

Correlation Tracking: Using simulations to interpolate highly correlated particle tracks

Ella M. King¹, Zizhao Wang², David A. Weitz,^{1,2} Frans Spaepen,² and Michael P. Brenner^{1,2}

¹Physics Department, Harvard University, USA

²School of Engineering and Applied Sciences, Harvard University, USA



(Received 25 January 2022; accepted 6 April 2022; published 25 April 2022)

Despite significant advances in particle imaging technologies over the past two decades, few advances have been made in particle tracking, i.e., linking individual particle positions across time series data. The state-of-the-art tracking algorithm is highly effective for systems in which the particles behave mostly independently. However, these algorithms become inaccurate when particle motion is highly correlated, such as in dense or strongly interacting systems. Accurate particle tracking is essential in the study of the physics of dense colloids, such as the study of dislocation formation, nucleation, and shear transformations. Here, we present a method for particle tracking that incorporates information about the correlated motion of the particles. We demonstrate significant improvement over the state-of-the-art tracking algorithm in simulated data on highly correlated systems.

DOI: [10.1103/PhysRevE.105.044608](https://doi.org/10.1103/PhysRevE.105.044608)

I. INTRODUCTION

Accurately tracking particle positions over time remains a major challenge in advancing colloidal physics research. Single-particle level tracking is critical to studies of the mechanisms of crystal nucleation [1], melting [2], glass formation and dislocations [3], and a variety of other phenomena in dense colloid systems. While advanced confocal microscopy and image analysis techniques have improved the accuracy of the extraction of particle positions from experiments [4], less progress has been made in accurate tracking of particles exhibiting highly correlated particle motion, such as dense or strongly interacting systems.

Existing tracking algorithms are based on the fundamental work of Crocker and Grier [5], in which positions of particles undergoing Brownian motion are linked between time steps by finding the closest candidates in successive frames. Importantly, this method has two conditions for accurate reconstruction of the tracks. First, the particle displacement between two successive time steps should be much smaller than the interparticle separation. Second, each particle should diffuse independently. If these two conditions are not met, the trajectory of each particle can often not be distinguished from that of its neighbors. These conditions are violated in systems of strongly interacting particles, which limits the accuracy of the Crocker-Grier algorithm.

Although enhancements to Crocker and Grier's original algorithm have been made, such as handling uniform flow [6,7], the fundamental requirement that particle displacement be far less than the interparticle separation remains a significant limitation. This problem is particularly acute in three-dimensional (3D) imaging due to limitations in the frame rate.

In this work, we introduce an algorithm that incorporates information about correlations in particle motion by using

simulations to improve estimates of tracking probability distributions. The algorithm achieves up to 33% more accurate tracking than Crocker and Grier in simulated systems that go beyond the bounds of the algorithm introduced by Crocker and Grier. We also demonstrate robust results on noisy experimental data.

II. TRACKING CORRELATED MOTION

A. The Crocker-Grier algorithm

The core assumption of the Crocker-Grier tracking algorithm is that a given particle at position x_0 will be found at a position x after a time t has passed with probability

$$P(x, t) = \frac{1}{(4\pi Dt)^{\frac{d}{2}}} e^{-\frac{(x-x_0)^2}{4Dt}}, \quad (1)$$

where D is the diffusion coefficient and d is the dimension. This is the Green's function solution to the single-particle diffusion equation [5]. The tracking algorithm is accurate for systems in which particles diffuse independently. However, for systems in which the motion of a group of particles is correlated, the assumption underlying Crocker-Grier breaks down.

In particular, once multiparticle correlations are introduced, the single-particle diffusion equation is no longer an accurate model of the system.

B. The correlation tracking algorithm

The proposed tracking algorithm approximates multiparticle correlations using simulated data to inform predictions. We refer to this method as “correlation tracking.” Inspired by prior work showing that particle interactions can directly tune dynamics [8], we use prior knowledge about particle interactions to make more accurate predictions about the dynamics.

By incorporating this physical information into the algorithm, we are able to increase tracking accuracy.

Consider first a pair of frames, x_0 and x_1 . The frame x_i contains a list of particle positions at time i . The aim of particle tracking is to match particle labels between frame x_0 and frame x_1 . Because we know the interaction potential, we can simulate the motion beginning at position x_0 and let it evolve for the same length of time as between the pair of frames. These simulations are stochastic, and thus we run them multiple times to extract M replicates that approximate a distribution of potential particle positions at time x_1 . Using this set of potential particle positions, we fit multivariate Gaussian distributions to subsets of the particles.

Consider a set of k particles (the $k - 1$ particles nearest to a central particle) in frame x_0 . These particles are in a bath of $N - k$ particles. To track these k particles in the subsequent frame, we perform the simulations described above. We refer to the resulting simulations as $x_{1,\text{sim}}$. All N particles are simulated in each of the M replicates. The M replicates provide mean positions $\mu_i = \langle x_{1,\text{sim}}^i \rangle$ for each particle i in the set of k particles. Additionally, we compute correlations between each particle in the set of k :

$$C_{ij} = \langle (x_{1,\text{sim}}^i - \mu_i) \cdot (x_{1,\text{sim}}^j - \mu_j) \rangle.$$

Together, the computed means and correlations populate a multivariate Gaussian distribution that describes the probability of finding a set of k particles in a position x :

$$P(x|x_0, \alpha) = \frac{1}{\sqrt{(2\pi)^d |C|}} \exp -\frac{1}{2}(x - \mu)^T C^{-1}(x - \mu), \quad (2)$$

where x is the set of positions of the k particles we are interested in tracking, α is the set of parameters describing the interaction potential, d is the dimension, and μ and C are defined above.

The probability function $P(x|x_0, \alpha)$ differs for each set of k particles. To link particles in one frame to the next, we compute the probability function for each set of k particles and use it to determine which set of k particles in the next frame is most likely to be the same set. Because it is computationally intractable to test every possible ordering of k particles in the frame, we select a set of candidate particles to test. These candidate particles are the p closest particles to the set of k in the first frame.

III. RESULTS ON SIMULATED DATA

We demonstrate our method in several simulated systems. We focus on systems that exhibit long-range attraction because these are systems that do not meet the conditions for accurate tracking with the Crocker-Grier algorithm.

In Fig. 1, we show the results of the correlation tracking method when applied to a Morse pair potential of the form

$$U(r) = D(e^{-2\alpha(r-r_0)} - 2e^{-\alpha(r-r_0)}), \quad (3)$$

with $D = 1$, $r_0 = 1$, and $\alpha = 1$. In all the data shown below, the algorithm was tested on simulated data with 100 particles undergoing Brownian motion with periodic boundary conditions. The volume (or area) fraction of the system was set to 0.5, with r_0 as the particle diameter. Molecular dynamics

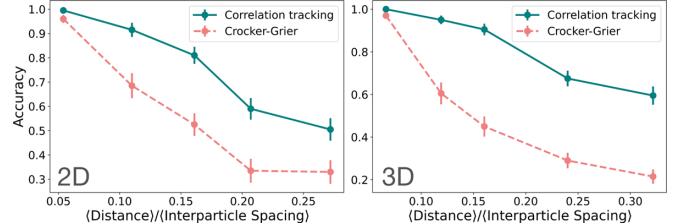


FIG. 1. Comparison of the correlation tracking results to those of the Crocker-Grier algorithm in simulated data using a Morse pair potential [Eq. (3)]. The x axis shows the mean particle displacement divided by the mean interparticle separation in units of r_0 . This metric serves as a proxy for the time and enables ready comparison between simulation and experiment. The y axis shows the tracking accuracy, or the fraction of correctly identified particle labels. The uncertainties are standard errors based on 25 independent simulations.

simulations were performed using JAX-MD. When tracking sets of k particles with p candidate particles, we specified $k = 5$ and $p = 15$ for all data. To estimate the correlation matrix for a given pair of frames, we averaged over 10 000 independent simulations.

As shown in Fig. 1, we observe significant improvement in the tracking accuracy for longer gaps between frames over the state-of-the-art algorithm in both two and three dimensions. We define tracking accuracy as the fraction of correctly identified particle labels.

We hypothesize that the success of our algorithm in this system is owed to the large range of interaction. The greater the interaction range, the more correlations can be established. To demonstrate this, we compare in Fig. 2 the accuracy of our method to that of the Crocker-Grier algorithm as a function of the range of interaction, which is varied by tuning the parameter α in the Morse potential of Eq. (3). Note that α varies inversely with interaction range. The figure shows that a larger interaction range indeed leads to improved performance

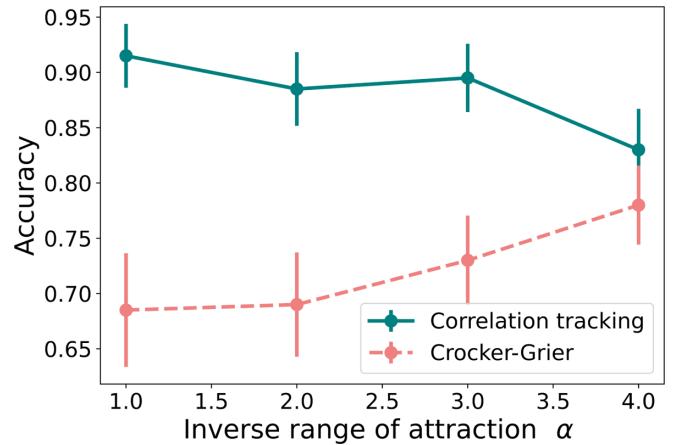


FIG. 2. Comparison of the correlation tracking algorithm to the Crocker-Grier algorithm as a function of interaction range. The results are computed on simulated data using a Morse potential [Eq. (3)]. The x axis shows the parameter α in the Morse potential which varies inversely with interaction range. The y axis shows the fraction of correctly identified particle labels. Error bars are based on 25 independent simulations.

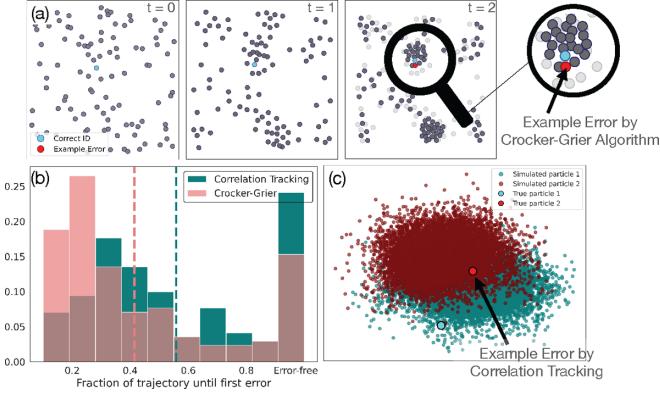


FIG. 3. (a) Sample of a trajectory in which the correlation tracking algorithm makes accurate predictions but the Crocker-Grier algorithm makes an error on the third frame. The particle we are tracking is shown in blue, and the incorrectly identified particle is shown in red. The final frame is overlaid on the previous frame (partially transparent). (b) Histogram showing the length of a given trajectory before the first error is made. The blue data points are results for correlation tracking whereas the pink results are for the Crocker-Grier algorithm. The means of each distribution are given by the dashed vertical lines. (c) Sample error made by the correlation tracking algorithm. Distributions of simulated particle positions are shown in blue and red, respectively, while the true particle positions are given in brighter shades of blue and red (overlaid). The correlation tracking algorithm mistakes the blue particle for the red particle.

by the correlation tracking method and worse performance by the Crocker-Grier algorithm. When the range of interaction is sufficiently small, the accuracies of the correlation tracking algorithm and the Crocker-Grier algorithm converge.

Thus far we have shown results on pairs of simulated frames. To better understand the errors in entire trajectories, we show a histogram of the length of a trajectory before the first tracking error is made in Fig. 3. We perform this test on 10 frame trajectories in two dimensions (2D), with long-range attraction ($\alpha = 1.0$) and with 500 time steps per frame. The conditions are otherwise the same as in the previous simulations. Each data point represents the tracking of a single particle in an independent simulation. We observe that correlation tracking can process longer trajectories without error and has a higher rate of perfect trajectory predictions. We additionally show a segment from a sample trajectory in which correlation tracking makes no errors, but the Crocker-Grier algorithm makes an error in the final frame. Commonly, errors accumulate in regions of increasing density. We thus include a sample trajectory in which aggregation occurs. Correlation tracking correctly predicts motion towards the region with a higher density of particles.

Additionally, we examined the errors made by the correlation tracking algorithm, and find that they commonly arise when the distributions over the simulated positions of two particles strongly overlap. This is illustrated in Fig. 3(c). We note, however, that rarely are the two particles confused for one another. Rather, because we only attempt to track sets of k particles in the system at a time, errors occur when a particle outside the set of k has a distribution of simulated positions that overlaps with a particle within the set of k . Although k

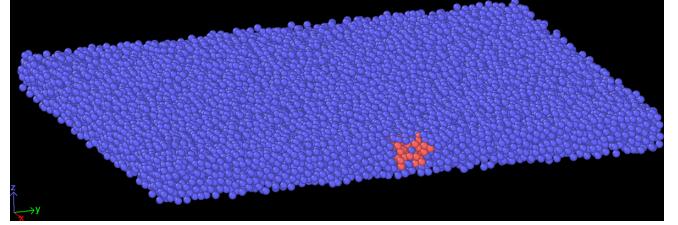


FIG. 4. 3D rendered image of the colloidal particles in the first frame of the tracking experiment. The front half of the sample has been removed to reveal the red particles at the center, which are used for the tracking analysis. The particle diameter is $1.75 \mu\text{m}$.

is currently limited by computational constraints, if the method were made more computationally efficient, a simple way to improve accuracy would be to increase the size of the set of k particles to match the correlation length of the system.

An additional potential source of error in the correlation tracking algorithm occurs if the posterior distribution over particle positions has multiple peaks. Our implementation of the algorithm assumes a normally distributed posterior distribution, and thus cannot currently predict such a distribution. Extending the algorithm to accommodate more general posterior distributions is a potential avenue for future work.

IV. RESULTS ON EXPERIMENTAL DATA

To verify the behavior on experimental data, we tested our tracking algorithm on experimental hard sphere colloids. In this system, since there are no long-range interactions and the motion is largely uncorrelated, we expect that the algorithm should match the performance of the Crocker-Grier algorithm.

The experimental system consists of core-shell colloidal particles undergoing Brownian motion in a density-matched and refractive-index-matched solvent. The core and shell of the particles have the same composition: poly trifluoroethyl methacrylate (pTFEMA) and poly tert-butyl methacrylate (pTBMA). The cores are fluorescently dyed. The core diameter is $0.8 \mu\text{m}$ and the particle diameter is $1.75 \mu\text{m}$. This core-shell design separates fluorescent spots in confocal imaging and ensures accurate particle locating in a dense colloidal suspension. A polyelectrolyte brush is grown on the particle surface to prevent aggregation. The particles are suspended in a mixture of urea and formamide, with tetrabutylammonium bromide (TBAB) added as salt to screen charges and to ensure that the particle interaction closely approximates that of hard spheres. This colloidal solution is sealed in a glass chamber with dimensions $20 \text{ mm} \times 10 \text{ mm} \times 0.4 \text{ mm}$. The particles take up 43% of the total volume.

We use fluorescent confocal microscopy to image the particles. Figure 4 shows part of the imaged volume, which measures $209 \mu\text{m} \times 118 \mu\text{m} \times 10 \mu\text{m}$ and contains about 38 000 particles. The imaged volume is far from the sample boundaries to prevent edge effects. The particle centers are located and tracked using TRACKPY [7], a python package based on the Crocker-Grier algorithm. Because some particles may enter and leave the imaging region through the experiment, we chose 55 particles at the center of the image for analysis, marked in red in Fig. 4. This ensures that the ana-

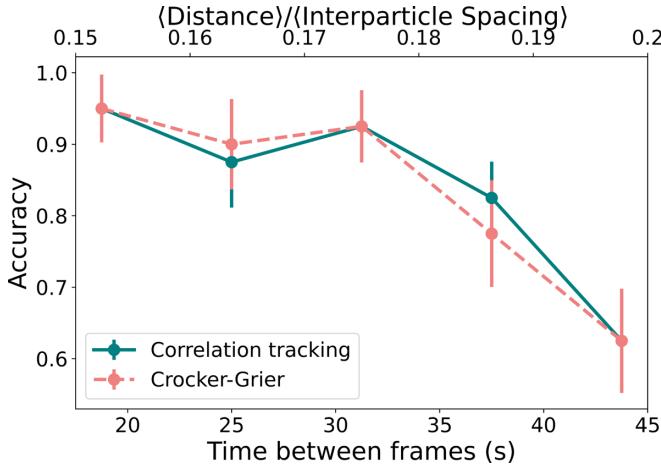


FIG. 5. Comparison of correlation tracking to Crocker-Grier in a three-dimensional experimental hard-sphere system. The primary x axis (bottom) shows the time between frames measured in seconds. The secondary x axis (top) shows the mean particle displacement in units of the particle diameter. The secondary axis is included for easier comparison to simulation. The y axis shows the fraction of correctly labeled particles. The error bars show the standard error, computed for 10 independent tracking events that use different central particles and different starting frames.

lyzed particles are present through all 100 frames and that the size of the experimental dataset matches that of the simulation dataset. The frame rate is 1.25 s per 3D stack. This high frame rate allows accurate tracking by the Crocker-Grier algorithm. We compare frames at different time intervals to test changes in tracking accuracy.

We approximate the hard-sphere behavior in the simulations with a Weeks-Chandler-Andersen (WCA) potential [9], which consists of the repulsive portion of a Lennard-Jones potential. We use this potential instead of an exact hard-sphere interaction to simplify the calculation of the dynamics. We first simulate an ensemble of trajectories starting from the positions in the first frame of the experimental data. This ensemble is used to parametrize a probability distribution over particle positions at a later time. We then label the particles at this later time using the probability distribution we have computed from the simulations. Figure 5 shows the results.

V. RESULTS ON COLORED PARTICLES

Lastly, we consider systems with multiple particle types, such as colloids of different colors. Knowledge of these types can be used to inform both tracking algorithms and offer more accurate predictions. Importantly, introducing more particle types is time-intensive. Below, we show that, by using the correlation tracking algorithm, we reach the same levels of accuracy as the Crocker-Grier algorithm with fewer particle types. Thus, we move human effort (time-intensive experiments) into computer effort (higher computational cost).

First, consider the case of N particles interacting by a Morse potential. Suppose each particle is dyed one of M possible colors. If $M = 1$, we recover the results found above. As M approaches N , all the particles become distinguishable, and it is always possible to track every particle accurately.

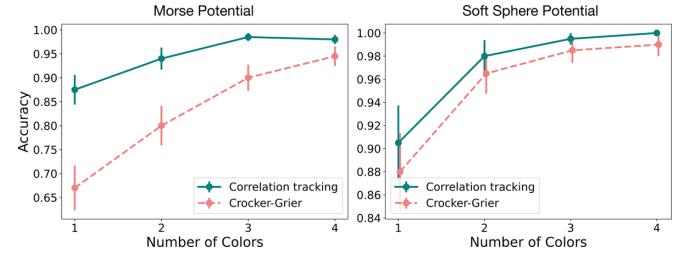


FIG. 6. Comparison of the correlation tracking algorithm to the Crocker-Grier algorithm in simulated data on particles with multiple colors, interacting via a Morse potential (left) and a soft sphere potential (right). The particles were allowed to evolve such that their distance, measured in units of r_0 per average interparticle separation, was 0.11 for the Morse potential, and 0.069 for the soft-sphere potential. The y axis shows the fraction of correctly labeled particles in that frame. Error bars are based on 25 independent simulations.

In the intermediate range, more colors lead to an effectively reduced density without interfering with correlated particle motion. Thus, noting that introducing N independent colors is rarely experimentally feasible, we examine the trade-off between limiting the number of colors and the accuracy of tracking using both the correlated tracking algorithm and the Crocker-Grier algorithm in Fig. 6.

To simulate adding independent colors, we added the concept of species to our simulation. We assume that if a particle is dyed a given color, it may only be mapped to a particle of the same color in the next frame. To enforce this constraint, we treat each color as a different species and track it separately, which prevents errors in labeling across colors in both the correlation and Crocker-Grier tracking algorithms. In this example, the species argument is only relevant to the tracking analysis and not to the simulations themselves; here we assume that differently colored particles interact by a Morse potential in the same way as particles with the same color.

Figure 6 shows that increasing the number of colors increases tracking accuracy for both algorithms. Adding too many colors leads to diminishing returns in tracking accuracy. For the Morse potential, Crocker-Grier achieves very high accuracies with four colors. Correlation tracking improves this and achieves nearly perfect tracking accuracy with only three colors. For the soft sphere potential, Crocker-Grier achieves high accuracy with three colors. The correlation tracking algorithm matches the same level of accuracy Crocker-Grier achieves with only two colors.

VI. DISCUSSION

Our results on experimental data indicate that the correlation tracking algorithm is robust. Even though we used an imperfect description of the hard-sphere potential in the experimental system, the algorithm recovers Crocker-Grier's performance in the diffusive regime. Many algorithms, particularly machine-learning algorithms, require extensive modeling of noise sources in the experimental system to achieve accurate results on experimental data [10–12]. For instance, computer vision algorithms often require modeling noise due to the type of camera and the method of transmitting data [10]. In contrast, we are able to accurately

track experimental data without introducing any noise modeling. Therefore, our model is robust to noise sources in the experimental data, such as polydispersity and positional uncertainty. We note, however, that the algorithm is sensitive to certain types of errors. In particular, because simulations with fixed particle labels are used to interpolate between images, the algorithm relies upon accurate detection and localization of particles. Errors in detection and localization are not currently accounted for in the correlation tracking algorithm. Furthermore, the correlation tracking method relies on an accurate description of the potential. However, we noted that our description of the potential is imperfect, yet we are still able to track particles accurately. As long as the potential is accurate over the range and timescale of the measurements, correlation tracking will be accurate.

The increase in tracking accuracy due to the correlation tracking algorithm comes at a significant computational cost. We mitigate that cost by taking advantage of the automatic vectorization and compilation offered by the package JAX [13,14], so that simulations are run simultaneously. Tracking a set of five particles across one pair of frames with a batch size of 10 000 simulations used to estimate the correlation matrix takes approximately four minutes on a K80 GPU on Google Colab, whereas performing the same analysis with the Crocker-Grier method takes only a few seconds. However, the proof of principle of the correlation tracking algorithm leaves a lot of room for improvements in computational efficiency. For instance, for commonly used implementations of the Crocker-Grier algorithm such as TRACKPY, the computational cost grows only quadratically with the number of particles, whereas the naive implementation would grow factorially. Implementing similar improvements in correlation tracking may significantly reduce computational cost.

Although our method increases computation time, it may reduce the overall experimental effort. As discussed in Sec. III, one method of increasing tracking accuracy is to introduce multiple colors for particles in an experimental system. Creating each color, however, comes at a high cost in experiment time. Since the correlation tracking method significantly reduces the number of colors needed to reach the highest levels of tracking accuracy, it may save significant time.

The correlation tracking algorithm uses simulation in a fundamentally new way. To date, simulations have been used to discover features and mechanisms of material and biological systems that cannot yet be observed in experiment [15,16], to deepen our understanding of experimental results [17,18], or to guide future directions for experimental exploration [19]. This algorithm, however, intertwines simulation and experiment in the data analysis procedure. The simulated data is used to directly increase the accuracy of particle tracking by rooting the data analysis method in the physics of the system. There is a rich space to explore by bringing simulation and experiment in direct contact, a space that we anticipate will contain new lines of research on both sides.

As the correlation tracking algorithm increases tracking accuracy, it allows accurate tracking for longer image acquisition intervals in experiments. This enables a taller imaging region in 3D confocal imaging. For example, in the experiments of this paper, the microscope can reach 40 frames per second while maintaining the necessary image resolution in a xy plane for accurate particle locating. The step size in the z direction needs to be $0.18 \mu\text{m}$ for accurate particle locating in the z dimension. The dynamics of the colloidal system limit the 3D imaging interval to less than 1.5 s to ensure accurate particle tracking using the Crocker-Grier algorithm. These restrictions give a maximum of 60 xy planes or $10.8 \mu\text{m}$ at each time step. With this limitation, only seven particle layers can be scanned for the $1.75\text{-}\mu\text{m}$ -sized particles used in this experiment. A significant portion of the particles moves out of the imaging region in a longer experiment. This makes it difficult to track each particle over a long time and to study kinetic phenomena. However, with correlation tracking, more particle layers can be imaged at each time step, and long-time particle tracking becomes more feasible.

The correlation tracking method promises to be of greatest use in studies with highly correlated motion, such as studies of crystal nucleation or shear thickening in attractive colloidal systems. The algorithm can be applied to any experimental system for which the particle interactions are well characterized. Moreover, recent work has had significant success in approximating interaction potentials from simulated data using machine-learning algorithms [20,21]. If these interaction potentials were learned from experimental data, the correlation tracking algorithm could be used to provide accurate tracking. These approximations need only be accurate over short distance and timescales to be useful for the correlation tracking algorithm. Furthermore, the correlation tracking algorithm could be inverted to provide estimates of interaction potentials in experimental systems: given data with accurate particle tracks, the correlation tracking algorithm would provide a probability that the tracks are well explained by a given interaction potential. One could converge on an accurate potential by proposing a given interaction potential, evaluating the probability that it explains the experimental particle tracks and accordingly updating the proposed potentials, for example, with gradient-based optimization methods.

ACKNOWLEDGMENTS

We thank Peter Lu for many valuable discussions and experimental guidance. This material is based on work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE1745303, the Harvard Materials Research Science and Engineering Center (DMR 20-11754), the Office of Naval Research (ONR N00014-17-1-3029) and the Simons Foundation, awarded to M.P.B., by the Simons Foundation Investigator Award.

[1] L. Rossi, S. Sacanna, W. T. Irvine, P. M. Chaikin, D. J. Pine, and A. P. Philipse, Cubic crystals from cubic colloids, *Soft Matter* **7**, 4139 (2011).

[2] Z. Wang, F. Wang, Y. Peng, Z. Zheng, and Y. Han, Imaging the homogeneous nucleation during the melting of superheated colloidal crystals, *Science* **338**, 87 (2012).

[3] P. Schall, D. A. Weitz, and F. Spaepen, Structural rearrangements that govern flow in colloidal glasses, *Science* **318**, 1895 (2007).

[4] C. Martin, B. Leahy, and V. N. Manoharan, Improving holographic particle characterization by modeling spherical aberration, *Opt. Express* **29**, 18212 (2021).

[5] J. C. Crocker and D. G. Grier, Methods of digital video microscopy for colloidal studies, *J. Colloid Interface Sci.* **179**, 298 (1996).

[6] Y. Gao and M. L. Kilfoil, Accurate detection and complete tracking of large populations of features in three dimensions, *Opt. Express* **17**, 4685 (2009).

[7] D. B. Allan, T. Caswell, N. C. Keim, C. M. van der Wel, and R. W. Verweij, soft-matter/trackpy: Trackpy v0.5.0 (2021).

[8] C. P. Goodrich, E. M. King, S. S. Schoenholz, E. D. Cubuk, and M. P. Brenner, Designing self-assembling kinetics with differentiable statistical physics models, *Proc. Natl. Acad. Sci. USA* **118**, e2024083118 (2021).

[9] H. C. Andersen, J. D. Weeks, and D. Chandler, Relationship between the hard-sphere fluid and fluids with realistic repulsive forces, *Phys. Rev. A* **4**, 1597 (1971).

[10] Z. Yue, H. Yong, Q. Zhao, D. Meng, and L. Zhang, Variational denoising network: Toward blind noise modeling and removal, *Advances in neural information processing systems* **32**, 1 (2019).

[11] J. Wu, S. Ruan, C. Lian, S. Mutic, M. A. Anastasio, and H. Li, Active learning with noise modeling for medical image annotation, in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)* (IEEE, Washington, DC, 2018), pp. 298–301.

[12] S. Gupta and A. Gupta, Dealing with noise problem in machine learning data-sets: A systematic review, *Procedia Comput. Sci.* **161**, 466 (2019).

[13] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, JAX: composable transformations of Python + NumPy programs (2018).

[14] S. S. Schoenholz and E. D. Cubuk, Jax MD: A framework for differentiable physics, in *Advances in Neural Information Processing Systems* (Curran Associates, Inc., 2020), Vol. 33.

[15] C. L. Porter, S. L. Diamond, T. Sinno, and J. C. Crocker, Shear-driven rolling of dna-adhesive microspheres, *Biophys. J.* **120**, 2102 (2021).

[16] C. X. Du, G. van Anders, R. S. Newman, and S. C. Glotzer, Shape-driven solid–solid transitions in colloids, *Proc. Natl. Acad. Sci.* **114**, E3892 (2017).

[17] L. Li, C. Goodrich, H. Yang, K. R. Phillips, Z. Jia, H. Chen, L. Wang, J. Zhong, A. Liu, J. Lu *et al.*, Microscopic origins of the crystallographically preferred growth in evaporation-induced colloidal crystals, *Proc. Natl. Acad. Sci.* **118**, e2107588118 (2021).

[18] D. J. Audus, S. Ali, A. M. Rumyantsev, Y. Ma, J. J. de Pablo, and V. M. Prabhu, Molecular Mass Dependence of Interfacial Tension in Complex Coacervation, *Phys. Rev. Lett.* **126**, 237801 (2021).

[19] O. Kimchi, C. P. Goodrich, A. Courbet, A. I. Curatolo, N. B. Woodall, D. Baker, and M. P. Brenner, Self-assembly–based posttranslational protein oscillators, *Sci. Adv.* **6**, eabc1939 (2020).

[20] J. Vandermause, Y. Xie, J. S. Lim, C. J. Owen, and B. Kozinsky, Active learning of reactive Bayesian force fields: Application to heterogeneous hydrogen-platinum catalysis dynamics, *arXiv:2106.01949*.

[21] H. Chan, B. Narayanan, M. J. Cherukara, F. G. Sen, K. Sasikumar, S. K. Gray, M. K. Chan, and S. K. Sankaranarayanan, Machine learning classical interatomic potentials for molecular dynamics from first-principles training data, *J. Phys. Chem. C* **123**, 6941 (2019).