Modeling heat conduction with dual-dissipative variables: A mechanism-data fusion method

Leheng Chen,[1] Chuang Zhang,[2,*] and Jin Zhao[3,4,†]
[1]*School of Mathematical Sciences, Peking University, Beijing 100871, China*
[2]*Department of Physics, Hangzhou Dianzi University, Hangzhou 310018, China*
[3]*Academy for Multidisciplinary Studies, Capital Normal University, Beijing 100048, China*
[4]*Beijing National Center for Applied Mathematics, Beijing 100048, China*

(Received 2 May 2024; accepted 12 July 2024; published 7 August 2024)

Many macroscopic non-Fourier heat conduction models have been developed in the past decades based on Chapman-Enskog, Hermite, or other small perturbation expansion methods. These macroscopic models have achieved great success in capturing non-Fourier thermal behaviors in solid materials, but most of them are limited by small Knudsen numbers and incapable of capturing highly nonequilibrium or ballistic thermal transport. In this paper, we provide a different strategy for constructing macroscopic non-Fourier heat conduction modeling, that is, using data-driven deep-learning methods combined with nonequilibrium thermodynamics instead of small perturbation expansion. We present the mechanism-data fusion method, an approach that seamlessly integrates the rigorous framework of conservation-dissipation formalism (CDF) with the flexibility of machine learning to model non-Fourier heat conduction. Leveraging the conservation-dissipation principle with dual-dissipative variables, we derive an interpretable series of partial differential equations, fine tuned through a training strategy informed by data from the phonon Boltzmann transport equation. Moreover, we also present the inner-step operation to narrow the gap from the discrete form to the continuous system. Through numerical tests, our model demonstrates excellent predictive capabilities across various heat conduction regimes, including diffusive, hydrodynamic, and ballistic regimes, and displays its robustness and precision even with discontinuous initial conditions.

DOI: 10.1103/PhysRevE.110.025303

I. INTRODUCTION

Heat conduction in solid materials is generally described by the Fourier's law in our daily life, which implies a diffusive phonon transport process [1,2]. The heat flux is proportional to the temperature gradient and the bulk thermal conductivity only depends on the materials' components and temperature in this classic empirical formula. However, with the rapid development of advanced manufacturing, ultrafast lasers, nanomaterials and other technologies [1,3,4], specifically, the laser response time has been shortened from microseconds to picoseconds or even femtoseconds [5], and with the emergence of low-dimensional materials [6–8] such as graphene and carbon nanotubes, typical macroscopic models or theories are difficult to accurately describe the thermal transport phenomenon at the micro- and nanoscale [9,10]. When the system characteristic length or time is comparable to or smaller than the phonon mean free path or relaxation time, or the momentum-conserved normal scattering process dominates heat conduction, the Fourier's law will be broken. Lots of non-Fourier heat conduction phenomena have been found [1,3,4,9–11], for example, wavelike propagation of heat with finite speed or heat wave [5,11], size effects [3,4], and hydrodynamic phonon transport with sufficient normal scattering process [6–8]. To study the non-Fourier heat conduction processes in solid materials, many theoretical and numerical methods have been developed as well as experimental measuring techniques [2–4,7,12–14]. Compared to microscopic or mesoscopic methods, macroscopic equations have fewer degrees of freedom, higher computational efficiency, and can efficiently solve the engineering multiscale heat transfer problems by introducing some empirical correction terms or coefficients so that they have a wider audience until today and many researchers are still keen on macroscopic non-Fourier heat conduction modeling [5,15–17].

In the past decades, many macroscopic heat conduction models have been developed for describing non-Fourier heat conduction phenomena by introducing phase lag, nonlocal, nonlinear, fractional, or other complex high-order terms to reflect the actual relationship between heat flux and temperature at the micro- and nanoscales [9,15,17–20]. However, the task of building models that capture the fundamental aspects of the underlying physics in simple, understandable, and reliably universal forms continues to be a complex and challenging endeavor [21]. This challenge is dual faceted: On one hand, the deliberate pace of theoretical advancement in physics encumbers rapid modeling innovation; on the other hand, rigid physical axioms severely constrain the creation of computationally feasible models which, compounded by the inevitable distortions wrought by mathematical abstractions, circumscribe their applicability. For example, many macroscopic non-Fourier heat conduction models are derived by

*Contact author: zhangc520@hdu.edu.cn
†Contact author: zjin@cnu.edu.cn

the low-order Chapman-Enskog expansion of the Boltzmann transport equation, which requires that the system characteristic size or time should be larger than the mean free path or relaxation time [9].

In this paper, we provide a different strategy for constructing macroscopic non-Fourier heat conduction modeling, that is, using data-driven deep-learning methods [22–26] combined with nonequilibrium thermodynamics instead of typical small perturbation expansion [5,15–17]. First, we theoretically derive a macroscopic heat conduction equation with unknown functions or parameters which is valid at any scale, and second, using data-driven deep-learning methods, we train and learn these unknown functions or parameters. We present the mechanism-data fusion method, an innovative approach that seamlessly integrates the rigorous framework of conservation-dissipation formalism (CDF) with the flexibility of machine learning to model non-Fourier heat conduction. Through comprehensive numerical tests, our model demonstrates superior predictive capabilities across various heat conduction regimes, including diffusive, hydrodynamic, and ballistic regimes, and displays its robustness and precision even with discontinuous initial conditions.

The paper is organized as follows. The theoretical derivations of the heat conduction model with two dissipative variables are presented in Sec. II. In Sec. III, we show how to learn the unknown functions and optimize the parameters, and the inner-step operation (ISO) method is presented here. Additionally, the training data generated by the phonon Boltzmann transport equation (BTE) is introduced in this section. Many numerical tests and discussions are conducted to validate our model in Secs. IV and V. Conclusions and remarks are given in Sec. VI.

II. MODEL

In this section, we use the framework of conservation-dissipation formalism (CDF) to derive the model of heat conduction with two dissipative variables, and the CDF can guarantee that this model naturally satisfies the first and second laws of thermodynamics. With reasonable assumption for the entropy of the system, we can obtain the specific form of the model.

A. Model derivation

For the heat conduction in solid materials without external heat source, the first law of thermodynamics reads as $\partial_t u + \nabla \cdot \mathbf{q} = 0$, where u is the internal energy and $\mathbf{q}$ is the corresponding heat flux. From the mathematical point of view, Fourier's law is $\mathbf{q} = -\kappa \nabla \theta$, with bulk thermal conductivity κ and temperature θ. The goal of this work is devoted to integrating the mechanism and data to model the non-Fourier's heat conduction. To begin with, we present the derivation of the model by using CDF. Specifically, we introduce two dissipative variables $\mathbf{w}$, $\mathbf{Q}$ whose counterpart is the conservative variable u; $\mathbf{w}$ is a vector which has the same dimension as $\mathbf{q}$, and $\mathbf{Q}$ is a symmetric tensor. The system is assumed to have the following entropy:

$$s = s(u, \mathbf{w}, \mathbf{Q}) = s^{eq}(u) + s^{\mathrm{neq}}(\mathbf{w}, \mathbf{Q}). \tag{2.1}$$

Here, s^{eq} and s^{neq} are both concave functions and correspond to the respective entropy equilibrium part and nonequilibrium part. Based on the generalized Gibbs relations [27,28], we deduce the evolution of the entropy,

$$\begin{aligned}
\partial_t s &= s_u \partial_t u + s_{\mathbf{w}} \cdot \partial_t \mathbf{w} + s_{\mathbf{Q}} : \partial_t \mathbf{Q} \\
&= -\theta^{-1} \nabla \cdot \mathbf{q} + \mathbf{q} \cdot \partial_t \mathbf{w} + s_{\mathbf{Q}} : \partial_t \mathbf{Q} \\
&= -\nabla \cdot (\theta^{-1} \mathbf{q} + \gamma s_{\mathbf{Q}} \cdot \mathbf{q}) \\
&\quad + \mathbf{q} \cdot (\nabla \theta^{-1} + \gamma \nabla \cdot s_{\mathbf{Q}} + \partial_t \mathbf{w}) + s_{\mathbf{Q}} : (\gamma \nabla \mathbf{q} + \partial_t \mathbf{Q}) \\
&= -\nabla \cdot (\theta^{-1} \mathbf{q} + \gamma s_{\mathbf{Q}} \cdot \mathbf{q}) + \mathbf{q} \cdot (\nabla \theta^{-1} + \gamma \nabla \cdot s_{\mathbf{Q}} + \partial_t \mathbf{w}) \\
&\quad + s_{\overset{\circ}{\mathbf{Q}}} : (\gamma \overset{\circ}{\nabla} \mathbf{q} + \partial_t \overset{\circ}{\mathbf{Q}}) + s_{\dot{Q}} : (\frac{\gamma}{N} \nabla \cdot \mathbf{q} + \partial_t \dot{Q}) \\
&=: -\nabla \cdot \mathbf{J} + \sigma.
\end{aligned} \tag{2.2}$$

Here, $\theta^{-1} := s_u$, and γ is a positive constant, and CDF suggests that $\mathbf{q} = s_{\mathbf{w}}$. The tensor is decomposed as $\mathbf{Q} = \overset{\circ}{\mathbf{Q}} + \dot{Q}\mathbf{I}$, with $\mathrm{tr}(\overset{\circ}{\mathbf{Q}}) = 0$, and $\overset{\circ}{\nabla}\mathbf{q}$ is symmetric where $\mathrm{tr}(\overset{\circ}{\nabla}\mathbf{q}) = 0$. $\mathbf{J} = \theta^{-1}\mathbf{q} + \gamma s_{\mathbf{Q}} \cdot \mathbf{q}$ is the entropy flux and $\sigma = \mathbf{q} \cdot (\nabla\theta^{-1} + \gamma \nabla \cdot s_{\mathbf{Q}} + \partial_t \mathbf{w}) + s_{\overset{\circ}{\mathbf{Q}}} : (\gamma \overset{\circ}{\nabla}\mathbf{q} + \partial_t \overset{\circ}{\mathbf{Q}}) + s_{\dot{Q}} : (\frac{\gamma}{N}\nabla \cdot \mathbf{q} + \partial_t \dot{Q})$ is the corresponding entropy production.

For simplicity, we only consider the quasi-one-dimensional case in this paper so that many of the vectors mentioned above can be approximated as scalars. Since the entropy flux σ is non-negative according to the second law of thermodynamics, we follow the CDF [29,30] to obtain that

$$\partial_t u + \partial_x q = 0, \tag{2.3a}$$

$$\partial_t w + \partial_x \theta^{-1} + \gamma \partial_x s_Q = M_0 q, \tag{2.3b}$$

$$\partial_t Q + \gamma \partial_x q = M_1 s_Q. \tag{2.3c}$$

Here, $M_0 = M_0(u, w, Q)$ and $M_1 = M_1(u, w, Q)$ are two positive functions since σ is non-negative according to the second law of thermodynamics.

B. Specific model and discrete version

A specific form of the entropy is considered in this paper, that is, the entropy of the nonequilibrium part has a quadratic form of the dissipative variables w, Q,

$$s^{\mathrm{neq}}(w, Q) = -\frac{w^2}{2\alpha} - \frac{Q^2}{2\beta}, \tag{2.4}$$

where β is a positive constant and $\alpha = \alpha(u)$ is positive. Thus, (2.3) becomes

$$\partial_t u + \partial_x q = 0, \tag{2.5a}$$

$$\partial_t(\alpha q) - \partial_x \theta^{-1} + \frac{\gamma}{\beta}\partial_x Q = -M_0 q, \tag{2.5b}$$

$$\partial_t Q + \gamma \partial_x q = -\frac{M_1}{\beta} Q. \tag{2.5c}$$

Here we have used that $w = -\alpha q$ (i.e., $q = s_w$) and thus $M_0 = M_0(u, q, Q)$ and $M_1 = M_1(u, q, Q)$. Noticing (2.1) and $\theta^{-1} := s_u = s_u^{eq}(u)$, we can assert that (2.5) is well defined with three unknowns: u, q, and Q.

Due to the discreteness of the training data (data exist only at discrete space-time points), an alternative discrete version

of (2.5) should be taken into account,

$$\begin{aligned}
&\frac{u_i^{n+1}-u_i^n}{\Delta t}+\frac{q_{i+1}^n-q_{i-1}^n}{2\Delta x}=0,\\
&(\alpha)_i^n\frac{(q)_i^{n+1}-(q)_i^n}{\Delta t}-\frac{(\theta^{-1})_{i+1}^n-(\theta^{-1})_{i-1}^n}{2\Delta x}\\
&\quad+\frac{\gamma}{\beta}\frac{(Q)_{i+1}^n-(Q)_{i-1}^n}{2\Delta x}=-(M_0q)_i^n,\\
&\frac{(Q)_i^{n+1}-(Q)_i^n}{\Delta t}+\gamma\frac{(q)_{i+1}^n-(q)_{i-1}^n}{2\Delta x}=-\frac{1}{\beta}(M_1Q)_i^n, \qquad (2.6)
\end{aligned}$$

where $(\cdot)_i^n$ denote the $(\cdot)$'s value at space-time points $(i\Delta x, n\Delta t)$, in which i and n are the indexes of the discretized cell and time step, respectively.

It is remarkable that (2.5) satisfies the conservation-dissipation principle and thereby the structure of the system is globally symmetrizable hyperbolic [30–32] and can be easily solved by traditional numerical methods. Further, we should point out that (2.5) is a generalized model since, on the one hand, it is only constrained by basic laws of physics, i.e., the first and second laws of thermodynamics, and, on the other hand, the model is independent of the data which could come from numerical results, experimental measurements, or real-world monitoring.

III. LEARN THE UNKNOWN FUNCTIONS AND PARAMETERS

In the last section, we derived the model of heat conduction with unknown functions and parameters in quasi-one-dimensional systems by using the CDF. In this section, we will show how to learn the unknown functions and parameters in (2.5) or (2.6) using deep neural networks, and propose an innovative method, i.e., the inner-step operation (ISO), to diminish the extra errors caused by the chosen discrete version. The current idea of macroscopic heat conduction modeling is the biggest innovation of the present paper, that is, first theoretically deriving a macroscopic heat conduction equation with unknown functions or parameters which is valid at any scale, and then using data-driven deep-learning methods to train and learn these unknown functions or parameters.

A. Training methods

We rewrite the last two equations of (2.6) in the following abstract form [24]:

$$\begin{aligned}
(\alpha)_i^n(q)_i^{n+1} &= (\alpha q)_i^n+\frac{\Delta t}{2\Delta x}\big[(\theta^{-1})_{i+1}^n-(\theta^{-1})_{i-1}^n\big]\\
&\quad-\frac{\gamma}{\beta}\frac{\Delta t}{2\Delta x}\big[(Q)_{i+1}^n-(Q)_{i-1}^n\big]-\Delta t(M_0q)_i^n\\
&=:\mathcal{S}[\alpha,\beta,\gamma,M_0]\big(V_{i-1}^n,V_i^n,V_{i+1}^n;\Delta t,\Delta x\big), \qquad (3.1a)\\
(Q)_i^{n+1} &= (Q)_i^n-\gamma\frac{\Delta t}{2\Delta x}\big[(q)_{i+1}^n-(q)_{i-1}^n\big]-\frac{\Delta t}{\beta}(M_1Q)_i^n\\
&=:\mathcal{S}[\beta,\gamma,M_1]\big(V_{i-1}^n,V_i^n,V_{i+1}^n;\Delta t,\Delta x\big). \qquad (3.1b)
\end{aligned}$$

Here, $V_i^n=((u)_i^n,(q)_i^n,(Q)_i^n)$, and $\mathcal{S}[\alpha,\beta,\gamma,M_0]$ denotes that $[\alpha,\beta,\gamma,M_0]$ need to be learned via machine learning, and similarly for $\mathcal{S}[\beta,\gamma,M_1]$. To be specific, $[\alpha,M_0,M_1]$ are approximated by the respective neural networks and $[\beta,\gamma]$ are optimized as parameters. For (3.1a) and (3.1b), the loss function is defined as

$$\begin{aligned}
L_1^1 &= \sum_{\text{training data}}\big|(\alpha)_i^n(q)_i^{n+1}-\mathcal{S}[\alpha,\beta,\gamma,M_0]\\
&\quad\times\big(V_{i-1}^n,V_i^n,V_{i+1}^n;\Delta t,\Delta x\big)\big|^2, \qquad (3.2a)\\
L_1^2 &= \sum_{\text{training data}}\big|(Q)_i^{n+1}-\mathcal{S}[\beta,\gamma,M_1]\\
&\quad\times\big(V_{i-1}^n,V_i^n,V_{i+1}^n;\Delta t,\Delta x\big)\big|^2, \qquad (3.2b)
\end{aligned}$$

where the mean squared error (MSE) is used. In addition, inspired by the "warm-up" technique [33,34] used in the training process, we design loss functions consisting of multistep time information,

$$\begin{aligned}
L_k^1 &= \sum_{\text{training data}}\big|(\alpha)_i^n(q)_i^{n+k}-\mathcal{S}[\alpha,\beta,\gamma,M_0]\\
&\quad\times\big(\tilde V_{i-1}^{n+k-1},\tilde V_i^{n+k-1},\tilde V_{i+1}^{n+k-1};\Delta t,\Delta x\big)\big|^2, \qquad (3.3a)\\
L_k^2 &= \sum_{\text{training data}}\big|(Q)_i^{n+k}-\mathcal{S}[\beta,\gamma,M_1]\\
&\quad\times\big(\tilde V_{i-1}^{n+k-1},\tilde V_i^{n+k-1},\tilde V_{i+1}^{n+k-1};\Delta t,\Delta x\big)\big|^2. \qquad (3.3b)
\end{aligned}$$

Here, $\tilde{}$ denotes the outputs of the neural networks, that is,

$$\begin{aligned}
\tilde V_i^{n+k-1} &= \big(\tilde u_i^{n+k-1},\tilde q_i^{n+k-1},\tilde Q_i^{n+k-1}\big), \qquad (3.4)\\
\tilde u_i^{n+k-1} &= \tilde u_i^{n+k-2}-\frac{\Delta t}{2\Delta x}\big(\tilde q_{i+1}^{n+k-2}-\tilde q_{i-1}^{n+k-2}\big), \qquad (3.5)\\
\tilde q_i^{n+k-1} &= \mathcal{S}[\alpha,\beta,\gamma,M_0]\\
&\quad\times\big(\tilde V_{i-1}^{n+k-2},\tilde V_i^{n+k-2},\tilde V_{i+1}^{n+k-2};\Delta t,\Delta x\big)\big/\alpha_i^{n-1}, \qquad (3.6)\\
\tilde Q_i^{n+k-1} &= \mathcal{S}[\beta,\gamma,M_1]\big(\tilde V_{i-1}^{n+k-2},\tilde V_i^{n+k-2},\tilde V_{i+1}^{n+k-2};\Delta t,\Delta x\big), \qquad (3.7)\\
\tilde V_i^1 &\equiv V_i^1. \qquad (3.8)
\end{aligned}$$

Additionally, we also define

$$L_k^3=\sum_{\text{training data}}\Big|u_i^{n+k}-\tilde u_i^{n+k-1}-\frac{\Delta t}{2\Delta x}\big(\tilde q_{i+1}^{n+k-1}-\tilde q_{i-1}^{n+k-1}\big)\Big|^2. \qquad (3.9)$$

Therefore, the total loss function is obtained by combining (3.3a), (3.3b), and (3.9),

$$L=\sum_{k=1}^{K}\big(\lambda_1L_k^1+\lambda_2L_k^2+\lambda_3L_k^3\big), \qquad (3.10)$$

where $\lambda_{1,2,3}$ are coefficients.

B. ISO

To diminish the extra errors caused by the selected discrete version, (2.6), we present a method, i.e., the inner-step operation (ISO),

$$(\alpha)_i^{n+\frac{k-1}{N}}(q)_i^{n+\frac{k}{N}} = (\alpha q)_i^{n+\frac{k-1}{N}} + \frac{\Delta t}{2N\Delta x}\left[(\theta^{-1})_{i+1}^{n+\frac{k-1}{N}} - (\theta^{-1})_{i-1}^{n+\frac{k-1}{N}}\right] - \frac{\gamma}{\beta}\frac{\Delta t}{2N\Delta x}\left[(Q)_{i+1}^{n+\frac{k-1}{N}} - (Q)_{i-1}^{n+\frac{k-1}{N}}\right] - \frac{\Delta t}{N}(M_0 q)_i^{n+\frac{k-1}{N}}, \tag{3.11a}$$

$$(Q)_i^{n+\frac{k}{N}} = (Q)_i^{n+\frac{k-1}{N}} - \gamma\frac{\Delta t}{2N\Delta x}\left[(q)_{i+1}^{n+\frac{k-1}{N}} - (q)_{i-1}^{n+\frac{k-1}{N}}\right] - \frac{\Delta t}{N\beta}(M_1 Q)_i^{n+\frac{k-1}{N}}, \tag{3.11b}$$

where N is the inner-steps number and $k = 1, \ldots, N$. Furthermore, by this method, Δt is divided into $\Delta t/N$, and thus it can reduce the dependency between the learned models and the data with an extremely small time step (Δt).

C. Training data

As mentioned before, the model is independent of the training data which could come from numerical results, experimental measurements, or real-world monitoring. However, limited by the difficulty of obtaining large amounts of real-world monitoring or experimental data for the present research group, the training data of the model in the present paper are obtained by numerically solving the phonon Boltzmann transport equation (BTE) under the Callaway approximation [11,35,36], which can describe the heat conduction in different regimes [6,37–40] and has a good agreement with the experimental results [7,41–43],

$$\frac{\partial e}{\partial t} + v_g \boldsymbol{s}\cdot\nabla_{\boldsymbol{x}} e = \frac{e_R^{eq} - e}{\tau_R} + \frac{e_N^{eq} - e}{\tau_N}. \tag{3.12}$$

Here, $e = e(\boldsymbol{x}, \boldsymbol{s}, t)$ is the phonon distribution function of the energy density depending on spatial position $\boldsymbol{x}$, unit directional vector $\boldsymbol{s}$, time t, and group velocity v_g. e_R^{eq} and e_N^{eq} are the equilibrium state of the momentum-conserved normal scattering process (N-process) and momentum-destroying resistive scattering process (R-process), respectively. τ_R and τ_N are the associated relaxation time, respectively.

In this work, the phonon gray model and linear phonon dispersion are used, and the wave vector in three-dimensional materials is assumed to be isotropic. The temperature θ is assumed that $|\theta - \theta_0| \ll \theta_0$ so that the equilibrium distribution function can be written as follows:

$$e_R^{eq}(\theta) \approx C\frac{\theta - \theta_0}{4\pi}, \quad e_N^{eq}(\theta, \mathbf{v}) \approx C\frac{\theta - \theta_0}{4\pi} + C\theta\frac{\boldsymbol{s}\cdot\mathbf{v}}{4\pi v_g},$$

where $C = C(\theta_0)$ is the specific heat at reference temperature θ_0, and $\mathbf{v}$ is the drift velocity. The temperature θ, heat flux $\mathbf{q}$, and the flux of heat flux $\boldsymbol{Q}$ can be calculated as the moments of distribution functions:

$$\theta = \theta_0 + \frac{\int e d\Omega}{C}, \quad \boldsymbol{q} = \int v_g \boldsymbol{s} e d\Omega, \quad \boldsymbol{Q} = \int v_g^2 \boldsymbol{s}\boldsymbol{s} e d\Omega,$$

where the integral is carried out in the whole solid angle space $d\Omega$. Here we intend to specify the heat conduction regimes which change with the dimensionless Knudsen numbers, defined as

$$Kn_R^{-1} = \frac{L}{v_g\tau_R}, \quad Kn_N^{-1} = \frac{L}{v_g\tau_N},$$

which indicates the strength of the R- and N-process, respectively, and L is the normalized spatial coordinates. Three distinct types of phonon transport are listed below and shown in Fig. 1:

(i) hydrodynamics: $Kn_R^{-1} \ll 1 \ll Kn_N^{-1}$;
(ii) ballistic: $Kn_R^{-1} \ll 1$ and $Kn_N^{-1} \ll 1$;
(iii) diffusive: $Kn_R^{-1} \gg 1$.

In addition, the Guyer-Krumhansl (GK) equation can be derived from Eq. (3.12) by the eigenvalue analysis method or Champman-Enskog method [9,12,44] in the phonon hydrodynamics regime with sufficient N-processes [7,11,37,40], which is a macroscopic heat conduction equation and plays the similar role as the Navier-Stokes equation in fluid hydrodynamics. The equations reads as

$$C\partial_t\theta + \nabla\cdot\mathbf{q} = 0,$$
$$\tau_R\partial_t\mathbf{q} + \mathbf{q} = -\kappa\nabla\theta + l^2[\nabla^2\mathbf{q} + 2\nabla(\nabla\cdot\mathbf{q})], \tag{3.13}$$

where the internal energy $u = C\theta$, $\kappa = (1/3)Cv_g^2\tau_R$ is the thermal conductivity, and $l^2 = (1/5)v_g^2\tau_R\tau_N$. In this paper, we use dimensionless parameters and set $C = v_g = 1$, $\tau_R = 0.1, 1.0, 10$, $\tau_N = 0.1, 1.0, 10$ for pairwise combination.

D. Testing

Once the unknown functions and parameters in the model (2.5) are trained well, Eq. (2.5) can be solved by traditional numerical methods, such as the finite-element method,

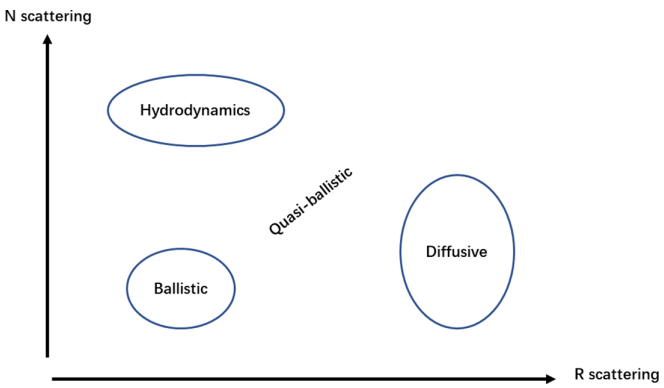

FIG. 1. A schematic of phonon transport regimes [6,7,12,44].

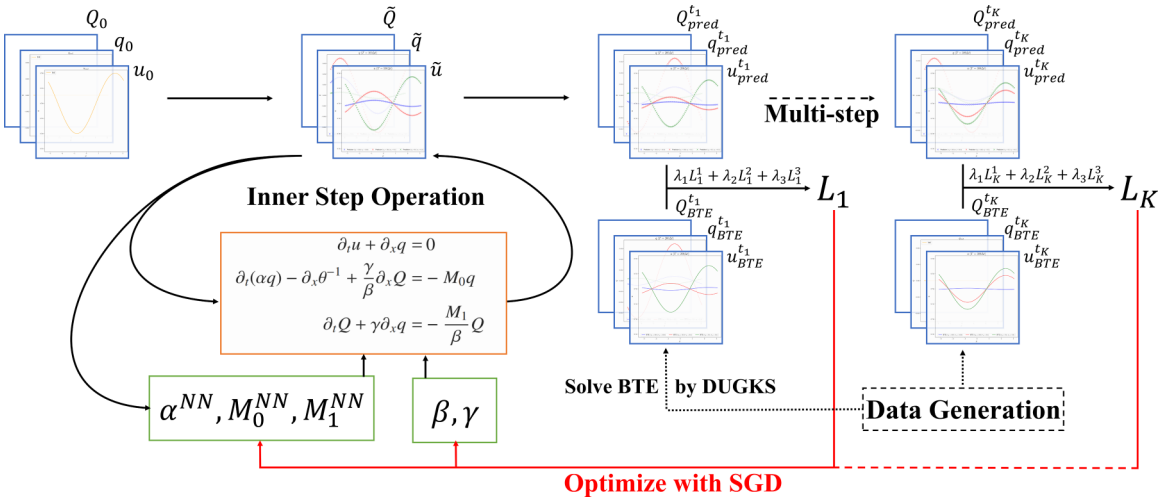

FIG. 2. The overview of the framework of the entire training process.

discontinuous Galerkin method, and so on. In this paper, we use the finite-difference method. The numerical results are validated by the solutions from (3.12), which are regarded as ground-truth data. Additionally, we also solve the GK model (3.13) numerically, in contrast to our model, for purposes of comparison with our model. We close this section by presenting the overview of the framework of the entire training process in Fig. 2.

IV. NUMERICAL RESULTS

To show the performance of the present model, the transient heat conduction in quasi-one-dimensional system is studied with different N or R scattering rates. Similar to the transient thermal grating [7,11] or time-domain thermoreflectance experiments [45], an ultrafast heat pulse is implemented on the materials at the initial moment, so that there is a spatial cosine temperature distribution or a hot spot at the center [36]. After the external heat source is removed, the temperature will propagate or dissipate along the x direction.

A. Generating training data and training setting

Generating training data. We solve the phonon BTE (3.12) by the discrete unified gas kinetic scheme (DUGKS) [11,46] in the domain $(x,t) \in [-\pi,\pi] \times [0,T]$ with periodic boundary conditions and the initial data constructed below. In this paper, we take $\Delta x = \frac{2\pi}{80}$ and $\Delta t = CFL \times \Delta x$ with $CFL = 0.5$. The data are saved every time step, $t = 0, \Delta t, 2\Delta t, \ldots, 600\Delta t$. The initial values are constructed as follows:

$$\begin{aligned} \theta(x,t=0) &= \alpha\theta_1(x) + (1-\alpha)\theta_2(x), \\ \theta_1(x) &= \theta_0 + \Delta\theta\cos(x - x_{j_1}), \\ \theta_2(x) &= \theta_0 + \Delta\theta\cos(x - x_{j_2}), \end{aligned} \tag{4.1}$$

with $x_{j_1} = -\pi + j_1 \times \Delta x$, $x_{j_2} = -\pi + j_2 \times \Delta x$, $\theta_0 = 0.6 : 0.1 : 1.0$, $\Delta\theta = 0.01 : 0.02 : 0.09$, and $\alpha = 0 : 0.2 : 0.8$.

Training setting. With respect to the Knudsen number pair, we set three neural networks to approximate α, M_0, M_1 in (3.1), respectively. All of the neural networks have four hidden layers with 50 neurons in each layer. In order to ensure the positivity of α, M_0, M_1, we take the *softplus* function as the activation function in the output player, while the *sin* function is used in other layers. We also use adaptive activation functions to accelerate the training [47]. The back propagation (BP) with the stochastic gradient descent (SGD) algorithm is selected as the optimizer and *CosineAnnealingLR* is used to adjust the learning rate *lr*. In the training process, the parameters that need to be optimized are initialized as $\beta = 10$, $\gamma = 1$. For the loss function, we take $K = 1$ to "warm up" the neural networks and, after a number of epochs, we fix $K = 4$.

Once the model is learned well, we solve the (2.5) by the finite-difference method [48].

B. Validation of accuracy, long-term stability, and generalization

Validation of accuracy and kong-term stability. We first show the high accuracy of our learned model (2.5) and also sketch the solutions of the GK model (3.13) for comparison. The initial values are set so that $\theta_0 = 0.75$, $\Delta\theta = 0.01$ in (4.1) and $q = 0$, which is not in the training data. Notice that although the training data are restricted to $[0, 600]\Delta t$, we solve the learned model up to $900\Delta t$.

The results with respect to diffusive ($\tau_R = 0.1, \tau_N = 10$), hydrodynamics ($\tau_R = 10, \tau_N = 0.1$), and ballistic ($\tau_R = 10, \tau_N = 10$) regimes are plotted in Fig. 3. Here the solutions of the BTE are taken as a benchmark, and the snapshots at $100\Delta t$, $300\Delta t$, and $900\Delta t$ are shown. It is easy to see that our model can predict the solutions with high accuracy, and performs better than those of the GK model.

Validation of generalization. The performance of the learned model with discontinuous initial values is explored in this numerical simulation. The results are demonstrated in Fig. 4, of which the first row is about the initial values of u and q. For comparison, we also plot the results of the GK model. We should point out again that the training data only consist of smooth cases. From Fig. 4, we can say that

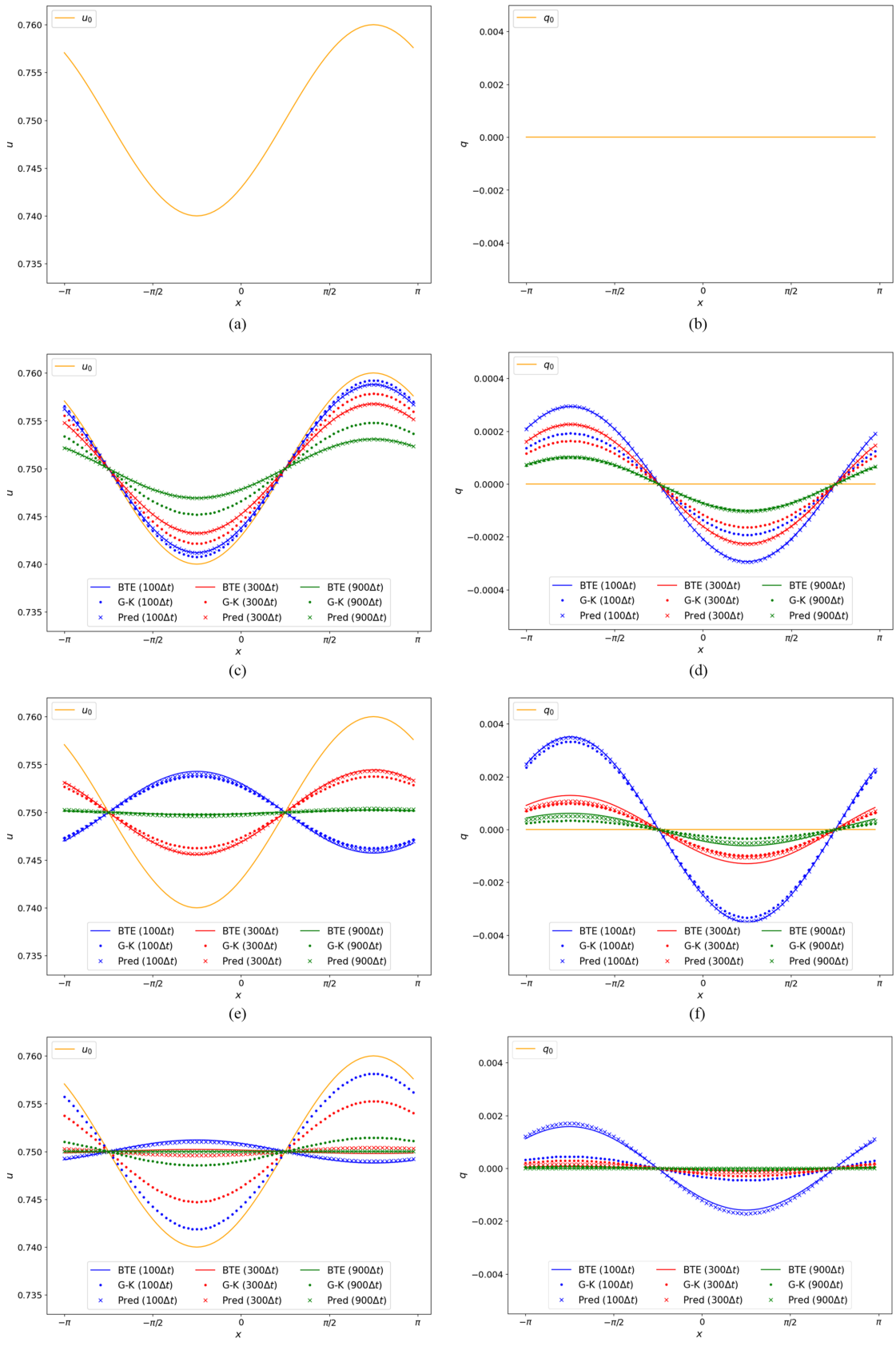

FIG. 3. The comparison of solutions of our model and the GK model at $100\Delta t$, $300\Delta t$, and $900\Delta t$. The baseline is the solutions of the BTE. (a), (b) The initial data about u and q, respectively. (c), (d) Diffusive $\tau_R = 0.1,\ \tau_N = 10$, (e), (f) hydrodynamics $\tau_R = 10,\ \tau_N = 0.1$, and (g),(h) ballistic $\tau_R = 10,\ \tau_N = 10$.

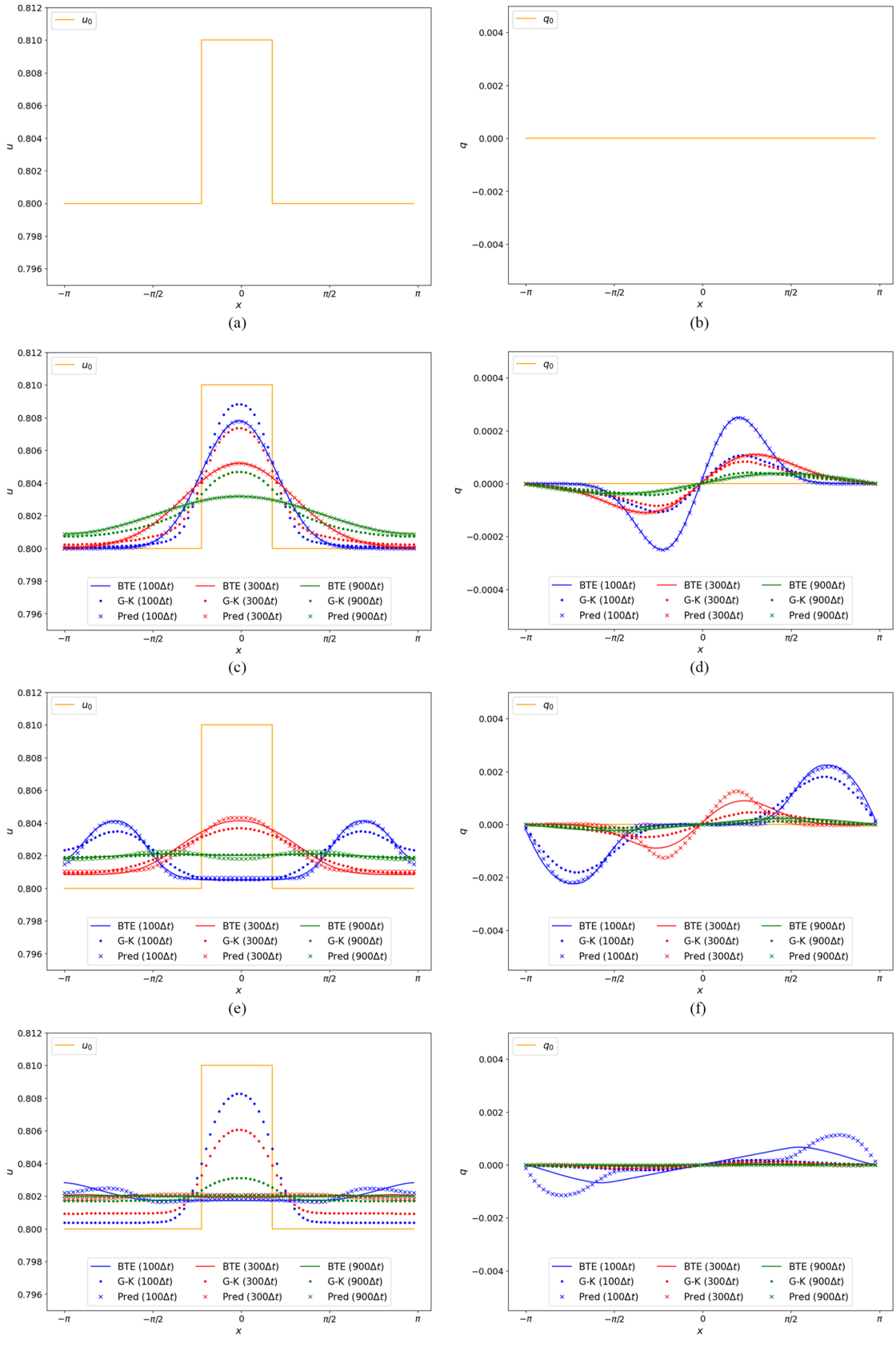

FIG. 4. The results of our model with discontinuous initial values at $100\Delta t$, $300\Delta t$, and $900\Delta t$. The results of the GK model are also plotted for comparison, and the baseline is the solutions of the BTE. (a), (b) The initial data about u and q, respectively. (c), (d) Diffusive $\tau_R = 0.1,\ \tau_N = 10$, (e), (f) hydrodynamics $\tau_R = 10,\ \tau_N = 0.1$, and (g), (h) ballistic $\tau_R = 10,\ \tau_N = 10$.

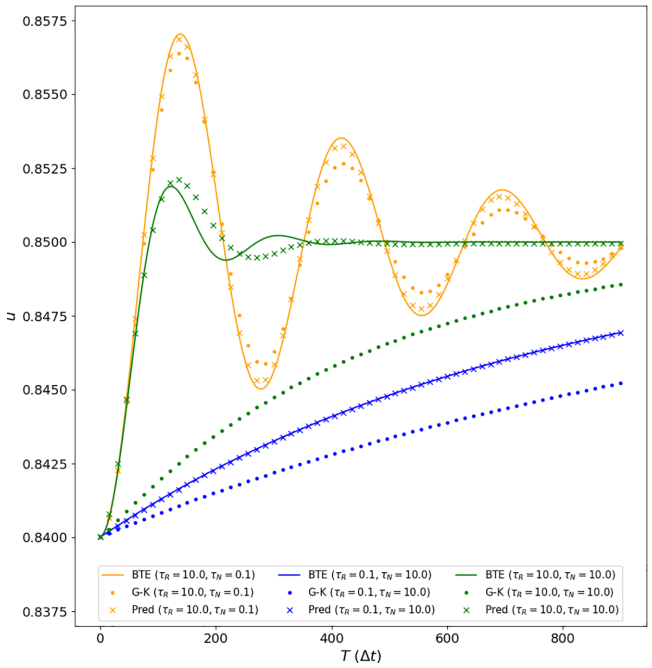

FIG. 5. The heat waves of $u(-\pi + 30\Delta t)$ with different Knudsen numbers. We take BTE as a baseline, and also show the behaviors of the GK model.

the results of our model agree with the exact ones very well, especially in the ballistic case, while the GK model disaccords with the BTE results. This implies that our model is valid with discontinuous initial data and has good generalization.

C. Heat waves

In this section, we test the behaviors of predicting the heat waves by the model with dual-dissipative variables, such as diffusive, hydrodynamic, and ballistic types, the last two of which cannot be captured by the model with one-dissipative variable [26]. The total results are drawn in Fig. 5. When $\tau_R = 0.1$, the heat conduction is in the diffusive regime, and the internal energy u decays over time; when $\tau_N = 0.1,\ \tau_R = 10$, the heat conduction is in the hydrodynamic regimes, and the internal energy u decays slowly with large heat waves; when $\tau_N = 10,\ \tau_R = 10$, the heat conduction is in the ballistic regimes, and the internal energy u decays quickly with small heat waves. From Fig. 5, it is easy to see that the results of our model are in consistent with those of the BTE.

D. ISO

In this section, we will demonstrate the power of the inner-step operation (ISO) in the mechanism-data fusion modeling process. The ISO is advanced and plays an important role in diminishing the extra errors caused by the alternative discrete version. For this purpose, we conduct the following numerical simulation: we only choose 60 points uniformly in $[0, T]$ to train the model, not the original 600 points, i.e., $t = 0, 10\Delta t, 20\Delta t, \ldots, 600\Delta t$. In the training process, we use the ISO with inner step $= 1,\ 10,\ 40$, and we also show the original results for comparison. Here we define the mean-L^2 relative error as

$$L = \frac{1}{N}\sum_{i=1}^{N} \frac{\sqrt{\sum_{j=1}^{80}(u - u_{BTE})^2(x_j, t_i)}}{\sqrt{\sum_{j=1}^{80} u_{BTE}^2(x_j, t_i)}},$$

$$N = 1, 2, \ldots, 900.$$

The corresponding results are demonstrated in Fig. 6, which shows that the model trained by coarse data with the ISO method can be competitive with that trained with fine data. Besides, the ISO method performs better with larger inner steps.

Moreover, we test the case in which the data consist of coarser time steps, $t = 0, 100\Delta t, 200\Delta t, \ldots, 600\Delta t$, i.e., six points in $[0, T]$. In this situation, the model cannot be trained well without the ISO method. However, with the ISO method, the model still has excellent performance. From Fig. 7, it is obvious that the ISO method plays a vital role in the model training process, and the prediction with the ISO method is even better than the original one.

V. ANALYSIS AND DISCUSSIONS

In this section, we summarize, analyze, and discuss the proposed model and training methods. Different from most previous macroscopic non-Fourier heat conduction models obtained by Chapman-Enskog, Hermite, or other small perturbation expansion methods, the heat conduction model built by the present framework is not limited by small Knudsen numbers. The current framework of macroscopic heat conduction modeling is the biggest innovation of the present paper; that is, first theoretically deriving a macroscopic heat conduction equation with unknown functions or parameters which is valid at any scale, and then using data-driven deep-learning methods to train and learn these unknown functions or parameters.

We advocate for a groundbreaking mechanism-data fusion method (MDFM), designed to model heat conduction using a dual-dissipative variables approach. The meteoric rise in computational processing capabilities and exponential growth in data storage capacities present opportunities for synergizing mechanism-based modeling with data-driven discoveries of nascent physical principles [49–54]. Advances in statistical methodologies coupled with machine learning's evolution have considerably enhanced the potency of data-driven models. The ascendancy of these computational tools, such as deep neural networks (DNNs), recurrent neural networks (RNNs), and convolutional neural networks (CNNs), have been notable. However, that most machine-learning-based models lack explicit expressions creates a barrier for the interpretability of the physical models owing to the reputation of being referred to as "black box."

We amalgamate machine learning methodologies with the conservation-dissipation formalism (CDF) [29,30], thus deriving an explicit and interpretable series of partial differential equations (PDEs) that elucidate both Fourier and non-Fourier modes of heat conduction. Utilizing the CDF, we introduce dual-dissipative variables to formulate a system characterized by first-order hyperbolic PDEs. Anchored in the second law of thermodynamics, the CDF ensures that the resultant PDE system adheres to the conservation-dissipation

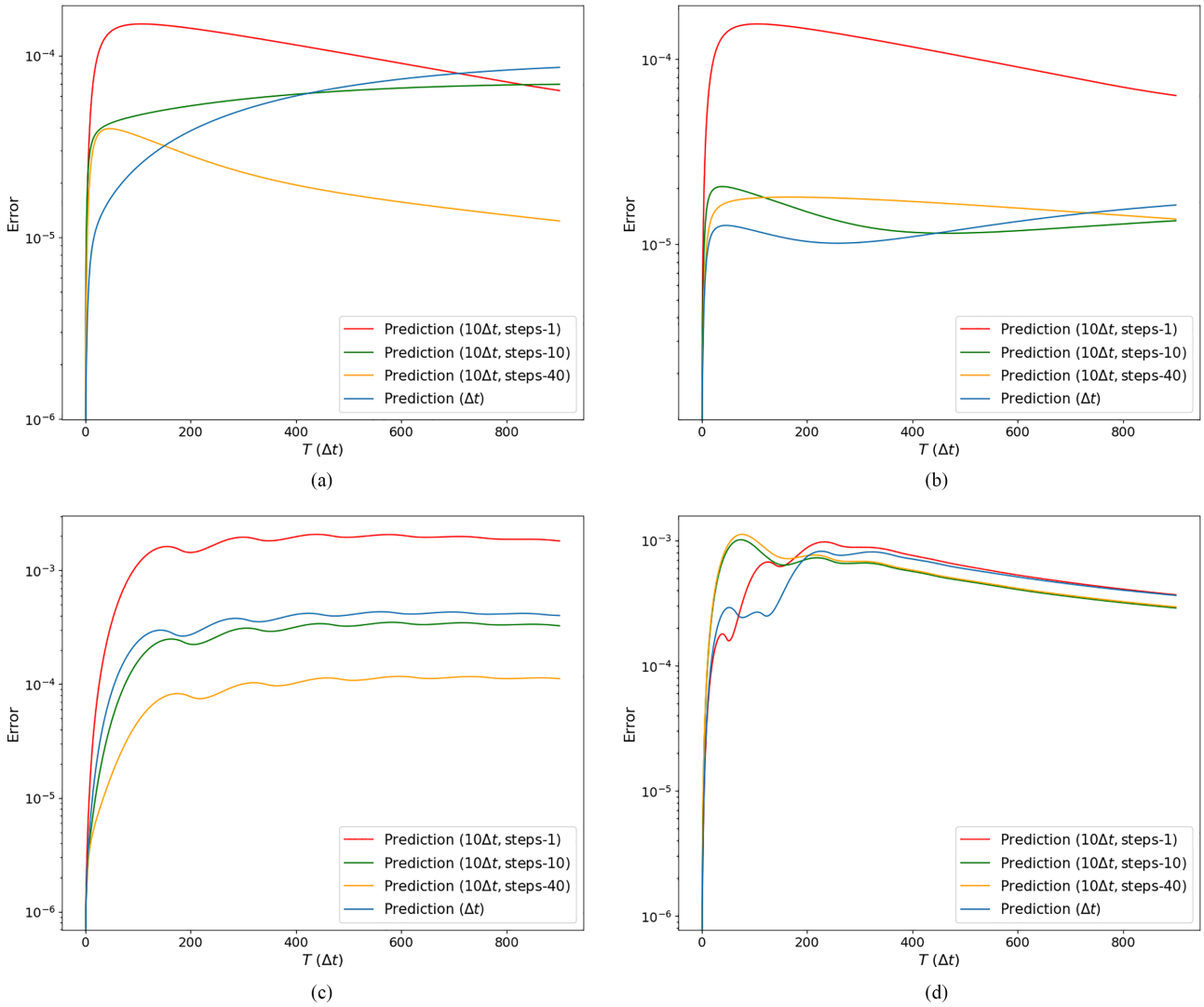

FIG. 6. The mean-L^2 relative errors of u with different Knudsen numbers. Here, step-$*$ is the ISO method with $*$ inner steps. Prediction (Δt) denotes that the model is trained by original data of every time step (Δt). (a) $\tau_R = 0.1,\ \tau_N = 0.1$, (b) $\tau_R = 0.1,\ \tau_N = 10$, (c) $\tau_R = 10,\ \tau_N = 0.1$, (d) $\tau_R = 10,\ \tau_N = 10$.

principle [31] and manifests as a globally symmetrizable hyperbolic structure [32]. Subsequent to this theoretical foundation, we employ deep neural networks (DNNs) to train the unknown functions within these PDEs, utilizing a preliminary "warm-up" procedure [33] to streamline the integration of temporal data series. Notably, while the data utilized for training in this study are derived from the phonon Boltzmann transport equation (BTE) [11], the training paradigm itself remains indifferent to the PDE derivation process, thereby accommodating data from a spectrum of sources, including numerical analyses, empirical simulations, or direct empirical observation.

In the training procedure, we propose the inner-step operation (ISO), an ingenious strategy engineered to bridge the gap between discrete formulations and the continuous paradigm inherent in system modeling. The discrete nature of our training data necessitates the adoption of an approximate discrete representation of the governing PDEs, which invariably introduces numerical discrepancies. Given DNN's insensitivity to higher-order discretizations, which paradoxically may induce instabilities, our ISO methodology seeks to attenuate the extraneous errors engendered by the choice of discretization schemes. Empirical evidence amassed through a plethora of numerical tests corroborates our assertion that the ISO methodology not only significantly mitigates these errors, but also lessens the model's sensitivity to data characterized by infinitesimal time-step magnitudes.

Many numerical tests are carried out to show the performance of the proposed model. It bears mentioning that the learned model is amenable to resolution via traditional numerical techniques, such as the finite-element method (FEM), finite-difference method (FDM), finite volume method (FVM), alongside more contemporaneous approaches such as the deep Ritz method (DRM) [55], deep Galerkin method (DGM) [56], and physics-informed neural networks (PINNs) [50]. An attribute of our endeavor is the model's proficiency in accurately capturing the thermal wave behavior characteristic of hydrodynamic and ballistic conduction regimes, which cannot be replicated in models confined to a one-dissipative variable framework [26]. Moreover, the

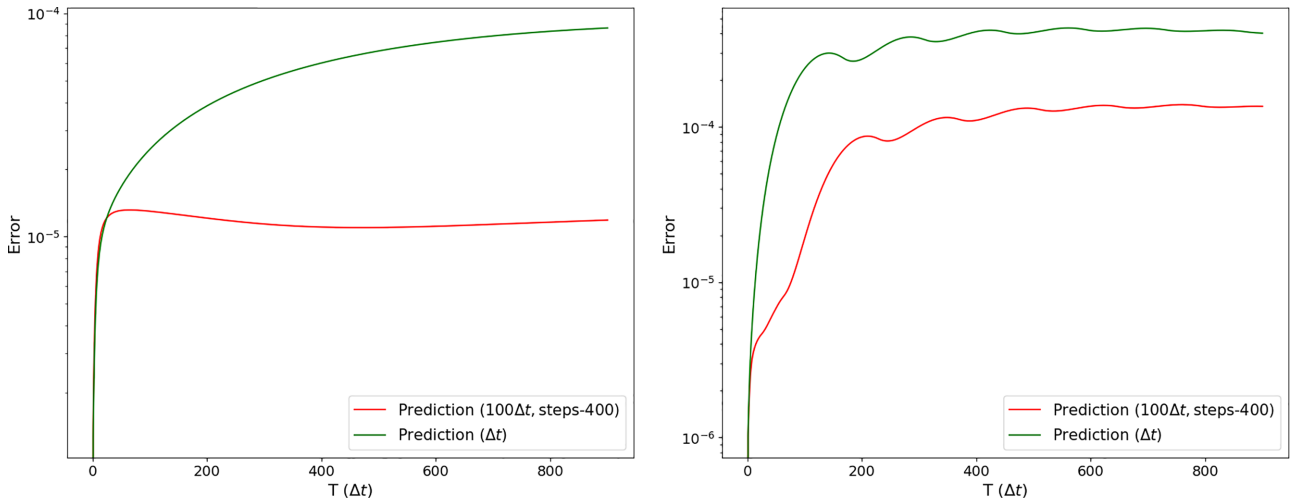

FIG. 7. The mean-L^2 relative errors of u with the ISO method. Here step-$*$ is the ISO method with $*$ inner steps. Prediction (Δt) denotes that the model is trained by original data of every time step (Δt). (a) $\tau_R = 0.1,\ \tau_N = 0.1$ and (b) $\tau_R = 10.0,\ \tau_N = 0.1$

model assures not only long-term stability, but also exhibits an enhanced accuracy over a broad spectrum of Knudsen numbers in comparison to the typical Guyer-Krumhansl (GK) model. The model's performance, particularly with discontinuous initial conditions despite being calibrated solely on smooth initial conditions, further underscores its versatility.

In addition, we should point out that there are many "schools" [57] for modeling heat conduction, such as classical irreversible thermodynamics (CIT) [27], extended irreversible thermodynamics (EIT) [28], and general equation for nonequilibrium reversible-irreversible coupling (GENERIC) [58]. The CDF allows more freedom, which can also be determined by the deep neural networks, and thus, we choose CDF as the modeling method in this work, which is clarified in our previous paper [26].

A. Coupled model

It is noticed that (2.3) is not unique, and we also investigate the coupled model as follows:

$$\begin{aligned} \partial_t u + \partial_x q &= 0, \\ \partial_t w + \partial_x \theta^{-1} + \gamma \partial_x s_Q &= M'_{11} q + M'_{12} s_Q, \\ \partial_t Q + \gamma \partial_x q &= M'_{21} q + M'_{22} s_Q. \end{aligned} \tag{5.1}$$

Here,

$$M' = M'(u, q, Q) = \begin{pmatrix} M'_{11} & M'_{12} \\ M'_{21} & M'_{22} \end{pmatrix}$$

is a positive matrix, and thus (5.1) satisfies the entropy law. Figure 8 shows the numerical results of (5.1) with the comparison of (2.5). It is concluded that the specific model (2.5) illustrates the ballistic and hydrodynamic non-Fourier heat conduction very well.

B. Other discrete versions

The proposed model is trained by the discrete version (2.6), which is based on a discrete time version, and thus some other discrete versions can be studied. The results of the original version, Lax-Wendroff scheme, Lax-Friedrichts scheme, implicit scheme, and the ISO with less data are plotted in Fig. 9. This shows that higher-order discrete versions seem not to make great improvements to the results. In addition, the training method of the continuous-time version is also used for (2.5), such as the physics-informed neural network (PINN) [50]. However, we do not achieve an acceptable result for the time being, and perhaps it is left for future investigations.

VI. CONCLUSIONS AND REMARKS

In this paper, we propose a mechanism-data fusion method (MDFM) for modeling heat conduction with dual-dissipative variables. This method inherits advantages of mathematical

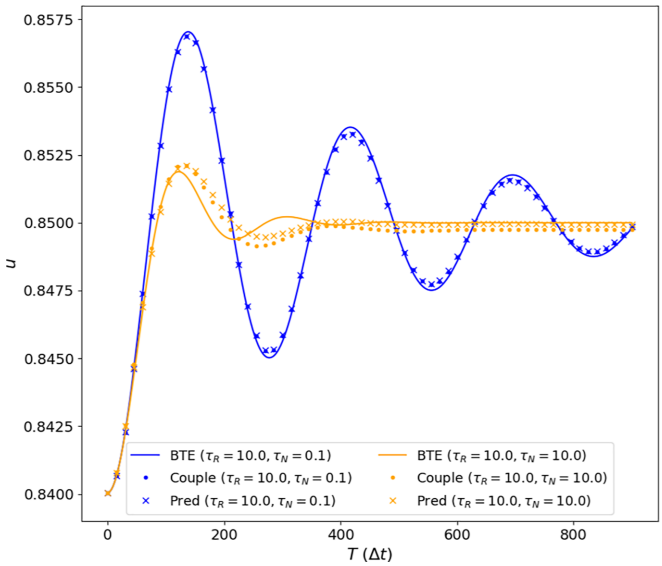

FIG. 8. The numerical results $u(-\pi + 30\Delta t)$ of (5.1) and (2.5) about $\tau_R = 10,\ \tau_N = 0.1$ and $\tau_R = 10,\ \tau_N = 10$.

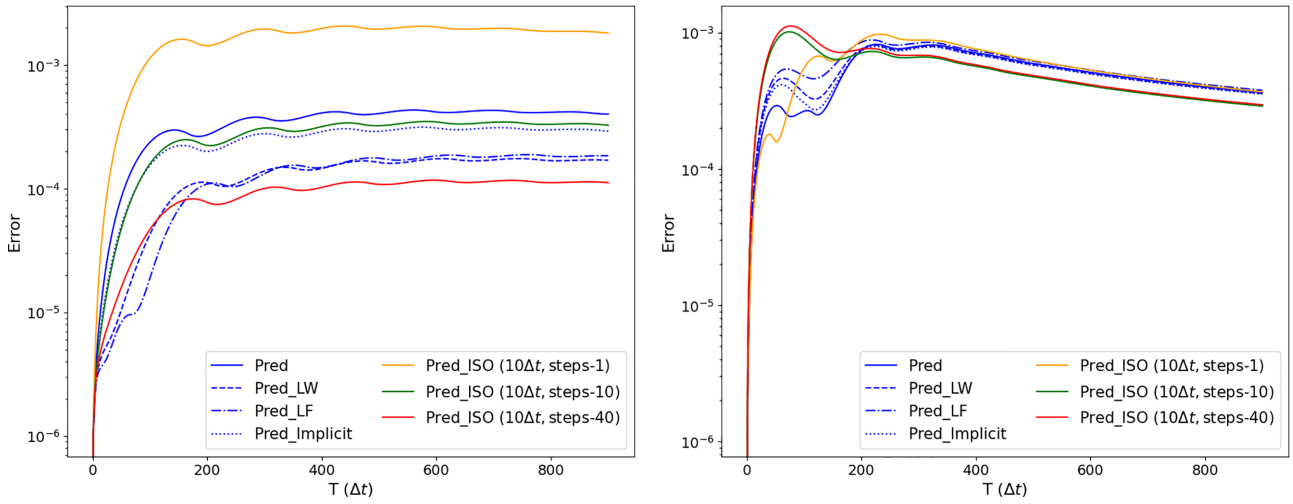

FIG. 9. The mean-L^2 relative errors of u with the original version, Lax-Wendroff (LW) scheme, Lax-Friedrichts (LF) scheme, implicit scheme, and the ISO with $10\Delta t$ about (a) $\tau_R = 10,\ \tau_N = 0.1$ and (b) $\tau_R = 10,\ \tau_N = 10$.

rigor and adaptable machine learning and, further, this model can be solved by conventional numerical methods directly. Specifically, we use the conservation-dissipation formalism (CDF) to derive an interpretable system of partial differential equations (PDEs) for heat conduction, which obeys the first and second laws of thermodynamics. The PDEs are macroscopic heat conduction equations for all phonon transport regimes, which are not limited by small Knudsen numbers. Next, we train the unknown functions in this PDE system with deep neural networks (DNNs); this involves a "warm-up" technique which prepares the connection of several time series. Moreover, we propose a method, i.e., the inner-step operation (ISO), to diminish the extra errors caused by the selected discrete version. Many numerical tests are conducted to show that the proposed model can well predict the heat conduction in diffusive, hydrodynamic, and ballistic regimes, the last two of which cannot be captured by the model with one-dissipative variable [26]. The model displays long-term stability and demonstrates higher accuracy under a wider range of Knudsen numbers than the Guyer-Krumhansl (GK) model.

In this paper, we select DNNs for our modeling due to their substantial computational capabilities and flexibility. Notably, one of the most advantageous features of DNNs is their capacity for transfer learning [59], which enables them to seamlessly adapt to new data. This adaptability makes them especially suitable for dynamic environments where data continuously evolve, thereby enhancing the model's performance and ensuring its sustained relevance. Future updates to our model will leverage this capacity to further refine its accuracy and expand its applicability across varying datasets. In addition, it is remarkable that due to simplicity and the strictness of the training data, we only consider the quasi-one-dimensional case in this current work. Actually, the derivation (2.2) is generic for multidimensional cases, but it does not contain the boundary condition which is necessary for the multidimensional cases. Investigation into modeling for heat conduction in multidimensions, including presenting a compatible boundary condition, is our ongoing and future work and, in turn, the result of the current work indicates that the MDFM is reasonable and powerful, which opens up a feasible and different way for the modeling method.

The codes used in this study have been deposited in the public Github without any restrictions (see Ref. [60]).

The data, including training data and testing data, have been deposited in the public netdisk without any restrictions (see Ref. [61]).

ACKNOWLEDGMENTS

We thank Y. Hu and B. Dong for useful discussions. This work is supported by the National Natural Science Foundation of China (Grants No. 12301520 and No. 12147122).

[1] G. Chen, *Nanoscale Energy Transport and Conversion: A Parallel Treatment of Electrons, Molecules, Phonons, and Photons* (Oxford University Press, Oxford, 2005).
[2] G. Chen, Nat. Rev. Phys. **3**, 555 (2021).
[3] D. G. Cahill, W. K. Ford, K. E. Goodson, G. D. Mahan, A. Majumdar, H. J. Maris, R. Merlin, and S. R. Phillpot, J. Appl. Phys. **93**, 793 (2003).
[4] 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.*, Appl. Phys. Rev. **1**, 011305 (2014).
[5] D. D. Joseph and L. Preziosi, Rev. Mod. Phys. **61**, 41 (1989).
[6] S. Lee, D. Broido, K. Esfarjani, and G. Chen, Nat. Commun. **6**, 6290 (2015).

[7] S. Huberman, R. A. Duncan, K. Chen, B. Song, V. Chiloyan, Z. Ding, A. A. Maznev, G. Chen, and K. A. Nelson, Science **364**, 375 (2019).

[8] K. Ghosh, A. Kusiak, and J.-L. Battaglia, J. Phys.: Condens. Matter **34**, 323001 (2022).

[9] Y. Guo and M. Wang, Phys. Rep. **595**, 1 (2015).

[10] P. Torres, A. Ziabari, A. Torelló, J. Bafaluy, J. Camacho, X. Cartoixà, A. Shakouri, and F. X. Alvarez, Phys. Rev. Mater. **2**, 076001 (2018).

[11] C. Zhang, S. Huberman, and L. Wu, J. Appl. Phys. **132**, 085103 (2022).

[12] R. A. Guyer and J. A. Krumhansl, Phys. Rev. **148**, 766 (1966).

[13] M. Simoncelli, N. Marzari, and A. Cepellotti, Phys. Rev. X **10**, 011019 (2020).

[14] Z. Guo, ES Energy. Environ. **1**, 4 (2018).

[15] C. Cattaneo, Atti Sem. Mat. Fis. Univ. Modena **3**, 83 (1948).

[16] M. E. Gurtin and A. C. Pipkin, Arch. Ration. Mech. Anal. **31**, 113 (1968).

[17] L. Sendra, A. Beardo, P. Torres, J. Bafaluy, F. X. Alvarez, and J. Camacho, Phys. Rev. B **103**, L140301 (2021).

[18] S. Li and B. Cao, AIP Adv. **10**, 105004 (2020).

[19] Y. Guo, D. Jou, and M. Wang, Phys. Rev. B **98**, 104304 (2018).

[20] M. Xu and X. Li, Intl. J. Heat Mass Transf. **55**, 1905 (2012).

[21] E. Weinan, Not. Am. Math. Soc. **68**, 565 (2021).

[22] W. E, J. Han, and L. Zhang, Phys. Today **74**(7), 36 (2021).

[23] J. Han, L. Zhang, and W. E, arXiv:2006.02619.

[24] J. Han, C. Ma, Z. Ma, and W. E, Proc. Natl. Acad. Sci. USA **116**, 21983 (2019).

[25] J. Huang, Z. Ma, Y. Zhou, and W.-A. Yong, J. Non-Equilib. Thermodyn. **46**, 355 (2021).

[26] J. Zhao, W. Zhao, Z. Ma, W.-A. Yong, and B. Dong, Intl. J. Heat Mass Transfer **185**, 122396 (2022).

[27] S. R. De Groot and P. Mazur, *Non-equilibrium Thermodynamics* (Courier Corporation, NY, USA, 2013).

[28] D. Jou, J. Casas-Vázquez, and G. Lebon, *Extended Irreversible Thermodynamics* (Springer, New York, 1996).

[29] Y. Zhu, L. Hong, Z. Yang, and W.-A. Yong, J. Non-Equilib. Thermodyn. **40**, 67 (2015).

[30] W.-A. Yong, Philos. Trans. R. Soc. A **378**, 20190177 (2020).

[31] W.-A. Yong, J. Math. Phys. **49**, 033503 (2008).

[32] X. Huo, Modeling and analysis of viscoelastic fluids, Ph.D. dissertation, Tsinghua University, 2017.

[33] Z. Long, Y. Lu, X. Ma, and B. Dong, in *Proceedings of the 35th International Conference on Machine Learning*, Proceedings of Machine Learning Research, edited by J. Dy and A. Krause (PMLR, 2018), Vol. 80, pp. 3208–3216.

[34] Z. Long, Y. Lu, and B. Dong, J. Comput. Phys. **399**, 108925 (2019).

[35] J. Callaway, Phys. Rev. **113**, 1046 (1959).

[36] C. Zhang and Z. Guo, Intl. J. Heat Mass Transfer **181**, 121847 (2021).

[37] X.-P. Luo, Y.-Y. Guo, M.-R. Wang, and H.-L. Yi, Phys. Rev. B **100**, 155401 (2019).

[38] C. Hua, L. Lindsay, X. Chen, and A. J. Minnich, Phys. Rev. B **100**, 085203 (2019).

[39] R. J. Hardy and D. L. Albers, Phys. Rev. B **10**, 3546 (1974).

[40] Z. Ding, J. Zhou, B. Song, V. Chiloyan, M. Li, T.-H. Liu, and G. Chen, Nano Lett. **18**, 638 (2018).

[41] X. Gu, Y. Wei, X. Yin, B. Li, and R. Yang, Rev. Mod. Phys. **90**, 041002 (2018).

[42] A. J. Minnich, J. A. Johnson, A. J. Schmidt, K. Esfarjani, M. S. Dresselhaus, K. A. Nelson, and G. Chen, Phys. Rev. Lett. **107**, 095901 (2011).

[43] A. J. Minnich, Phys. Rev. Lett. **109**, 205901 (2012).

[44] R. A. Guyer and J. A. Krumhansl, Phys. Rev. **148**, 778 (1966).

[45] P. Jiang, X. Qian, and R. Yang, J. Appl. Phys. **124**, 161103 (2018).

[46] Z. Guo and K. Xu, Int. J. Heat Mass Transfer **102**, 944 (2016).

[47] A. D. Jagtap, K. Kawaguchi, and G. E. Karniadakis, J. Comput. Phys. **404**, 109136 (2020).

[48] R. J. LeVeque, Conservative methods for nonlinear problems, in *Numerical Methods for Conservation Laws*, edited by O. E. Lanford (Birkhäuser Basel, Basel, 1990), pp. 122–135.

[49] S. L. Brunton, J. L. Proctor, and J. N. Kutz, Proc. Natl. Acad. Sci. USA **113**, 3932 (2016).

[50] M. Raissi, P. Perdikaris, and G. E. Karniadakis, J. Comput. Phys. **378**, 686 (2019).

[51] M. Edalatifar, M. B. Tavakoli, M. Ghalambaz, and F. Setoudeh, J. Therm. Anal. Calorim. **146**, 1435 (2021).

[52] K. Bao, W. Yao, X. Zhang, W. Peng, and Y. Li, Struct. Multidisc. Optim. **65**, 302 (2022).

[53] S. Manavi, T. Becker, and E. Fattahi, Intl. Commun. Heat Mass Transfer **142**, 106662 (2023).

[54] S. Durgam, A. Bhosale, V. Bhosale, R. Deshpande, P. Sutar, and S. Kamble, Heat Transfer **50**, 7817 (2021).

[55] W. E and B. Yu, Commun. Math. Stat. **6**, 1 (2018).

[56] J. Sirignano and K. Spiliopoulos, J. Comput. Phys. **375**, 1339 (2018).

[57] W. Muschik, Forsch. Ingenieurwes. **71**, 149 (2007).

[58] M. Pavelka, V. Klika, and M. Grmela, *Multiscale Thermodynamics: Introduction to GENERIC* (Walter de Gruyter GmbH & Co. KG, Boston, Berlin, 2018).

[59] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, in *Artificial Neural Networks and Machine Learning–ICANN 2018*, edited by V. Kůrková, Y. Manolopoulos, B. Hammer, L. Iliadis, and I. Maglogiannis (Springer International, Cham, 2018), pp. 270–279.

[60] Refer to the public Github site https://github.com/LehengChen/HeatModelMDFM.

[61] L. Chen, C. Zhang and J. Zhao, BTEData, Peking University public netdisk, 2024, https://disk.pku.edu.cn/link/AAF733D3A09571462A9D55C29127541EC6.