration algorithms. The solution of the original system can be derived by \( X = VX \).

**4 Experimental Results**

The proposed simulator with IEKS method was implemented with C++ language, and executed on a notebook with a 1.6 GHz Pentium 4 processor and 640 MB memory.

The extraction runtime of equivalent thermal circuits is not included in the runtime comparison. Figure 5 and Table 3 show the runtime comparison of the thermal simulation with the IEKS method and Tanner T-Spice for 1000 iterations. In the case of circuit-8 with 72428 nodes, the simulation with IEKS method took 221.21 sec. It is about 15 times faster than T-Spice.

**Table 3. Runtime of Thermal Circuits**

<table>
<thead>
<tr>
<th>Circuit</th>
<th>Node number</th>
<th>IEKS (s)</th>
<th>T-Spice (s)</th>
<th>Speedup (X)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Circuit-1</td>
<td>897</td>
<td>0.260</td>
<td>7.730</td>
<td>29.731</td>
</tr>
<tr>
<td>Circuit-2</td>
<td>3217</td>
<td>1.823</td>
<td>25.720</td>
<td>14.109</td>
</tr>
<tr>
<td>Circuit-3</td>
<td>7067</td>
<td>6.750</td>
<td>68.350</td>
<td>10.126</td>
</tr>
<tr>
<td>Circuit-4</td>
<td>12180</td>
<td>15.733</td>
<td>165.490</td>
<td>10.319</td>
</tr>
<tr>
<td>Circuit-5</td>
<td>19228</td>
<td>31.355</td>
<td>346.220</td>
<td>11.042</td>
</tr>
<tr>
<td>Circuit-6</td>
<td>27668</td>
<td>53.397</td>
<td>595.060</td>
<td>11.012</td>
</tr>
<tr>
<td>Circuit-7</td>
<td>97608</td>
<td>85.392</td>
<td>1059.080</td>
<td>12.381</td>
</tr>
<tr>
<td>Circuit-8</td>
<td>72428</td>
<td>221.210</td>
<td>3384.450</td>
<td>15.300</td>
</tr>
</tbody>
</table>

**Figure 5. Runtime comparison of the simulation with IEKS method and T-Spice.**

Next consider a chip with size \( 11.3 \text{mm} \times 14.4 \text{mm} \) and total power consumption \( 48W \). The chip is discretized with \( \Delta x = 120 \mu m, \Delta y = 120 \mu m, \) and \( \Delta z = 20 \mu m \). The effective heat transfer coefficients, \( h_i^c \), are supposed to be \( 7 \times 10^3 \text{ W/m}^2\text{K} \) in the primary heat transfer path, and no heat transfer around the side faces. The simulation runs 1000 iterations with time increment \( \Delta t = 10^{-3} \text{ sec} \), and the results are shown in Figure 6. The average temperature is \( 52.41^\circ C \), but the highest temperature is about \( 75^\circ C \) which is influenced by the effective heat transfer coefficients and the location of power sources. There are three main parameters that affect the temperature: the board-level component population (thermal loading), the heat sink style and design, and the air velocity on the components and/or the heat sink [16]. From the results, we can observed that there are hot spots in the chip even though the average temperature may be low due to the improvement of the cooling technologies. These hot spots are the reliability concerns.

The transient simulation with IEKS method and T-Spice is tested and is shown in Figure 7. The difference of the results...
5 Conclusion
In this paper, we propose a SPICE-compatible method of thermal simulation with lumped thermal circuit modeling for interconnect reliability analysis based on the IEKS method. Unlike the traditional model reduction method whose runtime is heavily dependent on the number of input ports, the runtime of the proposed method is independent of that. Therefore, the simulation runtime is improved significantly. In addition, the lumped thermal circuit modeling with adaptive approach reduces the formulated problem size. The proposed method can also be used for other interconnect related thermal applications.

6 Acknowledge
This work was partially funded by National Science Foundation under grants CCR-0093309 & CCR-0204468 and National Science Council of Taiwan, R.O.C. under grant NSC 92-2218-E-002-030. The authors would like to thank Tsung-Hao Chen for the help of solver development and Yu-Min Lee for the help of IEKS.

References


