Electrothermal Simulation and Vertical Interconnect Planning for Integrated Chiplets

2025 International Symposium of Electronics Design Automation (ISEDA) | 979-8-3315-3696-1/25/$31.00 ©2025 IEEE | DOI: 10.1109/ISEDA65950.2025.11100500

Siyuan Miao[1], Lingkang Zhu[1], Wenkai Yang[1], Teng Lu[1], Yanze Zhou[1], Chen Wu[2,3], Zhiping Yu[4], Ting-Jung Lin[2,3]*, Lei He[5]*
[1]*BTD Technology, Ningbo, China*
[2]*Ningbo Institute of Digital Twin, Eastern Institute of Technology, Ningbo, China*
[3]*Engineering Research Center of Chiplet Design and Manufacturing of Zhejiang Province, Ningbo, Zhejiang, China*
[4]*Tsinghua University, Beijing, China* [5]*Eastern Institute for Advanced Study, Eastern Institute of Technology, Ningbo, China*
tlin@idt.eitech.edu.cn, lhe@eitech.edu.cn

***Abstract*—Chiplets are emerging as novel solutions for high-performance AI computing processors. Vertical interconnects (VICs) including μbumps, C4 bumps and through-silicon vias (TSVs) in chiplets are critical as they not only carry signals and power supplies but also transfer heat efficiently. Due to the need of fine-grained VIC modeling, existing thermal tools are ineffective for VIC-embedded chiplets. Moreover, electrothermal analysis in previous architectural simulators does not consider temperature dependence for short-circuit power, which is non-trivial in our experiments. To address the above problems, this paper proposes SYSgen, a framework for accurate, location-based temperature-dependent power profiling and VIC planning for integrated chiplets. SYSgen achieves a 97.77× speedup with a maximum error below 1.2°C when the chiplet temperature is around 100°C compared to COMSOL. It also reduces VIC number by 21.7% and 12.4% compared to two existing papers with same constraints on signal and power routing and maximum temperature.**

***Index Terms*—Integrated chiplets, vertical interconnect planning, through-silicon via (TSV)**

I. Introduction

As Moore's Law [1] slows down, traditional monolithic chips are facing their limitations in performance and cost-effectiveness [2] [3]. Under these circumstances, chiplets are drawing extensive attention [4] [5] for lower design costs and better performance. Chiplets can be integrated through advanced packaging, such as 2.5D and 3D packaging [6]. As illustrated in Fig. 1, in 2.5D packaging, chiplets are placed on an interposer and connected with microbumps (μbumps). The interposer is bonded to the substrate using C4 bumps, while redistribution layers (RDLs) and through-silicon vias (TSVs) provide electrical connections between μbumps and C4 bumps. In 3D packaging, chiplets are stacked vertically and connected to a base die, which serves a function similar to a silicon interposer.

One feature of 2.5D and 3D packaging is the vertical interconnects, referred to as VICs in this paper, including μbumps, TSVs and C4 bumps. VICs significantly influence the heat dissipation of chiplet architectures. The bumps and TSVs form heat transfer channels and conduct heat from lower chiplets to upper layers. As shown in Fig. 2(a), the TSVs exhibit higher temperatures compared to the surrounding chiplet area. Compared with Fig. 2(c) to Fig. 2(d), VICs reduce chiplet temperature from 431K to 355K, showing their strong capability for heat dissipation. In addition to thermal integrity, VIC planning also significantly influences signal connectivity and power supply [7], making it a critical consideration in chiplet design.

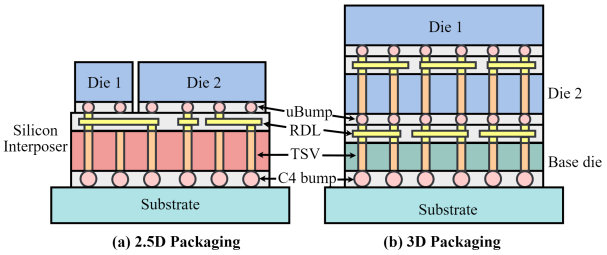

Fig. 1. Structure of (a) 2.5D packaging (b) 3D packaging. The VICs are aligned only for illustration; they can be unaligned when manufacturing.

However, VIC planning is challenging primarily due to the distinct characteristics of chiplet architectures compared to monolithic chips. **First, previous thermal analyzers are either too slow or ineffective when simulating VIC-embedded chiplets.** Incorporating VICs in thermal simulations remains challenging as the small physical sizes of VICs require very fine granularity of thermal tools to capture their effects. Finite Element Method (FEM) solvers like COMSOL and Celsius are effective, but they require hours to calculate the huge number of mesh cells. On the other hand, fast thermal analysis tools like HotSpot [8] fails to capture the influence of VICs due to the coarse modeling granularity, leading to large simulation errors. Without efficient and accurate thermal analysis, the effectiveness of thermal management for chiplet-based designs is highly limited.

Second, existing simulators fail to accurately estimate the power consumption of chiplet architectures. Both CPU-based simulators such as GEM5 [9] and PTscalar [10] and AI-computing-based simulators like OPU [11] disregard the physical implementations of chiplet architectures, which differ significantly from monolithic chips and considerably influence

*Corresponding authors. This work was partially supported by "Science and Technology Innovation in Yongjiang 2035" (2024Z283) and by research support from BTD Inc.

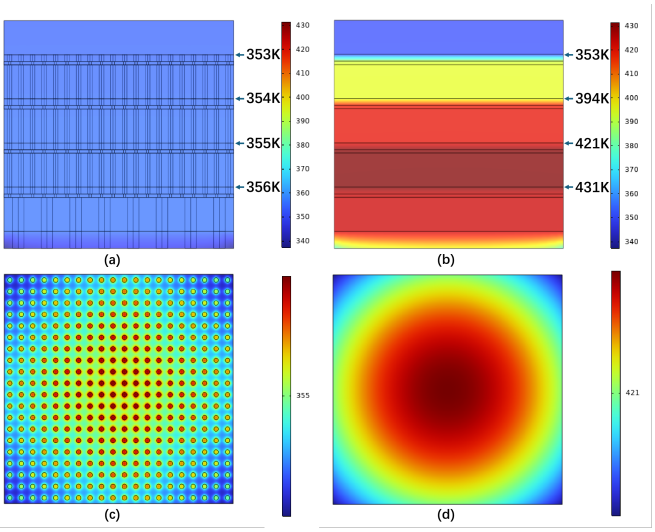

Fig. 2. (a) Vertical temperature distribution of 3D-IC with VICs, (b) vertical temperature distribution of 3D-IC without VICs, (c) temperature distribution of the bottom surface of Die3 (the third die from top to bottom) with VICs, (d) temperature distribution of the bottom surface of Die3 without VICs

the performance and power of interconnects. Furthermore, electrothermal analysis in existing simulators is either absent or inadequate. Although some works [10] [12] consider leakage power, none of them accounts for the temperature effect on dynamic power. Experimental results indicate that dynamic power increases by 10∼15% as temperature rises from 25°C to 100°C, leading to an overall power increase of approximately 10%. This increase is primarily due to the rise in short-circuit power with temperature. The lack of accurate power profiling directly leads to inefficient or even ineffective VIC planning for thermal integrity and power delivery.

Third, current electronic design automation (EDA) tools for VIC planning at the early design stage is inefficient. Floorplet [13] proposes a framework to optimize chiplet performance, but its scope is limited to floorplanning and does not extend to physical implementation. Other work introduces thermal VICs to improve heat dissipation [14] [15], while [16] further incorporates power integrity into via allocation. Additionally, [17] adopts sequential and simultaneous optimization during non-signal VIC design. However, the overlook of electrothermal coupling results in overdesign of thermal VICs. Furthermore, their thermal management strategies are built on abstract models of 3D packaging, which is inconsistent with the practical implementation of µbumps and TSVs.

To address these limitations, we propose SYSgen, a framework for electrothermal simulation and VIC planning for integrated chiplets, with the main contributions as follows:

- *VIC-aware thermal analysis:* We apply the effective thermal conductivity model in [18] and develop an efficient yet accurate thermal tool for 2.5D and 3D packaging, considering the influence of VICs. Our approach achieves a 30∼250× speedup compared to COMSOL, while maintaining a maximum error of below 1.2°C when the chiplet temperature is around 100°C.
- *Chiplet architecture simulation:* We propose a simulator based on OPU-sim that enables accurate temperature-dependent power profiling considering both microarchitecture and physical implementation of chiplets. Features of chiplets, such as protocol-based interconnections and HBM are supported as built-in libraries.
- *Vertical interconnect planning:* Our VIC planning minimizes the number of interconnects under the constraints of signal and power routing and thermal integrity. Compared to [16] and [12], SYSgen results in 21.7% and 12.4% less VICs, respectively. The advantages come from SYSgen's consideration of electrothermal coupling and non-uniform power distributions, which prevents overdesign of thermal VICs.

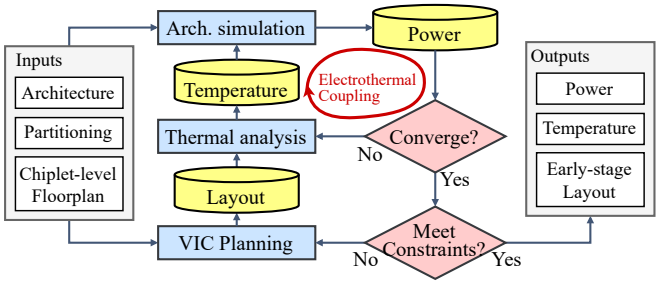

Fig. 3. Overview of SYSgen framework.

II. Preliminaries

A. Advanced Packaging and Vertical Interconnects

Packaging plays a critical role in chiplet-based designs, influencing system performance, area efficiency, heat dissipation, and overall cost. As illustrated in Fig. 1, advanced packaging technologies are generally classified into two categories: 2.5D packaging in which chiplets are placed on a silicon interposer for high-bandwidth interconnects, and 3D packaging where chiplets are stacked vertically together with bumps and RDLs for higher area utilization and integration density. However, 3D stacking also obstacles vertical heat dissipation, making thermal management a critical issue. Fortunately, the µbumps, TSVs and RDLs form heat dissipation channels, facilitating heat transfer from lower chiplet layers to upper layers. Thus, an effective strategy to mitigate thermal issues is to introduce additional VICs specifically for thermal purposes, referred to as thermal VICs. Beyond thermal considerations, VICs must also meet requirements for signal and power routing [19]. These challenges call for a comprehensive strategy for VIC planning.

B. Electrothermal Coupling

Existing studies on electrothermal coupling mainly focus on the leakage power. The work in [20] describes the dependency between leakage power and temperature with a third-order polynomial,

$$\frac{I_{\text{leakage}}(T)}{I_{\text{leakage}}(T_0)} = 1+\alpha_1(T-T_0)+\alpha_2(T-T_0)^2+\alpha_3(T-T_0)^3. \quad (1)$$

The coefficient α_1, α_2 and α_3 varies with fabrication process.

However, ignorance of the temperature effect on dynamic power leads to incomplete electrothermal coupling analysis.

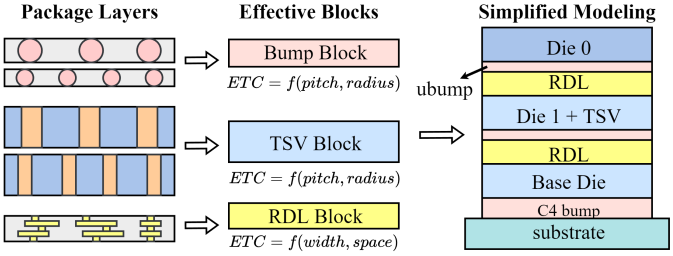

Fig. 4. Thermal macro model and an application on 3D integration.

Dynamic power consists of switching power and internal power (also known as short-circuit power), the former of which is independent of temperature. However, the short-circuit power is influenced by temperature according to

$$P_{\text{sc}} = f \int_0^\infty I_{\text{sc}}(t) V_{\text{DD}}\, dt \approx V_{\text{DD}} I_{\text{peak}} \left(\frac{t_r + t_f}{2} \right) f, \qquad (2)$$

where I_{sc} is the short-circuit current, I_{peak} is the peak short-circuit current, t_r and t_f are rising time and falling time. When temperature rises, t_r and t_f increase, and P_{sc} increases, leading to an increase in dynamic power. In SYSgen, we consider the temperature effect on both dynamic power and leakage power.

C. Thermal Macro Model

We classify the layers in 2.5D and 3D packaging into four categories according to their physical structures: die layer, TSV layer, bump layer and RDL layer. Notably, the TSV layer refers to silicon layers incorporating TSVs, including the base die and TSV-embedded dies in 3D packaging, as well as the interposer in 2.5D packaging. These layers can be mapped to effective thermal blocks with homogeneous effective thermal conductivity (ETC) in the vertical direction. These effective blocks can be stacked for temperature calculations in 2.5D and 3D integration, as shown in Fig. 4. Experiments show that our macro model obtains a maximum error of less than 1.2°C when the chiplet temperature is at 100°C for such packages. In this paper, we focus on the validation of this approach and its implementation for VIC planning, while the detailed methodologies are originally presented in a separate paper [18]. All implementations and experiments conducted in this paper are original.

III. SYSGEN FRAMEWORK

A. Overview of SYSgen Framework

Fig. 3 shows the overview of the proposed SYSgen. Given the architecture, partitioning and chiplet-level floorplan, the tool performs electrothermal analysis and VIC planning, generates an early-stage layout, and reports chiplet power and temperature. The framework incorporates novel tools for chiplet architecture simulation, thermal analysis, and VIC planning, which are elaborated in the following sections.

B. Thermal Analysis

In this section, we first introduce the implementation of the thermal macro model. We then develop a thermal-RC solver that performs thermal calculations based on the equivalent thermal resistance and capacitance network, serving as a faster alternative to finite element analysis (FEA) tools such as COMSOL. Finally, we present the validation result of the proposed macro+RC approach.

1) Implementation: The effectiveness of the macro model highly relies on an accurate ETC. The key of macro model is to determine the ETC, which is influenced by different parameters in different package layers:

$$ETC = \begin{cases} h_{bump}(\text{pitch}, \text{radius}), & \text{for bump layers} \\ h_{TSV}(\text{pitch}, \text{radius}), & \text{for TSV layers} \\ h_{RDL}(\text{width}, \text{space}), & \text{for RDL layers} \end{cases}. \qquad (3)$$

To obtain h_{bump}, h_{TSV} and h_{RDL}, we conduct experimental measurements with the following assumptions:

- Bumps consist of copper pillars and Sn-Pb solder, with epoxy resin as the underfill. TSVs are treated as copper cylinders encapsulated in a silicon dioxide layer and embedded in silicon. RDLs are represented as copper lines and pillars, forming a uniform mesh network.
- Bumps and TSVs are assumed to be uniformly distributed within the block, with each TSV positioned directly beneath a bump. At each intersection of copper lines in the RDL, a TSV is placed.

We conduct experiments on a block consisting of bump/TSV/RDL structures to determine the ETCs, with the following geometric parameters and boundary conditions: For the bump layer, the μbump has a radius ranging from 20 μm to 50 μm and the pitch varies between 50 μm and 330 μm. The C4 bump has a radius ranging from 50 μm to 80 μm, with pitch values between 100 μm and 330 μm. The TSV has a radius ranging from 10 μm to 40 μm, and the pitch varies between 50 μm and 220 μm. The RDL has a width ranging from 20 μm to 60 μm, with the corresponding spacing varying between 20 μm and 180 μm. The ambient temperature is set to 293.15K, the side surfaces are adiabatic, the top surface is maintained at a fixed temperature of 373.15K, and the bottom surface has a heat transfer coefficient of 20W/(m²·K).

The ETCs are calculated according to the following equation:

$$\text{ETC} = \frac{h_1 \cdot l \cdot |T_1 - T_\infty|}{|T_2 - T_1|}, \qquad (4)$$

where T_1 is the temperature at the top and T_2 is the temperature at the bottom, T_∞ is the indoor temperature. h_1 and l are parameters representing the heat transfer coefficient at the upper surface and the height of the layer, respectively. The ETCs are stored in LUTs for ease of use.

To solve the thermal macro model, we develop a thermal-RC solver with a cube-mesh. The modeling here has been simplified to macros with homogeneous ETCs, enabling straightforward computation of thermal resistance and capacitance of each cube using formulas the same as [21]. The resulting thermal-RC circuits are then solved using HSPICE.

2) Validation: We validate our macro+RC approach on four cases, two of which are 2.5D integrations and the others are 3D integrations. The structural parameters of these cases are summarized in Table I, with chiplet power densities ranging $0.1\sim0.4\text{W}/\text{mm}^2$. The golden standard is the detailed modeling of advanced packaging combined with COMSOL

TABLE I
STRUCTURAL PARAMETERS OF 2.5D AND 3D INTEGRATION CASES

Cases	Die layer	μBump layer			TSV layer			C4bump layer			RDL layer		
	height	height	radius	pitch	height	radius	pitch	height	radius	pitch	height	width	space
3D-2Die	200μm	40μm	30μm	200μm	200μm	30μm	200μm	100μm	60μm	600μm	20μm	30μm	170μm
3D-4Die	200μm	40μm	40μm	300μm	200μm	40μm	300μm	100μm	80μm	300μm	20μm	40μm	260μm
2.5D-2Die	200μm	40μm	20μm	100μm	200μm	20μm	100μm	100μm	50μm	300μm	20μm	20μm	80μm
2.5D-4Die	200μm	40μm	20μm	100μm	200μm	20μm	100μm	100μm	50μm	300μm	20μm	20μm	80μm

TABLE II
EXECUTION TIME COMPARISON BETWEEN MACRO+RC AND GOLDEN

Cases	Runtime (s)			Speedup‡
	Golden	Macro+COMSOL†	**Macro+RC**	
2.5D	217	10	**8**	27.13×
3D-2Die	238	11	**8**	29.75×
3D-3Die	380	12	**9**	42.22×
3D-4Die	960	13	**9**	106.67×
3D-5Die	2507	16	**10**	250.70×
Average	-	-	-	97.77×

†Simplified macro modeling + COMSOL solver.
‡Speedup of our macro+RC approach against golden.

TABLE III
ACCURACY COMPARISON BETWEEN MACRO+RC AND GOLDEN

#Cases	Approaches	Max (K)	Min (K)	MAE (K)	RMSE (K)
3D-2Dies	Golden	324.427	297.614	-	-
	Macro+RC	**325.048**	**296.517**	**0.419**	**0.454**
3D-4Dies	Golden	359.039	303.125	-	-
	Macro+RC	**360.744**	**300.625**	**0.967**	**1.104**
2.5D-2Dies	Golden	312.580	294.749	-	-
	Macro+RC	**312.660**	**294.049**	**0.389**	**0.392**
2.5D-4Dies	Golden	323.177	295.633	-	-
	Macro+RC	**322.642**	**294.488**	**0.596**	**0.597**

simulations. The errors of maximum and minimum temperatures, along with root mean square error (RMSE) values, across all four cases are reported in Table III. The runtime of our approach and speedup over golden is shown in Table II. Our macro+RC approach achieves an average 97.77× speedup over golden, while maintaining an limited RMSE of less than 1.2°C.

C. Chiplet Architecture Simulation

Our simulator is built upon the OPU-sim [11], which was originally developed under the TVM (Tensor Virtual Machine) framework [22]. We extend its functionality to support chiplets. For interconnects, we support protocol-based connections such as UCIe [23] and PCIe [24], and direct connections through wires and VICs. We also integrate HBM which is preferred in AI computing chips. These features are supported as built-in libraries in our simulator.

A key feature of our simulator is temperature-dependent power estimation, formulated as:

$$P(T) = P(T_0) \cdot f_s(T, T_0), \tag{5}$$

where f_s represents the temperature scaling function. Circuits with different structures exhibit distinct temperature-dependent behaviors [10]. Hence, we categorize circuits into analog and digital, with digital circuits further divided into functional units (FUs) and memory blocks. Accordingly, we define separate scaling functions: f_a for analog circuits, f_{FU} for functional units and f_{mem} for memory components.

Since FUs contribute to the majority of the power consumption, here we present a detailed explanation of f_{FU}. The following methodology is applied to both leakage power and dynamic power. Developing individual scaling function f_{FU_i} for every FU (denoted as FU_i) is neither practical nor necessary as defining unique scaling functions for all structural variations would be infeasible, and determining new rules for every newly introduced FU is unrealistic.

Fortunately, as illustrated in Fig. 5, FUs tend to exhibit similar temperature-dependent behaviors, so we derive a generalized f_{FU} as a weighted sum of f_{FU_i}, within a set of basic FUs ($\mathbb{F}_{basic}$). Then we derive the generalized scaling function f_{FU} as:

$$f_{FU}(T, T_0) = \sum_i w_{FU_i} \cdot f_{FU_i}(T, T_0), \tag{6}$$

$$w_{FU_i} = \frac{n_{FU_i} \cdot P_{FU_i}}{\sum_k n_{FU_k} \cdot P_{FU_k}}, \tag{7}$$

where w_{FU_i} represents the weight of FU_i, proportional to the product of its occurrence frequency (n_{FU_i}) in processors and its power consumption (P_{FU_i}). Since FUs are composed of standard cells (abbreviated as cells), we approximate f_{FU_i} using a combination of cell-level scaling functions f_{cell_j} rather than directly simulating FUs. This approach enhances extensibility, allowing new FUs to be supported as long as their cell composition is known. Instead of using the entire cell library in process design kit (PDK), we focus on a core subset ($\mathbb{C}_{basic}$) which is sufficient to construct any FU and universally included in PDKs. Thus, we approximate f_{FU_i} as follows:

$$f_{FU_i}(T, T_0) = \sum_j w^{(i)}_{cell_j} \cdot f_{cell_j}(T, T_0), \tag{8}$$

$$w^{(i)}_{cell_j} = \frac{n^{(i)}_{cell_j} \cdot P_{cell_j}}{\sum_k n^{(i)}_{cell_k} \cdot P_{cell_k}}, \tag{9}$$

where f_{cell_j} is the scaling function for $\mathrm{cell}_j \in \mathbb{C}_{basic}$. The occurrence frequency of cell_j in $\mathrm{FU_i}$ ($n^{(i)}_{cell_j}$) is determined through logic synthesis, while f_{cell_j} can be derived through cell library characterization under different temperatures.

We implement our methodology on two PDKs, TSMC28 and SMIC40, to prove its feasibility and generality. The temperature range is set as 273K to 393K. The $\mathbb{F}_{basic}$ includes combinational circuits from the EPFL benchmark [25] and manually designed sequential circuits with equivalent functions. The $\mathbb{C}_{basic}$ consists of 24 cells, including inverters, buffers, NAND, NOR, AND, OR, XOR, XNOR gates, D-flip flops and latches, with two drive strengths per cell type.

To determine f_{cell_j}, we perform library characterization on basic cells under different temperatures, assuming a fan-out of four (FO4) load capacitance. This process utilizes

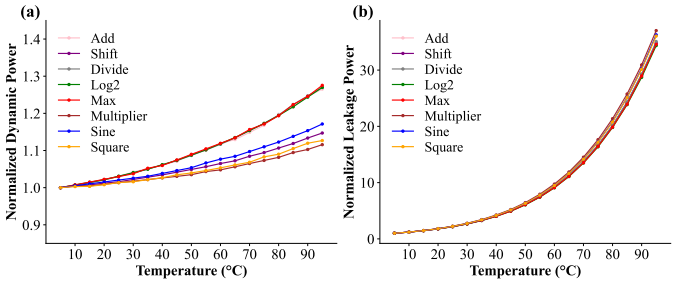

Fig. 5. Temperature scaling of (a) dynamic power (b) leakage power of functional units in TSMC28 process.

BTDcell and BTDsim, commercial tools from BTD Tech [26], functionally equivalent to SiliconSmart and HSPICE. Leakage power and dynamic power values are extracted from the characterization results (.lib files). To obtain $n^{(i)}_{cell_j}$, we conduct logic synthesis on each Verilog-described $\mathrm{FU_i}$ using Cadence Design Compiler [27], with a clock frequency of 300MHz and a temperature of 338K. Table IV shows the RMSE and R^2 of our approximation compared to ground truth, which is established through simulations of the synthesized circuits under different temperatures.

Finally, we derive the general scaling function f_{FU} according to Equation 6. Referring to the statistics in OPU, we set the occurrence frequency, n_{FU_i}, as 1024 for adders and multipliers and 16 for other FUs. Users can modify n_{FU_i} for customized simulations. For memories and analog circuits, we only consider the temperature effect on leakage power due to their low overall power contribution [28]. We integrate Cacti7.0 [29] to determine f_{mem} for SRAMs and DRAMs, while f_a is derived using the same approach as PTscalar [10]. The scaling functions f_{FU}, f_{mem} and f_a are stored as lookup tables (LUTs) in SYSgen.

D. Electrothermal Coupling

The electrothermal coupling is conducted through an iterative simulation of our chiplet architecture simulator and thermal analyzer, as illustrated in Fig. 3. The iteration continues until the temperature difference between two consecutive iterations is less than 0.1K. To be specific, chiplet architecture simulator conducts event-driven simulations and reports power, which is then passed to the thermal analyzer for temperature calculation. Incorporating the given structural parameters of packages, the thermal analyzer computes the temperature, which is fed back to the chiplet architecture simulator. The power consumptions of architectural components are adjusted according to the scaling functions f_{FU}, f_{mem} and f_a, using Lagrange interpolation when referencing the LUTs. In most cases, the iteration converges within ten rounds.

E. Vertical Interconnect (VIC) Planning

In this section, we propose a comprehensive VIC planning strategy that accounts for power delivery, signal integrity, and thermal management. Our approach aims to minimize the number of VICs while ensuring the following constraints:

TABLE IV
CONFIGURATIONS OF BASIC FUS AND APPROXIMATION ERRORS OF f_{FU_i} FOR LEAKAGE POWER AND DYNAMIC POWER

Functions	Input Bits	Output Bits	Leakage Approx.		Dynamic Approx.	
			RMSE	R^2	RMSE	R^2
S40* Add	16	17	0.0125	0.9999	0.0029	0.9481
S40 Multiply	16	32	0.0213	0.9999	0.0041	0.8995
S40 Divide	16	32	0.0080	0.9999	0.0028	0.9473
S40 Max	16	16	0.0244	0.9999	0.0026	0.9564
S40 Shift	16	16	0.0571	0.9999	0.0034	0.9414
S40 Sqrt	16	8	0.0205	0.9999	0.0019	0.9768
S40 Square	16	32	0.0054	0.9999	0.0038	0.9253
S40 Sine	16	16	0.0708	0.9999	0.0021	0.9739
S40 Average	-	-	0.0233	0.9999	0.0034	0.9356
T28 Average	-	-	0.0607	0.9998	0.0107	0.8834

*S40 for SMIC40, T28 for TSMC28

- *Signal Connectivity:* Chiplets typically connect with other chiplets or off-chip memory. These connections are established via μbumps and RDLs in the silicon interposer (for 2.5D packaging) or base die (for 3D packaging). Notably, signals do not traverse TSVs within the silicon interposer/base die or the underlying C4 bumps. To guarantee reliable interconnections, a sufficient number of VICs must be allocated. Such VICs for signal transmission are referred to as signal VICs.
- *Power Routing:* Power is supplied from C4 bumps, propagating through TSVs and μbumps and reaching the chiplets. Due to electromigration (EM) constraints, each bump and TSV has an upper current limit, denoted as I_{max}. The issue of power delivering is particularly critical in 3D integration, where bottom-layer bumps and TSVs must sustain the cumulative current of all upper-layer chiplets. To prevent excessive current density, an adequate number of VICs must be allocated. We denote the VICs for power/ground networks as power VICs.
- *Thermal Integrity:* Thermal integrity issues commonly arise in 3D packaging, where chiplets are stacked vertically, making heat dissipation ineffective for chiplets in the lower layers. As demonstrated in Section II, μbumps and TSVs can form heat channels, transferring heat from bottom to upper layers. While signal VICs and power VICs inherently contribute to heat dissipation, thermal VICs, which refers to those VICs solely for thermal management, should be added when necessary to enhance heat dissipation efficiency. Sufficient VICs must be allocated to ensure that the maximum chiplet temperature remains below a predefined threshold T_t, for chiplet functionality.

Denoting the set of μbump layer, C4 bump layer and TSV layer as $\mathbb{L}_{\mu b}$, $\mathbb{L}_{C4b}$ and $\mathbb{L}_{TSV}$, the issue of bump & TSV allocation can be represented in such a mathematical programming (MP) problem:

$$\begin{aligned} \min \quad & \alpha N_{\mu b} + \beta N_{C4b} + \gamma N_{TSV} \\ s.t. \quad & N^{(i)}_{vi} \geq N^{(i)}_{sig}, \quad \text{for layer}_i \in \mathbb{L}_{\mu b} \cup \mathbb{L}_{C4b} \cup \mathbb{L}_{TSV} \\ & I \leq I_{max}, \quad \text{for all bumps and TSVs} \\ & T_{max} \leq T_t, \quad \text{for all chiplets} \end{aligned} ,$$

TABLE V
VIC PLANNING RESULTS FOR 2.5D AND 3D PACKAGING

Case	Layers	Singal VICs	Power VICs	Thermal VICs	Total
3D	µbump-Die1	1989	352	1229	3570
	µbump-Die2	1989	528	1053	3570
	µbump-Die3	1989	1232	349	3570
	µbump-Die4	1989	1408	173	3570
	Die1	-	-	-	-
	Die2+TSV	1989	352	1229	3570
	Die3+TSV	1989	528	1053	3570
	Die4+TSV	1989	1232	349	3570
	BaseDie+TSV	-	136	-	136
	C4bump	-	136	-	136
2.5D	µbump-Die1	1989	176	-	2165
	µbump-Die2	1989	352	-	2341
	µbump-Die3	1989	176	-	2165
	µbump-Die4	1989	704	-	2693
	Interposer+TSV	-	136	-	136
	C4bump	-	136	-	136

Chiplet Temperature in 3D case (K)				Chiplet Temperature in 2.5D case (K)			
Die1	Die2	Die3	Die4	Die1	Die2	Die3	Die4
319.53	328.76	336.3	337.97	323.16	322.97	322.95	322.97

where the objective function is a combination of the number of µbumps, C4 bumps and TSVs. α, β and γ are determined according to the manufacturing cost. $N_{vi}^{(i)}$ is the number of vertical connections (µbumps, C4 bumps or TSVs) in the ith layer. $N_{sig}^{(i)}$ is the number of required signal channels. T_t is the target temperature. The MP is solved through an existing MP solver, while the T_{max} is calculated using our thermal model.

To enhance usability, SYSgen incorporates pin maps for commonly used I/O ports as built-in libraries, including PCIe [24], UCIe [23], HBM [30], and Ethernet [31]. The signal connectivity requirements can be determined accordingly.

IV. EXPERIMENTS AND CASE STUDIES

In this section, we evaluate the proposed VIC planning strategy in two scenarios: one for 3D integration and the other for 2.5D integration. We report the number of VICs, along with the corresponding chiplet temperatures. Furthermore, we compare the total number of VICs with existing approaches to demonstrate the advantages of SYSgen.

A. VIC Planning for 2.5D and 3D Packaging

We test our VIC planning strategy on the MCore-OPU architecture [32], which consists of four logic cores. The four cores are partitioned into four chiplets. In the case of 3D packaging, these chiplets are stacked vertically, whereas in 2.5D packaging, the chiplet-level floorplan is generated using Floorplet [13], forming a 2×2 array. The chiplets are interconnected via a ring structure [33] using UCIe protocol. For memory access, four chiplets share a common HBM. The average power density of the four chiplets is around 0.2W/mm^2, while the exact power trace is profiled using our chiplet architecture simulator. The area of each chiplet is 70.76mm^2.

The height of µbumps, C4 bumps, TSVs and RDLs are set as 40µm, 100µm, 200µm and 20µm, respectively. The radius of µbumps, C4 bumps and TSVs are defined as 20µm, 50µm, and 20µm. The maximum current density of µbumps, TSVs and C4 bump is set as 50mA, 20mA and 200mA. The RDL line width is 20µm, and the pitches of VICs are optimized in SYSgen. The target temperature T_t is set as 338K.

TABLE VI
COMPARISON OF OUR VIC PLANNING AGAINST PREVIOUS WORKS

#Stacks	Area (mm²)	Total number of VICs			Reduction of VICs	
		[16]	[12]	Ours	v.s. [16]	v.s. [12]
2-stacks	32.54	4296	3711	3123	-27.3%	-15.9%
2-stacks	70.76	9342	8067	6792	-27.2%	-15.8%
2-stacks	147.23	19437	16785	14574	-25.0%	-13.2%
4-stacks	32.54	13896	12756	11389	-18.0%	-10.7%
4-stacks	70.76	29720	27740	24992	-15.9%	-9.9%
4-stacks	147.23	62870	57716	52568	-16.4%	-8.9%

When planning VICs, we divide an entire package layer into multiple blocks and consider VICs to be block-wise uniform, which means VICs are uniformly distributed within each block while the density may vary across different blocks.

The results of SYSgen VIC planning are summarized in Table V. In the case of 3D package, Die1 is the chiplet at the top and Die4 is the chiplet at the bottom. µbump-Die1 refers to the µbump layer exactly beneath Die1. The corresponding chiplet temperatures are shown in Table V. In the 3D case, the addition of thermal VICs reduces the temperature of Die4 from 379.23K to 337.97K, showing their effectiveness for improving heat dissipation.

B. Comparisons with Previous Work

We also compare our VIC planning with previous work [16] and [12]. While [16] considers both spatial and temporal variations in power, but it does not incorporate electrothermal analysis. In contrast, [12] considers temperature effect only on leakage power. Since both studies focus only on thermal management in 3D packaging, we limit our experiments to 3D cases for a fair comparison. For the test cases, we scale the chiplet in Section IV-A to a smaller version (32mm^2) and a larger version (147mm^2), maintaining an approximate power density of 0.2W/mm^2. The experiments are conducted on 2-stack and 4-stack configurations, with all other settings remaining consistent with those in Section IV-A.

The experimental results and comparisons are shown in Table VI. For simplicity, we count the total number of µbumps, TSVs and C4 bumps for comparison. Tested on six cases, our VIC planning achieves an average reduction of 21.7% and 12.4% in VIC counts compared to [16] and [12], respectively. The advantage comes from SYSgen's consideration of non-uniform power distributions and temperature effect on both leakage power and dynamic power.

V. CONCLUSIONS

This paper presents SYSgen, a framework for VIC planning under electrothermal constraints. We developed a chiplet architecture simulator that enables accurate, location-based temperature-dependent power profiling and supports chiplet-specific features such as UCIe and HBM. Additionally, we built an efficient thermal tool that simplifies complex modeling of VICs. The proposed thermal solver achieved a 97.77× speedup with a maximum error below 1.2°C for chiplet temperature around 100°C, compared to COMSOL. Compared to [16] and [12], our VIC planning reduced the number of VICs by 21.7% and 12.4%, respectively.

REFERENCES

[1] G. Moore, "Cramming More Components Onto Integrated Circuits," *Proceedings of the IEEE*, vol. 86, no. 1, pp. 82–85, 1998.

[2] Y. Hao *et al.*, "Recent Progress of Integrated Circuits and Optoelectronic Chips," *Science China Information Sciences*, vol. 64, no. 10, pp. 201401–, 2021.

[3] J. L. Hennessy *et al.*, "A New Golden Age for Computer Architecture," *Commun. ACM*, vol. 62, p. 48–60, Jan. 2019.

[4] S. Naffziger *et al.*, "Pioneering Chiplet Technology and Design for the AMD EPYC™ and Ryzen™ Processor Families : Industrial Product," in *2021 ACM/IEEE 48th Annual International Symposium on Computer Architecture (ISCA)*, pp. 57–70, 2021.

[5] D. Suggs *et al.*, "The AMD "Zen 2" Processor," *IEEE Micro*, vol. 40, no. 2, pp. 45–52, 2020.

[6] J. H. Lau, "Recent Advances and Trends in Advanced Packaging," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 12, no. 2, pp. 228–252, 2022.

[7] S. Pasricha, "Exploring Serial Vertical Interconnects for 3D ICs," in *Proceedings of the 46th Annual Design Automation Conference*, DAC '09, (New York, NY, USA), p. 581–586, Association for Computing Machinery, 2009.

[8] W. Huang *et al.*, "HotSpot: A Compact Thermal Modeling Methodology for Early-Stage VLSI Design," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 14, no. 5, pp. 501–513, 2006.

[9] N. Binkert *et al.*, "The gem5 simulator," *SIGARCH Computer Architecture News*, vol. 39, p. 1–7, Aug. 2011.

[10] W. Liao *et al.*, "Temperature and Supply Voltage Aware Performance and Power Modeling at Microarchitecture Level," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 24, no. 7, pp. 1042–1053, 2005.

[11] Y. Yu *et al.*, "OPU: An FPGA-Based Overlay Processor for Convolutional Neural Networks," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 28, no. 1, pp. 35–47, 2020.

[12] K. Wang *et al.*, "Rethinking Thermal Via Planning with Timing-Power-Temperature Dependence for 3D ICs," in *Proceedings of the 16th Asia and South Pacific Design Automation Conference*, ASPDAC '11, p. 261–266, IEEE Press, 2011.

[13] S. Chen *et al.*, "Floorplet: Performance-Aware Floorplan Framework for Chiplet Integration," *Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 43, p. 1638–1649, June 2024.

[14] J. Cong *et al.*, "Thermal Via Planning for 3-D ICs," in *Proceedings of the 2005 IEEE/ACM International Conference on Computer-Aided Design*, ICCAD '05, (USA), p. 745–752, IEEE Computer Society, 2005.

[15] B. Goplen *et al.*, "Thermal Via Placement in 3D ICs," in *Proceedings of the 2005 International Symposium on Physical Design*, ISPD '05, (New York, NY, USA), p. 167–174, Association for Computing Machinery, 2005.

[16] H. Yu *et al.*, "Thermal Via Allocation for 3D ICs Considering Temporally and Spatially Variant Thermal Power," in *ISLPED'06 Proceedings of the 2006 International Symposium on Low Power Electronics and Design*, pp. 156–161, 2006.

[17] H. Yu *et al.*, "Simultaneous Power and Thermal Integrity Driven Via Stapling in 3D ICs," in *2006 IEEE/ACM International Conference on Computer Aided Design*, pp. 802–808, 2006.

[18] X. Meng *et al.*, "Stackable Thermal Model for 3D Integration," in *2025 International Symposium of Electronics Design Automation (ISEDA)*, pp. 1–7, 2025.

[19] W. Yao *et al.*, "Modeling and Application of Multi-Port TSV Networks in 3-D IC," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 32, no. 4, pp. 487–496, 2013.

[20] H. Hua *et al.*, "Exploring Compromises Among Timing, Power and Temperature in Three-Dimensional Integrated Circuits," in *2006 43rd ACM/IEEE Design Automation Conference*, pp. 997–1002, 2006.

[21] Z. Liu *et al.*, "Compact Lateral Thermal Resistance Model of TSVs for Fast Finite-Difference Based Thermal Analysis of 3-D Stacked ICs," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 33, no. 10, pp. 1490–1502, 2014.

[22] T. Chen *et al.*, "TVM: An Automated End-to-End Optimizing Compiler for Deep Learning," in *13th USENIX Symposium on Operating Systems Design and Implementation (OSDI 18)*, (Carlsbad, CA), pp. 578–594, USENIX Association, Oct. 2018.

[23] D. Das Sharma *et al.*, "Universal Chiplet Interconnect Express (UCIe): An Open Industry Standard for Innovations With Chiplets at Package Level," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 12, no. 9, pp. 1423–1431, 2022.

[24] PCI-SIG, "PCI Express ® Base Specification Revision 5.0 Version 1.0," May 2019.

[25] L. Amarú *et al.*, "The EPFL Combinational Benchmark Suite," in *Proceedings of the 24th International Workshop on Logic & Synthesis (IWLS)*, 2015.

[26] BTD-Technology, *BTD-Sim*. Shenzhen, Guangdong, China.

[27] I. Cadence Design Systems, *Genus Synthesis Solution*. San Jose, CA, USA.

[28] T. Vogelsang, "Understanding the Energy Consumption of Dynamic Random Access Memories," in *2010 43rd Annual IEEE/ACM International Symposium on Microarchitecture*, pp. 363–374, 2010.

[29] R. Balasubramonian *et al.*, "CACTI 7: New Tools for Interconnect Exploration in Innovative Off-Chip Memories," *ACM Transactions on Architecture and Code Optimization (TACO)*, vol. 14, June 2017.

[30] J. S. S. T. Association, "High Bandwidth Memory (HBM) DRAM JESD235D," March 2021. Standard No. JESD235D.

[31] I. L. S. Committee, "IEEE Standard for Ethernet," *IEEE Std 802.3-2022 (Revision of IEEE Std 802.3-2018)*, pp. 1–7025, 2022.

[32] S. Lu *et al.*, "An FPGA-based Multi-Core Overlay Processor for Transformer-based Models," in *2024 2nd International Symposium of Electronics Design Automation (ISEDA)*, pp. 697–702, 2024.

[33] S. Bourduas *et al.*, "Modeling and Evaluation of Ring-Based Interconnects for Network-on-Chip," *Journal of Systems Architecture*, vol. 57, no. 1, pp. 39–60, 2011. Special Issue On-Chip Parallel And Network-Based Systems.