

Monte Carlo physics-informed neural networks for inverse multiscale heat conduction problems via the phonon Boltzmann transport equation

Qingyi Lin^a, Chuang Zhang^b, Xuhui Meng^{a,1}, Zhaoli Guo^a

^a*Institute of Interdisciplinary Research for Mathematics and Applied Science, School of Mathematics and Statistics, Huazhong University of Science and Technology, Wuhan 430074, China*

^b*Institute of Fundamental and Transdisciplinary Research, Zhejiang University, Hangzhou 310058, China*

Abstract

Inferring thermal fields and thermophysical properties from limited measurements is a fundamental challenge in micro- and nanoscale heat conduction, where the classical Fourier law breaks down and the phonon Boltzmann transport equation (BTE) must be employed to capture non-diffusive transport effects. In this work, we extend the Monte Carlo physics-informed neural networks (MC-PINNs), originally developed for forward phonon BTE problems [*J. Comput. Phys.* 542, 114364, 2025], to solve inverse multiscale heat conduction problems. Two representative classes of inverse problems are considered: (i) reconstructing the full thermal field from sparse interior temperature measurements when boundary conditions are unknown, and (ii) simultaneously inferring the unknown relaxation time, which characterizes the transport regime, together with the thermal field. Problem-specific MC-PINNs architectures and training strategies are designed for each class. The mesh-free Monte Carlo sampling strategy, inherited from the forward formulation, enables a unified treatment across diffusive, transitional, and ballistic transport regimes without requiring a priori knowledge of the relaxation time. The proposed method is evaluated on quasi-one-dimensional, quasi-two-dimensional, and three-dimensional benchmark problems covering a wide range of Knudsen numbers, as well as on a realistic 3D fin field-effect transistor (FinFET) structure. Results demonstrate that MC-PINNs consistently outperform purely data-driven deep neural networks, particularly in the sparse-data regime, and can accurately infer spatially uniform relaxation times. For spatially varying relaxation times, the inferred distributions capture the dominant thermal response, and numerical simulations using the recovered parameters reproduce the macroscopic fields with good accuracy. These findings establish MC-PINNs as an effective and physically consistent framework for inverse thermal analysis at micro- and nanoscales.

Keywords: inverse problems, physics-informed neural networks, multiscale heat conduction, phonon Boltzmann transport equation

PACS: 0000, 1111

2000 MSC: 0000, 1111

¹Corresponding author: xuhui_meng@hust.edu.cn (Xuhui Meng).

1. Introduction

Understanding and controlling heat conduction at micro- and nanoscales is essential for a wide range of applications, including thermal management of microelectronic devices, defect detection via thermography, and design of nanoscale materials [1–4]. While the classical Fourier law provides an accurate description of heat conduction in most macroscopic systems, its validity breaks down when the characteristic length or time scale approaches the mean free path or relaxation time of energy carriers [5–7]. Under such conditions, encountered at cryogenic temperatures, in nanoscale structures, or under ultrafast laser heating, non-diffusive and ballistic transport effects become significant, and the phonon Boltzmann transport equation (BTE) provides a more fundamental description of the heat conduction process [7–9].

Conventional numerical methods for solving the phonon BTE, such as the discrete ordinates method and Monte Carlo simulations, generally require complete knowledge of the thermophysical properties as well as precise boundary and initial conditions [10–12]. In many practical scenarios, however, such information is incomplete or unavailable due to complex geometries, interfacial effects, or experimental limitations [13, 14]. At the same time, partial temperature measurements inside the domain can often be acquired through advanced sensing and imaging techniques [15]. This combination of incomplete governing information and sparse observational data naturally leads to inverse heat conduction problems, in which one seeks to infer the full thermal field and/or unknown physical parameters from limited measurements.

Classical approaches to field reconstruction from sparse data, such as interpolation, example-based internal learning, and neighbor embedding [16, 17], typically rely on assumptions of spatial smoothness or local correlations, which limits their ability to resolve sharp thermal gradients or capture nonlocal transport effects characteristic of the BTE. In recent years, deep learning methods have emerged as powerful tools for tackling such inverse problems by learning complex nonlinear mappings from data [18–20]. Convolutional neural networks and generative adversarial networks, for instance, have been successfully applied to super-resolution and high-fidelity reconstruction across various physical domains [21–27]. However, purely data-driven models typically require large, high-resolution training datasets that are rarely available in practical thermal measurements, and the absence of explicit physical constraints may lead to predictions that violate the underlying transport physics.

Physics-informed neural networks (PINNs) [28, 29] offer an attractive alternative by embedding governing equations directly into the training objective through automatic differentiation [30]. This enables the inference of physically consistent solutions from sparse data without requiring large training sets, making PINNs particularly well-suited for reconstructing physical fields and identifying unknown parameters from limited or noisy measurements [31–35]. Recent studies have demonstrated the effectiveness of PINNs for thermal field reconstruction and parameter identification [36–41]. Nevertheless, most existing PINN-based thermal studies have focused on macroscopic heat conduction governed by Fourier’s law, and multiscale problems that require the phonon BTE remain largely unexplored in the inverse setting.

The application of PINNs to the phonon BTE has so far been limited to forward modeling. Li *et al.* [42, 43] demonstrated that PINNs can achieve satisfactory accuracy for forward phonon BTE problems across transport regimes ranging from diffusive to ballistic. However, their formulation relies on a priori discretization of the solid angular space, which requires a large number of angular collocation points to resolve the ballistic and near-ballistic regimes, limiting scalability to high-dimensional problems. More critically, when thermophysical properties are unknown, as is typically the case in inverse problems, an appropriate angular discretization

cannot be prescribed in advance, because the transport regime is not known a priori.

To overcome these limitations, Lin *et al.* recently proposed Monte Carlo PINNs (MC-PINNs) [44] for the forward phonon BTE. The method employs a two-step Monte Carlo sampling strategy, randomly drawing collocation points in both physical and angular spaces at each training iteration. This mesh-free approach removes the need for predefined spatial or angular grids, enables efficient exploration of the computational domain across all transport regimes, and reduces computational cost compared with grid-based PINN formulations. These properties make MC-PINNs a natural foundation for inverse phonon BTE problems, where the transport regime is unknown and a flexible, grid-free treatment is essential.

In the present work, we extend the MC-PINNs to inverse multiscale heat conduction problems governed by the phonon BTE. Two representative classes of inverse problems are investigated, each reflecting a distinct source of ill-posedness commonly encountered in micro- and nanoscale thermal measurements:

- (1) **Thermal field reconstruction with known relaxation time.** The governing equation is fully known but boundary and/or initial conditions are incomplete. Given sparse interior temperature measurements, the objective is to reconstruct the full thermal field over the entire domain.
- (2) **Simultaneous inference of the thermal field and unknown relaxation time.** The boundary and initial conditions are prescribed, but the relaxation time τ is unknown. Two sub-cases are examined: (2a) τ is spatially uniform but unknown, and (2b) τ is an unknown spatially varying function $\tau(\mathbf{x})$.

For each problem class, we develop tailored neural network architectures and training strategies within the unified MC-PINNs framework. The framework’s effectiveness is systematically evaluated on quasi-one-dimensional (quasi-1D), quasi-two-dimensional (quasi-2D), and three-dimensional (3D) test cases spanning diffusive to ballistic transport regimes, including a realistic 3D FinFET structure representative of state-of-the-art microelectronic devices.

The remainder of the paper is organized as follows. Section 2 introduces the phonon BTE, formulates the two classes of inverse problems, and presents the MC-PINNs for solving them. Section 3 presents comprehensive numerical experiments for thermal field reconstruction and relaxation time inference. Section 4 summarizes the findings and discusses directions for future work.

2. Methodology

2.1. The phonon Boltzmann transport equation and inverse problem formulation

We adopt the gray-model approximation [8] to describe multiscale heat conduction in solids, following the formulation in Ref. [44]. Under this approximation, phonon dispersion and polarization are neglected, and all phonons are assumed to share the same group speed. The dimensionless gray phonon BTE reads

$$\partial_t e'' + \mathbf{v} \cdot \nabla e'' = \frac{1}{\tau} [e^{eq} - e''] + \dot{S}, \quad (1)$$

where e'' denotes the phonon energy density per unit solid angle, and $\mathbf{v} = v_g \mathbf{s}$ is the group velocity with group speed v_g and unit directional vector $\mathbf{s} = (\cos \theta, \sin \theta \cos \phi, \sin \theta \sin \phi)$. Here, $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi)$ are the polar and azimuthal angles, respectively; τ is the relaxation

time; and $\dot{S}$ is the volumetric heat source [10–12]. The equilibrium phonon energy density e^{eq} is given by

$$e^{eq}(t, \mathbf{x}, \mathbf{s}) = \frac{C_V T}{4\pi}, \quad (2)$$

where C_V is the volumetric heat capacity. In this work, the volumetric heat capacity and group velocity are normalized by their respective reference values, yielding dimensionless values $C_V = 1$ and $v_g = 1$. The macroscopic temperature T and heat flux vector $\mathbf{q}$ are obtained by taking angular moments of e'' :

$$T = \frac{\int e'' d\Omega}{C_V}, \quad \mathbf{q} = \int \mathbf{s} e'' d\Omega. \quad (3)$$

The thermalization boundary condition is applied to close Eq. (1) [45]. Under this condition, phonons incident on a boundary are absorbed, while emitted phonons are assumed to be in equilibrium with the wall temperature, leading to the Dirichlet condition

$$e''(t, \mathbf{x}_w, \mathbf{s}) = \frac{C_V T_w}{4\pi}, \quad \mathbf{s} \cdot \mathbf{n} > 0, \quad (4)$$

where $\mathbf{x}_w$ denotes a boundary point, $\mathbf{n}$ is the outward unit normal, and T_w is the prescribed wall temperature. The initial condition for a system initially at thermal equilibrium with temperature T_0 is [46]

$$e''(t_0, \mathbf{x}, \mathbf{s}) = \frac{C_V T_0}{4\pi}. \quad (5)$$

The transport regime is characterized by the Knudsen number $\text{Kn} = \lambda/L_0$, where $\lambda = v_g \tau$ is the phonon mean free path and L_0 is the characteristic length scale. In the diffusive limit ($\text{Kn} \ll 1$), the BTE reduces to Fourier’s law. As Kn increases, nonlocal and ballistic effects become increasingly important, and the system enters the transitional and ultimately the ballistic regime ($\text{Kn} > 10$). Since L_0 and v_g are prescribed in each test case, the relaxation time τ serves as a proxy for Kn , with different values of τ corresponding to distinct transport regimes.

We now formulate the two classes of inverse problems addressed in this work. In contrast to the forward setting considered in Ref. [44], where all thermophysical properties, boundary conditions, and initial conditions are fully specified, inverse problems are characterized by incomplete information about the governing equation or its auxiliary conditions. Such situations frequently arise in practice due to inaccessible boundaries, uncertainties in material properties, or limitations of experimental characterization techniques [47].

- (1) **Thermal field reconstruction with known relaxation time.** The relaxation time τ is known, but the boundary and/or initial conditions are missing. Partial temperature measurements from sensors distributed within the domain are available, and the objective is to reconstruct the full thermal field. This setting represents situations where the governing physics is well characterized but boundary conditions are difficult to specify, for instance due to complex device geometries or inaccessible interfaces.
- (2) **Simultaneous inference of the thermal field and unknown relaxation time.** The boundary and initial conditions are prescribed, but the relaxation time τ , which determines the transport regime, is unknown. Given sparse temperature and/or heat flux

measurements, the goal is to jointly reconstruct the thermal field and identify τ . Two sub-cases are distinguished: (a) τ is spatially uniform but unknown; and (b) τ is an unknown spatially varying function $\tau(\mathbf{x})$. This setting is representative of material characterization problems where the thermophysical properties of the sample are not known a priori.

These two problem classes represent different sources of ill-posedness. In Problem (1), the governing equation is known but the missing boundary conditions prevent a direct forward solution. In Problem (2), the boundary conditions are available but an unknown coefficient in the governing equation must be inferred from sparse observations together with the solution field. This structural difference motivates the distinct solution strategies developed below.

2.2. MC-PINNs for inverse multiscale heat conduction

We extend the MC-PINNs framework developed in Ref. [44] to the inverse problems formulated above. We first briefly summarize the forward MC-PINNs formulation, which serves as the foundation, and then describe how it is adapted to each class of inverse problems.

2.2.1. Forward MC-PINNs formulation

In the forward setting, the solution to the phonon BTE is approximated by a deep neural network $e_{\text{NN}}(t, \mathbf{x}, \mathbf{s}; \boldsymbol{\theta})$ with trainable parameters $\boldsymbol{\theta}$, which takes the temporal, spatial, and angular coordinates as inputs and predicts the phonon energy density. The network is trained by minimizing a composite loss function that enforces the governing equation together with the prescribed boundary and initial conditions:

$$\mathcal{L}_{\text{fwd}} = \omega_{Eq} \mathcal{L}_{Eq} + \omega_{\text{BC}} \mathcal{L}_{\text{BC}} + \omega_{\text{IC}} \mathcal{L}_{\text{IC}}, \quad (6)$$

where the individual terms are defined as

$$\begin{aligned} \mathcal{L}_{Eq} &= \frac{1}{N} \sum_{i=1}^N |R_i(t, \mathbf{x}, \mathbf{s})|^2, \\ \mathcal{L}_{\text{BC}} &= \frac{1}{N_{\text{BC}}} \sum_{j=1}^{N_{\text{BC}}} |e_{\text{NN},j}(t, \mathbf{x}_w, \mathbf{s}) - e_j^{eq}(t, \mathbf{x}_w, \mathbf{s})|^2, \\ \mathcal{L}_{\text{IC}} &= \frac{1}{N_{\text{IC}}} \sum_{k=1}^{N_{\text{IC}}} |e_{\text{NN},k}(t_0, \mathbf{x}, \mathbf{s}) - e_k^{eq}(t_0, \mathbf{x}, \mathbf{s})|^2, \end{aligned} \quad (7)$$

with the PDE residual

$$R = \partial_t e_{\text{NN}} + \mathbf{v} \cdot \nabla e_{\text{NN}} - \frac{1}{\tau} (e^{eq} - e_{\text{NN}}) - \dot{S}. \quad (8)$$

Here, N , N_{BC} , and N_{IC} denote the numbers of residual, boundary, and initial collocation points, and ω_{Eq} , ω_{BC} , and ω_{IC} are user-specified loss weights. For steady-state problems, the initial condition term $\mathcal{L}_{\text{IC}}$ is omitted.

A defining feature of MC-PINNs is the two-step Monte Carlo sampling strategy used to construct training points at each iteration. In the first step, $B_{t,\mathbf{x}}$ points are randomly drawn in the temporal-spatial domain; in the second step, $B_{\mathbf{s}}$ points are independently drawn in the solid angular domain. The training points are then formed as the Cartesian product of these two sets. Since the collocation points are regenerated at each training step, the angular

space is explored stochastically over the course of optimization, ensuring sufficient directional coverage without requiring a fixed angular grid. This sampling strategy eliminates the need for predefined discretizations, naturally handles complex geometries, mitigates the curse of dimensionality, and also achieves a unified, mesh-free treatment across all transport regimes [44].

2.2.2. Extension to inverse problems

To address the inverse problems formulated in Sec. 2.1, the forward MC-PINNs framework is augmented with a data-fidelity term $\mathcal{L}_\Phi$ that incorporates available measurements. The composite loss function becomes

$$\mathcal{L} = \omega_\Phi \mathcal{L}_\Phi + \omega_{Eq} \mathcal{L}_{Eq} + \omega_{BC} \mathcal{L}_{BC} + \omega_{IC} \mathcal{L}_{IC}, \quad (9)$$

with

$$\mathcal{L}_\Phi = \frac{1}{N_{\text{sen}}} \sum_{m=1}^{N_{\text{sen}}} |\Phi_{\text{sen},m}(t, \mathbf{x}) - \Phi_{\text{pred},m}(t, \mathbf{x})|^2, \quad (10)$$

where Φ denotes the observable macroscopic quantities (temperature T and/or heat flux $\mathbf{q}$), N_{sen} is the number of sensor locations, and $\Phi_{\text{sen},m}$ and $\Phi_{\text{pred},m}$ are the measured and predicted values, respectively. The predicted macroscopic quantities are computed from the network output e_{NN} via the moment integrals in Eq. (3), with the angular integrals evaluated by Monte Carlo quadrature using the sampled angular points. The loss terms $\mathcal{L}_{Eq}$, $\mathcal{L}_{BC}$, and $\mathcal{L}_{IC}$ retain the definitions in Eq. (7); any term for which the corresponding information is unavailable is simply dropped. Unless otherwise noted, unit loss weights ($\omega_\Phi = \omega_{Eq} = \omega_{BC} = \omega_{IC} = 1$) are used throughout, as they are found to yield satisfactory accuracy empirically.

The solution strategy is tailored to each inverse problem class, as illustrated in Fig. 1.

Problem (1) – Unknown boundary conditions. The main network $e_{\text{NN}}(t, \mathbf{x}, \mathbf{s})$ approximates the energy density distribution as in the forward setting. Since boundary conditions are unavailable, $\mathcal{L}_{BC}$ is omitted, and the loss function consists of the PDE residual, the initial condition (for transient problems), and the data-fidelity term evaluated at the sensor locations. The known relaxation time τ is directly used in the scattering term of the residual.

Problem (2a) – Spatially uniform unknown τ . The main network architecture remains the same, but the unknown relaxation time τ is treated as a trainable parameter and optimized jointly with the network weights. The boundary/initial conditions are fully prescribed, so the loss function includes all four terms in Eq. (9), with $\mathcal{L}_\Phi$ providing the additional constraint needed to identify τ .

Problem (2b) – Spatially varying unknown τ . An auxiliary sub-network is introduced to approximate the spatially varying relaxation time $\tau(t, \mathbf{x})$. This sub-network takes $(t, \mathbf{x})$ as inputs and outputs τ , with an exponential activation function applied at the output layer to enforce positivity. The inferred τ is dynamically embedded into the scattering term of the main network’s PDE residual, and both networks are trained jointly.

The mesh-free Monte Carlo sampling strategy is especially useful in the inverse setting. Since τ is either unknown or its value cannot be used to determine an appropriate angular discretization a priori, the grid-free nature of MC-PINNs ensures that the angular space is adequately explored regardless of the transport regime. This unified treatment across diffusive, transitional, and ballistic conditions is an advantage over grid-based PINNs or conventional deterministic numerical methods, which generally require the angular discretization to be chosen according to the expected Knudsen number.

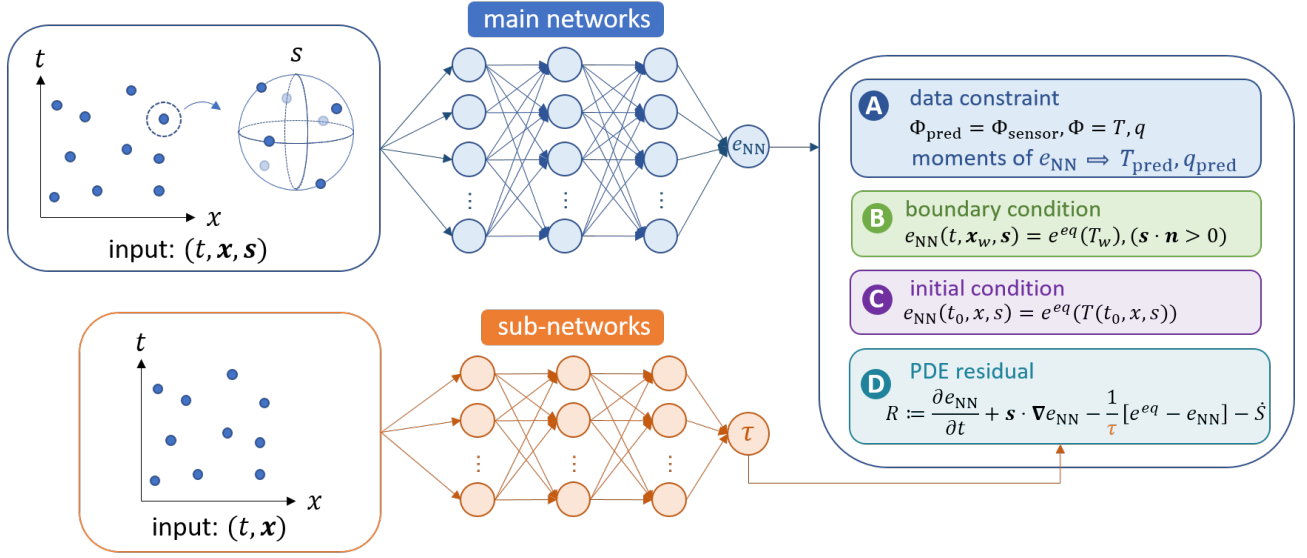


Figure 1. Schematic of the MC-PINNs framework for solving inverse problems governed by the phonon BTE. The main network takes $(t, \mathbf{x}, s)$ as inputs and outputs the phonon energy density e_{NN} . An optional sub-network is employed to approximate spatially varying relaxation times when required. The loss function combines observational data with the available physical constraints, including the governing equation and boundary/initial conditions.

3. Results and Discussion

This section presents a comprehensive evaluation of the proposed MC-PINNs framework on the two classes of inverse problems introduced in Sec. 2. In all test cases, reference solutions are obtained by solving the corresponding forward phonon BTE problems under complete boundary/initial conditions using the numerical methods of Refs. [12, 48, 49]. Sensor measurements used for training are mimicked by sampling from these reference solutions at specified locations. The network architectures, activation functions, and detailed computational settings for each case are provided in Appendix A. The prediction accuracy is quantified by the relative L_2 error,

$$E_{\Phi} = \frac{\|\Phi_{\text{ref}} - \Phi_{\text{pred}}\|_2}{\|\Phi_{\text{ref}}\|_2}, \quad \Phi = T, q, \text{ or } \tau, \quad (11)$$

where Φ_{pred} denotes the prediction (from either MC-PINNs or a purely data-driven deep neural network, DNN, used for comparison) and Φ_{ref} is the reference solution.

3.1. Thermal field reconstruction with known relaxation time

We first consider Problem (1), in which the relaxation time τ is known but the boundary conditions are unavailable, and the task is to reconstruct the full thermal field from sparse interior temperature measurements. Two representative configurations are studied: quasi-2D steady-state heat conduction in a square domain, and 3D heat conduction in a bulk FinFET structure. For each case, MC-PINNs are compared with purely data-driven DNNs trained on the same sensor measurements, to assess the effect of embedding physical constraints.

3.1.1. Quasi-2D heat conduction in a square domain

We begin with quasi-2D steady-state heat conduction in a square domain of side length $L = 1$. Four relaxation times, i.e., $\tau = 0.01, 0.1, 1.0$, and 10.0 , are examined, spanning

transport regimes from near-diffusive to ballistic. For each case, $B_x = 100$ spatial points and $B_s = 100$ angular points are randomly sampled to construct the training points used in the physics-informed loss at each training step. In this particular case, both MC-PINNs and DNNs are trained using $N_{\text{sen}} = 20$ sensor measurements randomly sampled from the reference solutions [12].

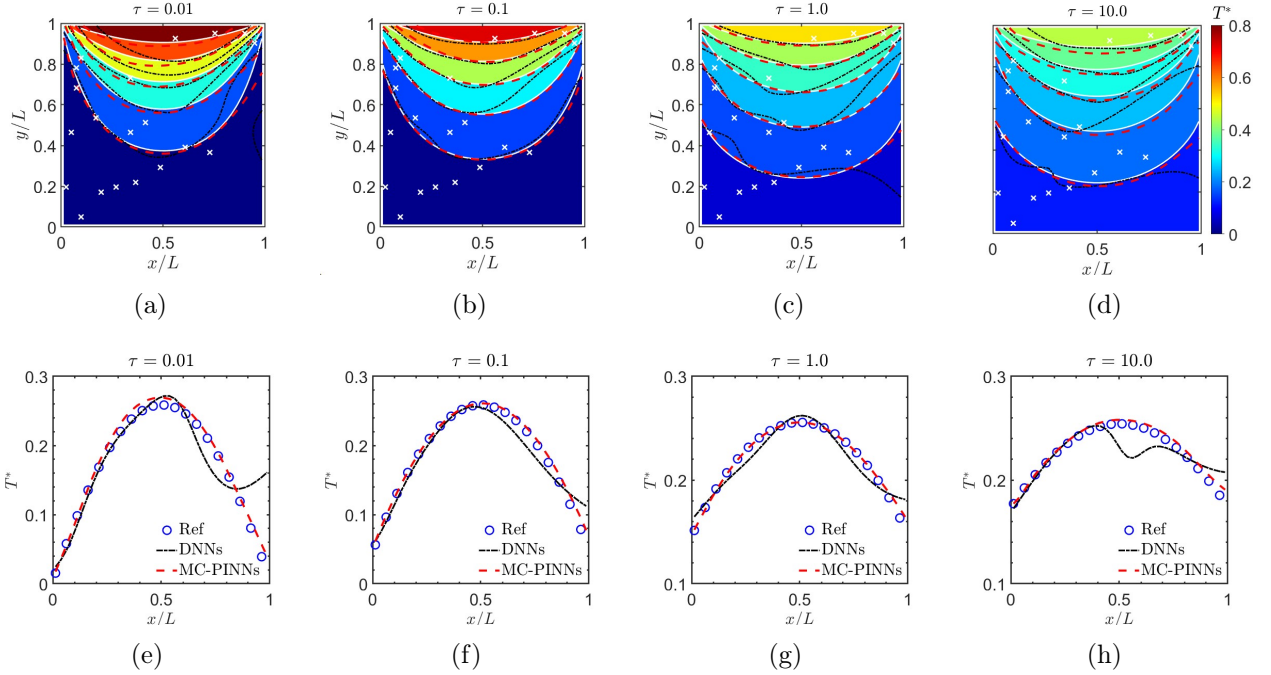


Figure 2. Quasi-2D multiscale heat conduction: reconstructed temperature fields (top row) and centerline temperature profiles at $y = 0.5$ (bottom row) obtained by DNNs (black dashed-dotted line) and MC-PINNs (red dashed line) for different relaxation times: (a, e) $\tau = 0.01$, (b, f) $\tau = 0.1$, (c, g) $\tau = 1.0$, and (d, h) $\tau = 10.0$. Colored background with white solid lines in (a–d): reference solutions [12]; white crosses: sensor locations.

Figure 2 compares the reconstructed temperature fields and centerline profiles at $y = 0.5$ obtained by the two methods. Both the contour plots and the centerline profiles show that MC-PINNs accurately recover the temperature distributions across the full range of relaxation times, whereas the DNN reconstructions exhibit noticeable errors near the boundaries and in regions with sparse observational coverage. The centerline profiles highlight this difference clearly: MC-PINNs faithfully reproduce the expected heat conduction behavior, while DNNs produce unphysical deviations in poorly observed regions.

To quantify the impact of data sparsity, we evaluate the reconstruction error across varying numbers of sensor points ($N_{\text{sen}} = 10, 20, 40$, and 80) and relaxation times. For each configuration, five independent runs are performed with randomly regenerated sensor locations, so that both the mean accuracy and the sensitivity to sensor placement can be assessed.

Figure 3 and Table 1 present the relative L_2 errors for both methods as functions of the number of sensor points. While the errors of both methods decrease with increasing sensor density, MC-PINNs consistently yield lower mean errors than DNNs across nearly all tested configurations. The advantage is most pronounced in the sparse-data regime ($N_{\text{sen}} = 10$ or 20), where DNN errors often reach or exceed 10%, whereas MC-PINNs maintain errors at approximately 5% or below. Moreover, the error bars of DNNs are substantially wider, indicating a strong dependence on the specific placement of sensors. When observations are sparse or

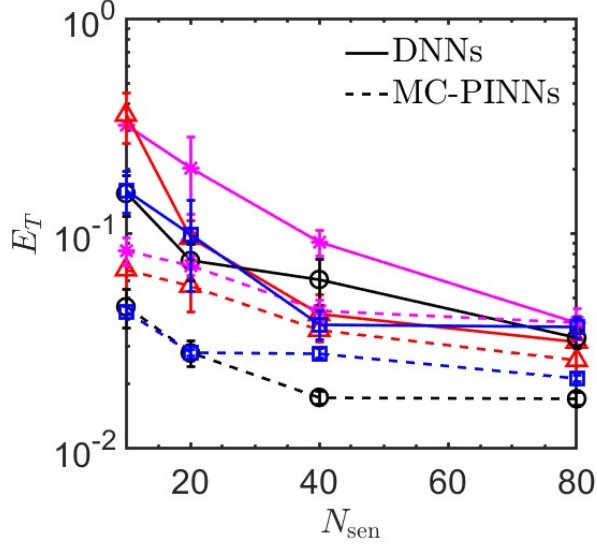


Figure 3. Quasi-2D heat conduction with spatially uniform τ : relative L_2 errors of the reconstructed temperature field obtained by DNNs (solid lines) and MC-PINNs (dashed lines) with different numbers of sensor points. Magenta: $\tau = 0.01$; black: $\tau = 0.1$; blue: $\tau = 1.0$; red: $\tau = 10.0$. Error bars denote the standard deviation over 5 independent runs.

Table 1. Mean relative L_2 error E_T (mean $\pm$ standard deviation) of DNNs and MC-PINNs with different numbers of sensor points.

τ	Method	E_T			
		$N_{\text{sen}} = 10$	$N_{\text{sen}} = 20$	$N_{\text{sen}} = 40$	$N_{\text{sen}} = 80$
0.01	DNNs	0.3208 ± 0.0587	0.2022 ± 0.0790	0.0910 ± 0.0120	0.0384 ± 0.0030
	MC-PINNs	0.0832 ± 0.0123	0.0710 ± 0.0073	0.0440 ± 0.0048	0.0386 ± 0.0063
0.1	DNNs	0.1539 ± 0.0331	0.0749 ± 0.0148	0.0612 ± 0.0149	0.0327 ± 0.0035
	MC-PINNs	0.0457 ± 0.0095	0.0279 ± 0.0037	0.0173 ± 0.0012	0.0171 ± 0.0012
1.0	DNNs	0.1592 ± 0.0353	0.0984 ± 0.0446	0.0379 ± 0.0057	0.0366 ± 0.0043
	MC-PINNs	0.0431 ± 0.0028	0.0280 ± 0.0010	0.0277 ± 0.0014	0.0212 ± 0.0007
10.0	DNNs	0.3564 ± 0.0936	0.0954 ± 0.0194	0.0422 ± 0.0101	0.0314 ± 0.0012
	MC-PINNs	0.0681 ± 0.0131	0.0571 ± 0.0138	0.0356 ± 0.0043	0.0258 ± 0.0033

unevenly distributed, DNNs produce large and unstable errors. In contrast, the narrow error bars of MC-PINNs indicate robust performance and reduced sensitivity to sensor placement. This robustness is a direct consequence of the physical constraints that regularize the solution in unobserved regions.

We further investigate the influence of the number of angular sampling points B_s on reconstruction accuracy. Four relaxation times ($\tau = 0.01, 0.1, 1.0$, and 10.0) are considered, with B_s varied as 8, 16, 32, 64, 128, and 256. The number of spatial points is fixed at $B_x = 100$, and $N_{\text{sen}} = 20$ sensor points are used.

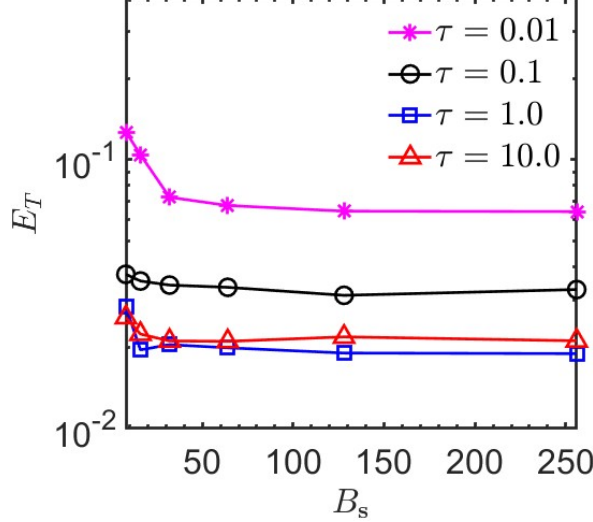


Figure 4. Quasi-2D heat conduction with spatially uniform τ : relative L_2 errors of the reconstructed temperature field for MC-PINNs with different numbers of angular sampling points. Magenta: $\tau = 0.01$; black: $\tau = 0.1$; blue: $\tau = 1.0$; red: $\tau = 10.0$.

As shown in Fig. 4, the testing error from MC-PINNs generally decreases with increasing B_s for all values of τ . The improvement saturates once $B_s \geq 32$, indicating that relatively sparse angular sampling is sufficient to capture the essential directional information in the inverse reconstruction. For $\tau = 0.1, 1.0$, and 10.0 , the relative L_2 errors fall below 5% with $B_s \geq 32$. The case $\tau = 0.01$ (near-diffusive) shows slightly higher errors, which is expected since the diffusive regime is characterized by near-isotropic angular distributions that require somewhat finer angular resolution. Note that similar results have also been reported in [44].

3.1.2. Spatially varying relaxation time

We next consider a quasi-2D case in which the relaxation time varies spatially, to evaluate the robustness of MC-PINNs for a governing equation with heterogeneous thermophysical parameters. The relaxation time varies only in the x -direction according to

$$\tau(x, y) = \frac{\tau_{\max} + \tau_{\min}}{2} - \frac{\tau_{\max} - \tau_{\min}}{2} \tanh\left(\frac{x - x_c}{2d}\right), \quad (12)$$

with $\tau_{\max} = 1.0$, $\tau_{\min} = 0.1$, $x_c = L/2$, and $d = 0.01$, corresponding to a sharp transition near the domain center. The prescribed range of τ lies in the transitional regime, where the thermal field is highly sensitive to spatial variations in the relaxation time. In the computations, $B_x = 200$ spatial points and $B_s = 100$ angular points are used, with $N_{\text{sen}} = 20$ sensor measurements for

both methods. The effect of sensor density is also evaluated with $N_{\text{sen}} = 10, 20, 40$, and 80 , using five independent runs per configuration.

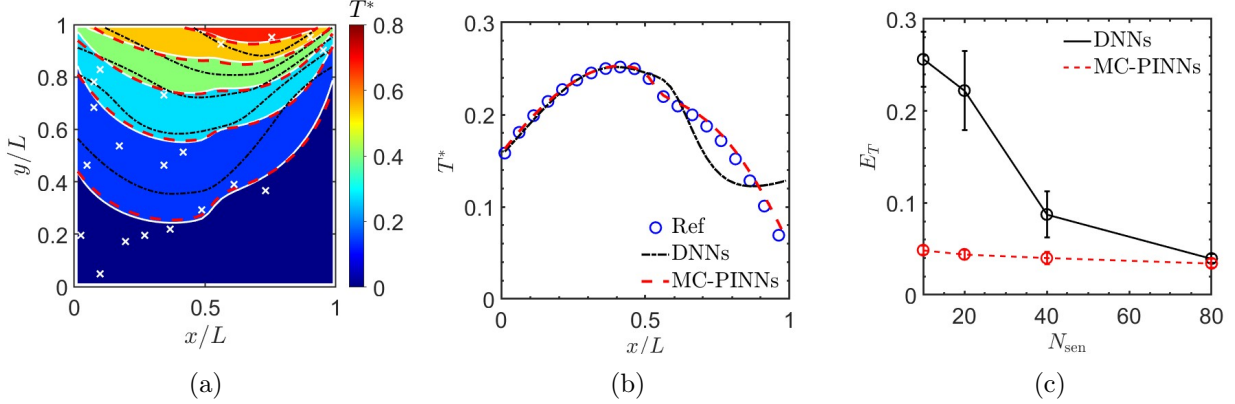


Figure 5. Quasi-2D heat conduction with spatially varying τ : (a) reconstructed temperature fields and (b) centerline temperature profiles at $y = 0.5$ obtained by DNNs (black dash-dotted) and MC-PINNs (red dashed). (c) Relative L_2 error E_T of MC-PINNs (dashed) and DNNs (solid) with different numbers of sensor points. Colored background with white solid lines in (a): reference solution [12]; white crosses: sensor locations. Error bars: standard deviation over 5 independent runs.

Figures 5(a) and 5(b) display the reconstructed temperature fields and centerline profiles at $y = 0.5$, respectively. The sharp variation in τ near $x = 0.5$ induces a pronounced change in the temperature gradient. MC-PINNs accurately capture this local feature, whereas the purely data-driven DNNs produce an overly smooth solution that fails to resolve the underlying physics. The relative L_2 error comparisons in Fig. 5(c) further quantify this advantage: across all tested sensor configurations, MC-PINNs achieve lower reconstruction errors with significantly smaller variance.

3.1.3. 3D heat conduction in FinFET structures

To assess the capability of MC-PINNs for realistic device-scale problems, we consider 3D steady-state heat conduction in a bulk FinFET structure, following the configuration in Ref. [49]. All geometric dimensions are nondimensionalized using the nanometer as the reference length scale.

The structure, illustrated in Fig. 6, comprises four bulk FinFETs with symmetrically arranged heat sources embedded in the fin regions. Exploiting geometric symmetry, only one quarter of the full structure is used as the computational domain. The bottom surface is held at a fixed temperature $T_w = 1$, while all other surfaces are treated as adiabatic and diffusely reflecting boundaries. The volumetric heat generation is prescribed as

$$\dot{S} = \begin{cases} 0.005, & \text{if } x \in [10, 18], y \in [50, 82], z \in [28, 38], \\ 0, & \text{otherwise.} \end{cases} \quad (13)$$

Two cases with different relaxation times, $\tau = 10.0$ and $\tau = 100.0$, are considered, with the corresponding Knudsen numbers lying in the near-diffusive to transitional regimes (using $L_c = 82$, the maximum device dimension along the primary heat-conduction path). In each test case, we assume $N_{\text{sen}} = 200$ sensors are randomly placed in the device to provide in situ measurements of the temperature.

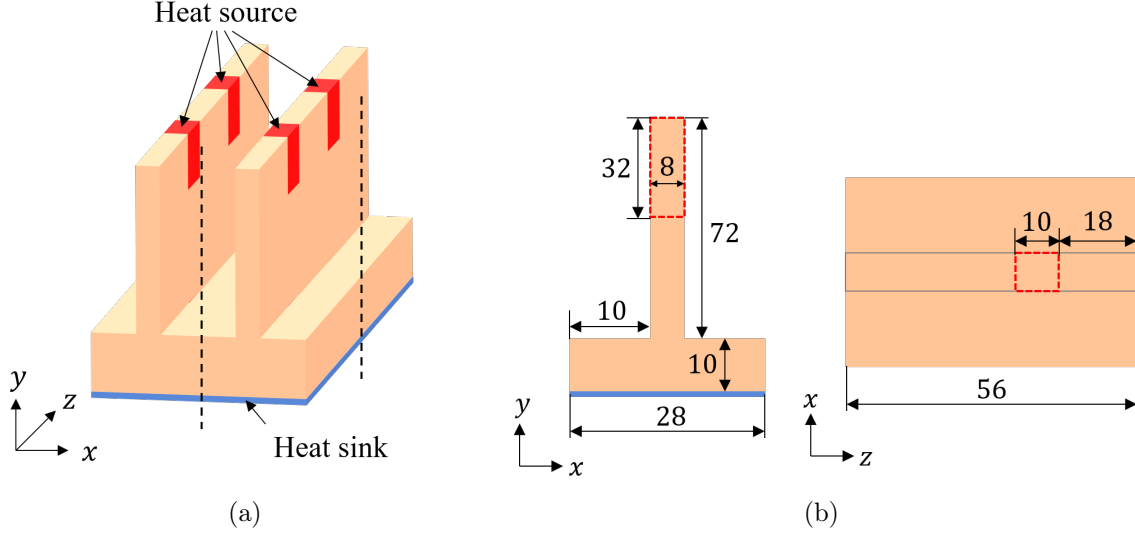


Figure 6. Schematic of the bulk FinFET structure and computational domain. (a) Three-dimensional view of four bulk FinFETs, with gray regions representing SiO_2 and light orange regions indicating Si. (b) Cross-sectional views of the computational domain in the x - y and x - z planes. Red dashed boxes indicate the heat-source regions.

For the MC-PINN computations, $B_{\mathbf{x}} = 500$ spatial points and $B_{\mathbf{s}} = 100$ angular points are randomly sampled at each iteration, with $N_{\text{sen}} = 200$ sensor points used for training. The reconstructed 3D thermal field is visualized through temperature slices in the x - y plane at $z = 29$ and in the y - z plane at $x = 14.5$. As shown in Fig. 7 and Table 2, MC-PINNs consistently achieve closer agreement with the reference solutions than DNNs for both values of τ , with relative L_2 errors approximately 2-3 times smaller. These results confirm that the proposed framework scales effectively to three-dimensional, device-scale geometries.

Table 2. 3D FinFET heat conduction: relative L_2 errors of the temperature field reconstructed by DNNs and MC-PINNs.

	$\tau = 10.0$	$\tau = 100.0$
$E_T(\text{DNNs})$	0.948%	0.807%
$E_T(\text{MC-PINNs})$	0.380%	0.309%

Several observations merit further discussion. First, examining the temperature slices in Fig. 7, both MC-PINNs and DNNs capture the overall thermal structure, including the elevated temperature in the fin regions where the heat sources are located and the gradual temperature decay toward the isothermal bottom surface. However, the DNN reconstructions exhibit visible discrepancies in the y - z plane (Figs. 7b and 7d), particularly in the lower portion of the domain near the isothermal boundary, where the temperature contours predicted by DNNs deviate from the reference isotherms. In contrast, the MC-PINNs contours closely follow the reference isotherms throughout the domain, including in regions far from sensor locations. This behavior is consistent with the observations in the quasi-2D cases: purely data-driven models struggle to extrapolate accurately beyond the vicinity of the sensors, whereas the embedded BTE residual in MC-PINNs propagates information along characteristic directions and regularizes the solution in unobserved regions.

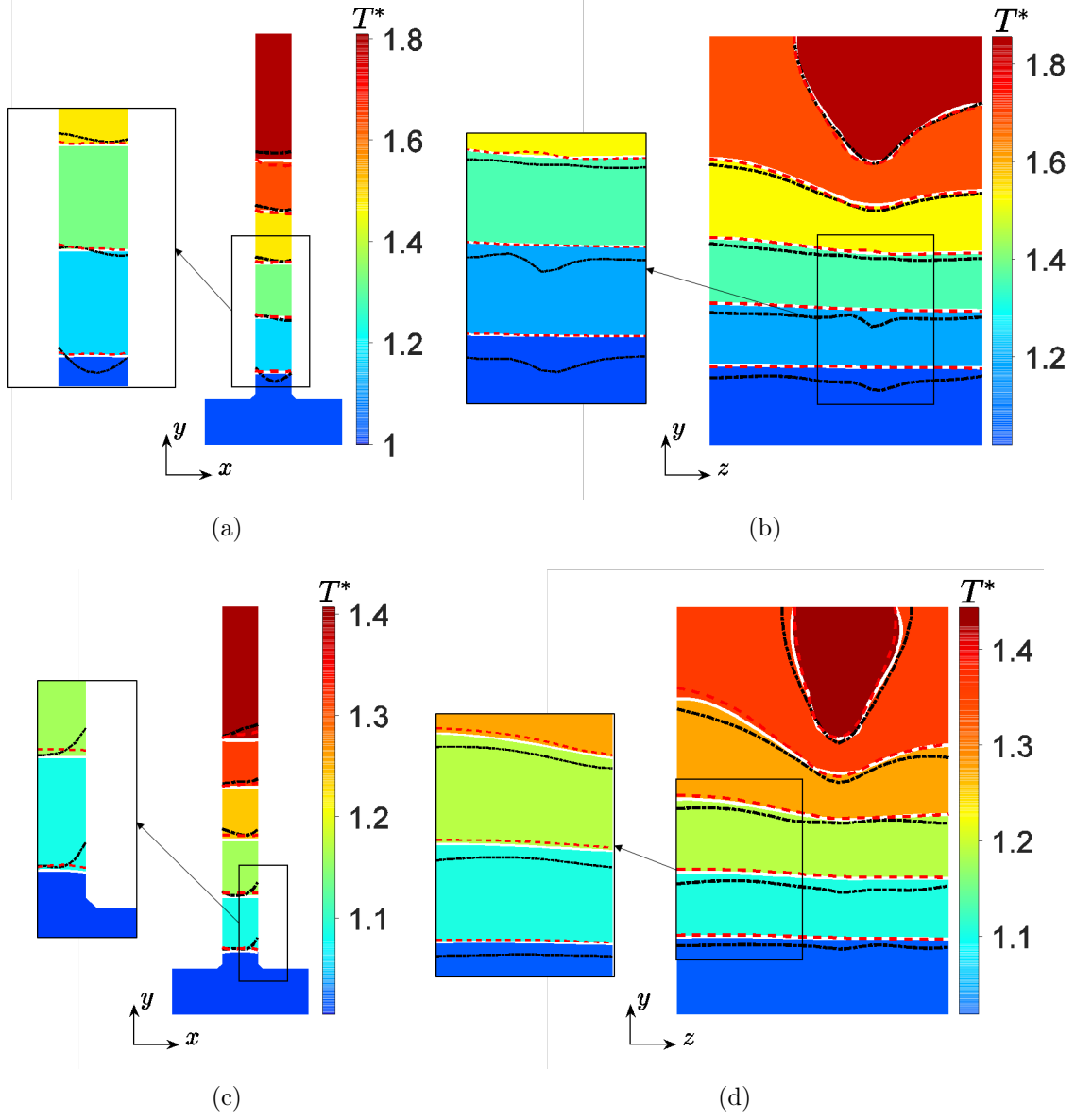


Figure 7. 3D heat conduction in a FinFET structure: reconstructed temperature slices in the x - y plane at $z = 29$ and the y - z plane at $x = 14.5$ for (a, b) $\tau = 10.0$ and (c, d) $\tau = 100.0$. Colored background with white solid lines: reference solutions [49]; black dotted-dashed: DNNs; red dashed: MC-PINNs.

Second, both methods achieve lower relative errors for $\tau = 100.0$ than for $\tau = 10.0$ (Table 2). Using the characteristic length $L_c = 82$ and the relation $\text{Kn} = v_g \tau / L_c$ with $v_g = 1$, the corresponding Knudsen numbers are $\text{Kn} \approx 0.12$ for $\tau = 10.0$ and $\text{Kn} \approx 1.22$ for $\tau = 100.0$. Both cases lie within the near-diffusive to transitional range, but their transport characteristics differ in ways that affect reconstructability. For $\tau = 10.0$ ($\text{Kn} \approx 0.12$, near-diffusive), phonon scattering is frequent and heat conduction is predominantly local, following an approximately Fourier-like behavior. The temperature field is characterized by relatively sharp thermal gradients confined near the embedded heat sources in the fin regions. For $\tau = 100.0$ ($\text{Kn} \approx 1.22$, transitional), the phonon mean free path ($\lambda = 100$) is comparable to the device dimensions, and non-local transport effects become significant. Phonons emitted from the heat sources travel further before thermalizing, distributing thermal energy more broadly across the domain. The resulting macroscopic temperature field exhibits less localized gradients and is therefore easier to reconstruct from sparse measurements for both methods. A closer examination of the temperature slices in Fig. 7 reveals that the DNN predictions exhibit noticeable discrepancies in two regions in particular. The first is near the heat sources (the fin regions in the x - y plane at $z = 29$), where the DNN contours deviate from the reference isotherms around the high-temperature zones, indicating that the purely data-driven model fails to adequately resolve the localized thermal gradients in the vicinity of the sources. The second is near the isothermal bottom boundary (visible in the y - z plane at $x = 14.5$), where the DNN-predicted temperature contours drift away from the reference solution in the lower portion of the domain. Both types of discrepancy are more pronounced for $\tau = 100.0$ (Figs. 7c and 7d): in the transitional regime, the DNN struggles to extrapolate from interior sensor measurements to the unobserved boundary, and the reconstructed temperature near the bottom surface deviates visibly from the prescribed isothermal condition. In contrast, the MC-PINNs reconstructions remain in close agreement with the reference solution in both the near-source and near-boundary regions across both values of τ . This can be attributed to the BTE residual in the loss function, which enforces the governing transport physics throughout the domain and propagates information from the interior sensors toward the boundaries along the characteristic directions of the kinetic equation, compensating for the absence of boundary data in Problem (1).

Third, it is instructive to consider the ratio of sensor points to the problem dimensionality in this 3D case. With $N_{\text{sen}} = 200$ sensor points distributed in a 3D domain of approximately $82 \times 82 \times 40$ (in dimensionless units), the effective sensor density is substantially lower than in the quasi-2D cases considered earlier. For the quasi-2D case with $N_{\text{sen}} = 20$ in a 1×1 domain, the linear sensor density is approximately 4.5 per unit length. In the FinFET case, with $N_{\text{sen}} = 200$ distributed in a domain of characteristic dimension $L_c = 82$ along the primary heat-conduction direction, the corresponding density is only about 1.4 per unit length, roughly three times sparser. Despite this reduced effective sensor density, MC-PINNs achieve relative errors below 0.4%, while DNNs, which lack physical constraints, incur errors near 1%. This comparison shows that embedding physical knowledge is more beneficial when the ratio of observational data to problem size is small, a regime of considerable practical relevance where measurements are inherently limited.

Fourth, the FinFET geometry introduces geometric complexity absent in the quasi-2D square domain: multiple material interfaces (Si and SiO_2), embedded heat sources of finite extent, and non-rectangular boundaries. In grid-based PINN formulations for the BTE, such geometric complexity would necessitate careful mesh generation and potentially local grid refinement near interfaces and small features. The MC-PINNs framework circumvents these difficulties through its mesh-free sampling strategy: collocation points are simply drawn from

the computational domain without regard to geometric features, and the two-step sampling naturally handles the higher-dimensional temporal-spatial-angular space. This demonstrates a practical advantage of the Monte Carlo approach for device-scale thermal analysis, where geometric complexity is the norm rather than the exception.

Finally, we note that the 200 sensor points used here represent only a modest fraction of the total degrees of freedom in the problem. The reference solution, obtained using the discrete ordinates method with 48×48 angular quadrature points and a spatial resolution of $dx = dy = 1$ and $dz = 2$, involves on the order of 10^5 grid points, each with angular degrees of freedom. That MC-PINNs can reconstruct the thermal field to within 0.4% accuracy using measurements at only 200 spatial locations, without knowledge of the boundary conditions, is due to the strong regularization provided by the BTE residual. The governing equation constrains the admissible solution to a low-dimensional manifold embedded in the high-dimensional discretized function space, and the sensor measurements serve primarily to select the correct point on this manifold. This property is general to physics-informed inverse methods and is not specific to the FinFET configuration, but the 3D device-scale example illustrates its practical value.

3.2. Simultaneous inference of thermal field and unknown relaxation time

We now turn to Problem (2), where the boundary conditions are prescribed but the relaxation time τ is unknown and must be inferred jointly with the thermal field. This setting is adopted to reduce the degree of ill-posedness: if both the boundary conditions and τ were unknown, sparse temperature measurements alone would generally be insufficient to uniquely determine all unknowns.

3.2.1. Inferring spatially uniform τ as a trainable parameter

We first examine the case where τ is spatially uniform but unknown, treated as a trainable parameter optimized jointly with the network weights. Steady-state heat conduction in a quasi-1D film of length $L = 1$ is considered, with ground-truth relaxation times $\tau = 0.1, 1.0$, and 10.0 . The temperatures at the left and right boundaries ($x = 0$ and $x = L$) are fixed at $T_h = 1$ and $T_c = 0$, respectively, and the thermalization boundary condition is imposed. For each case, $N_{\text{sen}} = 2$ sensor points, $B_{\mathbf{x}} = 40$ interior collocation points, and $B_{\mathbf{s}} = 64$ angular samples are used.

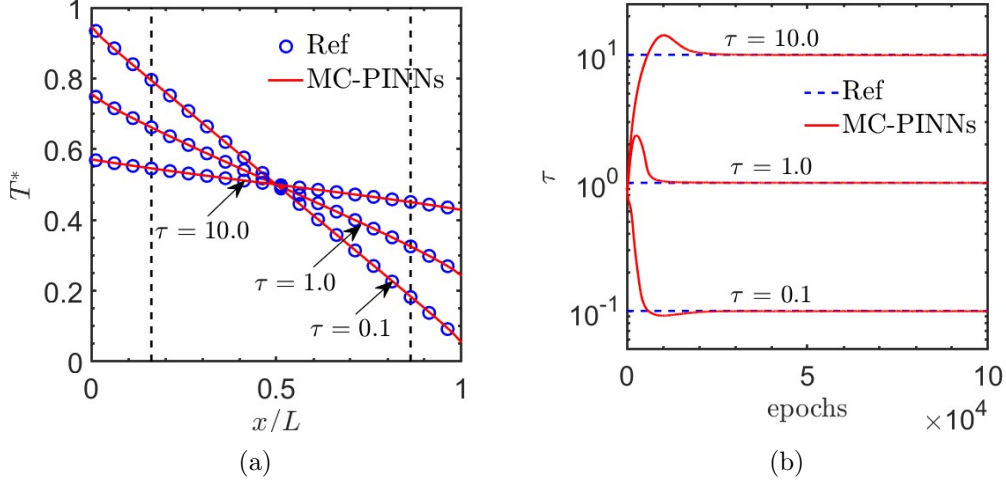


Figure 8. Quasi-1D heat conduction with unknown spatially uniform τ : (a) reconstructed temperature fields by DNNs (black dashed-dotted) and MC-PINNs (red solid) for different ground-truth τ . Black vertical dashed lines indicate sensor locations. (b) Inferred τ values compared with ground truth. Blue dashed: reference solutions [48]; red solid: MC-PINNs prediction.

Table 3. Quasi-1D heat conduction with unknown spatially uniform τ : relative L_2 errors of the reconstructed temperature and inferred τ obtained by MC-PINNs.

	$\tau = 0.1$	$\tau = 1.0$	$\tau = 10.0$
E_T	0.288%	0.079%	0.068%
E_τ	0.500%	0.100%	0.400%

Figure 8(a) shows that the temperature profiles reconstructed by MC-PINNs closely match the reference solutions [48] for all three cases, despite using only two sensor points. The predicted distributions correctly capture the boundary-layer effects that produce nonlinear temperature profiles near the walls (most evident at $\tau = 1.0$). Figure 8(b) and Table 3 show that the framework also infers τ with high accuracy: the relative errors in τ are below 0.5% for all cases, and the temperature errors remain below 0.3%.

We further examine the influence of angular sampling size B_s under the same quasi-1D setting, with $N_{\text{sen}} = 2$ and $B_x = 40$ fixed, while B_s is varied as 8, 16, 32, 64, 128, and 256.

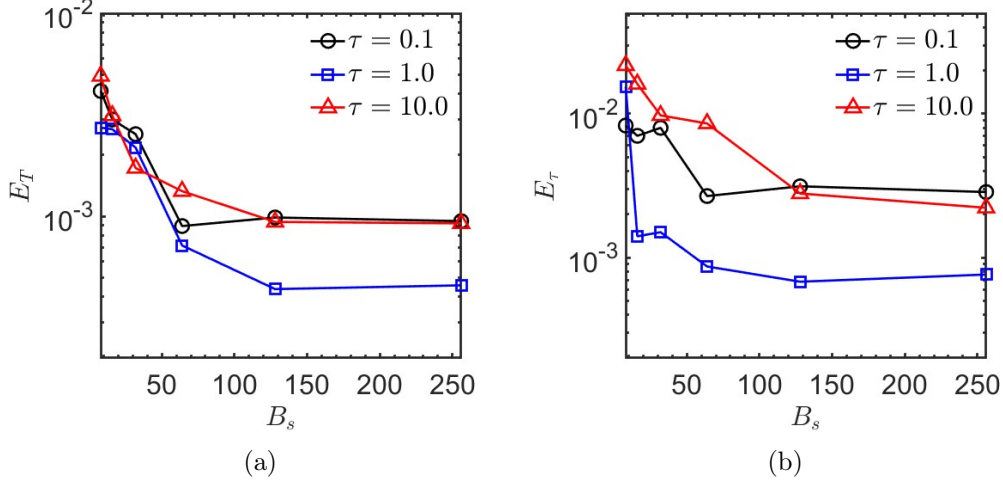


Figure 9. Quasi-1D heat conduction with unknown spatially uniform τ : relative L_2 errors of (a) the reconstructed temperature and (b) the inferred τ obtained by MC-PINNs with different numbers of angular sampling points. Black, blue, and red lines correspond to $\tau = 0.1$, 1.0 , and 10.0 , respectively.

As shown in Fig. 9, increasing B_s generally improves both temperature reconstruction and τ inference. Accurate results are already achieved with relatively modest angular sampling: for $B_s \geq 32$, the temperature error remains below 0.5% and the τ error below 1% in most cases. This insensitivity to the angular resolution is particularly valuable for inverse problems, where the transport regime is not known a priori. In grid-based PINNs or conventional deterministic numerical methods, the angular discretization must be chosen according to the expected Knudsen number, but this choice becomes ambiguous when τ is unknown. MC-PINNs avoid this difficulty through the mesh-free sampling strategy, which provides a uniform treatment across transport regimes.

A quasi-2D case is also examined with reference τ values of 0.1 and 1.0. For both cases, $N_{\text{sen}} = 40$ sensor points for the temperature are randomly sampled, with $B_x = 100$ interior collocation points and $B_s = 100$ angular points per training step. The top boundary is maintained at $T_h = 1$, while the remaining boundaries are set to $T_c = 0$.

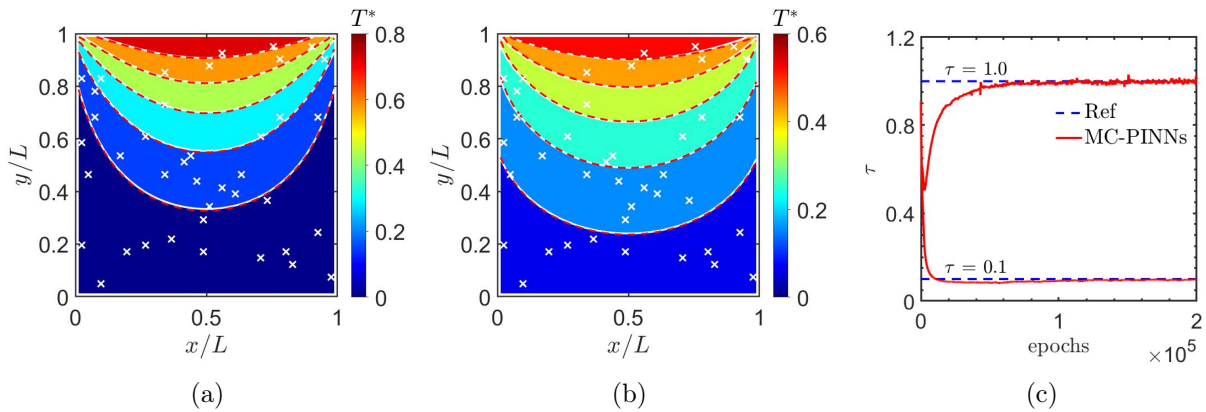


Figure 10. Quasi-2D heat conduction with unknown spatially uniform τ : reconstructed temperature fields by MC-PINNs (red dashed) for (a) $\tau = 0.1$ and (b) $\tau = 1.0$. (c) Inferred τ compared with ground truth. Colored background with white solid lines in (a, b): reference solutions [12]; white crosses: sensor locations.

Table 4. Quasi-2D heat conduction with unknown spatially uniform τ : relative L_2 errors of reconstructed temperature and inferred τ obtained by MC-PINNs.

	$\tau = 0.1$	$\tau = 1.0$
E_T	1.090%	1.771%
E_τ	0.891%	0.449%

Figure 10 and Table 4 confirm that MC-PINNs accurately reconstruct the temperature field from limited sensor data in the quasi-2D setting, even when the relaxation time is unknown. The inferred τ values are in close agreement with the ground truth, and the reconstructed solutions are consistent with the reference results. These findings demonstrate that the embedded physical constraints effectively regularize the inverse problem, enabling simultaneous field reconstruction and parameter identification with sparse measurements.

3.2.2. Inferring spatially varying τ using sub-networks

We now apply the MC-PINNs to the more challenging scenarios where the relaxation time varies spatially and is modeled as a function $\tau(\mathbf{x})$ using a sub-network, as described in Sec. 2.2. Both quasi-1D and quasi-2D steady-state problems are considered.

Quasi-1D case. Heat conduction across a film of length $L = 1$ is considered, with prescribed boundary temperatures $T_c = 0$ at $x = 0$ and $T_h = 1$ at $x = L$. The ground-truth $\tau(x)$ varies smoothly according to Eq. (14),

$$\tau(x) = \frac{\tau_{\max} + \tau_{\min}}{2} - \frac{\tau_{\max} - \tau_{\min}}{2} \tanh\left(\frac{x - x_c}{2d}\right), \quad (14)$$

with $\tau_{\min} = 0.1$, $\tau_{\max} = 1.0$, $x_c = L/2$, and $d = 0.01$. A total of $N_{\text{sen}} = 40$ sensor points are randomly distributed, with $B_{\mathbf{x}} = 100$ spatial points and $B_{\mathbf{s}} = 64$ angular points sampled at each training iteration.

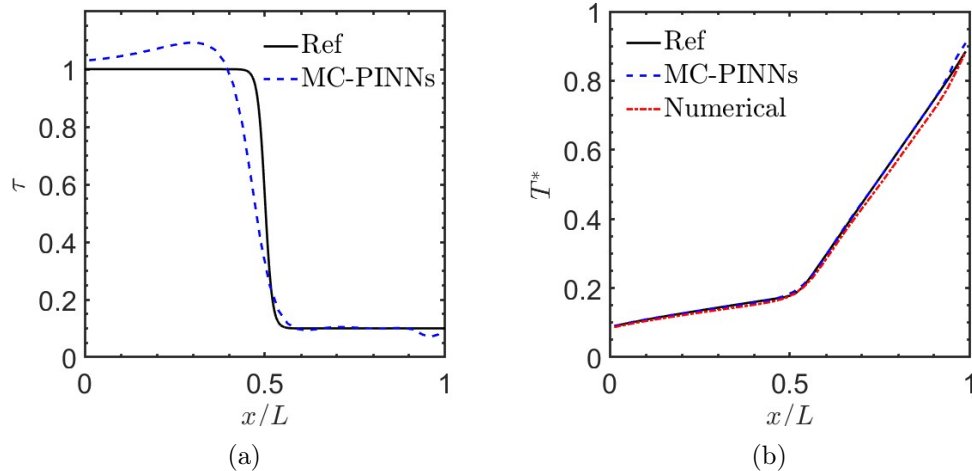


Figure 11. Quasi-1D heat conduction with unknown spatially varying τ : (a) inferred $\tau(x)$ compared with ground truth, and (b) temperature fields reconstructed by MC-PINNs and recomputed by the numerical solver using the inferred $\tau(x)$.

As shown in Fig. 11(a), the inferred $\tau(x)$ agrees well with the ground truth on the right side of the film ($\tau \approx 0.1$), but exhibits noticeable discrepancies on the left side and near the

center, where τ is larger or varies sharply. Despite these local deviations, the inferred profile captures the general variation trend and correctly identifies the location of the transition layer. Figure 11(b) shows that the temperature field reconstructed by MC-PINNs agrees well with the reference solution [48], despite the local errors in τ . Quantitatively, $E_\tau = 44.6\%$, while the directly reconstructed temperature field has an error of $E_T(\text{MC-PINNs}) = 1.472\%$. When the inferred $\tau(x)$ is used in the numerical solver, the recomputed temperature field has an error of $E_T(\text{Numerical}) = 3.284\%$. These results indicate that accurate pointwise identification of $\tau(x)$ remains challenging, particularly in regions where the macroscopic field is insensitive to variations in τ , while the inferred distribution nonetheless captures the dominant thermal response.

The larger deviations in $\tau(x)$ on the left side and near the transition are attributable to two factors. First, near the center, the sharp spatial variation in τ is difficult to resolve from sparse sensor data. Second, in regions where τ is large, the scattering term becomes relatively insensitive to changes in τ (the collision frequency scales as $1/\tau$), so variations in τ have only a weak influence on the distribution function and the resulting temperature field. Compared with the spatially uniform case (Sec. 3.2), the spatially varying $\tau(x)$ introduces a much larger solution space, increasing the possibility of non-unique reconstructions from sparse macroscopic observations. A detailed sensitivity analysis is provided in Appendix B.

Quasi-2D case. We further evaluate the sub-network approach in a quasi-2D setting, with reference $\tau = 0.1$ and $\tau = 1.0$ uniform in the ground truth, but treated as unknown spatially varying functions during inference. The temperature is set to $T_h = 1$ at the top boundary and $T_c = 0$ at the remaining boundaries. A total of $N_{\text{sen}} = 40$ sensor points are randomly distributed, with $B_x = 200$ spatial points and $B_s = 100$ angular directions per training step. The losses associated with q_x and q_y in $\mathcal{L}_\Phi$ are assigned an additional weight of $w = 10$ due to their relatively small magnitudes compared with temperature.

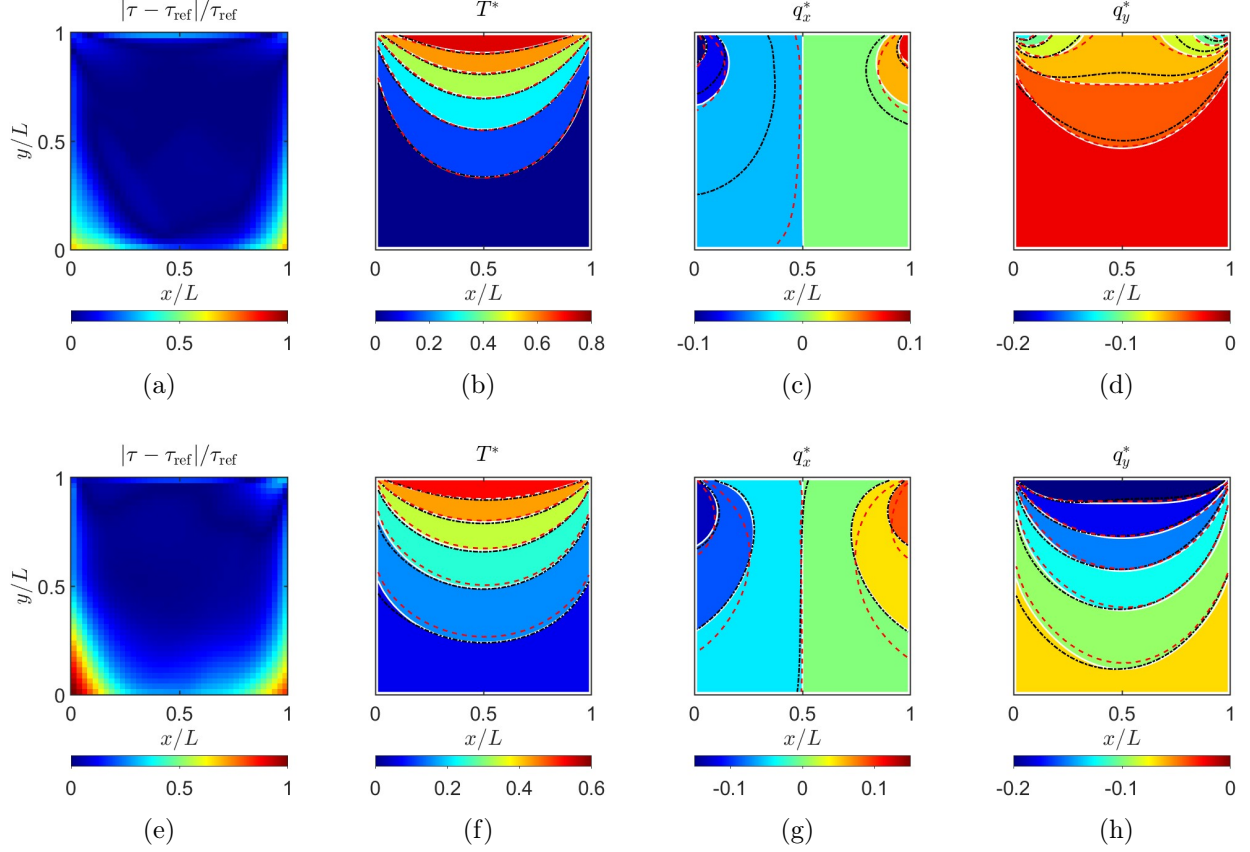


Figure 12. Quasi-2D heat conduction with unknown spatially varying τ inferred using sub-networks. Ground-truth $\tau = 0.1$ (a–d) and $\tau = 1.0$ (e–h). (a, e) Relative discrepancy between inferred and true $\tau(x, y)$; (b, f) reconstructed temperature fields from MC-PINNs (red dashed) and numerical results using inferred $\tau(x, y)$ (black dash-dotted); (c, d, g, h) reconstructed heat fluxes from MC-PINNs (red dashed) and numerical results using inferred $\tau(x, y)$ (black dash-dotted). Colored background with white solid lines: reference solutions [12].

Table 5. Quasi-2D heat conduction with unknown spatially varying τ : relative L_2 errors of inferred $\tau(x, y)$, reconstructed temperature, and heat flux from MC-PINNs.

	E_τ	E_T	E_{q_x}	E_{q_y}
$\tau_{\text{ref}} = 0.1$	32.950%	0.981%	13.582%	2.779%
$\tau_{\text{ref}} = 1.0$	20.593%	1.963%	18.763%	1.385%

Table 6. Quasi-2D heat conduction with unknown spatially varying τ : relative L_2 errors of temperature and heat flux recomputed by the numerical solver using the τ inferred by MC-PINNs.

	E_T	E_{q_x}	E_{q_y}
$\tau_{\text{ref}} = 0.1$	1.605%	10.283%	7.337%
$\tau_{\text{ref}} = 1.0$	1.170%	5.448%	3.187%

Figures 12(a) and 12(e) show that the inferred $\tau(x, y)$ agrees well with the reference in the interior but deviates near the domain boundaries. The reconstructed temperature and heat flux fields are compared in Figs. 12(b)–12(d) and 12(f)–12(h), with quantitative errors

summarized in Tables 5 and 6. Although the relative errors in τ are substantial (21%–33%) due to localized boundary deviations, the reconstructed temperature fields maintain high accuracy, with relative L_2 errors below 2% in both cases. The heat fluxes q_x and q_y exhibit larger relative errors, primarily because their small magnitudes amplify relative discrepancies, though the overall trends and distributions are captured reasonably well. Numerical simulations using the inferred $\tau(x, y)$ reproduce the macroscopic temperature and heat flux fields with moderate accuracy (Table 6), confirming that the inferred relaxation time captures the dominant thermal response despite local inaccuracies.

The larger boundary errors in τ can be attributed to two factors. First, the sensor points are distributed in the interior, leaving the boundaries without direct observational constraints. Second, as discussed in Appendix C, the phonon energy density e'' may exhibit angular discontinuities at boundaries under thermalization conditions, which complicates the inverse inference of local τ from the governing-equation residual.

4. Summary

We have extended the Monte Carlo physics-informed neural networks (MC-PINNs) developed for solving the phonon Boltzmann transport equation in [44] to solve inverse multiscale heat conduction problems governed by the phonon Boltzmann transport equation. By embedding the phonon BTE directly into the neural networks, the MC-PINN reconstructs thermal fields and identifies unknown thermophysical parameters from limited measurements while maintaining consistency with the governing transport physics. The mesh-free Monte Carlo sampling strategy, inherited from the forward MC-PINNs formulation [44], enables a unified treatment across diffusive, transitional, and ballistic transport regimes without requiring a priori knowledge of the relaxation time.

Two representative classes of inverse problems were investigated. For thermal field reconstruction under unknown boundary conditions, MC-PINNs consistently outperformed purely data-driven DNNs across quasi-2D and 3D benchmarks spanning a wide range of Knudsen numbers. The advantage is most pronounced in regions where DNNs struggle to extrapolate from sparse interior measurements, particularly near heat sources with localized thermal gradients and near unobserved boundaries. These discrepancies are especially evident for the transitional-regime FinFET case ($\tau = 100.0$). The BTE residual effectively propagates information from the sensor locations along the characteristic directions of the kinetic equation, compensating for the absence of boundary data and enabling accurate reconstruction even in device-scale 3D geometries with geometric complexity. For simultaneous inference of the thermal field and unknown relaxation time, the framework accurately identified spatially uniform τ with relative errors below 1% in both quasi-1D and quasi-2D settings, even with as few as two sensor points. When τ was treated as a spatially varying function and modeled by a sub-network, the inferred distributions captured the dominant thermal response: numerical simulations using the recovered $\tau(\mathbf{x})$ reproduced the macroscopic temperature and heat flux fields with good accuracy, despite localized pointwise deviations in τ near boundaries and in regimes of weak physical sensitivity.

The results for spatially varying τ highlight a broader open problem: the robust pointwise identification of spatially heterogeneous transport coefficients from sparse macroscopic observables. The difficulty is fundamental rather than architectural: the mapping from $\tau(\mathbf{x})$ to $(T, \mathbf{q})$ is non-injective and ill-conditioned, especially near the ballistic and diffusive limits where the macroscopic fields become insensitive to variations in τ . Addressing this challenge

motivates the integration of complementary data sources, such as surface measurements of the distribution function, multi-frequency observations, or material priors, into the MC-PINNs. More broadly, the present study shows that MC-PINNs can serve as a physically consistent framework for inverse thermal analysis at micro- and nanoscales, with potential applications in thermal imaging, hotspot detection, material characterization, and temperature monitoring in microelectronic devices.

Acknowledgments

Q. L., X. M., and Z. G. acknowledge the support of the National Natural Science Foundation of China (No. 12201229) and the Interdisciplinary Research Program of HUST (No. 2024JCYJ003 and 2023JCYJ002). C. Z. acknowledges the support of the National Natural Science Foundation of China (52506078) and Zhejiang Provincial Natural Science Foundation of China under Grant No. LMS26E060012. X. M. also acknowledges the support of the Xiaomi Young Talents Program. The authors acknowledge Beijing PARATERA Tech. CO., Ltd. for the HPC resources.

Appendix A. Computational Details

In all test cases presented in Sec. 3, the Adam optimizer with a learning rate of 10^{-3} is employed for training MC-PINNs. The detailed configurations are summarized in Table A.1. For quasi-2D cases, 128 points are used for Monte Carlo integration of the equilibrium energy density, which is found to be sufficient for accurate quadrature.

Table A.1. Network architectures and sensor points used by MC-PINNs in each case.

Section		Width $\times$ Depth	Activation function	Sensor points (N_{sen})
3.1 (quasi-2D)	Uniform τ	30×4	tanh	20
	Varying τ	30×4	tanh	20
3.1 (3D)	FinFET	50×4	tanh	200
3.2 (uniform)	Quasi-1D	20×3	tanh	2
	Quasi-2D	30×4	tanh	40
3.2 (varying)	Quasi-1D	$20 \times 3, 20 \times 2$	tanh, exp	40
	Quasi-2D	$50 \times 4, 30 \times 3$	tanh, exp	40

Details of the numerical grids used to generate the reference solutions are provided in Table A.2. Uniform spatial grids are adopted in all cases, with Gauss–Legendre quadrature applied in the solid angular domain.

Table A.2. Grid sizes used by the numerical methods to generate reference solutions.

Section		Space ($\mathbf{x}$)	Solid angular domain ($\mathbf{s}$)
3.1 (quasi-2D)	Uniform τ [12]	$100 \times 100, 40 \times 40$	$32 \times 16, 100 \times 100$
	Varying τ [12]	80×80	100×100
3.1 (3D)	FinFET [49]	$dx = dy = 1, dz = 2$	48×48
3.2 (uniform)	Quasi-1D [48]	40	100
	Quasi-2D [12]	40×40	100×100
3.2 (varying)	Quasi-1D [48]	40	100
	Quasi-2D [12]	40×40	100×100

Appendix B. Sensitivity Analysis of the Relaxation Time

To elucidate the challenges encountered in identifying spatially varying relaxation times (Sec. 3.2), we analyze the sensitivity of the macroscopic temperature field to perturbations in τ using the quasi-1D benchmark configuration. Two sets of comparisons are designed to examine the system response near the ballistic and diffusive limits. In the first set, the baseline relaxation times are $\tau_{\max} = 10.0$ (left) and $\tau_{\min} = 0.1$ (right). With $\tau_{\min}$ fixed, $\tau_{\max}$ is increased by 50% to assess the influence of large τ on the temperature distribution. In the second set, the baseline values are $\tau_{\min} = 0.01$ and $\tau_{\max} = 1.0$. Here, $\tau_{\max}$ is fixed while $\tau_{\min}$ is reduced by 50% to assess the sensitivity to small relaxation times.

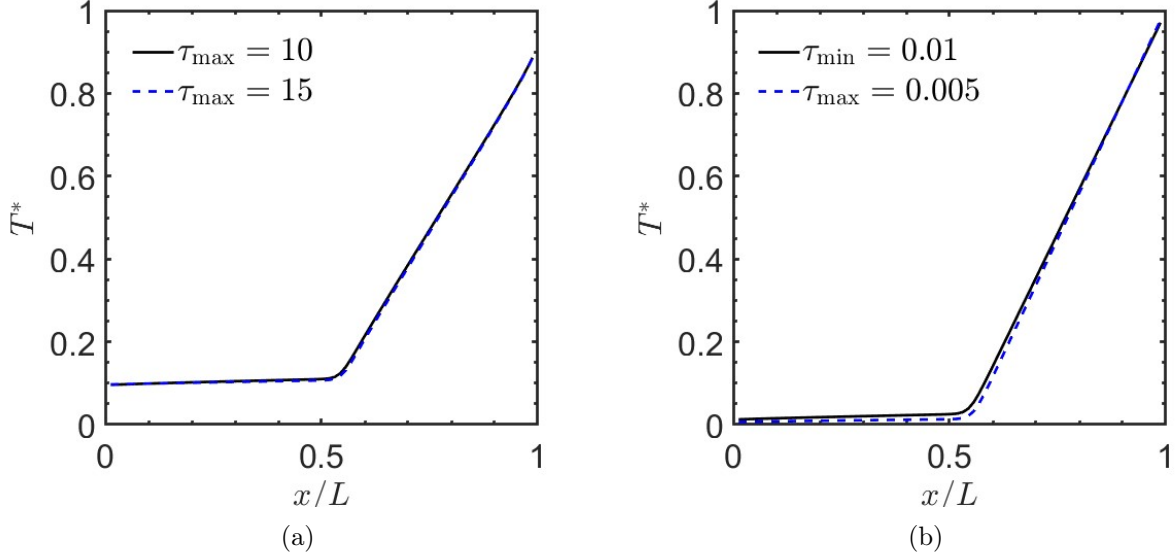


Figure B.1. Sensitivity of the temperature field to variations in the relaxation time τ . (a) Effect of increasing $\tau_{\max}$ by 50% with fixed $\tau_{\min}$. (b) Effect of decreasing $\tau_{\min}$ by 50% with fixed $\tau_{\max}$.

Figure B.1 presents the resulting temperature distributions. As shown in Fig. B.1(a), when τ is already large ($\tau > 10$, approaching the ballistic regime), a further increase produces a negligible change in the temperature field. Similarly, Fig. B.1(b) shows that when τ is sufficiently small ($\tau < 0.01$, approaching the diffusive regime), further reduction has minimal impact. In both limits, the macroscopic field is primarily governed by the boundary constraints and is

therefore insensitive to τ over a relatively wide range. This weak sensitivity reduces identifiability from macroscopic data and can lead to non-unique or inaccurate parameter estimates.

To further assess sensitivity in the transitional regime ($\tau \in [0.1, 1.0]$), we consider a baseline with $\tau_{\min} = 0.1$ and $\tau_{\max} = 1.0$, and introduce $\pm 50\%$ perturbations at both ends.

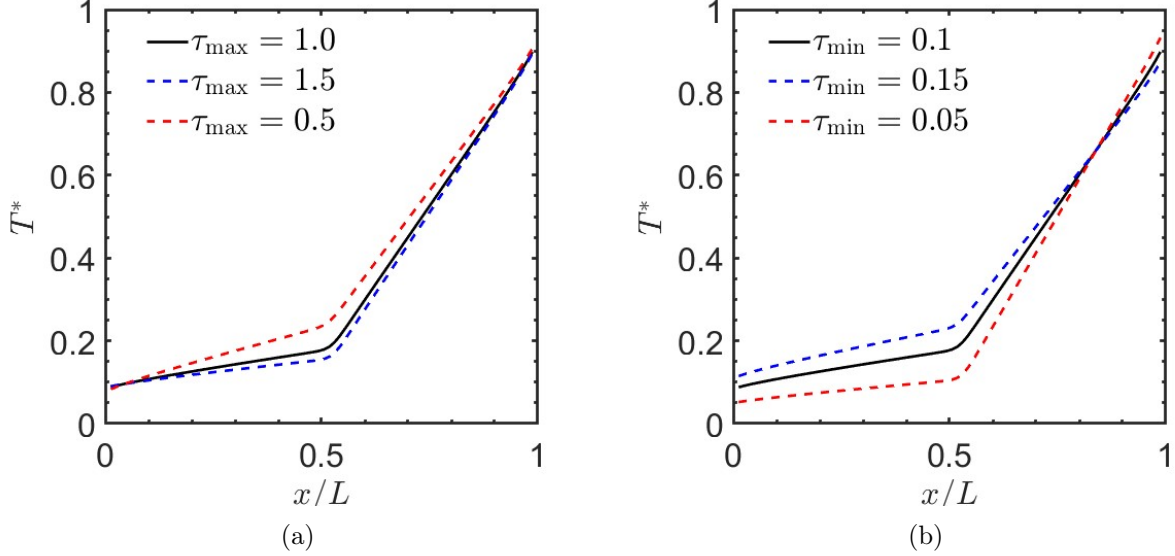


Figure B.2. Sensitivity of the temperature field to variations in τ within $\tau \in [0.1, 1.0]$. (a) Effect of $\pm 50\%$ perturbation in $\tau_{\max}$ with fixed $\tau_{\min}$. (b) Effect of $\pm 50\%$ perturbation in $\tau_{\min}$ with fixed $\tau_{\max}$.

Figures B.2(a) and B.2(b) show that perturbations in τ within the transitional regime produce more noticeable changes in the temperature field compared with the limiting regimes. Moreover, changes in the smaller relaxation time exert a stronger influence than those in the larger one, consistent with the $1/\tau$ scaling of the collision frequency in the scattering term. For large τ , the collision frequency is already low, and further changes produce relatively weak modifications of the scattering process. Consequently, the macroscopic field is inherently less sensitive to perturbations in regions with large τ , making accurate estimation in such regions more difficult. This is consistent with the inversion results in Fig. 11(a), where larger deviations from the reference occur in the high- τ region. The weak sensitivity implies that multiple relaxation-time distributions may be compatible with the same observed macroscopic temperature field, which is an inherent limitation of inferring spatially varying transport coefficients from sparse, macroscopic data alone.

Appendix C. Discontinuity of the Phonon Energy Density at Boundaries

Under thermalization boundary conditions, the phonon energy density e'' may exhibit discontinuities at solid boundaries. The incoming phonon distribution at the wall is prescribed as the equilibrium distribution at the wall temperature, while the outgoing distribution is determined by the solution of the BTE inside the domain, naturally introducing angular discontinuities.

To illustrate this effect, we compare e'' near the top boundary for the quasi-2D heat conduction problem discussed in Sec. 3.2.

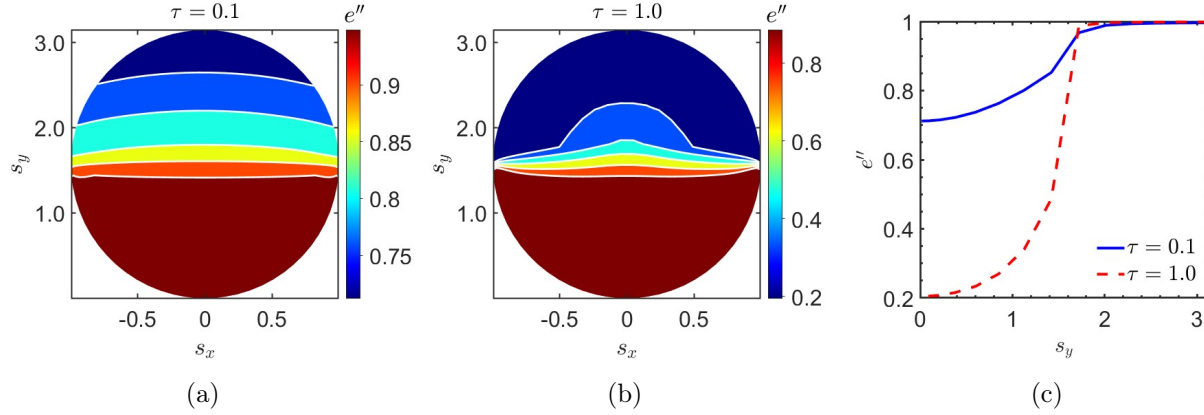


Figure C.1. Phonon energy density e'' in angular space at the top boundary center ($x = 0.5, y = 1.0$): (a) $\tau = 0.1$, (b) $\tau = 1.0$, and (c) angular profiles along s_y at $s_x = 0.0$ for both cases.

Figure C.1 displays the angular distribution of e'' at the center of the top wall ($x = 0.5, y = 1.0$). For $\tau = 0.1$ (near-diffusive, Fig. C.1(a)), the distribution remains relatively smooth, as scattering is still dominant. For $\tau = 1.0$ (transitional, Fig. C.1(b)), ballistic effects become pronounced, and a clear discontinuity appears between the imposed incoming distribution and the computed outgoing distribution. This contrast is further illustrated in Fig. C.1(c), which shows the density profiles along s_y at $s_x = 0.0$ for both cases.

Since MC-PINNs approximate the solution using smooth neural networks, such sharp boundary discontinuities can induce larger local approximation errors, which propagate into the governing-equation residual and reduce the accuracy of τ recovered from the residual at the boundary. This provides a plausible explanation for the larger boundary errors observed in the inferred $\tau(x, y)$ in Figs. 12(a) and 12(e) of the main text.

References

- [1] W. Liu, B. Yang, Thermography techniques for integrated circuits and semiconductor devices, *Sensor Review* 27 (4) (2007) 298–309.
- [2] I. J. Aldave, P. V. Bosom, L. V. González, I. L. De Santiago, B. Vollheim, L. Krausz, M. Georges, Review of thermal imaging systems in composite defect detection, *Infrared Physics & Technology* 61 (2013) 167–175.
- [3] J. Christofferson, K. Maize, Y. Ezzahri, J. Shabani, X. Wang, A. Shakouri, Microscale and nanoscale thermal characterization techniques, *Journal of Electronic Packaging* 130 (4) (2008) 041101.
- [4] X. Li, H. Shao, S. Lu, J. Xiang, B. Cai, Highly efficient fault diagnosis of rotating machinery under time-varying speeds using lsismm and small infrared thermal images, *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 52 (12) (2022) 7328–7340.
- [5] G. Benenti, D. Donadio, S. Lepri, R. Livi, Non-fourier heat transport in nanosystems, *La Rivista del Nuovo Cimento* 46 (3) (2023) 105–161.
- [6] G. Chen, Non-fourier phonon heat conduction at the microscale and nanoscale, *Nature Reviews Physics* 3 (8) (2021) 555–569.

- [7] M. Wang, N. Yang, Z.-Y. Guo, Non-fourier heat conductions in nanomaterials, *Journal of Applied Physics* 110 (6) (2011) 064310.
- [8] J. Kaiser, T. Feng, J. Maassen, X. Wang, X. Ruan, M. Lundstrom, Thermal transport at the nanoscale: A fourier’s law vs. phonon Boltzmann equation study, *Journal of Applied Physics* 121 (4) (2017) 044302.
- [9] A. J. McGaughey, A. Jain, H.-Y. Kim, B. Fu, Phonon properties and thermal conductivity from first principles, lattice dynamics, and the Boltzmann transport equation, *Journal of Applied Physics* 125 (1) (2019) 011101.
- [10] Y. Guo, M. Wang, Lattice Boltzmann modeling of phonon transport, *Journal of Computational Physics* 315 (2016) 1–15.
- [11] C. Zhang, Z. Guo, S. Chen, Unified implicit kinetic scheme for steady multiscale heat transfer based on the phonon Boltzmann transport equation, *Physical Review E* 96 (6) (2017) 063311.
- [12] C. Zhang, S. Huberman, X. Song, J. Zhao, S. Chen, L. Wu, Acceleration strategy of source iteration method for the stationary phonon Boltzmann transport equation, *International Journal of Heat and Mass Transfer* 217 (2023) 124715.
- [13] W. F. Ames, *Nonlinear partial differential equations in engineering*, Academic press, 1965.
- [14] J. Tu, G. H. Yeoh, C. Liu, Y. Tao, *Computational fluid dynamics: a practical approach*, Elsevier, 2023.
- [15] A. Reihani, Y. Luan, S. Yan, J. W. Lim, E. Meyhofer, P. Reddy, Quantitative mapping of unmodulated temperature fields with nanometer resolution, *ACS nano* 16 (1) (2021) 939–950.
- [16] S. Fadnavis, Image interpolation techniques in digital image processing: an overview, *International Journal of Engineering Research and Applications* 4 (10) (2014) 70–73.
- [17] K. I. Kim, Y. Kwon, Example-based learning for single-image super-resolution, in: *Joint Pattern Recognition Symposium*, Springer, 2008, pp. 456–465.
- [18] H. K. Aggarwal, M. P. Mani, M. Jacob, Modl: Model-based deep learning architecture for inverse problems, *IEEE Transactions on Medical Imaging* 38 (2) (2018) 394–405.
- [19] Y. Bai, W. Chen, J. Chen, W. Guo, Deep learning methods for solving linear inverse problems: Research directions and paradigms, *Signal Processing* 177 (2020) 107729.
- [20] S. Arridge, P. Maass, O. Öktem, C.-B. Schönlieb, Solving inverse problems using data-driven models, *Acta Numerica* 28 (2019) 1–174.
- [21] M. T. McCann, K. H. Jin, M. Unser, Convolutional neural networks for inverse problems in imaging: A review, *IEEE Signal Processing Magazine* 34 (6) (2017) 85–95.
- [22] G. Ongie, A. Jalal, C. A. Metzler, R. G. Baraniuk, A. G. Dimakis, R. Willett, Deep learning techniques for inverse problems in imaging, *IEEE Journal on Selected Areas in Information Theory* 1 (1) (2020) 39–56.

- [23] K. H. Jin, M. T. McCann, E. Froustey, M. Unser, Deep convolutional neural network for inverse problems in imaging, *IEEE Transactions on Image Processing* 26 (9) (2017) 4509–4522.
- [24] Z. Wei, X. Chen, Physics-inspired convolutional neural network for solving full-wave inverse scattering problems, *IEEE Transactions on Antennas and Propagation* 67 (9) (2019) 6138–6148.
- [25] Y. Xie, A. Franz, M. Chu, N. Thuerey, tempogan: A temporally coherent, volumetric gan for super-resolution fluid flow, *ACM Transactions on Graphics (TOG)* 37 (4) (2018) 1–15.
- [26] B. Liu, J. Tang, H. Huang, X.-Y. Lu, Deep learning methods for super-resolution reconstruction of turbulent flows, *Physics of Fluids* 32 (2) (2020) 025105.
- [27] K. Fukami, K. Fukagata, K. Taira, Super-resolution reconstruction of turbulent flows with machine learning, *Journal of Fluid Mechanics* 870 (2019) 106–120.
- [28] 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, *Journal of Computational physics* 378 (2019) 686–707.
- [29] S. Cai, Z. Mao, Z. Wang, M. Yin, G. E. Karniadakis, Physics-informed neural networks (PINNs) for fluid mechanics: A review, *Acta Mechanica Sinica* 37 (12) (2021) 1727–1738.
- [30] A. G. Baydin, B. A. Pearlmutter, A. A. Radul, J. M. Siskind, Automatic differentiation in machine learning: a survey, *Journal of Machine Learning Research* 18 (153) (2018) 1–43.
- [31] Y. Chen, L. Lu, G. E. Karniadakis, L. Dal Negro, Physics-informed neural networks for inverse problems in nano-optics and metamaterials, *Optics express* 28 (8) (2020) 11618–11633.
- [32] A. D. Jagtap, Z. Mao, N. Adams, G. E. Karniadakis, Physics-informed neural networks for inverse problems in supersonic flows, *Journal of Computational Physics* 466 (2022) 111402.
- [33] L. Bian, H. Song, L. Peng, X. Chang, X. Yang, R. Horstmeyer, L. Ye, C. Zhu, T. Qin, D. Zheng, et al., High-resolution single-photon imaging with physics-informed deep learning, *Nature Communications* 14 (1) (2023) 5902.
- [34] H. Gao, L. Sun, J.-X. Wang, Super-resolution and denoising of fluid flow using physics-informed convolutional neural networks without high-resolution labels, *Physics of Fluids* 33 (7) (2021) 073603.
- [35] S. Cai, Z. Wang, F. Fuest, Y. J. Jeon, C. Gray, G. E. Karniadakis, Flow over an espresso cup: inferring 3-d velocity and pressure fields from tomographic background oriented schlieren via physics-informed neural networks, *Journal of Fluid Mechanics* 915 (2021) A102.
- [36] S. Cai, Z. Wang, S. Wang, P. Perdikaris, G. E. Karniadakis, Physics-informed neural networks for heat transfer problems, *Journal of Heat Transfer* 143 (6) (2021) 060801.

- [37] 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, *Physics of Fluids* 34 (11) (2022) 116116.
- [38] X. Liu, W. Peng, Z. Gong, W. Zhou, W. Yao, Temperature field inversion of heat-source systems via physics-informed neural networks, *Engineering Applications of Artificial Intelligence* 113 (2022) 104902.
- [39] Z. Li, Y. Wang, H. Liu, Z. Wang, B. Dong, Solving the Boltzmann equation with a neural sparse representation, *SIAM Journal on Scientific Computing* 46 (2) (2024) C186–C215.
- [40] H. Tang, A. Melnikov, M. Liu, S. Sfarra, H. Zhang, A. Mandelis, Physics informed neural networks for solving inverse thermal wave coupled boundary-value problems, *International Journal of Heat and Mass Transfer* 245 (2025) 126985.
- [41] S. Koric, D. W. Abueidda, Data-driven and physics-informed deep learning operators for solution of heat conduction equation with parametric heat source, *International Journal of Heat and Mass Transfer* 203 (2023) 123809.
- [42] R. Li, E. Lee, T. Luo, Physics-informed neural networks for solving multiscale mode-resolved phonon Boltzmann transport equation, *Materials Today Physics* 19 (2021) 100429.
- [43] R. Li, J.-X. Wang, E. Lee, T. Luo, Physics-informed deep learning for solving phonon Boltzmann transport equation with large temperature non-equilibrium, *npj Computational Materials* 8 (1) (2022) 29.
- [44] Q. Lin, C. Zhang, X. Meng, Z. Guo, Monte carlo physics-informed neural networks for multiscale heat conduction via phonon boltzmann transport equation, *Journal of Computational Physics* (2025) 114364.
- [45] M. J. Fryer, H. Struchtrup, Moment model and boundary conditions for energy transport in the phonon gas, *Continuum Mechanics and Thermodynamics* 26 (2014) 593–618.
- [46] Z. Guo, K. Xu, Discrete unified gas kinetic scheme for multiscale heat transfer based on the phonon Boltzmann transport equation, *International Journal of Heat and Mass Transfer* 102 (2016) 944–958.
- [47] D. G. Cahill, P. V. Braun, G. Chen, D. R. Clarke, S. Fan, K. E. Goodson, P. Keblinski, W. P. King, G. D. Mahan, A. Majumdar, et al., Nanoscale thermal transport. ii. 2003–2012, *Applied physics reviews* 1 (1) (2014).
- [48] C. Zhang, Z. Guo, S. Chen, An implicit kinetic scheme for multiscale heat transfer problem accounting for phonon dispersion and polarization, *International Journal of Heat and Mass Transfer* 130 (2019) 1366–1376.
- [49] C. Zhang, Z. Xin, Q. Lou, H. Liang, Effects of heating strategies and ballistic transport on the thermal conduction in fin field-effect transistors, *Applied Thermal Engineering* 271 (2025) 126293.