ATSim3D: Towards Accurate Thermal Simulator for Heterogeneous 3D-IC Systems Considering Nonlinear Leakage and Conductivity

Qipan Wang
School of Integrated Circuits, Peking University
Academy for Advanced Interdisciplinary Studies
qpwang@pku.edu.cn

Tianxiang Zhu
School of Integrated Circuits
Peking University
txzhu@pku.edu.cn

Yibo Lin*
School of Integrated Circuits, Peking University
Institute of Electronic Design Automation, Wuxi
yibolin@pku.edu.cn

Runsheng Wang
School of Integrated Circuits, Peking University
Institute of Electronic Design Automation, Wuxi
r.wang@pku.edu.cn

Ru Huang
School of Integrated Circuits, Peking University
Institute of Electronic Design Automation, Wuxi
ruhuang@pku.edu.cn

Abstract—**Thermal simulation plays a fundamental role in the thermal design of integrated circuits, especially 3D-ICs. Current simulators require significant runtime for high-resolution simulation, and dismiss the complex nonlinear thermal effects, such as nonlinear thermal conductivity and leakage power. To address these issues, we propose ATSim3D, a thermal simulator for simulating the steady-state temperature profile of nonlinear and heterogeneous 3D-IC systems. We utilize the global-local approach, combining a compact thermal model at the global level, and a finite volume method at the local level. We tackle the nonlinear effects with Kirchhoff transformation and iteration. ATSim3D enables local-level parallelization that helps achieve an average speedup of 40× compared to COMSOL, with a relative error ¡3% and a state-of-the-art resolution of 4096 × 4096, holding promise for enhancing thermal-aware design in 3D-ICs.**

Index Terms—**Thermal Simulation; 3D-IC; Parallel Computing; Nonlinear Thermal Effects**

I. INTRODUCTION

Ever-increasing integration and power consumption of integrated circuits (ICs) lead to severe thermal issues, deteriorating the performance and reliability of modern chips [1]. Thus, thermal issues have become one of the most crucial factors in VLSI chip design and runtime management. Thermal simulation serves as an essential element to enable thermal optimization and management.

Nowadays, many thermal simulation algorithms have been proposed [2]. Among them, numerical methods, encompassing finite difference methods (FDM), finite volume methods (FVM), and finite element methods (FEM), receive popularity for their high accuracy and applicability to various scenarios. Hence commercial thermal simulators are usually based on discretization-based numerical methods, especially FEM, including COMSOL [3], Celsius [4], etc. However, FEM is computationally expensive and memory intensive. To overcome the efficiency issue, many open-sourced simulators like HotSpot [5] and 3D-ICE [6], construct and solve compact thermal models (RC network) based on FDM. One step further, PACT [7] exploits a SPICE solver for parallel high-resolution simulation.

However, we identify there exist three challenges in existing simulators. Firstly, an efficient fine-grained (standard-cell level) simulator is still missing. Most simulators feature grid resolutions ranging from 32×32 to 256×256 (block level), while thermal-aware physical design requires partitioning the active layers into 1024×1024 grids and beyond [7], [8], since modern ICs have reached sizes of several square millimeters. Although PACT can simulate with the state-of-the-art (SOTA) resolution of 1024×1024, it has already reached the limit and can hardly further scale up with grid sizes.

The second challenge lies in the heterogeneous nature of real-world chips, especially 3D-ICs. As shown in Fig. 2, 3D-ICs stack multiple active layers and can provide better performance than 2D counterparts [9]. Through silicon vias (TSV), composed of

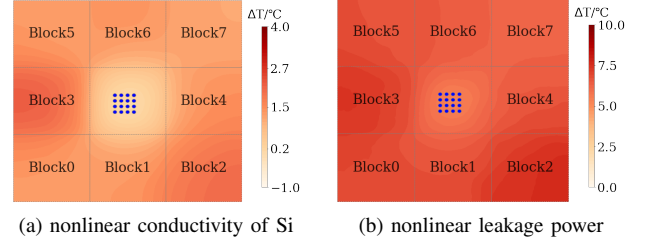(a) nonlinear conductivity of Si   (b) nonlinear leakage power

Fig. 1. Temperature difference induced by considering the (a) nonlinear thermal conductivity of Si and (b) nonlinear leakage power in a TSV-based 3D-IC (a 4×4 TSV array in the middle, shown by blue circles), compared to the linear case.

copper or tungsten, is introduced in 3D-ICs for conducting signal and power, as shown in Fig. 2(b). Previous chip-level thermal simulators often assume uniform materials within each layer and struggle to simulate the heterogeneous systems, especially when the TSV radius is small ($5 \sim 20\mu m$ typically).

The third challenge is that various nonlinear thermal mechanisms, such as nonlinear thermal conductivity and nonlinear leakage power accompany thermal hotspots and further exacerbate the high temperature in ICs. We illustrate these effects in Fig. 1, by simulating the temperature difference induced by considering the non-linear conductivity of Si and non-linear leakage power in a TSV-based 3D-IC, compared to the case with constant conductivity and power. The temperature will increase by at most $10°C$ if both mechanisms are taken into account. Although some previous efforts [10]–[12] have investigated either one of the mechanisms, a comprehensive simulation tool is lacking.

Confronted with these challenges, we propose ATSim3D[1], an accurate thermal simulator for nonlinear and heterogeneous simulation of 3D-IC systems. We summarize the capabilities of other relevant simulators in Table I, and compare them with our ATSim3D. In this way, we conclude our major contributions:

- We propose an efficient and accurate steady-state thermal simulator, considering both the nonlinear thermal conductivity and leakage power in 3D-ICs.
- We propose a global-local approach for simulation with ultra-high resolution (4096 × 4096 or even higher) and high efficiency by local-level parallelization.
- We propose utilizing the Kirchhoff transformation to address the nonlinear conductivity and prove the continuity condition for the transformed equation.
- We validate the ATSim3D for both 2D and 3D (Mono3D and TSV-based) ICs. It exhibits high accuracy (relative error $< 3\%$, max error $< 3°C$), and efficiency (40× acceleration in average) compared to rigorous simulator COMSOL.

[1]Binary available: https://github.com/Brilight/ATSim3D_pub.git

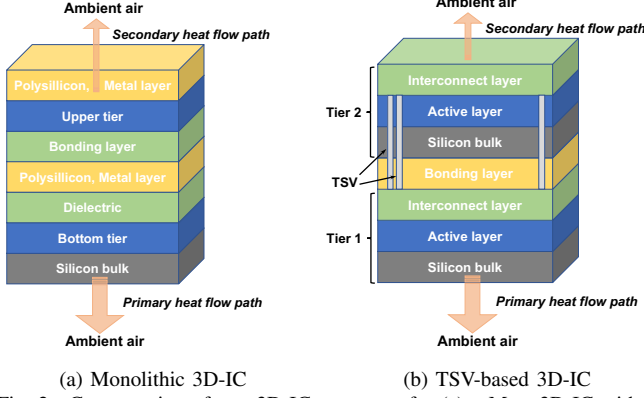(a) Monolithic 3D-IC   (b) TSV-based 3D-IC

Fig. 2. Cross-section of two 3D-IC structures for (a) a Mono3D-IC with 2 active layers, (b) a two-tier TSV-based 3D-IC.

The rest of this paper is organized as follows. Section II reviews the basic thermal model and formulates the accuracy criteria. Section III provides a thorough explanation of the proposed simulator. Section IV demonstrates the power of ATSim3D with comprehensive results, followed by the conclusion in Section V.

TABLE I
COMPARISON OF THERMAL SIMULATORS' FEATURES.

Simulator	Algorithm	Resolution	Efficiency	Nonlinear Leakage	Nonlinear Conductivity
COMSOL [Commercial]	FEM	**High**	Low	✓	✓
MTA [12]		Medium	**Fast**	✓	✓
HotSpot [5]		Low	Medium	✓	×
PACT [7]	FDM	**High**	**Fast**	×	×
3D-ICE [6]		Low	Medium	×	×
ATSim3D (ours)	FVM	**High**	**Fast**	✓	✓

II. PRELIMINARIES

In this section, we first introduce the 3D-IC configurations (II-A). Then the thermal models (II-B) are constructed, and different simulators are compared. Finally, we formulate the problem in II-C.

A. 3D-IC Configurations

Fig. 2 shows the simplified 3D-IC structures, including both (a) monolithic 3D-IC [9] and (b) TSV-based 3D-IC [13]. Layers in the Mono3D-IC are fabricated sequentially.Power sources are located at the top of the upper and bottom tiers, above which the metal layers are stacked. In TSV-based 3D-ICs, the active layers contain functional blocks as power sources, while the interconnect layers are modeled to be homogeneous with an effective thermal conductivity since the detailed routing information is agnostic. TSVs punch from the top of tier 1 up to the top of tier 2, as shown in Fig. 2(b).

There exist two heat dissipation paths: the primary heat flow path, composed of heat spreader, heat sink, and package, and the secondary heat flow path, including package substrate and PCB. As a common practice [7], we simplify the primary path to dissipate heat into the ambient through convection and ignore the secondary path. Namely, we impose the adiabatic boundary condition on each surface except the bottom, ignoring intricate mechanisms.

B. Thermal Models

1) Nonlinear thermal PDE: Since thermal designs typically concern steady-state temperature profiles under certain worst cases, we focus only on steady-state thermal analysis. The governing heat diffusion equation reads:

$$\nabla \cdot (\kappa(\mathbf{r}, T)\nabla T(\mathbf{r})) = -\mathbf{P}(\mathbf{r}, T), \tag{1}$$

subject to the boundary conditions (B.C.):

$$\vec{n} \cdot \nabla T|_{\text{Lateral\&Top}} = 0 \text{ (Adiabatic B.C.)} \tag{2}$$

$$-(\kappa\vec{n} \cdot \nabla T)|_{\text{Bottom}} = h(T - T_{amb}) \text{ (Convection B.C.)}, \tag{3}$$

where $T(\mathbf{r})$ is the 3D temperature $[unit\ K]$ profile over the location $\vec{r} = (x, y, z)$, $\kappa(\mathbf{r}, T)$ is the temperature-dependant heterogeneous conductivity $[W/m \cdot K]$, $\mathbf{P}(\mathbf{r}, T) = \mathbf{P}_{\text{dyn}}(\mathbf{r}) + \mathbf{P}_{\text{leak}}(\mathbf{r}, T)$ is the sum of dynamic and leakage power densities $[W/m^3]$. According to [14], the conductivity of common materials in IC (Si, Cu, GaAs, etc.) follows a power law with the temperature:

$$\kappa(T) = \kappa_0(T_0/T)^\alpha, \tag{4}$$

where κ_0 is the conductivity at temperature T_0, and α is a power law constant. $\alpha \sim 1.5$ for silicon, while $\alpha \sim 0$ for copper in the typical temperature range of IC. As for the power, the dynamic power of blocks exhibits little dependence on the temperature, while leakage power shows explicit reliance, following the relation [15]:

$$\mathbf{P}_{\text{leak}} = \mathbf{P}_{\text{L0}} \cdot e^{\beta(T-T_0)}, \tag{5}$$

where $\mathbf{P}_{\text{L0}}$ is the base leakage power, and β, T_0 the temperature coefficient and base temperature.

2) Previous works: The nonlinear leakage and conductivity mechanisms [10], [16] are typically tackled separately, where Newton methods are employed. To speed up, direct methods are proposed based on certain assumptions. For nonlinear conductivity, Bonani *et al.* [14] proposes to transform the Eq. 1 into the linear equation by the Kirchhoff transformation, based on the hypothesis that the temperature-dependent behavior for all materials is the same, which does not stand firm for TSV-based systems. In [17], virtual power sources are introduced to replace the TSV with homogeneous Si, which does not apply to large-scale heterogeneous systems. For nonlinear power profiles, approximate forms of leakage power, including linear model or piecewise linear model [18], are proposed to speed up calculation, at the cost of accuracy.

C. Problem Formulation

In this work, we focus on solving the nonlinear equation Eq. 1 under the B.C.s Eq. 2 and 3, with the temperature dependence of conductivity and leakage power defined in Eq. 5 and 4. To measure the accuracy of the simulation results, we define three error metrics: mean absolute error (MAE = $\textbf{Mean}\{\|T_{AT} - T_{ref}\|\}$), maximum error (MaxE = $\textbf{Max}\{\|T_{AT} - T_{ref}\|\}$), and mean absolute relative error (MARE = $\textbf{Mean}\{\frac{T_{AT}-T_{ref}}{T_{ref}-T_{\text{amb}}}\}$), where T_{AT}, T_{ref} are the temperature profile calculated by ATSim3D and rigorous simulators, and the ambient temperature T_{amb} is introduced in MARE for regularization.

III. WHOLE FRAMEWORK

In this section, we illuminate the algorithm of ATSim3D. We first introduce the motivation (III-A), then detail the global (III-B) and local (III-C) simulation, and end up with the whole iteration loop (III-D).

A. Motivation

Thermal simulation is a multi-scale problem since the package typically features mm size, while the blocks and cells are of μm size, and the thickness of layers may vary in the range of $0.1 \sim 100\ \mu m$. As a result, main-stream solvers that directly discretize the whole system will consume too much computational burden to achieve high resolution. To alleviate this problem, we adopt a simplified global-local approach [19]. It first solves the global solution on a coarse, global mesh. Then small subdomains are partitioned and their boundary conditions are extracted from
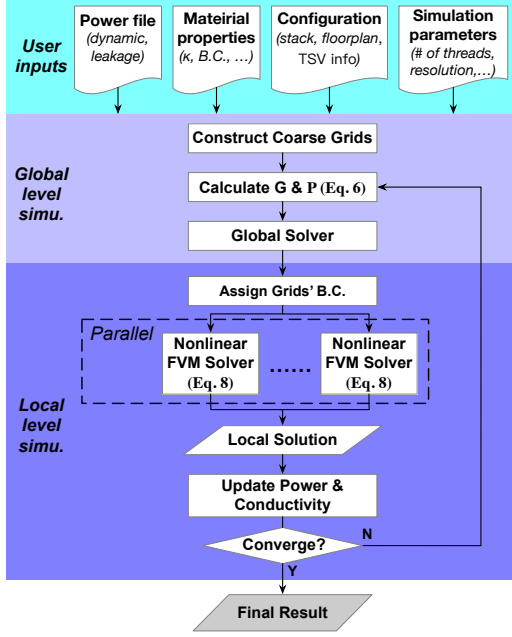
Fig. 3. The ATSim3D simulation flow, consisting of iterations between the coarse level and fine level simulation.
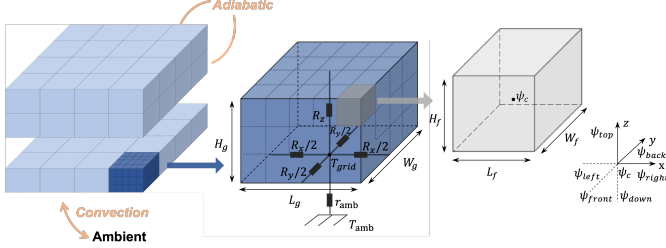

Fig. 4. The hierarchical partition of the whole system. From left to right: whole package, global coarse grid, and local fine grid. The geometric dimensions and unknown variables are indicated.

the global solution. Local solutions can be calculated afterward in all subdomains in parallel. The Eq. 1 is naturally suitable for applying this global-local approach since the local temperature depends mainly on local properties, aka. the localized heating effect in [20].

What's more, we exploit the observation that the temperature is jointly determined by two components: one that generates heat, and another dissipates. Local fine-level simulation helps delve into active layers that generate heat where we care most. In contrast, the simulation of the heat dissipation components considers the macro scale information, which can be satisfied by the global simulation. Based on these considerations, we propose ATSim3D, featuring a global-local method as illustrated in Fig. 3. It conducts global level simulation with compact thermal models, and then local level parallel simulation with finite volume method. By iteration, the detailed power and conductivity distribution are updated after each step, and the iteration stops when convergence is reached.

B. Global Simulation

The principle behind the global simulation is the duality between the thermal resistance network and the electrical counterpart [5]. Given the geometrical configuration and material parameters, we first construct coarse uniform grids for the whole system, as shown in Fig. 4. For simplicity, we assume a system of cubic domain, while the method can be generalized to other configurations. In Fig. 4, each coarse grid is associated with thermal resistances in the x,y, and z directions, connecting to

other coarse grids or the ambient. For a grid composed of uniform material, the resistances are:

$$R_x = L_g/(\kappa H_g W_g), R_y = W_g/(\kappa H_g L_g), R_z = H_g/(\kappa L_g W_g). \tag{6}$$

The geometry size L_g, W_g, H_g are attached in the figure. However, the case gets complicated if the coarse grid is composed of heterogeneous material or material with nonlinear conductivity. Previous methods often calculate the effective conductivity in a volume-averaged manner [21], which will reduce to a simple average for the composite conductivity case (TSV array, for example). Instead, we employ the formulation as in [22]. Taking the x direction as an example, the fine grid array inside the coarse grid, as shown in Fig. 4, is viewed as a connection of parallel lines. Each line represents a series of resistances in the x-direction, allowing the presence of non-uniform materials. In this way, the effective resistance is:

$$R_x = L_f / \left(W_f H_f \cdot \Sigma_{j,k}(1/\Sigma_i \kappa_{i,j,k}^{-1}) \right), \tag{7}$$

where L_f, W_f, H_f are the size of each fine grid, and $\kappa_{i,j,k}$ represents the value of conductivity of the fine grids, indexed in the order of x,y,z. The expressions of R_y, R_z can be derived in the same manner, and they reduce to Eq. 6 in the homogeneous case.

Now we can set up the matrix equation for global simulation:

$$\mathbf{G} \cdot \mathbf{T}_g = \mathbf{P}, \tag{8}$$

here $\mathbf{G}$ is the conductivity matrix composed of effective resistances, the elements in vector $\mathbf{T}_g, \mathbf{P}$ are the global temperature distribution at the bottom surface, and the total power in each coarse grid. The spatial resolution ranges in $50 \sim 1000$ μm in this stage.

C. Local Simulation

The global simulation simulates the coarse-level heat generation and dissipation process but dismisses the details inside coarse grids. So we conduct local simulation afterward to derive detailed results, especially in the active layer(s), and simulate the nonlinear thermal effects. In this step, we discretize the coarse grids of interest into fine grid arrays (as mentioned above, see Fig. 4). The detailed boundary conditions of the fine grids are calculated by linearly interpolating the global solution. A complete global-local approach involves iterative refinements of the interface conditions between the global and local solutions [23]. Although ATSim3D does not adhere to this principle as iterations are time-consuming, the solutions obtained are still accurate empirically. In the following, we will show how to solve the sub-problems inside each coarse grid in parallel. The sub-problem is to solve the nonlinear equation (Eq. 1) with the Dirichlet B.C. calculated from the global solution.

1) Kirchhoff transformation: To tackle the nonlinear equation, we impose Kirchhoff transformation to the temperature in Eq. 1:

$$\psi = \frac{1}{\kappa_b} \int_{T_g}^{T} \kappa(\eta) d\eta, \tag{9}$$

where $\kappa_b = \kappa(T_g)$ (Eq. 4) and T_g is the base temperature of the coarse grid, derived from the global solution. For constant conductivity $\psi = T - T_g$, and for the conductivity following Eq. 4, $\psi = \frac{T_0^\alpha}{1-\alpha}(T^{1-\alpha} - T_g^{1-\alpha})$ ($\alpha \neq 1$). Eq. 1 is then transformed to be (given detailed power distribution in each iteration):

$$\nabla \cdot (\kappa_b(\mathbf{r}) \nabla \psi(\mathbf{r})) = -\mathbf{P}(\mathbf{r}). \tag{10}$$

For homogeneous systems, the equation reduces to solving $\nabla^2 \psi = -\mathbf{P}(\mathbf{r})/\kappa_b$. However, for heterogeneous systems, we have to pay extra attention to the continuity conditions at material interfaces.

2) Continuity Conditions: Let us consider two regions (I, II) separated by interface Σ. The continuity conditions require both the temperature and heat flow to be continuous across the interface:

$$T_I|_\Sigma = T_{II}|_\Sigma, \quad -\kappa(T)_I \nabla T_I|_\Sigma = -\kappa(T)_{II}\nabla T_{II}|_\Sigma. \quad (11)$$

The second condition still holds for the transformed variable, since $\kappa_{b,\ i}\nabla\psi_i \equiv \kappa(T)_i\nabla T_i$ $(i = I, II)$. Conversely, the first condition does not hold at first glance, as the relation between ψ and temperature varies between materials. In [14], the authors show that to alleviate the jump of ψ at the interface, the $\kappa(T)_I/\kappa(T)_{II}$ must be a constant, which means the temperature dependence of the conductivity of all materials must be the same. However, this assumption does not hold in most cases, especially for the TSV-based 3D-ICs, composed of copper (constant conductivity) and silicon ($\alpha = 1.5$). Confronted with this problem, we prove that the first continuity condition will hold approximately in our global-local approach.

Proof. Consider the interface of copper and silicon, the temperature ($T_{\mathrm{Si}}, T_{\mathrm{Cu}}$) around the interface Σ satisfies $|T_{\mathrm{Si}}|_\Sigma - T_g| << T_g(T_g$ in Eq. 8), since the detailed temperature profile inside will not vary dramatically. Then the transformed variable satisfies:

$$\psi_{\mathrm{Si}}|_\Sigma = \frac{T_0^\alpha T_g^{1-\alpha}}{1-\alpha}[(1 + \frac{T_{\mathrm{Si}}|_\Sigma - T_g}{T_g})^{1-\alpha} - 1] \quad (12)$$

$$(\textit{To first order}) \approx \frac{T_0^\alpha T_g^{1-\alpha}}{1-\alpha}[1 + (1-\alpha)\frac{T_{\mathrm{Si}}|_\Sigma - T_g}{T_g} - 1] \quad (13)$$

$$= (\frac{T_0}{T_g})^\alpha (T_{\mathrm{Si}}|_\Sigma - T_g). \quad (14)$$

When further choose $T_0 = T_g(T_0$ in Eq. 4) in the global simulation, it concludes that $\psi_{\mathrm{Si}}|_\Sigma \approx T_{\mathrm{Si}}|_\Sigma - T_g = T_{\mathrm{Cu}}|_\Sigma - T_g = \psi_{\mathrm{Cu}}|_\Sigma$. $\square$

Now the continuity conditions can be satisfied for the variable ψ at interfaces, and the Eq. 10 inside the whole heterogeneous system can be discretized and solved by FVM.

3) Finite Volume Method: For each coarse grid, we assign the variable $\psi_{i,j,k}$ to the cubic center of each fine grid, as shown in Fig. 4. FVM integrates the Eq. 10 over the fine grid domain Ω:

$$\Sigma_{\gamma=x,y,z} \int_\Omega \partial_\gamma(\kappa_b \partial_\gamma \psi_{i,j,k})\mathrm{dxdydz} = -\mathbf{P}_{i,j,k}L_cW_cH_c. \quad (15)$$

Integrating the x-direction component of the left term we can derive $[(\kappa_b\partial_x\psi)_{i+0.5,j,k} - (\kappa_b\partial_x\psi)_{i-0.5,j,k}]W_cH_c$. For the fine grid not at the array boundary, the expression can be transformed as:

$$[\tilde{\kappa}_{i+0.5,j,k}(\psi_{i+1,j,k}-\psi_{i,j,k})-\tilde{\kappa}_{i-0.5,j,k}(\psi_{i,j,k}-\psi_{i-1,j,k})]W_cH_c/L_c.$$

While for the fine grid just at the boundary, like $i = 0$, the gradient term is calculated according to the Dirichlet B.C.:

$$[\tilde{\kappa}_{0.5,j,k}(\psi_{1,j,k}-\psi_{0,j,k})-2\times\kappa_{0,j,k}(\psi_{0,j,k}-\psi_{j,k}|_{\mathrm{BC}})]W_cH_c/L_c.$$

Here $\tilde{\kappa}_{i\pm0.5,j,k}$ represents the effective conductivity at the interface of fine grid (i, j, k) and $(i + 1, j, k)$. Rather than simply calculating the mean of $\kappa_{i,j,k}$ and $\kappa_{i+1,j,k}$, we instead use their harmonic average to keep the heat flow conservative across the interface [24]:

$$\tilde{\kappa}_{i+0.5,j,k} = 2/(\kappa_{i,j,k}^{-1} + \kappa_{i+1,j,k}^{-1}). \quad (16)$$

Summing over all three directions in Eq. 15, and incorporating all boundary values, the nonlinear equation Eq. 1 now reduces to a linear matrix equation of $\psi_{i,j,k}$. Local temperature results can be derived by the inverse Kirchhoff transformation easily from ψ.

4) Parallelization: In the local simulation, only the coarse grids that generate heat or feature nonlinear thermal effects will be simulated in detail. The sub-problems inside each coarse grid are independent after the B.C. is calculated from the global solution. Consequently, the local simulation can be conducted in a highly parallel manner for each coarse gird. Combining the results from the global and local simulation, we can derive the multi-scale temperature distribution of the whole system at the end of each global-local simulation step.

D. Iteration Loop

For systems that do not consider nonlinear thermal mechanisms, simply applying the aforementioned global-local simulation once is sufficient to derive an accurate result. However, iterations are necessary when nonlinear mechanisms are considered. As shown in Fig. 3, starting from a preset initial temperature (the ambient temperature or higher), we calculate the conductivity and leakage power distribution, based on which the global solution T_g and local solution are calculated sequentially. Then we update the leakage power within each block/unit according to the detailed local temperature profile. Different forms of leakage power are supported in this framework, and we choose Eq. 5 for all blocks for simplicity.

Subsequently follows another new iteration step. It first updates the conductivity distribution and corresponding thermal resistances and then conducts the global-local simulation. The loop is terminated when the difference between the temperature distribution of the current and the previous step is small enough, typically less than $0.1°C$. In summation, by employing iterative refinement, we can deduce the multi-scale temperature profile of non-linear and heterogeneous 3D-ICs.

IV. EXPERIMENTAL RESULTS

A. Experimental Setup

We arrange experiments for three IC structures, including a 2D Intel Multi-core CPU (labeled as **2DIC**) [7], a 7-layer complex Mono3D-IC with 2 active layers (**Mono3D**) [9], and a two-tier TSV-based 3D-IC (**TSV3D**) [13]), and four configurations of parameters (based on whether two nonlinear effects are considered or not, categorized in Table II). The floorplan and power maps are adopted from relevant works. As shown in Table V, the **Mono3D** features a random power map with 2500 units, and there exists a 4×4 TSV array (cylinder shape, radius $15\mu m$, pitch $50\mu m$) in the middle of **TSV3D**, with an oversized keep-out-zone. ATSim3D is written in Python, and the parallel computing is realized by the *multiprocessing.Pool* module, allowing the tasks to be offloaded to multiple cores. We perform the simulation on a Linux server with Intel Xeon 2.10 GHz processors with a maximum of 64 cores and 128 GB memory.

TABLE II
THE PARAMETER CONFIGURATIONS USED IN THE EXPERIMENTS,
CONCERNING THE NONLINEAR CONDUCTIVITY AND LEAKAGE POWER.

Ambient $40°C$	$\kappa_{Si} = 150\ W/(m\cdot K)$	$\kappa_{Si} = 153\cdot(300K/T)^{1.5}$
Constant Leakage @ $85°C$	I	II
β=0.015, $T_0 = 85°C$ (Eq. 5)	III	IV

B. Accuracy Analysis

We first validate the accuracy of ATSim3D against COMSOL. The error metrics defined in Section II-C are calculated for the z-axis central cross of the top active layer and shown in Table III and IV. The global and local grid resolutions are labeled by $Reso_{global}^2, Reso_{local}^2$ in the **Reso** columns. Local grid resolution represents the total number of fine grids in each layer.

In Table III, we first compare the results between three typical numerical methods in Table I: COMSOL (FEM), PACT (FDM), and ATSim3D (FVM) on configuration **I**. ATSim3D can achieve relative errors ¡3%, comparable to PACT. Note that PACT cannot handle TSV and it also fails on **2DIC-I** with 2048×2048 grids and **Mono3D-I** with 800×800 grids after using up all the memory. In contrast, ATSim3D is memory-efficient. In Table IV, we compare COMSOL and ATSim3D on configurations **II-IV**. We do not compare with PACT as it can not support nonlinear configurations. We can see that ATSim3D shows a mean absolute error of $< 1^\circ C$, a maximum error of $3^\circ C$, and a relative error of $< 3\%$. Additionally, the nonlinear thermal effects are reflected by the difference between maximum temperature values of **TSV3D-IV** and **TSV3D-I**, as high as $10^\circ C$.

Simulation results of ATSim3D and COMSOL and the error distribution are provided in Table V. We can observe a close match between the simulated results, while the error concentrates in the area of high temperature or close to the boundary, a result stemming from the different ways of mesh (grid) partitioning.

Fig. 5 provides further analysis of the factors affecting the error in **TSV3D-IV**. We can see that the error first decreases then saturates, as the number of iterations (Fig. 5(a)) and global grid resolution (Fig. 5(b)) increase. ATSim3D requires just 3-4 iterations to get converged, and it gets a mesh-independent solution at the grid resolution of $\sim 6\mu m$ as the power map is coarse. Higher resolution of $\sim 1\mu m$ is achieved for the **Mono3D** in Table III. To further reduce the error, attention should be paid to polishing the B.C. in the local simulation [25], which lacks some high frequency and local information since it is calculated by interpolating the global solution.

TABLE III
PERFORMANCE COMPARISON BETWEEN COMSOL, PACT, AND ATSIM3D FOR THE LINEAR CONFIGURATION **I**. **RESO** AND **RESO**$_{local}$ REPRESENT THE TOTAL NUMBER OF FINE GRIDS IN THE ACTIVE LAYERS FOR PACT AND ATSIM3D, RESPECTIVELY.

| Structure	COMSOL			PACT				ATSim3D			
	$T_{max}/^\circ C$	Memory	Time/s	Reso	Memory	Time/s	MARE/%	Reso$_{local}$	Memory	Time/s	MARE/%
2DIC-I	69.32	3.0 GB	60	256^2	590 MB	8	2.38	256^2	121 MB	6	2.41
				512^2	2.5 GB	55	2.40	512^2	209 MB	8	2.07
				1024^2	11.9 GB	450	2.40	1024^2	419 MB	11	2.09
				2048^2	*Not enough memory*			2048^2	1.1 GB	34	2.09
				4096^2	*Not enough memory*			4096^2	4.3 GB	94	2.09
Mono3D-I	89.61	12.9 GB	1410	200^2	6.5 GB	255	0.53	200^2	412 MB	15	0.55
				400^2	39.2 GB	2980	0.53	400^2	434 MB	18	0.54
				800^2	*Not enough memory*			800^2	536 MB	23	0.54
TSV3D-I	119.58	5.5 GB	77	*Unable to handle TSV*				640^2	337 MB	10	0.14

TABLE IV
PERFORMANCE COMPARISON BETWEEN ATSIM3D AND COMSOL FOR THE NONLINEAR CONFIGURATIONS (**II-IV**).

| Configuration	COMSOL		ATSim3D			Error Metric		
	$T_{max}/^\circ C$	Time/s	Time/s	Reso$_{global}$	Reso$_{local}$	MARE/%	MaxE/$^\circ C$	MAE/$^\circ C$
2DIC-II	69.86	110	18			1.45	1.02	0.12
2DIC-III	67.13	101	14	64^2	512^2	1.47	0.86	0.12
2DIC-IV	67.29	117	25			2.69	1.54	0.18
Mono3D-II	91.40	2930	37			0.69	0.96	0.28
Mono3D-III	89.07	3966	25	50^2	400^2	0.90	0.58	0.13
Mono3D-IV	90.60	5541	53			1.06	1.15	0.41
TSV3D-II	121.78	118	30			0.30	1.26	0.21
TSV3D-III	126.96	106	25	40^2	640^2	0.65	1.40	0.49
TSV3D-IV	130.16	109	39			1.15	2.71	0.92

C. Efficiency Analysis

In this section, we analyze the efficiency of ATSim3D. Firstly, the runtime is compared in Table III and IV w.r.t COMSOL and PACT, with a maximum of 16 CPU cores used. ATSim3D achieves a speedup of $3 \sim 150\times$ ($40\times$ in average) compared to COMSOL, benefiting from two factors: the simplified global-local algorithm and the parallelization utilized in the local simulation. COMSOL consumes much time for the **Mono3D** due to the requirement for fine-resolution mesh division. ATSim3D is also faster than PACT
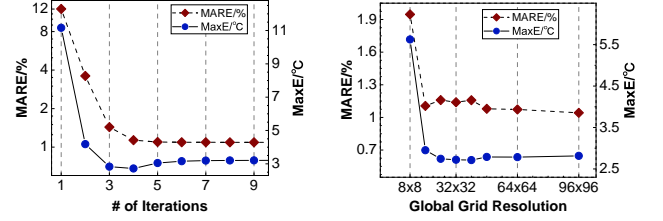
(a) Error — number of iterations   (b) Error — global grid resolutions

Fig. 5. The relation between the error of ATSim3D in **TSV3D-IV** and (a) the number of iterations, as well as (b) the global grid resolutions, with each coarse grid partitioned into 16×16 fine grids.
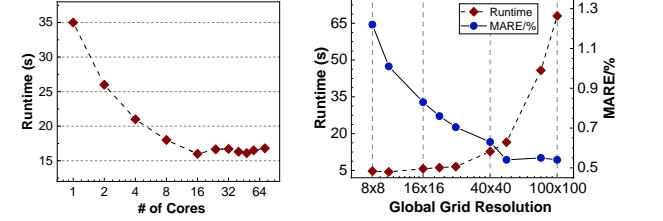
(a) Runtime — number of cores   (b) Runtime — grid resolutions

Fig. 6. The relation between the runtime of ATSim3D in **Mono3D-I** and (a) the number of cores, as well as (b) the grid resolutions, with the local resolution constant to be 400^2.

(at most $40\times$), and overcomes the scalability challenge of PACT, achieving a SOTA resolution of 4096×4096 in the active layer.
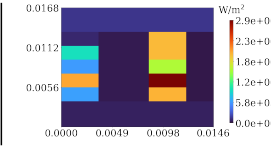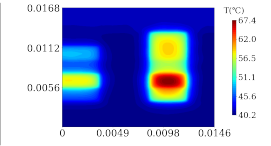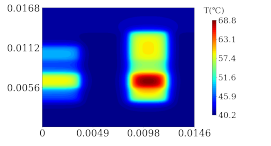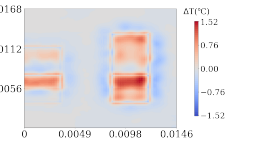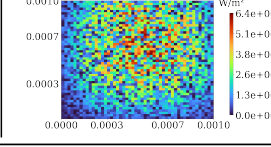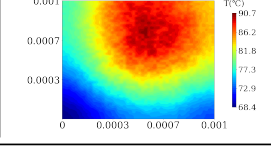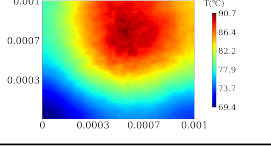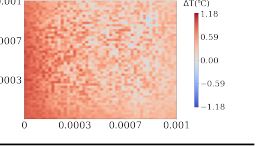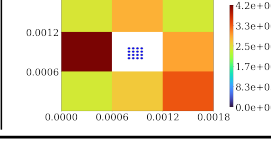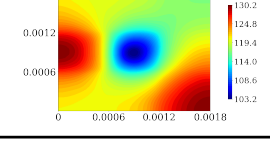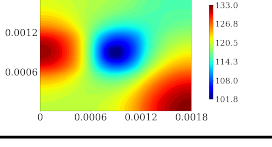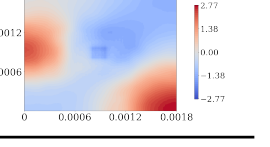
In the following, we illustrate in Fig. 6 the relation between the runtime in **Mono3D-I** and the number of cores employed, as well as the grid resolutions. The runtime first decreases and then saturates at a large number of cores (¿16), as shown in Fig. 6(a). This comes from the fact that runtime for the local simulation decreases while the cost of creating, managing, and terminating the process pools increases when more cores are used. Fig. 6(b) shows the relation between the runtime and global grid resolution for fixed local resolution. With the increase of global grid resolution, the global simulation time increases, the number of local subdomains increases, and the runtime for solving each sub-problem first decreases and then saturates. These factors collectively lead to the runtime exhibiting the trends in the figure. What's more, the error gets smaller with the increase in the number of global grids, indicating a trade-off between accuracy and efficiency when choosing the global grid resolution.

V. CONCLUSION

As 3D-ICs are attracting more and more attention, related thermal simulation has met new challenges, including standard-cell level resolution, complex thermal effects (nonlinear leakage power and conductivity), and the heterogeneous material systems. However, current simulators require significant runtime for high-resolution simulation and dismiss these nonlinear effects. To this end, we propose ATSim3D, an accurate thermal simulator for steady-state simulation of nonlinear and heterogeneous 3D-IC systems. We utilize the global-local approach, combining a compact thermal model at the global level, and a finite volume method at the local level. We tackle the nonlinear effects by Kirchhoff transformation and iterative updating of leakage and conductivity distribution. By parallelization in the local level simulation, ATSim3D achieves an average speedup of $40\times$ compared to COMSOL, with the relative error ¡3% and max error $< 3^\circ C$ for the test structures of both 2D and 3D (Mono3D and TSV-based) ICs. A SOTA resolution of 4096×4096 is also demonstrated, along with a maximum acceleration of $40\times$ compared to PACT.

TABLE V

RESULTS OF ATSIM3D AND COMSOL FOR THE NONLINEAR CONFIGURATION IN THE TABLE IV, AND CORRESPONDING ERROR DISTRIBUTION.

Structure	Power Map	COMSOL	ATSim3D	Error ($T_{AT} - T_{comsol}$)
2DIC-IV				
Mono3D-IV				
TSV3D-IV				

We believe that our simulator holds the promise for enhancing future thermal-aware design in 3D-ICs.

ACKNOWLEDGE

REFERENCES

[1] M. Pedram and S. Nazarian, "Thermal modeling, analysis, and management in vlsi circuits: Principles and methods," *Proceedings of the IEEE*, vol. 94, no. 8, pp. 1487–1501, 2006.

[2] H. Sultan, A. Chauhan, and S. R. Sarangi, "A survey of chip-level thermal simulators," *ACM Computing Surveys (CSUR)*, vol. 52, no. 2, pp. 1–35, 2019.

[3] "Comsol Multiphysics."

[4] "Cadence Celcius Thermal Solver."

[5] M. R. Stan, K. Skadron, M. Barcella, W. Huang, K. Sankara-narayanan, and S. Velusamy, "Hotspot: A dynamic compact thermal model at the processor-architecture level," *Microelectronics Journal*, vol. 34, no. 12, pp. 1153–1165, 2003.

[6] F. Terraneo, A. Leva, W. Fornaciari, M. Zapater, and D. Atienza, "3d-ice 3.0: efficient nonlinear mpsoc thermal simulation with pluggable heat sink models," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 41, no. 4, pp. 1062–1075, 2021.

[7] Z. Yuan, P. Shukla, S. Chetoui, S. Nemtzow, S. Reda, and A. K. Coskun, "Pact: An extensible parallel thermal simulator for emerging integration and cooling technologies," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 41, no. 4, pp. 1048–1061, 2021.

[8] Z. Hassan, N. Allec, F. Yang, L. Shang, R. P. Dick, and X. Zeng, "Full-spectrum spatial–temporal dynamic thermal analysis for nanometer-scale integrated circuits," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 19, no. 12, pp. 2276–2289, 2011.

[9] P. Shukla, A. K. Coskun, V. F. Pavlidis, and E. Salman, "An overview of thermal challenges and opportunities for monolithic 3d ics," in *Proceedings of the 2019 on Great Lakes Symposium on VLSI*, 2019, pp. 439–444.

[10] C. Yan, H. Zhu, D. Zhou, and X. Zeng, "An efficient leakage-aware thermal simulation approach for 3d-ics using corrected linearized model and algebraic multigrid," in *Design, Automation & Test in Europe Conference & Exhibition (DATE), 2017.* IEEE, 2017, pp. 1207–1212.

[11] K. R. Bagnall, Y. S. Muzychka, and E. N. Wang, "Application of the kirchhoff transform to thermal spreading problems with convection boundary conditions," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 4, no. 3, pp. 408–420, 2013.

[12] S. Ladenheim, Y.-C. Chen, M. Mihajlović, and V. F. Pavlidis, "The mta: An advanced and versatile thermal simulator for integrated systems," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 37, no. 12, pp. 3123–3136, 2018.

[13] Q. HANHUA, "Thermal management of 3d integrated circuits with microchannel-based liquid cooling," Ph.D. dissertation, Nanyang Technological University, 2014.

[14] F. Bonani and G. Ghione, "On the application of the kirchhoff transformation to the steady-state thermal analysis of semiconductor devices with temperature-dependent and piecewise inhomogeneous thermal conductivity," *Solid-State Electronics*, vol. 38, no. 7, pp. 1409–1412, 1995.

[15] L.-Y. Lu and L.-Y. Chiou, "Temperature gradient-aware thermal simulator for three-dimensional integrated circuits," *IET Computers & Digital Techniques*, vol. 11, no. 5, pp. 190–196, 2017.

[16] A. Ramalingam, F. Liu, S. R. Nassif, and D. Z. Pan, "Accurate thermal analysis considering nonlinear thermal conductivity," in *7th International Symposium on Quality Electronic Design (ISQED'06)*. IEEE, 2006, pp. 6–pp.

[17] D. Oh, C. C. P. Chen, and Y. H. Hu, "Efficient thermal simulation for 3-d ic with thermal through-silicon vias," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 31, no. 11, pp. 1767–1771, 2012.

[18] H. Sultan and S. R. Sarangi, "A fast leakage-aware green's-function-based thermal simulator for 3-d chips," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 28, no. 11, pp. 2342–2355, 2020.

[19] A. K. Noor, "Global-local methodologies and their application to nonlinear analysis," *Finite Elements in Analysis and Design*, vol. 2, no. 4, pp. 333–346, 1986.

[20] J. Wen, S. Pan, N. Chang, W.-T. Chuang, W. Xia, D. Zhu, A. Kumar, E.-C. Yang, K. Srinivasan, and Y.-S. Li, "Dnn-based fast static on-chip thermal solver," in *2020 36th Semiconductor Thermal Measurement, Modeling & Management Symposium (SEMI-THERM)*. IEEE, 2020, pp. 65–75.

[21] J. Meng, K. Kawakami, and A. K. Coskun, "Optimizing energy efficiency of 3-d multicore systems with stacked dram under power and thermal constraints," in *DAC Design Automation Conference 2012*, 2012, pp. 648–655.

[22] C. Nie, Q. Xu, C. Wang, H. Cao, J. Liu, and Z. Li, "Efficient transient thermal analysis of chiplet heterogeneous integration," *Applied Thermal Engineering*, vol. 229, p. 120609, 2023.

[23] A. El Kerim, P. Gosselet, and F. Magoulès, "Asynchronous global–local non-invasive coupling for linear elliptic problems," *Computer Methods in Applied Mechanics and Engineering*, vol. 406, p. 115910, 2023.

[24] J. W. Thomas, *Numerical partial differential equations: finite difference methods*. Springer Science & Business Media, 2013, vol. 22.

[25] P. Gosselet, M. Blanchard, O. Allix, and G. Guguin, "Non-invasive global–local coupling as a schwarz domain decomposition method: acceleration and generalization," *Advanced Modeling and Simulation in Engineering Sciences*, 2018.