Subordination-based neural network approach for non-Markovian Fokker-Planck equations

Gege Wang [1,2,3] Wei Wang [2] Xiaolong Wang [3,4] Ralf Metzler [2,5] and Yong Xu [1,3,*]

[1] *School of Mathematics and Statistics, Northwestern Polytechnical University, Xi'an 710072, China*

[2] *Institute of Physics and Astronomy, University of Potsdam, 14476 Potsdam-Golm, Germany*

[3] *MOE Key Laboratory for Complexity Science in Aerospace, Northwestern Polytechnical University, Xi'an 710072, China*

[4] *School of Mathematics and Statistics, Shaanxi Normal University, Xi'an 710119, China*

[5] *Asia Pacific Centre for Theoretical Physics, Pohang 37673, Republic of Korea*

(Received 6 February 2026; revised 19 May 2026; accepted 25 June 2026; published 31 July 2026)

The non-Markovian Fokker-Planck equation (FPE) provides an effective description of anomalous diffusion within the classical continuous time random walk framework. However, its nonlocal memory kernel leads to strong non-Markovian behaviors, making the non-Markovian FPE difficult to solve. Conventional numerical approaches must rely on intricate discretized implementations of fractional-calculus differentiation to approximate the fractional operators, leading to substantial computational and storage costs. To overcome these challenges and obtain the probability density function (PDF) across the entire temporal regime, we propose a subordination-based neural network (SNN) that leverages an equivalent subordinated integral representation of non-Markovian dynamics. This representation transforms the original non-Markovian FPE into a Markovian FPE formulated in an operational time where all memory effects are encapsulated by the subordination mechanism, enabling the solution via a deep neural network and direct transfer to FPEs with different memory kernels that share the same auxiliary Markovian FPE, without retraining. We validate the method for non-Markovian FPEs under nonlinear external forces, general boundary conditions, and with arbitrary memory kernels. With four representative examples, the SNN maintains high accuracy in both the central and tail regions of the PDFs, as well as in additional statistical quantities, e.g., survival probability and mean velocity, that are relevant to practical applications. The SNN approach provides a promising route for future extensions to higher-dimensional systems.

DOI: 10.1103/k8yg-mkml

I. INTRODUCTION

Einstein's 1905 formulation of the diffusion equation, a drift-free case of the ordinary Fokker-Planck equation (FPE) [1,2], demonstrated that Brownian motion (BM) exhibits classical diffusion with a mean-squared displacement (MSD) growing linearly in time [3]. Since then, theoretical frameworks of BM and the FPE have become foundational tools in the study of stochastic dynamics [4–8]. With advances in microscopic techniques, anomalous diffusion has been widely found in a broad range of biological, soft-matter, ecological, and quantum systems [9–19]. It is typically characterized by the power-law scaling of the MSD [20,21] $\langle x^2 \rangle = 2K_\alpha t^\alpha$, where K_α is the generalized diffusion coefficient and α is the anomalous diffusion exponent. The anomalous diffusion exponent classifies the type of transport: subdiffusion occurs for $0 < \alpha < 1$, superdiffusion for $\alpha > 1$, while $\alpha = 1$ and $\alpha = 2$ correspond to normal and ballistic diffusion, respectively.

Fractional derivatives provide a powerful tool to characterize anomalous diffusion. They were first incorporated into the diffusion equation by Schneider and Wyss [22], who introduced a fractional diffusion equation based on the Riemann-Liouville (RL) fractional differential operator with a power-law memory kernel. The corresponding equation admits solutions in terms of Fox H-functions [23]. It can be interpreted as the continuum limit of continuous-time random walks (CTRWs) featuring a scale-free waiting time probability density function (PDF) $\varphi(t) \simeq t^{-1-\alpha}$ with anomalous exponent $0 < \alpha < 1$ and jump length densities with the finite variance [24–28]. These features give rise to subdiffusive dynamics. CTRWs with power-law waiting time PDF have been revealed in living cells [17,29], and charge carrier transport in disordered and amorphous semiconductors [30,31]. In movement ecology, weak ergodicity breaking has been observed in the flight trajectories of kites and other foraging animals and linked to power-law-distributed resting times [19]. Similar long-tailed trapping times are observed in DNA conformational dynamics [32] and in chemical tracer transport in heterogeneous subsurface systems [33–36], where CTRW models naturally capture the resulting non-Fickian dispersion [20]. When an external potential is present, this fractional diffusion equation naturally extends to the fractional FPE [37–40]. The time-fractional FPE generalizes the ordinary FPE framework by introducing fractional time operators that encode memory effects and nonlocal transport, leading to intrinsically non-Markovian dynamics. In addition to the widely used RL and Caputo derivatives,[1] several

*Contact author: hsux3@nwpu.edu.cn

[1] The fractional FPEs with RL and Caputo derivatives are equivalent under suitable initial conditions.

 ©2026 American Physical Society

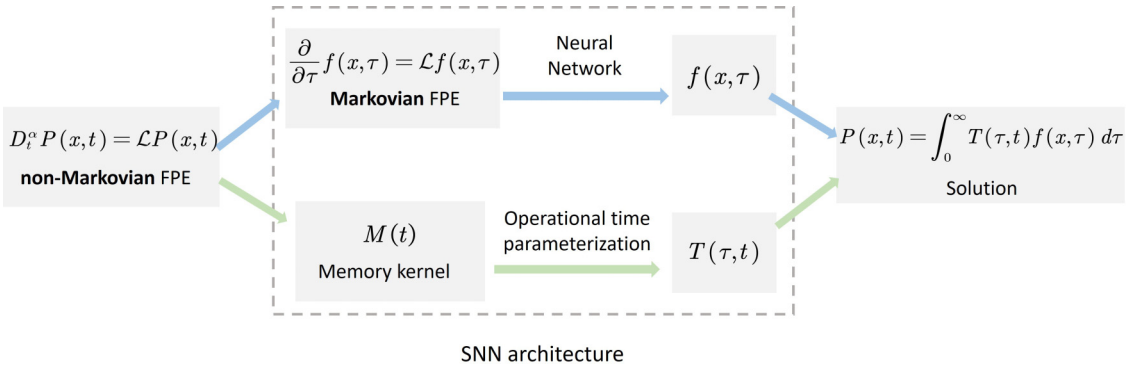


FIG. 1. Illustration of the workflow for solving non-Markovian FPEs using a subordination-based neural network (SNN). The approach addresses non-Markovian FPEs by decoupling the problem into two aspects: solving the underlying Markovian FPE via a neural network, and deriving the PDF of the operational time $\tau(t)$. The non-Markovian FPE solution is then reconstructed through the subordinated integral representation.

important generalizations employ more complex memory kernels, including the Mittag-Leffler functions underlying the Atangana-Baleanu (AB) derivative [41,42] and the generalized Mittag-Leffler forms associated with the Prabhakar derivative [43–45]. Evidence indicates that the formulation in terms of the AB derivative leads to a nonzero value of the MSD at short times, touching on some open questions concerning the physical application of this operator [42]. In this work, we focus on time-fractional FPEs with general memory kernels, i.e., on a broad class of non-Markovian FPEs. Recently, memory-dependent FPEs have also been developed for systems driven by fractional Gaussian noise and Gaussian white noise, characterizing the probability structure associated with fractional Brownian motion [46,47].

To compute the solution, i.e., positional PDF $P(x,t)$ of non-Markovian FPEs, a wide variety of numerical schemes have been proposed. Most existing approaches discretize fractional operators, whose intrinsic temporal nonlocality leads to substantial computational and memory demands when solved iteratively. Representative techniques include finite-difference [48–50] and spectral methods [51–56]. Finite-difference schemes often suffer from high computational cost due to the dense temporal coupling generated by the nonlocal memory kernel. Spectral methods, while efficient for smooth solutions and simple domains, become difficult to apply for FPEs with reflecting or absorbing boundaries or with non-power-law memory kernels, since boundary conditions and nonlocal operators destroy the structure that spectral methods rely on for efficiency.

Recently, machine learning approaches have been successfully employed for the analysis of anomalous diffusion [57–63], solving differential equations [64–67] and inferring hidden parameters in noisy data [68,69]. A widely used method for solving differential equations is the physics-informed neural network (PINN), a mesh-free deep learning framework that enforces residuals of the governing equations together with initial and boundary conditions through a loss function constructed via automatic differentiation [65]. PINNs avoid the memory-intensive grids required by traditional numerical schemes. Xu *et al.* further improved the original PINN formulation by introducing a normalization on the FPE's solution, thereby preventing convergence to the trivial zero solution [66,67]. To address the normalization issue from a different perspective, another research direction approximates the solution of FPEs using Gaussian-mixture representations [70–72]. Deep learning techniques have also been extended to non-Markovian FPEs [73–76]; however, these methods require approximating the fractional operator using intricate discretization schemes prior to applying neural networks. Taken together, these advances demonstrate that deep neural networks are highly effective for ordinary Markovian FPEs. Yet they remain limited when extended to fractional or more general non-Markovian FPEs, where nonlocal memory kernels introduce dense temporal coupling that standard neural frameworks cannot efficiently accommodate, highlighting the need for alternative formulations.

To address this difficulty, we introduce a subordination-based neural network (SNN), as illustrated in Fig. 1, designed for a broad class of non-Markovian FPEs. The approach is based on a subordinated process [37–39,77–82] $X[\tau(t)]$, which is built from a parent process $X(\tau)$ and operational time process $\tau(t)$, reproducing the dynamics of the target non-Markovian FPE. The parent process $X(\tau)$ obeys a memoryless (Markovian) FPE, which can be effectively approximated using a deep neural network. By combining the solution of the parent process with the statistical information encoded in the operational time process, the SNN reconstructs the PDF of the non-Markovian FPE. In our previous works [66,67], deep neural networks were successfully employed to solve Markovian FPEs and to obtain statistical solutions of stochastic differential equations. Building on these studies, we incorporate the established subordination formalism [37–39] into the neural-network framework and develop an SNN method for non-Markovian FPEs with memory effects. In contrast to existing neural-network-based methods for fractional PDEs, which typically include fractional derivatives directly in the residual loss and approximate them by numerical discretization [73–76], the SNN exploits the subordinated Markovian representation of non-Markovian dynamics: the neural network learns the auxiliary Markovian dynamics, while the memory effect is introduced through the subordination kernel. This structural decomposition makes the SNN transferable across different memory kernels or kernel parameters sharing the same auxiliary Markovian process, and provides a

framework for non-Markovian FPEs with complex potentials, boundary conditions, and memory kernels.

The paper is structured as follows. In Sec. II, we briefly introduce the general form of non-Markovian FPEs and review several typical fractional operators we used in this paper. Section III establishes the conditions under which a subordination representation exists for a given non-Markovian FPE, and then a detailed description of the proposed neural network architecture is provided. Section IV validates the SNN method through numerical experiments and demonstrates the accuracy and robustness. Finally, Sec. V summarizes the main findings and outlines potential directions for future research.

II. NON-MARKOVIAN FOKKER-PLANCK EQUATION

We consider the anomalous transport of particles in complex media confined to a finite domain $[c, d]$ under the influence of an external force, as governed by the non-Markovian FPE for the position PDF $P(x, t)$,

$$\int_0^t M(t-t')\frac{\partial}{\partial t'}P(x,t')dt' = \mathcal{L}P(x,t), x \in [c,d], \tag{1a}$$

$$P(x,0) = \rho_0(x), \tag{1b}$$

$$\mathcal{B}_{1,x}[P]|_{x=c} = 0, \quad \mathcal{B}_{2,x}[P]\big|_{x=d} = 0, \tag{1c}$$

where $M(t-t')$ is the memory kernel with physical dimension time$^{-\alpha}$, and $\rho_0(x)$ is a prescribed and normalized function specifying the initial condition given by Eq. (1b). The boundary operators $\mathcal{B}_{1,x}$ and $\mathcal{B}_{2,x}$ in Eq. (1c) are linear operators acting only on the spatial variable x, defined as

$$\mathcal{B}_{i,x} = \beta_i + \zeta_i \frac{\partial}{\partial x}, \quad i = 1, 2, \tag{2}$$

where the constants β_i and ζ_i specify the type of boundary condition: Dirichlet/absorbing ($\beta_i = 1, \zeta_i = 0$), Neumann/reflecting ($\beta_i = 0, \zeta_i = 1$), and Robin/reactive ($\beta_i < 0, \zeta_i = 1$). For transport in one-dimensional unbounded space, the natural boundary condition requires the PDF to vanish at infinity, i.e., $P(x,t) \to 0$ as $x \to \pm\infty$.

The Fokker-Planck operator $\mathcal{L}$ is defined as

$$\mathcal{L} = -\frac{\partial}{\partial x}\frac{F(x)}{m\gamma_\alpha} + K_\alpha \frac{\partial^2}{\partial x^2}, \tag{3}$$

where m is the particle mass, γ_α is a friction coefficient of dimension time$^{\alpha-2}$, $F(x)$ denotes the external force field, and the coefficient K_α is a constant with physical dimension length^2/time$^\alpha$. In this paper, two specific memory kernels are considered in Eq. (1a), corresponding to the Caputo and AB fractional derivatives, respectively.

The Caputo fractional derivative of the order $\alpha \in (0, 1]$ is defined as [83]

$${}^{C}D_t^\alpha P(x,t) \equiv \int_0^t M_C(t-t')\frac{\partial P(x,t')}{\partial t'}dt', \tag{4}$$

with the power-law memory kernel

$$M_C(t-t') = \frac{1}{\Gamma(1-\alpha)}(t-t')^{-\alpha}. \tag{5}$$

In definition (4), the sequence of temporal integration and differentiation is reversed with respect to the RL derivative [84].

In addition to the power-law kernel, we consider the AB fractional derivative with order $\alpha \in (0, 1]$, defined as

$${}^{AB}D_t^\alpha P(x,t) \equiv \int_0^t M_{AB}(t-t')\frac{\partial P(x,t')}{\partial t'}dt', \tag{6}$$

with the Mittag-Leffler memory kernel

$$M_{AB}(t-t') = \frac{1}{(1-\alpha)\tau^\alpha}E_\alpha\left(-\alpha\frac{(t-t')^\alpha}{(1-\alpha)\tau^\alpha}\right), \tag{7}$$

where $E_\alpha(-z) = \sum_{k=0}^{\infty}(-z)^k/\Gamma(1+\alpha k)$ is the (single-parameter) Mittag-Leffler function [85]. Note that we introduced the timescale τ to get dimensions correct. In contrast to the Caputo derivative, the AB derivative employs a nonsingular memory kernel.

III. SUBORDINATION-BASED NEURAL NETWORK

Non-Markovian FPEs contain memory kernels that introduce long-range temporal correlations, making the evolution at any time depend on the entire history of the system. This temporal nonlocality leads to substantial computational and storage costs for traditional numerical methods and often produces sharp features, e.g., distinct cusps of the PDF around the initial position [23], which can severely impact the performance and stability of neural network solvers. To overcome these challenges, we propose an SNN framework that combines an equivalent subordinated integral representation of the non-Markovian dynamics via transforming the original non-Markovian FPE into a Markovian FPE formulated in an operational time such that all memory effects are encapsulated by the subordination mechanism. This reformulation avoids directly discretizing the temporal convolution terms and allows the neural network to approximate the solution of the auxiliary Markovian FPE. As a result, the method provides a practical and flexible route for solving non-Markovian FPEs, particularly when sharing the same underlying Markovian dynamics. In the following, we first present the subordinated integral representation underlying the SNN method, followed by the neural network architecture and training algorithm.

A. Subordination scheme

Here, we aim to show that the PDF of the non-Markovian FPE (1) can be expressed by the PDF of a subordinated process $X[\tau(t)]$, where $\tau(t)$ is the operational time process parametrized by the physical time t. The process $X(\tau)$ is referred to as the parent process, while $\tau(t)$ serves as the subordinator [37–39]. The PDF of $X[\tau(t)]$ is completely determined by the two PDFs $f(x,\tau)$ and $T(\tau,t)$ of the parent process $X(\tau)$ and the subordinator $\tau(t)$, respectively, through

$$P(x,t) = \int_0^\infty T(\tau,t)f(x,\tau)\,d\tau. \tag{8}$$

Two remarks on the representation in Eq. (8) are worth emphasizing:

Remark 1. A valid subordination representation in Eq. (8) is not always in one-to-one correspondence with the non-Markovian FPE in Eq. (1). Such a representation is valid only when $T(\tau,t)$ is a well-defined PDF. Equivalently, its Laplace transform in Eq. (11) must be a completely monotone function

(CMF) [86]. In Appendix B, we show that this condition is typically satisfied by memory kernels $M(t)$ in the non-Markovian FPE (1) that are positive, monotonically decaying, and have alternating derivative signs, such as exponential and power-law kernels; see the discussion of Bernstein functions below. By contrast, when the memory kernel $M(t)$ is sign-changing or oscillatory, the completely monotone condition may fail. As a result, the corresponding subordination kernel $T(\tau,t)$ may lose its non-negativity, and the representation in Eq. (8) no longer guarantees a well-defined PDF. We also provide a concrete example in Appendix B, based on the telegrapher's equation with finite propagation speed, for which the complete monotonicity of Eq. (11) is absent.

Remark 2. In general, $T(\tau,t)$ is not available in closed form and may only be accessible in the Laplace domain. In this case, the solution of the non-Markovian FPE can be obtained in the Laplace domain as

$$\tilde{P}(x,s)=\int_0^\infty \tilde{T}(\tau,s)f(x,\tau)d\tau, \tag{9}$$

followed by a numerical inverse Laplace transform to recover $P(x,t)$.

In Ref. [39], the conditions guaranteeing the equivalence between the PDFs of the non-Markovian FPE (1) in free space (with natural boundary conditions) and those of the corresponding subordinated process (8) were established. Here, we further extend this equivalence to more general settings, including arbitrary boundary conditions and higher-dimensional domains (the proof is similar to that in Ref. [39] and is omitted for brevity). Specifically, the PDF $f(x,\tau)$ of the parent process corresponds to the solution of the Markovian FPE,

$$\frac{\partial}{\partial\tau}f(x,\tau)=\mathcal{L}f(x,\tau),\quad x\in[c,d],\ \tau\in[0,\infty), \tag{10a}$$

$$f(x,0)=\rho_0(x), \tag{10b}$$

$$\mathcal{B}_{1,x}[f]\big|_{x=c}=0,\quad \mathcal{B}_{2,x}[f]\big|_{x=d}=0, \tag{10c}$$

where the operators $\mathcal{B}_{1,x}$, $\mathcal{B}_{2,x}$, $\mathcal{L}$ act solely on the spatial variable x given by Eqs. (2) and (3), which are identical to those appearing in the non-Markovian FPE (1). Importantly, the non-Markovian FPE (1) and the auxiliary Markovian FPE (10) share the same initial and boundary conditions. Consequently, non-Markovian FPEs that differ only in the choice of the fractional derivative (i.e., the memory kernel) correspond to the same auxiliary Markovian FPE. For the power-law kernel (5), the relation between the solution $P(x,t)$ and the Markovian parent process was established in Laplace space for complex force field and boundary conditions [87], and the corresponding subordination formula was presented in [23,88].

Meanwhile, the PDF $T(\tau,t)$ of the subordinator $\tau(t)$ is related to the memory kernel $M(t)$ of the non-Markovian FPE (1) through its Laplace-domain form (also see the detailed derivation in [38,39]),

$$\tilde{T}(\tau,s)=\tilde{M}(s)\exp(-\tau s\tilde{M}(s)), \tag{11}$$

where $\tilde{T}(\tau,s)=\int_0^\infty e^{-st}T(\tau,t)\,dt$ and $\tilde{M}(s)=\int_0^\infty e^{-st}M(t)\,dt$. Since $T(\tau,t)$ is a normalized PDF, according to Bernstein's theorem [78], its Laplace transform $\tilde{T}(\tau,s)$ must be a CMF function of s, a condition that must be verified beforehand for any given memory kernel. This requirement implies that not every non-Markovian FPE admits a corresponding subordination scheme. As a special case, when $M(t-t')=\delta(t-t')$, one can immediately see that $T(\tau,t)=\delta(\tau-t)$ from Eq. (11), and the system possesses no memory, and the non-Markovian FPE (1) reduces to the ordinary Markovian FPE.

For the power-law memory kernel (5) and the Mittag-Leffler memory kernel (6), their corresponding $T(\tau,t)$ in Laplace-domain representations are given by

$$\tilde{T}_C(\tau,s)=s^{\alpha-1}\exp(-s^\alpha\tau), \tag{12a}$$

$$\tilde{T}_{AB}(\tau,s)=\frac{s^{\alpha-1}}{(1-\alpha)s^\alpha+\alpha}\exp\left(-\frac{s^\alpha}{(1-\alpha)s^\alpha+\alpha}\tau\right), \tag{12b}$$

indicating that both are CMFs (see Appendix A for a detailed justification). Moreover, for the power-law memory kernel in Eq. (5), applying the inverse Laplace transform to Eq. (12a) yields [37]

$$T_C(\tau,t)=\frac{1}{\alpha}\frac{t}{\tau^{1+1/\alpha}}l_\alpha\left(\frac{t}{\tau^{1/\alpha}}\right), \tag{13}$$

where $l_\alpha(z)$ denotes a one-sided Lévy stable PDF [89,90]. This result and the corresponding physical interpretation were extensively discussed in the seminal work [23,37,87,88].

From the above discussion, solving the non-Markovian FPE (1) reduces to determining the two PDFs $T(\tau,t)$ and $f(x,\tau)$. The former follows directly from the prescribed memory kernel $M(t)$, while the latter can be efficiently computed using a deep neural network, especially in the presence of nonlinear forces, complex boundary conditions, or higher dimensions.

B. Neural network architecture

Here, we focus on the one-dimensional case, while extensions to higher dimensions are discussed in detail in Sec. V. The training dataset is constructed as follows. For the 1D case, the spatial domain is defined over $\mathcal{S}=[c,d]$ and has n_s points with spacing Δx. However, special care must be taken in the temporal discretization. Since the initial condition is usually given as $\delta(x-x_0)$, where x_0 denotes the initial position, we approximate it by a short-time Gaussian distribution at a small time τ_ε, denoted as $f_0(x,\tau_\varepsilon)$. A similar strategy has been adopted in Refs. [91,92]. At very short times, the solution exhibits rapid variations. If the temporal resolution in this regime is too coarse, the numerical accuracy may deteriorate significantly. Conversely, at long times, the evolution of the solution becomes smoother and does not require such fine sampling. In order to accurately capture the solution across multiple timescales, we adopt a hybrid time-sampling strategy. Logarithmic spacing is employed in the short-time regime to provide finer solution where rapid variations occur, and linear spacing is applied in the long-time regime to balance computational efficiency. The temporal domain $\mathcal{T}=[\tau_\varepsilon,\tau_c]\cup[\tau_c,\tau_{\max}]$, where $\tau_{\max}$ denotes the end of the simulation time, is discretized into n_τ points. Accordingly, the temporal domain is discretized logarithmically in the short-time regime and linearly in the long-time regime, as expressed

by

$$\mathcal{T} = \left\{\tau_i^{(\log)}\right\}_{i=1}^{N_{\log}} \cup \left\{\tau_j^{(\lin)}\right\}_{j=1}^{N_{\lin}}, \tag{14}$$

where

$$\tau_i^{(\log)} = 10^{\log_{10}\tau_\varepsilon + \frac{i-1}{N_{\log}-1}(\log_{10}\tau_c - \log_{10}\tau_\varepsilon)}, \quad i = 1, \ldots, N_{\log}, \tag{15a}$$

$$\tau_j^{(\lin)} = \tau_c + (j-1)\Delta\tau, \quad j = 1, \ldots, N_{\lin}, \tag{15b}$$

where $N_{\log}$ and $N_{\lin}$ are the numbers of points in the logarithmic and linear regimes, respectively. Here, τ_c marks the interface between the logarithmic and linear discretization regimes, and Δ_τ corresponds to the linear time step.

The neural network used in this study consists of L_{total} layers in total, including the input layer, L_{hidden} hidden layers, and the output layer, with n_l nodes in each hidden layer. Its trainable parameters are collectively denoted by $\boldsymbol{\theta}$. The overall available domain is $\mathcal{S} \times \mathcal{T} = [c, d] \times [\tau_\varepsilon, \tau_{\max}]$. The training set $\Omega = \{X_i^D = (x_{i_s}, \tau_{i_\tau}) \,|\, x_{i_s} \in [c, d], \quad \tau_{i_\tau} \in [\tau_\varepsilon, \tau_{\max}], \quad i = 1, \ldots, N_D\}$ comprises $N_D = n_s \times n_\tau$ points. Two additional subsets of the training data should be considered. The boundary set is defined as $\partial\Omega = \{X_i^B = (x_{i_s}, \tau_{i_\tau}) \,|\, x_{i_s} \in \{c, d\}, \ \tau_{i_\tau} \in [\tau_\varepsilon, \tau_{\max}]\}$, containing N_B points along the two edges. The initial set contains N_I points and is defined as $\mathcal{I} = \{X_i^I = (x_{i_s}, \tau_\varepsilon) \mid x_{i_s} \in [c, d]\}$. Each hidden layer utilizes the tanh activation function, while the output layer employs the SoftPlus function to guarantee the non-negativity of the predicted $\hat{f}(X_i^D)$ [66]. The loss function is therefore defined as [66,93]

$$\begin{aligned}\mathscr{L}_{\text{Loss}} = a_1 \cdot \frac{1}{N_D}\sum_{i=1}^{N_D} \left|\hat{f}_\tau\left(X_i^D|\boldsymbol{\theta}\right) - \mathcal{L}\left(\hat{f}\left(X_i^D|\boldsymbol{\theta}\right)\right)\right|^2 \\ + a_2 \cdot \frac{1}{N_B}\sum_{i=1}^{N_B} \left|\hat{f}\left(X_i^B|\boldsymbol{\theta}\right)\right|^2 \\ + a_3 \cdot \frac{1}{N_I}\sum_{i=1}^{N_I} \left|\hat{f}\left(X_i^I|\boldsymbol{\theta}\right) - f_0(x, \tau_\varepsilon)\right|^2,\end{aligned} \tag{16}$$

where the first term enforces the residual of the governing equations, and the remaining terms enforce the boundary and initial conditions. The coefficients $a_i (i = 1, 2, 3)$ are penalty weights whose values are chosen according to the auxiliary Markovian FPE being solved.

Algorithm 1 summarizes the overall workflow. The SNN solves the non-Markovian FPE by first addressing its Markovian counterpart, and for non-Markovian FPEs that share the same auxiliary Markovian formulation but differ in their memory kernels, the trained network can be reused without retraining.

ALGORITHM 1. The SNN algorithm.

Require: The training, boundary, and initial sets of Eq. (1); the number of training iterations $\mathscr{M}$.
1: Initialize all trainable neural network parameters $\boldsymbol{\theta}$.
2: **for** $j = 1, 2, \ldots, \mathscr{M}$ **do**
3: Input the training data and boundary data to obtain the approximate solution $f(x, \tau \mid \boldsymbol{\theta}_j)$.
4: Using automatic differentiation, compute $f_\tau(x, \tau \mid \boldsymbol{\theta}_j)$, $f_x(x, \tau \mid \boldsymbol{\theta}_j)$, and $f_{xx}(x, \tau \mid \boldsymbol{\theta}_j)$.
5: Construct the loss function according to Eq. (16), and update the network parameters by gradient-based optimization with the Adam optimizer:
$$\boldsymbol{\theta}_{j+1} \leftarrow \boldsymbol{\theta}_j - \nabla_{\boldsymbol{\theta}_j}\mathscr{L}_{\text{Loss}}(\boldsymbol{\theta}_j).$$
6: **end for**
7: Obtain the solution $f(x, \tau \mid \boldsymbol{\theta})$.
8: Compute the final solution $P(x, t)$ using Eq. (8).
9: **return:** Solution of Eq. (1)

IV. NUMERICAL EXAMPLES

In this section, we examine four representative one-dimensional numerical examples to evaluate the performance of the proposed SNN framework. The first two examples involve force-free diffusion governed by the Caputo and AB fractional derivatives in unbounded space, for which analytical solutions are known. The remaining examples incorporate more realistic features, for instance, external (non-)linear potentials and absorbing boundaries, making analytical solutions unavailable. For comparison with our SNN results, we additionally provide Monte Carlo (MC) simulations based on the stochastic representation of the subordinated process; see [81] for the detailed algorithm. We acknowledge that, for the one-dimensional problems considered here, the SNN approach does not yet show a substantial advantage over classical finite-difference methods in terms of CPU time or memory usage. The main aim of the present study is to provide an instructive computational framework for one-dimensional non-Markovian dynamics involving complex boundary conditions, external forces, and memory effects. In future work, this framework will be applied to higher-dimensional cases.

A consistent neural network architecture is adopted in this paper. It consists of $L_{\text{total}} = 5$ layers, including $L_{\text{hidden}} = 3$ hidden layers, with $n_l = 50$ neurons in each hidden layer, and a detailed sensitivity analysis of the neural network architecture (e.g., how the number of layers/neurons affects the convergence of the subordination integral) is further provided in Appendix C. All models are trained for 80 000 iterations to ensure convergence. On a desktop workstation equipped with an NVIDIA GeForce RTX 4090 GPU and an Intel Xeon Platinum 8352V CPU, the computational time ranged from 20 to 30 minutes. The remaining parameters for constructing the training dataset are selected as $\tau_\varepsilon = 0.001$, $\tau_c = 0.1$, $\Delta x = 0.05$. For simplicity and without loss of generality, we take the diffusion coefficient to be $K_\alpha = 1$. These parameters are consistently kept for all examples, unless stated otherwise. For the numerical treatment of models defined on unbounded (Examples 1 and 2) or semi-infinite spatial domains (Example 4), a suitable domain truncation is required. In our simulations, the computational domain is chosen as a finite and sufficiently large interval so that the PDF decays to a negligibly small value at the artificial boundaries within the time window considered, while keeping the computational cost manageable. We have also checked that further enlarging the computational domain leads to negligible changes in the reported PDF profiles and statistical quantities. Therefore,

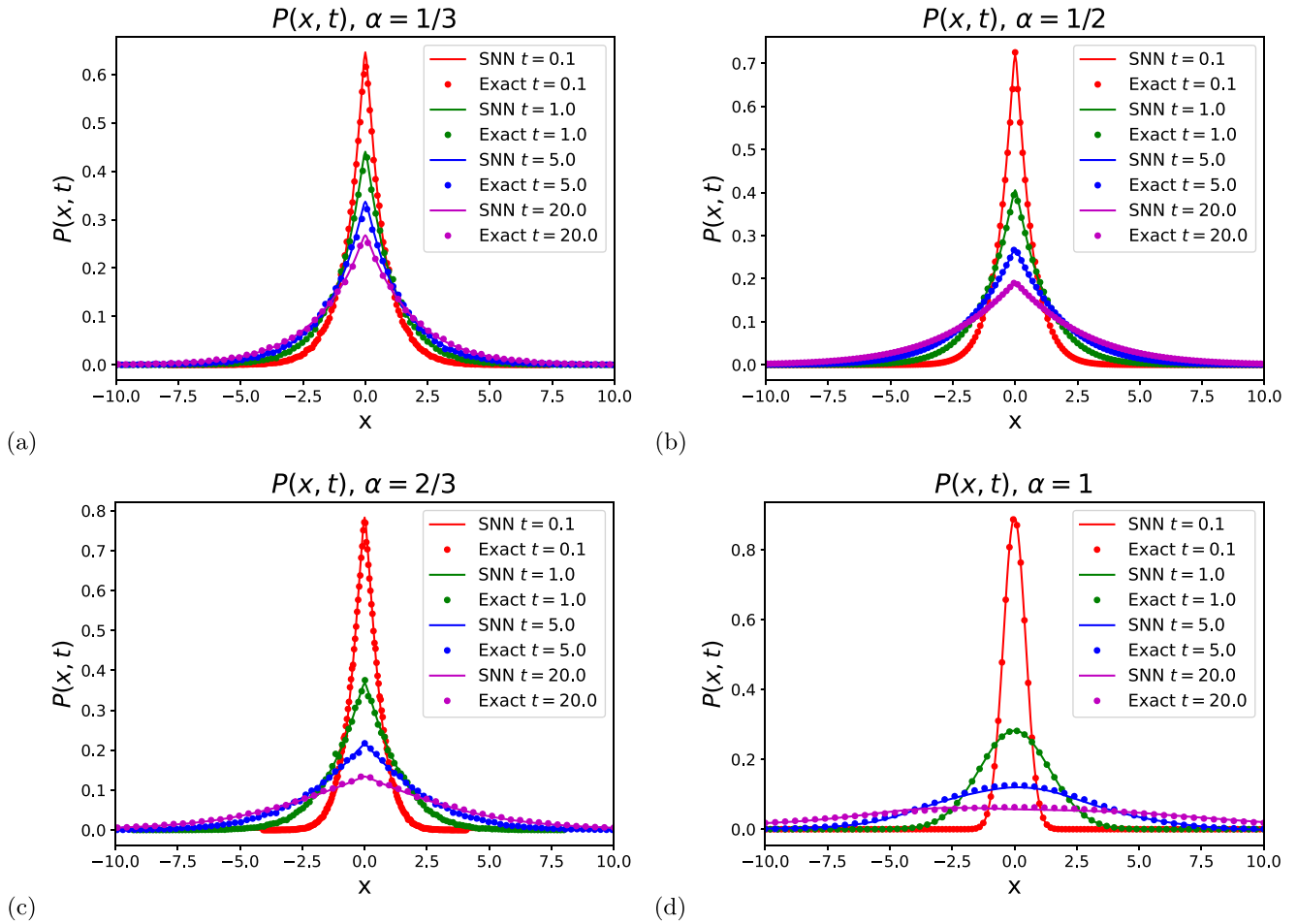


FIG. 2. SNN predictions (colored curves) for the PDFs of the non-Markovian FPE with Caputo derivative (17) at times $t = 0.1, 1, 5, 20$ with exponents: (a) $\alpha = 1/3$, (b) $\alpha = 1/2$, (c) $\alpha = 2/3$, and (d) $\alpha = 1$. For comparison, the corresponding analytical solutions given by Eq. (18) are shown as colored circles.

the artificial boundaries do not affect the numerical results presented in this work.

A. Example 1

Consider the one-dimensional non-Markovian FPE with a Caputo fractional derivative, subject to a Dirac delta initial condition and natural boundary conditions

$$\begin{aligned} {}^{C}D_t^{\alpha} P(x,t) &= K_\alpha \frac{\partial^2}{\partial x^2} P(x,t), \\ P(x,0) &= \delta(x), \\ P(\pm\infty, t) &= 0. \end{aligned} \tag{17}$$

Its analytical solution is given by [22,23]

$$P(x,t) = \frac{1}{\sqrt{4\pi K_\alpha t^\alpha}} H_{1,2}^{2,0}\left[\frac{x^2}{4K_\alpha t^\alpha} \,\middle|\, \begin{matrix} (1-\alpha/2, \alpha) \\ (0,1), (1/2, 1) \end{matrix} \right], \tag{18}$$

expressed in terms of the Fox H-function [94]. The asymptotic expansion of the H-function for $|x|/\sqrt{4K_\alpha t^\alpha} \gg 1$ yields the stretched Gaussian behavior of the PDF [23]

$$\begin{aligned} P(x,t) \sim \frac{1}{\sqrt{4\pi K_\alpha t^\alpha}} \frac{1}{\sqrt{2-\alpha}} \left(\frac{2}{\alpha}\right)^{\frac{1-\alpha}{2-\alpha}} \times \left(\frac{|x|}{\sqrt{K_\alpha t^\alpha}}\right)^{-\frac{1-\alpha}{2-\alpha}} \\ \times \exp\left(-\frac{2-\alpha}{2} \times \left(\frac{\alpha}{2}\right)^{\frac{\alpha}{2-\alpha}} \left(\frac{|x|}{\sqrt{K_\alpha t^\alpha}}\right)^{\frac{2}{2-\alpha}} \right). \end{aligned} \tag{19}$$

The training dataset is generated as follows. The spatial domain is truncated to $\Omega = [-10, 10]$ with $\Delta x = 0.05$. We verified numerically that this truncation has a negligible effect on the results for this example, as discussed above. The temporal domain $\mathcal{T} = [\tau_\varepsilon, 20]$ is constructed by combining a logarithmic grid of $N_{\log} = 20$ points for the short-time regime with a subsequent linear grid of $N_{\text{lin}} = 100$ points, using a linear time step $\Delta\tau = 0.2$. The penalty coefficients in the loss function are selected $a_1 = 10$, $a_2 = 1$, and $a_3 = 100$.

Figure 2 provides a detailed comparison between the SNN predictions and the analytical solution (18) at fixed times $t = 0.1$, 1.0, 5.0, and 20.0 for $\alpha = 1/3$, $1/2$, $2/3$, and 1. These results show that the SNN model accurately reproduces the solution over both short- and long-time regimes for all

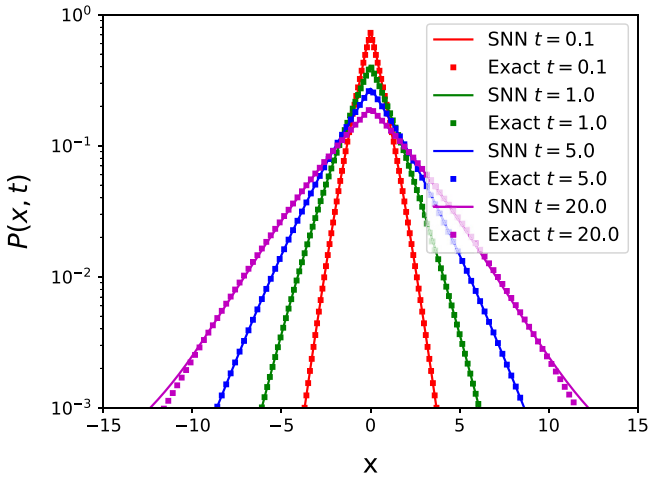


FIG. 3. SNN predictions (colored curves) for the PDFs of the non-Markovian FPE with Caputo derivative (17) at times $t = 0.1, 1, 5, 20$ with exponent $\alpha = 1/2$, in a logarithmic-linear scale. The predictions exhibit excellent agreement with the stretched-Gaussian form of the analytical solution (19), represented by colored circles. The data are identical to those shown in Fig. 2(b).

tested values of α, achieving high accuracy near the sharp central cusps. As illustrated in Fig. 2(d), when $\alpha = 1$, the equation reduces to the classical diffusion equation, and the solution becomes the Gaussian density with a smooth peak. To further verify the stretched Gaussian tail behavior for $\alpha \neq 1$, we examine the PDFs in a logarithmic-linear scale, as shown in Fig. 3. The SNN predictions remain in excellent agreement with the analytical results given by Eq. (18), demonstrating the robustness and accuracy of the proposed method.

B. Example 2

Next, we consider the non-Markovian FPE governed by a Mittag-Leffler memory kernel, also with a Dirac delta initial condition and natural boundary conditions [42]

$$\begin{aligned} {}^{AB}D_t^{\alpha} P(x,t) &= K_\alpha \frac{\partial^2}{\partial x^2} P(x,t), \\ P(x,0) &= \delta(x), \\ P(\pm\infty, t) &= 0. \end{aligned} \tag{20}$$

In the Laplace domain, Eq. (20) has the exact solution

$$\tilde{P}(x,s) = \frac{1}{2s}\phi(s)\exp\left(-\phi(s)|x|\right), \tag{21}$$

where $\phi(s) = \sqrt{\frac{s^\alpha}{(1-\alpha)s^\alpha + \alpha}}$.

Since the FPEs with the AB derivative (20) and the Caputo derivative (17) differ only in their memory kernels while sharing the auxiliary Markovian FPE with the same initial and boundary conditions, the neural network trained in the first example can be directly reused for the present case. This illustrates a key advantage of the proposed SNN framework. Applying the Laplace transform of the Mittag-Leffler memory kernel (12b), one can obtain the prediction of the SNN approach. We first present Fig. 4, which shows the heatmaps of the SNN prediction and the analytical solution of the FPE with the AB derivative in the Laplace domain, obtained from Eqs. (9) and (21), as well as the absolute error. The results show that the SNN prediction agrees very well with the analytical solution in the Laplace domain. Figure 5 then presents the corresponding comparison in the time domain, where both results are obtained by applying the numerical inverse Laplace transform. Excellent agreement is also observed after the inversion.

It is worth emphasizing again that when applying the SNN to predict solutions of a non-Markovian FPE with a different memory kernel, the previously trained network can be directly reused as long as the operator $\mathcal{L}$ and the initial and boundary conditions remain unchanged, i.e., the auxiliary Markovian FPEs are identical. In this case, the solution for the new kernel is obtained simply by updating the function $T(\tau, s)$ via Eq. (11) associated with the memory kernel, which leads to a substantial gain in computational efficiency.

C. Example 3

Consider the following non-Markovian FPE with a Caputo derivative and absorbing boundary conditions at $x = c$ and $x = d$, describing particle transport initialized at $x = 0.2$ within a tilted periodic potential [48,95–97],

$$\begin{aligned} {}^{C}D_t^{\alpha} P(x,t) &= -\frac{\partial}{\partial x}\left[\frac{F(x)}{m\gamma_\alpha} P(x,t)\right] + K_\alpha \frac{\partial^2}{\partial x^2} P(x,t), \\ P(x,0) &= \delta(x - 0.2), \\ P(c,t) &= P(d,t) = 0, \end{aligned} \tag{22}$$

where $F(x) = F_0[\sin(x/l) + \beta]$.

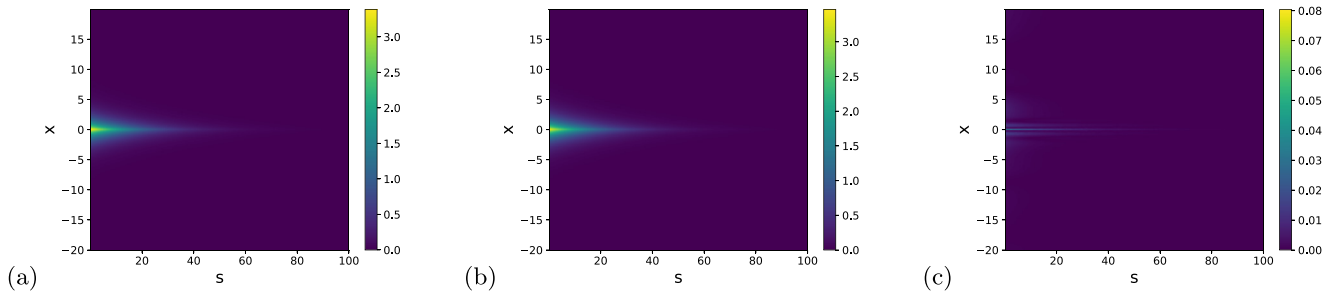


FIG. 4. Heatmaps in the Laplace domain for the non-Markovian FPE (20) with exponent $\alpha = 1/2$: (a) SNN prediction obtained from Eq. (9), (b) analytical solution given by Eq. (21), and (c) the corresponding absolute error between them. The deviations between the SNN prediction and analytical results remain small over the displayed (x, s) domain.

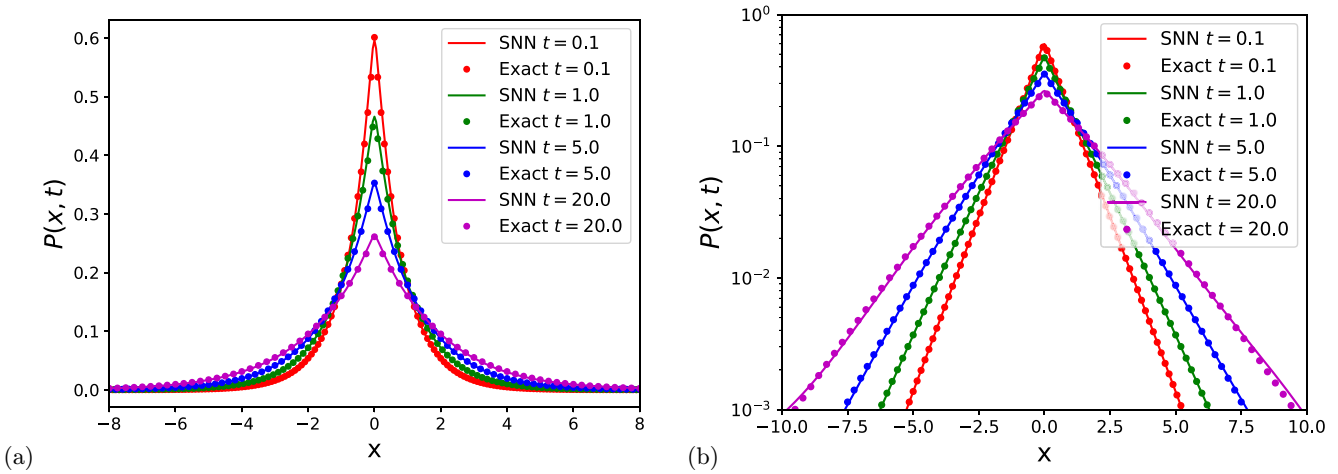


FIG. 5. SNN predictions (colored curves) for the PDFs of the non-Markovian FPE with AB derivative (20) at times $t = 0.1, 1, 5, 20$ with exponent $\alpha = 1/2$: (a) in a linear-linear scale, (b) in a logarithmic-linear scale. The SNN predictions and analytical solutions are obtained by performing the inverse Laplace transform of Eqs. (9) and (21). The predicted PDFs agree remarkably well with the analytical solution (21), indicated by the colored circles, capturing both the central behavior and the tail regions with high accuracy.

In this example, we set $K_\alpha = 2$, $F_0 = m\gamma_\alpha = 1$, $\beta = 6$ and domain boundaries $c = -2.8$ and $d = 3.2$. The temporal grid uses a hybrid scheme consisting of $N_{\log} = 30$ logarithmically spaced points and $N_{\text{lin}} = 80$ linearly spaced points with step size $\Delta\tau = 0.1$. The penalty factors are fixed to $a_1 = 4$, $a_2 = 20$, and $a_3 = 5$. Since the analytical solution of the FPE (22) is difficult to obtain, we here provide MC simulations for the FPE (22) based on the stochastic representation of the subordinated process [81]. A numerical predictor-corrector scheme for discretizing the FPE on a mesh grid can be found in [48].

As illustrated in Fig. 6, the SNN predictions for the PDFs remain highly accurate even under nonlinear force and show excellent agreement with the simulation results. In particular, when $\alpha = 1$, the process reduces to ordinary BM in the same confined potential: at short times, the PDF exhibits a Gaussian shape centered at $x = 0.2$, while at longer times it becomes increasingly asymmetric due to the influence of the external force. In contrast, for $\alpha \neq 1$, the PDFs are already asymmetric at short times. It is in fact connected to the inapplicability of the CTRW approximation for this short-time regime. Instead, the motion should physically be described by a telegrapher's type equation with finite propagation speed [98]. In addition, the survival probability $S(t) = \int_c^d P(x, t)\, dx$, defined as the integral of the PDF, is shown in Fig. 7 based on the SNN predictions and compared with MC simulations. As expected, for $\alpha = 1$, the survival probability decays exponentially, whereas for $\alpha \neq 1$, it follows an inverse power-law behavior, $S(t) \simeq t^{-\alpha}$.

Due to the absorbing boundaries, the particle's probability decreases over time, i.e., its survival probability tends to zero at long times. Obtaining a smooth and accurate PDF in MC simulations therefore requires a large number of samples, on the order of 10^5 or more, which leads to a substantial increase in computational cost. In contrast, the SNN maintains high accuracy and stability while requiring considerably less computational effort.

D. Example 4

The final example we consider is the Scher-Montroll model, which describes carrier transport in disordered materials such as amorphous semiconductors, polymers, and glassy solids [30,31,99–103]. Its central premise is that deep trapping events generate a heavy-tailed waiting-time distribution, causing carriers to exhibit strongly non-Gaussian and inherently nonequilibrium transport behavior. The transport dynamics can be described by a non-Markovian FPE with a Caputo derivative and a constant drift velocity $v_\alpha = F/(m\gamma_\alpha) > 0$, subject to a one-sided absorbing boundary at $x = a > 0$ and initialized at $x = 0$ [37]

$$
\begin{aligned}
{}^{C}D_t^{\alpha} P(x, t) &= -\frac{\partial}{\partial x} v_\alpha P(x, t) + K_\alpha \frac{\partial^2}{\partial x^2} P(x, t), \\
P(x, 0) &= \delta(x), \\
P(a, t) &= 0, P(-\infty, t) = 0.
\end{aligned} \tag{23}
$$

The system parameters are set to $a = 5$ and $K_\alpha = v_\alpha = 1$. The spatial domain is truncated to $[-6, 5]$, and as we verified, further enlargement of this domain has a negligible effect on the results in this example. The temporal domain consists of 60 logarithmic and 200 linear points with a step size of $\Delta_\tau = 0.1$. The penalty terms are weighted by $a_1 = 4$, $a_2 = 1$, and $a_3 = 1$.

The SNN predictions of the PDF are shown in Fig. 8, with excellent performance. The fact that the peak of the solution remains fixed at its initial position is fully consistent with a previous study [30]. At $t = 50$, as illustrated in Fig. 8, the peak value falls below 0.05, yet the fit remains excellent.

In the Scher-Montroll model, the transient current $I(t)$ (or equivalently, the mean carrier velocity $\langle v(t) \rangle$) demonstrates two power-law decay regimes [37]

$$
\langle v(t) \rangle \simeq \begin{cases} t^{-(1-\alpha)}, & t \ll t_\tau, \\ t^{-(1+\alpha)}, & t \gg t_\tau, \end{cases} \tag{24}
$$

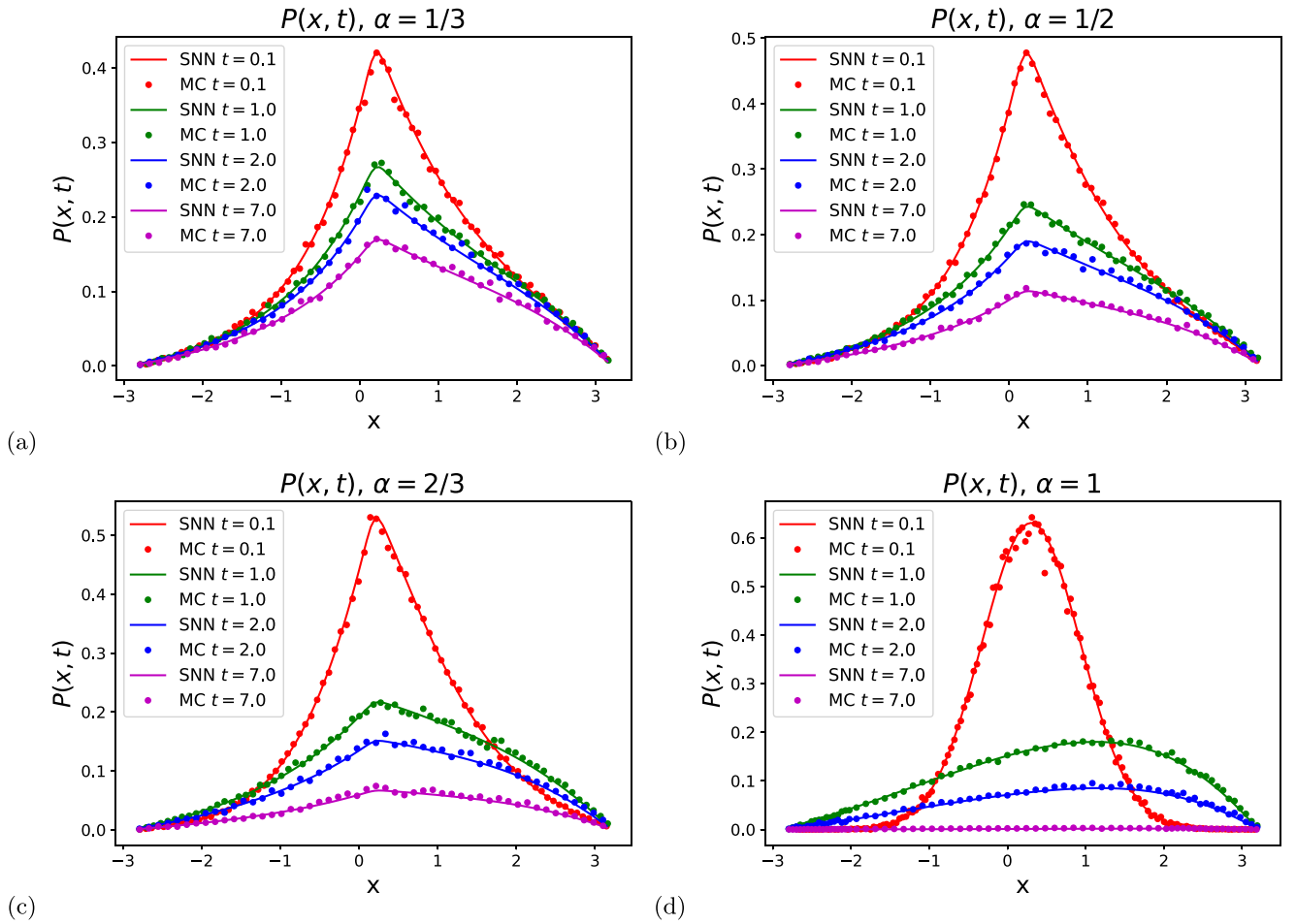


FIG. 6. SNN predictions (colored curves) for the PDFs of the non-Markovian FPE with a Caputo derivative and absorbing boundaries at $x = -2.8$ and $x = 3.2$ in a tilted periodic potential defined in Eq. (22) are shown at different times for (a) $\alpha = 1/3$, (b) $\alpha = 1/2$, (c) $\alpha = 2/3$, and (d) $\alpha = 1$. The corresponding MC simulation results for the same process are plotted as colored circles. For $\alpha = 1$ (normal diffusive particles), the PDF is approximately Gaussian and centered at $x = 0.2$ at short times, while it becomes non-Gaussian at long times. For $\alpha \neq 1$ (subdiffusive particles), the PDFs deviate from the Gaussian behavior already at the same short times as the influence of the external force becomes apparent much earlier.

with the characteristic crossover timescale

$$t_\tau = \left(\frac{\Gamma(1+\alpha)}{2\Gamma(1-\alpha)}\right)^{1/2\alpha} \left(\frac{a}{v_\alpha}\right)^{1/\alpha}. \tag{25}$$

Here, the mean carrier velocity $\langle v(t)\rangle$ is the time derivative of the mean displacement $\langle x(t)\rangle$ given by [37]

$$\langle x(t)\rangle = \int_{-\infty}^{a} xP(x,t)dx + [1 - S(t)]a, \tag{26}$$

where $S(t) = \int_{-\infty}^{a} P(x,t)\,dx$ is the survival probability. The first term in Eq. (26) represents the contribution from particles that have not yet reached the absorbing boundary, while the second term accounts for particles that have already been absorbed and remain at $x = a$.

In addition, when $\alpha = 1$, the exact survival probability can be obtained by the image method [104]

$$S(t) = \Phi\left(\frac{a - v_1 t}{\sqrt{2K_1 t}}\right) - e^{\frac{av_1}{K_1}} \Phi\left(-\frac{a + v_1 t}{\sqrt{2K_1 t}}\right), \tag{27}$$

where $\Phi(z)$ is a Gaussian cumulative distribution function $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{y^2}{2}}\, dy$ [105], and based on the formula (26), the mean velocity is given by

$$\langle v(t)\rangle = v_1 S(t), \tag{28}$$

which approaches v_1 when $t \to 0$ and decays to zero in the long-time limit $t \to \infty$.

The mean velocity $\langle v(t)\rangle$ for $\alpha = 1/2$ and 1.0 from the SNN predictions, are shown in Fig. 9. For $\alpha = 1/2$, the SNN predicted mean velocity exhibits the expected power-law scaling $\simeq t^{-(1-\alpha)}$ at short times and $\simeq t^{-(1+\alpha)}$ at long times, in agreement with the analytical expression in Eq. (24). The predicted mean velocity for $\alpha = 1$ displays a short-time plateau, with value equal to v_1, as particles have not yet reached the absorbing boundary and undergo normal diffusive motion with $\langle x(t)\rangle \simeq t$. As time advances and the absorbing boundary begins to take effect, the number of surviving particles decreases, causing $\langle v(t)\rangle$ to rapidly approach zero after a characteristic timescale, as expected from Eq. (28). Minor deviations observed at long times originate from numerical

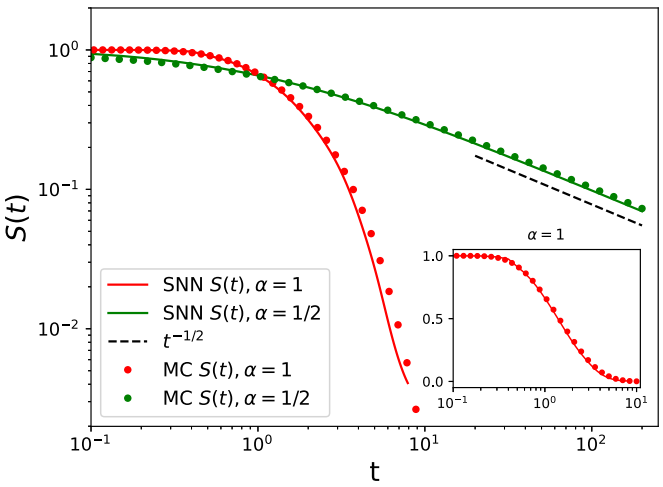


FIG. 7. SNN predictions (colored curves) for the survival probability $S(t)$ of the non-Markovian FPE with a Caputo derivative and absorbing boundaries in a tilted periodic potential defined in Eq. (22). The corresponding MC simulation results for the same process are plotted as colored circles. For $\alpha = 1$, the survival probability exhibits an exponential decay while it becomes inverse power-law $\simeq t^{-\alpha}$ for $\alpha \neq 1$.

differentiation in the low-probability regime, where the probability density becomes exceedingly small due to the absorbing boundary.

V. DISCUSSION AND CONCLUSIONS

In this work, we propose an effective machine learning framework, SNN, to solve non-Markovian FPEs with arbitrary memory kernels under complex nonlinear forces and boundary conditions. By leveraging the subordination scheme, the PDF of the non-Markovian process is reformulated through the PDF of a subordinated process, characterized by two components: one determined by the memory kernel and the other governed by an auxiliary Markovian FPE, which can be accurately and effectively approximated using a deep neural network. We validate the SNN approach through four representative examples from physics and engineering to demonstrate the high numerical accuracy and robustness in solving non-Markovian FPEs with complex memory kernels, nonlinear forces, and boundaries. Table I further quantifies the accuracy of the proposed SNN framework by reporting the relative errors defined by

$$\epsilon_r = \frac{\|P_{\rm SNN} - P_{\rm ref}\|}{\|P_{\rm ref}\|}. \tag{29}$$

Here, $\|\cdot\|$ denotes the discrete L^2 norm evaluated on the spatial grid, $P_{\rm SNN}$ denotes the SNN prediction, and $P_{\rm ref}$ denotes the corresponding reference solution, given by the analytical solution when available and by Monte Carlo simulations otherwise. Here, t_i $(i = 1, 2, 3, 4)$ denote the prescribed time points considered in each example, listed in ascending order.

Our SNN approach offers three key advantages. First, the SNN decomposes the solution of a non-Markovian FPE into a Markovian counterpart solved via a neural-network-based solver and an operational time $T(\tau, t)$ determined by the memory kernel, thereby avoiding spatial discretization and the computational complexity induced by long-range memory effects while providing a physically interpretable representation of the dynamics. Second, for non-Markovian FPEs that differ only in their memory kernels or parameter regimes but share the same force field and boundary conditions, the neural network needs to be trained only once, since the underlying auxiliary Markovian FPE remains identical, resulting in substantial savings in computational effort and memory. In Examples 1 and 2, a single neural network is trained for the auxiliary Markovian dynamics and reused across different parameter settings of the same underlying process. Third, the combined logarithmic-linear time-sampling strategy enables our proposed SNN to accurately capture transient dynamics across the entire temporal domain, including the challenging short-time regime. However, it should also be noted that the SNN framework has certain limitations and cannot be applied universally. In particular, the method requires that the PDF $T(\tau, t)$, determined by the memory $M(t)$ in the Laplace domain given by Eq. (11), remains completely monotonic. When this condition is violated, the SNN approach is no longer applicable. Moreover, in such cases, the boundary conditions of the corresponding non-Markovian FPE given by Eq. (1c) may become explicitly time-dependent.

TABLE I. Relative errors for each example at different times for $\alpha = 1/2$.

Example	t_1	t_2	t_3	t_4
1	2.92×10^{-2}	1.18×10^{-2}	6.43×10^{-3}	8.78×10^{-3}
2	2.95×10^{-2}	1.70×10^{-2}	1.05×10^{-2}	1.27×10^{-2}
3	4.80×10^{-2}	3.66×10^{-2}	3.32×10^{-2}	4.58×10^{-3}
4	1.12×10^{-2}	9.38×10^{-3}	1.50×10^{-3}	2.77×10^{-3}

Although the proposed SNN is able to obtain the solution of the non-Markovian FPEs, the result corresponds only to the one-point marginal PDF $P(x, t)$. For a genuinely non-Markovian stochastic process, however, the one-point PDF is insufficient to characterize the full dynamics: a complete statistical description requires multipoint joint PDFs, which encode temporal correlations, memory effects, and aging. Since these joint PDFs cannot be inferred from marginal distributions alone, the SNN framework cannot recover the full statistical structure of the underlying non-Markovian process. Similar limitations are also encountered in studies of time-dependent probability density evolution equations [47]. More advanced strategies will have to be developed for this purpose.

The SNN approach provides a potential route for tackling non-Markovian dynamics in high dimensions. Formally, the subordination scheme in Eq. (8) can be extended by replacing the one-dimensional spatial variable x with the vector $\mathbf{x} \in \mathbb{R}^d$, where the auxiliary density $f(\mathbf{x}, \tau)$ is governed by the corresponding d-dimensional Markovian FPE. For classical difference-based methods applied to integer-order FPEs, the computational complexity is typically of order $O(N_x^d N_t)$, where N_t and N_x^d denote the numbers of discretized grid points in time and in a d-dimensional space, respectively. For fractional or more general non-Markovian FPEs, how-

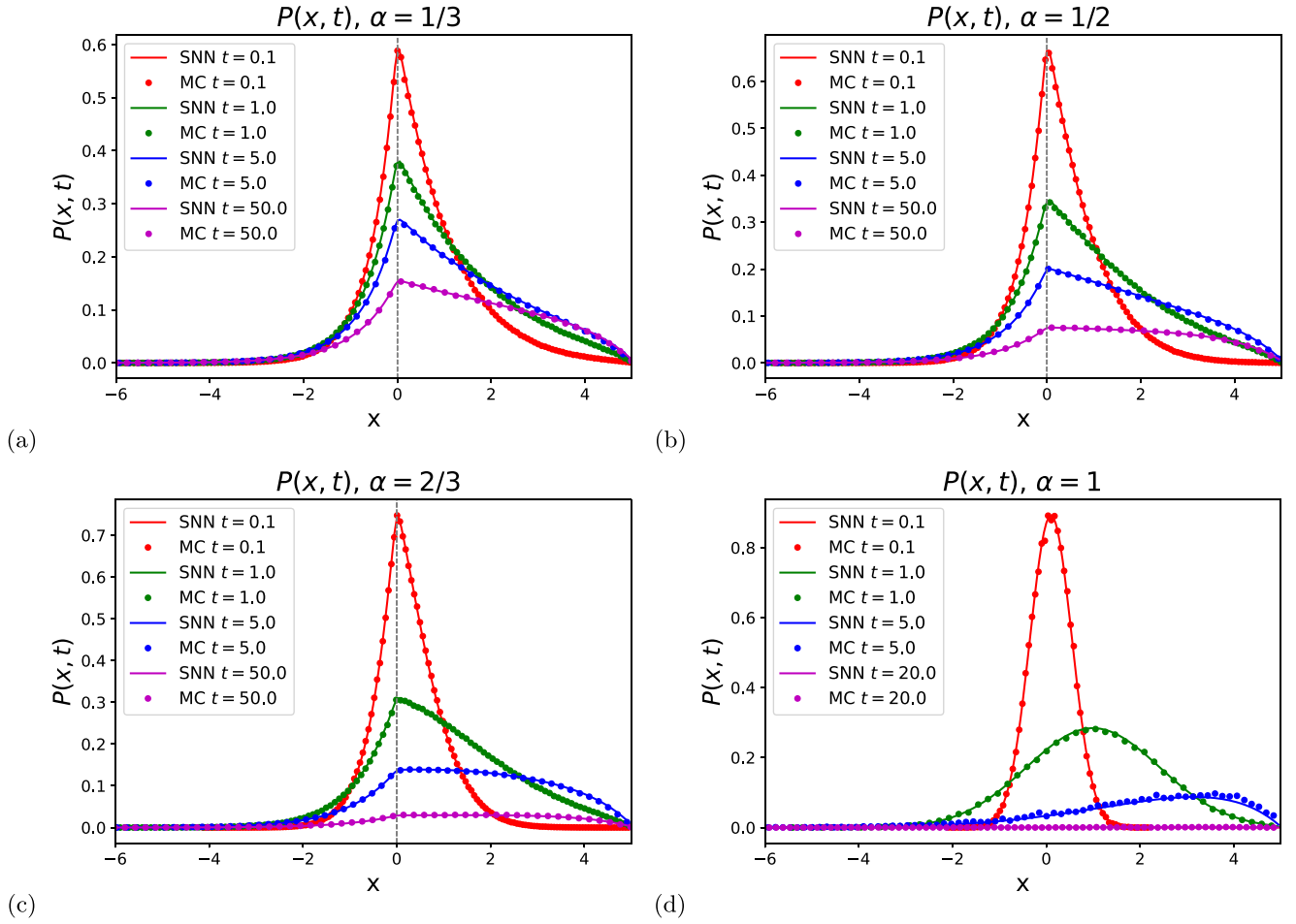


FIG. 8. SNN predictions (colored curves) for the PDFs of the non-Markovian FPE with a Caputo derivative, an absorbing boundary at $x = 5$, initialization at $x = 0$, and constant drift velocity $v_\alpha = 1$, as defined in Eq. (23), are shown at different times for (a) $\alpha = 1/3$, (b) $\alpha = 1/2$, (c) $\alpha = 2/3$, and (d) $\alpha = 1$. For comparison, the corresponding MC simulation results for the same process are plotted as colored circles.

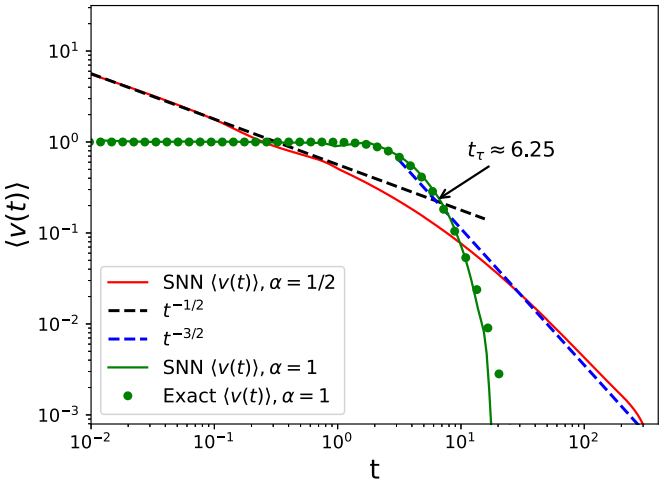


FIG. 9. Mean velocity $\langle v(t)\rangle$ versus time predicted by the SNN approach (colored solid curves) for the non-Markovian FPE (23) with $\alpha = 1$ and $\alpha = 1/2$. The exact mean velocity for $\alpha = 1$ given by Eq. (28) is represented by green circles. Black and blue dashed lines indicate the short-time scaling $\simeq t^{-(1-\alpha)}$ and the long-time scaling $\simeq t^{-(1+\alpha)}$. Arrows mark the crossover time t_τ given by Eq. (25).

ever, the complexity may increase to $O(N_x^d N_t^2)$ due to the long-range memory kernel, since the solution history over all previous time steps has to be stored and repeatedly evaluated. In contrast, the SNN approach avoids directly discretizing this memory term: it learns the auxiliary Markovian dynamics and incorporates the memory effect through the subordination kernel. Therefore, its computational cost is mainly associated with the neural-network training procedure, which depends on the network architecture, collocation points, and training iterations. Neural-network-based approaches for high-dimensional Markovian FPEs have been extensively studied, including Deep KD-tree methods [106], low-rank separated representations [107], score-based PINNs [108], and transfer learning strategies [109]. These developments provide useful foundations for extending the present SNN framework to high-dimensional non-Markovian FPEs. Nevertheless, several open issues remain. Compared with the present one-dimensional examples, high-dimensional problems involve more complex auxiliary Markovian dynamics, sparser spatial sampling, and more difficult boundary and tail treatments. As a result, local approximation errors may be amplified through the subordinated integral representation in Eq. (8). The overall error may

also arise from the numerical evaluation of the subordination kernel, as well as quadrature and truncation errors in the final integral, especially when the kernel has no closed-form expression. In addition, for $d = 2, 3$, the PDF associated with the fractional diffusion equation becomes singular at the origin [26], a delicate point requiring additional care. Therefore, a systematic study of error propagation, kernel approximation, and quadrature accuracy in high-dimensional non-Markovian systems remains an important direction for future work.

The SNN framework could be extended to FPEs with a distributed-order exponent α, in which the anomalous diffusion exponent varies with time according to a prescribed distribution $\lambda(\alpha)$ with $\int_0^1 \lambda(\alpha)d\alpha = 1$. Distributed-order diffusion equations and FPEs have been widely used to model subdiffusive and ultraslow diffusive processes, whose MSD grows logarithmically in time [110–112]. Since Ref. [110] establishes the connection between distributed-order time-fractional diffusion equations and more general CTRW frameworks, it is reasonable to expect that an SNN formulation for distributed-order FPEs may be developed in future studies. Another natural extension is to incorporate heterogeneous or spatially structured media, where the diffusion coefficient becomes position-dependent $D(x)$. Such systems arise in, for example, two-phase media or biological condensates that exhibit distinct diffusion regimes across different spatial regions, leading to PDFs with markedly different shapes in each phase [113–118]. Integrating these spatial heterogeneities would further broaden the applicability of the deep learning approach to realistic complex environments.

ACKNOWLEDGMENTS

We thank Andrey Cherstvy for valuable discussions. This work was partly supported by the NSF of China (Grant No. 12120101002). R.M. acknowledges financial support from the German Science Foundation (DFG Grants No. ME 1535/22-1 and No. 318763901, CRC1294 Project B10) and NSF-BMBF CRCNS (Grant No. 2112862/STAXS). G.W. would like to thank the China Scholarship Council (CSC) for its support.

DATA AVAILABILITY

The data that support the findings of this article are openly available [119].

APPENDIX A: CMFS AND BFS

Here, we first recall some definitions and properties of complete monotone functions (CMFs) and Bernstein functions (BFs) [86].

A CMF $\mathcal{M}(s)$ is defined on the non-negative semiaxis and has the property

$$(-1)^n \mathcal{M}^{(n)}(s) \geqslant 0 \quad \text{for all } n \in \mathrm{N}_0 \text{ and } s \geqslant 0. \tag{A1}$$

CMFs also have the following properties:

(a) A linear combination $a_1\mathcal{M}_1(s) + a_2\mathcal{M}_2(s)$ $(a_1, a_2 \geqslant 0)$ of CMFs $\mathcal{M}_1(s)$ and $\mathcal{M}_2(s)$ is a CMF as well.

(b) The product $\mathcal{M}(s) = \mathcal{M}_1(s)\mathcal{M}_2(s)$ of CMFs $\mathcal{M}_1(s)$ and $\mathcal{M}_2(s)$ is again a CMF.

A BF $\mathcal{B}(s)$ is a non-negative function if

$$(-1)^{(n)}\mathcal{B}^{(n+1)}(s) \geqslant 0, \text{ for all } n \in \mathrm{N}_0 \text{ and } s \geqslant 0. \tag{A2}$$

BFs have the following properties:

(c) A linear combination $a_1\mathcal{B}_1(s) + a_2\mathcal{B}_2(s)$ $(a_1, a_2 \geqslant 0)$ of BFs $\mathcal{B}_1(s)$ and $\mathcal{B}_2(s)$ is a BF.

(d) If $\mathcal{B}(s)$ is a BF, then $\mathcal{B}(s)/s$ is a CMF.

(e) A composition $\mathcal{M}(\mathcal{B}(s))$ of a CMF $\mathcal{M}(s)$ and a BF $\mathcal{B}(s)$ is a CMF.

(f) A composition $\mathcal{B}_1(\mathcal{B}_2(s))$ of BFs is again a BF.

(g) From these properties it follows that the function $e^{-u\,\mathcal{B}(s)}$ is a CMF for all $u > 0$ if $\mathcal{B}(s)$ is a BF.

With these properties, we prove that the Laplace-domain representations of Eqs. (12a) and (12b) are both CMFs with respect to s. For the power-law memory kernel equation (5), its Laplace transform is $\tilde{M}_C(s) = s^{\alpha-1}$, which is a CMF. Since s^α is a BF, using property (g), $e^{-\tau s^\alpha}$ is a CMF. Therefore, the product of CMFs is again a CMF, that is, $\tilde{T}_C(\tau, s) = s^{\alpha-1}e^{-\tau s^\alpha}$ is a CMF.

For the Mittag-Leffler memory kernel equation (7), we have

$$\tilde{M}_{AB}(s) = \frac{s^{\alpha-1}}{(1-\alpha)s^\alpha + \alpha}, \quad 0 < \alpha < 1. \tag{A3}$$

Let $h(s) = 1/\tilde{M}_{AB}(s)$, the first- and second-order derivatives of $h(s)$ are given by

$$h'(s) = (1-\alpha)\left[1 + \alpha s^{-\alpha}\right] > 0, \tag{A4}$$

$$h''(s) = -\alpha^2(1-\alpha)s^{-\alpha-1} < 0. \tag{A5}$$

By induction, the higher-order derivatives take the form

$$h^{(n)}(s) = \alpha(1-\alpha)(-1)^{n-1}\frac{\Gamma(\alpha+n-1)}{\Gamma(\alpha)}s^{-\alpha-(n-1)}, \tag{A6}$$

for $n \geqslant 2$. It follows that $(-1)^{n-1}h^{(n)}(s) \geqslant 0$ and hence $h(s)$ is a BF. Consequently, using property (e), $\tilde{M}_{AB}(s) = 1/h(s)$ is a CMF.

Defining $f(s) = s\,\tilde{M}_{AB}(s)$, one can rewrite it as

$$f(s) = \frac{1}{1-\alpha} - \frac{\alpha}{1-\alpha}\frac{1}{(1-\alpha)s^\alpha + \alpha}. \tag{A7}$$

Since $(1-\alpha)s^\alpha + \alpha$ is a BF, it follows that $\frac{1}{(1-\alpha)s^\alpha+\alpha}$ is a CMF. Hence $f(s)$ is a BF, and therefore $e^{-\tau f(s)}$ is a CMF. Consequently, the product

$$\tilde{T}_{AB}(\tau, s) = \tilde{M}_{AB}(s) \times \exp[-\tau f(s)], \tag{A8}$$

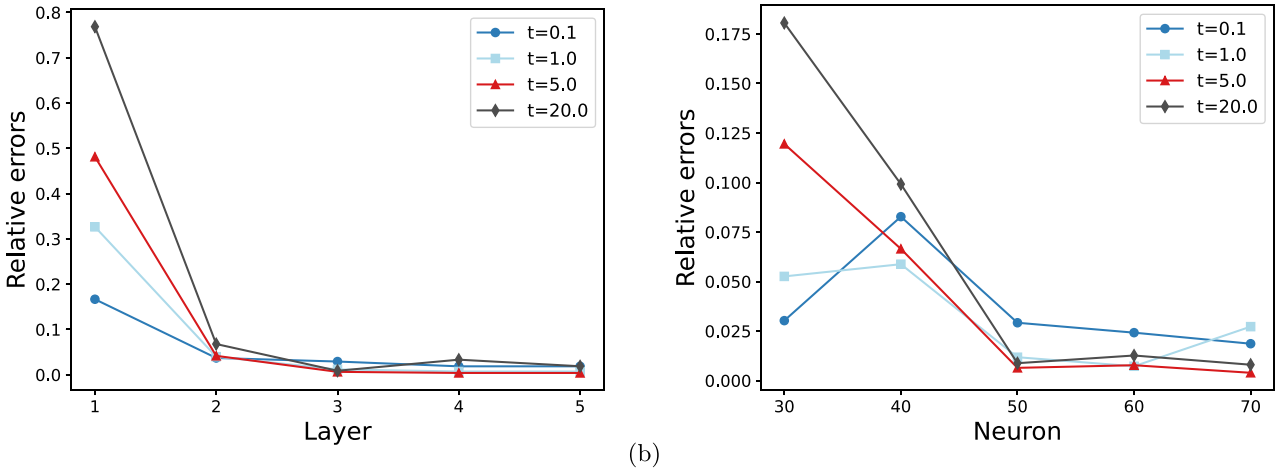
is also a CMF.

APPENDIX B: VALIDITY OF THE SUBORDINATION KERNEL

For the subordination kernel used in the present SNN framework,

$$\tilde{T}(\tau, s) = \tilde{M}(s)\exp(-\tau s\tilde{M}(s)), \tag{B1}$$

the essential requirement is that $\tilde{T}(\tau, s)$ is a CMF with respect to s. This guarantees that its inverse Laplace transform $T(\tau, t)$ is non-negative and can therefore be interpreted as a PDF of



(a) (b)

FIG. 10. Sensitivity of the PDF prediction accuracy to the neural network architecture for the non-Markovian FPE with the Caputo derivative (17), with $\alpha = 1/2$ in Example 1. The accuracy is measured by the relative error defined in Eq. (29). (a) Relative errors for different numbers of hidden layers L_{hidden}, with the number of neurons fixed to 50. (b) Relative errors for different numbers of neurons per hidden layer with the number of hidden layers fixed to three.

the operational time. In other words, the subordinated representation (8) preserves its meaning only when $T(\tau, t) \geqslant 0$ and $\int_0^\infty T(\tau, t)\, d\tau = 1$.

A convenient sufficient condition ensuring that $\tilde{T}(\tau, s)$ is a CMF can be formulated directly in terms of the time-domain memory kernel $M(t)$ appearing in the non-Markovian FPE (1). Suppose that $M(t)$ is a CMF, i.e.,

$$(-1)^n M^{(n)}(t) \geqslant 0, \quad n = 0, 1, 2, \ldots. \tag{B2}$$

This condition implies

$$M(t) \geqslant 0, \quad M'(t) \leqslant 0, \quad M''(t) \geqslant 0, \ldots, \tag{B3}$$

namely, the memory kernel is positive, monotonically decreasing, and has alternating derivative signs. Such kernels are typical of passive relaxation or trapping-type memory effects, including exponential and power-law kernels.

For such a completely monotone kernel $M(t)$, we can show that its Laplace transform $\tilde{M}(s)$ is a CMF with respect to s because $(-1)^n \hat{M}^{(n)}(s) = \int_0^\infty t^n e^{-st} M(t)\, dt \geqslant 0$.

Moreover, Bernstein's theorem [86] gives the representation $M(t) = \int_0^\infty e^{-rt}\, \mu(dr)$, with a non-negative measure μ. Hence $\hat{M}(s) = \int_0^\infty (s + r)^{-1}\, \mu(dr)$, and therefore $s\hat{M}(s) = \int_0^\infty s/(s + r)\, \mu(dr)$. For each $r \geqslant 0$, the function $s/(s + r)$ is a BF, since its derivative $r/(s + r)^2$ is a CMF. Therefore, the function

$$f(s) = s\tilde{M}(s), \tag{B4}$$

being a non-negative superposition of BFs, is also a BF.

Using property (g) in Appendix A, it follows that

$$\exp(-\tau s\tilde{M}(s)) = \exp\left(-\tau f(s)\right), \tag{B5}$$

is a CMF in s. Moreover, by property (b), the product of two CMFs is again a CMF. Hence

$$\tilde{T}(\tau, s) = \tilde{M}(s) \exp(-\tau s\tilde{M}(s)), \tag{B6}$$

is a CMF with respect to s.

This sufficient condition also clarifies the restriction of the method. When the memory kernel $M(t)$ is oscillatory, sign-changing, or involves active feedback, the function $f(s) = s\tilde{M}(s)$ may fail to be a BF. In such cases, the inverse kernel $T(\tau, t)$ is not guaranteed to be non-negative, and the subordinated SNN representation may lose its probabilistic interpretation. Consequently, although a formal integral representation may still be written, the resulting solution may fail to exist or be a valid PDF.

A representative example is provided by the FPE in the form

$$\frac{\partial}{\partial t} p(x, t) = \int_0^t \Phi(t - t') D \frac{\partial^2}{\partial x^2} p(x, t')\, dt', \tag{B7}$$

with diffusion coefficient D and exponential kernel

$$\Phi(t - t') = re^{-r(t-t')}. \tag{B8}$$

Equation (B7) can be transformed into the telegrapher's equation with finite propagation speed by differentiating both sides with respect to t [38,39]. Although Eq. (B7) differs from the form of Eq. (1), it can be recast into Eq. (1) with the corresponding memory kernel $M(t)$ given in the Laplace domain as

$$\tilde{M}(s) = 1/\tilde{\Phi}(s) = \frac{r + s}{r}. \tag{B9}$$

Using Eq. (B1), one obtains the subordination kernel in the Laplace domain

$$\tilde{T}(\tau, s) = (1 + sr^{-1}) \exp(-\tau s(1 + sr^{-1})). \tag{B10}$$

This function is not a CMF with respect to s, as its higher-order derivatives do not satisfy the required sign alternation. This indicates that the inverse Laplace transform $T(\tau, t)$ is not guaranteed to remain non-negative for all $t \geqslant 0$. Moreover, as shown in Ref. [39], the Laplace transform of any non-negative function normalized to unity cannot decay faster than exponentially as $s \to \infty$, which is not true for Eq. (B10).

APPENDIX C: PARAMETERS OF THE NEURAL NETWORK ARCHITECTURE

In our numerical implementation, the choice of network architecture is not arbitrary but is systematically examined. Figure 10 shows the influence of the number of hidden layers and the number of neurons per hidden layer on the relative error defined in Eq. (29) for predicting the PDFs of the non-Markovian FPE with the Caputo derivative, Eq. (17), with $\alpha = 1/2$ in Example 1. As shown in Fig. 10(a), once the number of hidden layers exceeds three, the relative error exhibits only minor variations. Since increasing the number of hidden layers also increases the number of trainable parameters and the training cost, three hidden layers are sufficient to obtain satisfactory accuracy. We further examine the effect of the number of neurons per hidden layer while fixing the number of hidden layers to three. The results in Fig. 10(b) show that using 50 neurons per hidden layer provides a good balance between accuracy and computational cost, yielding satisfactory performance.

Similar analyses were also carried out for Example 1 with other values of α, as well as for Examples 2–4. The same architecture, consisting of three hidden layers ($L_{\text{total}} = 5$) with 50 neurons per hidden layer, was found to be adequate in all cases. This consistency is expected because, in the SNN framework, the neural network approximates the auxiliary Markovian FPE, while the memory effects are introduced afterward through the subordination integral. Since all examples are one-dimensional and involve auxiliary Markovian problems of comparable complexity, the required network capacity remains similar across the different cases.

[1] R. Zwanzig, *Nonequilibrium Statistical Mechanics* (Oxford University Press, New York, 2001).

[2] H. Risken, *The Fokker–Planck Equation: Methods of Solution and Applications*, 2nd ed. (Springer-Verlag, Berlin, Heidelberg, 1996).

[3] A. Einstein, Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen, Ann. Phys. **322**, 549 (1905).

[4] D. Zhao, Y. Li, Y. Xu, Q. Liu, and J. Kurths, Probabilistic description of extreme oscillations and reliability analysis in rolling motion under stochastic excitation, Sci. China Technol. Sci. **66**, 2586 (2023).

[5] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 2007).

[6] H. Qian, Open-system nonequilibrium steady states: Statistical thermodynamics, fluctuations, and chemical oscillations, J. Phys. Chem. B **110**, 15063 (2006).

[7] M. Jeanblanc, M. Yor, and M. Chesney, *Mathematical Methods for Financial Markets* (Springer-Verlag London, London, 2009).

[8] J. A. Krommes, Fundamental statistical descriptions of plasma turbulence in magnetically confined plasmas, Phys. Rep. **360**, 1 (2002).

[9] M. Weiss, M. Elsner, F. Kartberg, and T. Nilsson, Anomalous subdiffusion is a measure for cytoplasmic crowding in living cells, Biophys. J. **87**, 3518 (2004).

[10] I. Golding and E. C. Cox, Physical nature of bacterial cytoplasm, Phys. Rev. Lett. **96**, 098102 (2006).

[11] J.-H. Jeon, V. Tejedor, S. Burov, E. Barkai, C. Selhuber-Unkel, K. Berg-Sørensen, L. Oddershede, and R. Metzler, *In vivo* anomalous diffusion and weak ergodicity breaking of lipid granules, Phys. Rev. Lett. **106**, 048103 (2011).

[12] D. Valverde-Mendez, A. M. Sunol, B. P. Bratton, M. Delarue, J. L. Hofmann, J. P. Sheehan, Z. Gitai, L. J. Holt, J. W. Shaevitz, and R. N. Zia, Macromolecular interactions and geometrical confinement determine the 3D diffusion of ribosome-sized particles in live Escherichia coli cells, Proc. Natl. Acad. Sci. USA **122**, e2406340121 (2025).

[13] F. Höfling and T. Franosch, Anomalous transport in the crowded world of biological cells, Rep. Prog. Phys. **76**, 046602 (2013).

[14] K. Norregaard, R. Metzler, C. M. Ritter, K. Berg-Sørensen, and L. B. Oddershede, Manipulation and motion of organelles and single molecules in living cells, Chem. Rev. **117**, 4342 (2017).

[15] S. Janušonis, N. Detering, R. Metzler, and T. Vojta, Serotonergic axons as fractional Brownian motion paths: Insights into the self-organization of regional densities, Front. Comput. Neurosci. **14**, 56 (2020).

[16] R. Klages, S. S. G. Gallegos, J. Solanpää, M. Sarvilahti, and E. Räsänen, Normal and anomalous diffusion in soft Lorentz gases, Phys. Rev. Lett. **122**, 064102 (2019).

[17] A. V. Weigel, B. Simon, M. M. Tamkun, and D. Krapf, Ergodic and nonergodic processes coexist in the plasma membrane as observed by single-molecule tracking, Proc. Natl. Acad. Sci. USA **108**, 6438 (2011).

[18] O. Vilk, E. Aghion, T. Avgar, C. Beta, O. Nagel, A. Sabri, R. Sarfati, D. K. Schwartz, M. Weiss, D. Krapf, R. Nathan, R. Metzler, and M. Assaf, Unravelling the origins of anomalous diffusion: From molecules to migrating storks, Phys. Rev. Res. **4**, 033055 (2022).

[19] O. Vilk, Y. Orchan, M. Charter, N. Ganot, S. Toledo, R. Nathan, and M. Assaf, Ergodicity breaking in area-restricted search of avian predators, Phys. Rev. X **12**, 031005 (2022).

[20] R. Metzler, J.-H. Jeon, A. G. Cherstvy, and E. Barkai, Anomalous diffusion models and their properties: Non-stationarity, non-ergodicity, and ageing at the centenary of single particle tracking, Phys. Chem. Chem. Phys. **16**, 24128 (2014).

[21] E. Barkai, Y. Garini, and R. Metzler, Strange kinetics of single molecules in living cells, Phys. Today **65**(8), 29 (2012).

[22] W. R. Schneider and W. Wyss, Fractional diffusion and wave equations, J. Math. Phys. **30**, 134 (1989).

[23] R. Metzler and J. Klafter, The random walk's guide to anomalous diffusion: A fractional dynamics approach, Phys. Rep. **339**, 1 (2000).

[24] R. Hilfer and L. Anton, Fractional master equations and fractal time random walks, Phys. Rev. E **51**, R848 (1995).

[25] A. Compte, Stochastic foundations of fractional dynamics, Phys. Rev. E **53**, 4191 (1996).

[26] E. Barkai, R. Metzler, and J. Klafter, From continuous time random walks to the fractional Fokker-Planck equation, Phys. Rev. E **61**, 132 (2000).

[27] R. Metzler, E. Barkai, and J. Klafter, Deriving fractional Fokker–Planck equations from a generalised master equation, Europhys. Lett. **46**, 431 (1999).

[28] R. Metzler, Non-homogeneous random walks, generalised master equations, fractional Fokker–Planck equations, and the generalised Kramers–Moyal expansion, Eur. Phys. J. B **19**, 249 (2001).

[29] S. M. A. Tabei, S. Burov, H. Y. Kim, A. Kuznetsov, T. Huynh, J. Jureller, L. H. Philipson, A. R. Dinner, and N. F. Scherer, Intracellular transport of insulin granules is a subordinated random walk, Proc. Natl. Acad. Sci. USA **110**, 4911 (2013).

[30] H. Scher and E. W. Montroll, Anomalous transit-time dispersion in amorphous solids, Phys. Rev. B **12**, 2455 (1975).

[31] M. Schubert, E. Preis, J. C. Blakesley, P. Pingel, U. Scherf, and D. Neher, Mobility relaxation and electron trapping in a donor/acceptor copolymer, Phys. Rev. B **87**, 024203 (2013).

[32] W. Reisner, K. J. Morton, R. Riehn, Y. M. Wang, Z. Yu, M. Rosen, J. C. Sturm, S. Y. Chou, E. Frey, and R. H. Austin, Statics and dynamics of single DNA molecules confined in nanochannels, Phys. Rev. Lett. **94**, 196101 (2005).

[33] B. Berkowitz, J. Klafter, R. Metzler, and H. Scher, Physical pictures of transport in heterogeneous media: Advection-dispersion, random-walk, and fractional derivative formulations, Water Resour. Res. **38**, 9 (2002).

[34] R. Metzler, A. Rajyaguru, and B. Berkowitz, Modelling anomalous diffusion in semi-infinite disordered systems and porous media, New J. Phys. **24**, 123004 (2022).

[35] A. Rajyaguru, R. Metzler, I. Dror, D. Grolimund, and B. Berkowitz, Diffusion in porous rock is anomalous, Environ. Sci. Technol. **58**, 8946 (2024).

[36] A. Rajyaguru, R. Metzler, A. G. Cherstvy, and B. Berkowitz, Quantifying anomalous chemical diffusion through disordered porous rock materials, Phys. Chem. Chem. Phys. **27**, 9056 (2025).

[37] E. Barkai, Fractional Fokker–Planck equation, solution, and application, Phys. Rev. E **63**, 046118 (2001).

[38] I. M. Sokolov, Solutions of a class of non-Markovian Fokker–Planck equations, Phys. Rev. E **66**, 041101 (2002).

[39] A. Chechkin and I. M. Sokolov, Relation between generalized diffusion equations and subordination schemes, Phys. Rev. E **103**, 032133 (2021).

[40] R. Zwanzig, Memory effects in irreversible thermodynamics, Phys. Rev. **124**, 983 (1961).

[41] A. Atangana and D. Baleanu, New fractional derivatives with nonlocal and non-singular kernel: Theory and application to heat transfer model, Therm. Sci. **20**, 763 (2016).

[42] Q. Wei, W. Wang, H. Zhou, R. Metzler, and A. Chechkin, Time-fractional Caputo derivative versus other integrodifferential operators in generalized Fokker-Planck and generalized Langevin equations, Phys. Rev. E **108**, 024125 (2023).

[43] T. R. Prabhakar, A singular integral equation with a generalized Mittag-Leffler function in the kernel, Yokohama Math. J. **19**, 7 (1971).

[44] A. A. Kilbas, M. Saigo, and R. K. Saxena, Generalized Mittag-Leffler function and generalized fractional calculus operators, Integr. Transforms Special Funct. **15**, 31 (2004).

[45] Ž. Tomovski, J. L. A. Dubbeldam, and J. Korbel, Applications of HP operator to option pricing financial model, Fract. Calc. Appl. Anal. **23**, 996 (2020).

[46] L. Feng, B. Pei, and Y. Xu, The memory-dependent FPK equation for fractional Gaussian noise, Probab. Eng. Mech. **82**, 103856 (2025).

[47] B. Pei, L. Feng, Y. Li, and Y. Xu, Non-Markovian dynamics: The memory-dependent probability density evolution equations, Nonlinear Dyn. **113**, 12589 (2025).

[48] W. Deng, Numerical algorithm for the time fractional Fokker–Planck equation, J. Comput. Phys. **227**, 1510 (2007).

[49] W. Deng, Finite element method for the space and time fractional Fokker–Planck equation, SIAM J. Numer. Anal. **47**, 204 (2009).

[50] Z. Odibat and S. Momani, Numerical solution of Fokker–Planck equation with space- and time-fractional derivatives, Phys. Lett. A **369**, 349 (2007).

[51] M. Zheng, F. Liu, I. Turner, and V. Anh, A novel high order space-time spectral method for the time fractional Fokker–Planck equation, SIAM J. Sci. Comput. **37**, A701 (2015).

[52] W. Zeng, J. He, and A. Xiao, A space-time Galerkin Müntz spectral method for the time fractional Fokker–Planck equation, Int. J. Comput. Math. **101**, 407 (2024).

[53] A. M. S. Mahdy and G. M. A. Marai, Numerical solution for the time-fractional Fokker–Planck equation using fractional power series method and the shifted Chebyshev polynomials of the third kind, Int. J. Appl. Eng. Res. **13**, 366 (2018).

[54] A. S. Mohamed, A. M. S. Mahdy, and A. H. Mtawa, Approximate analytical solution to a time-fractional Fokker–Planck equation, Bothalia **45**, 57 (2015).

[55] A. M. S. Mahdy, Numerical solutions for solving model time-fractional Fokker–Planck equation, Numerical Methods Partial Diff. Eq. **37**, 1120 (2021).

[56] A. M. S. Mahdy, K. A. Gepreel, Kh Lotfy, and A. El-Bary, Reduced differential transform and Sumudu transform methods for solving fractional financial models of awareness, Appl. Math. J. Chin. Univ. **38**, 338 (2023).

[57] G. Muñoz-Gil *et al.*, Objective comparison of methods to decode anomalous diffusion, Nat. Commun. **12**, 6253 (2021).

[58] G. Muñoz-Gil *et al.*, Quantitative evaluation of methods to analyze motion changes in single-particle experiments, Nat. Commun. **16**, 6749 (2025).

[59] G. Muñoz-Gil, M. A. Garcia-March, C. Manzo, J. D. Martín-Guerrero, and M. Lewenstein, Single trajectory characterization via machine learning, New J. Phys. **22**, 013010 (2020).

[60] H. Seckler and R. Metzler, Bayesian deep learning for error estimation in the analysis of anomalous diffusion, Nat. Commun. **13**, 6717 (2022).

[61] H. Seckler, J. Szwabiński, and R. Metzler, Machine-learning solutions for the analysis of single-particle diffusion trajectories, J. Phys. Chem. Lett. **14**, 7910 (2023).

[62] P. Kowalek, H. Loch-Olszewska, and J. Szwabiński, Classification of diffusion modes in single-particle tracking data: Feature-based versus deep-learning approach, Phys. Rev. E **100**, 032410 (2019).

[63] N. Granik, L. E. Weiss, E. Nehme, M. Levin, M. Chein, E. Perlson, Y. Roichman, and Y. Shechtman, Single-particle diffusion characterization by deep learning, Biophys. J. **117**, 185 (2019).

[64] 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, J. Comput. Phys. **378**, 686 (2019).

[65] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, Physics-informed machine learning, Nat. Rev. Phys. **3**, 422 (2021).

[66] Y. Xu, H. Zhang, Y. Li, K. Zhou, Q. Liu, and J. Kurths, Solving Fokker–Planck equation using deep learning, Chaos **30**, 013133 (2020).

[67] H. Zhang, Y. Xu, Y. Li, and J. Kurths, Statistical solution to SDEs with α-stable Lévy noise via deep neural network, Int. J. Dyn. Control. **8**, 1129 (2020).

[68] S. L. Brunton, J. L. Proctor, and J. N. Kutz, Discovering governing equations from data by sparse identification of nonlinear dynamical systems, Proc. Natl. Acad. Sci. USA **113**, 3932 (2016).

[69] K. Champion, B. Lusch, J. N. Kutz, and S. L. Brunton, Data-driven discovery of coordinates and governing equations, Proc. Natl. Acad. Sci. USA **116**, 22445 (2019).

[70] X. Wang, J. Feng, G. Wang, T. Li, and Y. Xu, The pseudo-analytical probability solution to parameterized Fokker–Planck equations via deep learning, Eng. Appl. Artif. Intell. **157**, 111344 (2025).

[71] X. Wang, J. Jiang, L. Hong, L. Chen, and J. Q. Sun, On the optimal design of radial basis function neural networks for nonlinear stochastic systems, Probab. Eng. Mech. **73**, 103470 (2023).

[72] W. Ye, L. Chen, J. Qian, and J. Q. Sun, RBFNN for calculating the stationary response of SDOF nonlinear systems excited by Poisson white noise, Int. J. Struct. Stab. Dyn. **23**, 2350019 (2023).

[73] J. L. Wei, G. C. Wu, B. Q. Liu, and Z. Zhao, New semi-analytical solutions of the time-fractional Fokker–Planck equation by the neural network method, Optik **259**, 168896 (2022).

[74] J. Hou, Z. Ma, S. Ying, and Y. Li, HNS: An efficient Hermite neural solver for time-fractional PDEs, Chaos Solitons Fract. **181**, 114637 (2024).

[75] S. Wang and G. E. Karniadakis, GMC-PINNs: A new general Monte Carlo PINNs method for solving fractional partial differential equations on irregular domains, Comput. Methods Appl. Mech. Eng. **429**, 117189 (2024).

[76] G. Pang, L. Lu, and G. E. Karniadakis, fPINNs: Fractional physics-informed neural networks, SIAM J. Sci. Comput. **41**, A2603 (2019).

[77] A. Baule and R. Friedrich, Joint probability distributions for a class of non-Markovian processes, Phys. Rev. E **71**, 026101 (2005).

[78] W. Feller, *An Introduction to Probability Theory and its Applications* (John Wiley & Sons, New York, 1968), Vols. 1 and 2.

[79] S. Bochner, Diffusion equation and stochastic processes, Proc. Natl. Acad. Sci. USA **35**, 368 (1949).

[80] M. Magdziarz and K. Weron, Anomalous diffusion schemes underlying the Cole–Cole relaxation: The role of the inverse-time α-stable subordinator, Physica A **367**, 1 (2006).

[81] M. Magdziarz, A. Weron, and K. Weron, Fractional Fokker–Planck dynamics: Stochastic representation and computer simulation, Phys. Rev. E **75**, 016708 (2007).

[82] M. Magdziarz, A. Weron, and J. Klafter, Equivalence of the fractional Fokker–Planck and subordinated Langevin equations: The case of a time-dependent force, Phys. Rev. Lett. **101**, 210601 (2008).

[83] M. Caputo, Linear models of dissipation whose Q is almost frequency independent, Ann. Geophys. **19**, 383 (1966).

[84] K. S. Miller and B. Ross, *An Introduction to the Fractional Calculus and Fractional Differential Equations* (Wiley-Blackwell, New York, 1993).

[85] R. Gorenflo, A. A. Kilbas, F. Mainardi, and S. V. Rogosin, *Mittag-Leffler Functions, Related Topics and Applications* (Springer, Berlin, 2020).

[86] R. L. Schilling, R. Song, and Z. Vondracek, *Bernstein Functions* (De Gruyter, Boston, 2010).

[87] R. Metzler, E. Barkai, and J. Klafter, Anomalous diffusion and relaxation close to thermal equilibrium: A fractional Fokker-Planck equation approach, Phys. Rev. Lett. **82**, 3563 (1999).

[88] E. Barkai and R. J. Silbey, Fractional Kramers equation, J. Phys. Chem. B **104**, 3866 (2000).

[89] K. A. Penson and K. Górska, On the properties of Laplace transform originating from one-sided Lévy stable laws, J. Phys. A: Math. Theor. **49**, 065201 (2016).

[90] A. Saa and R. Venegeroles, Alternative numerical computation of one-sided Lévy and Mittag-Leffler distributions, Phys. Rev. E **84**, 026702 (2011).

[91] J. Q. Sun, Random vibration analysis of a non-linear system with dry friction damping by the short-time Gaussian cell mapping method, J. Sound Vib. **180**, 785 (1995).

[92] X. Wang, J. Jiang, L. Hong, and J. Q. Sun, Stochastic bifurcations and transient dynamics of probability responses with radial basis function neural networks, Int. J. Non-Linear Mech. **147**, 104244 (2022).

[93] Y. Zhang, Z. Jin, L. Wang, K. Zheng, and W. Jia, Stochastic dynamics of aircraft ground taxiing via improved physics-informed neural networks, Nonlinear Dyn. **112**, 3163 (2024).

[94] A. M. Mathai, R. K. Saxena, and H. J. Haubold, The H-function, in *Theory and Applications* (Springer, New York, 2010).

[95] H. Sakaguchi, Generalized Einstein relation for Brownian motion in tilted periodic potential, J. Phys. Soc. Jpn. **75**, 124006 (2006).

[96] G. A. Pavliotis, A multiscale approach to Brownian motors, Phys. Lett. A **344**, 331 (2005).

[97] W. T. Coffey, Y. P. Kalmykov, S. V. Titov, and L. Cleary, Smoluchowski equation approach for quantum Brownian motion in a tilted periodic potential, Phys. Rev. E **78**, 031114 (2008).

[98] D. S. Grebenkov, R. Metzler, and G. Oshanin, Fastest first-passage time for multiple searchers with finite speed, arXiv:2602.15627.

[99] G. Pfister and H. Scher, Time-dependent electrical transport in amorphous solids: As_2Se_3, Phys. Rev. B **15**, 2062 (1977).

[100] G. Pfister and H. Scher, Dispersive (non-Gaussian) transient transport in disordered solids, Adv. Phys. **27**, 747 (1978).

[101] H. Krüsemann, A. Godec, and R. Metzler, First-passage statistics for aging diffusion in systems with annealed and quenched disorder, Phys. Rev. E **89**, 040101 (2014).
[102] H. Krüsemann, R. Schwarzl, and R. Metzler, Ageing Scher–Montroll transport, Transp. Porous Media **115**, 327 (2016).
[103] H. Scher, G. Margolin, R. Metzler, J. Klafter, and B. Berkowitz, The dynamical foundation of fractal stream chemistry: The origin of extremely long retention times, Geophys. Res. Lett. **29**, 5 (2002).
[104] S. Redner, *A Guide to First-Passage Processes* (Cambridge University Press, Cambridge, 2001).
[105] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (Dover, New York, 1972).
[106] H. Zhang, Y. Xu, Q. Liu, X. Wang, and Y. Li, Solving Fokker–Planck equations using deep KD-tree with a small amount of data, Nonlinear Dyn. **108**, 4029 (2022).
[107] H. Zhang, Y. Xu, Q. Liu, and Y. Li, Deep learning framework for solving Fokker–Planck equations with low-rank separation representation, Eng. Appl. Artif. Intell. **121**, 106036 (2023).
[108] Z. Hu, Z. Zhang, G. E. Karniadakis, and K. Kawaguchi, Score-based physics-informed neural networks for high-dimensional Fokker–Planck equations, SIAM J. Sci. Comput. **47**, C680 (2025).
[109] G. Wang, X. Wang, Q. Liu, J. Kurths, and Y. Xu, A transfer learning method to solve Fokker–Planck equation based on the equivalent linearization, Chaos **35**, 083119 (2025).
[110] A. V. Chechkin, R. Gorenflo, and I. M. Sokolov, Retarding subdiffusion and accelerating superdiffusion governed by distributed-order fractional diffusion equations, Phys. Rev. E **66**, 046129 (2002).
[111] T. Sandev, A. V. Chechkin, N. Korabel, H. Kantz, I. M. Sokolov, and R. Metzler, Distributed-order diffusion equations and multifractality: Models and solutions, Phys. Rev. E **92**, 042117 (2015).
[112] A. V. Chechkin, R. Gorenflo, I. M. Sokolov, and V. Yu Gonchar, Distributed order time fractional diffusion equation, Fract. Calc. Appl. Anal. **6**, 259 (2003).
[113] A. Pacheco-Pozo, M. Balcerek, A. Wyłomanska, K. Burnecki, I. M. Sokolov, and D. Krapf, Langevin equation in heterogeneous landscapes: How to choose the interpretation, Phys. Rev. Lett. **133**, 067102 (2024).
[114] S. Bo, L. Hubatsch, J. Bauermann, C. A. Weber, and F. Jülicher, Stochastic dynamics of single molecules across phase boundaries, Phys. Rev. Res. **3**, 043150 (2021).
[115] A. G. Cherstvy, A. V. Chechkin, and R. Metzler, Anomalous diffusion and ergodicity breaking in heterogeneous diffusion processes, New J. Phys. **15**, 083039 (2013).
[116] T. J. Doerries, A. V. Chechkin, and R. Metzler, Apparent anomalous diffusion and non-Gaussian distributions in a simple mobile–immobile transport model with Poissonian switching, J. R. Soc. Interface **19**, 20220233 (2022).
[117] T. J. Doerries, A. V. Chechkin, R. Schumer, and R. Metzler, Rate equations, spatial moments, and concentration profiles for mobile-immobile models with power-law and mixed waiting time distributions, Phys. Rev. E **105**, 014105 (2022).
[118] T. Akimoto, J.-H. Jeon, R. Metzler, T. Miyaguchi, T. Uneyama, and E. Yamamoto, Anomalous statistics in the Langevin equation with fluctuating diffusivity: From Brownian yet non-Gaussian diffusion to anomalous diffusion and ergodicity breaking, Rep. Prog. Phys. **89**, 014602 (2026).
[119] https://github.com/wgg26/SNN_for_Non-Markovian_FPE.