

Machine learning for parameter auto-tuning in molecular dynamics simulations: Efficient dynamics of ions near polarizable nanoparticles

The International Journal of High Performance Computing Applications 2020, Vol. 34(3) 357–374
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1094342019899457
journals.sagepub.com/home/hpc



JCS Kadupitiya[✉], Geoffrey C Fox and Vikram Jadhao[✉]

Abstract

Simulating the dynamics of ions near polarizable nanoparticles (NPs) using coarse-grained models is extremely challenging due to the need to solve the Poisson equation at every simulation timestep. Recently, a molecular dynamics (MD) method based on a dynamical optimization framework bypassed this obstacle by representing the polarization charge density as virtual dynamic variables and evolving them in parallel with the physical dynamics of ions. We highlight the computational gains accessible with the integration of machine learning (ML) methods for parameter prediction in MD simulations by demonstrating how they were realized in MD simulations of ions near polarizable NPs. An artificial neural network–based regression model was integrated with MD simulation and predicted the optimal simulation timestep and optimization parameters characterizing the virtual system with 94.3% success. The ML-enabled auto-tuning of parameters generated accurate dynamics of ions for ≈ 10 million steps while improving the stability of the simulation by over an order of magnitude. The integration of ML-enhanced framework with hybrid Open Multi-Processing / Message Passing Interface (OpenMP/MPI) parallelization techniques reduced the computational time of simulating systems with thousands of ions and induced charges from thousands of hours to tens of hours, yielding a maximum speedup of ≈ 3 from ML-only acceleration and a maximum speedup of ≈ 600 from the combination of ML and parallel computing methods. Extraction of ionic structure in concentrated electrolytes near oil–water emulsions demonstrates the success of the method. The approach can be generalized to select optimal parameters in other MD applications and energy minimization problems.

Keywords

Machine learning, nanoscale simulation, energy minimization, hybrid OpenMP/MPI, parallel computing, auto-tuning

1. Introduction

Many biological and synthetic nanoparticle (NP) systems are polarized in the presence of electric fields generated by surrounding ions and other macromolecular charged species (Abreu et al., 2008; Clapham, 2007; Levin, 2005). Examples include proteins and DNA in an aqueous cellular medium, emulsions where oil and water are partitioned, and gold NPs dispersed in water. Accurate knowledge of ionic structure near the surface of these NPs enables the understanding of many nanoscale phenomena associated with these materials such as protein conformational changes (Honig and Nicholls, 1995), DNA precipitation (Raspaud et al., 1998), spontaneous emulsification (Sacanna et al., 2007), and NP self-assembly (Levin, 2005). Extracting this structure by simulating the dynamics of ions in the presence of polarizable NPs using coarse-grained models is challenging due to the need to compute

polarization (induced) charges to propagate the ion configuration (Allen et al., 2001; dos Santos et al., 2011; Fahrenberger et al., 2014; Marchi et al., 2001). This computation typically involves solving the second-order Poisson differential equation in three-dimensional space at each simulation timestep, making the use of conventional nanoscale simulation methods very time-consuming and inefficient. Because of these computational challenges, the problem of extracting ionic structure near polarizable NPs has been a subject of intense research (Allen et al., 2001; Boda et al., 2004; dos Santos et al., 2011; dos Santos and Netz,

Intelligent Systems Engineering, Indiana University, Bloomington, IN, USA

Corresponding author:

Vikram Jadhao, Intelligent Systems Engineering, Indiana University, 700 N. Woodlawn Avenue, Bloomington, IN 47408, USA.
Email: vjadhao@iu.edu

2018; Fahrenberger et al., 2014; Gan et al., 2015; Jadhao et al., 2012, 2013; Marchi et al., 2001; Qin et al., 2016).

The problem is often recasted in terms of energy minimization for which different candidate functionals and associated minimization methods have been proposed (Allen et al., 2001; Attard, 2003; Barros et al., 2014; Marchi et al., 2001). Among these techniques, a molecular dynamics (MD) method based on the dynamical optimization of an energy functional enabled the replacement of the expensive solution of the Poisson equation at each simulation step with an on-the-fly computation of surface polarization charges (Jadhao et al., 2012, 2013; Jing et al., 2015). The main focus of this article is to highlight the computational gains accessible by integrating machine learning (ML) methods for parameter auto-tuning in MD simulations by demonstrating how these gains were realized in the MD simulations of ions near polarizable NPs based on the dynamical optimization framework.

In the dynamical optimization framework, inspired by the Car–Parrinello method for simulating ion–electron systems (Car and Parrinello, 1985), an energy functional of the induced charge density is dynamically optimized resulting in the physical dynamics of ions in parallel with the update of the virtual variables characterizing the induced charge density. The virtual system is evolved in a manner that keeps the induced charges close to the free-energy minimum (“ground state”) corresponding to the evolving ionic configuration. The advantages associated with the on-the-fly computation of polarization effects in conjunction with the reduction in computational costs achieved by solving for the scalar induced charge density variable have enabled the study of electrolyte solutions near polarizable NPs using this framework (Jadhao et al., 2012; Jing et al., 2015). However, the applicability of the original method is limited by the absence of a framework that automates the process of selecting the “good” parameters characterizing the virtual system as well as the optimal simulation timestep. These quantities determine the stability, accuracy, and overall efficiency of the dynamical optimization framework and they are found by a tedious process of trial and error that is informed by domain experience. Further, these parameters are selected at the start of the simulation and held fixed throughout the simulation, to often relatively conservative values, to ensure the long-time stability of the dynamics of ions.

Recent years have witnessed a remarkable growth in the use of ML to enhance computational methods aimed at understanding phenomena in materials science, biology, neuroscience, and physics (Bartók et al., 2017; Balakrishnan and Puthusserypady, 2005; Ch’ng et al., 2017; Ferguson, 2017; Liu et al., 2017; Long et al., 2015; Schoenholz, 2018; Ward et al., 2018a). ML has been applied to identify interesting parameter spaces (Spellings and Glotzer, 2018), predict parameters (Balakrishnan and Puthusserypady, 2005), update configurations (Botu and Ramprasad, 2015), infer assembly landscapes (Ferguson, 2017; Long

et al., 2015), predict the properties of materials (Ward et al. 2016, 2018b), and classify phases of matter (Ch’ng et al., 2017). Inspired by these recent developments, we describe an approach to integrate ML and MD methods to predict and auto-tune relevant parameters and simulation timestep. This approach is applied to the dynamical optimization framework to predict on-the-fly the virtual system parameters and simulation timestep that keeps the polarization charge density close to the ground state determined by the evolving ionic configuration at all times during the simulation. The demonstration of the use of ML to predict and tune the MD simulation timestep has broad applicability. Similarly, we expect that the idea of using ML for predicting the virtual system parameters can be extended to enhance the original Car–Parrinello molecular dynamics (CPMD) techniques (Car and Parrinello, 1985) for simulating ion–electron systems.

The use of ML to enhance the performance of the dynamical optimization framework is demonstrated using an $O(n^2)$ algorithm to propagate the dynamics of ions and virtual system variables that is accelerated by implementing a hybrid Open Multi-Processing / Message Passing Interface (OpenMP/MPI) parallelization approach to reduce the computing time associated with the evaluation of the forces and energies. The target applications of the framework are systems where the effects of NP surface charge and ion correlations typically lead to ion distributions that reach constant bulk value within a few nanometers of the NP surface such that a comprehensive study of ion densities near NP surfaces can be performed by including thousands of ions in a large simulation cell with reflective boundaries (Boda et al., 2004; dos Santos et al., 2011; Messina, 2002). Many synthetic and biological systems including oil–water emulsions, gold NPs, and globular proteins exhibit this scenario, and ion distributions in these systems have been analyzed using $O(n^2)$ methods that are competitive with $O(n \log n)$ methods for these moderately sized systems (Allen et al., 2001; Boda et al., 2004; dos Santos et al., 2011; Hatlo and Lue, 2008; Jadhao et al., 2012; Messina, 2002). Further, attempts to ameliorate this scaling via the use of Ewald sums (Deserno and Holm, 1998) or multigrid methods (Sagui and Darden, 2001) would introduce more variables and parameters into the system making the assessment of the coupling of the ML-enabled parameter selection process with the simulation of ions and virtual system difficult. The use of such methods is thus avoided in this first study of developing ML-based enhancements for the treatment of polarizability effects in ionic soft-matter simulations; future work will include integrating this approach with fast Ewald solvers for incorporating long-range effects.

As we discuss later in the results section, the ML-enhanced dynamical optimization framework leads to an increase in both the efficiency and the stability of the associated MD simulations, while retaining the accuracy of the unautomated framework. This combination of ML

and parallel computing in the context of nanoscale simulation of ions is the first of its kind and paves the way for developing online applications for web-based platforms like nanoHUB (Klimeck et al., 2008), where the user engages with the simulation software under limited interaction with the developer and/or domain expert. An application that simulates the self-assembly of ions near polarizable NPs by employing the unique features of this framework has been recently deployed on the nanoHUB cloud (Kadupitiya et al., 2018a). As is evident by the use of ML in numerous commercial platforms, scientific simulation workflow and software applications will increasingly employ an ML layer in the future. Understanding the integration of ML in scientific applications is thus critical; the work presented here contributes toward this goal.

2. Background and related work

2.1. Model and the energy functional

The problem of evaluating polarization effects in simulation of charged systems has been extensively explored by several research groups using different approaches (Allen et al., 2001; Barros et al., 2014; Boda et al., 2004; dos Santos et al., 2011; Gan and Xu, 2011; Jadhao et al., 2012; Marchi et al., 2001; Tyagi et al., 2010; Wynveen and Bresme, 2006). Explicit simulation of solvent (environment) and NPs is possible (Wynveen and Bresme, 2006) using advanced computational techniques such as fast multipole methods and local electrostatics algorithms (Rottler and Maggs, 2004); the role of solvent is also crucial in understanding properties of ion dynamics at the molecular scale (Beckstein et al., 2004). However, many phenomena cannot be suitably investigated using fully atomistic models due to the prohibitively large number of degrees of freedom associated with such systems. This has led to the study of coarse-grained models that treat ions explicitly but replace the molecular structure of the solvent and the NP with continuous dielectric environments. Systems where the different material parts are adequately captured by piecewise-uniform dielectric permittivities (e.g. NP and solvent, protein and cellular medium) have attracted particular attention (Allen et al., 2001; Barros and Luijten, 2014; Boda et al., 2004; dos Santos et al., 2011; Jadhao et al., 2012, 2013; Tyagi et al., 2010; Xu, 2013). For these model systems, solving for the induced charge density reduces the computational costs because the unknown induced charge density resides only on the two-dimensional interface (boundary) between the NP and the surrounding medium. We work with such a coarse-grained model.

The dynamical optimization framework for extracting ion distributions from simulations of the coarse-grained model that treats the solvent and NP as dielectric continua is based on the true energy functional of the induced charge density introduced in Jadhao et al. (2012):

$$\mathcal{F}[\omega] = \frac{1}{2} \int \int \rho_{\mathbf{r}} G_{\mathbf{r},\mathbf{r}'} (\rho_{\mathbf{r}'} + \Omega_{\mathbf{r}'}[\omega]) d\mathbf{r}' d\mathbf{r} - \frac{1}{2} \int \int \Omega_{\mathbf{r}}[\omega] G_{\mathbf{r},\mathbf{r}'} (\omega_{\mathbf{r}'} - \Omega_{\mathbf{r}'}[\omega]) d\mathbf{r}' d\mathbf{r}, \quad (1)$$

where ρ and ω are the ion and induced charge densities, respectively. The function $G(\mathbf{r}, \mathbf{r}') = |\mathbf{r} - \mathbf{r}'|^{-1}$ is the Green's function and Ω is given by

$$\Omega_{\mathbf{r}}[\omega] = \nabla \cdot \left(\chi_{\mathbf{r}} \nabla \int (\rho_{\mathbf{r}'} + \omega_{\mathbf{r}'}) d\mathbf{r}' \right), \quad (2)$$

where $\chi(\mathbf{r})$ is the dielectric susceptibility. $\chi(\mathbf{r})$ is related to the spatially varying dielectric permittivity $\varepsilon(\mathbf{r})$ via the relation $\varepsilon = 1 + 4\pi\chi$. The minimization of $\mathcal{F}[\omega]$ leads to the equation:

$$\omega = \Omega. \quad (3)$$

Solving this equation is equivalent to solving the Poisson equation; its solution produces the correct induced charge density (Jadhao et al., 2012, 2013). At its minimum, $\mathcal{F}[\omega]$ evaluates to the true electrostatic energy of the system. These features allow $\mathcal{F}[\omega]$ to be optimized dynamically as the ions move to their new positions in a simulation.

The functional $\mathcal{F}[\omega]$ can be transformed into a functional of only the surface (two-dimensional)-induced charge density for the case of polarizable NP in a solvent where the NP and the solvent are modeled as materials of different, but uniform, permittivities (Jadhao et al., 2012, 2013). The discretized form of this transformed functional obtained by meshing the NP surface into M finite elements is given as follows:

$$\mathcal{F}[\{\omega_k\}] = \frac{1}{2} \sum_{i=1}^N \sum_{j \neq i}^N q_i K_{\mathbf{r}_i, \mathbf{r}_j} q_j + \frac{1}{2} \sum_{i=1}^N \sum_{k=1}^M q_i K_{\mathbf{r}_i, \mathbf{s}_k} \omega_k a_k + \frac{1}{2} \sum_{k=1}^M \sum_{l=1}^M \omega_k K_{\mathbf{s}_k, \mathbf{s}_l} \omega_l a_k a_l \quad (4)$$

where ω_k , \mathbf{s}_k , and a_k are, respectively, the induced charge, position vector, and area associated with the k th finite element. Here, N is the total number of ions, and q_i and \mathbf{r}_i are the charge and position vector of the i th ion, respectively.

The terms K , K , and K in (4) are the effective potentials of interaction between two ions, between an ion and an induced charge, and between two induced charges; explicit expressions of these functions can be found in the original papers (Jadhao et al., 2012, 2013). $\mathcal{F}[\{\omega_k\}]$ can be minimized on-the-fly using MD methods that treat the induced charges on the surface as dynamic variables; the details of this dynamical optimization framework are provided in Section 3. We note that for the sake of brevity we have restricted the presentation of the above discretized form of the free-energy functional (equation (4)) to systems where the dynamics of ions is probed near a single dielectric interface. Extension of the approach to multiple dielectric surfaces corresponding to more complex systems (e.g.

multiple ion-containing oil nanodroplets in water or electrolyte ions in water confined by material surfaces of different permittivities) is straightforward; results from these investigations have appeared in previous papers (Jing et al., 2015; Shen et al., 2017).

2.2. Nanoscale simulation of ions near polarizable materials

We review the techniques of computing the ion distributions in systems described by the coarse-grained model of ions near the dielectric interface separating NP and solvent (Allen et al., 2001; Barros et al., 2014; Boda et al., 2004; Jadhao et al., 2012, 2013; Jing et al., 2015; Tyagi et al., 2010). Here, NP and solvent are characterized with different (but uniform) dielectric permittivities. In the light of the approach presented in Section 3, we focus on methods that are broadly applicable and are not limited by the choice of NP geometry or dielectric permittivity profile.

We first outline the methods based on variational approaches to the problem of evaluating the polarization effects as these techniques are most closely related to the work presented here. In this approach, one transforms the original problem of solving the Poisson differential equation into an optimization problem. A variety of functionals employing various electrostatic quantities as field variables have been proposed to formulate the variational optimization problem (Allen et al., 2001; Attard, 2003; Felderhof, 1977; Jackson, 1999; Lipparini et al., 2010; Marcus, 1956; Nakano et al., 1994; Reiner and Radke, 1990; Rottler and Maggs, 2004; Villasenor and Buneman, 1992; York and Karplus, 1999). Allen et al. (2001) performed an explicit (static) optimization of a functional of the induced surface charge density $\omega(\mathbf{s})$ at each MD step to solve the Poisson equation and propagate ions. Marchi et al. (2001) worked with a true energy functional of the polarization vector and implemented a dynamical optimization framework to propagate ion dynamics in parallel with the evaluation of polarization vector fields. However, the choice of the polarization vector as the variable field needed a three-dimensional specification leading to increased computational costs that can be avoided by choosing the induced charge density $\omega(\mathbf{s})$ as the variational field.

Another class of methods for computing $\omega(\mathbf{s})$ transform the problem into a matrix formulation (Barros et al., 2014; Boda et al., 2004; Tyagi et al., 2010). The induced charge computation methods (Boda et al., 2004) use matrix inversions to solve for $\omega(\mathbf{s})$. Matrix inversion operations involve $O(M^3)$ calculations where M is the number of surface mesh elements. Techniques to improve upon this scaling have been subsequently developed (Tyagi et al., 2010). Alternatively, iterative methods to solve the matrix equation have been proposed (Barros and Luitjen, 2014). In particular, the generalized minimum residual method solves the matrix equation without explicitly constructing the inverse matrix and yields a converged result for $\omega(\mathbf{s})$ at each simulation timestep within four to five iterations (Gan et al., 2015).

The evaluation of $\omega(\mathbf{s})$ in all the above approaches requires the ionic configuration to be static at each simulation step to guarantee the overall stability of the simulation. In Section 3, we present the details of a recently developed dynamical optimization framework that enables the simultaneous (on-the-fly) updates of $\omega(\mathbf{s})$ and the ionic configuration in the same simulation step (Jadhao et al., 2012, 2013).

2.3. Parameter prediction using ML

ML abstractions for parameter prediction and tuning have been extensively employed in the performance enhancement of bigdata or deep learning frameworks. Denil et al. (2013) used artificial neural network (ANN) and convolutional deep learning neural network to predict the parameters found in image classification tasks. The ANN was able to obtain an accuracy of 95%. Yigitbasi et al. (2013) employed ML-based auto-tuning for diverse MapReduce applications and cluster configurations in Hadoop framework. Their work showed that support vector regression (SVR) exhibits good accuracy while being computationally efficient for performance modeling of MapReduce applications.

Regression-based prediction schemes have been employed in different domain areas (Balachandran et al., 2016; Chen and Yu, 2014; Cherkassky and Ma, 2004; Eng et al., 2014; Kazemi and Sullivan, 2014; Quan et al., 2014; Yadav et al., 2016). Eng et al. (2014) used random forest regression algorithm to predict host tropism of influenza A virus proteins with an accuracy above 96%. Similarly, ensemble of regression trees were employed to perform face alignment for real-time applications (in 1 ms) by Kazemi and Sullivan (2014). SVR has been used for wind speed prediction by Chen and Yu (2014). ANN-based regression has been studied by Quan et al. (2014) to yield short-term load prediction of electrical power systems based on wind power forecasting. Yadav et al. (2016) have employed ANN-based regression for forecasting solar radiation.

In recent years, ML methods have been applied to enhance computational techniques aimed at understanding material phenomena; ML has been used to predict parameters, generate configurations in material simulations, and classify material properties (Behler, 2016; Botu and Ramprasad, 2015; Ch'ng et al., 2017; Ferguson, 2017; Fox et al., 2019a, 2019b; Guo et al., 2018; Kadupitige et al., 2019; Kadupitiya et al., 2018b, 2019; Liu et al., 2017; Long et al., 2015; Morningstar and Melko, 2017; Schoenholz, 2018; Spellings and Glotzer, 2018; Ward et al., 2016). Spellings and Glotzer (2018) applied a simple feedforward ANN to discover interesting areas of parameter space corresponding to crystal formation in the self-assembly of colloidal building blocks. Botu and Ramprasad (2015) employed kernel ridge regression to accelerate the ab initio MD method for nuclei–electron systems by learning the selection of probable configurations in MD simulations.

Liu et al. (2017) employed an ANN to select efficient updates for Monte Carlo simulations of classical Ising spin models. Balachandran et al. (2016) have used SVR to create an adaptive ML model to aid the design of new materials with desired elastic properties and enhanced long-term performance using minimum number of iterations.

These explorations have inspired us to use ML to design an adaptive MD-based dynamical optimization framework that updates the simulation timestep and auto-tunes the virtual parameters characterizing the dynamics of ions near polarizable NPs to yield a more stable and efficient simulation. Related work in the area of adapting timestep in a simulation has involved using analytical approaches to multiple timestep integration (Luehr et al., 2014; Tuckerman et al., 1992). Recent work has also focused on adaptive ensemble simulations to enhance the computational efficiency of biomolecular simulations (Kasson and Jha, 2018). We also note the development of auto-tuning technology for high-performance computing applications to reduce execution time and enhance programmer productivity (Whaley and Dongarra, 1998). Here, auto-tuning relates to the automatic generation of a search space of possible kernels for a computational task to identify the best possible kernel, with recent work involving the use of ML-based approaches for identifying the search space (Balaprakash et al., 2018). In Section 4, we describe the results of our experiments with different regression-based ML models to identify and tune optimal simulation parameters in MD simulations based on the dynamical optimization framework.

3. Dynamical optimization framework for simulating ions near polarizable NPs

In this section, we provide the details of the dynamical optimization framework for simulating ions in the presence of polarizable NPs. This framework uses CPMD technique (Car and Parrinello, 1985; Fois et al., 1993) to dynamically optimize an energy functional of the polarization charge density that results in the propagation of the ionic configuration in tandem with an accurate update of the polarization charges (Jadhao et al., 2012, 2013). These details will help clarify the use of the ML-based enhancement strategies outlined in Section 4.

3.1. Extended Lagrangian, equations of motion, and CPMD simulation

To implement the dynamical optimization of $\mathcal{F}[\{\omega_k\}]$, the induced charges $\{\omega_k\}$ are treated as dynamic virtual variables. A fictitious kinetic energy is associated with this virtual system:

$$\mathcal{K} = \sum_{k=1}^M \frac{1}{2} \mu_k \dot{\omega}_k^2, \quad (5)$$

where μ_k is the mass of the k th virtual variable ω_k , and M is the number of mesh points discretizing the NP surface. The extended Lagrangian \mathcal{L} with $\mathcal{F}[\{\omega_k\}]$ as its electrostatic potential energy is constructed by including \mathcal{K} as an additional term:

$$\mathcal{L} = \mathcal{K} + \sum_{i=1}^N \frac{1}{2} m_i \dot{\mathbf{r}}_i^2 - \mathcal{F}[\{\omega_k\}] - \mathcal{H}[\{\mathbf{r}_i\}]. \quad (6)$$

In (6), the second term is the usual total kinetic energy associated with N ions (physical system), with m_i being the mass of the i th ion. The final term contains a set of Lennard–Jones (LJ) potentials to model the ion–ion and ion–NP steric interactions. Note that $\mathcal{F}[\{\omega_k\}] \equiv \mathcal{F}[\{\omega_k\}, \{\mathbf{r}_i\}]$ is also a function of the set of ion positions $\{\mathbf{r}_i\}$.

The Lagrangian \mathcal{L} yields the following Euler–Lagrange equations of motion:

$$m_i \ddot{\mathbf{r}}_i = -\nabla_{\mathbf{r}_i} \mathcal{F}[\{\omega_k\}, \{\mathbf{r}_i\}], \quad (7)$$

$$\mu_k \ddot{\omega}_k = -\nabla_{\omega_k} \mathcal{F}[\{\omega_k\}, \{\mathbf{r}_i\}], \quad (8)$$

for the i th ion and the k th induced charge, respectively. These equations are used to evolve the induced charge configuration *on the fly* using the CPMD method. Following (7), each ion is moved by the force $-\nabla_{\mathbf{r}_i} \mathcal{F}[\{\omega_k\}, \mathbf{r}_i]$ in a timestep Δ during which each induced charge is updated via the force $-\nabla_{\omega_k} \mathcal{F}[\{\omega_k\}, \mathbf{r}_i]$ following (8). To simulate the behavior of the ions at temperature T , the extended system of ions and virtual variables is coupled to a set of Nosé–Hoover thermostats (this coupling modifies the equations of motion (7) and (8) similar to a canonical MD routine). This two-temperature approach is a standard feature of CPMD (Blöchl and Parrinello, 1992; Fois et al., 1993; Sprik, 1991). The ions couple to a thermostat at temperature T , while the virtual system is coupled to one at T_v .

Velocity-Verlet algorithm is used to generate the dynamics of the extended system. The dynamics is associated with a conserved quantity, the total energy of the extended system:

$$\mathcal{E} = \sum_{i=1}^N \frac{1}{2} m_i \dot{\mathbf{r}}_i^2 + \mathcal{K} + \mathcal{F}[\{\omega_k\}] + \mathcal{H}[\{\mathbf{r}_i\}] + \mathcal{T} + \mathcal{T}_v. \quad (9)$$

Here, \mathcal{T} and \mathcal{T}_v are the energy terms associated with the thermostats controlling the temperature of the physical and virtual systems, respectively. The extended energy and \mathcal{K} are monitored at periodic intervals during a CPMD simulation to assess the stability and accuracy of the simulation.

Virtual masses μ_k are chosen to be proportional to the areas of the mesh points. The value of the proportionality constant μ depends on the attributes of the physical system (e.g. NP charge, dielectric profile, ion valencies) as well as the simulation timestep Δ . The parameters T_v and μ are optimized to ensure the stability and accuracy of the simulation (see Section 3.3); further technical details of the method can be found in Jadhao et al. (2012).

3.2. OpenMP/MPI hybrid parallelization

A system with N ions near an unpolarizable NP effectively translates into a system with M additional dynamical variables in the case of a polarizable NP within the dynamical optimization framework. Due to the long-range nature of the electrostatic interactions, the associated computational costs scale roughly as $O((N + M)^2)$, with a prefactor that can be large owing to the complexity of the terms involved in the expressions for forces derived from $\mathcal{F}[\{\omega_k\}]$. Indeed, performance profiling report generated using Performance Counters for Linux showed that the sequential program spends the largest amount of computation time (64.33% of the total) calculating the forces between the ions for each step of the simulation. To reduce the computing time associated with the evaluation of these forces and enhance the performance of the simulation framework, a hybrid OpenMP/MPI parallel programming model is adopted. The hybrid model has advantages over pure MPI or pure OpenMP, when cache performance is taken into consideration. This strategy provides nonuniform memory access traffic and internode communication (Rabenseifner et al., 2009) to support maximum access locality and minimum number of cache misses. We note that the simulations associated with the dynamical optimization framework presented in the previous publications employed the OpenMP (shared memory) parallelization model and consequently, were limited in their application scope (Jadhao et al., 2012, 2013; Jing et al., 2015).

The hybrid masteronly model is implemented by combining the distributed memory MPI approach and the shared memory OpenMP approach (Rabenseifner et al., 2009) and is applied to the force and energy calculation subroutines in the dynamical optimization framework. The model uses one MPI process per node and OpenMP on the cores of the node, with no MPI calls inside the parallel regions. The domain decomposition is enabled under a two-level mechanism. On the MPI level, a coarse-grained domain decomposition is performed using boundary conditions as explained in Figure 1. The second level of domain decomposition is achieved through OpenMP loop level parallelization inside each MPI process.

3.3. Key framework features

We identify two key features of the dynamical optimization framework that encode the accuracy and stability of the simulation and guide the process of designing the ML-based enhancement strategies presented in Section 4. The first key feature is the conservation of the extended energy \mathcal{E} , given by equation (9), and the approximate conservation of the energy of the physical system that is captured by demanding that the kinetic energy of the virtual system nearly vanishes:

$$\mathcal{K} \approx 0. \quad (10)$$

In other words, the framework ensures that the physical system remains unaffected as much as possible by the presence of the virtual system.

The energy profiles of a typical, successful CPMD simulation of ions near a polarizable, spherical NP at room temperature are shown in Figure 2: the extended energy \mathcal{E} is constant and the total virtual kinetic energy \mathcal{K} stays stable and close to 0 throughout the entire simulation (for ≈ 10 ns). In practice, this feature is incorporated in the simulation by appropriately choosing values of simulation timestep Δ , virtual variable mass μ , and the virtual system temperature $T_v \ll T$ ($T_v \approx 0$). These parameters are selected to control large, abrupt rise in the kinetic energy associated with the virtual system as the simulation progresses.

This feature is encoded in the quantity R which measures the ratio of the fluctuations in \mathcal{E} and the fluctuations in the kinetic energy of the physical system. R determines the stability and the accuracy of the simulation. For good energy conservation (constant \mathcal{E}), it is demanded that the simulations satisfy the condition $R < 0.05$ as noted in the literature (Marchi et al., 2001). The latter inequality implicitly satisfies the requirement that \mathcal{K} is kept close to the value dictated by the low temperature T_v .

The second important feature considers the effectiveness of the framework to reproduce the induced charge distribution accurately at each simulation step. At regular intervals during the course of the simulation, the ion coordinates and induced charge densities on the NP surface are stored. Then, an ordinary (static) minimization of the functional \mathcal{F} is carried out to explicitly determine the (numerically) exact induced charge density. The tracking of the induced charge density distributions on the NP surface can be assessed by evaluating the matching of \mathcal{F} optimized on the fly with the electrostatic energy value obtained by optimizing the functional explicitly (Figure 3). This functional matching is the second key feature. In practice, we compute the functional deviation, f_d , which measures the average difference between the dynamically optimized functional \mathcal{F} and the energy functional obtained via direct (static) minimization. To pass the test of stability and accuracy, we enforce $|f_d| < 1\%$.

R and f_d are central to the success of the simulations based on this framework and determine the associated “good” virtual system parameters μ and T_v . In general, higher μ leads to better energy conservation and lower R , while lower T_v keeps the virtual system from excessive heating and generates lower f_d . Having just these two features biases the prediction of μ (T_v) toward higher (lower) values for a system of ions characterized with a generic input parameter pattern. Very high values of μ and/or very low values of T_v can affect the overall stability of the simulation as the virtual system can be prohibitively slow (due to the “heavy” virtual masses and “cooler” associated temperature) to react to the evolving ionic configuration, resulting in inaccurate induced charge updates. An experienced domain expert would typically avoid these choices.

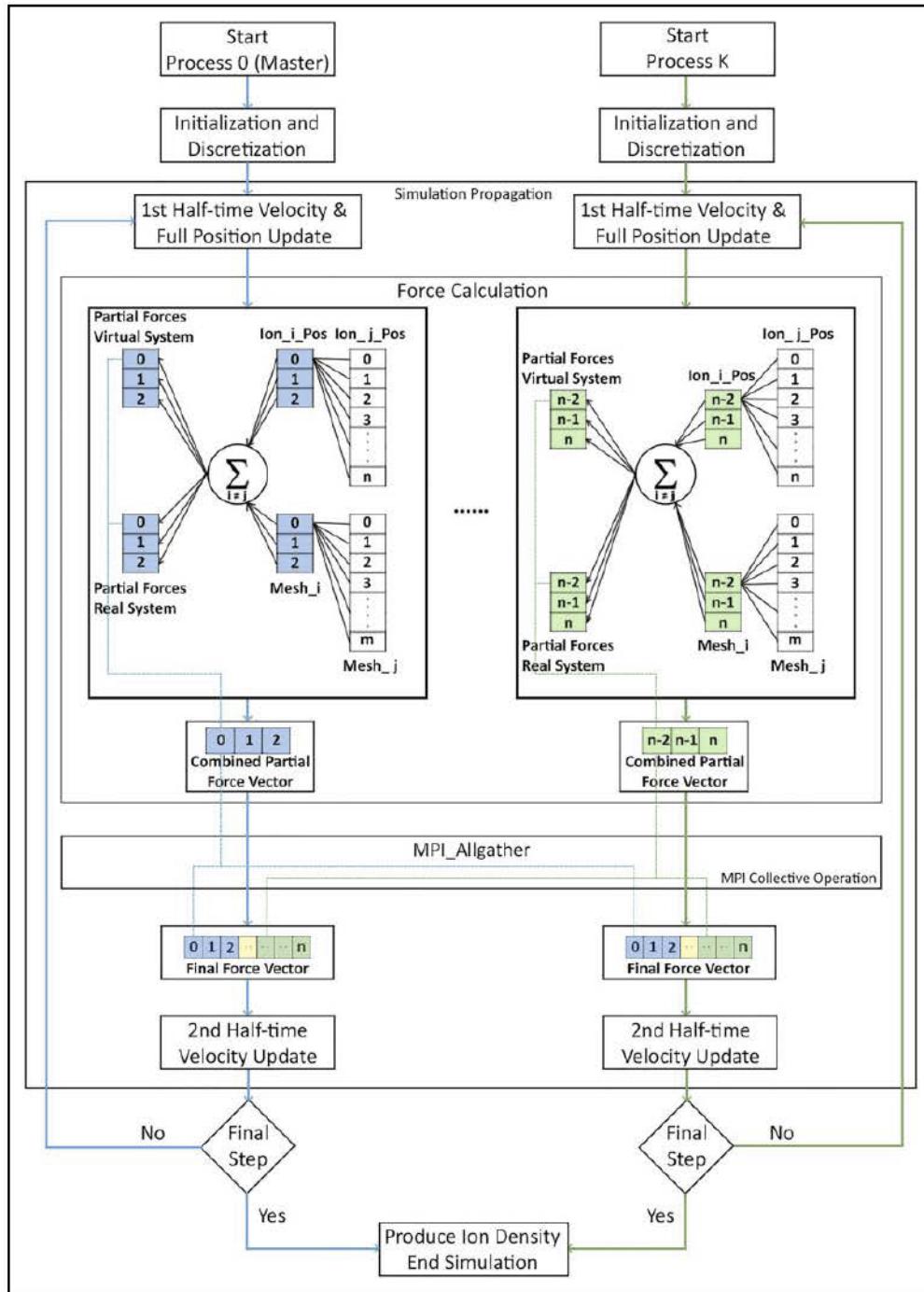


Figure 1. Hybrid model (employed inside the Force Calculation block) with distributed and shared memory parallelization techniques implemented using MPI and OpenMP.

To enhance the stability of the simulation and bias the selection of the virtual parameters toward those picked by a domain expert, another quantity, R_v , is introduced. R_v is the ratio of the fluctuations in \mathcal{E} and the fluctuations in the kinetic energy of the virtual system \mathcal{K} . Lower R_v implies that fluctuations in \mathcal{K} (that can arise from lower μ or higher T_v) are sufficiently strong to endow the virtual system with the necessary dynamics to adapt to the evolving ionic system. Unlike R and f_d that exhibit universal

bounds informed by the physical dynamics of ions, the bound on R_v depends on the set of systems investigated and is informed largely by past domain experience. For counterion-only systems used as training set for the ML-based methods, $R_v < 0.15$ is enforced. For systems characterized with electrolytes that are expected to exhibit a greater number of ions (with both positive and negative valencies), R_v can assume larger values depending on the salt concentration.

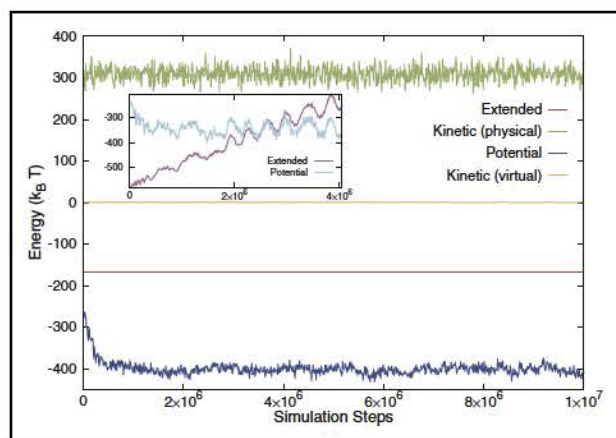


Figure 2. Energy profiles of 206 ions near a polarizable NP (whose surface is meshed with 1082 grid points). (Outset) Data are shown for ≈ 10 ns of simulated physical time (10 million simulation steps). Conservation of the extended (total) energy and nearly vanishing (≈ 0) virtual kinetic energy enabled by the selection of “good” virtual parameters ($\mu = 15$, $T_v = 0.001$) highlight the first key feature of the dynamical optimization framework. (Inset) Energy profiles for the same system when a set of “bad” virtual parameters are selected ($\mu = 6$, $T_v = 0.05$). This selection produces unphysical dynamics of the ionic system with the extended energy not getting conserved and the potential energy exhibiting artificial oscillations.

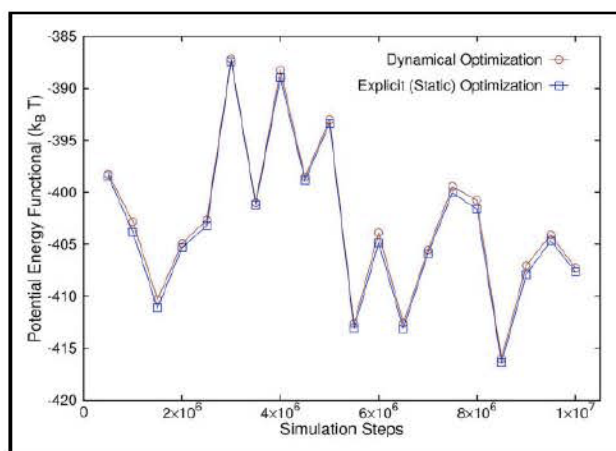


Figure 3. Comparison of the functional optimized dynamically (circles) and the functional optimized at regular intervals keeping the ionic configuration static during the optimization process (squares). Results are shown for the same system as in Figure 2. The matching of the two functionals illustrates the second key feature of the framework: the accurate tracking of the induced charge density.

Quantities R , f_d , and R_v determine the choice of the optimal virtual parameters. However, these quantities and the success of the simulation depend critically on another important parameter: the simulation timestep Δ . The simulation timestep in the CPMD-based dynamical optimization framework depends on both the physical and the virtual system parameters, the latter being unknown a priori. Conversely, the optimal values of virtual parameters (μ , T_v) that

ensure a longtime stable simulation are dependent on Δ . This complicates the process of choosing a reliable, yet efficient, value for Δ , μ , and T_v , and one typically chooses a conservative Δ that is smaller than the value used in conventional MD simulations ($\Delta \approx 1 - 5$ fs for an MD simulation of monovalent electrolyte ions in water at room temperature).

4. ML-based enhancement of the dynamical optimization framework

We now describe an ML-based procedure to increase the efficiency and improve the stability of the dynamical optimization framework while retaining the simulation accuracy. We present an ML technique that uses the aforementioned key features encoded in quantities R , f_d , and R_v to enhance the performance of the dynamical optimization framework by (1) predicting and auto-tuning the optimal virtual parameters μ and T_v and (2) adapting the timestep to the largest allowable value during the simulation. The ML technique is combined with OpenMP/MPI hybrid parallel programming model, described in Section 3, to carry out the simulation.

Figure 4 shows the overview of the enhanced framework. ML-based parameter prediction was implemented using two ML models (ML models I and II). First, the ion and NP model attributes, as well as the initial timestep $\Delta = \Delta t_0$, were fed to ML model 1 to predict the initial virtual system parameters μ and T_v . The predicted parameters and Δt_0 were used to start the simulation that was parallelized using the OpenMP/MPI hybrid programming model. At intermediate times t_n during the simulation, ML model II was used to predict the new timestep Δt_{n+1} and associated virtual parameters that continue the simulation for the subsequent time block (t_n, t_{n+1}) . The ion distributions near the polarizable NP were sampled during the simulation run and the ion densities were stored as simulation output. ML model II also checked if the simulation was successful up to t_n before dynamically tuning the parameters for the next iteration. The program aborted and called the error handler to display appropriate error messages if the simulation failed due to the imposed R , f_d , and R_v criteria. In addition, during the simulation, the quantities R , f_d , and R_v were computed and saved as output for retraining both ML models after a set number of simulation runs were executed. For every 1000 new simulation runs, both models were retrained.

After reviewing and experimenting with many ML techniques for parameter tuning and prediction including polynomial regression, SVR, decision tree regression, and random forest regression (Sections 2.3 and 5.1), the ANN was adopted for enhancing the dynamical optimization framework. Figure 5 shows the details of the ANN-based ML model II employed to predict the virtual system parameters and the adaptive timestep for simulations based on the framework; the ANN was trained to select the largest allowable Δ that satisfies the tests of stability and accuracy

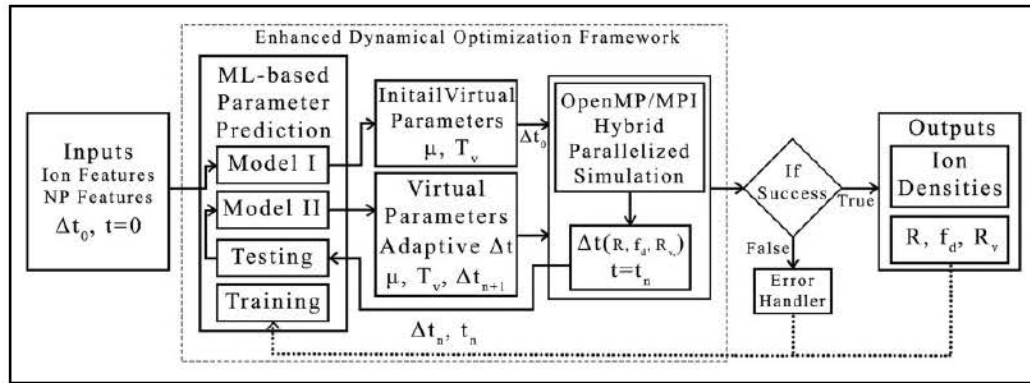


Figure 4. System overview of the ML-enhanced dynamical optimization framework.

encoded in the ML features. ML model I exhibits a similar process but is trained without the time and timestep parameters. The data preparation and preprocessing techniques, feature extraction and regression techniques as well as their validation for both models are discussed below.

4.1. Data preparation and preprocessing

Counterion-only systems (no added electrolyte) were considered for generating the training set. The polarization effects are expected to be the strongest in these systems as added electrolyte screens the ion–NP electrostatic interactions. Further, counterion-only systems are relatively smaller and enable a broader exploration of parameter space to train the ML models. Interestingly, as we discuss later, results from counterion-only systems were employed to successfully extract and infer ionic distributions associated with electrolyte systems for up to $O(0.1)$ M concentration, exhibiting the transfer learning aspects of the ML-based procedure employed.

Prior domain experience and backward elimination using the adjusted R^2 was considered for creating the training data set. Using this process, five input parameters that significantly affect the polarization charges on the NP surface were identified: NP permittivity ϵ_{NP} , solvent permittivity ϵ_S , NP charge Q (in units of electronic charge $|e|$), counterion valency ν , and NP mesh size M . While the temperature of the physical system and the size of the NP and ions affect the polarization charges and associated ionic distributions, in this initial study, these were considered fixed; NP diameter was taken to be ≈ 7.5 times the ion diameter (0.357 nm), and temperature was fixed at 298 K. Despite the aforementioned potential for transfer learning associated with the ML procedure, additional parameters such as salt concentration and co-ion attributes should be included in the training set to predict the optimal parameters associated with the simulation of electrolyte systems with higher accuracy; future work will explore the training with these additional input parameters. Further, to enable accurate prediction of virtual parameters and generation of stable ion dynamics near NPs of different shapes (e.g. discs, rods) or NPs present in media characterized with

a more complex permittivity profile (e.g. NPs stabilized near an oil–water interface), the training data must include NP shape characteristics and the relevant permittivity profile characterizing the system. Virtual parameter mass μ and virtual system temperature T_v were selected as the output parameters. Few discrete values for each of the input/output parameters were experimented with and swept over to create and run 13,600 simulations for training the ML model I. The range for different ionic system parameters was selected based on physically meaningful and experimentally relevant values: $\epsilon_{NP} \in (2, 160)$; $\epsilon_S \in (2, 160)$; $Q \in (-20, -100)$; $\nu \in 1, 2, 3$. For the mesh size and the virtual system parameters, the range was chosen based on previous trial and error procedure: $M \in (132, 1692)$, μ was swept from 1 to 40 using random discrete values to cover the range, and T_v was swept from 0.001 to 0.005. All simulations were performed for ≈ 1 ns.

To support on-the-fly tuning of Δ and associated selection of μ, T_v during the simulation, ML model II was trained with two additional parameters. Simulation time $t \equiv t_n$ and timestep $\Delta \equiv \Delta t_{n \rightarrow n+1}$ were added as input and output parameters, respectively, to the system parameters explored in ML model I. Twenty discrete simulation time values $t_n \approx 0.1, 0.2, \dots, 2$ ns and four discrete timestep values $\Delta = 0.001, 0.002, 0.003, 0.004$ were swept to generate 54,400 simulation configurations. In our current approach, the time interval (and related frequency) between updates of the simulation timestep and virtual parameters is $\delta = 0.1$ ns. This update frequency is largely limited presently by the computing costs associated with the size of the data set needed to compute reliable estimates of R, R_v , and f_d at δ time intervals. A higher frequency (lower δ) will require a larger training data set to be generated, adding to the computational costs. Further, for good stability profile, maximum update frequency should be smaller than characteristic timescales of fluctuations in the induced charge density (indicative of changes in the virtual system energy) in response to the dynamics of ions near polarizable surfaces.

As described in Section 3.3., R, f_d , and R_v encode the key features of energy conservation and accurate tracking of the induced charges that measure the success of the dynamical optimization framework. Acceptable threshold values were

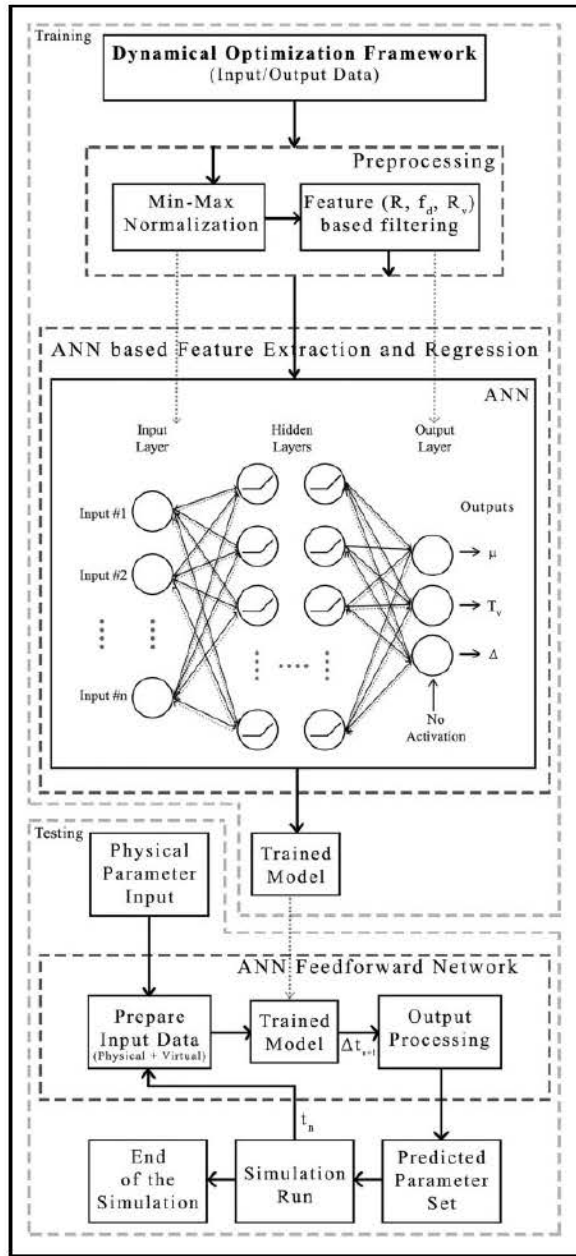


Figure 5. ML procedure (model II) for determining the parameters of the virtual system and the adaptive timestep used in the dynamical optimization framework.

identified for R , f_d , and R_v as 0.05, 1%, and 0.15, respectively, for the range of systems included in the training set. These quantities were treated as output features to filter the data sets to only keep the input parameter configurations resulting in successful simulation runs. From the data set for initial parameter prediction, 4530 input/output configurations were selected as successful. From the data set for adaptive timestep prediction, 15,640 input/output configurations were selected based on the same R , f_d , and R_v criteria. Each of these data sets were separated as training and testing using a ratio of 0.7:0.3. Min-max normalization filter was applied to normalize the input data in the data preprocessing stage.

Table 1. Comparison of regression models for the prediction of initial virtual parameters.

Model	Testing sets		Random sets
	Success %	MSE	Success %
Polynomial	45.7	12.25	16.5
Support vector	78.3	4.56	58.4
Decision tree	70.4	6.93	48.1
Random forest	75.3	4.88	55.1
ANN-based	94.3	0.56	87.2

MSE: mean square error; ANN: artificial neural network.

4.2. Feature extraction and regression

The ANN algorithm with two hidden layers (Figure 5) was implemented in Python for regression of two continuous variables in ML model I and for regression of three continuous variables in ML model II. In both models, outputs of the hidden layers were wrapped with the *relu* function; the latter was found to converge faster compared to the sigmoid function. No wrapping functions were used in the output layers of the algorithm.

By performing a grid search, hyper-parameters such as the number of first hidden layer units, second hidden layer units, batch size, and number of epochs were optimized to 13, 8, 25, and 150, respectively. Adam optimizer was used as the back propagation algorithm. The weights in the hidden layers and in the output layer were initialized to random values using a normal distribution at the beginning. The mean square loss function was used for error calculation in both ML models. We use dropout regularization mechanism (where hidden layer neurons are turned off at a dropout rate of 0.2) to prevent overfitting which was evaluated using the k-fold cross-validation technique. ANN implementation, training and testing was programmed with the aid of keras and sklearn ML libraries (Abadi et al., 2016; Buitinck et al., 2013; Chollet et al., 2015).

5. Results and discussion

5.1. Initial virtual parameter prediction

Several regression models were implemented and tested to predict the initial virtual parameters μ and T_v . These models were tested on 1359 input parameter sets comprising of values within the range for which the models were trained for. Additionally, the models were tested on 450 completely random input sets of parameters, which included some selections beyond the range of training data set. Table 1 shows the success rate and mean square error (MSE) for testing data sets as well as random input data sets. Success rate was calculated based on R , f_d , and R_v . Reported MSE values are calculated using k-fold cross-validation techniques with $k = 10$. ANN-based regression model predicted the initial virtual system parameters correctly with 94.3% success rate (MSE of 0.56), outperforming other nonlinear regression models as evident from

Table 2. Predicted parameters by ML model I, and simulation accuracy and stability.

Inputs	Prediction		Results		
	μ	T_v	R	f_d	R_v
$\varepsilon_{NP}, \varepsilon_S, Q, v$					
2, 10, -60, 1	9	0.001	0.001	-0.01	0.12
2, 78.5, -30, 3	7	0.002	0.003	-0.6	0.13
50, 78.5, -60, 2	18	0.001	0.001	-0.4	0.06
80, 160, -90, 3	30	0.002	0.002	-0.7	0.09
100, 120, 30, 2	36	0.005	0.002	-0.1	0.10
5, 71, -24, 2	42	0.001	0.007	-0.6	0.13
44, 37, -114, 1	38	0.006	0.002	-0.11	0.11
30, 35, -108, 3	43	0.007	0.002	-0.12	0.05
15, 78, -102, 1	17	0.025	0.005	-0.27	0.11

ML: machine learning.

Table 1. ANN-based regression model was also able to achieve a success rate of 87.2% on completely random input parameters. This ANN-based regression model was adopted as ML model I.

Table 2 shows the predicted μ and T_v for selected systems along with the quantities R , f_d , and R_v that characterize the key features of the framework: energy conservation and tracking of the induced charges. The predicted μ and T_v values produced stable and accurate dynamics of ions near polarizable NP as evidenced by the values of R , f_d , and R_v that lie within the allowed ranges ($R < 0.05$, $|f_d| < 1\%$, $R_v < 0.15$).

We note that when the ANN was trained utilizing only R and f_d quantities, higher μ and lower T_v values were predicted as expected from the arguments presented in Section 3.3. These virtual parameter choices are not optimal for the stability of the simulation and will not be picked by an experienced, domain expert. Inclusion of R_v as another feature for training the model enabled the ANN to predict the virtual system parameters that were likely to enhance the stability of the simulation and be selected by an expert.

5.2. Auto-tuning CPMD simulation parameters

Similar to ML model I, ML model II employed the ANN-based regression model trained with two additional parameters: simulation timestep Δ and time t . This model was trained to infer the largest allowed Δ and auto-update the associated optimal virtual parameters μ, T_v at t_n for the simulation during the interval (t_n, t_{n+1}) based on the ML output features R, f_d , and R_v in the time interval (t_n, t_{n+1}) from the training data.

Figure 6 illustrates the computational gains resulting from the auto-tuning of Δ using ML model II for systems of counterions near an NP. NPs of different dielectric permittivities (2, 40, and 60) in water (dielectric permittivity 78.5) are considered. The effect of varying ion valency ($v = 1, 2$) is also probed. Other input parameters are NP charge $Q = -100$, NP mesh size $M = 1272$, and $\varepsilon_S = 78.5$. Symbols indicate the gains associated with the

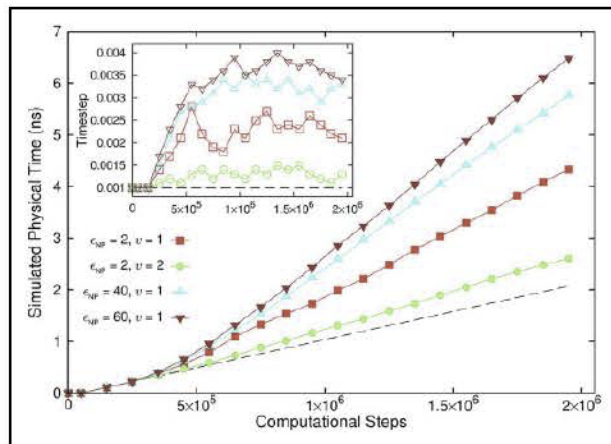


Figure 6. Simulated physical time t associated with the dynamics of ions as a function of computational steps S for a system of counterions near a polarizable NP. The legend denotes NP permittivity and ion valency ($\varepsilon_{NP}; v$). Symbols are the results from using ML-enabled tuning of simulation timestep Δ and the dashed line is the result for the nonadaptive case. (Inset) Δ in units of ≈ 1.09 ps. Open symbols correspond to the systems denoted with the closed symbols in the legend; black dashed line denotes $\Delta = 0.001$ associated with the nonadaptive case. Δ values represent the average timestep over a period of $t_n \rightarrow t_{n+1}$.

Table 3. Performance comparison of different ion–NP systems simulated for 2 million steps (≈ 20 h walltime) on 4 MPI nodes each with 16 OMP threads.

Physical system	Time (ns)	Speedup	Stability
Nonadaptive			
2, 78.5, -100, 1	2.18	1.00 \times	0.01552
ML-based adaptive			
2, 78.5, -100, 1	4.56	2.09 \times	0.00048
2, 78.5, -100, 2	2.75	1.26 \times	0.00088
40, 78.5, -100, 1	6.12	2.81 \times	0.00076
60, 78.5, -100, 1	6.85	3.14 \times	0.00063

NP: nanoparticle.

enhanced framework with adaptive timestep. The dashed line is the result from the nonadaptive simulation with static timestep for $\varepsilon_{NP} = 2, v = 1$ case (other systems also yield the same result when nonadaptive model is used). Compared to the simulation with nonadaptive Δ (performed using only ML model I), the auto-tuning of Δ extended the simulation of the ionic system to a longer physical time for the same number of simulation steps; a speedup of $\approx 1.25 - 3$ was observed depending on the system configuration, also see Table 3. For the system with $\varepsilon_{NP} = 2, v = 1$, the auto-tuning yielded a total simulated physical time of 4.5 ns compared to the 2.2 ns obtained with nonadaptive simulation. Figure 6 (inset) shows the variation in Δ as a function of the computational steps for the same systems. The tuning of Δ changed with the attributes of the ions and NP. Generally, longer Δ values (and associated higher speedup) were obtained for systems exhibiting a weaker dielectric contrast.

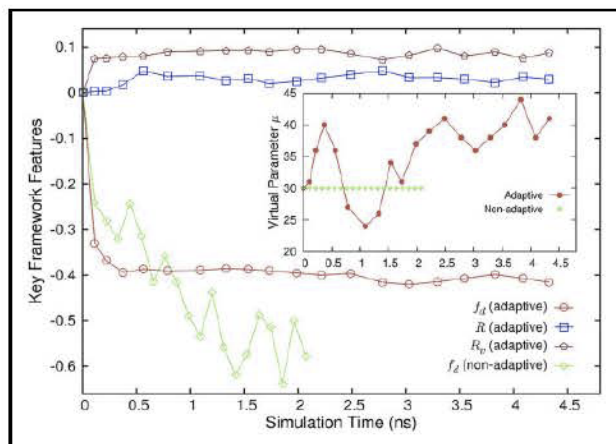


Figure 7. Key output features R , f_d , R_v as a function of the simulated physical time t for the counterion–NP system characterized with NP charge $Q = -100$, ion valency $\nu = 1$, NP permittivity $\epsilon_{NP} = 2$, solvent permittivity $\epsilon_S = 78.5$, and mesh size $M = 1272$. ML-enabled auto-tuning of timestep Δ and virtual parameters produces enhanced stability (diminished fluctuations in f_d) compared to the nonadaptive case (stronger fluctuations in f_d). (Inset) ML-enabled auto-tuning of the virtual parameter μ for the same system (closed circles) and fixed μ for the nonadaptive case (closed diamonds).

Table 3 quantifies the speedup by showing the performance comparison of the aforementioned ion–NP systems for 2 million steps on 4 MPI nodes each with 16 OpenMP threads and a fixed walltime of ≈ 20 h. ML-based adaptive timestep tuning enabled the simulation of the system for a longer physical time compared to the nonadaptive case under fixed compute resources and walltime. For the nonadaptive case, the virtual parameters $\mu = 30$ and $T_v = 0.002$ were selected using ML model I. A maximum speedup of 3.15 was achieved for the ion–NP system defined with the input parameters: $\epsilon_{NP} = 60$, $\epsilon_S = 78.5$, $Q = -100$, $\nu = 1$, and $M = 1272$, without adjusting any MPI or OpenMP parameters.

The ML-enhanced framework with adaptive timestep also improved the overall stability of the MD simulation. Figure 7 shows the key output features associated with the simulation of the ion–NP system characterized with $\epsilon_{NP} = 2$, $\nu = 1$ (other input parameters being the same as in Figure 6). Fluctuations in the output feature f_d , which measures the deviation of the on-the-fly optimized functional from the statically optimized functional, illustrate the stability of the simulation. Auto-tuning of parameters produced diminished fluctuations in f_d compared to the nonadaptive framework. Figure 7 (inset) shows the variation of virtual parameter (μ) with simulated physical time for the same ion–NP system. By definition, the nonadaptive model produced constant μ . On the other hand, the adaptive model produced the auto-tuning of μ which was correlated with the more stable dynamics (red circles characterizing f_d in the outset of Figure 7). Indeed, the variance in f_d data for the nonadaptive model ($f_{d|\sigma^2} = 0.01552$) was much higher than that for the adaptive model ($f_{d|\sigma^2} = 0.00048$). Table 3

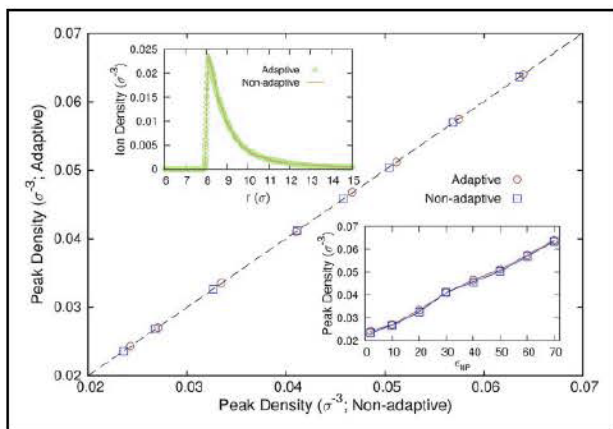


Figure 8. Correlation between the peak densities associated with the distribution of counterions for ion–NP systems characterized by different NP permittivities ϵ_{NP} ; other parameters are the same as listed in the caption of Figure 7. Blue squares are values from the nonadaptive simulation, red circles are results from the ML-enabled adaptive simulation. (Top-left inset) Density distribution of counterions for the system with $\epsilon_{NP} = 2$; symbols are the results from the adaptive model, line corresponds to the nonadaptive case. (Bottom-right inset) Peak densities from the two models as a function of ϵ_{NP} .

shows the variance of f_d for the same ion–NP systems analyzed above; in all cases, the variance was found to be significantly smaller by over an order of magnitude (indicating higher stability) for the adaptive model compared to the nonadaptive case. This enhanced stability can be attributed to the optimal updates of parameters μ , T_v during the intermediate times of the simulation.

In addition to increasing the efficiency and stability of the simulation, the framework with ML-enabled auto-tuning of parameters (adaptive framework) retained the accuracy associated with the framework using nonadaptive timestep and virtual parameters (nonadaptive framework). The accuracy can be assessed by comparing the density profiles of the ions computed using the two approaches. For different ϵ_{NP} values (other input parameters same as above), the peak densities computed using simulations based on adaptive framework were found to be in agreement with those calculated using the nonadaptive framework as shown in Figure 8; data from either approach fall on the dashed line which indicates linear correlation. Top-left inset of Figure 8 shows the variation of the counterion density (in σ^{-3} , where $\sigma = 0.357$ nm is the ion diameter) as a function of the distance from the NP (of radius 7.5σ) for the specific case of $\epsilon_{NP} = 2$. The density profiles extracted from the adaptive and nonadaptive frameworks were found to be in good agreement (relative error in either distributions was found to be $\approx 1\%$). Bottom-right inset of Figure 8 shows the variation of the peak density of counterions as a function of the dielectric permittivity ϵ_{NP} of the NP. Both approaches yield similar peak densities. Lowering ϵ_{NP} leads to an increase in the repulsive force on the counterions due to the induced charges on the NP surface, leading

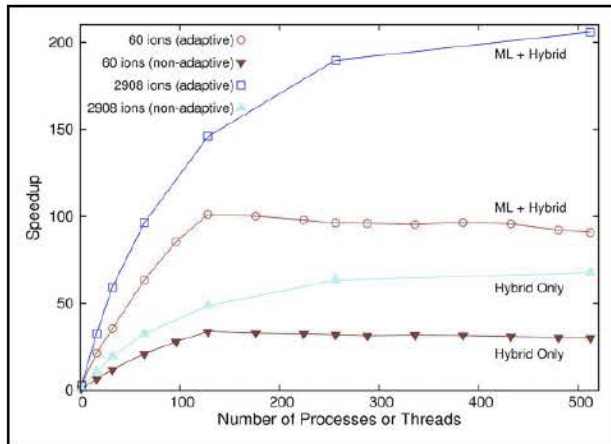


Figure 9. Strong scaling plot of the performance of the OpenMP/MPI hybrid technique with ML-enabled auto-tuning of the simulation timestep (open symbols) compared to the case with no auto-tuning (closed symbols). Data are shown for 60 and 2908 ions; in both systems, NP is meshed with 1082 mesh points (induced charges). For both systems, the combined ML and hybrid method outperforms the hybrid-only implementation.

to the reduction in the peak density of counterions near the NP surface.

5.3. Benchmarking ML-enhanced simulations

The enhanced dynamical optimization framework was benchmarked using BigRed2 cluster nodes. These nodes have maximal achieved performance of 596.4 teraFLOPS and feature a hybrid architecture based on two Cray, Inc., 344 XE6 (CPU-only) compute nodes, providing a total of 1020 compute nodes, 21,824 processor cores, and 43,648 GB of RAM. Each XE6 node has two AMD Opteron 16-core Abu Dhabi x86-64 CPUs and 64 GB of RAM; each XK7 node has one AMD Opteron 16-core Interlagos x86-64 CPU and 32 GB of RAM.

Figure 9 compares the strong scaling plot of the performance of the adaptive and the nonadaptive dynamical optimization framework parallelized using the OpenMP/MPI hybrid model. The reported speedup is defined as the ratio of serial run-time to the time taken by the parallelized simulations (with and without ML-enabled auto-tuning) to simulate a set t_b nanoseconds of ionic dynamics. In Figure 9, results are shown for simulation of a system of 60 ions and 1082 mesh points as well as a larger system of 2908 ions and 1082 mesh points for $t_b \approx 2$ ns. The larger system of 2908 ions is comprised of both positive and negative ions characterized by an electrolyte concentration $c \approx 0.2$ M (with 60 counterions, 1424 positive electrolyte ions, and 1424 negative electrolyte ions). While the ML models were not trained with c as an input parameter, simulations of the electrolyte systems using ML-predicted timestep and optimal virtual parameters associated with counterion-only systems were successful for up to $c \approx 0.4$ M (with varying t_b ; see Section 5.4.), enabling the benchmarking of simulations of larger systems.

With nonadaptive timestep, the hybrid model produced the maximum speedup of 33.80 with 128 processes (8 MPI nodes and 16 OpenMP threads inside each MPI node) for the smaller system. For the same configuration, the hybrid model with ML-enabled auto-tuning of the simulation timestep was able to achieve a maximum speedup of 101.07 with 128 processes. Thus, the runtime for this system was reduced from 55 h to 30 min (68 min without adaptive time-stepping). We note that the maximum speedup was calculated without considering the execution time reduction gained from the memory optimization techniques. For the hybrid model, we found that the optimal configuration of OpenMP threads is socket bound as noted in the literature (Rabenseifner et al., 2009). As a result, the number of optimal OpenMP threads in our experiment was 16 for any number of MPI processes.

When implemented to a larger system with a total number of 2908 ions and 1082 mesh points exhibiting induced surface charges, the combination of the hybrid methodology and the ML-based selection of adaptive timestep reduced the execution time of simulation for 2 ns from 88 days to 15.3 h (32 h without adaptive time-stepping) with a speedup of over 200. Clearly, the optimum number of MPI processes are proportional to the problem size when OpenMP thread affinity is set to the socket resulting in a well weak scaling system. The maximum speedup of 620.76 was obtained for 1024 processes executing a simulation of 5816 ions and 1272 mesh points for $t_b \approx 0.5$ ns.

5.4. Application: Concentrated electrolytes near an oil–water emulsion droplet

The ML-enhanced framework was applied to compute the distribution of monovalent electrolyte ions outside a charged oil-in-water emulsion droplet (Bier et al., 2008; de Graaf et al., 2008) at room temperature $T = 298$ K. Positive and negative ions were considered to be of the same size to simplify the system and focus on analyzing the effects of polarization charges on the density distributions. Such model systems have been considered in previous numerical studies of electrolyte ions near polarizable nanospheres (dos Santos et al., 2011; Messina, 2002; Shen et al., 2017). All ions were modeled as LJ spheres of diameter $\sigma = 0.357$ nm. The oil–water emulsion droplet was modeled as a spherical, dielectric interface with surface charge $Q = -60e$ and radius $a = 7.5\sigma \approx 2.7$ nm. The whole system of ions and the droplet was taken to be in a large spherical simulation cell of radius $b = 40\sigma \approx 14$ nm. The emulsion surface and the simulation cell boundary were modeled as spherical LJ walls. All excluded-volume interactions were modeled using the repulsive 6–12 LJ potential with $\epsilon_{LJ} = 1 k_B T$ and cutoff $r_c = 2^{1/6}\sigma$.

The dielectric permittivity of oil was taken to be $\epsilon_o = 2$, while water was associated with $\epsilon_w = 78.5$. The difference in the polarizable properties of oil and water lead to induced charges on the oil–water interface. Electrostatic interactions arising from the bare and induced charge

interactions in the system were modeled using the forces originating from the functional described in equation (4). The oil–water dielectric interface (which can be considered as the surface of the NP) was meshed with $M = 1082$ mesh points; higher M values were found to yield similar densities indicating that $M = 1082$ was large enough to obtain converged density profiles. In addition to the system with no added electrolyte, systems with electrolytes characterized by concentration $c \approx 0.02, 0.1$ M were considered to analyze the effects of changing c on the ionic distributions. Together with the 60 counterions (associated with the charged oil–water droplet), these concentrated electrolytes correspond to systems with a total of 350 and 1514 ions, respectively. Simulation of the smallest system (60 ions) assuming nonpolarizable NP surface was also performed for assessing the role of surface polarizability.

It should be noted that the ML procedure was only trained for smaller systems with counterions ($c = 0$, thus in the absence of co-ions). Further the training was performed for relatively smaller computational time (up to 2 million steps). With the application of the ML-enhanced framework to the aforementioned electrolyte systems, we are elucidating the transferability of the features learned for the smaller system to different, larger physical systems (with additional co-ions and salt counterions, and longtime dynamics). Such an extended application of the developed ML models is possible, in part, because the addition of electrolytes weakens the effective interaction between counterions and oil–water surface as a result of the screened electrostatic forces.

The aforementioned attributes of the physical system supply the input parameters for the enhanced dynamical optimization framework. Following the process elucidated in Figure 4, these input parameters were first passed to the ML model I to predict the required virtual system parameters to kick-start the simulation. Two protocols were followed: in one case, the auto-tuning using ML model II was performed for the entire duration of the simulation by repeating the optimal timestep and virtual parameter pattern inferred by the ANN for two million steps interval for subsequent cycles of two million steps. In the other approach, ML model II was employed to auto-tune the timestep and virtual parameters up to the first two million steps and subsequent evolution was performed with fixed values of these parameters predicted by ML model I (nonadaptive). Both approaches yielded the same results for the densities within the error bars. The total number of steps S were selected based on the convergence of the density distributions; S was system-dependent and converged results were obtained after $\approx 7 - 20$ ns depending on the electrolyte concentration.

In all simulations, regardless of system sizes and presence/absence of electrolytes, good energy conservation was observed with $R < 0.05$. Similarly, the induced charges on the oil–water dielectric interface were accurately tracked by the on-the-fly optimization framework ($f_d < 1\%$; inset in Figure 10 shows the accurate tracking of the functional for the $c = 0.02$ M case). These two key features demonstrated

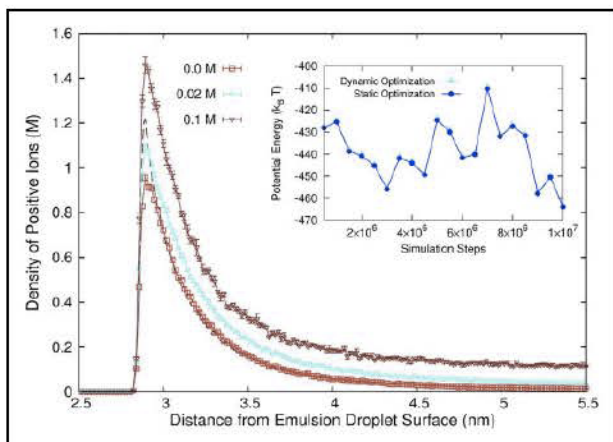


Figure 10. Ionic density profiles extracted from ML-enhanced MD simulations based on the dynamical optimization framework. Outset shows the density of positive ions for electrolytes of concentration $c \approx 0.0, 0.02, 0.1$ M near a negatively charged ($-60e$) oil–water emulsion with 60 associated counterions. The dielectric permittivity of oil and water is 2 and 78.5, respectively. Black dashed line refers to the result for the emulsion assumed to be unpolarizable (with oil permittivity as 78.5) at $c = 0$ M. (Inset) Comparison of the functional optimized dynamically (triangles) and the functional optimized at regular intervals keeping the ionic configuration static during the optimization process (circles) for the $c \approx 0.02$ M system.

the success of the ML-based virtual parameter selection process. As noted before, the bound on R_v is system-size and system-feature-dependent; for counterion-only systems, $R_v < 0.15$ was recorded as per the expected limit while for electrolyte systems the bound on R_v was higher and increased with increasing c . These findings demonstrate that ML models could be trained on smaller systems and applied to larger systems to obtain efficient and stable dynamics of ions in the latter case.

Figure 10 shows the density profiles of the positive ions associated with the aforementioned systems. For all concentrations, the densities reach a constant value in the bulk away from the polarizable oil–water surface (negative ion densities, not shown, also reach a constant value in the bulk). Positive ions are found to accumulate near the dielectric interface, with the peak density increasing with c . Negative ions are depleted near the interface due to the repulsion from the bare charge on the oil–water surface as well as the induced charge. Comparison of the no electrolyte (counterion-only) result with the case where the surface is considered to be unpolarizable (with a permittivity equal to that of water) is also shown. The polarization charges on the surface lead to depletion of ions from the interface. Increasing electrolyte concentration leads to the rise in the peak density; the overall behavior is determined by the competition between the ion-induced charge repulsion near the surface and the ion–ion electrostatic and steric correlations. The position of the peak density remains relatively unaltered regardless of the c value, as observed previously for monovalent electrolytes (Messina, 2002).

The inset in Figure 10 shows the comparison of the potential energy functional optimized dynamically with the energy obtained after optimizing at regular intervals keeping the ionic configuration static during the optimization process for the $c \approx 0.02$ M system. The stability and accuracy evident from this plot demonstrates the success of the ML-based parameter selection process; the potential energy and the associated induced charges it characterizes were accurately tracked at all times for up to 10 million steps. Other systems showed a similar agreement between dynamically optimized and statically optimized energy functionals.

6. Conclusion and outlook

We illustrated the computational gains accessible by integrating ML methods for parameter auto-tuning in MD simulations by demonstrating the enhancement of stability and efficiency of MD simulations of ions near polarizable NPs based on the dynamical optimization framework. The ANN-based ML model yielded the highest success rate among the nonlinear regression models employed to predict the virtual system parameters at the start of the simulation. When integrated with the MD simulation, the ANN model predicted the timestep and the associated optimal virtual parameters with 94.3% success rate. The auto-tuning of the simulation timestep resulting from the ML-enhanced, adaptive simulation framework enabled the simulation of ions for a longer physical time with a net speedup of $\approx 1.25 - 3$ (depending on the system configuration) compared to the nonadaptive simulation model. The combination of the ML procedure with the hybrid parallelization method generated stable dynamics of thousands of ions in the presence of polarizable NPs with computational time reducing from thousands of hours to tens of hours yielding a maximum speed up of ≈ 600 . Compared to the nonadaptive simulation with static initial virtual parameters, the stability of the adaptive framework increased by over an order of magnitude.

This enhanced simulation framework has many applications and we demonstrated its utility by generating stable, accurate dynamics of ions in the presence of a polarized nanoemulsion droplet for up to ≈ 10 million simulation steps. Additionally, we showed the broad applicability of the approach by demonstrating that the ML models trained on a smaller system can be applied successfully to produce accurate and stable dynamics of larger systems characterized by new attributes such as electrolyte concentration. At the same time, the approach reveals a limit on the electrolyte concentration and physical time that one can simulate based on training a counterion-only system. Future efforts will involve exploring the training of ML models with electrolyte concentration and attributes as input parameters. We will also explore integrating the current ML-enabled enhanced framework with fast Ewald solvers to support periodic boundary conditions and reduce the $O(n^2)$ scaling in system size to $O(n \log n)$ (Barros et al., 2014; Marchi et al., 2001). Further, future work will

involve exploring the training and design of ML models to predict the virtual parameters that generate stable ion dynamics near NPs of different shapes (e.g. discs, rods) (Brunk and Jadhaio, 2019) as well as in systems where multiple dielectric interfaces are present (Shen et al., 2017).

The use of ML to enhance the simulation framework enables users across the globe, with a diversity of domain experience, to simulate ions near polarizable NPs via the use of web-based applications hosted on services like nanoHUB. A tool powered by this enhanced framework was recently published on nanoHUB (Kadupitiya et al., 2018a). We will extend our framework to enable the process of using the data generated by this nanoHUB application for continuous training of the ML-based parameter prediction procedure. The approach presented here to integrate ML-based parameter prediction methods with MD-based simulations can be extended to other energy minimization problems (Brunk and Jadhaio, 2019). The implications of ML-enabled tuning of simulation timestep also suggest new avenues of exploring ML to advance simulation methods, such as ML-informed dynamic grid sizes in mesh-based problems.

Acknowledgements

The authors thank XSEDE startup allocation through Grant TG-DMR170089 and BigRed2 supercomputer at Indiana University for computing resources.



Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Science Foundation under the Network for Computational Nanotechnology (NCN) program through Award 1720625 (Network for Computational Nanotechnology—Engineered nanoBIO Node).

ORCID iD

JCS Kadupitiya  <https://orcid.org/0000-0003-2053-8441>
Vikram Jadhaio  <https://orcid.org/0000-0002-8034-2654>

References

- Abadi M, Barham P, Chen J, et al. (2016) Tensorflow: a system for large-scale machine learning. *OSDI* 16: 265–283.
- Abruña HD, Kiyá Y and Henderson JC (2008) Batteries and electrochemical capacitors. *Physics Today* 61(12): 43.
- Allen R, Hansen JP and Melchionna S (2001) Electrostatic potential inside ionic solutions confined by dielectrics: a variational approach. *Physical Chemistry Chemical Physics* 3: 4177–4186.
- Attard P (2003) Variational formulation for the electrostatic potential in dielectric continua. *The Journal of Chemical Physics* 119(3): 1365–1372.

- Balachandran PV, Xue D, Theiler J, et al. (2016) Adaptive strategies for materials design using uncertainties. *Scientific Reports* 6: 19660.
- Balakrishnan D and Puthusserypady S (2005) Multilayer perceptrons for the classification of brain computer interface data. In: *Proceedings of the IEEE 31st Annual Northeast Bioengineering Conference*, Hoboken, NJ, USA, 2-3 April 2005, pp. 118–119. IEEE.
- Balaprakash P, Dongarra J, Gambin T, et al. (2018) Autotuning in high-performance computing applications. *Proceedings of the IEEE* 106(99): 1–16.
- Barros K and Luijten E (2014) Dielectric effects in the self-assembly of binary colloidal aggregates. *Physical Review Letters* 113: 017801.
- Barros K, Sinkovits D and Luijten E (2014) Efficient and accurate simulation of dynamic dielectric objects. *The Journal of Chemical Physics* 140(6): 064903.
- Bartók AP, De S, Poelking C, et al. (2017) Machine learning unifies the modeling of materials and molecules. *Science Advances* 3(12): e1701816.
- Beckstein O, Tai K and Sansom MS (2004) Not ions alone: barriers to ion permeation in nanopores and channels. *Journal of the American Chemical Society* 126(45): 14694–14695.
- Behler J (2016) Perspective: machine learning potentials for atomistic simulations. *The Journal of Chemical Physics* 145(17): 170901.
- Bier M, Zwanikken J and van Roij R (2008) Liquid-liquid interfacial tension of electrolyte solutions. *Physical Review Letters* 101: 046104.
- Blöchl PE and Parrinello M (1992) Adiabaticity in first-principles molecular dynamics. *Physical Review B* 45: 9413–9416.
- Boda D, Gillespie D, Nonner W, et al. (2004) Computing induced charges in inhomogeneous dielectric media: application in a Monte Carlo simulation of complex ionic systems. *Physical Review E* 69(4): 046702.
- Botu V and Ramprasad R (2015) Adaptive machine learning framework to accelerate ab initio molecular dynamics. *International Journal of Quantum Chemistry* 115(16): 1074–1083.
- Brunk NE and Jadhao V (2019) Computational studies of shape control of charged deformable nanocontainers. *Journal of Materials Chemistry B* 7: 6370–6382.
- Buitinck L, Louppe G, Blondel M, et al. (2013) API design for machine learning software: experiences from the scikit-learn project. *arXiv preprint arXiv:1309.0238*.
- Car R and Parrinello M (1985) Unified approach for molecular dynamics and density-functional theory. *Physical Review Letters* 55(22): 2471–2474.
- Ch'ng K, Carrasquilla J, Melko RG, et al. (2017) Machine learning phases of strongly correlated fermions. *Physical Review X* 7: 031038.
- Chen K and Yu J (2014) Short-term wind speed prediction using an unscented Kalman filter based state-space support vector regression approach. *Applied Energy* 113: 690–705.
- Cherkassky V and Ma Y (2004) Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Networks* 17(1): 113–126.
- Chollet F, et al. (2015) Keras. GitHub, source code, Available at: <https://github.com/fchollet/keras>, (accessed 28 December 2019).
- Clapham DE (2007) Calcium signaling. *Cell* 131(6): 1047–1058.
- de Graaf J, Zwanikken J, Bier M, et al. (2008) Spontaneous charging and crystallization of water droplets in oil. *The Journal of Chemical Physics* 129(19): 194701.
- Denil M, Shakibi B, Dinh L, et al. (2013) Predicting parameters in deep learning. In: *Advances in Neural Information Processing System*, Babak Shakibi, October 2013, pp. 2148–2156.
- Deserno M and Holm C (1998) How to mesh up Ewald sums. I. A theoretical and numerical comparison of various particle mesh routines. *The Journal of Chemical Physics* 109(18): 7678–7693.
- dos Santos AP, Bakhshandeh A and Levin Y (2011) Effects of the dielectric discontinuity on the counterion distribution in a colloidal suspension. *The Journal of Chemical Physics* 135(4): 044124.
- dos Santos AP and Netz RR (2018) Dielectric boundary effects on the interaction between planar charged surfaces with counterions only. *The Journal of Chemical Physics* 148(16): 164103.
- Eng CL, Tong JC and Tan TW (2014) Predicting host tropism of influenza A virus proteins using random forest. *BMC Medical Genomics* 7(3): S1.
- Fahrenberger F, Xu Z and Holm C (2014) Simulation of electric double layers around charged colloids in aqueous solution of variable permittivity. *The Journal of Chemical Physics* 141(6): 064902.
- Felderhof BU (1977) Fluctuations of polarization and magnetization in dielectric and magnetic media. *The Journal of Chemical Physics* 67(2): 493–500.
- Ferguson AL (2017) Machine learning and data science in soft materials engineering. *Journal of Physics: Condensed Matter* 30(4): 043002.
- Fois ES, Penman JI and Madden PA (1993) Control of the adiabatic electronic state in ab initio molecular dynamics. *The Journal of Chemical Physics* 98(8): 6361–6368.
- Fox G, Glazier JA, Kadupitiya J, et al. (2019a) Learning everywhere: pervasive machine learning for effective high-performance computation. *arXiv preprint arXiv:1902.10810*.
- Fox G, Glazier JA, Kadupitiya J, et al. (2019b) *Learning everywhere: Pervasive machine learning for effective high-performance computation: Application background*. Technical report, Indiana University, February 2019. Available at: <http://dsc.soic> (accessed 28 December 2019).
- Gan Z, Wu H, Barros K, et al. (2015) Comparison of efficient techniques for the simulation of dielectric objects in electrolytes. *Journal of Computational Physics* 291: 317–333.
- Gan Z and Xu Z (2011) Multiple-image treatment of induced charges in Monte Carlo simulations of electrolytes near a spherical dielectric interface. *Physical Review E* 84: 016705.
- Guo AZ, Sevgen E, Sidky H, et al. (2018) Adaptive enhanced sampling by force-biasing using neural networks. *The Journal of Chemical Physics* 148(13): 134108.
- Hatlo MM and Lue L (2008) The role of image charges in the interactions between colloidal particles. *Soft Matter* 4: 1582–1596.

- Honig B and Nicholls A (1995) Classical electrostatics in biology and chemistry. *Science* 268(5214): 1144–1149.
- Jackson JD (1999) *Classical Electrodynamics*, 3rd ed. New York: Wiley.
- Jadhao V, Solis FJ and Olvera de la Cruz M (2012) Simulation of charged systems in heterogeneous dielectric media via a true energy functional. *Physical Review Letters* 109: 223905.
- Jadhao V, Solis FJ and Olvera de la Cruz M (2013) A variational formulation of electrostatics in a medium with spatially varying dielectric permittivity. *The Journal of Chemical Physics* 138(5): 054119.
- Jing Y, Jadhao V, Zwanikken JW, et al. (2015) Ionic structure in liquids confined by dielectric interfaces. *The Journal of Chemical Physics* 143(19): 194508.
- Kadupitige JC, Fox G and Jadhao V (2019) Machine learning for auto-tuning of simulation parameters in Car-Parrinello molecular dynamics. In: *APS Meeting Abstracts*. p. 672.
- Kadupitiya J, Brunk N, Ali S, et al. (2018a) Nanosphere electrostatics lab. DOI: doi:10.4231/D3WP9T88 W. Online on nanoHUB; source code on GitHub at github.com/softmaterialslab/np-electrostatics-lab. (accessed 28 December 2019)
- Kadupitiya J, Fox GC and Jadhao V (2018b) Machine learning for parameter auto-tuning in molecular dynamics simulations: efficient dynamics of ions near polarizable nanoparticles. *Indiana University, Nov* .
- Kadupitiya J, Fox GC and Jadhao V (2019) Machine learning for performance enhancement of molecular dynamics simulations. In: *International Conference on Computational Science*, Faro, Portugal, 12–14 June 2019, pp. 116–130. Springer.
- Kasson PM and Jha S (2018) Adaptive ensemble simulations of biomolecules. *Current Opinion in Structural Biology* 52: 87–94.
- Kazemi V and Sullivan J (2014) One millisecond face alignment with an ensemble of regression trees. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Washington, DC, 23–28 June 2014, pp. 1867–1874.
- Klimeck G, McLennan M, Brophy SP, et al. (2008) nanohub.org: advancing education and research in nanotechnology. *Computing in Science Engineering* 10(5): 17–23.
- Levin Y (2005) Strange electrostatics in physics, chemistry, and biology. *Physica A: Statistical Mechanics and its Applications* 352(1): 43–52.
- Lipparini F, Scalmani G, Mennucci B, et al. (2010) A variational formulation of the polarizable continuum model. *The Journal of Chemical Physics* 133(1): 014106.
- Liu J, Qi Y, Meng ZY, et al. (2017) Self-learning Monte Carlo method. *Physical Review B* 95: 041101.
- Long AW, Zhang J, Granick S, et al. (2015) Machine learning assembly landscapes from particle tracking data. *Soft Matter* 11(41): 8141–8153.
- Luehr N, Markland TE and Martínez TJ (2014) Multiple time step integrators in ab initio molecular dynamics. *The Journal of Chemical Physics* 140(8): 084116.
- Marchi M, Borgis D, Levy N, et al. (2001) A dielectric continuum molecular dynamics method. *The Journal of Chemical Physics* 114(10): 4377–4385.
- Marcus RA (1956) On the theory of oxidation-reduction reactions involving electron transfer. I. *The Journal of Chemical Physics* 24(5): 966–978.
- Messina R (2002) Image charges in spherical geometry: application to colloidal systems. *The Journal of Chemical Physics* 117(24): 11062–11074.
- Morningstar A and Melko RG (2017) Deep learning the Ising model near criticality. *arXiv preprint arXiv:1708.04622*.
- Nakano A, Vashishta P and Kalia RK (1994) Massively parallel algorithms for computational nanoelectronics based on quantum molecular dynamics. *Computer Physics Communications* 83(2): 181–196.
- Qin J, de Pablo JJ and Freed KF (2016) Image method for induced surface charge from many-body system of dielectric spheres. *The Journal of Chemical Physics* 145(12): 124903.
- Quan H, Srinivasan D and Khosravi A (2014) Short-term load and wind power forecasting using neural network-based prediction intervals. *IEEE Transactions on Neural Networks and Learning Systems* 25(2): 303–315.
- Rabenseifner R, Hager G and Jost G (2009) Hybrid MPI/OpenMP parallel programming on clusters of multi-core smp nodes. In: *2009 17th Euromicro International Conference on Parallel, Distributed and Network-based Processing*, Weimar, Germany, January 2009, pp. 427–436. IEEE.
- Raspaud E, Olvera de la Cruz M, Sikorav J, et al. (1998) Precipitation of DNA by polyamines: a polyelectrolyte behavior. *Biophysical Journal* 74(1): 381–393.
- Reiner ES and Radke CJ (1990) Variational approach to the electrostatic free energy in charged colloidal suspensions: general theory for open systems. *Journal of the Chemical Society, Faraday Transactions* 86: 3901–3912.
- Rottler J and Maggs AC (2004) Local molecular dynamics with coulombic interactions. *Physical Review Letters* 93(17): 170201.
- Sacanna S, Kegel WK and Philipse AP (2007) Thermodynamically stable pickering emulsions. *Physical Review Letters* 98: 158301.
- Sagui C and Darden T (2001) Multigrid methods for classical molecular dynamics simulations of biomolecules. *The Journal of Chemical Physics* 114(15): 6578–6591.
- Schoenholz SS (2018) Combining machine learning and physics to understand glassy systems. *Journal of Physics: Conference Series* 1036(1): 012021.
- Shen M, Li H and De La Cruz M (2017) Surface polarization effects on ion-containing emulsions. *Physical Review Letters* 119(13): 138002.
- Spellings M and Glotzer SC (2018) Machine learning for crystal identification and discovery. *AIChE Journal* 64(6): 2198–2206.
- Sprk M (1991) Computer simulation of the dynamics of induced polarization fluctuations in water. *The Journal of Physical Chemistry* 95(6): 2283–2291.
- Tuckerman M, Berne BJ and Martyna GJ (1992) Reversible multiple time scale molecular dynamics. *The Journal of Chemical Physics* 97(3): 1990–2001.
- Tyagi S, Suzen M, Sega M, et al. (2010) An iterative, fast, linear-scaling method for computing induced charges on arbitrary

- dielectric boundaries. *The Journal of Chemical Physics* 132(15): 154112.
- Villasenor J and Buneman O (1992) Rigorous charge conservation for local electromagnetic field solvers. *Computer Physics Communications* 69(2): 306–316.
- Ward L, Agrawal A, Choudhary A, et al. (2016) A general-purpose machine learning framework for predicting properties of inorganic materials. *NPJ Computational Materials* 2: 16028.
- Ward L, Dunn A, Faghaninia A, et al. (2018a) Matminer: an open source toolkit for materials data mining. *Computational Materials Science* 152: 60–69.
- Ward L, O’Keeffe SC, Stevick J, et al. (2018b) A machine learning approach for engineering bulk metallic glass alloys. *Acta Materialia* 159: 102–111.
- Whaley RC and Dongarra JJ (1998) Automatically tuned linear algebra software. In: *IEEE/ACM Conference on Supercomputing, 1998. SC98*, Orlando, FL, USA, 7-13 November 1998, pp. 38–38. IEEE.
- Wynveen A and Bresme F (2006) Interactions of polarizable media in water: a molecular dynamics approach. *The Journal of Chemical Physics* 124(10): 104502.
- Xu Z (2013) Electrostatic interaction in the presence of dielectric interfaces and polarization-induced like-charge attraction. *Physical Review E* 87: 013307.
- Yadav AK, Malik H and Chandel S (2016) Selection of most relevant input parameters using WEKA for artificial neural network based solar radiation prediction models. *Renewable and Sustainable Energy Reviews* 31: 509–519.
- Yigitbasi N, Willke TL, Liao G, et al. (2013) Towards machine learning-based auto-tuning of MapReduce. In: *2013 IEEE 21st International Symposium on Modeling, Analysis & Simulation of Computer and Telecommunication Systems (MASCOTS)*, San Francisco, CA, USA, 14–16 August 2013, pp. 11–20. IEEE.
- York DM and Karplus M (1999) A smooth solvation potential based on the conductor-like screening model. *The Journal of Physical Chemistry A* 103(50): 11060–11079.

Author biographies

JCS Kadupitiya was born in Sri Lanka in 1990 and completed his bachelor’s degree in Electrical and Information

Engineering from the University of Ruhuna, Sri Lanka, in 2015. He completed a Master’s degree in Computer Science and Engineering at the University of Moratuwa, Sri Lanka, in 2017. He also completed his second Master’s degree in Computer Engineering from Indiana University Bloomington in 2019, and he is currently enrolled in Indiana University Bloomington as a PhD student in the Department of Intelligent Systems Engineering. His research interests lie primarily in scientific machine learning and distributed computing.

Geoffrey C Fox received a PhD in Theoretical Physics from Cambridge University where he was Senior Wrangler. He is now a distinguished professor of Engineering, Computing, and Physics at Indiana University. He previously held positions at Caltech, Syracuse University, and Florida State University after being a postdoc at the Institute for Advanced Study at Princeton, Lawrence Berkeley Laboratory, and Peterhouse College Cambridge. He has supervised the PhD of 73 students with a h-index of 78 and over 36,000 citations. He is a Fellow of APS (Physics) and ACM (Computing) and works on the interdisciplinary interface between computing and applications.

Vikram Jadhao is an assistant professor of Intelligent Systems Engineering at Indiana University in Bloomington. Prior to joining Indiana University, he held postdoctoral fellowships in the Department of Physics and Astronomy at Johns Hopkins University as well as the Department of Materials Science and Engineering at Northwestern University. He received his PhD and MS in Physics from the University of Illinois at Urbana-Champaign and his BS in Physics from the Indian Institute of Technology at Kharagpur. His current research interests are in nanoscale self-assembly, soft-matter simulation, biomimetic materials, and machine learning. He is a recipient of an NSF CAREER award.