


Bayesian modeling of pattern formation from one snapshot of patternNatsuhiko Yoshinaga^{1,2,*} and Satoru Tokuda^{3,2}¹*WPI-Advanced Institute for Materials Research, Tohoku University, Sendai 980-8577, Japan*²*MathAM-OIL, AIST, Sendai 980-8577, Japan*³*Research Institute for Information Technology, Kyushu University, Kasuga 816-8580, Japan* (Received 12 January 2022; revised 22 June 2022; accepted 7 November 2022; published 7 December 2022)

Partial differential equations (PDEs) have been widely used to reproduce patterns in nature and to give insight into the mechanism underlying pattern formation. Although many PDE models have been proposed, they rely on the pre-request knowledge of physical laws and symmetries, and developing a model to reproduce a given desired pattern remains difficult. We propose a method, referred to as Bayesian modeling of PDEs (BM-PDEs), to estimate the best dynamical PDE for one snapshot of a objective pattern under the stationary state without ground truth. We apply BM-PDEs to nontrivial patterns, such as quasicrystals (QCs), a double gyroid, and Frank-Kasper structures. We also generate three-dimensional dodecagonal QCs from a PDE model. This is done by using the estimated parameters for the Frank-Kasper A15 structure, which closely approximates the local structures of QCs. Our method works for noisy patterns and the pattern synthesized without the ground-truth parameters, which are required for the application toward experimental data.

DOI: [10.1103/PhysRevE.106.065301](https://doi.org/10.1103/PhysRevE.106.065301)**I. INTRODUCTION**

The design of structures of materials is one of the most important issues in various fields of physical science, as their structures are related to their physical properties. The structures are often characterized by periodic or quasiperiodic order. These ordered structures, which we call a *pattern*, are ubiquitous in nature ranging from fluid convection [1] to the microphase separation of block copolymers [2,3] and atomic and molecular crystals [4,5]. Surprisingly, the same pattern appears in different systems with completely different length scales [6]. Complex patterns such as quasicrystal (QC), double gyroid (DG), and Frank-Kasper (FK) phases appear not only in metallic alloys [7,8] but also in soft materials such as block copolymers [2,9], biomaterials [10], surfactants [11], liquid crystals [12], and colloidal assemblies [13].

Understanding a generic mechanism of symmetry selection is an important step to understand structural pattern formation. A continuum approach, using nonlinear partial differential equations (PDEs), is useful for this purpose [5,6]. For example, it was shown that at least two length scales are necessary for the formation of QCs by using a phenomenological model [14]. Once we find the phenomenological PDE model, it also gives an insight into microscopic pictures required for the pattern, such as interactions between particles [15]. These studies have clarified generic pictures on a specific pattern. However, when encountering a new pattern, we do not know the interactions leading to the pattern and, therefore, need to find a

governing equation and parameters. In ideal cases, PDE models can be derived from microscopic models, but this is limited due to technical difficulties. Therefore, finding the governing equation requires a sophisticated guess and trial and error. In this study, we develop a systematic method to estimate the best model for the objective pattern we want to reproduce.

Recent developments in imaging techniques give us various structural information, and thus it is desired to understand pattern formation and the design principle of a desired structure [16]. There are two challenging issues to consider the inverse structural design applied to the real-world data. First, the structural data are often stationary because images with a high spatiotemporal resolution during structural formation are hard to acquire. Experimentally obtained structure is stable, or at least metastable. We want to estimate the dynamical model that produces a stable stationary pattern that best matches the snapshot given as an objective pattern. The estimated model may give us an insight into the dynamics of the pattern formation. Second, there is no ground truth of the objective structure; the true model to reproduce the experimental data is not available. Still, finding a surrogate model or phenomenological model is helpful to clarify underlying mechanisms of the structural formation [6]. This issue is called model inadequacy and is one of the biggest challenges in model estimation [17,18].

To overcome these issues, in this study, an inverse problem is formulated to reproduce a given snapshot of a pattern, after which we propose a method, referred to as Bayesian modeling of partial differential equations (BM-PDEs), to identify the best dynamical PDE model and its parameters. We apply our method to the problem without ground truth. We prepare the objective pattern according to crystallographic symmetry, which is independent of any candidate models, and then perform an estimation of the best model. We demonstrate the BM-PDE for complex patterns such as QC, DG, and FK

*Corresponding author: yoshinaga@tohoku.ac.jp

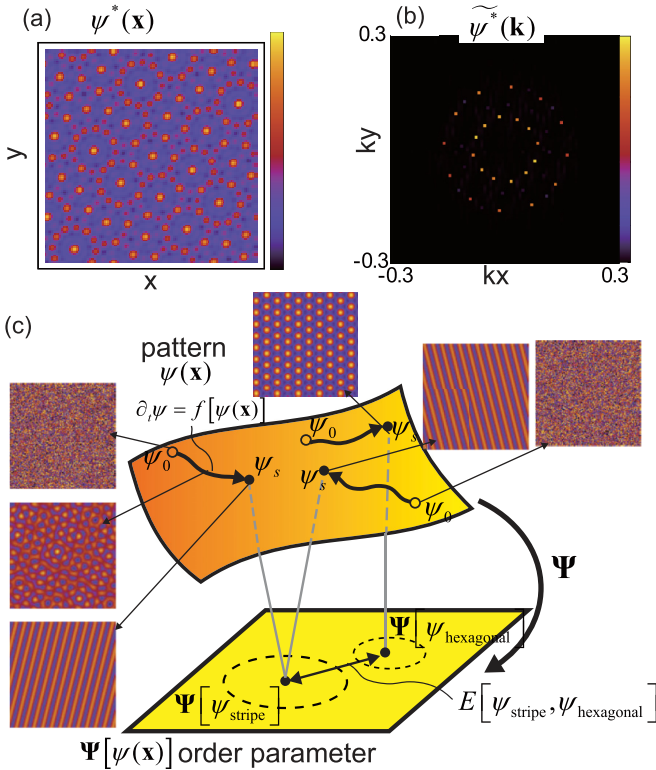


FIG. 1. Schematic illustration of Bayesian modeling of partial differential equations (BM-PDEs). (a), (b) The example of the objective pattern of a dodecagonal QC produced from a numerical result (a) and its Fourier transform (b). The color bar indicates $[-1.5, 3.5]$ in (a) and $[0, 3000]$ in (b). (c) The space of patterns $\psi(\mathbf{x})$ and order parameters $\Psi[\psi(\mathbf{x})]$. The PDE is solved with the initial condition ψ_0 taken from random variables. For each trajectory, the translational position and orientation of the pattern ψ_s , would change even under exactly the same parameters and the same model. The order parameter identifies the two patterns by extracting symmetries of the pattern. The distance between two patterns is quantified by E .

A15 patterns. We also demonstrate that from the estimated parameters, three-dimensional dodecagonal QCs can be generated. The success shows a potential application of BM-PDEs to understand the mechanism of structural formation of novel materials.

II. BASIC FORMULA

We consider a pattern (or crystalline structures) expressed by the scalar density field $\psi(\mathbf{x})$. An example of a two-dimensional dodecagonal QC (DDQC) is shown in Figs. 1(a) and 1(b). Higher density spots may be considered as a position of particles. We estimate a dynamical model to reproduce an objective pattern $\psi^*(\mathbf{x})$ as a stable pattern at the steady state $\psi_s(\mathbf{x})$ of a nonlinear partial differential equation $\partial_t \psi(\mathbf{x}) = f_\mu[\psi(\mathbf{x})]$ [Fig. 1(c)]. If the PDE and its parameters μ are ground truth for the objective pattern, $\partial_t \psi^*(\mathbf{x}) = f_\mu[\psi^*(\mathbf{x})] = 0$ is satisfied. Our objective pattern is *one* snapshot and has information only about the stationary state (Fig. 2). Its transient structure from the initial state to the stationary state is not available. We assume the stationary state is stable in the sense that the pattern is generated from a

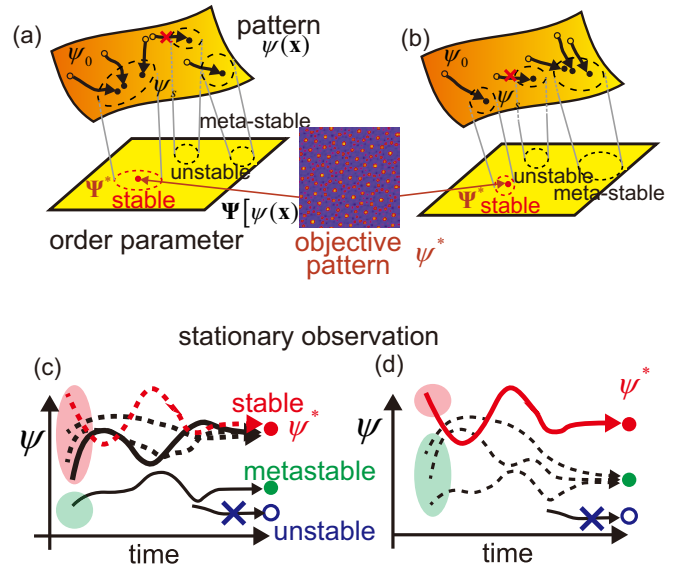


FIG. 2. Stability of stationary states and initial conditions. A pattern may be stable, metastable, or unstable. The unstable pattern cannot be generated from an initial condition. The better model (a), (c) has broader initial conditions that generate stable patterns, than the worse model (b), (d).

broad range of the initial conditions (Fig. 2). This assumption is natural when the objective pattern is obtained from an experimental result; the pattern should be reproducible.

At first look, the estimation for the stationary data is impossible. When $f_\mu[\psi^*(\mathbf{x})] = 0$ is a true model, we may have a series of *equally true* models, such as $f_\mu[\psi^*(\mathbf{x})]^2 = 0$, and $(f_\mu[\psi^*(\mathbf{x})] + 1)f_\mu[\psi^*(\mathbf{x})] = 0$. Therefore, the estimation is not unique. Nevertheless, as we will see later, the estimation of a model that reproduces the objective pattern as a stable structure plays a role as regularization (see also Appendix A4 for the comparison with other approaches). We should note that our problem is not the parameter estimation for $f_\mu[\psi^*(\mathbf{x})] = 0$ in which the stability of the stationary state is not guaranteed.

To see the difficulty of estimating a dynamical equation that reproduces the objective pattern only from its stationary data, it is instructive to consider the state-space model, widely used in data assimilation [19,20]. The cost function E consists of measurement (observation) and model errors and is expressed as

$$E[\mu, \psi(\mathbf{x})] = \frac{1}{2} \|\psi^*(\mathbf{x}) - \psi(\mathbf{x})\|^2 + \frac{1}{2} \|\partial_t \psi(\mathbf{x}) - f_\mu[\psi(\mathbf{x})]\|^2. \quad (1)$$

In many cases, the norm $\|\cdot\|$ is chosen to be the square norm. If the observation contains an error, we have to estimate both the parameters μ and the state $\psi(\mathbf{x})$. When the model represents a deterministic system, the state is described by its initial condition $\psi_0(\mathbf{x})$, and accordingly the cost function becomes $E[\mu, \psi_0(\mathbf{x})]$. When the observation does not contain noise, the first term in Eq. (1) vanishes, and the problem falls into a simple regression (see also Appendix A4). In the conventional data assimilation, both μ and ψ_0 are estimated by minimizing $E[\mu, \psi_0(\mathbf{x})]$ [20]. However, in the problem of pattern formation, the specific initial condition to produce

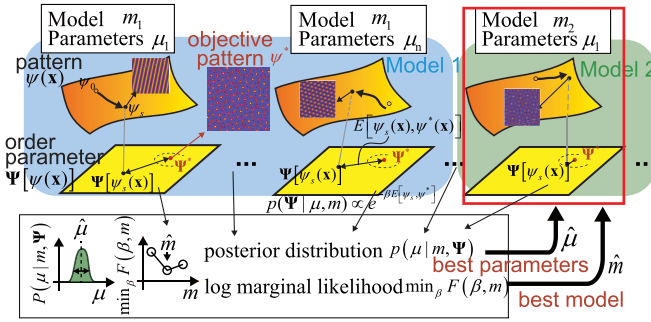


FIG. 3. Schematic illustration of Bayesian modeling of partial differential equations (BM-PDEs). For each model and each set of parameters, there is a stationary pattern ψ_s . The cost function $E[\psi^*, \psi_s]$ (energy) is calculated from the order parameters of the objective and generated patterns. From the posterior distribution, the best parameters and their errors are estimated. The distribution of the cost function gives the log marginal likelihood of the model from which the selection of models can be made.

the pattern is not of our interest. Moreover, there are many initial conditions that asymptotically reproduce the same pattern (Fig. 2). Therefore, the estimation under a given stationary objective pattern cannot be unique. When time-series data during pattern formation is available, the estimation is possible because each trajectory is from a different initial condition. In our method, we sample the stationary pattern by solving the model under each parameter and each realization from the random initial condition. By marginalizing the initial conditions, we may obtain a unique estimation.

The basic structure of our estimation is schematically shown in Fig. 3. Parameters are estimated from the posterior distribution under the objective pattern, whereas the best model is estimated from the log marginal likelihood. For a given model m_i and parameters μ in a PDE, the stationary pattern is uniquely determined under each initial condition, $\psi_0(\mathbf{x})$ [Fig. 1(c)]. We treat an initial pattern as a latent variable that is marginalized using a random variable for the initial condition. This is because the obtained pattern may be translationally shifted or rotated by changing an initial pattern [see Fig 1(c)]. We quantify the similarity between two patterns ψ_1 and ψ_2 by the distance between them defined as

$$E[\psi_1, \psi_2] = |\Psi[\psi_1] - \Psi[\psi_2]|^2. \quad (2)$$

Our objective pattern is ordered and has many invariants; the pattern must be identified under change by translation and rotation, and also, the pattern does not change by the action of the symmetry group that the pattern has [see Figs. 1(a) and 1(b)]. Here we introduce *order parameter* $\Psi[\psi(\mathbf{x})] = (\Psi_1[\psi(\mathbf{x})], \dots, \Psi_{l_0}[\psi(\mathbf{x})])$, which maps the pattern onto the feature space and eliminates the redundant information of the ordered pattern $\psi(\mathbf{x})$ due to symmetry [see Fig. 1(c) and Sec. VIII]. The distance defined by Eq. (2) identifies the patterns up to symmetry transformation thanks to the order parameter.

When $\psi_1(\mathbf{x})$ is a solution of a PDE $\psi_s(\mathbf{x})$ under fixed parameters μ and $\psi_2(\mathbf{x})$ is an objective pattern $\psi^*(\mathbf{x})$, Eq. (2) plays a role of a cost function in our problem (see also Sec. II A). It corresponds to the first term in (1). Since we

solve a PDE for each μ , the second term in (1) approximately vanishes. Note that each pattern in Eq. (2) is mapped onto the space of order parameters as in Fig. 1(c), whereas, in (1), the mapping was taken to be identity.

A. Bayesian modeling

Our goal is to find the most probable model \hat{m} described by a PDE and its parameters $\hat{\mu}$ for a given objective pattern $\psi^*(\mathbf{x})$. We also want to quantify the uncertainty of the estimation. To achieve this, we use the cost function $E[\psi_s, \psi^*]$ in Eq. (2), also called the energy, expressed by the order parameter Ψ , and compute the distance from the objective pattern, $\Psi^* = \Psi[\psi^*(\mathbf{x})]$, to the numerically generated stationary pattern for each model and parameter set, $\Psi[\psi_s(\mathbf{x})]$. Our purpose is not to estimate specific initial states ψ_0 for the objective pattern ψ^* , but to estimate the best model that could generate patterns similar to ψ^* independent of the initial state. Therefore, our best parameter set $\hat{\mu}$ is defined by the mean of the marginal probability distribution under a model m :

$$p(\mu | \Psi^*, \beta, m) = \int p(\psi_0) p(\mu | \Psi^*, \psi_0, \beta, m) d\psi_0. \quad (3)$$

The integral over the initial conditions ψ_0 implies that the posterior distribution of the parameters is chosen so that the estimated parameters can generate the objective pattern from a wide range of the initial conditions. We may avoid the parameters that can generate the objective pattern only for a specific initial condition [Figs. 2(b) and 2(d)]. Following Bayes' theorem, the posterior distribution under a fixed ψ_0 is given by

$$p(\mu | \Psi^*, \psi_0, \beta, m) = \frac{p(\Psi^* | \psi_0, \mu, \beta, m) p(\mu | m)}{p(\Psi^* | m, \beta)}. \quad (4)$$

The likelihood is represented by $p(\Psi^* | \mu, m, \beta) \propto e^{-\beta E[\psi_s, \psi^*]}$, where the inverse temperature β is associated with the variance of the observation noise. This likelihood implies that the error in the measurement is given by $\Psi^* = \Psi + \xi$ with the Gaussian noise ξ with zero mean and its variance β^{-1} . The prior distributions $p(\psi_0)$ and $p(\mu | m)$ are assumed as the uniform distribution. The normalization factor $p(\Psi^* | \beta, m)$, or the log marginal likelihood (free energy) $F(\beta, m) \equiv -\log p(\Psi^* | \beta, m)$, is one of the criteria of model selection and hyperparameter estimation [21–24]. Both $p(\Psi^* | \beta, m)$ and $F(\beta, m)$ play a role as a probability density of the models and hyperparameters (see Sec. VIII C). Therefore, our best model m and inverse temperature β are both determined by maximizing $p(\Psi^* | \beta, m)$, or equivalently, minimizing $F(\beta, m)$ [25].

B. Objective pattern without ground truth

In this work, we consider two types of objective patterns; (i) a numerical solution of the PDE model that we use and (ii) a pattern synthesized by superposition of plane waves as Eq. (19). The former has ground truth, whereas the latter does not. The synthesized pattern is independent of our PDE models, and therefore, it is not necessarily a solution for the PDEs. We will demonstrate that our approach still works for the problem without ground truth. This problem is an intermediate step between the estimation of the problem with ground-truth and experimental data. We do not know the true

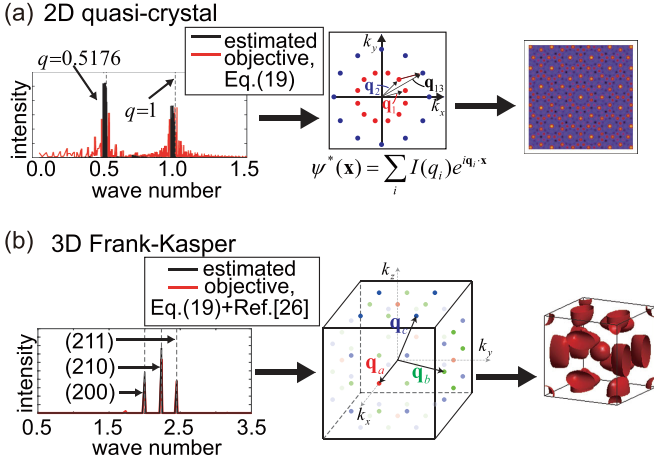


FIG. 4. The way to synthesize the objective patterns without ground truth by Eq. (19). (a) Two-dimensional dodecagonal QC and (b) FK A15. Left: The diffraction patterns of the objective and estimated patterns. Middle: From the peaks of the diffraction pattern, the spots in the Fourier space are identified for each wave mode. In (a) the spots correspond to $\mathbf{q}_i = q_0(\cos(2i + 1)\pi/12, \sin(2i + 1)\pi/12)$ for $i = 1, \dots, 12$, and $\mathbf{q}_i = q_1(\cos i\pi/6, \sin i\pi/6)$ for $i = 13, \dots, 24$. The ratio between the two wavelength $q_0/q_1 = 1/[2 \cos(\pi/12)] = 0.51\dots$. In (b), the spots indicate $\mathbf{q}_a = (2, 0, 0), (0, 2, 0), \dots$ in red points, $\mathbf{q}_b = (2, 1, 0), (1, -2, 0), \dots$ in green points, and $\mathbf{q}_c = (2, 1, 1), (1, 1, -2), \dots$ in blue points. Right: By superposition of the waves in the Fourier space, the real space objective patterns are synthesized by $\psi(\mathbf{x}) = \sum_i I(q_i)e^{i\mathbf{q}_i \cdot \mathbf{x}}$, where $I(q_i)$ is the intensity of the wave of \mathbf{q}_i , and the sum is taken over all the spots.

PDE for experimental data. Still, we would like to estimate the best PDE model to explain the data to understand the mechanism of the structural pattern formation. Our approach has the advantage in that we do not know *a priori* the model that reproduces the objective pattern. In most of the model identification, the problem with ground truth has been used.

To synthesize the objective structure by Eq. (19), positions of the peaks and their amplitude in the Fourier space are required [Fig. 4(a)]. These quantities can be directly measured in the scattering experiments in two dimensions and three dimensions. The positions and the amplitude in the Fourier space can be reproduced from the structure factor, as shown in Fig. 4. We use this information to synthesize the objective pattern. For the objective patterns of Figs. 7(a)–7(d) below, we artificially choose the value of the amplitude of the structure factor [Fig. 4(a)]. The structure factor $S_k = |\tilde{\psi}(\mathbf{k})|$ of the objective pattern of synthesized dodecagonal QC is shown in Fig. 4(a). The structure factor of the generated pattern with the estimated parameters has peaks at the same position as the objective pattern [see also Fig. 7(c), Fig. 17, and Table II]. The detailed method to synthesize the objective structure is shown in Sec. VIII B.

To claim that our approach can be applied to the experimental data, we demonstrate the estimation of PDEs for the FK A15 structure taken from the experimental data [26] [Fig. 4(b)]. The structure factors of the objective pattern and the generated pattern with the estimated parameters are shown in Fig. 4(b) [see also Fig. 7(e), Fig. 25, and Table VI]. The

peaks for the two patterns are similar not only in their wave numbers, but also in their relative amplitudes.

III. THE ESTIMATION FOR PHASE-FIELD CRYSTAL MODELS

We demonstrate our method for a class of phase-field crystal (PFC) models. The PFC models are conserved versions of the Swift-Hohenberg (SH) equation. The SH equation has one characteristic length in it and has been used as a basic phenomenological equation of periodic pattern formation [1]. We make a family of the models by changing mainly the linear part of Eq. (5). The family generalizes the PFC models to have multiple length scales (see Appendixes A 1 and B). Our approach is not limited to the choice of this family. In fact, we can replace Eq. (5) with any other family of models. A possible extension of this approach, including the estimation of nonlinear terms, is discussed in Sec. V. We consider a model, called $m \in M$, expressed by a nonlinear PDE of the form

$$\partial_t \psi(\mathbf{x}) = \mathcal{L}_\mu^{(m)} \psi(\mathbf{x}) + \mathcal{N}[\psi(\mathbf{x})] \quad (5)$$

with a set of parameters μ . The PDE is decomposed into two parts. The linear term is expressed by the linear operator, $\mathcal{L}_\mu^{(m)}$, acting on $\psi(\mathbf{x})$. Because we are interested in patterns in bulk, not affected by boundaries, we use periodic boundary conditions.

We make a family of models, $M = \{m_i\}_{i=1,2,\dots,i_{\max}}$. In model m_i the linear operator has i peaks in its spectrum [see Figs. 5(f)–5(h)]. Our family of models is designed to have a physical interpretation that the system has i length scales for model m_i because the number of peaks in the spectrum corresponds to the number of length scales. Each length scale is characterized by its wavelength q_i and the value of its spectrum at the wavelength a_i [see Figs. 5(f)–5(h)]. We also use the mean density $\bar{\psi}$ and system sizes in each direction as parameters. Note that because we consider conserved dynamics of the PFC model of Eq. (12), $\bar{\psi}$ is constant in time and can be used as a parameter of the estimation. We also note that $\bar{\psi}$ is identical to the parameter for the nonlinear ψ^2 term in Eq. (5) [see also Eq. (14)]. We fixed the number of mesh points in each spatial direction, but the mesh size is varied as a parameter during the estimation (see Sec. VIII A 1). This procedure corresponds to using the system size as a parameter of estimation.

A. Two-dimensional QC with ground truth

To give better insight into the BM-PDE, we first focus on an example of a two-dimensional QC with 12-fold symmetry (DDQC) shown in Fig. 1(a). This pattern has been studied using a model with two length scales [14]. The objective pattern in this section is numerically produced with a set of parameters μ^* . The two-length-scale model is used, that is, $m^* = m_2$.

For $m = m_2$, the cost function $E[\psi_s, \psi^*]$ decreases during the sampling from the posterior distribution, and the generated stationary patterns from Eq. (5) converge to QCs which are similar to the objective pattern in Fig. 5(b) (see also Fig. 16 in the Appendix). The estimated parameters well agree with

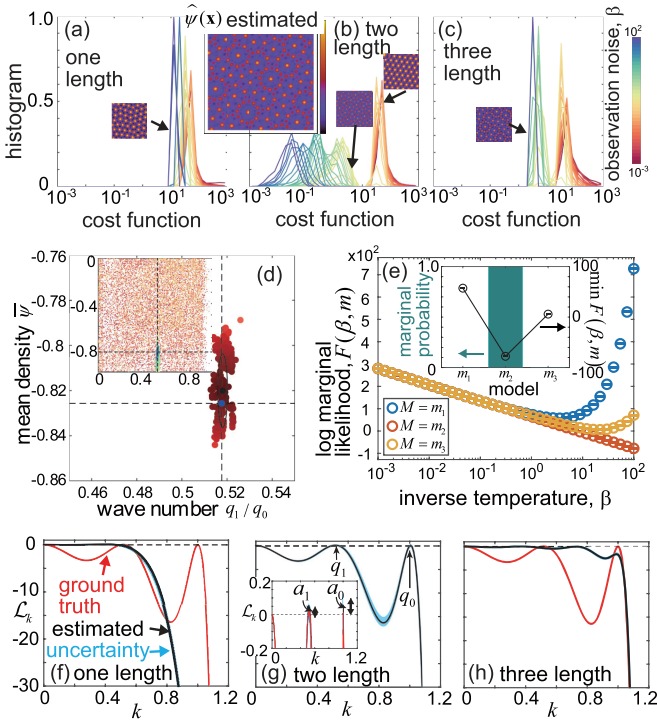


FIG. 5. Model selection and parameter estimation for the objective pattern of two-dimensional QC pattern with 12-fold (dodecagonal) symmetry. (a)–(c) The histograms of the cost function, $E[\psi^*, \psi_s]$, during the sampling. The horizontal axis is shown in the logarithmic scale. The generated pattern from the estimated PDE is shown in the insets. Typical patterns at each energy range are also shown in the insets with arrows. (d) The estimated parameters in the space spanned by q_1 and $\bar{\psi}$. The color indicates a histogram where darker red corresponds to a higher probability. The mean and standard deviation of the estimated parameters are shown by the black point and the black line, respectively. The ground-truth parameter values are shown in dashed lines and the blue point. The inset shows the same plot under various β in the range of parameters used for the prior distribution. The same color code as in (a)–(c) is used. (e) Model selection is made by the log marginal likelihood (free energy) calculated from the steady-state energy distribution for m_1 (a), m_2 (b), and m_3 (c). The inset shows the probability of each model marginalized for all β . The minimal free energy of each model is also shown with error bars, which overlap with the points. (f)–(h) The linear spectrum (black lines) as a function of wave number from the estimated parameters for m_1 (f), m_2 (g), and m_3 (h). The ground truth is shown in red lines. The uncertainty of the estimation is shown by the range in light blue. Note that in (g) the estimated line and ground truth are overlapped. The inset in (g) shows the same plot near the peaks.

the parameters that we used to generate the objective pattern [see Fig. 5(d) and Table I]. The estimated length scale is $q_1 = 0.52$ with the second length scale $q_0 = 1$. The ratio between them agrees to $q_0/q_1 = 2 \cos(\pi/12) \approx 1.9319$, which is the known value to generate this pattern [14]. The BM-PDE automatically estimates this ratio starting from uniform prior distribution of the wavelength. The estimation also works for other parameters (Table I and Fig. 16) in the Appendix.

Using the estimated parameters, we may generate a pattern similar to the objective pattern from uniform random initial

density [see the inset of Fig. 5(b)]. To see the quality of the estimation, we measure the steady distribution of the cost function. Figure 5(b) shows that there are two distinct states: one has a higher cost function $E \gtrsim 10^2$ and the other has a lower cost function $E \lesssim 10^2$. The latter corresponds to QC, whereas hexagonal patterns mainly dominate the former. The gap between the two states indicates that the QC patterns require a high resolution in the parameter search. The large step in the parameter space cannot find the optimal parameters because their range is narrow in the prior range of the parameters, as in the inset of Fig. 5(d). Therefore, the conventional gradient method with a fixed step size either cannot find the QC when the step size is large, or is impractical when it is small. The BM-PDEs use hierarchical sampling, such as replica exchange Monte Carlo (REMC) [27,28]. In this method, the knowledge of the step size corresponding to the observation noise β is not a prerequisite for the parameter search.

The same algorithm is performed for the models with one length scale $m = m_1$ and three length scale $m = m_3$. It is not surprising that the cost function E of the one-length-scale model is much higher than that of m_2 because it cannot reproduce a QC pattern [Fig. 5(a)]. In fact, the model $m = m_1$ with the estimated parameters produces hexagonal patterns rather than QCs. The three-length-scale model does reproduce QCs, which are comparable to the objective pattern [Fig. 5(c)]. However, our Bayesian estimation selects the two-length-scale model. Figure 5(e) demonstrates that the minimum free energy, that is, the log marginal likelihood, is lower for $m = m_2$. Therefore, we estimate that this QC pattern is described by the interaction of two modes with different length scales.

As the structure generated from the estimated parameters has a low-cost function of $E \approx 10^{-2}$, we may say the generated structure is close to the objective pattern. The next question is to what extent our model reproduces the underlying model for the objective pattern. To answer the question, first, we compare ensembles of patterns generated from ground-truth and estimated parameters. Figure 6(a) shows the histogram of the cost function for the ground-truth and estimated parameters. Even under the ground truth, the generated patterns are slightly different from the objective pattern because the initial conditions are different. The histogram shows that there is a peak for both cases at the same range of the cost function corresponding to the QCs. The probability of QCs is higher for the estimated parameters. This is because our method chooses the estimated parameters so that structures close to the objective pattern appear from a wide range of initial conditions (see Sec. II and Appendix A).

We also study the dynamics of the formation of QCs. We consider three cases: trajectory with the ground-truth parameters and from the ground-truth initial condition, trajectories with the ground-truth parameters from different initial conditions, and trajectories with the estimated parameters. Here the ground-truth initial condition means that the objective pattern in Fig. 1(a) is generated. In all cases, when QCs are formed, a random initial condition first gets smoother, then high-density spots appear, and finally, the QCs are formed [see Fig 6(b)]. Figures 6(c) and 6(d) show the structure factor $S_{q_{0,1}} = \langle |\hat{\psi}(\mathbf{k})|_{q_{0,1}} \rangle$ during the formation of QCs. These structure factors correspond to inhomogeneity at the two

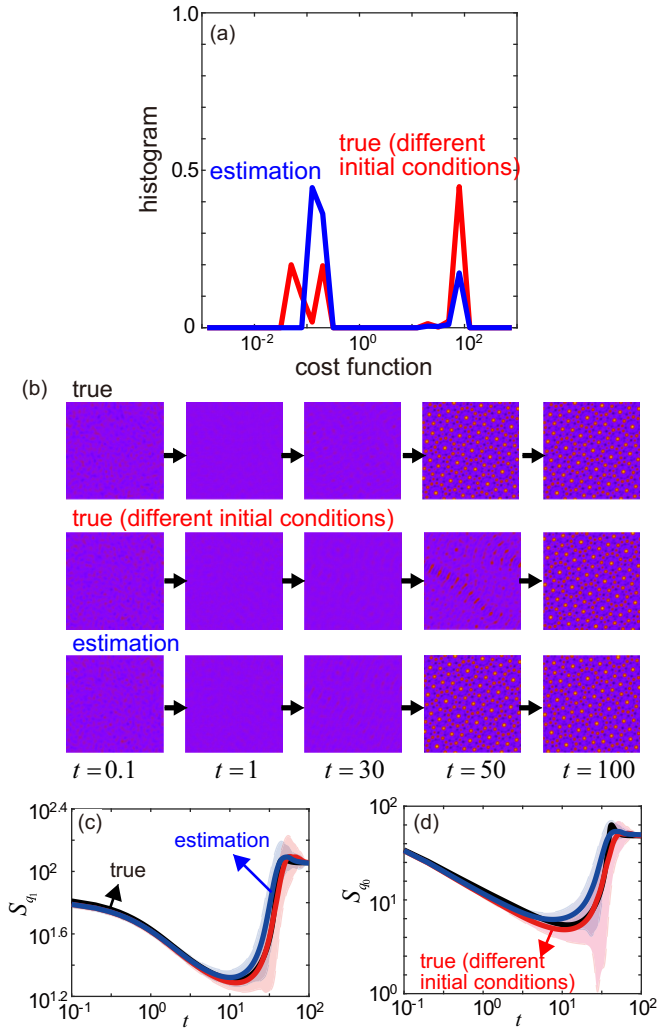


FIG. 6. Comparison of the estimated model with the ground truth. (a) The histogram of the cost function for the pattern generated from the model with the estimated parameters (blue) and with ground-truth parameters (red). (b)–(d) Dynamics of the formation of QCs using the estimated (blue) and ground truth (black and red) models. (b) Snapshots at each time step. (c), (d) The structure factors at $|\mathbf{k}| = q_0$ (d) and $|\mathbf{k}| = q_1$ (c). The trajectory for the model with ground-truth parameters and the initial condition is shown in a solid black line. The ensembles of trajectories for the model with the estimated parameters and ground truth from various random initial conditions are shown in solid blue and red lines, respectively. The standard deviation of each plot is shown as a shaded region.

length scales q_0 and q_1 . The average is taken over $|\mathbf{k}| = q_1$ or $|\mathbf{k}| = q_0$. Starting from random distribution, the pattern shows broad peaks at $|\mathbf{k}| = q_0, q_1$ at the early stage of the formation of QCs. Then sharper 12 peaks appear at each length scale, and finally, dodecagonal QCs appear.

With the ground-truth parameters but with different initial conditions, QCs can form. Depending on the initial conditions, the structures at the steady states are different, and accordingly, their trajectories are different. Still, the trajectory from the ground-truth initial condition is within the distribution of the trajectories from different initial conditions [Figs. 6(c) and 6(d)]. The trajectories generated from

the estimated parameters are also close to the ground-truth trajectories. These analyses indicate that our estimated model can not only generate a similar pattern to the objective one, but also reproduce the trajectory close to the one for the objective pattern. We stress that the uncertainty is not only given for the final pattern, but also for the dynamic evolution. This advantage of our method may be helpful to analyze patterns in experiments without ground truth, and to select a better model to interpret the patterns. We should remark that the dynamics are dependent on the system size. In a larger system size, the formation of QCs shows nucleation and growth [29]. Our system size during the estimation is small so that these processes are not seen in the dynamics. Still, if we use a larger system size, both estimated and ground-truth parameters show nucleation and growth.

B. Various objective patterns without ground truth

The BM-PDE is not restricted to the estimation of the parameters that are used to generate the objective pattern. Using a two-dimensional DDQC, we demonstrate that the BM-PDE successfully estimates the best model and approximated parameters for the objective pattern without ground truth. The DDQC is synthesized by the superposition of 12 plane waves in Eq. (19) (see Sec. VIII). The pattern is similar to the numerically produced QC used in the previous section [see Figs. 1(b) and 7(c)], but in the current case, there are no ground-truth parameters and a true model. The objective pattern can only approximately be accessed by one of the models in Eq. (5). In contrast with the numerically produced pattern, estimated parameters do not reproduce *exactly* the same pattern as the objective pattern, and therefore the cost function is relatively high (Fig. 17 in the Appendix). Nevertheless, both two-length-scale and three-length-scale models reproduce DDQC patterns. The estimated parameters reproduce the inherent ratio of the length scales $q_0/q_1 = 1.948 \approx 2 \cos(\pi/12)$. The marginal likelihood indicates that the two length scale is favorable [Fig. 7(c)].

We summarize the results of a variety of patterns in Fig. 7. For each objective pattern, we can reproduce visually similar patterns, and the most probable number of length scales. In two-dimensional systems, stripe and hexagonal are the two most popular patterns under one length scale. These patterns are obtained from the conventional phase-field crystal model [30]. The marginal likelihood calculated in the BM-PDE indeed estimates that one length scale is favorable [Figs. 7(a) and 7(b)]. The difference between stripe and hexagonal patterns appears in the mean density $\bar{\psi}$. A quadratic nonlinear term is necessary to reproduce hexagonal patterns, and this implies that it appears at $|\bar{\psi}| \gg 0$ [30]. When $\bar{\psi} \simeq 0$, stripe patterns appear. The estimated values of the mean density are consistent with the results from the phase diagram reported in the literature [30]; we obtain the estimated mean density $\hat{\psi} \simeq -0.23$ and $\hat{\psi} = -0.05$ for the hexagonal and stripe objective patterns, respectively (see also Figs. 20 and 22 in the Appendix). The BM-PDE automatically estimates appropriate parameters from an artificially synthesized snapshot of the objective pattern.

In the estimation of models with different length scales in Fig. 7, some objective patterns, such as QCs, have a unique

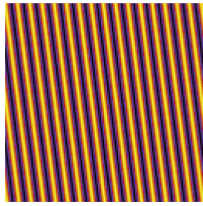
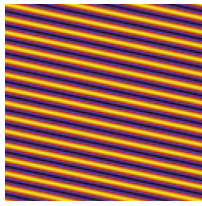
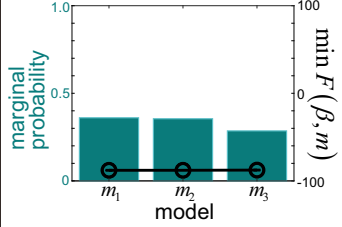
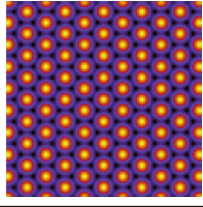
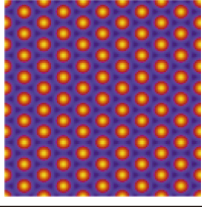
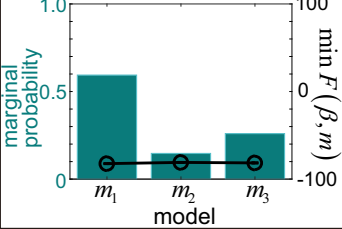
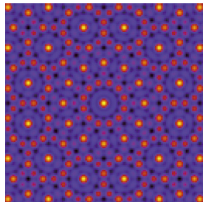
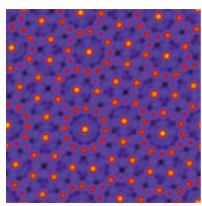
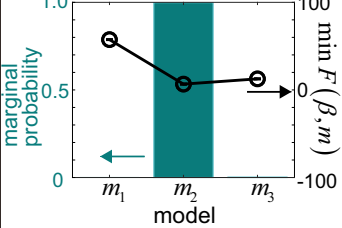
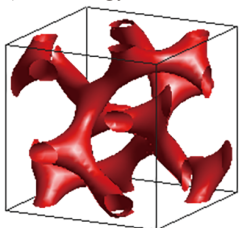
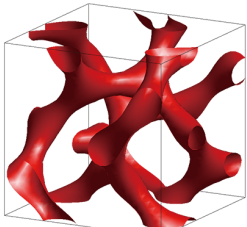
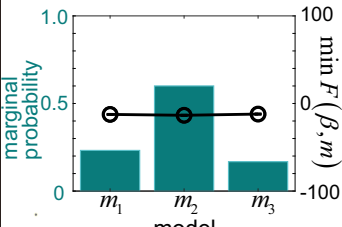
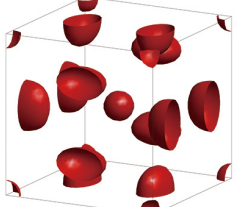
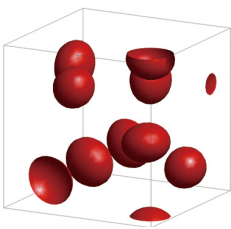
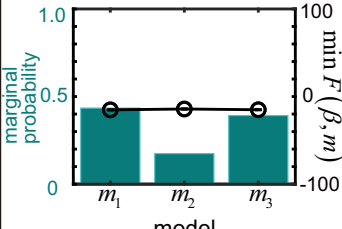
objective pattern	generated pattern from estimated parameters	length scale estimation from log marginal likelihood	estimated parameters
(a) stripe 			one length scale $\bar{\psi} = -0.0471 \pm 0.0351$ $q_0 = 0.704 \pm 0.103$
(b) hexagonal 			one length scale $\bar{\psi} = -0.226 \pm 0.054$ $q_0 = 0.503 \pm 0.004$
(c) quasi-crystal 			two length scale $\bar{\psi} = -0.748 \pm 0.019$ $q_0 / q_1 = 0.514 \pm 0.004$
(d) double gyroid 			two length scale $\bar{\psi} = -0.090 \pm 0.027$ $q_1 / q_0 = 0.419 \pm 0.0187$
(e) Frank Kasper A15 			one length scale $\bar{\psi} = -0.341 \pm 0.024$ $q_0 = 0.421 \pm 0.019$

FIG. 7. Summary of objective and estimated patterns for stripe (a), hexagonal (b), 12-fold symmetric QC (c), DG (d), and FK A15 (e). For each pattern, the free energy is evaluated for a model with one-, two-, and three-length scales. The best model is selected from the minimum of the log marginal likelihood (free energy). The model uncertainty is quantified by the marginal probability of each model obtained from the free energy marginalized for all temperatures. All the objective structures are synthesized from analytical functions by Eq. (19); they do not have ground truth in the model. The objective structure of FK A15 is synthesized based on the data in experiments in [26].

estimated model, whereas other patterns, such as stripes, have a similar probability among the models. The reason is that, first, the one-length-scale model cannot reproduce the QCs, and therefore, the marginal probability of m_1 in Fig. 7(c) should be zero. The QCs are available for limited models. The second reason is that for the stripe pattern, the log marginal likelihood has its minimum at the largest β , that is, the noise

associated with the difference between (synthesized) objective patterns and generated patterns is estimated to be small. The stripe pattern is less complex in that generated patterns using the PFC model are close to the synthesized objective pattern without ground truth, if not the same. Therefore, for the stripe pattern, the objective pattern has less error (see the discussions in Sec. IV). This is the case for whatever model we use

among the candidates, and thus, the marginal probability is less dependent on the models.

On the other hand, for the QCs, the optimal β is larger for the two-length-scale model [see Fig. 8(c)]. This arises from a different distribution of the cost function (see Appendix C 2). In Fig. 17(b) QCs appear from a broader range of β that is large enough, whereas, in Fig. 17(c), they disappear as β becomes slightly smaller. At the optimal β , the estimated parameters of the two-length-scale model are more likely to generate QCs than those of the three-length-scale model. Therefore, our estimation chooses the two-length-scale with a higher probability.

C. Double gyroid and Frank-Kasper patterns

We discuss an application of the BM-PDE to nontrivial three-dimensional patterns. The objective patterns are double gyroid (DG), shown in Fig. 7(d) and Frank-Kasper (FK) A15, shown in Fig. 7(e). The DG structure has two separate domains, each of which has degree-three vertices [10]. The DG structure was originally observed in self-assemblies of block copolymers, which is reproduced theoretically by the self-consistent field theory [31]. The simplified model under the random phase approximation is expressed by a PDE [32]. This model is a different class from PFC models. The DG structure was also reproduced by PFC models theoretically [33,34] and numerically [35]. These PFC models are called Landau-Brazovskii theory. FK A15 has been studied in metallic alloy and soft materials [9]. For example, the self-consistent field theory, designed to describe block copolymers, is capable of reproducing this pattern, but to our knowledge, this structure has not been reported within the framework of PFC.

In three dimensions, the order parameter may be defined similarly to that in two dimensions (see Sec. VIII B 2). Structural candidates in three dimensions are far richer than two-dimensional patterns. In fact, during the sampling, cylindrical patterns with a hexagonal lattice, BCC, and other patterns appear. These patterns can be stable under a certain region of parameter space [36]. The BM-PDE can reproduce both the DG and FK A15 patterns in all models ($m = m_1, m_2, m_3$). The generated structure is the same as the objective pattern, which is evident from the peaks in the Fourier space [Figs. 7(d) and 7(e) and see Appendix C 4]. The real space structures of the two patterns are overlapped by translation. The log marginal likelihood in Figs. 7(d) and 7(e) shows that the objective patterns of DG and FK A15 are expressed most likely by the two-length scale and one-length scale, respectively.

The DG structure has two length scales with their ratio of 1.15 (see also Appendix C 4). In the one-length-scale model, by taking $a_0 \gg 0$, several modes with slightly different length scales may become unstable so that the DG pattern is realized. In fact, it was discussed that DG appears not at $a_0 = \epsilon \approx 0$ with the small number ϵ , but at $a_0 \gtrsim O(1)$ [33,34]. In addition, this pattern appears between the stripe patterns and cylindrical (hexagonal) patterns in the phase diagram, namely $1 \gg |\bar{\psi}| \gtrsim 0$. On the other hand, in the model of block copolymers, the FK A15 phase appears near BCC patterns in the phase diagram [9]. This suggests that $|\bar{\psi}| \gg 0$ to obtain FK A15. The estimated mean density $\bar{\psi}$ well agrees with

this tendency. (Figs. 24 and 25 in the Appendix). The two models m_1 and m_2 have a comparable probability for both patterns. The DG pattern chooses m_2 possibly because m_2 has a wider range of spectrum amplitude in the phase diagram. The objective structure of FK A15 has ellipsoidal domains. The generated structures with the estimated parameters are also ellipsoidal (see Fig. 26 in the Appendix for quantitative results on sphericity). Such deformation is also reported for FK A15 structure made of block copolymers [37].

We should note that the objective structure of the FK A15 is taken from the experimental result [26]. From the X-ray scattering data, the position and amplitude of peaks in the Fourier space are identified [Fig. 4(b)]. We use the positions of the peaks and their amplitude in the data in [26].

IV. ROBUSTNESS AGAINST NOISE

We hypothesize that the robust estimation of the model and parameters for the objective pattern without ground truth originates from robustness against noise. To see this, we add Gaussian white noise to the objective pattern in Fig. 1(a) and study its effect on the estimation of the parameters. We focus on the two-length-scale model $m = m_2$ and the estimation of wavelength because this parameter has a narrow acceptable range. For the DDQC, the wave number is required to be close to 0.51, and otherwise the pattern cannot be generated because the mode coupling between two length scales does not occur. Figure 8(a) shows that even when the amplitude of the noise is about 20% of the amplitude of the density field of the objective pattern, the estimation of the parameter successfully works. Beyond the noise amplitude, the error of the estimated wave number becomes large, and the fraction of the patterns different from the objective pattern increases.

We compare the BM-PDE with the standard regression methods in which parameters μ are estimated by minimizing the cost function $\|\partial_t \psi - f(\psi; \mu)\|$ under an appropriate norm (see Sec. VIII D). This method relies on the state noise added in the dynamical equation, and, thus, does not give a good estimate for the measurement noise [38] (see also Appendix A). In fact, Fig. 8(a) shows that the estimated wave number deviates from ground truth even under 0.1% noise amplitude.

In contrast with the noiseless objective pattern, which has its log marginal likelihood $F[\beta]$ monotonically decreasing as β increases in Fig. 5(e), the log marginal likelihood for the objective pattern with noise has a minimum at the finite β as in Fig. 8(b). The minimum demonstrates the optimal noise corresponding to the noise amplitude in the objective pattern [25]. The noise level β^{-1} at the minimum of log marginal likelihood increases as the amplitude of noise in the objective pattern increases. At the large noise amplitude $\gtrsim 30\%$, the minimum β reaches a comparable value with the cost function at the gap between two structures in Fig. 5(b): QC and hexagonal patterns. Thus, DDQC can no longer be reproduced. Interestingly, the minimum of the log marginal likelihood is also observed for the objective pattern synthesized by the function of Eq. (19) [see Figs. 4(a) and 8(c)]. In this case, there is no ground-truth parameter that reproduces exactly the same pattern as the objective pattern. The optimal β at the minimum of $F[\beta]$ describes the deviation of the objective

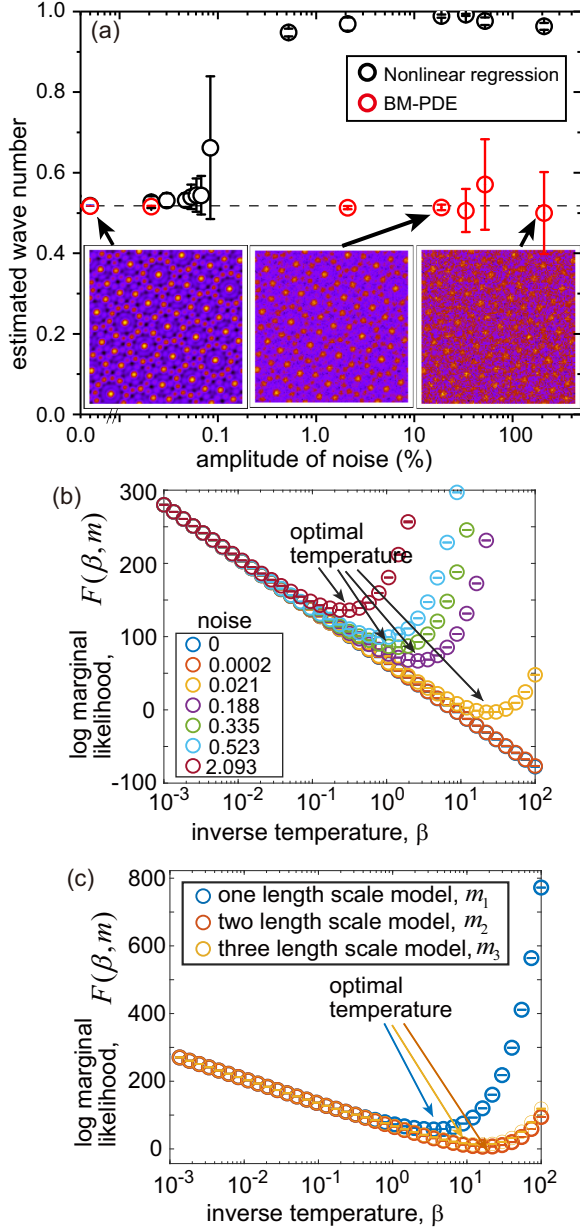


FIG. 8. (a) Estimated wave numbers for numerically generated QC by BM-PDE Eq. (2) (red points) and conventional regression method (black points) under Gaussian white noise added on the objective pattern [see Eq. (43)]. Noise amplitude with respect to the variance of the noiseless pattern is defined as $\sigma^2/\text{Var}[\psi^*]$ where σ^2 is the variance of added noise. The noise amplitude is shown in %. The horizontal dashed line indicates the ground truth of the wave number $q_1 = 0.51764$. The objective patterns under the different noise amplitude are shown in the insets. (b) The log marginal likelihood at each inverse temperature β for the objective pattern generated by the numerical simulation with noise corresponding to (a). The minimums of the log marginal likelihood are shown by arrows. (c) The free energy at each inverse temperature β in REMC under the models m_1 , m_2 , and m_3 for the objective pattern synthesized by the function of Eq. (19).

pattern from the patterns that the models can reach. Without the Bayesian inference, we cannot successfully estimate the uncertainty, which plays a similar role to noise.

The BM-PDE works for the damaged objective structure in which some spatial information is lost. We demonstrate the estimation for the damaged data in Appendix D. The BM-PDE does not rely on the time and spatial derivatives of the objective structure. Therefore, the damage in the data can be randomly distributed, namely, the information of neighboring data points is not necessary. This is not the case if the information of spatial derivatives is necessary for the estimation.

V. GENERALIZATION OF BM-PDE

In this section, we discuss the possible extension of our method. We have considered the family of phase-field crystal models with different length scales, and demonstrated in the previous sections the estimation of the parameters in the linear operators. Our approach is not limited to the choice of this family. In fact, we can arbitrarily choose a family of models. In this section, we consider other classes of models to show the generality of our method.

The objective pattern in Secs. III and IV was uniform in space. This is because the order parameter in (2) is defined for each structure. We will demonstrate that with certain pre-processing, we may estimate a model for an inhomogeneous pattern with defects.

A. Expansion of the linear operator by orthogonal bases

In [39,40] it was proposed that the directed correlation function is expanded by the Hermite polynomials. This method is appealing because the polynomial forms orthogonal bases, and therefore, its coefficient must be independent. In our formula this method corresponds to different parametrization of the linear operator $\mathcal{L}_\mu^{(m)}$ in Eq. (5), or \mathcal{L}_k in Eq. (15), by the Hermite polynomials as Eq. (6). Instead of (16), we may expand the linear operator in (11) and (15) as

$$\mathcal{L}(k) = -a_0 k^2 + k^2 \sum_{n=1}^{n_{\max}} a_n \frac{e^{-\frac{1}{2} \frac{k}{k_0}}}{(2^n n! \sqrt{\pi})^{1/2}} H_n(k/k_0), \quad (6)$$

where $H_n(x)$ is the Hermite polynomial. The reference wave number is denoted by k_0 . We set to be $k_0 = 0.25$, and $n_{\max} = 10$. We note that the first term with a_0 appears from the conserved system. The linear operator in the Fourier space diverges as $\mathcal{L}(k \rightarrow \infty) \sim -k^2$, which make sure that fluctuations at the higher wave numbers decay quickly. In the method shown in the previous section, the scale of the divergence is different, for example, $\mathcal{L}(k \rightarrow \infty) \sim -k^{10}$, but the qualitative shape of the linear spectrum of the two methods is the same [see Figs. 5(g) and 9(d)].

For the objective pattern shown in Figs. 9(a) and 9(b), we estimate the parameters in Eq. (6). The estimated parameters are shown in Fig. 9(e). Most of the parameters are close to zero, and there are two finite modes. Using the estimate parameters, we may generate QCs, as shown in Fig. 9(c). The estimated linear spectrum is shown in Fig. 9(d). The qualitative shape of the linear spectrum is the same as Fig. 5(g) used in the previous section. In fact, there are two peaks whose ratio is $1/[2 \cos(\pi/12)]$. The ratio is consistent with the result in the previous sections.

The benefit of the expansion by the basis functions is that there is no need to consider different models with the

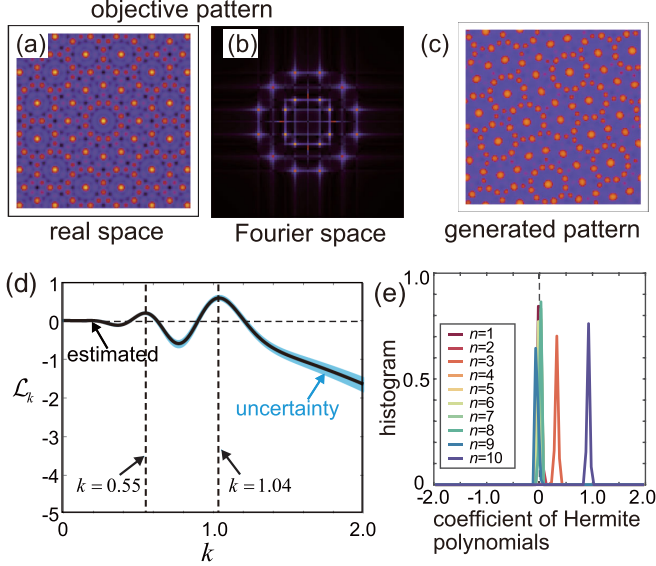


FIG. 9. (a), (b) The objective pattern of the synthesized dodecagonal QC (a) and the intensity of its Fourier transformation (b). (c)–(e) The estimation for the model using Hermite polynomials. (c) The generated pattern from the estimated parameters for the model using the Hermite polynomials. (d) The linear spectrum (black lines) as a function of wave number from the estimated parameters for the Hermite polynomials. The uncertainty of the estimation is shown by the range in light blue. The two solid black lines show the wave numbers at the peaks. (e) The histogram of the estimated coefficients of the Hermite polynomials.

different number of length scales. The arbitrary number of length scales can be expressed by the appropriate superposition of the bases. The disadvantage of this method is that the position of the characteristic wavelength, characterized by the peak in the linear spectrum, cannot be imposed as a parameter.

B. Design of nonlinear terms

Our approach is not limited to the estimation of linear parts in the model. Instead of Eq. (5), we may consider the two-component model described by the two density fields $\psi_1(\mathbf{x})$ and $\psi_2(\mathbf{x})$. Both the fields have their own single length scales q_1 for $\psi_1(\mathbf{x})$ and q_2 for $\psi_2(\mathbf{x})$, which may be different from each other. The model is expressed by

$$\partial_t \psi_1(\mathbf{x}) = \mathcal{L}_{\mu_1} \psi_1(\mathbf{x}) + \Delta \mathcal{N}_1[\psi_1, \psi_2], \quad (7)$$

$$\partial_t \psi_2(\mathbf{x}) = \mathcal{L}_{\mu_2} \psi_2(\mathbf{x}) + \Delta \mathcal{N}_2[\psi_1, \psi_2], \quad (8)$$

where the linear operators \mathcal{L}_{μ_1} and \mathcal{L}_{μ_2} are the same as the operator of the one-length-scale model in Sec. III [see Fig. 5(f)]. The position and height of a peak in \mathcal{L}_{μ_1} and \mathcal{L}_{μ_2} are different. The nonlinear terms $\mathcal{N}_1[\psi_1, \psi_2]$ and $\mathcal{N}_2[\psi_1, \psi_2]$ are given by polynomials of ψ_1 and ψ_2 , and are parametrized by b_n as

$$\begin{aligned} \mathcal{N}_1[\psi_1, \psi_2] = & \psi_1^3 - 2b_1\psi_1\psi_2 - b_2\psi_2^2 - b_3\psi_2^3 \\ & - 2b_4\psi_1\psi_2^2 - 3b_5\psi_1^2\psi_2, \end{aligned} \quad (9)$$

$$\begin{aligned} \mathcal{N}_2[\psi_1, \psi_2] = & \psi_2^3 - b_1\psi_1^2 - 2b_2\psi_1\psi_2 - 3b_3\psi_1\psi_2^2 \\ & - 2b_4\psi_1^2\psi_2 - b_5\psi_1^3. \end{aligned} \quad (10)$$

The first terms in (9) and (10) are the same as the families of models in the previous sections. For simplicity, we assume the gradient structure of the model, that is, there exists the free energy functional F so that the model is expressed by $\partial_t \psi_{1,2} = \Delta \frac{\delta F}{\delta \psi_{1,2}}$. The free energy functional contains the terms of coupling between ψ_1 and ψ_2 up to the fourth order of the polynomials. In this model, the parameters are the system size in x and y directions, mean density $\bar{\psi}$ which is assumed to be the same for the two fields, the two wave numbers q_1 and q_2 , and their corresponding heights of the linear spectrum a_1 in \mathcal{L}_{μ_1} and a_2 in \mathcal{L}_{μ_2} , and the coefficients b_1, \dots, b_5 of nonlinear coupling terms. Here q_1 is fixed to be 1 without loss of generality. We use the periodic boundary conditions.

The goal of the estimation in this model is to specify the most relevant nonlinear term among the candidates. To do this, we use sparse Bayesian formula [41]. Instead of the uniform prior distribution, we use the prior distribution $\exp(-\kappa \sum_{i=1}^5 |b_i|)$ for the coefficients of the nonlinear terms. When $\kappa > 0$, due to the prior distribution, finite values of the coefficients of nonlinear terms are penalized in the posterior distribution. As a result, only a few of b_i becomes nonzero. We have performed the parameter estimation both with ($\kappa = 5.0$) and without ($\kappa = 0$) sparsity. Without sparsity ($\kappa = 0$), the prior distribution is uniform and does not penalize the coefficients of nonlinear terms.

We use the same objective pattern shown in Figs. 9(a) and 9(b) as used in Fig. 7(c), and then perform the parameter estimation. The generated pattern from the estimated parameters is shown in Fig. 10(a). The pattern has dodecagonal symmetry as in the objective pattern. The posterior distributions of the parameters are shown in Figs. 10(c) and 10(d). Without sparsity ($\kappa = 0$), all the nonlinear coupling terms contribute to the estimation. On the other hand, with sparsity ($\kappa = 5.0$), only the term ψ_2^3 in the $\mathcal{N}_1[\psi_1, \psi_2]$ is nonzero, and other coefficients become close to zero. Therefore, the term ψ_2^3 is the most relevant nonlinear term.

Surprisingly, the estimated wave number is $q_2 \approx 1/\sqrt{2}$ with respect to $q_1 = 1$. This ratio q_2/q_1 is different from that in the two-length-scale model used in the previous section ($q_0/q_1 \approx 1/[2 \cos(\pi/12)]$). This can be understood by a quadrilateral in the Fourier space shown in Fig. 10(b). In the linear spectrum, the modes at the wave numbers q_1 and q_2 dominate. The nonlinear terms affect the pattern selection when triangle [second-order terms of $\psi_{1,2}$ in Eqs. (9) and (10)] or quadrilateral (third-order terms of $\psi_{1,2}$) appear to connect the spots in Fig. 10(b). In the previous section, the nonlinear term $\bar{\psi} \psi^2$ determines the stability of the dodecagonal QC. To draw a triangle to connect the spots in the inner and outer circles in Fig. 10(b), the angle of the triangle should be $\pi/6$. This is the reason why the ratio $q_0/q_1 = 1/[2 \cos(\pi/12)]$ appears [14].

In the two-component system given by Eqs. (9) and (10), the dominant nonlinear term is third-order ψ_2^3 in $\mathcal{N}_1[\psi_1, \psi_2]$ ($\psi_1\psi_2^2$ in $\mathcal{N}_2[\psi_1, \psi_2]$). Therefore, a quadrilateral consists of four wave vectors $\mathbf{q}_1, \dots, \mathbf{q}_4$ make a contribution when $\mathbf{q}_1 + \mathbf{q}_2 + \mathbf{q}_3 + \mathbf{q}_4 = 0$. From the geometry shown in Fig. 10(b), the ratio between the two wave numbers must be $1/\sqrt{2}$. To our knowledge, this is the first dodecagonal QC generated from PDEs with the ratio $1/\sqrt{2}$.

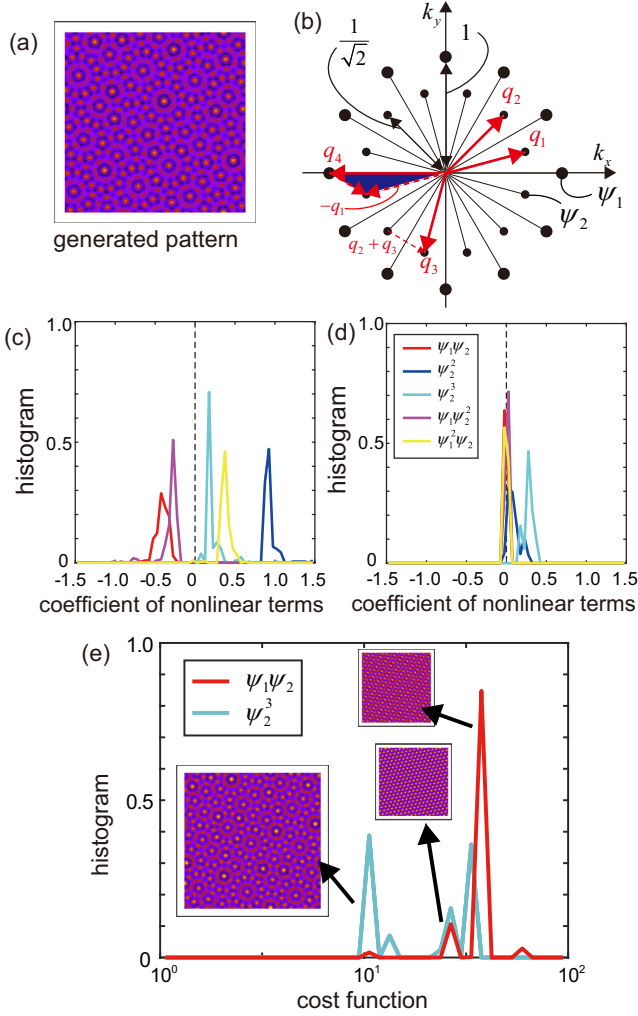


FIG. 10. The estimation for the two-component model. (a) The generated pattern from the estimated parameters for the two-component model. (b) The schematic illustration of the intensity in the Fourier space of the estimated dodecagonal QC. The four vectors \mathbf{q}_1 – \mathbf{q}_4 are the example of the dominant coupling term. The blue triangle corresponds to the three vectors \mathbf{q}_1 , \mathbf{q}_4 , and $\mathbf{q}_2 + \mathbf{q}_3$. (c, d) The estimated parameters of the nonlinear coupling terms between the two-component ψ_1 and ψ_2 with (d) and without (c) sparsity. The color legend indicates the nonlinear terms for ψ_1 . (e) The histogram of the cost function under the model only with $\psi_1\psi_2$ (black) and ψ_2^3 (red) in Eq. (9).

Figure 10(e) shows the histogram of the cost function of the generated patterns from the models only with the nonzero coefficient of the coupling term b_1 or b_5 in Eqs. (9) and (10). The parameters are fixed in each model, but each stationary pattern is generated from a different initial condition. Both models can generate the DDQC patterns in two dimensions. The cost function of the DDQC is $E \simeq 10$. The histogram shows that the model with $b_5 \neq 0$ is more stable in the sense that the DDQC is generated from a wide range of initial conditions. Even though the model with $b_2 \neq 0$ can generate the DDQC, most of the initial conditions lead to the hexagonal patterns. The result demonstrates that our method may choose parameters so that the model can reproduce the objective pattern from a wide range of initial conditions.

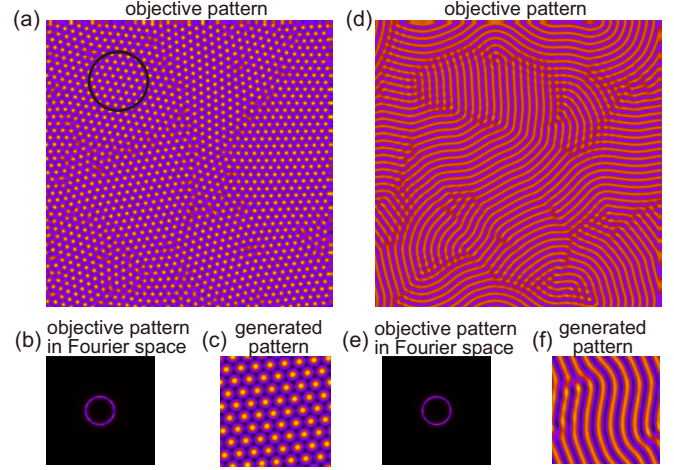


FIG. 11. The objective (a), (d) and generated (c), (f) patterns including defects for hexagonal (a)–(c) and stripe (d)–(f) structures. The objective structures in Fourier space are also shown in (b), (e). The circle in (a) indicates the size of the Gaussian filter.

C. Objective patterns with inhomogeneity of defects

We consider the objective patterns including defects, such as dislocations and grain boundaries. The objective structures are numerically generated for a larger system size $N = 512$ using the one-length-scale model. The objective hexagonal and stripe patterns are shown in Fig. 11. These patterns show polycrystalline structures. The structures in Fourier space show a ring at the characteristic length scale of the model. Six sharp peaks for hexagonal and two peaks for stripe are not visible from the objective patterns. The order parameter computed from the whole structure cannot result in a successful estimation. Therefore, we use a Gaussian filter with the width $x = 30$ centered at each mesh point in space, compute local order parameters, and take an average over the different spatial points.

The generated patterns using the estimate parameters of the one-length-scale model successfully reproduce hexagonal and stripe structures as in Figs. 11(c) and 11(f). The ground-truth parameters are $a_0^* = 0.1$ for both hexagonal and stripe, and $\bar{\psi}^* = -0.3$ for hexagonal and $\bar{\psi}^* = -0.01$ for stripe. The estimated parameters are $(\hat{a}_0, \hat{\psi}) = (0.0731, -0.1731)$ for hexagonal and $(\hat{a}_0, \hat{\psi}) = (0.0676, -0.0740)$, which are reasonably close to the ground truth.

VI. THREE-DIMENSIONAL DODECAGONAL PATTERN EXPRESSED BY A PDE

Using the estimated parameters, we may investigate additional physical insights of the model. To demonstrate it, we consider DDQC in three dimensions [Figs. 12(b) and 12(c)]. Although the icosahedral QCs have been found in the PDE model using two-length-scale PFC [42], other three-dimensional structures are largely unexplored [43]. The axisymmetric dodecagonal structure has 12-fold symmetry along the axis (c axis) of axisymmetry [Figs. 12(b) and 12(c)].

To generate this structure, larger system size is required, but it is computationally demanding. Moreover, as we will

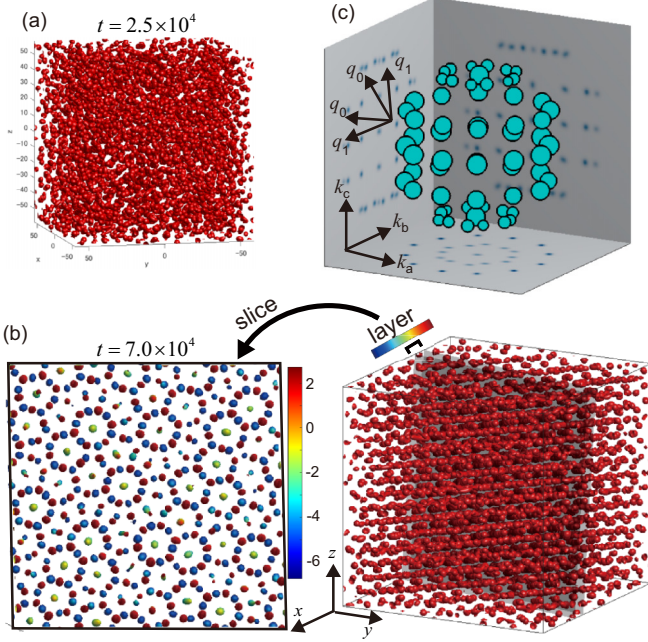


FIG. 12. Structures of the DDQC and their fluctuations. (a) Metastable micelle-like structure. (b) The DDQC (right) and its slice perpendicular to the axis of 12-fold symmetry in the region of the box in black. The color indicates the position along the axis. Therefore, the red and blue domains in the dodecagonal ring are located in different layers. The yellow domains are located between the two layers. (c) The DDQC in the Fourier space. The points size are proportional to the amplitude of the peak of $|\tilde{\psi}_{\mathbf{k}}|$. The structure has 12-fold symmetry around the axis of k_c , and k_a and k_b are perpendicular axes to k_c . The projection of $|\tilde{\psi}_{\mathbf{k}}|$ on each plane is also shown. Two lengths of wave vectors q_0 and q_1 are shown. (d) A local structural change in the DDQC. The slice perpendicular to the symmetry axis is shown at two different times. The insets show an enlarged structure along the symmetry axis.

discuss later, we found that the kinetics of the formation of the DDQC in three dimensions is fundamentally different from other structures. Because the FK A15 is known as the simplest approximant of the DDQC [44], it is fair to assume that the DDQC may appear near the estimated parameters of the FK A15. Therefore, we focus on the estimated parameters and solve the estimated models with larger systems size in a longer timescale.

The structure discussed in the previous sections, including the DDQC in two dimensions, may appear from a random initial condition by quenching. The random initial structure reorganizes to the desired pattern without noise. In addition, the system reaches its stationary state before $t = 10^2 - 10^3$. Note that the timescale $t = 1$ corresponds to the relaxation time of the structure when $a_i = 1$. The formation of the DDQC in three dimensions is entirely different from the two-dimensional structure. It requires not only a larger system size but also an annealing process. Without the state noise, the initial random structure results in a micelle-like isotropic structure [Fig. 12(a)], which is frozen. This is particularly the case for the system size $N > 64$. To overcome the trap at the metastable structure, we add the state noise into the PDE model (see Sec. VIII A 2), and anneal the structure by

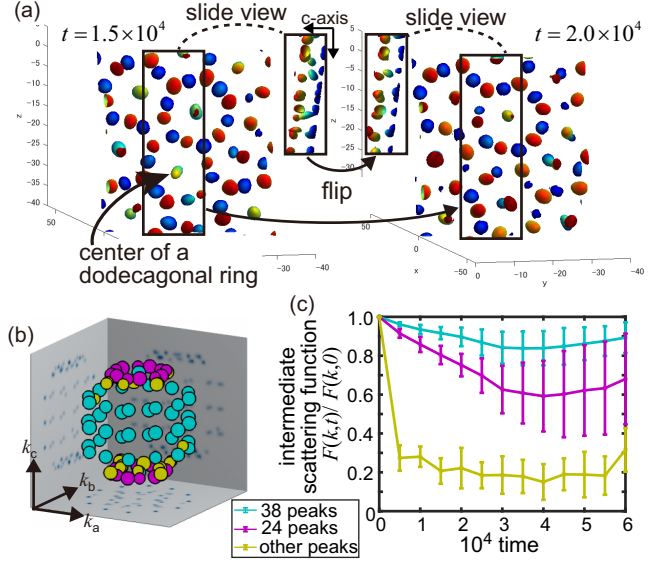


FIG. 13. Fluctuations of the DDQC. In (b), the peaks in the Fourier space are divided into three groups by their amplitude. The 38 cyan peaks have the largest amplitude, and the 24 magenta peaks have the next largest. The other peaks are shown in yellow. The 38 cyan peaks and the 24 magenta peaks form the main peaks in Fig. 12(c). The normalized intermediate scattering function is shown for each group of peaks in (c). Each color corresponds to $F(k, t)$ of the peaks in the same color. The points are the mean values in each group, and the error bars are the standard deviation in each group.

decreasing the noise amplitude. We found that the annealing process is very slow; the DDQC appear around $t \gtrsim 5 \times 10^4$.

Figure 12(b) shows the generated structure for the system size $N = 256$. The generated structure is periodic along the c axis. We denote the other two axes as the a axis and the b axis, and in the Fourier space, these axes are denoted by k_a , k_b , and k_c . As in Fig. 12(c), the structure has 12-fold symmetry in the plane perpendicular to the c axis. The 12-fold symmetry arises from the layered structure in which two layers with the hexagonal symmetry rotate $\pi/6$ with each other [Fig. 12(b)]. The center of the dodecagonal ring is located between the two layers. There are 62 main peaks in the Fourier space. Along the symmetry axis, there are five 12-fold rings and two spots along the c axis. From the peaks in the Fourier space shown in Fig. 12(c), it is evident that there are two length scales q_0 and q_1 whose ratio is $q_1/q_0 = 2/\sqrt{5}$. This value is, in fact, close to the estimated ratio of parameters $\hat{q}_1/\hat{q}_0 = 0.889$. The layered structure with the dodecagonal symmetry has been reported in particle-based simulations [43,45]. We have tried both one-length-scale and two-length-scale models with the estimated parameters for the FK A15 structure. All the results shown in Fig. 12 are by the two-length-scale model. The one-length-scale model cannot reproduce the QC. This is natural because more than two length scales are necessary to form QCs.

Using the obtained stable DDQC, we also study their fluctuations under the fixed amplitude of the noise. Figure 13(a) demonstrates the reconstruction of a dodecagonal ring. This process is due to the change of the position of the center along the c axis to the ring [see the insets of Fig. 13(a)]. This reconstruction has also been observed in a particle-

based model [46]. The fluctuation is characterized by the intermediate scattering function, $F(\mathbf{k}, t) = \langle \psi_{\mathbf{k}}(0) \psi_{\mathbf{k}}^{\dagger}(t) \rangle$ for the wave vectors corresponding to the peaks in the Fourier space [Fig. 13(b)]. Here \dagger indicates complex conjugate. The relaxation time of $F(\mathbf{k}, t)$ indicates the diffusive timescale of the structure at \mathbf{k} [Fig. 13(c)]. The peaks other than the main peaks of the DDQC shown in Fig. 12(c) decay quickly in a short timescale of $t \sim 10^3$. Among the main peaks, the larger peaks [cyan in Fig. 13(b)] are stable against noise even in the longer timescale of $t \gtrsim 10^4$, whereas the smaller peaks [magenta in Fig. 13(b)] show diffusive behavior. This slower diffusion at $t \gtrsim 10^4$ corresponds to phason flips. In fact, from the inverse Fourier transformation, we found the peaks with the intermediate amplitude [magenta in Fig. 13(b)] correspond to the centres of the dodecagonal rings. We stress that these analyses of fluctuations are achieved by the estimation of the *dynamical* PDE from the stationary objective pattern in the previous section.

VII. DISCUSSION AND CONCLUSION

Automatic discovery of a model equation is a recent key topic in data-driven science [47–49], and several methods have been proposed for PDEs of time-series data [38,39,50,51]. These approaches are designed to handle time-series data and cannot be applied to estimation for a stationary pattern, because it does not have information about trajectories from their initial conditions. In addition, the previous studies focus on the estimation for the data with ground truth. The data are generated typically from numerical results of a known model with known parameters, and then those parameters are estimated from the data. Therefore, the two issues, the estimation of a dynamical equation from stationary data and data without ground truth, remain elusive even in a technical aspect of model discovery (see also Appendix A 4).

In this work we propose the inverse problem of equation discovery for a stationary pattern without ground truth. We successfully estimate the best PDE for complex patterns. Our approach has several key features compared with the previous inference methods; it is designed for the data of a stationary pattern without ground truth. In addition, the method has generality to apply a wide class of PDE models and objective patterns. Here we summarize these features.

The estimation for stationary data: The previous approaches to estimate the PDE from data are the estimation of *dynamical equations* from *dynamical nonstationary data*. In this case, the data has information about trajectories of dynamics. In this work, we may estimate *dynamical equations* from *stationary data*, which is one snapshot of the stationary pattern. Such estimation may look impossible for two reasons: One is the lack of information on transient dynamics. The second reason is that in pattern formation, the dynamical equation is nonlinear, and therefore, $f(\psi) = 0$ has multiple stationary solutions. Only one solution from the multiple solutions is not enough to reconstruct $f(\psi)$ uniquely. We take the inverse problem for the stationary pattern as an estimation of a *dynamical* PDE from marginalized initial conditions (see Appendix A 4 for comparison to other methods). The stable structure should be generated from a wide range

of initial conditions. In the BM-PDE, the estimation of the model reproduces an objective pattern as a stable solution by marginalizing the initial conditions as in Eq. (3) [see also Figs. 6(a) and 10(e) for a demonstration]. This is in contrast with the method to estimate a dynamical equation from time-series data, for example, using data assimilation, in which a single initial condition is estimated from data [19].

The order parameters play an important role in the BM-PDE. The order parameters can extract symmetries of the pattern and identify the two patterns that are generated with the same parameters but the different initial conditions. This identification is a necessary step to marginalise the initial patterns.

Our method is designed so that the estimated model reproduces a pattern close to the objective pattern from as many initial conditions as possible. This is particularly the case when generated patterns in two models are equally close to the objective pattern. When the true model has several coexisting patterns, the estimated model may be biased in the sense that we choose a model that prefers one of the several coexisting patterns. Nevertheless, we believe our method is natural because we do have only one pattern as data. A possible extension, when a true model has several coexisting patterns and we have data of all the patterns, is to revise the cost function to measure the distance from generated patterns of estimated parameters to all the objective patterns.

The estimation without ground truth: The BM-PDE works not only for a numerically produced pattern, which has ground truth of parameters, but also for an synthesized pattern by superposition of plane waves of Eq. (19) (see Sec. II B). In the latter case, our family of models does not include ground truth. To demonstrate the ability to estimate the model for the structure without ground truth, we have used experimental data [Fig. 4(b)]. Even in this case, we have successfully estimated the best PDE model and its parameters. This is because the BM-PDE quantifies the uncertainty by estimating the optimal noise amplitude, and is robust against noise. A promising direction is to use our method for the estimation of a PDE for a pattern obtained by molecular dynamics simulations.

Our method may be used to find a structure that we do not know how to reproduce by a PDE model. We have found the three-dimensional DDQC from the estimated parameters for its approximant. To our knowledge, this is the first three-dimensional DDQC found in the PDE model.

The generality of BM-PDE: In the BM-PDE, the choice of models and the cost function are independent. Therefore, we may replace the models with any other PDE models. To demonstrate this feature, we have shown in Sec. V the estimation of parameters not only in the linear operator but also in nonlinear terms. The choice of the objective pattern is also flexible. By comparing two patterns in the space of the order parameter, we can estimate the model that can reproduce a similar pattern in the sense that it has the same symmetry as the objective pattern. Due to this feature, the BM-PDE works for the noisy objective pattern.

VIII. METHODS

The main framework of BM-PDE consists of three parts: a family of PDE models, characterization of a structure with

order parameter, and statistical inference. Here we summarize their basic steps.

A. PDE models

We consider a pattern described by a scalar density field $\psi(\mathbf{x})$ in a box with a periodic boundary condition, and $\mathbf{x} \in [-L_x/2, L_x/2] \times [-L_y/2, L_y/2]$ in two dimensions and $\mathbf{x} \in [-L_x/2, L_x/2] \times [-L_y/2, L_y/2] \times [-L_z/2, L_z/2]$ in three dimensions. The density field $\psi(\mathbf{x})$ is obtained by an unknown model of a partial differential equation. In this study, we focus on a family of the phase-field crystal (PFC) models given by

$$\partial_t \psi = \mathcal{L}\psi + \Delta\psi^3 \quad (11)$$

with the linear operator denoted by \mathcal{L} and the nonlinear term is given in the second term. The family is constructed by modifying the linear operator \mathcal{L} so that the system has one or more length scales. The simplest PFC model is a conserved version of the SH equation [5]. The equation is given by

$$\partial_t \psi = \Delta[-\epsilon\psi + (q_0 + \Delta)^2\psi + \psi^3], \quad (12)$$

where the total mass is conserved as

$$\bar{\psi} = \frac{1}{V} \int \psi(\mathbf{x}) d\mathbf{x}. \quad (13)$$

Here V is the total volume (area in two dimensions) in the system, and q_0 is a characteristic wave number corresponding to the length scale $2\pi/q_0$. The parameter ϵ controls whether the uniform state $\psi(\mathbf{x}) = \bar{\psi}$ is stable $\epsilon \lesssim 0$ or unstable $\epsilon \gtrsim 0$. The precise value of the threshold is dependent on other parameters and a type of patterns. The PFC equation reproduces a stripe (also called lamellar or smectic) and hexagonal patterns in two dimensions [30,52], and a lamellar, hexagonal cylinder, BCC, and hexagonal closed packing patterns [35,36]. A finite mean density $\bar{\psi}$ plays a role as the quadratic nonlinear term in Eq. (12). This can be seen by subtracting the mean density as $\psi \rightarrow \psi + \bar{\psi}$ in Eq. (12) as

$$\partial_t \psi = \Delta[-\epsilon + 3\bar{\psi}^2]\psi + (q_0 + \Delta)^2\psi + 3\bar{\psi}\psi^2 + \psi^3. \quad (14)$$

The PFC equation has a single characteristic length at which a real part of the eigenvalue is positive (or, at least, negative but close to zero). The linear spectrum is shown in Fig. 15 in the Appendix. To extend Eq. (12) for the arbitrary number of length scales, we use m characteristic wave numbers, q_0, q_1, \dots, q_{m-1} , and the values of the spectral, a_0, a_1, \dots, a_{m-1} , at the wave number $k = q_i$. A family of our models is conveniently described by the equations for the Fourier transform of the density field, that is, $\tilde{\psi}(\mathbf{k}) = \mathcal{F}[\psi(\mathbf{x})]$. Our models corresponding to Eq. (11) are given by

$$\partial_t \tilde{\psi}(\mathbf{k}) = \mathcal{L}_k \tilde{\psi}(\mathbf{k}) + \mathcal{F}[\Delta\psi(\mathbf{x})^3], \quad (15)$$

and its linear operator is expressed in the Fourier space as

$$\mathcal{L}_k = a_0 S_0(k) + a_1 S_1(k) + \dots + a_{m-1} S_{m-1}(k) + k^2(q_0^2 - k^2)^2(q_1^2 - k^2)^2 \dots (q_{m-1}^2 - k^2)^2. \quad (16)$$

The function $S_i(k)$ for $i \in [0, m-1]$ is chosen so that the coefficient a_i corresponds to the peak as a function of k of \mathcal{L}_k at $k = q_i$ [see Figs. 5(f)–5(h)]. Since we may freely choose a unit of length scale, we fix to be $q_0 = 1$ when $i \geq 2$ in $m = m_i$.

This implies that we impose the length scale $2\pi/q_i$. The concrete form of $S_i(k)$ is shown in Appendix B 1. The parameter s_0 describes the sharpness of the peaks. To make the spectrum sharp enough, we chose the parameter to be $s_0 = 100$ for the one-length and two-length models, and $s_0 = 2000$ for the three-length model. Both a_i and q_i are chosen as parameters to be optimized. We have other parameters such as the system size L in each direction and the mean density of a pattern $\bar{\psi}$. In this study, we use the periodic boundary condition. A set of parameters is thus $\mu = \{L_\alpha, \bar{\psi}, a_0, q_0, \dots, a_{l-1}, q_{l-1}\}$ with $\alpha \in [1, \dots, d]$ in space dimension d .

In order to make a pattern, higher-order spatial derivatives are necessary. Polynomial expansion in terms of the wave number k instead of Eq. (16) may be available to make several length scales (see Fig. 15). Nevertheless, it suffers from the large value of the coefficient of each order in the polynomials because we may not use prior distribution to confine the parameter space (see also Appendix B 2).

1. Numerical simulations of PDE models

Numerical simulations of the PDEs are performed using the pseudospectral method in which the linear terms are computed in the Fourier space, and the nonlinear terms are computed in real space. Since our PDEs contain higher-order derivatives, we use the operator-splitting method [53,54]. Both real and Fourier spaces are discretized into N^d meshes in d -dimensional space. Instead of changing the system size L_i in each dimension $i \in [1, d]$ under the periodic domain, we change the mesh size dx_i so that the system size becomes $L_i = (N-1)dx_i$. In the estimation of this study, we estimate both $dx_1 (= dx)$ and $dx_2 (= dy)$ for numerically generated DDQC in two dimensions, but we assume $dx_1 = dx_2 = dx_3$ for other cases.

The number of mesh points is fixed to be $N = 128$ in two dimensions and $N = 32$ in three dimensions. The larger N is better in terms of accuracy, but computational time is scaled roughly as $O(N^d)$. In REMC, we need to simulate it in N_{rep} replicas, and therefore, the choice of N is made by the balance between accuracy and realistic computational time. In addition, the larger system size suffers from longer relaxation time and a higher probability that topological defects such that dislocations and disclinations appear. We set the total number of steps to be 10^5 with a time step $dt = 0.01$. We confirmed this is enough to obtain the stationary patterns studied in this work, but may be changed depending on the pattern of interest. Note that statistical inference in this study is entirely independent of the algorithm of numerical simulations solving a PDE. An efficient algorithm would improve the performance of estimation, and one may replace the numerical scheme suit for one's purpose.

2. Formation of three-dimensional DDQC

To generate the DDQC in three dimensions, we have to add noise in Eq. (5),

$$\partial_t \psi(\mathbf{x}, t) = \mathcal{L}_\mu^{(m)} \psi(\mathbf{x}, t) + \mathcal{N}[\psi(\mathbf{x}, t)] + \gamma(t)\xi(\mathbf{x}, t). \quad (17)$$

Note that the noise is state noise, and has nothing to do with the observation noise. In the annealing process, the amplitude of the noise $\gamma(t)$ is decreased in time. The annealing schedule

is chosen as $\gamma(t) = 0.198 / \log(t + \tau_a)$ with $\tau_a = 10^4$. This choice ensures $\gamma(t) \sim 1 / \log t$ for large t [55]. The Gaussian white noise with the zero mean $\langle \xi(\mathbf{x}, t) \rangle = 0$ and its variance

$$\langle \xi(\mathbf{x}, t) \xi(\mathbf{x}', t') \rangle = -2\Delta \delta(\mathbf{x} - \mathbf{x}') \delta(t - t') \quad (18)$$

is used. The Laplacian in Eq. (18) ensures the conservation of the density and ensures to reach the equilibrium state. We may also use other statistical properties of the noise as long as the density conservation is ensured. For example, we have tested $\langle \xi(\mathbf{x}, t) \xi(\mathbf{x}', t') \rangle = 2\delta(\mathbf{x} - \mathbf{x}') \delta(t - t')$ with $\int \xi(\mathbf{x}', t') d\mathbf{x} = 0$. We have confirmed in all cases, the DDQC appears when the amplitude of the noise γ is decreased.

Fluctuation of the DDQC is studied by fixed γ . We set it to be $\gamma = 0.171$. Note that the DDQC is destroyed for the noise with the amplitude of $\gamma^2 = 0.198$.

B. Objective patterns and characterization of patterns

1. Objective pattern

An objective pattern $\psi^*(\mathbf{x})$ is prepared in two ways. One is a numerical solution of Eq. (5) for given parameters under a given model. The resulting pattern is numerically transformed into the Fourier space, and then the order parameter Ψ_l^* is calculated.

The second way is completely independent of the models. The objective pattern is expressed as a density field $\psi^*(\mathbf{x})$ that is a superposition of the cosine function. The simplest case is a stripe pattern, which is described only by one wave in one direction in a two-dimensional space. Similarly, a hexagonal pattern is expressed by two-dimensional waves in three directions. The objective pattern is expressed as

$$\psi^*(\mathbf{x}) = \sum_i b_i \cos(\mathbf{q}_i^* \cdot \mathbf{x}), \quad (19)$$

where the wave vectors \mathbf{q}_i^* are chosen at the position appropriate to express symmetries of the objective pattern. The amplitude of each mode b_i is also chosen properly. We numerically make the Fourier transform of Eq. (19) to obtain $\tilde{\psi}^*(\mathbf{k}) = \mathcal{F}[\psi^*(\mathbf{x})]$ and calculate the Fourier spectrum $|\tilde{\psi}^*(\mathbf{k})|$ from which we obtain the order parameter. The Fourier transform of Eq. (19) defined in the infinite domain is expressed by the superposition of the delta function at the position of \mathbf{q}_i^* . Nevertheless, the numerical Fourier transform in the bounded domain results in peaks smeared around \mathbf{q}_i^* . To remove the artifact, we set $|\psi^*(\mathbf{k})| = 0$ except for the region $|\psi^*(\mathbf{k})| > \alpha \max |\psi^*(\mathbf{k})|$. Here the value of α is chosen so that the peaks of the minimal height are left. We choose $\alpha = 0.6$ for the two-dimensional objective patterns, whereas $\alpha = 0.01$ for the three-dimensional objective patterns.

The dodecagonal QC pattern in two dimensions is synthesized by $\psi = \sum_{i=1}^{12} \cos(\mathbf{q}_i^* \cdot \mathbf{x})$ in which the wave vectors \mathbf{q}_i^* are chosen at the position of the vertices of the hexagon with a radius $|\mathbf{q}_1^*| = 2\pi / \sqrt{2 + \sqrt{3}}$ and the hexagon with a radius $|\mathbf{q}_2^*| = 2\pi$ rotated by $\pi/12$. The DG pattern is expressed by 24 wave vectors of $\mathbf{q}^* = (\pm 2, \pm 1, \pm 1)$ and 12 wave vectors of $\mathbf{q}^* = (\pm 2, \pm 2, 0)$ with their permutation along the x, y, z directions [56,57]. The amplitude of the latter wave vector is $\sqrt{8/6} \simeq 1.15$ times longer than the former waves. The FK A15 pattern is expressed by 24 wave vectors $\mathbf{q}^* = (\pm 2, \pm 1, 0)$, 24 wave vectors $\mathbf{q}^* = (\pm 2, \pm 1, \pm 1)$, 6 wave

vectors of $\mathbf{q}^* = (\pm 2, 0, 0)$ with their permutation along the x, y, z directions [58]. The amplitude of the peaks in the structure factor of FK A15 is chosen as the values obtained in the experiments of [26].

2. Order parameters

To assure translational invariance, we use a spectrum of the Fourier transform of the pattern and expand it by the basis functions, each of which expresses certain point group symmetries. In two dimensions, the basis functions reflect n -fold rotational symmetry, as shown in Fig. 1(b), whereas in three dimensions, spherical harmonics expansion is used. Projection of the Fourier spectrum of the pattern onto the basis function is given by $A_{l,\pm}$ in two dimensions, and A_{lm} in three dimensions.

The order parameter is a rotational invariant form of the quantity A_{lm} with $l \in [0, l_0]$ and $m \in \{\pm 1\}$ in two dimensions and $m \in [-l, l]$ in three dimensions. In two dimensions, $A_{l,\pm 1}$ is described by

$$A_{l,\pm 1}[\psi] = \int |\psi(\mathbf{k})| \begin{pmatrix} \cos l\theta_k \\ \sin l\theta_k \end{pmatrix} d\mathbf{k}^2, \quad (20)$$

where $+1$ (-1) corresponds to $\cos l\theta_k$ ($\sin l\theta_k$), respectively, and θ_k is a polar angle in the Fourier space. We use the Fourier transform of the pattern as

$$\tilde{\psi}(\mathbf{k}) = \int \psi(\mathbf{x}) e^{i\mathbf{k} \cdot \mathbf{x}} d\mathbf{x}, \quad (21)$$

$$\psi(\mathbf{x}) = \int_{\mathbf{k}} \tilde{\psi}(\mathbf{k}) e^{-i\mathbf{k} \cdot \mathbf{x}}, \quad (22)$$

where the volume in the Fourier space is $\int_{\mathbf{k}} = \frac{1}{(2\pi)^d} d^d \mathbf{k}$. We denote $A_{l,\pm}$ in the vector form as $\mathbf{A}_l = (A_{l,+}, A_{l,-})$. The maximum mode is denoted by l_0 . In three dimensions, A_{lm} is given by

$$A_{lm}[\psi] = \int |\psi(\mathbf{k})| Y_l^m(\theta_k, \varphi_k) d\mathbf{k}^3 \quad (23)$$

with spherical harmonics $Y_l^m(\theta_k, \varphi_k)$ in the spherical coordinates of the Fourier space (k, θ_k, φ_k) . Note that m in the superscript of $Y_l^m(\theta_k, \varphi_k)$ and subscript of A_{lm} should not be confused by m describing a model in M . The zeroth mode $l = 0$ corresponds to the mean amplitude of $\tilde{\psi}(\mathbf{k})$, which is independently considered by $\tilde{\psi}$. We, therefore, use the sum for $l \in [1, l_0]$ in the cost function. The maximum mode is denoted by l_0 . We use the convention of spherical harmonics

$$Y_l^m(\theta, \varphi) = \sqrt{\frac{(2l+1)(l-m)!}{4\pi(l+m)!}} P_l^m(\cos \theta) e^{im\varphi}, \quad (24)$$

where $P_l^m(\cos \theta)$ is associated Legendre polynomial with integers l and $m \in [-l, l]$. Any continuous function on a unit sphere may be expanded.

We define the order parameter $\Psi[\psi(\mathbf{x})] = \{\Psi_l[\psi(\mathbf{x})]\}_{l=1}^{l_0}$ by a rotationally invariant form of the coefficients, $A_{l,\pm}$ or A_{lm} ,

$$\Psi_l = \|\mathbf{A}_l\| \equiv \sqrt{A_{l,+}^2 + A_{l,-}^2}, \quad (25)$$

in two dimensions, and

$$\Psi_l = \|A_{lm}\| \equiv \sqrt{\frac{4\pi}{2l+1}} \sqrt{\sum_{m=-l}^l (-1)^m A_{l,m} A_{l,-m}} \quad (26)$$

in three dimensions. Here the prefactor is included because $(Y_l^0)^2 + \sum_{m=1}^l Y_l^m(\theta, \varphi) Y_l^{m*}(\theta, \varphi) = (2l+1)/(4\pi)$, and the sum of $|A_{lm}|^2$ scales $2l+1$.

We numerically evaluate A_l for patterns $\psi(\mathbf{x})$ and A_l^* for an objective pattern $\psi^*(\mathbf{x})$. Since both real space and Fourier space density fields are expressed by values at a finite number of mesh points, the range of a mode l is truncated at the maximum mode l_0 . The larger mode extracts a finer structure in the Fourier spectrum, and the structure finer than the mesh size is invalid. We thus take $l_0 = N$. Note that for odd l , $A_{lm} \simeq 0$ and therefore, the dimension of Ψ is $l_0/2$.

C. Statistical inference

1. Bayesian formulation

We may extend our model naturally toward the Bayesian formulation, which enables us not only to choose the optimal PDE, i.e., parameters and the number of the characteristic length scales, but also to evaluate their uncertainty [23,24]. To do this, we assume the order parameter $\Psi[\psi_s]$ of the stationary pattern ψ_s is observed as that of the objective pattern ψ^* with the additive noise ξ_l :

$$\Psi_l^* = \Psi_l(\psi_s; \psi_0, \mu) + \xi_l, \quad (27)$$

where ξ_l is the random variable for each mode l distributed according to zero-mean Gaussian distribution with variance $\beta^{-1} \geq 0$. The noise ξ_l and the corresponding inverse temperature β play a role of the uncertainty of the measurement. Here ψ_s and ψ_0 are the stationary and initial states of ψ in Eq. (11), respectively, and μ is the set of parameters. The dependence of $\Psi_l[\psi_s]$ on ψ_0 and μ is explicitly represented by $\Psi_l(\psi_s; \psi_0, \mu)$. The assumption Eq. (27) is equivalently represented as the conditional probability density

$$p(\Psi_l^* | \psi_0, \mu, \beta) = \sqrt{\frac{\beta}{2\pi}} \exp \left\{ -\frac{\beta}{2} [\Psi_l^* - \Psi_l(\psi_s; \psi_0, \mu)]^2 \right\}. \quad (28)$$

We consider the parameter estimation of μ by marginalizing ψ_0 . Hereafter the discretization $\psi_0 = \{\psi_0^{(j)}\}_{j=1}^{N_d}$ and the reparametrization $\psi_0^{(j)} \rightarrow \bar{\psi} + \psi_0^{(j)}$ are also considered, where $\bar{\psi}$ and $\psi_0^{(j)}$ are the mean density and the (relative) density at the mesh point j , respectively. By Bayes' theorem, the conditional joint probability density of ψ_0 and μ under given $\{\Psi_l^*\}_{l=1}^{l_0}$, β and the model class m is represented as

$$\begin{aligned} p(\mu | \{\Psi_l^*\}_{l=1}^{l_0}, \psi_0, \beta, m) \\ &= \frac{p(\mu | m)}{p(\{\Psi_l\}_{l=1}^{l_0} | \beta, m)} \prod_{l=1}^{l_0} p(\Psi_l^* | \psi_0, \mu, \beta), \\ &\propto \exp \left[-\frac{\beta}{2} E(\psi^*, \psi_s; \psi_0, \mu) \right], \end{aligned} \quad (29)$$

where m denotes the number of the characteristic length scales such as Eq. (B1), (B2), or (B5). Here $p(\mu | m)$ and $p(\psi_0)$ are the *prior* distribution defined as the uniform distribution, and

the marginal likelihood $p(\{\Psi_l\}_{l=1}^{l_0} | \beta, m)$ is given by

$$\begin{aligned} p(\{\Psi_l\}_{l=1}^{l_0} | \beta, m) \\ &= \left(\frac{\beta}{2\pi} \right)^{\frac{l_0}{2}} \int \exp \left[-\frac{\beta}{2} E(\psi^*, \psi_s; \psi_0, \mu) \right] \\ &\quad \times p(\psi_0) p(\mu | m) d\psi_0 d\mu. \end{aligned} \quad (30)$$

The dependence of $E[\psi^*, \psi_s]$ on ψ_0 and μ is explicitly represented by $E(\psi^*, \psi_s; \mu, \psi_0)$. Note that we assume the *causality*

$$\begin{aligned} p(\psi_0, \mu | \{\Psi_l^*\}_{l=1}^{l_0}, \beta, m) \\ &= p(\psi_0) p(\mu | \{\Psi_l^*\}_{l=1}^{l_0}, \psi_0, \beta, m), \end{aligned} \quad (31)$$

which ignores (i) the dependence of ψ_0 on $\{\Psi_l^*\}_{l=1}^{l_0}$ and β and (ii) the correlation between ψ_0 and μ . This assumption reflects our ansatz that ψ_0 is not uniquely determined only by ψ^* (or $\{\Psi_l^*\}_{l=1}^{l_0}$). Here ψ_0 is treated as a latent variable. By marginalizing out ψ_0 , the *posterior* distribution of μ is given by

$$p(\mu | \{\Psi_l^*\}_{l=1}^{l_0}, \beta, m) = \int p(\psi_0, \mu | \{\Psi_l^*\}_{l=1}^{l_0}, \psi_0, \beta, m) d\psi_0. \quad (32)$$

The posterior mean estimator $\hat{\mu}$, i.e., the mean of $p(\mu | \{\Psi_l^*\}_{l=1}^{l_0}, \beta, m)$, is adopted as our best parameter set. The standard deviation of $p(\mu | \{\Psi_l^*\}_{l=1}^{l_0}, \beta, m)$ plays a role of the error in $\hat{\mu}$.

We consider both the hyperparameter estimation of β and model selection of m [21,22,25,59,60]. By Bayes' theorem, the joint probability density of β and m under given $\{\Psi_l^*\}_{l=1}^{l_0}$ is represented as

$$p(\beta, m | \{\Psi_l^*\}_{l=1}^{l_0}) = \frac{p(\{\Psi_l^*\}_{l=1}^{l_0} | \beta, m) p(\beta) p(m)}{p(\{\Psi_l^*\}_{l=1}^{l_0})}, \quad (33)$$

where $p(\{\Psi_l^*\}_{l=1}^{l_0})$ is the normalization constant. Here $p(\beta)$ and $p(m)$ are the prior distributions defined as the uniform distribution. The maximum *a posteriori* estimator, or equivalently the empirical Bayes estimator in this setup, is adopted as the pair of our optimal model and temperature

$$(\hat{\beta}, \hat{m}) = \operatorname{argmax}_{\beta, m} p(\beta, m | \{\Psi_l^*\}_{l=1}^{l_0}) \quad (34)$$

$$= \operatorname{argmax}_{\beta, m} p(\{\Psi_l^*\}_{l=1}^{l_0} | \beta, m). \quad (35)$$

For convenience, the Bayes free energy $F(\beta, m) = -\log p(\{\Psi_l^*\}_{l=1}^{l_0} | \beta, m)$ is defined. Using the Bayes free energy, we may see the optimal model and temperature $(\hat{\beta}, \hat{m})$ minimize $F(\beta, m)$. If $\partial F / \partial \beta = 0$ is satisfied at $\beta = \hat{\beta}$, then we obtain the self-consistent equation

$$\hat{\beta} = \frac{1}{\langle E(\psi^*, \psi_s; \psi_0, \mu) \rangle_{\hat{\beta}}}, \quad (36)$$

where

$$\langle \dots \rangle_{\beta} = \int (\dots) p(\psi_0, \mu | \{\Psi_l^*\}_{l=1}^{l_0}, \beta, m) d\psi_0 d\mu. \quad (37)$$

By marginalizing out β , we can also evaluate the uncertainty of m as the probability

$$p(m | \{\Psi_l^*\}_{l=1}^{l_0}) = \int p(\beta, m | \{\Psi_l^*\}_{l=1}^{l_0}) p(\beta) d\beta. \quad (38)$$

Note that $p(m_1 | \{\Psi_l^*\}_{l=1}^{l_0}), \dots, p(m_{i_{\max}} | \{\Psi_l^*\}_{l=1}^{l_0})$ demonstrate the probability of each model $m_1, \dots, m_{i_{\max}}$, respectively, based on the observations $\{\Psi_l^*\}_{l=1}^{l_0}$.

2. Setup of a prior distribution

We assume no prior information about parameters and latent variables except their range. The prior density of each variable is defined by the continuous uniform distribution whose support equals the domain of each variable. The prior density of ψ_0 is defined by

$$p(\psi_0) = \prod_{j=1}^{N^d} \varphi(\psi_0^{(j)}), \quad (39)$$

where $\varphi(\psi_0^{(j)})$ is the continuous uniform distribution whose support is $\psi_0^{(j)} \in [-0.1, 0.1]$. For equation Eq. (28) with m length scales, the set of parameters is defined by

$$\mu = \{dx, dy, \bar{\psi}, a_0, q_1, a_1, q_2, a_2, \dots, q_{m-1}, a_{m-1}\}, \quad (40)$$

where $\dim(\mu) = 2m + 2$ in two dimensions. In three dimensions, the mesh size along the z axis dz is added in the parameters. The prior density of μ is also defined by

$$p(\mu | m) = \varphi(dx)\varphi(dy)\varphi(\bar{\psi})\varphi(a_0) \prod_{i=1}^{m-1} \varphi(a_i)\varphi(q_i), \quad (41)$$

where $\varphi(dx)$, $\varphi(dy)$, $\varphi(\bar{\psi})$, $\varphi(a_i)$, and $\varphi(q_i)$ are the continuous uniform distributions, whose supports are respectively $dx \in [1 - (1/q^*N), 1 + (1/q^*N)]$, $dy \in [1 - (1/q^*N), 1 + (1/q^*N)]$, $\bar{\psi} \in [-1, 0]$, $a_i \in [-0.2, 0.2]$, and $q_i \in [0, 1]$. Here $2\pi/q^*$ is the wavelength that is used to synthesize the objective pattern.

We also assume no prior information for the model and hyperparameter; the prior distribution of each variable is defined by the discrete uniform distribution. The prior distribution $p(\beta)$ is also defined by the discrete uniform distribution with $\beta \in \{\beta_\alpha\}_{\alpha=0}^{N_{\text{rep}}-1}$, where $\beta_0 = 0$ and

$$\beta_\alpha = 10^{\log_{10} \beta_{\min} + \frac{\alpha-1}{N_{\text{rep}}-1} \log_{10}(\beta_{\max}/\beta_{\min})} \quad (42)$$

for $\alpha \in \{1, 2, \dots, N_{\text{rep}} - 1\}$. Here we set as $N_{\text{rep}} = 40$, $\beta_{\min} = 10^{-3}$ and $\beta_{\max} = 10^2$. Equation (42) means that discretization of β is finer at the large β (lower variance of noise). The prior distribution $p(m)$ is defined by the discrete uniform distribution at $m \in \{m_i\}_{i=1,2,\dots,i_{\max}}$. Each grid point (m_i, β_α) can be regarded as the candidate of model selection equally possible in prior.

3. Sampling from a posterior distribution with Monte Carlo method

The realization of $p(\mu | \Psi^*, \psi_0, m, \beta)$ is carried out by Monte Carlo (MC) sampling in the parameter space. For each point in the parameter space, we compute a stationary pattern ψ_s in the model of Eq. (5) under the randomly chosen initial

condition ψ_0 . Then, the parameters are changed stochastically according to the Metropolis criterion defined by the cost function (energy) $E[\psi, \psi^*]$ and inverse temperature β . We use the REMC method [27,28] with different inverse temperatures β in parallel. The REMC is an efficient method for the estimation of the optimal variance $\hat{\beta}^{-1}$ of the observation noise because the method enables us to sample parameters simultaneously under various β . The method also makes an efficient sampling to avoid a local trap in the parameter space. Bridge sampling [61,62] is used to calculate $F(m, \beta)$ for each m . The error bars of $F(m, \beta)$ are calculated by the bootstrap resampling [63].

The joint distribution $p(\psi_0, \mu | \{\Psi_l^*\}_{l=1}^{l_0}, \beta, m)$ is realized by the Gibbs sampling based on the relation of Eq.(31), i.e., the alternately iterative sampling from $p(\psi_0)$ and $p(\mu | \{\Psi_l^*\}_{l=1}^{l_0}, \psi_0, \beta, m)$. The sampling from $p(\psi_0)$ simply follows Eq. (39). The sampling from $p(\mu | \{\Psi_l^*\}_{l=1}^{l_0}, \psi_0, \beta, m)$ follows the procedure below. First, we solve the model of Eq. (11) under a given initial state ψ_0 and parameters μ of Eq. (40) for a model m . Then, the similarity of an obtained pattern ψ_s as a stationary state and an objective pattern ψ^* is evaluated by the cost function $E(\psi^*, \psi_s; \psi_0, \mu)$, describing the distance shown in Eq. (2) in the space of the order parameter Ψ_l [see Eqs. (25) or (26)]. Changing μ , we may iterate numerical simulations and evaluation of the similarity between them. Following the Metropolis criterion at an inverse temperature β , we compare a current cost function with a cost function in a previous step, and decide whether a current set of parameters is accepted or not.

By using the replica-exchange Monte Carlo (REMC) method, we sample ψ_0 and μ from $p(\psi_0, \mu | \{\Psi_l^*\}_{l=1}^{l_0}, \beta_\alpha, m)$ for N_{rep} replicas in parallel [27,28]. At higher temperature (smaller β), the motion of one MC step in the parameter space is large, whereas, at the lower temperature (larger β), each motion is small so that it intensively samples parameters near the minimum of the cost function. For every two steps, the parameter sets of neighboring β were exchanged following the Metropolis criterion. This process enables us to sample parameters weighed with likelihood effectively [27,28].

The initial parameter set is sampled from the prior distribution of Eq. (41). The lowest cost function is typically achieved by 1000–2000 MC steps. In one MC step, the Gibbs sampling is used to perform motion in the parameter space in all directions one by one. After finding the lowest cost function, we restart REMC from the initial parameter set of the lowest energy state to sample its steady state. This is because the relaxation under smaller β is much faster than the larger β . After 1000 MC steps, we cut the initial burn-in steps and compute the statistical quantities after 200 MC steps. Bridge sampling was used to calculate $F(\beta, m)$ for each m [61,62]. The error bars of $F(\beta, m)$ were calculated by the bootstrap resampling [63].

D. Regression method for noisy data

In the BM-PDE, the estimation from data with noise is performed by adding zero-mean Gaussian noise ξ in the objective pattern, namely, $\psi^*(\mathbf{x}) \rightarrow \psi^*(\mathbf{x}) + \xi$. In addition to this estimation using BM-PDE, another parameter estimation method is tested. We performed parameter estimation using

the regression method in which the following cost function was used

$$E = \frac{1}{2} \int [\psi^* - f_\mu^{dt}(\psi^*)]^2 d\mathbf{x}. \quad (43)$$

Here the objective pattern ψ^* is numerically forward in dt by the model of Eq. (5) with a parameter set $\{\mu\}$ as $f_\mu^{dt}[\psi(t)] = \psi(t + dt)$. If the objective pattern is the stationary solution of the model, namely, if the parameters are ground truth to obtain the objective pattern, the cost function must be zero. This approach is philosophically the same as the regression method in previous studies in which the cost function is a difference between the *left-hand side* (time derivative) and the *right-hand side* (force to change ψ), that is, $E = (1/2)\|\partial_t \psi - f_\mu(\psi)\|$, under an appropriate norm $\|\cdot\|$ [48,64]. The norm is typically chosen as the L^2 norm

$$E = \frac{1}{2} \int [\partial_t \psi - f_\mu(\psi)]^2 d\mathbf{x}. \quad (44)$$

In the current system, our model is no longer linear in the parameters, and therefore, we cannot use linear regression (including conventional sparse regression). In order to carry out nonlinear regression, we used the REMC method to minimise the cost function, Eq. (43). The method is similar to our main algorithm in which we sample parameters and estimate the optimal noise by temperature β . Following Bayes' theorem, we estimate the best parameters by the sampled values and their error by the standard deviation.

ACKNOWLEDGMENTS

The authors are grateful to Edgar Knobloch, Yasumasa Nishiura, An-Chuang Shi, and Philippe Marcq for helpful discussions. The authors acknowledge the support by JSPS KAKENHI Grants No. 17K05605, No. 20H05259, and No. 20K03874 to N.Y. and No. 20K19889 to S.T. This work is also supported by JST FOREST Program (Grant No. JPMJFR2140, Japan). Numerical simulations in this work were carried out in part by AI Bridging Cloud Infrastructure (ABCI) at National Institute of Advanced Industrial Science and Technology (AIST), and by the supercomputer system at the information initiative center, Hokkaido University, Sapporo, Japan.

APPENDIX A: RELATED WORK

Our method covers interdisciplinary research fields such as materials science of soft materials and crystalline structures, nonlinear dynamics of pattern formation, and data-driven science techniques of machine learning and inverse problems. In this section, we review related works.

1. Pattern formation and crystalline structure

First, we give a brief review of nonlinear dynamics on pattern formation and self-assembly of crystalline-like structures. The nonlinear partial differential equations, such as the Swift-Hohenberg (SH) equation [1], have been used to describe a periodic pattern as a minimal model. SH was originally derived for stationary finite number instability of fluid

convection [1,65–68], and then has later been applied to optics [69], ecosystem [70], and other seemingly distinct problems. The Ohta-Kawasaki-Oono (OKO) equation, which is similar to the SH equation, was proposed to describe the micro-phase separation of block copolymers [71,72], whereas the SH equation may also have been used for the same problem [33,73]. In the theory of block copolymers, the SH equation is instead called the Landau-Brazovskii theory. These equations reproduce a stripe (also called lamellar or smectic) and hexagonal patterns in two dimensions [52], and lamellar, hexagonal cylinders, BCC, hexagonal closed packing patterns, also gyroid patterns in three dimensions [32,33,36]. The essence of these models is that there is a specific length scale, or wave number, at which the uniform state becomes unstable. Recently, the conserved version of the SH equation has been used for the description of crystalline structures [4,5]. This model is specifically called the phase-field crystal (PFC) model. The models that we used in this work are extensions of the PFC model.

All these models are based on nonlinear partial differential equations (PDEs). This approach has an advantage and disadvantage compared to particle-based simulations, such as molecular dynamics and Monte Carlo simulations that are widely used to describe crystalline structures. The advantage is that the system is expressed using a continuum (mostly scalar) field in which the position of a *particle* is identified by a peak of the field. Because the density field is defined everywhere in the system, it is easier to analyze structures (symmetry, length scale, and so on) and deviation from a perfect crystalline structure (dislocation and disclination). These topological defects have been studied not only in solids [74], but also in soft materials [3,75]. The dynamics of defects have been studied by using PDEs [5,52]. The PFC is a natural extension of the conventional phase-field model, and is capable of describing diffusive timescales as well as local crystalline order [4]. Therefore, the PFC captures both elasticity and plasticity, and long-time behaviours of defect dynamics. The disadvantage is that there is no explicit form of the interaction potential between particles, and efforts are necessary to map molecular interactions to the associated linear and nonlinear terms in PDEs [76].

The extension of the PFC (or SH) has been proposed in several ways. One is to introduce two length scales to destabilize the uniform state [14,77]. This results in stable quasicrystalline patterns such as decagonal (10-fold) and dodecagonal (12-fold) QCs in two dimensions. This model also reproduces an icosahedral pattern in three dimensions [42], although in this case, the two length scales are not linearly unstable but close to neutral stability. The QC pattern appears due to a nonlinear selection mechanism. The second type of extension is to use many order parameters as a model of a multicomponent system. Along this line, QC patterns are reproduced in two dimensions [78] and in three dimensions [79]. The third type of extension is the weak crystallization theory based on the Landau-Brazovskii theory [80]. The simplest version of this theory is nothing but the SH equation. In the weak crystallization theory, anisotropic nonlinear terms are included to make complex patterns to reproduce, for example, QCs in two dimensions [81]. The model using the anisotropic nonlinear terms corresponds to an anisotropic interaction in a

microscopic model, for example, using Janus particles, patchy particles [82,83], or polyhedral-shaped particles [84].

2. Inverse structural optimization

Studies on inverse structural design have been performed mostly using particle-based models. The two major directions of the research is a control of the structure through the external template, and the design of potential interaction between particles. The template method is more intuitive than the potential design, but it requires a good template beforehand. On the other hand, the design of potential interactions does not need the prescribed structure, and the structure appears spontaneously from a uniform disordered state. This self-assembly of discrete particles has an advantage against the template approach because after optimization, the structures are reproduced without external aids, and thus it is easy to scale up the structures [85–87]. Most of the studies along this line are discrete particles using either Monte Carlo, molecular dynamics simulations, or an extension of them. Examples include simulated annealing for the parameters in Lennard-Jones-type isotropic interactions using the difference of the particle position from the objective lattice in real space as a cost function [88,89] (see also [90,91] and references therein). These studies initiate the inverse structural design of the self-assembly of materials [92,93]. The design is based on various optimization techniques such as gradient method [94,95], generic algorithm [96], swarm optimization [97,98], the covariance matrix adaption-evolutionary strategy (CMA-ES) [99,100], and statistical physics-inspired inverse design (SP-ID) [100,101].

Recently, the optimization of interaction potentials by minimization of the relative entropy combined with Monte Carlo simulations has been proposed [94,95,102–106]. In those studies, parameters in a potential function are optimized by a gradient of pair-potential with respect to the parameters using the difference between the radial distribution functions in objective and generated patterns. The objective structure is given by the positions of particles, and by adding fluctuation around the positions, the radial distribution function for the structure is obtained. This method treats the probability of the microscopic structures for given parameters by the Boltzmann distribution. The inverse structural design has been applied to various soft materials such as colloidal assembly [94,95,102–104,106], block copolymers [97,98], and granular media [107]. These methods, thus far, are the point estimate in which only the most likely parameters are estimated. In the BM-PDE, we compute the posterior distribution from which the uncertainty of the parameters can be evaluated. We also optimize the noise amplitude described by the inverse temperature β . This optimization has not been performed in the previous works, and it is essential to estimate the best parameters for the objective pattern with noise and without ground-truth parameters. Lack of optimization of the noise amplitude results in overfitting of the parameters for the objective pattern under noise [25]. Besides, the BM-PDE is an inverse structural design of the model described by PDEs. The optimization problem of PDEs has focused on time-series data, and therefore, for the snapshot of a stationary structure, the inverse problem has not been formulated.

3. Parameter estimation of governing equations

The automatic discovery of the governing equations from data is a relevant topic in data-driven science. Parameter estimation of nonlinear PDEs has been studied [39,50,108] along the lines of system identification, optimization, and control of nonlinear dynamical systems [38]. Initiated by the success of physical laws for double-pendulum [109,110], it is becoming feasible to estimate an equation of motion from data ψ . Recently, the method of sparse linear regression was successfully applied to ordinal differential equations (ODEs) [48] and partial differential equations (PDEs) for time-series data [51,64]. The key idea is to minimize the error of $\|\partial_t \psi(t) - f[\psi(t)]\|$ under a certain norm $\|\cdot\|$ with a regularization term. The function $f(\psi)$ is expanded in terms of polynomials of ψ with their coefficients. The polynomials may be replaced by a list of candidate terms of ψ . Then the problem reduces to an estimation of their coefficients. Estimation based on sparse regression helps to make many coefficients zero, so that only a few terms remain in the estimated governing equation. Parsimony is an underlying philosophy of the method; the governing equation *should* be described by minimal terms in ψ and minimal terms in spatial derivatives in the case of PDEs. The idea of sparsity has also been combined with neural networks to estimate symbolic equations [111].

Our problem to estimate the governing PDE from a given objective pattern is fundamentally different from those studies in several respects. First, the methods mentioned above are based on the regression for the time-series data and therefore demand an accurate observation of $\partial_t \psi$ [38,50,112,113]. The objective pattern in our problem is only one snapshot ideally satisfying $\partial_t \psi = 0$ without noise, and therefore, we cannot use optimization with respect to $\partial_t \psi(t) - f[\psi(t)]$. Moreover, the information on the snapshot is far less than time-series data because, in the latter case, there are data of $\psi(t)$ under different time t . Second, our interest is to estimate not only parameters but also the best model. Third, our objective pattern is not necessarily produced from a numerical result but synthesized by a function of Eq. (19) that is independent of the models.

Another feature of BM-PDE is uncertainty quantification. The statistical inference has been widely used to estimate parameters with their errors (uncertainty) [22,114]. This approach is well established in the linear regression problems. However, fewer studies have been made in uncertainty quantification for the estimation of governing equations, particularly for nonlinear PDEs [40,115–117]. The majority of the methods are based on point estimation in which parameters are estimated by minimizing the cost function (energy) [48,51,64]. To estimate parameters in ODEs, statistical inference has been used for time-series data using approximate Bayesian computation [118].

4. State-space model

To summarize the related studies on the parameter estimation of the dynamical equations, it is instructive to consider the state-space model. We first explain the estimation from time-series data, and then discuss the difficulty in estimating the stationary problem. In the state-space model, the cost

function is given by

$$E = \frac{1}{2} \sum_{i,\alpha} [\psi^*(\mathbf{x}_i, t_\alpha) - \psi(\mathbf{x}_i, t_\alpha)]^2 + \frac{\lambda}{2} \int \{\partial_t \psi(\mathbf{x}, t) - f_\mu[\psi(\mathbf{x}, t)]\}^2 d\mathbf{x} dt, \quad (\text{A1})$$

where the indices i and α are measurement points in space and time, respectively [see also Eq. (1)]. Within this model, the data at hand are ψ^* , and the model is parametrized by μ . The first term in Eq. (A1) makes the pattern $\psi(\mathbf{x}, t)$ closer to the data (measurement). The second term penalises the pattern that deviated from the model $f_\mu(\cdot)$ under the parameters μ (model). Here λ sets a balance between the two terms in Eq. (A1). The cost function equation (A1) implies that both measurement and model have noise,

$$\psi^*(\mathbf{x}, t) = \psi(\mathbf{x}, t) + \xi_1 \quad (\text{A2})$$

$$\partial_t \psi(\mathbf{x}, t) = f_\mu[\psi(\mathbf{x}, t)] + \xi_2. \quad (\text{A3})$$

When the noise ξ_1 and ξ_2 are taken from the normal distribution, the likelihood is given by e^{-E} .

When the model is expressed by a deterministic equation, the trajectory of the solution $\psi(\mathbf{x}, t)$ is described by the flow map from the initial condition, $\psi(\mathbf{x}, t) = \Phi_t[\psi_0(\mathbf{x})]$. In this case, the cost function is $E[\psi^*(\mathbf{x})|\psi_0(\mathbf{x}), \mu]$ [20]. Then, using Bayes' theorem, we may consider the posterior distribution from three choices. The first choice is the posterior distribution of parameters $p(\mu|\psi^*(\mathbf{x}), \psi_0(\mathbf{x}))$ from the prior distribution of the parameters $p(\mu)$ with a fixed initial condition. The second choice is $p(\psi_0(\mathbf{x})|\psi^*(\mathbf{x}), \mu)$ from the prior distribution of the initial conditions $p[\psi_0(\mathbf{x})]$ with a fixed parameter. The third choice is $p(\psi_0(\mathbf{x}), \mu|\psi^*(\mathbf{x}))$ from both the prior distribution of parameters and the initial conditions. Even in the first and the second choices, the parameters and the initial conditions may be estimated from hyperparameter estimation [22]. In data assimilation, the initial condition is often estimated, and the parameters are estimated by hyperparameter estimation (second choice) [119]. When both the initial condition and the parameters are estimated (third choice), state augmentation has been used [19]. In this method, the parameters are treated as state variables. In [120], parameters in the phase-field model are estimated, and the initial condition is estimated by hyperparameter estimation. All the choices discussed above estimate the unique initial condition. However, for the stationary data in pattern formation, the same pattern appears from different initial conditions. It is rather desired that the objective pattern can be generated from a wide range of the initial conditions. In our method, we consider the posterior distribution of parameters $p(\mu|\psi^*(\mathbf{x}), \psi_0(\mathbf{x}))$ with a fixed initial condition (first choice), but the posterior distribution is marginalized about the initial conditions. This marginalization is a core of our approach, and it enables us to avoid nonunique estimation for stationary data, and to estimate the model that generates the objective pattern as a stable state.

When there is no measurement noise, that is $\xi_1 = 0$, and the measurement is dense in time so that the time derivative $\partial_t \psi$ is accurately measured [see Fig. 14(a)], we may drop the

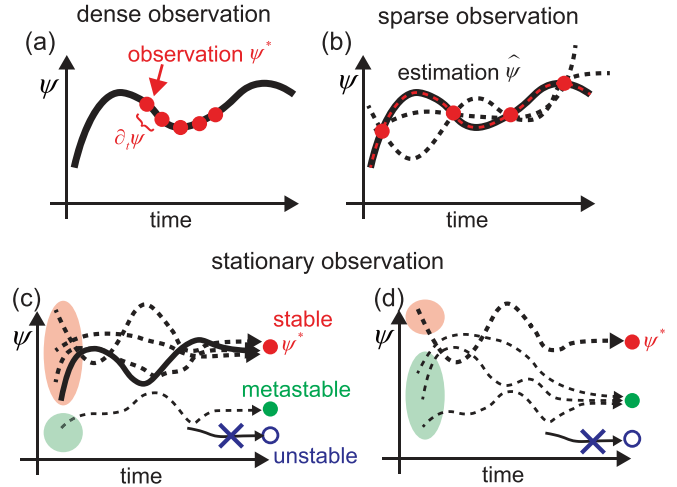


FIG. 14. Schematic estimated trajectories for dense observation (a), sparse observation (b), and stationary data (c), (d). The observations ψ^* are shown in red points. The true trajectory is shown in the solid curve, whereas the estimated trajectories $\hat{\psi}$ are shown in dashed curves. For the stationary data, two models are shown. (c) The stationary data are generated from broad initial conditions. (d) The stationary data are generated from narrower initial conditions. In this case, most of the initial conditions lead to the metastable state. In both cases, the unstable solution cannot be achieved from the initial conditions, even though it satisfies $f(\psi) = 0$.

first term in Eq. (A1), and the cost function becomes

$$E_{\text{reg}} = \frac{1}{2} \sum_{i,\alpha} \{\partial_t \psi(\mathbf{x}_i, t_\alpha) - f_\mu[\psi(\mathbf{x}_i, t_\alpha)]\}^2. \quad (\text{A4})$$

We call this cost function the regression model, which has been used to estimate μ for a PDE model in [38,50,108]. By adding the L^1 regularization term in E_{reg} , the sparse estimation of the parameters was demonstrated in [48,51,64,121]. With this regularization, only a few parameters become nonzero. This sparse regression method is appealing because even when the dynamical equation $f(\psi)$ is nonlinear, the model can be parameterized by linear terms in the parameters, such as $\mu_2 \psi^2$, $\mu_{2,1} \nabla \psi^2$, and so on [51]. Therefore, the estimation of the parameters falls into the linear regression problem, which can be handled in various ways [122]. The uncertainty of the estimation is also obtained once the linear regression problem is reformulated in the Bayesian framework [114,123]. Sparsity may be included in the prior distribution [121]. Nevertheless, in the regression model, the error is included only in the model. Therefore, when there is measurement noise in data, the estimation does not work even for the linear regression problem [38]. This issue has been studied by the total least square [38].

When the measurement is sparse in time, ψ in Eq. (A1) is available only at sparse points as ψ^* [Fig. 14(b)]. In this case, both the parameters μ and the unknown solution of the PDE $\psi(\mathbf{x}, t)$ need to be estimated. The simplest way is to interpolate the unobserved data in time and space, for example, by the spline method [124,125]. Clearly, to interpolate accurately, the observation point should be dense if not complete, and

the noise should be small [50,108]. If the interpolation works well, the estimation of the parameters can be made in the same way as the above-mentioned regression method.

When the observation is so sparse that the interpolation does not work, both the parameters μ and the solution of the PDE $\psi(\mathbf{x}, t)$ must be estimated. Even if the observation is linear as $\psi^* = \psi + \xi_2$, and the model is assumed to be linear in the parameters, the estimation for the sparse observation is inevitably nonlinear. This is in contrast with the regression method discussed above. The reason is that the dynamical equation $f_\mu(\psi)$ is nonlinear, and the solution of the PDE is a nonlinear function of the parameters μ . Therefore, the estimation under the sparse observation is far more difficult than that under the dense observation. To handle this issue, various methods have been proposed. Below we discuss these methods briefly and point out the difficulty for applying them to stationary data.

To our knowledge, all the studies on the estimation of governing equations described by PDEs are based on time-series data. Our problem is to estimate the best model and parameters from one snapshot of a pattern at the stationary state. Trajectory matching cannot be used for this problem because there is no information about the initial condition. We need to eliminate the dependence on the initial condition. To do this, we marginalize the initial condition, which is taken from a random distribution. We also have to identify the same patterns which have a different orientation. The different orientation arises from different initial conditions under the same model and the same parameters. This argument is the reason why the order parameter has to be introduced in BM-PDE.

a. Gaussian process

First of all, when both the observation and the model are linear, the Gaussian process is a very powerful tool [126]. Because the time and spatial derivatives are linear operations, the pattern $\psi(\mathbf{x}, t)$ taken from the Gaussian prior distribution remains Gaussian. Therefore, the posterior distribution may be computed by inversion of the matrix of the kernel. This method has been used for linear ODEs and PDEs [127–130]. The method using the Gaussian process is robust against noise [127,130] and can handle hidden variables [128] and boundary conditions [129], but the method is limited to the linear models. For nonlinear problems, the Gaussian process may be used by marginalizing the time derivative of the variables [131–133]. When the time derivative of the solution $\psi(\mathbf{x}, t)$ can be measured in the dense observation, the model can be linearized between the small time steps between the two observations. In this case, the Gaussian process works even for a nonlinear model [134]. The drawback of the Gaussian process is the interpretability of the estimated model. The estimated model can reproduce the observation, but its physical meaning is hardly understood from the estimation results. Another issue is computational cost. For the PDEs, the method assigns the Gaussian prior at each spatial and time point. The matrix inversion requires $O(M^3)$ computation where M is the total spatial points $M \sim N^d$ where N is the number of mesh points in each spatial direction. Therefore, for higher dimensional systems, this method is not feasible.

b. Shooting method

The second method is the shooting method (also multiple shooting and extended multiple shooting method) [50,135,136]. The idea is to solve Eq. (A1) explicitly by using the measured data as an initial condition. In most cases, solving the model is performed numerically. The estimated solution of $\psi(\mathbf{x}, t)$ is obtained under the given parameters. By minimizing the first term in Eq. (A1) with respect to the parameters, the best parameters may be estimated.

In the simplest shooting method, we use only one measurement for the initial condition and use the rest of the data to compute the cost function. When the model is sensitive to the initial condition, such as chaotic dynamical systems, the error grows exponentially in time. Therefore, this method results in a larger error away from the initial condition. To handle such data, in the multiple shooting method, several data are used as initial conditions, and ψ is estimated in a short period of time for each of the initial conditions.

c. Filtering and sequential data assimilation

Sequential data assimilation is widely used to fit data to a model at hand [19,20]. It consists of three steps: prediction, filtering, and smoothing. In prediction, the model is evaluated, and the state at the next step in time $\psi(\mathbf{x}, t + \Delta t)$ is predicted from the current time $\psi(\mathbf{x}, t)$. In filtering, the predicted value $\psi(\mathbf{x}, t + \Delta t)$ is revised to fit the measurement ψ^* . These steps are carried out sequentially in time. Finally, in smoothing, the estimated $\hat{\psi}(\mathbf{x}, t)$ is further revised to fit all the past measurements.

For the deterministic model, the value of $\psi(\mathbf{x}, t)$ at any time t may be expressed by the flow map $\psi(\mathbf{x}, t) = \Phi_t(\psi_0)$ for the initial condition ψ_0 . Therefore, sequential assimilation is often used to estimate the initial conditions. The estimation of the parameters can also be included in the sequential data assimilation.

d. Adjoint method and sensitivity analysis

The adjoint method is called the variational method or 4DVAR in data assimilation. The main idea is to introduce a Lagrange multiplier λ and define the cost function as

$$E_{adj} = \frac{1}{2} \sum_{i,\alpha} [\psi^*(\mathbf{x}_i, t_\alpha) - \psi(\mathbf{x}_i, t_\alpha)]^2 + \int \lambda(\mathbf{x}, t) \{ \partial_t \psi(\mathbf{x}, t) - f_\mu[\psi(\mathbf{x}, t)] \} d\mathbf{x} dt. \quad (\text{A5})$$

The variation of the cost function with respect to the parameters results in the gradient dynamics of both the parameters and the Lagrange multiplier. The equation of the Lagrange multiplier is nothing but solving the model backwards to correct the error between the estimated trajectory and the measurement. This method was successfully applied to pattern-forming PDEs [39,40]. However, in this method, the initial condition has to be fixed because the method makes an estimation by solving the model in forward and backward directions in time. Estimating the initial condition may be possible by the hyperparameter estimation. We are not aware of the discussion about the marginalization of initial conditions

and the stability of the estimated model within the adjoint method.

5. Estimation of the model with a stationary state

When the observation is made only at the stationary state, the data does not have information about trajectories before reaching the stationary state [Figs. 14(c) and 14(d)]. As we have seen in the previous section, the various methods for the parameter estimation of the state-space model rely on time series. Clearly, the method using the time derivative or its interpolation cannot be applied to the stationary problem. Only the stationary data are the *sparsest* measurement in the time series. The shooting method, the sequential data assimilation, and the adjoint method essentially estimate trajectories between two measurements in time. In the stationary problem, there are many initial conditions leading to the same stationary state. We are interested in the model that can reproduce the stationary objective pattern from a wider range of the initial conditions. Then the estimated model is expected to reproduce the objective pattern as a stable state.

Besides such obvious differences from the estimation from time series, the estimation from the stationary data has several difficulties. There are two types of problems for the estimation from the stationary data; the first is to find parameters μ for

$$f_\mu(\psi^*) = 0. \quad (\text{A6})$$

The second is to find the parameters for ψ^* as a solution of

$$\partial_t \psi = f_\mu(\psi) \quad (\text{A7})$$

at $t \rightarrow \infty$. Our method, BM-PDE, focuses on the second problem. The advantage is that it is possible to make sure that the estimated model can reproduce approximately the objective pattern as a stable solution. This is done by marginalizing the initial conditions. When the solution is obtained from a wide range of the initial conditions, the model should reproduce the pattern in a stable manner.

The first problem does not care about the initial conditions. When the model $f_\mu(\psi)$ is nonlinear, there may be several stable and unstable solutions. For this problem, it is not guaranteed whether the objective pattern is a stable or unstable solution of the estimated model. In addition, the estimation of parameters is not unique; when $f_\mu(\psi^*) = 0$, a close neighborhood, $\mu + \delta\mu$, of μ may also satisfy $f_\mu(\psi^*) = 0$ as long as f is not at a bifurcation point. To make the estimation unique, we need a regularization term. The ridge regression (L^2 regularization) or sparse regression (L^1 regularisation) may fix the estimation, but they are independent of the stability of the objective pattern. Therefore, these regularization methods do not guarantee that the estimated model reproduces the objective pattern as a stable state.

APPENDIX B: FAMILY OF MODELS

1. Phase-field crystal model and its generalization

For a single wavelength, we have $m = 1$, and we recover the conserved SH equation (the simplest PFC model). In order to control the position and amplitude of unstable mode, we use

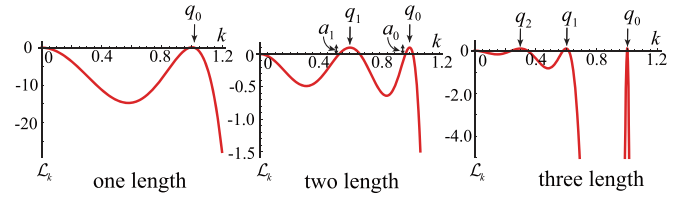


FIG. 15. Linear stability of multi-length-scale phase-field crystal equations. The plots show eigenvalues as a function of the wave number. Positive eigenvalues imply the uniform state is unstable.

the following form:

$$\mathcal{L}_k = -\frac{a_0}{q_0} k^2 (k^2 - 2q_0^2) - s_0 k^2 (q_0^2 - k^2)^2. \quad (\text{B1})$$

With this form, patterns with $k = q_0$ appear, and its stability is denoted by a_0 (see Fig. 15). Here s_0 expresses sharpness of the peak in the spectrum. For larger s_0 , the peak becomes sharper (Fig. 15).

For two length scales, we use another length scale $k = q_1$ in addition to $k = q_0$, and the linear stability around the uniform state for the two length scales, a_0 for $k = q_0$ and a_1 for $k = q_1$ as shown in Fig. 15. This model has been used to reproduce QC patterns [42]. The explicit form of the linear operator is given by

$$\mathcal{L}_k = a_0 S_0(k) + a_1 S_1(k) - \frac{s_0 k^2}{q_1^4} (q_0^2 - k^2)^2 (q_1^2 - k^2)^2, \quad (\text{B2})$$

$$S_0(k) = \frac{1}{q_0^4 (q_0^2 - q_1^2)^3} k^2 (q_1^2 - k^2)^2 \times [(q_1^2 - 3q_0^2)k^2 + q_0^2(-2q_1^2 + 4q_0^2)], \quad (\text{B3})$$

$$S_1(k) = \frac{1}{q_1^4 (q_0^2 - q_1^2)^3} k^2 (q_0^2 - k^2)^2 \times [(3q_1^2 - q_0^2)k^2 + q_1^2(2q_0^2 - 4q_1^2)]. \quad (\text{B4})$$

To make sharp enough peaks, we use $s_0 = 100$.

For three length scales, we may extend the above-discussed expressions and obtain

$$\mathcal{L}_k = a_0 S_0(k) + a_1 S_1(k) + a_2 S_2(k) - s_0 k^2 (q_0^2 - k^2)^2 (q_1^2 - k^2)^2 (q_2^2 - k^2)^2, \quad (\text{B5})$$

$$S_0(k) = \frac{1}{q_0^4 (q_0^2 - q_1^2)^3 (q_2^2 - q_0^2)^3} k^2 (k^2 - q_1^2)^2 (k^2 - q_2^2)^2 \times [q_0^2(-6q_0^4 + 4q_0^2 q_2^2 + 4q_0^2 q_1^2 - 2q_1^2 q_2^2) + k^2(5q_0^4 - 3q_0^2 q_2^2 - 3q_0^2 q_1^2 + q_1^2 q_2^2)], \quad (\text{B6})$$

$$S_1(k) = \frac{1}{q_1^4 (q_0^2 - q_1^2)^3 (q_1^2 - q_2^2)^3} k^2 (k^2 - q_0^2)^2 (k^2 - q_2^2)^2 \times [q_1^2(-2q_0^2 q_2^2 + 4q_0^2 q_1^2 + 4q_1^2 q_2^2 - 6q_1^4) + k^2(q_0^2 q_2^2 - 3q_0^2 q_1^2 - 3q_1^2 q_2^2 + 5q_1^4)], \quad (\text{B7})$$

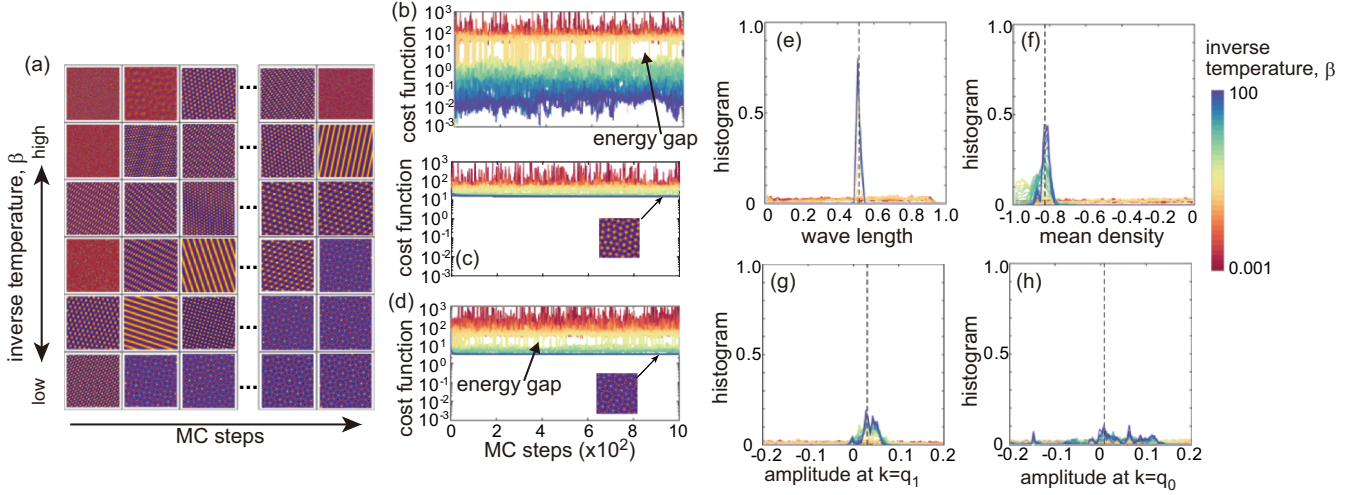


FIG. 16. (a) Snapshots of two-dimensional pattern with two length scales under different inverse temperature β during the Monte Carlo simulations for the objective pattern of numerically produced dodecagonal QC. (b)–(d) The cost function during the steps in replica-exchange Monte Carlo (REMC) simulations in the parameter space at the steady state with one (c), two (b), and three (d) length scales, respectively. (e)–(h) Histogram of the estimated parameters during REMC steps with vertical dashed lines indicating their ground truth. The different colors show different inverse temperature, β .

$$\begin{aligned}
 S_2(k) = & \frac{1}{q_2^4(q_2 - q_0)^3(q_1 - q_2)^3} k^2 (k^2 - q_0^2)^2 (k^2 - q_1^2)^2 \\
 & \times [q_2^2(-2q_0^2q_1^2 + 4q_0^2q_2^2 + 4q_1^2q_2^2 - 6q_2^4) \\
 & + k^2(q_0^2q_1^2 - 3q_0^2q_2^2 - 3q_1^2q_2^2 + 5q_2^4)]. \quad (\text{B8})
 \end{aligned}$$

The linear spectrum is shown in Fig. 15.

2. Polynomial expansion of the linear operation with respect to wave numbers

Our family of models is based on the number of length scales. This is demonstrated in the functional form of the linear operator in the Fourier space shown in Fig. 15 and Eqs. (15) and (16). The benefit of this approach is that parameters μ have clear physical meaning: the wave numbers q_i correspond to the characteristic length scales and stability at the wave numbers a_i . The disadvantage of this approach is that we have to treat the different number of length scales independently. This implies that the two-length-scale model

$m = m_2$ does not include the one-length-scale model. If q_0 and q_1 are overlapped in $m = m_2$, the operator \mathcal{L} diverges, as seen, for example, in Eq. (B3). On the other hand, we may consider a family of the model by expanding \mathcal{L}_k in Eq. (16) by a polynomial expansion of k . In this section, we discuss the drawbacks of this approach and explain why we use our family of models by Eq. (16).

Within the approach of polynomial expansion, parameters $\{\mu\}$ are chosen as coefficients of a polynomial expansion of $\mathcal{L}_k = \sum_i \mu_i k^i$. The advantage of this expansion is that we may express \mathcal{L}_k of different models with the different number of length scales in Fig. 15 in a unified way. Once we truncate the expansion up to k^{14} , we may express a three-length-scale model, and when the coefficients of k^{12} and k^{14} are identically zero, the model describes two length scales. We may use sparse regression by adding regularization and use a particular type of prior distribution of the parameters [137].

The disadvantage of this method is that the coefficients of the expansion in terms of polynomials of k do not have explicit physical meanings. In practical terms, the range of the parameters is too broad, so that estimation is not

TABLE I. The ground-truth and estimated parameter values for numerically produced 12-fold QCs. Errors are evaluated from the standard deviation of all data.

Parameter	Ground truth	Estimated (one length)	Estimated (two length)	Estimated (three length)
dx	1.020	0.987 ± 0.00615	1.0217 ± 0.0024	1.003 ± 0.00033
dy	1.026	0.994 ± 0.0135	1.0232 ± 0.00319	1.0040 ± 0.0033
ψ	-0.826	-0.389 ± 0.0272	-0.820 ± 0.0163	-0.682 ± 0.018
a_0	0.00794	0.055 ± 0.0113	0.0391 ± 0.0594	-0.150 ± 0.0081
q_1	0.518	–	0.518 ± 0.00214	0.729 ± 0.0023
a_1	0.0306	–	0.0375 ± 0.0140	0.0383 ± 0.020
q_2	–	–	–	0.368 ± 0.0018
a_2	–	–	–	-0.136 ± 0.011

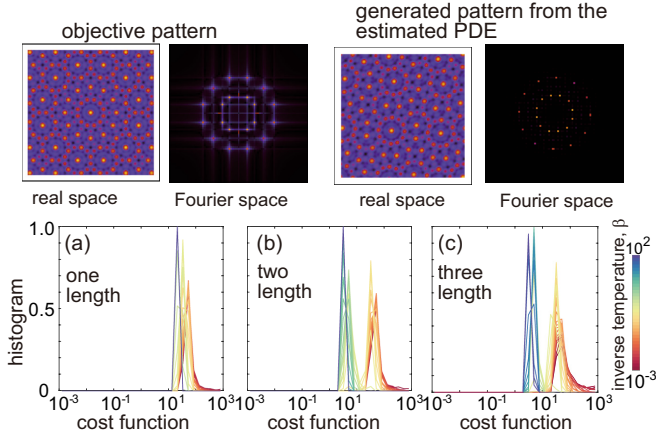


FIG. 17. Top: The objective pattern and the generated pattern from the estimated PDE in real and Fourier space. Bottom: Histograms of the cost function of the REMC sampling for the objective pattern of dodecagonal QC that is synthesized by the function of Eq. (19). The horizontal axis is shown in the logarithmic scale.

possible. For example, the numerically produced dodecagonal QC discussed in Fig. 5 is expressed as

$$\mathcal{L}_k \approx -99.7k^2 + 945.6k^4 - 2984.8k^6 + 3531.8k^8 - 1392.8k^{10}, \tag{B9}$$

and the precision of 0.1 is necessary to obtain QC. Therefore, the range of the parameters should be $[-4000.0, 4000.0]$. Practically, it is impossible to get a reasonable estimate from such wide range of the parameters, and even if possible, uncertainty is certainly much larger than 0.1, and thus there is no guarantee to obtain a QC within the error.

Finally, we should point out that the sparse regression does not work when parameters are correlated [138]. In the system that we are studying, the parameters must be correlated because the amplitude and position of the unstable wave number are set by the peak of \mathcal{L}_k , and it is determined by the balance of several terms in the polynomial expansion.

APPENDIX C: EXAMPLES

1. Objective pattern produced by a numerical simulation: Two-dimensional dodecagonal QC

The process of parameter estimation is shown in Fig. 16(a). At the initial parameters, estimation is poor in all temperatures, β (corresponding to the observation noise). After

several MC steps, the cost function decreases drastically at the lower temperature, and reaches its steady state. At the steady states, sampled patterns are very close to the objective pattern. At the higher temperature, the parameter change in each MC step is too large to converge to a unique pattern; rather, it exhibits nearly random motion in the parameter space. In Fig. 16(b) we demonstrate successful sampling in the parameter space by showing the cost function, or we call it energy, covers a wide range of its values. The distribution of the cost function has two regions connected by the energy gap. The lower cost function corresponds to a QC, which is qualitatively similar to the objective pattern, while the higher cost function is dominated by hexagonal patterns. This result implies that there are multiple minima of the cost function. This means that the standard MC simulation does not converge to the ground truth unless an initial parameter set is chosen nicely,

The same algorithm is used for the different models with one length scale or three length scales [Figs. 16(c) and 16(d)]. Both REMC simulations converge to their steady histograms of the cost function, both of which are higher than the histogram of the two-length-scale model. This result means the two-length-scale model is indeed most likely for the objective pattern. In fact, the marginal likelihood, which we also call free energy $F(\beta, m)$, shown in Fig. 5(e) has the lowest value for the two-length-scale model $m = m_2$. Because this objective pattern has corresponding ground-truth parameter values and the model of Eq. (5) is deterministic, the minimum of the free energy is, strictly speaking, achieved at $\beta \rightarrow \infty$. Nevertheless, we obtain a clear deviation of the free energy of the wrong models $m = m_1$ and $m = m_3$, as they increase at large β . This is because the wrong models do not have exactly the same pattern as the objective pattern in their solution, and therefore, the best estimation is made at the finite noise level, resulting in the minimum at the finite β . In those models, estimation of parameters is carried out at the optimal temperature β at which the marginal likelihood has its minimum value.

The probability of each model is calculated by marginalizing the probability $P \propto e^{-F[\beta]}$ at each temperature, β , about the whole temperature range. This probability is also well captured by the minimal value of the marginal likelihood (free energy) of each model. The result is shown in the inset of Fig. 5(e). In this example, the probability of selecting the two-length-scale model is almost 100%.

The estimate parameters are shown in Table I, together with the uncertainty of the estimation. The uncertainty

TABLE II. The estimated parameter values for functionally synthesized dodecagonal QCs. Errors are evaluated from the standard deviation of each sample from the REMC at the optimal temperature.

Parameter	Estimated (one length)	Estimated (two length)	Estimated (three length)
$dx = dy$	0.986 ± 0.0092	0.980 ± 0.0018	0.992 ± 0.0039
$\bar{\psi}$	-0.183 ± 0.0497	-0.748 ± 0.0192	-0.613 ± 0.0149
a_0	0.010 ± 0.0120	-0.132 ± 0.0443	-0.0581 ± 0.0213
q_1	—	0.514 ± 0.0037	0.916 ± 0.0174
a_1	—	0.0153 ± 0.0085	-0.155 ± 0.0214
q_2	—	—	0.507 ± 0.0029
a_2	—	—	0.0245 ± 0.00762

TABLE III. The estimated parameters values for functionally synthesized stripe pattern. Errors are evaluated from the standard deviation of each sample from the REMC at the optimal temperature.

Parameter	Estimated (one length)	Estimated (two length)	Estimated (three length)
$dx = dy$	0.992 ± 0.0153	0.995 ± 0.0119	0.988 ± 0.0146
$\bar{\psi}$	-0.0471 ± 0.0351	-0.0648 ± 0.0426	-0.0390 ± 0.0281
a_0	0.0329 ± 0.0115	-0.0337 ± 0.0952	-0.0877 ± 0.0788
q_1	–	0.571 ± 0.127	0.728 ± 0.240
a_1	–	0.0112 ± 0.0420	-0.0781 ± 0.0756
q_2	–	–	0.636 ± 0.187
a_2	–	–	0.0097 ± 0.0451

quantification is made by the standard deviation of the posterior probability distribution of the parameters around their mean value. The ground truth is indeed within the error.

Histograms of the parameters at each β under $m = m_2$ are shown in Figs. 16(e)–16(h). As we can see, at a lower temperature (higher β in the blue line), the distribution accumulates around the ground truth whereas, at the higher β (low β shown in the red line), the distribution covers the whole range of the parameters. From the high β to the low β , the distribution gradually becomes shrinking. This suggests that REMC can well sample the true parameter distribution.

The width of the histograms at the low β varies from parameter to parameter. For the wavelength q_1 , the distribution is narrow around the ground truth, as demonstrated in the small standard deviation in Table I. The mean density $\bar{\psi}$ and the spectrum amplitude at $q = q_1$ (see Fig. 15) has broader distribution, but still, there is a clear peak near the ground truth. On the other hand, the distribution of the spectrum amplitude at $q = q_0$ is very broad even at the lowest β . In fact, the error in the estimation of a_0 is comparable to its estimated value, suggesting that we have a poor estimation on a_0 . We suspect that this is because the model does not care about the value of a_0 once the mode at the wavelength q_1 becomes unstable as its amplitude $a_1 > 0$. Due to non-linear interaction between different modes, even the mode with a negative spectrum, which is linearly stable, becomes unstable and grows. Therefore, the mode at the wavelength $q_0 = 1$ may have arbitrary amplitude as long as its spectrum is close to zero that it can be destabilized by the mode at q_1 .

The estimated parameters in the one-length-scale and three-length-scale models deviate from the ground truth. One length scale is not enough to reproduce a quasicrystalline

pattern, and therefore, the estimated parameters correspond to hexagonal patterns. The three-length-scale model, on the other hand, does reproduce a quasicrystalline pattern, which is similar to the objective pattern. Therefore, the estimated parameters, particularly the wavelength q_1 and q_0 satisfy the ratio $q_0/q_1 \simeq 1.997$, which is close to $2 \cos(\pi/12)$. Nevertheless, the cost function is higher than that of the two-length-scale model because the three-length-scale model is not a true model, and the sampled pattern is quantitatively different from the objective pattern. We note that the cost function of the three-length-scale model is still lower than the one-length-scale model in which the estimation is qualitatively wrong.

2. Objective pattern synthesized by a function: Two-dimensional dodecagonal QC

The same algorithm was used for the objective pattern synthesized by functions by Eq. (19) for the two-dimensional stripe, hexagonal, and dodecagonal QC patterns, and the three-dimensional DG pattern. The results are summarized in Fig. 7.

The objective pattern of a dodecagonal QC is synthesized by the function of superposition of 12 plane waves as Eq. (19) (see Sec. VIII B). The pattern has 12-fold rotational symmetry, as demonstrated in the pattern in the Fourier space (Fig. 17). We may generate a similar pattern from the estimated PDE to the objective pattern, if not exactly the same.

The obtained histogram of the cost function is qualitatively similar to the objective pattern of the numerically produced dodecagonal QC [see Figs. 17 and 5(a)–5(c)]. The two-length-scale model has two distinct energy scales associated with QC and hexagonal patterns. At the lower cost function, we obtain

TABLE IV. The estimated parameter values for a functionally synthesized hexagonal pattern. Errors are evaluated from the standard deviation of each data.

Parameter	Estimated (one length)	Estimated (two length)	Estimated (three length)
$dx = dy$	0.990 ± 0.00778	1.007 ± 0.00642	1.03 ± 0.00518
$\bar{\psi}$	-0.226 ± 0.0536	-0.171 ± 0.0445	-0.170 ± 0.0399
a_0	0.0408 ± 0.0126	0.0438 ± 0.0341	-0.121 ± 0.0450
q_1	–	0.492 ± 0.0162	0.635 ± 0.113
a_1	–	0.0534 ± 0.00713	-0.102 ± 0.0408
q_2	–	–	0.483 ± 0.00261
a_2	–	–	0.0424 ± 0.00586

TABLE V. The estimated parameters values for functionally synthesized DG pattern in three dimensions. Errors are evaluated from the standard deviation of each sample from the REMC at the optimal temperature.

Parameter	Estimated (one length)	Estimated (two length)	Estimated (three length)
$dx = dy = dz$	0.528 ± 0.0261	0.561 ± 0.00380	0.565 ± 0.00995
$\bar{\psi}$	-0.0559 ± 0.0326	-0.090 ± 0.0266	-0.165 ± 0.0102
a_0	0.0548 ± 0.0149	0.127 ± 0.0236	0.0713 ± 0.0123
q_1	–	0.838 ± 0.0187	0.349 ± 0.188
a_1	–	0.112 ± 0.0208	-0.0604 ± 0.0336
q_2	–	–	0.855 ± 0.0156
a_2	–	–	0.146 ± 0.0398

QCs. The three-length-scale model also reproduces QCs, but the cost function is higher than that of the two-length-scale model. Therefore, the two-length-scale model will most likely give the objective pattern.

The log marginal likelihood (free energy) at each temperature, β , is calculated for each model [see Fig. 8(c)]. The models with one- and three-length scales show qualitatively the same β dependence of the log marginal likelihood with the results for numerically produced objective pattern [Fig. 5(e)]. On the other hand, the two-length-scale model $m = m_2$ shows qualitatively different β dependence, namely, there is a minimum at $\beta \approx 16$. This is because there is no ground truth in this objective pattern. Intuitively, this situation is similar to the objective pattern with noise. In both cases, the model of Eq. (5) cannot reproduce exactly the same pattern as the objective pattern. The Bayesian modeling is suitable in such cases because it gives us an optimal noise level [25]. Even though the cost function, or energy, is lower at the lower temperature, β , the posterior distribution is too narrow so that a slightly higher β than the lowest β gives the best estimate. This is exactly how our method estimates the observation noise in the objective pattern, and avoids overfitting. When the problem does not have the ground truth, there is a nonzero observation error in the objective pattern. By estimating the optimal β , we can quantify how the best-estimated model is close to the objective pattern.

The estimated parameters are shown in Table II, and the corresponding histograms for each β are shown in Fig. 18. The histograms of the sampled parameters are qualitatively similar to the results of the numerically produced objective pattern. The distribution of the estimated wave number is well accumulated around the value that is necessary to make QCs, that is, $q_1 \approx 0.51$, which is close to $1/[2 \cos(\pi/12)]$ under

$q_0 = 1$. The width of the distribution is sharp because if the ratio q_1/q_0 does not satisfy the appropriate value, the coupling between the two modes cannot occur. Other parameters do not have specific values that they should satisfy, and therefore, show broader distribution than that of the wavelength. In fact, there is a region where the dodecagonal QC can appear in the parameter space spanned by mean density $\bar{\psi}$ and the spectrum amplitude a_1 . These parameters are distributed inside the region. In particular, the mean density is a good quantity separating the dodecagonal QC from the stripe and hexagonal patterns. This is because $\bar{\psi}$ manifests the strength of coupling among three modes, that is, the coupling of two modes affects another mode in the dynamic equation [see the quadratic term in Eq. (14)]. The stripe pattern has a symmetry of $\psi \rightarrow -\psi$, and therefore, this coupling must be weak. The hexagonal pattern does not have this symmetry, and this requires stronger coupling; the dodecagonal QC requires even far stronger coupling [14]. Therefore, $|\bar{\psi}| \gg 0$. This is in agreement with the result of the estimation in Fig. 18(b). The distribution of a_0 is broad for the same reason in the case of numerically produced dodecagonal QC pattern.

3. Objective pattern synthesized by a function: Stripe and hexagonal patterns

We apply our method to stripe and hexagonal patterns. These patterns are simpler than the dodecagonal QC, and the one length scale is enough to reproduce the patterns. For the stripe pattern, the histograms of the cost function and the estimated parameters are shown in Fig. 19 and Fig. 20, respectively. For the hexagonal pattern, they are shown in Fig. 21 and Fig. 22. In contrast with the dodecagonal QCs, the cost function of these patterns can be very low up to about

TABLE VI. The estimated parameters values for functionally synthesized FK A15 pattern in three dimensions from experimental data. Errors are evaluated from the standard deviation of each sample from the REMC at the optimal temperature.

Parameter	Estimated (one length)	Estimated (two length)	Estimated (three length)
$dx = dy = dz$	0.533 ± 0.0364	0.461 ± 0.0037	0.466 ± 0.00705
$\bar{\psi}$	-0.341 ± 0.0235	-0.335 ± 0.0357	-0.384 ± 0.0361
a_0	0.0131 ± 0.005	0.101 ± 0.0483	0.118 ± 0.0377
q_1	–	0.889 ± 0.0270	0.919 ± 0.0264
a_1	–	0.0673 ± 0.0721	0.0532 ± 0.0267
q_2	–	–	0.599 ± 0.0533
a_2	–	–	-0.140 ± 0.0473

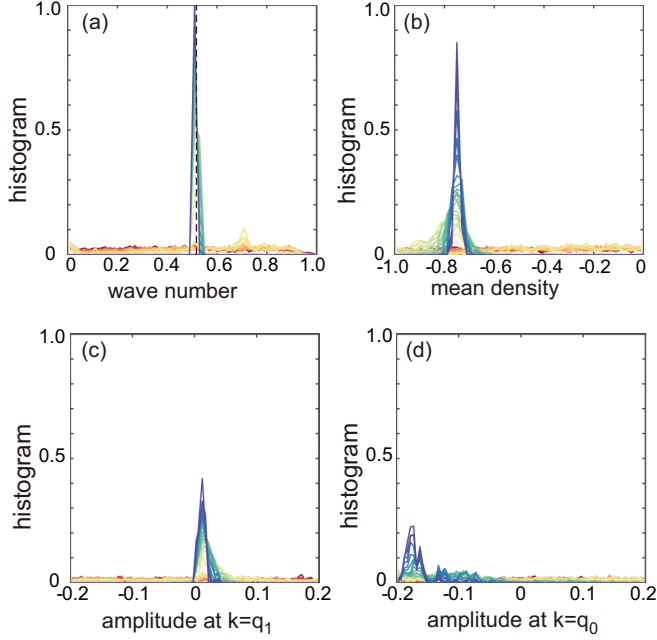


FIG. 18. Histogram of the estimated parameters for the objective pattern of dodecagonal QC that is synthesized by the function of Eq. (19). The vertical dashed line in the histogram of wave number indicates $q_1 = \sqrt{2 - \sqrt{3}}$, which is necessary to reproduce the dodecagonal pattern.

$E \approx 10^{-5} - 10^{-4}$. In addition, the shapes of the histogram under different models are qualitatively similar. Accordingly, the minimum log marginal likelihood and probability of each model are comparable among different models. Nevertheless, we estimate the one-length-scale model $m = m_1$ is the best model in both stripe and hexagonal patterns [Figs. 7(a) and 7(b)]. The dependence of the log marginal likelihood on β is indistinguishable among the three models. The marginal likelihood monotonically decreases as the temperature decreases (or β increases), and therefore, the minimum log marginal likelihood is attained at the highest β (lowest temperature). The difference of the log marginal likelihood among the three models is small, but still, by looking at marginalized probability, we find the one-length-scale model is more likely than other models, as shown in Figs. 7(a) and 7(b).

The generated patterns from the estimated PDE are similar to the objective pattern in both stripe (Fig. 19) and hexagonal patterns (Fig. 21). We stress that thanks to the order parameter Ψ , we are able to identify two patterns up to translation and rotation. In fact, the generated pattern from the estimated parameters in Fig. 19 has a different orientation.

The estimated parameters under the one-length-scale model are shown in Fig. 20 and Table III for the stripe pattern and in Fig. 22 and Table IV for the hexagonal pattern. The estimated wave number is close to that of the objective pattern $q_0 = 0.5$ in both patterns. The relevant parameter to distinguish between the stripe and hexagonal patterns is the mean density $\bar{\psi}$. It is known that the stripe pattern appears near $\bar{\psi} = 0$ because in this region, the pattern has parity symmetry, $\psi = -\psi$ [4]. The hexagonal pattern breaks this symmetry, and also it requires the three-body interaction, which is the

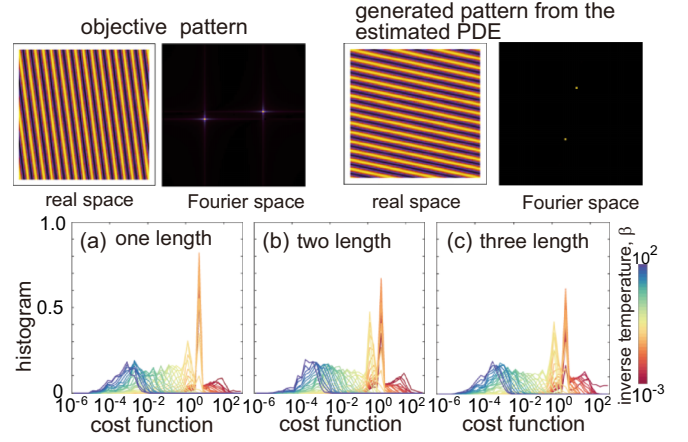


FIG. 19. Top: The objective pattern and the generated pattern from the estimated PDE in real and Fourier space. Bottom: Histograms of the cost function of the REMC sampling for the objective pattern of a stripe pattern that is synthesized by the function of Eq. (19). The horizontal axis is shown in the logarithmic scale.

quadratic term in ψ in the dynamical Eq. (5). The hexagonal pattern is expressed by the superposition of three plain waves $\psi_1 = e^{ix}$ and $\psi_{2,3} = e^{i(-\frac{1}{2}x \pm \frac{\sqrt{3}}{2}y)}$. The coupling between ψ_2 and ψ_3 gives rise to e^{-ix} , which is the complex conjugate to ψ_1 , and thus this quadratic term appears in the equation of ψ_1 . In fact, the stripe pattern appears near $\bar{\psi} = 0$, whereas the hexagonal pattern appears $\bar{\psi} \gg 0$.

4. Objective pattern synthesized by a function: Double gyroid and Frank-Kasper A15

Estimation of a PDE that generates a three-dimensional pattern is performed similarly to a two-dimensional pattern. The only difference is the definition of order parameter Ψ . As discussed in Sec. VIII B 2, three-dimensional patterns may have several invariants for each l . We expect including all the invariants is necessary to classify complex patterns, but in this study, we consider only one invariant similar to the two-dimensional problem. We found this assumption works at least for the patterns that we study here: DG and FK A15. We have confirmed our method also works for a simpler pattern such as lamellar pattern (stripes in three dimensions) and cylindrical pattern (hexagonal pattern in three dimensions).

In Fig. 23 the histogram of the cost function of each model is shown. The DG pattern is reproduced by all models ($m =$

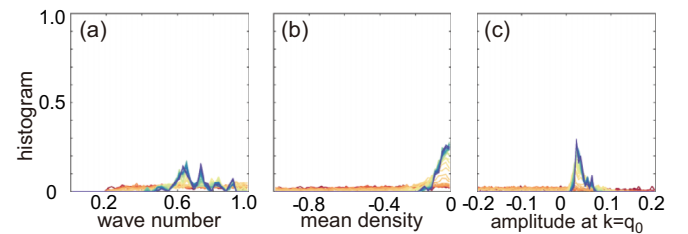


FIG. 20. Histogram of the estimated parameters for the objective pattern of a stripe pattern that is synthesized by the function of Eq. (19).

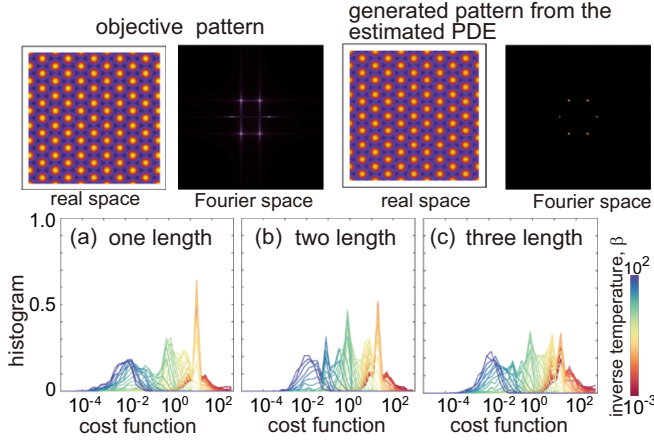


FIG. 21. Top: The objective pattern and the generated pattern from the estimated PDE in real and Fourier space. Bottom: Histograms of the cost function of the REMC sampling for the objective pattern of a hexagonal pattern that is synthesized by the function of Eq. (19). The horizontal axis is shown in the logarithmic scale.

m_1, m_2, m_3). In fact, the generated pattern from the estimated PDE is identical to the objective pattern up to translation, and they have almost the same patterns in the Fourier space (Fig. 23). The histograms of the cost function of all models are qualitatively similar, but $m = m_2$ has slightly smaller energy on average. The probability of the two models shown in Fig. 7(d) is comparable, but $m = m_2$ is chosen. The log marginal likelihood decreases as the temperature decreases (β increases), and therefore, within the current choice of the range of β , the minimum of the log marginal likelihood is not attained. If we use the temperature range containing larger β , the log marginal likelihood may attain its minimum.

Estimated parameters are shown in Fig. 24 and Table V. The estimated wave number distributes near the wavelength that we have imposed. The mean density $\bar{\psi}$ is accumulated near $\bar{\psi} = 0$ but with a slight deviation. This is consistent with previous theoretical and numerical results; the DG pattern, in fact, appears between lamellar (stripe) and cylinder (hexagonal) patterns [33,34,36]. As discussed in Appendixes C2 and C3, the stripe pattern appears near $\bar{\psi} = 0$ including $\bar{\psi} = 0$, whereas the hexagonal pattern appears at $|\bar{\psi}| \gg 0$. The region where the DG pattern appears is between the two regions, and thus the mean density should be $1 \gg |\bar{\psi}| > 0$. This is exactly observed in Fig. 24(b).

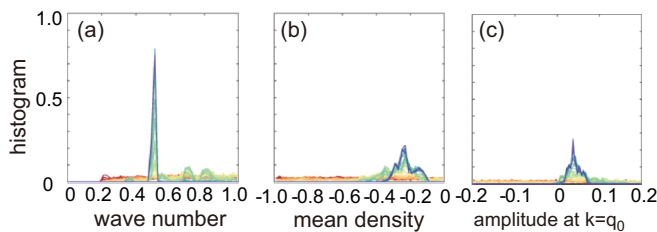


FIG. 22. Histogram of the estimated parameters for the objective pattern of a hexagonal pattern that is synthesized by the function of Eq. (19).

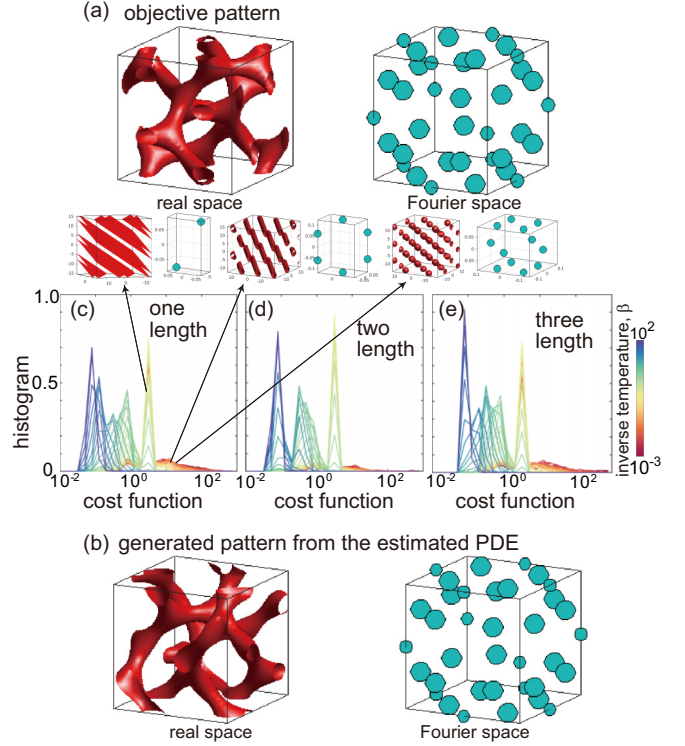


FIG. 23. (a), (b) The objective pattern and the generated pattern from the estimated PDE in real $\psi(\mathbf{x})$ and Fourier $|\hat{\psi}(\mathbf{k})|$ space for DG patterns. In Fourier space, each point corresponds to a peak of $|\hat{\psi}(\mathbf{k})|$, and the intensity is described by the size of the points. (c)–(e) Histogram of the cost function of the REMC sampling for the objective pattern of a DG pattern that is synthesized by the function as Eq. (19). The horizontal axis is shown in the logarithmic scale. Patterns with higher energy are also shown in the insets.

The FK A15 pattern is expressed by 24 wave vectors $\mathbf{q}^* = (\pm 2, \pm 1, 0)$, 24 wave vectors $\mathbf{q}^* = (\pm 2, \pm 1, \pm 1)$, six wave vectors of $\mathbf{q}^* = (\pm 2, 0, 0)$ with their permutation along the x, y, z directions [58]. In the unit cell, a center particle is surrounded by eight particles at the position of the corners, and two particles are located at each face [Fig. 25(a)]. In Fig. 25 the histogram of the cost function for each model is shown. The FK A15 pattern is reproduced by all models ($m = m_1, m_2, m_3$). The generated pattern from the estimated PDE is similar to the objective pattern up to translation, and they have almost the same patterns in the Fourier space [Fig. 25(b)]. The similarity is also demonstrated in their side views. The histograms of the cost function are shown in

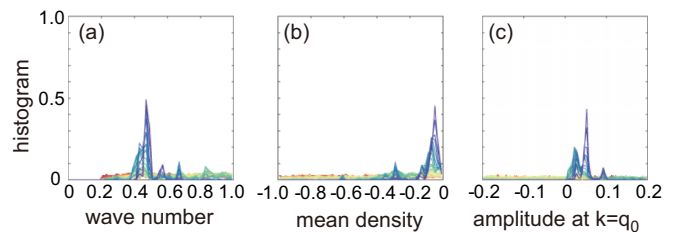


FIG. 24. Histogram of the estimated parameters for the objective pattern of a DG pattern that is synthesized by the function of Eq. (19).

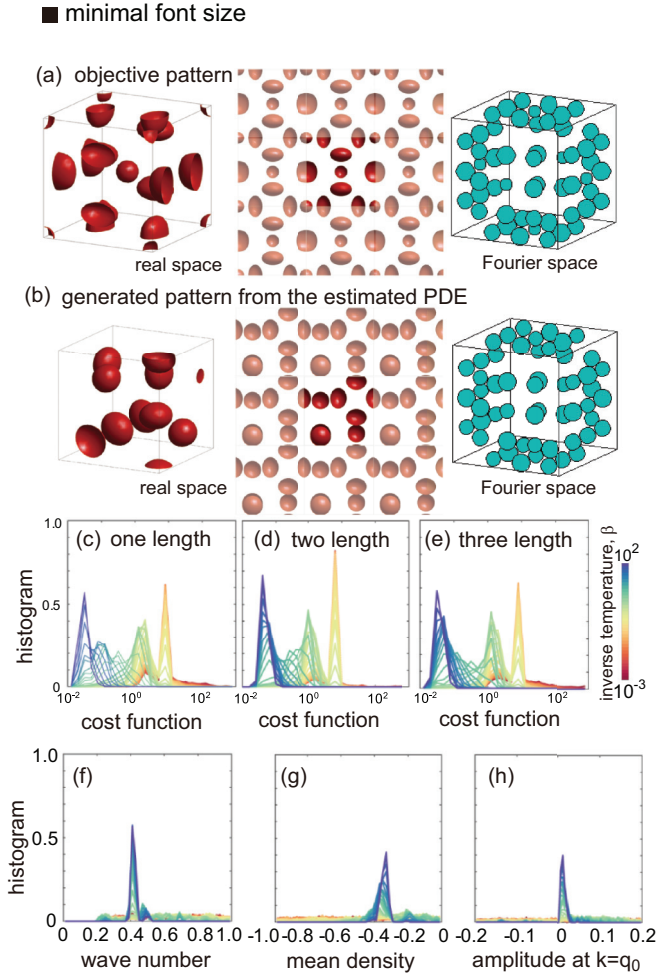


FIG. 25. (a), (b) The objective (a) and estimated (b) patterns in real $\psi(\mathbf{x})$ and Fourier $|\psi(\mathbf{k})|$ space for the objective pattern of FK A15. In Fourier space, each point corresponds to a peak of $|\psi(\mathbf{k})|$, and the intensity is described by the size of the points. The side view of the real-space images is also shown in the middle panels. In the side view, neighboring domains of the periodic boundary are added around the main domain to clarify periodicity. (c)–(e) Histogram of the cost function in the REMC sampling for the objective pattern of a FK A15 pattern that is synthesized by the function of Eq. (19). The horizontal axis is shown in the logarithmic scale. (f)–(h) Histogram of the estimated parameters for the objective pattern of a FK A15 pattern that is synthesized by the function.

Figs. 25(c)–25(e). The probability of the two models shown in Fig. 7(e) is comparable, but $m = m_1$ is chosen for this objective pattern. The log marginal likelihood decreases as the temperature decreases, and therefore, within the current choice of the range of temperature, the minimum of the log marginal likelihood is not attained. If we use the temperature range containing larger β , the log marginal likelihood may attain its minimum.

Estimated parameters are shown in Figs. 25(f)–25(h) and Table VI. The estimated wave number distributes near the wavelength that we have imposed $q = 0.5$. In contrast with the DG pattern, the mean density $\bar{\psi}$ is away from $\bar{\psi} = 0$. In fact, FK A15 has been found in higher resolution of $\bar{\psi} - a_0$

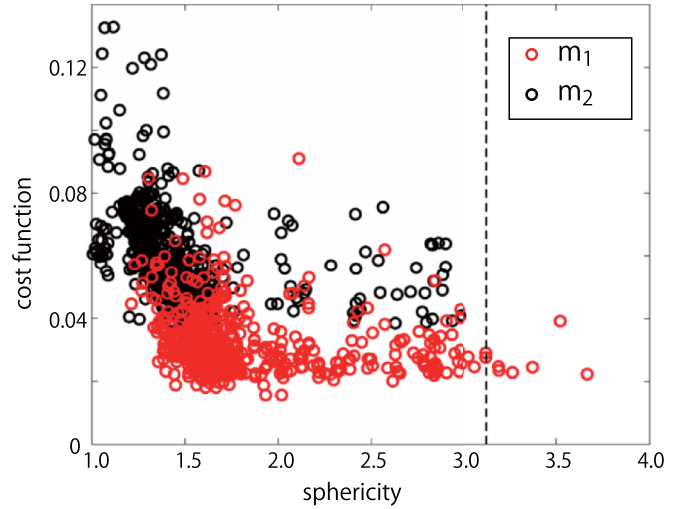


FIG. 26. The correlation between sphericity and the cost function for the generated patterns using the estimated parameters for one-length-scale model m_1 and two-length-scale model m_2 . The sphericity of the objective pattern is shown in a vertical dashed line.

phase diagram using the self-consistent field theory, which describes block copolymers [9]. The position of FK A15 in the phase diagram used to be BCC, but recently several FK phases have been found in this region. The BCC pattern has been found at $\bar{\psi} \gg 0$, and thus our estimated $\bar{\psi}$ is consistent with the observation. We note that, to our knowledge, FK A15 has not been reported in the framework of PFC. Because PFC is considered as an approximation of various model, including self-consistent field theory and density functional theory of atomic alloy, this pattern should appear in various pattern-forming systems.

Each domain of the objective pattern of FK A15 is deformed [Fig. 25(a)]. Figure 25(b) shows that the generated patterns using the estimated parameters are also deformed. To see the deviation from a spherical shape quantitatively, we extract the center of each domain in the density field

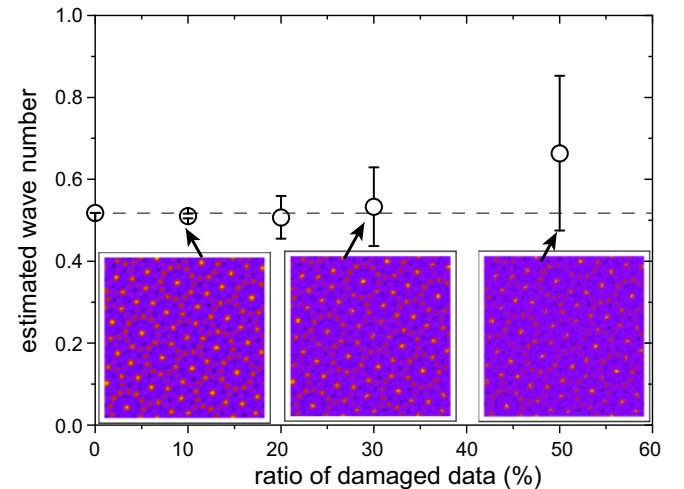


FIG. 27. Estimated wave numbers for numerically generated QC by BM-PDE equation (2) for the damaged objective pattern.

$\psi(\mathbf{x})$ of the generated patterns. The density field is binarized using the threshold of half of the maximum density. Then, the connected components are extracted from the binary field, and the center of each connected component is computed. We also compute a gyration tensor of each domain, from which we obtain three eigenvalues associated with lengths of three axes of an ellipsoid. Sphericity is defined by the ratio between the maximum and minimum eigenvalues. When the domain is a spherical shape, sphericity is one, whereas it is larger than one when the shape is deformed. Figure 26 shows the correlation between sphericity and the cost function. The one-length-scale model m_1 generates patterns that are more deformed than those generated by the two-length-scale model

m_2 . Accordingly, the cost functions of the generated patterns by the one-length-scale model are lower.

APPENDIX D: DAMAGED OBJECTIVE PATTERN

In addition to the estimation for the objective pattern with noise in the text, we perform the estimation for the damaged objective pattern. The damaged pattern is made by setting $\psi = 0$ at randomly distributed spatial points. We vary the fraction of the damaged point from 0 to 50%, and estimate the wave number, which is the most relevant parameter to reproduce dodecagonal QC. Figure 27 shows that up to 30% of the damage, the correct wave number can be estimated.

-
- [1] J. Swift and P. C. Hohenberg, Hydrodynamic fluctuations at the convective instability, *Phys. Rev. A* **15**, 319 (1977).
- [2] F. S. Bates and G. H. Fredrickson, Block copolymer thermodynamics: Theory and experiment, *Annu. Rev. Phys. Chem.* **41**, 525 (1990).
- [3] C. Harrison, D. H. Adamson, Z. Cheng, J. M. Sebastian, S. Sethuraman, D. A. Huse, R. A. Register, and P. M. Chaikin, Mechanisms of ordering in striped patterns, *Science* **290**, 1558 (2000).
- [4] K. R. Elder, M. Katakowski, M. Haataja, and M. Grant, Modeling Elasticity in Crystal Growth, *Phys. Rev. Lett.* **88**, 245701 (2002).
- [5] N. Provatas and K. Elder, *Phase-Field Methods in Materials Science and Engineering* (John Wiley & Sons, Weinheim, 2011).
- [6] M. C. Cross and P. C. Hohenberg, Pattern formation outside of equilibrium, *Rev. Mod. Phys.* **65**, 851 (1993).
- [7] F. C. Frank and J. S. Kasper, Complex alloy structures regarded as sphere packings. II. Analysis and classification of representative structures, *Acta Crystallogr.* **12**, 483 (1959).
- [8] D. Shechtman, I. Blech, D. Gratias, and J. W. Cahn, Metallic Phase with Long-Range Orientational Order and No Translational Symmetry, *Phys. Rev. Lett.* **53**, 1951 (1984).
- [9] M. W. Bates, J. Lequeieu, S. M. Barbon, R. M. Lewis, K. T. Delaney, A. Anastasaki, C. J. Hawker, G. H. Fredrickson, and C. M. Bates, Stability of the a15 phase in diblock copolymer melts, *Proc. Natl. Acad. Sci. USA* **116**, 13194 (2019).
- [10] M. R. J. Scherer, *Double-Gyroid-Structured Functional Materials: Synthesis and Applications* (Springer, Heidelberg, 2013).
- [11] K. Yue, M. Huang, R. L. Marson, J. He, J. Huang, Z. Zhou, J. Wang, C. Liu, X. Yan, K. Wu, Z. Guo, H. Liu, W. Zhang, P. Ni, C. Wedemiotis, W.-B. Zhang, S. C. Glotzer, and S. Z. D. Cheng, Geometry induced sequence of nanoscale Frank–Kasper and quasicrystal mesophases in giant surfactants, *Proc. Natl. Acad. Sci. USA* **113**, 14195 (2016).
- [12] X. Zeng, G. Ungar, Y. Liu, V. Percec, A. E. Dulcey, and J. K. Hobbs, Supramolecular dendritic liquid quasicrystals, *Nature (London)* **428**, 157 (2004).
- [13] A.-P. Hynninen, J. H. J. Thijssen, E. C. M. Vermolen, M. Dijkstra, and A. van Blaaderen, Self-assembly route for photonic crystals with a bandgap in the visible region, *Nat. Mater.* **6**, 202 (2007).
- [14] R. Lifshitz and D. M. Petrich, Theoretical Model for Faraday Waves with Multiple-Frequency Forcing, *Phys. Rev. Lett.* **79**, 1261 (1997).
- [15] K. Barkan, M. Engel, and R. Lifshitz, Controlled Self-assembly of Periodic and Aperiodic Cluster Crystals, *Phys. Rev. Lett.* **113**, 098304 (2014).
- [16] S. V. Kalinin, B. G. Sumpter, and R. K. Archibald, Big-deep-smart data in imaging for guiding materials design, *Nat. Mater.* **14**, 973 (2015).
- [17] M. C. Kennedy and A. O’Hagan, Bayesian calibration of computer models, *J. R. Stat. Soc. B* **63**, 425 (2001).
- [18] R. C. Smith, *Uncertainty Quantification: Theory, Implementation, and Applications*, the Computational Science and Engineering Series Vol. 12 (SIAM, Philadelphia, 2013).
- [19] G. Evensen, *Data Assimilation: The Ensemble Kalman Filter* (Springer, Berlin, 2009).
- [20] K. Law, A. Stuart, and K. Zygalakis, *Data Assimilation* (Springer, Cham, 2015).
- [21] D. J. C. MacKay, Bayesian interpolation, *Neural Comput.* **4**, 415 (1992).
- [22] C. M. Bishop, *Pattern Recognition and Machine Learning*, Vol. 1 (Springer, New York, 2006).
- [23] L. Wasserman, Bayesian model selection and model averaging, *J. Math. Psychol.* **44**, 92 (2000).
- [24] J. O. Berger, L. R. Pericchi, J. K. Ghosh, T. Samanta, F. De Santis, J. O. Berger, and L. R. Pericchi, Objective Bayesian methods for model selection: Introduction and comparison, in *Lecture Notes-Monograph Series 38 Model Selection*, edited by P. Lahiri (Institute of Mathematical Statistics, 2001), pp. 135–207.
- [25] S. Tokuda, K. Nagata, and M. Okada, Simultaneous estimation of noise variance and number of peaks in Bayesian spectral deconvolution, *J. Phys. Soc. Jpn.* **86**, 024001 (2017).
- [26] A. Jayaraman and M. K. Mahanthappa, Counterion-dependent access to low-symmetry lyotropic sphere packings of ionic surfactant micelles, *Langmuir* **34**, 2290 (2018).
- [27] C. J. Geyer, Markov chain Monte Carlo maximum likelihood, in *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface* (American Statistical Association, 1991), p. 156.
- [28] K. Hukushima and K. Nemoto, Exchange Monte Carlo method and application to spin glass simulations, *J. Phys. Soc. Jpn.* **65**, 1604 (1996).
- [29] C. V. Achim, M. Schmiedeberg, and H. Löwen, Growth Modes of Quasicrystals, *Phys. Rev. Lett.* **112**, 255501 (2014).
- [30] K. R. Elder and M. Grant, Modeling elastic and plastic deformations in nonequilibrium processing using phase field crystals, *Phys. Rev. E* **70**, 051605 (2004).

- [31] M. W. Matsen and M. Schick, Microphases of a diblock copolymer with conformational asymmetry, *Macromolecules* **27**, 4014 (1994).
- [32] M. Nonomura and T. Ohta, Kinetics of morphological transitions between mesophases, *J. Phys.: Condens. Matter* **13**, 9089 (2001).
- [33] V. E. Podneks and I. W. Hamley, Landau-Brazovskii theory for the $ia\bar{3}d$ structure, *J. Exp. Theor. Phys. Lett.* **64**, 617 (1996).
- [34] A.-C. Shi, Nature of anisotropic fluctuation modes in ordered systems, *J. Phys.: Condens. Matter* **11**, 10183 (1999).
- [35] P. Zhang and X. Zhang, An efficient numerical method of Landau-Brazovskii model, *J. Comput. Phys.* **227**, 5859 (2008).
- [36] A. Jaatinen and T. Ala-Nissila, Extended phase diagram of the three-dimensional phase field crystal model, *J. Phys.: Condens. Matter* **22**, 205402 (2010).
- [37] S. Lee, C. Leighton, and F. S. Bates, Sphericity and symmetry breaking in the formation of Frank-Kasper phases from one component materials, *Proc. Natl. Acad. Sci. USA* **111**, 17723 (2014).
- [38] H. U. Voss, J. Timmer, and J. Kurths, Nonlinear dynamical system identification from uncertain and indirect measurements, *Int. J. Bifurcation Chaos* **14**, 1905 (2004).
- [39] H. Zhao, B. D. Storey, R. D. Braatz, and M. Z. Bazant, Learning the Physics of Pattern Formation from Images, *Phys. Rev. Lett.* **124**, 060201 (2020).
- [40] H. Zhao, R. D. Braatz, and M. Z. Bazant, Image inversion and uncertainty quantification for constitutive laws of pattern formation, *J. Comput. Phys.* **436**, 110279 (2021).
- [41] R. Tibshirani, Regression shrinkage and selection via the lasso, *J. R. Stat. Soc. B* **58**, 267 (1996).
- [42] P. Subramanian, A. J. Archer, E. Knobloch, and A. M. Rucklidge, Three-Dimensional Icosahedral Phase Field Quasicrystal, *Phys. Rev. Lett.* **117**, 075501 (2016).
- [43] P. F. Damasceno, S. C. Glotzer, and M. Engel, Non-close-packed three-dimensional quasicrystals, *J. Phys.: Condens. Matter* **29**, 234005 (2017).
- [44] C. R. Iacovella, A. S. Keys, and S. C. Glotzer, Self-assembly of soft-matter quasicrystals and their approximants, *Proc. Natl. Acad. Sci. USA* **108**, 20935 (2011).
- [45] M. Dzugutov, Formation of a dodecagonal quasicrystalline phase in a simple monatomic liquid, *Phys. Rev. Lett.* **70**, 2924 (1993).
- [46] M. Dzugutov, Phason dynamics and atomic transport in an equilibrium dodecagonal quasi-crystal, *Europhys. Lett.* **31**, 95 (1995).
- [47] B. C. Daniels and I. Nemenman, Automated adaptive inference of phenomenological dynamical models, *Nat. Commun.* **6**, 8133 (2015).
- [48] S. L. Brunton, J. L. Proctor, and J. N. Kutz, Discovering governing equations from data by sparse identification of nonlinear dynamical systems, *Proc. Natl. Acad. Sci. USA* **113**, 3932 (2016).
- [49] S. L. Brunton and J. N. Kutz, *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control* (Cambridge University Press, Cambridge, 2019).
- [50] T. G. Müller and J. Timmer, Parameter identification techniques for partial differential equations, *Int. J. Bifurcation Chaos* **14**, 2053 (2004).
- [51] S. H. Rudy, S. L. Brunton, J. L. Proctor, and J. N. Kutz, Data-driven discovery of partial differential equations, *Sci. Adv.* **3**, e1602614 (2017).
- [52] L. M. Pismen, *Patterns and Interfaces in Dissipative Dynamics* (Springer, Berlin, 2006).
- [53] S. M. Cox and P. C. Matthews, Exponential time differencing for stiff systems, *J. Comput. Phys.* **176**, 430 (2002).
- [54] Igor S. Aranson, editor, *Physical Models of Cell Motility* (Springer, Cham, 2015).
- [55] S. Geman and D. Geman, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE PAMI* **6**, 721 (1984).
- [56] H. G. von Schnering and R. Nesper, Nodal surfaces of fourier series: Fundamental invariants of structured matter, *Z. Phys. B* **83**, 407 (1991).
- [57] K. Yamada, M. Nonomura, and T. Ohta, Kinetics of morphological transitions in microphase-separated diblock copolymers, *Macromolecules* **37**, 5762 (2004).
- [58] M. Impéror-Clerc, Three-dimensional periodic complex structures in soft matter: Investigation using scattering methods, *Interface Focus* **2**, 589 (2012).
- [59] B. Efron and C. Morris, Stein's estimation rule and its competitors—An empirical Bayes approach, *J. Am. Stat. Assoc.* **68**, 117 (1973).
- [60] H. Akaike, Likelihood and the Bayes procedure, in *Selected Papers of Hirotugu Akaike*, edited by E. Parzen, K. Tanabe, and G. Kitagawa (Springer, New York, 1998), pp. 309–332.
- [61] X.-L. Meng and W. H. Wong, Simulating ratios of normalizing constants via a simple identity: A theoretical exploration, *Statistica Sinica* **6**, 831 (1996).
- [62] A. Gelman and X.-L. Meng, Simulating normalizing constants: From importance sampling to bridge sampling to path sampling, *Stat. Sci.* **13**, 163 (1998).
- [63] B. Efron, Bootstrap methods: Another look at the jackknife, *Ann. Statist.* **7**, 1 (1979).
- [64] H. Schaeffer, Learning partial differential equations via data discovery and sparse optimization, *Proc. R. Soc. A* **473**, 20160446 (2017).
- [65] Lord Rayleigh, LIX. on convection currents in a horizontal layer of fluid, when the higher temperature is on the under side, *Lond. Edinb. Dubl. Phil. Mag. J. Sci.* **32**, 529 (1916).
- [66] G. I. Taylor, VIII. stability of a viscous liquid contained between two rotating cylinders, *Philos. Trans. R. Soc. Lond. A* **223**, 289 (1923).
- [67] J. T. Stuart, On the non-linear mechanics of hydrodynamic stability, *J. Fluid Mech.* **4**, 1 (1958).
- [68] J. Watson, On the non-linear mechanics of wave disturbances in stable and unstable parallel flows part 2. The development of a solution for plane poiseuille flow and for plane Couette flow, *J. Fluid Mech.* **9**, 371 (1960).
- [69] J. Lega, J. V. Moloney, and A. C. Newell, Swift-Hohenberg Equation for Lasers, *Phys. Rev. Lett.* **73**, 2978 (1994).
- [70] R. Lefever, N. Barbier, P. Couteron, and O. Lejeune, Deeply gapped vegetation patterns: On crown/root allometry, criticality and desertification, *J. Theor. Biol.* **261**, 194 (2009).
- [71] T. Ohta and K. Kawasaki, Equilibrium morphology of block copolymer melts, *Macromolecules* **19**, 2621 (1986).
- [72] M. Bahiana and Y. Oono, Cell dynamical system approach to block copolymers, *Phys. Rev. A* **41**, 6763 (1990).

- [73] G. H. Fredrickson and E. Helfand, Fluctuation effects in the theory of microphase separation in block copolymers, *J. Chem. Phys.* **87**, 697 (1987).
- [74] M. Kléman, *Points, Lines and Walls in Liquid Crystals, Magnetic Systems, and Various Disordered Media* (Wiley, New York, 1983).
- [75] M. Kléman and O. D. Lavrentovich, *Soft Matter Physics: An Introduction* (Springer, New York, 2003).
- [76] K. R. Elder, N. Provatas, J. Berry, P. Stefanovic, and M. Grant, Phase-field crystal modeling and classical density functional theory of freezing, *Phys. Rev. B* **75**, 064107 (2007).
- [77] S. Savitz, M. Babadi, and R. Lifshitz, Multiple-scale structures: From Faraday waves to soft-matter quasicrystals, *IUCrJ* **5**, 247 (2018).
- [78] H. W. Müller, Model equations for two-dimensional quasipatterns, *Phys. Rev. E* **49**, 1273 (1994).
- [79] N. D. Mermin and S. M. Troian, Mean-Field Theory of Quasicrystalline Order, *Phys. Rev. Lett.* **54**, 1524 (1985).
- [80] E. I. Kats, V. V. Lebedev, and A. R. Muratov, Weak crystallization theory, *Phys. Rep.* **228**, 1 (1993).
- [81] B. A. Malomed, A. A. Nepomnyashchii, and M. I. Tribelskii, Two-dimensional quasiperiodic structures in nonequilibrium systems, *Sov. Phys. JETP* **69**, 388 (1988).
- [82] Z. Zhang and C. Glotzer, Sharon, Self-assembly of patchy particles, *Nano Lett.* **4**, 1407 (2004).
- [83] U. T. Lieu and N. Yoshinaga, Topological defects of dipole patchy particles on a spherical surface, *Soft Matter* **16**, 7667 (2020).
- [84] P. F. Damasceno, M. Engel, and S. C. Glotzer, Predictive self-assembly of polyhedra into complex structures, *Science* **337**, 453 (2012).
- [85] G. M. Whitesides and B. Grzybowski, Self-assembly at all scales, *Science* **295**, 2418 (2002).
- [86] F. Li, D. P. Josephson, and A. Stein, Colloidal assembly: The road from particles to colloidal molecules and crystals, *Angew. Chem. Int. Ed.* **50**, 360 (2011).
- [87] M. A. Boles, M. Engel, and D. V. Talapin, Self-assembly of colloidal nanocrystals: From intricate structures to functional materials, *Chem. Rev.* **116**, 11220 (2016).
- [88] M. C. Rechtsman, F. H. Stillinger, and S. Torquato, Optimized Interactions for Targeted Self-assembly: Application to a Honeycomb Lattice, *Phys. Rev. Lett.* **95**, 228301 (2005).
- [89] M. Rechtsman, F. Stillinger, and S. Torquato, Designed interaction potentials via inverse methods for self-assembly, *Phys. Rev. E* **73**, 011406 (2006).
- [90] S. Torquato, Inverse optimization techniques for targeted self-assembly, *Soft Matter* **5**, 1157 (2009).
- [91] S. Torquato, Optimal design of heterogeneous materials, *Annu. Rev. Mater. Res.* **40**, 101 (2010).
- [92] A. Jain, J. A. Bollinger, and T. M. Truskett, Inverse methods for material design, *AIChE J.* **60**, 2732 (2014).
- [93] Z. M. Sherman, M. P. Howard, B. A. Lindquist, R. B. Jadrich, and T. M. Truskett, Inverse methods for design of soft materials, *J. Chem. Phys.* **152**, 140902 (2020).
- [94] B. A. Lindquist, R. B. Jadrich, and T. M. Truskett, Communication: Inverse design for self-assembly via on-the-fly optimization, *J. Chem. Phys.* **145**, 111101 (2016).
- [95] U. T. Lieu and N. Yoshinaga, Inverse design of two-dimensional structure by self-assembly of patchy particles, *J. Chem. Phys.* **156**, 054901 (2022).
- [96] C. L. Tsai, K. T. Delaney, and G. H. Fredrickson, Genetic algorithm for discovery of globally stable phases in block copolymers, *Macromolecules* **49**, 6558 (2016).
- [97] S. P. Paradiso, K. T. Delaney, and G. H. Fredrickson, Swarm intelligence platform for multiblock polymer inverse formulation design, *ACS Macro Lett.* **5**, 972 (2016).
- [98] M. R. Khadilkar, S. Paradiso, K. T. Delaney, and G. H. Fredrickson, Inverse design of bulk morphologies in multiblock polymers using particle swarm optimization, *Macromolecules* **50**, 6702 (2017).
- [99] G. S. Khaira, J. Qin, G. P. Garner, S. Xiong, L. Wan, R. Ruiz, H. M. Jaeger, P. F. Nealey, and J. J. de Pablo, Evolutionary optimization of directed self-assembly of triblock copolymers on chemically patterned substrates, *ACS Macro Lett.* **3**, 747 (2014).
- [100] R. Kumar, G. M. Coli, M. Dijkstra, and S. Sastry, Inverse design of charged colloidal particle interactions for self assembly into specified crystal structures, *J. Chem. Phys.* **151**, 084109 (2019).
- [101] M. Z. Miskin, G. Khaira, J. J. de Pablo, and H. M. Jaeger, Turning statistical physics models into materials design engines, *Proc. Natl. Acad. Sci. USA* **113**, 34 (2016).
- [102] B. A. Lindquist, R. B. Jadrich, and T. M. Truskett, Assembly of nothing: Equilibrium fluids with designed structured porosity, *Soft Matter* **12**, 2663 (2016).
- [103] W. D. Piñeros, B. A. Lindquist, R. B. Jadrich, and T. M. Truskett, Inverse design of multicomponent assemblies, *J. Chem. Phys.* **148**, 104509 (2018).
- [104] W. D. Piñeros, R. B. Jadrich, and T. M. Truskett, Design of two-dimensional particle assemblies using isotropic pair interactions with an attractive well, *AIP Adv.* **7**, 115307 (2017).
- [105] C. S. Adorf, J. Antonaglia, J. Dshemuchadse, and S. C. Glotzer, Inverse design of simple pair potentials for the self-assembly of complex structures, *J. Chem. Phys.* **149**, 204102 (2018).
- [106] B. A. Lindquist, R. B. Jadrich, and T. M. Truskett, Communication: From close-packed to topologically close-packed: Formation of laves phases in moderately polydisperse hard-sphere mixtures, *J. Chem. Phys.* **148**, 191101 (2018).
- [107] H. M. Jaeger and J. J. de Pablo, Perspective: Evolutionary design of granular media and block copolymer patterns, *APL Mater.* **4**, 053209 (2016).
- [108] M. Bär, R. Hegger, and H. Kantz, Fitting partial differential equations to space-time dynamics, *Phys. Rev. E* **59**, 337 (1999).
- [109] J. Bongard and H. Lipson, Automated reverse engineering of nonlinear dynamical systems, *Proc. Natl. Acad. Sci. USA* **104**, 9943 (2007).
- [110] M. Schmidt and H. Lipson, Distilling free-form natural laws from experimental data, *Science* **324**, 81 (2009).
- [111] M. Cranmer, A. Sanchez Gonzalez, P. Battaglia, R. Xu, K. Cranmer, D. Spergel, and S. Ho, Discovering symbolic models from deep learning with inductive biases, *Adv. Neural Inf. Proc. Syst.* **33**, 17429 (2020).
- [112] H. Schaeffer and S. G. McCalla, Sparse model selection via integral terms, *Phys. Rev. E* **96**, 023302 (2017).
- [113] H. Xu, H. Chang, and D. Zhang, DL-PDE: Deep-learning based data-driven discovery of partial differential equations

- from discrete and noisy data, *Commun. Comput. Phys.* **29**, 698 (2021).
- [114] J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*, Vol. 160 in Applied Mathematical Sciences (Springer Science & Business Media, New York, 2006).
- [115] X. Xun, J. Cao, B. Mallick, A. Maity, and R. J. Carroll, Parameter estimation of partial differential equation models, *J. Am. Stat. Assoc.* **108**, 1009 (2013).
- [116] E. Campillo-Funollet, C. Venkataraman, and A. Madzvamuse, Bayesian parameter identification for Turing systems on stationary and evolving domains, *Bull. Math. Biol.* **81**, 81 (2019).
- [117] A. Kazarnikov and H. Haario, Statistical approach for parameter identification by Turing patterns, *J. Theor. Biol.* **501**, 110319 (2020).
- [118] T. Toni, D. Welch, N. Strelkowa, A. Ipsen, and M. P. H. Stumpf, Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems, *J. R. Soc. Interface* **6**, 187 (2009).
- [119] A. Carrassi, M. Bocquet, A. Hannart, and M. Ghil, Estimating model evidence using data assimilation, *Q. J. R. Meteorol. Soc.* **143**, 866 (2017).
- [120] S.-i. Ito, H. Nagao, T. Kurokawa, T. Kasuya, and J. Inoue, Bayesian inference of grain growth prediction via multi-phase-field models, *Phys. Rev. Mater.* **3**, 053404 (2019).
- [121] Y. Yuan, J. Li, L. Li, F. Jiang, X. Tang, F. Zhang, S. Liu, J. Goncalves, H. U. Voss, X. Li, J. Kurths, and H. Ding, Machine discovery of partial differential equations from spatiotemporal data, [arXiv:1909.06730](https://arxiv.org/abs/1909.06730).
- [122] A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation* (Society for Industrial and Applied Mathematics, Philadelphia, 2005).
- [123] L. Tenorio, *An Introduction to Data Analysis and Uncertainty Quantification for Inverse Problems*, in Mathematics in Industry, Vol. 3 (SIAM, 2017).
- [124] J. M. Varah, A spline least squares method for numerical parameter estimation in differential equations, *SIAM J. Sci. Stat. Comput.* **3**, 28 (1982).
- [125] J. O. Ramsay, G. Hooker, D. Campbell, and J. Cao, Parameter estimation for differential equations: A generalized smoothing approach, *J. R. Soc. Soc. B* **69**, 741 (2007).
- [126] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning* (MIT Press, Cambridge, MA, 2003).
- [127] T. Graepel, Solving noisy linear operator equations by Gaussian processes: Application to ordinary and partial differential equations, in *Proceedings of the Twentieth International Conference on International Conference on Machine Learning, ICML'03* (AAAI Press, USA, 2003), pp. 234–241.
- [128] M. A. Álvarez, D. Luengo, and N. D. Lawrence, Latent force models, in *Proceedings of the International Conference on Artificial Intelligence and Statistics*, edited by D. van Dyk and M. Welling (JMLR, Clearwater Beach, Florida, 2009), pp. 9–16.
- [129] A. Melkumyan, Operator induced multi-task Gaussian processes for solving differential equations, in *Neural Information Processing Systems (NIPS) Workshop: New Directions in Multiple Kernel Learning* (2012).
- [130] S. Atkinson and N. Zabaras, Structured Bayesian Gaussian process latent variable model: Applications to data-driven dimensionality reduction and high-dimensional inversion, *J. Comput. Phys.* **383**, 166 (2019).
- [131] B. Calderhead, M. Girolami, and N. D. Lawrence, Accelerating Bayesian inference over nonlinear differential equations with gaussian processes, in *Advances in Neural Information Processing Systems*, edited by D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, Vol. 21 (Curran Associates, Inc., 2008), pp. 217–224.
- [132] F. Dondelinger, D. Husmeier, S. Rogers, and M. Filippone, ODE parameter inference using adaptive gradient matching with Gaussian processes, *J. Machine Learn. Res.* **31**, 216 (2013).
- [133] N. S. Gorbach, S. Bauer, and J. M. Buhmann, Scalable variational inference for dynamical systems, *Adv. Neural Inf. Process Syst.* **30**, 4806 (2017).
- [134] M. Raissi, P. Perdikaris, and G. Karniadakis, Numerical Gaussian processes for time-dependent and nonlinear partial differential equations, *SIAM J. Sci. Comput.* **40**, A172 (2018).
- [135] E. Baake, M. Baake, H. G. Bock, and K. M. Briggs, Fitting ordinary differential equations to chaotic data, *Phys. Rev. A* **45**, 5524 (1992).
- [136] T. G. Müller and J. Timmer, Fitting parameters in partial differential equations from partially observed noisy data, *Physica D* **171**, 1 (2002).
- [137] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing* (Springer Science & Business Media, New York, 2013).
- [138] H. Zou and T. Hastie, Regularization and variable selection via the elastic net, *J. R. Stat. Soc. B* **67**, 301 (2005).