Capturing non-Markovian dynamics in non-equilibrium stochastic systems using flow matching

Bhargav Sriram Siddani
Lawrence Berkeley National Laboratory
bsiddani@lbl.gov

John B. Bell
Lawrence Berkeley National Laboratory
jbbell@lbl.gov

Alejandro L. Garcia
San Jose State University
alejandro.garcia@sjsu.edu

Ishan Srivastava
Lawrence Berkeley National Laboratory
isriva@lbl.gov

Abstract

Hydrodynamic models of stochastic particle systems represented by coarse-grained stochastic partial differential equations (SPDE), such as the regularized Dean-Kawasaki (DK) equation, do not accurately capture the short-time system dynamics that is dominated by non-Markovian effects, and low particle density regimes where the distributions are highly non-Gaussian. We develop a generative flow matching method that directly models the probability distribution of fluxes from particle simulations that explicitly incorporates non-Markovian and non-Gaussian effects. As a demonstration, we use this method to simulate the Kramers first passage time problem for a system of non-interacting Brownian particles. We show the model accurately captures the short-time behavior and provides better predictions of the statistical moments of the number density when compared against the solution of the Markovian baseline, regularized DK equation.

1 Introduction

A wide range of phenomena in areas as diverse as fluid dynamics, chemistry, biology and the social sciences can be described by stochastic particle systems. For small to modest sized systems, the dynamics can be modeled directly using particle-based simulations, such as molecular dynamics. However, for very large systems, direct simulation of particle dynamics becomes prohibitively expensive. One approach to reducing the computational cost of many-particle system simulations, originally introduced by Dean [1] and Kawasaki [2], is to coarse-grain the particle dynamics to obtain a stochastic partial differential equation (SPDE) for the number density, similar to the formulation of the dynamic density functional theory. Development and analysis of these types of coarse-grained SPDE models has been an active area of research for the past several decades. See [3, 4] for recent reviews of work in this area.

Here, we consider a representative system of N non-interacting Brownian particles under the influence of an external potential, $V(x)$. In the overdamped limit, the dynamics of the position, $\boldsymbol{X}_i(t)$, of the ith particle is given by the stochastic ODE

$$d\boldsymbol{X}_i = -\mu \nabla V(\boldsymbol{X}_i)dt + d\boldsymbol{B}_t^i,\ i = 1, 2, ..., N, \tag{1}$$

where μ is a Stokes mobility and the $\boldsymbol{B}_t$ are independent Brownian motions. An SPDE that describes the evolution of the particles can be formally derived using Itô's formula [1] but the resulting system,

Accepted to the Conference on Physics and AI at Stanford University (PAI 2026).

which includes the divergence of space-time white noise, is highly irregular. Typically, a coarse-grained regularized version of the SPDE is introduced by incorporating a high-frequency cutoff to the white noise. This results in a hydrodynamic equation for number density, q, of the form

$$dq = \nabla \cdot \frac{1}{2} \nabla q \, dt + \nabla \cdot \sqrt{q} \, d\boldsymbol{W}^M + \nabla \cdot (\mu \, q \nabla V) \, dt \,, \tag{2}$$

where $d\boldsymbol{W}^M$ represents the regularized noise. Here we regularize the noise by discretizing Eq. (2) on a one-dimensional finite volume mesh using an Euler-Maruyama discretization that gives

$$\frac{q_j^{n+1} - q_j^n}{\Delta t} = -\frac{(\bar{\boldsymbol{F}}^B_{j+1/2} - \bar{\boldsymbol{F}}^B_{j-1/2})}{\Delta x} - \frac{(\tilde{\boldsymbol{F}}^B_{j+1/2} - \tilde{\boldsymbol{F}}^B_{j-1/2})}{\Delta x} - \frac{(\boldsymbol{F}^V_{j+1/2} - \boldsymbol{F}^V_{j-1/2})}{\Delta x} \tag{3}$$

where the numerical fluxes are given by

$$\bar{\boldsymbol{F}}^B_{j+1/2} = -\frac{q^n_{j+1} - q^n_j}{2\Delta x}, \; \tilde{\boldsymbol{F}}^B_{j+1/2} = -A(q^n_j, q^n_{j+1}) \frac{Z^n_{j+1/2}}{\sqrt{\Delta x \Delta t}}, \; \boldsymbol{F}^V_{j+1/2} = -\frac{q^n_j + q^n_{j+1}}{2} \mu V'(x_{j+1/2}).$$

Here $Z^n_{j+1/2}$ are normally distributed random numbers, and

$$A(q_1, q_2) = \sqrt{\frac{\max(q_1, 0) + \max(q_2, 0)}{2}}$$

approximates $\sqrt{q}$ at cells faces while avoiding numerical issues if q becomes negative. Further details regarding the regularization and finite-volume discretization are discussed in [5].

The regularized DK equation has been shown to correctly predict long-time system statistics, compared to particle simulations, when the number of particles is sufficiently large in every finite-volume cell. However, there are significant differences in statistics when there are regions with low particle counts [5]. Moreover, it is now well-known that the dynamics of such diffusive, stochastic systems are significantly influenced by memory effects where the initial conditions impact the evolution of the system trajectories at early times [6, 7, 8]. Classical Mori-Zwanzig theory from statistical physics predicts coarse-graining to introduce a memory term into the dynamics (cf. [9]). In the present context, the memory reflects the sensitivity of the coarse-graining of the particles dynamics to initial conditions, which is lost in the regularized coarse-grained DK setting.

2 Background

In the present work, we alleviate the shortcomings of the regularized DK equation through data-driven learning for the Brownian diffusion flux, $\boldsymbol{F}^B \equiv \bar{\boldsymbol{F}}^B + \tilde{\boldsymbol{F}}^B$. Specifically we investigate the impact of including memory on the predictive capability of the machine learning (ML) model. Recent studies have used machine learning to model stochastic processes in dynamical systems. For example, diffusion models were used to learn Lagrangian trajectories of particles in turbulence [10, 11], and a statistics-informed neural network (SINN) framework [12] was developed to learn non-Markovian stochastic dynamics using recurrent neural networks trained with statistics-based loss functions. Furthermore, machine learning methods have been used to learn and parameterize the memory kernels of non-Markovian dynamics in the context of generalized Langevin equations [13, 14, 15].

3 Method

We define the current, $J(\boldsymbol{x}, t)$, as the net number of particles crossing a face between two adjacent computational cells from time step t to step $t + \Delta t$, and which is related to the flux $\boldsymbol{F}$ as $J = (\boldsymbol{F} \cdot \boldsymbol{A}) \, \Delta t$, where $\boldsymbol{A}$ is the face area. The probability density function (PDF) of J depends on the number of particles in the cells to the left and right of the face, N_L and N_R respectively, and exhibits non-Gaussian properties when either value is low [5]. Furthermore, the current $J(\boldsymbol{x}, t)$ is also non-Markovian, with probability distribution that depends on the history of the system [6, 7, 8]. However, the stochastic flux, $\tilde{\boldsymbol{F}}^B$, used in the regularized DK equation (Eq. 2) is a Gaussian distribution that is uncorrelated in space and time. Therefore, we use generative modeling based on the flow matching (FM) framework [16, 17, 18] to approximate the non-Markovian and non-Gaussian $J(\boldsymbol{x}, t)$ distribution from particle simulations to accurately predict short-time dynamics.

Flow matching overview Flow matching builds a probability path p_τ, $0 \leq \tau \leq 1$, from a known source distribution $p_0 = s$ to the target distribution $p_1 = r$. The goal of flow matching is to learn the parameters θ of a velocity field neural network, u_τ^θ. Once trained, the velocity field u_τ^θ can be used to convert a sample from the source distribution s to a sample from the target distribution r by solving the ordinary differential equation (ODE) from $\tau = 0$ to $\tau = 1$. The neural network u_τ^θ is trained using the conditional flow matching loss,

$$\mathcal{L}_{\text{CFM}}(\theta) = \mathbb{E}_{\tau, Z_\tau, Z_1} ||u_\tau^\theta(Z_\tau) - u_\tau(Z_\tau|Z_1)||^2, \tag{4}$$

where $\tau \sim U[0,1]$ (uniform distribution), $Z_0 \sim s$, $Z_1 \sim r$, $Z_\tau = a_\tau Z_0 + b_\tau Z_1$, and $u_\tau(Z_\tau|Z_1) = \dot{a}_\tau Z_0 + \dot{b}_\tau Z_1$ is called conditional velocity field.

Problem specific details We use flow matching to learn the PDF of $J(\boldsymbol{x}, t)$ conditioned upon the history of the local state of the system from k previous time steps, i.e., $(\boldsymbol{N}_L^k, \boldsymbol{N}_R^k, \boldsymbol{J}^k)$, where

$$\begin{aligned}
\boldsymbol{N}_L^k &= \{N_L(\boldsymbol{x}, t), N_L(\boldsymbol{x}, t - \Delta t), ..., N_L(\boldsymbol{x}, t - k\Delta t)\}, \\
\boldsymbol{N}_R^k &= \{N_R(\boldsymbol{x}, t), N_R(\boldsymbol{x}, t - \Delta t), ..., N_R(\boldsymbol{x}, t - k\Delta t)\}, \\
\boldsymbol{J}^k &= \{J(\boldsymbol{x}, t - \Delta t), J(\boldsymbol{x}, t - 2\Delta t), ..., J(\boldsymbol{x}, t - k\Delta t)\}.
\end{aligned}$$

Therefore, u_τ^θ takes $(\boldsymbol{N}_L^k, \boldsymbol{N}_R^k, \boldsymbol{J}^k)$ as additional inputs to capture the non-Markovian effects of $J(\boldsymbol{x}, t)$. To capture the large tails of the target distribution, we chose the source distribution s to be the Student's t-distribution with a zero mean, unity scale, and a value of 4 for degrees of freedom, $a_\tau = \cos\left(\frac{1}{2}\pi\tau\right)$ and $b_\tau = \sin\left(\frac{1}{2}\pi\tau\right)$. Each mini-batch of training samples Z_1 was generated from a one-dimensional, equilibrium distribution of Brownian walkers across 186 computational cells, where the average number of particles per cell varied between 1 and 50, and the history length k was sampled uniformly between 0 and 10. The neural network architecture uses two separate branches depending on k. A DeepONet-based architecture [19] was used for $k = 0$, whereas a Transformer-based architecture [20] was used for $k > 0$.

Statistical reflection symmetry Symmetry of particle current dictates that the PDF of $J(\boldsymbol{x}, t, N_L = a, N_R = b)$ is the same as the PDF of $-J(\boldsymbol{x}, t, N_L = b, N_R = a)$. In the FM model this implies that

$$u_\tau^\theta(Z_\tau, \boldsymbol{N}_L^k, \boldsymbol{N}_R^k, \boldsymbol{J}^k) = -u_\tau^\theta(-Z_\tau, \mathcal{R}(\boldsymbol{N}_L^k), \mathcal{R}(\boldsymbol{N}_R^k), -\boldsymbol{J}^k), \tag{5}$$

where $\mathcal{R}(\boldsymbol{N}_L^k) = \boldsymbol{N}_R^k$, and $\mathcal{R}(\boldsymbol{N}_R^k) = \boldsymbol{N}_L^k$. This property was enforced by reconstructing u_τ^θ as:

$$u_\tau^\theta(Z_\tau, \boldsymbol{N}_L^k, \boldsymbol{N}_R^k, \boldsymbol{J}^k) = 0.5\left[v_\tau^\theta(Z_\tau, \boldsymbol{N}_L^k, \boldsymbol{N}_R^k, \boldsymbol{J}^k) - v_\tau^\theta\left(-Z_\tau, \mathcal{R}(\boldsymbol{N}_L^k), \mathcal{R}(\boldsymbol{N}_R^k), -\boldsymbol{J}^k\right)\right], \tag{6}$$

where v_τ^θ has the optimizable parameters.

4 Results

We tested our numerical method on a variant of the Kramers first passage time problem [21]. We simulated 5120 independent realizations of a one-dimensional system consisting of 100 finite-volume cells of size $\Delta x = 1 \times 10^{-2}$. The time step for evolving both the SPDE and the Brownian particle simulation was chosen as $\Delta t = 3 \times 10^{-6}$ with homogeneous Dirichlet boundary conditions at the two ends of the domain. The dynamics includes an external potential of the form $V(x) = (x-\alpha)^2(x-\beta)^2$, where $\alpha = 0.3$ and $\beta = 0.7$. The external potential has minima at $x = 0.3$ and $x = 0.7$, and a local maximum at $x = 0.5$. The system was initialized with 10^3 total particles that were distributed in the region from $x = 0.2$ to $x = 0.4$ based on an equilibrium distribution under the external potential, $\frac{1}{\mathcal{Z}}\exp(-2\mu V(x))$, where $\mathcal{Z}$ is the normalization constant and $\mu = 2 \times 10^3$. The primary objective in this numerical experiment was to track the rate at which particles transition from the region around $x = \alpha$ into the region around $x = \beta$, across the energy barrier of height V_b at $x = 0.5$. We computed the particle statistics in the target region around $x = \beta$ where $V(x) \leq \frac{1}{2}V_b$.

Each system was simulated for $\sim 3 \times 10^3$ steps using the three models: Brownian random-walkers, regularized DK model, and the FM model where the SPDE flux $\boldsymbol{F}^B$ is generated by the FM method. To emphasize non-Markovian effects, the simulations were run with the FM models both with a history length of $k = 10$, and the Markovian limit without history, i.e., $k = 0$. In the FM method, a uniform 20 step Euler time discretization was used in the ODE solver for sampling the target

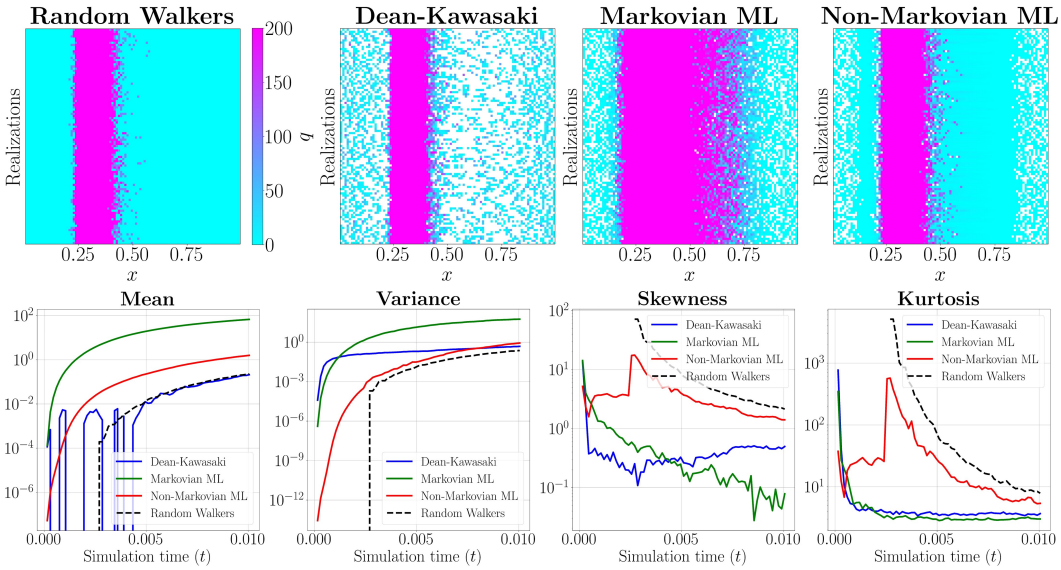

Figure 1: The top row shows a subset of 100 independent one-dimensional realizations (along vertical direction) at time $t = 5.1 \times 10^{-3}$ simulated using Brownian random walkers, the regularized Dean–Kawasaki (DK) model, and the two flow matching (FM) models: Markovian and non-Markovian. The white patches denote negative values in the simulation. The bottom row compares the time evolution of the statistics of particle count in the target region around $x = 0.7$.

distribution. All the simulations were performed on CPUs, and the most expensive non-Markovian FM model scaled well up to 100 cells per CPU, and averaging 9 seconds per time step.

Figure 1 depicts independent realizations of an early-time state of the system. The regularized DK model results in high occurrence of nonphysical negative densities as compared to the ML models. While the DK model well-predicts mean particle densities in the target region, the non-Markovian ML model more accurately captures the spatial particle distribution and also the higher-order statistical moments (see bottom row), thus demonstrating significant non-Markovian effects at early times.

Code and data availability: The code and data are open-source and available at the following links: code and data.

5 Conclusion

In this work, we used a generative flow matching framework to model non-Gaussian and non-Markovian effects (i.e., memory) in coarse-grained stochastic partial differential equations that represent stochastic particle systems. Using non-interacting Brownian particles as an example, we showed that the proposed approach more accurately captures short-time dynamics than Markovian models such as the DK equation. Furthermore, our flow matching based method reproduced higher-order statistics reasonably well at early times of the system evolution. However, at later times, the model predictions began to deviate significantly from the particle simulations, with the error increasing with time. While the current methodology is more computationally expensive than non-interacting Brownian dynamics simulations, it is expected to be significantly more efficient for stochastic particle systems with complex, long-ranged interactions, where costly neighbor-list construction and time step constraints incur significant computational costs.

Acknowledgments and Disclosure of Funding

This work was supported by the LDRD program of Lawrence Berkeley National Laboratory under U.S. DOE Contract No. DE-AC02-05CH11231, and by the U.S. DOE, Office of Science (SC), Office of ASCR, Applied Mathematics Program under Contract No. DE-AC02-05CH11231. This research used the resources of the NERSC, a DOE-SC User Facility, under Contract No. DE-AC02-05CH11231.

References

[1] David S Dean. Langevin equation for the density of a system of interacting Langevin processes. *Journal of Physics A: Mathematical and General*, 29(24):L613–L617, December 1996.

[2] Kyozi Kawasaki. Stochastic model of slow dynamics in supercooled liquids and dense colloidal suspensions. *Physica A: Statistical Mechanics and its Applications*, 208(1):35–64, 1994.

[3] Pierre Illien. The dean–kawasaki equation and stochastic density functional theory. *Reports on Progress in Physics*, 88(8):086601, 2025.

[4] Michael Te Vrugt and Raphael Wittkowski. Perspective: New directions in dynamical density functional theory. *Journal of Physics: Condensed Matter*, 35(4):041501, 2023.

[5] Ana Djurdjevac, Ann Almgren, and John Bell. A hybrid algorithm for systems of noninteracting particles with an external potential. *Communications in Applied Mathematics and Computational Science*, 20(1):147–174, 2025.

[6] Bernard Derrida and Antoine Gerschenfeld. Current fluctuations in one dimensional diffusive systems with a step initial density profile. *Journal of Statistical Physics*, 137(5):978–1000, 2009.

[7] N Leibovich and E Barkai. Everlasting effect of initial conditions on single-file diffusion. *Physical Review E*, 88(3):032107, 2013.

[8] Tirthankar Banerjee, Robert L Jack, and Michael E Cates. Role of initial conditions in one-dimensional diffusive systems: Compressibility, hyperuniformity, and long-term memory. *Physical Review E*, 106(6):L062101, 2022.

[9] Hermann Grabert. *Projection operator techniques in nonequilibrium statistical mechanics*. Springer, 2006.

[10] Tianyi Li, Luca Biferale, Fabio Bonaccorso, Martino Andrea Scarpolini, and Michele Buzzicotti. Synthetic lagrangian turbulence by generative diffusion models. *Nature Machine Intelligence*, 6(4):393–403, 2024.

[11] Tianyi Li, Samuele Tommasi, Michele Buzzicotti, Fabio Bonaccorso, and Luca Biferale. Generative diffusion models for synthetic trajectories of heavy and light particles in turbulence. *International Journal of Multiphase Flow*, 181:104980, 2024.

[12] Yuanran Zhu, Yu-Hang Tang, and Changho Kim. Learning stochastic dynamics with statistics-informed neural network. *Journal of Computational Physics*, 474:111819, 2023.

[13] Pinchen Xie, Roberto Car, and Weinan E. Ab initio generalized langevin equation. *Proceedings of the National Academy of Sciences*, 121(14):e2308668121, 2024.

[14] Zhiyuan She, Pei Ge, and Huan Lei. Data-driven construction of stochastic reduced dynamics encoded with non-markovian features. *The Journal of Chemical Physics*, 158(3), 2023.

[15] Hardeep Bassi, Yuanran Zhu, Senwei Liang, Jia Yin, Cian C Reeves, Vojtěch Vlček, and Chao Yang. Learning nonlinear integral operators via recurrent neural networks and its application in solving integro-differential equations. *Machine Learning with Applications*, 15:100524, 2024.

[16] Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*, 2023.

[17] Yaron Lipman, Marton Havasi, Peter Holderrieth, Neta Shaul, Matt Le, Brian Karrer, Ricky T. Q. Chen, David Lopez-Paz, Heli Ben-Hamu, and Itai Gat. Flow matching guide and code, 2024.

[18] Nanye Ma, Mark Goldstein, Michael S. Albergo, Nicholas M. Boffi, Eric Vanden-Eijnden, and Saining Xie. Sit: Exploring flow and diffusion-based generative models with scalable interpolant transformers. In Aleš Leonardis, Elisa Ricci, Stefan Roth, Olga Russakovsky, Torsten Sattler, and Gül Varol, editors, *Computer Vision – ECCV 2024*, pages 23–40, Cham, 2024. Springer Nature Switzerland.

[19] Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via deeponet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3):218–229, March 2021.

[20] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[21] Hendrik Anthony Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7(4):284–304, 1940.