



Physics-informed neural networks for thermal spreading of multilayer multichip power modules

Yonghun Kim^a, Changhyeon Yoon^a, Haeun Lee^{b,c}, Seonu Bae^d, Nana Kang^{d,e},
Changwoo Han^e, Dongmin Shin^e, Jungwan Cho^f, Sooyoung Lee^{a,*}, Hyoungsoon Lee^{a,*}

^a School of Mechanical Engineering, Chung-Ang University, Seoul 06974, Republic of Korea

^b Department of Automotive Engineering, Yeungnam University, Gyeongsan 38541, Republic of Korea

^c Department of Mechanical Engineering, Stanford University, Stanford 94305, United States

^d Department of Intelligent Energy and Industry, Chung-Ang University, Seoul 06974, Republic of Korea

^e Research & Development Division, Hyundai Motor Group, Hwaseong 18280, Republic of Korea

^f School of Mechanical Engineering, Sungkyunkwan University, Suwon, Gyeonggi-do 16419, Republic of Korea

ARTICLE INFO

Keywords:

Physics-informed neural networks
Gaussian process surrogate modeling
Thermal spreading resistance
Sub-network architecture
Thermal design optimization

ABSTRACT

Efficient thermal management in multilayer, multi-source power modules remains a critical barrier to advancing high-performance electronic systems. Analytical models are limited to highly simplified geometries, while finite volume method (FVM) requires extensive meshing and prohibitive computational cost for iterative design exploration. Here, we present a physics-informed neural network (PINN) integrated with a Gaussian process (GP) surrogate to enable mesh-free, physically consistent thermal prediction and rapid design optimization. The PINN achieves prediction errors below 2.4% for single-source validation and below 2.0% for a realistic seven-layer, multi-source power module, while the complete PINN-GP workflow reduced the estimated design-exploration time from approximately 15,000 to 200 h compared with sequential FVM evaluation. Coupling the PINN with a GP surrogate enables efficient identification of optimal geometric parameters, achieving up to 4.1% reduction in thermal spreading resistance. This PINN-GP framework provides a scalable pathway for thermal analysis and design optimization in next-generation power modules.

1. Introduction

Thermal management has become a fundamental constraint in modern high-power electronic systems as device architectures transition toward chiplet-based and heterogeneous integration. In these architectures, multiple localized heat sources are densely integrated within multilayer substrates composed of dissimilar materials, producing complex internal conduction networks. As power density continues to rise, junction temperature is no longer governed primarily by the effectiveness of external cooling hardware but by the ability of heat to redistribute within the package itself. This shift marks a structural transition in thermal design, in which internal heat spreading increasingly dictates thermal performance, reliability, and lifetime [1].

Over the past decade, significant advances have been achieved in external thermal management technologies, including forced-air heat sinks [2], metal foams [3], micro pin-fin arrays [4,5], microchannels [6,7], manifold microchannels [8,9], jet impingement [10], and two-

phase cooling systems [11]. These approaches have substantially increased achievable heat transfer coefficients and reduced convective resistance. However, in multichip and multilayer power modules, further improvements in external cooling yield diminishing reductions in junction temperature. As convection becomes increasingly effective, the dominant thermal bottleneck shifts inward, where localized heat fluxes must spread through heterogeneous solid layers before reaching cooled boundaries.

This internal bottleneck is governed by thermal spreading resistance, which arises when heat flows from confined source regions into larger conduction areas. Spreading resistance depends sensitively on source geometry, substrate thickness, material heterogeneity, and boundary conditions. In chiplet-based and heterogeneous integration platforms, these effects are amplified by the close proximity of multiple heat sources and the presence of material interfaces, leading to strong thermal crosstalk and spatially nonuniform temperature fields [12,13]. Although a variety of modeling approaches have been developed, accurately characterizing and minimizing thermal spreading resistance

* Corresponding authors.

E-mail addresses: lees@cau.ac.kr (S. Lee), leeh@cau.ac.kr (H. Lee).

<https://doi.org/10.1016/j.icheatmasstransfer.2026.112739>

Nomenclature		Greek symbols	
A	area (mm^2)	α	length ratio of heat source and spreader, a/L
a	x-direction, rectangular heat source dimension	β	width ratio of heat source and spreader, b/W
b	y-direction, rectangular heat source dimension	Γ	boundary surface of the solution domain (–)
B	frequency matrix	λ	loss weight
Bi	Biot number (–)	σ	uncertainty parameters
DBC	direct-bonded-copper	θ	signal variance
FVM	Finite Volume Method	$\mathcal{L}$	loss
GP	Gaussian Process	$\mathcal{R}$	residual
h	heat-transfer coefficient ($\text{W}\cdot\text{m}^{-2}\cdot\text{K}^{-1}$)	τ	thickness ratio of spreader, t/L
$IGBT$	insulated-gate bipolar transistor	Ψ	dimensionless spreading resistance
k	thermal conductivity ($\text{W}\cdot\text{m}^{-1}\cdot\text{K}^{-1}$)	Ω	solution domain (–)
L	x-direction, rectangular spreader dimension	Subscripts	
L_c	characteristic length	avg	average
LHS	Latin hypercube sampling	amb	ambient
MLP	multilayer perceptron	b	baseplate
n	outward normal vector	bc	boundary condition
Nu	Nusselt number (–)	$cond$	conduction
P	Perimeter (mm)	$conv$	convection
PDE	partial differential equation	$cool$	cooling surface
$PINNs$	Physics-Informed Neural Networks	e	energy conservation
Q	heat rate (W)	eff	effective
q'	heat flux ($\text{W}\cdot\text{m}^{-2}$)	f	fluid
R	Thermal resistance ($\text{K}\cdot\text{W}^{-1}$)	$heat$	heated surface
R_{cond}	Thermal conductive resistance ($\text{K}\cdot\text{W}^{-1}$)	i	layer index
R_{conv}	Thermal convective resistance ($\text{K}\cdot\text{W}^{-1}$)	$i + 1$	adjacent layer index
R_{sp}	Thermal spreading resistance ($\text{K}\cdot\text{W}^{-1}$)	if	interface
S_x	x-direction spacing between heat sources	PDE	partial differential equation
S_y	y-direction spacing between heat sources	sp	spreading
T	temperature	tot	total
t	thickness		
W	y-direction, rectangular spreader dimension		

in realistic multichip architectures remains challenging. Analytical formulations provide physical insight but rely on simplified geometries and single-source assumptions [14–16]. Experimental studies offer direct measurements in systems such as LED packages and multichip arrays [17,18], yet their cost and fabrication effort preclude systematic exploration of high-dimensional design spaces. High-fidelity numerical simulations based on finite element or computational fluid dynamics methods can accommodate complex geometries and heterogeneous materials [19,20], but their computational expense renders iterative thermal design optimization impractical.

Recent advances in machine learning have introduced alternative strategies for accelerating thermal prediction. Data-driven neural networks have been applied to infer heat transfer behavior from high-dimensional or image-based representations [21–23], yet these models often generalize poorly outside the training domain and may violate fundamental physical constraints. Physics-informed neural networks address these limitations by embedding governing partial differential equations directly into the learning process [24]. PINNs have demonstrated promising performance in steady-state heat conduction problems and can eliminate the need for mesh generation [25,26]. Domain-decomposition PINNs, including cPINN and XPINN, have extended physics-informed learning to partitioned domains through coupled subnetworks and interface constraints, while multi-domain PINNs have also been applied to steady-state heat conduction in multilayer media [26–28]. Accordingly, the layer-specific subnetworks employed here follow the established domain-decomposition principle of coupled sub-domain networks and interface constraints; they serve as an enabling formulation for the present multilayer thermal-spreading and GP-based design-optimization workflow rather than as a new decomposition

algorithm. However, in practical power-module architectures, thermal spreading is simultaneously governed by material interfaces, interactions among multiple localized heat sources, package geometry, and external cooling strength. Studies that address spreading resistance as a design metric under these coupled factors in heterogeneous multilayer, multichip power modules remain limited. This gap motivates the direct integration of layer-resolved physics-informed prediction with thermal-spreading analysis and design optimization.

Here we develop an integrated framework that combines domain-partitioned physics-informed learning with Gaussian process surrogate modeling to predict and optimize thermal spreading in multilayer multichip power modules. The advance lies not in a new PINN or GP algorithm, but in their application-specific integration for thermal design in heterogeneous power-module architectures. Dedicated subnetworks resolve the temperature field within individual material layers while enforcing the governing heat-conduction equation, boundary conditions, interfacial temperature and heat-flux continuity, and global energy balance, establishing the methodological basis for heterogeneous multilayer prediction. The resulting three-dimensional temperature fields are coupled to GP-based optimization of chip spacing and baseplate thickness across different cooling conditions, linking physics-informed field prediction directly to thermal design. Applied from a canonical single-layer benchmark to a seven-layer four-chip automotive power module used in production vehicles, the framework further provides an application-specific means of resolving how heat-source interaction, internal heat spreading, and external cooling jointly shape the preferred package geometry.

2. Physics-informed framework for thermal spreading analysis

This study develops a physics-informed framework for predicting and optimizing thermal spreading behavior in multilayer, multichip power modules, where internal heat redistribution increasingly governs thermal performance under strong external cooling. The methodology is structured to resolve two coupled challenges inherent to such systems: accurately capturing thermal spreading across heterogeneous architectures with localized heat sources, and efficiently exploring high-dimensional design spaces without reliance on prohibitively expensive numerical simulations. To this end, a canonical single-layer configuration is first established to isolate the fundamental physics of thermal spreading and to provide a controlled basis for validating physical consistency and numerical fidelity. The framework is then extended to a realistic multilayer power-module architecture in which material discontinuities, interfacial coupling, and multi-source interactions dominate heat transport. Finally, a Gaussian process surrogate model is integrated with the physics-informed neural network to transform physically consistent thermal predictions into a data-efficient design-space exploration workflow.

2.1. Problem formulation and thermal domains

Thermal spreading in multichip power modules arises from the coupled effects of localized heat generation, heterogeneous material stacks, and boundary-driven heat removal. In realistic device architectures, these factors are strongly intertwined, obscuring the physical mechanisms that govern spreading resistance and limiting direct interpretation. To address this challenge, the problem formulation adopts a staged representation that separates the fundamental physics of thermal spreading from geometric and material complexity, and then systematically reintroduces these effects under realistic constraints (Fig. 1). This formulation enables the governing spreading mechanisms to be identified and validated in isolation before being examined in regimes where multi-source interaction and interfacial coupling dominate thermal behavior.

2.1.1. Single-layer canonical configuration

The single-layer configuration is designed to isolate thermal spreading behavior from confounding effects associated with material discontinuities and multi-source interactions. By considering steady-state heat conduction within a homogeneous, isotropic plate subjected to a localized heat source and convective cooling on the opposing surface, this domain captures the essential physics of lateral heat redistribution while suppressing secondary complexities. The single-layer geometry and associated boundary conditions are defined in Fig. 1a–c, providing a minimal yet complete representation of thermal spreading suitable for direct comparison with analytical solutions and high-fidelity numerical simulations.

The geometry follows the dimensional proportions of commercial Infineon IGBT modules (FF1800XTR17T2P5 and FF150R12W2T7E_B11), with a representative planform aspect ratio of 1:1.5. A uniform heat flux is applied over a central heat-source region on the bottom surface, while the top surface is subjected to convective cooling. All remaining boundaries, including the unheated portion of the bottom surface and the sidewalls, are treated as adiabatic. This boundary configuration suppresses parasitic heat losses and ensures that temperature gradients arise primarily from lateral spreading within the plate rather than from external heat extraction pathways.

To systematically characterize the influence of geometry and convection on spreading resistance, the problem is formulated using four nondimensional parameters: the length ratio ($\alpha = a/L$), width ratio ($\beta = b/W$), thickness ratio ($\tau = t/L$), and Nusselt number ($Nu = hL_c/k_f$). The ranges of these parameters are summarized in Table 1, and the reference case ($\alpha = 0.25$, $\beta = 0.25$, $\tau = 0.1$, $Nu = 40$) is adopted from G.N. Ellison [29]. Independent variation of each parameter establishes a physically transparent benchmark for evaluating how geometric scaling and boundary conditions govern thermal spreading resistance.

2.1.2. Multilayer power-module configuration

While the single-layer domain isolates the physics of thermal spreading, practical power modules operate in regimes where multiple localized heat sources, heterogeneous material stacks, and interfacial

Fig. 1. Canonical and realistic thermal domains used to isolate and analyze thermal spreading behavior. a–c, Single-layer geometry and boundary conditions defining the canonical domain used to isolate fundamental thermal spreading mechanisms. d–f, Multilayer power-module configuration comprising seven layers and four heat sources, representing a realistic regime in which interfacial coupling and multi-source interactions govern heat redistribution.

Table 1

Non-dimensional parameters and ranges for the single-layer case.

Parameters	Definition	Values
Length ratio (α)	a/L	0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5
Width ratio (β)	b/W	0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5
Thickness ratio (τ)	t/L	0.01, 0.025, 0.05, 0.075, 0.1, 0.25, 0.5, 1, 5, 10
Nusselt number (Nu)	hL_c/k_f	4, 20, 40, 200, 400, 1000, 2000, 4000, 10000, 20000

heat transfer collectively govern temperature distribution. To capture this regime, a multilayer configuration representative of an automotive inverter module is constructed. The multilayer module architecture and chip arrangement are defined in Fig. 1d–f, representing operating conditions under which further improvements in external cooling yield diminishing returns and internal heat redistribution becomes the dominant thermal bottleneck.

The module consists of seven stacked layers, including SiC chips, solder layers, a direct-bonded-copper (DBC) substrate composed of Cu/AlN/Cu, and an aluminum baseplate. Each layer introduces distinct thermal conductivities and thicknesses, resulting in strong interfacial coupling and anisotropic heat-flow paths. The material properties and layer dimensions are listed in Table 2. For SiC and AlN, thermal conductivity is modeled as temperature dependent; the functional forms and associated temperature range used in this work are specified in Section 2.3. Four SiC chips are symmetrically placed on the DBC surface, each dissipating a uniform heat flux consistent with high-power operation. Convective cooling is applied to the outer surface of the aluminum baseplate, while all other boundaries are treated as adiabatic, ensuring that heat removal occurs predominantly through the designated cooling interface.

To examine how internal geometry governs spreading resistance in this multilayer regime, three design parameters are selected: the chip-to-chip spacings in the x- and y-directions (S_x and S_y), and the baseplate thickness (t_b). The ranges of these parameters are summarized in Table 3. Together, they define a design space in which thermal interactions among heat sources, interlayer conduction, and boundary-driven heat removal are strongly coupled. This configuration therefore represents a stringent test case for evaluating the ability of the proposed

Table 2

Packaging materials, thermal properties, and layer thicknesses [30–33].

Component	Material	Density ($\text{kg}\cdot\text{m}^{-3}$)	Specific heat ($\text{J}\cdot\text{kg}^{-1}\cdot\text{K}^{-1}$)	Thermal conductivity ($\text{W}\cdot\text{m}^{-1}\cdot\text{K}^{-1}$)	Size (mm)
Chip	SiC	2330	710	$61,100/(T-115)$	$6.4 \times 5 \times 0.048$
Chip solder	SAC 305	7265	227	58.2	$6.4 \times 5 \times 0.08$
DBC copper	Cu	8940	390	393	$58.3 \times 40.5 \times 0.3$
DBC ceramic	AlN	3260	730	$285(T/300)^{1.25}$	$58.3 \times 40.5 \times 0.63$
Base plate solder	SAC 305	7265	227	58.2	$58.3 \times 40.5 \times 0.2$
Base plate	Al1060	2705	900	234.2	$58.3 \times 40.5 \times 1$

* T denotes the absolute temperature in K.

Table 3

Dimensional parameters and chip-spacing ranges for the multilayer case.

Parameters	Values	Unit
x-direction spacing (S_x)	5, 10, 15, 20	mm
y-direction spacing (S_y)	5, 10, 15, 20	mm
Base plate thickness (t_b)	0.5, 1, 1.5, 2	mm

physics-informed framework to predict and optimize thermal spreading resistance under realistic, multichip operating conditions.

2.2. Governing thermal quantities and operating regimes

Thermal behavior in conjugate heat transfer systems is governed by the interplay between internal heat conduction within solid structures and external heat removal through convection. As external cooling becomes increasingly effective, the contribution of convective resistance to junction temperature diminishes, and internal conduction and heat spreading within the solid domain emerge as the dominant thermal bottlenecks. To capture this transition explicitly, the operating conditions and governing thermal quantities are defined to separate convection-controlled effects from geometry-dependent internal heat redistribution.

The total thermal resistance between the heat source and the ambient is decomposed into convective, conductive, and spreading components,

$$R_{tot} = R_{conv} + R_{cond} + R_{sp}, \quad (1)$$

where R_{conv} represents resistance associated with external heat removal, R_{cond} accounts for through-thickness conduction, and R_{sp} characterizes the lateral redistribution of heat from localized sources into surrounding solid regions. Under steady-state operating conditions with fixed heat dissipation and boundary temperatures, thermal spreading resistance can be isolated by removing the contributions of external convection and one-dimensional conduction,

$$R_{sp} = \frac{T_j - T_{amb}}{Q_{tot}} - R_{conv} - R_{cond}, \quad (2)$$

where T_j denotes the junction temperature, T_{amb} the ambient reference temperature, and Q the total heat dissipation. Among the resistance components in Eq. (1), R_{sp} uniquely depends on geometric configuration and internal heat-flow pathways, making it the primary quantity of interest in the present study.

External cooling conditions are characterized using the Nusselt number,

$$Nu = \frac{hL_c}{k_f}, \quad (3)$$

which provides a nondimensional measure of convective heat transfer strength relative to internal conduction. Here, h is the convective heat transfer coefficient, and k_f the thermal conductivity of the fluid. The characteristic length is defined as $L_c = 4A_{cool}/P_{cool}$, where A_{cool} and P_{cool} denote the cooled surface area and its perimeter, respectively. The range of Nusselt numbers considered in this study spans operating conditions representative of contemporary electronic cooling technologies, including micro pin-fin heat sinks, manifold pin-fin microchannels, jet impingement cooling, and two-phase cooling systems [10,11,34–36]. This range enables systematic traversal from conduction-limited to convection-enhanced heat-transfer regimes.

For the single-layer configuration, the characteristic length L_c is uniquely defined by the plate geometry, enabling the use of nondimensional quantities to generalize spreading behavior across geometric scales and boundary conditions. In contrast, the multilayer power-module configuration involves multiple materials and thicknesses, each associated with distinct characteristic length scales and thermal

conductivities, rendering a single nondimensional formulation physically ambiguous. In this regime, heterogeneous conduction effects are incorporated through an effective thermal conductivity that collapses the multilayer stack into an equivalent conduction metric while preserving its net thermal response. The detailed nondimensional formulations and data reduction procedures are provided in Supplementary Note S1.

These operating conditions and governing thermal quantities establish a unified analytical framework that isolates geometry-dependent thermal spreading from external cooling effects. This framework provides the necessary foundation for physics-informed prediction and optimization of thermal spreading resistance across both canonical and realistic power-module configurations.

2.3. High-fidelity reference simulations

High-fidelity reference data were constructed to validate the predictive accuracy of the physics-informed neural network and to assess the physical consistency of the subsequent Gaussian-process-based optimization. The objective of these simulations is not to reproduce the full operational complexity of practical power modules, but rather to generate physically exact reference solutions that isolate thermal spreading behavior under well-controlled conditions. This reference dataset therefore serves as a numerical ground truth against which data-driven predictions can be rigorously evaluated.

Three-dimensional steady-state heat conduction analyses were performed using ANSYS Fluent 2024 R2. All FVM calculations were performed on a workstation equipped with an AMD Ryzen Threadripper 7970 × 32-core processor and 128 GB of system memory, with 32 CPU cores assigned to each Fluent calculation. The numerical domains reproduced the single-layer and multilayer configurations introduced in Section 2.1, with identical geometric dimensions, material properties, and boundary conditions. The governing equation corresponds to the steady-state heat diffusion equation for an isotropic solid without internal volumetric heat generation,

$$\nabla \cdot (k \nabla T) = 0, \quad (4)$$

which represents the minimal physical model capable of preserving thermal spreading behavior. More complex effects, such as transient heat transfer and radiation, were intentionally excluded, as they do not alter the steady-state spreading resistance that constitutes the primary quantity of interest in this study. Temperature dependence was neglected for materials whose thermal conductivity varied only marginally over the operating range, but it was retained for SiC and AlN because of their substantial conductivity variation with temperature.

A hexahedral-polyhedral hybrid mesh was employed: the baseplate, solder, copper, and chip layers were meshed with mapped hexahedral cells, whereas the AlN ceramic and the chip-side copper, whose in-plane footprints differ, were meshed with tetrahedral cells and converted to polyhedra to reduce the cell count. Local refinement was applied near material interfaces, and at least two control-volume layers were maintained across thin chip and solder layers to resolve through-thickness temperature gradients. The small through-thickness cell size required for these thin layers also constrained the in-plane cell dimensions to avoid excessively large cell aspect ratios, thereby contributing to the relatively high overall cell count. Mesh independence was verified for both the single-layer and multilayer configurations using a stringent criterion of less than 0.1% variation in spreading resistance because the FVM solutions served as high-fidelity references for PINN validation. This verification ensures that the extracted spreading resistance reflects a physical property of the system rather than a numerical artifact. Details of the mesh-independence study and grid specifications are provided in Supplementary Note S2, Fig. S1, and Tables S1, S2.

Boundary conditions were prescribed to match the operating scenarios defined in Section 2.2. A uniform heat flux was applied to the

bottom surfaces of the heat-source regions, while convective cooling was imposed on the designated cooling surfaces, representing coolant flow at 30 °C with a constant heat transfer coefficient. All remaining boundaries, including unheated bottom regions and lateral walls, were treated as adiabatic. Under these conditions, heat transport is governed by lateral redistribution within the solid domain coupled with external convection at the cooling interface. Thermal conductivity was treated as constant for all materials except SiC and AlN. For the remaining constituents, the conductivity varied by less than 3% over the operating window bounded by the coolant temperature of 30 °C and the peak SiC junction temperature of 175 °C, so temperature dependence was neglected. For SiC, the thermal conductivity was prescribed as a temperature-dependent function following Nilsson et al. [32],

$$k_{\text{SiC}}(T) = \frac{61100}{T - 115}, \quad (5)$$

with T in K. For AlN, the intrinsic conductivity data reported by Slack et al. [33] were fitted over the temperature range of interest, yielding

$$k_{\text{AlN}}(T) = 285 \left(\frac{T}{300} \right)^{-1.25}. \quad (6)$$

In contrast, variations exceeding 40% were reported for SiC and AlN over the same temperature range; therefore, their thermal conductivities were implemented as temperature-dependent functions. Because this temperature dependence was retained only for SiC and AlN, the governing problem is weakly nonlinear rather than strictly linear. Accordingly, representative heat dissipation values were selected consistently across all simulations to ensure numerically stable and physically meaningful comparison of thermal spreading resistance.

For the single-layer configuration, the per-chip power dissipation was fixed at $Q = 100$ W, corresponding to a heat flux of approximately $267 \text{ W} \cdot \text{cm}^{-2}$ for a footprint of $7.5 \times 5.0 \text{ mm}^2$. For the multilayer power-module configuration, each chip dissipated $Q = 283.5$ W with a footprint of $6.4 \times 5.0 \text{ mm}^2$, resulting in a heat flux of approximately $885.9 \text{ W} \cdot \text{cm}^{-2}$ applied to four chips. These values ensure stable numerical convergence while remaining representative of high-power operating conditions. All simulations were iterated until the residuals of the governing equations decreased below 10^{-10} . Numerical convergence was further confirmed by verifying that the global energy imbalance was less than 0.1%. These criteria ensure that the computed temperature fields and spreading-resistance values are fully converged steady-state solutions.

The resulting temperature distributions and thermal spreading resistances constitute the high-fidelity reference dataset used to validate the physics-informed neural network and to verify the physical plausibility of the GP-optimized designs. By deliberately adopting a minimal yet physically exact numerical model, this reference dataset provides a robust and interpretable benchmark for evaluating data-driven thermal predictions in both canonical and realistic power-module configurations.

2.4. Physics-informed neural network for thermal spreading

A physics-informed neural network (PINN) integrated with a Gaussian-process (GP) surrogate was developed to model thermal spreading behavior and enable data-efficient design optimization. The PINN was constructed to predict temperature distributions and thermal spreading resistance in both single-layer and multilayer systems by embedding the governing steady-state heat-conduction equation and boundary conditions directly into the training process. After validation against the high-fidelity reference data described in Section 2.3, the PINN predictions were used to construct a physically consistent GP surrogate, allowing efficient exploration of the design space and identification of optimal geometric and thermal parameters. All models were implemented in Python 3.10 using PyTorch 2.7.1 and trained on an

NVIDIA RTX 4090 GPU.

The network architecture was designed to solve the steady-state heat-conduction problem in both canonical single-layer domains and realistic multilayer power-module configurations. The overall architecture and domain-specific implementations are summarized in Fig. 2. For the single-layer configuration, a single multilayer perceptron was employed to facilitate direct validation against analytical solutions and high-fidelity numerical results. For the multilayer configuration, layer-specific subnetworks were adopted to represent the thermal behavior of each material layer individually, enabling explicit treatment of strong material discontinuities and interfacial coupling. Each network comprised six hidden layers with 256 nodes per layer, and all subnetworks employed smooth activation functions (SiLU) for stable training and accurate approximation of the smooth temperature fields governed by elliptic heat conduction [37]. Training was performed using the Adam optimizer [38] with a cosine-annealing learning-rate schedule [39], where the learning rate was reduced from 1×10^{-3} to 1×10^{-6} to improve convergence stability.

In the multilayer configuration, the predicted temperature field of each layer was sequentially coupled to adjacent layers, ensuring continuity of temperature and heat flux across all interfaces. This hierarchical learning structure allows interfacial thermal interactions to be incorporated explicitly while maintaining numerical stability and physical consistency, and it enables the PINN to capture layer-dependent thermal behavior and interfacial heat-transfer characteristics that cannot be reliably represented by a single monolithic network.

Physical consistency of the PINN predictions was enforced through a composite training objective that incorporates local constraints associated with the governing heat-diffusion equation, boundary conditions, and interfacial continuity, as well as a global energy-balance constraint that enforces conservation of energy across the computational domain. The residuals of the individual constraints are defined as

$$\mathcal{R}_{\text{PDE}}(\mathbf{x}) = \nabla \cdot (\mathbf{k}_i \nabla T_i), \mathbf{x} \in \Omega_i \quad (7)$$

$$\mathcal{R}_{\text{bc}}(\mathbf{x}) = \begin{cases} -k \frac{\partial T}{\partial n} + q'', \mathbf{x} \in \Gamma_q \\ -k \frac{\partial T}{\partial n} - h(T - T_\infty), \mathbf{x} \in \Gamma_h \\ \frac{\partial T}{\partial n}, \mathbf{x} \in \Gamma_a \end{cases} \quad (8)$$

$$\mathcal{R}_{\text{if},T}(\mathbf{x}) = T_i - T_{i+1}, \mathbf{x} \in \Gamma_{i,i+1} \quad (9)$$

$$\mathcal{R}_{\text{if},q}(\mathbf{x}) = k_i \frac{\partial T_i}{\partial n} - k_{i+1} \frac{\partial T_{i+1}}{\partial n}, \mathbf{x} \in \Gamma_{i,i+1} \quad (10)$$

$$\mathcal{R}_e = Q_{\text{in}} - \int_{\Gamma_h} h(T - T_\infty) dA \quad (11)$$

where Ω_i denotes the i -th material layer; Γ_q , Γ_h , and Γ_a denote the heated, convectively cooled, and adiabatic boundaries, respectively; $\Gamma_{i,i+1}$ denotes the interface between adjacent layers i and $i+1$; n is the outward unit normal; q'' is the imposed chip heat flux; h and T_∞ are the heat transfer coefficient and coolant temperature; and Q_{in} is the total heat dissipated by the chips. The interface residual is thus split into two separate conditions, temperature continuity in Eq. (9) and normal heat-flux continuity in Eq. (10), so that both are enforced independently. Each residual is evaluated at N_j collocation points sampled in the corresponding region, and the associated loss term is defined as the mean squared residual over those points,

$$\mathcal{L}_j = \frac{1}{N_j} \sum_{i=1}^{N_j} |\mathcal{R}_j(\mathbf{x}_i)|^2, j \in \{\text{PDE, bc, if, } T, \text{ if, } q, e\}, \mathcal{L}_{\text{if}} = \mathcal{L}_{\text{if},T} + \mathcal{L}_{\text{if},q} \quad (12)$$

The two interface contributions are combined into a single interface loss, so that the total training objective consists of four loss groups; their weighted combination is given in Supplementary Note S3.

To avoid manual tuning of competing physical constraints, an uncertainty-based adaptive weighting strategy was employed [40]. In

Fig. 2. Domain-partitioned PINN-GP framework for multilayer thermal prediction and design optimization. Layer-specific subnetworks map the spatial coordinates (x, y, z) to the temperature field in each material layer, with automatic differentiation providing the spatial derivatives required to enforce the governing heat-conduction equation, boundary conditions, interfacial temperature and heat-flux continuity, and global energy balance. The trained PINN is subsequently used to generate thermal responses for GP surrogate modeling, which maps the design and cooling parameters (S_x, S_y, t_b, h) to the spreading resistance R_{sp} and enables rapid exploration of the design space and identification of the optimum.

addition, feature encoding was used to mitigate spectral bias in neural network training and to improve resolution of steep temperature gradients near heat sources and material interfaces [41,42]. Further details of the adaptive loss weighting, feature mapping, and training strategy are provided in Supplementary Note S3. The convergence histories of the total loss and the individual physics-informed loss components are additionally presented in Supplementary Fig. S2 to demonstrate the training behavior of the proposed PINN implementation.

By construction, the proposed PINN functions as a physics-informed approximation of the governing heat-conduction problem. Rather than acting as a purely data-driven interpolator, the network reproduces the governing physical behavior of steady-state thermal spreading under the same constraints imposed on the reference numerical model. A conventional scalar-output MLP can predict thermal metrics such as R_{sp} or junction temperature but does not reconstruct the full three-dimensional temperature field. Although a data-driven network can be formulated for field prediction, it requires labeled temperature-field data and does not explicitly enforce temperature and heat-flux continuity across material interfaces; in contrast, the present PINN directly imposes these constraints, which is particularly important for multilayer systems with discontinuous thermal conductivities [43]. A supplementary comparison with a conventional supervised MLP trained for the same geometry and boundary conditions showed cross-interface heat-flow mismatches of 54.0–281.3% for the MLP but only 0.124–0.343% for the PINN (Supplementary Note S4 and Table S3). This property enables seamless integration with the GP-based optimization framework introduced in the following section, where repeated evaluations of thermal spreading resistance are required across a high-dimensional design space.

2.5. Gaussian-process-based surrogate for design optimization

A Gaussian process (GP) regression model was developed to leverage the physics-informed dataset generated by the PINN for efficient prediction and optimization of multilayer thermal designs. The role of the GP surrogate is not to replace the physics-informed model, but to enable rapid and uncertainty-aware exploration of the design space once a physically consistent dataset has been established. By placing a probabilistic prior over functions, the GP provides joint predictions characterized by both predictive means and calibrated uncertainties through its mean and covariance functions [44].

The PINN-predicted spreading resistance values were treated as low-noise observations for GP training. Under this assumption, the GP posterior yields predictive estimates accompanied by credible intervals in unobserved regions of the design space, enabling explicit quantification of epistemic uncertainty associated with model generalization. Owing to its nonparametric Bayesian formulation, the GP does not require additional neural-network retraining and remains reliable in low-data regimes, which is particularly advantageous for high-dimensional thermal design problems where generating new training data is computationally expensive.

To construct the dataset for GP surrogate modeling, a total of 100 DOE cases were selected to cover the parameter space defined by S_x , S_y , t_b , and h . A full-factorial evaluation of the four levels of each geometric variable and the three convection conditions would require 192 cases. To reduce the data-generation cost while retaining physically interpretable coverage, the DOE included the representative baseline condition ($S_x = S_y = 15$ mm, $t_b = 1$ mm, and $h = 25,000$ W·m⁻²·K⁻¹), one-factor variations, and boundary cases, while the remaining design space was sampled using Latin hypercube sampling (LHS). The sample size was selected to exceed the commonly used heuristic of approximately 10 samples per input dimension for Gaussian process emulators [45], and its adequacy was subsequently assessed through the learning-curve analysis presented in Supplementary Note S5 and Fig. S3. For each DOE point, the validated PINN was used to obtain the corresponding thermal response and spreading resistance R_{sp} . The resulting 100 samples were partitioned into 75 training, 15 validation, and 10

independent test samples, with 75 samples used to train the GP model. All input and output variables were standardized using Z-score normalization to zero mean and unit variance prior to training, and GP predictions were subsequently transformed back to physical units for interpretation. A radial basis function (RBF) kernel with automatic relevance determination (ARD) was employed to model smooth thermal responses while capturing dimension-specific sensitivities. The kernel function is expressed as

$$\mathcal{K}(\mathbf{x}, \mathbf{x}') = \theta^2 \exp \left(- \sum_{d=1}^D \frac{(\mathbf{x}_d - \mathbf{x}'_d)^2}{2\ell_d^2} \right), \quad (13)$$

where $\mathcal{K}$ denotes the covariance function, θ^2 is the signal variance, and ℓ_d represents the characteristic length scale for the d -th design dimension. The kernel hyperparameters were determined by maximizing the log marginal likelihood using maximum likelihood estimation.

Within the 75-sample training set, five-fold cross-validation [46] was performed to evaluate the stability and predictive performance of the GP surrogate, yielding a coefficient of determination of $R^2 = 0.98$ across all folds. The validation and independent test subsets were retained separately to evaluate the predictive performance for design conditions not used during GP training. Together, the space-filling DOE, cross-validation, and hold-out evaluation were used to assess the adequacy of the adopted dataset for surrogate construction.

Once trained, the GP surrogate accepts a candidate set of design variables and returns the predicted spreading resistance together with its associated epistemic uncertainty. This capability enables efficient exploration of the high-dimensional design space without requiring additional high-fidelity simulations or repeated PINN evaluations. To ensure broad sampling coverage, a total of 30,000 design candidates were generated using Latin hypercube sampling, corresponding to 10,000 candidates for each of the three convective conditions ($h = 250, 25000, 100000$ W·m⁻²·K⁻¹). Each candidate was then evaluated using the GP surrogate to obtain its predictive mean and uncertainty. The optimization objective was to minimize the thermal spreading resistance in the multilayer structure, and the GP predictions were analyzed to determine the optimal design.

3. Results and discussion

The results presented in this section reveal how thermal spreading behavior evolves across operating regimes spanning conduction-limited to convection-enhanced heat transfer. Through systematic analysis of canonical single-layer configurations and realistic multilayer power modules, we show that increasing external cooling strength progressively diminishes the influence of convective resistance while amplifying the role of geometry-dependent internal heat spreading. Under sufficiently strong convection, thermal spreading resistance emerges as the dominant bottleneck governing junction temperature, rendering further improvements in external cooling increasingly ineffective. Building on this physical understanding, the latter part of this section demonstrates how physics-informed surrogate models can be leveraged to efficiently explore the design space and identify geometries that minimize thermal spreading resistance.

3.1. Nu-dependent thermal spreading behavior in a canonical single-layer system

We first examine a canonical single-layer configuration that isolates internal heat redistribution from multilayer coupling effects. This simplified geometry provides a controlled framework for identifying how thermal spreading resistance evolves across operating regimes spanning conduction-limited to convection-enhanced heat transfer. Particular attention is given to the role of the Nusselt number, which parametrizes the relative strength of external convection and serves as a key control variable for distinguishing cooling-dominated and

spreading-dominated thermal behavior.

3.1.1. Baseline thermal spreading behavior and cross-model benchmarking

In the single-layer configuration, thermal spreading behavior exhibits a systematic dependence on both geometric parameters and external cooling conditions. To quantify this behavior in a physically consistent manner, thermal spreading resistance is expressed using the dimensionless parameter Ψ_{sp} , which captures the contribution of lateral heat spreading normalized by the corresponding one-dimensional conduction resistance. The analysis considers four governing dimensionless parameters, namely the length ratio α , width ratio β , thickness ratio τ , and the Nusselt number Nu . Fig. 3 presents a comparative assessment of Ψ_{sp} obtained from high-fidelity FVM simulations, physics-informed neural network predictions, and analytical reference models over a broad parameter space. The parameter sweep is anchored to the reference configuration reported by G.N. Ellison [29] with $\alpha = 0.25$, $\beta = 0.25$, $\tau = 0.1$, and $Nu = 40$, from which each parameter is varied independently to isolate its influence on baseline spreading behavior. For analytical reference, classical formulations developed by Ellison [29] and by Yovanovich [14] are included as established benchmarks for thermal spreading resistance under idealized boundary conditions. These models provide closed-form solutions that have long served as theoretical limits for spreading behavior in simplified geometries. Detailed governing equations and assumptions for both analytical models are provided in Supplementary Note S6, allowing the present results to be interpreted relative to both numerical and idealized analytical references.

Across the full parameter sweep, the physics-informed neural network closely reproduces the FVM results, with a maximum deviation of 2.4% and an average deviation of 0.56%. In contrast, the analytical predictions exhibit larger discrepancies of 4.28% for the Ellison model and 5.97% for the Yovanovich model. These deviations originate primarily from truncation of the eigenfunction series near the isoflux boundary, which can bias the area-averaged source temperature in finite heat-source geometries. Because the PINN enforces the Neumann boundary condition pointwise across the heat-source region, it remains less sensitive to this truncation effect.

Figs. 3a and 3b illustrate the dependence of Ψ_{sp} on heat-source expansion along the in-plane length and width directions. In both cases, the response is non-monotonic. Ψ_{sp} initially increases as localized temperature gradients intensify near the heat source, and subsequently decreases as the effective lateral conduction area expands. The reduction is more pronounced with increasing α , reflecting the longer in-plane distance heat must traverse before reaching the cooled boundary and the resulting enhancement of lateral redistribution. By contrast, increasing β produces a weaker reduction because the shorter conduction path enables rapid heat extraction, limiting further in-plane spreading.

The combined effects of plate thickness and external convection are shown in Figs. 3c and 3d. At fixed Nu , increasing τ initially elevates Ψ_{sp} due to the longer through-thickness conduction path near the heat source, and subsequently reduces it once lateral diffusion within the plate becomes more effective. Increasing Nu consistently lowers Ψ_{sp} by strengthening heat removal at the cooling surface. A clear crossover emerges near $Nu \approx 500$, marking the transition between conduction-

Fig. 3. Variation of thermal spreading resistance with geometric and thermal parameters in the single-layer configuration. a, Variation with length ratio α , b, Variation with width ratio β , c, Variation with thickness ratio τ , and d, Variation with Nusselt number Nu . The PINN closely matches FVM across all trends, with τ and Nu showing the strongest influence and revealing a conduction-to-convection transition near $Nu \approx 500$.

dominated and convection-enhanced regimes. Below this threshold, thicker plates promote more uniform heat redistribution within the solid, whereas under strong convection thinner plates become favorable because heat is extracted before extensive in-plane spreading develops.

This baseline establishes how thermal spreading resistance depends on geometry and operating regime in the canonical single-layer configuration and provides the reference required to interpret the regime transitions induced by increasing external cooling strength in the following subsection.

3.1.2. Coupled influence of thickness and convection strength

The coupled effects of plate thickness and external convection strength on thermal spreading behavior are examined using the reference geometry with $\alpha = \beta = 0.25$. The thickness ratio τ and the Nusselt number Nu are varied independently to determine how internal heat redistribution responds to simultaneous changes in geometry and cooling intensity. The analysis spans convection strengths representative of cooling technologies ranging from forced-air convection to liquid and two-phase microchannel systems [11,36,47], corresponding to heat-transfer coefficients between 250 and 100,000 $\text{W}\cdot\text{m}^{-2}\cdot\text{K}^{-1}$.

Fig. 4 reports the dependence of the dimensionless spreading resistance Ψ_{sp} on plate thickness across multiple convection regimes, allowing direct comparison of thickness sensitivity as external cooling is progressively strengthened. At low Nu , weak convection limits heat removal at the cooling surface and thermal energy accumulates within the spreading layer. In this regime, increasing τ initially raises Ψ_{sp} because the longer through-thickness conduction path delays lateral heat redistribution. As τ increases further, enhanced lateral diffusion within the plate becomes dominant and Ψ_{sp} decreases as heat spreads more uniformly across the substrate.

As Nu increases, convective heat removal becomes more effective and the influence of plate thickness diminishes. Near $Nu \approx 500$, internal conduction and external convection contribute at comparable levels under the present operating conditions. In this intermediate regime, variations in τ produce only modest changes in Ψ_{sp} , indicating a transition in the dominant thermal resistance mechanism.

Under strong convection at high Nu , thinner plates consistently yield lower Ψ_{sp} values. Heat is extracted rapidly through the cooling surface before substantial lateral spreading can develop, and additional thickness introduces redundant conduction pathways that slightly increase spreading resistance. Although Ψ_{sp} continues to exhibit weak dependence on τ , the effect of thickness becomes secondary once external convection dominates.

3.1.3. Temperature field validation

The ability to resolve the full temperature field is essential for assessing thermal performance in configurations where localized hot-spots and steep gradients govern reliability. Unlike analytical formulations, the PINN solves the governing heat-conduction equation directly at collocation points distributed throughout the spatial domain, enabling reconstruction of the temperature field at a fidelity comparable to FVM.

The temperature field is examined for the reference configuration with $\alpha = 0.25$, $\beta = 0.25$, $\tau = 0.1$, and $Nu = 40$ using four representative planes passing through the heat-source centroid. These planes include the heated bottom surface, the top cooling surface, and two orthogonal vertical cross-sections aligned with the x and y directions. Fig. 5 presents the corresponding temperature distributions predicted by the PINN alongside the FVM reference.

On the heated bottom surface, shown in Fig. 5a, sharp temperature gradients develop in the vicinity of the heat source. In this region, the PINN prediction exhibits a local deviation of approximately 2.4% relative to FVM, while the area-averaged error remains below 1%. These localized discrepancies coincide with regions of strongest gradient, where accurate enforcement of boundary flux conditions is most critical. On the top cooling surface, shown in Fig. 5b, the temperature field remains nearly uniform, reflecting efficient heat removal and weak lateral gradients.

The vertical cross-sections shown in Figs. 5c and 5d further confirm that the PINN accurately reproduces temperature variations throughout the thickness and lateral extent of the domain. Both gradual temperature transitions and steep gradients near the heat-source edges are captured

Fig. 4. Interplay between plate thickness and convection strength in single-layer thermal spreading. Ψ_{sp} as a function of τ and Nu under the reference geometry ($\alpha = \beta = 0.25$). Stronger convection diminishes the influence of τ , while weak convection favors thicker plates for enhanced lateral diffusion, revealing the transition between conduction- and convection-dominated regimes.

Fig. 5. Comparison of temperature fields predicted by the PINN and FVM in the single-layer configuration. a, Heated bottom surface, b, Cooling surface, c, Cross-section along x , and d, Cross-section along y for the reference geometry with $\alpha = 0.25$, $\beta = 0.25$, $\tau = 0.1$, and $Nu = 40$. The PINN closely matches the FVM temperature fields, capturing spatial gradients and hot-spot distributions with average errors below 1%.

without numerical smoothing. This spatial consistency indicates that the embedded physical constraints preserve energy balance across the domain rather than relying on implicit numerical diffusion.

The close agreement between the predicted and simulated temperature fields demonstrates that the PINN retains physical fidelity beyond scalar metrics such as spreading resistance. By resolving detailed spatial temperature distributions with accuracy comparable to FVM, the model provides a reliable foundation for subsequent surrogate modeling and optimization in more complex multilayer configurations, where internal hotspots and interlayer heat-flow pathways play a dominant role.

3.2. Multilayer thermal spreading in a realistic power-module configuration

The multilayer configuration containing four SiC chips is analyzed to examine how geometric parameters and external convection jointly govern thermal spreading under realistic power-module conditions. Unlike the canonical single-layer system, this configuration introduces strong interlayer coupling, heterogeneous material stacks, and multi-source thermal interaction, which collectively reshape internal heat-

flow pathways. A convective heat-transfer coefficient of $25,000 \text{ W m}^{-2} \text{ K}^{-1}$ and a 1 mm aluminum baseplate are adopted as representative operating conditions, consistent with manifold pin-fin microchannel cooling reported in prior studies [36]. The symmetric four-chip layout with 15 mm spacing provides a realistic model to quantify how conduction and convection jointly govern thermal resistance in multilayer systems.

3.2.1. Influence of chip-to-chip spacing

Fig. 6 shows how the dimensionless spreading resistance Ψ_{sp} changes with chip spacing along the long-axis direction S_x and short-axis direction S_y . Increasing S_x produces a monotonic reduction in Ψ_{sp} as thermal coupling between neighboring heat sources weakens and the effective lateral diffusion region expands. As the distance between chips increases, localized temperature accumulation is alleviated and heat spreads more uniformly through the multilayer stack.

In contrast, variation of S_y yields a non-monotonic response. At small S_y , strong thermal interference between adjacent chips elevates spreading resistance. As S_y increases, Ψ_{sp} decreases as source interaction weakens and lateral diffusion becomes more effective. Beyond an

Fig. 6. Effect of chip-to-chip spacing on thermal spreading resistance in the multilayer configuration. a, Variation along the long-axis spacing S_x , b, Variation along the short-axis spacing S_y . Wider spacing along S_x lowers the spreading resistance, while spacing along S_y produces a non-monotonic response arising from competing effects of source interaction and boundary proximity.

intermediate spacing, further increases in S_y shift the heat sources closer to the substrate boundary, restricting the available diffusion area and causing Ψ_{sp} to rise again. This competition between inter-source interaction and boundary confinement governs the optimal spacing regime in the transverse direction. Additional discussion of boundary-limited spacing behavior is provided in Supplementary Fig. S4.

3.2.2. Coupled influence of base-plate thickness and convection strength

The coupled effects of base-plate thickness and convection strength are examined with chip spacing fixed at $S_x = S_y = 15$ mm. Fig. 7 summarizes how Ψ_{sp} varies with base-plate thickness across convection regimes relevant to high-performance electronic cooling. At low Nu , heat removal through convection is weak and temperature build-up occurs near the spreading layer. Increasing τ in this regime enhances lateral diffusion, which reduces Ψ_{sp} and promotes more uniform temperature distribution. As Nu increases, convection becomes stronger, and the influence of thickness gradually diminishes since heat is removed through the cooling surface before substantial in-plane diffusion develops. Around $Nu \approx 900$, the dependence on t_b nearly disappears as

conduction and convection contribute comparably. Beyond this crossover, further increasing t_b leads to a slight rise in Ψ_{sp} because excessive thickness adds conduction paths that no longer assist heat spreading once convection dominates.

The results demonstrate that plate thickness and convection intensity jointly define the transition between conduction- and convection-dominated regimes. Under weak convection, a thicker plate enhances in-plane diffusion and stabilizes temperature uniformity, while under strong convection, minimizing thickness facilitates faster heat extraction and reduces unnecessary conduction paths. This coupling establishes a clear design principle: the optimal plate geometry should scale with the effective cooling capability and applied heat flux, ensuring balanced thermal spreading and minimal resistance across operating conditions.

3.2.3. Temperature-field consistency in multilayer systems

Fig. 8 shows a comparison between the temperature fields predicted by the PINN and those obtained from FVM simulations for the reference case, where S_x and S_y were 15 mm, the base-plate thickness was 1 mm, and the convective heat-transfer coefficient was $25,000 \text{ W} \cdot \text{m}^{-2} \cdot \text{K}^{-1}$. The

Fig. 7. Interplay between base-plate thickness and convection strength in the multilayer configuration ($S_x = S_y = 15$ mm). a, Variation with plate thickness t_b at fixed convection conditions. b, Variation with convection intensity for different t_b values. A crossover near $Nu \approx 900$ indicates the shift from conduction- to convection-dominated behavior that determines the thickness most effective for heat spreading.

Fig. 8. Comparison of temperature fields predicted by the PINN and FVM in the multilayer configuration. a, Heated surface. b, Cooling surface. c, Cross-section along x. d, Cross-section along y through the heat-source centroids for the reference geometry with $S_x = S_y = 15$ mm, $t_b = 1$ mm, and $h = 25,000$ W·m⁻²·K⁻¹. The PINN closely matches the FVM temperature fields with average errors below 1%, accurately capturing three-dimensional gradients across the multilayer stack.

comparison included four planes through the heat-source centroids, corresponding to the heated surface, the cooling surface, and the vertical cross-sections normal to the x- and y-axes.

The overall temperature distribution predicted by the PINN closely matched the FVM results, with an average error below 1%. On the heated surface, shown in Fig. 8a, steep temperature gradients near the chip boundaries led to local deviations of about 2%, while the central region maintained excellent agreement. These localized errors arise because temperature gradients are strongest near the source edges, where rapid spatial variation makes numerical convergence more sensitive. On the cooling surface, shown in Fig. 8b, both the temperature level and the lateral distribution were reproduced accurately, indicating that the convective boundary condition was properly captured. The vertical cross-sections in Figs. 8c and 8d further confirm that both lateral and vertical heat transport were accurately resolved across multiple layers, including the material interfaces with distinct thermal conductivities.

The strong agreement between the PINN and FVM across all planes demonstrates that the PINN can reconstruct three-dimensional temperature fields in multilayer and multi-source configurations with high fidelity. The model successfully resolves both smooth and steep gradients without mesh-based discretization, validating its reliability for thermal

analysis within the multilayer power-module configuration and parameter domain examined in this study.

3.3. Physics-informed surrogate-guided optimization of multilayer thermal designs

The optimization framework coupled the PINN with a GP surrogate model to identify optimal thermal designs for multilayer, multi-chip structures under varying convection strengths. Three representative cooling conditions were examined with heat-transfer coefficients of 250, 25000, and 100,000 W·m⁻²·K⁻¹, corresponding to low, medium, and high convection regimes. For each case, ten thousand design candidates were generated by varying chip spacing (S_x, S_y) and plate thickness t_b , and the surrogate model was trained on the PINN dataset to locate the combination minimizing the spreading resistance. The resulting optimal points are summarized in Fig. 9 and Table 4. To clarify the physical origin of these optima, the following discussion is organized around two coupled mechanisms: the in-plane chip spacings govern inter-source thermal coupling and boundary confinement, whereas the base-plate thickness governs the balance between lateral heat spreading and through-thickness heat extraction.

At the lowest convection level (250 W·m⁻²·K⁻¹), shown in Fig. 9a,

Fig. 9. PINN-GP-guided optimization of multilayer multichip thermal designs under different cooling conditions. a, Low convection ($h = 250 \text{ W}\cdot\text{m}^{-2}\cdot\text{K}^{-1}$). b, Medium convection ($h = 25,000 \text{ W}\cdot\text{m}^{-2}\cdot\text{K}^{-1}$). c, High convection ($h = 100,000 \text{ W}\cdot\text{m}^{-2}\cdot\text{K}^{-1}$). The optimal designs marked in red reveal a clear shift from conduction- to convection-dominated behavior as cooling strength increases.

Table 4
Optimal designs predicted by the PINN-GP framework.

	HTC ($\text{W}\cdot\text{m}^{-2}\cdot\text{K}^{-1}$)	S_x (mm)	S_y (mm)	Thickness (mm)	Max T_j ($^{\circ}\text{C}$)	Avg T_j ($^{\circ}\text{C}$)	Max T_{cool} ($^{\circ}\text{C}$)	Avg T_{cool} ($^{\circ}\text{C}$)	R_{tot} ($\text{K}\cdot\text{W}^{-1}$)	R_{sp} ($\text{K}\cdot\text{W}^{-1}$)	Ψ_{sp}
Initial optimum	250	20	20	2	2208.8	2164.5	2008.2	1951.1	1.8823	0.1681	0.1922
PINN-GP		19.73	14.48	1.98	2195.4	2151.8	1994.2	1951.1	1.8710	0.1569	0.1794
FVM val		19.73	14.48	1.98	2200.7	2156.6	2005.9	1951.1	1.8753	0.1611	0.1843
Error (%)					0.24%	0.22%	0.28%	–	0.22%	2.65%	2.65%
Enhancement					8.05 $^{\circ}\text{C}$	7.87 $^{\circ}\text{C}$	8.54 $^{\circ}\text{C}$	–	0.37%	4.13%	4.13%
Initial optimum		20	15	1	165.1	145.1	87.2	49.2	0.1014	0.0644	0.0737
PINN-GP	25,000	19.73	13.04	1.92	161.7	142.7	82.8	49.2	0.0994	0.0624	0.0713
FVM val		19.73	13.04	1.92	161.7	142.7	82.1	49.2	0.0994	0.0624	0.0713
Error (%)					0.04%	0.00%	0.83%	–	0.01%	0.01%	0.01%
Enhancement					3.49 $^{\circ}\text{C}$	2.38 $^{\circ}\text{C}$	5.08 $^{\circ}\text{C}$	–	2.07%	3.25%	3.25%
Initial optimum		20	20	0.5	131.6	114.2	65.5	34.8	0.0743	0.0499	0.0571
PINN-GP		19.16	12.08	0.52	131.2	113.4	65.0	34.8	0.0735	0.0492	0.0563
FVM val	100,000	19.16	12.08	0.52	131.3	113.8	65.1	34.8	0.0739	0.0495	0.0567
Error (%)					0.10%	0.38%	0.08%	–	0.52%	0.77%	0.77%
Enhancement					0.32 $^{\circ}\text{C}$	0.39 $^{\circ}\text{C}$	0.47 $^{\circ}\text{C}$	–	0.46%	0.69%	0.69%

conduction and lateral diffusion dominate heat transfer. The optimal design emerges near $t_b \approx 2 \text{ mm}$, where a larger diffusion volume reduces Ψ_{sp} . Increasing S_x monotonically decreases Ψ_{sp} as thermal coupling between neighboring sources weakens, whereas S_y exhibits a weaker and non-monotonic influence. When S_x is small, lateral conduction fields from neighboring chips overlap significantly, effectively behaving as a combined heat source, and heat primarily spreads along y , making S_y variation negligible. As S_y widens, lateral diffusion improves until the chips approach the plate boundary, where confinement causes a modest rise in resistance. The balance among inter-source thermal coupling, boundary confinement, and available diffusion area defines the low-convection optimum.

At medium convection ($25,000 \text{ W}\cdot\text{m}^{-2}\cdot\text{K}^{-1}$), conduction and convection contribute comparably (Fig. 9b). The optimal point still appears near the upper thickness limit, yet the improvement range narrows as convection accelerates heat removal, with the absolute Ψ_{sp} reduction decreasing from approximately 0.09 at $h = 250 \text{ W}\cdot\text{m}^{-2}\cdot\text{K}^{-1}$ to about 0.02. Stronger cooling suppresses inter-source thermal coupling so that Ψ_{sp} decreases more uniformly with S_y , and independent heat-source behavior replaces collective diffusion. Nonetheless, when chip spacing approaches the plate edge, a mild resistance increase persists, showing that boundary confinement remains relevant even under forced-convection conditions.

At high convection ($100,000 \text{ W}\cdot\text{m}^{-2}\cdot\text{K}^{-1}$), convective extraction dominates (Fig. 9c). Ψ_{sp} becomes nearly insensitive to geometry, changing by less than 7% across the design space, and the absolute Ψ_{sp}

reduction further contracts to approximately 0.005. The optimal design shifts toward the thinnest configuration, confirming that high-intensity cooling favors shorter through-thickness heat-transfer paths and that geometric tuning provides limited additional benefit once convection fully controls heat transport. The specific geometric arrangements of the heat sources and the corresponding temperature contours, which shift according to the inter-chip spacing, are further detailed in Supplementary Fig. S5.

The GP surrogate, built upon the PINN and validated against FVM benchmarks (maximum deviation $< 3\%$), enabled the global optimization of 30,000 design points. Relative to the pre-optimization reference, the optimized designs reduced the spreading resistance R_{sp} by $\approx 4.1\%$ in the low h case and $\approx 3.3\%$ in the medium h case, which lowered the maximum junction temperature by $\approx 8.1^{\circ}\text{C}$ and $\approx 3.5^{\circ}\text{C}$, respectively, and the maximum cooling-surface temperature by $\approx 8.5^{\circ}\text{C}$ and $\approx 5.1^{\circ}\text{C}$, respectively. As h was fixed within each case, these decreases in R_{sp} directly translate into a reduced total thermal resistance R_{tot} . The engineering significance of these improvements should be considered within the constrained design space of a mature production power module. In the medium-convection case, the maximum junction temperature decreased from 165.1 to 161.7 $^{\circ}\text{C}$, increasing the thermal margin relative to the 175 $^{\circ}\text{C}$ criterion from 9.9 to 13.3 $^{\circ}\text{C}$, which corresponds to an approximately 34% increase. Based on the reported activation energy of approximately 1.1 eV for threshold-voltage instability in SiC MOSFETs [48], an Arrhenius-based estimate further suggests that the approximately 3.5 $^{\circ}\text{C}$ temperature reduction could increase the characteristic

degradation timescale for this mechanism by approximately 25%.

Computational cost was assessed on a wall-clock basis for the complete design-space exploration workflow. A representative multilayer FVM calculation, performed using the numerical settings and convergence criteria described in Section 2.3, required approximately 30 min on the AMD Ryzen Threadripper 7970× workstation. Sequential FVM evaluation of the 30,000 design candidates considered in this study was therefore estimated to require approximately 15,000 h. This value represents an extrapolated estimate and does not correspond to the measured duration of an actual 30,000-case FVM sweep. In comparison, the complete offline construction of the PINN-GP framework, including PINN training, surrogate-data generation, and GP training, required approximately 200 h. Once this offline stage was completed, the trained GP surrogate evaluated all 30,000 design candidates in approximately 10 min. The complete PINN-GP workflow therefore required approximately 200 h, corresponding to an estimated 75-fold reduction in practical wall-clock cost relative to a sequential FVM-only evaluation using the high-fidelity reference mesh adopted in this study. Because the FVM and PINN calculations were performed using CPU- and GPU-based computational platforms, respectively, this comparison represents the practical wall-clock costs of the workflows implemented in this study rather than a hardware-normalized comparison of intrinsic solver efficiency.

4. Conclusion

The transition toward chiplet-based and heterogeneous integration architectures has shifted the dominant thermal bottleneck in high-power electronic systems from external cooling hardware to internal heat-spreading pathways within multilayer packages. In this work, we addressed this emerging thermal design challenge by introducing a physics-informed framework that enables accurate prediction and efficient optimization of thermal spreading resistance under realistic multichip conditions. By coupling a domain-partitioned physics-informed neural network with a Gaussian process surrogate, the proposed approach preserves physical consistency while transforming internal thermal analysis from a simulation-driven, trial-and-error process into a data-efficient design-space exploration workflow. The framework achieves FVM-level accuracy with errors below a few percent while reducing computational cost by nearly two orders of magnitude, and it identifies geometric configurations that yield measurable reductions in spreading resistance even under strong convection, where further improvements in external cooling alone provide diminishing returns. Beyond the specific power-module architectures examined here, this methodology establishes a scalable pathway for physics-consistent thermal design in next-generation heterogeneous electronic systems, with natural extensions toward operator-learning frameworks capable of real-time optimization across broader classes of coupled multiphysics problems. In the present study, direct generalization has been validated for the investigated module architecture within the sampled design domain of chip spacing, baseplate thickness, and cooling conditions. Within this domain, interpolation to unseen design combinations is provided by the GP surrogate, whereas the present PINN remains configuration-specific. Because the physics-informed formulation does not rely on labeled temperature fields, configurations that preserve the layer topology can be treated by updating the material properties and heat-source conditions and retraining the PINN, whereas changes in the number or arrangement of layers can be accommodated by modifying the subnetworks and interface conditions before retraining. Future work will therefore investigate transfer-learning efficiency across variations in material properties, chip configurations, and layer topology, together with parameterized or operator-learning formulations for broader thermal-design generalization.

CRedit authorship contribution statement

Yonghun Kim: Writing – original draft, Methodology, Investigation, Formal analysis. **Changhyeon Yoon:** Methodology, Investigation. **Haeun Lee:** Supervision, Investigation. **Seonu Bae:** Validation, Writing – review & editing. **Nana Kang:** Formal analysis, Data curation. **Changwoo Han:** Writing – review & editing, Project administration. **Dongmin Shin:** Writing – review & editing, Supervision, Project administration. **Jungwan Cho:** Writing – review & editing, Project administration. **Sooyoung Lee:** Writing – review & editing, Supervision. **Hyoungsoon Lee:** Writing – review & editing, Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Hyundai Motor Company, the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2024-00411577), and the Chung-Ang University Graduate Research Scholarship in 2024.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.icheatmasstransfer.2026.112739>.

Data availability

Data will be made available on request.

References

- [1] J. Mathew, S. Krishnan, A review on transient thermal management of electronic devices, *J. Electron. Packag.* 144 (1) (2022) 010801.
- [2] V. Egan, J. Stafford, P. Walsh, E. Walsh, An experimental study on the design of miniature heat sinks for forced convection air cooling, *ASME J. Heat Transf.* 131 (7) (2009) 071402, <https://doi.org/10.1115/1.3110005>.
- [3] S. Mancini, C. Zilio, A. Diani, L. Rossetto, Air forced convection through metal foams: experimental results and modeling, *Int. J. Heat Mass Transf.* 62 (2013) 112–123.
- [4] Y. Yan, T. Zhao, Z. He, Z. Yang, L. Zhang, Numerical investigation on the characteristics of flow and heat transfer enhancement by micro pin-fin array heat sink with fin-shaped strips, *Chem. Eng. Process. Process Intensif.* 160 (2021) 108273.
- [5] P. Bhandari, K.S. Rawat, Y.K. Prajapati, D. Padalia, L. Ranakoti, T. Singh, Design modifications in micro pin fin configuration of microchannel heat sink for single phase liquid flow: a review, *J. Energy Storage* 66 (2023) 107548.
- [6] A.C. Kheirabadi, D. Groulx, Experimental evaluation of a thermal contact liquid cooling system for server electronics, *Appl. Therm. Eng.* 129 (2018) 1010–1025.
- [7] X. Zhang, Z. Ji, J. Wang, X. Lv, Research progress on structural optimization design of microchannel heat sinks applied to electronic devices, *Appl. Therm. Eng.* 235 (2023) 121294.
- [8] R. Van Erp, R. Soleimanzadeh, L. Nela, G. Kampitsis, E. Mantioli, Co-designing electronics with microfluidics for more sustainable cooling, *Nature* 585 (7824) (2020) 211–216.
- [9] D. Kong, H. Kwon, B. Jang, H.J. Kwon, M. Asheghi, K.E. Goodson, H. Lee, Extreme heat flux cooling from functional copper inverse opal-coated manifold microchannels, *Energy Convers. Manag.* 315 (2024) 118809.
- [10] T. Wei, H. Oprins, V. Cherman, Z. Yang, K.C. Rivera, G. Van der Plas, M. Baelmans, Experimental and numerical study of 3-D printed direct jet impingement cooling for high-power, large die size applications, *IEEE Trans. Compon. Packag. Manuf. Technol.* 11 (3) (2020) 415–425.
- [11] B. Markal, A. Evcimen, O. Aydin, Effect of inlet temperature on flow boiling behavior of expanding micro-pin-fin type heat sinks, *Int. Commun. Heat Mass Transf.* 149 (2023) 107143.
- [12] P. Sun, N. Lu, L. Li, Y. Li, H. Wang, H. Lv, M. Liu, Thermal crosstalk in 3-dimensional RRAM crossbar array, *Sci. Rep.* 5 (1) (2015) 13504.
- [13] C.C. Hsieh, C.H. Wu, D. Yu, Analysis and comparison of thermal performance of advanced packaging technologies for state-of-the-art mobile applications, in: 2016

- IEEE 66th Electronic Components and Technology Conference (ECTC), IEEE, 2016, pp. 1430–1438.
- [14] M.M. Yovanovich, Y.S. Muzychka, J.R. Culham, Spreading resistance of isoflux rectangles and strips on compound flux channels, *J. Thermophys. Heat Transf.* 13 (4) (1999) 495–500.
 - [15] T.Q. Feng, J.L. Xu, An analytical solution of thermal resistance of cubic heat spreaders for electronic cooling, *Appl. Therm. Eng.* 24 (2–3) (2004) 323–337.
 - [16] D. Guan, M. Marz, J. Liang, Analytical solution of thermal spreading resistance in power electronics, *IEEE Trans. Compon. Packag. Manuf. Technol.* 2 (2) (2012) 278–285.
 - [17] K.S. Yang, C.H. Chung, C.W. Tu, C.C. Wong, T.Y. Yang, M.T. Lee, Thermal spreading resistance characteristics of a high power light emitting diode module, *Appl. Therm. Eng.* 70 (1) (2014) 361–368.
 - [18] K.B. Abdelmleik, Z. Araoud, K. Charrada, G. Zissis, Optimization of the thermal distribution of multi-chip LED package, *Appl. Therm. Eng.* 126 (2017) 653–660.
 - [19] S. Dong, Q. Zhou, M. Wang, X. Jiang, J. Yang, Analysis of thermal spreading resistance in high power LED package and its design optimization, in: 2011 12th International Conference on Electronic Packaging Technology and High Density Packaging, IEEE, 2011, pp. 1–5.
 - [20] H. Dang, Y. Lu, Y. Du, X. Zhang, Q. Zhang, W. Ma, X. Zhang, A detailed thermal resistance network analysis of FCBGGA package, *J. Therm. Sci.* 33 (1) (2024) 18–28.
 - [21] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *Nature* 521 (7553) (2015) 436–444.
 - [22] H. Wei, S. Zhao, Q. Rong, H. Bao, Predicting the effective thermal conductivities of composite materials and porous media by machine learning methods, *Int. J. Heat Mass Transf.* 127 (2018) 908–916.
 - [23] H. Lee, G. Lee, K. Kim, D. Kong, H. Lee, Multimodal machine learning for predicting heat transfer characteristics in micro-pin fin heat sinks, *Case Stud. Therm. Eng.* 57 (2024) 104331.
 - [24] M. Raissi, P. Perdikaris, G.E. Karniadakis, Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, *J. Comput. Phys.* 378 (2019) 686–707.
 - [25] S. Manavi, T. Becker, E. Fattahi, Enhanced surrogate modelling of heat conduction problems using physics-informed neural network framework, *Int. Commun. Heat Mass Transfer* 142 (2023) 106662.
 - [26] B. Zhang, G. Wu, Y. Gu, X. Wang, F. Wang, Multi-domain physics-informed neural network for solving forward and inverse problems of steady-state heat conduction in multilayer media, *Phys. Fluids* 34 (11) (2022).
 - [27] A.D. Jagtap, E. Kharazmi, G.E. Karniadakis, Conservative physics-informed neural networks on discrete domains for conservation laws: applications to forward and inverse problems, *Comput. Methods Appl. Mech. Eng.* 365 (2020) 113028.
 - [28] A.D. Jagtap, G.E. Karniadakis, Extended physics-informed neural networks (XPINNs): a generalized space-time domain decomposition based deep learning framework for nonlinear partial differential equations, *Commun. Comput. Phys.* 28 (5) (2020).
 - [29] G.N. Ellison, Maximum thermal spreading resistance for rectangular sources and plates with nonunity aspect ratios, *IEEE Transact. Components Packag. Technol.* 26 (2) (2003) 439–454.
 - [30] S.S. Sinha, P. Zaghari, J.E. Ryu, D.C. Hopkins, A 400 W, 250 kHz (2 kW peak) integrated GaN half bridge power module in a non-isolated buck converter, in: 2024 IEEE Applied Power Electronics Conference and Exposition (APEC), IEEE, 2024, pp. 168–174.
 - [31] E. Gurpinar, R. Sahu, B. Ozpineci, Heat sink design for WBG power modules based on Fourier series and evolutionary multi-objective multi-physics optimization, *IEEE Open J. Power Electron.* 2 (2021) 559–569.
 - [32] O. Nilsson, H. Mehling, R. Horn, J. Fricke, R. Hofmann, S. Müller, D. Hofmann, Determination of the thermal diffusivity and conductivity of monocrystalline silicon carbide (300–2300K), *High Temp.-High Press.* 29 (1) (1997) 73–79.
 - [33] G.A. Slack, R.A. Tanzilli, R.O. Pohl, J.W. Vandersande, The intrinsic thermal conductivity of AlN, *J. Phys. Chem. Solids* 48 (7) (1987) 641–647.
 - [34] A. Siu-Ho, W. Qu, F. Pfefferkorn, Experimental study of pressure drop and heat transfer in a single-phase micropin-fin heat sink, *ASME J. Electron. Packag.* 129 (4) (2007) 479–487, <https://doi.org/10.1115/1.2804099>.
 - [35] D. Mei, X. Lou, M. Qian, Z. Yao, L. Liang, Z. Chen, Effect of tip clearance on the heat transfer and pressure drop performance in the micro-reactor with micro-pin-fin arrays at low Reynolds number, *Int. J. Heat Mass Transf.* 70 (2014) 709–718.
 - [36] Y.T. Shen, Y.H. Pan, H. Chen, W.L. Cheng, Experimental study of embedded manifold staggered pin-fin microchannel heat sink, *Int. J. Heat Mass Transf.* 226 (2024) 125488.
 - [37] P. Ramachandran, B. Zoph, Q.V. Le, Searching for activation functions, *arXiv preprint arXiv:1710.05941* (2017).
 - [38] D.P. Kingma, Adam: a method for stochastic optimization, *arXiv preprint arXiv:1412.6980* (2014).
 - [39] I. Loshchilov, F. Hutter, SGDR: stochastic gradient descent with warm restarts, *arXiv preprint arXiv:1608.03983* (2016).
 - [40] A. Kendall, Y. Gal, R. Cipolla, Multi-task learning using uncertainty to weigh losses for scene geometry and semantics, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7482–7491.
 - [41] N. Rahaman, A. Baratin, D. Arpit, F. Draxler, M. Lin, F. Hamprecht, A. Courville, On the spectral bias of neural networks, in: *International Conference on Machine Learning*, PMLR, 2019, pp. 5301–5310.
 - [42] M. Tancik, P. Srinivasan, B. Mildenhall, S. Fridovich-Keil, N. Raghavan, U. Singhal, R. Ng, Fourier features let networks learn high frequency functions in low dimensional domains, *Adv. Neural Inf. Process. Syst.* 33 (2020) 7537–7547.
 - [43] J. Xu, H. Wei, H. Bao, Physics-informed neural networks for studying heat transfer in porous media, *Int. J. Heat Mass Transf.* 217 (2023) 124671.
 - [44] C. Williams, C. Rasmussen, Gaussian processes for regression, *Adv. Neural Inf. Process. Syst.* 8 (1995).
 - [45] J.L. Loepky, J. Sacks, W.J. Welch, Choosing the sample size of a computer experiment: a practical guide, *Technometrics* 51 (4) (2009) 366–376.
 - [46] T. Fushiki, Estimation of prediction error by using K-fold cross-validation, *Stat. Comput.* 21 (2011) 137–146.
 - [47] P. Skubisz, H. Adrian, Estimation of heat transfer coefficient of forced-air cooling and its experimental validation in controlled processing of forgings, *Numer. Heat Transf. A Appl.* 73 (3) (2018) 163–176.
 - [48] A.J. Lelis, R. Green, D.B. Habersat, M. El, Basic mechanisms of threshold-voltage instability and implications for reliability testing of SiC MOSFETs, *IEEE Trans. Electron Devices* 62 (2) (2015) 316–323.