Supplementary information for "A machine learning approach to drawing phase diagrams of topological lasing modes"

Stephan Wong,[1,2] Jan Olthaus,[3] Thomas K. Bracht,[3] Doris E. Reiter,[3,4] and Sang Soon Oh[1,*]

[1] *School of Physics and Astronomy, Cardiff University, Cardiff CF24 3AA, UK*
[2] *Center for Integrated Nanotechnologies, Sandia National Laboratories, Albuquerque, New Mexico 87185, USA*
[3] *Institut für Festkörpertheorie, Universität Münster, 48149 Münster, Germany*
[4] *Condensed Matter Theory, Department of Physics, TU Dortmund, 44221 Dortmund, Germany*

(Dated: May 2, 2023)

Supplementary note 1. BASIS GENERATION METHODS

This supplementary section review few methods for generating a basis from a time series. There are many ways to generate the low-order model of a given dynamical behaviour. An important quantity is the so-called data matrix X built from the data at hand. The data matrix is a $(N_s \times N_t)$-matrix that collects the N_t data snapshots $x(t_i)$ into columns:

$$X = [x(t_0), x(t_1), \ldots, x(t_{N_t})]\,. \tag{S1}$$

Here, the vector $x(t_i)$ is chosen to be the complex-valued amplitudes of the modes at the A and B sites. Other "observables", such as the absolute values or the total intensity per sublattices, can be used. This may give different (better or worse) results and is left for a future study. The bases can then be constructed using dimensional reduction techniques on the data matrix. Here, we will cover different methods in order to highlight their importance and limitations for the classification scheme used.

A. Proper orthogonal decomposition

Proper orthogonal decomposition (POD) [1] is a commonly used tool for dimensional reduction of physical systems. This decomposition relies on the singular value decomposition (SVD) of the data matrix, given by:

$$X = U\Sigma V^\dagger \tag{S2}$$

where U and $V^\dagger$ are $(N_s \times N_s)$ and $(N_s \times N_t)$ unitary matrices, respectively. Σ is a diagonal $(N_0 \times N_0)$-matrix diag $= (\sigma_1, \ldots, \sigma_{N_0})$, with $N_0 = \min(N_s, N_t)$. The diagonal entries of Σ are the so-called singular values and are ordered in ascendant order $\sigma_1 > \sigma_2 > \ldots > \sigma_{N_0} \geq 0$. The SVD gives us two orthonormal bases U and $V^\dagger$ since the matrices U and $V^\dagger$ are unitary matrices. The columns of U are ordered according to the variance σ_i they capture in the data matrix and are called the singular vectors: these are the POD modes that are used in the basis Φ. Moreover, the POD modes are complex because of the complex data X.

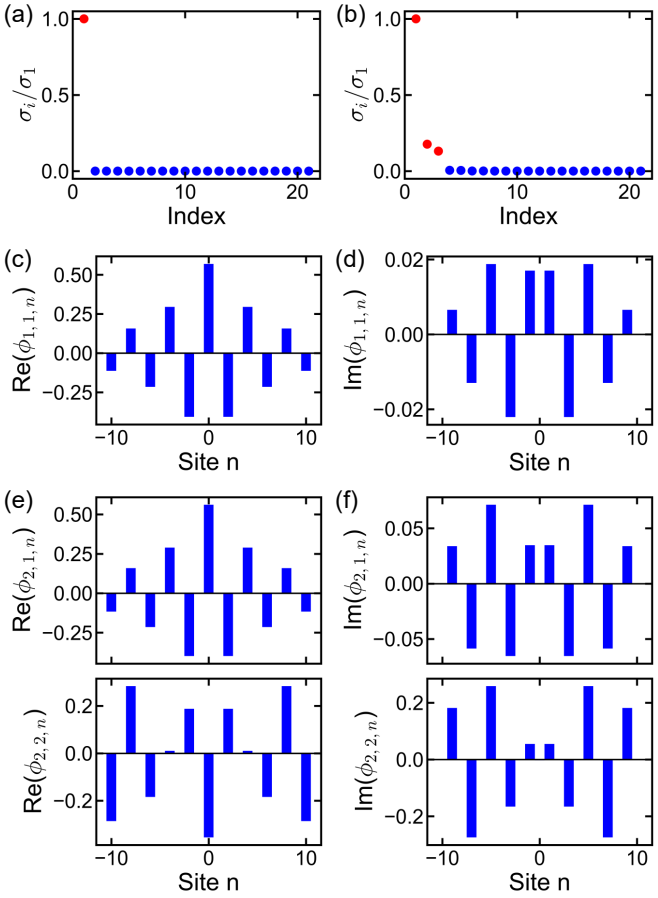

FIG. S1. **POD method.** Singular values of the (a) non-oscillating and (b) oscillating regimes. The red dots correspond to the singular values accumulating 99% of the total variance of the data. (c) Real and (d) imaginary parts of the field profile of the first POD mode for the non-oscillating regime. (e) Real and (f) imaginary parts of the field profile of the (top) first and (bottom) second POD mode for the oscillating regime. The POD bases have been generated from the time series starting at the 1800-th time step.

For a low-dimensional attractor, the POD basis can be safely truncated at a cut-off value r while retaining the main information of the data matrix. Explicitly, the

* Email: ohs2@cardiff.ac.uk

SVD reads:

$$X_{im} = \sum_{n=0}^{N_0} U_{in}\sigma_n V^\dagger_{nm}, \tag{S3}$$

and keeping only the r highest terms in the decomposition [Eq. S3], we have the approximation:

$$X_{im} \simeq \sum_{n=0}^{r} U_{in}\sigma_n V^\dagger_{nm}. \tag{S4}$$

This is re-written, in a matrix form, as:

$$X \simeq U_r \Sigma_r V_r^\dagger \tag{S5}$$

where U_r, Σ_r and $V_r^\dagger$ are the truncated matrix of U, Σ and $V^\dagger$, respectively. Although the cut-off value r can be chosen based on different criteria [2], r is typically chosen so that the POD modes retain a certain amount of the variance (or energy) σ_X in the data, namely:

$$\sum_{i=0}^{r} \sigma_i > \sigma_X. \tag{S6}$$

Figure S1 displays the POD method of the non-oscillating and oscillating regimes in the domain-wall SSH lattice with saturable gain [Fig. 1]. The truncation has been chosen such that $\sigma_X = 99\%$ of the total variance is retained. We observe in Figs. S1(a) and S1(b) the normalised singular values, and that a single POD mode is retained for the zero-mode-like, whereas three POD modes are needed for the oscillating regime, as marked by the red dots. The real and imaginary parts of the field profile of the corresponding first few POD modes are plotted in Figs. S1(c) and S1(d) and Figs. S1(e) and S1(f) for the zero-mode-like non-oscillating and oscillating regimes, respectively. One can see that the main spatial feature of the zero-mode is captured in the single POD mode obtained after truncation, where the majority of its amplitudes are on the A sublattice. On the other hand, the POD modes of the oscillating dynamical regime also capture part of the information with some finite amplitudes on the A and B sublattices.

Importantly, in this decomposition [Eq. S3], SVD is implicitly doing a space-time separation of the data matrix, where the POD modes U contain the spatial information while V have the temporal information at each spatial grid point. Therefore, the POD modes give a static basis and do not explicitly model the temporal dynamics of the time series. This method will therefore most likely fail to identify the correct dynamical regime in the classification step.

B. Dynamical mode decomposition

Dynamical mode decomposition (DMD) [1, 3, 4] is an alternative to the POD method for learning the dynamics of non-linear systems. DMD is an explicitly temporal

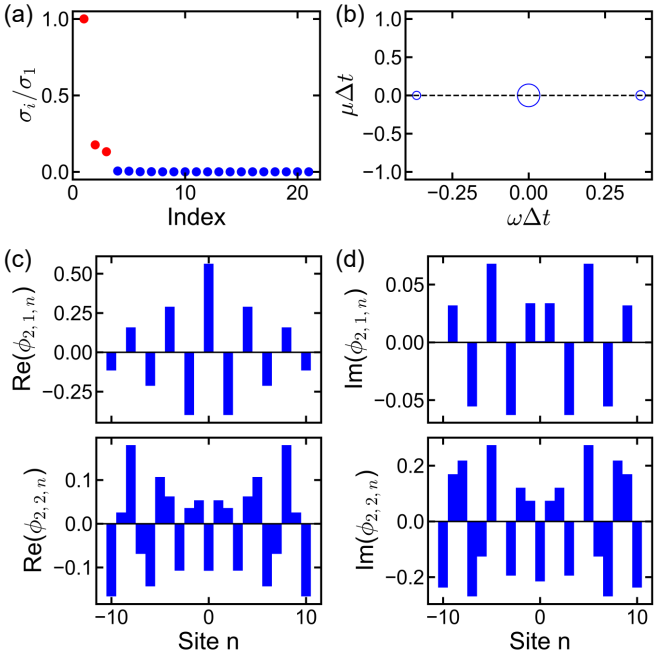

FIG. S2. **DMD method.** (a) Singular values of the oscillating regime in the DMD. The red dots correspond to the singular values accumulating 99% of the total variance of the data. (b) Plot of the logarithm of the DMD values $\ln(\lambda)$ in the complex plane. The size of the open circle is proportional to their corresponding singular values. (c) Real and (d) imaginary parts of the field profile of the (top) first and (bottom) second DMD mode for the oscillating regime. The DMD basis has been generated from the time series starting at the 1800-th time step.

decomposition that takes the sequences of snapshots into account, and is able to derive the spatio-temporal patterns of the data matrix X. The dynamics of the system are taken into account by considering a linear matrix A that maps a data matrix X_1, starting at some time steps (t_1), to the data matrix X_2, starting at the next time step (t_2). The matrix A is thus defined as:

$$X_2 = AX_1, \tag{S7}$$

and the corresponding data matrices are given by:

$$X_1 = [x(t_1), x(t_2), \ldots, x(t_{N_t-1})] \tag{S8}$$

and

$$X_2 = [x(t_2), x(t_3), \ldots, x(t_{N_t})]\,. \tag{S9}$$

Interestingly, Equation S7 is similar to a linear stability analysis formulation for discrete maps if we think of the stability matrix as the linear matrix A. The DMD method thus consists of solving the following eigenvalues problem:

$$A\Phi = \Phi\Lambda \tag{S10}$$

where the columns of Φ are the DMD modes ϕ_i and the corresponding DMD eigenvalues λ_i are the diagonal entry

of Λ. The DMD modes ϕ_i give us the spatial eigenmodes while their corresponding eigenvalues λ_i have their temporal information. Using a change of units from the data snapshots, observed at every Δt, to units in time, the eigenvalues are complex-valued scalars:

$$\frac{\ln(\lambda_i)}{\Delta t} = \mu_i + i\omega_i \tag{S11}$$

where μ_i gives the growth (decay) rate if $\mu_i > 0$ ($\mu_i < 0$) and ω_i the oscillation frequency of the DMD modes ϕ_i.

However, the size of the data matrix usually makes the eigendecomposition not feasible. The goal, here, is therefore to approximate the eigenvalues and eigenvectors of A, using only the data matrices X_1 and X_2. The idea is to start by the truncated SVD of $X_1 = U_r \Sigma_r V_r^\dagger$ in which Eq. S7 becomes:

$$X_2 = A U_r \Sigma_r V_r^\dagger. \tag{S12}$$

Then the linear matrix A is reduced by considering its projection onto the truncated POD subspace:

$$A_r := U_r^\dagger A U_r = U_r^\dagger X_2 V_r \Sigma^{-1}. \tag{S13}$$

The eigenvalue problem for A_r is solved with:

$$A_r W = W\Lambda, \tag{S14}$$

from which we have the relation:

$$\Phi = X_2 V \Sigma^{-1} W. \tag{S15}$$

The key feature of the DMD method is that it decomposes the data into a set of coupled spatio-temporal modes. The DMD resembles a mixture of the POD in the spatial domain and the discrete Fourier transform (DFT) in the time domain. Figure S2 shows the DMD results for the oscillating regime. We can, indeed, see in Fig. S2(a) that the singular values are similar to that of the POD. Besides, we observe in Figs. S2(c) and S2(d) that the field profile of the DMD modes closely resembles the POD modes in Fig. S1. The largest DMD modes not only look similar to the POD modes, but they also contain the oscillation frequencies from ω_i, as in DFT. The DMD even goes beyond DFT by giving an estimate of the growth (decay) rate in time via $\mu_i > 0$ ($\mu_i < 0$). This can be seen by plotting the DMD modes, scaled by their contribution in the decomposition σ_i, in the frequency plane of λ_i. We can see in Fig. S2(b) that the dynamical regime has a single DMD mode with $\omega_i = 0$ akin to the offset of the oscillation amplitudes, and two DMD modes with opposite $\omega_i \neq 0$ corresponding to the oscillating behaviours. All the above three DMD modes have vanishing growth or decay rate $\mu_i = 0$.

C. Time-augmented dynamical mode decomposition

Although the DMD gives the temporal behaviours of the non-linear system, the temporal information is not fully incorporated into the DMD basis Φ since only the DMD modes are used. Exploiting the time evolution in the dynamical regime requires the use of DMD modes along with their eigenvalues. The idea is therefore to incorporate the dynamic information by augmenting the basis Φ [5]. This time-augmented DMD will be denoted by aDMD in the remaining of this chapter. Using the defining relation of the eigenvalues λ_i as similar to a time evolution operator, *i.e.* multiplying by λ is the same as shifting by one time step, we have for a given DMD mode ϕ_i, its evolution given by $\lambda_i^{N_w}\phi_i$ at N_w time step ahead in time. Therefore, the time-augmented basis vector reads:

$$\begin{bmatrix} \phi_i \\ \lambda_i \phi_i \\ \vdots \\ \lambda_i^{N_w}\phi_i \end{bmatrix}. \tag{S16}$$

By considering a time window N_w, the time-augmentation of the DMD basis provides us with the dynamical information of the non-linear regime. Figure S3 shows the aDMD results for the same oscillating regime as in Fig. S2. We can see that the singular values and aDMD eigenvalues plots [Figs. S3(a) and S3(b)]

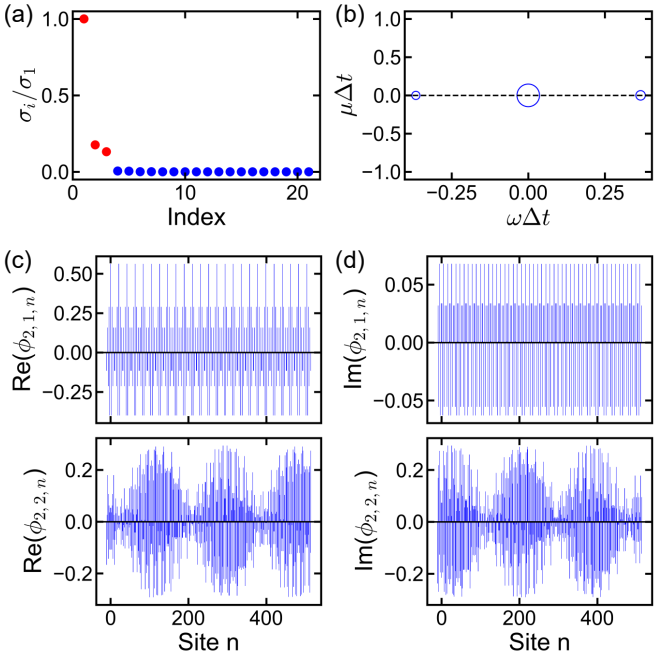

FIG. S3. **aDMD method.** (a) Singular values of the oscillating regime in the aDMD. The red dots correspond to the singular values accumulating 99% of the total variance of the data. (b) Plot of the logarithm of the aDMD values $\ln(\lambda)$ in the complex plane. The size of the open circle is proportional to their corresponding singular values. (c) Real and (d) imaginary parts of the "field profile" of the (top) first and (bottom) second aDMD mode for the oscillating regime. Here, by "Site n", we mean the n-th entry of the eigenvector. The aDMD basis has been generated with $N_w = 25$ from the time series starting at the 1800-th time step.

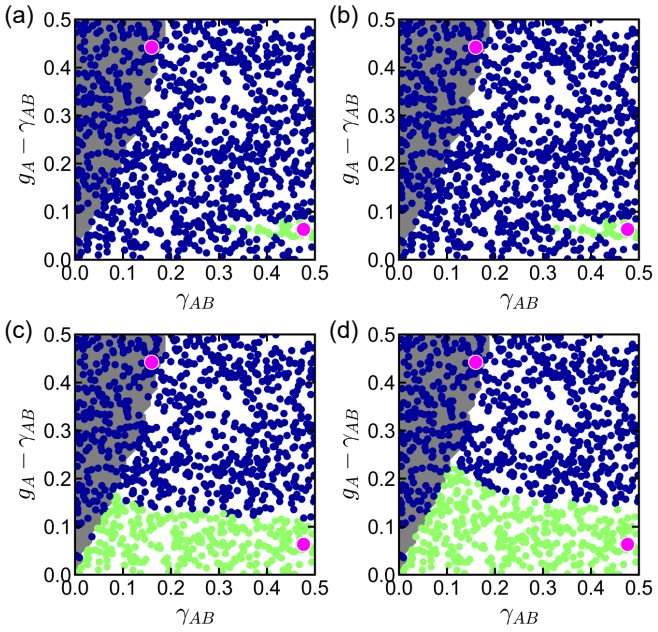

FIG. S4. **Phase diagram derived using different decomposition methods.** Phase diagrams obtained from a library composed of two regimes (one non-oscillating and one oscillating) from which the bases have been generated using the (a) POD, (b) DMD, (c) aDMD with $N_w = 5$ and (d) aDMD with $N_w = 25$. The green and blue dots correspond respectively to the identified non-oscillating and oscillating regimes. The magenta dots represent the regimes used for the construction of the library. These are located at $(\gamma_{AB}, g_A - \gamma_{AB}) = (0.48, 0.06)$ and $(0.16, 0.44)$. The white and grey areas are overlays of the referenced phase diagram obtained in Fig. 1. The bases have been generated from the time series starting at the 1800-th time step.

are the same as in the DMD algorithm [Figs. S2(a) and S2(b)] whereas the "field profile" of the aDMD modes [Figs. S3(c) and S3(d)] carry some temporal evolution information. In particular, we observe in the top panel of Figs. S3(c) and S3(d) the static behaviours of the aDMD mode with $\omega_i = 0$. On the other hand, we can see, in the bottom panel of Figs. S3(c) and S3(d), one of the first aDMD modes with $\omega_i \neq 0$ featuring some oscillating behaviour in time. The size of the basis mode is larger than the plain DMD, and can exhibit its time evolution. Nevertheless, the graphs do not exactly plot the temporal evolution of the DMD modes since the first N_s entry of the basis state is for the N_s sites; the next N_s for again the N_s sites but at the next time step, etc.

D. Classification results from different decomposition methods

The classification results [Eq. (7)] from the decomposition methods reviewed is shown in Fig. S4. We observe in Figs. S4(a) and S4(b) that the phase diagrams fail to correctly predict the distinct dynamical regimes. This is expected since the POD or DMD modes do not contain enough information about the temporal behaviours. Besides, the classification for these diagrams is based on a single snapshot ($N_w = 0$). Thus, it is expected the classification fails to capture the correct dynamics since a single snapshot only relies on the spatial pattern of the regime. On the other hand, we can see in Figs. S4(c) and S4(d) that the derived phase diagrams have better accuracy when the bases are time-augmented, or equivalently when the classification scheme uses several snapshots. By increasing the time window in the classification, the derived phase diagram is even better as illustrated in Figs. S4(c) and S4(d) for $N_w = 5$ and $N_w = 25$, respectively. The phase diagram will get improved until the time window is large enough to capture the dynamic behaviour.

Supplementary note 2. SPARSE MEASUREMENT

This section detailed the slight change in the methodology in case of sparse sensing. Sparse sensing is often desirable since the measurement and the data collection can be expensive for a complex system if the space grid is too fine, *i.e.* if N_s is very large. The compressed measurement $y(t)$ is derived from the full-state measurement $x(t)$ and the measurement matrix C:

$$y(t) = Cx(t), \tag{S17}$$

where C is a matrix of size $(N_p \times N_s)$ with N_p the number of measurements. Although the measurement matrix C can be represented by some advance and complex mapping [6], here we focus on point-wise measurements, namely the C_{ij} entry in the matrix measurement corresponds to the i-th measurement at the j-th spatial grid point. Therefore the compressed basis is given by:

$$\Theta = C\Phi \tag{S18}$$

where Φ is the basis obtained from the full-state data collection. The library of bases for the J distinct dynamical regimes is similarly re-written as:

$$\mathcal{L} = \{\Theta_1, \ldots, \Theta_J\}. \tag{S19}$$

Nevertheless, the size of the current SSH lattice is, here, reasonable and allows us to choose the matrix measurement as the identity matrix $C = 1_{N_s}$. We will thus use the full-state instead of sparse measurements, but retain the Θ and $y(t)$ notations to keep the general formalism.

Supplementary note 3. TOP-DOWN LIBRARY GENERATION PRINCIPLE

This section supplement the top-down library generation principle. The general idea of the top-down construction of the library is to start with a library made of a

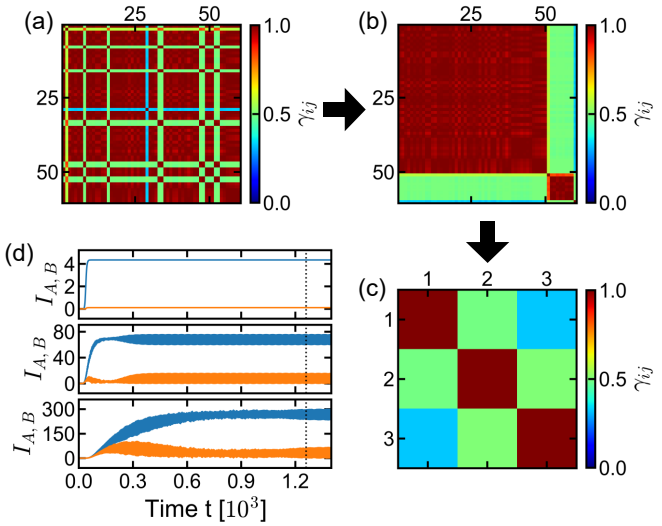

FIG. S5. **Principle of the adaptively top-down library generation.** (a) Subspace alignment matrix for an initial library composed of $J = 60$ regimes. (b) Subspace alignment matrix from the same library as in (a) but grouped into equivalent regimes. (c) Subspace alignment matrix from the reduced library after the equivalent regimes are merged with $\gamma_{th} = 0.75$ in (b). (d) Representative total intensity I_A (and I_B) of the A (and B) sublattice in blue (and orange) for the different regimes. The black vertical dotted line indicates the starting time from which the bases are generated. The aDMD bases have been generated with $N_w = 25$ from the time series starting at the 1800-th time step.

high number of bases, generated from the time series randomly chosen in the given parameter space region, and then merge them into groups of equivalent regimes. Figure S5 illustrates the top-down algorithm. Starting with a library composed of J bases (here $J = 60$), the subspace alignment γ_{ij} is computed [Fig. S5(a)] and then grouped into equivalent regimes according to Eq. (9) in the main text [Fig. S5(b)]. Each representative of the regimes is then randomly selected within each group [Fig. S5(c)]. We plot in Fig. S5(d) the time series of the representative of each regime. These regimes are the non-oscillating and oscillating regimes, as well as a third regime which may correspond to a transient regime. The vertical dashed line in the time series represents the initial time used for constructing the bases in the library.

Supplementary note 4. BOTTOM-UP LIBRARY GENERATION PRINCIPLE

This section supplement the bottom-up library generation principle. Figure S6 illustrates the bottom-up methodology proposed. We start with a single sample in the library, randomly chosen in the parameter space region [Fig. S6(a)]. The library is then adaptively constructed according to the relative reconstruction error ϵ [Fig. S6(b)]. Finally, with the large library at hand, the top-down approach is used to reduce the library size by merging equivalent regimes [Fig. S6(c)]. The representative of the regimes is plotted in Fig. S6(d) and corresponds to the non-oscillating, oscillating and transient regimes, respectively.

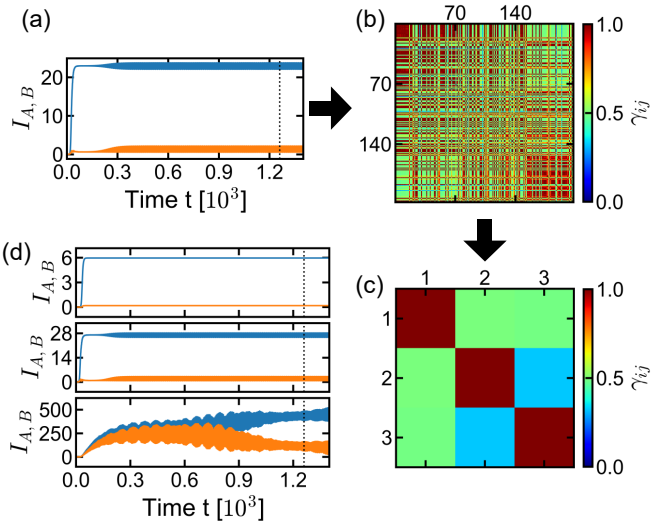

FIG. S6. **Principle of the adaptively bottom-up library generation.** (a) Total intensity I_A (and I_B) on the A (and B) sublattice in blue (and orange) for the single and randomly chosen regime composing the library. (b) Subspace alignment matrix from the adaptively constructed library with the single initial regime in (a) and with $\epsilon_{th} = 0.01$. (c) Subspace alignment matrix from the reduced library after the equivalent regimes are merged with $\gamma_{th} = 0.75$ in (b). (d) Representative total intensity I_A (and I_B) of the A (and B) sublattice in blue (and orange) for the different regimes. The black vertical dotted line in (a) and (d) indicates the starting time from which the bases are generated. The aDMD bases have been generated with $N_w = 25$ from the time series starting at the 1800-th time step.

SUPPLEMENTARY REFERENCE

[1] J. L. Proctor, S. L. Brunton, B. W. Brunton, and J. N. Kutz, Exploiting sparsity and equation-free architectures in complex systems, Eur. Phys. J. Spec. Top. **223**, 2665 (2014).

[2] M. Gavish and D. L. Donoho, The optimal hard threshold for singular values is $4/\sqrt{3}$, IEEE Trans. Inf. Theory **60**, 5040 (2014).

[3] P. J. Schmid, Dynamic mode decomposition of numerical and experimental data, J. Fluid Mech. **656**, 5 (2010).

[4] J. H. Tu, C. W. Rowley, D. M. Luchtenburg, S. L. Brunton, and J. N. Kutz, On dynamic mode decomposition: Theory and applications, J. Comput. Dyn. **1**, 391 (2014).

[5] B. Kramer, P. Grover, P. Boufounos, S. Nabi, and M. Benosman, Sparse sensing and DMD-based identification of flow regimes and bifurcations in complex flows, SIAM J. Appl. Dyn. Syst. **16**, 1164 (2017).

[6] E. J. Candès, *Proc. Int. Congr. Math. Madrid, August 22–30, 2006* (European Mathematical Society Publishing House, 2006) pp. 1433–1452.