

Bayesian Modelling of Pattern Formation from One Snapshot of Pattern

Natsuhiko Yoshinaga*

*WPI - Advanced Institute for Materials Research,
Tohoku University, Sendai 980-8577, Japan and
MathAM-OIL, AIST, Sendai 980-8577, Japan*

Satoru Tokuda

*Research Institute for Information Technology, Kyushu University, Kasuga 816-8580, Japan and
MathAM-OIL, AIST, Sendai 980-8577, Japan*

CONTENTS

SI. Related works	1
A. Pattern formation and crystalline structure	1
B. Inverse structural optimisation	2
C. Symbolic regression of governing equations	3
D. State noise and measurement noise	3
SII. Family of models	4
A. Phase-field crystal model and its generalisation	4
B. Polynomial expansion of the linear operation with respect to wavenumbers	5
SIII. Order parameters	6
A. Properties of the order parameters	8
SIV. Statistical inference	9
A. Bayesian formulation	9
B. Setup of prior distribution	10
C. Details of Monte Carlo simulations	11
SV. Examples	11
A. Target pattern produced by a numerical simulation: dodecagonal quasi-crystal	11
B. Target pattern synthesised by a function: dodecagonal quasi-crystal	13
C. Target pattern synthesised by a function: stripe and hexagonal patterns	16
D. Target pattern synthesised by a function: double gyroid and Frank Kasper A15	20
SVI. Robustness against noise	24
References	26

SI. RELATED WORKS

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.

* corresponding authors: yoshinaga@tohoku.ac.jp

A. 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 equation (SH) [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[1–5], and then has later been applied to optics[6], ecosystem[7], 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[8, 9], whereas SH may also have been used for the same problem[10, 11]. 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[12], and lamellar, hexagonal cylinders, BCC, hexagonal closed packing patterns, also gyroid patterns in three dimensions [11, 13, 14]. The essence of these models is that there is a specific length scale, or wavenumber, at which the uniform state becomes unstable. Recently, the conserved version of SH has been used for the description of the crystalline structures [15, 16]. This model is specifically called the phase-field crystal model (PFC). The models that we used in this work are extensions of the PFC.

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 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 analyse 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[17], but also in soft materials[18, 19]. The dynamics of defects has been studied by using PDEs[12, 16]. The PFC is a natural extension of the conventional phase-field model, and is capable of describing diffusive time scale as well as local crystalline order[15]. 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 an interaction potential between particles, and efforts are necessary to map molecular interactions to the associated linear and nonlinear terms in PDEs[20].

The extension of the PFC (or SH) has been proposed in several ways. One is to introduce two length scales to destabilise the uniform state[21, 22]. This results in stable quasi-crystalline patterns such as decagonal (10-fold) and dodecagonal (12-fold) quasi-crystals in two dimensions. This model also reproduces an icosahedral pattern in three dimensions[23] although in this case, the two length scales, are strictly speaking, not unstable but close to neutral stability. The quasi-crystal pattern appears due to nonlinear selection mechanism. The second type of the extension is to use many order parameters as a model of a multi-component system. Along this line, quasi-crystal patterns are reproduced in two dimensions[24] and in three dimensions[25]. The third type of the extension is the weak crystallisation theory based on the Landau-Brazovskii theory [26]. The simplest version of this theory is nothing but the SH equation. In the weak crystallisation theory, anisotropic nonlinear terms are included to make complex patterns to reproduce, for example, quasi-crystals in two dimensions[27]. The model using the anisotropic nonlinear terms corresponds to an anisotropic interaction in a microscopic model, for example, using Janus particles, patchy particles[28], or polyhedral-shape particles[29].

B. Inverse structural optimisation

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 uniform disordered state. This self-assembly of discrete particles has an advantage against the template approach because after optimisation the structures are reproduced without external aids, and thus it is easy to scale up the structures[30–32]. 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 target lattice in real space as a cost function [33, 34] (See also [35, 36] and references therein). These studies initiate inverse structural design of the self-assembly of materials[37, 38]. The design is based on various optimisation techniques such as gradient method[39], generic algorithm[40], swarm optimisation[41, 42], the Covariance matrix adaption-evolutionary strategy (CMA-ES)[43, 44] and statistical physics-inspired inverse design (SP-ID) [44, 45].

Recently, the optimisation of interaction potentials by minimisation of the relative entropy combined with Monte Carlo simulations has been proposed [39, 46–50]. In those studies, parameters in a potential function are optimised by

a gradient of pair-potential with respect to the parameters using the difference between the target and target radial distribution functions. The target structure is given by 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[39, 46–48, 50], block copolymers[41, 42], and granular media[51]. 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 uncertainty of the parameters can be evaluated. We also optimise the noise amplitude described by the inverse temperature β . This optimisation has not been performed in the previous works, and it is essential to estimate the best parameters for the target pattern with noise and without ground truth parameters. Lack of the optimisation of the noise amplitude results in overfitting of the parameters for the target under noise[52]. Besides, the BM-PDE is an inverse structural design of the model described by PDEs. The optimisation 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.

C. Symbolic regression 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[53–55] along the line of system identification, optimisation, and control of nonlinear dynamical systems[56]. Initiated by the success of physical laws for double-pendulum [57, 58], 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)[59] and partial differential equations (PDEs) for time-series data [60, 61]. The key idea is to minimise the error of $\|\partial_t\psi(t) - f(\psi(t))\|$ under a certain norm $\|\cdot\|$ with a regularisation 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 in ψ . Then, the problem reduces to an estimation of the 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.

Our problem to estimate the governing PDE from a given target 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 demands an accurate observation of $\partial_t\psi$ [54, 56, 62, 63]. The target pattern in our problem is only one snapshot ideally satisfying $\partial_t\psi = 0$ without noise, and therefore, we cannot use optimisation 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 target pattern is not necessarily produced from a numerical result, but synthesised by a function that is independent of the models.

Another feature of BM-PDE is uncertainty quantification. Statistical inference has been widely used to estimate parameters with their errors (uncertainty)[64, 65]. 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[66, 67]. Majority of the methods are based on point estimation in which parameters are estimated by minimising the cost function (energy)[59–61]. To estimate parameters in ODEs, statistical inference has been used for time-series data using approximate Bayesian computation [68].

D. State noise and measurement noise

To our knowledge, all the studies on the estimation of governing equations described by PDEs are based on time-series data. Two major approaches for such data are the regression and trajectory matching[54, 56]. The regression minimises $|\partial_t\psi(\mathbf{x}) - f[\psi(\mathbf{x})]|^2$ with respect to parameters μ in $f[\psi(\mathbf{x})]$ [53]. The sparse regression adds a regularisation term such as $|\mu|$ in the cost function. This method implies the dynamics is governed by the stochastic process $\partial_t\psi(\mathbf{x}) = \mathbf{f}[\psi(\mathbf{x})] + \xi(\mathbf{x}, t)$ where $\xi(\mathbf{x}, t)$ is the state noise. In the simplest case, the state noise is taken from the normal distribution. The advantage of this method is its low computational cost and independence on initial conditions. However, this method does not suit for the measurement noise $\psi(\mathbf{x}) = \psi^*(\mathbf{x}) + \xi(\mathbf{x})$ because the nonlinear terms in $f[\psi(\mathbf{x})]$ may amplify or reduce the noise[56].

The second method is trajectory matching[54, 67]. In this method, the estimation of parameters is made for a specific pattern generated numerically under specific initial condition. The optimisation problem is to minimise $|\psi(\mathbf{x}, t) - \psi^*(\mathbf{x}, t)|$ with respect to parameters which change $\psi(\mathbf{x}, t)$. Here, $\psi^*(\mathbf{x}, t)$ is data and $\psi(\mathbf{x}, t)$ is generated, typically using numerical simulations, from the initial condition under a candidate of parameters. This method may

be used not only for the state noise, but also for the measurement noise $\psi(\mathbf{x}) = \psi^*(\mathbf{x}) + \xi(\mathbf{x})$. The initial condition is either taken from the data (and fixed), or is treated as parameters to be optimised [54, 67].

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. The regression method is appealing in this respect, but in our problem, uncertainty arises from the target pattern, and not from the dynamical equations. Therefore, we have to consider the measurement noise rather than the state noise. We need to eliminate the dependence on the initial condition. To do this, we marginalise the initial condition, which is taken from a random distribution. We also have to identify the same patterns which have 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.

III. FAMILY OF MODELS

A. Phase-field crystal model and its generalisation

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

$$\partial_t \psi = \mathcal{L}\psi + \Delta\psi^3 \quad (\text{S1})$$

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 Swift-Hohenberg (SH) equation [16]. The equation is given by

$$\partial_t \psi = \Delta \left[-\epsilon\psi + (q_0 + \Delta)^2 \psi + \psi^3 \right] \quad (\text{S2})$$

where the total mass is conserved as

$$\bar{\psi} = \frac{1}{V} \int \psi(\mathbf{x}) d\mathbf{x}. \quad (\text{S3})$$

Here, V is the total volume (area in two dimensions) in the system, and q_0 is a characteristic wavenumber 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 [12, 69], and a lamellar, hexagonal cylinder, BCC, and hexagonal closed packing patterns [13, 70]. A finite mean density $\bar{\psi}$ plays a role as the quadratic nonlinear term in equation (S2). This can be seen by subtracting the mean density as $\psi \rightarrow \psi + \bar{\psi}$ in equation (S2) as,

$$\partial_t \psi = \Delta \left[(-\epsilon + 3\bar{\psi}^2) \psi + (q_0 + \Delta)^2 \psi + 3\bar{\psi}\psi^2 + \psi^3 \right]. \quad (\text{S4})$$

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. 1(e). To extend equation (S2) for the arbitrary number of length scales, we use m characteristic wavenumbers, q_0, q_1, \dots, q_{m-1} , and the values of the spectral, a_0, a_1, \dots, a_{m-1} , at the wavenumber $k = q_i$. 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 equation (S1) are given by

$$\partial_t \tilde{\psi}(\mathbf{k}) = \mathcal{L}_k \tilde{\psi}(\mathbf{k}) + \mathcal{F} [\Delta\psi(\mathbf{x})^3], \quad (\text{S5})$$

and its linear operator is expressed in the Fourier space as

$$\mathcal{L}_k = a_0 S_0(k) + a_1 S_1(k) + \dots + a_m S_m(k) + k^2 (1 - k^2)^2 (q_1^2 - k^2)^2 \dots (q_m^2 - k^2)^2. \quad (\text{S6})$$

To specify the position of peaks and their heights in the spectrum, $S_i(k)$ is chosen to be a function of k and q_i satisfying

$$\frac{d\mathcal{L}_{k=q_i}}{dk} = 0 \quad (\text{S7})$$

$$\mathcal{L}_{k=q_i} = a_i, \quad (\text{S8})$$

for $i \in [0, m - 1]$. The wavelength is limited in the range

$$q_i \in (0, 1] \quad (\text{S9})$$

and a_i is sufficiently close to 0, but not necessarily positive. Since we may freely choose a unit of length scale, we fix to be $q_0 = 1$ when $m \geq 2$.

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 the following form,

$$\mathcal{L}_k = -\frac{a_0}{q_0^4} k^2 (k^2 - 2q_0^2) - s_0 k^2 (q_0^2 - k^2)^2 \quad (\text{S10})$$

With this form, patterns with $k = q_0$ appear, and its stability is denoted by a_0 (see Fig. 1(e) in the main text). Here, s_0 expresses sharpness of the peak in the spectrum. For larger s_0 , the peak becomes sharper (Fig. 1(e)).

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. 1(e). This model has been used to reproduce quasi-crystal patterns [23]. The explicit form of the linear operator is given by

$$\mathcal{L}_k = a_0 S_0(k) + a_1 S_1(k) - \frac{s_0 k^2}{q_1^4} (q_0^2 - k^2)^2 (q_1^2 - k^2)^2 \quad (\text{S11})$$

$$S_0(k) = \frac{1}{q_0^4 (q_0^2 - q_1^2)^3} k^2 (q_1^2 - k^2)^2 ((q_1^2 - 3q_0^2)k^2 + q_0^2(-2q_1^2 + 4q_0^2)) \quad (\text{S12})$$

$$S_1(k) = \frac{1}{q_1^4 (q_0^2 - q_1^2)^3} k^2 (q_0^2 - k^2)^2 ((3q_1^2 - q_0^2)k^2 + q_1^2(2q_0^2 - 4q_1^2)). \quad (\text{S13})$$

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

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

$$\mathcal{L}_k = a_0 S_0(k) + a_1 S_1(k) + a_2 S_2(k) - s_0 k^2 (q_0^2 - k^2)^2 (q_1^2 - k^2)^2 (q_2^2 - k^2)^2 \quad (\text{S14})$$

$$S_0(k) = \frac{1}{q_0^4 (q_0^2 - q_1^2)^3 (q_2^2 - q_0^2)^3} k^2 (k^2 - q_1^2)^2 (k^2 - q_2^2)^2 \times (q_0^2(-6q_0^4 + 4q_0^2 q_2^2 + 4q_0^2 q_1^2 - 2q_1^2 q_2^2) + k^2(5q_0^4 - 3q_0^2 q_2^2 - 3q_0^2 q_1^2 + q_1^2 q_2^2)) \quad (\text{S15})$$

$$S_1(k) = \frac{1}{q_1^4 (q_0^2 - q_1^2)^3 (q_1^2 - q_2^2)^3} k^2 (k^2 - q_0^2)^2 (k^2 - q_2^2)^2 \times (q_1^2(-2q_0^2 q_2^2 + 4q_0^2 q_1^2 + 4q_1^2 q_2^2 - 6q_1^4) + k^2(q_0^2 q_2^2 - 3q_0^2 q_1^2 - 3q_1^2 q_2^2 + 5q_1^4)) \quad (\text{S16})$$

$$S_2(k) = \frac{1}{q_2^4 (q_2^2 - q_0^2)^3 (q_1^2 - q_2^2)^3} k^2 (k^2 - q_0^2)^2 (k^2 - q_1^2)^2 \times (q_2^2(-2q_0^2 q_1^2 + 4q_0^2 q_2^2 + 4q_1^2 q_2^2 - 6q_2^4) + k^2(q_0^2 q_1^2 - 3q_0^2 q_2^2 - 3q_1^2 q_2^2 + 5q_2^4)). \quad (\text{S17})$$

The linear spectrum is shown in Fig. 1(e).

B. Polynomial expansion of the linear operation with respect to wavenumbers

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. 1(e) and equations (S5)-(S6). The benefit of this approach is that parameters μ have clear physical meaning: the wavenumbers q_i correspond to the characteristic length scales and stability at the wavenumbers 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 in equation (S11). On the other hand, we may consider a family of the model by expanding \mathcal{L}_k in equation (S6) 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 equation (S6).

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 different number of length scales in Fig. 1(e) in a unified way. Once we truncate the expansion up to k^{14} , we may express 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 regularisation and use a particular type of prior distribution of the parameters[71].

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 a practical term, the range of the parameters is too broad so that estimation is not possible. For example, the numerically produced dodecagonal quasi-crystal discussed in Sec. SV A is expressed as

$$\mathcal{L}_k \approx -99.7k^2 + 945.6k^4 - 2984.8k^6 + 3531.8k^8 - 1392.8k^{10}; \quad (\text{S18})$$

and the precision of 0.1 is necessary to obtain quasi-crystal. 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 quasi crystal within the error.

Finally, we should point out that the sparse regression does not work when parameters are correlated [72]. In the system that we are studying, the parameters must be correlated because the amplitude and position of the unstable wavenumber are set by the peak of \mathcal{L}_k , and it is determined by the balance of several terms in the polynomial expansion.

SIII. ORDER PARAMETERS

To quantify the distance between a target pattern, $\psi^*(\mathbf{x})$, and a pattern $\psi(\mathbf{x})$ that we get from the model equation (S1), we consider the cost function, which we also refer to as energy. The cost function is zero when the two patterns are identical while it is larger when the current pattern is far from the target pattern. The naive cost function may be the average difference of *local* density

$$E \stackrel{?}{=} \int [\psi(\mathbf{x}) - \psi^*(\mathbf{x})]^2 d\mathbf{x}. \quad (\text{S19})$$

This is a counterpart of the average difference of position of particles with respect to positions in a target pattern in discrete particle systems [33, 34]. This cost function has also been used in the parameter estimation of a PDE for time-series data[55, 56]. This cost function is not appropriate for our purpose because a stationary target pattern has translational and rotational invariance. The energy should not change by translation and rotation of the whole pattern (see Fig. S1). We, therefore, consider the cost function by comparing two patterns up to global symmetry transformation \mathcal{T}_λ parametrised by λ as

$$E = \min_{\lambda} \int [\mathcal{T}_\lambda \circ \psi(\mathbf{x}) - \psi^*(\mathbf{x})]^2 d\mathbf{x}. \quad (\text{S20})$$

Here, we consider the action of the Euclidean group $E(d)$ and the global scale transformation. The naive way to compute the cost function is brute force matching, that is, the two patterns are compared after translating and rotating one of the pattern. This is impossible in practice, and therefore, we use an approximate form of equation (S20).

To do this, we used the Fourier transform of the pattern as

$$\tilde{\psi}(\mathbf{k}) = \int \psi(\mathbf{x}) e^{i\mathbf{k}\cdot\mathbf{x}} d\mathbf{x} \quad (\text{S21})$$

$$\psi(\mathbf{x}) = \int_{\mathbf{k}} \tilde{\psi}(\mathbf{k}) e^{-i\mathbf{k}\cdot\mathbf{x}} \quad (\text{S22})$$

where the volume in the Fourier space is $\int_{\mathbf{k}} = \frac{1}{(2\pi)^d} d^d \mathbf{k}$. The cost function similar to equation (S19) may be defined in the Fourier space. Yet this energy does not work well because the peaks of the pattern are very sharp and localised at some position in k -space (see the spectrum in Fig. S1).

We define the order parameter $\Psi[\psi] = (\Psi_1[\psi], \Psi_2[\psi], \dots, \Psi_l[\psi])$ to characterise *pattern* in the Fourier space, and use the following cost function,

$$E[\psi^*, \psi] = \sum_{l=1}^{l_0} |\Psi_l^* - \Psi_l[\psi]|^2 \quad (\text{S23})$$

where l denotes a mode of the order parameter and $\Psi_l^* = \Psi_l[\psi^*]$. 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. Note that

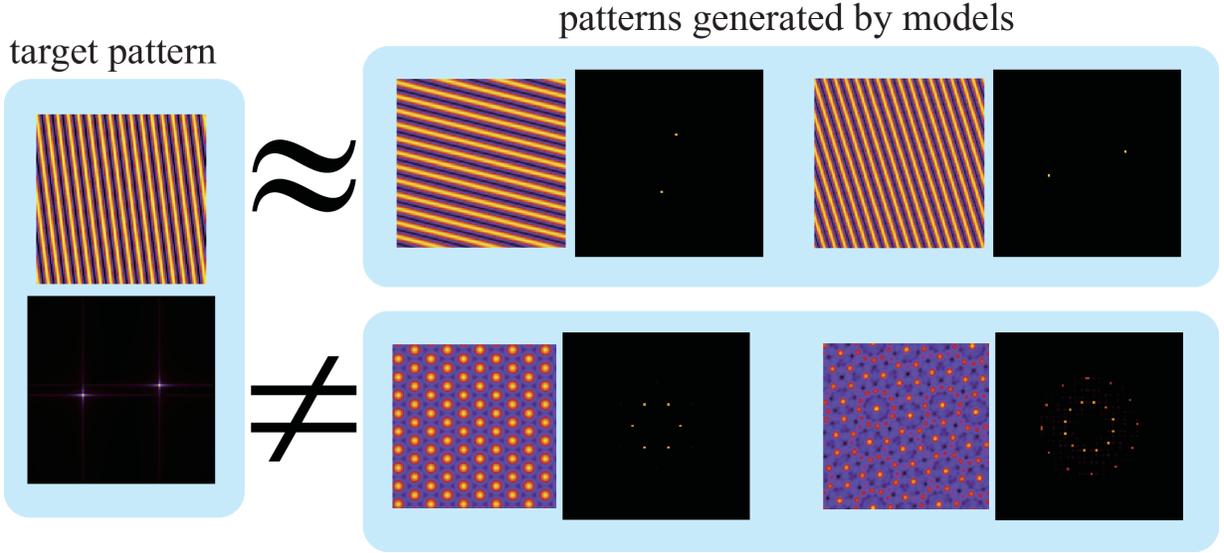


FIG. S1. Comparison between patterns. When the target pattern is a stripe (left), translationally shifted and/or rotated patterns are also regarded as a stripe (right top). However, hexagonal and dodecagonal quasi-crystal patterns (right bottom) are different from the stripe pattern. In addition to the patterns in real space, corresponding patterns in the Fourier space are shown.

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 . In three dimensions, A_{lm} is described by

$$A_{lm}[\psi] = \int W(k) |\tilde{\psi}(\mathbf{k})| Y_l^m(\theta_k, \varphi_k) d\mathbf{k}^3 \quad (\text{S24})$$

with spherical harmonics $Y_l^m(\theta_k, \varphi_k)$ in the spherical coordinates of the Fourier space (k, θ_k, φ_k) , and, in two dimensions, by

$$A_{l, \pm 1}[\psi] = \int W(k) |\tilde{\psi}(\mathbf{k})| \begin{pmatrix} \cos l\theta_k \\ \sin l\theta_k \end{pmatrix} d\mathbf{k}^2 \quad (\text{S25})$$

where $+1(-1)$ denotes $\cos l\theta_k(\sin l\theta_k)$, respectively, and θ_k is a polar angle in the Fourier space. In two dimensions, we denote $A_{l, \pm}$ in the vector form as $\mathbf{A}_l = (A_{l,+}, A_{l,-})$. Here, $W(k)$ is a weight function in the Fourier space, depending only on $k = |\mathbf{k}|$. In this work, we set to be $W(k) = 1$ for simplicity. The zeroth mode $l = 0$ corresponds to mean amplitude of $\tilde{\psi}(\mathbf{k})$, which is independently considered by $\bar{\psi}$. We, therefore, use the sum for $l \in [1, l_0]$ in equation (S23). The maximum mode is denoted by l_0 . We use the convention of spherical harmonics

$$Y_l^m(\theta, \varphi) = \sqrt{\frac{(2l+1)(l-m)!}{4\pi(l+m)!}} P_l^m(\cos \theta) e^{im\varphi} \quad (\text{S26})$$

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 as

$$f(\theta, \varphi) = \sum_{l=0}^{\infty} \sum_{m=-l}^l f_{lm} Y_l^m(\theta, \varphi) \quad (\text{S27})$$

and, in turn, the coefficients are expressed by

$$f_{lm} = \int f(\theta, \varphi) Y_l^{m*}(\theta, \varphi) d\theta d\varphi \sin \theta. \quad (\text{S28})$$

We consider appropriate norm that is invariant under rotation,

$$\Psi_l = \|A_{lm}\| = \sqrt{A_{l,+}^2 + A_{l,-}^2} \quad (\text{S29})$$

in two dimensions, and

$$\Psi_l = \|\mathbf{A}_l\| = \sqrt{\frac{4\pi}{2l+1}} \sqrt{\sum_{m=-l}^l (-1)^m A_{l,m} A_{l,-m}} \quad (\text{S30})$$

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 a target 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 $l_0/2$.

A. Properties of the order parameters

We summarise the properties of our order parameters defined in equation (S29) in two dimensions and equation (S30) in three dimensions. First, we show the order parameter is invariant with translation

$$\mathcal{T}_{\mathbf{a}} \circ \psi(\mathbf{x}) = \psi(\mathbf{x} + \mathbf{a}). \quad (\text{S31})$$

By the Fourier transform of the pattern, $\tilde{\psi}(\mathbf{k}) = \mathcal{F} \circ \psi(\mathbf{x})$, the translated pattern is expressed as

$$\begin{aligned} \mathcal{F} \circ \mathcal{T}_{\mathbf{a}} \circ \psi(\mathbf{x}) &= \int \psi(\mathbf{x} + \mathbf{a}) e^{i\mathbf{k} \cdot \mathbf{x}} d\mathbf{x} \\ &= e^{i\mathbf{k} \cdot \mathbf{a}} \int \psi(\mathbf{x}) e^{i\mathbf{k} \cdot \mathbf{x}} d\mathbf{x} \\ &= e^{i\mathbf{k} \cdot \mathbf{a}} \mathcal{F} \circ \psi(\mathbf{x}) \end{aligned} \quad (\text{S32})$$

and therefore, $|\mathcal{F} \circ \mathcal{T}_{\mathbf{a}} \circ \psi(\mathbf{x})| = |\mathcal{F} \circ \psi(\mathbf{x})|$.

Next, we consider the global rotation around the axis \mathbf{n} with an angle ϕ as

$$\mathcal{T}_{\phi, \mathbf{n}} \circ \psi(\mathbf{x}) = \psi(\mathcal{R}_{\phi, \mathbf{n}}(\mathbf{x})), \quad (\text{S33})$$

where $\mathcal{R}_{\phi, \mathbf{n}}$ is the rotation matrix around the axis \mathbf{n} with an angle ϕ . Note that in two-dimensions $\mathbf{n} = \hat{\mathbf{e}}_z$ and therefore the rotation is characterised only by ϕ . Because the rotation matrix is unitary transformation and commute with the Fourier transform, the rotational transformation may be considered as rotation in the Fourier space. Using the polar coordinates in the Fourier space, the transformation in two dimensions is given by

$$\mathcal{T}_{\phi} \circ |\mathcal{F} \circ \psi(\mathbf{x})| = |\tilde{\psi}(k, \theta_k + \phi_k)|. \quad (\text{S34})$$

Because equation (S25) is essentially the Fourier transform in the polar direction, its amplitude equation (S29) is invariant the shift in the polar direction, that is, invariant under the global rotation.

In three dimensions, under rotation, equation (S30) is transformed as

$$\mathcal{T} \circ \Psi_l = \sqrt{\frac{4\pi}{2l+1}} \sqrt{\sum_{m, m_1, m_2} \mathcal{D}_{mm_1}^{(l)} A_{lm_1} \left(\mathcal{D}_{mm_2}^{(l)} A_{lm_2} \right)^*} = \sqrt{\frac{4\pi}{2l+1}} \sqrt{\sum_m A_{lm} A_{lm}^*} = \Psi_l \quad (\text{S35})$$

which is invariant under rotation. Here, \mathcal{D}_{mm_1} is Wigner D-matrix rotating the spherical harmonics[73]. Note that the rotation does not change l , but do change m .

In contrast with the order parameter in two dimensions, equation (S30) is not the only quantity that is invariant under rotation. There is another invariant, under given l ,

$$\Psi_l^{(3)} = \sum_{m_1, m_2, m_3} \begin{pmatrix} l & l & l \\ m_1 & m_2 & m_3 \end{pmatrix} A_{lm_1} A_{lm_2} A_{lm_3} \quad (\text{S36})$$

This was proposed to characterise the structure of supercooled liquid in [74]. Algebraic structure behind this argument is that, in two dimensions, the coupling two same modes (say m) produce only 2 and 0, but do not produce m itself

resulting from trigonometric product. On the other hand, in three dimensions, the coupling between two modes l do produce the mode l . Here we have used the $3j$ -symbol counts only $m_1 + m_2 + m_3 = 0$. Notice that this quantity is identically zero for $l = 1$, for there is only one invariant in this case. The $l = 1$ is isomorphic to a vector, whose invariant is length (amplitude) of it.

Another property of the order parameter is about the cost function between a pattern $\psi(\mathbf{x})$ and its rotation by $2\pi/n$

$$E = \sum_{l=1}^{l_0} |\Psi_l - \mathcal{R}_{2\pi/n} \circ \Psi_l|. \quad (\text{S37})$$

When $A_{l,\pm 1}$ is finite only for $l = n, 2n, 3n, \dots$, the transformation does not change the order parameter and thus $E = 0$.

SIV. STATISTICAL INFERENCE

A. 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. To do this, we assume the order parameter $\Psi_l[\psi_s]$ of the stationary pattern ψ_s is observed as that of the target pattern ψ^* with the additive noise ξ_l :

$$\Psi_l^* = \Psi_l(\psi_s; \psi_0, \mu) + \xi_l, \quad (\text{S38})$$

where ξ_l is the random variable 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 equation (S1), 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 equation (S38) is equivalently represented as the conditional probability density

$$p(\Psi_l^* | \psi_0, \mu, \beta) = \sqrt{\frac{\beta}{2\pi}} \exp \left\{ -\frac{\beta}{2} [\Psi_l^* - \Psi_l(\psi_s; \psi_0, \mu)]^2 \right\}. \quad (\text{S39})$$

We consider the parameter estimation of μ by marginalising 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 respectively, the mean density and the (relative) density at the mesh point j . 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

$$\begin{aligned} p(\mu | \{\Psi_l^*\}_{l=1}^{l_0}, \psi_0, \beta, m) &= \frac{p(\mu | m)}{p(\{\Psi_l\}_{l=1}^{l_0} | \beta, m)} \prod_{l=1}^{l_0} p(\Psi_l^* | \psi_0, \mu, \beta), \\ &\propto \exp \left[-\frac{\beta}{2} E(\psi^*, \psi_s; \psi_0, \mu) \right] \end{aligned} \quad (\text{S40})$$

where m denotes the number of the characteristic length scales such as equations (S10), (S11) or (S14). 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

$$p(\{\Psi_l\}_{l=1}^{l_0} | \beta, m) = \left(\frac{\beta}{2\pi} \right)^{\frac{l_0}{2}} \int \exp \left[-\frac{\beta}{2} E(\psi^*, \psi_s; \psi_0, \mu) \right] p(\psi_0) p(\mu | m) d\psi_0 d\mu. \quad (\text{S41})$$

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*

$$p(\psi_0, \mu | \{\Psi_l^*\}_{l=1}^{l_0}, \psi_0, \beta, m) = p(\psi_0) p(\mu | \{\Psi_l^*\}_{l=1}^{l_0}, \psi_0, \beta, m), \quad (\text{S42})$$

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 marginalising out ψ_0 , the *posterior* distribution of μ is given by

$$p(\mu \mid \{\Psi_l^*\}_{l=1}^{l_0}, \beta, m) = \int p(\psi_0, \mu \mid \{\Psi_l^*\}_{l=1}^{l_0}, \psi_0, \beta, m) d\psi_0. \quad (\text{S43})$$

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

We consider both the hyperparameter estimation of β and model selection of m [52, 64, 75–77]. By Bayes' theorem, the joint probability density of β and m under given $\{\Psi_l^*\}_{l=1}^{l_0}$ is represented as

$$p(\beta, m \mid \{\Psi_l^*\}_{l=1}^{l_0}) = \frac{p(\{\Psi_l^*\}_{l=1}^{l_0} \mid \beta, m) p(\beta) p(m)}{p(\{\Psi_l^*\}_{l=1}^{l_0})}, \quad (\text{S44})$$

where $p(\{\Psi_l^*\}_{l=1}^{l_0})$ is the normalising 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

$$(\hat{\beta}, \hat{m}) = \arg \max_{\beta, m} p(\beta, m \mid \{\Psi_l^*\}_{l=1}^{l_0}) \quad (\text{S45})$$

$$= \arg \max_{\beta, m} p(\{\Psi_l^*\}_{l=1}^{l_0} \mid \beta, m). \quad (\text{S46})$$

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

$$\hat{\beta} = \frac{1}{\langle E(\psi^*, \psi_s; \psi_0, \mu) \rangle_{\hat{\beta}}}, \quad (\text{S47})$$

where

$$\langle \cdots \rangle_{\beta} = \int (\cdots) p(\psi_0, \mu \mid \{\Psi_l^*\}_{l=1}^{l_0}, \beta, m) d\psi_0 d\mu. \quad (\text{S48})$$

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

$$p(m \mid \{\Psi_l^*\}_{l=1}^{l_0}) = \int p(\beta, m \mid \{\Psi_l^*\}_{l=1}^{l_0}) p(\beta) d\beta. \quad (\text{S49})$$

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

B. Setup of 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 to the domain of each variable. The prior density of ψ_0 is defined by

$$p(\psi_0) = \prod_{j=1}^{N^d} \varphi(\psi_0^{(j)}), \quad (\text{S50})$$

where $\varphi(\psi_0^{(j)})$ is the continuous uniform distribution whose support is $\psi_0^{(j)} \in [-0.1, 0.1]$. For equation (S39) with m length scales, the set of parameters is defined by

$$\mu = \{dx, dy, \bar{\psi}, a_0, q_1, a_1, q_2, a_2, \dots, q_{m-1}, a_{m-1}\}, \quad (\text{S51})$$

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

$$p(\mu | m) = \varphi(dx)\varphi(dy)\varphi(\bar{\psi})\varphi(a_0) \prod_{i=1}^{m-1} \varphi(a_i)\varphi(q_i), \quad (\text{S52})$$

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 synthesise the target pattern (see Methods in the main text).

We also assume no prior information for 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

$$\beta_\alpha = 10^{\log_{10} \beta_{\min} + \frac{\alpha-1}{N_{\text{rep}}-1} \log_{10}(\beta_{\max}/\beta_{\min})} \quad (\text{S53})$$

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 (S53) 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 $m \in \{m_i\}_{i=1}^{i_{\max}}$. Each grid point (m_i, β_α) can be regarded as the candidate of model selection equally possible in prior.

C. Details of Monte Carlo simulations

The joint distribution $p(\psi_0, \mu | \{\Psi_l^*\}_{l=1}^{l_0}, \beta, m)$ is realised by the Gibbs sampling based on the relation of equation (S42), 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 equation (S50). 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 equation (S1) under a given initial state ψ_0 and parameters μ of equation (S51) for a model m . Then, the similarity of an obtained pattern ψ_s as a stationary state and a target pattern ψ^* is evaluated by the cost function $E(\psi^*, \psi_s; \psi_0, \mu)$, describing the distance shown in equation (S23) in the space of the order parameter Ψ_l (see equation (S29) or equation (S30)). 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 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 [78, 79]. At higher temperature, the motion of one MC step in the parameter space is large whereas at the lower temperature, 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 neighbouring temperature were exchanged following the Metropolis criterion. This process enables us to sample parameters weighed with likelihood effectively [78, 79].

The initial parameter set is sampled from the prior distribution of equation (S52). The lowest energy state 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 energy state, we restart REMC from the initial parameter set of the lowest energy state to sample its steady state. This is because the relaxation under higher temperature is much faster than the lower temperature. 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 [80, 81]. The error bars of $F(\beta, m)$ were calculated by the bootstrap resampling [82].

SV. EXAMPLES

A. Target pattern produced by a numerical simulation: dodecagonal quasi-crystal

The process of parameter estimation is shown in Fig. S2(a). At the initial parameters, estimation is poor in all temperatures. After several MC steps, the cost function decreases drastically at the lower temperature, and reach its steady state. At the steady states, sampled patterns are very closed to the target 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. S2(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 energy has two regions connected by the energy gap. The lower energy corresponds to a quasi-crystal, which is qualitatively similar

to the target pattern, while the higher energy is dominated by hexagonal patterns. This result implies that there are multiple minima of the energy distribution. This means that the standard MC simulation does not converge to the ground truth unless an initial parameter set is chosen nicely,

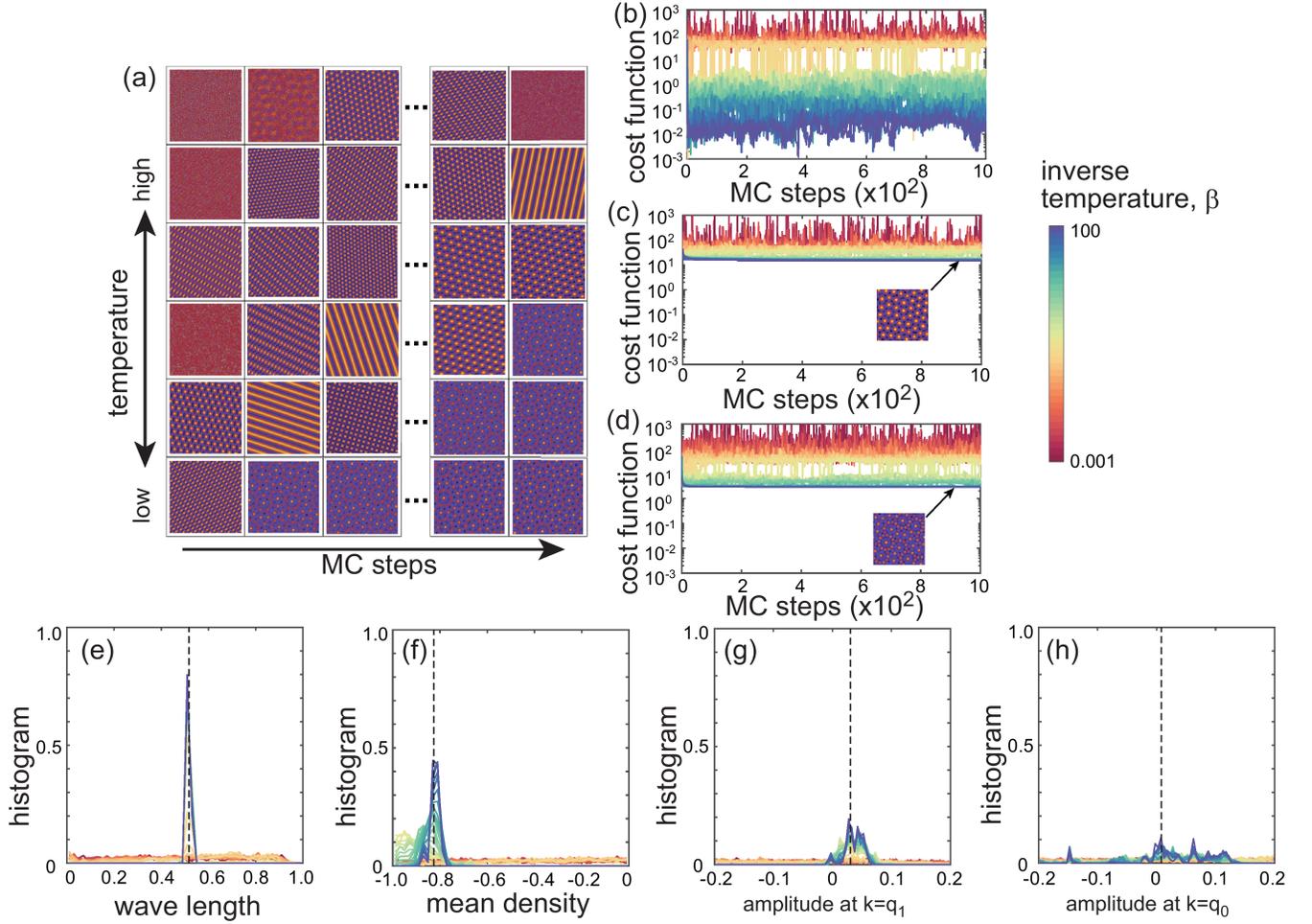


FIG. S2. (a) Snapshots of two-dimensional pattern with two length scales under different inverse temperature β during the Monte Carlo simulations for the target pattern of numerically produced dodecagonal quasi-crystal. (b-d) The cost function during the steps in replica exchange Monte Carlo (REMC) simulations 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 colours show different temperature.

The same algorithm is used for the different models with one length scale or three length scales (Fig. S2(c,d)). Both REMC simulations converge to their steady energy histograms, both of which are higher than the energy histogram of the two-length-scale model. This result means the two-length-scale model is indeed most likely for the target pattern. In fact, the marginal likelihood, which we call free energy $F(\beta, m)$, shown in Fig.2(e) in the main text has the lowest value for the two-length-scale model $m = m_2$. Because this target pattern has corresponding ground truth parameter values and the model of equation (2) 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 to the target patten 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 $\hat{\beta}$ at which the free energy has its minimum value.

The probability of each model is calculated by marginalising the probability at each temperature $P \propto e^{-F[\beta]}$ about the whole temperature range. This probability is also well captured by the minimal value of the free energy of each model. The result is shown in the inset of Fig.2(e). In this example, the probability of selecting the two-length-scale

TABLE S1. The ground truth and estimated parameter values for numerically produced 12-fold quasi-crystal. Errors are evaluated from the standard deviation of each data.

parameters	ground truth	estimated (one length)	estimated (two length)	estimated (three length)
dx	1.020	0.985 ± 0.000	1.0217 ± 0.0024	1.002 ± 0.00036
dy	1.026	1.0006 ± 0.000	1.0232 ± 0.00319	1.0019 ± 0.0002
$\bar{\psi}$	-0.826	-0.436 ± 0.000	-0.820 ± 0.0163	-0.677 ± 0.0024
a_0	0.00794	0.0648 ± 0.000	0.0391 ± 0.0594	-0.157 ± 0.00475
q_1	0.518	-	0.518 ± 0.00214	0.728 ± 0.00077
a_1	0.0306	-	0.0375 ± 0.0140	0.0282 ± 0.0020
q_2	-	-	-	0.367 ± 0.0004
a_2	-	-	-	-0.136 ± 0.00246

model is almost 100%.

The estimate parameters are shown in Table S1, together with the uncertainty of the estimation. The uncertainty quantification is made by the standard deviation of the 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 Fig.S2(e-h). As we can see, at lower temperature (higher β in the blue line), the distribution accumulates around the ground truth whereas, at the higher temperature (low β shown in the red line), the distribution covers the whole range of the parameters. From the high temperature to the low temperature, the distribution gradually becomes shrinking. This suggests that RMEC can well sample the true parameter distribution.

The width of the histograms at the low temperature 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 S1. The mean density $\bar{\psi}$ and the spectrum amplitude at $q = q_1$ (see Fig. 1(e)) 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 temperature. In fact, the error in the estimation of a_0 is comparable to its estimated value, suggesting that we have 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 nonlinear 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 destabilised 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 quasi-crystalline pattern, and therefore, the estimated parameters correspond to hexagonal patterns. The three-length-scale model, on the other hand, does reproduce a quasi-crystalline pattern, which is similar to the target 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 energy 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 target pattern. We note that the energy of the three-length-scale model is still lower than the one-length-scale model in which the estimation is qualitatively wrong.

B. Target pattern synthesised by a function: dodecagonal quasi-crystal

The same algorithm was used for the target pattern synthesised by functions for the two-dimensional stripe, hexagonal, and dodecagonal quasi-crystal patterns, and the three-dimensional double-gyroid pattern. The results are summarised in Fig. 3 in the main text.

The target pattern of a dodecagonal quasi-crystal that is synthesised by the function of superposition of twelve plane waves (see Methods in the main text). The pattern has twelve-fold rotational symmetry, as demonstrated in the pattern in the Fourier space (Fig. S3). We may generate a similar pattern from the estimated PDE to the target pattern, if not exactly the same.

The obtained energy histogram is qualitatively similar to the target pattern of the numerically produced dodecagonal quasi-crystal (see Fig. S3). The two-length-scale model has two distinct energy scales associated with quasi-crystal and hexagonal pattern. At the lower energy, we obtain quasi-crystals. The three-length-scale model also reproduces

quasi-crystals, but the energy is higher than that of the two-length-scale model. Therefore, the two-length-scale model will most likely give the target pattern.

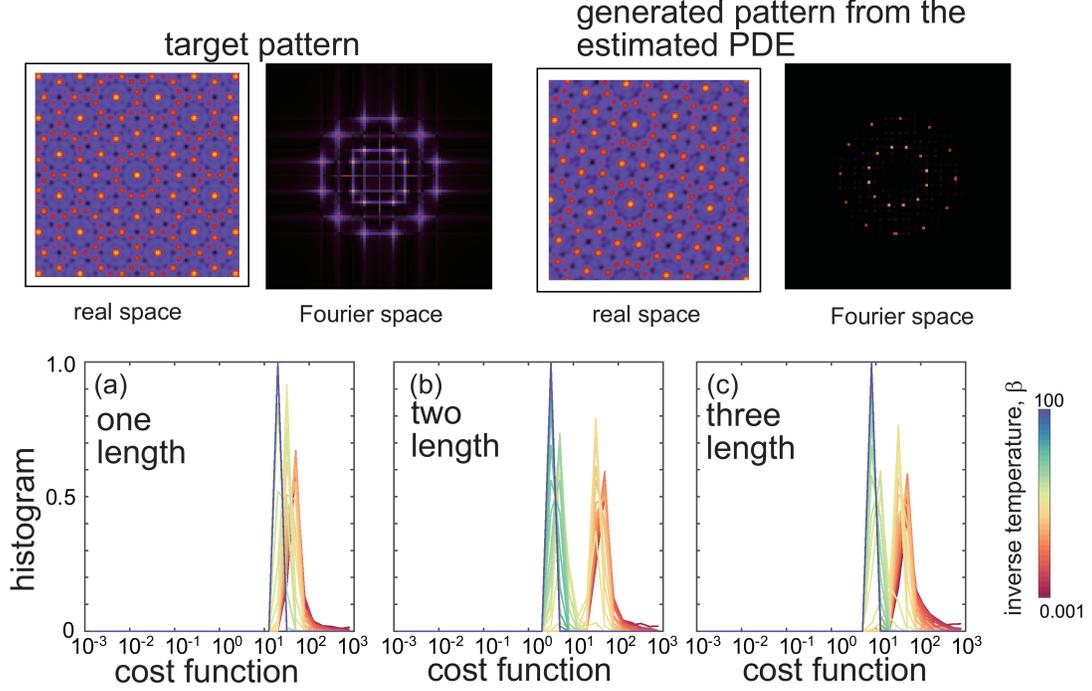


FIG. S3. (Top) The target pattern and the generated pattern from the estimated PDE in real and Fourier space. (Bottom) Energy histograms of the REMC sampling for the target pattern of dodecagonal quasi-crystal that is synthesised by the function. The horizontal axis is shown in the logarithmic scale.

The log marginal likelihood (free energy) at each temperature is calculated for each model (see Fig. 4(c)). The models with one- and three-length scales show qualitatively same temperature dependent of the log marginal likelihood with the results for numerically produced target pattern (Fig.2(e)). On the other hand, the two-length scale model $m = m_2$ shows qualitatively different target temperature dependence, namely there is a minimum at $\beta \approx 16$. This is because there is no ground truth in this target pattern. Intuitively, this situation is similar to the target pattern with noise. In both cases, the model of equation (2) cannot reproduce exactly the same pattern as the target. The Bayesian modelling is suitable in such cases because it gives us an optimal noise level[52]. Even though the cost function, or energy, is lower at the lower temperature, the posterior distribution is too narrow so that a slightly higher temperature than the lowest temperature gives the best estimate. This is exactly how our method avoids overfitting.

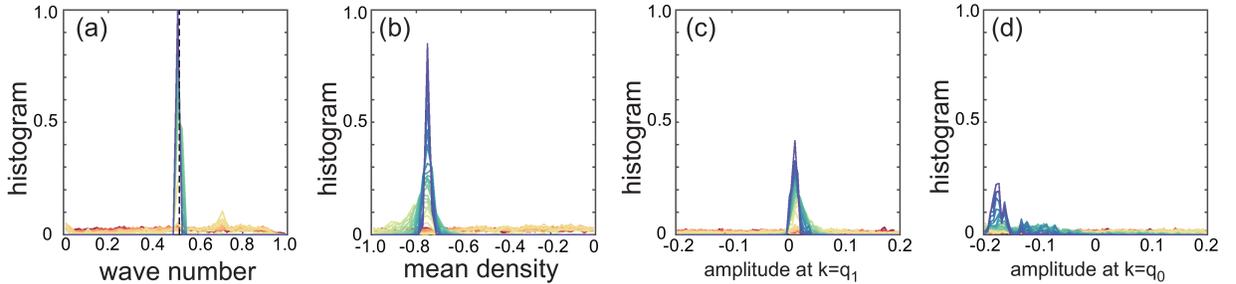


FIG. S4. Histogram of the estimated parameters for the target pattern of dodecagonal quasi-crystal that is synthesised by the function. The vertical dashed line in the histogram of wavenumber indicates $q_1 = \sqrt{2 - \sqrt{3}}$, which is necessary to reproduce the dodecagonal pattern.

The estimated parameters are shown in Table S2 and the corresponding histograms for each β are shown in Fig. S4. The histograms of the sampled parameters are qualitatively similar to the results of the numerically produced target pattern. The distribution of the estimated wavenumber is well accumulated around the value that is necessary to make quasi-crystals, 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. In fact, as we will discuss at the end of this section, there is a region where the dodecagonal quasi-crystal can appear in the parameter space spanned by mean density $\bar{\psi}$ and the spectrum amplitude a_1 . Other parameters show broader distribution than that of the wavelength. These parameters are distributed inside the region. In particular, the mean density is a good quantity separating the dodecagonal quasi-crystal from 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 equation (S4)). The stripe pattern has a symmetry of $\psi \rightarrow -\psi$, and therefore, this coupling must be weak. The hexagonal pattern does not have the symmetry and this requires stronger coupling; the dodecagonal quasi-crystal requires even far stronger coupling[21]. Therefore, $|\bar{\psi}| \gg 0$. This is in agreement with the result of the estimation in Fig. S4(b). The distribution of a_0 is broad by the same reason in the case of numerically produced dodecagonal quasi-crystal pattern.

TABLE S2. The estimated parameters values for functionally synthesised dodecagonal quasi-crystals. Errors are evaluated from the standard deviation of each sample from the REMC at the optimal temperature.

parameters	estimated (one length)	estimated (two length)	estimated (three length)
$dx = dy$	0.986 ± 0.0092	0.980 ± 0.0018	1.02 ± 0.0046
$\bar{\psi}$	-0.183 ± 0.0497	-0.748 ± 0.0192	-0.475 ± 0.00165
a_0	0.010 ± 0.0120	-0.132 ± 0.0443	-0.0459 ± 0.0433
q_1	-	0.514 ± 0.0037	0.666 ± 0.0052
a_1	-	0.0153 ± 0.0085	0.007 ± 0.0121
q_2	-	-	0.342 ± 0.0076
a_2	-	-	0.110 ± 0.0697

Once the best parameters are obtained, a larger scale pattern may be obtained as in Fig. S5. The dodecagonal quasi-crystal indeed spans the larger space. This pattern is also obtained from the random initial condition. Note that the estimated parameters do not imply that the dodecagonal quasi-crystal always appear. From the random initial condition, a solution of PFC may be trapped at metastable states. In a larger system, a pattern is more likely to be trapped because the energy barrier between the metastable states is larger. Topological defects are also likely to appear. Nevertheless, our method is capable of estimating the model that can reproduce the target pattern from a random initial condition. This is ensured at least at the system size we used for the estimate, but here we demonstrate the target pattern is also reproduced in the larger system.

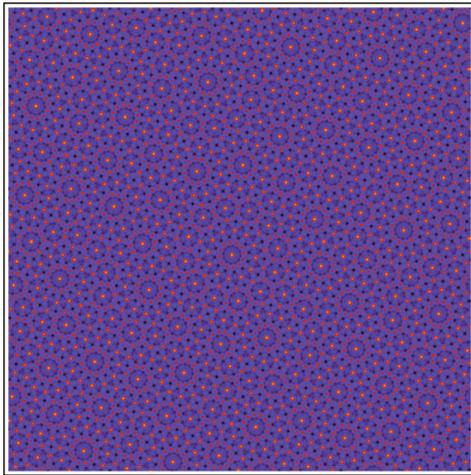


FIG. S5. A pattern from two-length-scale model under the estimated parameters in the system size of 512^2 .

To confirm that the BM-PDE indeed reproduces the target pattern, we compare the estimated parameters with the phase diagram. The phase diagram is based on a variational function $\mathcal{H}[\psi(\mathbf{x})]$, whose minimum implies most stable pattern. Our family of PFC models has $\mathcal{H}[\psi(\mathbf{x})]$ and thus equation (S1) is rewritten as

$$\partial_t \psi = \Delta \frac{\delta \mathcal{H}[\psi(\mathbf{x})]}{\delta \psi}. \quad (\text{S54})$$

In Fig. S6, the phase boundary is calculated by comparing $\mathcal{H}[\psi(\mathbf{x})]$ under each pattern. Each pattern is approximated under the one-mode approximation; namely, the pattern is expressed only by the single mode such as for the stripe pattern

$$\psi_{\text{stripe}}(\mathbf{x}) = A_{\text{stripe}}^* \sum_i \cos \mathbf{q}_i \cdot \mathbf{x} \quad (\text{S55})$$

where A_{stripe}^* is amplitude of the pattern[69, 83]. The wavenumber \mathbf{q}_i is chosen appropriately to reproduce the pattern (see Methods). At the stable state, the amplitude is, in fact, a function of the mean density $\bar{\psi}$ and the spectrum amplitude a_i at the characteristic wavenumber k_i .

For each pattern of the stripe, hexagonal, and dodecagonal quasi-crystal patterns, most of the sample of estimated parameters from the REMC at the optimal temperature are distributed inside each region in the phase diagram. This result suggests that the generated patterns from the estimated PDE are indeed the same as the target pattern. Note that between the two phases, there are coexistence regions where both patterns are possible; one is most stable and the other is meta-stable.

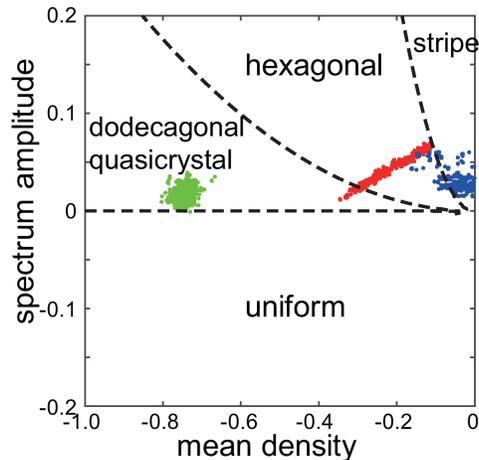


FIG. S6. The estimated parameters for the target patterns of stripe (blue), hexagonal (red), and dodecagonal quasi-crystals (green) in the parameter space spanned by the mean density $\bar{\psi}$ and spectrum amplitude (a_0 for stripe and hexagonal patterns, and a_1 for the quasi-crystal patterns). The target patterns are synthesised by functions, and therefore, there is no point of ground truth in the phase diagram. The phase boundaries are obtained by the condition of minimum \mathcal{H} in equation (S54) under the one-mode approximation.

C. Target pattern synthesised by a function: stripe and hexagonal patterns

We apply our method to stripe and hexagonal patterns. These patterns are simpler than the dodecagonal quasi-crystal, and the one length scale is enough to reproduce the patterns. The energy histogram is shown in Fig. S7 for the stripe pattern and in Fig. S9 for the hexagonal pattern. In contrast with the dodecagonal quasi-crystals, the energy of these patterns can be very low up to about $E \approx 10^{-5} - 10^{-4}$. In addition, the shapes of energy histogram under different models are qualitatively similar. Accordingly, the minimum free energy 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 (Table 3(a,b)). Temperature dependence of the free energy is indistinguishable among the three models. The free energy monotonically decreases as the temperature decreases (or β increases), and therefore, the minimum free energy is attained at the highest β (lowest temperature). The difference of the free

energy among the three models is small, but still by looking at marginalised probability, we find the one-length-scale model is more likely than other models as shown in Table 3(a,b).

The generated patterns from the estimated PDE are similar to the target pattern in both stripe (Fig. S7) and hexagonal patterns (Fig. S9). 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. S7 has a different orientation.

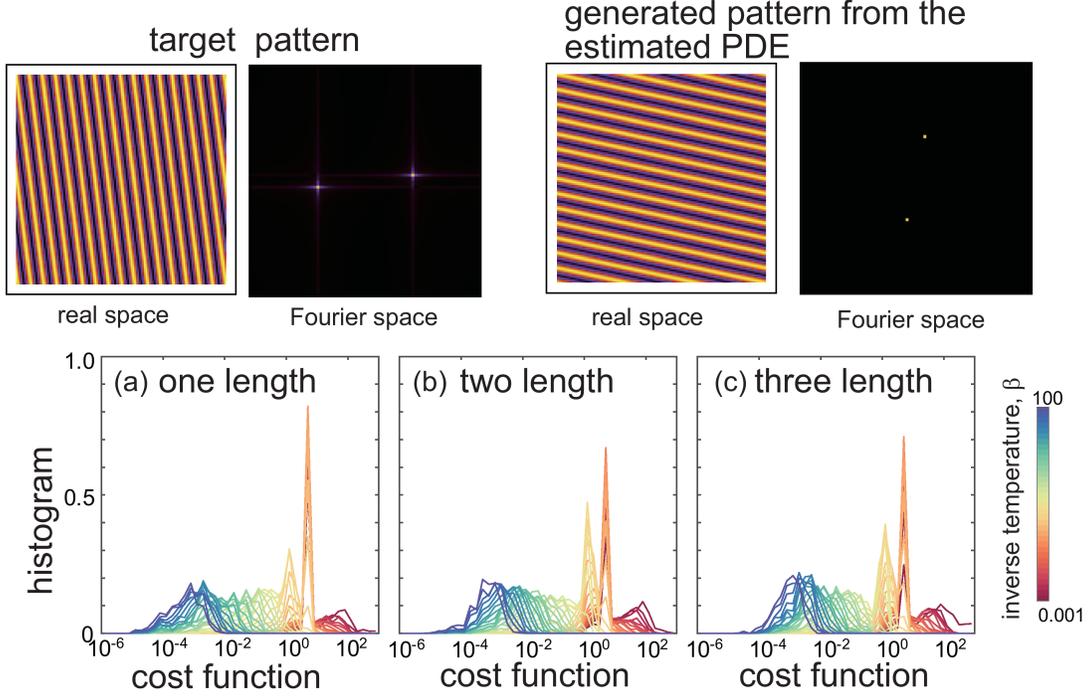


FIG. S7. (Top) The target pattern and the generated pattern from the estimated PDE in real and Fourier space. (Bottom) Energy histograms of the REMC sampling for the target pattern of a stripe pattern that is synthesised by the function. The horizontal axis is shown in the logarithmic scale.

The estimated parameters under the one-length-scale model are shown in Fig. S8 for the stripe pattern and in Fig. S10 for the hexagonal pattern. The estimated wavenumber is close to the choice of that of the target 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, $\psi = -\psi$ [15]. The hexagonal pattern breaks this symmetry and also it requires the three-body interaction, which is the quadratic term in ψ in the dynamical equation (2). The hexagonal pattern is expressed by 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 appear in the equation of ψ_1 . In Fig. S6, the phase diagram based on the one-mode approximation is shown. In fact, the stripe pattern appears near $\bar{\psi} = 0$, whereas the hexagonal pattern appears $\bar{\psi} \gg 0$. As in the case of the dodecagonal quasi-crystal, the estimated parameters are more or less inside each region in the phase diagram. This argument also justifies the successful estimation of the parameters also for the stripe and hexagonal target patterns.

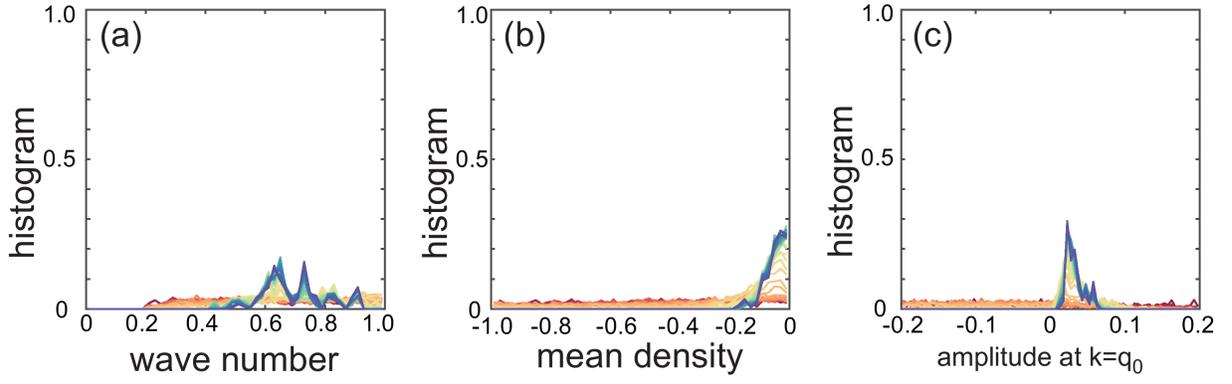


FIG. S8. Histogram of the estimated parameters for the target pattern of a stripe pattern that is synthesised by the function.

TABLE S3. The estimated parameters values for functionally synthesised stripe pattern. Errors are evaluated from the standard deviation of each sample from the REMC at the optimal temperature.

parameters	estimated (one length)	estimated (two length)	estimated (three length)
$dx = dy$	0.992 ± 0.0153	0.995 ± 0.0119	0.999 ± 0.0157
$\bar{\psi}$	-0.0471 ± 0.0351	-0.0648 ± 0.0426	-0.0447 ± 0.0310
a_0	0.0329 ± 0.0115	-0.0337 ± 0.0952	-0.0943 ± 0.0635
q_1	-	0.571 ± 0.127	0.669 ± 0.142
a_1	-	0.0112 ± 0.0420	-0.0490 ± 0.0944
q_2	-	-	0.387 ± 0.296
a_2	-	-	-0.0144 ± 0.0966

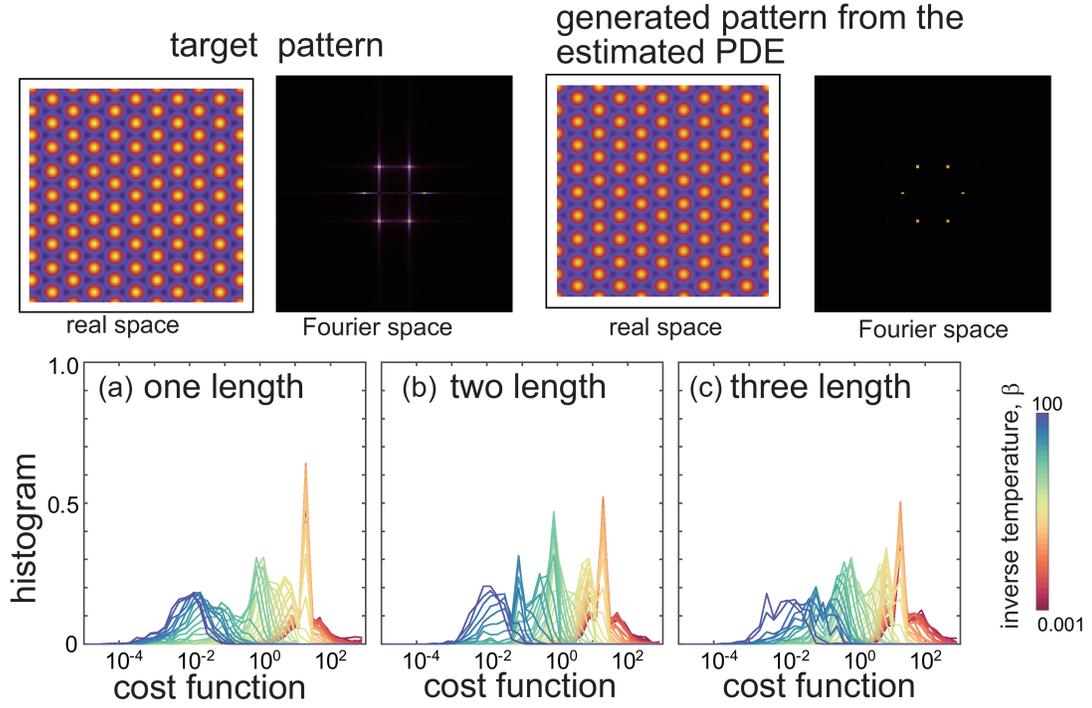


FIG. S9. (Top) The target pattern and the generated pattern from the estimated PDE in real and Fourier space. (Bottom) Energy histograms of the REMC sampling for the target pattern of a hexagonal pattern that is synthesised by the function. The horizontal axis is shown in the logarithmic scale.

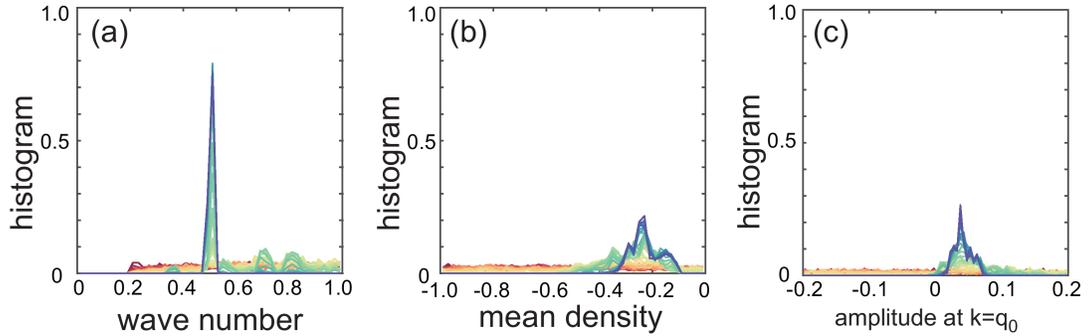


FIG. S10. Histogram of the estimated parameters for the target pattern of a hexagonal pattern that is synthesised by the function.

TABLE S4. The estimated parameters values for functionally synthesised hexagonal pattern. Errors are evaluated from the standard deviation of each data.

parameters	estimated (one length)	estimated (two length)	estimated (three length)
$dx = dy$	0.990 ± 0.00778	1.00720 ± 0.00642	0.991 ± 0.0149
$\bar{\psi}$	-0.226 ± 0.0536	-0.171 ± 0.0445	-0.215 ± 0.0729
a_0	0.0408 ± 0.0126	0.0438 ± 0.0341	-0.0240 ± 0.0800
q_1	-	0.492 ± 0.0162	0.288 ± 0.102
a_1	-	0.0534 ± 0.00713	-0.0245 ± 0.0970
q_2	-	-	0.431 ± 0.0953
a_2	-	-	-0.0789 ± 0.106

D. Target pattern synthesised by a function: double gyroid and Frank Kasper A15

Estimation of a PDE that generate 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 Se. SIII, 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: double gyroid (DG) and Frank Kasper A15 (FKA15). We have confirmed our method also works for simpler pattern such as lamellae pattern (stripe in three dimensions) and cylindrical pattern (hexagonal pattern in three dimensions).

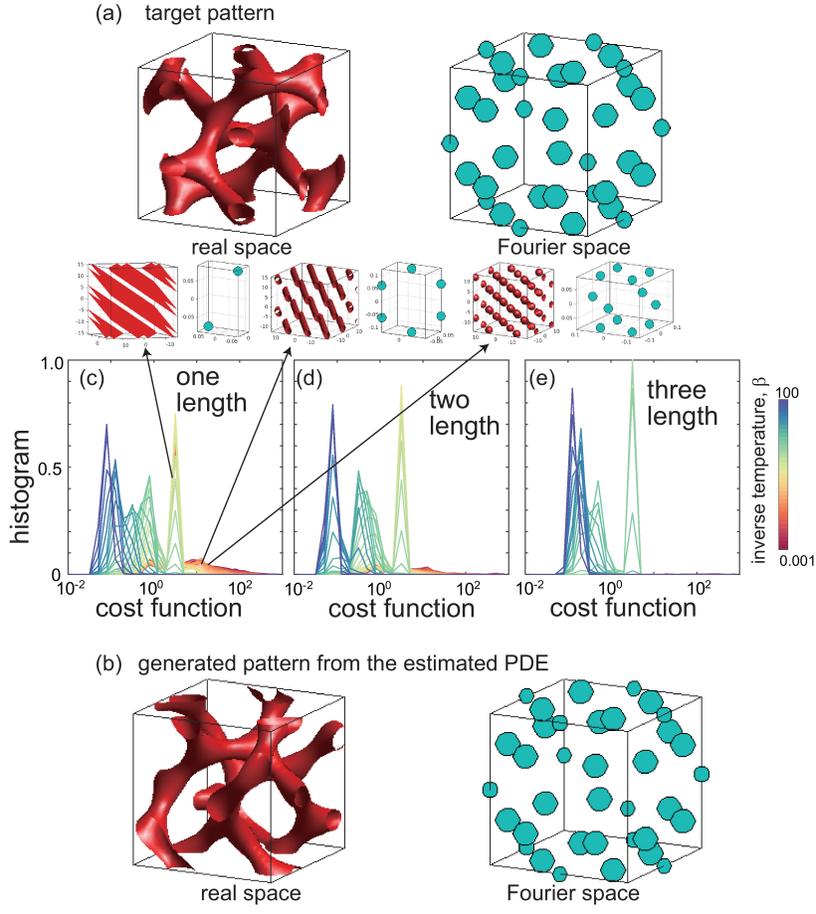


FIG. S11. (a,b) The target pattern and the generated pattern from the estimated PDE in real $\psi(\mathbf{x})$ and Fourier $|\tilde{\psi}(\mathbf{k})|$ space for the target pattern of double gyroid. In Fourier space, each point corresponds to a peak of $|\tilde{\psi}(\mathbf{k})|$, and the intensity is described by the size of the points. (c-e) Energy histogram of the REMC sampling for the target pattern of a double gyroid pattern that is synthesised by the function. The horizontal axis is shown in the logarithmic scale. Patterns with higher energy are also shown in the insets.

In Fig. S11, the energy histogram of each model is shown. The DG pattern is reproduced by one- ($m = m_1$) and two-length-scale ($m = m_2$) models. In fact, the generated pattern from the estimated PDE is identical to the target pattern up to translation, and they have almost the same patterns in the Fourier space (Fig. S11). The energy histograms of $m = m_1$ and $m = m_2$ are qualitatively similar, but $m = m_2$ has slightly smaller energy on average. The probability of the two models shown in Fig. 3(d) is comparable, but $m = m_2$ is chosen. 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.

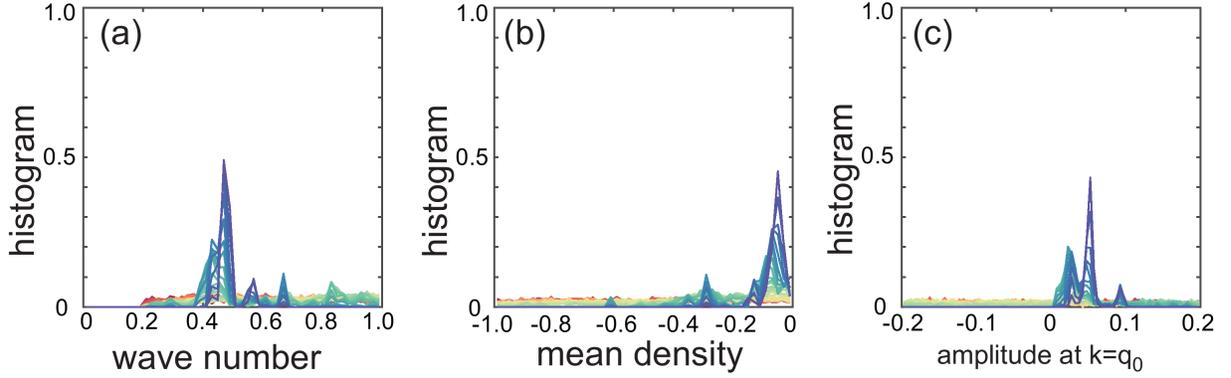


FIG. S12. Histogram of the estimated parameters for the target pattern of a double gyroid pattern that is synthesised by the function.

Estimated parameters are shown in Fig. S12 and Table. S5. The estimated wavenumber distributes near the wavelength that we have imposed. The mean density $\bar{\psi}$ is accumulated near $\bar{\psi} = 0$ but with slight deviation. This is consistent from previous theoretical and numerical results; the double gyroid pattern, in fact, appears between lamellae (stripe) and cylinder (hexagonal) patterns[11, 13, 84]. As discussed in Sec. SV B and Sec. SV C, 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 double gyroid 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.S12(b).

TABLE S5. The estimated parameters values for functionally synthesised double gyroid pattern in three dimensions. Errors are evaluated from the standard deviation of each sample from the REMC at the optimal temperature.

parameters	estimated (one length)	estimated (two length)	estimated (three length)
$dx = dy = dz$	0.528 ± 0.0261	0.561 ± 0.00380	0.541 ± 0.00264
$\bar{\psi}$	-0.0559 ± 0.0326	-0.090 ± 0.0266	-0.915 ± 0.0414
a_0	0.0548 ± 0.0149	-0.127 ± 0.0236	-0.127 ± 0.0369
q_1	-	0.838 ± 0.0187	0.0336 ± 0.00221
a_1	-	-0.112 ± 0.0208	-0.0797 ± 0.0410
q_2	-	-	0.974 ± 0.00233
a_2	-	-	0.157 ± 0.0236

The FKA15 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[85]. In the unit cell, a centre particle is surrounded by 8 particles at the position of the corners, and 2 particles are located at each face (Fig. S13(a)). In Fig. S13, the histogram of the cost function for each model is shown. The FKA15 pattern is reproduced by one- ($m = m_1$) and two-length-scale ($m = m_2$) models. The generated pattern from the estimated PDE is similar to the target pattern up to translation, and they have almost the same patterns in the Fourier space (Fig. S13(b)). The similarity is also demonstrated in their side views. The energy histograms are shown in Fig. S13(c-e). The probability of the two models shown in Fig. 3(e) is comparable, but $m = m_1$ is chosen for this target 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 Fig. S13(f-h). The estimated wavenumber 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, FKA15 has been found in higher resolution of $\bar{\psi} - a_0$ phase diagram using the self-consistent field theory, which describes block copolymers[86]. The position of FKA15 in the phase diagram used to be BCC, but recently several Frank Kasper 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, FKA15 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.

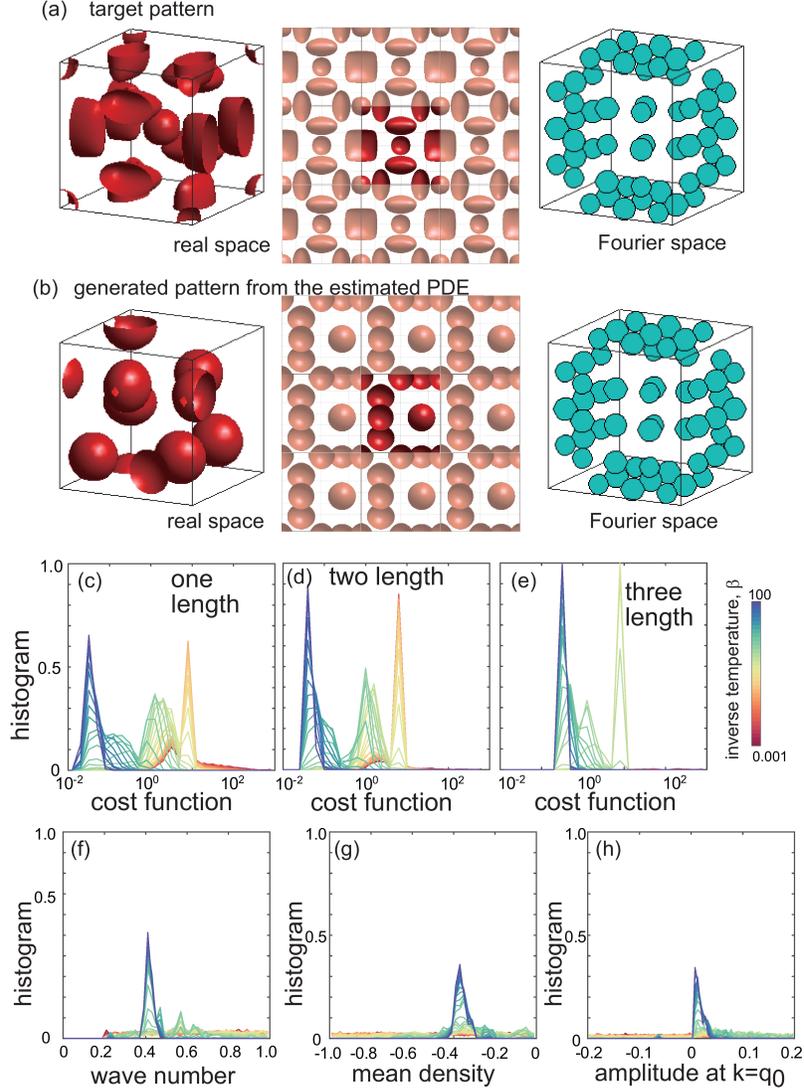


FIG. S13. (a,b) The target (a) and estimated (b) patterns in real $\psi(\mathbf{x})$ and Fourier $|\psi(\mathbf{k})|$ space for the target pattern of Frank Kasper A15. In Fourier space, each point corresponds to a peak of $|\psi(\mathbf{k})|$, and the intensity is described by the size of the points. The side view of the real-space images is also shown in the middle panels. In the side view, neighbouring 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 target pattern of a Frank Kasper A15 pattern that is synthesised by the function. The horizontal axis is shown in the logarithmic scale. Patterns with higher energy are also shown in the insets. (f-h) Histogram of the estimated parameters for the target pattern of a Frank Kasper A15 pattern that is synthesised by the function.

TABLE S6. The estimated parameters values for functionally synthesised Frank Kasper A15 pattern in three dimensions. Errors are evaluated from the standard deviation of each sample from the REMC at the optimal temperature.

parameters	estimated (one length)	estimated (two length)	estimated (three length)
$dx = dy = dz$	0.529 ± 0.0142	0.459 ± 0.00378	0.545 ± 0.00136
$\bar{\psi}$	-0.365 ± 0.0243	-0.377 ± 0.0265	-0.392 ± 0.090
a_0	0.0144 ± 0.007	-0.106 ± 0.0373	-0.106 ± 0.0616
q_1	-	0.885 ± 0.0278	0.0225 ± 0.0133
a_1	-	-0.0394 ± 0.0599	-0.0229 ± 0.0676
q_2	-	-	0.961 ± 0.0115
a_2	-	-	0.109 ± 0.0708

SVI. ROBUSTNESS AGAINST NOISE

We believe that the robustness is supported by our order parameter and marginalisation for an initial condition of pattern formation. To confirm this, we performed parameter estimation using the regression method in which the following cost function was used

$$E = \frac{1}{2} \int [\psi^* - f_{dt}(\psi^*; \mu)]^2 d\mathbf{x}. \quad (\text{S56})$$

Here, the target pattern ψ^* is numerically forward in dt by the model of equation (2) with a parameter set $\{\mu\}$ as $\psi(t + dt) = f_{dt}(\psi(t); \mu)$. If the target pattern is the stationary solution of the model, namely if the parameters are ground truth to obtain the target 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 *left-hand side* (time derivative) and *right-hand side* (force to change ψ), that is, $E = (1/2) \|\partial_t \psi - f(\psi; \mu)\|$, under an appropriate norm $\|\cdot\|$ [59, 60]. The norm is typically chosen as the L^2 norm

$$E = \frac{1}{2} \int [\partial_t \psi - f(\psi; \mu)]^2 d\mathbf{x}. \quad (\text{S57})$$

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, equation (S56). The method is similar to our main algorithm to sample parameters and to 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.

The cost function, equation (S56), implies that uncertainty arises from *state noise* added in the equation

$$\partial_t \psi = f(\psi; \mu) + \xi(\mathbf{x}, t) \quad (\text{S58})$$

where the noise ξ is taken from a normal distribution with zero mean and variance σ^2 . This is different from measurement noise

$$\psi = \psi^* + \xi \quad (\text{S59})$$

where the $\psi(\mathbf{x})$ is a pattern reproduced by a model and ψ^* is a target pattern. In this case, uncertainty arises from incomplete measurement and the fact that the model does not contain ground truth and cannot reproduce exactly the same pattern as the target pattern. This uncertainty results in correlation and nonlinear coupling by the nonlinear function of $f(\psi)$ with respect to ψ . Therefore, its likelihood is not necessarily given by the Gaussian distribution as equation (S57).

To overcome measurement noise, trajectory matching, referred to as the shooting method or the dynamical method, was proposed [54, 56]. In this method, $\psi(\mathbf{x}, t)$ is evaluated under the estimated parameters μ and initial conditions $\psi(\mathbf{x}, 0)$, is compared with data ψ^* , and then minimise the cost function $\|\psi^* - \psi\|$ with respect to parameters μ and the initial conditions $\psi(\mathbf{x}, 0)$. This method works only for time-series data because a stationary pattern does not have information about its initial condition. We would like to estimate the PDE that can generate patterns similar to the target pattern from a wide range of initial conditions. If the patterns are generated from very narrow initial conditions, they are hardly accessed. For example, a given target stripe pattern can be rotated, and we obtain another stripe

pattern (see Fig. 1(c) and Fig. S1). The rotated pattern is a solution of the same model with the same parameters under a different initial condition. For time-series data, the original stripe and rotated patterns are different because their initial conditions are different, and their structures during the relaxation process are also different. Estimation of a model for a stationary pattern does not care about such structures during relaxation, but does care only the model that results in the stationary pattern. This consideration suggests two things; one is to marginalise initial conditions and the second is to define the cost function that is independent of the initial conditions. In BM-PDE, the initial conditions are taken from random distribution each time a solution of a PDE model is evaluated. This process marginalise probability with respect to the initial conditions. BM-PDE also uses the cost function based on the order parameter, which identifies patterns under translation and rotation. These ingredients are necessary to consider an inverse problem of a stationary problem.

The result of the regression method is shown in the black points in Fig. 4(a) and Fig. S14). This method works only when the noise is small and it gives wrong estimation for larger noise. This result does not improve with the choice of the initial condition of the parameters; even when we choose the ground truth parameters as an initial condition, this method gives wrong estimates for the larger noise. This is because the noise is not along the relaxation path of the pattern for the ground truth, and therefore, the pattern significantly changes even at the ground truth parameters. As a result, the optimisation process tries to find the parameters to minimise the change, but those parameters are very different from the ground truth parameters. We should, however, note that this nonlinear regression method is much faster than our main algorithm, because we do not need to solve the PDE until a pattern reaches its stationary state.

To confirm the role of the cost function, we consider the following cost function:

$$E = \sum_l \left[\Psi_l^* - \Psi_l \left[f_{dt}^{\{\mu\}}(\psi^*) \right] \right]^2. \quad (\text{S60})$$

In this case, the target pattern is evolved in time by dt , we compute its cost function, and then we compare it by the cost function of the target pattern. This cost function is similar to the nonlinear regression method in equation (S56), but the order parameter is used instead of the density field itself. We should note that this method is also based on the regression method, and relies on the state noise. In fact, the cost function, equation (S60), implies that uncertainty arises from noise in the dynamical equation of Ψ , namely, $\partial_t \Psi = f_\Psi(\Psi) + \xi$ where the function f_Ψ is obtained from $f(\psi)$ by nonlinear transformation. To avoid an artificial estimate near $q_i \approx 0$, we use the prior uniform distribution $q_i \in [0.1, 1.0]$ in this case.

The result is shown in blue points in Fig. S14. The estimated wavenumbers are improved compared with the regression based on equation (S56). Still, the estimation deviates from the acceptable range around 2% of the noise amplitude. The acceptable range of the parameter is evaluated from the standard deviation of the estimated parameters in the noiseless case of the BM-PDE (see Fig. 2(b)) at which the cost function is below the gap. These parameters are taken from the samples in all temperatures as long as the corresponding cost function is $E \leq 10$. We can check all the sampled patterns below the gap are dodecagonal quasi-crystals. Note that this regression method using the order parameter works relatively well only for the target pattern with ground truth. When the target pattern is synthesised by the function, the method does not work and gives an entirely wrong estimate around $\hat{q}_1 \approx 0.3$.

Finally, we note that the robust estimation of the BM-PDE is further improved by using a pre-process, which is used for the target pattern synthesised by a function (see Methods). Under the application of the filter for the target pattern, estimation works even at 100% noise amplitude (the green points in Fig. S14). In the current case, all the peak of $\tilde{\psi}(\mathbf{k})$ have similar amplitude, and therefore we use $\alpha = 0.6$ in the pre-processing. When the target pattern has several peaks with different amplitudes, the choice of the parameter in the pre-process may affect the estimation because it may eliminate relevant peaks.

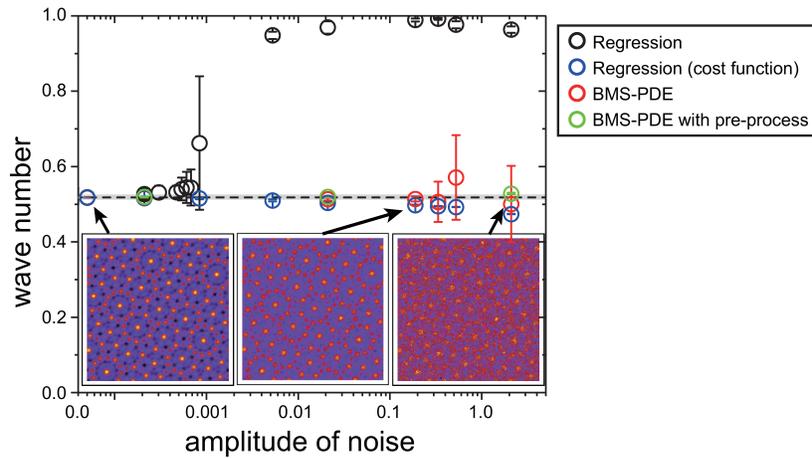


FIG. S14. (a) Estimated wavenumbers for numerically produced quasi-crystal by BM-PDE (red and green points) and conventional regression method by using equation (S56) (black points) and equation (S60) (blue points) under Gaussian white noise added on the target pattern. In addition to Fig. 4(a), BM-PDE with pre-process (green points) and conventional regression method using the cost function (blue points) are added. Noise amplitude with respect to the variance of the noiseless pattern is defined as $\sigma^2/\text{Var}[\psi^*]$ where σ^2 is variance of added noise. The horizontal dashed line indicates the ground truth of the wavenumber $q_1 = 0.51764$. The acceptable range of the parameter is shown by the gray area. The target patterns under the different noise amplitude is shown in the insets.

-
- [1] Swift, J. & Hohenberg, P. C. Hydrodynamic fluctuations at the convective instability. *Phys. Rev. A* **15**, 319–328 (1977).
 - [2] Lord Rayleigh, O. F. Lix. on convection currents in a horizontal layer of fluid, when the higher temperature is on the under side. *Lond. Edinb. Dubl. Phil. Mag.* **32**, 529–546 (1916).
 - [3] Taylor, G. I. Viii. stability of a viscous liquid contained between two rotating cylinders. *Phil. Trans. Royal Soc. A* **223**, 289–343 (1923).
 - [4] Stuart, J. T. On the non-linear mechanics of hydrodynamic stability. *J. Fluid Mech.* **4**, 1–21 (1958).
 - [5] Watson, J. 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**, 371–389 (1960).
 - [6] Lega, J., Moloney, J. V. & Newell, A. C. Swift-hohenberg equation for lasers. *Phys. Rev. Lett.* **73**, 2978–2981 (1994).
 - [7] Lefever, R., Barbier, N., Couteron, P. & Lejeune, O. Deeply gapped vegetation patterns: On crown/root allometry, criticality and desertification. *J. Theor. Biol.* **261**, 194 – 209 (2009).
 - [8] Ohta, T. & Kawasaki, K. Equilibrium morphology of block copolymer melts. *Macromolecules* **19**, 2621–2632 (1986).
 - [9] Bahiana, M. & Oono, Y. Cell dynamical system approach to block copolymers. *Phys. Rev. A* **41**, 6763 (1990).
 - [10] Fredrickson, G. H. & Helfand, E. Fluctuation effects in the theory of microphase separation in block copolymers. *J. Chem. Phys.* **87**, 697–705 (1987).
 - [11] Podneps, V. E. & Hamley, I. W. Landau- brazovskii theory for the $ia\bar{3}d$ structure. *J. Exp. Theor. Phys. Lett.* **64**, 617–624 (1996).
 - [12] Pismen, L. M. *Patterns and interfaces in dissipative dynamics* (Springer, 2006).
 - [13] Jaatinen, A. & Ala-Nissila, T. Extended phase diagram of the three-dimensional phase field crystal model. *J. Phys. Cond. Mat.* **22**, 205402 (2010).
 - [14] Nonomura, M. & Ohta, T. Kinetics of morphological transitions between mesophases. *J. Phys. Cond. Mat.* **13**, 9089–9112 (2001).
 - [15] Elder, K. R., Katakowski, M., Haataja, M. & Grant, M. Modeling elasticity in crystal growth. *Phys. Rev. Lett.* **88**, 245701 (2002).
 - [16] Provatas, N. & Elder, K. *Phase-field methods in materials science and engineering* (John Wiley & Sons, 2011).
 - [17] Kléman, M. *Points, Lines and Walls: in Liquid Crystals, Magnetic Systems, and Various Disordered Media* (Wiley, New York, 1983).
 - [18] Harrison, C. *et al.* Mechanisms of ordering in striped patterns. *Science* **290**, 1558–1560 (2000).
 - [19] Kléman, M. & Laverntovich, O. D. *Soft matter physics: an introduction* (Springer, 2003).
 - [20] Elder, K. R., Provatas, N., Berry, J., Stefanovic, P. & Grant, M. Phase-field crystal modeling and classical density functional theory of freezing. *Phys. Rev. B* **75**, 064107 (2007).
 - [21] Lifshitz, R. & Petrich, D. M. Theoretical model for faraday waves with multiple-frequency forcing. *Phys. Rev. Lett.* **79**,

- 1261–1264 (1997).
- [22] Savitz, S., Babadi, M. & Lifshitz, R. Multiple-scale structures: from Faraday waves to soft-matter quasicrystals. *IUCrJ* **5**, 247–268 (2018).
- [23] Subramanian, P., Archer, A. J., Knobloch, E. & Rucklidge, A. M. Three-dimensional icosahedral phase field quasicrystal. *Phys. Rev. Lett.* **117**, 075501 (2016).
- [24] Müller, H. W. Model equations for two-dimensional quasipatterns. *Phys. Rev. E* **49**, 1273–1277 (1994).
- [25] Mermin, N. & Troian, S. M. Mean-field theory of quasicrystalline order. *Phys. Rev. Lett.* **54**, 1524 (1985).
- [26] Kats, E., Lebedev, V. & Muratov, A. Weak crystallization theory. *Phys. Rep.* **228**, 1–91 (1993).
- [27] Malomed, B., Nepomnyashchii, A. & Tribelskii, M. Two-dimensional quasiperiodic structures in nonequilibrium systems. *Sov. Phys. JETP* (1988).
- [28] Zhang, Z. & Glotzer, C., Sharon. Self-assembly of patchy particles. *Nano Lett.* **4**, 1407–1413 (2004).
- [29] Damasceno, P. F., Engel, M. & Glotzer, S. C. Predictive self-assembly of polyhedra into complex structures. *Science* **337**, 453–457 (2012).
- [30] Whitesides, G. M. & Grzybowski, B. Self-assembly at all scales. *Science* **295**, 2418–2421 (2002).
- [31] Li, F., Josephson, D. P. & Stein, A. Colloidal assembly: The road from particles to colloidal molecules and crystals. *Angew. Chem. Int. Edit.* **50**, 360–388 (2011).
- [32] Boles, M. A., Engel, M. & Talapin, D. V. Self-assembly of colloidal nanocrystals: From intricate structures to functional materials. *Chem. Rev.* **116**, 11220–11289 (2016). PMID: 27552640.
- [33] Rechtsman, M. C., Stillinger, F. H. & Torquato, S. Optimized interactions for targeted self-assembly: Application to a honeycomb lattice. *Phys. Rev. Lett.* **95**, 228301 (2005).
- [34] Rechtsman, M., Stillinger, F. & Torquato, S. Designed interaction potentials via inverse methods for self-assembly. *Phys. Rev. E* **73**, 011406 (2006).
- [35] Torquato, S. Inverse optimization techniques for targeted self-assembly. *Soft Matter* **5**, 1157–1173 (2009).
- [36] Torquato, S. Optimal design of heterogeneous materials. *Annu. Rev. Mater. Res.* **40**, 101–129 (2010).
- [37] Jain, A., Bollinger, J. A. & Truskett, T. M. Inverse methods for material design. *AIChE J.* **60**, 2732–2740 (2014).
- [38] Sherman, Z. M., Howard, M. P., Lindquist, B. A., Jadrlich, R. B. & Truskett, T. M. Inverse methods for design of soft materials. *J. Chem. Phys.* **152**, 140902 (2020).
- [39] Lindquist, B. A., Jadrlich, R. B. & Truskett, T. M. Communication: Inverse design for self-assembly via on-the-fly optimization. *J. Chem. Phys.* **145**, 111101 (2016).
- [40] Tsai, C. L., Delaney, K. T. & Fredrickson, G. H. Genetic algorithm for discovery of globally stable phases in block copolymers. *Macromolecules* **49**, 6558–6567 (2016).
- [41] Paradiso, S. P., Delaney, K. T. & Fredrickson, G. H. Swarm intelligence platform for multiblock polymer inverse formulation design. *ACS Macro Letters* **5**, 972–976 (2016).
- [42] Khadilkar, M. R., Paradiso, S., Delaney, K. T. & Fredrickson, G. H. Inverse design of bulk morphologies in multiblock polymers using particle swarm optimization. *Macromolecules* **50**, 6702–6709 (2017).
- [43] Khaira, G. S. *et al.* Evolutionary optimization of directed self-assembly of triblock copolymers on chemically patterned substrates. *ACS Macro Lett.* **3**, 747–752 (2014).
- [44] Kumar, R., Coli, G. M., Dijkstra, M. & Sastry, S. Inverse design of charged colloidal particle interactions for self assembly into specified crystal structures. *J. Chem. Phys.* **151**, 084109 (2019).
- [45] Miskin, M. Z., Khaira, G., de Pablo, J. J. & Jaeger, H. M. Turning statistical physics models into materials design engines. *Proc. Nat. Acad. Sci. USA* **113**, 34–39 (2016).
- [46] Lindquist, B. A., Jadrlich, R. B. & Truskett, T. M. Assembly of nothing: equilibrium fluids with designed structured porosity. *Soft Matter* **12**, 2663–2667 (2016).
- [47] Piñeros, W. D., Lindquist, B. A., Jadrlich, R. B. & Truskett, T. M. Inverse design of multicomponent assemblies. *J. Chem. Phys.* **148**, 104509 (2018).
- [48] Piñeros, W. D., Jadrlich, R. B. & Truskett, T. M. Design of two-dimensional particle assemblies using isotropic pair interactions with an attractive well. *AIP Adv.* **7**, 115307 (2017).
- [49] Adorf, C. S., Antonaglia, J., Dshemuchadse, J. & Glotzer, S. C. Inverse design of simple pair potentials for the self-assembly of complex structures. *J. Chem. Phys.* **149**, 204102 (2018).
- [50] Lindquist, B. A., Jadrlich, R. B. & Truskett, T. M. Communication: From close-packed to topologically close-packed: Formation of laves phases in moderately polydisperse hard-sphere mixtures. *J. Chem. Phys.* **148**, 191101 (2018).
- [51] Jaeger, H. M. & de Pablo, J. J. Perspective: Evolutionary design of granular media and block copolymer patterns. *APL Mater.* **4**, 053209 (2016).
- [52] Tokuda, S., Nagata, K. & Okada, M. Simultaneous estimation of noise variance and number of peaks in bayesian spectral deconvolution. *J. Phys. Soc. Japan* **86**, 024001 (2017).
- [53] Bär, M., Hegger, R. & Kantz, H. Fitting partial differential equations to space-time dynamics. *Phys. Rev. E* **59**, 337–342 (1999).
- [54] Müller, T. G. & Timmer, J. Parameter identification techniques for partial differential equations. *Int. J. Bifurcat. Chaos* **14**, 2053–2060 (2004).
- [55] Zhao, H., Storey, B. D., Braatz, R. D. & Bazant, M. Z. Learning the physics of pattern formation from images. *Phys. Rev. Lett.* **124**, 060201 (2020).
- [56] Voss, H. U., Timmer, J. & Kurths, J. Nonlinear dynamical system identification from uncertain and indirect measurements. *Int. J. Bifurcat. Chaos* **14**, 1905–1933 (2004).
- [57] Bongard, J. & Lipson, H. Automated reverse engineering of nonlinear dynamical systems. *Proc. Nat. Acad. Sci. USA* **104**,

- 9943–9948 (2007).
- [58] Schmidt, M. & Lipson, H. Distilling free-form natural laws from experimental data. *Science* **324**, 81–85 (2009).
- [59] Brunton, S. L., Proctor, J. L. & Kutz, J. N. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc. Nat. Acad. Sci. USA* **113**, 3932–3937 (2016).
- [60] Schaeffer, H. Learning partial differential equations via data discovery and sparse optimization. *Proc. Royal Soc. A* **473**, 20160446 (2017).
- [61] Rudy, S. H., Brunton, S. L., Proctor, J. L. & Kutz, J. N. Data-driven discovery of partial differential equations. *Sci. Adv.* **3**, e1602614 (2017).
- [62] Schaeffer, H. & McCalla, S. G. Sparse model selection via integral terms. *Phys. Rev. E* **96**, 023302 (2017).
- [63] Xu, H., Chang, H. & Zhang, D. Dl-pde: Deep-learning based data-driven discovery of partial differential equations from discrete and noisy data. *arXiv:1908.04463* (2019).
- [64] Bishop, C. M. *Pattern recognition and machine learning I*, vol. 1 (Springer New York, 2006).
- [65] Kaipio, J. & Somersalo, E. *Statistical and computational inverse problems*, vol. 160 (Springer Science & Business Media, 2006).
- [66] Xun, X., Cao, J., Mallick, B., Maity, A. & Carroll, R. J. Parameter estimation of partial differential equation models. *J. Am. Stat. Assoc.* **108**, 1009–1020 (2013).
- [67] Campillo-Funollet, E., Venkataraman, C. & Madzvamuse, A. Bayesian parameter identification for Turing systems on stationary and evolving domains. *Bull. Math. Biol.* **81**, 81–104 (2019).
- [68] Toni, T., Welch, D., Strelkowa, N., Ipsen, A. & Stumpf, M. P. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J. R. Soc. Interface* **6**, 187–202 (2009).
- [69] Elder, K. R. & Grant, M. Modeling elastic and plastic deformations in nonequilibrium processing using phase field crystals. *Phys. Rev. E* **70**, 051605 (2004).
- [70] Zhang, P. & Zhang, X. An efficient numerical method of Landau-Brazovskii model. *J. Comp. Phys.* **227**, 5859 – 5870 (2008).
- [71] Foucart, S. & Rauhut, H. *A Mathematical Introduction to Compressive Sensing* (Springer Science & Business Media, New York, NY, 2013).
- [72] Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *J. Royal Stat. Soc. B* **67**, 301–320 (2005).
- [73] Edmonds, A. *Angular momentum in quantum mechanics* (Princeton Univ. Press, 1957).
- [74] Steinhardt, P. J., Nelson, D. R. & Ronchetti, M. Bond-orientational order in liquids and glasses. *Phys. Rev. B* **28**, 784–805 (1983).
- [75] Efron, B. & Morris, C. Stein’s estimation rule and its competitors—an empirical Bayes approach. *J. Am. Stat. Assoc.* **68**, 117–130 (1973).
- [76] Akaike, H. Likelihood and the Bayes procedure. In *Selected papers of Hirotugu Akaike*, 309–332 (Springer, 1998).
- [77] MacKay, D. J. Bayesian interpolation. *Neural Comput.* **4**, 415–447 (1992).
- [78] Geyer, C. J. Markov chain Monte Carlo maximum likelihood. *Computing science and statistics: Proceedings of the 23rd Symposium on the Interface, American Statistical Association* **156** (1991).
- [79] Hukushima, K. & Nemoto, K. Exchange Monte Carlo method and application to spin glass simulations. *J. Phys. Soc. Japan* **65**, 1604–1608 (1996).
- [80] Meng, X.-L. & Wong, W. H. Simulating ratios of normalizing constants via a simple identity: a theoretical exploration. *Statistica Sinica* 831–860 (1996).
- [81] Gelman, A. & Meng, X.-L. Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical Science* 163–185 (1998).
- [82] Efron, B. Bootstrap methods: another look at the jackknife. In *Breakthroughs in statistics*, 569–593 (Springer, 1992).
- [83] Robbins, M. J., Archer, A. J., Thiele, U. & Knobloch, E. Modeling the structure of liquids and crystals using one- and two-component modified phase-field crystal models. *Phys. Rev. E* **85**, 061408 (2012).
- [84] Shi, A.-C. Nature of anisotropic fluctuation modes in ordered systems. *J. Phys. Cond. Mat.* **11**, 10183 (1999).
- [85] Impéror-Clerc, M. Three-dimensional periodic complex structures in soft matter: investigation using scattering methods. *Interface Focus* **2**, 589–601 (2012).
- [86] Bates, M. W. *et al.* Stability of the a15 phase in diblock copolymer melts. *Proc. Nat. Acad. Sci. USA* **116**, 13194–13199 (2019).