

Studying first passage problems using neural networks: A case study in the slit-well microfluidic device

Andrew M. Nagel , Martin Magill, and Hendrick W. de Haan*

Faculty of Science, University of Ontario Institute of Technology, 2000 Simcoe St N, Oshawa, Ontario, Canada L1H7K4



(Received 27 April 2022; accepted 28 July 2022; published 11 August 2022)

This study presents deep neural network solutions to a time-integrated Smoluchowski equation modeling the mean first passage time of nanoparticles traversing the slit-well microfluidic device. This physical scenario is representative of a broader class of parametrized first passage problems in which key output metrics are dictated by a complicated interplay of problem parameters and system geometry. Specifically, whereas these types of problems are commonly studied using particle simulations of stochastic differential equation models, here the corresponding partial differential equation model is solved using a method based on deep neural networks. The results illustrate that the neural network method is synergistic with the time-integrated Smoluchowski model: together, these are used to construct continuous mappings from key physical inputs (applied voltage and particle diameter) to key output metrics (mean first passage time and effective mobility). In particular, this capability is a unique advantage of the time-integrated Smoluchowski model over the corresponding stochastic differential equation models. Furthermore, the neural network method is demonstrated to easily and reliably handle geometry-modifying parameters, which is generally difficult to accomplish using other methods.

DOI: [10.1103/PhysRevE.106.025311](https://doi.org/10.1103/PhysRevE.106.025311)

I. INTRODUCTION

Micro- and nanofluidic devices (MNFDs) are tools that can be used to manipulate or detect molecules with high precision [1–5]. For instance, the slit-well MNFD was proposed by Han and Craighead [6] as a tool for sorting otherwise free-draining polymers, such as DNA, according to chain length. The same device has also been shown to induce separation of free-draining nanoparticles by size [7,8]. The slit-well is operated by electrophoretically forcing analytes across a periodic array of deeper regions (wells) and shallower regions (slits) between two fixed planes (see Fig. 1). Its sorting effect has been comprehensively studied through theoretical, numerical, and experimental investigations, which have identified a variety of distinct mechanisms that are relevant in different operational regimes. At a high level, the sorting effect depends nonlinearly on the size of the analytes and the magnitude of the applied electric field, as well as the shape and size of the device's wells and slits [2,7–11]. In particular, depending on the choice of these parameters, the mobility of analytes can be made either increasing or decreasing with respect to molecule size.

The practical relevance of biotechnologies such as MNFDs has been stressed in the last year; for instance, Berkenbrock *et al.* [12] surveyed the potential of microfluidics as a means of rapidly testing large numbers of people for COVID-19 infections. Shepherd *et al.* [13] studied a parallelized MNFD that generated scalable lipid nanoparticle formulations needed for applications in RNA therapeutics and vaccines. Nonetheless, the design and optimization of MNFDs is often challenging

because it entails simultaneously considering the influence of many design parameters (e.g., operating voltage, solvent composition, device geometry, etc.) on multiple nonlinearly interdependent phenomena.

In many cases, important biological phenomena can fruitfully be modeled as first passage processes [14]. Moreover, in the study of MNFDs, key transport phenomena are often captured by only the first few moments of an appropriate first passage time distribution. For example, the translocation of a polymer through a nanopore is aptly described as a first passage process, and the mean translocation time is a widely studied metric [15–17]. Magill *et al.* [18] showed that, for the special class of MNFDs with periodic geometries featuring small bottlenecks, the long-term dynamics of molecules driven through the system depend exclusively on the first and second moments of their first passage times across one subunit of the device. The ability to focus on a handful of first passage time moments can greatly simplify the problem of characterizing and designing MNFDs.

This emphasis on the first few moments of the first passage time is of particular interest in light of a convenient mathematical property of the Smoluchowski equation¹ that describes the motion of analytes through MNFDs. For instance, the dynamics of nanoparticles electrophoretically driven through a MNFD can be modeled by the Smoluchowski equation as

$$\rho_t = \nabla \cdot (D \nabla \rho - \mu \vec{E} \rho), \quad (1)$$

¹Note that the Smoluchowski equation is also variously known as the Kolmogorov forward equation, the Fokker-Planck equation, or the convection-diffusion equation, with certain names more common in certain areas of application.

*Hendrick.deHaan@uoit.ca

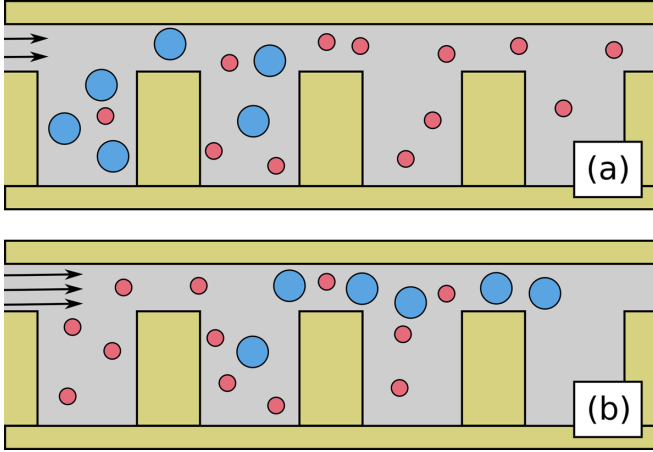


FIG. 1. A schematic of electrophoretic sorting of particles by size in the slit-well device. (a) A weak field causes small particles (red) to traverse the device more quickly on average. (b) A strong field causes large particles (blue) to traverse the device more quickly on average.

where ρ is the position distribution of the particles over space and time, D and μ are the diffusion and free-solution mobility coefficients of the particles, and $\vec{E}$ is the applied electric field. In first-passage problems where the domain geometry and applied fields are time-invariant, Eq. (1) can be integrated over time to obtain the time-integrated Smoluchowski equation [19]

$$-\rho_0 = \nabla \cdot (D\nabla g_0 - \mu\vec{E}g_0), \quad (2)$$

where ρ_0 is the initial condition for ρ . The new field g_0 is defined as

$$g_0(x, y) := \int_0^\infty \rho(x, y, t) dt. \quad (3)$$

The integral of g_0 in any region is the average residence time of particles in that region between initialization and absorption. In particular, it therefore has the property that

$$\int_\Omega g_0 dx = \langle \tau \rangle, \quad (4)$$

when Ω is the entire spatial domain, τ is the stochastic first passage time of the particles to the absorbing boundary conditions, and $\langle \tau \rangle$ is the mean first passage time (MFPT). Moreover, this formulation can be extended recursively to all higher-order moments as well. For instance, the field g_1 satisfying

$$-g_0 = \nabla \cdot (D\nabla g_1 - \mu\vec{E}g_1) \quad (5)$$

has the property that it integrates over the spatial domain to yield the second moment of the first passage time. A more comprehensive discussion of these moment equations can be found in standard references such as Redner [19].

Since the first few moments of first passage time distributions are so important to MNFD phenomena, it is natural to wonder whether solving the moment equations directly might be a useful line of investigation. In practice, however, it appears that this is rarely done. Redner [19] shows the power

of the moment equations for theoretical analysis of first passage problems, especially in the purely diffusive regime where direct analogies with electrostatics can be made. In the context of MNFDs, Magill *et al.* [18] showed that measuring g_0 approximately via particle simulations can aid in understanding the effect of design parameters on system dynamics. Similarly, Wang *et al.* [8] analyzed plots of the time-integrated particle position densities in a periodic model of the slit-well device; however, these maps were constructed in a manner subtly different from g_0 and in particular do not have the property of integrating to the MFPT. The authors are unaware of other studies in which the moment equations are solved numerically towards the goal of understanding the effect of MNFD design parameters on first passage time behavior. Moreover, even though the Smoluchowski equation is also an important mathematical model to study first passage time problems outside biophysics [20–23], we have found no examples in which the g_0 equation (nor any of the higher moment equations) were studied numerically in applied contexts.

A major barrier to the goal of solving the moments equations numerically in biophysics is the so-called curse of dimensionality. That is, for most common numerical methods for partial differential equations (PDEs), the computational cost grows exponentially in the dimensionality of the underlying domain. Thus, whereas highly effective techniques like the finite element method (FEM) can be used to solve PDEs in simple biophysical scenarios, like that of noninteracting nanoparticles, they fail when applied to the high-dimensional PDEs describing the dynamics of many-body systems such as polymers. Indeed, particle-based simulation methods do not exhibit the curse of dimensionality, and this can be seen as a major reason for the dominance of particle simulations over PDE-based calculations in biophysics.

In this work, we investigate a numerical method for PDEs that does not suffer from the curse of dimensionality. The technique, which we refer to as the neural network method (NNM), is inspired by the success of deep learning at solving high-dimensional problems in machine learning, such as image processing and natural language processing [24–26]. A growing body of theoretical and numerical evidence suggests that it can robustly solve high-dimensional PDEs [27–41]. In particular, the NNM has already been used to study high-dimensional problems in biophysics [38].

The NNM has also been shown to solve parametrized problems directly across a continuous range of parameter values [28,42]. As the number of parameters increases, the problem of solving a highly parametrized PDE can exhibit yet another curse of dimensionality. Because the neural network method shares information across parameter space, it is also able to overcome this computational challenge [43–45].

Note that parametrized solutions to PDEs typically cannot be obtained using the FEM, particle simulations, or similar methods. Rather, this goal is usually accomplished using reduced order modeling (ROM) techniques [46,47]. ROM methods typically interpolate between a relatively small number of high-fidelity solutions computed at a handful of reference points in parameter space in order to approximate solutions at new points in parameter space. Whereas most classical ROM methods interpolate to new parameter choices via a linear combination of the reference solutions, the NNM

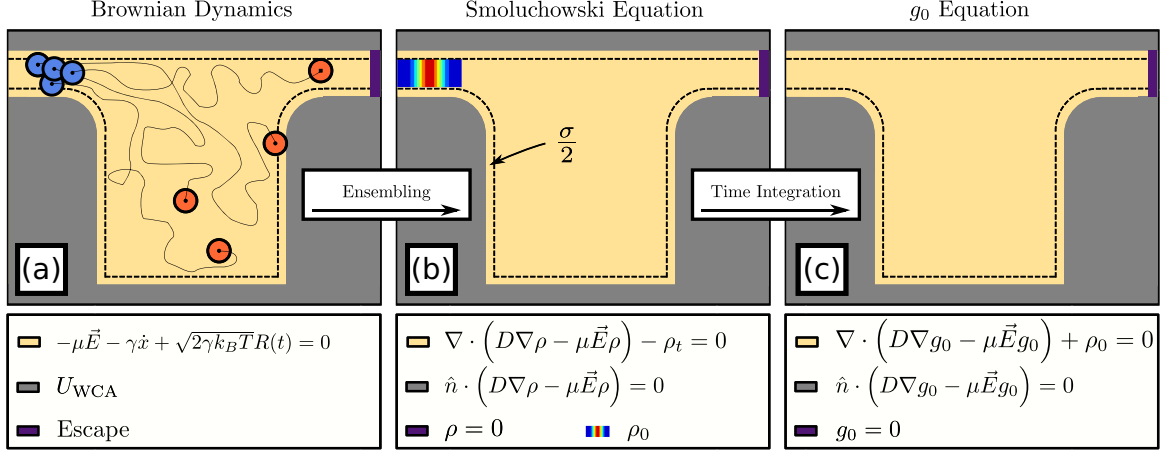


FIG. 2. Schematic of a single periodic subunit of the slit-well device to illustrate passage time models used in this study. (a) Particles are initialized in the left slit (blue particles), confined in the device via a WCA potential on gray walls, undergo Brownian dynamics in the yellow interior (trajectories denoted by black lines) until escaping device at the purple boundary. Dotted line denotes the region of the interior that cannot be occupied by the center of mass of particles. (b) A PDE model of the escape process where the solution ρ satisfies the Smoluchowski equation in the yellow interior. The initial Gaussian band source is located in the left slit, an absorbing boundary in the right slit (purple), and no flux conditions applied on gray walls that move inward to the dotted line to model particle size. (c) A PDE model of the escape process where the solution g_0 satisfies the g_0 equation (first moment) in the yellow interior region. An absorbing boundary is applied at the right slit wall (purple), and no flux conditions are applied on gray boundaries which move inward to the dotted line to model particle size..

is intrinsically nonlinear. Other nonlinear ROMs based on deep neural networks have been proposed in the literature [48–52]. However, these methods require that a database of FEM or other classical solutions be computed prior to training, whereas the NNM simultaneously solves the target PDE and acts as a ROM method over parameter space. In addition, dealing with parameters that modify the domain geometry using classical reduced-order methods can be challenging because these are typically constructed using mesh-based approaches. Although special ROMs can be developed for geometric parameters in some cases [53–55], the mesh-free nature of the NNM is intrinsically advantageous for this application [42,56].

II. PROBLEM DESCRIPTION

The primary goal of this paper is to study the effectiveness of the NNM as a tool for solving the g_0 equation in MNFDs by focusing on a sufficiently complicated representative device, as shown in Fig. 2. The specific problem under consideration is as follows: for an ensemble of noninteracting thermal particles initially located in the left slit of one periodic subunit of the slit-well device, compute the MFPT of these particles to the right slit. Here, the particles are driven by an electric field $\vec{E} = \lambda\vec{E}_0$ for a field strength constant λ and a baseline electric field $\vec{E}_0$. The baseline electric field is a solution to Laplace’s equation for a voltage drop of 2 V across the domain. It was obtained using the NNM in the manner described in Magill *et al.* [57], and plots of $\vec{E}_0$ are included here in Appendix B. The particles represent nanoparticles with diameters σ , diffusion coefficients D , and free-solution mobilities μ . The nanoparticles are assumed to be free-draining, with μ independent of σ , so that separation by size would not occur in free solution. Conversely, the diffusion coefficient is assumed to emerge from Stokes’ law and the fluctuation-dissipation

theorem, so that $D \propto \sigma^{-1}$. For simplicity, these behaviors are implemented as $\mu = 1$ and $D = \sigma^{-1}$. The g_0 equation is thus reduced to

$$F[g_0] \equiv \nabla \cdot \left(\frac{1}{\sigma} \nabla g_0 - \lambda \vec{E}_0 g_0 \right) + \rho_0 = 0, \quad (6)$$

where the particle size σ and field strength λ are the two free parameters, and $\vec{E}_0$ is the reference electric field.

The problem geometry is shown in Fig. 2 along with depictions of the particle-based, Smoluchowski, and g_0 representations of the problem. The domain is meant to represent a single periodic subunit of the slit-well device illustrated in Fig. 1. In the particle-based model of the problem [Fig. 2(a)], an ensemble of noninteracting nanoparticles are initially located in the left slit, and then these particles proceed to move under a combination of thermal diffusion and electrophoretic drift until reaching the far right purple wall in the right slit. In the Smoluchowski model [Fig. 2(b)], individual particles are eschewed, and the time evolution of the entire distribution of particle positions is modeled instead. Here, the initial position of particles is modeled by the initial condition $\rho_0(x; \sigma)$, located in the left slit. Finally, in the g_0 equation [Fig. 2(c)], the time-dependence of the Smoluchowski is accounted for implicitly by integration over all time. Here, the initial condition $\rho_0(x; \sigma)$ now appears as a source term in the (time-independent) PDE.

In each schematic of Fig. 2, the gray regions represent physical walls. These were modeled as short-range repulsive boundaries (in the particle model; see Appendix A) or no-flux boundary conditions in the continuum models [equations defined in the legends of Figs. 2(b) and 2(c)]. As a result of excluded volume interactions, the particle centers cannot come closer than a distance of roughly $\sigma/2$ from the repulsive boundaries. This exclusion zone is depicted by the dashed black line in Fig. 2. To model this in the continuum models,

the no-flux boundary conditions are applied at the boundary of the exclusion zone (i.e., along the dotted black lines in Fig. 2), rather than at the nominal boundaries (i.e., along the gray walls in Fig. 2).

The nominal dimensions of the domain Ω_0 are the same as those described in Magill *et al.* [57]. In particular, the topmost and bottommost walls are a distance $2L_y = 6.25$ apart, the leftmost and rightmost boundaries are $2L_x = 10$ apart, and the curvature of the re-entrant corners is set to $R = 1$ (see below). The total horizontal lengths of the slits and the well were set equal, to L_x , and the slits were given a height of $h_{\text{slit}} = L_x/4 = 1.25$.

As discussed in Magill *et al.* [57], the standard formulation of the NNM struggles to solve problems exhibiting singularities. For this reason, the re-entrant corners of the slit-well device geometry have been rounded (i.e., represented by circular arcs of finite curvature). Similarly, the NNM was found to perform poorly when the initial distribution of particles was too sharp. Instead, particles were initialized in a Gaussian band in the left slit, given by uniform distribution in y multiplied by a Gaussian distribution in x :

$$\rho_0 = \frac{1}{\sqrt{2\pi}r_s h_{\text{slit}}} \exp\left(-\frac{(x - x_s)^2}{2r_s^2}\right). \quad (7)$$

Here $r_s = 0.25$ is the width of the Gaussian band in the x direction, $h_{\text{slit}} = L_y - y_{\text{slit}} - \sigma$ is the height of the band in the y direction, and $x_s = -L_x + 1$ is the center of the band. Technically, ρ_0 requires a correction factor to be properly normalized over this bounded domain, as the Gaussian distribution in x is normalized over the entire real line, but the discrepancy is numerically insignificant.

The first passage time of the particles is computed when their centers cross the rightmost boundary of the domain for the first time (purple in Fig. 2). In the continuum models, this is represented by an absorbing boundary condition (i.e., a homogeneous Dirichlet condition). Physically, this boundary corresponds to the interface between consecutive periodic subunits of the slit-well, and not to a physical wall. As such, in contrast to the no-flux boundary condition on the gray walls, the placement of the absorbing boundary does not depend on σ .

As a simplifying assumption, particles were prevented from moving through the leftmost boundary of the domain. Mathematically, this was imposed by a no-flux boundary condition. Physically, this corresponds to the synthetic condition that particles cannot move against the direction of the imposed electric field into the previous periodic subunit of the slit-well; we will refer to this as the no-backflow condition. The location of this no-backflow boundary condition was fixed independently of σ .

Of course, in the actual slit-well device there is always a nonzero probability of particle backflow. The simplification was made here because it allows the g_0 equation to be posed in a much simpler domain (i.e., a single periodic subunit). However, as a result of this modeling choice, there will be discrepancies between the MFPT results reported in this paper and the results of previous studies of the slit-well device (such as Cheng *et al.* [7] and Wang *et al.* [8]), especially at low electric field strengths. Nevertheless, as the results in Sec. IV

will indicate, the major features of the slit-well system are preserved despite the no-backflow condition. Furthermore, the simplified model still contains several mathematical features that are expected to be common to many MNFDs and particularly difficult for the NNM to resolve: re-entrant corners, a nonuniform electric field, and nontrivial dependence on physical and geometric problem parameters. As stated above, the purpose of this paper is to study the performance of the NNM when solving a problem with the characteristic features of a typical MNFD problem. Certain features, such as the highly skewed geometry of the fully periodic slit-well and the singularities associated with the fully sharp re-entrant corners, are more technically challenging and relegated to future work.

III. METHODOLOGY

A. Neural network method

The NNM implementation used for this work was similar to that previously described by Magill *et al.* [57]. In the fixed parameter experiments (Sec. IV B 1), the true solution $g_0(\mathbf{x})$ of the g_0 equation [Eq. (6)] was approximated by a deep neural network $\tilde{g}_0(\mathbf{x})$ trained to minimize a composite loss functional

$$\mathcal{L} = \mathcal{L}_{\text{PDE}} + \mathcal{L}_{\text{BC}} + \mathcal{L}_{\text{norm}}. \quad (8)$$

The first loss term consisted of

$$\mathcal{L}_{\text{PDE}}[\tilde{g}_0] = \int_{\Omega} (F[\tilde{g}_0])^2 dA, \quad (9)$$

where F is the operator in the g_0 equation [Eq. (6)]. Thus, $\mathcal{L}_{\text{PDE}}[\tilde{g}_0]$ quantifies the extent to which $\tilde{g}_0$ satisfied the g_0 equation [Eq. (6)] throughout the domain Ω . Note that, as discussed in Sec. II, Ω depends on σ . The second loss term was defined as

$$\mathcal{L}_{\text{BC}}[\tilde{g}_0] = \int_{\partial\Omega} (B[\tilde{g}_0])^2 ds, \quad (10)$$

where $B[\tilde{g}_0]$ defines no-flux or absorbing boundary conditions (BCs), as appropriate, on each part of the boundary of the domain [see Fig. 2(c)]. Thus, $\mathcal{L}_{\text{BC}}[\tilde{g}_0]$ quantifies the extent to which $\tilde{g}_0$ satisfied the BCs over the domain boundary $\partial\Omega$.

The final term was given by

$$\mathcal{L}_{\text{norm}}[\tilde{g}_0] = \left[\int_{\Omega} (F[\tilde{g}_0]) dA \right]^2. \quad (11)$$

Similarly to $\mathcal{L}_{\text{PDE}}$, the last loss term $\mathcal{L}_{\text{norm}}$ quantifies the extent to which the approximate solution satisfies the PDE inside the domain Ω . However, whereas $\mathcal{L}_{\text{PDE}}$ is a *local* measure of the residual of Eq. (6), $\mathcal{L}_{\text{norm}}$ is a *global* measure. Specifically, $\mathcal{L}_{\text{PDE}}$ is the mean of the squared residual, while $\mathcal{L}_{\text{norm}}$ is the square of the mean residual. In theory, $\mathcal{L}_{\text{norm}}$ is a redundant loss term and simply setting $\mathcal{L}_{\text{PDE}}$ to zero is sufficient to ensure that $\tilde{g}_0$ satisfies the g_0 equation [Eq. (6)]. In practice, however, training without $\mathcal{L}_{\text{norm}}$ was found to produce approximate solutions that captured the shape of the true solution fairly accurately, but struggled to converge on the correct magnitude (i.e., differed from the true solution by a small multiplicative factor).

We note that our use of $\mathcal{L}_{\text{norm}}$ is similar to the normalization process used by Al-Arabi *et al.* [58] in solving the time-dependent Smoluchowski equation. The use of $\mathcal{L}_{\text{norm}}$

can also be contrasted with previous work by Avrutskiy [59]. There, Avrutskiy [59] showed benefits to adding redundant loss terms that encourage the solution to satisfy the derivative of the PDE operator $F' = 0$ over the spatial domain. Here, the term $\mathcal{L}_{\text{norm}}$ is a redundant loss term that encourages $\tilde{g}_0$ to agree with the *integral* of the PDE operator F over the spatial domain Ω . These additional loss terms can be thought of as soft constraints on the training process, or equivalently as regularization terms constructed out of prior knowledge of the target problem.

In Secs. IV B 2–IV B 4, the neural network was parametrized with respect to field strength λ , particle size σ , or both. Thus, the loss terms were redefined as

$$\mathcal{L}_{\text{PDE}}[\tilde{g}_0] = \left\langle \int_{\Omega_\sigma} (F_{\lambda,\sigma}[\tilde{g}_0])^2 dA \right\rangle_{\lambda,\sigma}, \quad (12)$$

$$\mathcal{L}_{\text{BC}}[\tilde{g}_0] = \left\langle \int_{\partial\Omega_\sigma} (B_{\lambda,\sigma}[\tilde{g}_0])^2 ds \right\rangle_{\lambda,\sigma}, \quad (13)$$

$$\mathcal{L}_{\text{norm}}[\tilde{g}_0] = \left\langle \left[\int_{\Omega_\sigma} (F_{\lambda,\sigma}[\tilde{g}_0]) dA \right]^2 \right\rangle_{\lambda,\sigma}. \quad (14)$$

The notations Ω_σ , $F_{\lambda,\sigma}$, and $B_{\lambda,\sigma}$ indicate that the domain changes with σ , and the PDE and BC operators change with both λ and σ . The angled brackets indicate averages over the parameter values. In other words, the loss used for the parametrized neural networks is identical to that used for the fixed parameter experiments, with the additional step of averaging the loss over parameter space.

Note that the electric field $\vec{E}$ was obtained by computing the electric potential u using the NNM methodology of Magill *et al.* [57]. Of course, this is not strictly necessary because u could just as easily be approximated by some other method (e.g., FEM). However, the intention was to illustrate the ease with which previously computed NNM solutions can be fed into the loss functional of new NNM solutions. A contour plot illustrating both u and $\vec{E}$ is included in Appendix B (Fig. 7). Note that the electric potential is defined on the nominal domain Ω_0 corresponding to $\sigma = 0$, which differs from the actual domain Ω on which g_0 is defined.

All of the NNM experiments in this work were conducted with fully connected feedforward neural networks of depth $d = 3$ and width $w = 50$. The hyperbolic tangent was used for activation functions in the hidden layers, while the output layer was linear. To solve the nonparametrized problems (Sec. IV B 1), the approximate solution $\tilde{g}_0$ was constructed as

$$\tilde{g}_0(\mathbf{x}) = f_{d+1} \circ f_d \circ \cdots \circ f_1(\mathbf{x}), \quad (15)$$

with

$$f_1(\mathbf{x}) = \tanh(W_1 \mathbf{x} + \mathbf{b}_1), \quad (16)$$

$$f_i(\mathbf{x}) = \tanh(W_i f_{i-1}(\mathbf{x}) + \mathbf{b}_i), \quad i = 2, \dots, d, \quad (17)$$

$$f_{d+1}(\mathbf{x}) = W_{d+1} f_d(\mathbf{x}) + b_{d+1}, \quad (18)$$

where $W_1 \in \mathbb{R}^{w \times 2}$, $W_i \in \mathbb{R}^{w \times w}$ for $i = 2, \dots, d$, and $W_{d+1} \in \mathbb{R}^{1 \times w}$ are the network's weight matrices, while $\mathbf{b}_i \in \mathbb{R}^w$ for $i = 1, \dots, d$, and $b_{d+1} \in \mathbb{R}$ are its biases.

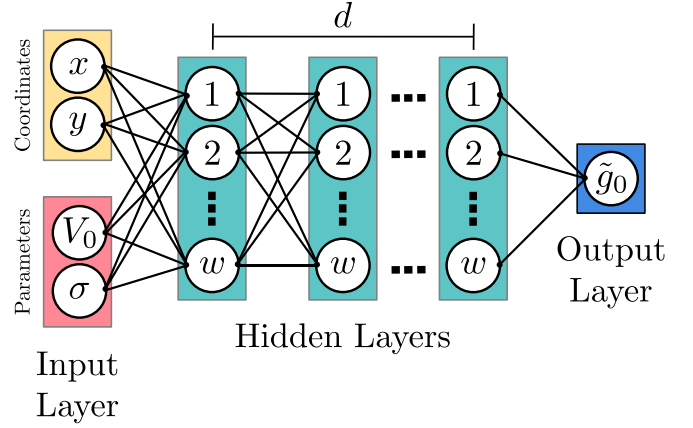


FIG. 3. Fully connected feedforward neural network of width w and depth d mapping coordinates (x, y) and problem parameters (λ, σ) to an output $\tilde{g}_0(x, y; \lambda, \sigma)$. At each node, a weighted sum of the incoming arrows and a bias is computed and passed through an activation function. The network's parameters are optimized such that $\tilde{g}_0(x, y; \lambda, \sigma)$ approximately satisfies the target PDE and BCs.

The experiments in Secs. IV B 2–IV B 4 considered parametrized neural networks, where one or both of the problem parameters λ and σ were included as additional inputs to the network. In these cases, the networks were defined as

$$\tilde{g}_0(\mathbf{x}; \mathbf{m}) = f_{d+1} \circ f_d \circ \cdots \circ f_1(\mathbf{x}; \mathbf{m}), \quad (19)$$

with f_i defined as before for $i = 2, \dots, d + 1$, but with f_1 adjusted to

$$f_1(\mathbf{x}) = \tanh(W_1^{(x)} \mathbf{x} + W_1^{(m)} \mathbf{m} + \mathbf{b}_1), \quad (20)$$

where $W_1^{(x)} \in \mathbb{R}^{w \times 2}$ and $W_1^{(m)} \in \mathbb{R}^{w \times m}$, where m is the length of the parameter vector $\mathbf{m}$. In other words, the parameters (λ or σ or both) were concatenated to the end of the input vector of the network, and the weight matrices were adjusted accordingly. This is illustrated schematically in Fig. 3. The same approach was used by Sirignano and Spiliopoulos [28] and Hennigh *et al.* [42] but can be contrasted with the recently proposed DeepONet architecture of Lu *et al.* [60].

Training was conducted in TENSORFLOW [61] version 1.15 with all unspecified hyperparameters set to their default values. The weights were initialized using the Glorot method [62], and biases were initialized to zero. Weights were iteratively updated using the Adam optimizer [63] to minimize $\mathcal{L}$ with the learning rate set to 10^{-3} in Sec. IV B 1, and to 10^{-4} in Secs. IV B 2–IV B 4. In each iteration, the integrals in $\mathcal{L}$ were approximated by Monte Carlo sampling using the same procedure described in Magill *et al.* [57]. Specifically, rejection sampling was applied to 10 000 nominal samples generated in the bounding box $[-5, 5] \times [-3.125, 3.125]$, and each smooth subunit of the boundary was randomly sampled with a linear density of about 13 points per unit length. For the parametrized network experiments (Secs. IV B 2–IV B 4), the relevant problem parameters were also sampled randomly in each training iteration. These samples were generated uniformly at random, with λ drawn from [5, 50] and σ drawn from [0.125, 0.625]. In particular, it was necessary to sample the parameter σ before sampling points in Ω , since the extent

of Ω varies with σ . One random parameter vector was drawn per training iteration.

The testing loss was evaluated every 1000 training iterations, using ten times more samples than during a training step. In the parametrized network experiments (Secs. IV B 2–IV B 4), the testing loss was averaged across 100 random parameter vectors. Training was continued for a fixed number of iterations (600 000 epochs for the fixed parameter experiments, and 30 000 000 epochs for the parametrized network experiments). The final network was taken as that which achieved the lowest testing loss across all iterations.

B. Finite element method

The MFPT problem in the slit-well domain cannot be solved exactly in closed form due to the complex nature of the geometry. Instead, approximate ground truth solutions to the problem were obtained using the finite element method (FEM). Following Magill *et al.* [57], the problem for g_0 was solved using a mixed FEM formulation implemented in FEniCS [64]. The electric field $\vec{E}$ included in the PDE [Eq. (6)] was obtained by also approximating the electric potential u by a mixed FEM formulation. As stated above, u is defined on the nominal domain Ω_0 , whereas g_0 is defined on a smaller domain depending on σ . Thus, for the FEM solutions it was necessary to first solve u and $\vec{E}$ on a discretization of Ω_0 , project $\vec{E}$ onto a discretization of the appropriate Ω , and then define the variational problem for g_0 on Ω .

The mesh decomposition of the domain was conducted using the MSHR package in FENICS. The resolution parameter was set to 200, and the re-entrant corners were approximated linearly by 400 segments each. The same mesh parameters were used for all values of σ , and for the nominal domain, Ω_0 , on which u was solved.

IV. RESULTS

This section details results obtained using the NNM to solve the g_0 equation modeling the MFPT of nanoparticles driven through the slit-well device (described in Sec. II). The focus throughout is on the relationship between key problem parameters and observables of physical interest, where g_0 acts as a proxy between the two. The first observable of interest is, naturally, the mean first passage time $\langle\tau\rangle$. As described in Eq. (4), $\langle\tau\rangle$ can be obtained by integrating g_0 over the domain Ω . Throughout this paper, the integration of g_0 to estimate $\langle\tau\rangle$ is accomplished using the same Monte Carlo procedure described for $\mathcal{L}_{\text{PDE}}$ in Sec. III.

In practice, an observable of greater interest than the mean first passage time itself is the net electrophoretic mobility of the nanoparticles through the slit-well device over long timescales [7,8]. In particular, the electrophoretic mobility is typically defined as

$$\mu_{\text{electro}} := \lim_{t \rightarrow \infty} \frac{\langle x \rangle_t}{E_c t}, \quad (21)$$

where $\langle x \rangle_t$ is the ensemble average of the x position at time t , and E_c is a characteristic scale for the applied electric-field strength. It is not clear whether μ_{electro} can be inferred directly from the g_0 problem being solved here. Instead, the present

paper will investigate a similar observable of interest, which will be called the effective mobility

$$\mu_{\text{eff}} := \frac{L_0 / \langle\tau\rangle}{E_c} = \frac{1}{\lambda \langle\tau\rangle}, \quad (22)$$

where L_0 is the mean horizontal distance from ρ_0 to the absorbing wall. The characteristic field strength is chosen of the form $E_c = V_c / L_c$, where V_c is a characteristic voltage drop and L_c is a characteristic length scale. Since the overall voltage drop across the system is of order one and proportional to the field strength λ , we choose $V_c = \lambda$. For numerical simplicity, we also choose $L_c = L_0$, thus obtaining the final equality in Eq. (22). The effective mobility is expected to exhibit similar features to the electrophoretic mobility because both consist of characteristic particle velocities divided by characteristic electric-field strengths. A comprehensive exploration of the relationship between the two mobility definitions is left to future work.

A. Characteristics of g_0

Figure 4 shows contour plots of g_0 solutions computed using the NNM, with the corresponding estimates of $\langle\tau\rangle$ and μ_{eff} shown in the legends. The four subplots correspond to the four essential parameter regimes alluded to in Fig. 1. Note that the magnitude of the color scale varies across the four subplots.

First, consider the solution of g_0 in Fig. 4(a) corresponding to small particles ($\sigma = 0.125$) driven by a weak field ($\lambda = 5.0$). Here, g_0 has a maximum in the left slit near the peak of the initial particle distribution ρ_0 . Naturally, since the particle positions are initialized according to ρ_0 , the average residence time in that region is relatively high; this feature is common to all four subplots in Fig. 4. Outside the left slit, g_0 decreases nearly monotonically from left to right, eventually reaching a value of zero on the absorbing boundary. The shape of this function is nearly visually indistinguishable from the solution with $\sigma = 0.125$ and $\lambda = 0$ (not shown) and is characteristic of predominantly diffusive dynamics in all regions of the domain.

Figure 4(b) again shows g_0 for small particles ($\sigma = 0.125$), but now driven by a much stronger field ($\lambda = 50.0$). In contrast with the monotonically decreasing solution in Fig. 4(a), in this scenario g_0 is relatively constant throughout most of the domain until a boundary layer near the absorber. In fact, here g_0 even exhibits some minor nonmonotonic features: a shadow is evident in the bottom-left of the well, and a local maximum is attained at the entrance to the right slit. Drift and diffusion effects are relatively balanced in this case, with the uniformity in x reflecting strongly driven motion in the horizontal direction, and the uniformity in y reflecting rapid diffusion in the vertical direction.

Figure 4(c) shows g_0 for large particles ($\sigma = 0.625$) driven by a weak field ($\lambda = 5.0$). Notice that the walls of the domain are shifted inward by 0.5σ , reflecting the reduced area that can be occupied by the center of mass of larger particles (Sec. II). In this scenario, the smaller diffusion coefficient of the larger particles balances the weaker field, resulting in a solution that more closely resembles that in Fig. 4(b) than that in Fig. 4(a). However, the solution in Fig. 4(c) is visibly

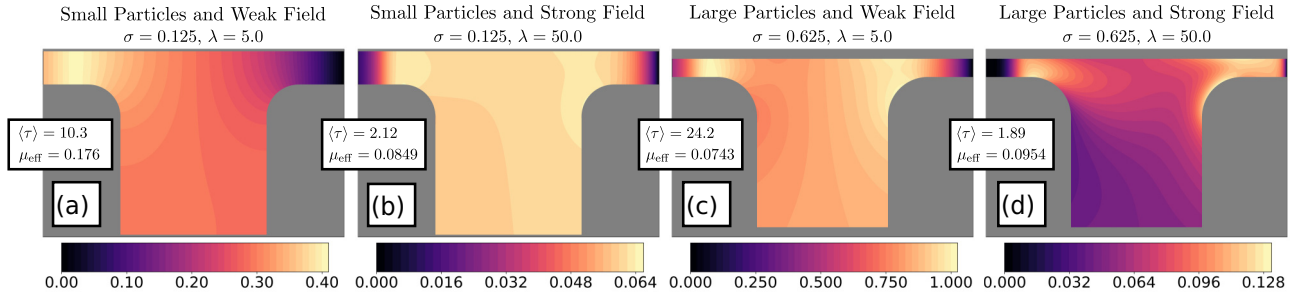


FIG. 4. NNM solutions to the g_0 equation subject to (a) a small particle size and a weak electric field, (b) a small particle size and a strong electric field, (c) a large particle size and a weak electric field, and (d) a large particle size and a strong electric field.

increasing from left to right across the well, in contrast with both the solutions in Figs. 4(a) and 4(b). As was the case in Fig. 4(b), drift and diffusion are of comparable importance; the differences between the two solutions are primarily due to the modifications to the domain geometry.

Finally, Fig. 4(d) shows g_0 computed for large particles ($\sigma = 0.625$) subject to a strong electric field ($\lambda = 50.0$). Here, the shape of the solution differs significantly from those in all of Figs. 4(a)–4(c). In Fig. 4(d), g_0 takes on very small values throughout the entire well, and decreases substantially from the top of the well to its bottom. The combination of the low diffusion coefficient and the very strong driving force causes the large particles to remain primarily streamlined in the upper region of the well as they move rapidly from ρ_0 to the absorber.

The MFPTs $\langle \tau \rangle$ and effective mobilities μ_{eff} in the four scenarios of Fig. 4 are consistent with the expected sorting mechanisms in each regime [7]. When the field is strong, smaller particles have a larger $\langle \tau \rangle$ and lower μ_{eff} than larger particles. The converse is true at weak fields.

Future work should explore the relationship of g_0 , $\langle \tau \rangle$, and μ_{eff} with standard explanations for these phenomena, such as the entrance effect [7,65]. The purpose of the discussion in this section was to illustrate the variety of complicated behavior that arise in g_0 solutions across the different physically meaningful parameter regimes in the slit-well. In Sec. IV B, parametrized NNM solutions will be trained to interpolate nonlinearly between all four solutions in Fig. 4. Ultimately, in Sec. IV B 4 this will yield continuously differentiable mappings between both problem parameters λ and σ and both key physical observables $\langle \tau \rangle$ and μ_{eff} , thereby capturing the entirety of this rich sorting mechanism in a single numerical solution.

B. Benchmarking the neural network method against the finite element method

In this section, g_0 will be leveraged as a proxy for the calculation of the metrics $\langle \tau \rangle$ and μ_{eff} . In practice, it is common in MNFD research and development (and scientific research more broadly) to study how such key metrics change in response to variations in the system parameters. The simplest approach to characterizing this variation is to compute or measure the metrics independently for a large number of parameter choices. In Sec. IV B 1, the NNM is applied to

precisely this task of calculating $\langle \tau \rangle$ and μ_{eff} for many combinations of particle size σ and field strength λ .

The above approach, however, requires repeated calculation of the key metrics which can be expensive when considering many independent parameters. As discussed in Sec. I, the NNM can be leveraged to solve such parametrized problems directly across continuous ranges of parameter values. The high-dimensional function $g_0(x, y; \lambda, \sigma)$ implicitly encodes $\langle \tau \rangle$ and μ_{eff} as continuously differentiable functions of σ and λ . The NNM is used to approximate this function directly in Secs. IV B 2–IV B 4, for g_0 solutions parametrized directly by λ , σ , or both simultaneously.

Throughout Sec. IV B, four quantities are used to characterize the performance of the NNM across parameter space. These quantities are all plotted in Fig. 5, with each column corresponding to one of the four NNM formulations discussed above. Naturally, both the MFPT $\langle \tau \rangle$ and the effective mobility μ_{eff} are included in the analysis. These are plotted in the first two rows [Figs. 5(a)–5(d) and 5(e)–5(h)], respectively, alongside the reference values computed using FEM. The NNM results are indicated by lines, and the corresponding FEM results are included as stars. Dotted lines in Figs. 5(a) and 5(e) connect values that are only computed at discrete parameter choices, whereas solid lines used everywhere else indicate values that are computed over continuous parameter ranges.

Next, in order to quantify the accuracy of the $\langle \tau \rangle$ values obtained via the NNM, the relative error ε with respect to the ground truth FEM solution is computed. Specifically, ε is defined as the relative error of $\langle \tau \rangle$ with respect to $\langle \tau \rangle_{\text{FEM}}$, i.e.,

$$\varepsilon = \frac{|\langle \tau \rangle - \langle \tau \rangle_{\text{FEM}}|}{\langle \tau \rangle_{\text{FEM}}}, \quad (23)$$

where $\langle \tau \rangle$ and $\langle \tau \rangle_{\text{FEM}}$ are the MFPTs computed by the NNM and FEM, respectively. The relative errors ε are plotted in Figs. 5(i)–5(l). Here, circular markers indicate the discrete parameter choices at which ε was computed. Additionally, the plots in Figs. 5(i)–5(l) contain a dotted black line at 10^{-2} , corresponding to a relative error of 1%. This is representative of a relative error threshold that is typically attainable and acceptable in MNFD research. Indeed, Appendix A describes standard particle simulations that were used to approximate $\langle \tau \rangle$ with relative errors comparable to or below 1% for all choices of parameters λ and σ .

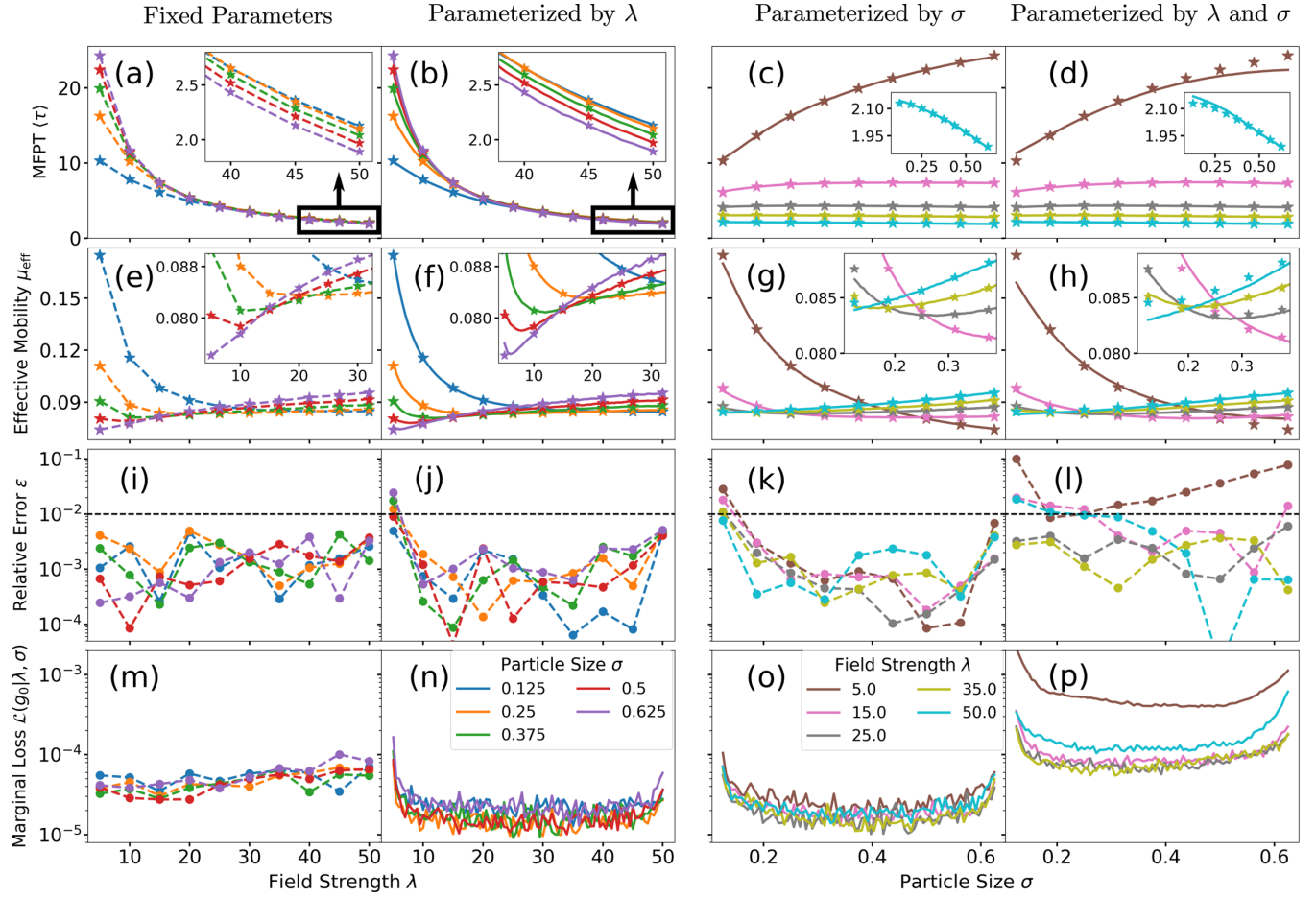


FIG. 5. Analysis of g_0 solutions computed using the NNM. Mean first passage times $\langle \tau \rangle$ for the NNM (a) with fixed parameters, and parametrized by (b) λ , (c) σ , and (d) both λ and σ . Star markers denote values obtained using the FEM, and insets display behavior at high field strengths. Effective mobilities μ_{eff} for the NNM (e) with fixed parameters, and parametrized by (f) λ , (g) σ , and (h) both λ and σ . Star markers denote values obtained using FEM, and insets zoom in on the minimum of the curves. Relative errors ε computed against the FEM for the NNM (i) with fixed parameters, and parametrized by (j) λ , (k) σ , and (l) both λ and σ . Dotted black line denotes 1% error baseline computed by particle simulations. (m) Testing loss of the NNM with fixed parameters, and marginal loss of the NNM parametrized by (n) λ , (o) σ , and (p) both λ and σ .

The final quantity included in the analysis is similar to the loss functional $\mathcal{L}$ used during the NNM training process [Eq. (9)]. However, the total loss provides only a single characterization of a network's performance over its entire domain. When the NNM was used to solve parametrized g_0 problems, it was valuable to evaluate the relative performance of these solutions at different points in parameter space. To this end, we defined the marginal loss $\mathcal{L}(g_0|\lambda, \sigma)$, a parameter-dependent generalization of the total loss. As with the true loss, $\mathcal{L}(g_0|\lambda, \sigma)$ is the sum of the L_2 norms of the PDE-, BC-, and norm-based residuals. However, whereas the total loss averages these quantities over all choices of λ and/or σ [Eqs. (12)–(14)], the corresponding terms in $\mathcal{L}(g_0|\lambda, \sigma)$ were instead treated as functions of λ and σ . In the case of nonparametrized NNM solutions, the marginal loss definition simply reduces to the original total loss.

The marginal losses are plotted in Figs. 5(m)–5(p). The values in Fig. 5(m) correspond to NNM solutions trained at fixed parameter choices and are thus indicated by discrete circular markers. Conversely, since $\mathcal{L}(g_0|\lambda, \sigma)$ can be evaluated continuously for parametrized solutions, the corresponding

marginal loss values shown in Figs. 5(n)–5(p) are indicated by solid lines sampled finely throughout the parameter space.

1. Neural network method with fixed parameters

This section contains a discussion of the results in Figs. 5(a), 5(e), 5(i), and 5(m). Here, the NNM was applied repeatedly to solving the g_0 equation for fixed choices of the problem parameters: field strength λ and particle size σ . This NNM formulation is the same as the one used in Sec. IV A but is now applied to many more choices of the problem parameters. Specifically, the results are shown for ten choices of λ uniformly spaced from 5 to 50 and five choices of σ uniformly spaced from 0.125 to 0.625, with a distinct neural network used to approximate g_0 for each parameter combination.

For small values of λ in Fig. 5(a), $\langle \tau \rangle$ is monotonically increasing with σ , and for large values of λ (see the inset) the opposite is true. Moreover, the finer sampling of parameter space resolves new features that were not clear from examining only the four samples in Sec. IV A. For instance, Fig. 5(a)

shows that $\langle \tau \rangle$ decreases monotonically with λ for each choice of σ . In addition, the dependence of $\langle \tau \rangle$ on σ is much stronger at low field strengths.

The same sorting behavior can be viewed from a different perspective via μ_{eff} in Fig. 5(e). In addition, the crossover in sorting order around $\lambda \approx 25$ is better resolved by μ_{eff} than $\langle \tau \rangle$. Indeed, in Fig. 5(e) it is clear that there is no single value of λ for which $\langle \tau \rangle$ and μ_{eff} are entirely independent of σ . Of course, the results discussed above for $\langle \tau \rangle$ and μ_{eff} are not novel because they are consistent with published results on the slit-well device (see, e.g., Cheng *et al.* [7]). Rather, the purpose of this discussion is to illustrate two points: first, that valuable information can be extracted by studying the variation of key output metrics (here, $\langle \tau \rangle$ and μ_{eff}) as functions of the key input parameters (here, λ and σ) and, second, that the physical problem being studied in this paper (Sec. II) indeed captures essentially the same physical mechanisms expected for the actual slit-well system.

Before considering the benefits of the more ambitious parametrized NNM formulations, it is important to assess how accurately the NNM resolves $\langle \tau \rangle$ and μ_{eff} when applied to the simpler task of solving g_0 at a single point in parameter space. The accuracy is quantified in Fig. 5(i), which shows ε , the relative error in $\langle \tau \rangle$. In this plot, it appears that ε is roughly independent of both σ and λ , suggesting that the current implementation of the NNM is fairly robust throughout the problem parameter space. This is corroborated by the testing losses $\mathcal{L}(g_0|\lambda, \sigma)$ plotted in Fig. 5(m), which are also roughly independent of the problem parameters. Most importantly, for all choices of parameters λ and σ in Fig. 5(i), ε is well below the 1% error threshold indicated by the black line. In other words, the NNM is at least as effective at resolving $\langle \tau \rangle$ as the Brownian dynamics particle simulations included in Appendix A.

2. Neural network method parameterized by field strength

Whereas in Sec. IV B 1, 50 networks were used to obtain 50 different g_0 solutions, which were then integrated over their respective domains to produce 50 different $\langle \tau \rangle$ measurements, in this section only five networks are utilized to accomplish the same goal. Each of these five networks solves $g_0(x, y; \lambda)$ for $\lambda \in [5, 50]$ at a fixed choice of σ . As in Sec. IV B 1, the metrics $\langle \tau \rangle$, μ_{eff} , ε , and $\mathcal{L}(g_0|\lambda, \sigma)$ are computed from the solutions; these are plotted in Figs. 5(b), 5(f), 5(j), and 5(n), respectively. Comparing Figs. 5(b) with 5(a) and Fig. 5(f) with 5(e), it is clear that the NNM formulation parametrized by λ recovers the same results previously obtained by solving g_0 independently for many different parameter choices in Sec. IV B 1.

One advantage of the parametrized NNM formulation is evident in the inset of Fig. 5(f). For each choice of σ , there is a λ value for which μ_{eff} is minimal. When computing μ_{eff} only at discrete choices of the parameters [as in Fig. 5(e)], the exact location of these minima is not clear. Instead, the results in Fig. 5(f) illustrate that the parametrized NNM formulation naturally resolves the existence of local minima, since the solution is trained continuously for all parameter values in the training domain. The benefit of continuous mappings from problem parameters to key output metrics becomes more valu-

able as dimensionality of parameter space is increased (e.g., as explored in Sec. IV B 4).

The relative error ε and marginal loss $\mathcal{L}(g_0|\lambda, \sigma)$ in Figs. 5(j) and 5(n) quantify the accuracy of the $g_0(x, y; \lambda)$ solution. Here, both ε and $\mathcal{L}(g_0|\lambda, \sigma)$ are highest at the boundaries of the λ training range and fairly uniform throughout the majority of the interior of the training range. In particular, both are highest at the left boundary, $\lambda = 5$. This relationship between error and loss is similar to those studied in Magill *et al.* [57] and provide further justification for using the (marginal) loss as an *a posteriori* method for gauging the reliability of NNM solutions.

The deterioration in performance seen in Figs. 5(j) and 5(n) at the boundaries of the λ training range can likely be attributed to the uniform Monte Carlo sampling of λ during training. The exact endpoints have very low probabilities of being sampled directly; moreover, their neighborhoods are only sampled on one side, whereas the neighborhoods of points nearer to the middle of the λ training range are sampled thoroughly on both sides. This could effectively lead to an under-representation of the behavior near the endpoints in the training loss. Characterizing this tentative mechanism is beyond the scope of the present work.

Overall, only three of the 50 relative errors in Fig. 5(j) slightly exceed the 1% error threshold. Thus, the implementation of the parametrized NNM studied in this section meets the standard of accuracy typically attained by Brownian dynamics (BD) simulations. In the regions of parameter space where the relative error was not measured directly, the marginal loss [Fig. 5(n)] provides an *a posteriori* estimate of the error, suggesting that the NNM's performance is excellent except for λ values very close to the boundaries of the training range. Altogether, these results demonstrate that the NNM is a feasible technique for solving the g_0 problem over a continuous range of field strengths. Moreover, using g_0 as a proxy for $\langle \tau \rangle$ and μ_{eff} enables the NNM to resolve the behavior of these key output metrics continuously over the target parameter range.

3. Neural network method parameterized by particle size

To expand upon the unique strengths of the NNM, this section will consider the problem of solving g_0 as a function of the particle size σ . Here, since the diffusion coefficient is being modeled as $D = \sigma^{-1}$, the terms of the g_0 PDE depend directly on the parameter σ , just as they depend directly on λ . However, the location of the boundaries of the slit-well domain also depend explicitly on the parameter σ (Sec. II). Thus, whereas λ only modified the PDE terms, σ modifies both the PDE terms and the domain geometry. As described in Sec. I, it is challenging for classical reduced-order methods to deal with parametrized domain geometries. However, this section will demonstrate that the NNM can handle geometry-modifying parameters (σ) just as easily as parameters that do not modify the domain geometry (λ).

Once again, the MFPT $\langle \tau \rangle$, effective mobility μ_{eff} , relative error ε , and marginal loss $\mathcal{L}(g_0|\lambda, \sigma)$ are computed from the NNM solutions and plotted in Figs. 5(c), 5(g), 5(k), and 5(o), respectively. Whereas the results in Figs. 5(a), 5(e), 5(i), and 5(m) and Figs. 5(b), 5(f), 5(j), and 5(n) for Secs. IV B 1–IV B 2 were solved and plotted as functions of λ , the results

for this section are presented as functions of σ . Specifically, each curve in Figs. 5(c), 5(g), 5(k), and 5(o) represents a single neural network trained over the range $\sigma \in [0.125, 0.625]$ at a fixed choice of λ [indicated by the legend in Fig. 5(o)].

The $\langle \tau \rangle$ and μ_{eff} measurements in Figs. 5(c) and 5(g) indicate that the NNM parametrized by σ recovers the same physical properties observed for the NNM with fixed parameters (Sec. IV B 1) and the NNM parametrized by λ (Sec. IV B 2). For the $\langle \tau \rangle$ measurements in Fig. 5(c), the $\lambda = 5.0$ curve is monotonically increasing whereas the $\lambda = 50.0$ curve (enhanced in the inset) is monotonically decreasing. This implies that small particles have a lower MFPT at low field strengths, and large particles have a lower MFPT at high field strengths. Likewise, the $\lambda = 5.0$ curve for μ_{eff} in Fig. 5(g) indicates that small particles are more mobile at low field strengths, whereas the $\lambda = 50.0$ curve indicates that large particles are more mobile at high field strengths.

Visual comparison of the NNM results (solid lines) to the ground-truth FEM results (stars) in Figs. 5(c) and 5(g) suggests good agreement between the two through most of the parameter space. However, the effective mobilities computed by the NNM in Fig. 5(g) deviate noticeably from the FEM results at the left endpoint $\sigma = 0.125$. Accordingly, the relative errors and marginal losses plotted in Figs. 5(k) and 5(o) are also highest at $\sigma = 0.125$. In fact, ε and $\mathcal{L}(g_0|\lambda, \sigma)$ of the NNM solutions parametrized by σ [Figs. 5(k) and 5(o)] exhibit the same structure previously identified (Sec. IV B 2) in ε and $\mathcal{L}(g_0|\lambda, \sigma)$ of the NNM solutions parametrized by λ [Figs. 5(j) and 5(n)]. That is, ε and $\mathcal{L}(g_0|\lambda, \sigma)$ are roughly uniform for intermediate values of σ but increase sharply near the boundaries of the training domain. In particular, ε and $\mathcal{L}(g_0|\lambda, \sigma)$ are consistently higher at the low- σ endpoint than at the high- σ endpoint.

Overall, the relative error in Fig. 5(k) is well below the 1% error threshold for most of the training range. As was the case in Sec. IV B 2, at the few points where relative error exceeds 1%, it only does so by a small amount. The marginal loss continues to behave as an *a posteriori* measure of solution accuracy and suggests that the regions of high relative error are once again concentrated near the endpoints of the training range. Despite the fact that parameter σ directly changes the domain geometry in addition to modifying the terms of the PDE, the performance measured in this section is essentially the same as that reported in Sec. IV B 2, where the NNM was parametrized by the simpler parameter λ . Thus, it appears that the NNM can handle geometry-modifying parameters just as easily as parameters that do not modify domain geometry. This is particularly interesting given the difficulty of treating parametrized geometries with other reduced-order modeling techniques.

4. Neural network method parameterized by field strength and particle size

The results shown so far have established that the NNM can robustly solve the g_0 equation in the slit-well MNFD (Sec. IV B 1), and that the method can easily be extended to produce solutions parametrized by field strength λ (Sec. IV B 2) or particle size σ (Sec. IV B 3). Expanding upon this capability, in this section the NNM is used to approximate

g_0 as a function of both λ and σ simultaneously (Fig. 3). Specifically, a single neural network is trained to approximate the four-dimensional function $g_0(x, y; \lambda, \sigma)$ over the same parameter space previously spanned by five networks in Secs. IV B 2 and IV B 3 or 50 networks in Sec. IV B 1.

The MFPT $\langle \tau \rangle$, effective mobility μ_{eff} , relative error ε , and marginal loss $\mathcal{L}(g_0|\lambda, \sigma)$ are computed from the NNM solution $g_0(x, y; \lambda, \sigma)$ and plotted in Figs. 5(d), 5(h), 5(l), and 5(p). The lines are shown as functions of σ and evaluated at the same choices of λ used in Sec. IV B 3 [indicated by the legend in Fig. 5(o)]. The $\langle \tau \rangle$ and μ_{eff} values plotted in Figs. 5(d) and 5(h) closely match those in Figs. 5(c) and 5(g), demonstrating that the NNM parametrized by both λ and σ can resolve all the same major physical phenomena previously identified in Secs. IV B 2–IV B 3.

However, the accuracy of the solution $g_0(x, y; \lambda, \sigma)$ is slightly worse than that observed in the previous sections [Figs. 5(a)–5(c) and 5(e)–5(g)] as visible in $\langle \tau \rangle$ and μ_{eff} [Figs. 5(d) and 5(h)] and quantitatively confirmed by ε and $\mathcal{L}(g_0|\lambda, \sigma)$ [Figs. 5(l) and 5(p)]. This is not entirely surprising, since the four-dimensional problem here is intrinsically more difficult than the three-dimensional (Secs. IV B 2 and IV B 3) and two-dimensional formulations (Sec. IV B 1) of the problem. Moreover, the network depth and width were held constant over all experiments, and the training time was held constant for all the parametrized formulations (Sec. III A). Regardless, although $g_0(x, y; \lambda, \sigma)$ appears somewhat less accurate than the solutions from previous sections, it generally still meets the target 1% error threshold over most of its parameter training range.

An exception to this statement is presented by the results at $\lambda = 5$ [the brown lines in Figs. 5(d), 5(h), 5(l), and 5(p)], for which the error of $g_0(x, y; \lambda, \sigma)$ is greater than 1% over nearly the entire σ training range. The marginal loss also reflects this poor performance; for $\lambda = 5$, $\mathcal{L}(g_0|\lambda, \sigma)$ in Fig. 5(p) is more than an order of magnitude larger than $\mathcal{L}(g_0|\lambda, \sigma)$ from all previous experiments [i.e., those in Figs. 5(m)–5(o)], and several times larger than the other $\mathcal{L}(g_0|\lambda, \sigma)$ curves in Fig. 5(p). Conspicuously, the $\mathcal{L}(g_0|\lambda, \sigma)$ curves in Fig. 5(p) vary significantly with λ , whereas in Figs. 5(m)–5(o) very little variation was observed between the different $\mathcal{L}(g_0|\lambda, \sigma)$ curves.

Of course, the results in Fig. 5(p) differ fundamentally from those in Figs. 5(m)–5(o); whereas each curve in Figs. 5(m)–5(o) corresponds to one or more independent networks, all the curves in Fig. 5(p) are generated by a single network. In fact, the brown ($\lambda = 5$) and blue ($\lambda = 50$) curves, which exhibit the highest marginal losses in Fig. 5(p), lie directly on the boundary of the network's (λ, σ) training domain. When analyzing the NNM parametrized by λ or σ (Secs. IV B 2 and IV B 3), a substantial deterioration in accuracy was found to be highly localized near the boundaries of the parameter training range. If a similar boundary effect exists here for the $g_0(x, y; \lambda, \sigma)$ solution, then the results in Figs. 5(d), 5(h), 5(l), and 5(p) are not representative of the solution's overall accuracy over the entire problem parameter space, as essentially half of the data shown in those plots lie on the boundary of the network's parameter training space.

To investigate this possibility, the same metrics that are shown as discrete lines in Figs. 5(d), 5(h), 5(l), and 5(p) are

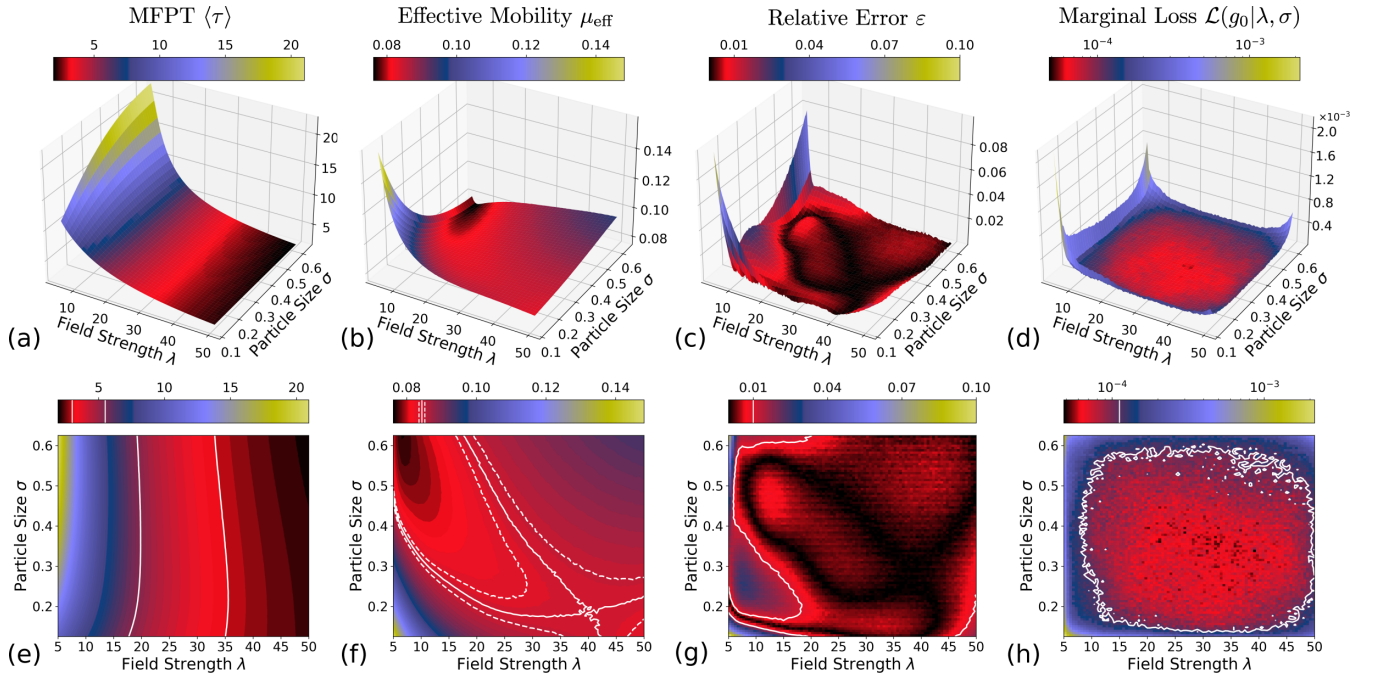


FIG. 6. Analysis of the $g_0(x, y; \lambda, \sigma)$ solution obtained using the NNM parametrized by both field strength λ and particle size σ . (a), (e) Mean first passage time $\langle \tau \rangle$ with white contours to show nonmonotonic sorting behavior. (b), (f) Effective mobility μ_{eff} with white contours to show saddle point. (c), (g) Relative error ε of the MFPT with white contour to denote 1% error threshold. (d), (h) Marginal loss $\mathcal{L}(g_0|\lambda, \sigma)$ with white contour to denote testing loss $\mathcal{L}$.

replotted in Fig. 6 as continuous functions of (λ, σ) . The first row [Figs. 6(a)–6(d)] shows three-dimensional plots of the metrics over parameter space, whereas the second row shows two-dimensional contour maps [Figs. 6(e) and 6(f)] and color maps [Figs. 6(g) and 6(h)] of the same metrics. As anticipated, the relative error ε [Figs. 6(c) and 6(g)] and marginal loss $\mathcal{L}(g_0|\lambda, \sigma)$ [Figs. 6(d) and 6(h)] are only large near the boundaries of the parameter training space. Indeed, ε in Fig. 6(g) is below the 1% threshold (indicated by the solid white line) throughout the majority of the parameter space, confirming the suspicion that the line plots in Fig. 5 provide a biased view of the $g_0(x, y; \lambda, \sigma)$ solution.

The boundary effect is particularly clear in $\mathcal{L}(g_0|\lambda, \sigma)$ in Fig. 6(d), which features a prominent convex shape. Here, $\mathcal{L}(g_0|\lambda, \sigma)$ is consistently higher along all the edges of the training parameter space and decreases monotonically and rapidly away from the boundary. In particular, $\mathcal{L}(g_0|\lambda, \sigma)$ is exceptionally large at the corners of the training space.

Note that the decay of marginal loss away from the boundaries of the parameter space is actually substantially sharper than it appears visually in Figs. 6(d) and 6(h). The color scales for Figs. 6(d) and 6(h) are logarithmic and the color map is not perceptually uniform: it exhibits far more variation in color and contrast near the lower end of the scale. These plotting choices make the subtle structure of the marginal loss more apparent, but give it the biased appearance of a gradual variation throughout the domain. In actuality, when plotted with a linear color scale and a perceptually uniform color map, the marginal loss appears essentially flat through most of the domain.

As expected, the relative error ε [Figs. 6(c) and 6(g)] is closely tied to the marginal loss $\mathcal{L}(g_0|\lambda, \sigma)$. Relative

error is uniformly low in the interior of the parameter space (roughly $(\lambda, \sigma) \in [15, 45] \times [0.2, 0.6]$), corresponding to the flat interior of $\mathcal{L}(g_0|\lambda, \sigma)$. Additionally, near the two corners at $\lambda = 5$ where $\mathcal{L}(g_0|\lambda, \sigma)$ is largest, ε also attains its highest values, approaching 10%. There is also a small peak in ε at the $(\lambda, \sigma) = (50, 0.125)$ corner, corresponding to an equally small peak in $\mathcal{L}(g_0|\lambda, \sigma)$ at the same corner. Surprisingly, although $\mathcal{L}(g_0|\lambda, \sigma)$ exhibits a clear peak at the $(\lambda, \sigma) = (50, 0.625)$ corner, ε does not. Therefore, the marginal loss $\mathcal{L}(g_0|\lambda, \sigma)$ once again appears to act as a conservative *a posteriori* estimator of relative error ε : high relative error occurs near regions of high marginal loss, although high marginal loss does not always imply high relative error.

As noted above, the performance of the $g_0(x, y; \lambda, \sigma)$ solution deteriorates even more significantly at the corners of the parameter training space than on its edges. This is more complicated than the boundary effect discussed for the solutions parametrized by just λ or σ and can be accounted for by extending the postulated mechanism from Secs. IV B 2 and IV B 3. There, it was argued that the deterioration in performance arises because the stochastic sampling used during training under-represents boundary points: whereas the neighborhoods of interior points are thoroughly sampled on all sides, this is not true for boundary points. In the two-dimensional parameter training space considered here, the corners and the edges of the boundary are under-represented to different extents by the stochastic sampling process. Whereas parameter values on the edges of the domain only have 50% as many neighboring points inside the training space as interior points, parameter values on the corners have only 25% as many. This tentatively explains why performance

is so much worse at corners of the parameter training space than it is on the edges.

If this mechanism extends to higher dimensional parameter spaces, it may eventually prove to be a dominant source of error: for instance, the corners of an n -dimensional hypercube have only $1/2^n$ as many neighbors inside the training space as interior points, and the number of boundary segments (corners, edges, faces, ...) grows rapidly with n . In fact, the fraction of a parameter space lying within a given distance of its boundary also increases with dimensionality. Altogether, these observations suggest the need for further investigation into this boundary effect, its possible connection to Monte Carlo sampling of the loss during training, and methods (such as low-discrepancy sampling methods [66,67]) for resolving the problem.

In contrast to the line plots in Fig. 5, the plots in Fig. 6 highlight the richness of information available through the $g_0(x, y; \lambda, \sigma)$ solution compared with the solutions parametrized only by λ (Sec. IV B 2), σ (Sec. IV B 3), or neither (Sec. IV B 1). For instance, although the NNM solutions parametrized by λ or σ (Secs. IV B 2 and IV B 3) suggested a nonmonotonic dependence of $\langle \tau \rangle$ and μ_{eff} with respect to σ for certain values of λ , they did not provide sufficient information to estimate the exact range of λ over which this behavior persists. Just as the NNM solutions parametrized by λ or σ (Secs. IV B 2 and IV B 3) are more helpful than the fixed parameter solutions (Sec. IV B 1) in localizing one-dimensional critical points, so is the NNM solution parametrized by both λ and σ more useful for delineating the nonmonotonic regions of parameter space.

The range of nonmonotonic behavior can be estimated visually from μ_{eff} in Figs. 6(b) and 6(f). Using a vertical line test, it is easy to see that nonmonotonic dependence of μ_{eff} on σ is present at voltages as low as $\lambda \approx 10$. In fact, μ_{eff} is doubly nonmonotonic with respect to both λ and σ in the large- σ , low- λ range [black region in Figs. 6(b) and 6(f)]. Although the same trends were suspected from the solutions discussed in Secs. IV B 2 and IV B 3, $g_0(x, y; \lambda, \sigma)$ resolves the features more completely.

Despite the usefulness of $g_0(x, y; \lambda, \sigma)$ for resolving critical points, the solution predicts a false saddle point in μ_{eff} at $(\lambda, \sigma) \approx (40, 0.2)$ [highlighted by the white solid line in Fig. 6(f)]. Additional FEM results (not shown) confirm that there is no saddle point anywhere in the parameter space under consideration. This error can be attributed to the fact that the true μ_{eff} changes extremely little in the high- λ , low- σ region of the domain. For illustration, the two dotted white lines in Fig. 6(f) indicate contours for μ_{eff} values 1% greater and smaller, respectively, than the value of μ_{eff} on the solid white line passing through the saddle point. Despite this range corresponding to a very small fraction of the total variation of μ_{eff} over the domain, the area between the dotted white lines account for roughly 25% of the total parameter training space, demonstrating that μ_{eff} is extremely flat throughout this entire region.

Although the presence of a false saddle point is a noticeable qualitative error, it corresponds to a very small quantitative error in the key output metrics $\langle \tau \rangle$ and μ_{eff} . In fact, the visual appearance of the saddle point in Fig. 6(f) is intentionally accentuated by the choice of color map, as discussed for

the marginal loss above. It is quite feasible that an error of such small magnitude could be resolved simply by increasing network capacity and/or training time.

Still, the question arises of whether and how the NNM can be used reliably in applications where these types of incorrect or ill-conditioned features may occur. The marginal loss $\mathcal{L}(g_0|\lambda, \sigma)$ provides one possible resolution to this concern. The region of increased $\mathcal{L}(g_0|\lambda, \sigma)$ near $(\lambda, \sigma) = (50, 0.125)$ in Fig. 6(h) coincides fairly closely with the right half of the saddle point in Fig. 6(f). Thus, $\mathcal{L}(g_0|\lambda, \sigma)$ correctly reflects that the solution is less reliable in this region, drawing into question the validity of the predicted saddle point.

Future work should elaborate on what quantitative predictions of solution quality can be based on the marginal loss, along the lines of the investigations of Magill *et al.* [57]. In the interim, we propose using the total loss [as indicated in Fig. 6(h) by the solid white line] as an approximate threshold between regions of relatively high and low expected accuracy. In fact, the marginal loss $\mathcal{L}(g_0|\lambda, \sigma)$ as defined here is likely a suboptimal tool for the detection of false critical points in parameter space because it does not directly measure gradient information with respect to (λ, σ) . Rather, it is only indirectly sensitive to the error in the shape of $\langle \tau \rangle$ and μ_{eff} insofar as it emerges from errors in the shape of $g_0(x, y; \lambda, \sigma)$. For applications in which the localization of ill-conditioned critical points is of interest, modified loss functions that incorporate the derivatives of the target PDE with respect to λ and σ (e.g., like those explored by Avrutskiy [59]) might be more relevant error estimators. This notion illustrates the potential benefits of customizing the NNM for specific PDEs and research questions, just as flux- or energy-conserving numerical methods are preferred for applications where those features are particularly important.

In summary, the results in this section demonstrate that the NNM can produce a robust approximation to the $g_0(x, y; \lambda, \sigma)$ solution. Here, $g_0(x, y; \lambda, \sigma)$ enables higher-dimensional visualization of $\langle \tau \rangle$ and μ_{eff} over λ and σ , resolving features in parameter space more accurately and completely than the solutions parametrized by only λ or σ . Furthermore, $g_0(x, y; \lambda, \sigma)$ accurately predicts the magnitude of $\langle \tau \rangle$ and μ_{eff} to within the 1% error threshold simultaneously over the majority of the parameter training space. Although $g_0(x, y; \lambda, \sigma)$ exhibits some regions of high relative error, $\mathcal{L}(g_0|\lambda, \sigma)$ once again provides a robust *a posteriori* estimator of the solution's reliability throughout the parameter space.

V. CONCLUSIONS

This work investigated the use of the neural network method to solve a parametrized time-integrated Smoluchowski equation describing nanoparticle passage through the slit-well microfluidic device. The g_0 solutions were solved for a variety of fixed choices of field strength λ and particle size σ using both the NNM and a standard FEM implementation. Additionally, the NNM was used to solve the equation directly as a function of λ and/or σ . Mean first passage time $\langle \tau \rangle$ and effective mobility μ_{eff} were studied as the primary output metrics of interest, with relative error ε and marginal loss $\mathcal{L}(g_0|\lambda, \sigma)$ used to characterize solution performance.

The qualitative examinations of g_0 in Sec. IV A revealed a wide variety of functional behavior over the region of (λ, σ) parameter space studied here. The four primary regimes underlying nanoparticle sorting in the slit-well (i.e., low and high fields for small and large particles) correspond to four significantly different g_0 solution types, each reflecting different interplays of drift, diffusion, and geometry. This highlights the challenging nature of the parametrized PDE problem studied in this work. Additionally, this analysis suggested that the g_0 solutions themselves may encode interesting and useful qualitative information about biophysical processes. Future work should examine how information encoded in g_0 may be complementary to qualitative information derived from stochastic particle trajectories.

Of course, qualitative insights aside, the most salient feature of g_0 is that it integrates to yield the mean first passage time $\langle \tau \rangle$. As noted, although $\langle \tau \rangle$ is a quantity of widespread interest in all first passage problems and is relevant to many MNFD design problems, it appears that numerical solutions of g_0 have rarely been leveraged for such applications. The results of this paper support that g_0 may be an undervalued tool in computational biophysics.

Although g_0 can be computed using many methods, such as FEM or particle simulations, this work focused on resolving g_0 using the NNM. When applied to fixed choices of problem parameters, the NNM consistently estimated $\langle \tau \rangle$ with errors below 1%. In particular, the NNM values were at least as accurate as typical particle simulations, which are the most common tool for studying first passage problems in biophysics. However, a proper comparison of runtime was not conducted in this work, and should be a major focus of future investigations.

The main appeal of the NNM is the unique ease with which it can be applied to parametrized g_0 problems. Via integration of g_0 , these solutions yield a direct mapping from key problem inputs (e.g., λ, σ) to key problem outputs (e.g., $\langle \tau \rangle, \mu_{\text{eff}}$). This is particularly appealing for the application of MNFD research, where essential phenomena often depend nontrivially on the coupling of many system parameters. The results in the current work demonstrate that the NNM can learn accurate approximations of g_0 parametrized by λ, σ , or both, all using a modest network size and even without careful hyperparameter optimization. Whereas classical ROM techniques typically require special considerations to handle geometry-modifying parameters like σ , the NNM was found to resolve $g_0(x, y; \sigma)$ just as easily as $g_0(x, y; \lambda)$. As discussed, parametrized solutions can be quite useful in characterizing entire regions of parameter space.

Although the NNM is expected to perform well on highly parametrized PDEs, the careful error analysis presented in the current study revealed several points of caution for future efforts in this direction. First, all parametrized solutions studied here exhibited a deterioration in accuracy near the boundaries of their parameter training space. Nonetheless, the predicted values of $\langle \tau \rangle$ were still mostly within the 1% margin of error. Moreover, the marginal loss functional $\mathcal{L}(g_0|\lambda, \sigma)$ proposed here was found to act as a conservative *a posteriori* estimator of the solution accuracy throughout parameter space.

The second point of caution that must be considered when applying the NNM to parametrized PDEs concerns the

interpretation of key features, such as critical points, that are identified using these solutions. For instance, in Sec. IV B 4, the NNM solution exhibited an erroneous saddle point in a flat region of μ_{eff} , which was an artifact that arose due to the ill-conditioning of the gradients of $\mu_{\text{eff}}(\lambda, \sigma)$. In fact, plots of ε showed no indication of errors in this region, as the mistake only manifested in the curvature of the mapping. However, once again the marginal loss $\mathcal{L}(g_0|\lambda, \sigma)$ did indicate that the NNM solution lost fidelity in this region of parameter.

In summary, the parametrized NNM solutions were generally accurate far from the training boundaries, and $\mathcal{L}(g_0|\lambda, \sigma)$ provided robust regions of confidence. Altogether, these results highlight the specific appeal of the NNM as a method for studying parametrized first passage problems via the time-integrated Smoluchowski model. We hope this work prompts further investigation into the use of g_0 with or without the NNM, and into the relationship of $\langle \tau \rangle$ and μ_{eff} to more standard MNFD metrics. Regarding the application of the NNM to such problems, future work should address technical challenges such as singularities posed by sharp corners, training difficulties for highly skewed geometries, and achieving competitive runtime.

ACKNOWLEDGMENTS

The authors gratefully acknowledge funding from Mitacs under the Accelerate Entrepreneur program (Ref. IT21168) and from Smart Computing for Innovation (SOSCIP Ref. 3-076). H.W.d.H. also acknowledges funding from the Natural Sciences and Engineering Research Council (NSERC) in the form of Discovery Grant No. 2020-07145. M.M. also acknowledges funding from the Ontario Graduate Scholarship (OGS) Program and the Vector Institute Postgraduate Affiliate Program.

APPENDIX A: COMPUTING MEAN FIRST PASSAGE TIMES WITH PARTICLE SIMULATIONS

This section contains a description of standard Brownian dynamics (BD) simulations used to measure the mean first passage times of nanoparticles traversing the slit-well microfluidic device (Sec. II). The BD simulations were initialized with $N = 100\,000$ noninteracting particles placed according to the distribution ρ_0 [Eq. (7)]. The position of the i th particle $\vec{x}_i$ was updated according to the discretized BD equation

$$\frac{\Delta \vec{x}_i}{\Delta t} = \sqrt{\frac{2D}{\Delta t}} \vec{R}(t) + \frac{q\lambda}{\gamma} \vec{E}_0 + \frac{1}{\gamma} \vec{F}_{\text{WCA}}. \quad (\text{A1})$$

In Eq. (A1), the particle properties are the diffusion coefficient D , the friction coefficient γ , and the particle charge q . As noted in Sec. II, both q and γ were set equal to the particle diameter σ , to capture free-draining behavior. The diffusion coefficient D was set to $1/\sigma$ and the time step was set $\Delta t = 10^{-5}$. The term $\vec{R}(t)$ in Eq. (A1) is a random driving force representing the thermal motion of an implicit solvent which was sampled from a uniform distribution of mean 0 and variance 1.

Rather than representing the interactions between particles and walls as perfectly rigid, the walls were implemented using

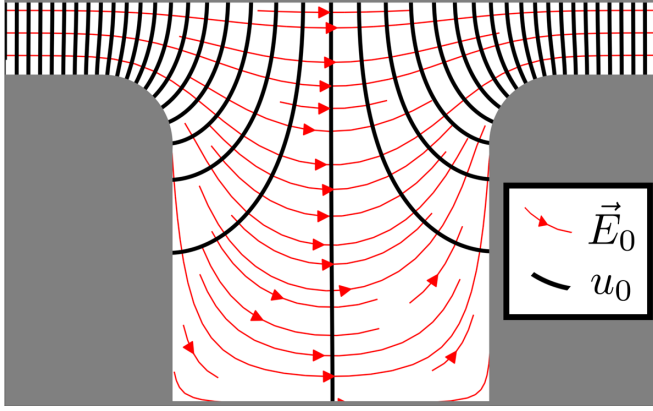


FIG. 7. Contour plots of the baseline electric potential u_0 (black) and field $\vec{E}_0$ (red) computed by the NNM.

a short-range repulsive shifted WCA force

$$\vec{F}_{\text{WCA}} = -\nabla U_{\text{WCA}}, \quad (\text{A2})$$

with

$$U_{\text{WCA}}(r_i) = \begin{cases} 4\epsilon \left[\left(\frac{\sigma_0}{r_i} \right)^{12} - \left(\frac{\sigma_0}{r_i} \right)^6 \right] + \epsilon, & r_i < r_{\text{cut}} \\ 0, & r_i \geq r_{\text{cut}} \end{cases} \quad (\text{A3})$$

where r_i is the minimum distance from particle i to the nearest reflective wall minus a distance $r_{\text{shift}} = 0.5(\sigma - \sigma_0)$. Here r_{shift} corresponds to the radius of the hard core of the particle, whereas $\sigma_0 = 0.125$ is the length over which the surface of the particle is partially compressible. The potential is zero beyond a cutoff distance $r_{\text{cut}} = 2^{1/6}\sigma_0$, so that if the center of the particle is farther than a distance $r_{\text{shift}} + r_{\text{cut}}$ from the wall there is no interaction. The energy scale of the repulsive interaction was set to $\epsilon = 0.125 = \sigma_0$.

Although this type of model is commonly used for particle-wall interactions, due to its improved numerical stability relative to perfectly rigid interactions, it introduces a small difference between the underlying physics of the BD simulations and the PDE models being solved in this work. For this reason, the MFPTs determined using particle simulations should not be expected to agree exactly with those obtained using the NNM and FEM methods, even in the limit of small Δt and large N . Nonetheless, as our results corroborate, the effect of this difference between the models is small.

The term $\vec{E}_0$ in Eq. (A1) corresponds to the baseline electric field in the slit-well domain (denoted by red in Fig. 7). This was solved for a voltage drop of two units from the leftmost to rightmost boundaries, as in Magill *et al.* [57]. The net electric-field strength was set by the parameter λ . $\vec{E}_0$ used here was the same one described in Sec. III B. As shown in Magill *et al.* [57], particle simulations conducted using an electric field solved with the NNM are nearly statistically indistinguishable from those conducted using a field solved with the FEM, so long as the NNM electric field exhibits a sufficiently small loss. The purpose of the present study is not to replicate this result, but to explore the computational advantages of the NNM over other techniques in parametrized problems. Thus, the particle simulations are conducted using

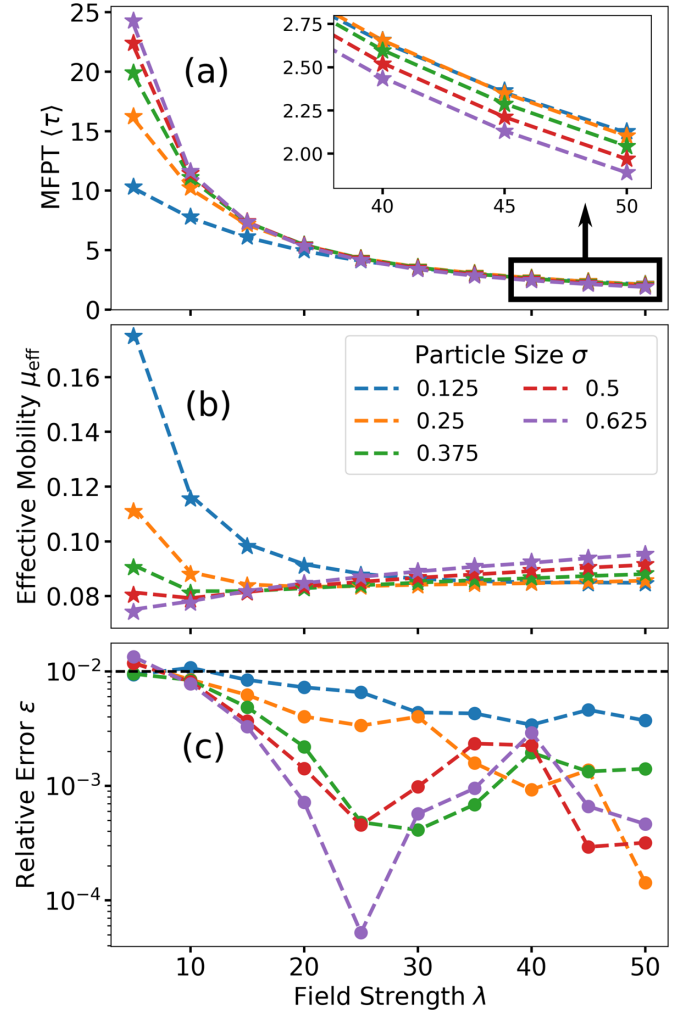


FIG. 8. Passage time properties of particles escaping the slit-well model computed using Brownian dynamics simulations. Star markers denote values obtained via FEM. (a) Mean first passage time $\langle \tau \rangle$. (b) Effective mobility μ_{eff} . (c) Relative error of $\langle \tau \rangle$ computed against the ground truth FEM solution.

the FEM electric field, which is taken as the reference ground truth.

Parallel to the analysis conducted in Sec. IV B, the mean first passage time $\langle \tau \rangle$ and effective mobility μ_{eff} are computed using the BD simulations for various choices of field strength λ and particle size σ . These values are plotted with dashed lines in Fig. 8(a) and 8(b) with star markers to denote $\langle \tau \rangle$ and μ_{eff} values obtained by FEM. Note that $\langle \tau \rangle$ and μ_{eff} are only solved for the same discrete choices of λ and σ that are also computed using FEM.

In addition, the relative error ϵ is computed using Eq. (23) where $\langle \tau \rangle$ and $\langle \tau \rangle_{\text{FEM}}$ are the MFPTs computed by BD and FEM, respectively. The values are plotted in Fig. 8(c) with circular markers denoting the parameter choices where the relative error was computed. All of the relative errors in Fig. 8(c) fall below 2%, with majority of the values being within 1% error. This establishes a 1% error baseline against the ground truth MFPT values computed by FEM, for which to benchmark the performance of the NNM.

TABLE I. Comparison of computational time (in minutes) of the NNM, BD, and FEM methods used to compute the MFPT at fixed parameter values.

Parameters (λ, σ)	Runtime (minutes)		
	NNM	BD	FEM
(5.0, 0.125)	136.95	5.02	2.25
(50.0, 0.125)	138.37	1.03	2.15
(5.0, 0.625)	126.57	12.80	2.18
(50.0, 0.625)	133.63	1.12	2.13

APPENDIX B: CONTOUR PLOTS OF ELECTRIC POTENTIAL AND FIELD

The baseline electric field $\vec{E}_0$ used to drive particle motion in the slit-well device [Eq. (6)] was computed using the NNM, as described in Magill *et al.* [57]. That is, the baseline electric potential u_0 was solved using Laplace's equation over a voltage drop of two units from the left slit wall to the right slit wall. The electric field was then computed using the relation $\vec{E}_0 = \nabla u_0$. The red and black contour lines in Fig. 7 correspond to the electric field $\vec{E}_0$ and electric potential u_0 , respectively, inside the slit-well MNFD.

APPENDIX C: RUNTIME COMPARISON

The MFPT and effective mobility of nanoparticles traversing the slit-well MNFD were obtained using BD simulations, the FEM, and the NNM. Table I shows the runtime, in minutes, of each method used to solve the MFPT at fixed choices parameter values. Four choices of the parameters are included,

TABLE II. Comparison of computational time (in days) of the various methods used to compute the MFPT over ranges of parameter space.

Method	Mean Runtime (days)
NNM parameterized by λ	6.33
NNM parameterized by σ	7.79
NNM parameterized by (λ, σ)	7.66
High-resolution FEM sampling	12.18

illustrating that runtimes were fairly independent of parameters for NNM and FEM but depended strongly on parameters for BD. Table II shows the runtime, in days, of each method used to solve the MFPT over large regions of parameter space. As implemented, the various parametrized NNM methods all have runtimes comparable to one another. Moreover, the total runtime of the high-resolution FEM sampling exceeds the mean runtime for the parametrized NNM methods. However, this runtime obviously depends on the number of points sampled. Here, 8099 parameter combinations were utilized in order to produce high-resolution maps of error over parameter space.

Optimizing runtime was not a goal of the current work. The implementations of each of the algorithms studied here (NNM, BD, FEM) can undoubtedly be improved upon to substantially decrease the runtimes from those reported in Tables I and II. Moreover, judicious use of parallelization across GPUs and/or CPUs, as applicable, could provide further improvements to each of the methods. Thus, the runtimes included here are provided for reference only, and a more careful comparison is left to future work.

- [1] S. L. Levy and H. G. Craighead, DNA manipulation, sorting, and mapping in nanofluidic systems, *Chem. Soc. Rev.* **39**, 1133 (2010).
- [2] K. D. Dorfman, DNA electrophoresis in microfabricated devices, *Rev. Mod. Phys.* **82**, 2903 (2010).
- [3] P. Abgrall and N. T. Nguyen, Nanofluidic devices and their applications, *Anal. Chem.* **80**, 2326 (2008).
- [4] H. Gardeniers and A. Van Den Berg, Micro-and nanofluidic devices for environmental and biomedical applications, *Int. J. Environ. Anal. Chem.* **84**, 809 (2004).
- [5] R. Mulero, A. S. Prabhu, K. J. Freedman, and M. J. Kim, Nanopore-based devices for bioanalytical applications, *J. Assoc. Lab. Anim.* **15**, 243 (2010).
- [6] J. Han and H. G. Craighead, Entropic trapping and sieving of long DNA molecules in a nanofluidic channel, *J. Vac. Sci. Technol., A* **17**, 2142 (1999).
- [7] K.-L. Cheng, Y.-J. Sheng, S. Jiang, and H.-K. Tsao, Electrophoretic size separation of particles in a periodically constricted microchannel, *J. Chem. Phys.* **128**, 101101 (2008).
- [8] H. Wang, H. W. de Haan, and G. W. Slater, Electrophoretic ratcheting of spherical particles in well/channel microfluidic devices: Making particles move against the net field, *Electrophoresis* **41**, 621 (2020).
- [9] J. Fu, P. Mao, and J. Han, Nanofilter array chip for fast gel-free biomolecule separation, *Appl. Phys. Lett.* **87**, 263902 (2005).
- [10] J. Fu, J. Yoo, and J. Han, Molecular Sieving in Periodic Free-Energy Landscapes Created by Patterned Nanofilter Arrays, *Phys. Rev. Lett.* **97**, 018103 (2006).
- [11] M. Langecker, D. Pedone, F. C. Simmel, and U. Rant, Electrophoretic time-of-flight measurements of single DNA molecules with two stacked nanopores, *Nano Lett.* **11**, 5002 (2011).
- [12] J. A. Berkenbrock, R. Grecco-Machado, and S. Achenbach, Microfluidic devices for the detection of viruses: Aspects of emergency fabrication during the covid-19 pandemic and other outbreaks, *Proc. R. Soc. London, Ser. A* **476**, 20200398 (2020).
- [13] S. J. Shepherd, C. C. Warzecha, S. Yadavali, R. El-Mayta, M.-G. Alameh, L. Wang, D. Weissman, J. M. Wilson, D. Issadore, and M. J. Mitchell, Scalable mRNA and siRNA lipid nanoparticle production using a parallelized microfluidic device, *Nano Lett.* **21**, 5671 (2021).
- [14] T. Chou and M. R. D'Orsogna, First passage problems in biology, in *First-passage Phenomena and their Applications* (World Scientific, Singapore, 2014), pp. 306–345.
- [15] M. H. Lam, K. Briggs, K. Kastitis, M. Magill, G. R. Madejski, J. L. McGrath, H. W. de Haan, and V. Tabard-Cossa, Entropic trapping of DNA with a nanofiltered nanopore, *ACS Appl. Nano Mater.* **2**, 4773 (2019).
- [16] K. Briggs, G. Madejski, M. Magill, K. Kastitis, H. W. de Haan, J. L. McGrath, and V. Tabard-Cossa, DNA translocations

- through nanopores under nanoscale preconfinement, *Nano Lett.* **18**, 660 (2018).
- [17] H. W. de Haan and G. W. Slater, Mapping the variation of the translocation α scaling exponent with nanopore width, *Phys. Rev. E* **81**, 051802 (2010).
- [18] M. Magill, Ed Waller, and H. W. de Haan, A sequential nanopore-channel device for polymer separation, *J. Chem. Phys.* **149**, 174903 (2018).
- [19] S. Redner, *A Guide to First-Passage Processes* (Cambridge University Press, Cambridge, 2001).
- [20] V. Kurella, J. C. Tzou, D. Coombs, and M. J. Ward, Asymptotic analysis of first passage time problems inspired by ecology, *Bull. Math. Biol.* **77**, 83 (2015).
- [21] S. Iyer-Biswas and A. Zilman, First passage processes in cellular biology, *Advances in Chemical Physics* **160**, 261 (2016).
- [22] Z. Ma, A. W. Krings, and R. C. Millar, Introduction of first passage time (FPT) analysis for software reliability and network security, in *Proceedings of the 5th Annual Workshop on Cyber Security and Information Intelligence Research: Cyber Security and Information Intelligence Challenges and Strategies* (Association for Computing Machinery, New York, 2009), pp. 1–6.
- [23] E. D. Filho and M. Araújo, Synaptic transmission and Fokker-Planck equation, in *ICORES* (SciTePress, Setúbal, Portugal, 2012), pp. 59–63.
- [24] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, ImageNet: A large-scale hierarchical image database, In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (Institute of Electrical and Electronics Engineers, New Jersey, 2009), pp. 248–255.
- [25] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, *Nature (London)* **521**, 436 (2015).
- [26] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, Natural language processing (almost) from scratch, *J. Mach. Learn. Res.* **12**, 2493 (2011).
- [27] Weinan E, J. Han, and A. Jentzen, Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations, *Commun. Math. Stat.* **5**, 349 (2017).
- [28] J. Sirignano and K. Spiliopoulos, DGM: A deep learning algorithm for solving partial differential equations, *J. Comput. Phys.* **375**, 1339 (2018).
- [29] V. R. Royo and C. Tomlin, Recursive regression with neural networks: Approximating the HJI PDE solution, *arXiv:1611.02739*.
- [30] V. I. Avrutskiy, Neural networks catching up with finite differences in solving partial differential equations in higher dimensions, *Neural Computing and Applications* (Springer, New York, 2020), pp. 1–16.
- [31] J. Han, A. Jentzen, and E. Weinan, Solving high-dimensional partial differential equations using deep learning, *Proc. Natl. Acad. Sci. USA* **115**, 8505 (2018).
- [32] C. Huré, H. Pham, and X. Warin, Deep backward schemes for high-dimensional nonlinear PDEs, *Mathematics of Computation* (American Mathematical Society, Providence, 2020).
- [33] S. Karumuri, R. Tripathy, I. Bilonis, and J. Panchal, Simulator-free solution of high-dimensional stochastic elliptic partial differential equations using deep neural networks, *J. Comput. Phys.* **404**, 109120 (2020).
- [34] H. Mutuk, Neural network study of hidden-charm pentaquark resonances, *Chin. Phys. C* **43**, 093103 (2019).
- [35] M. A. Nabian and H. Meidani, A deep learning solution approach for high-dimensional random differential equations, *Probab. Eng. Mech.* **57**, 14 (2019).
- [36] D. Zhang, L. Guo, and G. Em Karniadakis, Learning in modal space: Solving time-dependent stochastic PDEs using physics-informed neural networks, *SIAM J. Sci. Comput.* **42**, A639 (2020).
- [37] C. Beck, Weinan E, and A. Jentzen, Machine learning approximation algorithms for high-dimensional fully nonlinear partial differential equations and second-order backward stochastic differential equations, *J. Nonlinear Sci.* **29**, 1563 (2019).
- [38] Q. Wei, Y. Jiang, and J. Z. Y. Chen, Machine-learning solver for modified diffusion equations, *Phys. Rev. E* **98**, 053304 (2018).
- [39] P. Grohs, F. Hornung, A. Jentzen, and P. von Wurstemberger, A proof that artificial neural networks overcome the curse of dimensionality in the numerical approximation of Black-Scholes partial differential equations, *arXiv:1809.02362*.
- [40] A. Jentzen, D. Salimova, and T. Welti, A proof that deep artificial neural networks overcome the curse of dimensionality in the numerical approximation of Kolmogorov partial differential equations with constant diffusion and nonlinear drift coefficients, *Commun. Math. Sci.* **19**, 1167 (2021).
- [41] M. Hutzenthaler, A. Jentzen, T. Kruse, and T. A. Nguyen, A proof that rectified deep neural networks overcome the curse of dimensionality in the numerical approximation of semilinear heat equations, *SAM Research Report 2019* (ETH, Zürich, Switzerland, 2019).
- [42] O. Hennigh, S. Narasimhan, M. Amin Nabian, A. Subramaniam, K. Tangsali, Z. Fang, M. Rietmann, W. Byeon, and S. Choudhry, NVIDIA SimNet: An AI-accelerated multi-physics simulation framework, in *International Conference on Computational Science* (Springer, 2021), pp. 447–461.
- [43] J. Berner, M. Dablander, and P. Grohs, Numerically solving parametric families of high-dimensional Kolmogorov partial differential equations via deep learning, *arXiv:2011.04602*.
- [44] K. Glau and L. Wunderlich, The deep parametric PDE method: application to option pricing, *arXiv:2012.06211*.
- [45] Y. Khoo, J. Lu, and L. Ying, Solving parametric pde problems with artificial neural networks, *Eur. J. Appl. Math.* **32**, 421 (2021).
- [46] M. Abid Bazaz, S. Janardhanan *et al.*, A review of parametric model order reduction techniques, in *2012 IEEE International Conference on Signal Processing, Computing and Control* (IEEE, 2012), pp. 1–6.
- [47] K. Lu, K. Zhang, H. Zhang, X. Gu, Y. Jin, S. Zhao, C. Fu, and Y. Yang, A review of model order reduction methods for large-scale structure systems, *Shock and Vibration* (2021).
- [48] N. R. Franco, A. Manzoni, and P. Zunino, A deep learning approach to reduced order modelling of parameter dependent partial differential equations, *arXiv:2103.06183*.
- [49] M. F. Kasim, D. Watson-Parris, L. Deaconu, S. Oliver, P. Hatfield, D. H. Froula, G. Gregori, M. Jarvis, S. Khatiwala, J. Korenaga *et al.*, Building high accuracy emulators for scientific simulations with deep neural architecture search, *Mach. Learn.: Sci. Technol.* **3**, 015013 (2022).
- [50] H. Eivazi, H. Veisi, M. Hossein Naderi, and V. Esfahanian, Deep neural networks for nonlinear model order reduction of unsteady flows, *Phys. Fluids* **32**, 105104 (2020).

- [51] T. Daniel, F. Casenave, N. Akkari, and D. Ryckelynck, Model order reduction assisted by deep neural networks (rom-net), *Adv. Model. Simul. Eng. Sci.* **7**, 16 (2020).
- [52] S. Fresca *et al.*, A comprehensive deep learning-based approach to reduced order modeling of nonlinear time-dependent parametrized PDEs, *J. Sci. Comput* **87**, 61 (2021).
- [53] V. De La Rubia, A. Lamecki, and M. Mrozowski, Geometry parametric model order reduction with randomly generated projection bases, in *2016 IEEE MTT-S International Conference on Numerical Electromagnetic and Multiphysics Modeling and Optimization* (IEEE, 2016), pp. 1–2.
- [54] P. Heydari and M. Pedram, Model reduction of variable-geometry interconnects using variational spectrally-weighted balanced truncation, in *IEEE/ACM International Conference on Computer Aided Design. ICCAD 2001. IEEE/ACM Digest of Technical Papers (Cat. No. 01CH37281)* (IEEE, 2001), pp. 586–591.
- [55] L. Daniel, O. C. Siong, L. S. Chay, K. H. Lee, and J. White, A multiparameter moment-matching model-reduction approach for generating geometrically parameterized interconnect performance models, *IEEE Trans. Comput. Aid. D.* **23**, 678 (2004).
- [56] H. Gao, L. Sun, and J.-X. Wang, Phygeonet: Physics-informed geometry-adaptive convolutional neural networks for solving parameterized steady-state pdes on irregular domain, *J. Comput. Phys.* **428**, 110079 (2021).
- [57] M. Magill, A. M. Nagel, and H. W. de Haan, Neural network solutions to differential equations in nonconvex domains: Solving the electric field in the slit-well microfluidic device, *Phys. Rev. Research* **2**, 033110 (2020).
- [58] A. Al-Aradi, A. Correia, D. Naiff, G. Jardim, and Y. Saporito, Solving nonlinear and high-dimensional partial differential equations via deep learning, [arXiv:1811.08782](https://arxiv.org/abs/1811.08782).
- [59] V. I. Avrutskiy, Enhancing function approximation abilities of neural networks by training derivatives, *IEEE Trans. Neural Netw. Learn. Syst.* **32**, 916 (2020).
- [60] Lu Lu, P. Jin, and G. Em Karniadakis, DeepOnet: Learning nonlinear operators for identifying differential equations based on the universal approximation theorem of operators, [arXiv:1910.03193](https://arxiv.org/abs/1910.03193).
- [61] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser *et al.*, TensorFlow: Large-scale machine learning on heterogeneous systems, [arXiv:1603.04467](https://arxiv.org/abs/1603.04467) (2015).
- [62] X. Glorot and Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* (Proceedings of Machine Learning Research, Sardinia, Italy, 2010), pp. 249–256.
- [63] D. Kingma and J. Ba, Adam: A method for stochastic optimization, [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
- [64] M. Alnæs, J. Blechta, J. Hake, A. Johansson, B. Kehlet, A. Logg, C. Richardson, J. Ring, M. E. Rognes, and G. N. Wells, The FEniCS project version 1.5, *Archive of Numerical Software* **3**, 9 (2015).
- [65] I. V. Grigoriev, Y. A. Makhnovskii, A. M. Berezhkovskii, and V. Yu Zitserman, Kinetics of escape through a small hole, *J. Chem. Phys.* **116**, 9574 (2002).
- [66] J. Chen, R. Du, P. Li, and L. Lyu, Quasi-Monte Carlo sampling for machine-learning partial differential equations, [arXiv:1911.01612](https://arxiv.org/abs/1911.01612).
- [67] S. Mishra and T. K. Rusch, Enhancing accuracy of deep learning algorithms by training with low-discrepancy sequences, *SIAM J. Numer. Anal.* **59**, 1811 (2021).