



This is the accepted manuscript made available via CHORUS. The article has been published as:

Unsupervised machine learning for detecting mutual independence among eigenstate regimes in interacting quasiperiodic chains

Colin Beveridge, Kathleen Hart, Cassio Rodrigo Cristani, Xiao Li, Enrico Barbierato, and Yi-Ting Hsu

Phys. Rev. B **111**, L140202 — Published 25 April 2025

DOI: [10.1103/PhysRevB.111.L140202](https://doi.org/10.1103/PhysRevB.111.L140202)

Unsupervised machine learning for detecting mutual independence among eigenstate regimes in interacting quasiperiodic chains

Colin Beveridge,¹ Kathleen Hart,¹ Cassio Rodrigo Cristani,² Xiao Li,³ Enrico Barbierato,² and Yi-Ting Hsu¹

¹*Department of Physics, University of Notre Dame, South Bend, IN 46556 USA*

²*Department of Mathematics and Physics, Catholic University of the Sacred Heart, Brescia, Italy*

³*Department of Physics, City University of Hong Kong, Kowloon, HongKong, China*

(Dated: April 4, 2025)

Manybody eigenstates that are neither thermal nor manybody-localized (MBL) were numerically found in certain interacting chains with moderate quasiperiodic potentials. The energy regime consisting of these non-ergodic but extended (NEE) eigenstates has been extensively studied for being a possible manybody mobility edge between the energy-resolved MBL and thermal phases. Recently, the NEE regime was proposed to be a prethermal phenomenon that generally occurs when different operators spread at sizably different timescales. Here, we numerically examine the mutual independence among the NEE, MBL, and thermal regimes in the lens of eigenstate entanglement spectra (ES). Given the complexity and rich information embedded in ES, we develop an unsupervised learning approach that is designed to quantify the mutual independence among general phases. Our method is first demonstrated on an illustrative toy example that uses RGB color data to represent phases, then applied to the ES of an interacting generalized Aubry Andre model from weak to strong potential strength. We find that while the MBL and thermal regimes are mutually independent, the NEE regime is dependent on the former two and smoothly appears as the potential strength decreases. We attribute our numerically finding to the fact that the ES data in the NEE regime exhibits both an MBL-like fast decay and a thermal-like long tail.

Introduction— Eigenstates of an interacting one-dimensional (1D) chain with sufficiently weak random disorders are known to become thermalized even in the absence of a heat bath^{1–3}. As the disorder strength W increases, the system crossovers into a wide prethermal regime⁴ before transitioning into the manybody localized (MBL) phase⁵. This prethermal regime under an intermediate W still thermalizes (although after a longer thermalization time)^{6–10}, but simultaneously exhibits non-thermal behaviors, such as Poisson-like level statistics⁵. In addition to randomly disordered chains, interacting chains with deterministic quasiperiodic potentials were also numerically^{11–21} and experimentally^{22–24} found to exhibit thermal and MBL phases. Moreover, at certain moderate potential strength λ , numerical evidences have revealed an energy regime between the energy-resolved thermal and MBL phases where the manybody eigenstates are non-ergodic but extended. This non-thermal and non-MBL regime was dubbed a non-ergodic extended regime (NEE), and was first observed in a generalized Aubry Andre model called Ganeshan-Pixley-Das Sarma (GPD) model²⁵ in the presence of interaction^{12,14}. More generally, the NEE regime was recently proposed to be a generic prethermal phenomenon in finite-size systems that occurs when different operators spread at sizably different timescales²⁶. The identification of the NEE regime is thus crucial for further understandings of the prethermal physics in finite-size systems accessible by experimental and numerical studies.

Due to the lack of well-defined order parameters for NEE, its range (in energy E and λ) has been loosely identified as a non-thermal and non-MBL regime based on inconsistency among various inequivalent diagnostics, such as entanglement entropy^{12–14,16,27},

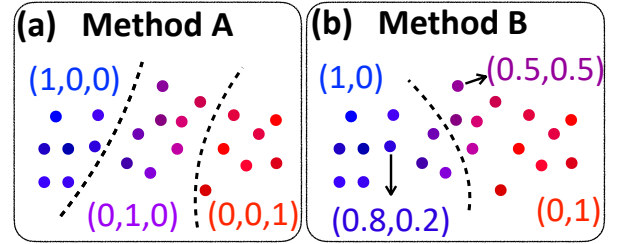


FIG. 1. Illustration for the difference between the two unsupervised clustering methods. The same set of input data represented by color dots are subjected to (a) K-means clustering, and (b) Fuzzy clustering, which forbid and allow fractional memberships, respectively. In this schematic scenario, Method A and B find three and two clusters, respectively. Each data point j is labeled by a membership vector $c_{j,k}$ for cluster k . In (a), the membership vector has $k = 3$ components and $c_{j,k} = 0, 1$. The purple data points forms a third cluster besides the blue and red clusters. In (b), the membership vector has $k = 2$ components and $0 < c_{j,k} < 1$. The purple data points are distributed into the blue or red cluster, depending on different degrees of fractional memberships.

dynamics^{13,16,26,27}, and fluctuations in observables within a small window^{12,14}. However, it remains puzzling how to precisely determine the number of eigenstate ‘phases’ (defined in finite-size systems) between thermal and MBL phases using multiple diagnostics. This is because different diagnostics generally exhibit different MBL-to-thermal transition energies, creating multiple energy regimes that do not behave consistently among all diagnostics. Therefore, some of us proposed in Ref. 17 to determine the number of eigenstate phases by identifying the number of patterns in a single informationally

rich quantity, the entanglement spectra (ES) of eigenstates. Given the complexity of ES, the determination was performed using a ‘self-supervised’ machine learning (ML) approach, which does not require prior knowledge about the number of ES patterns. This approach found three different regimes labeled by ES patterns in a GPD model with a weak to moderate potential strength λ , suggesting a phase diagram of energy-resolved MBL, NEE, and thermal phases.

How independent the NEE regime is of the more well-defined thermal and MBL phases, however, remains elusive. While prior works have shown that NEE exhibits a third type of distinct behavior that is neither MBL nor thermal^{17,27}, such behavior could also be a consequence of averaging over a mixture of MBL-like or thermal-like eigenstates.

In this work, we propose a novel numerical approach that determines the number of eigenstate phases and characterizes how independent these phases are by combining two *unsupervised* learning algorithms^{28,29}:

- Method A: K-Means clustering
- Method B: Fuzzy clustering.

As shown schematically in Fig. 1, both methods self-consistently distribute the input data into groups by minimizing the ‘intra-group distance’ and maximizing the ‘inter-group distance’. However, while Method A²⁸ assigns each data point to a definite cluster, Method B²⁹ allows *fractional memberships*. When applied together to determine a phase diagram of interest, Method A determines the qualitatively different regimes in some parameter space along with their phase boundaries, whereas Method B further determines if a phase is independent or dependent of the other phases. See [Supplementary Materials \(SM\) section III](#) for a brief review for these two well-established numerical methods.

In the following, as a benchmark, we will first examine the energy-resolved phase diagram of an interacting GPD model under different potential strengths λ using the well-developed self-supervised ML method¹⁷ (see Fig. 2a and b). Next, before investigating the GPD model using the unsupervised clustering approach, we will demonstrate our approach on an illustrative toy example that uses RGB color vectors to represent phases (see Fig. 3). In particular, in the blue-purple-red case, we find that Method A sees three clusters but Method B sees only two clusters since purple is a mixture of blue and red. Finally, we will apply our unsupervised approach to the interacting GPD model to examine the phase diagram. With ES being the input datasets, we find that each ES from the NEE regime exhibits features from both the MBL and thermal regimes. Benchmarking with the self-supervised phase diagram, our unsupervised results provide information about the nature of the NEE regime in an interacting quasiperiodic system.

Model— Our goal is to understand the manybody eigenstate phases in quasiperiodic chains with single-

particle mobility edges, under a short-range density-density interaction. As a case study, we consider a generalized Aubry-Andre model $H = H_0 + H_{\text{int}}$ called the GPD model, where the kinetic and interaction terms are given by

$$H_0 = \sum_{j=1}^L \left(-t(c_j^\dagger c_{j+1} + H.c.) + 2\lambda \frac{\cos(2\pi qj + \phi)}{1 - \alpha \cos(2\pi qj + \phi)} n_j \right)$$

$$H_{\text{int}} = V \sum_{j=1}^L n_{j+1} n_j, \quad (1)$$

respectively. Here, $n_j = c_j^\dagger c_j$ is the fermionic density operator at site j , t is the nearest-neighbor hopping strength, and V is the nearest-neighbor interaction strength. The quasiperiodic potential described by the second term in H_0 has a strength of 2λ , an irrational wave number $q = 2/(1 + \sqrt{5})$, and a chosen global phase $\phi \in [0, 2\pi)$. In the rest of this article, we express all energies in the unit of t and set the dimensionless parameter $\alpha = -0.8$ to obtain a substantial energy range for the NEE regime. Moreover, the ES of the manybody eigenstates of H are obtained using Lanczos method at a system size $L = 30$ with $1/6$ filling. Specifically, we cut the full spectrum into 125 bins, evenly distributed between the maximum and minimum eigenvalues of the given spectrum. We then sample 5 eigenstates per bin per ϕ , where ϕ takes 40 equally spaced values between 0 and 2π . We choose not to include all eigenstates in the datasets since a full diagonalization is memory prohibitive. At an intermediate potential strength $\lambda = 0.3 \equiv \lambda^*$, it was reported in Ref. 17 that the interacting manybody spectrum found by the self-supervised ML method contains three energy regimes in which the eigenstates show different ES patterns (see Fig. 2b).

Self-supervised phase diagram— As a benchmark for the unsupervised approach, we now apply the well-developed self-supervised ML method¹⁷ to examine how the three eigenstate phases at $\lambda = \lambda^*$ evolve with the potential strength λ . First, we obtain the three-phase network classifier that distinguishes the MBL, thermal, and NEE ES patterns at $\lambda = \lambda^*$ with a nearly perfect confidence level by reproducing the phase diagram reported in Ref. 17 (along the white dashed line at $\lambda = \lambda^*$ in Fig. 2b). Such a three-phase classifier is a feed-forward artificial neural network, where the input layer takes the ES of a manybody eigenstate and the output layer contains three neurons. The three output neurons characterize the probabilities p_i of the input ES being in phase $i = \text{MBL, thermal, NEE}$, with the normalization condition $\sum_i p_i = 1$ (see Fig. 2a). Note that without assuming there are three phases at $\lambda = \lambda^*$, we have checked but find low confidence levels for the two-phase and four-phase scenarios, where there are only MBL and thermal phases, and where there is an additional fourth phase, respectively. We therefore stick to the three-phase classifier trained at $\lambda = \lambda^*$. By feeding in the ES data from the full manybody spectra at different potential strengths

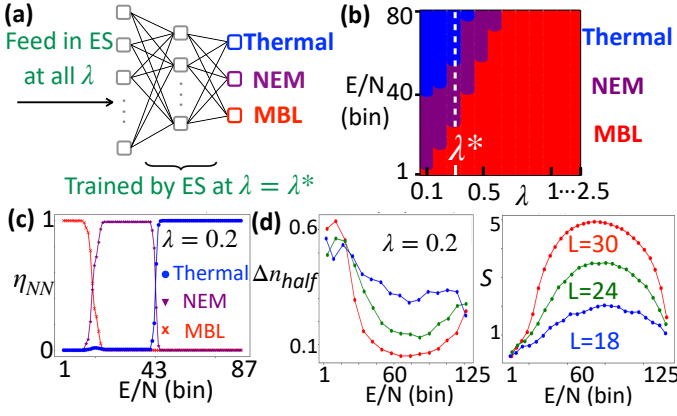


FIG. 2. (a) The schematics and (b) the resulting GPD model phase diagram of the self-supervised network approach¹⁷. The phases at potential strength $\lambda^* = 0.3$ along the white dashed line is consistent with that reported in Ref.¹⁷. (c) The confidence level $\eta_{NN} \in [0, 1]$ of the neural network in (a) at potential strength $\lambda = 0.2$, where substantial energy regimes for all three types of eigenstates are identified with nearly perfect confidence $\eta_{NN} \sim 1$. (d) The variance in the half-chain density Δn_{half} and the entanglement entropy S as a function of energy per electron E/N at different system sizes L . The calculated η_{NN} , Δn_{half} , and S for other potential strengths λ are shown in SM section I.

$\lambda \neq \lambda^*$, we obtain the phase diagram in Fig. 2b. Physically speaking, this phase diagram is generated by asking the network classifier how well a given ES from energy E at a general potential strength $\lambda \neq \lambda^*$ resembles the ES from the MBL, NEE, and thermal regimes at $\lambda = \lambda^*$.

In Fig. 2b, for a wide potential strength around λ^* , the low, intermediate, and mid-spectra³⁰ regimes exhibit MBL, NEE, and thermal-like ES behaviors, respectively. As λ decreases from λ^* , the three-phase scenario smoothly transitions towards a two-phase scenario that consists of predominantly NEE and thermal regimes. We attribute the tiny MBL-like regime on the spectrum edge at $\lambda = 0.1$ to the finite size effect. In contrast, as λ increases from λ^* , the three-phase scenario smoothly transitions into a two-phase scenario with only NEE and MBL regimes, then eventually into a one-phase scenario with only the MBL regime.

Note that instead of existing only right at λ^* , where the network classifier is trained, NEE extends to a finite range in λ with a nearly perfect confidence (Fig. 2c) and smoothly disappears. This shows that the NEE regime at λ^* is unlikely an artifact from the network convergence at λ^* . We further verify in Fig. 2d the extended and non-ergodic properties of the NEE regime at $\lambda \neq \lambda^*$ using the energy-resolved entanglement entropy $S(E)$ and the variance in half-chain density $n_{\text{half}}(E)$, respectively, consistent with previous findings¹². Specifically, we find that the eigenstates in the NEE regime exhibit a volume-law $S(E)$ and large variances of $n_{\text{half}}(E)$ within small energy windows. These two properties show that these metallic eigenstates are also non-ergodic. What remains elusive

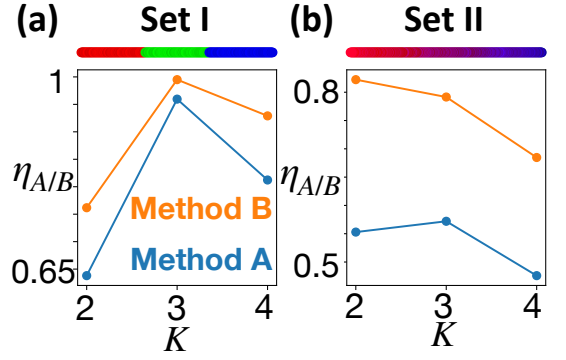


FIG. 3. Clustering results for the toy example that uses RGB color vectors as the input datasets. The confidence indicators $\eta_{A/B}$ obtained by Method A and B are shown for datasets (a) Set I and (b) Set II. The corresponding colors of the data points in each dataset are shown in the color bar. The indicators $\eta_{A/B}$ are obtained under the assumption of different cluster numbers $K = 2, 3, 4$. The K with the largest $\eta_{A(B)}$ is the cluster number that Method A (B) deems most probable.

from this self-supervised ML study is whether the NEE regime can be viewed as a wide transition regime between MBL and thermal phases, due to the nature of the transition or finite-size effects, or as a distinct phase with ES features fully independent of the other two.

Unsupervised clustering approach— We now introduce our clustering approach using an illustrative toy example, where we take the RGB color codes as the input datasets. Specifically, the RGB color code represents each color using a vector (r, g, b) , where r, g, b are real numbers between 0 and 1. We construct two datasets Set I and Set II: Set I consists of three groups of linearly independent colors, red, green, and blue, represented by $(r_1, 0, 0)$, $(0, g_1, 0)$, and $(0, 0, b_1)$, respectively. The data are generated by assigning $1 - \epsilon < r_1, g_1, b_1 < 1$ with a slightly varying small real number ϵ so that the data points visually appear as red, green, or blue with tiny variations. Set II consists of red, blue, and a wide range of purple, represented respectively by $(r_2, 0, 0)$, $(0, 0, b_2)$, and $(\tilde{r}_2, 0, \tilde{b}_2)$. The red and blue data points are generated in the same way as in Set I: $1 - \epsilon < r_2, b_2 < 1$, whereas the purple data points are generated by assigning $\tilde{r}_2$ and $\tilde{b}_2$ to be random values ranged from 0 to 1 (see SM section II). The two sets of colors are visually shown as color bars in Fig. 3a and b.

Next, we employ the two clustering methods Method A²⁸ and Method B²⁹ to determine the number of colors (‘phases’) in datasets Set I and Set II, and explain how the clustering results can identify the difference in nature of Set I and Set II. Clustering is an unsupervised learning method that separates input data into different clusters according to the “distance” among data points. Conventionally, the cluster number K is treated as a chosen hyperparameter, along with the initialization of cluster centers and the convergence tolerance. However, in our case, the cluster number K is the unknown information

that we want to identify. To this end, we first perform one clustering procedure for each possible K , then identify the most probable K as the cluster number with the highest ‘confidence indicator’ $\eta(K)$ for the convergence of the procedure. The precise definitions of the confidence indicators $\eta_{A/B}(K)$ for Method A and B are given in [SM section V](#).

The task is to calculate the confidence indicators $\eta_{A/B}(K)$ using Method A and B for the two datasets Set I and II. Method A is the Silhouette method in K-Means clustering²⁸. Given an assumed K , K-Means algorithm distributes all data points into K distinct clusters by iteratively updating cluster centers to simultaneously minimize and maximize the intra- and inter-cluster sum of squared distances, respectively. We then correspondingly update the integer membership degree $c_{j,k} = 0, 1$ of each data point j to each cluster $k = 1, \dots, K$. We repeat this procedure assuming different cluster numbers, $K = 2, 3, 4$, where the initial cluster centers are chosen randomly. At the end of each procedure, $c_{j,k}$ for all j and k will converge to either 0 or 1, and the convergence level will be quantified using Silhouette coefficient $\eta_A(K) \in [-1, 1]$. We then determine the most probable cluster number as the K that maximizes the average silhouette coefficient $\eta_A(K)$.

Method B is C-Fuzzy clustering²⁹, which is a K-Means-like algorithm that further accommodates noisy datasets, where data points might not exclusively belong to a single cluster. In practice, C-Fuzzy clustering assigns each data point j with a fractional membership degree across multiple clusters. In contrast to Method A, the membership degree $c_{j,k}$ is now a real number between 0 and 1, calculated by a membership function that depends on the distance between data point j and each cluster center k . The more a data point belongs to a cluster, the more influence it has on determining the cluster center in the convergence process. The algorithm iterates the steps of assigning membership degrees and updating the cluster centers until the cluster centers converge. Similar to Method A, in Method B we repeatedly perform the procedure assuming different cluster numbers $K = 2, 3, 4$. The convergence level of each procedure is labeled by a confidence indicator called fuzzy partition coefficient (FPC) index $\eta_B(K)$. The most probable cluster number is then determined by the K that maximizes $\eta_B(K)$.

The convergence indicators $\eta_A(K)$ and $\eta_B(K)$ for datasets Set I and II are shown in Fig. 3. For the red-green-blue dataset Set I, both $\eta_A(K)$ and $\eta_B(K)$ peak at $K = 3$ (see Fig. 3a). This shows that the clustering algorithm robustly finds three independent clusters regardless of whether a fractional membership is allowed, as expected for three groups of colors represented by orthogonal vectors. In sharp contrast, for the red-purple-blue dataset Set II, although $\eta_A(K)$ still peaks at $K = 3$, $\eta_B(K)$ now peaks at $K = 2$ (see Fig. 3b).

This discrepancy can be intuitively understood as follows. When a fractional membership is allowed, the purple points, which are linear combinations of blue and

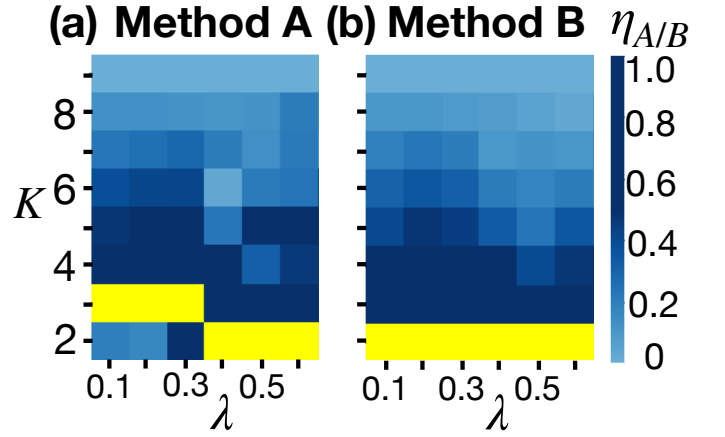


FIG. 4. Clustering results for the GPD model. The confidence indicators $\eta_{A/B}(K, \lambda)$ obtained by (a) Method A and (b) Method B are shown in color maps for different possible cluster numbers K and potential strengths λ . At each λ , the most probable cluster number $K_{A/B}(\lambda)$ is identified by the largest indicator $\eta_{A/B}$, which is colored in yellow. Explicit values of $\eta_{A/B}(K, \lambda)$ are shown in [SM section IV](#).

red, do not form a distinct cluster independent of red and blue clusters. We therefore propose that for a given dataset where the cluster number is unknown, by applying both unsupervised clustering algorithms Method A and Method B, one can determine (1) the number of distinct clusters of data, and (2) how many of which are independent clusters. Specifically, the discrepancy between Method A and B identifies ‘mixed’ clusters containing data points that emulate features from multiple other clusters. Such a mixed regime, which carries drastically fluctuating properties that emulate traits from multiple other regimes, could indicate a wide phase transition or a crossover in a physics system.

Non-ergodic metal regime— We now perform clustering Method A and B to determine the number of eigenstate phases in interacting GPD model in Eq. 1, benchmarking the self-supervised phase diagram we found in Fig. 2b. In particular, by comparing the number of phases found by the two methods, we can inspect how independent the NEE regime is from the MBL and thermal phases.

To construct the input datasets, we employ Lanczos method to diagonalize an $L = 30$ chain, filled with $N = 5$ electrons. For each potential strength λ , we generate the ES of manybody eigenstates sampled throughout the full spectrum E and sort each ES in descending order. Although each ES contains 4944 components, we extract the portion that are informationally rich to minimize the numerical noise and to improve the clustering performance. Specifically, we first omit the nearly zero elements within numerical noise range with a cutoff at 10^{-10} , then perform a dimensional reduction analysis on the remaining 250 ES components. For simplicity, we perform Principal Component Analysis (PCA)³¹ and

keep the two largest PCA eigen-modes. We have checked that the two largest eigen-modes are responsible for at least 80% of the data variation for most of the λ values so that essential information is likely preserved (see SM section III for a brief introduction to PCA and our data).

Method A is performed on this ES datasets to determine the most probable number of clusters $K_A(\lambda)$ at each potential strength λ in the following way. First, we systematically perform one K-Means clustering procedure for each $K = 2, 3, \dots, 9$. The cluster number K corresponds to the number of ES patterns, which also indicates the number of eigenstate phases. For each clustering procedure, we quantify the convergence level using the Silhouette coefficients $\eta_A(K, \lambda) \in [0, 1]$. At each λ , the most probable cluster number $K_A(\lambda)$ is determined by the K with the largest η_A (labeled in yellow in Fig. 4a). We find that $K_A(\lambda)$ in Fig. 4a is consistent with the number of eigenstate phases found in our self-supervised phase diagram in Fig. 2b at every potential strength λ smaller than 0.7, where there are more than one energy-resolved phases. For datasets containing only one cluster, clustering methods are generally not applicable since they tend to show low confidence for all K ³². The consistency between our self-supervised and unsupervised results confirms that there are indeed three ES patterns at weak to moderate potential strengths λ . Nonetheless, the independence among these three ES patterns is yet to be examined by Method B.

Next, we perform Method B on the same input ES dataset to determine the number of clusters $K_B(\lambda)$ at each potential strength λ , where fractional membership is allowed. Similar to Method A, we perform one Fuzzy-C clustering procedure for each possible cluster number $K = 2, 3, \dots, 9$, where the corresponding convergence level is quantified by the FPC index $\eta_B(K, \lambda) \in [0, 1]$. The most probable cluster number $K_B(\lambda)$ is then identified by the highest η_B at each λ . In Fig. 4b, we show the confidence indicator $\eta_B(K, \lambda)$, where the most probable number of clusters $K_B(\lambda)$ is highlighted in yellow. Although $K_B(\lambda)$ agrees well with $K_A(\lambda)$ at stronger potential strengths λ , they unexpectedly differ at weak to moderate λ 's. In contrast to the phase diagrams generated by the self-supervised learning and unsupervised Method A, we find that *there is no λ that exhibits three eigenstate phases* in the lens of Method B with fractional membership. Instead, Method B finds only two phases in that potential range, i.e. $K_B(\lambda) = 2$ whenever $K_A(\lambda) = 3$ (compare Fig. 4a and b).

This discrepancy could be intuitively understood by the fractional membership that is allowed only in Method B but not A. If a data point i has non-zero fractional membership $0 < c_{i,k} < 1$ for multiple clusters k , it implies that the corresponding ES exhibits features from more than one eigenstate phase. Therefore, when Method B sees fewer clusters than Method A, it implies that the clusters obtained from A are not all independent of each other. To verify this interpretation from the actual data, we now focus on the inconsistent regime

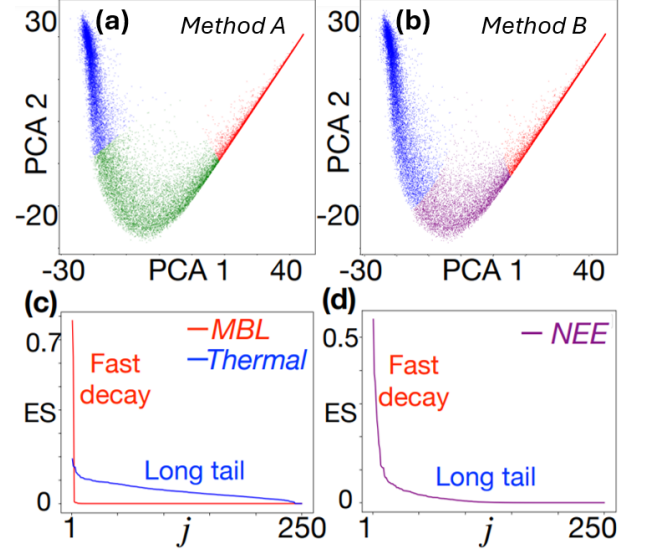


FIG. 5. (a)(b) The ES data of the interacting GPD model at $\lambda = 0.3$, constructed from the eigenstates throughout the spectrum, are displayed in the two-dimensional space spanned by the first two components in the PCA analysis. In (a), each data point is colored according to which of the three clusters found by Method A it belongs to. In (b), the data points with full memberships in the two clusters Method B found are colored in blue and red. The data points with fractional memberships are colored in purple. ES data for example data points chosen from the (c) red, blue, and (d) purple regimes in (b). In (c), the ES data of red and blue points feature a fast decay at small component i and a long tail at large i , respectively, which resemble the features of MBL and thermal eigenstates. In (d), the ES data of a purple point concurrently show both features.

$\lambda \in [0.1, 0.3]$ where Method A finds three phases but Method B finds two phases. Specifically, we examine the distribution of ES data in Method A and B by representing each ES with a point in the two-dimensional space spanned by the first two PCA eigen-components (see Fig. 5a and b).

In Fig. 5a with $K_A = 3$ clusters, we color a data point i in red, blue, or green if it carries a membership vector $\mathbf{c}_i = (c_{i,1}, c_{i,2}, c_{i,3}) = (1, 0, 0)$, $(0, 0, 1)$, or $(0, 1, 0)$, respectively, where $c_{i,k} = 0, 1$ labels the membership in cluster k . We identify the red cluster $k = 1$ as the MBL regime since the ES data therein (see Fig. 5c) are consistent with the typical fast power-law decay in an MBL ES³³. Similarly, we identify the blue cluster $k = 2$ as the thermal regime since the ES data therein have the typical Marchenko-Pastur-like form expected from a thermalized state³⁴. Therefore, the ES in the green cluster $k = 3$ are not MBL- nor thermal-like in the lens of Method A, and we attribute the green cluster to the NEE regime.

In Fig. 5b with $K_B = 2$ clusters, we color data point i in red and blue if it carries a full membership $\mathbf{c}_i = (c_{i,1}, c_{i,2}) = (1 - \epsilon, \epsilon)$ and $(\epsilon, 1 - \epsilon)$, respectively, with a threshold $\epsilon < 0.2$. For the rest of the points

with a fractional membership c_i with $\epsilon \geq 0.2$, we color them in purple. Note that the data distribution here in Fig. 5b nearly align with that in Fig. 5a when the green is replaced by the purple, besides points near the cluster boundaries. This shows that the two independent clusters that Method B identifies are the MBL (red) and thermal (blue) regimes. Moreover, the green NEE regime found by Method A mainly consists of and contains all the ES data points with fractional memberships, which implies that the ES data in the NEE regime exhibit certain MBL-like and thermal-like features simultaneously. Specifically, we find that the ES data with fractional membership all exhibit a faster decay at small component i than the Marchenko-Pastur distribution from the thermal-like ES. Moreover, at large component i , they all exhibit a longer tail than the MBL-like ES. Therefore, we attribute the fast decay and the long tail in ES to be the MBL-like and thermal-like features that Method B recognizes, respectively. These two features coexist in all the ES data points in the purple regime at different degrees.

Conclusion and discussion — Our unsupervised clustering method provides a general numerical approach for identifying the mutual independence among nearby phases. In this work, we demonstrate our method on a finite-length GPD chain by identifying the dependence of the NEE regime on energy-resolved MBL and thermal phases. We choose ES as the ML input data since ES contains nearly complete information about the eigen-

states without requiring much human pre-processing or assumptions. Moreover, ES is also computationally affordable when using the Lanczos algorithm for diagonalizing the Hamiltonian. These two advantages make ES an ideal input for unbiased ML analyses. Interesting future directions include applying our unsupervised and self-supervised ML approaches to investigate the independence among many-body eigenstate phases in other quasiperiodic or disordered systems with a single-particle mobility edge, **such as quasiperiodic models with longer range hopping**³⁵, as well as comparing the ML results to other well-known diagnostics, such as multifractal dimensions. Beyond this case study, our method provides a general numerical tool for studying the prethermal phenomena in non-equilibrium systems with either random disorders or quasiperiodic potentials based on ES and other eigenstate properties.

Acknowledgement — Y.-T.H. thanks Yi-Ting Tu and Dinhduy Vu for very helpful discussions. Y.-T.H. acknowledges support from NSF Grant No. DMR-2238748. This work was performed in part at Aspen Center for Physics, which is supported by National Science Foundation grant PHY-2210452. This work was supported in part by the National Science Foundation under Grant No. NSF PHY-1748958. X.L. is supported by the Research Grants Council of Hong Kong (Grants No. CityU 21304720, No. CityU 308 11300421, No. CityU 11304823, and No. C7012-21G), and City University of Hong Kong (Project No. 9610428).

-
- ¹ J. M. Deutsch, “Quantum statistical mechanics in a closed system,” *Phys. Rev. A* **43**, 2046–2049 (1991).
 - ² Mark Srednicki, “Chaos and quantum thermalization,” *Phys. Rev. E* **50**, 888–901 (1994).
 - ³ Marcos Rigol, Vanja Dunjko, and Maxim Olshanii, “Thermalization and its mechanism for generic isolated quantum systems,” *Nature* **452**, 854–858 (2008).
 - ⁴ David M. Long, Philip J. D. Crowley, Vedika Khemani, and Anushya Chandran, “Phenomenology of the prethermal many-body localized regime,” *Phys. Rev. Lett.* **131**, 106301 (2023).
 - ⁵ Alan Morningstar, Luis Colmenarez, Vedika Khemani, David J. Luitz, and David A. Huse, “Avalanches and many-body resonances in many-body localized systems,” *Phys. Rev. B* **105**, 174205 (2022).
 - ⁶ Jan Šuntajs, Janez Bonča, Tomaž Prosen, and Lev Vidmar, “Quantum chaos challenges many-body localization,” *Phys. Rev. E* **102**, 062144 (2020).
 - ⁷ Dries Sels and Anatoli Polkovnikov, “Dynamical obstruction to localization in a disordered spin chain,” *Phys. Rev. E* **104**, 054105 (2021).
 - ⁸ Dries Sels and Anatoli Polkovnikov, “Thermalization of dilute impurities in one-dimensional spin chains,” *Phys. Rev. X* **13**, 011041 (2023).
 - ⁹ Piotr Sierant, Dominique Delande, and Jakub Zakrzewski, “Thouless time analysis of anderson and many-body localization transitions,” *Phys. Rev. Lett.* **124**, 186601 (2020).
 - ¹⁰ Piotr Sierant, Dominique Delande, and Jakub Zakrzewski, “Thouless time analysis of anderson and many-body localization transitions,” *Phys. Rev. Lett.* **124**, 186601 (2020).
 - ¹¹ Shankar Iyer, Vadim Oganesyan, Gil Refael, and David A. Huse, “Many-body localization in a quasiperiodic system,” *Phys. Rev. B* **87**, 134202 (2013).
 - ¹² Xiaopeng Li, Sriram Ganeshan, J. H. Pixley, and S. Das Sarma, “Many-body localization and quantum nonergodicity in a model with a single-particle mobility edge,” *Phys. Rev. Lett.* **115**, 186601 (2015).
 - ¹³ Ranjan Modak and Subroto Mukerjee, “Many-body localization in the presence of a single-particle mobility edge,” *Phys. Rev. Lett.* **115**, 230401 (2015).
 - ¹⁴ Xiaopeng Li, J. H. Pixley, Dong-Ling Deng, Sriram Ganeshan, and S. Das Sarma, “Quantum nonergodicity and fermion localization in a system with a single-particle mobility edge,” *Phys. Rev. B* **93**, 184204 (2016).
 - ¹⁵ Vedika Khemani, D. N. Sheng, and David A. Huse, “Two universality classes for the many-body localization transition,” *Phys. Rev. Lett.* **119**, 075702 (2017).
 - ¹⁶ Sabyasachi Nag and Arti Garg, “Many-body mobility edges in a one-dimensional system of interacting fermions,” *Phys. Rev. B* **96**, 060203 (2017).
 - ¹⁷ Yi-Ting Hsu, Xiao Li, Dong-Ling Deng, and S. Das Sarma, “Machine learning many-body localization: Search for the elusive nonergodic metal,” *Phys. Rev. Lett.* **121**, 245701 (2018).
 - ¹⁸ Shenglong Xu, Xiao Li, Yi-Ting Hsu, Brian Swingle, and S. Das Sarma, “Butterfly effect in interacting aubry-andre

- model: Thermalization, slow scrambling, and many-body localization,” *Phys. Rev. Res.* **1**, 032039 (2019).
- ¹⁹ DinhDuy Vu, Ke Huang, Xiao Li, and S. Das Sarma, “Fermionic many-body localization for random and quasiperiodic systems in the presence of short- and long-range interactions,” *Phys. Rev. Lett.* **128**, 146601 (2022).
 - ²⁰ Ke Huang, DinhDuy Vu, Xiao Li, and S. Das Sarma, “Incommensurate many-body localization in the presence of long-range hopping and single-particle mobility edge,” *Phys. Rev. B* **107**, 035129 (2023).
 - ²¹ Yi-Ting Tu, DinhDuy Vu, and S. Das Sarma, “Avalanche stability transition in interacting quasiperiodic systems,” *Phys. Rev. B* **107**, 014203 (2023).
 - ²² Michael Schreiber, Sean S. Hodgman, Pranjal Bordia, Henrik P. Lüschen, Mark H. Fischer, Ronen Vosk, Ehud Altman, Ulrich Schneider, and Immanuel Bloch, “Observation of many-body localization of interacting fermions in a quasirandom optical lattice,” *Science* **349**, 842–845 (2015), <https://www.science.org/doi/pdf/10.1126/science.aaa7432>.
 - ²³ Henrik P. Lüschen, Pranjal Bordia, Sebastian Scherg, Fabien Alet, Ehud Altman, Ulrich Schneider, and Immanuel Bloch, “Observation of slow dynamics near the many-body localization transition in one-dimensional quasiperiodic systems,” *Phys. Rev. Lett.* **119**, 260401 (2017).
 - ²⁴ Thomas Kohlert, Sebastian Scherg, Xiao Li, Henrik P. Lüschen, Sankar Das Sarma, Immanuel Bloch, and Monika Aidelsburger, “Observation of many-body localization in a one-dimensional system with a single-particle mobility edge,” *Phys. Rev. Lett.* **122**, 170403 (2019).
 - ²⁵ Sriram Ganeshan, J. H. Pixley, and S. Das Sarma, “Nearest neighbor tight binding models with an exact mobility edge in one dimension,” *Phys. Rev. Lett.* **114**, 146601 (2015).
 - ²⁶ Yi-Ting Tu, David M. Long, and Sankar Das Sarma, “Interacting quasiperiodic spin chains in the prethermal regime,” *arXiv e-prints*, arXiv:2405.01622 (2024), arXiv:2405.01622 [cond-mat.dis-nn].
 - ²⁷ Yi-Ting Tu, DinhDuy Vu, and Sankar Das Sarma, “Localization spectrum of a bath-coupled generalized aubry-andré model in the presence of interactions,” *Phys. Rev. B* **108**, 064313 (2023).
 - ²⁸ Stuart Lloyd, “Least squares quantization in pcm,” *IEEE transactions on information theory* **28**, 129–137 (1982).
 - ²⁹ James C. Bezdek, *Pattern recognition with fuzzy objective function algorithms*, 2nd ed., Advanced applications in pattern recognition (Plenum Press, New York, 1981).
 - ³⁰ Eigenstates of the top band edge of the spectrum resemble those from the bottom band edge of the spectrum.
 - ³¹ Karl Pearson, “Liii. on lines and planes of closest fit to systems of points in space,” *The London, Edinburgh, and Dublin philosophical magazine and journal of science* **2**, 559–572 (1901).
 - ³² Robert Tibshirani, Guenther Walther, and Trevor Hastie, “Estimating the number of clusters in a data set via the gap statistic,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **63**, 411–423 (2001).
 - ³³ Maksym Serbyn, Alexios A. Michailidis, Dmitry A. Abanin, and Z. Papić, “Power-law entanglement spectrum in many-body localized phases,” *Phys. Rev. Lett.* **117**, 160601 (2016).
 - ³⁴ Zhi-Cheng Yang, Claudio Chamon, Alioscia Hamma, and Eduardo R. Mucciolo, “Two-component structure in the entanglement spectrum of highly excited states,” *Phys. Rev. Lett.* **115**, 267206 (2015).
 - ³⁵ Ke Huang, DinhDuy Vu, Xiao Li, and S. Das Sarma, “Incommensurate many-body localization in the presence of long-range hopping and single-particle mobility edge,” *Phys. Rev. B* **107**, 035129 (2023).