Thermal Implications of Non-Uniform Power in BSPDN-Enabled 2.5D/3D Chiplet-based Systems-in-Package using Nanosheet Technology

Yukai Chen*, Massimiliano Di Todaro*‡, Bjorn Vermeersch*, Herman Oprins*, Daniele Jahier Pagliari‡, Julien Ryckaert* Dwaipayan Biswas* James Myers†
*IMEC, Leuven 3000, Belgium
†IMEC, Cambridge CB1 2JD, United Kingdom
‡DAUIN, Politecnico di Torino, Turin 10129, Italy
Email: yukai.chen@imec.be

***Abstract*—Advances in nanosheet technologies have significantly increased power densities, exacerbating thermal management challenges in 2.5D/3D chiplet-based Systems-in-Package (SiP). While traditional thermal analyses often employ uniform power maps to simplify computational complexity, this practice neglects localized heating effects, leading to inaccuracies in thermal estimations, especially when comparing power delivery networks (PDN) in 3D integration. This work examines the thermal impact of non-uniform power distributions on SiPs utilizing frontside (FSPDN) and backside (BSPDN) power delivery approaches. Using high-resolution thermal simulations with non-uniform power maps at resolutions down to $5\,\mu\mathrm{m}$, we demonstrate that uniform power assumptions substantially underestimate peak temperatures and fail to reveal critical thermal differences between BSPDN and FSPDN configurations in 3D scenarios. Our results highlight that BSPDN configurations in 3D, although beneficial in simplified uniform scenarios, exhibit pronounced thermal penalties under realistic, localized workloads due to limited lateral heat spreading. These findings emphasize the necessity of adopting fine-grained, workload-aware power maps in early-stage thermal modeling to enable accurate PDN assessment and informed thermal-aware design decisions in advanced nanosheet-based 3D SiP.**

I. Introduction and Motivation

Semiconductor scaling into the nanosheet era has significantly increased computing core density, causing elevated power densities and heightened thermal management challenges in high-performance computing (HPC) systems. Chiplet-based Systems-in-Package (SiP), integrating multiple dies either horizontally (2.5D integration) or vertically (3D integration), compound these thermal challenges due to their compact and complex multi-layered structures. Consequently, effective heat dissipation has emerged as a critical design requirement, influencing system performance and reliability.

Backside Power Delivery Networks (BSPDN) have recently emerged to improve power integrity by delivering power directly from the wafer backside [1]. By relocating power rails to the backside, BSPDN reduces interconnect resistance, lowers IR drop, and enhances power efficiency, freeing frontside metal layers for optimized signal routing. However, BSPDN implementation involves extreme silicon substrate thinning, substantially affecting thermal pathways. Previous studies indicated that substrate thinning inherent to BSPDN significantly raises hotspot temperatures due to diminished lateral heat spreading. Although backside metal layers partially mitigate this effect, elevated peak temperatures remain compared to conventional Frontside Power Delivery Networks (FSPDN) [2]. While BSPDN's thermal impact has been evaluated in 2D System-on-Chip (SoC) scenarios, these findings do not generalize to 3D integrations, where vertical stacking introduces fundamentally different thermal dynamics.

Traditional early-stage thermal analyses frequently rely on simplified uniform power maps to reduce computational complexity. However, uniform power assumptions inadequately represent realistic chip operations, where computationally intensive workloads produce localized hotspots. Recent literature demonstrates significant underestimations of peak temperatures using uniform maps, particularly emphasizing these inaccuracies in 2D SoC with BSPDN [3]. Likewise, workload-based non-uniform power maps reveal that BSPDN exhibits hotspot temperatures up to 14° higher than comparable FSPDN designs in the 2D HPC SoC scenarios [4]. Such findings highlight the necessity of detailed, realistic non-uniform power distributions for accurate thermal modeling.

Nevertheless, most existing studies remain limited to 2D SoC with a single active power layer, leaving comprehensive thermal analyses for complex, vertically integrated 2.5D/3D chiplet-based SiPs largely unexplored. Addressing this critical gap, our study systematically investigates the thermal implications of employing fine-grained non-uniform power maps in BSPDN-enabled 2.5D/3D SiPs, explicitly focusing on 3D integration scenarios. Our study rigorously compares the thermal behavior of BSPDN and FSPDN under realistic workload-driven and synthetic power distributions, revealing critical hotspots and thermal gradients previously masked by uniform power assumptions in 3D chiplet integration.

In 3D design with uniform power maps, BSPDN may appear thermally superior to FSPDN, which is contradictory to the 2D analyses. However, our high-resolution thermal simulation reveals the opposite under realistic power distributions. By adopting non-uniform power maps with resolutions as fine

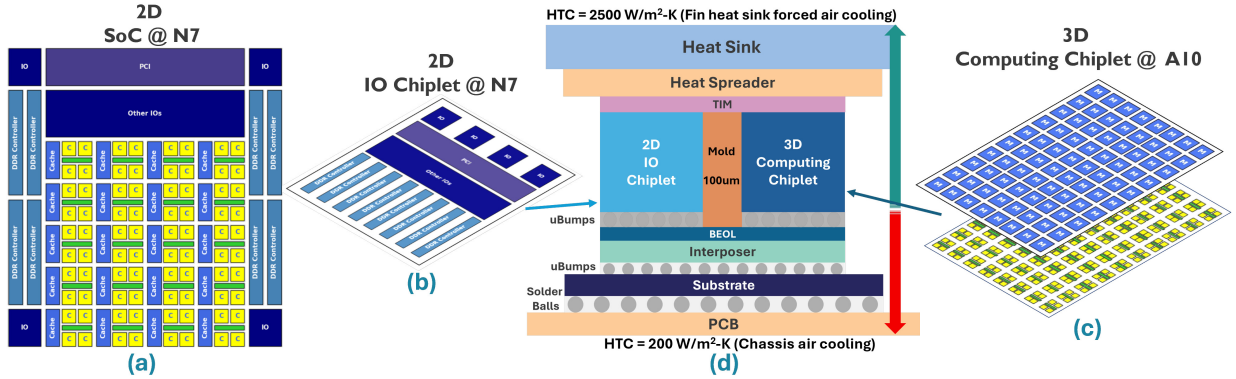

Figure 1: (a) The original 2D HPC SoC implemented in N7; (b) Floorplan of the 2D I/O chiplet disaggregated from the SoC implemented in N7; (c) Floorplan of the 3D computing chiplet implemented in A10; (d) 2.5D chiplet-based SiP thermal model.

as $5\,\mu\mathrm{m}$, we identify thermal penalties inherent to BSPDN configurations. The spatial granularity of power profiles fundamentally influences hotspot formation and thermal gradients, directly affecting the accuracy of thermal estimations and interpretations of PDN effectiveness. This study highlights that finer-grained non-uniform power maps are essential for accurate thermal modeling, thereby guiding the design and selection of effective power delivery and thermal management strategies in advanced nanosheet-based 2.5D/3D SiPs.

II. Chiplet-based SiP and Thermal Modeling

To accurately evaluate the thermal characteristics of chiplet-based SiP, we consider a representative HPC server SoC originally implemented in N7 FinFET node, referencing a contemporary industrial design [5]. Figure 1.(a) shows the mock-up floorplan of this 2D SoC design. Leveraging the disaggregation methodology outlined in [6], peripheral functional blocks, including PCI interfaces, I/O controllers, and DDR memory controllers, are grouped into a 2D IO chiplet fabricated in N7 technology for cost efficiency and moderate performance requirements (Figure 1.(b)). Compute-intensive components, such as cores, network-on-chip (NoC), and system-level caches, are integrated into a 3D computing chiplet using advanced A10 nanosheet technology for higher performance and density (Figure 1.(c)). This computing chiplet employs a 3D Memory-on-Logic (MoL) integration scheme, stacking system-level cache memory directly atop a logic die containing cores and NoC. The work [7] presents a thermal analysis based on a similar chiplet-based design, whereas it does not consider the BSPDN and non-uniform power maps.

Figure 1.(d) illustrates the resulting 2.5D SiP thermal model. The IO and computing chiplets connect through a passive silicon interposer chosen for its simplicity and cost-effectiveness, containing no active power-dissipating elements. The narrow inter-chiplet gap ($100\,\mu\mathrm{m}$) filled with a low thermal conductivity mold compound ($3\,\mathrm{W/(m\,K)}$) provides mechanical support. Two thermal dissipation paths are modeled. The primary thermal path is through the package's topside (green arrow in Figure 1.(d)), modeling a forced-air multi-fin heatsink with a high heat transfer coefficient (HTC = $2500\,\mathrm{W/(m^2\,K)}$). and a secondary thermal path through interposer and PCB, representing typical chassis-air cooling conditions (HTC = $200\,\mathrm{W/(m^2\,K)}$). Table I summarizes dimensions, thicknesses, and thermal conductivities for all stacking layers involved in both thermal dissipation paths.

Table I: Physical parameters of stacking layers.

Stacking Layers	Dimension (mm)	Thickness (um)	Thermal Conductivity (W/mK)
Heat Sink	100x100	3000	400
Heat Spreader	40x40	5000	400
TIM	32.9x22.6	250	30
IO Chiplet uBumps	19.2x22.6	10	3.5
Computing Chiplet uBumps	13.6x19.6	10	3.5
Interposer BEOL	32.9x22.6	5	1.2
Interposer	32.9x22.6	50	140
uBumps	32.9x22.6	100	6.0
Substrate	50x50	300	0.6
Solder Balls	50x50	100	8.0
PCB	100x100	800	5.0

Our thermal analysis specifically targets the 3D computing chiplet ($13.6\,\mathrm{mm}\times19.6\,\mathrm{mm}$), fabricated in A10 nanosheet technology, resulting in higher power densities compared to the thermally benign 2D IO chiplet ($19.2\,\mathrm{mm}\times22.6\,\mathrm{mm}$). Consequently, the IO chiplet primarily acts as an effective thermal spreader within the package. We focus our comparative thermal evaluation on two distinct power delivery network (PDN) architectures in the 3D computing chiplet: FSPDN and BSPDN. Both configurations employ face-to-face (F2F) hybrid bonding for vertical interconnections. Figure 2 details cross-sectional views of both stacks, including relevant thicknesses and thermal conductivities of stacking layers.

In stack 1 (FSPDN), power delivery is facilitated through a dedicated BEOL_MZ metallization stack layer, distributing power to both the logic die and the vertically integrated memory die. This configuration necessitates an ultra-thin silicon substrate ($5\,\mu\mathrm{m}$ thick) within the logic die, accommodating nTSVs for vertical interconnections. Conversely, stack 2 (BSPDN) fundamentally alters the power delivery pathway by directly supplying power through the backside of the logic die, eliminating the BEOL_MZ layer between logic and memory dies. The logic die's thinner silicon substrate is removed due to the backside contact implementation, and a BSPDN layer

is directly underneath the logic die's FEOL layer with a lower thermal conductivity ($71\,\mathrm{W/(m\,K)}$) compared to the thinner logic die's silicon substrate ($135\,\mathrm{W/(m\,K)}$) in stack 1. Transitioning from stack 1 (FSPDN) to stack 2 (BSPDN) introduces two key thermal implications: (a) **Improved vertical heat conduction** between the logic and memory dies due to the removal of the intermediate BEOL_MZ layer, thereby reducing thermal resistance at the memory-logic bonding interface; (b) **Degraded lateral heat spreading** beneath the FEOL as the high-conductivity silicon substrate is replaced by a lower-conductivity BSPDN layer, which restricts lateral heat dissipation beneath the FEOL, exacerbating localized hotspot temperatures. This fundamental trade-off highlights the complexity and importance of accurate thermal modeling in guiding PDN choices for advanced 3D SiP designs.

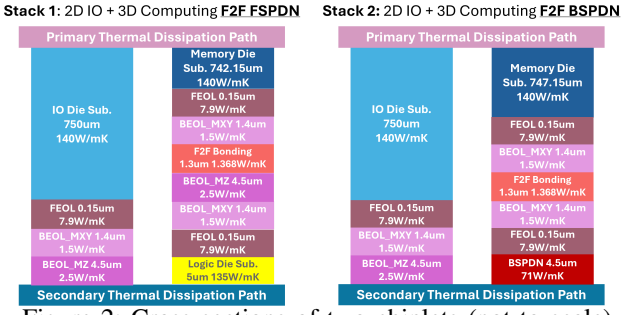

Figure 2: Cross sections of two chiplets (not to scale)

III. Results and Discussion

We evaluated the thermal impact of power map granularity and PDN architectures by conducting high-resolution thermal simulations using a hierarchical global-local methodology based on HotSpot [8]. The ambient temperature is set at $25\,^{\circ}\mathrm{C}$.

We initially performed baseline simulations employing uniform power maps at a coarse $200\,\mu\mathrm{m}$ resolution for each core within the 3D computing chiplet's logic die. Under these uniform conditions, as shown in Figure 3 left side, BSPDN appears to yield lower peak temperatures than FSPDN, primarily due to the reduced vertical thermal resistance between the memory and logic dies achieved by removing the BEOL_MZ layer in stack 2 (BSPDN). However, this advantage arises solely under idealized uniform power conditions, which do not realistically represent actual workloads.

To accurately capture realistic conditions, we selected the hottest core and applied a $5\,\mu\mathrm{m}$ resolution workload-annotated non-uniform power map using a global-local refinement flow. This methodology first applies a coarse-grained whole-chip simulation to extract boundary thermal conditions, followed by localized refinement incorporating interpolation, filtering and upscaling to preserve spatial detail. The resulting thermal map accurately captures core-level gradients while maintaining global context. Our results demonstrate that realistic, localized workloads significantly increase peak temperatures and fundamentally reverse the thermal trend observed under uniform power maps: BSPDN, which appears superior under uniform assumptions, consistently shows higher peak temperatures than FSPDN under realistic conditions. This reversal, extending and intensifying earlier 2D BSPDN observations [3], highlights the

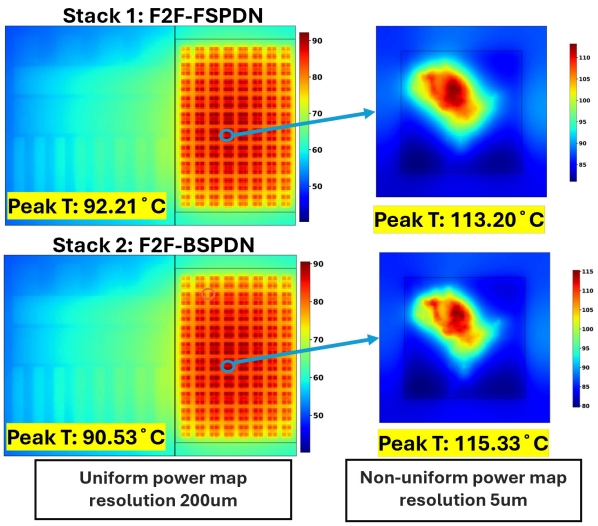

Figure 3: Thermal maps from uniform power and fine-grained Dhrystone workload-driven power map.

critical importance of spatial power modeling in evaluating PDN thermal performance in 3D chiplet systems.

The reversal of thermal ranking between BSPDN and FSPDN configurations in realistic scenarios is primarily attributed to the limited lateral heat spreading capability beneath the FEOL in BSPDN implementations. In FSPDN, the presence of a thin silicon substrate beneath the FEOL ensures effective lateral heat spreading, moderating local hotspot temperatures. Conversely, the BSPDN architecture replaces this high-conductivity substrate with a lower-conductivity BSPDN layer, severely restricting lateral heat diffusion. When power is spatially concentrated, this limitation becomes critical. Heat fails to diffuse laterally and instead accumulates beneath the hotspot in FEOL, intensifying local temperature rises. Although BSPDN benefits from improved vertical conduction due to a thinner bonding interface, the dominant heat escape path in 3D stacks remains through the top heatsink. Without effective in-plane diffusion, the vertical path becomes overwhelmed. This explains why BSPDN performs worse under realistic non-uniform power conditions and why its thermal behavior improves under uniform scenarios, allowing the heat to return more evenly through the heatsink.

To systematically explore how spatial power distribution influences thermal behavior, we conducted controlled single-core experiments using the same 3D stacking (Figure 2), eliminating thermal interference from neighboring cores. Four power maps shown in Figure 4 with identical total power were evaluated: 1) **Uniform**—power is evenly distributed across the entire core, reflecting a common early-stage modeling assumption; 2) **Realistic workload non-uniform**—extracted from workload-annotated post-PnR simulation, capturing localized activity patterns; 3) **Clustered synthetic**—the core is divided into multiple 25×25 sub-blocks, each forming a center-weighted micro-hotspot with randomized placement, emulating spatially distributed but localized compute regions; 4) **Center-focused synthetic**—high-power pixels are aggregated at the core center, creating a dominant hotspot to stress thermal

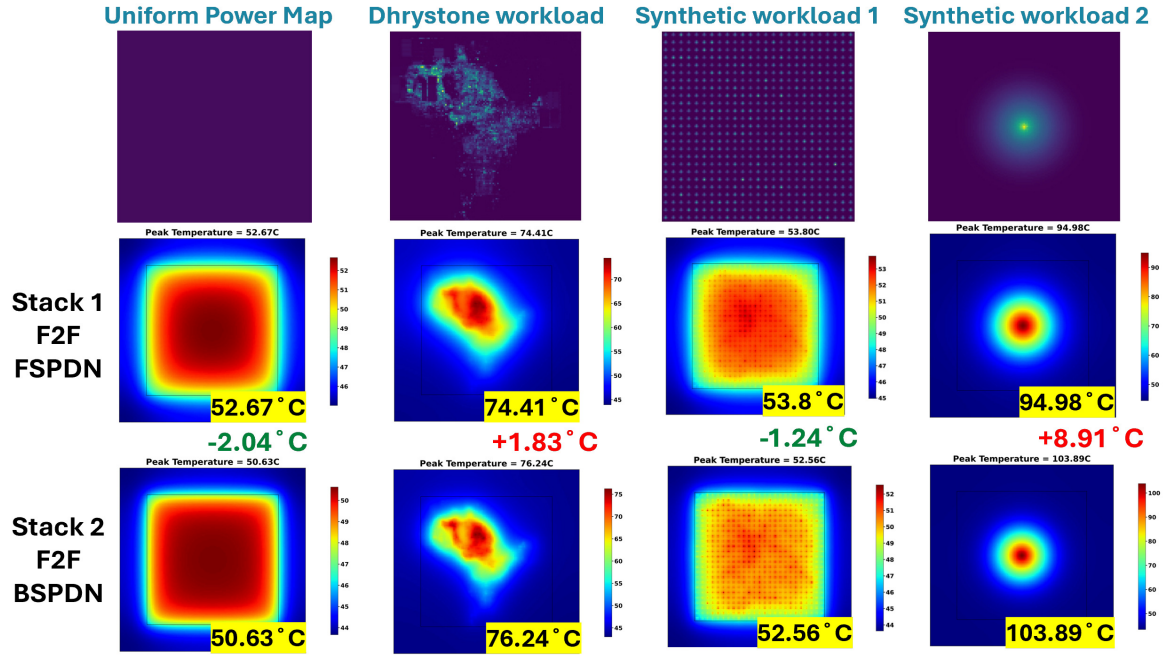

Figure 4: Four power maps used in single-core thermal analysis: (1) Uniform distribution, (2) Realistic workload non-uniform map, (3) Synthetic 1: clustered hotspots, (4) Synthetic 2: center-focused concentration. All cases have equal total power.

limits under extreme concentration.

These four maps span a controlled spectrum from uniform to highly localized power, enabling systematic analysis of how spatial shape and location of power influence thermal gradients across different PDNs. The results confirm that spatial power concentration directly amplifies BSPDN's thermal disadvantage. As the power distribution becomes more aggregated, BSPDN's peak temperature increases significantly due to the restricted lateral thermal diffusion in the low-conductivity BSPDN layer. FSPDN also exhibits peak temperature increases, but to a lesser extent, thanks to its silicon substrate's superior ability to dissipate heat laterally. The key insight is that the hotspot size and spatial concentration determine the extent to which the material beneath the FEOL can support lateral spreading. BSPDN's thermal penalty becomes pronounced when heat is confined to a limited area, overwhelming the vertical escape path to the heatsink.

Coarse-grained power assessments can underestimate peak temperatures and misrepresent PDN thermal behavior, which may lead to poor design choices. As power density increases in the nanosheet era and 3D integration becomes more complex, using detailed, high-resolution thermal modeling with workload-driven power maps is crucial for accurate early-stage thermal evaluations and informed PDN decisions.

IV. Conclusion

Our investigation of non-uniform, fine-grained power maps for BSPDN in 3D chiplet-based SiPs uncovers an often overlooked thermal penalty: when workloads are localized, BSPDN's constrained in-plane heat conduction can elevate hotspot temperatures beyond those in frontside-powered counterparts. This outcome contrasts sharply with the more benign predictions from uniform power assumptions. Our high-resolution thermal simulations enable the early identification of thermally critical regions, guiding effective PDN design by revealing subtle in-plane heat-spreading limitations. By eschewing coarse power assumptions and embracing fine-grained analysis, designers can more accurately gauge how localized workloads intensify hotspot formation. These insights underscore the urgency of co-optimizing power delivery and cooling strategies for nanosheet-based 2.5D/3D SiPs.

References

[1] E. Beyne, A. Jourdain, and G. Beyer, "Nano-through silicon vias (nTSV) for backside power delivery networks (BSPDN)," in *Symposium on VLSI Technology and Circuits (VLSI)*. IEEE, 2023, pp. 1–2.

[2] H. Oprins, J. L. R. Bohorquez, B. Vermeersch, G. Van der Plas, and E. Beyne, "Package level thermal analysis of backside power delivery network (BS-PDN) configurations," in *2022 21st IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (iTherm)*. IEEE, 2022, pp. 1–7.

[3] M. Lofrano, H. Oprins, V. Cherman, L. Witters, A. Jourdain, G. Van der Plas, and E. Beyne, "Block level and package level thermal assessment for back side power delivery network," in *74th Electronic Components and Technology Conference (ECTC)*. IEEE, 2024, pp. 1036–1043.

[4] B. Vermeersch, S. Mishra, M. Brunion, O. Zografos, M. Lofrano, H. Oprins, J. Myers, Z. Tokei, and G. Hellings, "Multiscale thermal impact of BSPDN: SoC hotspot challenges and partial mitigation," in *International Electron Devices Meeting (IEDM)*. IEEE, 2024, pp. 1–4.

[5] R. James, "The future of the high-performance semiconductor industry and design," in *2022 IEEE International Solid-State Circuits Conference (ISSCC)*, vol. 65. IEEE, 2022, pp. 32–35.

[6] Y. Chen, V. Sankatali, S. Mishra, J. Ryckaert, J. Myers, and D. Biswas, "Thermal implications in scaling high-performance server 3D chiplet-based 2.5D SoC from FinFET to Nanosheet," in *2024 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*. IEEE, 2024, pp. 45–50.

[7] N. Kumar, Y. Chen, V. Sankatali, S. Mishra, A. Gupta, P. Singh, F. Catthoor, J. Myers, J. Ryckaert, and D. Biswas, "Thermal insights into 3D packaging of a high performance server SoC in advanced nanosheet technology," in *2024 IEEE European Solid-State Electronics Research Conference (ESSERC)*. IEEE, 2024, pp. 337–340.

[8] R. Zhang, M. R. Stan, and K. Skadron, "Hotspot 6.0: Validation, acceleration and extension," in *University of Virginia, Tech. Rep*, 2015.