Image Inversion with Uncertainty Quantification for Pattern-Forming Systems

Hongbo Zhao,1 Richard D. Braatz,1 and Martin Z. Bazant1,2

1Department of Chemical Engineering, Massachusetts Institute of Technology
77 Massachusetts Avenue, Cambridge, MA 02139
2Department of Mathematics, Massachusetts Institute of Technology
77 Massachusetts Avenue, Cambridge, MA 02139

(Dated: October 22, 2020)

Extensive theoretical studies have led to the discovery of various continuum models for pattern-forming systems and computational advances have greatly empowered the tool set of simulating increasingly complex systems. Yet as a complement to forward problems, the numerical robustness of the inverse problem as well as its uncertainty quantification is less well understood. Here we use PDE-constrained optimization, Bayesian statistics, and analytical methods to infer the governing dynamics and constitutive relations from images of pattern formation and understand their uncertainties in different systems, operating conditions, and imaging conditions. We study the datasets and physical constraints needed to increase the inference accuracy and well-posedness of the inverse problem. We demonstrate the procedure and uncertainty of inversion under limited spatiotemporal resolution, unknown boundary condition, blurry initial conditions, and other nonideal situations. Phase field, reaction-diffusion, and phase field crystal models are used as model systems. This approach can be applied in the inference of unknown dynamics and physical properties, experimental design, and engineering of complex pattern-forming systems.

I. INTRODUCTION

Beyond the aesthetic value of pattern formation observed widely in many systems [1], their images can be harnessed to distill useful physical properties and test theoretical models quantitatively. In contrast to natural images and neural network models, images of pattern formation often lie on a manifold that can be described by relatively simple partial differential equations (PDE), despite extremely high degrees of freedom at the microscopic scale [2] and the richness of the pattern itself [3]. Identification of the governing equation behind snapshots of an evolving pattern helps to uncover its mechanism as well as its constitutive relations with a small number of experiments [4]. Physically, such identification reveals nonequilibrium behaviors difficult to compute from first principles, providing an alternative approach to modeling complex behavior. However, due to the nonlinearity of the PDEs, it remains unclear how reliably physics can be extracted from the images.

Analytical approaches have yielded valuable insights into the scaling laws of pattern formation [5] [6] and hence how the dynamics depend on key parameters. For spinodal decompositions, it is well known that the characteristic wavenumber of the initial spinodal pattern is related to the second derivative of the free energy [7], and when the coarsening is governed by diffusion, the domain size grows as $t^{1/3}$ in time [8] [10]. The growth law is found to vary with different models of concentration-dependent diffusivity [11]. For nonconserved fields, Ginzburg-Landau theory predicts a well-known $t^{1/2}$ growth law [12] [13]. The structure factor of the pattern at different stages of the coarsening has also been found in both simulations and experiments to collapse with the appropriate scaling [8] [12].

However, scaling and linear stability analysis do not offer detailed description of the morphology in real space. The sensitivity of the patterns with respect to constitutive relations (especially state-dependent or nonlocal constitutive relations) under different conditions remains unclear. The growth law only applies when the domain is sufficiently large and no boundary effects are present [9] [10]. In addition, for systems with competing dynamics such as a pattern-forming reaction-diffusion system, the pattern depends sensitively on diffusivity and reaction kinetics [3] [13], and does not necessarily follow a simple growth law. Hence, in this work, we identify the governing equation through PDE-constrained optimization and use uncertainty quantification to assess the accuracy of the constitutive relations inferred from images in real space.

Previous studies have demonstrated the possibility of inferring constitutive relations or governing equations from images [1] [14] [20]. However, for pattern-forming systems, where the morphology of the patterns can be strongly influenced by the specific conditions and governing dynamics, a systematic analysis of inversion based on images obtained under different conditions is lacking. Feature engineering [21] and dimensionality reduction techniques [22] compress the high-dimensional image data to a small number of parameters. Recently, it was found that the learned low-dimensional manifold for the spinodal decomposition images is well correlated with the free energy barrier and average concentration [23] [24]. Here we take the inverse problem approach: rather than creating a forward mapping from physical properties to patterns, we ask what physical properties can be accurately inferred from the patterns. In doing so, the full dataset is used to maximize the retrieval of useful information.

As with any regression problem, care must be taken with regard to generalization and extrapolation. Regu-
larization is needed for the inversion of functions (infinite-dimensional inverse problems) to ensure the problem is well-posed. For example, neural networks have been used to discover physical laws and constitutive relations by incorporating physical constraints \[25\,\,31\]. Sparse or symbolic regression are also commonly used to achieve parsimony and better physical interpretability \[32\,\,37\]. Our approach enforces the general form of the governing equation while allowing the unknown dynamics (such as reaction and diffusion) to be identified. We also allow the unknown constitutive relations to be nonlinearly or nonlocally dependent on the order parameters. Additionally, imposing symmetry in the constitutive relations and other prior knowledge such as the miscibility gap can narrow down the uncertainty and prevent overfitting.

Recently, surrogate models such as Gaussian processes \[38\], deep neural networks \[39\], generative adversarial networks \[40\], and physics-informed neural networks \[41\] have been used in inverse uncertainty quantification. Here we choose the full Bayesian approach \[17\,\,42\,\,43\] based on a single realization of snapshots and the full PDE model to infer the posterior distribution of the constitutive relations.

Our method offers a top-down approach to the construction of constitutive relations of complex systems directly from macroscopically observed fields. It is complementary to the bottom up approach, where scale bridging and the learning of closure models from statistical physics are used to perform multiscale simulations \[44\,\,47\]. Both approaches eventually converge to PDEs for macroscopic observables that satisfy physical constraints.

We develop methods for solving PDE-constrained optimization and uncertainty quantification in Section \[II\]. In Section \[III\] we apply the methods to different pattern-forming systems and operating or imaging conditions, and address the questions of what information and how accurately that information can be learned from the available images, whether inferred quantities are correlated and whether models can be distinguished. We study systems driven by diffusion and/or reaction, as they relax toward equilibrium or when they are chemically driven out of equilibrium. We also study the effect of imaging conditions including spatial and temporal resolutions, domain size, as well as blurring.

II. METHOD

We study a class of pattern-forming systems that has a nonconvex free energy \(F[c]\) as a functional of the order parameter field \(c(x)\). Phase separation occurs due to instability, that is, there exists \(\delta c\) such that \(\delta^2 F < 0\). When the order parameter is a conserved parameter such as concentration, the dynamics can be driven toward equilibrium by diffusion,

\[
\frac{\partial c}{\partial t} = \nabla \cdot (D(c)c \nabla \mu). \tag{1}
\]

where \(\mu = \delta F/\delta c\). In chemically driven systems where the interior of the system is in contact with the chemical reservoir, as found in surface adsorption and surface-reaction-limited nanoparticles \[48\], the governing equation is

\[
\frac{\partial c}{\partial t} = -R_0(c)\mu, \tag{2}
\]

Being in contact with a chemical reservoir, the free energy is now \(F[c] - \mu_{res} \int c dx\) with the addition of a Lagrange multiplier. The equation is applicable for small deviations from equilibrium where linear irreversible thermodynamics \[49\] holds.

Thermodynamically, the region of instability \(\delta^2 F < 0\) is prohibited unless the system is out of the equilibrium. Therefore, temporal evolution of the patterns gives us access to its properties in the unstable region. In the case of a phase-separating system, the free energy is typically described by a Ginzburg-Landau type double-well energy with a gradient penalty term,

\[
F[c(r)] = \int (g_h(c(r)) + \kappa |\nabla c|^2) d\mathbf{r}. \tag{3}
\]

When coupled with diffusion as in Eq. 1 this expression becomes the well-known Cahn-Hilliard equation used extensively in phase field models \[50\]. When coupled with Eq. 2 this expression is known as the Allen-Cahn equation. In this case, we attempt to learn the double-well free energy \(g_h(c)\) as well as the dynamics (diffusivity, and reaction kinetics). We define the homogeneous chemical potential to be \(\mu_h(c) = \frac{\partial g_h(c)}{\partial c}\). Hence the chemical potential is

\[
\mu = \mu_h(c) - \kappa \nabla^2 c. \tag{4}
\]

Another type of pattern-forming system has a more generic nonlocal form,

\[
F[c(r)] = \int g_h(c(r)) d\mathbf{r} + \int c(\mathbf{r}')C_2(|\mathbf{r}' - \mathbf{r}|)c(\mathbf{r}) d\mathbf{r} d\mathbf{r}'. \tag{5}
\]

The conserved diffusion equation Eq. 1 combined with Eq. 5 is also known as the phase field crystal equation (or dynamic density functional theory) \[51\]. The nonconserved Eq. 2 combined with Eq. 5 is known as the Swift-Hohenberg equation \[51\,\,52\]. Different direct correlation functions \(C_2\) can give rise to spatial patterns such as lamellar and crystal lattice structures. In this case, we are interested in the sensitivity of the pattern with respect to the direct correlation function.

Suppose the measurement noise of the images follows an independent Gaussian distribution, \(c(t, \mathbf{r}; \mathbf{p}) - c_{\text{data}}(t, \mathbf{r}') \sim \mathcal{N}(0, \sigma^2 \delta(t_j - t, \mathbf{r} - \mathbf{r}'))\), where \(\mathbf{p}\) are the parameters for the unknown constitutive relations. The conditional probability of the observed data, aka the likelihood, is

\[
P(c_{\text{data}} | \mathbf{p}) = \exp \left( -\frac{1}{2\sigma^2} \sum_{i=1}^{M} \int d\mathbf{r} \left( c(t_i, \mathbf{r}; \mathbf{p}) - c_{\text{data}}(t_i, \mathbf{r}) \right)^2 \right). \tag{6}
\]
Similarly, we can also define the likelihood when the observed data is discrete in space and the noise is spatially and temporally independent, where the integral becomes a summation. From a Bayesian perspective, the posterior distribution of the unknown parameters \( \mathbf{p} \) satisfies

\[
P(\mathbf{p}|\mathbf{c}_{\text{data}}) \propto P(\mathbf{c}_{\text{data}}|\mathbf{p}) P(\mathbf{p}),
\]

where \( P(\mathbf{p}) \) is the prior probability, which depends on the prior knowledge of the constitutive relations discussed in detail below.

For binary mixtures, we express the free energy as a sum of the ideal entropy of mixing to limit the concentration within \([0,1]\) and a non-ideal excess part. The corresponding chemical potential is

\[
\mu_h(c) = \mu_0(c) + \mu_{\text{ex}}(c) = \ln \frac{c}{1-c} + \sum_{n=1}^{N} a_n P_n(c),
\]

where the \( P_n \) are normalized Legendre polynomials defined on the interval \([0,1]\), and \( a_n \) are the coefficients to be determined. Similarly, the diffusivity (or reaction kinetic prefactor) can be parameterized as \( \ln D(c) = \sum_n b_n P_n(c) \) to ensure positivity. Assuming the prior for the non-ideal part of the (excess) chemical potential follows a Gaussian distribution \( \mu_{\text{ex}}(c) \sim N(0, \delta(c-c')) \), or similarly if the prior for diffusivity is \( \ln D(c) \sim N(0, \delta(c-c')) \), then

\[
P(\mathbf{p}) \propto \exp -\frac{1}{2} \| \mathbf{p} \|^2_2,
\]

where \( \mathbf{p} \) are the coefficients \( a_n \). The covariance for the priors may be defined by differential operators to penalize high-frequency components in the constitutive relations. For example, \( \ln D(c) \sim N(0, \mathcal{L}^{-1}) \), where \( \mathcal{L} \psi(c) = \frac{d}{dc} [c(1-c) \frac{d}{dc} \psi(c)] \). Legendre polynomials (defined on \([0,1]\)) are eigenfunctions of the differential operator, \( \mathcal{L} P_n(c) = 4n(n+1) P_n(c) \). Hence,

\[
P(\mathbf{p}) \propto \exp -\frac{1}{2} \sum_{i=1}^{\infty} \frac{p_i^2}{4n(n+1)}.
\]

Note that the coefficient for the constant term \( P_0(c) \) is not included in the sum above (a normal distribution degenerate in \( p_0 \)).

The direct correlation function in Eq. 5 is represented in Fourier space in the form

\[
\hat{C}_2(k) = \sum_{n=0}^{N} d_n (2^n n! \sqrt{\pi})^{-1/2} e^{-|k|^2/4} H_n(|k|),
\]

where \( H_n \) is the physicists’ Hermite polynomial and the basis functions are orthonormal. Hence, if the prior is \( \hat{C}_2(k) \sim N(0, \delta(k-k')) \), the prior probability distribution can be written as Eq. 9 where \( \mathbf{p} \) are the coefficients \( d_n \).

For a general prior \( \mathbf{p} \sim N(0, \mathbf{\Gamma}_{\text{pr}}) \), it is convenient to transform \( \mathbf{z} = \mathbf{\Gamma}_{\text{pr}}^{-1/2} \mathbf{a} \) so that \( \mathbf{z} \sim N(0, \mathbf{I}) \), which is shown below to be useful for the Markov chain Monte Carlo (MCMC) sampling and linear constraints.

In addition to a prior that promotes smoothness and penalizes high-order basis functions, priors can be modified to satisfy certain constraints. For example, in addition to the images that capture the transient behavior, the compositions of equilibrium phases \( (c_1,c_2) \) also known as the miscibility gap – can be easily accessible; hence the chemical potential is subject to the thermodynamic constraint,

\[
\int_{c_1}^{c_2} \mu_h(c) dc = \mu_h(c_1)(c_2 - c_1)
\]

\[
\mu_h(c_1) = \mu_h(c_2).
\]

For the Cahn-Hilliard equation, an arbitrary constant term can be added to \( \mu_h(c) \), so we can set \( \mu_h(c_1) = \mu_h(c_2) = 0 \) as linear constraints on the coefficients \( \mathbf{B} \mathbf{z} = \mathbf{d} \). For a general prior on the coefficients, it is convenient to decompose \( \mathbf{z} \) into two orthonormal spaces, \( \mathbf{A} \) and \( \mathbf{A}^\perp \), where \( \mathbf{A} \) is the null space of \( \mathbf{B} \), \( \mathbf{B} \mathbf{A} = 0 \). Then we have \( \mathbf{z} = \mathbf{A} \xi + \mathbf{A}^\perp \xi^\perp \). Thus the constraint is \( \mathbf{B} \mathbf{A}^\perp = \mathbf{d} \). It can be shown that the conditional prior is

\[
p(\mathbf{z}|\mathbf{B} \mathbf{z} = \mathbf{d}) \propto \exp -\frac{1}{2} \| \xi^\perp \|^2_2.
\]

According to Bayes theorem, the posterior of the unknown parameters is \( P(\mathbf{p}|\mathbf{c}_{\text{data}}) \propto P(\mathbf{c}_{\text{data}}|\mathbf{p}) P(\mathbf{p}) \). The maximum a posteriori estimate (MAP) is defined by minimizing the objective function

\[
S(\mathbf{p}) = \frac{1}{2\sigma^2} \left[ \sum_{i=1}^{M} \int d\mathbf{r} \left( c(t_i, \mathbf{r}; \mathbf{p}) - c_{\text{data}}(t_i, \mathbf{r}) \right)^2 \right] + \frac{1}{2} \mathbf{p}^T \mathbf{\Gamma}_{\text{pr}}^{-1} \mathbf{p}.
\]

The \( L_2 \) norm corresponds to the prior \( \mathbf{E} \) while the \( L_\infty \) norm corresponds to Eq. 10. Without loss of generality, 10 parameters are used for all functions studied here, that is, Legendre polynomials of order 1 to 10 are used for the chemical potential, and of order 0 to 9 are used for the diffusivity and reaction kinetics. When the thermodynamic constraint (Eq. 12) is applied, Legendre polynomials of order 1 to 13 are used.

The objective function is optimized using a gradient-based optimizer, which is fed

\[
\frac{\partial S}{\partial \mathbf{p}} = \frac{1}{\sigma^2} \sum_{i=1}^{M} \int d\mathbf{r} \frac{\partial c}{\partial \mathbf{p}} (c(t_i, \mathbf{r}; \mathbf{p}) - c_{\text{data}}(t_i, \mathbf{r})) + \mathbf{\Gamma}_{\text{pr}}^{-1} \mathbf{p}.
\]

The model sensitivity \( \partial c/\partial \mathbf{p} \) for each parameter can be computed while solving the forward problem,

\[
\frac{\partial}{\partial t} \left( \frac{\partial c}{\partial \mathbf{p}} \right) = \frac{\partial}{\partial \mathbf{p}} \left( \frac{\partial c}{\partial t} \right) \cdot \frac{\partial c}{\partial \mathbf{p}} + \frac{\partial}{\partial \mathbf{p}} \left( \frac{\partial c}{\partial t} \right),
\]

which is known as forward sensitivity analysis (FSA). From the FSA, we estimate the Hessian of the objective function using Gauss-Newton approximation [53],

\[
\mathbf{H}[S] = \frac{1}{\sigma^2} \sum_{i=1}^{M} \int d\mathbf{r} \left( \frac{\partial^2 c}{\partial \mathbf{p}^2} \right)^* \frac{\partial c}{\partial \mathbf{p}} + \mathbf{\Gamma}_{\text{pr}}^{-1}.
\]
The approximation becomes increasingly accurate as the difference between the model and data is decreased. An alternative to FSA is adjoint sensitivity analysis (ASA), which involves solving the adjoint linear sensitivity equation backward in time once to obtain the sensitivity of the objective function with respect to all parameters [54–56] (equivalent to backpropagation in the field of machine learning). This approach is particularly useful for a large number of parameters, such as a 2D field or weights in neural nets. When the number of parameters is small, such as in the case of parameterizing concentration-dependent functions, the benefit of using FSA to obtain an estimate of the Hessian outweighs its computational cost, as the convergence is much faster. The trust-region algorithm is used throughout.

Given the posterior distribution \( P(\mathbf{p}|\text{data}) \), we quantify the uncertainty by sampling the parameter space. The unknown functions are parameterized in a finite-dimensional space. For a large number of parameters, the traditional Markov chain Monte Carlo (MCMC) is often slow, even when the proposal distribution has a covariance that is updated adaptively from the chain samples [43]. The mixing quality of the Markov chain deteriorates rapidly with finer representations of the functions. The sensitivity with respect to high-order basis functions leads to poorer sampling, and the sampler stagnates for a prolonged period of time, which is especially problematic for computationally expensive model evaluations of PDEs that we study here. Here we use a dimension-independent and likelihood-informed (DILI) MCMC [57] that takes the local Hessian information and adopts an operator-weighted proposal distribution to achieve better sampling efficiency that is independent of the parameterization of the function. We start the algorithm from the optimal solution (MAP). The Hessian is computed periodically using Eq. [18] to determine the parameter subspace that are most informed by the model (likelihood-informed subspace). The posterior covariance \( \Sigma \) is more accurately estimated by combining the covariance-dependent functions, the benefit of using FSA to obtain an estimate of the Hessian outweighs its computational cost, as the convergence is much faster. The trust-region algorithm is used throughout.

The unknown functions are parameterized in a finite-dimensional space. For a large number of parameters, such as a 2D field or weights in neural nets. When the number of parameters is small, such as in the case of parameterizing concentration-dependent functions, the benefit of using FSA to obtain an estimate of the Hessian outweighs its computational cost, as the convergence is much faster. The trust-region algorithm is used throughout.

\[
\frac{d\mathbf{p}}{d\mathbf{p}} = -\Sigma \frac{\partial S}{\partial \mathbf{p}} d\tau + \sqrt{2\Sigma} d\mathbf{W},
\]

whose stationary distribution is the posterior distribution (Eq. [7]), and \( \mathbf{W} \) is random noise \( \langle \mathbf{W}(\tau_i) \mathbf{W}(\tau_j) \rangle = \delta(\tau_i - \tau_j) I \). As an alternative, we adopt the proposal distribution,

\[
\frac{d\mathbf{p}}{d\mathbf{p}} = -\Sigma^{-1} d\tau + \sqrt{2\Sigma} d\mathbf{W}
\]

based on gradient descent from the prior distribution only to avoid computing the model sensitivity at every step. This expression requires \( \Sigma^{-1} \) to be nonsingular. In practice, for a degenerate distribution such as Eq. [10] we set the variance of the degenerate component (the coefficient of the constant basis function \( R_0(c) \)) to be sufficiently large.

## III. APPLICATIONS

### A. Optimization and uncertainty quantification

We generate simulated data for the evolution of the Cahn-Hilliard equation with the boundary conditions of zero normal gradient for both and zero flux for the former [45]. Fig. 1 shows the evolution of the objective function in finding the MAP in different realizations of spinodal decomposition snapshots, with the same image resolution and physical parameters, while each starts from a different initial condition. These results demonstrate the robustness of the optimization algorithm. The optimizer converges to the truth within tens of iterations robustly with 5, or as few as 2 snapshots.

10 parameters are used for both \( \mu_0 \) and \( \ln D(c) \). The initial guess for \( \mu_0 \) and \( \ln D(c) \) is chosen such that it is far from the truth and generates a non-pattern-forming evolution \( (\mu_0(c) = \ln \frac{D}{D_{\text{true}}} + (1 - 2c) \text{ and } D(c)/L^2 = 0.1 \) \), where \( L \) is the domain size (see inset at iteration 1 in Fig. [1] these initial guesses are used throughout the text unless otherwise noted). We recommend choosing a small \( D(c) \) as the initial guess to “freeze” the pattern. If the pattern relaxes too fast to extend beyond the first frame, the sensitivity of the pattern with respect to all model parameters approaches zero, which is known as the vanishing gradient problem in machine learning. The initial guess for the gradient penalty is \( \kappa/\kappa_{\text{true}} = 0.05 \). The length scale of the diffuse interface is \( \kappa/\kappa_{\text{true}} = 0.045L \).

To find the MAP, \( L_2 \) norm is used as the regularizer for \( \mu_0(c) \) and \( L \) norm is used for \( \ln D(c) \), and regularization parameters for both are \( 10^{-5} \) (this set of parameters is used throughout unless otherwise noted). Due to the structure of Eq. [1], \( \mu_0(c) \) and \( D(c) \) can only be determined up to a constant scale; therefore, throughout the text, we report \( \mu_0(c)/\kappa_{\text{true}}/\kappa \) and \( D(c)/\kappa/\kappa_{\text{true}} \). In fact, the optimizer often converges to a different \( \kappa \), while the scaled quantities above are accurately recovered. Allowing \( \kappa \) to vary in the optimization generally speeds up the convergence.

The optimization allows us to quickly find the MAP, which serves as the starting point for MCMC. For the uncertainty analysis, we assume the measurement is a result and the performance of the MCMC algorithm. The shaded area corresponds to the confidence interval, while
FIG. 1. Training on 11 realizations of spinodal decomposition snapshots. Insets show the initial guess for \( \mu_h(c) \) and \( D(c) \), a typical state during the training, and its final convergence to the truth. Blue and orange curves in the main plot correspond to taking 5 and 2 snapshots as the training datasets, respectively, as highlighted by the outline of an example set of images on the right. The images are concentration fields \( c(z) \) with black and white corresponding to \( c = 1 \) and \( c = 0 \), respectively. The same colormap is used throughout this paper.

The solid line is the marginal mean at each \( c \). The measurement noise \( \sigma \) may be obtained from the knowledge of the instrument or inferred together with other parameters. From here on, unless otherwise noted, we assume \( \sigma^2 = 10^{-4} \) to illustrate the sensitivity more clearly. The Markov chain achieves good mixing within less than 100 steps for all parameters (20 in total) as shown by the autocorrelation of the chain in Fig. 2. Fig. 2 also shows the trajectory of one of the parameters, which is indicative of good sampling. \( 2 \times 10^4 \) samples are used unless otherwise noted.

If the system is non-phase-separating, \( \kappa = 0 \), Eq. [1] becomes

\[
\frac{\partial c}{\partial t} = \nabla \cdot (D(c)c \cdot \mu_h'(c) \nabla c).
\]  

(21)

Therefore, only the chemical diffusivity \( D_{chem} = cD\mu_h'(c) \) can be inferred and no thermodynamic information can be obtained from only the concentration field. However, the extraction of \( \mu_h(c) \) and \( D(c) \) is possible in phase-separating systems, and are mostly uncorrelated, as shown by the correlation coefficients between the parameters \( a_1, a_2, \ldots, a_{10} \) and \( b_0, b_1, \ldots, b_9 \) in Fig. S11. The correlation among odd/even polynomials of low order (of the same function) is strong, while the correlation between \( \mu_h(c) \) and \( D(c) \) is weak. The correlation between the thermodynamic and transport properties can also be quantified through certain scalars of interest. At \( \sigma^2 = 10^{-5}, -\mu_h'(c_0) \) and \( D(c_0) \) are weakly and negatively correlated (Fig. S12). As explained in further detail in Section III C, \( D(c_0)/[\mu_h'(c_0)]^2 \) determines the initial rate of spinodal decomposition, but information from the entire field decouples \( \mu_h(c) \) and \( D(c) \).

Another quantity of interest is the interfacial tension, which is defined to the energy of a flat interface at equilibrium \( \gamma = \int_{-\infty}^{\infty} \left[ g_h(c) + \frac{1}{2} \kappa (c_d)^2 \right] \). Using the equilibrium condition \( \mu(x) = \mu_0 \) where \( \mu_0 \) is the chemical potential of the two equilibrium phases, we have

\[
\gamma = \int_{c_t}^{c_0} \sqrt{2\kappa \Delta g(c)} dc
\]  

(22)

where \( \Delta g(c) = g_h(c) - g_h(c_1) - (c - c_1)\mu_0 \). At the late stage of coarsening, the domain growth rate is proportional to \( \gamma D(c) \). However, we find that \( \gamma \) and \( D(c_0) \) (or \( D(c_1) \)) are almost uncorrelated (Fig. S12), due to information at the diffuse interface and early-stage patterns.

For the phase field crystal models (Eq. [5]), previous studies have found correlation functions that generate certain crystal structures and elastic constants [58, 59]. In Fig. 2 we use MCMC to determine the uncertainty in the direct correlation function inferred from a set of images of nucleation that forms hexagonal crystal structure \( \sigma^2 = 10^{-4} \). Using the parameterization given by Eq. [11] and the \( L_2 \) norm, we find the extent that the function is allowed to vary while keeping the same pattern is very small around wavenumbers that correspond to the spatial wavelength \( k_0 \) observed in the image. Uncertainty increases for wavenumbers as moving away from \( k_0 \) due to the insensitivity of the model to these components and the lack of Fourier components at high wavenumbers in the images provided.

### B. Starting composition

Images taken at different operating conditions can reveal information with different levels of confidence. Patterns of spinodal decomposition at different average concentration range from tortuous regions near \( c_0 = 0.5 \) to dispersed droplets near \( c_0 = 0.3 \) and \( c_0 = 0.7 \). First, an optimization is performed to find the MAP, starting from the same initial guesses as Fig. 1 convergence to the truth is achieved within 20 iterations. Then MCMC is performed to study the uncertainty in the \( \mu_h(c) \) and \( D(c) \). Histograms of the images show a peak near the average concentration (also the initial condition) as strong as the peak at the miscibility gap near 0 and 1. With sufficient pixel values around \( c_0 \) and their dynamics driven by \( \mu_h'(c_0) \) and \( D(c_0) \) at the average concentration \( \mu_h(c_1) \) is fixed to be 0), the uncertainty consistently reaches a minimum for both constitutive relations and all concentrations studied. Overall, images of \( c = 0.5 \) reduce the
C. Model selection: Diffusion and reaction

When the governing dynamics of the system is unknown, or it is unknown whether the order parameter is locally conserved, we consider the possibility of both conserved and nonconserved dynamics, by generalizing Eqs. [1] and [2] and allow the magnitude of both dynamics to be inferred from the patterns,

$$\frac{\partial c}{\partial t} = (\nabla \cdot D(c)\nabla \mu) + R_0(c)(\mu_{\text{res}} - \mu), \quad (23)$$

where $\mu_{\text{res}}$ varies in time such that the average concentration trajectory follows the images, which is constant in this case. Reaction-diffusion models have been studied extensively in literature [13, 60, 61]. Here we focus on a one-component system where the reaction takes place between the system and reservoir. Lithium in lithium iron phosphate (LFP) platelet particles is known to undergo diffusion in the lateral direction via a surface layer while the platelet also exchanges lithium with the electrolyte reservoir [62]. The dynamics of a thin platelet particle may be modeled with Eq. [23] on a 2D plane [48].

The time scale of relaxation can be understood from the dispersion relation of Eq. [23] linearized around a homogeneous state $c(r) = c_0$, obtained by substituting a perturbation $e^{it+ikr}$ into the linearized equation [48, 63, 64], and together with the chemical potential defined by Eq. [3]

$$\omega(k) = -(c_0D(c_0)k^2 + R_0(c_0))(\mu'_h(c_0) - \kappa k^2), \quad (24)$$

where $k = |k|$. When diffusion dominates, $-c_0D(c_0)\mu'_h(c_0) > R_0(c_0)\kappa$, the maximum instability
growth rate is
\[ \omega_{\text{max}} = \max_k \omega(k) = \frac{(c_0 D(c_0) \mu'_h(c_0) - R_0(c_0) \kappa)^2}{4c_0 D(c_0) \kappa}, \] (25)

When there is no reaction, \( \omega_{\text{max}} = c_0 D(c_0)[\mu'_h(c_0)]^2/4\kappa. \) Otherwise, reaction dominates and
\[ \omega_{\text{max}} = -R_0(c_0) \mu'_h(c_0). \] (26)

When diffusion dominates, the maximum growth rate is obtained at a nonzero wavenumber, while when reaction dominates, the maximum growth rate corresponds to \( k = 0. \) Therefore the relaxation patterns are visually different but subtle to distinguish when both effects are important, which motivates a systematic approach of identifying the underlying dynamics via model selection.

We generate the images by varying the magnitude of the diffusivity while keeping the characteristic time scale \( \omega_{\text{max}} \) constant. The iso-\( \omega_{\text{max}} \) curve is shown in Fig. 2. The images used as the training data for each diffusivity value, ranging from diffusion only to reaction only, are taken at the same time and shown in Fig. 4. The chemical potential of the external reservoir adjusts passively while the total concentration is conserved. Training data and results are colored consistently based on its diffusivity in Fig. 4. The initial guess for \( \mu_h(c) \) and \( \kappa \) are the same as section III A The initial guesses for \( R_0(c) \) and \( D(c) \) are 0.1 and 0.01 respectively for all cases. 10 parameters are used each function. The residual plot Fig. 4c confirms that the truth model is found for all cases and that the images are sufficient for identifying the underlying dynamics. When the true \( D(c) \) or \( R_0(c) \) is zero, the solver converges to a vanishingly small value for \( D(c) \) and \( R_0(c) \).

Fig. 4 shows the uncertainty in the inferred functions. The uncertainty in the chemical potential, or free energy, inferred from reaction-controlled patterns are higher than diffusion-controlled patterns, which is also reflected in the uncertainty of the interfacial tension. For the five cases studied \( D(c_0)[\mu'_h(c_0)]^2/4\kappa = 0, 5, 10, 15, 20 \), the mean and 2 standard deviation of \( \gamma/\gamma_{\text{truth}} \) is 0.99 ± 0.12, 1.01 ± 0.079, 1.01 ± 0.066, 1.00 ± 0.057, 1.01 ± 0.059, respectively. The interfacial tension \( \gamma_{\text{truth}}/\sqrt{2\kappa} \) is 0.2. At the late stage of coarsening, the interface growth rate is proportional to the local curvature and independent of the interfacial tension (see Section 2.3 in Ref. [9]), which explains the increasing uncertainty in the free energy when reaction dominates. Early-stage snapshots within the miscibility gap are essential for inferring the free energy of an Allen-Cahn system.

When reaction dominates, the diffusivity becomes increasingly uncertain, and when the true diffusivity is zero, the inferred diffusivity can be anything below a threshold. Recall that the constant term in \( \ln D(c) \) has a degenerate prior; therefore the upper bound of \( D(c) \) is determined by the likelihood only. The same is true for reaction kinetics – when diffusion dominates, the inferred reaction kinetics becomes uncertain. While the patterns are not sensitive to the exact form of \( D(c) \) and \( R_0(c) \) when either reaction or diffusion dominates, their magnitudes can be identified relatively accurately. Regardless of the magnitude of diffusivity, the uncertainty in \( R_0(c) \) becomes increasingly large when \( c \) approaches 0 or 1. This effect can be understood from an analysis of the sensitivity of \( c(x) \) with respect to \( R_0(c) \). In the sensitivity equation Eq. 17 when close to the miscibility gap, \( \frac{\partial}{\partial c} \left( \frac{\delta c}{\delta R_0} \right) < 0 \) due to thermodynamic stability, the sensitivity of the reaction rate \( \delta R_0(c) \cdot (\mu_{\text{res}} - \mu(c)) \) and hence the sensitivity of \( c(x) \) becomes increasingly small as \( \mu \to \mu_{\text{res}} \). Similar to Section III A we observe a weakly negative correlation between \( -\mu'_h(c_0) \) and \( D(c_0) \) (or \( R_0(c_0) \)), which can be understood from their negative correlation when \( \omega_{\text{max}} \) is constant. The correlation between \( D(c_0) \) and \( R_0(c_0) \) is weak (see Fig. S13). The correlation among the three is at its maximum when \( -c_0 D(c_0) \mu'_h(c_0) = R_0(c_0) \kappa \), that is, the critical point where the system transitions from being reaction-dominated to diffusion-dominated.

D. Chemically driven system: the effect of autocatalysis

We study a system that is chemically driven by an external chemical reservoir at a constant average reaction rate, the concentration evolution follows [18]
\[ \frac{\partial c}{\partial t} = R_0(c)f(\mu_{\text{res}} - \mu). \] (27)

When driven far from equilibrium, \( f \) is no longer necessarily a linear function. Here we use \( f(x) = 2 \sinh(x/2) \), which is known as symmetric Butler-Volmer kinetics in electrochemistry [65]. The external chemical potential \( \mu_{\text{res}} \) varies in time, subject to the constraint of the average reaction rate \( R \),
\[ \int \frac{\partial c}{\partial t} dV = R \int dV. \] (28)

The linear stability and pattern formation of such a chemically driven system have been studied extensively [45, 47, 69, 68]. The dispersion relation of Eq. 27 is \( \omega = R_0 f - R_0 f' \cdot (\mu'_h - \kappa k^2) \). Depending on the magnitude and direction of the reaction rate \( R \), the state-dependent \( R_0(c) \) can alter the linear stability (and hence the pattern) to deviate from its thermodynamic stability as determined by \( \mu'_h(c) \).

In Fig. 3, we present an example where \( R_0(c) \) is asymmetric about \( c = 0.5 \). In the range of \( c \) where \( R'_0(c) < 0 \), the pattern is stabilized (destabilized) when \( f > 0 \) (\( f < 0 \)). With \( R_0(c) \) skewed to the left, the pattern under a positive reaction rate \( R \) is more homogeneous than a negative one of the same magnitude. We start from a random initial condition and consider two sets of snapshots where \( R = 0.08 \) and \( -1 \), respectively.

In Fig. 5, we compare the inversion results based on training data from a single or both directions. We find
that the systematic error between MAP and the truth is large when only images reacting in a single direction are used as the training data, even though the objective function $S$ is sufficiently small (root-mean-squared error of frames 2–5 is less than 0.5%). MAP becomes almost identical to the truth when both directions are used, and the strongly concentration-dependent reaction kinetics $R_0(c)$ can be captured. This effect is also reflected in the marginal mean of $\mu_h(c)$ and $R_0(c)$ from the MCMC result. The uncertainty of both functions are significantly reduced when both datasets are used, highlighting the necessity of datasets at different operating conditions in order to infer both thermodynamic and kinetic properties for a chemically driven system whose only observed information is the concentration field (assuming $\mu_{res}$ is unknown).

Fig. 4 shows the scatter plots of $\ln R_0(c_0)'$ and $-\mu_h'(c_0)$ ($c_0 = 0.5$) from the MCMC sample. Using the dataset with $R > 0$ ($R < 0$), the two quantities are negatively (positively) correlated. When both datasets are used, their correlation is reduced. The correlation can be understood from the dispersion relation mentioned above. Linearizing Eq. (27) around a uniform field $c(x) = c_0$, we obtain 

$$
\omega(k) = \left[ R_0' f - R_0 f' \cdot (\mu_h' + \kappa k^2) \right] \bigg|_{c=c_0} = R \left( \frac{d \ln R_0}{dc} - \frac{d \ln f}{dc} \cdot (\mu_h' + \kappa k^2) \right) \bigg|_{c=c_0} \tag{29}
$$

That is, if $c(x, t = 0) = c_0 + \nu e^{ikx}$, where $\nu \ll 1$ is a small perturbation, then $c(x, t) = c_0 + R t + \nu e^{\omega(k)t + ikx}$. Note that $R \rightarrow R_0(c) f(\mu_{res} - \mu_h(c_0))$ as $\nu \to 0$. Therefore, in the limit of $t \to 0$ and $\nu \to 0$, when multiple or a range of wavenumber $k$ exists, only $s/R = (\ln R_0)' - (\ln f)' \mu_h$ and $(\ln f)'$ at $c_0$ can be inferred from the pattern. Note that $s$, also known as the autocatalytic rate [64], is the key to determining the heterogeneity of the pattern: when $s > 0$, the pattern becomes linearly unstable and vice versa. Using Butler-Volmer kinetics,

$$
\frac{d \ln f}{dc} = \frac{1}{f} \sqrt{\left( \frac{f}{2} \right)^2 + 1} = \frac{1}{R} \sqrt{\left( \frac{R}{2} \right)^2 + R_0^2}. \tag{30}
$$

Therefore, when $R > 0$ ($R < 0$), $(\ln R_0)'$ and $-\mu_h'$ are negatively (positively) correlated. To further demonstrate the correlation between the two important quantities, we perform MCMC using two datasets, each of which contains two snapshots with an average concentration of 0.6 and 0.8, with $R = 0.2$ and $R = -0.2$. In Fig. 6 with very small observation noise $\sigma^2 = 10^{-8}$ and $10^{-10}$, we show that $\ln R_0(c_0)'$ and $-\mu_h'(c_0)$ ($c_0 = 0.7$) are strongly correlated with only one dataset used and much weaker correlation is observed when both are used. The solid lines are $(\ln R_0)' - (\ln f)' \mu_h = \text{const}$, where the constant and $(\ln f)'$ are determined by the known truth.

FIG. 4. Dynamics of spinodal decomposition driven by varying degrees of reaction and diffusivity. Five cases are colored consistently in all plots. (a) The training datasets and their corresponding physical time, scaled reaction kinetic prefactor and diffusivity. (b) The relationship between scaled reaction kinetic prefactor and diffusivity with constant instability growth rate. (c) The residual plot during the training process for all cases. (d) The uncertainty in chemical potential, free energy, scaled diffusivity and reaction kinetic prefactor for all cases. The scaling constants $\mu_h'(c_0)$ and $\kappa$ are the known truth.
FIG. 5. Learning of reaction kinetics and free energy from the spatial patterns of a system chemically driven at a constant average reaction rate. (a,b) Training data where spatial patterns of a system chemically driven at a constant exterior chemical potential, its growth rate is proportional to its difference with the exterior chemical potential and the detailed functional form of $\mu_h(c)$ is not important. Therefore, images of the dynamics within the miscibility gap are critical in measuring the free energy of a phase-separating system.

truth. The agreement of the uncertainty quantification with the analytical analysis demonstrates that reaction kinetics and thermodynamics are strongly coupled and only autocatalytic rate can be determined when only a dataset in a certain reaction direction is used. To infer both quantities separately, datasets in both directions are essential. Fig. S14 confirms that $R_0$ has a small posterior variance and is not correlated with $\mu'_h$, since $(\ln f)'$ can be determined independently.

E. Temporal resolution

The availability of snapshots over the course of spinodal decomposition and coarsening determines the uncertainty in the inferred parameters. Fig. 6 shows that with increasing number of snapshots (2, 3, and 5), the uncertainty decreases. These snapshots are evenly spaced in terms of the $L_2$ norm of the difference from the first snapshot. In Fig. 7 we define the distance between patterns to be $|\Delta| = \int (c(t, r) - c(t = 0, r))^2 dr$.

At the late stage of coarsening, most pixels in the image are found to be near the miscibility gap $c = c_1$ and $c_2$ (see histograms in Fig. 7), and the coarsening rate is most determined by diffusivity near $c_1$ and $c_2$. Whenever a late-stage snapshot is provided, we observe a local minimum in the uncertainty of $D(c)$ at $c_1$ and $c_2$, as highlighted by the vertical dashed lines. If only the early spinodal stage images are provided, the uncertainty for both $\mu_h(c)$ and $D(c)$ away from the initial concentration $c_0 = 0.5$ increases. Without the early-stage snapshots, the uncertainty of the chemical potential within the miscibility gap is high. While the value of $\mu_h(c)$ at each $c$ may be uncertain, the mean and 2 standard deviation of $\gamma/\gamma_{\text{true}}$ is $1.02 \pm 0.11$, $1.01 \pm 0.065$, $1.00 \pm 0.056$, $1.02 \pm 0.09$, $1.03 \pm 0.10$, respectively, which indicates that two late-stage snapshots are as useful as two early-stage snapshots in providing information on the interfacial tension.

In fact, it is known in phase field theory that the chemical potential of a sphere of radius $R$ is $\gamma/R$ (plus some constant), its growth rate is proportional to its difference with the exterior chemical potential and the detailed functional form of $\mu_h(c)$ is not important. Therefore, images of the dynamics within the miscibility gap are critical in measuring the free energy of a phase-separating system.
FIG. 7. The uncertainty in the chemical potential and diffusivity given snapshots taken at different time over the course of spinodal decomposition. The left columns show the $L_2$ norm of the difference between a pattern at time $t$ and that at $t = 0 (|\Delta|)$. The snapshots included in each row are highlighted as dots in the $|\Delta| - t$ plots whose color corresponds to that of the image outline in the far right column. From the top row to the bottom, 2, 3, and 5 snapshots equally spaced in $|\Delta|$, 2 early-stage snapshots, and 2 late-stage snapshots are taken as the training data, respectively. The insets show the histograms of combined training data of each case.

F. Spatial resolution

To study the effect of spatial resolution, images are taken from a subset of the field and down-sampled on a rectangular grid. The PDE is numerically solved on a finer grid to resolve the fine details. The first snapshot is linearly interpolated onto the finer simulation grid to be the initial condition. Since the boundary condition for the subset is unknown, all images are also interpolated in time and used as the boundary condition. Assuming that the noise at each pixel and snapshots are independent, the objective function is defined as

$$S(p) = \frac{1}{2\sigma_p^2} \left[ \sum_{i,j} (c(t_i, r_j; p) - c_{data}(t_i, r_j))^2 \right] + ||p||_L$$

(31)

where the pixel-wise variance is $\sigma_p^2 = 10^{-2}$. The likelihood function is defined similarly. Fig. 8a shows the MAP as well as the uncertainty from MCMC with different resolutions. Note that we place no constraint on $\mu_h(c)$ for finding the MAP but we fix the miscibility gap for MCMC.

Since the initial condition and boundary conditions are inaccurate at low temporal and spatial resolutions, MCMC predicts a systematic error compared to the truth. We also compare the MAP result when we enforce the known boundary condition (zero flux and no surface wetting) in Fig. 8b. In this case, the optimizer fails to get even close to the truth at low resolution. Therefore, when the spatial resolution is low, it is preferable to initialize from the first snapshot and impose boundary conditions from the data itself, despite the fact they are not accurately known. The inaccuracy in the initial condition can be partially compensated by the boundary condition. In fact, comparing the model prediction from the MAP results at different resolutions, the patterns are largely preserved. The optimizer fails to find a reasonable solution at a resolution of $8 \times 8$. Therefore, as a rule of thumb, at least 3 pixels per wavelength of spinodal pattern are needed.

Without constraining the miscibility gap, the optimizer converges to a $\mu_h(c)$ with smaller miscibility gap with decreasing spatial resolution. This occurs because the low sampling rate effectively filters out the high-frequency components, blurring the high contrast between the two phases.

G. Domain size

In cases where the field of view is limited to a subset of the entire domain, the boundary condition is unknown. Similar to Section III F, the concentration and its normal gradient from each snapshots are interpolated in time and used as the boundary conditions. Fig. 9 shows the training result for images of different domain sizes. The discrepancy between the truth and the MAP increases with decreasing domain size, suggesting that with the parameters become less identifiable with less information and local minima of the objective function is likely to be encountered. Only lower order polynomials of $D(c)$ can be inferred when the domain size are too small to contain large concentration variations. However, the computational cost for solving the PDE is greatly reduced if only a smaller subdomain in the training data is used. The regularization parameter is $10^{-5}$ times the domain size.

H. Blurring filter

Imaging systems have a certain point source function (PSF) that may spread over more than a single pixel. The images are a convolution of the object and the PSF. Therefore studying the inversion of these blurred images is important for practical imaging devices. In the example of spinodal decomposition used above, we show that
using blurred images directly will result in systematic error in the inferred chemical potential and diffusivity. However, it is possible to invert the characteristic length scale of the PSF and more accurate physical properties simultaneously, effectively leading to a physics-constrained deconvolution.

In Fig. 10, we generate a sequence of images in (a) and convolve them with Gaussian (b and d) and box-averaging PSF (c). The PSF is indicated by the red region in the upper right corner of the last image in the sequence. The Gaussian PSF is common and can be used to approximate an Airy disk, which is the PSF of a circular aperture. Box-averaging PSF is constant within a compact support. We adopt four inversion strategies, denoted in Fig. 10 as A) assume the images are not blurred; B) assume the images are blurred by a Gaussian PSF and invert its standard deviation $d$ together with $\mu_h(c)$ and $D(c)$; C) the same as B, except that $\mu_h(c)$ is constrained to a fixed miscibility gap, which may be measured more accurately after a long relaxation into two well-separated phases; D) the same as B, except that the first image is deconvolved and then used as the initial condition. In strategies A–C, the first given image (blurred) is used as the initial condition. In all strategies, the regularization parameter is $10^{-2}$, the objective function is a discrete summation of the squared error on five $50 \times 50$ images. The initial guess is $d/L = 0.1$, $\alpha = 0.1$ (noise-to-signal ratio for deconvolution, see below), $D(c) = 0.1$.

Strategy A (treating blurred images as the truth) underestimates the interfacial tension and the miscibility gap in all cases. The objective function value that the optimizer converged to is also significantly higher than other strategies. This effect becomes more severe when the length scale of the PSF increases and becomes comparable to the spinodal length scale, $l = \sqrt{2\kappa/\mu_h'(c_0)}$. The standard deviation of the Gaussian PSF is $d = 0.03L = 0.67l$ and $d = 0.06L = 1.34l$ respectively for Figs. 10(b) and 10(d), where $L$ is the image size. Hence $[-3d, 3d]$ extends to about one wavelength of the spinodal pattern $2\pi l$. The side length of the averaging box in Fig. 10(c) is $0.1L$.

When the PDE solution is convolved with a Gaussian PSF and $d$ is also optimized (strategy B), the inferred $\mu_h(c)$ and $D(c)$ are closer to the truth. For the sets of images blurred by a Gaussian PSF, the correct $d$ is found, while the solver converges to $d = 0.03$ for Fig. 10(c). For images whose unknown PSF is isotropic and decays to zero, the Gaussian PSF is often a good estimate. Here,
nonzero Fourier components of the images are concentrated in a narrow band around $l^{-1}$; blurring is only sensitive to the characteristic length scale of the PSF and not its details. Fig. 10 shows that an approximation with a Gaussian PSF is sufficient. In addition, Gaussian PSF is differentiable with respect to its parameter $d$, hence useful for gradient-based optimizers. Given the physical constraint of miscibility gap, strategy C shows that the objective function is further decreased.

Strategy D deconvolves the first image to use as the initial condition. To reduce the ringing effect of deconvolution due to discontinuity at the boundary, we perform a linear interpolation between the given image and the image convolved with the PSF, weighing the given image more in the interior and the convolved more in the boundary. We use $\left(1 - e^{-x^2/4d^2}\right)\left(1 - e^{-y^2/4d^2}\right)$, shifted so that $[x, y] = [0, 0]$ corresponds to the corner. We use Wiener deconvolution which in the Fourier space is $K^*(K^*K + \alpha)^{-1}$, where $K$ is the PSF in Fourier space, $\alpha$ is the noise-to-signal ratio and included as a variable to be optimized. The residual plots and plots of $\mu_h(c)$ and $D(c)$ show that this strategy is close to or sometimes worse than strategies B and C which do not deconvolve the first image, which shows that, while the initial condition can be polluted by noise and blurring, preprocessing may be unnecessary. In fact, the unknown boundary condition may introduce additional error in the process of deconvolution. Inversion of a PDE whose solution diverges from the perturbation of an initial condition is beyond the scope of this work.

The uncertainty using strategies B and C based on images in Fig. 10; and $\sigma_p^2 = 0.1$ is summarized in Fig. 10. The scatter plot shows a strong and negative correlation between the inferred Gaussian filter length $d$ and the width of the miscibility gap when $\mu_h(c)$ does not have any constraint, further demonstrating the importance of imposing the physical constraint. When such a constraint is imposed, the standard deviation of $d/d_0$ is reduced from 0.07 to 0.035, where $d_0$ is the truth. The miscibility gap $[c_1, c_2]$ is defined by Eq. 12 as well as $\mu_h'(c_1) > 0, \mu_h'(c_2) > 0$ and $c_1 \neq c_2$.

**IV. CONCLUSION AND OUTLOOK**

Using the approach of PDE-constrained optimization and Bayesian inference, we performed a systematic analysis of the inversion and the uncertainty quantification of the constitutive relations based on images of pattern formation. We also used analytical methods including scaling analysis and linearization to explain the correlation observed among these constitutive relations. Many of the complications studied here may arise in experimental settings, such as unknown boundary condition, inaccurate initial condition, limited spatial and temporal resolution and image blurring. We show that using the given image data as the initial condition and boundary condition of the model is usually sufficient and useful information can be extracted even under very limited resolution and blurring. Multiple datasets such as reaction at different rates and different starting compositions can enhance the accuracy of the inferred constitutive relations. Using prior knowledge such as the miscibility gap and other regularizable governing equation. When the imaging quality is limited by the instrument, such a PDE-constrained inversion effectively becomes a physics-informed superresolution imaging technique.

The methods and applications discussed here serve as a first step toward quantitative frame-by-frame and pixel-by-pixel matching between experiments and theoretical models, as excellent agreement has already been observed in many complex systems. Despite the lack of microscopic information, the macroscopic pattern formation dynamics can be described parsimoniously by
PDEs with relatively few parameters. The inversion of the PDE means that, from the images, these macroscopic physical properties can be measured that would otherwise be inaccessible especially for systems far from equilibrium such as active matter and biology, where free energy is ill-defined, and nonequilibrium thermodynamics is poorly understood. The uncertainty quantification can be applied to optimal design of experiments to carefully probe regions of higher uncertainty as informed by prior experimental data.

The inversion also enables a physically interpretable parametrization of complex systems, which may help in establishing a mapping between the macroscopic and microscopic parameters, and eventually engineering or controlling patterns by tuning physical properties of the constituents, as reported recently in the biological engineering of Turing patterns [74] and the design of PDEs to create desired patterns [75]. Our computational approach can be integrated into the loop to accelerate the search in a high dimensional parameter space by identifying the most important engineering handles.

While phase field (Cahn-Hilliard and Allen-Cahn models) and phase field crystal models were selected as the model systems in this article, the approach can be readily extended to other systems, such as for fluid dynamics and multi-component reaction-diffusion equations, where more complicated patterns may arise [1, 3, 76], and the sensitivity of patterns and bifurcation dynamics with respect to constitutive relations in a high-dimensional parameter space awaits exploration. For complex systems, further study is needed to quantify the range of phenomenon that a model can describe. Techniques in inverse problems, dynamical systems, and identifiability analysis should be employed when discrepancy between experiments and models arise due to nonidealties such as spatial heterogeneity as well as unknown hidden variables.

Appendix A: Correlation between thermodynamic and kinetic properties

Fig. S11 shows the matrix of correlation coefficients between parameters for \( \mu(\alpha_1, \alpha_2, \ldots, \alpha_{10}) \) and \( D(c) \) \((b_0, b_1, \ldots, b_6)\) inferred from spinodal decomposition snapshots (same as used by Fig. 2) and \( \sigma^2 = 10^{-4} \). The correlation coefficient between \( \alpha_i \) and \( \alpha_j \) is 
\[
\text{Cov}(\alpha_i, \alpha_j)/\sqrt{\text{Var}(\alpha_i)\text{Var}(\alpha_j)}
\]
where Cov and Var stand for the covariance and variance, respectively.

Using the same set of images, Fig. S12 shows the scatter plot of \(-\mu'(c_0)\) versus \(D(c_0)\) from the MCMC chain, as well as \(\gamma/\sqrt{2\kappa}\) versus \(D(c_0)\), where \(c_0 = 0.5\) is the average fraction. \(\rho\) is the correlation coefficient between the two parameters as defined above. The initial rate of decomposition is \(c_0 D(c_0)[\mu'(c_0)]^2/4\kappa\), which explains the negative correlation between \(-\mu'(c_0)\) and \(D(c_0)\). The contours are the 90% confidence region.

FIG. S11. Correlation coefficients between parameters for \( \mu(c) \) and \( D(c) \).

FIG. S12. Scatter plots and 90% confidence region of \(-\mu'(c_0)\) versus \(D(c_0)\) and \(\gamma/\sqrt{2\kappa}\) versus \(D(c_0)\) (20,000 samples). The colors of the contours match those of the scatter plots.

In a system where both reaction and diffusion are present, we study the pairwise correlation among thermodynamics, reaction kinetics, and diffusivity by showing the scatter plots of \(-\mu'(c_0)\), \(D(c_0)\), and \(R_0(c_0)\) in Fig. S13. The samples are drawn based on images in Fig. 2.

Using the chemically driven concentration fields studied in Fig. 6ab, Fig. S14 shows the scatter plots of \(R_0(c_0)\) and \(\mu'(c_0)\) where \(c_0 = 0.7\) and confirms that they are uncorrelated and \(R_0(c_0)\) can be determined with high accuracy.

ACKNOWLEDGMENTS

This work was supported by Toyota Research Institute through the D3BATT Center on Data-Driven-Design of Rechargeable Batteries.
FIG. S13. Scatter plots of $-\mu'_h(c_0)$, $D(c_0)$, and $R_0(c_0)$ (20,000 samples). Each row has a color that corresponds to the set of images having the same color in Fig. 4a.

FIG. S14. Scatter plots of $R_0(c_0)$ versus $\mu'_h(c_0)$ using datasets in Fig. 6 (20,000 samples), $c_0 = 0.7$. 
Bartelt, J. D. Sugar, K. R. Fenton, D. A. Cogswell, A. L. D. Kilcoyne, T. Tyliszczak, M. Z. Bazant, and W. C. Chueh, Nat. Mater. 13, 1149 (2014).

[63] P. Bai, D. A. Cogswell, and M. Z. Bazant, Nano Lett. 11, 4890 (2011).

[64] M. Z. Bazant, Faraday Discuss. 199, 423 (2017).

[65] J. Newman and K. E. Thomas-Alyea, Electrochemical systems (John Wiley & Sons, 2012).

[66] D. A. Cogswell and M. Z. Bazant, ACS Nano 6, 2215 (2012).

[67] H. Zhao and M. Z. Bazant, Phys. Rev. E 100, 012144 (2019).

[68] D. Fraggedakis and M. Z. Bazant, J. Chem. Phys. 152, 184703 (2020).

[69] J. Pathak, B. Hunt, M. Girvan, Z. Lu, and E. Ott, Phys. Rev. Lett. 120, 24102 (2018).

[70] J. Berry, S. C. Weber, N. Vaidya, M. Haataja, C. P. Brangwynne, and D. A. Weitz, Proc. Natl. Acad. Sci. U. S. A. 112, E5237 (2015).

[71] H. H. Wensink, J. Dunkel, S. Heidenreich, K. Drescher, R. E. Goldstein, H. Löwen, and J. M. Yeomans, Proc. Natl. Acad. Sci. U. S. A. 109, 14308 (2012).

[72] J. Dunkel, S. Heidenreich, K. Drescher, H. H. Wensink, M. Bär, and R. E. Goldstein, Phys. Rev. Lett. 110, 1 (2013).

[73] G. Liu, A. Patch, F. Bahar, D. Yllanes, R. D. Welch, M. C. Marchetti, S. Thutupalli, and J. W. Shaevitz, Phys. Rev. Lett. 122, 248102 (2019).

[74] M. Kryuchkov, O. Bilousov, J. Lehmann, M. Fiebig, and V. L. Katanaev, Nature 585, 383 (2020).

[75] N. Yoshinaga and S. Tokuda, (2020).

[76] S. Chandrasekhar, Hydrodynamic and Hydromagnetic Stability (Courier Corporation, 1981).