

UNIFORMLY ACCURATE STRUCTURE-PRESERVING NEURAL SURROGATES FOR RADIATIVE TRANSFER *

MENGJIA BAI[†], JINGRUN CHEN[‡], AND KEKE WU[§]

Abstract. In this work, we propose a uniformly accurate, structure-preserving neural surrogate for the radiative transfer equation with periodic boundary conditions based on a multiscale parity decomposition framework. The formulation introduces a refined decomposition of the particle distribution into macroscopic, odd, and higher-order even components, leading to an asymptotic-preserving neural network system that remains stable and accurate across all parameter regimes. By constructing key higher-order correction functions, we establish rigorous uniform error estimates with respect to the scale parameter ε , which ensures ε -independent accuracy. Furthermore, the neural architecture is designed to preserve intrinsic physical structures such as parity symmetry, conservation, and positivity through dedicated architectural constraints. The framework extends naturally from one to two dimensions and provides a theoretical foundation for uniformly accurate neural solvers of multiscale kinetic equations. Numerical experiments confirm the effectiveness of our approach.

Key words. Uniformly accurate, structure-preserving, asymptotic-preserving, radiative transfer

MSC codes. 82B40, 35Q20, 45K05, 82C40, 65N12

1. Introduction. Radiative transfer describes the propagation and interactions of radiation or particles within participating media [35, 24]. This field of study has broad applications across disciplines such as medical imaging [39, 23, 22], cancer therapy [5, 37, 4], and astrophysical flows [33]. It is also essential for analyzing energy transfer in engineering systems – for example, neutron transport in fission reactors [10] or photon-mediated heating in inertial confinement fusion capsules [25].

The numerical solution of the radiative transfer equation (RTE), however, poses several well-known challenges. First, the solution is defined over a high-dimensional phase space, typically encompassing three spatial coordinates and three momentum (or angular) variables, and it may also exhibit time dependence. Second, the interaction between radiation and matter can be highly complex, especially when the material properties themselves evolve. Third, spatial and temporal heterogeneities in collision rates, arising from strongly energy-dependent processes or material discontinuities, introduce multiple characteristic scales. To address these difficulties, various numerical strategies have been developed, broadly categorized into deterministic and stochastic methods. Among deterministic approaches, the discrete ordinates method (DOM/DVM), also known as the S_N method, is widely adopted. It discretizes the angular variables and solves the RTE along discrete directions [25, 2]. Spherical harmonics methods, by contrast, offer the advantage of rotational invariance and

*Submitted to the editors DATE: November 10, 2025.

Funding: Jingrun Chen is supported by NSFC Major Research Plan - Interpretable and General-purpose Next-generation Artificial Intelligence (No 92370205), NSFC grant 12425113, and the Key Laboratory of the Ministry of Education for Mathematical Foundations and Applications of Digital Technology, University of Science and Technology of China. Keke Wu is supported by the China Postdoctoral Science Foundation under Grant Number 2025M773105 and the Jiangsu Funding Program for Excellent Postdoctoral Talent.

[†]School of Mathematical Sciences, Fudan University, Shanghai, China (bai_mj@fudan.edu.cn).

[‡]School of Mathematical Sciences and Suzhou Institute for Advanced Research, University of Science and Technology of China, and Suzhou Big Data & AI Research and Engineering Center, China (jingrunchen@ustc.edu.cn).

[§]School of Artificial Intelligence and Data Science and Suzhou Institute for Advanced Research, University of Science and Technology of China, and Suzhou Big Data & AI Research and Engineering Center, China (wuokever@ustc.edu.cn).

have been extensively applied to radiative transfer problems [32, 11]. Given that the characteristic scale parameter (e.g., Knudsen number) in the RTE can span regimes from kinetic to diffusive, it is essential for numerical methods to handle multiscale behavior robustly. Two major classes of approaches have been proposed to address this challenge: domain decomposition-based methods and asymptotic-preserving (AP) schemes. Domain decomposition-based methods partition the computational domain into subregions, in which different governing equations are solved and coupled through interface conditions [13]. AP schemes, on the other hand, are designed to maintain uniform stability and accuracy across all scale regimes [17]. For stochastic approaches, the direct simulation Monte Carlo (DSMC) method has been widely used for RTE computations [3]. Numerous other computational techniques have also been proposed, addressing aspects such as implicit time integration, high-order accuracy, and efficient handling of multiscale effects [12, 28, 38, 42, 29, 14].

Recently, there has been growing interest in leveraging deep neural network methods to solve RTEs [34, 9, 31, 18, 19, 40, 26, 27, 7, 41]. When dealing with RTEs, the vanilla physics-informed neural networks (PINNs) [36] often suffer from numerical instability arising from the presence of small-scale features [18]. A key strategy for addressing the multiscale nature of RTEs lies in the careful design of a uniformly accurate loss function that captures the macroscopic limiting behavior. This approach, known as asymptotic-preserving neural networks (APNNs) [18], is illustrated in Fig. 1. The APNN framework has been further generalized to address a wide range of prob-

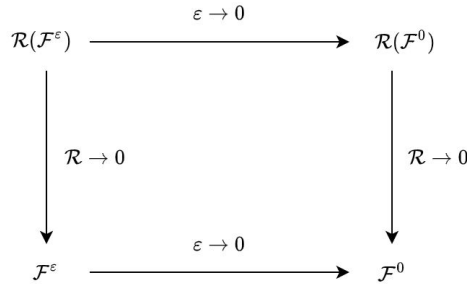


Fig. 1: Schematic diagram of the asymptotic-preserving neural networks (adapted from [18]). Assume $\mathcal{F}^\varepsilon$ is the multi-scale model that depends on the scale parameter ε and $\mathcal{F}^0$ is the corresponding asymptotic limit model as $\varepsilon \rightarrow 0$. Define $\mathcal{R}(\mathcal{F}^\varepsilon)$ as the neural network-based least-squares formulation of the model $\mathcal{F}^\varepsilon$. If $\mathcal{R}(\mathcal{F}^\varepsilon)$ converges to $\mathcal{R}(\mathcal{F}^0)$ as $\varepsilon \rightarrow 0$, and this limit is precisely the least-squares formulation of the limit model $\mathcal{F}^0$, then the method is called asymptotic-preserving.

lems, such as hyperbolic systems [6], the semiconductor Boltzmann equation [30], the Vlasov–Poisson–Fokker–Planck equation [20, 43], and the Boltzmann equation [8], thereby providing compelling evidence of its robustness and versatility.

While numerical methods have shown remarkable performance in multiscale kinetic problems, the lack of rigorous analytical justification underscores the need for a deeper theoretical understanding – this serves as the primary motivation for our study. To the best of our knowledge, existing theoretical analyses have predominantly been developed within the micro–macro decomposition framework. For instance, [31] established a uniform error estimate with respect to the Knudsen number for the steady

RTE, whereas [1] provided a rigorous proof of the AP property of APNNs applied to the Boltzmann equation. Other works, such as [9] and [26], studied the linear transport equation and gray RTEs, though their analyses are not uniform across regimes. More recently, [8] established rigorous statistical estimates for the multiscale control variate method, further advancing the theoretical understanding of such multiscale formulations.

While these efforts have significantly deepened insights into the micro–macro paradigm, our work instead draws inspiration from the parity decomposition framework. We introduce a novel multiscale parity decomposition and establish a uniform error estimate for the approximate solution with respect to ε . A key refinement in our formulation is the introduction of a new quantity of $\mathcal{O}(\varepsilon^2)$ – the deviation from local equilibrium, defined in the one-dimensional case as $\varepsilon^2 w = r - \rho$ (see Section 2). The two-dimensional extension involves additional subtleties, which are discussed in detail in Section 3. This refinement leads to a fundamental shift in the structure of our error analysis, where the central analytical object becomes the error in the refined variable set (ρ, w, j) .

1.1. Preliminaries. Consider the time-dependent RTE. Let $f(t, \mathbf{x}, \mathbf{v})$ be the probability density distribution for particles at space point $\mathbf{x} \in \mathcal{D} \subset \mathbb{R}^d$, time t , and travel in direction $\mathbf{v} \in \Omega \subset \mathbb{S}^d$, with $\int_{\Omega} d\mathbf{v} = S$. Here Ω is symmetric in $\mathbf{v}$, meaning that $\int_{\Omega} g(\mathbf{v}) d\mathbf{v} = 0$ for any function g odd in $\mathbf{v}$. Then f solves the RTE

$$(1.1) \quad \varepsilon \partial_t f + \mathbf{v} \cdot \nabla_{\mathbf{x}} f = \frac{1}{\varepsilon} \left(\frac{\sigma_S}{S} \int_{\Omega} f d\mathbf{v}' - \sigma f \right) + \varepsilon Q,$$

where $\sigma = \sigma(\mathbf{x})$ is the total transport coefficient, $\sigma_S = \sigma_S(\mathbf{x})$ is the scattering coefficient, $Q = Q(\mathbf{x})$ is the source term. In this model, the parameter $\varepsilon > 0$ is the Knudsen number, which denotes the ratio of the mean free path to a characteristic length. In particular, $\varepsilon \sim \mathcal{O}(1)$ refers to kinetic regime, and $\varepsilon \ll 1$ corresponds to the diffusive regime. Typically,

$$(1.2) \quad \sigma_S = \sigma - \varepsilon^2 \sigma_A,$$

where $\sigma_A = \sigma_A(\mathbf{x})$ is the absorption coefficient. Such an equation arises in neutron transport, wave propagation in random media. In all these applications, the scaling appeared in (1.1), and gives rise to a diffusion equation as $\varepsilon \rightarrow 0$, which is

$$(1.3) \quad \partial_t \rho = D \nabla_{\mathbf{x}} \cdot \left(\frac{1}{\sigma} \nabla_{\mathbf{x}} \rho \right) - \sigma_A \rho + Q, \quad \rho = \frac{1}{S} \int_{\Omega} f d\mathbf{v}.$$

Remark 1.1. Different diffusion coefficient D in (1.3) appears for different collision operators. For example, $D = 1/3$ in one-dimensional slab geometry as shown in Section 2, and $D = 1/2$ when Ω is a unit sphere in two dimensions, as will be discussed in Section 3.

In the next sections, for the sake of clarity, we will derive the frameworks for one-dimensional and two-dimensional equations in the simplified situation where $Q = 0$, $\sigma_A = 0$, and $\sigma = \sigma_S = 1$. The extension to the general case does not present additional difficulties and only needs some smooth assumptions.

1.2. Main results. Our main theoretical result is a uniform error estimate for the proposed multiscale parity decomposition framework.

THEOREM 1.2. *Let $\mathcal{R}_{\text{APNN}}^\varepsilon$ represent the physics-informed loss based on the residual of the multiscale parity-decomposition system, and let $\mathcal{R}_{\text{total}}^\varepsilon$ denote the error between the numerical solution and the exact solution. Then it follows that*

$$\mathcal{R}_{\text{total}}^\varepsilon = \mathcal{O}\left(\sqrt[3]{(\mathcal{R}_{\text{APNN}}^\varepsilon)^2}\right), \quad \text{uniformly in } \varepsilon.$$

The results conclusively demonstrate that the proposed multiscale parity decomposition framework converges uniformly in the Knudsen number ε .

In fact, as shown in Theorems 2.3 and 3.2, a direct numerical analysis gives the error estimate:

$$\mathcal{R}_{\text{total}}^\varepsilon = \mathcal{O}\left(\frac{\mathcal{R}_{\text{APNN}}^\varepsilon}{\varepsilon}\right) := \mathcal{E}_1.$$

However, this result exhibits a scaling of ε^{-1} , which becomes unbounded as $\varepsilon \rightarrow 0$. In contrast, the multiscale parity decomposition framework enables a refined analysis, leading to the estimate:

$$\mathcal{R}_{\text{total}}^\varepsilon = \mathcal{O}(\mathcal{R}_{\text{APNN}}^\varepsilon + \varepsilon^2) := \mathcal{E}_2,$$

established in Theorems 2.4 and 3.3. Both $\mathcal{E}_1$ and $\mathcal{E}_2$ are mathematically valid and hold simultaneously in this framework. A combination of the two estimates leads to a composite bound:

$$\mathcal{R}_{\text{total}}^\varepsilon \leq \min\{\mathcal{E}_1, \mathcal{E}_2\},$$

which achieves its upper bound at $\varepsilon = \mathcal{O}(\sqrt[3]{\mathcal{R}_{\text{APNN}}^\varepsilon})$, as illustrated in Fig. 2. Therefore, we obtain the uniform convergence result.

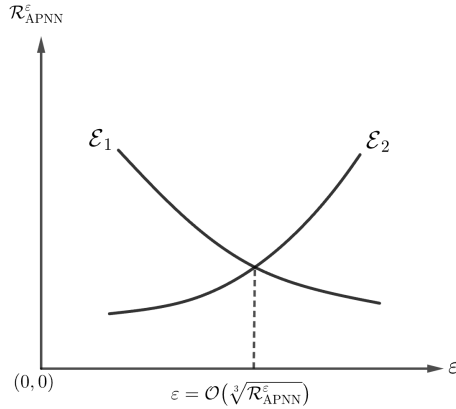


Fig. 2: Illustration of uniform convergence of the AP framework.

The rest of this paper is organized as follows. Section 2 details the one-dimensional multiscale parity decomposition framework and the corresponding uniform error estimates. Section 3 extends this formulation to two-dimensional problems. Section 4 discusses the structure-preserving mechanisms of our AP framework. Numerical results are provided in Section 5, and conclusions are given in Section 6.

2. One-dimensional method. In this section, we consider the one-dimensional RTE in the context of diffusive scaling, given by

$$(2.1) \quad \varepsilon \partial_t f + v \partial_x f = \frac{1}{\varepsilon} \left(\frac{1}{2} \int_{-1}^1 f \, dv' - f \right),$$

where $x \in \mathcal{D} := (x_L, x_R)$ with velocity $v \in \Omega := [-1, 1]$. We consider the periodic boundary condition

$$(2.2) \quad f(t, x_L, v) = f(t, x_R, v).$$

Furthermore, assume that the initial condition is given as a function of x and v :

$$(2.3) \quad f(0, x, v) := f_{IC}(x, v).$$

2.1. Multiscale parity decomposition method. We next demonstrate the multiscale parity decomposition technique and derive the APNN system for the RTE. Denote $\Omega^+ = [0, 1]$, and define the even and odd parities for $v \in \Omega^+$ as

$$(2.4) \quad \begin{cases} r(t, x, v) = [f(t, x, v) + f(t, x, -v)]/2, \\ j(t, x, v) = [f(t, x, v) - f(t, x, -v)]/2\varepsilon. \end{cases}$$

Inserting (2.4) into (2.1), one gets the system of equations as follows:

$$(2.5) \quad \begin{cases} \varepsilon^2 \partial_t r + \varepsilon^2 v \partial_x j = \rho - r, \\ \varepsilon^2 \partial_t j + v \partial_x r = -j, \end{cases}$$

where $\rho = \langle r \rangle := \int_0^1 r(t, x, v) dv$. To establish a consistent error analysis with respect to ε , we introduce ρ as a mediator between r and j within this system. Integrating the first equation of (2.5) over v yields the following equation for ρ :

$$(2.6) \quad \partial_t \rho + \langle v \partial_x j \rangle = 0.$$

Furthermore, we define $r = \rho + \varepsilon^2 w$, then subtract the first equation of (2.5) by (2.6) to deduce

$$(2.7) \quad \varepsilon^2 \partial_t w + w + v \partial_x j - \langle v \partial_x j \rangle = 0.$$

Therefore, we employ a multiscale parity decomposition of the distribution function $f(t, x, v)$, explicitly separating the macroscopic density $\rho(t, x)$ at $\mathcal{O}(1)$, the odd term $j(t, x, v)$ at $\mathcal{O}(\varepsilon)$, and higher-order corrections $w(t, x, v)$ at $\mathcal{O}(\varepsilon^2)$, each with definite parity properties, i.e.,

$$(2.8) \quad f(t, x, v) = \underbrace{\rho(t, x)}_{\mathcal{O}(1)} + \underbrace{\varepsilon j(t, x, v)}_{\mathcal{O}(\varepsilon)} + \underbrace{\varepsilon^2 w(t, x, v)}_{\mathcal{O}(\varepsilon^2)}.$$

To summarize, equations (2.5), (2.6) and (2.7) constitute the multiscale parity decomposition framework of (2.1) to be used for the APNN:

$$(2.9) \quad \begin{cases} \partial_t \rho + \langle v \partial_x j \rangle = 0, \\ \varepsilon^2 \partial_t j + v \partial_x \rho + \varepsilon^2 v \partial_x w + j = 0, \\ \varepsilon^2 \partial_t w + w + v \partial_x j - \langle v \partial_x j \rangle = 0, \end{cases}$$

with the constraint $\langle w \rangle = 0$. When $\varepsilon \rightarrow 0$, the above equation formally approaches

$$\begin{cases} \partial_t \rho + \langle v \partial_x j \rangle = 0, \\ v \partial_x \rho + j = 0, \\ w + v \partial_x j - \langle v \partial_x j \rangle = 0. \end{cases}$$

Substituting the second equation into the first and third equations to deduce

$$(2.10) \quad \begin{cases} w = (v^2 - \frac{1}{3}) \partial_{xx} \rho, \\ \partial_t \rho - \frac{1}{3} \partial_{xx} \rho = 0. \end{cases}$$

2.2. Loss function. Let $\rho_\theta, j_\theta, w_\theta$ be the neural network approximations to ρ, j, w of the (2.9), with θ denoting the trainable parameters of the network. For the APNN method, we propose the physics-informed loss based on the residual for the multiscale parity decomposition system (2.9) as the loss function:

$$\mathcal{R}_{\text{APNN}}^\varepsilon = \mathcal{R}_{\text{residual}}^\varepsilon + \mathcal{R}_{\text{initial}}^\varepsilon + \mathcal{R}_{\text{boundary}}^\varepsilon,$$

where $\mathcal{R}_{\text{residual}}^\varepsilon$, $\mathcal{R}_{\text{initial}}^\varepsilon$ and $\mathcal{R}_{\text{boundary}}^\varepsilon$ are defined as follows

$$\begin{aligned} \mathcal{R}_{\text{residual}}^\varepsilon &= \frac{1}{|\mathcal{T} \times \mathcal{D}|} \int_{\mathcal{T}} \int_{\mathcal{D}} |\partial_t \rho_\theta + \langle v \partial_x j_\theta \rangle|^2 dx dt \\ &\quad + \frac{1}{|\mathcal{T} \times \mathcal{D} \times \Omega^+|} \int_{\mathcal{T}} \int_{\mathcal{D}} \int_{\Omega^+} |\varepsilon^2 \partial_t j_\theta + v \partial_x \rho_\theta + \varepsilon^2 v \partial_x w_\theta + j_\theta|^2 dv dx dt \\ &\quad + \frac{1}{|\mathcal{T} \times \mathcal{D} \times \Omega^+|} \int_{\mathcal{T}} \int_{\mathcal{D}} \int_{\Omega^+} |\varepsilon^2 \partial_t w_\theta + w_\theta + v \partial_x j_\theta - \langle v \partial_x j_\theta \rangle|^2 dv dx dt, \\ \mathcal{R}_{\text{initial}}^\varepsilon &= \frac{\lambda_1}{|\mathcal{D}|} \int_{\mathcal{D}} |\rho_\theta - \rho_{\text{IC}}|^2 dx \\ &\quad + \frac{\lambda_2}{|\mathcal{D} \times \Omega^+|} \int_{\mathcal{D}} \int_{\Omega^+} \varepsilon^2 |j_\theta - j_{\text{IC}}|^2 + \varepsilon^4 |w_\theta - w_{\text{IC}}|^2 dv dx, \end{aligned}$$

and

$$\begin{aligned} \mathcal{R}_{\text{boundary}}^\varepsilon &= \frac{\lambda_3}{|\mathcal{T} \times \partial \mathcal{D}|} \int_{\mathcal{T}} \int_{\partial \mathcal{D}} |\rho_\theta - \rho_{\text{BC}}|^2 ds dt \\ &\quad + \frac{\lambda_4}{|\mathcal{T} \times \partial \mathcal{D} \times \Omega^+|} \int_{\mathcal{T}} \int_{\partial \mathcal{D}} \int_{\Omega^+} |w_\theta - w_{\text{BC}}|^2 + |j_\theta - j_{\text{BC}}|^2 dv ds dt. \end{aligned}$$

Here, λ_i ($i = 1, \dots, 4$) are hyperparameters as penalty terms, and $\mathcal{T}, \mathcal{D}, \Omega^+$ represent the bounded domains of time, space, and velocity space, respectively.

The AP property of this loss can be carried out by considering its behavior for ε small. One may only need to focus on the term $\mathcal{R}_{\text{residual}}^\varepsilon$. As $\varepsilon \rightarrow 0$, it yields that $\mathcal{R}_{\text{residual}}^\varepsilon$ converges to

$$\begin{aligned} \mathcal{R}_{\text{residual}}^0 &:= \frac{1}{|\mathcal{T} \times \mathcal{D}|} \int_{\mathcal{T}} \int_{\mathcal{D}} |\partial_t \rho_\theta + \langle v \partial_x j_\theta \rangle|^2 dx dt \\ &\quad + \frac{1}{|\mathcal{T} \times \mathcal{D} \times \Omega^+|} \int_{\mathcal{T}} \int_{\mathcal{D}} \int_{\Omega^+} |v \partial_x \rho_\theta + j_\theta|^2 dv dx dt \\ &\quad + \frac{1}{|\mathcal{T} \times \mathcal{D} \times \Omega^+|} \int_{\mathcal{T}} \int_{\mathcal{D}} \int_{\Omega^+} |w_\theta + v \partial_x j_\theta - \langle v \partial_x j_\theta \rangle|^2 dv dx dt, \end{aligned}$$

which is exactly the loss for the limiting equations (2.10). Thus, our designed loss function for APNN satisfies the AP property.

2.3. Convergence of the loss function. As mentioned above, we build the loss function based on the residual of the multiscale parity decomposition system (2.9). Therefore, our approximated solutions of the neural network are ρ, j and w , instead of f if we apply PINN to the model (2.1).

Let us first review an important result on the existence of the approximated neural network solution, namely the Universal Approximation Theorem.

LEMMA 2.1 ([16] Lemma A.1). *Let $\Omega = \prod_{i=1}^d [a_i, b_i]$. Suppose $f \in H^m(\Omega)$. Let $N > 5$ be an integer. Then there exists a tanh neural network $\hat{f}^N$ with two hidden layers, such that for any $k \in \{0, 1, \dots, m-1\}$,*

$$\|f - \hat{f}^N\|_{H^k(\Omega)} \leq C_{k,d,f,\Omega} (1 + \ln^k N) N^{-m+k}.$$

Here, the width of the tanh neural network depends on N .

Based on this approximation property, we can establish the convergence of the APNN loss function.

THEOREM 2.2. *Let $\rho \in H^2(\mathcal{T} \times \mathcal{D})$, $j, w \in H^2(\mathcal{T} \times \mathcal{D} \times \Omega^+)$ be the analytic solution to the multiscale parity decomposition system (2.9), and $N > 5$ be an integer. Then there exist APNN solutions $\rho_\theta = \hat{\rho}^N, j_\theta = \hat{j}^N, w_\theta = \hat{w}^N$ such that*

$$\mathcal{R}_{\text{APNN}}^\varepsilon \rightarrow 0, \quad \text{as } N \rightarrow \infty.$$

Proof. Since $\rho \in H^2(\mathcal{T} \times \mathcal{D})$, $j, w \in H^2(\mathcal{T} \times \mathcal{D} \times \Omega^+)$, then there exist neural network solutions $\rho_\theta = \hat{\rho}^N, j_\theta = \hat{j}^N, w_\theta = \hat{w}^N$, such that

$$(2.11) \quad \|\rho - \rho_\theta\|_{H^1(\mathcal{T} \times \mathcal{D})} \leq C(1 + \ln N) N^{-1},$$

and

$$(2.12) \quad \|u - u_\theta\|_{H^1(\mathcal{T} \times \mathcal{D} \times \Omega^+)} \leq C(1 + \ln N) N^{-1},$$

where $u = j, w$. For brevity of notation, we define

$$d_\theta^{(1)}(t, x) := \partial_t \rho_\theta + \langle v \partial_x j_\theta \rangle.$$

Since $\partial_t \rho + \langle v \partial_x j \rangle = 0$, subtract $d_\theta^{(1)}(t, x)$ by the above equation, then

$$(2.13) \quad d_\theta^{(1)}(t, x) = \partial_t(\rho_\theta - \rho) + \langle v \partial_x(j_\theta - j) \rangle.$$

By integrating $|d_\theta^{(1)}|^2$ over $\mathcal{T} \times \mathcal{D}$, it yields that

$$(2.14) \quad \begin{aligned} & \|\partial_t(\rho_\theta - \rho) + \langle v \partial_x(j_\theta - j) \rangle\|_{L^2(\mathcal{T} \times \mathcal{D})}^2 \\ & \leq \|\partial_t(\rho - \rho_\theta)\|_{L^2(\mathcal{T} \times \mathcal{D})}^2 + \|\partial_t(j - j_\theta)\|_{L^2(\mathcal{T} \times \mathcal{D} \times \Omega^+)}^2 \leq C(1 + \ln N)^2 N^{-2}, \end{aligned}$$

which corresponds to the estimate of the first term in the loss function $\mathcal{R}_{\text{residual}}^\varepsilon$. For the other terms, denote

$$(2.15) \quad d_\theta^{(2)}(t, x) := \varepsilon^2 \partial_t j_\theta + j_\theta + v \partial_x \rho_\theta + \varepsilon^2 v \partial_x w_\theta,$$

and

$$(2.16) \quad d_\theta^{(3)}(t, x) := \varepsilon^2 \partial_t w_\theta + w_\theta + v \partial_x j_\theta - \langle v \partial_x j_\theta \rangle.$$

Since ρ, j , and w are the solutions of the linear system (2.9), we subtract $d_\theta^{(2)}$ and $d_\theta^{(3)}$ from the second and last equations of (2.9), respectively, to obtain

$$\begin{cases} d_\theta^{(2)}(t, x) = \varepsilon^2 \partial_t (j_\theta - j) + (j_\theta - j) + v \partial_x (\rho_\theta - \rho) + \varepsilon^2 v \partial_x (w_\theta - w), \\ d_\theta^{(3)}(t, x) = \varepsilon^2 \partial_t (w_\theta - w) + (w_\theta - w) + v \partial_x (j_\theta - j) - \langle v \partial_x (j_\theta - j) \rangle. \end{cases}$$

Integrating $|d_\theta^{(2)}|^2$ and $|d_\theta^{(3)}|^2$ over $\mathcal{T} \times \mathcal{D} \times \Omega^+$, one gets

$$\begin{aligned} & \|\varepsilon^2 \partial_t (j_\theta - j) + (j_\theta - j) + v \partial_x (\rho_\theta - \rho) + \varepsilon^2 v \partial_x (w_\theta - w)\|_{L^2(\mathcal{T} \times \mathcal{D} \times \Omega^+)}^2 \\ & \leq \varepsilon^2 \|\partial_t (j_\theta - j)\|_{L^2(\mathcal{T} \times \mathcal{D} \times \Omega^+)}^2 + \|j_\theta - j\|_{L^2(\mathcal{T} \times \mathcal{D} \times \Omega^+)}^2 \\ & \quad + \|\partial_x (\rho_\theta - \rho)\|_{L^2(\mathcal{T} \times \mathcal{D})}^2 + \varepsilon^2 \|\partial_x (w_\theta - w)\|_{L^2(\mathcal{T} \times \mathcal{D} \times \Omega^+)}^2, \end{aligned}$$

and

$$\begin{aligned} & \|\varepsilon^2 \partial_t (w_\theta - w) + (w_\theta - w) + v \partial_x (j_\theta - j) - \langle v \partial_x (j_\theta - j) \rangle\|_{L^2(\mathcal{T} \times \mathcal{D} \times \Omega^+)}^2 \\ & \leq \varepsilon^2 \|\partial_t (w_\theta - w)\|_{L^2(\mathcal{T} \times \mathcal{D} \times \Omega^+)}^2 + \|w_\theta - w\|_{L^2(\mathcal{T} \times \mathcal{D} \times \Omega^+)}^2 + 2 \|\partial_x (j_\theta - j)\|_{L^2(\mathcal{T} \times \mathcal{D} \times \Omega^+)}^2. \end{aligned}$$

Thus from (2.12), we obtain

$$(2.17) \quad \|d_\theta^{(2)}\|_{L^2(\mathcal{T} \times \mathcal{D} \times \Omega^+)}^2 + \|d_\theta^{(3)}\|_{L^2(\mathcal{T} \times \mathcal{D} \times \Omega^+)}^2 \leq C(1 + \ln N)^2 N^{-2}.$$

Combining the estimates (2.14) and (2.17) to get

$$(2.18) \quad \mathcal{R}_{\text{residual}}^\varepsilon \leq C(1 + \ln N)^2 N^{-2}.$$

For the first term of the initial residual, using the trace inequality to get

$$\begin{aligned} \|\rho_\theta - \rho_{\text{IC}}\|_{L^2(\mathcal{D})} & \leq \|\rho_\theta - \rho\|_{L^2(\partial(\mathcal{T} \times \mathcal{D}))} \\ & \leq \|\rho_\theta - \rho\|_{H^1(\mathcal{T} \times \mathcal{D})} \leq C(1 + \ln N) N^{-1}. \end{aligned}$$

Similarly, we can deduce that

$$\varepsilon \|j_\theta - j_{\text{IC}}\|_{L^2(\mathcal{D} \times \Omega^+)} + \varepsilon^2 \|w_\theta - w_{\text{IC}}\|_{L^2(\mathcal{D} \times \Omega^+)} \leq C(1 + \ln N) N^{-1}.$$

Then it yields that

$$(2.19) \quad \mathcal{R}_{\text{initial}}^\varepsilon \leq C(1 + \ln N)^2 N^{-2}.$$

Moreover, we can also estimate the boundary residual using trace inequalities to get

$$(2.20) \quad \mathcal{R}_{\text{boundary}}^\varepsilon \leq C(1 + \ln N)^2 N^{-2}.$$

Combining the estimates (2.18), (2.19), and (2.20), we conclude that

$$\mathcal{R}_{\text{APNN}}^\varepsilon = \mathcal{R}_{\text{residual}}^\varepsilon + \mathcal{R}_{\text{initial}}^\varepsilon + \mathcal{R}_{\text{boundary}}^\varepsilon \leq C(1 + \ln N)^2 N^{-2}. \quad \square$$

Therefore, this result is proven.

2.4. Uniform error analysis. Next, we prove that the APNN converges to the analytic solution of the RTE (2.1) uniformly with respect to ε . To quantify the convergence behavior, we define the error between analytic solutions for system (2.9) and APNN solutions as $\mathcal{R}_{\text{total}}^\varepsilon$ in the following form:

$$(2.21) \quad \mathcal{R}_{\text{total}}^\varepsilon = \|f - f_\theta\|_{L^2(\mathcal{T} \times \mathcal{D} \times \Omega^+)}^2.$$

The following theorem establishes a bound for this total error using direct techniques.

THEOREM 2.3. *Let $\rho \in H^1(\mathcal{T} \times \mathcal{D})$, $j, w \in H^1(\mathcal{T} \times \mathcal{D} \times \Omega^+)$ be the analytic solution to the multiscale parity decomposition system (2.9). Then*

$$\mathcal{R}_{\text{total}}^\varepsilon \leq \frac{C}{\varepsilon} \mathcal{R}_{\text{APNN}}^\varepsilon.$$

Proof. Recall the definition of $d_\theta^{(1)}, d_\theta^{(2)}, d_\theta^{(3)}$ in (2.13), (2.15), and (2.16). Denote $\tilde{\rho} = \rho_\theta - \rho$, $\tilde{j} = j_\theta - j$ and $\tilde{w} = w_\theta - w$, then we have

$$\begin{cases} \partial_t \tilde{\rho} + \langle v \partial_x \tilde{j} \rangle = d_\theta^{(1)}(t, x), \\ \varepsilon^2 \partial_t \tilde{j} + \tilde{j} + v \partial_x \tilde{\rho} + \varepsilon^2 v \partial_x \tilde{w} = d_\theta^{(2)}(t, x), \\ \varepsilon^2 \partial_t \tilde{w} + \tilde{w} + v \partial_x \tilde{j} - \langle v \partial_x \tilde{j} \rangle = d_\theta^{(3)}(t, x). \end{cases}$$

Based on the decomposition of f in (2.8), we use the notation $\tilde{f} = f - f_\theta$ with

$$(2.22) \quad \tilde{f} = \tilde{\rho} + \varepsilon \tilde{j} + \varepsilon^2 \tilde{w},$$

which is governed by

$$(2.23) \quad \varepsilon^2 \partial_t \tilde{f} + \varepsilon v \partial_x \tilde{f} = (\tilde{\rho} - \tilde{f}) + \varepsilon^2 d_\theta^{(1)} + \varepsilon d_\theta^{(2)} + \varepsilon^2 d_\theta^{(3)}.$$

Multiply the above equation by $\tilde{f}$ and integrate in $\mathcal{D} \times \Omega^+$ to get

$$\begin{aligned} \frac{\varepsilon^2}{2} \frac{d}{dt} \|\tilde{f}\|_{L^2(\mathcal{D} \times \Omega^+)}^2 + \int_{\mathcal{D}} \int_{\Omega^+} \varepsilon v \partial_x \tilde{f} \tilde{f} dv dx \\ = \int_{\mathcal{D}} \int_{\Omega^+} (\tilde{\rho} - \tilde{f}) \tilde{f} dv dx + \int_{\mathcal{D}} \int_{\Omega^+} (\varepsilon^2 d_\theta^{(1)} + \varepsilon d_\theta^{(2)} + \varepsilon^2 d_\theta^{(3)}) \tilde{f} dv dx. \end{aligned}$$

By integration by parts, we have

$$(2.24) \quad \int_{\mathcal{D}} \int_{\Omega^+} \varepsilon v \partial_x \tilde{f} \tilde{f} dv dx = 0.$$

Since $\tilde{\rho} = \langle \tilde{f} \rangle$, it follows that

$$(2.25) \quad \int_{\mathcal{D}} \int_{\Omega^+} (\tilde{\rho} - \tilde{f}) \tilde{f} dv dx \leq 0.$$

In addition, by applying Young's inequality, we obtain

$$\begin{aligned} \int_{\mathcal{D}} \int_{\Omega^+} (\varepsilon^2 d_\theta^{(1)} + \varepsilon d_\theta^{(2)} + \varepsilon^2 d_\theta^{(3)}) \tilde{f} dv dx \\ \leq \varepsilon^2 \|\tilde{f}\|_{L^2(\mathcal{D} \times \Omega^+)}^2 + \varepsilon^2 \|d_\theta^{(1)}\|_{L^2(\mathcal{D})}^2 + \varepsilon \|d_\theta^{(2)}\|_{L^2(\mathcal{D} \times \Omega^+)}^2 + \varepsilon^2 \|d_\theta^{(3)}\|_{L^2(\mathcal{D} \times \Omega^+)}^2. \end{aligned}$$

Hence, we can conclude that

$$\frac{d}{dt} \|\tilde{f}\|_{L^2(\mathcal{D} \times \Omega^+)}^2 \leq 2\|\tilde{f}\|_{L^2(\mathcal{D} \times \Omega^+)}^2 + 2\|d_\theta^{(1)}\|_{L^2(\mathcal{D})}^2 + \frac{2}{\varepsilon} \|d_\theta^{(2)}\|_{L^2(\mathcal{D} \times \Omega^+)}^2 + 2\|d_\theta^{(3)}\|_{L^2(\mathcal{D} \times \Omega^+)}^2.$$

Therefore, by Gronwall's inequality, we prove this result. $\square$

Unfortunately, the above result indicates that the error may become unbounded as ε approaches zero. To overcome this limitation and obtain a uniform bound, we proceed with a more refined error analysis. Recall the decomposition of f in (2.8), and apply triangle inequality to get the relation:

$$\mathcal{R}_{\text{total}}^\varepsilon \leq \|\rho - \rho_\theta\|_{L^2(\mathcal{T} \times \mathcal{D})}^2 + \varepsilon^2 \|j - j_\theta\|_{L^2(\mathcal{T} \times \mathcal{D} \times \Omega^+)}^2 + \varepsilon^4 \|w - w_\theta\|_{L^2(\mathcal{T} \times \mathcal{D} \times \Omega^+)}^2.$$

Therefore, instead of estimating the total error directly, we estimate each term on the right-hand side of the above equation.

THEOREM 2.4. *Let $\rho \in H^1(\mathcal{T} \times \mathcal{D})$, $j, w \in H^1(\mathcal{T} \times \mathcal{D} \times \Omega^+)$ be the analytic solution to the multiscale parity decomposition system (2.9). Then*

$$\mathcal{R}_{\text{total}}^\varepsilon \leq C(\mathcal{R}_{\text{APNN}}^\varepsilon + \varepsilon^2),$$

where the constant $C > 0$ is independent of ε .

Proof. We follow the notation established in the proof of Theorem 2.3. For the sake of clarity, we recall the system of equations:

$$(2.26) \quad \begin{cases} \partial_t \tilde{\rho} + \langle v \partial_x \tilde{j} \rangle = d_\theta^{(1)}(t, x), \\ \varepsilon^2 \partial_t \tilde{j} + \tilde{j} + v \partial_x \tilde{\rho} + \varepsilon^2 v \partial_x \tilde{w} = d_\theta^{(2)}(t, x), \\ \varepsilon^2 \partial_t \tilde{w} + \tilde{w} + v \partial_x \tilde{j} - \langle v \partial_x \tilde{j} \rangle = d_\theta^{(3)}(t, x). \end{cases}$$

Here, the notations $d_\theta^{(1)}, d_\theta^{(2)}, d_\theta^{(3)}$ are defined in (2.13), (2.15), and (2.16). Multiply (2.26)₁ by $\tilde{\rho}$ and integrate over $\mathcal{D}$ to deduce that

$$(2.27) \quad \frac{1}{2} \frac{d}{dt} \|\tilde{\rho}\|_{L^2(\mathcal{D})}^2 + \int_{\mathcal{D}} \int_{\Omega^+} v \partial_x \tilde{j} \tilde{\rho} dv dx = \int_{\mathcal{D}} d_\theta^{(1)}(t, x) \tilde{\rho} dx.$$

Multiply (2.26)₂, (2.26)₃ by $\tilde{j}$, $\varepsilon^2 \tilde{w}$, respectively, and integrate over $\mathcal{D} \times \Omega^+$ to get

$$(2.28) \quad \begin{aligned} & \frac{\varepsilon^2}{2} \frac{d}{dt} \|\tilde{j}\|_{L^2(\mathcal{D} \times \Omega^+)}^2 + \|\tilde{j}\|_{L^2(\mathcal{D} \times \Omega^+)}^2 \\ & + \int_{\mathcal{D}} \int_{\Omega^+} v \partial_x (\tilde{\rho} + \varepsilon^2 \tilde{w}) \tilde{j} dv dx = \int_{\mathcal{D}} \int_{\Omega^+} d_\theta^{(2)}(t, x) \tilde{j} dv dx, \end{aligned}$$

and

$$(2.29) \quad \begin{aligned} & \frac{\varepsilon^4}{2} \frac{d}{dt} \|\tilde{w}\|_{L^2(\mathcal{D} \times \Omega^+)}^2 + \varepsilon^2 \|\tilde{w}\|_{L^2(\mathcal{D} \times \Omega^+)}^2 + \varepsilon^2 \int_{\mathcal{D}} \int_{\Omega^+} v \partial_x \tilde{j} \tilde{w} dv dx \\ & = \varepsilon^2 \int_{\mathcal{D}} \int_{\Omega^+} d_\theta^{(3)}(t, x) \tilde{w} dv dx + \varepsilon^2 \int_{\mathcal{D}} \int_{\Omega^+} \langle v \partial_x \tilde{j} \rangle \tilde{w} dv dx. \end{aligned}$$

Notice that combining the cross term of the left-hand side in the above three equalities

and deriving

$$\begin{aligned}
 (2.30) \quad & \int_{\mathcal{D}} \int_{\Omega^+} v \partial_x \tilde{j} \tilde{\rho} \, dv dx + \int_{\mathcal{D}} \int_{\Omega^+} v \partial_x (\tilde{\rho} + \varepsilon^2 \tilde{w}) \tilde{j} \, dv dx + \varepsilon^2 \int_{\mathcal{D}} \int_{\Omega^+} v \partial_x \tilde{j} \tilde{w} \, dv dx \\
 &= \int_{\mathcal{D}} \int_{\Omega^+} v \partial_x \tilde{j} (\tilde{\rho} + \varepsilon^2 \tilde{w}) \, dv dx + \int_{\mathcal{D}} \int_{\Omega^+} v \partial_x (\tilde{\rho} + \varepsilon^2 \tilde{w}) \tilde{j} \, dv dx \\
 &= \int_{\mathcal{D}} \int_{\Omega^+} v \partial_x (\tilde{j} (\tilde{\rho} + \varepsilon^2 \tilde{w})) \, dv dx = 0,
 \end{aligned}$$

where the last equality holds since the term vanishes under periodic boundary conditions. In addition, we estimate the right-hand terms of equations (2.27), (2.28), and (2.29) one by one. By Young's inequality, one gets

$$\begin{aligned}
 & \int_{\mathcal{D}} d_{\theta}^{(1)}(t, x) \tilde{\rho} \, dx + \int_{\mathcal{D}} \int_{\Omega^+} d_{\theta}^{(2)}(t, x) \tilde{\rho} \, dv dx + \varepsilon^2 \int_{\mathcal{D}} \int_{\Omega^+} d_{\theta}^{(3)}(t, x) \tilde{w} \, dv dx \\
 & \leq \frac{1}{2} \left(\|d_{\theta}^{(1)}(t, x)\|_{L^2(\mathcal{D})}^2 + \|d_{\theta}^{(2)}(t, x)\|_{L^2(\mathcal{D} \times \Omega^+)}^2 + \|d_{\theta}^{(3)}(t, x)\|_{L^2(\mathcal{D} \times \Omega^+)}^2 \right) \\
 & \quad + \frac{1}{2} \left(\|\tilde{\rho}\|_{L^2(\mathcal{D})}^2 + \|\tilde{j}\|_{L^2(\mathcal{D} \times \Omega^+)}^2 + \varepsilon^4 \|\tilde{w}\|_{L^2(\mathcal{D} \times \Omega^+)}^2 \right),
 \end{aligned}$$

and

$$\varepsilon^2 \int_{\mathcal{D}} \int_{\Omega^+} \langle v \partial_x \tilde{j} \rangle \tilde{w} \, dv dx \leq \frac{\varepsilon^2}{2} \|\partial_x \tilde{j}\|_{L^2(\mathcal{D} \times \Omega^+)}^2 + \frac{\varepsilon^2}{2} \|\tilde{w}\|_{L^2(\mathcal{D} \times \Omega^+)}^2.$$

Therefore, we can conclude that □

$$\begin{aligned}
 & \frac{d}{dt} \left(\|\tilde{\rho}\|_{L^2(\mathcal{D})}^2 + \varepsilon^2 \|\tilde{j}\|_{L^2(\mathcal{D} \times \Omega^+)}^2 + \varepsilon^4 \|\tilde{w}\|_{L^2(\mathcal{D} \times \Omega^+)}^2 \right) + \|\tilde{j}\|_{L^2(\mathcal{D} \times \Omega^+)}^2 + \varepsilon^2 \|\tilde{w}\|_{L^2(\mathcal{D} \times \Omega^+)}^2 \\
 & \leq \left(\|d_{\theta}^{(1)}(t, x)\|_{L^2(\mathcal{D})}^2 + \|d_{\theta}^{(2)}(t, x)\|_{L^2(\mathcal{D} \times \Omega^+)}^2 + \|d_{\theta}^{(3)}(t, x)\|_{L^2(\mathcal{D} \times \Omega^+)}^2 \right) \\
 & \quad + \varepsilon^2 \|\partial_x \tilde{j}\|_{L^2(\mathcal{D} \times \Omega^+)}^2 + \left(\|\tilde{\rho}\|_{L^2(\mathcal{D})}^2 + \varepsilon^4 \|\tilde{w}\|_{L^2(\mathcal{D} \times \Omega^+)}^2 \right).
 \end{aligned}$$

An application of Gronwall's inequality concludes the proof.

3. Two-dimensional method. Building upon the one-dimensional analysis in the previous section, we now generalize the framework to the two-dimensional case. Consider the two-dimensional RTE:

$$(3.1) \quad \varepsilon \partial_t f + \mathbf{v} \cdot \nabla_{\mathbf{x}} f = \frac{1}{\varepsilon} \left(\frac{1}{2\pi} \int_{|\mathbf{v}'|=1} f \, d\mathbf{v}' - f \right), \quad \mathbf{x} = (x_1, x_2) \in \mathcal{D} \subset \mathbb{R}^2, \quad |\mathbf{v}| = 1$$

with $\mathbf{v} = (\xi, \eta)$, $-1 \leq \xi, \eta \leq 1$, $\xi^2 + \eta^2 = 1$. The periodic boundary is given by $f(t, \mathbf{x}, \mathbf{v}) = f_{\text{BC}}(\mathbf{x}, \mathbf{v})$. We assume the initial condition that $f(0, \mathbf{x}, \mathbf{v}) = f_{\text{IC}}(\mathbf{x}, \mathbf{v})$.

3.1. Multiscale parity decomposition method. We now introduce the multiscale parity decomposition for the two-dimensional RTE. Similarly to the one-dimensional case, for convenience of representation, define the set of all positive ξ and η as

$$(3.2) \quad \Omega^+ = \{\xi > 0, \eta > 0 \mid \xi^2 + \eta^2 = 1, -1 \leq \xi, \eta \leq 1\}.$$

We rewrite (3.1) in $\xi, \eta \in \Omega^+$ only:

$$\begin{aligned}\varepsilon \partial_t f(\xi, \eta) + \xi \partial_x f(\xi, \eta) + \eta \partial_y f(\xi, \eta) &= \frac{1}{\varepsilon} (\rho - f(\xi, \eta)), \\ \varepsilon \partial_t f(-\xi, -\eta) - \xi \partial_x f(-\xi, -\eta) - \eta \partial_y f(-\xi, -\eta) &= \frac{1}{\varepsilon} (\rho - f(-\xi, -\eta)), \\ \varepsilon \partial_t f(\xi, -\eta) + \xi \partial_x f(\xi, -\eta) - \eta \partial_y f(\xi, -\eta) &= \frac{1}{\varepsilon} (\rho - f(\xi, -\eta)), \\ \varepsilon \partial_t f(-\xi, \eta) - \xi \partial_x f(-\xi, \eta) + \eta \partial_y f(-\xi, \eta) &= \frac{1}{\varepsilon} (\rho - f(-\xi, \eta)),\end{aligned}$$

where

$$\rho = \frac{1}{2\pi} \int_{|\mathbf{v}|=1} f \, d\mathbf{v}'.$$

Introducing the even and odd parities respect to the the vector (ξ, η) defined as follows

$$\begin{aligned}r_1(\xi, \eta) &= [f(\xi, -\eta) + f(-\xi, \eta)]/2, \\ r_2(\xi, \eta) &= [f(\xi, \eta) + f(-\xi, -\eta)]/2, \\ j_1(\xi, \eta) &= [f(\xi, -\eta) - f(-\xi, \eta)]/2\varepsilon, \\ j_2(\xi, \eta) &= [f(\xi, \eta) - f(-\xi, -\eta)]/2\varepsilon,\end{aligned}$$

thus we arrive at the system

$$(3.3) \quad \begin{cases} \varepsilon^2 \partial_t r_1 + \varepsilon^2 \xi \partial_x j_1 - \varepsilon^2 \eta \partial_y j_1 = \rho - r_1, \\ \varepsilon^2 \partial_t r_2 + \varepsilon^2 \xi \partial_x j_2 + \varepsilon^2 \eta \partial_y j_2 = \rho - r_2, \\ \varepsilon^2 \partial_t j_1 + \xi \partial_x r_1 - \eta \partial_y r_1 = -j_1, \\ \varepsilon^2 \partial_t j_2 + \xi \partial_x r_2 + \eta \partial_y r_2 = -j_2. \end{cases}$$

In addition, note the fact:

$$(3.4) \quad \rho = \frac{2}{\pi} \int_{\Omega^+} \left(\frac{r_1 + r_2}{2} \right) d\mathbf{v}' := \left\langle \frac{r_1 + r_2}{2} \right\rangle,$$

which satisfies the following equation:

$$(3.5) \quad 2\partial_t \rho + \langle \xi \partial_x (j_1 + j_2) + \eta \partial_y (j_2 - j_1) \rangle = 0.$$

Define $\frac{r_1 + r_2}{2} = \rho + \varepsilon^2 w$, then we have

$$(3.6) \quad \begin{aligned} 2\varepsilon^2 \partial_t w + (\xi \partial_x (j_1 + j_2) + \eta \partial_y (j_2 - j_1)) \\ - \langle \xi \partial_x (j_1 + j_2) + \eta \partial_y (j_2 - j_1) \rangle + 2w = 0. \end{aligned}$$

Furthermore, denote $\varphi = r_2 - r_1$, subtract (3.3)₂ from (3.3)₁ to deduce that

$$(3.7) \quad \varepsilon^2 \partial_t \varphi + \varepsilon^2 \xi \partial_x (j_2 - j_1) + \varepsilon^2 \eta \partial_y (j_1 + j_2) + \varphi = 0.$$

Adding and subtracting (3.3)₃ and (3.3)₄ yields

$$(3.8) \quad \varepsilon^2 \partial_t (j_1 + j_2) + (j_1 + j_2) + 2\xi \partial_x (\rho + \varepsilon^2 w) + \eta \partial_y \varphi = 0,$$

and

$$(3.9) \quad \varepsilon^2 \partial_t (j_2 - j_1) + (j_2 - j_1) + \xi \partial_x \varphi + 2\eta \partial_y (\rho + \varepsilon^2 w) = 0.$$

Thus, similarly to one dimension, we use the multiscale parity decomposition of the distribution function $f(t, \mathbf{x}, \mathbf{v})$, explicitly separating the macroscopic density $\rho(t, \mathbf{x})$, $\varphi(t, \mathbf{x}, \mathbf{v})$ at $\mathcal{O}(1)$, the odd term $j_1(t, \mathbf{x}, \mathbf{v}) + j_2(t, \mathbf{x}, \mathbf{v})$, $j_2(t, \mathbf{x}, \mathbf{v}) - j_1(t, \mathbf{x}, \mathbf{v})$ at $\mathcal{O}(\varepsilon)$, and higher-order corrections $w(t, \mathbf{x}, \mathbf{v})$ at $\mathcal{O}(\varepsilon^2)$, each with definite parity properties, that is,

$$(3.10) \quad f(t, \mathbf{x}, \mathbf{v}) = \underbrace{\rho(t, \mathbf{x}) + \frac{\varphi(t, \mathbf{x}, \mathbf{v})}{2}}_{\mathcal{O}(1)} + \underbrace{\varepsilon \left(\frac{j_2(t, \mathbf{x}, \mathbf{v}) + j_1(t, \mathbf{x}, \mathbf{v})}{2} + \frac{j_2(t, \mathbf{x}, \mathbf{v}) - j_1(t, \mathbf{x}, \mathbf{v})}{2} \right)}_{\mathcal{O}(\varepsilon)} + \underbrace{\varepsilon^2 w(t, \mathbf{x}, \mathbf{v})}_{\mathcal{O}(\varepsilon^2)}.$$

Hence, equations (3.5)-(3.9) constitute the multiscale parity decomposition framework of equation (3.1) to be used for the APNN:

$$(3.11) \quad \begin{cases} 2\partial_t \rho + \langle \xi \partial_x (j_1 + j_2) + \eta \partial_y (j_2 - j_1) \rangle = 0 \\ 2\varepsilon^2 \partial_t w + (\xi \partial_x (j_1 + j_2) + \eta \partial_y (j_2 - j_1)) \\ \quad - \langle \xi \partial_x (j_1 + j_2) + \eta \partial_y (j_2 - j_1) \rangle + 2w = 0 \\ \varepsilon^2 \partial_t \varphi + \varepsilon^2 \xi \partial_x (j_2 - j_1) + \varepsilon^2 \eta \partial_y (j_1 + j_2) + \varphi = 0 \\ \varepsilon^2 \partial_t (j_1 + j_2) + (j_1 + j_2) + 2\xi \partial_x (\rho + \varepsilon^2 w) + \eta \partial_y \varphi = 0 \\ \varepsilon^2 \partial_t (j_2 - j_1) + (j_2 - j_1) + \xi \partial_x \varphi + 2\eta \partial_y (\rho + \varepsilon^2 w) = 0 \end{cases}$$

with the constraint $\langle w \rangle = 0$.

As $\varepsilon \rightarrow 0$, one has

$$(3.12) \quad \varphi = 0, \quad j_1 + j_2 = -2\xi \partial_x \rho - \eta \partial_y \varphi, \quad j_2 - j_1 = -2\eta \partial_y \rho - \xi \partial_x \varphi,$$

and

$$(3.13) \quad 2w = -(\xi \partial_x (j_1 + j_2) + \eta \partial_y (j_2 - j_1)) + \langle \xi \partial_x (j_1 + j_2) + \eta \partial_y (j_2 - j_1) \rangle.$$

Applying (3.12) in (3.11)₁ gives

$$\partial_t \rho - \langle \xi^2 \partial_{xx} \rho + \eta^2 \partial_{yy} \rho \rangle = 0.$$

Adding the above two equations and integrating over $\xi^2 + \eta^2 = 1$, we obtain the diffusion equation

$$(3.14) \quad \partial_t \rho = \frac{1}{2} (\partial_{xx} \rho + \partial_{yy} \rho).$$

Furthermore, applying (3.12) and (3.14) in (3.13) yields

$$w = \left(\xi^2 - \frac{1}{2} \right) \partial_{xx} \rho + \left(\eta^2 - \frac{1}{2} \right) \partial_{yy} \rho.$$

3.2. Loss function. We denote by ρ_θ and φ_θ the network approximations to the macroscopic variables ρ and φ , and by $j_{1,\theta}$, $j_{2,\theta}$, w_θ the approximations to the microscopic fluxes j_1 , j_2 , and w , all corresponding to system (3.11). Here, θ collectively represents all trainable parameters. For the APNN method, the physics-informed loss is defined as the residual of the multiscale parity framework in system (3.11). Specifically, we denote

$$\begin{aligned} d_\theta^{(1)} &:= \partial_t \rho_\theta + \langle \xi \partial_x (j_{1,\theta} + j_{2,\theta}) + \eta \partial_y (j_{2,\theta} - j_{1,\theta}) \rangle, \\ d_\theta^{(2)} &:= 2\varepsilon^2 \partial_t w_\theta + (\xi \partial_x (j_{1,\theta} + j_{2,\theta}) + \eta \partial_y (j_{2,\theta} - j_{1,\theta})) \\ &\quad - \langle \xi \partial_x (j_{1,\theta} + j_{2,\theta}) + \eta \partial_y (j_{2,\theta} - j_{1,\theta}) \rangle + 2w_\theta, \\ d_\theta^{(3)} &:= \partial_t \varphi_\theta + \xi \partial_x (j_{2,\theta} - j_{1,\theta}) + \eta \partial_y (j_{1,\theta} + j_{2,\theta}) + \frac{1}{\varepsilon^2} \varphi_\theta, \\ d_\theta^{(4)} &:= \varepsilon^2 \partial_t (j_{1,\theta} + j_{2,\theta}) + (j_{1,\theta} + j_{2,\theta}) + 2\xi \partial_x (\rho_\theta + \varepsilon^2 w_\theta) + \eta \partial_y \varphi_\theta, \\ d_\theta^{(5)} &:= \varepsilon^2 \partial_t (j_{2,\theta} - j_{1,\theta}) + (j_{2,\theta} - j_{1,\theta}) + \xi \partial_x \varphi_\theta + 2\eta \partial_y (\rho_\theta + \varepsilon^2 w_\theta). \end{aligned}$$

The corresponding loss function is then defined as:

$$\mathcal{R}_{\text{APNN}}^\varepsilon = \mathcal{R}_{\text{residual}}^\varepsilon + \mathcal{R}_{\text{initial}}^\varepsilon + \mathcal{R}_{\text{boundary}}^\varepsilon,$$

where the terms on the right-hand side of the above equation are defined as

$$\mathcal{R}_{\text{residual}}^\varepsilon = \frac{1}{|\mathcal{T} \times \mathcal{D}|} \int_{\mathcal{T}} \int_{\mathcal{D}} |d_\theta^{(1)}|^2 d\mathbf{x} dt + \frac{1}{|\mathcal{T} \times \mathcal{D} \times \Omega^+|} \sum_{j=2}^5 \int_{\mathcal{T}} \int_{\mathcal{D}} \int_{\Omega^+} |d_\theta^{(j)}|^2 d\mathbf{v} d\mathbf{x} dt,$$

$$\begin{aligned} \mathcal{R}_{\text{initial}}^\varepsilon &= \frac{\lambda_1}{|\mathcal{D}|} \int_{\mathcal{D}} |\rho_\theta - \rho_{\text{IC}}|^2 d\mathbf{x} + \frac{\lambda_2}{|\mathcal{D} \times \Omega^+|} \int_{\mathcal{D}} \int_{\Omega^+} \varepsilon^4 |w_\theta - w_{\text{IC}}|^2 + |\varphi_\theta - \varphi_{\text{IC}}|^2 \\ &\quad + \varepsilon^2 (|(j_{1,\theta} + j_{2,\theta}) - (j_{1,\text{IC}} + j_{2,\text{IC}})|^2 + |(j_{2,\theta} - j_{1,\theta}) - (j_{2,\text{IC}} - j_{1,\text{IC}})|^2) d\mathbf{v} d\mathbf{x}, \end{aligned}$$

and

$$\begin{aligned} \mathcal{R}_{\text{boundary}}^\varepsilon &= \frac{\lambda_3}{|\mathcal{T} \times \partial \mathcal{D}|} \int_{\mathcal{T}} \int_{\partial \mathcal{D}} |\rho_\theta - \rho_{\text{BC}}|^2 ds dt \\ &\quad + \frac{\lambda_4}{|\mathcal{T} \times \partial \mathcal{D} \times \Omega^+|} \int_{\mathcal{T}} \int_{\partial \mathcal{D}} \int_{\Omega^+} |w_\theta - w_{\text{BC}}|^2 + |\varphi_\theta - \varphi_{\text{BC}}|^2 + |(j_{1,\theta} + j_{2,\theta}) \\ &\quad - (j_{1,\text{BC}} + j_{2,\text{BC}})|^2 + |(j_{2,\theta} - j_{1,\theta}) - (j_{2,\text{BC}} - j_{1,\text{BC}})|^2 d\mathbf{v} ds dt. \end{aligned}$$

Here, λ_i ($i = 1, \dots, 4$) are hyperparameters as penalty terms. As $\varepsilon \rightarrow 0$, $\mathcal{R}_{\text{residual}}^\varepsilon$ converges to

$$\begin{aligned} \mathcal{R}_{\text{residual}}^0 &:= \frac{1}{|\mathcal{T} \times \mathcal{D}|} \int_{\mathcal{T}} \int_{\mathcal{D}} |d_\theta^{(1)}|^2 d\mathbf{x} dt + \frac{1}{|\mathcal{T} \times \mathcal{D} \times \Omega^+|} \int_{\mathcal{T}} \int_{\mathcal{D}} \int_{\Omega^+} |\varphi_\theta|^2 d\mathbf{v} d\mathbf{x} dt \\ &\quad + \frac{1}{|\mathcal{T} \times \mathcal{D} \times \Omega^+|} \sum_{j=2,4,5} \int_{\mathcal{T}} \int_{\mathcal{D}} \int_{\Omega^+} \lim_{\varepsilon \rightarrow 0} |d_\theta^{(j)}|^2 d\mathbf{v} d\mathbf{x} dt, \end{aligned}$$

which is exactly the loss for the limit equations (3.5)–(3.9). Here

$$\begin{aligned}\lim_{\varepsilon \rightarrow 0} d_{\theta}^{(2)} &= (\xi \partial_x(j_{1,\theta} + j_{2,\theta}) + \eta \partial_y(j_{2,\theta} - j_{1,\theta})) \\ &\quad - \langle \xi \partial_x(j_{1,\theta} + j_{2,\theta}) + \eta \partial_y(j_{2,\theta} - j_{1,\theta}) \rangle + 2\omega_{\theta}, \\ \lim_{\varepsilon \rightarrow 0} d_{\theta}^{(4)} &= (j_{1,\theta} + j_{2,\theta}) + 2\xi \partial_x \rho_{\theta} + \eta \partial_y \varphi_{\theta}, \\ \lim_{\varepsilon \rightarrow 0} d_{\theta}^{(5)} &= (j_{2,\theta} - j_{1,\theta}) + \xi \partial_x \varphi_{\theta} + 2\eta \partial_y \rho_{\theta}.\end{aligned}$$

Analogous to the one-dimensional case, the approximation property likewise ensures the convergence of the APNN loss function in two dimensions.

THEOREM 3.1. *Let $\rho \in H^2(\mathcal{T} \times \mathcal{D})$, $\varphi, j_1, j_2, w \in H^2(\mathcal{T} \times \mathcal{D} \times \Omega)$ be the analytic solution to the multiscale parity decomposition system (3.11), and $N > 5$ be an integer. Then there exist APNN solutions $\hat{\rho}^N, \hat{\varphi}^N, \hat{j}_1^N, \hat{j}_2^N, \hat{w}^N$ such that*

$$\mathcal{R}_{\text{APNN}}^{\varepsilon} \rightarrow 0, \quad \text{as } N \rightarrow \infty.$$

The proof proceeds along similar lines to that of Theorem 2.2, and is therefore omitted for brevity.

3.3. Uniform error method. In the following, we demonstrate that the APNN solution converges uniformly to the analytical solution of the two-dimensional RTE (3.1). Parallel to our approach in one dimension, the total error $\mathcal{R}_{\text{total}}^{\varepsilon}$ for the 2D system (3.11) is defined as the difference between the analytic and APNN solutions:

$$(3.15) \quad \mathcal{R}_{\text{total}}^{\varepsilon} = \|f - f_{\theta}\|_{L^2(\mathcal{T} \times \mathcal{D} \times \Omega^+)}^2.$$

According to the decomposition (3.10), it yields that

$$\begin{aligned}\mathcal{R}_{\text{total}}^{\varepsilon} &\leq \|\rho - \rho_{\theta}\|_{L^2(\mathcal{D})}^2 + \|\varphi - \varphi_{\theta}\|_{L^2(\mathcal{D} \times \Omega^+)}^2 + \varepsilon^4 \|w - w_{\theta}\|_{L^2(\mathcal{D} \times \Omega^+)}^2 \\ &\quad + \varepsilon^2 \|(j_1 + j_2) - (j_{1,\theta} + j_{2,\theta})\|_{L^2(\mathcal{D} \times \Omega^+)}^2 + \varepsilon^2 \|(j_2 - j_1) - (j_{2,\theta} - j_{1,\theta})\|_{L^2(\mathcal{D} \times \Omega^+)}^2.\end{aligned}$$

Through a direct approach, a bound on the total error is established in the following theorem.

THEOREM 3.2. *Let $\rho \in H^1(\mathcal{T} \times \mathcal{D})$, $\varphi, j_1, j_2, w \in H^1(\mathcal{T} \times \mathcal{D} \times \Omega^+)$ be the analytic solution to the multiscale parity decomposition system (3.11). Then*

$$\mathcal{R}_{\text{total}}^{\varepsilon} \leq \frac{C}{\varepsilon} \mathcal{R}_{\text{APNN}}^{\varepsilon},$$

where the constant $C > 0$ is independent of ε .

Proof. Denote $\tilde{\rho} = \rho_{\theta} - \rho$, $\tilde{w} = w_{\theta} - w$, $\tilde{\varphi} = \varphi_{\theta} - \varphi$, and $\tilde{j}_i = j_{i,\theta} - j_i$ for $i = 1, 2$. Since ρ satisfies (3.5), we have

$$(3.16) \quad 2\partial_t \tilde{\rho} + \langle \xi \partial_x(\tilde{j}_1 + \tilde{j}_2) + \eta \partial_y(\tilde{j}_2 - \tilde{j}_1) \rangle = d_{\theta}^{(1)}.$$

Similarly, it yields that

$$\begin{aligned}(3.17) \quad 2\varepsilon^2 \partial_t \tilde{w} &+ (\xi \partial_x(\tilde{j}_1 + \tilde{j}_2) + \eta \partial_y(\tilde{j}_2 - \tilde{j}_1)) \\ &- \langle \xi \partial_x(\tilde{j}_1 + \tilde{j}_2) + \eta \partial_y(\tilde{j}_2 - \tilde{j}_1) \rangle + 2\tilde{w} = d_{\theta}^{(2)},\end{aligned}$$

$$(3.18) \quad \partial_t \tilde{\varphi} + \xi \partial_x (\tilde{j}_2 - \tilde{j}_1) + \eta \partial_y (\tilde{j}_1 + \tilde{j}_2) + \frac{1}{\varepsilon^2} \tilde{\varphi} = d_\theta^{(3)},$$

$$(3.19) \quad \varepsilon^2 \partial_t (\tilde{j}_1 + \tilde{j}_2) + (\tilde{j}_1 + \tilde{j}_2) + 2\xi \partial_x (\tilde{\rho} + \varepsilon^2 \tilde{w}) + \eta \partial_y \tilde{\varphi} = d_\theta^{(4)},$$

and

$$(3.20) \quad \varepsilon^2 \partial_t (\tilde{j}_2 - \tilde{j}_1) + (\tilde{j}_2 - \tilde{j}_1) + \xi \partial_x \tilde{\varphi} + 2\eta \partial_y (\tilde{\rho} + \varepsilon^2 \tilde{w}) = d_\theta^{(5)}.$$

Recalling the decomposition of f in (3.10), we use the notation $\tilde{f} = f - f_\theta$ with

$$\tilde{f} = \tilde{\rho} + \frac{\tilde{\varphi}}{2} + \varepsilon \left(\frac{\tilde{j}_2 + \tilde{j}_1}{2} + \frac{\tilde{j}_2 - \tilde{j}_1}{2} \right) + \varepsilon^2 \tilde{w}.$$

Hence, we deduce that

$$\varepsilon \partial_t \tilde{f} + \xi \partial_x \tilde{f} + \eta \partial_y \tilde{f} = \frac{1}{\varepsilon} (\tilde{\rho} - \tilde{f}) + \frac{1}{2} \left(\varepsilon \sum_{i=1}^3 d_\theta^{(i)} + d_\theta^{(4)} + d_\theta^{(5)} \right).$$

□

Observing that its structure is analogous to the one presented in the proof of Theorem 2.3, the desired estimate follows directly by analogous reasoning.

Building on this multiscale parity decomposition framework (3.11), we are motivated to establish an improved estimate for the error.

THEOREM 3.3. *Let $\rho \in H^1(\mathcal{T} \times \mathcal{D})$, $\varphi, j_1, j_2, w \in H^1(\mathcal{T} \times \mathcal{D} \times \Omega^+)$ be the analytic solution to the multiscale parity decomposition system (3.11). Then*

$$\mathcal{R}_{\text{total}}^\varepsilon \leq C(\mathcal{R}_{\text{APNN}}^\varepsilon + \varepsilon^2),$$

where the constant $C > 0$ is independent of ε .

Proof. Following Theorem 3.2, we continue to use the notations established in (3.16)-(3.19). Multiply (3.16) by $\tilde{\rho}$ and integrate over $\mathcal{D}$ to deduce that

$$(3.21) \quad \frac{d}{dt} \|\tilde{\rho}\|_{L^2(\mathcal{D})}^2 + \frac{2}{\pi} \int_{\mathcal{D}} \int_{\Omega^+} (\xi \partial_x (\tilde{j}_1 + \tilde{j}_2) + \eta \partial_y (\tilde{j}_2 - \tilde{j}_1)) \tilde{\rho} d\mathbf{v} d\mathbf{x} = \int_{\mathcal{D}} d_\theta^{(1)} \tilde{\rho} d\mathbf{x}.$$

We multiply (3.17), (3.18), (3.19), (3.20) by $\varepsilon^2 \tilde{w}$, $\tilde{\varphi}$, $\tilde{j}_1 + \tilde{j}_2$, and $\tilde{j}_2 - \tilde{j}_1$, respectively, and integrate over $\mathcal{D} \times \Omega^+$, we get

$$(3.22) \quad \begin{aligned} & \varepsilon^4 \frac{d}{dt} \|\tilde{w}\|_{L^2(\mathcal{D} \times \Omega^+)}^2 + 2\varepsilon^2 \|\tilde{w}\|_{L^2(\mathcal{D} \times \Omega^+)}^2 \\ & + \varepsilon^2 \int_{\mathcal{D}} \int_{\Omega^+} (\xi \partial_x (\tilde{j}_1 + \tilde{j}_2) + \eta \partial_y (\tilde{j}_2 - \tilde{j}_1)) \tilde{w} d\mathbf{v} d\mathbf{x} \\ & = \varepsilon^2 \int_{\mathcal{D}} \int_{\Omega^+} d_\theta^{(2)} \tilde{w} d\mathbf{v} d\mathbf{x} + \varepsilon^2 \int_{\mathcal{D}} \int_{\Omega^+} \langle \xi \partial_x (\tilde{j}_1 + \tilde{j}_2) + \eta \partial_y (\tilde{j}_2 - \tilde{j}_1) \rangle \tilde{w} d\mathbf{v} d\mathbf{x}, \end{aligned}$$

$$(3.23) \quad \begin{aligned} & \frac{1}{2} \frac{d}{dt} \|\tilde{\varphi}\|_{L^2(\mathcal{D} \times \Omega^+)}^2 + \frac{1}{\varepsilon^2} \|\tilde{\varphi}\|_{L^2(\mathcal{D} \times \Omega^+)}^2 \\ & + \int_{\mathcal{D}} \int_{\Omega^+} (\xi \partial_x (\tilde{j}_2 - \tilde{j}_1) + \eta \partial_y (\tilde{j}_1 + \tilde{j}_2)) \tilde{\varphi} d\mathbf{v} d\mathbf{x} = \int_{\mathcal{D}} \int_{\Omega^+} d_\theta^{(3)} \tilde{\varphi} d\mathbf{v} d\mathbf{x}, \end{aligned}$$

$$\begin{aligned}
(3.24) \quad & \frac{\varepsilon^2}{2} \frac{d}{dt} \|\tilde{j}_1 + \tilde{j}_2\|_{L^2(\mathcal{D} \times \Omega^+)}^2 + \int_{\mathcal{D}} \int_{\Omega^+} (2\xi \partial_x(\tilde{\rho} + \varepsilon^2 \tilde{w}) + \eta \partial_y \tilde{\varphi})(\tilde{j}_1 + \tilde{j}_2) \, d\mathbf{v} d\mathbf{x} \\
& + \|\tilde{j}_1 + \tilde{j}_2\|_{L^2(\mathcal{D} \times \Omega^+)}^2 = \int_{\mathcal{D}} \int_{\Omega^+} d_{\theta}^{(4)}(\tilde{j}_1 + \tilde{j}_2) \, d\mathbf{v} d\mathbf{x},
\end{aligned}$$

and

$$\begin{aligned}
(3.25) \quad & \frac{\varepsilon^2}{2} \frac{d}{dt} \|\tilde{j}_2 - \tilde{j}_1\|_{L^2(\mathcal{D} \times \Omega^+)}^2 + \int_{\mathcal{D}} \int_{\Omega^+} (\xi \partial_x \tilde{\varphi} + 2\eta \partial_y(\tilde{\rho} + \varepsilon^2 \tilde{w}))(\tilde{j}_2 - \tilde{j}_1) \, d\mathbf{v} d\mathbf{x} \\
& + \|\tilde{j}_2 - \tilde{j}_1\|_{L^2(\mathcal{D} \times \Omega^+)}^2 = \int_{\mathcal{D}} \int_{\Omega^+} d_{\theta}^{(5)}(\tilde{j}_2 - \tilde{j}_1) \, d\mathbf{v} d\mathbf{x},
\end{aligned}$$

We begin by estimating the cross-terms on the left-hand sides of (3.21)–(3.25). Define

$$\begin{aligned}
I_1 &:= 2(\xi \partial_x(\tilde{j}_1 + \tilde{j}_2) + \eta \partial_y(\tilde{j}_2 - \tilde{j}_1))\tilde{\rho} + 2\varepsilon^2(\xi \partial_x(\tilde{j}_1 + \tilde{j}_2) + \eta \partial_y(\tilde{j}_2 - \tilde{j}_1))\tilde{w} \\
&\quad + (\xi \partial_x(\tilde{j}_2 - \tilde{j}_1) + \eta \partial_y(\tilde{j}_1 + \tilde{j}_2))\tilde{\varphi}, \\
I_2 &:= (2\xi \partial_x(\tilde{\rho} + \varepsilon^2 \tilde{w}) + \eta \partial_y \tilde{\varphi})(\tilde{j}_1 + \tilde{j}_2) + (\xi \partial_x \tilde{\varphi} + 2\eta \partial_y(\tilde{\rho} + \varepsilon^2 \tilde{w}))(\tilde{j}_2 - \tilde{j}_1).
\end{aligned}$$

Then a direct computation shows

$$\begin{aligned}
I_1 + I_2 &= \xi \partial_x((\tilde{j}_2 - \tilde{j}_1)\tilde{\varphi}) + \eta \partial_y((\tilde{j}_1 + \tilde{j}_2)\tilde{\varphi}) \\
&\quad + 2\xi \partial_x((\tilde{\rho} + \varepsilon^2 \tilde{w})(\tilde{j}_1 + \tilde{j}_2)) + 2\eta \partial_y((\tilde{\rho} + \varepsilon^2 \tilde{w})(\tilde{j}_2 - \tilde{j}_1)),
\end{aligned}$$

and thus, by the divergence theorem and the periodic boundary condition, we have

$$(3.26) \quad \int_{\mathcal{D}} \int_{\Omega^+} I_1 + I_2 \, d\mathbf{v} d\mathbf{x} = 0.$$

Furthermore, by applying Young's inequality to the terms on the right-hand side of (3.21)–(3.23), we obtain

$$\begin{aligned}
& \pi \int_{\mathcal{D}} d_{\theta}^{(1)} \tilde{\rho} \, d\mathbf{x} + 2\varepsilon^2 \int_{\mathcal{D}} \int_{\Omega^+} d_{\theta}^{(2)} \tilde{w} \, d\mathbf{v} d\mathbf{x} + \int_{\mathcal{D}} \int_{\Omega^+} d_{\theta}^{(3)} \tilde{\varphi} \, d\mathbf{v} d\mathbf{x} \\
& \leq C \left(\|d_{\theta}^{(1)}\|_{L^2(\mathcal{D})}^2 + \|d_{\theta}^{(2)}\|_{L^2(\mathcal{D} \times \Omega^+)}^2 + \|d_{\theta}^{(3)}\|_{L^2(\mathcal{D} \times \Omega^+)}^2 \right) \\
& \quad + \left(\pi \|\tilde{\rho}\|_{L^2(\mathcal{D})}^2 + \varepsilon^4 \|\tilde{w}\|_{L^2(\mathcal{D} \times \Omega^+)}^2 + \|\tilde{\varphi}\|_{L^2(\mathcal{D} \times \Omega^+)}^2 \right),
\end{aligned}$$

and

$$\begin{aligned}
& 2\varepsilon^2 \int_{\mathcal{D}} \int_{\Omega^+} \langle \xi \partial_x(\tilde{j}_1 + \tilde{j}_2) + \eta \partial_y(\tilde{j}_2 - \tilde{j}_1) \rangle \tilde{w} \, d\mathbf{v} d\mathbf{x} \\
& \leq C\varepsilon^2 \left(\|\tilde{j}_1\|_{L^2(\mathcal{D} \times \Omega^+)}^2 + \|\tilde{j}_2\|_{L^2(\mathcal{D} \times \Omega^+)}^2 \right) + \varepsilon^2 \|\tilde{w}\|_{L^2(\mathcal{D} \times \Omega^+)}^2.
\end{aligned}$$

Similarly, for the right-hand side terms of (3.24), (3.25), we have

$$\begin{aligned}
& \int_{\mathcal{D}} \int_{\Omega^+} d_{\theta}^{(4)}(\tilde{j}_1 + \tilde{j}_2) \, d\mathbf{v} d\mathbf{x} + \int_{\mathcal{D}} \int_{\Omega^+} d_{\theta}^{(5)}(\tilde{j}_2 - \tilde{j}_1) \, d\mathbf{v} d\mathbf{x} \leq \frac{1}{2} \left(\|d_{\theta}^{(4)}\|_{L^2(\mathcal{D} \times \Omega^+)}^2 \right. \\
& \quad \left. + \|d_{\theta}^{(5)}\|_{L^2(\mathcal{D} \times \Omega^+)}^2 + \|\tilde{j}_1 + \tilde{j}_2\|_{L^2(\mathcal{D} \times \Omega^+)}^2 + \|\tilde{j}_2 - \tilde{j}_1\|_{L^2(\mathcal{D} \times \Omega^+)}^2 \right).
\end{aligned}$$

Hence, we conclude that

$$\begin{aligned} \frac{d}{dt}E(t) \leq & C \left(\|d_\theta^{(1)}\|_{L^2(\mathcal{D})}^2 + \sum_{j=2}^5 \|d_\theta^{(j)}\|_{L^2(\mathcal{D} \times \Omega^+)}^2 \right. \\ & \left. + \varepsilon^2 (\|\tilde{j}_1\|_{L^2(\mathcal{D} \times \Omega^+)}^2 + \|\tilde{j}_2\|_{L^2(\mathcal{D} \times \Omega^+)}^2) \right) + E(t), \end{aligned}$$

where

$$\begin{aligned} E(t) = & \pi \|\tilde{\rho}\|_{L^2(\mathcal{D})}^2 + \varepsilon^4 \|\tilde{w}\|_{L^2(\mathcal{D} \times \Omega^+)}^2 \\ & + \|\tilde{\varphi}\|_{L^2(\mathcal{D} \times \Omega^+)}^2 + \varepsilon^2 \|\tilde{j}_1 + \tilde{j}_2\|_{L^2(\mathcal{D} \times \Omega^+)}^2 + \varepsilon^2 \|\tilde{j}_2 - \tilde{j}_1\|_{L^2(\mathcal{D} \times \Omega^+)}^2. \end{aligned}$$

Therefore, an application of Gronwall's inequality concludes the proof. $\square$

4. Structure-preserving mechanism. The structure-preserving mechanism plays a crucial role in ensuring that the neural approximation retains the intrinsic symmetries and conservation properties of the AP system. In the context of neural networks, it is particularly important to guarantee that the learned representations are consistent with the analytical properties of the underlying kinetic model, especially in the diffusive limit $\varepsilon \rightarrow 0$.

4.1. One-dimensional parity symmetry, conservation and positivity.

For the AP system (2.9), the variables (ρ, j, w) correspond to the macroscopic density, the odd flux, and the higher-order even correction. The neural network approximations $(\rho_\theta, j_\theta, w_\theta)$ are constructed to respect the following relations

$$(4.1) \quad j_\theta(t, x, -v) = -j_\theta(t, x, v), w_\theta(t, x, -v) = w_\theta(t, x, v), \langle w_\theta(t, x, \cdot) \rangle = 0,$$

which are essential for preserving the equivalence between kinetic and macroscopic moments. These relations are explicitly embedded in neural architectures as follows:

- The flux network j_θ is enforced to be odd in v by the antisymmetric combination

$$j_\theta(t, x, v) = \frac{1}{2} [j_{\text{net}}(t, x, v) - j_{\text{net}}(t, x, -v)],$$

ensuring that the velocity reversal symmetry is exactly preserved at the network level.

- The correction term w_θ is constructed to be even in v and to satisfy the zero-average constraint by

$$w_\theta(t, x, v) = \frac{1}{2} [w_{\text{net}}(t, x, v) - \langle w_{\text{net}}(t, x, v) \rangle + w_{\text{net}}(t, x, -v) - \langle w_{\text{net}}(t, x, -v) \rangle],$$

which guarantees $\langle w_\theta \rangle = 0$ for all (t, x) .

- The macroscopic density ρ_θ is constrained to remain positive through the transformation $\rho_\theta = \log(1 + \exp(\rho_{\text{net}}))$. This use of the softplus activation function naturally bounds the gradient and mitigates numerical instabilities.

Here, $j_{\text{net}}, w_{\text{net}}, \rho_{\text{net}}$ denote the raw outputs of the neural submodules before enforcing these structures.

4.2. Two-dimensional parity symmetry, conservation and positivity.

For the AP system (3.11) with velocity variable (ξ, η) on the unit circle satisfying $\xi^2 + \eta^2 = 1$, the structure-preserving mechanism must account for the more complex

parity structure introduced by the decomposition with respect to the vector (ξ, η) . Recall that (3.11) involves five variables: the macroscopic density ρ , the variable φ , two odd fluxes j_1, j_2 , and the higher-order even correction w . The neural network approximations are constructed to satisfy the following symmetry and conservation properties:

$$(4.2) \quad \begin{aligned} \varphi_\theta(t, \mathbf{x}, \xi, \eta) &= \varphi_\theta(t, \mathbf{x}, -\xi, -\eta) = -\varphi_\theta(t, \mathbf{x}, -\xi, \eta) = -\varphi_\theta(t, \mathbf{x}, \xi, -\eta), \\ j_{1,\theta}(t, \mathbf{x}, \xi, \eta) &= -j_{1,\theta}(t, \mathbf{x}, -\xi, -\eta), \quad j_{2,\theta}(t, \mathbf{x}, \xi, \eta) = -j_{2,\theta}(t, \mathbf{x}, -\xi, -\eta), \\ w_\theta(t, \mathbf{x}, \xi, \eta) &= w_\theta(t, \mathbf{x}, -\xi, -\eta) = w_\theta(t, \mathbf{x}, -\xi, \eta) = w_\theta(t, \mathbf{x}, \xi, -\eta), \\ \langle w_\theta(t, \mathbf{x}, \xi, \eta) \rangle &= 0. \end{aligned}$$

These properties are enforced in the neural architecture as follows:

- The anisotropy term φ_θ is constructed to satisfy the required parity structure by combining evaluations at all four symmetric velocity points:

$$\begin{aligned} \varphi_\theta(t, \mathbf{x}, \xi, \eta) &= \frac{1}{2} \left[\varphi_{\text{net}}(t, \mathbf{x}, \xi, \eta) + \varphi_{\text{net}}(t, \mathbf{x}, -\xi, -\eta) \right. \\ &\quad \left. - \varphi_{\text{net}}(t, \mathbf{x}, -\xi, \eta) - \varphi_{\text{net}}(t, \mathbf{x}, \xi, -\eta) \right], \end{aligned}$$

ensuring $\varphi_\theta(t, \mathbf{x}, \xi, \eta) = \varphi_\theta(t, \mathbf{x}, -\xi, -\eta)$ and $\varphi_\theta(t, \mathbf{x}, \xi, \eta) = -\varphi_\theta(t, \mathbf{x}, -\xi, \eta)$.

- The fluxes $j_{1,\theta}$ and $j_{2,\theta}$ are enforced to be odd with respect to the transformation $(\xi, \eta) \rightarrow (-\xi, \eta)$ by

$$\begin{aligned} j_{1,\theta}(t, \mathbf{x}, \xi, \eta) &= \frac{1}{2} [j_{1,\text{net}}(t, \mathbf{x}, \xi, -\eta) - j_{1,\text{net}}(t, \mathbf{x}, -\xi, \eta)], \\ j_{2,\theta}(t, \mathbf{x}, \xi, \eta) &= \frac{1}{2} [j_{2,\text{net}}(t, \mathbf{x}, \xi, \eta) - j_{2,\text{net}}(t, \mathbf{x}, -\xi, -\eta)], \end{aligned}$$

which are consistent with the definitions from the previous decomposition.

- The higher-order correction w_θ is constructed to be even with respect to all velocity reflections and to satisfy the zero-average constraint. This is achieved through the symmetrized construction:

$$\begin{aligned} w_\theta(t, \mathbf{x}, \xi, \eta) &= \frac{1}{2} \left[w_{\text{net}}(t, \mathbf{x}, \xi, \eta) - \langle w_{\text{net}}(t, \mathbf{x}, \xi, \eta) \rangle \right. \\ &\quad + w_{\text{net}}(t, \mathbf{x}, -\xi, -\eta) - \langle w_{\text{net}}(t, \mathbf{x}, -\xi, -\eta) \rangle \\ &\quad + w_{\text{net}}(t, \mathbf{x}, -\xi, \eta) - \langle w_{\text{net}}(t, \mathbf{x}, -\xi, \eta) \rangle \\ &\quad \left. + w_{\text{net}}(t, \mathbf{x}, \xi, -\eta) - \langle w_{\text{net}}(t, \mathbf{x}, \xi, -\eta) \rangle \right], \end{aligned}$$

guaranteeing both the even parity with respect to all coordinate reflections and the constraint $\langle w_\theta \rangle = 0$. The zero-average property is essential for maintaining consistency with the decomposition $\frac{r_1 + r_2}{2} = \rho + \varepsilon^2 w$.

- As in the one-dimensional case, the macroscopic density ρ_θ is enforced to remain positive through the transformation $\rho_\theta = \log(1 + \exp(\rho_{\text{net}}))$, ensuring physical consistency by construction.

Here, φ_{net} , $j_{1,\text{net}}$, $j_{2,\text{net}}$, w_{net} , and ρ_{net} denote the raw outputs of the corresponding neural submodules before enforcing the structural constraints.

By integrating these constraints directly into the neural representation, our framework not only reproduces the correct asymptotic behavior but also retains the fundamental structure of the parity equations, thereby ensuring both accuracy and physical consistency across multiple scales.

5. Numerical experiments. In this section, we present a series of numerical experiments performed for both the rarefied and diffusive regimes of one- and two-dimensional RTEs, to demonstrate the effectiveness of the proposed framework and validate the theoretical analysis.

For these experiments, we employ a four-block AdaptiveResNet architecture, using 128 and 256 hidden units for the one- and two-dimensional problems, respectively. An L layer AdaptiveResNet is proposed as follows:

$$(5.1) \quad \begin{aligned} g_\theta^{[0]}(z) &= W^{[0]}z + b^{[0]}, \\ g_\theta^{[l]}(z) &= \text{AdaptiveResidualBlock}^{[l]}(g_\theta^{[l-1]}(z)), \quad 1 \leq l \leq L-1, \\ g_\theta(z) &= g_\theta^{[L]}(z) = W^{[L-1]}g_\theta^{[L-1]}(z) + b^{[L-1]}, \end{aligned}$$

where each adaptive residual block is defined as

$$(5.2) \quad \begin{aligned} \text{AdaptiveResidualBlock}^{[l]}(g_\theta^{[l-1]}(z)) &= \beta^{[l]} \cdot g_\theta^{[l-1]}(z) + (1 - \beta^{[l]}) \cdot h_\theta^{[l]}(z), \\ h_\theta^{[l]}(z) &= \sigma \circ \left(W_2^{[l-1]} \sigma \circ (W_1^{[l-1]} g_\theta^{[l-1]}(z) + b_1^{[l-1]}) + b_2^{[l-1]} \right) \end{aligned}$$

with $\beta^{[l]} \in [0.05, 1.0]$ being a learnable adaptive parameter for the l -th residual block. Each $\beta^{[l]}$ is initialized to 0.9 and optimized during training. To ensure numerical stability, the value of each $\beta^{[l]}$ is constrained to the interval $[0.05, 1.0]$ via a clamping operation. Here, $W_1^{[l]}, W_2^{[l]} \in \mathbb{R}^{m_{l+1} \times m_l}$ and $b_1^{[l]}, b_2^{[l]} \in \mathbb{R}^{m_{l+1}}$ are weight matrices and bias vectors, respectively. The input dimension is $m_0 = d_{\text{in}} = d$, and the output dimension is $m_L = d_0$. The activation function is chosen as $\sigma = \tanh(\cdot)$, which is applied in the element-wise sense, as indicated by “ $\circ$ ”. The collection of all network parameters, including $\{\beta^{[l]}\}_{l=1}^{L-1}$, is represented by θ .

We fix the spatial domain to $[0, 1]^d$, and an exact periodic boundary condition is enforced to improve the numerical performance. Taking the one dimensional case as an example, the ansatz is constructed based on a Fourier basis, where a transform $T : x \rightarrow \tilde{x} = \{\sin(2\pi kx), \cos(2\pi kx)\}_{k=1}^P$ is applied before the first layer of the deep neural network [15]. The extension to higher-dimensional cases follows naturally. The averaging operator is evaluated numerically using a 16-point Gauss–Legendre quadrature rule.

For network training, we employ the Adam optimizer with Xavier initialization. In each iteration, 4096 random sample points are drawn from the spatial domain and 1024 from the initial condition. Except for the initial condition term, which is assigned a weight of 10, all other loss terms are assigned a default weight of 1. To enhance training stability and convergence, we adopt an exponentially decaying learning rate schedule starting from $\eta_0 = 10^{-3}$ with a decay rate of $\gamma = 0.96$ and a decay step of $p = 2500$ iterations, $\eta_{it} = \eta_0 \cdot \gamma^{\lfloor \frac{it}{p} \rfloor}$, where it denotes the current iteration index, and the symbol $\lfloor \cdot \rfloor$ is the floor function.

The reference solutions are computed using a finite difference method [21], and we will check the relative ℓ^2 error of the solution $s(x)$ of our method, e.g., for the one-dimensional case,

$$(5.3) \quad \text{error} := \sqrt{\sum_j |s_\theta^j - s_{\text{ref}}^j|^2 / \sum_j |s_{\text{ref}}^j|^2},$$

where s_θ is the neural solution, and s_{ref} is the reference solution. Here, the superscript j indexes the collocation points at which both s_θ and s_{ref} are evaluated.

5.1. One-dimensional example. We first consider the one-dimensional RTE, testing the proposed framework across regimes ranging from the rarefied case ($\varepsilon = 1$) to the diffusive limit ($\varepsilon = 10^{-4}$). The computational domain is $x \in [0, 1]$ with periodic boundary condition, and the initial condition is $f_{IC}(x, v) = p(x) \exp(-v^2/2)/\sqrt{2\pi}$, where $p(x) = 1 + \cos(4\pi x)$.

Fig. 3 and Fig. 4 show the training history and the evolution of the macroscopic density $\rho(t, x)$ for $\varepsilon = 1$ at $t = 0.05$ and $t = 0.1$. The loss and relative ℓ^2 error decrease steadily as training proceeds, and the predicted densities closely match the reference finite difference solutions. This demonstrates that the proposed method effectively captures the kinetic behavior in the rarefied regime. Similarly, the results for the diffusive regime ($\varepsilon = 10^{-4}$) are shown in Fig. 5 and Fig. 6. As the regime approaches diffusion, the network continues to produce accurate solutions, maintaining numerical stability and precision even at smaller mean free path scales. The comparison between the neural and reference solutions again exhibits excellent agreement, verifying the robustness of the proposed method across multiscale regimes.

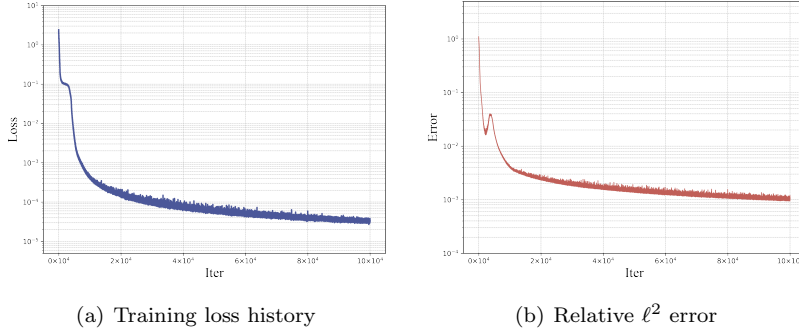


Fig. 3: Training loss and error evolution for the one-dimensional case ($\varepsilon = 1$).

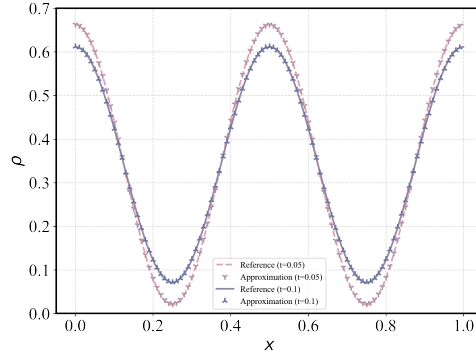


Fig. 4: Comparison between reference and neural network solutions at $t = 0.05$ and $t = 0.1$ ($\varepsilon = 1$).

5.2. Two-dimensional example. We next extend the analysis to the two-dimensional RTE under periodic boundary conditions. The initial distribution is

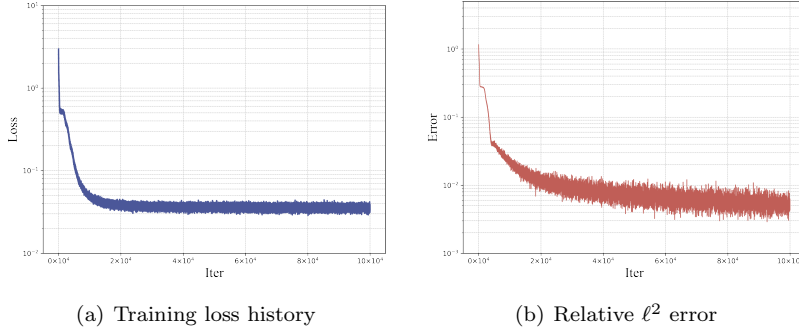


Fig. 5: Training loss and error evolution for the one-dimensional case ($\varepsilon = 10^{-4}$).

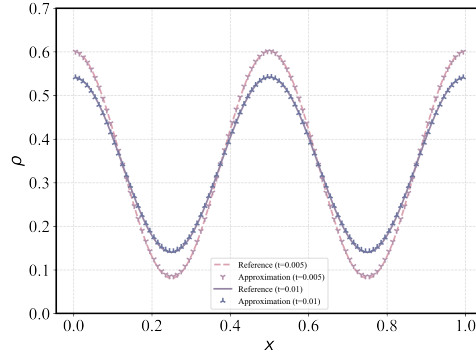


Fig. 6: Comparison between reference and neural network solutions at $t = 0.005$ and $t = 0.01$ ($\varepsilon = 10^{-4}$).

$f_{1C}(\mathbf{x}, \mathbf{v}) = p(\mathbf{x}) \exp(-|\mathbf{v}|^2/2)/\sqrt{2\pi}$, where $p(\mathbf{x}) = 1 + (\cos(2\pi x_1) + \cos(2\pi x_2))/2$.

Fig. 7 and Fig. 8 display the predicted and reference density fields $\rho(t, \mathbf{x})$ at time $t = 0.1$ for the rarefied regime ($\varepsilon = 1$) and the near-diffusive regime ($\varepsilon = 10^{-3}$), respectively. In both cases, our method accurately captures the spatial structure and amplitude of the density field, closely matching the reference solutions. These results confirm that the proposed adaptive framework generalizes effectively from one to two dimensions and maintains consistent performance across kinetic–diffusive transitions.

6. Conclusions. In this paper, we have proposed a multiscale parity decomposition framework for solving radiative transfer equations. Both the theoretical analysis and numerical experiments demonstrate that the proposed framework achieves uniform error control while preserving the intrinsic structure of the system, thereby enabling stable and accurate computation for multiscale radiative transfer problems. The main contributions of this work lie in the development of a structure-preserving and asymptotic-preserving neural network framework and the proof of uniform stability of the total error with respect to the Knudsen number, offering a viable approach for high-fidelity simulations of multiscale RTEs. The present study is limited to periodic boundary conditions; the extension to inflow and reflective boundary conditions

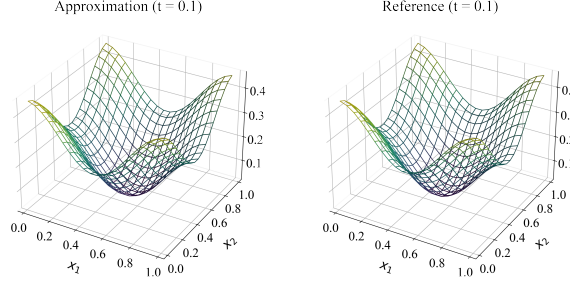


Fig. 7: Reference and approximate density fields at $t = 0.1$ ($\varepsilon = 1$).

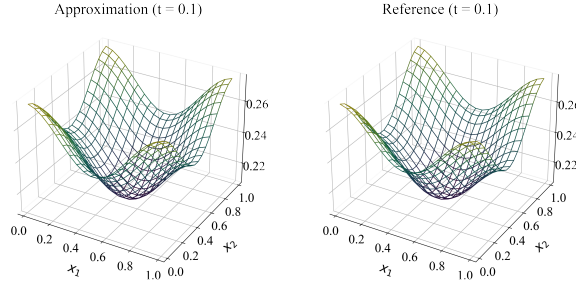


Fig. 8: Reference and approximate density fields at $t = 0.1$ ($\varepsilon = 10^{-3}$).

will be addressed in future work. Moreover, the present work has focused on the one- and two-dimensional cases following [21]; extending the approach to three dimensions poses additional challenges due to increased geometric complexity, although the underlying theoretical framework is expected to remain consistent with the lower-dimensional settings. Furthermore, applying the proposed method to nonlinear kinetic systems remains an important direction for continued research.

REFERENCES

- [1] E. ABDO, L. CHAI, R. HU, AND X. YANG, *Error estimates of physics-informed neural networks for approximating Boltzmann equation*, arXiv preprint arXiv:2407.08383, (2025).
- [2] M. L. ADAMS AND E. W. LARSEN, *Fast iterative methods for discrete-ordinates particle transport calculations*, Progress in Nuclear Energy, 40 (2002), pp. 3–159.
- [3] F. J. ALEXANDER AND A. L. GARCIA, *The direct simulation Monte Carlo method*, Computers in Physics, 11 (1997), 588 (6 pages).
- [4] S. R. ARRIDGE AND J. C. SCHOTLAND, *Optical tomography: forward and inverse problems*, Inverse Problems, 25 (2009), 123010 (69 pages).
- [5] R. BASKAR, K. A. LEE, R. YEO, AND K.-W. YEOH, *Cancer and radiation therapy: current advances and future directions*, International Journal of Medical Sciences, 9 (2012), 193 (7 pages).
- [6] G. BERTAGLIA, C. LU, L. PARESCHI, AND X. ZHU, *Asymptotic-preserving neural networks for*

- multiscale hyperbolic models of epidemic spread*, Mathematical Models and Methods in Applied Sciences, 32 (2022), pp. 1949–1985.
- [7] J. CHEN, Z. MA, AND K. WU, *A micro-macro decomposition-based asymptotic-preserving random feature method for multiscale radiative transfer equations*, Journal of Computational Physics, 537 (2025), 114103 (18 pages).
 - [8] W. CHEN, G. DIMARCO, AND L. PARESCHI, *Structure and asymptotic preserving deep neural surrogates for uncertainty quantification in multiscale kinetic equations*, arXiv preprint arXiv:2506.10636, (2025).
 - [9] Z. CHEN, L. LIU, AND L. MU, *Solving the linear transport equation by a deep neural network approach*, Discrete and Continuous Dynamical Systems-S, 15 (2022), pp. 669–686.
 - [10] B. DAVISON, J. B. SYKES, AND E. R. COHEN, *Neutron Transport Theory*, American Institute of Physics, 1958.
 - [11] K. F. EVANS, *The spherical harmonics discrete ordinate method for three-dimensional atmospheric radiative transfer*, Journal of the Atmospheric Sciences, 55 (1998), pp. 429–446.
 - [12] M. FRANK, A. KLAR, E. LARSEN, AND S. YASUDA, *Approximate models for radiative transfer*, Bulletin-Institute of Mathematics Academia Sinica, 2 (2007), 409 (24 pages).
 - [13] F. GOLSE, S. JIN, AND C. D. LEVERMORE, *A domain decomposition analysis for a two-scale linear transport problem*, ESAIM: Mathematical Modelling and Numerical Analysis, 37 (2003), pp. 869–892.
 - [14] H. HAN, M. TANG, AND W. YING, *Two uniform tailored finite point schemes for the two dimensional discrete ordinates transport equations with boundary and interface layers*, Communications in Computational Physics, 15 (2014), pp. 797–826.
 - [15] J. HAN, J. LU, AND M. ZHOU, *Solving high-dimensional eigenvalue problems using deep neural networks: A diffusion Monte Carlo like approach*, Journal of Computational Physics, 423 (2020), 109792 (13 pages).
 - [16] R. HU, Q. LIN, A. RAYDAN, AND S. TANG, *Higher-order error estimates for physics-informed neural networks approximating the primitive equations*, Partial Differential Equations and Applications, 4 (2023), 34 (35 pages).
 - [17] S. JIN, *Asymptotic preserving (AP) schemes for multiscale kinetic and hyperbolic equations: a review*, Lecture notes for summer school on methods and models of kinetic theory (M&MKT), Porto Ercole (Grosseto, Italy), (2010), pp. 177–216.
 - [18] S. JIN, Z. MA, AND K. WU, *Asymptotic-preserving neural networks for multiscale time-dependent linear transport equations*, Journal of Scientific Computing, 94 (2023), 57 (21 pages).
 - [19] S. JIN, Z. MA, AND K. WU, *Asymptotic-preserving neural networks for multiscale kinetic equations*, Communications in Computational Physics, 35 (2024), pp. 693–723.
 - [20] S. JIN, Z. MA, AND T.-A. ZHANG, *Asymptotic-preserving neural networks for multiscale Vlasov–Poisson–Fokker–Planck system in the high-field regime*, Journal of Scientific Computing, 99 (2024), 61 (21 pages).
 - [21] S. JIN, L. PARESCHI, AND G. TOSCANI, *Uniformly accurate diffusive relaxation schemes for multiscale transport equations*, SIAM Journal on Numerical Analysis, 38 (2000), pp. 913–936.
 - [22] A. D. KLOSE AND A. H. HIELSCHER, *Optical tomography using the time-independent equation of radiative transfer—Part 2: inverse model*, Journal of Quantitative Spectroscopy and Radiative Transfer, 72 (2002), pp. 715–732.
 - [23] A. D. KLOSE, U. NETZ, J. BEUTHAN, AND A. H. HIELSCHER, *Optical tomography using the time-independent equation of radiative transfer—Part 1: forward model*, Journal of Quantitative Spectroscopy and Radiative Transfer, 72 (2002), pp. 691–713.
 - [24] E. W. LARSEN, J. E. MOREL, AND W. F. MILLER JR, *Asymptotic solutions of numerical transport problems in optically thick, diffusive regimes*, Journal of Computational Physics, 69 (1987), pp. 283–324.
 - [25] E. LEWIS AND W. F. MILLER, *Computational Methods of Neutron Transport*, John Wiley and Sons, Inc., New York, NY, 1984.
 - [26] H. LI, S. JIANG, W. SUN, L. XU, AND G. ZHOU, *A model-data asymptotic-preserving neural network method based on micro-macro decomposition for gray radiative transfer equations*, Communications in Computational Physics, 35 (2024), pp. 1155–1193.
 - [27] H. LI, S. JIANG, W. SUN, L. XU, AND G. ZHOU, *Macroscopic auxiliary asymptotic preserving neural networks for the linear radiative transfer equations*, Journal of Scientific Computing, 103 (2025), 84 (32 pages).
 - [28] Q. LI AND L. WANG, *Implicit asymptotic preserving method for linear transport equations*, Communications in Computational Physics, 22 (2017), pp. 157–181.
 - [29] C. LIU, W. LI, Y. WANG, P. SONG, AND K. XU, *An implicit unified gas-kinetic wave-particle*

- method for radiative transport process*, Physics of Fluids, 35 (2023).
- [30] L. LIU, Y. WANG, X. ZHU, AND Z. ZHU, *Asymptotic-preserving neural networks for the semiconductor Boltzmann equation and its application on inverse problems*, Journal of Computational Physics, 523 (2025), 113669 (32 pages).
 - [31] Y. LU, L. WANG, AND W. XU, *Solving multiscale steady radiative transfer equation using neural networks with uniform stability*, Research in the Mathematical Sciences, 9 (2022), pp. 1–29.
 - [32] R. MARSHAK, *Note on the spherical harmonic method as applied to the Milne problem for a sphere*, Physical Review, 71 (1947), 443 (4 pages).
 - [33] D. MIHALAS AND B. W. MIHALAS, *Foundations of Radiation Hydrodynamics*, Courier Corporation, 2013.
 - [34] S. MISHRA AND R. MOLINARO, *Physics informed neural networks for simulating radiative transfer*, Journal of Quantitative Spectroscopy and Radiative Transfer, 270 (2021), 107705 (17 pages).
 - [35] M. F. MODEST AND S. MAZUMDER, *Radiative Heat Transfer*, Academic Press, 2021.
 - [36] M. RAISSI, P. PERDIKARIS, AND G. E. KARNIADAKIS, *Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations*, Journal of Computational Physics, 378 (2019), pp. 686–707.
 - [37] K. REN, G. BAL, AND A. H. HIELSCHER, *Frequency domain optical tomography based on the equation of radiative transfer*, SIAM Journal on Scientific Computing, 28 (2006), pp. 1463–1489.
 - [38] Y. SHI, P. SONG, AND T. XIONG, *An efficient asymptotic preserving Monte Carlo method for radiative transfer equations*, Journal of Computational Physics, 493 (2023), 112483 (19 pages).
 - [39] P. SUETENS, *Fundamentals of Medical Imaging*, Cambridge University Press, 2017.
 - [40] K. WU, X.-B. YAN, S. JIN, AND Z. MA, *Capturing the diffusive behavior of the multiscale linear transport equations by Asymptotic-Preserving Convolutional DeepONets*, Computer Methods in Applied Mechanics and Engineering, 418 (2024), 116531 (18 pages).
 - [41] Y. WU, J. HUANG, X. QIAN, AND W. SUN, *Physics informed neural networks with variable Eddington factor iteration for linear radiative transfer equations*, Computer Physics Communications, 318 (2026), 109879 (18 pages).
 - [42] T. XIONG, W. SUN, Y. SHI, AND P. SONG, *High order asymptotic preserving discontinuous Galerkin methods for gray radiative transfer equations*, Journal of Computational Physics, 463 (2022), 111308 (29 pages).
 - [43] T.-A. ZHANG AND S. JIN, *AP-MIONet: Asymptotic-Preserving multiple-input neural operators for capturing the high-field limits of collisional kinetic equations*, Journal of Scientific Computing, 104 (2025), 17 (36 pages).