

A Novel 2.5D Thermal Coupling Model for Device/Metal Reliability Applications

J.Y. Kuo*, Gary Lee, Kasa Huang, Jeffrey Tzeng, Y.S. Lin, S.H. Li, Y.L. Chang, H.P. Chien, P.C. Hsu, G.T. Lin, W.C. Chang, Eugene Chen, Y.C. Luo, C. C. Lai, Y.Y. Chang, Y.H. Tsai, C. Hsiao, W. Lee, S.H. Shen, C.W. Chang, Tony Yu, M.H. Hsieh, Clement Huang, K.Y. Su, Jim Liang, K.W. Su, and C.K. Lin

Taiwan Semiconductor Manufacturing Company, Hsinchu, Taiwan

*Email: jykuok@tsmc.com

Abstract—In this paper, we present a novel thermal model for transistor-level device and metal hotspot simulation that offers both high accuracy and simulation efficiency, accounting for complex layout-dependent thermal coupling effects in advanced technology nodes. Compared to traditional 3D FEM methods, which present a trade-off between accuracy and computational efficiency for large-scale circuits, our proposed 2.5D method, incorporating a coupling model approach, achieves superior simulation efficiency while maintaining high accuracy. To achieve this, our model leverages a machine-learning-assisted 2.5D thermal network extraction flow, which is seamlessly integrated within traditional EDA electrical parasitic R/C extraction tools. Consequently, the temperatures of devices and metals in FEOL/MOL/BEOL can be resolved concurrently with electrical simulations. Our thermal model demonstrates significant potential for reliability-aware co-optimization, spanning from layout design to final electrical performance and hardware reliability.

Keywords—thermal modeling, reliability

I. INTRODUCTION

Thermal management has become increasingly critical for device reliability due to higher power density, more confined structures, and reduced thermal conductivity in advanced technology nodes [1]-[6]. These factors exacerbate the self-heating effect within the device, which in turn influences the temperature of the metal segments above the device channel, potentially leading to reliability concerns such as electromigration. Therefore, an accurate thermal model is required for device and metal reliability evaluations.

Due to the continuous scaling of advanced technology nodes, decreased inter-device spacing leads to more pronounced thermal coupling effects, which in turn accelerate device aging. As illustrated in Fig. 1, devices with multiple active areas (ODs) suffer from more severe degradation due to elevated temperatures (Fig. 2). Furthermore, the thermal coupling effect between adjacent ODs is a complex, layout-dependent phenomenon (Table 1).

In addition, the thermal coupling between MOS devices and metal layers is significant and exhibits complex layout dependency. As illustrated in Fig. 3, the thermal coupling from the device channel temperature to the MG/MD/VD metals, and even to the M0/M1 layers, is substantial and non-negligible. Furthermore, the magnitude of this thermal coupling effect is highly sensitive to the layout configuration. Consequently, accurately predicting metal temperatures for electromigration (EM) assessment using conventional rule-based methods has become increasingly challenging.

While the 3D Finite Element Method (FEM) is a standard approach for thermal analysis, its application in circuit-level

reliability analysis involves a significant trade-off between accuracy and computational efficiency. Generally, hotspots within FEOL/MOL devices and metal layers are not fully resolvable using traditional 3D FEM at this scale in advanced technology nodes.

To address these challenges, a SPICE-based 2.5D thermal model is proposed [7]-[9], leveraging the fundamental analogy between thermal and electrical governing equations

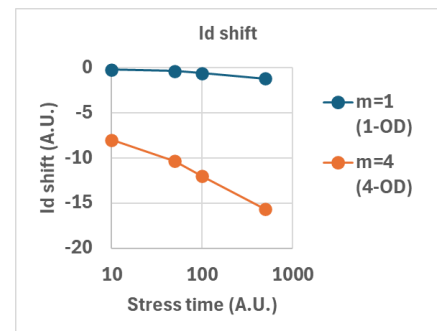


Fig. 1. The device with four active ODs ($m=4$) exhibits more severe I_d degradation than the device with a single active OD ($m=1$).

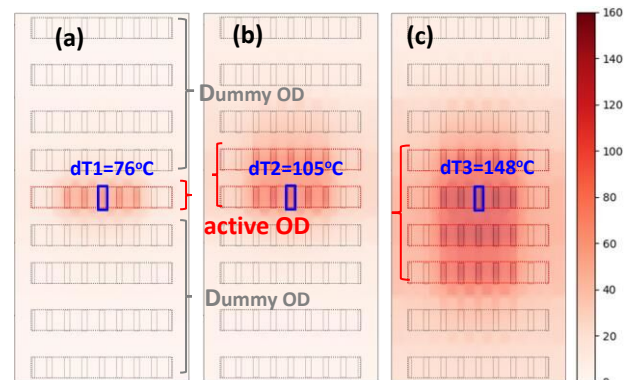


Fig. 2. Temperature profiles obtained via FEM thermal simulation for devices with (a) $m=1$, (b) $m=2$, and (c) $m=4$. The device with $m=4$ exhibits a significant temperature rise due to enhanced thermal coupling effects.

Table 1. MD/PO layout dependency of the thermal coupling ratio between two devices on separate ODs, obtained via FEM simulations.

	Case-1	Case-2	Case-3	Case-4
MD	Open	Short	Open	Short
PO	Open	Open	Short	Short
Coupling ratio ($dT2/dT1$)	0.06	0.35	0.29	0.39
Layout				

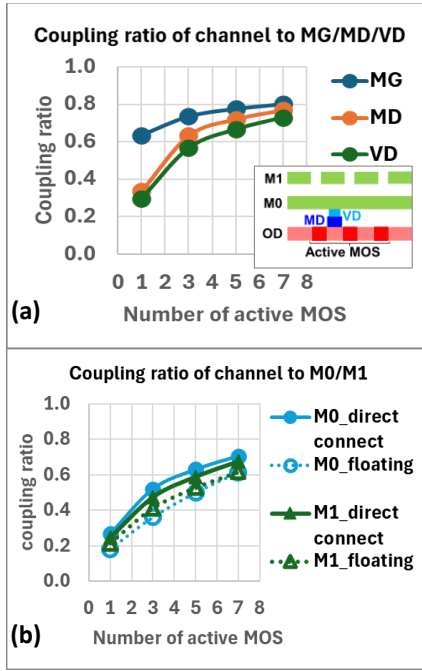


Fig. 3. Coupling ratios of channel temperature to (a) MG/MD/VD and (b) M0/M1, exhibiting layout dependency on the underlying active MOS devices as demonstrated in the inset via FEM thermal simulations.

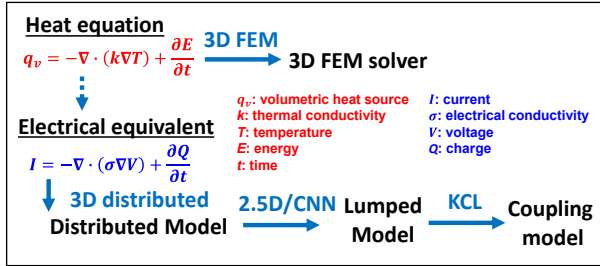


Fig. 4. Overview of thermal analysis methodologies, comparing traditional 3D FEM with the proposed SPICE-based framework, categorized into distributed, lumped, and coupling models.

which allows thermal behavior to be simulated using standard SPICE techniques. The 2.5D method, which strikes an optimal balance between the speed of 2D extraction and the accuracy of 3D modeling, is widely employed for electrical parasitic R/C extraction[10]. As illustrated in Fig. 4, the thermal network can be extracted as a lumped model using this approach. For reliability-aware simulations, the lumped model can be further simplified into a coupling model to significantly enhance simulation throughput.

II. 2.5D THERMAL MODEL AND COUPLING MODEL

Based on our previous work using the 2.5D method to extract electrical parasitic R/C[10], we extend the 2.5D methodology to the extraction of thermal resistances (R_{th}). A dedicated thermal technology file is developed to establish a comprehensive thermal network encompassing the FEOL, MOL, and BEOL structures. Unlike electrical current, which is confined to conductive paths, heat dissipation occurs not only within the metal interconnects but also through the surrounding dielectric layers. Therefore, it is imperative to extract the R_{th} associated with heat transfer from the metals across the dielectric media (e.g., as depicted in Fig. 5(d)).

To enhance the accuracy of R_{th} extraction, a thermal-aware convolutional neural network (CNN)-based model is applied

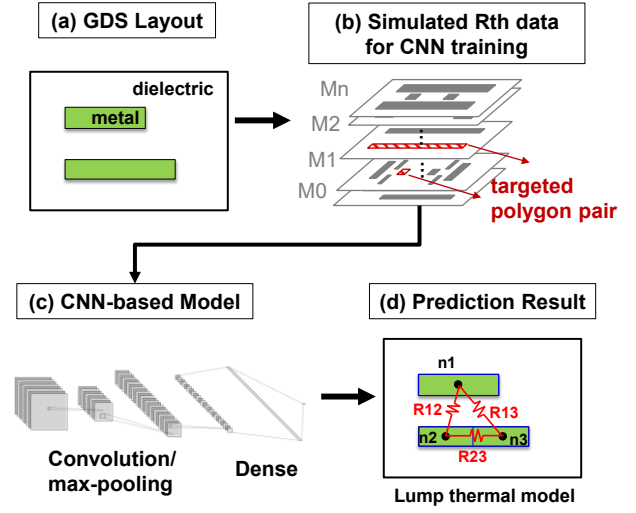


Fig. 5. The flow-chart of our CNN-based modeling methodology. As demonstrated in (d), there are 2 types of R_{th} , inside-metal R_{th} (e.g., R23) and cross-dielectric R_{th} (e.g., R12).

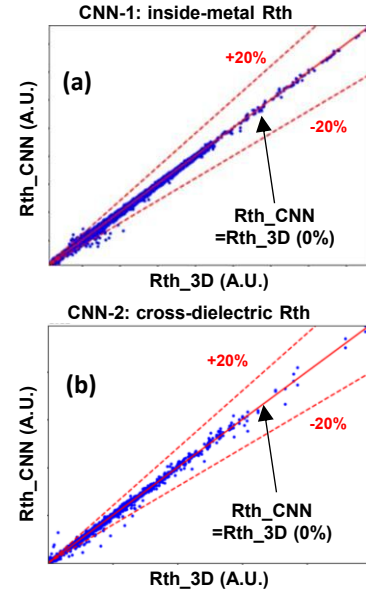


Fig. 6. The (a) inside-metal R_{th} and (b) cross-dielectric R_{th} can be modeled by CNN.

(Fig. 5) [11]. The model, leveraging the ResNet34 architecture, is trained using layout features derived from drawn GDS patterns and their corresponding ground-truth R_{th} values extracted from 3D simulations. The training dataset comprises 20,000 pairs of layout snapshots and simulated thermal results. It is worth noting that the profiles used in the 3D thermal simulations are consistent with the actual profile employed in the parasitic R/C extraction. As shown in Fig. 6, the R_{th} predicted by the CNN model (R_{th_CNN}) closely matches the R_{th} extracted by the 3D distributed model (R_{th_3D}).

To further enhance the computational efficiency of thermal simulation, the lumped model can be simplified into a compact coupling model, as demonstrated in Fig. 7. This approach leverages the principle of superposition, where the total temperature rise (dT_{total}) at a specific device or metal segment is expressed as the linear summation of thermal contributions from all relevant heat sources. Each contribution is weighted by a Coupling Ratio (CR), which quantifies the thermal influence of a source on a target location. Notably, these CRs are intrinsic to the thermal network's topology; once

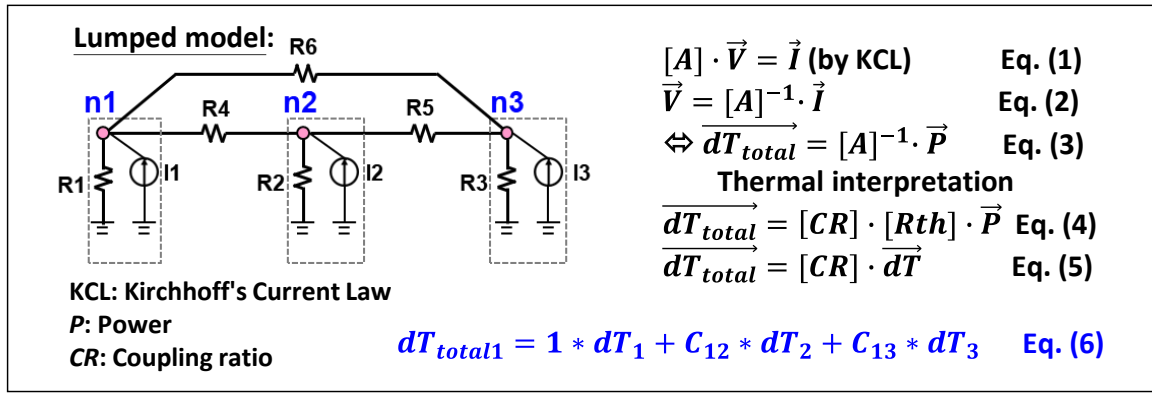


Fig. 7. Derivation of the coupling model from the lumped model. The coupling ratio (CR) matrix can be determined once the network is given.

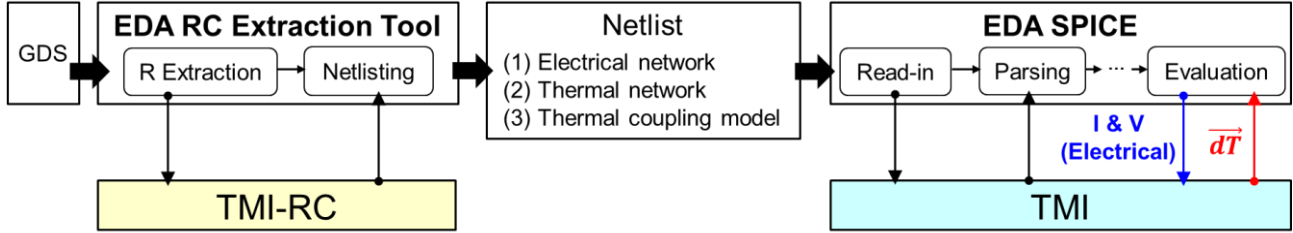


Fig. 8. Thermal simulation and 2.5D thermal network extractor are embedded in industry-standard EDA tools.

the RC network is extracted, the coupling ratios remain constant, allowing for near-instantaneous temperature evaluation without the need for repetitive and computationally expensive matrix solvers.

Our 2.5D thermal model is seamlessly integrated into the industry-standard EDA environment through the TSMC Model Interface (TMI), as illustrated in Fig. 8. During the RC extraction stage, TMI-RC is utilized to generate the thermal network with a specific thermal technology file. This process translates the layout geometries into R_{th} for both metal interconnects and dielectric layers, effectively mapping the electrical parasitics to their thermal counterparts. Simultaneously, a topologically invariant thermal coupling model is established to account for spatial heat spreading. Subsequently, the electrical netlist and the thermal coupling model are imported into the SPICE simulator. Leveraging the TMI framework, the simulator performs a self-consistent electro-thermal analysis, ultimately yielding a comprehensive and accurate temperature profile across both device and interconnect segments.

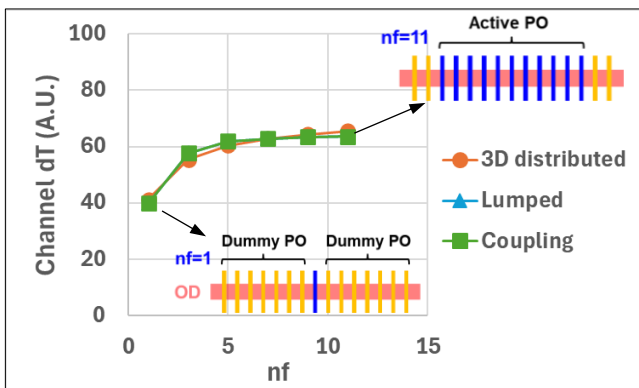


Fig. 9. A verification of the lumped model and coupling model. The device channel temperature at center of various finger (nf) devices can be accurately expressed by the lumped model and coupling model.

III. RESULT AND DISCUSSION

The accuracy of the proposed lumped and coupling models is validated by comparing their predictions with 3D distributed thermal simulation results, as shown in Fig. 9. The channel temperature rise at the center finger, which typically represents the worst-case thermal stress due to heat accumulation from neighboring fingers is examined. The results indicate that as the finger count (n_f) increases, the temperature rise predicted by both the lumped and coupling models exhibits good agreement with the 3D reference data.

The computational performance of the various thermal models is summarized in Table 2. While maintaining the high fidelity demonstrated in Fig. 9, the 2.5D thermal models exhibit vastly superior simulation efficiency. In contrast to the 3D distributed thermal model, the 2.5D models drastically reduce the number of simulation nodes, thereby minimizing both memory overhead and CPU time. Furthermore, within the 2.5D framework, the coupling model proves to be even more efficient than the lumped model, as it further streamlines thermal interactions into a compact mathematical form. This exceptional efficiency makes the coupling model an ideal candidate for large-scale, real-time reliability assessments and iterative design optimization.

Table 2. With the comparable accuracy shown in Fig. 9, 2.5D thermal models exhibit superior simulation efficiency as compared with the 3D distributed model. In addition, the coupling model is more efficient than the lumped model.

Method		# of Rth in the network	Runtime (sec)	Efficiency
3D distributed model		4,000,000	500	Ref (1X)
2.5D	Lumped model (electrical and thermal networks)	2,000	0.5	1000X
	Coupling model (electrical network with coupling ratio)	800	0.25	2000X

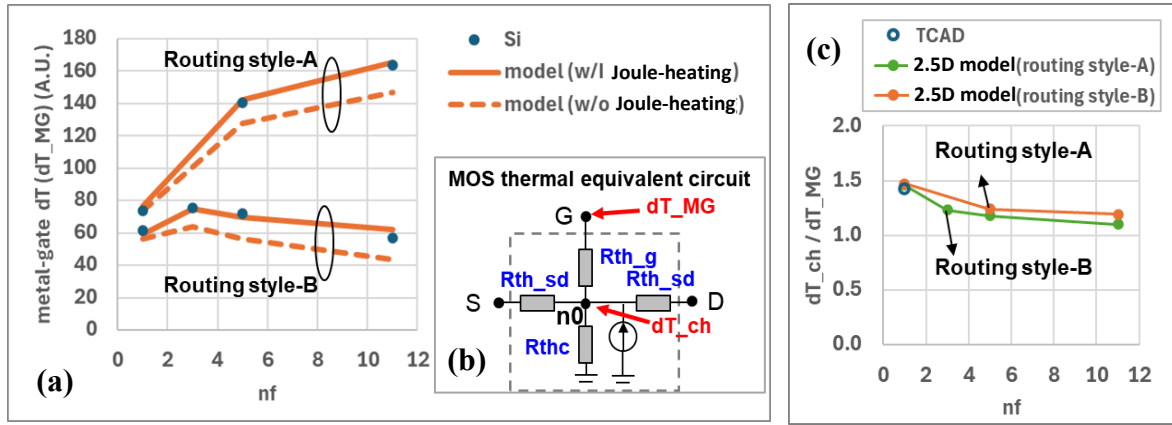


Fig. 10. (a) The 2.5D thermal model can accurately predict the Si data for various metal routing styles and with metal Joule heating considered. (b) A MOS thermal equivalent circuit is adopted to address the complicated 3D device profile. (c) The channel and metal gate temperature can be resolved as well.

Silicon validation of the proposed thermal model is further demonstrated in Fig. 10. Benchmarked against the metal-gate temperature extracted via the gate resistance method, the 2.5D model accurately predicts the center-finger temperature for various n_f devices across two distinct metal routing configurations (Routing Style-A and Style-B, as shown in Fig. 10(a)). By accounting for the varying power dissipation associated with different routing styles, the contribution of the metal Joule-heating effect to the overall gate temperature is effectively resolved. Notably, a MOS thermal equivalent circuit is implemented to represent the complex 3D device architecture (Fig. 10(b)). This equivalent circuit incorporates R_{thc} , $R_{th_{sd}}$, and R_{th_g} , which represent the intrinsic channel thermal resistance, the channel-to-source/drain coupling resistance, and the channel-to-gate coupling resistance, respectively. Consequently, both the channel and metal-gate temperatures can be precisely determined, as illustrated in Fig. 10(c). The extracted channel temperature is then integrated into the TMI framework to facilitate high-fidelity aging and reliability simulations.

Although 3D FEM remains the industry standard for chip and package-level thermal analysis, the proposed 2.5D methodology can complement 3D FEM by leveraging it for precise boundary condition determination. This hybrid approach enables high-fidelity resolution of localized hotspot regions where precision is paramount, while maintaining the computational efficiency necessary for large-scale layout assessments. Ultimately, this integrated framework provides a robust foundation for the comprehensive evaluation of device aging and long-term reliability under realistic electro-thermal operating conditions.

IV. CONCLUSIONS

In this paper, a novel 2.5D thermal simulation framework is proposed to address complex layout-dependent thermal effects in advanced technology nodes. By leveraging a topologically invariant coupling model, the proposed approach achieves high predictive fidelity while delivering substantial computational speedup over traditional 3D FEM simulations. The seamless integration of this framework into industry-standard EDA tools via the TMI environment enables the concurrent resolution of device and interconnect temperatures across the FEOL, MOL, and BEOL layers. Silicon validation rigorously confirms the model's capability

to capture localized self-heating and metal Joule-heating effects across diverse routing configurations. Ultimately, this work provides a scalable and efficient solution for reliability-aware co-optimization, effectively bridging the gap between layout design, electrical performance, and hardware reliability from the early design stages to final sign-off.

REFERENCES

- [1] W.-Y. Woon et al., "Integration and Characterization of High Thermal Conductivity Materials for Heat Dissipation in Stacked Devices," *2024 IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits)*, Honolulu, HI, USA, 2024, pp. 1-2.
- [2] T. Chou et al., "Transistor-to-Package Thermal Simulation with Hotspot Mitigation by Decoupling," *2025 IEEE International Electron Devices Meeting (IEDM)*, San Francisco, CA, USA, 2025, pp. 1-4.
- [3] D. Shin et al., "Device-To-Package Level Thermal Risk Analysis of the Back-Side Power Delivery Network," *2025 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, Grenoble, France, 2025, pp. 1-4.
- [4] B. Vermeersch et al., "Multiscale Thermal Impact of BSPDN: SoC Hotspot Challenges and Partial Mitigation," *2024 IEEE International Electron Devices Meeting (IEDM)*, San Francisco, CA, USA, 2024, pp. 1-4.
- [5] Y. Chen, M. Brunion, M. Walker, D. Biswas, J. Ryckaert and J. Myers, "A Multi-Die Self-Consistent Framework for Thermal Stability Analysis in Nanosheet 3D SiPs," *2026 IEEE International Reliability Physics Symposium (IRPS)*, Tucson, AZ, USA, 2026, pp. 1-6.
- [6] S. Wang et al., "Fast Full-Chip Thermal Assessment Methodology for Emerging BSPDN Architectures," *2026 IEEE International Reliability Physics Symposium (IRPS)*, Tucson, AZ, USA, 2026, pp. 1-5.
- [7] L. Jiang et al., "Thermal modeling of on-chip interconnects and 3D packaging using EM tools," *2008 IEEE-EPEP Electrical Performance of Electronic Packaging*, San Jose, CA, USA, 2008, pp. 279-282.
- [8] L. Jiang, Chuan Xu, H. Smith, B. Rubin, A. Deutsch and A. Caron, "Electrical modelling of temperature distributions in on-chip interconnects, packaging, and 3D integration," *2010 Asia-Pacific International Symposium on Electromagnetic Compatibility*, Beijing, China, 2010, pp. 625-628.
- [9] Ting-Yen Chiang, K. Banerjee and K. C. Saraswat, "Compact modeling and SPICE-based simulation for electrothermal analysis of multilevel ULSI interconnects," *IEEE/ACM International Conference on Computer Aided Design. ICCAD 2001*, San Jose, CA, USA, 2001, pp. 165-172.
- [10] O. C. H. Yu et al., "A Novel CNN-Based BEOL Contour Modeling Solution for Parasitic R/C Extraction," *2025 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, Grenoble, France, 2025, pp. 1-4.
- [11] M. L. Lu et al., "Applications of Machine learning techniques in the Compact Modeling Platform," *2026 International Compact Modeling Conference (ICMC)*, Long Beach, CA, USA, 2026, pp. 1-4.