THE UNIVERSITY OF CHICAGO

DATA-DRIVEN STATISTICAL MECHANICAL AND SYMMETRY INSIGHTS INTO COLLECTIVE COORDINATES IN SMALL OPTICAL MATTER CLUSTERS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF CHEMISTRY

BY
SHIQI CHEN

CHICAGO, ILLINOIS
AUGUST 2024

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor, Prof. Norbert Scherer, for his unwavering guidance, expertise, and patience. His insightful feedback and constant encouragement have been pivotal in shaping this research and guiding my path throughout graduate school. I am incredibly thankful to Norbert for all the opportunities, memories, and experiences.

I would also like to thank my co-advisor, Prof. Andrew Ferguson for mentoring me over the years and introducing me to the fascinating field of machine learning. Andy has provided me with many great insights that have helped me a lot during my research. Andy's passion and enthusiasm for my research has kept me motivated and inspired me to come up with new ideas.

I am grateful to Prof. Stuart Rice for sharing many important ideas and useful suggestions with me, which is crucial to the formation of the framework of my research. I want to thank Prof. Suriyanarayanan Vaikuntanathan for serving on my thesis committee.

All current and former members of the Scherer Lab have had a large impact on my research. Special thanks to John Parker and Curtis Peterson with whom I worked closely when I first joined the lab. I'm also deeply grateful to Emmanuel Valenton and Spoorthi Nagasamudram who shared their thoughts on my research. Yanzeng Li is a post-doc in the group and has been incredibly helpful, patient, and foundational to my learning of electrodynamics theory.

I am appreciative of the opportunity to interact with the great group of folks that were my colleagues in the Ferguson Lab. I would like to thank Mike Jones, Kirill Shmilovich, and Oliver Tang, for all of the discussions and collaborations we have had over the years. Many of you worked tirelessly in the lab testing and coding new machine learning algorithms for data analysis, and I have benefited from your hard labor.

The Research Computing Center (RCC) at the University of Chicago has been a very useful resource to me. The RCC's computing resources have provided me with over a million

compute hours that have made a large number of results and figures in this dissertation possible. Additionally, the RCC has held many seminars I have attended over the years that have taught me new computational and programming skills. I also want to thank the University of Chicago Department of Chemistry for the Otto H. & Valerie Windt Memorial Fellowship.

Finally, I'd like to thank my family and close friends for their invaluable support and encouragement. My parents' unwavering support, both in graduate school and in life, helped me to pursue my aspirations of a career in science and provided me with the foundation in life to do so. Their sacrifices and hard work have been a constant source of inspiration for me. My closest friend throughout graduate school Yu Jin helped make the hard times easier and the easier times more fun.

ABSTRACT

Optical matter (OM) systems are a class of active non-equilibrium materials. One of the most interesting variants consists of nano-particles (NPs) that form 2-dimensional ordered structures when illuminated and trapped by a focused laser beam. The force field developed by the electrodynamic interactions that hold the NPs together is non-conservative. Depending on the number of NPs and the phase, amplitude and polarization properties of the incident electromagnetic field, there are several different metastable ordered structures that can be formed. The relative stabilities of these structures can be tuned by adjusting the aforementioned laser beam properties. Therefore, the beam power, beam shape, spatial phase profile, and polarization of the light create a rich parameter subspace to explore stabilization, control and design of particular OM structures and their dynamics. Each of the one or more different ordered OM structures that form in a focused laser beam constitutes a (metastable) non-equilibrium steady state (NESS), which can, for example, be used to build optical matter machines that do mechanical work under a laser beam.

In order to study the mechanically dynamic and light scattering properties of OM systems, I have developed and employed a data-driven approach based on principal component analysis (PCA) and harmonic linear discriminant analysis (HLDA) to determine the collective modes of non-conservative and overdamped OM structures. The approach is demonstrated via electrodynamics-Langevin dynamics simulations six electrodynamically-bound nanoparticles coupled to an incident laser beam. I then use this data-driven approach to build the PCA-HLDA reaction coordinates between stable states connected by Markov state model (MSM), compute entropy production rate, and analyze light scattering properties as well as induced-polarization. These studies represent a systematic endeavor to understanding and eventually controlling optical matter systems. This approach is also promising to the study of other non-conservative and overdamped active matter systems.

CHAPTER 1

INTRODUCTION

Active matter is a type of systems that are intrinsically out of thermal equilibrium and consumes energy in order to maintain the motions in the system. For example, the kinesin that uses energy from adenosine triphosphate hydrolysis to move cargoes inside a cell is a type of active matter [4]. There are also synthetic molecular motors that are active matter driven by inter-molecular interactions [5]. In this thesis, I will focus on optical matter (OM) as a type of active matter system.

Optical matter (OM) is a type of material or molecule-like structure in which the constituents (e.g. nanoparticles (NPs) or micron-scale particles) are bound together by electrodynamic interactions [6, 7]. A fundamental aspect of OM structures is that they tend to occupy interparticle distances that are integer multiples of the wavelength of the incident optical field. This is particularly clear when the constituents are NPs that are smaller than the wavelength of light [7, 8]. The electrodynamic interaction that causes this behavior, known as optical binding [6, 9, 10], allows formation of regular OM configurations (e.g. 2D arrays with trigonal symmetry or anisotropic arrays with rectangular lattice configurations) with minimal optical information. Generally only the overall shape, polarization, phase, and power of the optical beam are controlled [9], and the resulting particle configurations are readily visualized by (dark-field) optical microscopy [7, 11–13]. These systems are of particular interest in optical and statistical physics due to their manifestation of non-reciprocal forces, collective (correlated) interactions and many-body effects in their electrodynamic interactions [14–18]. Furthermore, OM systems also serve as useful mesoscale analogs with which to study atomic-level chemical processes [19].

It is known that long range interactions manifest structural and dynamic correlations in driven dense colloidal solutions due to hydrodynamic interactions [20], and in quantum-dot perturbed molten salt solutions [21]. OM configurations self-organize in optical traps in

solution and achieve new collective properties due to their electrodynamic interactions and long range (periodic) potentials [6,7,9,10]. The long range interactions create a new richness for understanding isomerizations of OM configurations as they affect the energetics globally and imbue OM systems with many-body (N-body) interactions and physics. Therefore, rational control of the configuration of optical matter requires solving a many-body problem by effective and practical methods.

There are similarities between molecular and optical matter systems. The probability distribution of a molecular configuration in the neighborhood of the potential minimum can be well approximated by a multivariate Gaussian distribution with a covariance matrix governed by the Hessian matrix H at the minimum of the potential defined as:

$$H_{ij} = \frac{\partial U(\mathbf{r})}{\partial r_i \partial r_j} \tag{1.1}$$

where $\boldsymbol{r}$ corresponds to the Cartesian coordinates of the molecular configuration and U is the potential energy function. When the potential is harmonic as in the case of molecules, the normal modes of the system can be determined from the covariance matrix [22]. Consider the asymmetric stretching mode of a water molecule as a simple example. The coordinate displacement exemplified in this mode is involved in the dissociation and autoionization reactions of water [23]. Consider a second example of boat-to-chair isomerization of cyclohexane, in which particular normal modes dominate others in describing the isomerization transition [24]. Optical matter systems in optical traps in solution can undergo structural transitions, such as the transition from "triangle" to a "chevron" configuration of a 6-particle OM system, shown in Fig. 1.1. However, compared to molecular systems, the external electrodynamic forces acting on the OM system in solution are non-conservative and over-damped [25].

Normal modes are orthogonal collective motions of particles that carry independent contributions to the system energy. The conventional definition of normal modes is valid only

Figure 1.1: Experimental dark-field microscopy images of the structural transition from triangle to chevron of the 6-particle OM systems.

for harmonic particle-particle interactions. There is no formal mechanical definition of normal modes for overdamped and non-conservative systems. In other words, if the following Langevin equation is considered:

$$m\frac{d^2\mathbf{r}}{dt^2} = \mathbf{F}_{ext}(\mathbf{r},t) - \xi\frac{d\mathbf{r}}{dt} + \eta \tag{1.2}$$

where $\mathbf{r}$ is the position, m is the mass, $\mathbf{F}_{ext}$ is the external force field, ξ is the friction coefficient, and η is the random force, only when $\xi = 0$ and $\frac{\partial F_{ext,x}}{\partial y} = \frac{\partial F_{ext,y}}{\partial x}$ can normal modes be well-defined. It should be emphasized that the non-conservative nature of the OM system refers to the non-conservative (external) electrodynamic force field $\mathbf{F}_{ext}$ [25].

Renson and Kerschen have defined nonlinear normal modes in underdamped systems (ξ is not zero but not large enough to neglect the acceleration term) [26]. David and Jacobs have used principal component analysis (PCA) to study the large-scale fluctuations in molecular and colloidal systems [27]. Zaccone and co-workers have used instantaneous normal modes (INM) and the vibrational density of states in liquids to analyze none-affine dynamics of amorphous materials such as glassy polymers [28–31]. These molecular, colloidal, and liquid systems can be overdamped (i.e., ξ is large enough to neglect the acceleration term) but conservative (i.e., $\frac{\partial F_{ext,x}}{\partial y} = \frac{\partial F_{ext,y}}{\partial x}$). Chattoraj, et al. have analyzed the eigenvalues and eigenvectors of the J-matrix, the first derivative matrix of the external force field, and found oscillatory solutions of motion that are particularly useful for studies of non-conservative (i.e.,

$\frac{\partial F_{ext,x}}{\partial y} \neq \frac{\partial F_{ext,y}}{\partial x}$) but underdamped (i.e., ξ is not large enough to neglect the acceleration term) systems [32]. However, there exists no oscillatory solution in overdamped cases and the non-orthogonal eigenvectors of the J-matrix lead to intrinsically coupled collective modes that are complicated to analyze.

In the approach presented in Chapter 3 in this thesis, I define collective modes in overdamped and non-conservative systems (ξ is large enough to neglect the acceleration term; $\frac{\partial F_{ext,x}}{\partial y} \neq \frac{\partial F_{ext,y}}{\partial x}$) by carrying out PCA on the configurational trajectories based on the deviations of the OM constituent particles from a reference configuration. PCA diagonalizes the covariance matrix of the OM particle coordinates to define a linear transformation into a basis of non-local collective modes (principal components, PCs) ordered by the magnitude of configurational variance they contain [33]. The leading PCs correspond to collective degrees of freedom with large variance that typically characterize large-scale global rearrangements of the system, whereas the trailing PCs correspond to small-variance fluctuations around particular metastable configurations.

Chapter 2 introduces four important methods that are extensively used in this dissertation. First, the electrodynamic-Langevin dynamic (EDLD) method [34, 35] is introduced, which is used to simulate the OM system and generate time trajectories for further data analysis. Second, the "lattice fitting" algorithm is introduced, which is used to align the orientation of the structures in a simulation (or experimental) trajectory. Third, the principal component analysis (PCA) method [1] applied to simulation trajectories is introduced. PCA is used to generate collective fluctuation coordinates for stable OM structures. Last but not least, the diffusion k-means clustering method [2] is introduced, which is used to carry out microstate clustering on the OM simulation trajectory data. This method is a key aspect of creating Markov state models for the OM systems.

Chapter 3 shows that the leading PCs obtained from PCA carried out on the OM simulation trajectories serve as good descriptors for transitions between metastable OM configura-

tions. The PCs may further be formally converted into reaction coordinates using harmonic linear discriminant analysis (HLDA) [22]. This PCA-HLDA approach is demonstrated in an application to trajectories of the triangle-to-chevron transition, like the measured result of using combined EDLD simulations [15, 34] of a 6-particle OM system.

In Chapter 4, the dynamics of OM structural reconfiguration is studied by building Markov state models (MSMs) for the stable states of the OM system for different beam powers. In order to confront the permutational invariant nature of the OM nanoparticles, permutation-invariant nonlinear dimensionality reduction and spectral clustering are employed to perform data-driven identification of the metastable states corresponding to long-lived non-equilibrium OM configurations and the transition rates between them. By constructing MSMs for various powers of the incident laser beam, I construct empirical models for the relative stability of the metastable configurations and use these models to discover new beam conditions designed to preferentially stabilize particular OM configurations of interest. This methodology presents a transferable scheme that can be used to understand, design, and control the dynamics of permutation-invariant systems with conservative or non-conservative force fields prevalent in optical and active matter systems.

Chapter 5 develops approaches to compute the entropy production rate in numerical simulations of non-equilibrium steady states (NESSs) with internal dynamics, and to understand how those internal motions and the forces that generate the NESS affect the entropy production rate. The six particle triangle OM structure is used as the prototype NESS system investigated in detail where the calculated entropy production rates are compared to power dissipation. After decomposition of the external force field into conservative and non-conservative parts, I numerically compute the mean entropy production rate of the six particle triangle configuration using a linear approximation to and Helmholtz decomposition of the force field analyzed in terms of the 12 collective coordinates defined by PCA.

In Chapter 6, numerical simulations are conducted to study small OM structures with var-

ious symmetries to demonstrate the analogy between the Raman active modes of molecules and the OM fluctuation modes along which the derivative of the induced-polarization is non-zero. In order to analyze the mode symmetry, I adopt a broken symmetry perturbation generated by the chirality of the light polarization. I find that the OM fluctuation modes along which the derivative of their induced-polarization is non-zero correspond to the same irreducible representation of the structural symmetry point group as the Raman active modes that preserve the polarization of light in helicity-resolved Raman scattering.

Chapter 7 presents two types of OM system for which the collective mode analysis approach is not directly applicable. The first type is a small OM cluster that consists of more than one stable structure that undergoes frequent conversions between each other. The second type is a large OM cluster in which defects relative to the trigonal lattice structure frequently appear. The 8-particle OM structure that manifest pseudorotation is introduced as an example of the first type of system. The 19-particle 6-fold symmetric OM structure is introduced as an example of the second type of system. Several important and interesting properties, such as the structural and temperature dependence of the decay time scale of the system dynamics, are introduced for these two systems.

CHAPTER 2

METHODS

The codes for the methods introduced in this thesis can be found on the author's Github page [36].

2.1 Electrodynamics-Langevin Dynamics (EDLD) Simulation

The dynamical evolution of the optical matter (OM) system can be modeled by combining a finite-difference time-domain (FDTD) solution of the electrodynamic forces with a Langevin dynamics equation of motion for the particle positions [34]. Silver NPs with 150 nm diameter are used as the material constituents of the OM systems. In this study, the nanoparticles are illuminated with a defocused, converging circularly polarized Gaussian beam with tunable width and beam power. The defocus of the beam is equal to the Rayleigh range, $z = 0.5kw^2$, where w is the beam width, $k = 2\pi n_b/\lambda$, and n_b is the index of refraction. These field/beam conditions allowed formation of stable OM arrays even in the presence of thermal noise/forces.

However, the FDTD method is insufficiently efficient to access experimental timescales [35]. Therefore, an electrodynamics-Langevin dynamics (EDLD) approach was developed based on generalized multiparticle Mie theory (GMMT) [37,38], which characterizes the OM system model described above both efficiently and accurately enough for direct comparison with experiment [35]. The resulting EDLD solver performs a numerical Verlet integration of the following Langevin equation,

$$m\frac{d^2\boldsymbol{x}}{dt^2} = \boldsymbol{F}_{ED}(\boldsymbol{x}) - \boldsymbol{\nabla}U_{DL}(\boldsymbol{x}) - \zeta\frac{d\boldsymbol{x}}{dt} + \boldsymbol{f}^{(b)} \tag{2.1}$$

In Eq. 2.1, m is the mass of the particle; $\boldsymbol{x}$ contains the Cartesian coordinates of the particles; t is time; $\zeta = 6\pi\eta r$ is the friction coefficient specified by Stokes' Law where r is the

nanoparticle radius and η is the viscosity of the surrounding fluid; $\boldsymbol{f}^{(b)}$ is the bath random force term defined to be white noise satisfying the fluctuation dissipation relation at a specified temperature; $\boldsymbol{F}_{ED}$ is the sum of external electrodynamic force fields computed by the Maxwell stress tensor [37]. This EDLD method was encoded into a Python/C++ package termed MiePy [39] by Dr. John Parker, a former student in the Scherer group. The force field $\boldsymbol{F}$ computed is non-conservative due to the constant flux of power from the optical beam. The steady states reached by the OM system correspond to states in which the power input from the optical beam is balanced by frictional dissipation into the (assumed infinite) medium. In this dissertation, I assume that the solvent medium is water (refraction index $n_b = 1.33$, viscosity $\eta = 8 \times 10^{-4}$ Pa·s) and the wavelength $\lambda = 800$ nm.

An additional important element of the OM systems is the double layer electrostatic interaction U_{DL} [40]:

$$\begin{aligned} U_{DL}(\boldsymbol{x}) &= \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} V(r_{ij}) \\ V(r_{ij}) &= 32\pi\varepsilon_0\varepsilon_m \left(\frac{k_B T}{z_p e}\right)^2 \tanh^2\left(\frac{z_p e \psi}{4 k_B T}\right) \times \exp\left[-\kappa\left(r_{ij} - 2r\right)\right] \end{aligned} \tag{2.2}$$

where N is the number of particles; ε_m is the medium electric permittivity; $z_p = 1$ is the ion valence; $\psi = -77$ mV is the measured surface electrostatic potential for 150 nm dia. Ag nanoparticle [17]; $\kappa^{-1} = 100$ nm is the Debye screening length; r is the particle radius; $r_{ij} = \left|\boldsymbol{x}_i - \boldsymbol{x}_j\right|$ is the distance between the i^{th} and j^{th} particle. The surface electrostatic potential is chosen according to previous theoretical work [41]. The Debye screening length is determined by simulation experimental comparison [2]. Calculations performed on the University of Chicago Midway 2 computer using 28 $\times$ 2.4 GHz Intel E5-2680 v4 CPUs execute 2.5 seconds of simulation time per hour of wall clock time with a 5 μs simulation time step. The simulation takes 10 times if the simulation step is 0.5 μs. The electrodynamic force is updated for every Langevin dynamics step.

2.2 Lattice Fitting

Given a certain optical matter configuration with N particles, we want to find the best set of N sites on a 2-dimensional trigonal lattice pattern that is closest to the given configuration. Let the positions of the particles in the given configuration be $\mathbf{r}_1$, $\mathbf{r}_2$, ..., $\mathbf{r}_N \in \mathbb{R}^2$. Let the positions of the hexagonal lattice sites be $\mathbf{s}_1$, $\mathbf{s}_2$, ..., $\mathbf{s}_N \in \mathbb{R}^2$. Eq. 2.3 describes the approach for minimization of the fitting error, where Π is the set that contains all injections from 1, 2, ... N to $\mathbb{Z}^+$.

$$Err^* = \left[\min_{a\in\mathbb{R}^+, \mathbf{r}_0\in\mathbb{R}^2, \hat{\mathbf{R}}\in SO(2), \pi\in\Pi} \sum_{j=1}^{N} \left| \hat{\mathbf{R}}(\mathbf{r}_j + \mathbf{r}_0) - a\mathbf{s}_{\pi(j)} \right|^2 \right]^{\frac{1}{2}} \tag{2.3}$$

In Eq. 2.3, a is the lattice constant, r_0 is the translation vector, $\hat{\mathbf{R}} \in SO(2)$ is the 2x2 rotation matrix, and π is the assignment of particles to lattice sites. Considering the symmetry of the lattice sites, the parameter domain can be further limited by forcing $\pi(1) = 1$ and $\mathbf{r}_0$ within the primitive cell. After these steps, the parameter space is discretized and optimized to get the best assignment between the particles and the lattice sites, π^*.

With the assignment in hand an analytical solution of the optimal translation, rotation, and lattice constant can be found. Define vectors $\boldsymbol{p}$ and $\boldsymbol{q}$ such that $p_j = r_{j,1} + ir_{j,2}$ and $q_j = s_{\pi(j),1} + is_{\pi(j),2}$ ($j \in \{1, 2, ..., N\}$), so $\mathbf{p}, \mathbf{q} \in \mathbb{C}^N$. The fitting error expression can be transformed to Eq. 2.4, where p_0 is the translation, θ is the rotation angle, $\mathbf{1}$ is the vector of all ones, and a is the lattice constant. The superscript * indicates the optimal value.

$$Err^* = \left[\min_{a\in\mathbb{R}^+, p_0\in\mathbb{C}, \theta\in[0,2\pi)} \left| e^{i\theta}(\mathbf{p} + p_0\mathbf{1}) - a\mathbf{q} \right|^2 \right]^{\frac{1}{2}} \tag{2.4}$$

If translation is exempted from optimization, then $p_0^* = 0$. Otherwise:

$$p_0^* = \frac{\mathbf{1}^{\mathrm{T}}}{N}(\mathbf{q} - \mathbf{p}) \tag{2.5}$$

Let $\mathbf{p'} = \mathbf{p} + p_0^*\mathbf{1}$, then we have:

$$e^{i\theta^*} = \frac{(\mathbf{p'})^{\mathrm{H}}\mathbf{q}}{\left|\mathbf{q}^{\mathrm{H}}\mathbf{p}\right|} \tag{2.6}$$

If the lattice constant is exempted from optimization, then $a^* = 1$. Otherwise:

$$a^* = \frac{\left|\mathbf{q}^{\mathrm{H}}\mathbf{p}'\right|}{|\mathbf{q}|^2} \tag{2.7}$$

The detailed derivation of the results in Eqs. 2.5, 2.6, and 2.7 is the following:

First, we optimize the translation of the OM array with respect to the lattice by setting $\theta = 0$ and $a = 1$.

$$\begin{aligned}
&Err^2\Big|_{\theta=0,a=1} = |(\mathbf{p} + p_0\mathbf{1}) - \mathbf{q}|^2 = [(\mathbf{p} - \mathbf{q}) + p_0\mathbf{1}]^{\mathrm{H}}[(\mathbf{p} - \mathbf{q}) + p_0\mathbf{1}] \\
&= (\mathbf{p} - \mathbf{q})^{\mathrm{H}}(\mathbf{p} - \mathbf{q}) + (p_0\mathbf{1})^{\mathrm{H}}(p_0\mathbf{1}) + (p_0\mathbf{1})^{\mathrm{H}}(\mathbf{p} - \mathbf{q}) + (\mathbf{p} - \mathbf{q})^{\mathrm{H}}(p_0\mathbf{1}) \\
&= (\mathbf{p} - \mathbf{q})^{\mathrm{H}}(\mathbf{p} - \mathbf{q}) + Np_0\overline{p_0} + \overline{p_0}[\mathbf{1}^{\mathrm{H}}(\mathbf{p} - \mathbf{q})] + p_0\overline{[\mathbf{1}^{\mathrm{H}}(\mathbf{p} - \mathbf{q})]} \\
&= |\mathbf{p} - \mathbf{q}|^2 - \frac{1}{N}\left|(\mathbf{p} - \mathbf{q})^{\mathrm{H}}\mathbf{1}\right|^2 \\
&+ N\left\{p_0\overline{p_0} + \overline{p_0}\left[\frac{\mathbf{1}^{\mathrm{H}}}{N}(\mathbf{p} - \mathbf{q})\right] + p_0\overline{\left[\frac{\mathbf{1}^{\mathrm{H}}}{N}(\mathbf{p} - \mathbf{q})\right]} + \left[\frac{\mathbf{1}^{\mathrm{H}}}{N}(\mathbf{p} - \mathbf{q})\right]\overline{\left[\frac{\mathbf{1}^{\mathrm{H}}}{N}(\mathbf{p} - \mathbf{q})\right]}\right\} \\
&= |\mathbf{p} - \mathbf{q}|^2 - \frac{1}{N}\left|(\mathbf{p} - \mathbf{q})^{\mathrm{H}}\mathbf{1}\right|^2 + N\left|p_0 + \frac{\mathbf{1}^{\mathrm{H}}}{N}(\mathbf{p} - \mathbf{q})\right|^2
\end{aligned} \tag{2.8}$$

In order to minimize Eq. 2.8, one sets the last term to zero, which means:

$$p_0^* = \frac{\mathbf{1}^{\mathrm{H}}}{N}(\mathbf{q} - \mathbf{p}) \tag{2.9}$$

Note the expression for p_0^* (Eq. 2.9) is actually the difference between the centers of mass of the two configurations, $\mathbf{p}$ and $\mathbf{q}$. In addition, after $\mathbf{p}$ and $\mathbf{q}$ are translated to let both centers of mass overlap with the origin, p_0^* will stay unchanged no matter how p and q are rotated or linearly scaled. Therefore, after the choice of p_0^*, further optimization of θ and a does not

affect p_0^*, so that their optimization can be carried out successively and independently.

Next, we optimize the rotation and lattice constant. Letting $p' = p + p_0^*\mathbf{1}$ gives:

$$Err^* = \left[\min_{a\in\mathbb{R}^+,\theta\in[0,2\pi)} \left|e^{i\theta}\mathbf{p'} - a\mathbf{q}\right|^2\right]^{\frac{1}{2}} \tag{2.10}$$

and

$$\begin{aligned} Err^2\Big|_{p_0=p_0^*} &= \left|e^{i\theta}\mathbf{p'} - a\mathbf{q}\right|^2 = \left|\mathbf{p'} - ae^{-i\theta}\mathbf{q}\right|^2 = \left|\mathbf{p'}\right|^2 + a^2\left|\mathbf{q}\right|^2 - 2a\mathrm{Re}(\mathrm{e}^{\mathrm{i}\theta}\mathbf{q}^{\mathrm{H}}\mathbf{p'}) \\ &= a^2\left|\mathbf{q}\right|^2 - 2(a\left|\mathbf{q}\right|)\left[\frac{\mathrm{Re}(\mathrm{e}^{\mathrm{i}\theta}\mathbf{q}^{\mathrm{H}}\mathbf{p'})}{|\mathbf{q}|}\right] + \left[\frac{\mathrm{Re}(\mathrm{e}^{\mathrm{i}\theta}\mathbf{q}^{\mathrm{H}}\mathbf{p'})}{|\mathbf{q}|}\right]^2 - \left[\frac{\mathrm{Re}(\mathrm{e}^{\mathrm{i}\theta}\mathbf{q}^{\mathrm{H}}\mathbf{p'})}{|\mathbf{q}|}\right]^2 + \left|\mathbf{p'}\right|^2 \\ &= |\mathbf{q}|^2\left[a - \frac{\mathrm{Re}(\mathrm{e}^{\mathrm{i}\theta}\mathbf{q}^{\mathrm{H}}\mathbf{p'})}{|\mathbf{q}|^2}\right]^2 + \left\{\left|\mathbf{p'}\right|^2 - \left[\frac{\mathrm{Re}(\mathrm{e}^{\mathrm{i}\theta}\mathbf{q}^{\mathrm{H}}\mathbf{p'})}{|\mathbf{q}|}\right]^2\right\} \end{aligned} \tag{2.11}$$

In order to minimize Err, the first term in Eq. 2.11 is set to zero, and the second term should be as small as possible. This means that the real part of $e^{i\theta}\mathbf{q}^{\mathrm{H}}\mathbf{p'}$ is maximized to be $\left|\mathbf{q}^{\mathrm{H}}\mathbf{p'}\right|$.

Therefore:

$$e^{i\theta^*} = e^{-i\mathrm{arg}(\mathbf{q}^{\mathrm{H}}\mathbf{p'})} = \frac{(\mathbf{p'})^{\mathrm{H}}\mathbf{q}}{\left|\mathbf{q}^{\mathrm{H}}\mathbf{p'}\right|} \tag{2.12}$$

$$a^* = \frac{\mathrm{Re}(\mathrm{e}^{\mathrm{i}\theta}\mathbf{q}^{\mathrm{H}}\mathbf{p'})}{|\mathbf{q}|^2} = \frac{\left|\mathbf{q}^{\mathrm{H}}\mathbf{p'}\right|}{|\mathbf{q}|^2} \tag{2.13}$$

This is the optimal rotation angle and the optimal lattice constant.

2.3 Principal Component Analysis of Simulation Trajectories

The application of principal component analysis (PCA) on simulation trajectories involves three steps. The first step is to relate the eigenvectors of the covariance matrix of the probability distribution to the normal modes of a conservative system. Then, for a non-

conservative system, the physical meaning of the principal components generated by PCA can be interpreted as the orthogonal collective modes that characterize the independent degrees of freedom determined by the conservative part of the force field. The second step is the derivation of the Jacobian determinant that is used as the weight in the PCA algorithm. The third step is the complete weighted PCA algorithm from the input trajectory to the output collective modes.

2.3.1 Derivation of the relationship between conventional normal modes and statistical collective modes in conservative systems

Let us consider a system with N identical particles in $\mathbb{R}^d$. A d-dimensional position vector, $\mathbf{r}_k = [r_{k1}, r_{k2}, ..., r_{kd}]$ is assigned to the k^{th} particle. The configuration of the whole system is described by the Nd-dimensional configuration vector:

$$\mathbf{R} = (\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N) = [r_{11}, r_{12}, ..., r_{1d}, r_{21}, r_{22}, ..., r_{2d}, .., r_{N1}, r_{N2}, ..., r_{Nd}] \in \mathbb{R}^{Nd} \quad (2.14)$$

Next, consider a force field acting on this $\mathbb{R}^{Nd}$ configuration space. The force acting on a certain particle not only depends on the position of the particle itself, but also on the positions of the other particles in the system. The function $\mathbf{F}$ takes in the position vector of the N particles and produces the forces acting on these N particles. Since the force acting on a particle is a vector in $\mathbb{R}^d$, $\mathbf{F}$ maps the configuration space into the N-particle force space that is also Nd-dimensional.

$$\mathbf{F} : \mathbb{R}^{Nd} \longrightarrow \mathbb{R}^{Nd} \quad (2.15)$$

The force derivative matrix $\mathbf{K}$ is defined by taking derivatives over all the particle position

coordinates of the force function.

$$K_{ip,jq} = \frac{\partial F_{ip}}{\partial r_{jq}} \tag{2.16}$$

Since the system is conservative, the potential energy difference between two different configurations is well-defined as the integral of the force field over any path that connects them. We choose a reference configuration $\mathbf{R}^{(0)}$ and define its potential energy to be zero. Then for any other configuration $\mathbf{R}$, its potential function is defined as:

$$V(\mathbf{R}) = \int_{\mathbf{R}}^{\mathbf{R}^{(0)}} \mathbf{F}(\hat{\mathbf{R}}) \cdot d\hat{\mathbf{R}} \tag{2.17}$$

From Eq. 2.17, it is obvious that we can find the force function $\mathbf{F}$ from the potential function by taking the negative gradient:

$$\mathbf{F} = -\nabla V \tag{2.18}$$

Therefore, the force derivative matrix $\mathbf{K}$ is the negative of the Hessian matrix of V.

$$K_{ip,jq} = \frac{\partial F_{ip}}{\partial r_{jq}} = -\frac{\partial^2 V}{\partial r_{ip} \partial r_{jq}} \tag{2.19}$$

When a configuration is at a local minimum of the potential, its energy is at a local minimum. The system dynamics near the minimum is determined, to second order by

$$\begin{aligned} V(\mathbf{R}) =& V(\mathbf{R}_0) + \nabla V(\mathbf{R}_0) \cdot (\mathbf{R} - \mathbf{R}_0) + \frac{1}{2}(\mathbf{R} - \mathbf{R}_0)^{\mathrm{T}} [\nabla \nabla^{\mathrm{T}} V(\mathbf{R}_0)](\mathbf{R} - \mathbf{R}_0) \\ &+ o(\|\mathbf{R} - \mathbf{R}_0\|_2^2) \end{aligned} \tag{2.20}$$

Since V is at its local minimum at $\mathbf{R}_0$, the gradient of V at $\mathbf{R}_0$ is zero. Therefore, the first order term vanishes and the second order term dominates. In addition, we can replace the Hessian matrix of V by the negative of force derivative matrix $\mathbf{K}$.

$$V(\mathbf{R}) = V(\mathbf{R}_0) - \frac{1}{2}(\mathbf{R} - \mathbf{R}_0)^{\mathrm{T}} \mathbf{K} (\mathbf{R} - \mathbf{R}_0) + o(\|\mathbf{R} - \mathbf{R}_0\|_2^2) \tag{2.21}$$

At a non-zero temperature T, the system configuration $\mathbf{R}$ is a random variable that has a probability distribution. The probability density for a configuration $\mathbf{R}$ to appear at a non-zero temperature T follows the Boltzmann distribution:

$$p(\mathbf{R}) \propto e^{-\frac{V(\mathbf{R})}{k_B T}} \tag{2.22}$$

where k_B is the Boltzmann constant. Then,

$$p(\mathbf{R}) \propto e^{-\frac{V(\mathbf{R})-V(\mathbf{R}_0)}{k_B T}} = e^{\frac{(\mathbf{R}-\mathbf{R}_0)^{\mathrm{T}}\mathbf{K}(\mathbf{R}-\mathbf{R}_0)}{2k_B T}+o(\|\mathbf{R}-\mathbf{R}_0\|_2^2)} \tag{2.23}$$

Therefore, the probability distribution of configurations near the minimum energy configuration $\mathbf{R}_0$ is close to a Gaussian distribution. Let the covariance matrix of this Gaussian distribution be Σ. Then, according to the distribution function, we have

$$\mathbf{K}(\mathbf{R}_0)\Sigma = -k_B T \mathbf{I}_{Nd\times Nd} \tag{2.24}$$

Therefore, any normal mode of the system, which is one of the eigenvectors of the force derivative matrix $\mathbf{K}$, is also an eigenvector of the covariance matrix Σ. This is because, for any $\mathbf{v}$ that satisfies:

$$\mathbf{K}\mathbf{v} = \lambda \mathbf{v}, \tag{2.25}$$

we have:

$$\Sigma v = k_B T \mathbf{K}^{-1}\mathbf{v} = \frac{k_B T}{\lambda}\mathbf{v} \tag{2.26}$$

Note that the derivation above cannot be applied to non-conservative systems directly.

Since the potential is not well-defined in our system, the Hessian matrix should be replaced by taking the first derivative of the external force field, the J matrix, instead of taking the second derivative of the potential. The J-matrix approach [32] is an important method that can be compared to the present work. The J-matrix method leads to discovery

of oscillatory solutions [32], which indeed sounds promising in analyzing non-conservative systems, but is limited to underdamped systems. In the overdamped case that is relevant for optical matter systems, the Langevin equation is first order so that there is no oscillatory solution. The J-matrix method is still applicable to the overdamped OM system, but due to the absence of oscillatory solutions, its application is limited to the analysis of its eigenvectors and eigenvalues. Furthermore, since the J-matrix is asymmetric, its eigenvectors are not orthogonal to each other. Its left eigenvectors are different from its right eigenvectors, which induces coupling between the collective modes represented by its eigenvectors. We note that the covariance matrix analyzed using PCA gives orthogonal eigenvectors that represent an orthogonal basis much easier to manipulate. Therefore, I do not use the J-matrix approach, but use PCA on covariance matrices to obtain the collective modes for OM systems.

2.3.2 Derivation of the Jacobian determinant for each frame in a trajectory

2.3.2.1 Definition of collective PCA coordinates

The OM systems we study in this thesis are N-particle systems in a 2-dimensional space, described by the Cartesian coordinates $\boldsymbol{x} = [x_1, y_1, \cdots, x_N, y_N]^T$. We have a reference structure of which the Cartesian coordinate is $\boldsymbol{x}_0 = [a_1, b_1, \cdots, a_N, b_N]^T$. For an OM system prepared with circularly polarized light, the reference structure can be taken as the trigonal lattice structure; the field is cylindrically symmetric. This is in contrast with the case of OM systems formed with linearly (or elliptically) polarized optical beams that cause spatially anisotropic interactions between the nanoparticle constituents. The reference structure is selected using an iteration algorithm, which is stated in detail in Section 2.3.3. The rotation angle θ to align the two structures is computed by lattice fitting (described in Section 2.2) by fitting $\boldsymbol{x}$ to $\boldsymbol{x}_0$. The resulting expression of θ can be written in three parameters p, q, and d defined as:

$$p = \sum_{i=1}^{N}(a_i x_i + b_i y_i), \quad q = \sum_{i=1}^{N}(b_i x_i - a_i y_i), \quad d = \sqrt{p^2 + q^2} \tag{2.27}$$

Then θ can be determined by:

$$\cos\theta = \frac{p}{d}, \quad \sin\theta = \frac{q}{d} \tag{2.28}$$

Therefore, the Cartesian coordinate of the rotation-aligned structure $\tilde{\boldsymbol{x}} = [\tilde{x}_1, \tilde{y}_1, \cdots, \tilde{x}_N, \tilde{y}_N]^T$ can be written as:

$$\tilde{x}_i = x_i \cos\theta - y_i \sin\theta, \quad \tilde{y}_i = x_i \sin\theta + y_i \cos\theta \tag{2.29}$$

According to Section 2.2.1, a set of collective coordinates $\boldsymbol{r} = [r_1, r_2, \cdots, r_{2N}]^T$ can be obtained for a stable OM structure, which is similar to the molecular case. As eigenvectors of the population covariance matrix of the system, they can be written in the form of:

$$r_j = \sum_{i=1}^{N} \left[O_{2i-1,j}(\tilde{x}_i - a_i) + O_{2i,j}(\tilde{y}_i - b_i)\right] \tag{2.30}$$

where O_{ij} is the $2N \times 2N$ orthogonal matrix, with each column vector $\boldsymbol{O}_{\cdot,j}$ (i.e., the j^{th} column of the matrix $\boldsymbol{O}$) representing a collective mode. Note that O_{ij} is a constant matrix independent of the Cartesian coordinates, so its partial derivative over the Cartesian coordinates are zero. Without loss of generality, we choose $\boldsymbol{O}_{\cdot,2N}$ to be the rotation mode, so that $r_{2N} = 0$ due to the rotation alignment by lattice fitting. Therefore, the new set of $2N$-dimensional coordinate is $\{r_1, r_2, \cdots, r_{2N-1}, \theta\}$, which is a set of orthogonal coordinates.

2.3.2.2 Partial derivatives of the collective coordinates over the Cartesian coordinates

Taking the partial derivative of eq. 2.30 gives:

$$\begin{aligned}
\frac{\partial r_j}{\partial x_k} &= \frac{\partial}{\partial x_k}\left\{\sum_{i=1}^{N}\left[O_{2i-1,j}(\tilde{x}_i - a_i) + O_{2i,j}(\tilde{y}_i - b_i)\right]\right\} = \sum_{i=1}^{N}\left(O_{2i-1,j}\frac{\partial \tilde{x}_i}{\partial x_k} + O_{2i,j}\frac{\partial \tilde{y}_i}{\partial x_k}\right) \\
\frac{\partial r_j}{\partial y_k} &= \frac{\partial}{\partial y_k}\left\{\sum_{i=1}^{N}\left[O_{2i-1,j}(\tilde{x}_i - a_i) + O_{2i,j}(\tilde{y}_i - b_i)\right]\right\} = \sum_{i=1}^{N}\left(O_{2i-1,j}\frac{\partial \tilde{x}_i}{\partial y_k} + O_{2i,j}\frac{\partial \tilde{y}_i}{\partial y_k}\right)
\end{aligned} \tag{2.31}$$

In order to compute eq. 2.31, we have to compute the partial derivatives of $\tilde{x}_i$ and $\tilde{y}_i$ by taking the derivatives of eq. 2.29:

$$\begin{aligned}
\frac{\partial \tilde{x}_i}{\partial x_k} &= \frac{\partial}{\partial x_k}(x_i\cos\theta - y_i\sin\theta) = \frac{\partial x_i}{\partial x_k}\cos\theta - x_i\sin\theta\frac{\partial\theta}{\partial x_k} - y_i\cos\theta\frac{\partial\theta}{\partial x_k} \\
&= \delta_{ik}\cos\theta - \tilde{y}_i\frac{\partial\theta}{\partial x_k} \\
\frac{\partial \tilde{x}_i}{\partial y_k} &= \frac{\partial}{\partial y_k}(x_i\cos\theta - y_i\sin\theta) = -x_i\sin\theta\frac{\partial\theta}{\partial y_k} - \frac{\partial y_i}{\partial y_k}\sin\theta - y_i\cos\theta\frac{\partial\theta}{\partial y_k} \\
&= -\delta_{ik}\sin\theta - \tilde{y}_i\frac{\partial\theta}{\partial y_k} \\
\frac{\partial \tilde{y}_i}{\partial x_k} &= \frac{\partial}{\partial x_k}(x_i\sin\theta + y_i\cos\theta) = \frac{\partial x_i}{\partial x_k}\sin\theta + x_i\cos\theta\frac{\partial\theta}{\partial x_k} - y_i\sin\theta\frac{\partial\theta}{\partial x_k} \\
&= \delta_{ik}\sin\theta + \tilde{x}_i\frac{\partial\theta}{\partial x_k} \\
\frac{\partial \tilde{y}_i}{\partial y_k} &= \frac{\partial}{\partial y_k}(x_i\sin\theta + y_i\cos\theta) = x_i\cos\theta\frac{\partial\theta}{\partial y_k} + \frac{\partial y_i}{\partial y_k}\cos\theta - y_i\sin\theta\frac{\partial\theta}{\partial y_k} \\
&= \delta_{ik}\cos\theta + \tilde{x}_i\frac{\partial\theta}{\partial y_k}
\end{aligned} \tag{2.32}$$

In order to compute eq. 2.32, we have to compute the partial derivatives of θ to get the form of $\boldsymbol{\nabla}\theta$. We take the derivatives for eq. 2.28:

$$
\begin{aligned}
-\sin\theta\frac{\partial\theta}{\partial x_k} &= \frac{\partial}{\partial x_k}\left(\frac{p}{d}\right) = \frac{1}{d^2}\left(d\frac{\partial p}{\partial x_k} - p\frac{\partial d}{\partial x_k}\right) = \frac{1}{d^2}\left[d\frac{\partial p}{\partial x_k} - \frac{p}{d}\left(p\frac{\partial p}{\partial x_k} + q\frac{\partial q}{\partial x_k}\right)\right] \\
&= \frac{1}{d^2}\left[da_k - \frac{p}{d}\left(pa_k + qb_k\right)\right] \\
\frac{\partial\theta}{\partial x_k} &= -\frac{d}{q}\frac{1}{d^3}(d^2a_k - p^2a_k - pqb_k) = -\frac{1}{d^2}(qa_k - pb_k) = -\frac{1}{d}(a_k\sin\theta - b_k\cos\theta)
\end{aligned} \tag{2.33}
$$

$$
\begin{aligned}
-\sin\theta\frac{\partial\theta}{\partial y_k} &= \frac{\partial}{\partial y_k}\left(\frac{p}{d}\right) = \frac{1}{d^2}\left(d\frac{\partial p}{\partial y_k} - p\frac{\partial d}{\partial y_k}\right) = \frac{1}{d^2}\left[d\frac{\partial p}{\partial y_k} - \frac{p}{d}\left(p\frac{\partial p}{\partial y_k} + q\frac{\partial q}{\partial y_k}\right)\right] \\
&= \frac{1}{d^2}\left[db_k - \frac{p}{d}\left(pb_k - qa_k\right)\right] \\
\frac{\partial\theta}{\partial y_k} &= -\frac{d}{q}\frac{1}{d^3}(d^2b_k - p^2b_k + pqa_k) = -\frac{1}{d^2}(qb_k + pa_k) = -\frac{1}{d}(a_k\cos\theta + b_k\sin\theta)
\end{aligned} \tag{2.34}
$$

2.3.2.3 Jacobian matrix between $\boldsymbol{r}$ and $\boldsymbol{x}$

Eq. 2.32 can be rewritten in the matrix form:

$$
\begin{aligned}
\tilde{\boldsymbol{J}} &= [\boldsymbol{\nabla}\tilde{x}_1, \boldsymbol{\nabla}\tilde{y}_1, \cdots, \boldsymbol{\nabla}\tilde{x}_N, \boldsymbol{\nabla}\tilde{y}_N] \\
&= (\boldsymbol{\nabla}\theta)[-\tilde{y}_1, \tilde{x}_1, -\tilde{y}_2, \tilde{x}_2, \cdots, -\tilde{y}_N, \tilde{x}_N] + I_N \otimes \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \\
&= (\boldsymbol{\nabla}\theta)\left[R\left(\frac{\pi}{2}\right)\tilde{\boldsymbol{x}}\right]^T + \boldsymbol{R}(\theta)^T = (\boldsymbol{\nabla}\theta)\tilde{\boldsymbol{x}}^T R\left(-\frac{\pi}{2}\right) + \boldsymbol{R}(-\theta)
\end{aligned} \tag{2.35}
$$

where $\tilde{\boldsymbol{J}}$ is the Jocobian matrix between the two sets of coordinates: $[\tilde{x}_1, \tilde{y}_1, \cdots, \tilde{x}_N, \tilde{y}_N]$ and $[x_1, y_1, \cdots, x_N, y_N]$, and $\boldsymbol{R}(\theta)$ is the $2N \times 2N$ block-diagonal rotation matrix that rotates each of the N particles of angle θ in the counterclockwise direction with respect to the origin.

Eq. 2.33 and eq. 2.34 can be written in the matrix form:

$$\boldsymbol{\nabla}\theta = -\frac{1}{d}\boldsymbol{R}(-\theta)[-b_1, a_1, -b_2, a_2, \cdots, -b_N, a_N]^T = -\frac{1}{d}\boldsymbol{R}(-\theta)R\left(\frac{\pi}{2}\right)\boldsymbol{x}_0 \tag{2.36}$$

Plugging eq. 2.36 into eq. 2.35 gives:

$$\begin{aligned}\tilde{\boldsymbol{J}} &= -\frac{1}{d}\boldsymbol{R}(-\theta)R\left(\frac{\pi}{2}\right)\boldsymbol{x}_0\tilde{\boldsymbol{x}}^T R\left(-\frac{\pi}{2}\right) + \boldsymbol{R}(-\theta) = \boldsymbol{R}(-\theta)\left[I_{2N} - \frac{1}{d}R\left(\frac{\pi}{2}\right)\boldsymbol{x}_0\tilde{\boldsymbol{x}}^T R\left(-\frac{\pi}{2}\right)\right] \\ &= R\left(\frac{\pi}{2} - \theta\right)\left[I_{2N} - \frac{1}{d}\boldsymbol{x}_0\tilde{\boldsymbol{x}}^T\right]R\left(-\frac{\pi}{2}\right)\end{aligned} \tag{2.37}$$

Next, we shall relate $\tilde{\boldsymbol{J}}$ to $\boldsymbol{J_r}$; the latter is the Jacobian matrix between the coordinates $\boldsymbol{r}$ and $\boldsymbol{x}$. Taking the gradient of both sides of eq. 2.30 gives:

$$\begin{aligned}\boldsymbol{\nabla}r_j &= \boldsymbol{\nabla}\sum_{i=1}^{N}\left[O_{2i-1,j}(\tilde{x}_i - a_i) + O_{2i,j}(\tilde{y}_i - b_i)\right] = \sum_{i=1}^{N}\left(O_{2i-1,j}\boldsymbol{\nabla}\tilde{x}_i + O_{2i,j}\boldsymbol{\nabla}\tilde{y}_i\right) \\ &= [\boldsymbol{\nabla}\tilde{x}_1, \boldsymbol{\nabla}\tilde{y}_1, \cdots, \boldsymbol{\nabla}\tilde{x}_N, \boldsymbol{\nabla}\tilde{y}_N]\boldsymbol{O}_{\cdot,j}\end{aligned} \tag{2.38}$$

Therefore,

$$\boldsymbol{J_r} = \tilde{\boldsymbol{J}}\boldsymbol{O} \tag{2.39}$$

Plugging eq. 2.37 into eq. 2.39 gives:

$$\boldsymbol{J_r} = R\left(\frac{\pi}{2} - \theta\right)\left[I_{2N} - \frac{1}{d}\boldsymbol{x}_0\tilde{\boldsymbol{x}}^T\right]R\left(-\frac{\pi}{2}\right)O \tag{2.40}$$

Since $\boldsymbol{R}(\theta)$ and $\boldsymbol{O}$ are orthogonal matrices, the absolute value of their determinants are one, so we have:

$$|\det \boldsymbol{J_r}| = \left|\det\left[I_{2N} - \frac{1}{d}\boldsymbol{x}_0\tilde{\boldsymbol{x}}^T\right]\right| \tag{2.41}$$

Note that $\boldsymbol{x}_0\tilde{\boldsymbol{x}}^T$ is a rank-1 matrix with $\tilde{\boldsymbol{x}}^T\boldsymbol{x}_0$ as its unique non-zero eigenvalue and $\boldsymbol{x}_0$ as its corresponding eigenvector. Therefore,

$$|\det \boldsymbol{J_r}| = 1 - \frac{\tilde{\boldsymbol{x}}^T \boldsymbol{x}_0}{d} = 1 - \cos\theta \tag{2.42}$$

2.3.2.4 Jacobian determinant in the collective coordinate

The Jocobian matrix $\boldsymbol{J_r}$ defined in the previous section does not correspond to the measured difference between the original configuration space and the collective coordinate space with rotation θ aligned. The reason is that, with θ aligned, the last collective coordinate should be set as θ instead of r_{2N}. Therefore, we should use a new Jacobian matrix $\boldsymbol{J}$, whose only difference from $\boldsymbol{J_r}$ is that the last column of $\boldsymbol{J}$ is $\boldsymbol{\nabla}\theta$ while the last column of $\boldsymbol{J_r}$ is $\boldsymbol{\nabla} r_{2N}$.

First, we can get the expression of $\boldsymbol{O}_{\cdot,2N}$ according to the rotation direction of the reference structure $\boldsymbol{x}_0$ and its unit norm:

$$\boldsymbol{O}_{\cdot,2N} = R\left(\frac{\pi}{2}\right)\frac{\boldsymbol{x}_0}{\|\boldsymbol{x}_0\|} \tag{2.43}$$

Taking the gradient on both sides of eq. 2.30 gives:

$$\boldsymbol{\nabla} r_{2N} = \sum_{i=1}^{N}\left(O_{2i-1,2N}\boldsymbol{\nabla}\tilde{x}_i + O_{2i,2N}\boldsymbol{\nabla}\tilde{y}_i\right) = \tilde{\boldsymbol{J}}\boldsymbol{O}_{\cdot,2N} \tag{2.44}$$

Plugging eq. 2.37 and eq. 2.43 into eq. 2.44 gives:

$$\begin{aligned}\boldsymbol{\nabla} r_{2N} &= R\left(\frac{\pi}{2}-\theta\right)\left[I_{2N} - \frac{1}{d}\boldsymbol{x}_0\tilde{\boldsymbol{x}}^T\right]R\left(-\frac{\pi}{2}\right)R\left(\frac{\pi}{2}\right)\frac{\boldsymbol{x}_0}{\|\boldsymbol{x}_0\|} \\ &= \boldsymbol{R}(-\theta)R\left(\frac{\pi}{2}\right)\left[1 - \frac{\tilde{\boldsymbol{x}}^T\boldsymbol{x}_0}{d}\right]\frac{\boldsymbol{x}_0}{\|\boldsymbol{x}_0\|} = \left(\frac{\tilde{\boldsymbol{x}}^T\boldsymbol{x}_0 - d}{\|\boldsymbol{x}_0\|}\right)\left[-\frac{1}{d}\boldsymbol{R}(-\theta)R\left(\frac{\pi}{2}\right)\boldsymbol{x}_0\right]\end{aligned} \tag{2.45}$$

Comparing eq. 2.36 and eq. 2.45 gives:

$$\boldsymbol{\nabla} r_{2N} = \left(\frac{\tilde{\boldsymbol{x}}^T\boldsymbol{x}_0 - d}{\|\boldsymbol{x}_0\|}\right)\boldsymbol{\nabla}\theta = \frac{d}{\|\boldsymbol{x}_0\|}(\cos\theta - 1)\boldsymbol{\nabla}\theta \tag{2.46}$$

Therefore,

$$\begin{aligned} \det J &= \frac{\|\boldsymbol{x}_0\|}{d(\cos\theta - 1)} \det \boldsymbol{J_r} \\ |\det J| &= \frac{\|\boldsymbol{x}_0\|}{d(1-\cos\theta)} |\det \boldsymbol{J_r}| \end{aligned} \tag{2.47}$$

Plugging eq. 2.42 into eq. 2.47 gives:

$$\begin{aligned} |\det J| &= \frac{\|\boldsymbol{x}_0\|}{d(1-\cos\theta)} \times (1-\cos\theta) = \frac{\|\boldsymbol{x}_0\|}{d} \\ &= \frac{\sqrt{\sum_{i=1}^{N}\left(a_i^2 + b_i^2\right)}}{\sqrt{\left[\sum_{i=1}^{N}(a_i x_i + b_i y_i)\right]^2 + \left[\sum_{i=1}^{N}(b_i x_i - a_i y_i)\right]^2}} \end{aligned} \tag{2.48}$$

Note that the expression of $|\det J|$ in eq. 2.48 does not depend on the choice of the first $2N-1$ orthonormal collective coordinates $\{\boldsymbol{O}_{\cdot,1}, \cdots, \boldsymbol{O}_{\cdot,2N-1}\}$. This means that $|\det J|$ can be computed before carrying out PCA that is used to obtain the collective coordinates. Note that when $N = 1$, according to eqn. 2.48, $|\det J| = 1/\sqrt{x_1^2 + y_1^2}$, which is the Jacobian determinant of the 2-dimensional polar coordinate system.

2.3.3 Algorithm of obtaining collective coordinates using weighted PCA: from trajectories of the OM system to output collective modes

2.3.3.1 The necessity of weighted PCA

Due to the cylindrical symmetry of the system force field in the case of circularly polarized optical trapping beams, the rotational degree of freedom is not confined; i.e., the variance along the rotation mode is infinite. Therefore, in order to carry out PCA on the trajectory data, we have to align the rotation angle of the structures to a fixed reference frame before centralizing the data (i.e., the structures in the trajectory). However, the alignment of the rotation angle changes the probability distribution that we use to calculate the covariance matrix. As a result, weighted PCA is necessary to be adopted to the non-uniform nature

of the new probability distribution in the collective mode space. The non-uniform nature of the new probability distribution in the collective mode space is caused by the fact that the Jacobian determinant between the collective mode space and the Cartesian space is not uniform.

We start by examining conventional PCA. Suppose we have a data set of the coordinates of the structures in the trajectory $X_{n\times 2N}$ sampled from a random vector $\boldsymbol{X} = [X_1, X_2, \cdots, X_{2N}]$ (note that this is used as a simplified notation instead of $\{X_i, Y_i\}$), where n is the number of data points and $2N$ is the dimension of each data point. Let $\boldsymbol{\mu} = [\mu_1, \mu_2, ..\mu_{2N}]$ be the mean of $\boldsymbol{X}$ and Σ_{ij} as the covariance matrix of $\boldsymbol{X}$:

$$\mu_i = \int X_i p(\boldsymbol{X}) d^{2N}X \tag{2.49}$$

$$\Sigma_{ij} = \int (X_i - \mu_i)(X_j - \mu_j) p(\boldsymbol{X}) d^{2N}X \tag{2.50}$$

where $d^{2N}X$ denotes $dX_1 \cdots dX_{2N}$ and $p(\boldsymbol{X})$ is the probability distribution of $\boldsymbol{X}$. We take $\bar{X}_i = \sum_k X_{ki}/n$ as the estimator of μ_i. An estimator is a mathematical nomenclature which describes the estimate of a statistical quantity. Then we take:

$$S_{ij} = \frac{1}{n}\sum_{k=1}^{n}(X_{ki} - \bar{X}_i)(X_{kj} - \bar{X}_j) \tag{2.51}$$

as the estimator of the covariance matrix. Then we carry out sample PCA on the data set X by solving the eigenvalue decomposition of S_{ij} and take the eigenvectors and eigenvalues as the principle components and corresponding variances, respectively.

Next, we consider the case where the rotation angle θ is aligned. This gives another set of coordinates $\{r_1, \cdots, r_{2N-1}, \theta\}$ whose Jacobian determinant over the Cartesian coordinate is $|\det J|$, which was obtained in Section 2.3.2.4. In the following discussion, the probability distribution $p(\boldsymbol{X}) = p(r_1(\boldsymbol{X}), \cdots, r_{2N-1}(\boldsymbol{X}))$ is independent of θ.

Similarly, if we use the same approach to define the mean and covariance, we obtain:

$$\begin{aligned}\tilde{\mu}_i &= \int r_i(\boldsymbol{X})p(\boldsymbol{X})d^{2N}X = \int r_i p(\boldsymbol{X}(r_1 \cdots, r_{2N-1}))|\det J|^{-1} d^{2N-1}r d\theta \\ &= \int r_i p(\boldsymbol{X}(r_1 \cdots, r_{2N-1}))\frac{2\pi}{|\det J|}d^{2N-1}r\end{aligned} \tag{2.52}$$

$$\begin{aligned}\tilde{\Sigma}_{ij} &= \int (r_i - \mu_i)(r_j - \mu_j)p(\boldsymbol{X}(r_1 \cdots, r_{2N-1}))|\det J|^{-1} d^{2N-1}r d\theta \\ &= \int (r_i - \mu_i)(r_j - \mu_j)p(\boldsymbol{X}(r_1 \cdots, r_{2N-1}))\frac{2\pi}{|\det J|}d^{2N-1}\end{aligned} \tag{2.53}$$

where $d^{2N-1}r$ denotes $dr_1 \cdots dr_{2N-1}$.

We can see that the expressions of the mean and covariance in eq. 2.52 and eq. 2.53 are actually calculating the mean and variance over a new distribution $\tilde{p} = \frac{2\pi p}{|\det J|}$. This is not the variance and mean that correspond to the equations for the mechanics of the OM system; in the latter, we take the gradient of $\ln p$ (instead of $\ln \tilde{p}$) as the expression of the conservative part of force field. Therefore, the mean and variance that we want to compute should have the following form:

$$\mu_i = Z \int r_i p(\boldsymbol{X}(r_1 \cdots, r_{2N-1}))d^{2N-1}r \tag{2.54}$$

$$\Sigma_{ij} = Z \int (r_i - \mu_i)(r_j - \mu_j)p(\boldsymbol{X}(r_1 \cdots, r_{2N-1}))d^{2N-1}r \tag{2.55}$$

where:

$$Z = \left[\int p(\boldsymbol{X}(r_1 \cdots, r_{2N-1}))d^{2N-1}\right]^{-1} \tag{2.56}$$

With the expressions eq. 2.54 and eq. 2.55, we cannot use the conventional estimator of the mean and variance; if we use eq. 2.51 in this case, then what is estimated is eq. 2.53 instead of eq. 2.55. What we should do is rewrite eq. 2.54 and eq. 2.55 in the following forms:

$$\mu_i = \int \left[r_i \frac{Z|\det J|}{2\pi}\right] p(\boldsymbol{X}(r_1 \cdots, r_{2N-1}))\frac{2\pi}{|\det J|}d^{2N-1}r \tag{2.57}$$

$$\Sigma_{ij} = \int \left[(r_i - \mu_i)(r_j - \mu_j) \frac{Z|\det J|}{2\pi} \right] p(\boldsymbol{X}(r_1 \cdots, r_{2N-1})) \frac{2\pi}{|\det J|} d^{2N-1} r \tag{2.58}$$

Then, the proper expression of the mean and covariance estimators are:

$$\bar{r}_i = \frac{1}{n} \sum_{k=1}^{n} \left(r_{ki} \frac{Z \left| \det J^{(k)} \right|}{2\pi} \right) = \frac{Z}{2n\pi} \sum_{k=1}^{n} \left(r_{ki} \left| \det J^{(k)} \right| \right) \tag{2.59}$$

$$S_{ij} = \frac{1}{n} \sum_{k=1}^{n} \left[(r_{ki} - \bar{r}_i)(r_{kj} - \bar{r}_j) \frac{Z \left| \det J^{(k)} \right|}{2\pi} \right] = \frac{Z}{2n\pi} \sum_{k=1}^{n} \left[(r_{ki} - \bar{r}_i)(r_{kj} - \bar{r}_j) \left| \det J^{(k)} \right| \right] \tag{2.60}$$

where $\left| \det J^{(k)} \right| = \left| \det J(X_{k,1}, \cdots, X_{k,2N}) \right|$.

Next, we take a constant random variable to find the normalization coefficient for eq. 2.59 and eq. 2.60. Suppose the sampling is from a constant variable that generates $r_{ki} = r^*$ for all k, then obviously, $\bar{r}_i = r^*$. Plugging this into eq. 2.59 gives:

$$\begin{aligned} r^* &= \frac{Z}{2n\pi} \sum_{k=1}^{n} \left(r^* \left| \det J^{(k)} \right| \right) \\ Z &= \frac{2n\pi}{\sum_{k=1}^{n} \left| \det J^{(k)} \right|} \end{aligned} \tag{2.61}$$

Plugging eq. 2.61 in to eq. 2.59 and eq. 2.60 gives:

$$\bar{r}_i = \sum_{k=1}^{n} (w_k r_{ki}) \tag{2.62}$$

$$S_{ij} = \sum_{k=1}^{n} \left[w_k (r_{ki} - \bar{r}_i)(r_{kj} - \bar{r}_j) \right], \tag{2.63}$$

where

$$w_k = \frac{\left| \det J^{(k)} \right|}{\sum_{l=1}^{n} \left| \det J^{(l)} \right|} \tag{2.64}$$

2.3.3.2 Steps for generating PC collective modes from fluctuation simulation trajectories of a stable OM structure

1. Select the initial input of reference structure. For a stable OM configuration formed in a circularly polarized beam, the trigonal lattice structure is used as the initial reference structure.

2. Calculate the trajectory data of the nanoparticle constituents of the OM system with rotation angles of the OM structures aligned to the reference structure by lattice fitting. The post-processed trajectory data is denoted as $X_{n\times 2N}$.

3. Compute the weighted average of the data:

$$\overline{X} = \boldsymbol{w}^T X, \tag{2.65}$$

   where $\boldsymbol{w} = [w_1, \cdots, w_n]^T$ is defined in eq. 2.64.

4. Compare $\overline{X}$ to the reference structure. If the difference between them is small, then move on to step 5. If the difference between them is large, then choose $\overline{X}$ as the reference structure input and iterate from steps 2 to 4 until $\overline{X}$ converges; i.e., until the 2-norm of the difference between the values of $\overline{X}$ calculated by two adjacent iterations is smaller than a threshold value of 1 nm.

5. Centralize the data (i.e., subtract the mean from the data):

$$\Delta X = X - \mathbf{1}_n \overline{X} \tag{2.66}$$

   where $\mathbf{1}_n$ is the n-dimensional column vector with all entries equal to 1.

6. Compute the covariance matrix:

$$S = \Delta X^T D_w \Delta X, \tag{2.67}$$

where D_w is a $n \times n$ matrix with $(D_w)_{ij} = w_i \delta_{ij}$

7. Carry out eigenvalue decomposition for the matrix S. The eigenvalues are the variances, and the eigenvectors are the collective modes.

2.4 Diffusion K-means Clustering

Markov state model is used for describing the dynamics of a system with multiple states: $C_1, C_2, \cdots, C_m|$. The model is built upon a trajectory $\boldsymbol{x}(t)$ that adapts to a probability transition matrix $T_{ij}(\tau) = \text{Prob}\{\boldsymbol{x}(t+\tau) \in C_j | \boldsymbol{x}(t) \in C_i\}$ where τ is the lag time. Detailed explanation in properties and examples of MSM can be found in Chapter 4.

In order to build the Markov state model (MSM) to compute the transition rates between the structural states of an OM system, we need to perform featurization and clustering analysis on the data set of configuration trajectories. The first step in the MSM construction pipeline is to project the simulation trajectories into their leading slow modes to define a low-dimensional embedding conducive to identification of the metastable states of the system using clustering algorithms [42]. This is typically achieved using time lagged independent components (tICA) analysis [43] or its kernel [44] or deep [45–47] variants. However, there are technical challenges in applying these methods to systems exhibiting full permutation symmetry, such as the OM system, where all particles are identical. Therefore, in the featurization procedure, we instead use diffusion maps [48–51], a nonlinear manifold learning method that can generate permutation invariant coordinates for clustering.

2.4.1 Pairwise distance calculation

Let $(x_p^{(i)}, y_p^{(i)})$ denote the 2D Cartesian coordinates of particle p in configuration i. We can calculate the distance matrix $\boldsymbol{M}^{(i)}$ for each configuration i with matrix elements,

$$M_{pq}^{(i)} = \sqrt{\left(x_p^{(i)} - x_q^{(i)}\right)^2 + \left(y_p^{(i)} - y_q^{(i)}\right)^2}. \tag{2.68}$$

Let $\boldsymbol{e}_k$ denote the unit column vector with the k^{th} component unity and others zero. Then the permutation-invariant distance defined between a pair of configurations i and j is,

$$d_{ij} = \min_{\boldsymbol{P} \in S_n} \sqrt{\sum_{k=1}^{N} \min_{\boldsymbol{Q}_k \in S_n} \left\| \boldsymbol{Q}_k \boldsymbol{M}^{(i)} \boldsymbol{e}_k - \boldsymbol{M}^{(j)} \boldsymbol{P} \boldsymbol{e}_k \right\|_2^2}, \tag{2.69}$$

where N is the number of particles, and S_n is the set of all permutation matrices so that $\boldsymbol{P}$ and $\boldsymbol{Q}_k$ are the optimal permutation matrices that minimize d_{ij}. Here, $\boldsymbol{P}$ and $\boldsymbol{Q}_k$ are $(N+1)$ independent permutation matrices to be optimized, in which $\boldsymbol{P}$ corresponds to the inter-column permutation while $\boldsymbol{Q}_k$ corresponds to the intra-column permutations for all the columns so that the norm of the difference of $M^{(i)}$ and $M^{(j)}$ is optimized over all inter-column and intra-column permutations. Then d_{ij} is a permutation-invariant pairwise distance for the configurations that serves as a kernel for the diffusion map calculations.

2.4.2 Diffusion maps

Diffusion maps are a type of non-linear manifold learning method that take the input of pairwise distances of the configurations and generate a low-dimensional non-linear subspace of the configuration space [49,50]. A brief introduction to diffusion map methodology is provided below, while full details of this method applied to colloidal self-assembly are discussed elsewhere [52–54].

First, the kernel matrix $\boldsymbol{K}^{(0)}$ is calculated with elements,

$$K_{ij}^{(0)} = \exp\left(-\frac{d_{ij}^2}{2\epsilon^2}\right), \tag{2.70}$$

where d_{ij} is the permutation-invariant pairwise distance defined in Section 2.4.1 and ϵ is the kernel bandwidth parameter that characterizes the adjacency among the configurations. Next, $\boldsymbol{K}^{(0)}$ is normalized to $\tilde{\boldsymbol{K}}$ to gain correspondence to Langevin dynamics [55, 56],

$$\tilde{K}_{ij} = \frac{K_{ij}}{\sqrt{\left(\sum_k K_{ik}\right)\left(\sum_k K_{kj}\right)}}. \tag{2.71}$$

$\tilde{\boldsymbol{K}}$ is then used to calculate the right-stochastic Markov transition matrix (RSMTM) $\boldsymbol{T}$,

$$T_{ij} = \frac{\tilde{K}_{ij}}{\sum_{j'} \tilde{K}_{ij'}}, \tag{2.72}$$

with eigenvalues $\{\lambda_k\}$ and right eigenvectors $\{\psi_k\}$. Since the components of ψ_1 are all unity, $\{\psi_k\}_{k=2}^{m+1}$ is taken as the basis of the low-dimensional nonlinear configuration subspace. An appropriate value of m is identified based on a gap in the eigenvalue spectrum. Finally, we obtain,

$$\left\{\left(x_p^{(i)}, y_p^{(i)}\right)\right\}_{p=1}^{N} \longrightarrow \{\psi_k(i)\}_{k=2}^{m+1}, \tag{2.73}$$

which maps the Euclidean coordinates of each configuration to its corresponding diffusion map embedding. After obtaining this m-dimensional permutation-invariant reduction, configurations are clustered into microstates.

2.4.3 Diffusion k-means clustering

The k-means clustering algorithm is a widely-used unsupervised clustering method [57]. Chen and Yang introduced diffusion k-means, which maximizes the within-cluster connectedness

based on the diffusion distance [58]. The diffusion distance is defined as the Euclidean distance in the eigenvector space of diffusion map embedding [48–50]. In other words, diffusion k-means is k-means clustering applied to diffusion map embeddings. In this dissertation, diffusion k-means is used as the microstate clustering algorithm so the clusters generated by diffusion k-means are termed "microstates" while the clusters of microstates generated by Robust Perron Cluster Cluster Analysis (PCCA+) [59–61] are called "macrostates" which are also the states in the Markov state model. The eigenvectors of a diffusion map correspond to different eigenvalues that characterize the time scale of transition between macrostates. Therefore, instead of executing k-means clustering directly on the basis set $\{\psi_k\}_{k=2}^{m+1}$, we execute it on the basis set $\{\lambda_k^{\tilde{t}}\psi_k\}_{k=2}^{m+1}$, where λ_k is the eigenvalue corresponding to ψ_k and $\tilde{t}$ is a parameter that characterizes the time scale of diffusion distances encountered in the k-means clustering [50]. As $\tilde{t}$ becomes larger, the eigenvectors with large eigenvalues become more important in the clustering, leading to merging of regions discriminated by higher order eigenvectors and discriminating microstates largely on the basis of the leading eigenvectors. On the other hand, when $\tilde{t}$ is small or even negative, the eigenvectors with small eigenvalues become important, leading to microstates as well as macrostates connected by fast transitions and merging the regions connected by slow modes. Therefore, we can adjust the value of $\tilde{t}$ such that our k-means clustering identifies the macrostates characterized by the time scale we want for studying the system dynamics.

CHAPTER 3

DATA-DRIVEN REACTION COORDINATE DISCOVERY IN OVERDAMPED AND NON-CONSERVATIVE SYSTEMS: APPLICATION TO OPTICAL MATTER STRUCTURAL ISOMERIZATION

This chapter is based on the following publication:

Shiqi Chen, Curtis W. Peterson, John A. Parker, Stuart A. Rice, Andrew L. Ferguson, Norbert F. Scherer. "Data-driven reaction coordinate discovery in overdamped and non-conservative systems: application to optical matter structural isomerization" *Nature Communications*, 2021, **12** (1), 2548.

3.1 Abstract

Optical matter (OM) systems consist of (nano-)particle constituents in solution that can self-organize into ordered arrays that are bound by electrodynamic interactions. They also manifest non-conservative forces, and the motions of the nano-particles are overdamped; i.e., they exhibit diffusive trajectories. We propose a data-driven approach based on principal components analysis (PCA) to determine the collective modes of non-conservative overdamped systems, such as OM structures, and harmonic linear discriminant analysis (HLDA) of time trajectories to estimate the reaction coordinate for structural transitions. We demonstrate the approach via electrodynamics-Langevin dynamics simulations six electrodynamically-bound nanoparticles coupled to an incident laser beam. The reaction coordinate we discover is in excellent accord with a rigorous committor analysis, and the identified mechanism is in good agreement with the experimental observations. The PCA-HLDA approach to data-driven discovery of reaction coordinates aid in understanding and eventually controlling non-conservative and overdamped systems including optical and active matter systems.

3.2 Introduction

A major goal of chemical research is the determination of the details of atomic rearrangements, bonding and reaction pathways of molecules [62]. Contrary to the simplified "general chemistry" perspective, true molecular reaction pathways are multi-dimensional and challenging to represent and visualize. The committor probability for a two-state system – the statistical probability that a particular configuration will transition into the product basin before the reactant basin – is the optimal reaction coordinate in that it is perfectly correlated with, and indeed defines, the extent of reaction [63]. As a purely statistical measure, the committor does not provide any configurational or physical understanding of the reaction mechanism, and simplified, dimensionally-reduced, configuration-based reaction coordinates are valuable in defining and quantifying the important collective motions driving transitions between metastable configurations [64]. A traditional approach to quantifying the important dynamical fluctuations of a system in a metastable configuration is to project them onto the leading (vibrational) normal modes of the system as a basis set that allows characterization of the soft collective fluctuations [22]. Alternatively, transition path sampling (TPS) and related methods [65, 66] can determine reaction paths in high dimensional systems by identifying high-probability reactive paths.

In addition to the aforementioned methods for molecular systems, atomic and molecular van der Waals clusters made in molecular beams and colloidal clusters in solution represent classes of materials that form close-packed configurations that are held together by Van der Waals (i.e., dispersion) interactions [67], depletion forces [68–70], or Casimir type interactions [71]. Their configurations are well-determined by repulsive "hard-sphere" interactions as described by Weeks, Chandler and Anderson (WCA) theory [72]. Since the interactions are typically short ranged, the transitions between different structural isomers are often single particle moves [69, 71] or correlated few particle moves [73] that are often interpretable in terms of broken 'bonds' and involve a small number of degrees of freedom. Although the

transition state configurations are a challenge to study in atomic clusters due to their small spatial scale and short lifetime, the size of colloidal systems readily allow conventional optical microscopy and fast imaging to be used for visualization of particle trajectories.

Optical matter (OM) is a type of material or molecule-like structure in which the constituents (e.g. nanoparticles (NPs) or micron-scale particles) are bound together by electrodynamic interactions [6, 7]. A fundamental aspect of OM structures is that they tend to occupy interparticle distances that are integer multiples of the wavelength of the incident optical field. This is particularly clear when the constituents are NPs that are smaller than the wavelength of light [7,8]. This interaction, known as optical binding [6,9,10], allows formation of regular OM configurations (e.g. 2D arrays with trigonal symmetry or anisotropic arrays with rectangular lattice configurations) with minimal optical information, generally only the overall shape, polarization, phase, and power of the optical beam [9] and the resulting particle configurations are readily visualized by optical microscopy [7, 11–13]. These systems are of particular interest in optical physics due to their manifestation of non-reciprocal forces, collective (correlated) interactions and many-body effects in their electrodynamic interactions [14–18] and can also serve as useful mesoscale analogs with which to study atomic-level chemical processes [19].

It is known that long range interactions manifest structural and dynamic correlations in driven dense colloidal solutions due to hydrodynamic interactions [20], and in quantum-dot perturbed molten salt solutions [21]. OM configurations self-organize in optical traps in solution and achieve new collective properties due to their electrodynamic interactions and long range (periodic) potentials [6, 7, 9, 10]. The long range interactions create a new richness for understanding isomerizations of OM configurations as they affect the energetics globally and imbue OM systems with many-body interactions and physics. Rational control of the configuration of optical matter requires solving a many-body problem by effective and practical methods.

Fig. 3.1 demonstrates the similarities between molecular and optical matter systems. The probability distribution of a molecular configuration in the neighborhood of the potential minimum can be well approximated by a multivariate Gaussian distribution with a covariance matrix governed by the Hessian matrix H at the minimum of the potential defined as:

$$H_{ij} = \frac{\partial U(\mathbf{r})}{\partial r_i \partial r_j} \tag{3.1}$$

where $\mathbf{r}$ is the Cartesian coordinates of the molecular configuration and U is the potential function. When the potential is harmonic, the normal modes of the system can be determined from the covariance matrix [22] (a detailed derivation provided in Section 3.6.1). Consider the asymmetric stretching mode of a water molecule shown in Fig. 3.1a-d as a simple example. The coordinate displacement exemplified in this mode is involved in the dissociation and autoionization reactions of water [23]. Consider a second example of boat-to-chair isomerization of cyclohexane, in which particular normal modes dominate others in effecting the isomerization transition [24]. Optical matter systems in optical traps in solution can undergo structural transitions. For example, Fig. 3.1e-g illustrate a transition in a 6-particle OM system from a "triangle" to a "chevron" configuration through a transition state. (Experimental details for creating the OM system and optical microscopy visualization are given in Section 3.5.) However, compared to molecular systems, besides friction, the external electrodynamic forces acted on the OM system in solution is non-conservative and overdamped [25]. Is it possible, then, to find an analog of normal mode analysis to OM systems in solution? In answer, we propose that this transition can be described by the collective modes of the OM system. The primary goal of this paper is to develop a collective coordinate analysis for non-conservative, overdamped systems and demonstrate the approach in defining a reaction coordinate and transition mechanism in an OM system structural isomerization.

Normal modes are orthogonal collective motions of particles that carry independent contributions to the system energy. The conventional definition of normal modes is valid only

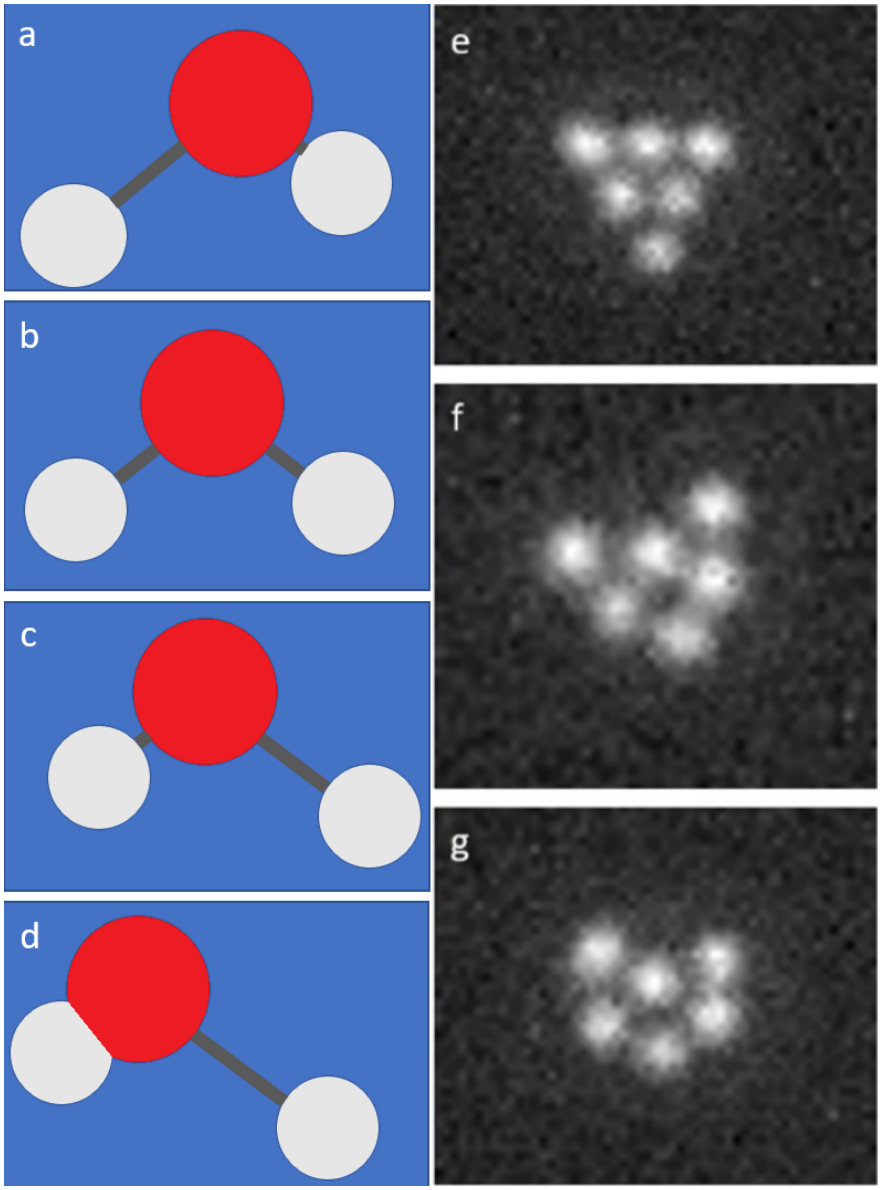

Figure 3.1: An analogy between molecular normal modes and the collective modes of optical matter systems. (a-d) The antisymmetric stretching mode of the water (H_2O) molecule. (e-g) Experimentally measured "instantaneous" OM configurations associated with a transition between triangle (e) and chevron (g) via and intermediate higher energy transition configuration (f) in the 6-particle optical matter system that forms and fluctuates in a converging Gaussian optical beam that is circularly polarized.

for harmonic particle-particle interactions. There is no formal mechanical definition of normal modes for overdamped and non-conservative systems. In other words, if the following Langevin equation is considered:

$$m\frac{d^2\mathbf{r}}{dt^2} = \mathbf{F}_{ext}(\mathbf{r},t) - \xi\frac{d\mathbf{r}}{dt} + \eta \tag{3.2}$$

where $\mathbf{r}$ is the position, m is the mass, $\mathbf{F}_{ext}$ is the external force field, ξ is the friction coefficient, and η is the random force, only when $\xi = 0$ and $\frac{\partial F_{ext,x}}{\partial y} = \frac{\partial F_{ext,y}}{\partial x}$ can normal modes be well-defined. It should be emphasized that the non-conservative nature of the OM system refers to the non-conservative (external) electrodynamic force field $\mathbf{F}_{ext}$ [25]. Renson and Kerschen have defined nonlinear normal modes in underdamped systems (ξ is not zero but not large enough to neglect the acceleration term) [26]. David and Jacobs have used principal component analysis (PCA) to study the large-scale fluctuations in molecular and colloidal systems [27]. Zaccone and co-workers have used instantaneous normal modes (INM) and the vibrational density of states in liquids to analyze none-affine dynamics of amorphous materials such as glassy polymers [28–31]. These two types of systems can be overdamped but conservative (ξ is large enough to neglect the acceleration term; $\frac{\partial F_{ext,x}}{\partial y} = \frac{\partial F_{ext,y}}{\partial x}$). See Section 3.6.1 for further discussion. Chattoraj, et al. have analyzed the eigenvalues and eigenvectors of J-matrix, the first derivative matrix of the external force field, and found oscillatory solutions of motion that are particularly useful for studies of underdamped non-conservative systems (ξ is not large enough to neglect the acceleration term; $\frac{\partial F_{ext,x}}{\partial y} \neq \frac{\partial F_{ext,y}}{\partial x}$) [32]. In overdamped cases, however, there exists no oscillatory solution while the non-orthogonal eigenvectors of the J-matrix lead to intrinsically coupled collective modes that are complicated to analyze. See Section 3.6.1 for additional discussion. In our approach we define collective modes in overdamped and non-conservative systems (ξ is large enough to neglect the acceleration term; $\frac{\partial F_{ext,x}}{\partial y} \neq \frac{\partial F_{ext,y}}{\partial x}$) by carrying out PCA on the configurational trajectories based on the deviations of the OM constituent particles from a reference configuration. PCA diagonalizes the covariance matrix of the OM particle coordinates to define a linear transformation into a basis of non-local collective modes (principal components, PCs) ordered by the degree of configurational variance they contain [33].

The leading PCs correspond to collective degrees of freedom with large variance that typically characterize large-scale global rearrangements of the system, whereas the trailing PCs correspond to small-variance fluctuations around particular metastable configurations. The leading PCs are therefore anticipated to serve as good descriptors for transitions between metastable system configurations. The PCs may further be formally converted into reaction coordinates using harmonic linear discriminant analysis (HLDA) [22].

In this work, we demonstrate this PCA-HLDA approach described above in an application to trajectories of the triangle-to-chevron transition, like the measured result of Fig. 3.1e-g, using combined electrodynamics-Langevin dynamics simulations [15, 34] of a 6-particle OM system. We note that the local fluctuation trajectory required for our PCA-HLDA approach must be adequately long with sufficiently fine time steps. Although this analysis could in principle be done by particle tracking analysis of experimental data [41], the transitions are rare and must be sampled at rates that are higher than can be readily obtained even in relatively high speed (100's fps) video measurements. We determine the contributions of each PCA collective mode to the transition, employ HLDA to formulate a reaction coordinate from these modes, validate the reaction coordinate using committor probability analysis, and use our results to define the transition state ensemble and reaction mechanism. This PCA-HLDA approach is analogous to those used to describe molecular reaction dynamics, but it is herein applied to an overdamped and non-conservative OM system [22].

3.3 Results

We demonstrate our methodology on an OM system, which is an open assembly of particles subject to a persistent flux of an external electromagnetic field that induces the interparticle interactions. The OM system is also overdamped owing to the mass of the 150 nm diameter silver nanoparticles such that the mass term in the Langevin equation becomes negligible. The primary data are from electrodynamics-Langevin dynamics (EDLD) simulations of a

6-particle optical matter system (see Section 3.5), and the results are corroborated with experimental data. We first describe the observed collective motions of the 150 nm diameter silver nanoparticles in a focusing (converging) optical trapping beam of this non-conservative system. Then we apply PCA to the deviations from the triangle (reactant) configuration obtained simulation trajectories to estimate the collective modes to serve as a basis to describe the transition to the chevron (product) configuration. Finally, we apply HLDA to the collective modes projected from a trajectory to obtain reaction coordinates to study the triangle-to-chevron configurational transition.

3.3.1 Large-scale Collective Modes.

The OM system consists of six silver nanoparticles of 150 nm diameter confined to a 2D plane and optically trapped by a circularly polarized focused laser beam, which makes the interactions isotropic in 2D [15,17]. The NPs tend to form metastable configurations that, to a first approximation, maximize the number of optical binding electrodynamic interactions, analogous to chemical bonds, with an inter-particle spacing of approximately one optical wavelength [74]. The particles therefore tend to adopt configurations based on a hexagonal lattice. While the optical binding energies can be several (2-10) k_BT units of thermal energy [7, 41, 74], the OM configurations undergo spontaneous configurational transitions between these metastable configurations driven by both thermal fluctuations and by non-conservative optical forces. The two most probable configurations for six particles under a circularly polarized laser beam are a "triangle" (Fig. 3.1e) and a "chevron" (Fig. 3.1g). Because energy is not well defined in the OM system, by saying a configuration is more stable we mean that it is more probable (more commonly observed in trajectories obtained from both simulation and experiment).

The EDLD simulations are relatively inexpensive to conduct, allowing us to obtain long trajectories of the OM system to obtain good statistical sampling of configurations with high

($\sim 1\mu$s) time resolution. The representative instantaneous configuration and deviations of the positions of the NPs from the stable triangle configuration shown in Fig. 3.2a-c indicates correlated motion of the particles over the course of the triangle-to-chevron configuration. In Fig. 3.2a-c we mark the positions of the NPs in the stable triangle configuration with red crosses and the instantaneous location of the particles (i.e., at a particular point in a trajectory) over the course of a representative transition from the triangle to the chevron configuration by black circles. Each configuration is rotated so that the sum of its squared deviations to the triangle reference configuration (red crosses) is minimized. Fig. 3.2d-f shows shows histograms of the particle positions collected in structural epochs during which the simulated 150 nm diameter silver nanoparticles transition from the triangle (reactant) configuration (Fig. 3.2d) to the chevron (products) configuration (Fig. 3.2f) via a transition state (Fig. 3.2e).The simulations performed here generate long trajectories of the 6-particle OM system that undergo many transitions between triangle and chevron. This transition conserves a mirror axis of symmetry in the OM structures. Therefore, a primitive and intuitive reaction coordinate to characterize the transition is the distance between the two particles on this mirror axis labeled "2" and "6" in Fig. 3.2a. By binning frames in the trajectories according to 20 nm increments of this distance we define 22 windows spanning the transition from the triangle to the chevron configuration. The empirical probability densities computed from histograms over all frames in each window are presented for the first, middle, and last window in Fig. 3.2d-f, respectively. The tight probability distributions in these probability density functions suggest that there are well-defined collective motions among the NPs as they execute the triangle to chevron transition. However, the "smearing" of some of the sites at the transition state and for the chevron make it ambiguous whether the reaction really proceeds along a simple 1D path, which motivates the PCA-HLDA approach. Fig. 3.7 in Section 3.6.1 shows the experimental counterpart of Fig. 3.2d-f obtained from microscopy measurements of a 6-particle OM system undergoing the analogous structural

transition.

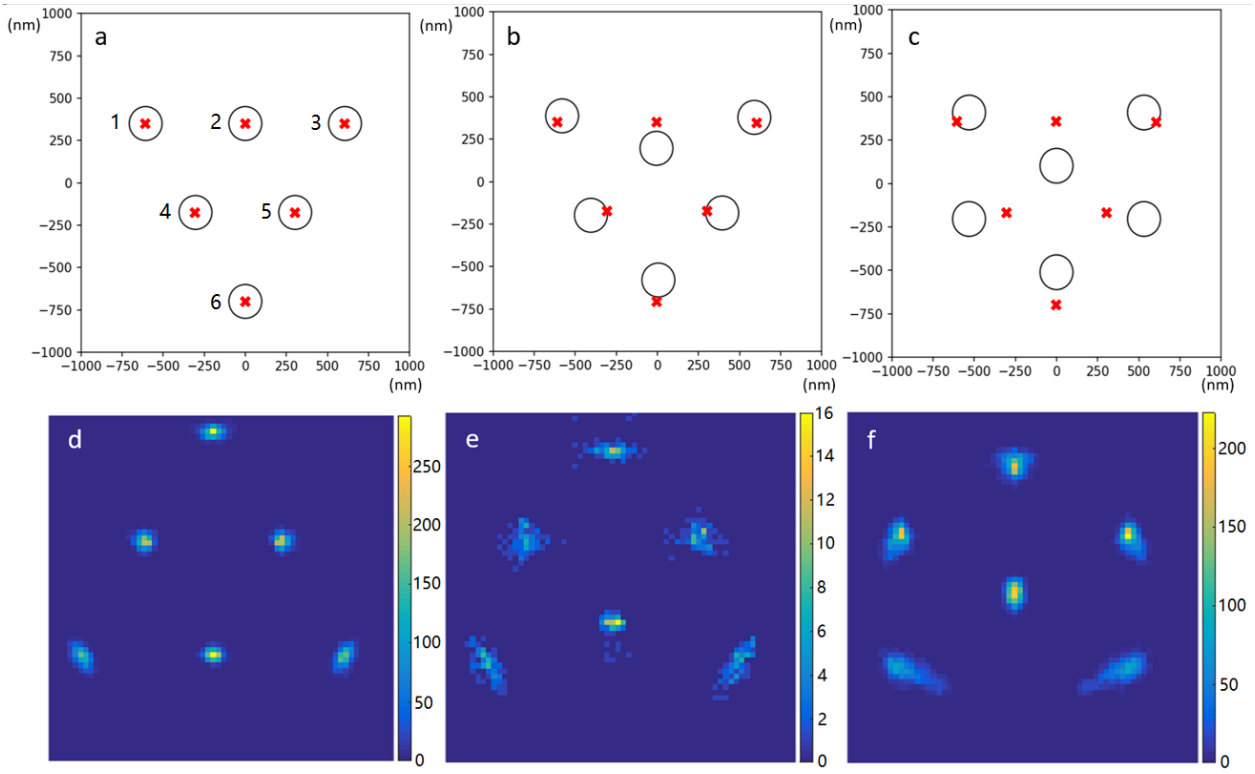

Figure 3.2: Particle positions and collective dynamics over the course of a triangle-to-chevron configurational isomerization. (a-c) Mean deviations of particle position(s) (black circles) from the metastable triangle configuration (red crosses) over the course of a representative transition. (d-f) Empirical probability densities of particle positions compiled from 21186 configurations at an optical power of 70 mW. The color scheme describes the number of configurations that contains a particle centered at a specific pixel. Configurations are binned into 22 windows of 20 nm in the distance between particles "2" and "6" (see panel a) and empirical probability density functions estimated by histogramming all configurations in each bin under a rotational and translational alignment to the mean particle positions (red crosses). The position probability density plots of the first, middle, and last window are presented in panels d-f, respectively.

Note that the initial perfect triangle structure corresponds to a trigonal lattice, which is taken as the reference structure for quantitative analysis of the particle fluctuations and deviations from the lattice sites. The distribution of lattice squared deviations shown in Fig. 3.3 also suggest the existence of collective modes in the OM system. Fig. 3.3a shows the cumulative density function (CDF) of lattice fitting deviation as the sum of squares of par-

ticle position deviations from the ideal triangle lattice sites (i.e., the red crosses in Fig. 3.2a) aggregated 2,087 simulation data points of the (local) fluctuations of the OM system in the triangle configuration at laser power 20 mW, 14,982 at 40 mW, 24,168 at 60 mW, and 39,172 at 100 mW. The simulation conditions are calibrated to actual laboratory experiments and correspond to an incident optical power of 100 mW of the trapping laser. The squares of particle position deviations were calculated after minimizing over all possible translations, rotations, and lattice constants such that four of the 12 degrees of freedom are eliminated during the lattice fitting. There are a total of $6 \times 2 = 12$ degrees of freedom for the 6-particle system constrained to a 2D plane. If the particle motions of a system are uncorrelated and described by independent identical Gaussian distributions (i.i.d. Gaussian), then the CDF of squared translationally and rotationally-minimized deviations should follow a χ^2 distribution with $(12 - 4) = 8$ degrees of freedom (black curve). Calculations in which we impose i.i.d. Gaussian fluctuations upon the six particles (red diamonds, inset) do indeed follow exactly this trend. On the other hand, the dynamics and CDF obtained from EDLD simulation (colored lines, main figure) exhibit significant deviations from this trend with increasing beam power. This result suggests the inference that the particle fluctuations around the metastable triangle configuration are not independent but are coupled into one or more collective mechanical modes by the electrodynamic interactions induced by the incident laser and coherent field. In Fig. 3.3b, we compute the deviation of the fitting displacement over 55,000 simulation data points harvested from the complete configurational space explored by the EDLD simulations (i.e., not just restricted to the local vicinity of the triangle configuration). Again, the fitting displacement distribution shows increasing deviations from the 8 degree of freedom χ^2 distribution as the intensity of the optical field increases. This further supports the result that the extent of correlated motion of the particles increases with the intensity of the optical trapping beam. We note that deviations about the lattice sites are, of course, partially due to random uncorrelated fluctuations and partially due to

correlated motions along collective coordinates and as the power increases the former become dominated by the latter.

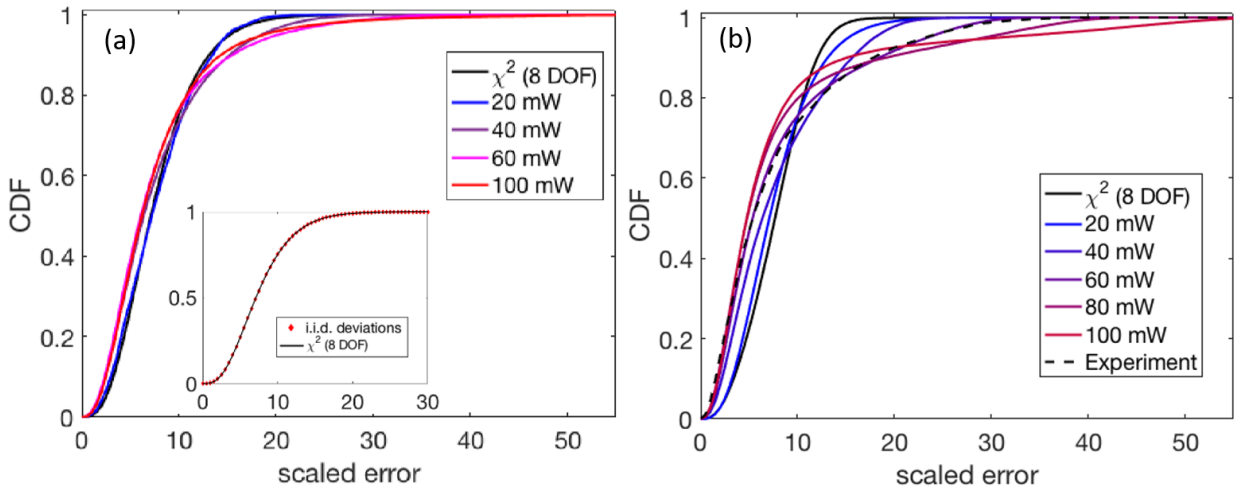

Figure 3.3: Cumulative density function (CDF) representation of lattice fitting displacement distributions of the 6-particle system in the local vicinity of the triangle configuration (a) and over the complete configuration space (b). The lattice fitting displacement is the sum of squares of particle position deviations of a configuration from the stable triangle lattice sites minimized over all possible translations, rotations, and the lattice constant. (a) The fitting displacement computed from EDLD simulation trajectories of local fluctuations in the vicinity of the triangle configuration deviates from the 8 degree of freedom χ^2 distribution that would be expected for particles executing uncorrelated i.i.d. Gaussian fluctuations (black curve) indicating the presence of collective structural modes. The inset shows a control simulation in which i.i.d. Gaussian fluctuations are imposed upon the particles (red diamonds) exactly follows the 8 DOF χ^2 distribution. The magnitude of the deviation of the CDF fitting displacement distribution from the 8 DOF χ^2 distribution (solid black curve) increases with optical trapping power in simulations conducted over the range 20-100 mW, indicating that the collective motions become more significant (and increase in magnitude) at higher optical powers. The corresponding probability density functions are shown in Fig. 3.8. (b) The fitting displacement computed from EDLD simulation trajectories over the whole configuration space (i.e., not just local fluctuations) of the 6-particle OM system deviates from the 8 degree of freedom χ^2 distribution. The magnitude of the deviation of the CDF fitting displacement distribution from the 8 DOF χ^2 distribution (solid black curve) increases with optical trapping power in simulations conducted over the range 20-100 mW (solid colored curves) and experimental data gathered at an optical power of 50 mW (dashed black curve), indicating that the collective motions become more significant (and increase in magnitude) at higher optical powers.

3.3.2 Principal Component Analysis and Definition of Collective Modes.

Now that we have confirmed the collective motions in the OM systems, we wish to study them quantitatively. We perform PCA on a single simulation trajectory that kept the lattice fitting displacement from the ideal triangle lattice to less than 250 nm to quantify the local collective fluctuations of the OM system around the triangle (reactant) configuration. Before PCA is carried out, each configuration is rotated according to the center of the field (i.e., the focused laser beam in the experiment) so that its fitting displacement with respect to the ideal triangular lattice is minimized. If this is not done the leading PCs will be contaminated by trivial rotations. The data set is then centered to a common origin by subtracting from the particle positions, the location of the center of the stable triangle configuration. The reason why other degrees of freedom such as translations are preserved is that the symmetry of the field is such that it is only invariant with rotation. However, with respect to the analogy with molecular systems that discard translations, even if unimportant degrees of freedom are taken into account, the result will not be harmed. Therefore, to be safe, only rotation is eliminated in data preprocessing.

The PCA identifies a linear transformation of the 12 degrees of freedom of the 6-particle system constrained to a 2D plane into a new basis of 12 orthogonal PCs arranged in order of decreasing configurational variance. The 12 PCs are illustrated in Fig. 3.4 are ranked in order of largest to the smallest eigenvalues. PCs 1, 2, and 12 correspond to rigid body transformations: PC 1 and 2 correspond to global translations in directions indicated by the black arrows and PC 12 to global rotation. The remaining PCs correspond to the collective fluctuations around the (metastable) triangle configuration. Some modes, such as PC 3, are nearly symmetric, while others, such as PC 4, are nearly antisymmetric. The symmetry axes of symmetric modes are indicated by dashed lines.

We verified that the discovered PCs represent stable collective modes by showing that they are preserved upon analyzing long simulation trajectories. Fig. 3.5a presents a heatmap

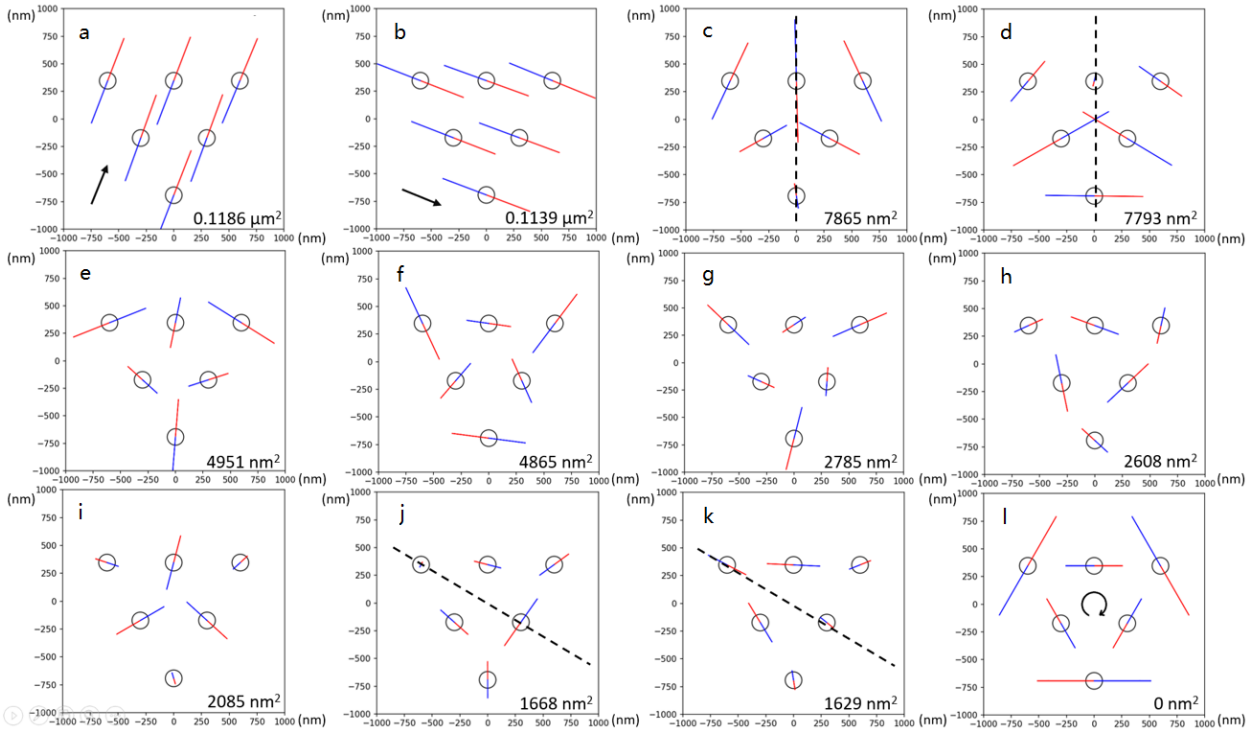

Figure 3.4: Principal components of the 6-particle triangle configuration. Panels a to l correspond to PC 1 to 12, respectively. The colored solid lines depict the directions and magnitudes of the collective particle motions, and the color defines the sense (i.e. phase) of the motion; i.e., particles simultaneously move in the indicated directions for the same color. PCs 1 and 2 correspond to rigid translation and PC 12 to rigid rotation. For the non-rigid transformations, PCs 3 to 11, the length of the solid lines is proportional to the PCA eigenvalues, λ_i, which are reported in the bottom right of each panel. The value of the 12^{th} eigenvalue is exactly zero due to elimination of rotational motion in application of PCA. Dashed black lines represent axes of symmetry. Black arrows indicate global translational or rotational motions.

reporting the inner product norms between the PCs computed from a 2-million-frame data set from a shorter trajectory and those computed from a 5-million-frame data set from a longer simulation. The matrix element (i,j) represents the norm of the inner product between the i^{th} PC of the 2-million-frame trajectory and the j^{th} PC of the 5-million-frame trajectory. If the PCs computed over the two data sets are identical, we would observe the identity matrix with elements (i,j) = 1 for i=j, and 0 otherwise. Therefore, the closer to the identity matrix Fig. 3.5a is, the better converged we may assess the modes to be and therefore

stable with increasing trajectory length. The matrix in Fig. 3.5a is close to identity with the exception of large off-diagonal elements between four pairs of PCs: (1,2), (3,4), (5,6), and (10,11). These deviations from the identity matrix arise from the degenerate character of these four pairs (see Fig. 3.4) so that the corresponding PCs are resolved only up to an arbitrary angle within the eigenspace. The off-diagonal couplings we observe in Fig. 3.5a are due to linear mixing within this degenerate subspace. As such, the four *degenerate pairs* are robust between the 2-million-frame and 5-million-frame data sets, although the individual modes within these pairs are not, due to the arbitrary breaking of the degeneracy induced by PCA.

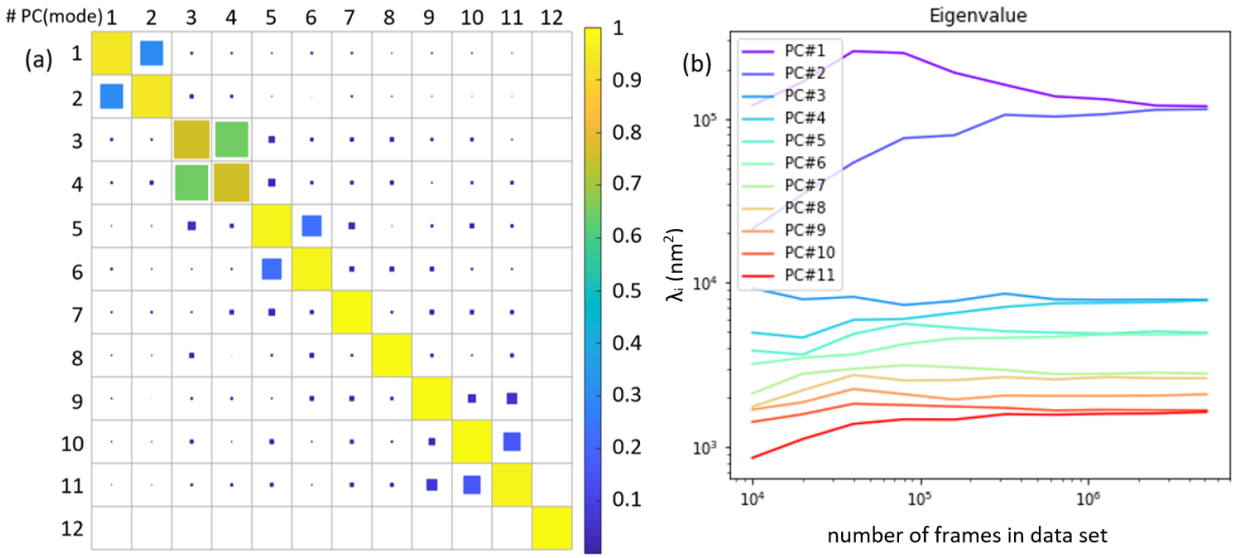

Figure 3.5: Convergence test of the PCs. (a) The norm of the inner-product between right singular vectors from PCA performed on a 2-million-frame (rows) and a 5-million-frame (columns) trajectory. (b) The eigenvalues λ_i of the PCs i=1-11 plotted against the number of frames in the analyzed data set.

Fig. 3.5b shows how the eigenvalues of the PCA converge as the size of the data set increases from 10,000 to 5 million frames. We observe convergence of all eigenvalues $\{\lambda_i\}$ to stable values by around 4 million frames, indicating that our 5 million frame data set is sufficiently large to obtain converged and stable PCs. We also observe that the eigenvalues of the four degenerate PC pairs (1,2), (3,4), (5,6), and (10,11) converge to identical values

within 4% error. Therefore, in light of the results in Fig. 3.5, we assert that the convergence test is successful, and that we have demonstrated that the PCs are indeed well-defined collective modes of the system.

3.3.3 HLDA Definition of Reaction Coordinates.

Having validated the PCs as proper collective modes, we seek to convert them into reaction coordinates for the triangle to chevron configurational isomerization using harmonic linear discriminant analysis (HLDA) [22]. The reaction coordinate is valuable in illuminating the transition mechanism, identifying the transition state ensemble, and providing a physically-motivated measure of reaction progress that is vastly cheaper to compute than the committor probability and more configurationally informative.

The reaction coordinate generated by HLDA, for the transition from the triangle configuration to the chevron configuration, is defined as [22],

$$s_{HLDA}(\mathbf{R}) = (\mu_A - \mu_B)^T(\Sigma_A^{-1} + \Sigma_B^{-1})d(\mathbf{R}) \tag{3.3}$$

where d is the 12-by-12 linear transformation matrix that converts particle positions $\mathbf{R}$ into the collective mode basis defined by the collective modes, and μ and Σ are the mean and the covariance matrices of the collective modes. The subscripts A and B represent the triangle (reactant) and chevron (product) configurations, respectively. This expression can be considered as the projection of the coordinate in the collective mode basis onto the vector $\mathbf{W}^*$ that maximizes the Rayleigh ratio:

$$\mathbf{W}^* = \arg\max_{\mathbf{W}} \frac{\mathbf{W}^{\mathrm{T}}\mathbf{S}_b\mathbf{W}}{\mathbf{W}^{\mathrm{T}}\mathbf{S}_w\mathbf{W}} = (\Sigma_A^{-1} + \Sigma_B^{-1})(\mu_A - \mu_B) \tag{3.4}$$

where $\mathbf{S}_b = (\mu_A - \mu_B)(\mu_A - \mu_B)^{\mathrm{T}}$ is the between class scatter matrix, and $\mathbf{S}_w = (\Sigma_A^{-1} + \Sigma_B^{-1})^{-1}$ is the within class scatter matrix [22]. The vector $\mathbf{W}^*$ can therefore be interpreted

as the direction along which the two classes are best separated [22].

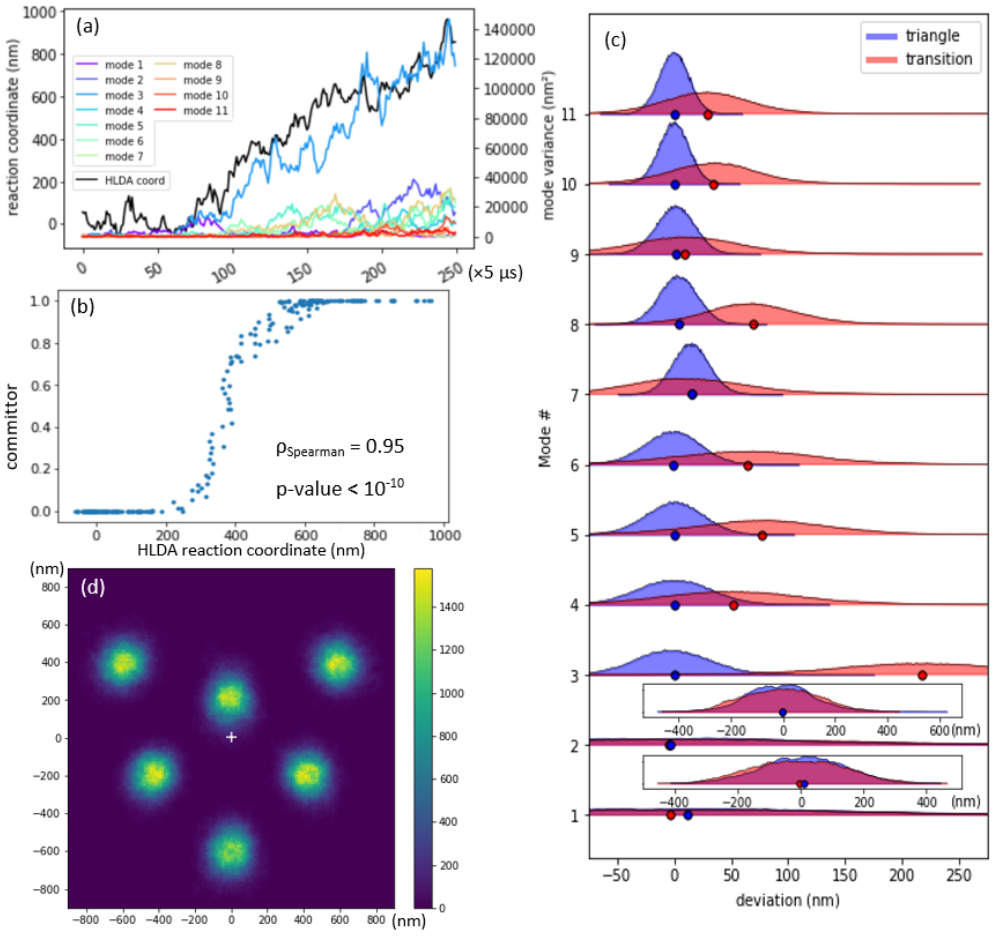

Figure 3.6: Validation and analysis of the PCA-HLDA reaction coordinate. (a) Comparison of the HLDA reaction coordinate s_{HLDA} and the variance of particle deviations along each of the 12 PCs over the course of a single triangle to chevron isomerization transition. (b) Calibration plot of the committor probability against the HLDA reaction coordinate for 250 selected configurations. The high rank-order correlation ($\rho_{spearman} = 0.95$) and low scatter validates the PCA-HLDA reaction coordinate as a good measure of reaction progress. (c) Distribution of particle deviations from the metastable triangle configuration projected onto all 11 non-trivial (omitting rigid rotation) PCs over the course of the transition for configurations with committor values 0.1-0.9 (red) and in the metastable triangle configuration with committor values < 0.1 (blue). Means of each distribution are indicated by colored dots. PC 3 changes most markedly in moving from the metastable reactant basin of the triangle configuration to executing the isomerization transition. (d) Probability distribution of aligned particle positions computed over 395,860 configurations in the transition region with 300 nm $< s_{HLDA}(\mathbf{R}) <$ 500 nm (committor probabilities in the range 0.1-0.9); the white cross indicates the average center of mass of these configurations. The color scheme describes the number of configurations that contains a particle centered at a specific pixel.

Fig. 3.6a shows the HLDA reaction coordinate, $s_{HLDA}(\mathbf{R})$, and the square of the projection of position deviation vector (projected variance) onto each of the collective modes for one specific transition trajectory. Analogous plots for other representative trajectories and transitions are presented in Fig. 3.9. The presented trajectory, which is typical of the majority of the transitions observed in simulation of the triangle-to-chevron transition, reveals a strong correlation between the HLDA coordinate (black curve) and PC 3 (blue curve) with Spearman correlation coefficient 0.93 and p-value less than 10^{-10}, implying that PC 3 is the dominant collective mode that contributes to the HLDA reaction coordinate characterizing the triangle to chevron isomerization. This is intuitively reasonable considering the motions embodied in PC 3 (see Fig. 3.4 and Fig. 3.2d-f).

The committor probability associated with a particular configuration is calculated by computing the probability that a trajectory initialized from that specific configuration arrives at the chevron configuration before arriving at the triangle configuration [63, 75]. A committor of zero indicates that the configuration is bound to arrive in the reactant basin before the product, a committor of unity means that it will first arrive in the product basin, and a committor of 0.5 that it has equal chance of first arriving in the reactant or product basin and is, by definition, a member of the transition state ensemble (TSE). We estimated the value of the committor for each OM structural configuration by initializing 420 trajectories at each of 250 configurations in a triangle-to-chevron transition trajectory, spanning the range of the HLDA reaction coordinate and computing the fraction that arrive in the chevron before the triangle. Fig. 3.6b shows that the committor agrees well with the HLDA reaction coordinate because the committor changes monotonically with the HLDA reaction coordinate and the scatter in the plot has small variance. We identify from the plot, for example, that configurations with a $s_{HLDA}(\mathbf{R}) < 300$ nm in the vicinity of reactant, $s_{HLDA}(\mathbf{R}) > 500$ nm in the vicinity of product, and 300 nm $< s_{HLDA}(\mathbf{R}) < 500$ nm as in the transition region. This committor analysis validates the HLDA reaction coordinate as a reliable

structural measure of reaction progress, a useful means to identify the TSE, and as a tool to understand the isomerization mechanism.

We use the committor probability to perform two additional analyses of the configurational mechanism of the isomerization transition. First, to further quantify and illuminate the significance of PC 3 within the HLDA reaction coordinate, we extracted 1000 transition trajectories from the EDLD simulation trajectories and extracted configurations with committor probabilities in the range 0.1-0.9. Since the PCs form a complete basis set of the space of position deviation vectors, the position deviation vector of each configuration can be written as a linear combination of the PCs, in which the coefficients can be obtained by the following orthogonal projection:

$$\mathbf{r} - \mathbf{r}_0 = \sum [\mathbf{v}_i^{\mathrm{T}}(\mathbf{r} - \mathbf{r}_0)]\mathbf{v}_i \tag{3.5}$$

where, the $\mathbf{v}_i$'s are the principal components obtained from PCA. $\mathbf{r}_0$ is the coordinate of the triangle configuration, and $\mathbf{r}$ is the coordinate of a specific configuration in the trajectory. We then project the deviation vector of these transition configurations onto the 11 non-trivial collective modes – omitting the trivial rotation mode that was eliminated in our PCA analysis – to identify the distribution of configurational deviation from the triangle pattern in each of these modes over the course of the transition (Fig. 3.6c, red shading). We compare these distributions in each collective mode to those harvested from local fluctuation around the metastable triangle (Fig. 3.6c, blue shading). We plot the mean values of each distribution as a dot and present the precision of the mean values in Fig. 3.10. This analysis clearly illustrates that PC 3 dominates among the linear combination coefficients when configurational deviations are projected onto the PCs. In other words, configurational deviations in PC 3 change most markedly in moving from the metastable triangle configuration to executing the isomerization transition while the other PCs only a small coefficient. Note that the variance shown in Fig. 3.4 and the projection magnitudes shown in Fig. 3.6c are

descriptions for two different processes. Fig. 3.4 describes a trajectory that contains only the triangle configuration and its local fluctuations with no pattern transition. Fig. 3.6c describes an ensemble of trajectories of a specific pattern transition from triangle to chevron. Therefore, the large variance of PC 4 described in Fig. 3.4 (local fluctuation of triangle) has no direct relation to whether it leads to the transition considered in Fig. 3.6c (transition from triangle to chevron). In addition, Fig. 3.6c only describes one specific transition starting from triangle (to chevron), so it is totally possible that mode 4 may dominate transitions to other states. Second, we collect the 395,860 configurations in the transition region with 300 nm $< s_{HLDA}(\mathbf{R}) <$ 500 nm and present the probability density of aligned particle positions in Fig. 3.6d. The transition state is identifiable a configuration intermediate to the triangle and chevron configurations accessed by a collective motion along PC 3 as the dominant contributor to the HLDA reaction coordinate (cf. Fig. 3.4, red arrows).

3.4 Discussion

Normal mode analysis is often used to study the statistics of the system configurations for conservative and undamped systems. However, this approach cannot be used for systems that are non-conservative, overdamped, and in which the configurational transformation involves large particle displacements. In this paper, we report an approach based on PCA to identify important collective fluctuations in non-conservative and overdamped systems and then use HLDA to transform these into a reaction coordinate for a configurational transition. We show that the PCs are stable collective modes and provide an interpretable basis for constructing and understanding the reaction coordinate. We demonstrate our approach in the triangle to chevron transition in numerical simulations of a 6-particle optical matter system and show that our results are consistent with experimental observations of the system. The HLDA reaction coordinate is shown to be valuable in resolving the transition mechanism and is validated as a robust reaction coordinate by committor analysis.

Data-driven discovery of reaction coordinates and kinetic transition rates is the first step in defining a kinetic network of the OM dynamics characterizing the metastable states and inter-state transition rates for the system. A number of methods and tools have been developed to understand the effect of incident field on configurations, stability and non-conservative dynamics of OM arrays [76–78], including the spin angular momentum of light in optical tweezers used to introduce the driven spin of individual NPs [79–83], and interparticle electrodynamic interactions that create orbital rotation in OM arrays [15, 16, 84]. Therefore, determining the kinetic network and quantifying the effect of these interventions upon the OM kinetic networks offers a new route to engineer and control the stability and transitions of particular OM structural isomers with applications in optical matter machines [17].

In future work, we anticipate that the PCA-HLDA approach to reaction coordinate identification can be used to understand and quantify other configurational transitions in diverse optical matter systems and to construct kinetic networks for the global system dynamics. We anticipate that PCA-HLDA will prove particularly valuable for systems containing large numbers of particles where human intuition can often fail. There are several challenges associated with broader implementation of the PCA-HLDA method. The first is the magnitude of the data required. We show in Fig. 3.5b that the eigenvalues of the PCs take around 2 million frames to converge for this system (with a convergence resolution 4%). While it is technically possible to acquire this number of experimental frames (i.e. images of individual configurations) with the correct number of particles, a highly automated acquisition and analysis process would be required. Also, since our current implementation of PCA-HLDA was developed for short simulated trajectories where each particle is assigned to a specific lattice site while the experimental data will contain several rearrangements where lattice assignment will switch, developing a method to consistently assign experimentally obtained particle positions to specific lattice sites is not trivial. Therefore, we also anticipate that PCA-HLDA may, at least in the case of slow transitions that can be adequately character-

ized, be applied directly to experimental data sets by developing a method to consistently assign experimentally obtained particle positions to specific lattice sites for the trajectories obtained from experiments that involve rather frequent particle rearrangements.

We also anticipate a number of elaborations and improvements of PCA-HLDA. First, the method requires definition of a reference configuration to which the trajectory snapshots are aligned prior to application of PCA. In this work, we adopt the triangle configuration that is on hexagonal lattice sites as this reference is defined by a local minimum in the non-conservative force field. As the reactant configuration for the isomerization transition, this represents a natural choice, but we could also have adopted the product chevron configuration for this purpose. Determination of reference configurations that do not lie on well defined lattice sites or exhibit high variance around a marginally metastable mean may be challenging, so that the PCA-HLDA approach may not be directly applicable for these configurations. Second, we see profitable integrations of PCA-HLDA with unsupervised nonlinear dimensionality reduction and clustering techniques, to first learn the metastable configurations and define which pairs are connected by configurational transitions in a data-driven manner and then use PCA-HLDA to identify reaction coordinates for transitions between each reactive pair.

Transitions involving indirect pathways, as shown in Fig. 3.11, are also observed for the triangle-to-chevron isomerization. However, they are infrequent and are also not included in the HLDA analysis presented. Such pathways with intermediates will be considered in detail in future work. Apart from optical matter systems, the approaches presented here are promising in other non-conservative and overdamped systems such as active matter systems. Compared to optical matter, active matter is driven out of thermal equilibrium by stored or locally supplied free energy [18]. Approaches by Speck have used the work required to deform a certain volume of active matter to derive the dynamics of the active matter system [85], and work by Takatori and Brady has focused on an effective free energy for active particles [86].

With our research, it can be expected that the dynamics of active matter can be further explored.

3.5 Methods

Simulations. Electrodynamics-Langevin dynamics (EDLD) simulations were performed with the Generalized Multiparticle Mie Theory (GMMT) using the MiePy software developed by our group [17,39]. Silver NPs with 150 nm diameter were used as the material constituents of the optical matter configurations. The nanoparticles were illuminated with a defocused, converging right-hand circularly (RHC) polarized Gaussian beam with a width $w = 2500$ nm, power $P = 50$ mW (except Fig. 3.3, where P is varied), and defocus equals to the Rayleigh range, $z = 0.5kw^2$, where $k = 2\pi n_b/\lambda$ and n_b is index of refraction. These field/beam conditions allowed formation of stable 6-nanoparticle optical matter (OM) arrays even in the presence of thermal noise/forces. The electrodynamic forces were passed into an overdamped Langevin equation to integrate the equation of motion for the OM array with a 5 μs time-step using a simple first-order Euler integrator [87]. Two hours are required for a one-million-time-step trajectory on a 2.4 GHz Intel E5-2680 v4 CPU.

Experiments. Experiments were conducted using a single-beam circularly polarized optical tweezer in an inverted microscope setup as described in [7, 15]. A dilute water solution with a mixture of polyvinyl pyrrolidone (PVP)-coated 150 nm Silver NPs was used. A continuous wave (CW) Ti-Sapphire laser beam ($\lambda = 800$ nm) was focused near the glass cover-slip using a 60x microscope objective, pushing a small number of NPs toward the glass surface. The laser power was 200 mW before entering the microscope, where additional power is lost before focusing towards the sample. A spatial light modulator (SLM) was used to slightly defocus the trapping beam such that it was converging at the sample. Electrostatic repulsion between the ligands on the NPs and the glass cover-slip balances the radiation pressure, resulting in a 2D trapping envrionment.

Lattice Fitting. Given a certain optical matter configuration with the number of particles known as N, we want to find the best set of N sites on a 2d hexagonal lattice pattern that is closest to the given configuration. Let the positions of the particles in the given configuration be $\mathbf{r}_1$, $\mathbf{r}_2$, ..., $\mathbf{r}_N \in \mathbb{R}^2$. Let the positions of the hexagonal lattice sites be $\mathbf{s}_1$, $\mathbf{s}_2$, ..., $\mathbf{s}_N \in \mathbb{R}^2$. The formula of minimization of the fitting error is shown below, where Π is the set that contains all injections from 1, 2, ... N to $\mathbb{Z}^+$.

$$Err^* = \left[\min_{a\in\mathbb{R}^+, \mathbf{r}_0\in\mathbb{R}^2, \hat{\mathbf{R}}\in SO(2), \pi\in\Pi} \sum_{j=1}^{N} \left| \hat{\mathbf{R}}(\mathbf{r}_j + \mathbf{r}_0) - a\mathbf{s}_{\pi(j)} \right|^2 \right]^{\frac{1}{2}} \tag{3.6}$$

Here, a is the lattice constant, r_0 is the translation vector, $\hat{\mathbf{R}}$ is the rotation matrix, and π is the lattice assignment. Considering the symmetry of the lattice sites, the parameter domain can be further limited by forcing $\pi(1) = 1$ and $\mathbf{r}_0$ within the primitive cell. Then, the parameter space is discretized and optimized to get the best assignment between the particles and the lattice sites, π^*.

Next, with the assignment in hand, an analytical solution of the optimal translation, rotation, and lattice constant can be found. Let $p_j = \mathbf{r}_j[1] + \mathbf{r}_j[2]i$ and $q_j = \mathbf{s}_{\pi(j)}[1] + \mathbf{s}_{\pi(j)}[2]i$ ($j \in \{1, 2, ..., N\}$), so $\mathbf{p}, \mathbf{q} \in \mathbb{C}^N$. The fitting error expression can be transformed to the one below, where p_0 is the translation, θ is the rotation angle, $\mathbf{1}$ is the vector of all ones, and a is the lattice constant. The superscript "*" indicates the optimal value. The detailed derivation of the results below is presented in Section 3.6.2.

$$Err^* = \left[\min_{a\in\mathbb{R}^+, p_0\in\mathbb{C}, \theta\in[0,2\pi)} \left| e^{i\theta}(\mathbf{p} + p_0\mathbf{1}) - a\mathbf{q} \right|^2 \right]^{\frac{1}{2}} \tag{3.7}$$

If translation is exempted from optimization, then $p_0^* = 0$. Otherwise:

$$p_0^* = \frac{\mathbf{1}^{\mathrm{T}}}{N}(\mathbf{q} - \mathbf{p}) \tag{3.8}$$

Let $\mathbf{p'} = \mathbf{p} + p_0^*\mathbf{1}$, then we have:

$$e^{i\theta^*} = \frac{(\mathbf{p'})^{\mathrm{H}}\mathbf{q}}{\left|\mathbf{q}^{\mathrm{H}}\mathbf{p}\right|} \tag{3.9}$$

If lattice constant is exempted from optimization, then $a^* = 1$. Otherwise:

$$a^* = \frac{\left|\mathbf{q}^{\mathrm{H}}\mathbf{p'}\right|}{|\mathbf{q}|^2} \tag{3.10}$$

3.A Supplementary Information

The supplementary information contains figures that show comparisons between experimental and simulation results and provide supporting details to the figures and discussion of the main text.

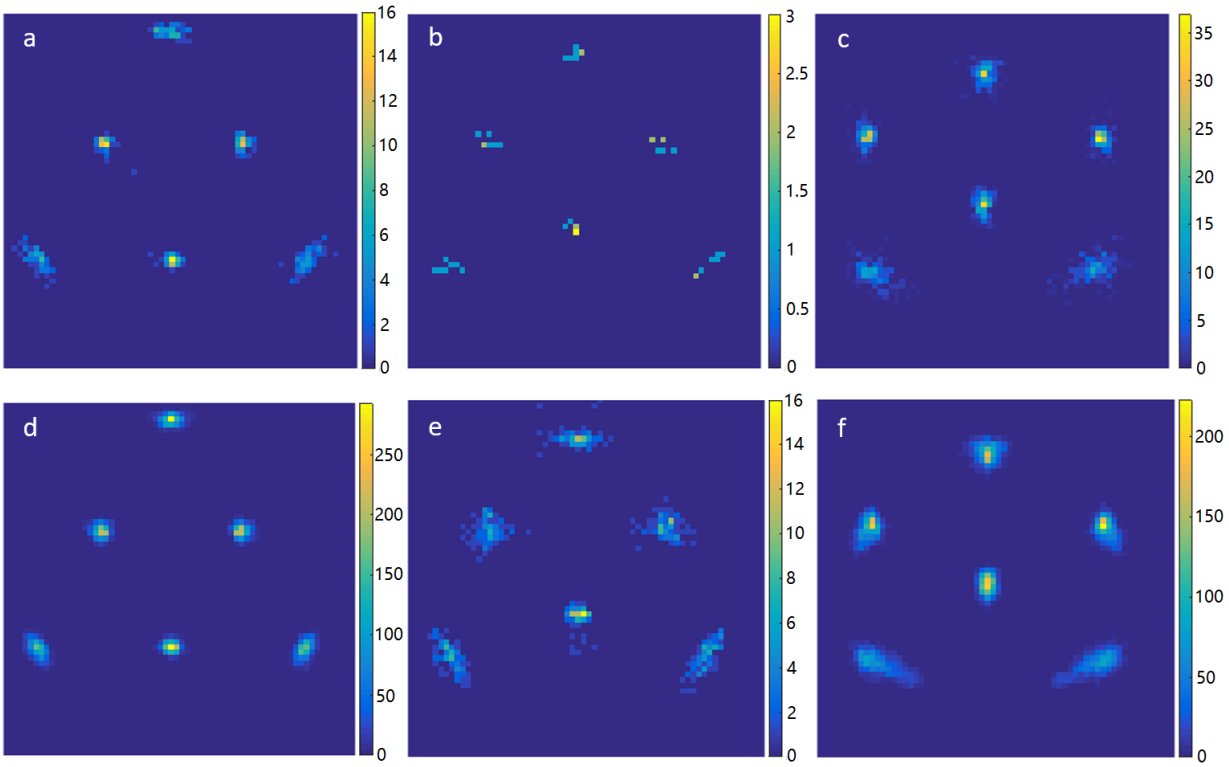

Figure 3.7: Conditional probability distributions of constituent nanoparticles of 6-particle optical matter clusters conditioned on the distance between the two particles on the (vertical) symmetry axis (cf. Fig. 3.2). The color scheme describes the number of configurations that contains a particle centered at a specific pixel. (a-c) Experimental distributions collected over 1686 configurations at an optical power of 50 mW. (d-f) Simulation result collected 21186 configurations at an optical power of 70 mW.

Fig. 3.5 of the main text presents a detailed HLDA analysis of a single triangle-to-chevron transition of a 6-particle optical matter cluster as well as statistical analysis of many transitions observed in simulations. Of course, each transition behaves differently, so Fig. 3.9 in Section 3.A shows some of the variations for representative transitions. In general, PC mode 3 is dominant for almost all transitions, but different modes can contribute significantly at

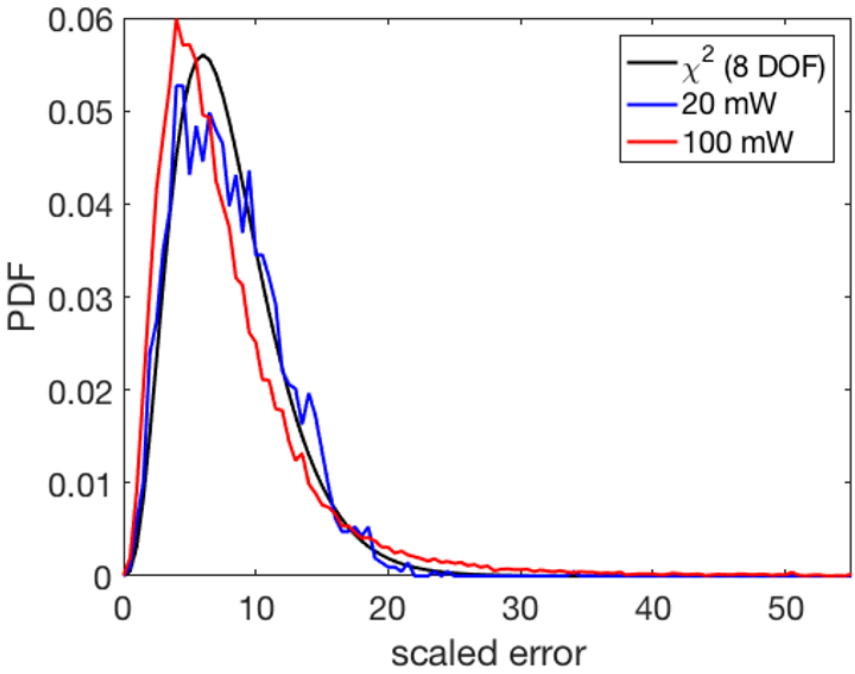

Figure 3.8: Probability density function (PDF) representation of lattice fitting displacement distributions of the 6-particle optical matter system corresponding to the cumulative density functions shown in Figure 3. The lattice fitting displacement is the sum of squares of particle position deviations of a configuration from the stable triangle lattice sites minimized over all possible translations, rotations, and the lattice constant (i.e., lattice site spacing). The fitting displacement computed from electrodynamics-Langevin dynamics (EDLD) simulation trajectories of fluctuations in the vicinity of the triangle configuration deviates from the 8 degree of freedom χ^2 distribution. The magnitude of the deviation of the PDF fitting displacement distribution from the 8 DOF χ^2 distribution (solid black curve) increases with optical trapping power in simulations conducted over the range 20-100 mW, indicating that the collective motions become more significant (and increase in magnitude) at higher optical powers (i.e. for stronger simulated optical trapping fields).

various times during the structural transition.

3.A.1 Supplementary Discussion

Derivation of the relationship between conventional normal modes and statistical collective modes in conservative systems.

Let us consider a system with N identical particles in $\mathbb{R}^d$. A d-dimensional position vector, $\mathbf{r}_k = [r_{k1}, r_{k2}, ..., r_{kd}]$ is assigned to the k^{th} particle. The configuration of the whole system

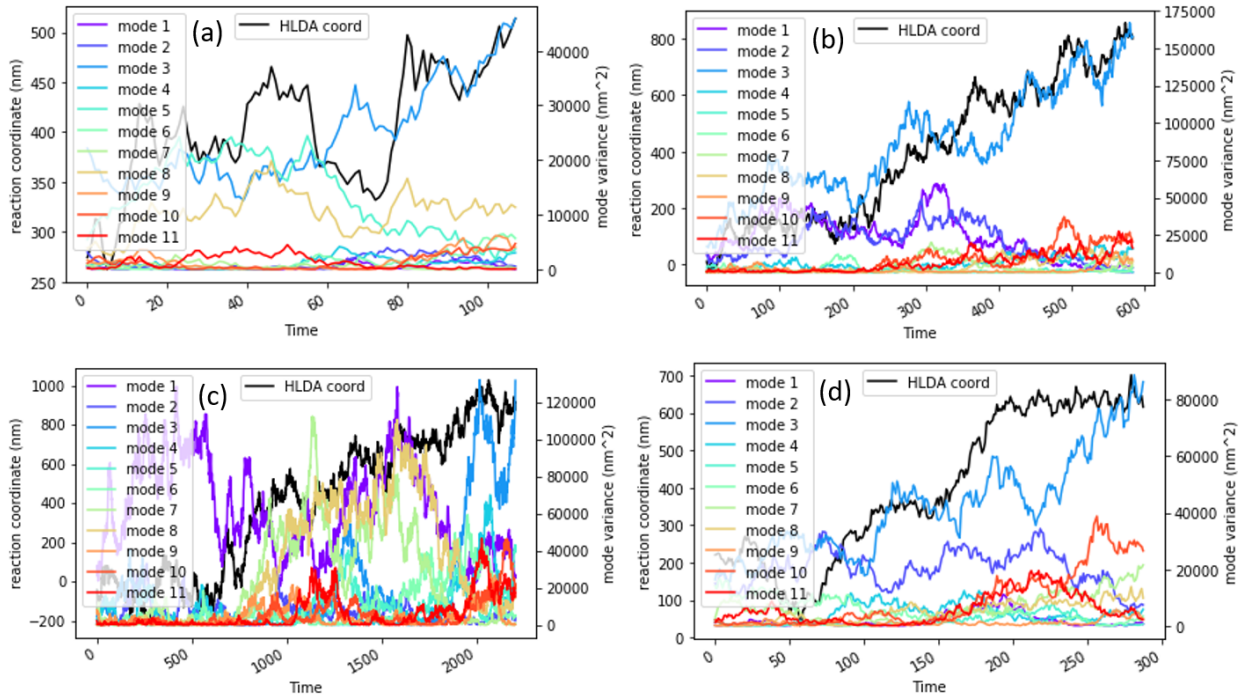

Figure 3.9: Comparison of the HLDA reaction coordinate s_{HLDA} and the variance of particle deviations along each of the 12 collective modes over the course of four different triangle-to-chevron isomerization transitions (cf. Fig. 3.6a). Some transitions occur quickly like Fig. 3.9a, but some are more prolonged like Fig. 3.9c. However, at the end of the transition, mode 3 always dominates.

is described by the Nd-dimensional configuration vector:

$$\mathbf{R} = (\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N) = [r_{11}, r_{12}, ..., r_{1d}, r_{21}, r_{22}, ..., r_{2d}, .., r_{N1}, r_{N2}, ..., r_{Nd}] \in \mathbb{R}^{Nd} \quad (3.11)$$

Next, consider a force field acting on this $\mathbb{R}^{Nd}$ configuration space. The force acting on a certain particle not only depends on the position of itself, but also on the positions of other particles. The function $\mathbf{F}$ takes in the position vector of the N particles and produces the forces acting on these N particles. Since the force acting on a particle is a vector in $\mathbb{R}^d$, $\mathbf{F}$ maps the configuration space into the N-particle force space that is also Nd-dimensional.

$$\mathbf{F} : \mathbb{R}^{Nd} \longrightarrow \mathbb{R}^{Nd} \quad (3.12)$$

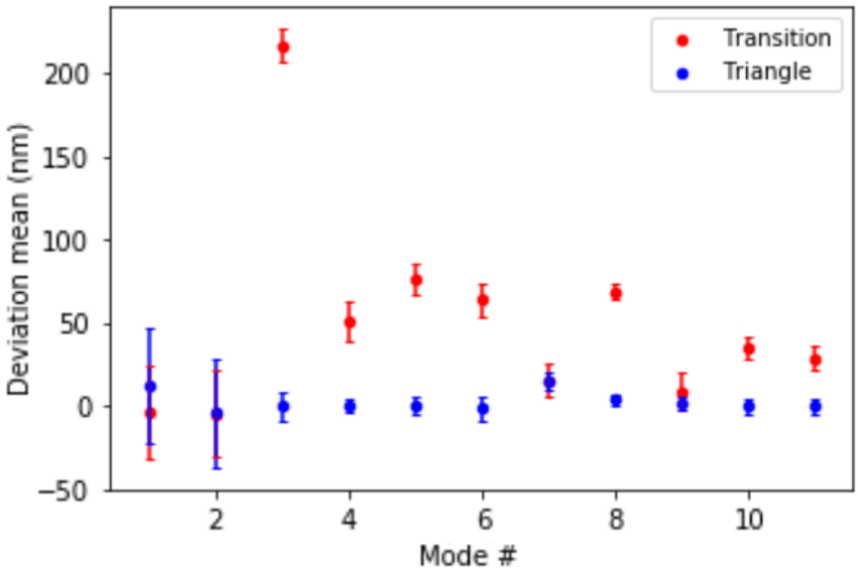

Figure 3.10: The error bars plotted for the mean of the deviations projected on to the 11 modes for triangle and transition configurations plotted in Fig. 3.6c. The error bars correspond to the standard deviations of the distributions shown in Fig. 3.6c

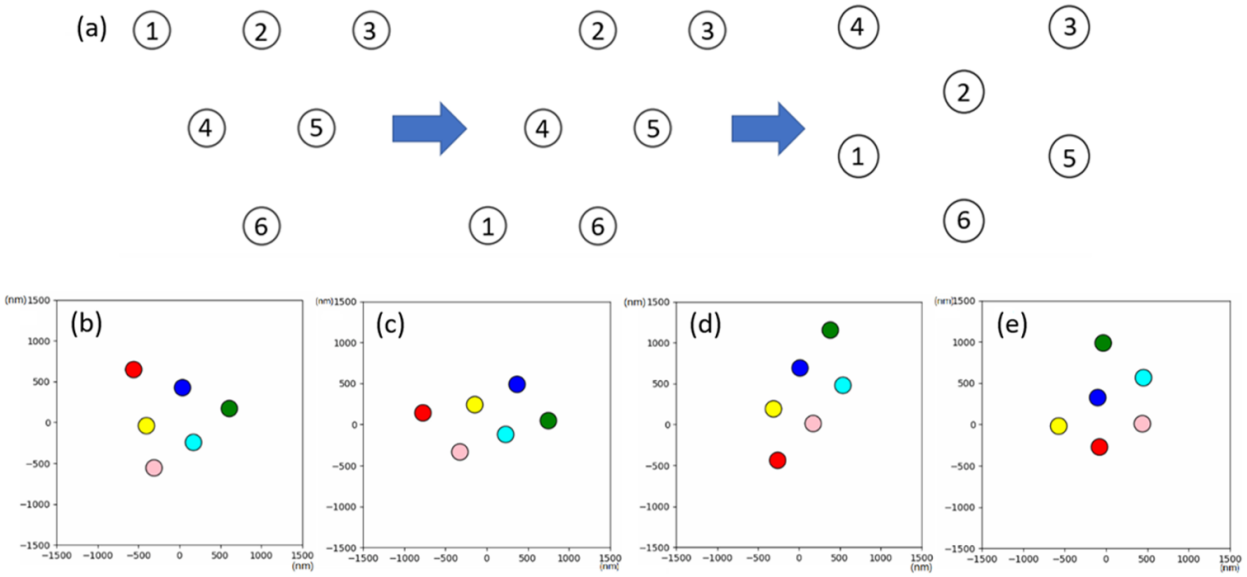

Figure 3.11: An indirect triangle-chevron transition. **a** The transition scheme with labeled particles. **b-c** Initial and final states of first half of transition. **d-e** Initial and final states of second half of transition. In this case, the intermediate OM structure has a parallelogram shape, which is different from the triangle "reactant" and chevron "product".

The force derivative matrix **K** is defined by taking derivatives over all position coordinates

of the force function.

$$K_{ip,jq} = \frac{\partial F_{ip}}{\partial r_{jq}} \tag{3.13}$$

Since the system is conservative, the potential difference between two different configurations is well-defined as the integral of the force field over any path that connects them. We choose a reference configuration $\mathbf{R}^{(0)}$ and define its potential to be zero. Then for any other configuration $\mathbf{R}$, its potential is defined as:

$$V(\mathbf{R}) = \int_{\mathbf{R}}^{\mathbf{R}^{(0)}} \mathbf{F}(\hat{\mathbf{R}}) \cdot d\hat{\mathbf{R}} \tag{3.14}$$

From the equations above, it is obvious that we can find the force function $\mathbf{F}$ from the potential function by taking negative gradient.

$$\mathbf{F} = -\nabla V \tag{3.15}$$

Therefore, the force derivative matrix $\mathbf{K}$ is the negative of the Hessian matrix of V.

$$K_{ip,jq} = \frac{\partial F_{ip}}{\partial r_{jq}} = -\frac{\partial^2 V}{\partial r_{ip} \partial r_{jq}} \tag{3.16}$$

When a configuration is at a local minimum of the potential, its energy is at a local minimum. The system dynamics near the minimum is determined, to second order by

$$V(\mathbf{R}) = V(\mathbf{R}_0) + \nabla V(\mathbf{R}_0) \cdot (\mathbf{R} - \mathbf{R}_0) + \frac{1}{2}(\mathbf{R} - \mathbf{R}_0)^{\mathrm{T}} [\nabla \nabla^{\mathrm{T}} V(\mathbf{R}_0)](\mathbf{R} - \mathbf{R}_0) + o(\|\mathbf{R} - \mathbf{R}_0\|_2^2) \tag{3.17}$$

Since V is at its local minimum at $\mathbf{R}_0$, the gradient of V at $\mathbf{R}_0$ is zero. Therefore, the first order term vanishes and the second order term dominates. In addition, we can replace the

Hessian matrix of V by the negative of force derivative matrix $\mathbf{K}$.

$$V(\mathbf{R}) = V(\mathbf{R}_0) - \frac{1}{2}(\mathbf{R} - \mathbf{R}_0)^{\mathrm{T}}\mathbf{K}(\mathbf{R} - \mathbf{R}_0) + o(\|\mathbf{R} - \mathbf{R}_0\|_2^2) \tag{3.18}$$

At a non-zero temperature T, the system configuration $\mathbf{R}$ is a random variable that has a probability distribution. The probability density for a configuration $\mathbf{R}$ to appear at a non-zero temperature T follows the Boltzmann distribution:

$$p(\mathbf{R}) \propto e^{-\frac{V(\mathbf{R})}{k_B T}} \tag{3.19}$$

k_B is the Boltzmann constant. Then,

$$p(\mathbf{R}) \propto e^{-\frac{V(\mathbf{R}) - V(\mathbf{R}_0)}{k_B T}} = e^{\frac{(\mathbf{R}-\mathbf{R}_0)^{\mathrm{T}}\mathbf{K}(\mathbf{R}-\mathbf{R}_0)}{2k_B T} + o(\|\mathbf{R}-\mathbf{R}_0\|_2^2)} \tag{3.20}$$

Therefore, the probability distribution of configurations near the stable configuration $\mathbf{R}_0$ is close to a Gaussian distribution. Let the covariance matrix of this Gaussian distribution be Σ. Then according to the distribution function, we will have the following relationship.

$$\mathbf{K}(\mathbf{R}_0)\Sigma = -k_B T \mathbf{I}_{Nd \times Nd} \tag{3.21}$$

Therefore, any normal mode of the system, which is one of the eigenvectors of the force derivative matrix $\mathbf{K}$, is also an eigenvector of the covariance matrix Σ. This is because, for any $\mathbf{v}$ that satisfies:

$$\mathbf{K}\mathbf{v} = \lambda \mathbf{v} \tag{3.22}$$

we have:

$$\Sigma v = k_B T \mathbf{K}^{-1}\mathbf{v} = \frac{k_B T}{\lambda}\mathbf{v} \tag{3.23}$$

Note that the derivation above cannot be applied to non-conservative systems directly due

to reasons stated below. Since the potential is not well-defined in our system, the Hessian matrix should be replaced by taking the first derivative of the external force field, the J matrix, instead of taking the second derivative of the potential. The J-matrix approach is an important method that can be compared to the present work. The J-matrix method leads to discovery of oscillatory solutions [32], which indeed sounds promising in analyzing non-conservative systems, but is limited to underdamped systems. In the overdamped case relevant for optical matter systems, the Langevin equation is first order so that there is no oscillatory solution. In the overdamped OM system, the J-matrix method is still applicable, but since the absence of oscillatory solutions, the application is limited to the analysis of its eigenvectors and eigenvalues. Furthermore, since the J-matrix is asymmetric, its eigenvectors are not orthogonal to each other and its left eigenvectors are different from its right eigenvectors, which induces coupling between the collective modes represented by its eigenvectors. We note that the covariance matrix analyzed using PCA gives orthogonal eigenvectors that make an orthogonal basis much easier to manipulate.

3.A.2 Supplementary Methods

Derivation of the analytical expressions of optimal lattice fitting parameters given lattice assignment.

$\mathbf{1}$ is the vector of all ones. Given $\mathbf{p}, \mathbf{q} \in \mathbb{C}^N$ (configuration coordinates), solve the optimal parameter set $\left\{a^*, \theta^*, p_0^*\right\}$ for:

$$Err^* = \left[\min_{a\in\mathbb{R}^+, p_0\in\mathbb{C}, \theta\in[0,2\pi)} \left| e^{i\theta}(\mathbf{p} + p_0\mathbf{1}) - a\mathbf{q} \right|^2 \right]^{\frac{1}{2}} \tag{3.24}$$

We first optimize the translation, fixing $\theta = 0$ and $a = 1$.

$$\begin{aligned}
Err^2\Big|_{\theta=0,a=1} &= |(\mathbf{p} + p_0\mathbf{1}) - \mathbf{q}|^2 = [(\mathbf{p} - \mathbf{q}) + p_0\mathbf{1}]^{\mathrm{H}}[(\mathbf{p} - \mathbf{q}) + p_0\mathbf{1}] \\
&= (\mathbf{p} - \mathbf{q})^{\mathrm{H}}(\mathbf{p} - \mathbf{q}) + (p_0\mathbf{1})^{\mathrm{H}}(p_0\mathbf{1}) + (p_0\mathbf{1})^{\mathrm{H}}(\mathbf{p} - \mathbf{q}) + (\mathbf{p} - \mathbf{q})^{\mathrm{H}}(p_0\mathbf{1}) \\
&= (\mathbf{p} - \mathbf{q})^{\mathrm{H}}(\mathbf{p} - \mathbf{q}) + Np_0\overline{p_0} + \overline{p_0}[\mathbf{1}^{\mathrm{H}}(\mathbf{p} - \mathbf{q})] + p_0\overline{[\mathbf{1}^{\mathrm{H}}(\mathbf{p} - \mathbf{q})]} \\
&= |\mathbf{p} - \mathbf{q}|^2 - \frac{1}{N}\left|(\mathbf{p} - \mathbf{q})^{\mathrm{H}}\mathbf{1}\right|^2 \\
&+ N\left\{p_0\overline{p_0} + \overline{p_0}\left[\frac{\mathbf{1}^{\mathrm{H}}}{N}(\mathbf{p} - \mathbf{q})\right] + p_0\overline{\left[\frac{\mathbf{1}^{\mathrm{H}}}{N}(\mathbf{p} - \mathbf{q})\right]} + \left[\frac{\mathbf{1}^{\mathrm{H}}}{N}(\mathbf{p} - \mathbf{q})\right]\overline{\left[\frac{\mathbf{1}^{\mathrm{H}}}{N}(\mathbf{p} - \mathbf{q})\right]}\right\} \\
&= |\mathbf{p} - \mathbf{q}|^2 - \frac{1}{N}\left|(\mathbf{p} - \mathbf{q})^{\mathrm{H}}\mathbf{1}\right|^2 + N\left|p_0 + \frac{\mathbf{1}^{\mathrm{H}}}{N}(\mathbf{p} - \mathbf{q})\right|^2
\end{aligned} \tag{3.25}$$

In order to minimize the expression above, one sets the last term of the expression above to zero, which means:

$$p_0^* = \frac{\mathbf{1}^{\mathrm{H}}}{N}(\mathbf{q} - \mathbf{p}) \tag{3.26}$$

It should be noticed that the expression of p_0^* is actually the difference between the centers of mass of the two configurations, $\mathbf{p}$ and $\mathbf{q}$. In addition, after $\mathbf{p}$ and $\mathbf{q}$ are translated to let both centers of mass overlap with the origin, p_0^* will stay unchanged no matter how p and q are rotated or linearly scaled. Therefore, after the choice of p_0^*, further optimization of θ and a does not affect p_0^*, so that their optimization can be carried out successively and independently.

Next, we optimize the rotation and lattice constant. Let $p' = p + p_0^*\mathbf{1}$, we have:

$$Err^* = \left[\min_{a\in\mathbb{R}^+,\theta\in[0,2\pi)}\left|e^{i\theta}\mathbf{p}' - a\mathbf{q}\right|^2\right]^{\frac{1}{2}} \tag{3.27}$$

$$
\begin{aligned}
Err^2\Big|_{p_0=p_0^*} &= \left|e^{i\theta}\mathbf{p}' - a\mathbf{q}\right|^2 = \left|\mathbf{p}' - ae^{-i\theta}\mathbf{q}\right|^2 = \left|\mathbf{p}'\right|^2 + a^2\left|\mathbf{q}\right|^2 - 2a\mathrm{Re}(\mathrm{e}^{\mathrm{i}\theta}\mathbf{q}^{\mathrm{H}}\mathbf{p}') \\
&= a^2\left|\mathbf{q}\right|^2 - 2(a\left|\mathbf{q}\right|)\left[\frac{\mathrm{Re}(\mathrm{e}^{\mathrm{i}\theta}\mathbf{q}^{\mathrm{H}}\mathbf{p}')}{|\mathbf{q}|}\right] + \left[\frac{\mathrm{Re}(\mathrm{e}^{\mathrm{i}\theta}\mathbf{q}^{\mathrm{H}}\mathbf{p}')}{|\mathbf{q}|}\right]^2 - \left[\frac{\mathrm{Re}(\mathrm{e}^{\mathrm{i}\theta}\mathbf{q}^{\mathrm{H}}\mathbf{p}')}{|\mathbf{q}|}\right]^2 + \left|\mathbf{p}'\right|^2 \\
&= |\mathbf{q}|^2\left[a - \frac{\mathrm{Re}(\mathrm{e}^{\mathrm{i}\theta}\mathbf{q}^{\mathrm{H}}\mathbf{p}')}{|\mathbf{q}|^2}\right]^2 + \left\{\left|\mathbf{p}'\right|^2 - \left[\frac{\mathrm{Re}(\mathrm{e}^{\mathrm{i}\theta}\mathbf{q}^{\mathrm{H}}\mathbf{p}')}{|\mathbf{q}|}\right]^2\right\}
\end{aligned}
\tag{3.28}
$$

In order to minimize Err, the first term in the equation above should be set to zero, and the second term should be as small as possible, meaning that the real part of $e^{i\theta}\mathbf{q}^{\mathrm{H}}\mathbf{p}'$ is maximized to be $\left|\mathbf{q}^{\mathrm{H}}\mathbf{p}'\right|$.

Therefore:

$$
e^{i\theta^*} = e^{-i\mathrm{arg}(\mathbf{q}^{\mathrm{H}}\mathbf{p}')} = \frac{(\mathbf{p}')^{\mathrm{H}}\mathbf{q}}{\left|\mathbf{q}^{\mathrm{H}}\mathbf{p}'\right|} \tag{3.29}
$$

$$
a^* = \frac{\mathrm{Re}(\mathrm{e}^{\mathrm{i}\theta}\mathbf{q}^{\mathrm{H}}\mathbf{p}')}{|\mathbf{q}|^2} = \frac{\left|\mathbf{q}^{\mathrm{H}}\mathbf{p}'\right|}{|\mathbf{q}|^2} \tag{3.30}
$$

This is the optimal rotation angle and the optimal lattice constant.

CHAPTER 4

UNDERSTANDING AND DESIGN OF NON-CONSERVATIVE OPTICAL MATTER SYSTEMS USING MARKOV STATE MODELS

This chapter is based on the following publication:

Shiqi Chen, John A. Parker, Curtis W. Peterson, Stuart A. Rice, Norbert F. Scherer, Andrew L. Ferguson. "Understanding and design of non-conservative optical matter systems using Markov state models" *Molecular Systems Design & Engineering*, 2022, **7** (10), 1228-1238.

4.1 Abstract

Optical matter (OM) systems consist of nano-particle constituents in solution that, when illuminated with a laser beam, can self-organize into ordered arrays bound by electrodynamic interactions. OM systems are intrinsically non-equilibrium due to the incident electromagnetic flux and may manifest non-conservative forces and interconversion among structural isomers. Rational design of desired configurations and transitions requires quantitative understanding of the relation between the incident beam and the emergent metastable states and isomerization dynamics. We report a data-driven approach to build Markov state models appropriate to non-conservative and permutation-invariant systems. We demonstrate the approach in electrodynamics-Langevin dynamics simulations of six electrodynamically-bound nanoparticles. The Markov state models quantify the relative stability of competing metastable states and the transition rates between them as a function of incident beam power. This informs the design and testing of new beam conditions to stabilize desired nanoparticle configurations. The technique can be generalized to understand and control non-conservative and permutation-invariant systems prevalent in optical and active matter.

4.2 Introduction

The self-organization of nanoparticles (e.g., gold, silver, silicon, etc.) to fabricate metamaterials is a promising approach to create new functional materials [88–98]. Doing so requires knowledge of and control over the interactions between constituent elements. Liquid crystals, a well-studied example where anisotropic molecules can be manipulated to form ordered phases that can be controlled by temperature and/or external fields, have become a pillar of the information technology revolution by way of their integration into displays and other devices [99]. Given the tremendous impact of these materials for specific applications, it is highly desirable to create new classes of self-organizing materials with engineered properties.

This goal requires using and expanding the principles of physical chemistry and condensed matter physics. If we envision material formation as a kinetic assembly process, then the averaged microscopic dynamics underlying macroscopic rate laws can be inferred from ensemble (e.g., spectroscopic or far-field flux) measurements [100–104]. However, such measurements cannot provide information about individual particle motions and the forces leading to them.

Optical matter (OM) is a class of materials formed by self-organization of its particle constituents into ordered structures [6, 105, 106]. OM structures form in focused optical beams (i.e., optical traps or tweezers), [9, 107] without explicit external control of particle positions. Once the nanoparticles (e.g., Au, Ag, Si, etc.) and optical beam properties are chosen, the many body electrodynamic interparticle interactions and forces established in the system generate structures. The electrodynamic interactions, termed optical binding, range from a few to many k_BT units of thermal energy, so the OM structures can undergo structural rearrangements. Fig. 4.1 shows the transformation of the 6-particle OM system among its three most stable structures resolved by dark-field optical microscopy microscopy.

OM systems are fundamentally non-equilibrium due to a continuous flux of optical beam power through the material. The electrodynamic interactions amongst the nanoparticles

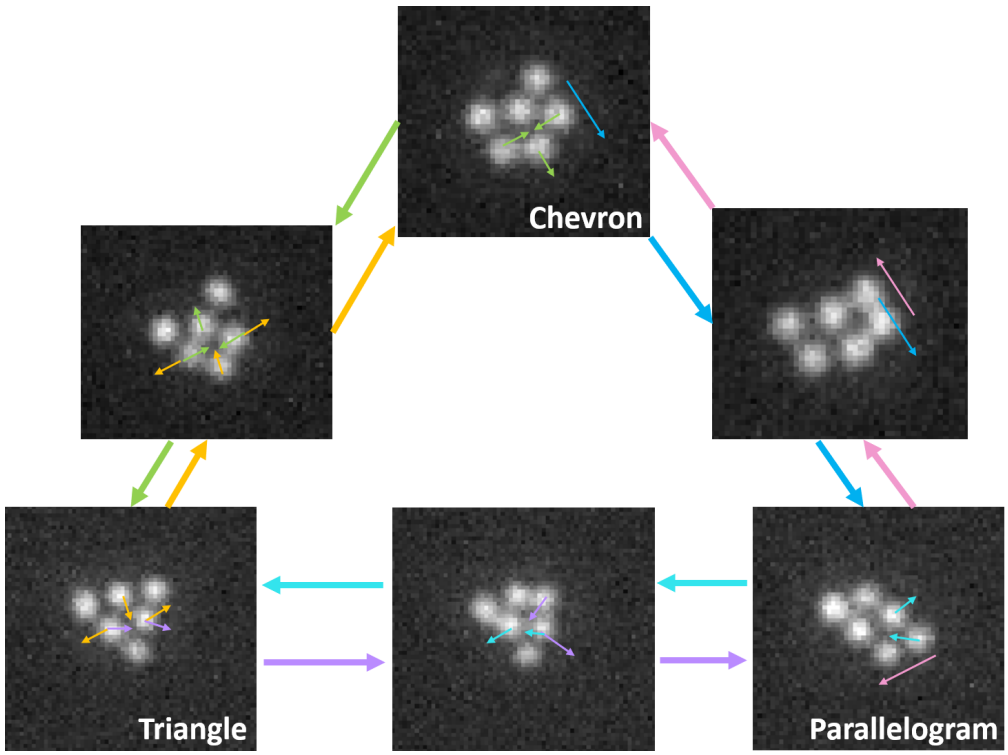

Figure 4.1: Dark-field microscopy image of optical matter structures of 150 nm diameter silver particles. Six silver nanoparticles were drawn into a focused Gaussian beam (i.e., optical trap). Most nearest neighbor interparticle spacings are around 600 nm while some are closer due to near-field interactions. The chevron, triangle, and parallelogram are three commonly observed stable structures, while the other three structures correspond to transition states along the structural transformation coordinates. The colored arrows in the dark-field images sketch the structure change of the transition depicted by the arrows in the same color between the dark-field images.

are complex but controllable. Optical beam shape, the polarization of the light, the spatial phase profile of the optical field, [41, 108–111] as well as the nature (e.g., elemental makeup, dielectric, etc.) and shape (e.g., spheres, ellipsoids, rods, wires, cubes, platelets, etc.) [11, 112–115] of the constituent particles can all be selected or tuned, allowing one to explore a large space of pairwise and many-body interactions. This large parameter space makes accessible a wide range of phenomena (e.g., dynamics of self-assembling and driven active matter, negative torque, spectroscopy of collective excitations, etc.) [16, 77, 115, 116]. Recent work has shown that the dynamic behavior of OM arrays is related to their shape and symmetry [12, 14, 113, 117–122]. Light scattering from nanoparticle arrays can bring about

unusual phenomena such as non-reciprocal forces [14, 121], negative torque, [15, 84, 123] and non-conservative forces [25].

While understanding of how incident fields can affect the dynamics and structure of OM arrays is improving [77, 78, 124], and efficient computer simulation techniques exist to model OM dynamics [1, 17, 34, 37, 38], we still lack quantitative theoretical models of how to modulate the properties of the particles and incident light to stabilize particular desired OM states and transitions. This presents an opportunity for the development of data-driven models to learn empirical mappings from the system properties to the emergent dynamical behaviors and inform the design of steady-state and dynamic control strategies to stabilize various OM (non-ground state) structures or to drive OM isomerization.

The primary objective of the present work is to devise an analytical scheme with which we understand and control the dynamics and metastable states in numerical simulations of 6-particle OM systems. In order to achieve this, we build Markov state models (MSMs) as a powerful approach to infer long-time kinetic models from simulation trajectories [42,125–128]. In order to build a MSM, we have to carry out featurization and clustering analysis on the simulation trajectory and convert the configuration trajectories to trajectories of cluster labels that are the direct input of MSM construction. Due to the permutation symmetry resulting from the fungible nature of identical particles, the featurization procedure must take permutation invariance into account. Our analysis uses a permutation invariant pairwise metric that we supply as the kernel to perform nonlinear dimensionality reduction using diffusion maps [49]. We use diffusion k-means [1] to define the microstate clustering, and Robust Perron Cluster Cluster Analysis (PCCA+) [59–61, 129, 130] to define the macrostate clustering and build MSMs. We test the Markovianity of the macrostate MSMs using the Chapman-Kolmogorov (CK) test to verify that they are valid kinetic models of the non-equilibrium OM system and provide *post hoc* support for the use of permutationally-invariant diffusion map embeddings to identify and resolve microstates. By constructing MSMs at a

variety of beam powers we quantify how the intensity of the incident light controls the relative stabilities of and transition rates between the metastable OM configurations. We then use these models to guide the design and testing of new beam conditions to preferentially stabilize desired OM states.

4.3 Methods

4.3.1 Langevin dynamics simulations of optical matter

The dynamical evolution of the OM system can be modeled by combining a finite-difference time-domain (FDTD) solution of the electrodynamic forces with a Langevin dynamics equation of motion for the particle positions [34]. However, this is insufficiently efficient to access experimental timescales. Therefore, we developed an electrodynamics-Langevin dynamics (EDLD) approach based on generalized multiparticle Mie theory (GMMT) [37, 38]. The resulting EDLD solver performs a numerical Verlet integration of the following Langevin equation,

$$m\frac{d^2\boldsymbol{r}}{dt^2} = \boldsymbol{F}(\boldsymbol{r},t) - \lambda_v\frac{\boldsymbol{r}}{dt} + \boldsymbol{\eta}, \tag{4.1}$$

where m is the nanoparticle mass, $\boldsymbol{r}$ is its location, t is time, $\lambda_v = 6\pi\nu R$ is the friction coefficient specified by Stokes' Law where R is the nanoparticle radius and ν is the viscosity of the surrounding fluid, $\boldsymbol{\eta}$ is a stationary Gaussian random variable with zero mean and a standard deviation that satisfies the fluctuation-dissipation relation at the specified temperature, and $\boldsymbol{F}$ is the net force experienced by the particle comprising electrodynamic contributions computed from GMMT and electrostatic and interparticle interactions. The force field $\boldsymbol{F}$ computed is non-conservative due to the input power from the optical beam. The steady states reached by the system correspond to states in which the power input from the optical beam is balanced by frictional dissipation into the medium. Simulations

are conducted using the MiePy software developed in the Scherer group [17]. Calculations on 28 × 2.4 GHz Intel E5-2680 v4 CPUs execute 2.5 seconds of simulation time per hour of wall clock time with a 5 μs simulation time step.

The simulation is carried out for six spherical silver nano-particles with 150 nm diameter under several beam powers ranging from 40 mW to 90 mW. For each beam power, we simulate 100 trajectories, each of which is 100 seconds in length. Data is collected every 200 simulation time steps, so the frame-wise time step is 1 ms. We assume that the solvent medium is water (refraction index $n_b = 1.33$, viscosity $\eta = 8 \times 10^{-4}$ Pa·s), temperature $T = 300$ K, the beam width $w = 2.5$ μm, the wavelength $\lambda = 800$ nm. Based on our experience with previous EDLD simulations using the MiePy package, [17] we add a defocus equal to the Rayleigh range, $z = 0.5kw^2$ where $k = 2\pi n_b/\lambda$, and an electrostatic double layer potential with particle surface potential 77 mV and Debye screening length 27.6 nm according to previous theoretical work [41].

4.3.2 Nonlinear manifold learning

Before building the MSM and computing the transition rates between the structural states, we need to perform featurization and clustering analysis on the data set of configuration trajectories. The first step in the MSM construction pipeline is to project the simulation trajectories into their leading slow modes to define a low-dimensional embedding conducive to identification of the metastable states of the system using clustering algorithms [42]. This is typically achieved using time lagged independent components (tICA) analysis [43] or its kernel [44] or deep [45–47] variants. However, there are technical challenges in applying these methods to systems exhibiting full permutation symmetry, such as the OM system, where all particles are identical. Therefore, in the featurization procedure, we instead use diffusion maps [48–51], a nonlinear manifold learning method that can generate permutation invariant coordinates for clustering. At first glance, this is possibly problematic since diffusion maps

are designed to identify high-variance as opposed to slow collective modes and thus may not provide optimal embeddings for clustering into metastable states. However, the diffusion map is dynamically meaningful and the eigenfunctions are known to be identical to those of the Langevin equation for conservative systems equipped with the common Euclidean distance without permutation invariance [55]. There is no known proof that this eigenvector correspondence continues to hold for non-conservative systems with permutation symmetry, but we conjecture that the leading diffusion map eigenvectors may nevertheless provide approximations for the slow modes of the Langevin equation governing the OM dynamical evolution and may therefore offer a useful embedding for clustering the metastable states. We provide *post hoc* support for this conjecture by validating that the MSMs generated using diffusion map embeddings are valid kinetic models that pass all the numerical tests of Markov properties and convergence of implied time scales.

4.3.2.1 Pairwise distance calculation

Let $(x_p^{(i)}, y_p^{(i)})$ denote the 2D Cartesian coordinates of particle p in configuration i. We can calculate the distance matrix $\boldsymbol{M}^{(i)}$ for each configuration i with matrix elements,

$$M_{pq}^{(i)} = \sqrt{\left(x_p^{(i)} - x_q^{(i)}\right)^2 + \left(y_p^{(i)} - y_q^{(i)}\right)^2}. \tag{4.2}$$

Let $\boldsymbol{e}_k$ denote the unit column vector with the k^{th} component unity and others zero. Then the permutation-invariant distance defined between a pair of configurations i and j is,

$$d_{ij} = \min_{\boldsymbol{P} \in S_n} \sqrt{\sum_{k=1}^{N} \min_{\boldsymbol{Q}_k \in S_n} \left\| \boldsymbol{Q}_k \boldsymbol{M}^{(i)} \boldsymbol{e}_k - \boldsymbol{M}^{(j)} \boldsymbol{P} \boldsymbol{e}_k \right\|_2^2}, \tag{4.3}$$

where N is the number of particles, and S_n is the set of all permutation matrices so that $\boldsymbol{P}$ and $\boldsymbol{Q}_k$ are the optimal permutation matrices that minimize d_{ij}. Here, $\boldsymbol{P}$ and $\boldsymbol{Q}_k$ are

$(N+1)$ independent permutation matrices to be optimized, in which $\boldsymbol{P}$ corresponds to the inter-column permutation while $\boldsymbol{Q}_k$ corresponds to the intra-column permutations for all the columns so that the norm of the difference of $M^{(i)}$ and $M^{(j)}$ is optimized over all inter-column and intra-column permutations. Then d_{ij} is a permutation-invariant pairwise distance for the configurations that serves as a kernel for the diffusion map calculations.

4.3.2.2 Diffusion maps

Diffusion maps are a type of non-linear manifold learning method that take the input of pairwise distances of the configurations and generate a low-dimensional non-linear subspace of the configuration space [49,50]. A brief introduction to diffusion map methodology is provided below, while full details of this method applied to colloidal self-assembly are discussed in previous work [52–54].

First, the kernel matrix $\boldsymbol{K}$ is calculated with elements,

$$K_{ij} = \exp\left(-\frac{d_{ij}^2}{2\epsilon^2}\right), \tag{4.4}$$

where d_{ij} is the permutation-invariant pairwise distance defined previously and ϵ is the kernel bandwidth parameter that characterizes the adjacency among the configurations. Next, $\boldsymbol{K}$ is normalized to $\tilde{\boldsymbol{K}}$ to gain correspondence to the Langevin dynamics [55,56],

$$\tilde{K}_{ij} = \frac{K_{ij}}{\sqrt{\left(\sum_k K_{ik}\right)\left(\sum_k K_{kj}\right)}}. \tag{4.5}$$

$\tilde{\boldsymbol{K}}$ is then used to calculate the right-stochastic Markov transition matrix (RSMTM) $\boldsymbol{T}$,

$$T_{ij} = \frac{\tilde{K}_{ij}}{\sum_{j'} \tilde{K}_{ij'}}, \tag{4.6}$$

with eigenvalues $\{\lambda_k\}$ and right eigenvectors $\{\psi_k\}$. Since the components of ψ_1 are all unity,

$\{\psi_k\}_{k=2}^{m+1}$ is taken as the basis of the low-dimensional nonlinear configuration subspace. An appropriate value of m is identified based on a gap in the eigenvalue spectrum. Finally, we obtain,

$$\left\{\left(x_p^{(i)}, y_p^{(i)}\right)\right\}_{p=1}^{N} \longrightarrow \{\psi_k(i)\}_{k=2}^{m+1}, \tag{4.7}$$

which maps the Euclidean coordinates of each configuration to its corresponding diffusion map embedding. After obtaining this m-dimensional permutation-invariant reduction, configurations are clustered into microstates.

4.3.2.3 Nyström extension

The time and memory complexity of diffusion maps scale quadratically [131] with the number of data points n. The Nyström extension is an out-of-sample extension technique that scales linearly with n and can be used to embed a new point to a pre-existing diffusion map embedding [132–135]. In this case, we can choose n' ($n' < n$) representative data points (termed "pivots") from the trajectory, calculate the diffusion map on these n' points, and then use Nyström extension to embed the remaining $(n - n')$ points. The pivot points must cover the configuration space of the entire data set so that all points to be embedded are within ϵ (the kernel bandwidth) of at least one pivot point to assure that the new points are accurately interpolated [136, 137]. This approach is known as pivot diffusion maps [136]. Given a new point and the n'-point diffusion map previously constructed, pivot diffusion maps compute the distance between the new point and the n' existing points $\{d_{0,j}\}_{j=1}^{n'}$ where subscript 0 denotes the new point. Next, we compute and append a new row to the kernel matrix $\boldsymbol{K}$ corresponding to the new point, apply the Langevin normalization, and then calculate the corresponding RSMTM row vector,

$$K_{0j} = \exp\left(-\frac{d_{0j}^2}{2\epsilon^2}\right), \quad \tilde{K}_{0j} = \frac{K_{0j}}{\sqrt{\left(\sum_k K_{0k}\right)\left(\sum_k K_{kj}\right)}}, \quad T_{0j} = \frac{\tilde{K}_{0j}}{\sum_{j'} \tilde{K}_{0j'}}, \tag{4.8}$$

and the embedding of the new point is given by,

$$\psi_k(0) = \frac{1}{\lambda_k} \sum_{j=1}^{n'} T_{0j} \psi_k(j), \quad k = 2, 3, ..., m+1. \tag{4.9}$$

The representative set of the n' pivot points is generated to assure good coverage of the configurational phase space. First, we perform EDLD simulations for 70 mW beam power using a periodic temperature profile with a period of 7000 simulation time steps: 2000 steps at 300 K and 5000 steps at 100 K. The pairwise distances among the particles of the last configuration of every 1000-step segment are computed. The total number of degrees of freedom is $(2N-3)$, where we have N particles moving in the plane subject to two translational constraints and one rotational constraint. We impose a condition in order not to get too far away from the relevant portion of configurational space: if the number of the pairwise distances less than 1.5 optical wavelengths in the medium (i.e., 900 nm) is less than $(2N-4)$, the simulation is restarted. This is because the number of pairwise distances that are less than 1.5 optical wavelengths can be regarded as the number of particles pairs that are at the first optical binding sites of each other. Since the total number of degrees of freedom is $(2N-3)$, at most $(2N-3)$ first optical bindings can be formed. Just as formation of bonds lowers the potential energy of molecular systems, the more first optical bindings the more stable the OM structure is. If the number of first optical bindings is less than $(2N-4)$, the structure is not stable. We apply 13 temperature cycling periods in a single simulation and 100 single simulations are carried out in parallel. Next, we iterate through all the configurations in the trajectories and add them one by one to a pivot set in which all pairwise distances among the configurations are larger than 470 nm. Then, we repeat the simulation and addition of points to the pivot set 15 times before the number of points in the pivot set converges to include 545 configurations. We enrich these pivots with 19,500 configurations randomly selected from the simulation trajectories to form the terminal pivot

set. We have verified that the pivot set constructed according to this procedure provides complete coverage of this space such that all remaining data points lie within ϵ of at least one pivot point.

4.3.2.4 Density-adaptive diffusion maps

Diffusion maps may not simultaneously resolve the region of configuration space with high density of points and the sparse connectivity region with low density of points. The density adaptive variant of diffusion maps was developed to address this challenge [138]. Instead of using the distance directly in the Gaussian kernel in eq. 4.4, this method parameterizes the kernel matrix elements as,

$$K_{ij}(\alpha) = \exp\left(-\frac{d_{ij}^{2\alpha}}{2\epsilon^2}\right) = \exp\left(-\frac{d_{ij}^2}{2\left[\epsilon d_{ij}^{(1-\alpha)}\right]^2}\right). \tag{4.10}$$

In effect, the kernel bandwidth ϵ is scaled according to the pairwise distance by a factor $d_{ij}^{1-\alpha}$. When $\alpha = 1$, we recover the original diffusion map with a constant kernel bandwidth. When $\alpha = 0$, the kernel bandwidth is proportional to the pairwise distance between any pair of points and K_{ij} becomes a constant value for any pair of configurations (i, j). Clearly, α should be chosen from $(0, 1]$ to make the diffusion map adapt to the density of the configuration space. In this work, we choose $\alpha = 0.1$ and $\epsilon = 2$ nm$^{0.1}$, which is motivated by the guidelines based on the work of Wang and co-workers [138]. We provide post hoc validation that this choice of α and ϵ generates diffusion map embedding and clustering leading to MSMs that pass all of our numerical validations. The embedding plot of the density-adaptive diffusion map for the 50 mW beam power simulation data is shown in Fig. 4.7 in the Section 4.A. As discussed in the section of results and discussion, the diffusion map embedding provides good discrimination between the metastable macrostates.

4.3.3 Diffusion k-means clustering into microstates

The k-means clustering algorithm is a widely-used unsupervised clustering method [57]. Chen and Yang introduced diffusion k-means, which maximizes the within-cluster connectedness based on the diffusion distance [58]. The diffusion distance is defined as the Euclidean distance in the eigenvector space of diffusion map embedding [48–50]. In other words, diffusion k-means is k-means clustering applied to diffusion map embeddings. In the present work, diffusion k-means is used as the microstate clustering algorithm so the clusters generated by diffusion k-means are termed "microstates" while the clusters generated by Robust Perron Cluster Cluster Analysis (PCCA+) [59–61] are called "macrostates". The eigenvectors of a diffusion map correspond to different eigenvalues that characterize the time scale of transition between macrostates. Therefore, instead of executing k-means clustering directly on the basis set $\{\psi_k\}_{k=2}^{m+1}$, we execute it on the basis set $\{\lambda_k^{\tilde{t}}\psi_k\}_{k=2}^{m+1}$, where λ_k is the eigenvalue corresponding to ψ_k and $\tilde{t}$ is a parameter that characterizes the time scale of diffusion distances encountered in the k-means clustering [50]. As $\tilde{t}$ becomes larger, the eigenvectors with large eigenvalues become more important in the clustering, leading to merging of regions discriminated by higher order eigenvectors and discriminating microstates largely on the basis of the leading eigenvectors. (On the other hand, when $\tilde{t}$ is small or even negative, the eigenvectors with small eigenvalues become important, leading to microstates as well as macrostates connected by fast transitions and merging the regions connected by slow modes.) Here, we employ an empirical procedure to select $\tilde{t}$ such that our k-means clustering identifies the long-lived macrostates and that the results are not sensitive to the precise value of this parameter.

4.3.4 Clustering into macrostates and Markov state model

Each MSM is built upon a data set containing 100 trajectories that are 100 seconds long and contain 100,000 configurations for a specific optical beam power. The MSM is built

using the PyEMMA software (http://www.emma-project.org/latest/) [139]. We use diffusion k-means with $k = 1000$ and maximum iteration number of 200 in the microstate clustering. The microstates are clustered into macrostates using the Robust Perron Cluster Cluster Analysis (PCCA+) algorithm [59–61] that determines the stable states of the OM system. Mathematically, we construct the elements of the microstate transition matrix $\mathbf{\Gamma}$ as $\Gamma_{ij} = c_{ij}(\tau)/\sum_k c_{ik}(\tau)$, where $c_{ij}(\tau)$ are the counts of transition events between microstates i and j at a lag time of τ [140]. Due to the non-conservative nature of OM systems, $\mathbf{\Gamma}$ is not guaranteed to obey detailed balance and therefore may not possess strictly real eigenvalues and eigenvectors as required by the PCCA+ clustering algorithm employed to cluster microstates into macrostates [140]. As such, we adopt the conventional pragmatic solution of symmetrizing the count matrices under the operation $c_{ij}(\tau) \leftarrow \frac{1}{2}\left(c_{ij}(\tau) + c_{ji}(\tau)\right)$ to enforce detailed balance within $\mathbf{\Gamma}$ [140]. Physically, this corresponds to collating counts over the forward and reverse trajectories, although we observe more sophisticated techniques based on likelihood maximization and Koopman reweighting also exist [140]. We perform this "reversibilization" to furnish real eigenvectors as required by PCCA+, but since the OM systems are non-equilibrium and therefore not constrained to obey detailed balance, there is a concern that the reversibilized eigenvectors may substantially deviate from, and therefore not be representative of, those of the original non-reversibilized system. We test this by computing the cosine similarity between the first 10 eigenvectors of the reversibilized and original microstate transition matrices for the 50 mW beam power and illustrate in Fig. 4.8 in Section 4.A that they are very similar. The reversibilization procedure is justified since a PCCA+ clustering into $(n_M + 1)$ macrostates uses only the leading n_M ($n_M = 6$ in the current case) right eigenvectors. Finally, we estimate our macrostate MSM by computing count matrices and (non-reversibilized) transition matrices from our simulation trajectory data over the macrostates. Importantly, since we do not enforce detailed balance within the macrostate MSM, the microstate reversibilization procedure may be viewed purely as a

means to aid in the definition of the macrostates and which has no bearing on the subsequent specification of the macrostate transition kinetics. We observe that the macrostate count matrix is naturally very close to a symmetric matrix whereas the microstate count matrix is substantially asymmetric. An analysis of the count matrix symmetry is discussed in Section 4.A below Fig. 4.8.

There are three key hyperparameters within this protocol that must be manually selected: the parameter $\tilde{t}$ within the diffusion k-means clustering; the number of macrostates n_M within the PCCA+ clustering; and the lag time τ of the MSM. We self-consistently specify these hyperparameters by analyzing the assignment matrices and implied time scales as described below.

4.4 Results and Discussion

We now proceed to construct MSMs for our 6-particle OM system as a function of beam power. These MSMs represent data-driven models that we can use to predictively link beam power to the identity and stability of emergent macrostates of the OM system and can be used to guide the design of beam powers to preferentially stabilize desired microstates and transitions. We illustrate the hyperparameter tuning procedure for the MSM fitted to 6-particle OM system for a 50 mW beam power. Analogous protocols are followed for other beam powers considered.

4.4.1 MSM hyperparameter optimization

To execute our analysis, we must first tune the hyperparameters $(\tilde{t}, n_M, \tau)$. Fig. 4.2a shows the implied time scale (ITS) plot for the MSM constructed for the 50 mW 6-particle OM system with $\tilde{t} = 8.3$. According to this implied time scale plot, when $\tilde{t} = 8.3$, five implied time scales can be resolved at a lag time $\tau = 10$ ms, meaning that $n_M = 6$ macrostates are identified for that lag time. Fig. 4.2b shows the silhouette score [141] plotted against

the number of clusters n_M, in which $n_M = 4$ has the largest score and $n_M = 6$ the second largest. However, according to Fig. 4.2a, at least four time scales are resolved at 10 ms lag time, so the number of macrostates should be greater than or equal to five. Therefore, we select $n_M = 6$.

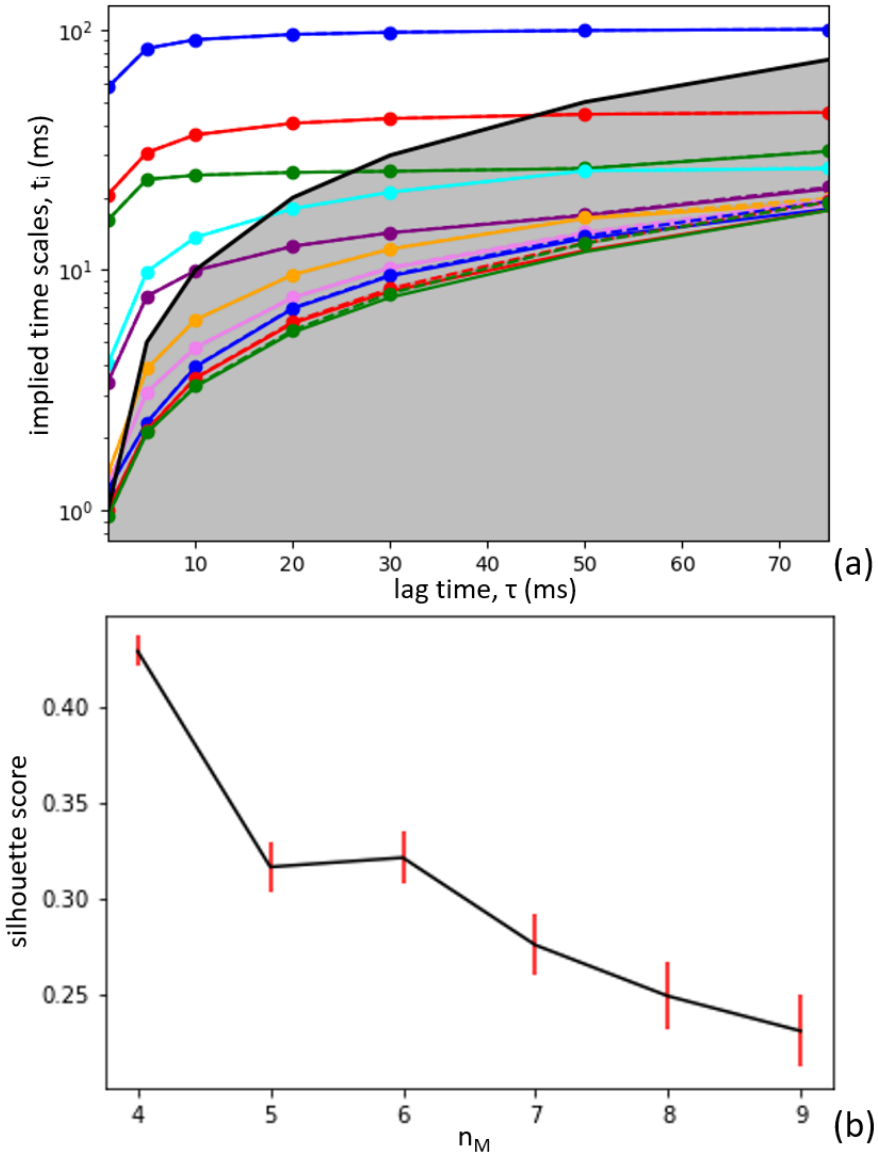

Figure 4.2: Analysis used for determination of number of macrostates at $\tilde{t} = 8.3$ and $\tau = 10$ ms for the 6-particle OM system under 50 mW beam power. (a) Implied time scale plot. The shaded grey area demarcates the region where the lag time exceeds the implied time scale. Implied time scales falling into this region cannot be distinguished within the time resolution of the resulting MSM. (b) Silhouette score plot against number of clusters n_M.

In order to evaluate the clustering results, we compare the predictions of the MSM metastable states with physical intuition. We know that OM structures with particles on hexagonal lattice sites typically form when the incident optical trapping beam is circularly polarized [11]. Therefore, given a certain OM structure with N particles we seek the best set of N sites on a 2D hexagonal lattice (i.e., a lattice pattern such as triangle, chevron, parallelogram, etc.) that is closest to the given configuration. We then categorize the particle configurations using the corresponding lattice patterns. The details of this lattice fitting method have been reported elsewhere [1].

We plot the Frobenius norm (F-norm) of the row-normalized assignment matrix (RNAM) for $\tau = 10$ ms with n_M and $\tilde{t}$ varying; the results are shown in Fig. 4.9 in Section 4.A. The (i, j) entry of the assignment matrix is the number of frames that is put into the i^{th} macrostate by the clustering method while classified into the j^{th} lattice pattern by lattice fitting. This matrix displays the matching relation between the clustering result and the lattice fitting result. Then, the RNAM can be generated by dividing each row by its row sum. (Analogously, the column-normalized assignment matrix (CNAM) can be generated by dividing each column by its column sum.) We can see from Fig. 4.9 in Section 4.A that the F-norm of the RNAMs converges as $\tilde{t}$ increases, indicating the stable identification of metastable clusters by diffusion k-means for sufficiently large $\tilde{t}$. It is clear that $\tilde{t} = 8.3$ lies within the converged region. The F-norm of the RNAM increases when n_M increases since there are more entries in RNAM, but this trend exhibits a knee at $n_M = 6$ and begins to fail to stably resolve a sufficient number of modes for $n_M > 6$, leading us to select $n_M = 6$ for the number of clusters. From this analysis, we identify $(\tilde{t}, n_M, \tau) = (8.3, 6, 10 \text{ ms})$ is a reasonable and robust tuning of the three hyperparameters for 50 mW beam power. As a final *post hoc* validation, we return to Fig. 4.7 to observe that the diffusion map embedding nicely distinguishes and separates the various macrostates and that the macrostates are in good agreement with the lattice pattern labels. We follow an analogous procedure to tune

the hyperparameters for the other beam powers.

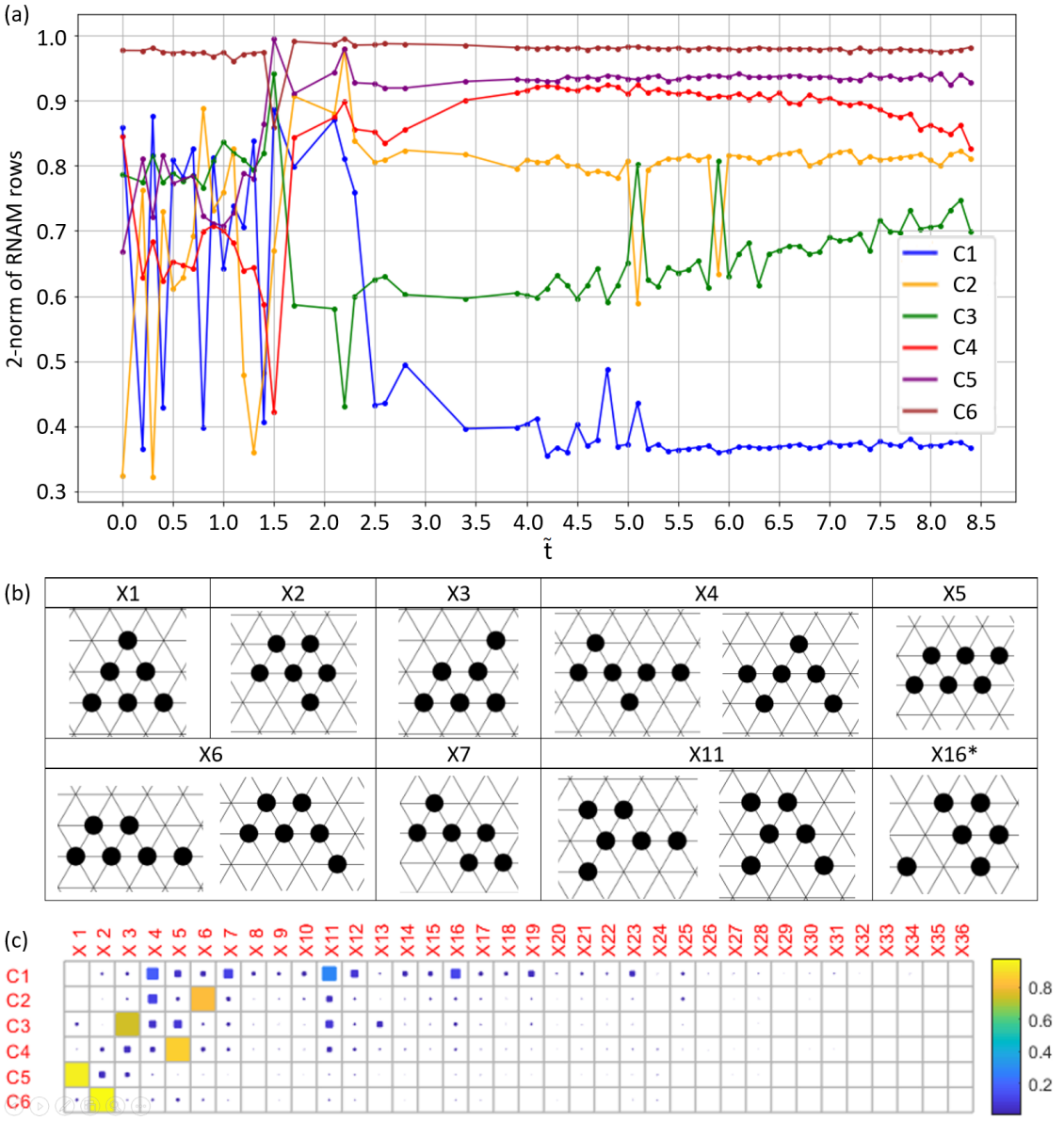

Figure 4.3: Details and interpretation of the clustering result for the MSM constructed at 50 mW beam power. (a) Plot of 2-norms of rows of the row-normalized assignment matrix (RNAM) against the diffusion k-means parameter $\tilde{t}$ for $n_M = 6$ and $\tau = 10$ ms. (b) The lattice patterns of some important lattice labels. The asterisk on X16 means that it corresponds to more than two lattice patterns where one particle is separated from the other five particles that are gathered together. (c) Illustration of the RNAM illustrating the assignment probabilities of each hexagonal lattice pattern (columns, X1-36) to each of the six macrostates within the learned MSM (rows, C1-6). The pattern of matrix elements indicates that C2-6 are high-purity macrostates comprising of largely a single lattice label, whereas C1 contains a mixture of lattice labels.

4.4.2 Analysis of MSM macrostate clustering

Fig. 4.3a shows the plot of 2-norms of the rows of the RNAM against the parameter $\tilde{t}$ for 10 ms lag time for 50 mW beam power for each of the MSM macrostates C1-6. The closer the norm is to 1, the better the clustering agrees with lattice fitting. We see that from the 2-norms of the RNAM rows for C2-6 converge to values in excess of 0.7 as $\tilde{t}$ increases, whereas that for C1 remains at a low value of only 0.4. This indicates that five of the six clusters well agree with lattice fitting and are quite insensitive to $\tilde{t}$. Fig. 4.3c presents the RNAM indicating the assignment probabilities of each lattice pattern (X1-36) to each of the six MSM macrostates (C1-6). The important idealized nanoparticle structures on a lattice and their lattice labels are shown in Fig. 4.3b, where X4, X6, and X11 correspond to two lattice patterns while X16 corresponds to many lattice patterns with one particle separated from the other five particles that are gathered compactly. The other lattice patterns are shown Fig. 4.10a in Section 4.A. Macrostates C2-6 are composed largely of a single lattice pattern, whereas C1 contains a mixture of patterns. This rationalizes the trends observed in Fig. 4.3a and leads us to expect that our MSM will contain five configurationally "pure" macrostates containing structures with a single long-lived OM lattice label, and a mixed macrostate containing structures with a mixture of lattice labels that rapidly interconvert on time scales shorter than the MSM lag time. Fig. 4.10b in Section 4.A displays the entries of the CNAM, showing that the five lattice patterns corresponding to the five stable macrostates are indeed not contaminated by other macrostates.

The Chapman-Kolmogorov (CK) test assesses the Markovianity of a fitted MSM and therefore determines whether or not it is a valid kinetic model [139]. The $(\tilde{t}, n_M, \tau) = (8.3,$ 6, 10 ms) macrostate MSM for a beam power of 50 mW satisfactorily passes the CK test as illustrated in Fig. 4.11 in Section 4.A. Passing the CK test also provides *post hoc* validation of our non-canonical use of permutationally-invariant diffusion map embeddings and reversibilization of the microstate transition matrix within our MSM pipeline, demonstrat-

ing that our approach provides a satisfactory means to construct valid macrostate MSMs for non-conservative and permutational-invariant systems.

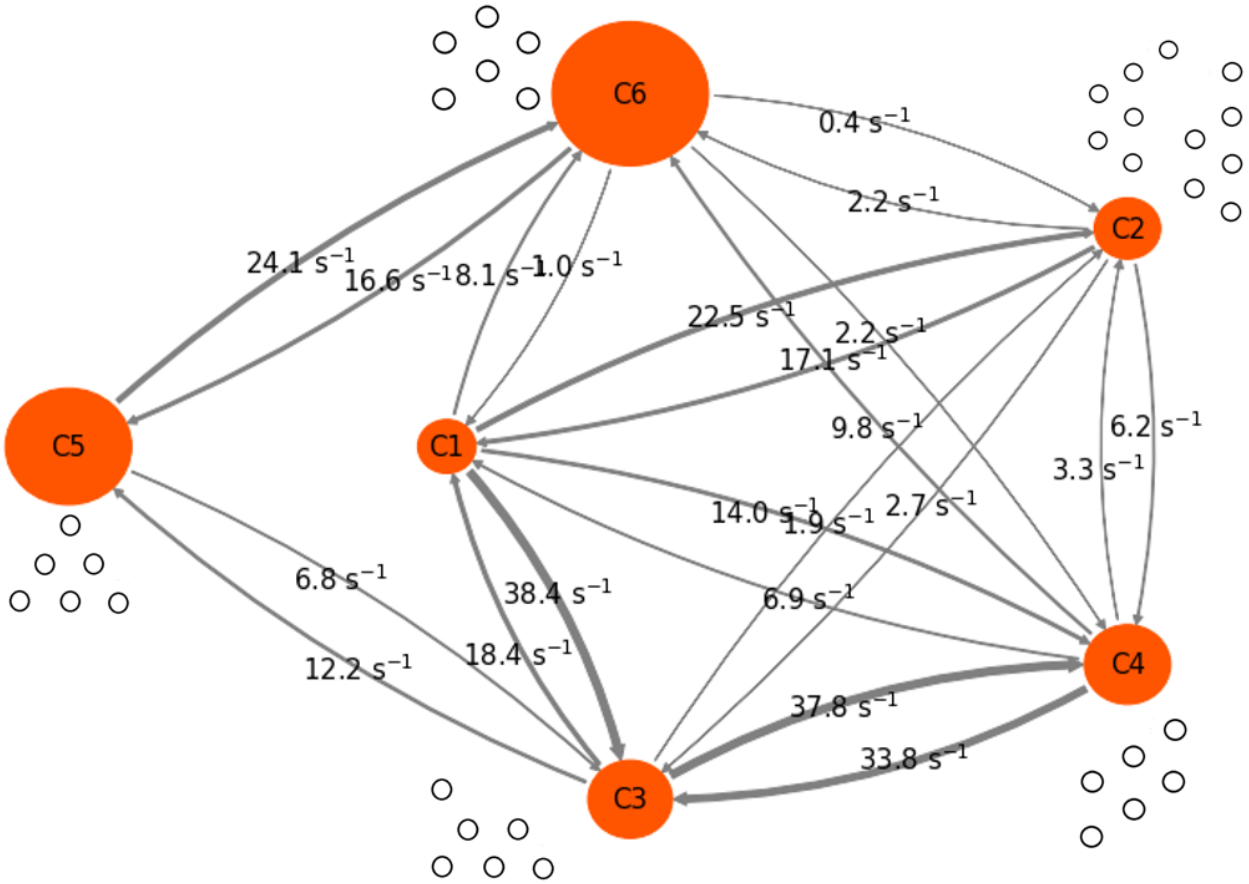

Figure 4.4: The state map of the Markov state model (MSM) built for the 6-particle OM system with $\tau = 10$ ms, $n_M = 6$ and (beam power, $\tilde{t}$)=(50 mW, 8.3). The sizes of the orange circles are proportional to the probability distributions of the macrostates. The thickness of the connecting curves is in accord with the magnitude of the transition rates. Representative lattice patterns are shown next to each macrostate; C1 contains a mixture of lattice patterns too numerous to display.

The macrostate MSM shown in Fig. 4.4 is the primary result of our analysis for the 50 mW beam power, and provides a wealth of interpretable and quantitative information on the metastable states and isomerization dynamics of the OM system. The sizes of the orange circles are proportional to the stationary distributions of the macrostates and the thickness of the connecting curves reflects the rate constants for transitions between macrostates. We

illustrate next to each macrostate a schematic representation of the representative lattice patterns corresponding to the long-lived lattice labels contained within each macrostate. C3-6 essentially contain a single lattice pattern. C2 contains two lattice patterns that interconvert on time scales below the MSM lag time. C1 contains a mixture of lattice patterns too numerous to display. We identify the chevron (C6), triangle (C5), and parallelogram (C4) states that have been previously observed and reported in experimental studies of this 6-particle OM system (cf., Fig. 4.1).

4.4.3 Beam power dependence of the dynamics of optical matter systems

In addition to the MSM constructed for the 6-particle OM system for 50 mW beam power, we employed an analogous approach to construct MSMs for beam powers of 40, 60, 70, 80, and 90 mW. The complete set of macrostate MSMs is presented in Fig. 4.12 in Section 4.A. By constructing MSMs over a range of beam powers we can analyze the ensemble of MSMs to extract trends in the relative stabilities of and transition rates between the various macrostates as a function of beam power.

Fig. 4.5 displays the changes with beam power of the stationary distribution probabilities of the six macrostates C1-6 and the rate constants of three selected macrostate-to-macrostate transitions (C4 $\rightarrow$ C3, C6 $\rightarrow$ C5, C5 $\rightarrow$ C6). Analogous plots for all 30 possible transitions are presented in Fig. 4.13 in Section 4.A. Focusing on the five configurationally pure macrostates we see that the abundances of macrostates C2 and C3 – each corresponding to states with a single unstable (i.e., "dangling") particle – decrease as the beam power increases. Similar trends are observed for macrostate C4 corresponding to the parallelogram state. On the other hand, the abundance of the triangle macrostate C5 increases as the beam power increases, and that of the chevron, macrostate C6, first increases and then decreases with increasing beam power achieving a maximum at around 60 mW. These quantitative trends inform us that we need to further increase the beam power to stabilize triangle structure, whereas we

need to decrease the beam power to further stabilize the parallelogram macrostate. We can make the chevron macrostate maximally important by tuning the beam power close to 60 mW.

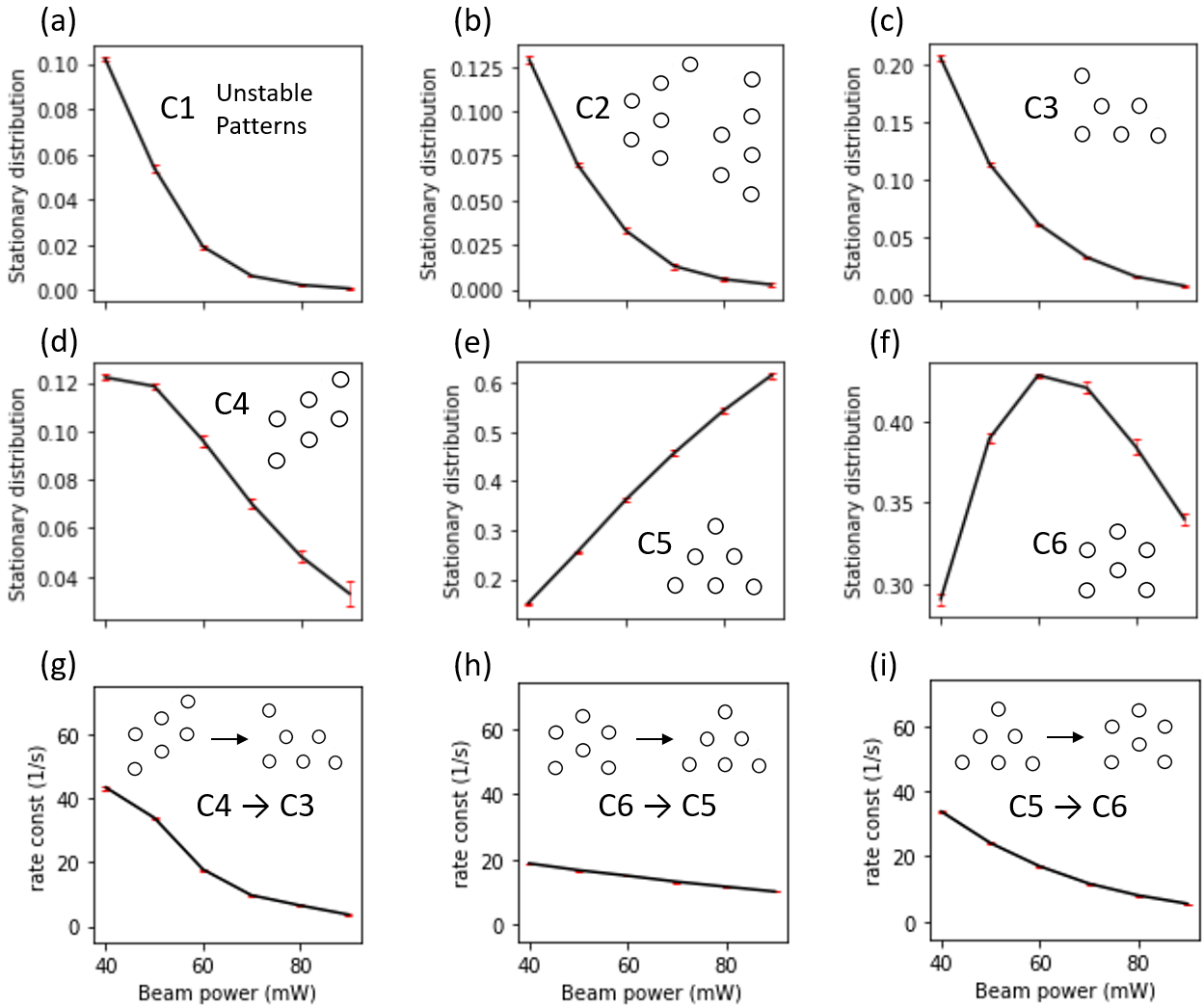

Figure 4.5: Plot of (a-f) the stationary distribution probabilities for the six macrostates C1-6 and (g-i) the rate constants of 3 selected macrostate-to-macrostate transitions (C4 → C3, C6 → C5, C5 → C6) as a function of beam power for the 6-particle OM system. Error bars represent standard errors in the mean estimated by five-fold block averaging.

The rates of the macrostate-to-macrostate transitions generally decrease when the beam power increases because the constraint exerted on the particles by the laser beam becomes larger with increasing beam power, leading to less freedom in the particle movement and

thus smaller transition rate constants. There are exceptions when the beam power drops to less than or equal to 40 mW, because the constraint exerted on the particles by the laser beam is then not large enough to stabilize the OM structures for sufficiently long periods of time.

4.4.4 Beam power design to achieve the maximum population for the chevron state

From Fig. 4.5f, we can see that the stability of the chevron pattern within macrostate C6 exhibits a non-monotonic behavior with respect to beam power. As an illustration of the value of our data-driven MSMs to inform control of the OM system, we adopt as our design goal maximal stabilization of the chevron pattern as a function of beam power. To do so, we carry out polynomial fitting of the chevron stationary distribution data for beam powers over the range 40 mW to 90 mW. Next, the Akaike information criterion (AIC) [142] is calculated for each fit, shown in the inset of Fig. 4.6, from which we can see that a 4^{th} order polynomial corresponds to the smallest AIC, and is therefore the fit most supported by the data. (The maximum degree of the polynomial fit used for AIC calculation is four because AIC cannot be calculated for higher order polynomials given only six data points.) The analysis of the 4^{th} order fit identifies a global maximum at a beam power of 62.94 mW, which represents our estimate of the beam power that maximally promotes stability of the chevron pattern within macrostate C6. To test this prediction, we carry out a simulation for a beam power of 62.94 mW, construct the corresponding macrostate MSM, and extract the stationary distribution of the C6 macrostate, which we plot as the red dot on Fig. 4.6. We can see that the predicted beam power indeed corresponds to a larger chevron population compared to other beam powers ranging from 40 mW to 90 mW. We could, of course, use the new 62.94 mW data point to further refine our beam power predictions by repeating this fitting and analysis approach. However, a fourth order polynomial fit to the new data

predicts the maximum to lie at 62.93 mW, which is within 0.01 mW of our existing estimate of the optimal beam power. Our ensemble of MSM models enables analogous optimizations to be performed to maximally promote macrostates or transitions of interest.

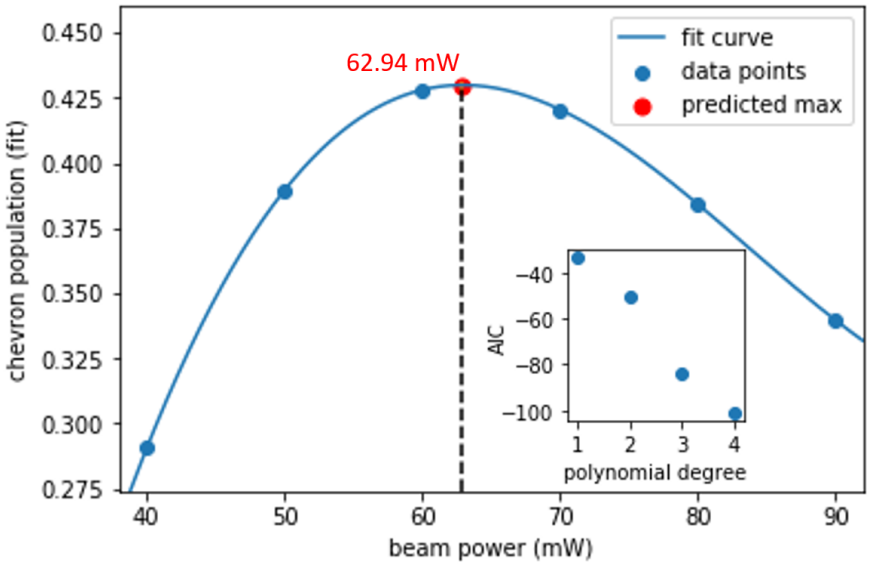

Figure 4.6: Dependence of the stationary distribution of the C6 (chevron) macrostate as a function of beam power. Blue points correspond to the stationary probabilities extracted from the corresponding macrostate MSMs constructed at beam powers of 40, 50, 60, 70, 80, and 90 mW. The blue line represents the best 4^{th} order polynomial fit to these data. The red point is the stationary distribution calculated from a MSM constructed at a beam power of 62.94 mW residing at the peak of the 4^{th} order polynomial fit. (Inset) Scatter plot of Akaike information criterion (AIC) against polynomial degree for fits to the six initial beam powers.

4.5 Conclusion

We have developed a MSM construction method for non-conservative systems with permutational invariance using permutationally-symmetrized diffusion maps and reversibilized microstate transition matrix construction. We applied this approach to non-conservative OM systems to understand how the stability of the various macrostates and transition rates depend on beam power. We found that as beam power increases, the stability of most macrostates decreases while the stability of the triangle state increases and that of the

chevron state first increases then decreases. A meta-analysis of our MSM models at various beam powers enables the rational control of the system via the design of beam powers to maximally promote particular self-assembled OM states or transitions. We found that the chevron macrostate reaches its maximum stability at a beam power of 62.94 mW.

The present paper represents a first proof of principle for this MSM construction method for the understanding and control of OM systems. In future work we plan to extend our analysis to additional controllable aspects of the incident beam including its phase profile and beam width in order to explore stabilization of additional self-assembled OM structures. We can also apply our analyses to more complex OM systems including those containing more particles, non-spherical particles, particles made of various other materials, or particle mixtures that have richer landscapes of self-assembled configurations. It is anticipated that this approach can help optimize the performance of optical matter machines [17]. We also anticipate that the approach may be extended to the analysis of molecular self-assembly where issues of permutational invariance and breaking of detailed balance must often be engaged in the construction of kinetic models [143].

Future improvement of the method can include determination of dynamic, as opposed to static, control policies that can wield tighter control on the stable state and transitions through simple feedback controllers that respond to the instantaneous state of the system and take the appropriate corrective action. These approaches can then be applied to construct MSMs directly from experimental as opposed to simulation data and use these models to guide experimental control strategies such as creating new stable OM structures and directing the transition of one structure to the other by on-the-fly adjustment of the beam parameters.

4.A Supplementary Information

For the case with 50 mW beam power demonstrated in section 4.4.1 and 4.4.2, let the macrostate count matrix be C_{macro} and the microstate count matrix be C_{micro}. We have,

$$\begin{aligned} \frac{\|C_{micro} - C_{micro}^T\|_F}{\|C_{micro}\|_F} &= 0.226 \\ \frac{\|C_{macro} - C_{macro}^T\|_F}{\|C_{macro}\|_F} &= 0.016 \end{aligned} \tag{4.11}$$

where $\|\cdot\|_F$ denotes a Frobenius norm. These magnitudes of the asymmetric component of the two count matrices illustrate that the macrostate count matrix is close to a symmetric matrix (within $\sim$1.6%) while the microstate count matrix is significantly asymmetric ($\sim$23% deviation).

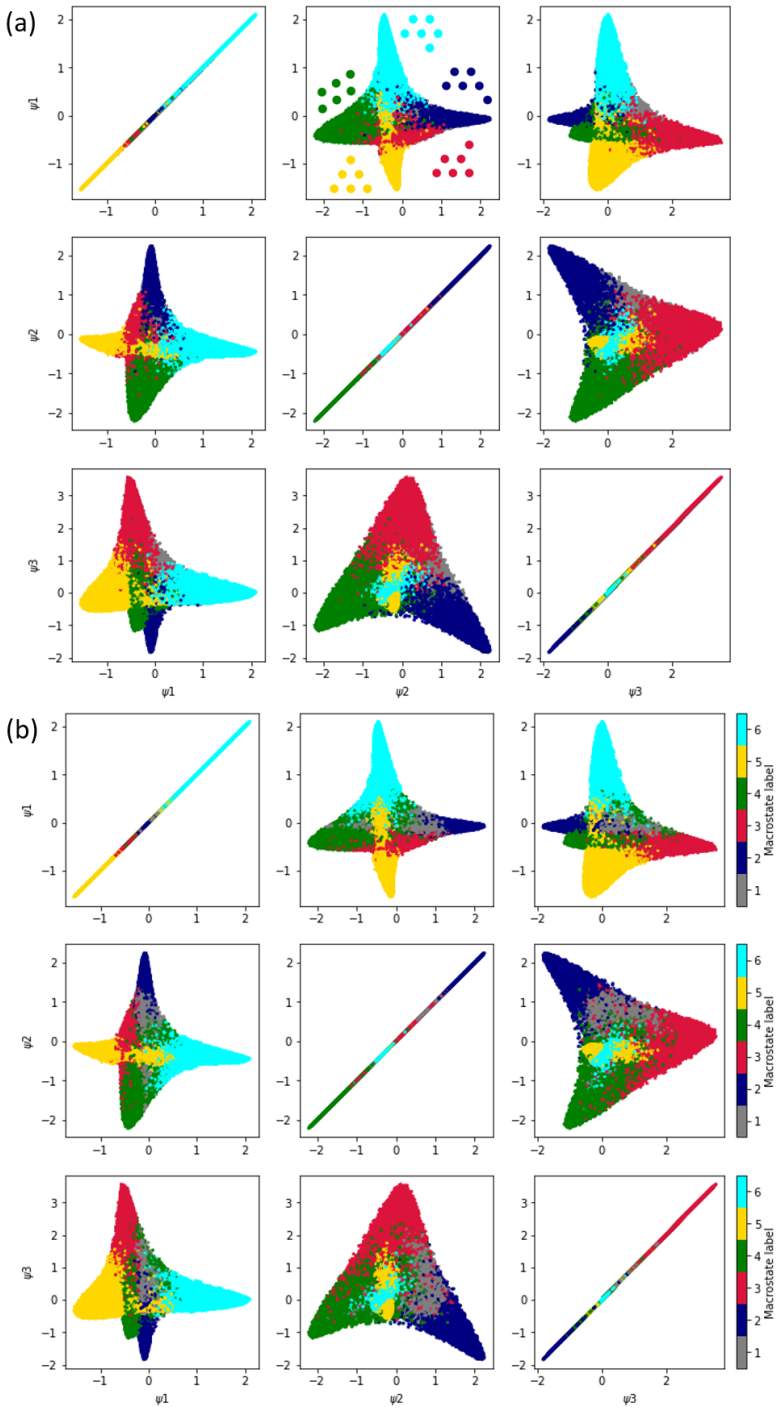

Figure 4.7: Diffusion map embedding plot colored according to (a) lattice pattern (b) macrostate labels generated for $(t, n_M, \tau) = (8.3, 6, 10$ ms) of one simulation trajectory for 50 mW beam power. The grey spots in panel (a) represents the configurations that are not identified as any of the 5 lattice patterns shown. ψ_i is the i^{th} non-trivial eigenfunction of the diffusion map. The diffusion map embedding distinguishes the various macrostates and the macrostates are in good agreement with the lattice labels.

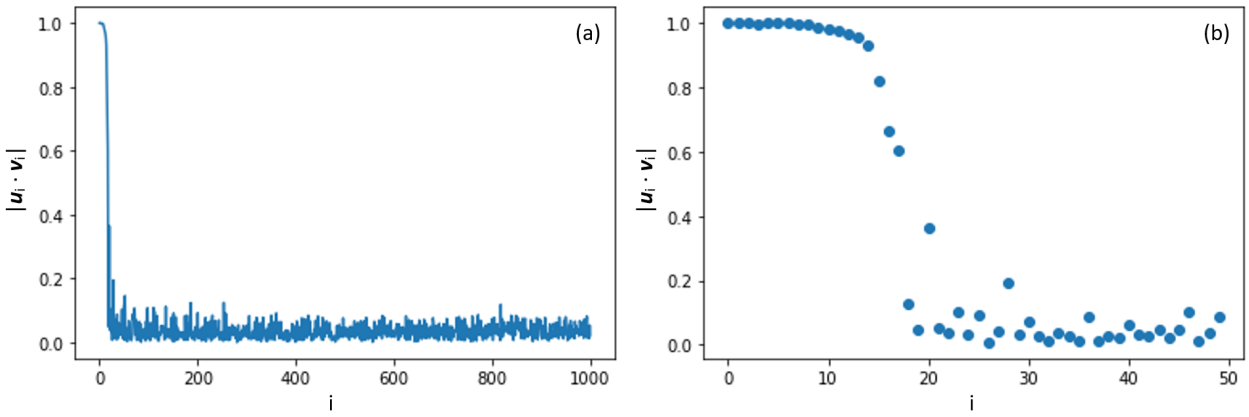

Figure 4.8: Plot of the absolute values of the dot products (cosine similarity) of the normalized right eigenvectors of the reversibilized and original microstate transition matrices for the 6-particle OM system under 50 mW with $(\tilde{t}, \tau)$=(8.3, 10 ms). $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ are the reversibilized and original normalized right eigenvectors, respectively. i denotes the eigenvector labels. The cosine similarity scores for the top $\sim$10 eigenvectors lie very close to unity, indicating close correspondence of these eigenvectors between the original and reversibilized systems.

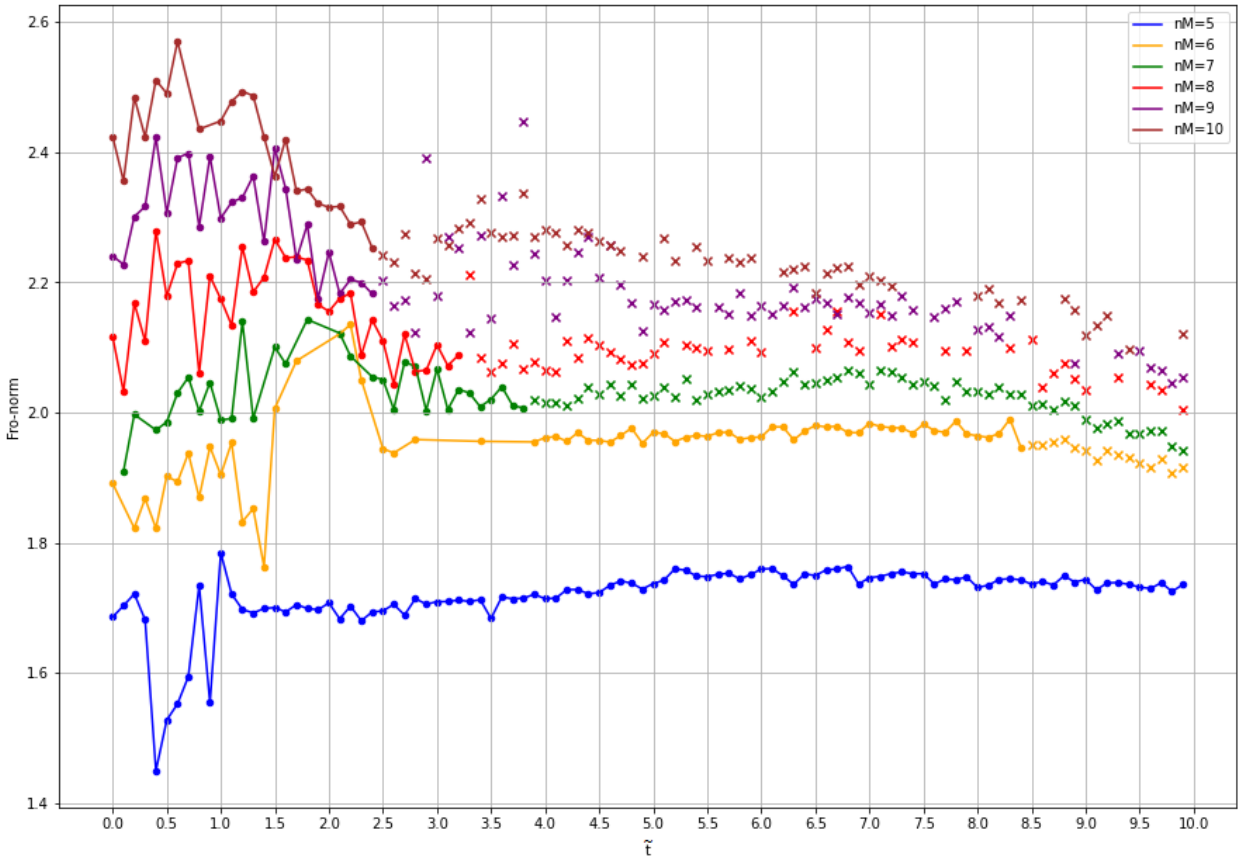

Figure 4.9: Plot of the Frobenius norm of the row-normalized assignment matrices against diffusion k-means parameter $\tilde{t}$ for various number of macrostates with lag time $\tau = 10$ ms for 6-particle OM system at a 50 mW beam power. For each value of $(n_M, \tilde{t})$ we render the corresponding data point in one of three ways. (1) If at least $(n_M - 1)$ implied time scales are resolved (i.e., are greater than the $\tau = 10$ ms lag time and lie above the grey area in Fig. 4.2a) then we plot using a circle. (2) If fewer than $(n_M - 1)$ implied time scales are resolved, we plot using a cross to indicate that insufficiently many modes are resolved to identify n_M clusters and that this combination of $(n_M, \tilde{t})$ is invalid. (3) If any of the n_M clusters is empty (i.e., no trajectory frames are assigned to this cluster), we decline to plot this point at all, again indicating an invalid $(n_M, \tilde{t})$ combination.

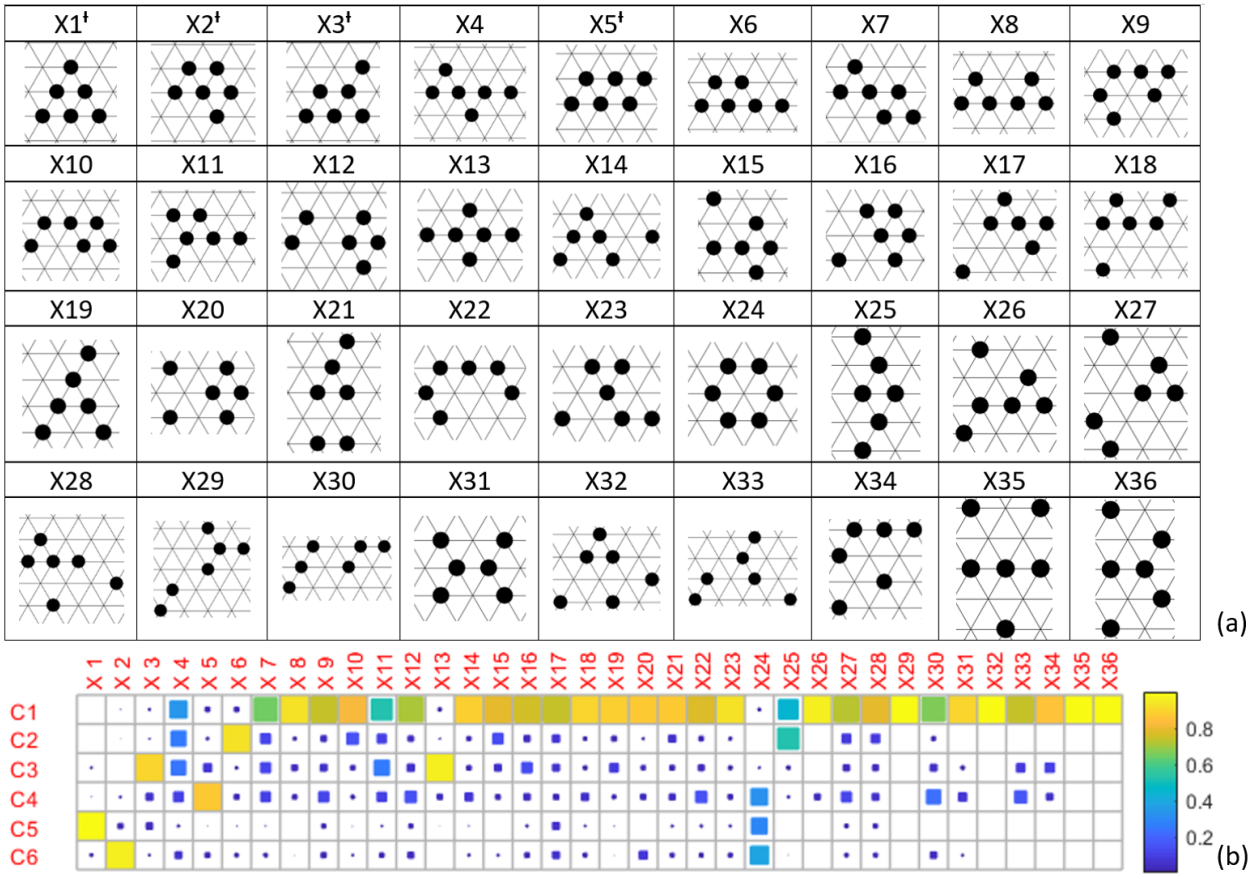

Figure 4.10: (a) The representative lattice patterns of the lattice labels X1-36. The lattice labels with the superscript † correspond to a unique lattice pattern while other lattice labels correspond to more than one lattice pattern. (b) Illustration of the column normalized assignment matrix (CNAM) of the clustering result for the MSM constructed at the 50 mW beam power, illustrating the assignment probabilities of each of the six macrostates within the learned MSM (rows, C1-6) to each hexagonal lattice pattern (columns, X1-36). The pattern of matrix elements indicates that the lattice patterns X6, X3, X5, X1, and X2 are mostly contributed by high-purity macrostates C2, C3, C4, C5, and C6, respectively.

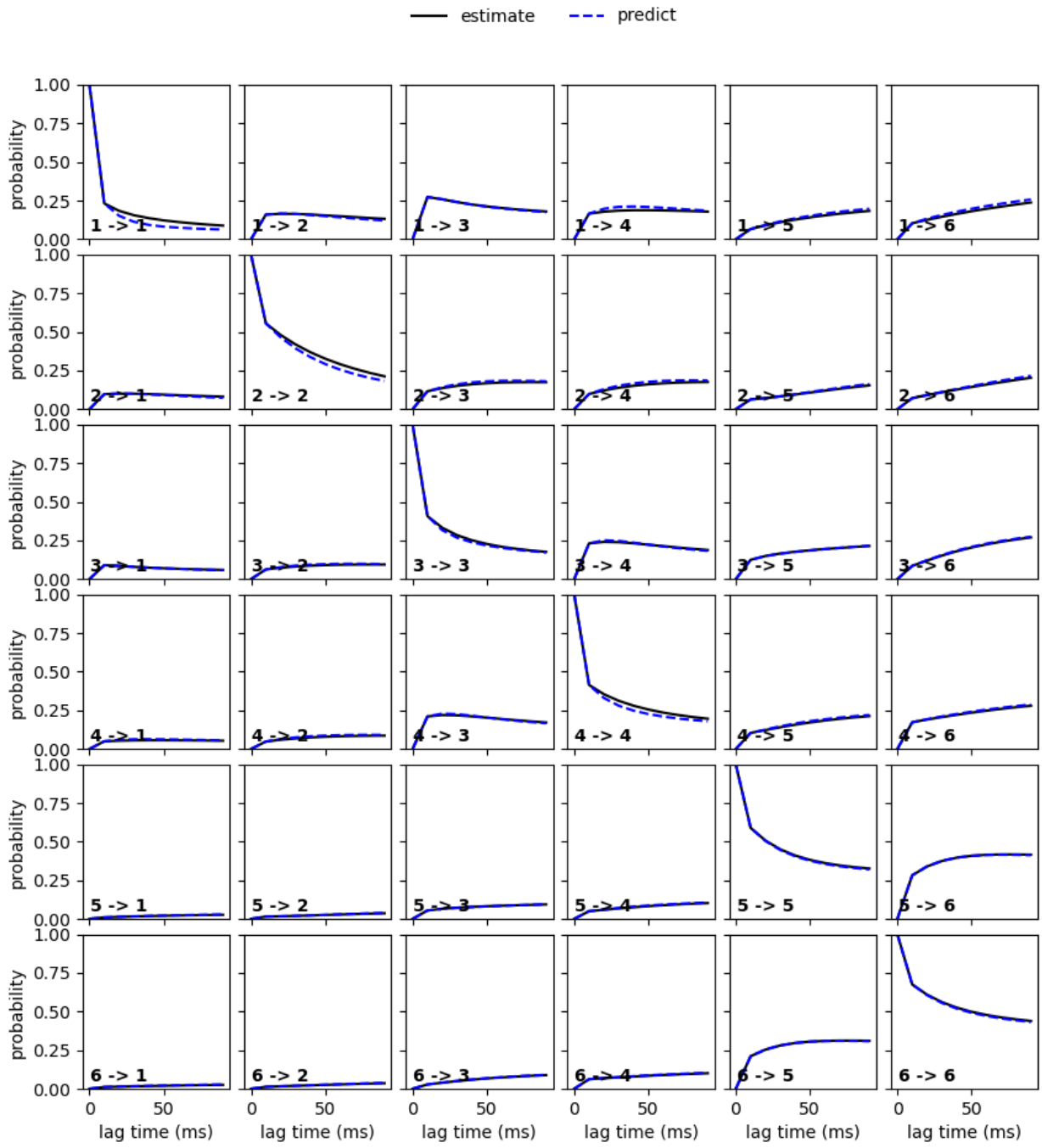

Figure 4.11: The Chapman-Kolmogorov (CK) test result computed for the Markov state model built for the 6-particle OM system under 50 mW beam power with parameter set $(\tilde{t}, n_M, \tau)$ =(8.3, 6, 10 ms). The CK test compares the transition probability elements of a macrostate transition matrix $\boldsymbol{T}(k\tau)$ computed at a lag time of $k\tau$ with those of a macrostate transition matrix computed at a lag time of τ taken to the k^{th} power $\boldsymbol{T}^k(\tau)$. If τ is sufficiently large for the system to be Markovian, then $T_{ij}(k\tau) \approx T_{ij}^k(\tau)$ and there should be agreement between all (i, j) matrix elements. The excellent agreement between $\boldsymbol{T}(k\tau)$ (black solid line) and $\boldsymbol{T}^k(\tau)$ (blue dashed line) indicates that this MSM is a valid kinetic model.

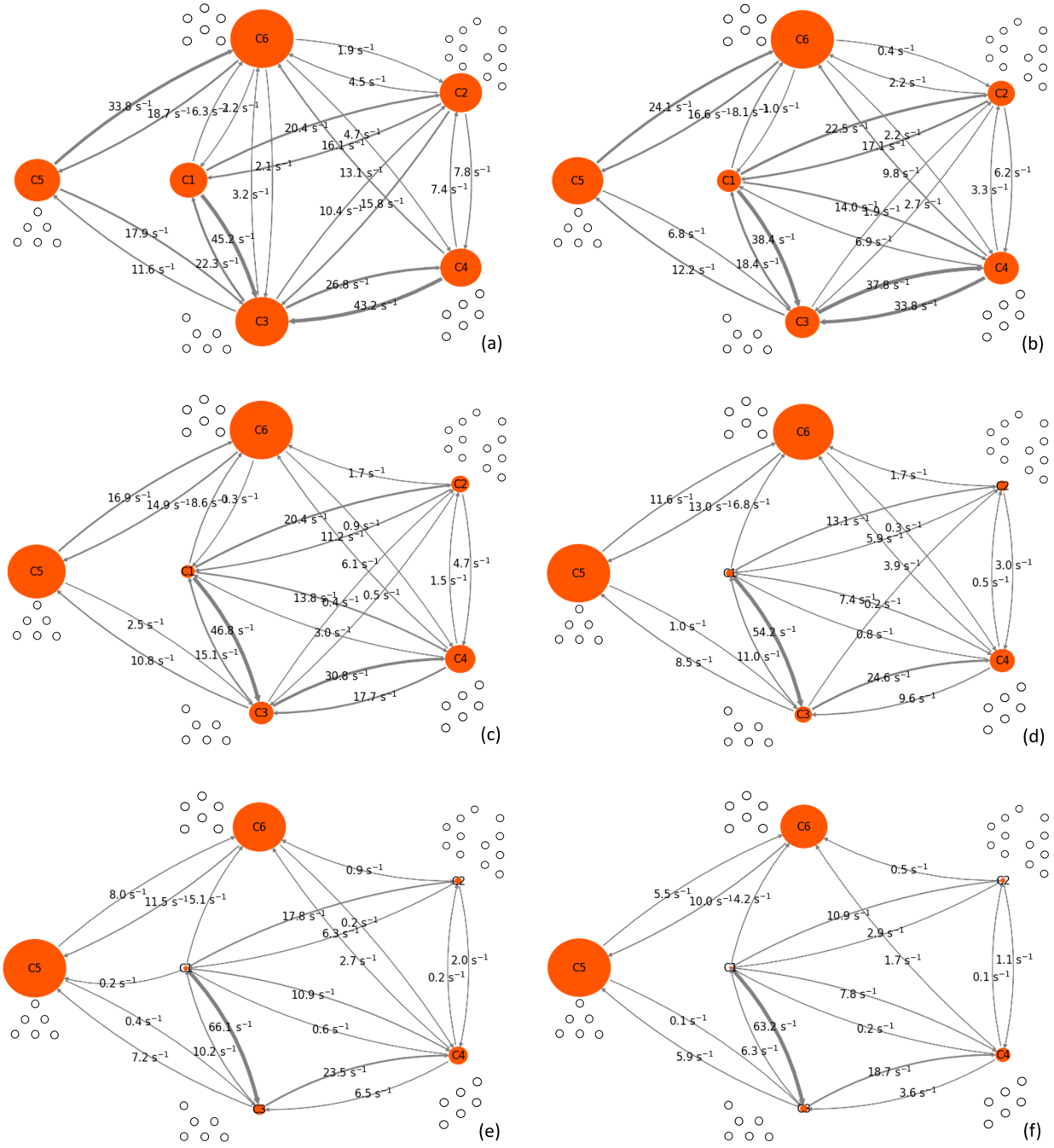

Figure 4.12: The state maps of the Markov state models built for the 6-particle OM system with $\tau = 10$ ms, $n_M = 6$. The beam powers and the values of $\tilde{t}$ corresponding to the panels are: (a) (40 mW, 6.3); (b) (50 mW, 8.3); (c) (60 mW, 7.3); (d) (70 mW, 5.4); (e) (80 mW, 5.3); (f) (90 mW, 4.5). The sizes of the orange circles are proportional to the stationary probability distributions of the macrostates. The thickness of the arrows is in accord with the magnitude of the transitions. Representative lattice patterns are shown next to each macrostate, except when the number of corresponding lattice patterns is too numerous to compactly display.

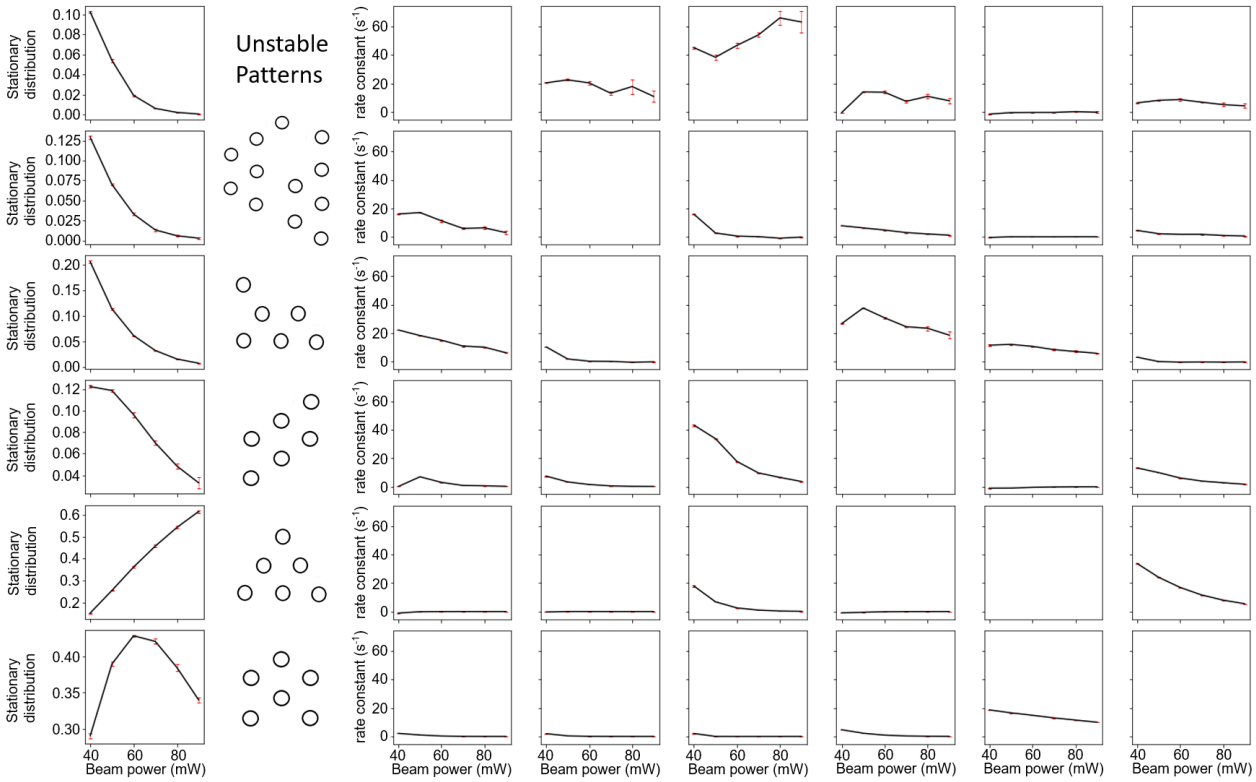

Figure 4.13: Stationary distribution probabilities of the macrostates and the rate constants of the transitions among the macrostates as a function of beam power for 6-particle OM system. The left-most column contains the stationary distribution probabilities and are annotated by the corresponding lattice patterns. The (i, j) entry of the matrix on the right indicates the rate constant of the transition from macrostate i to macrostate j. Error bars represent standard errors in the mean estimated by five-fold block averaging.

CHAPTER 5

POWER DISSIPATION AND ENTROPY PRODUCTION RATE OF HIGH-DIMENSIONAL OPTICAL MATTER SYSTEMS

This chapter is based on the following manuscript under review:

Shiqi Chen, Emmanuel Valenton, Grant M. Rotskoff, Andrew L. Ferguson, Stuart A. Rice, Norbert F. Scherer. "Power dissipation and entropy production rate of high-dimensional optical matter systems" *Physical Review E*, 2024, under review.

5.1 Abstract

Entropy production is an essential aspect of maintaining nonequilibrium systems and their relaxation. However, calculation of the entropy production (rate) is challenging for high dimensional systems so it has only been reported for simple systems. Moreover, there is a dearth of nontrivial experimental systems where precise measurements of entropy production rate and characterization of the NESS are simultaneously possible. Optical matter (OM) systems can be used to build optical matter machines that do mechanical work under a laser beam. They consist of (nano-)particles organized into ordered arrays that are bound by electrodynamic interactions associated with the scattering, interference and induced polarizations in and amongst the particles in (focused) coherent optical beams. The flux of laser light into OM systems necessitates that they dissipate energy and produce entropy, and thereby can relax to a non-equilibrium steady state (NESS) condition. The NESS may have several ordered particle configurations that can interchange by barrier crossing processes. Understanding the power dissipation and entropy production rate of a NESS in an OM system along different (collective) modes of motion can advance understanding of the relative stability of the NESSs as well as inform design and control of OM structures. We report an approach to calculate the entropy production rate of overdamped, non-conservative sys-

tems and demonstrate this on a 6-particle triangle OM structure as a non-trivial example. In particular, we compute the components of power dissipation along the previously published collective coordinates of that structure. The approach obtains the entropy production rate and power dissipation numerically from OM NESS trajectory data. By verifying the Seifert relation [U. Seifert Rep. Prog. Phys. 75 126001 (2012)] for these overdamped non-conservative OM systems, we demonstrate that they can support phase transitions and other collective phenomena.

5.2 Introduction

Systems in non-equilibrium steady states (NESS) are defined by a time-independent probability density distribution in its configuration space. Maintaining a NESS requires dissipating power and thus continuously producing entropy [144]. The role of the entropy production rate in determining the stability of a NESS has long been debated [145]. Unfortunately, the calculation of the entropy production rate requires knowledge of the flux density or power dissipation [144], both of which require extensive sampling in the system phase space that is very demanding for multi-particle systems. Therefore, it is useful to find an approach to calculation of entropy production rate by making approximations on the flux density.

There is a dearth of nontrivial experimental systems where precise measurements of entropy production rate *and* characterization of the NESS are simultaneously possible. Previous theoretical treatments of the design and control of a NESS either study systems with low dimensional configuration spaces [146, 147], or rely on an explicit expression of the force field [148]. Entropy production rate was quantified in previous works within single-particle non-conservative force fields using upper bounds [149]. These approaches cannot be used to study the structure of systems where the force field computation is implicit. Furthermore, they do not scale well to N-particle systems where the dimensionality of configuration space is 2N.

Optical matter (OM) systems are a class of active non-equilibrium materials [6,9,17,105]. One of the most interesting variants consists of nano-particles (NPs) that form 2-dimensional ordered structures when illuminated and trapped by a focused laser beam [9, 106, 107]. The force field developed by the electrodynamic interactions that hold the NPs together is non-conservative. Depending on the number of NPs and the phase, amplitude and polarization properties of the incident electromagnetic field, there are several different metastable ordered structures that can be formed. The relative stabilities of these structures can be tuned by adjusting the aforementioned laser beam properties [2]. Therefore, the beam power, beam shape, spatial phase profile, and polarization of the light [41,108–111] create a rich parameter subspace to explore stabilization, control and design of particular OM structures and their dynamics [2, 74, 150]. Each of the one or more different ordered OM structures that form in a focused laser beam constitutes a (metastable) NESS, which can be used to build optical matter machines that do mechanical work under a laser beam [17].

In this paper we use our realistic numerical simulations, which helped generate reaction coordinates and build Markov state models for OM system in previous works [1,2], to provide new understanding of the experimental systems by probing entropy production to quantify the NESS. Here we report an efficient approach for the calculation of the entropy production rate in optical matter NESS's. We use the six-particle OM triangle structure held together by electrodynamic interactions as an example. With no transition to a different OM NESS as an assumption for simulation, we computed the entropy production rate of the the six-particle OM triangle structure by analyzing the simulation trajectory and compared it to its mean power dissipation, which is generally easy to compute but relies on large data set size to achieve sufficient precision. Also, the power dissipation is structure- and path-dependent as we show in Fig. 5.1.

The primary objectives of this paper are to develop approaches to compute the entropy production rate in numerical simulations of NESS's with internal dynamics, and to under-

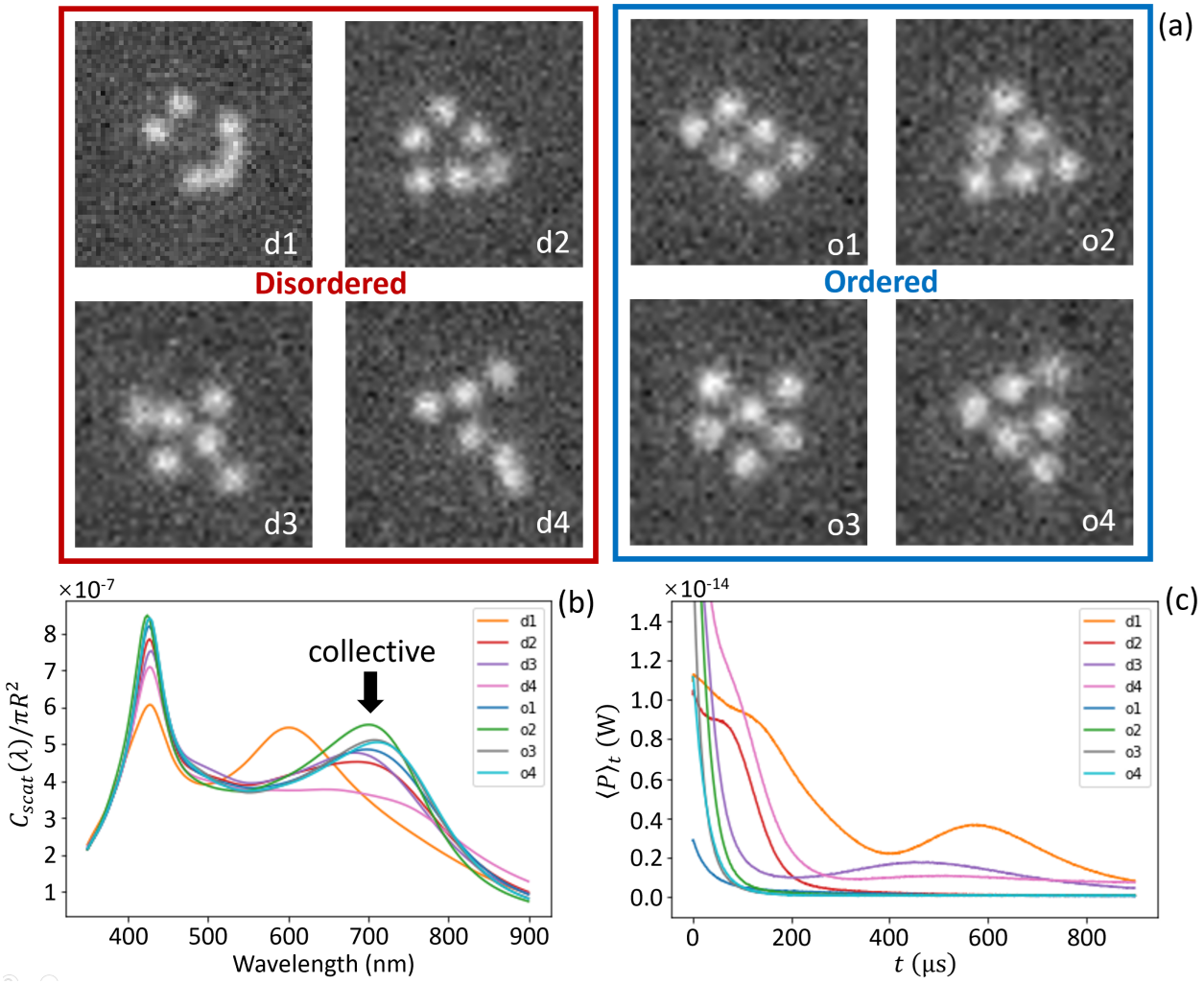

Figure 5.1: Demonstration of the difference in the collective scattering resonance and power dissipation between disordered and ordered 6-particle OM structures. (a) Representative experimental (dark-field microscopy) images taken (with 90× magnification) of an OM system with 6 silver 150 nm dia. NPs trapped in a focused right-hand circularly polarized laser beam. (b) Plot of per particle scattering cross section (normalized by πR_p^2, where $R_p = 75$ nm is the radius of a single NP) against the wavelength of light from a probe source. (c) Plot of power dissipation averaged over 2.5 million Langevin dynamics simulation trajectories as a function of the elapsed time of the simulation. Each trajectory consists of 10,000 time steps of 0.5 μsec/step. (Note: Matlab uses the convention "e^{-x}" for exponentials with base 10, which we adopted in our previous manuscripts [1, 2] Here we use "$\times 10^{-x}$" instead to avoid confusion.)

stand how those internal motions and the forces that generate the NESS affect the entropy production rate. The vehicle for our calculations is the six particle triangle configuration

NESS, one of the stable structures of the 6-particle OM system; an example of a multi-particle high-dimensional system [151, 152]. We use this OM NESS system to compare the entropy production rate to power dissipation. After decomposition of the external force field into conservative and non-conservative parts, we numerically compute the mean entropy production rate of the six particle triangle configuration using a linear approximation to, and Helmholtz decomposition of, the force field, analyzed in terms of the 12 collective coordinates defined by principal components of the triangle configuration we reported previously [1]. We also compute the dissipated power along the collective coordinates. We find that the the rotation mode of the OM structure (an explicitly non-conservative dynamics) dominates the power dissipation. The collective coordinates in this system are determined by the conservative part of the force field, but the non-conservative part of the total force still affects the collective coordinates through symmetry breaking of the system. We find a negligible effect of the non-conservative part of the force field of the system on the position auto-correlation function. Understanding and controlling the OM system can help us create more OM structures with different optical trapping conditions.

5.3 Methods

5.3.1 Experimental

Experimental studies of the 6-particle OM system were conducted using a single-beam circularly polarized optical tweezer in an inverted microscope set up as described previously [7,15]. The experimental sample was a dilute water solution with a mixture of polyvinyl pyrrolidone (PVP)-coated 150-nm diameter silver NPs. A continuous wave Ti-Sapphire laser beam (λ = 800 nm) was focused near the top glass cover-slip using a 60x microscope objective, pushing a small number of NPs toward the upper glass surface of the aqueous sample cell. The laser power was 200 mW before entering the microscope, wherein some power is lost

before focusing. The beam width is 1.29 μm. A spatial light modulator was used to slightly defocus the trapping beam to be converging at the sample. Thereby applying an inward-directed phase gradient and phase gradient force [11]. Electrostatic repulsion between the ligands on the NPs and the glass cover-slip balances the radiation pressure, resulting in a 2D trapping condition.

5.3.2 Simulation

Electrodynamics-Langevin dynamics (EDLD) simulations were performed using Generalized Multiparticle Mie Theory [37,38] as input forces $\boldsymbol{F}_{ED}(\boldsymbol{x})$ to the Langevin equation. The code is MiePy software developed by the Scherer Lab [39]. The resulting EDLD solver performs a numerical integration of the following Langevin equation:

$$m\frac{d^2\boldsymbol{x}}{dt^2} = \boldsymbol{F}_{ED}(\boldsymbol{x}) - \boldsymbol{\nabla}U_{DL}(\boldsymbol{x}) - \zeta\frac{d\boldsymbol{x}}{dt} + \boldsymbol{f}^{(b)} \tag{5.1}$$

where m is the mass of the particle; $\boldsymbol{x}$ is the Cartesian coordinate of the structure; t is time; $\zeta = 6\pi\eta r$ is the friction coefficient for particle radius $r = 75$ nm; η is the viscosity of the medium (taken to be water at 298 K); $\boldsymbol{f}^{(b)}$ is the bath random force term defined to be white noise satisfying the fluctuation dissipation relation at a specific temperature; $\boldsymbol{F}_{ED}$ is the sum of external electrodynamic force fields computed by the MiePy package [39]. When beam power is 200 mW, the average beam power density is 8.3×10^{10} W/m^2 in the circle with the radius equal to the standard deviation of the Gaussian intensity profile of the laser beam. The simulation time step between a pair of frames is 0.5 μs. U_{DL} is the double layer electrostatic interaction [40]:

$$
\begin{aligned}
U_{DL}(\boldsymbol{x}) =& \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} V(r_{ij}) \\
V(r_{ij}) =& 32\pi\varepsilon_0\varepsilon_m \left(\frac{k_B T}{z_p e}\right)^2 \tanh^2\left(\frac{z_p e\psi}{4k_B T}\right) \\
& \times \exp\left[-\kappa\left(r_{ij} - 2r\right)\right]
\end{aligned}
\tag{5.2}
$$

where N is the number of particles; ε_m is the medium electric permittivity; $z_p = 1$ is the ion valence; $\psi = -77$ mV is surface electrostatic potential; $\kappa^{-1} = 100$ nm is the Debye screening length; r is the particle radius; $r_{ij} = \left|\boldsymbol{x}_i - \boldsymbol{x}_j\right|$ is the distance between the i^{th} and j^{th} particle. The Debye screening length was determined in a previous simulation experimental comparison [2].

5.4 Results

5.4.1 Ordered and disordered OM structures

Fig. 5.1a shows 4 representative disordered structures and 4 representative ordered OM structures, labeled d1-d4 and o1-o4, respectively, from a 6-nano-particle OM experiment. We analyzed these structures by computer simulations that took them as initial configurations. Fig. 5.1b shows that the "collective" scattering resonances of the disordered structures have shorter wavelengths (near 200 nm) than those of the ordered structures [3, 17]. In fact, the scattering spectrum of d1 is more reminiscent of the scattering of individual 150 nm dia. Ag NPs than the lattice resonance of well-ordered structures [3]. Fig. 5.1c shows the mean power dissipation plotted against time elapsed in the Langevin dynamics simulation trajectory initiated at each of the eight structures shown; the power dissipation of the ordered structures is less than that of the disordered structures. This observation suggests that a deeper understanding of the relationship between power dissipation and entropy production rate can potentially lead to new approaches to design and control OM systems.

The large range of relaxation behaviors shown in Fig. 5.1c can arise because the initial condition (structure) for each simulation is from experiment, and are variously close to a stable (minimum energy) structure. Therefore, it takes time for the initial configuration to relax [153]. The power dissipation decreases during this relaxation period since it takes more power to maintain an unstable structure than a stable one. Also, the mean power dissipation does not go to zero at long simulation times, but asymptotes to a value several orders of magnitude smaller than the initial value. Therefore, for short simulation times, the particle configuration is still displaced from a stable structure, so the power dissipation is large. The small but non-zero power dissipation of the OM system near a stable structure is described in detail later (in Fig. 5.5) for the six-particle OM triangle structure.

5.4.2 Collective coordinates

Analogous to the vibrational normal modes of molecules, collective modes can be defined for the stable OM structures [1]. However, the electrodynamic forces exerted on the OM system are non-conservative, hence the gradient matrix of the force field acting on an OM NESS is not symmetric, and its eigenvectors are not orthogonal to each other. In order to ensure orthogonality among the collective coordinates of the OM structure(s) observed in experiments, we carried out EDLD simulations of the fluctuations of each OM NESS. The deviation between the nanoparticle locations at each time step of the trajectory and the reference structure, $\left\|\boldsymbol{x}(t) - \boldsymbol{x}_{ref}\right\|_2$, is determined in the fluctuation simulation. If the deviation is larger than a threshold, set as 275 nm for the 6-particle triangle state, the simulation is restarted from 1000 frames earlier so as to ensure that the OM structure simulated does not undergo transitions to other stable OM structures. This threshold is chosen according to two factors. First, the most probable state (structure) that the triangle structure can transform to is the chevron [2]. Second, 275 nm corresponds to a committor equal to 0.5 for the transition between triangle and chevron [1]. To account for the rotational symmetry

of the force field, the rotation angles of all the structures in the sequence of trajectory time steps are aligned, after which a principal component analysis (PCA) is performed on the trajectory data set. The principal components (PCs) generated by the PCA are taken as the collective coordinates (basis) of the NESS.

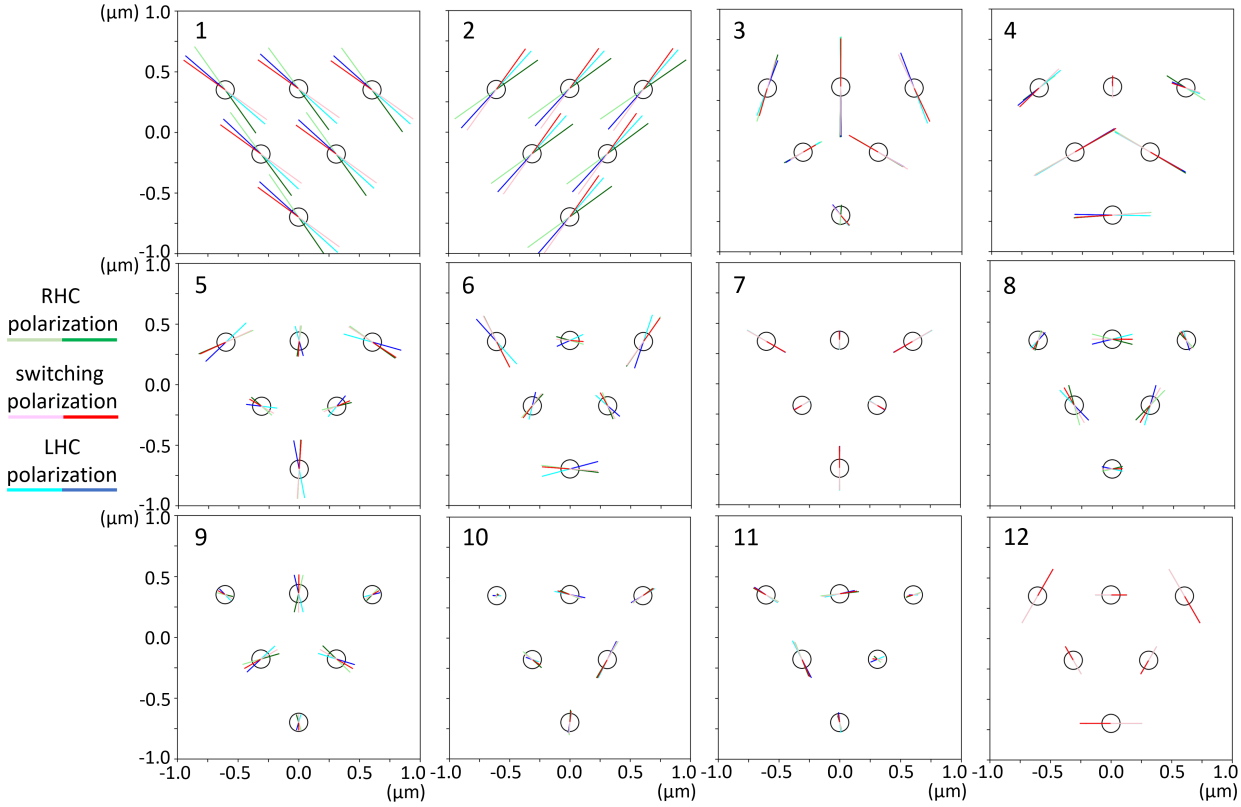

Figure 5.2: Twelve principal components (modes) of the triangle structure of the 6-particle OM system. The colored solid lines depict the magnitudes and directions of the collective particle motions, and the color defines the direction (i.e., phase) of the collective motion; i.e., particles simultaneously move in the indicated directions for the same color. PCs 1 and 2 correspond to translation and PC 12 to rotation. PCs 3-11 correspond to motions that are analogous to vibration while the length of the solid lines of these PCs is proportional to the standard deviations along these PCs. The green and blue solid lines represent fluctuation amplitudes for right- and left-hand circularly polarization laser beams, while the red lines also show the effect of switching between right- and left-hand circular polarization alternately for adjacent simulation time steps. Note that PC modes 1 and 2 are termed translations, but are more akin to vibrations for the OM structure in a Gaussian (confining) beam vs. plane wave illumination.

Fig. 5.2 shows the principal components computed for the six particle triangle structure,

one of the metastable structures of the 6-particle OM system. The length of the fluctuation simulation is 25 million frames (i.e., time steps). See Fig. 5.7 in Section 5.D for the results of convergence tests of the PCs and the variances along the PCs. In these calculations the temperature was taken to be 298 K. Noting that the chirality of the force field breaks the symmetry of the system, we examined several different laser polarizations to study the relationship between the collective modes and the non-conservative force in the system.

The 12 PCs shown in Fig. 5.2 are obtained from PCA applied to the fluctuation data of the 6-particle triangle OM structure. See Fig. 5.8 and Fig. 5.9 in Section 5.D for the demonstration that these PCs are uncorrelated and that the marginal probability distribution of the simulation data along these PCs are very close to Gaussian. The pairs of PCs (1,2), (3,4), (5,6), and (10,11) correspond to the 2-dimensional irreducible representation of the D3 2-dimensional point group that describes the symmetry of the triangle structure: the two PCs in each pair are degenerate, hence the exact choice of the two basis vectors in each degenerate eigenspace is sensitive to random error of the data set. The rotation mode (# 12) is aligned during pre-processing of the data due to the rotational symmetry of the system, hence its standard deviation is zero.

Of the 12 PCs, modes 7, 8, and 9 shown in Fig. 5.3 carry the most interesting information. These three modes are non-degenerate and demonstrate the influence of the symmetry breaking due to the chirality of the force field. With the laser changes from left- to right-hand circular polarization, the collective motion along these three modes changes into its mirror image. If the laser is switched rapidly between left- and right-hand circular polarization, these three PCs achieve mirror symmetry, and their displacement directions lie between those of the purely left- and right-hand circularly polarized displacement directions.

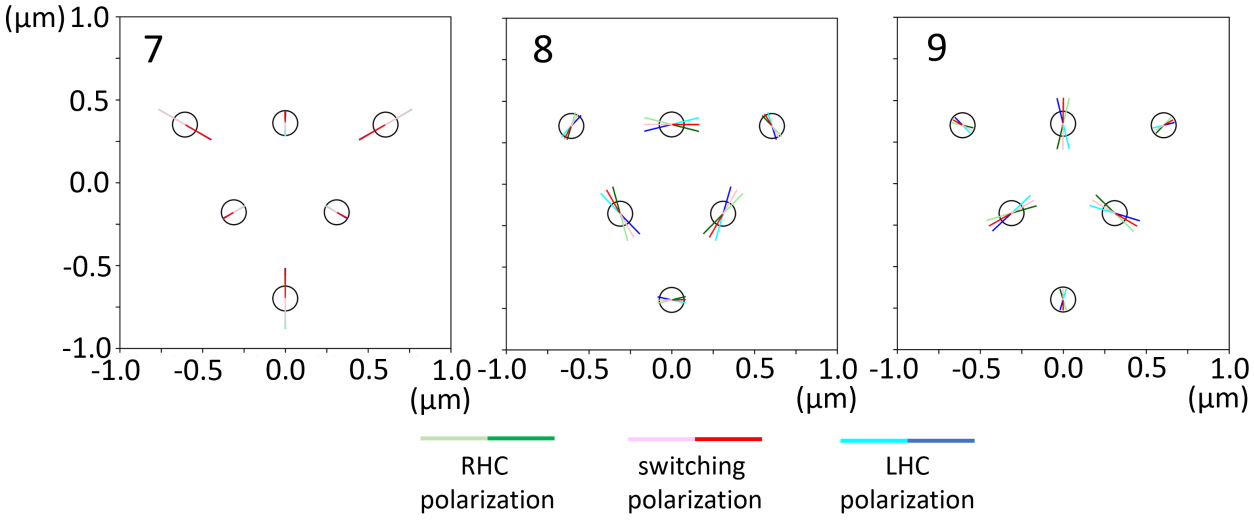

Figure 5.3: Detailed view of the non-degenerate principal components of the triangle structure of the 6-particle OM system. The colored solid lines depict the magnitudes and directions of the collective particle motions, and the color defines the direction of the collective motion; i.e., particles simultaneously move in the indicated directions for the same color. The lengths of the solid lines of these PCs are proportional to the standard deviations along these PC coordinates. The green and blue solid lines represent right- and left-hand circularly polarization laser beams while the red lines also show the effect of switching between right- and left-hand circular polarization for adjacent simulation time step.

5.4.3 Damping condition analysis

We now establish that the ordered OM structures we observe are overdamped multiparticle oscillators. We used the data from simulations at 273 K and 200 mW incident laser power to compare the single particle damping rate $\zeta/m = 6\pi\eta r/m$ to the collective coordinate vibration frequencies

$$f_i = \frac{1}{2\pi}\sqrt{\frac{k_B T}{m\sigma_i^2}} \tag{5.3}$$

where σ_i^2 is the variance along the i^{th} collective coordinate. This comparison is important because the damping rate must be much larger than the vibrational frequencies for an overdamped Langevin dynamics computation to be valid.

The comparison, shown in Fig. 5.4a, is that ζ/m is more than 2 orders of magnitude larger

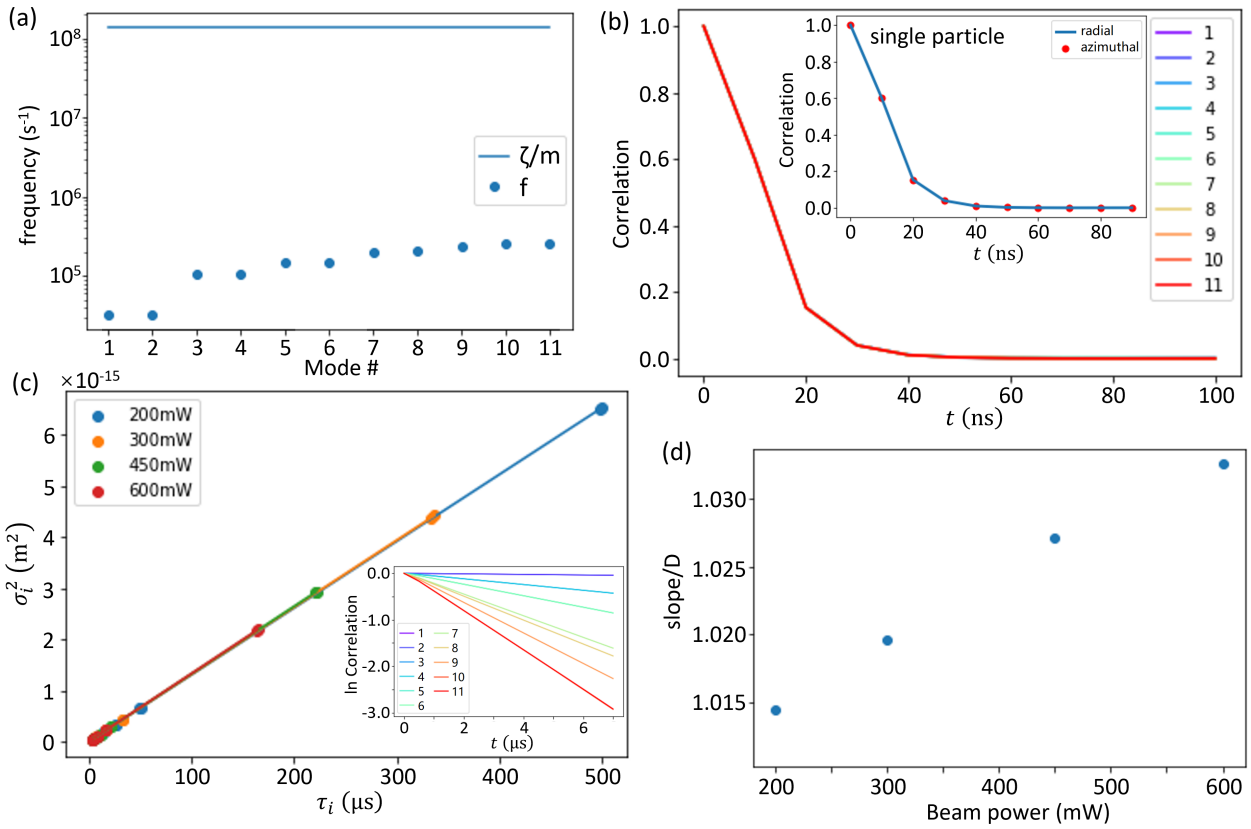

Figure 5.4: Damping condition analysis for the triangle structure of the 6-particle OM system. (a) Plot of damping frequency ζ/m (solid blue line) and the oscillator frequencies (blue dots) along the collective coordinates for $T = 273$ K and beam power 200 mW. (b) Velocity auto-correlation functions of single particle (in the inset) and auto-correlation function of the velocity projected onto the collective coordinates for the triangle structure with $T = 273$ K and beam power 200 mW. Note that all curves superimpose. (c) At 373 K, the variances along the collective coordinates is plotted against the decay time scale of the auto-correlation function of the position component along the collective coordinates. The logarithm is plotted in the inset for 600 mW beam power. (d) Plot of the slope of the linear fit of mode variances against projected position auto-correlation function time scales for various beam powers.

than any of the collective mode vibration frequencies. This demonstrates that the system can be studied using overdamped Langevin dynamics. In addition, we have computed the auto-correlation functions of the velocity components along each collective coordinate:

$$corr_i(t) = \frac{\langle \dot{x}_i(t)\dot{x}_i(0)\rangle}{\langle \dot{x}_i^2\rangle} \tag{5.4}$$

where $\dot{x}_i$ is the velocity component projected onto the i^{th} collective coordinate. The decay

time scale of the velocity auto-correlation is important because the simulation time step of the Langevin dynamics should be set much smaller. Fig. 5.4b shows the auto-correlation functions of $\dot{x}_i$ for the 6-particle triangle structure. The auto-correlation functions of the $\dot{x}_i$ for the single particle case is shown in the inset. The velocity auto-correlation functions of the collective coordinates vanish after 50 ns. From these calculations we infer that simulations of the system with 500 ns time steps sample the overdamped motions in the system.

The time dependence of an overdamped multi-dimensional harmonic oscillator with equilibrium position at the origin is described, using the overdamped Langevin equation, by

$$\ln \frac{\langle x_i(t)x_i(0)\rangle}{\langle x_i^2\rangle} = -\frac{Dt}{\sigma_i^2} = -\frac{t}{\tau_i} \tag{5.5}$$

(see Section 5.A). In Eq. 5.5, $D = k_BT/\zeta$ while $\tau_i = \sigma_i^2/D$ is the decay time of the auto-correlation function of the position coordinate along the i^{th} mode. We note that for harmonic oscillators the decay time scale is proportional to the mode variances and the proportionality constant is the diffusion coefficient.

Fig. 5.4c shows that Eq. 5.5 holds for the triangle structure of the 6-particle OM system at 273 K. The inset of Fig. 5.4c shows that the logarithms of the auto-correlation functions of the collective coordinates are linear for 200 mW beam power, while the main panel of Fig 5.4c shows that the relation between the mode variance and decay time is also linear for the OM triangle structure for various beam powers. Both of these results resemble those for the overdamped harmonic oscillator. Fig. 5.4d further supports the resemblance to the overdamped harmonic oscillator by showing that the slope of the fit of the mode variance against decay time is very close to the diffusion coefficient D for all the beam powers that we sampled. We infer, then, that the position auto-correlation function is very little affected by the non-conservative nature of the force experienced by the system. In fact, most of the non-conservative force is associated with the rotational motion of the entire OM structure (i.e., PC 12).

5.4.4 Power dissipation along the collective coordinates

The triangle OM structure is a NESS that requires power input to be maintained and therefore constantly dissipates energy to the medium. As discussed is association Fig. 5.1c, the mean power dissipation of the ordered NESS's of the OM system is significantly smaller than that of the disordered structures. We now examine the mean power dissipation along the collective coordinates defined by the PCs of the system.

The mean power dissipation along each of the first 11 collective coordinates was computed using the Stratonivich discretization:

$$\langle P_i \rangle = \frac{1}{N} \sum_{j=1} N F_i \left(\frac{\boldsymbol{x}^{(j)} + \boldsymbol{x}^{(j+1)}}{2} \right) \frac{x_i^{(j+1)} - x_i^{(j)}}{\Delta t} \qquad (i = 1 \sim 11) \tag{5.6}$$

where $\langle P_i \rangle$ is the mean power dissipation along the i^{th} collective coordinate, N is the total number of frames, j is the time (i.e., frame number) in the simulation trajectory, Δt is the simulation time step, $F_i(\boldsymbol{x})$ is the force component along the i^{th} collective coordinate analyzed at position $\boldsymbol{x}$, and x_i is its i^{th} component in the collective coordinate basis. A detailed description of the numerical procedure we have employed is presented in Section 5.D. Fig. 5.10 demonstrates that the mean power dissipation is not sensitive to the choice of simulation time step. Since the collective coordinate is generated from the trajectory data set with rotation angles aligned, the power dissipation along the rotation mode (#12) cannot be calculated directly, but it can be computed as the difference between the mean total power dissipation and the power dissipation along the first 11 collective coordinates:

$$
\begin{aligned}
\langle P_{12} \rangle &= \langle P_{tot} \rangle - \sum_{i=1}^{11} \langle P_i \rangle \\
&= \frac{1}{N} \sum_{j=1}^{N} \left[\boldsymbol{F} \left(\frac{\boldsymbol{x}^{(j)} + \boldsymbol{x}^{(j+1)}}{2} \right) \frac{\boldsymbol{x}^{(j+1)} - \boldsymbol{x}^{(j)}}{\Delta t} \right. \\
&\quad \left. - \sum_{i=1}^{11} F_i \left(\frac{\boldsymbol{x}^{(j)} + \boldsymbol{x}^{(j+1)}}{2} \right) \frac{x_i^{(j+1)} - x_i^{(j)}}{\Delta t} \right]
\end{aligned}
\tag{5.7}
$$

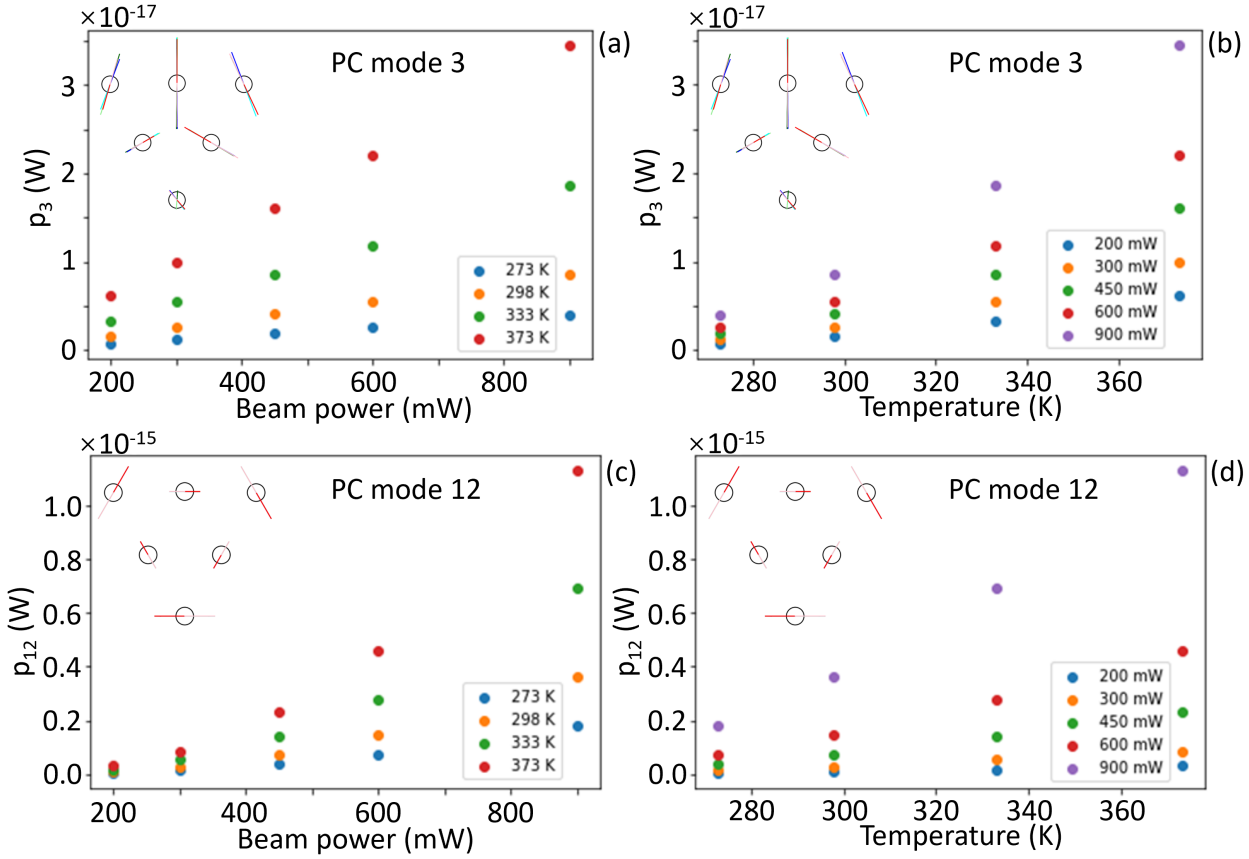

Figure 5.5: Plot of the mean power dissipation along the third collective coordinate of the triangle configuration: (a) as a function of beam power at various temperatures; (b) as a function of temperature for various beam powers. Plot of the mean power dissipation along the rotation coordinate of the triangle configuration: (c) as a function of beam power at various temperatures; and (d) as a function of temperature under various beam powers.

The results of these calculations are shown in Fig. 5.5. Among the 12 collective coordinates, PC 3 and PC 12 (rigid rotation) are chosen for further examination. See Fig. 5.11

and Fig. 5.12 in Section 5.D for the power dissipation components along all 12 collective coordinates. When beam power and temperature are increased, the mean power dissipation along all the collective coordinates increases. PC 3 is relevant to the transition between the triangle structure to the chevron structure, another NESS of the 6-particle OM system [1]. Indeed, the chevron is the most probable structure from transformations of the 6 particle triangle [2]. However, the mean power dissipation along the rotation coordinate is two orders of magnitudes larger than that along PC 3 and other coordinates shown in Fig. 5.11 and Fig. 5.12 in Section 5.D: that is power dissipation associated with the rigid rotation of the structure dominates over structural transitions, "vibrational" motions, and translation. Whereas the rotation motion (mode 12) is driven by a non-conservative force [17], mode 3-11 manifest a "housekeeping" power dissipation associated with maintaining the NESS in the presence of the random noise force of the thermal environment.

5.4.5 Total power dissipation and entropy production rate

The entropy production rate of a NESS can be computed from the Seifert relation [144]:

$$\begin{aligned} \dot{S}_m &= \frac{\langle \boldsymbol{F} \cdot \dot{\boldsymbol{x}} \rangle}{T} = \frac{1}{T} \int dx \boldsymbol{F}(\boldsymbol{x}) \cdot \boldsymbol{j}_s(\boldsymbol{x}) \\ \dot{S}_{tot} &= \frac{1}{T} \int dx \frac{\zeta j_s^2(\boldsymbol{x})}{p_s(\boldsymbol{x})} \end{aligned} \tag{5.8}$$

where $\dot{S}_m$ is the entropy production rate of the surrounding medium, $\dot{S}_{tot}$ is the total entropy production rate, $\boldsymbol{j}_s(\boldsymbol{x})$ is the steady state probability current density, and $p_s(\boldsymbol{x})$ is the steady state probability density [154, 155]. Since the probability density of a NESS is stationary, the difference between S_{tot} and S_m is the system entropy, which is equal to $k_B \int dx p_s \ln p_s$ and independent of time, so that $\dot{S}_m$ is equal to $\dot{S}_{tot}$:

$$\frac{\langle \boldsymbol{F} \cdot \dot{\boldsymbol{x}} \rangle}{T} = \dot{S} = \frac{1}{T} \int dx \frac{\zeta j_s^2(\boldsymbol{x})}{p_s(\boldsymbol{x})} \tag{5.9}$$

It can be shown that $p_s\boldsymbol{F} = k_BT\boldsymbol{\nabla}p_s + p_s\boldsymbol{F'}$, where $\boldsymbol{\nabla}\cdot\left(p_s\boldsymbol{F'}\right) = 0$, is a Helmholtz decomposition of the vector field $p_s\boldsymbol{F}$ and that $\boldsymbol{j}_s = p_s\boldsymbol{F'}/\zeta$. Here we define $\boldsymbol{F'}$ and $\boldsymbol{F}-\boldsymbol{F'} = k_BT\boldsymbol{\nabla}\ln p_s$ as the non-conservative and conservative parts of the force field, respectively. If we rewrite Eq. 5.9 using the relationship between $\boldsymbol{j}_s$ and $\boldsymbol{F'}$, we obtain:

$$\begin{aligned}\frac{\langle\boldsymbol{F}\cdot\dot{\boldsymbol{x}}\rangle}{T} = \dot{S} = \frac{\left\langle F'^2\right\rangle}{\zeta T} &= \frac{\left\langle\left\|\boldsymbol{F}-k_BT\boldsymbol{\nabla}\ln p_s\right\|^2\right\rangle}{\zeta T} \\ &\approx \frac{1}{\zeta T}\sum_{i=1}^{2N-1}\left\langle\left(F_i + k_BT\frac{x_i}{\sigma_i^2}\right)^2\right\rangle\end{aligned} \tag{5.10}$$

The detailed derivation of Eq. 5.10 is given in Section 5.B. For the single particle OM system, the $\boldsymbol{F'}$ obtained is equal to the azimuthal component of the force field, and the derivation is given in Section 5.C. We report the results of our calculations as the ratio between mean power dissipation $\langle\boldsymbol{F}\cdot\dot{\boldsymbol{x}}\rangle$ and $T\dot{S}$, which, if Eq. 5.9 is obeyed, should be unity.

Fig. 5.6 shows the ratio $z = \langle\boldsymbol{F}\cdot\dot{\boldsymbol{x}}\rangle/T\dot{S}$ computed for three OM systems with 1, 2, and 6 particles. The analysis of each case is based on a data set containing ten 10-million-frame trajectories where the time step between adjacent frames is 0.5 μs. All the trajectories start at the reference structure with all particles on trigonal lattice sites and lattice spacing equal to the laser wavelength in the medium (water). All the data points in Fig. 5.6 use the complete trajectories except the single particle case at 273 K and 200 mW. In this case, ten 11-million-frame trajectories are simulated and the last 10 million frames in each of the ten trajectories are taken as the data set for analysis in order to avoid bias introduced by the initial condition. See Fig. 5.13 in Section 5.D for detailed reasons for selection of this data set.

Fig. 5.6a shows the ratio $z = \langle\boldsymbol{F}\cdot\dot{\boldsymbol{x}}\rangle/T\dot{S}$ for the single-particle OM system. The value of z is unity within our estimated error for for all temperatures and beam powers we have sampled, in agreement with Seifert's analysis [155]. However, as is shown in Fig. 5.6b and

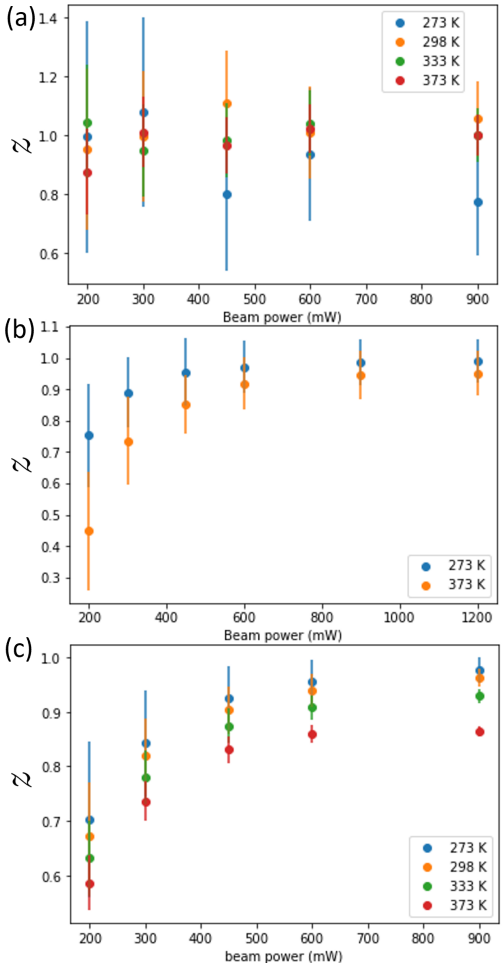

Figure 5.6: Plot of z, the ratio between mean power dissipation $\langle \boldsymbol{F} \cdot \dot{\boldsymbol{x}} \rangle$ and $T\dot{S}$ and temperature multiplied by entropy production rate $T\dot{S}$ for (a) single-particle OM system; (b) 2-particle OM system; (c) 6-particle triangle OM structure.

Fig. 5.6c, when $T > 273$ K, z is less than 1 in the 2-particle OM system and the 6-particle triangle OM structure for a considerable range of smaller beam power (less than about 600 mW). Although z does increase as the beam power increases, it remains less than unity when temperature is high or beam power is low, and we only find $T\dot{S}$ equal to the mean power dissipation (i.e., $z = 1$) at high beam power and low temperature. We provide a conjecture concerning the deviation of z from 1 in the next section.

5.5 Discussion

The results of our power dissipation and entropy production rate calculations along the collective coordinates of the (metastable) 6-particle OM NESS call for a nuanced interpretation of the Seifert relation. That equality, Eq. 5.9, between the power dissipated in maintaining a NESS and the rate of entropy production in that state is exact provided that the probability density distribution for the NESS is truly time-independent and the force acting on the NESS affects all particles in the configuration that define it. These conditions are apparently well defined and easily achievable for a single Brownian particle in a well-characterized external field. They are less precisely delineated when the NESS is defined by the criterion that its probability density distribution is time-independent, and the basic force-flux relations that maintain the NESS are, fundamentally, just conservation conditions. Interesting NESS's will, typically, have many particles in both ordered and disordered states, thereby requiring that calculation of the entropy production rate accurately account for the forces on every particle. The OM system considered in this paper satisfies this requirement.

A comparison of the principal findings from our simulation studies and the experimental results reported by Han and Yan [74] are of interest. We have shown that, on the time scale that the metastable 6-particle OM triangle can be treated as in a stationary state, it's internal motions can be represented with 12 orthonormal collective coordinates derived from a principal component analysis of the 12-particle trajectories. The power dissipation along the collective rotation coordinate is significantly larger than that along the other collective coordinates. This observation is consistent with the report by Han and Yan that the more ordered the OM structure the larger the angular velocity associated with its rigid body rotation [74]. They refer to the change between the ordered and disordered OM structures as a "phase transition" and use the angular velocity of the particle cluster as a parameter that identifies the "phase change". This choice of "phase change" order parameter ignores details associated with changes in participation of the internal motions that accompany changes

in OM structure. Our results agree with their work in that the triangle configuration is a metastable ordered OM NESS that dissipates the most power along rotation. In general, we expect that quantities such as power dissipation and entropy production rate will be useful tools for the theoretical characterization of the NESS's of optical matter. Of course, changes in entropy and chemical potential are the classic approach to describing phase transitions [64, 156], but they cannot be applied to the OM system since there is not a well-defined potential and entropy is constantly being produced.

Consider now the principal component-defined collective coordinates; these are determined by the conservative part of the force field as they only depend on the probability density distribution. Nevertheless, the non-conservative part of the force field still affects them by breaking the symmetry of the system. Since the non-conservative force intrinsically comes from the power input of the laser, scattered photons are of particular interest vis a vis the relationship between the collective coordinates and the non-conservative force. A study of the correlation between the collective coordinates and angular scattering modes [17] is a natural complement to the work reported here.

Power dissipation projected onto the collective coordinates are computed. It might seem intuitive that power dissipation projected onto the collective coordinate with larger variance will be larger. However, as is shown in Fig. 5.11 and Fig. 5.12 in Section 5.D, the power dissipation is smaller along collective coordinates 7 and 8 compared to that collective coordinates 9, 10, and 11, but collective coordinates 7 and 8 have larger variance compared to collective coordinates 9, 10, and 11. This is because it is the non-conservative part of the force field that contributes to the power dissipation, while the collective coordinates and their variances are determined by the conservative part of the force field. Whereas the rotation motion (mode 12) is driven by a non-conservative force [17], mode 3-11 manifest a "housekeeping" power dissipation associated with maintaining the NESS in the presence of the random noise force of the thermal environment.

5.6 Conclusions

The system we have studied is a 6-particle OM structure that supports several metastable ordered structures (triangle, chevron, ...) and exhibits transformations between those ordered structures [1, 2]. The proper NESS for this OM system, defined by the requirement that the probability density distribution is time-independent, is the ensemble of 6-particle structures including structural transformations. Even if the properties of an OM structure are sampled on a time scale shorter than its lifetime, the existence of leaks to other structures affects the phase space that is sampled.

Overall, we conclude that:

1. We single out for attention only one of the structures in the 6-particle OM aggregate, which generates a "metastable" NESS. The "metastable" NESS is stabilized by high beam power and low temperature. This observation is relevant because the strength of optical binding in each metastable NESS increases with increasing laser power, leading to increasing collective mode frequency with increasing incident power. When the power dissipation and entropy production rate are calculated as described in Section 5.7, with the approximation that the probability density distribution is Gaussian, we find that in the one particle system there is a deviation from $z = 1$ at low incident laser power and at high temperature that is attributable to inadequate sampling of the initial condition for the probability density function, specifically the deviation of the initial distribution of sample points from those of the true stationary probability density function.

2. The internal dynamics of the OM aggregate are strongly influenced by the non-electromagnetic interactions that act between the nanoparticles. In our case these are the electrostatic double layer interactions of the nanoparticles. We illustrate this dependence in Fig. 5.14 in Section 5.D. Since the double-layer electrostatic interaction plays an important role

in the entropy production rate calculation, analysis of experimental data from real OM systems in solution (i.e., overdamped conditions) may require more accurate treatment of the double-layer electrostatic interaction than is provided by the form used in our simulations [157]. Note however, that our choice of model does not affect our results since the simulation data are numerically accurate, but the model and parameter values do affect the stability of OM structures that form [2].

3. We have shown that the mean power dissipation and entropy production rate are generally equal to each other. However, the power dissipation is smaller for multi-particle cases where the Gaussian approximation is applied to probability distribution at high temperature or low beam power because the deviation of the OM structure from the reference structure is larger at high temperature and/or low beam power. The further the OM system fluctuates from the reference structure the larger the error of the Gaussian approximation. The reason we do not suffer from this issue for the single particle case is that the force decomposition can be carried out analytically (see Section 5.C). The entropy production rate calculated in the present paper is the total entropy production rate. Although in theory the entropy production rate in the medium is equal to the total entropy production rate for NESS's, the Gaussian approximation is the source of greater error in the calculation of the entropy production in the medium (see Fig. 5.16 in the Section 5.D).

An important goal is learning how to control the stability of the NESS's and the transition rates between them in multi-particle OM systems by tuning the parameters of the input laser beam. Generically, such control problems are very computation expensive because a simulation of the whole configuration space is needed for each set of control parameters. On the other hand, controlling some functions in the configuration space such as entropy production rate and scattering that are cheaper to compute, would allow searching for relationships that connect these quantities to NESS stability, and thereby new routes for control. We believe

that decomposition of the internal motions of OM systems, and better understanding of the rates of entropy production of those internal motions, will play an important role in those investigations. Apart from OM, other overdamped systems with non-conservative force field can also be studied using this strategy.

5.A The position auto-correlation function of overdamped multi-dimensional harmonic oscillators

The forces exerted along the principal directions of a multi-dimensional harmonic oscillator whose equilibrium position is at the origin of the Cartesian coordinate, are

$$F_i = k_i x_i \quad (i = 1, 2, ...) \tag{5.11}$$

where the k_i's are the force constants. The probability density of the particle positions at a given temperature T is:

$$p(\boldsymbol{x}, T) = \prod_i p_i(x_i, T) \tag{5.12}$$

where

$$p_i(x_i, T) = \sqrt{\frac{k_i}{2\pi k_B T}} e^{-\frac{k_i x_i^2}{2k_B T}} = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{x_i^2}{2\sigma_i^2}} \tag{5.13}$$

where σ_i^2 is the variance along the i^{th} principal direction. Then according to Eq. 5.13, we have:

$$k_i = \frac{k_B T}{\sigma_i^2} \tag{5.14}$$

The Langevin equation for an overdamped multi-dimensional harmonic oscillator is:

$$0 = -k_i x_i(t) - \zeta \frac{dx_i(t)}{dt} + f_i^{(b)} \tag{5.15}$$

where ζ is the friction coefficient and $f_i^{(b)}$ is the random bath force which is white (noise) and homogeneous in time. Multiplying both sides of Eq. 5.15 by $x_i(0)$ and taking the ensemble average, we have:

$$0 = -k_i \langle x_i(0) x_i(t) \rangle - \zeta \frac{d}{dt} \langle x_i(0) x_i(t) \rangle + x_i(0) \langle f_i^{(b)} \rangle \tag{5.16}$$

Since $f_i^{(b)}$ is a time-homogeneous white noise, $\langle f_i^{(b)} \rangle = 0$, so Eq. 5.16 can be rearranged to:

$$\frac{d}{dt}\langle x_i(0)x_i(t)\rangle = -\frac{k_i}{\zeta}\langle x_i(0)x_i(t)\rangle \tag{5.17}$$

the solution of which is

$$\ln \frac{\langle x_i(0)x_i(t)\rangle}{\langle x_i^2\rangle} = -\frac{k_i}{\zeta}t \tag{5.18}$$

Therefore, the decay time scale along the i^{th} principal direction can be defined as $\tau_i = \frac{\zeta}{k_i}$. According to Eq. 5.14 and the fluctuation-dissipation theorem [156], we have:

$$\ln \frac{\langle x_i(0)x_i(t)\rangle}{\langle x_i^2\rangle} = -\frac{Dt}{\sigma_i^2} = -\frac{t}{\tau_i} \tag{5.19}$$

where $D = \frac{k_B T}{\zeta}$ is the diffusion coefficient.

5.B Entropy production rate calculation

The probability density distribution of a system described by overdamped Langevin dynamics follows the Fokker-Planck equation:

$$\frac{\partial p}{\partial t} = \zeta^{-1} k_B T \triangle p - \zeta^{-1} \boldsymbol{\nabla} \cdot (p\boldsymbol{F}) \tag{5.20}$$

where $\boldsymbol{F}$ is the external force field. For a NESS, the steady state distribution can be defined:

$$p_s(\boldsymbol{x}) = \lim_{t\to\infty} p(\boldsymbol{x}, t) \tag{5.21}$$

Replacing p with p_s in Eq. 5.20 and rearranging the equation give:

$$0 = \frac{\partial p_s}{\partial t} = -\zeta^{-1} \boldsymbol{\nabla} \cdot [p_s \left(\boldsymbol{F} - k_B T \boldsymbol{\nabla} \ln p_s\right)] \tag{5.22}$$

Now let $\boldsymbol{F'} = \boldsymbol{F} - k_B T \boldsymbol{\nabla} \ln p_s$. Then according to Eq. 5.22, $\boldsymbol{\nabla} \cdot (p_s \boldsymbol{F'}) = 0$. In addition, since

$$\boldsymbol{\nabla} \times (p_s k_B T \boldsymbol{\nabla} \ln p_s) = k_B T \boldsymbol{\nabla} \times \boldsymbol{\nabla} p_s = 0 \tag{5.23}$$

then

$$p_s \boldsymbol{F} = p_s k_B T \boldsymbol{\nabla} \ln p_s + p_s \boldsymbol{F'} \tag{5.24}$$

is a Helmholtz decomposition of the force. The decomposition is unique. As a result, the steady state probability current density is:

$$\boldsymbol{j}_s = \zeta^{-1} p_s \boldsymbol{F} - \zeta^{-1} k_B T \boldsymbol{\nabla} p_s = \zeta^{-1} p_s \boldsymbol{F'} \tag{5.25}$$

When Eq. 5.25 is substituted in Eq. 5.7, we find

$$\dot{S} = \frac{1}{T} \int dx \frac{\zeta j_s^2}{p_s} = \frac{1}{T} \int dx \frac{F'^2 p_s}{\zeta} = \frac{\langle F'^2 \rangle}{\zeta T} \tag{5.26}$$

Since the steady state probability distribution is very close to a multi-dimensional Gaussian along the collective coordinates for the OM NESS,

$$\frac{\partial \ln p_s}{\partial x_i} \approx \sigma_i^{-2} x_i \tag{5.27}$$

where σ_i^2 is the variance along the i^{th} collective coordinate. Substitution of Eq. 5.27 together with the definition of $\boldsymbol{F'}$ back to Eq. 5.26 leads to:

$$\dot{S} = \frac{\langle F'^2 \rangle}{\zeta T} = \frac{1}{\zeta T} \sum_{i=1}^{2N-1} \left\langle \left(F_i + k_B T \frac{x_i}{\sigma_i^2} \right)^2 \right\rangle \tag{5.28}$$

where N is the number of particles and F_i is the external force component along the i^{th} collective coordinate.

5.C Single-particle force field decomposition

For a single-particle OM system, the force field $\boldsymbol{F}(\boldsymbol{x})$ is a 2-dimensional vector field defined on a 2-dimensional configuration space, so its Helmholtz decomposition can be accurately computed without using the Gaussian distribution approximation. Due to the cylindrical symmetry of the system, the steady-state probability density distribution $p_s(\boldsymbol{x})$ and the magnitude of the force field do not depend on the polar angle, so the radial component of the force field can be written as:

$$\boldsymbol{F}_r(r,\theta) = F_r(r)\hat{\boldsymbol{r}} = F_r(r)\cos\theta\hat{\boldsymbol{x}} + F_r(r)\sin\theta\hat{\boldsymbol{y}} \tag{5.29}$$

In Eq. 5.29, F_r is the magnitude of $\boldsymbol{F}_r$. $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$, and $\hat{\boldsymbol{r}}$ are the unit vectors along x, y, and radial directions, respectively. Likewise, the azimuthal component of the force field can be written as:

$$\boldsymbol{F}_a(r,\theta) = F_a(r)\hat{\boldsymbol{\theta}} = -F_a(r)\sin\theta\hat{\boldsymbol{x}} + F_a(r)\cos\theta\hat{\boldsymbol{y}} \tag{5.30}$$

Let $g_r(r) = p_s(r)F_r(r)$ and $g_a(r) = p_s(r)F_a(r)$, and we have:

$$\begin{aligned}
\boldsymbol{\nabla}\times(p_s\boldsymbol{F}_r) &= \frac{\partial}{\partial x}\left(g_r\sin\theta\right) - \frac{\partial}{\partial y}\left(g_r\cos\theta\right) \\
&= \left(\frac{\partial g_r}{\partial r}\frac{\partial r}{\partial x}\sin\theta + g_r\cos\theta\frac{\partial\theta}{\partial x}\right) \\
&\quad - \left(\frac{\partial g_r}{\partial r}\frac{\partial r}{\partial y}\cos\theta - g_r\sin\theta\frac{\partial\theta}{\partial y}\right) \\
&= 0
\end{aligned} \tag{5.31}$$

and

$$\begin{aligned}\boldsymbol{\nabla}\cdot(p_s\boldsymbol{F}_a) &= \frac{\partial}{\partial x}\left(-g_a\sin\theta\right)+\frac{\partial}{\partial y}\left(g_a\cos\theta\right)\\ &= \left(-\frac{\partial g_a}{\partial r}\frac{\partial r}{\partial x}\sin\theta - g_a\cos\theta\frac{\partial\theta}{\partial x}\right)\\ &+ \left(\frac{\partial g_a}{\partial r}\frac{\partial r}{\partial y}\cos\theta - g_a\sin\theta\frac{\partial\theta}{\partial y}\right)\\ &= 0\end{aligned} \tag{5.32}$$

Therefore, $\boldsymbol{\nabla}\times(p_s\boldsymbol{F}_r)=0$ and $\boldsymbol{\nabla}\cdot(p_s\boldsymbol{F}_a)=0$, so that

$$p_s\boldsymbol{F} = p_s\boldsymbol{F}_r + p_s\boldsymbol{F}_a \tag{5.33}$$

Eq. 5.33 is a Helmholtz decomposition of the vector field $p_s\boldsymbol{F}$. According to the uniqueness of Helmholtz decompositions as well as Eq. 5.24 and Eq. 5.26, we have:

$$\dot{S} = \frac{\langle F_a^2\rangle}{\zeta T} \tag{5.34}$$

Thus the entropy production rate of a single-particle OM system can be calculated by computing the mean square of the azimuthal component of the force field without the Gaussian distribution approximation.

5.D Supplementary Information

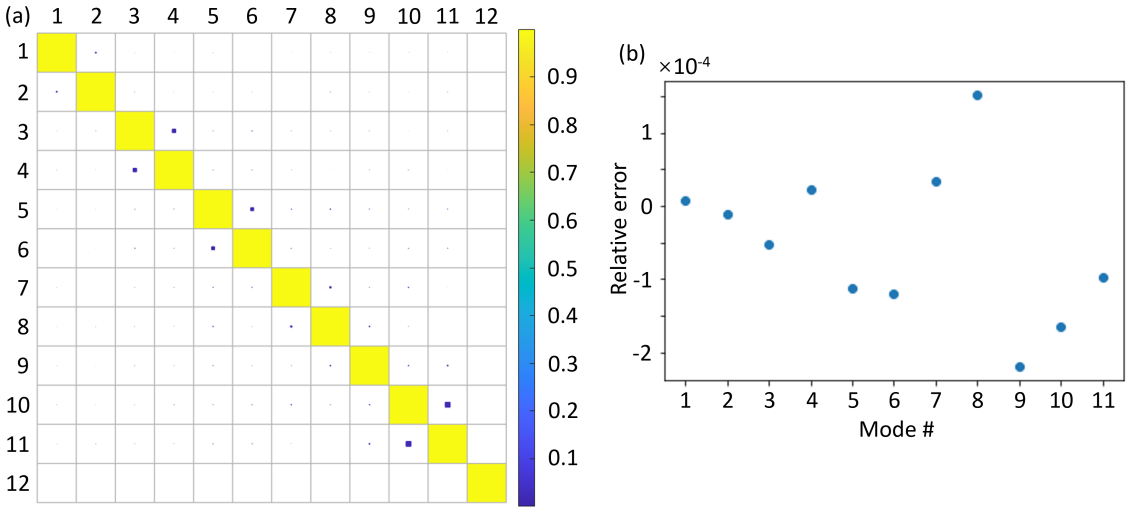

Figure 5.7: Convergence test of the PCs generated by Principle Component Analysis (PCA) applied to the rotationally aligned fluctuations of the nanoparticle constituents from electrodynamics - Langevin dynamics (EDLD) simulation trajectory data set of the triangle structure 6-particle OM system. (a) The norm of the inner product between right singular vectors (PCs) from PCA performed on a 10-million-frame (rows) and a 25-million frame (columns) trajectory. The diagonal nature of the matrix indicates both convergence and orthonormality of the numerically estimated PCs. (b) The relative difference between the variances of PCs 1-11 of the 10-million-frame (rows) and a 25-million frame (columns) trajectory. The very small relative errors provide further support of convergence of the numerically estimated PCs. The initial condition of the simulation is the triangle structure on trigonal lattice sites whose lattice spacing is the wavelength of the laser in water. Other parameters of the simulation are stated in the Method section of the main text.

From Fig. 5.7, we can see that when the electrodynamics - Langevin dynamics (EDLD) simulation trajectory length of the triangle structure 6-Ag nanoparticle optical matter (OM) system is larger than 10 million frames (i.e., time steps), both the principal components (PCs) and the variances along the PCs converge. Therefore, the PCs that have been determined and shown in Fig. 5.2 of the main text are numerically meaningful.

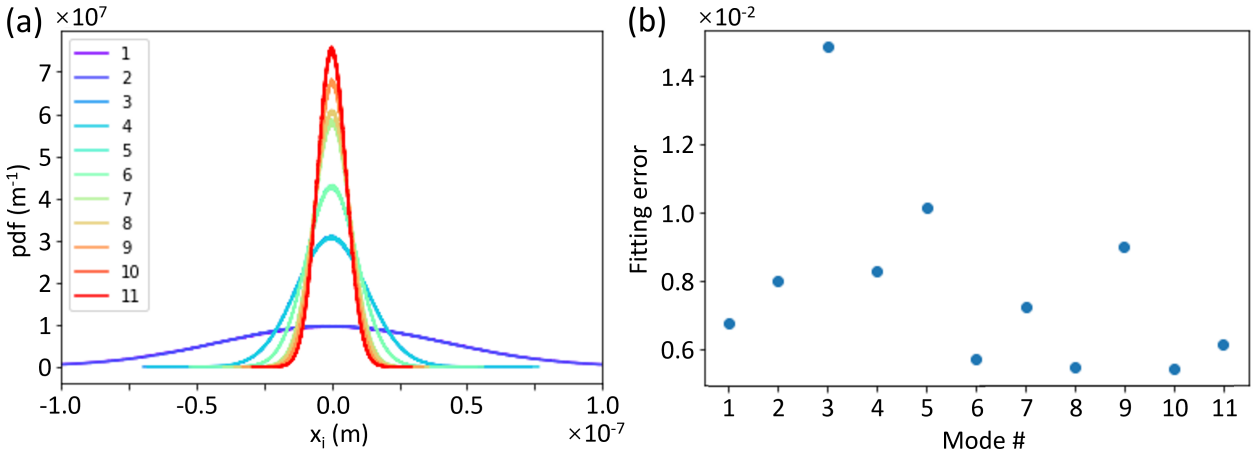

Figure 5.8: Projection of the EDLD simulation trajectory of the 6-particle triangle OM structure at 298 K and 600 mW beam power into the basis of the numerically estimated PCs. (We omit PC 12 corresponding to rigid rotation from this analysis.) (a) The marginal probability density distribution of the collective coordinates. (b) The fitting error of the marginal probability density distributions compared to a best fit Gaussian distribution. The fitting error is $\int \left| p_i - p_i^{\text{Gaussian fit}} \right| dx_i$. The results demonstrate that each projected mode follows an approximately Gaussian distribution.

Fig. 5.13 shows the analysis of entropy production rate times temperature, $T\dot{S}$ and mean power dissipation $\langle P \rangle$ for different 1-million-frame segment in the EDLD simulation trajectory for 1-particle and 2-particle cases. For both cases, 100 10-million-frame trajectories are simulated and each trajectory is divided into 10 1-million-frame segments labeled 1 to 10 according to the order in the trajectory. The segments with the same label are collected together for the analysis. Fig. 5.13a,b show that $T\dot{S}$ and $\langle P \rangle$ of the first 1-million-frame segment deviates from the rest, meaning that there is an initial condition bias in this segment. Therefore, the 1-particle case analysis in the main text is performed using the last 10 million frames of 11-million-frame trajectories. On the other hand, this bias is not important for the 2-particle case shown by Fig. 5.13c and 5.13d , so the data analysis for the 2-particle OM system is based on 10-million-frame trajectories that start at the reference structure. Other simulation parameters are the same as Fig. 5.11.

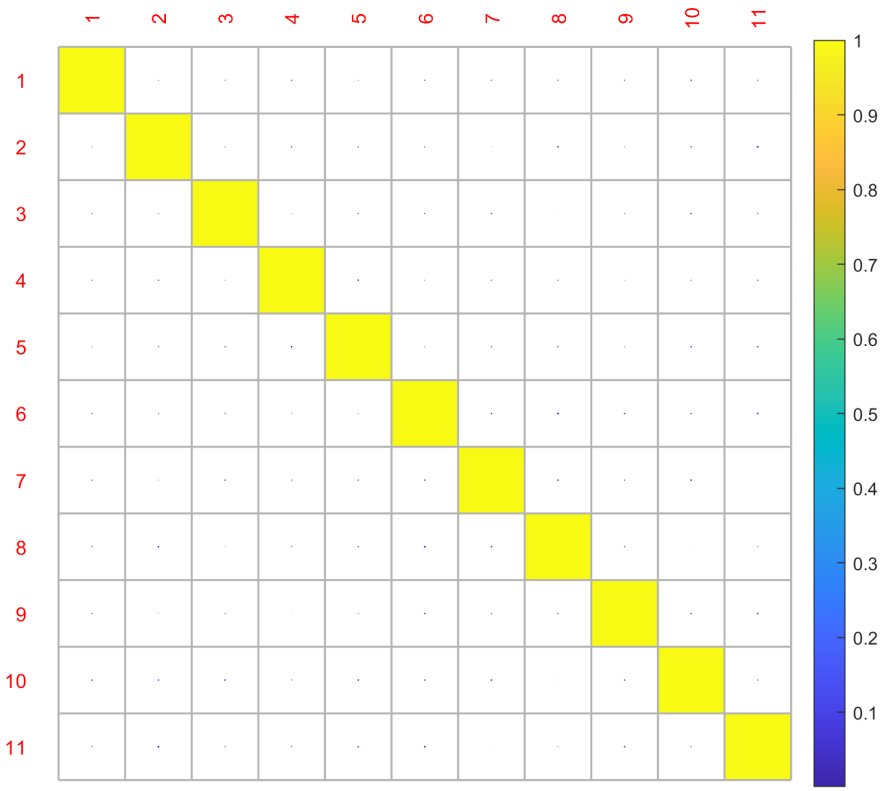

Figure 5.9: The pairwise correlation plot of the PCs for the 6-particle triangle OM structure at 298 K and 600 mW beam power. The diagonal nature of the plot indicates the orthonormality of the PC modes. (We omit PC 12 corresponding to rigid rotation from this analysis.) The initial condition of the simulation is the triangle structure on trigonal lattice sites whose lattice spacing is the wavelength of the laser in water. Other parameters of the simulation are stated in the Method section of the main text.

Fig. 5.16a and 5.16b show that a very small error introduced by the linear approximation of the conservative part of the force field, $\boldsymbol{F}_0$, will cause a large error in the medium entropy production rate, $\dot{S}_m$, for the single particle OM system in a Gaussian beam. This means that the computation of $\dot{S}_m$ is sensitive to the error of $\boldsymbol{F}_0$. The sensitivity of $\dot{S}_m$ compared to the total entropy production rate $\dot{S}_{tot}$ is analytically shown below. Without loss of generality, we can assume that the particle is on x-axis due to the cylindrical symmetry of the force

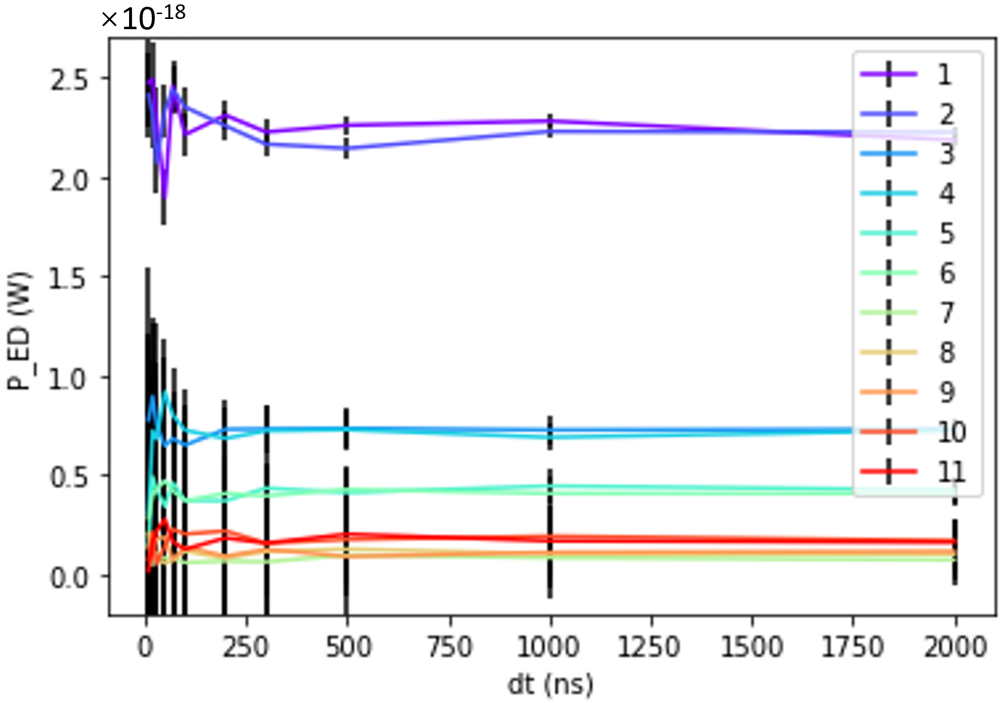

Figure 5.10: The mean power dissipation components $P_{ED,i}$ where i=1-11 are the first 11 PC modes, computed using Stratonovich discretization, along the collective coordinates for the 6-particle triangle OM structure. Each colored result is obtained at 298 K and 600 mW beam power and shown as a function of simulation time step. Other EDLD simulation parameters are the same as Fig. 5.9. The calculated power dissipation components are quite insensitive to the choice of simulation time step.

field. As shown in Section 5.C in the main text, the conservative part of the force, $\boldsymbol{F}_0$, is along the x-axis while the nonconservative part of the force, $\boldsymbol{F}'$, is along the y-axis. Then $T\dot{S}_m$ and $T\dot{S}_{tot}$ can be written as:

$$T\dot{S}_m = \frac{\boldsymbol{F}\cdot\boldsymbol{F}'}{\zeta} = \frac{F_y^2}{\zeta}, \qquad T\dot{S}_{tot} = \frac{\boldsymbol{F}'\cdot\boldsymbol{F}'}{\zeta} = \frac{F_y^2}{\zeta} \tag{5.35}$$

where:

$$\boldsymbol{F} = \boldsymbol{F}_0 + \boldsymbol{F}', \qquad \boldsymbol{F}_0 = F_x\hat{\boldsymbol{x}}, \qquad \boldsymbol{F}' = F_y\hat{\boldsymbol{y}} \tag{5.36}$$

, and ζ is the friction coefficient. Then linear approximation is applied to the calculation of $T\dot{S}_m$ and $T\dot{S}_{tot}$:

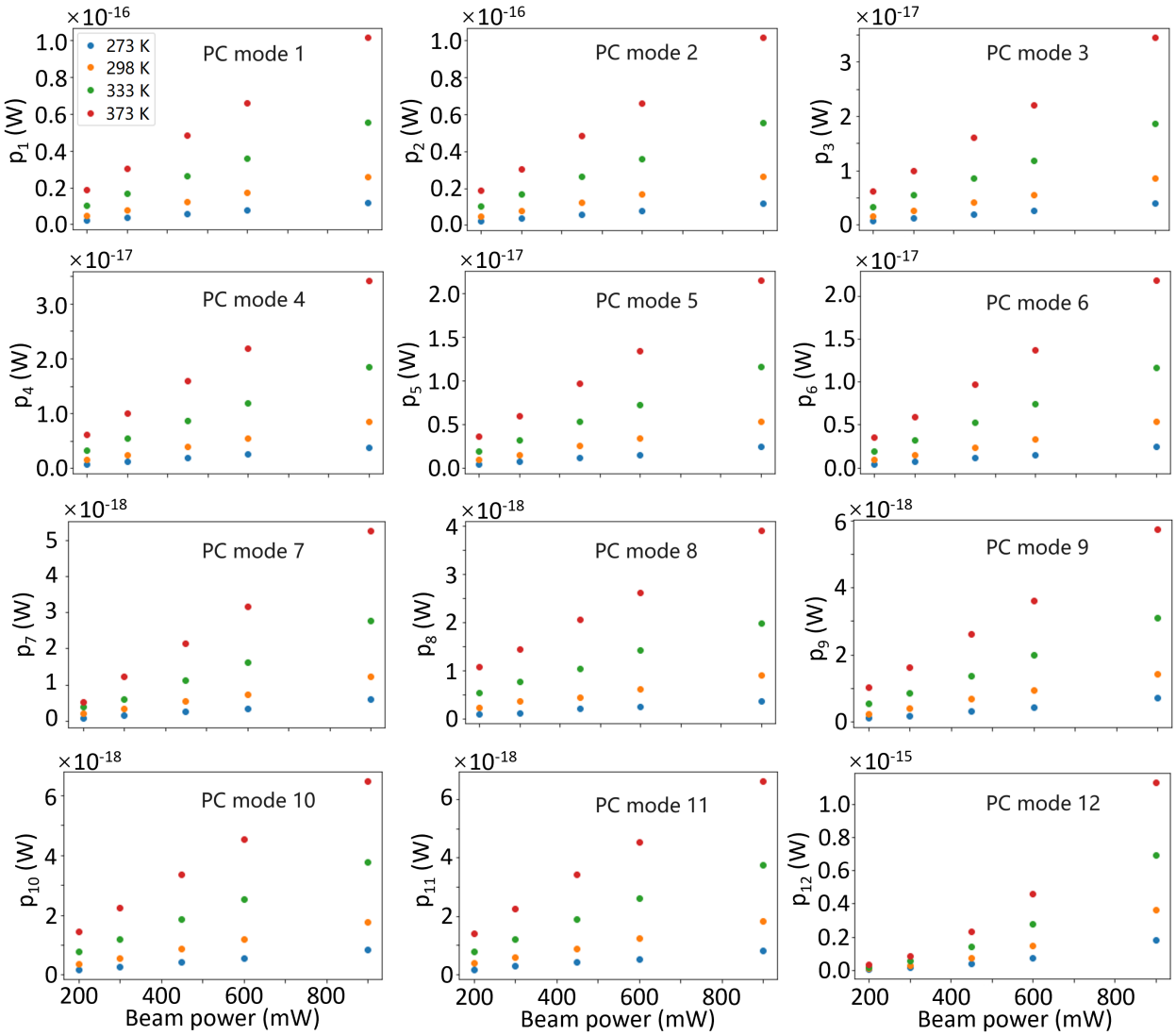

Figure 5.11: Plot of the mean power dissipation components along each collective coordinate of the 6-particle triangle OM structure as a function of beam power at various temperatures. The initial condition of the simulation is the triangle structure on trigonal lattice sites whose lattice spacing is the wavelength of the laser in water. Other parameters of the simulation are stated in the Method section of the main text. The scale of the power dissipation along PC mode 12 is 1 to 3 orders of magnitude larger than the power dissipation along the rest of the modes.

$$T\dot{\tilde{S}}_m = \frac{\boldsymbol{F}\cdot\tilde{\boldsymbol{F}}'}{\zeta} = \frac{F_x(F_x+kx)+F_y^2}{\zeta} = T\dot{S}_m + \frac{F_x(F_x+kx)}{\zeta} \tag{5.37}$$

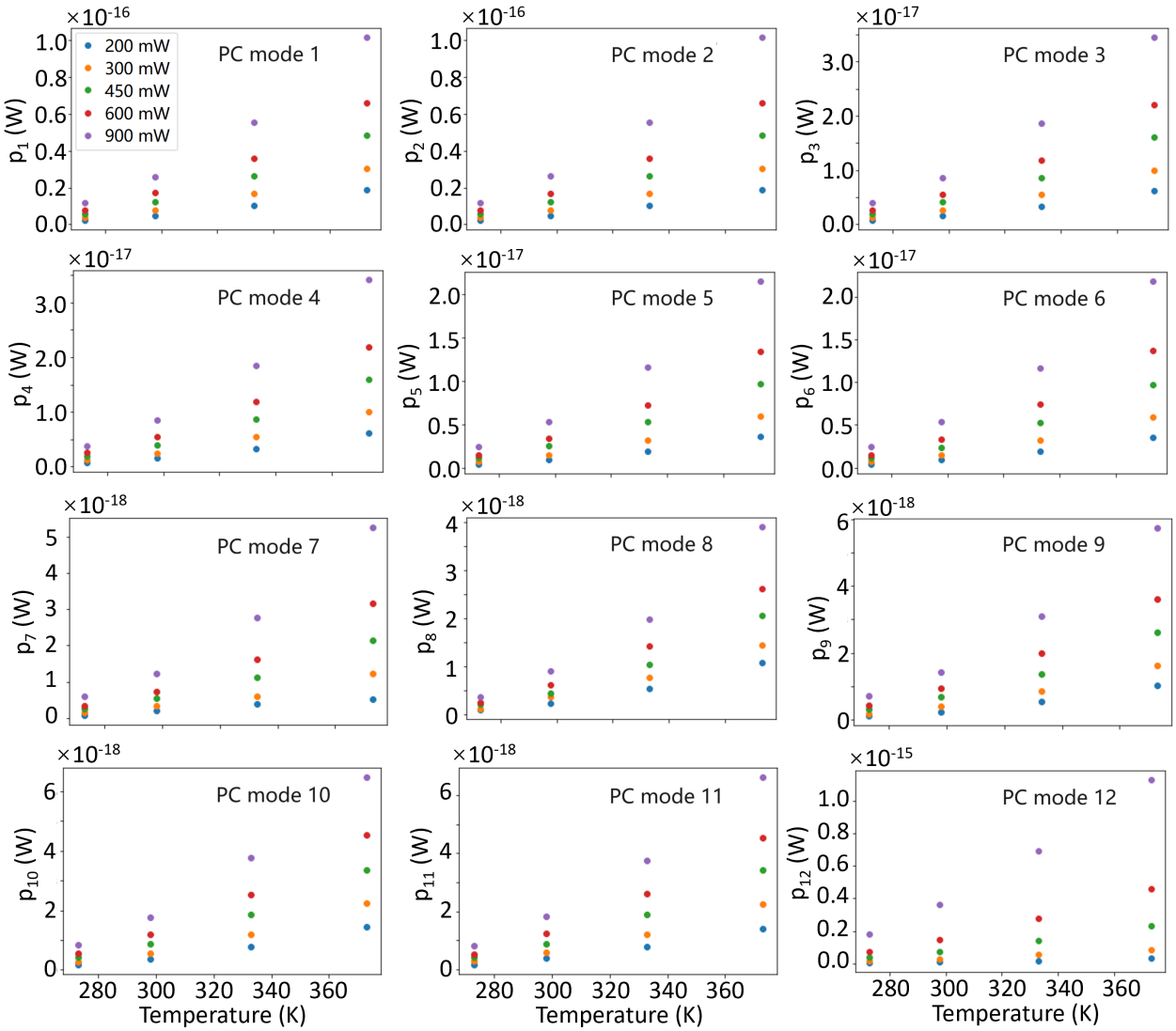

Figure 5.12: Plot of the mean power dissipation components along each collective coordinate of the 6-particle triangle OM structure as a function of temperature under various beam powers. The initial condition of the simulation is the triangle structure on trigonal lattice sites whose lattice spacing is the wavelength of the laser in water. Other parameters of the simulation are stated in the Method section of the main text. The scale of the power dissipation along PC mode 12 is 1 to 3 orders of magnitude larger than the power dissipation along the rest of the modes.

$$T\dot{\tilde{S}}_{tot} = \frac{\tilde{\boldsymbol{F}}' \cdot \tilde{\boldsymbol{F}}'}{\zeta} = \frac{(F_x + kx)^2 + F_y^2}{\zeta} = T\dot{S}_{tot} + \frac{(F_x + kx)^2}{\zeta} \tag{5.38}$$

where:

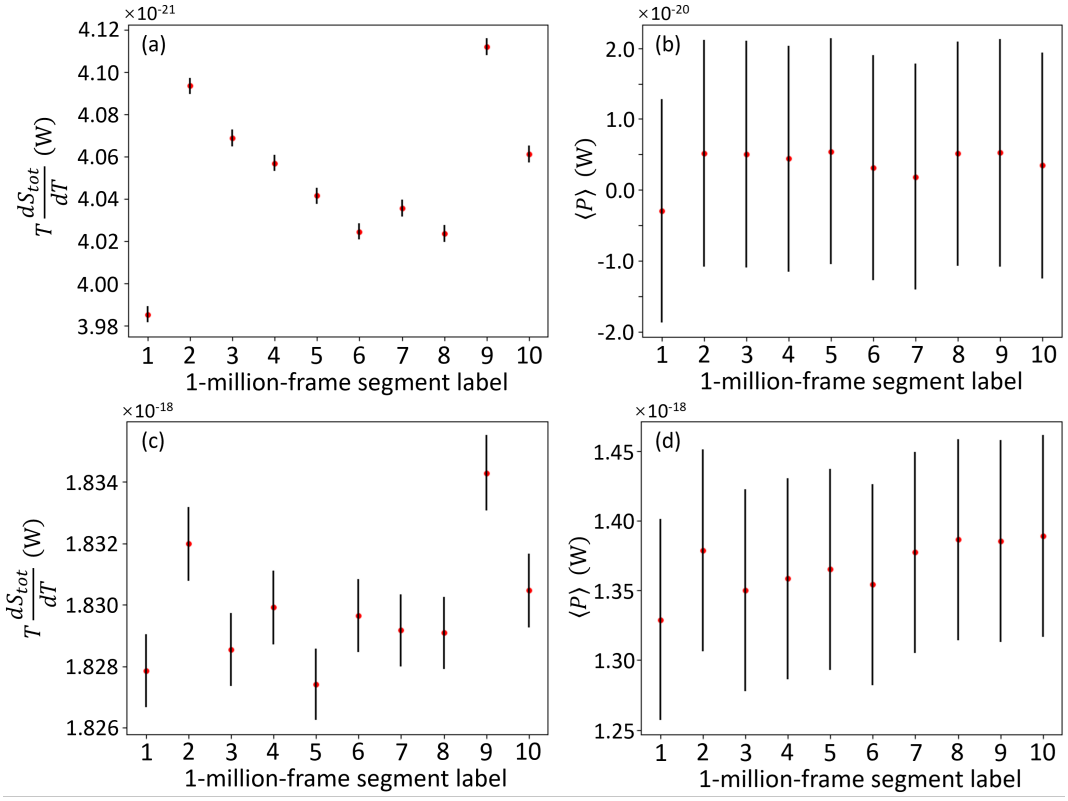

Figure 5.13: Plot of entropy production rate times temperature, $T\dot{S}$, and mean power dissipation, $\langle P \rangle$, as a function of 1-million-frame segments of 10-million-frame trajectories at 273 K and 200 mW beam power. (a) $T\dot{S}$ of 1-particle case. (b) $\langle P \rangle$ of 1-particle case. (c) $T\dot{S}$ of 2-particle case. (d) $\langle P \rangle$ of 2-particle case. The mean entropy production rate of the first 1-million-frame segment deviates from the those of the rest 1-million-frame segments, so the mean entropy production rate analyzed the main text is computed from the trajectory without the first 1-million-frame segment.

$$\boldsymbol{F} = \tilde{\boldsymbol{F}}_0 + \tilde{\boldsymbol{F}}', \qquad \tilde{\boldsymbol{F}}_0 = -kx\hat{\boldsymbol{x}}, \qquad \tilde{\boldsymbol{F}}' = (F_x + kx)\hat{\boldsymbol{x}} + F_y\hat{\boldsymbol{y}} \tag{5.39}$$

We can see from Eq. 5.37 and Eq. 5.38 that, when calculated using linear approximation, the error term of the medium entropy production rate is the first order term of the error in the calculation of force $F_x + kx$, while that of the total entropy production rate is the second order term. Therefore, compared to $\dot{S}_{tot}$, $\dot{S}_m$ is more sensitive to the force error introduced in the linear approximation.

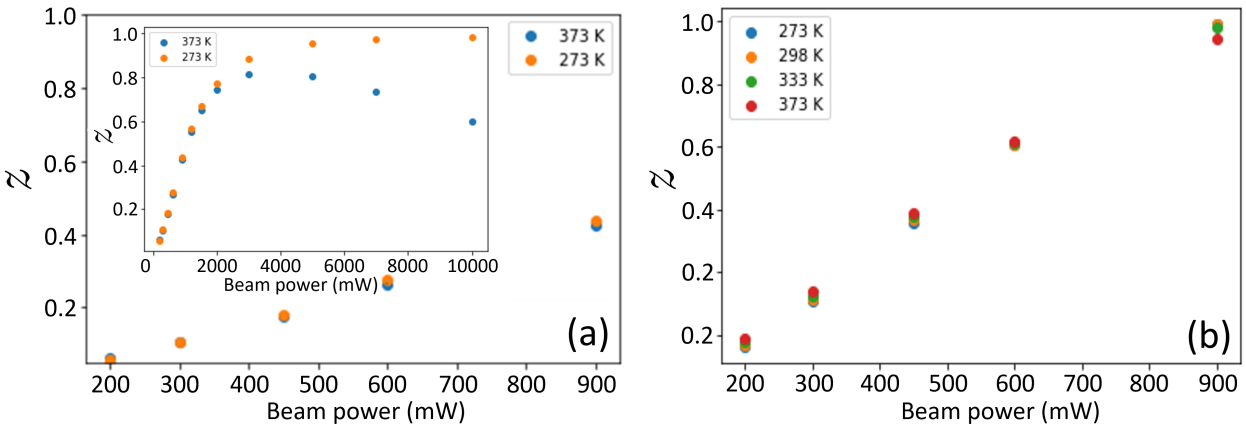

Figure 5.14: Plot of z, the ratio between mean power dissipation, $\langle \boldsymbol{F} \cdot \dot{\boldsymbol{x}} \rangle$, and temperature multiplied by entropy production rate, $T\dot{S}$, without taking double-layer electrostatic interaction into consideration for (a) 2-particle OM system; (b) 6-particle triangle OM structure. The initial condition of the simulation is the triangle structure on trigonal lattice sites whose lattice spacing is the wavelength of the laser in water. Other parameters of the simulation are stated in the Method section of the main text. The double-layer electrostatic interaction has a large influence on the entropy production rate calculation, so it cannot be neglected.

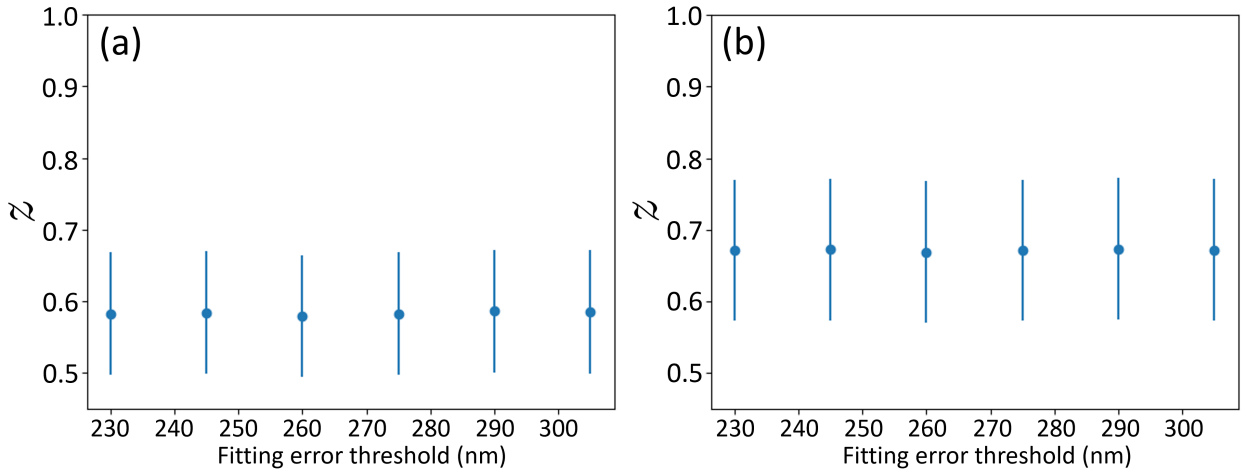

Figure 5.15: Plot of z, the ratio between mean power dissipation, $\langle \boldsymbol{F} \cdot \dot{\boldsymbol{x}} \rangle$, and temperature multiplied by entropy production rate, $T\dot{S}$, for the 6-particle triangle OM structure at 298 K and 200 mW beam power; (a) medium entropy production rate; (b) total entropy production rate. The initial condition of the simulation is the triangle structure on trigonal lattice sites whose lattice spacing is the wavelength of the laser in water. Other parameters of the simulation are stated in the Method section of the main text. The choice of the fitting error threshold in the local fluctuation simulation does not affect the entropy production rate calculation.

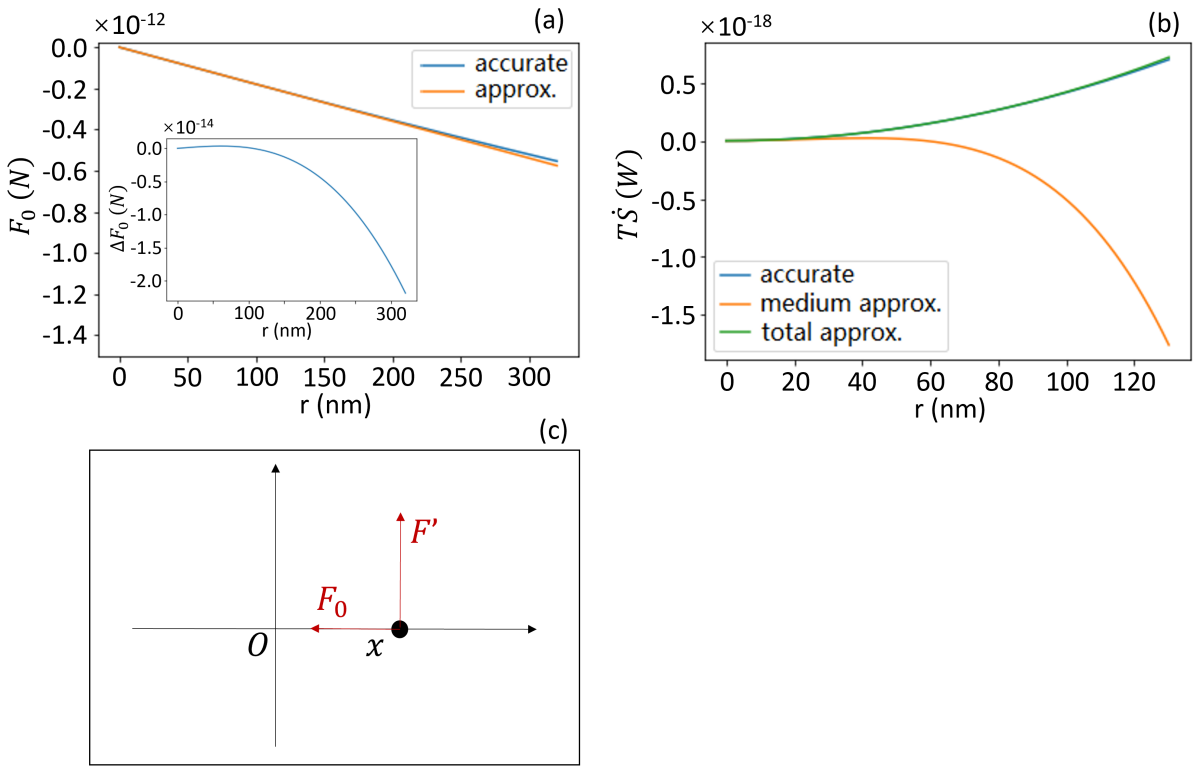

Figure 5.16: Support for the claim that the difference between the power dissipation and $T\dot{S}$ computed for a 6-nanoparticle OM is caused by the error of the linear approximation of the conservative part of the force field $\boldsymbol{F}_0$. (a) Plot of the accurate $\boldsymbol{F}_0$ and the linearly approximated $\boldsymbol{F}_0$ for a single Ag nanoparticle for 600 mW Gaussian beam. (b) Plot of the accurate $T\dot{S}$ as well as the medium and total $T\dot{S}$ computed at specific positions using a linearly approximated $\boldsymbol{F}_0$ against the radial distance for the single particle OM for 600 mW Gaussian beam. (c) Sketch of the force decomposition in the single particle case. The computation parameters of the force calculation are the same as those of dynamic simulations, stated in the Method section of the main text.

CHAPTER 6

RAMAN EFFECT-INSPIRED INSIGHTS INTO COLLECTIVE FLUCTUATION MODE-DEPENDENT LIGHT SCATTERING OF OPTICAL MATTER SYSTEMS

In this chapter, sections 6.1 to 6.7 are based on the following manuscript under review: Shiqi Chen, Andrew L. Ferguson, Stuart A. Rice, Norbert F. Scherer. "Raman Effect-Inspired Insights into Collective Fluctuation Mode-Dependent Light Scattering of Optical Matter Systems" *J. Phys. Chem. C*, 2024, under review.

6.1 Abstract

Raman scattering is widely used to study *underdamped* vibrational normal modes of a molecule, their intra-molecular potentials and molecular structures via the change of polarizability associated with structural deformation. Inspired by the classical Raman effect, this paper presents the relationship of the symmetry of *overdamped* collective fluctuation modes (analogous to underdamped normal modes) of optical matter (OM) arrays to their coordinate-dependent changes in induced-polarization and light scattering. OM systems consist of (nano-)particles that form 2-dimensional ordered (often trigonal lattice) structures due to electromagnetic interactions in coherent fields. The interactions among the particles are mediated by incident and scattered coherent light. While the influence of incident light on the dynamics of the OM system has been explored and the relationships of electrodynamic interference and induced-polarization to the scattered light have been studied, there is less appreciation of the relationship of the scattered light and the nanoparticle dynamics of OM structures. We demonstrate a homomorphism between the monotonic fluctuation mode-dependence of the scattering properties of OM systems and the monotonic coordinate-dependent polarizability of Raman active modes. We show that the induced-polarization and

total scattering cross-section changes monotonically along modes with high symmetry but non-monotonically along lower symmetry modes. Furthermore, the broken symmetry due to the chirality of the incident circularly polarized light vs. the rotationally symmetric trigonal lattice structure(s) can be described by changes in the irreducible representations such that some "Raman inactive" modes become active. Our analysis opens new opportunities to explore dynamics of N-body systems.

6.2 Introduction

Determination of intra-molecular potentials and structure from internal motions has long been exemplified by infra-red absorption and Raman scattering spectroscopies [158, 159]. These spectroscopic measurements report a response of the coordinate-dependent internal (e.g. vibrational) motions of the system in some basis such as normal modes to the applied field. Inspired by classical Raman scattering theory (more precisely elastic Rayleigh scattering) [158, 159], this paper develops a symmetry-dependent analysis for overdamped collective mode internal motions in electrodynamically bound optical matter (OM) systems [6, 105–107]. OM is a particle assembly created when polarizable (nano-)particles trapped in a coherent electromagnetic (laser) beam [6, 105–107] having electrodynamic interactions mediated by incident and scattered coherent light [9] between the particles that are stronger than random (e.g., thermal) forces of the environment. The OM assemblies can be ordered (i.e., rigid body-like) or disordered. To date, most studies of optical matter have focused on the influence of incident light on the stability of and/or the transitions between small ordered OM structures (i.e., isomerizations) without detailed analysis of possible relationships between the particle dynamics and light scattered by the OM structure [3]. Here we demonstrate that the elastically scattered light contains information about the internal motions of the OM structures and develop a group theoretic analysis thereof.

The classical Raman effect involves inelastic scattering of light due to its coupling with

changes in, e.g., the vibrational modes of an illuminated molecule or crystalline material [160]. Information about the symmetry of the molecular vibrational modes or the phonon modes of a crystal can be obtained by analyzing the scattered light [160]. One of the important features of Raman active modes is that the polarizability of the molecule or the crystal changes when it deforms along the mode [161]. For example, both the A_{1g} mode (D-band) and the E_{2g} mode (G-band) in graphene shown in Figures 6.1a-c are Raman active [162]. Furthermore, the helicity (i.e., left- or right-hand circular polarization) of the incident and scattered light involved in Raman scattering is particularly useful for studying 2-dimensional (or layered) materials [163, 164]. The helicity of the circularly polarized scattered light in helicity-resolved Raman scattering [162,165,166] does not change for some phonon modes but changes for others, hence can be used to determine the symmetry and polarization changes corresponding to the phonon modes in 2-D materials [167]. For example, the handedness of the polarization of helicity-resolved Raman scattered light of the A_{1g} mode (D-band) of graphene is the same as the input light polarization while the handedness of scattered light polarization of the E_{2g} mode (G-band) is the opposite to the input light [162].

The OM system studied in this paper consists of silver nano-particles in an experimental or simulated Gaussian optical trapping beam. The individual nano-particle motions are overdamped because they are suspended in water or simulated using the frictional damping thereof. Three experimental examples are shown (as dark-field microscopy images averaged over 157 frames) in Fig. 6.1d-f termed hexagon, triangle, and chevron, formed in a focused Gaussian laser (i.e., trapping) beam [1,3,17]. The electrodynamic interactions that bind the nano-particles together arise from the light scattered between and the induced-polarization created in each particle [3,9,168]. The momentum change between the incident and scattered light and conservation of momentum can cause the overdamped OM structure to rotate [15, 17].

Previous studies of the far-field scattering and induced-polarization of OM systems only

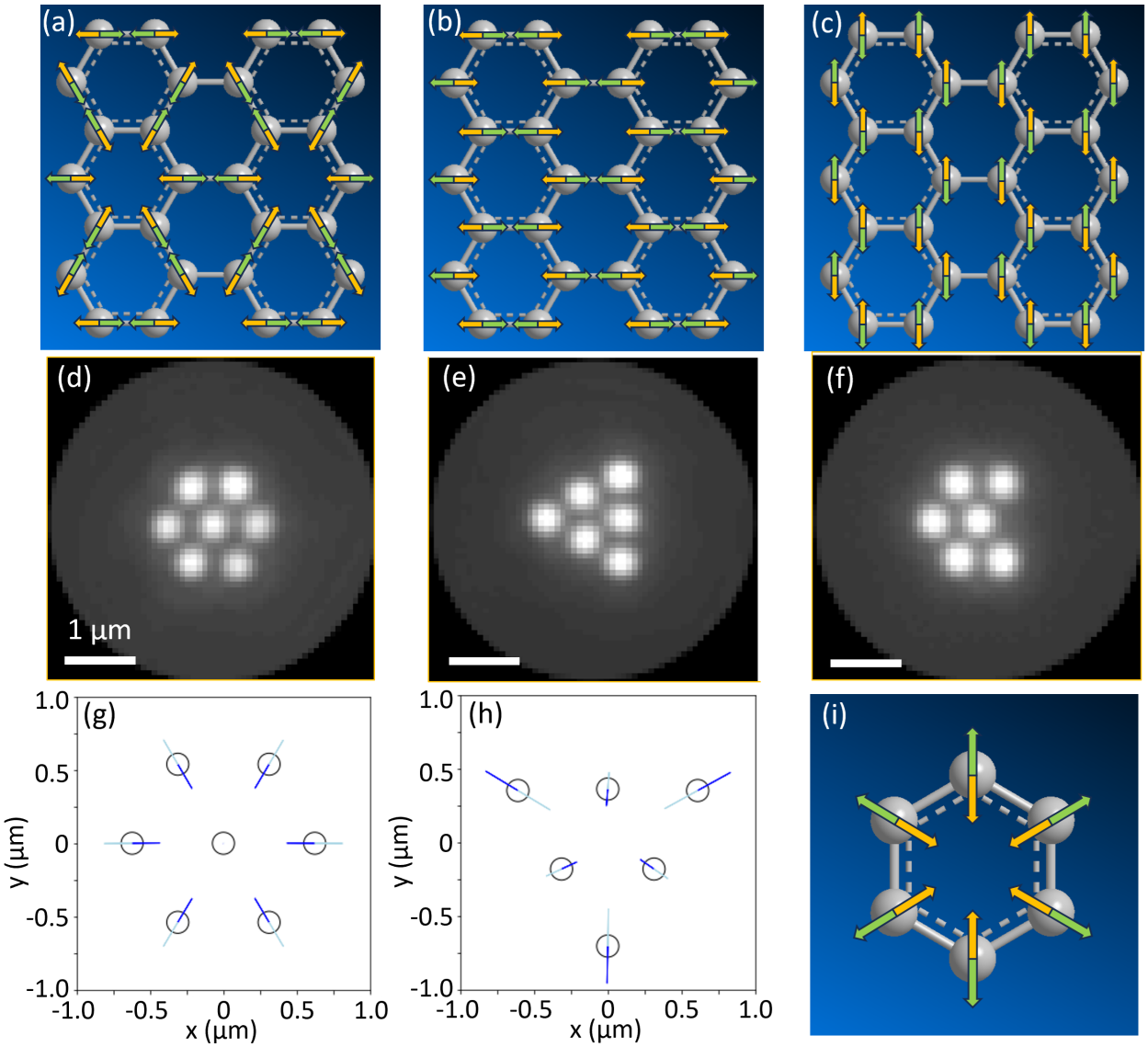

Figure 6.1: Representative vibrational and fluctuation modes with high symmetry. (a-c) Raman active phonon modes of graphene corresponding to the D_{6h} point group representations (a) A_{1g} and (b-c) E_{2g}. (d-f) Dark-field microscopy images aligned and averaged over 157 frames for (d) 6-fold symmetric, (e) triangle, and (f) chevron OM structures. (g) The breathing fluctuation mode of the 7-particle 6-fold symmetric OM structure corresponding to D_6 point group representation A_1. (h) One of the three-fold symmetric fluctuation modes of the 6-particle triangle OM structure corresponding to D_3 point group representation A_1. (i) The Raman-active breathing normal mode (corresponding to D_{6h} point group representation A_{1g}) of a planar hexagonal molecule with 6-fold rotational symmetry such as a united-atom model of benzene. Panels (d-f) reproduced with permission from ref. [3]. Copyright American Chemical Society.

considered the nearest neighbor spacing of the particle constituents as variables to study the behavior of the scattering cross-section and the induced-polarization [3]. However, OM systems are dynamic, manifesting structural fluctuations arising from the thermal energy of their envrionment. The collective particle fluctuation modes of "stable" OM structures can be obtained by principal component analysis (PCA); these PC modes resemble the normal modes of molecules but are appropriate for the overdamped (single and) collective particle motions [1]. For small structurally well-defined OM arrays, it is more intuitive to make analogies between the Raman active modes of molecules and the OM collective fluctuation modes with high symmetry. As shown in Fig. 6.1i, the symmetric breathing mode for a hexagonal molecule with six-fold rotational symmetry (e.g. a united atom model of benzene) is Raman active because the polarizability of the molecule changes when it deforms along this mode, a property of Raman active modes. Therefore, the symmetric breathing mode of the 7-particle 6-fold symmetric OM structure shown in Fig. 6.1g is of interest.

In this paper, we conduct numerical simulations to study small OM structures with various symmetries. Since the OM structure scatters light from its polarizable particle constituents (vis a vis the induced-polarizations and collective scattering modes [3,17]) , the present paper establishes that the symmetry analysis of the vibration modes used in Raman spectroscopy can be adopted to study the PC mode-dependent scattering of stable 2-D OM structures. To be clear, the present work is not about Raman enhancement in the near-field of nanostructures [169,170], but developing the analogy of light scattering from the collective modes of optical matter structures themselves to the classical Raman effect. We demonstrate that *the homomorphism holds when the induced-polarization changes monotonically along particular collective modes*. For stable OM structures with lower symmetry such as the 6-particle triangle structure shown in Fig. 6.1h, which does not have a symmetric breathing mode as a collective fluctuation PC mode, there exists a three-fold symmetric collective mode with a non-zero derivative of the induced-polarization. The 6-particle chevron structure is an

example of an OM structure with lower symmetry, and therefore less contrasting symmetry properties, that is presented in Section 6.A. Finally, recognizing that the chirality of the polarization state of the incident light causes symmetry breaking, reveals that the perturbed OM fluctuation modes along which the derivative of their induced-polarization is non-zero correspond to the same irreducible representation of the structural symmetry point group as the Raman active modes that preserve the polarization of light in helicity-resolved Raman scattering [165–167].

6.3 Methods

Electrodynamics-Langevin dynamics (EDLD) simulations were performed with the electrodynamic forces calculated by Generalized Multiparticle Mie Theory [37,38] using the MiePy software developed by the Scherer Lab [17,39]. Full details of the simulation approach have been reported [171]. Briefly, the EDLD solver performs numerical integration of the following Langevin equation:

$$m\frac{d^2\boldsymbol{x}}{dt^2} = \boldsymbol{F}_{ED}(\boldsymbol{x}) - \boldsymbol{\nabla}U_{DL}(\boldsymbol{x}) - \zeta\frac{d\boldsymbol{x}}{dt} + \boldsymbol{f}^{(b)} \tag{6.1}$$

In Eq. 6.1, m is the mass of the particle; $\boldsymbol{x}$ is the Cartesian coordinate of the structure; t is time; $\zeta = 6\pi\eta r$ is the friction coefficient for particle radius $r = 75$ nm; η is the viscosity of the medium (taken to be water); $\boldsymbol{f}^{(b)}$ is the bath random force term defined to be white noise satisfying the fluctuation dissipation relation at a specified temperature; $\boldsymbol{F}_{ED}$ is the sum of external electrodynamic force fields computed by the Maxwell stress tensor using MiePy [39]. The simulation time step between a pair of frames is 0.5 μs. U_{DL} is the double layer electrostatic interaction [40]:

$$U_{DL}(\boldsymbol{x}) = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} V(r_{ij})$$
$$V(r_{ij}) = 32\pi\varepsilon_0\varepsilon_m \left(\frac{k_B T}{z_p e}\right)^2 \tanh^2\left(\frac{z_p e\psi}{4k_B T}\right) \times \exp\left[-\kappa\left(r_{ij} - 2r\right)\right] \tag{6.2}$$

where N is the number of particles; ε_m is the medium electric permittivity; $z_p = 1$ is the ion valence; $\psi = -77$ mV is surface electrostatic potential; $\kappa^{-1} = 100$ nm is the Debye screening length; r is the particle radius; $r_{ij} = \left|\boldsymbol{x}_i - \boldsymbol{x}_j\right|$ is the distance between the i^{th} and j^{th} particle. The Debye screening length was determined in a previous simulation experimental comparison [2].

6.4 Results

6.4.1 Collective fluctuation modes and symmetry of the triangular OM structure

In this section, we show the 12 collective fluctuation modes of the planar 6-particle triangle OM structure and the influence of the optical chiral broken symmetry on them.

Fig. 6.2 shows the collective fluctuation modes computed for the triangle structure, one of the stable structures of the 6-particle OM system. The detailed algorithm for generating the collective fluctuation modes is the same as that used in our previous publication [1]: they are obtained by carrying out principal component analysis (PCA) on simulation trajectories of the local nanoparticle spatial fluctuations with the rotation angle of the OM structures aligned to a particular orientation (thereby removing the driven rotation of the whole OM array). The principal components (PCs) generated are taken as the collective fluctuation modes, which are labeled according to their decreasing variance. In Fig. 6.2, the red (blue) sticks represent the collective motion of the particles when the incident light is left (right) circularly polarized. The dark (light) shade of the sticks shows the direction (i.e., phase) of

the particle motion when the structure deforms towards the positive (negative) direction of the collective coordinate.

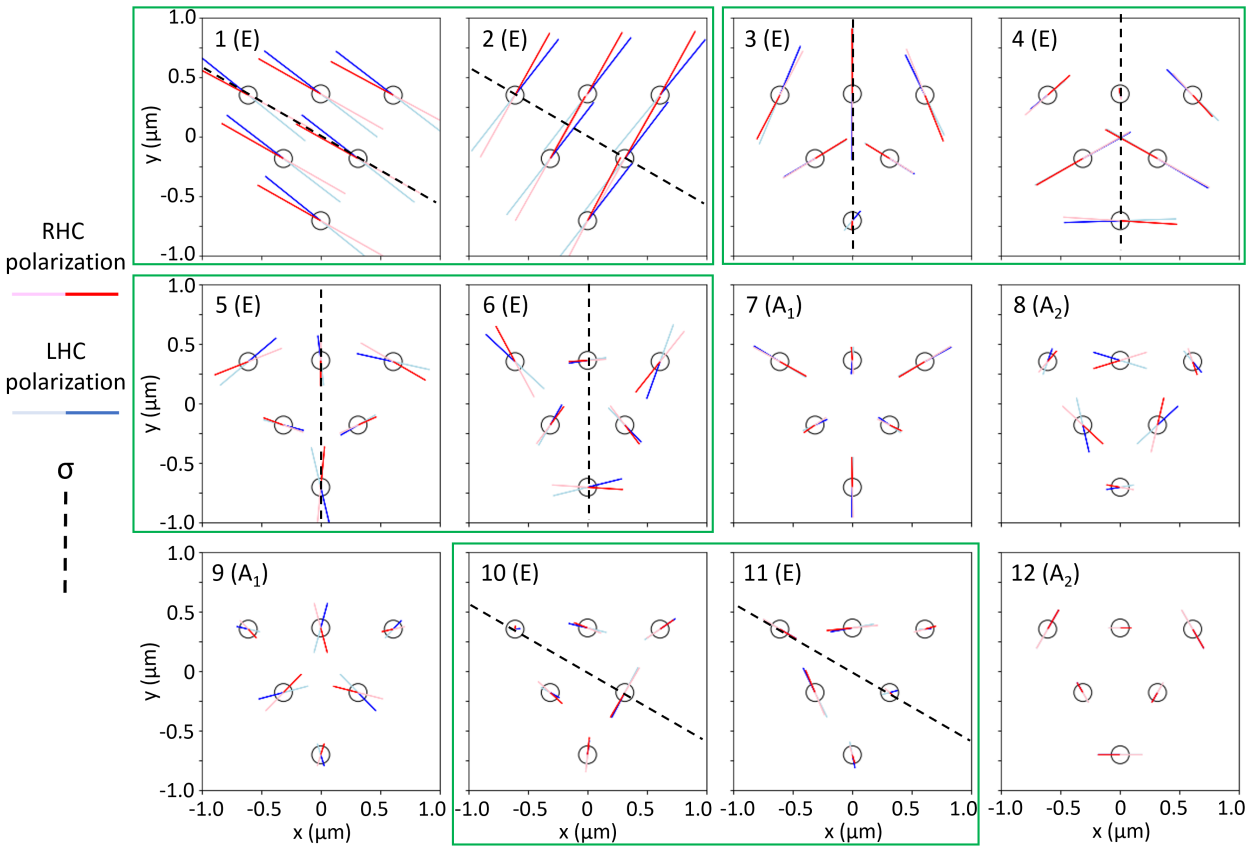

Figure 6.2: Twelve collective fluctuation modes of the triangle structure of the 6-particle OM system. The colored solid lines depict the magnitudes and directions of the collective particle motions. The color defines the direction of the collective motion; i.e., particles simultaneously move in the indicated directions for the same color. Modes 1 and 2 correspond to translation and mode 12 to rotation. Modes 3-11 correspond to collective vibration-like motion and the length of the solid lines of these modes is proportional to the standard deviations along these coordinates. The red and blue solid lines represent the motions for right- and left-hand circularly polarized laser beams. The dark (light) color corresponds to the positive (negative) direction of the mode coordinate. The pair of modes in each green box are degenerate and their mirror axes are shown by the black dashed line. The irreducible representations (i.e., E, A_1, A_2) of the structure symmetry point group D_3 are given in parentheses.

The symmetry of the collective PC modes in Fig. 6.2 does not rigorously follow the symmetry of the corresponding irreducible representation of the structure symmetry point group D_3 because the chirality of the incident light polarization breaks the mirror symmetry.

This broken symmetry occurs in both the degenerate modes and the non-degenerate modes. For each pair of degenerate modes, the choice of the two basis vectors is supposed to be arbitrary, but Fig. 6.2 shows that modes 1, 3, 5, and 11 are symmetric, while modes 2, 4, 6, and 10 are anti-symmetric. Furthermore, the non-degenerate modes 7, 8, and 9 become mirror images when the chiral perturbation of the incident light switches from left-hand to right-hand.

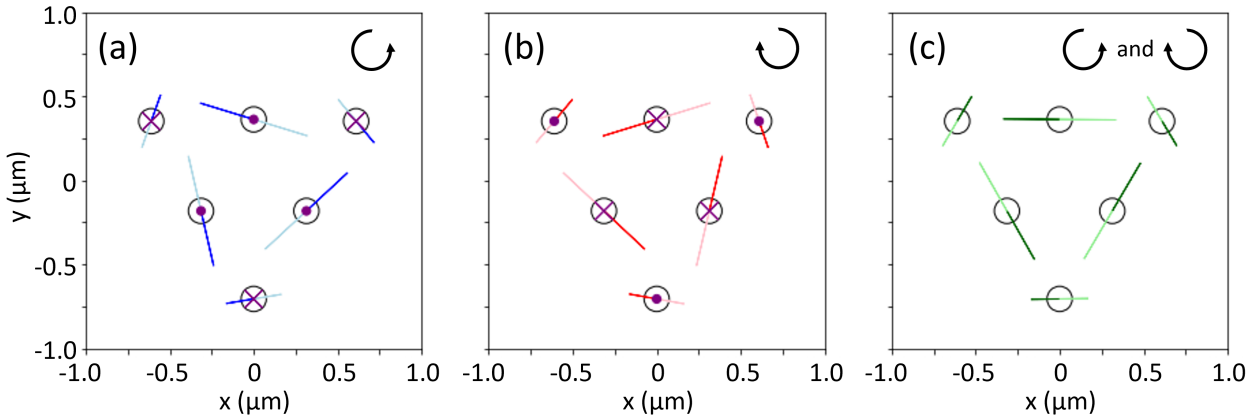

Figure 6.3: The 8^{th} collective fluctuation mode of the triangle configuration of the 6-particle OM system. The colored solid lines depict the magnitudes and directions of the collective particle motions. The dots and crosses labeled on the particles correspond to the direction of the cross-product between the radial and the azimuthal component of the particle motion. Specifically, dots correspond to pointing outward and crosses correspond to pointing inward. (a) The incident light is left-hand circularly polarized. (b) The incident light is right-hand circularly polarized. (c) The polarization of the incident light is switched between left- and right-hand circular polarization for each successive simulation time step, hence there is no (net) symmetry breaking.

In order to address this last point more clearly, mode 8 is taken as an example and shown in detail in Fig. 6.3. The 8th modes for left- and right-hand circularly polarized incident light shown in Figures 6.3a and 6.3b do not have mirror symmetry, but they are mirror images of each other because the circular polarization of the incident light acts as a symmetry breaking perturbation that affects the direction (orientation) and symmetry of the modes. If we eliminate this broken symmetry by switching the incident light between right- and left-handed circular polarization in successive simulation time steps, the resulting helical

handedness averaged mode 8 shown in Fig. 6.3c gains mirror symmetry. Therefore, the mode shown in Fig. 6.3(c) has all the symmetry elements for the A_2 irreducible representation of the structure symmetry point group D_3. For the helical field perturbed cases in Figures 6.3a and 6.3b, the symmetry of the system is the structure symmetry perturbed by the incident light symmetry, which is $D_3 \cap C_\infty = C_3$, a subgroup of the structure symmetry point group. Therefore, the irreducible representation of the modes shown in Fig. 6.3a and 6.3b is reduced from A_2 of the structure symmetry point group D_3 to the irreducible representation A of the subgroup C_3. However, the symmetry breaking perturbation is small enough that the variances of the degenerate mode pairs are close and that the modes are close to their symmetric unperturbed versions. This is the reason why the broken symmetry is treated as a perturbation in the following discussion.

6.4.2 Collective fluctuation modes and symmetry of the hexagonal OM structure

This section shows the collective fluctuation modes of the 7-particle 6-fold symmetric OM structure and their comparison to the Raman activity of the normal modes of a united-atom model of benzene.

Fig. 6.4 shows the collective fluctuation modes computed for the 7-particle 6-fold symmetric OM system. The inversion symmetry restriction in the modes that belong to the E_2 irreducible representation as well as the 3-fold rotational symmetry restriction in all non-degenerate modes that belong to the 1-dimensional irreducible representations of the D_6 point group means that central particle motion can only be significant in the E_1 modes. In addition, E_1 is a Raman-inactive irreducible representation for D_6 as a point group for the 2-dimensional OM system in the x-y plane, so that all the collective fluctuation modes that correspond to Raman-active irreducible representations do not involve central particle motion. Since there is no central particle in the middle of the planar hexagons in graphene or a

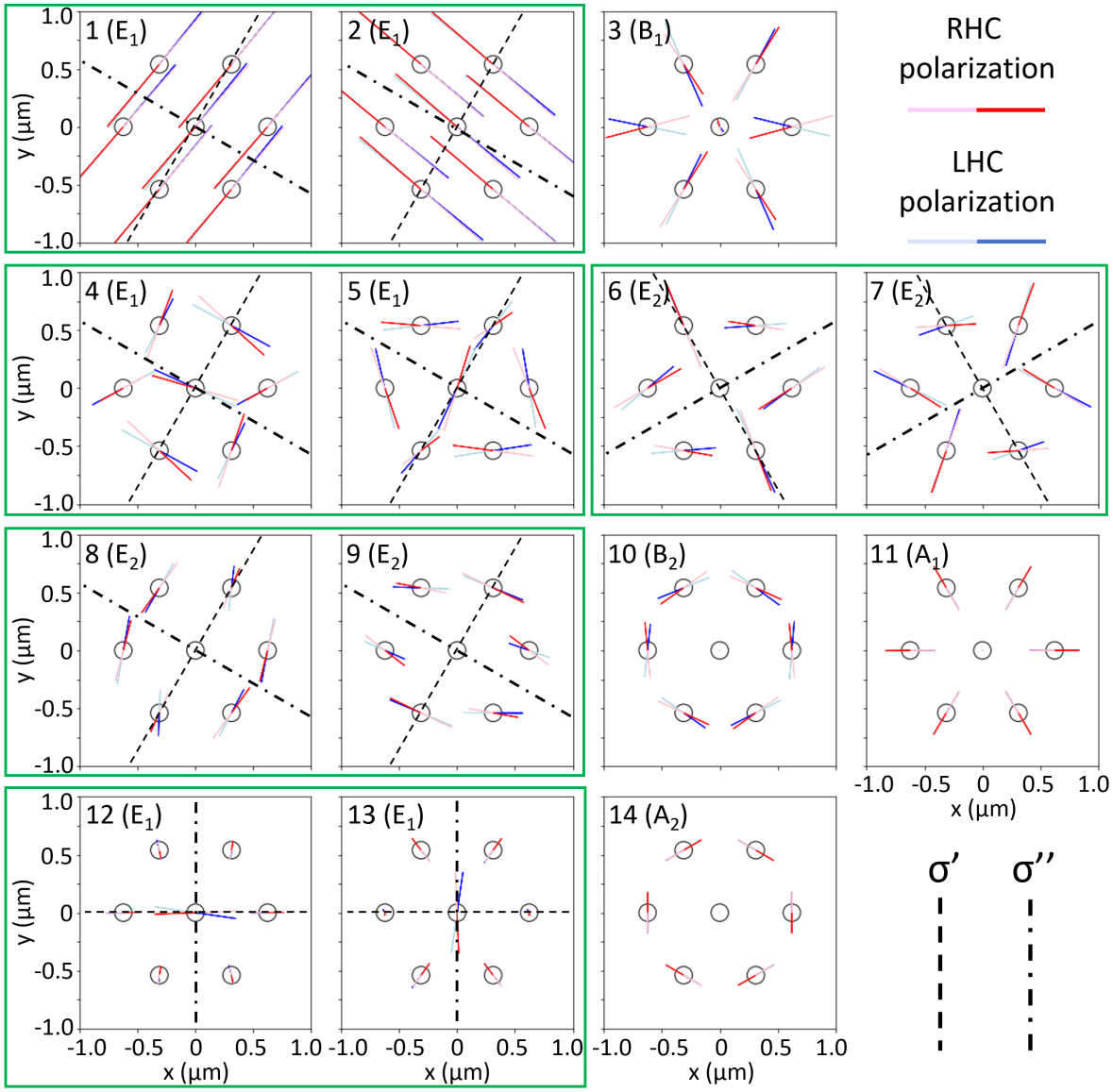

Figure 6.4: Fourteen collective fluctuation modes of the 6-fold symmetric structure of the 7-particle OM system. The colored solid lines depict the magnitudes and directions of the collective particle motions, and the color defines the direction of the collective motion; i.e., particles simultaneously move in the indicated directions for the same color. Modes 1 and 2 correspond to translation and mode 14 to rotation. Modes 3-13 correspond to vibration-like motions and the length of the solid lines of these modes is proportional to the standard deviations along these modes. The red and blue solid lines represent the motions for right- and left-hand circularly polarized laser beams. The dark (light) color corresponds to the positive (negative) direction of the mode coordinate. The pair of modes in each green box are degenerate and their mirror axes are shown by two different styles of black dashed lines according to the conjugacy class they belong to. The irreducible representations of the structure symmetry point group D_6 are labeled in parentheses.

united-atom model of benzene, the analogy to Raman activity between these 2-dimensional 6-fold symmetric hexagonal structures and the 7-particle 6-fold symmetric OM structure is reasonable; this analogy cannot be affected by the extra central particle in the 7-particle 6-fold symmetric structure. Furthermore, similar to the 6-particle triangle structure case, the symmetry of the modes in Fig. 6.4 does not rigorously follow but is close to the symmetry of the corresponding irreducible representation of the structure's symmetry point group D_6. This is because the chirality of the incident light polarization causes broken mirror symmetry that can be taken as a perturbation due to its moderate significance compared to the structure symmetry. Each green-boxed pair of degenerate modes in Fig. 6.4 consist of one symmetric and one anti-symmetric mode.

6.4.3 Groups and conjugacy classes

This section shows detailed differences among the mirror symmetries of the collective fluctuation modes of the 7-particle 6-fold symmetric OM structure by introducing conjugacy classes in a point group.

A point group consists of group elements that represent symmetry operations. The multiplication adapted to the point group acting on group elements pairs corresponds to successive symmetry operations. Given a group G, two group elements of G, g and g', are conjugate to each other if there exists a group element h such that $g' = hgh^{-1}$ holds. A conjugacy class of group G is a subset of G such that each pair of elements in the conjugacy class are conjugate to each other.

Compared to the D_3 point group in the case of the triangle OM structure for which all the mirror symmetries belong to a single conjugacy class, the mirror symmetries in D_6 point group belong to 2 conjugacy classes, which complicates the mirror symmetries in the collective fluctuation modes generated by PCA. The three mirrors that pass through two opposite particles belong to one conjugacy class, in which the three mirror symmetries are

referred to as σ'. The other three mirrors that bisects two opposite vertices belong to the other conjugacy class, in which the three mirror symmetries are referred to as σ''. Among the degenerate pairs that belong to the irreducible representation E_1, modes 1, 5, and 12 are symmetric with respect to a σ' but anti-symmetric with respect to a σ'', while modes 2, 4, and 13 are anti-symmetric with respect to a σ' but symmetric with respect to a σ''. On the other hand, among the degenerate pairs that belong to the irreducible representation E_2, with respect to both a σ' and a σ'', modes 6 and 8 are symmetric while modes 7 and 9 are anti-symmetric.

6.4.4 Initial structural relaxation by mode and symmetry

This section shows the mode- and symmetry-dependence of the electrostatic repulsion of the charged nanoparticles in OM structures by analyzing their relaxed structures with different powers of the optical beam. We find that the structural deviation of the relaxed structure in response to the power change of the optical beam is only significant along the collective fluctuation modes corresponding to the A-type irreducible representations.

Apart from the electrodynamic interactions amongst the nanoparticle constituents of an OM structure that are related to light scattering, the other interaction among the particles is the double-layer Coulomb potential between colloidal particles. Therefore, before analyzing the scattering properties, we first explain the relation between the double-layer potential and the collective coordinates by analyzing the collective fluctuation mode components of the structural deviations between the reference OM structure and OM structures in the deterministic simulation trajectory initiated at the reference structure evaluated for different beam powers. The reference structure is taken as the mean structure of the fluctuation simulation trajectory at 100 K temperature and 200 mW beam power, which is also used as the data set for the PCA that allows generating the collective fluctuation modes.

The analysis of the collective fluctuation mode components of the structural deviations

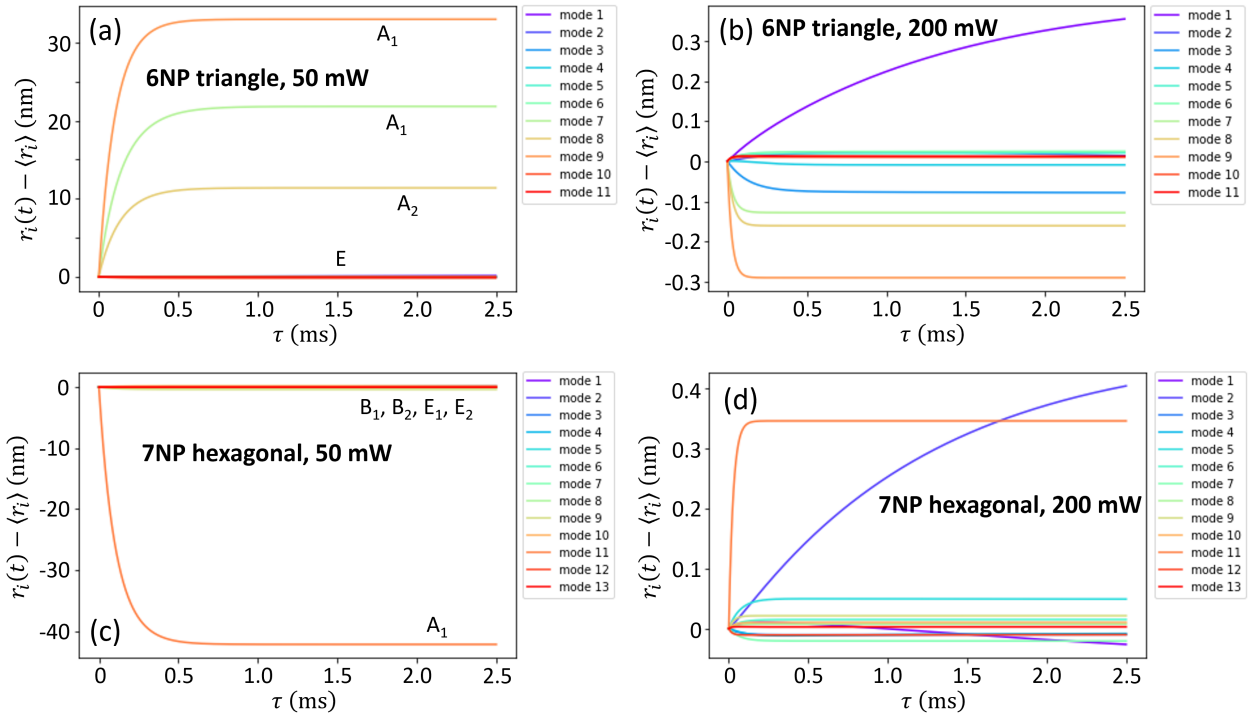

Figure 6.5: Symmetry analysis of electrostatic properties using structural deviations between the reference structure and the OM structures in the trajectory of the deterministic simulation. (a) Plot of the collective fluctuation mode components of the structural deviations between the reference structure of 6-particle triangle structure (termed X) and the OM structures in the trajectory of the deterministic simulation with 50 mW beam power (Y) initiated at the reference structure as a function of simulation elapsed time. The reference structure is taken as the mean structure of the fluctuation simulation at 100 K temperature and 200 mW beam power. (b) X = 6-particle triangle structure, Y = 200 mW. (c) X = 7-particle 6-fold symmetric structure, Y = 50 mW. (d) X = 7-particle 6-fold symmetric structure, Y = 200 mW.

between the reference OM structure and OM structures in the deterministic simulation trajectory initiated at the reference structure and evaluated for different beam powers is based on the fact that the electrodynamic interaction changes proportionally to the change of optical beam power, while the double-layer potential is independent of beam power. Therefore, the difference between the converged structure in the deterministic simulation and the reference structure will be negligibly small if the double-layer potential is absent [171]. Comparing Fig. 6.5b to 6.5a, or Fig. 6.5d to 6.5c, we see that if the deterministic simulation is at the same beam power as the fluctuation simulation used to obtain the reference structure, the

structural difference between the two is very small compared to structural difference between the deterministic simulation and the reference structure at different beam powers.

However, in the cases where the deterministic simulation and the reference structure are at different beam powers, the structural difference is significant only along the collective fluctuation modes corresponding to the A-type of irreducible representation of the structure symmetry point group. In the triangle case shown in Fig. 6.5a, modes 7, 8, and 9 that correspond to the A_1 and A_2 irreducible representation of the D_3 point group dominate the structural difference. Note that the structural difference along the A_2 mode is less than that of the two A_1 modes. This is caused by the chiral perturbation of the optical field and explained in detail in the Discussion Section. The only A_2 mode in the 6-fold symmetric case shown in Fig. 6.5c is rotation. However, rotation is eliminated before carrying out the PCA, so only mode 11, which corresponds to the A_1 irreducible representation of the D_6 point group, contributes to the structural difference.

6.4.5 Scattering cross-section vs. mode displacement

We show that the mode-dependence of the scattering cross-section is monotonic along A-type collective fluctuation modes and not monotonic along other modes for two representative OM structures.

Apart from the double layer potential, the symmetry of the collective fluctuation modes also plays an important role in the scattering properties of the OM structures. This is the key idea of this paper. Fig. 6.6a shows the normalized scattering cross-section as a function of structural displacement along the collective fluctuation modes with respect to the reference structure of the 6-particle triangle OM structure. The scattering cross-section is normalized by division by πR^2, one-fourth of the nanoparticle surface area [17]. The normalized scattering cross-section only changes monotonically about the mean position (origin) for modes 7, 8, and 9. These modes correspond to the A-type irreducible representation. In other words,

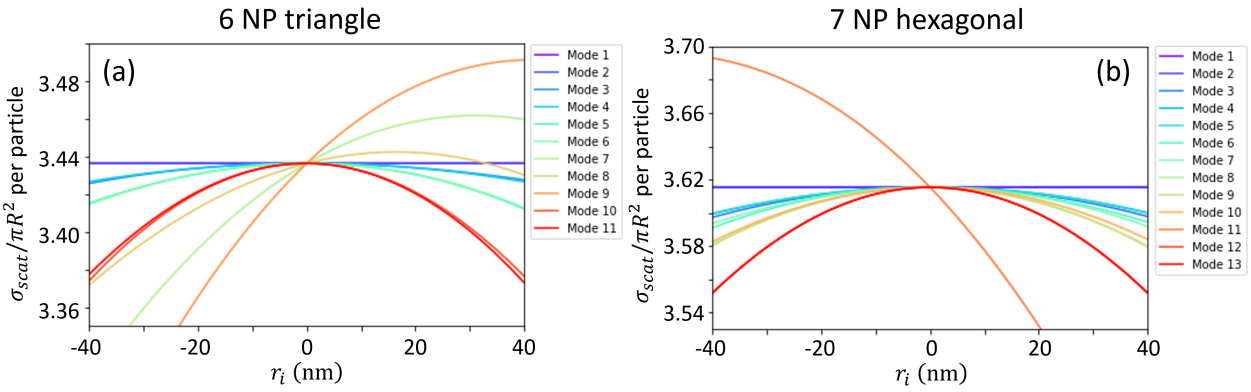

Figure 6.6: Scattering cross-sections as a function of structural displacement along the collective fluctuation modes with respect to the reference structure of the: (a) 6-particle triangle OM structure for which the scattering cross-section only changes monotonically with displacement along mode 7, 8, and 9; (b) 7-particle 6-fold symmetric OM structure for which the scattering cross-section only changes monotonically with displacement along mode 11. The scattering cross-sections are normalized by πR^2, where $R = 75$ nm is the radius of a single Ag nanoparticle.

the partial derivatives of the scattering cross-section analyzed at the origin along these three modes (7, 8, and 9) are non-zero, but are zero along other modes. Likewise, Fig. 6.6b shows the same properties for the 7-particle 6-fold symmetric OM structure. The partial derivative of the scattering cross-section analyzed at the origin is non-zero only along mode 11, which is the only fluctuation mode that corresponds to the A type irreducible representation in the 7-particle 6-fold symmetric OM structure.

6.4.6 Induced-polarization vs. mode displacement

We show that the mode-dependence of the induced-polarization is monotonic along A-type collective fluctuation modes and not monotonic along other modes for two representative OM structures.

In addition to the scattering cross-section, we examined the averaged induced-polarization of the OM structure, which is related to the electrodynamic interaction among the particles [3]. Fig. 6.14 in Section 6.A shows the particle-wise induced-polarization as a function of structural displacement along the collective fluctuation modes. Fig. 6.7 shows that the

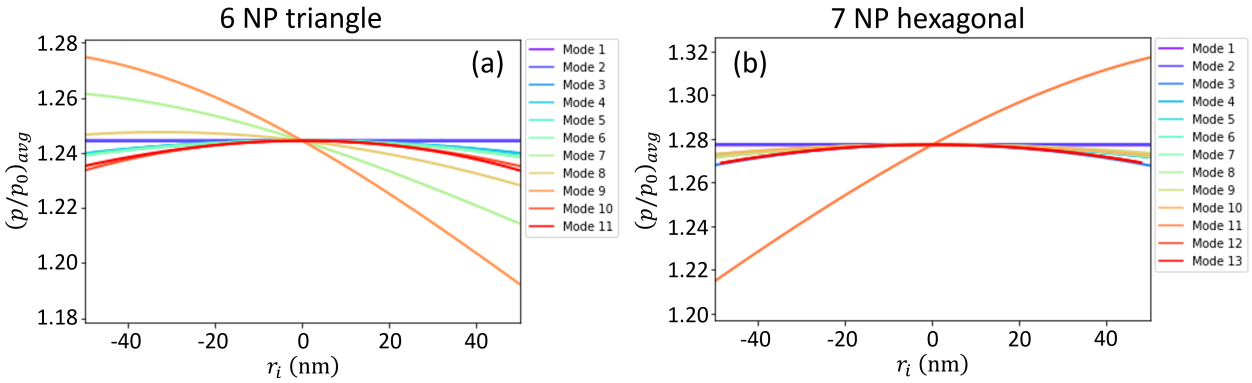

Figure 6.7: Average induced-polarization as a function of structural displacement along the collective fluctuation modes with respect to the reference structure of the: (a) 6-particle triangle OM structure where the average induced-polarization changes monotonically with displacement along mode 7, 8, and 9; (b) 7-particle 6-fold symmetric OM structure where the averaged induced-polarization changes monotonically with displacement along mode 11.

symmetry of the collective fluctuation modes also play an important role in the average induced-polarization of the OM structure. Fig. 6.7a shows the average induced-polarization as a function of structural displacement along the collective fluctuation modes with respect to the reference structure of the 6-particle triangle OM structure. At the origin, the average induced-polarization only changes monotonically along modes 7, 8, and 9. These modes correspond to the A-type irreducible representation. Likewise, Fig. 6.7b shows the same properties for the 7-particle 6-fold symmetric structure. The partial derivative of the average induced-polarization analyzed at the origin is non-zero only along mode 11, which is the only collective fluctuation mode that corresponds to the A-type irreducible representation in the 7-particle 6-fold symmetric structure.

6.4.7 Chevron OM structure

The same set of analyses described in Figures 6.5-6.7 are also carried out for the 6-particle chevron OM structure. The chevron is an example of an OM structure with low symmetry (C_s point group). The results of this symmetry analysis on its collective fluctuation modes are provided in Fig. 6.11, Fig. 6.12, and Fig. 6.13 in Section 6.A. Our conclusions for the

chevron are:

1. There are only 2 irreducible representations in the C_s point group, namely the symmetric A_1 and the antisymmetric A_2 representations.

2. Both A_1 and A_2 irreducible representations of the C_s point group turns into the same A irreducible representation of the C_1 point group, which enables the A_2 modes to interact with A_1 irreducible representation.

3. The collective modes that belong to either the A_1 irreducible or the A_2 irreducible representations manifest significant differences between the converged structures of the deterministic simulation and the reference structure at different beam powers. The collective modes that belong to either the A_1 irreducible or the A_2 irreducible representations also manifest non-zero derivatives of the scattering cross section and averaged induced-polarization at the origin.

6.5 Discussion

From the symmetry analysis of both the double-layer electrostatic and optical electrodynamic interactions along the collective fluctuation modes, we see that only the modes corresponding to the A-type irreducible representation manifest significant differences between the converged structure of the deterministic simulation and the reference structure at different optical beam powers, and therefore manifest non-zero derivatives of the scattering cross-section and averaged induced-polarization about the origin. This observation is the same as one obtains from symmetry analysis of the Raman activity of molecular vibration modes; the coordinate-dependent derivative of the molecular polarizability near its equilibrium structure is non-zero along Raman active modes but is zero along Raman inactive modes.

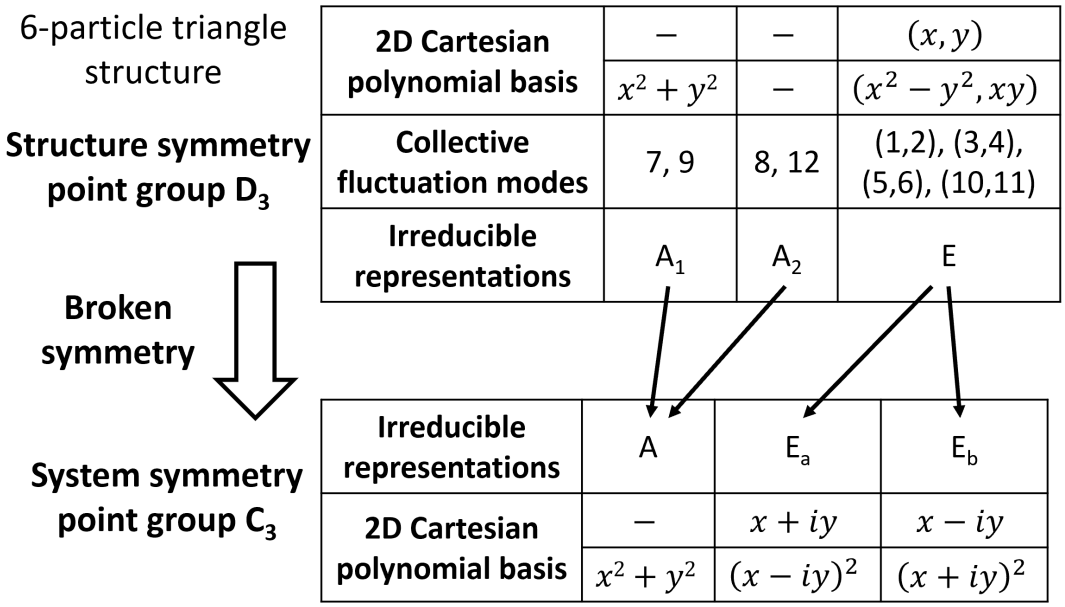

Figure 6.8: Assignment of the collective fluctuation modes of the 6-particle triangle OM structure to the irreducible representation of point groups C_3 and D_3. The solid arrows indicate the change in the irreducible representations due to the chiral broken symmetry caused by the circular polarization of the incident light. Note that the Cartesian polynomial basis only includes x and y coordinate because the dynamics of the OM system we focus on is in 2-dimensional space.

The Raman active molecular vibration modes correspond to irreducible representations that have a quadratic functional basis, so for the D_3 point group shown in Fig. 6.8 the modes corresponding to the A_1 and E irreducible representations are Raman active. On the other hand, in the case of the 6-particle triangle OM structure, A_1 and A_2 modes manifest a non-zero coordinate-dependent derivatives of their scattering cross-section and their induced-polarization. The differences in the Raman activity of the A_2 mode in the analyses of molecular vibration modes and OM fluctuation modes are due to the chiral symmetry breaking in the OM system that reduces the OM system point group symmetry from D_3 to C_3. Therefore, both A_1 and A_2 irreducible representations of the D_3 point group turn into the same A irreducible representation of the C_3 point group. This change of symmetry enables the A_2 modes to interact with the A_1 irreducible representation. This chiral perturbation-induced symmetry breaking is illustrated by the smaller structural difference

along the A_2 mode relative to the A_1 mode shown in Fig. 6.5a and by the smaller absolute value of the derivative of the scattering cross-section (in Fig. 6.6a) and induced-polarization (in Fig. 6.7a) along the A_2 mode relative to the A_1 mode.

The difference of the two analyses in the E mode is due to the polarization difference between the light used in the OM system and conventional Raman scattering. While conventional Raman scattering does not specify the polarization of the light, the optical beam used to assemble the 2-dimensional OM structure studied here is circularly polarized. Therefore, helicity-resolved Raman scattering [165,166] is a more relevant for comparison. Previous work [165,166] has shown that when Raman scattering is carried out for 2-dimensional MoS_2, a layered material with D_3 type of symmetry (like the case of graphene in Fig. 6.1a-c), the polarization of the scattered light related to the A_1 mode is the same as the polarization of the incident light, while the polarization of the scattered light related to the E mode is opposite to the polarization of the incident light. As a result, the symmetry analysis of the electrostatic and electrodynamic properties along the collective fluctuation modes of the 6-particle triangle OM structure can be compared to the helicity-resolved Raman scattering analysis when the polarization of incident and scattered light are the same.

Similar results can be obtained from the symmetry analysis of the scattering properties and the collective fluctuation modes of the 7-particle 6-fold symmetric OM structure. As is shown in Fig. 6.9, modes corresponding to the A_1 and E_2 irreducible representations of the D_6 point group are Raman active. This is different from the properties obtained for the 6-fold symmetric OM structure where only the A_1 modes manifest a non-zero derivative of the scattering cross-section and the induced-polarization. Like in the case of D_3 symmetry discussed above, the ideas from helicity-resolved Raman can be used to explain the behavior of the E_2 mode in the D_6 symmetry case. Previous publications [162,166] showed that when helicity-resolved Raman scattering is carried out for graphene, a 2-dimensional material with D_6 type of symmetry, the polarization of the scattered light related to the G-band

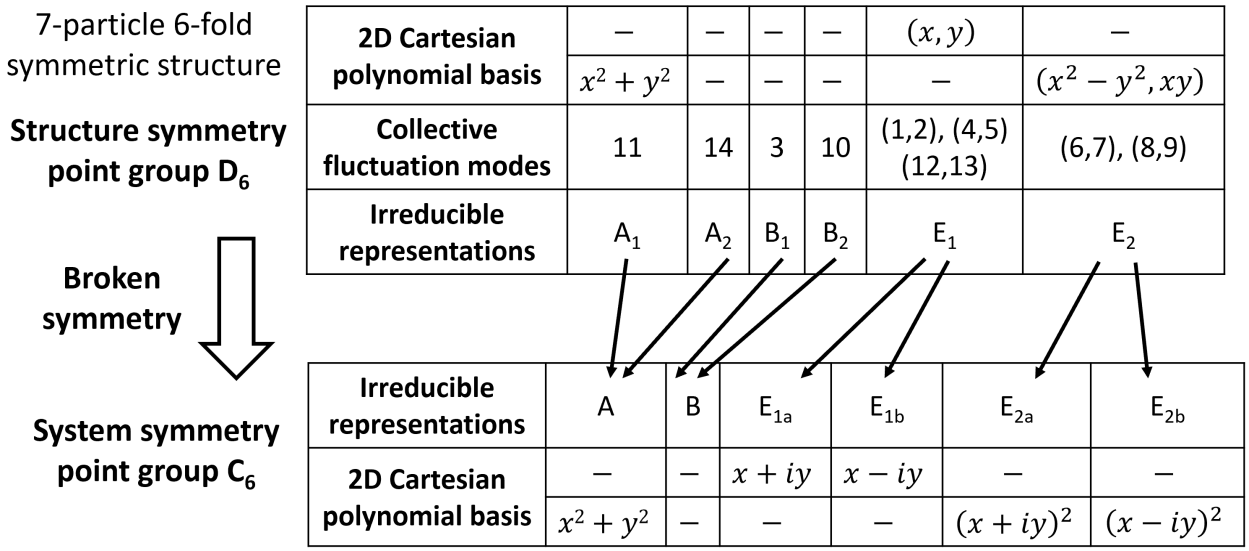

Figure 6.9: Assignment of the collective fluctuation modes of the 7-particle 6-fold symmetric OM structure to the irreducible representation of point groups C_6 and D_6. The solid arrows indicate the change in the irreducible representations due to the chiral broken symmetry caused by the circular polarization of the incident light. Note that the Cartesian polynomial basis only includes x and y coordinate because the dynamics of the OM system we focus on is in 2-dimensional space.

phonon with E_2 symmetry is opposite to the polarization of the incident light. Therefore, the symmetry analysis of the electrostatic and electrodynamic properties along the collective fluctuation modes of the 7-particle 6-fold symmetric OM structure can also be compared to the helicity-resolved Raman scattering analysis when the polarization of incident and scattered light are the same.

6.6 Conclusions

This study has demonstrated the symmetry correspondence of collective mode-dependent light scattering from optical matter (OM) systems and (helicity-resolved) Raman scattering. The key idea is that the monotonic dependence of the induced-polarization and scattering of specific symmetry overdamped fluctuation modes in OM systems is homomorphic to the monotonic coordinate-dependent polarizability of Raman active (e.g., normal) modes.

Collective coordinates are obtained for the 6-particle triangle and the 7-particle 6-fold symmetric OM structures by carrying out a principal component analysis on fluctuation trajectories of the reference OM structure. For both OM structures, we calculate the structural difference between the reference trigonal lattice structure and the relaxed structure of a deterministic simulation, as well as the scattering cross-section and averaged induced-polarization along the collective fluctuation modes. In both 6-particle triangle and 7-particle 6-fold symmetric OM structure, we find that the collective fluctuation modes corresponding to the A-type irreducible representations of the structure symmetry point groups behave differently from the rest of the collective fluctuation modes. Using an analysis of the broken symmetry introduced by the chiral nature of the circular polarization of the incident light, we find that the OM collective fluctuation modes along which the scattering cross-section and induced-polarization change monotonically correspond to the same group representations of the helicity-resolved Raman active modes of a 2-dimensional material with the same symmetry.

The same type of analyses of the collective fluctuation mode-dependent properties is carried out for the 6-particle chevron structure as an example of an OM structure with lower symmetry. The results for the chevron OM structure are shown in Figs. 6.11, 6.12, and 6.13 in Section 6.A. The 6-particle chevron OM structure corresponds to the low symmetry C_s point group, which stands in contrast to the high symmetry and analysis of the 6-particle triangle and 7-particle 6-fold symmetric OM structure in the main text. There are only 2 irreducible representations in the C_s point group, namely the symmetric A_1 and the antisymmetric A_2 representations. Though only the A_1 irreducible representation corresponds to a quadratic function basis in the x-y plane of the OM system, both A_1 and A_2 modes manifest significant differences between the relaxed structures of a deterministic simulation and the reference structure at different beam powers. Both A_1 and A_2 modes also manifest non-zero derivatives of the scattering cross section and averaged induced-polarization at the origin. This is

because both A_1 and A_2 irreducible representations of the C_s point group turn into the same A irreducible representation of the C_1 point group, which enables the A_2 modes to interact with A_1 irreducible representation.

These findings will be useful in computing the polarization of the light scattered light from an OM structure, which will enable the analysis of the orbital momentum in the OM scattering process that is similar to the pseudoangular momentum analysis for the helicity-resolved Raman scattering of 2-dimensional materials [172, 173]. Our insights suggest future (experimental) studies that utilize strategies analogous to helicity-resolved Raman scattering to study degenerate modes of overdamped OM systems. This analysis will enable deeper understanding of the information contained in the scattered light about the response of overdamped internal motions of various systems, e.g., from colloids to plasmas, to external applied fields.

6.A Supporting Information

Because the OM system is overdamped, the collective PC modes shown should not be mistaken for or overinterpreted as underdamped deterministic harmonic motions in the actual EDLD simulations or in experimental studies of OM systems in solution.

The 6-particle chevron OM structure corresponds to the low symmetry C_s point group in contrast with the analysis of the 6-particle triangle 3-fold symmetric and 7-particle 6-fold symmetric OM structure in the main text. There are only 2 irreducible representations in the C_s point group, namely the symmetric A_1 and the antisymmetric A_2 representations. Only the A_1 irreducible representation corresponds to a quadratic function basis in the 2D x-y plane of the OM structure. However, both A_1 and A_2 irreducible representations manifest significant differences between the converged structures of the deterministic simulation and the reference structure at different beam powers. In addition, the derivatives of the scattering cross section and averaged induced-polarization at the origin are non-zero along the modes with A_1 and A_2 irreducible representations. This is because the chirality of the circular polarization of the optical beam reduces the system symmetry from C_s to C_1. As a consequence, both A_1 and A_2 irreducible representations of the C_s point group turn into the same A irreducible representation of the C_1 point group, which is illustrated by the smaller structural difference along the A_2 modes shown in Fig. 6.11a as well as the smaller absolute value of the derivative of the mode-dependent scattering cross-sections (in Fig. 6.12) and induced-polarizations (in Fig. 6.13) along the A_2 mode.

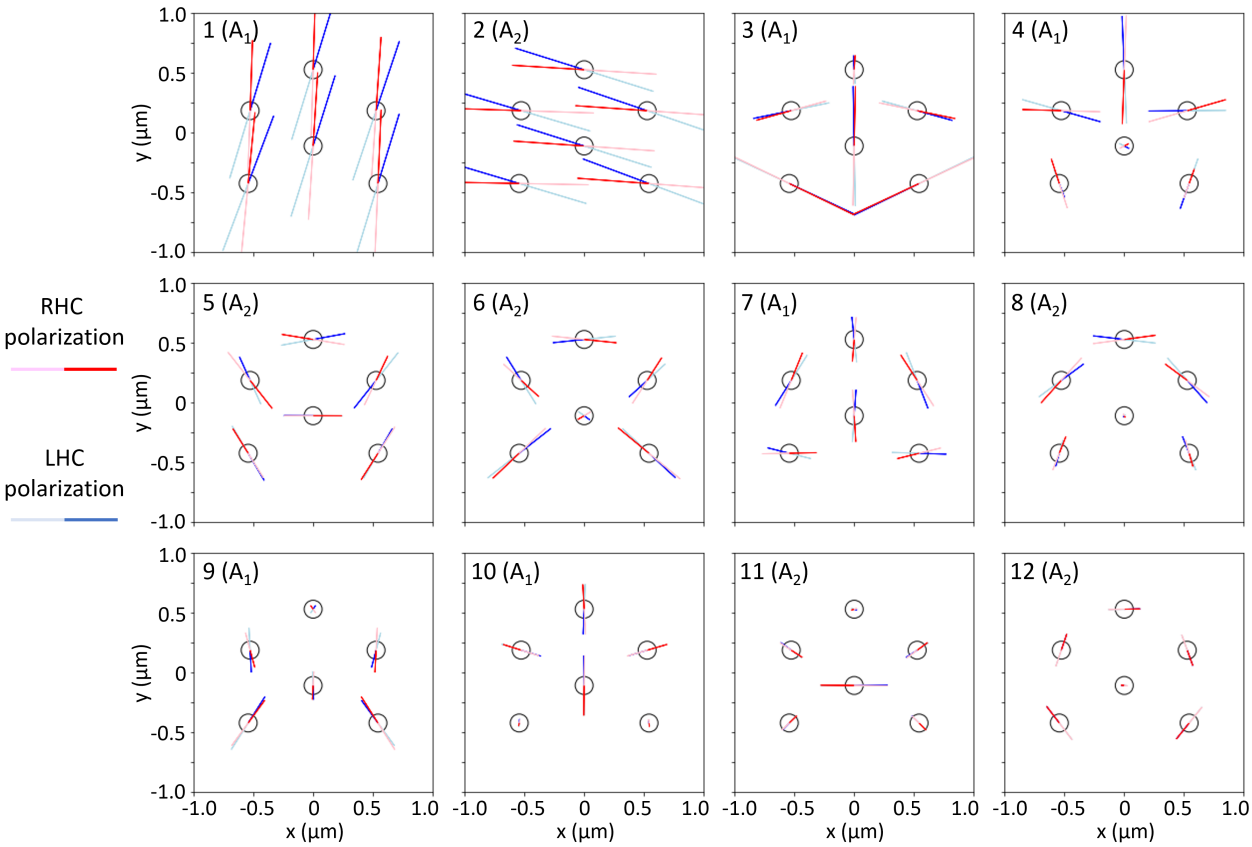

Figure 6.10: Twelve collective fluctuation modes of the chevron structure of the 6-particle OM system. The colored solid lines depict the magnitudes and directions of the collective particle motions. The color defines the direction of the collective motion; i.e., particles simultaneously move in the indicated directions for the same color. Modes 1 and 2 correspond to translation and mode 12 to rotation. Modes 3-11 correspond to "vibrations" and the lengths of the solid lines showing these modes are proportional to the standard deviations along these modes. The red and blue solid lines represent right- and left-hand circularly polarized laser beam illuminations. The dark (light) color corresponds to the positive (negative) direction of the mode coordinate. The irreducible representations of the structure symmetry point group C_s are labeled in the parentheses.

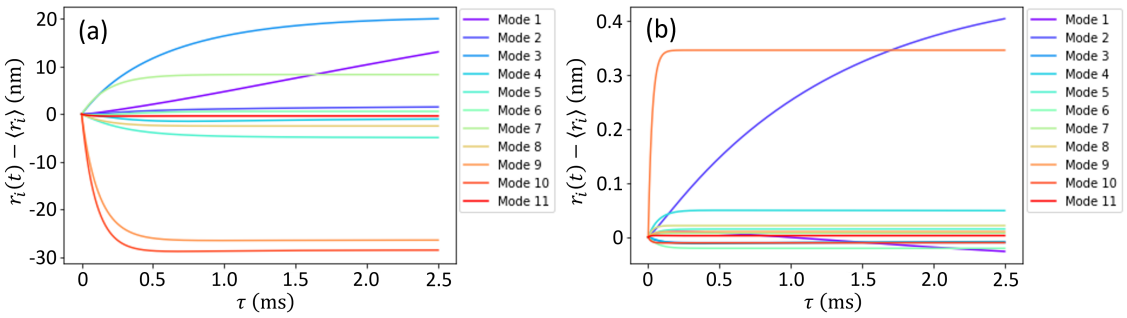

Figure 6.11: Plots of the collective fluctuation mode components of the difference between the reference structure of the 6-particle chevron OM structure and the trajectory of the deterministic simulation with (a) 50 mW and (b) 200 mW beam power initiated at the reference structure as a function of (simulation) time. The reference structure is taken as the mean structure of the fluctuations of the 6-particle chevron structure obtained in simulations at 100 K temperature and 200 mW beam power.

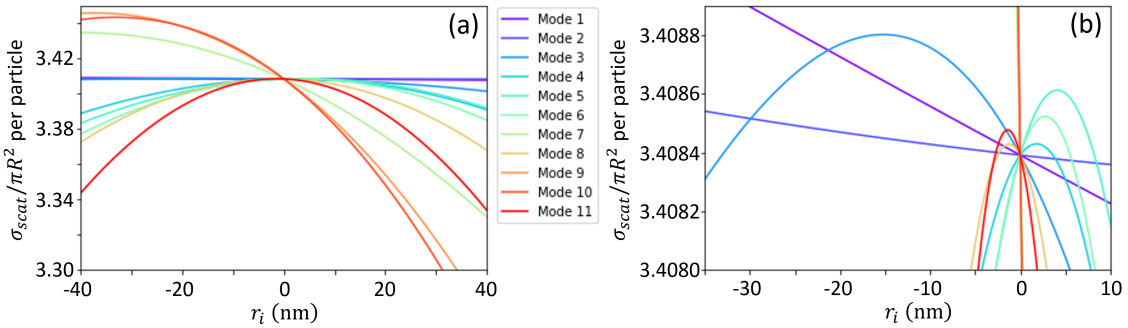

Figure 6.12: Total scattering cross-section as a function of structural deviation along the collective fluctuation modes with respect to the reference structure of 6-particle chevron OM structure. The scattering cross sections are normalized by πR^2, where $R = 75$ nm is the radius of a single nanoparticle. Panel (b) is a zoomed-in version of panel (a).

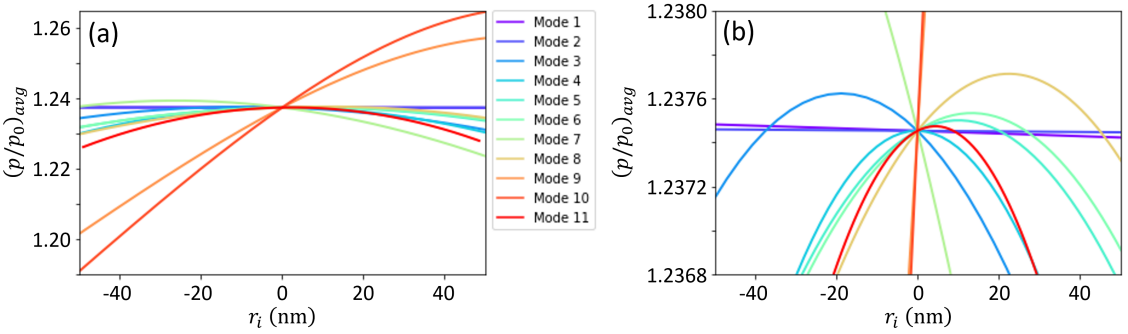

Figure 6.13: Average induced-polarization as a function of structural deviation along the collective fluctuation modes with respect to the reference structure of of 6-particle chevron OM structure. Panel (b) is a zoomed-in version of panel (a).

Fig. 6.14, which shows the single-particle induced-polarizations as functions of structural displacement along the PCs of the OM triangle structure, is not the result of an additional analysis. Rather, Fig. 6.14 shows in detail the single particle induced-polarizations that are used to compute the average induced-polarizations plotted in Fig. 6.7. The induced-polarization plotted in the main text is the average of the single-particle induced-polarization. Note that due to the 3-fold rotation symmetry of modes 7, 8, and 9, the single-particle induced-polarization plotted along these three modes contains 3-fold overlaps of the curves in the single-particle induced-polarization plots.

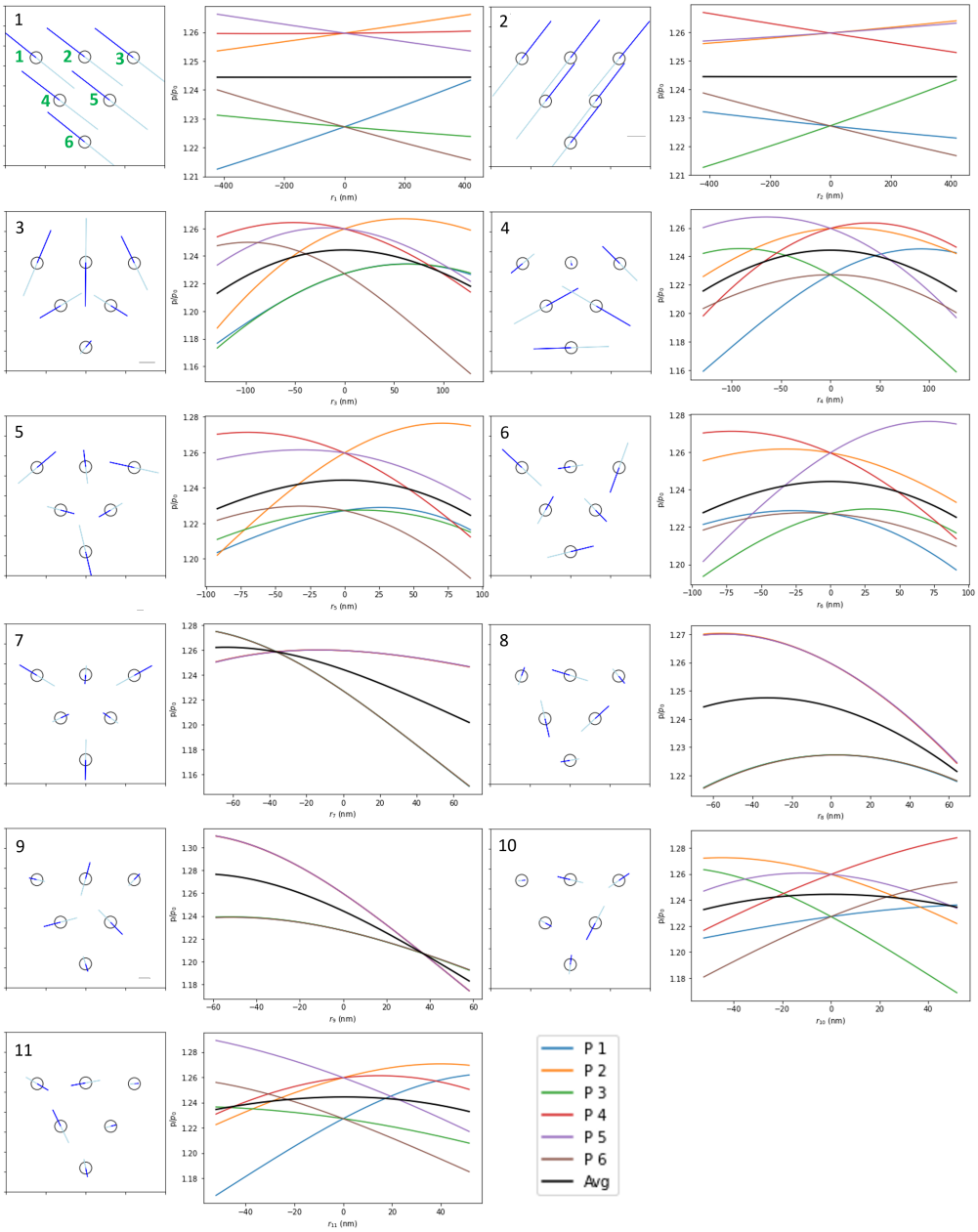

Figure 6.14: Induced-polarization of different particles as a function of deterministic structural deviation along the collective modes of the 6-particle OM system. The diagrams in the square boxes show the collective modes; the mode number is shown in the upper left corners. The colored solid lines depict the magnitudes and directions of the collective particle motions, and the color defines the direction of the collective motion; i.e., particles simultaneously move in the indicated directions for the same color. The dark (light) color corresponds to the positive (negative) direction of the mode coordinate. The diagrams in the rectangular boxes are the average induced-polarization for the fluctuation mode (black) and the induced-polarization of each particle (colored) plotted as a function of the collective coordinates. The particle labels are the same for each mode, and are shown in the diagram of mode 1.

6.B Helicity-resolved Analysis Using Symmetry Group Representations

There is not a direct analogy to the Raman tensor for OM systems, so we use a symmetry method instead that only focuses on the irreducible representations of the three terms in the bra·ket product of the polarization state of incident light, Raman tensor, and the polarization state of outgoing light, based on the system symmetry point group for the 2D lattice materials [158, 159]. Therefore, instead of using Raman tensors, we can use irreducible representations to check the helicity-resolved Raman activity of the A_1 and E modes of the systems with D_3 symmetry as well as that of the A_1 and E_2 modes of the systems with D_6 symmetry.

The two analyses of the collective modes of the 6-particle triangle and 7-particle hexagonal OM structures discussed in this chapter demonstrate the monotonic behavior of scattering cross-section and induced-polarization of the OM structure when it deforms along a collective mode. This behavior is a homomorphism of the classical Raman effect and activity of normal modes of a molecule. This section explains why these two analysis gives different results for the E mode in the D_3 symmetry case (6-particle triangle OM structure and 2D MoS_2) and the E_2 mode in D_6 symmetry.

While these modes in the molecular case are Raman active, the change in scattering cross-section and induced-polarization in the OM systems are not monotonic along these modes. This difference for the E modes of the OM systems is due to the polarization difference between the light used in the OM system and conventional Raman scattering. While conventional Raman scattering does not specify the polarization of the light, the optical beam used to assemble the 2-dimensional OM structures studied here is circularly polarized. Therefore, helicity-resolved Raman scattering [165, 166] is a more relevant for comparison.

Previous work [165, 166] has shown that when Raman scattering is carried out for 2-dimensional MoS_2, a layered material with D_3 type of symmetry (like the case of graphene

in Fig. 6-1a-c), the polarization of the scattered light related to the E mode is opposite to the polarization of the incident light. Similar results are obtained for the 2D material graphene with D_6 symmetry; when helicity-resolved Raman scattering is carried out for graphene, the polarization of the scattered light related to the G-band phonon with E_2 symmetry is opposite to the polarization of the incident light [162, 166].

In helicity-resolved Raman analysis, Raman tensors are used to compute the Raman cross-section $A\sum_j |\langle\varepsilon_o|R_j|\epsilon_i\rangle|^2$, where A is a constant coefficient, R_j is the Raman tensor of mode j, and $|\varepsilon_i\rangle$, $|\varepsilon_o\rangle$ are the states of incoming and outgoing light [165]. A matrix product is used to compute this bra·ket product for each mode j: $\langle\varepsilon_o|R_j|\epsilon_i\rangle = \boldsymbol{\sigma}_o^\dagger \boldsymbol{R}_j \boldsymbol{\sigma}_i$ where $\boldsymbol{R}_j$ is a 3x3 matrix and $\boldsymbol{\sigma}$ is the Jones vector [165]. For circularly polarized light, the Jones vector is $\boldsymbol{\sigma}_\pm = [1, \pm i, 0]/\sqrt{2}$, where +(-) corresponds to left (right) circular polarization. If the matrix multiplication result $\boldsymbol{\sigma}_o^\dagger \boldsymbol{R}_j \boldsymbol{\sigma}_i$ is non-zero, mode j is helicity-resolved Raman active when the polarization of incoming and outgoing light matches $\boldsymbol{\sigma}_i$ and $\boldsymbol{\sigma}_o$, respectively, where the value of $\boldsymbol{\sigma}_i$ and $\boldsymbol{\sigma}_o$ can be chosen from $\boldsymbol{\sigma}_\pm = [1, \pm i, 0]/\sqrt{2}$ [165].

Since the Raman tensor is determined by the vibrational mode, its irreducible representation is the same as that of the vibration mode, and denoted as $\Gamma_j^{(m)}$. If the irreducible representations of the incoming and outgoing light are denoted as Γ_i and Γ_o, respectively, then the irreducible representation of the matrix product $\langle\varepsilon_o|R_j|\epsilon_i\rangle$ can be written as a direct product $\Gamma_o^\dagger \otimes \Gamma_j^{(m)} \otimes \Gamma_i$. If $\langle\varepsilon_o|R_j|\epsilon_i\rangle$ is non-zero, then according to the symmetry of the system, $\Gamma_o^\dagger \otimes \Gamma_j^{(m)} \otimes \Gamma_i$ should contain the total symmetric irreducible representation A [158, 159].

For the 2D lattice systems with D_3 symmetry, the circular polarization of light breaks the mirror symmetries so that the system symmetry is reduced from D_3 to C_3. The left (right) circularly polarized light corresponds to irreducible representation E_a (E_b). The irreducible representations of the collective modes change according to the symmetry point group change given by Fig. 6.8. The mode corresponding to A_1 representation in D_3 corresponds to A

in C_3, while the mode corresponding to E representation in D_3 corresponds to E_a or E_b representation in D_3. Then we can compute the direct product $\Gamma_o^\dagger \otimes \Gamma_j^{(m)} \otimes \Gamma_i$ for all possible choices of Raman active modes and different circular polarizations of incoming and outgoing light. The result is shown in Table 6.1.

Table 6.1: The result of the direct product of irreducible representations $\Gamma_o^\dagger \otimes \Gamma_j^{(m)} \otimes \Gamma_i$ for all possible choices of Raman active modes and different circular polarizations of incoming and outgoing light for 2D materials with D_3 symmetry such as MoS2.

$\Gamma_o^\dagger \otimes \Gamma_j^{(m)} \otimes \Gamma_i$	$\Gamma_j^{(m)} = \mathrm{A}$	$\Gamma_j^{(m)} = \mathrm{E_a}$	$\Gamma_j^{(m)} = \mathrm{E_b}$
$\Gamma_i = \mathrm{E_a},\ \Gamma_o = \mathrm{E_a}$	A	E_a	E_b
$\Gamma_i = \mathrm{E_a},\ \Gamma_o = \mathrm{E_b}$	E_b	A	E_a
$\Gamma_i = \mathrm{E_b},\ \Gamma_o = \mathrm{E_a}$	E_a	E_b	A
$\Gamma_i = \mathrm{E_b},\ \Gamma_o = \mathrm{E_b}$	A	E_a	E_b

Table 6.1 shows that for Raman active vibration A_1 (i.e., $\Gamma_j^{(m)} = A$), only when the polarizations of incoming and outgoing light are the same ($\Gamma_i = \Gamma_o$) can the result of $\Gamma_o^\dagger \otimes \Gamma_j^{(m)} \otimes \Gamma_i$ be A, the total symmetric irreducible representation. For Raman active vibration E (i.e., $\Gamma_j^{(m)} = E_a$ or E_b), only when the polarizations of incoming and outgoing light are opposite ($\Gamma_i = \Gamma_o^\dagger$) can the result of $\Gamma_o^\dagger \otimes \Gamma_j^{(m)} \otimes \Gamma_i$ be A. Therefore, the analysis using irreducible representations shown in Table 6.1 gives the same result as that given by Raman tensor analysis for 2D lattice materials with D_3 symmetry. Therefore, we can use irreducible representations to analyze the 6-particle triangle OM structure.

The circular polarization of light breaks the mirror symmetries of 2D materials with D_6 symmetry, reducing the system symmetry from D_6 to C_6. Left (right) circularly polarized light corresponds to irreducible representation E_{1a} (E_{1b}). The irreducible representations of the collective modes change according to the symmetry point group change given by Fig. 6.9. The mode corresponding to A_1 representation in D_6 corresponds to A in C_6, while the mode corresponding to E_2 representation in D_6 corresponds to E_{2a} or E_{2b} representation in D_3. Therefore we can compute the direct product $\Gamma_o^\dagger \otimes \Gamma_j^{(m)} \otimes \Gamma_i$ for all possible choices of Raman

active modes and different circular polarizations of incoming and outgoing light. The result is shown in Table 6.2.

Table 6.2: The result of the direct product of irreducible representations $\Gamma_o^\dagger \otimes \Gamma_j^{(m)} \otimes \Gamma_i$ for all possible choices of Raman active modes and different circular polarizations of incoming and outgoing light for 2D lattice systems with D_6 symmetry.

$\Gamma_o^\dagger \otimes \Gamma_j^{(m)} \otimes \Gamma_i$	$\Gamma_j^{(m)} = \mathrm{A}$	$\Gamma_j^{(m)} = \mathrm{E_{2a}}$	$\Gamma_j^{(m)} = \mathrm{E_{2b}}$
$\Gamma_i = \mathrm{E_{1a}},\ \Gamma_o = \mathrm{E_{1a}}$	A	$\mathrm{E_{2a}}$	$\mathrm{E_{2b}}$
$\Gamma_i = \mathrm{E_{1a}},\ \Gamma_o = \mathrm{E_{1b}}$	$\mathrm{E_{2a}}$	$\mathrm{E_{2b}}$	A
$\Gamma_i = \mathrm{E_{1b}},\ \Gamma_o = \mathrm{E_{1a}}$	$\mathrm{E_{2b}}$	A	$\mathrm{E_{2a}}$
$\Gamma_i = \mathrm{E_{1b}},\ \Gamma_o = \mathrm{E_{1b}}$	A	$\mathrm{E_{2a}}$	$\mathrm{E_{2b}}$

Table 6.2 shows that for Raman active vibration $\mathrm{A_1}$ (i.e., $\Gamma_j^{(m)} = A$), only when the polarizations of incoming and outgoing light are the same ($\Gamma_i = \Gamma_o$) can the result of $\Gamma_o^\dagger \otimes \Gamma_j^{(m)} \otimes \Gamma_i$ be A, the total symmetric irreducible representation. For Raman active vibration $\mathrm{E_2}$ (i.e., $\Gamma_j^{(m)} = \mathrm{E_{2a}}$ or $\mathrm{E_{2b}}$), only when the polarizations of incoming and outgoing light are opposite ($\Gamma_i = \Gamma_o^\dagger$) can the result of $\Gamma_o^\dagger \otimes \Gamma_j^{(m)} \otimes \Gamma_i$ be A. Therefore, the analysis using irreducible representations shown Table 6.2 gives the same result as that given by Raman tensor analysis for 2D lattice materials with D_6 symmetry. Therefore, we can use irreducible representations to analyze the 7-particle 6-fold symmetric OM structure.

Here is a summary of the analysis in Tables 6.1 and 6.2. The Raman tensor criterion normally used for the helicity-resolved Raman active modes (phonons) in 2D materials is $\langle \varepsilon_o | R_j | \epsilon_i \rangle \neq 0$. We now propose a new criterion based on group representation that $\Gamma_o^\dagger \otimes \Gamma_j^{(m)} \otimes \Gamma_i$ contains the totally symmetric representation, which is found to be equivalent to the Raman tensor criterion for the helicity-resolved Raman active modes (phonons) in 2D materials. Therefore, this group representation we propose can be taken as a common criterion for the helicity-resolved Raman activity of the modes of the 2D materials and the OM systems. Since the group representations of the incoming and outgoing light are 1-dimensional, the group representation criterion can be rewritten as: $\Gamma_j^{(m)}$ contains Γ_{io},

where $\Gamma_{io} = \Gamma_i^\dagger \otimes \Gamma_o$.

Next, we want to find a separate relationship for the three representations: Γ_o, $\Gamma_j^{(m)}$, and Γ_i. In order to find this separate relationship, we apply the angular momentum conservation to the scattering process. Previous studies on the angular momentum conservation mainly focus on the conversion from the change in spin angular momentum between incident and scattered photons to the orbital angular momentum along the rotation mode of the OM structure, which leads to a net torque exerted on the OM structure [15, 17, 84]. Besides rotation, we hypothesize that the incident photons can also transfer momentum to other collective fluctuation modes of OM systems, of which the angular momentum conservation can be written as:

$$m_o + m_{ph} - m_i = 0, \tag{6.3}$$

where $m_i(m_o)$ is the spin angular momentum of the incoming (outgoing) light and m_ph is the angular momentum of the phonon (vibration) mode. We then relate the angular momentum eigenstates of photons and phonons to the irreducible representations of the symmetry point group. We demonstrate this relationship by taking the C_3 and C_6 point group as an example, shown in Table 6.3.

From Table 6.3, we can see two relationships between the irreducible representations and the angular momentum eigenstates. First, the zero angular momentum eigenstate corresponds to the totally symmetric irreducible representation. Second, the sum operation of angular momentum corresponds to the direct product of irreducible representation. In detail, given two angular momentum eigenstate $|m_1\rangle$ and $|m_2\rangle$, their angular momentums are m_1 and m_2, respectively. We can combine $|m_1\rangle$ and $|m_2\rangle$ using direct product to obtain a new state: $|m_1, m_2\rangle = |m_1\rangle \otimes |m_2\rangle$. The angular momentum of this new state is the sum of the angular momentums of the two original states $|m_1\rangle$ and $|m_2\rangle$. In addition, the irreducible representation of the new state is the direct product of the irreducible representation of the

Table 6.3: Demonstration of the relationship between the angular momentum eigenstates and the irreducible representations of point group C_3 and C_6 using analogy to the angular part of the atomic orbitals of hydrogen, $Y_{lm}^{(H)}$, where l is the angular quantum number and m is the magnetic quantum number. σ and m_{ph} are the angular momentum of photon and phonon, respectively. Γ is the corresponding irreducible representation of the angular momentum eigenstates. Only phonons corresponding to vibrations in the x-y plane are taken into consideration.

l in $Y_{lm}^{(H)}$	m in $Y_{lm}^{(H)}$	σ	m_{ph}	Γ in C_3	Γ in C_6	Cartesian polynomial basis
1	+1	$+\hbar$	$+\hbar$	E_a	E_{1a}	$x+iy$
1	0	–	–	A	A	z
1	-1	$-\hbar$	$-\hbar$	E_b	E_{1b}	$x-iy$
2	+2	–	$+2\hbar$	E_b	E_{2a}	$(x+iy)^2$
2	+1	–	–	E_a	E_{1a}	$z(x+iy)$
2	0	–	0	A	A	z^2, x^2+y^2
2	-1	–	–	E_b	E_{1b}	$z(x-iy)$
2	-2	–	$-2\hbar$	E_a	E_{2b}	$(x-iy)^2$

two original states $|m_1\rangle$ and $|m_2\rangle$.

Therefore, Eq. 6.3 can be written in the form of irreducible representations stating that $\Gamma_o \otimes \Gamma_j^{(m)} \otimes \Gamma_i^\dagger$ contains the totally symmetric representation, which is the group representation expression of angular momentum conservation. Since Γ_o and Γ_i are 1-dimensional representations, this group representation expression can be rewritten as $\Gamma_j^{(m)}$ contains $\Gamma_{io}^\dagger$ where $\Gamma_{io} = \Gamma_i^\dagger \otimes \Gamma_o$.

Eventually, we have two constraints on $\Gamma_{io}(= \Gamma_i^\dagger \Gamma_o)$ and $\Gamma_j^{(m)}$. The first constraint is obtained from Raman tensor criterion for helicity-resolved Raman activity, which is $\Gamma_j^{(m)}$ contains Γ_{io}. The second constraint is obtained from angular velocity conservation, which is $\Gamma_j^{(m)}$ contains $\Gamma_{io}^\dagger$. Therefore, $\Gamma_j^{(m)}$ contains both Γ_{io} and $\Gamma_{io}^\dagger$. We now continue the analysis by considering two situations:

1. If the polarization of incoming and outgoing light are the same, $\Gamma_i = \Gamma_o$, which means $\Gamma_{io} = \Gamma_{io}^\dagger$=A is a real representation. In this case, $\Gamma_j^{(m)}$ contains A representation, which is consistent to the helicity-resolved Raman active representation A that corre-

sponds to the same polarization for incoming and outgoing light.

2. If the polarization of incoming and outgoing light are opposite, $\Gamma_i = \Gamma_o^\dagger$, which means Γ_{io} is a complex representation. In this case, $\Gamma_j^{(m)}$ contains $\Gamma_{io} \oplus \Gamma_{io}^\dagger$, meaning that the representation of phonon, $\Gamma_j^{(m)}$ is at least 2-dimensional. Therefore, for 2D materials with symmetry point group D_3 or D_6 that contain Raman active modes that correspond to 2-dimensional irreducible representations, there exist helicity-resolved Raman active modes that manifest opposite polarization for incoming and outgoing light. In contrast, for OM structures with symmetry point group C_3 or C_6 that only contains 1-dimensional irreducible representations, there cannot exist helicity-resolved Raman active modes that manifest opposite polarization for incoming and outgoing light.

In conclusion, by applying Raman tensor criterion and angular momentum conservation, two constraints are obtained for the irreducible representations of the state of incoming light Γ_i, the state of outgoing light Γ_o, and the collective fluctuation mode $\Gamma_j^{(m)}$. The two constraints can be stated as $\Gamma_j^{(m)}$ contains both Γ_{io} and $\Gamma_{io}^\dagger$, where $\Gamma_{io} = \Gamma_i^\dagger \Gamma_o$. When the polarizations of incoming and outgoing light are the same, Γ_{io} is the real totally symmetric representation, which corresponds to the A-type helicity-resolved Raman active modes that exists in both 2D material and OM systems. When the polarizations of incoming and outgoing light are opposite, Γ_{io} is a complex representation, which corresponds to the E-type helicity-resolved Raman active modes that exists only in 2D material but not in OM systems, because the E-type irreducible representation is split into 1-dimensional representations in OM systems due to the chiral broken symmetry.

CHAPTER 7

OPTICAL MATTER SYSTEMS TO WHICH COLLECTIVE MODE ANALYSIS IS NOT DIRECTLY APPLICABLE

The 150 nm dia. Ag nanoparticle OM systems discussed in the previous Chapters have several features in common. First, their stable structures are close to trigonal lattice structures. Second, the probability distribution in their configuration space has only one local maximum in the configuration space while the lifetime of the system fluctuating near the local probability maximum is long. However, not all OM structures have these two properties. In this Chapter, two OM systems that lack these two properties are introduced: the 8-particle structure that manifests "pseudorotation" behavior (8P-SMP) and the 19-particle 6-fold symmetric structure. It is shown that there are two local maxima that are very close to each other in the 8P-SMP, which leads to obstacles to collective mode analysis. It is also shown that the disordered structure, encountered when studying the transition of the 19-particle 6-fold symmetric OM system from an ordered to disordered structure, does not manifest a trajectory that fluctuates near a stable configuration. It appears to be amorphous. Interesting properties of these two systems and the related data analysis methods are introduced in this Chapter.

7.1 8-particle Optical Matter Structure that Manifests Pseudorotation

For most of the OM systems, such as the 6- and 7-particle systems introduced in the previous Chapters, their stable structures are found to lie on 2-dimensional trigonal lattice sites. Therefore, stable structures with 4-fold symmetry do not form in beams with circularly polarized light, although OM structures with 2- and 4-fold symmetry can form in linearly polarized beams [116]. However, there are exceptions to forming trigonal lattice structures of

150 nm dia. Ag nanoparticles in optical beams with circularly polarized light, one of which is the 8-particle structure that manifests pseudorotation (8P-SMP), shown in Fig. 7.1a, an experimental picture of the 8P-SMP taken using dark-field microscopy imaging [3].

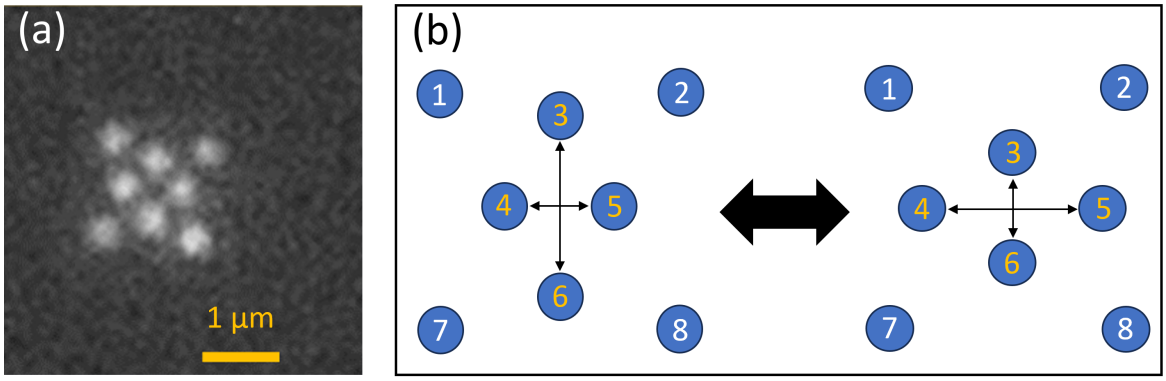

Figure 7.1: The 8-particle structure that manifests pseudorotation (8P-SMP) of the OM system. (a) An experimental dark-field microscopy image of the 8P-SMP. (b) Sketch of the pseudorotation, the structural isomerization between two stable configurations of the 8P-SMP. Both structures are created in sketch to represent the concept of pseudorotation. Neither structure is obtained by accurate computation.

The 4-fold symmetric structure shown in Fig. 7.1a is not the stable configuration of the 8P-SMP. It is a transition state between two structures shown in Fig. 7.1b, which is explained in detail in section 7.1.1. If we neglect the particle labels and do not track the particles, then by rotating one of the structures in Fig. 7.1b by 90 degrees one obtains the other structure. Therefore, we name the structural isomerization shown in Fig. 7.1b pseudorotation in reference to the phenomenon discovered for molecular systems [174, 175].

7.1.1 Important non-equilibrium features

The 8P-SMP has two important features that not only distinguish it from the OM structures on trigonal lattice sites, but also highlight its non-equilibrium nature. The first feature has to do with the temperature dependence of the argument maximum of the probability density in the configuration space. The second non-equilibrium feature of the 8P-SMP is that at any finite temperature, there are frequent transitions between the two relaxed structures

obtained by deterministic simulations.

The two interparticle distances d_1 and d_2 shown in Fig. 7.2a are used as descriptors to characterize the pseudorotation of 8P-SMP, so that the structure on the left in Fig. 7.1b manifests $d_1 < d_2$, while the structure on the right in Fig. 7.1b manifests $d_1 > d_2$. Then, the argument maximum of the probability density function of d_1 and d_2, $p(d_1, d_2)$, should correspond to the stable configuration of 8P-SMP. The plot of $p(d_1, d_2)$ is shown by the 2-dimensional histogram of d_1 and d_2 generated by EDLD simulation trajectories at 50 mW beam power, 27.6 nm Debye screening length, and temperatures ranging from 0.1 K to 200 K, shown in Fig. 7.2b-f.

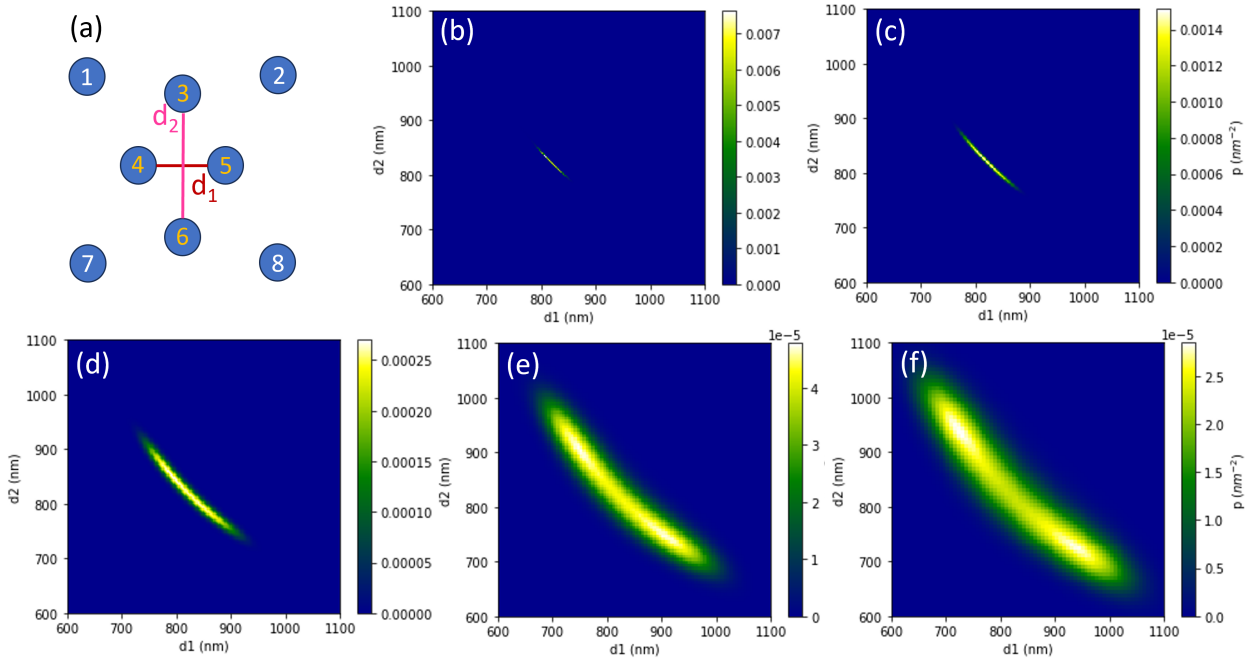

Figure 7.2: Demonstration of the temperature dependence of the argument maximum of the probability density $p(d_1, d_2)$ in the configuration space of the 8P-SMP. (a) The schematic definition of the two interparticle distances d_1 and d_2. (b-f) 2-dimensional histogram of d_1 and d_2 for the 8P-SMP trajectory obtained from EDLD simulation with 50 mW beam power, 27.6 nm Debye screening length, and at temperatures: (b) 0.1 K; (c) 1 K; (d) 10 K; (e) 100 K; (f) 200 K. The viscosity of water at 298 K and 1 atm is used as the solvent viscosity.

Fig. 7.2 shows that we can see that there are two argument maxima of the $p(d_1, d_2)$ when $T \geq 100$ K while there is one argument maximum of the $p(d_1, d_2)$ when $T \leq 10$ K. In

addition, the two argument maxima (i.e., bimodal distribution) of the $p(d_1, d_2)$ at $T \geq 100$ K corresponds to configurations with either d_1 or d_2 close to 900 nm. However, when $T \leq 1$K, the value of $p(d_1, d_2)$ at d_1 or d_2 close to 900 nm is very small as indicated by the dark blue color near $d_1 = 900$ nm and $d_2 = 900$ nm in Fig. 7.2b,c. This indicates that the argument maximum of the probability density of the 8P-SMP depends on temperature. This is a pure non-equilibrium property because the argument maximum of the probability density is independent of temperature for a conservative system; i.e.,

$$\operatorname{argmax} p = \operatorname{argmax} e^{-\beta U} = \operatorname{argmin} U \tag{7.1}$$

where $\beta = 1/k_B T$ and U is the T-independent potential function of the conservative system. Also, the two argument maxima of probability density differs the 8P-SMP from most other OM structures with one argument maximum of probability density.

The second non-equilibrium feature of the 8P-SMP is that as long as the temperature is non-zero, no matter how low the temperature is, there are frequent transitions between the two relaxed structures obtained by deterministic simulations. For conservative systems, there are activation energy barriers E_a between two relaxed structures obtained by deterministic simulations so that the transition between the two relaxed structures will become rare if temperature is significantly lower than E_a/k_B [64]. Fig. 7.3a shows the d_1 and d_2 of a 8P-SMP structure relaxed in the deterministic simulation. Since the d_1 and d_2 converge to different values, there are two relaxed structures for the deterministic simulation; one of them has $d_1 > d_2$ and the other with $d_1 < d_2$. Fig. 7.3b shows the d_1 and d_2 values vs. time of the 8P-SMP structures in a simulation trajectory at $T = 0.1$ K. We can see that there are frequent transitions between the two relaxed structures even if the temperature is very low. Also, the d_1 and d_2 curves that correspond to 25 mW, 50 mW, and 100 mW beam power crosses each other multiple times in Fig. 7.3b, meaning that beam power does not have a significant influence on the frequency of transition between two relaxed structures.

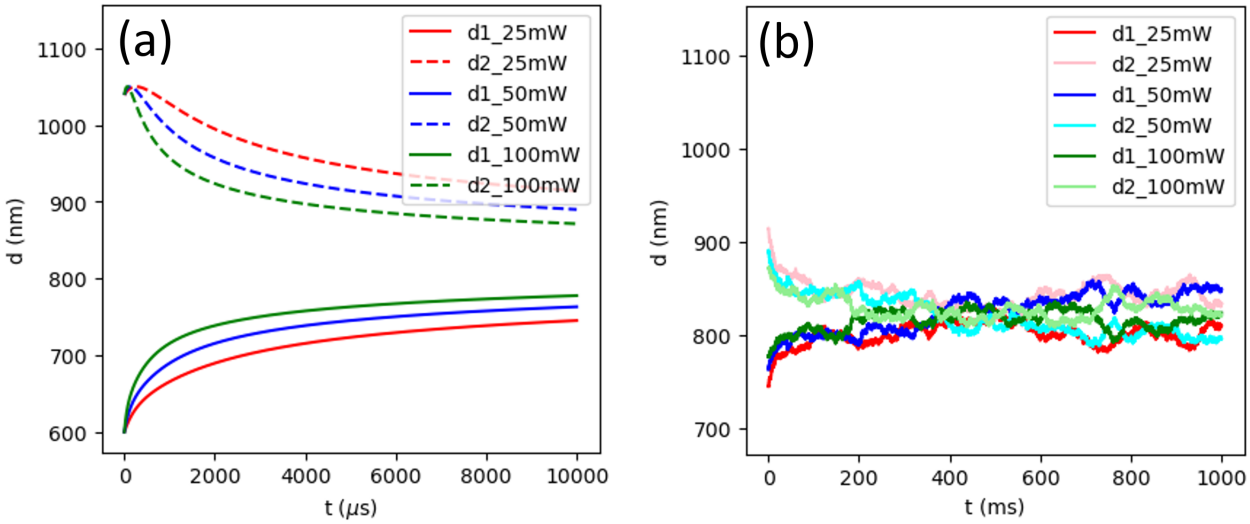

Figure 7.3: Plots of interparticle distances d_1 and d_2 for the 8-particle OM structure that manifests pseudorotation trajectories generated by (a) deterministic simulation; (b) simulation at $T = 0.1$ K.

7.1.2 Collective mode analysis

Due to the frequent transitions between the two argument maxima of the probability density $p(d_1, d_2)$ of the 8P-SMP, the reference structure of 8P-SMP is not determined, so that principal component analysis (PCA) that is applied to OM structures on trigonal lattice sites is not applicable to the 8P-SMP. Therefore, weighted principal component analysis (w-PCA) in proposed in order to analyze the collective modes of the 8P-SMP..

7.1.2.1 Weighted principal component analysis (w-PCA)

The algorithm for w-PCA is listed below. See section 2.3.2 for detailed derivations.

1. Select the initial input of the reference structure as one of the argument maxima of the 8P-SMP.

2. Let $[a_1, b_1, a_2, b_2, \ldots, a_N, b_N]$ ($N = 8$ for 8P-SMP) denote the coordinate of the reference structure. Obtain the trajectory data with the rotation angle of the OM structure aligned to the reference structure by lattice fitting (see section 2.2). The trajectory

data with the rotation angle aligned is denoted as $X_{n\times 2N}$, where n is the length of the trajectory.

3. For each time step in the trajecotry of the OM system with coordinate $\left[x_1^{(k)}, y_1^{(k)}, x_2^{(k)}, y_2^{(k)}, \ldots, x_N^{(k)}, y\right.$ in $X_{n\times 2N}$, compute its weight w_k $(1 \leq k \leq n)$:

$$w_k = \frac{\left|\det J^{(k)}\right|}{\sum_{l=1}^{n}\left|\det J^{(l)}\right|} \tag{7.2}$$

where

$$|\det J^{(k)}| = \frac{\sqrt{\sum_{i=1}^{N}\left(a_i^2 + b_i^2\right)}}{\sqrt{\left[\sum_{i=1}^{N}\left(a_i x_i^{(k)} + b_i y_i^{(k)}\right)\right]^2 + \left[\sum_{i=1}^{N}\left(b_i x_i^{(k)} - a_i y_i^{(k)}\right)\right]^2}} \tag{7.3}$$

4. Compute the weighted average of the data:

$$\overline{X} = \boldsymbol{w}^T X \tag{7.4}$$

5. Compare $\overline{X}$ to the reference structure by lattice fitting. If the difference between them is small, then move on to step 6. If the difference between them is large, then choose $\overline{X}$ as the reference structure input and iterate from steps 2 to 5 until $\overline{X}$ converge; i.e. the 2-norm of the difference between the values of $\overline{X}$ calculated by two adjacent iterations is smaller than an empirically chosen threshold value.

6. Centralize the data:

$$\Delta X = X - \mathbf{1}_n \overline{X} \tag{7.5}$$

where $\mathbf{1}_n$ is the n-dimensional column vector with all entries equal to 1.

7. Compute the covariance matrix:

$$S = \Delta X^T D_w \Delta X \tag{7.6}$$

where D_w is a $n \times n$ matrix with $(D_w)_{ij} = w_i \delta_{ij}$

8. Carry out eigenvalue decomposition for the matrix S. The eigenvalues are the variances, and the eigenvectors are the collective modes.

The w-PCA identifies a linear transformation of the 16 degrees of freedom of the 8-particle system constrained to a 2D plane into a new basis of 16 orthogonal PCs (taken as collective modes) arranged in order of decreasing configurational variance. The 16 PCs shown in Fig. 7.4 are ranked in order of largest to smallest configurational standard deviations. PCs 1, 2, and 16 correspond to rigid body transformations: PC 1 and 2 correspond to translations and PC 16 to rotation. The remaining PCs correspond to the collective fluctuations around the reference configuration, the converged $\overline{X}$ given by Step 5 of the w-PCA algorithm. This converged reference structure is shown by the 8 circles within each panel of Fig. 7.4. This reference structure is actually the 4-fold symmetric transition state between the two argument maxima of the probability density in the 8P-SMP configuration space.

Among all 16 PCs, PC 3 corresponds to the structural isomerization (pseudorotation) between the two argument maxima of the probability density in the 8P-SMP configuration space. This means that the probability distribution along the pseudorotation mode (PC 3) has two peaks, which is illustrated by Fig. 7.5a where we can also see that the probability distributions along all the other modes are Gaussian at $T = 200$ K. In fact, the probability distributions of all the other modes superimpose and are shown as the curve in red. Fig. 7.5b shows that the standard deviation along pseudorotation (PC 3) (blue), σ_{psrot}, is proportional to $T^{0.25}$. On the other hand, the standard deviations along the other PCs are proportional to $T^{0.5}$. The exponent $T^{0.5}$ is the relation between standard deviation and temperature

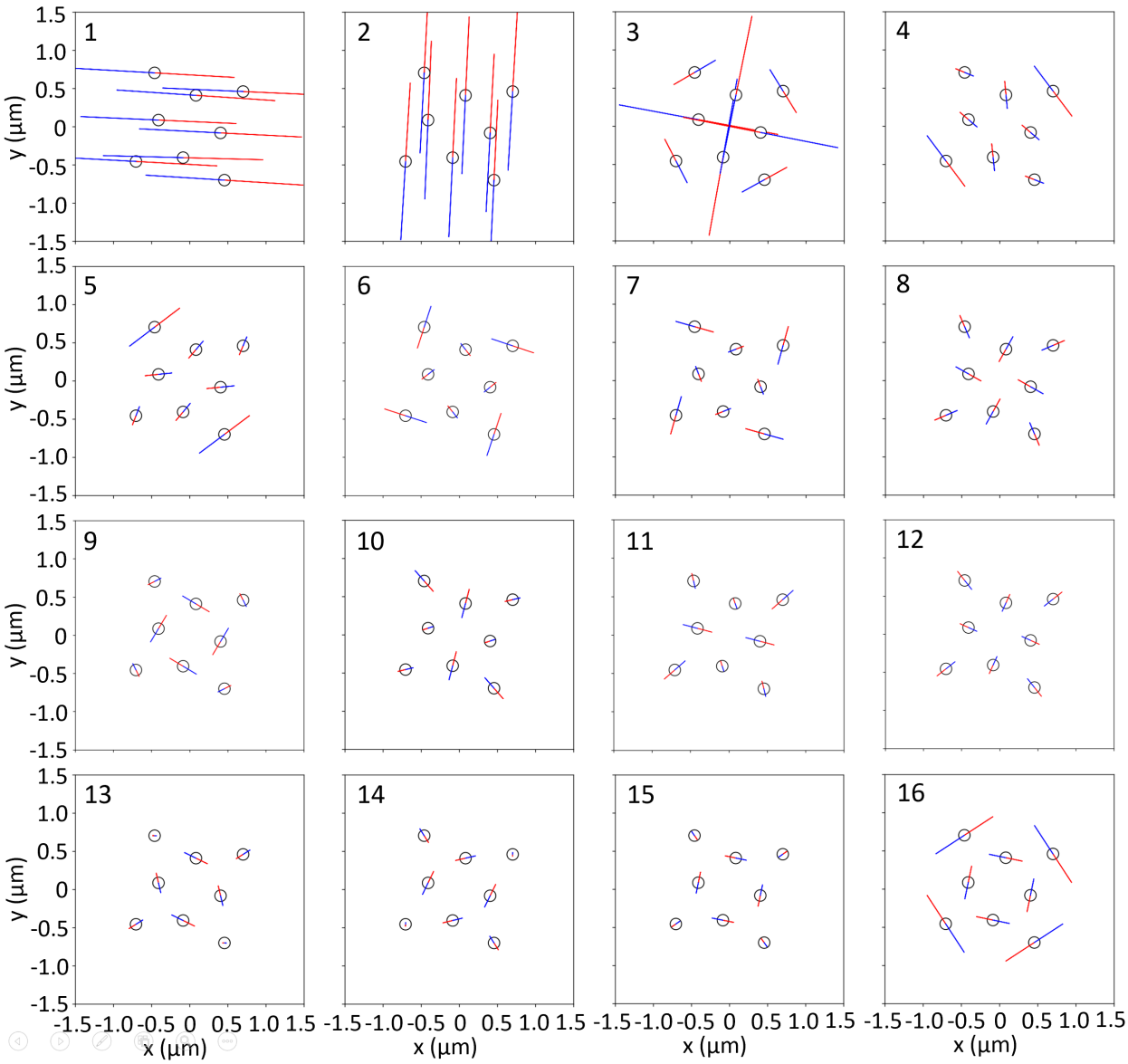

Figure 7.4: Principal components (PCs) of the 8-particle OM structure that manifests pseudorotation computed by weighted principal component analysis (w-PCA). The PCs are labeled in descending order of their configurational standard deviations. The colored solid lines depict the directions and magnitudes of the collective particle motions, and the color defines the sense (i.e. phase) of the motion; i.e., particles simultaneously move in the indicated directions for the same color. PCs 1 and 2 correspond to rigid translation and PC 16 to rigid rotation. For the non-rigid transformations, PCs 3 to 15, the length of the solid lines is proportional to their corresponding configurational standard deviation. Note the large variance of the central 4 nanoparticles of mode 3 that correspond to d_1 and d_2

expected for Boltzmann distribution of a quadratic potential. Fig. 7.5b also shows that σ_{psrot} is larger than the standard deviation of translation at $T < 50$ K and that σ_{psrot} is less than the standard deviation of translation at $T > 50$ K.

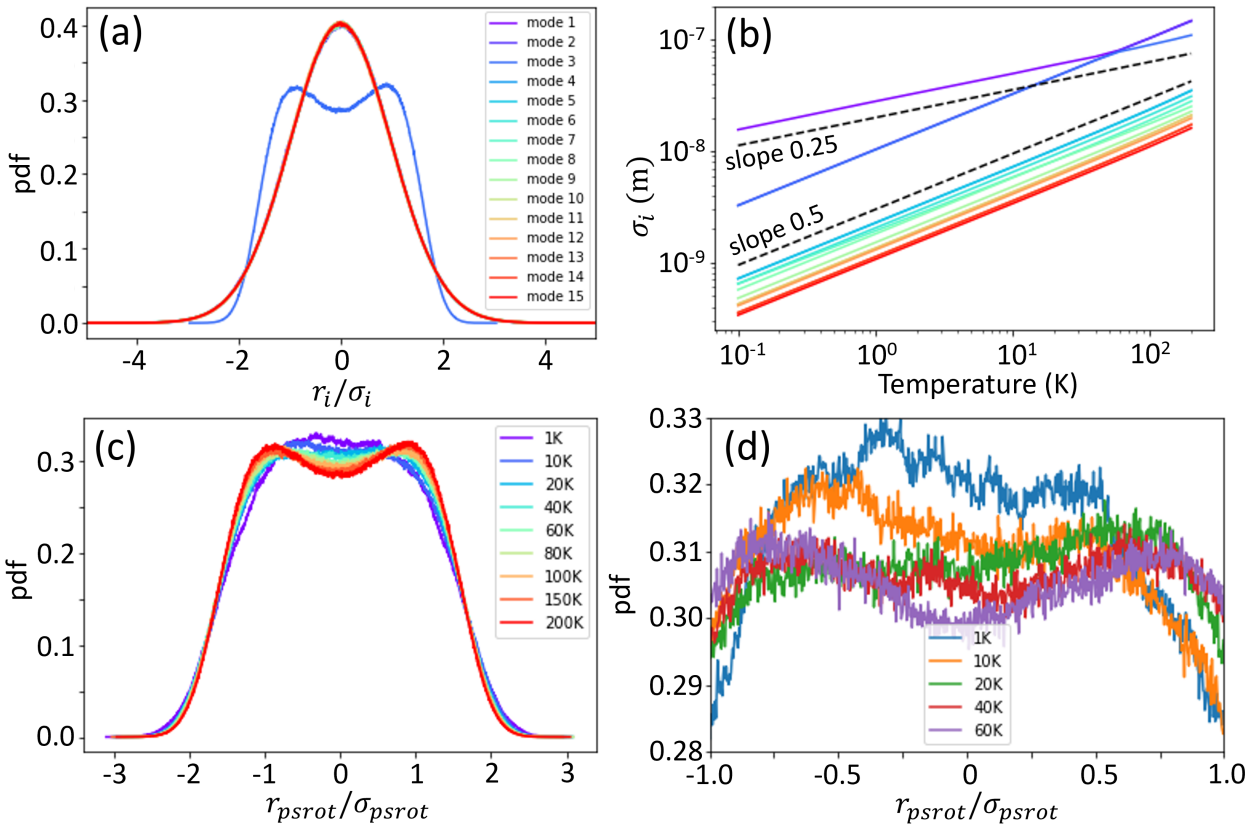

Figure 7.5: Statistical results of the 15 principal components (PCs) (with rotation excluded) of the 8-particle structure that manifests pseudorotation (8P-SMP) of OM. (a) Plot of the probability distribution along the 15 PC coordinates normalized by standard deviation for the 8P-SMP at 200 K. The curves corresponds to the modes other than mode 3 superimpose each other so that only 2 colors are shown. (b) The logarithm plot of the standard deviations along the 15 PC coordinates as a function of temperature with two dashed lines representing slopes 0.25 and 0.5. Panels (a) and (b) share the same color labels for the w-PCA modes. (c-d) Plot of the probability distribution along PC 3 with displacement normalized by standard deviation along PC 3 for the 8P-SMP at various temperatures. (d) An expanded view of (c).

Fig. 7.5c,d shows the T-dependence of the probability distribution along the pseudorotation mode, p_{psrot}, plotted against the ratio between structural displacement and σ_{psrot} along the pseudorotation mode. We can see that when T decreases, the two peaks of p_{psrot} turn into a single plateau in the vicinity of the origin. When T further decreases, the single

plateau of p_{psrot} becomes narrower. This is consistent with the frequent transitions between the two argument maxima of the probability density in the 8P-SMP configuration space at very low T. We can see from Fig. 7.5d that the threshold temperature is between 40 K and 60 K for the transition between the double-peak distribution and the single-plateau distribution along the pseudorotation mode. This is consistent with the threshold $T = 50$ K at which σ_{psrot} is equal to the standard deviation of translation.

7.1.2.2 Explanation of the temperature dependence of the argument maximum of the probability density

The analysis of Fig. 7.5c,d shows the temperature dependence of the argument maxima of probability density. In this section, the detailed derivation of the temperature dependence of the argument maxima of probability density is provided using the Fokker-Planck equation and force field decomposition. First of all, the Fokker-Planck equation, which describes the time evolution of the probability distribution, has the form of [144]:

$$\frac{\partial p}{\partial t} = D[\Delta p - \beta \nabla \cdot (p\boldsymbol{F})] \tag{7.7}$$

where p is probability distribution, t is time, D is diffusion coefficient, $\beta = 1/k_B T$, and $\boldsymbol{F}$ is the external force field that is non-conservative for the OM system. Eq. 7.7 can be rewritten as:

$$\frac{\partial p}{\partial t} = -\nabla \cdot (\beta D p\boldsymbol{F} - D\nabla p) = -\nabla \cdot \boldsymbol{j}, \tag{7.8}$$

where

$$\boldsymbol{j} = \beta D p\boldsymbol{F} - D\nabla p = \beta D p \left(\boldsymbol{F} - \frac{\nabla \ln p_2}{\beta_2} \right) \tag{7.9}$$

is the probability density current of the system.

Now let $p_1(\boldsymbol{x})$ be the stationary distribution at β_1 and define the effective potential $U(\boldsymbol{x})$

so that $p_1(\boldsymbol{x})$ has the form of Boltzmann distribution:

$$p_1 = Z_1 e^{-\beta_1 U}, \tag{7.10}$$

where Z_1 is the normalization constant. Since p_1 is stationary, if the system evolves with an initial condition $p(t=0,\boldsymbol{x}) = p_1(\boldsymbol{x})$, when according to Eq. 7.7 we have:

$$\begin{aligned} 0 = \frac{\partial p}{\partial t} &= D[\Delta p_1 - \beta_1 \nabla \cdot (p_1 \boldsymbol{F})] \\ &= D[\Delta p_1 + \beta_1 \nabla \cdot (p_1 \nabla U)] - \beta_1 D \nabla \cdot [p_1(\boldsymbol{F} + \nabla U)] \\ &= -\beta_1 D \nabla \cdot [p_1(\boldsymbol{F} + \nabla U)] = -\nabla \cdot \boldsymbol{j}_1 \end{aligned} \tag{7.11}$$

Now let the temperature change to β_2 and, again, define $p_2(x)$ according to Boltzmann distribution:

$$p_2 = Z_2 e^{-\beta_2 U} \tag{7.12}$$

where Z_2 is the normalization constant. Apparently, if the system is conservative, we have $\boldsymbol{F} = -\nabla U$, so that according to Boltzmann distribution theory [64], $p_2(\boldsymbol{x})$ is also stationary, which means that $p_1(\boldsymbol{x})$ and $p_2(\boldsymbol{x})$ share the same argument maximum that is also the argument minimum of $U(\boldsymbol{x})$. This is consistent with the fact that the argument maximum of the probability distribution is independent of temperature. Therefore, we want to find out if this is still the case for the non-conservative system.

Let $Z = Z_2/Z_1$ and the initial condition be $p(t=0,\boldsymbol{x}) = p_2(\boldsymbol{x})$. Then, evolving the system using Eq. 7.7 gives:

$$
\begin{aligned}
\frac{\partial p}{\partial t} &= D[\Delta p_2 - \beta_2 \nabla \cdot (p_2 \boldsymbol{F})] \\
&= D[\Delta p_2 + \beta_2 \nabla \cdot (p_2 \nabla U)] - \beta_2 D \nabla \cdot [p_2(\boldsymbol{F} + \nabla U)] \\
&= -\beta_2 D \nabla \cdot [p_2(\boldsymbol{F} + \nabla U)] = -\beta_2 D \nabla \cdot \left[Z e^{(\beta_1 - \beta_2)U} p_1 (\boldsymbol{F} + \nabla U) \right] \\
&= -\beta_2 D Z \left\{ e^{(\beta_1 - \beta_2)U} \nabla \cdot [p_1(\boldsymbol{F} + \nabla U)] + p_1(\boldsymbol{F} + \nabla U) \cdot \nabla e^{(\beta_1 - \beta_2)U} \right\}
\end{aligned}
\tag{7.13}
$$

From Eq. 7.11, we know that $\nabla \cdot [p_1(\boldsymbol{F} + \nabla U)] = 0$, so that Eq. 7.13 can be rewritten as:

$$
\begin{aligned}
\frac{\partial p}{\partial t} &= -\beta_2 D Z p_1 (\boldsymbol{F} + \nabla U) \cdot \nabla e^{(\beta_1 - \beta_2)U} \\
&= -\beta_2 D Z p_1 (\boldsymbol{F} + \nabla U) \cdot e^{(\beta_1 - \beta_2)U} (\beta_1 - \beta_2) \nabla U \\
&= \beta_2 D p_2 (\beta_2 - \beta_1)(\boldsymbol{F} + \nabla U) \cdot \nabla U
\end{aligned}
\tag{7.14}
$$

Next, we rewrite $\nabla \cdot [p_1(\boldsymbol{F} + \nabla U)] = 0$ as:

$$
\begin{aligned}
\nabla \cdot [p_1(\boldsymbol{F} + \nabla U)] &= 0 \\
p_1 \nabla \cdot (\boldsymbol{F} + \nabla U) + (\boldsymbol{F} + \nabla U) \cdot \nabla p_1 &= 0 \\
-(\boldsymbol{F} + \nabla U) \cdot \frac{\nabla p_1}{p_1} &= \nabla \cdot (\boldsymbol{F} + \nabla U) \\
\beta_1 (\boldsymbol{F} + \nabla U) \cdot \nabla U &= (\boldsymbol{F} + \nabla U) \cdot \nabla U \\
(\boldsymbol{F} + \nabla U) \cdot \nabla U &= \beta_1^{-1} (\boldsymbol{F} + \nabla U) \cdot \nabla U
\end{aligned}
\tag{7.15}
$$

Plugging Eq. 7.15 into Eq. 7.14 gives:

$$
\frac{\partial p}{\partial t} = \frac{\beta_2}{\beta_1} D(\beta_2 - \beta_1) \nabla \cdot (\boldsymbol{F} + \nabla U) \tag{7.16}
$$

According to the force field decomposition, $\boldsymbol{F} + \nabla U$ is the non-conservative part of the external force field $\boldsymbol{F}$ (see section 5.4.5 for detailed derivation of force field decomposition). Therefore, $\nabla \cdot (\boldsymbol{F} + \nabla U)$ is the divergence of the non-conservative part of the force field,

which is not necessarily zero for a non-conservative system. As a consequence, $p_2(\boldsymbol{x})$ is not stationary at β_2. This means that the stationary distribution at β_2, $\tilde{p}_2 = \tilde{Z}_2 e^{-\beta_2 \tilde{U}}$ corresponds to an effective potential $\tilde{U}(\boldsymbol{x})$ that is different from $U(\boldsymbol{x})$, the effective potential corresponding to the stationary distribution at β_1. In other words, the effective potential depends on temperature, so that the argument maximum of the stationary distribution depends on temperature.

7.1.3 Koopman analysis

Koopman analysis is a spectral representation method that describes the dynamics of a system using a low dimensional space spanned by the eigenfunctions of the Koopman operator with longest decay time scales [176]. This method arises by solving the Fokker-Planck equation of the time evolution of the system probability distribution (Eq. 7.7) using spectral decomposition. First, Eq. 7.7 is rewritten using the operator $\mathcal{L}$:

$$\frac{\partial p}{\partial t} = D[\Delta p - \beta \nabla \cdot (p\boldsymbol{F})] = \mathcal{L}^* p \tag{7.17}$$

$$\mathcal{L}^* \phi_k(\boldsymbol{x}) = \lambda_k \phi_k(\boldsymbol{x}) \tag{7.18}$$

where $\mathcal{L}^*$ is defined in Eq. 7.17 and $\mathcal{L}$ is the adjoint operator of $\mathcal{L}^*$; λ_k and ϕ_k are the eigenvalues and left eigenfunctions of $\mathcal{L}$, respectively. Then, the solution of Eq. 7.7 can be written as:

$$p(\boldsymbol{x}, \tau) = e^{\tau \mathcal{L}^*} p(\boldsymbol{x}, 0) = \sum_k c_k e^{\lambda_k \tau} \phi_k(\boldsymbol{x}) \tag{7.19}$$

where $p(\boldsymbol{x}, 0)$ is the initial condition. The Koopman operator is defined as $\mathcal{K}_\tau = e^{\tau \mathcal{L}}$ so that its adjoint operator $e^{\tau \mathcal{L}^*}$ corresponds to time evolution by operating on the system probability distribution according to Eq. 7.19. Eq. 7.19 shows that the time evolution of the probability distribution can be written as a linear combination of the left eigenfunctions of

$\mathcal{K}_\tau$, $\phi_k(\boldsymbol{x})$. Also, the eigenvalues λ_k corresponds to the reciprocal decay time scale of each mode.

While the left eigenfunctions of $\mathcal{K}_\tau$, $\phi_k(\boldsymbol{x})$ characterizes the time evolution of probability distribution, the right eigenfunctions of $\mathcal{K}_\tau$, denoted as $\psi(\boldsymbol{x})$, characterize the time evolution of the expectation of system observables [177]. The Koopman analysis uses the first m right eigenfunctions with longest decay time scales, $\psi_k(\boldsymbol{x})$ $(k = 1, 2, \cdots, m)$, as the basis to span a m-dimensional latent space to describe the slow modes of the system dynamics. Because the basis functions $\{\psi_k(\boldsymbol{x})\}$ correspond to the longest time scales, the variational approach to Markov process (VAMP) is normally adopted to generate the $\{\psi_k(\boldsymbol{x})\}$ [47, 176, 177]:

$$\begin{gathered}\psi_k\,(k \geq 1) = \underset{\psi \in \Omega_k}{\operatorname{argmax}} \frac{\langle \psi | \mathcal{K}_\tau \psi \rangle}{\langle \psi | \psi \rangle} = \underset{\psi \in \Omega_k}{\operatorname{argmax}} \frac{\langle \psi(t=0) | \psi(t=\tau) \rangle}{\langle \psi | \psi \rangle} \\ \Big(\psi_0 = \mathbf{1}, \quad \Omega_k = \operatorname{span}^{\perp}\{\psi_0, \psi_1, \cdots, \psi_{\mathrm{k}-1}\} \Big)\end{gathered} \tag{7.20}$$

where $\langle f_1 | f_2 \rangle$ denotes the ensemble mean of the inner product between the two functions f_1 and f_2. The term $\langle \psi(t=0) | \psi(t=\tau) \rangle$ is computed by means of trajectory sampling [47]. One way of carrying out VAMP is to use a neural network to fit the $\{\psi_k\}$ by minimizing the loss function in the form of VAMP-2 score [46, 47, 176–178]:

$$L = g\left(\sum_{k=1}^{m} \left(\frac{\langle \psi(t=0) | \psi(t=\tau) \rangle}{\langle \psi | \psi \rangle} \right)^2 \right) \tag{7.21}$$

where g is a decreasing function. This neural network approach to solve VAMP is termed as VAMPnet [46, 47, 176, 178].

The VAMPnet approach described above operates on a system that obeys detailed balance because orthogonality between the eigenfunctions of $\mathcal{K}_\tau$ is required for carrying out the variational approach in Eq. 7.20. However, eigenfunctions of $\mathcal{K}_\tau$ of a (non-equilibrium) system that breaks detailed balance are not necessarily orthogonal to each other. Therefore, given the non-conservativeness of the OM system, we use singular value decomposition

(SVD) [177,179] instead of eigenvalue decomposition to carry out VAMPnet, which gives the orthogonal right singular vectors of $\mathcal{K}_\tau$.

For the 8P-SMP OM system, we take $m = 5$ and compute the first 5 non-trivial right singular vectors, ψ_k ($k = 1 \cdots 5$) using VAMPnet with lag time $\tau = 10$ μs at various temperatures. The computation is accomplished by means of the state-free non-reversible VAMPnet (SNRV) python package developed by the Ferguson lab [180]. We compute the correlation between the $\{\psi_k\}$ and three PC collective coordinates including two translations (r_1 and r_2) shown in Figs. 7.6a,b and the pseudorotation (r_3) shown in Fig. 7.6c. The two translations mostly correlate with ψ_4 and ψ_5 at 1 K, with ψ_2 and ψ_3 at low T but above 1 K, and with ψ_1 at high T. The pseudorotation mostly correlates with ψ_1 at low and with ψ_3 at high T.

Figs. 7.6d shows the implied time scale, $|\lambda_k^{-1}|$, as a function of temperature; the dots are colored according to the PC mode that the corresponding ψ_k correlates with. We can see that the implied time scales of the translations are almost independent of temperature, while the implied time scale of the pseudorotation is proportional to $T^{-0.48}$ according to the linear fit shown by the green dashed line. The implied time scales of translation and pseudorotation cross at around 50 K, which is also the threshold temperature at which σ_{psrot} is equal to the standard deviation along translation. The implied time scale of translation and pseudorotation are very close in the vicinity of the crossing temperature, which causes coincident degeneracy to their corresponding ψ_k's. Therefore, there exist arbitrary choices of two basis functions in the degenerate eigen-subspaces. These arbitrary choices of basis functions in the degenerate eigen-subspaces mix the ψ_k's that correlate with pseudorotation and translation. These mixed modes are shown as purple dots in Fig. 7.6d.

In conclusion, the singular vectors ψ_k's given by SNRV can characterize the time scale of the dynamics along the collective fluctuation modes for the 8P-SMP OM system. In addition, the trend of the time scales implied by SNRV is consistent with the variances given

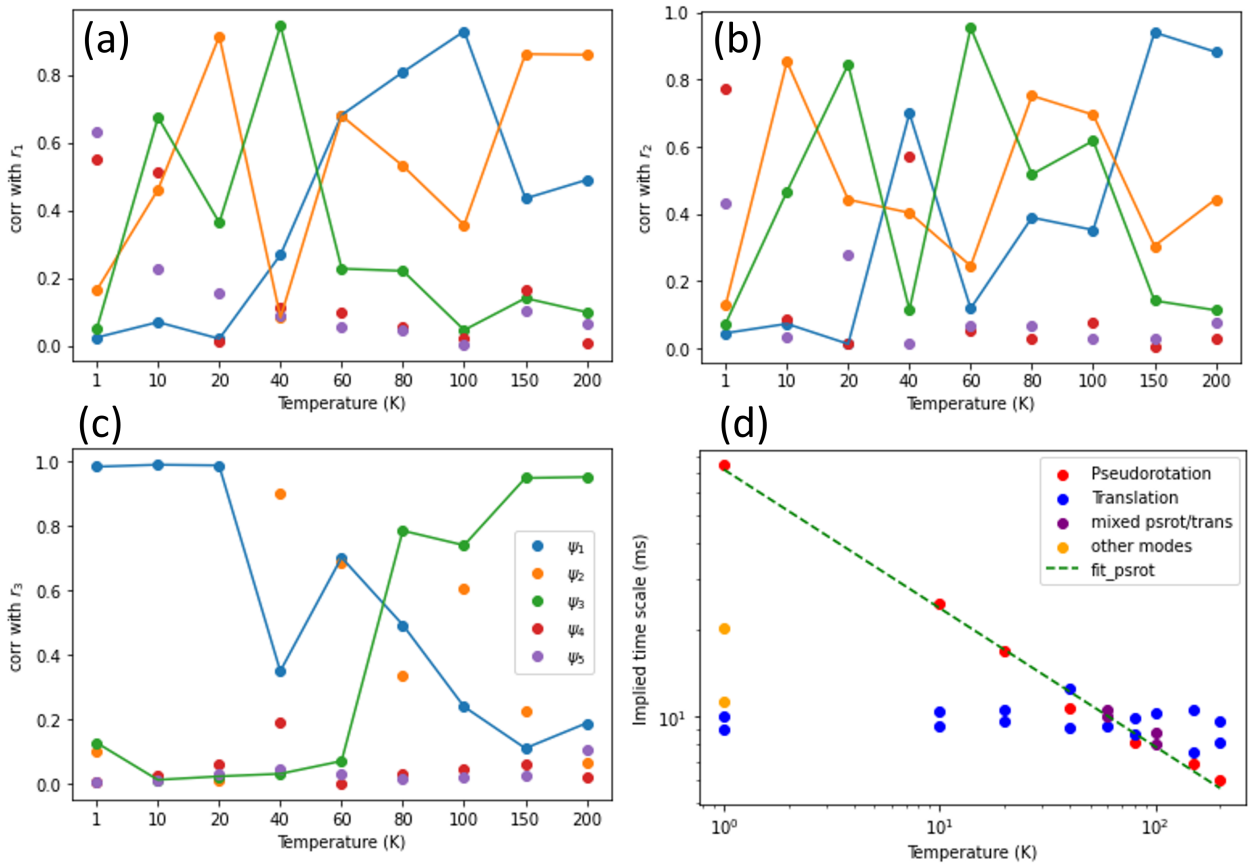

Figure 7.6: Interpretation and implied time scales of the right singular functions $\{\psi_k\}$ ($1 \leq k \leq 5$) obtained by state-free non-reversible VAMPnet (SNRV). (a-c) Correlations between the ψ_k and (a,b) the two translation modes (c) the pseudorotation mode of the 8P-SMP. (d) Plot of the temperature dependence of the implied time scales corresponding to the right singular functions $\{\psi_k\}$ that correlate with the translation and pseudorotation modes of the 8P-SMP. The green dashed line (slope = -0.48) is the linear fit of the implied time scale of the pseudorotation mode as a function of temperature (on a logarithm scale).

by the w-PCA that is used to generate collective fluctuation modes. Therefore, SNRV is promising for further study on the non-equilibrium features of the 8P-SMP by looking into other collective fluctuation modes in the system.

7.2 19-particle 6-fold symmetric Optical Matter Structure

7.2.1 Results

The 19-particle 6-fold symmetric OM structure is a stable structure of the OM system of 150 nm dia. Ag nanoparticles. The inset in Fig. 7.7c shows an experimental image of the 19-particle 6-fold symmetric structure obtained by dark-field microscopy. The 19 particles in this OM structure can be divided into three categories: one center particle, six middle-layer particles, and twelve outer-layer particles. A 10-million-frame EDLD simulation trajectory is obtained at each beam power where the ordered 19-particle 6-fold symmetric OM array is the initial structure. The simulation is carried out at 300 K and 1 μs simulation time step.

At 300 K, the 19-particle OM structure is in its ordered state at 60 mW beam power, illustrated by the sharp peaks of the pairwise distance distribution shown in Fig. 7.7c, which is the feature of lattice structures. The 19-particle OM structure becomes disordered at 20 mW beam power, illustrated by the broader peaks of the pairwise distance distribution shown in Fig. 7.7a. Fig. 7.7b shows the pairwise distance distribution at 40 mW, which is the transition point between ordered and disordered structures.

Fig. 7.7d shows the total scattering-cross section normalized by $N\pi R^2$, where $R = 75$ nm is the radius of the nanoparticles and N is the number of particles. The total scattering-cross section has a peak close to 800 nm wavelength for the ordered structure at high beam power. This spectroscopic feature is the collective resonance of the trigonal lattice OM structure [17]. When the beam power decreases from 60 mW to 20 mW, the 19-particle OM structure turns from ordered to disordered, and the collective resonance peak at 800 nm wavelength diminishes while the single-particle scattering peak at 600 nm wavelength [17] increases.

Fig. 7.8a shows the probability distribution of the ratio (mean angular velocity)/(beam power) (averaged in 1 ms) at various beam powers. We can see that when beam power

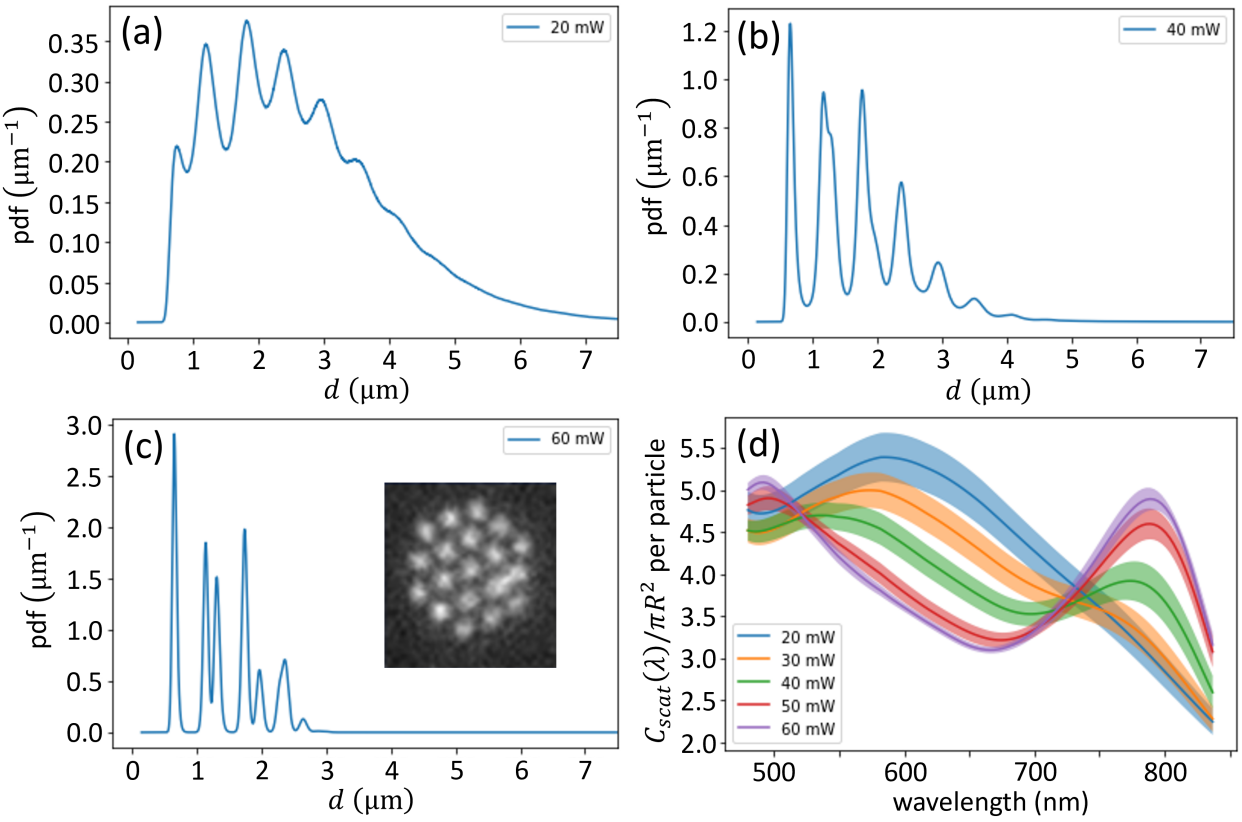

Figure 7.7: The pairwise distance distribution and the total scattering cross-section of the 19-particle 6-fold symmetric OM structure at 300 K and various beam powers. (a-c) Plot of the pairwise distance distribution of the nanoparticles in the 19-particle 6-fold symmetric OM structure at 300 K and (a) 20 mW (b) 40 mW (c) 60 mW beam power. The inset in panel c is the dark-field microscopy image of the 19-particle 6-fold symmetric OM structure at 300 K. (d) Plot of the normalized total scattering cross-section as a function of wavelength of the incident optical beam and 6 different beam powers (colors). The viscosity is taken as the viscosity of water at 300 K and 1 atm. The Debye-screening length is 100 nm.

increases from 20 mW to 40 mW, where the structure shifts from disordered to an intermediate state between disordered and ordered structures, the mean angular velocity becomes positive and reaches the largest positive value at 40 mW. When the beam power is increased from 40 mW to 60 mW, where the structure shifts to becoming ordered, the mean angular velocity changes from positive to negative. From Fig. 7.8b, we can see that if the beam power increases above 60 mW the mean angular velocity becomes more and more negative. Therefore, the angular velocity can characterize the transition between disordered and or-

dered structure of the 19-particle OM system A similar finding has been reported recently in a study using Ag nanoparticles [74]. Fig. 7.8b also shows that the standard deviation of the magnitude of the angular velocity decreases as the beam power decreases, which means that the mean angular velocity is a reliable metric for characterizing disordered-to-order transition.

Because the initial configuration of each simulation trajectory is the ordered 19-particle 6-fold symmetric OM structure, it takes some time for the statistical quantities of the system to relax for the trajectories obtained at beam powers less than 60 mW. In order to study the relaxation of the mean angular velocity, we evenly divide the 10-million-frame simulation trajectory at each beam power into 10 1-million-frame segment and compute the probability distribution of the ratio between angular velocity and beam power for each segment separately. The result is shown in Fig. 7.8c-f, where we can see that the distribution of angular velocity has a relaxation for low beam powers, but the angular velocity does not relax for beam powers at 50 mW and higher because the simulation starts at the ordered structure that is stable under high beam powers.

In addition to the angular velocity of the structure, we also compute the root mean square displacement (RMSD) as a function of lag time τ for each nanoparticle:

$$\begin{aligned} RMSD_i(\tau = m\Delta t) &= \Big\langle [(x_i(t+\tau) - x_i(t)]^2 + [(y_i(t+\tau) - y_i(t)]^2 \Big\rangle^{\frac{1}{2}} \\ &= \left\{ \frac{1}{n-m} \sum_{k=1}^{n-m} [x_i((k+m)\Delta t) - x_i(k\Delta t)]^2 + [y_i((k+m)\Delta t) - y_i(k\Delta t)]^2 \right\}^{\frac{1}{2}} \end{aligned} \tag{7.22}$$

where RMSD_i denotes the RMSD of the i^{th} particle, $n = 10^7$ is the total number of frames in each trajectory, $\Delta t = 1\,\mu$s is the simulation time step, and $[x_i(t), y_i(t)]$ is the 2D Cartesian coordinate of the i^{th} particle at time t. From this, the average RMSD is computed over all particles:

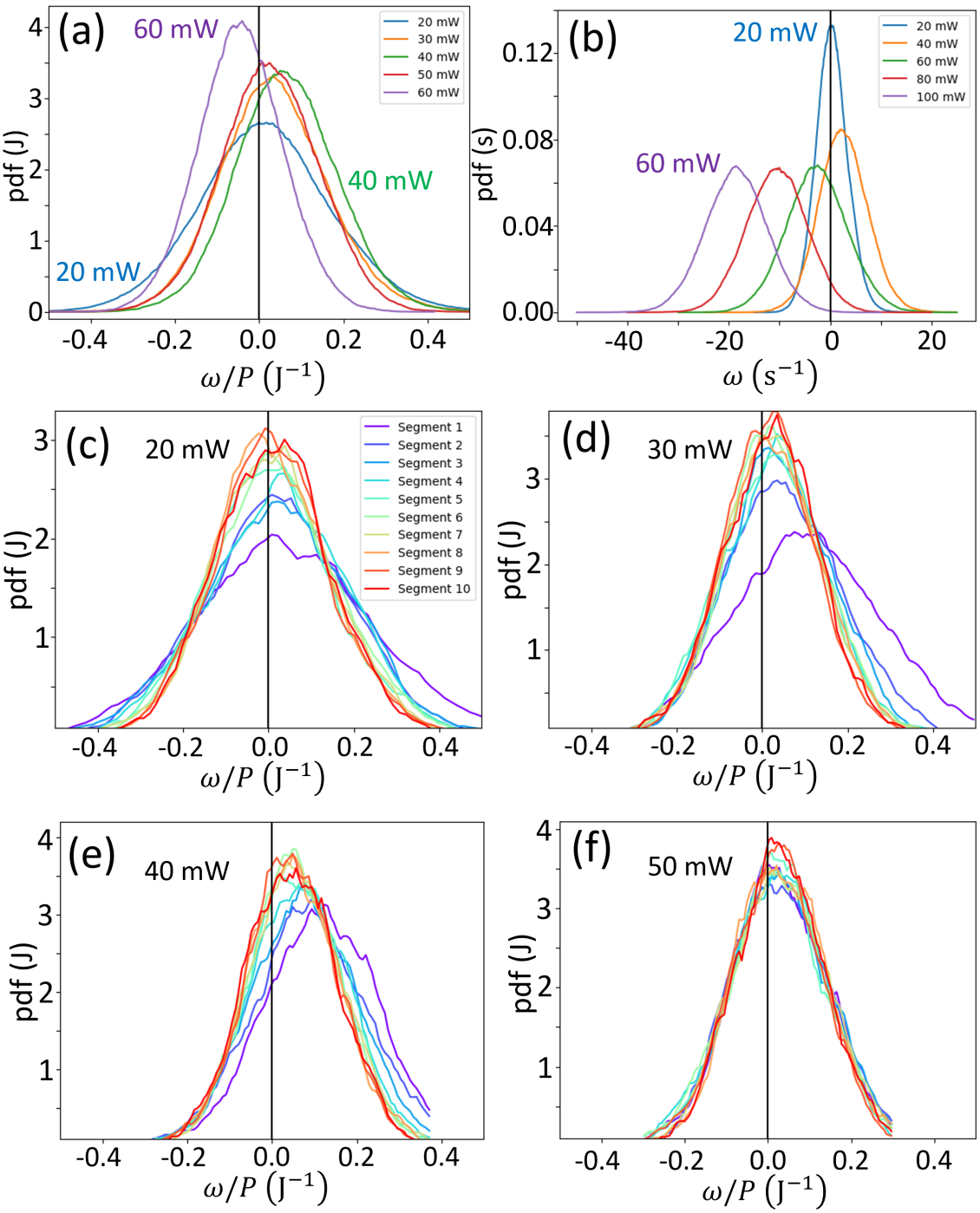

Figure 7.8: Analysis of the mean angular velocity (averaged in 1 ms) of the 19-particle OM system. (a-b) The probability distribution of (a) ratio between mean angular velocity and beam power (b) mean angular velocity at various beam powers P. (c-f) the probability distribution of the ratio between mean angular velocity and by beam power of different 1-million-frame segments of a 10-million-frame trajectory initiated at the ordered 19-particle 6-fold symmetric structure at (c) 20 mW (d) 30 mW (e) 40 mW (f) 50 mW beam power.

$$\overline{RMSD} = \sum_{i=1}^{N} RMSD_i \tag{7.23}$$

where $n = 19$ is the number of particles in the system. The average RMSD is plotted as a function of lag time τ for various beam powers, shown in Fig. 7.9a. We can see that the plots are linear with a slope of 0.5 when lag time τ is small for all beam powers. However, at large lag times, the plot is still linear for small beam powers, but bends down at large beam powers. This means that the random motions of the particles are more diffusive (Brownian) for low beam powers when the structure is disordered. When the structure becomes more ordered when beam power increases, where the slope of average RMSD drops below 0.5 in the logarithm scale. This transition to apparent subdiffusive behavior reflects the fact that the particle motions are more constrained (confined) by electrodynamic interactions that are stronger than $k_B T$.

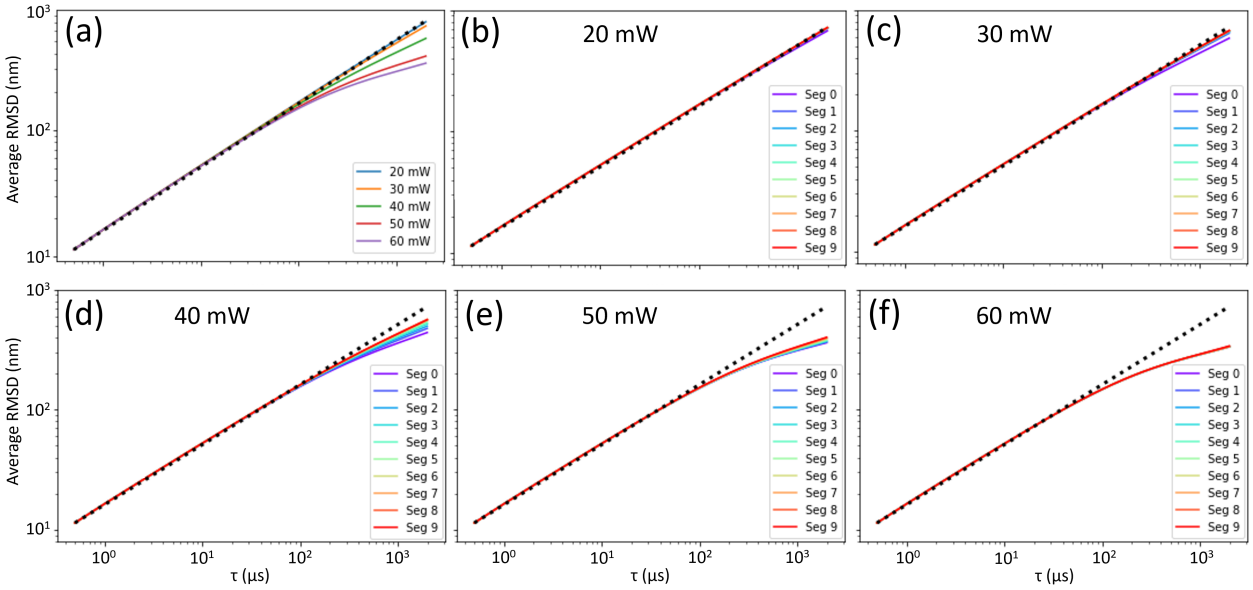

Figure 7.9: Plots of the root mean square displacement (RMSD) averaged over the nanoparticles as a function of lag time at various beam powers. (a) Plot of average RMSD of the 10-million frame simulation trajectory at each beam power as a function of lag time. (b-f) Plot of average RMSD of the 10 evenly divided 1-million-frame segments of the simulation trajectory as a function of lag time at (b) 20 mW (c) 30 mW (d) 40 mW (e) 50 mW (f) 60 mW beam power.

In order to study the relaxation of the system in the simulation, we computed the average RMSD for the 10 evenly divided 1-million-frame segments of the 10-million-frame trajectory at each beam power. The results are shown in Fig. 7.9b-f. We can see that when the beam power is 20 mW or 60 mW, i.e., when the OM structure is in completely disordered or ordered, respectively, the relaxation is fast since the RMSD curve of the 1^{st} segment overlaps with the RMSD curve of the 10^{st} segment. On the other hand, the relaxation becomes slower when the beam power is closer to 40 mW; this is the case when the OM system is in an intermediate condition between disordered and ordered.

In order to study the dynamics of the particles in different parts of the 19-particle OM structure, we classify the 19 particles into three types: 1 center particle, 6 middle layer particles, and 12 outer layer particles. These three classes are illustrated in the insets in Fig. 7.10a-c. The main results shown in Figs. 7.10a-c are the derivatives with respect to lag time τ of the average mean square displacement (MSD) (i.e., square of RMSD) scaled by $4D$. Here τ is lag time and $D = k_B T/6\pi\eta R$ is the single-particle diffusion coefficient where η is viscosity and R is particle radius (75 nm).

Since the MSD of a purely diffusive particle in 2D space is $4D\tau$, the derivative $\mathrm{dMSD}/\mathrm{d}\tau$ over $4D$ is 1 when τ is small. When τ increases, the derivative of MSD decreases because the particles in the OM system interact with each other, their motions are not purely diffusive but constrained. When τ increases, the decreasing value of $\mathrm{dMSD}/\mathrm{d}\tau$ depends on the beam power and the position of the particles in the array. When the beam power is higher or the particle is closer to the center of the 19-particle OM structure, the $\mathrm{dMSD}/\mathrm{d}\tau$ decreases faster as τ increases. Fig. 7.10d shows the difference between the average derivative of MSD of the outer layer particles and the center particle. At 20 mW beam power, the difference between the average derivative of MSD of the outer layer particles and the center particle is small, which means that the center particle and outer layer particles are not distinguishable in their dynamics for disordered structures. When beam power increases from 20 mW to 40

mW, the 19-particle OM structure manifests structural and dynamic properties intermediate between ordered and disordered, while the difference between the average derivative of MSD of the outer layer of particles and the center particle becomes larger. When the beam further increases from 40 mW to 60 mW, the 19-particle OM structure becomes ordered and the difference between the average derivative of MSD of the outer layer particles and the center particle first increases then decreases as τ increases.

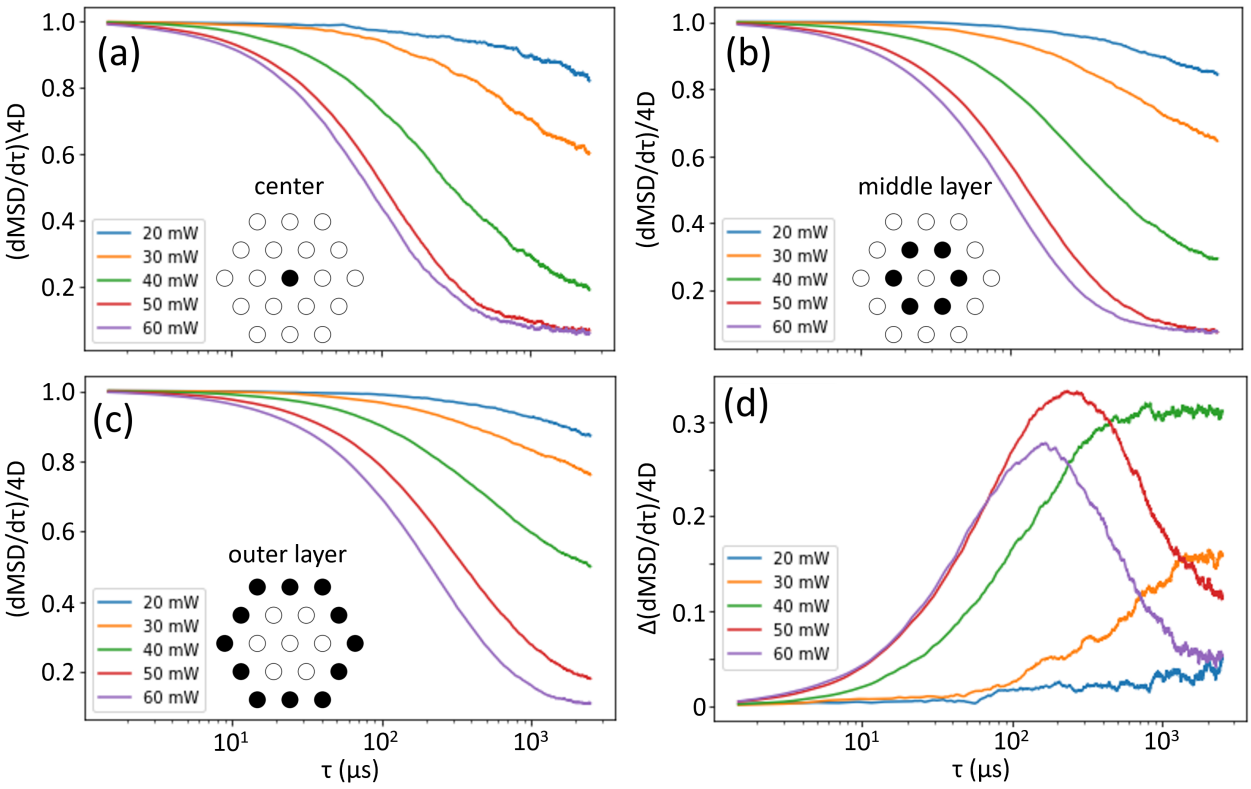

Figure 7.10: Plot of the derivative of average mean square displacement (MSD) with respect to lag time τ scaled by $4D$ where D is single-particle diffusion coefficient and τ is the lag time of the 19-particle OM system. (a-c) Scaled derivative of (a) the center particle (b) the average for the middle layer of particles (c) the average of the outer layer of particles. (d) Plot of the difference of the derivative dMSD/dτ over $4D$ between the outer layer of particles and the center particle.

Therefore, we can see from Fig. 7.10 that the central particle relaxes faster than the particles in the middle and outer layers. This is different from the Onsager inverted snowball model [181] which claims that the relaxation of the liquid molecules closer to the center of

the molecular cluster is slower than the ones distant to the center.

7.2.2 Discussion and conclusions

The structural transition between the ordered and disordered state of large OM clusters was referred to as "phase transition" by Han and his coworkers [74]. They analyzed the statistics of three order parameters for the transition: ψ_6 that characterize the structural deviation from trigonal lattice cites, pairwise distance distribution, and the angular velocity distribution. They also analyzed two mechanical properties of the system. One of them is the 2D potential which is the work done by moving one particle with other particles fixed. This 2D potential only considers the conservative part of the force field and neglects the non-conservativeness of the OM system. The other mechanical property is the sum of pairwise interactions of the particles in the OM system. Using pairwise interactions to describe OM structures neglects the many-body nature of the electrodynamic interactions among the particles, which is most essential for large OM clusters. In addition, Han and his coworkers did not connect the statistical analysis to the mechanical quantities they computed. Therefore, their work did not give a statistical mechanical description of the "phase transition" between the disordered and ordered OM clusters.

In contrast, my work in Section 7.2 looks into the mean square displacement (MSD) of the 19-particle OM structure. MSD and its derivative over lag time are statistical properties of the OM system that can characterize ordered and disordered structures. Moreover, as functions of lag time, MSD and its derivative over lag time can be connected to force field of the OM system using equations that characterizes diffusion such as Fokker-Planck equation. In this way, we can connect the statistics (MSD) of the system to the mechanics (force-field) of the system and thus, come up with a more rigorous description of the "phase transition" between the disordered and ordered OM clusters in the language of statistical mechanics.

CHAPTER 8

CONCLUSIONS AND FUTURE WORK

This dissertation focuses on the collective fluctuation modes of the optical matter (OM) systems, including the method of generating the fluctuation modes and the their applications. The applications of the collective fluctuation modes studied in this dissertation include: (i) obtaining the reaction coordinate between stable OM structures in the Markov state model constructed for the 6-particle OM system; (ii) computation of the entropy production rate of OM systems by introducing a linear approximation and comparing this to power dissipation; (iii) introducing a fluctuation mode-dependent symmetry analysis of the scattering cross-section and induced-polarization of OM structures in analogy to the classical Raman effect; (iv) investigating the non-equilibrium features of the 8-particle OM structure that manifests "pseudorotation"; and (v) examining OM systems where collective modes are not directly applicable such as the 19-particle OM system. In conclusion, the collective fluctuation mode methods developed in this dissertation are anticipated to offer useful new tools in studying the dynamics and scattering properties of OM systems. Furthermore, the collective mode approach does not rely on the electrodynamics interactions, so that this approach can be applied other overdamped non-equilibrium steady states of which the fundamental interaction is not electrodynamic.

In this chapter, I propose further research on four topics related to this dissertation: (i) application of machine learned policy beams to stabilize particular OM structures or transitions; (ii) computation of entropy production rate of the 19-particle OM structure; (iii) analysis of the role non-equilibrium effects play in the structural isomerization of the 8-particle OM structure that manifests pseudorotation; and (iv) verification of the fluctuation-mode-dependent torque direction predicted by the helicity-resolved symmetry analysis of the scattering properties of the 6-particle triangle and the 7-particle 6-fold symmetric structures. The following elaborates these three directions of future work in detail.

First, in Chapter 3, the optical beam power is the only parameter that is adjusted when the Markov state models are built for the 6-particle OM system. However, apart from beam power, there are other tunable parameters for the optical beam such as beam width, intensity profile, and phase profile. The phase profile of the incident beam can be experimentally controlled by tuning the Zernike coefficients in the spatial light modulator. Reinforcement learning can be adopted to control the stability of and the transitions between the stable states of the OM system by learning the policy beams that correspond to certain sets of beam parameters satisfying the goal of control.

Second, the entropy production rate is a thermodynamic quantity that is promising in describing the transition of between the disordered and ordered states of larger OM systems; the 19-particle OM system is a particular example. However, the collective fluctuation mode approach to compute the entropy production rate is not applicable to this 19-particle OM system because the disordered 19-particle OM system lacks a stable reference structure that is essential for obtaining the collective fluctuation modes of the system. Therefore, a different method such as transition path sampling (TPS) [182–184] could be used to compute the entropy production rate of this system. TPS estimates the entropy production rate from temporal averages [183]. Compared to the spatial average approach used in this thesis that involves Helmholtz decomposition of the force field, a temporal averaging approach uses the generated reversed-time trajectories and reweights them in a forward-time ensemble. This reweighting shows how the comparison of forward and reversed noise values has to do with a force dotted into a displacement, which gives the temporal entropy production rate [182] that can be averaged to generate the ensemble entropy production rate of the system. The precision of the entropy production rate computed by temporal averaging is controlled by the magnitude of the probability current fluctuation [184].

Entropy production rate calculation can serve as an intermediate quantity that links stability of OM structures and optical beam parameters. Taking advantage of the collec-

tive modes and Helmholtz decomposition of the force field, the entropy production rate can be computed efficiently for stable OM structures as non-equilibrium steady states (NESS). This computation efficiency enables data collection and training for reinforcement learning that eventually lead to control the OM entropy production rate tuning beam parameters. Next, taking advantage of the HLDA reaction coordinate introduced in Chapter 3, the TPS approach enables entropy production rate calculation along trajectories that manifest transitions between various pairs of stable OM structures. In this way, we will be able to relate the entropy production rate of different OM structures to their structural stability. To summarize, TPS and HLDA coordinate links structural stability to entropy production rate of OM structure, while reinforcement learning and the Helmholtz decomposition based on collective modes links entropy production rate to optical beam parameters. This is the idea of two-step control and design of stable OM system structures by means of entropy production rate.

Third, the helicity-resolved analysis of the scattering properties of symmetric OM structures in Section 6.B using irreducible representations predicts the torque direction for the so called Raman active degenerate mode pairs. It is necessary to verify this prediction by analyzing the torque in the Raman active degenerate subspaces of the two OM systems studied; the E representation subspaces for the 6-particle triangle OM structure and the E_2 representation subspaces for the 7-particle 6-fold symmetric structure. However, because these two structures are stable in the corresponding OM configuration space, the magnitude of the radial motion in the Raman active degenerate subspace is much larger than the azimuthal motion. This decreases the signal-to-noise ratio of the computed torque since only the azimuthal part of the motion characterizes the torque and extracting the signal is challenged by the significant noisy (thermal) background of radial motions. In future work, I believe that one can compute the torque in the Raman active degenerate subspace and average it throughout the simulation trajectory. In this way, only the azimuthal component of the force field will be taken into consideration in the Raman active degenerate subspace, which

is orthogonal to the radial force that generates the noise.

Finally, two important non-equilibrium features of the 8-particle OM system that manifests a pseudorotation fluctuation (8P-SMP) were introduced in section 7.1. Weighted principal component analysis (w-PCA) was used to generate the collective fluctuation coordinates of the 8P-SMP. In addition, a derivation starting from the Fokker-Planck equation explained the temperature dependence of the argument maximum of the probability distribution in the OM configuration space in Section 7.1.2.2. However, the role that the non-equilibrium nature of the electrodynamic interactions in OM systems plays in the frequent transitions between the two argument maxima of the probability distribution is still not clear. Here, I propose three directions for future work on this topic:

1. In section 7.1.3, we used Koopman analysis to determine that the implied time scale of the pseudorotation depends on temperature while the implied time scales of the other modes are independent of temperature. We also found that there exists a common threshold temperature for the marginal probability distribution along the collective modes in the configuration space (Fig. 7.5c,d), standard deviation (Fig. 7.5b), and the implied times scales (Fig. 7.6d) of pseudorotation configurations and dynamics. However, we do not know whether these non-conservative features are caused by a barrier crossing process along the pseudorotation coordinate or by the non-conservative part of the force field. Therefore, I propose carrying out the Koopman analysis using state-free non-reversible VAMPnet (SNRV) (introduced in detail in Section 7.1.3) and w-PCA for two more systems. One of them is the 6-particle triangle OM structure while the other is a modified 8-particle system for which the external force field is the conservative part of the force field of the OM 8P-SMP. If the Koopman analysis result of the 6-particle triangle OM system resembles one of the non-conservative features of the 8P-SMP, then that means that the feature is caused by the non-conservative part of the force field. If the Koopman analysis result of the modified 8-particle system

(for which the external force field is the conservative part of the force field of the OM 8P-SMP) resembles one of the non-conservative features of the full 8P-SMP, then that feature is caused barrier crossing along the pseudorotation coordinate. Therefore, by comparing the Koopman analysis results of these two systems and the results of the full 8P-SMP, we will obtain a better understanding of the cause of the non-conservative features in 8P-SMP dynamics.

2. Different from the OM systems that form trigonal lattice structures, the broken symmetry for modes 6 and 7 shown in Fig. 7.4 for the 8P-SMP is too large to be considered a perturbation. From irreducible representation analysis, we can find that modes 6 and 7 correspond to the B representation of the C_4 point group after circularly polarized field-induced symmetry breaking. Therefore, before symmetry breaking, the structure has D_4 symmetry and either mode 6 or 7 should correspond to irreducible representation B_1 (with a mirror symmetry axis) and the other should correspond to B_2 (with an anti-symmetric mirror axis). However, modes 6 and 7 shown in Fig. 7.4 deviate significantly from mirror symmetry or mirror anti-symmetry. Furthermore, modes 6 and 7 are mirror symmetry replicas of each other, but Fig. 7.5b shows that the standard deviations of modes 6 and 7 are different. The latter means that these two modes are more affected by the chiral symmetry breaking compared to other modes. In addition, the standard deviations of modes 7 and 8 cross each other at around 50 K, shown by Fig. 7.5b. This is the same threshold temperature as the crossing of translation and pseudorotation standard deviations. Therefore, it would be important to study modes 6 and 7 by looking at their correlation with the singular functions given by SNRV.

3. Raman activity of the normal modes of molecules [158, 159] includes two parts. One is the mode symmetry, which is studied in the Raman effect-inspired symmetry analysis introduced in Chapter 6. The other part is the selection rule; that is the vibration quantum number difference between the initial and final states can only be ± 1.

The selection rule comes from the harmonic approximation of the vibration potential along each normal mode. Therefore, it should be emphasized that the Raman effect-inspired symmetry analysis introduced in Chapter 6 does not include the quantum number difference, and thus does not include the harmonic approximation of the modes. Therefore, the non-harmonicity of the pseudorotation mode will not affect the Raman effect-inspired symmetry analysis so that the symmetry analysis analogous to Raman scattering can also be applied to the 8P-SMP to study its changes of scattering cross-section and induced-polarization along its collective fluctuation modes. By comparing the result of the change of scattering cross-section induced-polarization along the collective modes of the 8P-SMP to that of the 6-particle triangle and 7-particle 6-fold symmetric OM systems, we will obtain a better understanding of the relation between the scattering properties and the non-conservative features in the 8P-SMP OM system.

REFERENCES

[1] S. Chen, C. W. Peterson, J. A. Parker, S. A. Rice, A. L. Ferguson, and N. F. Scherer. Data-driven reaction coordinate discovery in overdamped and non-conservative systems: application to optical matter structural isomerization. Nat. Commun., 12(1):2548, 2021.

[2] S. Chen, J. A. Parker, C. W. Peterson, S. A. Rice, N. F. Scherer, and A. L Ferguson. Understanding and design of non-conservative optical matter systems using markov state models. Mol. Syst. Des. Eng., 7(10):1228–1238, 2022.

[3] C. W. Peterson, J. A. Parker, E. Valenton, Y. Yifat, S. Chen, S. A. Rice, and N. F. Scherer. Interference and induced-polarization in nanoparticle-based optical matter arrays. J. Phys. Chem. C, 128(18):7560–7571, 2024.

[4] H. Y. K. Kaan, D. D. Hackney, and F. Kozielski. The structure of the kinesin-1 motor-tail complex teveals the mechanism of autoinhibition. Science, 333(6044):883–885, 2011.

[5] C.-A. Palma, J. Bj ork, F. Rao, D. K uhne, F. Klappenberger, and J. V. Barth. Topological dyanmics in supramolecular rotors. Nano Lett., 14(8):4461–4468, 2014.

[6] M. M. Burns, J. Fournier, and J. A. Golovchenko. Optical matter: Crystallization and binding in intense optical fields. Science, 249(4970):749–754, 1990.

[7] Z. Yan, R. A. Shah, G. Chado, S. K. Gray, M. Pelton, and N. F. Scherer. Guiding spatial arrangements of silver nanoparticles by optical binding interactions in shaped light fields. ACS Nano, 7(2):1790–1802, 2013.

[8] V. Demergis and E. Florin. Ultrastrong optical binding of metallic nanoparticles. Nano Lett., 12(11):5756–5760, 2012.

[9] K. Dholakia and P. Zemánek. Colloquium: Gripped by light: Optical binding. Rev. Mod. Phys., 82(2):1767–1791, 2010.

[10] K. A. Forbes, D. S. Bradshaw, and D. L. Andrews. Optical binding of nanoparticles. Nanophotonics, 9(1):1–17, 2020.

[11] Z. Yan, M. Sajjan, and N. F. Scherer. Fabrication of a material assembly of silver nanoparticles using the phase gradients of optical tweezers. Phys. Rev. Lett., 114(14):143901, 2015.

[12] F. Nan, F. Han, N. F. Scherer, and Z. Yan. Dissipative self-assembly of anisotropic nanoparticle chains with combined electrodynamic and electrostatic interactions. Adv. Mater., 30(45):1803238, 2018.

[13] F. Nan and Z. Yan. Synergy of intensity, phase, and polarization enables versatile optical nanomanipulation. Nano Lett., 20(4):2778–2783, 2020.

[14] Y. Yifat, D. Coursault, C. W. Peterson, J. Parker, Y. Bao, S. K. Gray, S. A. Rice, and N. F. Scherer. Reactive optical matter: Light-induced motility in electrodynamic asymmetric nanoscale scatterers. Light: Sci. & Appl., 7(1):1–7, 2018.

[15] F Han, J. A. Parker, Y. Yifat, C. W. Peterson, S. K. Gray, N. F. Scherer, and Z. Yan. Crossover from positive to negative optical torque in mesoscale optical matter. Nat. Commun., 9(1):4897, 2018.

[16] N. Sule, Y. Yifat, S. K. Gray, and N. F. Scherer. Rotation and negative torque in electrodynamically bound nanoparticle dimers. Nano Lett., 17(11):6548–6556, 2017.

[17] J. A. Parker, C. W. Peterson, Y. Yifat, S. A. Rice, Z. Yan, S. K. Gray, and N. F. Scherer. Optical matter machines: Angular momentum conversion by collective modes in optically bound nanoparticle arrays. Optica, 7(10):1341–1348, 2020.

[18] T. Vicsek and A. Zafeiris. Collective motion. Phys. Rep., 517(3):71–140, 2012.

[19] P. Figliozzi, C. W. Peterson, S. A. Rice, and N. F. Scherer. Direct visualization of barrier crossing dynamics in a driven optical matter system. ACS Nano, 12(6):5168–5175, 2018.

[20] N. Kamaly, G. Fredman, M. Subramanian, S. Gadde, A. Pesic, L. Cheung, Z. A. Fayad, R. Langer, I. Tabas, and O. C. Farokhzad. Development and in vivo efficacy of targeted polymeric inflammation-resolving nanoparticles. PNAS, 110(16):6506–6511, 2013.

[21] V. Kamysbayev, V. Srivastava, N. B. Ludwig, O. J. Borkiewicz, H. Zhang, J. Ilavsky, B. Lee, K. W. Chapman, S. Vaikuntanathan, and D. V. Talapin. Nanocrystals in molten salts and ionic liquids: Experimental observation of ionic correlations extending beyond the debye length. ACS Nano, 13(5):5760–5770, 2019.

[22] D. Mendels, G. Piccini, and M. Parrinello. Collective variables from local fluctuations. J. Phys. Chem. Lett., 9(11):2776–2781, 2018.

[23] H. Seenivasan, B. Jackson, and A. K. Tiwari. Water dissociation on ni(100), ni(110), and ni(111) surfaces: Reaction path approach to mode selectivity. J. Chem. Phys., 146(7):074705, 2017.

[24] K. Kakhiani, U. Lourderaj, W. Hu, D. Birney, and W. L. Hase. Cyclohexane isomerization. unimolecular dynamics of the twist-boat intermediate. J. Phys. Chem. A, 113(16):4570–4580, 2009.

[25] S. Sukhov and A. Dogariu. Non-conservative optical forces. Rep. Prog. Phys., 80(11):112001, 2017.

[26] L. Renson and G. Kerschen. Nonlinear Normal Modes of Nonconservative Systems. In: Kerschen G., Adams D., Carrella A. (eds) Topics in Nonlinear Dynamics, Volume

1. Conference Proceedings of the Society for Experimental Mechanics Series. Springer, New York, 2013.

[27] C. C. David and D. J. Jacobs. Principal component analysis: a method for determining the essential dynamics of proteins. Methods Mol Biol., 1084:193–226, 2014.

[28] R. Milkus, C. Ness, V. V. Palyulin, J. Weber, A. Lapkin, and A. Zaccone. Interpretation of the vibrational spectra of glassy polymers using coarse-grained simulations. Macromolecules, 51(4):1559–1572, 2018.

[29] V. V. Palyylin, C. Ness, R. Milkus, R. M. Elder, T. W. Sirk, and A. Zaccone. Parameter-free predictions of the viscoelastic response of glassy polymers from non-affine lattice dynamics. Soft Matter, 14(42):8475–8482, 2018.

[30] I. Kriuchevskyi, V. V. Palyulin, R. Milkus, R. M. Elder, T. W. Sirk, and A. Zaccone. Scaling up the lattice dynamics of amorphous materials by orders of magnitude. Phys. Rev. B, 102(2):024108, 2020.

[31] A. Zaccone and M. Baggioli. Universal law for the vibrational density of states of liquids. PNAS, 118(5):e2022303118, 2021.

[32] J. Chattoraj, O. Gendelman, M. P. Ciamarra, and I. Procaccia. Oscillatory instabilities in frictional granular matter. Phys. Rev. Lett., 123(9):098003, 2019.

[33] H. Abdi and L. J. Williams. Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics, 2:433–459, 2010.

[34] N. Sule, S. A. Rice, S. K. Gray, and N. F. Scherer. An electrodynamics-langevin dynamics (ed-ld) approach to simulate metal nanoparticle interactions and motion. Optics Express, 23(23):29978–29992, 2015.

[35] J. A. Parker. Collective Electrodynamic Excitations and Non-conservative Dynamics in Optical Matter and Meta-Atom Systems. PhD thesis, University of Chicago, 2020.

[36] https://github.com/shiqic/PhD-Thesis.

[37] Y.-I. Xu. Electromagnetic scattering by an aggregate of spheres. Appl. Opt., 34(21):4573–4588, 1995.

[38] J. A. Lock and G. Gouesbet. Generalized lorenz-mie theory and applications. J. Quant. Spectrosc. Radiat. Transf., 110(11):800–807, 2009.

[39] Parker, J. A., Miepy package, 10.5281/zenodo.4606745 (2021). https://github.com/johnaparker/miepy.

[40] H.-J. Butt, K. Graf, and M. Kappl. Physics and chemistry interfaces. John Wiley & Sons, Hoboken, New Jersey, 2013.

[41] P. Figliozzi, N. Sule, Z. Yan, Y. Bao, S. Burov, S. K. Gray, S. A. Rice, S. Vaikuntanathan, and N. F. Scherer. Driven optical matter: Dynamics of electrodynamically coupled nanoparticles in an optical ring vortex. Phys. Rev. E, 95(2):022604, 2017.

[42] V. S. Pande, K. Beauchamp, and G. R. Bowman. Everything you wanted to know about markov state models but were afraid to ask. Methods, 52(1):99–105, 2010.

[43] S. Schultze and H. Grubmüller. Time-lagged independent component analysis of random walks and protein dynamics. J. Chem. Theory Comput., 17(9):5766–5776, 2021.

[44] C. R. Schwantes and V. S. Pande. Modeling molecular kinetics with tica and the kernel trick. J. Chem. Theory Comput., 11(2):600–608, 2015.

[45] G. Andrew, R. Arora, J. Bilmes, and K. Livescu. Deep canonical correlation analysis. PMLR, 28(3):1247–1255, 2013.

[46] A. Mardt, L. Pasquali, H. Wu, and F. Noé. Vampnets for deep learning of molecular kinetics. Nat Commun., 9(1):4443, 2018.

[47] W. Chen, H. Sidky, and A. L. Ferguson. Nonlinear discovery of slow molecular modes using state-free reversible vampnets. J. Chem. Phys., 150(21):214114, 2019.

[48] R. R. Coifman and S. Lafon. Diffusion maps. Appl. Comput. Harmon. Anal., 21(1):5–30, 2006.

[49] R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker. Geometric diffusions as a tool for harmonic analysis and structure definition of data: diffusion maps. Proc. Natl. Acad. Sci. U. S. A., 102(21):7426–7431, 2005.

[50] B. Nadler, S. Lafon, R. R. Coifman, and I. G. Kevrekidis. Diffusion maps, spectral clustering and reaction coordinates of dynamical systems. Appl. Comput. Harmon. Anal., 21(1):113–127, 2006.

[51] A. Singer, R. Erban, I. G. Devrekidis, and R. R. Coifman. Detecting intrinsic slow variables in stochastic dynamical systems by anisotropic diffusion maps. Proc. natl. Acad. Sci. U. S. A., 106(38):16090–16095, 2009.

[52] A. W. Long and A. L. Ferguson. Nonlinear machine learning of patchy colloid self assembly pathways and mechanisms. J. Phys. Chem. B, 118(15):4228–4244, 2014.

[53] A. W. Long, J. Zhang, S. Granick, and A. L. Ferguson. Machine learning assembly landscapes from particle tracking data. Soft Matter, 11(41):8141–8153, 2015.

[54] A. W. Long, C. L. Phillips, E. Jankowksi, and A. L. Ferguson. Nonlinear machine learning and design of reconfigurable digital colloids. Soft Matter, 12(34):7119–7135, 2016.

[55] Z. Trstanova, B. Leimkuhler, and T. Lelièvre. Local and global perspectives on diffusion maps in the analysis of molecular systems. Proc. R. Soc. A, 476(2233):20190036, 2020.

[56] L. Boninsegna, G. Gobbo, F. Noé, and C. Clementi. Investigating molecular kinetics by variationally optimized diffusion maps. J. Chem. Theory Comput., 11(12):5947–5960, 2015.

[57] J. B. MacQueen. Some methods for classification and analysis of multivariate observations, pages 281–297. Proc. Fifth Berkeley Sympos. Math. Statist. and Probability, 1967.

[58] X. Chen and Y. Yang. Diffusion k-means clustering on manifolds: Provable exact recovery via semidefinite relaxations. Appl. Comput. Harmon. Anal., 52(3):303–347, 2021.

[59] S. Röblitz and M. Weber. Fuzzy spectral clustering by pcca+: application to markov state models and data classification. Adv. Data Anal. Classif., 7(2):147–179, 2013.

[60] P. Deuflhard and M. Weber. Robust perron cluster analysis in conformation dynamics. Linear Algebra Appl., 398(1):161–184, 2005.

[61] S. Kube and M. Weber. A coarse graining method for the identification of transition rates between molecular conformations. J. Chem. Phys., 126(2):024103, 2007.

[62] G. N. Simm, A. C. Vaucher, and M. Reiher. Exploration of reaction pathways and chemical transformation networks. J. Phys. Chem. A, 123(2):385–399, 2019.

[63] R. Elber, J. M. Bello-Rvas, P. Ma, A. E. Cardenas, and A Fathizadeh. Calculating iso-committor surfaces as optimal reaction coordinates with milestoning. Entropy, 19(5):219–235, 2017.

[64] R. S. Berry, S. A. Rice, and J. Ross. Physical Chemistry. Oxford Univ. Press, Oxford, 2000.

[65] P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler. Transition path sampling: Throwing ropes over rough mountain passes. Annu. Rev. Phys. Chem., 53(1):291–318, 2002.

[66] A. Dickson and A. R. Dinner. Enhanced sampling of nonequilibrium steady states. Annu. Rev. Phys. Chem., 61(1):441–459, 2010.

[67] S. Sun and E. R. Bernstein. Aromatic van der waals clusters: Structure and nonrigidity. J. Phys. Chem., 100(32):13348–13366, 1996.

[68] P. A. Pincus and T. A. Witten. Structured Fluids: Polymers, Colloids, Surfactants Ch. 5. Oxford Univ. Press, Oxford, 2004.

[69] V. N. Manoharan. Colloidal matter: Packing, geometry, and entropy. Science, 349(6251):1253751, 2015.

[70] D. G. Grier. Optical tweezers in colloid and interface science. Current Optinion in Colloid & Interface Science, 2(3):264–270, 1997.

[71] M. Z. Lim, A. Souslov, V. Vitelli, and H. M. Jaeger. Cluster formation by acoustic forces and active fluctuations in levitated granular matter. Nat. Phys., 15(5):460–464, 2019.

[72] H. C. Andersen, D. Chandler, and J. D. Weeks. Roles of repulsive and attractive forces in liquids: the optimized random phase approximation. J. Chem. Phys., 56(8):3812–3823, 1972.

[73] R. S. Berry and B. M. Smirnov. Configurational transitions in processes involving metal clusters. Physics Reports, 527(4):205–250, 2013.

[74] F. Han and Z. Yan. Phase transition and self-stabilization of light-mediated metal nanoparticle assemblies. ACS Nano, 14(6):6616–6625, 2020.

[75] F. A. Escobedo, E. E. Borrero, and J. C. Araque. Transition path sampling and forward flux sampling: Applications to biological systems. J. Phys.: Condens. Matter, 21(33):333101, 2009.

[76] Y. Roichman and D. G. Grier. Three-dimensional holographic ring traps. Proc. SPIE, 6483:64830F, 2007.

[77] J. Damková, L. Chvátal, J. Ježek, J. Oulehla, O. Brzobohatý, and P. Zemánek. Enhancement of the 'tractor-beam' pulling force on an optically bound structure. Light: Sci. & Appl., 7(1):17135, 2018.

[78] C. W. Peterson, J. A. Parker, S. A. Rice, and N. F. Scherer. Controlling the dynamics and optical binding of nanoparticle homodimers with transverse phase gradients. Nano Lett., 19(2):897–903, 2019.

[79] A. La Porta and M. D. Wang. Optical torque wrench: Angular trapping, rotation, and torque detection of quartz microparticles. Phys. Rev. Lett., 92(19):190801, 2004.

[80] F. Pedaci, Z. Huang, M. Van Oene, S. Barland, and N. H. Dekker. Excitable particles in an optical torque wrench. Nat. Phys., 7:259–264, 2011.

[81] J. Inman, S. Forth, and M. D. Wang. Passive torque wrench and angular position detection using a single-beam optical trap. Opt. Lett., 35(17):2949–2951, 2010.

[82] L. Tong, V. D. Miljkovic, and M. Kall. Alignment, rotation, and spinning of single plasmonic nanoparticles and nanowires using polarization dependent optical forces. Nano Lett., 10(1):263–273, 2009.

[83] Z. Yan and N. F. Scherer. Optical vortex induced rotation of silver nanowires. J. Phys. Chem. Lett., 4(17):2937–2942, 2013.

[84] J. Chen, J. Ng, K. Ding, K. H. Fung, Z. Lin, and C. T. Chan. Negative optical torque. Sci. Reports, 4:6386, 2014.

[85] T. Speck. Stochastic thermodynamics for active matter. EPL, 114(3):30006, 2016.

[86] S. C. Takatori and J. F. Brady. Towards a thermodynamics of active matter. Phys. Rev. E, 91(3):032117, 2015.

[87] U. M. Ascher. Computer Methods for Ordinary Differential Equations and Differential Algebraic Equations. Society for Industrial and Applied Mathematics, Philadelphia, 1998.

[88] F. Shafiei, F. Monticone, K. Q. Le, X.-X. Liù, T. Hartsfield, A. Alu, and X. Li. A subwavelength plasmonic metamolecule exhibiting magnetic-based optical fano resonance. Nat. Nanotechnol., 8(2):95–99, 2013.

[89] S. N. Sheikholeslami, H. Alaeian, A. L. Koh, and J. A. Dionne. A metafluid exhibiting strong optical magnetism. Nano Lett., 13(9):4137–4141, 2013.

[90] A. Kuzyk, R. Schreiber, H. Zhang, A. O. Govorov, T. Liedl, and N. Liu. Reconfigurable 3d plasmonic metamolecules. Nat. Mater., 13(9):862–866, 2014.

[91] S. Yang, X. Ni, X. Yin, B. Kante, P. Zhang, J. Zhu, Y. Wang, and X. Zhang. Feedback-driven self-assembly of symmetry-breaking optical metamaterials in solution. Nat. Nanotechnol., 9(12):1002–1006, 2014.

[92] Z. Qian, S. P. Hastings, C. Li, B. Edward, C. K. McGinn, N. Engheta, Z. Fakhraai, and S.-J. Park. Raspberry-like metamolecules exhibiting strong magnetic resonances. ACS Nano, 9(2):1263–1270, 2015.

[93] J. Sharma, R. Chhabra, A. Cheng, J. Brownell, Y. Liu, and H. Yan. Control of self-assembly of dna tubules through integration of gold nanoparticles. Science, 323(5910):112–116, 2009.

[94] L. Malassis, P. Massé, M. Tréguer-Delapierre, S. P. Mornet, P. Weisbecker, V. Kravets, A. Grigorenko, and P. Barois. Bottom-up fabrication and optical characterization of dense films of meta-atoms made of core-shell plasmonic nanoparticles. Langmuir, 29(5):1551–1561, 2013.

[95] J. Lee, J. H. Huh, K. Kim, and S. Lee. Dna origami-guided assembly of the roundest 60-100 nm gold nanospheres into plasmonic metamolecules. Adv. Funct. Mater., 28(15):1707309, 2018.

[96] M. Kanahara, H. Satoh, T. Higuchi, A. Takahara, H. Jinnai, K. Harano, S. Okada, E. Nakamura, Y. Matsuo, and H. Yabu. Fabrication of nir-excitable sers-active composite particles composed of densely packed au nanoparticles on polymer microparticles. Part. Part. Syst. Charact., 32(4):441–447, 2015.

[97] U. Manna, J.-H. Lee, T.-S. Deng, J. Parker, N. Shepherd, Y. Weizmann, and N. F. Scherer. Selective induction of optical magnetism. Nano Lett., 17(12):7196–7206, 2017.

[98] A. Sánchez-Iglesias, M. Grzelczak, T. Altantzis, B. Goris, J. Perez-Juste, S. Bals, G. Van Tendeloo, S. H. Donaldson, B. F. Chmelka, and J. N. Israelachvili. Hydrophobic interactions modulate self-assembly of nanoparticles. ACS Nano, 6(12):11059–11065, 2012.

[99] M. Stalder and M. Schadt. Linearly polarized light with axial symmetry generated by liquid-crystal polarization converters. Opt. Lett., 21(23):1948–1950, 1996.

[100] A. H. Zewail. Femtochemistry: Atomic-scale dynamics of the chemical bond. J. Phys. Chem. A, 104(24):5660–5694, 2000.

[101] N. F. Scherer, L. R. Khundkar, R. B. Bernstein, and A. H. Zewail. Real-time picosecond clocking of the collision complex in a bimolecular reaction: The birth of oh from h+co_2. J. Chem. Phys., 87(2):1451–1453, 1987.

[102] N. F. Scherer, D. M. Jonas, and G. R. Fleming. Femtosecond wave packet and chemical reaction dynamics of iodine in solution: Tunable probe study of motion along the reaction coordinate. J. Chem. Phys., 99(1):153–168, 1993.

[103] Y. Lee and Y. Shen. Studies with crossed laser and molecular beams. Phys. Today, 33(11):52–59, 1980.

[104] R. D. Levine and R. B. Bernstein. Molecular reaction dynamics and chemical reactivity. Phys. Today, 41(10):90, 1988.

[105] M. M. Burns, J.-M. Fournier, and J. A. Golovchenko. Optical binding. Phys. Rev. Lett., 63(12):1233, 1989.

[106] S. Tatarkova, A. Carruthers, and K. Dholakia. One-dimensional optically bound arrays of microscopic particles. Phys. Rev. Lett., 89(28):283901, 2002.

[107] A. Ashkin, J. M. Dziedzic, J. E. Bjorkholm, and S. Chu. Observation of a single-beam gradient force optical trap for dielectric particles. Opt. Lett., 11(5):288–290, 1986.

[108] Y. Roichman, B. Sun, Y. Roichman, J. Amato-Grill, and D. G. Grier. Optical forces arising from phase gradients. Phys. Rev. Lett., 100(1):013602, 2008.

[109] M. Pelton, M. Liu, H. Y. Kim, G. Smith, P. Guyot-Sionnest, and N. F. Scherer. Optical trapping and alignment of single gold nanorods by using plasmon resonances. Opt. Lett., 31(13):2075–2077, 2006.

[110] J. Barton, D. Alexander, and S. Schaub. Theoretical determination of net radiation force and torque for a spherical particle illuminated by a focused laser beam. J. Appl. Phys., 66(10):4594–4602, 1989.

[111] A. Yevick, D. J. Evans, and D. G. Grier. Photokinetic analysis of the forces and torques exerted by optical tweezers carrying angular momentum. Philos. Trans. Royal Soc. A: Mathematical, Physical and Engineering Sciences, 375(2087):20150432, 2017.

[112] S. E. S. Spesyvtseva and K. Dholakia. Trapping in a material world. ACS Photonics, 3(5):719–736, 2016.

[113] S. H. Simpson, P. Zemánek, O. M. Maragò, P. H. Jones, and S. Hanna. Optical binding of nanowires. Nano Lett., 17(6):3485–3492, 2017.

[114] S. Kuhn, A. Kosloff, B. A. Stickler, F. Patolsky, K. Hornberger, M. Arndt, and J. Millen. Full rotational control of levitated silicon nanorods. Optica, 4(3):356–360, 2017.

[115] D. Coursault, N. Sule, J. Parker, Y. Bao, and N. F. Scherer. Dynamics of the optically directed assembly and disassembly of gold nanoplatelet arrays. Nano Lett., 18(6):3391–3399, 2018.

[116] Z. Yan, S. K. Gray, and N. F. Scherer. Potential energy surfaces and reaction pathways for light-mediated self-organization of meta nanoparticle clusters. Nat. Commun., 5(1):1–7, 2014.

[117] J. Ng, Z. Lin, C. Chan, and P. Sheng. Photonic clusters formed by dielectric microspheres: Numerical simulations. Phys. Rev. B, 72(8):075130, 2005.

[118] J. Taylor and G. Love. Spontaneous symmetry breaking and circulation by optically bound microparticle chains in gaussian beam traps. Phys. Rev. A, 80(5):053808, 2009.

[119] S. Albaladejo, J. J. Sáenz, and M. I. Marqués. Plasmonic nanoparticle chain in a light field: A resonant optical sail. Nano Lett., 11(11):4597–4600, 2011.

[120] L. Chvátal, O. Brzobohatý, and P. Zemánek. Binding of a pair of au nanoparticles in a wide gaussian standing wave. Opt. Rev., 22(1):157–161, 2015.

[121] S. Sukhov, A. Shalin, D. Haefner, and A. Dogariu. Actio et reactio in optical binding. Opt. Express, 23(1):247–252, 2015.

[122] V. Karásek, M. Siler, O. Brzobohatý, and P. Zemaánek. Dynamics of an optically bound structure made of particles of unequal sizes. Opt. Lett., 42(7):1436–1439, 2017.

[123] J. Chen, S. Wang, X. Li, and J. Ng. Mechanical effect of photonic spin-orbit interaction for a metallic nanohelix. Opt. Express, 26(21):27694–27704, 2018.

[124] Y. Roichman and D. G. Grier. Complex light and optical forces. International Society for Optics and Photonics, 6483:64830F, 2007.

[125] U. Sengupta, M. Carballo-pacheco, and B. Strodel. Automated markov state models for molecular dynamics simulations of aggregation and self-assembly. J. Chem. Phys., 150(11):115101, 2019.

[126] J. D. Chodera and F. Noé. Markov state models of biomolecular conformational dynamics. Curr. Opin. Struct. Biol., 25:135–144, 2014.

[127] B. E. Husic and V. S. Pande. Markov state models: From an art to a science. J. Am. Chem. Soc., 140(7):2386–2396, 2018.

[128] C. Wehmeyer, M. K. Scherer, T. Hempel, B. E. Husic, S. Olsson, and F. Noé. Introduction to markov state modeling with the pyemma software. Living J. Comput. Mol. Sci., 1(1):5965, 2019.

[129] M. Weber and T. Galliat. Characterization of transition states in conformational dynamics using fuzzy sets. Konrad-Zuse-Zentrum f ur Informationstechnik Berlin, 12:1–12, 2002.

[130] J. D. Chodera, N. Singhal, V. S. Pande, K. A. Dill, and W. C. Swope. Automatic discovery of metastable states for the construction of markov models of macromolecular conformational dynamics. J. Chem. Phys., 126(15):155101, 2007.

[131] J. Hu and A. L. Ferguson. Global graph matching using diffusion maps. Intelligent Data Analysis, 20(3):637–654, 2016.

[132] A. L. Ferguson, A. Z. Panagiotopoulos, I. G. Kevrekidis, and P. G. Debenedetti. Nonlinear dimensionality reduction in molecular simulation: The diffusion map approach. Chem. Phys. Lett., 509(1):1–11, 2011.

[133] C. Fowlkes, S. Belongie, F. Chung, and J. Malik. Spectral grouping using the nyström method. IEEE Trans. Pattern Anal. Mach. Intell., 26(2):214–225, 2004.

[134] S. Lafon, Y. Keller, and R. R. Coifman. Data fusion and multicue data matching by diffusion maps. IEEE Trans. Pattern Anal. Mach. Intell., 28(11):1784–1797, 2006.

[135] B. E. Sonday, M. Haataja, and I. G. Kevrekidis. Coarse-graining the dynamics of a driven interface in the presence of mobile impurities. Phys. Rev. E, 80(3):031102, 2009.

[136] J. Wang and A. L. Ferguson. A study of morphology, dynamics, and folding pathways of ring polymers with supramolecular topological constraints using molecular simulation and nonlinear manifold learning. Macromolecules, 51(2):598–616, 2018.

[137] A. W. Long and A. L. Ferguson. Landmark diffusion maps (l-dmaps): Accelerated manifold learning out-of-sample extension. Appl. Comput. Harmon. Anal., 47(1):190–211, 2019.

[138] J. Wang, M. A. Gayatri, and A. L. Ferguson. Mesoscale simulation and machine learning of asphaltene aggregation phase behavior and molecular assembly landscapes. J. Phys. Chem. B, 121(18):4923–4944, 2017.

[139] M. K. Scherer, B. Trendelkamp-Schroer, F. Paul, G. Pérez-Hernández, M. Hoffmann, N. Plattner, C. Wehmeyer, J.-H. Prinz, and F. Noé. Pyemma 2: a software package for estimation, validation, and analysis of markov models. J. Chem. Theory Comput., 11(11):5525–5542, 2015.

[140] H. Wu, F. Nüske, F. Paul, S. Klus, P. Koltai, and F. Noé. Variational koopman models: Slow collective variables and molecular kinetics from short off-equilibrium simulations. J. Chem. Phys., 146(15):154104, 2017.

[141] P. J. Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math., 20:53–65, 1987.

[142] H. Akaike. A new look at the statistical model identification. IEEE Trans. Automat. Contr., 19(6):716–723, 1974.

[143] X. Zeng, B. Li, Q. Qiao, L. Zhu, Z. Lu, and X. Huang. Elucidating dominant pathways of the nano-particle self-assembly process. Phys. Chem. Chem. Phys., 18(34):23494–23499, 2016.

[144] U. Seifert. Stochastic thermodynamics, fluctuation theorems and molecular machines. Rep. Prog. Phys., 75(12):126001, 2012.

[145] I. Prigogine. Étude thermodynamique des phénomènes irréversibles. Bull. Acad. Roy. Belg., 31:600, 1945.

[146] P. R. Zulkowski, D. A. Sivak, G. E. Crooks, and M. R. DeWeese. Geometry of thermodynamic control. Phys. Rev. E, 86(4):041148, 2012.

[147] N. S. Wadia, R. V. Zarcone, and M. R. DeWeese. Solution to the fokker-planck equation for slowly driven brownian motion: Emergent geometry and a formula for the corresponding thermodynamic metric. Phys. Rev. E, 105(3):034130, 2022.

[148] W. D. Pineros and T. Tlusty. Inverse design of nonequilibrium steady states: A large-deviation approach. Phys. Rev. E, 103(2):022101, 2021.

[149] T. Nishiyama and Y. Hasegawa. Upper bound for entropy production in markov processes. Phys. Rev. E, 108(1):044139, 2023.

[150] F. Nan, X. Li, S. Zhang, J. Ng, and Z. Yan. Creating stable trapping force and switchable optical torque with tunable phase of light. Sci. Adv., 8(46):eadd6664, 2022.

[151] É. Fodor, C. Narkini, M. E. Cates, J. Tailleur, P. Visco, and F. van Wijland'. How far from equilibrium is active matter? Phys. Rev. Lett., 117(3):038103, 2016.

[152] S. Shankar and M. C. Marchetti. Hidden entropy production and work fluctuations in an ideal active gas. Phys. Rev. E, 98(2):020604, 2018.

[153] F. Nan and Z. Yan. Probing spatiotemporal stability of optical matter by polarization modulation. Nano Lett., 18(2):1396–1401, 2018.

[154] H. Qian. Nonequilibrium steady-state circulation and heat dissipation functional. Phys. Rev. E, 64(2):022101, 2001.

[155] U. Seifert. Entropy production along a stochastic trajectory and an integral fluctuation theorem. Phys. Rev. Lett., 95(4):040602, 2005.

[156] D. Chandler. Introduction to Modern Statistical Mechanics. Oxford University Press, Oxford, 1987.

[157] B. Beresford-Smith, D. Y. C. Chan, and J. D. Mitchell. The electrostatic interaction in colloidal systems with low added electrolyte. J. Colloid Interface Sci., 105(1):216–234, 1985.

[158] G. Herzberg. Molecular Spectra and Molecular Structure Vol. II Infrared and Raman Spectra of Polyatomic Molecules. D. van Nostrand, New york, 1945.

[159] D. C. Harris and M. D. Bertolucci. Symmetry and Spectroscopy: An Introduction to Vibrational and Electronic Spectroscopy. Oxford University Press, Dover, 1945.

[160] R. Loudon. The raman effect in crystals. Adv. Phys., 50(52):813–864, 2001.

[161] M. Tanaka and R.J. Young. Review polarised raman spectroscopy for the study of molecular orientation distributions in polymers. J. Mater. Sci., 41(3):963–991, 2006.

[162] S. G. Drapcho, J. Kim, X. Hong, C. Jin, S. Shi, S. Tongay, J. Wu, and F. Wang. Apparent breakdown of raman selection rule at valley exciton resonances in monolayer mos_2. Phys. Rev. B, 95(16):165417, 2017.

[163] K. Y. Bliokh, F. J. Rodríguez-Fortuño, F. Nori, and A. V. Zayats. Spin-orbit interactions of light. Nat. Photonics, 9(12):796–808, 2015.

[164] K. Y. Bliokh and F. Nori. Transvers and longitudinal angular momenta of light. Phys. Rep., 592:1–38, 2015.

[165] S.-Y. Chen, C. Zheng, M. S. Fuhrer, and J. Yan. Helicity-resolved raman scattering of mos_2, $mose_2$, ws_2, and wse_2 atomic layers. Nano Lett., 15(4):2526–2532, 2015.

[166] Y. Tatsumi and R. Saito. Interplay of valley selection and helicity exchange of light in raman scattering for graphene and mos_2. Phys. Rev. B, 97(11):115407, 2018.

[167] R. Saito, Y. Tatsumi, S. Huang, X. Ling, and M. S. Dresselhaus. Raman spectroscopy of transition metal dichalcogenides. J. Phys.: Condens. Matter, 28(35):353002, 2016.

[168] D. S. Bradshaw and D. L. Andrews. Optically induced forces and torques: Interactions between nanoparticles in a laser beam. Phys. Rev. A, 72(3):033816, 2005.

[169] S. Raza and A. Kristensen. Raman scattering in high-refractive-index nanostructures. Nanophotonics, 10(4):1197–1209, 2020.

[170] G. P. Zograf, D. Ryabov, V. Rutckaia, P. Voroshilov, P. Tonkaev, D. V. Permyakov, Y. Kivshar, and S. V. Makarov. Stimulated raman scattering from mie-resonant subwavelength nanoparticles. Nano Lett., 20(8):5786–5791, 2020.

[171] S. Chen, E. Valenton, G. M. Rotskoff, A. L. Ferguson, S. A. Rice, and N. F. Scherer. Power dissipation and entropy production rate of high-dimensional optical matter systems. Power Dissipation and Entropy Production Rate of High-dimensional Optical Matter Systems. *Phys. Rev. E*, 2024, under review.

[172] S. Streib. Difference between angular momentum and pseudoangular momentum. Phys. Rev. B, 103(10):L100409, 2021.

[173] Y. Tatsumi, T. Kaneko, and R. Saito. Conservation law of angular momentum in helicity-dependent raman and rayleigh scattering. Phys. Rev. B, 97(19):195444, 2018.

[174] R. S. Berry. Correlation of rates of intramolecular tunneling processes, with application to some group v compounds. J. Chem. Phys., 32(3):933–938, 1960.

[175] M. E. Cass, K. K. Hii, and H. S. Rzepa. Mechanisms that interchange axial and equatorial atoms in fluxional processes: Illustration of the berry pseudorotation, the turnstile, and the lever mechanisms via animation of transition state normal vibration modes. J. Chem. Educ., 83(2):336, 2006.

[176] F. Noé. Machine Learning Meets Quantum Physics, Ch. 16 Machine Learning for Molecular Dynamics on Long Timescales. Spring Nature Switzerland AG, Cham, Switzerland, 2020.

[177] H. Wu and F. Noé. Variational approach for learning markov processes from time series data. J. Nonlinear Sci., 30:23–66, 2020.

[178] H. Sidky, W. Chen, and A. L. Ferguson. High-resolution markov state models for the dynamics of trp-cage miniprotein constructed over slow folding modes identified by state-free reversible vampnets. J. Phys. Chem. B, 123(38):7999–8009, 2019.

[179] F. Paul, H. Wu, M. Vossel, B. L. de Groot, and F. Noé. Identification of kinetic order parameters for non-equilibrium dynamics. J. Chem. Phys., 150(16):164120, 2019.

[180] https://github.com/Ferg-Lab/snrv.

[181] J. Jortner and A. Gaathon. Effects of phase density on ionization processes and electron localization in fluids. Can. J. Chem., 55(11):1801–1819, 1977.

[182] T. R. Gingrich, J. M. Horowitz, N. Perunov, and J. L. England. Dissipation bounds all steady-state current fluctuations. Phys. Rev. Lett., 116(12):120601, 2016.

[183] J. Li, J. M. Horowitz, T. R. Gingrich, and N. Fakhri. Quantifying dissipation using fluctuating currents. Nat. Commun., 10(1):1666, 2019.

[184] A. Albaugh, R.-S. Fu, G. Gu, and T. R. Gingrich. Limits on the precision of catenane molecular motors: Insights from thermodynamics and molecular dynamics simulations. J. Chem. Theory Comput., 20(1):1–6, 2024.