Soft Matter

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/softmatter

Interaction Potentials from Arbitrary Multi-Particle Trajectory Data

Ian C. Jenkins, John C. Crocker and Talid Sinno[1]

Department of Chemical and Biomolecular Engineering,
University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

Understanding the complex physics of particle-based systems at the nanoscale and mesoscale increasingly relies on simulation methods, empowered by exponential advances in computing speed. A major impediment to progress lies in reliably obtaining the interaction potential functions that control system behavior—which are key inputs for any simulation approach—and which are often difficult or impossible to obtain directly using traditional experimental methods. Here, we present a straightforward methodology for generating pair potential functions from large multi-particle trajectory datasets, with no operational constraints regarding their state of equilibration, degree of damping or presence of hydrodynamic interactions. Using simulated datasets, we demonstrate that the method is highly robust against trajectory perturbations from Brownian motion and common errors introduced by particle tracking algorithms. Given the recent rapid pace of advancement in high-speed and three-dimensional microscopy and associated particle tracking algorithms, we anticipate a near future experimental regime where easily collected high-dimensional trajectory sets can be rapidly converted to the detailed interaction and hydrodynamic force fields required to replicate the system's physics in simulation.

[1] Corresponding author: talid@seas.upenn.edu, Tel: (215) 898-2511

I. INTRODUCTION

The ensemble of trajectories for a system of interacting particles contains, at least in principle, detailed information about particle-particle, particle-boundary, and particle-external field interactions. Extracting *useful* information from this very high dimensional dataset – three times the number of particles in the most general case – is a fundamental 'big-data' challenge. The forward problem, which comprises the mapping of high-dimensional particle trajectories onto macroscopically meaningful, low-dimensional properties such as diffusion coefficients, density distribution functions, or thermodynamic properties, is well established within the realm of statistical mechanics. In this case, a suitable average over a portion of the degrees-of-freedom directly produces the desired property. While some quantities are more difficult to compute, e.g., free energies, the procedure is nonetheless quite straightforward.

By contrast, the inverse problem associated with determining the inter-particle interaction potential of a given system from measures of its structural and thermodynamic properties is far more challenging[1-4]. For atomic and molecular systems, the canonical approach is to postulate a functional form for the potential based on an understanding of the interaction physics and then apply a database of properties (e.g., cluster energies, cohesive energies, elastic constants, phase stability ordering, defect energies, etc.), obtained by experimental measurements and/or electronic structure calculations[2, 5-9], to fit multiple adjustable parameters. The potential may be expressed analytically in closed-form based on some physical arguments[7, 10], or by sets of interpolating functions or tables[11-13], or some combination of both. This overall strategy for atomic/molecular potential development has led to the development of a vast array of highly successful empirical and semi-empirical potentials for an immense number of material systems.

For disordered atomic systems like vapors, liquids, and glasses, the ability to measure experimentally, e.g., with x-ray or neutron diffraction, the structure factor, $S(q)$, and therefore its real-space equivalent, the pair distribution function, offers a more 'direct' approach for potential construction[14-18]. In this regard, a large number of methods have been proposed to iteratively compute an

interaction (usually pairwise) potential by comparing a simulated pair distribution function to the target one. Many variations on this general theme have been proposed, which differ in the initially assumed form of the potential, the nature of additional information input in the form of constraints, and the particular details of the iterative strategy used to converge to a final pair potential[1, 19-21]. Overall, these approaches have been successful in generating pair potentials that reproduce both structural and thermodynamic measures of the target system. However, they also are generally quite computationally demanding because the pair distribution function, by its very nature, requires extensive configurational sampling, and therefore long simulations, to determine accurately as the trial potential is being refined. Moreover, uncertainties in the input pair distribution function also pose a problem as good potential extraction requires that the target pair distribution function be known well across the interaction range[18, 22]. It should also be noted that the satisfactory (implicit) capture of many-body interactions with effective pair-potentials generated from pair distributions is far from guaranteed, and in any case requires input that may be difficult to obtain experimentally or by other means.

At the colloidal scale, inter-particle interactions have classically been inferred from experiments on analogous macroscopic surfaces in a surface force apparatus[23]. More recently, the interactions of multi-micron sized particles (with each other or with flat walls) have been measured directly using an AFM cantilever[24], or using thermal fluctuations with total internal reflection microscopy (TIRM)[25-29]. Colloidal interactions also can be inferred from optical tracking of pairs of micron-sized particles in optical tweezers[30]; specifically from their equilibrium separation distributions in line optical tweezers[31-35] or from their non-equilibrium motion in blinking optical tweezers[36, 37]. These methods come with technical challenges: the AFM or optical instrumentation required to make these measurements is often rather complex compared to a simple imaging system, many interesting particles have sizes, shapes or compositions that are not amenable to optical manipulation or have interactions that are too strong or weak to be readily measured with one or more of the above methods. Finally, analogous to the methods for disordered atomic systems, colloidal interactions also may be inferred from equilibrium distributions

of ensembles of particles[38-40]. However, these approaches are limited to thermal energy scale potentials, require careful treatment of liquid structure effects, and exhibit demanding constraints on statistical sampling.

Following this long line of methodological developments, here we describe a new approach for extracting interaction potentials from particle trajectory data alone. In essence, particle trajectories are used to compute numerical estimates of positional derivatives, i.e., velocities and/or accelerations, which are then used to infer forces as a function of inter-particle separation. The method does not place a constraint on the number of interacting particles being observed, nor on whether the trajectories represent equilibrium or non-equilibrium conditions, and only requires that the particle positions be recorded as a function of time with sufficient fidelity. As we will demonstrate throughout the paper, the current approach is flexible enough to enable the consideration of various dynamical situations, as long as some information regarding the nature of the dynamics is available.

The approach described here shares some important aspects with, and differs in key ways from, the force-matching (FM) technique originally proposed by Ercolessi and Adams[41] for the parametrization of empirical potentials using *ab initio* data, and later generalized into a powerful coarse-graining framework by Voth and coworkers[42-44] and others[45-48]. Essentially, both the present approach and the numerous FM variants seek to fit a pairwise force field using information obtained from some reference system. However, while the FM technique uses forces computed from configurations obtained from simulations performed with a reference (known) force-field to fit a simpler one, the present approach numerically estimates forces from trajectories of particles that are subject to some *unknown* interparticle interaction force-field. Moreover, here we consider the possibility that the particle trajectories used to approximate forces maybe additionally be impacted by thermal fluctuations, measurement uncertainty, and hydrodynamic effects.

The remainder of the paper is structured as follows. The general methodological details are presented in Section II. In Section III, we first consider the case of noiseless trajectories in both the fully

inertial and overdamped limits. In Section IV, the impact of noise on particle trajectories is considered. We first discuss noise produced by thermal fluctuations (IV.A) and then consider noise introduced by positional measurement uncertainty (IV.B). In Section V, we address the impact of hydrodynamic coupling between particle trajectories. Finally, conclusions are presented in Section VI.

2. METHOD

Consider a system of two particles, *i* and *j*, with known trajectories, interacting through a pairwise potential, $U(r_{ij})$, where r_{ij} is the particle center-to-center separation. If these particles are otherwise isolated, this pairwise interaction is entirely responsible for the net force, $f_{i,\alpha}$, that each particle, *i*, experiences along each coordinate direction, α. For this simple two-particle system, the force profile as a function of interparticle separation distance is given by

$$F(r_{ij}) = -\frac{\partial U}{\partial r_{ij}} = f_{i,\alpha}\frac{r_{ij}}{r_{i,\alpha} - r_{\mathrm{j},\alpha}}, \tag{1}$$

where $F(r_{ij})$ is the force along the center-to-center direction and $r_{i,\alpha}$ is the α-coordinate of particle *i*. Repeating this calculation for many separation distances and binning $F(r_{ij})$ over discrete values of $r_{i,\alpha}$ produces an approximation to the complete force profile. As shown in Eq. (1), the interaction potential energy function then may be generated by integrating over the discretized force profile.

Now consider a system of three or more particles, again with known trajectories, where the total force acting on each particle is attributed to multiple interactions. To extract a pairwise force profile from such a system we first rewrite Eq. (1) for a particle experiencing an arbitrary number of interactions, i.e.,

$$f_{i,\alpha} = \sum_{\substack{j=1 \\ j\neq i}}^{n} \frac{r_{i,\alpha} - r_{j,\alpha}}{r_{ij}} F(r_{ij}), \tag{2}$$

which may be restated as the system of linear equations,

$$\mathbf{f}_\alpha = \mathbf{C}_\alpha \mathbf{F} . \tag{3}$$

Here, $\mathbf{f}_\alpha \in \mathbb{R}^n$ represents the forces acting on all n particles along a single Cartesian direction α, $\mathbf{F} \in \mathbb{R}^{n(n-1)}$ is the vector of pairwise forces between each particle pair, and $\mathbf{C}_\alpha \in \mathbb{R}^{n \times n(n-1)}$ is a matrix with coefficients

$$C^{\alpha}_{i,(jk)} = \begin{cases} \dfrac{r_{i,\alpha} - r_{j,\alpha}}{r_{ij}} & k = i \\ 0 & k \neq i \end{cases} . \tag{4}$$

In principle, the system of equations (3) may be written and solved for $\mathbf{F}$ independently along each direction α, or as a combined system along all 3 directions simultaneously, i.e.,

$$\mathbf{f} = \mathbf{C}\mathbf{F} . \tag{5}$$

However, while the number of constraints (rows in $\mathbf{C}$) increases linearly with particle count, the number of unknowns (columns) scales as the number of particle *pairs*. Consequently, eq. (3) is, in general, singular for n>3. As described in the force matching literature[41], the problem may be recast so that the unknowns are the coefficients, g_m, of a discretized approximate function, $\tilde{F}(r_{ij})$, where

$$\tilde{F}\left(r_{ij}\right) = \sum_{m=1}^{M} g_m \phi_m \left(r_{ij}\right), \tag{6}$$

and $\phi_m \left(r_{ij}\right)$ are a set of M basis functions used to construct the approximation. The choice of basis for $\tilde{F}(r_{ij})$ has been discussed in some detail in Ref. [45]. Here, we consider, for simplicity, a "square wave" basis in which the basis functions are constant over each interval between adjacent discretization points, i.e.,

$$\phi_m^{sq}(r) = 1 , \tag{7}$$

as well as a linear basis (defined on the unit interval [-1,1]),

$$\begin{aligned}\phi_{m,1}^{lin}(r) &= \frac{1}{2}(1-r)\\ \phi_{m,2}^{lin}(r) &= \frac{1}{2}(1+r)\end{aligned}. \tag{8}$$

Equation (2) now becomes

$$f_{i,\alpha} = \sum_{\substack{j=1\\ j\neq i}}^{n}\left[\frac{r_{i,\alpha}-r_{j,\alpha}}{r_{ij}}\sum_{m=1}^{M} g_m \phi_m\left(r_{ij}\right)\right], \tag{9}$$

or

$$\mathbf{f} = \mathbf{C}^{*}\mathbf{G}, \tag{10}$$

where $\mathbf{f} \in \mathbb{R}^{3N}$, $\mathbf{G} \in \mathbb{R}^{M}$ and $\mathbf{C}^{*} \in \mathbb{R}^{3N\times M}$. The system of equations (10) is over-constrained for $M<3N$ so that $\mathbf{G}$ may be obtained using the least-squares approximation, i.e.,

$$\mathbf{G} = \left(\mathbf{C}^{*T}\mathbf{C}^{*}\right)^{-1}\mathbf{C}^{*T}\mathbf{f}. \tag{11}$$

In the case where the force vector, $\mathbf{f}$, is obtained from simulation configurations using a known interaction potential, Eq. (10) corresponds to the well-established technique of force matching[42-48], which is usually applied to match a coarse-grained pair force model, $\tilde{F}(r_{ij})$, to forces generated by a more detailed interaction model. Here, however, we consider a fundamentally different application for Eq. (10): *can one robustly and accurately calculate a pairwise force model from an experimentally measured set of particle trajectories subject to unknown interparticle interactions*? Consider first the general case of a system of particles that are subject to Langevin dynamics dictated by a combination of interparticle forces as well as forces due to the presence of an implicit solvent, i.e.,

$$m\ddot{\mathbf{r}} = \mathbf{f} - \frac{k_B T}{D}\dot{\mathbf{r}} + \left(\sqrt{2\gamma m k_B T}\right)\mathbf{R}, \tag{12}$$

where $\mathbf{r}$ is the time-dependent vector of particle positions, $\mathbf{f}$ represents interparticle forces, D is the single particle diffusivity, k_BT/D is the damping, or friction, coefficient ($k_BT/D \equiv \gamma$) and $\mathbf{R}(t)$ is a random Brownian force modeled as a delta-correlated Gaussian process with zero mean so that $\langle \mathbf{R}(t) \rangle = 0$ and $\langle \mathbf{R}(t)\mathbf{R}(t') \rangle = \delta(t-t')$.

Assuming, for example, that the dynamics are noise-free and fully inertial, so that $\mathbf{R} = 0$ and $k_BT/D = 0$, particle forces therefore are dictated by Newton's Second Law of Motion, $\mathbf{f} = m\ddot{\mathbf{r}}$. Substitution into Eq. (11) then gives

$$\mathbf{G} = \left(\mathbf{C}^{*T}\mathbf{C}'\right)^{-1}\mathbf{C}^{*T}\left(m\ddot{\mathbf{r}}\right), \tag{13}$$

in which the vector of particle accelerations, $\ddot{\mathbf{r}}$, may be estimated from three sequential snapshots of the trajectories using a second-order central difference approximation,

$$\ddot{r}_{i,\alpha}(t) \sim \frac{r_{i,\alpha}(t-\Delta t) - 2r_{i,\alpha}(t) + r_{i,\alpha}(t+\Delta t)}{\Delta t^2} + O\left(\Delta t^2\right) \tag{14}$$

where $r_{i,\alpha}(t)$ is the position of particle i in direction α at time t, and Δt is the time interval between sequential snapshots. In practice, the minimum appropriate time interval between trajectory observations is dictated by trajectory fluctuations due to measurement error and/or Brownian motion; the impact of trajectory noise is discussed in Section IV. Once $\ddot{\mathbf{r}}$ has been calculated, Eq. (13) is solved for the pairwise force profile and interaction potential. Equation (11) may be similarly applied to overdamped systems where $\mathbf{R} = 0$ and $\ddot{\mathbf{r}} = 0$, so that $\mathbf{f} = \gamma\dot{\mathbf{r}}$. Once again, substitution of the equation of motion into Eq. (11) gives a system of equations that may be used to extract the force and interaction potential profiles, i.e.,

$$\mathbf{G} = \left(\mathbf{C}^{*T}\mathbf{C}^{*}\right)^{-1}\mathbf{C}^{*T}\left(\gamma\dot{\mathbf{r}}\right), \tag{15}$$

Note that only two snapshots are required to compute the velocity, $\dot{\mathbf{r}}$ in Eq. (15), in contrast to the three snapshots needed for calculation of the acceleration.

3. NOISELESS DYNAMICS

In order to demonstrate the application of Eqs. (13) and (15) for noiseless trajectories, we consider a periodic system containing 64 Lennard-Jones (LJ) particles at reduced particle density, $\rho^* \equiv \rho\sigma^3 = 0.04$, and reduced temperature, $T^* \equiv k_B T / \varepsilon = 1$; these conditions correspond to a homogeneous fluid phase. The LJ potential function is truncated at 4.0σ, and a cubic polynomial function is used to smoothly zero the potential at 4.5σ. Newtonian and overdamped dynamics were simulated in the NVE (constant density and Energy) and NVT (constant density and Temperature) ensembles, respectively. Particle configuration snapshots were stored at $\Delta t = 0.001\sqrt{m\sigma^2/\varepsilon}$ intervals (particle mass, *m*, and the potential parameters, σ and ε, are all set to unity throughout). The impact of the time step is discussed in the context of other experimental constraints in Section 4.C. The pairwise force profile and associated interaction potential function were computed with Eq. (13) for inertial dynamics and Eq. (15) for overdamped dynamics with $\gamma = 1$). The basis function set was chosen to be a series of 60 square waves over the interval $0 \le r \le 4.5\sigma$, each with width 0.075σ. The profiles were smoothed by repeating the force evaluation over 500 sets of sequential snapshots, each time shifting the trajectory observations by Δt. The extracted profiles for both the inertial and overdamped cases are shown in Figure 1 (symbols), along with the corresponding input profiles (lines); the agreement generally is excellent across the range of the interaction.

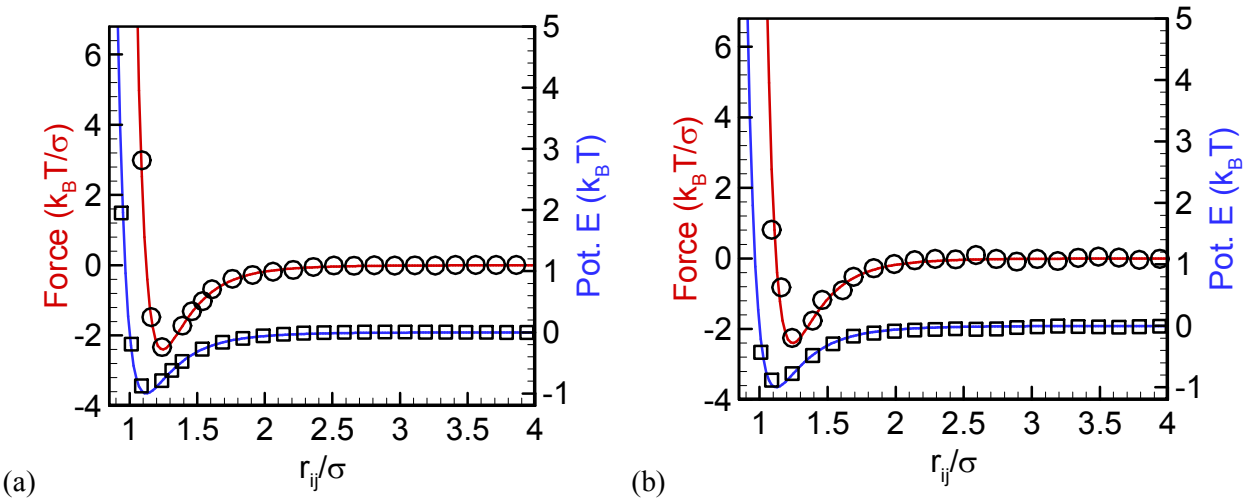


Figure 1: Pair interaction potentials (blue lines and squares) and force profiles (red lines and circles) extracted from observing a system of 64 Lennard-Jones particles evolving via (a) inertial dynamics and (b) overdamped dynamics. Extracted profiles, which are generated using 60 0.075σ-width square wave basis functions, are shown by symbols; input profiles are denoted by the solid lines. 500 force evaluations were used to construct the profiles in each case.

The *numerical* error associated with extracting a force profile from a noiseless trajectory (i.e., without Brownian fluctuations or any measurement uncertainties) is dependent on two primary factors: (1) the type and number of basis functions used to discretize the force profile, and (2) the number of force evaluations over which the extracted force profile is averaged. The latter factor is relevant because as the number of samples increases, the average particle pair separation sampled in each discrete interval converges to the interval center. The dependence of the numerical error in the force profile as a function of these factors is shown in Figure 2. As expected, the error decreases as the number of force evaluations increases. However, as with any spatial discretization technique, the minimum (systematic) error achieved depends on the number (and type) of basis functions used. Indeed, as the number of intervals is increased from 5 to 60, the residual numerical error decreases from ~1 to 0.03. Also shown in Figure 2, given an identical number of basis functions and force evaluations, the linear basis set always produces a

more accurate force profile then the square wave basis set. Importantly, the linear basis set error also converges to its minimum value with approximately an order-of-magnitude fewer force evaluations than the square wave case. Interestingly, the number of particles considered in the system does not appear to significantly impact the statistical quality of the extracted profiles, presumably because additional particles (beyond the 64 considered here) do not sample distinct configurations. We emphasize again that the errors shown in Figure 2 are intrinsic to the numerical procedure used to extract the force profile from exact, noiseless particle trajectories; additional errors due to experimental measurement uncertainties and thermal trajectory fluctuations are addressed in detail below.

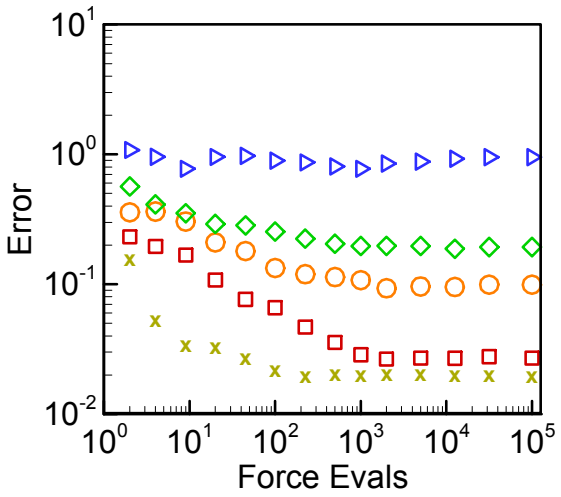


Figure 2: Error as a function of total trajectory data points for a system of 64 Lennard-Jones particles evolving via inertial dynamics. Error is calculated as $\left\|\mathbf{F}^{*}-\mathbf{F}\right\|_{2}/\sqrt{M}$, where $\mathbf{F}^{*}$ contains the force calculated from Eq. (13) at the midpoint of each basis function, $\mathbf{F}$ is the actual force at each of these points, and M is the number of comparison points (bins). Error is computed over the range $1.2\sigma \le r \le 4.5\sigma$, which is sampled by all trajectories. Four discretization levels for the square wave case were considered: 60 basis functions (red squares), 20 basis functions (orange circles), 10 basis functions (green diamonds) and 5 basis functions (blue triangles) over the interval $0 \le r \le 4.5\sigma$. Also shown is the error for the 60 line-segment basis function set (gold crosses).

4. TRAJECTORY NOISE

A. Thermal Fluctuations

In many cases (e.g., nanoparticles in solution), particle trajectories are subject to solvent-induced thermal fluctuations. Returning to the Langevin equation [Eq. (12)], we now consider the same system used to generate the dynamics for Figure 1, but with non-zero Brownian fluctuations and with friction coefficient $\gamma = 1$, so that the governing Brownian dynamics are given by $\mathbf{f} = \gamma \dot{\mathbf{r}} - \mathbf{R}$. Shown in Figure 3 are the force profile and potential function extracted from a set of BD trajectory data. The quality of agreement between the input and output force and interaction potentials is very similar to what is seen in the noiseless overdamped case. Note that we do not consider here the case of correlated Brownian fluctuations (i.e., Stokesian dynamics[49]), where systematic impacts on the extracted profiles may be present. The inset of Figure 3 shows the (subtractive) difference in the error between the noiseless, overdamped and the BD cases, which represents the scatter in the extracted profiles due entirely to the thermal fluctuations in the particle trajectories. For small force evaluation counts (<50), the presence of thermal fluctuations in the trajectories does lead to higher statistical errors, but the same residual (numerical) error as in the noiseless case can be achieved once $O(10^2)$ force evaluations are used. Further discussion of the errors associated with thermal fluctuations is provided along with the discussion of measurement uncertainties in the following sections.

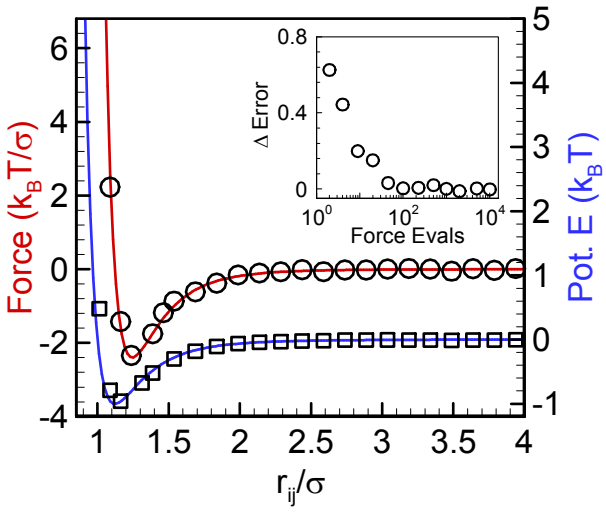


Figure 3: Potential function (blue line and squares) and force profile (red line and circles) extracted from a system of 64 Lennard-Jones particles evolving via Brownian dynamics. Extracted profiles, which are generated using 60 0.075σ-wide square wave basis functions, are shown by symbols; input profiles are denoted by solid lines. 500 force evaluations were used to construct the profiles. Inset: Error difference in the force profiles extracted from overdamped (fluctuation free) and Brownian dynamics trajectories.

B. Measurement Uncertainty

The practical application of the present method to experimental observations of particle trajectories in a wide variety of settings depends crucially on its robustness with respect to measurement uncertainty. The difference between measurement uncertainty and thermal fluctuations is qualitatively apparent by considering that thermal fluctuations act cumulatively over time to progressively alter particle trajectories, while measurement uncertainty is reset every snapshot. Here, we consider a situation in which each snapshot of particle positions is subjected to artificial 'measurement uncertainty' by displacing each particle a different random vector with an average magnitude of 3%, 30%, and 150% of the average particle displacement between sequential snapshots.

Shown in Figure 4 are force profiles, extracted from inertial, thermal fluctuation-free dynamics trajectories (all system parameters remain as given previously) for each of the three 'measurement uncertainty' amplitudes, extracted using either 500 or 10,000 force evaluations. Two important features are apparent. First, as expected, the scatter in the extracted profiles grows as the trajectory uncertainty grows – it becomes difficult to even discern any type of force profile with 500 evaluations for the 150% uncertainty case. However, as the number of evaluations is increased to 10,000, a high-quality force profile is once again obtained.

Interestingly, as the measurement uncertainty magnitude increases, it becomes evident that the extracted force profile does not converge to the input profile and becomes increasingly distorted, exhibiting a deeper attraction well (dashed blue lines). In fact, the converged force profiles are empirically found to correspond to LJ force profiles multiplied, or magnified, by a constant larger than unity that depends on the magnitude of trajectory uncertainty (see Figure 5). The source of the force profile magnification may be qualitatively understood by considering that particle forces are calculated from observed particle displacements between trajectory snapshots. When particle coordinates are perturbed randomly, which is the case for measurement error, the average distance a particle is observed to travel between snapshots increases on average. Put another way, although the random perturbation averages to zero at a given time instant, the net perturbation *across a time interval* does not. This effect causes the apparent acceleration and velocity of the particle to increase, which is in turn reflected by the magnification of the extracted force profile.

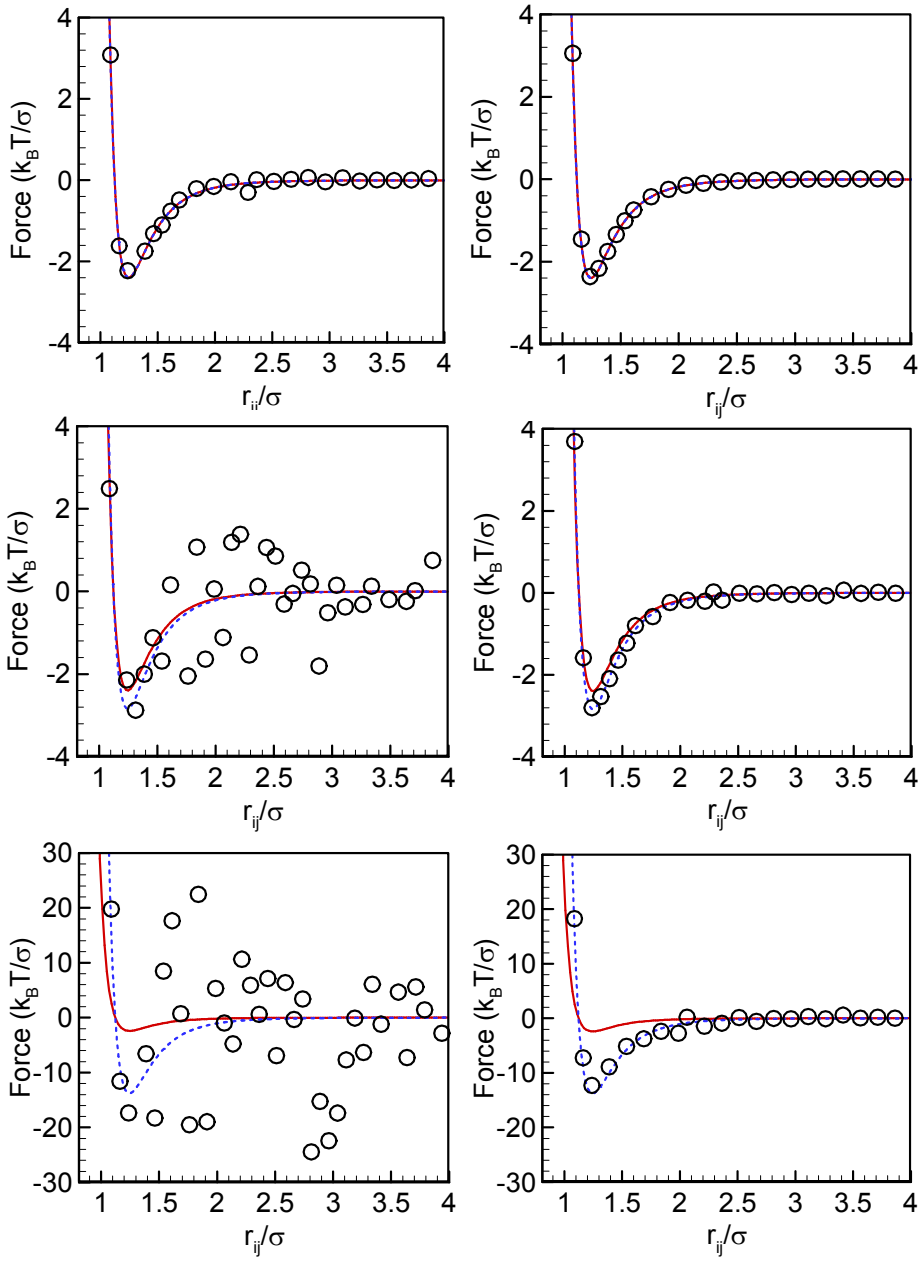


Figure 4: Force profiles extracted from inertial trajectories free of thermal fluctuations but subject to measurement uncertainty using 500 (left column) or 10000 (right column) force evaluations. Measurement uncertainty magnitude is 0.03 (top row), 0.3 (middle row), and 1.5 (lower row) of the mean particle displacement between two successive observations. In each panel, the input force profile is shown as a solid red line. The dashed blue line represents the best-fit LJ force profile using a single

scalar multiplier. All extracted profiles are generated using 60 square wave basis functions of width 0.075σ.

A key question now arises: how do we extract the undistorted force profile in the (usual) situation for which the measurement uncertainty is not *a priori* known? We define μ as the multiplicative factor (>1) that the actual force profile needs to be multiplied by in order to correspond the extracted force. As shown in Figure 5, the reduced force magnification factor, $\mu^* \equiv \mu - 1$, scales as the square of the noise amplitude, so that

$$\mu^* = a\eta^2, \tag{16}$$

where η is the average effective particle displacement due to measurement error relative to mean particle displacements magnitude between sequential snapshots.

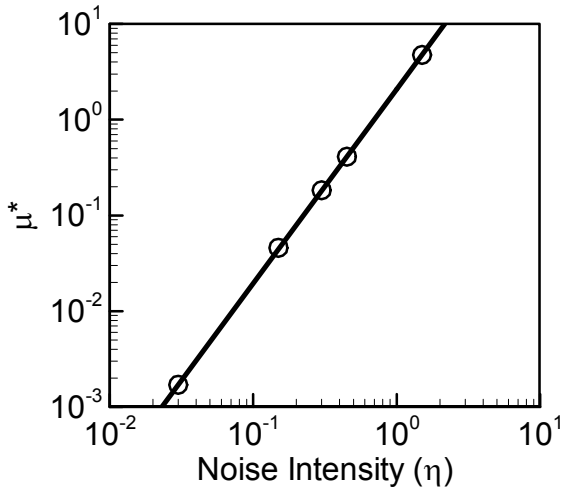


Figure 5: Reduced magnification factor as a function of noise intensity amplitude; black line shows quadratic fit (see text for details).

Now consider a sequence of particle positions measured at regular time intervals along some trajectory that are used to compute two force profiles – one in which the data at every interval is used and

a second in which only every other observation is used, effectively doubling the time between snapshots. It then follows that η in the second case will be reduced by a factor of two, i.e., $\eta_2 = 0.5\eta_1$, assuming that the particle velocity is constant across the time step interval. The validity of this assumption is subject to constraints on the time step size (or the imaging framerate; see Section 4.C) and the rate of velocity change or acceleration, i.e., $a\Delta t / v = \Delta v / v << 1$. The relative magnification effect between the two extracted profiles (which is known) is then given by

$$\mu_r = \frac{\mu_2^* + 1}{\mu_1^* + 1} = \frac{1 + a\left(\eta_1/2\right)^2}{1 + a\eta_1^2}. \tag{17}$$

Rearranging for the reduced magnification factor for the first profile then gives

$$\mu_1^* = a\eta_1^2 = \frac{\mu_r - 1}{1/4 - \mu_r} - 1. \tag{18}$$

Equation (18) was tested by considering a relative noise intensity of 0.525 (corresponding to a reduced magnification factor of 0.58) applied to the inertial dynamics situation considered in Figure 4. Shown in Figure 6 are force profiles extracted with (squares) and without (circles) correcting for measurement uncertainty using Eq. (18). It is readily apparent that the magnification correction provided by Eq. (18) robustly accounts for the distortion, although some scatter remains in the extracted profile – this is addressed next.

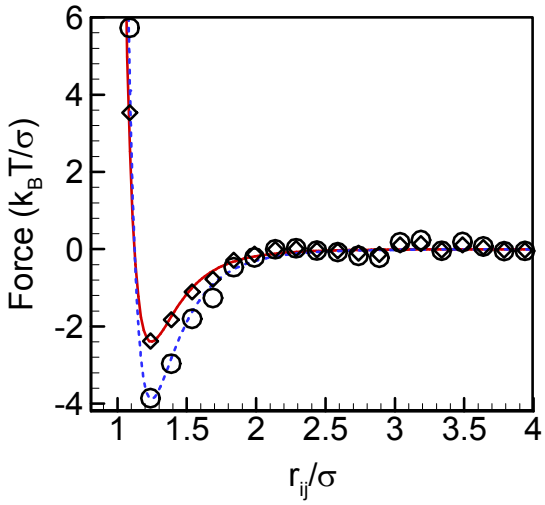


Figure 6: Force profiles extracted from observing a system of 64 Lennard-Jones particles evolving via inertial dynamics subject to measurement uncertainty of amplitude 0.525. Extracted profiles are shown by symbols (uncorrected force – circles, corrected force – diamonds), the input force profile is denoted by the solid red line. The dashed blue line represents the best-fit LJ force profile for the uncorrected force assuming that the input force is scaled by a single multiplier of 1.58. Both extracted profiles are generated using a set of 60 square wave basis functions of width $0.075\,\sigma$.

The residual impact of measurement uncertainty on the error in the extracted force may be assessed by computing the difference in the force profile error obtained for noisy (Figure 4) and exact (Figure 1(a)) inertial dynamics trajectories. Figure 7 shows the subtractive error difference between these two cases as a function of evaluation count for several noise amplitudes. As in the case for thermal fluctuations (see inset of Figure 3), measurement uncertainty or trajectory noise leads to additional scatter in the extracted profiles that may be systematically reduced with additional force evaluations.

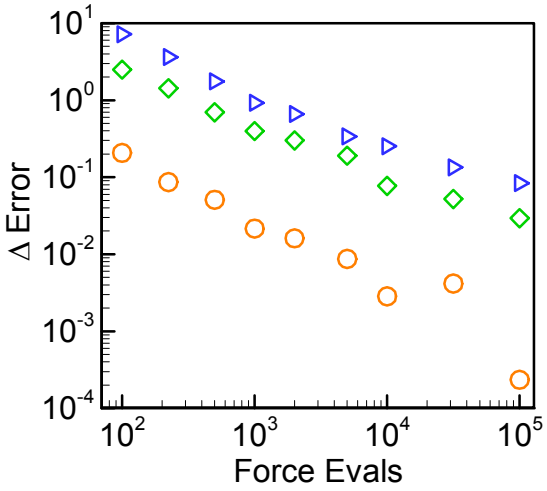


Figure 7: Difference in the force profile error extracted from noisy and exact inertial dynamics trajectories as a function of the force evaluation count calculated for several measurement uncertainty amplitudes (relative to average particle displacement): orange circles – 3%, green diamonds – 30%, blue triangles – 150%. Note that the errors shown here are exclusive of the magnification effect described in the text (which has been corrected for using Eq. (18)).

C. Error Analysis in the Context of Experiment

The preceding analyses demonstrate that both thermal fluctuations and trajectory measurement noise increase scatter in the extracted force profiles, but which may be reduced by increasing the number of force evaluations used to construct the profile. The impact of such scatter in the context of realistic experimental constraints is the subject of this section. We begin by considering the trajectory impact of Brownian diffusion and measurement uncertainty, relative to that of interparticle interactions (the signal). In this context, two noise-to-signal ratios may be defined,

$$\alpha \equiv \frac{\sqrt{Df}}{v}, \tag{19}$$

and

$$\beta \equiv \frac{e_M f}{v}, \tag{20}$$

where f is the framerate at which images may be recorded, e_M is the trajectory displacement due to measurement error, D is Brownian diffusivity, and v is the signal, or drift, velocity, i.e., the velocity due to interparticle interactions. Consequently, α represents the ratio of apparent diffusion velocity (over a time interval dictated by the framerate) and the drift velocity which may represent either overdamped or inertial particle dynamics. Similarly, β represents the ratio of an effective velocity due measurement uncertainty relative to the signal velocity, v.

For reference, consider a system of 1 μm beads suspended in water at 300K being video imaged at a framerate of 50 Hz. For such a system, a typical measurement uncertainty of 0.01 μm is expected (~1% of the particle size), along with a Brownian diffusivity of ~0.5 $\mu m^2/s$. Consequently, for a single force evaluation the conditions $\alpha,\beta<<1$ require that $v>>1$ μm/s, i.e., the interparticle force must be many $k_B T$ in magnitude to be captured accurately. However, as evidenced by the convergence of the profiles in Figures 4, even if α and/or β are not small, the associated error may be systematically reduced by increasing the number of force evaluations. The 150% uncertainty case in Figure 7, for example, corresponds to $\alpha=0$, $\beta=1.5$, and shows clearly that the force profile may be recovered well beyond the O(1) error expected from Eq. (20), as long as a sufficient number of force evaluations are used. The convergence of force profiles extracted from noisy trajectories with respect to the number of evaluations is readily understood by considering Eqs. (19) and (20) – the averages of both D and e_M tend to 0 as the number of samples increases.

The expressions in Eqs. (19) and (20) suggest that α and β also may be arbitrarily reduced by simply lowering the imaging framerate. In fact, this is not the case – the distance a particle travels between sequential snapshots (either by drift or by diffusion) also must be small relative to the length

scale of the potential function, L, otherwise the extracted potential will be blurred. Two additional constraints for extracting accurate force profiles may therefore be identified as

$$\gamma \equiv \frac{v}{Lf} << 1, \tag{21}$$

$$\delta \equiv \frac{1}{L}\sqrt{\frac{D}{f}} << 1. \tag{22}$$

Here, γ and δ represent ratios of the drift and diffusion distances, respectively, relative to the interparticle potential length scale, L. For v~1 μm/s, D~1 μm^2/s, and a potential function of order the particle size (1 μm), the lower bound on the framerate due to drift and diffusion is f~1 s^{-1}. However, as the potential becomes shorter ranged, this lower bound becomes more severe: for example, a DNA-mediated interaction potential is characterized by L~10 nm, necessitating a minimum framerate of ~100 s^{-1}.

The impact of framerate on the error was investigated using the exact, noiseless (inertial) trajectories considered in Figure 1(a). Shown in Figure 8 is the error, as defined in Figure 2, as a function of the time interval size; recall that all preceding analysis was performed with a time step, $\Delta t = 0.001\sqrt{m\sigma^2/\varepsilon}$, which corresponds to a (dimensionless) framerate of 1000. For time interval sizes less than ~0.05-0.1, the error is insensitive to the time step size and corresponds to the minimum error attainable with the discretization level and basis function choices. Above this time step size, the error increases rapidly to O(1), where the force profile details are essentially lost.

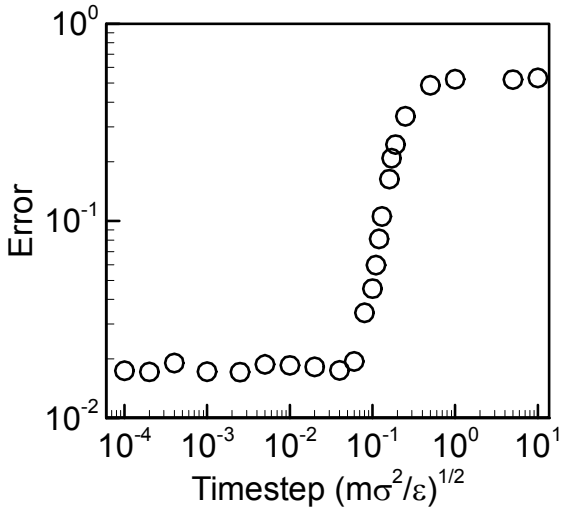


Figure 8: Error as a function of time step size for a system of 64 Lennard-Jones particles evolving via exact, noiseless, inertial dynamics. Extracted force profiles, are generated using 60 0.075σ-wide line segment basis functions. Error is calculated as $\left\|\mathbf{F}^* - \mathbf{F}\right\|_2 / \sqrt{M}$, where $\mathbf{F}^*$ contains the force calculated from Eq. (13) at the midpoint of each basis function, $\mathbf{F}$ is the actual force at each of these points, and *M* is the number of comparison points (bins).

5. HYDRODYNAMIC CORRELATIONS

Colloidal particle trajectories are unavoidably measured in some liquid medium, typically water. As each particle moves through the fluid, it perturbs it and creates a flow field that in turn impacts surrounding particles' trajectories[50-52]. Collectively, these hydrodynamic couplings lead to an effectively many-body interaction between the particles (or between particles and a wall), and must be accounted for if the intrinsic inter-particle interactions are to be isolated. Our aim here is not to extract a many-body interaction potential that mimics the hydrodynamic forces on each particle, although even this may be, in principle, possible. Instead, we seek to demonstrate that the hydrodynamic coupling can be effectively removed from the trajectory observations, allowing for the intrinsic pair-interaction (i.e., the LJ potential) to be recovered from the particle trajectories. Hydrodynamic correlations in incompressible fluids may be

included in Eq. (12) by replacing the single particle diffusivity, D, with the full mobility tensor, $\mathbf{D}$. The individual components of the mobility tensor used here are summarized in the Supplementary Materials.

Consider an overdamped system subject to hydrodynamic correlations and with no Brownian fluctuations such that the dynamics are given by $\mathbf{f} = k_B T \mathbf{D}^{-1} \dot{\mathbf{r}}$. Substitution into Eq. (11) then results in the following system of equations which may be solved to extract a force profile from a hydrodynamically correlated system.

$$\mathbf{G} = \left(\mathbf{C}^{*T} \mathbf{C}^{*} \right)^{-1} \mathbf{C}^{*T} \left(k_B T \mathbf{D}^{-1} \dot{\mathbf{r}} \right). \tag{23}$$

Note that the mobility tensor must be recalculated for every particle configuration, as it is dependent on particle coordinates. An example of a profile extracted using Eq. (19) is shown in Figure 9 for the case where the viscosity of the ambient fluid, η, is $0.07\sigma^2/\sqrt{\varepsilon m}$ and the particle radius is 0.45σ. These parameters were chosen such that γ is unchanged relative to the case shown in Figure 1(b), while introducing a sufficiently large particle radius to ensure significant hydrodynamic correlations between the particles. All other parameters were set to be the same as those used in Figure 1(b). As shown in Figure 9(a), ignoring the hydrodynamic correlations, i.e., using Eq. (15) instead of Eq. (19) when extracting the force profile, leads to large errors. Although not shown, these errors are configuration-dependent and in general cannot be removed once the force profile has been extracted from the particle trajectories. In Figure 9(b), the correct force and potential function are obtained by accounting for hydrodynamic correlations using Eq. (19). Statistical and systematic errors are similar to the cases in which hydrodynamic correlations are not present.

It should be noted that, as formulated, our method requires the mobility tensor, $\mathbf{D}$, as input in order to separate out the influence of hydrodynamics from the intrinsic inter-particle interactions. While this is straightforward for simple, unbounded geometries, it becomes more challenging for situations in which the medium is constrained by interfaces or walls. In such cases, complementary approaches such

as the methods in refs. [37, 40] may be employed to determine the relevant mobility tensor. In these studies, a Smoluchowski equation is fit to equilibrium distributions of particle-wall or particle-particle separations in order to simultaneously extract the interaction potentials and hydrodynamic mobilities.

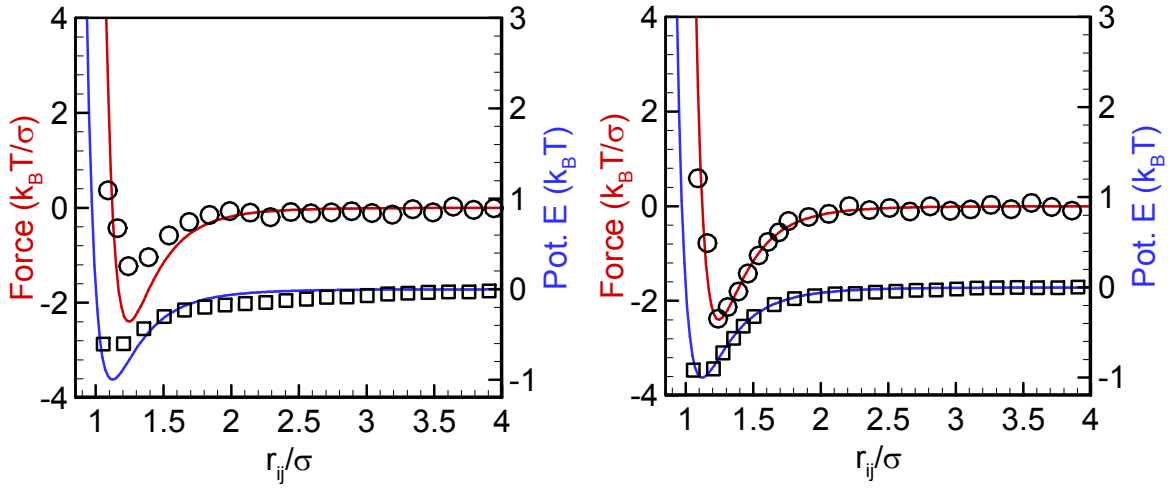


Figure 9: Pair potential functions and force profiles extracted from a system of 64 Lennard-Jones particles evolving via overdamped dynamics with hydrodynamic correlations. The profiles shown on the left were extracted assuming simple overdamped dynamics with no hydrodynamic correlations ($D_{i,j\neq i} = 0$). The profiles shown on the right were calculated while including hydrodynamic corrections. Extracted profiles are shown by symbols (force – circles, potential energy – squares), input profiles are denoted by solid lines. A total of 500 trajectory snapshots were used to extract the profiles. Both extracted profiles are generated using a set of 60 square wave basis functions of width 0.075σ.

VII. CONCLUSIONS

In this paper we have presented a computationally efficient and robust method for reliably extracting pair potential functions from arbitrary sets of multiple particle trajectories with no special constraints on particle configuration or system equilibration. We expect that this flexibility will greatly increase the scope of the systems that are amenable to interparticle interaction analysis, especially in

situations where equilibration is difficult to confirm or achieve. Our approach relies on some knowledge of the equations of motion that govern the particle trajectories but otherwise places no assumptions on the nature of the (pairwise) inter-particle interactions. Importantly, the same mathematical approach can be applied to systems having different fundamental equations of motion, i.e. underdamped, overdamped, with hydrodynamic interactions, or without. It is shown to be robust with respect to particle tracking error and the presence of Brownian motion. The latter can be mitigated by simply averaging over additional uncorrelated trajectory data, while the former, which is shown to introduce systematic errors, can be eliminated using a simple procedure without any *a priori* knowledge of the tracking error.

Further developments will be required to address additional relevant situations including particle anisotropy[53], many-body interactions[32], particle polydispersity[54], and hydrodynamic effects in more general situations[37]. One notable limitation with regards to hydrodynamic correlations is that our approach requires that the motion be a function only of the particle positions and its derivative—this will not be the case for particles in viscoelastic fluids, for instance. In the context of the first two issues, we note that the method does not formally require a pairwise target potential, and can be readily generalized for use on more complex, anisotropic or multi-body potential functions, with no fundamental methodological changes necessary. That said, increasing the dimensionality of the variable space upon which the potential function is dependent (e.g. orientation for anisotropic potentials) will place increased demands on the tracking data.

Acknowledgements: This work was supported by NSF awards CBET-113386 and CBET-1403237.

REFERENCES:

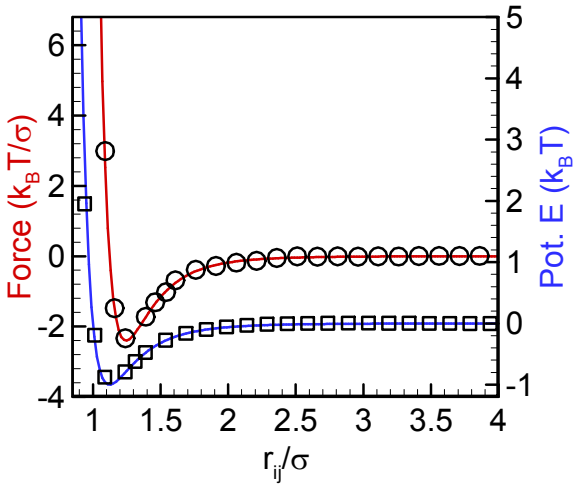
1 N. G. Almarza and E. Lomba, Phys Rev E **68**, 011202 (2003).
2 J. Tersoff, Phys Rev Lett **56**, 632 (1986).
3 A. J. Cohen, P. Mori-Sánchez, and W. Yang, Chemical Reviews **112**, 289 (2011).
4 O. Schueler-Furman, C. Wang, P. Bradley, K. Misura, and D. Baker, Science **310**, 638 (2005).
5 A. Skinner and D. Pettifor, Journal of Physics: Condensed Matter **3**, 2029 (1991).
6 J. Tersoff, Phys Rev Lett **56**, 632 (1986).
7 K.-P. Schröder and J. Sauer, The Journal of Physical Chemistry **100**, 11043 (1996).
8 M. Finnis and J. Sinclair, Philosophical Magazine A **50**, 45 (1984).
9 H. M. Hulburt and J. O. Hirschfelder, The Journal of Chemical Physics **9**, 61 (1941).
10 G. Casanova, R. Dulla, D. Jonah, J. Rowlinson, and G. Saville, Mol Phys **18**, 589 (1970).
11 J. Dymond and B. Alder, The Journal of Chemical Physics **51**, 309 (1969).
12 H. Böhm, I. McDonald, and P. Madden, Mol Phys **49**, 347 (1983).
13 D. Gough, E. Smith, and G. Maitland, Mol Phys **27**, 867 (1974).
14 M. Dharma-Wardana and G. Aers, Phys Rev B **28**, 1701 (1983).
15 M. Brennan, P. Hutchinson, M. Sangster, and P. Schofield, Journal of Physics C: Solid State Physics **7**, L411 (1974).
16 M. Bellissent-Funel, P. Chieux, D. Levesque, and J. Weis, Phys Rev A **39**, 6310 (1989).
17 N. G. Almarza, E. Lomba, and D. Molina, Phys Rev E **70**, 021203 (2004).
18 K. Watanabe and M. L. Klein, Chemical Physics **131**, 157 (1989).
19 K. Vondermassen, J. Bongers, A. Mueller, and H. Versmold, Langmuir **10**, 1351 (1994).
20 L. Reatto, D. Levesque, and J. J. Weis, Phys Rev A **33**, 3451 (1986).
21 F. Barocchi, M. Zoppi, and P. A. Egelstaff, Phys Rev A **31**, 2732 (1985).
22 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, The Journal of chemical physics **79**, 926 (1983).
23 J. N. Israelachvili and G. E. Adams, J Chem Soc Farad T 1 **74**, 975 (1978).
24 W. A. Ducker, T. J. Senden, and R. M. Pashley, Nature **353**, 239 (1991).
25 D. C. Prieve and N. A. Frej, Langmuir **6**, 396 (1990).
26 J. Y. Walz, Current opinion in colloid & interface science **2**, 600 (1997).
27 M. A. Bevan and D. C. Prieve, Langmuir **15**, 7925 (1999).
28 H.-J. Wu, T. O. Pangburn, R. E. Beckham, and M. A. Bevan, Langmuir **21**, 9879 (2005).
29 G. E. Fernandes, D. J. Beltran-Villegas, and M. A. Bevan, Langmuir **24**, 10776 (2008).
30 E. M. Furst and A. P. Gast, Phys Rev E **62**, 6916 (2000).
31 J. C. Crocker, J. A. Matteo, A. D. Dinsmore, and A. G. Yodh, Phys Rev Lett **82**, 4352 (1999).
32 M. Brunner, J. Dobnikar, H.-H. von Grünberg, and C. Bechinger, Phys Rev Lett **92**, 078301 (2004).
33 M. Polin, Y. Roichman, and D. G. Grier, Phys Rev E **77**, 051401 (2008).
34 Y. P. Kong and R. Parthasarathy, Langmuir **26**, 10541 (2010).
35 W. B. Rogers and J. C. Crocker, Rev Sci Instrum **85**, 043704 (2014).
36 J. C. Crocker and D. G. Grier, Phys Rev Lett **73**, 352 (1994).
37 S. K. Sainis, V. Germain, and E. R. Dufresne, Phys Rev Lett **99**, 018303 (2007).
38 G. M. Kepler and S. Fraden, Phys Rev Lett **73**, 356 (1994).
39 Y. L. Han and D. G. Grier, Phys Rev Lett **91**, 038302 (2003).
40 C. R. Iacovella, R. E. Rogers, S. C. Glotzer, and M. J. Solomon, The Journal of chemical physics **133**, 164903 (2010).
41 F. Ercolessi and J. B. Adams, EPL (Europhysics Letters) **26**, 583 (1994).
42 S. Izvekov and G. A. Voth, The Journal of chemical physics **123**, 134105 (2005).

[43] S. Izvekov, M. Parrinello, C. J. Burnham, and G. A. Voth, The Journal of chemical physics **120**, 10896 (2004).
[44] S. Izvekov, A. Violi, and G. A. Voth, The Journal of Physical Chemistry B **109**, 17019 (2005).
[45] A. Das and H. C. Andersen, The Journal of chemical physics **131**, 034102 (2009).
[46] Y. Umeno, T. Kitamura, K. Date, M. Hayashi, and T. Iwasaki, Computational materials science **25**, 447 (2002).
[47] Y. Li, D. J. Siegel, J. B. Adams, and X.-Y. Liu, Phys Rev B **67**, 125101 (2003).
[48] D. Wei, Y. Song, and F. Wang, The Journal of chemical physics **134**, 184704 (2011).
[49] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press Inc., New York, 1987).
[50] R. Schmitz and B. U. Felderhof, Physica A **116**, 163 (1982).
[51] P. Reuland, B. U. Felderhof, and R. B. Jones, Physica A **93**, 465 (1978).
[52] P. Hoogerbrugge and J. Koelman, EPL (Europhysics Letters) **19**, 155 (1992).
[53] L. Botto, E. P. Lewandowski, M. Cavallaro, and K. J. Stebe, Soft Matter **8**, 9957 (2012).
[54] T. O. Pangburn and M. A. Bevan, The Journal of chemical physics **124**, 054712 (2006).

Interaction Potentials from Arbitrary Multi-Particle Trajectory Data

Ian Jenkins, John C. Crocker and Talid Sinno

Department of Chemical and Biomolecular Engineering,
University of Pennsylvania,
220 S. 33rd St. Philadelphia, Pennsylvania 19104, USA



Robust and computationally efficient extraction of interparticle forces and potentials from non-equilibrium, multiparticle trajectories.