Learning Activator-Inhibitor Dynamics at the Cell Cortex with Neural Likelihood Ratio Estimation

Ondrej Maxian [1,*] Edwin Munro,[2,3] and Aaron R. Dinner [3,4,5]

[1] *Department of Applied and Computational Mathematics and Statistics, University of Notre Dame, Notre Dame, Indiana 46556, USA*

[2] *Department of Molecular Genetics and Cell Biology, University of Chicago, Chicago, Illinois 60637, USA*

[3] *Institute for Biophysical Dynamics, University of Chicago, Chicago, Illinois 60637, USA*

[4] *Department of Chemistry, University of Chicago, Chicago, Illinois 60637, USA*

[5] *James Franck Institute, University of Chicago, Chicago, Illinois 60637, USA*

(Received 6 May 2026; accepted 16 September 2026; published 3 November 2026)

A key question in cell biology is how cell-scale organization emerges from a given set of molecular players and rules of interaction. Given its multiscale nature, addressing this question requires a combination of experimental perturbation, mathematical modeling, and parameter inference. We leverage recent advances in each of these fields, focusing, in particular, on neural-network methods for simulation-based inference, to study how cell-scale patterns of Rho guanosine triphosphatase (GTPase) activity are defined by molecular-scale activator-inhibitor interactions with filamentous actin. Using an existing model of this interaction, we demonstrate that an overexpressive, regularized classification neural network can approximate the likelihood of data arising from a particular parameter set. We show that variations in F-actin assembly dynamics can be inferred directly from experimental data, but only if the network is made less sensitive to model misspecification. We use our approach to interpret perturbation experiments in which increasing Rho GTPase-activating protein (RhoGAP) coexpression increases the frequency and coherence of Rho activity waves in frog eggs. After showing that the known functions of RhoGAP are insufficient to explain experimentally observed dynamics, we use neural methods to suggest an alternative pathway by which RhoGAP could decrease filament nucleation rates to sustain waves. Our work yields specific, experimentally testable predictions and illustrates how a combination of traditional forward simulation models and modern inference tools can aid in unraveling mechanisms of self-organization.

DOI: 10.1103/5l5r-fvjd

I. INTRODUCTION

In cell biology, a central question is how cell-scale dynamics emerge from microscopic interactions among constituents [1–3]. Given that experimental systems are complex [4,5], mathematical modeling is often used to correlate differences in molecular parameters with changes in cell-scale patterns [6–10]. As some of the molecular processes driving self-organization can now be directly characterized experimentally [11–15], model parameters can be validated for the first time, making it all the more important to utilize accurate and efficient methods for parameter inference.

Since mathematical modeling involves simplifying choices, there is almost never a one-to-one correspondence between the model and the experimental system at the molecular level. As a result, multiple parameter combinations may fit the data comparably well, and it is useful to quantify how many combinations do so. For this reason, it is more appropriate to infer *distributions* of model parameters, rather than values. Model and experimental outputs are complex, however, and it is usually not possible to write a closed-form mapping from parameters to data. Thus, it is difficult to assign a probability (or "likelihood") of observing the data given a particular parameter set (with a probability assumed "prior" to comparison with the data), or a ("posterior") probability that the data come from a specific set of parameters. The only plausible approach is therefore to use repeated simulations, or "simulation-based" inference (SBI) [16], to compare the model output to the data.

Patterning of Rho guanosine triphosphatases (GTPases) on the cell membrane for cytokinesis, cell migration, and morphogenesis serves as a paradigm of self-organization in development [17–22]. One stunning example is the interaction between active membrane-bound RhoA (hereon Rho) and F-actin/RhoGAP (Rho GTPase-activating protein) at the cell cortex [Fig. 1(a)], in which active Rho autocatalyzes its own activation through the Rho guanine nucleotide exchange factor (RhoGEF) Ect2, while simultaneously initiating the assembly of actin filaments (F-actin) through downstream effectors; to complete the activator-inhibitor loop, F-actin recruits RhoGAP (RGA-3/4) to inactivate Rho [23,24]. Despite conservation of the key molecular players across cell types [23–29], the dynamics of Rho activity in starfish and frog cells (traveling waves) [24,27] differ sharply from those in *C.*

*Contact author: omaxian@nd.edu

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

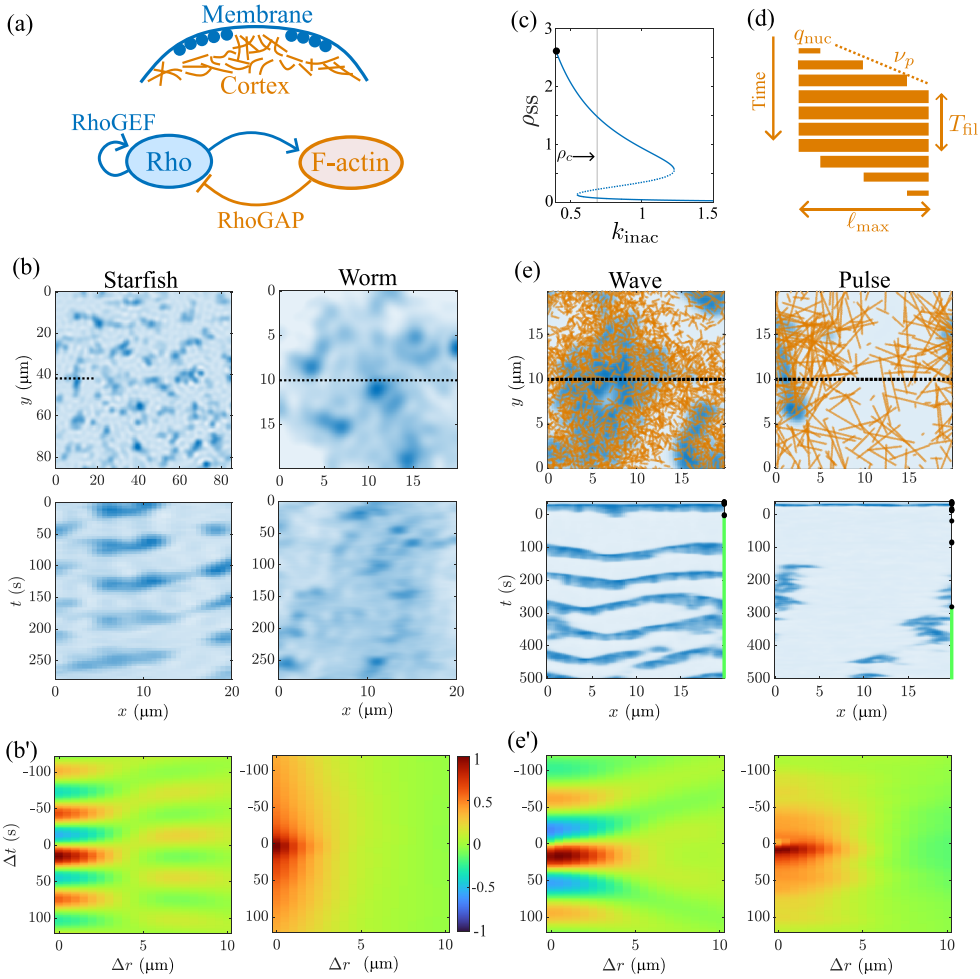


FIG. 1. Rho and F-actin form a versatile activator-inhibitor system at the cell cortex, which can be understood using cross-correlation and computational modeling. (a) System schematic. Active membrane-bound Rho signals the assembly of cortical actin filaments, which recruit RhoGAP to inhibit Rho. (b) The dynamics of Rho in starfish [Fig. 1(c) in Ref. [24]] and worm (*C. elegans*) [26] embryos. Top plots show stills; bottom plots show kymographs along a 20 μm width (the dashed lines). By slight abuse of notation, x refers to the x coordinate in these plots, and is distinct from $\boldsymbol{x}$ that refers to data (usually cross correlation) throughout the paper. These data are preprocessed to accentuate the Rho signal (see Appendix). (b′) Corresponding cross-correlation functions, computed on the filtered data using Eq. (1) (color bar applies to all cross-correlation functions throughout this paper). (c)–(e) Hybrid model of Rho/F-actin dynamics. (c) Rho is represented as a continuum according to Eq. (A1a). This diagram shows the steady states in the absence of diffusion with all parameters fixed except $k_{\text{inac}} = k_{\text{inac}}^{(0)} + k_{\text{GAP}} f$. By default, $k_{\text{inac}}^{(0)} = 0.7$ so that the Rho dynamics lie in a bistable regime. (d) F-actin is represented using discrete filaments, which nucleate with rate q_{nuc}, assemble with speed v_p to length $\ell_{\max}$, remain fixed for time T_{fil}, and then disassemble with rate $v_d = v_p$. These filaments are convolved with a Gaussian to define a continuum field (Fig. S1 of the Supplemental Material [63]). (e) Sample snapshots and kymographs of simulations representing waves and pulses. If all excitations extinguish, we reinitiate (force) them by switching from $k_{\text{inac}}^{(0)} = 0.7$ to $k_{\text{inac}}^{(0)} = 0.4$ [see panel (c) and Appendix 2 c]; switching points are marked by black symbols. (e′) Simulated cross-correlation functions, computed over the longest simulation period without forcing [green lines in kymographs in panel (e)].

elegans (transient pulses) [23,25,26]. As the dynamics in both cases can be manipulated by perturbing certain components of the circuit (e.g., Rho effectors and RhoGAP) [24,27], the broader question becomes how each molecular player shapes the overall patterns of Rho activity *in vivo*.

We recently developed a mathematical model that captures the spectrum of behaviors from traveling waves to transient pulses [30]. Our model represents colocalized F-actin and RhoGAP explicitly in filamentous form, imparting the system with a stochasticity and anisotropy that is key to reproducing the behavior observed in *C. elegans*. To infer parameters, we previously employed approximate Bayesian computation (ABC) to approximate the likelihood function by errors in summary statistics [31–33]. In addition to making sampling specific to the dataset being studied, this approach prevents full characterization of the likelihood function, instead only identifying the parameter sets that best reproduce the data [16,34,35] (see also Sec. IV B). Thus, our previous work was neither sufficiently high throughput to examine multiple sets of perturbation experiments, nor sufficiently accurate to compare multiple hypotheses for how pattern variation could emerge.

Modern machine learning algorithms have the potential to revolutionize SBI [16,36–41], but their rapid development has

made it unclear how best to incorporate them in complex inference problems. Neural techniques for inference vary from unsupervised learning of the posterior (e.g., via normalizing flows) [42–45] to supervised learning of likelihood ratios between parameter sets [36]. The past three years alone have seen the application of neural techniques for SBI to synthetic data in economics and finance [46], astrophysics [47,48], cardiovascular modeling [49], and cell spreading [50]. Extending these approaches to experimental data is more challenging because the data do not directly come from the model (there is some misspecification), which can result in overconfident and incorrect predictions [51–53] (see also Sec. IV C). Applications to experimental data have consequently been relatively limited but do include both static structural data on cell [54–57] and organismal scales [58], as well as dynamic time series data [55,59–61].

Here, we make use of neural SBI to perform high-throughput, accurate inference on our model of Rho GTPase dynamics, making it possible to identify and differentiate hypotheses for how perturbations in molecular players yield cell-scale changes in Rho activity. Our primary innovation is to adapt neural likelihood ratio estimation (NLRE) [34], to biological models where randomness plays a significant role, training a discriminative classifier to learn the mapping from data and parameters to the corresponding likelihood-to-evidence ratio (LER). In the first half of the paper, we validate this approach on both synthetic and previously studied experimental data [30], in the process demonstrating the sensitivity of posterior estimation to classifier regularization and the dimension of the training data. Our validation also shows how NLRE provides complementary information to errors in summary statistics.

In the second half of the paper, we use our approach to understand how RhoGAP expression levels could tune the frequency and coherence of waves that arise in activated frog embryos [24]. Beginning with a model that accounts for the known interactions in this system, we show that increasing GAP activity with other variables fixed does not reproduce the experimental data. Using the robustness of NLRE, we identify two other parameters (filament lifetime and nucleation rate) that could change in response to additional RhoGAP activity. We then use the accuracy of NLRE to infer the more likely of the two, hypothesizing that the amount of F-actin nucleated independently of Rho must decrease as GAP activity increases. This prediction, which does not require any additional simulations once the network is pretrained, is experimentally testable using fluorescence microscopy and particle tracking [14,62], and provides an alternative to previous predictions that utilized more traditional approaches [24].

II. SELF-ORGANIZATION OF RHO AND F-ACTIN AT THE CELL CORTEX

The spatiotemporal dynamics of Rho (and F-actin/RhoGAP) can be visualized by imaging fluorescently tagged biosensors that bind active Rho (F-actin/RhoGAP). Using this strategy, transient cortical waves of Rho activity were first observed in starfish and frog oocytes at the beginning of cytokinesis, just after anaphase onset [27]. Quasisteady versions of these dynamics can be reconstituted in earlier stages of the cell cycle by coexpressing additional copies of RhoGEF (Ect2) and RhoGAP (RGA-3/4) [24,27], as shown in Fig. 1(b).

In *C. elegans*, in contrast, the dynamics of Rho are characterized by transient, meandering pulses of activity [23,25,64]. Previous work showed that these pulses are *not* a result of contractile instabilities [23,25], but rather result from the intrinsic dynamics of the activator-inhibitor circuit of RhoA and F-actin. In fact, the pulsatile dynamics in wild-type and myosin-depleted embryos are indistinguishable [26,30], and so for simplicity we show an embryo depleted of myosin in Fig. 1(b) (see Appendix for preprocessing steps).

A. Cross-correlation function quantifies spatiotemporal activator/inhibitor dynamics

Because the experimental dynamics are stochastic, reproducing individual trajectories of Rho amounts to overfitting noise, and the overall spatiotemporal dynamics are best understood through system-level statistics. We make use of the cross-correlation function between Rho and F-actin, which has been used extensively in previous work [24,27,30] [Fig. 1(b′)]. At time lag Δt and radial distance Δr, the cross-correlation function is defined as

$$R_{\rho f}(\Delta r, \Delta t) = \frac{1}{Z}\sum_{j,k}(\rho(r_j, t_k) - \bar{\rho}) \times (f(r_j + \Delta r, t_k + \Delta t) - \bar{f}), \quad (1)$$

where $\bar{\rho}$ and $\bar{f}$ are the mean amounts of Rho and F-actin, respectively, and the normalizing factor sets the maximum absolute value $|R_{\rho f}| = 1$. For the starfish data, the cross-correlation function gives information about the temporal frequency and spread of waves, as it is oscillatory with a period of roughly 60 s (corresponding to the period observed in kymographs), and $R_{\rho f}(0, \Delta t < 0) < 0$ results when Rho spreads into regions that lack F-actin (making for a negative correlation with F-actin in the past). In *C. elegans*, Rho and F-actin are positively correlated for $\Delta t = 0$ and $\Delta r \leqslant 3$ μm, corresponding to the typical size of a pulse. As we observed previously [30], the cross-correlation in *C. elegans* lacks the signature of traveling waves observed in starfish and is more characteristic of standing patterns (being roughly constant in time without any negative values).

B. Forward model that reproduces variation across cell types

We previously showed that the differences in Rho activity across cell types can be explained by accounting for changes in F-actin assembly dynamics [30]. Specifically, while in the starfish system F-actin is relatively homogeneous, fluorescent movies of F-actin (LifeAct/Utrophin) in *C. elegans* [23,26,65] clearly show a heterogeneous meshwork with visible bundles and holes. Motivated by this, we developed a model that explicitly accounts for the *filamentous* nature of F-actin [Figs. 1(c)–1(e)] [30]. In this two-dimensional model [Figs. 1(c)–1(e)], active Rho evolves according to a reaction-diffusion equation [Eq. (A1a)], where the reaction terms are parametrized to give low and high stable steady states [Fig. 1(c)]. The dynamics of Rho are coupled to a discrete network of actin filaments that nucleate, grow to a fixed length,

stay standing for some time, and then shrink [Fig. 1(d)]. To encode the activator-inhibitor interaction, actin filaments are preferentially nucleated in regions of high Rho activity, and F-actin acts as an inhibitor of Rho, with inhibition strength k_{GAP} modeling the RhoGAP concentration in the system (see Appendix for equations).

To examine how F-actin dynamics dictate Rho activity, we fix the values of all parameters *except* those that determine F-actin assembly and inhibition strength. We reduce F-actin dynamics to one spatial and one temporal parameter by fixing the growth and shrink rates, and scanning over filament length $\ell_{\max}$ and lifetime T_{fil}. Defining parameters f_b and f_ρ as the rates of basal and Rho-mediated filament nucleation (see Appendix) gives a total of five unknowns ($k_{\mathrm{GAP}}, \ell_{\max}, T_{\mathrm{fil}}, f_b, f_\rho$), which set the overall model behavior. As a large amount of actin leads to a complete loss of Rho activity, we constrain the amount of F-actin by computing a binary five-dimensional interpolating function $S(k_{\mathrm{GAP}}, \ell_{\max}, T_{\mathrm{fil}}, f_b, f_\rho)$ that assigns parameter sets that are unable to sustain excitation a prior probability (of observing the parameters) of zero (see Appendix).

III. NEURAL LIKELIHOOD RATIO ESTIMATION

Once the prior and model are defined, the full inference problem comes into focus. Given experimental data $\boldsymbol{x}$, we would like to predict the probability that the data come from the model with parameters $\boldsymbol{\theta} = (k_{\mathrm{GAP}}, \ell_{\max}, T_{\mathrm{fil}}, f_b, f_\rho)$. This is the posterior density $p(\boldsymbol{\theta}|\boldsymbol{x})$, which by Bayes theorem is given by

$$p(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int p(\boldsymbol{x}|\boldsymbol{\theta}')p(\boldsymbol{\theta}')d\boldsymbol{\theta}'} := \frac{p(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{x})} := r(\boldsymbol{x},\boldsymbol{\theta})p(\boldsymbol{\theta}), \tag{2}$$

where $p(\boldsymbol{x}|\boldsymbol{\theta})$ is the likelihood of observing the data from a given set of parameters, $p(\boldsymbol{\theta})$ is the prior probability of observing those parameters, and $p(\boldsymbol{x})$ is known as the evidence. Thus, $r(\boldsymbol{x}|\boldsymbol{\theta})$ is the likelihood-to-evidence ratio. We set $\boldsymbol{x}$ equal to a low-dimensional representation of the cross-correlation function, which is obtained by performing principal component analysis (PCA) on a set of 10 000 cross-correlation functions (see Appendix and Fig. S2 of the Supplemental Material [63]). We use the cross-correlation function to represent the data because it is reproducible across experimental and simulated samples (unlike pointwise trajectories), and we compress it to focus the learning on coarse features rather than fine details (which are less robust across replicates).

In ABC, the posterior density is estimated by drawing a parameter set $\boldsymbol{\theta} \sim p(\boldsymbol{\theta})$ from the prior, running the forward model to obtain a simulated dataset $\boldsymbol{x}_s$, and then estimating $p(\boldsymbol{x}|\boldsymbol{\theta}) \approx \hat{\mathcal{L}}(\boldsymbol{x}_s(\boldsymbol{\theta}), \boldsymbol{x})$ using an approximate likelihood based on summary statistics. In addition to requiring one simulation per sample, this approach does not adapt well to new data, as it requires an additional scan of parameter space for each new dataset. In contrast, in NLRE, the likelihood-to-evidence ratio $r(\boldsymbol{x}|\boldsymbol{\theta})$ is approximated by a classification neural network, which is pretrained to pair parameter sets with data. This requires a high upfront cost for training, but a reduced cost thereafter, since the "amortized" likelihood [39] (neural network) is much cheaper to query than the forward model and can be used for multiple datasets at once without having to resample/retrain.

To perform NLRE, we make use of the so-called "likelihood ratio trick" [34,36] to approximate $r(\boldsymbol{x}|\boldsymbol{\theta})$ in Eq. (2) via a classification neural network. We specifically follow the approach of Hermans *et al.*, who train a classifier on pairs of samples $\boldsymbol{x}$ and parameters $\boldsymbol{\theta}$ by (somewhat unusually) injecting the parameters $\boldsymbol{\theta}$ as features [34]. The job of the classifier is then to distinguish *dependent* sample-parameter pairs $(\boldsymbol{x}, \boldsymbol{\theta}) \sim p(\boldsymbol{x}, \boldsymbol{\theta})$ (labeled by $y = 1$) from *independent* sample-parameter pairs $(\boldsymbol{x}, \boldsymbol{\theta}) \sim p(\boldsymbol{x})p(\boldsymbol{\theta})$ (labeled by $y = 0$); i.e., separate parameters $\boldsymbol{\theta}$ that generate $\boldsymbol{x}$ from those that do not (Fig. 2 and see Appendix for more details on classifier training). Once the classifier is trained, a new dataset $\boldsymbol{x}$ and potential parameter match $\boldsymbol{\theta}$ can be assessed by computing the probability that the data come from that parameter set (equal to the probability $y = 1$). From this, the likelihood-to-evidence ratio can be approximated as

$$\hat{r}(\boldsymbol{x},\boldsymbol{\theta}) := \frac{p(y=1)}{1-p(y=1)} \approx \frac{p(\boldsymbol{x},\boldsymbol{\theta})}{p(\boldsymbol{x})p(\boldsymbol{\theta})} = \frac{p(\boldsymbol{x}|\boldsymbol{\theta})}{p(\boldsymbol{x})} = r(\boldsymbol{x}|\boldsymbol{\theta}). \tag{3}$$

This result is proportional to the posterior density $p(\boldsymbol{\theta}|\boldsymbol{x})$ for parameters with nonzero prior (since the prior is constant where it is nonzero; see Appendix for more details on inference). Throughout the paper, we use a hat to refer to quantities *approximated* by the neural estimator.

IV. RESULTS

Our ultimate goal is to use NLRE to infer the minimal set of parameters responsible for changes in cell-scale Rho activity. Prior to doing this, however, we must be confident that the likelihood ratio estimator $\hat{r}(\boldsymbol{x},\boldsymbol{\theta})$ provides a good approximation to the model's true likelihood function $p(\boldsymbol{x}|\boldsymbol{\theta})$. Thus, we first validate NLRE on synthetic data in Sec. IV A by choosing a parameter set $\boldsymbol{\theta}^*$, generating samples from the true likelihood function $p(\boldsymbol{x}|\boldsymbol{\theta}^*)$, and comparing to the approximate likelihood function $p(\boldsymbol{x})\hat{r}(\boldsymbol{x},\boldsymbol{\theta}^*)$ generated by NLRE [34,36]. In Sec. IV B, we show further that the likelihood ratio estimate tracks with the error in the cross-correlation function but penalizes errors in some PCA modes more than others. In Sec. IV C, we move on to experimental data, finding that additional regularization is essential to prevent overfitting when there is model misspecification. In Sec. IV D, we validate the inferred model parameters (with the additional regularization) against kinetic data from *C. elegans* [11,23,62]. We also show how NLRE can efficiently survey the posteriors for the starfish and worm data and interpolate between them. Finally, in Sec. IV E, we apply our methodology to infer how changing RhoGAP expression alters the frequency and coherence of waves of Rho activity in activated frog oocytes. After demonstrating that existing knowledge of how RhoGAP enters the circuit [Fig. 1(a)] is insufficient to reproduce the data, we show that there must be negative feedback of RhoGAP activity on basal filament nucleation rates to sustain traveling waves.

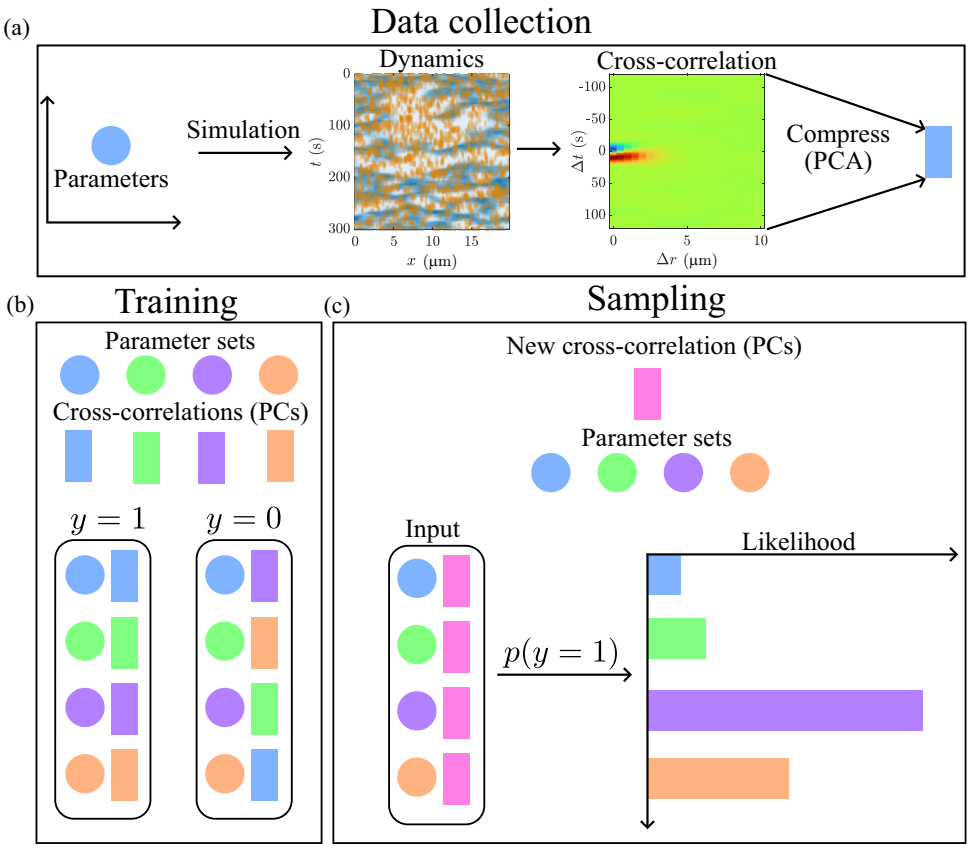


FIG. 2. Deploying NLRE [34] to estimate the likelihood-to-evidence ratio. (a) To construct a dataset, we sample parameter space by running the forward model and computing a compressed version of the cross-correlation function for each parameter set. (b) In the training phase, we train the classifier to distinguish between dependent and independent parameter-sample pairs. (c) In the sampling phase, we input a new sample (e.g., experimental data), and then compute the probability that this sample is associated with each parameter set, from which the likelihood can be computed by Eq. (3). In this schematic, the purple parameter set "most resembles" the pink new sample, and so the classifier assigns it the highest likelihood.

A. Validation on synthetic data reveals importance of classifier architecture and regularization

Following previous work in NLRE [34,36], we validate the LER [Eq. (3)] by evaluating the identity $p(\boldsymbol{x}|\boldsymbol{\theta}^*) \approx p(\boldsymbol{x})\hat{r}(\boldsymbol{x}, \boldsymbol{\theta}^*) =: \hat{p}(\boldsymbol{x}|\boldsymbol{\theta}^*)$ for a fixed parameter set $\boldsymbol{\theta}^*$. The distribution $p(\boldsymbol{x}|\boldsymbol{\theta}^*)$ can be sampled by repeatedly running the forward model with $\boldsymbol{\theta} = \boldsymbol{\theta}^*$ to generate sample cross-correlation functions (and representations in PCA space). Meanwhile, samples from $p(\boldsymbol{x})$ can be obtained by running the forward model with a new parameter set chosen randomly from the prior at each sample. Weighting these samples by the ratio $\hat{r}(\boldsymbol{x}, \boldsymbol{\theta}^*)$ then gives $\hat{p}(\boldsymbol{x}|\boldsymbol{\theta}^*)$. Column 1 of Fig. 3 gives a visual representation of this procedure in the first two components of PCA space: Samples from $p(\boldsymbol{x}|\boldsymbol{\theta}^*)$ are shown using black dots, while samples from $p(\boldsymbol{x})$ are colored by the estimated likelihood ratio $\hat{r}(\boldsymbol{x}, \boldsymbol{\theta}^*)$. The ratio estimator is accurate when black samples from $p(\boldsymbol{x}|\boldsymbol{\theta}^*)$ overlap with the highest-probability regions of the likelihood (red points), and when there are no red points outside the support of the black samples. The first part of this section develops some basic metrics to quantify this comparison, while the second part discusses results for specific classifier architectures.

1. Metrics for quantifying accuracy of ratio estimator

Previous work compared $p(\boldsymbol{x}|\boldsymbol{\theta}^*)$ and $\hat{p}(\boldsymbol{x}|\boldsymbol{\theta}^*)$ by training a classifier to distinguish between samples from $\hat{p}(\boldsymbol{x}|\boldsymbol{\theta}^*)$ and $p(\boldsymbol{x}|\boldsymbol{\theta}^*)$ [34,36]. Paradoxically, success is defined by a *failure* to classify the samples correctly, meaning that the distributions are indistinguishable. In our case, approximating the likelihood is challenging because of randomness in the simulated outputs, which arises from the stochastic nature of the F-actin dynamics (as *in vivo*). Even in local regions of two-dimensional parameter space, significant variation of the first two cross-correlation PCA coordinates is observed (Fig. S3 of the Supplemental Material [63]), and the true likelihood function $p(\boldsymbol{x}|\boldsymbol{\theta})$ changes its width depending on the location of $\boldsymbol{\theta}$ in parameter space (for a few examples, see the spread of the black dots in column 1 of Fig. S4 of the Supplemental Material [63]). Thus, we do not think that it is realistically attainable, nor necessary, for the distributions $\hat{p}(\boldsymbol{x}|\boldsymbol{\theta})$ and $p(\boldsymbol{x}|\boldsymbol{\theta})$ to be indistinguishable in our system, and we ask instead for $\hat{p}(\boldsymbol{x}|\boldsymbol{\theta})$ to be a "good enough" approximation to the likelihood function.

To quantify "good enough," we compute the one-dimensional marginal likelihood over each dimension of PCA

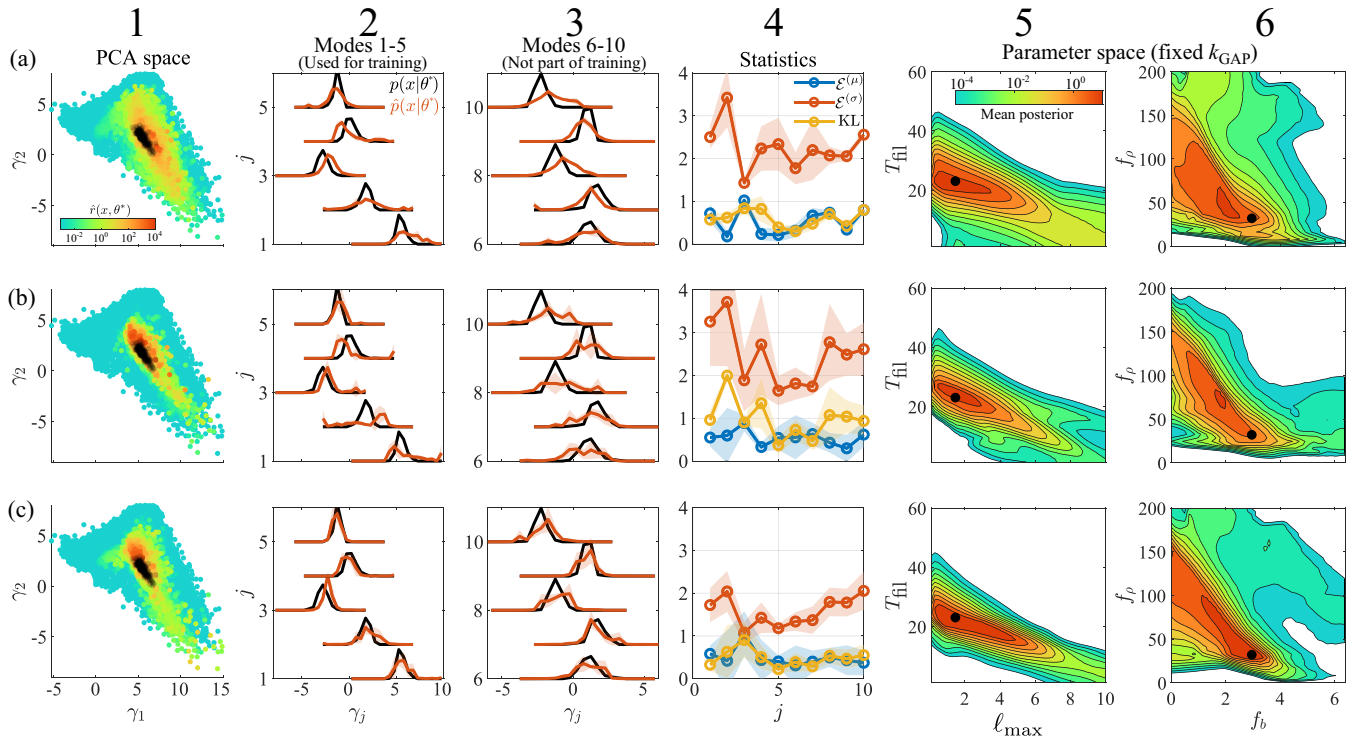


FIG. 3. NLRE on synthetic data is most successful with overexpressive, regularized neural networks. Column 1: Samples from $p(\boldsymbol{x}|\boldsymbol{\theta}^*)$ in the first two components (γ_1, γ_2) of PCA space (black), compared to likelihood ratio estimates $\hat{r}(\boldsymbol{x}, \boldsymbol{\theta}^*)$ for random samples from $p(\boldsymbol{x})$. Column 2: One-dimensional histograms of distributions $p(\boldsymbol{x}|\boldsymbol{\theta}^*)$ (black) and $\hat{p}(\boldsymbol{x}|\boldsymbol{\theta}^*) = p(\boldsymbol{x})\hat{r}(\boldsymbol{x}, \boldsymbol{\theta}^*)$ (red) in PCA space for the first five modes used to train the classifier. Column 3: Same histograms, but for modes 6–10, which are not used to train the classifier. Column 4: Summary statistics in Eq. (4) comparing the one-dimensional distributions for each mode. The relative error in the mean (blue), ratio of standard deviations (red), and KL divergence (yellow) are shown. Columns 5 and 6: Two-dimensional projections of the posterior $p(\boldsymbol{\theta}|\langle \boldsymbol{x}(\boldsymbol{\theta}^*)\rangle)$, where the black dot is the true parameter $\boldsymbol{\theta}^*$ and $k_{GAP} = 0.25$ is fixed to the true value to make the F-actin densities (f_b, f_ρ) meaningful. (a) Classifier has one hidden layer of 40 nodes with no L^2 weight penalty. (b) Classifier has three hidden layers of 20, 40, and 20 nodes, with an L^2 weight penalty of $\lambda = 10^{-5}$. (c) Classifier has three hidden layers of 20, 40, and 20 nodes with an L^2 weight penalty of $\lambda = 10^{-4}$.

space and plot the true marginals together with their approximation in columns 2 and 3 of Fig. 3. To quantify the comparison in each PCA dimension γ_j, we consider the Kullback–Leibler (KL) divergence of the two distributions, as well as the relative error in their means (μ_j and $\hat{\mu}_j$) and ratio of their standard deviations (σ_j and $\hat{\sigma}_j$):

$$\mathcal{E}_j^{(\mu)} = \frac{|\hat{\mu}_j - \mu_j|}{\hat{\sigma}_j}, \quad \mathcal{E}_j^{(\sigma)} = \frac{\hat{\sigma}_j}{\sigma_j},$$

$$\mathrm{KL}_j = \sum_k p\big(\gamma_j^{(k)}|\boldsymbol{\theta}^*\big) \ln \frac{p\big(\gamma_j^{(k)}|\boldsymbol{\theta}^*\big)}{\hat{p}\big(\gamma_j^{(k)}|\boldsymbol{\theta}^*\big)}, \tag{4}$$

where the KL divergence is computed by summing over discrete $\gamma_j^{(k)}$ spaced 0.5 apart in PCA coordinate space. To estimate the mean and variance in Eq. (4), we construct four different approximations $\hat{p}(\boldsymbol{x}|\boldsymbol{\theta}^*)$ using two different sets of 20 000 samples from $p(\boldsymbol{x})$ and two different instantiations of the classifier. The comparison of statistics is shown in column 4 of Fig. 3.

An alternative way to validate the ratio estimator is to perform inference in the reverse direction. That is, instead of fixing $\boldsymbol{\theta}^*$ and checking the probability of different cross-correlation functions $\boldsymbol{x}$, we could fix $\boldsymbol{x} = \bar{\boldsymbol{x}}(\boldsymbol{\theta}^*)$ (the empirical mean cross-correlation function for $\boldsymbol{\theta} = \boldsymbol{\theta}^*$), and then compute the likelihood ratio $\hat{r}(\bar{\boldsymbol{x}}(\boldsymbol{\theta}^*), \boldsymbol{\theta})$ for an arbitrary parameter set $\boldsymbol{\theta}$. For a "good" ratio estimator, the posterior probability $\hat{r}(\bar{\boldsymbol{x}}, \boldsymbol{\theta})p(\boldsymbol{\theta})$ should peak around the original parameters $\boldsymbol{\theta}^*$ used to generate $\bar{\boldsymbol{x}}$. This version of validation is less rigorous because there is no access to a ground truth, as many (potentially unknown) parameter sets could generate cross-correlation functions similar to $\bar{\boldsymbol{x}}$ (for instance, halving k_{GAP} but doubling the amount of F-actin leaves the dynamics unchanged in the mean field limit). Yet, it is still useful because fixing the data and scanning over the parameters is how the ratio estimator will be used in practice. Therefore, we make the amount of F-actin unique by fixing $k_{GAP} = k_{GAP}^*$ and sample a $50 \times 50 \times 50 \times 50$ grid in the four-dimensional space of remaining parameters. We project the approximate posterior $\hat{r}(\bar{\boldsymbol{x}}, \boldsymbol{\theta})p(\boldsymbol{\theta})$ onto two dimensions by averaging over the other parameters and overlay the true parameters $\boldsymbol{\theta}^*$ (black dot) with the average posteriors in columns 5 and 6 of Fig. 3.

2. *Results for different classifier architectures*

We now use our metrics to determine the performance of different classifier architectures for our specific model. Because there are only five parameters, we expect the likelihood of a particular parameter set to be determined approximately

by the first five PCA modes of the cross-correlation, so we begin by training the classifier on the first five modes only. A fully connected neural network has an input layer with size equal to the number of inputs (size of $\boldsymbol{x}$ + size of $\boldsymbol{\theta}$, which equals 10 in our case), an output layer with size equal to the number of outputs (the number of classes, which is 2 here since $y = 0$ or 1), and an arbitrary number of "hidden layers" in between (each of arbitrary size). Our goal in this section is to identify the simplest hidden layer architecture that reduces the error metrics in Eq. (4).

We begin with a single hidden layer of 40 nodes. When we train the neural network without an L^2 penalty on the weights, we still obtain overly wide likelihood functions $\hat{p}(\boldsymbol{x}|\boldsymbol{\theta}^*)$, with $\mathcal{E}^{(\sigma)} \in [2, 3]$ [columns 2–4 of Fig. 3(a)]. Previous work [34] showed empirically that these erroneously wide likelihoods can result when the classifier lacks a sufficient number of parameters to fully describe the likelihood function; in other words, the classifier might not be sufficiently expressive. Based on this reasoning, we expand the network to three hidden layers with 20, 40, and 20 nodes. Even when trained with a small L^2 penalty on the weights, this classifier generates estimates for the likelihood that are high variance [note the large error bars in columns 2–4 of Fig. 3(b)] and have significantly higher KL divergences than the less expressive classifier. An examination of PCA space [column 1 of Fig. 3(b)] reveals that high variance predictions come from isolated points that are erroneously assigned high likelihood, a classic symptom of overfitting [66–68]. Increasing the L^2 penalty on the classifier weights results in approximations with desirable metrics: KL $\leqslant 1$, $\mathcal{E}^{(\mu)} \leqslant 1$, and $\mathcal{E}^{(\sigma)} \leqslant 2$ (column 4 of Fig. 3). Though less rigorous, it is also noteworthy that the true parameter set sits in the region of highest posterior probability only for this last classifier [columns 5 and 6 of Fig. 3(c)].

Interestingly, though the classifier is trained only on modes 1–5, the likelihood estimates for modes 6–10 are also good approximations of the truth [column 3 of Fig. 3(c)], with the errors $\mathcal{E}^{(\mu)}$, $\mathcal{E}^{(\sigma)}$, and KL all lower for modes 6–10 than for some of the modes we used for training [column 4 of Fig. 3(c)]. In fact, of the first 40 modes, only mode 28 has a KL divergence larger than 1, and only mode 34 has a standard deviation ratio larger than 3 (not shown). Thus, in the context of synthetic data, getting the first five modes right is sufficient to reproduce the entire spectrum.

Summing up, we find that a regularized, expressive classifier, with only the first five PCA modes used as input, is sufficient to approximate $p(\boldsymbol{x}|\boldsymbol{\theta}^*)$. Although our approximate likelihood is distinguishable from the true likelihood, the peak in each dimension is at most one standard deviation away from the true peak, and the approximate standard deviation in each dimension is at most twice the true standard deviation. Figure S4 of the Supplemental Material [63] shows that these observations are robust to several different choices of $\boldsymbol{\theta}^*$ [different widths of the true likelihood $p(\boldsymbol{x}|\boldsymbol{\theta}^*)$], implying that the ratio estimator is robust and accurate for synthetic data.

B. Likelihood ratio estimate complements cross-correlation error

The main advantage of NLRE is the amortized likelihood function, which allows for inference of parameters from new data without having to perform additional simulations. However, how does the approximate likelihood compare to errors in summary statistics? To answer this question, we compare the output of NLRE to errors in the cross-correlation function, which we employed in previous work [30].

To compare the two metrics, we sample 5000 cross-correlation functions $\boldsymbol{x} \sim p(\boldsymbol{x})$, compute the mean error relative to 200 samples $\boldsymbol{x}^* \sim p(\boldsymbol{x}|\boldsymbol{\theta}^*)$ (using the Frobenius norm), and compare this to the mean estimated LER $\hat{r}(\boldsymbol{x}, \boldsymbol{\theta}^*)$ (Fig. 4 shows the same $\boldsymbol{\theta}^*$ as Fig. 3 and Fig. S5 of the Supplemental Material [63] shows an additional $\boldsymbol{\theta}^*$). Generally speaking, parameter sets with low cross-correlation error are deemed high likelihood, while parameter sets with high cross-correlation error are deemed low likelihood [Fig. 4(a) and Fig. S5(a) of the Supplemental Material [63]]. There are, however, a few samples that have high likelihood and high cross-correlation error, or low likelihood and low cross-correlation error. These samples merit further examination, as the inferred likelihood depends on the choice of method.

We first isolate samples with low cross-correlation errors and varying estimated likelihood by considering the left-most region [vertical rectangle in Fig. 4(a)] where cross-correlation error $\bar{e}$ is roughly constant (at most 10% difference in relative error), but the estimated likelihood $\bar{\ell}$ changes significantly (by at least 3 orders of magnitude). Samples of these cross-correlation functions (spaced uniformly in $\bar{\ell}$) demonstrate subtle changes that move the cross-correlation farther from the true mean as the likelihood decreases [Fig. 4(b) and Fig. S5(b) of the Supplemental Material [63]]. For an oscillatory cross-correlation, the central positive band appears to decrease in width [moving from top to bottom in Fig. 4(b)], while for a diffuse positive cross-correlation, a negative patch appears near $(\Delta r, \Delta t) = 0$ [bottom panel of Fig. S5(b)]. The change in likelihood is best understood by examining the cross-correlation functions in PCA space [Fig. 4(b′)], where low-likelihood samples are positioned closer to the edges of the true likelihood distribution [black distributions in Fig. 4(b′)], with mode 5 appearing particularly sensitive in both parameter sets [Fig. 4(b′) and Fig. S5(b′) of the Supplemental Material [63]]. Since the fifth mode has the highest peak in the marginal likelihood, we conclude that the classifier penalizes sharply constrained modes more than others, thus overcoming some of the limitations associated with a single summary statistic (the error).

We next isolate samples with high estimated likelihood and varying error by considering the topmost region [horizontal rectangle in Fig. 4(a)] where the likelihood estimate $\bar{\ell}$ is roughly constant (within one order of magnitude), but cross-correlation error $\bar{e}$ changes significantly (by at least 0.5). Interestingly, these cross-correlation functions [Fig. 4(c) and Fig. S5(c) of the Supplemental Material [63], approximately uniformly spaced in $\bar{e}$] all resemble the original function but typically have more extreme features as the error increases. Indeed, it is possible to increase the error by over 50% while remaining within the support of the likelihood function [$\bar{e} = 0.7$ in Fig. 4(c′)], which suggests that error can be an unnecessarily harsh statistic. On the other hand, very high errors [$\bar{e} = 1$ in Fig. 4(c′)] fall outside of the true likelihood (for mode 2 especially), which implies that the estimated likelihood might be erroneously large in these cases.

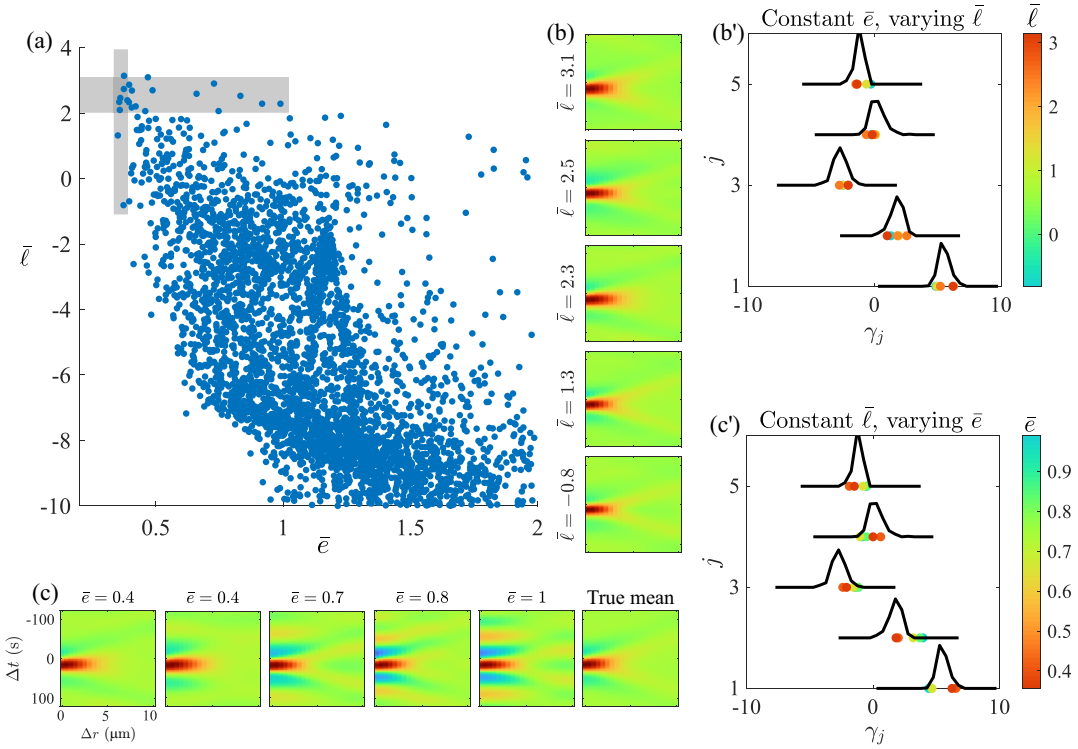


FIG. 4. Likelihood ratio estimate and cross-correlation error contain complementary information. (a) Scatter plot comparing the mean error in cross-correlation $\bar{e}(\boldsymbol{x}) = (\langle \|\boldsymbol{x} - \boldsymbol{x}(\boldsymbol{\theta})\|_F^2 \rangle / \langle \|\boldsymbol{x}(\boldsymbol{\theta})\|^2 \rangle)^{1/2}$, with the mean log-likelihood ratio estimate $\bar{\ell} = \log_{10} \langle \hat{r}(\boldsymbol{x}, \boldsymbol{\theta}) \rangle$. Gray rectangles denote high-probability regions (size 0.1 in cross-correlation and 1 in log-likelihood space) where there is variation in one quantity but not the other. (b) Samples of constant cross-correlation error and varying log-likelihood. (b′) Corresponding samples in PCA space, colored by the log-likelihood estimate, and compared to the true likelihood $p(\gamma_j|\boldsymbol{\theta}^*)$. (c) Samples of constant log-likelihood estimate but varying cross-correlation error. (c′) Corresponding samples in PCA space, colored by the cross-correlation error and compared to the true likelihood $p(\gamma_j|\boldsymbol{\theta}^*)$. The empirical mean cross-correlation $\bar{\boldsymbol{x}}(\boldsymbol{\theta}^*) \approx \langle \boldsymbol{x} p(\boldsymbol{x}|\boldsymbol{\theta}^*) \rangle$ is shown at the bottom of panel (b) and right of panel (c).

Overall then, we can say that cross-correlation errors and estimated likelihoods are *complementary* statistics. Looking only at cross-correlation errors can discount the nonzero width of the true likelihood, especially when some modes are more sharply constrained than others. At the same time, however, abnormally large cross-correlation errors are a sign the estimated likelihood function is erroneously wide.

C. Inference on experimental data requires additional regularization and higher-dimensional training data

Having validated the model on synthetic data, we now transition to experimental data, first considering the cross-correlation functions from starfish and worms [Fig. 1(b′)]. As we previously used ABC to identify parameter regions matching these datasets [30], this represents a midpoint between synthetic data (where rigorous validation is possible) and completely new experimental data (where we have yet to determine the model parameters). In the context of NLRE, the data (which we denote by capital $\boldsymbol{X}_d$) are given, and the only quantity that we can evaluate is the likelihood ratio estimate $\hat{r}(\boldsymbol{X}_d, \boldsymbol{\theta})$ for model parameters $\boldsymbol{\theta}$. Multiplying by the prior then gives the estimated posterior $\hat{p}(\boldsymbol{\theta}|\boldsymbol{X}_d) = \hat{r}(\boldsymbol{X}_d, \boldsymbol{\theta}) p(\boldsymbol{\theta})$. Figures 5(a) and 5(e) show two-dimensional projections of the posterior for starfish and worm data, respectively, using the optimal classifier settings from the synthetic case.

As the ratio estimator is pretrained on synthetic data, the posterior estimate might be poor when the data $\boldsymbol{X}_d$ do not resemble the training data. To get a sense of the accuracy of the posterior, we identify (from simulations of 50 000 parameter sets with $k_{\mathrm{GAP}} = 0.4$ fixed) the simulated data $\boldsymbol{x}_m(\boldsymbol{\theta}_m)$ that minimize $\|\boldsymbol{X}_d - \boldsymbol{x}_m\|$, and then plot the corresponding parameters $\boldsymbol{\theta}_m$ on top of the estimated posterior distributions (first two columns of Fig. 5). Since minimal errors should correlate with maximal likelihood estimates (Fig. 4), $\boldsymbol{\theta}_m$ (indicated by the black dot) ought to be near the peak posterior $\boldsymbol{\theta}_{\mathrm{mp}}$ (indicated by the white dot). As an additional check, we simulate the model twice with $\boldsymbol{\theta} = \boldsymbol{\theta}_{\mathrm{mp}}$ [two samples of $\boldsymbol{x} \sim p(\boldsymbol{x}|\boldsymbol{\theta}_{\mathrm{mp}})$] to compute an average cross-correlation function (third column of Fig. 5). Ideally, this should resemble the experimental data. In addition to these simpler tests, in Fig. S6 of the Supplemental Material [63] we also define an estimated probability distribution $\hat{p}(\boldsymbol{x}|\boldsymbol{X}_d) := \hat{r}(\boldsymbol{X}_d, \boldsymbol{\theta}) p(\boldsymbol{x}|\boldsymbol{\theta})$ that is based on sampling $p(\boldsymbol{x}|\boldsymbol{\theta})$ for random $\boldsymbol{\theta}$, and then weighting the result by $\hat{r}(\boldsymbol{X}_d, \boldsymbol{\theta})$. Ideally, $\hat{p}(\boldsymbol{x}|\boldsymbol{X}_d)$ should peak near $\boldsymbol{X}_d$, or at least near $\boldsymbol{x}_m$, the model's closest approximation to $\boldsymbol{X}_d$.

Unfortunately, directly applying the same classifier architecture and settings used for synthetic data to experimental data yields a posterior that fails to satisfy these tests [Figs. 5(a) and 5(e)]. For both the starfish and worm data, the minimum-error parameter set falls at least an order of magnitude away

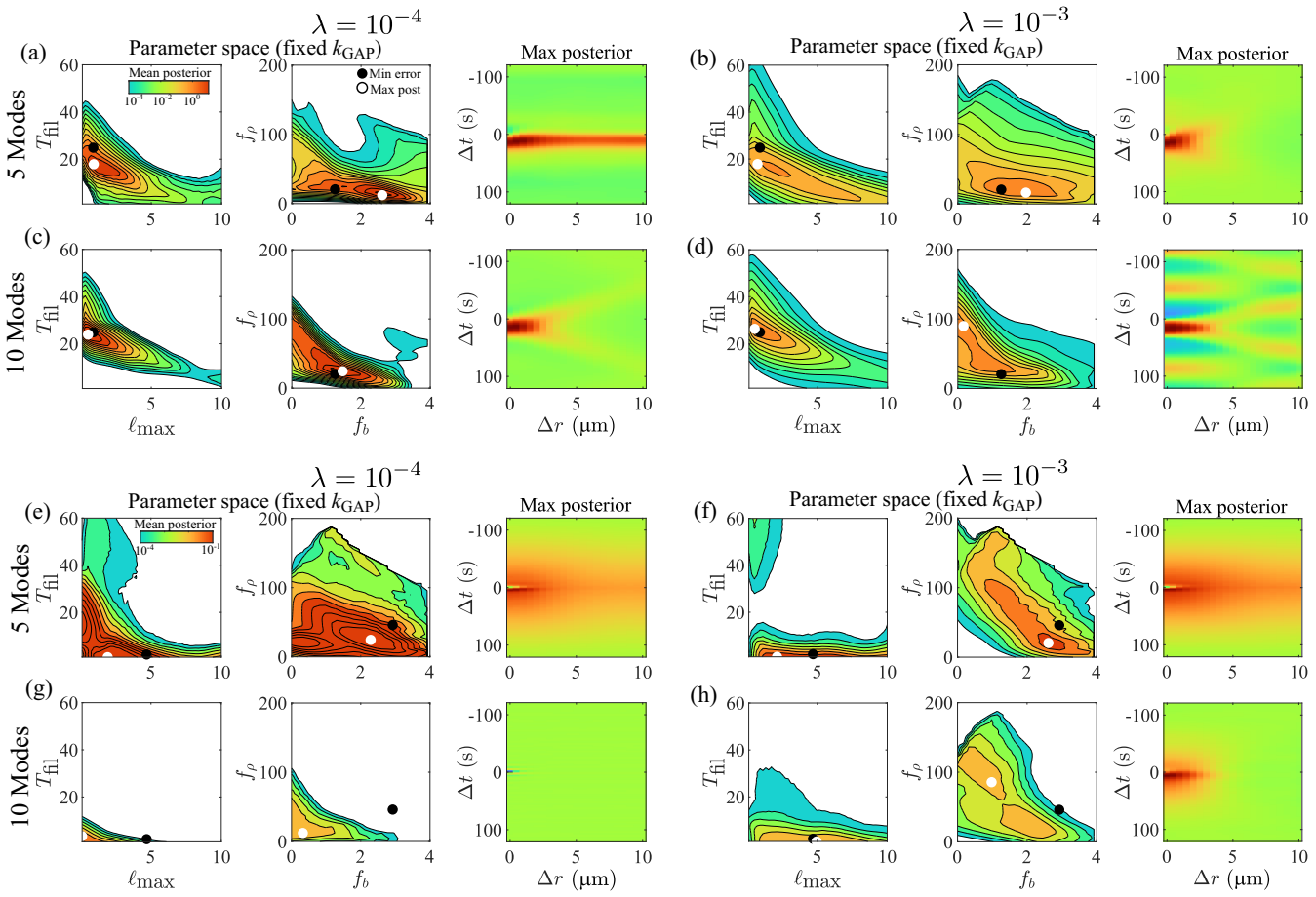


FIG. 5. Experimental data require additional regularization strategies due to model misspecification. First two columns of each panel: Two-dimensional projections of the estimated posterior $\hat{p}(\boldsymbol{\theta}|\boldsymbol{X}_d) = \hat{r}(\boldsymbol{X}_d, \boldsymbol{\theta})p(\boldsymbol{\theta})$ when $k_{\mathrm{GAP}} = 0.4$ is fixed and $\boldsymbol{X}_d$ is the (a)–(d) starfish or (e)–(h) worm data [see Fig. 1(b)]. Black dots: Parameters that minimize cross-correlation error. White dots: Parameters that maximize the posterior (occasionally not at the peaks due to two-dimensional projection). Third column: The average simulated cross-correlation function obtained using the parameters that maximize the posterior. (a), (e) Default classifier parameters ($\lambda = 10^{-4}$, five PCA modes) from synthetic validation. (b), (f) Additional regularization ($\lambda = 10^{-3}$) with five PCA modes. (c), (g) More (ten) PCA modes, with regularization held at $\lambda = 10^{-4}$. (d), (h) More PCA modes and additional regularization ($\lambda = 10^{-3}$, ten modes).

from the peak posterior, and the peak posterior parameters generate cross-correlation functions that do not resemble the corresponding experimental ones. This behavior contrasts with our analysis of synthetic data and thus must result because the physical model, which enters through the training data, does not capture the true data-generating process (*in vivo* dynamics). This is defined as *model misspecification* in the SBI literature [51,53,69]. Since the minimum-error cross-correlation functions $\boldsymbol{x}_m(\boldsymbol{\theta}_m)$ lie close to experimental data in PCA space (Fig. S6 of the Supplemental Material [63]), the model can be made to resemble the starfish and worm data in certain parameter regimes [30]. Thus, the issue is not with the model but with the neural estimator's ability to find those parameter regimes given experimental data.

We consider two nonexclusive possibilities for why the estimator performs poorly under model misspecification: failure of extrapolation and lack of parameter identifiability. A failure of extrapolation can result when the classifier estimates $\hat{r}(\boldsymbol{X}_d, \boldsymbol{\theta})$ for data outside the training set even if there is no overfitting inside the training set. This phenomenon manifests as high variance in the estimated likelihood ratios [see the shading and blue arrows in Figs. S6(a) and S6(e) of the Supplemental Material [63]] and can be mitigated by increasing the regularization parameter λ in the classifier loss function. On the other hand, a lack of identifiability occurs when the modes used are insufficient for the classifier to relate experimental data to training data. This can occur because the PCA modes are computed using *simulated* data and thus need not be representative of experimental data. In this case, parameters can be made more identifiable by increasing the number of modes used for inference.

Figure 5 shows that the combination of more regularization and more PCA modes in training yields reliable posterior estimates for both the starfish [panels (a)–(d)] and worm [panels (e)–(h)] data. In both parameter and PCA space, more regularization alone gives higher relative probabilities of the minimum-error parameter sets, but does little to exclude cross-correlation functions that do not resemble the data [Figs. 5(b) and 5(f) and see also Figs. S6(b) and S6(f) of the Supplemental Material [63] that show overly wide distributions $\hat{p}(\boldsymbol{x}|\boldsymbol{X}_d)$ in PCA space]. Alternatively, increasing the number of modes alone sharpens the posteriors and

makes the parameters identifiable but still gives erroneously high posterior estimates in some regions of parameter space [Figs. 5(c) and 5(g) and see also Figs. S6(c) and S6(g) of the Supplemental Material [63]]. Combining more regularization with more PCA modes gives the best outcomes: Posteriors assign high probability to the minimum-error parameter sets, and maximal-posterior parameters generate cross-correlation functions that resemble the data. In PCA space [Figs. S6(d) and S6(h) of the Supplemental Material [63]], the probability distributions $\hat{p}(\boldsymbol{x}|\boldsymbol{X}_d)$ peak near the data and/or its best model approximation [compare, for example, green arrows in Figs. S6(b) and S6(d) of the Supplemental Material [63]]. Interestingly, the minimum-error parameter set for the worm data falls far from the maximum posterior [black dot in Fig. 5(h)], but the minimum-error cross-correlation function is also far from the data in the first PCA mode [see the blue star in Fig. S6(h) of the Supplemental Material [63]], so we do not view this as a shortcoming of the ratio estimator. Summing up, the optimal classifier training settings ($\lambda = 10^{-3}$ and ten PCA modes) combine more regularization (to address overfitting when $\boldsymbol{X}_d$ is not part of the training data) with more modes (to sharpen identifiability). Figure S7 of the Supplemental Material [63] shows that using these classifier settings on synthetic data widens the estimated marginal likelihoods in some cases, without inducing strong changes in the location of the peaks or KL divergence relative to the true likelihoods.

D. Validated NLRE reveals how changing F-actin assembly parameters shifts dynamics from waves to pulses

We previously showed that the difference between the worm and starfish data can be explained by choosing parameters in diametrically opposite regions of F-actin parameter space, with starfish having long-lived, short F-actin, and worms having short-lived, long F-actin [30]. A more complete understanding of the transition between the two types of behavior can be obtained by using NLRE to sample the complete parameter space, in this case by fixing $k_{\mathrm{GAP}} = 0.4$ and sampling the posterior on a $50 \times 50 \times 50 \times 50$ grid in the space of $\ell_{\max}$, T_{fil}, f_b, and f_ρ [Fig. 6(a) shows the average posterior in two-dimensional subsets of parameter space]. The posterior estimates clearly show that the two parameters of F-actin *architecture* ($\ell_{\max}$ and T_{fil}) are most distinguishable across datasets. Interestingly, the likelihood ratios for the worm data are significantly smaller than those for the starfish, which reflects the relative impacts of model misspecification discussed previously (Fig. 5 and Fig. S6 of the Supplemental Material [63]). Despite this, independent kinetic measurements of F-actin dynamics in *C. elegans* have found maximum filament lengths of 6 μm [62] and subunit turnover times $T_{\mathrm{sub}} = \ell_{\max}/v_p + T_{\mathrm{fil}}$ on the order of 8–10 s [11] (which translates to $T_{\mathrm{fil}} = 2$ s). These direct parameter measurements fall squarely within the high-likelihood regime of our inference [blue dot in Fig. 6(a)], thus validating our ability to predict the microscopic dynamics responsible for pattern formation.

Reducing the parameter space to subunit lifetime alone further accentuates the distinction between starfish and worms, and allows us to find parameter sets that interpolate between the two. Plotting the likelihood-to-evidence ratio estimate for

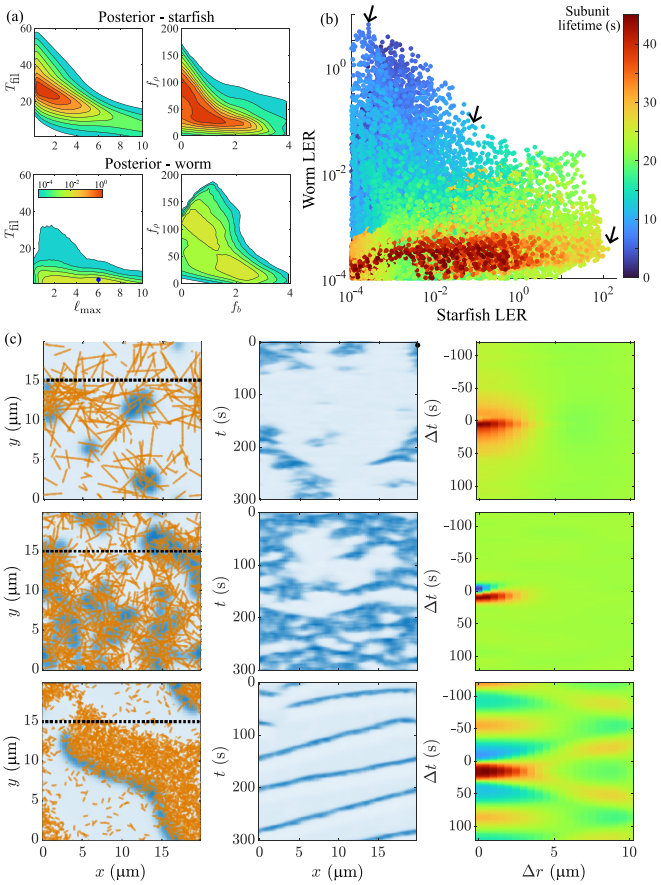


FIG. 6. Posterior distributions reveal continuous transition between starfish and worm data, depending on the actin subunit lifetime. (a) Two-dimensional projections of the posterior for starfish (top) and worm (bottom) data with fixed $k_{\mathrm{GAP}} = 0.4$ (sampled at 50^4 parameters). (b) Scatter plot of the LER $\hat{r}(\boldsymbol{x}_{\mathrm{star}}, \boldsymbol{\theta})$ against $\hat{r}(\boldsymbol{x}_{\mathrm{worm}}, \boldsymbol{\theta})$ for 10% of the samples (number reduced for clarity, samples taken only at points away from the boundary of the prior, measured as probability $\geqslant$75% of sustaining excitation). (c) Sampled dynamics at the arrows indicated in panel (B), from top to bottom. Left panels show snapshots at $t = 60$, middle panels show kymographs about $y = 15$, and right panels show cross-correlation [cf. data in Fig. 1(b′)].

starfish $\hat{r}(\boldsymbol{x}_{\mathrm{star}}, \boldsymbol{\theta})$ against that for worms $\hat{r}(\boldsymbol{x}_{\mathrm{worm}}, \boldsymbol{\theta})$ and coloring points by subunit lifetime [Fig. 6(b)] shows that subunit lifetimes larger than 20 s are favorable for the starfish data alone, while subunit lifetimes smaller than 10 s are favorable for the worm data alone. Simulations using the maximum likelihood parameter set for each dataset strongly resemble the original data [Fig. 6(c)], allowing us to assert with renewed confidence that changes in F-actin assembly dynamics drive the change from Rho waves in starfish to Rho pulses in *C. elegans* [30]. Wavelike dynamics correlate with short, long-lived filaments, which lag spreading Rho wave fronts. Pulselike dynamics correlate with long, short-lived filaments, which polymerize rapidly out of excitations, allowing them to spread only slightly before being extinguished.

A key advantage of NLRE is that we can classify any parameter set as more wavelike or pulselike, depending on the comparison of likelihood against the two types of experi-

mental data. Figure 6(b) shows that the dividing line between wavelike and pulselike data occurs around subunit lifetime of 15 s. To visualize the dividing line, we find a value of $c = 10^{-1.12}$ where there is exactly one parameter set with both $\hat{r}(\boldsymbol{x}_{\text{worm}}, \boldsymbol{\theta}) > c$ and $\hat{r}(\boldsymbol{x}_{\text{star}}, \boldsymbol{\theta}) > c$ [middle black arrow in Fig. 6(b)]. A corresponding simulation from this parameter set [middle panels of Fig. 6(c)] shows an interpolation between the wavelike and pulselike patterns. The dynamics still have the signature of wave spread, with negative cross-correlation on $\Delta t < 0$, but there are no long-time oscillations in the cross-correlation function, which is more similar to the worm data. We emphasize again that the parameter sets in Fig. 6 are *not* part of the training data. Thus, NLRE is able to predict new dynamics that resemble worms, starfish, and in between. This allows us to be more confident in parameter inference when there appears to be a mixture of waves and pulses, as in the data we consider next.

E. Inferring how RhoGAP could modulate wave dynamics

An important advantage of NLRE is the amortized likelihood function; since the classifier is already pretrained, we can infer parameters of new experimental dynamics without having to perform more simulations. NLRE therefore provides an opportunity to examine data from perturbation experiments that do not lend themselves to straightforward interpretation. In a typical perturbation experiment, the expression of one of the circuit's molecular players is modified (e.g., through RNA interference or microinjected RNA), and the resulting dynamics are read out on the cell scale. Such a perturbation can be challenging to interpret because a molecule can play multiple roles, and feedbacks make them hard to parse. Our approach provides an opportunity to infer how the observed patterns could result from changes to specific model parameters, which we can then connect to the molecular players under scrutiny.

We use as an example data collected from activated frog oocytes [24], where extra Ect-2 (RhoGEF) and RGA-3/4 (RhoGAP) are coexpressed from microinjected mRNA: Ect-2 at constant level (200 ng/μl) and RGA-3/4 at varying levels. Representative trajectories for RGA-3/4 levels ranging from 33 to 1000 ng/μl [Fig. 7(a)] reveal a continuous increase in the frequency and coherence of waves as RhoGAP expression levels increase (data for 0 ng/μl are shown in Fig. S8 of the Supplemental Material [63]) [24]. Previous continuum-level modeling reproduced many of the features of these dynamics by incorporating molecular noise (modeled as a spatially and temporally correlated Gaussian stochastic field) into the F-actin disassembly rate [24]. Our approach is to ground this molecular noise in our model of F-actin assembly dynamics, and then use NLRE to infer how changes to model parameters lead to the observed changes in wave propagation.

1. Increasing GAP strength alone cannot explain the data

Given that only the GAP expression levels are varied in the data, a sensible first step is to fix the parameters of F-actin assembly and increase the GAP strength k_{GAP} in the model. When the GAP strength is too low, the F-actin/RhoGAP levels are not high enough to push Rho activity to the lower steady state, and Rho activity levels are high everywhere [left panel of Fig. 7(b)]. A slight increase in k_{GAP} destabilizes the uniform state, giving rise to coherent waves with sharp, oscillatory cross-correlation functions [second panel from left, Fig. 7(b)]. As the GAP strength increases further, the waves become less coherent as stronger inhibition pushes Rho down to its lower steady state in more places. When the GAP strength is sufficiently high, Rho activity ceases unless actively restarted [right panel of Fig. 7(b)].

As the F-actin length is small [0.5 μm in Fig. 7(b)], it is not surprising that our results match those of previous continuum models [24]. Similar to those results, the dynamics of Rho below a critical point are dominated by fluctuations around a uniform state. This is supported by the data without any RhoGAP coexpression, though the cross-correlation in that case comes from averaging the random fluctuations, and is consequently more difficult to interpret (Fig. S8 of the Supplemental Material [63]). Above the critical k_{GAP}, waves emerge, and further increases in k_{GAP} beyond this point only lower the maximum wave amplitude and temporal width (defined as the time between half maxima for a given point in space); cf. Figs. 7(b)–7(d) in Ref. [24].

At the same time, however, analysis of cross-correlation functions [in cases when they are informative, Figs. 7(a′) and 7(b′)] shows that the model does not fully reproduce the data. In the data, waves emerge gradually, with *more* coherent waves as GAP strength increases, while in the model the waves emerge suddenly, and then become *less* coherent as GAP strength increases. Furthermore, in the data, the wave frequency (as seen in the cross-correlation function) increases as GAP expression increases, while in the model the wave frequency is approximately constant as GAP strength increases (the central band in the cross-correlation has approximately the same width, regardless of GAP strength). This pattern holds not just for the illustrative (fixed) F-actin assembly parameters in Fig. 7(b), but also for F-actin assembly parameters chosen to maximize the likelihood while ensuring that the mean inferred k_{GAP} increases as RGA-3/4 expression levels increase (Fig. S9 of the Supplemental Material [63]). Thus, the assumption that RhoGAP levels affect only the strength of inhibition and not any of the other model parameters cannot be correct. This conclusion is not model dependent: If filament assembly rates are the same, adding additional RhoGAP should only extinguish Rho activity, not sharpen it.

2. NLRE predicts an essential role for nucleation rates

Given that changing only the GAP strength cannot reproduce the data, a logical next step is to look for another parameter that must change as the GAP strength increases. This also provides an opportunity to leverage the power of NLRE; since the likelihood is a function of all five parameters, we can easily fix a subset and examine which parameters must change to maintain a high likelihood.

There are four candidate parameters that could change alongside GAP strength: the filament lifetime T_{fil}, the filament length ℓ_{max}, and the basal and Rho-mediated nucleation rates q_b and q_ρ (which we report in terms of nucleated amounts f_b and f_ρ). For each candidate parameter (e.g., T_{fil}), we identify potential fixed values of the other parameters (ℓ_{max}, q_b, and q_ρ) by imposing two constraints: First, with those values fixed,

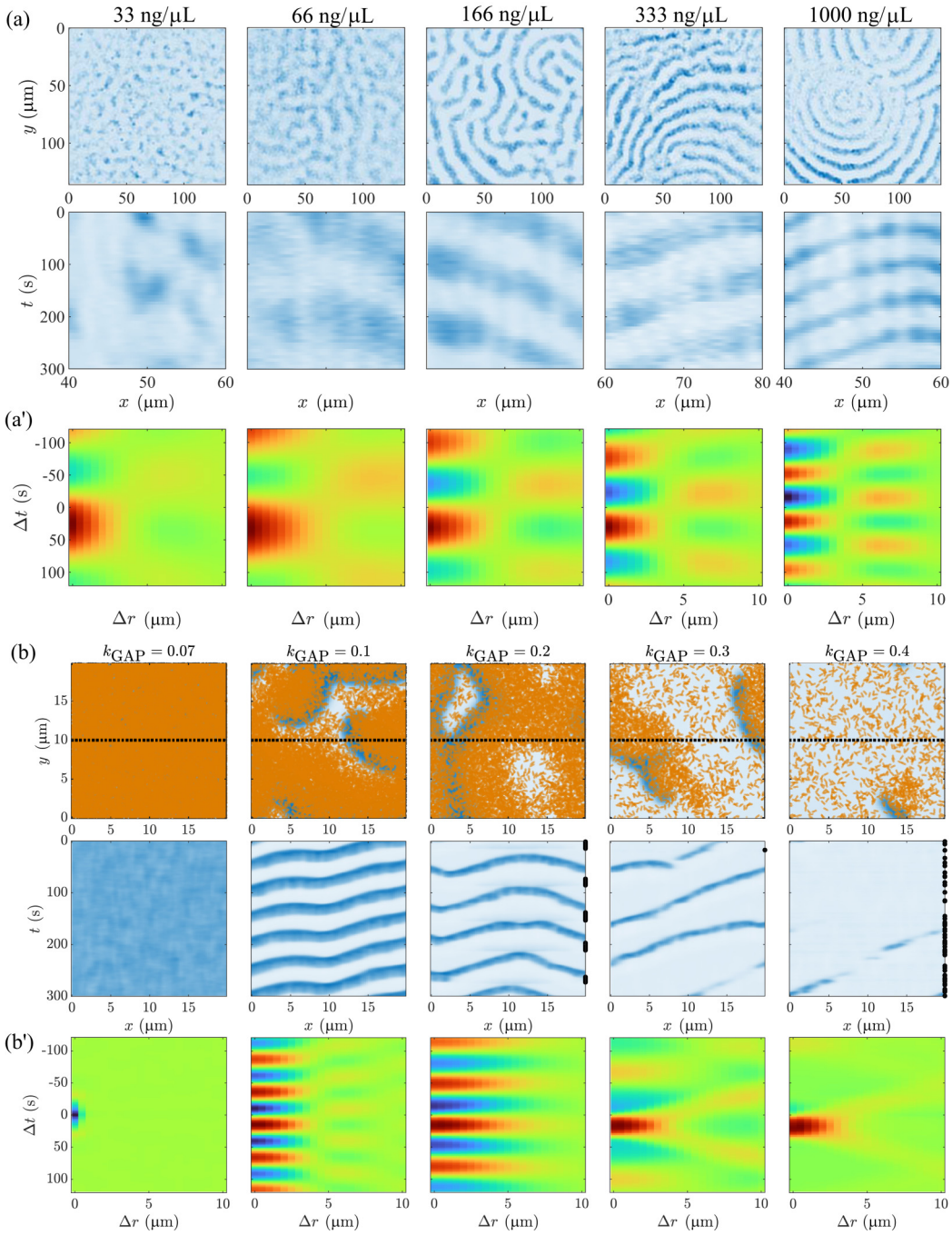


FIG. 7. Increasing RhoGAP coexpression leads to faster, more coherent waves *in vivo*, which cannot be explained by increasing GAP strength alone in the model. (a) Snapshots and kymographs (along 20 μm distance) in activated frog eggs coexpressing 200 ng/μl Ect2 and varying levels of RhoGAP (RGA-3/4) (indicated in the titles). (a′) Corresponding cross-correlation functions. (b) Model simulations with increasing (from left to right) GAP strength k_{GAP} and constant filament assembly dynamics and nucleation rates ($T_{fil} = 30$ s, $\ell_{max} = 0.5$ μm, $f_b = 1$ μm^{-1}, $f_\rho = 80$ μm^{-1}). (b′) Corresponding cross-correlation functions. In the model, highly coherent waves emerge suddenly once the GAP strength surpasses a threshold, after which coherence decreases as the GAP strength increases.

each dataset must have at least one (k_{GAP}, T_{fil}, ℓ_{max}, q_b, q_ρ) combination where the data favor the parameters (likelihood-to-evidence ratio larger than 1). Second, the mean k_{GAP} (measured via the first moment of the posterior distribution) must be an increasing function of RhoGAP expression levels [see Appendix and Eq. (A3) for more details]. We plot the mean posterior in two dimensions, averaged over all potential values of the three fixed parameters, then choose a specific parameter set that balances high posterior probability with increasing mean k_{GAP} values [via maximizing Eq. (A4)]. For this parameter set, we plot the corresponding two-dimensional posterior over the varying parameters (Fig. S11 of the Supplemental Material [63]), and then simulate with parameters chosen from the highest likelihood regions (and such that

k_{GAP} and the other varied parameters vary monotonically with RhoGAP expression).

This procedure identifies two candidate parameters that could be changing with k_{GAP}, with all other candidates failing to satisfy our two criteria. A first possibility is the filament lifetime T_{fil}, which could decrease with increasing RhoGAP levels [Fig. 8(a)]. While it is possible to see waves of increasing coherence in this case [Fig. 8(c)], the optimal k_{GAP} varies over a small range, from $k_{GAP} = 0.3$ for the lowest coexpression levels (33 ng/μl) to $k_{GAP} = 0.35$ for the highest ones (1000 ng/μl). Thus, an increase of 30-fold RNA corresponds to only a 17% increase in k_{GAP}. Simulations with representative parameter sets [Fig. 8(c)] suggest that, when inferred k_{GAP} levels do not vary much, NLRE is focusing on common features of all five cross-correlation functions, and the increases in the mean k_{GAP} with RGA-3/4 coexpression are coincidental. Based on these considerations, as well as a Bayes factor calculation that shows a relatively low probability for this model (Appendix and Fig. S10 of the Supplemental Material [63]), we disfavor a mechanism in which T_{fil} alone changes with k_{GAP}.

The second parameter that satisfies our criteria is the basal nucleation rate q_b (and by extension the amount of basal filaments f_b). In this case, the inferred parameter sets span a broad range in (k_{GAP}, f_b) space, with the smallest expression levels clustering around ($k_{GAP} = 0.2$, $f_b = 2$), and the largest around ($k_{GAP} = 0.6$, $f_b = 0$). The remaining fixed parameters ($T_{fil} = 44$ s, $\ell_{max} = 0.5$ μm, and $f_\rho = 78$/μm is a representative example) are consistent with our existing knowledge of the frog system; fluorescence recovery after photobleaching (FRAP) experiments show F-actin recovery rates on the order of 40 s [70], and filaments must be short and homogeneous to generate waves [30]. Simulations along the most likely path in (k_{GAP}, f_b) space reproduce many aspects of the data: Waves become more coherent, higher frequency, and have a lower temporal width, and cross-correlation functions become more oscillatory in time, with the central positive band shrinking in size as RhoGAP coexpression levels increase (though not as significantly as in the data). Thus, the model clearly suggests that sharper waves of Rho activity occur when basal nucleation rates are also reduced in response to increasing RhoGAP activity.

Mechanistically, the output of our inference can be understood in terms of the role of basal and Rho-mediated F-actin [Fig. 8(e)]. Typically, Rho-mediated actin assembles behind wave fronts of Rho (and stays there for quite some time), effectively "pushing" the wave front in one direction. Basal F-actin, in contrast, inhibits spreading excitation, blocking the formation of coherent traveling waves. In the model, a high amount of basal F-actin is necessary to reproduce the incoherent waves that arise at low RhoGAP levels. As RhoGAP levels increase, inhibition becomes more intense, and so holding the basal assembly levels constant makes it impossible for waves to maintain their structure, much less become more coherent. Basal filament levels must therefore decrease with RhoGAP. In the case when Rho-mediated nucleation levels are constant, increasing RhoGAP makes the trailing inhibition more powerful, which pushes the wave at a faster speed, decreasing the temporal width in a given location.

3. Potential molecular mechanisms to tune nucleation

The result that basal nucleation must decrease with RhoGAP expression suggests a number of possible molecular mechanisms, depending on how we interpret basal filament nucleation. In the simplest interpretation, all of the F-actin is still Rho mediated, and basal assembly occurs from local fluctuations in Rho density, which are not included in the model (and are averaged over in the cross-correlation analysis). RhoGAP would then act to suppress these fluctuations, without altering the rate of Rho-mediated nucleation. Alternatively, basal filaments could be interpreted as F-actin nucleated independently of Rho via a basal rate of formin activity. In this case, RhoGAP would inhibit formin directly (as has been shown in other systems [71]) and Rho-mediated nucleation would also likely decrease with increases in RhoGAP. Finally, interpreting basal nucleation as independent of the Rho pathway yields the hypothesis that RhoGAP inhibits other GTPases and their effectors (e.g., CDC-42/WASP/Arp2/3 [72,73]). If the amount of F-actin on the cell cortex is conserved [74–76], a reduction in CDC-42-mediated F-actin could cause an increase in Rho-mediated F-actin. Given that the dynamics in Fig. 8(d) can also be reproduced when Rho-mediated assembly decreases or increases in response to RhoGAP (Fig. S12 of the Supplemental Material [63]), these hypotheses would be best differentiated by performing live imaging of formin molecules, which would yield local rates of Rho-mediated and formin-mediated nucleation.

V. CONCLUSION

In this work, we applied methods for parameter inference to learn how the activator-inhibitor interaction between Rho and F-actin dictates cell-scale pattern formation. Our methodological contribution was to adapt NLRE [34] to quantify the correspondence between models and data via classification neural networks, appropriately regularized in the presence of significant model stochasticity. Using the high-throughput likelihood function, we mapped how varying Rho activity patterns could emerge from changes in F-actin assembly dynamics in starfish, worm, and frog cells. We found that the change from pulselike behavior to cohesive waves could be accomplished by making actin filaments shorter and longer lived [30], and we validated the inferred parameters from NLRE by comparing to *in vivo* kinetic measurements from *C. elegans* [11,62]. We also found that more cohesive waves could be achieved with higher RhoGAP levels, but only when basal rates of F-actin assembly are reduced, and suggested experiments to test this prediction.

On the methods side, our work revealed the advantages of neural SBI, while also demonstrating the challenges that arise when trying to apply these methods to noisy biological data. Even for low-dimensional representations of the data, our synthetic model exhibited marginal likelihoods with relatively large variance in some regions of parameter space and small variance in others. We showed that an overexpressive classifier, with loss function regularized to prevent overfitting, can approximately learn this space, with approximate marginal likelihoods that are at most twice as wide as the true ones. The approximate likelihood provides complementary information

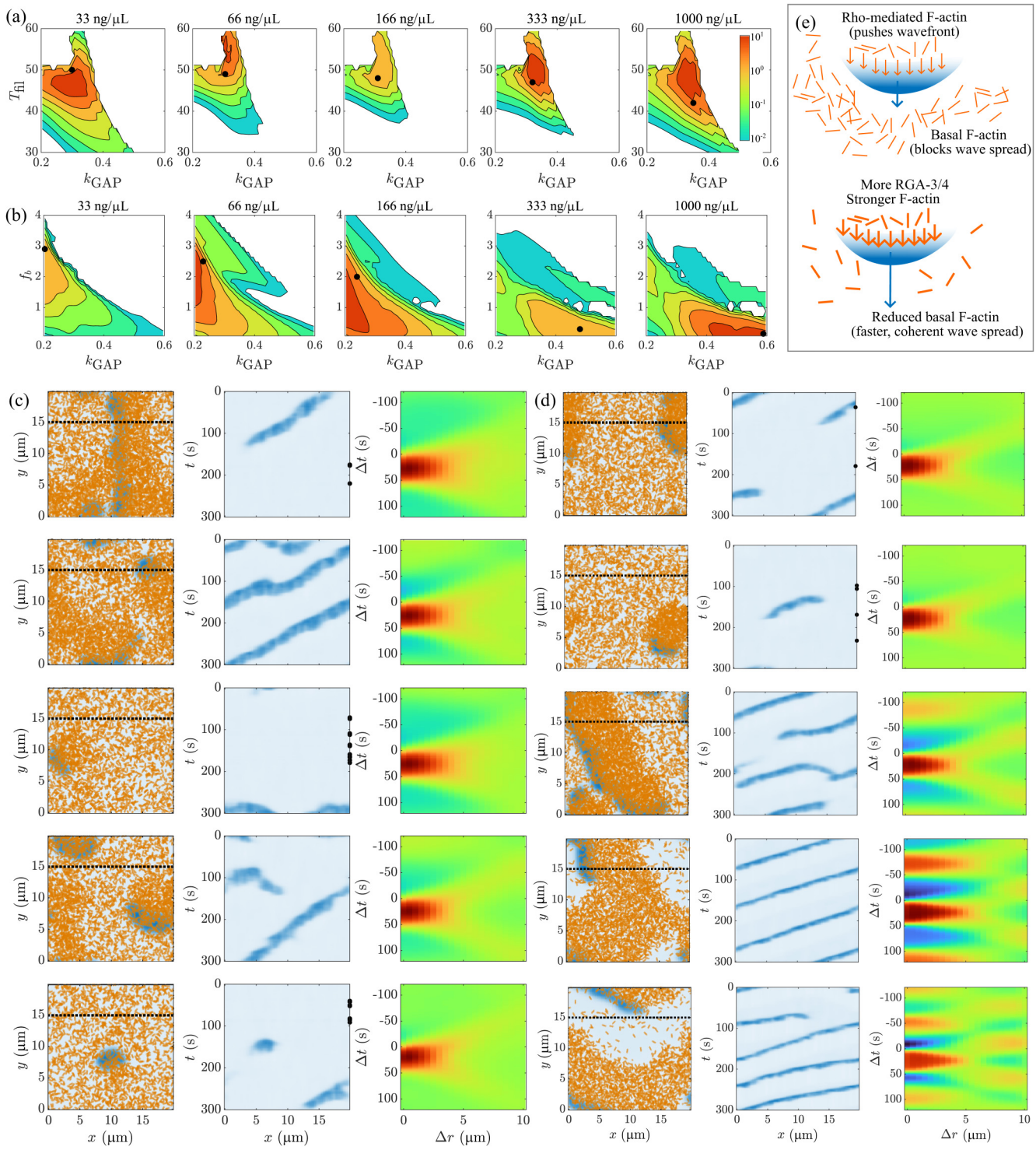


FIG. 8. To reproduce data, basal F-actin assembly must be reduced as GAP strength increases. (a), (b) Average estimated posteriors for each coexpression level when exactly one parameter is varied in addition to k_{GAP}. (a) Varying T_{fil} alone in response to k_{GAP} gives a narrow band of plausible parameter values. Black dots show samples in panel (c), taken for a specific set of fixed parameters that maximize the posterior while also respecting increases in k_{GAP} [see Eq. (A4) and Fig. S11 of the Supplemental Material [63]]. (b) Varying the basal assembly rate f_b gives a rich distribution of probability that shifts appropriately as expression levels increase. Black dots show samples in panel (d). (c) Sample dynamics for the varying lifetime model [GAP strength increases from top to bottom; fixed parameters: $\ell_{\text{max}} = 0.50$, $f_b = 0.096\ell_{\text{max}}(2\ell_{\text{max}} + T_{\text{fil}})$, and $f_b = 1.1\ell_{\text{max}}(2\ell_{\text{max}} + T_{\text{fil}})$]. (d) Sample dynamics for the varying basal nucleation model (GAP strength increases from top to bottom; fixed parameters: $T_{\text{fil}} = 44$, $\ell_{\text{max}} = 0.50$, and $f_\rho = 78$). (e) Mechanistic diagram explaining the role of Rho-mediated F-actin (assembles behind Rho wave front and sets direction of propagation) and basal F-actin (assembles in the bulk, blocking wave spread). Waves spread faster and become more coherent if basal levels decrease when RhoGAP levels increase.

to the errors in the cross-correlation function, as it penalizes some aspects of the error more than others. By leveraging the amortized likelihood function, we used 10^5 samples of the model to predict the posterior for $50^5 = 3.1 \times 10^8$ samples, which gives each sample an effective power of over 1000. Still, 10^5 samples is quite large, and it will be important to implement some of the recent "cost-aware" improvements to neural SBI as our models move beyond just Rho and F-actin [38,44,77–79].

When we attempted to apply NLRE directly to experimental data, we found that even more regularization was required to avoid overconfident posteriors that often arise from model misspecification [52,80]. Recent work has addressed this issue by developing a model for the misspecification, potentially based on a small set of measurements where the true parameters are known, and then inverting the error model prior to inferring the likelihood [51,53]. This approach could be promising here given that kinetic data are increasingly available in organisms like *C. elegans* [11,62]. Of course, developing a better model of the dynamics that reduces misspecification (perhaps by incorporating more molecular-scale features) is a direction to pursue simultaneously.

While we always visualized the model dynamics in concert with the cross-correlation function (e.g., Fig. 8), only a compressed version of the latter was actually used to perform inference, and we found that the number of included modes influenced the results. This raises the question: What information is lost when reducing complex dynamics to summary statistics? For instance, while experimental measurements are subject to noise, there is also randomness in the true dynamics that is averaged over when computing the cross-correlation function. A helpful next step would therefore be to optimize the representation of the data, for instance, using feature extraction with neural networks [60]. Though not strictly necessary, it would be ideal if the compressed representation of the data were still interpretable. A straightforward way to ensure this is to augment the cross-correlation function with additional summary statistics designed to detect other kinds of spatiotemporal phenomena, similar to how the cross-correlation function is optimal for detecting traveling waves,

On the biology side, our main contribution was to show how increasing RhoGAP levels could give rise to waves of increasing coherence and frequency in activated frog eggs. Similar to previous work [24], we first showed that increasing k_{GAP} (the model's GAP strength parameter) can transition the dynamics from a high uniform steady state of Rho (with small fluctuations) to waves with oscillatory cross-correlation functions (Fig. 7 and Fig. S9 of the Supplemental Material [63]). Unlike in the data, in the model increases in k_{GAP} only suppress this peak excitation, as stronger filaments make waves less coherent and ultimately nonexistent (Fig. 7 and Fig. S9 of the Supplemental Material [63]). Previous work [24] interpreted the disagreement with the experimental data as evidence for saturation of k_{GAP} with increasing amounts of microinjected RNA, which could occur if expression levels are limited by translation, or by crowding along actin filaments [see Fig. 8(g) in Ref. [24], where the GAP strength parameter is limited to small values for frog data]. In this interpretation, the full spectrum of dynamics is explained by a noisy transition from a uniform steady state to a bistable state with propagating waves and saturated k_{GAP}. In contrast, our NLRE analysis found that k_{GAP} need not saturate, so long as increases in k_{GAP} are paired with *systematic* decreases in the basal nucleation rate. In this interpretation, coherent waves are sustained by reducing the amount of inhibition (F-actin) *in the bulk*, which allows waves to propagate without disruption [Fig. 8(e)]. We offered several hypotheses for how these dynamics could arise molecularly, including the possibilities that RhoGAP suppresses fluctuations in Rho density, or that it directly inhibits formins or other GTPases. In this way, our combination of modeling and neural inference yielded specific, testable predictions that were overlooked previously.

Since our inference procedure is now high throughput, there is no limit to the amount of experimental data we can analyze—including knockdown experiments that examine how accessory proteins such as anillin and septin [81,82], or anterior PARs [83,84] impact GTPase activity by modulating F-actin assembly dynamics. Inference is only useful if the model includes the data, and so learning the full scope of dynamics requires accounting for myosin-mediated contractility, which is also downstream of Rho [26,85,86]. When paired with neural SBI, our model can determine how a balance of reaction, diffusion, and contractility shapes patterns of Rho GTPases, and how each of these is in turn modulated by the key cytoskeletal molecular players.

On a broader level, our use of neural SBI to "learn" the relevant mathematical model and parameters provides a roadmap for future work across other self-organizing systems. Whether in cells [87,88], tissues [89,90], or ecosystems [91], connecting small and large scales requires a combination of versatile mathematical models and tools for parameter inference. Neural techniques can identify both relevant models and parameters faster than ever before, but it remains unclear how best to deploy them with experimental data. Our work is a step toward answering this question.

ACKNOWLEDGMENTS

We thank G. Louppe for providing details on how to validate the likelihood ratio, B. Bement and A. Michaud for providing the data on activated frog oocytes, and A. Mogilner and A. Lindsay for feedback on the manuscript. O.M. thanks D. Schiavazzi for suggesting neural methods for SBI. Simulations were carried out on the Research Computing Center Midway cluster at the University of Chicago and the Center for Research Computing cluster at the University of Notre Dame. O.M. acknowledges partial support from the Physics Frontier Center for Living Systems (CLS) Spark Fellowship, funded by the National Science Foundation (NSF PHY-2317138); A.R.D. and E.M. acknowledge support from the CLS as well as the "From Cells to Organs with AI" Initiative at the University of Chicago.

The authors declare that they have no conflict of interest.

DATA AVAILABILITY

Code for simulations, data analysis, and inference, as well as the cross-correlation data from experiments, is available on GitHub [92].

APPENDIX

1. Data analysis

Similar to previous work [30], we perform preprocessing to better visualize experimental data and extract cross-correlation. In all movies, we first extract a square domain in the middle of the oocyte/embryo. For starfish and frog data, we perform difference subtraction, subtracting the last frame from the others, to remove static Rho signal [24,27]. For starfish and *C. elegans*, we also correct for photobleaching by setting the mean intensity of each frame to be equal to the global mean. For *C. elegans* (where the signal-to-noise ratio is significantly lower), we smooth the signal in time by applying a moving average filter with a width of 20 frames (12 s). For all movies, we remove high-frequency noise in space by applying a Fourier filter to retain the 20 (for *C. elegans*), 50 (starfish), or 100 (for frogs) lowest-frequency modes. In all cases, cross-correlation functions are computed on the filtered data (although the cross-correlation functions on the raw data are similar).

2. Mathematical model

a. Governing equations

The dynamics of Rho in our model are governed by the two-dimensional PDE:

$$\partial_t \rho = D_\rho \nabla^2 \rho + k_\mathrm{b} + k_\mathrm{ac}\left(\frac{\rho^3}{\rho^3 + K_\mathrm{sat}}\right) - \left(k_\mathrm{inac}^{(0)} + k_\mathrm{GAP} f\right)\rho, \tag{A1a}$$

with periodic boundary conditions. Equation (A1a) accounts for basal Rho activation and autocatalytic excitation (through the cubic Hill function), as well as basal Rho inactivation and enhanced inactivation by an inhibitor f, which represents colocalized F-actin/RhoGAP [23,24]. As justified in Ref. [30], our model only considers the active form of Rho, as experimental evidence [82,93] has suggested that inactive Rho is not significantly depleted.

At each point in space, the concentration of Rho sets a nucleation rate of individual actin filaments,

$$q_\mathrm{nuc}(\boldsymbol{x}) = q_b + q_\rho\left(\frac{\rho(\boldsymbol{x})}{\rho_\mathrm{max}}\right)^2, \tag{A1b}$$

where ρ_max is the maximum steady state of Rho in Eq. (A1a) when $f = 0$. Upon nucleation, each filament is assigned a random direction of growth on the unit circle, along which it polymerizes with rate v_p up to length ℓ_max [see kymograph in Fig. 1(d)]. Once the filament reaches the maximum length, it remains fixed in space for time T_fil, after which it depolymerizes from the older end with rate v_dp. The continuum field $f(\boldsymbol{x})$ is then defined by convolving the discrete filament centerline positions with a Gaussian point spread function (Fig. S1 of the Supplemental Material [63]).

As the nucleation rates are not easily measurable experimentally, we convert these quantities into approximate densities of basal (f_b) and Rho-mediated (f_ρ) F-actin per domain area by setting $f_b = q_b \ell_\mathrm{max}/k_\mathrm{diss}$ and $f_\rho = q_\rho \ell_\mathrm{max}/k_\mathrm{diss}$, where the effective filament turnover rate is defined as

$$k_\mathrm{diss} = \left(2\frac{\ell_\mathrm{max}}{v_p} + T_\mathrm{fil}\right)^{-1}, \tag{A2}$$

and the factor of 2 accounts for growth and shrinkage. For basal nucleation, f_b is a good measure of the filament density produced independently of Rho, while for Rho-mediated nucleation, f_ρ expresses the filament density that would result if $\rho = \rho_\mathrm{max}$ uniformly.

b. Parametrization

We previously showed [30] that the dynamics of Rho must be in a bistable regime where Rho excitations can be initiated via diffusive flux from neighboring regions (e.g., traveling waves), but not automatically *de novo* in the absence of F-actin. We use these observations to constrain the parameters $k_\mathrm{inac}^{(0)}$, k_b, k_ac, and K_sat in Eq. (A1a). Fixing k_b, k_ac, and K_sat, we choose $k_\mathrm{inac}^{(0)} = 0.7$, so that in the absence of F-actin ($f = 0$), the total inhibition $k_\mathrm{inac} = k_\mathrm{inac}^{(0)}$ is such that there are two stable steady states of Rho [Fig. 1(c)]. The diffusivity of Rho ($D_\rho = 0.1$ μm^2/s) is fixed in accordance with experimental data [23].

In previous work [30], we demonstrated that the dynamics of F-actin are effectively governed by two parameters: one having to do with the spatial extent of growth and the other the timescale of filament turnover. We therefore reduce the dynamics to one spatial and one temporal parameter by fixing $v_p = v_\mathrm{dp} = 1$ μm/s (which corresponds to polymerization speeds in a variety of cell types [62,94]), and scanning over k_GAP, ℓ_max, T_fil, f_b, and f_ρ. Table S1 of the Supplemental Material [63] summarizes the fixed and varied parameters.

c. Dynamic steady state and statistics

When there are two steady states of Rho in the absence of F-actin, a uniformly high F-actin concentration could extinguish *all* of the excitation, causing Rho to sit at the lower steady state everywhere. This prevents a quasisteady state, since an absence of excitation at one time implies an absence of excitation for all time. To induce a dynamic steady state, our approach is to define a threshold Rho concentration, ρ_c, equal to the arithmetic mean of the saddle point and higher steady state when $k_\mathrm{inac}^{(0)} = 0.7$ [arrow in Fig. 1(c)]. When the concentration is everywhere below this level, we instantaneously bump $k_\mathrm{inac}^{(0)} = 0.4$ everywhere [black circle in Fig. 1(c)], until at least one point of Rho activity exceeds the threshold level [in kymographs of Rho activity, e.g., Fig. 1(e), black dots denote regions of time when $k_\mathrm{inac}^{(0)} = 0.4$]. Starting from an initial condition where Rho is at the lower steady state everywhere, we track statistics over the longest self-sustaining simulated interval [the green regions in Fig. 1(e)].

d. Prior on varied parameters

To approximate the prior $p(\boldsymbol{\theta})$, we train a discriminative classifier to identify the region in five-dimensional (k_GAP, ℓ_max, T_fil, f_b, f_ρ) parameter space that is able to sustain excitation (by sampling uniformly within the ranges in Table S1 of the Supplemental Material [63] and recording whether two out of two samples of the model are able to sustain excitations for 120 s or longer). Sampling from the prior then amounts to sampling uniformly from the five-dimensional space (k_GAP, ℓ_max, T_fil, f_b, f_ρ), then eliminating any parameters with a prior value of zero. The prior for the remaining

parameter sets is uniform. As the boundary of being able to sustain excitation is complex and not fully resolved by running only two trials per parameter set, our prior is sometimes inaccurate (classifier has 95% validation accuracy). In general, the marginal posterior distributions do not depend significantly on the threshold used to assign a nonzero prior. However, in cases where we are interested in visualizing high-likelihood simulations (e.g., Fig. 6), we reduce false positives by only considering parameters where the score is 0.75 or higher (75% chance of being able to sustain excitation).

e. Compressed cross-correlation functions

As discussed in Sec. II A, pointwise trajectories of Rho and F-actin are of little use, since reproducing them often amounts to reproducing noise. One work-around to this is to use a black-box-style approach, where the pointwise trajectories $\boldsymbol{x}$ are passed to an autoencoder that is trained simultaneously with the likelihood estimator, so that statistics are learned naturally [39,95–97]. In our case, this presents some challenges because the data are complex and the learned statistics are typically not interpretable, making the output hard to analyze. Our approach is instead to leverage the cross-correlation function (Sec. II A), which we already know is a good representation of the full spatiotemporal dynamics. However, the two-dimensional matrix of cross-correlation functions [Fig. 1(b′)] can be arbitrarily large, depending on the spatial and temporal resolution of the experimental data, and so we still require a means of reducing the problem dimensionality.

To compress the cross-correlation while still retaining as much information as possible, we interpolate each function at $\Delta t = -120, -118, \ldots, 120$ s and $\Delta r = 0, 0.5, 1, \ldots, 10$ μm (so that the cross-correlation function is a 121×21 matrix), and then use PCA to convert each cross-correlation function to principal components (Fig. S2 of the Supplemental Material [63]). To compute the principal components, we first build a matrix of many cross-correlation functions (where each one is stacked as a single column vector)

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{R}_{\rho f}^{(1)} & \boldsymbol{R}_{\rho f}^{(2)} & \cdots \end{pmatrix}.$$

The singular value decomposition (SVD) $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$ gives a list of "components" (the columns of $\boldsymbol{U}$) and scores $\boldsymbol{\Sigma}\boldsymbol{V}^T$. Compressing the cross-correlation into a subset of modes then amounts to setting $\boldsymbol{\gamma}^{(i)} = \boldsymbol{U}^T\boldsymbol{R}_{\rho f}^{(i)}$ (to obtain the scores of a single cross-correlation), extracting the first c components of $\boldsymbol{\gamma}$ (denote this $c \times 1$ vector as $\boldsymbol{\gamma}_c$). If necessary for visualization, the image can then be decompressed by multiplying $\boldsymbol{U}_c\boldsymbol{\gamma}_c := \widetilde{\boldsymbol{R}}_{\rho f}$, where $\boldsymbol{U}_c$ is the matrix containing the first c columns of $\boldsymbol{U}$.

3. Neural likelihood ratio estimation

a. Training the classification neural network

To train the neural likelihood ratio estimator in our context, we perform the following set of steps:

(1) Sample uniformly over parameter space to obtain a set of parameters $\boldsymbol{\theta}_i$ and simulated datasets $\boldsymbol{x}_i$ [Fig. 2(a)].

(2) Remove any parameter sets from the list where at least one of the runs does not sustain excitation. The number of remaining parameter data pairs is denoted by N_p.

(3) For parameter set $\boldsymbol{\theta}_i$, input to the classifier $(\boldsymbol{x}_i, \boldsymbol{\theta}_i, y = 1)$ and $(\boldsymbol{x}_i, \boldsymbol{\theta}_j, y = 0)$ for $j \neq i$ [Fig. 2(b)]. Repeat this for all N_p parameter sets, so that there are a total of $2N_p$ observations passed to the classifier.

(4) Train the classifier. We use the `ClassificationNeuralNetwork` object in Matlab R2024b with standardized predictors and an iteration limit of 15 000. As noted in the main text, we vary the size of the hidden layer and the regularization parameter in the cross-entropy loss function, which is minimized by the BFGS algorithm.

In step 1, we choose a total of 2×10^5 parameter sets, where for each parameter set we obtain two cross-correlation functions. Thus, the total initial number of parameter-data pairs is 4×10^5. Removing parameter sets where excitation is not sustained for both simulations gives $N_p = 1.4 \times 10^5$ pairs.

b. Using the classification neural network for inference

Given new data $\boldsymbol{x}_n$, we perform the following set of steps to estimate the posterior $p(\boldsymbol{\theta}|\boldsymbol{x}_n)$ (assuming all five parameters can vary):

(1) Sample 50^5 parameter sets by uniformly gridding five-dimensional parameter space over the ranges given in Table S1 of the Supplemental Material [63].

(2) Use the trained classifier to predict whether these parameters will sustain excitation. Assign prior $p(\boldsymbol{\theta}) = 0$ and posterior $p(\boldsymbol{\theta}|\boldsymbol{x}_n) = 0$ for any parameters predicted not to sustain excitation.

(3) For each parameter set $\boldsymbol{\theta}_i$ with $p(\boldsymbol{\theta}_i) \neq 0$, input $(\boldsymbol{x}_n, \boldsymbol{\theta}_i)$ to the classifier.

(4) The output of the classifier is the probability $y = 1$, which gives $\hat{r}(\boldsymbol{x}_n, \boldsymbol{\theta}_i)$ by Eq. (3) [Fig. 2(c)]. Assign the estimated posterior $\hat{p}(\boldsymbol{\theta}_i|\boldsymbol{x}_n) = \hat{r}(\boldsymbol{x}_n, \boldsymbol{\theta}_i)$ for these parameter sets (the constant prior is normalized out in the next step).

(5) To compute the marginal posterior in two dimensions (e.g., over the first two parameters), set

$$\bar{p}(\theta_1, \theta_2) = \frac{1}{50^3} \sum_i \hat{p}(\boldsymbol{\theta}_i|\boldsymbol{x}_n),$$

where the sum is over all $\boldsymbol{\theta}_i$ with first two components equal to θ_1 and θ_2.

c. Two-dimensional posteriors for frog data

To identify a potential variable θ_v that changes with k_{GAP}, with the other three variables $\boldsymbol{\theta}_f$ fixed, we scan over $\boldsymbol{\theta}_f$ by sampling a $50 \times 50 \times 50$ grid in the regions of parameter space given in Table S1 of the Supplemental Material [63]. For each $\boldsymbol{\theta}_f$, we compile a two-dimensional posterior $\hat{p}(k_{\mathrm{GAP}}, \theta_v, \boldsymbol{\theta}_f)$ by scanning over 50×50 values of $(k_{\mathrm{GAP}}, \theta_v)$. We say that the fixed parameters $\boldsymbol{\theta}_f$ are allowable if

$$\hat{r}_i^{(m)} := \max_{k_{\mathrm{GAP}}, \theta_v} \hat{r}\left(\boldsymbol{x}_i | k_{\mathrm{GAP}}, \theta_v, \boldsymbol{\theta}_f\right) \geqslant 1 \quad \forall i, \tag{A3a}$$

$$\bar{p}_{i+1} \geqslant \bar{p}_i \quad i = 1, \ldots, D-1,$$

$$\text{where} \quad \bar{p}_i(\boldsymbol{\theta}_f) = \frac{\int_{k_{\mathrm{GAP}}} \int_{\theta_v} k_{\mathrm{GAP}} \hat{p}(k_{\mathrm{GAP}}, \theta_v, \boldsymbol{\theta}_f | \boldsymbol{x}_i)}{\int_{k_{\mathrm{GAP}}} \int_{\theta_v} \hat{p}(k_{\mathrm{GAP}}, \theta_v, \boldsymbol{\theta}_f | \boldsymbol{x}_i)}, \tag{A3b}$$

where i indexes the datasets (RhoGAP coexpression increases with i). Equation (A3a) says that the maximum likelihood-to-evidence must be larger than 1 for at least one parameter set $(k_{\mathrm{GAP}}, \theta_v)$, with $\boldsymbol{\theta}_f$ fixed, and Eq. (A3b) says that the first moment of k_{GAP} must increase with RhoGAP coexpression (indexed by i). The posterior we show for a given $(k_{\mathrm{GAP}}, \theta_v)$ is the total posterior $\hat{p}(k_{\mathrm{GAP}}, \theta_v, \boldsymbol{\theta}_f)$ summed over all $\boldsymbol{\theta}_f$ satisfying Eq. (A3), divided by the number of such $\boldsymbol{\theta}_f$. This amounts to marginalizing only over fixed parameters that could reasonably fit all five datasets.

To select a particular fixed parameter set $\boldsymbol{\theta}_f$ for plotting [Figs. 8(c) and 8(d)], we identify fixed parameters that always increase the first moment $\bar{p}$ by a substantial amount (quantified by the *minimum* change in first moment), and have a high probability for all datasets (quantified by the *minimum* peak probability). The parameter set is therefore chosen by maximizing the score function

$$\mathcal{S}(\boldsymbol{\theta}_f) = \min_{1\leqslant i\leqslant D} \hat{r}_i^{(m)}(\boldsymbol{\theta}_f) + 5 \min_{1\leqslant i\leqslant D-1} \bar{p}_{i+1}(\boldsymbol{\theta}_f) - \bar{p}_i(\boldsymbol{\theta}_f), \tag{A4}$$

where the factor of 5 is chosen empirically to balance the two terms.

d. Bayes factor analysis

In Fig. S10 of the Supplemental Material [63], we use the Bayes factor, which measures the relative probabilities of two models, to compare the model where the filament lifetime T_{fil} varies with the GAP strength to the model where the basal nucleaton rate q_b varies with the GAP strength (Fig. 8). To compute a Bayes factor, we compute the total evidence for varying parameter θ_v by integrating over fixed parameter sets $\boldsymbol{\theta}_f$ that satisfy Eq. (A3),

$$P_i(\theta_v) = \int \hat{p}(k_{\mathrm{GAP}}, \theta_v, \boldsymbol{\theta}_f | \boldsymbol{x}_i)\delta(\boldsymbol{\theta}_f)\, dk_{\mathrm{GAP}} d\theta_v d\boldsymbol{\theta}_f, \tag{A5}$$

where $\delta(\boldsymbol{\theta}_f) = 1$ if Eq. (A3) is satisfied and zero otherwise. Substituting $\theta_v = T_{\mathrm{fil}}$, q_b into Eq. (A5) and taking the ratio gives $B_i = P_i(q_b)/P_i(T_{\mathrm{fil}})$, which is the Bayes factor comparing the model with varying T_{fil} to that with varying q_b. A number greater than 1 favors the model in the numerator, and a number less than 1 favors the model in the denominator.

[1] J. Halatek and E. Frey, Rethinking pattern formation in reaction–diffusion systems, Nat. Phys. **14**, 507 (2018).

[2] J. Halatek, F. Brauns, and E. Frey, Self-organization principles of intracellular pattern formation, Philos. Trans. R. Soc. B **373**, 20170107 (2018).

[3] P. Caldarelli, A. Chamolly, A. Villedieu, O. Alegria-Prévot, C. Phan, J. Gros, and F. Corson, Self-organized tissue mechanics underlie embryonic regulation, Nature (London) **633**, 887 (2024).

[4] E. Karsenti, Self-organization in cell biology: A brief history, Nat. Rev. Mol. Cell Biol. **9**, 255 (2008).

[5] R. Wedlich-Söldner and T. Betz, Self-organization: The fundament of cell biology, Philos. Trans. R. Soc. B: Biol. Sci. **373**, 20170103 (2018).

[6] M. F. Staddon, E. M. Munro, and S. Banerjee, Pulsatile contractions and pattern formation in excitable actomyosin cortex, PLoS Comput. Biol. **18**, e1009981 (2022).

[7] A. Mogilner, J. Allard, and R. Wollman, Cell polarity: Quantitative modeling as a tool in cell biology, Science **336**, 175 (2012).

[8] Q. Ni and G. A. Papoian, Turnover versus treadmilling in actin network assembly and remodeling, Cytoskeleton **76**, 562 (2019).

[9] S. Bhattacharya, T. Banerjee, Y. Miao, H. Zhan, P. N. Devreotes, and P. A. Iglesias, Traveling and standing waves mediate pattern formation in cellular protrusions, Sci. Adv. **6**, eaay7682 (2020).

[10] I. Hecht, D. A. Kessler, and H. Levine, Transient localized patterns in noise-driven reaction-diffusion systems, Phys. Rev. Lett. **104**, 158301 (2010).

[11] F. B. Robin, W. M. McFadden, B. Yao, and E. M. Munro, Single-molecule analysis of cell surface dynamics in *Caenorhabditis elegans* embryos, Nat. Meth. **11**, 677 (2014).

[12] S. Yamashiro, S. Tanaka, L. M. McMillen, D. Taniguchi, D. Vavylonis, and N. Watanabe, Myosin-dependent actin stabilization as revealed by single-molecule imaging of actin turnover, Mol. Biol. Cell **29**, 1941 (2018).

[13] S. Yamashiro, H. Mizuno, M. B. Smith, G. L. Ryan, T. Kiuchi, D. Vavylonis, and N. Watanabe, New single-molecule speckle microscopy reveals modification of the retrograde actin flow by focal adhesions at nanometer scales. Mol. Biol. Cell **25**, 1010 (2014).

[14] E. D. White, Y. Li, R. Kadzik, Y. Qiu, A. R. Dinner, and E. M. Munro, Microscopic control of cortical flows in polarized *C. elegans* zygotes, bioRxiv (2025).

[15] S. Glenn, J. Cisewski-Kehe, J. Zhu, and W. M. Bement, Tracking temporal evolution of topological features in image data, arXiv:2508.17530.

[16] K. Cranmer, J. Brehmer, and G. Louppe, The frontier of simulation-based inference, Proc. Natl. Acad. Sci. USA **117**, 30055 (2020).

[17] A. Michaud, Z. T. Swider, J. Landino, M. Leda, A. L. Miller, G. von Dassow, A. B. Goryachev, and W. M. Bement, Cortical excitability and cell division, Curr. Biol. **31**, R553 (2021).

[18] M. Graessl, J. Koch, A. Calderon, D. Kamps, S. Banerjee, T. Mazel, N. Schulze, J. K. Jungkurth, R. Patwardhan, D. Solouk, N. Hampe, B. Hoffmann, L. Dehmelt, and P. Nalbant, An excitable Rho GTPase signaling network generates dynamic subcellular contraction patterns, J. Cell Biol. **216**, 4271 (2017).

[19] A. Hall, Rho GTPases and the actin cytoskeleton, Science **279**, 509 (1998).

[20] W. M. Bement, A. B. Goryachev, A. L. Miller, and G. von Dassow, Patterning of the cell cortex by Rho GTPases, Nat. Rev. Mol. Cell Biol. **25**, 290 (2024).

[21] E. M. Ozbudak, A. Becskei, and A. van Oudenaarden, A system of counteracting feedback loops regulates Cdc42p activity during spontaneous cell polarization, Dev. Cell **9**, 565 (2005).

[22] S. Y. S. Fung, S. Xiao, Y. Bao, M. Graham, M. Su, X. Liu, J. Bewersdorf, and M. Wu, Mitotic Cdc42 waves encode PI (3, 4)

P_2 signaling and Golgi morphological state to control spindle scaling, Sci. Adv. **12**, eaec7705 (2026).

[23] J. B. Michaux, F. B. Robin, W. M. McFadden, and E. M. Munro, Excitable RhoA dynamics drive pulsed contractions in the early *C. elegans* embryo, J. Cell Biol. **217**, 4230 (2018).

[24] A. Michaud, M. Leda, Z. T. Swider, S. Kim, J. He, J. Landino, J. R. Valley, J. Huisken, A. B. Goryachev, G. von Dassow, and W. M. Bement, A versatile cortical pattern-forming circuit based on Rho, F-actin, Ect2, and RGA-3/4, J. Cell Biol. **221**, e202203017 (2022).

[25] M. Nishikawa, S. R. Naganathan, F. Jülicher, and S. W. Grill, Controlling contractile instabilities in the actomyosin cortex, eLife **6**, e19595 (2017).

[26] B. Yao, S. Donoughe, J. Michaux, and E. Munro, Modulating RhoA effectors induces transitions to oscillatory and more wavelike RhoA dynamics in *Caenorhabditis elegans* zygotes, Mol. Biol. Cell **33**, ar58 (2022).

[27] W. M. Bement, M. Leda, A. M. Moe, A. M. Kita, M. E. Larson, A. E. Golding, C. Pfeuti, K.-C. Su, A. L. Miller, A. B. Goryachev, and G. von Dassow, Activator–inhibitor coupling between Rho signalling and actin assembly makes the cell cortex an excitable medium, Nat. Cell Biol. **17**, 1471 (2015).

[28] S. Chen, D. S. Seara, A. Michaud, S. Kim, W. M. Bement, and M. P. Murrell, Energy partitioning in the cell cortex, Nat. Phys. **20**, 1824 (2024).

[29] J. Li, C.-W. J. Liu, M. Szurek, and N. Fakhri, Measuring irreversibility from learned representations of biological patterns, PRX Life **2**, 033013 (2024).

[30] O. Maxian, A. R. Dinner, and E. Munro, Actin network heterogeneity tunes activator–inhibitor dynamics at the cell cortex, Proc. Natl. Acad. Sci. USA **122**, e2520485122 (2025).

[31] M. Sunnåker, A. G. Busetto, E. Numminen, J. Corander, M. Foll, and C. Dessimoz, Approximate Bayesian computation, PLoS Comput. Biol. **9**, e1002803 (2013).

[32] K. Csilléry, M. G. B. Blum, O. E. Gaggiotti, and O. François, Approximate Bayesian computation (ABC) in practice, Trends Ecol. Evol. **25**, 410 (2010).

[33] D. Prangle, Lazy ABC, Stat. Comput. **26**, 171 (2016).

[34] J. Hermans, V. Begy, and G. Louppe, Likelihood-free MCMC with amortized approximate ratio estimators, in *International Conference on Machine Learning* (PMLR, Cambridge, MA, 2020), pp. 4239–4248.

[35] M. A. Beaumont, W. Zhang, and D. J. Balding, Approximate Bayesian computation in population genetics, Genetics **162**, 2025 (2002).

[36] K. Cranmer, J. Pavez, and G. Louppe, Approximating likelihood ratios with calibrated discriminative classifiers, arXiv:1506.02169.

[37] M. Gloeckler, M. Deistler, C. D. Weilbach, F. Wood, and J. H. Macke, All-in-one simulation-based inference, in *International Conference on Machine Learning* (PMLR, Cambridge, MA, 2024), pp. 15735–15766.

[38] A. Bharti, D. Huang, S. Kaski, and F.-X. Briol, Cost-aware simulation-based inference, in *International Conference on Artificial Intelligence and Statistics* (PMLR, Cambridge, MA, 2025), pp. 28–36.

[39] A. Zammit-Mangion, M. Sainsbury-Dale, and R. Huser, Neural methods for amortized inference, Annu. Rev. Stat. Appl. **12**, 311 (2025).

[40] A. Delaunoy, J. Hermans, F. Rozet, A. Wehenkel, and G. Louppe, Towards reliable simulation-based inference with balanced neural ratio estimation, Adv. Neural Inf. Process. Syst. **35**, 20025 (2022).

[41] V. Zaballa and E. Hui, Approximation of intractable likelihood functions in systems biology via normalizing flows, in *NeurIPS 2023 Generative A. I. and Biology (GenBio) Workshop* (Neural Information Processing Systems Foundation, Inc. (NeurIPS), San Diego, CA, 2023).

[42] C.-W. Huang, D. Krueger, A. Lacoste, and A. Courville, Neural autoregressive flows, in *International Conference on Machine Learning* (PMLR, Cambridge, MA, 2018), pp. 2078–2087.

[43] G. Papamakarios, T. Pavlakou, and I. Murray, Masked autoregressive flow for density estimation, in *Advances in Neural Information Processing Systems* (Neural Information Processing Systems Foundation, Inc. (NeurIPS), San Diego, CA, 2017), Vol. 30.

[44] Y. Wang, C. Chi, and A. R. Dinner, Mitigating mode collapse in normalizing flows by annealing with an adaptive schedule: Application to parameter estimation, J. Comput. Phys. **564**, 115115 (2026).

[45] J. Rudi, J. Bessac, and A. Lenzi, Parameter estimation with dense and convolutional neural networks applied to the FitzHugh–Nagumo ODE, in *Mathematical and Scientific Machine Learning* (PMLR, Cambridge, MA, 2022), pp. 781–808.

[46] J. Dyer, P. Cannon, J. D. Farmer, and S. M. Schmon, Black-box Bayesian inference for agent-based models, J. Econ. Dyn. Control **161**, 104827 (2024).

[47] M. Vasist, F. Rozet, O. Absil, P. Mollière, E. Nasedkin, and G. Louppe, Neural posterior estimation for exoplanetary atmospheric retrieval, Astron. Astrophys. **672**, A147 (2023).

[48] B. Tucci and F. Schmidt, EFTofLSS meets simulation-based inference: σ_8 from biased tracers, J. Cosmol. Astropart. Phys. 05 (2024) 063.

[49] A. Wehenkel, J. Behrmann, A. C. Miller, G. Sapiro, O. Sener, M. Cuturi, and J.-H. Jacobsen, Simulation-based inference for cardiovascular models, arXiv:2307.13918.

[50] A. Chakraborty, D. J. Nott, C. C. Drovandi, D. T. Frazier, and S. A. Sisson, Modularized Bayesian analyses and cutting feedback in likelihood-free inference, Stat. Comput. **33**, 33 (2023).

[51] A. Wehenkel, J. L. Gamella, O. Sener, J. Behrmann, G. Sapiro, J.-H. Jacobsen, and M. Cuturi, Addressing misspecification in simulation-based inference through data-driven calibration, in *Proceedings of the International Conference on Machine Learning* (ICML, Vancouver, CA, 2025).

[52] P. Cannon, D. Ward, and S. M. Schmon, Investigating the impact of model misspecification in neural simulation-based inference, arXiv:2209.01845.

[53] D. Ward, P. Cannon, M. Beaumont, M. Fasiolo, and S. Schmon, Robust neural posterior estimation and statistical model criticism, Adv. Neural Inf. Process. Syst. **35**, 33845 (2022).

[54] L. Dingeldein, D. Silva-Sánchez, L. Evans, E. D'Imprima, N. Grigorieff, R. Covino, and P. Cossio, Amortized template matching of molecular conformations from cryoelectron microscopy images using simulation-based inference, Proc. Natl. Acad. Sci. USA **122**, e2420158122 (2025).

[55] L. Dingeldein, P. Cossio, and R. Covino, Simulation-based inference of single-molecule experiments, Curr. Opin. Struct. Biol. **91**, 102988 (2025).

[56] E. Touron, P. Rodrigues, J. Arbel, N. Varoquaux, and M. Arbel, Simulation-based inference of yeast centromeres, in *NeurIPS 2025-39th Conference on Neural Information Processing Systems Workshop: The 3rd Workshop on Imageomics: Discovering Biological Knowledge from Images Using AI* (Neural Information Processing Systems Foundation, Inc. (NeurIPS), San Diego, CA, 2025), pp. 1–13.

[57] J. Boelts, P. Harth, R. Gao, D. Udvary, F. Yáñez, D. Baum, H.-C. Hege, M. Oberlaender, J. H. Macke, and E. C. Robinson, Simulation-based inference for efficient identification of generative models in computational connectomics, PLoS Comput. Biol. **19**, e1011406 (2023).

[58] H. Hishinuma, H. Takigawa-Imamura, and T. Miura, Data-driven discovery and parameter estimation of mathematical models in biological pattern formation, PLoS Comput. Biol. **21**, e1012689 (2025).

[59] L. Dingeldein, P. Cossio, and R. Covino, Simulation-based inference of single-molecule force spectroscopy, Mach. Learn.: Sci. Technol. **4**, 025009 (2023).

[60] J. C. J. Heyn, M. A. Juanatey, M. Falcke, J. O. Rädler, and M. Kordrostami, Cell-mechanical parameter estimation from 1D cell trajectories using simulation-based inference, PLoS One **20**, e0310669 (2025).

[61] J. Arruda, E. Alamoudi, R. Mueller, M. Vaisband, R. Molkenbur, J. Merrin, E. Kiermaier, and J. Hasenauer, Simulation-based inference of cell migration dynamics in complex spatial environments, NPJ Syst. Biol. Appl. **12**, 20 (2026).

[62] V. Costache, S. P. Garcia, C. N. Plancke, J. Li, S. Begnaud, S. K. Suman, A. Reymann, T. Kim, and F. B. Robin, Rapid assembly of a polar network architecture drives efficient actomyosin contractility, Cell Rep. **39**, 110868 (2022).

[63] See Supplemental Material at http://link.aps.org/supplemental/10.1103/5l5r-fvjd for additional figures and tables.

[64] M. E. Werner, D. D. Ray, C. Breen, M. F. Staddon, F. Jug, S. Banerjee, and A. S. Maddox, Mechanical and biochemical feedback combine to generate complex contractile oscillations in cytokinesis, Curr. Biol. **34**, 3201 (2024).

[65] A.-C. Reymann, F. Staniscia, A. Erzberger, G. Salbreux, and S. W. Grill, Cortical flow aligns actin filaments to form a furrow eLife **5**, e17807 (2016).

[66] C. M. Bishop and N. M. Nasrabadi, *Pattern Recognition and Machine Learning* (Springer, New York, NY, 2006), Vol. 4.

[67] C. W. Groetsch and C. W. Groetsch, *Inverse Problems in the Mathematical Sciences* (Springer, New York, NY, 1993), Vol. 52.

[68] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems* (Springer Science & Business Media, New York, NY, 1996), Vol. 375.

[69] M. O'Callaghan, K. S. Mandel, and G. Gilmore, Misspecification-robust amortised simulation-based inference using variational methods, arXiv:2509.05724.

[70] B. M. Burkel, G. von Dassow, and W. M. Bement, Versatile fluorescent probes for actin filaments based on the actin-binding domain of utrophin, Cell Motil. Cytoskeleton **64**, 822 (2007).

[71] Y. Xu, J. Shen, H. Ruan, X. Qu, Y. Li, Y. Wang, P. Li, R. Yi, H. Ren, Y. Zhang, *et al.*, A RhoGAP controls apical actin polymerization by inhibiting formin in *Arabidopsis* pollen tubes, Curr. Biol. **34**, 5040 (2024).

[72] D. D. Tang and S. J. Gunst, The small GTPase Cdc42 regulates actin polymerization and tension development during contractile stimulation of smooth muscle, J. Biol. Chem. **279**, 51722 (2004).

[73] M.-F. Carlier, A. Ducruix, and D. Pantaloni, Signalling to actin: The Cdc42-N-WASP-Arp2/3 connection, Chem. Biol. **6**, R235 (1999).

[74] T. A. Burke, J. R. Christensen, E. Barone, C. Suarez, V. Sirotkin, and D. R. Kovar, Homeostatic actin cytoskeleton networks are regulated by assembly factor competition for monomers, Curr. Biol. **24**, 579 (2014).

[75] C. Suarez and D. R. Kovar, Internetwork competition for monomers governs actin cytoskeleton organization, Nat. Rev. Mol. Cell Biol. **17**, 799 (2016).

[76] L. Mohapatra, T. J. Lagny, D. Harbage, P. R. Jelenkovic, and J. Kondev, The limiting-pool mechanism fails to control the size of multiple organelles, Cell Syst. **4**, 559 (2017).

[77] J. Wildberger, M. Dax, S. Buchholz, S. Green, J. H. Macke, and B. Schölkopf, Flow matching for scalable simulation-based inference, Adv. Neural Inf. Process. Syst. **36**, 16837 (2023).

[78] G. Papamakarios, E. Nalisnick, D. J. Rezende, S. Mohamed, and B. Lakshminarayanan, Normalizing flows for probabilistic modeling and inference, Mach. Learn. Res. **22**, 1 (2021).

[79] H. Du, C. Krause, V. Mikuni, B. Nachman, I. Pang, and D. Shih, Unifying simulation and inference with normalizing flows, Phys. Rev. D **111**, 076004 (2025).

[80] I. J. A. Bongalonta, A. R. Dinner, and A. Tokmakoff, Using multiple isotope-labeled infrared spectra for the structural characterization of an intrinsically disordered peptide, J. Phys. Chem. B **129**, 11717 (2025).

[81] A. S. Maddox, L. Lewellyn, A. Desai, and K. Oegema, Anillin and the septins promote asymmetric ingression of the cytokinetic furrow, Dev. Cell **12**, 827 (2007).

[82] M. Lebedev, F.-Y. Chan, E. Rackles, J. Bellessem, T. Mikeladze-Dvali, A. Xavier Carvalho, and E. Zanin, Anillin mediates unilateral furrowing during cytokinesis by limiting RhoA binding to its effectors, J. Cell Biol. **224**, e202405182 (2025).

[83] E. Munro, J. Nance, and J. R. Priess, Cortical flows powered by asymmetrical contraction transport PAR proteins to establish and maintain anterior-posterior polarity in the early *C. elegans* embryo, Dev. Cell **7**, 413 (2004).

[84] Y. Chang and D. J. Dickinson, A particle size threshold governs diffusion and segregation of PAR-3 during cell polarization, Cell Rep. **39**, 110652 (2022).

[85] S. Schonegg and A. A. Hyman, CDC-42 and RHO-1 coordinate acto-myosin contractility and PAR protein localization during polarity establishment in *C. elegans* embryos, Development **133**, 3507 (2006).

[86] E. Zanin, A. Desai, I. Poser, Y. Toyoda, C. Andree, C. Moebius, M. Bickle, B. Conradt, A. Piekny, and K. Oegema, A conserved RhoGAP limits M phase contractility and coordinates with microtubule asters to confine RhoA during cytokinesis, Dev. Cell **26**, 496 (2013).

[87] P. Gross, K. V. Kumar, N. W. Goehring, J. S. Bois, C. Hoege, F. Jülicher, and S. W. Grill, Guiding self-organized pattern formation in cell polarity establishment, Nat. Phys. **15**, 293 (2019).

[88] J. C. Herron, S. Hu, B. Liu, T. Watanabe, K. M. Hahn, and T. C. Elston, Spatial models of pattern formation during phagocytosis, PLoS Comput. Biol. **18**, e1010092 (2022).

[89] S. A. Wickström and C. M. Niessen, Cell adhesion and mechanics as drivers of tissue organization and differentiation: Local cues for large scale organization, Curr. Opin. Cell Biol. **54**, 89 (2018).

[90] Y. Sasai, Cytosystems dynamics in self-organization of tissue architecture, Nature (London) **493**, 318 (2013).

[91] J. A. Bonachela, R. M. Pringle, E. Sheffer, T. C. Coverdale, J. A. Guyton, K. K. Caylor, S. A. Levin, and C. E. Tarnita, Termite mounds can increase the robustness of dryland ecosystems to climatic change, Science **347**, 651 (2015).

[92] RhoActinRD: GitHub repository for activator-inhibitor modeling, 2026, https://github.com/omaxian/RhoActinRD.

[93] D. Chomchai, M. Leda, A. Golding, G. Von Dassow, W. M. Bement, and A. B. Goryachev, Rho GTPase dynamics distinguish between models of cortical excitability, Curr. Biol. **35**, 1414 (2025).

[94] C. J. Staiger, M. B. Sheahan, P. Khurana, X. Wang, D. W. McCurdy, and L. Blanchoin, Actin filament dynamics are dominated by rapid growth and severing activity in the *Arabidopsis* cortical array, J. Cell Biol. **184**, 269 (2009).

[95] Y. Chen, D. Zhang, M. U. Gutmann, A. Courville, and Z. Zhu, Neural approximate sufficient statistics for implicit models, in *9th International Conference on Learning Representations* (International Conference on Learning Representations, Appleton, WI, 2021).

[96] Y. Chen, M. U. Gutmann, and A. Weller, Is learning summary statistics necessary for likelihood-free inference?, in *International Conference on Machine Learning* (PMLR, Cambridge, MA, 2023), pp. 4529–4544.

[97] M. Sainsbury-Dale, A. Zammit-Mangion, and R. Huser, Likelihood-free parameter estimation with neural Bayes estimators, Am Stat. **78**, 1 (2024).