

Supporting Information for Data-Driven Reaction Coordinate
Discovery in Overdamped and non-Conservative Systems:
Application to Optical Matter Structural Isomerization

Shiqi Chen^{1,3}, Curtis W. Peterson^{1,3}, John A. Parker^{2,3}, Stuart A. Rice^{1,3}, Andrew L.
Ferguson^{4,*}, and Norbert F. Scherer^{1,3,*}

¹*Department of Chemistry, University of Chicago, Chicago, IL 60637, USA*

²*Department of Physics, University of Chicago, Chicago, IL 60637, USA*

³*James Franck Institute, University of Chicago, Chicago, IL 60637, USA*

⁴*Pritzker School of Molecular Engineering, University of Chicago, Chicago, IL 60637, USA*

^{*}*nfschere@uchicago.edu, andrewferguson@uchicago.edu (orcid.org/0000-0002-8829-9726)*

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

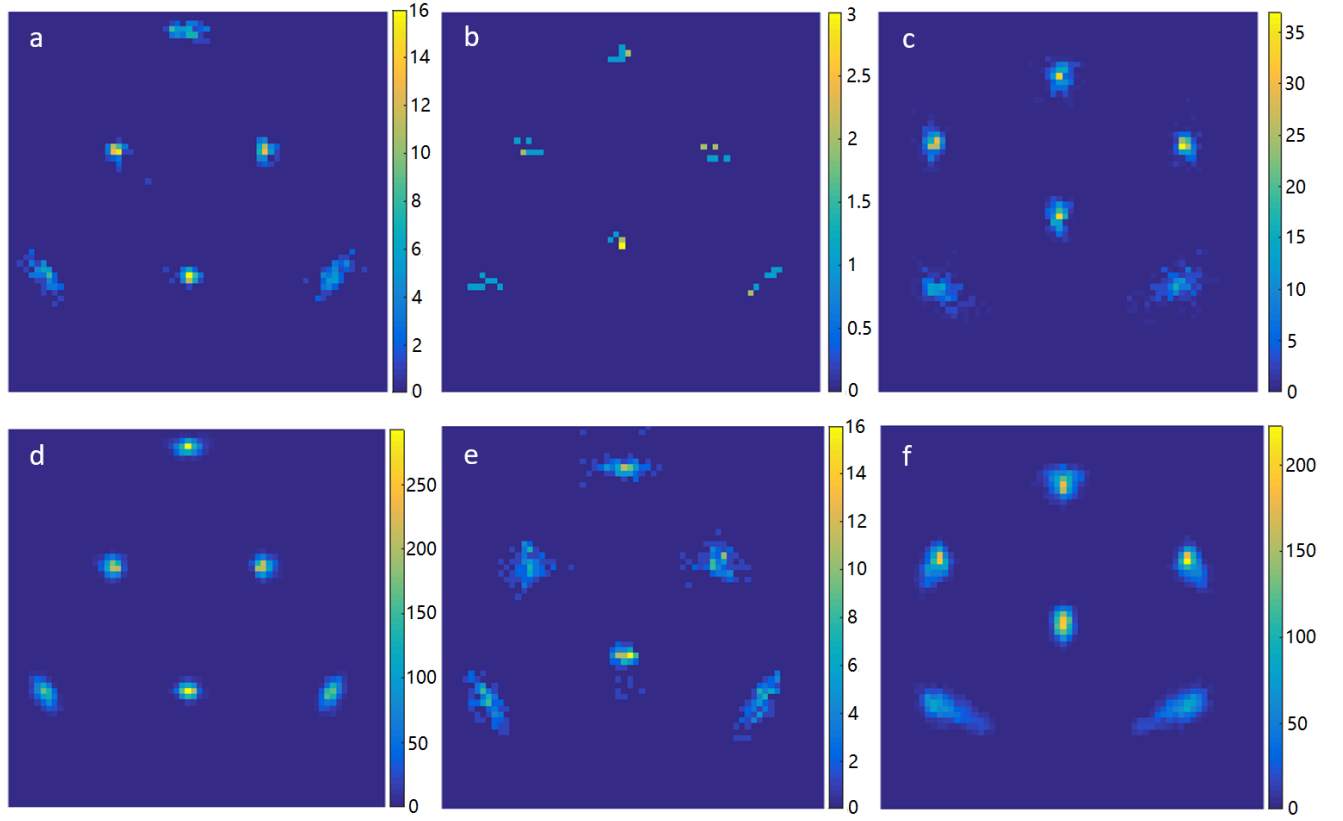


Figure S1: Conditional probability distributions of constituent nanoparticles of 6-particle optical matter clusters conditioned on the distance between the two particles on the (vertical) symmetry axis (cf. **Fig. 2**). **(a-c)** Experimental distributions collected over 1686 configurations at an optical power of 50 mW. **(d-f)** Simulation result collected 21186 configurations at an optical power of 70 mW.

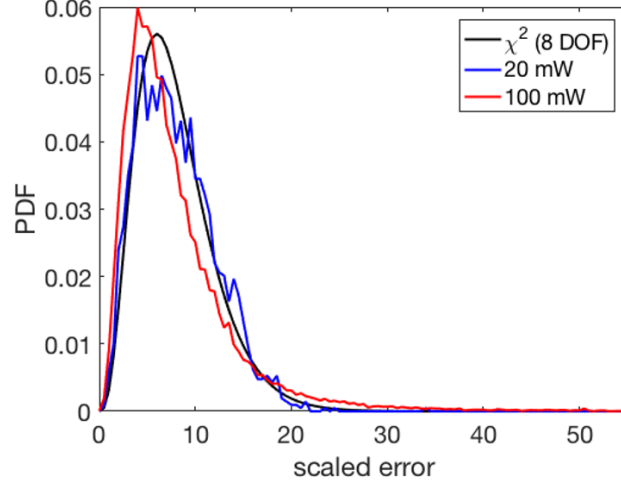


Figure S2: Probability density function (PDF) representation of lattice fitting displacement distributions of the 6-particle optical matter system corresponding to the cumulative density functions shown in Figure 3. The lattice fitting displacement is the sum of squares of particle position deviations of a configuration from the stable triangle lattice sites minimized over all possible translations, rotations, and the lattice parameter. 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).

Fig. 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. S3** shows some of the variations for representative transitions. In general, mode 3 is dominant for almost all transitions, but different modes can contribute significantly at various times for the structural transition.

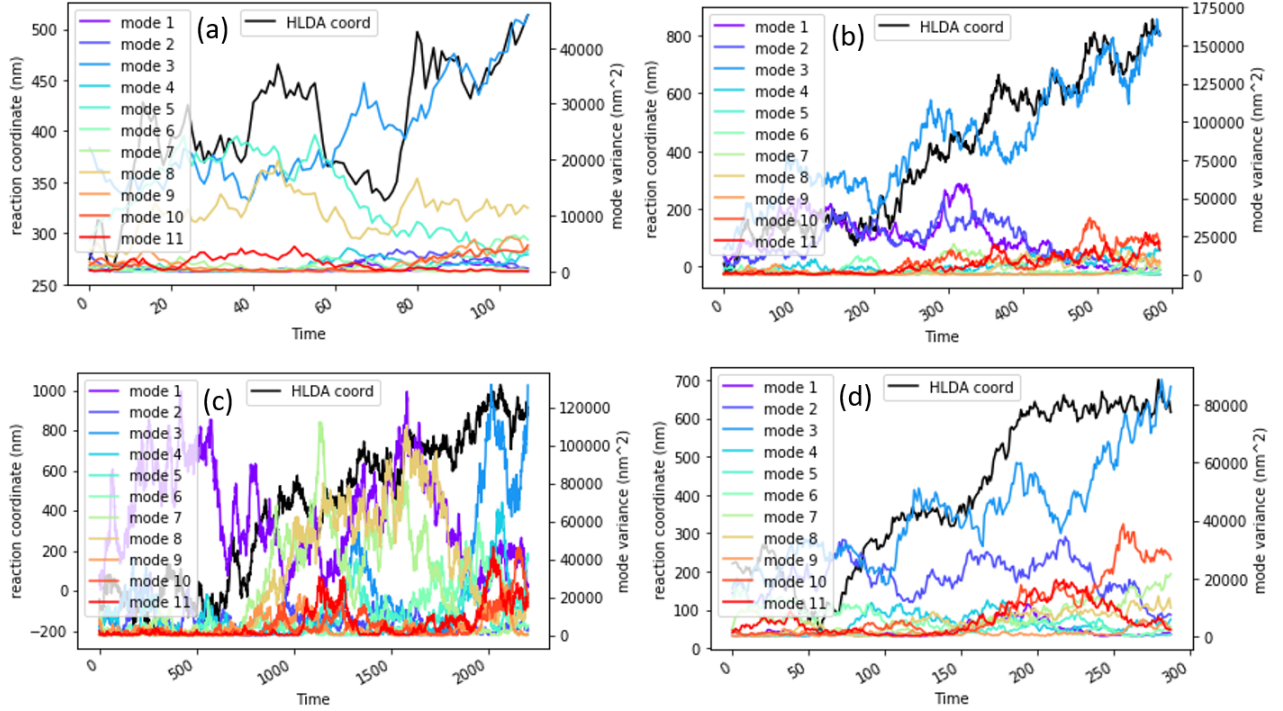


Figure S3: Comparison of the HLDA reaction coordinate s_{HLDA} and the variance of particle deviations along each of the 12 collective modes over the course of four additional triangle-to-chevron isomerization transitions (cf. **Fig. 6(a)**). Some extracted transitions occur quickly like **Fig. S3(a)**, but some are more prolonged like **Fig. S3(c)**. However, at the end of the transition, mode 3 always dominates.

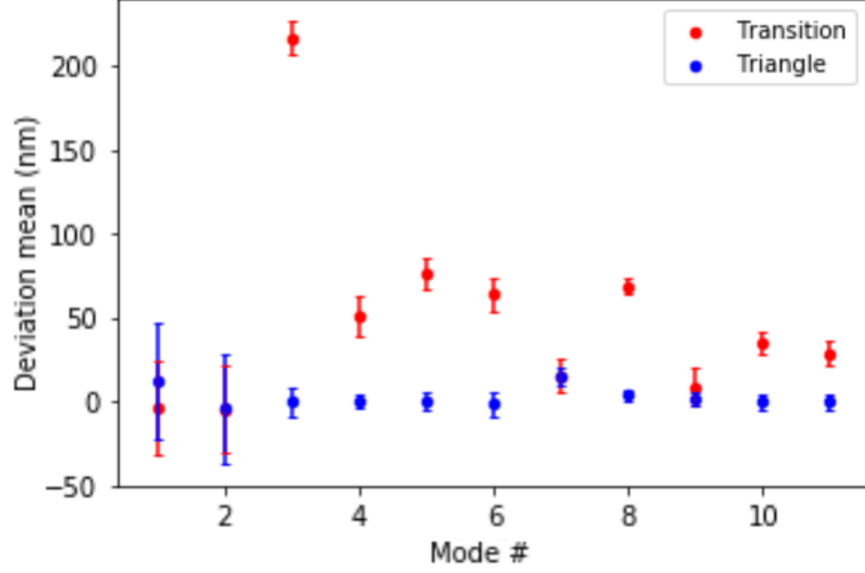


Figure S4: The error bars plotted for the mean of the deviations projected on to the 11 modes for triangle and transition configurations plotted in **Fig. 6(c)**.

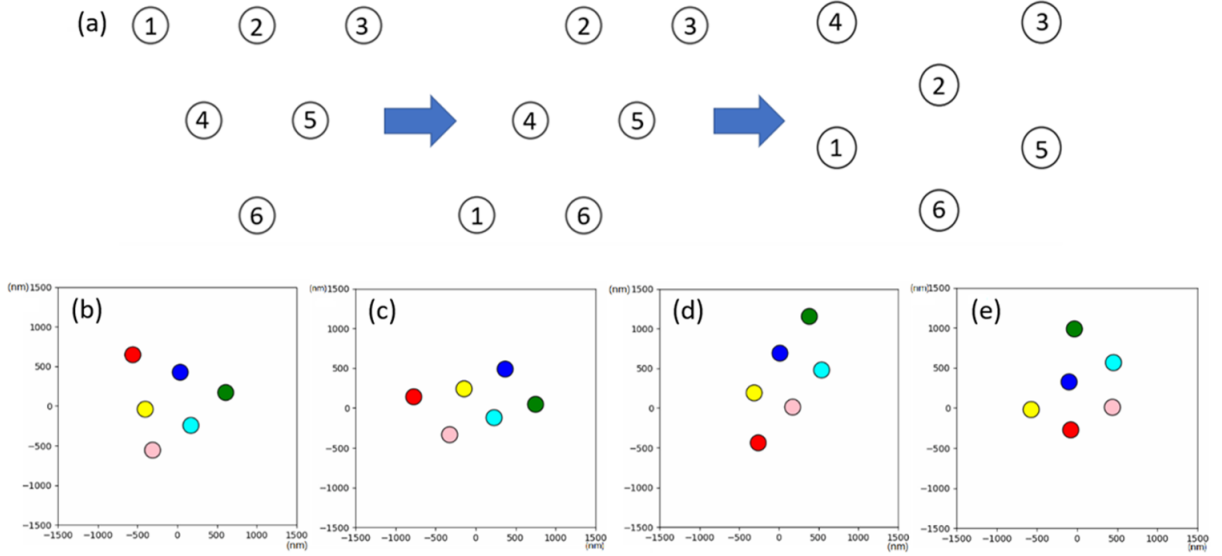


Figure S5: An indirect triangle-chevron transition. (a) The transition scheme with labeled particles. (b-c) Initial and final states of first half of transition. (d-e) Initial and final states of second half of transition.

Video S1. Experimental and simulation movies of the triangle-to-chevron transitions in the six-particle optical matter system.

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

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

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

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

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

In order to minimize the expression above, it is obvious to set the last term of the expression above to zero, which means:

$$p_0^* = \frac{\mathbf{1}^H}{N} (\mathbf{q} - \mathbf{p}) \quad (3)$$

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 $\mathbf{p}$ and $\mathbf{q}$ are rotated or linearly scaled. Therefore, after the choice of p_0^* , further optimization of θ and a has no perturbation on p_0^* , so that it can be carried out successively and independently.

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

$$Err^* = \left[\min_{a \in \mathbb{R}^+, \theta \in [0, 2\pi)} \left| e^{i\theta} \mathbf{p}' - a \mathbf{q} \right|^2 \right]^{\frac{1}{2}} \quad (4)$$

$$\begin{aligned}
Err^2|_{p_0=p_0^*} &= \left| e^{i\theta} \mathbf{p}' - a \mathbf{q} \right|^2 = \left| \mathbf{p}' - a e^{-i\theta} \mathbf{q} \right|^2 = |\mathbf{p}'|^2 + a^2 |\mathbf{q}|^2 - 2a \text{Re}(e^{i\theta} \mathbf{q}^H \mathbf{p}') \\
&= a^2 |\mathbf{q}|^2 - 2(a |\mathbf{q}|) \left[\frac{\text{Re}(e^{i\theta} \mathbf{q}^H \mathbf{p}')}{|\mathbf{q}|} \right] + \left[\frac{\text{Re}(e^{i\theta} \mathbf{q}^H \mathbf{p}')}{|\mathbf{q}|} \right]^2 - \left[\frac{\text{Re}(e^{i\theta} \mathbf{q}^H \mathbf{p}')}{|\mathbf{q}|} \right]^2 + |\mathbf{p}'|^2 \quad (5) \\
&= |\mathbf{q}|^2 \left[a - \frac{\text{Re}(e^{i\theta} \mathbf{q}^H \mathbf{p}')}{|\mathbf{q}|^2} \right]^2 + \left\{ |\mathbf{p}'|^2 - \left[\frac{\text{Re}(e^{i\theta} \mathbf{q}^H \mathbf{p}')}{|\mathbf{q}|} \right]^2 \right\}
\end{aligned}$$

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 $e^{i\theta} \mathbf{q}^H \mathbf{p}'$ is maximized to be $|\mathbf{q}^H \mathbf{p}'|$. Therefore:

$$e^{i\theta^*} = e^{-i\arg(\mathbf{q}^H \mathbf{p}')} = \frac{(\mathbf{p}')^H \mathbf{q}}{|\mathbf{q}^H \mathbf{p}'|} \quad (6)$$

$$a^* = \frac{\text{Re}(e^{i\theta} \mathbf{q}^H \mathbf{p}')}{|\mathbf{q}|^2} = \frac{|\mathbf{q}^H \mathbf{p}'|}{|\mathbf{q}|^2} \quad (7)$$