Supplementary Information: Understanding and Design of Non-Conservative Optical Matter Systems Using Markov State Models

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

[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)*

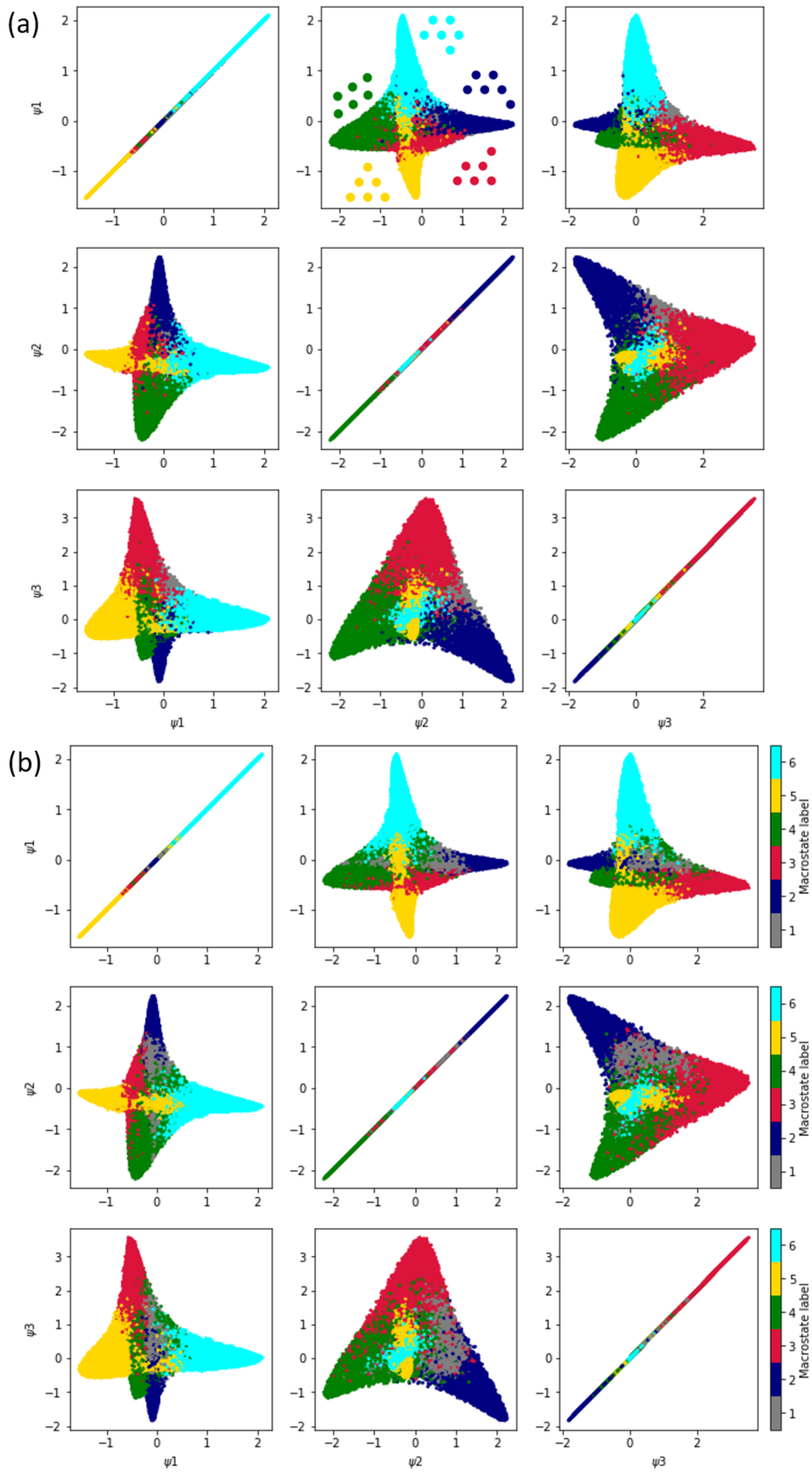

Figure S1. Diffusion map embedding plot colored according to (a) lattice pattern (b) macrostate labels generated for $(t, n_M, \tau) = (8.3, 6, 10\text{ ms})$ of one simulation trajectory for 50 mW beam power. The grey spots in panel (a) represents the configurations that are not identified as any of 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.

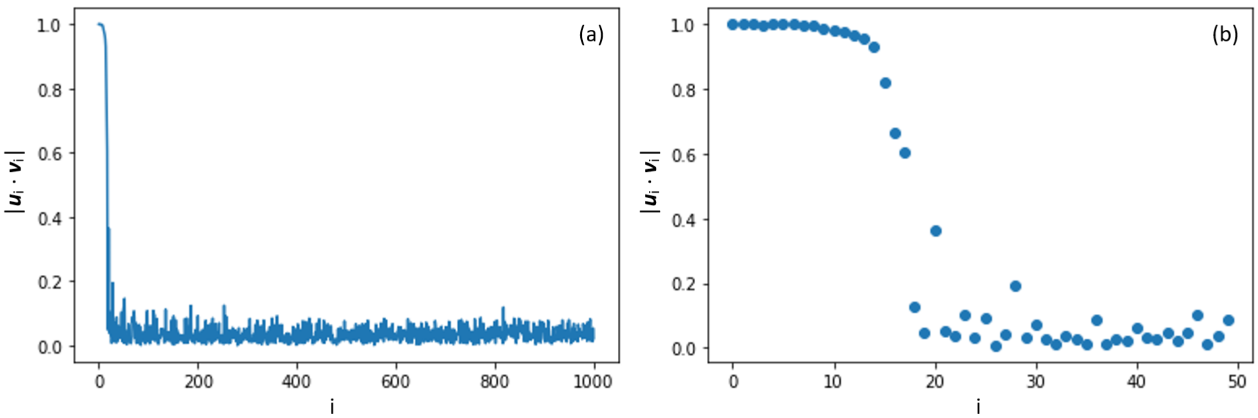

Figure S2. Plot of the absolute values of the dot products (cosine similarity) of the normalized right eigenvectors of the reversibilized and original microstate transition matrices for the 6-particle OM system under 50 mW with $(\tilde{t}, \tau)$=(8.3, 10 ms). $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ 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.

For the case with 50 mW beam power demonstrated in section 3.1 and 3.2, let the macrostate count matrix be C_{macro} and the microstate count matrix be C_{micro}. We have,

$$\begin{aligned} \frac{\|C_{micro} - C_{micro}^T\|_F}{\|C_{micro}\|_F} &= 0.226 \\ \frac{\|C_{macro} - C_{macro}^T\|_F}{\|C_{macro}\|_F} &= 0.016 \end{aligned} \tag{1}$$

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).

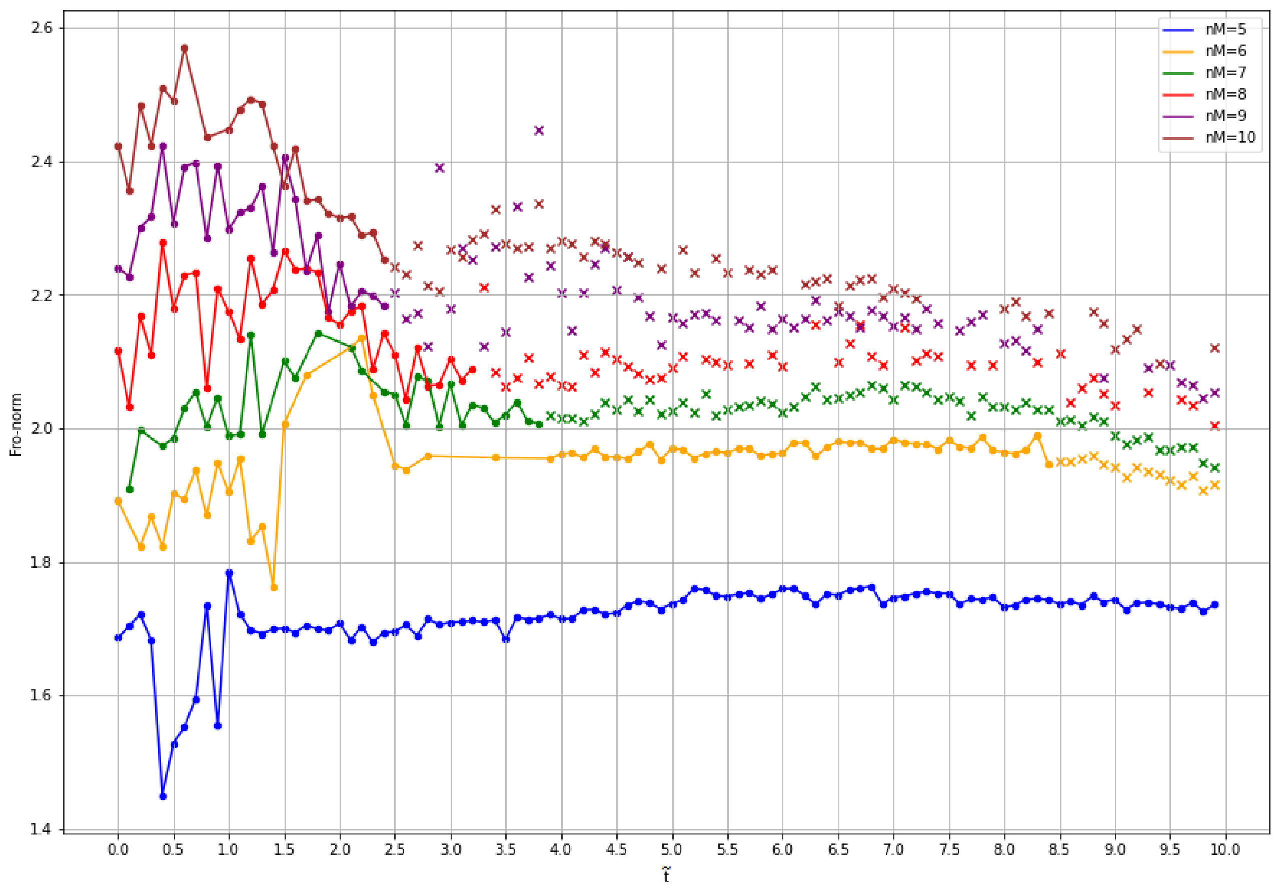

Figure S3. Plot of the Frobenius norm of the row-normalized assignment matrices against diffusion k-means parameter $\tilde{t}$ for various number of macrostates with lag time τ = 10 ms for 6-particle OM system at a 50 mW beam power. For each value of $(n_M, \tilde{t})$ we render the corresponding data point in one of three ways. (1) If at least $(n_M - 1)$ implied time scales are resolved (i.e., are greater than the τ = 10 ms lag time and lie above the grey area in Fig. 2(a)) 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.

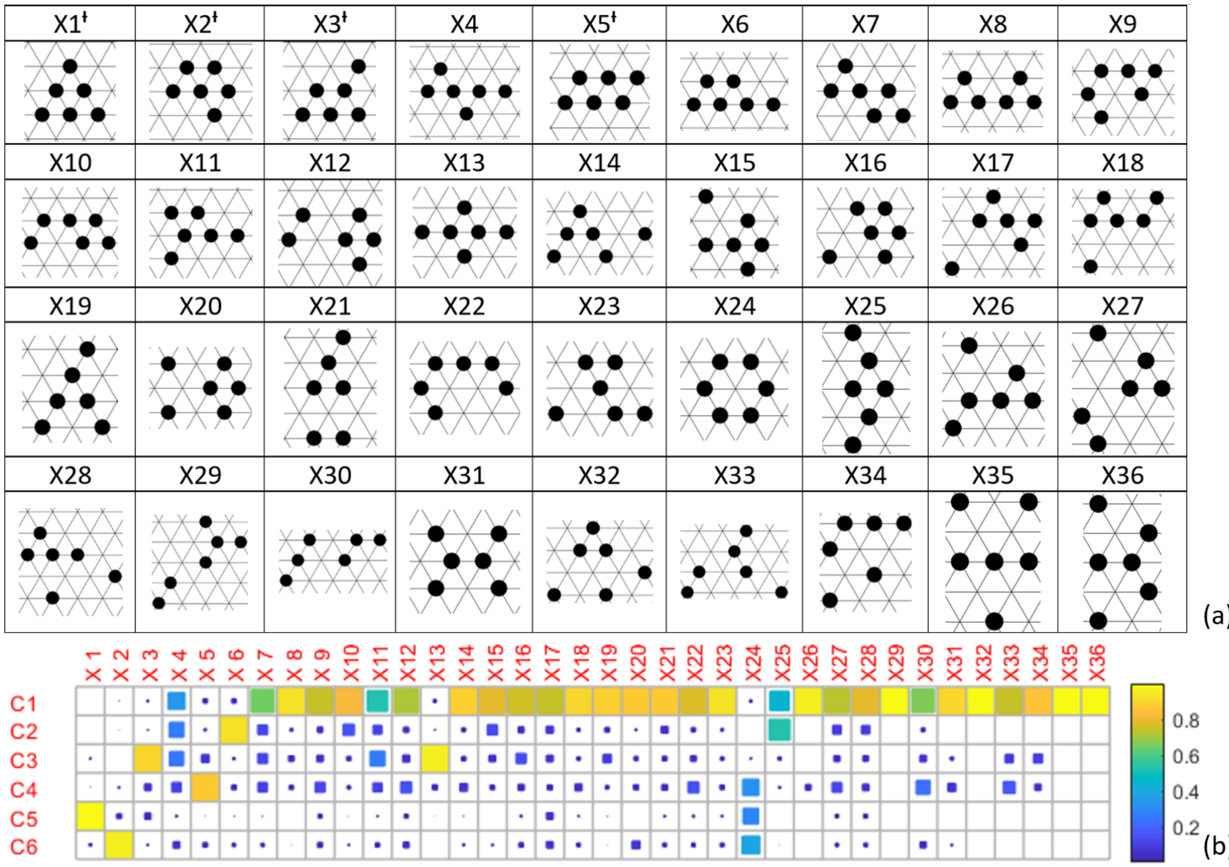

Figure S4. (a) The representative lattice patterns of the lattice labels X1-36. The lattice labels with the superscript † correspond to a unique lattice pattern while other lattice labels correspond to more than one lattice pattern. (b) Illustration of the column normalized assignment matrix (CNAM) of 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.

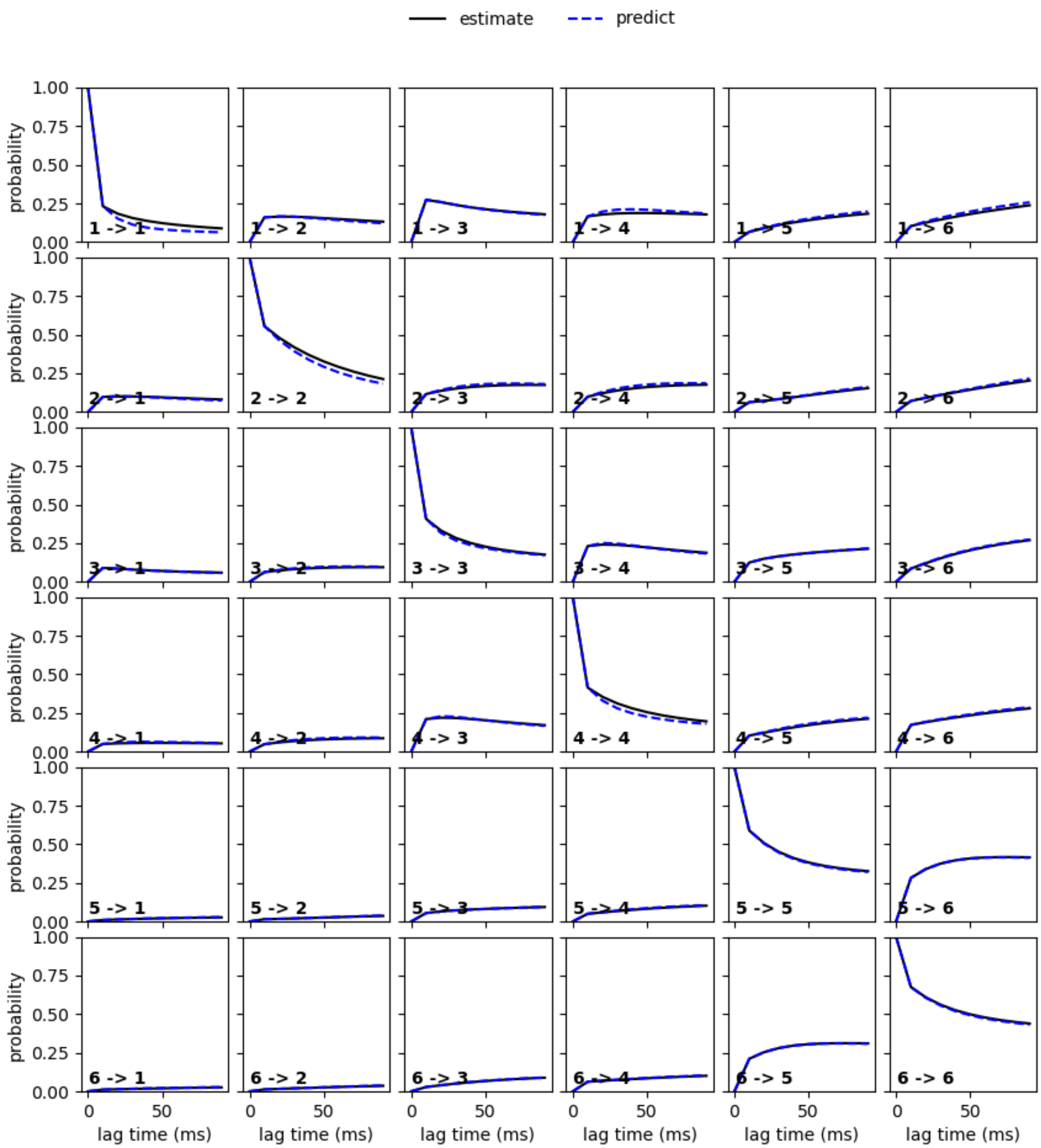

Figure S5. The Chapman-Kolmogorov (CK) test result computed for the Markov state model built for the 6-particle OM system under 50 mW beam power with parameter set $(\tilde{t}, n_M, \tau) = (8.3, 6, 10\text{ ms})$. The CK test compares the transition probability elements of a macrostate transition matrix $\boldsymbol{T}(k\tau)$ 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.

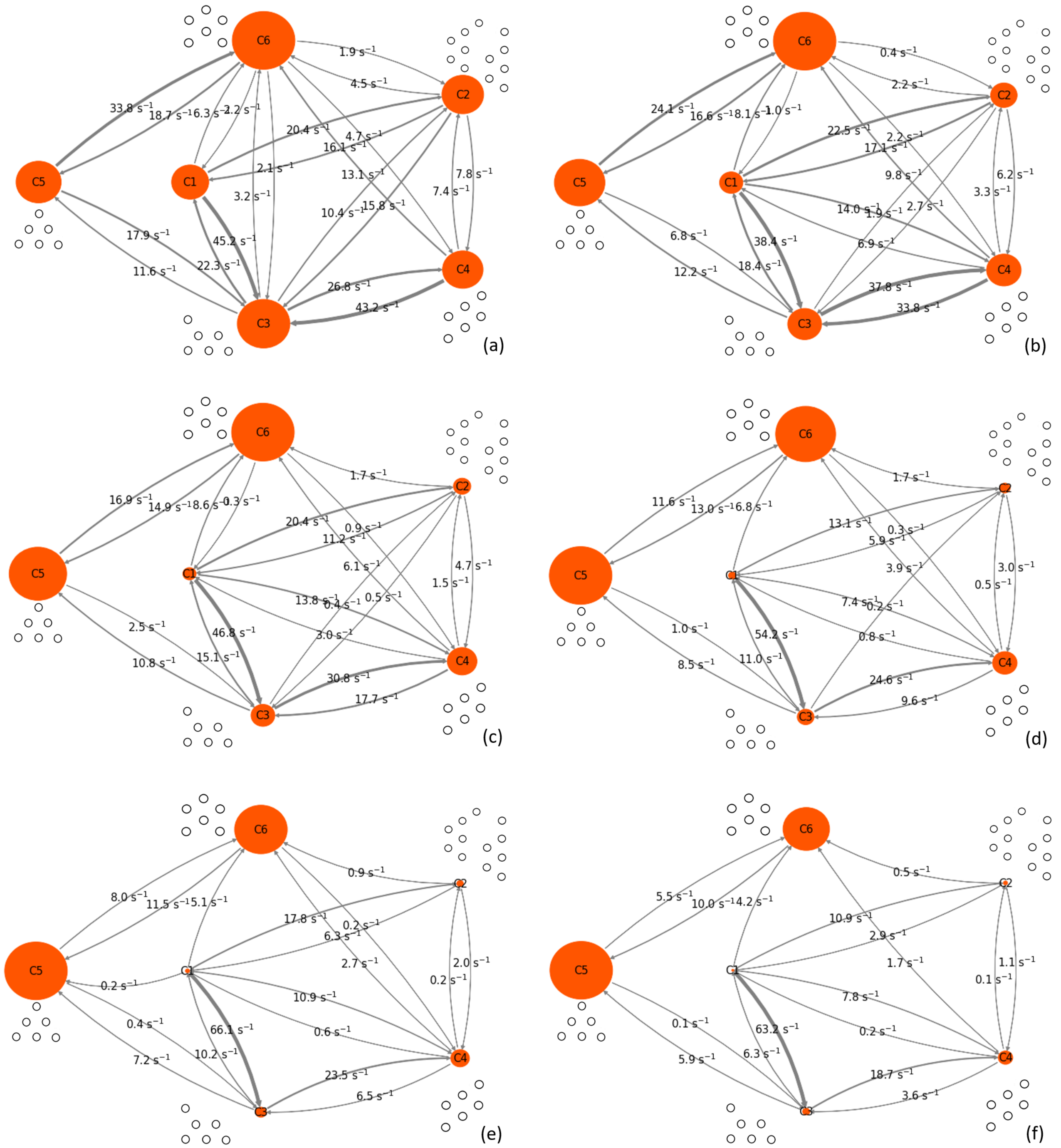

Figure S6. The state maps of the Markov state models built for the 6-particle OM system with $\tau = 10$ ms, $n_M = 6$. The beam powers and the values of $\tilde{t}$ corresponding to the panels are: (a) (40 mW, 6.3); (b) (50 mW, 8.3); (c) (60 mW, 7.3); (d) (70 mW, 5.4); (e) (80 mW, 5.3); (f) (90 mW, 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.

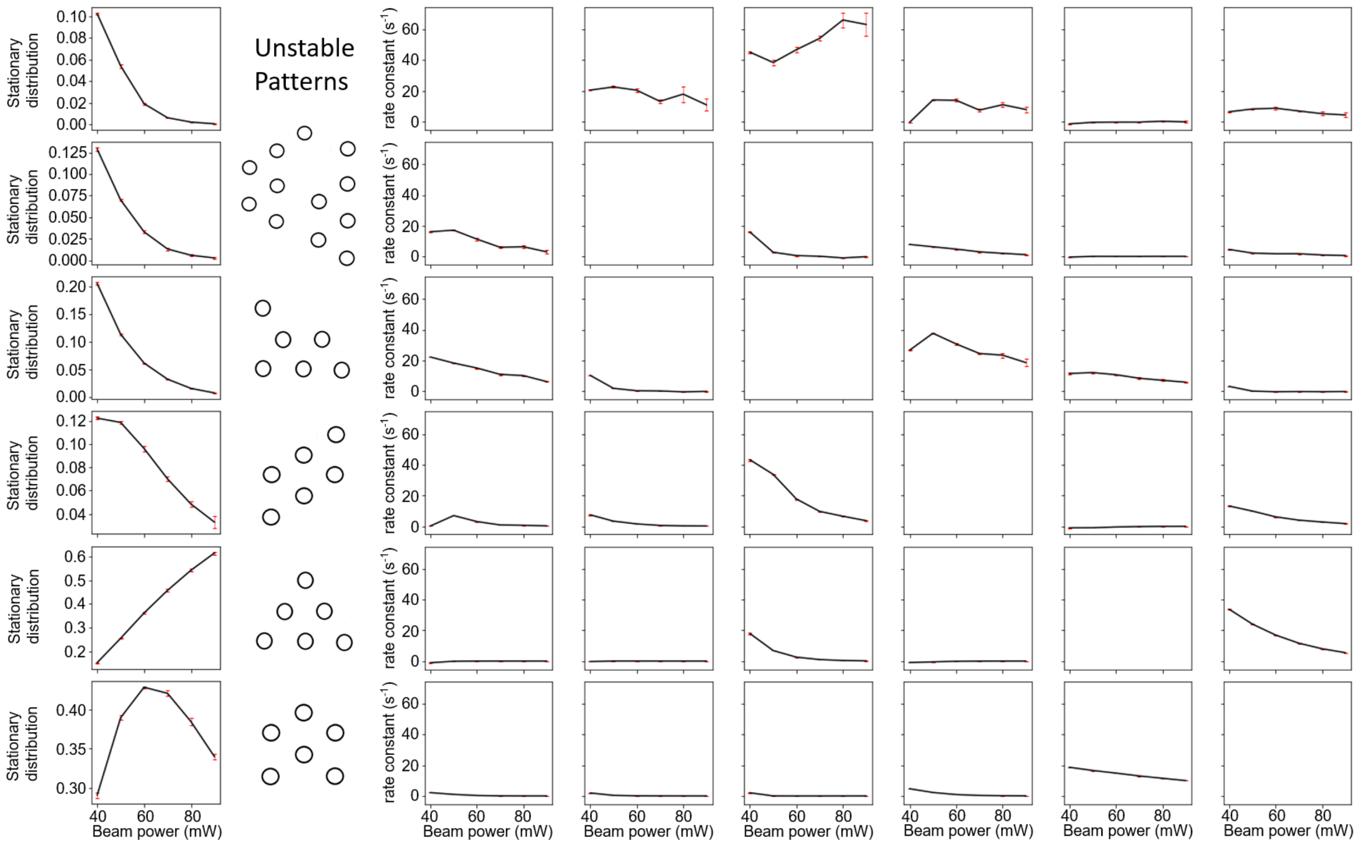

Figure S7. Stationary distribution probabilities of the macrostates and the rate constants of the transitions among the macrostates as a function of beam power for 6-particle OM system. The left-most column contains the stationary distribution probabilities and are annotated by the corresponding lattice 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.