Machine-learning classification of two-dimensional vortex configurations

Rama Sharma and Tapio P. Simula ⓞ

Optical Sciences Centre, Swinburne University of Technology, Melbourne, Victoria 3122, Australia

We consider computer-generated configurations of quantized vortices in planar superfluid Bose-Einstein condensates. We show that unsupervised machine-learning technology can successfully be used for classifying such vortex configurations to identify prominent vortex phases of matter. The machine-learning approach could thus be applied for automatically classifying large data sets of vortex configurations obtainable by experiments on two-dimensional quantum turbulence.

I. INTRODUCTION

Statistical mechanics is one of the cornerstones of modern physics [1–3]. The key underlying principle is statistical equivalence of the myriad of microstates that all share common thermodynamic properties such as configuration energy or temperature, and collectively represent the macroscopically observable phenomena. Typically, the macroscopic state variables are given as input parameters and the corresponding microstates are obtained as solutions to the underlying model Hamiltonian. In this work, we are interested in a reverse process where the microstates are given as measurement outcomes which we wish to categorize into distinct macrostates based on their statistical similarity, without any knowledge of the underlying Hamiltonian.

The Onsager model of two-dimensional turbulence [4] is a statistical mechanics description of pointlike vortices where each vortex configuration corresponds to a particular statistical microstate of the fluid. The model is particularly well suited for modeling vortices and their statistical behavior in superfluids where the vorticity of the fluid is quantized. The applicability of the Onsager model has been verified by recent experiments on two-dimensional quantum turbulence (2DQT) in superfluid Bose-Einstein condensates [5,6].

In typical cold-atom experiments the quantitative information of the physical system is extracted from images of the atom density distribution [7]. Each such experimental image corresponds to a single representative microstate of the system's thermodynamic state and if quantized vortices are present in the system, their positions can be read off from such images. This raises the question whether it would be possible to categorize such experimental data using machine-learning protocols into ensembles of statistically equivalent microstates, for instance, in order to detect distinct phases of matter supported by the system.

The remarkable success of artificial neural networks when applied to the problems of image recognition, image classification, and natural language processing has prompted interdisciplinary efforts to investigate how a broader range of scientific problems might benefit from deploying these new tools. This has led to implementations of machine-learning methods, for instance, to identify symmetry-broken phases

in the field of classical statistical physics [8–10], and in some cases neural networks have even been shown to be able to learn an order parameter or other thermodynamical parameters [8,10]. More recently, the machine-learning methodology has found applications in the realm of physics problems such as identifying phase transitions of many-body systems [11–22], topological systems [23–27], and finding quantum enhanced learning algorithms [28–30].

An appropriately designed and trained supervised deep learning procedure has been applied to spin and vortex configurations in the two-dimensional XY model [24] to identify the Kosterlitz-Thouless (KT) transition [31]. However, labeled training sets, which are mandatory for supervised learning, are not always easily attainable. For this reason, the unsupervised machine-learning methods, that do not rely on prior knowledge, may offer significant benefits over supervised learning methods [32,33]. Inspired by the application of the supervised learning to classify the KT transition [11,24], it is natural to ask whether unsupervised neural networks would be capable of identifying a variety of vortex phases of matter and quantum turbulent flow states in two-dimensional Bose-Einstein condensates. Few unsupervised learning techniques have been previously applied to the XY model. The principal component analysis (PCA) method [34] has been performed on spin configurations [35–37] but even when learning with the vorticity field directly, the PCA was found to be unable to identify the transition point corresponding to the vortex-antivortex unbinding [36]. Additionally, most of these previous approaches need the prior information before processing, for example, the previously known number of phases and approximate transition temperature value [17,23]. Hence such aspects can be complicated to implement the generalized idea via the machine-learning technique to detect these phase transitions. In contrast, our work presents a machine-learning approach which employs the bag-of-features function for feature extraction and unsupervised self-organizing-map (SOM) algorithm for classification and is able to detect the ordered and disordered sides of the vortex binding-unbinding transition without prior labeling.

In this paper, we apply an unsupervised machine-learning strategy to the task of identifying vortex phases of matter of

the two-dimensional Onsager point-vortex model. Our main goal is to test whether it is possible for an artificial neural network, trained only on the features extracted from the vortex configurations, to learn distinct vortex phases of matter that are thermodynamically defined in terms of external macrostate properties such as temperature. The rest of this paper is organized as follows. Section II provides the theoretical background on the Onsager point-vortex model of two-dimensional turbulence and outlines the numerical methods employed, such as the Monte Carlo method and machine-learning model. Section III begins by providing benchmark results using an unsupervised machine-learning approach showing successful classification of states of a system of two same-sign vortices. Specifically, we show that the unsupervised learning is able to identify the analytically predicted topological phase boundary of this system [38,39]. We then consider experimentally relevant point-vortex configurations of 20 polarized (single-sign circulation for all vortices) and 40 neutral (equal number of vortices and antivortices) vortex systems and show that the unsupervised machine learning is able to successfully identify both the positive-vortex-temperature Kosterlitz-Thouless transition as well as the negative-vortex-temperature Onsager vortex-condensation transition. These findings show promise for applying machine-learning models for exploring experimental data sets involving vortex configurations. We close the paper in Sec. IV by summarizing our findings with concluding remarks.

II. THEORY

Consider a Navier-Stokes equation,

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla)\mathbf{v} = -\frac{\nabla p}{\rho} + \nu \nabla^2 \mathbf{v}, \tag{1}$$

which describes the flow of a Newtonian incompressible fluid that satisfies the continuity equation $\partial_t \rho + \nabla \cdot (\rho \mathbf{v}) = 0$. In Eq. (1) $\mathbf{v}(\mathbf{r}, t)$ is the fluid velocity field, ρ is the fluid density, p is the fluid pressure, ∇^2 is the Laplacian operator, and ν is the kinematic viscosity. In rare cases, such as one-dimensional flow and creeping flow, the Navier-Stokes equation for $\mathbf{v}(\mathbf{r}, t)$ can be solved analytically. However, the nonlinearity in turbulent fluids arising due to the convective acceleration of the fluid makes analytical solutions impossible in general. Moreover, an accurate numerical solution of Eq. (1) is difficult to achieve due to the vast range of length scales and number of degrees of freedom involved [40]. Fortuitously, a dramatic simplification can be achieved by considering the fluid's vorticity field, $\omega(\mathbf{r}, t) = \nabla \times \mathbf{v}(\mathbf{r}, t)$, instead of the velocity field. In particular in the case of planar superfluids, the former can in many cases be well approximated by only a small number of point vortices.

A. Point-vortex dynamics

One of the conceptual benefits of two-dimensional fluid dynamics comes from the fact that the flow field is confined to a plane. Consequently, the vortices cannot bend or expand the way they can in three-dimensional flows. The vortices can be modeled as point sources of rotating fluid flow in the limit where the vortices are well separated. In such a case,

the vorticity $\omega(\mathbf{r})$, which is normally smoothly distributed over the fluid, may be obtained by summing over the vortices according to

$$\omega(\mathbf{r}) = \nabla \times \mathbf{v} = \sum_i^{N_v} \Gamma_i \hat{\mathbf{e}}_z \delta(\mathbf{r} - \mathbf{r}_i), \tag{2}$$

where the point vortices with circulations Γ_i are located at positions $\mathbf{r}_i$ for $i \in \{1, 2, \ldots, N_v\}$ with N_v the total number of vortices. The velocity field of a fluid flow around a single vortex at location r is

$$\mathbf{v}_i(\mathbf{r}) = \frac{\Gamma_i}{2\pi} \frac{1}{|\mathbf{r} - \mathbf{r}_i|} \hat{\theta}_i, \tag{3}$$

where the azimuthal coordinate axes $\hat{\theta}_i$ are centered on the vortex cores. The total fluid velocity field

$$\mathbf{v}(\mathbf{r}) = \sum_i^{N_v} \mathbf{v}_i(\mathbf{r}) + \mathbf{B} \tag{4}$$

due to many vortices is a simple superposition of the individual vortex velocity fields $\mathbf{v}_i$, supplemented by $\mathbf{B}$ which stands for terms that are nonzero in the presence of boundaries. The point-vortex approximation significantly simplifies the modeling of a two-dimensional (2D) fluid, as the velocity of the fluid at each point in space can be mapped by the net flow of all vortices within the fluid. The vortices are positioned in such a way that the superposition of their regular circulating flows better mimics the fluid's complete velocity field [41]. This description has the advantage of allowing the continuous fluid to be replaced by N_v zero-spatial-extent points (vortices) of well-defined locations (x_i, y_i) and circulations Γ_i, which carry the full fluid flow information. The dynamics of the vortices maps onto the dynamics of the fluid particles and each vortex moves with the fluid velocity induced by all other vortices within the fluid at its location. Hence, each vortex's motion is measured by the relative location and strength of all other vortices [42].

The point-vortex approximation is particularly well suited for modeling the dynamics of vortices in superfluid Bose-Einstein condensates (BECs) for which Eq. (2) is accurately satisfied. This leads to the point-vortex model (PVM) for the dynamics of the vortices, which utilizes a set of coupled ordinary differential equations to describe the vortex dynamics by considering the interactions among vortices and related boundary conditions in trapping potential. For most of this work we make the use of a *uniform trap* potential.

Vortex equations of motion

The so-called box potentials have become commonly utilized in cold-atom experiments [5,6,43,44]. Considering vortices in a unit disk geometry, each vortex is accompanied by a single image vortex of charge $\bar{s}_i = -s_i$ positioned outside the fluid boundary at a location $\bar{\mathbf{r}}_i = \mathbf{r}_i R^2 / |\mathbf{r}_i|^2$ [45]. The equations of motion for such point vortices are [4]

$$hs_i \frac{\partial x_i}{\partial t} = \frac{\partial H}{\partial y_i} \quad \text{and} \quad hs_i \frac{\partial y_i}{\partial t} = -\frac{\partial H}{\partial x_i}, \tag{5}$$

where the energy of the vortex configuration is

$$H = \alpha k_B \sum_{i=1}^{N_v} s_i^2 \ln\left(1 - r_i^2\right) - \alpha k_B \sum_{j<i}^{N_v} s_i s_j \ln\left(r_{ij}^2\right)$$
$$+ \alpha k_B \sum_{j<i}^{N_v} s_i s_j \ln\left(1 - 2x_i x_j - 2y_i y_j + r_i^2 r_j^2\right). \quad (6)$$

In Eq. (6), $\alpha = \rho_s \kappa^2 / 4\pi k_B$, where ρ_s is the superfluid density and $\kappa = h/m$ is the quantum circulation with h, m, and k_B denoting the Planck constant, the atom mass, and the Boltzmann constant, respectively. The $r_i^2 = x_i^2 + y_i^2$ with $x_i = \text{Re}(\mathbf{z}_i)$ and $y_i = \text{Im}(\mathbf{z}_i)$ are the Cartesian coordinates of the ith vortex and are expressed in terms of complex numbers $\mathbf{z}_i$ in a system of dimensionless radius $R = 1$ with circulation winding number $s_i = \pm 1$. The first term of Eq. (6) corresponds to the interaction of the vortices with their images. The second term describes the logarithmic long-range two-dimensional Coulomb interaction between the vortices, and last term is the interaction of the real vortices with the images of all other vortices in the system.

B. Monte Carlo thermodynamics

Since we are particularly interested in the statistical properties of two-dimensional vortex configurations, we have implemented a Monte Carlo sampling method following Refs. [46,47]. Briefly, the algorithm takes the initial positions and circulations of N_v vortices and a vortex temperature T_v as input parameters and returns a set of statistically equivalent equilibrium vortex configurations.

To initiate the algorithm we generate random initial positions for the fixed number of N_v vortices. In each step of the algorithm an attempted move of one randomly selected vortex is made and the move is accepted or rejected probabilistically based on a temperature-dependent weight function η [46,47]. In this work a Boltzmann factor $\eta = \exp(-H/k_B T_v)$, where k_B is the Boltzmann constant and T_v is the vortex temperature, is used. The inverse temperature $\beta = 1/(k_B T_v)$ characterizes the equilibrium vortex configurations. The energy H is determined from the point-vortex Hamiltonian [Eq. (6)] of a uniform fluid inside a circular domain of radius R. The core radius of the vortices is set to be $0.008R$. In order to account for the effect of finite core size of real superfluid vortices, a constraint was placed when generating the vortex configurations to prevent vortices from falling too close to each other or to the system boundary. A minimum intervortex separation of twice the vortex core radius, and a minimum vortex-boundary separation of one vortex core radius were implemented. To ensure fair sampling of the vortex configurations, a variety of observables were monitored, and the systems were deemed to have reached equilibrium after 10^5 steps. All Monte Carlo calculations were therefore first run for an initial burn in of 10^5 steps at each temperature. Subsequently, at every temperature, 1000 vortex configurations were sampled uniformly from the total of 10^6 microstates generated.

Clusters, dipoles, and free vortices

To quantitatively detect different vortex configurations we used a vortex classification algorithm [47,48]. The algorithm
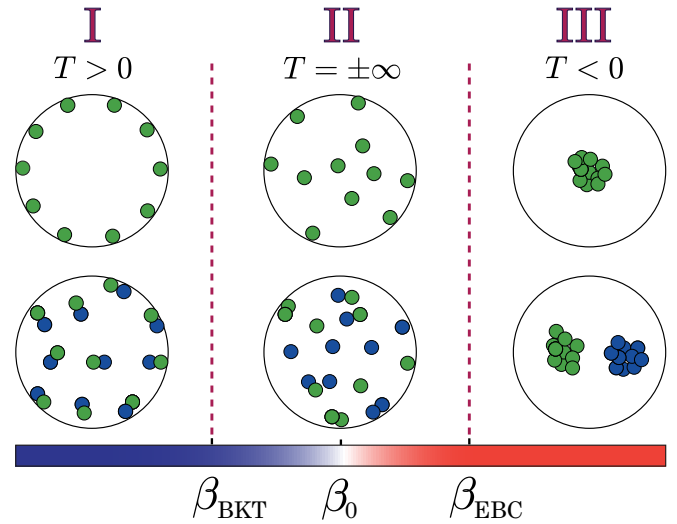

FIG. 1. Illustration of the three point-vortex phases of matter as functions of inverse temperature β. The top row shows configurations of 10 vortices all having the same sign of circulation. The bottom row shows configurations of 20 vortices with equal numbers of clockwise and anticlockwise circulations. The three phases shown are the low-entropy positive-temperature phase I, high-entropy disordered phase II, and low-entropy negative-temperature phase III. The two extremes are separated from the disordered phase by the critical temperatures β_{BKT} and β_{EBC}, which are marked by dashed vertical lines.

classifies vortices based on their spatial configuration by measuring all intervortex distances and assigning the vortices into three categories: same-sign clustered vortices, vortex dipoles, and free vortices.

C. Vortex fluid phase transitions

The point-vortex statistical thermodynamics sets a framework for understanding the vortex fluid behavior within the point-vortex model. The model has two prominent transition temperatures corresponding to the positive-temperature vortex dipole pair formation or breaking, and negative-temperature same-sign vortex cluster formation or breaking. A schematic summarizing the vortex phases is shown in Fig. 1. For a two-dimensional neutral Coulomb gas with equal number of vortices and antivortices without core structure (true point vortices) [49,50], the vortex dipole pair collapse transition occurs at a critical temperature $\beta_d = 4\pi / \rho_s \Gamma^2$, where ρ_s is the fluid density, $\Gamma = h/m$ the circulation, with h and m the Planck constant and atom mass, respectively. By accounting for the nonzero vortex core size, this transition temperature shifts [51], towards the Berezinskii-Kosterlitz-Thouless (BKT) critical temperature $\beta_{\text{BKT}} = 2\beta_d$ [31,52,53].

The negative temperature transition, referred to as supercondensation or Einstein-Bose condensation (EBC) [47,54,55], corresponds to the condensation of like-sign Onsager vortex clusters. The critical temperature of this transition is $\beta_{\text{EBC}} = -4\beta_d / N_v$ [51,55] for neutral vortex configurations and $\beta_{\text{EBC}} = -2\beta_d / N_v$ for a system of single-sign vortices.

These temperatures serve as significant reference points characterizing the vortex configurations illustrated in Fig. 1. In between the two low-entropy ordered phases the vortices are seemingly randomly distributed with their configurational entropy being maximized at $\beta = 0$. The top and bottom rows in Fig. 1 show representative vortex configurations for the cases of $N_v = 10$ all clockwise (single-sign) circulation and $N_v = 20$ with equal number of clockwise and anticlockwise (neutral) circulations, respectively.

Our main goal in this paper is to apply machine-learning technology to investigate if a simple unsupervised learning with neural networks is capable of correctly identifying the three significant vortex phases of matter summarized in Fig. 1.

D. Machine-learning classification of vortex configurations

Machine-learning algorithms are used for categorization of data sets based on occurrences of common sets of features. Possible features include continuous, binary, and categorical. Supervised machine learning requires additional knowledge (a training set) to supplement the data whereas in an unsupervised machine-learning approach the data set is provided for categorization as is without additional supporting information [56].

The SOM [57–59] is a case of unsupervised artificial neural networks (ANNs) successfully applied in areas of data clustering and complex data visualization, and for image processing. Here we apply a SOM algorithm to classify vortex configurations. For this purpose we employ a machine-learning framework implemented in MATLAB. In unsupervised learning, test images containing vortex positions are provided to the classifier and the self-organizing maps cluster the data based on the detected similarity and topology [57–59]. First, the test images that are used for training the classifier need to be preprocessed. For this purpose, we use the bag-of-features (BoF) model to construct the feature vector (a histogram of discrete features detected in an image). The feature vectors are then used for training the SOM classifier model. These machine-learning models used in this work are briefly described below.

1. Bag-of-features model

Detecting robust image features forms the basis for accurate object recognition [60]. The features of a digital image such as shapes, color, texture, and the locations of these local features inside the image are properties that allow the image to be differentiated from other images in the database. Representing an image through its pixel values results in a very high-dimensional matrix, which is not appropriate for image classification or recognition [61]. Therefore, we instead extract the local features of digital images containing vortex positions to classify them in categories.

The extracted features are influential when forming a signature, such as a bag-of-visual-words depiction, for an image [62]. Bag of visual words is a simplified approach for extracting image content for machine-learning classification [63]. It represents the images with orderless collections of local image features. Here we use the BoF model in MATLAB to construct the visual words for a given image. The concept of BoF is analogous to the *bag of words* (BoW) [64,65] used for representing a text document. The BoF model applies a similar methodology, but instead of *words* it uses the image features for analyzing the image. The key goal is to create a visual vocabulary known as codebook, where the most common and strong image features are coded as codewords or visual phrases. The image representation as a BoF is a histogram generated by a simple image codeword or visual word occurrences. This model treats each image as a visual word frequency histogram based on a visual vocabulary that measures the spatial characteristics of all images in the database.

The complete bag-of-visual-words process is accomplished by using an in-built MATLAB function, which operates in a step-by-step approach. First, it extracts the strongest image features using a speeded-up robust features (SURF) algorithm [66,67]. Second, the model constructs the visual vocabulary by considering 80% of the strongest features and clustering them into visual words using a k-means clustering algorithm [67,68]. The k-means clustering follows a heuristic approach to construct initial clustering by selecting random k-centroids from the data set, which represents the SURF features as points [69,70]. For each data point the clustering algorithm calculates the distance from all centroids and then assign its membership to the nearest k-centroid iteratively. After each iteration, the recalculation of the new k-centroid is done by averaging all data points that are assigned to the clusters and the process is repeated until convergence [67]. In this way the final centroids become visual words or discrete image features comprising the visual vocabulary [67], for the given image set. These discrete image features are used for training the (SOM) classifier model to reveal the possible classifications.

2. Self-organizing-map algorithm

A Kohonen SOM [57] is a popular unsupervised artificial neural network which is used to group similar patterns such as feature vectors or data items together [71]. It projects a high-dimensional input data onto a low-dimensional array of nodes (neurons) [72]. This mapping retains the topological relationships between the data domains. Consequently, the image of the data space tends to manifest clustering of input information and their relationships on the map. Initially a random weight is assigned for each neuron and is placed in the feature space containing the input vectors of the testing images. Then one of the input vectors is randomly selected. For each input vector, its Euclidean distance to every weight vector is calculated, and the neuron with the closest matching weight vector is moved towards the input vector in the feature space. Also the neighboring neurons within a certain radius are dragged toward the input vector [58,59]. This "nearest" neuron is called the best matching unit (BMU) or the winning neuron. The neurons' positions are updated in each iteration and the process is repeated for each input data and over all iterations. The magnitude of these displacements decreases with the distance from the BMU and as the iteration proceeds. After considering each neuron and all iterations, eventually the entire neural network tends to approximate the input vector distribution. Finally, the similar data are clustered together in one area and the dissimilar data are grouped in a separate area.

For analyzing the classification of vortices we deployed the bag-of-features function for extracting image features. We used a grid method for picking the key point location in the feature extraction mechanism and used a block width to specify the scale of the feature. We employed the grid method to optimize memory requirements and computational time [73], while maintaining the accuracy of classifier. However, the grid size selection is a crucial step as a scattered grid that corresponds to a low number of image features can lead to loss of the key information, and on the other hand a dense grid that corresponds to excessively many features becomes computationally demanding and may result in irrelevant information. For classifying two same-sign vortices (data set of 4000 images) we use the options GridStep = [8 8], BlockWidth = [32 64 96 128], and vocabulary size = 500 in MATLAB. For analyzing the classification of more vortices for polarized and neutral systems (larger data sets of $\approx$30 000 images) we customize the bag-of-features model in order to reduce memory consumption while maintaining the desired accuracy.

Specifically, in the context of this work for larger data sets ($\approx$30 000 images of size 291×291 pixels per image), the parameters in the deployed `bagOfFeatures` function call were VocabularySize = 250, StrongestFeatures = 0.8, PointSelection = Grid, GridStep = [20 20], and BlockWidth = [32 64 96 128]. We ran the SOM clustering algorithm for 800 training iterations and clustered the output typically into four different classes by setting the dimension of the network (number of neurons) accordingly.

3. Silhouette analysis for clustering accuracy

The clustering performance may be evaluated using a range of methods such as the Calisnki-Harabasz coefficient, Ball-Hall index, Hartigan index, Dunn index, and silhouette coefficient [74]. Here we use the most common silhouette coefficient method, which is a measure of how similar a data item is to its corresponding cluster, as compared to the different clusters [75]. The silhouette coefficient for data point k is $S_k = (b_k - a_k)/\max(a_k, b_k)$, where a_k is the average distance from the point k to all other points within the same cluster as k, and b_k is the minimum average distance from the point k to points in a different cluster as illustrated in the Supplemental Material [76]. The silhouette coefficient S for the entire data set is the average of the silhouette coefficients over the total number of data points. That is, if there are P data points or objects in total in the system then the silhouette coefficient of the full data set is $S = (\sum_{k=1}^{P} S_k)/P$.

The value of S ranges from -1 to $+1$. A high positive value indicates that the sample is similar to its own cluster and distinct from the other clusters. By contrast, a low or negative S value for many data points indicates poor cluster compliance.

In order to test our SOM clustering accuracy and to validate the choice of four clusters we performed the k-means cluster evaluation using the silhouette criterion where the cluster number with maximum silhouette coefficient is decided as the optimum cluster number. Although the k-means method is a popular, conceptually simple, and computationally efficient clustering algorithm, it has a feature that it may or may not

give the same outcome for repeated runs on the same input, since it is initiated with random assignments. Here we perform k-means cluster evaluation for ten different numerical experiments using one to ten clusters. We then average the silhouette value over the total number of experiments for each cluster number to determine the optimum number of clusters.

III. RESULTS

To investigate the feasibility of using machine learning to classify point-vortex configurations, we begin by considering the minimal system of two same-sign vortices. Using the obtained results as encouragement we then move on to consider separately larger polarized and neutral vortex systems.

A. Two same-sign vortices

To study the vortex dynamics of two corotating vortices in an oblate harmonically trapped Bose-Einstein condensate where the motion of the vortices is restricted within the Thomas-Fermi radius, $R > |z_i|$ of the BEC [38,77,78], we consider the equation of motion

$$-i\dot{\mathbf{z}}_i = R^2 \Omega_0 \frac{s_i \mathbf{z}_i}{R^2 - |\mathbf{z}_i|^2} + R^2 \Omega_{\text{int}} \sum_{j \neq i}^{N} s_j \frac{\mathbf{z}_i - \mathbf{z}_j}{|\mathbf{z}_i - \mathbf{z}_j|^2}, \quad (7)$$

where the two phenomenological parameters Ω_0 and Ω_{int} correspond to the orbital angular frequency of a unit-strength single vortex which orbits around the trap center, and the angular frequency that determines the intervortex interaction strength [38,79,80], respectively. To demonstrate this vortex dynamics we first consider a system of two vortices, $N_v = 2$, of unit circulation. The length and time are measured in the units of the system radius R and inverse angular frequency Ω_0^{-1}, respectively. All possible two-vortex configurations in this system are characterized by three observables: the vortex angular momentum $L = r_1^2 + r_2^2$ [41], not to be confused with the angular momentum of the fluid; the angle $\phi = \tan^{-1}(r_2/r_1)$; and the azimuthal angle $\theta_{21} = \theta_2 - \theta_1$, subtended by the position vectors of the two vortices. In our work we choose the value of the phenomenological parameters $\Omega_{\text{int}} = 0.01\omega_0$ and $\Omega_0 = 1\omega_0$, where ω_0 is an arbitrary frequency reference.

To compute the point-vortex dynamics in such a harmonic trap, Eq. (7) was integrated using MATLAB function `ode113` with relative error tolerance and absolute error tolerance both set to 10^{-13}. Each simulation was initialized by setting the initial positions and circulations of the vortices. The accuracy of the integration was confirmed by monitoring the values of the conserved quantities H and L.

Vortex dynamics of two same-sign vortices in Bose-Einstein condensates has been observed experimentally [38,81]. In a system with a circular boundary each possible two-vortex configuration is either of overlapping or nonoverlapping type [38,39].

Figures 2(a) and 2(d) show two such configurations. Visually, configurations in Figs. 2(a) and 2(d) look similar and their key difference is revealed by the dynamics shown in respective Figs. 2(b) and 2(e). In Fig. 2(b) the vortex paths never cross and the configuration in Fig. 2(a) corresponds to
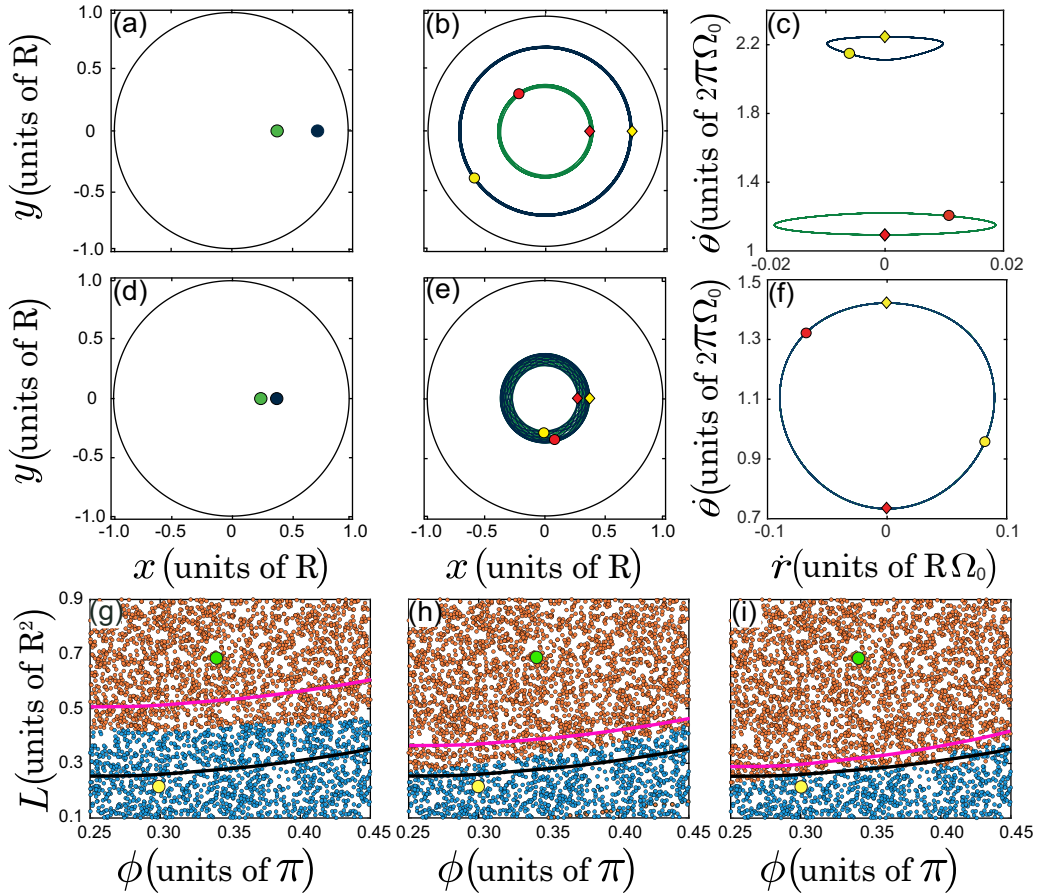
FIG. 2. Unsupervised machine-learning classification of vortex states for the case of two same-sign vortices. Vortex configurations corresponding to the initial conditions (a) $\phi/\pi = 0.349$, $L = 0.69R^2$, $\theta_{21} = \pi$ and (d) $\phi/\pi = 0.30$, $L = 0.22R^2$, $\theta_{21} = \pi$, respectively. (b, e) Vortex trajectories in real space corresponding to the initial configurations of (a) and (d) and integrated for the duration $60\Omega_0^{-1}$. (c, f) Velocity space representations of the vortex dynamics in (b) and (e), respectively. In (b), (c), (e), and (f) the initial and final vortex positions are shown using diamond and circular markers, respectively. The bottom row shows the machine-learning classification of the data based on the vortex position data (first column), the vortex trajectory data (second column), and the vortex velocity data (third column). The green and yellow markers in (g–i) label the two categories requested in the unsupervised machine-learning classification and are obtained for a set of 4000 initial conditions in (ϕ, L) space. The nonoverlapping vortex trajectories (b) are categorized as red markers and the overlapping trajectories (e) as blue markers. The black curves show the theory prediction for the phase boundary between the overlapping and nonoverlapping vortex states [39]. The pink curves are the boundaries generated using machine classification between the overlapping and nonoverlapping vortex states with 2000 initial configurations.

a nonoverlapping type. In Fig. 2(e) the vortex paths share a region of phase space and the configuration in Fig. 2(d) corresponds to an overlapping type. The topological distinctness of these two types of vortex configurations becomes even clearer when considering the velocity space representation [39], shown in Figs. 2(c) and 2(f).

To study the phase boundary between the overlapping and nonoverlapping phases [38,39], we generated 4000 initial vortex configurations in the (ϕ, L) space and used the unsupervised machine-learning approach to classify the vortex configurations. The images used for the classification in all cases were 221×221 pixels, to provide sufficient resolution for the machine-learning algorithms (especially for feature extraction process) to operate effectively. We trained the clustering SOM algorithm on the feature vectors with the image features extracted using the default parameters of the bag-of-features function.

Figure 2(g) shows the resulting unsupervised machine-learning classification with red and blue markers corresponding to the two classified categories. The green and yellow markers correspond to Figs. 2(a) and 2(d), respectively. The black curve shows the correct location of the phase boundary [38,39]. Although the configurations in Figs. 2(a) and 2(d) are classified correctly, the location of the phase boundary lies at a higher value of L, when compared with its correct value, for all considered values of ϕ.

In the second test, we trained the SOM clustering algorithm with images showing the full vortex trajectories, such as in Figs. 2(b) and 2(e), instead of the initial vortex configurations. The corresponding classification result shown in Fig. 2(h) is very similar to the case in Fig. 2(g) with an improvement in the accuracy of the location of the phase boundary. In the third case [Fig. 2(i)], we have taken the preprocessing of the data even further and have trained the SOM clustering
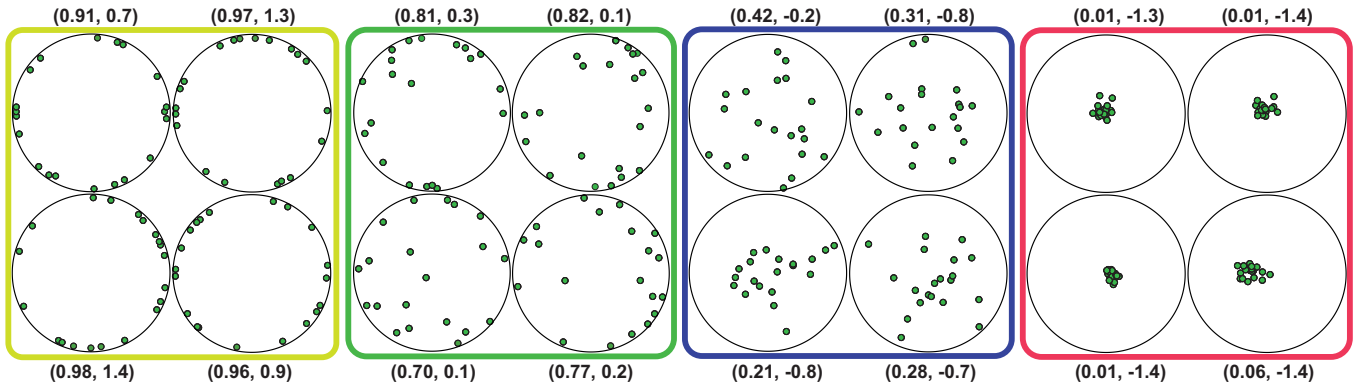
FIG. 3. Sixteen representative vortex configurations for the four color-coded categories in Fig. 4. The $(L/N_v, \beta)$ values of each vortex configuration are shown closest to the respective images.

algorithm with images showing the velocity space representation of the vortex trajectories. The classification in this case is in excellent agreement with the correct phase boundary. The two main conclusions from these test cases are that (i) the unsupervised machine-learning classification works for this problem remarkably well overall, and (ii) preprocessing the data before performing the classification can significantly improve the outcome of the classification. In addition, the number of sampled images affects the classification significantly, as demonstrated by the pink curves which show the identified boundary between the blue and red markers when 2000 initial configurations are used. When the number of samples is increased to 4000 the classified boundary shifts closer towards the correct (black curve) boundary [see Figs. 2(g)–2(i)]. However, it is not always the case that the classification accuracy would continually improve with the increasing number of training samples since excessive training may lead to overlearning complications.

Although using the velocity space representation is ideal for this two-vortex problem, our numerics indicates that similar benefit over the vortex trajectory representation in the case of many vortex configurations is not realized. Furthermore, when the number of vortices in the system increases, the trajectory images such as Figs. 2(b) and 2(e) become increasingly overcrowded and ultimately cannot be used for classification purposes as the whole image becomes covered densely for a fixed (resolution) trajectory linewidth. On the other hand, increasing the data resolution to resolve finer trajectory lines would rapidly lead to a memory bottleneck in computation. Consequently, for the remainder of this paper we are exclusively considering the plain vortex configurations such as Figs. 2(a) and 2(d), which also correspond to the most realistic experimental data in systems where real-time vortex tracking is not feasible.

B. Polarized vortex system

The success of the machine-learning approach in analyzing the topological phase boundary of a two-vortex system is encouraging and it motivates us to consider systems comprised of more vortices to assess the applicability of this machine model for classifying different phases of vortex matter.

For this purpose, we generate point-vortex configurations for $N_v = 20$ vortices of all same-sign circulation using the Monte Carlo simulation method. Considering different values of $N_v \geqslant 8$ yields qualitatively similar results. We produce a sample of 1000 images per temperature point representing the statistically equivalent vortex configurations. A total of 29 temperature points uniformly distributed over the ranges $\beta = [1.4, 0] \beta_{BKT}$ and $\beta = [0, -1.4] \beta_{EBC}$ are considered.

Figure 3 shows 16 example images of vortex configurations sampled at various temperatures. A collection of 29 000 such images were sampled from the Monte Carlo run and were fed into the machine-learning model. Then, we processed the images as per the procedure explained in the bag-of-features model in Sec. II D 3. Thereafter, the processed image features were used to train the clustering algorithm.

The classification result with four vortex phases of polarized vortices is demonstrated using the average vortex impulse per particle $(\bar{L}/N_v)$ as a function of inverse temperature (β) and the average energy per vortex $(\bar{E}/N_v)$ as shown in Figs. 4(a) and 4(b), respectively. In Fig. 4(a) the x axis is scaled by the critical inverse temperature $|\beta_{BKT}|$ and $|\beta_{EBC}|$ for positive and negative temperature, respectively. The error bars in Figs. 4(a) and 4(b) are one standard deviation of the statistical value of angular momentum per vortex, and the average energy per vortex, respectively. Samples of typical vortex configurations for the four color-coded categories of Fig. 4 are presented in Fig. 3 and are encapsulated in corresponding colored frames. The $(L/N_v, \beta)$ values of each of the 16 vortex configurations are shown closest to the respective images.

Figure 4 demonstrates that the unsupervised machine-learning approach is able to distinguish four different vortex phases according to their vortex temperature. At high positive inverse temperature the classified yellow category corresponds to the vortex dipole phase, where the real vortices and their image vortices are paired across the boundary of the circular BEC as shown in the four configurations in the yellow box (Fig. 3). On increasing the temperature the vortex pairs begin to unbind from their images. For zero core point vortices this transition occurs at critical temperature $\beta/|\beta_{BKT}| = 0.5$ and the machine-learning classification seems to capture this transition. The green category corresponds to the high-entropy, high-positive-temperature phase. When the
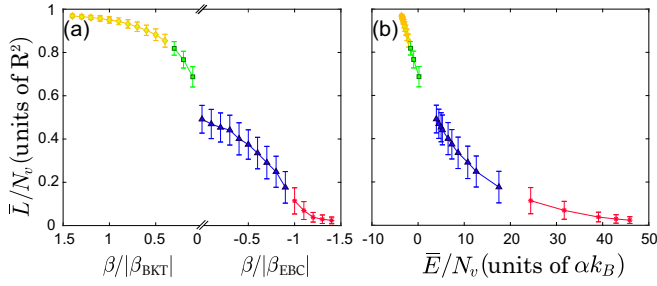
FIG. 4. Unsupervised machine-learning classification of point-vortex configurations into four categories, corresponding to the different marker styles and colors. (a, b) The vortex impulse per particle (L/N_v), as a function of inverse temperature (β) and the energy per vortex (E/N_v). In (a) the x axis is scaled by critical inverse temperature $|\beta_{EBC}|$ and $|\beta_{BKT}|$ for negative and positive temperatures, respectively. In both frames the y axis is representing the average vortex impulse per particle ($\bar{L}/N_v$) and in (b) the x axis is representing the average vortex energy per particle ($\bar{E}/N_v$). In both frames the classification is conducted for a system of $N_v = 20$ same-sign vortices using an ensemble of 1000 initial configurations. The error bars are one-standard-deviation statistical estimates.

temperature changes sign at $\beta = 0$ the output is classified in the blue category. Visually, many of the vortex configurations in the green and blue categories would be hard to distinguish whereas the machine-learning model has no difficulty in succeeding in this task. The fourth identified vortex matter phase is classified as a red category, which corresponds to the condensation of the well-defined Onsager vortex clusters taking place at critical temperature $\beta/|\beta_{EBC}| = -1$.

To obtain deeper understanding of these classification results, Fig. 5(a) shows a histogram that counts the number of images classified into each of the four categories at each value of β. In accordance with the histogram, we represent each data point of Fig. 4 using the majority color whose representation in the histogram exceeds 50%. To validate the SOM clustering
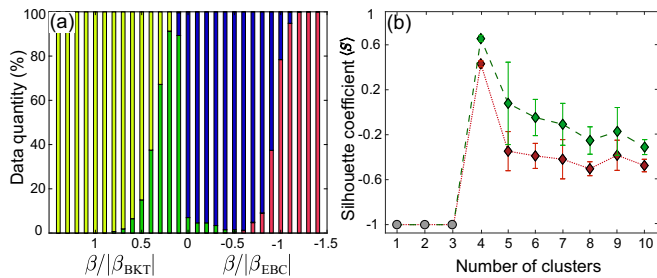

FIG. 5. (a) The percentage of images of the $N_v = 20$ polarized vortex system that were classified by SOM to belong to each of the four color-coded nodes as a function of the inverse temperature. (b) The average silhouette value of the data set computed using k-means clustering as a function of number of clusters over ten different experiments. The red dotted curve is the output of the same input data as in (a) and the green dashed curve corresponds to silhouette analysis of 20 same-sign vortices using an ensemble of 100 initial positions for one to ten clusters. The silhouette analysis produces an undefined value for cluster numbers 1–3. We set those points to the minimum silhouette value of -1 as shown by grey circular markers.

output for the optimal number of four clusters for the tested data, we create a silhouette criterion clustering evaluation using k-means clustering. The silhouette analysis provides the optimum cluster number (Silhouette Evaluation in MATLAB) and evaluates the quality of clustering [75].

Figure 5(b) presents the average silhouette coefficient $\langle S \rangle$ of clustering for one to ten clusters averaged over ten experiments. The red dotted curve is the average silhouette coefficient of clustering for the same data set that is clustered in Fig. 5(a) and the green dashed curve is the mean silhouette for the data clustering of a system of 20 same-sign vortices for 29 temperature points and an ensemble of 100 initial configurations at each temperature point. That is, for the later case, the total 2900 images were created and their features were extracted using the custom extractor function for the bagOfFeatures model in MATLAB. This function uses the default SURF feature extraction over a uniform grid of point locations at many scales. The used default values for this feature extraction for step size and multifeature scales are 8 and [1.6 3.2 4.8 6.4], respectively. Then using k-means clustering for those extracted features of 2900 images we evaluate the optimum cluster number by performing silhouette analysis. For both data sets we ran the experiment ten times for one to ten clusters and then averaged the silhouette value over the total number of experiments for each cluster number. The error bars are one-standard-deviation statistical estimates of silhouette coefficients.

In Fig. 5(b), the mean silhouette value is higher (green curve) for the smaller data set (preprocessed using a custom extractor function with dense grid for BoF) than the larger data set (preprocessed with grid step size [20 20]), but for both red and green curves the highest silhouette value occurs at four clusters, suggesting that the optimal number of clusters to be employed is four. This justifies our choice of four neurons in the SOM classification, as using either more or fewer clusters leads to poorer clustering as a result of lower silhouette value. As such, the unsupervised machine-learning classification is not only able to identify the boundaries of the four temperature regions but is also able to identify the correct number of physically meaningful regions.

C. Neutral vortex system

Having successfully classified the single sign vortices (polarized vortex fluid) into four temperature regions, we next repeat the analysis for the case of a neutral vortex system (with equal number of vortices and antivortices) having in total $N_v = 40$ vortices. Considering different values of $N_v \geqslant 14$ yields qualitatively similar results. As for the polarized vortex systems, the test images of the neutral vortex system using an ensemble of 1000 initial configurations were produced using Monte Carlo simulation at 31 temperature points. These temperature points over the ranges $\beta = [1.4, 0] \beta_{BKT}$ and $\beta = [0, -1.4] \beta_{EBC}$ were uniformly distributed with increments of 0.1 except for $\beta = [-0.8, -0.9] \beta_{EBC}$ and $\beta = [-1.1, -1.2] \beta_{EBC}$, for which an increment of 0.05 was used. For machine-learning classification the algorithm is trained on the feature vectors of images containing positions of vortices and antivortices plotted in green and blue circular markers as shown through 16 representative images in Fig. 6. Although
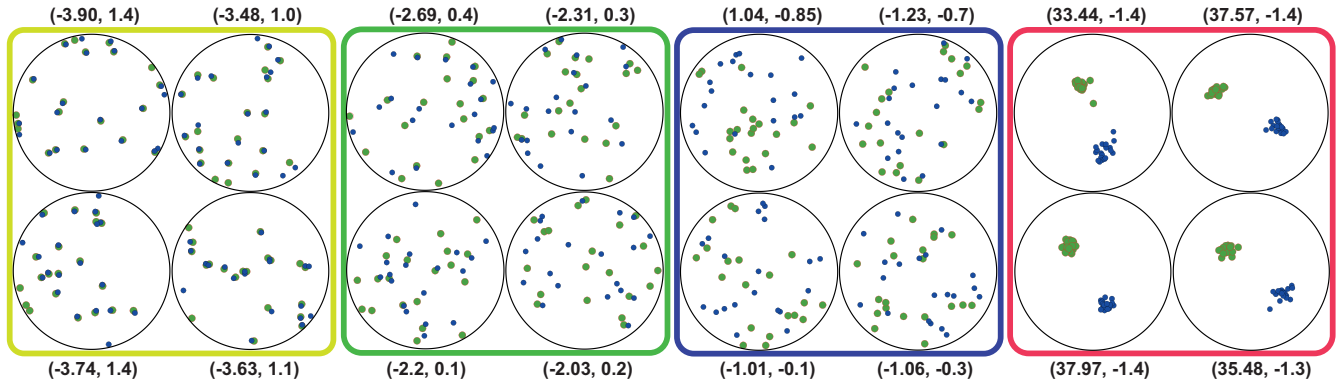
FIG. 6. Sixteen representative vortex configurations for the four color-coded categories in Fig. 7. The $(L/N_v, \beta)$ values of each vortex configuration are shown closest to the respective images.

these images carry the information of the vortex circulation sign in terms of the color for the machine-learning model, removing the color coding has surprisingly little influence on the outcome of the classification [76]. As in the case of the polarized vortex system, here we also clustered the data into four categories.

The quantitative classification results are again demonstrated using the average energy per particle $(\bar{E}/N_v)(\alpha k_B)$ and the average vortex cluster fraction $(\bar{C}_f(N_v))$ as a function of temperature (β) as shown in Figs. 7(a) and 7(b), respectively. The energy is in units of αk_B, where $\alpha = \rho_s \Gamma^2 / 4\pi k_B$. In both frames the x axis is scaled by critical inverse temperature $|\beta_{\text{BKT}}|$ and $|\beta_{\text{EBC}}|$ for positive and negative temperature ranges, respectively. The error bars in Figs. 7(a) and 7(b) are one standard deviation in statistical value of energy per vortex and cluster fraction as a function of temperature, respectively. The vortex configurations for the four color-coded categories in Fig. 7 are presented in Fig. 6 encapsulated by the corresponding colored boxes. The $(\beta, E/N_v)$ values of each vortex

configuration are shown closest to the respective images. The vortices and antivortices are indicated by green (bigger) and blue (smaller) circular markers, respectively.

The unsupervised SOM algorithm again distinguishes excellently the four temperature regions. At high positive inverse temperature the yellow category corresponds to the lowest energy configurations within the pair collapse phase, where the vortices and antivortices are now paired up in the bulk (instead of forming edge states as in the polarized vortex fluid). A sample of corresponding configurations is shown in the yellow box of Fig. 6. On increasing the temperature the transition from vortex pair collapse to vortex unbinding takes place at $\beta/|\beta_{\text{BKT}}| = 0.5$ and leads to the green category in Fig. 7. The negative-temperature configurations around at $\beta = 0$ are classified into the blue category which corresponds to the random negative absolute temperature vortex states. The red category is again identified as the Einstein-Bose condensate phase of Onsager vortices that emerges at $\beta/|\beta_{\text{EBC}}| = -1$.
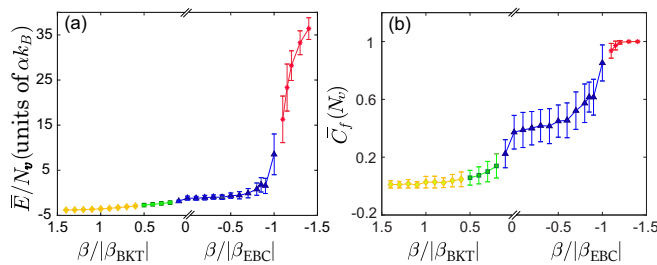

FIG. 7. Unsupervised machine-learning classification of point-vortex configurations into four categories, corresponding to the different marker styles and colors. (a) The energy per particle (E/N_v) and (b) the vortex cluster fraction as a function of inverse temperature (β). In both frames the x axis is scaled by critical inverse temperature $|\beta_{\text{EBC}}|$ and $|\beta_{\text{BKT}}|$ for negative and positive temperatures, respectively. In (a) the y axis is representing the average energy per particle $(\bar{E}/N_v)$ and in (b) the y axis is representing the average cluster fraction $(\bar{C}_f)$ in the units of number of vortices. In both cases the classification is conducted for a neutral vortex system of $N_v = 40$ vortices using an ensemble of 1000 initial configurations. The error bars are one-standard-deviation statistical estimates.
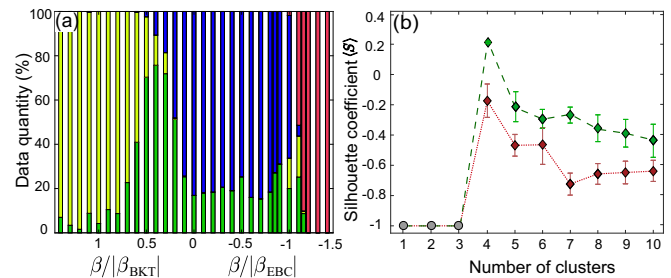

FIG. 8. (a) The fraction of images of a neutral vortex system classified by a SOM into the four nodes as a function of inverse temperature. The clustering was performed for a system of $N_v = 40$ vortices using an ensemble of 1000 initial configurations. (b) The average of silhouette value for simulation results of k-means clustering as a function of number of clusters over ten experiments. The red dotted curve is the output of the same input data as in (a) and the green dashed curve corresponds to silhouette analysis of a system of 40 neutral vortices using an ensemble of 100 initial positions. The silhouette analysis produces an undefined value for cluster numbers 1–3. We set those points to the minimum silhouette value of -1 as shown by grey circular markers.

The histogram of the classification is shown in Fig. 8(a) and, as for the polarized vortex fluid, this was used for assigning the color coding in Fig. 7 based on the 50% criterion. Although the overall classification is again very good, configurations belonging to the green category are now observed infrequently at nearly all temperatures. The choice of four data clusters is again justified by performing the silhouette analysis using k-means clustering for optimum cluster number evaluation. The resulting analysis of the average silhouette value for one to ten clusters over ten trials is shown in Fig. 8(b). The red dotted curve corresponds to the average silhouette value of clustering data (features extracted using [20 20] grid and 250 visual words in bag-of-features function from 31 000 input images) as used in Fig. 8(a). The green dashed curve corresponds to the average silhouette value of clustering data (3100 input images with 100 ensembles at each temperature point). For this smaller data set the features were extracted similarly using the custom feature extractor function as explained for the data set of 2900 images in the polarized vortex system. The result shows that the smaller data set whose features were extracted using a dense grid has a higher silhouette value than the larger data set (preprocessed with sparse grid), yet the highest silhouette coefficient occurs at four clusters in both cases. This justifies our choice of four neurons in SOM clustering as an optimal value for this data. In addition to this the low silhouette value (red curve) in Fig. 8(b) can be understood by inspecting the vortex configurations for larger data system. It is clear that a small fraction of vortex dipoles are present at nearly all temperatures, leading to the green category stretching over to the other temperature regions. To mitigate this issue, a further improvement of both training data and machine model, for example, high-quality vortex configuration images, additional thermalization points in the testing data preparation, or more precise feature extraction, could be performed before feeding them to the classifier model. However, even with the minimal amount of preprocessing, the classification results, as shown in Fig. 7, are remarkably good. Figure 8(b) shows that the highest silhouette value occurs at four clusters, illustrating that the four-neuron case is the optimal case also for the neutral vortex system.

IV. CONCLUSIONS

The main objective of this paper was to test the feasibility of using a simple unsupervised machine-learning approach to search for new vortex phases of matter and to identify the corresponding transition temperatures. We demonstrated the success of this approach using the vortex positions and

their circulation sign as the input information for the machine model. Surprisingly similar results were obtained also when the vortex sign information was hidden.

In the first part, we considered two same-sign vortices and trained a neural network using three types of preprocessed input images comprising the vortex positions, vortex trajectories, and velocity space images. In each case the boundary separating overlapped and nonoverlapped phase-space regions [39] was successfully detected and the sample size dependence of the location of the phase boundary result was demonstrated.

When considering larger numbers of vortices in both polarized and neutral vortex configurations, the unsupervised artificial neural network displayed consistent classification results when compared with previously known results. Specifically, both the positive-temperature Kosterlitz-Thouless transition in a two-dimensional Coulomb gas (point-vortex model) and the negative-temperature Einstein-Bose condensation of Onsager vortices transitions were successfully identified. The obtained results were also found to be robust with respect to fluctuations in the number of vortices considered [76].

Furthermore, with the aid of silhouette analysis, the unsupervised machine-learning model was also able to self-generate information on the optimal number of clusters to be employed for the classification and thereby the number of distinct temperature regions in the provided data set.

Considering that identification of the Kosterlitz-Thouless transition was found to be challenging even for complex convolutional neural network under a supervised approach [24], that our simplified unsupervised approach was able to detect the vortex binding-unbinding transition in this system shows promise for further applications of this methodology.

In light of the demonstrated performance, unsupervised machine learning has the potential to widen our understanding of topological phases of two-dimensional vortex matter, and may find applications in discovering exotic underlying vortex features both in theoretical models and in laboratory experiments, especially in the context of the research on two-dimensional quantum turbulence.

ACKNOWLEDGMENTS

[1] J. W. Gibbs, *Elementary Principles in Statistical Mechanics* (Dover Publications, Inc, Mineola, New York, 2014).

[2] C. Gogolin and J. Eisert, Rep. Prog. Phys. **79**, 056001 (2016).

[3] J. Sethna, *Statistical Mechanics: Entropy, Order Parameters, and Complexity*, Vol. 14 (Oxford University Press, New York, 2021).

[4] L. Onsager, Il Nuovo Cimento (1943-1954) **6**, 279 (1949).

[5] G. Gauthier, M. T. Reeves, X. Yu, A. S. Bradley, M. A. Baker, T. A. Bell, H. Rubinsztein-Dunlop, M. J. Davis, and T. W. Neely, Science **364**, 1264 (2019).

[6] S. P. Johnstone, A. J. Groszek, P. T. Starkey, C. J. Billington, T. P. Simula, and K. Helmerson, Science **364**, 1267 (2019).

[7] M. Inguscio, S. Stringari, and C. Wieman, in *Proceedings of the International School of Physics "Enrico Fermi," Course CXL*,

edited by M. Inguscio, S. Stringari, and C. E. Wieman (IOS Press, Amsterdam, 1999).

[8] J. Carrasquilla and R. G. Melko, Nat. Phys. **13**, 431 (2017).

[9] E. P. Van Nieuwenburg, Y.-H. Liu, and S. D. Huber, Nat. Phys. **13**, 435 (2017).

[10] S. J. Wetzel and M. Scherzer, Phys. Rev. B **96**, 184410 (2017).

[11] Y. Miyajima, Y. Murata, Y. Tanaka, and M. Mochizuki, Phys. Rev. B **104**, 075114 (2021).

[12] X.-Y. Dong, F. Pollmann, and X.-F. Zhang, Phys. Rev. B **99**, 121104(R) (2019).

[13] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Rev. Mod. Phys. **91**, 045002 (2019).

[14] A. A. Shirinyan, V. K. Kozin, J. Hellsvik, M. Pereiro, O. Eriksson, and D. Yudin, Phys. Rev. B **99**, 041108(R) (2019).

[15] N. Käming, A. Dawid, K. Kottmann, M. Lewenstein, K. Sengstock, A. Dauphin, and C. Weitenberg, Mach. Learn.: Sci. Technol. **2**, 035037 (2021).

[16] P. Zhang, H. Shen, and H. Zhai, Phys. Rev. Lett. **120**, 066401 (2018).

[17] W. Zhang, J. Liu, and T.-C. Wei, Phys. Rev. E **99**, 032142 (2019).

[18] K. Shiina, H. Mori, Y. Okabe, and H. K. Lee, Sci. Rep. **10**, 2177 (2020).

[19] G. Torlai and R. G. Melko, Phys. Rev. B **94**, 165134 (2016).

[20] G. Carleo and M. Troyer, Science **355**, 602 (2017).

[21] L. Wang, Phys. Rev. B **94**, 195105 (2016).

[22] M. S. Scheurer and R.-J. Slager, Phys. Rev. Lett. **124**, 226401 (2020).

[23] M. Richter-Laskowska, H. Khan, N. Trivedi, and M. M. Maśka, Condens. Matter Phys. **21**, 33602 (2018).

[24] M. J. S. Beach, A. Golubeva, and R. G. Melko, Phys. Rev. B **97**, 045207 (2018).

[25] Y. Zhang and E.-A. Kim, Phys. Rev. Lett. **118**, 216401 (2017).

[26] P. Huembeli, A. Dauphin, and P. Wittek, Phys. Rev. B **97**, 134109 (2018).

[27] J. F. Rodriguez-Nieva and M. S. Scheurer, Nat. Phys. **15**, 790 (2019).

[28] V. Dunjko and H. J. Briegel, Rep. Prog. Phys. **81**, 074001 (2018).

[29] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Nature (London) **549**, 195 (2017).

[30] M. Schuld, I. Sinayskiy, and F. Petruccione, Contemp. Phys. **56**, 172 (2015).

[31] J. M. Kosterlitz and D. J. Thouless, J. Phys. C: Solid State Phys. **6**, 1181 (1973).

[32] E. Greplova, A. Valenti, G. Boschung, F. Schäfer, N. Lörch, and S. D. Huber, New J. Phys. **22**, 045003 (2020).

[33] K. Fukushima, S. S. Funai, and H. Iida, arXiv:1908.00281.

[34] I. T. Jolliffe, in *Principal Component Analysis*, Springer Series in Statistics (Springer-Verlag New York, 2002), pp. 338–372, 2nd ed.

[35] S. J. Wetzel, Phys. Rev. E **96**, 022140 (2017).

[36] W. Hu, R. R. P. Singh, and R. T. Scalettar, Phys. Rev. E **95**, 062122 (2017).

[37] C. Wang and H. Zhai, Phys. Rev. B **96**, 144432 (2017).

[38] R. Navarro, R. Carretero-González, P. J. Torres, P. G. Kevrekidis, D. J. Frantzeskakis, M. W. Ray, E. Altuntaş, and D. S. Hall, Phys. Rev. Lett. **110**, 225301 (2013).

[39] A. V. Murray, A. J. Groszek, P. Kuopanportti, and T. Simula, Phys. Rev. A **93**, 033649 (2016).

[40] R. Cant, Combust. Flame **125**, 1361 (2001).

[41] H. Aref, J. Math. Phys. **48**, 065401 (2007).

[42] S. Weißmann, J. Nonlinear Sci. **24**, 359 (2014).

[43] L. Chomaz, L. Corman, T. Bienaimé, R. Desbuquois, C. Weitenberg, S. Nascimbene, J. Beugnon, and J. Dalibard, Nat. Commun. **6**, 6162 (2015).

[44] S. Tempone-Wiltshire, S. Johnstone, and K. Helmerson, Opt. Express **25**, 296 (2017).

[45] Y. B. Pointin and T. Lundgren, Phys. Fluids **19**, 1459 (1976).

[46] A. J. Groszek, M. J. Davis, D. M. Paganin, K. Helmerson, and T. P. Simula, Phys. Rev. Lett. **120**, 034504 (2018).

[47] R. N. Valani, A. J. Groszek, and T. P. Simula, New J. Phys. **20**, 053038 (2018).

[48] M. T. Reeves, T. P. Billam, B. P. Anderson, and A. S. Bradley, Phys. Rev. Lett. **110**, 104501 (2013).

[49] A. M. Salzberg and S. Prager, J. Chem. Phys. **38**, 2587 (1963).

[50] E. Hauge and P. Hemmer, Phys. Norv. **5**, 209 (1971).

[51] J. Viecelli, Phys. Fluids **7**, 1402 (1995).

[52] V. Berezinskii, Zh. Eksp. Teor. Fiz. **59**, 907 (1970) [Sov. Phys. JETP **32**, 493 (1971)].

[53] V. Berezinskii, Zh. Eksp. Teor. Fiz. **61**, 1144 (1971) [Sov. Phys. JETP **34**, 610 (1972)].

[54] R. H. Kraichnan, Phys. Fluids **10**, 1417 (1967).

[55] R. H. Kraichnan and D. Montgomery, Rep. Prog. Phys. **43**, 547 (1980).

[56] S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, in *Emerging Artificial Intelligence Applications in Computer Engineering*, edited by I. Maglogiannis *et al.* (IOS Press, Amsterdam, 2007), Vol. 160, p. 3.

[57] T. Kohonen, Biol. Cybern. **43**, 59 (1982).

[58] T. Kohonen and T. Honkela, Scholarpedia **2**, 1568 (2007).

[59] T. Kohonen, Neural Networks **37**, 52 (2013).

[60] M. G. Pena, Master's thesis, Utah State University, 2011.

[61] B. Jähne, H. Haussecker, and P. Geissler, *Handbook of Computer Vision and Applications* (Academic Press, New York, 1999), Vol. 2.

[62] N. S. Mansoori, M. Nejati, P. Razzaghi, and S. Samavi, in *2013 21st Iranian Conference on Electrical Engineering (ICEE)* (IEEE, New York, 2013), pp. 1–6.

[63] M. V. N. Bedo, D. P. Dos Santos, M. Ponciano-Silva, P. M. de Azevedo-Marques, A. P. D. L Ferreira de Carvalho, and C. Traina, J. Digital Imaging **29**, 22 (2016).

[64] V. Suma, J. Inf. Technol. Digital World **2**, 151 (2020).

[65] Y. Zhang, R. Jin, and Z.-H. Zhou, Int. J. Mach. Learn. Cybern. **1**, 43 (2010).

[66] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, Comput. Vision Image Understanding **110**, 346 (2008).

[67] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray, in *Workshop on Statistical Learning in Computer Vision, ECCV* (Prague, 2004), pp. 1–22.

[68] A. Likas, N. Vlassis, and J. J. Verbeek, Pattern Recognit. **36**, 451 (2003).

[69] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, IEEE Trans. Pattern Anal. Mach. Intell. **24**, 881 (2002).

[70] S. Shukla and S. Naganna, Int. J. Inf. Comput. Technol. **4**, 1847 (2014).

[71] M. Pacella, A. Grieco, and M. Blaco, Comput. Intell. Neurosci. **2016**, 5139574 (2016).

[72] J. Vesanto and E. Alhoniemi, IEEE Trans. Neural Networks **11**, 586 (2000).

[73] T. Chu, Y. Chen, L. Huang, Z. Xu, and H. Tan, Remote Sens. **12**, 3978 (2020).

[74] J.-O. Palacio-Niño and F. Berzal, arXiv:1905.05667.

[75] P. J. Rousseeuw, J. Comput. Appl. Math. **20**, 53 (1987).

[76] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevA.105.033301 for supplementary figures.

[77] S. Middelkamp, P. Torres, P. Kevrekidis, D. Frantzeskakis, R. Carretero-González, P. Schmelcher, D. Freilich, and D. Hall, Phys. Rev. A **84**, 011605 (2011).

[78] G. Moon, W. J. Kwon, H. Lee, and Y.-i. Shin, Phys. Rev. A **92**, 051601 (2015).

[79] A. L. Fetter and A. A. Svidzinsky, J. Phys. Condens. Matter **13**, R135 (2001).

[80] D. Freilich, D. Bianchi, A. Kaufman, T. Langin, and D. Hall, Science **329**, 1182 (2010).

[81] T. W. Neely, E. C. Samson, A. S. Bradley, M. J. Davis, and B. P. Anderson, Phys. Rev. Lett. **104**, 160401 (2010).