

RESEARCH ARTICLE | DECEMBER 15 2023

Predicting micro/nanoscale colloidal interactions through local neighborhood graph neural networks

Alexandra N. Filiatraut ; Jaber R. Mianroodi; Nima H. Siboni; Mehdi B. Zanjani  



J. Appl. Phys. 134, 234702 (2023)

<https://doi.org/10.1063/5.0175062>





APL Machine Learning

2023 Papers with Best Practices in Data Sharing and Comprehensive Background

Read Now

 AIP
Publishing

Predicting micro/nanoscale colloidal interactions through local neighborhood graph neural networks

Cite as: J. Appl. Phys. **134**, 234702 (2023); doi: [10.1063/5.0175062](https://doi.org/10.1063/5.0175062)

Submitted: 4 September 2023 · Accepted: 13 November 2023 ·

Published Online: 15 December 2023



Alexandra N. Filiatraut,¹  Jaber R. Mianroodi,² Nima H. Siboni,² and Mehdi B. Zanjani^{1,a)} 

AFFILIATIONS

¹Department of Mechanical and Manufacturing Engineering, Miami University, Oxford, Ohio 45056, USA

²Max-Planck-Institut für Eisenforschung, Düsseldorf, Germany

^{a)}Author to whom correspondence should be addressed: zanjanm@miamioh.edu

ABSTRACT

Understanding interparticle interactions has been one of the most important topics of research in the field of micro/nanoscale materials. Many significant characteristics of such materials directly stem from the way their building blocks interact with each other. In this work, we investigate the efficacy of a specific category of Machine Learning (ML) methods known as interaction networks in predicting interparticle interactions within colloidal systems. We introduce and study Local Neighborhood Graph Neural Networks (LN-GNNs), defined according to the local environment of colloidal particles derived from particle trajectory data. The LN-GNN framework is trained for unique categories of particle neighborhood environments in order to predict interparticle interactions. We compare the performance of the LN-GNN to a baseline interaction network with a simpler architecture and to an Instance-Based ML algorithm, which is computationally more expensive. We find that the prediction performance of LN-GNN measured as an average normalized mean absolute error outperforms the baseline interaction network by a factor of 2–10 for different local neighborhood configurations. Furthermore, LN-GNN's performance turns out to be very comparable to the instance-based ML framework while being an order of magnitude less expensive in terms of the required computation time. The results of this work can provide the foundations for establishing accurate models of colloidal particle interactions that are derived from real particle trajectory data.

Published under an exclusive license by AIP Publishing. <https://doi.org/10.1063/5.0175062>

I. INTRODUCTION

Interacting particles are present everywhere around us, from biological systems^{1–8} to synthetic materials of different sizes and scales.^{9–16} Therefore, elucidating the governing laws of particle interactions is a critical aspect of many scientific fields. Such interactions can be affected by a plethora of factors in complex ways depending on their specific composition and the overall environment surrounding them. For example, one of the important topics in the area of soft matter is understanding interparticle interactions between micro/nanoscale colloidal particles, which provide the building blocks for a variety of metamaterials with many applications in electronic, optical, and energy harvesting systems.^{17–22} The interactions between colloidal particles are crucial in directing their self-assembly process, as well as in determining structural features and the overall properties of colloidal metamaterials for various applications. Most

of the past studies on this topic have used empirical interaction potentials or simplified coarse-grained models of colloidal interactions.^{23–27} Such simplified interaction models are known to become less accurate as the relevant systems get more complex. Additionally, the use of more precise methods, such as first-principles calculations, becomes computationally impractical for particles with sizes in the range of a few hundred nanometers.

More recently, a number of studies have targeted the prediction of particle interactions based on particle trajectory data, i.e., measured spatial coordinates and velocities of the particles.^{28–33} Such methods typically aim to infer particle interaction profiles, or particle forces, from particle trajectories, which are, in principle, dictated by the physics of particle interactions. Data-driven and machine learning (ML)-based approaches have been shown to provide a versatile platform for this purpose. Recent studies in this area have introduced ML-based *Interaction Networks* (INs) that

20 June 2024 16:42:23

have demonstrated successful prediction of particle interactions for a variety of objects and systems, such as rigid body motion, deformable solids, and fluids.^{34–46}

Motivated by this recent pioneering work, we investigate the problem of predicting interparticle interactions within micro/nano-scale colloidal particle systems. We note that for colloidal particles, interparticle interactions are typically determined through surface functionalizing groups facilitating interactions that act over a relatively short range compared to particle sizes. Several examples of this interaction mechanism can be found within systems of DNA-coated^{18,47} and ligand-covered nanoparticles,¹⁰ which have been shown to create a wide variety of metamaterial structures. The nature of the interaction between these building blocks involves specific regions of the surface of each particle interacting with nearby surfaces of neighboring particles so that the surface functionalizing groups can facilitate contact between neighboring particles.^{10,16,48,49} This means particle interactions can be predicted through studying the local neighborhood of each particle. To this end, we define a machine learning framework that adopts a graph representation of colloidal particle systems and uses topological and compositional information about the local neighborhood of each particle, obtained from particle trajectory data, to train a graph neural network that can predict colloidal particle interactions. This framework, termed *Local Neighborhood Graph Neural Network* or *LN-GNN* in short, trains a ML algorithm to predict colloidal interaction forces by categorizing the local neighborhood of colloidal particles through defining physically relevant node and edge features to create a graph representation of the system. LN-GNN provides a concise framework that only utilizes information about the position of colloidal particles and provides a prediction of the net force acting on the particles. As a result, this method eliminates the need for pairwise calculations that are computationally demanding, and scales according to the cost of building a neighbor-list, i.e., $\sim O(N)$ for a system with N particles. Additionally, this framework is free of simplifying assumptions about particle interactions as it considers particle trajectories directly influenced by the physics of the systems. We note that such particle trajectories can be measured experimentally, as the use of real-time video microscopy has been a common practice in studying the assembly of micro/nanoscale colloids.⁵⁰ Analogous tools for probing self-assembly dynamics of micro/nanoscale objects have also been developed by recent advancements of the liquid-cell TEM technique,^{51–55} which can provide an opportunity to experimentally determine the trajectories of colloidal particles. Therefore, LN-GNN can provide the advantage of capturing interparticle interactions based on experimental data acquired from tracking of colloidal particle motion.

II. METHODOLOGY AND FRAMEWORK

We begin by first describing the overall scheme of the study. Considering a number of colloidal particles interacting with each other, our goal is to define and utilize a machine learning-based model, generally known as an Interaction Network (IN), to predict the colloidal particle interactions according to the surrounding environment of the particles. Ultimately, the output of the model will be the net force acting on a colloidal particle while the input is

composed of information about the local neighborhood of the particle. In other words, the trained ML model will predict the forces acting on a particle solely based on its surrounding environment. In order to provide the required training data for the ML framework, we use particle tracking information, i.e., the x , y , and z coordinates of each particle over time, which will be used to obtain particle accelerations and the desired data about particle interactions. Without loss of generality, we will consider particle tracking data obtained from Molecular Dynamics (MD) simulations; however, we note that particle tracking data obtained from experimental measurements, as mentioned above, can also be used within the same framework proposed in this study.

We now introduce the basics of the LN-GNN framework. Since particle interactions within systems of interest, i.e., micro/nanoscale colloidal particles with specific surface functionalization, are dominated by the interactions with neighboring particles, LN-GNN will derive predictions of interparticle interactions by relying on each particle's spatial configuration relative to its neighbors. Figure 1 shows a schematic demonstration of this framework, which involves inferring information about particle neighborhood and forces acting on each particle based on particle trajectories. Considering time-series information obtained from tracking particle coordinates, a dataset can be established that includes features about each particle and the list of its neighboring particles. These features can then provide input information to a machine learning algorithm that will be trained to predict forces acting on each particle. We note that in order to represent the interactions within a system of colloidal particles, a suitable framework can consist of a coarse-grained interaction model that delivers the *net force* acting on each specific particle, as sketched by the $\mathbf{F}_{net}$ vector for a sample particle colored in pink in Fig. 1. The choice of the net force on each particle, i.e., the cumulative force acting on the particle due to its interaction with all its neighbors, as the desired output of the ML-based framework will enable the prediction of the overall thermodynamic/kinetic behavior of such particles. Training LN-GNN requires a set of input data obtained through tracing particle positions over time. In order to generate an appropriate dataset for training this ML framework, we generated particle trajectories through molecular dynamics (MD) simulations of different mixtures of colloidal particles. The systems included in the dataset are one-, two-, and three-particle type systems with similar coarse-grained particle sizes. We additionally included a two-particle type system with two different particle sizes in the dataset. The MD simulations were conducted over 50 000 timesteps, providing tens of thousands of unique particle configurations for examination within this study. Detailed information about the dataset generation is provided in the [supplementary material](#).

Considering a group of colloidal particles interacting with each other, particle trajectories can be monitored as a function of time, i.e., over a number of time steps. For each particle, such as the central pink particle shown in the second panel of Fig. 2, we detect its *local neighborhood* by finding all the particles within a cutoff distance of 1.3 times the equilibrium average diameter of the two particle types. Each particle (i) can then be considered as a central node in a graph, connected to each of its neighboring particles (j) with an undirected edge, as shown in Fig. 2.

20 June 2024 16:42:23

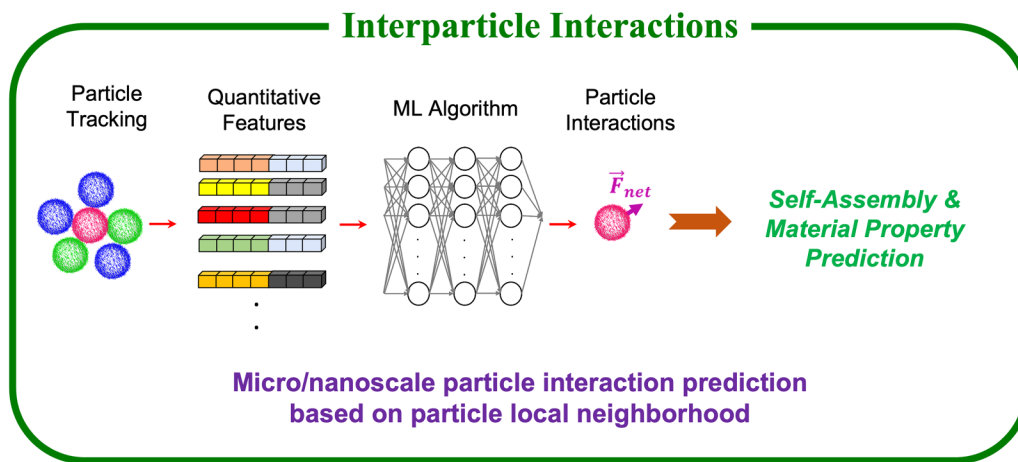


FIG. 1. Schematic of the framework for learning interparticle interactions of colloidal particles: prediction of net force on each particle based on its neighboring particles can lead to understanding assembly and properties of colloidal systems. Color coding reflects the variety of particle types and different features that can be studied within this framework.

In this representation, each neighbor particle j will be assigned a feature vector $\mathbf{v}_j$, composed of the average size of particles I and j , along with the relative position of j with respect to the central node I , i.e., $\mathbf{v}_j = \langle d_{ij}, x_j - x_I, y_j - y_I, z_j - z_I \rangle$ with d_{ij} representing the average diameter of the two particles. Each edge of the local neighborhood graph represents the connection between the central node (I) and one of its neighbors (j), which will be described through an edge feature vector $\mathbf{u}_{ij}$. The edge feature vector is used to describe the type of the interaction between particle I and its neighbor j . Without loss of generality, here we assume $\mathbf{u}_{ij}$ represents the type of the two particles I and j . This is specifically a good assumption for colloidal particle interactions. For example, in a binary mixture of colloidal particles of types A and B with two different types of surface functionalization,⁴⁸ there could be three interaction types

present in the system: $A-A$, $B-B$, and $A-B$. We will mark the different interaction types with a uniquely assigned numerical embedding value normalized between 0 and 1. As an example, in the case of a binary system of colloidal particles, $\mathbf{u}_{ij} = \langle 0.0 \rangle$, $\mathbf{u}_{ij} = \langle 0.5 \rangle$, and $\mathbf{u}_{ij} = \langle 1.0 \rangle$ will represent $A-A$ (green-green edges in Fig. 2), $A-B$ (green-blue edges in Fig. 2), and $B-B$ (blue-blue edges in Fig. 2), respectively. Once the node and edge feature vectors are set, we will combine all the data into a 2D feature matrix, where each row is made up of the concatenation of the corresponding node vector and the corresponding edge feature, i.e., $\mathbf{v}_j \oplus \mathbf{u}_{ij}$ for $j \in \{1, 2, \dots, N_I\}$, with N_I representing the number of the neighbors of particle I . This feature matrix is then fed through a convolutional layer, which utilizes a $1 \times n_{\text{feature}}$ kernel, with n_{feature} representing the total number of the concatenated node and

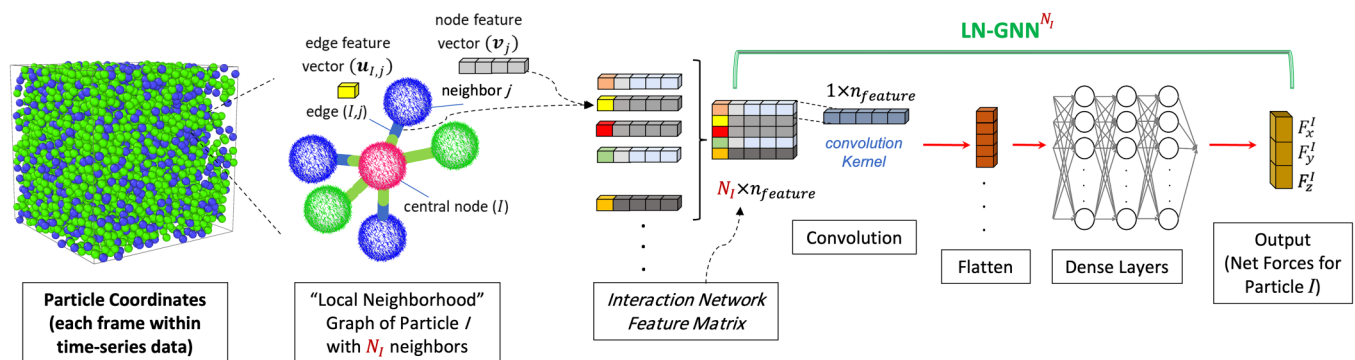


FIG. 2. Proposed framework of Local Neighborhood Graph Neural Networks (LN-GNNs) to learn interparticle interactions within colloidal particle systems. This sketch depicts the extraction of interaction feature matrices from particle coordinates and the model architecture that predicts the corresponding net force on each particle. The blue and green colors represent the two existing particle types in a sample binary system. The central particle, colored in pink for distinction, can represent any desired particle in the system.

edge features, i.e., the size of the $\mathbf{v}_j \oplus \mathbf{u}_{ij}$ vector. The result of the convolutional layer is then flattened and passed to a fully connected, multilayer perceptron (MLP) to generate the output layer representing the net force on the central particle, i.e., $\mathbf{F}_i = \langle F_x^i, F_y^i, F_z^i \rangle$. The LN-GNN framework may be utilized for predicting structure and properties of a variety of colloidal systems. For dilute solutions of colloidal particles, typically a lower number of neighboring particles are expected, and for crystalline phases, a larger number of neighbors are expected. Since LN-GNN is trained for each category (each value of N_i) based on exclusive training data, the framework can be effectively used for all different scenarios. Additionally, the same framework can be generalized to be used for other types of systems composed of discrete particles. It is noteworthy that for systems where the effective interaction range is larger than typical particle sizes, the input parameters of the LN-GNN algorithm need to be adjusted to include the features of additional neighboring particles beyond the nearest neighbors.

III. RESULTS AND DISCUSSION

After defining the LN-GNN method, we now discuss the training and evaluation process. First, the data generated earlier are sorted and separated based on the number of neighboring particles, i.e., the number of graph nodes in each case (N_i). We will then train and validate a LN-GNN for each N_i value using the corresponding dataset for the specific N_i value. It is important to note that any one central particle can have up to 12 nearby particle

neighbors due to the coordination number shared by all particles across all systems, i.e., for the systems we consider in this study $1 \leq N_i \leq 12$. As a result, 12 separate, yet architecturally similar LN-GNNs are created to handle the unique dimensionalities that local neighborhoods can take on, as shown in Fig. 3(a). For every system of interest, 12 sets of approximately 10 000 data points were collected, and each set was used for training and validation of the ML algorithm with the corresponding number of neighbors. During model training, 90% of data collected—roughly 9000 data points per model—were used for training of the model. The remaining 10%—roughly 1000 data points—were used for model validation.

After the ML algorithm is trained and validated, we used a separately generated dataset to evaluate the prediction accuracy of the LN-GNNs (see details in the [supplementary material](#)). Figure 3(b) demonstrates the prediction analysis scheme of the LN-GNN: for each single prediction attempt, a sample particle P is considered and its local neighborhood graph is established consisting of its N_P neighbors. Following this step, the corresponding trained algorithm with the same number of neighbors, i.e., LN-GNN N_P , is utilized to produce the prediction for the net forces acting on particle P . For evaluation of all the training, validation, and prediction steps, we use an average value of the mean absolute error defined as $MAE(\hat{\mathbf{F}}, \mathbf{F}) = \frac{1}{n_{data}} \frac{1}{n_p} \sum_{s \in D} \sum_{l=1}^{n_p} l_1(\hat{\mathbf{F}}_l^s, \mathbf{F}_l^s)$, where s represents a data point belonging to the dataset D of size n_{data} , and $l_1(\hat{\mathbf{F}}_l^s, \mathbf{F}_l^s)$ represents the sum of absolute differences between the x , y , and z

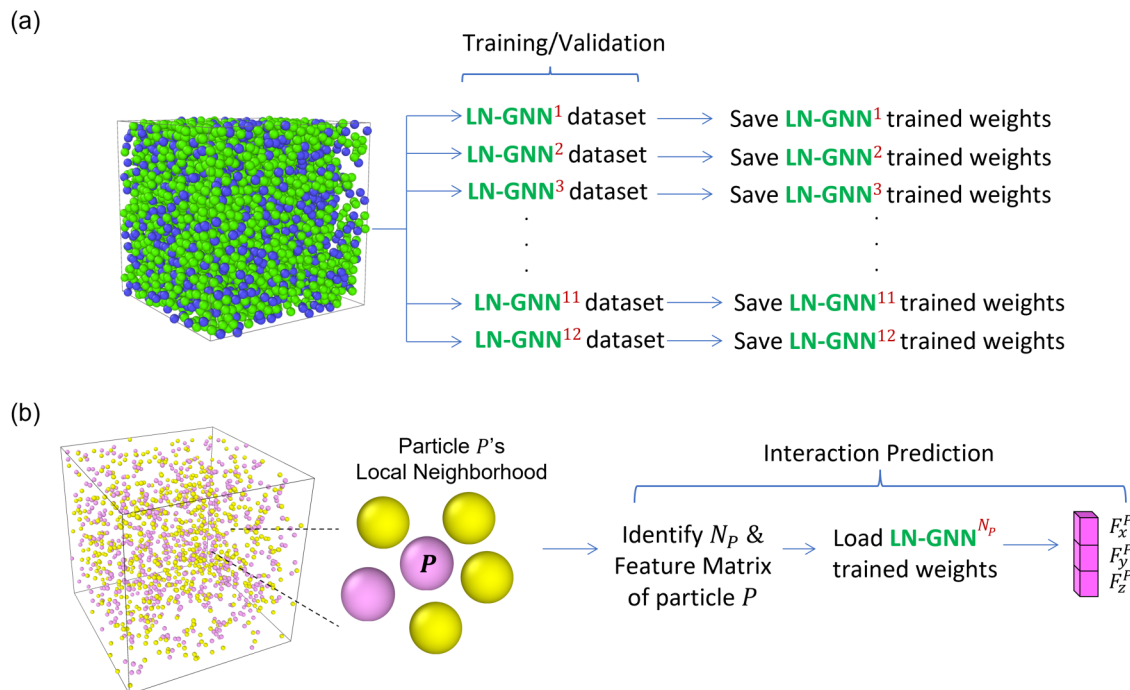


FIG. 3. Training/validation (a) vs prediction (b) schemes of interparticle interactions with the LN-GNN framework. The blue/green and pink/yellow colors represent the two existing particle types in sample binary systems.

20 June 2024 16:42:23

components of the true force $\hat{\mathbf{F}}_I^s$ and the predicted force $\mathbf{F}_I^s$ on particle I . Further details on the specific choices of parameters and various layer sizes of the LN-GNN as implemented in Tensorflow⁵⁶ are provided in the [supplementary material](#).

Following the framework explained above, we first train the different categories of LN-GNN ^{N_I} using the training and validation datasets. The average training and validation errors for all N_I values and all the four system types discussed above, i.e., single-type, binary, ternary, and binary with different particle sizes, were observed to be below 0.08, or below 8% of the maximum force, as detailed in the [supplementary material](#). We then used the independent prediction dataset to evaluate the prediction accuracy of the LN-GNN. The normalized-MAE values in [Fig. 4\(a\)](#) show the prediction accuracy for different LN-GNN ^{N_I} categories with $N_I = 1, 2, \dots, 12$. We observe that the prediction error is smaller for smaller N_I values, which is consistent with the fact that the dimensionality of a low neighbor network number is smaller than that of the higher neighbor network numbers. In essence, the spatial distribution of the neighboring particles and the resulting interactions are less complex when there are fewer neighbors of the central particle, thus resulting in a lower prediction error. The trends also show that in the case of a binary system of particles with different sizes, the prediction error is larger for smaller number of neighbors, e.g., $N_I = 5$ or 6, compared to other cases with same-size particles. This is due to the fact that variations in particle sizes complicate the morphology and the geometry of the local neighborhood of the central particle, which, in turn, results in a slightly larger prediction error. Furthermore, for a larger number of particle neighbors, such as $N_I = 11$ or 12, the prediction error turns out to be smaller for systems with two or three types of particles. This suggests that distinguishing between particle types provides a level of variety to the training dataset that results in a more accurate prediction by LN-GNN compared to the more uniform dataset of the single-type system. The overall average MAE values, shown in [Fig. 4\(b\)](#), show a value of around 0.02 for binary and ternary systems, while the single-type system MAE is around 0.03. The overall average MAE of the system with different particle sizes was largest at around 0.035, which relates to the additional complexity of the local neighborhood in the presence of particles with different sizes. In order to evaluate the performance of LN-GNN, which learns particle interactions based on categorizing their local neighborhood, we compare the accuracy of LN-GNN with a general baseline interaction network model. This baseline model, which we term *Basic-IN*, defines a framework that considers all categories of particles as part of the training dataset without distinguishing the local neighborhood directly. [Figure 5\(a\)](#) shows the baseline IN framework as compared to the LN-GNN method. In contrast to LN-GNN that analyzes a unique feature matrix based on the number of neighbors of the central particle, the baseline IN architecture passes the feature vector of each one of the neighbors of the central particle through a MLP and then performs a pooling operation on all of the resulting vectors, thus handling the interactions of particles with different numbers of the neighbors through the same ML structure. The baseline model was trained using the same overall dataset utilized to train LN-GNN. A separate dataset of 500 data points for each N_I value was generated and consistently

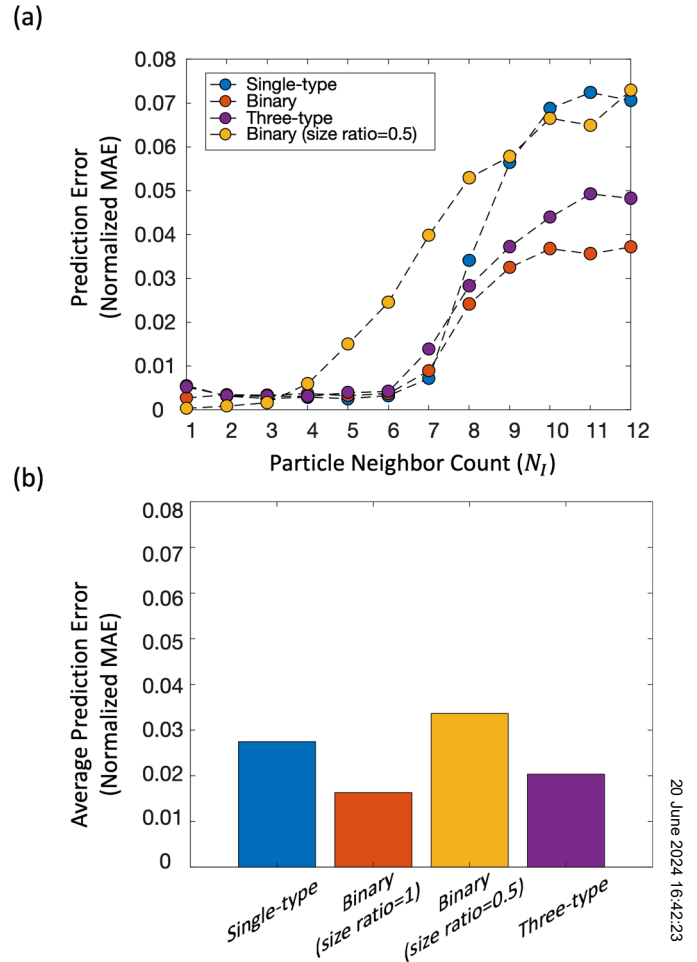


FIG. 4. Prediction error of the LN-GNN framework evaluated as a function of specific numbers of neighbors, N_I , in (a) and averaged over all specific cases in (b). Results were obtained on the test dataset, which was separate from training/validation dataset, for four different system types: single-type, binary, binary with different sizes, and three-type (ternary) system of particles.

used by LN-GNN and the baseline model to compare the prediction accuracy of the two schemes. [Figure 5\(b\)](#) shows the comparison between the accuracy of the LN-GNN and the Basic-IN models for sample cases studied in this work, including five different neighbor categories ($N_I = 4, 5, 6, 7$, or 8) and two system types, i.e., single-type and binary system of particles. The prediction MAE values show that the accuracy of the LN-GNN outperforms the accuracy of the baseline model across all number of neighbors, which confirm that the baseline model, by nature trained to predict particle interactions without directly distinguishing between particles' local neighborhood, will not perform as well as LN-GNN. This result suggests that categorizing the input data and adjusting the architecture of the ML framework based on particle local environment, although more cumbersome for coding and potentially

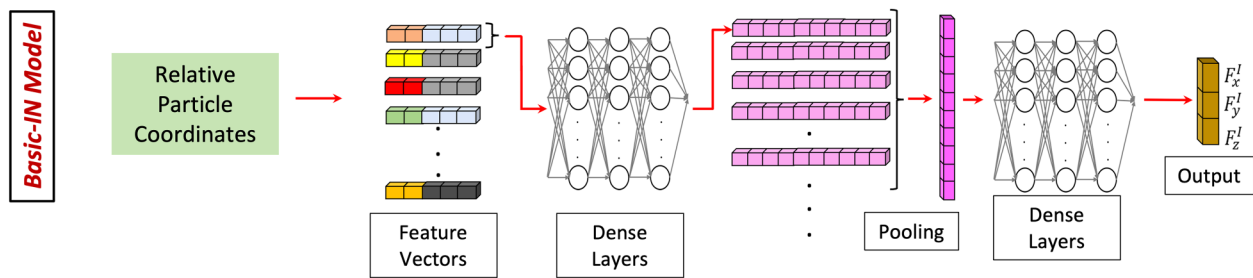
20 June 2024 16:42:23

requiring more computational resources, helps achieve more accurate predictions of particle-level interactions, especially in the case of micro/nanoscale building blocks whose behavior is dominated by short-range interactions. It is worth mentioning that the range of interparticle interactions can be of significance to the LN-GNN framework if particle sizes are much smaller than typical colloidal length scales. In such cases, one might need to look beyond the nearest neighbor particles in order to capture all the important factors needed for the successful training of the ML algorithm.

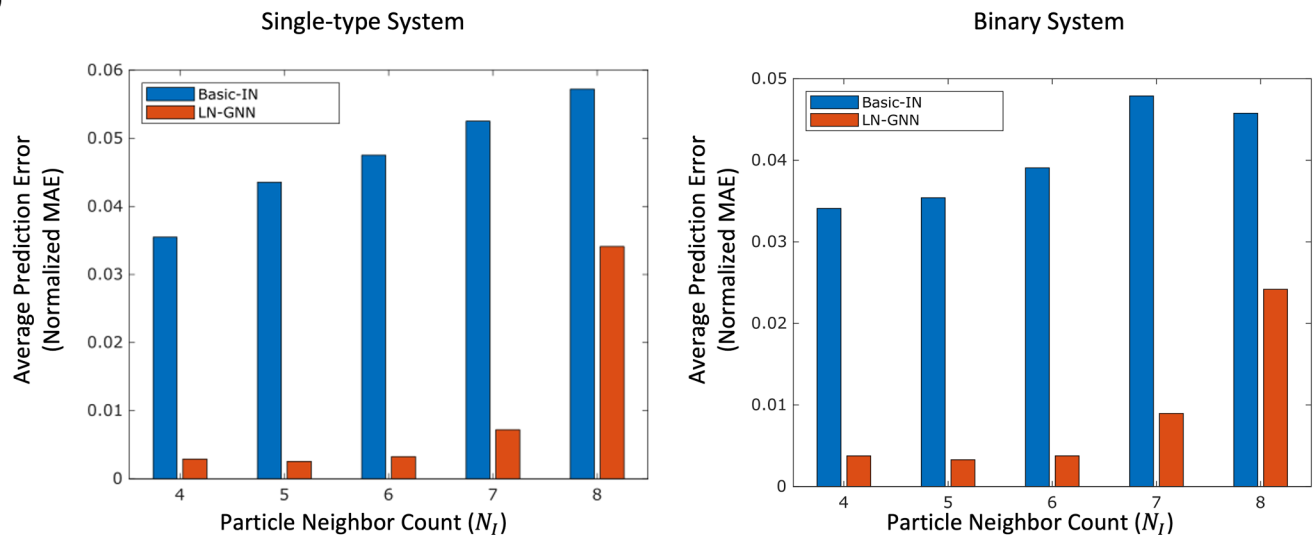
Additionally, we compare the performance of LN-GNN with an Instance-Based ML framework,⁵⁷ which employs a simple algorithm to make predictions of interparticle interactions investigated in this work, as shown in Fig. 6. The Instance-Based model follows a simple, but computationally expensive, algorithm by comparing the feature matrix corresponding to a particle of interest to the feature matrices of all the training configurations and identifying a “best match” average value for the prediction of net force acting on the particle. Specifically, we define a Euclidean distance between

the feature matrix of the prediction data point and each one of the training data points via $d_s = \sqrt{\sum (a_{ij} - a_{ij}^s)^2}$, where d_s represents a “distance” between the features of the point of interest and each training data point (s). a_{ij}^s and a_{ij} represent the individual components of the feature matrix for the training data point and prediction configuration, respectively. The output force values are determined as a distance-weighted average for k -nearest neighbors⁵⁷ of the point of interest from the training data, i.e., k points from the training data with the smallest d_s values. The details of this framework are sketched in Fig. 6(a). We selected $k = 6$ nearest neighbors for our calculations presented here. The average MAE values of the predicted forces, shown in Fig. 6(b), are fairly similar between LN-GNN and the Instance-Based models. On the other hand, the computation time scales proportional to the number of particles, $\sim O(N)$, for the LN-GNN method while it scales as $\sim O(N^2)$ for the Instance-Based method. Therefore, utilizing LN-GNN is generally preferable to simple Instance-Based

(a)



(b)



20 June 2024 16:42:23

FIG. 5. (a) The architecture of the baseline model (Basic-IN). (b) Prediction error (normalized-MAE) comparison between the baseline model and LN-GNN for a given set of data for particles with a specific number of neighbors, $N_I = 4, 5, 6, 7$, or 8.

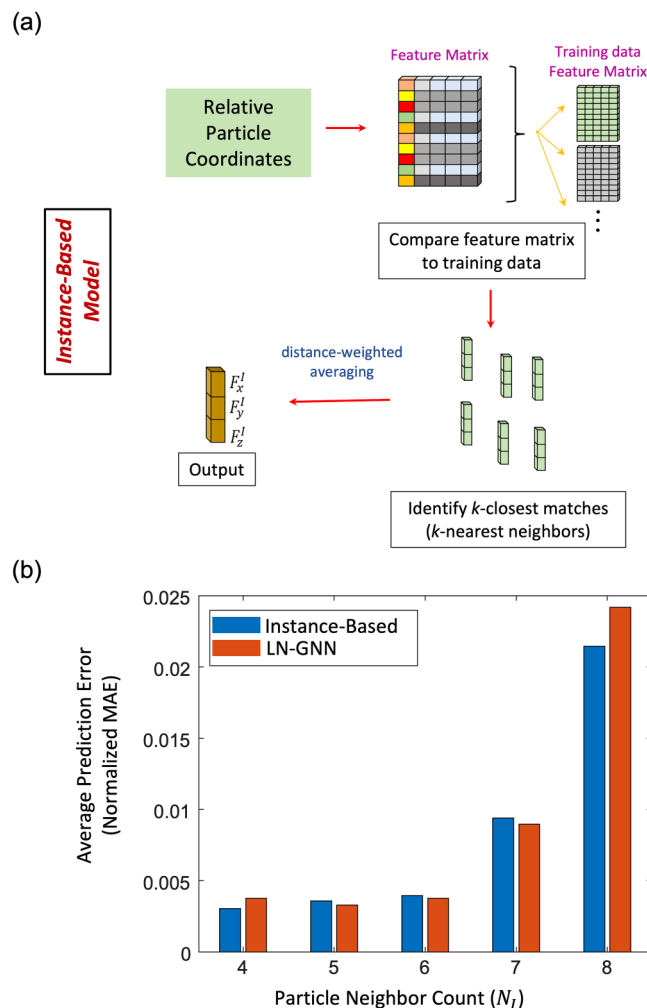


FIG. 6. (a) Description of a basic Instance-Based ML algorithm for predicting colloidal interactions. (b) Prediction error of the LN-GNN framework compared to the basic Instance-Based ML algorithm evaluated as a function of specific numbers of neighbors, N_l , calculated for the binary system of particles introduced earlier.

algorithms unless the number of particles in a system of interest is very small, e.g., finite-sized assemblies of colloidal particles. The combined results in Figs. 4–6 demonstrate that, while the accuracy of the ML framework is generally lower for cases with a larger number of neighbors, both LN-GNN and Instance-based methods provide considerably better predictions compared to the baseline IN framework. We note that such interaction network predictions are most accurate for colloidal particles that exist in dilute solutions, e.g., at early stages of colloidal self-assembly processes. However, the computational cost of using interaction network models is much smaller compared to using fully atomistic models that attempt to predict particle interactions in any scenario.

IV. CONCLUSIONS

In conclusion, we investigate a machine learning framework that predicts micro/nanoscale particle interactions based on data obtained from particle trajectories. This framework, named LN-GNN, considers the local neighborhood environment of a particle to define a representative graph with a central node depicting the particle of interest. This graph representation is utilized to predict the net forces on micro/nanoscale particles after LN-GNN is trained on appropriate datasets. The LN-GNN framework can practically be utilized in a wide range of problems related to the study of self-assembly and phase behavior of colloidal particles. This process involves three steps: (i) For training purposes, particle trajectory data, either from experiments or simulation data, can be fed to the LN-GNN algorithm according to the number of neighbors of each particle. (ii) Once the training is performed, the trained models can be saved and used for the prediction of forces on colloidal particles in other systems. (iii) The predicted net forces on each particle can then be utilized to calculate structural evolution and various properties of colloidal assemblies, as commonly performed with MD and BD simulations. One of the main impacts of using such a framework is that the use of the trained LN-GNN framework for the prediction of particle interactions will be computationally much cheaper compared to calculating the sum of pairwise interactions of particles or using fully atomistic models to predict particle interactions.

Understanding small-scale particle interactions is one of the fundamental topics in the design and discovery of advanced materials. Given the recent advancements in TEM technology and other relevant methods that can provide accurate experimental data of particle trajectories, LN-GNN and other ML-based schemes can pave the way for obtaining accurate, physics-informed, models of complex micro/nanoscale interparticle interactions in order to guide the design and development of a variety of new functional metamaterials.

SUPPLEMENTARY MATERIAL

See the supplementary material for the description of data generation and processing, details of machine learning architecture and parameters utilized within LN-GNN and baseline IN, and specifics of training and validation data for the LN-GNN framework.

ACKNOWLEDGMENTS

We acknowledge support from the Ohio Space Grant Consortium (No. OSGC 2022-2023) and the National Science Foundation (NSF) (Award No. 2025319). We also acknowledge computational resources provided by the Ohio Supercomputer Center (Grant No. PMIU0139) and the Redhawk Cluster at Miami University.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Alexandra N. Filiatraut: Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Validation (equal); Visualization (equal); Writing – original draft (equal). **Jaber R. Mianroodi:** Conceptualization (supporting); Methodology (lead); Supervision (supporting); Writing – review & editing (supporting). **Nima H. Siboni:** Conceptualization (supporting); Methodology (equal); Supervision (supporting); Writing – review & editing (supporting). **Mehdi B. Zanjani:** Conceptualization (lead); Data curation (equal); Formal analysis (equal); Funding acquisition (lead); Investigation (equal); Methodology (equal); Supervision (lead); Writing – original draft (equal).

DATA AVAILABILITY

The data that support the findings of this study are openly available in GitHub at <https://github.com/Zanjani-Lab/LN-GNN/tree/main>, Ref. 58.

REFERENCES

- ¹T. Gulik-Krzywicki, E. Shechter, V. Luzzati, and M. Faure, “Interactions of proteins and lipids: Structure and polymorphism of protein–lipid–water phases,” *Nature* **223**, 1116–1121 (1969).
- ²P. W. Rothmund, “Folding DNA to create nanoscale shapes and patterns,” *Nature* **440**, 297–302 (2006).
- ³E. P. Furlani, “Magnetic biotransport: Analysis and applications,” *Materials* **3**, 2412–2446 (2010).
- ⁴A. Marchand, “New protein–protein interactions designed by a computer,” *Nat. Biotechnol.* **38**, 426–432 (2020).
- ⁵C. Liu, P. Elvati, S. Majumder, Y. Wang, A. P. Liu, and A. Violi, “Predicting the time of entry of nanoparticles in lipid membranes,” *ACS Nano* **13**, 10221–10232 (2019).
- ⁶Z. Shaukat, S. Aiman, and C.-H. Li, “Protein–protein interactions: Methods, databases, and applications in virus–host study,” *World J. Virol.* **10**, 288 (2021).
- ⁷Z. Othman, B. C. Pastor, S. van Rijt, and P. Habibovic, “Understanding interactions between biomaterials and biological systems using proteomics,” *Biomaterials* **167**, 191–204 (2018).
- ⁸S. T. Johnston, M. Faria, and E. J. Crampin, “Understanding nano-engineered particle–cell interactions: Biological insights from mathematical models,” *Nanoscale Adv.* **3**, 2139–2156 (2021).
- ⁹W. A. Ducker, T. J. Senden, and R. M. Pashley, “Direct measurement of colloidal forces using an atomic force microscope,” *Nature* **353**, 239–241 (1991).
- ¹⁰E. V. Shevchenko, D. V. Talapin, N. A. Kotov, S. O’Brien, and C. B. Murray, “Structural diversity in binary nanoparticle superlattices,” *Nature* **439**, 55–59 (2006).
- ¹¹M. He, J. P. Gales, É. Ducrot, Z. Gong, G.-R. Yi, S. Sacanna, and D. J. Pine, “Colloidal diamond,” *Nature* **585**, 524–529 (2020).
- ¹²R. M. Erb, H. S. Son, B. Samanta, V. M. Rotello, and B. B. Yellen, “Magnetic assembly of colloidal superstructures with multipole symmetry,” *Nature* **457**, 999–1002 (2009).
- ¹³C. A. Silvera Batista, R. G. Larson, and N. A. Kotov, “Nonadditivity of nanoparticle interactions,” *Science* **350**, 1242477 (2015).
- ¹⁴Y. Lin, A. Böker, J. He, K. Sill, H. Xiang, C. Abetz, X. Li, J. Wang, T. Emrick, S. Long, and Q. Wang, “Self-directed self-assembly of nanoparticle/copolymer mixtures,” *Nature* **434**, 55–59 (2005).
- ¹⁵S. Y. Park, A. K. Lytton-Jean, B. Lee, S. Weigand, G. C. Schatz, and C. A. Mirkin, “DNA-programmable nanoparticle crystallization,” *Nature* **451**, 553–556 (2008).
- ¹⁶D. V. Talapin, E. V. Shevchenko, C. B. Murray, A. V. Titov, and P. Král, “Dipole-dipole interactions in nanoparticle superlattices,” *Nano Lett.* **7**, 1213–1219 (2007).
- ¹⁷W. P. McConnell, J. P. Novak, L. C. Brousseau, R. R. Fuijier, R. C. Tenent, and D. L. Feldheim, “Electronic and optical properties of chemically modified metal nanoparticles and molecularly bridged nanoparticle arrays,” *J. Phys. Chem. B* **104**, 8925–8930 (2000).
- ¹⁸A. J. Kim, R. Scarlett, P. L. Biancanello, T. Sinno, and J. C. Crocker, “Probing interfacial equilibration in microsphere crystals formed by DNA-directed assembly,” *Nat. Mater.* **8**, 52–55 (2009).
- ¹⁹K. Miszt, J. De Graaf, G. Bertoni, D. Dorfs, R. Brescia, S. Marras, L. Ceseracciu, R. Cingolani, R. Van Roij, M. Dijkstra, and L. Manna, “Hierarchical self-assembly of suspended branched colloidal nanocrystals into superlattice structures,” *Nat. Mater.* **10**, 872 (2011).
- ²⁰M. A. Boles, M. Engel, and D. V. Talapin, “Self-assembly of colloidal nanocrystals: From intricate structures to functional materials,” *Chem. Rev.* **116**, 11220–11289 (2016).
- ²¹D. Garcia-Lojo, S. Nunez-Sanchez, S. Gomez-Grana, M. Grzelczak, I. Pastoriza-Santos, J. Perez-Juste, and L. M. Liz-Marzan, “Plasmonic supercrystals,” *Acc. Chem. Res.* **52**, 1855–1864 (2019).
- ²²R. Mao, J. O’Leary, A. Mesbah, and J. Mittal, “A deep learning framework discovers compositional order and self-assembly pathways in binary colloidal mixtures,” *JACS Au* **2**, 1818–1828 (2022).
- ²³S. Masoumi, H. Valipour, and M. J. Abdolhosseini Qomi, “Interparticle interactions in colloidal systems: Toward a comprehensive mesoscale model,” *ACS Appl. Mater. Interfaces* **9**, 27338–27349 (2017).
- ²⁴R. Scarlett, J. C. Crocker, and T. Sinno, “Computational analysis of binary segregation during colloidal crystallization with DNA-mediated interactions,” *J. Chem. Phys.* **132**, 234705 (2010).
- ²⁵M. T. Casey, W. B. Rogers, R. Scarlett, I. Jenkins, T. Sinno, and J. C. Crocker, “Driving diffusionless transformations in colloidal crystals using DNA handshaking,” *Nat. Commun.* **3**, 1209 (2012).
- ²⁶M. B. Zanjani, I. C. Jenkins, J. C. Crocker, and T. Sinno, “Colloidal cluster assembly into ordered superstructures via engineered directional binding,” *ACS Nano* **10**, 11280–11289 (2016).
- ²⁷K. Aryana, J. B. Stahley, N. Parvez, K. Kim, and M. B. Zanjani, “Superstructures of multielement colloidal molecules: Efficient pathways to construct reconfigurable photonic and phononic crystals,” *Adv. Theory Simul.* **2**, 1800198 (2019).
- ²⁸I. C. Jenkins, J. C. Crocker, and T. Sinno, “Interaction potentials from arbitrary multi-particle trajectory data,” *Soft Matter* **11**, 6948–6956 (2015).
- ²⁹N. G. Almaraz and E. Lomba, “Determination of the interaction potential from the pair distribution function: An inverse Monte Carlo technique,” *Phys. Rev. E* **68**, 011202 (2003).
- ³⁰J. Tersoff, “New empirical model for the structural properties of silicon,” *Phys. Rev. Lett.* **56**, 632 (1986).
- ³¹K. Vondermassen, J. Bongers, A. Mueller, and H. Versmold, “Brownian motion: A tool to determine the pair potential between colloid particles,” *Langmuir* **10**, 1351–1353 (1994).
- ³²L. Reatto, D. Levesque, and J. Weis, “Iterative predictor-corrector method for extraction of the pair interaction from structural data for dense classical liquids,” *Phys. Rev. A* **33**, 3451 (1986).
- ³³F. Barocchi, M. Zoppi, and P. Egelstaff, “Density analysis of the neutron structure factor and the determination of the pair potential of krypton,” *Phys. Rev. A* **31**, 2732 (1985).
- ³⁴A. Radovic, M. Williams, D. Rousseau, M. Kagan, D. Bonacorsi, A. Himmel, A. Aurisano, K. Terao, and T. Wongjirad, “Machine learning at the energy and intensity frontiers of particle physics,” *Nature* **560**, 41–48 (2018).
- ³⁵J. Shlomi, P. Battaglia, and J.-R. Vlimant, “Graph neural networks in particle physics,” *Mach. Learn.: Sci. Technol.* **2**, 021001 (2020).
- ³⁶O. Méndez-Lucio, M. Ahmad, E. A. del Rio-Chanona, and J. K. Wegner, “A geometric deep learning approach to predict binding conformations of bioactive molecules,” *Nat. Mach. Intell.* **3**, 1033–1039 (2021).

20 June 2024 16:42:23

- ³⁷C. W. Park, M. Kornbluth, J. Vandermause, C. Wolverton, B. Kozinsky, and J. P. Mailoa, "Accurate and scalable graph neural network force field and molecular dynamics with direct force architecture," *npj Comput. Mater.* **7**, 73 (2021).
- ³⁸P. Battaglia, R. Pascanu, M. Lai, and D. Jimenez Rezende, "Interaction networks for learning about objects, relations and physics," in *Proceedings of the 30th Conference on Neural Information Processing Systems (NIPS)*, 2016).
- ³⁹A. Sanchez-Gonzalez, J. Godwin, T. Pfaff, R. Ying, J. Leskovec, and P. Battaglia, "Learning to simulate complex physics with graph networks," in *International Conference on Machine Learning (PMLR)*, 2020, pp. 8459–8468.
- ⁴⁰S. Matsumoto, S. Ishida, M. Araki, T. Kato, K. Terayama, and Y. Okuno, "Extraction of protein dynamics information from cryo-EM maps using deep learning," *Nat. Mach. Intell.* **3**, 153–160 (2021).
- ⁴¹T. Pfaff, M. Fortunato, A. Sanchez-Gonzalez, and P. W. Battaglia, "Learning mesh-based simulation with graph networks," *arXiv:2010.03409* (2020).
- ⁴²Z. Han, D. S. Kammer, and O. Fink, "Learning physics-consistent particle interactions," *PNAS Nexus* **1**, pgac264 (2022).
- ⁴³S. Thaler and J. Zavavlav, "Learning neural network potentials from experimental data via differentiable trajectory reweighting," *Nat. Commun.* **12**, 6884 (2021).
- ⁴⁴H. Ibayashi, T. M. Razakh, L. Yang, T. Linker, M. Olguin, S. Hattori, Y. Luo, R. K. Kalia, A. Nakano, K.-I. Nomura, and P. Vashishta, "Allegro-legato: Scalable, fast, and robust neural-network quantum molecular dynamics via sharpness-aware minimization," in *International Conference on High Performance Computing* (Springer, 2023), pp. 223–239.
- ⁴⁵O. T. Unke and M. Meuwly, "Physnet: A neural network for predicting energies, forces, dipole moments, and partial charges," *J. Chem. Theory Comput.* **15**, 3678–3693 (2019).
- ⁴⁶S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt, and B. Kozinsky, "E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials," *Nat. Commun.* **13**, 2453 (2022).
- ⁴⁷R. J. Macfarlane, B. Lee, M. R. Jones, N. Harris, G. C. Schatz, and C. A. Mirkin, "Nanoparticle superlattice engineering with DNA," *Science* **334**, 204–208 (2011).
- ⁴⁸W. B. Rogers, W. M. Shih, and V. N. Manoharan, "Using DNA to program the self-assembly of colloidal nanoparticles and microparticles," *Nat. Rev. Mater.* **1**, 16008 (2016).
- ⁴⁹W. B. Rogers and J. C. Crocker, "Direct measurements of DNA-mediated colloidal interactions and their quantitative modeling," *Proc. Natl. Acad. Sci. U.S.A.* **108**, 15687–15692 (2011).
- ⁵⁰J. C. Crocker and D. G. Grier, "Methods of digital video microscopy for colloidal studies," *J. Colloid Interface Sci.* **179**, 298–310 (1996).
- ⁵¹Q. Chen, H. Cho, K. Manthiram, M. Yoshida, X. Ye, and A. P. Alivisatos, "Interaction potentials of anisotropic nanocrystals from the trajectory sampling of particle motion using in situ liquid phase transmission electron microscopy," *ACS Cent. Sci.* **1**, 33–39 (2015).
- ⁵²N. De Jonge and F. M. Ross, "Electron microscopy of specimens in liquid," *Nat. Nanotechnol.* **6**, 695–704 (2011).
- ⁵³H. Zheng, R. K. Smith, Y.-W. Jun, C. Kisielowski, U. Dahmen, and A. P. Alivisatos, "Observation of single colloidal platinum nanocrystal growth trajectories," *Science* **324**, 1309–1312 (2009).
- ⁵⁴J. M. Yuk, J. Park, P. Ercius, K. Kim, D. J. Hellebusch, M. F. Crommie, J. Y. Lee, A. Zettl, and A. P. Alivisatos, "High-resolution EM of colloidal nanocrystal growth using graphene liquid cells," *Science* **336**, 61–64 (2012).
- ⁵⁵X. Ye, M. R. Jones, L. B. Frechette, Q. Chen, A. S. Powers, P. Ercius, G. Dunn, G. M. Rotskoff, S. C. Nguyen, V. P. Adiga, and A. Zettl, "Single-particle mapping of nonequilibrium nanocrystal transformations," *Science* **354**, 874–877 (2016).
- ⁵⁶M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems" (2015), software available from tensorflow.org.
- ⁵⁷T. Mitchell, *Machine Learning* (McGraw Hill, 1997).
- ⁵⁸A. Filatrat, J. R. Mianroodi, N. H. Siboni, and M. Zanjani, "Predicting micro/nanoscale colloidal interactions through local neighborhood graph neural networks" (2023), see <https://github.com/Zanjani-Lab/LN-GNN/tree/main>.