An Efficient Leakage-Aware Thermal Simulation Approach for 3D-ICs Using Corrected Linearized Model and Algebraic Multigrid

Chao Yan[1], Hengliang Zhu[*1], Dian Zhou[2] and Xuan Zeng[*1]
[1]State Key Lab of ASIC & System, Microelectronics Dept., Fudan University, Shanghai, China
[2]Electrical Engineering Dept., University of Texas at Dallas, Richardson, USA

***Abstract*—Thermal control has become a great challenge for 3D-ICs due to the ever increasing power density and 3D integration. Among techniques to address the problem, fast thermal simulation approach is basically required to accurately characterize the runtime temperature variations of 3D-ICs. In this paper, we propose an accurate and fast leakage-aware thermal simulation approach for 3D-ICs with consideration of both heatsink cooling and microfluidic cooling. First, the proposed approach is based on a corrected linearized model for leakage power approximation, which is proved to be equivalent to the Newton-Chord method for solving nonlinear algebra equations. A convergence comparison is presented in this paper to show that such approach is more efficient than other methods for leakage-aware thermal simulation. Second, an aggregation-based algebraic multigrid (AMG) preconditioned iterative linear solver is adopted that greatly reduces the computation time for solving the linear equations during calculation, which makes the proposed approach even more efficient. Numerical experiments show that the proposed approach can achieve $8\times - 139\times$ speedup in comparison with the state-of-the-art methods, and with almost negligible average temperature error no more than 0.025K and maximum temperature error no more than 0.095K.**

I. Introduction

With ever increasing demand for higher integration and performance, three-dimensional integrated circuits (3D-ICs) are envisioned as an attractive solution in nanometer scale design. By stacking device layers through direct vertical paths, 3D-ICs significantly reduce average wire delay and power consumption compared with traditional 2D chips. However, the stacked multi-layer 3D-IC structure poses great challenges in thermal control due to higher power density and large junction-to-ambient thermal resistance [1]. The considerable heat flux can lead to high temperature and thermal stress, thus limiting the performance and reliability of the devices. To address the thermal issue, heatsink-cooling using thermal through silicon via (TSV) [2] and microfluidic-cooling using inter-tier microchannel [3] are generally adopted in 3D-IC design. Thermal simulation approach with consideration of heatsink-cooling and microfluidic-cooling is basically needed to characterize the runtime temperature variations of 3D-ICs.

For integrated circuits with conventional heatsink cooling, thermal simulation approaches mostly fall into the categories of differential approaches and Green's function based approaches. Differential approaches, including finite difference (FD) method, finite volume (FV) method etc., are widely used in thermal simulation, which usually involve an analogy between heat and electrical conduction to transform the discretized heat diffusion equation into a compact thermal model for simulation. Typical examples for differential approaches include HotSpot [4] for steady-state thermal simulation and 3D Thermal-ADI [5] for transient thermal simulation. The Green's function based approaches use the convolution of Green's function and power map to evaluate the temperature profile of the chip, which are faster than differential approaches by taking advantages of the uniform structure of the chip. An efficient thermal simulation algorithm is proposed in [6], which is a combination of Green's function method, discrete cosine transform (DCT) and frequency domain computation. Power Blurring in [7] uses a matrix convolution technique in analogy with image blurring for fast thermal analysis of the packaged integrated circuits and power devices.

For 3D-IC with microfluidic cooling, FD approach is generally more efficient than Green's function based approaches, which is mostly due to the complex structure of the inter-tier microchannels. One of the well-developed thermal simulation tools for 3D-IC with microfluidic cooling is 3D-ICE in [8], which proposes a compact thermal model for simulation by FD discretization and modeling the heat convection of the liquid as temperature controlled heat source.

On the other hand, the leakage power in today's VLSI becomes comparable to dynamic power. According to [9], an estimated $30\% - 50\%$ of the total power is due to leakage in $45nm$ and beyond, which makes the consideration of leakage power indispensable. HotSpot [4] and LightSim [10] are among the few works that take leakage power into consideration. HotSpot adopts a simple iterative approach for leakage-aware thermal simulation by repeatedly evaluating the temperature and leakage power dissipation, but with relatively slow convergence rate. LightSim proposes a leakage-aware Green's function and uses Hankel transform to speed up the computation process, which is ultrafast for 2D chips' leakage-aware thermal simulation. The adoption of a universal leakage-aware Green's function requires that each chip layer is uniformly structured and the devices in the active layers are uniformly distributed with uniform linear leakage model, which makes it difficult to be applied in real applications of 3D-IC thermal simulation, especially for 3D-ICs with

[*]Corresponding authors: {hlzhu, xzeng}@fudan.edu.cn.

978-3-9815370-8-6/17/$31.00 ©2017 IEEE

microfluidic cooling.

In this paper, a fast and accurate leakage-aware thermal simulation approach is proposed for 3D-ICs with consideration of both heatsink cooling and microfluidic cooling. The proposed approach is based on the FD discretization, which is more flexible in handling the various kinds of complex 3D structures including inter-tier microchannels. To solve the heat convection-diffusion equation with leakage power dissipation, we propose an iterative approach based on a corrected linearized model for leakage power approximation, which is shown to be equivalent to the Newton-Chord method for solving nonlinear algebra equations. Convergence comparison presented in this paper will show that such approach is more efficient than other iterative methods for leakage-aware thermal simulation. To further improve the computational efficiency, we adopt an aggregation-based AMG preconditioned iterative linear solver to reduce the computation time for solving the linear equations during calculation. The proposed approach is shown to be not only accurate but also fast, which is $8\times-139\times$ faster than the conventional leakage-aware thermal simulators with almost negligible average temperature error no more than 0.025K.

The rest of the paper is organized as follows. Section II presents the problem definition of leakage-aware thermal simulation and the related works. The proposed approach is presented in Section III. Experimental results are shown in Section IV, and the conclusions are drawn in Section V.

II. Background and Related Work

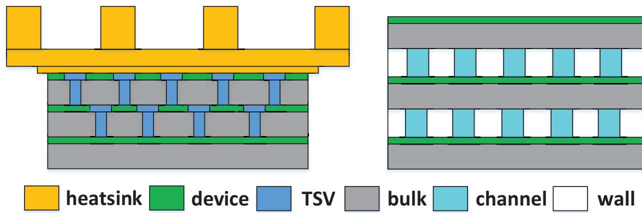

Fig. 1. 3D-ICs with heatsink cooling and microfluidic cooling.

A. Thermal Model for 3D-ICs

Fig. 1 shows the structures of 3D-ICs with heatsink-cooling and microfluidic-cooling architectures. In heatsink-cooled 3D-ICs, thermal TSVs are used for heat dissipation from each die to heatsink and spreader. In microfluidic-cooled 3D-ICs, the microchannel cavities are etched at the back of dies and liquid coolant flows through the channels carrying away the heat.

For 3D-ICs with heatsink cooling, heat diffusion equation in solid can be formulated as

$$C_v \frac{\mathrm{d}T}{\mathrm{d}t} + (-k\nabla^2 T) = \dot{q}, \tag{1}$$

where k is the thermal conductivity of the material, C_v is the corresponding volumetric specific heat, T is the temperature and $\dot{q}$ is the heat generation per unit volume.

For 3D-ICs with microfluidic cooling, the heat convection-diffusion equation in fluid is

$$C_v \frac{\mathrm{d}T}{\mathrm{d}t} + \nabla \cdot (-k\nabla T) + C_v \vec{u}\,\nabla T = \dot{q}, \tag{2}$$

where $\vec{u}$ is the velocity of the fluid.

By applying finite difference (FD) discretization, and with proper boundary conditions [8], we can have the following system of ordinary differential equations,

$$\mathbf{G}\mathbf{T}(t) + \mathbf{C}\dot{\mathbf{T}}(t) = \mathbf{P}(t), \tag{3}$$

where $\mathbf{T}(t) \in R^m$ is the unknown temperature at time t, $\mathbf{C} \in R^{m\times m}$ is a diagonal matrix of thermal capacitances at the grids, $\mathbf{P}(t) \in R^m$ is a vector of heat generation of the corresponding grid at time t, and m is the number of grids for discretization. $\mathbf{G} \in R^{m\times m}$ is called the thermal conductance matrix for the heat diffusion equation in (1), and for the heat convection-diffusion equation in (2), convection terms will be added into $\mathbf{G}$ which becomes an asymmetric matrix.

In the rest of the paper, we will use steady-state thermal simulation as the example to explain the idea of the proposed method. The problem of transient thermal simulation can be solved similarly. Thus (3) can be simplified as

$$\mathbf{G}\mathbf{T} = \mathbf{P}. \tag{4}$$

B. Problem Definition of Leakage-Aware Thermal Simulation

With shrinking technology node, dynamic power has substantially decreased due to the significantly reduced voltage, while leakage power is exacerbated and becomes one of the major sources of the power dissipation. Therefore, it's necessary to consider the contribution of leakage power when solving the equation (4), which will result in a nonlinear equation as

$$\mathbf{G}\mathbf{T} = \mathbf{P}_{dyn} + \mathbf{P}_{leak}(\mathbf{T}), \tag{5}$$

where $\mathbf{P}_{dyn}$ and $\mathbf{P}_{leak}(\mathbf{T})$ represent the dynamic power and leakage power at the corresponding position, respectively. For each grid at active layer of the chip, the accurate leakage power is generally given by the BSIM4 [11] model:

$$P_{leak}(T) = k v_T^2 e^{\frac{V_{GS}-V_{th}-V_{off}}{\eta v_T}} (1 - e^{\frac{-V_{DS}}{v_T}}), \tag{6}$$

where k is a technology-dependent constant, v_T is thermal voltage, V_{GS} is the gate-to-source voltage, V_{th} is the threshold voltage, V_{off} is the offset voltage in the subthreshold region, η is the subthreshold swing coefficient for the transistor and V_{DS} is the drain-to-source voltage.

C. Related Work

Since leakage power consumption influences temperature, which in turn affects leakage power, a conventional approach to consider leakage in thermal simulation is to get the temperature profile from an initial power map, then evaluate the corresponding leakage power, and repeat the process until the temperature and leakage power converge. Such *simple iterative approach* is adopted in many works, e.g. HotSpot [4], however converges slowly to the solution of (5).

LightSim [10] adopts a linear model for leakage power approximation and proposes a leakage-aware Green's function for thermal simulation, which is ultrafast for leakage-aware thermal simulation. However, the adoption of a universal

Green's function is based on the uniform structure of chip layers. Thus, when thermal TSVs are included, it needs multiple Green's functions to account for the difference between heat conduction of thermal TSVs and that of silicon [7]. And in microfluidic-cooled 3D-ICs, it is difficult to define a proper Green's function to account for the complex structure of microchannels. Moreover, the derivation of leakage-aware Green's function in LightSim further assumes that devices in each active layer are uniformly distributed over the whole layer and with the same linear leakage model. Consequently, for cases when there exist non-device areas (e.g. the areas around the thermal TSVs) in the active layer, or when there are different types of leakage models with different sets of model parameters in the design, the leakage-aware Green's function based approach could easily lead to large errors, as will be shown in Section IV.

Therefore, instead of adopting the Green's function based approach, we propose an efficient leakage-aware thermal simulation approach based on FD discretization and adopt an iterative approach using corrected linearized model and aggregation-based AMG to improve the computational efficiency for solving the nonlinear equation in (5).

III. Proposed Method

The proposed leakage-aware thermal simulation approach will be presented in this section. The major idea is to propose an iterative approach for solving (5) using a corrected linearized model for leakage power approximation, which will be presented in Subsection III-A and III-B. To further improve the computation efficiency, aggregation-based AMG method is adopted for solving the linear equations during calculation, which will be discussed in Subsection III-C.

A. Iterative Approach Using Corrected Linearized Model

1) Linearized Model for $\mathbf{P}_{leak}(\mathbf{T})$: One simple approach for solving the nonlinear equation in (5) is to find a linear approximation of the nonlinear term $\mathbf{P}_{leak}(\mathbf{T})$. By applying Taylor expansion to (6), $\mathbf{P}_{leak}(\mathbf{T})$ in (5) can be approximated by a linearized model, which can be expressed as

$$\mathbf{P}_{leak}(\mathbf{T}) \approx \widehat{\mathbf{P}}_{leak}(\mathbf{T}) = \mathbf{P}_{leak0} + \mathbf{C}_0\mathbf{T}, \qquad (7)$$

where

$$\mathbf{P}_{leak0} = \begin{bmatrix} P_{leak0,1} \\ P_{leak0,2} \\ \cdots \\ P_{leak0,n} \end{bmatrix} \mathbf{C}_0 = \begin{bmatrix} P'_{leak,1} & 0 & \cdots & 0 \\ 0 & P'_{leak,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & P'_{leak,n} \end{bmatrix}.$$

When there are devices in the grid i, $P_{leak0,i}$ and $P'_{leak,i}$ respectively denote the constant term and first order derivative term of the Taylor expansion, otherwise, they are equal to zeros. Substituting (7) into (5) yields a linear equation

$$(\mathbf{G} - \mathbf{C}_0)\mathbf{T} = \mathbf{P}_{dyn} + \mathbf{P}_{leak0}, \qquad (8)$$

which would be easier to solve.

Many existing approaches, e.g. LightSim [10], adopt a linearized model for leakage power approximation with one *single* expansion point for all devices, in other words, $P_{leak0,i}$ and $P'_{leak,i}$ are the same for every grid i in the active layers. However, due to the temperature variations, the accuracy of such linearized model is largely different for devices at different places of the chip.

In order to improve the accuracy, we adopt a linearized model with *multiple* expansion points to approximate $\mathbf{P}_{leak}(\mathbf{T})$. In other words, we choose different expansion points T_i's for different grids in the active layer, and as a result, $P_{leak0,i}$ and $P'_{leak,i}$ would be different for different i's. To determine the expansion points, a thermal simulation without leakage power dissipation is first performed to get a rough estimation of the temperature distribution $\mathbf{T}_0$. This process can be simplified by using a coarser grid discretization in (4), which will result in a smaller linear equation for thermal simulation.

2) Correction to the Linearized Model: Although multiple expansion points improve the model accuracy, the linearized model in (7) could still lead to large errors in thermal simulation. To resolve this problem, we propose a corrected linearized model for leakage power approximation as follows,

$$\mathbf{P}_{leak}(\mathbf{T}) \approx \widehat{\mathbf{P}}^c_{leak}(\mathbf{T}) = \mathbf{P}_{leak0} + \mathbf{C}_0\mathbf{T} + \boldsymbol{\gamma}(\mathbf{T}), \qquad (9)$$

where $\boldsymbol{\gamma}(\mathbf{T})$ represents the high-order term of the approximation to be determined.

Instead of using high-order terms in Taylor expansion, we propose an iterative approach to estimate $\boldsymbol{\gamma}(\mathbf{T})$ in $\widehat{\mathbf{P}}^c_{leak}(\mathbf{T})$. Substituting (9) into (5) yields

$$(\mathbf{G} - \mathbf{C}_0)\mathbf{T} = \mathbf{P}_{dyn} + \mathbf{P}_{leak0} + \boldsymbol{\gamma}(\mathbf{T}). \qquad (10)$$

At the first step of the iteration, the high-order term $\boldsymbol{\gamma}(\mathbf{T})$ is assumed to be zero, and (10) is reduced to the linear equation in (8), of which the solution can be denoted by $\mathbf{T}_1$. Then at the $k+1$ step, the high-order term $\boldsymbol{\gamma}(\mathbf{T})$ can be approximated by the accumulated residual of equation (5), i.e.

$$\boldsymbol{\gamma}(\mathbf{T}) \approx -\sum_{l=1}^{k} \mathbf{F}(\mathbf{T}_l), \qquad (11)$$

where

$$\mathbf{F}(\mathbf{T}) = \mathbf{G}\mathbf{T} - \mathbf{P}_{dyn} - \mathbf{P}_{leak}(\mathbf{T}). \qquad (12)$$

Substituting (11) into (10) for $\mathbf{T} = \mathbf{T}_{k+1}$ yields

$$(\mathbf{G} - \mathbf{C}_0)\Delta\mathbf{T}_k = -\mathbf{F}(\mathbf{T}_k), \qquad (13)$$

for $\Delta\mathbf{T}_k = \mathbf{T}_{k+1} - \mathbf{T}_k$, where the $\mathbf{P}_{leak}(\mathbf{T}_k)$ in $\mathbf{F}(\mathbf{T}_k)$ is given by the accurate BSIM model.

When the residual $\mathbf{F}(\mathbf{T}_n)$ at the nth iteration is small enough, e.g. $\|\mathbf{F}(\mathbf{T}_n)\| < \boldsymbol{\epsilon}$ for a predefined threshold $\boldsymbol{\epsilon}$, we can terminate the iteration and report the final solution as $\mathbf{T} = \mathbf{T}_n$, which actually is the solution of

$$(\mathbf{G} - \mathbf{C}_0)\mathbf{T}_n = \mathbf{P}_{dyn} + \mathbf{P}_{leak0} - \sum_{k=1}^{n-1} \mathbf{F}(\mathbf{T}_k). \qquad (14)$$

In the following subsection, we will prove such iterative approach using corrected linearized model is actually equivalent

to Newton-Chord method, which is a Quasi-Newton method for solving nonlinear equations. Since the leakage power in (6) is not a strongly nonlinear function, the above iterative process will be terminated in a few iterations and produce accurate results, which will be shown in Section IV.

B. Convergence Analysis

In this part, we will first prove that the proposed iterative approach using corrected linearized model is equal to Newton-Chord method. In Newton-Chord method, the updated solution for $\mathbf{g}(\mathbf{x}) = \mathbf{0}$ is given by the following equation [12]:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{J}_g(\mathbf{x}_0)^{-1}\mathbf{g}(\mathbf{x}_k). \tag{15}$$

For our leakage-aware thermal problem (12), the corresponding Jacobian matrix at $\mathbf{T}_0$ is given by

$$\mathbf{J}_F(\mathbf{T}_0) = \mathbf{G} - \mathbf{C}_0. \tag{16}$$

Therefore, (13) can be rewritten as:

$$\mathbf{T}_{k+1} = \mathbf{T}_k - \mathbf{J}_F(\mathbf{T}_0)^{-1}\mathbf{F}(\mathbf{T}_k), \tag{17}$$

which is equivalent to formula of Newton-Chord method in (15). Moreover, Newton-Chord method is proved to have linear convergence rate [12]. Thus, the proposed iterative approach using corrected linearized model is guaranteed to find an accurate result for leakage-aware thermal simulation.

In the following, we provide a convergence comparison of the proposed iterative approach using corrected linearized model, the simple iterative approach in HotSpot and Newton method for solving (5).

1) In HotSpot, the updated solution is derived by

$$\mathbf{G}\mathbf{T}_{k+1} = \mathbf{P}_{dyn} + \mathbf{P}_{leak}(\mathbf{T}_k),$$

which yields

$$\mathbf{G}\Delta\mathbf{T}_k = -\mathbf{F}(\mathbf{T}_k). \tag{18}$$

2) Newton method is a standard approach for solving nonlinear equations, where Jacobian matrix will be re-evaluated in each iteration, that is

$$(\mathbf{G} - \mathbf{C}_k(\mathbf{T}_k))\Delta\mathbf{T}_k = -\mathbf{F}(\mathbf{T}_k), \tag{19}$$

where $\mathbf{G} - \mathbf{C}_k(\mathbf{T}_k)$ is the Jacobian matrix $\mathbf{J}_F(\mathbf{T}_k)$.

Fig. 2 is a one-dimensional example showing how T is updated in three different approaches according to (13), (18) and (19). From Fig. 2, we can see that the slope rate of the update direction of Newton-Chord method is between those of simple iterative approach in HotSpot and Newton method. Therefore, among these three approaches, the simple iterative approach in HotSpot converges most slowly to the solution. It will be demonstrated in Section IV that for leakage-aware thermal simulation, the number of iterations for HotSpot is about two times of those for Newton and Newton-Chord methods, while the number of iterations for Newton and Newton-Chord methods are nearly the same. Although the convergence rate of Newton method is proved to be quadratic [12], it requires additional setup time for Jacobian matrix $\mathbf{G} - \mathbf{C}_k(\mathbf{T}_k)$ at each iteration. Therefore, with nearly the same number of iterations, Newton-Chord method is more efficient than Newton method.

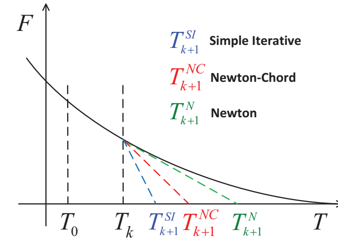

Fig. 2. Solution update in kth iteration for three methods.

C. Aggregation-Based Algebraic Multigrid Method

In the proposed approach for leakage-aware thermal simulation, an efficient linear solver is generally needed for (8) and (13). Conventional direct solvers such as KLU [13] and SuperLU [14] can be applied. However, for large scale problem, KLU or SuperLU is usually time-consuming and memory inefficient. Alternatively, we can adopt algebraic multigrid (AMG) approach, which is well known for its high computation efficiency and fast convergence rate. In this paper, we adopt the aggregation-based AMG preconditioned Conjugate Gradient and Generalized Conjugate Residual (AMG-PCG/AMG-PGCR) [15], [16], [17] to solve the large scale linear equations in (8) and (13) .

The framework of AMG-PCG/AMG-PGCR consists of three parts: aggregation-based coarsening, K-Cycle multigrid preconditioning and using the aggregation-based AMG as an implicit preconditioner for CG/GCR iterative method to solve the linear system.

1) Aggregation-Based Coarsening: The main idea of coarsening is to transfer the fine grid matrix to a coarser one by grouping the fine grid variables, which will produce a prolongation matrix P and a restriction matrix R for grid mapping. An aggregation algorithm is proposed in [15] which gives preference to the strongest negative couplings and is adopted in this paper. In thermal simulation problem, stronger negative coupling means larger thermal conductance, which also implies that the difference between the node temperatures would be smaller.

2) K-Cycle Multigrid Preconditioning: AMG is often adopted as an inexplicit preconditioner for CG/GCR method instead of standalone solver in order to improve the robustness of the approach. In K-Cycle Multigrid, Krylov-subspace smoothing is performed at the end of each level to improve the solution, as shown in Fig. 3, which can achieve guaranteed convergence according to theoretical analysis [16].

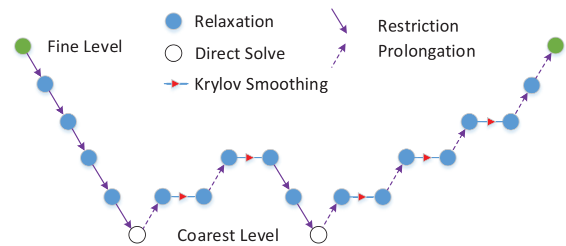

Fig. 3. K-Cycle Multigrid.

3) CG/GCR Iterative Method: The multigrid preconditioning is called by a main iterative routine which is based on CG method if the coefficient matrix is symmetric positive definite, and is based on GCR method in asymmetric cases. For heatsink-cooled 3D-ICs, the thermal conductance matrix $\mathbf{G}$ as well as $\mathbf{G} - \mathbf{C}_0$ are symmetric, and AMG-PCG solver will be adopted. For microfluidic-cooled 3D-ICs, $\mathbf{G} - \mathbf{C}_0$ is no longer symmetric due to the unidirectional flow of heat convection term [8], and AMG-PGCR solver will be used.

IV. Experimental Results

The leakage-aware thermal simulator is implemented in C++ and experiments are conducted on a Linux PC with Intel Core 2 Quad CPUs running at 2.30GHz clock frequency with 4GB memory. Test cases include five microfluidic-cooled 3D-ICs (Case I-V), which are the benchmarks of the 3D-ICON problem from ICCAD 2015 contest [18], and two heatsink-cooled 3D-ICs (Case VI and VII). Case I-III and V have 2 die tiers, Case IV and VI have 3 die tiers and Case VII has 5 die tiers. Detailed geometry and material properties of the chip can be found in [18]. The input pressure drop and liquid cooling network for Case I-V are designed to meet the specifications provided by ICCAD 2015 contest.

A. Limitation of Leakage-Aware Green's Function

In this subsection, we use two simple test cases to show that the assumptions for deriving leakage-aware Green's function as discussed in Section II could lead to large errors. In Case A, there are four types of devices with different leakage model parameters and in Case B, there is about 19.2% of the active layer set to be no-device area. We calculate the temperature distribution of these two cases using uniform linear leakage model, and report their errors in comparison with the real ones in Table III.

From results of Case A, it can be observed that because of the adoption of uniform linear model, the maximum error could be as large as 7.86K. In Case B, due to the existence of no-device area, the maximum error is 12.67K. As a result, for 3D-ICs like Case A and B, it is difficult for leakage-aware Green's function based approach to produce accurate results.

B. Comparison between Different Leakage Models

The accuracy of different linearized models for leakage power approximation against accurate BSIM leakage model is further compared in Table I, including the linearized model with single expansion point (LSEP), the linearized model with multiple expansion points (LMEP) and the proposed corrected linearized model (CLM).

From the first two columns of Table I, it can be observed that LSEP is not accurate enough, of which the maximum error could be as large as 7.49K and 5.52K for Case I and Case IV, respectively. Comparing the results of LMEP with that of LSEP, it can be concluded that LMEP performs better in most cases, and the reason is that LMEP takes the temperature variations among different places of the chip into consideration. However, LMEP still has a maximum error of

TABLE I
Comparison Between LSEP, LMEP, CLM against BSIM

case	LSEP/K		LMEP/K		CLM/K	
	MaxErr	AvgErr	MaxErr	AvgErr	MaxErr	AvgErr
I	7.49	3.21	4.98	2.00	0.095	0.025
II	0.48	0.11	0.54	0.15	0.001	0.0001
III	1.96	1.07	1.36	0.64	0.006	0.002
IV	5.52	2.03	4.77	1.75	0.084	0.020
V	0.42	0.16	0.10	0.03	0.002	0.0003
VI	0.005	0.001	0.05	0.04	0.001	0.0004
VII	2.58	1.91	0.74	0.65	0.002	0.001

4.98K and 4.77K in Case I and IV, respectively. The last two columns of the Table I show that the result of the CLM agrees very well with that of BSIM model, with maximum error no more than 0.095K and average temperature error no more than 0.025K, which is accurate enough for leakage modeling in 3D-ICs.

C. Comparison between Different Linear System Solvers

Here, the AMG preconditioned iterative solver is compared with direct solver KLU [13] and SuperLU [14] in solving a high-dimensional linear system equation. Table II shows the results. And the second column of the Table II presents the number of discretization grid of each test case.

TABLE II
Comparison Between Different Linear Solvers

case	#grid	AMG /s	KLU/s	SuperLU/s
I/II/III/V	0.05M	0.14	4.66	7.22
IV/VI	0.08M	0.28	19.39	33.30
VII	0.14M	0.37	127.08	293.38

From the results, we can observe that AMG preconditioned iterative method is much faster than conventional direct solver, with more than 33× speedup. And the speedup becomes even higher with the rise of the chip size, which is more than 343× for 0.14M grids. Therefore, AMG preconditioned iterative method is preferred for 3D-IC thermal simulation.

D. Accuracy and Efficiency of the Proposed Method

The accuracy and efficiency of different leakage-aware thermal simulation approaches are compared in Table IV, including Newton method, HotSpot (simple iterative with SuperLU) and the proposed CLM with KLU and AMG solvers. The threshold ϵ for convergence ($\|\mathbf{F}(\mathbf{T}_n)\| < \epsilon$) is set as 1e-3 for all these approaches, and the results of Newton method with ϵ =1e-8 are used as the baseline for evaluating the errors. The errors of Newton method with ϵ =1e-3 are small, which are not reported in Table IV, and the errors of CLM+KLU are listed in the last two columns of Table I.

From Table IV, we can observe that CLM+KLU and Newton methods have higher convergence rate than HotSpot. The numbers of iterations for CLM+KLU and Newton are almost 2 times less than that of HotSpot, and CLM+KLU is more than 4 times faster than HotSpot. For Case I-VII, the numbers of iterations for CLM+KLU and Newton are

TABLE III
ASSUMPTIONS FOR LEAKAGE-AWARE GREEN'S FUNCTION COULD LEAD TO LARGE ERRORS

case	Error/K	
	Max	Avg
A	7.86	1.97
B	12.67	4.09

TABLE IV
COMPARISON BETWEEN DIFFERENT LEAKAGE-AWARE THERMAL SIMULATION APPROACHES IN 3D-ICS

case	Newton+KLU		HotSpot (simple iterative)				CLM+KLU		CLM+AMG (Ours)			
	time/s	iter	time/s	iter	Error/K		time/s	iter	time/s	iter	Error/K	
					Max	Avg					Max	Avg
I	17.56	3	38.83	6	0.25	0.08	6.95	3	1.70	3	0.095	0.025
II	17.12	3	25.24	4	0.05	0.01	5.87	3	2.05	3	0.003	0.0001
III	18.05	3	31.95	5	0.07	0.03	6.04	3	2.23	3	0.005	0.001
IV	58.59	3	208.32	6	0.22	0.06	17.9	3	3.83	3	0.084	0.020
V	12.12	2	25.19	4	0.004	0.001	6.22	2	1.35	2	0.01	0.0003
VI	34.21	2	129.66	4	0.002	0.002	17.1	2	0.93	2	0.001	0.0004
VII	405.8	3	1187.23	5	0.03	0.02	146.9	3	1.74	3	0.002	0.001

the same, and CLM+KLU is about 2 times faster than Newton+KLU since additional setup time for Jacobian matrices is saved in CLM+KLU. Meanwhile, CLM+KLU produces highly accurate results, with no more than 0.095K maximum errors and no more than 0.025K average errors, which is more accurate than HotSpot.

Furthermore, with the adoption of AMG preconditioned iterative linear solver, the proposed method (CLM+AMG) can achieve more than $12\times$ and $8\times$ speedup over HotSpot and Newton+KLU respectively in microfluidic-cooled 3D-ICs, and more than $139\times$ and $36\times$ speedup over HotSpot and Newton+KLU respectively in heatsink-cooled 3D-ICs. Experimental results show that the CLM+AMG method can still maintain the same level of accuracy as CLM+KLU, with no more than 0.095K maximum errors and no more than 0.025K average errors, which is almost negligible.

V. CONCLUSIONS

In this work, we have proposed an efficient and accurate leakage-aware thermal simulation approach for 3D-ICs with consideration of both heatsink and microfluidic cooling. We have proposed an iterative approach based on the corrected linearized model for leakage power approximation, which is proved to be equivalent to Newton-Chord method and guarantees the accuracy and efficiency of leakage-aware thermal simulation. We have proposed to adopt aggregation-based algebraic multigrid preconditioned iterative solver for solving high-dimensional linear system equations during calculation, which further speeds up the overall process of leakage-aware thermal simulation. Numerical experiments based on heatsink-cooled and microfluidic-cooled 3D-ICs have shown that the proposed approach is not only fast but also highly accurate, which could be further used for leakage-aware design space explorations and early thermal-aware software development.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China research project under Grants 61376040 and 91330201, in part by the National Key Research and Development Program under Grant 2016YFB0201304.

REFERENCES

[1] G. L. Loi, B. Agrawal, N. Srivastava, S.-C. Lin, T. Sherwood, and K. Banerjee, "A thermally-aware performance analysis of vertically integrated (3-D) processor-memory hierarchy," in *Proceedings of the 43rd annual Design Automation Conference*. ACM, 2006, pp. 991–996.

[2] H. Yu, Y. Shi, L. He, and T. Karnik, "Thermal via allocation for 3-D ICs considering temporally and spatially variant thermal power," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 16, no. 12, pp. 1609–1619, 2008.

[3] D. B. Tuckerman and R. Pease, "High-performance heat sinking for VLSI," *IEEE Electron device letters*, vol. 2, no. 5, pp. 126–129, 1981.

[4] K. Skadron, M. R. Stan, W. Huang, S. Velusamy, K. Sankaranarayanan, and D. Tarjan, "Temperature-aware microarchitecture," in *ACM SIGARCH Computer Architecture News*, vol. 31, no. 2. ACM, 2003, pp. 2–13.

[5] T.-Y. Wang and C. C.-P. Chen, "3-D thermal-ADI: A linear-time chip level transient thermal simulator," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 21, no. 12, pp. 1434–1445, 2002.

[6] Y. Zhan and S. S. Sapatnekar, "A high efficiency full-chip thermal simulation algorithm," in *Proceedings of the 2005 IEEE/ACM International conference on Computer-aided design*. IEEE Computer Society, 2005, pp. 635–638.

[7] A. Ziabari, J.-H. Park, E. K. Ardestani, J. Renau, S.-M. Kang, and A. Shakouri, "Power blurring: Fast static and transient thermal analysis method for packaged integrated circuits and power devices," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 22, no. 11, pp. 2366–2379, 2014.

[8] A. Sridhar, A. Vincenzi, M. Ruggiero, T. Brunschwiler, and D. Atienza, "3D-ICE: Fast compact transient thermal modeling for 3D ICs with inter-tier liquid cooling," in *Proceedings of the International Conference on Computer-Aided Design*. IEEE Press, 2010, pp. 463–470.

[9] S. Borkar, "Low power design challenges for the decade (invited talk)," in *Proceedings of the 2001 Asia and South Pacific Design Automation Conference*. ACM, 2001, pp. 293–296.

[10] S. R. Sarangi, G. Ananthanarayanan, and M. Balakrishnan, "Lightsim: A leakage aware ultrafast temperature simulator," in *2014 19th Asia and South Pacific Design Automation Conference (ASP-DAC)*. IEEE, 2014, pp. 855–860.

[11] W. Liu, K. M. Cao, X. Jin, and C. Hu, "Bsim 4.0.0 technical notes," EECS Department, University of California, Berkeley, Technical Report UCB/ERL M00/39, 2000.

[12] Y. A. Kuznetsov, *Elements of applied bifurcation theory*. Springer Science & Business Media, 2013, vol. 112.

[13] T. A. Davis and E. Palamadai Natarajan, "Algorithm 907: KLU, a direct sparse solver for circuit simulation problems," *ACM Transactions on Mathematical Software (TOMS)*, vol. 37, no. 3, p. 36, 2010.

[14] J. W. Demmel, S. C. Eisenstat, J. R. Gilbert, X. S. Li, and J. W. Liu, "A supernodal approach to sparse partial pivoting," *SIAM Journal on Matrix Analysis and Applications*, vol. 20, no. 3, pp. 720–755, 1999.

[15] Y. Notay, "An aggregation-based algebraic multigrid method," *Electronic transactions on numerical analysis*, vol. 37, no. 6, pp. 123–146, 2010.

[16] A. Napov and Y. Notay, "An algebraic multigrid method with guaranteed convergence rate," *SIAM journal on scientific computing*, vol. 34, no. 2, pp. A1079–A1109, 2012.

[17] Y. Notay, "Aggregation-based algebraic multigrid for convection-diffusion equations," *SIAM journal on scientific computing*, vol. 34, no. 4, pp. A2288–A2316, 2012.

[18] (2015) 3D-ICON: 3D interlayer cooling optimized network. [Online]. Available: http://cad-contest.el.cycu.edu.tw/problem%5FA/default.html