

Time-embedded convolutional neural networks for modeling plasma heat transport

Mufei Luo^{1,*}, Charles Heaton¹, Yizhen Wang¹, Daniel Plummer¹, Mila Fitzgerald¹, Francesco Miniati²,
Sam M. Vinko¹ and Gianluca Gregori¹

¹*Department of Physics, University of Oxford, Parks Road, Oxford OX1 3PU, United Kingdom*

²*Mach42, Robert Robinson Avenue, Oxford Science Park, Oxford OX4 4GP, United Kingdom*



(Received 3 September 2025; accepted 10 February 2026; published 9 March 2026)

We introduce a time-embedded convolutional neural network (TCNN) for modeling spatiotemporal heat transport in plasmas, particularly under strongly nonlocal conditions. In our earlier work, the Luciani-Mora-Virmont (LMV) Informed Neural Network (LINN) (Luo *et al.*, *Phys. Rev. Res.* **7**, L042017 (2025)) combined prior knowledge from the LMV model with kinetic Particle-in-Cell (PIC) data to improve kernel-based heat-flux predictions. While effective under moderately nonlocal conditions, LINN produced physically inconsistent kernels in strongly time-dependent regimes due to its reliance on the quasistationary LMV formulation. To overcome this limitation, TCNN is designed to capture the coupled evolution of both the normalized heat flux and the characteristic nonlocality parameter using a unified neural architecture informed by underlying physical principles. Trained on fully kinetic PIC simulations, TCNN accurately reproduces nonlocal dynamics across a broad range of collisionalities. Our results demonstrate that the combination of time modulation, coupled prediction, and convolutional depth significantly enhances predictive performance, offering a data-driven yet physically consistent framework for multiscale plasma transport problems.

DOI: [10.1103/xx8k-xw91](https://doi.org/10.1103/xx8k-xw91)

I. INTRODUCTION

Accurate modeling of heat transport [1,2] in plasmas is fundamental to a variety of high-energy-density applications, including inertial confinement fusion [3] and astrophysical phenomena [4,5], and further in understanding how small-scale kinetic instabilities can impact heat fluxes [6,7]. In these systems, the degree of nonlocality is often characterized by the ratio between the electron mean free path, $\lambda_{\text{free}} = 4\pi\epsilon_0^2 m_e^2 v_{th}^4 / (Zn_e e^4 \ln \Lambda)$ [8], and the electron temperature gradient length, $L_T = T_e / |\nabla T_e|$. Here, T_e is the electron temperature, ϵ_0 is the vacuum permittivity, m_e is the electron mass, $v_{th} = \sqrt{T_e/m_e}$ is the electron thermal velocity, Z is the ion charge state, n_e is the electron number density, e is the elementary charge, and $\ln \Lambda$ is the Coulomb logarithm.

When $\lambda_{\text{free}}/L_T \ll 0.002$, electron heat conduction can be reliably described by classical, local models such as the Spitzer-Härm (SH) theory [9]. However, as this ratio increases, nonlocal effects emerge, leading to the breakdown of traditional hydrodynamic closures [10]. In response, several semiempirical models have been proposed, including the Luciani-Mora-Virmont (LMV) [11] and Schurtz-Nicolai-Busquet (SNB) [12] formulations, which aim to extend local theory by introducing an empirical nonlocal kernel [13,14]. In recent years, diffusion-based formulations have been re-

considered [15]. However, such approaches typically rely on the assumption of temporal stationarity [16]. Alternatively, empirical flux-limited diffusion schemes are often used in large-scale hydrodynamic modeling [17].

Lamy *et al.* [18] constructed databases from CHIC hydrodynamic simulations using the classical SNB [12] model, which restricted their neural network to reproducing an existing analytic model rather than learning directly from kinetic data. In contrast, our earlier work [19] introduced the Luciani-Mora-Virmont (LMV) informed neural network (LINN), a data-driven framework that integrates prior knowledge from the LMV model to learn nonlocal transport kernels. Trained on fully kinetic particle-in-cell (PIC) simulation data, LINN successfully captures both the nonlocal kernel structure and the associated heat flux under moderately nonlocal conditions and, therefore, improves the kernel-based heat flux model. However, as the degree of nonlocality increases, the initial temperature profile undergoes significant temporal evolution. Although LINN maintains accurate predictions of the heat flux in such regimes, it yields physically inconsistent kernels. This limitation arises from the model's adherence to the LMV formulation, which assumes quasistationary temperature profiles and therefore does not generalize to strongly time-dependent scenarios.

To address the challenges of modeling dynamic, strongly nonlocal heat transport, this study proposes a time-embedded convolutional neural network (TCNN) architecture. In contrast to LINN, which focuses on predicting the kernel and its associated heat flux, TCNN is designed to capture the spatiotemporal evolution of both the heat flux and the characteristic ratio $\lambda_{\text{free}}/L_T$ using a unified set of learnable parameters. This formulation reflects the underlying coupling between the heat flux and characteristic ratio, providing a

*Contact author: mufei.luo@physics.ox.ac.uk

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

more coherent and physically consistent modeling framework. At the same time, TCNN retains the foundational principle of nonlocal transport inspired by the LMV model, namely, that the heat flux at a given location is influenced by contributions from all spatial points within the system.

The remainder of this paper is organized as follows. Section II details the generation of kinetic data using PIC simulations, with an emphasis on distinguishing the characteristics of the data under local and nonlocal transport regimes. The architectural design of the TCNN is presented in Sec. III, followed by an in-depth evaluation of its predictive performance and comparisons with several architectural variants in Sec. IV. Finally, Sec. V offers concluding remarks and discusses the broader implications of this work.

II. KINETIC DATA GENERATION

To investigate heat transport across different regimes, we follow the same approach used previously [19] and employ one-dimensional (1D) PIC simulations using the OSIRIS code [20], equipped with physically validated binary collision models [21,22]. The system is initialized with a Maxwellian velocity distribution, and the spatial temperature profile is defined as follows:

$$T_e(x) = \frac{2T_{e0}(R-1)}{(R+1)[1+\exp(x/L)]} + \frac{2T_{e0}}{R+1}. \quad (1)$$

We set the central temperature to $T_{e0} = 1$ keV, a value characteristic of conduction-zone plasmas in laser-fusion targets, where typical electron temperatures lie between ~ 50 eV and 2 keV [23]. $R = T_h/T_c$ represents the hot-to-cold temperature ratio, and L defines the transition width between hot and cold regions. This profile establishes a temperature gradient in the box center, with a high-temperature region on the left and a low-temperature region on the right. The corresponding electron density is set to follow $n_e(x) \propto 1/T_e(x)$, approximating an isobaric condition characteristic of laser-heated conduction zones [23]. To vary the degrees of nonlocality, we adjust the background electron density n_{e0} , effectively modifying the mean free path λ_{free} relative to the temperature gradient length $L_T = 2L(R+1)/(R-1)$. This approach enables a controlled transition from classical transport behavior to regimes dominated by strong nonlocal effects.

We consider a fully ionized plasma composed of high- Z ions, specifically with $Z = 16$ to enhance collisionality. The Coulomb logarithm, $\ln \Lambda$, is computed at the center of the simulation domain and held constant within each run. We employ normalized units based on the plasma frequency $\omega_{\text{pe}} = (n_{e0}e^2/\epsilon_0 m_e)^{1/2}$ and wave number $k_{\text{pe}} = \omega_{\text{pe}}/c$, with c the light velocity in vacuum. The box length is fixed at $L_{\text{box}} \approx 1500$, with spatial resolution $\Delta x = 0.04$ and time step $\Delta t = 0.035$, satisfying the Courant-Friedrichs-Lewy condition ($\Delta t/\Delta x \approx 0.9$) and resolving the Debye length ($\lambda_{\text{De},0} = 1.2\Delta x$). Each cell contains 25 000 electron and 1600 ion macroparticles.

Figure 1 illustrates the initial electron temperature profiles used to control the nonlocality. The light-blue, blue, and dark-blue curves correspond to a fixed temperature ratio of $R_1 = 2$, with varying transition widths of $L = 16$, $L = 24$, and $L = 32$, respectively. These values yield corresponding

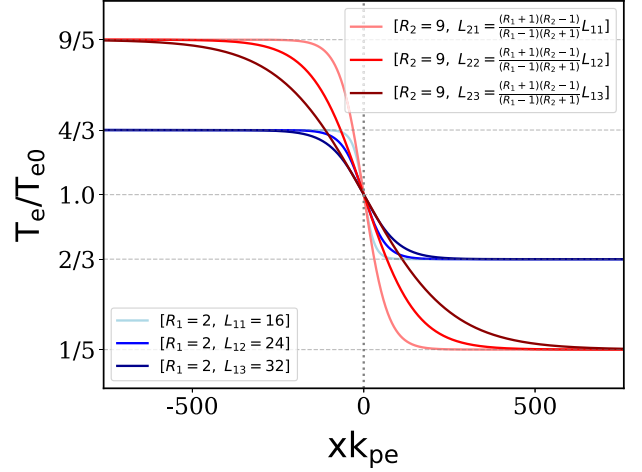


FIG. 1. Initial electron temperature profiles used to explore the effects of nonlocality and high-energy electrons on heat transport. The light-blue, blue, and dark-blue curves correspond to a fixed temperature ratio of $R_1 = 2$ and increasing transition widths $L = 16$, 24, and 32, resulting in temperature gradient lengths of $L_T = 96$, 144, and 192 at the center of the domain. To explore the influence of hot electrons while maintaining a constant gradient length, the temperature ratio R is increased from 2 to 9, and the transition width L is tuned according to $L_T = 2L(R+1)/(R-1)$, shown by the light-red, red, and dark-red curves.

temperature gradient lengths at the center of the domain of $L_T = 96$, $L_T = 144$, and $L_T = 192$. The electron mean free path λ_{free} is changed by adjusting the background electron density n_{e0} , thereby modifying the characteristic ratio $\lambda_{\text{free}}/L_T$. To examine the influence of hotter electrons on heat transport, we increase the temperature ratio $R = T_h/T_c$ while maintaining a constant temperature gradient length L_T at the center of the domain. This is achieved by appropriately adjusting the transition width L in accordance with the relation $L_T = 2L(R+1)/(R-1)$. As shown by the light-red, red, and dark-red curves in Fig. 1, these profiles preserve the same gradient length as their corresponding $R = 2$ curves at the domain center but exhibit progressively higher temperatures on the hot side. Accordingly, the temperature ratio R is varied from 2 to 9 to systematically assess the impact of increasingly energetic electrons on heat transport. The background electron density is varied across different simulations but is restricted to values not exceeding 10^{25} cm^{-3} to ensure that the plasma remains in a nondegenerate regime.

Figure 2 presents two representative cases illustrating heat transport behavior in the local and nonlocal regimes. Both scenarios are initialized with a temperature ratio of $R = 5$ and a temperature gradient length of $L_T = 192$. In the case shown in Figs. 2(a) and 2(b), the central electron density is set to $n_{e0} = 10^{25} \text{ cm}^{-3}$, corresponding to a characteristic ratio of $\lambda_{\text{free}}/L_T \approx 0.0004$. This places the system well within the collisional (local) transport regime. Figure 2(a) displays the evolution of the electron temperature (left y axis) at three distinct normalized times, $t/t_{\text{sa}} = 0$, $1/4$, and 1 , represented by the solid light-blue, blue, and dark-blue curves, respectively. t_{sa} denotes the saturation time specific to each case. The corresponding heat fluxes (right y axis) at $t/t_{\text{sa}} = 1/4$

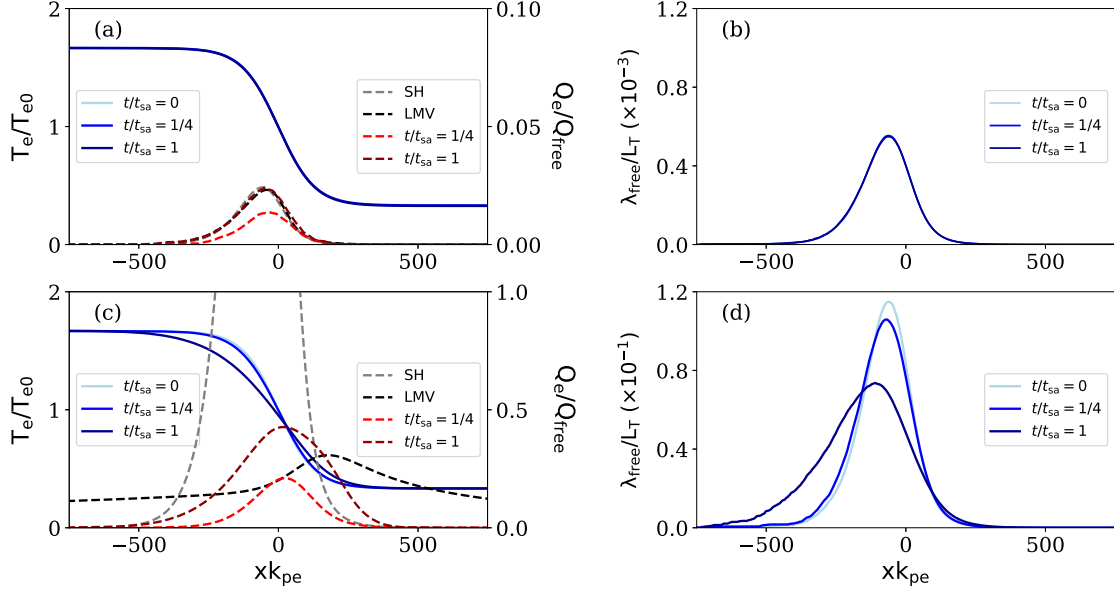


FIG. 2. Heat transport in the local and strongly nonlocal regimes, both initialized with a temperature ratio of $R = 5$ and gradient length $L_T = 192$. (a) For the local case with central density $n_{e0} = 10^{25} \text{ cm}^{-3}$ ($\lambda_{\text{free}}/L_T \approx 0.0004$), temperature profiles are shown at normalized times $t/t_{\text{sa}} = 0, 1/4$, and 1 (solid light-blue, blue, and dark-blue curves, left y axis), along with corresponding heat fluxes at $t/t_{\text{sa}} = 1/4$ and 1 (dashed red and dark-red curves, right y axis). Classical Spitzer-Härm (SH) and nonlocal LMV model predictions are shown for comparison (dashed gray and black curves). The saturated heat flux remains consistent with theoretical models and induces negligible temperature evolution. (b) The characteristic ratio $\lambda_{\text{free}}/L_T$ remains temporally invariant, as the normalized heat flux, Q_e/Q_{free} , is sufficiently small to induce negligible thermal evolution. (c) In the strongly nonlocal case ($n_{e0} = 10^{19} \text{ cm}^{-3}$, $\lambda_{\text{free}}/L_T \approx 0.084$), PIC simulations reveal significant suppression and preheating in the heat flux relative to SH predictions. LMV fails to capture the observed transport behavior. Strong heat flux leads to temporal evolution of the temperature profile and substantial variation in $\lambda_{\text{free}}/L_T$ shown in (d).

and 1 are indicated by the dashed red and dark-red curves. For comparison, theoretical predictions based on the SH and LMV models are overlaid using dashed gray and black curves, respectively.

The classical local heat transport prediction of SH flux [9] is given by $Q_{\text{SH}}/Q_{\text{free}} = a \lambda_{\text{SH}}/L_T$, where $a = 128(Z + 0.24)/[3\pi(Z + 4.2)]$ and $\lambda_{\text{SH}} = 3\lambda_{\text{free}}/\sqrt{\pi/2}$ is the classical Spitzer-Härm mean free path. The nonlocal LMV model is expressed as [11]

$$Q_e(x) = \int Q_{\text{SH}}(x') \mathcal{W}(x, x') dx', \quad (2)$$

where the kernel function $\mathcal{W}(x, x')$ quantifies the contribution of the local SH flux at position x' to the net flux at position x , thereby encapsulating nonlocal transport effects. It is defined as

$$\mathcal{W}(x, x') = \frac{1}{2\lambda(x')} \exp\left[-\frac{X(x, x')}{\lambda(x')}\right], \quad (3)$$

where X represents a density-weighted spatial separation between x and x' , computed as

$$X = \frac{1}{n_e(x')} \left| \int_x^{x'} n_e(x'') dx'' \right|, \quad (4)$$

and

$$\lambda(x') \simeq 30 \sqrt{Z + 1} \lambda_{\text{free}}(x'). \quad (5)$$

As shown in Fig. 2(a), the saturated heat flux exhibits strong agreement with both the SH and LMV model predictions. The temperature profile remains effectively stationary

throughout the simulation, as the normalized heat flux, Q_e/Q_{free} , is sufficiently small to induce negligible thermal evolution. Consequently, the characteristic ratio $\lambda_{\text{free}}/L_T$ remains stationary throughout the space, as shown in Fig. 2(b).

In contrast, Fig. 2(c) presents a scenario with a much lower central electron density, $n_{e0} = 10^{19} \text{ cm}^{-3}$, resulting in a characteristic ratio of $\lambda_{\text{free}}/L_T \approx 0.084$. This firmly places the system within the strongly nonlocal regime. Under these conditions, the heat flux obtained from the PIC simulation shows a marked suppression and evidence of preheating when compared to the classical SH prediction. Furthermore, the LMV model fails to accurately capture the heat flux dynamics in this regime. Owing to the relatively large normalized heat flux, Q_e/Q_{free} , the temperature profile evolves significantly over time. This dynamic behavior also induces substantial variation in the characteristic ratio $\lambda_{\text{free}}/L_T$, as illustrated in Fig. 2(d). Thus the onset of strong nonlocality introduces pronounced temporal dynamics into the system, conditions under which our previous model, LINN, demonstrates limited effectiveness. It is worth noting that the time-independent heat flux formulation proposed by Epperlein and Short [13] exhibits a higher degree of localization compared to the LMV model, particularly within the strongly nonlocal regime; however, it nevertheless underestimates the heat flux relative to kinetic simulations.

To facilitate direct comparison with previous studies on heat transport [21,24,25], Fig. 3 presents the normalized heat flux, Q_e/Q_{free} , plotted against the inverse characteristic ratio, $L_T/\lambda_{\text{free}}$. All cases shown share a fixed temperature ratio of $R = 5$ and a central temperature gradient length of $L_T = 192$,

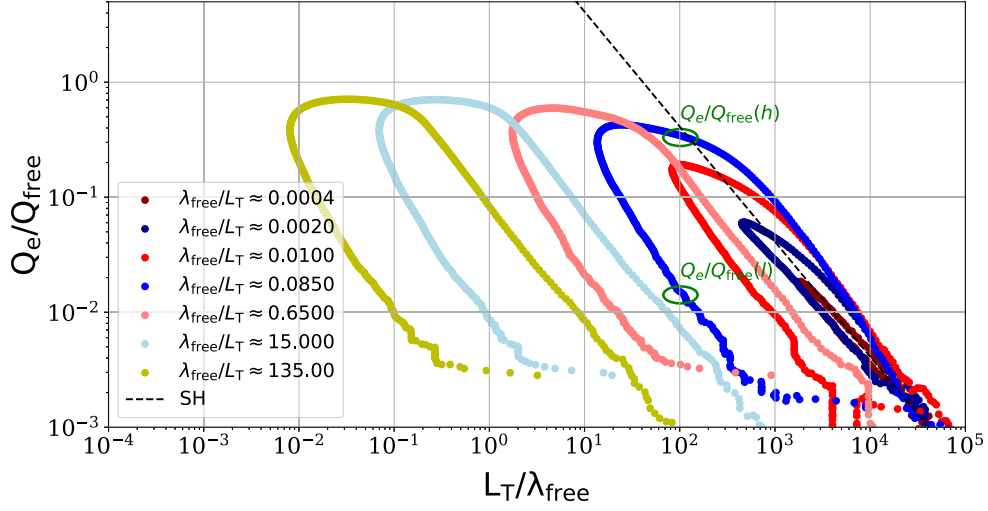


FIG. 3. Normalized heat flux Q_e/Q_{free} as a function of the inverse characteristic ratio $L_T/\lambda_{\text{free}}$. All cases are initialized with a fixed temperature ratio $R = 5$ and a central temperature gradient length $L_T = 192$, while varying the background electron density. Data points correspond to central densities of $n_{e0} = 10^{25}$ (dark red), 10^{23} (dark blue), 10^{21} (red), 10^{19} (blue), 10^{17} (light red), 10^{14} (light blue), and 10^{12} cm^{-3} (yellow), yielding characteristic ratios of $\lambda_{\text{free}}/L_T \approx 0.0004, 0.002, 0.01, 0.085, 0.65, 15$, and 135 , respectively. The dashed black line shows the classical SH prediction.

while differing in background electron density. The plotted data points correspond to central densities of $n_{e0} = 10^{25}$ (dark red), 10^{23} (dark blue), 10^{21} (red), 10^{19} (blue), 10^{17} (light red), 10^{14} (light blue), and 10^{12} cm^{-3} (yellow). These density options correspond to the characteristic ratio $\lambda_{\text{free}}/L_T \approx 0.0004$ (dark red), 0.002 (dark blue), 0.01 (red), 0.085 (blue), 0.65 (light red), 15 (light blue), and 135 (yellow). All data are taken after the system has reached saturation, although the saturation time differs among the examined cases. The dashed black line plots the SH flux.

Within each case examined in Fig. 3, for a given value of $L_T/\lambda_{\text{free}}$, two distinct values of the normalized heat flux Q_e/Q_{free} are identified: a higher value, $Q_e/Q_{\text{free}}(h)$, representing the heat flux in the downstream, and a lower value, $Q_e/Q_{\text{free}}(l)$, corresponding to the upstream one. This is clearly illustrated in Figs. 2(c) and 2(d). In particular, panel (d) shows that the same value of $\lambda_{\text{free}}/L_T$ can correspond to two different spatial positions, and these two positions can exhibit different heat fluxes, as shown in Fig. 2(c). The position associated with the lower heat flux is referred to as upstream, while the one with the higher heat flux is referred to as downstream. As a representative example, two green ellipses are marked to indicate $Q_e/Q_{\text{free}}(h)$ and $Q_e/Q_{\text{free}}(l)$ for $L_T/\lambda_{\text{free}} = 100$ within the blue scatter data. In the case of $\lambda_{\text{free}}/L_T \approx 0.0004$, the system across the space resides in the local transport regime. Both $Q_e/Q_{\text{free}}(h)$ and $Q_e/Q_{\text{free}}(l)$ closely align with the classical SH theory. However, for $\lambda_{\text{free}}/L_T \approx 0.002$, the local condition begins to break down on the hot side of the temperature gradient, where electrons become sufficiently energetic to exhibit nonlocal behavior. These nonlocal electrons propagate toward the colder region, resulting in enhanced heat flux in the downstream region; $Q_e/Q_{\text{free}}(h)$ exceeds SH predictions, while the upstream flux $Q_e/Q_{\text{free}}(l)$ is diminished. This asymmetry intensifies with increasing nonlocality. As $\lambda_{\text{free}}/L_T$ grows, the peak heat flux begins to fall below SH levels, indicating flux suppression due to nonlocal transport.

Furthermore, once $\lambda_{\text{free}}/L_T$ exceeds approximately 0.1 , the system transitions into a collisionless regime within the duration of the simulation ($T\omega_{pe} = 2500$), and no saturation state is reached. Consequently, the maximum achievable heat flux plateaus and shows no further enhancement despite increasing nonlocality. For this reason, the subsequent studies are confined to cases where the characteristic ratio satisfies $\lambda_{\text{free}}/L_T \leq 0.1$, ensuring that the system remains in a regime where meaningful saturation of the heat flux can be observed. Correspondingly, the considered range of central electron densities spans from 10^{19} to 10^{25} cm^{-3} .

III. TIME-EMBEDDED CONVOLUTIONAL NEURAL NETWORKS ARCHITECTURE

In our previous work [19], the LINN model was designed to incorporate as much theoretical structure from the LMV model as possible. Specifically, the elements of the nonlocal kernel $\mathcal{W}$ [Eq. (3)] were generated by mapping structured physical inputs—namely, X [Eq. (4)] and λ [Eq. (5)]—through the neural network architecture of LINN. The resulting kernel was then used to compute the nonlocal heat flux by convolving the classical SH flux Q_{SH} over space, as prescribed by Eq. (2). The SH flux itself was evaluated using the initial value of the characteristic ratio $\lambda_{\text{free}}/L_T$, maintaining consistency with the LMV-based nonlocal transport formulation.

However, as demonstrated in Figs. 2(c) and 2(d), under strongly nonlocal conditions, the normalized heat flux Q_e/Q_{free} becomes sufficiently large to drive a substantial evolution of the temperature profile. Consequently, the characteristic ratio $\lambda_{\text{free}}/L_T$ also undergoes temporal variation. In such dynamic scenarios, relying on the initial value of $\lambda_{\text{free}}/L_T$ to compute the nonlocal heat flux yields physically inconsistent kernels, though it can still maintain accurate predictions of the heat flux. From a physical perspective, the intrinsic

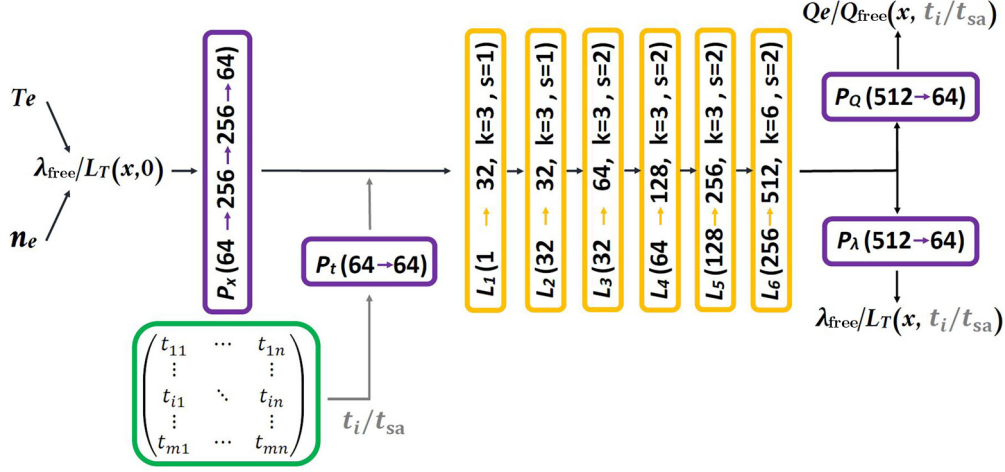


FIG. 4. Architecture of the time-embedded convolutional neural network (TCNN). The model predicts spatiotemporal quantities, characteristic ratio $\lambda_{\text{free}}/L_T(x, t)$, and normalized heat flux $Q_e/Q_{\text{free}}(x, t)$, conditioned on the spatial profile $\lambda_{\text{free}}/L_T(x, 0)$ derived from electron temperature T_e and density n_e . The spatial input is projected via a fully connected encoder P_x , while the time index is embedded and projected through P_t . These are combined and passed through a sequence of six 1D convolutional layers $L_1 - L_6$ with progressively increasing channels. Two projection heads P_λ and P_Q decode the latent representation into physically meaningful outputs. The time matrix (the element is normalized by the saturation time t_{sa}), highlighted within the green box, has dimensions (n, m) , where $n = 64$ and $m = 44$ denote the number of spatial and temporal steps, respectively.

coupling between the heat flux and the parameter $\lambda_{\text{free}}/L_T$ suggests that both quantities should be predicted simultaneously using a unified set of parameters within a neural network framework. To achieve this objective, we introduce a time-embedded convolutional neural network (TCNN) framework, presented in the following.

In contrast to LINN, which strictly adheres to the formalism of Eq. (2), the TCNN architecture relaxes this constraint to enable the discovery of alternative, data-driven representations of nonlocal heat transport. Nonetheless, it preserves the core physical principle that the heat flux at any given location is influenced by contributions from the entire spatial domain. Specifically, TCNN is designed such that the predicted heat flux at a point incorporates information about the spatial distribution of the parameter $\lambda_{\text{free}}/L_T$. We rewrite Eq. (2) as

$$Q_e(x, t)/Q_{\text{free},0} = \int \tilde{a}(x, x', t) \lambda_{\text{free}}(x')/L_T(x') dx'. \quad (6)$$

Here, $Q_{\text{free},0}$ denotes the free-streaming heat flux evaluated at the center of the domain, while the temporal information on λ_{free} and L_T has been implicitly encoded within the learned $\tilde{a}$.

Inspired by Eq. (6), several NN modeling strategies can be proposed in order of complexity:

- (1) Operating directly in real space by applying a single convolutional layer to convolve parameter $\lambda_{\text{free}}/L_T$ across the domain;
- (2) Lifting the input vector ($\lambda_{\text{free}}/L_T$) into a latent space, and applying a single convolutional layer to convolve the lifted vector across the latent domain;
- (3) Operating directly in real space, but applying multiple convolutional layers to convolve parameter $\lambda_{\text{free}}/L_T$ layer by layer across the domain;
- (4) Lifting the input vector ($\lambda_{\text{free}}/L_T$) into a latent space, before applying multiple convolutional layers to convolve the lifted vector across the latent domain.

Furthermore, in any of the above strategies, the temporal information must be introduced to capture the dynamics of heat transport. The first strategy directly adheres to the formulation of Eq. (6). However, in the context of nonlocal transport, the inherent nonlocal characteristics of the system could be more effectively captured within one latent space [26], which the model needs to learn. Furthermore, the use of multiple convolutional layers can enhance the model's ability to capture nonlocal and nonlinear features of the system. This aspect will be further elucidated in the following. Please note that once the nonlocal operator is learned through multiple convolutional layers acting on $\lambda_{\text{free}}/L_T$, rather than as an explicit linear kernel, it does not admit a direct interpretation as a physical LMV-type kernel.

Figure 4 shows the resulting TCNN architecture that captures both spatial correlations and temporal dependencies through explicit time embedding. The model operates on the initial profile of the characteristic ratio $\lambda_{\text{free}}/L_T$, which is discretized into a 64-point grid. It is conditioned on a discrete time index $t_i/t_{sa} \in \{1/44, 2/44, \dots, 1\}$, corresponding to one of 44 time steps (see the time matrix within the green box; the elements of this time matrix are trainable).

A learnable embedding layer encodes the discrete time step t_i by delivering the i th row of the time matrix to another trainable vector in $\mathbb{R}^{64}$ (see projection labeled “ P_t ”). This is projected through a fully connected (FC) network into a 64-dimensional latent space to match the spatial feature embedding, allowing time-aware conditioning of the model.

The spatial input vector $\lambda_{\text{free}}/L_T \in \mathbb{R}^{64}$ is first passed through an FC network (see projection labeled “ P_x ”):

$$\lambda_{\text{free}}/L_T \rightarrow \text{FC}(64 \rightarrow 256) \rightarrow \text{GELU} \rightarrow \text{FC}(256 \rightarrow 64)$$

to embed it into a latent feature representation. The time-conditioned embedding is added to this output to yield a time-aware spatial representation.

Subsequently, the resulting feature is processed through a series of six 1D convolutional layers to extract hierarchical spatial features, shown by the yellow boxes in Fig. 4. GELU activations are applied after each layer.

The final latent representation is passed through two separate linear heads to predict two target quantities (see projections labeled “ P_λ ” and “ P_Q ”):

- (i) $\lambda_{\text{free}}/L_T(x, t_i)$: the evolved characteristic ratio at time t_i ;
- (ii) $Q_e/Q_{\text{free}}(x, t_i)$: normalized heat flux at time t_i .

As illustrated in the flowchart of Fig. 4, the core mechanism for learning the system’s dynamics involves embedding temporal information by introducing distinct time vectors into the initial conditions. Conditioned on the initial characteristic ratio profile, $\lambda_{\text{free}}/L_T$, the introduced time vector serves to temporally modulate the convolutional layer parameters. With the exception of the final two FC layers (P_Q and P_λ), the model employs a largely shared set of parameters to jointly predict the evolution of both $\lambda_{\text{free}}/L_T$ and the normalized heat flux Q_e/Q_{free} —thereby preserving the inherent physical coupling between nonlocality and heat transport. Following the convolutional layers, the spatial domain is progressively reduced such that all spatial points are aggregated into a single point, while the feature dimension expands into a latent representation of 512 channels. This architectural design is consistent with the nonlocal LMV model, wherein the heat flux at any given location receives contributions from the entire spatial domain.

In addition to the model architecture, several strategies are required to address the imbalance in data distribution. As illustrated in Figs. 2 and 3, the values of $\lambda_{\text{free}}/L_T$ span approximately two orders of magnitude across the transition from local to strongly nonlocal regimes, while the normalized heat flux Q_e/Q_{free} varies by about one order of magnitude. Notably, $\lambda_{\text{free}}/L_T$ remains relatively stationary in the local regime but exhibits pronounced temporal dynamics under strongly nonlocal conditions. These disparities in scale and behavior necessitate tailored techniques to ensure stable and effective learning across regimes.

To address the challenges posed by uneven data distribution across local and nonlocal transport regimes, we introduce a weighted loss function defined as

$$\mathcal{L}_{\lambda, Q} = \left\langle \frac{1}{N_b} \sum_{i=1}^{N_b} \frac{(y_i - \hat{y}_i)^2}{(\max(y_i) + \varepsilon)^{p_1}} \right\rangle_x,$$

where y and $\hat{y}$ denote the ground truth and predicted values, respectively. $\mathcal{L}_\lambda$ corresponds to the loss for the characteristic ratio $\lambda_{\text{free}}/L_T$, and $\mathcal{L}_Q$ for the normalized heat flux Q_e/Q_{free} . The exponent p_1 adjusts the scaling of the loss to reflect the magnitude differences across regimes (e.g., $p_1 = 1.0$ for $\lambda_{\text{free}}/L_T$, and $p_1 = 0.75$ for Q_e/Q_{free}). The constant $\varepsilon = 10^{-8}$ prevents division by zero, and $N_b = 16$ is the patch size. Spatial averaging is applied to compute the loss, represented by the operator $\langle \rangle_x$.

To further encourage accurate modeling of temporal dynamics, we incorporate a trend-consistency loss that penalizes deviations in the first-order differences between the predicted

and ground truth sequences:

$$\mathcal{L}_{\text{trend}, (\lambda, Q)} = \left\langle \frac{1}{N_b} \sum_{t=t_2}^{t_{\text{sa}}} (\Delta \hat{y}_t - \Delta y_t)^2 \right\rangle_{x, t},$$

$$\Delta y_t = y_t - y_{t-1}, \quad \Delta \hat{y}_t = \hat{y}_t - \hat{y}_{t-1}.$$

This term is averaged over both spatial and temporal dimensions.

Additionally, to balance variations in the rate of change across regimes, we introduce a differential weighting loss based on the magnitude of changes in the ground truth:

$$\mathcal{L}_{\text{diff}, (\lambda, Q)} = \left\langle \frac{1}{N_b} \sum_{t=t_2}^{t_{\text{sa}}} \frac{(\hat{y}_t - \hat{y}_{t-1})^2}{(|y_t - y_{t-1}| + \varepsilon)^{p_2}} \right\rangle_{x, t},$$

with $p_2 = 0.8$ for both $\lambda_{\text{free}}/L_T$ and Q_e/Q_{free} , and averaging over space and time.

Combining these components, the total loss function for a given sequence is

$$\begin{aligned} \mathcal{L} = & \frac{1}{m} \sum_{t=t_1}^{t_{\text{sa}}} (0.5 \mathcal{L}_\lambda + 0.5 \mathcal{L}_Q) + \mathcal{L}_{\text{trend}, \lambda} + \mathcal{L}_{\text{trend}, Q} \\ & + 0.05 (\mathcal{L}_{\text{diff}, \lambda} + \mathcal{L}_{\text{diff}, Q}), \end{aligned} \quad (7)$$

where m denotes the total number of time steps. The model is trained using the Adam optimizer with an initial learning rate of 0.001. The ReduceLROnPlateau strategy monitors the loss in min mode and reduces the learning rate by a factor of 0.8 if no improvement is observed for five consecutive epochs.

IV. PERFORMANCE OF TCNN

We utilize the same dataset as in our previous study [19], partitioned into training, validation, and test subsets in a 70:15:15 ratio. To reduce numerical noise and limit the complexity of the neural network, the PIC results comprising 38 400 cells are coarse grained to 64 sampling points, while still preserving the heat-flux profile. The training process is depicted in Fig. 5(a), which shows the evolution of training and validation losses, calculated via Eq. (7), for the TCNN model. Both curves exhibit smooth convergence toward a plateau, indicating stable optimization and no evidence of overfitting.

Figure 5(b) assesses the model’s predictive performance on the test set without being weighted. Red markers represent the relative deviation between the predicted and ground-truth heat flux values, while blue markers correspond to the deviation in the characteristic ratio, $\lambda_{\text{free}}/L_T$. These deviations are quantified as follows:

$$\bar{\sigma}_Q = \frac{\sqrt{\langle (Q_{\text{PIC}} - Q_{\text{NN}})^2 \rangle_{x, t}}}{Q_{\text{PIC}, \max}},$$

and

$$\bar{\sigma}_\lambda = \frac{\sqrt{\langle \left(\left(\frac{\lambda_{\text{free}}}{L_T} \right)_{\text{PIC}} - \left(\frac{\lambda_{\text{free}}}{L_T} \right)_{\text{NN}} \right)^2 \rangle_{x, t}}}{\left(\frac{\lambda_{\text{free}}}{L_T} \right)_{\text{PIC}, \max}}.$$

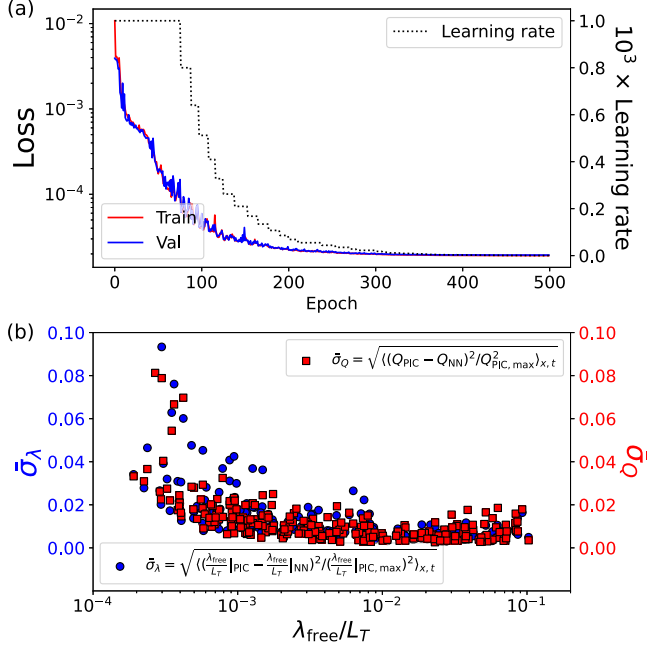


FIG. 5. Training and evaluation performance of the TCNN model shown in Fig. 4. (a) Training and validation loss curves exhibit smooth convergence, indicating stable optimization without signs of overfitting. (b) Relative prediction errors on the test set, normalized by the maximum values of Q_e/Q_{free} and $\lambda_{\text{free}}/L_T$ for each sample, where red markers represent the deviation in normalized heat flux Q_e/Q_{free} and blue markers indicate the deviation in the characteristic ratio $\lambda_{\text{free}}/L_T$. The horizontal axis denotes the initial value of $\lambda_{\text{free}}/L_T$ in the box center for each test case.

Here, $Q_{\text{PIC, max}}$ and $\lambda_{\text{free}}/L_T|_{\text{PIC, max}}$ denote the maximum values of the respective PIC-based quantities for each sample.

The horizontal axis in Fig. 5(b) represents the value of $\lambda_{\text{free}}/L_T$ at the center of each test sample. Results indicate that TCNN performs well in capturing the system's spatiotemporal evolution under nonlocal conditions. However, its predictive accuracy tends to degrade in the local regime. This degradation arises because, under highly collisional conditions, the system evolves slowly, and $\lambda_{\text{free}}/L_T$ remain nearly stationary, in particular. Consequently, the time-embedded neural architecture receives weak temporal signals, limiting its ability to update and adapt effectively in these near-static regimes, but it can still reproduce the SH behavior reliably.

The performance of four additional architectural variants, discussed following Eq. (6), is summarized in Fig. 6. Figure 6(a) illustrates a baseline model in which convolution is performed directly in real space without applying a spatial projection layer (P_x). A single convolutional layer with a kernel size of 64 is employed, allowing full spatial convolution. This setup maps the input from one channel to 5210 output features, yielding a comparable parameter count (332 800 parameters in the convolutional layer, and 655 488 in the output projections) to the full TCNN architecture (919 648 parameters in the convolutional layer, and 65 664 in the output projections) shown in Fig. 4. In Fig. 6(b), the architecture builds on (a) by incorporating a simple spatial projection layer $P_x = \text{FC}(64 \rightarrow 64)$, effectively transforming the input into a

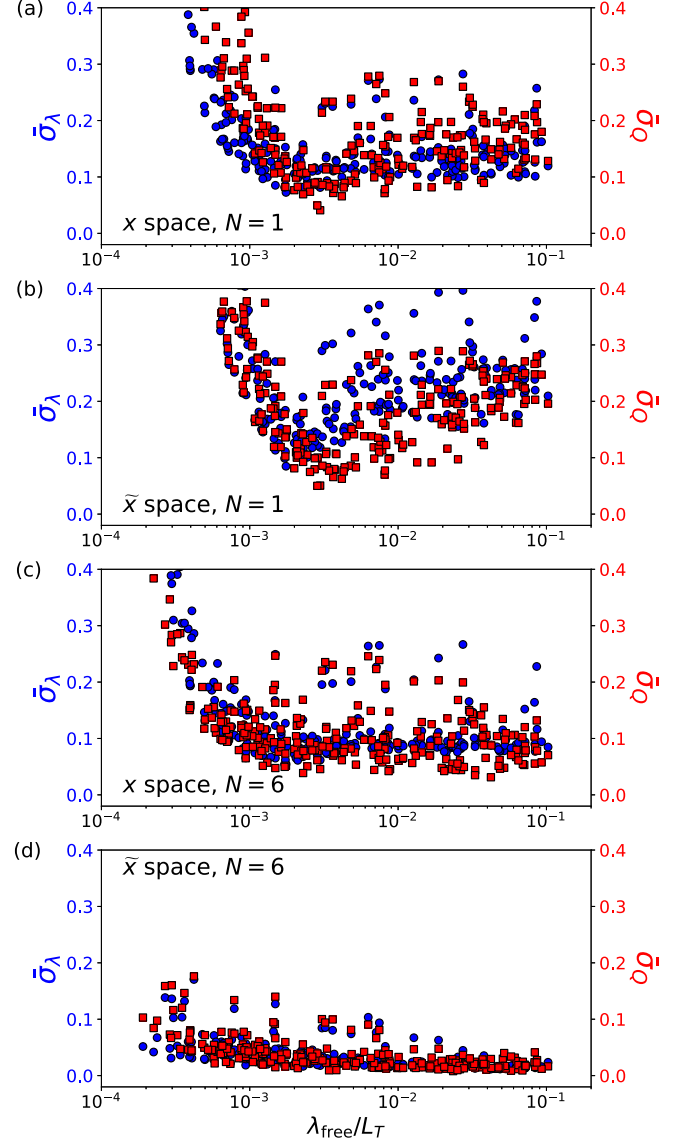


FIG. 6. Performance comparison of four architectural variants related to the TCNN framework. (a) A baseline model performing convolution directly in real space without spatial projection, using a single convolutional layer with a kernel size of 64. (b) Similar to (a), but includes a lightweight spatial projection layer $P_x = \text{FC}(64 \rightarrow 64)$, transforming inputs into a latent space $\tilde{x}$. (c) A deeper network with six convolutional layers operating in real space. (d) Combining both a lightweight spatial projection layer and six convolutional layers.

latent spatial framework $\tilde{x}$ but maintaining lower expressive capacity than the projection structure in Fig. 4. Figure 6(c) reverts to real space, but extends the convolutional depth to six convolutional layers, identical in depth and structure to those used in the TCNN model depicted in Fig. 4. Finally, Fig. 6(d) combines both architectural enhancements: a lightweight projection layer $P_x = \text{FC}(64 \rightarrow 64)$ and a six-layer convolutional stack.

The architectural variants depicted in Figs. 6(a) and 6(b) exhibit poor performance due to insufficient convolutional depth and limited spatial transformation capabilities, which

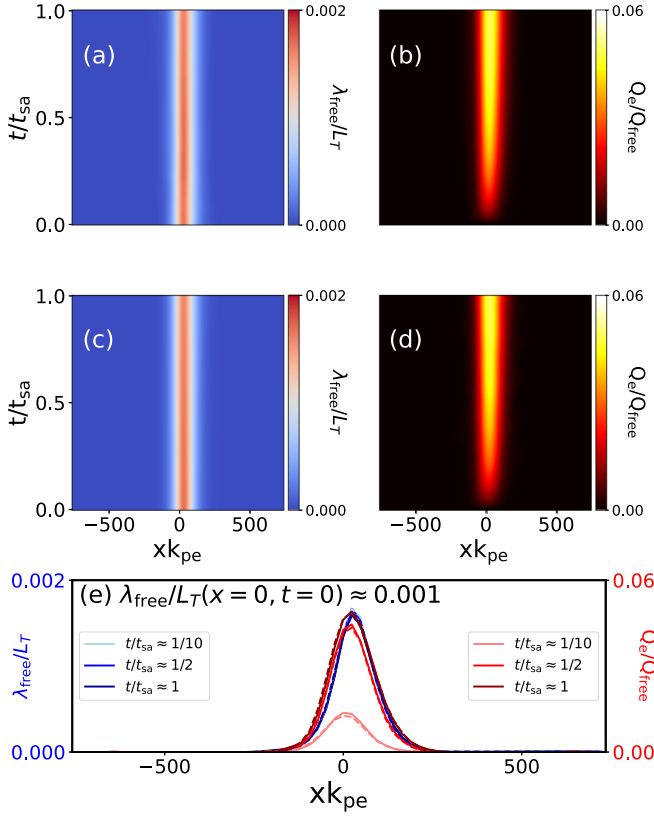


FIG. 7. Spatiotemporal prediction results by TCNN in the local regime with an initial characteristic ratio of $\lambda_{\text{free}}/L_T \approx 0.001$. Panels (a,b) show the model-predicted evolution of $\lambda_{\text{free}}/L_T$ and normalized heat flux Q_e/Q_{free} , respectively. Panels (c), (d) display the corresponding ground-truth quantities obtained from PIC simulations. Panel (e) presents a direct comparison of spatial profiles at three representative normalized times: $t/t_{\text{sa}} = 1/10, 1/2$, and 1 . The solid light-blue, blue, and dark-blue curves represent the TCNN-predicted spatial profiles of $\lambda_{\text{free}}/L_T$ (left y axis), while the solid light-red, red, and dark-red curves denote the corresponding predictions for the normalized heat flux Q_e/Q_{free} (right y axis). The dashed curves of matching colors indicate the ground truth results obtained from PIC simulations.

constrain their ability to model the intrinsic nonlinearity and nonlocality of heat transport. In contrast, the architecture in Fig. 6(c) demonstrates notable performance gains through the use of a deeper convolutional network. This capability is further significantly improved in Fig. 6(d), where convolution is performed within a latent spatial framework, enabling more expressive feature extraction and better generalization across regimes, although its performance remains slightly inferior to the full TCNN model shown in Fig. 4, whose performance is quantified in Fig. 5(b), owing to the latter's more expressive latent space.

Figures 7 and 8 showcase two representative examples of the spatiotemporal predictions made by the TCNN model. Figure 7 corresponds to a scenario within the local regime, characterized by an initial characteristic ratio of $\lambda_{\text{free}}/L_T \approx 0.001$, while Fig. 8 illustrates a strongly nonlocal case with $\lambda_{\text{free}}/L_T \approx 0.084$. Panels (a) and (b) in each figure depict the model-predicted evolution of $\lambda_{\text{free}}/L_T$ and the

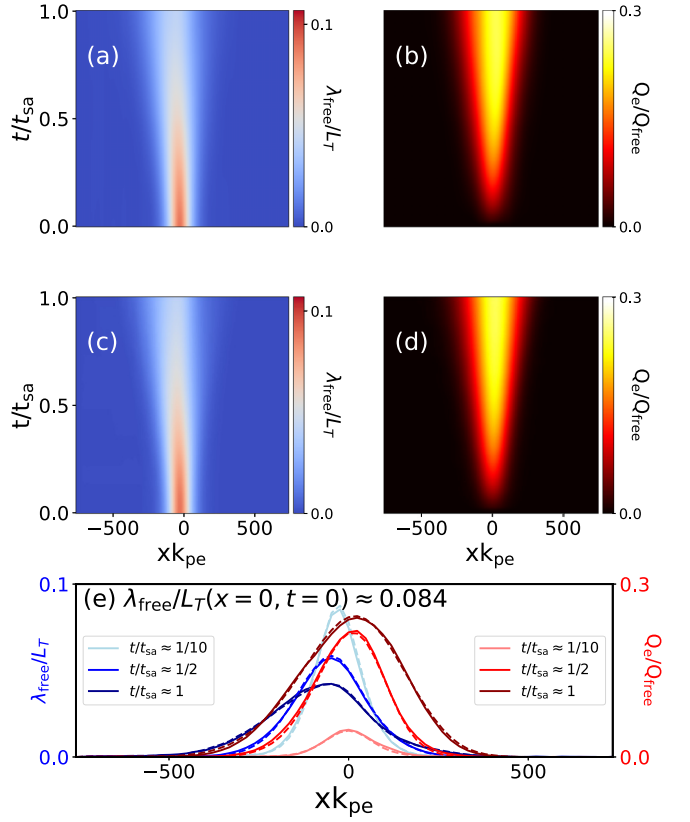


FIG. 8. Spatiotemporal prediction results by TCNN in the strongly nonlocal regime with an initial characteristic ratio of $\lambda_{\text{free}}/L_T \approx 0.084$. Panels (a), (b) show the model-predicted evolution of $\lambda_{\text{free}}/L_T$ and normalized heat flux Q_e/Q_{free} , respectively. Panels (c), (d) display the corresponding ground-truth quantities obtained from PIC simulations. Panel (e) presents a direct comparison of spatial profiles at three representative normalized times: $t/t_{\text{sa}} = 1/10, 1/2$, and 1 . The solid light-blue, blue, and dark-blue curves represent the TCNN-predicted spatial profiles of $\lambda_{\text{free}}/L_T$ (left y axis), while the solid light-red, red, and dark-red curves denote the corresponding predictions for the normalized heat flux Q_e/Q_{free} (right y axis). The dashed curves of matching colors indicate the ground truth results obtained from PIC simulations.

normalized heat flux Q_e/Q_{free} , respectively, across space and time. For comparison, panels (c) and (d) display the corresponding ground truth data generated via PIC simulations. Panel (e) of each figure presents spatial profiles of both quantities at three distinct normalized times, $t/t_{\text{sa}} = 1/10, 1/2$, and 1 . The solid light-blue, blue, and dark-blue curves represent the TCNN-predicted spatial profiles of $\lambda_{\text{free}}/L_T$ (left y axis), while the solid light-red, red, and dark-red curves denote the corresponding predictions for the normalized heat flux Q_e/Q_{free} (right y axis). The dashed curves of matching colors indicate the ground truth results obtained from PIC simulations. Across both the collisional and strongly nonlocal regimes, the TCNN model accurately reconstructs the spatial and temporal evolution of key transport quantities.

Finally, we examine the robustness of the TCNN to input noise. To generate noisy data, we perturb the initial condition of $\lambda_{\text{free}}/L_T$ according to

$$(\lambda_{\text{free}}/L_T)_{\text{noise}} = \lambda_{\text{free}}/L_T + \epsilon(\lambda_{\text{free}}/L_T)_{\text{max}} \xi,$$

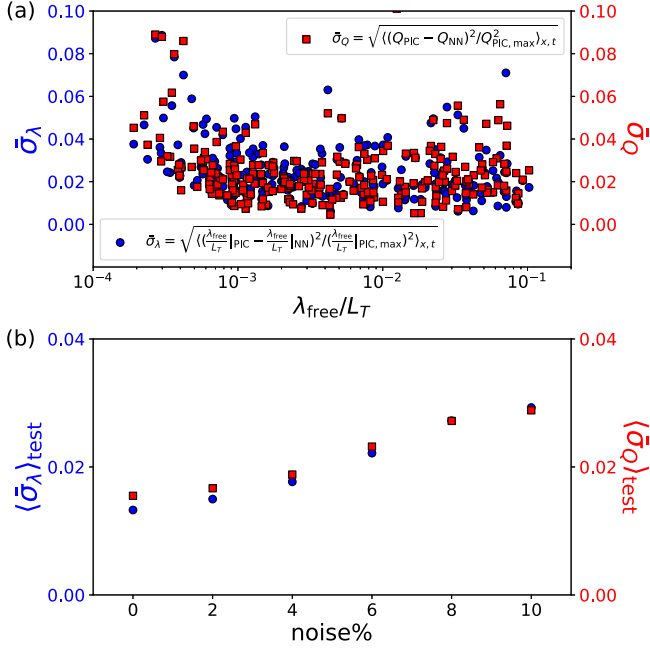


FIG. 9. Robustness of the TCNN to input noise. (a) Deviation of the TCNN predictions from the ground truth for $\lambda_{\text{free}}/L_T$ and the normalized heat flux Q_e/Q_{free} when 10% Gaussian noise is added to the initial condition of $\lambda_{\text{free}}/L_T$. (b) Averaged prediction deviations over all testing samples, $\langle \delta_\lambda \rangle_{\text{test}}$ and $\langle \delta_Q \rangle_{\text{test}}$, as functions of the imposed noise level ϵ .

where $\xi(x) \sim \mathcal{N}(0, 1)$ denotes an i.i.d. Gaussian random field with zero mean and unit variance at each spatial grid point, ϵ controls the noise level, and $(\lambda_{\text{free}}/L_T)_{\text{max}}$ is the maximum value of the initial condition of each sample over the space. By introducing random noise into the initial condition, we assess how such perturbations affect the subsequent evolution of $\lambda_{\text{free}}/L_T$ and the normalized heat flux Q_e/Q_{free} . Figure 9(a) shows the deviation of the model predictions from the ground truth when 10% noise is added to the input. Compared to Fig. 5(b), where no noise is applied, the model performance does not degrade significantly, and most testing samples exhibit only small deviations. The average standard deviation over all testing samples, i.e., $\langle \delta_\lambda \rangle_{\text{test}}$ and $\langle \delta_Q \rangle_{\text{test}}$, as a function of the noise level, is shown in Fig. 9(b). We find that while increasing noise levels lead to larger prediction errors, the error growth remains sublinear with respect to the noise amplitude, indicating good robustness of the TCNN to input noise.

V. CONCLUSION AND DISCUSSION

In this study, we presented a unique deep learning framework, the time-embedded convolutional neural network (TCNN), for modeling spatiotemporal heat transport in plasmas, particularly within the strongly nonlocal regime. This architecture is physics informed, incorporating the underlying principle that heat flux and nonlocality are intrinsically coupled, while allowing data-driven flexibility by relaxing the strict kernel-based formalism of traditional models such as the LMV theory.

A key feature of TCNN is its ability to jointly predict the normalized heat flux Q_e/Q_{free} and the characteristic ratio $\lambda_{\text{free}}/L_T$ using a unified set of convolutional parameters. This formulation reflects the underlying physical coupling between heat flux and nonlocality, providing a more coherent and physically consistent modeling framework. By embedding temporal information directly into the model through time-indexed convolutions, TCNN learns to represent dynamic system evolution. This time-embedded modeling allows for capturing the progression of heat transport processes across varying regimes of collisionality, outperforming prior models such as SH and LMV, especially under strongly nonlocal conditions. Although its accuracy degrades slightly in the local, quasistatic case due to diminished temporal variation, this limitation could be addressed by employing a hybrid of the TCNN and LNN [19] models. While integration into radiation-hydrodynamics codes would follow the same approach as the LMV or SH models, implementation is not our immediate goal. Our current focus is on developing a reliable, physics-informed, and stable neural network to establish the model, while exploring multiscale strategies, such as partitioning hydrodynamic domains into smaller grids treatable with PIC, to overcome scale limitations for future coupling with hydrodynamic codes. Although the PIC simulations are currently set up under an isobaric condition, the TCNN model is not limited by this assumption, since kernel-based nonlocal approaches are not restricted to this geometry. Meanwhile, the TCNN can be extended directly to multidimensional situations by using multidimensional convolutions, and would learn the corresponding nonlocal operator if multidimensional kinetic data were available. We note that kernel-based nonlocal models have been implemented in three-dimensional settings [5], indicating the validity of the multidimensional nonlocal kernels, but training on 2D or 3D kinetic data would be more computationally expensive because generating multidimensional datasets is costly.

We conducted an extensive evaluation of multiple architectural variants to dissect the impact of convolutional depth and latent spatial projections. Results show that shallow or unprojected models struggle to capture the complexity of nonlocal dynamics, while deeper convolutional stacks operating in latent space substantially improve performance. The full TCNN architecture, with its rich latent representation and depth, achieved sufficient accuracy, highlighting the importance of flexible and extensible design in deep learning models for physical systems. Meanwhile, since convolution in real space is dual to multiplication in Fourier space, the required TCNN capacity can be related to the degrees of freedom in Fourier space, consistent with Fourier-based closure approaches [27].

Beyond the immediate application to plasma heat transport, our methodology offers broader inspirations for applying neural networks to complex physical systems [28–31]. In particular, the integration of temporal modulation, joint multitarget prediction, and architecture-informed loss functions provides a practical framework for modeling multiscale, nonlinear, and nonlocal processes. We hope this work serves as a reference for researchers seeking to design models that are both physically grounded and data efficient, and helps bridge the gap between traditional physics-based modeling and modern machine learning [32–38]. While the present work focuses

on directly learning the nonlocal heat flux, an alternative approach would be to learn the underlying first-order spherical harmonic components and reconstruct the nonlocal heat flux indirectly. In particular, within a multiple-group framework inspired by the SNB model, such an approach could allow the nonlocal heat flux to be obtained as a correction to the SH, while simultaneously providing access to information on the electron energy distribution. We leave the exploration of this direction to our future work.

ACKNOWLEDGMENTS

The authors thank S. Hu, S. Nathaniel, S. Hüller, T. Fülöp, and C. Riconda for their useful discussions. They are

also grateful for the computing resources provided by the STFC Scientific Computing Department's SCARF cluster. This work was supported by EPSRC and First Light Fusion under the AMPLIFI prosperity partnership, Grant No. EP/X025373/1.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available upon publication because it is not technically feasible and/or the cost of preparing, depositing, and hosting the data would be prohibitive within the terms of this research project. The data are available from the authors upon reasonable request.

-
- [1] G. Gregori, S. H. Glenzer, J. Knight, C. Niemann, D. Price, D. H. Froula, M. J. Edwards, R. P. J. Town, A. Brantov, W. Rozmus, and V. Y. Bychenkov, Effect of nonlocal transport on heat-wave propagation, *Phys. Rev. Lett.* **92**, 205006 (2004).
 - [2] C. Arran, M. Holec, M. Sherlock, M. Marinak, and C. P. Ridgers, The accuracy of multi-group models for nonlocal electron transport in magnetized plasmas, *Phys. Plasmas* **32**, 043908 (2025).
 - [3] O. A. Hurricane, P. K. Patel, R. Betti, D. H. Froula, S. P. Regan, S. A. Slutz, M. R. Gomez, and M. A. Sweeney, Physics principles of inertial confinement fusion and u.s. program overview, *Rev. Mod. Phys.* **95**, 025005 (2023).
 - [4] B. Somov, *Plasma Astrophysics, Part I: Fundamentals and Practice*, Astrophysics and Space Science Library (Springer, New York, 2012).
 - [5] S. Silva, J. Santos, J. Büchner, and M. Alves, Nonlocal heat flux effects on temperature evolution of the solar atmosphere, *Astron. Astrophys.* **615**, A32 (2018).
 - [6] E. L. Yerger, M. W. Kunz, A. F. Bott, and A. Spitkovsky, Collisionless conduction in a high-beta plasma: a collision operator for whistler turbulence, *J. Plasma Phys.* **91**, E20 (2025).
 - [7] S. Komarov, A. A. Schekochihin, E. Churazov, and A. Spitkovsky, Self-inhibiting thermal conduction in a high- β , whistler-unstable plasma, *J. Plasma Phys.* **84**, 905840305 (2018).
 - [8] J. D. Huba, NRL Plasma Formulary, Naval Research Laboratory, Washington, DC, 2023 Report No. nRL/PU/6790–23-528.
 - [9] L. Spitzer and R. Härm, Transport phenomena in a completely ionized gas, *Phys. Rev.* **89**, 977 (1953).
 - [10] A. V. Brantov, V. Y. Bychenkov, V. T. Tikhonchuk, and W. Rozmus, Nonlocal electron transport in laser heated plasmas, *Phys. Plasmas* **5**, 2742 (1998).
 - [11] J. F. Luciani, P. Mora, and J. Virmont, Nonlocal heat transport due to steep temperature gradients, *Phys. Rev. Lett.* **51**, 1664 (1983).
 - [12] G. P. Schurtz, P. D. Nicolaï, and M. Busquet, A nonlocal electron conduction model for multidimensional radiation hydrodynamics codes, *Phys. Plasmas* **7**, 4238 (2000).
 - [13] E. M. Epperlein and R. W. Short, A practical nonlocal model for electron heat transport in laser plasmas, *Phys. Fluids B* **3**, 3092 (1991).
 - [14] O. V. Batishchev, V. Y. Bychenkov, F. Detering, W. Rozmus, R. Sydora, C. E. Capjack, and V. N. Novikov, Heat transport and electron distribution function in laser produced plasmas with hot spots, *Phys. Plasmas* **9**, 2302 (2002).
 - [15] O. Michel, R. Duclous, P.-E. Masson-Laborde, C. Enaux, and P. Lafitte, A nonlocal electron transport model in the diffusion scaling of hydrodynamics, *Phys. Plasmas* **30**, 022712 (2023).
 - [16] A. Chrisment, P. Loiseau, J.-L. Feugeas, P.-E. Masson-Laborde, J. Mathiaud, V. Tikhonchuk, and P. Nicolaï, Analysis of a kinetic model for electron heat transport in inertial confinement fusion plasmas, *Phys. Plasmas* **29**, 062301 (2022).
 - [17] W. Q. Yuan, Z. H. Zhao, S. P. Zhu, X. T. He, and B. Qiao, Space-time dependent non-local thermal transport effects on laser ablation dynamics in inertial confinement fusion, *Plasma Phys. Controlled Fusion* **66**, 045023 (2024).
 - [18] C. Lamy, B. Dubroca, P. Nicolaï, V. Tikhonchuk, and J.-L. Feugeas, Modeling of electron nonlocal transport in plasmas using artificial neural networks, *Phys. Rev. E* **105**, 055201 (2022).
 - [19] M. Luo, C. Heaton, Y. Wang, D. Plummer, M. Fitzgerald, F. Miniati, S. M. Vinko, and G. Gregori, Learning heat transport kernels using a nonlocal heat transport theory-informed neural network, *Phys. Rev. Res.* **7**, L042017 (2025).
 - [20] R. A. Fonseca, L. O. Silva, F. S. Tsung, V. K. Decyk, W. Lu, C. Ren, W. B. Mori, S. Deng, S. Lee, T. Katsouleas, and J. C. Adam, Osiris: A three-dimensional, fully relativistic particle in cell code for modeling plasma based accelerators, in *Computational Science—ICCS 2002*, edited by P. M. A. Sloot, A. G. Hoekstra, C. J. K. Tan, and J. J. Dongarra (Springer, Berlin, 2002), pp. 342–351.
 - [21] Y. Sentoku, K. Mima, Y. Kishimoto, and M. Honda, Effects of relativistic binary collisions on pic simulation of laser plasmas, *J. Phys. Soc. Jpn.* **67**, 4084 (1998).
 - [22] Y. Sentoku and A. Kemp, Numerical methods for particle simulations at extreme densities and temperatures: Weighted particles, relativistic collisions and reduced currents, *J. Comput. Phys.* **227**, 6846 (2008).
 - [23] S. X. Hu, K. A. Nichols, N. R. Shaffer, B. Arnold, A. J. White, L. A. Collins, V. V. Karasiev, S. Zhang, V. N. Goncharov, R. C. Shah, D. I. Mihaylov, S. Jiang, and Y. Ping, A review on charged-particle transport modeling for laser direct-drive fusion, *Phys. Plasmas* **31**, 040501 (2024).
 - [24] A. R. Bell, R. G. Evans, and D. J. Nicholas, Electron energy transport in steep temperature gradients in laser-produced plasmas, *Phys. Rev. Lett.* **46**, 243 (1981).

- [25] T. D. Arber, K. Bennett, C. S. Brady, A. Lawrence-Douglas, M. G. Ramsay, N. J. Sircombe, P. Gillies, R. G. Evans, H. Schmitz, A. R. Bell, and C. P. Ridgers, Contemporary particle-in-cell approach to laser-plasma modelling, *Plasma Phys. Controlled Fusion* **57**, 113001 (2015).
- [26] N. Kovachki, Z. Li, B. Liu, K. Azizzadenesheli, K. Bhattacharya, A. Stuart, and A. Anandkumar, Neural operator: Learning maps between function spaces with applications to PDEs, *J. Mach. Learn. Res.* **24**, 1 (2023).
- [27] C. Ma, B. Zhu, X.-Q. Xu, and W. Wang, Machine learning surrogate models for landau fluid closure, *Phys. Plasmas* **27**, 042502 (2020).
- [28] S. Hüller, A. Porzio, A. Héron, and P. Mora, Laser-driven ion acceleration in the presence of increasing heating of relativistic electrons at steep overdense plasma interfaces, *Phys. Plasmas* **31**, 103103 (2024).
- [29] S. Hüller, A. Porzio, J.-C. Adam, and A. Héron, On the non-thermal nature of distributions of electrons accelerated by high intensity lasers at the vacuum-plasma interface, *Phys. Plasmas* **26**, 083107 (2019).
- [30] M. Luo, C. Riconda, I. Pusztai, A. Grassi, J. S. Wurtele, and T. Fülöp, Control of autoresonant plasma beat-wave wakefield excitation, *Phys. Rev. Res.* **6**, 013338 (2024).
- [31] M. Luo, C. Riconda, A. Grassi, N. Wang, J. Wurtele, T. Fülöp, and I. Pusztai, Evolution of autoresonant plasma wave excitation in two-dimensional particle-in-cell simulations, *J. Plasma Phys.* **91**, E31 (2025).
- [32] F. Miniati and G. Gregori, Learning transport processes with machine intelligence, *Sci. Rep.* **12**, 11709 (2022).
- [33] E. P. Alves and F. Fiuza, Data-driven discovery of reduced plasma physics models from fully kinetic simulations, *Phys. Rev. Res.* **4**, 033192 (2022).
- [34] Z. Huang, C. Dong, and L. Wang, Machine-learning heat flux closure for multi-moment fluid modeling of nonlinear landau damping, *Proc. Natl. Acad. Sci. USA* **122**, e2419073122 (2025).
- [35] D. D. Carvalho, D. R. Ferreira, and L. O. Silva, Learning the dynamics of a one-dimensional plasma model with graph neural networks, *Mach. Learn.: Sci. Technol.* **5**, 025048 (2024).
- [36] C. J. McDevitt and X. Tang, A physics-informed deep learning description of Knudsen layer reactivity reduction, *Phys. Plasmas* **31**, 062701 (2024).
- [37] A. Döpp, C. Eberle, S. Howard, F. Irshad, J. Lin, and M. Streeter, Data-driven science and machine learning methods in laser-plasma physics, *High Power Laser Sci. Eng.* **11**, e55 (2023).
- [38] E. R. Ingelsten, M. C. McGrae-Menge, E. P. Alves, and I. Pusztai, Data-driven discovery of a heat flux closure for electrostatic plasma phenomena, *J. Plasma Phys.* **91**, E64 (2025).